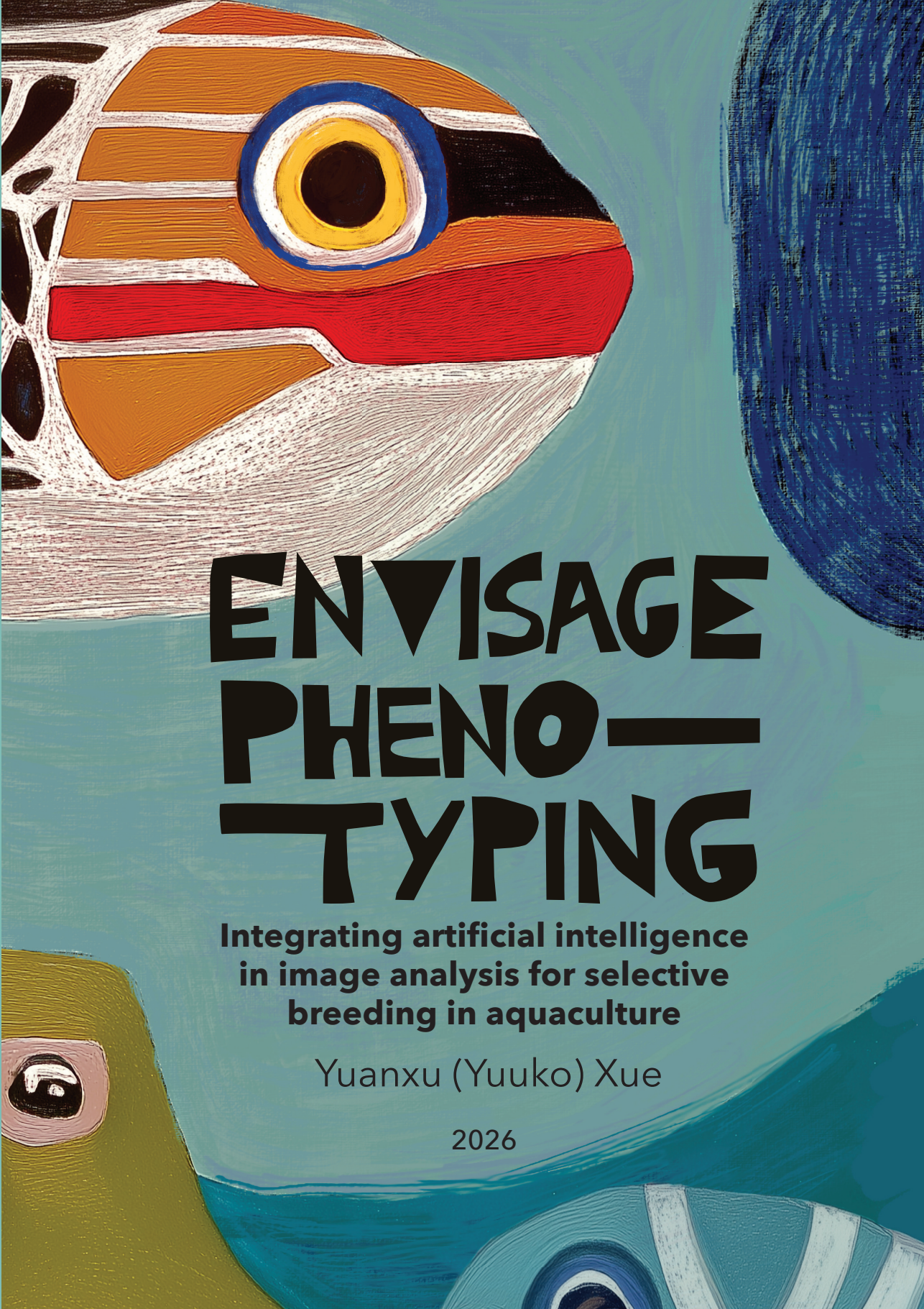


Integrating artificial intelligence in image analysis for selective breeding in aquaculture

Yuanxu (Yuuko) Xue

2026

Propositions

1. Acquiring images of selection candidates is essential for breeding programs in aquaculture.
   (this thesis)

2. Genetic analysis is necessary for the validation of phenotyping tools.
   (this thesis)

3. Involvement in data analysis requires participation in data collection.

4. The most important aspect of interdisciplinary communication is shared evaluation metrics.

5. LinkedIn is toxic for networking.

6. Tofu is not a meat replacer.

Propositions belonging to the thesis, entitled

Envisage phenotyping: integrating artificial intelligence in image analysis for selective breeding in aquaculture

Yuanxu (Yuuko) Xue

Wageningen,
15th January 2026

Envisage phenotyping

Integrating artificial intelligence in image analysis for selective breeding in aquaculture

Yuanxu (Yuuko) Xue

Thesis committee

Promotor

Prof. Dr *ir*. H. Bovenhuis

Professor in Animal Breeding and Genomics

Wageningen University & Research

Co-promotor

Dr *ir*. J.W.M. Bastiaansen

Research Associate, Animal Breeding and Genomics

Wageningen University & Research

Other members

Prof. Dr R. da Silva Torres, Wageningen University & Research

Dr J.J. Poos, Wageningen University & Research

Dr W.J. dos Santos Silva, Utrecht University

Dr G.F. Difford, NMBU, Ås, Norway

This research was conducted under the auspices of the Graduate School of Wageningen Institute of Animal Sciences (WIAS), the Netherlands

Envisage phenotyping

Integrating artificial intelligence in image analysis for selective breeding in aquaculture

Yuanxu (Yuuko) Xue

Thesis

Submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr C. Kroeze,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Thursday 15 January 2026
at 3.30 p.m. in the Omnia Auditorium.

Yuanxu (Yuuko) Xue

Envisage phenotyping: integrating artificial intelligence in image analysis for selective breeding in aquaculture

190 pages

PhD thesis, Wageningen University, Wageningen, the Netherlands (2026)

With references, with summary in English

DOI: http://doi.org/10.18174/681487

Sometimes I think, but then I forget.

Abstract

Xue, Y. (2026). Envisage phenotyping: integrating artificial intelligence in image analysis for selective breeding in aquaculture. PhD thesis, Wageningen University, the Netherlands.

Automated phenotyping is a rapidly evolving interdisciplinary field focused on high-throughput, non-invasive phenotype collection to support breeding and management decisions in aquaculture. Current development primarily targets production traits, leaving a gap in traits linked with metabolism, health and physiological well-being. These traits are complex and difficult to measure, but advancements in AI-based prediction models combined with imaging technologies provide opportunities. This thesis used fish image data collected in breeding facilities, to broaden the scope of predicted phenotypes that contribute to the breeding goals for these species.

For production traits, a structural framework was introduced to determine optimal image-based prediction strategies, demonstrated in gilthead seabream (*Sparus aurata*). The framework captures both simple and high-dimensional relationships between images and target traits. Interpretation of the model was provided by visualizing predictive imaginal features.

To dissect the genetic basis of swimming performance in rainbow trout (*Oncorhynchus mykiss*), a novel workflow identified accessible and heritable morphological indicators. Among fish of the same weight, relatively larger and broader epaxial muscles, larger heads, and smaller caudal fins were associated with poorer swimming performance. This study illustrates how images and AI can be interpreted physiologically using Explainable AI.

Individual re-identification is essential for longitudinal traits. Although image-based re-identification is promising, it remains infeasible for Atlantic salmon (*Salmo salar*) under realistic aquaculture conditions due to variation in the phenotyping environment, especially lighting, and limited phenotypic stability. In search for solutions, a review was conducted on the current state of vision-based re-identification in aquaculture. Limitations in data acquisition, data analysis, and practical implementation were identified. Without high-tech data acquisition, multi-dimensional data quality frameworks, and benchmarking methods, vision-based re-identification is found too immature to replace PIT tags. Future progress depends on dataset standardization, methodological refinement, and strong alignment with aquaculture systems.

The objectivity of qualitative phenotypic measurements using images was improved and applied to body shape in gilthead seabream (*Sparus aurata*). Two moderately heritable traits, “Distance to the Best” (DtB) and “Distance to the Worst” (DtW), were derived for the deviation of each fish from an ideal or undesirable shape. This

approach is objective, automated, visually intuitive, and flexible for updating target shapes.

This thesis emphasizes that successful integration of AI and image analysis requires an understanding of the entire breeding program, not just phenotypic data collection and analysis. This thesis contributes advanced phenotyping and decision-support AI tools to the ongoing integration of emerging AI trends to improve animal production, health, and welfare.

Contents

1. General introduction

The observation of animals has always been a remarkable source of inspiration for humanity. One of the most classic debates in Taoism emerged from observing fish in a river:

> *Zhuangzi and Huizi were crossing the Hao River by the dam. Zhuangzi said: See how free the fishes leap and dart: that is their happiness. Huizi replied: Since you are not a fish, how do you know what makes fishes happy?*
>
> *Zhuangzi said: Since you are not I, how can you possibly know that I do not know what makes fishes happy? Huizi argued: If I, not being you, cannot know what you know, it follows that you, not being a fish, cannot know what they know.*
>
> *Zhuangzi said: Let us get back to the original question. From the terms of your question, you evidently know I know what makes fishes happy. I know the joy of fishes in the river through my own joy, as I go walking along the same river.*
>
> *-- Zhuangzi, Chapter 17, Autumn Floods (476–221 BC), based on translation by Merton* (1969)

This debate highlights the consciousness of animals and whether it can be perceived by humans. While Zhuangzi's speculation on the happiness of fish reflects a subjective human perspective, it stems from a momentary observation of animals in their natural environment.

Over thousands of years, through domestication and cohabitation, more comprehensive observations of animals have been documented, often accompanied by philosophical insights. Nietzsche, for example, reflected on the contentment of animals through their behavioral pattern:

> *Observe the herd which is razing beside you. It does not know what yesterday or today is. It springs around, eats, rests, digests, jumps up again, and so from morning to night and from day to day, with its likes and dislikes closely tied to the peg of the moment, and thus neither melancholy nor weary … One day the man demands of the beast: "Why do you not talk to me about your happiness and only gaze at me?" The beast wants to answer, too, and say: "That comes about because I always immediately forget what I wanted to say."*

Nietzsche interpreted the forgetfulness of cattle (*the beast*) as a blessing, enabling them to find happiness in the repetition of daily life. As Nietzsche's attempt of verbal communication with the herd failed, he could only rely on his own reasoning to infer the herd's inconceivable experiences.

> *But by then the beast has already forgotten this reply and remains silent, so that the man wonders on once more* (Nietzsche, 1874).

Despite their seeming unknowability, the thoughts and feelings of animals have long been a topic of curiosity and importance. Animal sciences germinated and flourished, and humans grew into the role of caretakers, responsible for ensuring the welfare of farmed animals. This responsibility includes fulfilling animals' needs by providing conditions that allow them to thrive in optimal habitats. Understanding and nurturing animals requires *scientific evidence* that reflects their well-being, shifting from subjective human projections of emotion and experience to objective, observable indicators related to animals' behaviors and health.

1.1 Phenotyping

The word phenotype originates from the Ancient Greek φαίνω (to appear, show) and τύπος (mark, type), referring to the observable and measurable characteristics of an organism. Phenotyping, therefore, is to observe and measure such exhibited characteristics. Scientifically, phenotyping is the assessment of expressed traits, influenced by many factors such as genetics, the environment, and the interaction between an animal's nature and nurture.

Phenotyping has a long history in animal breeding and management. As early as 1859, Darwin described the phenotyping process in the selection of merino sheep:

> *the sheep are placed on a table and are studied, like a picture by a connoisseur; this is done three times at intervals of months, and the sheep are each time marked and classed, so that the very best may ultimately be selected for breeding* (Darwin, 1859).

Traditionally, phenotyping was carried out by experts (*connoisseurs*) who manually measured and recorded traits (Fig. 1.1, left). In dairy cattle, for example, common recorded phenotypes include morphological traits such as body weight and height; production traits such as milk yield and fertility; behavioral traits such as lameness; welfare traits such as mortality; and color pattern for individual identification. These traits inform decisions on breeding selection, herd management, and veterinary interventions (Fig. 1.1, right).

Figure 1.1 Left: A cow was extensively inspected and assessed by the inspector for studbook registration, 1955 *(de la Bruhèze, Lintsen, Rip, & Schot, 2000)*. **Right:** Calf sketches from the breeding register, including the spot pattern of both sides of a calf, together with data such as date of birth, parents, owner, breed, health condition. The sketch was used as the passport of the calf to prevent and trace the outbreak of diseases. Source: https://www.deverhalenvangroningen.nl/alle-verhalen/er-gaat-geen-koe-op-reis-zonder-schets-en-tbc-bewij & https://www.estersheem.nl

The relative importance of traits has often been tied to economic returns. Historically, traits with clear economic benefits were prioritized, such as milk yield in dairy cattle, egg number in laying hens, carcass weight in pigs, and fillet yield in fish. However, a narrow focus on maximizing production has sometimes led to undesirable consequences (D'Eath, et al., 2010; Rauw, 1998), urging phenotyping to expand its scope to include more health and welfare-related traits (Hu, Do, Gray, & Miar, 2020; Weimer, Mauromoustakos, Karcher, & Erasmus, 2020).

Assessing animal wellbeing is rarely a one-time task. Phenotyping is often repeated throughout an animal's lifetime to monitor changes in traits and ensure accurate assessments. While production traits are straightforward to acquire, animals nowadays are evaluated for more complex traits. For instance, measurements on physiology and metabolism can be invasive or post slaughter, such as white blood cell count in chicken (Ribeiro, et al., 2024), feed digestion via ruminal cannulation in cattle (Castillo & Hernández, 2021), and visceral lipid weight in fish (Kause, Paananen, Ritola, & Koskinen, 2007). Such phenotyping requires a high level of expertise in animal health and welfare, while repeated assessments increase the demand for manual labor. These challenges highlight the urgent need for more advanced and innovative phenotyping approaches (Neethirajan & Kemp, 2021).

1.2 Automated phenotyping

The development of automatic and robotic systems has enabled high-throughput data collection in livestock and aquaculture, generating large volumes of objective information on animal productivity and behaviors. These datasets are used to define and derive traits relevant for selective breeding, production, and management decisions. Such process is referred to as **automated phenotyping**.

For example, the collection of daily, individual milking records in dairy cattle has been made possible through the widespread adoption of automated milking systems,

with the main aim of monitoring milk production for management. With advances in statistical methods, these production and behavioral data can now generate additional insights, such as earlier detection of diseases, especially mastitis, monitoring of physical activity (Ozella, Brotto Rebuli, Forte, & Giacobini, 2023), and analysis of social dynamic of dairy cattle (Fadul-Pacheco, Liou, Reinemann, & Cabrera, 2021). Similarly, automated feeders in pig pens can record daily visits and feed intake, enabling improved estimations of feed efficiency (Garrido-Izard, Correa, Requejo, Villarroel, & Diezma, 2022) and optimization of feeding strategies (Marcon, Brossard, & Quiniou, 2015).

1.2.1 Data collection

In automatic systems, high-throughput data is primarily collected through sensors and cameras. Sensors are commonly employed to monitor environmental conditions in animal facilities, such as ammonia emissions in poultry houses and methane emission in cattle barns (Li, et al., 2023; Aldhafeeri, Tran, Vrolyk, Pope, & Fowler, 2020). On top of that, for individual-level measurements in livestock, sensors can be implemented as wearable devices to minimize disruption to the animals' daily activities. In aquaculture, sensors are typically used as implants to monitor complex traits, including heart rate, respiratory frequency, and other metabolic or health indicators (Abbink, Palstra, Agbeti, Lembo, & Komen, 2022; Martos-Sitcha, et al., 2019; Wu, et al., 2015). However, the application of sensors on individual level is often associated with high costs and maintenance, with many mainly suited for research and development purposes (Ellen, et al., 2019). Wearable sensors are also less suitable for small animals with high density, such as poultry compared to cattle. And for fish, the implantation and retrieval of sensors require invasive and often delicate procedures, which can limit their practicality.

In contrast, cameras provide a non-invasive, remote, and cost-effective approach of data collection that is broadly applicable across species, making them particularly attractive for both livestock and aquaculture systems.

Image data captured by cameras can take various forms, each offering unique advantages for phenotyping application. Standard RGB images consist of three-color channels (Red, Green, Blue) with the final color of each pixel defined by the intensity of these contributing colors. Videos are continuous sequences of images, allowing the capture of changes and movements. Three-dimensional (3D) images contain pixel-wise depth measurements that capture the volume of the object (Fig. 1.2). Spectral imaging extends beyond the visible RGB channels into other regions of the electromagnetic spectrum to reveal details that are invisible to the human eye.

Figure 1.2 Different forms of rainbow trout (*Oncorhynchus mykiss*) images. Left: Depth image, where each pixel contains distance from the camera to the fish. Middle: RGB image, a visual representation of the fish. Right: 3D hologram constructed with depth and RGB images.

To process high-throughput data, automated phenotyping also requires advanced data analysis and statistical methods. The development of artificial intelligence (AI) has provided powerful analytical tools to extract meaningful trait information from large and complex datasets (Pérez-Enciso & Steibel, 2021).

1.2.2 Machine learning

A key component of artificial intelligence applied in automated phenotyping is machine learning, including neural networks and deep learning. There are several learning paradigms, with the most common being supervised learning, unsupervised learning and transfer learning (Mahadevkar, et al., 2022).

Supervised learning involves training algorithms on labelled datasets, where the desired output is predefined and annotated by humans. Once trained, these algorithms can automatically determine the expected outputs for new, unseen data. Supervised learning is the most widely applied in animal research, with examples including tracking and behavior detection (Ariza-Sentís, Vélez, Martínez-Peña, Baja, & Valente, 2024; Liu, Li, Liu, Li, & Yue, 2024).

In contrast, unsupervised learning is particularly valuable in uncovering patterns in large and complex datasets without annotation. In animal studies, unsupervised learning is commonly applied in species classification (Guerrero, Bedoya, López, Daza, & Isaza, 2023), individual identification (Hossain, et al., 2022), and behavior patterns identification (McVey, Hsieh, Manriquez, Pinedo, & Horback, 2023), although it may not outperform supervised methods in similar tasks (Manohar, Kumar, & Kumar, 2016).

Transfer learning leverages the knowledge gained from one task to improve performance on another. It is becoming increasingly popular in animal studies, where dedicated datasets are often scarce, but large datasets in other fields are abundant and publicly available. For instance, a model trained to recognize human faces can transfer its understanding of basic facial structures, such as the relative positions of eyes, nose and mouth, to a task of identifying individual cows (Ruchay, Kolpakov, Guo, & Pezzuolo, 2024).

The advancement of machine learning provides innovative solutions in gaining detailed knowledge of individual animals, with an increasing focus on their health and welfare (Vinci, 2025; García, Aguilar, Toro, Pinto, & Rodríguez, 2020). These

developments have greatly contributed to the growth of precision farming, an interdisciplinary field that uses automated phenotyping to support data-driven decision-making in animal breeding and management (Morrone, Dimauro, Gambella, & Cappai, 2022; Norton, Chen, Larsen, & Berckmans, 2019).

1.2.3 Image analysis and computer vision

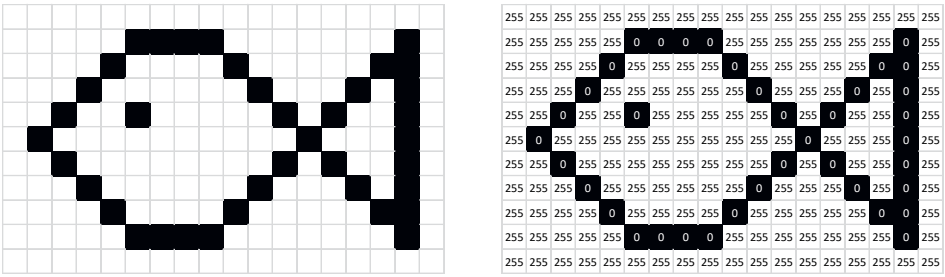


Figure 1.3 Illustration of hierarchy of process. **Left**: An 18x12 input image. **Right**: each pixel value of the input image. Low level process will show each pixel value in the image. Middle level process will output the coordinate of the black pixel as contour detection. High level process can recognize that the shape resembles a fish.

As mentioned earlier, image data plays a crucial role in precision farming, and artificial intelligence methods applied to image analysis are often referred to as computer vision. In practice, there are no clear boundaries between image processing, image analysis, and computer vision. Computer vision explicitly incorporates artificial intelligence and focuses on enabling machines to analyze, interpret, and understand visual information. Similarly, image analysis and image processing can also benefit from artificial intelligence to improve performance (Zhang & Dahu, 2019). Overall, computer vision, image analysis and image processing are interconnected yet hierarchical, organized by the level of processing involved and the increasing complexity of their outputs (Fig. 1.4) (Gonzales & Wintz, 1987).

This thesis emphasizes the involvement of artificial intelligence (AI) in image-based automated phenotyping, focusing on the extraction and interpretation of complex image features to produce traits and indicators relevant for animal breeding, through methodologies that cover both image analysis and computer vision (Fig. 1.4).

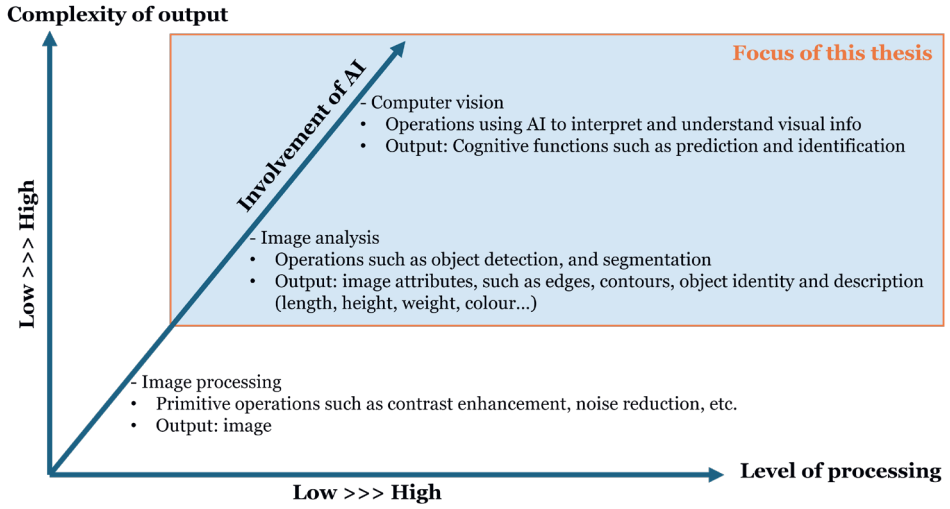


Figure 1.4 The relationship between image processing, image analysis and computer vision, with the scope of this thesis.

1.2.4 Phenotyping and computer vision in animal breeding

Animal breeding plays a crucial role in advancing animal production systems by exploiting genetic merit and managing genetic diversity through structured breeding programs (Granados Chapatte, 2021). Automated phenotyping has further contributed to this landscape by enabling the collection of high-throughput, objective data on individual animals.

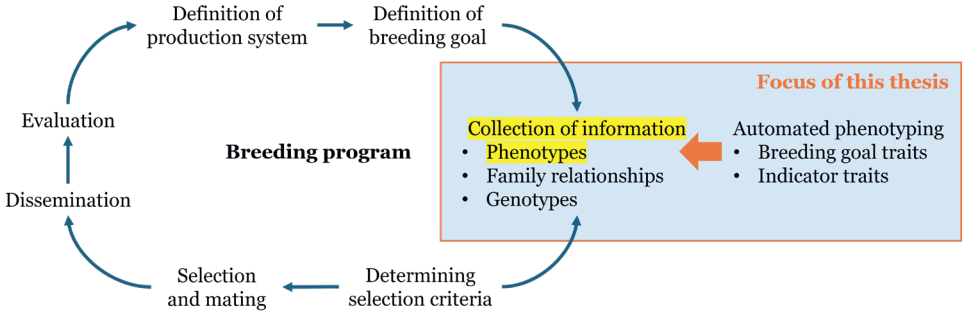


Figure 1.5 The structure of a breeding program *(Oldenbroek & Calus, 2024)* and the main contribution of automated phenotyping highlighted with yellow.

Automated phenotyping contributes to measuring both breeding goal traits and indicator traits in the breeding program (Fig. 1.5). It enables large-scale trait measurements, often through non-invasive, high-throughput methods (Difford, Boison, Khaw, & Gjerde, 2020). This is particularly beneficial for production traits such as body weight and carcass weight, opening the possibility to obtain very high selection intensities. Additionally, certain traits are only measurable invasively or

post-slaughter. Automated phenotyping can provide non-invasive, repeated measurements of these traits directly on the selection candidates (Ahmad, et al., 2025; Rezende, Perazza, Freitas, Hallerman, & Hilsdorf, 2023; Elalfy, et al., 2021).

Another major contribution of automated phenotyping lies in acquiring indicator traits, especially for complex breeding goals related to animal health and welfare. Recent technologies, such as pose estimation, can decompose behavior into fine-scale body movements by tracking joint positions, offering new insights into activity patterns and stress responses (Fodor, et al., 2023; Jiang, Lee, Teotia, & Ostadabbas, 2022). Furthermore, automated phenotyping can help recognizing and tracking individuals over time (Meng, et al., 2025) to potentially facilitate the collection of longitudinal data, such as growth curves. Automated phenotyping helps incorporate indicator traits with high genetic correlations to breeding goals to accelerate genetic gain (Billah, et al., 2025; Brito, et al., 2020)

1.3 Image-based automated phenotyping in aqua breeding

While livestock and their products remain the primary source of animal-based protein for the global population, aquaculture is the fastest-growing food production sector (Mair, Halwart, Derun, & Costa-Pierce, 2023). Worldwide, aquatic foods provide 15% of animal proteins (FAO, 2025). They account for at least 20% of the average per capita intake of animal protein for 3.2 billion people (Fig. 1.6), and are an essential source of key nutrients, including omega-3 fatty acids, minerals, and vitamins (FAO, 2024; Boyd, McNevin, & Davis, 2022).

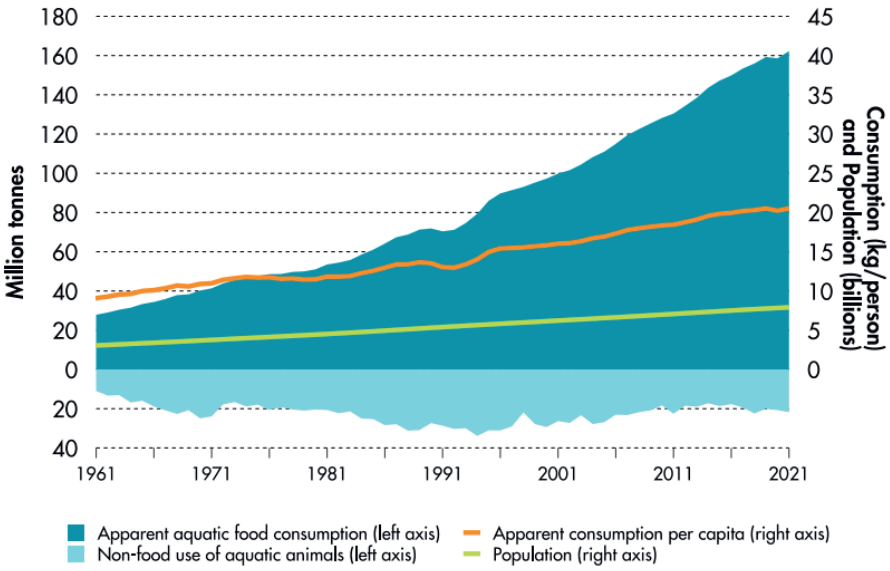


Figure 1.6 Aquaculture production & fisheries of aquatic animals: utilization and apparent consumption (FAO, 2025).

Genetic improvement is one of the important reasons for the growth of aquaculture production (Scanes, 2018). Still, only a small proportion of aquaculture production is based on genetically improved stocks (Gulzari B. , 2023; Gjedrem, Robinson, & Rye, 2012) and relatively few distinct strains or varieties are available per species in aquaculture compared to that of livestock (FAO, 2025), leaving considerable room for improvement by means of selective breeding (Boudry, et al., 2021).

An emerging challenge in aqua breeding and genetics is to safeguard production despite rapid climate change (Esposito, Carputo, Cardi, & Tripodi, 2019). This requires breeding programs to boost production by optimizing selective breeding for performance under variable conditions. Most fish breeding programs target economic production traits such as growth rate (GR) using direct measurements. However, body composition is an important component of GR and a promising predictor trait for feed efficiency (Knap & Kause, 2018). Body composition includes protein and fat content (Breck, 2014), typically measured invasively, and carcass traits such as carcass weight, offal weight, fillet yield, and abdominal fat measured post-slaughter (Vallecillos, et al., 2021). With increasing requirements to improve wellbeing, breeding programs must incorporate traits related to health and physiological wellbeing (Seibel, Weirup, & Schulz, 2020). Welfare indicators include stress resilience and susceptibility to diseases (Barreto, Rey Planellas, Yang, Phillips, & Descovich, 2022), which are measurable, but only invasively and lag the real time development of the animal. Future data collection strategies should target more direct measurements of complex phenotypes, such as disease resistance, but at the same time societal pressures make collecting such data more controversial. Such challenges can be answered with advancement in automated phenotyping (Mandal & Ghosh, 2024; Wang, et al., 2021).

To date, automated phenotyping in aquaculture has targeted image analysis for species identification (Miyazono & Saitoh, 2017), physical parameters and fillet quality. For instance, automated phenotyping can replace labor-intensive measurements such as body weights and other production parameters such as body length, width, and dimensions of body areas (Prchal, et al., 2020; Xue, 2019; Al-Jubouri, Al-Nuaimy, Al-Taee, & Young, 2017), and enable predictions for traits on whole fish that are otherwise measured invasively, such as fillet yield (Gulzari B. , 2023; Grassi, Casiraghi, & Alamprese, 2018), carcass yield (Vandeputte, et al., 2017) and feed intake (Difford, et al., 2023).

Challenges arise when extending the concept of precision farming to aquaculture (Føre & Alver, 2023), as the biology and housing systems of aquaculture species differ from those of terrestrial livestock. Breeding programs in aquaculture typically involve larger numbers of selection candidates than livestock systems, creating difficulties for both individual and group phenotyping. Furthermore, the unique biology and aquatic environments of fish result in distinct metabolism and behavior, which require alternative approaches for defining health and welfare indicators compared to those for terrestrial livestock species such as ruminants (cattle, sheep,

and goats) (Tzanidakis, Tzamaloukas, Simitzis, & Panagakis, 2023; Groher, Heitkämper, & Umstätter, 2020) and non-ruminants (broilers and laying hens) (Olejnik, Popiela, & Opaliński, 2022; Li, Ren, Li, & Zeng, 2020). Because fish are housed underwater, they are less accessible for direct observation by humans and often require more intensive handling during phenotyping. These factors pose unique challenges and opportunities for image-based automated phenotyping in aqua breeding, to develop tailored approaches supported by advancements in artificial intelligence.

1.4 Aim and overview of this thesis

In this thesis I study the integration of artificial intelligence in image-based automated phenotyping aimed at selective breeding in aquaculture. Image data was used that was collected on three common aquacultural species: gilthead seabream (*Sparus aurata*), Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*). These fish were selection candidates with pedigree information. All image data collection took place in breeding facilities.

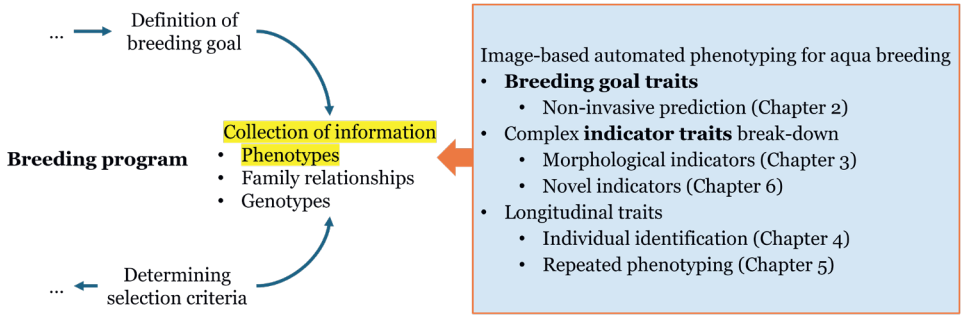


Figure 1.7 The structure of all chapters in this thesis.

To extract relevant information from image data for breeding applications, the objectives of automated phenotyping can be divided according to the types of traits it aims to deliver (Fig. 1.7). For some breeding goal traits, especially those related to production, images can be used as non-invasive alternatives for manual measurements. **Chapter 2** introduces a structural framework, including both numerical methods and deep learning, for optimizing the use of image data in predicting invasive production traits, illustrated through case studies on seabream.

As breeding goals become more complex, such as improving health and welfare, indicator traits are often employed by measuring aspects of metabolism and physiology. However, these indicators can also be challenging to measure. **Chapter 3** addresses this issue with an AI-driven workflow, to identify accessible and heritable morphological indicators that account for a proportion of the variance in metabolic traits, with a specific focus on swimming performance in rainbow trout. Defining indicator traits often involves manual annotation, and many of these traits are recorded categorically using scales derived from expert judgment. In **Chapter 6**

I use shape analysis in seabream as an example to propose an automated, image-based transformation to increase the resolution of shape-related traits, and to improve the genetic correlation assessment between shape and production trait such as body weight.

Finally, breeding value estimation benefits greatly from repeated measurements, and certain traits are longitudinal. **Chapter 4** explores the potential of image-based individual identification for tracking animals over time within breeding programs, while **Chapter 5** critically reviews previous research on re-identification. Based on existing research, I identify key challenges and propose an aquaculture-centered approach to guide future studies by improving data acquisition, enhancing data analysis, and emphasizing practical implementation.

The general discussion (**Chapter 7**) synthesizes all results and discussions of this thesis, including the challenges identified and the solutions proposed. In addition, the experience and knowledge gained from this work are examined both within the context of aquaculture and from the broader perspective of livestock systems. Together, these insights offer a vision for the future integration of image-based automated phenotyping with artificial intelligence in animal breeding.

2. An analytical framework to predict slaughter traits from images in fish

Yuuko Xue[1], John W.M. Bastiaansen[1], Haris Ahmad Khan[2], Hans Komen[1]

[1]Department of Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands

[2]Farm Technology Group, Wageningen University & Research, P.O. Box 16, 6700 AA Wageningen, The Netherlands

This chapter is based on:

Xue, Y., Bastiaansen, J. W., Khan, H. A., & Komen, H. (2023). An analytical framework to predict slaughter traits from images in fish. *Aquaculture*, *566*, 739175

Abstract

Accurate measurements of breeding traits on individuals are critical in aquaculture for obtaining breeding values and tracking the progress of the breeding program. Modern breeding programs prioritize not only production traits, but also complex traits related to production, product quality, body composition, disease resistance, and fish health, such as slaughter traits. Slaughter traits can be selected indirectly and incorporated into breeding programs. Indirect selection is cost-effective, but there is often little genetic correlation between measured and target traits. Accurate phenotypic prediction of the target traits using modern phenotyping technology can be game-changing in indirect selection. This paper proposes an analytical framework for predicting slaughter traits using images. The framework demonstrated that using images in addition to body weight improved fat percentage prediction accuracy from 0.4 to 0.7 when compared to a model that only used body weight and its numerical derivations. The framework also allowed for the interpretation of the prediction by providing imaginal features. In the case study, the dorsal side, the upper edge of the pectoral fin, and operculum edge were discovered to be the three regions on seabream that have properties that are negatively correlated with fillet fat percentage. The framework showed that both body weight and visceral weight are highly correlated with total fish body area. The framework also revealed that the lower edge of the pectoral fin, operculum edge, and anal fin are the regions with properties that explain variation in the visceral percentage. Future research will be required to segment and quantify each predictive imaginal feature to calculate its heritability. The framework can potentially predict other harvest, post-slaughter, and metabolic traits for aquacultural study.

Key words: Novel phenotyping; Machine vision; Non-invasive prediction; multi-input framework; Machine learning

2.1 Introduction

Accurate phenotyping is essential for selective breeding to be successful (Gjedrem & Robinson, 2014). Different traits, however, necessitate different phenotyping methods (Fu & Yuna, 2022). In the current practice, fish has to be sacrificed to determine a particular set of traits. These are primarily quality traits such as fillet yield and visceral weight (Gjedrem T. , 1997). Quality traits, as well as appearance (Salerno, Berlino, Mangano, & Sarà, 2021; Colihueque & Araneda, 2014; Violle, et al., 2007), metabolic (Metcalfe, 2016) and behavioral (Carter, Feeney, Marshall, Cowlishaw, & Heinsohn, 2013) traits, are regarded as complex because they involve many genetic and environmental factors, and their interactions (Gjedrem & Robinson, 2014). Direct measurements of these traits are typically invasive and stressful for fish. Stressful practices may have a negative impact on both product quality and animal welfare (Poli, Parisi, Scappini, & Zampacavallo, 2005). Phenotyping methods should accommodate the increasing diversity of traits emphasized by modern aquaculture breeding programs.

Sibling selection and indirect selection can be used in breeding programs to integrate traits that are invasive or difficult to measure on selection candidates. Siblings are used to provide data to estimate breeding values for the selected candidates on slaughter and survival traits (Gjedrem T. , 2010). Sibling selection is often combined with genomic selection (GS) to include the within-family genetic variation. However, the implementation of GS also requires large-scale phenotyping which can be very expensive for some species in aquaculture. Indirect selection, on the other hand, selects on correlated traits to improve the traits of interest. These correlated traits can be measured on the selection of candidates and can be used as predictors for the target traits. Although introduced a long time ago (Eknath & Doyle, 1985; Doyle, Singholka, & New, 1983), the indirect selection is not favored by breeders due to the typically low genetic correlations between the predictor and target traits. However, the strength of the genetic correlation can be improved by choosing a more appropriate phenotype as the predictor traits (Gulzari, Mencarelli, Roozeboom, Komen, & Bastiaansen, 2022; Kause, Paananen, Ritola, & Koskinen, 2007). The success of a cost-effective, indirect selection strategy, therefore, depends on the accuracy of the phenotypic prediction of the target traits using the predictor traits.

Emerging technologies enable advanced, accurate and automated phenotyping (Yue & Shen, 2022). For instance, computer vision and sensors allow high-throughput, non-invasive data collection on live fish (Saeed, Feng, Wang, Zhang, & Fu, 2022; Yang, et al., 2021; Saberioon, Gholizadeh, Cisar, Pautsina, & Urban, 2017). Based on the data collected, phenotypic prediction models are established using machine learning (ML) and deep learning (DL) approaches, targeting traits like body weight and carcass weight (Yang, et al., 2021; Fernandes, et al., 2020). Considering the success of large-scale phenotyping in plants (Yang, et al., 2020), there is a huge potential for the use of computer vision and machine learning techniques in aquaculture. The new phenotyping tool is expected to make it possible to evaluate

morphology and physiology across entire populations and throughout an individual's development. Slaughter traits are frequently used in practice to assess morphology and physiology. Accurate predictions of slaughter traits based on non-invasive traits can thus allow for individual, non-destructive, and repeated morphological and physiological assessments (Ventura, e Silva, Yáñez, & Brito, 2020). This study proposes an analytical framework for predicting slaughter traits using images. To best demonstrate the use of the framework, we predict three example traits: slaughter traits, fat percentage of the fillet, visceral weight, and visceral percentage of the fish. With the framework, we develop predictive models for slaughter traits by combining images and other non-invasive measurements. The framework can also extract specific imaginal features that contribute to the predictions, allowing for a better understanding of how such predictions work.

2.2 Materials and methods

The main contribution of this study is to propose an analytical framework to phenotype traits of interest by prediction. Once the prediction model has been established, it can be used for future phenotyping of the same trait.

2.2.1 Data preparation

To build a prediction model, the measured traits are divided into two categories: predictor trait(s) and target trait. Predictor traits, such as body weight, are acquired via non-invasive measurements. Images are also collected non-invasively and are listed as a separate input in the framework. The target trait is the trait of interest for prediction, mainly slaughter traits depending on the goal of the application. Predictor and target traits can both be categorical or continuous. The framework can only accept one target trait but multiple predictor traits.

Exploratory data analysis (EDA) is not included as a part of the framework but is a critical step in data preparation. Some of the framework's analytical methods are sensitive to outliers. When combined with domain knowledge, EDA can help identify and remove outliers caused by measurement errors during data collection. Descriptive statistics such as histograms and scatter plots were used and presented in this study.

After EDA and removal of outliers, the data is divided into three sets: training, validation, and test. Training and validation sets are combined into a single large training set when cross-validation is used. The three data sets are divided in a 3:1:1 ratio and used by the different framework sections (Fig. 2.1). Training and validation sets are used repeatedly, while the test set is used only once per method in model comparison, feature extraction, and visualization.

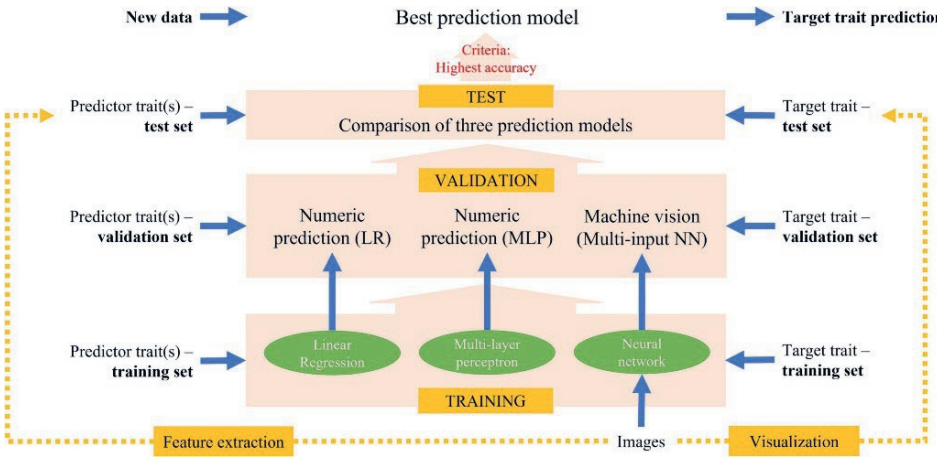


Figure 2.1 Analytical framework of traits prediction. To generate the best prediction model, data was processed in the order of training, validation, and testing. Each section uses a unique subset of the data, starting from the bottom of the framework. Three methods are included for training and validation. Linear regression and multi-layer perceptron both require only numeric inputs. Images are required for the neural network. The neural network also facilitates feature extraction and visualization once the prediction model is established for testing.

2.2.2 Overview of framework

Figure 2.1 presents the architecture of the framework.

Analytical methods

The framework includes three analytical methods: linear regression (LR), multi-layer perceptron (MLP), and neural network (NN). Each method generates a model that captures a different level of relationship that underlies the predictors and target trait.

Numeric models

Two of the three analytical methods, LR and MLP, requires only numeric inputs. LR assumes a linear relationship between the predictors and the target trait. In application, LR is combined with the K-fold cross-validation to avoid overfitting. The value of K is chosen from a range of 2 to 50 based on the highest adjusted coefficient of determination (adjusted R^2) on the validation set.

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

$$R^2_{adjusted} = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

Where R^2 is the sample coefficient of determination; N is sample size and p is the number of predictors.

Multi-layer perceptron (MLP) captures the more complex, non-linear relationship between the predictors and the target trait. Hornik et al. (1989) established that standard MLP can approximate any measurable function to any desired degree of accuracy. MLP is therefore considered as the universal approximator that can capture a wide variety of complex relationships. MLP consists of three types of layers: input, hidden, and output layers. The level of complexity that MLP can capture depends on the number of hidden layers. The number of the hidden layer is chosen from a range of 1 to 100. MLP constructed with the optimal number of layers produces the lowest mean squared error (MSE) and the highest adjusted R^2 between the predictors and the target trait on the validation set. At this point, adding more layers will result in approximately the same MSE and adjusted R^2 on the validation set. This is an important criterion to check if MLP truly captures the relationship between predictors and target traits. In this case, increasing model complexity will not further benefit the prediction but only increase the computing time and cost.

Machine vision

Convolutional neural network (CNN) is applied to incorporate images into prediction. The architecture of CNN was adapted from the work of the NVIDIA team (Bojarski, et al., 2016) on self-driving vehicles for the following reasons. First, most CNN applications aim for categorical prediction such as classification and identification, but like the study on self-driving vehicles, we aim for a continuous, numeric prediction from imaginal inputs. Second, we also want to avoid human-designated features. In the study of Bojarski et al. (2016) these features mainly included lane markings and other vehicles, while in this study these features are morphological measurements like length and width. Preliminary extraction of these traits, although proven to be predictive (Vandeputte, et al., 2017; Kora, et al., 2000) could narrow our view and limit a broader discovery of other possibly correlated traits. Instead, whole fish images, without biological landmarks, are used as input without any presumption on what imaginal features are most predictive for the traits of interest.

