A Review of Computer Vision Technologies for Fish Tracking

Zhenbo Li\textsuperscript{a,b,c,d,*}, Weiran Li\textsuperscript{a,b,c}, Fei Li\textsuperscript{a,b,c}

\textsuperscript{a}College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
\textsuperscript{b}Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture, Beijing 100083, China
\textsuperscript{c}Beijing Engineering and Technology Research Center for Internet of Things in Agriculture, Beijing 100083, China
\textsuperscript{d}National Innovation Center for Digital Fishery, China Agricultural University, Beijing 100083, China

Abstract: Fish tracking based on computer vision is a complex and challenging task in fishery production and ecological studies. Most of the applications of fish tracking use classic filtering algorithms, which lack in accuracy and efficiency. To solve this issue, deep learning methods utilized deep neural networks to extract the features, which achieve a good performance in the fish tracking. Some one-stage detection algorithms have gradually been adopted in this area for the real-time applications. The transfer learning to fish target is the current development direction. At present, fish tracking technology is not enough to cover actual application requirements. According to the literature data collected by us, there has not been any extensive review about vision-based fish tracking in the community. In this paper, we introduced the development and application prospects of fish tracking technology in last ten years. Firstly, we introduced the open source datasets of fish, and summarized the preprocessing technologies of underwater images. Secondly, we analyzed the detection and tracking algorithms for fish, and sorted out some transferable frontier tracking model. Thirdly, we listed the actual applications, metrics and bottlenecks of the fish tracking such as occlusion and multi-scale. Finally, we give the discussion for fish tracking datasets, solutions of the bottlenecks, and improvements. We expect that our work can help the fish tracking models to achieve higher accuracy and robustness.

Keywords: Fish tracking; Underwater image; Computer vision; Underwater data preprocessing.

1 Introduction

According to the “State of World Fisheries and Aquaculture” report\textsuperscript{(Stankus, 2021)}, the total global aquatic product production is expected to increase from 179 million tons in 2018 to 204 million tons in 2030. The consumption of aquatic products is expected to reach 21.5 kg, higher than the 2018 average of 20.5 kg per capita. During the forecast period, the world's trade in aquatic products for human consumption is expected to increase by 9% and exceed 54 million tons of fresh weight equivalent in 2030, occupying an important position in the world economy and trade. The tracking technology of fish can be applied to disease identification, hypoxic stress identification, etc., which is of great significance in fishery.

Fish identification and behavior analysis have attracted great attention in the scientific community. The identification and tracking methods of fish are mainly used in ocean observation, aquaculture, biological research, etc. Compared with the methods based on acoustics or sensors,
fish tracking methods based on computer vision have advantages of real-time, contactless, simple equipment requirements and do not affect the normal behavior of fish (Li, Wang, Wu, Miao, Du, Duan, 2020). Furthermore, image contains a various of information, the deep semantic information invisible can be obtained through data analysis.

Object tracking means to predict trajectories for the instances of interest in videos. Filters are widely used in fish tracking, which mainly are generating model methods such as Kalman filter and particle filter. We found that the combination of background subtraction method and Kalman filter takes the majority of fish tracking methods. The optical flow method uses the changes of pixel points between frames to calculate the motion parameters, which is also used in the small-scale tracking such as fish targets(Mohamed, Fadl, Anas, Wageeh, ElMasry, Nabil, Atia, 2020). In the literature compiled, we count the detectors and trackers used in fish tracking areas in last ten years, as shown in Figure 1.

Compared with traditional methods which are less robust to the complex water environments, the deep learning-based tracking technologies can fit the underwater data distribution better and process massive data better as well. Deep learning technology has great development prospects, and tracking algorithms used in pedestrian, vehicle and other fields are being updated frequently. The deep learning methods follow the discriminative tracking process, making full use of the background information of the data to distinguish targets. Novel tracking algorithms tend to merge the detection and tracking models into a unified framework to avoid the overall loss caused by sub-model errors, which is called Joint Detection and Embedding (JDE) model. However, tracking methods based on deep learning are not widely used in fish applications. Transfer the methods for pedestrians and vehicles to track fish targets and enhancing the real-time performance of the model are the main research points.

(a) Statistics of fish detectors  (b) Statistics of fish trackers

Figure 1: Statistics of fish detectors in recent years

This article summarizes and introduces the related methods of fish detection and tracking based on computer vision technology. The rest of the paper is organized as follows: The typical open-source fish datasets and underwater data preprocessing methods are introduced in Section 2. The detection and tracking models of fish is presented in Section 3. The applications and bottlenecks of fish tracking is drawn in Section 4. Then, some discussion about the status and difficulties of fish tracking is shown in Section 5. Finally, some conclusions are described in Section 6.
2 Preprocessing

In this section, we describe the related underwater datasets and processing methods in fish tracking domain. On the one hand, we compile several tracking datasets and related datasets, as shown in Table 1. On the other hand, we describe some underwater image enhancement and restoration methods.

The fishes video data, especially some research in marine ecosystems, is generally difficult to obtain due to the limitations of deep-underwater light source attenuation and underwater noise. What's more, it also has extremely high requirements for manual labeling. It is difficult to perform original video analysis and abstract the observed data. Through underwater data enhancement and restoration technology can significantly improve the quality of the degraded images. Therefore, public datasets and the enhancement and restoration techniques of underwater data play a very active role in new methods.
| Dataset               | Fish4Knowledge | LifeCLEF2014          | LifeCLEF2015          | Labeled Fishes in the Wild |
|----------------------|----------------|----------------------|----------------------|---------------------------|
| **Data type**        | Video / Image  | Video                | Video / Image        | Video / Image             |
| **Data volume**      | 200TB, 6,514 video clips / 27,370 images, 23 types | About 1,000 video clips | 20 videos / 20,000 images | An ROV video, 4096 images |
| **Characteristics**  | The data categories collected by the dataset are relatively unbalanced | From the Fish4Knowledge video dataset, including algae attachment data and turbid water data, difficult to identify | Both video and image data are clearly labeled, and there is still a problem of imbalance of data samples | Contains positive and negative datasets |
| **Website**          | https://groups.inf.ed.ac.uk/f4k/ | https://www.imageclef.org/2014/lifeclef/fish | https://www.imageclef.org/lifeclef/2015/fish | https://swfscdata.nmfs.noaa.gov/labeled-fishes-in-the-wild/ |
2.1 Fish Datasets

**Fish4Knowledge** The Fish4Knowledge (F4K) is an open-source dataset funded by the European Union's Seventh Framework Program. Taiwan Power Company, Taiwan Ocean Research Institute, and Kenting National Park shared video datasets observed from Nanwan, Lanyu, and Houbi Lake of Taiwan. Figure 3(a) is several video frames of F4K dataset. This project mainly studies the methods of capturing, storing, analyzing, and querying multiple video streams of seabed. The project focuses on collecting video data of marine ecosystems, focusing on environmental research tasks. The collected data can be used for fish detection and recognition, target tracking, and analysis. In the project, 10 cameras are used to collect more than 12 hours of available daylight data, and about 100TB of data is collected each year. A total of 23 types of fish species image datasets are collected, a total of 27,370 images, of which the number of different types is relatively not equal. This dataset is large in scale and is one of the most typical public datasets for fish targets, but the downside is the data types are not balanced obviously.

