Artificial vision, a new method to estimate the species composition of catches in tropical tuna purse seine fisheries

By X. Lekunberri[1], J. Ruiz[1], A. Kamal[1]& I. Quincoces[1]

1 AZTI Foundation, Txatxarramendi ugartea z/g 48395, Sukarrieta, Bizkaia, Spain

Abstract

Purse seine fisheries play a crucial role in global tuna fishing, accounting for approximately 66% of the world's tropical tuna catch. However, accurately estimating the specific composition of these catches per set and in real-time remains challenging. To enhance traditional data collection methods like onboard observers and port sampling, all tuna Regional Fisheries Management Organizations (RFMOs) have established minimum standards for utilizing Electronic Monitoring (EM). EM was developed to improve data collection efficiency and traceability. Despite its advantages, EM faces challenges such as the time required to review all data and the difficulties in accurately distinguishing and estimating the specific composition of purse seine catches based on visual reviews.

In response to these challenges, we are developing a pipeline to reduce the workload involved in reviewing EM footage. This pipeline utilizes video captured by the EM, along with several computer vision models. This setup enables the estimation of species composition of the catches with minimal human interaction. The target tuna species composition is estimated using standard 2D footage from the wells deck, where the fish are recorded on the conveyor belt before being stored in wells.

This document highlights several key points identified during the pipeline's development. Some issues to consider while collecting data aboard the vessel include ensuring adequate lighting on the fishing deck and utilizing dedicated global shutter camera hardware. Other considerations pertain to the development of AI models, such as creating custom datasets and specific model architectures. Addressing these factors will enhance the accuracy and efficiency of species composition estimation, ultimately improving the overall effectiveness of EM systems in fisheries management.

1. Background

In the tropical tuna purse seine fishery, catches at sea typically include several target species, such as yellowfin tuna (*Thunnus albacares*, YFT) and skipjack tuna (*Katsuwonus pelamis*, SKJ), along with non-targeted but significantly caught bigeye tuna (*Thunnus obesus*, BET) and varying proportions of other secondary species. During a fishing operation, onboard catch handling is much faster compared to other fisheries. Retained catch is stored directly in freezing tanks (wells) in brine without prior sorting to avoid sanitary and food safety issues. Sorting occurs as needed when certain non-marketed species are discarded. Thus, the estimation of catch per species is determined either through visual assessments or subsampling.

The species composition by set is recorded in the logbook; however, bias in logbooks has been evident since the inception of the tropical tuna purse seine fishery (Fonteneau 1976; Cayré 1984; Fonteneau 2007), primarily concerning small YFT and BET individuals. While the total weight can be accurately estimated since catches are weighed at landing, the species composition remains problematic (Lawson, 2009), preventing the use of these values as precise measures. Ruiz *et al*. (2021) concluded that, in many instances, acquiring accurate estimates of species composition per fishing operation may not be feasible due to three main causes: the inability to obtain an adequate minimum sample size, the challenges in ensuring random selection of this sample, and the added difficulty of distinguishing between juvenile YFT and BET. The accuracy of logbooks can vary, but despite the capacity and experience of the crew member completing them, discrepancies between logbook entries and actual landings have been demonstrated. Therefore, to determine the specific composition of tropical tuna landings, some port sampling can be conducted during the landing process. A sampling and data processing strategy for estimating species and size composition of catches in the European and Seychellois purse seine fisheries was developed more than two decades ago and is routinely employed by scientists based in the primary landing ports (i.e., Seychelles) (Duparc, 2018). However, this method is intended for generating global estimates, which means it is not suitable for obtaining estimates per set, thereby making it unusable for quota control, for example.

Technology (E-reporting, Vessel Monitoring Systems (VMS), Electronic Monitoring Systems (EMS)) has recently improved the monitoring of fishing activities. Several organizations, including tuna-RFMOs (Regional Fisheries Management Organizations), have promoted the use of EMS as a data collection tool. This could lead to greater and more efficient at-sea observation coverage and could complement (or even replace) some responsibilities and tasks traditionally associated with onboard human observers. The IOTC was one of the pioneering RFMOs when it adopted its minimum standards for the use of EMS in 2023 (IOTC Resolution 23/08), and now, several companies owning large purse seiners have voluntarily installed EM systems, presenting an opportunity to obtain species-specific catch estimates. From the footage captured on the conveyor belt in the bellow-deck area, a qualified operator could estimate the species composition of the catches. However, various pilot studies with EMS in purse seine fleets reveal that, besides being tedious and costly, EMS is not currently prepared to fulfil this function, and the target tuna catch estimates obtained by EM analysts are not accurate (Ruiz et al., 2015; Murua et al., 2021). The application of artificial vision techniques that could automate the identification of catches by species is seen to be the next step in automating EM data workflows, increasing the value of these electronic monitoring systems by enabling the collection of catch data in an objective and near real-time manner.

