INVESTIGATION IN YELLOWTAIL KINGFISH INDIVIDUAL IDENTIFICATION: THE LIMITATION IN WHOLE-FISH IMAGES PRE-TRAINED VGG-19 MODEL

STUDENT NAME: Xirui LIU
STUDENT NUMBER: 1047640

COURSE NAME: MSc Research Practice (24-36cr) ABG
COURSE CODE: ABG79324

SUPERVISOR: John Bastiaansen
Yuuko Xue

DATE OF SUBMISSION: 21th August 2023

Abstract

Research that accurately tracks specific individuals within fish cultivated groups is an emerging concern in modern aquaculture. The combination of deep learning models and computer vision methods presents the potential for fish farming to make informed decisions based on image data. Some studies have demonstrated the effectiveness of using pre-trained deep learning models to identify individuals by using distinctive patterns or facial features observed in these animals within the same species. However, methods for addressing individual fish recognition are still worthwhile to investigate in the current scenario. In this project, we investigated the feasibility of utilizing a pre-trained VGG19 deep learning model to identify individual Yellowtail Kingfish(*Seriola lalandi*) from whole-fish photographs. The image dataset comprised 803 individual fish and spanned a growth period of four months. This dataset was used as input for training and testing the pre-trained VGG19 model. Additionally, we investigated the influence of time longitudinal data on the model's performance. Finally, we investigated the interest region of manual identification. Our findings not only shed light on the limitations of the pre-trained VGG19 model in accurately identifying individual fish from whole-fish images, but also underscore the significance of considering temporal changes in fish appearance for robust individual identification. Particularly, as the time span increased, the model faced greater challenges in distinguishing between individual fish. This study emphasizes the need for further research into tailored approaches for precise and consistent individual fish recognition.

Keywords — Yellowtail Kingfish, Individual Identification, Deep learning model, VGG-19, CNNs

Content

1. Introduction

1.1 Background

The recognition of individual specimens has become an important issue in fish studies. Whether in zoological experiments or in the breeding industry, stable and accurate long-term recognition are the prerequisite for obtaining behavioral information, habitat or breeding environment preferences, distribution, population structure, and growth rate information[1,2,3]. In the aquaculture industry, accurate fish measurement, especially measuring at the individual level, is critical for fish breeders to assess the quality and economic benefits of specific fish[4]. Therefore, this puts forward diverse requisites for identification technology to achieve heightened accuracy, reduced costs, and minimal impact on measured animals.

A traditional approach frequently used in aquaculture to identify individual fish is tagging. Commonly used marking techniques are usually invasive, such as subcutaneous chemical markings, amputations, insertion of transponders and tattoos[5,6]. These methods often pose risks to animal health or survival, and raise animal welfare concerns[7]. Even non-invasive tagging methods such as external stains, labels, collars, can lead to issues like behavioral changes, reduced fitness, or loss of tags over time[8,9]. An alternative approach is to identify individuals based on natural markings, such as coloration, spots, or stripes. Recording these natural markings with image data not only avoids physical marking and reduces animal stress, but also retains image database that can be compiled in careful identification process. And this has led to extensive research on photo-identification technique, although this approach was initially predominantly applied to large terrestrial vertebrates[10,11]. However, an increasing number of research indicates that image recognition can not only be applied to some large terrestrial vertebrates, but also be feasible in some fishes. For example, Eurasian perch can be distinguished by the stripes on the body[12], Permanent melanocytic speckle pattern in *Salmo salar* with long-term stability, providing a strong reference for long-term individual recognition[13].

Deep learning(DL) approaches is a state-of-art method used in image analysis and computer vision(CV) domains, which have produced impressive results since it was combined. In contrast to manual image recognition, CV approaches which combined deep learning models can effectively address a multitude of recognition tasks, thus mitigating the issue of inefficiency and insufficient characteristics derived from human experience[14]. Recently, the bulk of solutions for fish identification make use of deep neural network(DNN) that are built on the framework for fish recognition. This framework consists of three processes: fish object detection, fish feature extraction, and fish feature comparison[14]. Convolutional neural network(CNN) is a class of DNN, mostly applied to visual analyses. CNN and its variants have the key advantage of automatically extracting and learning image features[15]. Moreover, a distinctive feature

of CNNs is that they can be fine-tuned for new datasets, such as fish datasets, after they have been trained on general objects datasets[16]. This adaptability of transferring their acquired features and knowledge across various datasets lends CNNs exhibit a robust capability when dealing with small-scale datasets. As a result, the good recognition performance and transfer learning capabilities of CNNs model have make us interest in whether they can address identification issues in fish studies.

1.2 Related Works

1.2.1 Different species identification in Fish

The computer vision approach, combined with the deep learning model, achieved a significant reduction in the error rate on the classification of different species of fish using images. By using a pre-trained convolutional neural network as a cross-layer pooling algorithm for generalized feature detectors, an accuracy of 94.3% was achieved in the fine-grained fish classification problem[17]. Another example is to construct a new deep learning algorithm framework that trained by independent datasets, fine-tuning is conducted for different species, assigning classification scores to each class in each image. Reducing the species misclassification rates from 22% to 2.98% after post-processing using this framework[18]. Furthermore, a study proposed a novel CNN consisting of three branches for classifying fish at the species, family, and order in a taxonomy. This approach aimed to enhance recognition accuracy for different species with similar features, which achieved an increase in accuracy from 0.86 to 0.96[19].

1.2.2 Individual identification in Fish

The classification of the same species of fish is more difficult to deal with than the classification of different fish. The classification problem of different species of fish can usually be solved by visual cues used in the classification, commonly used such as key points, color, texture, shape, etc. These features can be refined into a set of up to 47 features[20]. In the same species, most of the characteristics cannot be distinguished because of the phenotypic similarity. Studies have proved that long-term and stable pigmentation or pattern composed of scales is an important feature of individual recognition. Take an example, a study demonstrated that the scales of *C. carpio* vary greatly in number, size and combination, which allows these characteristics to be used for individual identification[21]. Moreover, an automated method for the identification of Lake Malawi cichlids using computer vision and geometric morphometry was presented. By extracting color and stripe features from photographic images, the model can classify various images belonging to 12 different classes in an average accuracy of

78%[22]. A recent development is to use skin dot patterns to enhance recognition through the images of Atlantic salmon (*Salmo salar*) scales[23], and the identification accuracy was 100% for 30 fish based on out of water images. These above researches indicates that individual recognition within the same species necessitates specific features among individuals to possess quantifiable or comparable differences, which form the foundation for achieving individual recognition.

In this project, the research animal is the yellowtail kingfish(*Seriola lalandi*). As a kind of new aquaculture fish with broad economic value and consumer market, the yellowtail kingfish has garnered significant attention from research area such as important economic traits[24], bioenergetic growth model[25], and health conditions under specific aquaculture conditions[26]. However, in Netherlands, there is currently no research on the individual identification of yellowtail kingfish. Therefore, a stable and long-term individual recognition system can greatly assist both researchers and fish breeders in gaining more detailed information about individual fish, helping them to make more accurate assessments.

The main goal of this project is to investigate the performance of pre-trained CNN models in addressing the yellowtail kingfish individual identification issues. The project is to adopt a pre-trained VGG-19 model trained by large-scale image dataset and to construct the preliminary fish individual identification models with whole-fish images from more than 800 fish at four different time point. Transfer learning methods are beneficial when the available training data is not large enough (as in this project), or when there are few examples of different variables[17]. The VGG-19 model is chosen as its capability have been proved in a wide range of applications in both fish species classification and individual recognition[27,28]. We also explored the potential for individual recognition within specific regions of whole-fish images in the yellowtail kingfish due to whole-fish model’s limitation.

2. Material and Methods

2.1 Data collection and Raw data

The experimental animal in this study is Yellowtail kingfish (*Seriola lalandi*). A total of 803 farmed yellowtail kingfish were invasive tagged in the experiment, with an initial average weight of 157 grams and an initial average length of 10 cm. The fish were tagged with TROVAN's microchip ID-00C for stable and long-term data collection. A total of four data collections were performed over a four-month period with one-month intervals. During each session, the out of water images of the fish were collected. The fish was caught and placed horizontally on a conveyor belt, then the fish was transported to the photographing area along with the conveyor belt. The photo area consists of a blue platform, and two shooting cameras. These two cameras capture the images of fish's head, and the images of fish side body. After scanning the microchip inside the fish to obtain its ID, the experimenter promptly operates the camera to capture photographs of the individual(as shown in the Figure 1.1 below), and generated two types of data, RGB images and depth images. After data collection was completed, all fish were returned to the tanks.

In Figure 1.2, it shows the RBG side images of one individual fish which collected at four time points, and the resolution of each image is 1280 x 720 pixels.

