ICICoS 2020

10 - 11 November 2020

Conference Program

The 4th International Conference on Informatics and Computational Sciences

Organized by:
Department of Informatics
Faculty of Science and Mathematics
Universitas Diponegoro

Sponsored by:
IEEE INDONESIA SECTION

COMMITTEES

Steering Committee:

- Prof. Widowati, Universitas Diponegoro, ID
- Prof. W. Jatmiko, Universitas Indonesia, ID
- Prof. J. Jamari, Universitas Diponegoro, ID
- R. Kusumaningrum, Universitas Diponegoro, ID
- Kurnianingsih, Politeknik Negeri Semarang, ID
- NA. Setiawan, Universitas

General Chair:

- DMK. Nugraheni, Universitas Diponegoro, ID

General Co-chairs:

- PW. Wirawan, Universitas Diponegoro, ID
- R. Saputra, Universitas Diponegoro, ID

Secretary:

- S. Adhy, Universitas Diponegoro, ID
- SN. Endah, Universitas Diponegoro, ID

Finance:

- B. Noranita, Universitas Diponegoro, ID
- Khadijah, Universitas Diponegoro, ID

Program Chair:

- FA. Nugroho, Universitas Diponegoro, ID

Publication Chair:

- Rismiyati, Universitas Diponegoro, ID
- E. Suharto, Universitas Diponegoro, ID

Technical Program Committee Chairs:

- A. Wibowo, Universitas Diponegoro, ID
- MA. Riyadi, Universitas Diponegoro, ID

Technical Program Committee Member:

- AH. Ali, Universiti Kuala Lumpur, MY
- A. Alsobeh, Yarmouk University, JOR
- A. Antonyova, University of Prešov in Prešov, SVK
- A. Azizan, Universiti Teknologi Malaysia (UTM), MY
- A. Bagwari, IEEE Member, UTU, IN
- A. Ghasempour, ICT Faculty, USA
- A. Hamad, University of Baghdad, IRQ

- A. Miry, Al-Mustansiriyah University, IRQ
- A. Mobashsher, The University of Queensland, AU
- AM. Nasrabadi, Shahed University, IRN
- A. Mukherjee, Jiangsu University, CN
- A. Najmurrokhman, Universitas Jenderal Achmad Yani, ID
- AP. Widodo, Diponegoro University, ID
- A. Rafiei, University of Technology Sydney, AU
- A. Sallam, Taiz University, YEM
- A. Sharma, Quantum University, Roorkee, Uttarakhand, IND
- AS. Shibghatullah, UCSI University, MY
- A. Sunyoto, Universitas Amikom Yogyakarta, ID
- A. Testa Ministry of Economy and Finance, ITA
- A. Wibowo, Diponegoro University, ID
- B. Jiang, The City University of New York, USA
- B. Warsito, Diponegoro University, ID
- C. Kerdvibulvech, National Institute of Development Administration, THA
- C. Sonagiri, Institute of Aeronautical Engineering, IND
- C. Yuen, Singapore University of Technology and Design, SG
- D. Andriana, Indonesian Institute of Sciences, ID
- D. Chisanga, The University of Zambia, ZMB
- D. Ciuonzo, University of Naples Federico II, ITA
- DH. Fudholi, Universitas Islam Indonesia, ID
- D. Krstić, Faculty of Electronic Engineering, University of Niš, SRB
- DMK. Nugraheni, Universitas Diponegoro, ID
- D. Nurjanah, Telkom University, ID
- D. Wijaya, Telkom University, ID
- E. Babulak, Liberty University, USA
- E. Jiménez Macías, University of La Rioja, ESP
- E. Kaburuan, BINUS University, ID
- E. Werbin, Universidad Nacional de Cordoba, ARG
- F. Aadil, COMSATS University Islamabad Attock Campus, PAK
- F. Farikin, Diponegoro University, ID
- G. Abbas, GIK Institute of Engineering Sciences & Technology, PAK
- G. Dekoulis, Aerospace Engineering Institute (AEI), CYP
- G. Palmerini, Sapienza Università di Roma, ITA
- G. Tambouratzis, Institute for Language & Speech Processing, GRC
- H. Abdulkareem, Al-Nahrain University, IRQ
- H. Alasadi, IRAQ- BASRA, IRQ
- H. El Abed, Technical Trainers College (TTC), SA
- H. Khattak, COMSATS University, Islamabad, PAK
- H. Ng, Multimedia University, MY
- H. Pan, New Jersey Institute of Technology, USA
- H. Pardede, Indonesian Institute of Sciences, ID
- HLH. Spits Warnars, Bina Nusantara University, ID
- H. Toba, Maranatha Christian University, ID