Artificial vision has gained significant popularity in recent years, thanks to advancements in deep learning algorithms and the enhanced computing capabilities available today. This technique, when applied to fisheries, is already being tested in various case studies to automate the image review process through the automatic detection of events or species, both in tuna fisheries (Lekunberri et al., 2021) and other fisheries (Rizwan et al., 2022; Kay et al., 2021).

The objective of this working document is to present the progress made by AZTI in this field (i.e. automatic identification of tropical tuna catches), as well as to propose a series of future recommendations that will pave the way to achieve this goal.

2. Methods

Figure 1 illustrates the complete pipeline, which begins with footage acquisition, proceeds to image annotation, and concludes with training and validation. The process consists of three independent models: the segmentation model, individual fish tracking, and species classification. Sections 2.1 to 2.5 detail each step within the pipeline.

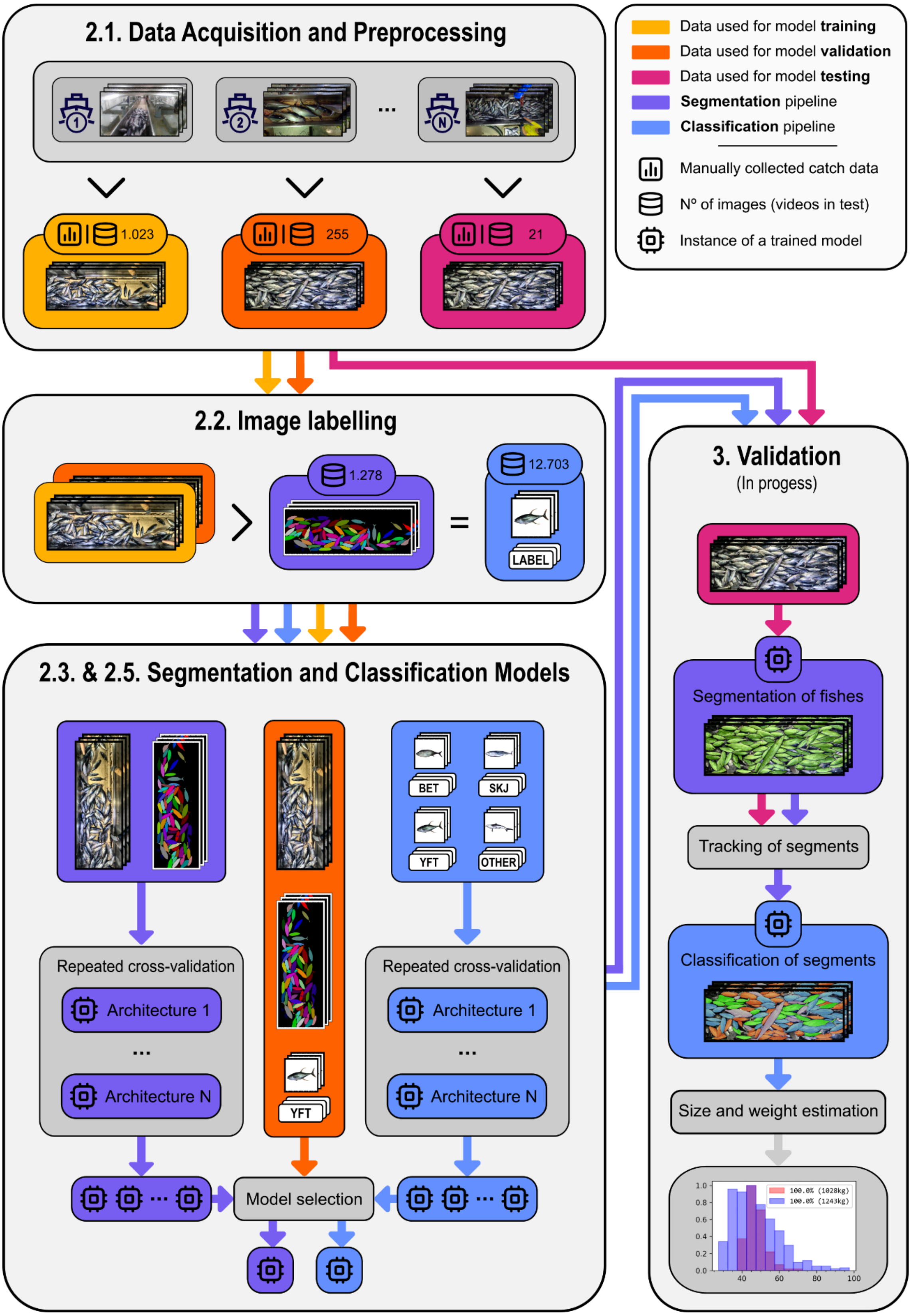

Figure 1. General pipeline. The process starts with the data collection. Footage is organized according to its intended use: training, validation, or testing. The training and validation images are annotated. These annotated images are then used to train a segmentation model and to obtain the individual fish segments required for training a classification model. The real-world validation step is performed using unannotated test images, where the ground truth is known.

