https://doi.org/10.1051/itmconf/20246701034

Fish grades identification system with ensemble-based key feature learning

Fityanul Akhyar[1*], *Ledya* Novamizanti[2], *Inung* Wijayanto[2], *Cahaya Irham* Wirawan[2], *Dede Chandra* Wijaya[2], *Agno* Fredigo[2], *Ferdi* Ramdhon[3], and *Chih-Yang* Lin[4]

[1]Universiti Tunku Abdul Rahman, Department of Computer and Communication Technology, 31900 Kampar Perak, Malaysia
[2]Telkom University, Bachelor of Telecomunication Engineering, 40257 Buah Batu Bandung, Indonesia
[3] PT. Aruna Jaya Nusantara, Technology Division, 12540 DKI Jakarta, Indonesia
[4]National Central University, Department of Mechanical Engineering, 320317 Zhongli Taoyuan, Taiwan

Abstract. Indonesia has already contacted the maritime nations due to its 5.8 million km² of coastline. Consequently, fish products are among the most important commodities. Moreover, fish grading is a crucial step in the process of exporting fisheries products. Currently, in Indonesia, the process itself is manually inspected by an expert. In addition, this paper proposes to assist the industry by suggesting a method for grading fish. This method involves combining two essential fish parts with different resolutions: the high-level feature (the body) and the low-level feature (the eye) serve as defining characteristics. These two main parts are accurately localized using a deep learning-based object detection model, specifically YOLOv7, and extracted with an efficient and adaptive learned classification model, namely EfficientnetV2S. In the final stage, the two extracted features are combined and learned with Dense Layers to generate three distinct fish grades. Based on the experimental results, the proposed work achieved an accuracy, F1 Score, and recall of 96.88%, 97%, and 97%, respectively. The proposed model outperformed the baseline model, which relies solely on deep learning-based classification, by a significant margin.

1 Introduction

Indonesia is a maritime nation with a total water area of 6,400,000 km2 [1]. This allows Indonesia to excel in the marine and fishing industries. In Indonesia, fisheries products are sold domestically and exported abroad. In order to stimulate marine economy growth in Indonesia, fishermen must ensure that the fish they intend to sell meet export quality standards [2]. In general, the appearance of the eye and body can be used to determine the quality of the fish. Currently, fish are graded based on their quality through physical inspection. It causes results to vary based on the knowledge or experience of each inspector.

*Corresponding author: fityanul@utar.edu.my

In recent years, researchers from around the world have conducted experiments using deep learning to successfully solve object recognition [3] and visual problems [4] problems. Positive results of deep learning in research augur positive applications of deep learning in a variety of industries, including marine and fisheries. Consequently, this study proposed a method for grading fish by combining object detection and classification deep learning-based models. Specifically, in the first stage, the system uses the state-of-the-art object detection-based model or YOLOv7 [5] to locate the key feature with different resolution by cropping two main parts of the raw image region, for taking the low level and high level feature. In the second step, the cropped images are fed to the classification-based model or EfficientnetV2S [6] for feature extraction. In addition, the extracted of two features are ensembled to produce the grade of fish in the final stage.

2 Related Works

Several researchers have previously utilized machine learning in the marine and fishing industries. A great deal of research has been conducted on fish, primarily to classify fish species [7-9] or even to monitor fish habitats [10]. For fish grading itself, a few researchers have done it before. The experiment described in [11] used a simple machine-learning algorithm called Support Vector Machine to process and grade images of fish. The work described in [12] used multispectral imaging to grade fish using the flesh as the input. CNN was used to classify fish into two categories with an accuracy ranging from 90.6% to 99.9%, depending on the species. On the other hand, some more recent models have utilized object detection models to identify fish and generate valuable information. For instance, the work known as FishNet [13], along with others [14-16], performs object detection for fish counting according to specific species. The work in [17] employed one of the state-of-the-art object detection models, well-known as YOLO version five or YOLOv5, to monitor fish in farms, and similarly, the work in [18] utilized YOLOv4 for sorting fish.