- I. Song, Korea Advanced Institute of Science and Technology, KOR
- I. Yulita, Universitas Padjadjaran, ID
- J. Ali, University of Technology, IRQ
- J. Chin, Multimedia University, MY
- J. Suba, University of the Assumption, PHL
- K. Al-Saedi, Mustansiriyah University, IRQ
- K. Hopkinson, Air Force Institute of Technology, USA
- K. Kabassi, Ionian University, GRC
- K. Majumder, West Bengal University of Technology, IND
- L. Boubchir, University of Paris 8, FR
- MA. Ahmad, Universiti Malaysia Pahang, MY
- MKI. Ahmad, Universiti Tun Hussein Onn Malaysia, my
- M. Chen, Electrical Engineering, Da-Yeh University, TW
- M. Hasibuan, University Gadjah Mada, ID
- M. Koyimatu, Universitas Pertamina, ID
- M. Mustaffa, Universiti Putra Malaysia, MY
- M. Nyirenda, University of Zambia, ZM
- M. Riasetiawan, Universitas Gadjah Mada, ID
- M. Riyadi, Diponegoro University, ID
- M. Tuloli, Universitas Negeri Gorontalo, ID
- M. Younis, University of Baghdad, IRQ
- M. Yusuf, University of Trunojoyo, ID
- N. Al-Azzawi, Baghdad University, IRQ
- NC. Basjaruddin, Politeknik Negeri Bandung, ID
- N. Che Pa, Universiti Putra Malaysia, MY
- N. Ifada, University of Trunojoyo Madura, ID
- N. Ojha, Sharda University, Greater Noida, UP, IND
- N. Prabaharan, SASTRA Deemed University, IND
- O. Ridha , University of Baghdad, IRQ
- O. Manad, UMANIS SA, FR
- P. Corral, Miguel Hernández University, ES
- P. Crippa, Università Politecnica delle Marche, ITA
- P. Mursanto, Universitas Indonesia, ID
- P. Zhang, Stony Brook University, USA
- R. Agustin, Universitas Pasundan, ID
- R. Fyath, Al-Nahrain University, College of Engineering, IRQ
- RC. Herrera Lara, National Polytechnic School, EC
- R. Kusumaningrum, Diponegoro University, ID
- R. Ngadiran, Universiti Malaysia Perlis, MY
- R. Nugroho, Soegijapranata Catholic University, AU
- R. Palma-Orozco, Instituto Politécnico Nacional, MX
- S. Alkassar, Ninevah University, IRQ
- S. Birla, Manipal University Jaipur, IND
- S. Biryuchinskiy, Vigitek, Inc, USA
- S. Horng , National Taiwan University of Science and Technology, TW

- S. Mardiyanto, Institut Teknologi Bandung, ID
- S. Nazir, University of Peshawar, PAK
- S. Ooi, Multimedia University, MY
- S. Sarkar, Vijaya Vittala Institute of Technology, IND
- S. Singh, Manipal Institute of Technology, IND
- S. Suryono, Diponegoro University, ID
- S. Tadisetty, Kakatiya University College of Engineering and Technology, IND
- S. Tan, Newcastle University, GB
- T. Chen, MAXEDA Technology, TW
- T. Manjunath, Dayananda Sagar College of Engineering, Bangalore, Karnataka, IND
- T. Nguyen, Purdue University Fort Wayne, USA
- T. Yingthawornsuk, King Mongkut's University of Technology Thonburi, THA
- V. Ramos, Universidad Autonoma Metropolitana, MX
- V. Omwenga, Strathmore University, KEN
- W. Sunindyo, Institut Teknologi Bandung, ID
- W. Suwarningsih, Indonesian Institute of Science, ID
- Y. Su, Shu-Te University, TW
- Z. Mekkioui, University of tlemcen, DZ

List of Abstracts

C.

D.

E.

F.

2020 4th International Conference on Informatics and Computational Sciences (ICICoS)

A Comparison of YOLO and Mask R-CNN for Segmenting Head and Tail of Fish

Eko Prasetyo
Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
Department of Informatics, Engineering Faculty
University of Bhayangkara Surabaya
Surabaya, Indonesia
eko@ubhara.ac.id

Nanik Suciati*
Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
*corresponding author :
nanik@if.its.ac.id

Chastine Fatichah
Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
chastine@if.its.ac.id

***Abstract*— The visual appearance of the fish's head and tail can be used to identify its freshness. A segmentation method that can well isolate those certain parts from a fish body is required for further analysis in a system for detecting fish freshness automatically. In this research, we investigated the performance of two CNN-based segmentation methods, namely YOLO and Mask R-CNN, for separating the head and tail of fish. We re-trained the YOLO and Mask R-CNN pre-trained models on the Fish-gres dataset consisting of images with high variability in the background, illumination, and overlapping objects. The experiment on 200 images containing 724 heads and 585 tails annotated manually indicated that both models work optimally. YOLO's performance was slightly better than Mask R-CNN, shown by 98.96% and 96.73% precision, and 80.93% and 75.43% recall, respectively. The experimental result also revealed that YOLO outperforms Mask R-CNN with mAP of 80.12% and 73.39%, respectively.**

