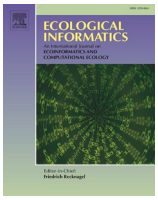
Ecological Informatics 67 (2022) 101495

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



Identification and measurement of tropical tuna species in purse seiner catches using computer vision and deep learning

Xabier Lekunberri [a,b,*], Jon Ruiz [a], Iñaki Quincoces [a], Fadi Dornaika [b,c], Ignacio Arganda-Carreras [b,c,d], Jose A. Fernandes [a]

[a] AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), Txatxarramendi ugartea z/g, 48395 Sukarrieta, Bizkaia, Spain
[b] University of the Basque Country (UPV/EHU), San Sebastian, Spain
[c] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
[d] Donostia International Physics Center (DIPC), San Sebastian, Spain

ARTICLE INFO

Keywords:
Computer vision
Deep learning
Fisheries
Electronic monitoring
Tropical tuna species identification

ABSTRACT

Fishery monitoring programs are essential for effective management of marine resources, as they provide scientists and managers with the necessary data for both the preparation of scientific advice and fisheries control and surveillance. The monitoring is generally done by human observers, both in port and onboard, with a high cost involved. Consequently, some Regional Fisheries Management Organizations (RFMO) are opting for electronic monitoring (EM) as an alternative or complement to human observers in certain fisheries. This is the case of the tropical tuna purse seine fishery operating in the Indian and Atlantic oceans, which started an EM program on a voluntary basis in 2017. However, even when the monitoring is conducted though EM, the image analysis is a tedious task manually performed by experts. In this paper, we propose a cost-effective methodology for the automatic processing of the images already being collected by cameras onboard tropical tuna purse seiners. Firstly, the images are preprocessed to homogenize them across all vessels and facilitate subsequent steps. Secondly, the fish are individually segmented using a deep neural network (Mask R-CNN). Then, all segments are passed through other deep neural network (ResNet50V2) to classify them by species and estimate their size distribution. For the classification of fish, we achieved an accuracy for all species of over 70%, i.e., about 3 out of 4 individuals are correctly classified to their corresponding species. The size distribution estimates are aligned with official port measurements but calculated using a larger number of individuals. Finally, we also propose improvements to the current image capture systems which can facilitate the work of the proposed automation methodology.

1. Introduction

Catches of tropical tuna purse seine fishery generally include several target species (*Thunnus albacares*, yellowfin tuna; *Katsuwonus pelamis*, skipjack tuna), not targeted but significatively caught (*Thunnus obesus*, bigeye tuna), accompanied in different proportions by other secondary species. These secondary species include both tuna-like species (neritic tunas) and other bycatch species (billfishes, sharks, or other bony fishes) (Amandè et al., 2010; Ruiz et al., 2018). Purse seine catches make up nearly 62% of the 4.2 million tons of tuna caught globally every year (Restrepo and Forrestal, 2012) and it is the fishing gear that contributes most to the catch of yellowfin and skipjack globally (Majkowski et al., 2011). Every time a school of tunas is caught, an 8 to 10 tons capacity brail is hoisted on board. Each brail is emptied onto a conveyor belt that ends in a well (tank) where the fish is frozen. The vessels have different independent wells, where the catches from one or more fishing sets are stored. During this process, an EM camera captures images of the conveyor belt every 2 seconds. Once the fishing operation is completed, species composition is estimated and reported by the crew in the onboard logbook. However, bias in logbooks has been evidenced since the beginning of the tropical tuna purse seine fishery, particularly for bigeye and yellowfin juveniles (Cayré, 1984; Fonteneau, 1976; Fonteneau, 2008). This is due to the difficulties in species identification (Lawson, 2009), which prevent their use as accurate values both for scientific and

* Corresponding author.
E-mail addresses: xlecumberri001@ikasle.ehu.eus (X. Lekunberri), jruiz@azti.es (J. Ruiz), iquincoces@azti.es (I. Quincoces), fadi.dornaika@ehu.eus (F. Dornaika), ignacio.arganda@ehu.eus (I. Arganda-Carreras), jfernandes@azti.es (J.A. Fernandes).

https://doi.org/10.1016/j.ecoinf.2021.101495
Received 16 August 2021; Received in revised form 19 November 2021; Accepted 21 November 2021
Available online 25 November 2021

compliance purposes. Consequently, these onboard estimations are reevaluated with species composition estimates based on at sea or port sampling schemes for management and scientific purposes (Duparc et al., 2019; Lawson, 2009). Despite this, such sampling methods have been also criticized and several limitations have been identified. Previous studies have shown that fish are not randomly selected for sampling either due to physical constraints, such as layering in the set, brail or well, or to behavior, such as samplers tending to select certain species and/or sizes of fish more than others (Lawson, 2009). Similarly, the excessively large spatiotemporal strata used to design the port sampling scheme has been identified as a potential bias to estimate the species composition (Duparc et al., 2018).

EM is an emerging field that has developed rapidly over the last decade, with high potential as a cost-effective tool to complement the current RFMO monitoring programs (Fujita et al., 2018; McElderry, 2008; Michelin et al., 2020; Van Helmond et al., 2020). It generally consists of several sensors such as a hydraulic pressure sensor or rotational sensor, GPS (Global Positioning System) and cameras, which allow the monitoring of fishing activity through heterogeneous data. This positioning and image data can be used to estimate the fishing effort, date and position of fishing operations, catches, discards, or interaction with protected species. The tropical tuna purse seine fishery has conducted several pilot studies to determine the effectiveness of EM technology (Briand et al., 2018; Gilman et al., 2019; Monteagudo et al., 2014; Murua et al., 2020b; Ruiz et al., 2015, 2016). In general, these experiences showed that species identification and their size composition are difficult to estimate through EM. Differentiating between small yellowfin from bigeye or the overlap of individuals on the conveyor belt are some of the main challenges. Also, vessels using EM equipment store imagery on hard drives, which are sent after the trip to land-based video analysis stations that process and extract key information. This time-consuming procedure can slow the data availability by several months, due to disk shipment and processing delay. This is a concern when near real time data is essential, particularly for management programs that have catch limits by species, such as the yellowfin tuna in the Indian Ocean (IOTC, 2019) or the bigeye tuna in the Atlantic Ocean (ICCAT, 2019).

The use of image analysis through artificial intelligence for automatic species identification has been identified as the next step in the use of EM that could increase the accuracy of catch estimates involved in tuna fisheries (Gilman et al., 2019), but has not yet been tested for tuna catches. Some work has been done on using deep learning techniques to automatically extract EM video segments of catch events from commercial fishing vessels, which substantially reduces storage space and review time by analysts (Qiao et al., 2020), but without identification of species. Most of the work on fish catch analysis has aimed for individual fish under a controlled environment where individuals are separated manually (Saberioon and Císař, 2018; Yu et al., 2020). For example, in Yu et al. (2020) each fish is digitalized individually, always at certain distance from the camera, with controlled lighting and fish position for size estimation, but without species identification. French et al. (2020) used a less controlled environment and aimed at the classification of species, but the accuracy across species varied greatly (between 17% and 90%). Successful applications of machine learning in marine science to identify organisms such as plankton (Bachiller and Fernandes, 2011; González et al., 2017; Hirata et al., 2016) or identify fish in uncontrolled environments such as underwater imaging (Cui et al., 2020; Rathi et al., 2017), show the potential to improve current tuna catch estimations by means of image analysis and machine learning.

Although the previous studies permit a degree of optimism about improving EM systems for the automatic identification of species, tuna on conveyor belts is more challenging since (1) the fish is highly aggregated in different layers, partially overlaps, (2) lighting is poor, and (3) current images are of low quality as the lens can be dirty and the cameras are placed at variable angles and distances. The first challenge for this type of work is the segmentation of the individual fishes. Image morphology methods such as watershed and K-means clustering (Chuang et al., 2013; Yao et al., 2013) were often used in the past. These methods take specific pixels of the image as starting points and advance iteratively based on the morphology of the elements of the image. They are usually easier to train than those currently used, but the results are not as good. Nowadays, approaches based on deep learning are being used more frequently (Alsmadi and Almarashdeh, 2020). These do not rely on hand-crafted descriptors of the fishes but usually require a lot more data to achieve good results. The Region-based Convolutional Neural Networks (R-CNN) family is one that is emerging in the state of the art (Girshick, 2015; Girshick et al., 2014; He et al., 2018; Ren et al., 2017; Zhao et al., 2019). To be specific, we used a Mask R-CNN, which is an extension of the Faster R-CNN and is able to perform instance segmentation (per-pixel level segmentation) of the objects. A Faster R-CNN is made up of two stages: (1) the Region Proposal Network (RPN), where candidate object bounding boxes are proposed and (2) a convolutional neural network, where features from each candidate box are extracted and the classification and bounding-box regression is performed. The Mask R-CNN extends this model by predicting a binary mask for each bounding box (and semantic class) in parallel to the second stage.

The aim of this study is to estimate the tropical tuna purse seine tuna catch composition and size distribution by fishing set using image analysis and automatic classification of catches digitalized with current EM systems installed onboard. This makes possible the use of a large set of historical and ground data information (port landings sampling) and the use of cheaper digitalization systems. We also suggest some cost-effective changes on currently installed EM systems that could improve the image analysis and automatic classification.

2. Methods

Historical records of images from an EM system during a whole fishing trip, about 30–40 days of vessel activity, consist of around one terabyte of images, but only a small number belongs to fishing events. The number of images captured in each set varies between 100 and 1500 images accordingly with the duration of the fishing operation and the number of fish captured. These images are captured with an image size of 1920 × 1080 pixels and a frame rate of one image every two seconds (0.5 fps) during fishing events. Only one photo per minute is captured when the vessel is not fishing. Analyzing this data can take up to 1 hour per fishing operation for an experienced analyst. This section details the process followed to estimate the distribution by size and species using image analysis and automatic classification of images captured by an EM system onboard a tuna purse seiner. This process and its validation are split in four major modules (Fig. 1, Table 1): A first module where appropriate images are selected, a second module about the segmentation and species classification models training and statistical validation using 5-fold cross-validation, a third module for the prediction of the models, and a last module where results are contrasted with independent "ground truth" based on port sampling of landings.

The *training/validation module* consists of the following steps: (1) the selected images are annotated by experienced analysts (delineating and assigning each fish to a class); (2) two different models are trained with these data: One for automatic fish segmentation and another for classification by species of these segments. For manual annotation, 100 images from four different fishing sets were annotated, obtaining a total of 975 individual fishes (segments).