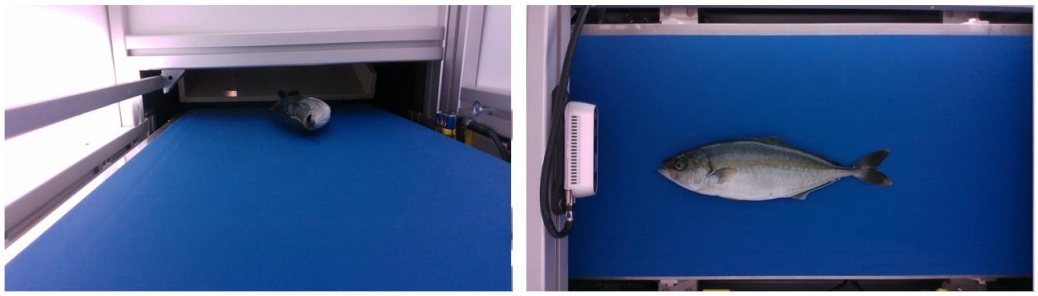

Figure 1.1: Example of images from data collection. Left image was collected from front camera, and right image was collected from top camera.

Figure 1.2: RBG profile images of individual fish that collected at four time points. Top left: February, top right: March, bottom left: April, bottom right: May. Resolution of each image is 1280 x 720 pixels.

The raw data generated after data collection includes RGB information and depth information of the head image for each fish among first two time points and side image for each fish among four time points. Typically, image data was recorded only once per fish per time point. The third and fourth data collections also included the image data of untagged fish, as shown in Figure 2. From April 2022, a very large number of image data of untagged individuals had been collected. And as time went by, as the fish died and the collected data was lost happened through the data collection process, the final long-term and stable image datasets have fewer than 803 fish individuals.

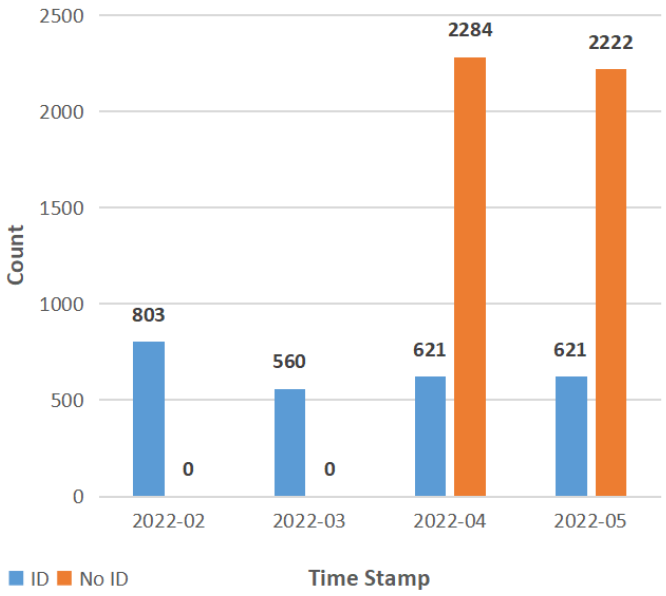

Figure 2: Visualization of the number of IDs distribution among four months.

2.2 Data Pre-processing

In this part, a comprehensive image pre-processing process was conducted to prepare the datasets for individual fish identification. The following steps are: removing irrelevant images; categorizing and integrating images from different time points based on their matched IDs; identifying fish subjects within the images and subsequently conducting automatic segmentation. Through these procedures, the raw datasets were transformed into a smaller, less noisy dataset that is conducive for recognition purposes. The overview of this process could be seen in Figure 3.

2.2.1 Images Removal and IDs matching

Three types of images that were not relevant to our identification task were removed. These images were:

- Depth Images: This project focuses on the identification of individual fish using visible RGB images. The depth image provides information about the distance of the object from the camera and is often used to represent the object's volume. Therefore, all depth images were discarded from the dataset.
- Front-side Images: Upon observing the images captured from the frontal view of the fish, it was noted that these images exhibited features such as the contour of the fish's head and mouth (Figure 1.1). Nevertheless, given their absence of essential distinguishing features for identification and the potential to introduce computational burden during model training, these images were also excluded.
- No IDs fish's Images: Images where the fish's identity was not clearly established, were removed from the dataset. As these images do not participate in the training process of the model in this project.

A Python script was written to deal with files containing image names and corresponding ID information. Integrated images from different time points based on their respective ID. Ultimately, each image was associated with its corresponding time point and ID.

2.2.2 Image Segmentation

Segmentation has been applied to enhance the image which aim at reducing the image size and filtering out noise. Edge detection is a tool which makes the process of image segmentation and pattern recognition more comfortable[29], as it offers significant feature parameters for recognition targets and interpretation from images. The OpenCV[30] package was used to achieve image segmentation, and the Canny algorithm was mainly applied for edge detection.

The below section outlines the core steps undertaken in this process:

1. Image crop and object identification:

The position of the cameras and the edge of the platform is fixed in all side images. This means that through the specified pixel coordinates, the camera and area out of platform can initially be removed from raw image.

The hue-saturation-lightness(HSL) color space is utilized to extract the saturation channel. A gaussian blur is applied to the saturation channel, followed by thresholding and edge detection using the Canny algorithm. The largest contour was selected as the contour of interest. After that, the 'boundingRect' function[30] was applied to this largest contour, which generated a rectangle close to this contour area and returned both the pixel coordinates of the top-left corner and the length and width of this rectangle.

2. Resizing and Relocating:

Background noise during model training can enhance the model's robustness[31]. Additionally, maintaining consistent aspect ratios for all images is desired. Thus, in this step, we introduced a fixed margin of 30 pixels to the rectangle generated previously and set a fixed aspect ratio of 2:1 for the final cropping rectangle. Based on the pixel coordinates of the top-left corner and the dimensions of the bounding rectangle returned by the 'boundingRect' function in the previous step, we calculated the new top-left corner coordinates and pixel dimensions of the final cropping rectangle. This ensures that the segmented image includes the entire fish side body along with a portion of the blue background.

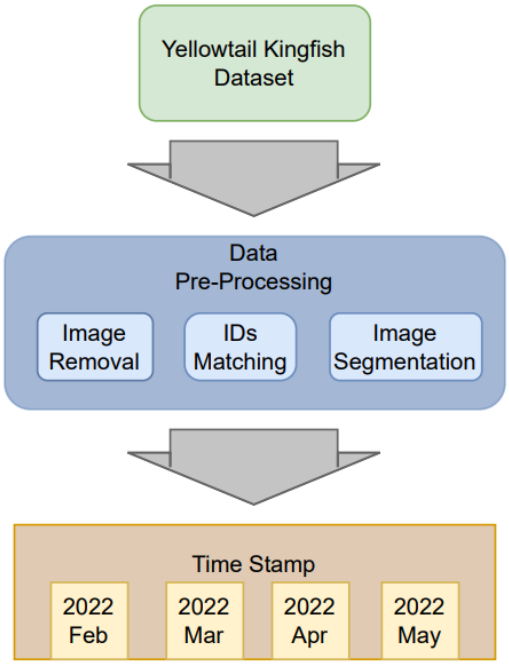

Figure 3: Block diagram of the pre-processing system.

Upon completion of these steps, the original data has transformed to retain only the RGB images of individual fish at four different time points, each associated with a unique identification (ID). These images were of reduced size, ensuring the preservation of the focal subject while incorporating a certain amount of blue background, thus forming a training image dataset containing both informative signal and noise.

2.3 VGG-19 Convolutional Neural Network Construction

2.3.1 Structure of VGG-19 model

Constructing a CNN model from scratch demands a substantial volume of data, a requirement that proves nearly unattainable in this project. Instead, an exceedingly promising approach is to use pre-trained models on extensive datasets such as VGG model. In this endeavor, we opted for the VGG-19 model. The pre-trained VGG-19 model encompasses approximately 143 million parameters, which were acquired through learning from the ImageNet dataset[32]. By invoking these parameters, our objective is to facilitate transfer learning of the model across diverse datasets.

As shown in Figure 4 below, VGG-19 encompasses 19 trainable layers, comprising convolutional, fully connected max-pooling and dropout layers. In our devised solution, we capitalized on the pre-trained convolutional base and tailor the classification component to our context, including a densely connected classifier and dropout layers for regularization. In this model configuration, most convolutional layers within VGG19 were frozen, and additional fully connected layers were trained on top of the model. A fully connected layer with 512 hidden units and ReLU activation function was introduced, followed by an output layer with 'Softmax' activation function at the top, catering to the multi-class classification task.