Keywords*— *segmentation, object detection, YOLO, Mask R-CNN, fish freshness, head and tail of fish

I. Introduction

A system that automatically recognizes fish freshness from its image is a smart solution to help people select good quality fish in easy, real-time, and non-destructive ways. Moreover, to classify five levels of fish freshness, we need a more accurate system [1]. An automatic recognition system also could be a simple, fast, and easy-to-use identification tool, but recognizing fish freshness using whole fish, fillets, or skin is not adequate as a basis for classifying freshness due to changes in physical properties [2]. Therefore, the classification of fish freshness should be more explored using certain body parts such as head or tail through imaging. To encourage high performance in classification, we have to segment the head and tail of fish as ROI (region of interest) using the proper segmentation method.

Image segmentation is one of the essential steps in the vision system [3] because it produces the main object in the problem solved. In particular, object detection-based segmentation using Convolutional Neural Network (CNN) serves up impressive results with high performance, such as YOLO (You Only Look Once) [4] and Mask R-CNN (Region-Based Convolutional Neural Networks) [5]. Research conducted by [6] customized YOLOv3 to detect and locate a single class (license plate of the vehicle). The system can detect objects with 98.22% accuracy and recognize the object with 78% accuracy by using 2049 images. Research conducted by [7] created Faster-YOLO that improved YOLO version 2 [8] using deep random kernel convolutional extreme learning machine (DRKCELM) and double hidden layer extreme learning machine auto-encoder (DLELM-AE) joint network as a feature extractor for classification and object detection. The result showed that Faster-YOLO gained more accurate 1.1% than YOLO version 2 and two times faster than YOLO version 3 [4]. In [9] also modified YOLOv3 for detecting apples with different maturity. The system is implemented on 480 apple images and is compared to state-of-the-art such as original YOLOv3 and Faster R-CNN. Research conducted by [10] combined Residual Network (ResNet) and DenseNet as a backbone of Mask R-CNN to segment overlapped apple images. The system achieved a precision of 97.31% and a recall of 95.70%; also, the model is faster than Mas R-CNN. The research by [11] used U-Net's backbone to improve Mask R-CNN object detection for detecting three growth levels of apple flowers. The model achieved and mean intersection over union (mIoU) 91.55% and mean average precision (mAP) 0.594. Research conducted in [12] created a CNN framework to segment body fish in underwater videos. Also, [13] uses Mask R-CNN to segment the fish body and pupil's eyes to generate morphological features. Much of the research on object detection that has been done by previous researchers is on specific domains. Likewise, in the case of detection of fish heads and tails, we need an object detection segmentation method with optimal performance to support the freshness classification of fish. Therefore it is necessary to evaluate the performance of the object detection-based segmentation method applied to fish heads and tails to gain a segmentation method that works optimally.

The head and tail are body parts used for non-destructive fish freshness classification. However, there is not yet research using these body parts as the basis for the study. Recognizing fish freshness automatically using the visual appearance of body parts raises a huge challenge for isolating those certain parts from a fish body and background. Nevertheless, there are several limitations in the automatic system through imaging, such as lighting, variations in background, and the number of fish in an image. Particularly the number of fish in an image, our dataset consists of single fish, multi-fish, and overlapping fish, we will describe in the next section.

978-1-7281-9526-1/20/$31.00 ©2020 IEEE

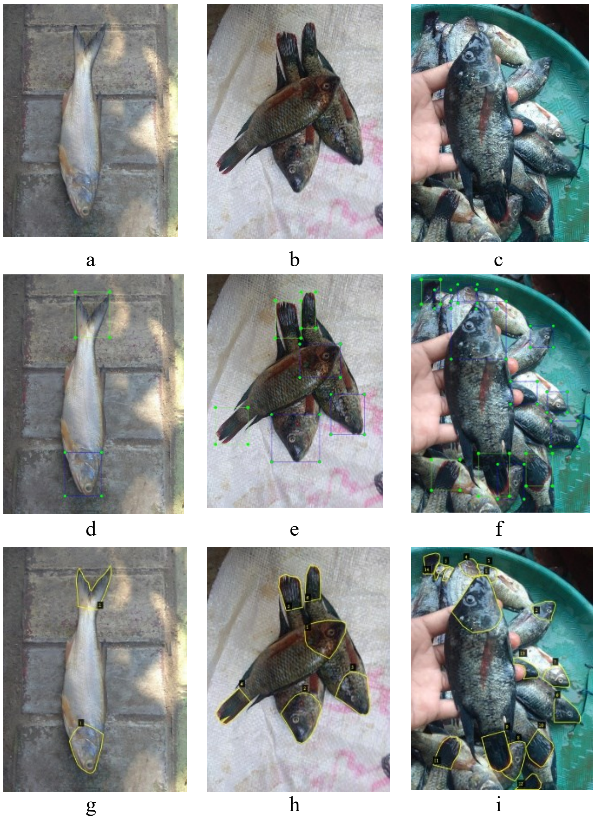

Fig. 1. Sample images; (a)-(c) Orginal images; (d)-(f) YOLO annotation; (g)-(i) Mask R-CNN annotation