The *prediction module* uses images that have not been previously utilized in the *train/validation module* from 22 fishing sets. First, these images are processed by the segmentation model in order to detect the segments (each individual fish) within each image; next, these new segments are processed by the classification model to identify the fish species; and finally, the segments are individually measured to obtain the species catch composition and size distribution by set (each fishing event).

Both modules share common steps such as image and segment

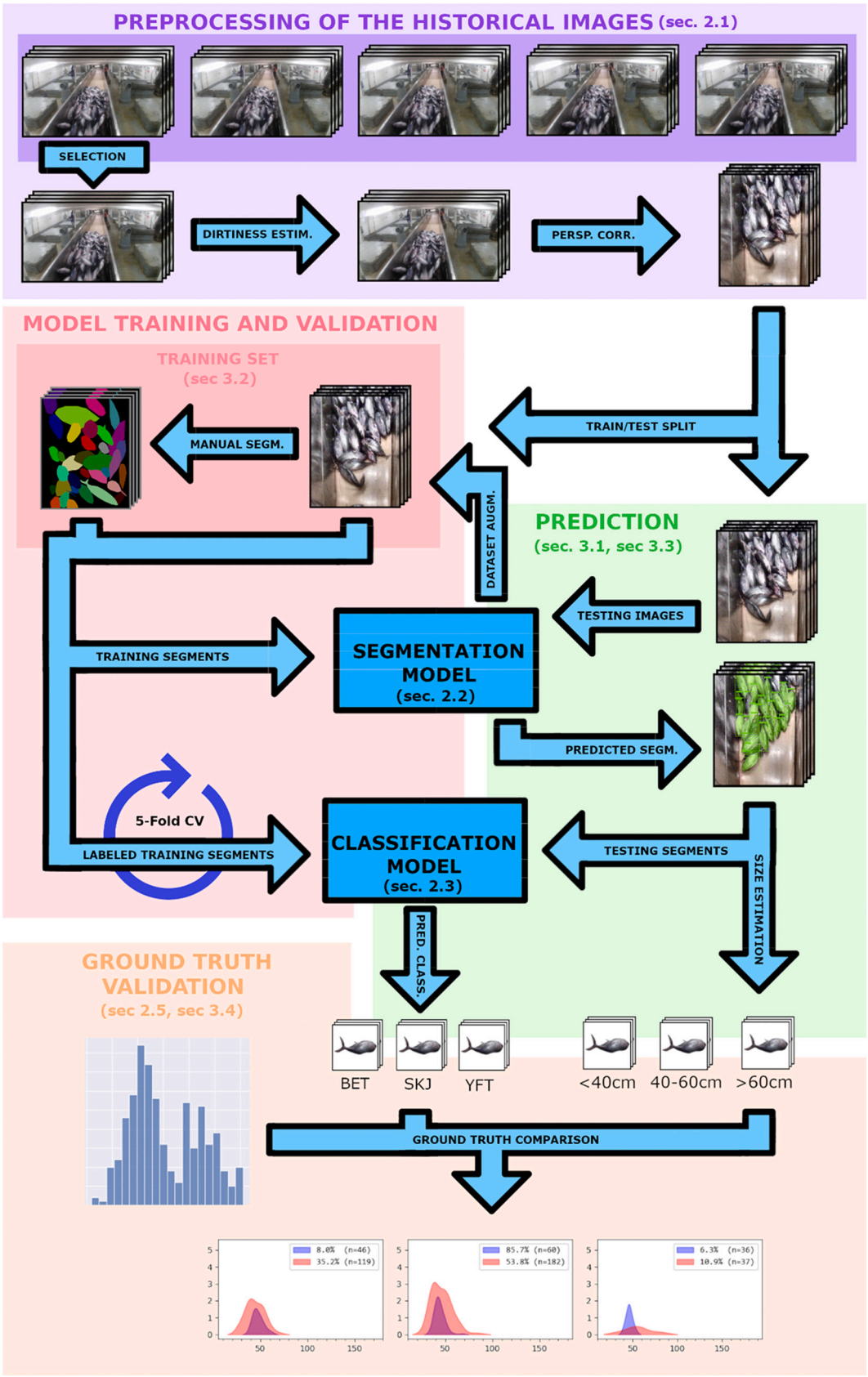


Fig. 1. Diagram of our methodology. There are four main modules: Preprocessing of historical images (purple, *sec.* 2.1), model training and validation (red, sec. 2.2 and sec. 2.3), prediction making (green, sec. 3.1 and sec. 3.3) and ground truth validation (orange, sec. 2.5 and sec. 3.4). The original image database (dark purple) and the training set (dark red, sec. 3.2) are also highlighted. Data used in each module is specified in Table 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Summary of the data processed in each of the modules and its source. More details of the modules in Fig. 1. The images used in this work are captured on fishing vessels equipped with electronic monitoring (EM) cameras. Once the analysts manually review them, the images are stored in a database and are organized by fishing sets (each fishing set has a correspondent image set).

Module	Input	Source of the input	Output
Preprocessing of the historical images (*sec.* 2.1)	Image database	EM cameras of the fishing vessels	Images of 30 fishing sets
Model training and validation (Segmentation, sec. 2.2)	100 images (975 segments) - Manually: 975	4 of the 30 fishing sets (partially annotated)	Trained model
Model training and validation (Classification, *sec.* 2.3)	718 images (86,202 segments) - Manually: 275 segments - Automatically: 6921 segments - Data augmentation: x12	4 of the 30 fishing sets (annotated)	Trained model
Prediction (sec. 3.1, *sec.* 3.3)	- 14,896 images - Two trained models	22 of the 30 fishing sets (unannotated)	Data of the 22 fishing sets - Annotations - Species prediction - Size distribution
Ground truth validation (sec. 2.5)	Data from 4 fishing sets	- Official sources - Estimations	Comparison

preprocessing, where the perspective and contrast of the images are corrected to obtain better results. However, the use of the model is different in both modules since the *training/validation module* builds and validates statistically the model with repeated cross-validation and the *prediction module* uses the full train data model to identify the species. The third module, *ground truth comparison*, compares the species composition obtained through the classification model and size distribution estimates, versus the routine sampling program done manually by scientific staff in the main landing ports. For this comparison, four fishing sets were used, where both the EM data and routine port sampling data was available.

2.1. Preprocessing

Lens dirtiness estimation is needed since the images are taken with a camera on top of a conveyor belt containing live fish where splashes of water drops and fish scales are common (Fig. 2a, Fig. 2b). The variance of each pixel in the image over an entire set is used as a method of determining the level of dirtiness of a camera. A dirty region will have a low variance since dirt accumulates but does not change position, while a clean region will have a high variance due to the movement of the conveyor belt transporting the fish. Knowing the dirtiness of a camera enables sets or images that will not provide useful segments to be discarded.

Once the images that are not useful are filtered, a perspective correction is applied. As these cameras have been placed for image review by humans, factors such as location and perspective are not optimal for automatic processing of the images (Fig. 2c, Fig. 2d). For this study the region of interest is the conveyor belt but simply discarding the rest of the image is not enough, as the image will still be distorted. The conveyor belt is rectangular, so using the algorithm described by Zhang and He (2007) it is possible to rectify it into a rectangle and to obtain an image that would resemble one captured by a zenith-angle camera. To make this transformation the four vertices (pixels) that create the trapezium of the conveyor belt were identified (Fig. 2c, blue dots), and transformed to each corner of the rectified image (Fig. 2d).

Contrast enhancement was done to make the automatic segmentation easier, as a higher contrast allows the outline of each individual to be clearer and easier to delimit. The technique applied is Contrast Limited Adaptive Histogram Equalization (CLAHE; (Reza, 2004)). This method equalizes the histogram of the image locally (in small regions) instead of globally (Fig. 2e, Fig. 2f). We chose a window size of 5 pixels and 2 as the histogram clip limit.

2.2. Segmentation and annotation of images

Manual segmentation and annotation of images was done by experienced onboard analysts using the open-source tool CVAT (Sekachev et al., 2019). It allows the images that are going to be annotated to be uploaded to a joint repository where the experts can work together. A total of 975 segments were extracted from the images (Fig. 3). In 275 of these segments species could be identified by the experts, whereas in the other 700 the species could not be identified. This first set of segments was used for two purposes: (1) as a reference for the segmentation model to learn what a fish is (all 975 segments), and (2) to train a species classification model (the 275 segments with class) distributed as yellowfin tuna (YFT) with 143 segments, skipjack tuna (SKJ) with 127 segments and, bigeye tuna (BET) with 5 segments.

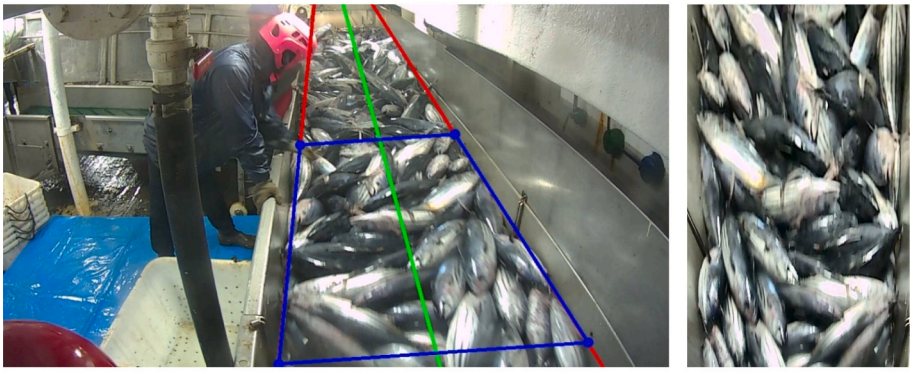
The Automatic segmentation uses a neural network-based method called Mask R-CNN (He et al., 2018), selected for its successful application in other similar domains (French et al., 2020; Yu et al., 2020). We have used a previously trained model from the Tensorflow Object Detection API, since it needs fewer images compared to a segmentation model trained from scratch (training parameters in Table 2). This model is known as "Mask R-CNN Inception ResNet V2 1024x1024" and has been trained with the COCO database (Huang et al., 2017). Transfer learning is based on employing a neural network trained on a dataset to recognize new objects that were not present on it by retraining it with these new objects (Torrey and Shavlik, 2010). In other words, transfer learning through fine-tuning was used. In order to fine-tune a neural network, all layers of the neural network are frozen except the last one. Freezing these layers prevents them from being updated during training, thus maintaining knowledge of the previous training. In this way only the final layer is updated, which is in charge of making the final prediction of fish species labels. This allows a network designed to segment generic objects to specialize in our fish segmentation problem.

