Journal of Environmental Research and Engineering

JERE, 2026, 2 (1): 7-19

Fish Instance Segmentation and Closed-Polygon Boundary Visualization Using YOLO Model

Ahmer İliulu [1*], Cemal Turan[2], Yakup Kutlu[1]

[1] Department of Computer Engineering, Iskenderun Technical University, Hatay, Türkiye.

[2] Department of Molecular Biology and Genetics, Atlas University, Istanbul, Türkiye

ABSTRACT

Automated fish detection and identification are critical for aquaculture and marine ecology research. However, the complexity of underwater environments and the difficulty of pixel-level labeling limit the performance of computer vision models. This study aims to detect and visualize the outer boundaries of fish using closed polygon coordinates with the YOLO architecture. A standardized, clean dataset of 2700 images containing 13 scientific fish species was created. The model was trained and generated, and predictions for determining the boundaries of fish, converted to a closed polygon labeling format, were rendered as closed edge contours on the original images. According to the validation set results, the model demonstrated high performance in the mask prediction task, achieving an F1 score of 0.934 and mAP50-95 values of 0.854. Furthermore, all fish outer boundaries were successfully drawn on the image using the established automated visualization pipeline.

ARTICLE INFO

Received 2026-05-17,

Accepted 2026-06-29,

Publication Date 2026-06-30

Keywords:

Segmentation, YOLO, Computer Vision, Object Detection

*Corresponding Author: *Ahmet İlisulu*, *E-mail*: ahmetilisulu.lee25@iste.edu.tr

YOLO Model Kullanarak Balık Örneği Bölümleme ve Kapalı Poligon Sınır Görselleştirme

ÖZET

Otomatik balık bulma ve tanıma, su ürünleri yetiştiriciliği ve deniz ekolojisi araştırmaları için kritik bir öneme sahiptir. Ancak su altı ortamlarının karmaşıklığı ve piksel düzeyinde etiketlemenin zorluğu, bilgisayarlı görü modellerinin performansını sınırlandırmaktadır. Bu çalışmada, YOLO mimarisi kullanılarak balıkların dış sınırlarının kapalı çokgen koordinatlarıyla tespit edilmesi ve görselleştirilmesi hedeflenmiştir. Çalışma kapsamında 13 bilimsel balık türünü içeren 2700 görüntülük standartlaştırılmış temiz bir veri kümesi oluşturulmuştur. Model, kapalı çokgen etiketleme formatına dönüştürülen eğitilmiş ve üretilen balıkların sınırlarının belirlenmesi için tahminler, orijinal görüntüler üzerine kapalı kenar konturları olarak işlenmiştir. Doğrulama kümesi sonuçlarına göre model, maske tahmini görevinde 0,934 F1 skoru ve 0,854 mAP50-95 değerlerine ulaşarak yüksek bir başarım sergilemiştir. Ayrıca kurulan otomatik görselleştirme hattı ile tüm balık dış sınırları başarılı bir şekilde resim üzerine çizilmiştir.

MAKALE BİLGİSİ

Keywords:

Örneklem Segmentasyonu, Bilgisayarlı Görü, Nesne Tespiti

GİRİŞ

Automatic recognition of fish populations in marine and freshwater resources through imagery plays a critical role in modern aquaculture and marine ecology applications, such as developing precise feeding strategies, optimizing stock density, and estimating growth rates (Wang et al., 2025). Traditional object detection methods, frequently used in computer vision studies, only produce rectangular bounding boxes when classifying and locating objects (Alsuwaylimi, 2024). However, instance segmentation offers much more stable and comprehensive solutions for examining the morphological structures of fish and precisely distinguishing anatomical regions such as body, fins, and tail at the pixel level (Alsuwaylimi, 2024; Wang et al., 2025).