In this research, we investigate object detection-based segmentation to gain a proper segmentation method. We carried out comparisons of object detection-based segmentation methods using CNN, such as YOLOv3 and Mask R-CNN, to establish which method provides optimum results for fish body parts segmentation. We used 200 randomly selected images from the Fish-gres dataset, split into 160 training images and 40 test images. The metrics used in performance comparison as follows, precision, recall, AP (average precision), and mAP (mean of average precision). From these comparisons, we expect to know the proper segmentation method to obtain fish heads and tails as the input for classifying fish freshness.

II. Research Methodology

A. Dataset

Fish-gres is a fish dataset collected in the conventional market in Gresik Regency, East Java, Indonesia, by freely photographing fish with variability in the backgrounds, the number of objects, illumination, and overlapping object [14]. The dataset consists of eight species (class), each species contains numerous images ranging from 240 images to 577 images with 624x832 pixel resolution. We randomly selected 200 images among them as the dataset for the segmentation of fish's heads and tails during CNN modeling. We picked images with various combinations, including single fish, multi-fish, and overlapping fish, as shown in Fig. 1 (a)-(c). Single fish is one fish object in an image; multi-fish is several fish objects in an image without overlapping both the head and tail of fish and overlapping fish is several fish objects with complicated positions and overlapping on the head or tail of fish. We split the dataset by the 80:20 ratio, where 160 images are used as training data and 40 images as test data. Each image contains ROI annotation of the head and fish, which we will explain in the next sub-section.

B. Annotation

Object detection-based segmentation requires annotation as a component in the form of a region of interest (ROI) as the classification and localization target. We annotate YOLO using the Labelimg tool to create a bounding box for each head, and tail fish object, the YOLO annotation format used is [x, y, h, h]. Where x and y are the center points of the bounding box, the h and w are the height and width of the bounding box. All values are relative to the image size. In Mask R-CNN segmentation, we also use the VGG Image Annotator (VIA) tool to create polygon annotation and class labels for each head and tail object. In our 200 images, we create 1309 annotations consists of 724 head and 585 tail annotations and split into 1073 annotations (591 heads and 482 tails) and 236 annotations (133 heads and 103 tails) for training and validation, respectively. The examples of the annotations are shown in Fig. 1(d-f) for the YOLO annotation, while Fig. 1(g-i) for the annotation of Mask R-CNN.

C. YOLO

YOLO (You Only Look Once) was developed by Redmon and Farhadi from the University of Washington, used for object detection based-segmentation [15]. YOLO evolved from version 1 [15] to version 4 [16] with various architectural improvements for optimal performance. YOLO also uses a darknet engine for class and bounding boxes prediction. We use the YOLOv3 [4] architecture with weight initialization from darknet53.conv.74.

D. Mask R-CNN

Mask R-CNN (Region-Based Convolutional Neural Networks) [5] was invented from Faster R-CNN [17], where Faster R-CNN uses ROI pooling that performed on region proposal from feature maps, then goes through the classification layer (using Fully Connected Layer) to produce class and bounding box of ROI. Parallel to the classification layer, the Mask R-CNN adds a branch at the end of the network for mask (instance) segmentation. We train the model using weights initialization from the pre-trained Mask R-CNN that were trained with the COCO dataset.

E. Training the Model

In our experiment, we use Colab with Tensorflow 2.0.8 and Keras 1.15.0 (according to the requirements of Mask R-CNN using Tensorflow version 2 and Keras version 1). We train the model of Mask R-CNN using hyperparameters as follows, backbone ResNet101, epoch 100, step-per-epoch 100, minimal confidence 0.9, batch size 1, learning rate 0.001. In the YOLO experiment, we also use Colab with Tensorflow 2.3.0 and Keras 2.4.3, the hyperparameters as follows, max batch 4000, steps 3200 and 3600, batch size 64, subdivision 16. During training, we utilize GPU from Colab with one GPU.

F. Metric Performance

The basic metrics used in object detection are Confidence and IOU (Intersection over Union), where confidence is the probability that an anchor box contains an object; it is typically predicted by a classifier of the method. We don't need to measure performance with this metric because confidence only decides whether an anchor is an object or not. Meanwhile, IOU is a metric for calculating the similarities between the anchor predicted bounding box and the ground truth bounding box, which is expressed using a percentage between the intersection and the union of the predicted boxes and the ground truth boxes. We set a threshold of 0.5 as the similarity boundary, which means all predicted bounding boxes> 0.5 will be considered as a predicted object.