Mask R-CNN can segment objects and classify them in the same process. However, we decided to split both steps since the manually annotated segments were not enough for the model to distinguish between different tuna species. Therefore, all these segments were temporarily relabeled to "fish" in this first segmentation process. This way the model can learn to segment correctly a fish without false positives regardless of its species by using fewer examples. A second model (a stand-alone ResNet) is used to classify those "fish" segments into the different tuna species. The first training set with manually labeled segments was expanded with monospecific fishing sets segmented using a model trained with the manually segmentation (see *sec.* 3.2). A monospecific set is a fishing event where a school of fishes containing a single species is captured (and digitalized), mainly of yellowfin tuna and skipjack to a lesser extent.

2.3. Fish classification by species

A classifier based on deep neural networks was also used for this task. In this case, the chosen architecture is a residual neural network, whose main characteristic is that it can skip some of its layers (He et al., 2016). In particular, the implementation of ResNet50V2 model in TensorFlow and pre-trained with the ImageNet dataset (training parameters in Table 2). A problem encountered in initial tests related to species

(a) Dirty camera lens (b) Clean camera lens



(c) Perspective correction guidelines (d) Corrected image

(e) Corrected image with normal contrast (f) Corrected image with enhanced contrast

Fig. 2. From left to right and top to bottom: A) Image taken from a camera when it is dirty. This is an extreme example of dirtiness. B) With the camera cleaned. The region of interest is the central conveyor belt. C) Guidelines for perspective correction. The inner area of the blue trapezoid will form the corrected image. D) Same image as before but with the perspective corrected. E) Corrected image with normal contrast and F) corrected image with enhanced contrast. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

classification was that poorly segmented fish made the classification inconsistent. The model was not able to work with defective segments. As the model is able to predict as many classes as necessary, these defective segments were used to add three new representative classes to our model, making a total of 6 classes. These new classes represent segments made of mostly fish heads (HEAD), segments of mostly fins (FIN) and other segments that do not fit these descriptions but are not whole individuals either (ART). For a species-only classification it may be possible to work with non-complete segments, but this separation also helps to make a correct size distribution estimation.

Data augmentation techniques were used to obtain a larger number of segments. For each of the original segments 12 new ones are generated, varying the orientation of the individual (image rotation), the contrast and the brightness of the image. To generate these augmentations, first each individual was placed so that its longest axis was horizontal. This way each fish could be found randomly in a left-right hand view combination and with its head up or down. By generating the three missing rotations, all the possible positions in which the model can find a fish are generated. Finally, a change in brightness and contrast was randomly applied to each of these four images. This two parameters were altered using Eq. 1. Each pixel x of the image is multiplied by α to alter the contrast and β is added to change the brightness. The possible values for α and β were $\{\alpha \in \mathbb{R} : 0.9 \leq \alpha \leq 1.1\}$ and $\{\beta \in \mathbb{N} : -4 \leq \beta \leq 4\}$.

$$g(x) = \alpha f(x) + \beta \tag{1}$$

2.4. Fish size estimation

As regards fish measurement, different sizing methods such as fork length (FL) or first predorsal length (LD1) can be used (ICCAT, 2006). Size estimation of each individual was based on the distance in pixels from the head to the tail, so this was assumed to be FL. The pixel-to-meter ratio used for size estimation was calculated based on the width of the conveyor belt (e.g., the width of the belt it is known in pixels and

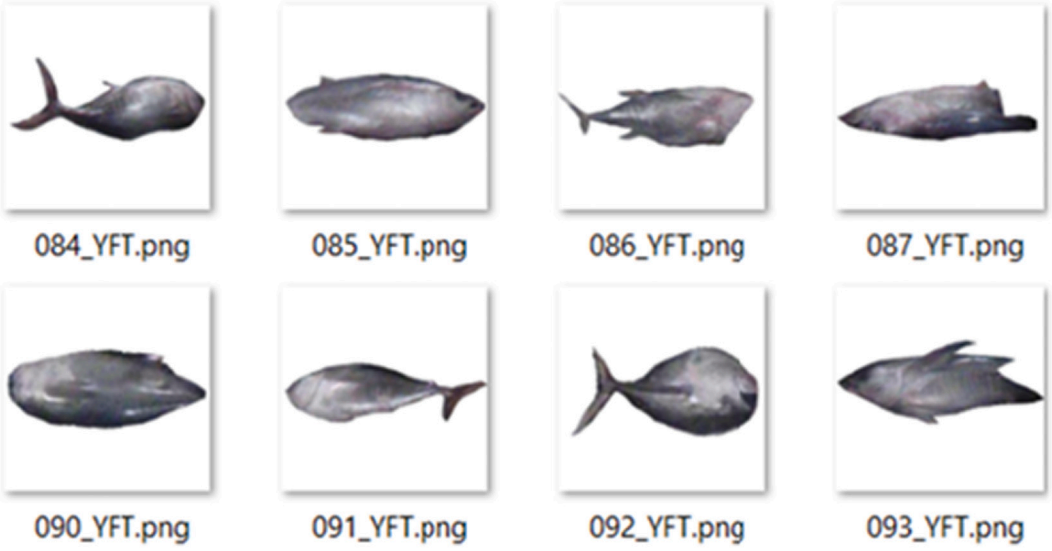


Fig. 3. Example of individual fishes once the segmentation is done. These fishes belong to a monospecific set of YFT.

Table 2
Computing infrastructure, search space and best assignments of hyperparameters of the neural networks.

(a) Hyperparameters used for the training of each neural network. The base models have been obtained from TensorFlow Object Detection API (segmentation) and the base implementation of TensorFlow (classification)

Hyperparameter	Module	Search space	Best assignment
Epochs	Segmentation	[1000−20,000]	15,000
	Classification	[500–5000]	500
Batch size	Segmentation	[2−32]	4
	Classification	[2–128]	128
Learning rate	Segmentation	[0.0001–0.01]	0.008
	Classification	[0.0001–0.01]	0.0001
Activation	Segmentation	Softmax	Softmax
	Classification	Softmax	Softmax
Train/test split	Segmentation	[0.9/0.1–0.8/0.2]	0.9/0.1
	Classification	[0.9/0.1–0.8/0.2]	0.8/0.2

(b) Computing infrastructure used to train both neural networks

Infrastructure	Nvidia Tesla V100 (16GB)	
Search strategy	Manual tuning	
Training duration	Segmentation	4 h 6 min 24 s
(best assignment)	Classification	9 h 22 min 34 s

meters). For the size distribution estimation to be as accurate as possible, the measured individuals must be fully visible from head to tail (Fig. 4). The creation of auxiliary classes at classification time was a first step to filter out invalid segments, but two other methods were also used to validate segments: setting a minimum size and a minimum major-to-minor axis ratio. The selected smallest size is 20 cm, as this is the lowest value for which ground truth measurements exist. The ratio selected has been 4, as this is the ratio established by the International Commission for the Conservation of Atlantic Tunas (ICCAT) in their manual (ICCAT, 2006). All individuals that do not meet these criteria were not used to compute the size distribution.

2.5. Ground truth validation

Finally, the species catch composition estimates obtained through the species classification model and the size distribution by species were compared with the ground truth. We understand ground truth as referring to the official estimates that are currently used for the management of the European tropical purse seine fishery, which are

Fig. 4. Overlapping individuals (red) and a fully visible individual (green). The red ones will be discarded for further estimations. The green square is the one used for the measurement. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

obtained through a port sampling program. This sampling strategy for estimating the composition of catches by species and sizes in the purse seine tropical tuna fisheries was established in 1998, and although it has been slightly adapted, it is still used today by European scientists (Duparc et al., 2019). When a tuna vessel arrives at port, sampling begins with the selection of the well to be sampled. Although some wells may contain catches from different sets, those that store catches from a single set are prioritized. For each well selected, a sample of 500 individuals is randomly selected and sampled (i.e., species id and size measurement). Sampling is conducted during the routine fish landing process in two batches while it is still frozen: 300 individuals first and 200 individuals one hour after the first batch. The measurements taken at port are made either to FL if the fish is small or to LD1 otherwise. Using LD1 to FL conversion factors (ICCAT, 2006) it is possible to transform all of them to FL. Thus, making the comparison between image analysis-based estimates and the port estimates possible. As individuals selected on each sample are random, is unlikely that one fish is selected in both. Therefore, it is not possible to evaluate our estimations accuracy.

3. Results

3.1. Automatic fish segmentation

In order to evaluate our segmentation model, we reported the mean average precision (mAP, ranging from 0 to 1) of the detections. To calculate this metric, it is necessary to define what is considered as a valid detection. This is done by setting an intersection over union (IoU) score that if it exceeded, the segment will be counted as valid. The IoU is measured as the area of the predicted bounding box that overlaps with the ground truth bounding box divided by the total area of both bounding boxes. Once the IoU is set, the mAP is calculated by measuring the area bellow the precision-recall curve of the detections. We calculated three scores for our model: (1) mAP for an IoU of 0.50, (2) mAP for an IoU of 0.75 and (3) the average mAP for the IoUs from 0.05 to 0.95 in increments of 0.05. The mAP was calculated for both bounding boxes and masks. The bounding box detection achieved mAP scores of (1) 0.95, (2) 0.89 and (3) 0.74; while mask detection achieved mAP scores of (1) 0.90, (2) 0.76 and (3) 0.66. Since the mask is calculated after the bounding box detection and depends on it, all mask measurements are slightly lower.

As with all machine learning models, the result achieved will be as good as the data fed to the system. If the information used to learn is biased, the system will learn this bias and apply it in future predictions. Given the nature of the images, one bias that we detected early on was that the analysts annotated individuals with light colors only (as they are the easiest ones to identify and annotate) so the first model we trained ignored a large proportion of the fish. By correcting this bias (annotating more dark-colored individuals), we were able to drastically increase the number of detected fish. Fig. 5 shows the segmentation in one of the test images once the bias was corrected. To further evaluate the results, two more metric were calculated: (1) the surface area of the image occupied by fishes and (2) the average number of fish detected per image.