2.1. Generating the image training set: Data acquisition and preprocessing

All images were captured with digital cameras on the conveyor belt in the well deck. Two different EMS systems were used, manufactured by Marine Instrument and Zunibal. Both systems feature different cameras, although they share similar technical characteristics (Vivotek FD-9366-HV model or similar). These are 2 MP, IP67 cameras that include enhanced Wide Dynamic Range technology.

Additionally, independent stereoscopic cameras (Stereolabs ZED2 and ZED X models) were punctually installed during two fishing trips. These cameras provide a 3D view of the conveyor belt and come with enhanced features. They utilize a global shutter, unlike the rolling shutter found in most 2D cameras. The global shutter eliminates distortion in images of fast-moving objects, ensuring uniform exposure across the entire frame. This makes them preferable for our specific case, where precision and clarity are required for moving objects (fish) under artificial and variable poor lighting conditions. In all instances, images were captured at 30 fps.

As the first step, preprocessing the images is crucial since they come from an uncontrolled environment. This ensures a set of images as homogeneous as possible, thereby facilitating subsequent automatic segmentation and classification. The preprocessing is carried out in two phases: perspective correction and contrast enhancement (Figure 2). In addition, it is important to take into account that when capturing images with a camera positioned above a conveyor belt, it's not unusual for the lens to gradually become soiled with water or blood droplets. Since dirt on a lens remains stationary while the region of interest is in constant motion, image dirtiness can be estimated by calculating the variance of each pixel over an entire fishing operation. A dirty region will have low variance, as the colour remains consistent, whereas a clear region will have higher variance. This method could be used to discard sets that do not reach a minimum image quality (Figure 2).

For fishing operations without significant dirt, the region of interest (RoI) in the images was first selected. This RoI usually covered the area of the conveyor belt closest to the camera; however, each camera underwent a calibration to define its own RoI.

Once the region of interest was identified, perspective correction was applied, discarding everything outside its boundaries. This correction was necessary because the cameras were not always positioned to capture a top-down view of the conveyor belt, causing fish size and shape to vary depending on their position along the belt. This correction aims to maintain uniform size throughout the image, facilitating segmentation and subsequent species classification.

Finally, the image contrast was enhanced to better differentiate between specimens. The technique used is known as CLAHE (Contrast Limited Adaptive Histogram Equalization), described by Reza, A. M. (2004), which increases image contrast locally (by different regions), highlighting the fish contours more effectively.

