Instance Segmentation in Mobile Computing Environments for Identification of Specific Characteristics in Endangered Species

Morgana C. O. Ribeiro[1], Rhayane S. Monteiro[1], Mário W. L. Moreira[1], Joel J. P. C. Rodrigues[2,3]

[1]Federal Institute of Education, Science and Technology of Ceará (IFCE), Aracati, CE, Brazil
[2]Federal University of Piauí (UFPI), Teresina, PI, Brazil
[3]Instituto de Telecomunicações, Portugal

morgana.cassia.oliveira05@aluno.ifce.edu.br; {rhayane.monteiro,mario.wedney}@ifce.edu.br; joeljr@ieee.org

***Abstract*—The segmentation of instances is a key topic in image processing and computer vision. There are numerous applications such as medical image analysis, video surveillance, image compression among others, in which its algorithms show significant results. Due to the COVID-19 pandemic, most countries have been affected mainly by their economy. In the food production sector, including fishing and aquaculture, it was no different. In this context, this research has as main objective to contribute to the 2030 Agenda for Sustainable Development suggested by the United Nations (UN), through a fish detection and segmentation model based on the framework Detectron2, optimizing the time of professionals in identifying specific characteristics of a particular species. To achieve this objective, this research seeks to facilitate the recognition of patterns of parts of the fish from the segmentation of instances and to stimulate scientific research in the area through the morphological information collection of certain species. The results present an accuracy, based on the Intersection over Union (IoU) indicator, of 88.4%, providing an effective solution for the collection of these characteristics.**

***Index Terms*— Mobile Computing, Pattern Recognition, Instance Segmentation, Convolutional Neural Networks, Detectron2.**

I. Introduction

Recently, the pandemic caused by COVID-19 has affected most countries worldwide, with severe impacts on the global economy and the food production and distribution sector, including fisheries and aquaculture [1]. The Food and Agriculture Organization of the United Nations (FAO) is monitoring the situation closely to assess the global impact of the pandemic on production, consumption, and trade in fisheries and aquaculture. Twenty-five years after the adoption of the Code of Conduct for Responsible Fishing[1], the importance of using fishing and aquaculture resources responsibly is now widely recognized and prioritized.

The document informs the development of international instruments, policies, and programs to support efforts responsible for global, regional, and national management. These efforts have been consolidated and prioritized since 2015 to address, in a coherent and coordinated way, the Sustainable Development Goal (SDG) 14, which talks about the conservation and sustainable use of oceans, seas, and marine resources for sustainable development and that it has clear and important implications for fisheries and aquaculture[2]. By extension, the achievement of this objective will bring progress concerning other SDG challenges for these activities. The main challenges are related to improving data collection, protecting endangered species, preventing fishing, and ensuring social sustainability in the value chain. Besides, sustainable development presents an international challenge that will require consistent, coherent, and effective cooperation among countries and institutions.

To this purpose, the United Nations 2030 Agenda for Sustainable Development was drafted in 2015. It is based on the Millennium Development Goals and provides a comprehensive set of goals on which companies, governments, and individuals can focus their efforts on social improvement.

Given the challenges faced by FAO to provide a development roadmap that is socially, environmentally and economically sustainable, and inclusive, the identification of species at risk of extinction in an automated way is considered one of the most challenging computer vision issues, since it quickly extract valuable information from images still presents several challenges to be considered. It is worth mentioning that this approach is based on certain criteria for dividing the input image into regions or categories, which correspond to different objects or parts of the objects, *i.e.*, each pixel in an image is allocated to one of several categories to extract the interesting areas [2].

In this context, this research proposes as a general objective to contribute to the 2030 Agenda, based on a fish detection and segmentation model based on the Detectron2 framework, optimizing the professional's time in identifying specific char-

[1]Available at: http://www.fao.org/3/i1900e/i1900e00.htm

[2]Available at: http://www.fao.org/documents/card/en/c/ca9229en/

acteristics of a specific species. As specific objectives, this study seeks to facilitate the recognition of patterns of parts of the fish from the segmentation of instances and to stimulate scientific research in the area through the morphological information collection of fish [3]. These objectives strengthen the justification for developing a model that is capable of segmenting parts of the fish's body, *i.e.*, developing new approaches for pattern recognition, in order to increase accuracy when recognizing characteristics of a particular fish species. In addition, the use of this type of approach to understanding the dynamics of populations at risk of extinction can contribute to overfishing prevention, guaranteeing the sustainability of the ecosystem, as suggested by the SDG 14.