Under ideal conditions, the surface area occupied by individuals in an image should be close to those obtained in manually labeled images (mean of 43.97% of the covered image, with a standard deviation of 12%). Although the values vary greatly from set to set, some of them exceed the above values. These values together with the dirtiness estimations provide useful information about the ideal conditions of the EM image capture system. When plotting the dirtiness of each set and the surface area occupied by individuals in an image, is obvious that lens cleanliness is an important factor. The trend shown in Fig. 6 confirms that dirtiness directly affects the quality of the results. Past a certain threshold, it will be necessary to discard the ongoing images and clean the camera, since it will not be possible to detect individuals.

Unfortunately, most of the images we used had some dirtiness, and although this is generally not a problem for detection, in some cases it is impossible to obtain satisfactory results. As expected, the values for the sets with the cleanest camera have a higher percentage of pixels assigned to fish. The number of pixels segmented as fish in these sets ranges from 46 to 67%, while the standard deviation is smaller than the ones with the

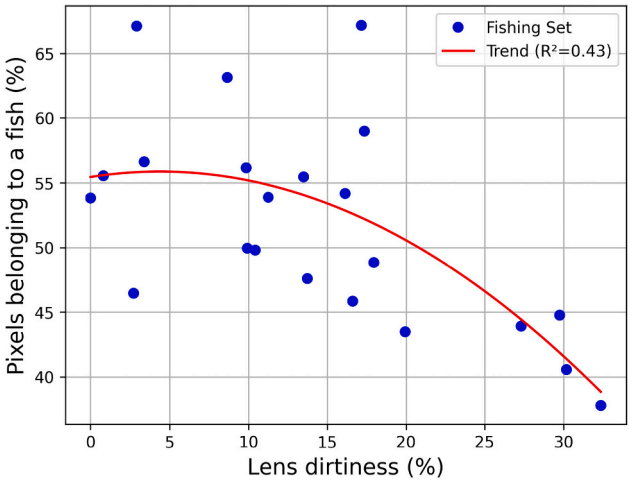


Fig. 6. Relationship between the dirtiness of a lens and the fish that the model is capable of detecting. Some dirtiness does not affect the quality of the results, but they quickly deteriorate if the dirtiness exceeds 10%.

(a) Original image overlapped by the mask (b) Predicted mask

Fig. 5. Automatic segmentation predicted by the trained model. Fig. 5a is used to visually validate our model. Fig. 5b is the one used to crop the images into segments.

dirtiest lenses.

The other estimator that we used to evaluate the quality of the model (number of detected fishes in each image) is shown in Table 3. When analyzing the table, several sets such as Mixed 1 and YFT stand out for the low number of detected fish. These estimator can also be affected by the same reasons as above (dirt, better model, etc.). However, an important element to consider along with the number of detections per image is the size of these segments. As more small fish fit in the same space than large fish, sets with larger segments have fewer fishes per image.

3.2. Improving the training set

A second set of segments was automatically generated the trained segmentation model. With a sufficiently accurate segmentation model and fishing sets where the capture composition is known, it is possible to use automatic segmentation to extend the training set. Three monospecific sets were used for this purpose, from which about 7000 unique segments were obtained. The class of each segment was known, since a monospecific set only contains one species of tuna. A manual review was necessary to classify the invalid segments into their corresponding artifact class. A larger set of images to work on was achieved by mixing the 275 manually annotated specimens that were marked with their species and the automatically segmented monospecific sets. This new dataset is in Table 4. The data augmentation techniques applied allowed us to increase 12 times the number of segments, so we were able to train the classification model with a total of 86,208 segments. Fig. 7 shows how the model is able to learn in fewer steps and with better accuracy if data augmentation is used.

3.3. Fish classification by species

Initially, only the three tuna species were used as target classes and the model managed to correctly classify 3 out of 4 individuals (Fig. 9a). As two datasets from different sources were used (sec. 3.2), the distribution of misclassified segments conditional to the dataset that each segment belongs was tested. This checked possible biases, verifying if the neural network learned to differentiate species, datasets or both. To test this, the model was trained from scratch as follows: (1) trained with manual segments and tested on automatic segments and (2) trained with automatic segments and tested on manual segments. The model trained with the manual dataset only (model 1, Fig. 8a) it was not able to learn or predict correctly, with a mean accuracy of 45%, likely due to the low number of images. Same occurs with model 2 (Fig. 8b), as predictions are not as accurate as training with all segments (Fig. 9a), but superior to model 1. However, model 1 is superior to model 2 where for YFT which is the only species where model 1 has more training data than model 2. Analyzing both models and Table 4 at the same time, looks like accuracy is dependent on the number of training segments per species and not dependent on the dataset itself.

In a second approach, three new auxiliary classes were added. There is no need for an expert to reclassify the segments into these new classes, as anyone can identify an incorrect segment. For this job, it is preferable to minimize false positives in the fish classes at the cost of increasing false negatives (i.e., discard invalid elements in the final set even if we lose part of the useful individuals). In this approach, the overall accuracy of our model was lower, but we manage to get rid of most of the invalid segments. As we will see in sec. 3.4, this allowed us to obtain more accurate estimates that are in agreement with official data.

Table 3
Number of fishes detected and size statistics for some of the sets used in the predict module. Mixed sets include all three species, while the sets labeled with a tuna species name only contains that specie.

Set	Fishes per image		Fish size (cm)	
	Mean	SD	Mean	SD
Mixed 1	4.09	1.89	51.11	24.37
Mixed 2	16.61	4.41	23.83	9.11
Mixed 3	21.6	3.93	36.79	14.7
Mixed 4	21.32	4.78	29.25	9.31
SKJ	10.92	3.54	30.2	11.94
YFT	3.16	1.4	62.28	20.35

Table 4
Number of segments for each of the labels used during the classification by species.

Label	Manual	Automatic	Data augm.	Total
BET	5	117	1342	1464
SKJ	127	314	4851	5292
YFT	143	64	2277	2484
HEAD	–	207	2277	2484
FIN	–	239	2629	2868
ART	–	5980	65,636	71,616

3.4. Comparison against official port measurements

In this section, we present a performance evaluation of the machine learning method using official port measurements. Comparing the degree of accuracy of our estimates fish by fish is laborious, even impossible when dealing with size estimates as we do not have the real size values. Thus, we compared our estimates to global indicators. These indicators are size distribution and species proportion. Fig. 10 is the result of the comparison between our results and official port measurements in four fishing sets. The first two samples (1 and 2) are from YFT monospecific free school (FSC) sets. The other two samples (3 and 4) belong to mixed sets (mostly SKJ), fished associated to floating objects (FOB). The results calculated by species are added to form a global size distribution with the aggregated data for all species.

For the BET class the model detects more fish than there are. This may be because the BET class has the fewest images of all classes, as it is the least fished tuna species. Moreover, this is the class most contaminated by other classes (Fig. 9). Something similar happened with the SKJ images, as our model underestimates the number of these individuals. In the trained model, this class is the one that lost the most images by classifying valid individuals as artifacts (up to 39%). Since the mixed validation sets are mostly made of SKJ, the drop in this percentage is notorious. The third class (YFT) is notable for having two modes. In the monospecific sets of YFT, an approximate mode of 150 cm is observed, while in the others it is about 50 cm. This occurs as the FSC sets are composed of mostly adult specimens and tend to be monospecific, while the FOB tend to catch more juvenile individuals and are usually mixed-species. There is a clear underestimation of size in the adult specimens, but we managed to identify one of the problems that was causing it. As these fish are so large, they do not overlap each other completely, they overlap only with the fins. This phenomenon causes our model to interpret some of the overlapped fish as if they were two independent segments, one of a valid fish and the other of an artifact. Nevertheless, the problem we encountered with the juvenile individuals was that we do not have monospecific sets of this size. So, training the model to correctly predict these two types of individuals as one class is hard. It is clear that for FSC there is still a lot of work to be done, but for the FOB fishing samples, the models' predictions agree with the official data (Fig. 10, last column, last two rows). The mode and shape of the distribution are almost identical. With our methodology, we are able to sample 3 to 4 times the number of fishes used in the official estimations.

4. Discussion

This work has three main contributions: the application of segmentation and classification algorithms to currently existing camera systems installed on vessels, and “ground truth” validation with reports based on

Table 5
Some statistics of the four samples used to compare the classification of species and size estimations performed by our model against official data. PS = Port Sampling, DLS = Deep learning sampling (this methodology). Paired sample *t*-test calculated $p < 0.10$ (*), $p < 0.05$ (†) and $p < 0.01$ (‡). See Fig. 10.

Sample number	Source	BET			SKJ			YFT			Total		
		n	Size (cm)		*n*	Size (cm)		*n*	Size (cm)		*n*	Size (cm)	
			Mean	SD		Mean	SD		Mean	SD		Mean	SD
1	PS	0	–	–	0	–	–	200	150.84	9.02	200	150.84	9.02
	DLS	10	95.15	34.59	14	106.86	28.66	111	119.67‡	20.39‡	135	116.53‡	23.82‡
2	PS	0	–	–	0	–	–	200	149.86	10.2	200	149.86	10.2
	DLS	16	105.69	41.75	16	109.47	21.11	62	119.68‡	25.66‡	94	115.56‡	29.0‡
3	PS	46	47.26	5.43	60	43.7	5.34	36	45.86	3.51	142	45.4	5.21
	DLS	119	44.01*	9.7*	182	45.8	10.98	37	55.73	14.97	338	46.26	11.59
4	PS	10	46.2	1.66	60	38.73	2.38	8	43.62	0.86	78	40.19	3.5
	DLS	183	40.91	14.11	225	40.21	9.94	26	56.0	15.98	434	41.45	12.81

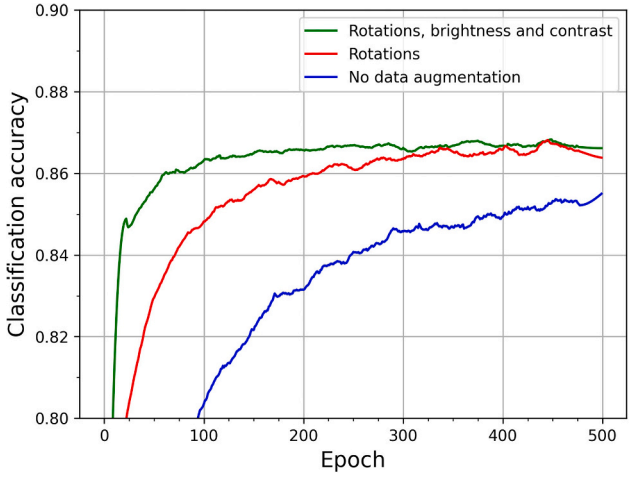


Fig. 7. Accuracy of different tuna species classification models trained with and without data augmentation.