![Fish4Knowledge Examples](image1)

*Figure 3: Open-source datasets samples*

**LifeCLEF** LifeCLEF aims to establish a comprehensive biological information database since 2011, including species, geographical distribution, evolution, etc. LifeCLEF proposed the fish datasets FishCLEF2014 and FishCLEF2015 and the marine life datasets SeaCLEF2016 and SeaCLEF2017 from 2014 to 2017. FishCLEF2014's underwater video dataset comes from Fish4Knowledge's video database. It contains about 700,000 10-minute video clips of Taiwan's coral reefs in the past five years, including videos recorded from sunrise to sunset. The dataset comprises turbid water background, and the data on the camera lens and the algae attached to the camera lens certain difficulties in identifying fish. FishCLEF2014 contains about 40GB of the training set and 8GB of the testing set, which are composed of four sub-datasets. Each sub-dataset is used for different research tasks. The sub-dataset of the image-based fish species recognition task is composed of images, the others are all video datasets. FishCLEF2015 is based on the Fish4Knowledge database like FishCLEF2014. The training dataset includes 20 manually annotated videos, a list of fish species, sample images of each fish. The dataset consists of more than 9,000 annotations and more than 20,000 sample images. The testing set includes 73 underwater videos, as well as a list of fish species. The dataset still has the problem of unbalanced data samples.
SeaCLEF2016 and SeaCLEF2017 are derived from the fish identification tasks of FishCLEF2014 and FishCLEF2015. Compared with the F4K dataset, FishCLEF2014 and FishCLEF2015 have been extensively expanded, considering general marine life, which make the training performance better.

**Labeled Fishes in the Wild** The Labeled Fishes in the Wild dataset is provided by National Marine Fisheries Service (NOAA) to encourage the development, testing, and evaluation of unconstrained automatic image analysis algorithms. The dataset mainly contains images of fish, invertebrate fish, seabed and corresponding annotation files, and annotated the location of the fish target. The dataset includes a positive dataset (for training and verification, fish data), a negative image set (for training and verification, non-fish data), and a testing image dataset as is shown in Figure 3(b). For the training set and the testing set, the specific location of the fish target object is marked with the size of the fish range. The dataset was taken by the Southwest Fisheries Science Center using a forward oblique digital camera deployed on a Remote Operated Vehicle (ROV). The positive dataset contains 929 image files and 1,005 related annotations, and the negative dataset contains 3,167 pictures. The testing dataset contains image sequences captured using ROV's HD cameras. A total of 2061 fish objects are marked in the annotated frame. The dataset is of good quality, clearly divided, and has corresponding annotations.

### 2.2 Enhancement and Restoration

Underwater images often suffer from severe quality degradation due to light absorption and scattering in water medium (Wang, Song, Fortino, Qi, Zhang, Liotta, 2019), so that the general image enhancement methods do not perform well on underwater data processing. The underwater image processing technology can be divided into image enhancement algorithm and image restoration algorithm according to whether it is based on an underwater imaging physical model. On the one hand, the illumination of the underwater environment is less stable than the atmospheric environment, and the scattering and refraction of light will cause the loss of brightness of the image. The color deviation is mainly caused by light absorption, blurring of details caused by light forward scattering and low contrast caused by light backward scattering. On the other hand, due to underwater suspended particles, there are more noises in underwater images and videos compared to the land. As a result, underwater images generally have low contrast, color distortion, noise, etc., and the enhancement technology will qualitatively improve tracking models.

**Enhancement** The image enhancement method aims to research and enhance the pixel points of the obtained underwater image. Underwater image enhancement methods include traditional technologies and data-driven technologies. Some physical models are widely used such as atmospheric scattering models, simplified underwater imaging models, and modified underwater imaging models, which comprehensively considering the influencing factors of image quality. On the contrary, the widely used deep learning methods only pay attention to color correction, so the precision is higher (Naik, Swarnakar, Mittal, 2021). Deep Convolutional Neural Networks (Deep-CNN) and Generative Adversarial Networks (GAN) are often used for image enhancement. Table 2 and Figure 4 list some deep networks for underwater data enhancement (Anwar, Li, 2020a). The CNN-based goal is to keep the original images, but the GAN-based goal is to improve the quality of the images.

| Network            | Example       |
|--------------------|---------------|
| Encoder-Decoder networks | P2P, UGAN |

Table 2: Deep underwater data enhancement networks
| Modular design networks | UWCNN, DenseGAN |
|------------------------|----------------|
| Multi-branch designs   | DUIENet, FGAN  |
| Depth-guided networks  | URCNN, WaterGAN|
| Dual Generator GANs    | UWGAN, MCycleGAN|

![Figure 4: Qualitative comparisons of various networks](image)

**Restoration**

The image restoration method is to obtain the real situation according to the imaging process of the underwater image, which requires information such as the optical parameters of the water, camera parameters, and the distance between the camera and the target object, and then uses the physical model for calculation. Underwater image restoration methods make the data have a better visibility, such as dehazing methods. Some methods based on the Retinex technology achieve a good performance. Retinex theory is based on color constancy, which can adaptively enhance different types of images and has a better balance effect: (Fu, Zhuang, Huang, Liao, Zhang, Ding, 2014) proposed a single underwater image enhancement method based on the Retinex. The method decomposes the reflectivity and brightness of a single underwater image, uses an alternate direction optimization strategy to solve the model, and carries out color correction and image brightness. This method enhances and restores better the edges and details of the image.
3 Fish Tracking Method

Object tracking methods can be divided into two types according to the initialization mode of the target: manually marking the target location and using detection algorithms for recognition. The former method is relatively fast, and the disadvantage is that it cannot track the new target following the video. The latter method is based on target detection, which recognizes the target of interest in the data sequence at first, and then pairs and associates direct targets in different frames. In Figure 5, we summarize and classify the methods in the field of fish tracking. Other division basis for tracking are listed in Table 3.

![Common fish detection and tracking methods](image)

Figure 5: Common fish detection and tracking methods
| Classification | Method | Introduction |
|----------------|--------|--------------|
| Task calculation | Online tracking | Real-time processing of tasks, using only current and past frames to track the position of objects on future frames |
|                | Offline tracking | Offline processing tasks, using past, present, and future frames for object position tracking, with high accuracy |
| Tracking task  | Single object tracking | Track the location of a given target |
|                | Multi-object tracking | Track the location of multiple targets |
| Method category | Re-ID | Considered as a sub-problem of image retrieval, judging the similarity with a given picture |
|                | MTMCT | Multi-target multi-camera tracking, considered as an extension of Re-ID |
|                | Generative model | Establish a target model or extract target features, search for similar features in subsequent frames, and iteratively achieve target positioning step by step, regardless of background information |
|                | Discriminative model | Consider the background information and target model, and detect the current frame position of the target through difference comparison |
|                | Tracking-by-detection | Take the current state of the relevant target as input, and use the tracking algorithm to predict the position |
| Model fusion degree | Joint-detection-and-prediction | Integrate detection and prediction methods into a single network, perform positioning and prediction, and associate targets in different frames |
|                | Joint-detection-and-embedding | Combine the two parts of detection and recognition into a first-level network |
In the field of underwater object tracking, the classic Tracking-by-detection (TBD) method is still widely used. This section mainly summarizes the algorithms from the two stages of both target detection and tracking.