The work is organized in sections, presenting the sequence, namely, Section II presents the important concepts for understanding the proposal; Section 3 discusses the works that relate to this article theme. The materials and methods are shown in Section 4 and Section 5 is used for results and discussions. Finally, in Section 6, the conclusion and proposal for future work are presented.

II. Theoretical Foundation

Deep learning (DP) image segmentation approaches have two subdivisions, namely, semantic segmentation (pixel classification problem with semantic labels) and instance segmentation (partitioning of individual objects) [4].

Semantic segmentation performs pixel-level labeling with a set of categories of objects [5]. This technique provides the same label to all pixels of both fish, *i.e.*, the label is a single color with information about the category, location, and form. Therefore, it is generally a more difficult undertaking than image classification, which provides for a single label for the entire image. Also, semantic segmentation differs from image classification, since it allows all parts of the object to interact, identifying and grouping pixels that are semantically together. The segmentation of instances further widens the segmentation scope, detecting and delineating each object of interest in the image.

However, the semantic segmentation results do not distinguish among different instances within the same category, resulting in limitations in the individual separation of objects. Therefore, this new problem consists not only in determining the pixels of a specific class (semantic segmentation), however also includes the discernment of different objects in the same category, obtaining the exact number of a given object in the image (instance segmentation). Therefore, instance segmentation consists of a novel paradigm and evolution of semantic segmentation, allowing a unique understanding of each object, counting the number of objects, and analyzing objects in conditions of occlusion and contact.

According to He *et al.*, instance segmentation algorithms have two main approaches, namely, *(i)* first segmentation strategy, in which segmentations occur before classification, and *(ii)* first instance strategy, parallel segmentation, and classification process [6]. In turn, the first segmentation strategy also has two approaches, the first of which is based on segments, *i.e.*, first candidates for segments are established and then the classification is performed and, based on semantic segmentation masks, this study seeks to separate the pixels of the same classes in different instances. First instance strategy methods have advantages because these methods are simpler and more flexible, allowing the algorithm to obtain bounding boxes and segmentation masks simultaneously. The main model based on the first instance strategy used during this research is called Mask-Region-based Convolutional Neural Network (Mask-RCNN).

III. Related Works

Several instance segmentation approaches that have achieved promising results have been proposed for the most diverse scenarios. In this section, some works related to the theme of this paper are presented.

Yu *et al.* proposed a method based on Mask-RCNN for recognizing the morphological characteristics of fish, *e.g.*, body length, body width, caudal peduncle length, pupil diameter, and eye diameter [7]. This study aimed to segment the image of the fish and measure the morphological indicators based on the Mask-RCNN, optimizing the measurement recognition of parts of the fish. The experimental results showed that the proposed scheme can segment the body of the fish in pure and complex environments with a good performance. In the set of images of fish with a pure bottom, the average relative errors (AREs) reached rates lower than 2.8%, and the AREs referring to the length and width of the body are less than 0.8%. In complex contexts, the AREs for all indicators are less than 3% and the AREs for body length and width are less than 1.8%. Although the morphological feature of measuring fish has been widely used in intelligent mariculture, the focus on making automated image acquisition from a device that provides data augmentation should be better thought out. Thus, the model could obtain better performance by receiving a higher number of images concerning the quantity acquired, 250 images with a clean background, and 250 images with a complex background (composed of objects present in boats). Also, the model could be more efficient on the high seas if a multiplatform version (web and mobile) and also offline was carried out so that the analysis of measurements can be made in real-time.