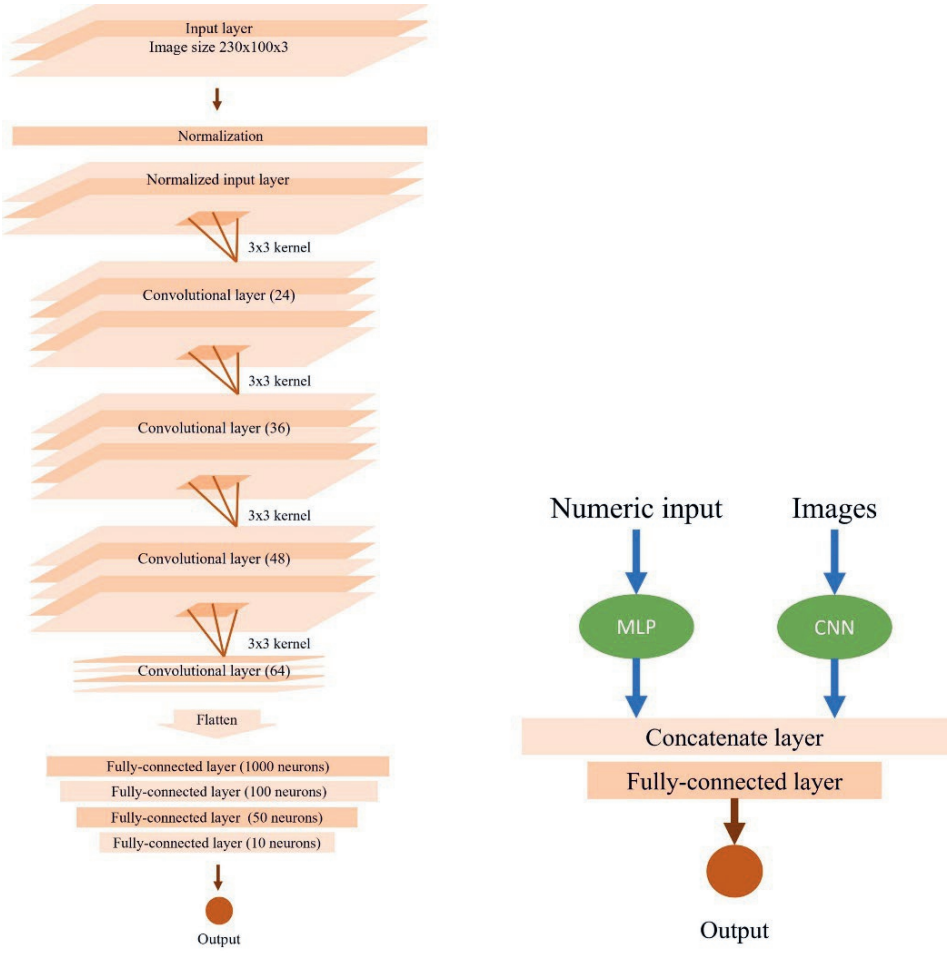
To avoid overfitting, the original CNN architecture has been simplified empirically and optimized through a series of experiments with various layer configurations (Fig. 2.2a). The final CNN includes an input layer, a normalization layer, and four convolutional blocks with 24, 36, 48, and 64 layers, respectively. The input layer takes RGB images with a resolution of 230x100. All convolutional layers use kernel size 3x3. The output from the 4th convolutional block is flattened and 4 fully connected layers are used before the final output. For visceral percentage prediction, due to the heavy overfitting, the fully connected block is modified. Each fully connected layer is followed by a dropout layer. The dropout ratio is empirically chosen as 0.5, meaning that the value of half of the neurons in the previous fully connected layer will be randomly changed to 0 in each training iteration. Dropout layers can force the neural network to reveal imaginal features that are of the most

relevance to the prediction. The final output layer performs regression through a rectified linear unit activation function.

When receiving numeric variables in addition to images, the framework combines CNN with MLP to construct a multi-input neural network (multi-input NN) that integrates data of different types. A multi-input NN can process images through CNN and simultaneously process numeric data through MLP. The outputs of these two compartments are then combined by a concatenated layer and passed down to a fully connected layer, and eventually, to the output (Fig. 2.2b).

In this study, NN was trained to minimize the MSE between the output and the observed value of the target trait. MSE and Pearson correlation coefficient (r), also addressed as the accuracy of prediction in this paper, are metrics to evaluate the result of each training round. During the training section, if NN gradually captures the relationship between predictors and target trait, MSE for training and validation will correspondingly decrease and eventually stabilize to a minimum, while r will increase and stabilize to a maximum. Visualization of changes in these metrics during training and validation is called the learning curve.

The NN was built under Python 3.8 (Van Rossum, Drake, & others, 1995) with Tensorflow 2.3 (Abadi, et al., 2016). The training and validation were performed on an HPC cluster with one GPU (NVIDIA V100 Tensor Core). The average training time of NN was 6 to 8 h per 100 epochs, with the aforementioned structure and configuration.



(a) CNN architecture. Numbers in parentheses indicate the number of layers in each convolutional block.

(b) Multi-input NN architecture

Figure 2.2 Neural network architecture.

Model comparison

Using the test data set, the performance of the three analytical methods is evaluated using the MSE and r between the predicted and observed value. Lower MSE and higher r therefore imply that a prediction model has better performance in terms of prediction accuracy. The best-performing prediction model is then presented to be used for future predictions of a specific target trait.

Feature extraction

Feature visualization shows the filtered features from each convolutional block which aids the human understanding of the training process of the NN. From the test set, we randomly choose an image, which is then fitted to the prediction model. This image will be converted into feature maps in the same manner as that of the training

and validation images. The more convolutional blocks there are, the more abstract and informative the feature will be, and the more challenging they are for humans to interpret. The visualized features can range from edges and textures to patterns and individual object parts.

Visualization

The feature extraction reveals regions of images that contribute to prediction. However, these features might have a different effect on the target trait. Phenotypic correlation between traits can be positive and negative. The direction of correlation matters because it has an impact on the selection of traits included in breeding programs. This section explains the connection between the features and the target trait, focusing in particular on how certain features affect the prediction. This is accomplished by the Score Class Activation Map, or score-CAM, a novel post-doc visualization method (Wang, et al., 2020).

The first step in score-CAM is to acquire the activation map. The activation map shows the regions in the image that were relevant to the prediction. We randomly select an image from the test set, fit it to the prediction model, and use the prediction value for the target trait as the output. The activation maps are then produced by adding a global average pooling layer after the last convolutional layer. The pooling layer eventually produces numerous activation maps for this prediction after repeatedly fitting the same image into the NN. These activation maps are then applied to the original image as a mask, so that only the overlapped areas of the activation maps and the original image are sent, again, repeatedly to the NN. Each activated region will consequently receive a score to indicate how important it is to the prediction. Ultimately, an average map is produced by linearly combining these scores with their corresponding activation maps as weights. An activated region that receives a higher score appears brighter and makes a greater contribution to the prediction value on the final activation map. By doing so, score-CAM not only shows the significant features but also their relative weight in predicting the target trait.

Score-CAM produces multiple scores for various prediction classes when used for classification. However, when used for regression as in this framework, there is one class with continuous prediction values. High-scoring regions, therefore, have higher prediction values. In other words, increasing certain properties of these regions could lead to a change of value in the target trait. We also modified the scores by taking their negative values before combining them with their corresponding activation maps. The value of certain properties of the features in these regions is presumably negatively correlated with the target trait. The properties of the features could be their size, length, shape, or even relative distance to other features. Understanding the direction of the correlation gives one the chance to speculate about the biological mechanism that links these imaginal regions or features to the target trait.

2.2.3. Case study: fat percentage, visceral weight, and visceral percentage prediction

To demonstrate the framework, we used data that was collected on 4,766 gilthead seabream (*Sparus Aurata*) at harvest. These fish were breeding candidates that were gathered from two different farms in Greece and Spain. Fish were all from the same population. However, since the main focus of this study is phenotype prediction, pedigree information was excluded from the analysis because it is not relevant at this stage. The data collection was conducted in sessions at different times between 2018 and 2020. Each seabream was first sedated and pictured in a lighting cabinet. The lighting cabinet has three LED panels on the top, side, and bottom. All three panels provide the light needed to take pictures of the object. The top horizontal panel is equipped with a built-in RGB camera (Logitech HD Pro Webcam C920 in Greece and Intel Realsense camera in Spain). Each seabream was laid with its lateral side up on the bottom horizontal panel. A grid paper was placed as the reference scale between the bottom panel and the seabream in Greece, and a blue paper as the background in Spain. Before being imaged, each seabream was weighed manually on a scale for body weight. Fat percentage was measured after imaging, using a hand-held device called Distell Fish Fatmeter (Brosset, et al., 2015; Kent, 1990), hereafter termed fatmeter, on sedated fish. Fat percentage is a potential indicator for fat composition, metabolic health (Óskarsson, 2008), and fitness of a live fish, and it influences the quality of the final fillet product. In total, 8 measurements of fat percentage were taken for each seabream: two times on the dorsal side and two times on the ventral side for each lateral side. Out of the 4,766 seabreams, 1,697 were slaughtered and measured for their visceral weight. Visceral weight is the amount of intestine, containing important organs such as guts, stomachs, and liver (Gjedrem T. , 1997). The visceral percentage was calculated by taking the ratio of visceral weight over body weight per fish.

Data was divided into 3 overlapping subsets, to model for the predictability of three traits: fat percentage, visceral weight, and visceral percentage separately. In each subset, data were further randomly divided into training, validation and test sets with a size ratio of 3:1:1. All three divisions of each subset contained no missing values, while the outliers were removed based on domain knowledge: records with negative body weight or visceral weight were removed; two fatmeter measurements with the value of 112.9% and 82% were removed, as a single fat percentage measurement should not be considerably higher than 20%. Images were connected with trait information by a unique pit-tag ID for each fish. Some fish had images but no traits information and vice versa. Records of these fish were also removed. Table 2.1 summarizes the variables within each modeling subset as well as the description of each variable.

Table 2.1 Division of Seabream data into three overlapping subsets for the prediction of fat percentage, visceral weight and visceral percentage, respectively.

Description	Group A	Group B	Group C
Prediction traits	Fat%	VW	Viscera%
Variable traits	BW	Fat%, BW	Fat%, BW
Number of images (train, validation & test sets)	1711, 679 & 679	1280, 209 & 208	1280, 209 & 208

BW: body weight, Fat%: fat percentage, VW: visceral weight, Viscera%: visceral percentage.

2.3 Results

Table 2.2 includes the descriptive statistics of the three test sets: fat percentage, visceral weight, and visceral percentage. Table 2.3 shows an overview of the prediction performance by fitting LR, MLP and multi-input NN separately on each test set, respectively.

Table 2.2 Descriptive statistics of test sets.

Description	Group A – Fat%	Group B – VW	Group C – Viscera%
Mean (unit)	11.52 (%)	40 (g)	8.37 (%)
Standard deviation	2.5	18	1.91

Fat%: fat percentage, VW: visceral weight, Viscera%: visceral percentage.

Table 2.3 Evaluation of prediction models on test sets.

	Group A - Fat%		Group B - VW		Group C - Viscera%	
	MSE (%²)	***r***	**MSE**(g²)	***r***	**MSE** (%²)	***r***
LR	4.86	0.48	63.26	0.90	2.13	0.64
MLP	5.26	0.48	75.41	0.89	2.30	0.64
Multi-Input NN (with images)	2.98	0.7	87.14	0.87	2.39	0.66

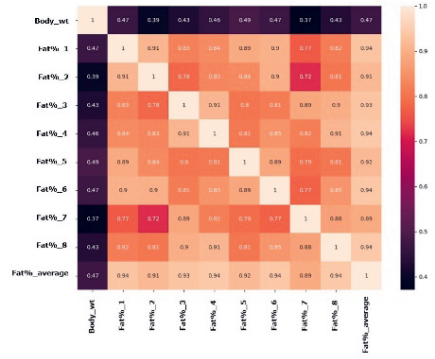
Fat%: fat percentage, VW: visceral weight, Viscera%: visceral percentage.

To summarize, the image model increased the prediction accuracy from 0.48 to 0.70 compared to both numeric models for fat percentage. For visceral weight, the accuracy of prediction is high and similar for all models, ranging from 0.87 to 0.90. For visceral percentage, the accuracy of prediction is moderate, around 0.6 for all three models. The following subsections will illustrate in detail how the models for all traits came to the prediction results, together with the feature extraction and visualization.

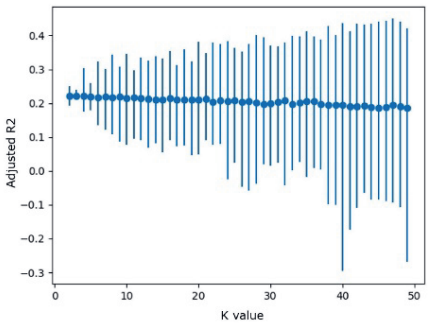
2.3.1 Fat percentage

Numeric models

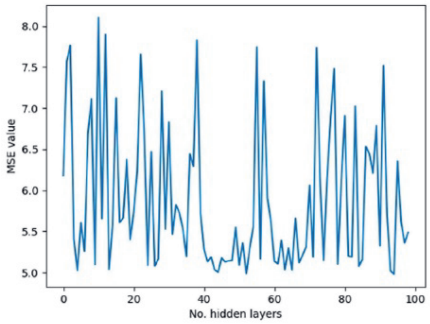
Exploratory data analysis of fat percentage (Fig. 2.3a) suggested strong correlations between the 8 individual fatmeter reads. Therefore, an average fat percentage was taken for modeling instead of 8 repeated records. However, none of the records, including the average, indicated a good correlation with body weight, with the average correlation being 0.47. The correlation between body weight and fat percentage is moderate. However, it is still questionable whether body weight as a single variable was sufficient to explain the variance in fat percentage. This is further confirmed by the low adjusted R^2 (0.22) from the K = 3 cross-validation of the linear model, suggesting that only around 20% of the variance in fat percentage was explainable by body weight (Fig. 2.3b). The presence of negative adjusted R^2 in Figure 2.3b during training also implies that body weight is negligible for fat percentage prediction, especially when the sample size is too small.



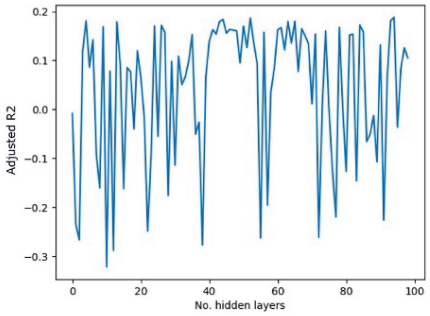
(a) Correlations between eight fat measurements, their average, and body weight



(b) Results of linear regression - cross-validation on training and validation set



(c) Training of MLP - validation MSE changes as the number of hidden layers increases



(d) Training of MLP - validation adjusted R^2 changes as the number of hidden layers increases

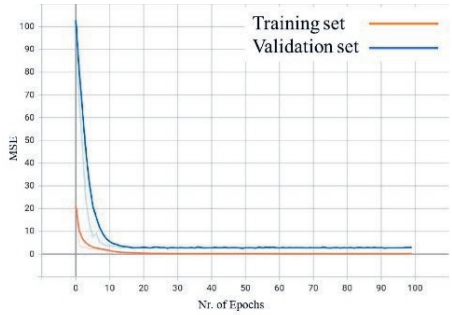
Figure 2.3 Exploratory data analysis, training, and validation of numeric models for fat percentage prediction.

The non-linear numeric model MLP showed similar results as the LR. A layer number of 50 was chosen based on the MSE (Fig. 2.3c) and adjusted R^2 (Fig. 2.3d)

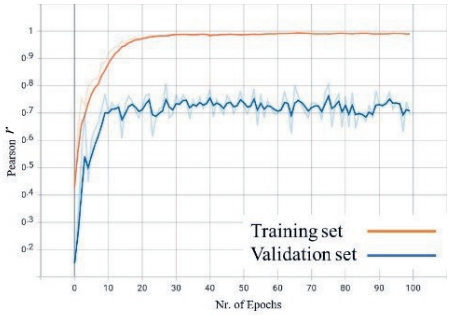
of the prediction on the validation set. However, as the layer number increased, the model performance failed to stabilize. Furthermore, the highest adjusted R^2 of 0.16 in all rounds of validation was lower compared to that of LR. By comparison, a linear model is as good as, if not better, than MLP in fat percentage prediction. The similarity of LR and MLP on prediction accuracy indicates that the non-linear, complex MLP failed to capture more information compared to a linear model. Given the instability of the MLP performance during training, it is arguable whether a high-dimensional relationship between body weight and fat percentage is formed.

Machine vision

On the other hand, the neural network quickly captured the relationship between image, body weight, and fat percentage after 20 epochs of training as the learning curves plateaued (Fig. 2.4). MSE on the validation set followed the same pattern and stabilized at a value slightly higher than that of the training set (Fig. 2.4a). Pearson r on the validation set fluctuated around 0.7 towards the end (Fig. 2.4b). The final test set resulted in an accuracy of 0.7, considerably higher than that of LR and MLP.



(a) Learning curve - loss value decreases with epochs



(b) Learning curve - Pearson r increases with epochs

Figure 2.4 Training and validation of neural network for fat percentage prediction.

Feature extraction

The features extracted by the convolutional layers are displayed in Figure 2.5. The first row illustrates that in the beginning, NN extracted whole-fish contours and contrasts between regions as informative features for fat percentage prediction. Several visually dark regions correspond to morphological locations like patch, eye and fin on the colored image. However, NN quickly abandoned these regions and focused on outlines and contours, shown by the features in the second row that corresponds to the second convolutional block. The final convolutional block in the last row comprises features mostly highlighted in visually dark color: the curvature of the dorsal side and the upper edge of the pectoral fins.

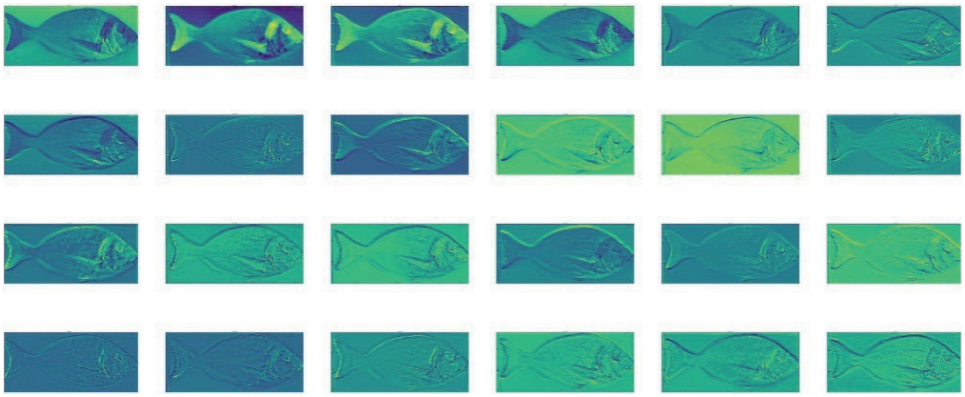
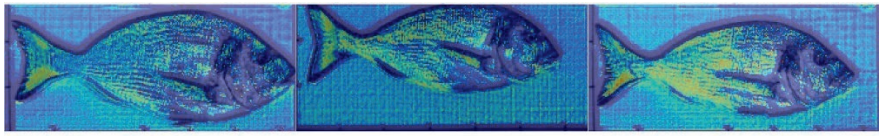


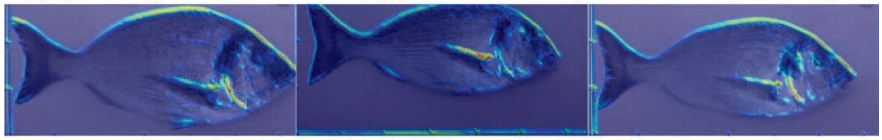
Figure 2.5 Extracted features by each convolutional block for fat percentage prediction. The four rows represent the four convolutional blocks shown in the NN architecture. Each row only contains features obtained from the layers in the same block.

Visualization

Three images were randomly selected from the test set. Visualization of the prediction using ScoreCAM pointed to the same direction as feature extraction in Figure 2.5. Figure 2.6 illustrates how certain regions in the images contribute to a prediction output. For fat percentage prediction, all highlighted regions are informative and important. However, they could affect the output in different ways. Figure 2.6a shows regions that cause an increase in output. In other words, properties of these regions which cover a major part of the fish body without the head are positively correlated with fat percentage. The ventral side region was also emphasized with strong brightness in all three images. While figure 2.6b accentuates three specific areas: the dorsal side, the upper edge of the pectoral fin, and the edge of the operculum. These regions have some properties that are negatively correlated with fat percentage. One possibility could be the size of the region surrounded by these three areas: the bigger this surrounding region becomes, the lower the fat percentage will be.



(a) Specified regions and features whose properties increase the output value in three random images.



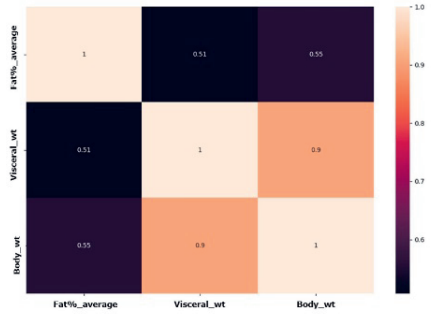
(b) Specified regions and features whose properties decrease the output value in three random images.

Figure 2.6 ScoreCAM feature visualization under two situations - increase and decrease of the output value for fat percentage prediction.

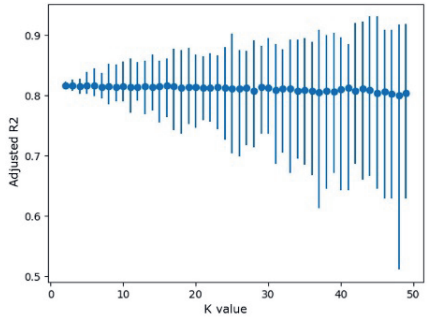
2.3.2 Visceral weight

Numeric models

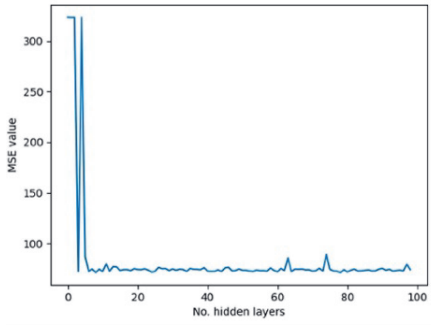
Exploratory data analysis showed a strong correlation between visceral weight and body weight and a poor correlation between average fat percentage and the other two traits (Fig. 2.7a). K = 10 was chosen for the LR model as it achieved the highest average adjusted R^2 of 0.81 (Fig. 2.7b). MSE stabilized after using 5 hidden layers in MLP (Fig. 2.7c). The adjusted R^2 reached around 0.8 (Fig. 2.7d), similar to that of the 10-fold LR. On the final test set, however, the accuracy of MLP was 0.89, slightly lower than the result of LR.



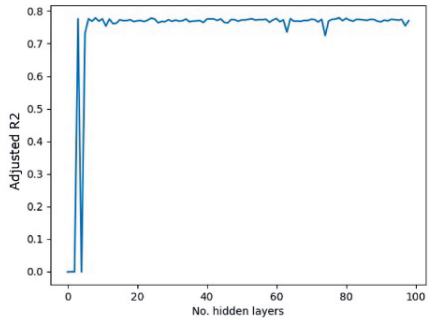
(a) Correlations between the average fat measurements, body weight, and visceral weight



(b) Results of linear regression - cross-validation on training and validation set



(c) Training of MLP - validation MSE changes as the number of hidden layers increases

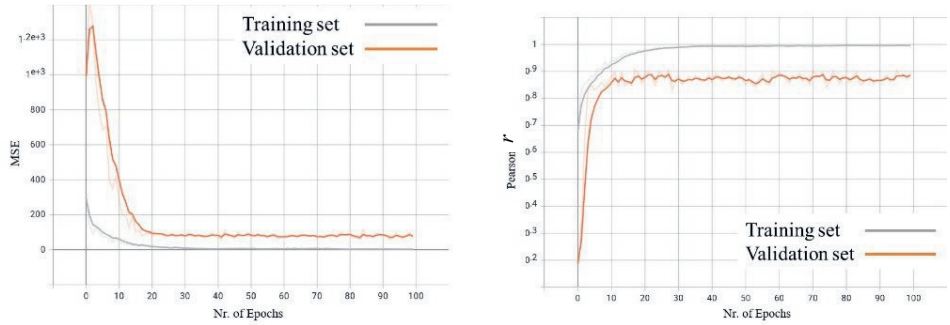


(d) Training of MLP - validation adjusted R^2 changes as the number of hidden layers increases

Figure 2.7 Exploratory data analysis, training, and validation of numeric models for visceral weight prediction.

Machine vision

Neural network performance plateaued (Fig. 2.8) after around 35 epochs of training. MSE on the validation set stabilized at a value slightly higher than that of the training set towards the end (Fig. 2.8a). Pearson r on the validation set did not fluctuate as much, varying in a range of 0.85 to 0.9 (Fig. 2.4b). The final test set resulted in an accuracy of 0.87, slightly lower than that of both LR and MLP.



(a) Learning curve - loss value decreases with epochs

(b) Learning curve - Pearson r increases with epochs

Figure 2.8 Training and validation of neural network for visceral weight prediction.

Feature extraction

Figure 2.9 displays the extracted features for visceral weight prediction. In the first row, whole-fish contours were already captured as informative features. Afterward, the attention of NN fixated on fish outlines and contours with different levels of contrast to their surroundings. The final convolutional block in the last row basically featured only the whole fish area imprinted on a background with different brightness.

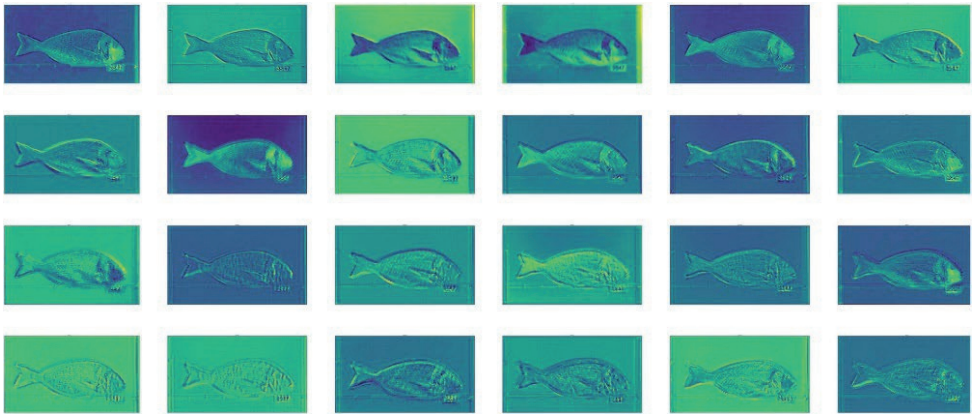
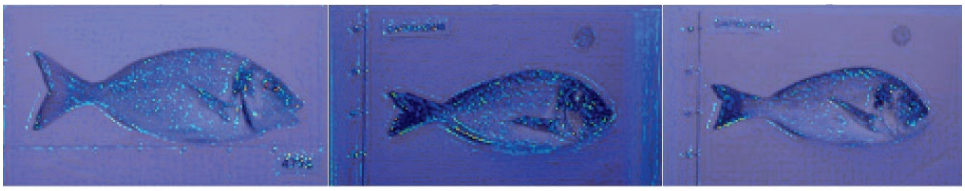


Figure 2.9 Extracted features by each convolutional block for visceral weight prediction. The four rows represent the four convolutional blocks shown in the NN architecture. Each row only contains features obtained from the layers in the same block.

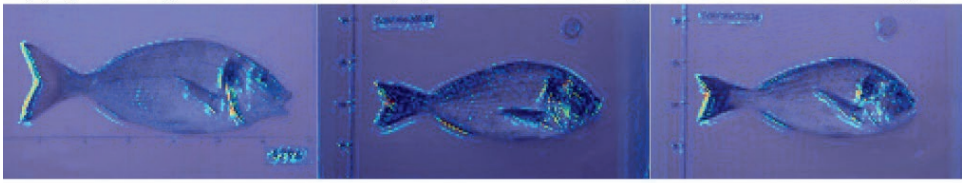
Visualization

Score-CAM visualization added interesting details to the importance of the whole fish area in visceral weight prediction. Figure 2.10a depicts dotted lines that either outline the shape of the whole fish or cover the majority of the body area. These are the regions and features whose properties contribute positively to visceral weight. One obvious possibility is that the bigger the fish is, the more visceral weight it will have. While Figure 2.10b highlighted three specific areas whose properties have a

negative effect on the output: the end of the head, the lower edge of the pectoral fin, and the back edge of the caudal fin.



(a) Specified regions and features whose properties increase the output value in three random images.



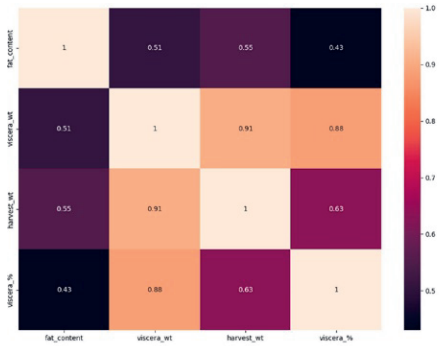
(b) Specified regions and features whose properties decrease the output value in three random images.

Figure 2.10 ScoreCAM feature visualization under two situations - increase and decrease of the output value for visceral weight prediction.

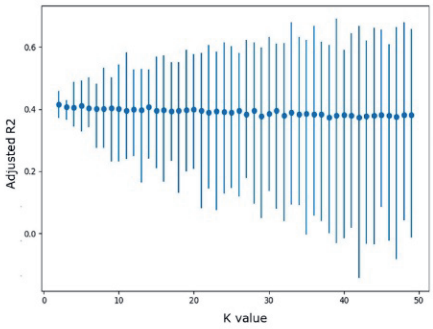
2.3.3 Visceral percentage

Numeric models

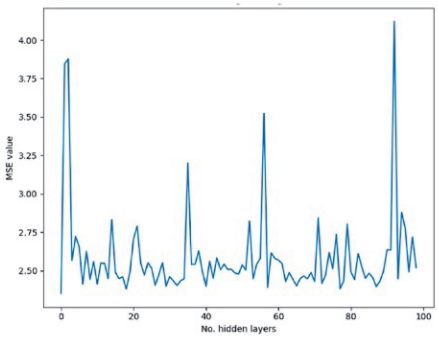
Exploratory data analysis (Fig. 2.11a) showed the correlation of visceral percentage with other traits. The visceral percentage is more correlated with visceral weight than with harvest body weight, although calculated from both traits. Therefore, to avoid overfitting, only fat percentage and harvest weight were included as numeric inputs to predict visceral percentage. In linear regression with 5-fold cross-validation, the adjustedR^2 was 0.41. This indicated that body weight and fat percentage in total can only explain 40% of the variance in visceral percentage (Fig. 2.11b).



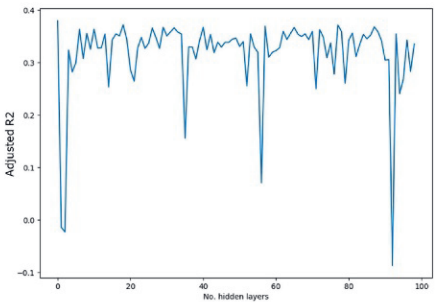
(a) Correlations between fat percentage, harvest body weight, visceral weight and visceral percentage



(b) Results of linear regression - cross-validation on training and validation set



(c) Training of MLP - validation MSE changes as the number of hidden layers increases



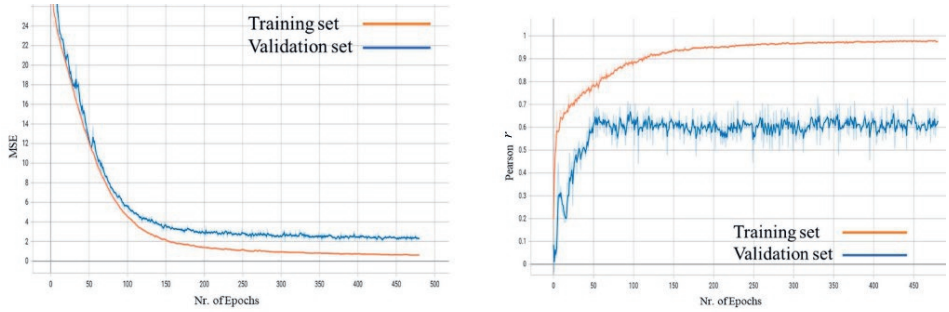
(d) Training of MLP - validation adjusted R^2 changes as the number of hidden layers increases

Figure 2.11 Exploratory data analysis, training, and validation of numeric models for visceral percentage prediction.

For multi-layer perceptron, a layer number of 5 was chosen based on the MSE (Fig. 2.11c) and adjusted R^2 (Fig. 2.11d) of the prediction on the validation set. Clearly, adding layers did not manage to stabilize MSE. Comparable to fat percentage prediction, MLP showed similar results as linear regression in visceral percentage prediction. It is obvious that a high-dimensional relationship does not exist between body weight, fat percentage, and visceral percentage.

Machine vision

The training of the neural network was not as smooth as the other two traits. To stabilize both the MSE and accuracy, the neural network for visceral percentage was trained 4 times more epochs than that of the other traits. After 150 to 200 epochs of training, the MSE finally plateaued (Fig. 2.12a). On the validation set, Pearson r showed a constant fluctuation around 0.6, roughly varying in a range of 0.55 to 0.65 (Fig. 2.12b).



(a) Learning curve - loss value decreases with epochs

(b) Learning curve - Pearson r increases with epochs

Figure 2.12 Training and validation of neural network for visceral percentage prediction.

Feature extraction

The extracted features for visceral percentage prediction in Figure 2.13 are alike that for visceral weight. In visceral weight prediction, whole-fish contours were gradually highlighted after two convolutional blocks. However, in the case of extracted features for visceral percentage, the outlines of the fish were captured at the very early stage of the convolutional process. Edges of the other objects in the images were also captured, including the printed coordinate on the background and the written fish ID on the side. These features increase the robustness of the network but could also be the major source of noise for prediction.

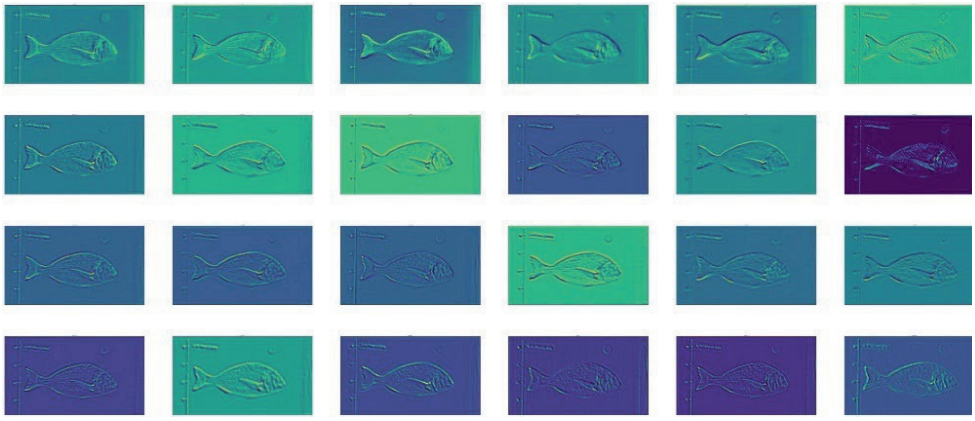
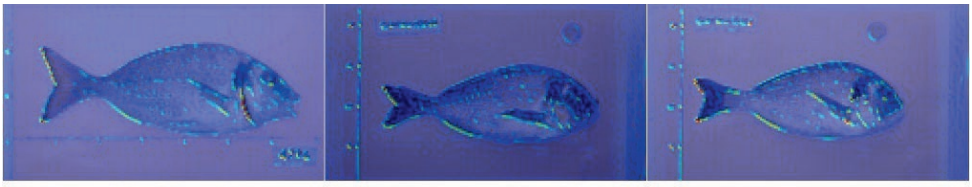


Figure 2.13 Extracted features by each convolutional block for visceral percentage prediction. The four rows represent the four convolutional blocks shown in the NN architecture. Each row only contains features obtained from the layers in the same block.

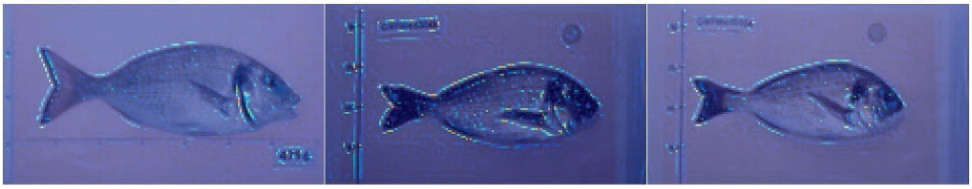
Visualization

Score-CAM of visceral percentage also reveals similar regions of interest compared to that of visceral weight. However, some differences can be observed. Figure 2.14a shows the regions and features whose properties contribute positively to the visceral percentage. Apart from the general contour, one interesting feature is the curve

connecting the lower beginning of the tail and the vent. This region was also vaguely highlighted by a dotted line for the visualization of visceral weight prediction. While in visceral percentage prediction, it is visually more clearly defined. Biologically this region corresponds to the position of the anal fin. Besides, the pectoral fin and the edge of the operculum were also highlighted as important features. While in Figure 2.14b, compared to that of the visceral weight, less emphasis was shown on the end of the caudal fin and the operculum.



(a) Specified regions and features whose properties increase the output value in three random images.



(b) Specified regions and features whose properties decrease the output value in three random images.

Figure 2.14 ScoreCAM feature visualization under two situations - increase and decrease of the output value for visceral percentage prediction.

2.4 Discussion

The framework provides a comprehensive, structural, and efficient strategy for predicting slaughter traits. Three novel aspects of this framework are highlighted here. The first is the direct use of images rather than the introduction of image-derived landmarks. Images are unstructured data, whereas structured data typically includes biologically intuitive descriptive information such as body length and width, as well as morphometrics indices (Nash, Valencia, & Geffen, 2006; Jones, Petrell, & Pauly, 1999). Common practice is to first extract structured information from unstructured data, which frequently requires extensive landmark annotation (Holmes & Jeffres, 2021; Prchal, et al., 2020; Navarro, et al., 2016; Haffray, et al., 2013). However, descriptive morphological features, particularly indices (McPherson, Slotte, Kvamme, Meier, & Marshall, 2011), are not always adequate for predicting slaughter traits (Wilder, Raubenheimer, & Simpson, 2016). Prediction accuracy is inherently limited by what can be derived from images and the biological understanding of both predictor and target traits. The proposed framework overcomes these constraints while prioritizing prediction accuracy. The framework begins with unstructured, whole fish images, moves on to predictive analysis, and then returns to descriptive insights (Vaughan, 2020). Descriptive information is then revealed through feature extraction and visualization.

Descriptive information is the second aspect of novelty: interpretability. Consider the outcome of the case study on fat percentage prediction. Visualization revealed an important region bordered by the edge of the operculum and pectoral fin. This region anatomically overlaps with the abdominal cavity and the location of the liver and the heart, both of which are important for fat metabolism. The ventral side is also highlighted as an important region for fat percentage. The ventral side of the fish contains more fat than the dorsal side on average, which may explain why this region is a good indicator of fat percentage. Although the fat percentage is measured mainly on the fillet, it is possible that there is enough exhibited variance along the dorsoventral axis within the lateral side. The study of Gulzari et al. (2022) supports similar results by showing both genetic and phenotypic correlations of fillet fat percentage and visceral weight. For visceral weight, a simple linear regression from body weight results in the best prediction. Nonetheless, feature extraction interprets the significance of size and shape in predicting visceral weight, as these two features are most frequently revealed by the convolutional blocks. Based on biological knowledge, size and shape are morphometrics that is strongly related to body weight. Previous research (Yang, et al., 2021) demonstrated the prediction of body weight from images. Therefore, it is reassuring that the framework confirms that the size and shape are sufficient to phenotype the traits of body weight and visceral weight for seabream. However, when it comes to visceral percentage, it might be beneficial to add different measurements since the visualization for visceral percentage highlighted the position of the anal fin as an informative feature.

The third aspect of novelty is the possibility of selecting different target traits for prediction. The framework was used in the case study to predict visceral weight, fat percentage, and visceral percentage. In principle, any other slaughter traits that are collected during harvest can be used as target traits for prediction. This increases the feasibility of developing models that can relate exterior traits with indicator traits on functions, metabolism and behaviors (Fu & Yuna, 2022), such as visceral fat. A next step would be to predict visceral fat from body weight and images. Visceral fat is a slaughter trait and also a key indicator of feed conversion rate. Fish with lower visceral fat percentage convert feed more efficiently and are therefore preferred in production (Besson, et al., 2022; García-Celdrán, et al., 2015). The Distell fat meter used in our study measures fat percentage in fillet and muscle tissue. Fat percentage in the fillet region may however be a poor indicator of visceral fat, which is found in the abdomen. The framework can find the promising features that can be applied as indicator traits for the reduction of visceral fat by breeding. In principle, the same applied for any other slaughter traits that are collected during harvest.

The framework has the ability to use machine vision for regression analysis. To the best of our knowledge, the majority of machine vision applications in aquaculture are for classification, including identification (Saberioon, Gholizadeh, Cisar, Pautsina, & Urban, 2017). In some livestock research studies (Shahinfar, Khansefid, Haile-Mariam, & Pryce, 2021), quantitative measurements are converted into

qualitative categories. The conversion of data types from continuous to categorical can result in a significant loss of explainable variance in the target traits. The framework takes whole-fish images as input so that more variances can be preserved. Using whole images also saves time and effort on data annotation prior to model training.

Studies on deep learning applications in aquaculture frequently include multiple algorithms as analytical methods and compare their performance (Palaiokostas, 2021), particularly in genomic prediction. These algorithms are often state-of-art in the field of machine learning. When a promising algorithm is applied to different data for different research questions, the accuracy of performance cannot be guaranteed. Taking this into account, our framework has a structure that includes 3 different analytical methods based on their ability to capture different levels of complexity in the relationship between the predictor and target traits. We attempted to include the most recent, cutting-edge analytical methods available at the time this study was conducted. These methods are modularized, including feature extraction and score-CAM for visualization. The framework is built by combining these fundamental analytical methods and has the potential to be upgraded. LayerCAM (Jiang, Zhang, Hou, Cheng, & Wei, 2021), for example, could replace score-CAM for visualization; in machine vision, prediction accuracy can potentially be increased by using a deeper NN with more convolutional blocks. It should be noted that the cost of computing power and data storage potentially rises as more updated analytical methods are integrated in the framework.

This framework can be improved further in two ways. First and foremost, data collection and utilization. The seabream included in the study came from various families, generations, and two different locations. This ensures that there is enough variation so that the predictions are as applicable to future data as possible. If new data is collected from the same location, it can be used directly for prediction as well as to improve the framework's performance. However, under certain conditions, slaughter traits may vary seasonally and inter-annually (Goñi & Arrizabalaga, 2010; Rodríguez, Fountoulaki, Grigorakis, Alexis, & Flos, 2010). When used in different circumstances, such as a new farming location, it is still unclear whether re-training of the framework is required, as the pre-trained framework may not cover the variance introduced by the geographical differences. Another option is to add location, season or family origin to the framework as a new input variable. It is also worth emphasizing the importance of images in breeding data collection. Incorporating imaging into the phenotyping routine can be difficult and expensive. Image quality varies, which potentially influences machine vision results. However, images play an important role in the framework. The framework proves that including whole-fish images can be most beneficial, potentially improving model accuracy and interpretability. Imaging is non-invasive and high throughput. In combination with the framework, it creates room for more efficient and innovative phenotyping practice. For instance, fat percentage can be measured using the hand-

held fatmeter device. The time and labor required in this process will be considerably reduced if the same measurement is replaced by image analysis. The same applies to visceral weight. Visceral weight can be measured on fish in a non-invasive or even contact-free manner using its correlation with the body weight, which can also be derived from images using the size and shape of the fish. Given this trade-off, we strongly advise breeders and other users of this framework to collect images in addition to other trait data.

Another improvement is the requirement for additional qualification on the visualized feature. Although features and regions revealed by feature extraction and visualization appear intuitive to human eyes, they require quantitative description. For example, the region bordered by the edge of the operculum and the pectoral fin is shown to harbor information that is negatively correlated with fat percentage. However, the precise properties of this region that are causing this prediction are unknown. It could be the angle of the pectoral fin against the operculum, or the distance between the pectoral fin and the dorsal edge of the operculum, or a very different characteristic. More research is needed to properly segment these features from whole-fish images and quantify the properties of these features. For visceral percentage, the position of the anal fin, together with the edge of the operculum, was shown to be the key to an improved prediction. Further work is required to transform these features into precise landmarks or generate the activation map on new images and use the color differences in the heat map to isolate these features. Furthermore, visualization only shows the strength of the correlation between features in the context of the same image. For example, the end of the caudal fin was strongly highlighted as being negatively correlated with visceral weight. According to the framework, body weight can explain the majority of the variance in visceral weight. There is no clear evidence that the shape of the caudal fin has a significant impact on body weight. However, the shape of the caudal fin is revealed to be accountable for visceral weight. Therefore, to reach a more general conclusion, it is necessary to test the significance of these features using other statistical analyses, such as stepwise selection, ridge regression, or lasso regression.

2.5 Conclusion

In this paper, we present a framework for predicting invasive or destructive traits based on non-invasive traits and images. The framework can capture various levels of relationship between the predictor and the target trait and generate the prediction model by prioritizing prediction accuracy. We provide a detailed interpretation of the model by extracting and visualizing predictive imaginal features. We show how to use the framework with three examples: visceral weights, fat percentage and visceral percentage in gilthead seabream. The framework could be applied to other harvest or post-slaughter traits such as visceral fat, liver weight, or even metabolic traits. Further research can expand on this framework progressively in these three aspects: 1. Predictive imaginal feature segmentation and quantification; 2. Including pedigree and genetic data to estimate the potential for indirect selection in breeding;

3. Using the framework on different invasive traits in different fish species, to build a more complete phenotypic profile of traits and their correlations.

Acknowledgements

This project receives financial support from the Topsector Agri & Food. The use of the HPC cluster has been made possible by CAT-AgroFood (Shared Research Facilities Wageningen UR).

3. Genetic analysis of swimming performance in rainbow trout (*Oncorhynchus mykiss*) using image traits derived from deep learning

Yuuko Xue[1], Arjan Palstra[1], Robbert Blonk[2], Hans Komen[1], Haris Ahmad Khan[3], Robert Mussgnug[1], and John Bastiaansen[1]

[1]Department of Animal Breeding and Genomics, Wageningen University & Research, Wageningen, 6700 AH, the Netherlands

[2]Hendrix Genetics Aquaculture BV, Boxmeer, 5831 CK, the Netherlands

[3]Agricultural Biosystems Engineering Group, Wageningen University & Research, Wageningen, 6700 AA, the Netherlands

This chapter is based on:

Xue, Y., Palstra, A. P., Blonk, R., Mussgnug, R., Khan, H. A., Komen, H., & Bastiaansen, J. W. (2025). Genetic analysis of swimming performance in rainbow trout (Oncorhynchus mykiss) using image traits derived from deep learning. Aquaculture, 606, 742607.

