IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 31 January 2023, accepted 10 February 2023, date of publication 22 February 2023, date of current version 28 March 2023.
Digital Object Identifier 10.1109/ACCESS.2023.3247143

RESEARCH ARTICLE

Fish Tracking and Continual Behavioral Pattern Clustering Using Novel Sillago Sihama Vid (SSVid)

S. SHREESHA[1], M. M. MANOHARA PAI[1], (Senior Member, IEEE), UJJWAL VERMA[2], (Senior Member, IEEE), AND RADHIKA M. PAI[3], (Senior Member, IEEE)
[1]Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
[2]Department of Electronics and Communication Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal 576104, India
[3]Department of Data Science and Computer Applications, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

Corresponding authors: M. M. Manohara Pai (mmm.pai@manipal.edu), Ujjwal Verma (ujjwal.verma@manipal.edu), and Radhika M. Pai (radhika.pai@manipal.edu)

This work was supported by the Department of Science and Technology, Government of India.

ABSTRACT Aquaculture provides food security to many developing countries and enhances the socio-economic conditions of the fishermen. To enhance the productivity of the aquaculture, it is necessary to maintain stress free controlled eco-system for the fishes. For recognising the stress in fishes, behaviour analysis of fishes via tracking is imperative. Early detection of stress in fish facilitates fishermen to take precautionary measures promptly. Computer vision-based fish behaviour analysis of economically important fish species is challenging due to the lack of datasets, occlusions, rapid changes in swim directions etc. The present study proposes a multiple fish video dataset of an economically important species in a controlled environment, namely **Sillago Sihama-Vid** with accurate annotations. The study emulates the natural environment of *Sillago Sihama* in a large aquarium. This work proposes a novel fish tracking algorithm that incorporates swim direction information in addition to temporal, appearance, and spatial information. The inclusion of swim direction information reduces the number of identity switches. Comparative performance analysis of the proposed tracking algorithm with the conventional methods on the developed dataset highlights the performance efficiency. The proposed method has a clear performance improvement in MOTA, MOTP, IDSW and MT with respect to the other compared methods. The study also presents a novel unsupervised continual behaviour modelling strategy to model the evolving behaviours of the fishes. Further, interpretation of fish behaviour from the proposed behaviour modelling is performed to highlight the reliability of the proposed method. The significance of the proposed method is that, it is independent of training and labelled data. In addition, the method represents an innovative alternative to capture all the non observable behaviours of the fishes. The proposed tracking and behaviour modelling strategy act as a benchmark for developing algorithms to study fish behaviour via tracking. Finally, the dataset provides an opportunity for developing computer vision-based models to analyse the different behaviours of fish *Sillago Sihama*.

INDEX TERMS Pattern analysis, fish behaviour analysis, fish tracking, fish trajectory dataset.

I. INTRODUCTION

Aquaculture farming has become more popular in the recent past as the demand for quality aquatic products has increased steadily [1]. In conventional farming, the fish farmers rely on manual measuring of growth rate, laboratory-based water quality monitoring, manual identification of abnormal behaviours and diseases in fishes, manual feeding etc. These methods are time-consuming, labour-intensive and lack efficiency, affecting both quality fish production and

The associate editor coordinating the review of this manuscript and approving it for publication was Claudio Zunino.

commercial profits [2], [3], [4]. As a result, they fail to keep pace with the growing demand. The shift from the traditional approach to technology use in fish farming can help the fish farmers to keep pace with the demand [5], [6], [7]. In the last few years, researchers used computer vision-based methods to analyse images/videos of fish for estimating the growth rate, wastage of feed, species identification, behaviour analysis etc. [8], [9], [10].

Analyzing the behaviour of fish involves understanding and interpreting their responses to a particular stimulus and dominant environment. The behavioural study helps maintain healthy fish growth, leading to good quality and productivity on fish farms. Typically, behavioural analysis focuses on the activity patterns of the fish. The generally observed patterns involving fishes are swimming, feeding, flight, resting and spawning. However, under stress, the fishes exhibit different movement patterns. For example, fish under stress may display rapid movements, rubbing their belly, gasping for air at the surface, staying at one location for a very long time, etc. Unless fishermen take appropriate measures to overcome these anomalies, the mortality rate may increase. On the other hand, timely corrective action to address anomalies will improve sustainability and productivity. Identifying these abnormal patterns can be effectively done by tracking fish movements in the videos.

Object tracking is defined as the process of estimating the position of the target object and maintaining the identities across frames. It further categorises into single and multiple object tracking as a function of the number of objects tracked. The current study considers a multiple-object tracking scenario as fish farms generally consists of a *n* number of fishes. Several studies have been carried out in the literature concerning single/multiple objects tracking [11], [12], [13]. However, these extensively focus on pedestrian/vehicle tracking. Fish tracking poses several challenges in comparison to pedestrian and vehicle tracking. One of the prominent differences is that pedestrians/vehicles typically have a linear movement while fish can move in three dimensions in the water volume. In addition, unlike pedestrians/vehicles distinctive characteristics such as colour, shape, etc., the same species of fish, regardless of gender, appear the same. Therefore, the current method of tracking pedestrians and vehicles will lead to the erroneous merging of multiple paths of different objects into one because of the similar appearance (identity switch). Accordingly, object tracking algorithms designed to track pedestrians and vehicles may not be effective at tracking fish. It is, therefore, necessary to develop a new monitoring strategy that will effectively address these challenges in monitoring fish of a single species. There exist fish tracking algorithms in the literature [9], [14], [15], but these algorithms mostly use either appearance-based models or motion-based models to track fishes. Motion only models do not effectively track fish since they change their swim directions rapidly. As well, the similar appearance of fishes makes appearance-based tracking models difficult. Hence, an efficient tracking algorithm that can address sudden changes in swim directions and appearance similarities of fishes is a compelling necessity.

Estimating the trajectory of fish and studying their behaviour requires good-quality annotated datasets with good resolution, sufficient illumination, and minimal movement of the camera [16], [17]. However, existing fish datasets are captured in the wild and include inedible fish species. Further, these datasets suffer from camera motion and poor illumination. Thus, a comprehensive dataset is the need of the hour for studying fish behaviour in fish farms. This research aims to study the behaviour of a single fish species (*Sillago Sihama*) grown in a large aquarium. This species harvested in the Indo-Pacific regions is economically significant. In recent years, *Sillago Sihama*, is cultured in cages as it is a highly nutritious fish with good sources of protein and essential amino acids [18]. In addition, cage farming offers an additional source of income for fishermen. The current study uses a large aquarium to emulate the behaviours of fish shown in a confined environment such as cages. The behaviour study of *Sillago Sihama* is essential to improve the harvest of cage cultivation, ensure food security and improve the socio-economic conditions of fishermen.

This work proposes a multi-object tracking algorithm for estimating the trajectory of multiple *Sillago Sihama* cultured in an aquarium. In addition to the motion and appearance features, the proposed Tracking By Detection (TBD) algorithm incorporates spatial and swim directions features. The appearance and motion similarity scores are refined using Gaussian kernels. The inclusion of Gaussian kernels and swim direction information reduces the number of identity switches that eventuate due to the similar appearance of the fishes. Further, the study proposes a novel continual fish behaviour modelling strategy. The proposed fish behaviour modelling strategy uses the estimated fish trajectory and optical flow information to model behaviour using an unsupervised learning approach. The principal advantages of the proposed method over the existing method is its in-dependency to training, labelled data and supervision. Further, the proposed method can capture all unobserved behaviour of fishes in a continual manner. The work also presents a new fish tracking dataset containing videos of *Sillago Sihama* cultured in an aquarium. The experimental setup emulates the natural environment for the fish *Sillago Sihama*. Finally, the study brings out the performance comparison of the proposed tracking algorithm with the existing methods on this dataset.

The major contributions of this work is as follows:

1) A novel dataset of economically important fish for studying the behaviour. The dataset contains videos of multiple *Sillago Sihama* along with manually annotated trajectories for more than 4000 frames.
2) A novel multiple fish tracking strategy with performance analysis and comparative study. The tracking

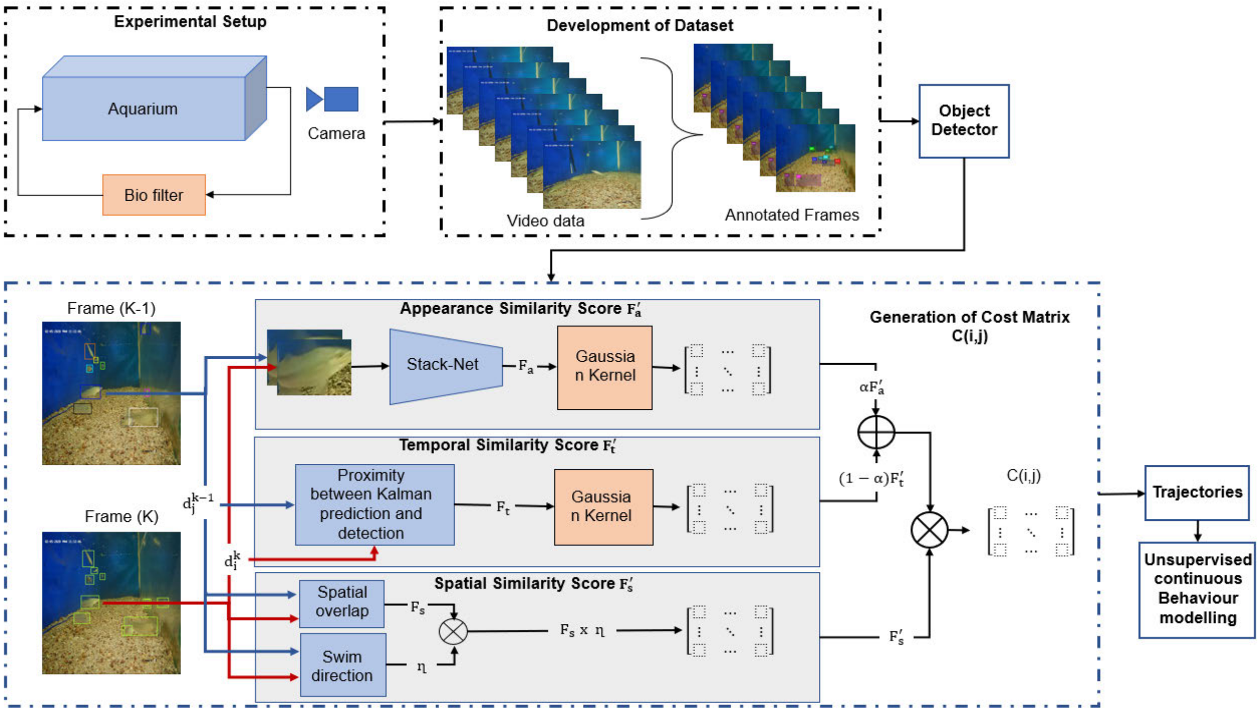


FIGURE 1. Overview of the proposed study. A new annotated dataset for tracking fishes is developed. Temporal, appearance and spatial information of the fishes are refined using Gaussian kernel and swim direction information. Final cost matrix $C(i, j)$ is generated from (F'_a), (F'_t) and (F'_s). Subsequently, data association is performed using linear sum assignment. The trajectory information of each fishes are used to perform continuous behaviour modelling.

strategy incorporates Gaussian kernel and optical flow-based swim direction information.

3) A novel unsupervised behaviour model to interpret the behaviour of fishes and validation of the model using proposed dataset.