In [8], the authors presented an automatic segmentation model based on DL for the location of lung lesions due to the Sars-Cov-2 coronavirus. This method is promising in the diagnosis and monitoring of patients with severe respiratory syndrome, mainly in assisting the medical team in monitoring the progression or regression of the lesions presented in computerized tomography (CT). In the proposed approach, radiologists marked regions of interest (ROI), to apply them later using image segmentation. A total of 2,469 cuts (1,402 abnormal cuts and 1,067 normal cuts) from the CT were included for training. The images were selected randomly from 55 patients with Sars-Cov-2, including all slices or slices from 10 sets of fully labeled images, and 41 healthy individuals. A total of 6,480 notes are available in the training data set. The Detectron2 platform was used for the image segmentation task. The accuracy, sensitivity, and specificity of the model trained in marking a single image cut with injury reached rates of 0.954, 0.928, and 0.961, respectively. It is worth mentioning that the model was able to diagnose patients with pneumonia caused by Sars-Cov-2 with a sensitivity of 98.2% and specificity of 88.5%. Besides, the average index of intersection over union (IoU) for the test data set was 0.865. As the model was evaluated using a patient data set from a single hospital, the metrics and implementation of the model can be negatively impacted if CT images of patients from different geographic regions are aggregated. Also, it is important to highlight the impact that this model can have by diversifying the data set, including different types of injuries, patients with different diagnoses of pulmonary diseases, and risk stratification at the admission time.

In [9], the authors state that one of the main problems of instance segmentation is that most algorithms use images in Red, Green, and Blue (RGB). Considering this, satellite images often have more channels that can be crucial to improving performance. Therefore, the study brings three contributions: *(i)* data conversion system in polygon format for formatting common objects in context (COCO) annotation; *(ii)* adaptation and application of the Detectron2 software source code in multichannel images; and *(iii)* mosaic of large images. Applying the instance segmentation procedure to a multichannel remote sensing data set to detect the center-pivot irrigation system (CPIS), with images from the National Water Agency (NWA) and Landsat-8 Operational Land Imager (OLI), which are 7 channels with a resolution of 30m. In this context, the research seeks to improve the segmentation of remote sensing instances. Among the objectives, this study seeks to develop a method to convert the remote sensing data with its respective vector data and track them to the COCO data format containing the JavaScript object notation (JSON) annotation file. Second, adapt the Detectron2 instance segmentation source code to allow the multispectral data set, which are the seven surface reflectance bands of the Landsat 8 image. Finally, a mosaic method was developed to classify images on a large scale. During training, the model employed patches of 512 x 512 pixels, in which it was tested on seven different backbone structures (ResNet50 - Feature Pyramid Network (FPN), ResNet50-D5, ResNet50-C4, ResNet101-FPN, and ResNext101-FPN). The evaluation of the model used metrics of the standard COCO (Average Precision (AP), AP$_{50}$, AP$_{75}$, AP$_{small}$, AP$_{medium}$, and AR$_{100}$). It is worth noting that the ResNeXt101-FPN had the best results, with an advantage of 3% over the second-best model (ResNet101-FPN) A comparison of the ResNeXt101-FPN model is also performed on seven-channel and RGB images, in which the multichannel model presented a 3% higher performance, showing improvements using a greater number of channels. This research used as an innovation a mosaicking algorithm in instance segmentation tasks tested on an image of 1536 x 1536 pixels using a non-axial suppression ordered by area method. remote and medical images that often have more channels than the 3 traditional RBG channels.

From the works presented, there are still challenges for the task of recognizing fish. They are *i)* datasets with a limited number of images for practices involving segmentation of instances; *ii)* the lack of models being trained, and subsequently placed in developed for web or mobile environments and can be used in real-time, and *iii)* restriction of using standard size images, limiting the segmentation model to have the ability and performance to connect with training with the most diverse image sizes. Therefore, this research objective is to propose a model, based on the Detectron2 framework, to support the recognition of endangered species for their morphological identification to avoid overfishing. Thus, contributing to overcoming the challenges listed.

IV. Materials ans Methods

A. Data Acquisition

The first step comprises the acquisition of the data that make up the dataset. For this, a database was built with 363 images, divided into 8 species of marine fish commonly found in the East Coast of Ceará, Brazil. The images were captured between August 2019 and August 2020 using Canon Camera devices, Smartphones iPhone 7 Plus, and Xiaomi RedMi Note 8. Besides, the photos were collected at fish sales outlets in the city of Aracati, CE, Brazil. Figure 1 shows some examples of images in the dataset. These images generally include elements other than fish, such as the human body, objects, among others. These variations encourage the model to segment parts of the fish's body precisely regardless of the orientation or focus of the image, *i.e.*, it will be segmented and labeled instances of a particular species.

Fig. 1. Example of dataset images.

