Automated fish detection and tracking system using pre-trained Mask R-CNN for Ecological biodiversity

Suja Cherukullapurath Mana (cmsuja@gmail.com)
Sathyabama Institute of Science and Technology
T. Sasipraba
Sathyabama Institute of Science and Technology

Research Article

Keywords: Pre-trained Mask-R-CNN, fish classifying, continuous automatic tracking system, computer-aided-automatic system and underwater surveillance

Posted Date: March 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1395108/v1

Abstract

In this paper, propose a new dynamic classifying algorithm which supports fully automatic fish detection and tracking system to identify fish species and to track fish activities for understanding synapomorphies characteristic simultaneously. The pre-trained Mask Regional Convolutional Neural Network (Mask-R-CNN) is involved for having well-enhanced feature vectors derived after undergone with number of test samples taken from captured video footage. Hence, the proposed framework is noticed as pre-trained Mask-R-CNN. It improves the system function of automatic fish detection and tracking to enhance the underwater surveillance for monitoring ecological biodiversity. The available ground–truth dataset is used to evaluate the system precisian, F1-score, and recall in terms of classifying and tracking mechanism. The comparative analysis is made with existing tracking R-CNN algorithms such as minimum output sum of squared errors (MOSSE), sequential non-maximum suppression (Seq-NMS) and Siamese mask (SiamMask). Simulation results conveys that the proposed algorithm support effective fish detection (i.e. around 120 out of 170 individual bream) and accuracy of the pre-trained Mask-R-CNN (87%) as compared with MOSSE (75%), Seq-NMS (78%), and SiamMask (84%) respectively. Thus, the evaluation result shows, the proposed pre-trained Mask-R-CNN achieves reasonable improvement in accuracy (detection and tracking) which give potential benefits to ocean ecosystem for establishing the ecological management.

I. Introduction

An Indian marine fishery is significantly increases the average production growth upto 92% including freshwater and inland fisheries from last two decades. Hence, it shows total annual fish market rate is reached around 10.88% as compared with subsequent financial years (i.e. 2014-15 to 2018-19 (as per National Fisheries Development Board)). According to NFDB, nearly 17 million metric tons have estimated during 2019-20 in the total fish production. In addition to this, 6.3% of the global fish production constituted in the same period. A marine fishery in India extends the Exclusive Economic Zone (EEZ) to 2.02 million square kilometers that support people around coastline for their food and employment [1]. However, the climatic variation disturbs the natural habitat of fishes that causes migration of species, lack of fish's density and absence of expected fishes at specific target etc. Hence, it degrades capturing ratio of marine fishery which may fail to sustain production growth continuously. Thereby, inland fisheries take higher revenue and contribute more percent to total fish production [2]. It is achieved by inventing habitat of aquaculture production that means, allocating areas for creating streams, ponds and rivers. It supports native fish species are easier to breed for reaching desired size and they are more resistant to local climate conditions [3]. In such aquaculture fish production, follows daily monitoring on water purity, salty content and cultivating a new species etc. It is a prescribe pattern need to adapt for cultivating native fish breeds in the controlled environment for reaching maximum profits [4]. In recent times, the fisheries research community decided to develop a new methodology for improving marine fisheries by using high definition sensors, cameras mounted on underwater vehicles which are operated under influence of artificial intelligence system in order to track the fish movement and classify

the species. Thus, increases large-scale commercial fish production and retain the efficiency of marine fishery industry. In [5], describes remote sensing technique for continuous tracking of fish colonies in the various ocean depth regions. Consequentially, detect the specific fish species, classify and count the fish in the underwater [6]. Although this method is excellent in recognizing large size fish, it is ineffective in detecting tiny fish. That means, segmentation is done based on edge thresholding which is capable for capturing underwater species like sharks, however, it fails to accurate in the murky water [7]. For freshwater fishes, multi-level perceptron (MLP) detection was used on sonar images with improved accuracy and more run time complexity [8]. Thereby, HAAR characteristics are extracted by using an adaboost method reduces the run time complexity of detecting the fish, but the results were not as promising for real-world deployment [9]. The use of sonar images used for detecting and tracking fish in real time using conventional neural networks yielded promising results, but the captured image was obtained from an underwater camera that could not be connected to the remote vehicles [10]. It is observed that false detection occurs whenever overlapping fish objects are happened using the deep learning architecture, which includes fish recognition model and the support vector machine (SVM). The development and deployment of automated technology (fish detection and tracking mechanism) can help us understand fish species migration pattern across a wide range of spatiotemporal scales and ecological hierarchies (e.g., location, species crowed and communities), overcoming these limitations.

In this study, describes about the functionality of automatic fish detection and tracking system to enhance the underwater surveillance for monitoring ecological biodiversity. Recently, pre-trained Mask-R-CNN is involved for accurate object detection especially fish species and tracks the movement for studying behavioral characteristic continuously. Figure 1 shows the Fish feature extraction using mask-RCNN for species detection and classification. In this regard, propose a new dynamic classifying algorithm which supports an automatic fish identification system in terms of fish abundance, identify fish species, and follow fish activity. It is a derived version of existing deep learning algorithm namely minimum output sum of squared errors (MOSSE), sequential non-maximum suppression (Seq-NMS) and Siamese mask (SiamMask). The project target is to give potential benefits to ocean ecosystem for establishing the ecological management.

The major contribution of this paper is given as follows:

- To develop a dynamic classifying algorithm for detecting and tracking of the small size object (fish). Pre-trained Mask-R-CNN is used to derive the feature datasets which transfer to the dynamic classifying algorithm. Thus, enhance the tracking and detecting ability of the underwater surveillance monitoring system.
- It enables quality integration of different monitoring system for knowledge sharing of feature datasets.
- It obtains trusty and well sufficient fish features vector utilizes continuous tracking by placing acoustic sensor to detect motion.

- It detects the unauthorized fish species and ensures follow-up access and send quick control message shared to control unit from satellite services.

The organization of paper is given as follows: Section II describes contributions of recent articles towards object detection and object tracking. In section III, authors elaborates the data preparation, training phase, detection and tracking phase, dataset creation and system interface using pre-trained mask-R-CNN proposed in the research elaborately. Section IV briefs the simulation results of the proposed algorithm in comparison with other traditional methods availed in the literature. Finally, conclusion and future scope is given in section V.

Ii. Related Works

A. Object Detection

It is a branch of computer vision concerned with finding instances of objects in images and videos [12]. Traditionally, an image processing algorithms [13], as well as deep learning techniques [14], are commonly used for object detection. It influences a subset of machine learning that employs neural networks to learn higher-dimensional representations and recognize patterns in unstructured data [15]. Recently, the most promising structure is used for fish detection and tracking is referred as Mask Regional Convolutional Neural Network (Mask R-CNN) due to feature extraction purpose [16]. It efficiently locates and classifies objects of interest using open-access deep learning models [17]. To enhance the fish detection model, the sample training videos are collected from aquariums for extracting the fish features in order to train the Mask-R-CNN which is suitable for real time fish detection in the complex scenario of the ocean outlet. The six high definition cameras (1080p Haldex Sports Action Cam HD) are used to record the fish movement in 1-hour duration (catla and white mirgal) in various positions (Up, Right, Left and Down) by adjusting the camera angle to catch a variety of backgrounds and fish views. In order to accommodate more feature extraction, we cut the original 1-hour recordings into snippets where bream were present. After then, the snippets were transformed into still frames at a rate of 5 frames per second. The training videos consists of two fish species for better feature extraction which will be transfer to pre-trained mask-R-CNN for continuous tracking on target fishes even in densely populated scenario. Initially, annotation is manual marked across the subsequent frames for accurate feature learning. In the proposed pre-trained mask-R-CNN reaches with a learning rate of 0.0025 to train the model. For training phase, we employed a 92% random selection of sample used for annotated dataset, with the remaining 10% used for validation. Moreover, the early-stopping strategy to reduce overfitting by assessing proper validation in between set of intervals during iterations and determining when performance began to deteriorate. To ensure the overlapping tolerability extended upto 50% of segmented mask of the outline of the fish. We chose object detection outputs when the model was 80% or more confidence criterion. The proposed pre-trained mask-R-CNN provide flexible functionality of automatic fish detection and tracking system to enhance the underwater surveillance for monitoring ecological biodiversity.