3.1 Detection for TBD

3.1.1 Subtraction

The performance of the detector will directly affect the tracking results. For fish detection, traditional methods mostly use the background subtraction method or the improved ones.

**Background Subtraction Method** Due to the poorer quality of underwater videos and images illuminance and clarity compared with the general data, the background subtraction method is one of the most widely used fish target extraction methods, which can produce good results on the data captured by a still camera. Gaussian of Mixture Models (GMM) is a kind of background subtraction method, which uses Gaussian probability density function to accurately quantify the target and decomposes it into several models based on Gaussian probability density function curves. These methods have a better processing effect on slow-moving objects and robust against complex multi-model backgrounds, but the computational cost is higher (Friedman, Russell, 2013; Stauffer, Grimson, 1999; Zivkovic, 2004).

**Otsu Adaptive Segmentation** The Maximum Between-Class Variance Method (Otsu) is an adaptive segmentation algorithm based on the background subtraction method. Otsu is often used to automatically generate thresholds in the background subtraction method (Otsu, 1979), and further obtain the tracking feature parameters of the fish according to the geometric characteristics of the fish. But it will affect the tracking performance when the water quality is poor or the overlap problem is serious, as is shown in Figure 6.

![Otsu adaptive segmentation algorithm](image1)

![Simple threshold segmentation](image2)

**Figure 6: Otsu adaptive segmentation method comparison**

**Inter-frame Difference Method** The inter-frame difference is a method of moving target detection based on consecutive frames in the video sequence. The inter-frame difference method can realize fast detection of moving fish under different environmental conditions due to the simple principle. In addition, it only uses the difference comparison between adjacent frames, so the computational complexity is quite low. But when the moving target changes between adjacent frames is not obvious, the target will overlap between different frames. Therefore, if only the inter-frame difference method is used to detect targets, it often produces holes in this time.

**Combined Difference Method** In applications of fish detection, the inter-frame difference method and the background subtraction method are often combined. (Wu, Xiao, Zhang, Chen, Zhu, Li, 2011) proposed a fish school tracking method based on Delaunay triangulation, which uses a background update method that combines two methods of frame subtraction and background...
subtraction (Salman, Maqbool, Khan, Jalal, Shafait, 2019). This method divides the pixels of the frame to be updated into the foreground, ghost, foreground aperture, etc., and uses different update factors to update each part, which has the advantages of shorter processing time and more accurate background update. (Nguyen, Huynh, Vo, Van Pham, 2015) also used the method of GMM and the inter-frame difference method. When using this method for fish detection, the three-frame difference algorithm (Singla, 2014) is applied, it is no longer just to use the adjacent two frames of image for difference. For the search of moving objects, each two frame difference images of adjacent three frames of pictures are used to perform a union operation to solve the difficult problem of detection of similar colors or slow motion. Combining the background difference method with the inter-frame difference method has achieved good results in the detection of fish targets.

3.1.2 Based on Deep Learning

Traditional methods are commonly considered as less robust ones due to poor regional selection strategies, high time complexity, and poor robustness of manually extracted features. Deep learning methods gradually replace them by the strong learning capabilities, strong portability, the characteristic of data-driven.

Two-stage Method: R-CNN Series Region-Convolutional Neural Network (R-CNN) follows the idea of the traditional algorithms, which sequentially perform candidate region generation, feature extraction, feature classification, and non-maximum suppression (Girshick, Donahue, Darrell, Malik, 2014). It uses a selective search method to select candidate regions, and generates 1,000 to 2,000 candidate frames of different sizes and shapes on a given picture. The algorithm uses a convolutional neural network (CNN) to extract deep features of the candidate frames, support vector machine (SVM) for feature classification, and the non-maximum suppression method to select the borders with higher SVM scores to obtain the detection results. The disadvantage is when generating candidate frames and extracting features, a large number of repetitive calculations are performed, and the image size requirements are expensive. After R-CNN, optimization algorithms such as SPPNet, Fast R-CNN (Girshick, 2015; He, Zhang, Ren, Sun, 2015), Faster R-CNN (Ren, He, Girshick, Sun, 2015), Mask R-CNN (He, Gkioxari, Doll A R, Girshick, 2017) appeared. Taking Faster-RCNN as an example, Faster-RCNN used Region Proposal Network (RPN) algorithm instead of the original Selective Search method to generate candidate frames, and applied the CNN network to select candidate frames and detect target. This reduces the number of candidate frames and the quality of the suggested frames is also substantially improved. The R-CNN series of algorithms is applied in the target detection stage of the dataset processing. However, these methods are more inclined to the accuracy of object detection. In fish tracking applications, due to the high requirements for real-time detection, this type of algorithm is not commanded.

One-stage Method: YOLO and SSD The You Look Only Once (YOLO) series of algorithms is a typical one-stage algorithm based on deep learning methods. Compared with the R-CNN series of methods, the classification of object categories and the positioning of the bounding box are unified into a regression calculation. They achieved high efficiency and strong real-time performance. So far, the YOLO series of algorithms have updated YOLOv1 (Redmon, Divvala, Girshick, Farhadi, 2016), YOLOv2 (Redmon, Farhadi, 2017), YOLOv3 (Redmon, Farhadi, 2018), YOLOv4 (Bochkovskiy, Wang, Liao, 2020), YOLOv5 (Jocher, Stoken, Borovec, NanoCode, Chaurasia, TaoXie, Changyu, V, Laughing, Tkianai, YxNONG, Hogan, Lorenzomammmana, AlexWang, Hajek, Diaconu, Marc. Kwon, Oleg, Wanghaoyang, Defretin, Lohia, Ah, Milanko, Fineran, Khromov, Yiwei, Doug, Durgesh, Ingham, 2021), and other generations of algorithms. The
core idea of the YOLO series is to use the entire image as a network and perform a one-time regression of bounding box and class in the output layer. Compared with other algorithms, YOLO has the advantages of fast speed which can achieve real-time detection, and uses the whole image as context information, the probability of incorrectly identifying the background as an object is small, and the generalization ability is strong. It can be directly performed end-to-end optimization according to the detection performance. Single Shot MultiBox Detector (SSD) (Liu, 2016) is also a typical One-stage algorithm. Compared with YOLO, SSD is directly detected through CNN and is not detected after the fully connected layer. It also proposes feature maps of different scales and aspect ratios, the detection and positioning of small targets are more accurate.