B. The COCO Annotation Format

Due to the purpose of the present project in segmenting fish body parts, it was necessary to use the instance segmentation algorithm. However, it requires a double effort in formatting labeling and annotation, requiring that each instance in a sample image of the training process needs a unique value. For example, an instance segmentation mask with ten fish fins of different classes needs different values for each type of caudal fin. Most instance segmentation algorithms follow the COCO annotation format (data format widely used by the instance segmentation and object detection community). The preparation and conversion procedure, respectively, uses two programs, namely, first, the labeling format and annotation of the image of the labelme format is performed, using the labeling software called Labelme, as shown in Figure 2. For each formatted image, a respective .json file for each image that makes up the dataset, *i.e.*, `ariaco1.png` and `ariaco1.json`. Then, a script is executed, namely, `labelme2coco.py`, which generates a new JSON file with the COCO annotation format. In this way, the dataset becomes suitable for training at Detectron2.

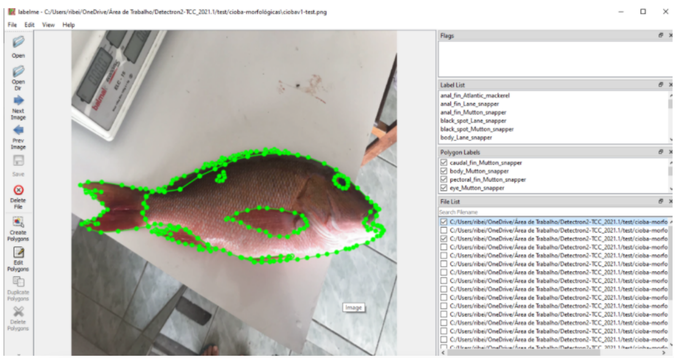

Fig. 2. Labeling of the Mutton snapper class fish parts.

C. Division of data

There is no predetermined separation of training/validation/testing in the literature. Through this, a Script was elaborated, to separate the data set automatically. This dataset consists of 256 images for training and 54 images for testing and validation (approximately 70%/15%/15%), formed, respectively, by 1,000 instances, 450 instances and 430 instances. The RGB training images have a dimension of 512 x 512 pixels, resulting in the input form 512 x 512 x 3 channels, according to Figure 2. In summary, this study analyzed 3 types of families distributed in 8 types of species. It is important to note that species from the same family have very similar characteristics. However, from the morphological segmentation, it is possible to identify precisely the peculiarities of each species. Besides, with the limited number of images in the dataset and avoiding overfitting, data augmentation strategies were used, including random rotation, random flip, brightness, and contrast adjustment techniques, to improve the results of the model final metrics.

D. Detectron2 Training

Detectron2 is a state-of-the-art Facebook AI Research (FAIR) open-source system that implements object detection and segmentation algorithms (WU et al., 2020). This platform is a modular library that implements object detection algorithms, such as Mask-RCNN. Also, it is based on PyTorch 1.6, a fast and effective tensor library used to structure deep learning and calculations in a Central Process Unit (CPU) and Graphics Processing Unit (GPU) environment. Besides, it can be used as a library to promote support for different projects implemented based on it. During the segmentation training of the fish parts, Mask-RCNN was used with the backbone `mask-rcnn_R_50_FPN_3x.yml` and `mas-rcnn_R_101_FPN_3x.yml`. The model was executed in a cloud computing service (Google Colab), in an approximate time of 17 minutes.

Detectron was developed in 2018 and has since become one of the widely adopted open-source projects to further accelerate research in the areas of object detection, *e.g.*, road damage detection and classification [10], segmentation, and the estimation of human posture [11]. It is worth mentioning that it provides support for the implementation and evaluation of novel computer vision research [12]. Detectron has changed over the past two years, developers have implemented the following object detection algorithms on this platform, to know, Mask R-CNN, RetinaNet, Faster R-CNN, Fast R-CNN, RPN, Cascade R-CNN, TensorMask, PointRend, DensePose, Point Read, DeepLab, among others. This framework supports detection tasks, such as detection of bounding boxes and instance segmentation masks, semantic segmentation, and finally, panoramic segmentation. Besides, this approach provides an intuitive and easy programming environment that allows researchers to design new models and experiment with them, and also, it has a modular design that allows users to modify each module independently, making an algorithm more efficient and flexible. It is important to note that this environment is scalable and performs very fast calculations, and also provides different sets of recognized data.