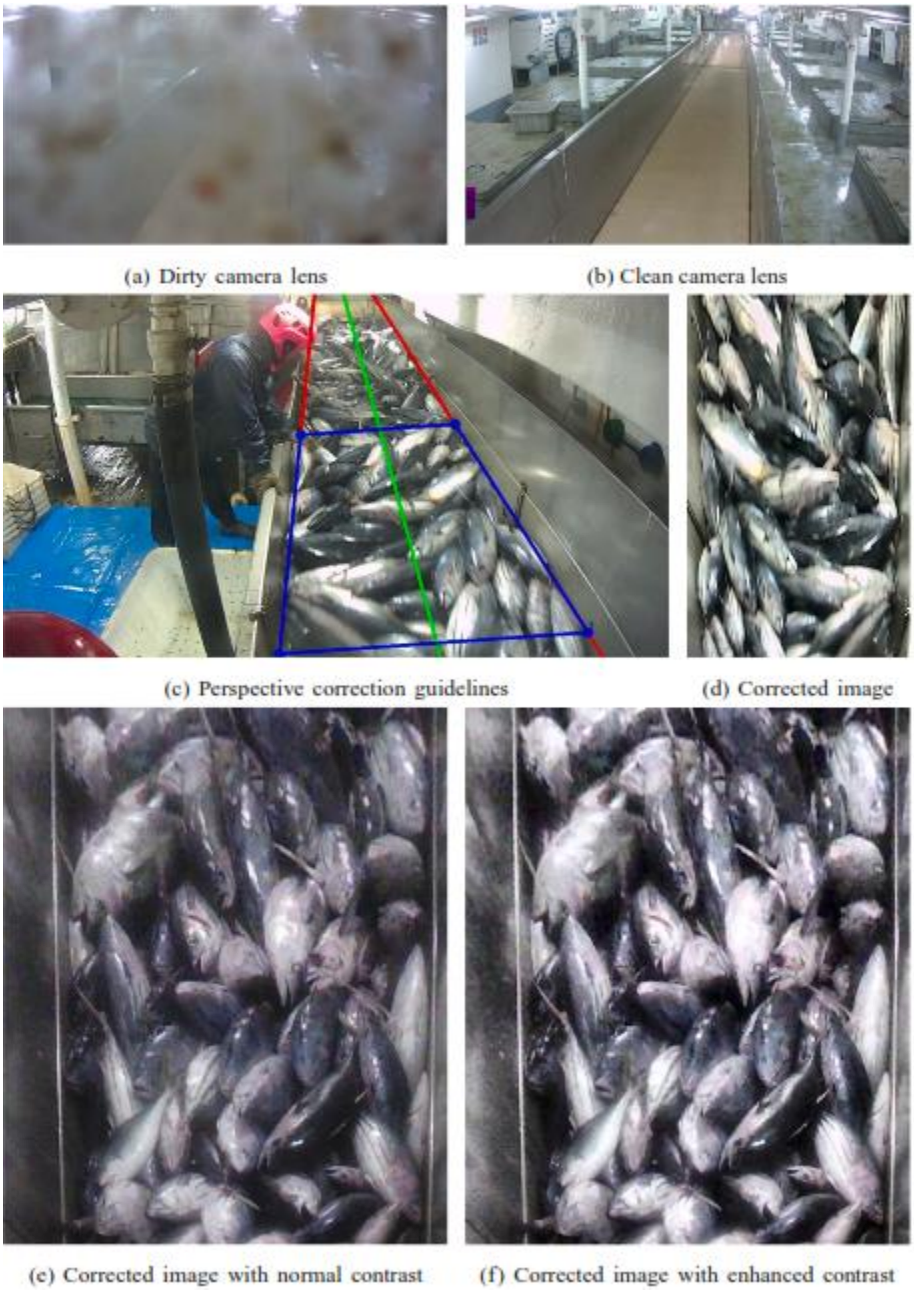

Figure 2. Examples of image preprocessing: dirty and unusable image (a) versus clear and clean image (b). Original image (c) and perspective correction (d). Normal contrast (e) versus enhanced contrast(f).

2.2. Generating the image training set: Image labelling

Experienced onboard observers manually segmented and annotated images using the open-source tool CVAT. This tool enables the upload of images to a shared repository, facilitating collaborative work among experts. All fishes were labelled according to species. In some cases, where it was not possible to reach the species level, the label “no-target” was used for all bycatch, including neritic tunas. Conversely, the label “target” was used for the targeted species. In instances where distinguishing between target and non-target species was not feasible, the label "fish" was utilized. Fishes with non-species labels are only used for the segmentation model or the first steps of the hierarchical classification model (see sections 2.3 and 2.5).

It is well known that the classification of YFT and BET is not easy, especially in their juvenile stage. Therefore, although at the beginning of the project this classification was based on the criterion of the land-based observers annotating the images, the current criterion is that an onboard observer validates the species: after identifying the specimens in situ (onboard), they place them in separate batches of BET and YFT in front of the camera for later annotation. This criterion was established after experts showed low species identification agreement when using only photographs in an *ad hoc* workshop.

In 2022, AZTI conducted a test that revealed the visual identification of YFT and BET based on EM images is challenging and uncertain. Therefore, the YFT and BET specimens used in the annotation process have now been previously identified onboard by experienced observers before being placed in front of the camera. As part of the 2022 test, an exercise was organized for labelling YFT and BET specimens. Besides AZTI experts, eight external experts from the four tuna RFMOs were invited to participate, including CSIRO (Australia), SPC (South Pacific), University of Hawaii, IATTC (USA), Datafish (Spain), and IRD (France). They were provided with 40 EM images that came from different vessels and fishing operations, all taken on the conveyor belt (Figure 3). The exercise focused exclusively on YFT and BET, where each expert selected only those individuals whose species identification was possible (decided individually by each expert). The photos varied in quality (dirty, blurry, etc.), representing a sample of what a set of images might show during a real fishing operation. Each identified fish was subsequently assigned a final label (YFT, BET, or UNK - unknown). To decide if a fish was truly BET or YFT, at least one-third of the experts (three out of nine) needed to identify the same fish, and at least 75% of the experts who identified it labelled it as the same species. Among all the experts, a total of 257 specimens of YFT or BET were identified. However, when comparing the species assigned by each expert, 181 of these fish could not be confidently assigned to either species.

