



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** X **Month of publication:** October 2026

DOI: <https://doi.org/10.22214/ijraset.2026.85116>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Analysis of Deep Learning Models for Underwater Species Detection and Classification

Kamali B¹, Kishore A², Harshavardhan T³, Sakthivel Murugan S⁴

^{1,2,3} Dept. of Electronics and Communication Engineering, Sri Venkateswara College of Technology, Sriperumbudur

⁴ Dept. of Electronics and Communication Engineering, National Institute of Technical Teachers Training and Research, Chennai, India

Abstract: Underwater fish detection is challenging. Fish can be very small in images. The contrast in scenes is often weak. Turbid water can hide details. When fish overlap it becomes harder to locate each one. This study compares deeplearning object detectors using an underwater fish dataset. The focus is on both detection quality and how well the models work in video processing. The models tested are YOLO26, RF-DETR and SSD300 with a backbone. In the evaluations YOLO26s achieved a mAP@50 of 0.875 and a mAP@50:95 of 0.499.

RF-DETR Base reached 0.4465 and 0.3085 respectively. The SSD300 model showed a precision of 0.8611, recall of 0.1975, mAP@50 of 0.1436 and mAP@50:95 of 0.0607. These results show that high precision alone does not mean detection performance if recall is low. The comparison also looks at class-label consistency and differences, in settings. The model runs were not all done under the conditions. Still the results give a view of what each detector can do. They highlight the strengths and weaknesses of YOLO26, RF-DETR and SSD300 for detecting fish in underwater.

Keywords: Underwater Fish Detection, Deep Learning, Object Detection, YOLO, YOLO26, RF-DETR, SSD, SSD300, , Fish Classification, Fish Counting, mAP, Precision, Recall, Underwater Imaging.

I. INTRODUCTION

Trustworthy fish detection, in images and video can help ecological surveys, aquaculture monitoring and automated fish classification. A detector must find every fish. Give the right class while the underwater world adds extra visual challenges. Small fish take up a few pixels, scattering and colour distortion lower contrast and when fish are partly hidden it makes fish species recognition hard. These challenges mean that the best fish detector depends not on its design but also on the dataset and on the purpose it is meant for.

These records examined in this project serve two goals: correctly separating fish classes and running video at a useful speed. YOLO-style detectors. Classify objects in one step. DETR-family models rely on representations, for detecting objects. SSD is another single-stage detector that predicts from several feature scales. Comparing these detectors gives a wider view of how traditional and newer detectors work on the same underwater problem.

Earlier underwater studies looked at enhancement, attention, multi-scale feature extraction and class-aware learning to handle visibility, small targets and class imbalance. This current work does not say that a single architecture is always the best. Instead it reviews the experimental records and checks where each model does well or faces difficulties.

II. LITERATURE SURVEY

Underwater computer vision research has looked into learning methods for finding and sorting fish in images and videos. Underwater pictures often have contrast, cloudy water, fish that overlap, messy backgrounds and species that look alike. These issues can make it harder to find and name the fish correctly.

YOLO-based tools are often used when quick results are needed. In the project records YOLO-based models show performance for checking each video frame one by one. RF-DETR uses a way that is like a transformer and is helpful for looking at how good the location is compared to how fast the system works.

SSD, which stands for Single Shot MultiBox Detector is another type of object finder that works in one step. SSD gives the type of object. Where it is by looking at different parts of the image. Using sizes helps with finding objects in general but very small or unclear objects can still be hard to spot. In this study SSD300 with a setup is added as an example so that it can be compared with YOLO26 and RFDETR for finding fish underwater.

Other tools like YOLOv8-TF, HRA-YOLO, YOLO-JUICE. These tools try to solve problems like having few examples of some fish not enough light and finding tiny fish. They offer ideas for future work but their results are not added to the local testing numbers because they were tested on different pictures and, with different rules.

III. PROPOSED SYSTEM

This system. Identifies fish in underwater pictures and videos with several object-detection models. The video is broken down into frames. Each frame goes through a trained detector that gives the box coordinates the guessed class and the confidence score for every fish it finds.