The paper is organised as follows: Section I provides the introduction of the paper. Section II summarizes existing object tracking and behaviour analysis methods along with a brief discussion on existing fish datasets. Section III presents description of the experimental setup and fish tracking dataset proposed in this work. Section IV presents the details of the proposed tracking algorithm and unsupervised continuous modelling of fish behaviour. Experiments and results are presented in Section V and the paper is concluded in Section VI. The sample results are available at *https://github.com/shreesha-sur/Sillago-Sihama-Vid*.

II. RELATED WORK

Computer vision techniques are widely used to process videos and images for performing semantic segmentation, re-identification, object tracking and object detection [5], [19], [20], [21], [22], [23], [24], [25]. This section summarizes the existing fish tracking and behaviour analysis methods. Further, it presents existing publicly available standard datasets of fishes.

A. FISH TRACKING DATASETS

Most of the datasets of fishes available in the literature are collected from natural habitats. The authors generated these datasets for developing algorithms such as tracking, object detections, species identification etc. The *National Oceanic and Atmospheric Administration*(NOAA) has provided image dataset of wild fishes [17]. These videos are collected in Southern California Bight using a Remotely Operated Vehicle (ROV). Their main purpose is to encourage the development, testing and performance verification of algorithms for recognition, detection and tracking in underwater videos. Another dataset (*Fish4knowledge dataset*) [16] contains two subsets of underwater videos, one of 10 minutes duration and the other of full-day duration. These videos also contain ground truth information for fish detection and fish trajectory for fishes in the wild. Further, *Life CLEF (2014, 2015)* challenge uses the Fish4Knowledge dataset to develop automatic fish identification algorithms for estimating the existence and quantity of fishes [26]. *Sea clef (2016,2017)* originated from *Life CLEF (2014, 2015)* with focus on fish species and sea organism in general. The dataset contains videos and images of marine organisms. The purpose is mainly on ecological surveillance and biodiversity monitoring. *OzFish dataset* is an open-source dataset mainly developed for classifying fishes. The authors used Baited Remote Underwater Video Stations (BRUVS) to collect the images. This dataset collected in the wild consists of more than 507 different species of fishes [27]. *A Large-scale fish dataset* consists of images of nine distinct types of fish from the fish market [28]. The purpose of the dataset is the quality assessment of seafood using segmentation and classification. *Aquarium dataset* consists of 638 images of the fishes from 2 aquariums in USA [29].

The authors developed the dataset to conserve coral reefs, monitor environmental health, swimmer safety etc. *3D-ZeF* is a 3D RGB publicly available dataset for studying the behaviour of fishes [30]. The species used in the study is Zebrafish. The dataset contains 8 sequences of $15-120$ seconds duration with $1-10$ fishes. The authors intend to encourage the behaviour analysis of fishes as these fishes are used to study drug addiction, neurological disorders etc. In [31], the authors developed the Sablefish startle behaviour dataset for automatic behaviour detection. Their purpose is to aid the researchers to study climate change by analysing the behaviour of Sablefish. It is also observed from the literature that there exist several behaviour studies of the fishes in aquariums and tanks [32], [33], [34], [35], [36], [37], [38], [39]. However, these datasets are not available to the public domain for further studies.

The existing datasets collected from the natural habitat suffer from illumination variations, bright light intensity and resolution. In addition, they are not focused on farmed fish species but instead consist of random fish species present in the wild. Studying the behaviour of farmed fish species can enhance harvest, ensuring food security. Fish behave differently in confined habitats such as cages, as they do not interact with other fish species and have restricted movements. A controlled environment is essential for the behaviour study of caged fishes. Therefore, to study the behaviour of farmed fish, it is necessary to develop a fish dataset in a controlled environment. To this end, the present study uses a large aquarium. Further, the study chooses a single consumable fish species for observation in the aquarium.

B. OBJECT TRACKING METHODS

There are several published works on object tracking algorithms and most of these object tracking algorithms focus on tracking pedestrian/vehicles [40], [41], [42], [43]. The standard frame-work for object tracking followed in the literature is TBD [13], [44]. However, this approach is dependent on an efficient object detector. Yolo V3, Yolo V4, Single Shot Detector, Faster R-CNN, etc., are commonly explored for detecting the target objects [45], [46], [47], [48], [49]. Subsequently, similarities between the two target objects is computed using CNN based architectures. To this end, the authors frequently used StackNet, Siamese-net [22], [44] etc. Alternatively, authors extracted deep features from the target objects and computed the similarity scores using cosine distance [13]. Thus generated optimization problem is often solved using Hungarian algorithm, Graph-Cut [13], [44] etc. Also, there are tracking models which track the target objects directly in an end-to-end manner [40], [50], [51].

The existing fish detection and tracking methods in the literature predominantly use conventional machine learning based approaches such as Histogram of Oriented Gradients (HOG), background subtraction, Gaussian Mixture Models, Kalman filter, Hungarian algorithm and Viola Jones based methods [8], [10], [14], [15], [34], [52], [53], [54], [55], [56], [56], [57], [58]. Kalman filter performs better when the motion of the objects are linear. However, fishes have non linear motion which limits the effectiveness of Kalman filter [59]. To this end, in the present study, an algorithm is designed to incorporate swim direction information to reduce identity switches. Background subtraction works efficiently with the constant background. However, it does not perform well when the water is murky and the video quality is poor. The application of these non-CNN based approaches is not efficient in extracting the trajectories of fishes. CNN based methods can perform well in these situations. Recently, there has been an upsurge in the usage of these models to track animals [30], [31], [60], [61]. Generally, motion or appearance models are popularly adopted to track the target objects. However, the similar appearance and heterogeneous movement patterns of fishes often produce higher identity switches and reduce tracking accuracy. In this regard, the present study proposes a novel method to capture the trajectories of fishes using appearance, temporal and spatial information. Further, it devises an approach to refine the appearance, temporal and spatial similarity scores.

C. FISH BEHAVIOUR ANALYSIS USING TRACKING

There exist several work in the literature concerning the fish behaviour study [32], [33], [34], [35], [36], [37], [38], [39]. Most common approach in the literature to study the behaviour of fish is to extract the trajectories and analyse it. The studies mostly used zebra and tilapia fish species [14], [15], [38], [57], [30], [35]. It is interesting to note that these studies were conducted using tanks or aquarium [57], [58], [62]. The authors studied various distinct behaviours of fishes such as speed of swim, escape behaviour, surface swimming behaviour, effects of ambient light, collective behaviour etc., [56], [57], [58], [62], [33], [60]. Most of these behavioural analyses are performed manually or by simple thresholding. Also, some works focus on classifying the behaviours of fishes as normal and abnormal based on movement patterns [9], [14]. To this end, features are extracted and clustered using clustering algorithms. Support Vector Machine(SVM) or hierarchical classifiers were used to classify the extracted features as outliers. Further, understanding the fish's behaviour from its movement patterns is a challenging task as there might be very minute changes from one behaviour to another. These minute changes are hard to differentiate by human eyes. In addition, supervised and semi-supervised behaviour modelling methods fail due to a lack of data with ground truths from domain experts. Further, no work exists that models the continually changing behaviours in an unsupervised manner. Thus, an accurate unsupervised method of modelling fish behaviour is the need of the hour for a better understanding of the culture systems.

The present work develops a standard fish tracking video dataset of *Sillago Sihama* in a confined environment. Besides, it proposes a multi-object tracking algorithm to address the issue of identity switch in tracking fish. Further, the study


(a)


(b)

FIGURE 2. Experimental setup for collecting the dataset. (a) Sample image of aquarium with camera highlighted in the red circle. (b) Sample image of aquarium with blue HDPE sheets and sand bed.

presents a novel unsupervised behaviour modelling strategy to model the evolving behaviour of the fish continuously. The overview of the proposed study is shown in Figure 1.

III. SILLAGO SIHAMA VID (SSVid): FISH TRACKING DATASET

This section describes the details of the proposed fish tracking dataset. The dataset contains videos of *Sillago Sihama* cultured in an aquarium (Section III-A). The details of the videos collected are presented in Section III-B. Section III-C presents the annotation protocol followed for tracking *Sillago Sihama* in the acquired videos.

A. EXPERIMENTAL SETUP

An aquarium of size $2.4m \times 0.9m \times 1.22m$ is constructed in the premises using acrylic material. The experimental setup uses biofilters to reduce the amount of ammonia, nitrite, and nitrate produced by fish excrement in the system. Also, it utilizes a mechanical filter for removing the solid wastes generated in the aquarium by the fishes. The aeration for the fishes is provided by adopting necessary procedures in the biofilter, consequently avoiding the usage of an aerator and minimizing the disturbances in the water. The high-quality camera is set up on one side of the aquarium for capturing the movements of the fishes. The camera is fixed at this end to avoid any shake in the captured videos. However, the rest of the sides are covered with blue HDPE sheets for blocking the surrounding scene information in the collected video data. Figure 2 shows the experimental setup.

The camera captures the video at 12 Frames per second (FPS) with 1920×1080-pixel resolution. The frames have been resized to 800×800- pixel resolution during annotation. The experimental set-up includes lights to provide the necessary illumination during the video capturing process. The study utilizes *Sillago Sihama* fish species caught from the back-waters of Sita-Swarna River, Udupi district, Karnataka, India to populate the aquarium. It is a marine fish capable of changing its colour to match the environment. In addition, it is a bottom-dwelling fish with high nutritional value. The experimental set-up includes a sand bed at the bottom of the aquarium to create a natural environment for the *Sillago Sihama* fishes. In the present study, 8 to 10 fishes are grown in the aquarium. The motion information present in the video is solely due to the activity of fishes.

B. DATA COLLECTION

The videos are acquired starting from the 1^{st} of February 2020 till the end of March 2020 using the experimental set-up described above. To maintain the randomness in the collected video data, 10 videos are collected in the month of February (5^{th}, 6^{th}, 7^{th}, 8^{th}, 9^{th} of February) and the rest 6 videos are collected in the month of March (10^{th}, 11^{th}, 12^{th} of March). Each day's data contains two video sequences of $20-30$ seconds, one collected in the morning and the other in the evening. Thus, the dataset contains a total of 16 video sequences. Further, it is split into training and testing with 8 sequences each. The data collection process came across several challenges such as reflections, appearance variations and lack of data during night. The details of these challenges has been discussed in the Appendix A.

C. DATA ANNOTATION

The ground truth annotation representing the trajectory of the fishes observed in the videos was manually determined by following the rules adopted from MOT16 [63]. The frames of the video sequences are carefully annotated, which enables accurate and precise computation of CLEAR metrics of tracking algorithms [64]. Fishes appearing on the acrylic aquarium panel are not annotated since these are just reflections. The frames of the video are resized to 800×800 during the annotation procedure. A total of more than 4000 frames are annotated. Table 1 provides the details regarding the experimental setup and dataset. The developed tracking algorithm should track the target object moment it is visible in the frame. Usually, when 10% of the target object is visible, the algorithm should start tracking the target object. It must be tracked until it is unable to locate the target object. This is the main criteria for evaluating the performance of the tracking

TABLE 1. Details of the experimental setup and dataset.

Parameter	Specification
Dimension of the aquarium	2.4m $\times$ 0.9m $\times$ 1.22m
Number of cameras	1
Total number of videos	16
Number of videos for training	8
Number of videos for testing	8
Total number of frames annotated	$\geq$ 4000
Resolution of the video	800 $\times$ 800
Frames per second	12
Number of fishes	8,9,10
Fish species	Sillago-sihama
Annotation format	Pascal VOC

algorithm. All the fishes, irrespective of size, are annotated. Every target object is annotated through the occlusion. More details on the data annotations and challenges in data annotations are provided in the Appendix A.