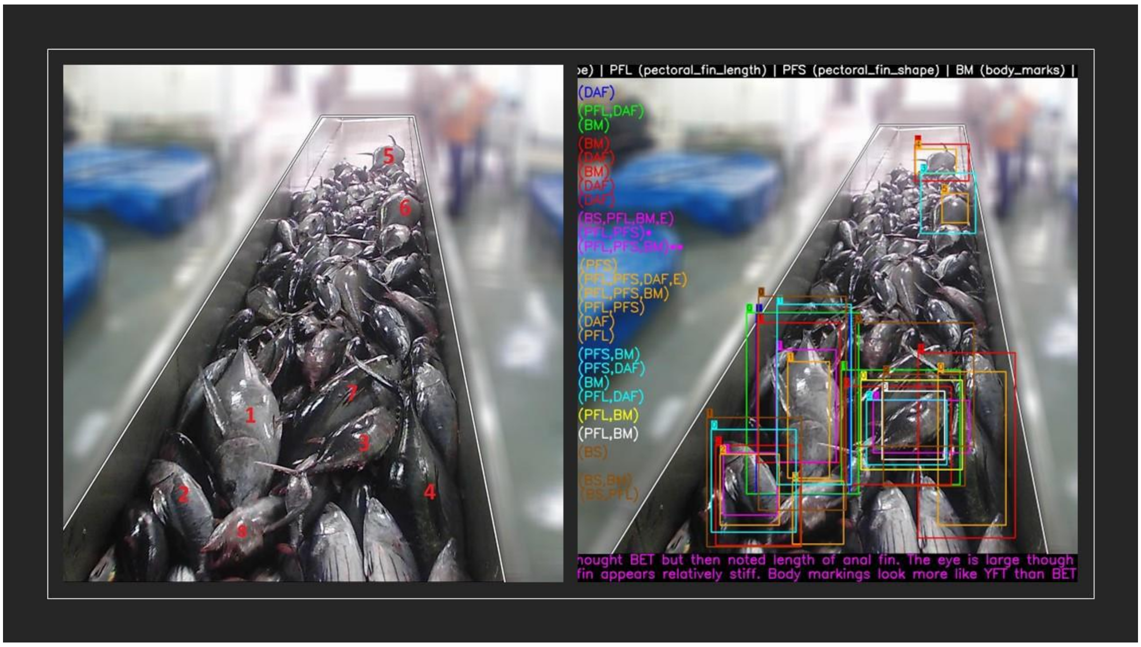

Figure 3. Example of an image used in the expert labelling exercise. Original (left) and labelled (right). Each colour represents the individuals selected by each expert.

2.3. Segmentation model

After annotating the frames, the segmentation and classification models (Section 2.5) were trained in parallel. Segmentation was addressed separately due to the morphological similarities among the three target species (BET, SKJ, and YFT). Consequently, all labels were combined into a single "FISH" super-label for segmentation, thereby reducing the complexity of the task. Previously labelled “target” and “fish” labels were also added to this new super-label. The model underwent validation using a repeated cross- validation strategy.

A segmentation model based on neural networks, utilizing the Mask R- CNN structure (He et al., 2017), was initially trained (see Lekunberri et al., 2022). This model is one of the most advanced in the field of machine learning for image segmentation. Starting from a model previously trained for general-purpose segmentation of everyday objects, our model was retrained to segment fish effectively. Mask R- CNN first analyses the image and detects up to 2000 areas that might contain objects (in this case, fish). It groups similar areas to form larger regions, identifying parts of the image where objects are likely to be found.

The network then extracts features from these grouped regions and compares them to features it has learned previously. If a region contains an object of interest, such as a fish, the network labels it accurately.

One challenge with this process is that it requires many labelled images and significant time for the model to learn how to separate different objects correctly. To reduce this cost, a technique called transfer learning is employed. This involves taking a neural network already trained for a specific task and adapting its knowledge to a new, similar task. In this instance, the "Mask R-CNN Inception ResNet V 2" model, trained with the ImageNet dataset, was modified to detect fish. Fish detection is achieved by changing the final layer of the network and partially retraining it. This approach allows the neural network to retain its ability to differentiate individual objects while becoming specialized in segmenting fish.

Other segmentation approaches, such as a combination of YOLO (Redmon et al., 2016) and SAM 2 (Ravi et al., 2024), are also being tested to identify the most accurate model for individual fish detection.

2.4. Tracking algorithms

Considering that the fish on the conveyor belt are moving, and each one will appear in different frames of the video, it is essential to identify each individual throughout the sequence of frames. The main objective is to avoid counting any fish more than once. Additionally, identifying each fish individually in various frames allows us to obtain a better estimate of the species, as the classification model will have multiple opportunities to make predictions from different angles, lighting, or occlusion. Therefore, the probability of a fish belonging to a specific species will be based on multiple predictions.

As with the segmentation model, several algorithms have been tested. ByteTrack (Zhang et al., 2022) has proven to be the approach that best fits our specific case. We opted to implement ByteTrack for its consideration of both high and low confidence predictions, making it more robust to occlusions during tracking. As an added feature, optical flow was integrated into the ByteTrack pipeline to ensure tracking only occurs when the conveyor belt is moving in one direction and not the other (when the algorithm detects the conveyor belt moving backwards, the tracking stops).

2.5. Classification model

A classification model was trained on data labelled for each targeted tuna species, namely skipjack tuna (SKJ), yellowfin tuna (YFT), and bigeye tuna (BET). For this task, various state-of-the-art models were considered, with RegNet (Radosavovic et al., 2020) being chosen for its efficient extraction of fine image features. RegNet is an image classification model that is designed to automatically find the best way to analyze images for subtle details. This allows it to accurately distinguish between similar-looking objects in an image by focusing on fine visual differences between them making it ideal for distinguishing between tuna species with similar features.

Initially, single-stage classification approach was considered: outputs from the segmentation model were directly classified into the three targeted tuna species. However, with the aim of improving the model's performance, a hierarchical classification pipeline was adopted. Hierarchical classification works by breaking down the problem, allowing the model to learn specific features at each stage, and improve accuracy for similar classes. The classification took place in three stages. Firstly, the detections would be classified in target (tuna) vs non-target (bycatch) species. Then, the skipjack tuna, characterized by its distinct features, namely the dark stripes on its belly, was separated from other tuna species. Finally, having similar visual features, bigeye and yellowfin tuna were classified in the third stage of the classification.