Moreover, YOLO is a highly regarded algorithm for object detection in general. It was implemented in a variety of real-world applications. For instance, in [19, 20], the YOLO family was utilized to construct a successful detection system. In addition, we anticipate that the current version of YOLO [5] can also adapt to the task of fish grading. On the other hand, Efficientnet [6] is one of the state-of-the-art classification models that effectively balances the network depth, width and resolution. As well as YOLO, the Efficientnet will be utilized in the fish application task. Therefore, this work aims to unite these two models to achieve the best performance.

3 Proposed Work

Figure 1 illustrates the architecture of the proposed system for the fish grade identification system. We present a method for identifying the grade of fish. According to the review in [21], the YOLOv7 shows better accuracy on the MS COCO dataset rather than the other two current version YOLO or YOLOv6 [22] and YOLOv8 [23]. Using YOLOv7 [5], the high-level feature or fish's body and the low-level feature or eye are extracted from the raw image as the first step in the process. Moreover, these two keys that have different resolutions were identified by the discussion with the expert inspector from PT. Aruna Jaya Nusantara. Then, this stage can help to get these two important factors directly and eliminate the background as un-useful information. Next, the two cropped images are loaded into an equal two-feature extraction EfficientnetV2S [6]. Specifically, this model has an effective compound coefficient that can be used to uniformly scales the depth, width, and resolution dimensions that effectively learn different level features while improving the efficiency of speed

 https://doi.org/10.1051/itmconf/20246701034

performance. The final step is to apply a Dense Neural Network to the ensembled or concatenated and learned features from the two images. In this section, detailed information on the methodology is provided.

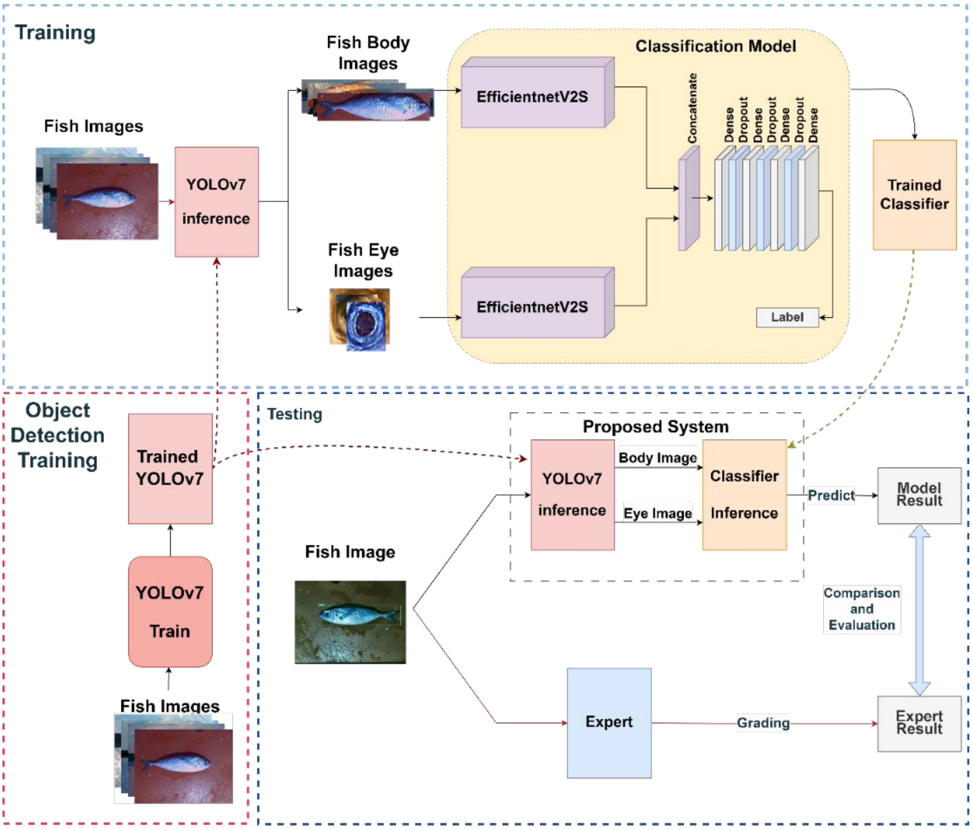


Fig. 1. The Architecture of the Proposed Fish Grading System.