Abstract

The physical and physiological condition of fish directly influences their swimming performance, which is crucial for their health and survival. This study explored how physical characteristics affect swimming performance in rainbow trout. 3D images were used to capture the morphology of fish and assess its impact on critical swimming speed (U_{crit}), measured via individual swim tests. A convolutional neural network (CNN) was utilized to predict U_{crit} from the images. Using Gradient-weighted Class Activation Maps (GradCAM), image regions that contributed to U_{crit} predictions were visualized. These regions were further refined into areas that are biologically relevant to U_{crit}, leading to the definition of four swim traits: head volume, caudal fin volume, epaxial muscle volume, and shape. Our findings indicated that U_{crit} is moderately heritable. Genetically, heavier fish demonstrated poorer swimming performance; among fish of the same weight, those with larger and broader epaxial muscles, larger heads, and smaller caudal fins performed worse. Although genetic improvement of U_{crit} is feasible, caution is advised because of potential correlated responses that reduce the body volume and epaxial muscle volume. The interdisciplinary workflow (data collection, model construction, visualization, interpretation, definition, and evaluation) in this study demonstrated how image-based deep learning can be used as a hypothesis-free approach to deepen the understanding of the genetic background of complex traits. Additionally, it highlights the value of genetic analysis to validate the physiological interpretation of Explainable AI, broadening the opportunities to discover novel phenotypes in aquaculture.

Key words: Deep learning; Critical swimming speed; Image analysis; Explainable AI

3.1 Introduction

The swimming performance of fish is an indicator of health and is crucial to survival. One of the most common estimates (Farrell, 2008; Drucker, 1996; Beamish, 1978) of swimming performance is the measurement of the critical swimming speed (Brett, 1964). Critical swimming speed, abbreviated as U_{crit}, is the maximum velocity a fish can maintain for a specific period of time. Often used as an indicator when investigating behavior (Dehaan, 2019), production, health (Castro, et al., 2013; Tierney & Farrell, 2004), and environmental resilience (Athammer, 2023; Silva, Alexandre, Quintella, & de Almeida, 2021; Mengistu, et al., 2021; Vandeputte, et al., 2016), U_{crit} is determined by subjecting fish to an exhaustive swimming test, where swimming speeds incrementally increase at predefined intervals until the fish fatigues.

Two major factors influence U_{crit}: the physiological and physical condition of the fish. Researchers primarily monitor the underlying physiological parameters contributing to U_{crit}, such as heart rate and oxygen consumption (Farrell, et al., 2003; Altimiras, et al., 2002). The relation between physical condition and U_{crit} is usually characterized by morphological traits such as body length and mass, although muscle composition and structure also receive attention for their critical function in cardio activity (Hachim, et al., 2021; Claireaux, et al., 2005; Rome, Choi, Lutz, & Sosnicki, 1992; Farrell, Johansen, & Suarez, 1991). Even after adjusting U_{crit} for body length, there is a significant relationship with body mass (Cai, Chen, Johnson, Tu, & Huang, 2020; Rubio-Gracia, García-Berthou, Guasch, Zamora, & Vila-Gispert, 2020; Cano-Barbacil, et al., 2020). Fish of the same length and weight can still exhibit different swimming abilities and considerable diversity in their morphology such as girth, volume, and shape (Svozil, Baumgartner, Fulton, Kopf, & Watts, 2020; Yan, He, Cao, & Fu, 2013). Given that, research is required to thoroughly investigate and quantify the relationship of U_{crit} with finer-scale morphological units of fish.

Using conventional measurement techniques for morphology has two drawbacks. First, these measurements are invasive, time-intensive, prone to errors, and may require anatomical dissection (Moya, et al., 2019). 3D image analysis provides a promising, non-lethal solution to extract morphological parameters. Trait extraction using 3D imaging has demonstrated higher accuracy and reduced bias compared to manual measurements (Gulzari B. , 2023), with success in livestock research for assessing muscle depth, back fat (Fernandes, et al., 2020; Miller, et al., 2019; McPhee, et al., 2017) and body condition score (Shi, et al., 2023), and in aquaculture for body mass, length, and girth (Li & Du, 2022). The second drawback of the conventional measurements is the difficulty identifying all relevant morphological traits for U_{crit}. Although the number of measurable traits on fish is limited, analyzing U_{crit} in relation to morphology still requires iterative hypothesis formation and testing, as U_{crit} is a complex trait influenced by many factors that may also manifest as subtle morphological variation (Videler, 1993).

To address these challenges a hypothesis-free approach, as part of the analytical framework proposed in (Xue, Bastiaansen, Khan, & Komen, 2023), was employed in an attempt to predict U_{crit} by incorporating whole, unannotated fish images and visualizing the most predictive image features through class activation maps (CAMs). Deep learning is frequently criticized for being a 'black box', making it challenging to interpret its results. Explainable artificial intelligence (xAI) is developed to increase the transparency and interpretability of deep learning systems (Saranya & Subhashini, 2023). Among various methods of xAI, CAMs are designed to visualize regions in images that support the decision-making of deep learning models (Zhou, Khosla, Lapedriza, Oliva, & Torralba, 2016). The 3D morphology can be efficiently narrowed down to specific morphological traits that are most relevant to U_{crit}.

While CAMs enhance model interpretability, their effectiveness relies on alignment with domain knowledge, which requires further evaluation by the intended audience and their expertise (Saeed & Omlin, 2023). To our knowledge, studies that utilized CAMs in livestock research have only implied such practice (Garcia, 2022). The relevance of CAM features to the prediction still needs to be tested through proper quantitative evaluation methods (Nauta, et al., 2023). such as genetic analysis.

This study aims to dissect the genetic basis of swimming performance and understand its subtle relationship with morphology in rainbow trout (*Oncorhynchus mykiss*). We investigate the feasibility of using genetic analysis to assess CAM - derived traits in enhancing the interpretability of image-based deep learning models within the aquaculture domain. To achieve this, a novel approach of "data collection - model construction - visualization - interpretation - definition - evaluation" is introduced, aiming at unveiling finer-scale morphological traits that enhance our comprehension of swimming performance.

3.2 Materials and methods

3.2.1 Ethical statement

The experiment was conducted at Troutlodge (Idaho, USA) under the farm's established Veterinary Health Management Plan (VHMP) and in accordance with RSPCA welfare standards. All procedures adhered to the local legislation regarding animal welfare.

This study is structured as follows: two experiments were conducted to phenotype individual fish for their morphology and swimming performance. A convolutional neural network (CNN) was then constructed using 3D images to predict Ucrit. Afterward, we used the gradient-weighted class activation map (GradCAM) to visualize what the prediction model has learned and transformed the GradCAM results into swim traits with clear definitions. Lastly, we estimated the genetic properties of the swim traits and Ucrit. See the flowchart in Figure 3.1 for an overview.

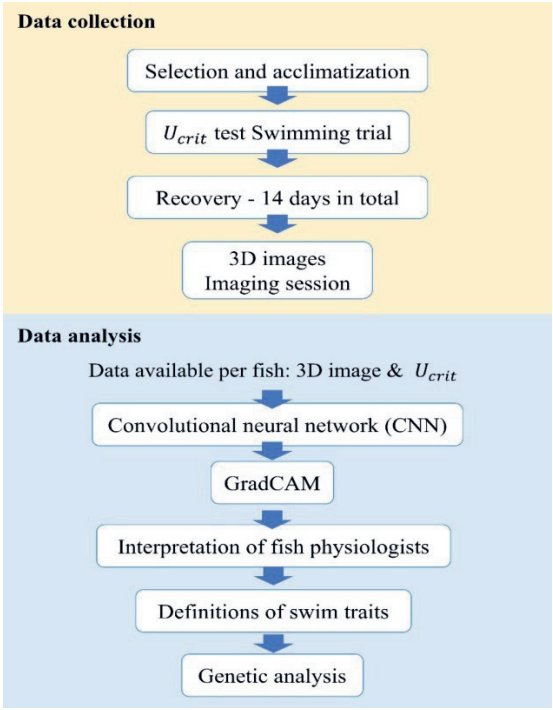


Figure 3.1 Overview of data collection and analysis. Arrows indicate the order in which the methodologies were applied.

3.2.2. Materials

100 full-sib families of rainbow trout with full pedigree (15 generations) were provided by Hendrix Genetics (Troutlodge, Idaho, USA). From each family, a group of 12 fish was selected at random. The bulk of fish were between 10 and 14 cm, and body length ranged from 8.4 cm to 15.2 cm, measured from the tip of the upper lip to the tip of the upper caudal lobe. The fish (n = 1200) were placed in a holding tank within the same facility and left to acclimatize for 48 h prior to the experiment. Fish were fed at the maintenance level once a day.

3.3 Methods

3.3.1 Swim test

For the swimming performance, fish were tested in groups of 50 with a Brett-type, portable swim gutter (2420 mm length, 950 mm width) to measure U_{crit} (Fig. 3.2A). Water flow was generated by an electrically powered propeller. Minimum water depths in the propeller chamber were held at 40 cm at low flows and 50 cm at high flows, to prevent the suction of air into the propeller chamber which could cause cavitation and irregular velocities. Swim tests consisted of fish swimming for exactly 30 min at a starting motor speed of 10 Hz and a flow speed of 45 cm s – 1. The motor speed was then increased by 2 Hz every 30 min until reaching a maximum of 26 Hz

and a flow speed of 104 cm s − 1. Fish were immediately removed from the gutter upon fatigue when they lay flat and still against the back fence of the swim chamber, and the fatigue time was recorded individually. Fatigued fish were allowed to recover, were lightly sedated, weighed, tissue sampled, measured for standard length (SL), and PIT tagged before being placed in a recovery tank. Each day, 2 groups were tested. It took 12 days for the swimming test to be completed. U_{crit} was calculated as:

$$U_{crit} = U_{-1} + (t/\Delta t)\,\Delta U / SL$$

where U_{-1} is the highest velocity (cm/s) maintained for the prescribed period, t is the time until fatigue from the start of the final velocity (min), Δt is the time increment (min), and ΔU is the velocity increment (cm/s). The U_{crit} (SL/s) is the critical swimming speed (cm/s) divided by standard length (cm) (Brett, 1964).

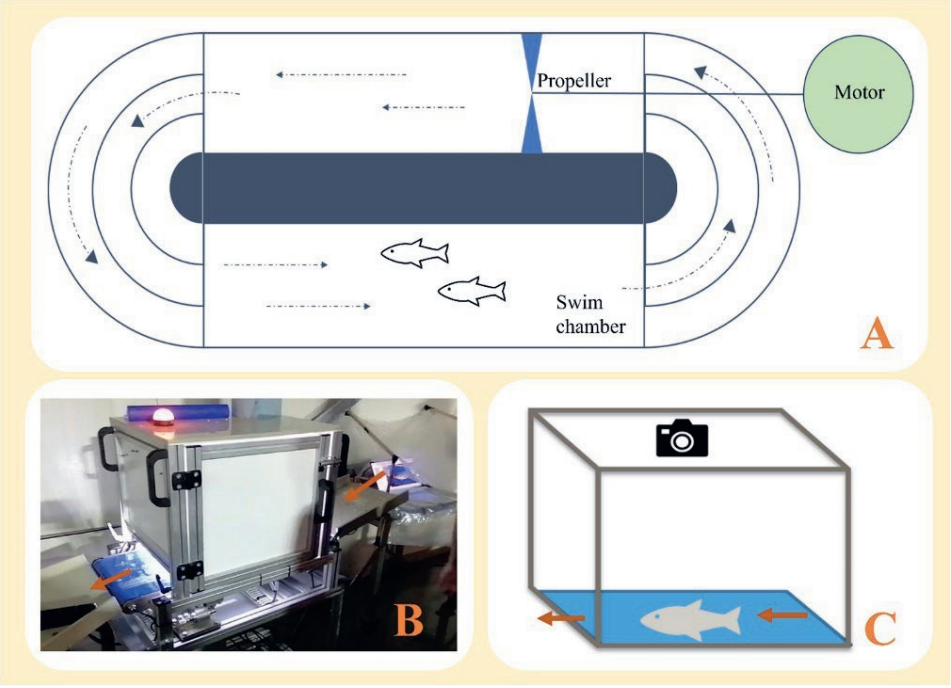


Figure 3.2 Illustration of experimental equipment in swim test and imaging session. A: Graphic top view of the swim gutter. The dotted arrows show the direction of the water flow generated by the motor-controlled propeller.; B: The imaging cabinet. The orange arrows indicate the directions in which a fish was sent in and removed from the cabinet by the conveyor belt at the bottom; C: Graphic representation of the inside of the imaging cabinet, with the camera mounted on top. Sedated fish was laid left lateral side up with the direction of movement of the conveyor belt.

3.3.2 Imaging

14 days after the first swim test, all fish were subsequently imaged using a 50 cm high imaging cabinet (Fig. 3.2B). The cabinet had white plastic walls and ceiling, and a blue anti-reflective plastic conveyor belt that moves the fish into the imaging box. LED strips (OSRAM GmbH, Munich, Germany) inside the cabinet provided

consistent white light. The ceiling housed an intel® RealSense™ D435 Depth Camera (Intel Corp., Santa Clara, California, USA), positioned to face the blue bottom (Fig. 3.2C). Fish were placed individually with the left lateral side facing up and captured simultaneously in two images: one RGB image and one depth image. Both images had a resolution of 1280 × 720 pixels and were registered under the same ID. PIT tag information for 163 fish out of the initial 1200 swim tested was lost and their images were excluded. The final dataset contained 1037 rainbow trout, each with a unique ID, two images (colored and depth), body length, body weight, and a record of U_{crit}.

3.3.3 Image preprocessing

All images underwent preprocessing for dimension reduction to improve computing cost. Images were cropped into 400 × 200 pixels, with fish present around the center. The colored image was combined with the depth image per fish. Colored images consist of three channels: hue, saturation, and value, of which hue does not encode any morphological information. Therefore, the hue channel was excluded and replaced by depth in the merging of the two image sets. This resulted in a pseudo image that retains the same level of information as a 3D image with only three channels in a 2D format (Fig. 3.3).

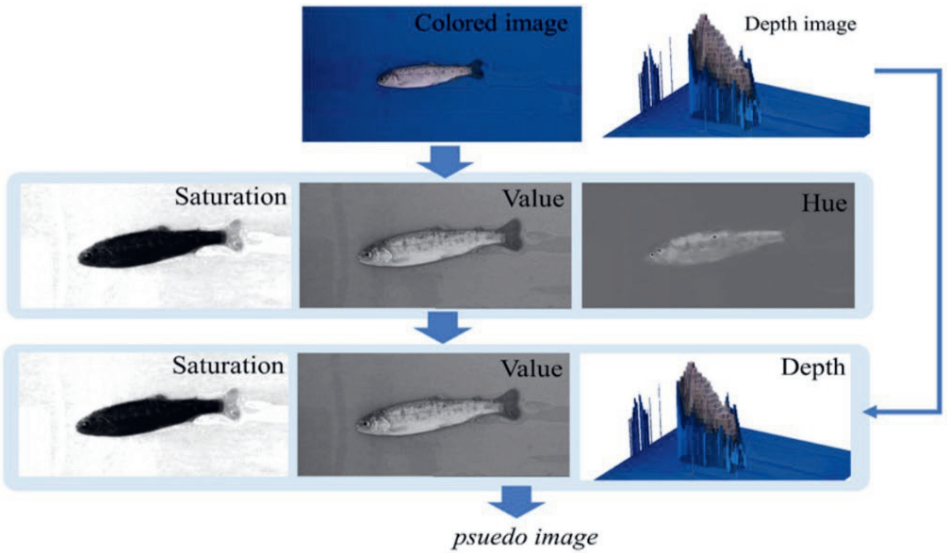


Figure 3.3 Dimension reduction. The hue channel of the colored image was replaced by depth information, resulting in a pseudo image.

3.3.4 Convolutional neural network

Convolutional neural networks (CNNs) are a type of deep learning model that takes images as input for prediction. The applied CNN in this study comprises 5 convolutional blocks, with 24, 36, 48, 64, and 64 layers per block, respectively. A dropout rate of 0.3 follows each convolutional block. The dataset of 987 pseudo

images was divided into training, validation, and test sets following a ratio of 3:1:1. The training and validation of the CNN used the mean squared error (MSE) as the loss function and the correlation coefficient (r) as metric. The input images were multiplied by the weights in the convolutional layers and transformed into information-dense features. To find the weights and features that minimize the loss function, CNN uses backpropagation to compute the gradient of a loss function with respect to the current weights per layer and to update the weights along the steepest descent direction of the gradient. This optimization method is called gradient descent.

The CNN was built under Python 3.8 (Van Rossum, Drake, & others, 1995) with Tensorflow 2.3 (Abadi, et al., 2016). The training and validation were performed on an HPC cluster with one GPU (NVIDIA V100 Tensor Core). The average training time of NN was 6 to 8 h per 100 epochs, with the aforementioned structure and configuration.

3.3.5 Class activation map

We employed the gradient-weighted class activation map (GradCAM) (Selvaraju, et al., 2017) to visualize the learning process of the prediction model. GradCAM utilizes gradients of predictions with respect to the weights and features in the final convolutional layer to generate a heatmap. This heatmap localizes predictive features to different regions of the input image with color coding. On a color scale from red to blue, regions overlapping with features associated with a steeper descent in gradient appeared redder (Fig. 3.4). A steeper descent in gradient signifies higher predictive capability for Ucrit.

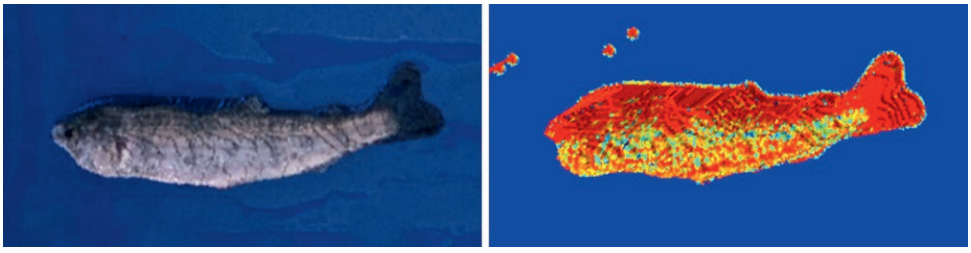


Figure 3.4 Left: visualization of an example of the original 3D image as a reference; Right: visualization of GradCAM on the original 3D image. In this example, the red region in GradCAM on the right corresponds to several regions in the original image on the left: the body outline, the head, the caudal fin, and a region extending along the dorsal side, overlapping with the epaxial muscle region.

3.3.6 Definitions of swim traits

We associated the imaginal regions highlighted by GradCAM with morphological regions that are physiologically related to U_{crit}. Based on physiological interpretation, the general red regions were refined into three specific regions: the head, caudal fin, and epaxial muscle. Prior to annotation, we performed object segmentation to remove the background of both colored and depth images (see Supplement). The

regions of the head and caudal fin were then annotated on the colored image (Fig. 3.5A & B). To annotate epaxial muscle, the value of the pixels of the fish body in the depth image was standardized between 0 and 255, resulting in a grayscale contour map (Fig. 3.5C). The ventral edge of the epaxial muscle was identified by finding the lowest gradient magnitude on the lateral side of the fish (Fig. 3.5D). We later removed overlapping regions with the head and caudal fin based on annotations from the colored image, resulting in an approximated region of the epaxial muscle (Fig. 3.5E).

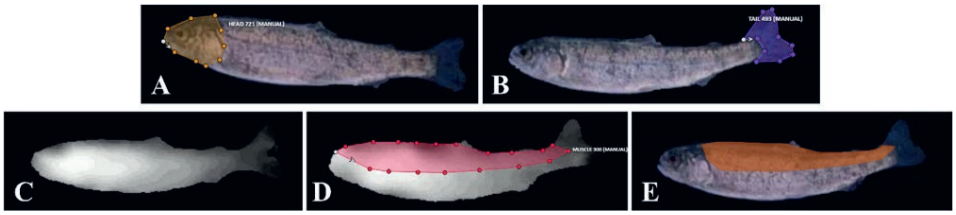


Figure 3.5 Examples of the annotations of morphological regions. A: Head region; B: Caudal fin region; C: Standardized grayscale depth image; D: Grayscale annotation that covers the area of the epaxial muscle and part of the head region. E: Epaxial muscle region.

The regions were manually annotated using key points on images for all individuals using CVAT.ai (Sekachev, et al., 2020). Visual inspection during the manual annotation process guaranteed that all key points were annotated precisely. This was needed because the small size of the fish resulted in low resolution in the depth images. The annotation result was exported as an extensible markup language (XML) file including for each fish their ID, and the coordinates of key points on each image.

To further interpret the prediction model based on fish physiology, we extracted in total four phenotypes that are associated to U_{crit}, which we call swim traits: head volume, caudal fin volume, epaxial muscle volume and epaxial muscle shape. Head, caudal fin, and muscle volume were calculated using the annotated regions and their corresponding depth values from the depth image. The shape of the epaxial muscle was described as the coefficient of variation (CV) of depth values in the epaxial muscle region. The CV measures the relative variability of a distribution, calculated as the standard deviation divided by the mean. A higher CV indicates greater variability within the distribution, allowing for more extreme values and a faster decline moving away from the maximum (Fig. 3.6). Therefore, a higher CV in the depth values in the epaxial muscle region indicates a steeper transition from the furthest lateral points to the dorsal side, resulting in a thicker epaxial muscle.

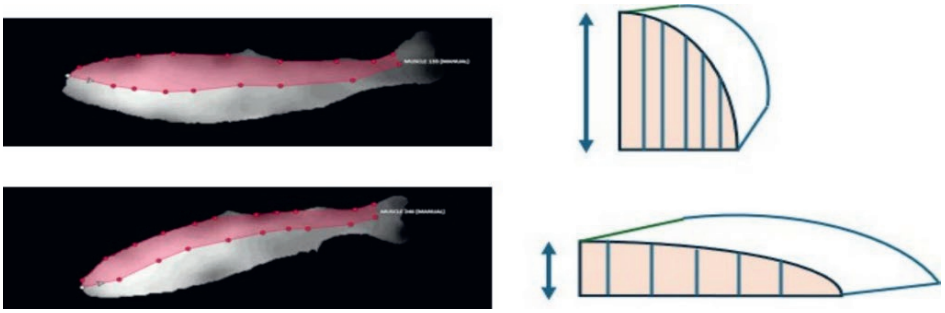


Figure 3.6 Comparison of 2 different epaxial muscle shapes using coefficient of variation (CV) of depth values. The blue lines indicate the depth measurements that each muscle region consists of. If the muscle is laterally flatter, the blue lines are of similar values, indicating lower CV of all depth values within this region. Therefore, a higher CV in depth results in a thicker muscle.

Besides the four swim traits, the volume of the region that was not highlighted in the image was calculated by subtracting the volume of swim traits from the body volume for each fish.

3.3.7 Animal model

We calculated the value of swim traits for all individuals (n = 1037) from the annotated images using OpenCV (Bradski, Kaehler, & others, 2000) in Python 3.8 (Van Rossum, Drake, & others, 1995). 50 fish with body weights that were more than 4 standard deviations from the mean were removed as outliers, resulting in 987 individual records. All traits were standardized with a mean of 0 and a standard deviation of 1. The outlier removal and standardization were performed using R 4.2.1 (R Core Team, 2019).

Genetic parameters were estimated using ASReml-R 4.2 (Butler, 2021). We used the univariate linear mixed model:

$$y = X\beta + Zv + e$$

to estimate the heritability of all traits, where y is a vector of phenotypes, β is the vector of fixed effects with the mean and for the swim traits the covariate body volume, υ is the vector of random animal additive genetic effects ~ (0, $A\sigma_a^2$), where A is the relationship matrix and σ_a^2 is the additive genetic variance of the trait, and e is the vector of random residual effects ~ (0, $I\sigma_e^2$), where I is an identity matrix and σ_e^2 is the residual variance of the trait. X and Z are design matrices that relate observations to the fixed and additive genetic effect of animals, respectively. The phenotypic (r_p) and genetic (r_g) correlations of U_{crit} with swim traits were estimated using the bivariate model:

$$y = X\beta + Zv + e$$

where y is the concatenated vector of phenotypes of the two traits, υ is the vector of random additive genetic effects $\sim \left(\begin{bmatrix}0\\0\end{bmatrix}, A \otimes \begin{bmatrix}\sigma^2_{a,1} & r_{a,12}\sigma_{a,2}\sigma_{a,1}\\ r_{a,12}\sigma_{a,2}\sigma_{a,1} & \sigma^2_{a,2}\end{bmatrix}\right)$, where A is the genetic relationship matrix, $\sigma^2_{a,1(2)}$ is the additive genetic variance of trait 1(2), $r_{a,12}$ is the genetic correlation between traits 1 and 2, and e is the vector of residual effect $\sim \left(\begin{bmatrix}0\\0\end{bmatrix}, I \otimes \begin{bmatrix}\sigma^2_{e,1} & r_{e,12}\sigma_{e,2}\sigma_{e,1}\\ r_{e,12}\sigma_{e,2}\sigma_{e,1} & \sigma^2_{e,2}\end{bmatrix}\right)$, where I is an identity matrix, $\sigma^2_{e,1(2)}$is the residual variance of trait 1(2), $r_{e,12}$ is the residual correlation between traits 1 and 2. X and Z are design matrices that assign measurements to the fixed and random additive genetic effects, respectively.

In the genetic analysis of U_{crit}, we anticipated date or group as a significant fixed effect. However, during the experiment, bigger fish were prioritized due to the risk of exceeding the maximum body length for the swim test, causing a correlation between groups, testing dates and body weight. As body weight was also correlated with body volume (0.785), we therefore include only body volume as fixed effect for both univariate and bivariate model. The same fixed effects were applied for the genetic analysis of all swim traits.

3.4 Results

3.4.1 Prediction of U_{crit} from images

The CNN was applied to the dataset of 987 fish with their pseudo images and was run for 100 epochs. The learning curve of the CNN shows no sign of overfitting as the loss values on the training and validation set gradually decreased and then stabilized (Fig. 3.7). The final evaluation of CNN on the test set resulted in a correlation coefficient (r) of 0.12.

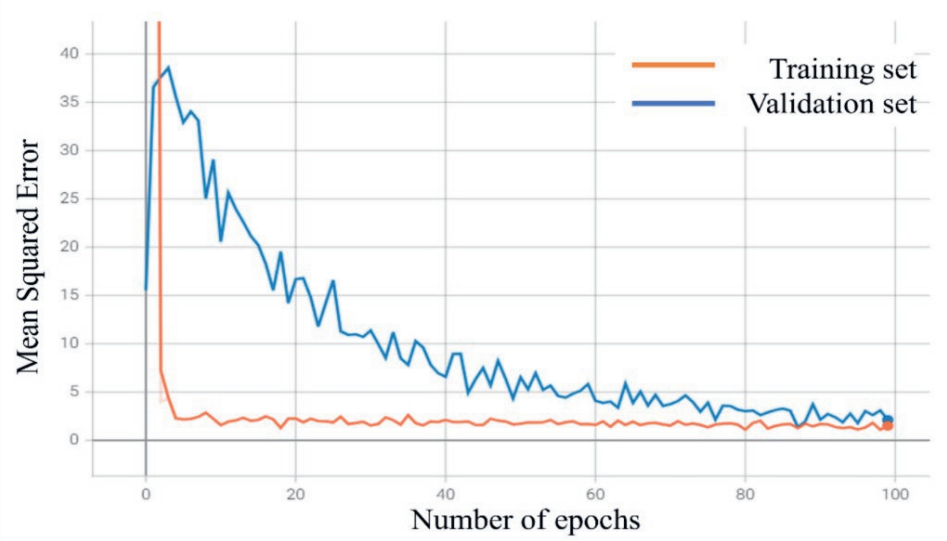


Figure 3.7 The learning curve of CNN using 3D images for U_{crit} prediction. The learning curve represents the performance of the model by calculating the MSE between prediction and real U_{crit} over epochs of training and validation. The training curve is plotted in orange and the validation curve in blue.

3.4.2 Genetic analysis of U_{crit} and swim traits

We derived swim traits from features identified by CNN and highlighted with GradCAM, focusing on the head, caudal fin, and epaxial muscle in the red-coded regions (Fig. 3.8).

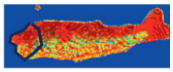
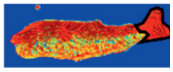
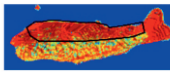
Trait name	Definition	Region in image	Region in GradCAM
Head volume	Volume of the head		
Caudal fin volume	Volume of the tail		
Epaxial muscle volume	Volume of the epaxial muscle		
Epaxial muscle shape	Coefficient of variation (CV) of the depth values of all pixels in the epaxial muscle region		

Figure 3.8 A graphic overview of all swim traits, their definitions and corresponding regions on the RGB image and GradCAM.

The swim traits, together with U_{crit}, were used to estimate their phenotypic and genetic relationship (Tab. 3.1). Phenotypically, only body volume had a significant ($p < 0.05$), negative correlation with U_{crit}. Genetically, all swim traits showed low to moderate heritability. Epaxial muscle volume had the highest negative genetic

correlation with U_{crit}, while caudal fin volume was the only swim trait that positively correlated with U_{crit}. The genetic correlation between the volume of the not highlighted region and U_{crit} was −0.26 ± 0.19.

3.5 Discussion

To our knowledge, this is the first animal study that evaluates the results of CNN and GradCAM by applying domain specific knowledge of physiology and a quantitative genetic method. Based on our methods and findings, there are several interesting points of discussion.

First, we examined the CNN model's performance. The Pearson correlation (r) between U_{crit} and the CNN- predicted values was 0.12. We attribute the low r to the biological complexity of U_{crit}, which is also influenced by multiple factors such as temperature and handling. The variance in U_{crit} is unlikely to be fully explained by morphological variation alone. Consequently, an image-based model may not be expected to achieve a high Pearson correlation with actual measurements. While incorporating other factors in addition to the images could improve r, this study does not aim to maximize prediction accuracy. Instead, our goal was to assess whether a prediction model, visualized through GradCAM, could assist in forming hypotheses on relevant morphological traits that affect U_{crit}. Despite the low Pearson r, GradCAM provided meaningful validation, as the swim traits exhibited stronger genetic correlations (absolute values between 0.38 and 0.48) with U_{crit} than the not highlighted region (r_g = -0.26±0.19).

The interpretation of CNN model results requires domain knowledge. While GradCAM efficiently identifies image regions associated with U_{crit}, understanding their biological mechanism required physiological interpretation, manual annotation, and genetic analysis. These steps show that the highlighted regions may be causally linked to U_{crit} rather than coincidentally associated, eventually leading to the definition of swim traits.

Table 3.1 Results of genetic analysis of U_{crit} and swim traits. Diagonal: heritability. Above the diagonal: phenotypic correlations. Below the diagonal: genetic correlations. Plus-minus sign (±) indicates the standard error of the estimate.

	Ucrit	**Body volume**	**Head volume**	**Caudal fin volume**	**Epaxial muscle volume**	**Epaxial muscle shape**
Ucrit	**0.24±0.07**	-0.05±0.02	-0.007±0.017*	0.006±0.02*	0.003±0.01*	-0.04±0.02
Body volume	-0.23±0.18	**0.60±0.10**	0.44±0.02	0.28±0.02	NE**	0.45±0.02
Head volume	-0.40±0.20*	0.97±0.02	**0.35±0.09***	-0.12±0.01*	-0.01±0.009*	-0.07±0.01*
Caudal fin volume	0.38±0.21*	0.48±0.19	-0.41±0.19*	**0.23±0.07***	-0.02±0.01*	0.01±0.01*
Epaxial muscle volume	-0.48±0.29*	NE**	-0.88±0.21*	0.59±0.44*	**0.07±0.04***	NE**
Epaxial muscle shape	-0.42±0.19	0.97±0.02	-0.90±0.02*	0.55 ±0.26*	NE**	**0.37±0.08**

*Estimated with body volume as the fixed effect.

**NE: Not estimated. Models did not converge.

Swim traits provide better understanding on the genetic basis of U_{crit}, as they explain additional genetic variation in U_{crit} compared to body size alone. After adjusting for body length, U_{crit} demonstrated a moderate heritability of 0.24±0.07, consistent with findings in other aquacultural species like Nile Tilapia (*Oreochromis niloticus*) (Mengistu, et al., 2021) and European sea bass (*Dicentrarchus labrax*) (Vandeputte, et al., 2016). The strongest genetic correlation of -0.48±0.29 was observed between U_{crit} and epaxial muscle volume. Interestingly, all volume-related traits were negatively correlated with U_{crit}, except caudal fin volume which was positively correlated.

Caudal fin volume has a positive genetic correlation of 0.38±0.21 with U_{crit}. After correction for body volume, the caudal fin exhibits positive correlations with other swim traits except head volume. We speculate that the positive correlation of caudal fin volume with U_{crit} is due to its biological relevance in swimming. Trout tails are characterized roughly by a flat rectangular shape, which facilitates the generation of propulsion through a pushing action that directs water obliquely backward (Videler, 1993). Consequently, a genetically larger caudal fin likely favors better swimming performance.

Head volume exhibited negative genetic correlations with U_{crit} and with all other swim traits, as smaller heads corresponded to better swimmers. As the fish head enters the undisturbed water first, a more voluminous head creates increased drag on the sides, perpendicular to the swimming direction. A faster swimmer would therefore exhibit a flatter and smaller head.

Epaxial muscle volume and shape, which measure different aspects of the same region, revealed interesting insights. Epaxial muscle shape showed a negative genetic correlation of -0.42±0.19 with U_{crit}. This echoes the genetic correlation of epaxial muscle volume (-0.48±0.29), suggesting that laterally thinner fish swim to a higher U_{crit}. This is contrary to the expectation that a relatively larger muscle mass would convey higher swimming ability. The effects of increased drag may outweigh the benefits of additional muscle mass, and additional muscle would only lead to higher U_{crit} if the volume-length relationship remains unchanged.

Notably, both epaxial muscle volume and shape demonstrate higher genetic correlations with U_{crit} compared to body volume. This indicates that there is potential value in investigating additional traits that capture different aspects of the epaxial muscle. Trout utilizes various muscle types for different swimming performances, categorized into prolonged swimming, sustained swimming, and burst swimming (Beamish, 1978). U_{crit} is the indicator of prolonged swimming involving both white and red muscle. Sustained swimming refers to a long period of steady movement powered by red skeletal muscle and aerobic metabolism, whereas burst swimming involves short sprints at maximum speed, utilizing white skeletal muscle and anaerobic metabolism (McKenzie, 2011; Plaut, 2001). The genetic relationships between epaxial muscle traits and swimming speed may differ across these

swimming types. Future studies could incorporate white-to-red muscle ratios to capture more precise traits such as muscle composition, and explore alternative forms of U_{crit}, such as residual U_{crit} (Palstra, Kals, Böhm, Bastiaansen, & Komen, 2020) to account for size-related effect in swimming performance, especially when conducting experiments on fish of larger range of sizes.

Beyond their biological relevance, swim traits hold potential for various domain applications. For aquacultural breeding, the genetic relationships between swim traits and U_{crit} provide new direction for selection. However, careful consideration is required when including U_{crit} in the breeding objective. Breeding for increased U_{crit} should proceed with caution due to its moderate negative genetic correlation with epaxial muscle volume. Prioritizing higher U_{crit} may reduce epaxial muscle size and overall body volume which contributes to body weight and fillet yield, both important production traits. Conversely, direct selection on U_{crit} would be more effective for genetic improvement of swimming performance given its heritability of 0.24±0.07, which is relatively higher compared to the expected correlated response from selection on swim traits.

More accurate assessment of the estimated breeding value (EBV) of U_{crit} requires more frequent measurements and might demand additional manual effort to ensure accuracy. Additionally, swim tests are constrained by fish size and age. A higher precision of swim trait measurements through non-invasive imaging techniques could improve EBV accuracy. In this study, swim traits had to be annotated and calculated manually due to the quality of depth image, which was used to annotate epaxial muscle. As the fish were small in volume, the depth image has a low resolution which is visible in Figure 3.5C. Using deep learning for annotation was too risky to get an accurate measurement, which would sequentially be a problem for the genetic analysis. However, with more time and effort to collect images, a subset of manually annotated images could be used to train and fine-tune deep learning models, such as YOLO (Redmon, Divvala, Girshick, & Farhadi, 2016) for landmark detection or region segmentation algorithms like SAM (Kirillov, et al., 2023), further improving precision in swim trait measurements.

This study challenges the prevailing focus on prediction accuracy, where aquacultural research often prioritizes collecting more data to improve models. In recent years, deep phenotyping has emerged as a valuable approach for directly measuring key physiological indicators, such as U_{crit} in swim tests. Beyond physiology, deep phenotyping is increasingly used for comprehensive assessments of metabolism and health, aided by sensor technologies and computer vision. Machine learning, a form of predictive AI, has become a powerful statistical tool capable of integrating diverse data sources to model metabolic and health traits. While these AI-driven approaches show promise, they should not replace conventional measurements such as swim tests, which, despite being labor-intensive, remain essential for reliable metabolic assessments. Instead, AI should complement

traditional methods, providing additional insights rather than substituting well-established techniques.

Through our workflow of data collection, model construction, visualization, interpretation, definition, and evaluation, we demonstrate how genetic analysis of relevant traits can be independent from reliance on highly accurate prediction models, with the help of GradCAM and domain knowledge. Shifting the focus from predictive accuracy to genetic component estimation could also avoid the risk of overfitting models to environmental variances.

3.6 Conclusion

Advancements in technology have enabled extensive data collection, while traditional hypothesis-driven approaches risk overlooking important details and may be influenced by experiential bias. This study demonstrates how an image-based deep learning model can generate data-driven hypotheses and explain critical swimming speed (U_{crit}) in relation to morphology. Genetic analysis not only unravels the genetic links of U_{crit} with fine-scale morphological traits but also validates the physiological interpretations derived from Explainable AI results. The overall interdisciplinary workflow highlights the potential of deep learning to discover novel traits and enhance the understanding of complex traits in aquaculture.

Acknowledgement

This project receives financial support from Topsector Agri & Food. The use of the HPC cluster has been made possible by CAT-AgroFood (Shared Research Facilities Wageningen UR).

Supplement

Object segmentation - RGB (red, green, and blue) images were converted to HSV (hue, saturation, and value) format to segment the fish from the blue background using the following thresholds: lower: 10, 188, 83; upper: 179, 25, 255. for hue, saturation, and value, respectively. Gaussian blur, dilation, and Canny edge detection were applied to refine the segmentation and remove background noise. The outlined fish region was copied from the colored image to a blank image of the same dimensions where the value of all pixels was 0 (Fig. 3.9). The same segmentation was performed on the depth image. The initial object segmentation was performed on the colored images, and then the coordinates obtained from this segmentation were used to locate and segment the fish in the corresponding depth images. This approach works because both the colored and depth images share the same dimensions. All preprocessing was done with OpenCV in Python 3.8.

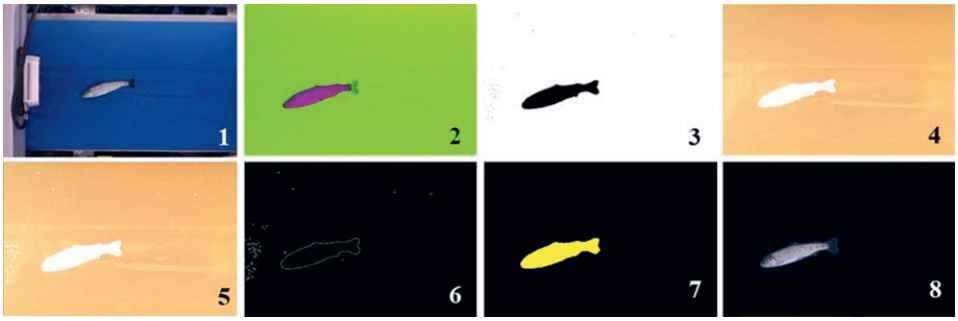


Figure 3.9 Overview of the image preprocessing using a single image as an example. 1. The original colored image; 2. Transformation from RGB to HSV color space; 3. Removal of the background by applying specific thresholds (lower: [10, 188, 83], upper: [179, 25, 255]) in the hue, saturation, and value channels; 4. Setting the background value to 0 using a bit-wise transformation; 5. Enhancing the main object's edges and background noise through Gaussian blur and dilation; 6. Detecting all contours using Canny edge detection; 7. Creating a mask with the largest contour; 8. Segmenting the fish from the original colored image using the generated mask.

4. Individual re-identification of Atlantic salmon (*Salmo salar*) using images taken 5-months apart

Yuuko Xue[1], Gert Kootstra[2], Hans Komen[1] ,Stephen Tapping[3], John W.M. Bastiaansen[1]

[1]Department of Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands

[2]Department of Agricultural Biosystems Engineering, Wageningen University & Research, P.O. Box 16, 6700 AA Wageningen, The Netherlands

[3]Landcatch Natural Selection Ltd., Ormsary Fish Farm, PA 31 8PE, Lochgilphead, Argyll, Scotland

Abstract

This study investigates whether image-based re-identification (re-ID) of Atlantic salmon is feasible under realistic aquaculture breeding conditions using Siamese network. Two images of ~1,500 fish were taken, five months apart. Images were preprocessed and augmented, and three regions of interest (ROIs) were extracted. Re-ID was performed using Siamese networks with InceptionV3 and EfficientNetB7 architecture trained via triplet loss. Despite reproducing earlier models, top 1 accuracy reached only 26.2% using an image-level split, and performance fell to near-random levels when using ID-level splits. We suggest that previous benchmarks of 98% were overly optimistic for re-identification of large numbers of fish and long intervals between images. Results identify environmental variance, particularly lighting, and limited phenotypic stability as major challenges, and highlight the need for standardized imaging protocols and phenotypic features with longitudinal stability to enable effective image-based re-identification in aquaculture systems.

Key words: Individual re-identification; Siamese network; Convolutional neural network; Triplet loss; Triplet variation; image quality

4.1 Introduction

In aquaculture, phenotypes are typically measured at the individual level for breeding and sometimes also for production management. Accurately recording multiple measurements for each individual enables essential practices such as tracking growth, assessing maturation, and monitoring health and welfare parameters. For selective breeding, individual identification is crucial for connecting phenotypes to the pedigree, and for selection of candidates after breeding value estimation (Rasal, Patnaik, Murmu, Sundaray, & Das Mahapatra, 2021).

Currently, re-identification is achieved through the use of Passive Integrated Transponder (PIT) tags, which assign each fish a unique ID. Upon retrieval, PIT tags are scanned to re-identify the same individual. While PIT tags are efficient and widely used, their insertion and manual reading are labor-intensive and stressful for the fish. Studies have also reported adverse effects such as increased mortality (Macaulay, et al., 2021; Vollset, et al., 2020). Under challenging environmental conditions, reading PIT tags can be difficult, and tag loss may occur (Thorstad, Rikardsen, Alp, & Økland, 2013). Consequently, researchers are exploring less invasive alternatives for re-identifying fish.

With the developments in computer vision, image-based individual re-identification is becoming increasingly popular. Images have been used to measure morphological traits (Yu, et al., 2021), and to predict invasive or complex trait based on fish's exterior features (Xue, et al., 2025; Prchal, et al., 2020; Vandeputte, et al., 2017). Some exterior features, such as scale shape (Huntingford, Borçato, & Mesquita, 2013) or melanin dots (Merz, et al., 2012), can exhibit subtle, individual differences that can potentially be used for re-identification. Building on research in human re-identification, which focuses on features such as facial structure, posture, and appearance like body shape and clothing, animal sciences have started to adopt vision-based technology in animal health and production systems (Schneider, 2020), including aquaculture. This method has shown success in several marine species, including giant sunfish (*Mola mola*) (Pedersen, Nyegaard, & Moeslund, 2023), basking shark (*Cetorhinus maximus*) (Gore, Frey, Ormond, Allan, & Gilkes, 2016) and undulate skate (*Raja undulata*) (Gómez-Vargas, Alonso-Fernández, Blanquero, & Antelo, 2023).

Several studies have explored image-based re-identification in common aquaculture species such as Atlantic salmon (*Salmo salar*), European seabass (*Dicentrarchus labrax*), brown trout (*Salmo trutta*) and brook trout (*Salvelinus fontinalis*) (Bekkozhayeva & Cisar, 2022; Zhou, Hitt, Letcher, Shi, & Li, 2022; Cisar, Bekkozhayeva, Movchan, Saberioon, & Schraml, 2021; Pedersen & Mohammed, 2021; Zhao, Pedersen, Hardeberg, & Dervo, 2019). However, the applicability of these methods to real-life scenarios remains limited due to either being successful only when a small number of individuals is involved or when images are collected frequently at short time intervals. To be useful, image-based re-identification must

be successful when the application is aligned with common aquacultural practices. For example, in breeding programs, images are typically captured during routine phenotyping to minimize handling stress. These sessions occur with intervals of several to many months, depending on species and age, and are applied in groups consisting of thousands of fish.

This study investigates whether image-based re-identification is feasible in a real-life aquaculture breeding scenario. To date, the FishNet model by Mathisen et al. (2020) has achieved the highest reported accuracy (98%) for re-identification on 715 individuals taken at a short time interval (up to 9 minutes). The longest documented interval between imaging sessions was 10 months, reported by Stien et al. (Stien, et al., 2017) who achieved accuracy of 85% on 30 individuals. Both studies used images of Atlantic salmon (*Salmo salar*). Building on these earlier efforts, our primary objective is to investigate the accuracy of individual re-identification of over 1000 adult salmon using images taken five months apart.

4.2 Materials and methods

4.2.1 Ethical statement

Image data collection was conducted at Landcatch Natural Selection Ltd, Ormsary, Scotland under the farm's established Veterinary Health Management Plan (VHMP) and in accordance with RSPCA welfare standards. All procedures adhered to the local legislation regarding animal welfare.

4.2.2 Materials

Atlantic salmon (*Salmo salar*) (n = 1,485) were kept in flow-through freshwater for 1 year, before being transferred to seawater in land-based tanks. Eggs were fertilized in December 2020 and grow until 12 months old when they were individually handled for weight and length measurements. Afterwards, they underwent 24 hours light regime for 6 weeks to induce smoltification. Once smoltification was confirmed by blood chloride testing, fish were transferred to broodstock tanks and further distributed into more tanks to reduce stocking density. Two imaging sessions took place during such moments of redistribution: the first time was 10 months after transfer to seawater, when also fin clipping was done for genotyping, and the second time was after 15 months.

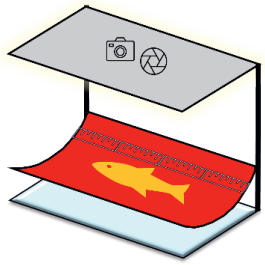


Figure 4.1 Illustration of the phenotyping device.

Salmon were imaged with a purpose-built portable device (Fig. 4.1). The exterior of the device, both the bottom and top module, is made of high-density polyethylene to provide a non-stick, low-friction surface that prevents injury to the fish. The top edge of the surface also had a numerical scale to provide quick reading of the body length

of the fish. The top module is equipped with a HQ camera (CAM) consisting of a 12 Megapixel sensor with fixed zooming and a vision module capturing the entire bottom surface of the device with a field of view of 1.1 m. The bottom surface is a red, curved tray to offer a contrasting background to the fish.