2.6. Size sampling and LWR

We need the final estimates of the specific composition of the catch in terms of weight by species. Therefore, it will be necessary to determine the weight of each identified individual. Since we don't have an effective way to weigh the specimens, we'll accomplish this by measuring them and then converting those measurements to weight using the specific length-weight relationship (LWR).

Once the image perspective has been corrected, estimating sizes becomes straightforward since the dimensions of each individual remain consistent throughout the image. For the most accurate approximation, it's crucial to measure only those individuals where occlusion is insignificant. These are individuals that are fully visible from head to tail (Figure 4). One way to check if an individual is overlapped is to measure the ratio between its major and minor axes, i.e., horizontal and vertical length. A complete individual will have a higher ratio, while overlapped ones will have a lower ratio. This ratio can be specified by species, or a generic one can be used if the differences between tuna species are minimal.

Figure 4. Non-occluded fish selected for measurement (green), and incomplete fish not selected for the size sample (red)

Then, with the selected individuals, the measurements are performed. To simplify this, we use the width of the conveyor belt, which remains constant throughout its length, both in the image and in reality. Knowing the width of the belt both in centimetres and in pixels, we only need to calculate the pixel distance from end to end of the major axis of each individual and apply a proportion to approximate its real size.

3. Results and validation

The pipeline tuning tasks are still ongoing, so the results presented in this document should be considered preliminary. The training image set continues to grow, as do the tests with different models in various phases (segmentation and classification).

Currently, we have 12,703 individual segments derived from 1278 independent annotated images. Table 1 below shows the number of segments by label. Keep in mind that labels referring to species were merged into the corresponding super-labels for each of the hierarchical classification models.

<table>
<tr><th colspan="5">Segments by label</th></tr>
<tr><th colspan="3">TARGET</th><th>NO_TARGET</th><th>FISH</th></tr>
<tr><td colspan="3">6204</td><td rowspan="5">1114</td><td rowspan="5">4928</td></tr>
<tr><th>SKJ</th><th colspan="2">BET or YFT</th></tr>
<tr><td rowspan="3">3590</td><td colspan="2">2614</td></tr>
<tr><th>BET</th><th>YFT</th></tr>
<tr><td>303</td><td>2274</td></tr>
</table>

Table 1. The segmented images and the total number of segments used for model training are categorized hierarchically. The top-level labels are TARGET and NO_TARGET, representing species targeted by the fishery (BET, SKJ, and YFT) and non-target species (all others). Based on morphological similarities, target species are further divided into SKJ and BET_YFT. Finally, BET_YFT is split into the respective species

All our models were validated using a robust 5-fold stratified cross-validation. Figure 5 shows the confusion matrices for each of the classification steps. Results show that for the first two levels, an acceptable level of accuracy is achieved, while for the third and final level, the results have a large margin for improvement. At the initial level, a substantial proportion of the individuals belonging to the target species (BET, SKJ, and YFT) are accurately identified, with a 99% accuracy rate. 18% of the bycatch that is erroneously identified as tuna is added to this group of fish that passes to the next step. It is worth noting that this 18% refers only to the bycatch, not the total catch. Considering that bycatch accounts for a small proportion of the total catch, this 18% does not represent a significant error in the final estimates. Each stage of the hierarchical classification system has been trained exclusively with its respective labels. No analyses have been conducted to assess the impact of misclassified individuals as they progress to the next stage. The second level, where skipjack is differentiated from the other two species, performs even better, with an accuracy of 99% for SKJ identifications and 97% for BET and YFT. As previously mentioned, the distinction based on visual characteristics of these species facilitates classification. However, the final stage of the classification exhibits a large margin for improvement, as only 72% of the BET and 87% of the YFT are correctly classified. While these percentages might be considered high for other applications, they are below the minimum acceptable threshold before using them both for science and compliance purposes.