In the area of fish target detection, the YOLO series of algorithms are well applied. (Liu, Li, Gao, Cai, Nian, Li, Yan, Lendassse, Liu, Li, Liu, Deng, Liu, Zhai, 2018) proposed an online fish detection and tracking method combining YOLOv3 and parallel correlation filters. In the detection stage, YOLOv3 is used to predict the score of each bounding box using logistic regression. Multi-label classification predicts possible categories, but the method does not effectively deal with fish occlusion. (Liu, Li, Liu, Deng, Liu, Zhai, 2021) proposed a method for real-time fish detection and tracking based on YOLOv4, which can generate population statistics of a variety of fish. The method can achieve better fish detection and tracking in a relatively complex underwater environment. However, the generalization of water quality conditions is hard, and the accuracy of identifying other water conditions has significant decrease. (Mohamed, Fadl, Anas, Wageeh, ElMasry, Nabil, Atia, 2020) proposed a fish trajectory detection technology based on the YOLOv3. Combining the YOLO and the optical flow method to obtain the motion trajectories of multiple fishes has better detection accuracy in the case of turbid water. Due to the small-scale of training datasets, this method cannot effectively ensure the accuracy of fish detection under complex conditions. In general, the YOLO series of algorithms also have general defects of insufficient accuracy in fish tracking.

3.2 Tracking Method

Generally speaking, tracking methods can be divided into two categories: (1) One method is to initialize the instance first, which means to manually select the tracking target in the first frame, and then use a specific tracking algorithm to track. This method is relatively fast but cannot track the new instance in the subsequent frame; (2) The other method uses detection algorithms to recognize the interest targets, and then use tracking methods to pair and associate targets between different frames.

3.2.1 Association Method

Most of the mainstream methods adopt a tracking-by-detection paradigm to track fish targets (Held, 2016a; Henriques, Caseiro, Martins, Batista, 2014; Lantsova, Voitiuk, Zudilova, Kaarna, 2016; Li, Wei, Huang, Nie, Zhang, Wang, 2018; Palconit, Almero, Rosales, Sybingco, Bandala, Vicerra, Dados, 2020a; Sharif, Galip, Guler, Uyaver, 2015; Shijun, Yao, 2012), as shown in Figure 7. This type of algorithm takes the current state of the target as input, predicts the position of the following frame, and then uses the Hungarian algorithm to make corrections.
Fish tracking models are usually based on the method of kinematics for algorithm design. Kinematics-based target tracking methods such as particle filter, Kalman filter, Kernel Correlation Filter (KCF) algorithm, etc., have extensive applications.

**Filter** Filtering algorithms are classic methods for fish target tracking. Common filters include particle filter, Kalman filter, Kernel Correlation Filter, etc. Particle filter is a sequential estimation method, which has superiority in processing nonlinear and non-Gaussian motion models (Pinkiewicz, Williams, Purser, 2008; Ristic, Arulampalam, Gordon, 2004). Proposed an enhanced tracking algorithm based on particle filtering, which used an adaptive partition scheme to analyze fish movement through data association of the global nearest neighbor method. However, the robustness of this method needs to be improved. Kalman filter is a popular and effective method of system state estimation. It can make educated predictions in dynamic systems with uncertain information. A single Kalman filter is used to track the center of each fish, but some tracking will be significantly affected under high turbulence conditions (Lantsova, Voitiuk, Zudilova, Kaarna, 2016; Sharif, Galip, Guler, Uyaver, 2015; Shijun, Yao, 2012). Kernel Correlation Filter (KCF) algorithm (Henriques, Caseiro, Martins, Batista, 2014) belongs to discriminant tracking methods. It performs well in terms of speed and accuracy. The main problem of filters is that a large number of samples are needed to approximate the posterior probability density of the system well (Henriques, Caseiro, Martins, Batista, 2014; Li, Wei, Huang, Nie, Zhang, Wang, 2018).

**Optical flow** Optical flow refers to the instantaneous velocity of a moving object in space when observing the movement of pixels on the imaging plane. It uses the change of a certain pixel in the image sequence on the time axis and the correlation between adjacent frames to find the correspondence between two frames, and then handle the object between adjacent frames. The calculation of motion information has been widely used in detection. (Mohamed, Fadl, Anas, Wageeh, ElMasry, Nabil, Atia, 2020) uses the optical flow method to recognize the movement of the object, passes the center point of the detection frame to the optical flow method to track the movement of the fish, and combines the Retinex color enhancement algorithm with YOLO to solve the turbid water body to a certain extent. It is difficult to detect the size, quantity, and motion characteristics of fish targets.

**SORT and DeepSORT** SORT and DeepSORT methods are widely used in tracking technology,
which algorithm is relatively simple but has an outperforming. SORT (Simple Online and Realtime Tracking) (Palconit, Almero, Rosales, Sybingco, Bandala, Vicerra, Dadios, 2020a) and DeepSORT (Held, 2016a) are the key algorithms of Multiple object tracking (MOT) in engineering applications. The method used Faster-RCNN is used for detection and Kalman filtering to predict the state. SORT is based on the Hungarian algorithm using the position of the detection frame and intersection over union (IOU). The method obtained good performance at a high frame rate, but the ID switching is too frequent, which has fewer practical applications. DeepSORT integrates appearance information to improve the performance of SORT, which can track objects through longer occlusions. DeepSORT effectively reduced the number of ID switching.

Although the performance of the above model has been improved, the accuracy of any sub-models of tracking and detection will decrease in the final performance. What’s more, the network cannot realize sharing, there is a difficulty to breakthrough in terms of computing speed.

3.2.2 Based-on Deep Learning

Deep learning algorithm-based models in fish tracking is in exploratory, some improved algorithms based on Recurrent Neural Network (RNN) and GoTurn have been developed.

**LSTM in fish tracking**  Long Short-Term Memory (LSTM) is a type of time cyclic network, which solves the long-term dependency problem of general RNN, and effectively prevents the gradient disappearance or explosion in the long sequence training process. (Palconit, Almero, Rosales, Sybingco, Bandala, Vicerra, Dadios, 2020b) proposes a model that uses genetic algorithms and LSTM to predict fish trajectories. It tentatively utilizes genetic algorithms to predict the shortest path of fish which have obtained a promising a result. The detection is based on the background difference method and blob analysis to extract the information such as bounding box, location, area of the fish target. The genetic algorithm employs linear regression as the fitness function, tourism selection to find the best coordinates and select the optimal solution. In the LSTM network, it uses the detected centroid positions of the fish target in three adjacent frames as input for trajectory prediction. The method has good processing performance in video data under turbid waters. But it needs to be improved in accuracy because it does not effectively utilize the marking of fish.