Training deep learning-based sample segmentation models involves a rather laborious and time-consuming labeling process, requiring the generation of pixel-level boundary coordinates (polygons) or masks for each object sample ("Automated Mask Generation," 2024). This further increases the risk of labeling errors in challenging underwater optical conditions such as light refraction, color distortion, water turbidity, and fish occlusion ("Automated Mask Generation," 2024; Alsuwaylimi, 2024). To meet real-time segmentation requirements and overcome these operational challenges, single-stage YOLO (You Only Look Once) architectures offering high throughput are widely preferred (Alsuwaylimi, 2024). The YOLOv11 series, a product of recent advancements in architecture, provides a balance of high accuracy and processing speed in underwater object and keypoint detection tasks with its optimized parameter structure and enhanced feature extraction capabilities (Wang et al., 2025).

Early studies on the automatic recognition of fish species relied heavily on morphometric measurements and classical machine learning classifiers. For example, Kutlu et al. (2017a) classified 78 morphometric distances extracted from 13 anatomical points (landmarks) in fish images using Deep Belief Networks (DBN) and distinguished three species belonging to the Triglidae family, which were quite similar in shape, color, and fin type, with 97.61% accuracy. Similarly, Kutlu et al. (2017b) achieved approximately 99% accuracy by developing a morphometry-based multi-stage classification system on a dataset of 1000 images covering 25 fish species and 13 families. However, these approaches are time-consuming as they require manual marking of anatomical points in each image and do not allow for pixel-level identification of the fish's outer boundary.

Advances in deep learning have made fish detection and segmentation learnable end-to-end, reducing reliance on manual feature extraction. However, underwater and aquaculture environments limit model performance due to challenges such as light refraction, color distortion, turbidity, low resolution, and especially fish occlusion (Zhang et al., 2026; Qin et al., 2026). Qin et al. (2026) highlighted that complex backgrounds, variable postures, and dense overlap make detection difficult in aquaculture; Zhang et al. (2026) developed a YOLOv11-based architecture (YOLOv11-SAS) for detecting dense schools of fish in sonar images and demonstrated its effectiveness in noisy underwater conditions. Sample segmentation offers richer information for morphological analyses as it identifies objects not only with rectangular boxes but also with pixel-level masks or polygons; however, this significantly increases the cost of labeling. Suchodolak et al. (2026) notes that manual mask production for sample segmentation can be more costly and time-consuming than model development itself, and that data collection and labeling can account for more than 80% of the total cost of a computer vision system. Single-stage segmentation models from the YOLO family are preferred in many fields, from medical imaging to autonomous vehicles, due to their balance of speed and accuracy (Suchodolak et al., 2026; Liu et al., 2026). YOLOv11 offers higher performance compared to previous versions (YOLOv8, YOLOv9, YOLOv10) by preserving fine details through architectural improvements such as C3k2 blocks and a position-aware self-attention (PSA) mechanism (Liu et al., 2026).

Sample segmentation offers richer information for morphological analysis because it identifies objects not only with rectangular boxes but also with pixel-level masks or polygons; however, this significantly increases the cost of labeling. Suchodolak et al. (2026) state that manual mask generation for sample segmentation can be more costly and time-consuming than model development itself, with data collection and labeling accounting for over 80% of the total cost of a computer vision system. Single-stage segmentation models from the YOLO family are preferred in many fields, from medical imaging to autonomous vehicles, due to their balance of speed and accuracy (Suchodolak et al., 2026; Liu et al., 2026). YOLOv11 offers higher performance compared to previous versions (YOLOv8, YOLOv9, YOLOv10) by preserving fine details through architectural improvements such as C3k2 blocks and a position-aware self-attention (PSA) mechanism (Liu et al., 2026).

