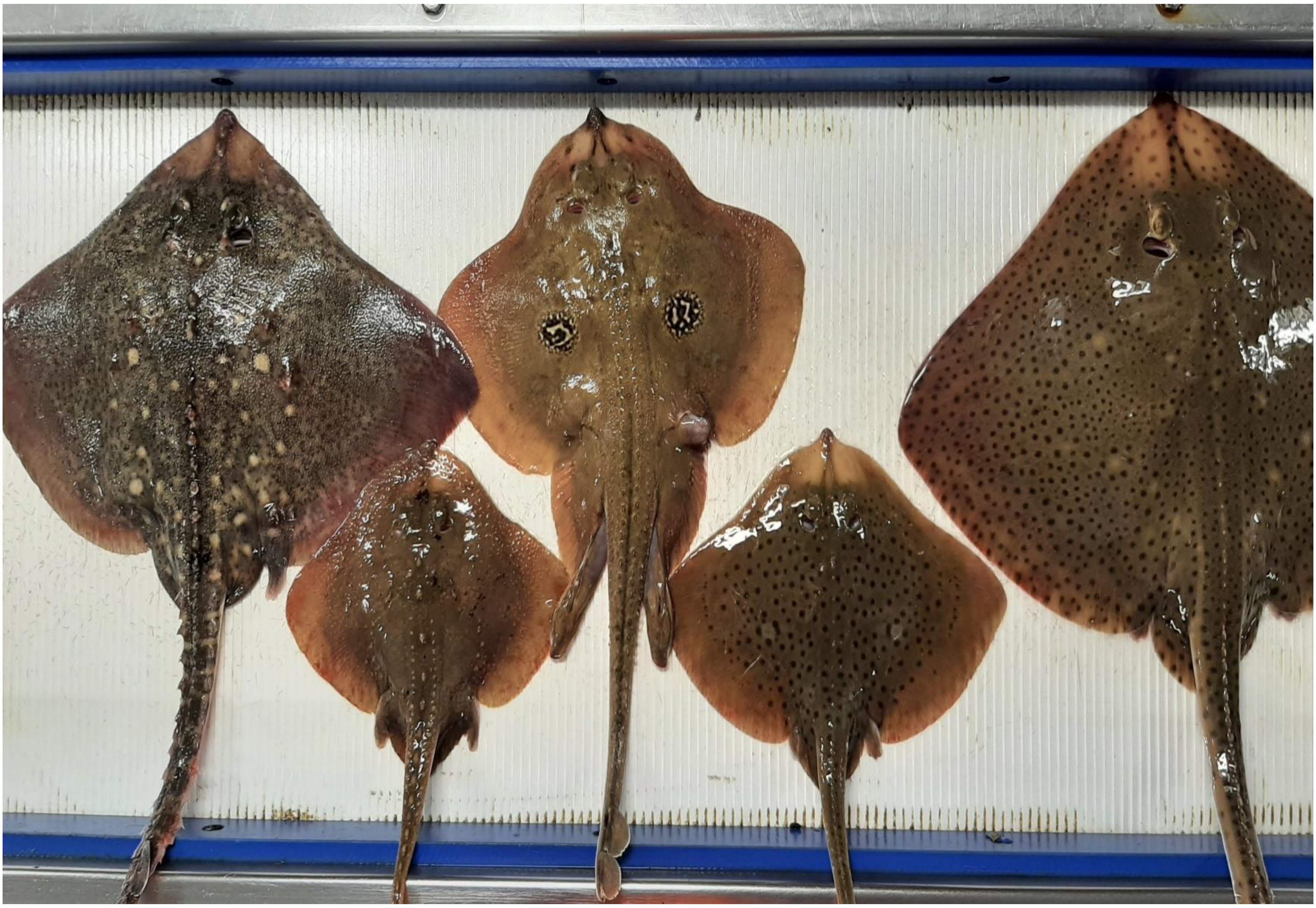

Can Remote Electronic Monitoring be used to monitor catches of skates and rays?

Authors: Jurgen Batsleer, Karolina Molla Gazi, Katinka Bleeker, Tomasz Zawadoski, Edwin van Helmond, Angelo Mencarelli and Gert Kootstra

Wageningen University & Research report C093/22

Can Remote Electronic Monitoring be used to monitor catches of skates and rays?

Authors: Jurgen Batsleer[1], Karolina Molla Gazi[1], Katinka Bleeker[1], Tomasz Zawadoski, Edwin van Helmond[1], Angelo Mencarelli[2] and Gert Kootstra[2]

[1] Wageningen Marine Research

[2] Wageningen Plant Research

Wageningen Marine Research
30 December 2022

Wageningen Marine Research report C093/22

Keywords: Elasmobranchs, Computer vision, discards, stock assessment

Client: Rijksdienst voor Ondernemend Nederland
Postbus 93144
2509 AC Den Haag

This report can be downloaded for free from https://doi.org/10.18174/585311
Wageningen Marine Research provides no printed copies of reports

Wageningen Marine Research is ISO 9001:2015 certified.

Photo cover: Edward Schram

European Maritime and Fisheries Fund

KvK nr. 09098104,
WMR BTW nr. NL 8113.83.696.B16.
Code BIC/SWIFT address: RABONL2U
IBAN code: NL 73 RABO 0373599285

A_4_3_2 V32 (2021)

Contents

Summary

Skates and rays are managed under the European Union total allowable catch (TAC) and quota regulation since 1999. Since its introduction, the TAC has gradually been reduced, and has been constraining landings for many fisheries. Consequently, under the current Common Fisheries Policy (CFP) skates and rays are one of the main "choke species" under the landing obligation. While skates and rays have a temporary high-survivability exemption from the landing obligation, better data collection of catch quantities and composition are needed. Such data will contribute to improve stock assessments and thus improve management of skates and rays in the North Sea.

The OSW 2.1 Innorays project aims to improve the knowledge base for skate and ray stocks in the North Sea. The project is financed from the Science and Fisheries Research Collaboration scheme under the Dutch operational program for the European Maritime and Fisheries Fund (EMFF). Here, the potential of video-based monitoring on board fishing vessels, commonly described as Electronic Monitoring (EM), to estimate catches is evaluated. EM systems allow continuous catch monitoring over extended periods without requiring additional on-board personnel. In addition, EM can provide more representative coverage of the fleet than any other observer programme.

In 2019, two beam trawlers and one twin rigger were equipped with an EM system (Anchor Lab) used for catch recording. The EM system consists of CCTV cameras, winch sensors, a GPS antenna and a 4G-LTE antenna mounted on board a fishing vessel. Catch processing on board the vessel is recorded through five CCTV cameras; one mounted on the conveyor belt and two on each sorting belt. During the entire project, EM systems have been on board for 368 trips of which 218 trips (59%) resulted in valid trips that where thus useable for video analysis. The latter consists of recording counts of the different skate and ray species in the catch on a haul by haul basis.

In the project nine observer trips divided over the three participating vessels where planned. These were carried out by on-board observers appointed by the fishing sector. Observers were obliged to participate and pass a species identification test. In 2019, a species identification test was organised starting with a test followed by a workshop to demonstrate and discuss the distinctive characteristics of the different species. Three of the seven participants passed and were allowed to conduct an observer trip independently.

The observer trips were planned to have a ground truth and allow validation of the counts made by the manual video review versus the number of skates and rays observed by the on-board observers. Outcomes demonstrate there is a significant difference in the numbers counted between the video review and observer, whereby the number of rays counted in the video review is higher. When reviewing footage of the observer trips, the video reviewers are helped by the handling of the rays of the observer. The reviewer can identify when the observer is picking up an individual from the sorting belt and can use the software to pause and replay video footage. In real-time, on board, processing is a continuous process; a ray that is not picked up from the conveyer belt because an observer does not see it or is 'too late' is not counted. As opposed to the observers, the video reviewers could, however, not identify all rays in the catch, especially individuals with the ventral side up to the camera. As observers on board can handle the fish (i.e., turn around to see the dorsal side), species-specific identification is better compared to the video reviewer. Consequently, the percentage of unidentified rays was high for each trip, accounting for more than 50% of the individuals observed in the video footage. In addition, the data of the observer trips were used to estimate the number of hauls to be reviewed to have an accurate estimate of ray catches in a trip. This was estimated to range between 44% and 57% of total hauls, relating to 17 to 23 hauls per trip, that need to be reviewed in order to achieve 80% certainty in catch estimates.

ICES provides single stock advice. In this context, it's required to have species-specific catch data. For rays which could not be identified through video review, the proportion of species composition from

roundfish area 5 and 6 from obtained from the data collected within the Dutch discard self-sampling program was used. Depending on the fishing location, the proportion of species composition of either roundfish area 5 or 6 was used to extrapolate to the unidentified rays within a haul. This assumes that the species-specific distribution of rays within the INNORAYS project is similar to the distribution found within the discards self-sampling program. Most of the unidentified rays are allocated to thornback ray and spotted ray, which were also present in highest numbers within the identified individuals.

This project demonstrates that manual review of EM data requires human observations which is error prone, labour intensive and results in relatively high costs. These are limiting factors to implement the system on a broader scale in commercial fisheries. Here, the technical feasibility of automated image recognition as a solution to fully automatically record the number and species of rays present in the (by)catch has been examined. Images of three ray and four flatfish species were collected with three levels of complexity in composition. Outcomes indicate that a computer can make reliable judgements on detecting and identifying a species, especially with a low complexity, but slowly decreasing when the composition of fish become more complex. Yet, even in the most complex situation, i.e. rays being occluded, the performance of the network is good. These outcomes demonstrate that computer vision technology may contribute to increase monitoring coverage of fishing activities and may ease registration of catches on-board. The development of computer vision technology in Dutch demersal fisheries is currently ongoing within the EMFF funded ‘Fully Documented Fisheries” project.

To conclude, the use of video-based monitoring on-board fishing vessels is a way to improve our knowledge on catches in commercial fisheries. EM could be a tool leading to better registration of catches and thus estimates of fishing mortality on a stock. Consequently, improved data could allow an increase of quota and reduce the risk of these species being a “choke species” under the landing obligation. Especially, the automated registration of catches by species, which was explored in this project, may contribute to the accuracy of catch estimates for ”data-limited stocks”.

Samenvatting

Roggen worden sinds 1999 beheerd onder de TAC en quota verordening van de Europese Unie. Sinds de invoering, is de TAC systematisch verlaagd, waardoor de aanvoer voor veel visserijen beperkt is geworden. In het kader van het huidige gemeenschappelijk visserijbeleid (GVB) zijn roggen dan ook een van de belangrijkste "choke species" onder de aanlandplicht. Hoewel roggen een tijdelijke uitzondering op basis van "hoge' overleving hebben, is een betere dataverzameling over de omvang en de samenstelling van de vangsten nodig.

Voor beter beheer van de roggen populaties in de Noordzee zijn betere gegevens voor de bestandsschattingen nodig. Het OSW 2.1 INNORAYS project heeft als doel de kennis- en gegevensbasis voor roggen in de Noordzee te verbeteren. Het project is gefinancierd uit de regeling Samenwerkingsprojecten Wetenschap en Visserij in het kader van het Nederlandse operationele programma voor het Europees Fonds voor Maritieme Zaken en Visserij.

In deze studie is gekeken naar de inzet van 'Electronic Monitoring' (EM) aan boord van visserschepen om de schattingen van roggenvangsten in de visserij te verbeteren. EM kan de ruimtelijk en periodieke dekking van een monitoringsprogramma aanzienlijk vergroten zonder dat daar hogere kosten aan verbonden zijn of extra personeel aan boord voor nodig is en is dus mogelijk effectiever dan enig ander monitoringsprogramma.

Twee boomkor kotters en één twin rigger werden uitgerust met een EM-systeem van Anchor Lab in 2019. Het systeem wordt enkel gebruikt voor vangstregistratie van roggen. Het EM-systeem bestaat uit CCTV-camera's, liersensoren, een GPS-antenne en een 4G-LTE-antenne die aan boord van een vissersvaartuig gemonteerd zijn. De vangstverwerking aan boord van het vaartuig wordt geregistreerd via vijf CCTV-camera's. Eén camera is op de transportband gericht en twee cameras zijn boven elke sorteerband geinstalleerd zodat het hele verwerkingsproces in beeld gebracht wordt. Gedurende de looptijd van het project zijn de EM-systemen tijdens 368 reizen in werking geweest, waarvan er voor 218 reizen (59%) bruikbaar beeldmateriaal beschikbaar waren voor videoanalyse. De videoanalyse zelf omvat het op trekniveau registreren en tellen van de verschillende soorten rog in de vangst.

Voor het project waren negen waarnemersreizen gepland, verdeeld over de drie deelnemende schepen. Tijdens de waarnemersreizen werden per trek alle roggen gepakt en op soort geidentificeerd. Deze reizen zijn belangrijk om een 'ground truth' te verkrijgen en de validatie van de telling uit de videoanalyses mogelijk te maken. Waarnemersreizen werden uitgevoerd door sectoropstappers. Opstappers waren verplicht deel te nemen aan en te slagen voor een soort-identificatie toets georganiseerd door WMR in 2019. De toets omvatte een test, gevolgd door een workshop om de onderscheidende kenmerken van de verschillende soorten te bespreken. In totaal slaagden drie van de zeven deelnemers waardoor zij zelfstandig een waarnemersreis uit mochten voeren.

Resultaten van de vergelijking tussen validatiereizen en video-review tonen aan dat er een significant verschil is in de tellingen tussen de waarnemersreizen en videoreviewers, waarbij het aantal roggen geobserveerd in de videoreview hoger is. Bij het bekijken van de videobeelden van de waarnemersreizen worden de videoreviewers geholpen door handelingen van de waarnemer aan boord. De videoreviewer ziet wanneer de waarnemer een individu van de sorteerband oppakt, daarnaast kan de reviewer videobeelden pauzeren en terugspoelen Aan boord gaat het verwerkingsproces gewoon door, een rog die niet van de sorteerband gepakt wordt omdat deze niet wordt opgemerkt of omdat de opstapper 'te laat' is, wordt dan ook niet geregistreerd. De videoreviewers zijn daarentegen niet in staat alle roggen in de vangst op soort te identificeren. Dit is met name het geval wanneer individele roggen met de witte buikzijde naar boven en dus de camera gericht zijn. Het percentage niet op soort gebrachte roggen bedroeg meer dan 50% van de op de videobeelden waargenomen individuen. In tegenstelling tot de videoreviewers kunnen de waarnemers aan boord de roggen omdraaien, bekijken, en op soort identificeren. Naast de validatie van de videoreview werden de gegevens van de waarnemersreizen gebruikt om een schatting te maken van het aantal te reviewen trekken per visreis om zodoende een nauwkeurige schatting te maken van de

totale hoeveelheid roggenvangsten. Om met 80% zekerheid de vangsten van rog te schatten, moeten ongeveer 44% tot 57% van de totale aantal trekken, d.w.z. 17 tot 23 trekken per reis geanalyseerd worden.

ICES geeft vangstadviezen voor individuele roggenbestanden. Om de gegevensbasis voor bestandsschattingen te verbeteren is het dan ook vereist om soortspecifieke vangstgegevens te verkrijgen. Wanneer een rog niet op soort geidentificeerd kon worden in de videoanalyse, werd een soort toegekend op basis van de bekende soortensamenstelling uit rondvisgebied 5 en 6. Hiervoor zijn gegevens verzameld binnen het Nederlandse discard zelfbemonsteringsprogramma gebruikt. Afhankelijk van de vislocatie werd het aandeel van de soortensamenstelling van rondvisgebied 5 of 6 gebruikt om te extrapoleren naar de niet-geïdentificeerde roggen binnen een trek. Hierbij is de aanname dat de soortspecifieke verdeling van roggen binnen een reis in het INNORAYS-project vergelijkbaar is met de verdeling waargenomen in het discard zelfbemonsteringsprogramma. De meeste ongeïdentificeerde roggen werden toegewezen aan stekelrog en gevlekte rog.