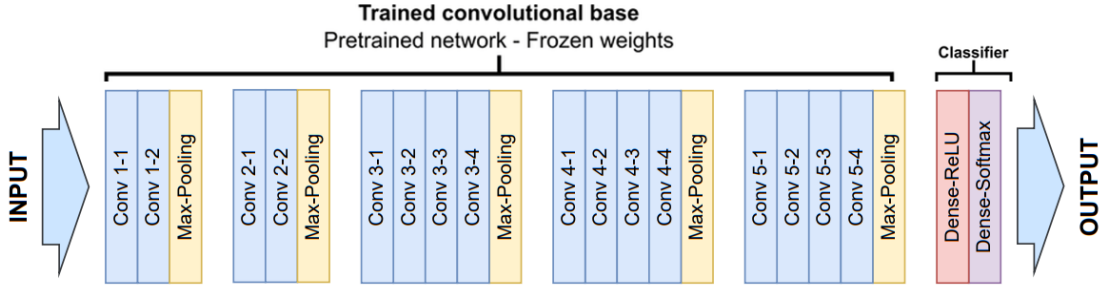

Figure 4: Overview of the VGG-19 network architecture with description of layers.

2.3.2 Image Resizing

To facilitate compatibility with the pre-trained VGG19 model, all the segmented images were resized to a standardized format of 224 x 224 pixels. This resizing ensured uniformity and consistency in the dataset, as VGG19 requires inputs of this specific size for accurate feature extraction and recognition.

2.3.3 Data Split Strategy

To effectively train and evaluate the model, three specific recognition tasks were performed using the recorded datasets as below:

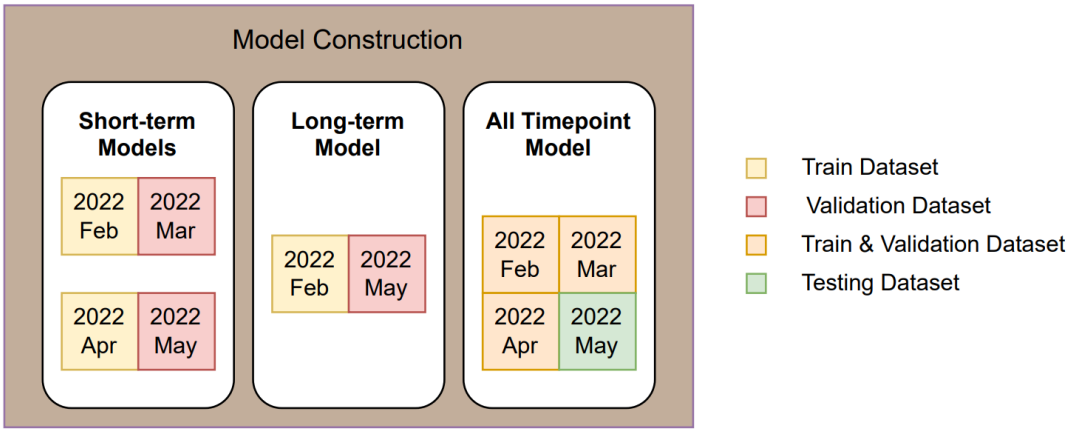

Figure 5: Block diagram of the data split strategy base on three specific identification tasks. The meanings of blocks of different colors can be seen from the right legend.

a. Short-term Models:

The first focus was on testing the uniqueness in whole-fish images for individual identification, referred to as the short-term pattern. In this mode, the dataset from four time points was divided into two groups with a one-month interval: February and March; April and May. The dataset of February and April were used as the training dataset, while the image data from the subsequent month served as the validation dataset. This approach aimed to enable the model to capture short-term variations and trends in whole fish image.

b. Long-term Model:

The long-term model aims at exploring potential patterns that are stable throughout the 4 months period within whole-fish images. The image data collected at the longest intervals, specifically those from February and May, are utilized for the construction and training of this model. Through such design, we expect to test whether the model can overcome the time longitudinal variations present in the dataset due to fish growth

over three months. By training on images of different individuals at the same time point and validating on matching ID images three months later. This enables it to learn underlying stable features inherent in these images. Furthermore, this approach serves to evaluate the model's ability to generalize over an extended period.

c. Full Time Points Model:

The third mode utilizes a dataset encompassing all time points, ensuring comprehensive training and evaluation across the entire dataset. The images from May were designated as the testing dataset, while those from February to April were employed for training and validation. For those images which collected from February to April, one time point was randomly chosen, and the image matched to that time point was used as part of validation dataset in model construction. And the images which collected in the rest two time points were employed for training. This strategy was implemented to ensure the training and validation dataset coverage across the first three time points.

2.3.4 Data Augmentation

Due to the limited size of the original dataset, data augmentation was crucial to enhance the model's performance and generalization capability. Data augmentation involves creating more training samples by applying various transformations to the existing images.

In this project, data augmentation was implemented using various techniques as below:

- Rotation: A random group of images were rotated at a random angle in the range of 0 to 40 degrees, to simulate different views and orientations of the fish.
- Width and Height Shift: The images were randomly shifted horizontally and vertically based on the 10% fraction of total width to introduce small displacements in the fish's position.
- Zooming: The images were randomly enlarged or zoomed out by 20% to simulate variations in the fish's size and scale.
- Horizontal Flip: The images were horizontally flipped.
- Brightness Changing: A random group of images were selected to adjusted darken or lighten to simulate varying lighting conditions.
- Rescale: By this process, the pixel values in images were converted from the [0, 255] range to the [0, 1] range. This process is to avoid high pixel range images have a large weight in model training.

We expected this process to increase the dataset size and diversity. And enables the model to learn from a more diverse set of images, leading to improved performance on unseen data.

2.3.5 Model Training

The proposed and describe VGG-19 model took 224 × 224 pixel preprocessed RGB images as inputs and generated predicted accuracy indicating individual fish.

For multi-class classification problem, the categorical cross-entropy loss function, also referred to as 'Softmax' loss[33] has been applied. The 'Softmax' formula and Cross-Entropy formula are given by Figure 6.1. Figure 6.2 demonstrates the specific stages at which these formulas are applied. The number in Figure 6.2 are only shown as example for better explanation.

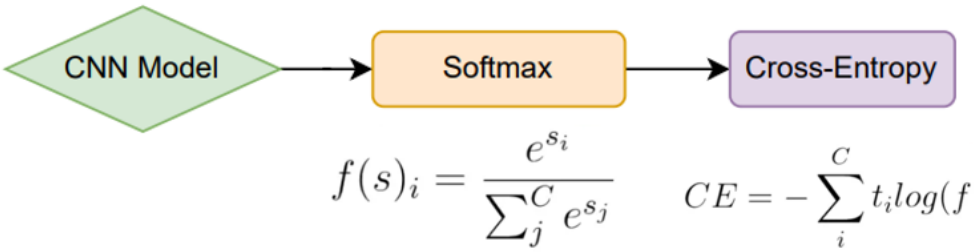

Figure 6.1: The 'Softmax' formula and 'Cross-Entropy Loss' formula.

C: number of classes.

si: input vector to a 'Softmax' function which consist of *C* elements for *C* classes.

sj: scores inferred by the net for each class in *C*.

ti: ground truth vector.

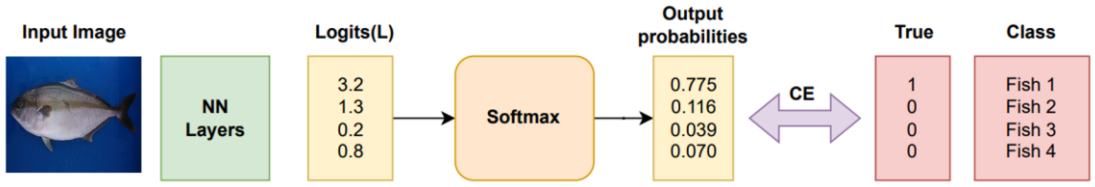

Figure 6.2: The application of 'Softmax' formula and 'Cross-Entropy Loss' formula.

As shown in Figure 6.2, the fully connected layer of the CNN produces a logits vector L, also is *si* in figure 6.1, which is then transformed into probabilities through the 'Softmax' formula. The Cross-Entropy takes the output probabilities (P) for each class and measures the distance from the truth value. Finally, calculate the accumulation of cross-entropy values for all classes to generate one of the values for assessing model performance: the loss value.

The training process commenced with predefined settings, including various parameters such as learning rate(0.0001), batch size(32), and number of epochs(100).

2.4 Manual recognition of specific features

After exploring and testing models trained on whole-fish images based on three strategies, the potential features for manual identification were further considered and the specific regions indicated by these features were extracted. It is necessary to select features that persist for a sufficiently long period of time (i.e. usually the entire life of the fish). In this process, three features of yellowtail kingfish, lateral line, shape, and the yellowish region located in the middle of the body, were selected as specific features for identification. As shown in Figure 7, the manual recognition of three features is performed on the image data collected at the time point of May 2022 as an example.