B. Object Tracking

Due to the 3D medium that continuous fish migration tracking is difficult because of significant variance in the shape, size and texture of the objects which can be covered by floating objects [19]. These challenges are being addressed by advances in object tracking, and maintain consistently despite natural fluctuations in object size, shape and location [20]. In the proposed pre-trained mask-R-CNN includes fish detection architecture, it is being activate once the trained feature set matches with the instant video sequence. Subsequent ally, track the target fish species upto the segmented mask thresholding is defined by dynamic classifying algorithm which derived version of the deep learning algorithms such as minimum output sum of squared errors (MOSSE), sequential non-maximum suppression (Seq-NMS) and Siamese mask (SiamMask). Usually, it is very much difficult to demonstrate the underwater scenario through computer vision ability to find the specific fish spices in a multispecies assemblage and estimate movement direction. Figure 2 shows Architecture of continuous fish tracking model after superimposed the trained features. The six high definition cameras (1080p Haldex Sports Action Cam HD) are used to record the fish movement in 1-hour duration (catla and white mirgal) in various positions by adjusting the camera angle to catch a variety of backgrounds and fish views. Hence, it ensures to determine horizontal movement (left or right) of fish through the camera arrangement. The USV camera adjustment in all possible directions (up, down, right, left) for fish detection and tracking continuous on two datasets (Set 1: facing centre (Catla Fish) and Set 2: facing south (White Mirgal)). In order to accommodate more feature extraction, we cut the original 1-hour recordings into snippets where bream were present. After then, the snippets were transformed into still frames at a rate of 5 frames per second. The training videos consists of two fish species for better feature extraction which will be transfer to pre-trained mask-R-CNN for continuous tracking on target fishes even in densely populated scenario. Initially, annotation is manual marked across the subsequent frames for accurate feature learning. The evaluation matrices of the fish detection and tracking of Catla and Mirgal Fish species using dynamic classifying algorithm superimposed in pre-trained Mask-R-CNN architecture for every batch of frames such Frame (1–5) ,Frame (6–13), Frame (14–19), Frame (20–24), and Frame (25–30) respectively.

Iii. Pre-trained Mask-r-cnn Architecture

The pipeline of fish detection and continuous tracking using proposed architecture which includes four different stages is shown in Fig. 3.

Step: 1 It involves the use of an unmanned aerial vehicle (UAV) to acquire fish data, followed by data preparation. To create candidate fish zones, the detector is continuously run on the processing frame.

Step: 2 It creates candidate regions of interest (ROIs) are supplied into mask generation and categorization for future analysis.

The target (ROIs) is continuously tracked with help of the dynamic classifying algorithm at any point of the time instant.

A. Data Acquisition

In this proposed pre-trained Mask-R-CNN, the fish detection model is trained with the sample training raw videos are collected from aquariums (http://www.cs.toronto.edu/~dross/ivt/) and (https://deepblue.lib.umich.edu/data/concern/data_sets/bg257f267?locale=en) for extracting the fish features in order to train the Mask-R-CNN which is suitable for real time fish detection in the complex scenario of the ocean outlet. The six high definition cameras (1080p Haldex Sports Action Cam HD) are used to record the fish movement in 1-hour duration (catla and white mirgal) in various positions (Up, Right, Left and Down) by adjusting the camera angle to catch a variety of backgrounds and fish views. The training dataset of aquariums fisheries are considered for extracting the feature sets of two specific fish species namely (catla and white mirgal) placed in the summing pool which is exactly replicate the underwater scenario with clean water and with sediments water as shown in Fig. 3(a). Fish in a glass container submerged beneath the ocean's surface in water containing decaying organic materials as shown in Fig. 3(b).

B. Data pre-processing

Initially, trained the fish dataset is used to train the deep learning algorithm and must be processed to refine, the training dataset into 1200 frames. Manually analyzing and resizing the fish ROIs to a standard resolution of 640x320 pixels [6]. Additionally, annotation is manual marked across the subsequent frames for accurate feature learning using the open source.

C. Training Phase

The pre-trained model (ResNet 101) is used to recognize the fish features in the subsequent frames of the targeted ROIs and areas of interest are created by sliding windows at the last layer of the output [24, 25]. Both classifier and regresses layers confirm the available of target fish spices which are calculated by finding the likelihood of target fish in a suggested location. Figure 6 shows the pipeline of fish detection and continuous tracking architecture includes four different stages. The overview of working operation of proposed Mask-R-CNN is described as follows. First of all, underwater surveillance video is undergo data processing function by using gaussian filter in order to extract the salient features of training videos set as a sample trained feature sets. It is being compared with the real time underwater surveillance video, frame by frame detect the fish species based on mapping feature from trained feature sets. Then, it followed by segmenting specific fish species through clustering strategy in order to acquire features in the depth of penetration among densely populated zone as well as fish textures constraints. After evaluating its euclidean distance, the threshold point of tracking the individual fish is updated. In addition to this, R-CNN provides favorable and unfavorable criterion which may helpful to reach minimum membership function much faster. Then, selected segmented region further undergoes first level of fish spice classification is carried by dynamic classifying algorithm. Thus, it clearly shown in Fig. 4 and Fig. 5 for both fish species (catla and white mirgal). It gives two optimistic parameter values which includes local best and global best parameters. Thereby, set the threshold value for obtaining best pixel position which can trace out the entire fish region in the densely covered by other spices zone. Further, it provides assured optimistic parameter for estimating actual region of interest along with spices count rate present

in the targeted location. The proposed pre-trained Mask-R-CNN combine's fish detection and tracking algorithms (MOSSE, Seq-NMS and SiamMask) for better detection of specific fish spices and radiologist make easy to operate to find densely populated fish location for accruing more capturing ratio. When compared to other existing approach, the proposed dynamic classifying algorithm on the pre-trained Mask-R-CNN takes less computational complexity and high accuracy in terms of fish detection and tracking.