4.2.3 Methods

Imaging

Before the images were taken, fish were handled first into a sedation container with Tricaine mesylate and, once in a much calmer docile state, placed onto the imaging device and then into the non-crowded area of the tank to recover from anesthetic effects until fully conscious. During the operation, each fish would have been handled by 3 different people, for checking the anesthetic status, placing on and removing from the imaging device, respectively. With the imaging device, fish were out of water for only 12 seconds, and the crowding took less than 3 hours in total, in accordance with the RSPCA standards. After imagining the fish were placed back into clean seawater.

Images collected from the device were named after the PIT tag per fish and stored per session. The dimension of images was 4056x3040 pixels. In total, 1,485 images were collected in February session (10-month after transferred to seawater) and 1,478 in July session (15-month after the transfer).

Image preprocessing

Image preprocessing was performed using OpenCV (Bradski, Kaehler, & others, 2000), a computer vision library in Python 3.9, and included the following methods (See also Table 4.1, row 1-3):

Cropping

In addition to the fish, the raw images included a screen display on the top of the view. This part of the image was cropped out using a fixed ratio of the height of the image.

Distortion correction

The camera lens introduced perspective distortion toward the edges of the images. Since the lens parameters were not available, distortion was corrected using a transformation matrix, with parameters empirically determined through iterative experimentation and visualization. The numerical scale on the top edge of the image was used as a reference. Correction was considered finished once the scale was visually in parallel to the length of the image.

Segmentation

Fish were segmented from the background using the Segment Anything Model (SAM) (Kirillov, et al., 2023), which generates regional masks by detecting differences in color, texture, and other imaginal features. A bounding box was then derived from

the mask of the fish, defined by x (longitude) and y (latitude) coordinates, with its edges tightly enclosing the segmented fish. The segmented fish were then placed on a 3600 × 1600-pixel black background, aligning the bounding box center with the center of the new image. Segmentation failed in cases where fish were not visible due to handling or extreme lighting conditions. These images were subsequently removed after visual inspection.

The segmented images were also used to calculate the body size of each salmon in pixel number as part of exploratory data analysis. Salmon were expected to exhibit increased body size after five months of growth, except in cases of deformity or early maturation. Individuals showing signs of either condition were excluded from the dataset, as they can significantly alter external phenotypes.

Data quality assessment

The images were captured outdoors in two sessions separated by 5 months. As a result, lighting conditions varied significantly between sessions. Also, within each session conditions varied due to changes in natural daylight. To mitigate the impact of these variations, we performed a quality assessment of each image using its color properties. Specifically, we computed the average hue, saturation, and lightness (HSL) values per image. Images who's average HSL values deviated by more than three standard deviations from the session mean were identified as outliers and excluded from the dataset.

Data augmentation

We applied several data augmentation techniques (Shorten & Khoshgoftaar, 2019) as each fish was only imaged twice.

Artificial background

The segmented fish image was merged with an underwater image, where the black background was replaced with an underwater image. The underwater background image was obtained from a public domain repository, with the original source no longer accessible. Then the brightness of the image was adjusted by taking the average brightness of the fish image and the background image (Tab. 4.1, row 4).

Linear combination

Two images of the same fish with random backgrounds were aligned using the left edge of the bounding box. The bounding box of a fish image from July was resized to match that of the February image. Both images were then divided into 100-pixel wide vertical strips. Two linear combination augmented images were created in this manner, by taking the odd number strips of one image and the even number strips of the other to construct a new image, and vice versa (Tab. 4.1, row 5).

Taking average images

Two images of the same fish with random backgrounds were aligned using the center of the bounding box of the fish and resized to match the larger bounding box. Then

an average image was generated by taking the average value of the two images across all pixels (Tab. 4.1, row 6).

After data augmentation, the number of images per fish increased from 2 to 9. This includes 2 original images, 2 with artificial background and 2 with black backgrounds, 2 images generated through linear combination and 1 average image.

Regions of interest

Three ROIs were extracted separately from all the images. While using the whole fish image provides more information on the external pattern, the size of the image increases computational cost, and resizing to smaller size compromises resolution. Previous research (Cisar, Bekkozhayeva, Movchan, Saberioon, & Schraml, 2021; Stien, et al., 2017) suggested that melanin dots patterns on the operculum and the dorsal side were informative for re-identification. Therefore, we extracted the following regions (Tab. 4.1): **a)** Anterior 50% of the whole fish, **b)** Anterior 10%, including the head and part of the pectoral fin and **c)** Operculum region only. ROIs were identified and cropped based on the bounding box coordinates and the total fish length, which was measured by the length of the bounding box. Apart from improving computational efficiency, the use of ROIs also helped the algorithm focus on the more informative regions in the images.

Table 4.1 Data augmentation with example images.

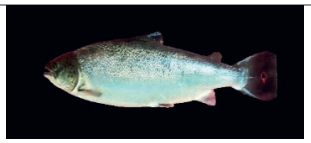
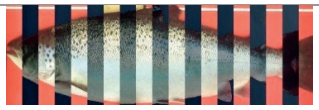
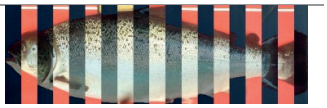
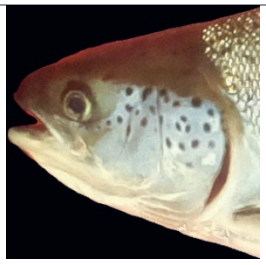
Original image (Left: February; Right: July)		
Cropping & distortion correction		
Segmentation		
Artificial background		
Linear combination		
Average image		

Table 4.2 Regions of interest identified and cropped.

a) Anterior 50%	b) Anterior 10%	c) Operculum region
		

Siamese network

A Siamese network (Chopra, Hadsell, & LeCun, 2005) consists of two identical neural networks with shared weights, designed to learn a similarity metric between image pairs. Unlike traditional classification or regression models that make predictions based on images, Siamese networks learn to distinguish between images of different objects (individuals in this case). The images are mapped into a high-dimensional embedding space, where the proximity of embeddings reflects their image similarity.

To ensure the model generalizes well to new data, either a new image of an already known fish or images of a new fish, we applied two different data splitting approaches: image-level split and ID-level split. For image-level split, data augmentation was performed first, followed by random splitting of all images into training, validation and tests, similar to the approach in FishNet (Mathisen, Bach, Meidell, Måløy, & Sjøblom, 2020). For ID-level split, data was first split according to the individual fish IDs before applying augmentation, so that no images of the same fish appeared in both training/validation and test sets. For both methods, data was divided into 60% training, 20% validation, and 20% test sets.

The training of Siamese networks requires image triplets. Each triplet contains: a reference image of an individual fish (anchor), another image of the same fish (positive), and an image of a different fish (negative). The Siamese network is trained to minimize the distance between embeddings of the anchor and positive pair while maximizing the distance between embeddings of the anchor and negative pair. This forces the model to learn meaningful, discriminative features that remain consistent across images of the same fish while differing between images of different fish. For this, we used the triplet loss function, defined as:

$$Loss = \sum_{i=1}^{N} \max\left(d(A,P) - d(A,N) + \alpha, 0\right)$$

Where $d(A,P)$ is the Euclidean distance between the anchor and the positive image of the same fish, $d(A,N)$ is the Euclidean distance between the anchor and the negative image of a different fish, α is a margin value that ensures a sufficient separation between positive and negative pairs, and was set to 0.2, a value commonly used in re-identification tasks, and the max function ensures that the weights are updated until $d(A,P)$ remains smaller than $d(A,N)$ by the margin. A well-trained model should eventually produce embeddings where images of the same fish cluster together, while those of different fish are well-separated in the high-dimensional space (Fig. 4.2).

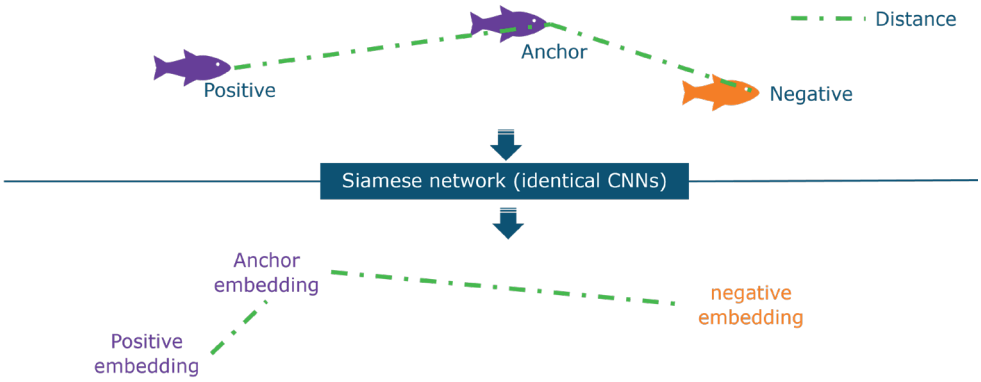


Figure 4.2 Illustration of Siamese network and its function.

We tested two convolutional neural network (CNN) (LeCun, Bengio, & Hinton, 2015) architectures to generate embeddings: Inception V3 (Xia, Xu, & Nan, 2017) and EfficientNet B7 (Tan & Le, 2019). Inception V3 was selected based on the study of FishNet and EfficientNetB7, an upgrade of Inception V3, was chosen because it is one of the most advanced CNN architectures. Both models were initialized with pre-trained ImageNet (Deng, et al., 2009) weights before continuing training on our dataset. In both models, the final classification layer was replaced with a 128-dimensional dense layer, representing the learned embedding vector. To further ensure that embeddings learned by the Siamese network remain meaningful and stable, we applied L2 normalization to the output embeddings by dividing the embedding vector with their Euclidean norm.

The models were built under Python 3.9 with Tensorflow 2.3. Each model was trained for 100 epochs, with a batch size of 32. The training and validation were performed on a high-performance cluster (HPC) using one GPU (NVIDIA V100 Tensor Core).

After training, we evaluate the learned embedding by Euclidean distance using top 1 and top 5 accuracy. A top 1 match is considered correct if the nearest embedding comes from a different image of the same fish. For top 5 accuracy, the result is considered correct if at least one of the five closest embeddings corresponds to the same fish.

Finally, to visualize the decision-making process of the trained model, we employed Gradient-weighted Class Activation Mapping (GradCAM) (Selvaraju, et al., 2017). GradCAM generates heatmaps by utilizing the gradients of predictions with respect to the weights and features in the final convolutional layer. These heatmaps highlight regions in the input image that contributed most to the model's decision. The color coding follows a red-to-blue scale, where red regions indicate areas with the highest importance in determining similarity and blue regions indicate less relevant regions.

By applying Grad-CAM, we could visually inspect whether the model was focusing on regions with biological features, such as the melanin patterns on the operculum or dorsal side.

Triplet structure

One of the main challenges in this study was the limited number of two available images per fish. We first assessed the structure of the dataset by analyzing Euclidean distances between the embeddings of images before training the model. Specifically, for each fish's July image, we computed its embedding using the pre-trained model and visualized its Euclidean distance to the embeddings of all February images.

There were three types of triplets when training with triplet loss function: Easy triplets, where the positive is already closer to the anchor than the negative by the margin, i.e. $d(A,P)$ + margin < $d(A,N)$; Semi-hard triplets, where the negative is closer than the positive pair but the difference is within the margin, i.e. $d(A,P) < d(A,N) < d(A,P)$ + margin; And hard triplets, where the negative pair is closer than the positive pair. To explore the structure of the embeddings in the triplets before training, we visualized the Euclidean distances between the two session images for randomly selected fish. As there is only one positive pairs, we selected both semi-hard and hard triplets for training simply by excluding all easy triplets.

Research strategy

A series of experiments were designed and conducted to investigate how regions of interest (ROIs), convolutional neural network (CNN) architecture of the Siamese network, data augmentation, and dataset splitting strategy affect the performance of image-based re-identification. These experiments are summarized in Table 4.3.

Table 4.3 Summary of re-identification experiment using Siamese network.

EXPERIMENT	ROI	CNN	Augmentation	Data split	Data size
A	Anterior 50%	InceptionV3	No	Image-level	1449
B	Anterior 50%	InceptionV3	Yes	Image-level	1449
C	Anterior 10%	EfficientNetB7	Yes	Image-level	1449
D	Anterior 10%	EfficientNetB7	Yes	ID-level	1449
E	Operculum	EfficientNetB7	Yes	ID-level	1449

Experiment A and B were based on the work of Mathisen et. al (2020), using the same network architecture, InceptionV3. The key difference between these two experiments is the use of data augmentation: **Experiment A** used the original 2 images per fish, while **Experiment B** increased the number of images to 9 through augmentation. Both experiments used the anterior 50% as ROI, slightly broader than the region proposed by Mathisen et. al (2020), to include additional informative regions, such as dorsal patterns, suggested by Cisar et al. (2021).

Experiment C introduced two changes: it replaced the CNN with the more advanced EfficientNetB7 architecture and reduced the ROI to the anterior 10% of the fish. The reduction of the ROI was emphasized in both Mathisen et al. (2020) and Stien et al. (2017). Like A and B, **Experiment C** used an image-level data split.

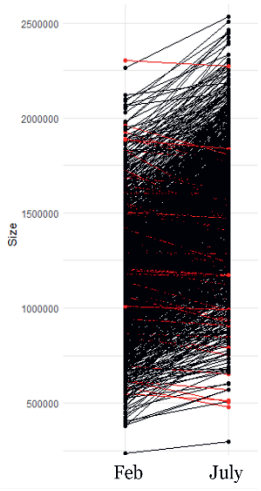
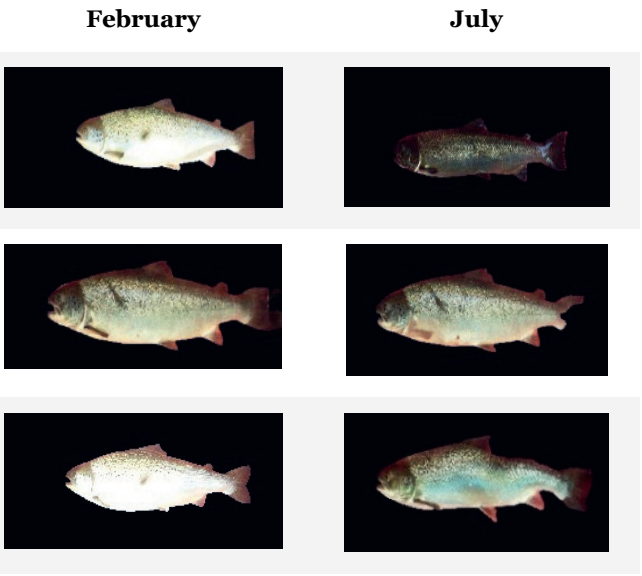
Experiments D and E adopted an ID-level data split, where images of the same individual are not allowed to appear in both training and test sets. **Experiment D** used the same anterior 10% as Experiment C. In **Experiment E** we isolated the operculum region.

4.3 Results

4.3.1 Exploratory data analysis

The biological development of salmon (n=1,485) in 5 months was assessed using image-based measurements (Tab. 4.5). The number of pixels in each segmented fish image was used as a proxy for body weight, given the expected strong correlation between body size and weight (Balaban, Ünal Şengör, Soriano, & Ruiz, 2010). While most individuals showed an increase in body size over the five-month period, some fish showed a reduction. Upon visual inspection, we identified three likely causes: early maturation, lesions or bleeding indicating disease or injury, and deformities (Tab. 4.5, Right). Among these, early maturation and severe deformities visibly altered the fish's morphology and was likely to affect re-identification performance. Lesions and bleeding, on the other hand, did not noticeably affect the melanin patterns on the head and operculum.

Table 4.5 Summary of the exploratory data analysis on fish growth in 5 months. left: graph with the growth trend where x-axis is the time point, and y axis is the size calculated by pixel numbers. right: image example of likely causes of growth size decrease. From top to bottom, each row represents early maturation, lesions and bleeding, and deformity, respectively.

GROWTH TREND	Image example
	

4.3.2 Triplet structure

In total, 36 images were removed as outliers that were overexposed and lacked visible melanin patterns (Fig. 4.3). Afterwards, 1,449 individuals remained in the dataset.

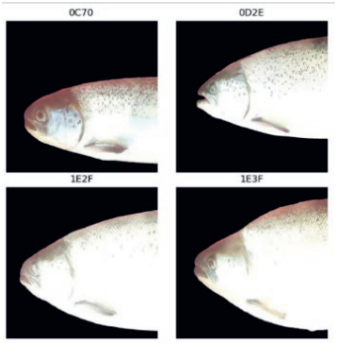


Figure 4.3 Examples of four outlier images that suffer from overexposure.

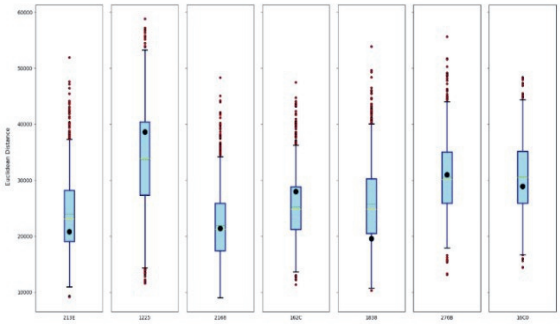


Figure 4.4 Euclidean distances between 7 randomly selected individual images from July and all the images in February. X-axis: image ID, y-axis: Euclidean distance. Black dot: positive image.

To explore the structure of the embeddings in the triplets before training, we visualized the Euclidean distances between July and February images for seven randomly selected fish (Fig. 4.4). In each panel, the y value of the black dot represents the distance between the positive pair (i.e., the same individual in the other session), while the boxplot illustrates the distribution of distances to images of all other individuals in the other session (negative pairs).

4.3.3 Siamese network performance

The results are presented in the order of the experiments, with emphasis on how changes in the experimental setup affected the outcomes.

In **Experiment A**, the network achieved a top 5 accuracy of 3% and a top 1 accuracy below 1%, slightly above the expected random pairing accuracy of 0.35%. To better understand the performance of the model, we generated GradCAM visualizations for both a successful case, where the true match was among the top 5, and a failed case (Fig. 4.5). The successful match appeared to focus on the head and dorsal melanin pattern, whereas the failed case did not utilize the hypothesized regions of interest.

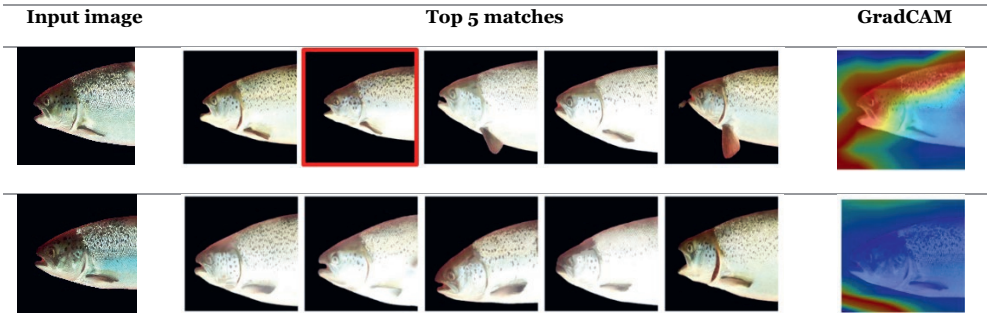


Figure 4.5 First row is a correct pairing example, where the correct match is indicated with the red square and the GradCAM covering mainly the head and dorsal melanin pattern; Second row is an example of failed pairing, no correct match was in the top 5 matches, and the GradCAM was highlighting the background.

In **Experiment B**, after applying data augmentation, the top 5 accuracy increased to 4%, while top 1 accuracy remained unchanged at 1%.

To provide the network with more focused visual input, **Experiment C** used only the anterior 10% of the fish body, combined with data augmentation and a deeper CNN architecture. This led to a substantial improvement, with top 5 accuracy reaching 30.7% and top 1 accuracy 26.2%. GradCAM results highlighted two key regions for successful re-

identification: the area surrounding the eye (including the operculum) and the pectoral fin (Fig. 4.6).

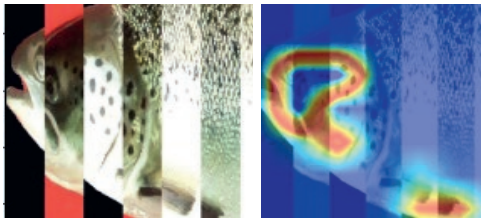


Figure 4.6 An example of the input anterior 10% augmented image and its GradCAM where top 1 match was correct.

We suspected the results of Experiment C might be artificially inflated due to applying an image-level data split. Data augmentation was applied *before* splitting, potentially allowing images of the same individual to appear in both the training and test sets. To address this, **Experiment D** used the same setup as **Experiment C** but performed an ID-level data split to avoid data leakage. Accuracy dropped dramatically, with both top 5 and top 1 accuracy approaching zero, comparable to the 0.5% random pairing level.

In **Experiment E**, we further constrained the input by using only the operculum region, with the objective to reduce potentially irrelevant information other than melanin spot pattern, but performance was not improved.

In summary, the Siamese network was able to re-identify fish when an image level split was applied but was unable to re-identify fish when trained with an ID-level data split. The model failed to exceed random performance across all experiments.

4.4 Discussion

To investigate whether image-based re-identification is feasible in a real-life aquaculture breeding scenario, this study made a two-stage effort. The first stage (Experiments A, B, and C) applied methods from previous successful studies using image data collected in breeding facilities. The second stage (Experiments D and E) aimed to mimic real-life application by testing whether the trained models could be generalized to images of new fish.

Two key results emerged from our findings. First, based on Experiment A to C, we find that the InceptionV3 architecture applied with an image level split of the data may not be sufficient for long-term re-identification. While the same method showed high accuracy before (Mathisen et al., 2020), this was in a dataset that was based on short video clips of less than 9 minutes, while our experimental interval was 5 months. For our dataset, the same network architecture yielded only 4% top 5 accuracy (Experiment B). Upgrading to EfficientNetB7 in Experiment C improved top 5 accuracy to 31% in our dataset spanning over 5 months, yet this remained far below the 98% accuracy previously reported using short video.

Additionally, it is likely that results in experiments A to C and the results by Mathisen et al. (2020) were inflated by data leakage, as an image level data split was performed after data augmentation, not based on the fish ID. This allows near-duplicate augmented images of the same fish to appear in both training and test sets, thus boosting performance.

In experiment D we applied ID level split which dramatically reduced accuracy to near zero. This contrast between image-level and ID-level splits is the second key result and illustrates the difficulty of identifying stable, time-invariant features across individuals. In image-level training, the model can associate identity with features that remain stable across different image conditions (e.g., lighting, background). However, in ID-level training, the model is required to learn and rely on features that are time-invariant across different individuals and imaging conditions. Although Stien et al. (2017) demonstrated that operculum melanin spots can remain stable for up to 10 months, we speculate that variation from uncontrolled imaging conditions between sessions, particularly in lighting, can outweigh this biological stability. As a result, the model may fail to prioritize biologically meaningful features and therefore struggle to generalize to new individuals.

For a Siamese network to succeed, intra-class variation (same individual imaged at different times) must be minimized while inter-class variation (differences between individuals) remains distinct. Ideally, intra-class variation is low because unique biological features should remain visually stable. However, imaging inconsistencies such as angle, lighting, and zoom can inflate intra-class variation, sometimes to the point where it overlaps with or exceeds inter-class variation. When this occurs, the model cannot reliably separate individuals.

Triplet loss addresses this by encouraging positive pairs (same fish under different conditions) to remain close in the embedding space while pushing negative pairs (different individuals) apart (Fig. 4.2). Triplet mining improves this process by selecting semi-hard or hard negatives, which are different individuals that appear similar to the anchor. This forces the network to learn discriminative features under challenging conditions (Lamping, Kootstra, & Derks, 2025). When natural intra-class variation is minimal, as in study using short video sequences where highly similar frames for each individual are produced (Mathisen, Bach, Meidell, Måløy, & Sjøblom, 2020), data augmentation is critical for simulating variation in lighting and angle to improve model's robustness on new data (Kumar, Asiamah, Jolaoso, & Esiowu, 2025; Rebuffi, et al., 2021; Khosla & Saini, 2020).

By comparison, our dataset exhibited substantial intra-class variation (Fig. 4.4) caused by uncontrolled imaging conditions, which likely obscured stable biological features. With only one positive pair per individual, the network had limited opportunity to learn consistent identity representations. Consequently, intra-class variation caused by imaging noise overlapped with inter-class variation driven by biological features, making them inseparable. For example, two different fish could appear identical due to overexposure, while the same fish could appear different across sessions due to lighting changes.

In addition to the stability of image conditions, the effectiveness of biological features also matters. The melanin spot pattern on the operculum is informative for humans to distinguish individuals and has been successfully used in both studies of

Stien et al. (2017) and Mathisen et al. (2020). Given its stability and distinctiveness, we hypothesized that narrowing the ROI to the operculum would enhance model performance by focusing on this biologically relevant feature. Comparison of Experiments A and C suggests a positive effect from focusing on the operculum, but this could also be due to the deeper CNN architecture. We speculate that the model in Experiment C also relied on positional information from nearby landmarks such as the mouth and pectoral fin, as also highlighted in Figure 4.6.

This study combined two challenging factors: a high number of individuals and a long interval between images. High accuracy of re-identification was reported when these two factors were separated. Our results with data from a practical aquaculture breeding environment revealed that image-based re-identification was not yet feasible in this scenario. Nevertheless, our findings provide valuable insights into both data quality and methodological requirements.

Future improvements ultimately center on managing intra- and inter-class variation. A primary step is to enhance image quality by ensuring consistent lighting conditions so that biological features contributing to inter-class variation remain visually stable and unobscured. Increasing the frequency of imaging sessions or capturing multiple images per fish can further reduce intra-class variation by providing a wider range of biological features under varying conditions, potentially with higher longitudinal stability. For this reason, image acquisition protocols are important, in particular under standardized lighting, ideally within a closed imaging system. However, repeated handling of fish for imaging raises ethical and logistical concerns. For re-identification in the domain of marine conservation, at least 8–10 images per individual were recommended for reliable model performance (Bouma, Pawley, Hupman, & Gilman, 2018). Whether similar thresholds are applicable to aquaculture species remains to be researched.

In salmonids, spatial shifts in operculum spot patterns are known (Stien, et al., 2017) and may serve as dynamic biometric markers, provided they are properly quantified and modeled over time. In human facial recognition, changes in identity-related features are often accounted for using functional growth matrices (Ramanathan & Chellappa, 2008). Growth curves and more sophisticated longitudinal patterns could enable the modeling of intra-class variation as a continuous, predictable biological development.

A promising development for the advancement of re-identification is the broader integration of imaging technologies into aquaculture phenotyping. Images already enable automated trait extraction and support longitudinal monitoring of growth and early detection of abnormalities, including lesions and deformities, as illustrated in Table 4.1. Incorporating imaging into routine phenotyping will lead to the accumulation of datasets that are essential for training more robust re-identification models. However, accumulating sufficient data takes time and knowledge on the minimum image quality requirements.

4.5 Conclusion

We remain optimistic that Siamese networks can serve as a reliable method for aquaculture re-identification. Unlike traditional identification models that require labeling for every new individual, Siamese networks learn the similarity between image pairs, enabling them to recognize individuals based on comparative features without then need for any fixed IDs. Therefore, Siamese network is well-suited for aquaculture settings, where new fish are regularly introduced, and exhaustive labeled datasets are difficult to obtain. Moreover, Siamese networks are capable of leveraging a wide range of phenotypic features including visually subtle traits, which means they may also be suitable for fish species that do not exhibit obvious individual differences like spot patterns.

5. End of the tag era? Critical review on vision-based individual re-identification for aquaculture

Yuuko Xue[1], John W.M. Bastiaansen[1], Hans Komen[1]

[1]Department of Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands

Abstract

This chapter critically reviews the development of vision-based individual re-identification technologies in aquaculture as a potential alternative to PIT tags. While the non-invasive nature of vision-based methods is appealing, based on a review of 23 studies, we identify multiple limitations and challenges in current implementations regarding data acquisition, data analysis, and practical implementation. To address these gaps, we recommended a new aquaculture-centered flowchart emphasizing high-tech data acquisition, adapting multi-dimensional data quality frameworks, and benchmarking. Importantly, we call for a realistic view of technology's potential: while vision-based methods hold long-term promise, they are currently insufficient to replace PIT tags. Future advancements depend on coordinated efforts in dataset standardization, methodological refinement, and the alignment of model development with practical aquacultural systems.

Key words: Machine vision; Fish re-identification; PIT tag replacement; Non-invasive; Precision aquaculture

5.1 Introduction

Individual re-identification is essential in many aspects of aquacultural research, such as breeding and behavior studies. Different from identification, which often refers to distinguishing between different species and classifying the individual in the correct species, re-identification tracks the same individual over time with a consistent ID. This process connects repeated observations on the same individual, allowing researchers to monitor changes over time and detect deviations that may signal stress or health issues (Weirup, Schulz, & Seibel, 2022). For example, the growth curve derived from repeated measurements can predict expected body weights. Reduced growth rate can result from stress (Sopinka, Donaldson, O'Connor, Suski, & Cooke, 2016), and the variance of individual deviation from expected weight serves as a key indicator for health and resilience (Mengistu, et al., 2022).

The most common method for individual re-identification involves the use of tags, such as Passive Integrated Transponder (PIT) tags. Widely used for registering individual records in aquacultural breeding, production, and management, PIT tags are stable, recyclable, and available in large quantities. However, PIT tags also present several challenges. Studies report higher mortality rates among tagged fish, raising welfare concerns (Mahapatra, et al., 2001; Vollset, et al., 2020; Dheeran, Varghese, & Salimkumar, 2022). Practically, re-identification with PIT tags requires inserting a small chip into each fish, followed by manual handling and scanning each time the fish ID needs to be retrieved. Tagging and re-identification, including sedation and out-of-water transport during measurements, increase labor costs and induce stress for fish.

With the developments in computer vision, image- and video-based individual re-identification based on exterior features is becoming increasingly popular (Ye, et al., 2022; Zahra, Perwaiz, Shahzad, & Fraz, 2023). Building on research in human re-identification, which focuses on features such as facial structure, posture, and appearance like body shape and clothing, animal sciences have started to adopt vision-based technology in animal health and production systems (Schneider, 2020), including aquaculture. Non-invasive, high-throughput vision data, such as images and videos, enable real-time monitoring of individual and group phenotypes with minimal disturbance to the animals. Several fish species have distinctive exterior phenotypes such as pigment patterns or scale shapes (Stien, et al., 2017; Huntingford, Borçato, & Mesquita, 2013; Merz, et al., 2012), which suggests the possibility to replace PIT tags with vision-based re-identification if these distinctive phenotypes would exhibit unique, identifiable features of individual fish.

Despite the widespread use of re-identification technologies in security and surveillance of humans (Vezzani, Baltieri, & Cucchiara, 2013; Leng, Ye, & Tian, 2019), their application in aquaculture lags behind. At this stage, it remains unclear whether aquaculture is ready to replace PIT tags with non-invasive, vision-based re-identification, as the related challenges have not yet been systematically identified,

nor have potential solutions been fully explored. This review addresses these questions by identifying the obstacles to vision-based re-identification and proposes recommendations to improve its adoption in aquaculture.

5.2 Vision-based fish re-identification

We conducted a literature search via Google Scholar and Elsevier Scopus using combinations of the following keywords: individual, recognition, (re-)identification, aquaculture, fish, images, videos, and vision. A total of 23 studies were identified that used images or videos for individual re-identification in both captive and wild, for aquacultural, and ornamental species (Tab. 5.1).

Species were categorized according to their common usage, as well as the descriptions provided in each study's materials and methods section. For instance, although Arctic charr (*Salvelinus alpinus*) is typically associated with aquaculture, the individuals studied by Dębicki (2021) were found in spring-fed lava caves, suggesting a wild origin. Similarly, common carp (*Cyprinus carpio*) appeared in both aquaculture and ornamental studies, depending on the specific research setup. In several cases, the distinction between farmed and wild fish was unclear. For instance, Pedersen (2021) used out-of-water images of 39 brown trout of unspecified origin.

Given such ambiguities present in several reviewed studies, categorizing species by their usage provides a structured basis to evaluate methods with the highest potential for practical implementation. The following sections focus on the specific challenges of individual fish re-identification in aquaculture, while drawing comparative insights from wild, ornamental, and experimental setting.

Table 5.1 Vision-based fish re-identification studies on common aquacultural species, wild, ornamental, and experimental fish.

Species	**Aim**
Aquaculture	
Atlantic Salmon (*Salmo salar*)	Large (c. 3kg) salmon individual re-identification in sea cages 10 months later (Stien, et al., 2017)
	Iris recognition as a long-term biometric method for individual identification in salmon (Foldvik, Jakobsen, & Ulvan, Individual recognition of Atlantic Salmon using iris biometry, 2020)
	Non-invasive fish identification using biometric characteristics of fish iris (Schraml, et al., 2020)
	Fully automatic long-term individual identification of Atlantic salmon (Cisar, Bekkozhayeva, Movchan, Saberioon, & Schraml, 2021)
	Non-invasive individual identification to facilitate life-long tracking and monitoring of salmon (Mathisen, Bach, Meidell, Måløy, & Sjøblom, 2020)
Brook trout (*Salvelinus fontinalis*)	Novel visual learning framework for individual identification to improve identification accuracy (Zhou, Hitt, Letcher, Shi, & Li, 2022)
	Determine whether individual spot patterns are unique for short duration mark-recapture re-identification (Haxton, 2021)
European seabass (*Dicentrarchus labrax*) and common carp (*Cyprinus carpio*)	Show the possibility of automatic image-based ID based on the pattern of the scale positions (Bekkozhayeva & Cisar, 2022)
Chinook salmon (*Oncorhynchus tshawytscha*)	Identify individual salmon during migration based on spots pattern and development (Merz, et al., 2012)
Brown trout (*Salmo trutta*)	Automatic image-based individual recognition of brown trout (*Salmo trutta*) in the wild (Zhao, Pedersen, Hardeberg, & Dervo, 2019)
	To match the same brown trout in a set of other images based on individual image for tag-free, non-invasive recognition (Pedersen & Mohammed, 2021)

Wild	
Undulate skate (*Raja undulata*)	Non-invasive photo-identification via deep learning to discriminate between individuals (Gómez-Vargas, Alonso-Fernández, Blanquero, & Antelo, 2023)
Giant sunfish (*Mola alexandrini*)	Refined automated pipeline to re-identify giant sunfish (Pedersen, Nyegaard, & Moeslund, 2023)
Freckled hawkfish (*Paracirrhites forsteri*)	To assess the efficacy and accuracy of computer-aided photo identification of individual freckled hawkfish in the wild (McInnes, et al., 2020)
Stingray (*Megatrygon microps*)	To establish whether individuals could be reliably identified using natural spot patterns (Keeping, et al., 2020)
Basking Shark (*Cetorhinus maximus*)	Individual recognition over time from dorsal fin images for population estimation for conservation (Gore, Frey, Ormond, Allan, & Gilkes, 2016)
Freshwater armored catfish (*Rineloricaria aequalicuspis*)	To test the feasibility of photo-identification for individual recognition (Dala-Corte, Moschetta, & Becker, 2016)
Arctic charr (*Salvelinus alpinus*)	Photographic individual re-identification method that uses spot constellations in images (Dębicki, et al., 2021)
Corkwing wrasse (*Crenilabrus melops)*	To test the applicability of image-based re-ID as potential tagging replacement (Olsen, et al., 2023)
Experiment, ornamental and other fish	
Zebrafish (*Danio rerio*)	Individual fish identification based on the HSV color model (Al-Jubouri, Al-Azawi, Al-Taee, & Young, 2018)
	Combine different tracklets through individual re-identification (Bruslund Haurum, Karpova, Pedersen, Hein Bengtson, & Moeslund, 2020)
Medaka (*Oryzias latipes*)	Non-invasive and experimentally practical identification of medaka to improve experiment design (Osada, Yasugi, Yamamoto, Ito, & Fukamachi, 2024)
Crucian carp (*Carassius carassius*) and common carp (*Cyprinus carpio*)	Robust and accurate multi-view fish re-identification framework (Fan, Song, Feng, & Yu, 2024)

5.3 Vision-based re-identification in aquaculture

The current challenges in re-identification in aquaculture are examined from three perspectives: data acquisition, data analysis, and practical implementation.

5.3.1 Challenges in data acquisition

In the development of the concepts and practical guidelines for Precision Fish Farming (PFF) (Føre & Alver, 2023), the setup of a vision-based re-identification system should meet the criteria that address the specific requirements for aquaculture, which include:

Operational compatibility

The re-identification system should consider how specialized equipment and protocols can be integrated into real-world aquaculture environments that are built for fish production.

Minimal intervention

The introduction of new technology should minimize the interventions in regular farm operations. Possible difficulties come from increased time and labor requirements, and a higher frequency of fish handling. Given that farms usually do not have personnel dedicated to vision data collection, the technology setup should prioritize autonomy, robustness, and automation.
Most studies employed unique setups to capture vision data.

Overall, capturing exterior traits such as skin or scale patterns appears more feasible for practical implementation than imaging a biometric trait such as iris patterns. Imaging devices, for example as proposed by Cisar (2021), and others referenced in Chapters 1-3, could facilitate the collection of clear out-of-water images by stabilizing fish against a fixed background. Another common approach is to isolate fish in separate aquariums to ensure individual image capture. Both approaches, however, may increase labor demands or require personnel training. Moreover, frequent handling of fish raises welfare concerns due to the stress induced and can increase production costs. Data acquisition requires more automated, economical and fish-friendly design for data collection in aquaculture.

Among all approaches, using underwater cameras offers the least stressful and most practical solution for image acquisition by comparison, particularly when implemented as underwater video collection, although it may lead to higher costs related to device maintenance, data transfer, storage and pre-processing.

5.3.2 Challenges in data analysis

Based on the reviewed studies, we attributed a repeatable re-identification result to two aspects: data quality and applied methodologies.

Inadequate assessment of data quality

The reviewed studies used datasets of varying sizes, collected under different conditions (e.g. natural vs. induced lighting, underwater vs. out-of-water). Most studies acknowledged common image quality issues, such as low resolution, motion blur, incomplete fish in the view, lighting variation, and proposed technical solutions. However, their assessments were limited to the visual characteristics of individual images, whereas a broader and more systematic assessment of data quality was largely absent.

Although image quality is an important aspect for vision-based studies, data quality consists of multiple dimensions (Miller, et al., 2024) that are crucial for re-identification. This review highlights *timeliness* and *volume* as particularly relevant, based on the data quality framework proposed by Black & Nederpelt (2020).

Timeliness refers to how well the data represents the biological state of the fish at a specific point in time. In aquaculture, where many species (especially salmonids) undergo significant morphological changes throughout development, the timing of data collection is crucial. If datasets do not align with these changes, the corresponding re-identification methods are also limited. Most reviewed studies did not evaluate this aspect adequately. For instance, Mathisen (2020) used approximately 9 minutes of video, while Pedersen (2021) used just 24 seconds. Despite sufficient image resolution, it remains unclear whether such short recordings are sufficient for tracking individuals over time or across developmental transitions. Image datasets often focus on a single developmental stage, such as pre- or post-smoltification. Without continuous coverage across development periods, these methods may fail when applied to fish at different biological stages.

Volume refers to the number of fish included in the dataset and is critical for assessing model scalability and practical feasibility. This issue is addressed in more detail in the following section, which focuses on implementation challenges.

Lack of benchmarks in applied methodologies

Table 5.2 summarized the methods applied in reviewed studies.

Table 5.2 Methods used in individual re-identification studies in aquaculture.

Methodologies	Reference
Feature extraction	
Low-level features	
Blob extraction	Merz (2012)
Key points annotation	Stien (2017), Haxton (2021)
Histogram of Oriented Gradients (HOG)	Cisar (2021), Bekkozhayeva (2022)
Local binary pattern (LBP)	Pedersen (2021)
Speeded Up Robust Features (SURF)	Zhao (2019)
High-level features	
Convolutional neural network (CNN)	Schraml (2020), Cisar (2021), Zhou (2022)
Bidirectional feature pyramid networks (BiFPN)	Pedersen (2021)
Siamese network	Mathisen (2020)
Matching	
Cosine similarity	Schraml (2020), Pedersen (2021)
Euclidean distance	Cisar (2021), Bekkozhayeva (2022), Haxton (2021)
Angles between vectors	Stien (2017)
Coordinate matching	Merz (2012)
Support vector machine (SVM)	Zhao (2019)
Loss function (i.e. triplet loss)	Pedersen (2021), Mathisen (2020)

Re-identification methods typically involve two steps: feature extraction and feature matching. Feature extraction can be divided into low and high level. Low-level features refer to basic image characteristics such as edges, color histograms, and textures. These features are often pixel-based, relatively simple to compute, and require minimal processing power. In some studies, these features were annotated manually, as shown in the work by Stien (2017) , or extracted using semi-automated systems like the Interactive Individual Identification System (I3S spot) (Van Tienhoven, Den Hartog, Reijns, & Peddemors, 2007), which has been widely applied in the re-identification of wild species (Table , section Wild). In addition, automated image processing algorithms such as Histogram of Oriented Gradients (HOG) (Dalal & Triggs, 2005), Local binary pattern (LBP) (Ojala, Pietikäinen, & Harwood, 1996), Speeded Up Robust Features (SURF) (Bay, Tuytelaars, & Van Gool, 2006) are commonly used to extract low-level features efficiently.

High-level features, in contrast, are more abstract and often derived from deep learning models. Convolutional neural networks (CNNs) (LeCun, Bengio, & Hinton, 2015) are the most common approach for extracting these features, and several architectures, such as Faster R-CNN (Ren, He, Girshick, & Sun, 2015) and ResNet (He, Zhang, Ren, & Sun, 2015), have been used in the studies reviewed, such as Zhou (2022). Among the CNN-based approaches, two specific types of models were particularly notable. Bidirectional feature pyramid networks (BiFPN) (Tan, Pang, & Le, 2020) can fuse information in different layers of CNN to capture both low-level features like edges and textures and high-level features like the contours. Siamese network, or sometimes call twin neural network, transform the image into embeddings that contain both similar and dissimilar features of the two images simultaneously. Siamese networks (Chopra, Hadsell, & LeCun, 2005), also known as twin neural networks, transform image pairs into high-dimensional vectors where both similarity and dissimilarity can be evaluated to facilitate robust identity comparisons.

Feature matching, the second step in the re-identification process, involves comparing the extracted features to determine the likelihood that two images represent the same individual. This comparison can be done through traditional similarity metrics such as Euclidean or cosine distance (Malkauthekar, 2013), or through classification techniques such as support vector machines (Hearst, Dumais, Osuna, Platt, & Scholkopf, 1998). In studies where dot patterns served as identifying features, similarity was evaluated by assessing the alignment of these dots coordinates across images. In deep learning-based frameworks, matching is often integrated into the learning process itself. For instance, Siamese networks frequently use contrastive loss (Wang & Liu, 2021) or triplet loss (Schroff, Kalenichenko, & Philbin, 2015) functions to optimize the feature space such that similar vectors are drawn closer together while dissimilar ones are pushed further apart.

Evidence suggests that high-level, CNN-extracted features outperformed low-level features extracted from simple algorithm (Pedersen & Mohammed, 2021) in terms of re-identification accuracy. Furthermore, integrating multiple feature levels through feature fusion strategies has also shown to enhance model performance (Zhou, Hitt, Letcher, Shi, & Li, 2022). Another critical methodological advance is the use of transfer learning, where networks are initialized with weights pre-trained on large, general-purpose datasets and subsequently fine-tuned on species-specific data (Zhou, Hitt, Letcher, Shi, & Li, 2022; Mathisen, Bach, Meidell, Måløy, & Sjøblom, 2020).

Despite the availability of state-of-the-art algorithms, methodological challenges persist, primarily due to the absence of standardized benchmarks. As a result, method selection is often empirical and lacks theoretical justification. For example, Pedersen et al. (2023) and Zhao et al. (2019) both utilized the TROUT39 dataset. However, their studies focused on entirely different regions of interest, making direct comparison of results not feasible. The variation in datasets, in combination with

inconsistent or insufficient data quality description, hinders the objective evaluation of algorithm performance across reviewed studies. This issue extends beyond feature extraction and also affects the matching. While feature extraction plays a dominant role in determining performance, it would be naïve to overlook the influence of the chosen matching method on the overall accuracy of re-identification. For example, although not yet evaluated in the context of fish re-identification, studies have shown that cosine similarity between embeddings can produce inconsistent or arbitrary results when assessing semantic similarity (Steck, Ekanadham, & Kallus, 2024; Srivastava, 2023).

The absence of standardized benchmarks also leads to problematic analytical practice, such as data leakage. While it does not necessarily lead to inaccurate results, data leakage can inflate model performance by allowing the network to train on information it should not have. For example, the study of Mathisen et al. (2020) on salmon re-identification used a total of 225,000 images from 715 individuals, or about 314 images per fish. The dataset was split into 90% for training and 10% for testing, meaning that even in the test set, each salmon would occur more than 30 times. This raises serious concerns about the model's ability to generalize to data from fish that are not seen during training, as high accuracy may result from memorizing melanin patterns rather than identifying distinctive individual features.

Another related issue is the possible inconsistency in the definition of accuracy across studies. For example, Pedersen & Mohammed (2021) assessed accuracy using a binary metric to determine whether a given image pair belongs to the same individual. In datasets containing N individuals, this approach requires $N(N-1)/2$ pairwise comparisons to estimate re-identification success. In contrast, studies like Mathisen et al. (2020) evaluated accuracy by directly matching predicted identities to known individual IDs, where the final accuracy reflects the overall identification performance. While the difference between these evaluation methods may seem subtle, it introduces challenges for benchmarking, as the metrics are not directly comparable and may reflect different aspects of model performance.

In summary, challenges in data analysis stem from both data quality and methodological constraints. Data quality is oversimplified as image quality and not sufficiently assessed within an aquacultural framework. Methodologies employed are mainly developed for human re-identification without benchmarking in aquacultural datasets, making it difficult to identify optimal strategies for combining feature extraction and matching.

5.3.3 Challenges in practical implementation

Although vision-based re-identification always has the general aim to distinguish individual fish based on images, its practical implementation in aquaculture is for two purposes:

Replacement of tagging

In systems where fish are currently tagged individually, such as breeding candidates, the fish are manually handled for tag insertion and later for scanning to retrieve the tag ID. Using image or video data reduces stress due to handling and labor costs in this situation. Additionally, vision-based re-identification can help mitigate the risk of tag loss over long periods.