IV. PROPOSED METHOD

A. OVERVIEW OF THE PROPOSED TRACKING AND BEHAVIOUR MODEL

The proposed tracking algorithm follows the tracking by detection framework. It consists of two main steps, namely detection and data association. The first step identifies all the target objects (fishes in the current study) in the given frame (Section IV-B1). In the second step, these detections are associated with the corresponding tracklets. The proposed method performs the data association by constructing an assignment problem that can be subsequently solved using linear sum assignment. The similarity scores for the assignment problem is computed between every previous frame detections and the new detections. The study proposes a novel similarity score estimator that incorporates appearance, temporal, spatial and swim direction information (Section IV-B2). Further inspired from [65], the proposed method uses Gaussian kernels to refine the appearance and temporal similarity scores. It enables effective differentiation between non-similar fishes. In addition, the proposed method weighs the spatial similarity scores using the swim direction information. It further reduces the identity switches and facilitates accurate fish tracking. This simple integration of refining the similarity scores, using Gaussian kernels and swim direction assists effectively in the generation of the cost matrix, which is subsequently solved using linear sum assignment.

In addition to the tracking algorithm, the work proposes motion information based unsupervised continuous behaviour modelling of the fishes to detect unique patterns. To this end, the track information is obtained from the above-noted tracking method to identify the corresponding fish. Subsequently, the optical flow is computed in relation to the corresponding fish to capture motion information such as magnitude and swimming angle of each individual fish. The study then uses a mixture of Gaussian kernels to model the behavioural patterns of fishes. In addition, it formulates a continuous learning strategy to model new unique and evolving behaviour patterns of fish. Section IV-C discusses the details of the same.

B. TRACKING ALGORITHM

1) OBJECT DETECTOR

The proposed method uses Yolo V3 object detector algorithm to detect fish. Yolo V3 makes predictions at three different scales by downsampling the dimensions of the input image, thus making it ideal for detecting smaller target objects. In the present study, the algorithm is trained with 9 anchor boxes, three boxes per scale. It ensures the detection of all sizes of fish in a given frame. The study identifies the size of the anchor boxes by clustering the bounding box dimensions from the ground truth of the training set. To this end, the K-means clustering algorithm is used with $K = 9$. The value of $k = 9$ ensures that maximum detections are obtained and all the target objects are identified by the Yolo V3 algorithm. Let us represent i^{th} object (fish) in the k^{th} frame as d_i^k, where $i = 1, 2 \ldots m$ and m is the total number of fish detected in the frame. Likewise, j^{th} object (fish) in the $(k-1)^{th}$ frame is represented as d_j^{k-1}, where $j = 1, 2 \ldots n$ and n is the total number of fish detected in the frame.

2) GENERATION OF COST MATRIX

Cost matrix plays a significant role in assigning a trajectory to a particular object. The cost matrix contains the similarity score between the objects (fishes) of the previous and the current frame. Typically, the similarity scores are computed based on the appearance, motion cues, etc. [13], [22], [44]. However, these cues may not be adequate when the target objects are deformable, similar in appearance and change directions rapidly, such as fishes. It is seen from the literature that, Gaussian kernels are popularly employed to control the degree of similarity between feature vectors, pixels, etc [65]. The present study considers a similar approach to penalize the dissimilar matches and reduce ambiguity. Thus, the proposed method uses appearance, temporal and spatial cues to generate a cost matrix. Further, the appearance and temporal cues are weighed using Gaussian kernels and the spatial cues are weighed using swim direction information.

Appearance similarity score (F_a): The proposed method uses StackNet [44] to compute the appearance similarity score of the fishes. StackNet takes in a pair of objects $(d_j^{k-1}, d_i^k \quad \forall i, j)$ stacked one behind the other along the RGB color channels and produces an output representing the similarity score between the object pairs of two consecutive frames. The input dimension for the StackNet is $(h, w, 6)$, where h, w are the height and width of the image frame. The utilized StackNet incorporates Resnet-50 as the backbone CNN architecture for feature extraction. Finally, it uses a soft-max as the last layer, which provides the probability of image pairs being similar. For object pair d_j^{k-1}, d_i^k as the input, the appearance similarity score $F_a^k(i, j)$ is defined as output of the StackNet network. Further, this appearance

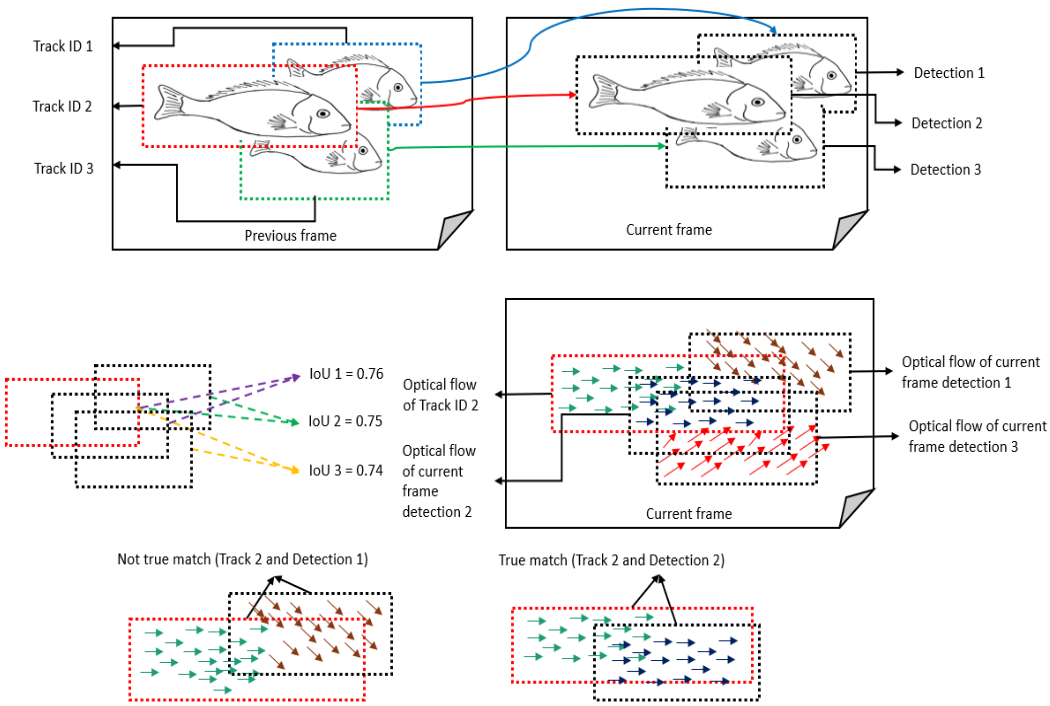


FIGURE 3. **Limitations of IoU based object tracking algorithm under multiple occlusions. IoU scores are closer to each other which increases the ambiguity during data association. In this work, swim direction is computed for every bounding box and the true matches are the bounding box with same swim direction.**

similarity score is scaled exponentially using the Gaussian kernel which is given as follows:

$$F'_a(i,j) = \exp\left(\frac{-(1-F_a(i,j))}{2\lambda^2}\right) \tag{1}$$

where, λ represents the model parameter that controls the degree of nearness. This appearance similar score would be high for a similar looking fish.

Temporal similarity score (F_t): The study attempts to capture the temporal similarity score of the fishes using the procedure described in [13]. For every object detections d_j^{k-1} in the previous frame, a prediction p_j^k of its possible location in the current frame is made, presuming the object is moving with constant velocity. This presumption is valid since the object's motion across consecutive frames is small. The predictions are computed with Kalman filter (constant velocity) assuming spatial location of the detection as the state space. Subsequently, for the current frame, true measurements $\hat{d}_i^k$ are obtained from the object detector. Thereupon, Mahalanobis distance is computed between p_j^k and $\hat{d}_i^k$, $\forall i,j$. This distance is defined as the temporal similarity score $F_t^k(i,j)$. Similar to appearance similarity score, the temporal similarity score is also scaled exponentially using the Gaussian kernel. It is given as follows:

$$F'_t(i,j) = \exp\left(\frac{-F_t(i,j)}{2\lambda^2}\right) \tag{2}$$

Unlike appearance similarity score, temporal similarity score produces a value closer to zero for similar fishes. As a result, in Equation 2, F_t is considered directly.

Spatial similarity score (F_s): The spatial similarity score is computed using an approach similar to that of an Intersection over Union (IoU) based tracker [23]. The IoU based tracker computes IoU metric between all the pairs of object detection in the two consecutive frames. The use of IoU metric ensures the detection are associated with a trajectory based on the spatial overlap between detection and trajectory. The spatial overlap information reduces the number of identity switches [66]. In this work, the spatial similarity score between the object detection $d_j^{k-1} d_i^k$ is defined:

$$F_s^k(i,j) = \frac{Area(d_j^{k-1}) \cap Area(d_i^k)}{Area(d_j^{k-1}) \cup Area(d_i^k)} \tag{3}$$

where, $Area(d_i^k)$ is the area of the bounding box representing the object detection d_i^k and $Area(d_j^{k-1})$ is the area of the bounding box representing the object detection d_j^{k-1}. Note that Spatial similarity score would be higher if the detections ($d_j^{k-1} d_i^k$) corresponds to the same trajectory.

In case of a highly occluded or crowded scene, tracking based on spatial information (IoU) would result in multiple identity switch. Figure 3 shows an example scenario. Given the trajectories till previous frame, and detections in the current frame, the objective is to assign current frame detections to the trajectories. However, when there is occlusion among

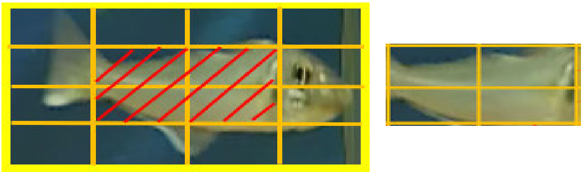


FIGURE 4. Sample detection divided into 4 × 4 grids. optical flow is computed for the center 2 × 2 grids indicated by red shaded region.

multiple (3 or more) fishes, the computed IoU scores can be very close. This scenario is common among the fishes as they swim together. Thus generated cost matrix may not be explicit enough to be solved using linear sum assignment. Consequently, it leads to identity switches reducing the accuracy of the tracking algorithm. To overcome this challenge, this work incorporates swim direction information. It identifies the true match while penalising the non-true matches based on the swim direction of fishes. To this end, the proposed methodology first computes the average fish swim direction for *each* detection using optical flow. However, instead of utilizing the optical flow of entire image, a masked optical flow is computed which focuses on the motion information of each individual fish. Let us represent optical flow vector for the entire image as $OF = [u_i^k, v_i^k]$. The bounding box representing each detection is further sub-divided into 16 grids as shown in Figure 4. A binary mask M_i^k where the pixels inside the 4 inner grids (shaded in Red) are represented as 1, while other pixels as 0. The masked optical flow is then defined as $\hat{OF}_i^k = OF \times M_i^k$. This masked optical flow vector $\hat{OF}_i^k = [\hat{u}_i^k, \hat{v}_i^k]$ is then utilized to compute the average displacement. Specifically, for a given detection d_i^k, the average displacement direction (z_i^k) of this detection is computed from the masked optical flow vector $\hat{OF}_i^k$ as follows:

$$z_i^k = \frac{2 \times 2}{h_i^k w_i^k} \sum_x \sum_y (tan^{-1}(\hat{v}_i^k / \hat{u}_i^k)) \tag{4}$$

where h_i^k, w_i^k are the height and width of the bounding box representing the detection d_i^k. Similarly, the average displacement direction (z_j^{k-1}) for the detection (d_j^{k-1}) is computed as:

$$z_j^{k-1} = \frac{2 \times 2}{h_j^{k-1} w_j^{k-1}} \sum_x \sum_y (tan^{-1}(\hat{v}_j^{k-1} / \hat{u}_j^{k-1})) \tag{5}$$