The pre-trained mask R-CNN includes localization, segmentation and classification are required for proper trained feature sample extraction.

$$F = F_{clc} + F_{box} + F_{mask}$$

1

Where $F_{clc}$and F_{box}be the clustering and boundary boxing for target fish detection at any dense zone respectively. It is calculated from Equ. (2), which may helpful to reach minimum membership function much faster R-CNN fish detection [26].

$$F\left(P_i, t_i\right) = 1 \left/ N_{clc}\left[\sum_i F_{clc}\left(P_i, P_i\right) + \lambda\left(1 \left/ N_{box}\sum_i P_i F_{box}\left(t_i, t_i\right)\right.\right)\right.\right]$$

2

Where, P_i is indicates targeted region whose projected probability lies either 1 or 0 depends up on the positive and negative anchor respectively. N_{clc} represents the number of mini-batch allocation. For each ROIs of target specified fish class, the fish mask branch constructs a mask of dimension$(M \times M)$. Suppose, if K classes are identified then the output will come in the size of$k - m^2$. Hence, the model lags to fetch mask for each targeted class, as result, there is no competition arise between the classes when it comes to creating masks. The average entropy of k-th mask ROIs associated with the ground truth class K. It is denoted asF_{mask}

$$F_{mask} = -1 \left/ m^2 \sum_{1 \leqslant i,j \leqslant m}\left[f_{i,j}\log\left(y_{i,j}^k\right) + \left(1 - f_{i,j}\right)\log\left(1 - \hat{f}_{i,j}^k\right)\right]\right.$$

3

D. Detection and Tracking Phase

It is critical to keep track of the discovered fish in order to increase run time efficiency. The proposed pre-trained Mask-R-CNN quickly learns the trained dataset for reliable fish tracking and detection. To enhance counting process of the tracked fish, this is subjected to generate a track id, and the tracking details updating in every 12 frames. The proposed pipeline works well in the real time scenario for fish detection, tracking, and counting by using the targeted ROIs as input for subsequent frames. Thus, it improves an efficiency of the pre-trained mask-R-CNN.

Iv. Results And Discussion

In this section explores the qualitative and quantitative evaluation of the automatic fish detection and tracking system to enhance the underwater surveillance for monitoring ecological biodiversity. The proposed pre-trained Mask-R-CNN framework is accurately identified the fish species and track widely in densely populated region by enhancing the visual quality of segmented mask through dynamic classifying algorithm which is the derived version of existing deep learning algorithm namely minimum output sum of squared errors (MOSSE), sequential non-maximum suppression (Seq-NMS) and Siamese mask (SiamMask). The experimental output is comprehensively compared with other existing MOSSE, Seq-NMS and SiamMask results to understand the progressive improvement achieved by proposed pre-trained Mask-R-CNN framework towards the fish detection and tracking functions. The region of interest is usually targeted the fish occupancy of large number which is effectively classified by using dynamic classifying algorithm. The deadly inactive region undergone several feature manipulation process and further develop into new segmented mask for accurate fish detection. It is very much difficult to find fishes in the densely level of the populated zone in these multifaceted layers. Thereby, it is necessary to have an efficient soft computing technique which can identify the fish features depth level as well as fish size accurately. In this paper, a pre-trained Mask-R-CNN framework is proposed to enhance the accuracy of fish detection process in the segmented mask and also fulfilled the continuous fish tracking of the target fish detected efficiently. For analysis purpose, the sample training raw videos are collected from aquariums (http://www.cs.toronto.edu/~dross/ivt/) for extracting the fish features in order to train the Mask-R-CNN which is suitable for real time fish detection in the complex scenario of the ocean outlet. The proposed work mainly focuses on supporting an automatic fish identification system in terms of fish abundance, identify fish species, and follow fish activity. The project target is to give potential benefits to ocean ecosystem for establishing the ecological management. The output of the proposed framework compared with the traditional methods using key confusion metrics of the fish detection and tracking of Catla and Mirgal Fish species using dynamic classifying algorithm superimposed in Mask-R-CNN architecture. The proposed method using ResNet101 shown better performance as compared to global-ROI-based methods and there by achieved effectiveness in terms of computational complexity and accuracy. These are five different set of frames cross validation is performed to evaluate the proposed pre-trained Mask-R-CNN framework as shown in Fig. 7 and Fig. 8, and the confusion matrices for each frames set shown in Fig. 9 for reference.

The confusion metric can measure closer similarity appear in the low level intensity region of both segmented mask and instant video sequence. It gives additional weight factor to ensure the accuracy of correlated coefficient values obtained when having closer similarity associated in the intensity scale. The correlated coefficient values changes between 0 and 1. Where, ‘0’ denotes low similarity index and 1 denotes high similarity index. It exhibits that the proposed pre-trained mask-R-CNNN framework produces the best ROI value of 92% (expressed in percentage) compared to other traditional methods specified in this work. Similarly, accuracy of the proposed pre-trained mask-R-CNNN framework has produced maximum accuracy (87%) in percentage as compared with MOSSE (75%), Seq-NMS (78%), and SiamMask (84%) respectively.

Table 1
The comparison chart of Mask R-CNN architecture using trained aquarium datasets

Method	Catla Fish		White Mirgal Fish		Accuracy
	Avg. Precision	Avg. Recall	Avg. Precision	Avg. Recall	
MOSSE	73.54 ± 3.63	73.08 ± 2.48	73.16 ± 2.05	73.44 ± 4.45	73.28 ± 2.22
Seq-NMS	74.12 ± 6.16	74.24 ± 3.16	74.18 ± 2.08	73.46 ± 8.47	73.86 ± 3.36
Siam Mask	84.72 ± 2.65	84.98 ± 3.72	84.20 ± 2.94	84.78 ± 3.22	84.40 ± 2.18
Hybrid algorithm	**87.02 ± 4.99**	**86.94 ± 4.79**	**87.18 ± 3.25**	**86.54 ± 6.21**	**86.72 ± 0.80**

Table 2
The comparison chart of Mask R-CNN architecture using underwater surveillance video datasets

Method	Catla Fish		White Mirgal Fish		Accuracy
	Avg. Precision	Avg. Recall	Avg. Precision	Avg. Recall	
MOSSE	73.46 ± 3.62	74.62 ± 7.56	74.82 ± 6.21	73.08 ± 4.40	73.86 ± 3.75
Seq-NMS	74.74 ± 2.79	73.06 ± 4.16	73.46 ± 3.43	75.00 ± 3.15	74.02 ± 2.77
Siam Mask	85.98 ± 7.02	85.78 ± 7.02	86.24 ± 4.51	85.00 ± 8.98	85.40 ± 1.42
Hybrid algorithm	**87.02 ± 4.99**	**86.94 ± 4.79**	**87.18 ± 3.25**	**86.54 ± 6.21**	**86.72 ± 0.80**

In order to ensure the quality measurement of fish detection and tracking functions based on the effective classifiers. It is being tested by calculating the three important factors such as average precision and average recall that can be evaluated by using Equ. (4), (5) and (6).

Precision = TP/TP + FP' (4)
Sensitivity=TP/TP+FN' (5)
Specificity=TN/TP+FP' (6)
Where ‘FP’, ‘TP’, and ‘TN’ indicates false positives rate, true positives rate and true negatives rate respectively. Table.1 shows the comparison chart of Mask R-CNN architecture using trained aquarium datasets that can segment target region accurately. Similarly, Table.2 shows the comparison chart of Mask R-CNN architecture using underwater surveillance video datasets.

The fish detection and tracking architectures is quite difficult but it is possible continuous frames tracking are established for four frames after the initial detection (MOSSE and SiamMask). The movement was calculated for Seq-NMS by calculating the space vectors between two detections. The fish detection model spotted a (catla and white mirgal) fishes bream and was carried for all frames, and films, resulting in an interaction between detections and tracker that lasted the entire length of a video. For each frame where the fish detection and tracking interactions were successful, all trackers gave a movement

direction. It obtains trusty and well sufficient fish features vector utilizes continuous tracking by placing acoustic sensor to detect motion and enables quality integration of different monitoring system for knowledge sharing of feature datasets. It is clearly evaluated by adjusting USV camera in all possible directions (up, down, right, left) for fish detection and tracking continuous on two datasets (Set 1: facing centre (Catla Fish) and Set 2: facing south (White Mirgal)) as shown in Fig. 10.