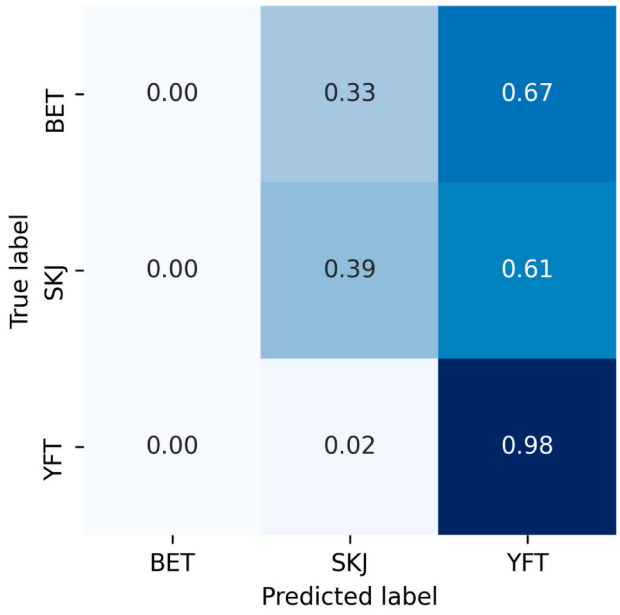
port manual estimations.

The methodology developed allows for the automatic processing of the captured images. This methodology allows the size of a higher number of individuals compared to port sampling to be estimated. The goal is to find the catch composition of the three target tuna species and not the absolute number of individuals with current low-cost camera systems. Although the images used are of low quality due to heterogeneous angles and poor maintenance of the cameras, more than 30,000 images were processed, segmenting a large part of the fishes captured. Normally, image analysis studies in marine science are small scale proof-of-concept in controlled environments, studies with low-cost approaches applied or with the potential to be applied at large scale are sparse (Irigoien et al., 2009). Other researchers have proposed similar segmentation methodologies to this (French et al., 2020; Garcia et al., 2020), but none of them attempted to do it with highly aggregated individuals. Both use the Mask R-CNN architecture (He et al., 2018). The former takes the images in a controlled environment, so the correct illumination and low aggregation of individuals allows it to segment the fish almost perfectly. On the other hand, French et al. (2020) works with images from a conveyor belt, so the lighting is not as controlled and the overlap between individuals is bigger. They monitor the fish discards, so the fishes do not overlap as much as in our case. To the best of our knowledge, the combination of uncontrolled environment and highly aggregated fishes makes our work a challenge that was never previously addressed in fisheries image applications.

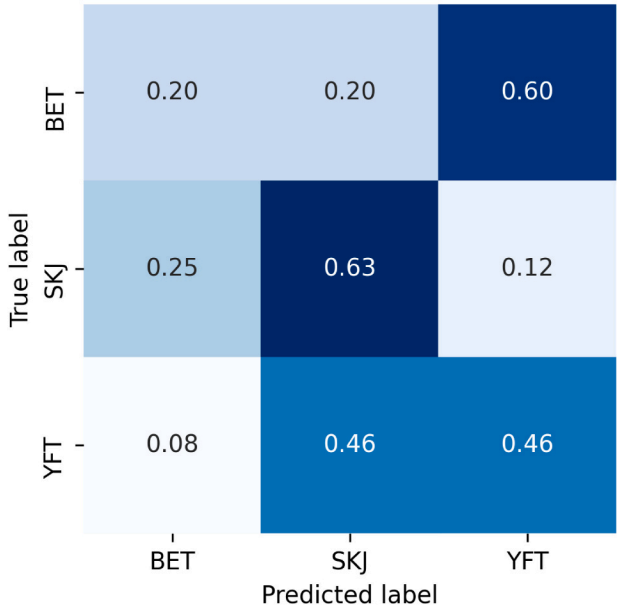
Despite a mere average of 200 specimens for each species were-Correccion? labeled for the classification algorithm, competitive performance was achieved. Similar studies have used a labeled dataset with 6 times more data (French et al., 2020; Rathi et al., 2017). As with segmentation, direct comparison of our results with previous studies is not possible, as the methodology used differs. Since there are no examples that use new classes for segment validation we must rely solely on the predicted species. French et al. (2020) classifies up to nine different fish species, obtaining a mean accuracy of 62.55% with 4-fold cross validation. Rathi et al. (2017) classifies 21 species and goes as high as 96.29%, but the images are captured underwater and they do not cross validate. We achieved a mean accuracy of 77.66% for the classification by species. More robust scheme validation reduces performance variance, but also accuracy (Fernandes et al., 2010). Training sets that are biased towards good quality images can achieve higher accuracies, but may not be representative of the data to be analyzed under real conditions. For example, in our case, accuracy dropped from 77.66% to 54.50% when artifacts were added to our training set, but the final comparison improved.

One of the key points for these results has been the preprocessing of the images. In machine learning the preprocessing often has more impact in the results than the model selected (Bora, 2017; Fernandes et al., 2010). Much of this preprocessing is specific to the current conditions of the EM system, requiring a change in the event of an improvement of the EM system. The dirtiness detection module developed is independent of these conditions and it would not need to be changed. This module can be integrated in the commercial software used to digitalize the images to send an alarm to demand cleaning. The methodology used for perspective correction depends on the type of work to be performed (Jagannathan and Jawahar, 2005). In our case the most complicated part was to find the boundaries of our initial images, but once the procedure is set up, only small tweaks are needed for each vessel. Contrast enhancement has also proven to be vital for computer vision (Vijayalakshmi et al., 2020). In our case, the segmentation of individuals would have been much more difficult without a correct histogram equalization.

Predictions have been compared with official data from manual counting in port (Sarralde et al., 2009). But it must be considered that the subsample taken by the camera and selected by the port sampler are unlikely to contain the same individuals. This comparison made it possible to detect strengths and weaknesses in the methodology. As the behavior of fish changes with maturity, it is expected that the individuals caught with different fishing techniques belong to a different growth stage. A pattern in the yellowfin tuna, with two size modes, can also be found in other studies on tropical tunas (Floch et al., 2018). The other tuna species had more complications, especially bigeye, which is the least caught target tuna species and is greatly overrepresented by the model in the final comparison. On average, bigeye catches make up only 2.5% of each fishing set, whereas our model predicts 25.5%, a tenfold increase. The validation method used shows that the most commonly confused species are also bigeye and skipjack, but it does not suggest that the differences with the official data would be so large. Therefore, it is necessary to continue developing our methodology to reduce these errors. It should be noted that both methods have their own biases, as

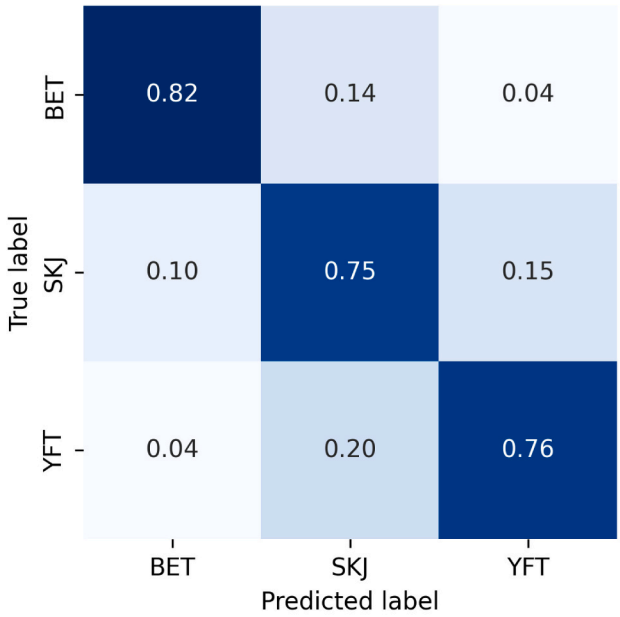


(a) Model trained with manual segments and tested on automatic segments.

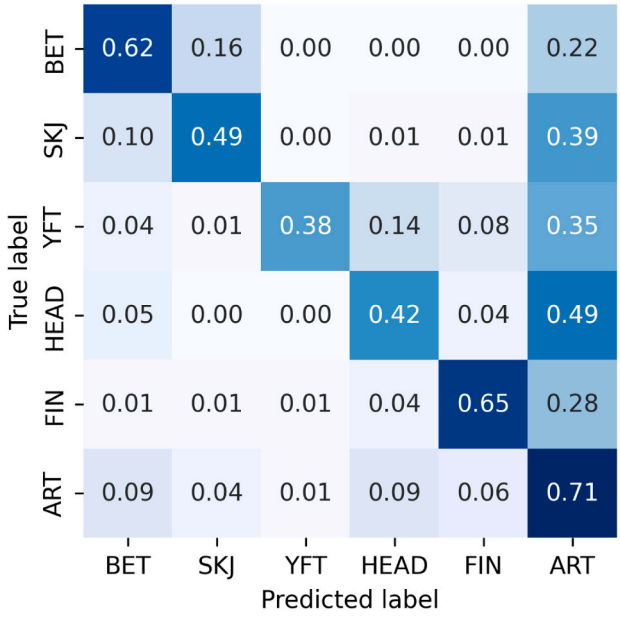


(b) Model trained with automatic segments and tested on manual segments.

Fig. 8. Confusion matrices for the two models used to test possible biases in our model. The vertical axis represents the true species of the segments and the horizontal axis represents the species predicted by the model.