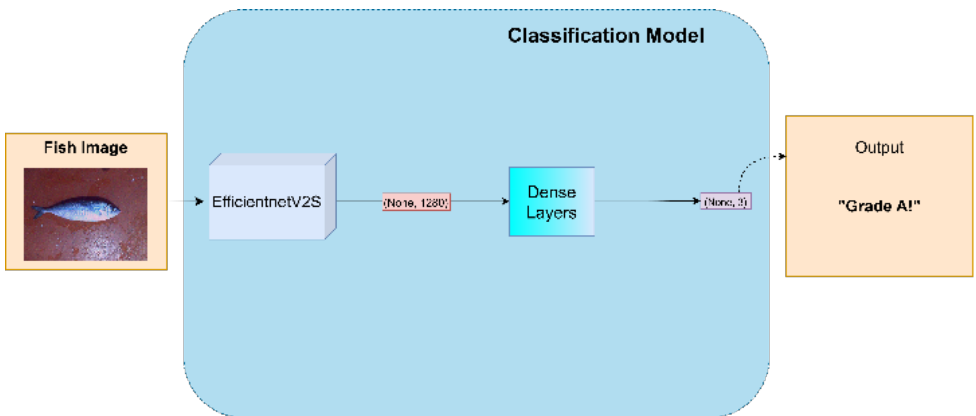
3.1 Detection Stage

The first step is to prepare the dataset for the detection stage by assigning two labels for the fish body and eye to each image. Following labelling, the data are augmented to create a more diverse dataset of fish. In this stage, albumentations [24] is utilized. The resampling enhancement was implemented while also considering bounding-box-safe enhancements. The system used the YOLOv7-based algorithm with the default configuration, or the training process began from scratch with the input image 640x640, followed by mosaic and mix-up augmentation with probability 1.0 and 0.15, respectively.

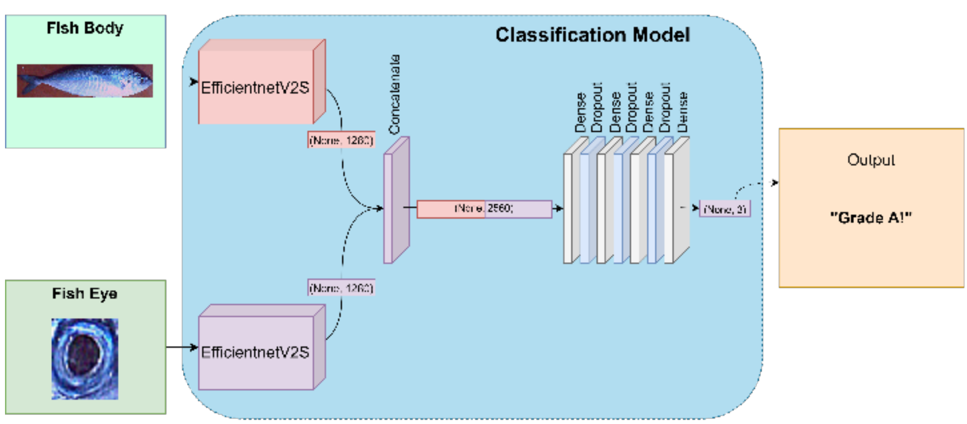
3.2 Classification Stage

The classification stage employs the most advanced model for convolutional neural networks. In this system, we employ transfer learning to utilize a previously constructed or trained deep learning model. EfficientnetV2S [6] is the pre-trained model of choice. As shown in Table 1, this model achieved the highest accuracy among the deep-learning models that were evaluated on nine classes of a Large-Scale Fish dataset. All of the Efficientnet model's layers are not frozen; otherwise, the entire model would be trainable.

As depicted in Figure 2, we enhance the classification baseline model depicted in Figure 2(a) with the proposed classification procedure depicted in Figure 2(b). The proposed method begins with feature extraction using a pre-trained EfficientnetV2S model. Images of the body and eye cropped from object detection will be the input for classification. Only images with detected bodies and eyes will be accepted as input for classification. If neither is detected, the image will be discarded. Each image of the body and eye is reshaped into (224, 224, 3) images before being paired with an EfficientnetV2S model for feature extraction. This procedure generates two output tensors of size (None, 1280) for the body image and eye image, respectively. The next subnetwork describes the grading procedure itself. This process begins with a concatenate layer that receives two inputs from the two EfficientnetV2S. The layers must ensemble the tensors before passing them to the Dense Neural Network. The network has three outputs, one for each of the three fish grades. In addition, we add a three-dropout layers before the final layer to aid in the model's regularization. The layers and details of the model are depicted in Table 2.