**GoTurn in fish tracking**  The GoTurn (Held, 2016b) is an offline tracker with high speed of processing which based on deep learning method. The offline training needs labeled video data, while the algorithm uses only simple forward propagation to develop the tracking speed which reaches 100fps. The method effectively prevented over-fitting because learning a certain generalized motion feature can also achieve effective tracking of new objects that have not appeared before. Due to the complex underwater environment, the high false negative rate of traditional tracking methods often leads to the fish tracking failures. (Arvind, Prajwal, Bhat, Sreedevi, Prabhudeva, 2019) conducted a comparative study on fish detection and tracking in parallel processing of high-resolution image data in multiple regions, using Mask R-CNN and GoTurn tracking algorithms for real-time fish detection and tracking. It runs the detector on the data captured by the drone to generate candidate fish areas, generates and classifies the mask, and then uses the GoTurn algorithm to track candidate routes, which improves the efficiency of the algorithm operation. The method can achieve real-time detection, tracking, and counting of fish, but under turbid water may cause missing of fish, the UAV is hard to get clear images in the murky conditions as well.

Due to the difficulty in obtaining tracking datasets, most of the latest tracking methods are
trained based on the open source datasets such as VOT Challenges (Visual Object Tracking) and MOT Challenges (Multiple Object Tracking), which makes the performance more representative in model evaluation.

**Siamese network**  Siamese network (Li, Wu, Wang, Zhang, Xing, Yan, 2019) based trackers with well-balanced tracking accuracy and efficiency received significant attentions recently. The Siamese network turns the target tracking problem into a patch block matching problem. The development timeline is shown in Figure 8. SiamRPN++ (Li, Wu, Wang, Zhang, Xing, Yan, 2019) uses a deeper network (ResNet) to improve the tracking performance. They proposed a new layered and deep aggregation model architecture, which assembles the hierarchy of connections to aggregate different levels of representation and a depth-wise correlation layer. SiamGAT (Guo, Shao, Cui, Wang, Zhang, Shen, 2021) proposed a graph attention module to match the different part for the information embedding. Compared with the previous works, SiamGAT is adaptive to the size and aspect-ratio variations. SiamMOT (Shuai, Berneshawi, Li, Modolo, Tighe, 2021) uses area-based trackers to model instance-level motion. Since it’s difficult to distinguish the characteristics of different targets in fish, the Siamese network is still to be developed in practical applications. Siamese network can reach the filtering series method in performance, with great potential in the data-driven research.

![Development timeline of tracking method based on Siamase network](image)

**GNN-based model**  Graph Neural Network (GNN) has begun to be applied in tracking due to its advantages in relational modeling. (Li, Gao, Jiang, 2020) designed a nearly online MOT tracking method based on GNN, and proposed a missed detection strategy to notify the detector of defects. At the same time, an update mechanism is introduced to update subsequent nodes, edges, and node variables. (Jiang, Li, Li, Zhen, 2019) designed an end-to-end framework combining affinity learning and optimization modules. They used the GNN to deal with the problem of data association of online multi-target tracking. (Bras O, Leal-Taix E, 2020) used the classical network flow formulation of MOT to define a fully differentiable framework based on Message Passing Networks (MPNs). This method performs global inference on the entire detection set and predicts the final solution. (Dai, Weng, Choi, Zhang, He, Ding, 2021) designed a proposal-based learning framework. This paradigm models MOT as three steps: generating proposals, scoring proposals, and inferring trajectories on an affinity graph. The method uses iterative graph clustering to generate proposals and graph convolutional networks to learn the structural patterns of proposals, scoring and sorting. As a result, the framework improves the quality of proposals, significantly reduces calculation costs, and enables more accurate predictions. GNN has advantages in extracting associations of instances, and can effectively improve tracking accuracy in multi-target tracking.

**Transformer-based model**  Transformer-based model was first applied in the field of natural language processing. It has been successfully applied to image recognition, target detection, segmentation, super-resolution and other fields (Khan, Naseer, Hayat, Zamir, Khan, Shah, 2021). Transformer's success factors are the self-supervision mechanism and self-attention mechanism. For
example, (Chen, Yan, Zhu, Wang, Yang, Lu, 2021) proposed a new attention-based feature fusion network, which effectively combines the template and search area features. The network includes an ego-context augment module based on self-attention and a cross-feature augment module based on cross-attention. Specific to multi-fish tracking, the effect of introducing the Query-Key mechanism is very poor, and the problem of missed detection is obvious, because the characteristics of the new target appearing in the current frame certainly do not exist in the query, and the key cannot be obtained, resulting in the loss of the new target. When the detection and Re-ID tasks are merged into the same network, the efficiency of the model can be significantly improved. Transtrack (Sun, Jiang, Zhang, Xie, Cao, Hu, Kong, Yuan, Wang, Luo, 2020) adopts the JDE paradigm for multi-object tracking, introduces transformer to the MOT field for the first time. It uses the Query-Key mechanism to design a learning branch to query the target position from the key for detection. And in tracking branch, the target position in the current frame is queried from the key to obtain tracking boxes, and finally uses IOU matching is performed to complete the tracking. It has a faster processing speed, but it still lacks in ID switch (IDS) indicators. (Sun, Jiang, Zhang, Xie, Cao, Hu, Kong, Yuan, Wang, Luo, 2020) introduced a tracking-by-attention paradigm called TrackFormer, achieved detection and data association. The method used autoregressive track queries, embed an object’s spatial position and identity to track it in space and time. (Anwar, Li, 2020b) proposed the module named Global Context Disentangling (GCD), to decouple the features as detection-specific and ReID-specific one. And they used the encoder and deformable attention of transformer to develop the Guided Transformer Encoder (GTE) module, consider the global semantic relation, which is usually ignored.

4 Bottlenecks, Metrics and Applications

In practical applications, one example system designed (as shown in Figure 9) is used to track fish targets. Cameras are set up at different positions of the tank to shoot fish behavior videos, and the data is synchronized to the computer for further processing.

![Figure 9: Tracking system for fish tank](image-url)
4.1 Processing of Occlusion and Multi-Scale

Networks Design Deep learning technology can effectively handle scale issues. For example, Faster R-CNN (Anwar, Li, 2020b) uses anchors of different areas, and Feature Pyramid Network (FPN) combines deep feature maps with strong semantic information and shallow feature maps with high resolution. (Zhou, Koltun, Kr A Henb U Hi, 2020) follows the Anchors-Free detection method of CenterNet(Zhou, Wang, Kr A Henb U Hi, 2019). It utilizes the heatmap to extract to center of the object, which has better robustness for the various scales of interested targets. The model achieves a balance between speed and accuracy, and can be introduced to the 3D applications. The Anchors-Free tracking method avoids the frequent changes of scale, and corrects the matching error of overlap. What’s more, the distribution of high density can also be effectively identified, which will help solving the condition such as feeding and scared.