(a) Classification model without artifacts

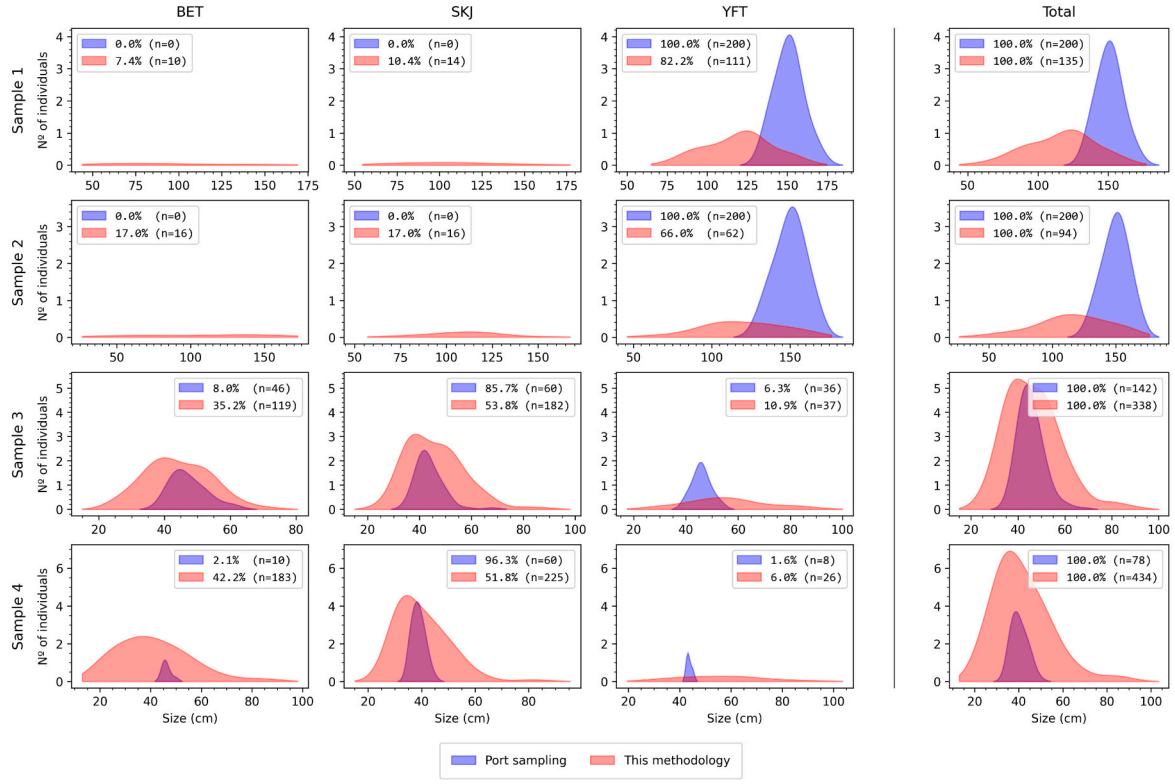


(b) Classification model with artifacts

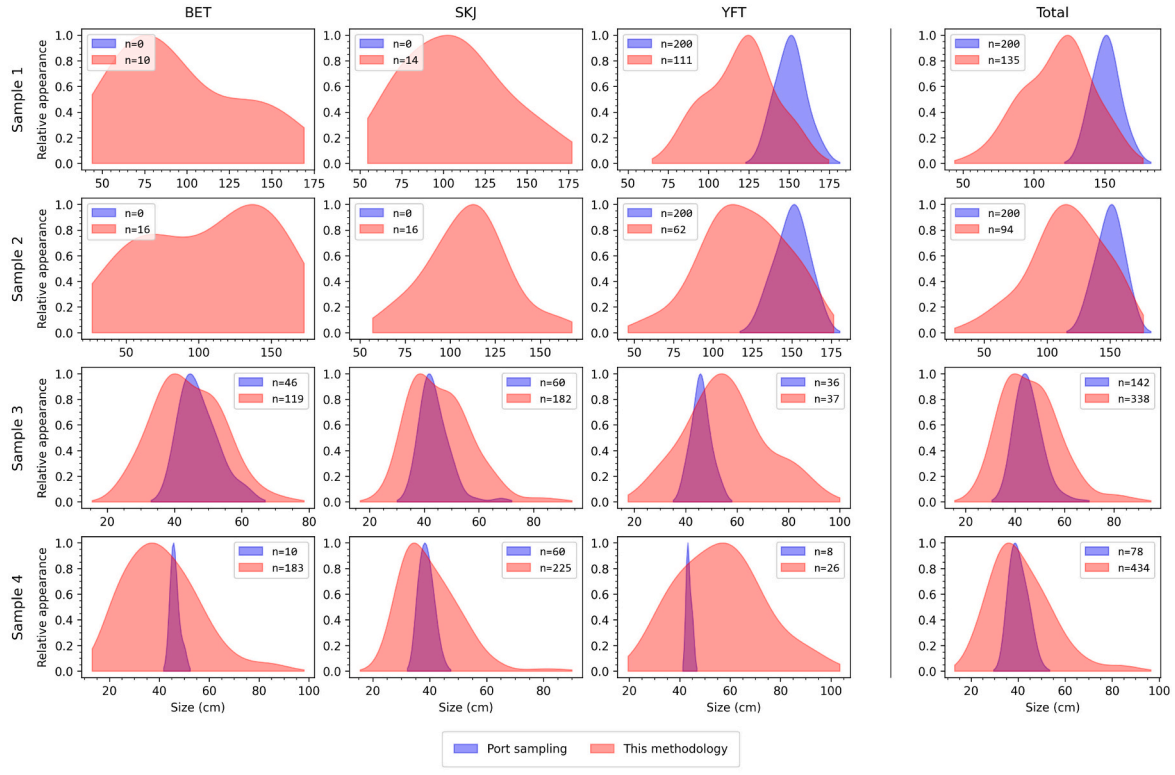
Fig. 9. Confusion matrices for the two approaches used for the classification model (with and without artifacts). The vertical axis represents the true species of the segments and the horizontal axis represents the species predicted by the model.

seen in (Lawson, 2009) and *sec*. 3.1. The accuracy of these estimations will improve and converge to the real catch distribution as more individuals are selected for sampling. As sampling all individuals is not feasible, finding the optimum number between not selecting enough fishes and too many and detecting new biases is something that is worth studying in future work.

In this regard, all tuna Regional Fisheries Management Organizations (RFMOs), as well as several national administrations, are in the process of defining the minimum standards that an EM system should meet for the correct monitoring of fishing activity (Murua et al., 2020a). The following recommendations should be considered when defining the standards to monitor the tropical tuna purse seine fishery: (1) The inclusion of a dedicated camera for automatic species identification and measurement. By positioning a camera at a zenithal angle to the conveyor belt, the perspective correction would become unnecessary, speeding up the process and removing any possible errors introduced by doing so. A camera like the current ones would be more than enough to capture ideal images, but the use of other types of cameras (higher resolution or 3D, for example) that have proven to be useful in other cases for the segmentation of aggregate elements can be of great help (Garcia et al., 2020; Xia et al., 2015). (2) Evenly illuminating the area targeted by the camera is also crucial (Bachiller and Fernandes, 2011;



(a) Absolute number of fishes



(b) Relative appearance of each size

Fig. 10. Four samples have been used to compare with official data. Port samplings are in blue and our estimates are in red. Fig. 10a: The label on each graph represents the percentage of occurrence of that species in that set (the number of specimens is also included). See Table 5 for more data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Xiang et al., 2011). Some fish are currently overexposed and unrecognizable in the images, so a well-lit environment can facilitate recognition of the fishes. The boundaries between different individuals would also become much easier to identify, easing the segmentation. (3) One way of avoiding large accumulations of fishes would be to use a hopper, like the ones currently used by some tuna fishing fleets operating in the Pacific Ocean (Grande et al., 2019; Murua et al., 2020c). It consists of an intermediate hopper that is placed in the access hatch to the fishing park over which the brailer is pulled. This device has a tray and an opening/closing flap at the end of it, allowing regulation of the flow of fish falling onto the belt. This system, or a similar one, would favor the separation between individuals and therefore improve segmentation and automatic classification.

We understand that all these changes to an established system take time and cost money, but assuming a fishing trip of 30 days and 2 fishing operations per day, more than 60 hours of work must be done to process all the data of a single vessel in a single fishing trip. Furthermore, to process this data is required to invest time in the training of an analyst and continue this time-investment with every fishing operation. On the other hand, once the artificial intelligence model is properly trained it does not require expensive and time-costly manual labor to operate. This paper shows that is technically feasible to do similar work as an analyst in near real time. Ensuring a regular maintenance of the system and doing the following tweaks to our methodology may be enough to greatly improve the results: (1) Creating a custom-made dataset. Whereas analysts have years of experience in tuna species classification, it is much easier for them to do their work onboard and seeing the fish directly. It is possible to create a better dataset and take advantage of this. The ideal way to generate this new dataset is by making "virtual fishing sets", where the composition of the set is known. For such a set, the analyst will select in advance the individuals that pass through the conveyor belt to be photographed. This way, we gain much more knowledge about the images that are used to train the model. (2) New classification algorithms. We tested a classification based on neural networks, but since the number of valid segments obtained per image may not be sufficient for this type of algorithms, it may be useful to try other approaches. Some of the methods that have already proven their utility in making good predictions in image classification are support vector machines (Chandra and Bedi, 2018), random forest (Sheykhmousa et al., 2020) or even the combination of several methods (Agarap, 2019). If it is possible to increase the dataset as described above, it would also be interesting to test other neural network architectures (Wang and Yang, 2019). Other type of validations such as an out-of-distribution approach (Hendrycks and Gimpel, 2018) may also be useful. Stricter thresholds for our predictions were tested, but too many segments were discarded. If enough segments are provided, predictions should improve. (3) Testing with more ground truth data. Another way to detect possible biases that have been ignored so far is to use more ground truth data in the comparison. By combining this image analysis methodology with logbooks or number of brails recorded at every fish event, it would be possible to have tons of catches with its composition for each fishing event and the accumulated total catches. For this new methodology to be accepted by the industry, it would be necessary to perform the same validation comparing catch composition estimates at the fleet level for a complete year. Reporting this data annually by fleet is a RFMO requirement for stock assessment, so it would be possible to estimate the catches distribution by species for the purse-seine fleet in an ocean and compare it with the official estimates presented at the RFMO level.