V. Conclusion

In this paper, a new dynamic classifying algorithm is proposed which supported fully automatic fish detection and tracking system in order to identify fish species and to track fish activities for understanding synapomorphies characteristic simultaneously. The pre-trained Mask-R-CNN is involved for having well-enhanced feature vectors derived after undergone with number of test samples taken from captured video footage. It has improved the system function of automatic fish detection and tracking to enhance the underwater surveillance for monitoring ecological biodiversity. The comparative analysis is made with existing tracking R-CNN algorithms such as minimum output sum of squared errors (MOSSE), sequential non-maximum suppression (Seq-NMS) and Siamese mask (SiamMask). Simulation results conveys that the proposed algorithm support effective fish detection (i.e. around 120 out of 170 individual bream) and accuracy of the pre-trained Mask-R-CNN (87%) as compared with MOSSE (75%), Seq-NMS (78%), and SiamMask (84%) respectively. Thus, the evaluation result shows, the proposed pre-trained Mask-R-CNN achieves reasonable improvement in accuracy (detection and tracking) which give potential benefits to ocean ecosystem for establishing the ecological management. Moreover, it has enabled quality integration of different monitoring system for knowledge sharing of feature datasets and obtained trusty and well sufficient fish features vector utilizes continuous tracking by placing acoustic sensor to detect motion.

Declarations

Funding : Not Applicable

Conflicts of interest :

The authors declare that they have no conflict of interest. This article does not contain any studies with human participants or animals performed by any of the authors.

Availability of data and material : Not Applicable

Code availability : Not Applicable

References

1. https://nfdb.gov.in/about-indian-fisheries

2. Teame (2017) Ghirmai Some Aspects of Sustainable Management of Resources and Environment in India
3. Berman B, David J, Sanchez-Jerez, Armando P, Armando M (2018) Detection of aquaculture structures using Sentinel-1 data
4. Abrantes KG, Barnett A, Baker R, Sheaves M (2015) Habitat-specific food webs and trophic interactions supporting coastal-dependent fishery species: An Australian case study. Rev Fish Biol Fish 25(2):337–363
5. Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, Hasan M, van Essen BC, Awwal AAS, Asari V (2019) K. A state-of-the-art survey on deep learning theory and architectures. Electronics 8(3):292
6. Ditria EM, Lopez-Marcano S, Sievers M, Jinks EL, Brown CJ, Connolly RM (2020) Automating the analysis of fish abundance using object detection: Optimizing animal ecology with deep learning. Front Mar Sci 7:429
7. Shrivakshan GT, Chandrasekar C (2012) A Comparison of Various Edge Detection Techniques used in Image Processing.IJCSI_ International Journal of Computer Science Issues;pp.272–276
8. Massa F, Girshick R (2018) Maskrcnn-benchmark: Fast, modular reference implementation of instance segmentation and object detection algorithms in PyTorch
9. Kim B (2017) Son-Cheol Yu. Imaging SONAR Based Real Time Underwater Object Detection utilising AdaBoost Method.IEEE Underwater Technology (UT)
10. Kim J, Yu S-C (2016) Convolutional neural network-based real-time ROV detection using forward-looking sonar image. IEEE/OES Autonomous Underwater Vehicles
11. Moniruzzaman M, Islam SMS, Bennamoun M, Lavery P (2017) Deep learning on underwater marine object detection: A survey. Paper presented at the International Conference on Advanced Concepts for Intelligent Vision Systems
12. Zhao Z-Q, Zheng P, Xu S-T, Wu X (2019) Object detection with deep learning: A review. ArXiv
13. Lecun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444. https://doi.org/10.1038/natur e14539
14. Cui S, Zhou Y, Wang Y, Zhai L (2020) Fish detection using deep learning. Applied Computational Intelligence and Soft Computing, 2020, 3738108. https://doi.org/10.1155/2020/3738108
15. Ditria EM, Lopez-Marcano S, Sievers M, Jinks EL, Brown CJ, Connolly RM (2020) Automating the analysis of fish abundance using object detection: Optimizing animal ecology with deep learning. Front Mar Sci 7. https://doi.org/10.3389/fmars.2020.00429
16. Ditria E, Sievers M, Lopez-Marcano S, Jinks EL, Connolly RM (2020a) Deep learning for automated analysis of fish abundance: The benefits of training across multiple habitats. https://doi.org/10.1101/2020.05.19.105056. bioRxiv
17. Guo S, Xu P, Miao Q, Shao G, Chapman CA, Chen X, He G, Fang D, Zhang HE, Sun Y, Shi Z, Li B (2020) Automatic identification of individual primates with deep learning techniques. iScience

23(8):101412. https://doi.org/10.1016/j.isci.2020.101412
18. Prechelt L (2012) Early stopping —but when? In: Montavon G, Orr GB, Müller K-R (eds) Neural networks: Tricks of the trade, 2nd edn. Springer, Berlin Heidelberg, pp 53–67
19. Sidhu R (2016) Tutorial on minimum output sum of squared error filter. Degree of Master of Science, Colorado State University
20. Bolme DS, Beveridge JR, Draper BA, Lui YM (2010) Visual object tracking using adaptive correlation filters. IEEE Conference on Computer Vision and Pattern Recognition, 2544–2550
21. Han W, Khorrami P, Le Paine T, Ramachandran P, Babaeizadeh M, Shi H, Huang T (2016) Seq-NMS for video object detection.ArXiv, 3
22. Cheng JC, Tsai YH, Hung WC, Wang SJ, Yang MH (2018) Fast and accurate online video object segmentation via tracking parts. IEEE Conference on Computer Vision and Pattern Recognition, pp. 7415–7424
23. Ahilan Y, Adityan AV, Kailash S (2015) Efficient Utilization of Unmanned Aerial Vehicle (UAV) for Fishing through Surveillance for Fishermen.International Journal of Aerospace and Mechanical Engineering
24. Ren S, He, Kaiming G, Sun R, Jian. Faster R-CNN (2015) Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell 39. 10.1109/TPAMI.2016.2577031
25. Image Recognition. IEEE Conference on Computer Vision and Patten Recognition (2016)
26. Annotation tool for fish annoation. Available online: http://www.robots.ox.ac.uk/~vgg/software/via/