Dit project heeft aangetoond dat het handmatig uitvoeren van videoanalyses van de EM-beelden foutgevoelig en arbeidsintensief is en relatief hoge kosten met zich meebrengt. Dit zijn beperkende factoren om het systeem op grotere schaal toe te kunnen passen in de commerciële visserij. In het project is een technische haalbaarheidsstudie naar automatische beeldherkenning opgezet. Automatische beeldherkenning zou een oplossing kunnen bieden om roggenvangsten per soort te registreren. Uit de resultaten blijkt dat een computer roggen kan detecteren en accuraat op soort kan identificeren. De betrouwbaarheid van automatische beeldherkenning is erg hoog bij een lage complexiteit (als de rog duidelijk zichtbaar is). De betrouwbaarheid neemt geleidelijk af naarmate vissen elkaar overlappen de de samenstelling dus complexer wordt. Toch zijn de prestaties van automatische beeldherkenning in de meest complexe situatie goed. Deze resultaten tonen aan dat deze technologie bij kan dragen tot een betere monitoring van visserijactiviteiten en registratie van vangsten aan boord.

Concluderend, het gebruik van EM aan boord van vissersschepen kan bijdragen om kennis over de roggenvangsten in de commerciële visserij te verbeteren. Verdere ontwikkeling van het EM-systeem leidt mogelijk tot een betere registratie van de vangsten, waardoor er meer inzicht komt in de visserij-effecten wat vervolgens ook de bestandsschattingen kan beinvloeden. Een betere gegevensbasis voor belangrijke commerciele soorten zoals stekelrog zou mogelijk tot een verhoging van de quota kunnen leiden en het risico dat deze soorten "choke species" worden in het kader van de aanlandingsverplichting verminderen. Met name de verdere ontwikkeling van automatische beeldherkenning van vangsten per soort kan bijdragen tot nauwkeurigere vangstschattingen voor datagelimiteerde bestanden. Op dit moment vindt de verdere ontwikkeling van automatische beeldherkenning in de Nederlandse demersale visserij plaats in het door het EFMZV gefinancierde project "Fully Documented Fisheries " (FDF).

1 Introduction

Sharks, skates, and rays (elasmobranchs) are characterized by specific biological traits including being long-lived, showing slow growth, late sexual maturity and producing a small number of young per year. These traits make elasmobranchs vulnerable to fishing, pollution, and changes in essential habitats, especially spawning and nursery areas (Stevens *et al.*, 2000, Schindler *et al.*, 2002, Heessen, 2010). Worldwide, several populations have undergone sharp declines under the influence of anthropogenic activities such as fishing, large-scale coastal infrastructure, and pollution (Brander, 1981, Walker and Heessen, 1996, Dulvy *et al.*, 2008, Dulvy *et al.*, 2014, Sguotti *et al.* 2016). While data from scientific survey programs show an increase in populations for several species in European waters since 2010 (ICES, 2021), there is still much concern that current assessments and management do not offer adequate protection for elasmobranchs (STECF, 2017).

In the North Sea, most elasmobranch stocks are classified as category 3 stocks, for which ICES advice is based on an indicative trend from available survey data. Scientific surveys such as the International Bottom Trawl Survey (IBTS) and Beam Trawl Surveys (BTS) are the primary source of fisheries-independent data for elasmobranch stock assessment. These surveys, however, were initiated primarily to estimate the recruitment of the main exploited stocks and were not primarily designed to inform on the populations of demersal elasmobranchs. Hence, gears used, timing of the surveys and distribution of sampling stations may not be optimal for informing on elasmobranch species and/or life-history stages. This is problematic and impedes the use of analytical stock assessments for these stocks.

Analytical stock assessments generally rely on population models integrating biological, survey and fisheries data including fishing mortality and catch estimates (Beverton and Holt, 1957; Punt *et al.*, 2006). In this context, accurate catch estimates are important to achieve sustainable fisheries, especially for sensitive species such as skates and ray. Unfortunately, good catch estimates are lacking for most skates and ray stocks. More specifically, data on the discarded part of the catch is uncertain due to species misidentification, insufficient sampling effort, variable raising factors, varying discard retention patterns, and high expected discard survival (ICES, 2021). While much effort has been done to improve discard estimations (ICES 2017, 2020), several Member States still report incomplete data which is not species specific affecting the catch statistics provided to ICES. As such, data from national discard programs has, to date, mostly been used in exploratory and descriptive analyses.

In 1999, the European Commission introduced a combined Total Allowable Catch (TAC) for skates and rays, in the North Sea, meaning several species are managed by a single TAC. Since its introduction, the TAC has gradually been reduced, and landings of skates and rays in the North Sea have been at or above the TAC since 2006. To keep landings within the national quota, Dutch Producer Organisations have implemented landing restrictions including a minimum landing size of 55cm total length and trip limits to control quota uptake Furthermore, the constraining quota for skates and rays make them one of the main “choke species” under the landing obligation, meaning the fisheries for flatfish is halted in case all skate and ray must be landed and the ray quota has been fully exhausted. Skates and rays are exempt from the landing obligation based on high survival up until 2023. Yet, improved management of skates and rays in the North Sea is urgent and can be supported by developing better data collection of catch quantities and composition.

One way to improve our knowledge on catch in the Dutch fleet, is through the use of video-based monitoring on board fishing vessels. This is commonly described as Electronic Monitoring (EM), also described as Remote Electronic Monitoring (REM), allowing catches to be observed remotely by human experts without requiring additional on-board personnel (McElderry *et al.*, 2003; Kindt-Larsen *et al.*, 2011; Stanley *et al.*, 2015; Hold *et al.*, 2015; van Helmond *et al.*, 2017). EM systems enable continuous catch monitoring over extended periods, making them more suitable for monitoring discards from commercial fishing vessels than human on-board observers; they can provide more representative coverage of the fleet than any other observer programme (van Helmond *et al.*, 2020). Hence, this part of the INNORAYS project aims to explore the potential of EM to monitor skate and ray catches in the Dutch demersal fisheries. Because analysis of EM video data requires human observations, costs are still relatively high, which, together with the number of human resources needed, is a limiting factor in the uptake of EM (Needle *et al.*, 2015; Mortensen *et al.*, 2017). To reduce the workload and improve the sampling frequency, a reform of EM data processing is necessary. As such the project also explored the potential of automated image recognition to improve accuracy in catch estimates on board.

The project covered a four-year period and was funded by the Science and Fisheries Cooperation Projects Scheme (*Partnerschappen Wetenschap en Visserij*) under the Dutch Operational Program for the European Maritime and Fisheries Fund (EMFF).

2 Assignment

2.1 Aim

The objective of the project is to improve the quantity and quality of data for the Data Limited Stocks (DLS) of skates and rays in the North Sea by using on-board Electronic Monitoring (EM) and by applying an innovative genetic tool (close-kin Mark recapture) to estimate population structure and size. Improved catch estimates from EM and knowledge on the population size (through DNA-analysis) can be used to estimate the effects of fishing on skate and rays in the North Sea. Hence, the project will contribute to better advice on fishing opportunities and reducing fishing impacts on skates and rays as an important species in the North Sea ecosystem.

This report focuses on applying EM to evaluate observations of skates and rays in the catch of Dutch demersal fisheries. Working with EM has the advantage that it gives a larger spatial and temporal distribution of catch data allowing to improve our knowledge on the catch composition and estimation of catch quantities in the context of sustainable management of the North Sea. INNORAYS also explores potential of using these techniques (if successful) to additionally improve data collection and quality of other North Sea DLS fish stocks and links to the development of Fully Documented Fisheries (FDF) as an instrument in the implementation of the Common Fisheries Policy.

2.2 Work packages

The main question in the project was to explore whether Electronic Monitoring can be used to estimate ray catches in the Dutch demersal fisheries. Improved catch data will contribute to better advice on fishing opportunities and thereby reduce the fishing impact on skates and rays, important species in the North Sea ecosystem. Three work packages were defined to address the question.

Work package A focussed on the ground truth for manual video review by planning observer trips which are required to validate counts made by video reviewers. Work package A consisted of:

1. Improving species identification by organising a species-identification workshop with participating skippers, industry observers and video reviewers.
2. Recruiting vessels and installation of hard- and software on-board.
3. Conducting 10 observer trips on-board participating vessels, required to have allow comparison of on-board and video footage observations (i.e., the ground truth).

Work package B focused on the implementation of EM systems and validating the effectiveness of manual video review to estimate ray catches. Work package B consisted of:

1. Validation of EM by comparing catch estimates from manual video review with observed catches in the observer trips.
2. Data analysis to develop a method to accurately estimate ray catches using video monitoring.
3. Reviewing video footage of participating vessels and apply method to estimate ray catches within a trip.

Work package C is focussing on the development of computer vision technology. Work package C includes:

1. Performing a pilot on the technical feasibility of automated computer vision for skates and rays.

3 Materials and Methods

3.1 Species identification workshop

Because the project aims to obtain a better understanding of the catch composition and catch estimates by species, species identification is a critical component for the quality assurance of the data collection within the INNORAYS project. More specifically, video reviewers, skippers and crew as well as on-board observers are required to accurately identify the different species and thus must be tested on their species identification skills.

In 2019 Wageningen Marine Research (WMR) organised an elasmobranch species identification test and workshop. The workshop was compulsory for WMR employees participating in fish surveys on board of research vessels, sampling of landings and discards from commercial fishing vessels and those responsible for the review of video footage. In addition, externally hired personnel including skippers, crew and on-board observers appointed by the sector parties were obliged to participate and pass the test. Because the initial score of the industry observers was low, a second test was organised in April. This allowed them to absorb the information of the workshop and gave additional time to increase their species-identification skills.

The identification workshop was split in two parts, starting with a test followed by a workshop to demonstrate and discuss the distinctive characteristics of the different species. In total, ten elasmobranch species were used in the test, in total 20 specimens. When species were identified correctly to the lowest taxonomic level, 1 point was assigned. Wrong identification or empty fields were scored as 0. When the main identification criteria for two similar species were put on the list, this was scored as 0.5. Ambiguous or incomplete naming of species (e.g., 'starry smooth hound' instead of '*Mustelus* sp.') was scored as 0. During the test it was not allowed to use any reference material for species identification. Participants were encouraged to mention on their forms distinctive identification criteria when being in doubt between two species, as a measure for their knowledge of distinctive species characteristics. For the project, video reviewers of WMR and observers appointed by the sector parties had a limit on species identification set to >80% before being allowed to identify species independently and join the observer trips.

3.2 Vessel monitoring

In Dutch fisheries beam trawlers and twin rig fisheries targeting flatfish are responsible for respectively 87.6 % and 3.8 % of Dutch skate and ray catches. Initially, the aim was to involve four vessels in the project, but due to developments in European fisheries policy and the apprehensiveness of the fishing industry towards on-board camera systems(i.e., for control and enforcement purposes) only three vessels were found willing to participate on a voluntary basis. Two of these vessels are beam trawlers and one is a twin rigger.

In 2019, the three vessels were equipped with an electronic monitoring system (BlackBox VX, Anchor Lab) used for catch recording. The BlackBox is a system consisting of CCTV cameras, winch sensors, a GPS antenna and a 4G-LTE antenna mounted on board a fishing vessel. It also uses a software programme to analyse the data recorded on board. The winch sensors and GPS antenna on board the vessel provides reliable positioning recording and time synchronization from the location of fishing hauls. Catch processing on board the vessel is recorded through five CCTV cameras; one mounted on the conveyor belt and two on each sorting belt. The cameras on the sorting belt only show what is on the belt, due to privacy reasons other areas of the image have been automatically blacked out on the recording and is therefore not recorded (figure 3.1).

Figure 3.1: *Screenshot of the Blackbox software. At the top, vessel speed is depicted by the blue line which provides an indication of the activity (setting or hauling the net or fishing). The bottom figure shows the video footage of the 4 cameras installed above the sorting belt.*

3.2.1 Manual review

Video images are stored on board and sent through WIFI to a database ashore at the end of the trip when the vessel enters the 4G network range. While the video images are owned by the vessel owners, the database was managed by VisNed. WMR was given access to the images for quality control and analysis. Access was under strict conditions, which included a restriction of data use to the project itself and that video images gathered within the project may only be used for the implementation of the project itself and that only trained researchers can access the images. Furthermore, review was only to be done in a locked room, and images were removed immediately after video analysis.

The video analysis consists of recording the catch within a haul. Images taken by the CCTV cameras on board the vessel were manually reviewed using the BlackBox software programme. The species observed on the camera images were recorded on a datasheet including information on counts by species on a haul by haul basis.

3.2.2 Observer trips

A total of 9 observer trips (3 in 2019 and 5 in 2021) were caried out. Due to COVID-19 one of the planned observer trips was not done. Furthermore, one trip had a duration of 10 days and was split to count this trip as two separate observer trips. An observer appointed by the fishing sector recorded the species and number of individuals in all hauls within an observer trip. During these trips, also video footage from the sorting belt was routinely collected via the EM system for all the hauls within a trip. The camera images are viewed on land by experienced observers who make estimates from the video footage (see 3.2.1 manual review). The video footage from the REM system is validated by the data collected on board by the observer. Furthermore, the results were used to estimate how many hauls within a trip must be video reviewed in order to effectively determine the catch composition and quantities. A bootstrap approach was applied with a 20% confidence interval to calculate the number of hauls to be video reviewed. The number of hauls to be reviewed in order to achieve 80% certainty in number of rays caught varied between vessels and trips, ranging from 44% to 57% of total hauls. This relates to 17 to 23 hauls per trip.

3.3 Catch composition and catch estimation

The effectiveness of the manual video review to estimate ray catches was first validated by comparing those hauls in which ray catches were also recorded by an on-board observer, which is not covered in the current section. This section only covers trips for which no observer was present and only video footage was analysed. It was decided for each vessel to randomly select and review two trips by quarter for 2021. For every selected trip the camera images of the sorting process were manually reviewed.

In order to get species-specific data from video reviews, it is necessary that rays, and their specific morphological traits needed to identify a species, are visible on the conveyer belt. For rays that cannot be identified, data collected within the Dutch discard self-sampling program (Bleeker *et al.*, 2022), enforced through the Data Collection Framework (DCF) of the European Commission, is used. The spatial distribution of five ray species; thornback ray, spotted ray, blonde ray, cuckoo ray and starry ray within the discard self-sampling between 2018 and 2021 was used to calculate the proportion present of each species (Figure 3.2). To incorporate a potential spatial effect of the distribution of species, it was chosen to use ICES Roundfish areas as cut off point. We looked at the proportion of species between Roundfish area 5 and 6 as the spatial distribution of these areas overlap with the spatial distribution of the trips in which the REM system was used. There was a minor difference in proportions between the areas. For each haul in which unidentified rays were found, the proportion of species composition, of either area 5 or 6, was used to extrapolate to the unidentified rays proportion. This assumes that the species-specific distribution of rays within the INNORAYS project is similar to the distribution found within the discards self-sampling program.