Represented by a red-highlighted curve in the top image of Figure 7, the lateral line is a prominent black line extending along the midsection of the fish, starting from the gill cover and continuing towards the tail. Its distinctive curvature and shape make it a significant characteristic for individual differentiation. The shape of the fish is highlighted by the yellow line as shown in the middle part of Figure 7. In this project, this outline is the outermost shape of the fish, including the shape of dorsal fin, caudal fin, anal fin and pelvic fin. The above-mentioned part is also commonly used as the characteristics to identify individual fish[20]. This makes the outline's shape become a characteristic we are interested in, and we speculate different individuals can be distinguished by it. Finally, the green frame shown in the bottom of Figure 7 circles a yellowish band from the snout to near the upper region of the caudal peduncle. As one of the most recognizable features of the yellowtail kingfish, the individual recognition performance of this feature was also explored in this project.

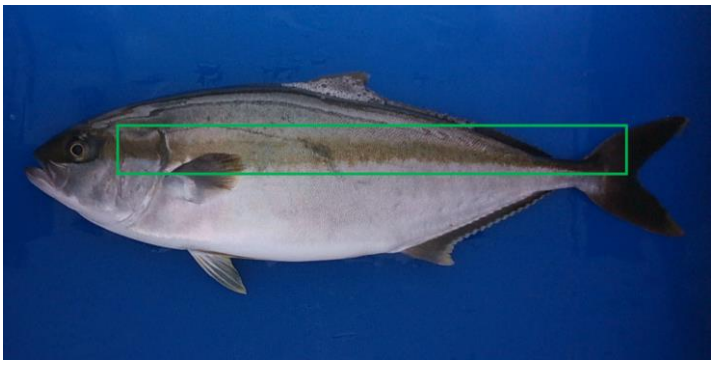

Figure 7: Display of the selected three features in the image. Top: Lateral Line. Middle: Shape. Bottom: Yellowish Band.

Traditional manual recognition only relies on human eyes for photo comparison, which is inefficient. Many studies have proved that the accuracy of computer-aided visual recognition is improved[34] compare to that of manual identification. In this process, we used manual identify and highlighting approach for fish lateral line and utilized image processing technology from OpenCV package to highlight the remaining two features to optimize the process of manual recognition.

2.4.1 Lateral Line

The lateral line enables fish to detect motion and pressure gradients in the water, and always could be seen from fish body's side. The manually identified recognition region was defined on the upper half of the fish's body. The distinct black scale regions visible on the fish's lateral side was delineated. Upon zooming in on these regions, the continuous or intermittent black scales were accurately recognized and highlighted in red. The resulting curved lines with distinctive curvature were considered and compared as features for distinguishing different fish individual.

2.4.2 Shape

The image processing package OpenCV was used in this step to automatically extract the outer contour of the fish from the image. The input RGB image was transformed into the hue-saturation-lightness (HSL) color space. The saturation channel, known for its ability to emphasize object boundaries, was isolated from the HSL image. A gaussian blur was applied to the Saturation channel to reduce noise, followed by the utilization of the Canny edge detection algorithm. This produces an edge map highlighting the boundaries of the main object. By employing the 'findContours' function, a series of contours were extracted from the edge map, and they were colored as green. Among these contours, the one with the biggest area was identified as the outer contour of the main object, which is the fish shape.

2.4.3 Yellowish Band

Within the HSL color space, the Hue channel encapsulates color information, making it instrumental in distinguishing specific color ranges. The extracted Hue channel was visualized, aiding in the interpretation of color distribution. By thresholding the Hue channel within a predefined yellow color range, the yellow regions on the fish's body were isolated. These regions correspond to the distinct yellow areas of interest.

In the final comparative analysis, two IDs were randomly selected from the data sets, and two times were randomly selected from the four time points. Four images corresponding to these selected IDs and time points were retrieved. Visual enhancements were applied to the three types of features mentioned above in the images. Subsequently, a image map was generated to visually showcase the variations among different fish individuals in these three manual identification features.

3. Results

3.1 Dataset Curation

3.1.1 IDs matching and image segmentation

As shown in Table 1, the table presents the retained number of IDs and the corresponding count of images at the four time points after removing not relevant images and images without identifiable IDs from the original dataset. The table reveals that the total number of images exceeds the number of IDs, indicating that there is at least one available side-view RGB image of a fish for each ID at every time point. February, serving as the initial time point for image collection, retains the highest number of available images and IDs. As time progresses, the number of collected IDs gradually decreases. Among these, the month of March has the least collected IDs and image data. This reduction may potentially be attributed to data collection oversights rather than loss of data due to fish mortality.

Table1:The available IDs and available images retained at four time points.

Time Point	February	March	April	May
IDs Num	803	560	621	570
Images Num	830	576	661	578

Figure 8 provides a more intuitive depiction of the overlap of these IDs across the four different time points. The Venn diagram vividly illustrates that a total of 496 fish have been consistently and longitudinally captured in available images over the four months. Among these, 169 fish, for which image data were collected in February, are absent from image collection in the subsequent three months. Additionally, image data for over 70 fish appear to be overlooked or lost during the month of March, as the image data for these fish reappear in April and May.

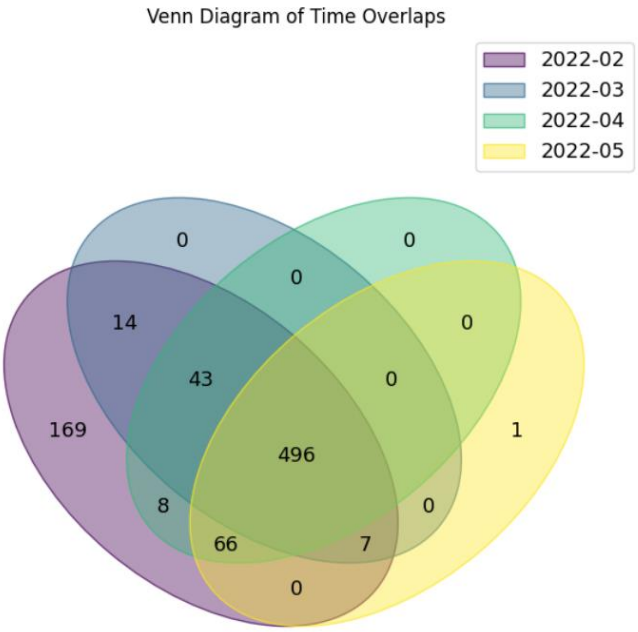

Figure 8: Venn diagram of ID overlaps situation among four-time stamps.

The sequential visualization outcomes of the automated image segmentation process are presented in Figure 9. The leftmost image within the figure depicts the original image. The yellow dashed rectangle delineates the region chosen based on fixed pixel coordinates, excluding the white sensor and blue platform. The green dashed rectangle illustrates the region proximate to the maximum contour, which obtained through the 'findContours' function automated extraction of the primary contour within the image. The red solid-line rectangle is derived from the green dashed rectangle, incorporating a fixed 30-pixel margin, and adhering to a 2:1 aspect ratio in length and wide for the rectangle. This final segmented image comprises both the main object and part of background noise. Taking this image data as an example, the automated segmentation process transforms the image size from its original 2.63MB to a reduced 598KB.

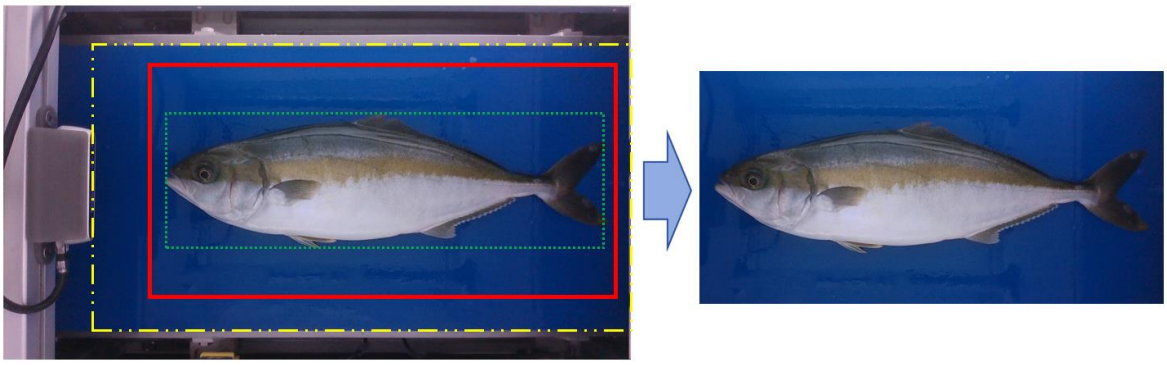

Figure 9: Visualization of automated cropping of image. Yellow dashed box: initial cropping result with fixed pixel coordinates. Green dashed box: rectangle generated through 'findContours' function applied to maximum contour. Red line box: final cropping rectangle.