During comparison performance, we used metrics as follows, TP (true positive), FP (false positive), Precision, Recall, AP (Average Precision), and mAP (mean of Average Precision). TP is ROI in ground truth that is successfully detected; usually, this is an object detected with IOU > 0.5. FP is a non-ground-truth object and has IOU > 0.5 detected. We do not use TN (true negative) because it is not appropriate to discuss the ROI that is not detected, and while evaluation with Recall can be measured without using FN. TN (true negative) is also not used since it is an object not to be detected and should not be discussed.

Precision is the system's strength to present correct data, measured as a percentage between the number of true positives and the number of true positives and false positives. A recall is the capacity of the system to predict positive data, which is expressed as the percentage between the number of true positive data and the ground truth data (sum of true positive and false negatives). We employ the following equation.

$$p = \frac{TP}{TP + FP} \tag{1}$$

$$r = \frac{TP}{TP + TN} = \frac{TP}{number\ of\ ground\ truth} \tag{2}$$

Average precision (AP) is the generally accepted standard for measuring the performance of object detection systems [18], besides that, it is also harder to compare two detectors using precision and recall metrics, so in this study, we used AP as the key metric for comparing YOLO and Mask R-CNN performance. AP is the precision averaged across all unique recall levels. To soften the wiggles in the curve, we interpolate the precision at multiple recall levels, then calculating AP. The interpolated precision p_{interp} at a certain level of recall r is defined as the highest precision found for any recall level $r' \geq r$:

$$p_{interp}(r) = \max_{r' \geq r} p(r') \tag{3}$$

AP is calculated from a precision-recall curve where the value is limited from 0 to 1, using the equation below:

$$AP = \int_0^1 p(r)\, dr \tag{4}$$

This integral equation is approximated closely by a sum over the precision at any possible threshold value, multiplied by the change within recall as the expression below:

$$AP = \sum_{i=1}^{n=1} (r_{i+1} - r_i)\, p_{interp}(r_{i+1}) \tag{5}$$

We measure the AP on the head and tail objects, then calculate mAP (mean of average accuracy) as the final performance metric, where the mAP is estimated from the mean of the AP across all K classes. We use the equation below:

$$mAP = \frac{\sum_{i=1}^{K} AP_i}{K} \tag{6}$$

III. Results and Discussions

We assess the performance of both models during training by showing the loss values. YOLO demonstrates a significant reduction in loss from the beginning until around epoch 1350, and the loss hits 1. We continue training until epoch 4000, while YOLO keeps fluctuating around value loss 1, as shown in Fig. 2 (a). We measure the Mask R-CNN 's performance during the training by showing the loss values for both class loss, bounding box loss, mask loss, and overall loss. As seen in the graph in Fig. 2 (b), the average loss value decreases during training, where it drops rapidly at the beginning and drops slowly at the end until it reaches 0.169. This chart trend is also followed by both class loss, bounding box loss, and mask loss. It can be said that the model can improve themselves during training.

The validation results using 40 images (133 heads and 103 tail annotations), as shown in Table 1, indicate that YOLO can detect 115 heads and 76 tails while the R-CNN mask can detect 115 heads and 63 tails as well. Also, both methods detected some objects incorrectly; YOLO detected two objects as heads, while Mask R-CNN detected one object as head, and five objects as tails. YOLO thus achieves precision for head and tail over class 98.29% and 100%, respectively. Though for head and tail, Mask R-CNN has obtained 99.14% and 92.65%, respectively. Compared to ground truth, YOLO gain recall 86.47% and 73.79% for head and tail, respectively, while Mask R-CNN gain 86.47% and 61.17% for head and tail, respectively. We resume all results to be global precision and global recall where YOLO gain 98.96% and 80.93%, respectively, while Mask R-CNN gain 96.73% and 75.43%, respectively. We see that the precision and recall performance of YOLO outperformed Mask R-CNN in all performance metrics except precision-over-class, where YOLO obtained false positives for two objects while Mask R-CNN just one; we show the false-positive result below. Thus, according to precision and recall performance, YOLO promises better segmentation results.

The object detection performance obtained by YOLO with 100% precision over class at a glance shows perfect results, but we shouldn't just look at the percentage. We have to believe that this 100% means that no false-positive tails were present during detection; of course, this is great. However, we also need to look deeper at how many objects should be detected; we will see the 73.79% recall, meaning that 26.21% of objects are not detected. Therefore we ought to evaluate it with other appropriate opinions as well.

We also use AP (average precision) to compare performance, which calculates using a precision-recall curve, as shown in Fig. 3. Due to one FP result, YOLO shows a flat head curve and one step-down because there is one FP, while the tail curve shows only a flat curve because no FP reduces recall performance. The head curve is greater than 0.8, while the tail reaches 0.7 (Fig. 3 (a)-(b)). The mask R-CNN reveals a different curve, where the head curve is closer to the YOLO's head curve, but the tail curve has more step-downs and reaches 0.6 recall (Fig. 3 (c)-(d)). Lots of FP gained causes the curve to have multiple step-downs. Next, we quantify the AP of all the curves we are presenting using equation (4) or (5) where AP is identical to the area under the precision-recall curve.