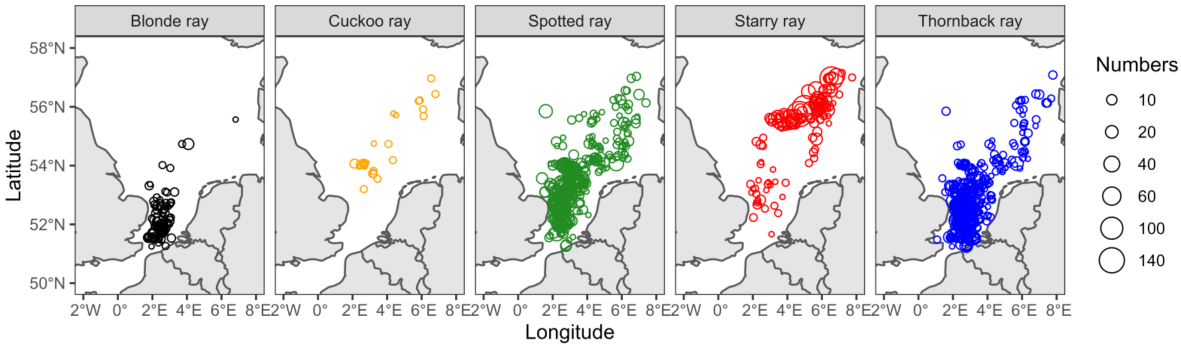

Figure 3.2: *Species-specific spatial distribution of rays collected in the Dutch discard self-sampling program for years 2018-2021 (Bleeker et al., 2022).*

Table 3.1: *mean weight (g) for the five main ray species based on data from the Dutch beam trawl survey (2020)*

Species	Mean weight
Thornback ray	**435.5**
Spotted ray	**275.2**
Blonde ray	**756.6**
Cuckoo ray	**291.4**
Starry ray	**239.1**

Total number of individuals for each species was raised from sampled hauls to trip level using the ratio between the total duration of the trip to the sampled duration. These ratio's where between 1.1 and

2.9 for all trips. A mean weight for each species (table 3.1), calculated based on data from the Dutch beam trawl survey (2020) was used for total catch estimation for each species and trip.

3.4 Automated image recognition

Manual review of EM data requires human observations which is labour intensive and results in relatively high costs. These are limiting factors to implement the system on a broader scale. To reduce the workload and improve sampling frequency and accuracy, a reform of EM data processing using automated image recognition is necessary.

Training a computer to recognize fish during the sorting process on board a fishing vessel is challenging, due to variability in fish appearance. For example, fish from the same species are not identical in size, colour, and patterning, while different ray species also share similarities, e.g., all species have a white ventral side. Also, catch is often loaded in bulk on a sorting belt for processing on board, and consequently, fish are randomly positioned on the belt, often overlapping each other, resulting in severe occlusions. The monitoring system needs to deal with partly visible specimens and must be able to identify multiple individuals and species in a single image frame.

Here, the technical feasibility of automated image recognition as a solution to fully automatically record the number and species of rays present in the (by)catch has been examined. To do so "computer vision" and "deep learning" techniques were used to train a neural network that can automatically recognise three ray species (blonde ray, *Raja brachyura*; spotted ray, *Raja montagui*; and thornback ray, *Raja clavata*) in images, and keep track of the number per species.

Training a system to recognise ray species in catches on board of fishing vessels in demersal mixed fisheries requires representative image data. Image data was created by placing fish on a blue background in a lighting box with LED strips on the topside. A camera (Nikon D5300) was mounted above the fish, capturing images with a resolution of 6000 x 4000. Composition of collected images was comparable to the top view of a conveyor belt during the sorting process on a fishing vessel. Since the fish catch consists not only of ray species, image data were also collected of other demersal fish species. Specimens of seven fish species were collected. Three were ray species: thornback ray, spotted ray, and blonde ray; and four flatfish species, plaice, dab, turbot, and sole.

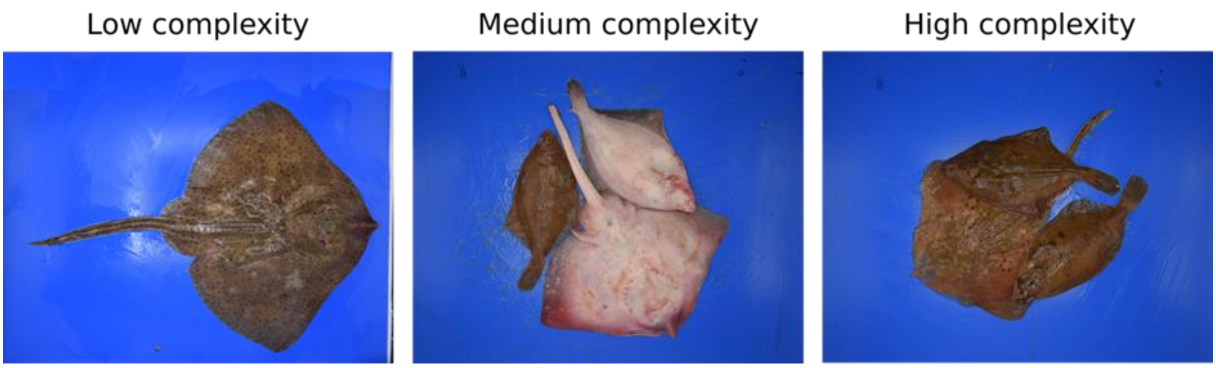

Figure 3.3 *Sample images of possible compositions that could occur on a conveyor belt. From left to right: a single fish (low complexity), multiple fish lying adjacent to each other (medium complexity), or fish overlapping each other (high complexity).*

To mimic the sorting process on a conveyor belt, images were collected with different compositions. The possible compositions were defined in three levels of complexity (Figure 3.3). At low complexity, images contain single fish completely visible (1108 images). The view of both dorsal and ventral sides of the fish were photographed. At medium complexity, images contain multiple fish lying close and adjacent to each other (253 images with 871 fish). Lastly, at high complexity, images contain multiple overlapping fish (229 images with 834 fish). In medium and high complexity, the combinations of fish were randomly selected with the only requirement that at least two fish are in the frame. Also, the dorsal and ventral view of each fish in the image was randomly selected. In all situations, the fish were randomly positioned. The total data set consists of 1591 images containing 2813 individuals of seven species.

Deep neural networks allow the end-to-end processing of data, meaning that the raw images are taken as input, which are processed by the network to provide the required output. This study uses YOLO version 3 (Redmon and Farhadi, 2018) as a tool, providing bounding boxes of the objects in the image as output, together with the corresponding object classes. Additionally, the network provides a confidence score for each of the object detections.

Training a YOLO network requires a dataset in which every object of interest in the images is annotated. An annotation includes the coordinate and size of the object's bounding box and the corresponding class, which in our case is the name of the fish species. Examples of annotations can be observed in Figure 3.4. The performance of the models has been evaluated using the precision, and recall. *Precision* represents the proportion of correct detections over all detections by the network (Eq. 1) and *recall* represents the proportion of correct detection that could be retrieved out of all actual detections (Eq. 2). Furthermore, a confusion matrix was made to summarise the number of correct and incorrect classified fishes. A more detailed description of the method can be found in van Essen *et al.* 2021.

$$Precision = \frac{TP}{TP+FP} \qquad \text{(Eq. 1)}$$

$$Recall = \frac{TP}{TP+FN} \qquad \text{(Eq. 2)}$$

When the actual fish is correctly detected, the detection is defined as a true positive (*TP*). All fish in images that are not detected are labelled as false negatives (*FN*), and all detections that do not correspond with a fish or have the wrong class are labelled as false positives (*FP*). Full description of the method can be found in Annex 1.

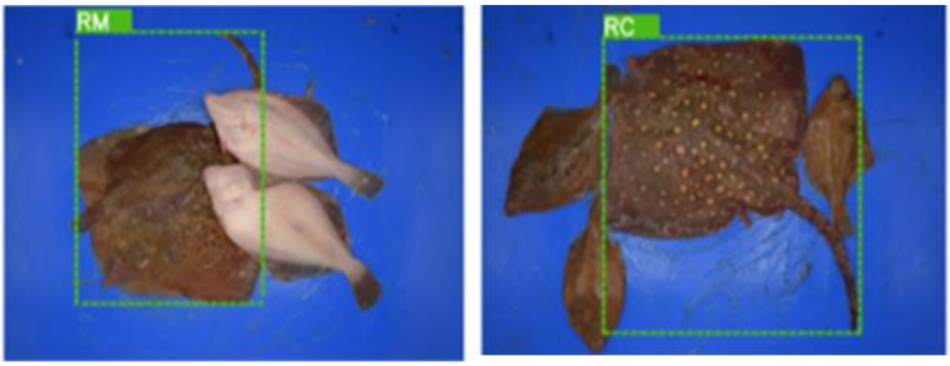

Figure 3.4: *Annotation example. Bounding boxes are drawn around the object of interest for each detection task. RM= spotted ray, RC=thornback*

4 Results

4.1 Species identification workshop

Two species-identification workshops were organised in 2019. The maximum possible score in both workshops was 20 (the number of specimens to identify). The scores of the first workshop showed that the industry observers' species knowledge was insufficient to allow them to carry out the work independently (Table 4.1). During this workshop, much attention was given to clarifying the species-specific characteristics. In addition, participants were referred to available literature to aid them in memorizing the different characteristic by species.

The second workshop was held in April 2019 to re-evaluate species identification skills of the industry observers. The outcomes showed that three of the five observers are able to identify elasmobranch species independently and were therefore allowed to conduct an observer trip independently (Table 4.2). Two participants from the first workshop were unable to attend.

Table 4.1: *Outcomes of the elasmobranch identification workshop in February 2019. Percentage = % good, score = number good, Level = level against which tested (3=ability to independently identify species as used by WMR), Level_2 and level_3 contain the threshold values for the level scores (60% and 80% respectively), The column passed indicates whether someone meets the required level. Names of candidates have been anonymised for privacy reasons.*

Percentage	Score	Participant	Level	Level_2	Level_3	Passed
15	3	A	3	12	16	N
25	5	B	3	12	16	N
10	2	C	3	12	16	N
20	4	D	3	12	16	N
25	5	E	3	12	16	N
30	6	F	3	12	16	N
15	3	G	3	12	16	N

Table 4.2: *Outcomes of the elasmobranch identification workshop in April 2019. Percentage = % good, score = number good, Level = level against which tested (3=ability to independently identify species as used by WMR), Level_2 and level_3 contain the threshold values for the level scores (60% and 80% respectively), The column passed indicates whether someone meets the required level. Names of candidates have been anonymized for privacy reasons.*

Percentage	Score	Participant	Level	Level_2	Level_3	Passed
70	14	A	3	12	16	N
90	18	B	3	12	16	Y
95	19	D	3	12	16	Y
100	20	E	3	12	16	Y
55	11	G	3	12	16	N

4.2 Vessel monitoring

In March 2019, two vessels were equipped with an electronic monitoring system. The third vessel could only be equipped with the EM system in July 2019. During the entire project, EM systems have been on board for 368 trips of which 218 trips (59%) resulted in valid trips that where thus useable for further analysis. Video reviewers of WMR regularly checked the quality of the video footage and noticed differences in quality by vessel. While vessel two has 101 valid trips (79%), vessels 1 and 3 have 75 (58%) and 42 valid trips (37%), respectively (Table 4.3). The small number of valid trips in vessel 3 was caused by technical failures of the EM system on board. Unfortunately, due to covid, there was a long delay before the system could be repaired.

Beyond technical issues, the quality of the video footage is a key factor to determine their usability for analysis. The quality of the available video material depends on several factors in which the maintenance of the system by the crew is crucial. The lenses from the onboard cameras focused on the sorting belt need to be cleaned on a regular basis. Hauls of which the image quality is low due to e.g., drops or other dirt on the lens cannot be used for further analysis because this prevents proper detection and identification of the fish on the sorting belt. Furthermore, technical issues may occur such as video footage being displayed in black and white, making species identification difficult, or video footage stopping to record in the sorting process preventing the video reviewer to analyse the full haul. Finally, factors like fishing grounds and catch volume are also important aspects allowing reviewers to correctly analyse the video footage. For instance, it can be difficult to identify a species when the catch is covered in mud or peat or when the rays are occluded by other fish or benthos.

Despite checks on quality, the project did not entirely succeed in properly engaging the participating vessels. Meetings in both 2020 and early 2021 with skippers and crew have been postponed due to COVID-19. This was not ideal as contact about their input, functioning and output of the project was of upmost importance to continue to have support and contributions to the project.

Table 4.3: *By vessel, the gear (TBB = beam trawl, OTB = twin rig), period of having EM system on board, Total number of weeks with EM system and valid weeks.*

Vessel	Gear	Period	Total weeks (Number)	Usable weeks (Number)
V1	OTB	March 2019 – October 2021	128	75
V2	TBB	March 2019 – October 2021	128	101
V3	TBB	July 2019 – October 2021	112	42

Table 4.4: *number of hauls and rays observed for both video reviewer and observer for each vessel en trip*

Vessel	Trip ID	Number of hauls	Number of rays observed	
			Video review	Observer
V1	1	27	203	189
V2	1	22	2691	1733
V2	2	30	509	435
V2	3	27	1686	1692
V3	1	46	440	467
V3	2	28	648	694
V3	3	27	826	876
V3	4	28	2125	694

4.2.1 Paired observer – video review trips

The number and species of rays caught per haul were recorded by an observer on board and by manually reviewing the video footage of these hauls for eight trips. In three out of eight trips, all hauls were video reviewed, whereas for the remainder of these trips the last hauls were systematically excluded. For the reviewed hauls the whole duration of the video footage was annotated. When it was not possible to identify an individual, it was recorded as *Rajidae* (unidentified ray). Table 4.4 shows for each trip, the number of hauls and the total number of rays observed by both the video reviewer and the on-board observer.

Figure 4.1 *shows per trip and haul the number of rays observed by both the video reviewer and observer. Visual inspection of the total number of rays counted per haul between the observer and the video reviews show that for most of the trips similar numbers were observed. In two trips (V2_1, V3_4) the number of rays observed in the video reviews are notably higher than the observer. Trip V3_4 is excluded from further analysis as the number of rays observed in the video review were 3 times higher than the observer data.*

Figure 4.1: *the number of rays by haul for each observer trip. Red bars indicate the observer count and grey bars the video reviewer.*

4.2.2 Video review validation

Preliminary analysis involved the comparison of the number of ray individuals observed in the video footage and by the on-board observers. For eight trips with paired haul observations, Pearson's R was used to determine the strength of the relationship between the two count variables. R values close to 0 indicate a weak relationship, whereas a value of 1 indicates a strong relationship. Figure 4.2 shows the correlation between the number of rays counted by the observer (x-axis) and video reviewer (y-axis) for all paired haul observations within each observer trip. All trips, except for one, show a strong correlation with R values ranging between 0.89-1. In one of the trips, the number of rays counted in the video review is much higher than those counted by the on-board observer, which results in a weak correlation between the variables (R = 0.3). Next, only trips with a significant relationship ($p < 0.05$) were included for further analysis (Figure 4.2). Then a parametric paired t-test was applied to all trips except V2_2, and a significant difference was found in the numbers counted between the video review and observer. For trip V2_2, a Wilcoxon rank sum test was applied, as the data did not meet the assumptions for a t-test and showed a significant difference.