Multi-view Design The commonly used methods to deal with the occlusion problem are based on the video capture method of the master-slave camera, or using the camera with mirror, etc. (Xiao, Fan, Mao, Cheng, Zhong, Li, 2016) utilizes a camera and a mirror to obtain the 3D coordinates of fish targets. It calculates the center coordinates of the fish body through a multi-object correlation algorithm, and combined the 2D plane information obtained by the mirror or the depth information to calculate the 3D coordinate point of fish targets. In the calculation process, the quantitative comparison of the difference between fish amount of the two perspective images is performed to determine whether the occlusion occurs. Although it can deal with the occlusion problem of physical or virtual fish better, there is still a deviation of the center position of the corresponding relationship, and the information is underutilized. (Jia-Fa, Gang, Wei-Guo, Liu, 2015) utilized the principle of plane mirror to derive a theoretical model of parallel occlusion tracking for both physical fish and virtual fish. The method realizes real-time 3D tracking of unobstructed fish. What’s more, it achieves higher accuracy with less data. However, it uses only a single camera, which makes the processing of occlusion more cumbersome, and robustness is relatively poor. (Wang, Liu, Zhao, Liu, Chen, 2016) proposed a limited main view tracking method based on the master-slave camera structure, using Eyes-focused Gaussian Mixture Model (E-GMM) for fish extraction, and 2D main view for tracking. The eye focus detector is used to locate the slave view and correlate the data to reconstruct the three-dimensional trajectory. This method requires two slave cameras and the amount of data collected is relatively large. (Liu, Yue, Shi, Qian, 2019) proposed a 3D tracking method based on fish bones. In the skeleton analysis, the top view and side view are used to simplify the fish target into the representation of feature points, which are associated with the movement trajectory of the top view. The continuity constraint matches the feature points of the two views to obtain a 3D trajectory, but it has limitations due to using asymmetric strip structure. In conclusion, the calculation and facility layout with the help of master-slave cameras or mirroring are complicated. It is more favored by researchers to predict and estimate from single-camera video.
In this subsection, we summarized the evaluation indicators used for multiple fish tracking. MOT Challenge (Dendorfer, Rezatofghii, Milan, Shi, Cremers, Reid, Roth, Schindler, Leal-Taix E, 2020) provides multi-target pedestrian datasets, its classic MOT16, MOT17, MOT20 datasets are most commonly used benchmarks. The typical measurement indicators of the MOT algorithm are given, such as: MOTA, IDF1, HOTA, MT, ML, etc. Evaluation index is shown in Table 4.

### MOTA
Multiple object tracking accuracy (MOTA) is an indicator to measure the accuracy of single-camera multi-target tracking, and the formula is expressed as:

\[
MOTA = 1 - \frac{\sum_t(FN_t + FP_t + IDSW_t)}{\sum_t GT_t}
\]  

\(FN\) is false negative, \(FP\) is false positive, \(IDSW\) is identification switch, and \(GT\) is the number of ground truth objects. MOTA considers errors of object matching throughout the whole frames, which is an intuitive measure of the tracker's performance. At the same time, it is unrelated to the estimation accuracy of the object's position.

### IDF1
Identification-Score (IDF1) comprehensively considers identification (IDP) precision and identification recall rate (IDR), and the equation is shown as:

\[
IDF_1 = \frac{TP}{TP + 0.5FP + 0.5FN}
\]

\(TP\) is true positive. IDF1 can capture ID more accurately, to evaluate the tracker's ability to determine whether the trajectory is the same target.

### HOTA
Higher order tracking accuracy (HOTA) (Luiten, Ošep, Dendorfer, Torr, Geiger, Leal-Taixé, Leibe, 2020) is metrics making the evaluation scores closer to visual sense of human. The formulas about HOTA are listed as:

\[
HOTA_\alpha = \sqrt{\frac{\sum_{c} \left( TP(c) \right) A(c)}{TP + FN + FP}}
\]

\[
A(c) = \sqrt{TPA(c)TPA(c) + FNA(c) + FPA(c)}
\]

HOTA solves the problems of overemphasizing detection or correlation, which becomes a widely accepted interpretable evaluation indicators in multi-target tracking.

### Table 4: MOT Evaluation Measures

| Metrics | Description | Better |
|---------|-------------|--------|
| MOTA   | Multi-Object Tracking Accuracy, involving false positives, missed targets and identity switches. | Higher |
| IDF1   | The ratio of correctly identified detections over the average number of ground-truth and computed detections. | Higher |
| HOTA   | Higher-Order Tracking Accuracy. Geometric mean of detection accuracy and association accuracy. | Higher |
| MT       | Mostly tracked targets. The trajectory prediction covers those ground-truth accounts for more than 80% of the total. | Lower |
|---------|------------------------------------------------------------------------------------------------------------------|-------|
| ML      | Mostly lost targets. The trajectory prediction covers those ground-truth accounts for less than 20% of the total.  | Lower |
| Rcll    | Ratio of correct detections to total number of ground-truth.                                                      | Higher|
| ID Sw.  | Number of Identity Switches.                                                                                    | Lower |

4.3 Fish Behavior Analysis

The tracking methods finally generate the trajectory data of the fish, which marking the location of various fish individuals in each frame. In practical applications, biological knowledge is used to identify abnormal fish behavior trajectories. The key point is to establish the abnormal evaluation model, to judge whether the fish is currently in hunger, parasites, hypoxia or other abnormalities, which is of great significance in fishery production, marine environment monitoring, and biological research. For example, (Shreesha, MM, Verma, Pai, 2020) studied the three behavior patterns of Sillago-sihama and developed the Sillago-sihama decision support system based on the motion information obtained from the tracking model. However, the model is complex and requires a larger dataset for training to improve accuracy.

What’s more, fish have high sensitivity in smell and taste, and changes in temperature, and the behavior parameters are highly sensitive and relatively convenient to obtain. Salinity, pH, and dissolved oxygen under different water quality will cause abnormal fish behavior. By analyzing the behavior of fish, characteristic variables that respond to changes in water quality can be extracted, which is one of the significant solutions of water quality detection.

Finally, fish tracking technology also provides data support for the number, species, and behavior characteristics of fish for the ocean observation network. (Nian, Wang, Che, He, Xu, Li, Lendasse, 2017) proposed a method of combining variable fish bodies and compressed sensing, using multi-scale variable fish bodies to describe mixed fish bodies, and developed an ocean observation network portable device with online fish detection and tracking strategies.

5 Discussion

Although object tracking algorithms based on computer vision that have achieved the accuracy of human eye recognition to some extent, most of them focus on objects on land. Fish tracking based has received attention from smart aquaculture researchers. Because underwater vision has certain characteristics of poor visibility, non-rigid deformation, and high-frequency appearance changes under certain lighting as well as viewpoints, underwater fish tracking still has great development potential in the fields of biological research, fish farming, and environmental detection. This section mainly focuses on a brief analysis of the challenges in the fish field and the future development prospects.
5.1 Datasets

We analyze the problem of fish tracking datasets from two aspects: data acquisition and frame quality.

On the one hand, fish datasets are extremely rare due to the tiny scale and high density, which makes the label of images more challenging. The quality of open source fish datasets still cannot satisfy the requirements of the deep learning-based methods, especially in resolution and frames. There is still a lack of high-quality open source fish tracking datasets. Most researches used the self-made datasets for the experiment and model evaluation, which lacking of comparisons in commonality between various latest models.