Following these steps, the final dataset used for model construction was significantly refined compared to the initial dataset, resulting in a reduction in the number of images from 28,992 to 2,645 and a decrease in data size from 62GB to 1GB. This reduction in dataset size not only contributes to mitigating the training time required for the VGG-19 model but also has implications for optimizing computational resources and enhancing the efficiency of both model training and prediction processes.

3.2 Models Training and Performance

3.2.1 Data Augmentation visualization

As shown in Figure10, a 4x4 augmented image map is presented, displaying sixteen images generated through data augmentation using a generator applied to an image from the training dataset. From the images, it is evident that the original image undergoes a series of modifications, including rotations, flips, cropping, scaling, and changes in brightness. These alterations give rise to new versions of the image, each exhibiting subtle variations from the original.

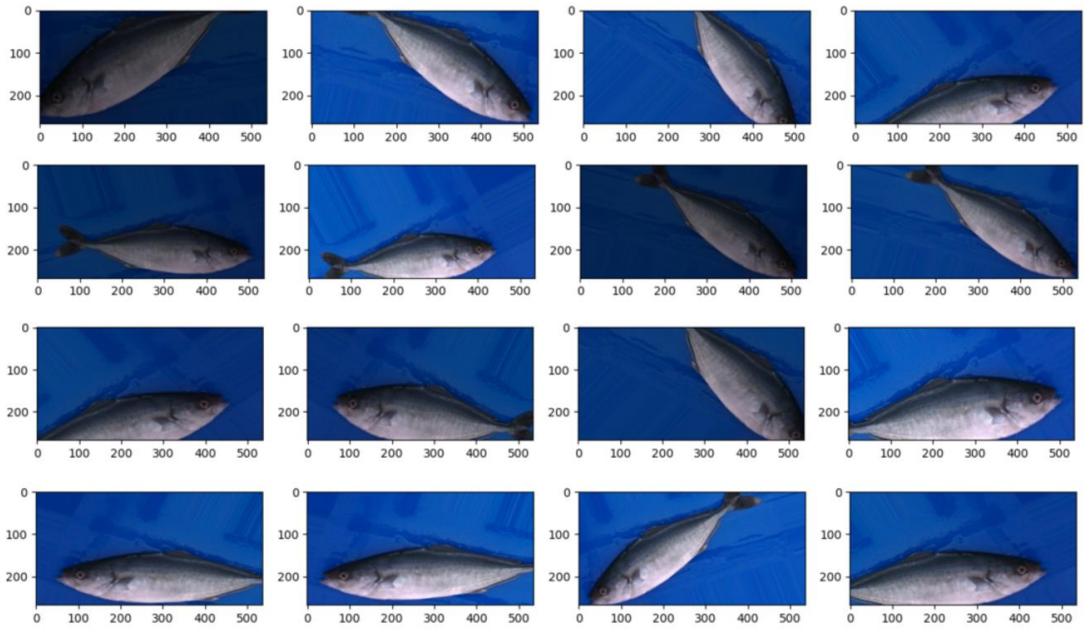

Figure 10: Visualization of image augmentation.

3.2.2 Learning curve for short-term models

Figures 11.1 and 11.2 depict the learning curves of two separate short-term models trained for 100 epochs each, based on loss and accuracy metrics. Figure 11.1 presents the learning curve of a model trained on February image data and validated using March image data. Notably, both the training loss and validation loss exhibit an amplifier-shaped divergence, starting from approximately the same loss value and gradually expanding with increasing epochs. This trend indicates that as the model learns from the February image data, it becomes more prone to errors when validating against the March image data. While the training accuracy of this model improves with epoch progression, the validation accuracy demonstrates minimal observable growth when the trained model is applied to the validation dataset. This lack of improvement in validation accuracy aligns with the consistent rise in validation loss, indicating persistent misclassifications of the trained model on the validation dataset. The learning curve shows that the model exhibits overfitting on the training dataset.

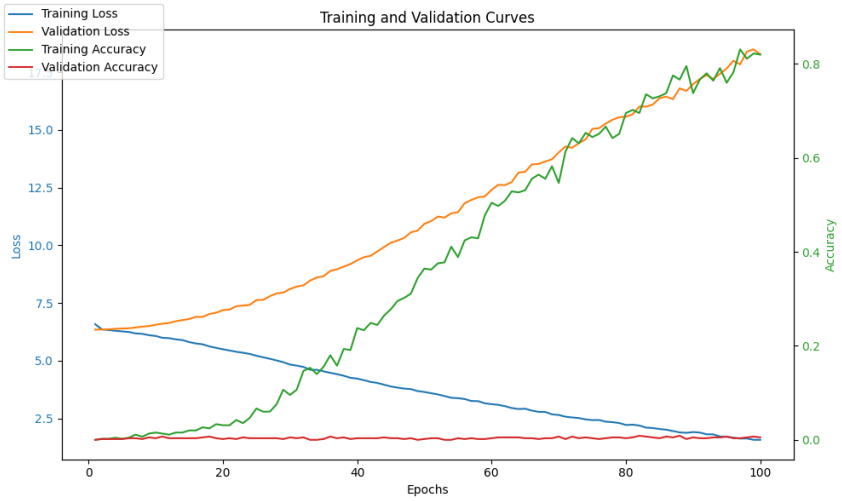

Figure 11.1: Learning curve which shows short-term model that trained by February time point dataset and validated by March time point dataset.

Similar trends are observed in the other short-term model. In Figure 11.2, both the training loss and validation loss similarly diverge with increasing epochs. However, by comparing the gap between the train loss and validation loss of these two short-term models, we observed that although the gap of both models increased with the growth of epochs, the model constructed using the April and May datasets exhibited a lower gap compared to the other short-term model. This indicates a reduced accumulation of errors. It must be pointed out that the training loss has a huge drop in this model, almost down to close to 0. Additionally, for this model, the training accuracy experiences a more rapid improvement as epochs progress after being trained on April image data. Notably, when this trained model is applied to validate against May image data, there is a subtle improvement in validation accuracy compared to the complete absence of change observed in the accuracy curve of the other short-term model.

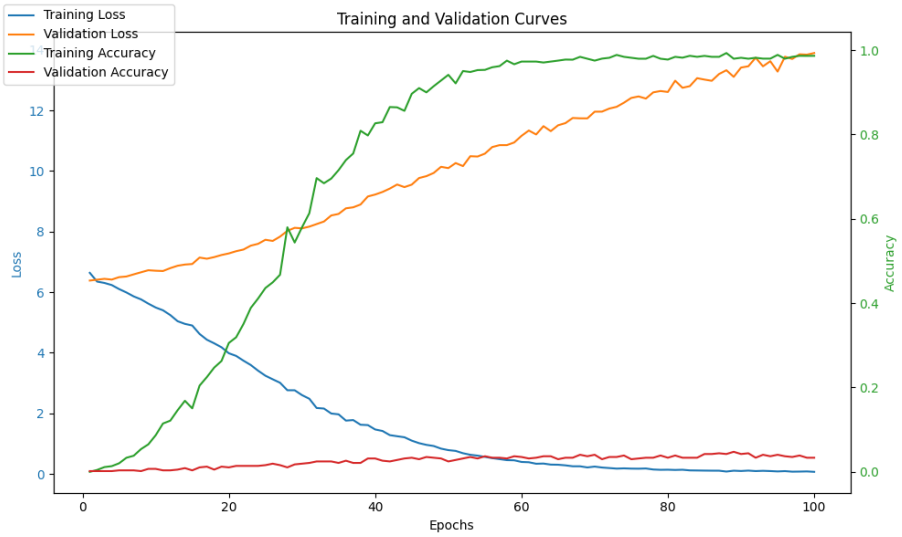

Figure 11.2: Learning curve which shows short-term model that trained by April time point dataset and validated by May time point dataset.

3.2.3 Long-term Model

Through Figure 12, it can be observed that when the model was trained using the image data from February and validated using the image data from May, the learning curve of the model reflected suboptimal performance. As evident from the graph, with increasing epochs, the gap between training loss and validation loss widened, indicating that the model experienced overfitting on the training dataset. Additionally, it can be noted that while the training accuracy of the model increased with epochs, the trained model was unable to make meaningful predictions on the validation dataset. This suggests that the model trained with February data exhibited poor generalization performance on the May dataset.