(a)



(b)

Fig. 2. (a) The Classification Baseline Model. (b). The Proposed Classification Model.

Table 1. Model Testing on A Large-Scale Fish Dataset.

Model	Accuracy (%)
ShuffleNet [25]	93.02
RepVGG [26]	93.02
TNT [27]	86.05
VAN [28]	74.42
MobileNetV2 [29]	90.70
NasNet [30]	95.35
EfficienNetV2S [6]	**97.67**

Table 2. The Classifier Specification.

Layer	Parameters	Output Shape
Input 1	0	(None, 224, 224, 3)
Input2	0	(None, 224, 224, 3)
Efficientnet 1	20331360	(None, 1280)
Efficientnet 2	20331360	(None, 1280)
Concatenate	0	(None, 2560)
Dense	163904	(None, 64)
Dropout	0	(None, 64)
Dense	2080	(None, 32)
Dropout	0	(None, 32)
Dense	1056	(None, 32)
Dropout	0	(None, 32)
Dense	99	(None, 3)

4 Ablation and Results

A Large-Scale Fish Dataset [31] is the dataset used in this study. The dataset is a publicly available dataset containing images of nine classes of fish species. We reorganized the data into three random groups of fish images. Figure 3 depicts how we modified the data based on a discussion with an expert from PT. Aruna Jaya Nusantara by adding marks to the fish images to distinguish grades A, B, and C. The images of grade A contain no marks or alterations. The grade B images were marked with red on the fish's body or eye. In grade C images, red marks were placed on the fish's body and eye. We split the dataset which contains 319 images into training, validation, and testing with a ratio of 77%, 23%, and 10%, respectively. Besides YOLOv7 as the detection model was trained from scratch, the process for the feature extraction model will be full training as well. Both Efficientnet layers and top layers that we made from scratch will be trained. We assume retraining the Efficientnet model will help our model to get a better result, considering our low number of images in our dataset.

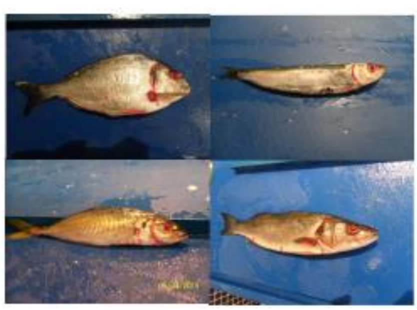
For the hyperparameters as shown on Table 3, we applied adam optimizer with 1e-3 learning rate with categorical cross-entropy loss. We train each model for 80 epochs which we assume will be enough considering the number of data, parameters, and learning rate.

Table 3. The Classifier Specification.

Description	Specification
Optimizer	Adam
Learning rate	0.003
Epochs	80
Batch size	16

(a)

(b)

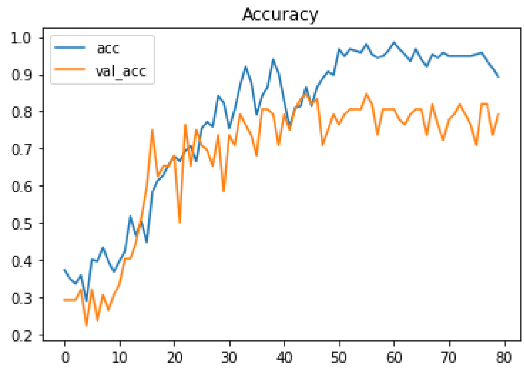


(c)

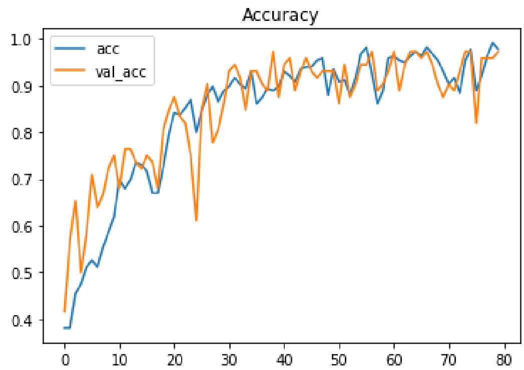
Fig. 3. The fish Dataset. (a) Grade A. (b) Grade B. (c) Grade C.