Tag-free systems

In cases where tagging is impractical, such as for juvenile fish, vision-based re-identification enables tracking of individuals. Application of tracking is seen in experimental fish studies (Table 5.1, section *Experiment, ornamental and other fish*), but its application in aquaculture would present additional challenges if there were a need to scale up to track large populations over extended periods.

Table 5.3 summarizes the re-identification results of reviewed studies on aquaculture species, based on three key factors: the number of individuals in the study, the duration between initial imaging and re-identification, and accuracy. Some studies report top i accuracy, where i represents the number of candidates matches among which the correct individual is present. For example, top 1 accuracy is the most practical for real-world deployment, top 5 accuracy can assist in narrowing down potential matches if further refinement opportunities exist, such as expert verification.

The reported accuracy in most of these studies is relatively high, however there are some limitations. Traditional tagging is typically done for tracking individuals over a timescale of weeks, months or even years. Most studies with high accuracy either included a small number of fish or covered only short time intervals. The longest time interval was 10 months (Stien, et al., 2017) with an 85% accuracy of re-id across 30 individuals. The largest number of fish was 715 (Mathisen, Bach, Meidell, Måløy, & Sjøblom, 2020) with a 96% accuracy of re-id over a 9-minute time interval. In systems that aim to replace traditional tagging, such as selective breeding programs or growth and feed conversion rate (FCR) assessment, nearly 100% accuracy is required for up to thousands of individuals. For tag-free systems, even larger sample sizes may be encountered.

While some studies describe the potential for upscaling by increasing image frequency, extending video durations, or including more fish, there is no evidence in the literature that accuracy of re-identification will remain high with a larger number of fish over a longer time period. As a result, vision-based individual re-identification in aquaculture remains challenging, and its feasibility as a replacement for tagging is still doubtful.

To summarize, while all studies on fish re-identification use vision data aim to develop non-invasive solutions, they approach the problem without fully integrating an aquacultural perspective in data acquisition, analysis and practical

implementation. For instance, description on the life stage and living condition of the fish involved were overlooked in many studies; Another example is the study by Pedersen & Mohammed (2021), which explored image-based recognition as an alternative to tagging juvenile fish. However, the data analysis excluded video frames containing these fish, likely due to low visibility. While this improved model performance, it ultimately misaligned the study with its original purpose.

Table 5.3 PIT tag performance for each species compared to the performance of vision-based re-identification.

Species	Reference	Nr. of re-ID individuals	Accuracy (Top 1)	Time interval	Are fish tagged?	PIT tag performance benchmarks
Atlantic Salmon (*Salmo salar*)	Stien, et al. (2017)	30	85%	10 months (12-month to 22-month salmon)	Yes	91% retention over 533-day (pre-smolt to one year adult) (Foldvik & Kvingedal, 2018)
	Foldvik, et al. (2020)	14	100%	124 days	Yes	
	Schraml, et al. (2020)	30	28%-80%	6 months	Not reported	
	Cisar, et al (2021)	30	36% (out of water) & 70% (aquarium)	6 months	Yes	
		328	100%	Within 5 days	No	
	Mathisen, et al. (2020)	715	96%	≤9 minutes	No	
Brook trout (*Salvelinus fontinalis*)	Zhou, et al. (2022)	Not reported	48%-97%	Not reported	Not reported	24% loss rate 70% retention (intraperitoneal) & 100% (dorsal musculature) over 2 months (Dieterman & Hoxmeier, 2009)
	Haxton (2021)	109	Not reported	≤2 days	No	
		315	Not reported	≤5 days	No	
Chinook salmon (*Oncorhynchus tshawytscha*)	Merz, et al. (2012)	105-295	0	3 months (FL from 69.7 to 122.1mm)	Yes	2% loss rate 18% retention in fish returning 6 months to 4 years (Knudsen, et al., 2009)
		106	100%	3 months (FL from 122.1 to 139.2mm)	Yes	
Brown trout (*Salmo trutta*)	Zhou, et al. (2022)	30	74%	Not reported	Not reported	32% loss rate 56% retention (intraperitoneal) & 95% (dorsal musculature) over 2 months (Dieterman & Hoxmeier, 2009)
	Pedersen & Mohammed (2021)	49	87.4% - 97.4%	Within 5 frames of 24-second video	No	
		39	71.8%	Not reported	Not reported	

***No PIT tag performance benchmark was found for European seabass (*Dicentrarchus labrax*) and common carp (*Cyprinus carpio*)**

5.4 Recommendation on future work

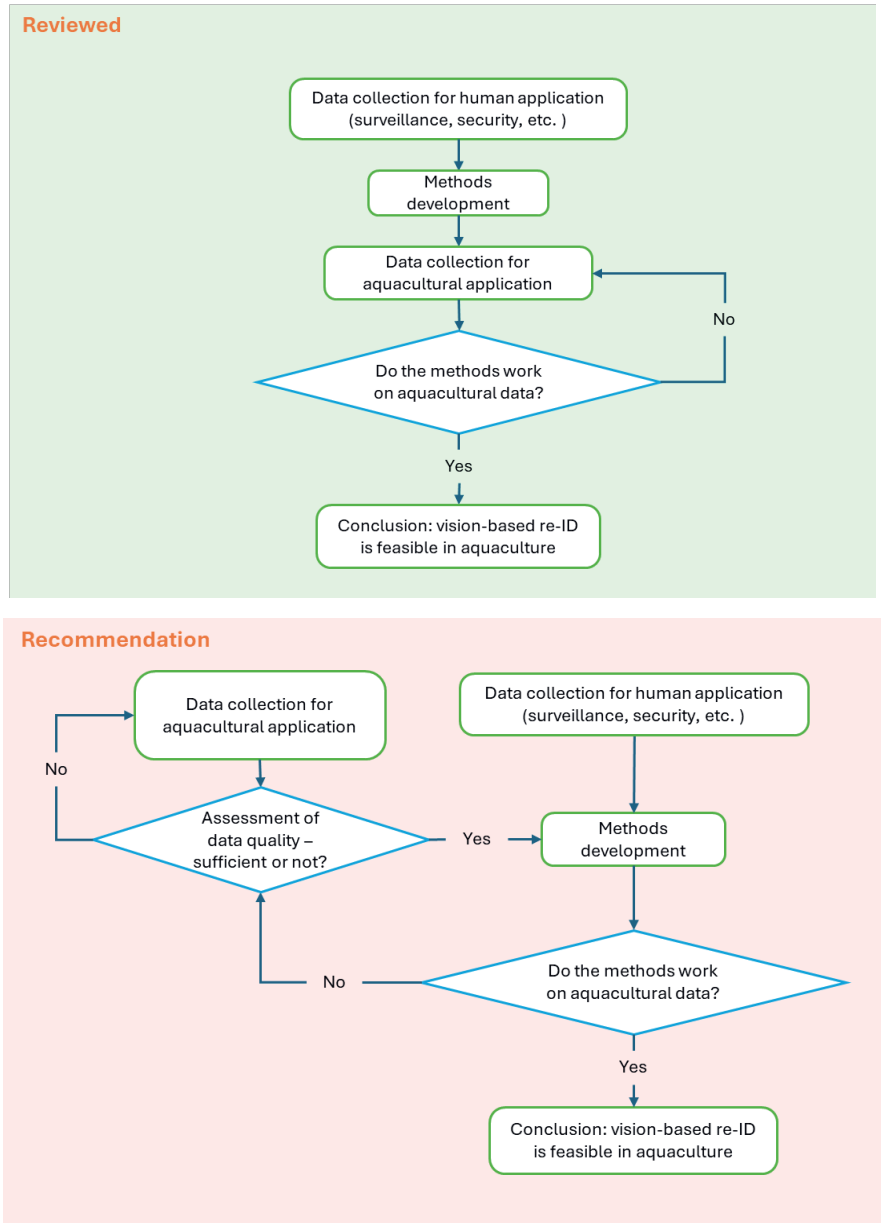


Figure 5.1 Top: flowchart of the structure of reviewed studies. Bottom: flowchart of the recommended steps with integrated aquacultural perspectives.

The current challenges in data acquisition, analysis and practical implementation are interconnected. Therefore, as a potential solution, we recommend a fundamental shift in how vision-based re-identification is approached in aquaculture. Figure 5.1 presents two flowcharts: one summarizing the structure of the reviewed studies and

the other illustrating our recommended approach. In the following section, we expand on these recommendations, using the flowchart as a guide for future research directions.

5.4.1 Recommendation on data acquisition

The first and foremost consideration in data acquisition is the environment or setting in which vision-based re-identification is intended to be conducted. Most aquaculture takes place outdoors, in ponds, cages, lakes, etc., and some in controlled facilities. Given such diversity in environments, we recommend two aspects of improvement in data acquisition considering the specific challenges facing in aquaculture.

First, **increasing automation enhances operational compatibility**. The increasing adoption of automated phenotyping in aquacultural research (Føre & Alver, 2023) is expected to drive the development of specialized imaging devices that are more portable, user-friendly, and resistant to harsh conditions such as high humidity and salinity. These devices can help optimize technical parameters of images, for instance, by offering stable lighting conditions and auto-adjusting zoom or focus, but their operation may still require active human involvement.

Second, **passive vision enables minimal intervention**. Passive vision (Jacobs, Souvenir, & Pless, 2009) utilizes data from continuously operating camera networks. Although such infrastructure is not yet widespread in aquaculture, it is feasible to begin deploying fixed cameras in both indoor and outdoor facilities to passively capture images over extended periods.

Cameras can be installed above the water to monitor surface behavior or underwater to directly observe fish activity. Efficient data acquisition through passive vision involves strategically placing cameras to take advantage of biological knowledge of the aquaculture system. For instance, in sea cages, fish distribution and behavior vary with depth (Banno, et al., 2025; Sauphar, Stolz, Tuene, Gansel, & Aas, 2024). Combining camera data from different depths can increase the likelihood of capturing identifiable individuals.

An effective example of passive vision is the study by Pedersen & Mohammed (2021) on brown trout, where video data was recorded as fish migrated through a fish ladder. Since all fish were guaranteed to pass through the ladder, this setup maximized the chances of obtaining individual images without actively capturing and releasing. A similar concept has also been applied in existing public camera systems, the Fish Doorbell (Utrecht, HDSR, & van Heukelum, 2025), an interactive webcam to observe fish migration. Such public sources could potentially be used to extract repeated image data of individuals.

Nevertheless, while passive vision reduces intervention related to fish handling, it may increase operational demands for device setup, maintenance, and data management, along with associated costs.

Setting up criteria for high quality data is the first step towards optimizing data collection. The following section provides a detailed discussion of this process.

5.4.2 Recommendation on data analysis

Many of the reviewed studies recommend collecting and testing on larger image datasets as a direction for future research. While we agree with this general approach, we argue that simply increasing dataset size is not sufficient. To ensure meaningful progress, we emphasize the need for more nuanced improvements, including the implementation of systematic data quality assessments and the establishment of standardized benchmarks for methodology evaluation.

Data quality assessment

In section 5.3.2, the assessment framework developed by Black & Nederpelt (2020) was used to evaluate the quality of datasets in the reviewed studies. Currently, no such framework has been applied to data quality in aquaculture. Accordingly, we recommend adapting this well-established framework to the aquacultural context to enable general data quality assessment.

Black & Nederpelt (2020) define a total of 65 dimensions of data quality. We screened these dimensions and selected those most relevant to aquacultural applications (see Table 5.4), along with practical examples specific to aquaculture. We also recommend applying these criteria during data acquisition to ensure quality control.

Table 5.4 Adapted Data Quality Dimensions from Black & Nederpelt, (2020) for Aquaculture Data quality assessment.

Dimensions	Definition	Examples in aquaculture data
Accessibility	Whether data can be obtained when needed	High accessibility for public datasets like ImageNet; private farm data may be restricted
Accuracy	Data correctly represents real-world entities	High-resolution images capture fish morphology accurately
Appropriateness	Data volume/type is suitable for task	Adult fish images may be inappropriate for juvenile re-ID
Availability	Whether authorized users/applications can retrieve the data	Mathisen et al. (2020) dataset is no longer accessible (expired hyperlink)
Clarity	Data presentation is understandable	Clear fish images aid interpretation: murky or dark images reduce clarity
Coherence	Logical consistency for combining data	ROIs from same fish should cohere better than varied ROIs from same image
Comparability	Cross-context matching potential	Comparing melanin patterns across individuals or over time
Completeness	All relevant data is collected	A dataset excluding early life stages lacks completeness
Concurrency	Simultaneous or time-matched data consistency	Images of 10-month-old salmon from multiple farms
Consistency	Absence of discrepancies across repeated representations	Stable lighting and background ensure image consistency
Equivalence	Conceptual equality of data in multiple places	Adjacent frames in video may show equivalent views of the same fish
Homogeneity	Consistency in terminology and scope across studies	Varying definitions of 'short-term' re-ID in reviewed studies reduce homogeneity
Integrity	Data is complete and without missing components	Uneven image counts across time points lowers dataset integrity
Metadata Compliance	Adherence to agreed definitions of attributes	Fork length measured consistently across studies meets metadata compliance
Precision	Degree of detail or granularity in data	Age in months is more precise than 'juvenile'/'adult' labels
Relevance	Applicability of data to the intended purpose	Salmon data can be irrelevant for seabass re-identification
Reproducibility	Ability to replicate dataset with same results	Changing camera models may reduce image reproducibility
Timeliness	Data reflects the correct time frame	Images taken 10 months apart may not reflect current fish appearance
Traceability	Data is well-documented, verifiable, and linked to source	Dataset without timestamps or locations has low traceability
Validity	Data conforms to its intended use domain	Images of juveniles are valid only for juvenile re-identification
Variety	Data from multiple sources, formats, or perspectives	Includes multiple angles, lighting conditions, and devices

Under such assessment criteria, what are the minimum requirements for a high-quality dataset? We recommend that datasets systematically report and document two key aspects: individual data and environmental data, as illustrated by the example in Figure 5.2.

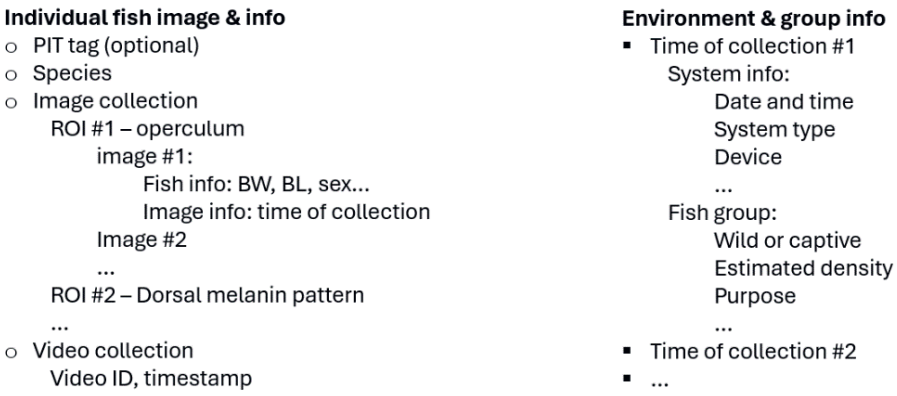


Figure 5.2 An example snip of what information should be included in the datasets collected for individual re-identification for aquaculture.

At the individual level, vision data should be recorded in relation to continuous biological indicators. Fish data is typically classified by species and sometimes broad life stages. However, such classifications can be insufficient for covering the lifelong biological variation relevant to individual re-identification. A potential solution is offered in Merz's study (Merz, et al., 2012) which uses fork length (FL) as a quantitative indicator of growth.

Currently, most studies rely on melanin patterns for identification, as they offer a promising balance between detectability, stability, and uniqueness but can be influenced by fish health, life stage, and environmental conditions. Operculum melanin spots, for example, have been shown to maintain stable locations and sizes over ten months in adult salmon (Stien, et al., 2017). Body melanin and scale patterns are less visible, particularly in underwater imaging. Additionally, skin spots may change in response to stress (Kittilsen, et al., 2009).

To address these limitations, iris biometrics have been explored as a complementary approach. Iris patterns are highly unique to each fish at certain life stage but show lower stability over time compared to melanin-based features.

In human biometric datasets, features are categorized by type (e.g., fingerprints, facial features, full-body images). A similar approach should be adopted for fish. Since different ROIs exhibit different levels of changes at different biological development, combining ROIs per individual could enhance success rates in re-identification. For example, melanin patterns on Atlantic salmon's body change pre-

and post-smoltification, making exterior patterns less reliable for re-identification across these stages. In such cases, iris biometrics may be a more reliable alternative, albeit with its own drawbacks.

Therefore, vision data intended for developing re-identification methods should include measurable, quantitative biological indicators to track developmental stages and should categorize images by distinct ROIs along with the whole fish image.

Apart from individual information, data regarding the environment is also crucial. This includes fish density at the time of data acquisition, descriptions of the aquaculture system, and background information such as whether fish are wild-origin or captive-bred, and their intended purpose (e.g., breeding, restocking, or grow-out).

Methods benchmarks

Benchmarking re-identification methods becomes more feasible as the availability of larger and public datasets increases. This includes implementing different combinations of feature extraction and matching methods on the same datasets or conducting replication studies using alternative datasets. For instance, in Chapter 3, we applied the materials proposed by Stein et. al (2017) and the methodological approach of Mathisen et al. (2020) to our own dataset, by combining two key factors from each study: a high number of individuals and a long interval between images. High accuracy of re-identification was reported when these two factors were separated. Our results with data from a practical aquaculture breeding environment revealed that image-based re-identification was not yet feasible in this scenario, with re-identification accuracy approximated zero.

Based on our review, current studies focus only on features extracted from static image or video frame. However, as more data annotated with biological development become available, longitudinal features that capture growth and morphological changes over time can potentially improve individual uniqueness. Fish re-identification may not be based solely on still exterior patterns but could also incorporate time-series features that reflect and predict how an individual's appearance changes with growth and development over time.

5.4.3 Recommendation on practical implementation

Challenges in practical implementation result from the uncertainty associated with upscaling. While these challenges can be mitigated through standardized datasets and benchmarked methods, additional considerations remain critical when transitioning re-identification approaches from experimental development to real-world application.

First, the intended purpose of re-identification must be clearly defined. For instance, if the goal is to fully replace PIT tags, as may be the case in experimental or ornamental fish tracking (Table , section Experiment, ornamental and other fish),

the expected accuracy must be very high. Optionally, if the goal is to complement existing PIT tag systems, such as providing backup identification when tags fall out during growth or migration, a lower accuracy may still provide valuable utility. Understanding this distinction allows for realistic expectations regarding performance and acceptable error margins.

Second, practical deployment scenarios often diverge significantly from the controlled conditions in which models are trained and evaluated. Environmental variability, device limitations, and handling constraints may all undermine the scientific validity and performance of a model. These real-world uncertainties mean that, beyond a certain point, investing additional time and resources to achieve marginal gains in accuracy may yield declining returns and become economically inefficient. For instance, cameras installed in fish farms and occasionally left unattended can experience blurred views due to water droplets from splashes, while underwater cameras used for tracking algorithms may have reduced lifespan in high-salinity environments. Addressing these challenges requires iterative experimentation and rigorous testing under real-world conditions to refine models and optimize their performance within practical constraints.

By taking these considerations into account early on, the implementation of re-identification systems can be better aligned with operational needs and cost-benefit strategies in aquaculture settings.

5.5 Conclusion

The non-invasive nature of vision-based data holds great value and potential in individual re-identification. However, the practice of aquaculture can be humbling. Facing many challenges regarding data acquisition, data analysis and practical implementation, the development of vision-based individual re-identification is currently too immature to be a replacement for the PIT tag. In response, we propose a new framework for method development that places strong emphasis on data quality and addresses these challenges from an aquaculture-centered perspective. Our recommendations prioritize realistic, practical implementation to ensure that vision-based individual re-identification can be meaningfully integrated into real-world aquacultural systems.

6. Distance to perfect shape in gilthead seabream (*Sparus Aurata*) is moderately heritable when quantified objectively by automated image analysis

Yuuko Xue[1], Gert Kootstra[2], Hans Komen[1], Stelios Karapanagiotis[3], Kalliopi Tsakoniti[3], John W.M. Bastiaansen[1]

[1]Department of Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands

[2]Department of Agricultural Biosystems Engineering, Wageningen University & Research, P.O. Box 16, 6700 AA Wageningen, The Netherlands

[3] Galaxidi Marine Farm S.A., Galaxidi, Greece

Abstract

Body shape is an economically and biologically important trait in aquaculture breeding programs, yet it remains challenging to measure shape objectively. In this study, we propose a novel method for quantifying body shape in gilthead seabream (*Sparus aurata*) using image-derived traits. Expert-based visual scores were collected on five morphological regions for 3,421 breeding candidates together with images. Using elliptic Fourier descriptors and generalized Procrustes analysis, we visualized both ideal and less desirable reference shapes as average best and worst contours. Two quantitative traits, Distance to the Best (DtB) and Distance to the Worst (DtW), were derived to measure how much each fish deviated from these reference shapes. Genetic analyses showed moderate heritability for DtB (0.33 ± 0.04), and lower heritability for DtW (0.23 ± 0.03) and the score traits. The pattern of phenotypic averages, and the genetic correlations indicated that a lower DtB is associated with higher expert scores, while DtW captured a wider range of less desirable shape variations. This approach provides an automated, objective, and visually intuitive method for integrating preference on body shape into aquaculture breeding programs, with potential for economic valuation.

Key words: Score traits; Objectivity; Body shape; Contour analysis; Elliptic Fourier Descriptor

6.1 Introduction

Body shape of fish is a critical trait in aquacultural breeding and management, as it reflects the biological performance (Xue, et al., 2025) and influences the market value (Mørkøre, et al., 2001). Fish of the same species can display substantial variation in shape due to environmental factors (Smith, Costa, Kristjánsson, & Parsons, 2024; Ndiwa, Nyingi, & Claude, 2016; Collin & Fumagalli, 2011; Marcil, Swain, & Hutchings, 2006; Wimberger, 1992), or due to domestication that further alters shape through selection. For example, wild and domesticated fish can exhibit different body forms (McMaster, 2023; Ahmet Doğdu & Turan, 2021). In fish species that are sold whole such as sole and seabream, shape is a key commercial trait valued by both breeding companies and consumers (Colihueque & Araneda, 2014).

Body shape is a complex trait. Although it can be simply defined as the external form of the fish body, shape is not a single measurement but a collective expression of various morphological attributes. In flatfish such as sole, shape can be approximated with simple geometric measurements such as fitting an ellipse (Blonk, Komen, Tenghe, Kamstra, & Van Arendonk, 2010). For round fish such as gilthead seabream, measuring shape becomes more challenging and typical relies on multiple indicators derived from landmarks, dimensions and their ratios (Freitas, et al., 2023; Fragkoulis, Kerasovitis, Batargias, & Koumoundouros, 2021).

Selective breeding for shape can utilize indicator traits, but this is only effective when genetic correlations are sufficiently strong. Landmark-derived indicators might only capture part of the variance in shape, therefore exhibiting a low correlation with the visually perceived shape. Another approach of measuring shape is an expert-based grading system, where shape is scored visually according to pre-defined criteria (Fragkoulis, et al., 2017). Compared to using indicator measurements, this method directly reflects the desired shape, with higher scores indicating preference. However, such approach is labor intensive, prone to subjectivity and bias. Moreover, scoring systems result in categorical traits, which limits the coefficient of variation or has low resolution in the collected data (Blonk, Komen, Tenghe, Kamstra, & Van Arendonk, 2010).

To answer the need for a better definition, measurement and quantification of shape for selective breeding, we proposed an algorithm for automated, quantitative analysis of body shape in gilthead seabream. Visual scores given to selection candidates were used to derive two novel traits: Distance to the Best (DtB) and Distance to the Worst (DtW). These traits quantify how closely each individual resembles the ideal or less desirable shapes. Furthermore, we investigate the genetic variance of both traits and the potential to integrate these automated measurements into breeding programs for shape improvement.

6.2 Materials & Methods

6.2.1 Materials

Images were collected for 3,792 harvest size gilthead seabream (*Sparus aurata*), all of which were PIT-tagged breeding candidates with known pedigree information. Data collection was conducted in sessions at 10 different times between 2018 and 2020. Prior to imaging, each seabream was sedated, weighted and placed in a lighting cabinet with three LED panels positioned at the top, side, and bottom. The top horizontal panel is equipped with a built-in RGB camera (Logitech HD Pro Webcam C920). Each seabream was positioned with its right lateral side up on the bottom panel, with a grid paper placed underneath as a reference scale (Tab. 6.1). All images were taken with the same device, but at different times of the year. The zooming setting, background and image scaling are the same within a batch but may be different between batches.

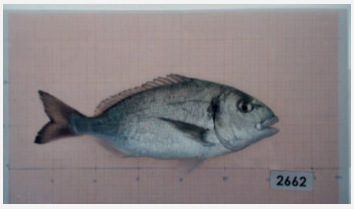
6.2.2 Methods

To distinguish between visual scores used to describe preference in shape and traits derived from image analysis, we refer to the former as *score traits* and the latter as *shape traits*.

Definition of score traits

Using the corresponding image, each fish was visually scored by at least two breeding experts on five morphological traits: head, dorsal side, caudal peduncle, caudal fin and ventral side (Tab.6.1). Each trait was assessed on a scale from 1 to 3, where a score of 3 indicates a favorable shape for selection and a score of 1 an unfavorable shape. The score for each trait added up to total score.

Table 6.1 Score traits and their definitions, with fish image as reference.



Traits name	Definition
Head	Shape of the head
Dorsal side	Curviness of the dorsal side
Caudal peduncle	Width and length of the caudal peduncle
Caudal fin	Completeness and openness of caudal fin
Ventral side	Curviness of the ventral side
Total score	The sum score of all traits above

Definition of shape traits

All image analyses were conducted using Python 3.9 (Van Rossum, Drake, & others, 1995). To ensure consistency across images, a standardized image processing workflow was implemented. See Algorithm 6.1 for the pseudo codes.

Segmentation

We applied the Segment Anything Model (SAM) (Kirillov, et al., 2023) to segment the fish from their respective backgrounds. Following segmentation, each fish mask

that contained the whole fish surface was centrally aligned and pasted onto a white background of fixed dimensions (640 × 360 pixels) to create the segmented fish images.

Contour detection

The segmented fish images were processed using OpenCV (Bradski, Kaehler, & others, 2000) contour detection to extract the fish outline. The largest detected contour was retained as the representation of the fish body shape, and its coordinates were saved for subsequent analysis.

Elliptic Fourier descriptor

Using the elliptic Fourier descriptors (Kuhl & Giardina, 1982) analysis, the extracted contours were decomposed into a series of ellipses where each captures a different level of details of the contour.

Elliptic Fourier descriptors (EFDs) consist of harmonics and Fourier coefficients that can be used to reconstruct the original curve. Each harmonic in the EFD framework uses a pair of sine and cosine functions to describe the periodic changes in both the x and y coordinates of the contour over time(t). Specifically, the parametric equations are defined as:

$$x(t) = a_0 + \sum_{n=1}^{N} (a_n \cos(n\omega t) + b_n \sin(n\omega t))$$

$$y(t) = c_0 + \sum_{n=1}^{N} (c_n \cos(n\omega t) + d_n \sin(n\omega t))$$

Where t is the normalized time parameter ($t \in [0, 1]$) along the closed contour, ω is the fundamental angular frequency equal to 2π, n is the harmonic number and a_n, b_n, c_n, d_n are the corresponding coefficients. While a_0 and c_0 are the average x and y position over t, these terms are often set to 0 to center the shape in the image. Besides, the first harmonic is often standardized by rescaling all coefficients by its amplitude. The first harmonic is then rotated to align the shape to a common orientation, and its phase is adjusted to fix a consistent starting point along the contour. These steps ensure EFDs invariance to scale, rotation, and starting point.

Contour reconstruction

The original fish contour can be reconstructed and visualized using harmonics (Fig. 6.1). While using more harmonics will capture more details in shape and better resemble the original contour, it increases the risk of including noise, such as fin protrusions.

To balance these two aspects, we assess the proportion of shape variance captured by a given number of harmonics in EFDs. The power or contribution of each harmonic can be quantified as the sum of squares of its coefficients, reflecting the harmonic's amplitude:

$$P_n = a_n^2 + b_n^2 + c_n^2 + d_n^2$$

The total harmonic power across all N harmonics considered is then:

$$P_{total} = \sum_{n=1}^{N} P_n$$

To evaluate the cumulative proportion of variance explained by the first k harmonics, the normalized cumulative power is computed as:

$$Explained\ Variance_k = \frac{\sum_{n=1}^{k} P_n}{P_{total}}$$

For this study, 30 harmonics $(N = 30)$ were used in total and 7 ($k = 7$) were used for reconstruction. This choice was based on the visual resemblance of the reconstructed and original contour (Fig. 6.1). The first 7 harmonics explain 99% of the variance out of 30 harmonics of the original contour.

Each reconstructed contour was resampled with 1,000 equally spaced points. Using the same number and equal spacing of points allows comparison of corresponding positions along the contours of different fish.

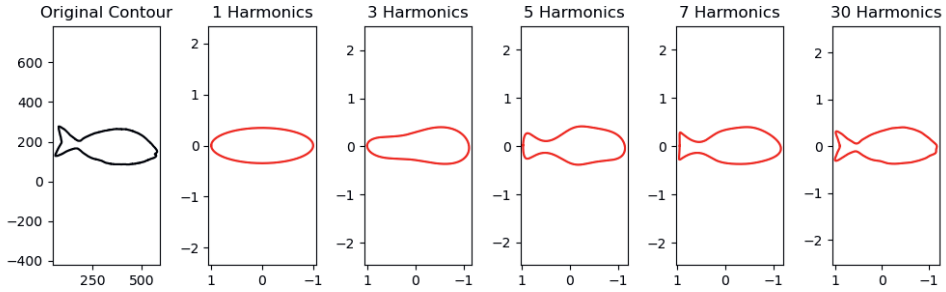


Figure 6.1 Original and reconstructed contours using 1, 3, 5 7 and 30 harmonics.

Python package pyefd (Blidh, 2021) was used to generate both EFDs and reconstructed contours.

Contour alignment

To ensure directional consistency across all fish, each reconstructed contour was aligned based on its principal axes using Principal Component Analysis (PCA) (Karl Pearson, 1901). The directions of the first and second principal components of the reconstructed contours were used to orient them, ensuring all fish appeared facing

right, with the dorsal side upwards. The starting point of each contour was reindexed to begin at the most anterior point of the fish and proceed clockwise along the contour.

Python package scikit-learn (Pedregosa, et al., 2011) was used for principal component analysis.

Generalized Procrustes analysis

We selected all fish that received the maximum value (3) for all score traits and used their reconstructed contours to represent the ideal shape. Generalized Procrustes analysis (GPA) (Gower, 1975) was used to calculate the *average best* contour. Same was applied to fish that received low values for all score traits. Since there were only 3 fish with 1 for all traits, we included all fish with a total score smaller than 8, to calculate the *average worst* contour. Both average contours served as geometric references for the derivation of shape traits.

Calculation of shape traits

Procrustes distance (Mitteroecker & Gunz, 2009) refers to the minimized sum of squared Euclidean distances between all corresponding points along the contour. The Procrustes distances between each individual fish contour and the two average contours derived from GPA were computed. These two distances, referred to as *Distance to the Best* (DtB) and *Distance to the Worst* (DtW), quantify how far each fish deviates from the two references. A lower DtB, or higher DtW, means that the shape of this individual is closer to the ideal, or further from the less desired shape, and therefore more favorable for selection. Python package Scipy (Virtanen, et al., 2020) was used for both generalized Procrustes analysis and the calculation of Procrustes distance.

Algorithm 6.1 Computation of Distance to the Best (DtB) & Distance to the Worst (DtW)

Input: Set of fish images I = {I_1, I_2, ..., I_n} with associated scores
Input: target total score
Initialize empty lists:
    Contours ← [ ]
    Scores ← [ ]
For image I_i in I **do**:
    Segment fish using SAM (Segment Anything Model)
    Center segmented fish on a 640×360 white background
    Detect contour using OpenCV
    Compute Elliptic Fourier Descriptors (EFD)
    Reconstruct contour using first 7 harmonics (1000 key points)
    Align contour orientation using PCA
    Append aligned contour to Contours
    Append the sum of all associated scores to Scores
End for
BestSet ← {contour ∈ Contours | score = 3 for all traits}
WorstSet ← {contour ∈ Contours | score = 1 for all traits}
BestContour ← GeneralizedProcrustesAnalysis(BestSet)
Average best contour <-Plot(BestContour)
WorstContour ←GeneralizedProcrustesAnalysis (WorstSet)
Average worst contour <-Plot(WorstContour)
Initialize empty lists:
    DtB ← [ ]
    DtW ← [ ]
For contour in Contours **do**:
    d_best ← ProcrustesDistance(contour, BestContour)
    d_worst ← ProcrustesDistance(contour, WorstContour)
    Append d_best to DtB
    Append d_worst to DtW
End for
Return DtB and DtW

Agreement between score traits and shape traits

After calculating the Distance to the Best (DtB) and Distance to the Worst (DtW), outlier removal and distribution transformation were performed. Most (99.0%) of the DtB and DtW values were between 0 and 0.06. However, distance values were disproportionately high for images where contour detection failed, ranging from 0.49 to 0.95, with a mean of 0.63. To ensure these failed detections were excluded, an empirical threshold of 0.1 was applied, which effectively removed all such cases. Subsequently, both DtB and DtW were transformed from a Gamma distribution to a normal distribution using function gamma2norm in R package faux (DeBruine,

2025). Both traits were also standardized with a mean of 0 and a standard deviation of 1. The outlier removal, distribution transformation and standardization were performed using R 4.2.1 (R Core Team, 2019).

To assess how well the shape traits align with score traits, boxplots of DtB and DtW values were plotted against the total scores to examine the relationship of distance and score traits. This was performed in Python 3.9 using package Seaborn (Waskom, 2021).

Genetic analyses of all traits were conducted using ASReml-R 4.2 (Butler, 2021). A univariate linear mixed model was fitted to estimate the heritability of all traits, represented by the equation:

$$y = X\beta + Z\upsilon + e$$

where y is a vector of phenotypes, β is the vector of fixed effect batch number, υ is the vector of random animal additive genetic effects ~ $N(0, A\sigma_a^2)$, where A is the relationship matrix and σ_a^2 is the additive genetic variance of the trait, and e is the vector of random residual effects ~ $N(0, I\sigma_e^2)$, where I is an identity matrix and σ_e^2 is the residual variance of the trait. X and Z are the design matrices that relate observations to the fixed effects and additive genetic effects, respectively. Heritability is calculated as the proportion of total phenotypic variance explained by additive genetic variance:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$$

The phenotypic (r_p) and genetic (r_g) correlations of shape traits with score traits were estimated using the bivariate model:

$$y = X\beta + Z\upsilon + e$$

where y is the concatenated vector of phenotypes of the two traits, β is the vector of fixed effect batch number, υ is the vector of random additive genetic effects ~ $N(\begin{bmatrix}0\\0\end{bmatrix}, A \otimes \begin{bmatrix}\sigma_{a,1}^2 & r_{a,12}\sigma_{a,2}\sigma_{a,1}\\ r_{a,12}\sigma_{a,2}\sigma_{a,1} & \sigma_{a,2}^2\end{bmatrix})$, where A is the genetic relationship matrix, $\sigma_{a,1(2)}^2$ is the additive genetic variance of trait 1(2), $r_{a,12}$ is the genetic correlation between traits 1 and 2, and e is the vector of residual effect ~ $N(\begin{bmatrix}0\\0\end{bmatrix}, I \otimes \begin{bmatrix}\sigma_{e,1}^2 & r_{e,12}\sigma_{e,2}\sigma_{e,1}\\ r_{e,12}\sigma_{e,2}\sigma_{e,1} & \sigma_{e,2}^2\end{bmatrix})$, where I is an identity matrix, $\sigma_{e,1(2)}^2$ is the residual variance of trait 1(2), $r_{e,12}$ is the residual correlation between traits 1 and 2. X and Z are the design matrices that relate observations to the fixed effects and additive genetic effects, respectively. The genetic correlation between trait 1 and 2 is calculated as:

$$r_g = \frac{\sigma_{a,12}}{\sqrt{\sigma_{a,1}^2 \cdot \sigma_{a,2}^2}}$$

Where $\sigma_{a,12}$ is the genetic covariance between trait 1 and 2.

The founders of the population originated from 23 different sources, and their genetic contributions to the offspring within each batch are uneven. Therefore, for both the univariate and bivariate models, the genetic analysis was performed with genetic groups to account for differences between the sources of founders. Batch number was tested as fixed effect and was included for traits where this effect was significant.

6.3 Results

6.3.1 Descriptive statistics

Table 6.2 For each trait, the number of fish per score, mean and standard deviation.

Traits	Number of fish			Mean (unit: score)	Standard deviation (unit: score)
	Score 1	Score 2	Score 3		
Head	352	1957	1482	2.30	0.63
Dorsal side	308	1288	2196	2.50	0.64
Caudal peduncle	301	2029	1461	2.31	0.61
Caudal fin	220	1740	1832	2.43	0.60
Ventral side	783	2059	950	2.04	0.67

Table 6.2 summarizes the basic statistics for each score trait. Overall, the distribution of fish across scores was unbalanced, with fewer individuals receiving a score of 1. A similar pattern was observed for the total score (Fig. 6.2), where most fish were scored around the average. Only three fish received a score of 1 on all traits, while 129 fish achieved the maximum score of 3 on all traits.

In total, image analysis failed for 234 images, primarily due to issues with segmentation, contour detection, or reconstruction, and no DtB/DtW values were calculated for these cases. Figure 6.3 shows the Gamma and transformed normal distribution of DtB and DtW. Subsequently, outliers exceeding four standard deviations from the mean DtB and DtW were removed, resulting in a total of 3,452 fish included in the estimation of genetic parameters.

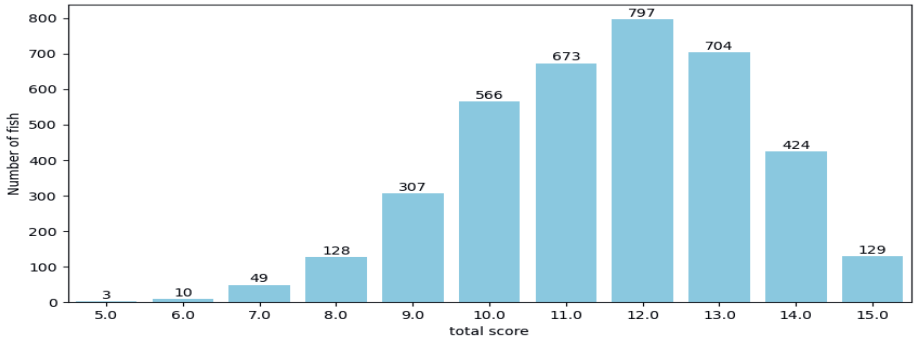


Figure 6.2 The number of fish per total score, from 5 to 15.

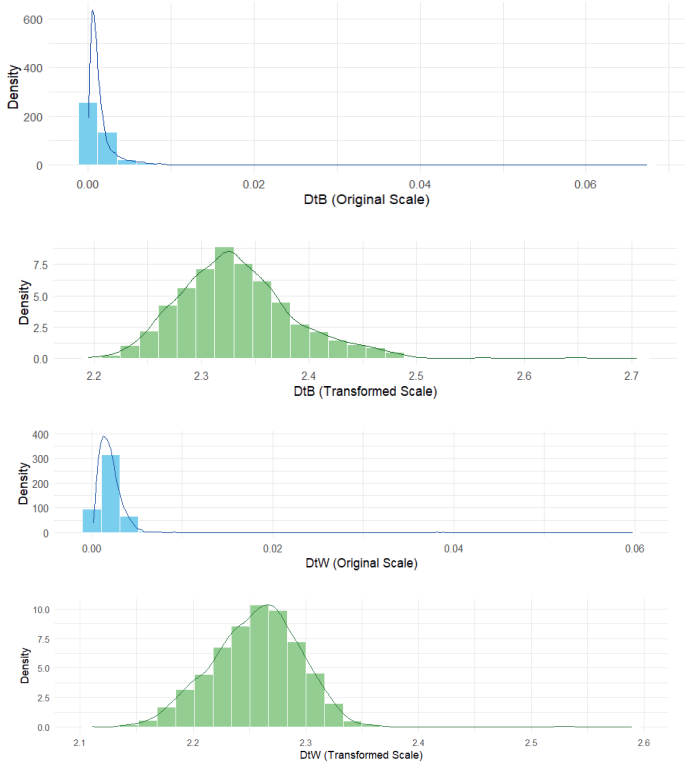


Figure 6.3 The distribution transformation of DtB and DtW. From top to bottom: Gamma distribution of DtB; transformed normal distribution of DtB; Gamma distribution of DtW; transformed normal distribution of DtW.

6.3.2 The average best and average worst contour

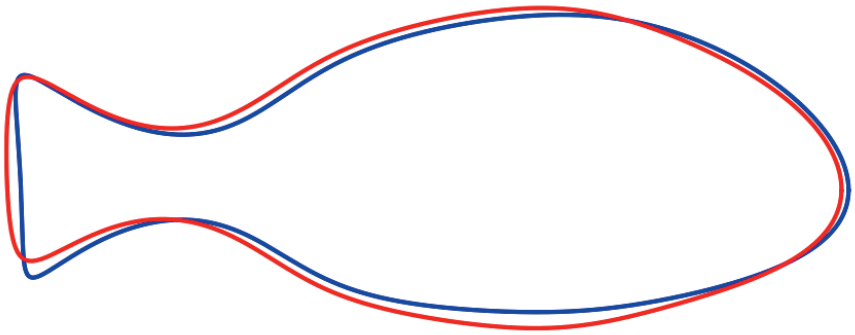


Figure 6.4 Average best contour of the fish with score 3 (blue) and average worst contour of the fish with score 1 (red). Both contours are standardized and center aligned.

Figure 6.4 shows the visualization of average best (N = 129) and average worst (N = 62) contours. The largest deviation between the two contours is observed at the ventral side of the fish, indicating that a rounder ventral side is less favorable. Additionally, the caudal peduncle of the fish with lower scores appears to have a wider, or less straight, connection with the body. However, the width of this region, when defined as the narrowest part of the peduncle, does not seem to differ between low and high-scoring fish.

Low and high-scoring fish exhibit similar roundness on the dorsal side. Nevertheless, the transition from the dorsal to the head region is more gradual in high-scoring fish. In contrast, this transition is steeper in low-scoring fish.

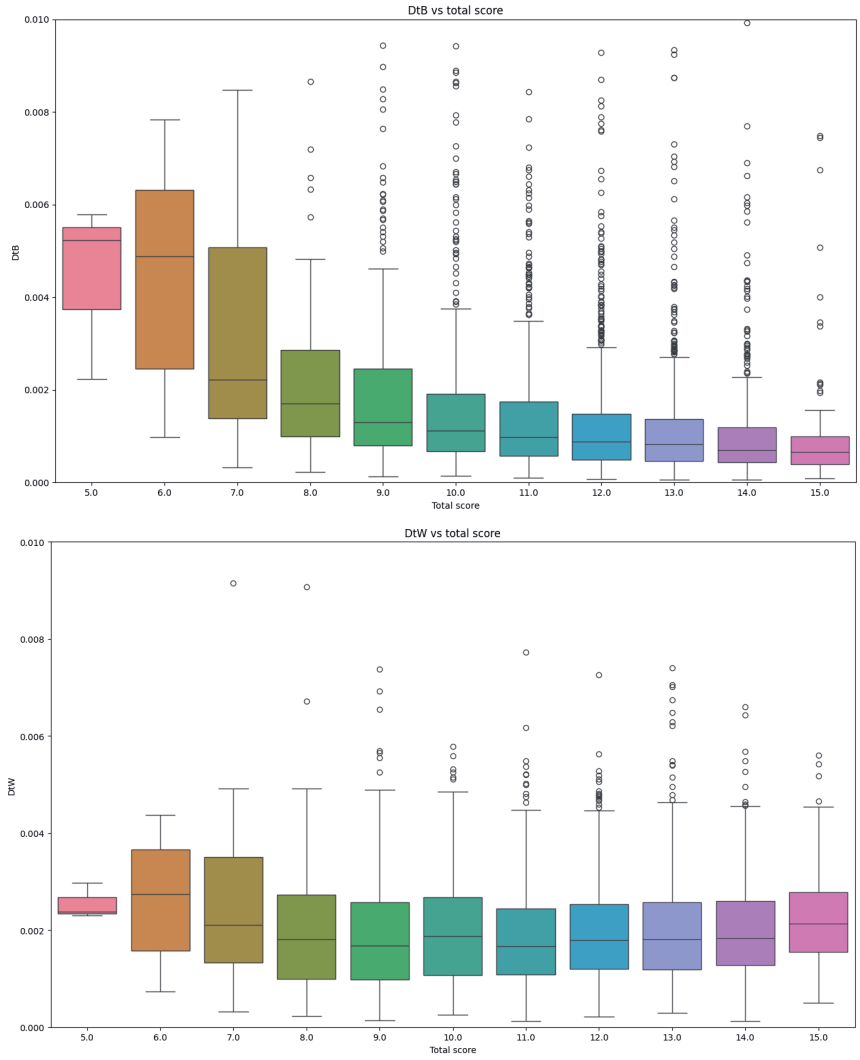


Figure 6.5 Top: the relationship between the total score scores and DtB. Bottom: the relationship between the total score and DtW.

The phenotypic relation between Distance to the Best (DtB), Distance to the Worst (DtW), and the score traits are visualized in Figure 6.5. For the mean value of DtB across the total scores, a clear trend is evident: as the total score increases, DtB decreases, indicating that the corresponding fish shape is closer to the average best shape. However, such a trend is less clear for DtW. The differences in average DtW are small across all the classes of total score and the minimum DtW is seen with intermediate scores of 8 and 9. When the overall score is above 10, there is a tendency for DtW to increase with higher scores, suggesting that the corresponding fish shape is further from the average worst shape. Conversely, for scores ranging from 5 to 10, DtW decreases as the score increases.

Additionally, the variance of DtB decreases when the overall score increases. However, such decrease in variance between scores is much smaller in DtW: the variance is low for total score 5, higher for score 6 and 7, and similar for overall scores between 8 and 15.