E. Mask-Region-Based Convolutional Neural Network

One of the most powerful frameworks for instance segmentation is Mask-RCNN. It was introduced by FAIR, which combines object detection and semantic segmentation, an evolution of the Regional Convolutional Network (R-CNN), and is an extension of Faster R-CNN. Proposed by He *et al.*, this model efficiently detects objects in an image while simultaneously generating a high-quality segmentation mask for each instance, making it useful for automated inspection applications. Mask-RCNN is essentially a Faster R-CNN with 3 output branches, as shown in Figure 3. The first calculates the coordinates of the bounding box, the second calculates the associated classes and the third provides the segmentation mask for each Region of Interest, *i.e.*, for each object in the image, in parallel with the existing branch for classification and regression of the bounding box [6].

Figure 3 shows the Mask-RCNN architecture for the segmentation of morphological characteristics is composed of four modules. The first module is a feature extraction module that generates a high-level characterized representation of the input image. The second module is a convolutional neural network (CNN), which proposes RoI in the image, based on the characterized image. The third module is a CNN that tries to classify the objects in each RoI. The fourth module performs

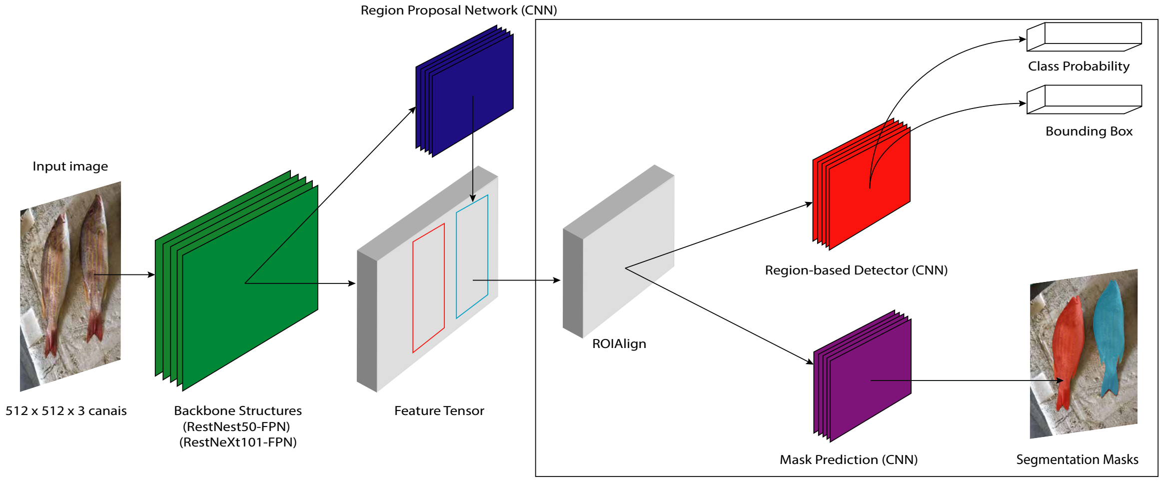

Fig. 3. Mask-RCNN architecture for instance segmentation.

the segmentation of the image, with the objective of generating a binary mask for each region.

1) Module of Backbone Structures: The first module, proposed for the model, transforms the pixels of the input image into a high-level representation. According to related works, the best results can be obtained with extractors with more modern characteristics [13]. In [9], it is observed that the object detection network with the ResNeXt-101-FPN feature extractor results in a higher prediction precision in the CPIS data set than the same detection network. objects with a ResNet101-FPN feature extractor. Therefore, the backbone structures used to compare their performances were ResNet50-FPN and ResNeXt-101-FPN, feature extraction modules.

2) RestNet e ResNeXt Architectures: Regarding the detection and segmentation of objects, the more layers, the longer it takes to train, however, the accuracy tends to be higher, especially in the detection of complex objects. In this study, ResNet and ResNeXt were used. ResNeXt mostly presents the best results when compared to ResNet, since it uses several parallel convolutional layers.