To compare the average displacement direction of two detections, this work computes the cosine of the absolute difference between the two directions. Specially, $\eta(i, j)$ can be computed as

$$\eta(i, j) = \cos \mid z_i^k - z_j^{k-1} \mid \tag{6}$$

If the two detections represents the same fish in two consecutive frames, the average displacement direction would be identical, which will result in $\eta = 1$. This formulation penalizes the detections where the average displacement direction is not identical for two consecutive frames, which is the case

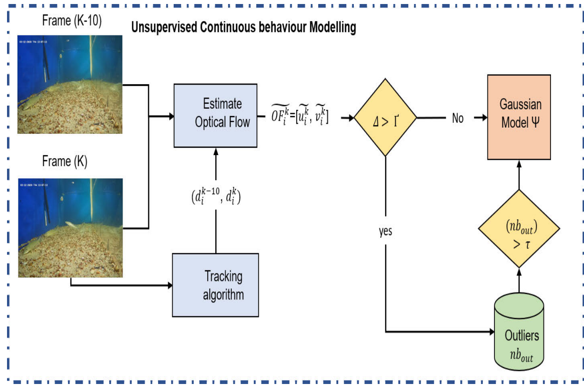


FIGURE 5. Overview of the proposed unsupervised continuous behaviour modelling.

for detection corresponding to two different fishes (Figure 3). Subsequently, the spatial similarity score (F_s) is weighted by η to incorporate swim direction information in the similarity score. Equation 7 gives the final computed spatial information of the proposed method.

$$F'_s(i, j) = F_s(i, j) \times \eta(i, j) \tag{7}$$

Thus, the final cost matrix $C(i, j)$ of the proposed method is computed as follows:

$$C(i, j) = (\alpha F'_a(i, j) + (1 - \alpha) F'_t(i, j))(F'_s(i, j)) \tag{8}$$

where, α is the hyper-parameter that decides the importance of appearance and temporal information. Note that the appearance and temporal similarity score is weighted by the proposed swim direction based spatial similarity score. Finally, the cost matrix computed using Equation 8 is solved using linear sum assignment. The novel integration of Gaussian kernels and swim direction of fishes in computing the cost matrix enables accurate data association, thereby increasing the performance of the tracking algorithm.

C. UNSUPERVISED CONTINUOUS FISH BEHAVIOUR MODELLING

The present study proposes a novel unsupervised fish behaviour modelling strategy that continually categorises the evolving behaviour. The behaviour is determined from motion information and the trajectories of the fishes. The tracking algorithm described in Section IV-B2 is used to identify two detection d_i^{k-10} and d_i^k corresponding to the same fish in k^{th} and $k - 10^{th}$ frame. Instead of consecutive frame, this work utilizes a difference of ten frames to capture sufficient motion information. Subsequently, the average displacement between these two detections are computed from the masked optical flow vector. Let us represent $\overline{OF} = [\overline{u}, \overline{v}]$ as the optical flow vector computed from k^{th} and $(k - 10)^{th}$ frames. The masked optical flow $\tilde{OF}_i^k = [\tilde{u}_i^k, \tilde{v}_i^k]$ is then computed as $\tilde{OF}_i^k = \overline{OF} \times M_i^k$, where M_i^k is the mask defined in Section IV-B2 (Figure 4). The average displacement $\tilde{\delta}_i^k$

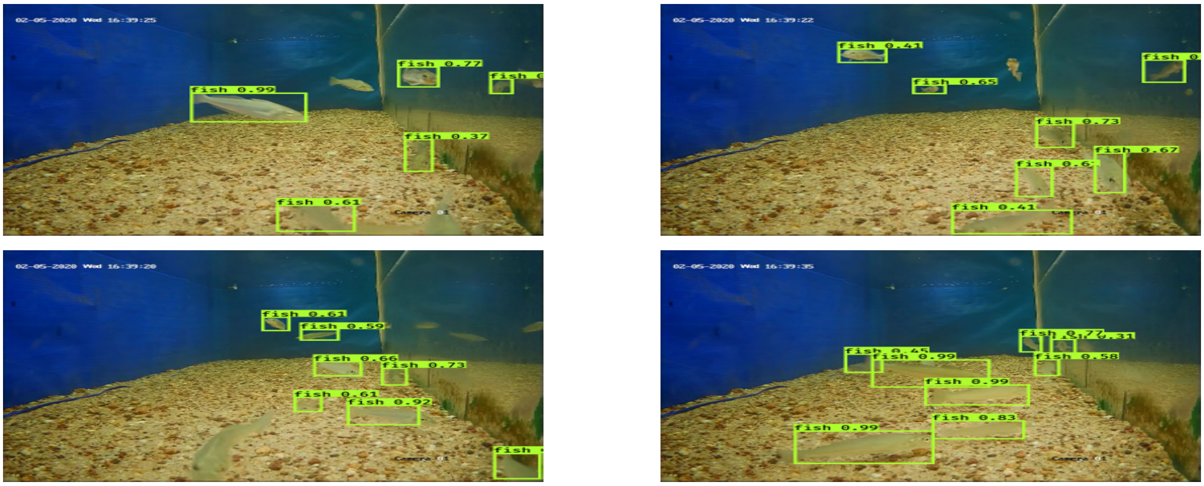


FIGURE 6. Sample results of Yolo V3 object detector on the proposed dataset.

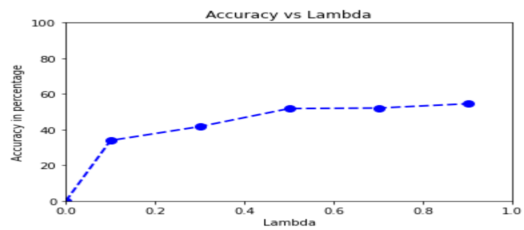
and direction $\tilde{z}_i^k$ for a particular fish is then computed as

$$\tilde{\delta}_i^k = \|\tilde{OF}_i^k\| \tag{9}$$

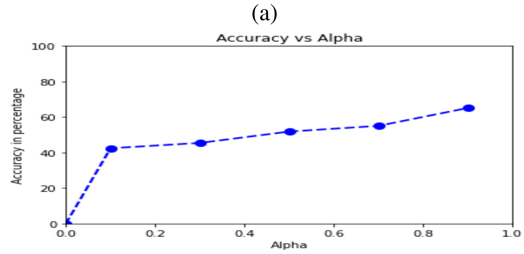
$$\tilde{z}_i^k = \frac{2 \times 2}{h_i^k w_i^k} \sum_x \sum_y (tan^{-1}(\tilde{v}_i^k / \tilde{u}_i^k)) \tag{10}$$

where $\|\tilde{OF}\|$ represents the magnitude of the optical flow vector $\tilde{OF}$.

Subsequently, a Gaussian Mixture Model (GMM) is estimated assuming displacement and direction features of each fish as the data points generated from mixture of Gaussian distribution. The number of Gaussian components is estimated using the Silhouette method. The number of components identified represents the different fish behaviour. However, it is possible that the behaviour not observed earlier is visible due to changes in the environmental conditions (water quality parameters). Therefore, this work proposes to continuous behaviour modelling which updates the GMM regularly. Let us suppose that Ψ_{init} is the initial GMM estimated using displacement and direction features extracted from the first few frames of the video. At the same time, the proposed method identifies a set of possible outliers. A data point is considered as a possible outlier if the distance (Δ) in the feature space between the mean for each cluster and the data point is above a threshold (Γ). The Γ regulates the sensitivity of the developed model. For subsequent frames from the video, new data points are examined for possible outliers. If the number of possible outliers (nb_{out}) increases beyond a threshold τ, the Gaussian Mixture Model is re-estimated (Ψ_{up}) for the new set of data points. The process of evaluating the new data points and updating the number of segments of GMM ensures the unobserved behaviour is represented as one of segments of GMM. The proposed method can also be extended to model the behaviour of single fish. By considering the motion information of single fish, the behaviour of single fish can be modelled explicitly. The overview of the proposed unsupervised continual behaviour modelling is shown in Figure 5.



(a)



(b)

FIGURE 7. (a)Variation of tracking accuracy for different values of lambda.(b)Variation of tracking accuracy for different values of alpha.

V. RESULTS AND DISCUSSION

A. IMPLEMENTATION DETAILS

The proposed tracking and behaviour modelling approach is implemented on Intel(R) Core(TM) i7-8550U CPU and NVIDIA 1650 GPU. The proposed tracking method is evaluated on the SSVid dataset (Section III), and compared with the existing tracking methods. For training the StackNet module (Section IV-B2), a subset of images from SSVid dataset is utilized. This subset contains 250 selected image (frame) pairs from SSVid dataset. Further, the bounding boxes representing the fishes are manually cropped for training StackNet. The StackNet model, trained with SGD optimizer and the learning

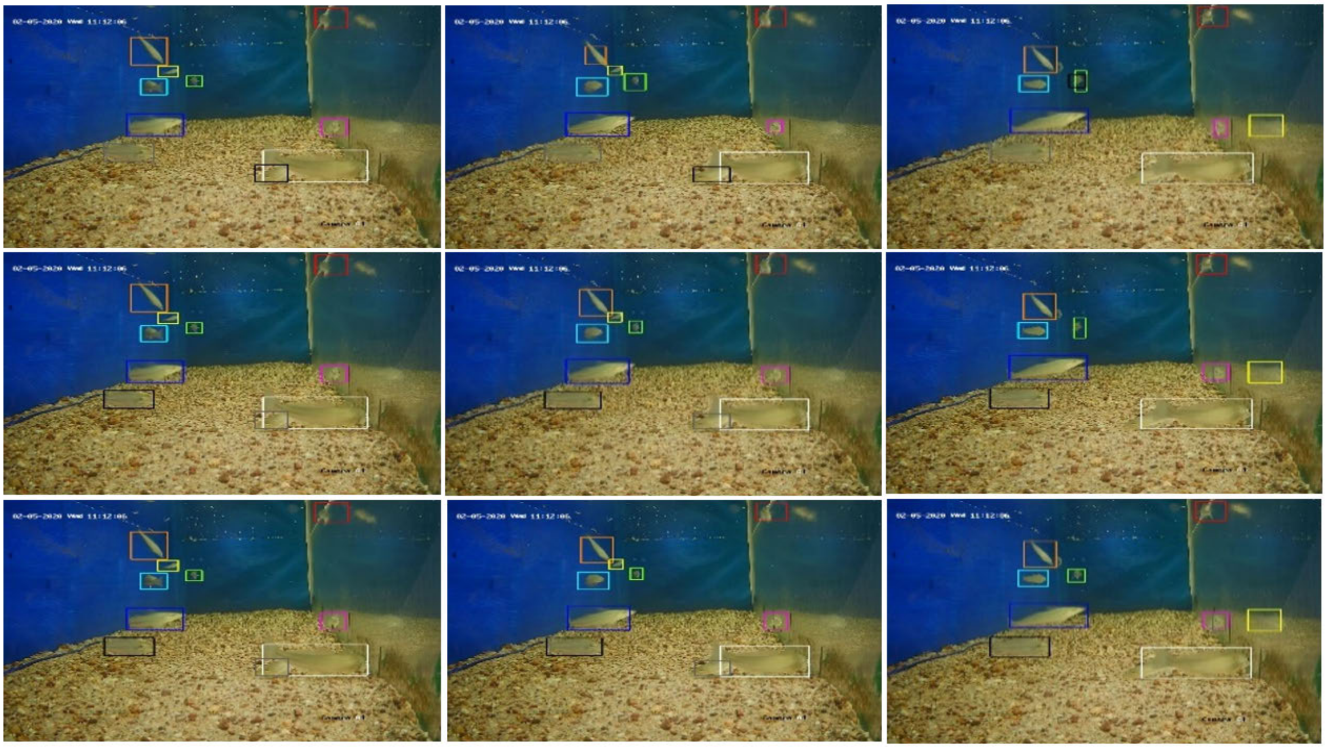


FIGURE 8. The results of tracking algorithms for 3 consecutive frames are shown. The results of Deep SORT, High speed IOU and the proposed tracking algorithm are given in the first, second and third row respectively. The bounding box corresponding to each individual trajectory are shown in different colors.

rate initialized to 0.0001 has softmax as the last layer. For training the object detector module (Yolo V3), a subset of 400 images with multiple fishes are extracted from the SSVid dataset. The bounding boxes representing these fishes are manually annotated for training the object detector module. The sample results of Yolo V3 object detector on the proposed dataset is shown in Figure 6.