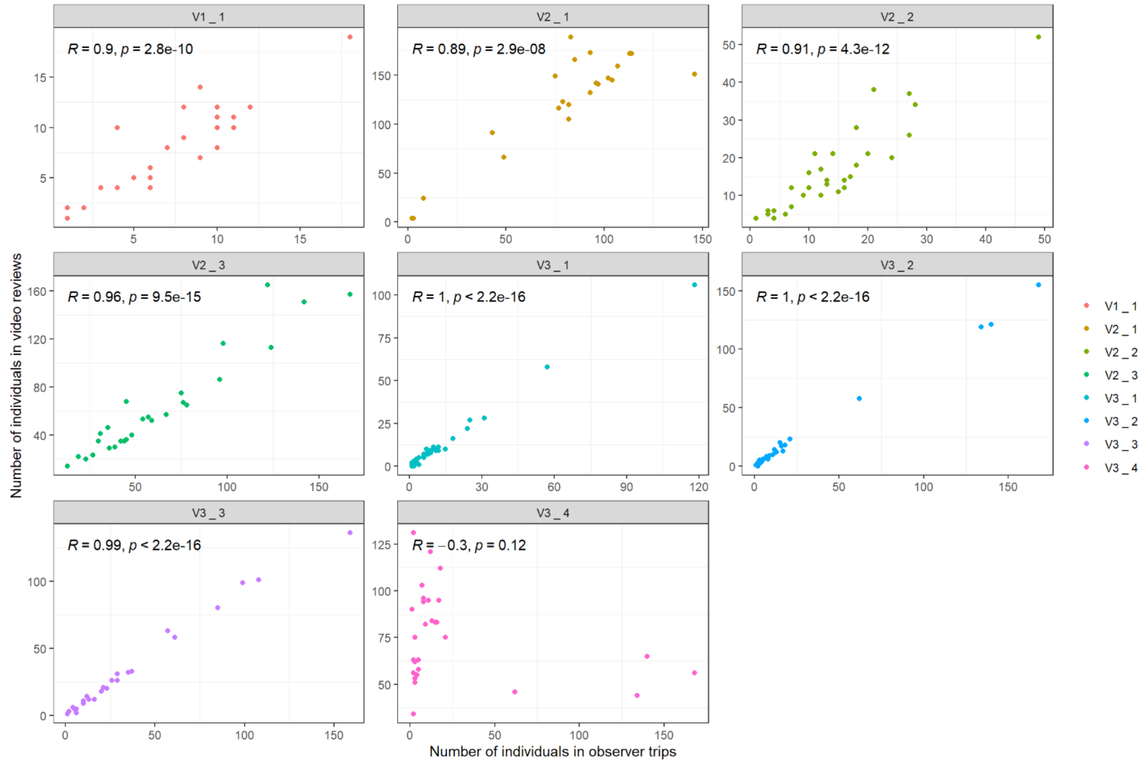

Figure 4.2: *Correlation (R) between number of individuals counted by the observer versus the video review for each observer trip. P-values denote the significance of the correlation.*

4.3 Catch composition and catch estimation

This section only covers trips for which no observer was present and only video footage was reviewed. A total of 15 trips were used in the analyses to gain insight into the potential of EM to estimate catches of skates and rays in Dutch demersal fisheries.

4.3.1 Species identification and allocation

The percentage of unidentified rays was high for each trip, accounting for more than 50% of the individuals observed in the video footage. The spatial distribution per species for all video-reviewed trips is shown in figure 4.3. Three species were identified in the video reviews: blond ray, spotted ray and thornback ray. Thornback ray is found in highest numbers, followed by spotted ray.

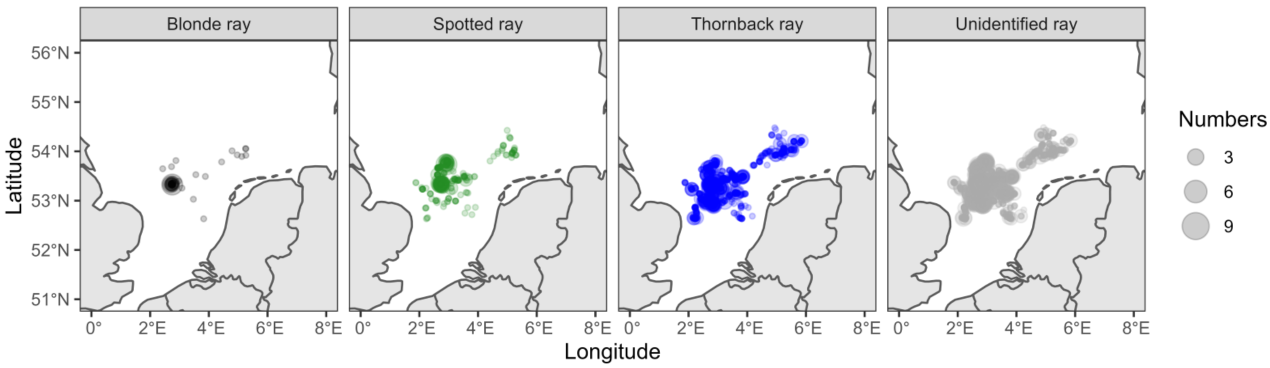

Figure 4.3: *Spatial distribution per species for all video-reviewed trips.*

The proportion of species composition from the Dutch discards self-sampling program (Bleeker *et al.*, 2022) was used to fill in the species composition of unidentified rays. Figure 4.4 shows the proportion of ray species with unidentified rays (upper panel) and the species distribution after allocating the unidentified fraction. Though cuckoo ray and starry ray were not present in the video footage, based on the spatial distribution and their presence in the discard self-sampling data the proportion of these two species is also used in the allocation of unidentified rays. Most of the unidentified rays are allocated to thornback ray and spotted ray, which were also present in highest numbers within the identified individuals.

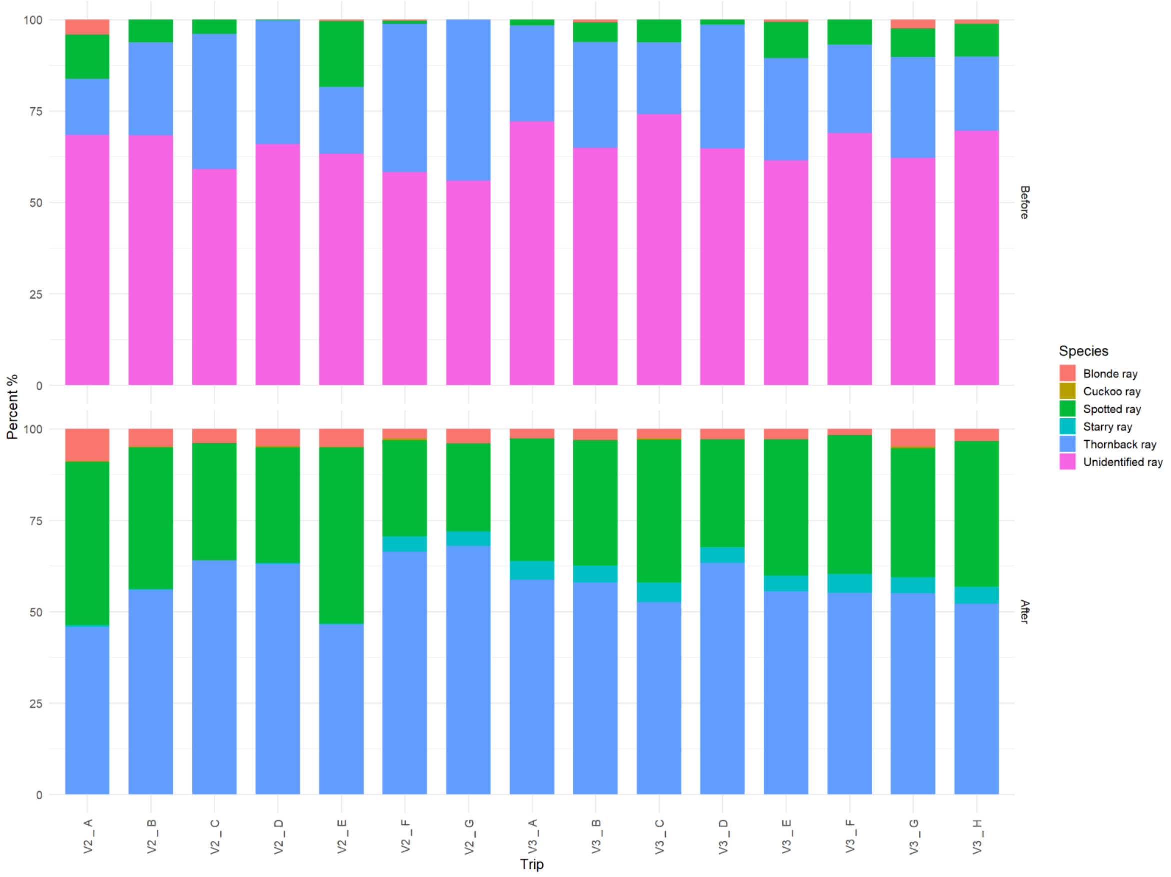

Figure 4.4: *Percentage of individuals counted by species within a trip. The upper panel shows the proportions including unidentified rays. The lower panel shows the proportions after allocating unidentified rays to species.*

4.3.2 Catch estimation

The total number of individuals from the sampled hauls for each species and trip can be found in figure 4.5 (red bars). For each of the 15 video-reviewed trips, the numbers were raised to estimate the total numbers by trip, using the ratio between the total duration of the trip and the sampled duration. The raised numbers are shown in green bars (figure 4.5). The ratio's for raising ranged between 1.1 and 2.9 between trips. For each species and trip, raised numbers were multiplied with the mean weight calculated for each species to get an estimate on catch weight (Figure 4.6).

Figure 4.5: *Counted numbers by species for the 15 video-reviewed trips, with red bars indicating the count of sampled hauls and green bars the numbers raised to total trip.*

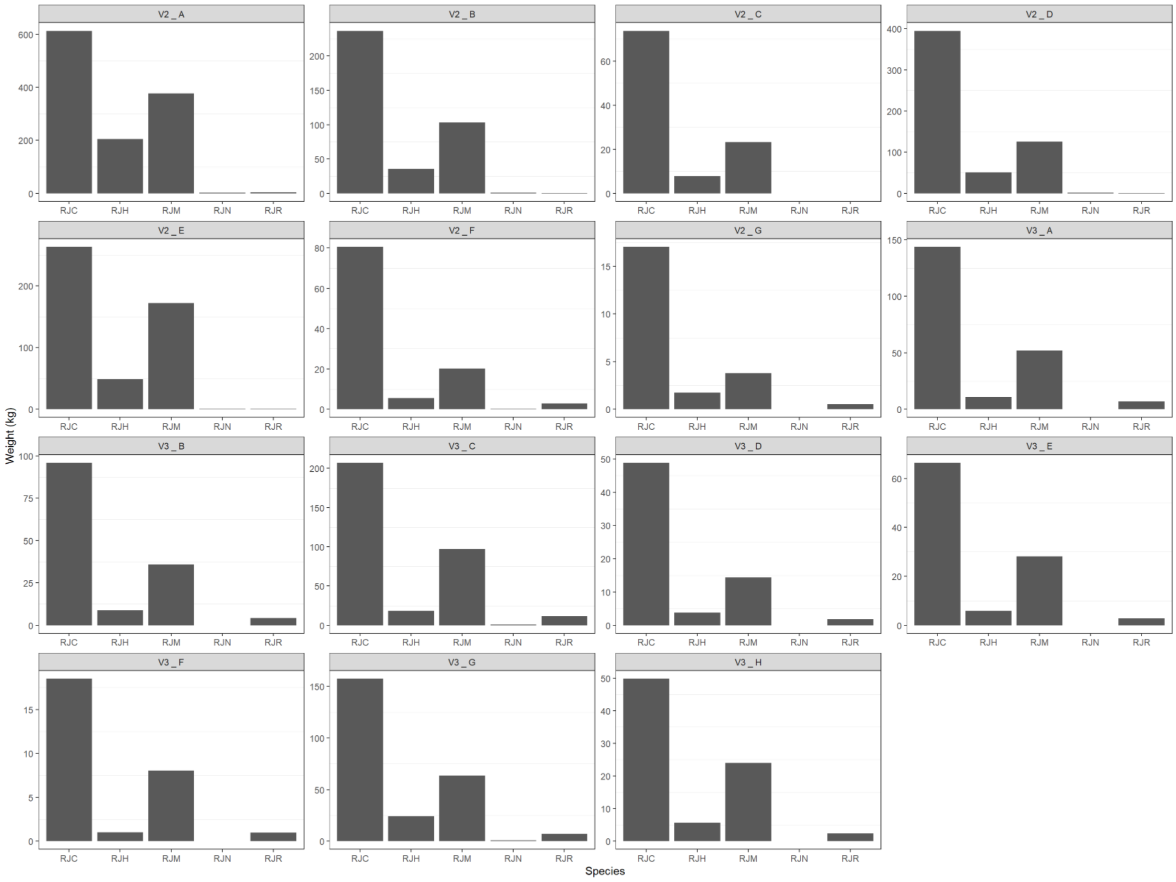

Figure 4.6: *Total catch weight (kg) per species for the 15 video-reviewed trips.*

4.4 Automated image recognition

To determine whether the network can correctly identify specific ray and the other fish species, a "confusion matrix" was created. The matrix contains the actual number of specimens per species against the predicted number, showing which prediction errors are made (Table 4.5). Each row in the confusion matrix represents the fish species and each column the predicted species. A class "background" is added. Horizontally, this means that a detection in reality is background, which happened only three times, while vertically, it means a fish was not detected by the network, being the case in 32 occasions. Ideally, all values should be on the diagonal, meaning a perfect prediction.

Table 4.5: *Confusion matrix of neural network with seven classes. Predicted class background class refers to false negatives, while actual class background refers to false positives. The green cells show the number of correct predictions.*

		Predicted class							
		Blonde ray	Spotted ray	Thornback ray	Dab	Plaice	Sole	Turbot	Background
Actual class	Blonde ray	30	5	1	0	0	0	0	2
	Spotted ray	1	95	9	0	0	0	0	6
	Thornback ray	0	7	79	0	1	0	0	3
	Dab	0	0	0	145	6	0	0	12
	Plaice	0	0	0	4	135	0	0	6
	Sole	0	0	0	1	0	6	0	3
	Turbot	0	0	0	0	0	0	6	0
	Background	0	2	0	0	1	0	0	NA

Performance of the neural network was quantified by "precision" (how many predictions are correct) and "recall" (how many of the rays are found). Figure 4.7 shows the precision and recall for the different image compositions (complexity) for rays only. The precision is high for all compositions, indicating that the network makes reliable judgements. The recall is very high with a low complexity and slowly decreases when the composition of fish become more complex. Yet, even in the most complex situation, i.e. rays being occluded, the performance of the network is good. Most errors were made by the network when the rays were pointing with the ventral side to the camera.