YOLO26 serves as the real-time detector. Its backbone pulls out features its neck merges data from multiple scales and its detection head produces the final predictions. RF-DETR is also used as a transformer-based comparison model. SSD300- is added as a standard single-stage baseline.

Once detection is done the models are checked by precision, recall, F1-score, mAP@50 and mAP@50:95 when those metrics can be computed. The time taken and frames, per second are also listed. MAP@50 shows how well the model detects objects when the overlap threshold is 0.50. MAP@50:95 averages results over thresholds giving a tougher test of how accurately the model localises fish.

After detection the fish can be counted. If you count fish frame by frame the same fish may be counted times so counting unique fish needs tracking across consecutive frames. This comparison concentrates on detection; counting by tracking is a separate task that needs its own verification.

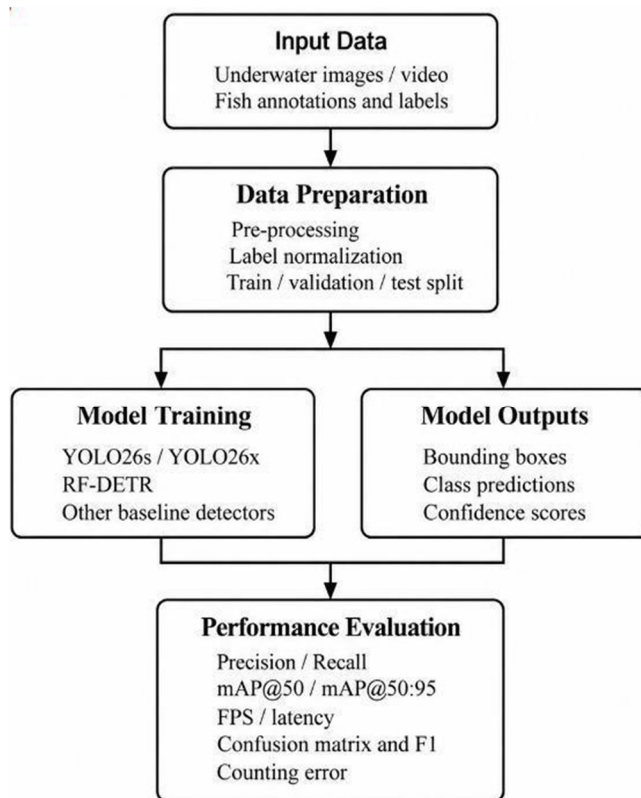


Fig. 1. This workflow is used for fish detection, model evaluation and subsequent counting.

IV. METHODOLOGY

A. Underwater Video and Frame Processing

The input is a video that shows many fish. The video is split into frames so that each frame can be examined by detection models. Underwater frames often have contrast, cloudiness, fish that overlap and very small fish. These underwater conditions are taken into account when reading the model results because they can cause missed detections and wrong class labels.

B. YOLO26-Based Fish Detection

YOLO26 gets an image and figures out where the fish are by drawing bounding boxes around them and identifying their types. The model has three parts: a backbone, a neck and a detection head. Handling features at scales is key because fish can look very different in size even when they are in the same picture. In the experiment, with YOLO26 the model was trained for 100 epochs using images that were 1280 pixels wide. The validation results showed a precision of 0.845, a recall of 0.839, an mAP@50 of 0.875 and an mAP@50:95 of 0.499.

C. RF-DETR-Based Fish Detection

RF-DETR is included as a transformer-based model for comparison. The RF-DETR Base experiment was trained over 100 epochs using CPU execution and an input size of 560 pixels. The precision recorded was 0.3437. The recall was 0.6254. The mAP@50 value was 0.4465. The mAP@50:95 score was 0.3085. Its processing speed was reported to be around 1.63 frames per second. These results are treated as an experimental record because the hardware setup and input configuration were different, from the YOLO26 run.

D. SSD300-Based Fish Detection

SSD300 with a backbone was tested as a single-stage detector. The model was run using CUDA execution and the best checkpoint available was saved at epoch 94. The validation set had 93 images for testing. The model gave a precision of 0.8611 and a recall of 0.1975. Its mAP@50 was 0.1436 and mAP@50:95 was 0.0607. The model found 31 positives, 5 false positives and 126 false negatives in the test. The big difference, between precision and recall is important: even though most of the predictions were correct many real fish cases were not found.