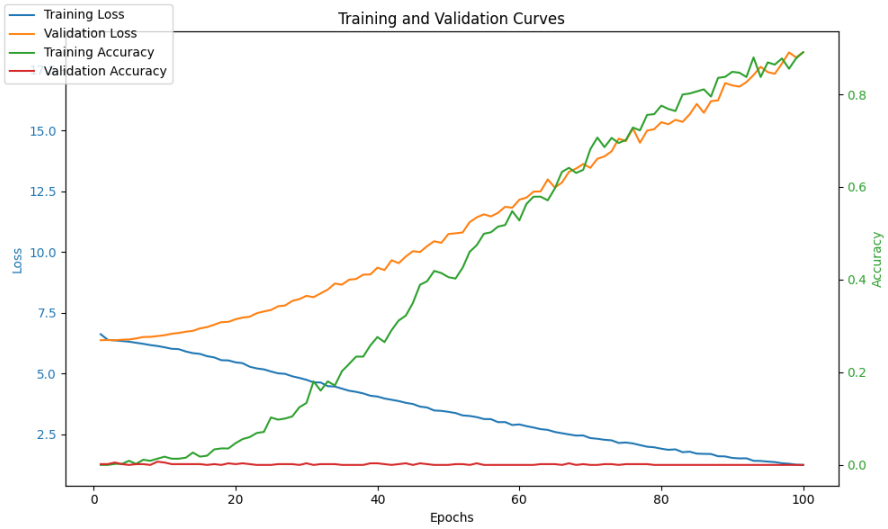

Figure 12: Learning curve which shows long-term model that trained by February time point dataset and validated by May time point dataset.

3.2.4 Full-time Model

In the final model, datasets from all time points were utilized. On one hand, this approach expanded the model's learning data, while on the other hand, it enhanced the model's ability to generalize to time longitudinal datasets. As shown in Figure 13, the learning curve provides insights into the similarities and differences between the full-time model, short-term model, and long-term model. Firstly, it can be observed that the training process of the full-time model is more challenging in the same 100-epoch training period, as indicated by the relatively smaller decrease in training loss and smaller increase in validation loss compared to the other models. Correspondingly, the accuracy of this model fluctuates within a relatively smaller range, with more noticeable fluctuations.

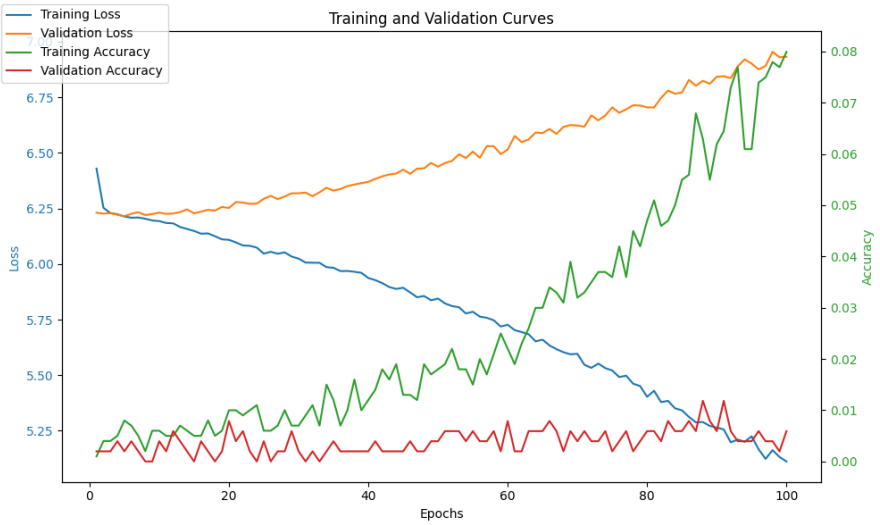

Figure 13: Learning curve which shows Full-time model.

3.3 Manual Identification

3.3.1 Manual features selection

Figure 14.1 presents the visual identification of the lateral line, a linear arrangement of black scales located in the middle section of the fish's side. To facilitate recognition, these regions were outlined and subsequently magnified for a clearer view of the black line contours. Although variations exist among different individuals, in most cases, the lateral line in these two regions, near the head and tail, is distinctly visible.

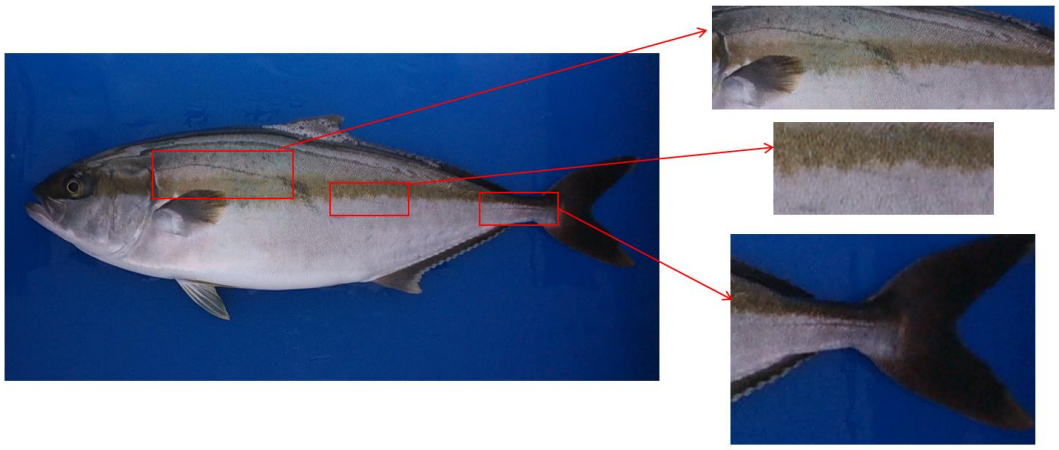

Figure 14.1: Display of the lateral line manual identification.

In Figure 14.2, the results of computer vision techniques enhancing the shape of the fish's lateral profile are displayed. The fish's outline is highlighted with fluorescent green lines. While the computer-aided method for shape recognition and enhancement can occasionally result in confusion in darker or shaded areas, as seen in the case of the fish's tail in this example, overall, the highlighted contour accurately encapsulates the outermost shape of the fish's lateral profile.

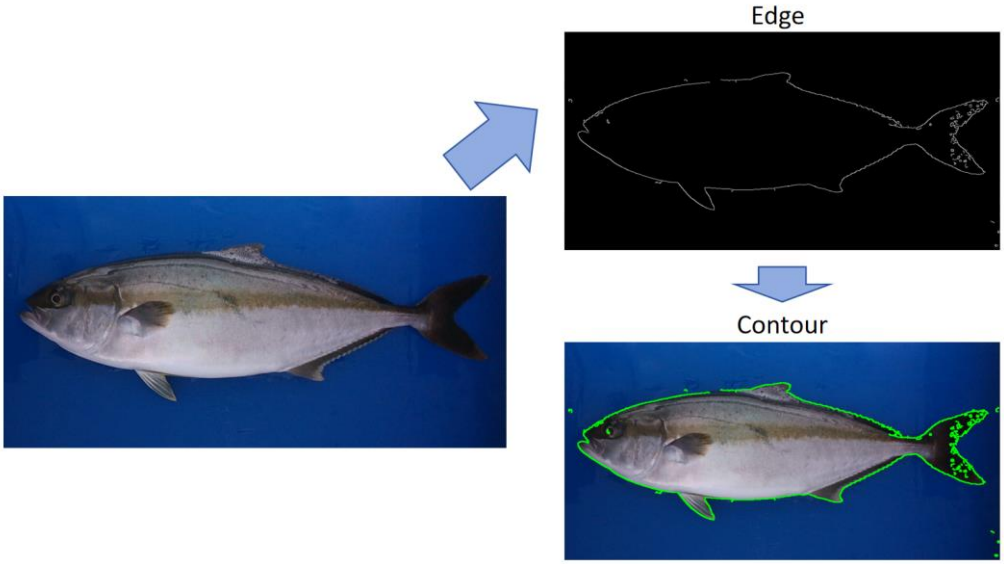

Figure 14.2: Display of the fish shape manual identification.

By extracting the hue channel from the image to transform the light yellow region into a darker one, the enhanced image result of the yellowish band region is displayed in Figure 14.3. The distinct yellowish band on the yellowtail kingfish's body, which is typically discerned from other parts of the lateral profile only by vague boundaries in human recognition, is strongly differentiated from the rest of the body using the hue channel approach. However, it is noteworthy that this image enhancement method also incorporates the pectoral fin into the yellow stripe region.