Despite all the improvements described above, our results suggest that improving the number of labeled segments and ensuring them to be representative of the variety in the captured images would be sufficient to have an effective and operational automated EM system. This will be achieved by having an expert observer and the computer engineer on board a fishing vessel during real operations to visually classify and afterwards digitalize. This process will make segments available in the training set that cannot be labeled in the images, but that have been classified to species level before digitalization. Therefore, this achieves both objectives, it increases the labeled segments and makes them more representative. Moreover, classification of FSC is the most challenging, but is also showing a downward trend with some fleets in the Indian Ocean having less than 5% of their fishing sets using this fishing strategy (Basurko et al., 2022; Floch et al., 2019). Having accurate and less cost-effective estimations of the catch composition and size distribution will allow for improvements in the stock assessment of the tropical tuna species (Uranga et al., 2017). Furthermore, the methodology presented in this study would allow catch estimates to be obtained in near real time and be associated with a specific vessel, this being a key point for the control of the fishery (i.e., quota monitoring). The current scientific sampling method used by the European purse seine fleet based on port sampling (Duparc et al., 2019) only obtains aggregated species composition estimates for the entire fleet during a period (for example, quarter), this being the main obstacle to its use as a control tool (i.e., for the control of the individual vessel quotas). The industry will also be able to have near real time catch monitoring to planning of their activities. There are also other applications that contribute to the environmental and economic sustainability of the tuna fisheries such as using the data for developing decision support systems that reduce fuel consumption (Granado et al., 2021).

The independent monitoring, control and surveillance of fishing activities is important in an era that requires transparency and reliable information for scientists and managers, thus guaranteeing sustainable exploitation. In the case of the RFMOs this is challenging, as they play a key role in managing highly migratory fish stocks, such as tuna, that span the jurisdictions of many countries as well as the high seas. Since the European Community, one of the leading agents in the control and scientific advice to management of fisheries, has started the process to amend its legislation to introduce the digitalization of the control and scientific sampling by means of electronic monitoring as can be seen in the European Parliament proposal for the amendment of Council Regulation (EC) No 1224/2009, and amending Council Regulations (EC) No 768/2005, (EC) No 1967/2006, (EC) No 1005/2008, and Regulation (EU) No 2016/1139 of the European Parliament and of the Council as regards fisheries control (Commission, 2018), the development of systems like the one presented here are becoming more and more important. Reliable catch information by species is a key element for tropical tuna assessment and management. Without well-grounded commercial catch data, stock assessments and quota control can clearly be undermined. Previous EM pilot studies have shown that accurate estimation of catch composition in tropical tuna purse seine is challenging even if done manually (Murua et al., 2020a; Ruiz et al., 2015), as large volumes of fish enter the conveyor belt at once. Thus, incorporating deep learning methodologies such as those developed in this study to EM systems would give them a much greater potential use.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This project is funded by the Basque Government, and the Spanish fisheries ministry through the EU next Generation funds. Jose A. Fernandes' work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 869342 (SusTunTech). This work is supported in part by the University of the Basque Country UPV/EHU grant GIU19/027. We want to thank the expert analysts who helped to annotate images with incredible effort: Manuel Santos and Iñigo Krug. We also like to extend our gratitude to Marine Instruments for providing the necessary equipment to

collect the data. This paper is contribution n° 1080 from AZTI, Marine Research, Basque Research and Technology Alliance (BRTA).

References

Agarap, A.F., 2019. An Architecture Combining Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for Image Classification. arXiv:1712.03541 [cs, stat]. URL: https://arxiv.org/abs/1712.03541.

Alsmadi, M.K., Almarashdeh, I., 2020. A survey on fish classification techniques. J. King Saud Univ. https://doi.org/10.1016/j.jksuci.2020.07.005. S1319157820304195 URL:

Amandè, M.J., Ariz, Javier, Chassot, E., Delgado, A., Gaertner, D., Murua, H., Pianet, R., Ruiz, J., Pierre, C., 2010. Bycatch of the European purse seine tuna fishery in the Atlantic Ocean for the 2003-2007 period. Aquat. Living Resour. 23, 353–362. https://doi.org/10.1051/alr/2011003.

Bachiller, E., Fernandes, J.A., 2011. Zooplankton image analysis manual: automated identification by means of scanner and digital camera as imaging devices. Rev. Investig. Mar. 18, 17–37. URL: https://www.azti.es/rim/wp-content/uploads/2010/05/Revista-Marina-18_2.pdf.

Basurko, O.C., Gabiña, G., Lopez, J., Granado, I., Murua, H., Fernandes, J.A., Krug, I., Ruiz, J., Uriondo, Z., 2022. Fuel consumption of free-swimming school versus FAD strategies in tropical tuna purse seine fishing. Fish. Res. 245, 106139. https://doi.org/10.1016/j.fishres.2021.106139.

Bora, D.J., 2017. Importance of Image Enhancement Techniques in Color Image Segmentation: A Comprehensive and Comparative Study. URL: https://arxiv.org/abs/1708.05081.

Briand, K., Bonnieux, A., Dantec, W.L., Couls, S.L., Bach, P., Maufroy, A., Relot-Stirnemann, A., Sabarros, P., 2018. Comparing Electronic Monitoring System with Observer Data for Estimating Non-target Species and Discards on French Tropical Tuna Purse Seine Vessels, pp. 3813–3831.

Cayré, P., 1984. Procédure suivie pour la révision de la composition spécifique des statistiques thonières, ICCAT, Madrid. In: Volume 21 of Collective Volume of Scientific Papers - ICCAT, pp. 102–107. URL: https://www.documentation.ird.fr/hor/fdi:010028715.

Chandra, M.A., Bedi, S.S., 2018. Survey on SVM and their application in image classification. Int. J. Inf. Technol. https://doi.org/10.1007/s41870-017-0080-1. URL:

Chuang, M.C., Hwang, J.N., Rose, C.S., 2013. Aggregated segmentation of fish from conveyor belt videos. In: In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, Vancouver, BC, Canada, pp. 1807–1811. https://doi.org/10.1109/ICASSP.2013.6637964.

Commission, E., 2018. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL amending Council Regulation (EC) No 1224/2009, and amending Council Regulations (EC) No 768/2005, (EC) No 1967/2006, (EC) No 1005/2008, and Regulation (EU) No 2016/1139 of the European Parliament and of the Council as Regards Fisheries Control. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018PC0368.

Cui, S., Zhou, Y., Wang, Y., Zhai, L., 2020. Fish Detection Using Deep Learning. Applied Computational Intelligence and Soft Computing 2020, pp. 1–13. URL: https://doi.org/10.1155/2020/3738108.

Duparc, A., Aragno, V., Depetris, M., Floch, L., Cauquil, P., Lebranchu, J., Gaertner, D., Marsac, F., Bach, P., 2018. Assessment of the Species Composition of Major Tropical Tunas in Purse Seine Catches: A New Modelling Approach for the Tropical Tuna Treatment Processing, p. 32.

Duparc, A., Aragno, V., Depetris, M., Floch, L., Cauquil, P., Lebranchu, J., Gaertner, D., Marsac, F., Bach, P., 2019. Assessment of the Species Composition of Major Tropical Tunas in Purse Seine Catches: A New Modelling Approach for the Tropical Tuna Treatment Process | IOTC. URL: https://www.iotc.org/documents/WPTT/21/10.

Fernandes, J.A., Irigoien, X., Goikoetxea, N., Lozano, J.A., Inza, I., Parez, A., Bode, A., 2010. Fish recruitment prediction, using robust supervised classification methods. Ecol. Model. 221, 338–352. https://doi.org/10.1016/j.ecolmodel.2009.09.020.

Floch, L., Hervé, A., Irié, D., Guillou, A., Depetris, M., Duparc, A., Lebranchu, J., Bach, P., 2018. Statistics of the French purse fishing fleet targeting tropical tunas in the Atlantic Ocean (1991-2017). In: ICCAT, Madrid. URL: https://www.documentation.ird.fr/hor/fdi:010076017. Accession number: fdi:010076017 source: IRD - base horizon / Pleins textes.

Floch, L., Depetris, M., Dewals, P., Duparc, A., Kaplan, D., Lebranchu, J., Marsac, F., Pernak, M., Bach, P., 2019. Statistics of the French Purse Seine Fishing Fleet Targeting Tropical Tunas in the Indian Ocean (1981–2018). URL: https://archimer.ifremer.fr/doc/00590/70259/.

Fonteneau, A., 1976. Note sur les problèmes d'identification du bigeye dans les statistiques de pêche. Col. Vol. Sci. Pap. ICCAT 5, 168–171.

Fonteneau, A., 2008. Species Composition of Tuna Catches Taken by Purse Seiners. Western and Central Pacific Fisheries Commission, Honolulu, HI, USA, p. 14.

French, G., Mackiewicz, M., Fisher, M., Holah, H., Kilburn, R., Campbell, N., Needle, C., 2020. Deep neural networks for analysis of fisheries surveillance video and automated monitoring of fish discards. ICES J. Mar. Sci. 77, 1340–1353. https://doi.org/10.1093/icesjms/fsz149.

Fujita, R., Cusack, C., Karasik, R., Takade-Heumacher, H., 2018. Designing and Implementing Electronic Monitoring Systems for Fisheries: A Supplement to the Catch Share Design Manual. Technical Report. Environmental Defense Fund, San Francisco. URL: https://www.edf.org/sites/default/files/oceans/EM_DesignManual.PDF.

Garcia, R., Prados, R., Quintana, J., Tempelaar, A., Gracias, N., Rosen, S., Vågstøl, H., Løvall, K., 2020. Automatic segmentation of fish using deep learning with application to fish size measurement. ICES J. Mar. Sci. 77, 1354–1366. https://doi.org/10.1093/icesjms/fsz186.

Gilman, E., Legorburu, G., Fedoruk, A., Heberer, C., Zimring, M., Barkai, A., 2019. Increasing the functionalities and accuracy of fisheries electronic monitoring systems. Aquat. Conserv. Mar. Freshwat. Ecosyst. 29, 901–926. https://doi.org/10.1002/aqc.3086.

Girshick, R., 2015. Fast R-CNN. URL: https://arxiv.org/abs/1504.08083.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Columbus, OH, USA, pp. 580–587. https://doi.org/10.1109/CVPR.2014.81.

González, P., Álvarez, E., Díez, J., López-Urrutia, N., Coz, J.J.D., 2017. Validation methods for plankton image classification systems. Limnol. Oceanogr. Methods 15, 221–237. https://doi.org/10.1002/lom3.10151.

Granado, I., Hernando, L., Galparsoro, I., Gabiña, G., Groba, C., Prellezo, R., Fernandes, J.A., 2021. Towards a framework for fishing route optimization decision support systems: review of the state-of-the-art and challenges. J. Clean. Prod. 320, 128661. https://doi.org/10.1016/j.jclepro.2021.128661.