The proposed tracking model contains two hyper-parameters namely: λ and α (Equations 1,2,8). λ controls the degree of similarity nearness of two target fish. α determines the preference given to the appearance and motion model. Initially, having set α to 0.5, λ is estimated. Thus, with equal preference to the appearance and motion model, the best value of λ is estimated. Figure 7(a) shows the multiple object tracking accuracy (MOTA) of the proposed model for various values of λ. It is observed from the Figure 7(a) that, MOTA is highest for $\lambda = 0.9$. Thus, in the present study, the value λ is set to 0.9. Henceforth, the value of α is estimated, with λ set to 0.9. Figure 7(b) shows the variation of MOTA for various values of α. It is seen that, the accuracy of tracking is highest for $alpha = 0.9$. Accordingly, the proposed method gives 90% importance to the appearance similarity score and 10% to the motion similarity score.

B. PERFORMANCE METRICS

Classification of Events, Activities and Relationships (CLEAR) metrics [64] have been used for performance evaluation of the proposed tracking algorithm. Essential parameters which quantify the performance of the tracking algorithm such as Multiple Object Tracking Accuracy (MOTA), Multiple Object Tracing Precision (MOTP), Identity Switch (IDSW), Identification F1 score (IDF1), Mostly Tracked (MT), Partially Tracked(PT) and Mostly Lost (ML) are computed for the proposed method on SSVid dataset and compated with existing methods. MOTA provides an accurate measure of the tracker's performance in detecting target objects and maintaining the trajectories. MOTP indicates the capability of the tracker to estimate the position of the target object. IDSW quantifies the number of times the algorithm shifted the identity from one target object to another target object. MT, PT and ML metrics quantify the duration up to which the tracker algorithm can track the target object.

C. PERFORMANCE ANALYSIS OF THE PROPOSED TRACKING ALGORITHM

Firstly, the study investigates the effectiveness of handcrafted and CNN-based features in generating the similarity scores between the target objects. To this end, it used HOG and Local Binary Patterns for the handcrafted features (TBD-HC) and VGG-16 for the CNN features (TBD-VGG). The present study collects these features between d_j^{k-1} and d_i^k. Subsequently, it computes the cosine similarity between the two feature vectors to compute the resemblances. From Table 2, it is seen that TBD-VGG outperforms TBD-HC. It indicates

TABLE 2. Tracking performance on the developed dataset of various tracking algorithms.

Tracking algorithms	IDF1	MOTA	MOTP	IDSW	MT	PT	ML
TBD-HC	50.82%	37.90%	0.58	351	65.20%	27.70%	7.10%
TBD-VGG	50.22%	42.20%	0.58	346	63.80%	27.70%	8.50%
High speed IoU [23]	**56.20%**	62.40%	0.59	127	81.4%	15.02%	1.58%
Tracktor [22]	50.10%	40.22%	0.58	355	68.05%	**30.50%**	1.45%
Deep SORT [13]	55.20%	47.60%	0.62	261	69.40%	25%	5.60%
Proposed	54.7%	**65.07%**	**0.63**	**110**	**84.72%**	13.80%	1.48%

TABLE 3. Tracking performance on the developed dataset of the proposed model with different configurations of features.

Features	IDF1	MOTA	MOTP	IDSW	MT	PT	ML
Appearance + Temporal	**55.20%**	47.60%	0.62	261	69.40%	25%	5.60%
Appearance + Spatial	52.47%	45.27%	0.56	331	64.92%	31.05%	4.23%
Temporal + Spatial	51.34%	43.63%	0.58	357	61.01%	31.31%	7.68%
Appearance + Temporal (with Gaussian kernel)	54.82%	52.78%	0.62	223	73.7%	23,5%	2.8%
Appearance + Temporal + Spatial (with Gaussian Kernel)	53.8%	57%	0.62	175	77.5%	20.72%	1.78%
Appearance + Temporal + Spatial (with Gaussian kernel and swim direction)	54.7%	**65.07%**	**0.63**	**110**	**84.72%**	13.80%	1.48%

that the CNN-based feature extractor could distinguish better between the target objects. In other words, it generates better similarity scores than the handcrafted features. Thus, the proposed method uses a CNN-based (StackNet) feature extractor while identifying the appearance similarity scores.

1) COMPARATIVE ANALYSIS

Besides, the proposed method is also compared with the existing methods viz, High Speed IOU [23], Tracktor [22], and DeepSort [13]. From Table 2 it is observed that the proposed method achieved a MOTA of 65.07%, which is significantly greater than the other existing methods. It signifies that the algorithm has minimum false positives, mismatches and misses. Also, High-speed IOU performs competitively with the proposed method (62.40%). These results offer compelling evidence for the effectiveness of spatial overlap information in improving the performance of tracking algorithms. Further, the inclusion of swim direction information in the proposed method enhanced its performance by 2.6%. As anticipated, the study shows that the Deep SORT tracking algorithm which incorporates motion and appearance information, performed poorly. It is significant as it highlights the effectiveness of spatial and swim direction information for tracking fishes.

The MOTP value for all tracking algorithms varies between 0.58 to 0.63. As all tracking algorithms use the same detection algorithm (Yolo V3), MOTP values are almost similar. MOTP metric quantifies the capacity of the tracking algorithm to estimate the next position of the target object. This value of the proposed and Deep SORT methods is high compared to the other algorithms as it uses Kalman predictions along with Yolo V3 for estimating the position of the target objects. Further, the proposed method proves the effectiveness of Gaussian kernels in computing the degree of motion similarity score.

One of the major challenges of tracking multiple fishes of the same species is the identity switch due to similar appearance. The challenge is further escalated as *Sillago Sihama* typically travels in a group. The identity switch (IDSW) is very low for the proposed method, demonstrating the better performance of the data association step. It highlights the effectiveness of motion, appearance, spatial and swim direction information in associating the data for tracking fishes. Even though the Deep SORT tracking algorithm utilizes a motion and appearance model, it suffers from an identity switch problem as highlighted by the high IDSW. These findings reinforce the in-effectiveness of motion and appearance alone features in addressing the identity switch problem when tracking multiple fishes. Also, High-speed IOU based tracking algorithm, which only uses spatial information during data association, performs comparatively better than the other existing algorithms. However, the proposed method outperforms the High-speed IOU based tracking algorithm, highlighting the usefulness of Gaussian kernels and swim direction information used in the algorithm.

The minimum MT value in Table 2 is 63.8%. It indicates that the compared algorithms can track 63% of the target object for more than 80% of its life span. It is also interesting to note that the ML value for the compared algorithms is less than 9%. This value is satisfactory for the proposed and compared methods. However, improving the data association step of the tracking algorithms can further reduce this. Figure 8 shows the results of proposed and existing tracking methods for 3 consecutive frames. It shows that the identity switches are lower in the proposed method as compared to other methods (Trajectory corresponding to the bounding box shown in Orange, Yellow and Green in Figure 8).

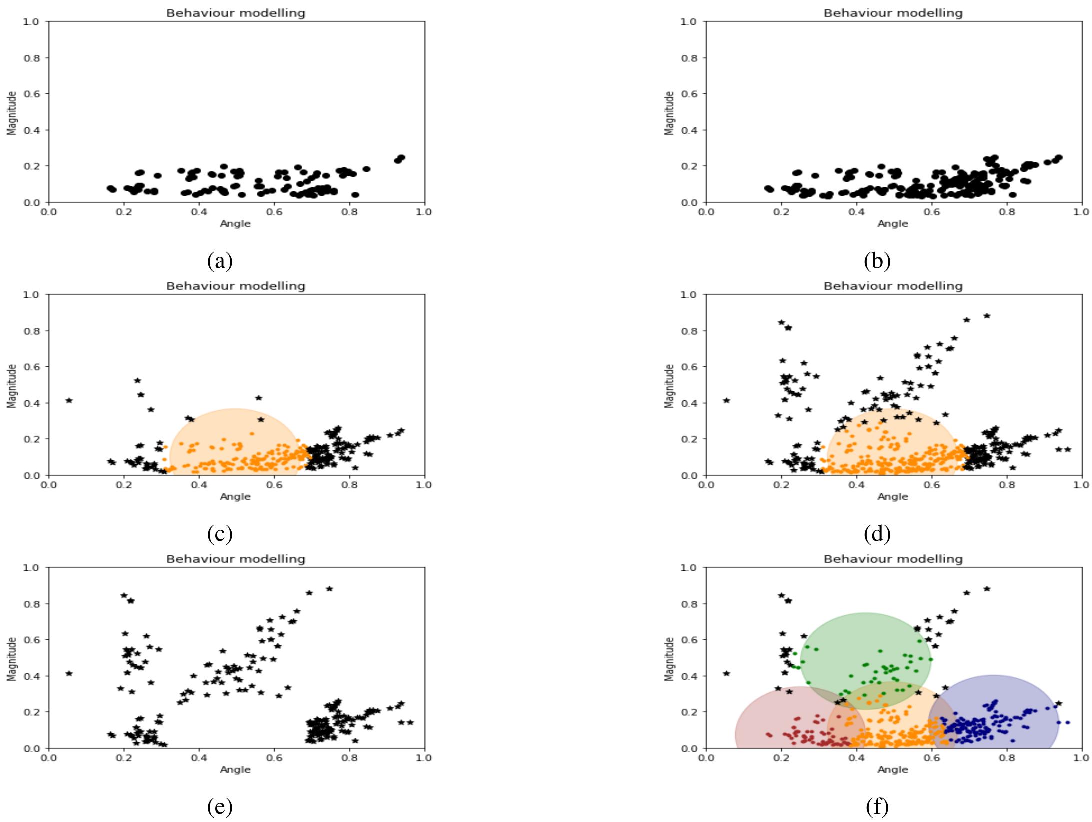


FIGURE 9. **Sample results of the proposed continual behaviour modelling strategy. (a), (b) Sample data points. (c) Gaussian modelling of data points at time $t + n$. A pattern identified is shown in orange colour cluster. (d) Newer data points along with the older data points are modeled. (e) All the unique patterns are shown in black data points. (f) Unique patterns are modelled using Gaussian. Three new cluster are identified by the proposed model as shown in green, red and blue coloured clusters.**

2) ABLATION STUDY

The study also investigates the effectiveness of different features for tracking fishes. The Table 3 shows the performance of tracking algorithm for various combinations of features. It is observed that, without Gaussian kernels, the model performance in terms of accuracy is well below 50%. This is indicated by the first three rows of Table 3. Also, it can be seen that there is increase in identity switches. However, there is considerable increase in the performance in terms of IDSW and MOTA by the inclusion of Gaussian kernels to appearance and temporal features. This highlights the effectiveness of Gaussian kernels in generating the cost matrix. The performance further improved with the insertion of spatial features. Spatial features improved the models capability to localize the target object more accurately. Finally, the addition of swim direction to spatial features further improved the models performance. Thus, the study clearly highlight the significance of Gaussian kernels and swim direction in refining the similarity scores for generating effective cost matrix.