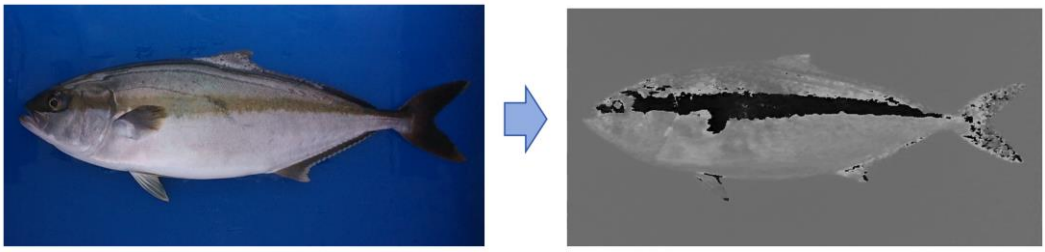

Figure 14.3: Display of the fish yellowish band region manual identification.

3.3.2 Comparison

As shown in Figure 15, the top row showcases four images of two distinct fish at different time points. The fish ID and time point are shown in the top of the Figure. Moving down from the top row, the corresponding image data have undergone enhancement through manual identification and computer vision techniques for three specific features: lateral line, fish shape, and yellowish band region. In this set of images, it is evident to the human eye that the left two columns represent image data from different time points of the same fish, and the same applies to the right two columns.

Notably, a significant distinction arises from the enhancement of the yellowish band region through hue channel manipulation. Through this enhancement, the distinct patterns of the fish's lateral markings become pronounced, revealing the trajectory of its side profile markings. Moreover, the fish's overall contour exhibits variations, with the comparative analysis in Figure 15 highlighting more prominent discrepancies in the curvature of the upper dorsal region and the shape of the dorsal fin. Conversely, differences in the lateral line are subtler, making it easier to miscalculate in shape and specific curvature recognition by human eye.

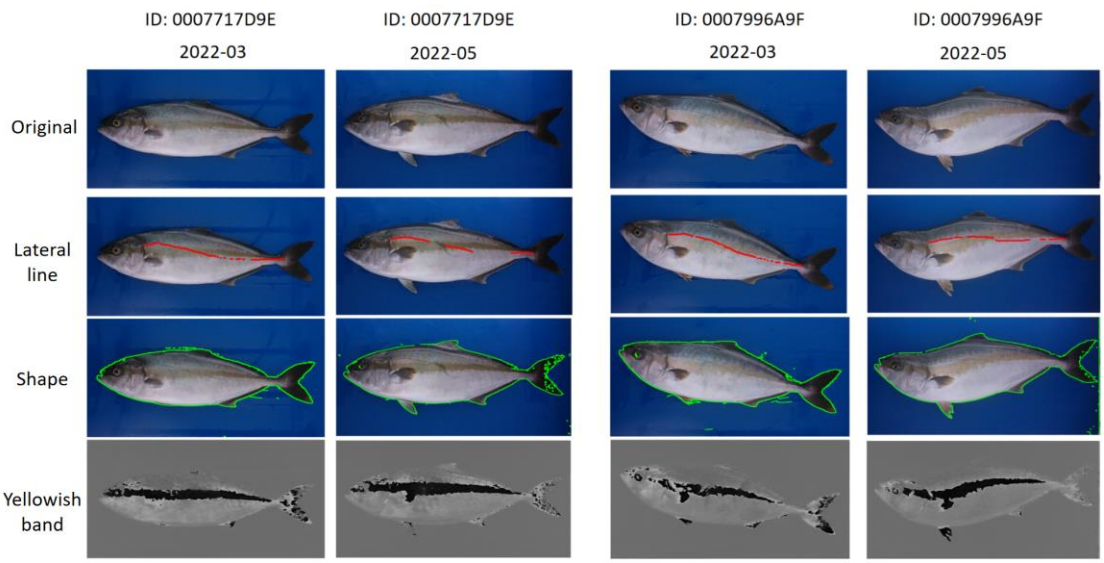

Figure 15: The enhanced features of distinct fish at different time points, highlighting the lateral line, fish shape, and yellowish band region.

4.Discussion

In this project, we investigated the application of state-of-the-art computer vision image recognition methods on yellowtail kingfish. This study represents the first utilization of the pre-trained VGG19 model for individual yellowtail kingfish recognition in a scenario where there is a small size of dataset. Three different whole-fish image recognition models were trained based on the characteristics of the dataset. And the potential of using manual identification to recognize three specific feature regions was explored in cases where the model's results were limited. In this section, we begin by discussing the performance of the three models and the potential impact of time longitudinal data present in the dataset. We then delve into the disparities between the full-fish image recognition models and manual identification, as well as the limitations of the dataset, to explore avenues for experiment refinement and future prospects.

4.1 Short-term recognition

Comparing the two models trained under short-term recognition patterns, it can be initially inferred that the pre-trained VGG-19 model performs better in recognizing individual fish based on the last two time points: April and May. The short-term models trained and validated in February and March data suggest that these models have difficulty to achieve accurate predictions on data from these two time points. However, models constructed using the April and May datasets exhibit better performance in the recognition process compared to those constructed using the February and March datasets.

This discrepancy at least indicates that yellowtail kingfish is better suited for individual recognition image acquisition during the later stages of growth, when their bodily features are more stable. Conducting individual recognition studies based on image data collected during the later growth stages could potentially yield superior results. Alternatively, we could consider more frequent image capture for fish in their early growth stages to reduce the morphological variations between each time point. We hypothesize that this phenomenon might be attributed to the relatively rapid growth rate of yellowtail kingfish during their initial stages, leading to more pronounced changes in their physical characteristics. Consequently, the application of computer vision methods during the early growth phases may not yield optimal performance. This conjecture is further corroborated by findings from growth model studies related to yellowtail kingfish[25].

4.2 Long-term model and Full-time point model

To observe the whole-fish images of yellowtail kingfish at four time points, it reveals both similarities and variations in images of the same individual across different time periods. These disparities encompass variations in body size, stretching and deformation generated by growth in different regions, and shifts in the position of fish scales resulting in differences in body pattern images. These variations are most likely to impact the features extracted by deep learning models from the images, which presents a major challenge for models when dealing with datasets containing time longitudinal aspects. By comparing the learning curves of the long-term model and full-time model, it becomes apparent, that the full-term model exhibits smaller disparities through the comparison of the differences in train loss and validation loss. This phenomenon is attributed, on one hand, to the increased dataset size providing the model with more learning material, and on the other hand, to the incorporation of additional time points in the dataset, thereby enhancing dataset generalization.

4.3 Future research

Irrespective of the specific task, individual fish recognition studies of the same species are significantly constrained by the quantity and availability of training data. Deep models require substantial data, and in our experiment, while we considered hundreds of fish images from four time points, the number of images captures for individual fish at each time point is extremely limited, with only one capture per fish. Given these constraints, employing a transfer learning strategy with pre-trained models is a reasonable approach. However, a limitation of pre-trained models compared to task-specific models is their better performance on smaller datasets with larger variations. Additionally, considering that our project used whole-fish images for re-training rather than specific feature areas, the model may be more inclined to learn similarities in image data, potentially leading to errors during the identification process.

a. Potential of manual identification

By exploring three features during the manual recognition phase, we demonstrated that human eyes can preliminarily differentiate feature variations among individual fish of the same species. Since the principles underlying the resolution of individual recognition problems through computer vision and deep learning models essentially simulate the human visual recognition process, we can consider utilizing the results of manual recognition to identify specific regions within the whole fish image. Alternatively, we can also construct CNN models to identify informative feature regions. Some studies have proposed schemes based on Mask R-CNN for fish image segmentation and measurement of fish morphological features, thus achieving the automatic extraction of fish morphological indicators[35]. Subsequently, CNN models

can be trained on these specific regions, and the predictions from these specialized feature recognition models can be combined to produce the final prediction outcome.

b. Exploration of other models

The considerations of dataset limitations and model performance invite a broader discussion on whether improving the dataset through increased data collection or enhancing model architectures would lead to more significant advancements in individual fish recognition. Further research is needed to assess the balance between these two avenues and identify the optimal approach for enhancing recognition accuracy and robustness.

If the limitation of image dataset will always exist, an alternative strategy based on the performance of the pre-trained VGG-19 model in this project would use other type of model, such as One-shot learning. One-shot learning for computer vision tasks involves a unique type of CNNs known as Siamese neural networks (SNNs)[36]. Siamese neural networks have the capability to learn from a minimal amount of data (only one to two images per class). This model and has been demonstrated to outperform other types of neural networks in terms of speed and accuracy, particularly in identifying images, faces, and other highly similar objects[37]. Moreover, this model can achieve superior generalization performance, especially when dealing with similar but distinct objects[38]. The drawback of this model primarily lies in its increased computational resource consumption when building the model with an equivalent-sized dataset.

5.Conclusion

This project has broadened the horizons of individual identification research for yellowtail kingfish. The outcomes of the project reveal the limitations of the VGG19 model in addressing the specific challenges of recognizing multiple classes (individuals) and low-data fish identification within the same species. Furthermore, we have demonstrated that, for yellowtail kingfish, utilizing images from time points where the individuals exhibit more stable growth and pronounced feature differences can enhance the model's recognition accuracy. We also explored the effectiveness of the specific regions identified through manual recognition, namely the lateral line, fish shape, and yellowish band region, in addressing the individual identification problem.