In the detection stage, our dataset is fed to YOLOv7. In addition, before entering this phase, the fish's body and eye, denoted by bounding boxes, adhere to the Pascal VOC format [32] With an IoU threshold of 0.5, YOLOv7 has achieved mAP accuracy of 99.1% based on experimental results.

Finally, in the classification stage, as shown in Table 4, after we performed the proposed model, the accuracy shows 91.74%, 92.97%, 97% and 97% for validation, testing, F1 Score and Macro Average Recall, respectively. These results are contrasted with the classification baseline model or single EfficientnetV2S that inputs a whole fish image which only gets an accuracy of 71.04%, 68.18%, 72%, and 70% for validation, testing, F1 Score and Macro Average Recall, respectively. Similarly, as shown in Figures 4(a) and (b), the validation curve of the proposed method achieved the highest accuracy than the baseline model. Based on these results, the proposed detection stage that direct captured two important key features that have different resolution, the high-level feature or the body and the low-level feature or the eye. Then extracted and grading these features in the classification stage, achieve significant accuracy.



(a)



(b)

Fig. 4. (a) Classification Baseline Model Accuracy. (b) Proposed Classification Model Accuracy.

Table 4. The Comparison Results

Metrics (%)	Base Model	Proposed Method
Validation Accuracy	79.17	97.22
Test Accuracy	71.88	96.88
F1 Score	72	97
Macro Average Recall	70	97

5 Conclusion

This study described a method for identifying fish grades using the deep learning object detection model YOLOv7 in the detection stage to capture the two essential fish parts. In addition, the proposed classification model with an ensemble of the two components as the key features achieves a testing accuracy of 96.88%, which is significantly better than the baseline model. In the near future, actual sites of PT. Aruna Jaya Nusantara will implement the proposed system for the automatic fish quality inspection system.

This work was supported in part by Universiti Tunku Abdul Rahman, Malaysia. Additionally, research funding was provided by the Directorate of Research and Community Service (DPPM) at Telkom University, under Grant No. 138/LIIT0/PM-LIT/2024.

References

1. K. R. Jones, C. J. Klein, H. S. Grantham, H. P. Possingham, B. S. Halpern, N. D. Burgess, S. H. Butchart, J. G. Robinson, N. Kingston, N. Bhola, One Earth, **2**(2), 188-196, (2020)
2. M. Doddema, G. Spaargaren, B. Wiryawan, S. R. Bush, Marine Policy, **119**, 104100, (2020)
3. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, Int. J. Comput. Vis., **115**, 211-252, (2015)
4. V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, nature, **518**, 529-533, (2015)
5. C.-Y. Wang, A. Bochkovskiy, H.-Y. M. Liao, *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*, In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 7464-7475 (2023)
6. M. Tan, Q. Le, *Efficientnetv2: Smaller models and faster training*, In International conference on machine learning, 10096-10106, (2021)
7. N. E. M. Khalifa, M. H. N. Taha and A. E. Hassanien, *Aquarium family fish species identification system using deep neural networks*, In International Conference on Advanced Intelligent Systems and Informatics, 347-356, (2019)
8. D. Rathi, S. Jain, S. Indu, *Underwater fish species classification using convolutional neural network and deep learning*, In 2017 Ninth international conference on advances in pattern recognition (ICAPR), 1-6, (2017)