Thus, the experimental results presented in this work demonstrates that deep features are more reliable in describing the target object in the dataset than the hand-crafted features. Explicit description of the target object can considerably improve the data association process, thereby increasing the performance of the tracking algorithms. The proposed method outperforms the conventional tracking algorithms regarding the accuracy, precision and identity switches. Taken together, these findings highlight the impact of swim direction and Gaussian kernels on the refinement of similarity scores in tracking fishes.

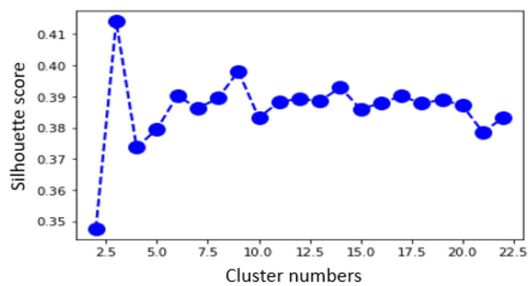


FIGURE 10. **Plot of Silhouette scores for different number of Gaussian clusters.**

D. PERFORMANCE ANALYSIS OF CONTINUAL BEHAVIOUR MODELLING

Meticulous observation of the fish movements in the aquarium reveals several unique swimming patterns. The experimental setup effectively captures these patterns. These patterns are specific to certain conditions in the ecosystem.

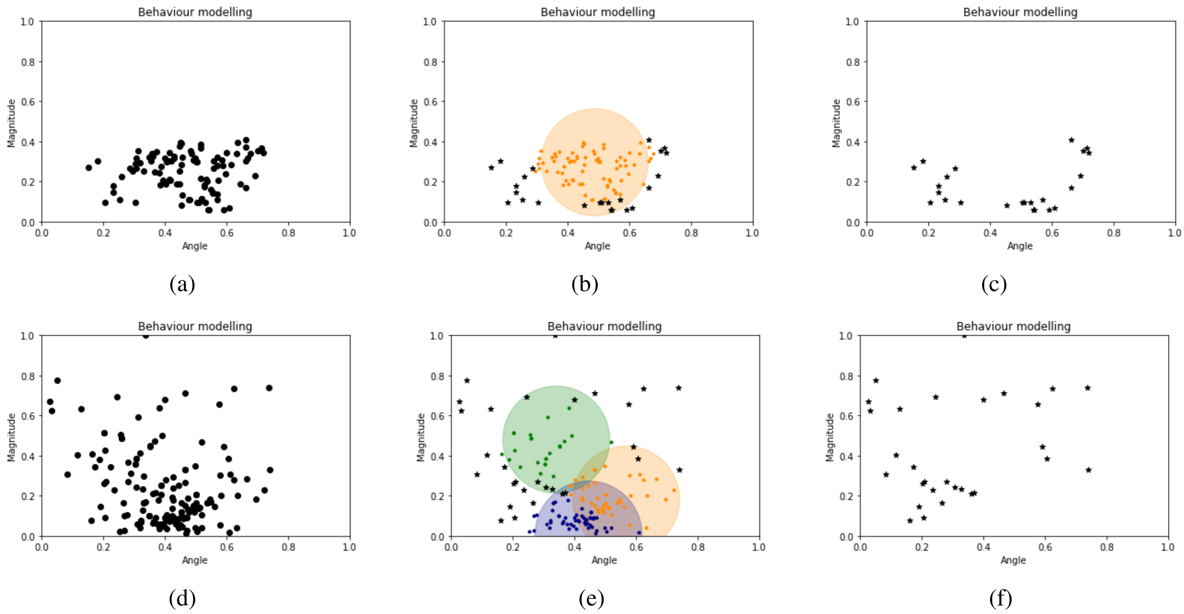


FIGURE 11. Sample results of the proposed continual behaviour modelling strategy for single fish. (a), (b), (c) shows the data points, unique patterns and outlier for one particular fish. Likewise, (d), (e), (f) shows the data points, unique patterns and outlier for another fish.

Thus, understanding these patterns can provide greater insights into the species under study.

The reliability of the proposed continual behaviour modelling strategy is substantiated on the dataset. Figure 9 shows the sample results of the continuous behaviour modelling on the dataset. Figure 9a and Figure 9b shows the sample data points collected from the video. The inference from the distribution of the data points is that the fishes are moving slowly as the magnitude is well below 0.3. The distributions extend from 0.1 to 0.9 along the x-axis implying fishes are swimming in all directions. Thus, these data points represent the normal motion of the fishes. Figure 9c shows the estimation of a Gaussian kernel on the data points. There are some outliers not grouped into the cluster as their distance to the cluster mean is greater than Γ. Figure 9d shows the growth of the cluster with the incoming new data points. Here, it is evident that fishes are exhibiting frantic movements as there are data points with magnitude above 0.5. Figure 9e shows only the unique patterns for a clear understanding of the modelling strategy. Subsequently, the number of Gaussian kernels required to model the data is estimated using the Silhouette method. Figure 10 shows the plot of Silhouette scores for a different number of clusters. Accordingly, the required number of Gaussian kernels are fit to these new unique patterns as illustrated in Figure 9f. The model identified 3 new cluster points highlighted in red, green and blue coloured clusters. Thus, the proposed unsupervised behaviour modelling strategy models the fish behaviour continuously. The Figure 11 shows the sample results for single fish. The Figure 11 presents 2 separate cases, with each row showing the motion information, behaviour clusters and outliers for single fish. Specifically, the proposed method identifies all the behaviour patterns as unique patterns and clusters them for further analysis by the domain experts.

E. AQUACULTURE VIEWPOINT

The development of a decision support system for the detection of fish behaviour requires close observation of the movements of each fish. To this end, tracking fishes provides its trajectories information for fish behaviour analysis. Consequently, as the accuracy of the tracking algorithms increases, the accuracy of behaviour analysis via tracking increases. Thus to accurately study the fish behaviour patterns, one must maintain the trajectories for more than 80% of the video duration. In this regard, the MT value of the tracking algorithm has to be as high as possible. It is preferred to have a small ID switch value as it can trigger the pattern analysis model. For example, an ID switch that occurs between a resting and a moving fish provides wrong information on the behaviour patterns. Thus, a robust tracking algorithm requires high MOTA, MOTP, IDF1, MT and low IDSW metrics.

The pattern analysis model thus developed via tracking is beneficiary for aquaculture. As an example, the proposed behaviour detection method can easily detect fish's swimming frantically. This pattern can be detected if data points are closer to 1 along the y-axis. Thus, greater insight into the fish behaviours obtained through the behaviour analysis model aids the farmers to understand the required ecosystem for aquaculture. Further, they benefit from a Decision Support System to effectively maintain their aquaculture systems. For example, they can plan appropriate preventive measures after witnessing anomalous behaviour. Thus, the mortality rate of the fish can be reduced, thereby increasing the culture harvest.

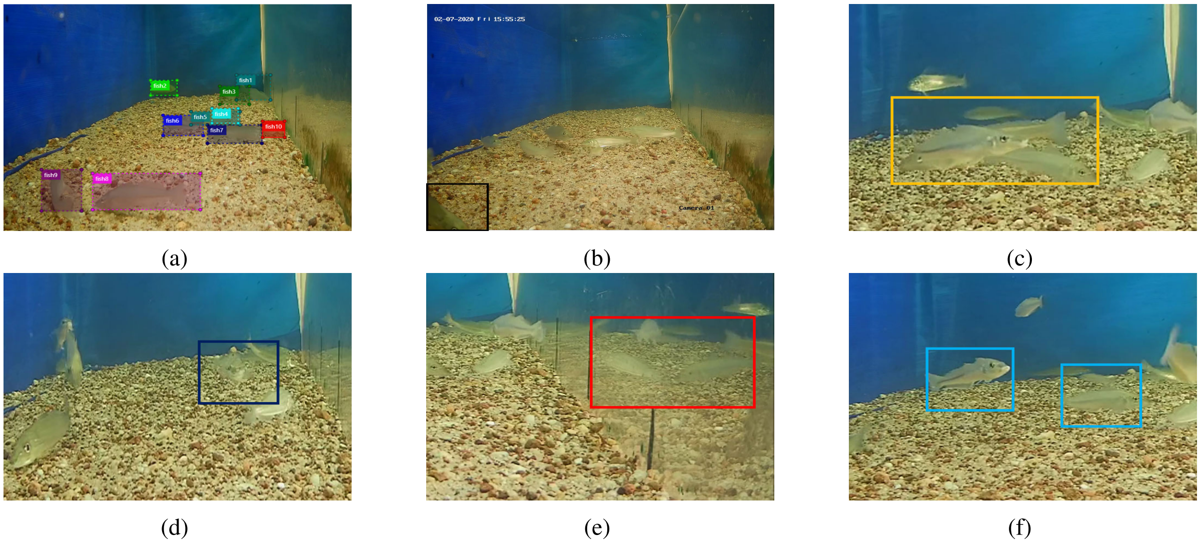


FIGURE 12. **Sample images showing the various challenges encountered during annotation. (a) Sample annotation in a frame. (b) Frame illustrating the cropped fish scenario – Black box. (c)Frame illustrating the occlusion among the fish scenario – Orange box. (d) Frame illustrating the camouflaged fish scenario – Dark blue box. (e) Frame illustrating the reflection of fishes on the acrylic panel – Red box. (f) Frame illustrating the similar appearance of fish scenario – Light blue box. Note that frames shown in (b) to (f) are enlarged for better visualization.**

VI. CONCLUSION

This paper proposes multiple fish tracking and unsupervised fish behaviour approach for an economically important fish species, namely *Sillago Sihama*. *Sillago Sihama* typically travels in a group which increases the probability of identity switches due to identical appearance. The proposed fish tracking method incorporates swim direction in addition to the appearance and temporal information in a tracking by detection framework. Further, the proposed tracking algorithm uses Gaussian Kernels for computing the weighted appearance and temporal scores. The use of swim direction information significantly reduces the number of identity switches. The proposed method is evaluated on a novel Sillago Sihama Vid Dataset, which consists of videos acquired of multiple *Sillago Sihama* grown in a large aquarium. The experimental set-up emulated the natural eco-system in the aquarium. The proposed tracking algorithm has a MOTA of 65.07%, MOTP of 0.63, IDSW of 110 and MT of 84.72%, which are higher than the other existing tracking algorithms. The results highlight the influence of the Gaussian kernel in refining the appearance and motion similarity scores and swim direction information in weighing spatial similarity scores.