3) Regional Proposal Network Module: The second module proposed for this architecture is the regional proposal network (RPN). The RPN receives a map of features of any size as input and produces a set of proposals for rectangular objects, each with a score describing the probability that the region contains an object or parts of the object. After RPN, it is passed by RoIAlign which quantifies free bilinear interpolation that preserves spatial information. These fixed-size RoIs enter into three parallel processes, to know, *(i)* class of the object and its respective probability; *(ii)* bounding box; and *(iii)* segmentation mask.

4) Region Based Detector: The region-based detector (RBD) can select a fixed number of regions from the original image. The RBD is used to classify the morphological characteristics of each class of fish, and to delimit the bounding box coordinates. The RBD is based on the Faster R-CNN object detection network.

5) Mask Prediction: The mask prediction is a mask that encodes the spatial layout of an input object, in this case, it would be the two objects (fish) present in the model entry photo, the branch of the mask receives positive RoI and predicts the mask using a fully convolutional network (FCN). Thus, unlike class labels and bounding boxes, they are inevitably encapsulated in short output vectors by fully connected layers. Extracting the spatial structure of the masks can be treated naturally by pixel-by-pixel correspondence during convolutions.

V. Analysis of Results and Discussion

Table 1 presents the COCO metrics regarding the segmentation of instances. As noted, during the training of the backbones, the ResNeXt101-FPN backbone presented the best results concerning the precision metrics (AP_{50}, AP_{75}, AP_{small}, AP_{medium}, and AR_{100}), distributed in the mask type and bounding box, being superior regarding the ResNet50-FPN backbone structure.

In addition, the task of inference on the test data set is performed, to analyze the performance of the COCO metrics visually. In Figure 4, there is a difference concerning the contour of the segmentation mask over the body of the red snapper fish and over its head, and the bounding boxes (known as box) present percentages (represent the accuracy of the bounding box on each instance). The IoU is a popular metric in segmentation calculated by the formula (1). The IoU is an evaluation metric used to measure the accuracy of an object detector (parts of the fish) on a specific set of data. A result of IoU > 0.5 is considered a good forecast.

$$IoU = \frac{area\ of\ overlap}{area\ of\ union} = \frac{\#TP}{\#TP + \#FP + \#FN} \quad (1)$$

Where respectively, true positive (TP), false negative (FP), and false negative (FN) pixel numbers are obtained from the test suite using the ResNeXt101-FPN, in which it reached an IoU-based accuracy of 0.884 in a limited resource scenario.

Backbone	Type	AP	$\mathbf{AP}_{50}$	$\mathbf{AP}_{75}$	$\mathbf{AP}_{small}$	$\mathbf{AP}_{medium}$	$\mathbf{AR}_{100}$
ResNet50-FPN	*Mask*	63.095	87.088	82.250	55.255	76.510	76.2
	Box	67.014	87.524	83.454	58.160	79.120	75.7
ResNeXt101-FPN	*Mask*	74.725	91.791	88.705	64.789	82.765	79.5
	Box	75.778	94.927	89.834	66.017	81.456	79.4

Figure 4 shows the segmentation process result using the ResNeXt101-FPN method.

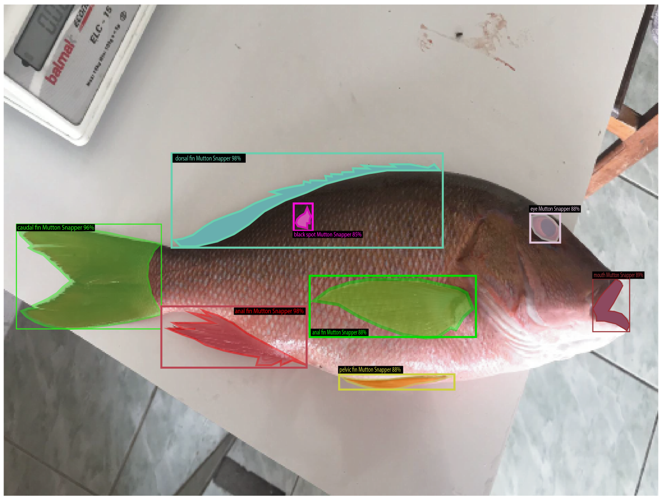

Fig. 4. Inference regarding the Mutton snapper class (ResNeXt101-FPN).