Figures

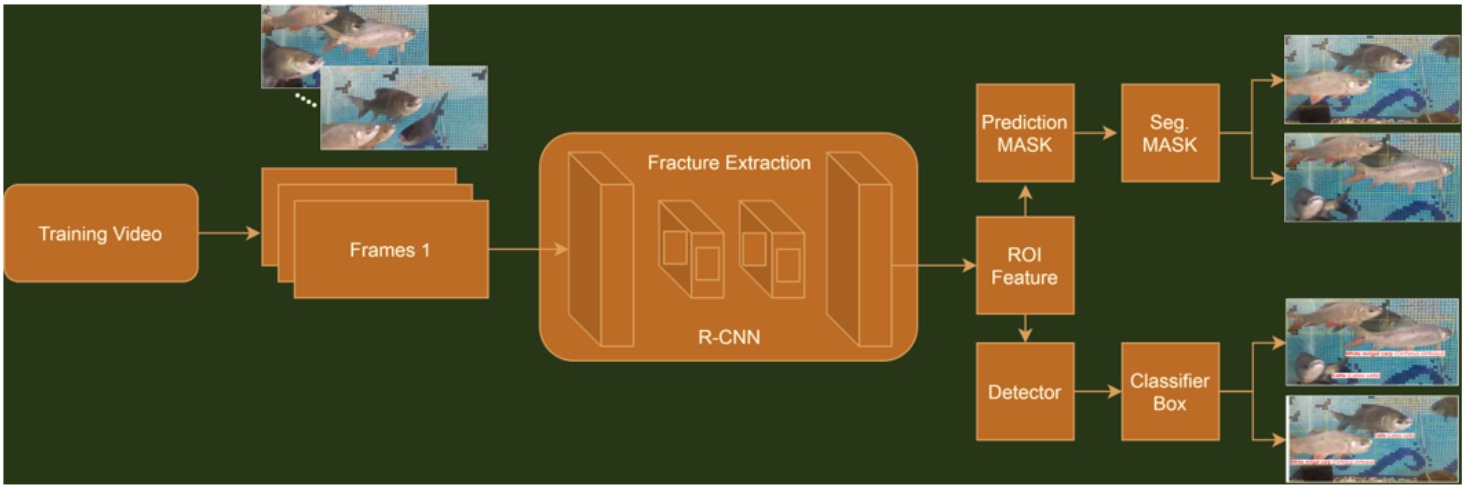


Figure 1

Fish feature extraction using mask-RCNN for species detection and classification. Further enhancing the running time efficiency a hybrid tracking algorithm is used for continuous tracking of the fish species.

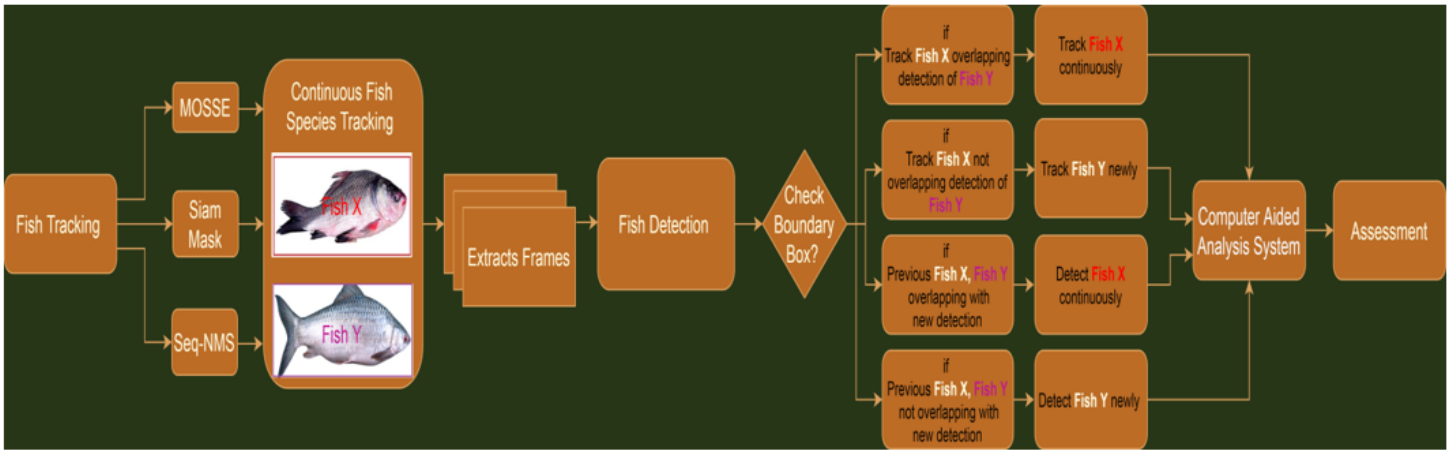


Figure 2

Architecture of continuous fish tracking model after superimposed the trained features.

Figure 3

(a) Data on fish was collected in artificial inland water with a distinct underwater background. (b) Fish in a glass box surrounded by organic stuff.

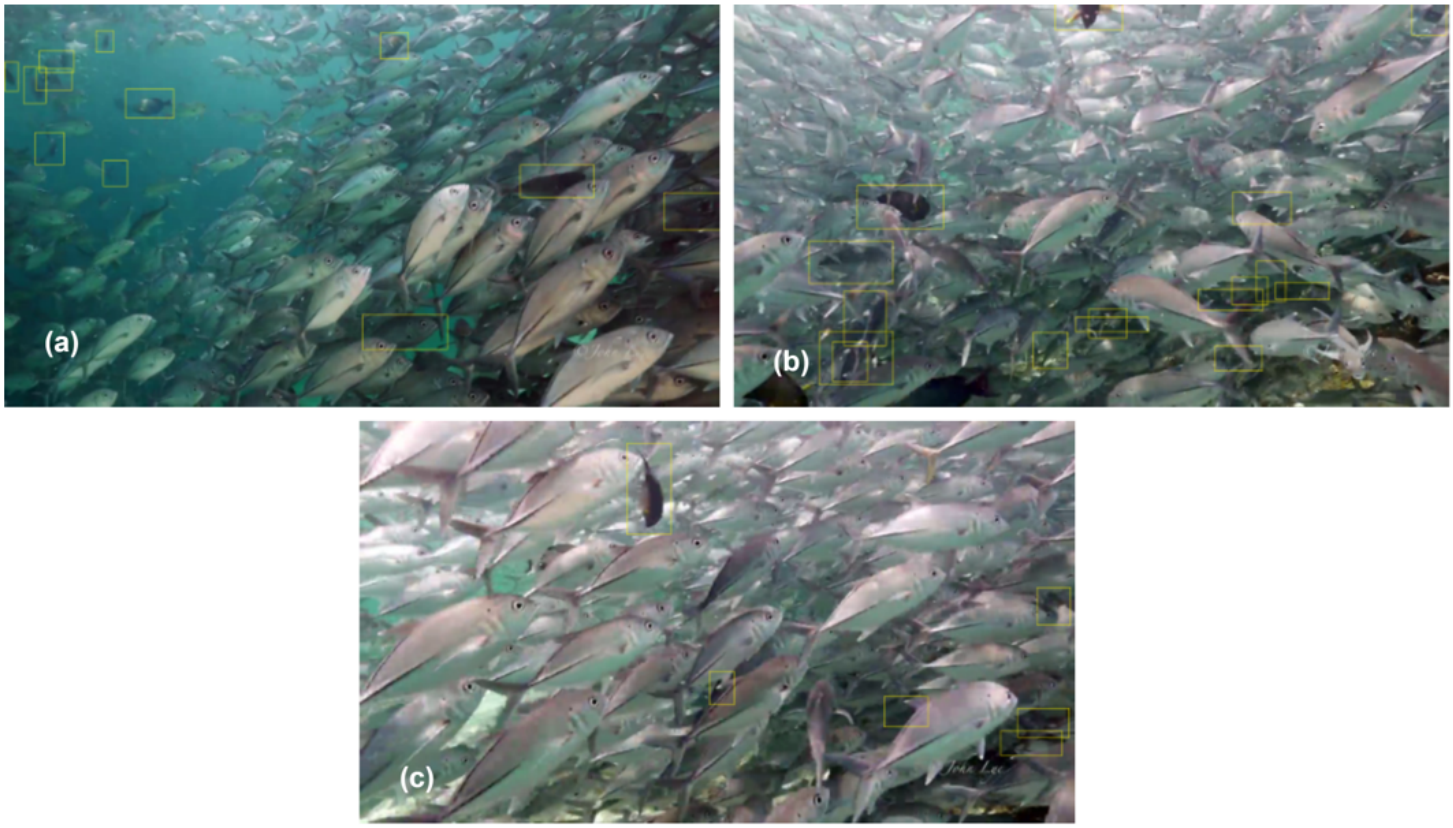


Figure 4

From the acquired fish dataset, a highlighted catla fish among other spices densely populated zone (a) Accurate fish detection and classiifcation at left side (b) Accurate fish detection and classiifcation at right side, (c) Accurate fish detection and classiifcation at down side