The class-wise mAP@50 results also show levels of performance. Black Surgeonfish had an AP of 43.1% while butterflyfish, damselfish_chromis and surgeonfish_tang had 0% in the class- test. These results show that the SSD model had trouble detecting all fish types.

V. RESULTS AND DISCUSSION

A. Model Comparison

The validation records clearly show differences between the detectors that were evaluated. YOLO26s achieved the highest mAP@50 compared to the two local runs. RFDETR achieved a mAP@50:95 than SSD. SSD300 recorded the precision, among the runs yet its recall and both mAP scores were much lower. This demonstrates why precision alone is not enough to judge an object detector.

Model	Precision	Recall	mAP@50	mAP@50:95	FPS
YOLO-26	0.845	0.839	0.875	0.499	17.7
RF-DETR	0.3437	0.6254	0.4465	0.3085	1.63
SSD300	0.8611	0.1975	0.1436	0.0607	25.83

Table 1. Comparison of the validation records that are available, for the detectors that were examined.

The table should be read while keeping the settings in mind. The runs were not carried out with input resolutions, hardware or training configurations. The values show how the recorded models performed in their experiments rather than proving a controlled architectural ranking.

B. SSD300 Validation Results

The SSD300 validation results show a precision of 86.11%. Recall was only 19.75%. The resulting F1-score was 32.12%. At the IoU threshold of 0.50 the model achieved 14.36% mAP, which fell to 6.07% when averaged over IoU thresholds from 0.50 to 0.95. The reduction at IoU thresholds indicates that the localization quality was also limited.

The SSD confusion matrix gives a detailed view of these results. Black Surgeonfish has 31 detections while 2 instances are predicted as surgeonfish_tang and 38 are missed as background. The butterflyfish row contains 18 missed instances. Surgeonfish_tang contains 68 correct detections, in the available matrix. The background entries also show that some fish predictions were confused with background. Overall the matrix supports the recall value reported for SSD and indicates that missed detections were a major source of error.

Looking at the class-wise AP@0.50 plot we see that Black Surgeonfish stands out as the class that achieved a measurable AP. That AP for Black Surgeonfish is 43.1 %. Butterflyfish, damsselfish_chromis and surgeonfish_tang each recorded zero AP, in the class-wise evaluation. The uneven performance indicates that the SSD model did not generalize equally across the fish categories represented in the evaluation set.

The SSD precision-recall curve shows a sharp change in precision over a relatively small recall range. This behaviour is consistent with the reported recall of 0.1975. The mAP-versus-IoU plot also shows a steady decline as the IoU threshold becomes stricter: mAP@0.50 is 14.36%, falling to approximately 6.07% at the 0.50:0.95 average. The result suggests that both missed detections and imperfect localization contribute to the weaker overall performance.

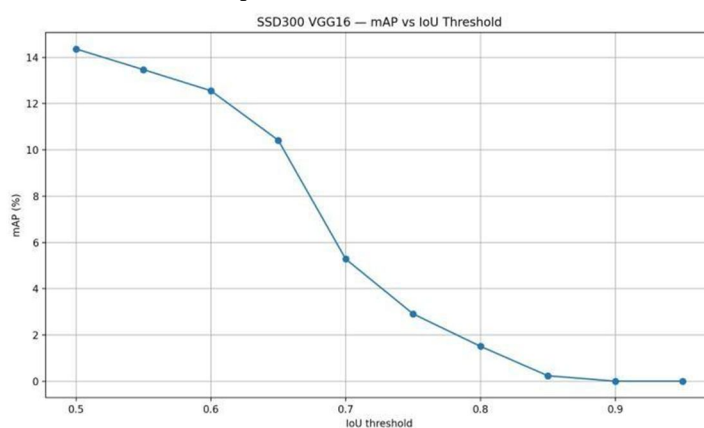


Fig. 5. SSD300 mAP variation with IoU threshold.

C. Representative SSD Detection Output

A sample prediction image is set below for the final SSD output. The image must display the model's bounding boxes, class labels and confidence values overlaid on a frame. It should be placed after the final prediction image has been chosen, ensuring the qualitative example matches the evaluation setup.