VI. Conclusion and Future Works

This paper presented a segmentation model of instances/morphological characteristics of fish species. From the above, it was possible to build a dataset with images in real environments. Still, after the implementation and comparison of the results of the backbones, it was found that the ResNeXt101-FPN architecture had better performance when comparing the metrics (AP, AP_{50}, AP_{75}, AP_{small}, AP_{medium}, and AR_{100}) with the ResNet50-FPN model. Given the above, it is concluded that the model can be applied to optimize the professional's time in identifying specific characteristics of a particular species. Also, this type of method contributes to the fish overfitting prevention.

Considering the difficulties for the model evaluation, the construction of the dataset is still being conducted. This task requires more time to collect the images. However, with the implementation of the data increase, it could optimize the performance of the evaluation metrics when prevent overfitting overlapping the dataset. Another challenge was the preparation of the images in the COCO format, using the Labelme tool, from the individual segmentation stage of the instances to the result analysis to understand the performance of each backbone.

Future work for the continuation of this proposal is based on increasing the results of the metrics related to the segmentation mask and bounding box and, next to that, it is expected to elaborate the mockups of the screens for coding the web application later. Thus, it is expected that the model can contribute to the recognition of patterns of morphological characteristics of species in fish, supporting society, industry, and government.

Acknowledgements

This work was supported by the Pró-reitoria de Pesquisa, Pós-graduação e Inovação (PRPI) of the IFCE through the Institutional Scholarship Program for Initiation in Technological Development and Innovation (PIBITI), Project nº 11.384, Edital nº 3/2020 PRPI/REITORIA-IFCE; by FCT/MCTES through national funds and when applicable co-funded EU funds under the Project UIDB/50008/2020; and by Brazilian National Council for Scientific and Technological Development - CNPq, via Grant No. 313036/2020-9.

References

[1] J. van Senten, M. A. Smith, and C. R. Engle, "Impacts of COVID-19 on US aquaculture, aquaponics, and allied businesses," *J. World Aquac. Soc.*, vol. 51, no. 3, p. 574, 2020.

[2] S. Yuheng and Y. Hao, "Image segmentation algorithms overview," *arXiv preprint arXiv:1707.02051*, 2017.

[3] Y. Wu, A. Kirillov, F. Massa *et al.*, "Detectron2 Beginner's Tutorial *On-line*," 2019. [Online]. Available: https://colab.research.google.com/drive/16jcaJoc6bCFAQ96jDe2HwtXj7BMD_-m5

[4] S. Pouyanfar, S. Sadiq, Y. Yan *et al.*, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–36, 2018.

[5] P. Wang, P. Chen, Y. Yuan *et al.*, "Understanding convolution for semantic segmentation," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*. Lake Tahoe, NV, USA: IEEE, Mar. 12–15 2018, pp. 1451–1460.

[6] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy: IEEE, Oct. 22-29 2017, pp. 2961–2969.

[7] C. Yu, X. Fan, Z. Hu *et al.*, "Segmentation and measurement scheme for fish morphological features based on Mask R-CNN," *Information Processing in Agriculture*, vol. 7, no. 4, pp. 523–534, 2020.

[8] Z. Ghomi, R. Mirshahi, A. Fattahpour *et al.*, "Segmentation of COVID-19 pneumonia lesions: A deep learning approach," *Medical Journal of The Islamic Republic of Iran (MJIRI)*, vol. 34, no. 1, pp. 1216–1222, 2020.

[9] O. L. Carvalho, O. A. Carvalho Júnior, A. O. Albuquerque *et al.*, "Instance segmentation for large, multi-channel remote sensing imagery using Mask-RCNN and a mosaicking approach," *Remote Sensing*, vol. 13, no. 1, p. 39, 2021.

[10] V. Pham, C. Pham, and T. Dang, "Road damage detection and classification with Detectron2 and faster R-CNN," *arXiv preprint arXiv:2010.15021*, 2020.

[11] B. Artacho and A. Savakis, "UniPose: Unified human pose estimation in single images and videos," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE, 2020, pp. 7035–7044.

[12] R. Girshick, I. Radosavovic, G. Gkioxari, P. Dollár, and K. He, "Detectron," https://github.com/facebookresearch/detectron, 2018.

[13] W. Huang, Y. Xue, L. Hu, and H. Liuli, "S-EEGNet: Electroencephalogram signal classification based on a separable convolution neural network with bilinear interpolation," *IEEE Access*, vol. 8, pp. 131 636–131 646, 2020.