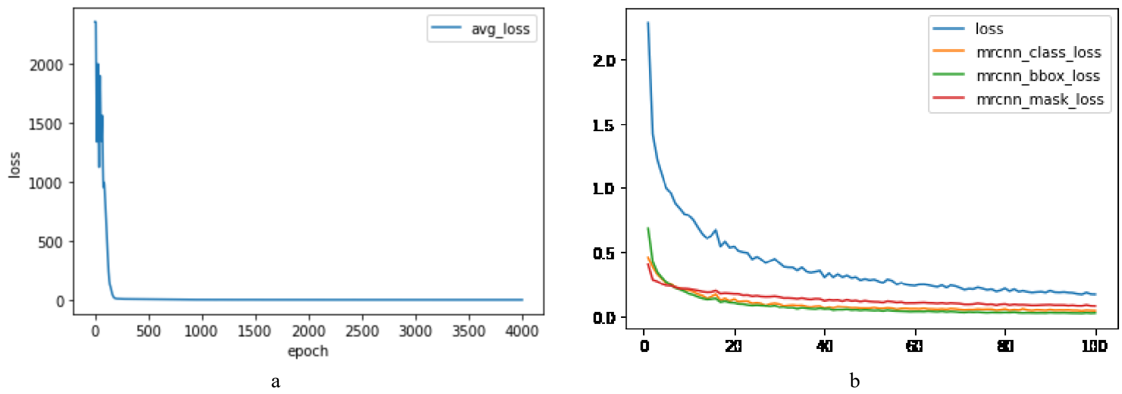

Fig. 2. Performance during training; (a) YOLO; (b) Mask R-CNN

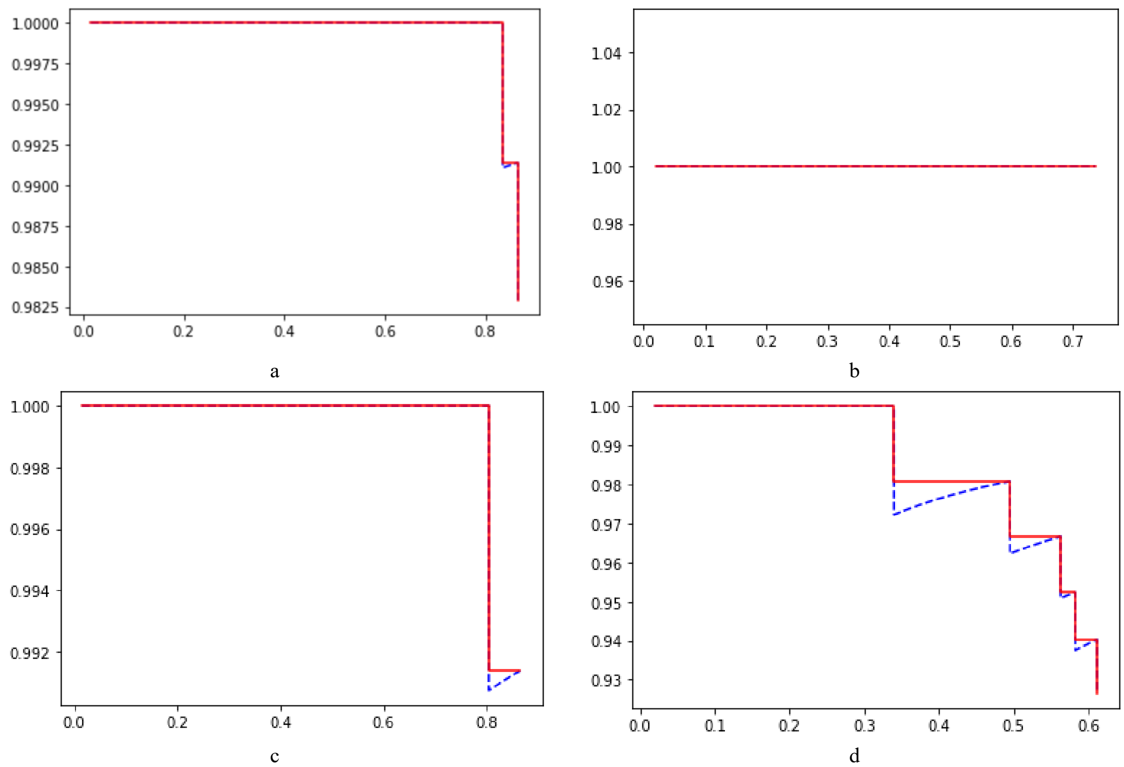

Fig. 3. Average precision of precision-recall curve of head and tail using validation data; (a)-(b) YOLO l; (c)-(d) Mask R-CNN