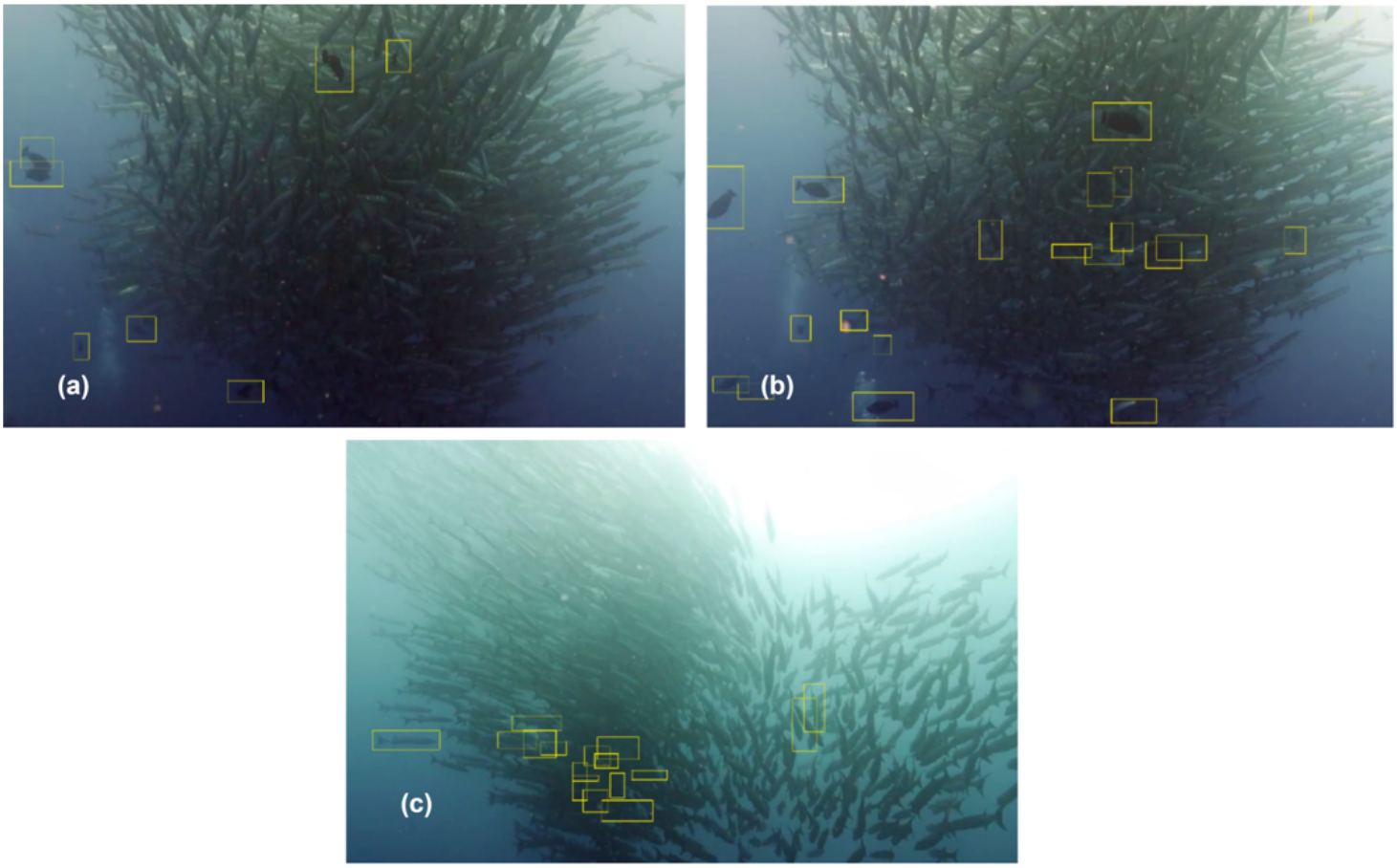


Figure 5

From the acquired fish dataset, a highlighted white mirgal fish among other spices densely populated zone (a) Accurate fish detection and classiifcation at top-left side (b) Accurate fish detection and classiifcation at centre, (c) Accurate fish detection and classiifcation at down side