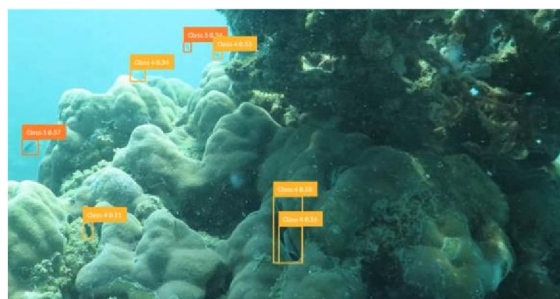
Fig. 6. Representative SSD300 detection output on an underwater fish image.

D. YOLO26 and RF-DETR outputs.

The confusion matrices from the runs that're available also show that performance is not the same, for all fish classes. In particular the project records contain duplicated labels because of different class strings. These Damsselfish labels should be normalized before any final species-level comparison is treated as definitive.



YOLO26s



RF-DETR

Fig. 7. Representative YOLO26s and RF-DETR output images.

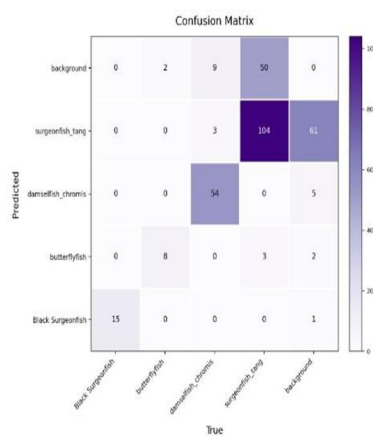
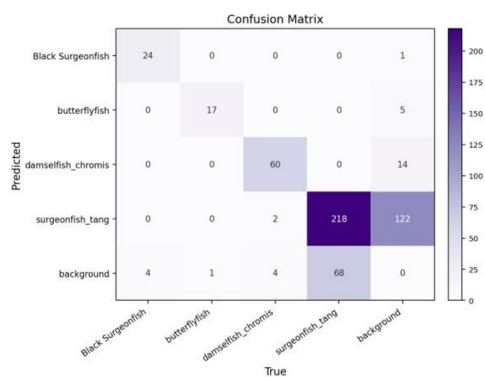


Fig. 8. YOLO26/FishModel1 confusion matrix.

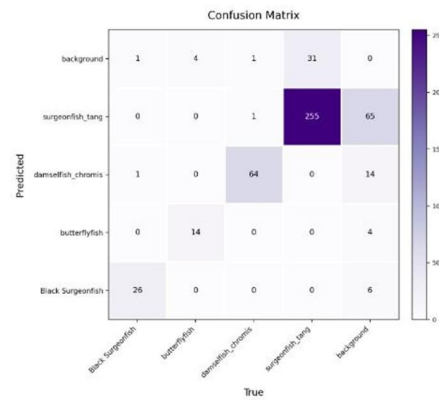


Fig. 9. FishModel2 confusion matrix.

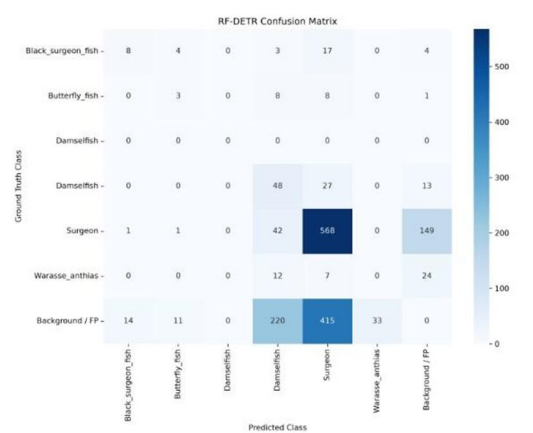


Fig. 10. RF-DETR class-wise confusion matrix.

VI. DISCUSSION

The expanded comparison gives a picture of the trade-offs between the three detector families. YOLO26s is the speed-oriented baseline in the available records and also has the highest mAP@50 among the local runs. RF-DETR provides a result at the stricter mAP@50:95 measure than SSD although its reported frame rate is much lower. SSD300 has a high precision of 0.8611 but the recall of 0.1975 shows that the model misses many fish instances. For a monitoring application where missed fish directly affect detection and counting this low recall is an important limitation.