In this study, in order to precisely isolate the outer boundaries of fish anatomy, the closed polygon coordinates determined during the labeling phase were directly converted to the YOLO segmentation format and incorporated into the training architecture. As part of the study, the raw dataset, which initially contained taxonomic labeling errors and irregularities, was standardized and a clean dataset (2700 images) was constructed under 13 scientific species (Apogon queketti, Apogon smithi, Argyrosomus regius, Caranx rhoncus, Chaetodon kleinii, Chanos chanos, Eleutheronema tetradactylum, Johnius trachycephalus, Rastrelliger faughni, Sarda sarda, Sciaena umbra, Umbrina cirrosa, and Upeneus moluccensis). Training was then conducted on this prepared dataset using the YOLOv11 sample segmentation architecture for 100 epochs. The numerical success of the trained model was demonstrated using validation metrics; Furthermore, the estimated polygon coordinates were plotted as closed contour lines on the original images, establishing a qualitative boundary visualization infrastructure.

MATERIALS AND METHODS

Dataset and Labeling

The dataset used in this study was reconstructed under 13 scientific fish species (Apogon queketti, Apogon smithi, Argyrosomus regius, Caranx rhoncus, Chaetodon kleinii, Chanos chanos, Eleutheronema tetradactylum, Johnius trachycephalus, Rastrelliger faughni, Sarda sarda, Sciaena umbra, Umbrina cirrosa, and Upeneus moluccensis) by standardizing raw data containing taxonomic labeling errors and irregularities, and cleaned to consist of a total of 2700 images; the dataset was divided into 2151 training and 549 validation images. In previous studies on morphometry-based recognition of fish species, the need to manually label anatomical points per image constituted a significant bottleneck (Kutlu et al., 2017a; Kutlu et al., 2017b; Kutlu and Turan, 2018). To overcome this bottleneck, the outer boundary of each fish sample was labeled with closed polygon coordinates, and these labels were directly converted to the YOLO segmentation format (normalized polygon coordinates). Polygon-based labeling represents the anatomical boundaries of the fish more accurately at the pixel level compared to rectangular bounding boxes (Suchodolak et al., 2026). In the mask/polygon transformation, the OpenCV findContours function can be used to extract the outer contours from binary masks, and the resulting coordinates are divided by the image dimensions and normalized to the range [0, 1] (Suchodolak et al., 2026).

The 13 fish species used for segmentation and classification in the dataset and their corresponding YOLO class indices are presented in Table 1.

Table 1. Fish species and YOLO class indices used in the dataset.

#	**Fish spacies**
1	Apogon queketti
2	Apogon smithi
3	Argyrosomus regius
4	Caranx rhoncus
5	Chaetodon kleinii
6	Chanos chanos
7	Eleutheronema tetradactylum
8	Johnius trachycephalus
9	Rastrelliger faughni
10	Sarda sarda
11	Sciaena umbra
12	Umbrina cirrosa
13	Upeneus moluccensis

Hardware and Software Environment

The training, hyperparameter optimization, and inference processes of the developed segmentation model were carried out on the NVIDIA graphics processing unit architecture, which provides hardware parallel computing support, and on the CUDA platform. In the software layer, PyTorch libraries were used as the deep learning backbone, and the Ultralytics YOLO framework was integrated for the model architecture and training pipeline. YOLO is a powerful architecture with a wide range of model applications (İlisulu, and Kutlu, 2025; Chowdhury et al. 2026). The processing of the obtained polygon predictions and edge lines onto image matrices was performed using the OpenCV library for computer vision operations.Model Training

A transfer learning strategy was applied for the sample segmentation task, based on pre-trained segmentation weights. Model training was performed on a dataset of 2700 images, 2151 of which were for training and 549 for validation, containing 13 validated scientific fish species (Apogon queketti, Apogon smithi, Argyrosomus regius, Caranx rhoncus, Chaetodon kleinii, Chanos chanos, Eleutheronema tetradactylum, Johnius trachycephalus, Rastrelliger faughni, Sarda sarda, Sciaena umbra, Umbrina cirrosa, and Upeneus moluccensis). The training process was structured with 100 epochs, 8 batch sizes, and $640 \times $640 pixel input resolution parameters via the developed training pipeline. Automatic Mixed Precision (AMP) was kept active to optimize memory and increase computational speed. The model's performance; Precision (P), Recall (R), F1 score, mAP50 and mAP50-95 metrics were tracked and evaluated separately at both the bounding box (Box) and pixel-based mask (Mask) levels. For bounding box regression and classification, the YOLOv11's predefined composite loss function (binary cross-entropy for classification; Distribution-Oriented Loss with CIoU for box regression) was used (Liu et al., 2026). In the literature, it has been shown that advanced IoU-based loss variants such as GIoU, DIoU, CIoU and EIoU improve localization accuracy to reduce positioning errors caused by variable orientation and overlap (Qin et al., 2026).