6.3.3 Genetic parameter estimates

Table 2 shows the results of the genetic analysis among all score traits, shape traits DtB and DtW, and harvest weight. The analysis revealed that DtB exhibited moderate heritability; 0.33±0.04. In contrast, all score traits, together with DtW, showed low heritabilities between 0.14±0.03 and 0.25±0.04. The total score trait also has a low heritability of 0.20±0.03. Harvest weight had the highest heritability 0.47±0.04.

Table 6.2 Genetic parameter estimates of score traits, image-derived shape traits, and harvest weight. diagonal: heritability. above the diagonal: phenotypic correlations. below the diagonal: genetic correlations. plus-minus sign (±) indicates the standard error of the estimate.

	HEAD	DORSAL SIDE*	CAUDAL PEDUNCLE	CAUDAL FIN	VENTRAL SIDE	TOTAL SCORE*	HARVEST WEIGHT*	DTB*	DTW*
HEAD	0.21±0.03	0.09±0.006	0.04±0.005	0.013±0.006	0.045±0.006	0.26±0.008	0.09±0.01	-0.001±0.001	0.0003±0.0007
DORSAL SIDE*	0.70±0.09	0.17±0.03	0.054±0.005	0.024±0.005	0.08±0.006	0.30±0.007	0.055±0.012	-0.003±0.001	-0.0003±0.0007
CAUDAL PEDUNCLE	0.17±0.13	-0.04±0.14	0.14±0.03	0.043±0.005	0.08±0.006	0.26±0.007	0.014±0.012	-0.012±0.001	0.0009±0.0007
CAUDAL FIN	-0.06±0.13	-0.05±0.14	0.29±0.13	0.17±0.03	0.01±0.006	0.19±0.008	0.08±0.01	-0.007±0.001	-0.005±0.0007
VENTRAL SIDE	0.20±0.12	0.20±0.12	0.49±0.11	0.02±0.13	0.25±0.04	0.30±0.001	-0.15±0.01	-0.01±0.001	0.007±0.0007
TOTAL SCORE*	0.69±0.07	0.61±0.08	0.61±0.09	0.34±0.11	0.69±0.06	0.20±0.03	0.03±0.01	-0.012±0.001	-0.0014±0.001
HARVEST WEIGHT*	0.27±0.10	0.23±0.10	-0.18±0.03	-0.14±0.11	-0.35±0.10	-0.10±0.10	0.44±0.04	0.004±0.001	-0.01±0.0007
DTB*	0.31±0.11	0.36±0.11	-0.34±0.11	-0.15±0.12	-0.35±0.10	-0.05±0.11	0.42±0.08	0.33±0.04	0.01±0.0004
DTW*	0.18±0.11	0.08±0.12	0.30±0.12	-0.06±0.13	0.41±0.10	0.35±0.11	-0.44 ±0.08	-0.06±0.11	0.23±0.03

* **Traits are estimated with fixed effect**

6.4 Discussion

This study introduced a novel transformation of five subjective qualitative measurements of fish shape into objective quantitative shape traits. The five score traits collectively captured the preferred body shape in the breeding program, each focusing on a different morphological region. In contrast, the image-derived shape traits were based on visual representations that summarize either the ideal or the less desired shapes of gilthead seabream.

Figure 6.5 shows the relationship between the total score and the two image-derived traits, Distance to the Best (DtB) and Distance to the Worst (DtW). Both the average of DtB and the variation in DtB trend to lower values when total score becomes higher. This suggests that with a higher total score the fish shape converges to a clear 'perfect' shape. In contrast, the average of DtW and the variation in DtW does not trend lower when total scores become smaller. Instead DtW showed similar values across total scores between 8 and 15 and trended higher for the lowest scores of 5 to 7. Only three individuals had total score of 5, but total scores 6 and 7 clearly show higher distance to DtW and higher variation.

These different patterns in average and variation between DtB and DtW suggest that the ideal shape is uni-directional: there is a clear 'perfect' shape that individuals converge toward when highly scored. The high and variable DtW for low scoring fish suggests that strong deviations from the perfect shape occur in multiple directions. In other words, while there may be one ideal way to have the 'right' shape, there are many ways to have the 'wrong' shape. This interpretation is supported when plotting the individual contours of the best and worst fish (Fig. 6.6). The best fish contours are much closer to their average than the worst fish' contours. Finally, the average worst shape closely resembles the average best shape (as shown in Fig. 6.6). This appears to be a result of the undesirable extremes cancelling each other out in the averaging process for the average worst shape. The genetic correlation between DtB and total score (-0.05±0.11) is smaller than that between DtW and total score (0.35±0.11), suggesting that DtW captured more genetic variance in the total score than DtB.

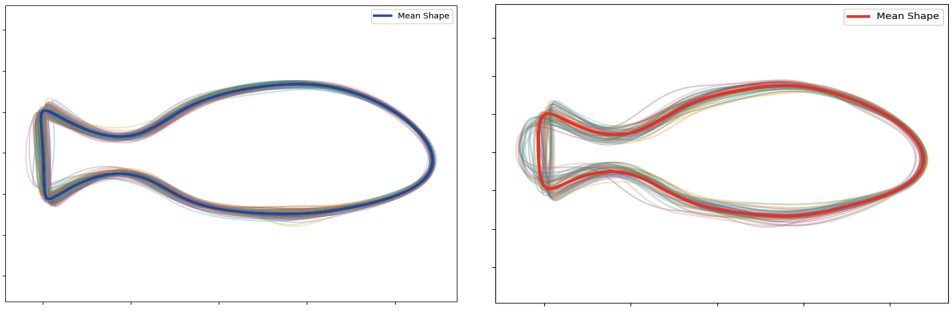


Figure 6.6 Left panel: contours of fish with score 3 for all traits, and the average best contour in blue, thick line. Right panel: contours of fish with score 1 for all traits, and the average worst contour in red, thick line.

Although the manual scoring system is prone to subjectivity, the genetic analysis revealed heritability for all score traits, indicating that the scores were consistent and repeatable. Genetic correlations were stronger between neighboring traits (ranging from −0.10 to 0.70, Tab. 6.2, green highlight) than between non-neighboring traits (−0.15 to 0.17). These correlations could have a biological explanation. For instance, a good ventral side often indicates low visceral fat content, so the caudal peduncle may also tend to be leaner and receive a higher score. However, we cannot rule out that the assessment of one region may also influence the scoring of adjacent regions.

Scores given to individual fish may become biased due to differences in the average level between batches. Batch effect was significant for dorsal side, total score, harvest weight, and both image-derived shape traits. In batches with a lower average quality, scores could be biased upwards. Such bias may lead to unequal numbers of score 3 and score 1 fish per batch contributing to the average best and average worst contour, respectively. Consequently, fish from certain batches might have lower DtB and higher DtW as their batch score is on average higher.

Notably, the image-derived shape trait Distance to the Best (DtB) exhibited the highest heritability among all score and shape traits. This suggests that DtB could offer greater potential for genetic gain if used as an indicator trait for selection, compared to individual score traits. However, in cases where the breeding goal targets the shape of one specific body area, it may still be more effective to use the corresponding regional score trait. Our results highlight an important point: combining scores from different regions may not realistically represent the overall shape evaluation and could lead to unfavorable selection consequences. For example, the genetic correlation between total score and harvest weight was close to 0, suggesting little effect on harvest weight when selecting for overall shape based on total score. In contrast, DtB shows a moderate positive genetic correlation (0.40 ± 0.08) with harvest weight, unfavorable for improving production together with shape.

This study employed an innovative approach to visualize how image analysis derives heritable, descriptive shape traits. Elliptic Fourier Descriptors (EFDs) are widely used in shape analysis to extract and summarize high-dimensional contour features (Valizadeh & Babapour Mofrad, 2022). While EFD coefficients can be visualized through contour reconstruction, their biological meaning is often difficult to interpret, as the coefficients describe abstract curvature frequencies rather than discrete morphological structures, and their genetic relevance is also not straightforward. Previous work has combined EFDs with principal component analysis (PCA) to efficiently summarize overall shape variance in fish (Gayo, Berbel, Korozi, Zerolo, & Manchado, 2023). By examining the effects of increasing component numbers, PCA can describe shape changes associated with specific morphological traits. However, traits derived from principal components are orthogonal in phenotypic variance and therefore expected to show little or no genetic correlation among them. In contrast, our contour-based approach integrates both the definition and visualization of overall shape, rather than dissecting shape into regional, or orthogonal indicators. Landmarks were also used to capture growth-related shape changes (Fragkoulis, Kerasovitis, Batargias, & Koumoundouros, 2021). While landmarks are interpretable and morphologically meaningful, they are limited to separate and predefined regions of interest and may miss capturing visually subtle shape variance. The increased heritability of DtB compared to that of score traits highlight the advantages of shifting from categorical to continuous trait definitions, which enhances data resolution. While total trait variance may be similar, low-resolution traits carry higher measurement error at the individual record level, leading to lower heritability estimates (Blonk, Komen, Tenghe, Kamstra, & Van Arendonk, 2010). Continuous traits reduce this error and allow for greater selection intensity (Smith, Potgieter, & Chapman, 2021; Crain, Reynolds, & Poland, 2017), as each individual measurement is more distinctive.

Both traits open new avenues for shape-related breeding, especially in the areas of automated phenotyping and economic evaluation.

In this study, DtB was derived from the average ideal shape calculated from the best shapes present in the population, based on expert annotation. As more images are collected, the algorithm could select new images based on the established preference and update the reference contour based on similarity to the current average best. This approach can reduce the need for iterative manual grading and annotation, which is often required for every batch and generation. In other situations, the target shape could be defined independently of the current population, for example, by referencing the natural shape of wild fish or shapes preferred in ornamental species.

Currently, DtW is defined by the distance to a single average worst. As shown in Figure 6.6, there is not a single undesirable shape, and this complicates the definition of a trait like DtW. By applying clustering methods on the images of fish with a low total score, we could potentially identify multiple subtypes of undesirable shapes,

such as specific deformities or disproportions. If clear clusters exist then the distances of individual fish to the respective cluster averaged contours could serve as a quantitative measure of deformity or other undesirable aspect of shape, thereby enabling selection against several distinct categories of poor morphology.

At present, the shape of fish lacks direct economic evaluation. Shape preferences differ between markets and among farmers, retailers, processors, and consumers (Mehar, Mekkawy, McDougall, & Benzie, 2020), making its economic value difficult to quantify. In this study the shape traits were derived based on expert scores, but the methodology can extend directly to applications that quantify shape differences between populations; a target shape based on consumer preference scores; or a target based on commercial data from product batches where shape or appearance plays a role in pricing. The visualization of average best and worst shapes allows breeders, as well as farmers and consumers, to perceive and understand the concept of shape ideals and to express their preferences. This intuitive, image-based approach can support surveys and other research into establishing economic values for shape. Alternatively, with our algorithm, any price-relevant preference for shape could be translated into quantitative traits, enabling estimation of genetic parameters and breeding values for selection candidates. Quantifying the value of shape differences would allow breeders to assign appropriate weights and formally incorporate shape into their selection index.

6.5 Conclusion

This study demonstrates the potential of integrating automated image analysis into aquaculture breeding by quantifying fish body shape into two novel traits, Distance to the Best (DtB) and Distance to the Worst (DtW). By leveraging expert-defined shape preferences and contour-based methods, we establish heritable, visually intuitive, and continuous traits. This approach provides an automated and objective method for integrating body shape into aquaculture breeding programs, with potential for economic valuation. The proposed algorithm also allows for updating target shapes, offering flexibility to incorporate alternatives, such as natural shape, or differences in shape ideals from farmers and consumers. Overall, this study represents a valuable step toward the data-driven definition of novel shape traits in aquaculture breeding.

7. General discussion

Aquaculture is the world's fastest growing food production sector with potential for expansion (FAO, 2025). Selective breeding plays an important role in this development, driving improvement in primarily production traits but also disease resistance, environmental resilience, and other new traits of commercial importance (Nguyen, Sonesson,, Houston, & Moghadam, 2022; Gjedrem & Rye, 2018; Janssen, Chavanne, Berentsen, & Komen, 2017; Sae-Lim, Kause, Mulder, & Olesen, 2017; Gjedrem & Robinson, 2014). Although aquaculture is trending towards greater species diversity, a few dominant species continue to account for most production, such as finfish species in the European Union (EU, 2025). This concentration in species facilitates rapid innovation and improvement of techniques and efficiency (Metian, Troell, Christensen, Steenbeek, & Pouil, 2020). Selective breeding itself has also evolved in the past decades, transitioning from conventional, pedigree-based methods to genomic prediction (Fu & Yuna, 2022). For both approaches, the availability of accurate individual phenotypes and the associated phenotyping techniques remain critical.

This general discussion reflects on the use of artificial intelligence (AI) in advancing image-based automated phenotyping to support selective breeding in aquaculture. While the thesis focused primarily on finfish species including gilthead seabream, Atlantic salmon, and rainbow trout, this discussion also aims to provide broader insights relevant to livestock breeding. To achieve this, I first highlight the main challenges encountered in applying AI to image-based phenotyping and examine their implications for selective breeding. Building on these challenges, I propose potential solutions based on the results and knowledge gained from this thesis. Finally, I will conclude with my perspective on the future of image-based automated phenotyping, emphasizing emerging trends in AI and the key developments required for AI-driven applications in animal breeding.

Section 1: Challenges in image-based automated phenotyping

With the advancement of computer vision, particularly through machine learning, breeding programs are increasingly adopting image data to capture phenotypic information.

Although automated phenotyping is often promoted for its non-invasive and high-throughput potential, it is difficult for aquaculture to take advantage of such qualities. Similar to traditional phenotyping methods, the application of automated phenotyping in aquaculture still requires handling and sedating fish. The aquatic environment poses technical challenges for underwater imaging, especially for reliably individual identification of fish from images. As a result, out-of-water imaging currently remains the most sensible practice, with ongoing development of technical devices targeting this purpose.

This thesis employed three portable imaging devices for data collection. These imaging devices were specifically designed for easy transport, installation, and operation within breeding facilities. All imaging sessions were conducted during

routine phenotyping events, when fish were already sedated for standard procedures, such as weighing. The imaging devices can provide immediate phenotypic outputs, including body length, size, and estimated weight. Based on hands-on experience during data collection, the imaging process was significantly faster than manual weighing. Weighing a sedated fish by hand typically takes twice as long as processing the fish through an imaging device. Additionally, the imaging system requires only a single handling event (placing the fish on the machine with a conveyor belt, Fig 7.1) compared to the two steps needed for manual weighing (placing on and removing from a scale). Although sedation and handling remain unavoidable for now and the near future, the development of automated phenotyping systems offers the potential to automate trait collection and reduce handling-related stress and injuries.

Furthermore, the collected images can be repurposed for research and development that do not require real-time output, such as body shape analysis to find novel traits for fish shape selection (Chapter 6), deformities and injuries detection, individual re-identification (Chapter 4) and many more, thereby improving the overall utility of the data.

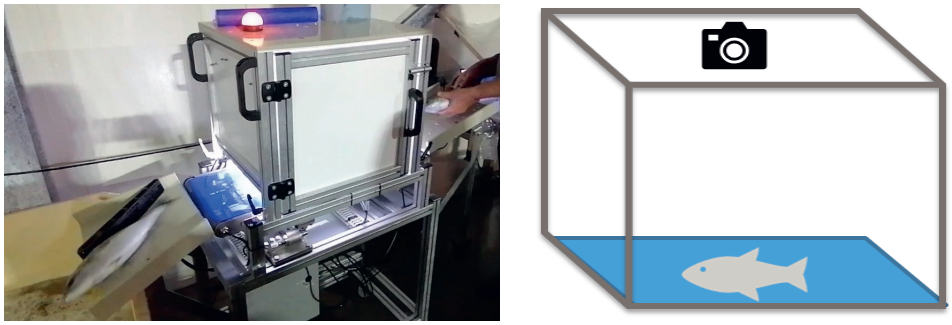


Figure 7.1 Left: Phenotyping device operating in a breeding facility. The picture shows a fish sliding out the machine after being imaged, while someone is waiting to enter a sedated fish into the machine. Right: Schematic illustration of the phenotyping device, where the bottom blue background resembles the conveyor belt that moves the fish in and out the devices.

While current technologies have made data collection increasingly practical and efficient, image-based automated phenotyping has not yet realized its full potential in contributing to selective breeding due to insufficient data quality, limiting phenotyping capacity, and the performance of prediction models. These aspects will be discussed in more detail in the following sections.

7.1.1 Data quality and quantity

Data quality can be assessed in terms of the images themselves and the associated phenotypic information, including both breeding goal traits and indicator traits.

Compared to underwater imaging, out-of-water images can capture the full body of the fish in high resolution and are of better quality. Still, the quality of these images often suffers from variation in lighting conditions, reduced visibility due to water

splashes, and blurring from unintended movement. These factors hinder the extraction of basic morphological features such as body length, height, and volume, as well as more detailed features like melanin patterns used for individual identification (Fig. 7.2. See also Fig. 4.3). Some technical improvements, such as enclosing the imaging device to provide stable lighting or capturing multiple frames in quick succession, can help to reduce these sources of error.

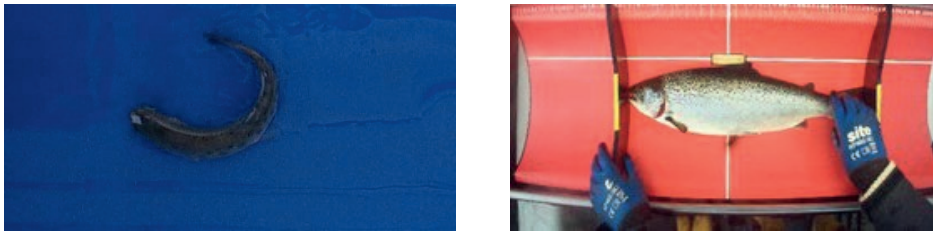


Figure 7.2 Two examples of bad image quality. Left: a rainbow trout moved during the imaging session. Instead of capturing its side lying flat, the image captures the above with a curved shape. Right: s salmon moved during the imaging session. Staff intervened and was captured in the image, blocking part of the fish body.

Another limitation lies in the lack of targeted phenotypic data linked to the images. For supervised machine learning, data annotation is essential for the model to learn the underlying patterns and predict the traits. However, two scenarios often hinder this process. First, the targeted phenotypic data is available, but not the images. For instance, measurements of metabolic traits such as visceral weight or muscle composition typically rely on sibling data from post-slaughter lab analysis (Chapter 2). However, these siblings might not have image records. Similar situations also happened in the experiment described in Chapter 3, where some fish were tested for swimming speed but missed out on the imaging session. This reduces the number of annotated images available to train the prediction model on the target traits. Second, images are available, but their labels lack accuracy. This can happen due to technical errors during data collection or transferring, which also occurred during this thesis, and highlight the need for strict version control and backup in data collection protocol. Besides, label quality can also suffer from subjectivity when based on human annotation. In Chapter 6, shape scores were based on five defined indicators for seabream, but further analysis (van den Berg, 2022) revealed bias: experts tended to under-score in batches with mostly high scores and over-score in batches with mostly low scores. Although these traits showed heritability's up to 0.25, which indicates repeatability in scoring, a convolutional neural network (CNN) trained on such data would "inherit" this scoring bias by the experts. Although standardization can adjust the data and mitigate the effect of bias, I also demonstrated in Chapter 6 that such bias in the scoring of shape was not consistent across batches or individuals. The bias underlay in the definition of the scoring system, that there was a clear standard for the ideal shape, but the less desirable shape varied and lacked clear definitions.

Data quantity can also become a challenge when using images to extract or predict target traits. The amount of data needed to train a machine learning model depends on the complexity of the prediction, but it can range from thousands to millions of data points (Aung, Abdul Razak, & Rahiman Bin Md Nor, 2025). In Chapter 2, it was noted that slaughter traits can vary seasonally and across years, demanding continuous data collection and post-slaughter analysis throughout the breeding cycle to capture such variation and include it in the model training. Optionally, retraining of the model is needed for each season. Moreover, scaling up data collection is sometimes not feasible. For instance, in Chapter 3, the swim tests to measure critical swimming speed are strictly limited by the age and size of the fish. In the next section, I will estimate the number of images needed to train a prediction model in detail.

7.1.2 From pixel information to phenotypes

Beyond data quality and quantity, prediction models also play a key role in extracting phenotypic information from images.

The most intuitive approach to transforming pixel data into phenotypes is through prediction models, for instance, estimating body length from pixel counts. Chapter 2 introduced an analytical framework for comparing different models based on prediction accuracy. This comparison demonstrated that certain traits can be predicted with high accuracy by leveraging their strong genetic correlation with other image-derived traits, such as estimating visceral weight non-invasively using its correlation with body weight, which itself can be derived from image-based size and shape measurements. Besides, images may also reveal additional variation in visceral weight that is independent of body weight, using feature extraction and visualization methods in the analytical framework.

Given that data quantity can be limited, another strategy is to increase the amount of data extracted per individual. This does not necessarily require new measurements of target traits but focuses on decomposing the existing variance in the images into additional, potentially informative indicators. Take the example of shape analysis in Chapter 6, with the target trait being the score representing the breeder's preference. To collect more image data, therefore, it also requires new experts scoring. When applying contour analysis, each contour is decomposed into 1,000 key points. Without acquiring new expert scoring, the key points and the variance between their distance across individuals can be used to derive new shape indicators.

However, a major limitation remains: the lack of defined and validated indicators. Complex trait influenced by many factors may also manifest subtle morphological variation. Identifying what traits are reliably reflected in external morphology, and to what degree, is still a challenging and unresolved issue. Chapter 2 explored the definition of such indicator traits by proposing a framework that uses images to increase prediction accuracy and identify morphological features that are most informative for predicting slaughter traits. Similarly, Chapter 3 applied this concept

to reveal indicators from image data that explain variance in complex traits like critical swimming speed.

Artificial intelligence can support this process of predicting both breeding goal traits and indicator traits. One example is the widely used approach: key point detection. Key point detection captures morphological indicators based on expert-defined anatomical landmarks. These indicators are then used to predict traits such as body weight (Holmes & Jeffres, 2021), slaughter yield (Prchal, et al., 2020), and growth patterns (Fragkoulis, Kerasovitis, Batargias, & Koumoundouros, 2021). In videos, key point detection enables the tracking of specific body parts over time to derive a range of movement-related indicators. These have been applied in tasks like lameness detection in dairy cattle (Russello, van der Tol, Holzhauer, van Henten, & Kootstra, 2024), gait scoring in broilers (Fodor, et al., 2023; Doornweerd, et al., 2021), and tail beat frequency in seabream (Arechavala-Lopez, Lankheet, Díaz-Gil, Abbink, & Palstra, 2021) and Atlantic salmon (Agbeti, et al., 2025). As AI-based tracking models become increasingly sophisticated (Hassan, Mujtaba, Rajput, & Fatima, 2024), this line of research offers promising opportunities for identifying new image-derived indicators (Taghavi, Russello, Ouweltjes, Kamphuis, & Adriaens, 2024; Russello, van der Tol, & Kootstra, 2022).

However, automated phenotyping risks becoming an AI-centered process, where the goal shifts toward technical optimization rather than the delivery of traits aligned with breeding goals. In the example of key points detection, adding more key points or improving tracking resolution may yield a greater number of indicators, but these indicators are not validated for their biological or genetical meaningfulness. Besides, some research proposed new algorithms to increase prediction accuracy without mentioning how such prediction can be used for aquaculture. For instance, a few reviewed papers in Chapter 5 only reported re-identification accuracy of the algorithms, but did not mention the number of individuals involved, nor the time interval between re-identification, making it difficult to judge if these algorithms are best suited for use in a commercial aquaculture setting.

Section 2: Integration of artificial intelligence (AI) for image-based automated phenotyping

In addressing the current challenges in automated phenotyping, it is tempting to view AI as a universal solution. Studies often demonstrate improved performance by applying increasingly advanced algorithms to experimental datasets (Chapter 5), fostering the impression that methodological sophistication alone will drive the field of automated phenotyping forward. While this perspective is understandable given the remarkable progress of AI over recent decades, I argue that such optimism is misplaced. This section aims to critically examine why this belief is misleading and to propose a more grounded perspective on how AI should be integrated into automated phenotyping for selective breeding.

7.2.1 Rationale

Automated phenotyping can be considered a technical *design* aimed at delivering individual-level information for breeding purposes, currently facing substantial challenges. For such design, philosopher Daniel Dennett proposed two metaphors for problem-solving strategies: the **skyhook** and the **crane** (Dennett, 2013).

The **skyhook** is an explanation that relies on a miraculous, unsupported idea:

> *The skyhook concept is perhaps a descendant of the deus ex machina of ancient Greek dramaturgy: when second-rate playwrights found their heroes into inescapable difficulties, they were often tempted to crank down a god onto the scene, like Superman, to save the situation supernaturally. … Skyhooks would be wonderful things to have, great for lifting unwieldy objects out of difficult circumstances, and speeding up all sorts of construction projects. Sad to say, they are impossible.*

Dennett contrasted this with **cranes**—grounded, incremental mechanisms built from existing structures:

> *Cranes can do the lifting work our imaginary skyhooks might do, and they do it in an honest, non-question-begging fashion. …They have to be designed and built, from everyday parts already on hand, and they have to be located on a firm base of existing ground.*
>
> *Each step has been accomplished by brute, mechanical, algorithmic climbing, from the base already built by the efforts of earlier climbing.*

These metaphors map directly onto how AI is currently integrated in breeding programs. Too often, models developed for other domains are applied with the hope they will function similarly in aquaculture breeding programs, without fully accounting for the breeding and aquaculture context. These risks bypassing the necessary "climbing", which is the slow, cumulative work of challenging AI with real-world breeding program constraints. Phenotyping is not an isolated task but one embedded in a structured hierarchy where data collection, trait definition, selection decisions and many other steps are deeply interconnected.

Chapter 4 of this thesis illustrates the limitations of the skyhook approach. When applying individual re-identification algorithms previously successful in experimental settings to data from breeding facilities, model accuracy approximated zero. Even with more advanced model structures, performance remained poor. Chapter 5 further investigated this by reviewing past work on individual re-identification. A common issue was that most studies avoided confronting practical challenges like motion blur or inconsistent lighting by excluding such flawed images. While this improves algorithmic performance, it comes at the cost of excluding individuals, potentially affecting the impact of breeding decisions. Such approaches

often follow the standards of computer vision research practice but fail to address the approaches' relevance for breeding programs.

Improving automated phenotyping, therefore, requires reflection on how AI fits into the different steps of breeding programs. What are the prerequisites for AI to be effective, and what impact do these tools have on selection outcomes?

As Chapter 4 demonstrated, confronting reality can yield disappointing results, which can discourage publication or erode researcher motivation. However, as Pirsig once wrote:

> *An experiment is never a failure solely because it fails to achieve predicted results.* And *...to design an experiment properly he has to think very rigidly in terms of what directly causes what. This you know from hierarchy* (Pirsig, 2014).

Here, Pirsig's "hierarchy" aligns closely with Dennett's crane metaphor. In a breeding program, this hierarchy includes all upstream and downstream steps of the collection of information. The integration of AI in automated phenotyping, therefore, must also align with the breeding program structure.

7.2.2 Opportunities to integrate AI in automated phenotyping for selective breeding

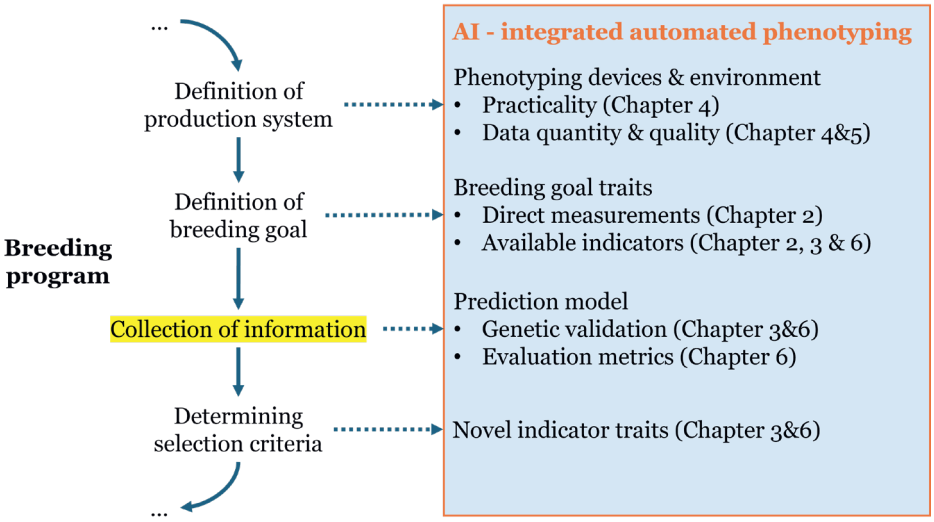


Figure 7.3 The integration of artificial intelligence in automated phenotyping in alignment with different steps of a breeding program. The highlighted step, collection of information, is where automated phenotyping conventionally focused on.

The integration of AI into automated phenotyping should not be limited to improving information collection alone; it must also extend to other critical steps in the breeding program regarding the production system, breeding goals and the

selection criteria (Fig 7.3), which will be discussed in more detail in the following sections.

Phenotyping devices and environment

To integrate AI into automated phenotyping, it is crucial to understand the practical constraints and their consequences for data quantity and quality. Most AI algorithms, particularly those in computer vision were originally developed for human-centered applications such as facial recognition and autonomous driving. Accordingly, these algorithms perform best in scenarios tailored to human subjects. It would be naive to assume that tools trained on human datasets will transfer seamlessly to animal data. While performance issues can be addressed through additional data collection and repeated trials, these efforts require time, resources, and domain-specific expertise, which are often limited in breeding facilities.

As previously discussed, phenotyping devices are helpful but remain in the early stages of development. Image collection in animal breeding is also more complex. For instance, while collecting a flower recognition dataset with contributions from anyone may be easy, access to selection candidates is limited, and associated trait data may be subject to ownership restrictions. Practical challenges in breeding facilities such as camera blockages, variable daylight in livestock housing, or humidity and salinity exposure in aquaculture, can further complicate image acquisition. Moreover, annotation of image data often requires invasive lab analysis or expert judgment, which contrasts with the relative ease of annotation in human-focused applications (Fig 7.4).

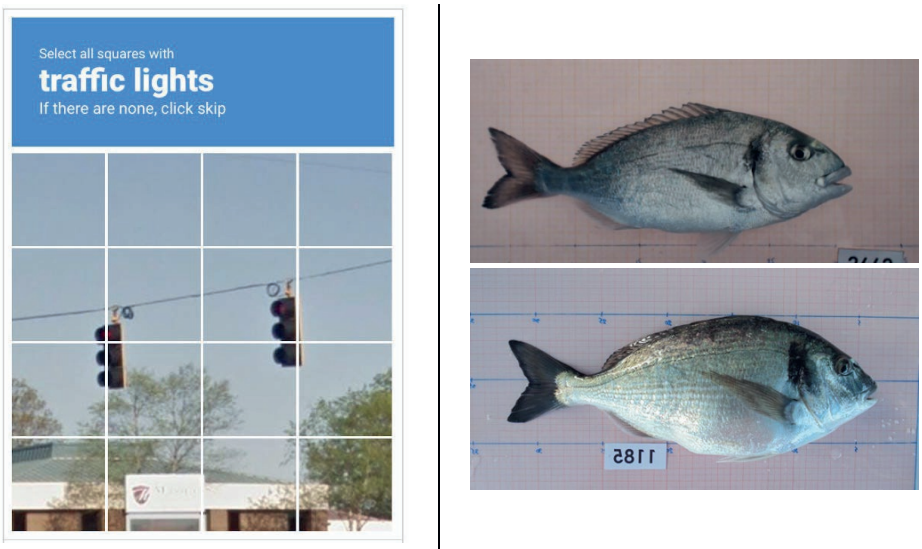


Figure 7.4 Two examples of annotation. Left: A CAPTCHA, a type of challenge–response Turing test. Often seen when accessing websites. The location of the traffic lights, annotated by websites users, are fast and only require common knowledge. Right: Fish scoring, where two fish need to be given scores (1-3) based on multiple shape indicators. Such assessment requires training to understand what the indicators are, how shape can differ and what the distribution should be per fish batch. These two examples, by contrast, show the differences in data annotation in human-centered application and breeding practice.

Therefore, understanding the definition of a production system is not about advancing imaging technologies or replicating algorithm development processes from human applications. Instead, it is about acknowledging these differences and making informed adjustments to the requirements for data quality and quantity, as well as to the expectations on AI model performance, in subsequent steps.

As for breeders, it is important to distinguish between the production environment and the phenotyping environment. The production environment refers to the actual setting where animals are kept, whereas the phenotyping environment includes more specific factors that influence the captured images rather than the phenotypes themselves, such as lighting conditions, camera angles, and image distortion. This distinction is discussed in more detail in a later section, on the genetic validation of the prediction model.

Breeding goal

To integrate AI in automated phenotyping, it is essential to understand the breeding goal clearly. While this might appear straightforward, the definition of breeding goal directly determines several downstream factors in the breeding program. As breeding goal is often not a single trait but a combination of traits, I divided this section into breeding goal traits and indicator traits which can be included in the selection index.

Breeding goal traits

Automated phenotyping can change the scope of breeding goal traits with broader possibilities of indicator traits. One major advantage of image-based phenotyping is its non-invasive nature, which may open opportunities to directly predict breeding goal traits that are traditionally invasive or expensive to measure. For instance, fillet yield in seabream can be predicted on selection candidates using image data (Gulzari B. , 2023), instead of using data from siblings post slaughter.

Indicator traits

The main contribution of AI lies in identifying indicator traits through prediction. Take the example in Chapter 3. Critical swimming speed (Brett, 1964), as one of the most common estimates of swimming performance, could serve as an indicator for health (Castro, et al., 2013; Tierney & Farrell, 2004). If selecting fish with higher swimming speed for health improvement, it becomes relevant to also consider morphological traits that physiologically influence swimming and are genetically correlated to production traits. The combination of convolutional neural network (CNN) and GradCAM can identify morphological features that explain swimming performance but also related to production traits such as fillet yield. Such traits like epaxial muscle volume can be a good indicator trait to include with swimming performance, if to select for fillet yield.

The genetic parameters of AI predicted traits need to be estimated to assess their value for selective breeding. This refers to the heritability of the AI predicted trait, genetic correlations of the AI predicted traits with breeding goal traits, and with indicator traits.

First, the predicted traits need to be heritable, otherwise their value as an information source for breeding goal traits is limited. Sometimes predicted traits exhibit higher heritability than the traditional indicators, increasing their value as indicator traits. Second, predicted traits need to show sufficiently strong genetic correlations with breeding goal traits to replace certain traits in the selection index that are expensive or invasive to measure. For instance, in Chapter 6, image-derived trait DtB (Distance to the Best) showed higher heritability than total score for shape measurements. Besides, with harvest weight, DtB also showed stronger, negative genetic correlation than that of total score. Therefore, DtB captured better the genetic consequence of shape if selected for harvest weight, and vice versa, and can be more suitable than total scores as a shape indicator.

Besides sufficient genetic correlation with breeding goal traits, predicted traits with clear economic benefits or technical accessibility can also be considered as indicator traits. For example, the morphological indicator epaxial muscle volume in Chapter 3. It showed higher genetic correlation with swimming speed than body volume, and it might also correlate with fillet yield, another important breeding goal trait. Therefore, including epaxial muscle as indicator can capture more of the complex relationship between breeding goal traits, and lead to better selection decisions.

Due to the lack of defined and validated indicators to be predicted from images, it is difficult to estimate the number of images needed to train a prediction model, as it depends on the complexity of the target traits. If predicted traits are estimated on their heritability and genetic correlation with breeding goal traits, the number of individuals required imaging can be deducted from the different scenarios of the genetic parameter estimation, to maximize the accuracy of subsequent breeding value estimation of the predicted traits.

For the very first session, all selection candidates should be imaged. For target traits that require invasive or post slaughter measurements, 20 full sibs per family are needed, preferably 50/50 in sex ratio, assuming no common environmental effect.

The number of selection candidates can vary per breeding program, but it should guarantee an accurate heritability estimation for breeding goal traits such as harvest weight and is therefore essential to be included in phenotyping. Consequently, the AI predicted traits will have records of own performance from all selection candidates. A baseline seabream breeding program can include 2,125 selection candidates (Janssen, Saatkamp, & Komen, 2018). The number of selection candidates in Chapters 3 and 6 are 1,200 and 3,421, respectively. In this thesis, the heritability was estimated at 0.24±0.07 for critical swimming speed (Chapter 3), and at 0.32±0.040 for the shape indicator (Chapter 6), suggesting the available amount of data was sufficient for accurate heritability estimation, with reasonably low standard errors.

Some phenotypes, such as diseases resistance, cannot be measured directly on selection candidates. Therefore, both the choice of which animals to image and the number of animals imaged are critical. The EBV accuracy of the predicted traits themselves can be used to assess the return on investment when phenotyping individuals that are not themselves selection candidates.

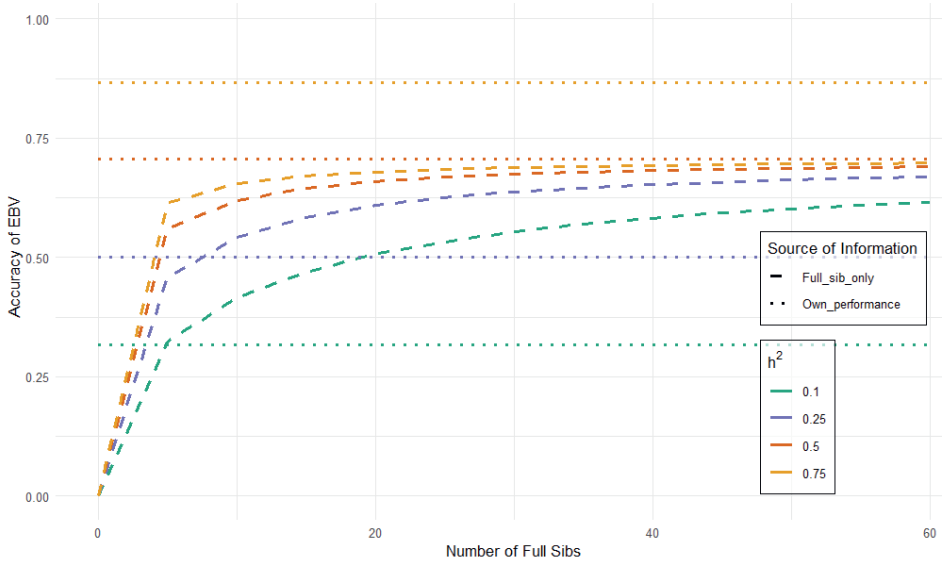


Figure 7.5 Accuracy of EBVs using full sibs and own performance under different heritability of the predicted traits.

Figure 7.5 illustrates the effect of phenotyping full sibs on the accuracy of EBVs of predicted traits with different heritability, compared to using animals' own performance. Including 20 full sibs in phenotyping gives the most improvement in EBV accuracy, especially for predicted traits with low heritability.

To summarize, the integration of AI into automated phenotyping must align the AI prediction output with breeding goal and the indicator traits. This is particularly important for broader breeding goals such as those that include health and welfare traits. Checking the heritability of predicted traits early on is advisable, as this may determine the usefulness and cost-effectiveness of phenotyping.

Prediction model

As mentioned earlier, AI applied to images has largely been developed with a focus on human applications. While automated phenotyping will continue to benefit from advances in AI-powered tools, it is also important to evaluate and adapt these tools within the context of animal breeding.

Previously, I discussed the output of prediction models and its utilization to describe breeding goals. In this section, I will highlight two key aspects in more detail in the following sections: the **genetic validation** of AI-based prediction models and the importance of **breeding-relevant evaluation metrics**.

Genetic validation

Let the target trait be decomposed as

$$P = G + E + \varepsilon$$

Where G is the genetic effect and E is the environmental effect. In the context of breeding, E represents specific, systematic environmental contributions such as geographical location and production system, and Var(E) is used to quantify the variance.

Now we define X, representing all image features AI model used to predict the target trait:

$$P_{prd.} = f(X)$$

These features can be as simple as raw pixel value, or actual measurable traits such as body length and width. More importantly, part of X reflects genetic component, through the genetic correlation between the predicted trait and the target trait, while other parts of X may correlate with environmental effects:

$$P_{prd.} = g_{ind.} + f(X_E) + \varepsilon$$

Where $g_{ind.}$ represents the genetic effect captured by the AI model, typically through indicator traits derived from image features, while$f(X_E)$ reflects variance attributable to environmental factors, such as differences in production systems or imaging artifacts (e.g., lighting conditions, camera angles), as there is no clear distinction between the sources of information. Moreover, a fundamental difference lies between a linear mixed model, also called animal model, and an AI model when it comes to $f(X_E)$. Animal models explicitly estimate environmental variance and effects, and model gene-by-environment interactions (G×E), allowing a clearer separation of genetic and environmental contributions. While in AI model, data augmentation is often used to introduce variance in imaging artifacts to reduce overfitting and force the model to focus on robust patterns, implicitly reducing the environmental effects on the final prediction.

Model performance is often optimized by minimizing the difference between P and $P_{prd.}$:

$$\min (P_{prd.} - P) = min(g_{ind.} - G) + min(f(X_E) - E)$$

However, reducing $(f(X_E) - E)$ alone can improve model performance while inadvertently increasing overfitting to specific environments. Therefore, the heritability of the predicted trait should also be considered when optimizing and validating prediction models, especially across different environments:

$$h^2_{prd.} = \frac{Var(g_{ind.})}{Var(P_{prd.})}$$

Heritability indicates how much of the predicted variance comes from genetics. If prediction accuracy improves but $h^2_{prd.}$ remains unchanged, it signals overfitting, meaning the model will not benefit from further training or using a more complex structure.

Note that there may be covariance between the genetic effect of predicted indicator $g_{ind.}$ and environmental variance captured by $fX(e)$, like the G×E effect. Such covariance would result in a portion of the genetic variance in predicted trait being incorrectly attributed to measurement error, reducing the accuracy of genetic parameter estimation. For example, Zhou et al. (2021) reported that genetic variation explained 27–59% of the measurement error in image-derived sorghum height, while Bi et al. (2025) found no such effect for pig body weight. These contrasting findings suggest that the presence and impact of this covariance may depend on the trait and species. It also may reflect systematic biases, such as certain genotypes being over- or under-estimated using image-based automated phenotyping.

Evaluation metrics

Metrics are a critical part of evaluating prediction model performance, especially for traits extracted from image data. However, conventional computer vision metrics may not always be optimal to assess the value of prediction models for making progress towards the breeding objective. I argue that the development of additional, breeding-specific metrics is essential for integrating AI into automated phenotyping.

Many computer vision metrics focus on technical accuracy. For example, if key points are missing or misplaced when detecting body landmarks, it is often seen as a failure. But in breeding, if the remaining points still produce reliable phenotypic estimates, such errors may be acceptable. Considering that, the study of Liu et al. (2024) proposed metrics which evaluate accuracy based on how well the phenotype can still be estimated given the possible key point detection error, rather than on key point detection alone.

Choosing breeding-related metrics is crucial. In Chapter 3, for example, image data was used to predict critical swimming speed. While the prediction accuracy appeared limited under standard metrics (e.g., Pearson correlation), the model was still able to identify body regions with higher or lower genetic correlation to the target trait. Therefore, even a “low accuracy” model by computer vision standards may still be valuable for genetic applications. Optionally, models can be evaluated based on the genetic correlation between the predicted traits and breeding goal traits, to maximize the utilization of potential indicator traits.

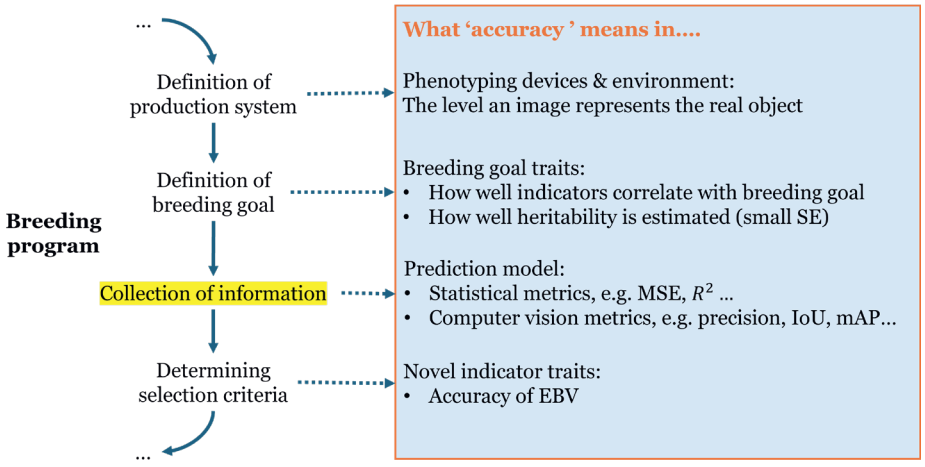


Figure 7.6 The meaning of 'accuracy' can differ at different stages of AI integration in breeding program.

Further development in automated phenotyping should therefore promote the use of breeding-oriented evaluation metrics. This also facilitates better interdisciplinary communication. For example, 'accuracy' is a commonly used term, but it can mean different things across disciplines. Breeders might expect a model to produce EBVs directly, while computer vision researchers focus on how well the detected object overlaps with the annotation (Fig. 7.6).

Some metrics are shared across the field, but ambiguity remains, particularly for individual re-identification (Chapter 5). Many studies report accuracy as a key metric, but definitions vary. One approach is to assess the percentage of individuals correctly re-identified within a given image set. Another approach assesses the proportion of image pairs correctly classified as belonging to the same individual. For breeding applications, it is essential to assess these methods under realistic conditions, such as longer time intervals between images or larger numbers of individuals to test scalability. However, current accuracy metrics often fail to reflect these considerations. As a result, some studies report re-identification accuracy without specifying the total number of fish, limiting their practical implementation. Metrics should therefore reward models that maintain performance over time or across larger datasets, enabling comparisons between studies and ensuring practical relevance for breeding. Compared to standard computer vision metrics, such breeding-specific metrics are more explicit, as they directly address the operational demands of breeding programs.

A critical metric that needs to be developed is the quantification of data leakage and its impact on model evaluation. In supervised learning, models typically require a training-validation-test split, often in a ratio of 3:1:1, with performance evaluated on the test set. Data leakage occurs when the same data points are used in both training and testing, leading to overly optimistic performance estimates. If this principle were

strictly applied to breeding programs, it would imply that none of the selection candidates' image could be used for training, tripling the number of individuals required for phenotyping. The impact of including predicted traits from both training and test set for genetic analysis is currently unknown and requires more investigation. Therefore, it is essential to develop metrics that quantify or model the impact of data leakage, ideally incorporating pedigree relationships. The same applies for genomic prediction, where data leakage in phenotypic prediction can affect the determination of reference population size.