9. A. Salman, A. Jalal, F. Shafait, A. Mian, M. Shortis, J. Seager, E. Harvey, L&O Methods, **14**, 570-585, (2016)
10. A. Saleh, M. Sheaves, D. Jerry and M. R. Azghadi, Expert Syst. Appl., (2022)
11. F. Odone, E. Trucco and A. Verri, Appl. Artif. Intell.,**15**, 735-745, (2001)
12. D. Jayasundara, L. Ramanayake, N. Senarath, S. Herath, R. Godaliyadda, P. Ekanayake, V. Herath, S. Ariyawansha, *Multispectral Imaging for Automated Fish Quality Grading*, In 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), 321-326, 121841, (2020)
13. M. Mots' oehli, A. Nikolaev, W. B. IGede, J. Lynham, P. J. Mous, P. Sadowski, FishNet: Deep Neural Networks for Low-Cost Fish Stock Estimation, arXiv preprint arXiv:2403.10916, (2024)
14. K. M. Babu, D. Bentall, D. T. Ashton, M. Puklowski, W. Fantham, H. T. Lin, N. P. Tuckey, M. Wellenreuther, L. K. Jesson, J. Roy. Soc. New Zeal., **53**(1), 52-68, (2023)
15. B. Wu, C. Liu, F. Jiang, J. Li, Z. Yang, Front. Environ. Sci., **11**, 1059217, (2023)
16. H. Yu, Z. Wang, H. Qin and Y. Chen, IEEE Access, (2023)
17. K. S. K. Patro, V. K. Yadav, V. Bharti, A. Sharma, A. Sharma, Natl. Acad. Sci. Lett., **46**, 407-412, (2023)
18. A. Kuswantori, T. Suesut, W. Tangsrirat, G. Schleining, N. Nunak, Appl. Sci., **13**(6), 3812, (2023)
19. F. Akhyar, L. Novamizanti, T. Putra, E. N. Furqon, M.-C. Chang, C.-Y. Lin, *Lightning YOLOv4 for a Surface Defect Detection System for Sawn Lumber*, In 2022 IEEE 5th International Conference on Multimedia Information Processing and Retrieval (MIPR) 184-189, (2022)
20. Y. Pratama, Y. Fujimura, T. Funatomi, Y. Mukaigawa, *Parasitic Egg Detection and Classification by Utilizing the YOLO Algorithm with Deep Latent Space Image Restoration and GrabCut Augmentation*, In 2022 IEEE International Conference on Image Processing (ICIP), 4311-4315, (2022)
21. J. Terven, D. Cordova-Esparza, Mach. learn. knowl., **5**(4), 1680-1716, (2023)
22. C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, YOLOv6: A single-stage object detection framework for industrial applications, arXiv preprint arXiv:2209.02976, (2022)
23. M. Sohan, T. Sai Ram, R. Reddy, C. Venkata, *A Review on YOLOv8 and Its Advancements*, In International Conference on Data Intelligence and Cognitive Informatics,529-545, (2024)
24. A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, A. A. Kalinin, Information, **11**(2), 125, (2020)
25. X. Zhang, X. Zhou, M. Lin, J. Sun, *Shufflenet: An extremely efficient convolutional neural network for mobile devices*, In Proceedings of the IEEE conference on computer vision and pattern recognition, 6848-6856, (2018)
26. X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, J. Sun, *Repvgg: Making vgg-style convnets great again*, In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 13733-13742, (2021)
27. K. Han, A. Xiao, E. Wu, J. Guo, C. Xu, Y. Wang, Adv. Neural. Inf. Process. Syst., **34**, 15908-15919, (2021)

28. M.-H. Guo, C.-Z. Lu, Z.-N. Liu, M.-M. Cheng, S.-M. Hu, Comput. Vis. Media., **9**(4), 733-752, (2022)
29. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, *Mobilenetv2: Inverted residuals and linear bottlenecks*, In Proceedings of the IEEE conference on computer vision and pattern recognition, 4510-4520, (2018)
30. B. Zoph, V. Vasudevan, J. Shlens, Q. V. Le, *Learning transferable architectures for scalable image recognition*, In Proceedings of the IEEE conference on computer vision and pattern recognition, 8697-8710, (2018)
31. O. Ulucan, D. Karakaya, M. Turkan, *A large-scale dataset for fish segmentation and classification*, In 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), 1-5, (2020)
32. M. Everingham, L. Van Gool, C. K. Williams, J. Winn, A. Zisserman, Int. J. Comput. Vis., **88**, 303-308, (2010)