Grande, M., Murua, J., Ruiz, J., Ferarios, J.M., Murua, H., Krug, I., Zudaire, I., Goñi, N., Santiago, J., 2019. Bycatch mitigation actions on tropical tuna purse seiners: Best practices program and bycatch releasing tools. In: Technical Report. IATTC 9th Meeting of the Working Group on Bycatch. San Diego, California.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Las Vegas, NV, USA, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2018. Mask R-CNN. URL: https://arxiv.org/abs/1703.06870.

Hendrycks, D., Gimpel, K., 2018. A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. URL: https://arxiv.org/abs/1610.02136.

Hirata, N.S.T., Fernandez, M.A., Lopes, R.M., 2016. Plankton image classification based on multiple segmentations. In: 2016 ICPR 2nd Workshop on Computer Vision for Analysis of Underwater Imagery (CVAUI), pp. 55–60. URL: https://doi.org/10.1109/CVAUI.2016.022.

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., Murphy, K., 2017. Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors. URL: https://arxiv.org/abs/1611.10012.

ICCAT, 2006. ICCAT Manual. International Commission for the Conservation of Atlantic Tuna. URL: https://www.iccat.int/en/iccatmanual.html.

ICCAT, 2019. Recommendation by ICCAT to Replace Recommendation 16-01 by ICCAT on a Multi-Annual Conservation and Management Programme for Tropical Tunas. Technical Report. URL: https://www.iccat.int/Documents/Recs/compendiopdf-e/2019-02-e.pdf.

IOTC, 2019. Resolution 19/01 on an Interim Plan for Rebuilding the Indian Ocean Yellowfin Tuna Stock in the IOTC Area of Competence | IOTC. Technical Report. URL: https://iotc.org/cmm/resolution-1901-interim-plan-rebuilding-indian-ocean-yellowfin-tuna-stock-iotc-area-competence.

Irigoien, X., Fernandes, J.A., Grosjean, P., Denis, K., Albaina, A., Santos, M., 2009. Spring zooplankton distribution in the Bay of Biscay from 1998 to 2006 in relation with anchovy recruitment. J. Plankton Res. 31, 1–17. https://doi.org/10.1093/plankt/fbn096.

Jagannathan, L., Jawahar, C.V., 2005. Perspective correction methods for camera-based document analysis. In: Proc. First Int. Workshop on Camera-based Document Analysis and Recognition, pp. 148–154.

Lawson, T., 2009. Selectivity Bias in Grab Samples and Other Factors Affecting the Analysis of Species Composition Data Collected by Observers on Purse Seiners in the Western and Central Pacific Ocean, p. 45.

Majkowski, J., Arrizabalaga, H., Carocci, F., Murua, H., 2011. Tuna and Tuna-Like Species, p. 17.

McElderry, H., 2008. At-Sea Observing Using Video-Based Electronic Monitoring.

Michelin, M., Sarto, N.M., Gillett, R., 2020. Roadmap for Electronic Monitoring in RFMOs, p. 42.

Monteagudo, J.P., Legorburu, G., Justel-Rubio, A., Restrepo, V., 2014. Preliminary Study About the Suitability of an Electronic Monitoring System to Record Scientific and Other Information from the Tropical Tuna Purse Seine Fishery, p. 20.

Murua, H., Fiorellato, F., Ruiz, J., Chassot, E., Restrepo, V., 2020a. Minimum Standards for Designing and Implementing Electronic Monitoring Systems in Indian Ocean Tuna Fisheries. URL: https://www.iotc.org/documents/SC/23/12E.

Murua, H., Herrera, M., Morón, J., Abascal, F., Legoburu, G., Roman, M., Moreno, G., Hosken, M., Restrepo, V., 2020b. Comparing electronic monitoring and human observer collected fishery data in the tropical tuna purse seine operating in the Pacific Ocean. In: IATTC SAC-11 INF-G. URL: https://www.iattc.org/Meetings/Meetings2020/SAC-11/Docs/_English/SAC-11-INF-G_Comparing%20Electronic%20Monitoring%20and%20human%20observer%20collected.pdf.

Murua, J., Moreno, G., Itano, D., Hall, M., Dagorn, L., Restrepo, V., 2020c. ISSF Skippers Workshops. Technical Report Round 9. URL: https://iss-foundation.org/knowledge-tools/reports/technical-reports/.

Qiao, M., Wang, D., Tuck, G.N., Little, L.R., Punt, A.E., Gerner, M., 2020. Deep learning methods applied to electronic monitoring data: automated catch event detection for longline fishing. ICES J. Mar. Sci. fsaa158 https://doi.org/10.1093/icesjms/fsaa158.

Rathi, D., Jain, S., Indu, S., 2017. Underwater fish species classification using convolutional neural network and deep learning. In: 2017 Ninth International

Conference on Advances in Pattern Recognition (ICAPR). IEEE, Bangalore, pp. 1–6. https://doi.org/10.1109/ICAPR.2017.8593044.
Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.
Restrepo, V., Forrestal, F., 2012. A Snapshot of the Tropical Tuna Purse Seine Large-Scale Fishing Fleets At the End of 2011. Technical Report.
Reza, A.M., 2004. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. J. VLSI Signal Process. Syst. Signal Image Video Technol. 38, 35–44. https://doi.org/10.1023/b:vlsi.0000028532.53893.82.
Ruiz, J., Batty, A., Chavance, P., McElderry, H., Restrepo, V., Sharples, P., Santos, J., Urtizberea, A., 2015. Electronic monitoring trials on in the tropical tuna purse-seine fishery. ICES J. Mar. Sci. 72, 1201–1213. https://doi.org/10.1093/icesjms/fsu224.
Ruiz, J., Krug, I., Gonzalez, O., Hammann, G., 2016. E-EYE PLUS: Electronic Monitoring Trial on the Tropical Tuna Purse Seine Fleet. ICCAT, Madrid, Spain, p. 3.
Ruiz, J., Abascal, F., Bach, P., Baez, J., Cauquil, P., Grande, M., Krug, I., Lucas, J., Murua, H., Alonso, M., Sabarros, P., 2018. Bycatch of the European, and Assocated Flag, Purse-Seine Tuna Fishery in the Indian Ocean for the Period 2008–2017. URL:. https://doi.org/10.13140/RG.2.2.11527.24482.
Saberioon, M., Císař, P., 2018. Automated within tank fish mass estimation using infrared reflection system. Comput. Electron. Agric. 150, 484–492. https://doi.org/10.1016/j.compag.2018.05.025.
Sarralde, R., Delgado de Molina, A., Ariz, J., Santana, J.C., Pallaés, P., Pianet, R., Dewals, P., Hervé, A., Dedo, R., Areso, J.J., 2009. Port Sampling Procedures for Tropical Tuna in the Atlantic and Indian Oceans. Technical Report ICCAT Manual. URL: https://www.iccat.int/en/iccatmanual.html.
Sekachev, B., Zhavoronkov, A., Manovich, N., 2019. Computer Vision Annotation Tool: A Universal Approach to Data. URL: https://www.intel.ai/introducing-cvat.
Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., Homayouni, S., 2020. Support vector machine versus random Forest for remote sensing image classification: a Meta-analysis and systematic review. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 13, 6308–6325. https://doi.org/10.1109/JSTARS.2020.3026724.
Torrey, L., Shavlik, J., 2010. Transfer learning. In: Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques. IGI Global, pp. 242–264.
Uranga, J., Arrizabalaga, H., Boyra, G., Hernandez, M.C., Goai, N., Arregui, I., Fernandes, J.A., Yurramendi, Y., Santiago, J., 2017. Detecting the presence-absence of bluefin tuna by automated analysis of medium-range sonars on fishing vessels. PLoS One 12, e0171382. https://doi.org/10.1371/journal.pone.0171382.
Van Helmond, A.T.M., Mortensen, L.O., Plet-Hansen, K.S., Ulrich, C., Needle, C.L., Oesterwind, D., Kindt-Larsen, L., Catchpole, T., Mangi, S., Zimmermann, C., Olesen, H.J., Bailey, N., Bergsson, H., Dalskov, J., Elson, J., Hosken, M., Peterson, L., McElderry, H., Ruiz, J., Pierre, J.P., Dykstra, C., Poos, J.J., 2020. Electronic monitoring in fisheries: lessons from global experiences and future opportunities. Fish Fish. 21, 162–189. https://doi.org/10.1111/faf.12425.
Vijayalakshmi, D., Nath, M.K., Acharya, O.P., 2020. A comprehensive survey on image contrast enhancement techniques in spatial domain. Sens. Imaging 21, 40. https://doi.org/10.1007/s11220-020-00305-3.
Wang, W., Yang, Y., 2019. Development of convolutional neural network and its application in image classification: a survey. Opt. Eng. 58, 1 https://doi.org/10.1117/1.OE.58.4.040901.
Xia, C., Wang, L., Chung, B.K., Lee, J.M., 2015. In situ 3D segmentation of individual plant leaves using a RGB-D camera for agricultural automation. Sensors 15, 20463–20479. https://doi.org/10.3390/s150820463.
Xiang, R., Ying, Y., Jiang, H., 2011. Research on image segmentation methods of tomato in natural conditions. In: 2011 4th International Congress on Image and Signal Processing. IEEE, Shanghai, China, pp. 1268–1272. https://doi.org/10.1109/CISP.2011.6100424.
Yao, H., Duan, Q., Li, D., Wang, J., 2013. An improved K -means clustering algorithm for fish image segmentation. Math. Comput. Model. 58, 790–798. https://doi.org/10.1016/j.mcm.2012.12.025.
Yu, C., Fan, X., Hu, Z., Xia, X., Zhao, Y., Li, R., Bai, Y., 2020. Segmentation and measurement scheme for fish morphological features based on mask R-CNN. Inform. Process. Agric. 7, 523–534. https://doi.org/10.1016/j.inpa.2020.01.002.
Zhang, Z., He, L.W., 2007. Whiteboard scanning and image enhancement. Digital Signal Process. 17, 414–432. https://doi.org/10.1016/j.dsp.2006.05.006.
Zhao, Z.Q., Zheng, P., Xu, S.T., Wu, X., 2019. Object Detection with Deep Learning: A Review. URL: https://arxiv.org/abs/1807.05511.