The SSD confusion matrix supports this interpretation through the large number of fish instances assigned to background. Its class-wise AP results also show that the model performs unevenly across categories. These observations are useful because they show why a single metric can give a picture. A model can produce detections when it makes a prediction while still failing to find a large portion of the fish present in an image.

FishModel1 and FishModel2 remain YOLO26 training records rather than additional controlled architectures. FishModel2 reports mAP@50 of 0.906 and mAP@50:95 of 0.592 compared with 0.860 and 0.494 for FishModel1. However differences in training configuration and runtime environment mean that these values should not be interpreted as a controlled model-, to-model improvement.

VII. LIMITATIONS

The comparison combines records made under different settings. Image size, training duration, hardware, dataset splits and model configurations are not the same on all runs. Also the SSD evaluation shown here does not contain a speed measurement for inference so no SSD FPS number is given. The class labels in the project data include entries because of inconsistent naming. These problems weaken the ability of the comparison to give a numeric ranking

A better future comparison should standardize all class names use one fixed train/validation/est split test every model , on the held-out images use the same input resolution when possible repeat the experiment and record inference time on the same hardware. For fish counting the detector should also connect to a tracking method. Use track-level counting metrics.

VIII. CONCLUSION AND FUTURE WORK

This study looked at YOLO26, RF-DETR and SSD300 using the underwater fish detection data. Among the models tested YOLO26 showed the mAP@50 and also had the fastest processing speed. RF-DETR performed better than SSD in terms of mAP@50:95, which means it did a job when the localization requirements were more strict. SSD300 had a precision value of 0.8611 but its recall was very low at 0.1975. This led to a mAP@50 of 0.1436 and an even lower mAP@50:95 of 0.0607. The confusion matrix and class-wise results show that SSD300 missed detections and had uneven performance across different classes.

The findings suggest that YOLO26 is currently the practical choice for real-time underwater video analysis. RF-DETR gives a comparison focused on accuracy. SSD300 serves as a baseline and helps highlight the difference between precision and recall in this specific task. These conclusions are limited to the data and conditions used in this study. A final decision about which model to deploy should only be made after rerunning all models under the evaluation protocol and testing the full detection, to-tracking counting pipeline.