To summarize, integrating AI into automated phenotyping requires aligning expectations across computer vision and selective breeding, and employing metrics that are interpretable, useful, and explicit in both fields.

Novel indicator traits

The most exciting opportunity brought by AI-integrated automated phenotyping lies in the definition and discovery of *novel indicator traits*. In this thesis, I explored this possibility in three main directions: non-invasiveness, objectivity, and longitudinal.

First, AI can help identify morphological indicators for traits that are conventionally measured invasively. Selecting indicators with high genetic correlation to the breeding goal trait can contribute to the improvement of genetic gain. Chapter 2 introduced an analytical framework that uses image data to predict slaughter traits in a non-invasive manner. Chapter 3 further developed a workflow to identify fine-scale morphological traits with the highest genetic correlation to complex traits.

The second direction is in the refinement of subjective, qualitative traits into objective, quantitative traits with higher heritability. In Chapter 6, I proposed two image-derived shape traits: Distance to the Best (DtB) and Distance to the Worst (DtW), through visualization of subtle variance in body shape based on five categorical shape indicators traditionally used. Moreover, both DtB and DtW demonstrated higher heritability than the original indicators, suggesting that the image-based automated method not only increases measurement precision but also the potential for genetic improvement.

The third direction regards longitudinal traits that are derived from repeated measurements over time. Longitudinal traits capture dynamic changes and are valuable for resilience and responses to environmental variation, such as recovery from stress challenges (Spiliopoulos, Brown, Hilder, Tilbrook, & Descovich, 2025),. Extracting these traits requires well-structured datasets. Chapter 5 outlines the essential steps for building such datasets, including defining essential data content, establishing robust data structures, and implementing thorough quality assessments.

Novel traits may further benefit breeding programs in two specific ways. First, they can offer biological and genetic insight by creating a more direct link between phenotypic variation and its genetic cause. Breeding goal traits such as growth rate, survival, or product quality often represent the combined outcome of many

underlying biological processes, each influenced by multiple loci and environmental effects. This complexity makes it difficult to detect individual QTL, as a single QTL only contributes a minor part of the total variation. Indicator traits, in contrast, are designed to capture a specific component of the breeding goal traits, such as a morphological feature, a physiological rate, or a behavioral pattern.

Second, novel traits can help translate phenotyping outcomes into economic value. For optimizing breeding programs, it is needed to know the economic importance of traits. For instance, the shape traits are currently defined by expert opinion on a scale of 1 to 3, which is not yet assigned economic market values. Chapter 6 proposed a visualization of the ideal shape for selection based on the expert scores, which could be extended to include preferences from farmer, processors, retailers and consumers (Mehar, Mekkawy, McDougall, & Benzie, 2020) and eventually translate these preferences into quantitative traits. These traits should not be used to select for a specific desirable shape, but rather to eliminate fish with extremely low market value due to unfavorable body shape, which may also indicate subtle deformities. Similarly, longitudinal traits facilitate market diversification. A good example is selective harvest in aquaculture, where different fish sizes are preferred by different markets and priced accordingly (Janssen, Berentsen, Besson, & Komen, 2017). Using longitudinal phenotypes, such as daily weight gain, breeding programs can be optimized to match multiple market strategies and improve economic return (Sun, et al., 2022).

In summary, the successful integration of AI in automated phenotyping relies not only on algorithmic performance but also on alignment with the broader breeding program. This includes understanding the production system and breeding goals, capturing accurate individual phenotypes through genetically validated and evaluated models. Novel traits serve as an important bridge, further linking AI with biological and economic perspective of the breeding goals.

Section 3: Towards smart observation of animals for genetic improvement

The successful integration of AI into automated phenotyping has the potential to support genetic improvement across various stages of the breeding program. To realize this potential, this section explores how advanced AI can further enhance selective breeding, particularly through smarter, more informative observations of individual animals. These developments remain largely unexploited in current phenotyping strategies and represent a promising frontier for future research. While the focus of this thesis has been on image data, the same principles can be extended to other data types such as sensor signals from movement detectors or other automated monitoring technologies.

7.3.1 Generative artificial intelligence

Earlier I discussed both the limitations and opportunities in data collection, quality, and quantity. Besides technical development of phenotyping devices, Chapter 2

advocates routine image collection as a standard practice within breeding programs, while Chapter 5 highlights the importance of data quality assessment and structured data organization in aquaculture. These requirements apply not only to image data but also to other data types used for animal breeding. Improving the description, structure, and accessibility of such datasets would greatly benefit communication and reuse across the research community. However, as most datasets are collected as part of commercial breeding program, the improvement on description, structure and accessibility should start with datasets that are already made available, for instance, through scientific publication.

An emerging trend besides conventional data collection outside the field of animal breeding is generative artificial intelligence (GAI), which can synthesize new data based on existing datasets. This is particularly promising for underrepresented or sensitive data, such as those related to certain behaviors or health, where labelled data are often scarce. For instance, GAI is used to generate images of diseased plants as training data for model to recognize and detect common diseases and their progress for plant phenotyping (Lokesh, et al., 2024; Klair, Agrawal, & Kumar, 2024; Liu B. , Tan, Li, He, & Wang, 2020). A further advantage of GAI is its potential to facilitate data anonymization. Much of the data related to automated phenotyping remains privatized, whether owned by breeding companies or farmers, complicating data sharing and aggregation across ownership boundaries. GAI offers a potential workaround by synthesizing new data that preserves the statistical relationships and structure of the original, without exposing the raw source data (Yoon, Drumright, & Van Der Schaar, 2020).

Nonetheless, the use of GAI for generating phenotype data presents its own risks (Pérez-Enciso, Zingaretti, & de Los Campos, 2025). A key concern is dataset bias, which can propagate into synthetic outputs. For example, when generative image models are asked to depict livestock farms, they may default to outdoor, pastoral scenes (Sheng, Tuyttens, & von Keyserlingk, 2025). To some extent, GAI reflects collective stereotypes of livestock and aquaculture (Fig. 7.7). Another concern is that synthetic data often exhibits reduced variance compared to real-world data, which can lead to overfitting and model collapse (Shumailov, et al., 2024).

Figure 7.7 ChatGPT 4.0 generated image using prompt 'salmon in a breeding facility'. While the salmon appear realistic in form, they are depicted as motionless, floating in the water, and display minimal visible phenotypic variation in traits such as size or body shape. The generated environment places the farm on a lowland plain surrounded by hills, which does not match the typical coastal landscapes where most salmon farms are located (Manolin, 2025).

7.3.2 Unsupervised learning

Currently, supervised learning remains the most applied machine learning approach. However, there is considerable potential in exploring unsupervised learning methods for automated phenotyping (Gladju & Kanagaraj, 2021).

One promising application of unsupervised learning is its ability to reveal hidden patterns within datasets. This is valuable when identifying indicator traits for behavior, based on continuously collected data on animal activity (McVey, Hsieh, Manriquez, Pinedo, & Horback, 2023; Ahn, Kim, & Jeong, 2023). It is especially relevant for detecting behaviors that are difficult to observe or annotate, such as feather pecking in poultry (Rossi, et al., 2025; Subedi, Bist, Yang, & Chai, 2023), tail biting in pigs (Paula, 2025; Wang Z. , 2025), grazing (Ditria, Jinks, & Connolly, 2021) and reduced foraging behavior after stress exposure in fish (Martins, et al., 2012). Developing algorithms for such behaviors often requires large datasets of the events in question, which is difficult to collect and unethical to induce for problematic behavior such as aggression. Moreover, behavior detection for management demands low latency, enabling farmers to act immediately. However, this is less critical for breeding purposes, where phenotypes only need to be available at the time of selection. In addition, management typically needs cohort-level data, whereas breeding requires phenotypes to be recorded at the individual level. For interactive behaviors such as pecking, precise labelling is required to distinguish the performer and the recipient for accurate trait assignment.

An alternative approach is to analyze the activity pattern before and after specific events at the individual level, using video analysis, location tracking, or the fusion of multiple data sources such as locomotion sensors. Unsupervised learning can be used to uncover latent sources of variance related to these behaviors. For example, cattle showing a gradual decline in daily activity or movement could signal early signs of lameness.

Unsupervised learning leverages the natural patterns within data and phenotypic variance without relying solely on human annotation. In this sense, it enables a shift from a human-centric to an animal-centric perspective: whereas supervised approaches model variance within a definition, unsupervised methods reveal sources of variance that then require biological or genetic interpretation. Large-scale activity data can serve as a baseline to define normal behavior profiles, and any deviations from these norms may represent indicators of health or welfare disturbance. These deviations, when consistently measurable across individuals, hold promise as source of variance for novel, unbiased phenotypic traits. A successful example is the log-transformed variance (lnVar) of daily deviation. LnVar measures the variation of deviations of a trait from its expected value. It captures the stability of an animal's performance over time using the expected value as a baseline, and is used as an indicator for resilience (Casto-Rebollo, Nuñez, Gol, Reixach, & Ibáñez-Escriche, 2025; Aththar, 2024; Poppe, 2022). However, deriving and defining traits from such variance requires validation of their biological and physiological relevance, as well as estimation of their heritability and genetic correlation with breeding goal traits, before considering their inclusion in the selection index.

7.3.3 Explainable artificial intelligence

The concept of explainable artificial intelligence (XAI) is that it aims to improve the interpretability and transparency of AI models, both in terms of how they function and how they produce results. By opening the black box to some extent, explainable AI can also help to build trust. This is, in many ways, like the development of selective breeding. Take neural networks, one of the most applied AI methods as example. Each neuron contains certain functions used to receive, process and pass information from neighboring neuros. Neural networks are widely applied in complex tasks such as object classification, facial recognition and self-driving vehicles. However, the exact process where imaginal input is transferred into certain decisions is numerically accessible but not transparent enough to interpret. Like that of selective breeding. Genetic improvement is a complex output resulting from the interaction of genes, many of whose functions remain undiscovered. However, the design of breeding programs to achieve genetic gains does not require breeders to be omniscient on every aspect of genes. By clarifying what information is collected, processed, and used at each step of a breeding program, the process becomes scientifically grounded and communicable to breeders in familiar terms, such as genetic gain or economic return. Similarly, AI models generate results by capturing variance patterns within the training data, with their reasoning encoded through

adjustments to thousands of model parameters. XAI provides a means to trace these learned patterns back to interpretable biological or morphological features, thereby bridging the gap between statistical abstraction and practical breeding insight.

To effectively integrate artificial intelligence into automated phenotyping, a similar level of transparency is essential. For instance, in Chapter 3, I used GradCAM, a widely used visualization tool in XAI, to identify the regions of the image that most strongly influenced the model's prediction of swimming performance. By annotating these regions and calculating their genetic correlation with swimming speed, the relevance of these image areas was transformed from abstract visual features in a computer vision model to biologically meaningful morphological indicators. Interestingly, these GradCAM-highlighted regions also showed stronger genetic correlations with swimming speed than other areas, making the revealed indicators more persuasive for breeding decisions. This process of translating both AI reasoning and outputs into biologically interpretable terms is a critical step toward a broader acceptance beyond the level of AI developers. It strengthens the credibility of automated phenotyping tools and encourages their integration into real-world breeding programs (Mallinger, et al., 2024; Hoxhallari, Purcell, & Neubauer, 2022).

All these AI advancements open new opportunities to extract complex, previously inaccessible information from animals, particularly for traits related to behavior, health, and welfare, which are still underrepresented in automated phenotyping or as general breeding goals. However, as with the natural patterns identified through unsupervised learning, such traits must be validated, both from a genetic perspective, in relation to other traits within the breeding program, and from biological and physiological perspectives, in relation to the overarching goals of improving health and welfare. Achieving this requires thorough interdisciplinary examination, involving expertise from genetics, physiology, ethology, artificial intelligence, and ethics to ensure that the traits are both biologically meaningful and genetically actionable.

Altogether, image-based automated phenotyping for selective breeding will continue to connect phenotypic description with genetic insight, shifting our understanding from animals as products to animals as cohabitants. From collective knowledge and expert experience to artificial intelligence, technology continues to shape the evolving relationship between animals and humans, and offers new possibilities for how we observe, understand, and care for them.

Conclusion

In this thesis, I focused on the integration of artificial intelligence into image-based automated phenotyping in aquaculture, addressing both breeding goal traits and indicator traits.

Artificial intelligence holds great promise for advancing automated phenotyping and expanding the scope of what can be recorded and understood about animals.

However, its full integration requires a deeper understanding of the breeding program. This includes considerations of the breeding system, the definition of breeding goals, the implementation of advanced machine learning models, and the genetic validation aligned with selection criteria. Other aspects of artificial intelligence, such as generative AI, explainable AI, and unsupervised learning, can further support key steps in the breeding program.

With such integration, automated phenotyping can become not only more accurate and efficient but also more transparent, biologically grounded, and economically justified. This will improve its credibility and facilitate its broader application across aquaculture, livestock, and beyond.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., . . . others. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

Abbink, W., Palstra, A., Agbeti, W., Lembo, G., & Komen, J. (2022). The possibilities of using electronic sensors in aquaculture breeding. In *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)* (pp. 2395--2398).

Agbeti, W. E., Palstra, A. P., Black, S., Magnoni, L., Lankheet, M., & Komen, H. (2025). Atlantic salmon (Salmo salar) under challenge: Heart rate and acceleration dynamics during exercise and stress. *Frontiers in Physiology*, 1562665.

Ahmad, A., Wold, J. P., Sonesson, A. K., Hatlen, B., Dagnachew, B. S., Berg, P., . . . Difford, G. F. (2025). Genetic and phenotypic validation of whole body fat content measured across production phases of Atlantic salmon using dielectric and near infrared Interactance spectroscopy. *Aquaculture*, 741747.

Ahmet Doğdu, S., & Turan, C. (2021). Genetic and morphological impact of the cultured gilthead sea bream (Sparus aurata Linnaeus, 1758) populations on wild stocks. *Egyptian Journal of Aquatic Biology and Fisheries*, 499--511.

Ahn, S.-H., Kim, S., & Jeong, D.-H. (2023). Unsupervised Domain Adaptation for Mitigating Sensor Variability and Interspecies Heterogeneity in Animal Activity Recognition. *Animals*, 3276.

Aldhafeeri, T., Tran, M.-K., Vrolyk, R., Pope, M., & Fowler, M. (2020). A review of methane gas detection sensors: Recent developments and future perspectives. *Inventions*, 28.

Al-Jubouri, Q., Al-Azawi, R., Al-Taee, M., & Young, I. (2018). Efficient individual identification of zebrafish using Hue/Saturation/Value color model. *Egyptian journal of aquatic research*, 271--277.

Al-Jubouri, Q., Al-Nuaimy, W., Al-Taee, M., & Young, I. (2017). An automated vision system for measurement of zebrafish length using low-cost orthogonal web cameras. *Aquacultural Engineering*, 155–162.

Altimiras, J., Axelsoon, M., Claireaux, G., Lefrancois, C., Mercier, C., & Farrell, A. (2002). Cardiorespiratory status of triploid brown trout during swimming at two acclimation temperatures. *Journal of Fish Biology*, 102--116.

Arechavala-Lopez, P., Lankheet, M. J., Díaz-Gil, C., Abbink, W., & Palstra, A. P. (2021). Swimming activity of Gilthead Seabream (Sparus aurata) in swim-tunnels: accelerations, oxygen consumption and body motion. *Frontiers in Animal Science*, 679848.

Ariza-Sentís, M., Vélez, S., Martínez-Peña, R., Baja, H., & Valente, J. (2024). Object detection and tracking in Precision Farming: A systematic review. *Computers and Electronics in Agriculture*, 108757.

Athammer, R. R. (2023). *Tolerance limits to fluctuating water currents in Atlantic salmon (Salmo salar): A novel method to simulate the impact of ocean waves on salmon welfare in the laboratory.* MSc thesis, The University of Bergen.

Aththar, M. H. (2024). *PhD thesis: Breeding for resilient growth in tilapia*. Wageningen University and Research.

Aung, T., Abdul Razak, R., & Rahiman Bin Md Nor, A. (2025). Artificial intelligence methods used in various aquaculture applications: A systematic literature review. *Journal of the World Aquaculture Society*, e13107.

Balaban, M. O., Ünal Şengör, G. F., Soriano, M. G., & Ruiz, E. G. (2010). Using image analysis to predict the weight of Alaskan salmon of different species. *Journal of food science*, E157--E162.

Banno, K., Gao, S., Stolz, C., Tuene, S. A., Aas, G. H., & Gansel, L. C. (2025). Multi-beam sonars for monitoring the distribution and welfare of fish reared in sea cages-strengths and challenges. *Aquaculture*, 742553.

Barreto, M. O., Rey Planellas, S., Yang, Y., Phillips, C., & Descovich, K. (2022). Emerging indicators of fish welfare in aquaculture. *Reviews in Aquaculture*, 343--361.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). Surf: Speeded up robust features. In *Computer Vision--ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9* (pp. 404--417). Springer.

Beamish, F. (1978). Swimming capacity: fish physiology. *Locomotion*, 101--187.

Bekkozhayeva, D., & Cisar, P. (2022). Image-based automatic individual identification of fish without obvious patterns on the body (scale pattern). *Applied Sciences*, 5401.

Besson, M., Rombout, N., Salou, G., Vergnet, A., Cariou, S., Bruant, J.-S., . . . others. (2022). Potential for genomic selection on feed efficiency in gilthead sea bream (Sparus aurata), based on individual feed conversion ratio, carcass and lipid traits. *Aquaculture Reports*, 101132.

Bi, Y., Huang, Y., Yu, H., & Morota, G. (2025). Impact of trait measurement error on quantitative genetic analysis of computer vision derived traits. *bioRxiv*, 2025--06.

Billah, M., Bermann, M., Hollifield, M., Tsuruta, S., Chen, C., Psota, E., . . . Lourenco, D. (2025). Genomic selection in the era of phenotyping based on digital images. *animal*, 101486.

Black, A., & Nederpelt, P. v. (2020). *Dimensions of Data Quality (DDQ) Research Paper*. DAMA NL Foundation.

Blidh, H. (2021). *pyefd 1.6.0*. Python packages. Retrieved from https://pypi.org/project/pyefd/

Blonk, R., Komen, J., Tenghe, A., Kamstra, A., & Van Arendonk, J. (2010). Heritability of shape in common sole, Solea solea, estimated from image analysis data. *Aquaculture*, 6--11.

Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., . . . others. (2016). End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*.

Boudry, P., Allal, F., Aslam, M. L., Bargelloni, L., Bean, T. P., Brard-Fudulea, S., . . . others. (2021). Current status and potential of genomic selection to improve selective breeding in the main aquaculture species of International Council for the Exploration of the Sea (ICES) member countries. *Aquaculture Reports*, 100700.

Bouma, S., Pawley, M. D., Hupman, K., & Gilman, A. (2018). Individual common dolphin identification via metric embedding learning. In *2018 international conference on image and vision computing New Zealand (IVCNZ)* (pp. 1--6). IEEE.

Boyd, C. E., McNevin, A. A., & Davis, R. P. (2022). The contribution of fisheries and aquaculture to the global protein supply. *Food security*, 805--827.

Bradski, G., Kaehler, A., & others. (2000). OpenCV. *Dr. Dobb's journal of software tools*.

Brett, J. R. (1964). The respiratory metabolism and swimming performance of young sockeye salmon. *Journal of the Fisheries Board of Canada*, 1183--1226.

Brito, L. F., Oliveira, H. R., McConn, B. R., Schinckel, A. P., Arrazola, A., Marchant-Forde, J. N., & Johnson, J. S. (2020). Large-scale phenotyping of livestock welfare in commercial production systems: a new frontier in animal breeding. *Frontiers in genetics*, 793.

Brosset, P., Fromentin, J.-M., Mé}nard, F., Pernet, F., Bourdeix, J.-H., Bigot, J.-L., . . . Saraux, C. (2015). Measurement and analysis of small pelagic fish condition: A suitable method for rapid evaluation in the field. *Journal of Experimental Marine Biology and Ecology*, 90--97.

Bruslund Haurum, J., Karpova, A., Pedersen, M., Hein Bengtson, S., & Moeslund, T. B. (2020). Re-identification of zebrafish using metric learning. *Proceedings*

of the IEEE/CVF winter conference on applications of computer vision workshops, 1--11.

Butler, D. (2021). asreml: Fits the linear mixed model. *R package version*.

Cai, L., Chen, J., Johnson, D., Tu, Z., & Huang, Y. (2020). Effect of body length on swimming capability and vertical slot fishway design. *Global Ecology and Conservation*, e00990.

Cano-Barbacil, C., Radinger, J., Argudo, M., Rubio-Gracia, F., Vila-Gispert, A., & García-Berthou, E. (2020). Key factors explaining critical swimming speed in freshwater fish: a review and statistical analysis for Iberian species. *Scientific reports*, 18947.

Carter, A. J., Feeney, W. E., Marshall, H. H., Cowlishaw, G., & Heinsohn, R. (2013). Animal personality: what are behavioural ecologists measuring? *Biological reviews*, 465--475.

Castillo, C., & Hernández, J. (2021). Ruminal fistulation and cannulation: a necessary procedure for the advancement of biotechnological research in ruminants. *Animals*, 1870.

Casto-Rebollo, C., Nuñez, P., Gol, S., Reixach, J., & Ibáñez-Escriche, N. (2025). Variability of daily feed intake as an indicator of resilience in Pietrain pigs. *animal*, 101415.

Castro, V., Grisdale-Helland, B., Jørgensen, S. M., Helgerud, J., Claireaux, G., Farrell, A. P., . . . Takle, H. (2013). Disease resistance is related to inherent swimming performance in Atlantic salmon. *BMC physiology*, 1.

Chopra, S., Hadsell, R., & LeCun, Y. (2005). Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (pp. 539--546). IEEE.

Cisar, P., Bekkozhayeva, D., Movchan, O., Saberioon, M., & Schraml, R. (2021). Computer vision based individual fish identification using skin dot pattern. *Scientific Reports*, 16904.

Claireaux, G., McKenzie, D. J., Genge, A. G., Chatelier, A., Aubin, J., & Farrell, A. P. (2005). Linking swimming performance, cardiac pumping ability and cardiac anatomy in rainbow trout. *Journal of Experimental Biology*, 1775--1784.

Colihueque, N., & Araneda, C. (2014). Appearance traits in fish farming: progress from classical genetics to genomics, providing insight into current and potential genetic improvement. *Frontiers in genetics*, 251.

Collin, H., & Fumagalli, L. (2011). Evidence for morphological and adaptive genetic divergence between lake and stream habitats in European minnows (Phoxinus phoxinus, Cyprinidae). *Molecular Ecology*, 4490--4502.

Crain, J., Reynolds, M., & Poland, J. (2017). Utilizing high-throughput phenotypic data for improved phenotypic selection of stress-adaptive traits in wheat. *Crop Science*, 648--659.

Dala-Corte, R. B., Moschetta, J. B., & Becker, F. G. (2016). Photo-identification as a technique for recognition of individual fish: a test with the freshwater armored catfish Rineloricaria aequalicuspis Reis \& Cardoso, 2001 (Siluriformes: Loricariidae). *Neotropical Ichthyology*, e150074.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (pp. 886--893). IEEE.

Darwin, C. (1859). *On the Origin of Species*.

de la Bruhèze, A., Lintsen, H., Rip, A., & Schot, J. (2000). Dieren en gewassen in een veranderende landbouw De georganiseerde rundveeverbetering. In *Techniek in Nederland in de twintigste eeuw. Deel 3. Landbouw, voeding* (pp. 137--138).

D'Eath, R., Conington, J., Lawrence, A., Olsson, I., Sand, P., & others. (2010). Breeding for behavioural change in farm animals: practical, economic and ethical considerations. *Animal welfare*, 17--27.

Dębicki, I. T., Mittell, E. A., Kristjánsson, B. K., Leblanc, C. A., Morrissey, M. B., & Terzić, K. (2021). Re-identification of individuals from images using spot constellations: a case study in Arctic charr (Salvelinus alpinus). *Royal Society Open Science*, 201768.

DeBruine, L. (2025). *faux: Simulation for Factorial Designs*. Zenodo. Retrieved from https://debruine.github.io/faux/

Dehaan, K. C. (2019). Critical Swimming Speed and Metabolic Activity as Predictors of Ecologically Relevant Behaviour in Juvenile Rainbow Trout (Onchorhynchus mykiss).

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248--255). IEEE.

Dennett, D. C. (2013). *Intuition pumps and other tools for thinking*. WW Norton & Company.

Dheeran, P., Varghese, B., & Salimkumar, A. V. (2022). Meta-analysis of PIT tagging effects on fish growth and mortality. *International Journal of Theoretical \& Applied Sciences*.

Dieterman, D. J., & Hoxmeier, R. J. (2009). Instream evaluation of passive integrated transponder retention in Brook Trout and Brown Trout: effects of season, anatomical placement, and fish length. *North American Journal of Fisheries Management*, 109--115.

Difford, G. F., Boison, S. A., Khaw, H. L., & Gjerde, B. (2020). Validating non-invasive growth measurements on individual Atlantic salmon in sea cages using diode frames. *Computers and Electronics in Agriculture*, 105411.

Difford, G. F., Hatlen, B., Heia, K., Bæverfjord, G., Eckel, B., Gannestad, K. H., . . . Gjerde, B. (2023). Digital phenotyping of individual feed intake in Atlantic salmon (Salmo salar) with the X-ray method and image analysis. *Frontiers in Animal Science*.

Ditria, E. M., Jinks, E. L., & Connolly, R. M. (2021). Automating the analysis of fish grazing behaviour from videos using image classification and optical flow. *Animal Behaviour*, 31--37.

Doornweerd, J. E., Kootstra, G., Veerkamp, R. F., Ellen, E. D., van der Eijk, J. A., van de Straat, T., & Bouwman, A. C. (2021). Across-species pose estimation in poultry based on images using deep learning. *Frontiers in Animal Science*, 791290.

Doyle, R. W., Singholka, S., & New, M. B. (1983). "Indirect selection" for genetic change: a quantitative analysis illustrated with Macrobrachium rosenbergii. *Aquaculture*, 237--247.

Drucker, E. G. (1996). The use of gait transition speed in comparative studies of fish locomotion. *American Zoologist*, 555--566.

Eknath, A. E., & Doyle, R. W. (1985). Indirect selection for growth and life-history traits in Indian carp aquaculture: 1. Effects of broodstock management. *Aquaculture*, 73--84.

Elalfy, I., Shin, H. S., Negrín-Báez, D., Navarro, A., Zamorano, M., Manchado, M., & Afonso, J. (2021). Genetic parameters for quality traits by non-invasive methods and their G x E interactions in ocean cages and estuaries on gilthead seabream (Sparus aurata). *Aquaculture*, 736462.

Ellen, E. D., Van Der Sluis, M., Siegford, J., Guzhva, O., Toscano, M. J., Bennewitz, J., . . . others. (2019). Review of sensor technologies in animal breeding: phenotyping behaviors of laying hens to select against feather pecking. *Animals*, 108.

Esposito, S., Carputo, D., Cardi, T., & Tripodi, P. (2019). Applications and trends of machine learning in genomics and phenomics for next-generation breeding. *Plants*, 34.

EU. (2025, March). *Aquaculture statistics*. Retrieved from Eurostat: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Aquaculture_statistics

Fadul-Pacheco, L., Liou, M., Reinemann, D. J., & Cabrera, V. E. (2021). A preliminary investigation of social network analysis applied to dairy cow behavior in automatic milking system environments. *Animals*, 1229.

Fan, S., Song, C., Feng, H., & Yu, Z. (2024). Take good care of your fish: fish re-identification with synchronized multi-view camera system. *Frontiers in Marine Science*, 1429459.

FAO. (2024, 06 07). *FAO Report: Global fisheries and aquaculture production reaches a new record high*. Retrieved from FAO.org: https://www.fao.org/newsroom/detail/fao-report-global-fisheries-and-aquaculture-production-reaches-a-new-record-high/en

FAO. (2025). *Aquatic biodiversity: underpinning aquatic food security*. Retrieved from FAO: aquatic genetic resourcces: https://www.fao.org/aquatic-genetic-resources/en/

FAO. (2025). *Fisheries and Aquaculture*. Retrieved from Sustainable aquaculture: https://www.fao.org/fishery/en/aquaculture

FAO. (2025). *Fishery and Aquaculture Statistics – Yearbook 2022*. Rome: FAO.

Farrell, A. (2008). Comparisons of swimming performance in rainbow trout using constant acceleration and critical swimming speed tests. *Journal of Fish Biology*, 693--710.

Farrell, A., Johansen, J., & Suarez, R. (1991). Effects of exercise-training on cardiac performance and muscle enzymes in rainbow trout (Oncorhynchus mykiss). *Fish Physiology and Biochemistry*, 303--312.

Farrell, A., Lee, C., Tierney, K., Hodaly, A., Clutterham, S., Healey, M., . . . Lotto, A. (2003). Field-based measurements of oxygen uptake and swimming performance with adult Pacific salmon using a mobile respirometer swim tunnel. *Journal of Fish Biology*, 64--84.

Fernandes, A. F., Dórea, J. R., Valente, B. D., Fitzgerald, R., Herring, W., & Rosa, G. J. (2020). Comparison of data analytics strategies in computer vision systems to predict pig body composition traits from 3D images. *Journal of Animal Science*, skaa250.

Fernandes, A. F., Turra, E. M., de Alvarenga, É. R., Passafaro, T. L., Lopes, F. B., Alves, G. F., . . . Rosa, G. J. (2020). Deep Learning image segmentation for

extraction of fish body measurements and prediction of body weight and carcass traits in Nile tilapia. *Computers and electronics in agriculture*, 105274.

Fodor, I., van der Sluis, M., Jacobs, M., de Klerk, B., Bouwman, A. C., & Ellen, E. D. (2023). Automated pose estimation reveals walking characteristics associated with lameness in broilers. *Poultry science*, 102787.

Foldvik, A., & Kvingedal, E. (2018). Long-term PIT tag retention rates in Atlantic salmon (Salmo salar). *Animal Biotelemetry*, 1--4.

Foldvik, A., Jakobsen, F., & Ulvan, E. M. (2020). Individual recognition of Atlantic Salmon using iris biometry. *Copeia*, 767-771.

Føre, M., & Alver, M. O. (2023). Precision Aquaculture. In *Encyclopedia of Smart Agriculture Technologies* (pp. 1--12). Springer.

Fragkoulis, S., Christou, M., Karo, R., Ritas, C., Tzokas, C., Batargias, C., & Koumoundouros, G. (2017). Scaling of body-shape quality in reared gilthead seabream Sparus aurata L. Consumer preference assessment, wild standard and variability in reared phenotype. *Aquaculture Research*, 2402--2410.

Fragkoulis, S., Kerasovitis, D., Batargias, C., & Koumoundouros, G. (2021). Body-shape trajectories and their genetic variance component in gilthead seabream (Sparus aurata L.). *Scientific reports*, 16964.

Freitas, M. V., Lemos, C. G., Ariede, R. B., Agudelo, J. F., Neto, R. R., Borges, C. H., . . . others. (2023). High-throughput phenotyping by deep learning to include body shape in the breeding program of pacu (Piaractus mesopotamicus). *Scientific reports*, 738847.

Fu, G., & Yuna, Y. (2022). Phenotyping and phenomics in aquaculture breeding. *Aquaculture and Fisheries*, 140--146.

Garcia, A. C. (2022). Utilizing Convolutional Neural Networks and Gradient-weighted Class Activation Mapping (Grad-CAM) for Dairy Cow Teat Image Classification.

García, R., Aguilar, J., Toro, M., Pinto, A., & Rodríguez, P. (2020). A systematic literature review on the use of machine learning in precision livestock farming. *Computers and Electronics in Agriculture*, 105826.

García-Celdrán, M., Ramis, G., Manchado, M., Estévez, A., Afonso, J., & Armero, E. (2015). Estimates of heritabilities and genetic correlations of carcass quality traits in a reared gilthead sea bream (Sparus aurata L.) population sourced from three broodstocks along the Spanish coasts. *Aquaculture*, 175--180.

Garrido-Izard, M., Correa, E. C., Requejo, J. M., Villarroel, M., & Diezma, B. (2022). Cleansing data from an electronic feeding station to improve estimation of feed efficiency. *Biosystems Engineering*, 361--369.

Gayo, P., Berbel, C., Korozi, E., Zerolo, R., & Manchado, M. (2023). Assessment of body shape variation using Elliptic Fourier descriptors and ellipse fitting estimators and their genetic estimates in the flatfish Senegalese sole. *Aquaculture*, 739948.

Gjedrem, T. (1997). Flesh quality improvement in fish through breeding. *Aquaculture international*, 197--206.

Gjedrem, T. (2010). The first family-based breeding program in aquaculture. *Reviews in Aquaculture*, 2--15.

Gjedrem, T., & Robinson, N. (2014). Advances by selective breeding for aquatic species: a review. *Agricultural Sciences*, 1152.

Gjedrem, T., & Rye, M. (2018). Selection response in fish and shellfish: A review. *Reviews in Aquaculture*, 168--179.

Gjedrem, T., Robinson, N., & Rye, M. (2012). The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. *Aquaculture*, 117--129.

Gladju, J., & Kanagaraj, A. (2021). Potential applications of data mining in aquaculture. In *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)* (pp. 1--5). IEEE.

Gómez-Vargas, N., Alonso-Fernández, A., Blanquero, R., & Antelo, L. T. (2023). Re-identification of fish individuals of undulate skate via deep learning within a few-shot context. *Ecological Informatics*, 102036.

Goñi, N., & Arrizabalaga, H. (2010). Seasonal and interannual variability of fat content of juvenile albacore (Thunnus alalunga) and bluefin (Thunnus thynnus) tunas during their feeding migration to the Bay of Biscay. *Progress in Oceanography*, 115--123.

Gonzales, R. C., & Wintz, P. (1987). *Digital image processing*. Addison-Wesley Longman Publishing Co., Inc.

Gore, M. A., Frey, P. H., Ormond, R. F., Allan, H., & Gilkes, G. (2016). Use of photo-identification and mark-recapture methodology to assess basking shark (Cetorhinus maximus) populations. *PloS one*, e0150160.

Gower, J. (1975). Generalized procrustes analysis. *Psychometrika*, 33--51.

Granados Chapatte, A. (2021). *What is animal breeding and why is it important for sustainable farming?* European Forum of Farm Animal Breeders (EFFAB). Retrieved from https://www.europarl.europa.eu/cmsdata/239025/Animal%20breeders%202%20-%20what%20is%20animal%20breeding-%20why%20is%20it%20important-%20EFFAB.pdf

Grassi, S., Casiraghi, E., & Alamprese, C. (2018). Fish fillet authentication by image analysis. *Journal of Food Engineering*, 16--23.

Groher, T., Heitkämper, K., & Umstätter, C. (2020). Digital technology adoption in livestock production with a special focus on ruminant farming. *Animal*, 2404--2413.

Guerrero, M. J., Bedoya, C. L., López, J. D., Daza, J. M., & Isaza, C. (2023). Acoustic animal identification using unsupervised learning. *Methods in Ecology and Evolution*, 1500--1514.

Gulzari, B. (2023). *All fish deserve a breeding program designing affordable genomic selection programs with automated image analysis*. Wageningen University, Animal breeding and genomics. Wageningen: Wageningen University.

Gulzari, B., Mencarelli, A., Roozeboom, C., Komen, H., & Bastiaansen, J. (2022). Prediction of production traits by using body features of gilthead seabream (Sparus aurata) obtained from digital images. In *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP) Technical and species orientated innovations in animal breeding, and contribution of genetics to solving societal challenges* (pp. 2412--2415). Wageningen Academic Publishers.

Hachim, M., Rouyer, T., Dutto, G., Kerzerho, V., Bernard, S., Bourjea, J., & McKenzie, D. J. (2021). Oxygen uptake, heart rate and activities of locomotor muscles during a critical swimming speed protocol in the gilthead sea bream Sparus aurata. *Journal of Fish Biology*, 886--890.

Haffray, P., Bugeon, J., Rivard, Q., Quittet, B., Puyo, S., Allamelou, J. M., . . . Dupont-Nivet, M. (2013). Genetic parameters of in-vivo prediction of carcass, head and fillet yields by internal ultrasound and 2D external imagery in large rainbow trout (Oncorhynchus mykiss). *Aquaculture*, 236--244.

Hassan, S., Mujtaba, G., Rajput, A., & Fatima, N. (2024). Multi-object tracking: a systematic literature review. *Multimedia Tools and Applications*, 43439--43492.

Haxton, T. (2021). Use of unique brook trout spot patterns over a short duration for a mark-recapture study. *Environmental Biology of Fishes*, 1391--1399.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *CoRR*.

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their applications*, 18--28.

Holmes, E. J., & Jeffres, C. A. (2021). Juvenile Chinook Salmon Weight Prediction Using Image-Based Morphometrics. *North American Journal of Fisheries Management*, 446--454.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 359--366.

Hossain, M. E., Kabir, M. A., Zheng, L., Swain, D. L., McGrath, S., & Medway, J. (2022). A systematic review of machine learning techniques for cattle identification: Datasets, methods and future directions. *Artificial Intelligence in Agriculture*, 138--155.

Hoxhallari, K., Purcell, W., & Neubauer, T. (2022). The potential of explainable artificial intelligence in precision livestock farming. University of Veterinary Medicine Vienna.

Hu, G., Do, D. N., Gray, J., & Miar, Y. (2020). Selection for favorable health traits: a potential approach to cope with diseases in farm animals. *Animals*, 1717.

Huntingford, F., Borçato, F., & Mesquita, F. (2013). Identifying individual common carp Cyprinus carpio using scale pattern. *Journal of fish biology*, 1453-1458.

Jacobs, N., Souvenir, R., & Pless, R. (2009). Passive vision: The global webcam imaging network. In *2009 IEEE Applied Imagery Pattern Recognition Workshop (AIPR 2009)* (pp. 1--8). IEEE.

Janssen, K., Berentsen, P., Besson, M., & Komen, H. (2017). Derivation of economic values for production traits in aquaculture species. *Genetics Selection Evolution*, 5.

Janssen, K., Chavanne, H., Berentsen, P., & Komen, H. (2017). Impact of selective breeding on European aquaculture. *Aquaculture*, 8--16.

Jiang, L., Lee, C., Teotia, D., & Ostadabbas, S. (2022). Animal pose estimation: A closer look at the state-of-the-art, existing gaps and opportunities. *Computer Vision and Image Understanding*, 103483.

Jiang, P.-T., Zhang, C.-B., Hou, Q., Cheng, M.-M., & Wei, Y. (2021). Layercam: Exploring hierarchical class activation maps for localization. *IEEE transactions on image processing*, 5875--5888.

Jones, R., Petrell, R., & Pauly, D. (1999). Using modified length--weight relationships to assess the condition of fish. *Aquacultural engineering*, 261--276.

Karl Pearson, F. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 559--572.

Kause, A., Paananen, T., Ritola, O., & Koskinen, H. (2007). Direct and indirect selection of visceral lipid weight, fillet weight, and fillet percentage in a rainbow trout breeding program. *Journal of animal science*, 3218--3227.

Keeping, J. A., Reeve-Arnold, K. E., Burns, N. M., Catarina, N., Bailey, D. M., & McNeill, D. C. (2020). Computer-aided photo-identification of a rare stingray, Megatrygon microps. *Journal of Fish Biology*, 815--819.

Kent, M. (1990). Hand-held instrument for fat/water determination in whole fish. *Food Control*, 47--53.

Khosla, C., & Saini, B. S. (2020). Enhancing performance of deep learning models with different data augmentation techniques: A survey. In *2020 international conference on intelligent engineering and management (ICIEM)* (pp. 79--85). IEEE.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., . . . others. (2023). Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 4015--4026).

Kittilsen, S., Schjolden, J., Beitnes-Johansen, I., Shaw, J. C., Pottinger, T. G., Sørensen, C., . . . Overli, O. (2009). Melanin-based skin spots reflect stress responsiveness in salmonid fish. *Hormones and behavior*, 292--298.

Klair, Y. S., Agrawal, K., & Kumar, A. (2024). Impact of generative ai in diagnosing diseases in agriculture. In *2024 2nd International Conference on Disruptive Technologies (ICDT)* (pp. 870--875). IEEE.

Knap, P. W., & Kause, A. (2018). Phenotyping for genetic improvement of feed efficiency in fish: lessons from pig breeding. *Frontiers in Genetics*, 184.

Knausgård, Ø. L. (2023). A contrastive learning approach for individual re-identification in a wild fish population. *arXiv*, 2301.00596.

Knudsen, C. M., Johnston, M. V., Schroder, S. L., Bosch, W. J., Fast, D. E., & Strom, C. R. (2009). Effects of passive integrated transponder tags on smolt-to-adult recruit survival, growth, and behavior of hatchery spring Chinook Salmon. *North American Journal of Fisheries Management*, 658--669.

Kora, H., Tsuchimoto, M., Miyata, K., Osato, S., Wang, Q., Apablaza, P. A., . . . Tachibana, K. (2000). Estimation of body fat content from standard body length and body weight on cultured red sea bream. *Fisheries Science*, 365--371.

Kuhl, F. P., & Giardina, C. R. (1982). Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing*, 236--258.

Kumar, S., Asiamah, P., Jolaoso, O., & Esiowu, U. (2025). Enhancing Image Classification with Augmentation: Data Augmentation Techniques for Improved Image Classification. *arXiv preprint arXiv:2502.18691*.

Lamping, C., Kootstra, G., & Derks, M. (2025). Transformer-based similarity learning for re-identification of chickens. *Smart Agricultural Technology*, 100945.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 436--444.

Leng, Q., Ye, M., & Tian, Q. (2019). A survey of open-world person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 1092--1108.

Li, D., & Du, L. (2022). Recent advances of deep learning algorithms for aquacultural machine vision systems with emphasis on fish. *Artificial Intelligence Review*, 4077--4116.

Li, N., Ren, Z., Li, D., & Zeng, L. (2020). Automated techniques for monitoring the behaviour and welfare of broilers and laying hens: towards the goal of precision livestock farming. *animal*, 617--625.

Li, T., Wang, C., Ji, W., Wang, Z., Shen, W., Feng, Y., & Zhou, M. (2023). Cutting-edge ammonia emissions monitoring technology for sustainable livestock and poultry breeding: A comprehensive review of the state of the art. *Journal of Cleaner Production*, 139387.

Liu, B., Tan, C., Li, S., He, J., & Wang, H. (2020). A data augmentation method based on generative adversarial networks for grape leaf disease identification. *Ieee Access*, 102188--102198.

Liu, W., Tan, J., Lan, G., Li, A., Li, D., Zhao, L., . . . Dong, N. (2024). Benchmarking Fish Dataset and Evaluation Metric in Keypoint Detection--Towards Precise Fish Morphological Assessment in Aquaculture Breeding. *arXiv preprint arXiv:2405.12476*.

Liu, Y., Li, W., Liu, X., Li, Z., & Yue, J. (2024). Deep learning in multiple animal tracking: A survey. *Computers and Electronics in Agriculture*, 109161.

Lokesh, G. H., Chandregowda, S. B., Vishwanath, J., Ravi, V., Ravi, P., & Al Mazroa, A. (2024). Intelligent Plant Leaf Disease Detection Using Generative Adversarial Networks: a Case-study of Cassava Leaves. *The Open Agriculture Journal*.

Ma, W., Sun, Y., Qi, X., Xue, X., Chang, K., Xu, Z., . . . Li, Q. (2024). Computer-vision-based sensing technologies for livestock body dimension measurement: A survey. *Sensors*, 1504.

Macaulay, G., Warren-Myers, F., Barrett, L. T., Oppedal, F., Føre, M., & Dempster, T. (2021). Tag use to monitor fish behaviour in aquaculture: a review of benefits, problems and solutions. *Reviews in Aquaculture*, 1565--1582.

Mahadevkar, S. V., Khemani, B., Patil, S., Kotecha, K., Vora, D. R., Abraham, A., & Gabralla, L. A. (2022). A review on machine learning styles in computer vision—techniques and future directions. *Ieee Access*, 107293--107329.

Mahapatra, K. D., Gjerde, B., Reddy, P., Sahoo, M., Jana, R., Saha, J., & Rye, M. (2001). Tagging: on the use of passive integrated transponder (PIT) tags for the identification of fish. *Aquaculture Research*, 47-50.

Mair, G. C., Halwart, M., Derun, Y., & Costa-Pierce, B. A. (2023). A decadal outlook for global aquaculture. *Journal of the World Aquaculture Society*, 196--205.

Malkauthekar, M. D. (2013). Analysis of euclidean distance and Manhattan Distance measure in face recognition. In *Third International Conference on Computational Intelligence and Information Technology (CIIT 2013)* (pp. 503-507).

Mallinger, K., Corpaci, L., Neubauer, T., Tikasz, I. E., Goldenits, G., & Banhazi, T. (2024). Breaking the barriers of technology adoption: Explainable AI for requirement analysis and technology design in smart farming. *Smart Agricultural Technology*, 100658.

Mandal, A., & Ghosh, A. R. (2024). Role of artificial intelligence (AI) in fish growth and health status monitoring: a review on sustainable aquaculture. *Aquaculture International*, 2791--2820.

Manohar, N., Kumar, Y. S., & Kumar, G. H. (2016). Supervised and unsupervised learning in animal classification. In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 156--161). IEEE.

Manolin. (2025). *Aquaculture Salmon Data Map*. Retrieved from https://globalsalmonmap.manolinaqua.com/

Marcil, J., Swain, D. P., & Hutchings, J. A. (2006). Genetic and environmental components of phenotypic variation in body shape among populations of Atlantic cod (Gadus morhua L.). *Biological Journal of the Linnean Society*, 351--365.

Marcon, M., Brossard, L., & Quiniou, N. (2015). Precision feeding based on individual daily body weight of group-housed pigs with an automatic feeder developed to allow for restricting feed allowance. *Precision livestock farming*, e601.

Martins, C. I., Galhardo, L., Noble, C., Damsgård, B., Spedicato, M. T., Zupa, W., . . . others. (2012). Behavioural indicators of welfare in farmed fish. *Fish Physiology and Biochemistry*, 17--41.

Martos-Sitcha, J. A., Sosa, J., Ramos-Valido, D., Bravo, F. J., Carmona-Duarte, C., Gomes, H. L., . . . Ferrer, M. Á. (2019). Ultra-low power sensor devices for

monitoring physical activity and respiratory frequency in farmed fish. *Frontiers in physiology*, 667.