Evaluation Metrics

The model's performance was evaluated using the metrics Precision, Recall, F1 score, and mean precision (mAP), which are commonly used in the object detection and segmentation literature (Suchodolak et al., 2026; Qin et al., 2026). Precision measures how many of the samples that the model positively predicted are actually correct; Sensitivity measures how many real objects were captured. The F1 score is the harmonic mean of these two metrics and penalizes unbalanced Precision–Sensitivity values. mAP50 is the mean precision calculated when the Intersection Over Union (IoU) threshold is taken as 0.50; mAP50-95 represents the average of thresholds taken in 0.05 increments from 0.50 to 0.95 and is a stricter measure (Zhang et al., 2026; Suchodolak et al., 2026). Segmentation quality is also quantified by IoU (Jaccard index), which measures the overlap between the predicted mask and the reference mask.

Segmentation

Following the completion of training, the highest-performing weights obtained were applied collectively to all images in the dataset through an inference architecture. A confidence threshold of 0.25 was set to filter out background noise and weak predictions during the inference phase; and a Non-Largest Suppression (NMS) IoU threshold of 0.70 was used to eliminate overlapping boxes. Polygon coordinates generated based on the presence of multiple objects in the images were normalized at the pixel level and exported in YOLO format and COCO JSON, a standard data exchange format.

Boundary Visualization

An automated visualization pipeline was established to visually validate the segmentation success of the model and produce contour outputs suitable for biological analyses. Polygon points obtained from the inference phase were sequentially linearly connected and overlaid on the original images in the form of a closed polygon. To maintain the clarity of the boundary lines in the visual outputs, the line thickness was set to 2 pixels, the inner fill transparency (alpha) was set to 30/255 to slightly highlight the fish body, and the scientific name of the class was printed on the coordinate of the geometric center mass of the relevant polygon. Unicode file path security was ensured by using in-memory buffering techniques in data reading and writing processes to prevent file access errors that may arise from operating system-level character incompatibilities.

RESULS AND CONCLUSIONS

Quantitative findings from the 100-epoch training performed with the developed model and the subsequent inference processes conducted on the entire dataset are presented in this section. Model Performance Metrics