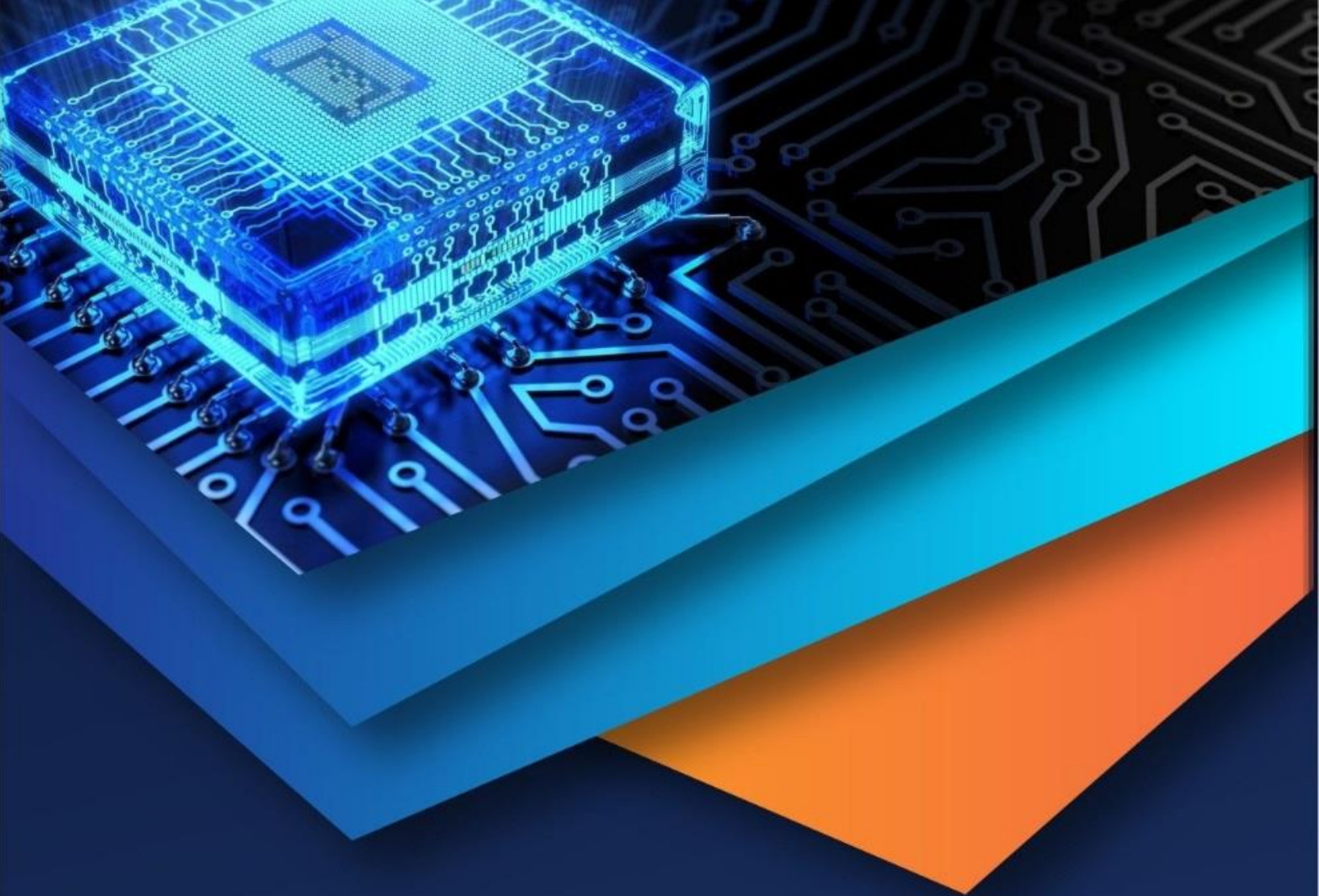
Future work should aim to improve detection of fish and fish that are partially hidden in low-visibility underwater settings. A bigger varied dataset could also help classify fish that look very similar. Future steps might add real-time tracking and counting of fish from videos and also test lightweight models to speed up processing on devices, with limited resources.

IX. ACKNOWLEDGMENT

This work was carried out as part of an internship at the Aazhi Research Laboratory, Department of Electronics and Communication Engineering, National Institute of Technical Teachers Training and Research(NITTTR),Chennai.

REFERENCES

- [1] C. Shah et al., "YOLOv8-TF: Transformer-Enhanced YOLOv8 for Underwater Fish Species Recognition with Class Imbalance Handling," *Sensors*, vol. 25, no. 6, 1846, 2025. <https://doi.org/10.3390/s25061846>
- [2] I. Robinson, P. Robicheaux, M. Popov, D. Ramanan, and N. Peri, "RF-DETR: Neural Architecture Search for Real-Time Detection Transformers," in *Proc. ICLR*, 2026. <https://arxiv.org/abs/2511.09554>
- [3] W. Liu et al., "SSD: Single Shot MultiBox Detector," in *Proc. ECCV*, 2016. <https://arxiv.org/abs/1512.02325>
- [4] Research paper comparison slides, "Underwater Fish Detection Model Comparison," version including RF-DETR and later YOLO-family model discussion, local research paper file, 2026.
- [5] Local training records for YOLO26s, RF-DETR Base, and the custom YOLOv10 small-dataset run, including results.csv, args.yaml, metrics.csv, and training_config.json, 2026.
- [6] YOLO26s 1280-pixel run output directory, including results.png, confusion matrices, and validation mosaics, local research paper files, 2026.
- [7] G. Jocher, J. Qiu, M. Liu, S. Lyu, F. C. Akyon, and M. E. Kalfaoglu, "Ultralytics YOLO26: Unified Real-Time End-to-End Vision Models," *arXiv:2606.03748*, 2026. <https://arxiv.org/abs/2606.03748>
- [8] Ultralytics, "YOLO26: Real-Time Vision Models," *Ultralytics Documentation*, 2026. <https://docs.ultralytics.com/models/yolo26>



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)