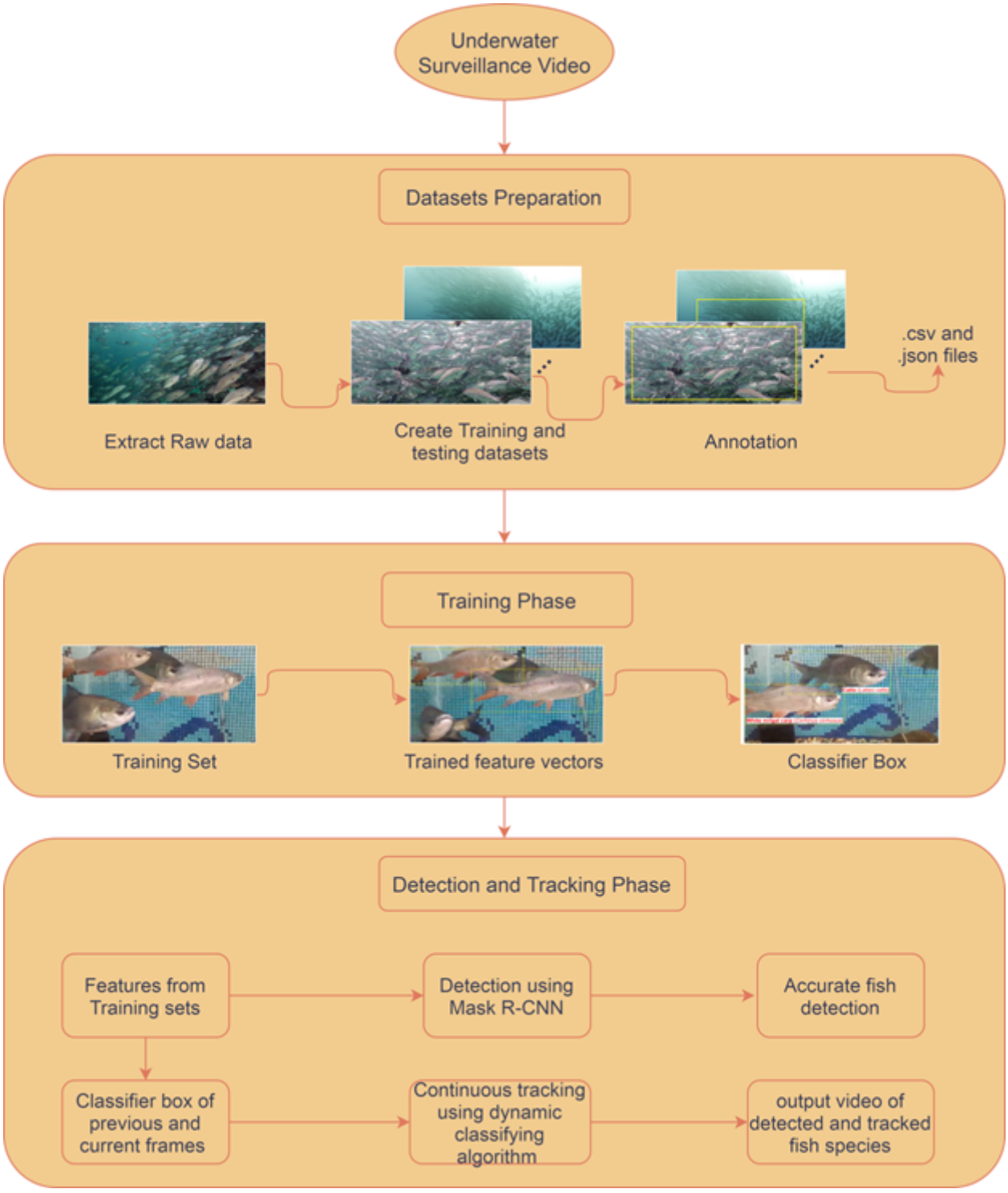


Figure 6

Pipeline of fish identifcation and continous tracking in four different scenarios. (i) Collection of fish data using Underwater surveillence video (ii) ROI by manual annotation for fish detection. (iii) Fish feactures training phase for accurate fish detection and classiifcation using hyper parameter turning. (iv) Continuous fish tracking model for real time computer aided interface.

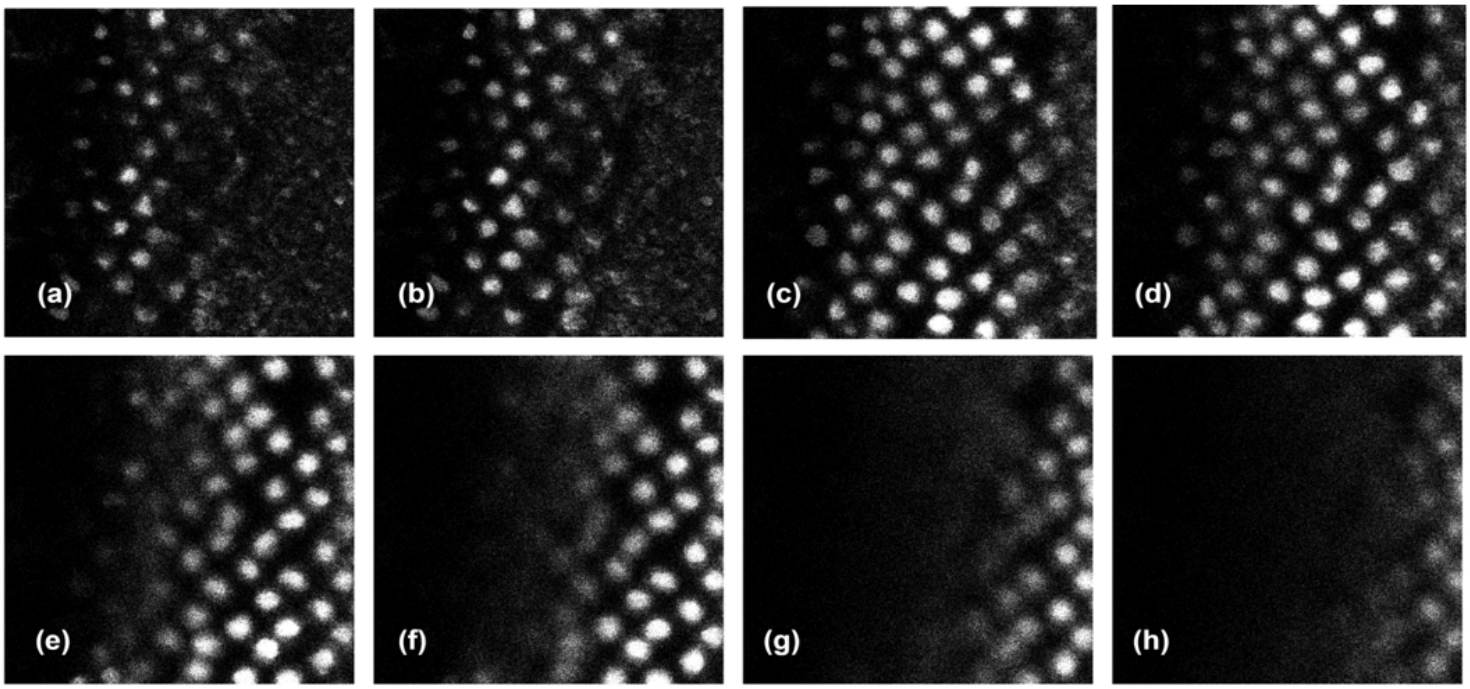


Figure 7

Fish feature extraction in the consecutive frames (top to bottom and left to right) from an underwater surveillance video in order to detect and track the fish species (catla). (a) Frame (1), (b) Frame (4), (c) Frame (8), (d) Frame (12), (e) Frame (16), (f) Frame (20), (g) Frame (24) and (h) Frame (28)