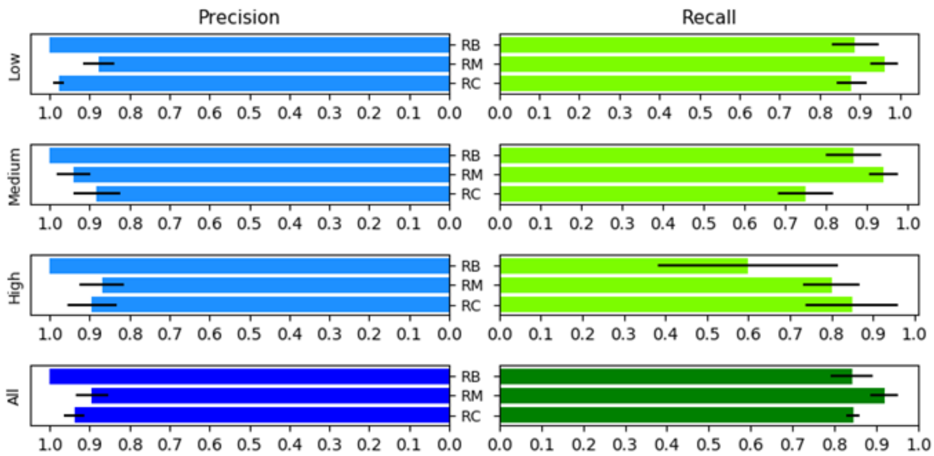

Figure 4.7: *Performance of the neural network. The data was separated by composition complexity (Low, Medium, and High). The ray species blonde ray (RB); spotted ray (RM); and thornback ray (RC) were evaluated separately. The error bars represent the standard error.*

Figure 4.8 *shows the results on the precision and recall when other fish species were included. The overall performance of detecting seven individual fish species shows no significant difference with the detection of the three ray species. This indicates that the performance of the deep neural network is not harmed by adding more fish species. Looking at the precision and recall, we can observe that the* precision is generally higher *than the recall, indicating that the most of detections made by the networks are correct and thus very reliable, but that some of the fish are not detected.*

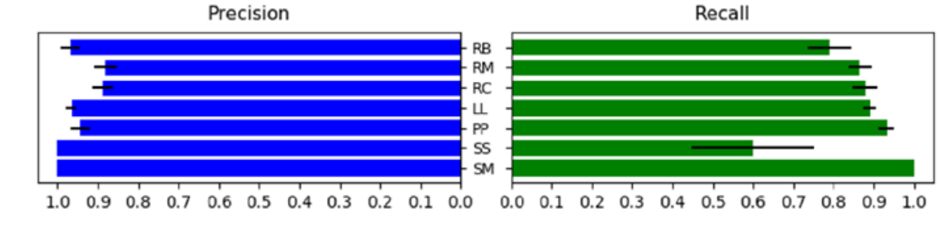

Figure 4.8: *Performance of a neural network on seven demersal fish species. The test data were separated by composition complexity (Low, Medium, and High). The seven fish species blonde ray (RB); spotted ray (RM); thornback ray (RC), dab (LL); plaice (PP); turbot (SM); and sole (SS) were evaluated separately. The error bars represent the standard error.*

In general, we can see that most of the detections are correct, and that misidentification mainly occur within the skates and rays' group and within the other demersal fish group. There is only one misidentification between these two groups. While blonde ray and spotted ray share similar patterns and are known to be sometimes misidentified in surveys or by observers, the errors occur mostly between spotted ray and thornback ray. While dab and plaice also look very similar, they're only sporadically confused by the network, which is encouraging for possible future extension to other fish species with many similarities between them.

The identification of skate and ray species based on the ventral side is difficult and often performed poorly by humans compared to the dorsal view due to the absence of well-identifiable patterns. This can also be seen in the performance of the network. Figure 4.9 shows the percentage of correct detections, misidentifications and non-detections for the dorsal side (left) and the ventral side (right). Misidentifications occur more on the ventral side, 11% compared to 3% in case of the dorsal side. Nevertheless, the network was still able to classify about 84% of the fish from the ventral side, compared to 92% on the dorsal side.

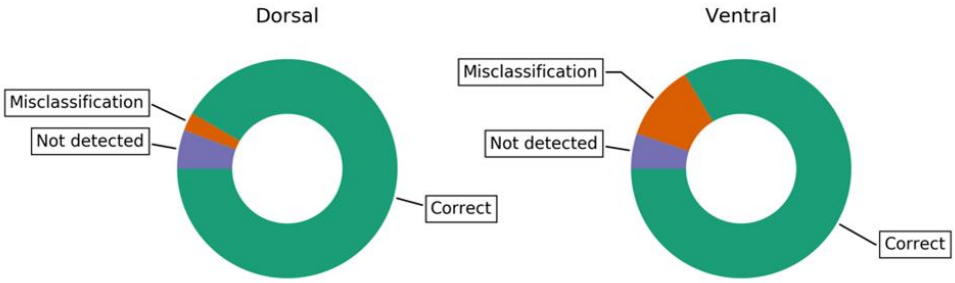

Figure 4.9: *Performance of the Fish-species network. The test set contain 332 dorsal views and 231 ventral views of the fish. Evaluation was made based on whether the detection was correct, misclassified, or the fish was not detected. For dorsal, 92% was correct, 3% was misclassified, and 6% was not detected. For ventral, 84% was correct, 11% was misclassified, and 6% was not detected.*

5 Conclusions and recommendations

The main question in this study was to explore whether EM can be used to estimate ray catches in the Dutch demersal fisheries. In total eight observer trips on board participating vessels were carried out to allow a comparison of on-board observations and video footage observations. The outcomes showed that the manual review of video footage resulted in a higher count of individual rays. When reviewing footage of the observer trips, the video reviewers are helped by the handling of the rays of the observer. The reviewer can identify when the observer is picking up an individual from the sorting belt. In addition, the reviewers can manipulate the review process by pausing and replaying footage allowing them to more accurately count the individuals compared to the observer who is limited by the sorting speed and volume of other fish on the sorting belt. However, video reviewers could not identify all rays in the catch, especially individuals with the ventral side up to the camera. As observers on board can handle the fish (i.e., turn around to see the dorsal side), species-specific identification is better compared to the video reviewer.

ICES advice for North Sea rays and skates is based on data-limited methods which only include trends in surveys. Catch estimates are not part of the current assessment because good estimates of ray catches, and especially the discarded part of the catch, are lacking. Given the common fisheries policy (CFP) requires fish stocks to be managed at maximum sustainable yield (by 2020), there is a growing demand for stock assessments which allow establishing reference points such as Bmsy and Fmsy. In this context, the use of surplus production models (e.g., SPiCT or JABBA) are recommended as these provide a quantitative estimate of the stock status. The application and parameterization of surplus production models would benefit from having reliable fisheries dependent data (catch data). Here, we have shown that EM can be a promising tool to register catches and inform stock assessment models in the future. However, EM as applied in this study, i.e., manual video review, still has some limitations.

The first limitation is the inability of the reviewers to identify a large part (>50%) of the individuals in the catch at the species level. Having species-specific catch data is required because ICES is providing single stock advice (e.g., Thornback ray (*Raja clavata*) in Subarea 4 and in divisions 3.a and 7.d (North Sea, Skagerrak, Kattegat, and eastern English Channel)). Here, the potential of data collected within the Dutch discard self-sampling program to fill in the unidentified skate and ray species was explored. Due to limitation in the spatial coverage of the sampling programme, data were aggregated at a large spatial scale (by roundfish area). However, skates and rays are known for their patchy distribution and the fishing activities and the distribution of a species within a roundfish area may not necessarily be consistent. For example, whereas starry ray (*Amblyraja radiata*) is widely distributed in the central North Sea it is only sporadically caught in the southern North Sea. Given roundfish area 6 covers both areas partially, the area will always contain a proportion of starry ray in the catch composition. This proportion will be extrapolated over the unidentified part of the catch even if the fishing trip has taken place in the southern North Sea, outside the distributional range of the species. This bias could be mitigated by deriving the proportions in species composition at a finer scale such as by ICES rectangle, if sampling coverage permits, or by extrapolating the proportion of known species within a haul to the unidentified species within the same haul.

The second limitation is the need to fully review a large number of hauls (i.e., 44-57%) to achieve a 80% certainty in the number of skates and rays caught within a trip. The number of hauls to review will increase if the accuracy needs to be improved. In addition, a representative part of the fleet will also have to contribute to the data collection if data are to be used in stock assessments. This could include a direct use of catch estimates but could possibly also include processing the data into a CPUE index as applied in turbot in the North Sea. Despite EM being increasingly used to monitor catches from commercial fisheries remotely by human experts, manual review of video footage requires a lot of manhours and thus still results in relatively high costs (Helmond *et al.* 2020). Ultimately, to reduce costs and the need of human resources, as well as allowing the technology to be expanded over a large part of the fleet would require to implementation of automated computer vision.

In this project the potential of using a deep-learning method which can deal with the complexities present in on-board catch monitoring was explored. Three conclusions can be drawn: 1) the number of fish species could be increased without loss in performance, 2) fish singulation is not necessary as the composition complexity could be increased in most cases without loss in performance, and 3) similarly looking skate and ray as well as flatfish species could be detected, even when only the white ventral side was visible to the camera. Although the method does make errors, the benefit is the potential to observe the full catch. Implementing such technology and applying this to a representative part of the fleet will increase the sample size for stock assessment by an order of magnitude compared to current practices. Of course, fish in video footages should be partly visibly to enable detection. Still, all video footage taken from the fishing vessels could be processed with known error margins, whereas current on-board or video observers generally identify species and record their choice. Since the error margins in the choices of observers is unknown, extrapolating count estimation based on small samples of the catch gives a false sense of accuracy. This is especially relevant in fisheries with rare species for which misidentification is common (van Helmond *et al.*, 2015; van Helmond *et al.*, 2017). To conclude, implementation of computer vision technology in commercial fisheries could provide an important improvement compared to current observer programs. An automated registration of catches will increase monitoring coverage of fishing activities and will reduce the errors of misidentification by allowing the probabilities of correct species identification. These probabilities can be propagated into the estimates of population size and mortality rates. Furthermore, computer vision could be expanded to acquire length and weight information from individuals to further improve the data input in more complex assessment models.

Overall, the use of video-based monitoring on-board fishing vessels is a way to improve our knowledge on catches in the Dutch fleet. Here, we aimed to improve the quantity and quality of data for the Data Limited Stocks (DLS) of skates and rays in the North Sea, using on-board EM. Such improvement in the on-board monitoring process could be relevant for a wide range of fisheries management measures (Catchpole *et al.*, 2005; Uhlmann *et al.*, 2013). This includes the landing obligation which obliges fishers to land the complete catch of species under quotas, including the undersized, unmarketable part of the catch. The landing obligation also applies to skates and rays which are managed under the quota regulation. Moreover, since the introduction of the TAC for skates and rays, the TAC has gradually been reduced and has become restrictive for most fisheries (ICES, 2021). Consequently, skates and rays are one of the main "choke species" for the Dutch fishery under the landing obligation. In this context, there is a real incentive to improve monitoring and data collection of skate and ray catches allowing the use of more quantitative assessments methods currently developed for "data-limited stocks" within ICES. These novel methods provide improved advice which falls within the ICES MSY framework. Yet, the ability of these methods to ensure fishing opportunities reflect the total catch of a stock depend on the ability to efficiently register catches on-board. Possibly, EM could be a tool leading to better registration of catches and thus estimates of fishing mortality on a stock. Consequently, improved data could allow an increase of quota and reduce the risk of these species being a "choke species" under the landing obligation. Especially, the automated registration of catches by species, which was explored in this project, may contribute to the accuracy of catch estimates for "data-limited stocks".

6 Acknowledgements

This project could not have been successfully executed without those involved in science-industry research collaboration. First and foremost, we thank the skippers and crew for their willingness to have an EM-system installed and their continuous support and work on keeping the system running. We also thank the representatives from the Nederlandse Vissersbond and VisNed. A special thanks to David Ras and Henk Brouwer for their flexibility in organising the observer trips and carrying out the work on board within the observer trips. Finally, we would like to thank our (former) colleagues Linh Nguyen, Morgane Amelot, Chun Chen, and Jan Jaap Poos for their valuable input.

7 Quality Assurance

Wageningen Marine Research utilises an ISO 9001:2015 certified quality management system. The organisation has been certified since 27 February 2001. The certification was issued by DNV.

If desired, information regarding the performance characteristics of the analytical methods is available.

If the quality cannot be guaranteed, appropriate measures are taken.

References

Bleeker, K., Van Overzee, H.M.J., Dammers, M. (2019). Discard self-sampling of Dutch bottom trawl and seine fisheries in 2021. CVO report: 22.009.

Brander, K. (1981). "Disappearance of common skate Raia batis from Irish Sea." Nature 290(5801): 48-49.

Catchpole, T., Frid, C., and Gray, T. 2005. Discarding in the English north-east coast nephrops norvegicus fishery: the role of social and environmental factors. Fisheries Research, 72: 45–54.

Dulvy, N.K., Dulvy, N.K, Baum, J.K., Clarke, S., Compagno, L.J.V., Cortés, E. *et al.* (2008). "You can swim but you can't hide: the global status and conservation of oceanic pelagic sharks and rays." Aquatic Conservation: Marine and Freshwater Ecosystems 18(5): 459-482.

Dulvy, N.K., Fowler, S.L., Musick, J.A., Cavanagh, R.D., Kyne, P.M. *et al.* 2014. Extinction risk and conservation of the world's sharks and rays. eLife 2014;3:e00590. DOI: 10.7554/eLife.00590

Heessen, H. J. L. (2010) State of the Art - Haaien en roggen in de Noordzee. Wageningen, IMARES Wageningen UR rapport C011/10.: 30 pp.

Hold, N., Murray, L. G., Pantin, J. R., Haig, J. A., Hinz, H., & Kaiser, M. J. (2015). Video capture of crustacean fisheries data as an alternative to on-board observers. ICES Journal of Marine Science, 72, 1811– 1821. https://doi.org/10.1093/icesjms/fsv030

ICES (2017): Report of the Workshop to compile and refine catch and landings of elasmobranchs (WKSHARK3). ICES Expert Group reports (until 2018). Report. https://doi.org/10.17895/ices.pub.19290452.v1

ICES (2020) Workshop on the distribution and bycatch management options of listed deep-sea shark species (WKSHARK6). ICES Scientific Reports. 2:76. 85 pp. http://doi.org/10.17895/ices.pub.7469

ICES (2021) Report of the Working Group on Elasmobranch Fishes (WGEF). ICES Scientific Reports. 4:74. 848pp. http://doi.org/10.17895/ices.pub.21089833

Kindt-Larsen, L., Kirkegaard, E., & Dalskov, J. (2011). Fully documented fishery: A tool to support a catch quota management system. ICES Journal of Marine Science, 68, 1606– 1610. https://doi.org/10.1093/icesjms/fsr065

McElderry, H., Schrader, J., & Illingworth, J. (2003). The efficacy of video-based monitoring for the Halibut Longline (p. 80). Victoria, Canada. Canadian Research Advisory Secretariat Research Document 2003/042.

Mortensen, L. O., Ulrich, C., Olesen, H. J., Bergsson, H., Berg, C. W., Tzamouranis, N., & Dalskov, J. (2017). Effectiveness of fully documented fisheries to estimate discards in a participatory research scheme. Fisheries Research, 187, 150– 157. https://doi.org/10.1016/j.fishres.2016.11.010

Needle, C.L., Dinsdale, R., Buch, T.B., Catarino, R.M.D., Drewery, J. and Butler, N. (2015). Scottish science applications of Remote Electronic Monitoring. ICES Journal of Marine Science, 72: 1214–1229.

Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Schindler, D. E., Essington, T. E., Kitchell, J. F., Boggs, C., Hilborn, R. (2002). "Sharks and tunas: fisheries impacts on predators with contrasting life histories." Ecological Applications 12(3): 735-748.

Scientific, Technical and Economic Committee for Fisheries (STECF) – Long-term management of skates and rays (STECF-17-21) (2017). Publications Office of the European Union, Luxembourg, 2017, ISBN 978-92-79-67493-8, doi:10.2760/44133, JRC109366

Sguotti, C., Lynam, C.P., García-Carreras, B. Ellis, J.R., Engelhard, G.H., (2016). "Distribution of skates and sharks in the North Sea: 112 years of change." Global Change Biology 22(8): 2729-2743.

Stevens, J. D., Bonfil, R., Dulvy, N.K., Walker, P.A., (2000). "The effects of fishing on sharks, rays, and chimaeras (chondrichthyans), and the implications for marine ecosystems." ICES Journal of Marine Science 57(3): 476-494

Stanley, R.D., Karim, T., Koolman, J., and McElderry, H. (2015). Design and implementation of electronic monitoring in the British Columbia groundfish hook and line fishery: a retrospective view of the ingredients of success. ICES Journal of Marine Science, 72: 1230–1236.