TABLE I. PERFORMANCE WITH PRECISION AND RECALL

Metric	YOLO		Mask R-CNN	
	Head	***Tail***	***Head***	***Tail***
TP	115	76	115	63
FP	2	0	1	5
Total detected	117	76	116	68
Precision over class	98.29	100	99.14	92.65
Recall over class	86.47	73.79	86.47	61.17
Global Precision	98.96		96.73	
Global Recall	80.93		75.43	

TABLE II. PERFORMANCE WITH AP AND MAP

Model	AP (%)		mAP (%)
	Head	***Tail***	
YOLO	86.44	73.80	80.12
Mask R-CNN	86.41	60.37	73.39

As shown in Table 2, the performance of YOLO and Mask R-CNN by AP shows that YOLO achieves AP 86.44% and 73.80% for head and tail, respectively, and outperforms Mask R-CNN where it reached 86.41% and 60.37% for head and tail, respectively. All models have almost the same AP performance for the head, while for tail YOLO outperforms Mask R-CNN. Ultimately, the mAP results reveal that YOLO outperforms Mask R-CNN by 80.12% and 73.39%, respectively, for YOLO and Mask R-CNN.

As reported by [19] [20] [21], for ecological data, Faster-RCNN and Mask-RCNN achieved the best results at IoU 0.7 and 0.5, respectively. Therefore, we investigate YOLO and Mask-RCNN also using IOU 0.5, which means all the detected bounding box with IOU greater than or equal to 0.5 will be considered the detection result. The smaller the IoU, the more objects will be detected, it means we allow Mask-RCNN to achieve many results.

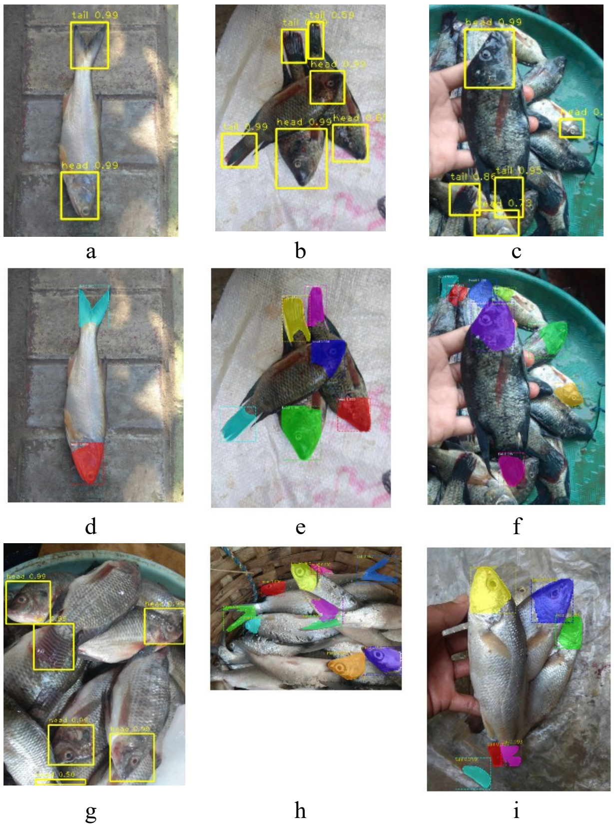

Fig. 4. Sample images; (a)-(c) Segmentation result by YOLO; (d)-(f) Segmentation result by Mask R-CNN; (g) False positive resulted by YOLO; (h-i) False positive resulted by Mask R-CNN

Our experimental results show a different performance where the mAP achieved by YOLO is higher than Mask-RCNN. We observe that the cattle detection carried out by [19] [20] [21] uses a dataset where the object has many variations in direction (captured from many directions). In contrast, our dataset (fish) is only two directions, right and left. Perhaps this dataset situation causes YOLO's performance to improve drastically compared to Mask-RCNN. Apart from achieving better performance, YOLO also has a simpler architecture because it uses a one-stage detector where the backbone and detector are integrated into one architecture and trained simultaneously. Even though the darknet as the backbone is usually pre-trained using imagenet, we retrain all parts of the layer when training YOLO.

The example result of segmentation using YOLO, as presented in Fig. 4 (a-c), shows that the head and tail of a single fish are segmented excellently with high confidence, which in Fig. 4 (a) both head and tail achieves confidence 0.99. The results of multi-fish segmentation (Fig. 4 (b)) shows impressive results where three heads and three tails were detected with confidence ranging from 0.59 to 0.99. The result of overlapping fish segmentation is that not all ROI was detected; a failure occurred in which the model succeeded in detecting only three out of nine heads and two out of four tails (Fig. 4 (c)). The rests were not detected, possibly because they overlapped with other objects.

The segmentation results by Mask R-CNN for single fish also showed excellent results where both head and tail were detected with confidence 1.00, respectively (Fig. 4 (d)). In a multi-fish image, the model can detect three heads and three tails, with confidence varying from 0.924 to 1.00. Meanwhile, the results of the overlapping fish segmentation (Fig. 4 (f)) are better than YOLO, where six out of nine heads and two out of four fish were detected by the model. We also show FP detected by YOLO and Mask R-CNN, as shown in Fig. 4(g), YOLO gets FP where one non-head object is detected with the confidence of 0.5 as the head. In contrast, Fig. 4(h-i) shows that Mask R-CNN gets several FP where one tail is detected as two tails (see Fig. 4(h)), and a non-tail object is also detected as a tail with the confidence of 0.969.