This work also proposes an unsupervised fish behaviour using Gaussian Mixture Model (GMM). The optical flow and the fish trajectory information is utilized to extract motion features for *each* individual fish present in the video. Subsequently, these features are then grouped using Gaussian Mixture Model. Besides, the number of components (groups) of the GMM are updated if the number of possible outliers increases over a certain threshold. This continual monitoring of the fish behaviour ensures that a behaviour not observed earlier will also be modelled in the proposed method. It is fundamental to note that, the proposed behaviour modelling strategy does not require training and labelled data. Further, the method is unsupervised making it ideal for modelling the behaviours of all fish species.

The proposed fish tracking and behaviour approach for *Sillago Sihama* will be extremely useful in studying fish behaviour of an economically important and high nutrient content fish. The proposed approach will also be helpful in identifying any anomalous fish behaviour for the fishes cultured in cage and pond aquaculture systems. This enhances the productivity per given area by reducing the the mortality of fishes. Besides, the standard fish tracking dataset (Sillago Sihama Vid) created as part of this work, will further drive the research in tracking multiple fishes.

APPENDIX A
SILLAGO SIHAMA VID (SSVID): FISH TRACKING DATASET

The details of the dataset generation is provided in the current section.

A. CHALLENGES IN DATA COLLECTION

Several challenges were encountered in the data collection process, such as a reflection on the panel, lack of data collection during the night and appearance variations of the target object. The current paragraph discusses these issues in detail.

The study uses an aquarium constructed using acrylic panels. It ensures the durability of the aquarium and the ease of viewing the fishes from outside. Also, the toughness of the acrylic panel helps for long term experimentation. However, this resulted in the formation of reflections on the panel. These reflections are avoided on the two sides of the aquarium panel by arranging HDPE sheets. The other two sides without HDPE sheets generate reflections of fish from the camera's point of view. These reflections can be deceiving for the computer vision-based algorithms and cause false triggering of object detection algorithms. Developed fish tracking and behaviour identification algorithms need to address this problems.

It is significant that fishes in the experimental setup experience natural day and night cycles. Exposing fishes to light 24 hours a day prevents capturing the normal behaviour of fishes shown when exposed to the day and night cycles. The fishes will be under stress if exposed to light 24 hours a day, eventually leading to their death. To avoid this situation, fishes are exposed to light for only 12 hours a day, during which the data have been captured. However, this limits the data collection process to daytime only. As a result, the dataset does not capture the unique patterns exhibited during night-time by the Sillago Sihama. The camera is fixed at one end of the aquarium, as shown in Figure 2. Therefore, fishes closer to the camera appear large, and the same fish appears smaller when it is at the other end of the aquarium. This fact also limits the usage of the proposed dataset for studying the growth rate of fishes.

B. DATA ANNOTATION

All frames of a video sequence are stored in PNG format and named sequentially to a 10-digit file name. (e.g., frame_0001.png). An XML file is generated for each frame of a video sequence. The imaginary boxes, outlining the object of interest defining its center with x and y coordinates, are referred to as bounding boxes. The bounding box is manually placed around the target object as accurately as possible. The bounding box contains all the pixels belonging to the target object. Additionally, the bounding box is as concise as possible for avoiding unnecessary pixels inside the bounding box. Furthermore, if the fishes are occluded, then the extent of the fish is estimated based on the factors such as size, shadow, previous frames and other cues. For the fishes cropped from the frame, only the visible portion of the fish is annotated.

C. CHALLENGES IN DATA COLLECTION

Data annotation encountered several challenges, the main being occlusions (Figure 12 (c)). It is observed that on several occasions in the videos, multiple fishes were occluding. There are a couple of reasons for severe occurrences of occlusions. The fishes considered in the present case being bottom-dwelling fishes cover lower 60% of the aquarium resulting in high occlusions. Further, the smaller fishes usually follow the bigger fishes increasing the occlusions. Yet, the dataset captures an accurate trajectory of each fish. Also, from the camera's point of view, reflections of fishes are formed on one of the aquarium panels. This is shown in Figure 12 (e). It proved challenging in differentiating between the true images and the reflections. The fishes being similar in appearance and colour, it is difficult to differentiate each fish. It has been a challenge to maintain the unique fish ID throughout the video sequence Figure 12 (f). It should also be noted that *Sillago Sihama* fish has the colour similar to sand. The camouflage nature of these fishes, makes it difficult to identify them in certain frames (Figure 12 (d)) even after maintaining the high resolution of the videos. The camera covers 98% of the aquarium. However, there is a small patch of the aquarium which is not covered by the camera. When the fishes go to these areas, they are lost completely from the view of the camera. This results in cropping of the fishes in the frame as shown in Figure 12 (b).

ACKNOWLEDGMENT

The authors would like to thank the Department of Science and Technology (DST), Government of India, under which the work has been carried out. Indian patent has been filed on the tracking algorithm (Application No: 202241010629).

REFERENCES

[1] M. Føre, K. Frank, T. Norton, E. Svendsen, J. A. Alfredsen, T. Dempster, H. Eguiraun, W. Watson, A. Stahl, L. M. Sunde, C. Schellewald, K. R. Skøien, M. O. Alver, and D. Berckmans, ''Precision fish farming: A new framework to improve production in aquaculture,'' *Biosyst. Eng.*, vol. 173, pp. 176–193, Sep. 2018.

[2] K. R. S. R. Raju and G. H. K. Varma, ''Knowledge based real time monitoring system for aquaculture using IoT,'' in *Proc. IEEE 7th Int. Advance Comput. Conf. (IACC)*, Jan. 2017, pp. 318–321.

[3] C. Encinas, E. Ruiz, J. Cortez, and A. Espinoza, ''Design and implementation of a distributed IoT system for the monitoring of water quality in aquaculture,'' in *Proc. Wireless Telecommun. Symp. (WTS)*, Apr. 2017, pp. 1–7.

[4] C. M. Jolly and H. A. Clonts, *Economics of Aquaculture*. Boca Raton, FL, USA: CRC Press, 1993.

[5] P. Edwards, ''Traditional Asian aquaculture,'' in *New Technologies in Aquaculture*. Amsterdam, The Netherlands: Elsevier, 2009, pp. 1029–1063.

[6] U. Jadhav, *Aquaculture Technology and Environment*. Delhi, India: PHI Learning Pvt. Ltd., 2009.

[7] M. B. Asif, F. I. Hai, W. E. Price, and L. D. Nghiem, ''Impact of pharmaceutically active compounds in marine environment on aquaculture,'' in *Sustainable Aquaculture*. Cham, Switzerland: Springer, 2018, pp. 265–299.

[8] M. Radolko, F. Farhadifard, and U. von Lukas, ''Change detection and blob tracking of fish in underwater scenarios,'' in *Proc. Int. Joint Conf. Comput. Vis., Imag. Comput. Graph.*, Cham, Switzerland: Springer, 2017, pp. 279–304.

[9] C. Beyan and R. B. Fisher, ''Detecting abnormal fish trajectories using clustered and labeled data,'' in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 1476–1480.

[10] E. Lantsova, T. Voitiuk, T. Zudilova, and A. Kaarna, ''Using low-quality video sequences for fish detection and tracking,'' in *Proc. SAI Comput. Conf. (SAI)*, Jul. 2016, pp. 426–433.

[11] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, ''ECO: Efficient convolution operators for tracking,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6638–6646.

[12] D. H. Lee, ''CNN-based single object detection and tracking in videos and its application to drone detection,'' *Multimedia Tools Appl.*, vol. 80, pp. 34237–34248, Oct. 2021.

[13] N. Wojke, A. Bewley, and D. Paulus, ''Simple online and realtime tracking with a deep association metric,'' in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3645–3649.

[14] H. AlZu'bi, W. Al-Nuaimy, J. Buckley, L. Sneddon, and I. Young, ''Real-time 3D fish tracking and behaviour analysis,'' in *Proc. IEEE Jordan Conf. Appl. Electr. Eng. Comput. Technol. (AEECT)*, Nov. 2015, pp. 1–5.

[15] Y.-X. Bai, S.-H. Zhang, Z. Fan, X.-Y. Liu, X. Zhao, X.-Z. Feng, and M.-Z. Sun, ''Automatic multiple zebrafish tracking based on improved HOG features,'' *Sci. Rep.*, vol. 8, no. 1, pp. 1–14, Jul. 2018.

[16] A. Faro, D. Giordano, S. Palazzo, and C. Spampinato, ''Fish4knowledge deliverable D1.1 fish detection and tracking,'' Fish4Knowledge, U.K., Techh. Rep. IST-257024-Fish4Knowledge, 2011.

[17] *Ozfish Dataset—Machine Learning Dataset for Baited Remote Underwater Video Stations*, U. O. W. A. Australian Institute of Marine Science (AIMS) and C. Univ., Crawley, WA, Australia, 2021.

[18] H. Hou, B. Li, X. Zhao, X. Shao, Z. Liu, Y. Yan, Z. Wang, Y. Tao, and X. Yang, "Proteins characteristics and lipid profiles of silver Sillago (Sillago Sihama)," in *Proc. Int. Conf. Remote Sens., Environ. Transp. Eng.*, Jun. 2011, pp. 7738–7741.

[19] S. Girisha, U. Verma, M. M. M. Pai, and R. M. Pai, "UVid-Net: Enhanced semantic segmentation of UAV aerial videos by embedding temporal information," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4115–4127, 2021.

[20] G. S, M. P. M. M, U. Verma, and R. M. Pai, "Semantic segmentation of UAV videos based on temporal smoothness in conditional random fields," in *Proc. IEEE Int. Conf. Distrib. Comput., VLSI, Electr. Circuits Robot. (DISCOVER)*, Oct. 2020, pp. 241–245.

[21] X. Li, Y. Qi, Z. Wang, K. Chen, Z. Liu, J. Shi, P. Luo, X. Tang, and C. C. Loy, "Video object segmentation with re-identification," 2017, *arXiv:1708.00197*.

[22] P. Bergmann, T. Meinhardt, and L. Leal-Taixe, "Tracking without bells and whistles," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 941–951.

[23] E. Bochinski, V. Eiselein, and T. Sikora, "High-speed tracking-by-detection without using image information," in *Proc. 14th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug. 2017, pp. 1–6.

[24] S. Girisha, M. P. MM, U. Verma, and R. M. Pai, "Semantic segmentation of UAV aerial videos using convolutional neural networks," in *Proc. IEEE 2nd Int. Conf. Artif. Intell. Knowl. Eng. (AIKE)*, Jun. 2019, pp. 21–27.

[25] X. Sun, P. Wu, and S. C. H. Hoi, "Face detection using deep learning: An improved faster RCNN approach," *Neurocomputing*, vol. 299, pp. 42–50, Jul. 2018.

[26] A. Salman, A. Jalal, F. Shafait, A. Mian, M. Shortis, J. Seager, and E. Harvey, "Fish species classification in unconstrained underwater environments based on deep learning," *Limnology Oceanography, Methods*, vol. 14, no. 9, pp. 570–585, Sep. 2016.

[27] H. Jiang, S. Fels, and J. J. Little, "A linear programming approach for multiple object tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[28] O. Ulucan, D. Karakaya, and M. Turkan, "A large-scale dataset for fish segmentation and classification," in *Proc. Innov. Intell. Syst. Appl. Conf. (ASYU)*, Oct. 2020, pp. 1–5.