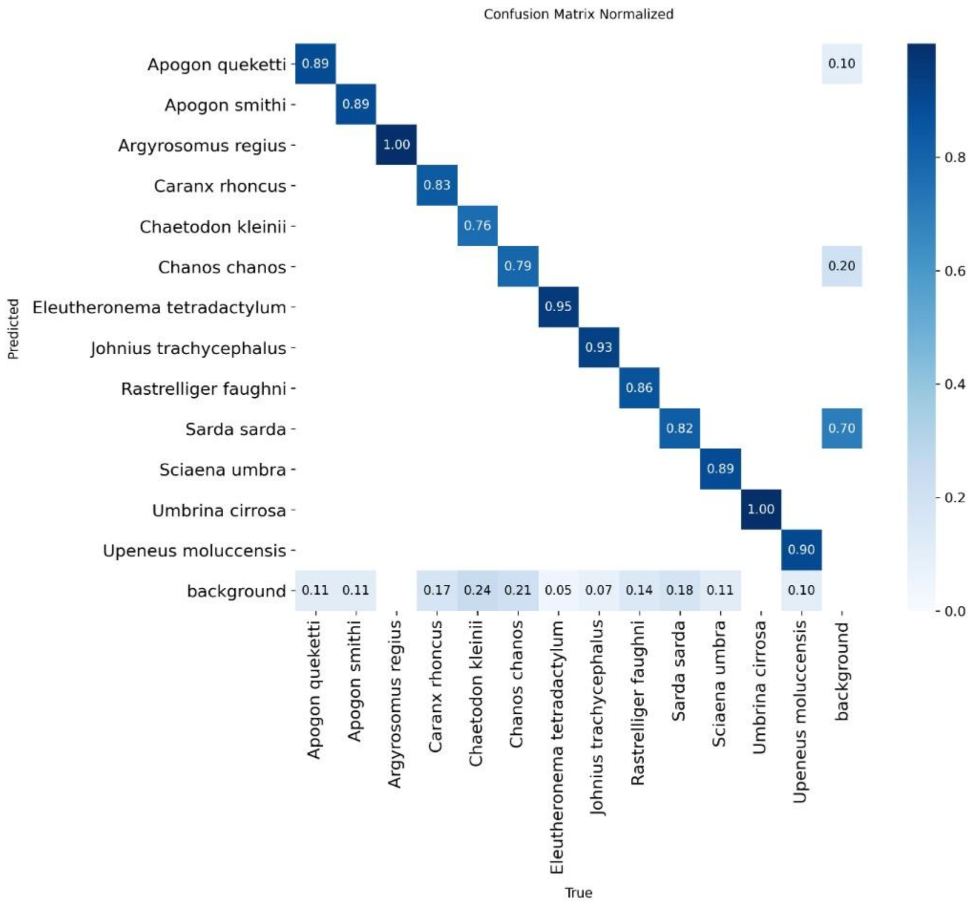
The object detection (Box) and pixel-level segmentation (Mask) success metrics of the weights obtained as a result of training on the validation dataset are presented in Table 2.

Tablo 2. Model başarı metrikleri

Metric	Box	Mask/Polygon
Precision	0,989	0,989
Recall	0,884	0,884
F1 Skoru	0,934	0,934
mAP50	0,895	0,888
mAP50-95	0,882	0,854

The species-based distribution of classification performance obtained on the validation set was evaluated using a normalized confusion matrix (Figure-2). Diagonal values indicate the classification accuracy rate for each species. 100% accuracy was achieved in Argyrosomus regius and Umbrina cirrosa; the lowest performance was observed in Chaetodon kleinii (76%), Chanos chanos (79%), and Sarda sarda (82%). Direct interspecies confusion remained limited; the majority of errors resulted from confusing fish samples with the background class. In particular, 70% of background samples were mistakenly classified as Sarda sarda.

Figure 1. Segmentation performance.



3.2. Inference and Edge Drawing Statistics

The segmentation inference performance of the model architecture on the entire data set and the number of successfully completed closed polygon edge line drawings are summarized in Table 3.

Qualitative segmentation results on selected examples are presented in Visual-3. In each row, the original image, the hand-drawn ground truth closed polygon boundary (green) and the model prediction boundary (blue) are compared, respectively. Model predictions show high overlap with ground truth limits in different morphology, color and background conditions; It can be seen that fine anatomical details, especially the fins and tail, are successfully monitored.

Table 3. Data Set Overall Inference and Boundary Plotting Statistics

Data Set Section	**Image Count**	**Number of Polygons**	**Estimates Edge Boundary**	**Skipped Images**
Train	2151	2231	2151	0
Validation	549	557	549	0
Toplam	2700	2788	2700	0

Figure 2. Example Result image.