On the other hand, the fish target datasets usually have low brightness and contrast, more noise, and serious color distortion problems, which makes greater challenges for the detection algorithms. The underwater image enhancement methods need to implement in the preprocessing steps for a precise detection, which makes the tracking result more reliable by inputting location with less offset. The general image enhancement technology has poor applicability in underwater data, and it is difficult in underwater images restoration. The GAN-based image enhancement method can effectively optimize the underwater datasets to highlight the contour and further interface higher quality trajectory results.

5.2 Methods

Although the JDE methods have been applied in the MOT tracking, the traditional two-step separation tracking models still dominate the fish tracking field. We summarized from the two stages of detector and tracker respectively.

In the detection stage, traditional methods already meet the location requirements due to the obvious shape, color characteristics of fish and the high discrimination from water. The traditional background difference method, inter-frame difference method or the combination of both are widely applied in the detection stage owing to their simple calculation. Based on the deep learning methods, the one-shot detector is more inclined to the real-time applications, which achieves the requirements of fishery production due to its balance of accuracy and speed, such as the YOLO and SSD. The one-shot algorithms gradually began to be applied to the detection stage to provide better detection results for the tracker.

In the tracking stage, the filters methods based on the kinematics are widely applied in fish tracking field. The typical tracking algorithms of SORT and DeepSORT still embedded in most tracking models. Although the deep learning-based methods have been applied in the detection stage, how to apply in tracking is still unresolved. Using data-driven models to inference the behavior trajectories will be a research hotspot for future development. Meanwhile, it is also considerable to reach the real-time prediction and maintain the accuracy as a result of the real-time requirements of fish tracking applications.

5.3 Occlusions, Accelerations and Scales

Occlusion is one of the key reasons for target miss in MOT. The dense distribution of fish leads to severe mutual occlusion. Compared with pedestrian’s re-identification, the individual characteristics of fish are more difficult to extract. Therefore, it is more difficult to retrieve the original target, resulting in frequent ID switch. The typical Re-ID methods for pedestrian are not applicable for fish tracking. For the short-term mutual occlusion, the fish individual kinematics
model should be established for matching the missing targets; for the problem of occlusion at a fixed point, using the master-slave cameras or camera with mirror to correlate the MTMCT information, and calculate the 3D coordinates.

Accelerations is often occurred when feeding or the fish scared, and explosive speed-up will cause target missing. The problem of target missing caused by acceleration can be effectively solved by recording higher frame rate video data. But this method will significantly increase the burden of the hardware device. In model training, the increase of frame will also lead to a decrease in training efficiency. Since the tracking model is often trained using the interval frame, it is also possible to obtain a better tracking result by adding the frames in these special scenes.

It is inevitable that the scale of fish is inconstant for the data captured by a fixed camera. Diverse anchors need to be used to identify targets at different distances. Using networks such as FPN, U-Net to extract multi-scale features can solve the problem of scale changes. What’s more, the no-rigid characteristic makes fish is deformed during movement and turning. It often uses different proportions of anchors or anchors-free methods to adapt the deformation. How to re-identify the target after scale change and deformation is still one of the bottlenecks in fish tracking.

6 Conclusion

In this paper, we summarized the open-source fish datasets, preprocessing methods, fish tracking methods, bottlenecks and applications. Because of the difficulties in labeling of the tracking information and acquisition of underwater data, there is a shortage of high-quality open-source fish tracking datasets. From the materials we have collected, traditional subtraction and filter are majority algorisms which applied in fish tracking, which based on the track-by-detection paradigm. The deep learning-based methods began to be applied in detection stage gradually with a higher accuracy. However, it is difficult to transfer the re-identification methods of pedestrians to small targets like fish, which is lack of visible features. We also analyzed the bottlenecks in fish tracking technology such as preprocessing, occlusion, scale and acceleration for future application and development. We hope that our work will provide more optimal innovation points for more optimization algorithms.

Acknowledgment

This study was supported by Key-Area Research and Development Program of Guangdong Province - Ecological engineering breeding technology and model in seawater ponds (2020B0202010009); National Key R&D Program of China - Integrated demonstration of seawater and brackish water fish ecological intensive intelligent breeding model and processing and circulation (2020YFD0900204).

References

Anwar, S., Li, C., 2020a. Diving deeper into underwater image enhancement: A survey. Signal Processing: Image Communication 89, 115978.
Anwar, S., Li, C., 2020b. Diving deeper into underwater image enhancement: A survey. Signal Processing: Image Communication 89, 115978.
Arvind, C.S., Prajwal, R., Bhat, P.N., Sreedevi, A., Prabhudeva, K.N., 2019. Fish detection and tracking in pisciculture environment using deep instance segmentation. IEEE, pp. 778-783.
Bochkovskiy, A., Wang, C., Liao, H.M., 2020. YOLOv4: Optimal Speed and Accuracy of Object
Detection.

Bras O., G., Leal-Taix E., L., 2020. Learning a neural solver for multiple object tracking, pp. 6247-6257.

Chen, X., Yan, B., Zhu, J., Wang, D., Yang, X., Lu, H., 2021. Transformer tracking, pp. 8126-8135.

Dai, P., Weng, R., Choi, W., Zhang, C., He, Z., Ding, W., 2021. Learning a proposal classifier for multiple object tracking, pp. 2443-2452.

Dendorfer, P., Rezatofghi, H., Milan, A., Shi, J., Cremers, D., Reid, I., Roth, S., Schindler, K., Leal-Taix E., L., 2020. MOT20: A benchmark for multi object tracking in crowded scenes. arXiv:2003.09003[cs].

Friedman, N., Russell, S., 2013. {Image Segmentation in Video Sequences: A Probabilistic Approach}. arXiv e-prints, 1302-1539.

Fu, X., Zhuang, P., Huang, Y., Liao, Y., Zhang, X., Ding, X., 2014. A retinex-based enhancing approach for single underwater image. IEEE, pp. 4572-4576.

Girshick, R., 2015. Fast r-cnn, pp. 1440-1448.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation, pp. 580-587.

Guo, D., Shao, Y., Cui, Y., Wang, Z., Zhang, L., Shen, C., 2021. Graph attention tracking, pp. 9543-9552.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE T PATTERN ANAL 37, 1904-1916.

Henriques, J.A.O.F., Caseiro, R., Martins, P., Batista, J., 2014. High-speed tracking with kernelized correlation filters. IEEE T PATTERN ANAL 37, 583-596.

Jia-Fa, M., Gang, X., Wei-Guo, S., Liu, X., 2015. A 3D occlusion tracking Model of the underwater fish targets. IEEE, pp. 82-86.

Jiang, X., Li, P., Li, Y., Zhen, X., 2019. Graph neural based end-to-end data association framework for online multiple-object tracking. arXiv preprint arXiv:1907.05315.

Jocher, G., Stoken, A., Borovec, J., NanoCode, Chaurasia, A., TaoXie, Changyu, L., V., A., Laughing, Tkianai, YxNONG, Hogan, A., Lorenzomammmana, AlexWang, Hajek, J., Diaconu, L., Marc, Kwon, Y., Oleg, Wanghaoyang, Defretn, Y., Lohia, A., Ah, M., Milanko, B., Fineran, B., Khromov, D., Yiwei, D., Doug, Durgesh, Ingham, F., 2021. {ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations}. Zenodo.