[29] *Aquarium Dataset*, Roboflow, Locust St, Des Moines, 2020.

[30] M. Pedersen, J. B. Haurum, S. H. Bengtson, and T. B. Moeslund, "3D-ZeF: A 3D zebrafish tracking benchmark dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2426–2436.

[31] D. McIntosh, T. P. Marques, A. B. Albu, R. Rountree, and F. D. Leo, "Movement tracks for the automatic detection of fish behavior in videos," 2020, *arXiv:2011.14070*.

[32] S. Butail and D. A. Paley, "Three-dimensional reconstruction of the fast-start swimming kinematics of densely schooling fish," *J. Roy. Soc. Interface*, vol. 9, no. 66, pp. 77–88, 2011.

[33] S. Shreesha, M. P. Mm, U. Verma, and R. M. Pai, "Computer vision based fish tracking and behaviour detection system," in *Proc. IEEE Int. Conf. Distrib. Comput., VLSI, Electr. Circuits Robot. (DISCOVER)*, Oct. 2020, pp. 252–257.

[34] Z.-M. Qian, S. H. Wang, X. E. Cheng, and Y. Q. Chen, "An effective and robust method for tracking multiple fish in video image based on fish head detection," *BMC Bioinform.*, vol. 17, no. 251, pp. 1–11, Jun. 2016.

[35] J.-H. Lee, M.-Y. Wu, and Z.-C. Guo, "A tank fish recognition and tracking system using computer vision techniques," in *Proc. 3rd Int. Conf. Comput. Sci. Inf. Technol.*, Jul. 2010, pp. 528–532.

[36] M. M. Saberioon and P. Cisar, "Automated multiple fish tracking in three-dimension using a structured light sensor," *Comput. Electron. Agricult.*, vol. 121, pp. 215–221, Feb. 2016.

[37] A. Rodriguez, A. J. Rico-Diaz, J. R. Rabunal, J. Puertas, and L. Pena, "Fish monitoring and sizing using computer vision," in *Proc. Int. Work-Conf. Interplay Between Natural Artif. Comput.*, Cham, Switzerland: Springer, 2015, pp. 419–428.

[38] R. Bhaskaran, R. K. Baskaran, and C. Vijayalakshmi, "Machine learning technique for analyzing the behavior of fish in an aquarium," in *Proc. Int. Conf. Modeling, Mach. Learn. Astron.*, Cham, Switzerland: Springer, 2019, pp. 55–65.

[39] V. M. Papadakis, I. E. Papadakis, F. Lamprianidou, A. Glaropoulos, and M. Kentouri, "A computer-vision system and methodology for the analysis of fish behavior," *Aquaculture Eng.*, vol. 46, no. 1, pp. 53–59, Jan. 2012.

[40] F. Zeng, B. Dong, Y. Zhang, T. Wang, X. Zhang, and Y. Wei, "MOTR: End-to-end multiple-object tracking with transformer," 2021, *arXiv:2105.03247*.

[41] P. Dai, R. Weng, W. Choi, C. Zhang, Z. He, and W. Ding, "Learning a proposal classifier for multiple object tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2443–2452.

[42] Y. Zhang, C. Wang, X. Wang, W. Zeng, and W. Liu, "Fairmot: On the fairness of detection and re-identification in multiple object tracking," *Int. J. Comput. Vis.*, vol. 129, no. 11, pp. 3069–3087, 2021.

[43] E. Yu, Z. Li, S. Han, and H. Wang, "RelationTrack: Relation-aware multiple object tracking with decoupled representation," *IEEE Trans. Multimedia*, early access, Feb. 10, 2022, doi: 10.1109/TMM.2022.3150169.

[44] S. Tang, M. Andriluka, B. Andres, and B. Schiele, "Multiple people tracking by lifted multicut and person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3539–3548.

[45] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[46] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2016, pp. 21–37.

[47] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[48] H. Wu, C. Du, Z. Ji, M. Gao, and Z. He, "SORT-YM: An algorithm of multi-object tracking with YOLOv4-tiny and motion prediction," *Electronics*, vol. 10, no. 18, p. 2319, Sep. 2021.

[49] G. Khan, Z. Tariq, M. U. G. Khan, P. Mazzeo, S. Ramakrishnan, and P. Spagnolo, "Multi-person tracking based on faster R-CNN and deep appearance features," in *Visual Object Tracking With Deep Neural Networks*. London, U.K.: IntechOpen, 2019, pp. 1–23.

[50] T. Meinhardt, A. Kirillov, L. Leal-Taixe, and C. Feichtenhofer, "TrackFormer: Multi-object tracking with transformers," 2021, *arXiv:2101.02702*.

[51] Y. Cui, D. Guo, Y. Shao, Z. Wang, C. Shen, L. Zhang, and S. Chen, "Joint classification and regression for visual tracking with fully convolutional Siamese networks," *Int. J. Comput. Vis.*, vol. 130, pp. 550–566, Jan. 2022.

[52] Y. Xiang, A. Alahi, and S. Savarese, "Learning to track: Online multi-object tracking by decision making," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4705–4713.

[53] M. Radolko, F. Farhadifard, and U. V. Lukas, "Change detection in crowded underwater scenes–via an extended Gaussian switch model combined with a flux tensor pre-segmentation," in *Proc. 12th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2017, pp. 405–415.

[54] E. Hossain, S. M. S. Alam, A. A. Ali, and M. A. Amin, "Fish activity tracking and species identification in underwater video," in *Proc. 5th Int. Conf. Informat., Electron. Vis. (ICIEV)*, May 2016, pp. 62–66.

[55] M.-C. Chuang, J.-N. Hwang, J.-H. Ye, S.-C. Huang, and K. Williams, "Underwater fish tracking for moving cameras based on deformable multiple kernels," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 9, pp. 2467–2477, Sep. 2016.

[56] T. H. Pinkiewicz, G. J. Purser, and R. N. Williams, "A computer vision system to analyse the swimming behaviour of farmed fish in commercial aquaculture facilities: A case study using cage-held Atlantic salmon," *Aquacultural Eng.*, vol. 45, no. 1, pp. 20–27, Jul. 2011.

[57] J. Xu, Y. Liu, S. Cui, and X. Miao, "Behavioral responses of tilapia (Oreochromis niloticus) to acute fluctuations in dissolved oxygen levels as monitored by computer vision," *Aquacultural Eng.*, vol. 35, no. 3, pp. 207–217, Oct. 2006.

[58] S. Butail, P. Salerno, E. M. Bollt, and M. Porfiri, "Classification of collective behavior: A comparison of tracking and machine learning methods to study the effect of ambient light on fish shoaling," *Behav. Res. Methods*, vol. 47, no. 4, pp. 1020–1031, Dec. 2015.

[59] I. A. Iswanto, T. W. Choa, and B. Li, "Object tracking based on meanshift and particle-Kalman filter algorithm with multi features," *Proc. Comput. Sci.*, vol. 157, pp. 521–529, Jan. 2019.

[60] L. Long, Z. V. Johnson, J. Li, T. J. Lancaster, V. Aljapur, J. T. Streelman, and P. T. McGrath, "Automatic classification of cichlid behaviors using 3D convolutional residual networks," *iScience*, vol. 23, no. 10, Oct. 2020, Art. no. 101591.

[61] A. Shirke, A. Saifuddin, A. Luthra, J. Li, T. Williams, X. Hu, A. Kotnana, O. Kocabalkanli, N. Ahuja, A. Green-Miller, I. Condotta, R. N. Dilger, and M. Caesar, "Tracking grow-finish pigs across large pens using multiple cameras," 2021, *arXiv:2111.10971*.

[62] V. M. Papadakis, A. Glaropoulos, and M. Kentouri, "Sub-second analysis of fish behavior using a novel computer-vision system," *Aquacultural Eng.*, vol. 62, pp. 36–41, Sep. 2014.
[63] A. Milan, L. Leal-Taixe, I. Reid, S. Roth, and K. Schindler, "MOT16: A benchmark for multi-object tracking," 2016, *arXiv:1603.00831*.
[64] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The CLEAR MOT metrics," *J. Image Video Process.*, vol. 2008, pp. 1–10, Feb. 2007.
[65] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 24, 2011, pp. 1–86.
[66] E. Bochinski, T. Senst, and T. Sikora, "Extending IOU based multi-object tracking by visual information," in *Proc. 15th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Nov. 2018, pp. 1–6.

S. SHREESHA received the B.E. degree from the Srinivas School of Engineering, VTU, Belgaum, and the master's degree in VLSI design and embedded systems from NMAMIT, Nitte, India. He is currently pursuing the Ph.D. degree in computer vision and machine learning with the Manipal Institute of Technology. His research interests include object detection, object tracking, and deep learning for computer vision.

M. M. MANOHARA PAI (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering. He is currently a Professor with the Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India. He has a rich experience of 31 years as a Professor. He holds seven patents to his credit and has published 131 papers in national and international journals/conference proceedings. He has published two books, guided six Ph.D.'s, and 85 master's thesis. His research interests include data analytics, cloud computing, the IoT, computer networks, mobile computing, scalable video coding, and robot motion planning. He is also a Life Member of ISTE and a Life Member of Systems Society of India. He is also a principal investigator for multiple industry/government research projects. He has been an Executive Committee Member of the IEEE Bangalore Section, Mangalore Subsection, and the past Chair of the IEEE Mangalore Subsection. He has recently received National award from ministry of education, Government of India for Best Technical Teacher (NTTA 2022).

UJJWAL VERMA (Senior Member, IEEE) received the Ph.D. degree in image analysis from Télécom ParisTech, University of Paris-Saclay, Paris, France, and the M.S. (Research) degree in signal and image processing from IMT Atlantique, France. He is currently an Associate Professor and the Head of the Department of Electronics and Communication Engineering, Manipal Institute of Technology, Bengaluru, India. His research interests include computer vision and machine learning, focusing on variational methods in image segmentation, deep learning methods for scene understanding, and semantic segmentation of aerial images. He was a recipient of the "ISCA Young Scientist Award (2017–2018)" by the Indian Science Congress Association (ISCA), a professional body under the Department of Science and Technology, Government of India. He is also the Co-Lead of the Working Group on Machine/Deep Learning for Image Analysis (WG-MIA) of the Image Analysis and Data Fusion Technical Committee (IADF TC) of the IEEE Geoscience and Remote Sensing Society. He is also a Guest Editor of Special Stream in IEEE Geoscience and Remote Sensing Letters and a Reviewer for several journals, such as IEEE Transactions on Image Processing, IEEE Transactions on Geoscience and Remote Sensing, and IEEE Geoscience and Remote Sensing Letters. He is also a Sectional Recorder of the ICT Section of the Indian Science Congress Association (2020–2022). He is a Life Member of the Indian Science Congress Association.

RADHIKA M. PAI (Senior Member, IEEE) received the Ph.D. degree from the National Institute of Technology Karnataka, Surathkal, India. She is currently a Professor and the Head of the Department of Data Science and Computer Applications, Manipal Academy of Higher Education, Manipal, India. She has a teaching and research experience of over 30 years. She has published 87 papers in national/international journals/conferences and has guided three Ph.D.'s and several master's thesis. Her research interests include data mining, big data analytics, character recognition, sensor networks, and e-learning. She was an Executive Committee Member of the IEEE Mangalore Subsection. She was a recipient of National Doctoral Fellowship from AICTE, Government of India.