Uhlmann, S. S., van Helmond, A. T. M., Stefánsdóttir, E. K., Sigurðardóttir, S.,Haralabous, J., Bellido, J.M., Carbonell, A., *et al.* 2013. Discarded fish in European waters: general patterns and contrasts. ICES Journal of Marine Science, 71: 1235–1245.

van Helmond, A. T. M., Chen, C., & Poos, J. J. (2017). Using electronic monitoring to record catches of sole (Solea solea) in a bottom trawl fishery. ICES Journal of Marine Science, 74, 1421– 1427. https://doi.org/10.1093/icesjms/fsw241

van Helmond A. T. M., Mortensen L. O., Plet-Hansen K. S., Ulrich C., Needle C. L., Oesterwind D., Kindt-Larsen L. *et al.* (2020) Electronic monitoring in fisheries: Lessons from global experiences and future opportunities. Fish and Fisheries, 21: 162–189.

Walker, P. A. and H. J. L. Heessen (1996). "Long-term changes in ray populations in the North Sea." ICES Journal of Marine Science 53(6): 1085-1093.

Justification

Report C093/22
Project Number: 4311400024

The scientific quality of this report has been peer reviewed by a colleague scientist and a member of the Management Team of Wageningen Marine Research

Approved: Jasper Bleijenberg
Fisheries researcher

Signature:

Date: 23 Januari 2023

Approved: Dr. Ir. T.P. Bult
Director

Signature:

Date: 23 Januari 2023

Annex 1

Do not site this manuscript without prior permission of the authors.

Automated detection of ray species in fish-catch monitoring using deep-learning techniques

Linh Nguyen[1], Edwin van Helmond[2], Jurgen Batsleer[2], Jan-Jaap Poos[1,2], Gert Kootstra[1]*

1. Wageningen University & Research, Wageningen, 6700 AA, The Netherlands

2. Wageningen IMARES, Institute for Marine Resources and Ecosystem Studies, PO Box 68, 1970 AB IJmuiden, The Netherlands

*Corresponding author: Tel: +31(0)317480302; e-mail: gert.kootstra@wur.nl,

Abstract

North-Sea rays are currently considered to be data-limited stocks as catch estimates are uncertain and the stock assessments are primarily based on indicative trends of survey data. There is great potential in using computer-vision technology for automated identification of ray species in commercial fish catches, which will lead to improved and more cost-efficient data collection. The objective of this study is to investigate whether a current deep-learning method can deal with the complexities present in on-board catch monitoring. We deployed deep-learning techniques for automated detection of three ray species and four flatfish species. To reflect the sorting process on board of a commercial fishing vessel, the collected data consisted of images containing different fish species with varying amount of overlap. Our results show: 1) no significant difference in performance when more species are added for detection, 2) the performance is not influenced by the amount of overlap, and 3) similar looking fish can be distinguished from each other, even when the white ventral side was faced up. Compared to current practice of random manual sampling, we expect our method to be able to improve stock assessments of ray species as well as other data-limited stocks.

Keywords: fish detection, computer vision, deep learning, neural networks, stock assessments

Introduction

Fisheries management generally relies on accurate estimates of fish abundance and fishing mortality. These are often derived from population models that integrates available fisheries data including catch estimates (Beverton and Holt, 1957; Punt *et al.*, 2006; Rijnsdorp *et al.*, 2007; Bradshaw *et al.*, 2018). Accurate catch estimates are thus a cornerstone of sustainable fisheries management, especially for species that are sensitive to fishing, such as sharks and rays (Stobutzki *et al.*, 2002; Broadhurst *et al.*, 2006; Harry *et al.*, 2011). However, factors such as species misidentification, low economic value, and low quotas could cause inaccuracy in catch estimates, which is the case for the ray population in the North Sea. Because of the low value and low quotas, a substantial fraction of ray catches are being thrown overboard (Stevens *et al.*, 2000; Oliver *et al.*, 2015) in fisheries that target other demersal fish stocks. The ray catches in the North Sea are mostly estimated from on-board observer programs with low sample sizes (Stratoudakis *et al.*, 1999; Dickey-Collas *et al.*, 2007; Poos *et al.*, 2013). As a consequence, stock assessments of rays are solely based on data from annual scientific surveys (ICES WGEF, 2018). To improve sustainable management of ray populations, better data collection of catch quantities per species is required.

Video-based monitoring on-board fishing vessels, commonly described as electronic monitoring (EM), allows observing catches without requiring additional on-board personnel (McElderry *et al.*, 2003; Ames *et al.*, 2006; Kindt-Larsen *et al.*, 2011; Stanley *et al.*, 2015; Hold *et al.*, 2015; van Helmond *et al.*, 2017). EM systems enable continuous catch monitoring over long periods, making it suitable for monitoring species of low abundance or patchy spatial distribution. Considering the dispersion patterns of North Sea rays, EM would significantly improve data collection of these ray species (Walker *et al.*, 1997; Ellis *et al.*, 2004; Daan *et al.*, 2005). However, as EM requires human observations, costs are still relatively high and the amount of human resources needed for video review is a limiting factor (Needle *et al.*, 2015; Mortensen *et al.*, 2017). To reduce the workload and improve the sampling frequency, a reform of EM data processing is necessary.

Automated image recognition is the logical next step to improve accuracy in catch estimates on board. However, training a computer to recognise rays during the sorting process on board a fishing vessel is challenging due to the variability in appearance. For example, rays from the same species are not identical in size, colour, and patterning, while different fish species also share similarities, e.g. all ray species have a white ventral side. Also, catch is often loaded in bulk on a sorting belt for processing on board, and consequently, fish are randomly positioned on the belt, often overlapping each other, resulting in severe occlusions. The monitor system needs to deal with partly visible specimens and must be able to identify multiple individuals and species in a single image frame.

Related work

Recognising multiple fish in images is an object-detection task, consisting of the subtasks, object localisation and object classification. Object localisation involves locating objects in an image, and object classification assigns a class label to that located object. Several approaches have been proposed to perform the detection task or one of the subtasks. Traditional approaches rely on tailor-made image-processing algorithms. Zion *et al.* (1999), and Storbeck and Daan (2001), for instance, constructed a system that performs image classification based on shape descriptors of the fish in the images. However, these systems are constrained by a fixed orientation of the fish. White *et al.* (2006) suggested a method that automatically rotates and orient fish in the images before classification as a solution to this constraint. A different approach was proposed by Hu *et al.* (2012), in which, colour and texture features were used instead of shape features, as fish are not always completely visible in images. The above systems were developed for images with single fish of a specific species. Marini *et al.* (2018) developed a framework that detects multiple fish in underwater videos using traditional image-processing techniques, such as colour thresholding, for localisation and a machine-learning based approach for binary classification (a fish or not a fish). All these approaches require dedicated feature extraction for specific species in images with specific conditions, and therefore, cannot be reused for other fish detection problems.

Developing tailor-made systems using handcrafted features for every particular problem is not practical and quite costly. An alternative approach is to use of deep-learning techniques, such as convolutional neural networks (CNNs) (LeCun *et al.,* 2015). Through feeding a large number of data examples, CNNs are trained to perform specific tasks by independently finding relevant features in the input data. At the beginning of the training, the network outputs random predictions. During training, the network adjusts the weights of the network connections based on the error in prediction. The goal is to show the network enough data until the prediction is a good approximation of the actual answer. Current, widely used object-detection CNN architectures are YOLO (Redmon *et al.*, 2016) and RCNN and its variations (Girshick *et al.*, 2013; Girshick, 2015; Ren *et al.*, 2015). YOLO performs the localisation and classification tasks jointly in one CNN, while RCNN splits the tasks in two stages where relevant regions in the images are first searched by a region-proposal network and then every proposed region is classified by a different network.

Several CNN-based approaches have been suggested for detecting or classifying fish in images. Shafait *et al.* (2016) and Siddique *et al.* (2018) developed an approach that classifies images containing single fish underwater, while Lu *et al.* (2019) designed a CNN that classifies images of fish catch landed on the deck of fishing vessels. Allken *et al.* (2018) implemented a CNN that classified images with multiple fish in a controlled environment. Various studies have implemented a variation of RCNN or YOLO to detect fish species in image datasets that are made publicly available for benchmark purposes, such as ImageCLEF (underwater images of fish) (Li *et al.*, 2016), or for a competition, such as the Nature Conservancy Fisheries Monitoring (images of fish on decks of fishing vessels) (Wang *et al.*, 2018).

Objectives

The main objective of this paper is to investigate whether a current deep-learning method can deal with the complexities that an on-board monitoring system faces. Specifically, we investigated three main factors: 1) the influence of training neural networks on different number of fish species, 2) the complexity of the composition of the fish on the conveyor belt, and 3) the main sources of error.

To this end, we applied deep-neural networks to localise and classify different fish species in images mimicking the sorting process on board of a fishing vessels. Rather than using EM videos, specimens were physically collected as species identification require specific expertise in order to get reliable ground truth, especially for the ray species. The images include overlapping fish of different species randomly positioned in the frame. We studied the performance of the networks to detect three ray species; thornback ray (*Raja clavata*), spotted ray (*Raja montagui*), and blonde ray (*Raja brachyura*), in combination with four demersal flatfish species, plaice (*Pleuronectes platessa*), dab (*Limanda limanda*), turbot (*Scophthalmus maximus*), and sole (*Solea solea)*. These fish are commonly caught in demersal-mixed fisheries in the North Sea.

Methods

Data collection

Training a system to recognise ray species in catches on board of fishing vessels in demersal mixed fisheries requires representative image data. Image data was created by placing fish on a blue background in a lighting box with LED strips on the topside. A Nikon D5300 camera was mounted above the fish, capturing images with a resolution of 6000 x 4000. Composition of collected images was comparable to the top view of a conveyor belt during the sorting process on a fishing vessel. Since the fish catch consists not only of ray species, image data were also collected of other demersal fish species. Specimens of seven fish species were collected. Three were ray species: thornback ray, spotted ray, and blonde ray; and four flatfish species, plaice, dab, turbot, and sole. The specimens were collected from four commercial fishing trips to the North Sea between June and November 2018 by a commercial fishing vessel. During these trips, the vessel deployed an 80 mm codend on a pulse trawl gear (de Haan *et al.*, 2016).

To mimic the sorting process on a conveyor belt, images were collected with different compositions. The possible compositions were defined in three levels of complexity (Figure 1). At low complexity, images contain single fish completely visible (1108 images). The view of both dorsal and ventral sides of the fish were photographed. At medium complexity, images contain multiple fish lying close and adjacent to each other (253 images with 871 fish). Lastly, at high complexity, images contain multiple overlapping fish (229 images with 834 fish). In medium and high complexity, the combinations of fish were randomly selected with the only requirement that at least two fish are in the frame. Also, the dorsal and ventral view of each fish in the image was randomly selected. In all situations, the fish were randomly positioned. The total data set consists of 1591 images containing 2813 individuals of seven species.

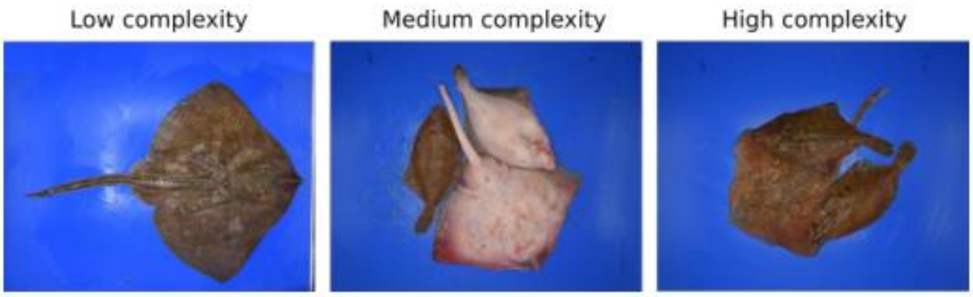

Figure 1 Sample images of possible compositions that could occur on a conveyor belt. From left to right: a single fish (low complexity), multiple fish lying adjacent to each other (medium complexity), or fish overlapping each other (high complexity).

Deep-learning method for fish localisation and classification

This study uses YOLO version 3 (Redmon and Farhadi, 2018) for ray species and flat fish species detection. Although there are other methods available for the task, such as Faster R-CNN (Ren *et al.*, 2015) and SSD (Liu *et al.*, 2016), YOLO has been used frequently for similar tasks and has a good trade-off between speed and accuracy. Differences between these methods, however, are marginally, and similar results can be expected for a different deep-learning method.

Deep neural networks allow the end-to-end processing of data, meaning that the raw images are taken as input, which are processed by the network to provide the required output. A YOLO network provides the bounding boxes of the objects in the image as output, together with the corresponding object classes. Additionally, the network provides a confidence score for each of the object detections. Hence, YOLO performs the subtasks of object localisation and object classification jointly in one network. The combined subtasks will be henceforward termed *object detection*. Figure 2 illustrates the architecture of YOLO. The YOLO network consists of a feature extractor backbone and three blocks that perform the bounding box detection at different spatial scales. The backbone has 53 convolutional layers with skip connections (residual blocks) to ensure minimal loss of information across the large number of layers. This residual learning was introduced by He *et al.* (2016). YOLO also adopted a Feature Pyramid network-like structure (Lin *et al.*, 2017) to allow for detection of object with different sizes on the images. In the backbone block, feature maps are extracted at three spatial scales and then merged to add more meaningful information from earlier feature map for better object detection. For each bounding box, YOLO outputs five components: *x*, *y*, *w*, *h*, and confidence class probability. The (*x*, *y*) coordinates represent the centre of the bounding box, while the (*w*, *h*) are the dimensions of the bounding box. For more details on the network, we refer to Redmon *et al.* (2016), Redmon and Farhadi (2017), and Redmon and Farhadi (2018).

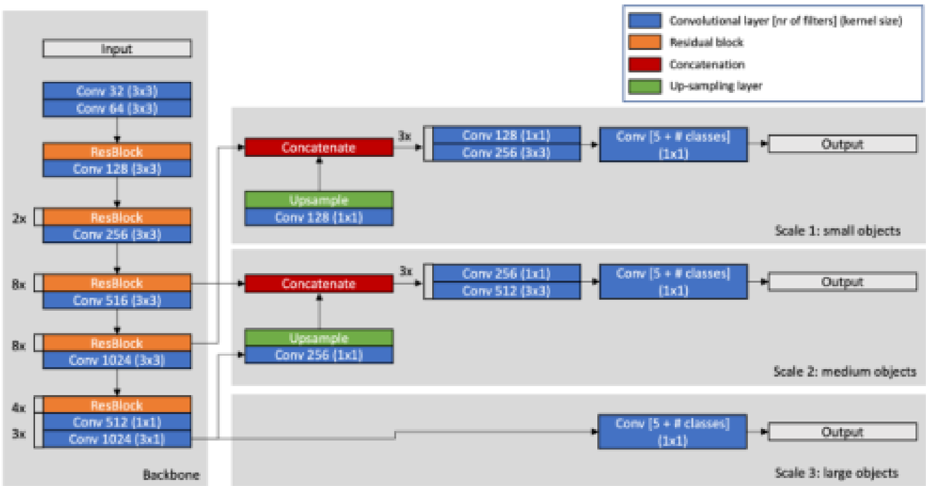

Figure 2 Visual representation of YOLOv3 network. It consists of a feature extractor backbone and three blocks for detecting objects in different scales.

Training a YOLO network requires a dataset in which every object of interest in the images is annotated. An annotation includes the coordinate and size of the object's bounding box and the corresponding class, which in our case is the name of the fish species. Examples of annotations can be observed in Figure 3 and the specific annotations used will be discussed in more detail in the subsection "Experiments". Training a neural network essentially comes down to tuning the network weights. The weights are adjusted based on the loss (or error) of the network. The loss is determined based on the location and size of the predicted bounding box and on the predicted class. Additionally, the network provides a confidence score on its predictions. Initially, the loss will be high, as the network is not yet able to correctly predict the location and class of the objects. Over time, the loss will decrease as the network learns to translate the image data into the desired output. During training, the network will learn to detect patterns in the image data that can be used to detect the fish in the images. The data set was split in a proportion of about 8:2 to generate a training and a test set respectively.