Khan, S., Naseer, M., Hayat, M., Zamir, S.W., Khan, F.S., Shah, M., 2021. Transformers in vision: A survey. arXiv preprint arXiv:2101.01169.

Lantsova, E., Voitiuk, T., Zudilova, T., Kaarna, A., 2016. Using low-quality video sequences for fish detection and tracking. IEEE, pp. 426-433.

Li, B., Wu, W., Wang, Q., Zhang, F., Xing, J., Yan, J., 2019. Siamrpn++: Evolution of siamese visual tracking with very deep networks, pp. 4282-4291.

Li, D., Wang, Z., Wu, S., Miao, Z., Du, L., Duan, Y., 2020. Automatic recognition methods of fish feeding behavior in aquaculture: A review. AQUACULTURE 528, 735508.

Li, J., Gao, X., Jiang, T., 2020. Graph networks for multiple object tracking, pp. 719-728.
Li, X., Wei, Z., Huang, L., Nie, J., Zhang, W., Wang, L., 2018. Real-time underwater fish tracking based on adaptive multi-appearance model. IEEE, pp. 2710-2714.

Liu, S., Li, X., Gao, M., Cai, Y., Nian, R., Li, P., Yan, T., Lendasse, A., Liu, T., Li, P., Liu, H., Deng, X., Liu, H., Zhai, F., 2018. Embedded online fish detection and tracking system via YOLOv3 and parallel correlation filter

Multi-class fish stock statistics technology based on object classification and tracking algorithm. IEEE, pp. 1-6.

Liu, T., Li, P., Liu, H., Deng, X., Liu, H., Zhai, F., 2021. Multi-class fish stock statistics technology based on object classification and tracking algorithm. ECOL INFORM 63, 101240.

Liu, W.A.A.D., 2016. SSD: Single Shot MultiBox Detector, Cham, pp. 21-37.

Liu, X., Yue, Y., Shi, M., Qian, Z., 2019. 3-D video tracking of multiple fish in a water tank. IEEE ACCESS 7, 145049-145059.

Luiten, J., O’sep, A., Dendorfer, P., Torr, P., Geiger, A., Leal-Taixé, L., Leibe, B., 2020. HOTA: A Higher Order Metric for Evaluating Multi-object Tracking. INT J COMPUT VISION 129, 548-578.

Mohamed, H.E., Fadl, A., Anas, O., Wageeh, Y., ElMasry, N., Nabil, A., Atia, A., 2020. MSR-YOLO: Method to Enhance Fish Detection and Tracking in Fish Farms. Procedia Computer Science 170, 539-546.

Naik, A.R., Swarnakar, A., Mittal, K., 2021. Shallow-UWnet : Compressed Model for Underwater Image Enhancement, pp. {}.

Nguyen, N.D., Huynh, K.N., Vo, N.N., Van Pham, T., 2015. Fish detection and movement tracking. IEEE, pp. 484-489.

Nian, R., Wang, X., Che, R., He, B., Xu, X., Li, P., Lendasse, A., 2017. Online fish tracking with portable smart device for ocean observatory network. IEEE, pp. 1-7.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics 9, 62-66.

Palconit, M.G., Almero, V.J., Rosales, M., Sybingco, E., Bandala, A., Vicerra, R., Dadios, E., 2020a. Towards Tracking: Investigation of Genetic Algorithm and LSTM as Fish Trajectory Predictors in Turbid Water, pp. {}.

Palconit, M.G., Almero, V.J., Rosales, M., Sybingco, E., Bandala, A., Vicerra, R., Dadios, E., 2020b. Towards Tracking: Investigation of Genetic Algorithm and LSTM as Fish Trajectory Predictors in Turbid Water, pp. {}.

Pinkiewicz, T., Williams, R., Purser, J., 2008. Application of the Particle Filter to Tracking of Fish in Aquaculture Research, pp. 457-464.

Redmon, J., Farhadi, A., 2017. YOLO9000: better, faster, stronger, pp. 7263-7271.

Redmon, J., Farhadi, A., 2018. YOLOv3: An Incremental Improvement.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection, pp. 779-788.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems 28, 91-99.

Ristic, B., Arulampalam, S., Gordon, N., 2004. Beyond the Kalman Filter: Particle Filters for Tracking Applications.

Salman, A., Maqbool, S., Khan, A.H., Jalal, A., Shafait, F., 2019. Real-time fish detection in complex backgrounds using probabilistic background modelling. ECOL INFORM 51, 44-51.

Sharif, M.H., Galip, F., Guler, A., Uyaver, S., 2015. A simple approach to count and track underwater
fishes from videos. IEEE, pp. 347-352.
Shijun, H., Yao, H., 2012. A study of fish velocity measurement base on video tracking, pp. 1898-1901.
Shreesha, S., MM, M.P., Verma, U., Pai, R.M., 2020. Computer vision based fish tracking and behaviour detection system. IEEE, pp. 252-257.
Shuai, B., Berneshawi, A., Li, X., Modolo, D., Tighe, J., 2021. SiamMOT: Siamese Multi-Object Tracking. pp. 12372-12382.
Singla, N., 2014. Motion Detection Based on Frame Difference Method.
Stankus, A., 2021. State of World Aquaculture 2020 and Regional Reviews: FAO Webinar Series. FAO Aquaculture Newsletter, 17-18.
Stauffer, C., Grimson, W.E.L., 1999. Adaptive background mixture models for real-time tracking. IEEE, pp. 246-252.
Sun, P., Jiang, Y., Zhang, R., Xie, E., Cao, J., Hu, X., Kong, T., Yuan, Z., Wang, C., Luo, P., 2020. Transtrack: Multiple-object tracking with transformer. arXiv preprint arXiv:2012.15460.
Wang, S.H., Liu, X., Zhao, J., Liu, Y., Chen, Y.Q., 2016. 3D tracking swimming fish school using a master view tracking first strategy. IEEE, pp. 516-519.
Wang, Y., Song, W., Fortino, G., Qi, L., Zhang, W., Liotta, A., 2019. An experimental-based review of image enhancement and image restoration methods for underwater imaging. IEEE ACCESS 7, 140233-140251.
Wu, J., Xiao, G., Zhang, Y., Chen, J., Zhu, L., Li, A., 2011. Fish group tracking based on delaunay triangulation network. IEEE, pp. 534-537.
Xiao, G., Fan, W., Mao, J., Cheng, Z., Zhong, D., Li, Y., 2016. Research of the Fish Tracking Method with Occlusion Based on Monocular Stereo Vision, pp. 581-589.
Zhou, X., Wang, D., Kr A Henb U Hl, P., 2019. Objects as points. arXiv preprint arXiv:1904.07850.
Zhou, X., Koltun, V., Kr A Henb U Hl, P., 2020. Tracking objects as points. Springer, pp. 474-490.
Zivkovic, Z., 2004. Improved adaptive Gaussian mixture model for background subtraction. IEEE, pp. 28-31.