The experiments conducted in this project come with certain limitations that need to be addressed in future research endeavors. These limitations include the choice of CNN model and the imperfect criteria for assessing the effectiveness of manual recognition. We opted for the pre-trained VGG-19 model due to the scarcity of data, without selecting a model specifically tailored for extremely limited datasets. In summary, this project marks a valuable step towards refining and advancing individual fish identification methods within the same species. Despite the current limitations, the investigation into deep learning models, data augmentation techniques, and innovative data collection methods holds promising avenues for the research on yellowtail kingfish and the future progress of aquaculture industry.

Reference:

1. Beaudouin, R. *et al.* An Individual-Based Model of Zebrafish Population Dynamics Accounting for Energy Dynamics. *PLoS One* **10**, e0125841 (2015).
2. Heupel, M. R. & Simpfendorfer, C. A. Using Acoustic Monitoring to Evaluate MPAs for Shark Nursery Areas: The Importance of Long-term Data. *Mar. Technol. Soc. J.* **39**, 10–18 (2005).
3. Radinger, J. *et al.* Effective monitoring of freshwater fish. *Fish Fish.* **20**, 729–747 (2019).
4. Føre, M. *et al.* Precision fish farming: A new framework to improve production in aquaculture. *Biosyst. Eng.* **173**, 176–193 (2018).
5. Wright, D. W., Stien, L. H., Dempster, T. & Oppedal, F. Differential effects of internal tagging depending on depth treatment in Atlantic salmon: a cautionary tale for aquatic animal tag use. *Curr. Zool.* **65**, 665–673 (2019).
6. Nova J. Silvy, Roel R.Lopez, and M. J. P. Techniques for marking wildlife. 25 (2012).
7. Cooke, S. J. *et al.* To Tag or not to Tag: Animal Welfare, Conservation, and Stakeholder Considerations in Fish Tracking Studies That Use Electronic Tags. *https://doi.org/10.1080/13880292.2013.805075* **16**, 352–374 (2013).
8. Thorstad, E. B., ØKland, F. & Heggberget, T. G. Are long term negative effects from external tags underestimated? Fouling of an externally attached telemetry transmitter. *J. Fish Biol.* **59**, 1092–1094 (2001).
9. Jepsen, N., Thorstad, E. B., Havn, T. & Lucas, M. C. The use of external electronic tags on fish: An evaluation of tag retention and tagging effects. *Anim. Biotelemetry* **3**, 1–23 (2015).
10. He, Q. *et al.* Distinguishing individual red pandas from their faces. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **11858 LNCS**, 714–724 (2019).
11. Shen, W. *et al.* Individual identification of dairy cows based on convolutional neural networks. *Multimed. Tools Appl.* **79**, 14711–14724 (2020).
12. Hirsch, P. E. & Eckmann, R. Individual identification of Eurasian perch Perca fluviatilis by means of their stripe patterns. *Limnologica* **54**, 1–4 (2015).
13. Stien, L. H. *et al.* Consistent melanophore spot patterns allow long-term individual recognition of Atlantic salmon Salmo salar. *J. Fish Biol.* **91**, 1699–1712 (2017).
14. Malik, H. *et al.* Multi-classification deep neural networks for identification of fish species using camera captured images. *PLoS One* **18**, (2023).
15. Rauf, H. T. *et al.* Visual features based automated identification of fish species using deep convolutional neural networks. *Comput. Electron. Agric.* **167**, 105075 (2019).
16. Saleh, A., Sheaves, M. & Rahimi Azghadi, M. Computer vision and deep learning for fish classification in underwater habitats: A survey. *Fish Fish.* **23**, 977–999 (2022).

17. Siddiqui, S. A. *et al.* Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data. *ICES J. Mar. Sci.* **75**, 374–389 (2018).
18. Villon, S. *et al.* A new method to control error rates in automated species identification with deep learning algorithms. *Sci. Rep.* **10**, (2020).
19. dos Santos, A. A. & Gonçalves, W. N. Improving Pantanal fish species recognition through taxonomic ranks in convolutional neural networks. *Ecol. Inform.* **53**, 100977 (2019).
20. Nery, M. *et al.* Determining the Appropriate Feature Set for Fish Classification Tasks Centro de Apoio a Pesquisas Sobre Televisão View project Serious game for old people View project Determining the appropriate feature set for fish classification tasks. doi:10.1109/SIBGRAPI.2005.25.
21. Huntingford, F. A., Borçato, F. L. & Mesquita, F. O. Identifying individual common carp Cyprinus carpio using scale pattern. *J. Fish Biol.* **83**, 1453–1458 (2013).
22. Joo, D. *et al.* Identification of Cichlid Fishes from Lake Malawi Using Computer Vision. *PLoS One* **8**, e77686 (2013).
23. Cisar, P., Bekkozhayeva, D., Movchan, O., Saberioon, M. & Schraml, R. Computer vision based individual fish identification using skin dot pattern. *Sci. Reports 2021 111* **11**, 1–12 (2021).
24. Whatmore, P. *et al.* Genetic parameters for economically important traits in yellowtail kingfish Seriola lalandi. *Aquaculture* **400**–**401**, 77–84 (2013).
25. Donohue, C. G., Partridge, G. J. & Sequeira, A. M. M. Bioenergetic growth model for the yellowtail kingfish (Seriola lalandi). *Aquaculture* **531**, 735884 (2021).
26. Horlick, J., Booth, M. A. & Tetu, S. G. Alternative dietary protein and water temperature influence the skin and gut microbial communities of yellowtail kingfish (Seriola lalandi). *PeerJ* **2020**, e8705 (2020).
27. Aruna, S. K., Deepa, N. & Devi, T. Underwater Fish Identification in Real-Time using Convolutional Neural Network. 586–591 (2023) doi:10.1109/ICICCS56967.2023.10142531.
28. Banerjee, A. *et al.* Deep Learning Based Identification of Three Exotic Carps. *Lect. Notes Networks Syst.* **480 LNNS**, 416–426 (2022).
29. Tian, B. & Wei, W. Research Overview on Edge Detection Algorithms Based on Deep Learning and Image Fusion. *Secur. Commun. Networks* **2022**, (2022).
30. Learning OpenCV: Computer Vision with the OpenCV Library - Gary Bradski, Adrian Kaehler - Google 图书. https://books.google.co.uk/books?hl=zh-CN&lr=&id=seAgiOfu2EIC&oi=fnd&pg=PR3&dq=computer+vision+opencv&ots=hVM2amgATh&sig=fv3NavFLu0S3MF82kWkTu0YlJV8&redir_esc=y#v=onepage&q=computer vision opencv&f=false.
31. Bishop, C. M. Training with Noise is Equivalent to Tikhonov Regularization. *Neural Comput.* **7**, 108–116 (1995).
32. Deng, J. *et al.* ImageNet: A large-scale hierarchical image database. 248–255

(2010) doi:10.1109/CVPR.2009.5206848.
33. Rubinstein, R. Y. & Kroese, D. P. The Cross-Entropy Method. (2004) doi:10.1007/978-1-4757-4321-0.
34. dala Corte, R. B., Moschetta, J. B. & Becker, F. G. Photo-identification as a technique for recognition of individual fish: a test with the freshwater armored catfish *Rineloricaria aequalicuspis* Reis & Cardoso, 2001 (Siluriformes: Loricariidae). *Neotrop. Ichthyol.* **14**, e150074 (2016).
35. Yu, C. *et al.* Segmentation and measurement scheme for fish morphological features based on Mask R-CNN. *Inf. Process. Agric.* **7**, 523–534 (2020).
36. Koch, G., Zemel, R. & Salakhutdinov, R. Siamese Neural Networks for One-shot Image Recognition.
37. Kumar, C. R., N, S., Priyadharshini, M., E, D. G. & M, K. R. Face recognition using CNN and siamese network. *Meas. Sensors* **27**, 100800 (2023).
38. Lee, K. & Lee, E. C. Siamese Architecture-Based 3D DenseNet with Person-Specific Normalization Using Neutral Expression for Spontaneous and Posed Smile Classification. *Sensors (Basel).* **20**, 1–19 (2020).

Appendix:

1. Data availability

Both the raw and processed datasets are stored in the Anunna High Performance Computer (HPC) at the Wageningen University & Research. Access permissions will be available upon request.

2. Code availability

The source code for data processing and model construction is located at ‘https://git.wur.nl/abg_rp/fish_individual_iden_yk’. Access permissions can be provided upon request.

List of Figures

List of Tables