To improve learning, two techniques were used in conjunction: transfer learning and data augmentation. Transfer learning is a technique that reuses a network that was trained on a related problem as a starting point (Bengio, 2012). This improves learning a task, especially if training data is limited. We initialised the networks with the weights of a network pre-trained on images from ImageNet (Krizhevsky *et al.*, 2012), a large online image data set containing many different classes of objects, and then fine-tuned them with our collected data set. With data augmentation, the original training data is randomly transformed to make a trained network more robust to variations in the appearance of objects in the images. We applied real-time augmentation where each image is randomly transformed on-the-fly at each iteration during the training procedure. This ensures that the network sees different images each epoch. The augmentations include transformations of the colour (hue, saturation, and brightness) of the image by changing the pixel values, and spatial transformations by randomly flipping, rotating, resizing and cropping the image (Table 1) (Redmon *et al.*, 2016). Images were resized from 6000 x 4000 to 416 x 277 to keep the aspect ratio, and then padded with black bars to create a 416 x 416 squared image.

Table 1 Data augmentation parameter settings

Parameter	*Value*
Hue	Initial value * [-0.1 – 0.1]
Saturation	Initial value * [1/1.5 – 1.5]
Brightness	Initial value * [1/1.5 – 1.5]
Flipping	Probability of 0.5

Rotating	[0-360] degrees
Cropping	+/- 30% of original image

All three networks were fine-tuned on a NVIDIA GeForce GTX 1080 Ti. We used a batch size of 64, the learning rate was set to 10^{-3}, the momentum was set to 0.90, and the decay rate was set to 5×10^{-3}. The networks were fine-tuned for 500 epochs (number of passes through the full training set).

Experiments

Three experiments were conducted to evaluate the performance of the YOLO network in detecting fish in camera images. In the first experiment, different levels of class complexity were examined. A first network, the "Ray network", was trained to detect rays as a single class. This network was evaluated to examine if rays, irrespective of species, can be detected. To examine if a network can distinguish the three ray species, the "Ray-species network" was trained to detect the three ray species, separately. To explore the possibility of extending the detection to more species, a third network, the "Fish-species network", was trained to detect the three ray species, as well as, the four flatfish species. The three networks were trained with differently annotated image data, see Figure 3. For the "Ray network", all rays in the images were annotated with a single class "Ray", irrespective of the specific ray species. The "Ray-species network" was trained with annotations where each ray species was provided a different class label. For the "Fish species network", all fish in the images were annotated and labelled with their corresponding species name. All described networks in this experiment were trained and tested with images of every composition complexity. In the second experiment, the influence of the composition complexity on the network performance was examined. To this end, the performance of the three networks was tested for images of low, medium and high complexity, as illustrated in Figure 1. In the third experiment, common errors made by the fish species network were examined.

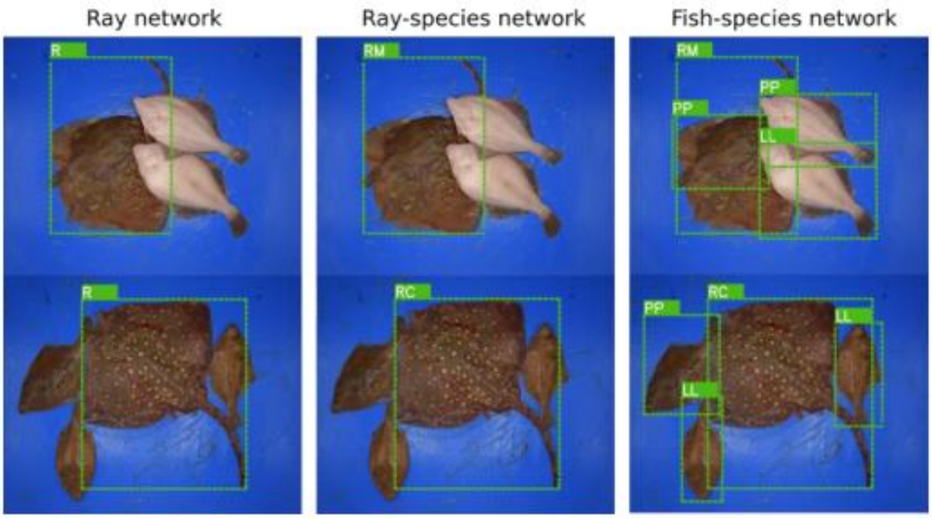

Figure 3 Annotation example for the same training image for three networks that have different detection tasks. Bounding boxes are drawn around the object of interest for each detection task. The first column contains training images for a network with a ray detection task. The second column contain images for a ray species task. The last column contains images for a fish species detection task. R=ray, RM= spotted ray, RC=thornback ray, PP=plaice, LL=dab.

Indicators of network performance

Detection of objects was determined based on the intersection-over-union metric (IoU) (Eq. 1), which represents how much overlap there is between the area of an actual bounding box, A^G, surrounding an object of interest, and a predicted bounding box, A^P. When the actual and predicted bounding box have an $\mathrm{IoU} \geq 0.5$ and have the same class, the detection is defined as a true positive (TP). All fish in images that are not detected are labelled as false negatives (FN), and all detections that do not correspond with a fish or have the wrong class are labelled as false positives (FP).

$$\mathrm{IoU} = \frac{|A^P \cap A^G|}{|A^P \cup A^G|} \qquad \text{(Eq. 1)}$$

The performance of the models has been evaluated using the precision, recall, and F1 metric. Precision represents the proportion of correct detections over all detections by the network (Eq. 2) and recall represents the proportion of correct detection that could be retrieved out of all actual detections (Eq.

3). The F1-score combines both precision and recall and computes the harmonic mean of the two (Eq. 4).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \qquad \text{(Eq.2)}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad \text{(Eq.3)}$$

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision + Recall}} \qquad \text{(Eq.4)}$$

To determine the confidence interval for every metric, the test set was randomly split in five groups. For every group, the mean and its 95%-confidence interval are provided. A two-sided dependent Student's t-test was used in experiment 1. In experiment 2, a two-sided independent Student's t-test was used. In both cases, an alpha-level of 0.05 was applied.

Results

Experiment 1

Comparisons were made between the Ray network, Ray-species network and Fish-species network. Figure 4 shows some examples of correct and incorrect detections made by the three networks on two test images.

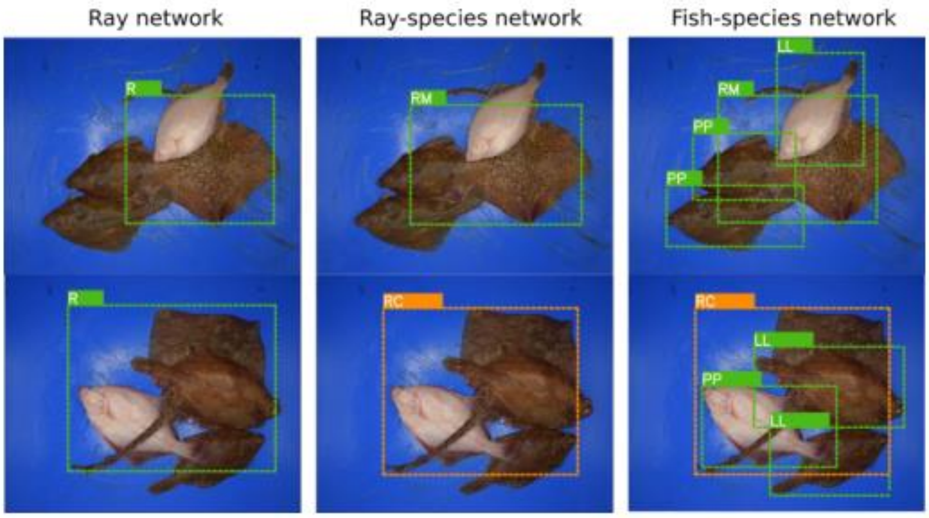

Figure 4 Detection example by three neural networks with different detection task. The first column are detections by the Ray network that was trained to detect ray species as a collective group. The second column are detections by the Ray-species network that was trained to detect three ray species. The last column are detections by the Fish-species network that was trained to detect seven fish species. Green boxes represent correct detections. Oranges boxes represent missing detection. The box labels describe the actual class: R=ray, RC=thornback ray, RM= spotted ray, PP= plaice, LL = dab.

Figure 5 shows the quantitative performance of each of the three networks. The plots show the precision, recall and F1-score for the different networks tested on the tasks of ray detection, ray-species detection and fish-species detection. Detecting ray species as a single class is performed by each network with the lowest F1 score of 0.97. No significant differences in performance were found between the network that was specifically trained for detecting rays and the networks that were trained to detect individual ray species and individual fish species. When testing the ability of the Ray-species and Fish-species networks to detection specific ray species, we observe lower average F1-scores (0.90 and 0.88 respectively). This indicates that the networks can detect rays in general very reliably, but that distinguishing individual ray species is more challenging. No significant difference was found between the Ray-species network and the Fish-species network, showing that increasing the number of classes during training does not affect the overall performance of ray-species detection.

Finally, looking at the task of fish-species detection, we see that the Fish-species network has an average F1-score of 0.91. The overall performance of detecting the seven individual fish species shows no significant difference with the detection of the three ray species. Indicating that the performance of the deep neural network is not harmed by adding more fish species. Looking at the precision and recall, we can observe that the precision is generally higher than the recall. This shows that the detections made by the networks are very reliable, but that some of the fish are not detected correctly.

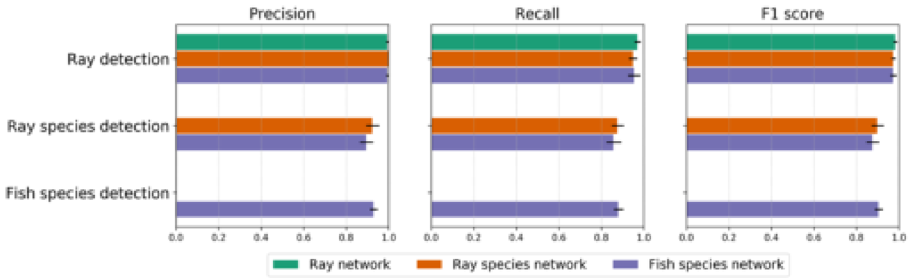

Figure 5 Performance of three neural networks on three different detection tasks. Each network was trained on the same data to detect different number of classes. The Ray network detects all ray species as one class, the Ray-species network detects the ray species as individual classes and the Fish-species network detects all seven fish species individually. The performance on ray detection was tested on all three networks, in which the individual ray species were not taking into account. The performance on detecting the three ray species was tested with the ray species and fish species networks. Finally, the detection of the individual fish species was tested only for the fish species network. The bars show the mean scores and the error bars represent 95% confidence intervals on the means.

Experiment 2

In this experiment, we investigate the performance of the networks depending on the complexity of the fish composition. Here, the three networks were trained on all training images and then tested on images with low, medium and high complexity. Figure 6 shows some correct and incorrect examples of detections by the Ray-species network for the different complexities.

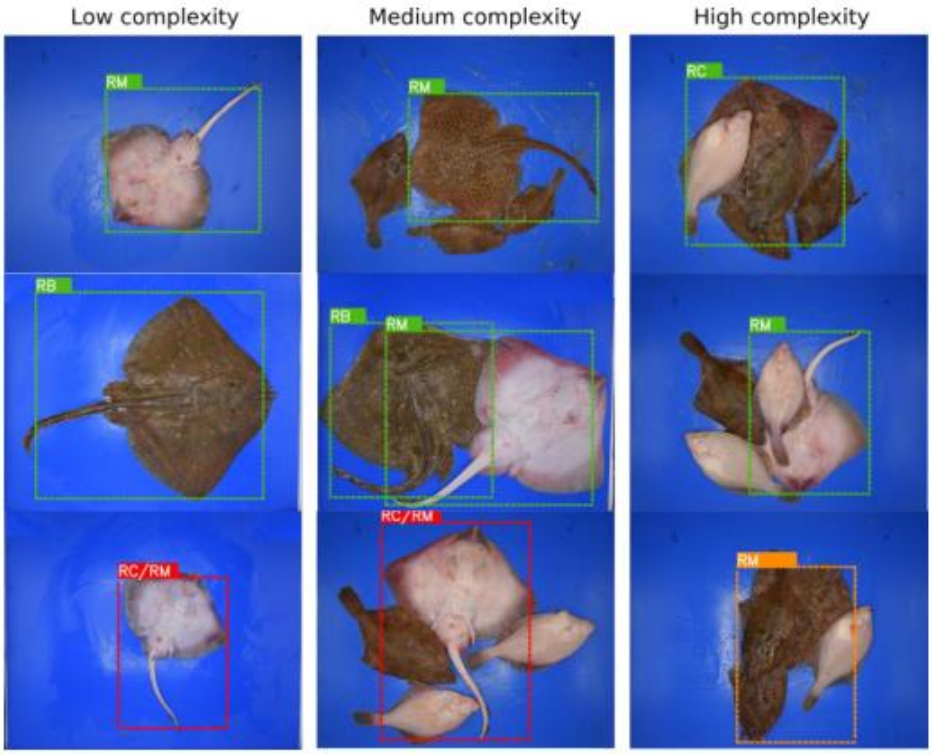

Figure 6 Examples of ray species detections by a neural network on images with different composition complexity. The network was trained to detect the blonde ray (RB), the spotted ray (RM), and the thornback ray (RC). The first column contains low complexity images, the second column contains medium complexity images, and the last column contains high complexity images. Green boxes represent correct detections, red boxes represent misclassifications, and oranges boxes represent missing detections. The labels of the green and orange boxes describe the actual class. The labels of the red boxes describe the actual class and the incorrect predicted class.

Figure 7 shows the F1-scores for the three different networks tested on the low, medium and high complexity. The Ray network has a perfect detection performance on the low complexity. For the medium and high complexity, some detection errors are visible. The Ray-species network and the Fish-species network show a lower performance on the low complexity and a small decrease for more complex compositions. In general, the results show a small decrease in performance for the more complex situations. However, performing a two-sided t-tests show no significant difference between the performance for the different composition complexities, apart from the medium and high complexity of the Fish-species network (p-value 0.03). That most differences are non-significant shows that the networks in general can deal with the complex situations with overlapping fish without a big loss in performance.

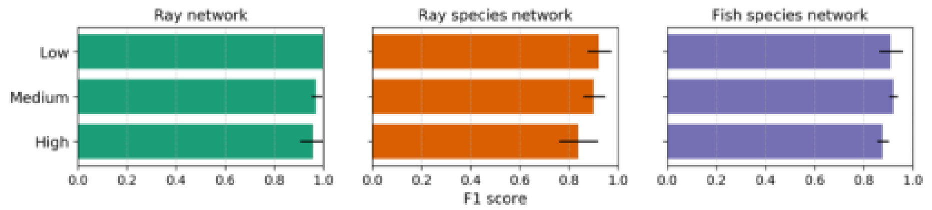

Figure 7 Evaluation of three neural networks with different detection tasks on different test data. The networks were trained on all complexity images but evaluated per complexity (low, medium, and high). Error bars represent 95% confidence intervals.