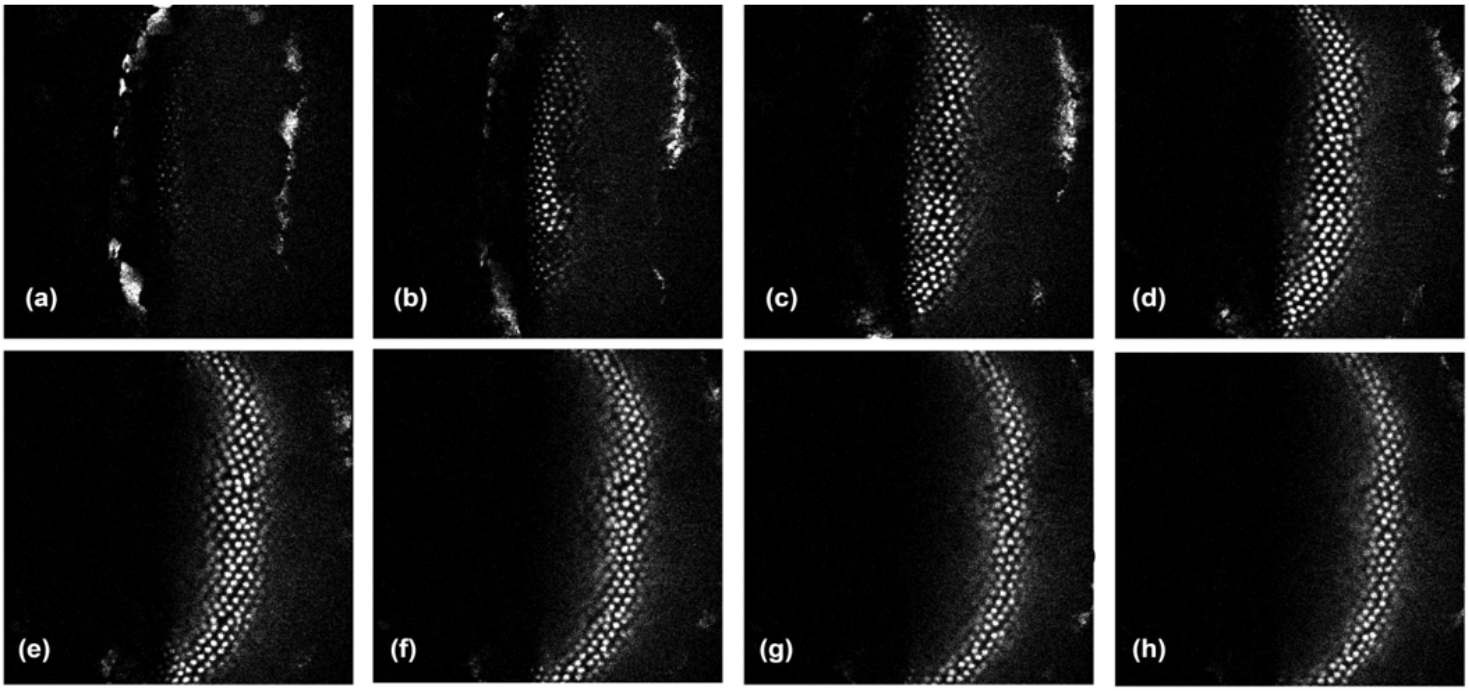


Figure 8

Fish feature extraction in the consecutive frames (top to bottom and left to right) from an underwater surveillance video in order to detect and track the fish species (white mirgal). (a) Frame (1), (b) Frame (4), (c) Frame (8), (d) Frame (12), (e) Frame (16), (f) Frame (20), (g) Frame (24) and (h) Frame (28)

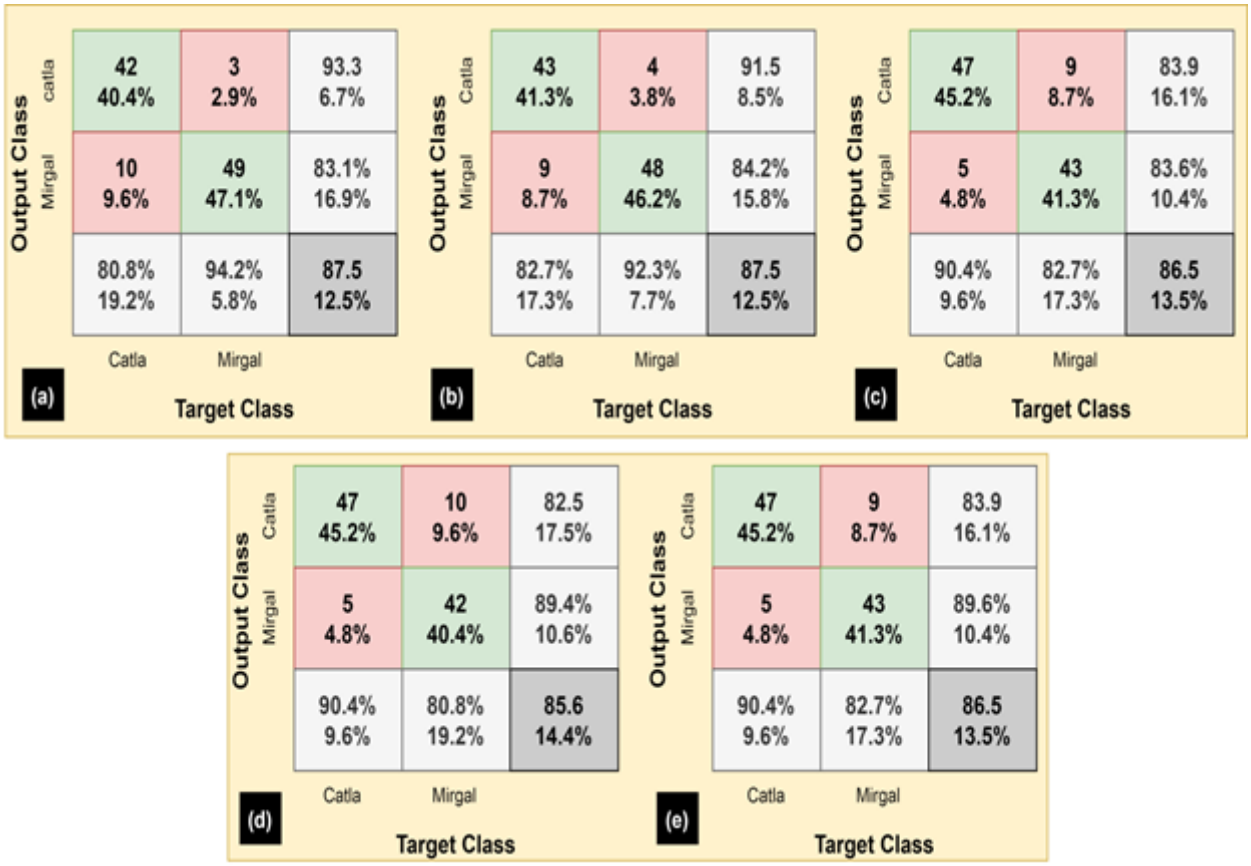
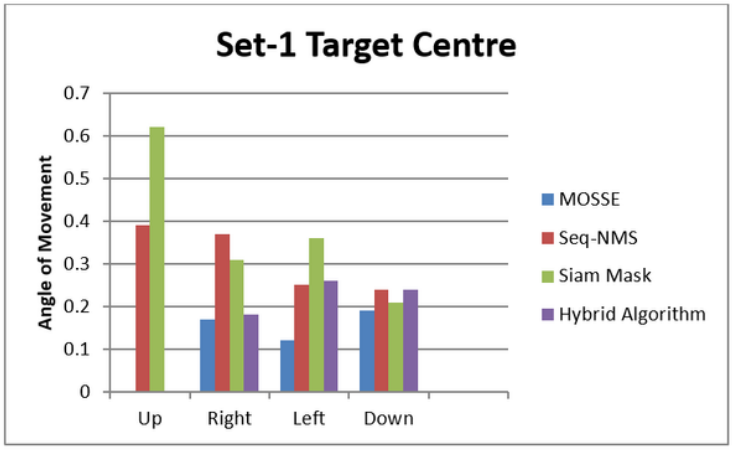
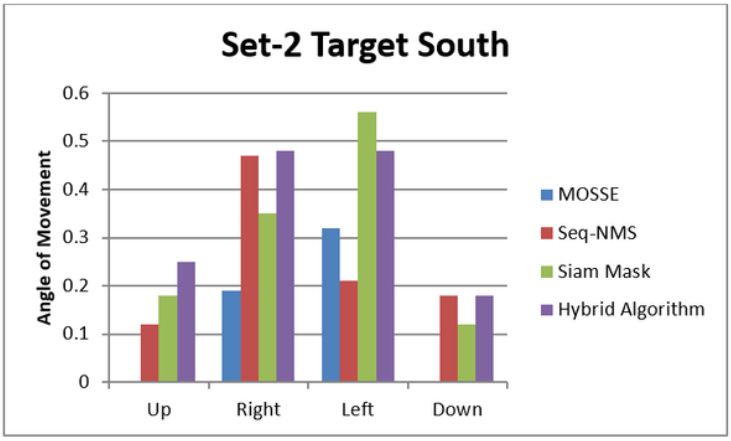


Figure 9

The confusion matrices of the fish detection and tracking of Catla and Mirgal Fish species using dynamic classifying algorithm superimposed in Mask-R-CNN architecture (a) Frame (1-5), (b) Frame (6-13), (c) Frame (14-19), (d) Frame (20-24), and (e) Frame (25-30).



(a)



(b)

Figure 10

USV camera adjustment in all possible directions (right, left, up, down,) for continuous fish detection and tracking on two datasets (Set 1: facing centre (Catla Fish) and Set 2: facing south (White Mirgal)).