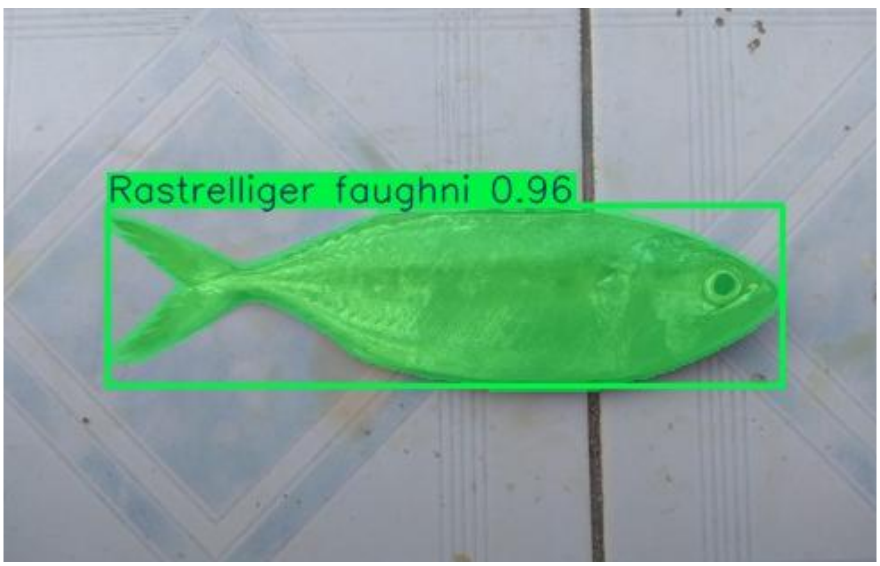
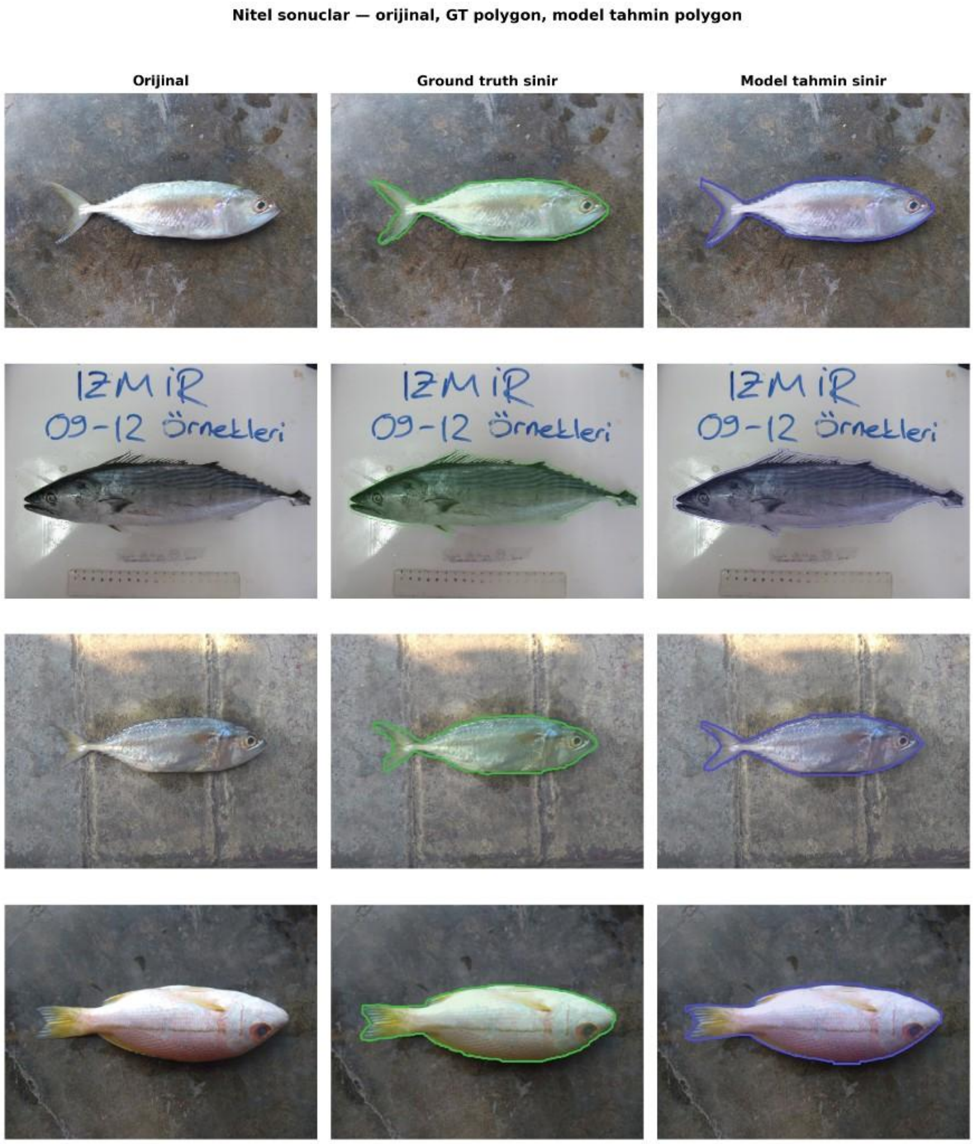