Analysis of errors

Errors made by the networks can be categorised as either no detection or misclassification of localised fish. Table 2 show the confusion matrix for the Fish-species network, indicating how often these errors occur. In total, the network made 566 detections, which happens to correspond to the same number of fish in reality. The actual classes are presented horizontally and the predicted classes vertically. A class "background" is added. Horizontally, this means that a detection in reality is background, which happened only thrice, while vertically, this means that a fish was not detected by the network, which was the case in 32 occasions. In general, we can see that most of the detections are correct. Misclassifications occur mainly within the ray species group and within the other group of demersal fish species. Only once, there is a misclassification between these two groups. Even though the blonde ray and spotted ray share similar back pattern (both have spots), the errors occur mostly between spotted ray and thornback ray. Dab and plaice are also sometimes mixed up. In total, 35 misclassifications were made on the test set.

Table 2 Confusion matrix of neural network with seven classes. Predicted class background class refers to false negatives, while actual class background refers to false positives. The green cells show the number of correct predictions.

		Predicted class							
		Blonde ray	Spotted ray	Thornback ray	Dab	Plaice	Sole	Turbot	Background
Actual class	Blonde ray	30	5	1	0	0	0	0	2
	Spotted ray	1	95	9	0	0	0	0	6
	Thornback ray	0	7	79	0	1	0	0	3
	Dab	0	0	0	145	6	0	0	12
	Plaice	0	0	0	4	135	0	0	6
	Sole	0	0	0	1	0	6	0	3
	Turbot	0	0	0	0	0	0	6	0
	Background	0	2	0	0	1	0	0	NA

From experience, identifying the species based on the ventral side of demersal fish is performed poorly by humans compared to the dorsal view, due to the absence of well-identifiable patterns on the ventral side. This can also be seen in the performance of the Fish-species network. Figure 8 shows the percentage of correct detections, misclassifications and non-detections for the dorsal side (left) and the ventral side (right). Misclassifications occur more on the ventral side, 11% compared to 3% in case of the dorsal side. Nevertheless, the network was still able to classify about 84% of the fish in the test set from the ventral side, compared to 92% on the dorsal side.

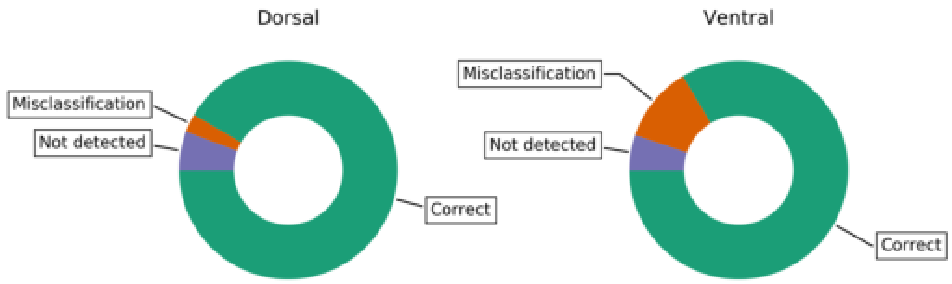

Figure 8 Performance of the Fish-species network. The test set contain 332 dorsal views and 231 ventral views of the fish. Evaluation was made based on whether the detection was correct, misclassified, or the fish was not detected. For dorsal, 92% was correct, 3% was misclassified, and 6% was not detected. For ventral, 84% was correct, 11% was misclassified, and 6% was not detected.

Conclusion and discussion

In this paper, we investigated a current deep-learning method can deal with the complexities present in on-board catch monitoring. Three conclusions can be drawn: 1) the number of fish species could be increased without loss in performance, 2) fish singulation are not necessary as the composition complexity could be increased in most cases without loss in performance, 3) similarly looking ray and flat fish species could be detected, even when only the white ventral side was visible.

Our collected data represent a simplification of on-board situation. Images of fish were taken in a controlled environment without benthos and debris, which is often part of the catch in demersal-mixed fisheries. More data shall likely further improve the results. This would especially be in the case of more complex situation. For an automated monitoring system on-board of a fishing vessel, more data is indeed required. However, the goal of this research is not to achieve optimal performance. Rather, we investigate the changes of the performance as a function of the complexity. In our work, neither the complexity in composition nor the number of detection of fish species are significant. Adding more training data will likely diminish the gap in performance between the simple and complex situations. Hence, our conclusions will not change by adding more training data.

Nevertheless, for human observers, recognising and identifying species in video footage with complex compositions is challenging, but for our method it is not. Although our method makes errors, the benefit is the potential to observe the full catch, thereby increasing the sample size for stock assessment by an order of magnitude compared to current practices. Of course, fish in video footages should at least be partly visibly to enable detection. Still, all video footages taken from the fishing vessels could be processed with known error margins, whereas current on-board or video observers generally identify species and record their choice. Since the error margins in the choices of observers is unknown, extrapolating count estimation based on small samples of the catch gives a false sense of accuracy. This is especially relevant in fisheries with rare species for which misidentification is common (van Helmond *et al.*, 2015; van Helmond *et al.*, 2017). Thus, deep-learning methods could provide an important element compared to observer programs: the probabilities of correct species identification.

In theory, these probabilities can be used to inform stock assessment models about the accuracy of the catch information, which can then be propagated into the estimates of population size and mortality rates.

Currently, our method detects fish in images, and counting fish in a sequence of images would be a simple extension. For most stock assessments these counts would have to be converted to weights in order to obtain estimates of total catch in weight. Previous studies have already developed approaches that estimate weights based on lengths (Robinson *et al.,* 2010; Froese *et al.*, 2013). In future research, fish length could be estimated based on images (White *et al.*, 2006), even if individuals are partly covered by other fish in the catch. Deep-learning methods can be developed to estimate size and weight simultaneously. Cumulating the body weights in the samples will give an estimate of the total catch in weight.

More complex assessments require length or age stratification of catches to disentangle the effects of birth, growth, and mortality on population dynamics (Beverton and Holt, 1957; Hillary *et al.*, 2010). Age information is generally obtained from growth rings in hard structures inside the body, and these cannot be obtained from the images. Meanwhile, deep-learning techniques could in theory be able to estimate age from morphological characteristics of the individuals beyond length, if such characteristics exist.

In conclusion, this study provides a crucial step in innovating data collection in commercial fisheries by exploring deep-learning techniques in fish-catch monitoring, which could contribute to stock assessments, and eventually, more sustainable fisheries.

References

Allken, V., Handegard, N.O., Rosen, S., Schreyeck, T., Mahiout, T., Malde, K. 2018. Fish species identification using a convolutional neural network trained on synthetic data. ICES Journal of Marine Science, 76: 342-349.

Bengio, Y. 2012. Deep learning of representations for unsupervised and transfer learning. Journal of Machine Learning Research – Proceedings track, 27: 17-36.

Beverton, R.J.H., Holt, S.J. 1957. On the Dynamics of Exploited Fish Populations. Her Majesty's Stationery Office, London (UK).

Bradshaw, C.J.A., Prowse, T.A.A., Drew, M., Gillanders, B.M., Donnellan, S.C., Huveneers, C. 2018. Predicting sustainable shark harvests when stock assessments are lacking. ICES Journal of Marine Science, 75: 1591-1601.

Broadhurst, M.K., Suuronen, P., Hulme, A. 2006. Estimating collateral mortality from towed fishing gear. Fish and Fisheries, 7: 180-218.

Daan, N., Heessen, H.J.L., and ter Hofstede, R. 2005. North Sea Elasmobranchs: distribution, abundance and biodiversity, ICES CM 2005/N:06.

Dickey-Collas, M., Pastoors, M.A., Van Keeken, O.A. 2007. Precisely wrong or vaguely right: Simulations of noisy discard data and trends in fishing effort being included in the stock assessment of North Sea plaice. ICES Journal of Marine Science, 64: 1641-1649.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 580-587.

Girshick, R. 2015. Fast r-cnn. Proceedings of the IEEE international conference on computer vision and Pattern Recognition, 1440-1448.

de Haan, D., Fosseidengen, J.E., Fjelldal, P.G., Burggraaf, D. and Rijnsdorp, A.D. 2016. Pulse trawl fishing: characteristics of the electrical stimulation and the effect on behaviour and injuries of Atlantic cod (Gadus morhua). ICES Journal of Marine Science, 73: 1557-1569.

Ellis, J.R., Cruz-Martinez, A., Rackham, B.D., and Rogers, S.I. 2005. The Distribution of Chondrichthyan Fishes Around the British Isles and Implications for Conservation. Journal of Northwest Atlantic Fishery Science, 35: 195-213.

Harry, A.V., Tobin, A.J., Simpfendorfer, C.A., Welch, D.J., Mapleston, A., White, J., Williams, A.J., Stapley, J. 2011. Evaluating catch and mitigating risk in a multispecies, tropical, inshore shark fishery within the Great Barrier Reef World Heritage Area. Marine and Freshwater Research, 62: 710-721.

van Helmond, A.T., Chen, C., and Poos, J.J. 2015. How effective is electronic monitoring in mixed bottom-trawl fisheries? ICES Journal of Marine Science, 72: 1192-1200.

van Helmond, A.T.M., Chen, C. and Poos, J.J. 2017. Using electronic monitoring to record catches of sole (Solea solea) in a bottom trawl fishery. ICES Journal of Marine Science, 74: 1421-1427.

Froese, R., Thorson, J. T., and Reyes Jr, R. B. 2014. A Bayesian approach for estimating length-weight relationships in fishes. Journal of Applied Ichthyology, 30: 78-85.

He, K., Zhang, X., Ren, S., and Sun, J. 2016. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770-778,

Hillary, R.M., Edwards, C.T.T., Mitchell, R.E., and Agnew, D.J. 2010. Length-based assessment for mackerel icefish (Champsocephalus gunnari) at South Georgia. CCAMLR science journal of the Scientific Committee and the Commission for the Conservation of Antarctic Living Resources, 17: 129-137.

Hold, N., Murray, L.G., Pantin, J.R., Haig, J.A., Hinz, H. and Kaiser, M.J. 2015. Video capture of crustacean fisheries data as an alternative to on-board observers. ICES Journal of Marine Science, 72: 1811–1821.

Hu, J., Li, D., Duan, Q., Han, Y., Chen, G. and Si, X. 2012. Fish species classification by color, texture and multi-class support vector machine using computer vision. Computers and electronics in agriculture, 88: 133-140.

ICES, 2018. Report of the Working Group on Elasmobranch Fishes (WGEF), 19-28 June 2018, Lisbon, Portugal. ICES CM 2018/ACOM:16.

Kindt-Larsen, L., Kirkegaard, E., and Dalskov, J. 2011. Fully documented fishery: a tool to support a catch quota management system. ICES Journal of Marine Science, 68: 1606-1610.

Krizhevsky, A., Sutskever, I., and Hinton, G.E. 2012. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 1097-1105.

LeCun, Y., Bengio, Y., and Hinton, G. 2015. Deep learning. Nature, 521:436.

Li, X., Shang, M., Hao, J., and Yang, Z. 2016. Accelerating fish detection and recognition by sharing CNNs with objectness learning. OCEANS 2016-Shanghai, 1-5.

Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. 2017. Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2117-2125).

Lu, Y., Chen, T., and Kuo, Y. 2019. Identifying the species of harvested tuna and billfish using deep convolutional neural networks. ICES Journal of Marine Science.

Marini, S., Fanelli, E., Sbragaglia, V., Azzurro, E., Fernandez, J. D. R., and Aguzzi, J. 2018. Tracking fish abundance by underwater image recognition. Scientific reports, 8: 13748.

Mortensen, L.O., Ulrich, C., Olesen, H.J., Bergsson, H., Berg, C.W., Tzamouranis, N., and Dalskov, J. 2017. Effectiveness of fully documented fisheries to estimate discards in a participatory research scheme. Fisheries Research, 187: 150-157.

Needle, C.L., Dinsdale, R., Buch, T.B., Catarino, R.M.D., Drewery, J. and Butler, N. 2015. Scottish science applications of Remote Electronic Monitoring. ICES Journal of Marine Science, 72: 1214–1229.

Oliver, S., Braccini, M., Newman, S.J., and Harvey, E.S. 2015. Global patterns in the bycatch of sharks and rays. Marine Policy, 54: 86-97.

Poos, J.J., Aarts, G., Vandemaele, S., Willems, W., Bolle, L.J., van Helmond, A.T.M. 2013. Estimating spatial and temporal variability of juvenile North Sea plaice from opportunistic data. Journal of Sea Research, 75: 118-128.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. 2016. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 779-788.

Redmon, J. and Farhadi, A. 2018. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Ren, S., He, K., Girshick, R., and Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems. 91-99.

Rijnsdorp, A.D., Daan, N., Dekker, W., Poos, J.J., and van Densen, W.L.T. 2007. Sustainable use of flatfish resources: addressing the credibility crisis in mixed fisheries management. Journal of Sea Research, 57: 114–125.

Robinson, L. A., Greenstreet, S. P. R., Reiss, H., Callaway, R., Craeymeersch, J., De Boois, I., Degraer et al. 2010. Length–weight relationships of 216 North Sea benthic invertebrates and fish. Journal of the Marine Biological Association of the United Kingdom, 90: 95-104.

Shafait, F., Mian, A., Shortis, M., Ghanem, B., Culverhouse, P.F., Edgington, D., Cline, D., Ravanbakhsh, M., Seager, J., and Harvey, E.S., 2016. Fish identification from videos captured in uncontrolled underwater environments. ICES Journal of Marine Science, 73: 2737-2746.

Siddiqui, S.A., Salman, A., Malik, M.I., Shafait, F., Mian, A., Shortis, M.R., and Harvey, E.S. 2017. Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data. ICES Journal of Marine Science, 75: 374-389.

Stanley, R.D., Karim, T., Koolman, J., and McElderry, H. 2015. Design and implementation of electronic monitoring in the British Columbia groundfish hook and line fishery: a retrospective view of the ingredients of success. ICES Journal of Marine Science, 72: 1230–1236.

Stevens, J.D., Bonfil, R., Dulvy, N.K., and Walker, P.A. 2000. The effects of fishing on sharks, rays, and chimaeras (chondrichthyans), and the implications for marine ecosystems. ICES Journal of Marine Science, 7: 476-494.

Stobutzki, I.C., Miller, M.J., Heales, D.S., Brewer, D.T. 2002. Sustainability of elasmobranchs caught as bycatch in a tropical prawn (shrimp) trawl fishery. Fishery Bulletin, 100: 800-821.

Storbeck, F. and Daan, B. 2001. Fish species recognition using computer vision and a neural network. Fisheries Research, 51: 11-15.

Stratoudakis, Y., Fryer, R. J., Cook, R. M., and Pierce, G. J. 1999. Fish discarded from Scottish demersal vessels: Estimators of total discards and annual estimates for targeted gadoids. ICES Journal of Marine Science, 56: 592–605.

Wang, M., Liu, M., Zhang, F., Lei, G., Guo, J., and Wang, L. 2018. Fast Classification and Detection of Fish Images with YOLOv2. 2018 OCEANS-MTS/IEEE Kobe Techno-Oceans, 1-4.

White, D.J., Svellingen, C., and Strachan, N. J. 2006. Automated measurement of species and length of fish by computer vision. Fisheries Research, 80: 203-210.

Zion, B., Shklyar, A., and Karplus, I. 1999. Sorting fish by computer vision. Computers and electronics in agriculture, 23:175-187.

Wageningen Marine Research
T +31 (0)317 48 7000
E: marine-research@wur.nl
www.wur.eu/marine-research

Visitors’ address
- Ankerpark 27 1781 AG Den Helder
- Korringaweg 7, 4401 NT Yerseke
- Haringkade 1, 1976 CP IJmuiden

With knowledge, independent scientific research and advice, **Wageningen Marine Research** substantially contributes to more sustainable and more careful management, use and protection of natural riches in marine, coastal and freshwater areas.

Wageningen Marine Research is part of Wageningen University & Research. Wageningen University & Research is the collaboration between Wageningen University and the Wageningen Research Foundation and its mission is: 'To explore the potential for improving the quality of life'