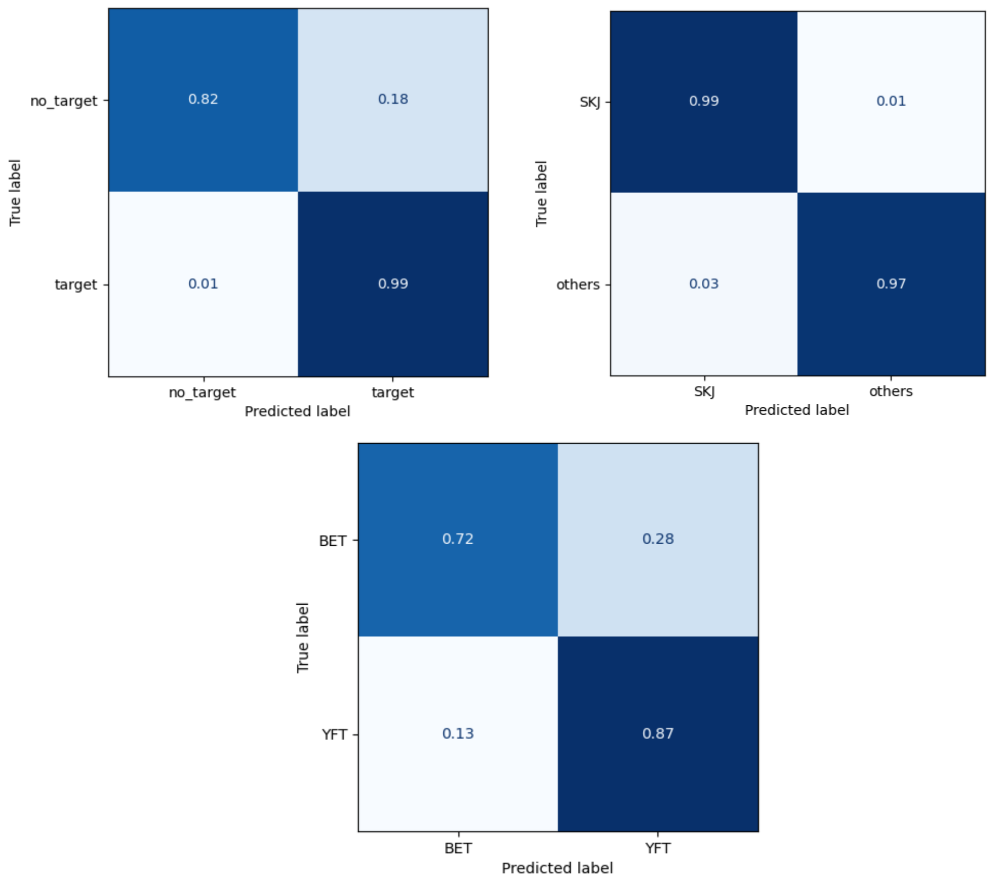

Figure 5: Confusion matrices for each of the steps in our hierarchical classification. Top left for the first step (target or no-target), top right for the second step (SKJ or others) and third step in the bottom (BET or YFT)

The second phase of validation is currently underway. This involves comparing our model's automatic estimates against the ground truth. Obtaining accurate data on the specific composition of a commercial fishing operation (i.e, ground truth) is challenging, mainly because the minimum sample size required is often unattainable. To address this difficulty and determine the ground truth, two onboard scientists from AZTI randomly selected a sample of 1.5 to 2 tons of fish from several commercial sets in the Indian Ocean. They measured (FL lower cm) and identified the species of each fish. Once the catch was stowed, this sample was randomly placed on the conveyor belt and filmed using the EM system. This comparison will not only allow us to assess the model's performance but also to study potential sources of bias, such as the implications of identifying/measuring only individuals from the surface layer on the conveyor belt or only whole individuals (without significant occlusion).

4. Recommendations and future developments

The work is ongoing, and new results will be presented at future IOTC meetings. However, the work done so far has allowed us to identify several strengths and weaknesses in the automatic identification of species on the conveyor belt in tropical tuna purse seine vessels.

- **Strengths**

    - *Larger Sample size*: This method allows for inferring the catch composition based on a much larger sample compared to what could be done by onboard personnel (whether an observer or a crew member). In small samples of 300 to 500 fish on the conveyor belt, the system successfully segments (samples) between 72% and 93% of the fish. However, in a commercial fishing set, where the belt contains multiple layers of fish, this percentage would significantly decrease

    - A catch composition estimation method based on AI would mean having a *standardized method* among CPCs that implemented EM in their purse seine fleets.

    - *Near real-time data*: Obtaining real-time or near real-time data would be of great help to complete the logbooks automatically, or at least for the captains to have another source of information beyond the sampling prepared by the crew.

- **Weaknesses and or challenges**

    - *Poor light conditions and movement*: The training and detection images are captured on the vessel's lower deck in low-light conditions. This poor lighting forces the camera to use longer exposure times. While this isn't an issue for still objects, the rapid movement of fish on the conveyor belt during these longer exposures causes significant blur in the resulting images, negatively impacting their quality.

    - *Inconsistency among installations*: Due to the unique characteristics of each boat, the images of the well deck can vary significantly from one boat to another. The position and angle of the cameras, as well as the lighting conditions, are unique for each boat. This variability poses a substantial challenge for AI models, as it can be difficult to learn the characteristic patterns of each species.

- *Occlusion*: The bulk delivery of tuna onto the conveyor belt often leads to significant occlusion, meaning parts of the fish cannot be seen in the image. This poses a considerable challenge, particularly when only a small portion like the tail is visible, rendering the accurate classification of the species difficult for experienced observers annotating the data, let alone a trained AI-model.

- *Unbalanced and small training set*: Compared to the other two target tuna species, bigeye tuna is caught in significantly lower numbers. This disparity in catch rates directly impacts the training data, creating an imbalance. This data bias is particularly problematic for distinguishing bigeye from yellowfin tuna, as they share many visual similarities, leading to frequent misclassifications of bigeye as yellowfin.

- *Dirty lenses*: Since the cameras used are placed very close to the conveyor, the fish themselves splash them as they pass through the conveyor. These contaminants are made up of blood, scales, water, and any other type of debris that the fish bring with them. If the camera lenses are not cleaned regularly, they become clogged to the point of being useless.