Figure 3. Overall results images.



TARTIŞMA

This study utilized the YOLO model. The segmentation architecture achieved an F1 score of 0.934 and a mAP50-95 value of 0.854 on a mask estimation task on a cleaned dataset containing 13 fish species (Apogon queketti, Apogon smithi, Argyrosomus regius, Caranx rhoncus, Chaetodon kleinii, Chanos chanos, Eleutheronema tetradactylum, Johnius trachycephalus, Rastrelliger faughni, Sarda sarda, Sciaena umbra, Umbrina cirrosa, and Upeneus moluccensis). This performance demonstrates that directly converting closed polygon coordinates to the YOLO segmentation format is effective in accurately isolating fish outer boundaries; qualitative comparisons in Figure 3 also visually confirm that the model predictions are consistent with ground truth limits.

The findings are consistent with other studies based on YOLOv11. Liu et al. (2026) reported that YOLOv11, thanks to its C3k2 block and PSA attention mechanism, provides approximately 1–2% improvement in mIoU and Dice scores compared to YOLOv8, YOLOv9, and YOLOv10 by better preserving fine boundary details; this supports the preference for YOLOv11 in the present study. Zhang et al. (2026) showed that the YOLOv11-based architecture provides a significant performance increase (up to 27.2% improvement in mAP50 in dense schools) for dense schools of fish in noisy sonar images compared to the base model. In the context of aquaculture, Qin et al. (2026) revealed that YOLO-based models are superior to transformer and anchor-free models in terms of accuracy-efficiency balance; this result justifies the selection of single-stage YOLOv11 in our study.

On the other hand, the Sensitivity value of the model (0.884) is lower compared to the Accuracy value (0.989); the confusion matrix in Figure 2 also supports this. Missed detections of 18–24% for the background class were observed in Chaetodon kleinii, Chanos chanos, and Sarda sarda species. Factors such as occlusion, variable orientations, and complex backgrounds in underwater images can be considered as the main reasons for this low sensitivity. Similarly, the literature reports that overlap and occlusion lead to missed detections and that this reduces Sensitivity (Qin et al., 2026; Zhang et al., 2026). Qin et al. (2026) showed that enhanced IoU variants such as EIoU improve localization accuracy to reduce positional errors caused by variable orientations; such attenuation functions offer a promising direction for improving Sensitivity in the future.

Another contribution of the study is that it offers a practical solution to the labeling bottleneck. Manual mask generation for sample segmentation is quite costly (Suchodolak et al., 2026); the direct conversion of closed polygon coordinates to YOLO format and the established automated visualization pipeline contribute to reducing this cost. Suchodolak et al. (2026) showed that a YOLOv11-seg based semi-automatic labeling pipeline could reduce the need for manual labeling by up to 20 times and achieve competitive performance (mAP50 > 0.80) even with only 1% labeled data; this finding suggests that a similar semi-automatic loop could be used in expanding our dataset in the future.