From the results presented above, we evaluate that both models offer optimum performance but with a disadvantage. In single and multi-fish images, YOLO and Mask R-CNN can work optimally, as evidenced by the high confidence for each detected ROI. On overlapping fish images, the two models can not work optimally. YOLO can detect several fish heads and tails, especially those that do not overlap with other objects, whereas the Mask R-CNN can detect several overlapping objects, but not all objects, both the fish's heads and tails, are detected.

IV. Conclusions

The experimental findings show that YOLO outperforms Mask R-CNN by mAP of 80.12% and 73.39%, respectively. In terms of head segmentation, both models work optimally with the limitation in segmenting overlapping objects. In terms of tail segmentation, YOLO outperformed Mask R-CNN with AP of 73.80% and 60.37%, respectively. Precision and recall performance also reveals that YOLO was slightly better than Mask R-CNN with 98.96% and 96.73% precision, respectively. The recall performance achieved by YOLO and Mask R-CNN is 80.93% and 75.43%, respectively. All of these performance results use 160 images as training data and 40 images as validation data. YOLO achieved 100% precision over tail class, but in fact, only detected 76 out of 102 tail annotations, no FP detected. Therefore, it is necessary to improve YOLO and Mask R-CNN performance both by adding data and by modifying the architecture.

References

[1] A. Taheri-Garavand, A. Nasiri, A. Banan, and Y. D. Zhang, “Smart deep learning-based approach for non-destructive freshness diagnosis of common carp fish,” *Journal of Food Engineering*, vol. 278, p. 109930, Aug. 2020.

[2] H. Mohammadi Lalabadi, M. Sadeghi, and S. A. Mireei, “Fish freshness categorization from eyes and gills color features using multi-class artificial neural network and support vector machines,” *Aquacultural Engineering*, vol. 90, p. 102076, Aug. 2020.

[3] F. Garcia-Lamont, J. Cervantes, A. López, and L. Rodriguez, “Segmentation of images by color features: A survey,” *Neurocomputing*, vol. 292, pp. 1–27, May 2018.

[4] J. Redmon and A. Farhadi, “Yolov3,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2017.

[5] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

[6] Hendry and R. C. Chen, “Automatic License Plate Recognition via sliding-window darknet-YOLO deep learning,” *Image and Vision Computing*, vol. 87, pp. 47–56, Jul. 2019.

[7] Y. Yin, H. Li, and W. Fu, “Faster-YOLO: An accurate and faster object detection method,” *Digital Signal Processing: A Review Journal*, vol. 102, p. 102756, Jul. 2020.

[8] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017.

[9] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," *Computers and Electronics in Agriculture*, vol. 157, pp. 417–426, Feb. 2019.

[10] W. Jia, Y. Tian, R. Luo, Z. Zhang, J. Lian, and Y. Zheng, "Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot," *Computers and Electronics in Agriculture*, 2020.

[11] Y. Tian, G. Yang, Z. Wang, E. Li, and Z. Liang, "Instance segmentation of apple flowers using the improved mask R–CNN model," *Biosystems Engineering*, vol. 193, pp. 264–278, May 2020.

[12] N. F. F. Alshdaifat, A. Z. Talib, and M. A. Osman, "Improved deep learning framework for fish segmentation in underwater videos," *Ecological Informatics*, vol. 59, p. 101121, Sep. 2020.

[13] C. Yu *et al.*, "Segmentation and measurement scheme for fish morphological features based on Mask R-CNN," *Information Processing in Agriculture*, Jan. 2020.

[14] E. Prasetyo, N. Suciati, and C. Fatichah, "Fish-gres Dataset for Fish Species Classification." Mendeley Data, 2020.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-December, pp. 779–788.

[16] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," Apr. 2020.

[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[18] K. Oksuz, B. C. Cam, E. Akbas, and S. Kalkan, "Localization Recall Precision (LRP): A New Performance Metric for Object Detection," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11211 LNCS, pp. 521–537, Jul. 2018.

[19] S. Schneider, G. W. Taylor, and S. Kremer, "Deep learning object detection methods for ecological camera trap data," in *Proceedings - 2018 15th Conference on Computer and Robot Vision, CRV 2018*, 2018, pp. 321–328.

[20] M. Danish, "Beef Cattle Instance Segmentation Using Mask R-Convolutional Neural Network Neural Netw," 2018.

[21] B. Xu *et al.*, "Automated cattle counting using Mask R-CNN in quadcopter vision system," *Computers and Electronics in Agriculture*, vol. 171, p. 105300, Apr. 2020.