Future work:

- *3D camera*: Besides the improved image quality from the 3D camera's global shutter, it also captures depth information, generating a 3D representation of the scene. By measuring the 3D distance across a reference object of known length within this 3D data (e.g. width of the conveyor belt), a real-world scale is established. This scale then allows for accurate measurement of the 3D distance between the fish's fork length end points, enabling precise calculation of its actual length (in cm) and, therefore, its weight.

- *Other models*: To potentially enhance the segmentation performance, preliminary experiments are currently being trialled. Real-time YOLO bounding box detection followed by SAM2 segmentation is being compared against the performance of the current method that uses Mask R-CNN. This approach leverages YOLO''s speed for initial object detection, feeding those results to SAM2 for instance segmentation, aiming for improved accuracy without significantly increasing the overall duration required to perform the detections.

5. Acknowledges

This work has received funding from various sources, including the Basque Government, the Spanish Fisheries Secretariat through Next Generation funds, the European project EVERYFISH

(Grant agreement ID: 101059892), and the company ZUNIBAL. The authors also wish to thank the Spanish purse seine fishing industry, especially the company PEVASA, for making their vessels available to AZTI scientists for conducting sea trials.

6. References

Cayré P. 1984. Procédure suivie pour la composition spécifique des statistiques Thonières. Col Vol Sci Pap ICCAT. 21(2):102–107.

Duparc A, Cauquil P, Depestris M, Dewals P, Gaertner D, Hervé A, Lebranchu J, Marsac F, Bach P. 2018. Assessment of accuracy in processing purse seine tropical tuna catches with the T3 methodology using French fleet data. In: Report of thte 20th session of the IOTC Working Party on Tropical Tunas. Victoria, Seychelles:IOTC. (IOTC-2018-WPTT20-16). p. 1–19.

Fonteneau A. 1976. Notes sur les problèmes d'indentification du Bigeye dans les statistiques de pêche ColVolSciPap ICCAT. 5(1):168–171.

Fonteneau A. 2007. Species composition of tuna catches taken by purse seiners. In: 3rd Regular Session of the Scientific Committee. Honolulu, HI, USA: Western and Central Pacific Fisheries Commission. (WCPFC-SC3-STSWG-IP-7). p. 1–13.

Kay, J., & Merrifield, M. (2021). The Fishnet Open Images Database: A Dataset for Fish Detection and Fine-Grained Categorization in Fisheries. ArXiv. https://arxiv.org/abs/2106.09178

Lekunberri X., Ruiz J., Quincoces I., Dornaika F., Arganda-Carreras I., Fernandes J. 2022. Identification and measurement of tropical tuna species in purse seiner catches using computer vision and deep learning, Ecological Informatics, Volume 67, 2022, 101495, ISSN 1574-9541, https://doi.org/10.1016/j.ecoinf.2021.101495.

M. Rizwan Khokher, L. Richard Little, Geoffrey N. Tuck, Daniel V. Smith, Maoying Qiao, Carlie Devine, Helen O'Neill, John J. Pogonoski, Rhys Arangio, and Dadong Wang. 2022. Early lessons in deploying cameras and artificial intelligence technology for fisheries catch monitoring: where machine learning meets commercial fishing. Canadian Journal of Fisheries and Aquatic Sciences. 79(2): 257-266. https://doi.org/10.1139/cjfas-2020-0446.

Murua, H., Fiorellato, F., Ruiz, J., Chassot, E., & Restrepo, V. (2020). Minimum standards for designing and implementing Electronic Monitoring systems in Indian Ocean tuna fisheries. IOTC–2020–SC23–12[E] rev2.

IOTC. (2023). Resolution 23/08 on electronic monitoring standards for IOTC fisheries. https://iotc.org/cmm/resolution-2308-electronic-monitoring-standards-iotc-fisheries.

Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K., & Dollár, P. (2020). Designing Network Design Spaces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N., Wu, C.-Y., Girshick, R., Dollár, P., & Feichtenhofer, C. (2024). SAM 2: Segment Anything in Images and Videos. In Proceedings of the International Conference on Learning Representations (ICLR)

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

Reza, A. M. (2004). Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement. Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology, 38, 35-44

Ruiz, J., Batty, A., Chavance, P., McElderry, H., Restrepo, V., Sharples, P., Santos, J., Urtizberea, A., 2015. Electronic monitoring trials in the tropical tuna purse-seine fishery. ICES J. Mar. Sci. 72, 1201–2013. https://doi.org/10.1038/278097a0

Ruiz J., Duparc A., Abascal F., Bach P., Baez JC., Pascual P., Gaertner D, Marsac F. and Santiago J. 2021. Comments on the assessment of catch by species in the tropical purse seine fishery. From: https://orthongel.fr/pages/tools/notes/declar/Catch_composition_assessment_forPS.pdf

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-Object Tracking by Associating Every Detection Box. In Proceedings of the European Conference on Computer Vision (ECCV)