Mathisen, B. M., Bach, K., Meidell, E., Måløy, H., & Sjøblom, E. S. (2020). FishNet: A unified embedding for Salmon recognition. *arXiv preprint arXiv:2010.10475*.

McInnes, M. G., Burns, N. M., Hopkins, C. R., Henderson, G. P., McNeill, D. C., & Bailey, D. M. (2020). A new model study species: high accuracy of discrimination between individual freckled hawkfish (Paracirrhites forsteri) using natural markings. *Journal of Fish Biology*, 831--834.

McKenzie, D. (2011). Swimming and other activities. *Encycl. Fish Physiol*, 1636--1644.

McMaster, A. C. (2023). *Genetic and environmental influences on morphological traits in salmonids*. University of British Columbia.

McPhee, M. J., Walmsley, B. J., Skinner, B., Littler, B., Siddell, J., Cafe, L. M., . . . Alempijevic, A. (2017). Live animal assessments of rump fat and muscle score in Angus cows and steers using 3-dimensional imaging. *Journal of animal science*, 1847--1857.

McPherson, L. R., Slotte, A., Kvamme, C., Meier, S., & Marshall, C. T. (2011). Inconsistencies in measurement of fish condition: a comparison of four indices of fat reserves for Atlantic herring (Clupea harengus). *ICES Journal of Marine Science*, 52--60.

McVey, C., Hsieh, F., Manriquez, D., Pinedo, P., & Horback, K. (2023). Invited Review: Applications of unsupervised machine learning in livestock behavior: Case studies in recovering unanticipated behavioral patterns from precision livestock farming data streams. *Applied Animal Science*, 99--116.

Mehar, M., Mekkawy, W., McDougall, C., & Benzie, J. A. (2020). Fish trait preferences: a review of existing knowledge and implications for breeding programmes. *Reviews in Aquaculture*, 1273--1296.

Meng, H., Zhang, L., Yang, F., Hai, L., Wei, Y., Zhu, L., & Zhang, J. (2025). Livestock Biometrics Identification Using Computer Vision Approaches: A Review. *Agriculture*, 102.

Mengistu, S. B., Mulder, H. A., Bastiaansen, J. W., Benzie, J. A., Khaw, H. L., Trinh, T. Q., & Komen, H. (2022). Fluctuations in growth are heritable and a potential indicator of resilience in Nile tilapia (Oreochromis niloticus). *Aquaculture*, 738481.

Mengistu, S. B., Palstra, A. P., Mulder, H. A., Benzie, J. A., Trinh, T. Q., Roozeboom, C., & Komen, H. (2021). Heritable variation in swimming performance in

Nile tilapia (Oreochromis niloticus) and negative genetic correlations with growth and harvest weight. *Scientific Reports*, 11018.

Merton, T. (1969). *The way of Chuang Tzu*. New Directions.

Merz, J. E., Skvorc, P., Sogard, S. M., Watry, C., Blankenship, S. M., & Van Nieuwenhuyse, E. E. (2012). Onset of melanophore patterns in the head region of Chinook Salmon: a natural marker for the reidentification of individual fish. *North American journal of fisheries management*, 806--816.

Metcalfe, N. a. (2016). Does individual variation in metabolic phenotype predict fish behaviour and performance? *Journal of fish biology*, 298--321.

Metian, M., Troell, M., Christensen, V., Steenbeek, J., & Pouil, S. (2020). Mapping diversity of species in global aquaculture. *Reviews in Aquaculture*, 1090--1100.

Miller, G. A., Hyslop, J. J., Barclay, D., Edwards, A., Thomson, W., & Duthie, C.-A. (2019). Using 3D imaging and machine learning to predict liveweight and carcass characteristics of live finishing beef cattle. *Frontiers in Sustainable Food Systems*, 30.

Miller, R., Whelan, H., Chrubasik, M., Whittaker, D., Duncan, P., & Gregório, J. (2024). A Framework for Current and New Data Quality Dimensions: An Overview. *Data*, 151.

Mitteroecker, P., & Gunz, P. (2009). Advances in geometric morphometrics. *Evolutionary biology*, 235--247.

Miyazono, T., & Saitoh, T. (2017). Fish species recognition based on CNN using annotated image. In *IT Convergence and Security 2017: Volume 1* (pp. 156--163).

Mørkøre, T., Vallet, J.-L., Cardinal, M., Gomez-Guillen, M., Montero, P., Torrissen, O., . . . Thomassen, M. (2001). Fat content and fillet shape of Atlantic salmon: relevance for processing yield and quality of raw and smoked products. *Journal of food science*, 1348--1354.

Morrone, S., Dimauro, C., Gambella, F., & Cappai, M. G. (2022). Industry 4.0 and precision livestock farming (PLF): an up to date overview across animal productions. *Sensors*, 4319.

Moya, A., Torrella, J., Fernández-Borràs, J., Rizo-Roca, D., Millán-Cubillo, A., Vélez, E., . . . Blasco, J. (2019). Sustained swimming enhances white muscle capillarisation and growth by hyperplasia in gilthead sea bream (Sparus aurata) fingerlings. *Aquaculture*, 397--403.

Nash, R. D., Valencia, A. H., & Geffen, A. J. (2006). The origin of Fulton's condition factor—setting the record straight. *Fisheries*, 236--238.

Nauta, M., Trienes, J., Pathak, S., Nguyen, E., Peters, M., Schmitt, Y., . . . Seifert, C. (2023). From anecdotal evidence to quantitative evaluation methods: A systematic review on evaluating explainable ai. *ACM Computing Surveys*, 1--42.

Navarro, A., Lee-Montero, I., Santana, D., Henríquez, P., Ferrer, M. A., Morales, A., . . . others. (2016). IMAFISH_ML: A fully-automated image analysis software for assessing fish morphometric traits on gilthead seabream (Sparus aurata L.), meagre (Argyrosomus regius) and red porgy (Pagrus pagrus). *Computers and Electronics in Agriculture*, 66--73.

Ndiwa, T. C., Nyingi, D. W., & Claude, J. a.-F. (2016). Morphological variations of wild populations of Nile tilapia (Oreochromis niloticus) living in extreme environmental conditions in the Kenyan Rift-Valley. *Environmental Biology of Fishes*, 473--485.

Neethirajan, S., & Kemp, B. (2021). Digital phenotyping in livestock farming. *Animals*, 2009.

Nguyen, N. H., S. A., Houston, R. D., & Moghadam, H. (2022). Applications of modern genetics and genomic technologies to enhance aquaculture breeding. *Frontiers in Genetics*, 898857.

Nietzsche, F. (1874). *On the Use and Abuse of History for Life*. (I. C. Johnston, Trans.) Untimely Meditations. Retrieved from https://la.utexas.edu/users/hcleaver/330T/350kPEENietzscheAbuseTableAll.pdf

Norton, T., Chen, C., Larsen, M. L., & Berckmans, D. (2019). Precision livestock farming: Building 'digital representations' to bring the animals closer to the farmer. *Animal*, 3009--3017.

Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 51--59.

Oldenbroek, k., & Calus, M. (2024). *Textbook Animal Breeding and Genetics for BSc students* (Second ed.). Wageningen University & Research, Animal Breeding and Genomics. Retrieved from Groen Kennisnet: https://wiki-groenkennisnet.atlassian.net/wiki/spaces/TAB/overview

Olejnik, K., Popiela, E., & Opaliński, S. (2022). Emerging precision management methods in poultry sector. *Agriculture*, 718.

Olsen, Ø. L., Sørdalen, T. K., Goodwin, M., Malde, K., Knausgård, K. M., & Halvorsen, K. T. (2023). A contrastive learning approach for individual re-identification in a wild fish population. *arXiv preprint arXiv:2301.00596*.

Osada, M., Yasugi, M., Yamamoto, H., Ito, A., & Fukamachi, S. (2024). Individual Identification of Medaka, a Small Freshwater Fish, from the Dorsal Side Using Artificial Intelligence. *Hydrobiology*, 119--133.

Óskarsson, G. (2008). Variation in body condition, fat content and growth rate of Icelandic summer-spawning herring Clupea harengus L. *Journal of Fish biology*, 2655--2676.

Ozella, L., Brotto Rebuli, K., Forte, C., & Giacobini, M. (2023). A literature review of modeling approaches applied to data collected in automatic milking systems. *Animals*, 1916.

Palaiokostas, C. (2021). Predicting for disease resistance in aquaculture species using machine learning models. *Aquaculture Reports*, 100660.

Palstra, A. P., Kals, J., Böhm, T., Bastiaansen, J. W., & Komen, H. (2020). Swimming performance and oxygen consumption as non-lethal indicators of production traits in Atlantic salmon and gilthead seabream. *Frontiers in Physiology*, 759.

Paula, A. M. (2025). *Computer vision tools to detect the precursors of tail biting in pigs: first steps towards methodological framework.* MSc thesis.

Pedersen, M., & Mohammed, A. (2021). Photo identification of individual Salmo trutta based on deep learning. *Applied Sciences*, 9039.

Pedersen, M., Nyegaard, M., & Moeslund, T. B. (2023). Finding nemo's giant cousin: Keypoint matching for robust re-identification of giant sunfish. *Journal of Marine Science and Engineering*, 889.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . others. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 2825-2830.

Pérez-Enciso, M., & Steibel, J. P. (2021). Phenomes: the current frontier in animal breeding. *Genetics Selection Evolution*, 22.

Pérez-Enciso, M., Zingaretti, L., & de Los Campos, G. (2025). Generative AI for predictive breeding: hopes and caveats. *Theoretical and Applied Genetics*, 147.

Pirsig, R. M. (2014). *Zen and the art of motorcycle maintenance.* Vintage.

Plaut, I. (2001). Critical swimming speed: its ecological relevance. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 41--50.

Poli, B., Parisi, G., Scappini, F., & Zampacavallo, G. (2005). Fish welfare and quality as affected by pre-slaughter and slaughter management. *Aquaculture international*, 29--49.

Poppe, M. (2022). *PhD thesis: Genetic improvement of resilience in dairy cattle using longitudinal data*. Wageningen University and Research.

Prchal, M., Kocour, M., Vandeputte, M., Kause, A., Vergnet, A., Zhao, J., . . . others. (2020). Morphological predictors of slaughter yields using 3D digitizer and their use in a common carp breeding program. *Aquaculture*, 734993.

R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.

Ramanathan, N., & Chellappa, R. (2008). Modeling shape and textural variations in aging faces. In *2008 8th IEEE International Conference on Automatic Face \& Gesture Recognition* (pp. 1--8). IEEE.

Rasal, A., Patnaik, M., Murmu, K., Sundaray, J., & Das Mahapatra, K. (2021). Importance of PIT Tagging in Aquaculture Selective Breeding Programs. *Vol. 52 No. 4, December*, 59--63.

Rauw, W. a.-S. (1998). Undesirable side effects of selection for high production efficiency in farm animals: a review. *Livestock production science*, 15--33.

Rebuffi, S.-A., Gowal, S., Calian, D. A., Stimberg, F., Wiles, O., & Mann, T. A. (2021). Data augmentation can improve robustness. *Advances in neural information processing systems*, 29935--29948.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779--788).

Ren, S., He, K., Girshick, R. B., & Sun, J. (2015). Faster {R-CNN:} Towards Real-Time Object Detection with Region Proposal. *CoRR*.

Rezende, C. E., Perazza, C. A., Freitas, R. T., Hallerman, E., & Hilsdorf, A. W. (2023). Use of ultrasonographic imaging for non-invasive carcass yield prediction in Nile tilapia (Oreochromis niloticus). *Aquaculture International*, 2764--2778.

Ribeiro, L. R., Sans, E. C., Santos, R. M., Taconelli, C. A., de Farias, R., & Molento, C. F. (2024). Will the white blood cells tell? A potential novel tool to assess broiler chicken welfare. *Frontiers in Veterinary Science*, 1384802.

Rodríguez, R., Fountoulaki, E., Grigorakis, K., Alexis, M., & Flos, R. (2010). Season and size effects: changes in the quality of gilthead sea bream (Sparus aurata L.). *Mediterranean Marine Science*, 117--132.

Rome, L. C., Choi, I.-H., Lutz, G., & Sosnicki, A. (1992). The influence of temperature on muscle function in the fast swimming scup: I. Shortening velocity and muscle recruitment during swimming. *Journal of Experimental Biology*, 259--279.

Rossi, F. B., Rossi, N., Orso, G., Barberis, L., Marin, R. H., & Kembro, J. M. (2025). Monitoring poultry social dynamics using colored tags: Avian visual perception, behavioral effects, and artificial intelligence precision. *Poultry Science*, 104464.

Rubio-Gracia, F., García-Berthou, E., Guasch, H., Zamora, L., & Vila-Gispert, A. (2020). Size-related effects and the influence of metabolic traits and morphology on swimming performance in fish. *Current Zoology*, 493--503.

Ruchay, A., Kolpakov, V., Guo, H., & Pezzuolo, A. (2024). On-barn cattle facial recognition using deep transfer learning and data augmentation. *Computers and electronics in agriculture*, 109306.

Russello, H., van der Tol, R., & Kootstra, G. (2022). T-LEAP: Occlusion-robust pose estimation of walking cows using temporal information. *Computers and Electronics in Agriculture*, 106559.

Russello, H., van der Tol, R., Holzhauer, M., van Henten, E. J., & Kootstra, G. (2024). Video-based automatic lameness detection of dairy cows using pose estimation and multiple locomotion traits. *Computers and Electronics in Agriculture*, 109040.

Saberioon, M., Gholizadeh, A., Cisar, P., Pautsina, A., & Urban, J. (2017). Application of machine vision systems in aquaculture with emphasis on fish: state-of-the-art and key issues. *Reviews in Aquaculture*, 369--387.

Saeed, R., Feng, H., Wang, X., Zhang, X., & Fu, Z. (2022). Fish quality evaluation by sensor and machine learning: A mechanistic review. *Food Control*, 108902.

Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowledge-based systems*, 110273.

Sae-Lim, P., Kause, A., Mulder, H., & Olesen, I. (2017). Breeding and genetics symposium: Climate change and selective breeding in aquaculture. *Journal of animal science*, 1801--1812.

Salerno, M., Berlino, M., Mangano, M. C., & Sarà, G. (2021). Microplastics and the functional traits of fishes: A global meta-analysis. *Global Change Biology*, 2645--2655.

Saranya, A., & Subhashini, R. (2023). A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. *Decision analytics journal*, 100230.

Sauphar, C., Stolz, C., Tuene, S. A., Gansel, L. C., & Aas, G. K. (2024). Atlantic salmon (Salmo salar) distribution and vertical size-stratification in a commercial sea cage: A case study. *Aquaculture*, 740356.

Scanes, C. G. (2018). Animal agriculture: Livestock, poultry, and fish aquaculture. In *Animals and human society* (pp. 133--179). Elsevier.

Schneider, S. (2020). *PhD thesis: Deep Learning Based Computer Vision for Animal Re-Identification*. Guelph: University of Guelph.

Schraml, R., Hofbauer, H., Jalilian, E., Bekkozhayeva, D., Saberioon, M., Cisar, P., & Uhl, A. (2020). Towards fish individuality-based aquaculture. *IEEE Transactions on Industrial Informatics*, 4356--4366.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815--823).

Seibel, H., Weirup, L., & Schulz, C. (2020). Fish welfare--between regulations, scientific facts and human perception. *Food ethics*, 4.

Sekachev, B., Manovich, N., Zhiltsov, M., Zhavoronkov, A., Kalinin, D., Hoff, B., . . . others. (2020). opencv/cvat: v1. 1.0. *Zenodo*.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618--626).

Shahinfar, S., Khansefid, M., Haile-Mariam, M., & Pryce, J. (2021). Machine learning approaches for the prediction of lameness in dairy cows. *Animal*, 100391.

Sheng, K., Tuyttens, F. A., & von Keyserlingk, M. A. (2025). The erasure of intensive livestock farming in text-to-image generative AI. *arXiv preprint arXiv:2502.19771*.

Shi, W., Dai, B., Shen, W., Sun, Y., Zhao, K., & Zhang, Y. (2023). Automatic estimation of dairy cow body condition score based on attention-guided 3D point cloud feature extraction. *Computers and Electronics in Agriculture*, 107666.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, 1--48.

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, 755--759.

Silva, S. S., Alexandre, C. M., Quintella, B. R., & de Almeida, P. R. (2021). Seasonal environmental variability drives the swimming performance of a resident Iberian fish. *Ecology of Freshwater Fish*, 366--374.

Smith, B. A., Costa, A. P., Kristjánsson, B. K., & Parsons, K. J. (2024). Experimental evidence for adaptive divergence in response to a warmed habitat reveals

roles for morphology, allometry and parasite resistance. *Ecology and Evolution*, e10907.

Smith, D. T., Potgieter, A. B., & Chapman, S. C. (2021). Scaling up high-throughput phenotyping for abiotic stress selection in the field. *Theoretical and Applied Genetics*, 1845--1866.

Sopinka, N. M., Donaldson, M. R., O'Connor, C. M., Suski, C. D., & Cooke, S. J. (2016). Stress indicators in fish. *Fish physiology*, 405--462.

Spiliopoulos, O., Brown, C., Hilder, P., Tilbrook, A., & Descovich, K. (2025). he Path to Resilience: Improving Welfare in Aquaculture Through Physical Exercise and Stressor Predictability Training. *Reviews in Aquaculture*, e70051.

Srivastava, R. P. (2023). A new measure of similarity in textual analysis: Vector similarity metric versus cosine similarity metric. *Journal of Emerging Technologies in Accounting*, 77-90.

Steck, H., Ekanadham, C., & Kallus, N. (2024). Is cosine-similarity of embeddings really about similarity? In *Companion Proceedings of the ACM Web Conference 2024* (pp. 887--890).

Stien, L. H., Nilsson, J., Bui, S., Fosseidengen, J. E., Kristiansen, T. S., Øverli, Ø., & Folkedal, O. (2017). Consistent melanophore spot patterns allow long-term individual recognition of Atlantic salmon Salmo salar. *Journal of Fish Biology*, 1699--1712.

Subedi, S., Bist, R., Yang, X., & Chai, L. (2023). Tracking pecking behaviors and damages of cage-free laying hens with machine vision technologies. *Computers and Electronics in Agriculture*, 107545.

Sun, P., Yu, J., Sun, R., Liu, G., Xi, X., & Tian, Y. (2022). The Effects of Selective Harvest on Exploited Population and Economic Benefits. *Frontiers in Marine Science*, 847840.

Svozil, D. P., Baumgartner, L. J., Fulton, C. J., Kopf, R. K., & Watts, R. J. (2020). Morphological predictors of swimming speed performance in river and reservoir populations of Australian smelt Retropinna semoni. *Journal of Fish Biology*, 1632--1643.

Taghavi, M., Russello, H., Ouweltjes, W., Kamphuis, C., & Adriaens, I. (2024). Cow key point detection in indoor housing conditions with a deep learning model. *Journal of Dairy Science*, 2374--2389.

Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105--6114). PMLR.

Tan, M., Pang, R., & Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10781--10790).

Thorstad, E. B., Rikardsen, A. H., Alp, A., & Økland, F. (2013). The use of electronic tags in fish research--an overview of fish telemetry methods. *Turkish Journal of Fisheries and Aquatic Sciences*, 881--896.

Tierney, K., & Farrell, A. (2004). The relationships between fish health, metabolic rate, swimming performance and recovery in return-run sockeye salmon, Oncorhynchus nerka (Walbaum). *Journal of fish diseases*, 663--671.

Tzanidakis, C., Tzamaloukas, O., Simitzis, P., & Panagakis, P. (2023). Precision livestock farming applications (PLF) for grazing animals. *Agriculture*, 288.

Utrecht, G., HDSR, & van Heukelum, M. (2025). *The fish Doorbell*. Retrieved from https://visdeurbel.nl/: https://visdeurbel.nl/

Valizadeh, G., & Babapour Mofrad, F. (2022). A comprehensive survey on two and three-dimensional Fourier shape descriptors: biomedical applications. *Archives of Computational Methods in Engineering*, 4643--4681.

Vallecillos, A., María-Dolores, E., Villa, J., Rueda, F. M., Carrillo, J., Ramis, G., . . . Armero, E. (2021). Phenotypic and genetic components for growth, morphology, and flesh-quality traits of meagre (Argyrosomus regius) reared in tank and sea cage. *Animals*, 3285.

van den Berg, D. (2022). *Subjectivity in gilthead sea bream (Sparus aurata) trait scoring*. Wageningen: Animal Breeding and Genomics, Wageningen University.

Van Rossum, G., Drake, F. L., & others. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Van Tienhoven, A., Den Hartog, J., Reijns, R., & Peddemors, V. (2007). A computer-aided program for pattern-matching of natural marks on the spotted raggedtooth shark Carcharias taurus. *Journal of Applied ecology*, 273--280.

Vandeputte, M., Porte, J., Aupérin, B., Dupont-Nivet, M., Vergnet, A., Valotaire, C., . . . Chatain, B. (2016). Quantitative genetic variation for post-stress cortisol and swimming performance in growth-selected and control populations of European sea bass (Dicentrarchus labrax). *Aquaculture*, 1--7.

Vandeputte, M., Puledda, A., Tyran, A. S., Bestin, A., Coulombet, C., Bajek, A., . . . others. (2017). Investigation of morphological predictors of fillet and carcass yield in European sea bass (Dicentrarchus labrax) for application in selective breeding. *Aquaculture*, 40--49.

Vaughan, D. (2020). *Analytical skills for AI and data science: building skills for an AI-driven enterprise*. O'Reilly Media.

Ventura, R. V., e Silva, F. F., Yáñez, J. M., & Brito, L. F. (2020). Opportunities and challenges of phenomics applied to livestock and aquaculture breeding in South America. *Animal frontiers: the review magazine of animal agriculture*, 45.

Vezzani, R., Baltieri, D., & Cucchiara, R. (2013). People reidentification in surveillance and forensics: A survey. *ACM Computing Surveys (CSUR)*, 1--37.

Videler, J. J. (1993). *Fish swimming*. Springer Science \& Business Media.

Vinci, C. (2025). *Transforming animal farming through artifical intelligence*. European Parliamentary Research Service.

Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., & Garnier, E. (2007). Let the concept of trait be functional! *Oikos*, 882--892.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., . . . others. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, 261--272.

Vollset, K. W., Lennox, R. J., Thorstad, E. B., Auer, S., Bär, K., Larsen, M. H., . . . Dohoo, I. (2020). Systematic review and meta-analysis of PIT tagging effects on mortality and growth of juvenile salmonids. *Reviews in Fish Biology and Fisheries*, 553--568.

Wang, C., Li, Z., Wang, T., Xu, X., Zhang, X., & Li, D. (2021). Intelligent fish farm—the future of aquaculture. *Aquaculture International*, 1--31.

Wang, F., & Liu, H. (2021). Understanding the behaviour of contrastive loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2495--2504).

Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., . . . Hu, X. (2020). Score-CAM: Score-weighted visual explanations for convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 24--25).

Wang, Z. (2025). *Towards genetic improvement of social behaviours in large groups of livestock: a case study with turkeys*. Animal breeding and genomics. Wageningen: Wageningen University.

Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 3021.

Weimer, S., Mauromoustakos, A., Karcher, D., & Erasmus, M. (2020). Differences in performance, body conformation, and welfare of conventional and slow-growing broiler chickens raised at 2 stocking densities. *Poultry Science*, 4398--4407.

Weirup, L., Schulz, C., & Seibel, H. (2022). Fish welfare evaluation index (fWEI) based on external morphological damage for rainbow trout (Oncorhynchus mykiss) in flow through systems. *Aquaculture*.

Wilder, S. M., Raubenheimer, D., & Simpson, S. J. (2016). Moving beyond body condition indices as an estimate of fitness in ecological and evolutionary studies. *Functional Ecology*, 108--115.

Wimberger, P. H. (1992). Plasticity of fish body shape. The effects of diet, development, family and age in two species of Geophagus (Pisces: Cichlidae). *Biological Journal of the linnean society*, 197--218.

Wu, H., Aoki, A., Arimoto, T., Nakano, T., Ohnuki, H., Murata, M., . . . Endo, H. (2015). Fish stress become visible: A new attempt to use biosensor for real-time monitoring fish stress. *Biosensors and Bioelectronics*, 503--510.

Xia, X., Xu, C., & Nan, B. (2017). Inception-v3 for flower classification. In *2017 2nd international conference on image, vision and computing (ICIVC)* (pp. 783--787). IEEE.

Xue, Y. (2019). *Quantitative and qualitative traits prediction from Gilthead Seabream (Sparus aurata) image analysis using machine learning and deep learning algorithms*. Wageningen, the Netherlands: Wageningen University.

Xue, Y., Bastiaansen, J. W., Khan, H. A., & Komen, H. (2023). An analytical framework to predict slaughter traits from images in fish. *Aquaculture*, 739175.

Xue, Y., Palstra, A. P., Blonk, R., Mussgnug, R., Khan, H. A., Komen, H., & Bastiaansen, J. W. (2025). Genetic analysis of swimming performance in rainbow trout (Oncorhynchus mykiss) using image traits derived from deep learning. *Aquaculture*, 742607.

Yan, G.-J., He, X.-K., Cao, Z.-D., & Fu, S.-J. (2013). An interspecific comparison between morphology and swimming performance in cyprinids. *Journal of Evolutionary biology*, 1802--1815.

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., . . . Yan, J. (2020). Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Molecular plant*, 187--214.

Yang, X., Zhang, S., Liu, J., Gao, Q., Dong, S., & Zhou, C. (2021). Deep learning for smart fish farming: applications, opportunities and challenges. *Reviews in Aquaculture*, 66--90.

Yang, Y., Xue, B., Jesson, L., Wylie, M., Zhang, M., & Wellenreuther, M. (2021). Deep convolutional neural networks for fish weight prediction from images. In

2021 36th International Conference on Image and Vision Computing New Zealand (IVCNZ) (pp. 1--6). IEEE.

Ye, M., Shen, J., Lin, G., iang, T., Shao, L., & Hoi, S. C. (2022). Deep Learning for Person Re-Identification: A Survey and Outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2872-2893.

Yoon, J., Drumright, L. N., & Van Der Schaar, M. (2020). Anonymization through data synthesis using generative adversarial networks (ADS-GAN). *IEEE journal of biomedical and health informatics*, 2378--2388.

Yu, C., Hu, Z., Han, B., Wang, P., Zhao, Y., & Wu, H. (2021). Intelligent measurement of morphological characteristics of fish using improved U-Net. *Electronics*, 1426.

Yue, K., & Shen, Y. (2022). An overview of disruptive technologies for aquaculture. *Aquaculture and Fisheries*, 111--120.

Zahra, A., Perwaiz, N., Shahzad, M., & Fraz, M. M. (2023). Person re-identification: A retrospective on domain specific open challenges and future trends. *Pattern Recognition*.

Zhang, X., & Dahu, W. (2019). Application of artificial intelligence algorithms in image processing. *Journal of Visual Communication and Image Representation*, 42--49.

Zhao, L., Pedersen, M., Hardeberg, J. Y., & Dervo, B. (2019). Image-based recognition of individual trouts in the wild. *2019 8th European Workshop on Visual Information Processing (EUVIP)*, 82--87.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2921--2929).

Zhou, Y., Kusmec, A., Mirnezami, S. V., Attigala, L., Srinivasan, S., Jubery, T. Z., . . . Schnable, P. S. (2021). Identification and utilization of genetic determinants of trait measurement errors in image-based, high-throughput phenotyping. *The Plant Cell*, 2562--2582.

Zhou, Z., Hitt, N. P., Letcher, B. H., Shi, W., & Li, S. (2022). Pigmentation-based visual learning for salvelinus fontinalis individual re-identification. *2022 IEEE International Conference on Big Data (Big Data)*, 6850--6852.

Summary

In this thesis, I explored the integration of artificial intelligence (AI) in image-based automated phenotyping for selective breeding in aquaculture. Automated phenotyping is a rapidly evolving, interdisciplinary field focused on advancing phenotype collection, to support breeding and management decisions. High-throughput, non-invasive imaging technologies create opportunities to expand the range of measurable traits. The current scope of automated phenotyping primarily targets production-related traits, leaving a gap in traits associated with metabolism, health and physiological well-being. These traits are often complex and challenging to measure, but advancements in AI-based prediction models combined with imaging technologies offer a promising solution.

I used image data collected on selection candidates from multiple fish species in real-world breeding facilities. The overarching objective was to broaden the scope of phenotypes predicted from images using AI and to deliver diverse traits that collectively contribute to the breeding goals for these species.
In **Chapter 2**, I introduced a structural framework to find the best strategy for using images to predict invasive traits, with the example of gilthead seabream (*Sparus aurata*). The framework can capture both simple, linear and complex, high-dimensional relationship between the predictor and the target trait and generate the best prediction model by prioritizing prediction accuracy. A detailed interpretation of the model was provided by extracting and visualizing predictive imaginal features. Besides the three cases I studied – fat percentage, visceral weight and visceral percentage – the framework could be applied to many harvest or post-slaughter traits. This chapter highlighted the benefits of incorporating imaging sessions in routine phenotyping.

Indicator traits are commonly used to predict breeding goal traits with a complex biological and physiological background, such as those related to metabolism and health. In **Chapter 3**, I introduced a novel workflow to dissect the genetic basis of swimming performance in rainbow trout (*Oncorhynchus mykiss*) using an image-based deep learning model. I identified accessible and heritable morphological indicators that explained more variance in critical swimming speed than whole-body volume. Genetically heavier fish showed poorer swimming performance. Among fish of the same weight, relatively larger and broader epaxial muscles, larger heads, and smaller caudal fins lead to worse swimming performance. This study demonstrated how the combination of images and AI can generate data-driven hypotheses by explaining complex trait performance based on the fish morphology, which was subsequently supported by physiological interpretation of the Explainable AI results.

Repeated individual measurements are required for longitudinal traits such as growth rate, making individual re-identification essential for such traits. In **Chapter 4**, I investigated image-based re-identification of Atlantic salmon (*Salmo salar*) under realistic aquacultural conditions and concluded that this is not yet feasible. The results highlighted two major challenges: variation in the phenotyping environment, particularly lighting, and limited phenotypic stability. The discussion

further emphasized the need for standardized imaging protocols, and the discovery of biologically unique and stable phenotypic features in individual salmon.

Building on the conclusions of Chapter 4, I reviewed, in **Chapter 5**, the current state of vision-based re-identification in aquaculture as a potential alternative to PIT tags. Based on this review of 23 studies, multiple limitations and challenges in current implementations were identified regarding data acquisition, data analysis, and practical implementation. Recommendations to address these gaps were presented in a new aquaculture-centered flowchart, emphasizing high-tech data acquisition, multi-dimensional data quality frameworks, and benchmarking methods. Importantly, in this chapter, I called for a realistic view that the development of vision-based individual re-identification is currently too immature to be a replacement for the PIT tag. In response, recommendations were made focusing dataset standardization, methodological refinement, and the alignment of model development with practical aquacultural systems.

In **Chapter 6** I proposed a novel method for quantifying body shape using image-derived traits in gilthead seabream (*Sparus aurata*). Two quantitative traits with moderate heritability, "Distance to the Best" (DtB) and "Distance to the Worst" (DtW), were derived from expert's scores given to images to measure how much each fish deviated from an ideal and less desirable shape, respectively. This approach is automated, objective, and visually intuitive, and has the flexibility to update target shapes. Alternative shapes could be the natural shape or based on the preferences for ideal shape obtained from farmers and consumers.

Finally, in **Chapter 7**, I synthesized findings across all chapters, putting them within the broader context of both aquaculture and livestock breeding. The discussion began by addressing fundamental challenges in image-based automated phenotyping, emphasizing that successful AI integration requires an understanding of the entire breeding program, not just phenotypic data collection and analysis. The chapter concluded with an outlook on emerging AI trends, highlighting opportunities to improve animal health and welfare through advanced phenotyping and other decision-support AI tools, across aquaculture, livestock, and beyond.

Appendices

About the author

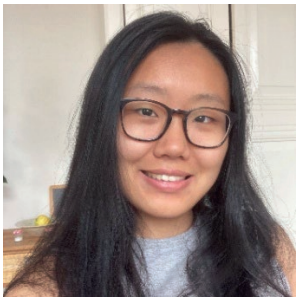
Publications and presentations

Training and supervision plan

Acknowledgement

Authorship statement

About the Author



Yuanxu (Yuuko) Xue was born on 21st January 1995 in Baoding, China. She completed her bachelor's degree in Animal Sciences at the College of Wildlife Resources, Northeast Forestry University in Harbin, China, from 2012 to 2016. After two internships, one as a lab assistant in animal reproduction and another in personalized genetic test marketing, she started her master's program in Animal Sciences at Wageningen University in 2017.

During her master, she specialized in Genetics and Biodiversity. She completed a minor thesis titled *"Application of Multi-Block Data Analysis Methods on Tomato Traits Prediction"* in Biometrics, and a major thesis titled *"Quantitative and qualitative traits prediction from Gilthead Seabream (Sparus aurata) image analysis using machine learning and deep learning algorithms"* in Animal Breeding and Genomics.

After graduating in 2019, she worked as a research assistant in the Department of Plant Sciences at Wageningen University, contributing to the development of a Digital Twin project for poultry farming. She later joined Wageningen Livestock Research as a researcher, focusing on video analysis of dairy cattle behavior, tracking, and database management for breeding programs. In 2020, she began her PhD project on image-based automated phenotyping in aquaculture breeding, with the results presented in this thesis.

Since August 2025, Yuuko has been a postdoctoral researcher at the Center for Quantitative Genetics and Genomics at Aarhus University, Denmark, where she continues to work on the genetic validation of phenotyping tools for selective breeding.

Publications and presentations

Peer reviewed publications

Xue, Y., Palstra, A. P., Blonk, R., Mussgnug, R., Khan, H. A., Komen, H., & Bastiaansen, J. W. (2025). Genetic analysis of swimming performance in rainbow trout (Oncorhynchus mykiss) using image traits derived from deep learning. Aquaculture, 742607.

Xue, Y., Bastiaansen, J. W., Khan, H. A., & Komen, H. (2023). An analytical framework to predict slaughter traits from images in fish. Aquaculture, 566, 739175.

Publications under review

Xue, Y., Bastiaansen, J. W., & Komen, H. (2025). End of the tag era? Critical review on vision-based individual re-identification for aquaculture. Aquaculture Reports. *Under review*.

Manuscripts in preparation

Xue, Y., Kootstra, G., Komen, H., Karapanagiotis, S., Tsakoniti, K., & Bastiaansen, J.W. (2025). Transforming categorical score to quantitative trait by automated image analysis increases fish shape heritability.

Contributions to conferences

Xue, Y., Kootstra, G., Bastiaansen, J.W., & Komen, H. (2025). Breed above the curve: quantitative genetics for shape-related traits using image analysis in gilthead seabream (*Sparus aurata*). European Aquaculture Society (EAS) Meeting Aquaculture Europe 2025, Valencia, Spain

Xue, Y., Palstra, A. P., Komen, H., Blonk, R., & Bastiaansen, J.W. (2025). Catch the wave: a workflow to implement images and artificial intelligence for selective breeding. European Aquaculture Society (EAS) Meeting Aquaculture Europe 2024, Copenhagen, Denmark

Xue, Y., Palstra, A. P., Blonk, R., Mussgnug, R., Khan, H. A., Komen, H., & Bastiaansen, J.W. (2023). Discover novel traits for breeding using image analysis and class activation maps in fish. EAS Meeting Aquaculture Europe 2023, Vienna, Austria

Xue, Y., Palstra, A.P., Khan, H.A., Komen, H., & Bastiaansen, J.W. (2022). Unravelling physical characteristics from three-dimensional image analysis affecting swimming performance of rainbow trout (*Oncorhynchus mykiss*). XIV International Symposium on Genetics in Aquaculture (ISGA), Puerto Varas, Chile

Xue, Y., Bastiaansen, J.W. & Komen, H. (2022). Prediction fat percentage and visceral weight from whole fish images with a multi-input neural network. The 12th

World Congress on Genetics Applied to Livestock Production (WCGALP), Rotterdam, the Netherlands

Xue, Y., Bastiaansen, J.W. & Komen, H. (2022). Tricks of traits: which gives better prediction on fat percentage of gilthead seabream (*Sparus aurata*), increased model complexity or better morphological predictors. The 27th WIAS Annual Conference 2022: Collective Action, Lunteren, the Netherlands

Training and Supervision Plan

Education and Training	Year	ECTs*
A. The Basic Package		**3**
WIAS Introduction Day	2021	0.3
WGS Scientific Integrity	2021	0.6
WGS Ethics in Animal Sciences	2021	0.8
WIAS Introduction course on Personal Effectiveness	2021	1.2
B. Disciplinary Competences		**13.9**
Research Workplan	2021	
Image and Video Analysis	2021	1.6
LAS Species Specific Course on Fish	2021	0.6
WIAS Associated PhDs (WAPS) Education Committee & Chair	2021-2022	5.0
Quantitative Discussion Group (QDG) organization	2021-2022	2.0
Genomic Prediction in Animal and Plant Breeding	2022	1.5
WIAS Annual Conference Organization Committee	2022	2.0
Genetics of Resilience and Trade-off	2024	1.2
C. Professional Competences		**10.5**
Research Data Management	2021	0.5
Project and Time Management	2021	1.5
Scientific Writing	2021	1.8
Scientific Publishing	2021	0.3
Efficient Writing Strategies	2022	1.3
Poster and Pitching	2022	1.0
Presenting with Impact	2023	1.0
Start to Teach at University	2023	0.9

Writing Propositions for your PhD	2023	
Career Perspectives	2023	1.6
Last Stretch of your PhD	2024	
WIAS course the Final Touch	2024	0.6
D. Societal Relevance		**2.5**
Societal Impact of your Research	2021	1.5
Making an Impact	2021	1.0
E. Presentation Skills		**4.0**
WIAS Annual Conference, Lunteren, the Netherlands (Pitch)	2022	
WCGALP, Rotterdam, the Netherlands (Pitch)	2022	1.0
4th CiBreed Workshop, University of Gottingen, Germany (Oral)	2022	
ISGA, Puerto Varas, Chile (Oral)	2022	1.0
Orientation day HAS, Wageningen, the Netherlands (Oral)	2023	
PLF Workshop, Copenhagen, Denmark (Oral)	2023	
Big Data Lunch Meeting, Wageningen, the Netherlands (Oral)	2023	
EAS, Vienna, Austria (Oral)	2023	1.0
PLF Workshop, Helsinki, Finland (Oral)	2024	
EAS, Copenhagen, Denmark (Oral)	2024	1.0
ASG festival, Wageningen, the Netherlands (Oral)	2024	
PLF Workshop, Wageningen, the Netherlands (Oral)	2025	
F. Teaching Competences		**6.0**
Supervise 2 MSc thesis students	2022-2023	2.0
Assisting course Animal Breeding and Genetics	2022	2.0
Assisting course Genetic Improvement of Livestock	2023	2.0
Total		**40**

Acknowledgement

I would like to thank everyone who has supported me over the last years, who has contributed to the project and my thesis.

First, I would like to thank my supervisors. Hans, I keep the sketch we made when you, John, and I discuss the proposal for my PhD project. It is a long way from that day. I would like to thank you for believing in me since the beginning, especially when I had a delay finishing my thesis. Thank you for always taking critical looks at my work while encouraging me to be creative and proactive, to explore and develop my own style, especially in presentations. John, I am very grateful to have you as my daily supervisor. When we were (virtually) at the Unilever Award ceremony, I said I looked forward to continuing working with you, and I would like to say it again. You are a trustworthy supervisor, always sincere about my progress and ideas, and always supportive. Henk, thank you for standing in at the end. I really appreciate your contribution to the thesis and the defense. The only thing I would regret is not having the opportunity to have such discussion with you earlier in my past years at ABG.

I would also like to thank Haris and Gert. Haris, I appreciate that you always give constructive feedback and challenge my methods of choice, which taught me how to use cross-disciplinary methods with understanding and responsibility. Gert, thank you for stepping in for Haris and continuing to advise on the project. I appreciate that you always bring in inspiration and perspectives, that discussions with you are like opening a treasure box with shining ideas. I was very lucky to have both of you supervised on my project.

Very importantly, I would like to thank the support I received from Hendrix Genetics. Robbert, thank you for making many things happen, including my monthly visit to the office and overseas data collections. Bram and Anatolli, thank you for the discussion during every meeting we have, especially the technical inputs. Stephen, thank you for your hospitality in Scotland and all the images collected. Marjolein, thank you for making it possible the visit to Belgium, and what a nice coincidence that we are colleagues now :)

I got to know many people before starting my PhD, and I would like to thank their contribution to this journey. Roel, Claudia, Wijbrand, Ina and Esther, thank you all for the valuable 15 months that I spent as a researcher in WLR, where I have learnt a lot. I would also like to thank Jochem and Angelo for your help during my MSc thesis and PhD application. Being in your lab has made my major thesis easier, and the phenotyper has helped me a lot in my PhD thesis. Finally, Galaxidi, especially Stelios and Poppi, for you had helped me a lot with the image data and the scores. The work on pretty fish is one of my favorite parts of my whole project. It would not have been possible without your contribution.

Doing a PhD was a fun experience for me, thanks to the lovely people in ABG. First, all the fish people. Special thanks to Mark, you helped me a lot in both my PhD and postdoc application, and I appreciate your trust in my knowledge of AI and image analysis. Another special thanks to Arjan for your help with the swimming paper, and your super chill aura when I was freaking out over the reviewers' comments. Farid and Meilisa, thank you for the great time we had during ISGA and EAS. Robert, thank you for traveling to the States for the experiments, and the detailed documentation on everything which practically saved my work later. I would also like to thank just all ABG colleagues, for your being there and giving feedback on every presentation, for every random discussion on science or life, every laugh at coffee break and lunchtime, every holiday events, for all the friendliness and kindness that makes ABG a great place which I will miss enormously. Thank you, Marielle and Fadma, for making all work-related procedures so easy, and Partycom for the activities that broke me free from my routine PhD life.

Especially, I would like to thank my paranymphs, Anouk and Jan Erik. Dank jullie wel voor de geweldige tijd in ons hoekje aan het einde van de gang, het gelach, de geheimen en de grappen naast JE's kabinet, die in vier jaar nooit is gevallen. Anouk, you will always have a spare key to my house. JE, you will always be my No. 1 Dutch knowledge reference. And Tobias, I'm still impressed by your contagious passion in breeding, and I'm very positive about future cross-countries bird watching events. I also hope to join Veluweloop for many years onwards, so this long run with ABG will never end. Outside ABG, I would like to thank WIAS. Besides all the training and courses, I really enjoyed participating in WAPS and meeting fellow PhDs with different backgrounds. Another community that I really appreciate is the PLF workshop. It is always a pleasure to meet passionate people with similar mindsets.

Overall, I think what I cherish the most is a sense of belonging. Thank you all for making my PhD a wonderful, wonderful journey.

Finally, I would like to thank several people in my personal life. Samuel, als ik aan je denk, denk ik aan een zonnige dag bij de rustige zee, frisse lucht met meeuwen die vrij rondvliegen. 什么想法也没有。舒服。因为你是我最好的朋友，感谢你选择我们成为家人。I also like to extend the appreciation to your family, for all the fun trips and gatherings, Eastern brunches, Pesach, Rosj Hasjana and Chanukah dinners. 我要感谢我的父母对我求学之路一直以来的鼓励和支持。薛润民先生，你对自由和平等坚持着与时代无关的、充满浪漫主义的向往和追求；马淑娜女士，你对生活保持着积极乐观的态度，和利他主义般的真诚和善良。你们独立的品质和美德影响了我成为今天的自己。I would also like to thank Keng-Wei Chang and Zhengrun Wang, for the conversations and discussions we have about almost everything. You help me build my assertiveness towards a future with uncertainty. Finally, Qifan Gao, Wen Gao, Qiyue Shao, Zhenyu Shao and Hanxiao Zhan, thank you for creating the safe place to cure and share. 愿我们都能在认清生活的真相后继续勇敢地热爱生活。

Authorship statement

Chapter 1 General introduction. I developed the research question from my MSc thesis and expanded it into a personal PhD grant proposal under the guidance of my promotor and supervisor. The research proposal was the basis of chapter 1. I defined the research question, explained its scientific relevance, and outlined its societal importance. I revised the text four times based on feedback from my supervisors.

Chapter 2 An analytical framework to predict slaughter traits from images in fish. I had the idea and developed it into a study. I proposed the methodology, performed the analysis, and interpreted the results. I wrote the original draft and revised it based on co-authors' feedback. I integrated the reviewer's feedback with extra analysis and results after peer review.

Chapter 3 Genetic analysis of swimming performance in rainbow trout (Oncorhynchus mykiss) using image traits derived from deep learning. I conceptualized the study, proposed the methodology, performed the analysis, and interpreted the results. I validated the findings after peer review. I wrote the original draft and revised it with co-authors' feedback.

Chapter 4 Individual re-identification of Atlantic salmon (Salmo salar) using images taken 5-months apart. I conceptualized the study with my supervisors and the industrial partners, contributed to data collection, proposed the methodology after discussion with my supervisor, performed the analysis, and interpreted the results. I wrote the original draft and revised it with co-authors' feedback.

Chapter 5 End of the tag era? Critical review on vision-based individual re-identification for aquaculture. I conducted the study based on my supervisors' suggestions. I developed the structure, asked for feedback, and implemented it throughout the process. I restructured the outline three times and revised the text seven times.

Chapter 6 Distance to perfect shape in gilthead seabream (Sparus Aurata) is moderately heritable when quantified objectively by automated image analysis. I conceptualized the study, proposed and developed the methodology, performed the analysis, and interpreted the results. I reviewed the results twice with my supervisors. I wrote the original draft and revised it three times based on co-authors' feedback.

Chapter 7 General discussion. I proposed the discussion topics and wrote the first draft. I implemented comments from my supervisors and revised the text three times after discussions with them. The discussion reflects my own opinions, which may differ from those of my supervisors.

In all chapters except chapter 2, I have used ChatGPT to provide language support, to enhance clarity and improve the flow of writing. I am fully responsible for the final content and interpretation of this thesis.

Colophon

The research described in this thesis was financially supported by the TKI Agri & Food project LWV20386 (Wageningen, the Netherlands) and the partner Hendrix Genetics (Boxmeer, the Netherlands)

The cover of this thesis was designed by Lotje van Lieshout

Printed by: Digiforce | Proefschriftmaken.nl