Compared to previous studies based on morphometry and classical classifiers (Kutlu et al., 2017a; Kutlu et al., 2017b), the proposed approach obtains the entire outer boundary of the fish at the pixel level without the need for manual marking of anatomical points. This provides a rich visual infrastructure for biological analyses (e.g., length/mass estimation, population and stock assessment, migration tracking) beyond species identification (Zhang et al., 2026). Future studies aim to expand the dataset with more species and real-world field images, integrate semi-automatic tagging cycles, and run the model in real-time on embedded/edge devices (Qin et al., 2026; Zhang et al., 2026).

SONUÇ

In conclusion, the YOLOv11 sample segmentation architecture, trained with closed polygon-based labeling, stands out as an effective and reliable tool for the highly accurate detection and visualization of fish outer boundaries in challenging underwater images encompassing 13 different fish species. With F1=0.934 and mAP50-95=0.854, the model produced successful results at both quantitative and qualitative levels; the established automated edge boundary visualization pipeline successfully projected all fish outer boundaries onto the images, providing outputs suitable for biological analysis.

KAYNAKÇA

Alsuwaylimi, A. A. (2024). Enhanced YOLOv8-Seg Instance Segmentation for Real-Time Submerged Debris Detection. IEEE Access, 12, 114825-114835.

Automated Mask Generation for Efficient YOLO-Based Instance Segmentation in Marine Environments for Fish Detection. (2024). Marine Vision and Technology Journal, 8(2), 142-155.

Chowdhury, S., Kutlu, Y., Goongoon, N. (2026). Comparative Evaluation of YOLO-Based Pretrained Models for Multi-Class Vegetable Detection. Tethys Environmental Science, 3(1), 16-30, doi : 10.71350/tethysjournal.2

İlisulu, A., Kutlu, Y. (2025). YOLO-Based Deep Learning Approach for Turkish Sign Language Recognition. Journal of Natural and Engineering Research, 1(1), 38-44, doi : 10.5281/zenodo.18664511

Kutlu, Y., Altan, G., İşçimen, B., Doğdu, S. A., & Turan, C. (2017a). Recognition of species of Triglidae Family using Deep Learning. Journal of the Black Sea/Mediterranean Environment, 23(1), 56-65.

Kutlu, Y., İşçimen, B., & Turan, C. (2017b). Multi-Stage Fish Classification System Using Morphometry. Fresenius Environmental Bulletin, 26(3), 1910-1916.

Kutlu, Y., & Turan, C. (2018). An Intelligent Software for Measurements of Biological Materials: BioMorph. Natural and Engineering Sciences, 3(2), 225-233. https://doi.org/10.28978/nesciences.424679

Liu, Y., Fan, H., Wang, S., Huang, Y., Luo, N., Ren, H., & Xu, H. (2026). Tongue Image Segmentation Using YOLOv11-based Object Detection and Mobile Segment Anything Model. In 2026 6th International Conference on Neural Networks, Information and Communication Engineering (NNICE) (pp. 411-417). IEEE.

Qin, H., Mei, Y., Yu, H., Xu, L., Chai, Y., He, Q., & Chen, Y. (2026). Improved YOLO-InceptionV3 network for shrimp detection in recirculation aquaculture system. Information Processing in Agriculture (Baskıda).

Suchodolak, M. A. R., Fraidenraich, G., de Lima, E. R., & de Oliveira, J. R. (2026). A Semi-Automatic Labeling Approach for Image Segmentation Using YOLOv11. IEEE Access, 14, 49908-49927.

Wang, J., Cheng, Z., Lin, M., Yang, R., & Huang, Q. (2025). FishKP-YOLOv11: An Automatic Estimation Model for Fish Size and Mass in Complex Underwater Environments. Animals, 15(19), 2862.

Zhang, H., Yang, X., Zhang, R., Gao, N., Wang, N., Zhang, Z., & Zhang, C. (2026). Detection of dense fish schools in sonar imagery with a novel YOLOv11-SAS model. Ecological Informatics, 94, 103646.