

Fish Tracking, Counting, and Behaviour Analysis in Digital Aquaculture: A Comprehensive Survey

Meng Cui¹, Xubo Liu¹, Haohe Liu¹, Jinzheng Zhao¹, Daoliang Li², Wenwu Wang¹

Abstract—Digital aquaculture leverages advanced technologies and data-driven methods, providing substantial benefits over traditional aquaculture practices. This paper presents a comprehensive review of three interconnected digital aquaculture tasks, namely, fish tracking, counting, and behaviour analysis, using a novel and unified approach. Unlike previous reviews which focused on single modalities or individual tasks, we analyse vision-based (i.e. image- and video-based), acoustic-based, and biosensor-based methods across all three tasks. We examine their advantages, limitations, and applications, highlighting recent advancements and identifying critical cross-cutting research gaps. The review also includes emerging ideas such as applying multi-task learning and large language models to address various aspects of fish monitoring, an approach not previously explored in aquaculture literature. We identify the major obstacles hindering research progress in this field, including the scarcity of comprehensive fish datasets and the lack of unified evaluation standards. To overcome the current limitations, we explore the potential of using emerging technologies such as multimodal data fusion and deep learning to improve the accuracy, robustness, and efficiency of integrated fish monitoring systems. In addition, we provide a summary of existing datasets available for fish tracking, counting, and behaviour analysis. This holistic perspective offers a roadmap for future research, emphasizing the need for comprehensive datasets and evaluation standards to facilitate meaningful comparisons between technologies and to promote their practical implementations in real-world settings.

Index Terms—Digital aquaculture, fish tracking, counting, behaviour analysis, multimodal fusion

I. INTRODUCTION

With the expansion of the global population and the degradation of the ecological environment, traditional fishing (i.e. capture fisheries) is no longer capable of meeting the growing human demand for fish products [1], [2]. Aquaculture has become the primary source of fish acquisition, and digital aquaculture is emerging as a promising approach to enhance the efficiency and sustainability of the industry [3].

As enabling technologies in digital aquaculture, fish tracking, counting, and behaviour analysis have been studied extensively. Traditional methods for fish tracking and behaviour analysis rely on the experience of human observers, and the observation results depend on their skills and knowledge, which are not always reliable [4], [5]. Similarly, manual fish counting methods involve removing fish from tanks, leading to stress, injury, and disease, negatively impacting fish welfare

M. Cui, X. Liu, H. Liu, J. Zhao, and W. Wang are with the Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, Guildford GU2 7XH, UK. (e-mail: [m.cui, xubo.liu, haohe.liu, j.zhao, w.wang]@surrey.ac.uk).

D. Li are with the National Innovation Center for Digital Fishery, China Agricultural University, China (e-mail: dliangl@cau.edu.cn).

and growth [6], [7]. The implementation of intelligent tracking, counting, and behaviour analysis technologies can help overcome these limitations, reducing the risk of fish mortality, improving feeding strategies, and promoting sustainable development in aquaculture [8]–[10]. This paper aims to provide a comprehensive survey of the research problems, existing technologies, their connections, and potential gaps in this field.

There are already literature reviews conducted previously in this field, such as [9], [11]–[15]. However, most of them focused on individual tasks or single modalities, primarily emphasizing computer vision technology as the main approach. In practice, these tasks are interconnected and form a coherent part of an overall digital aquaculture system. For example, accurate monitoring of these tasks is vital for detecting abnormal fish behaviour, estimating fish abundance, and formulating effective management strategies, ultimately improving fish welfare and economic outcomes [16]. The narrow focus in previous reviews is inherently limited in capturing the interconnected nature of these tasks in real-world aquaculture scenarios. Different from these reviews, our focus is not only on surveying the progress made for these tasks, but also on establishing the connections among them.

Various technologies such as vision-based, acoustic-based and biosensing-based methods are used for fish tracking, counting, and behaviour analysis in aquaculture. Vision-based sensors and computer vision technology have found widespread application due to advancements in optical imaging and computer vision. However, they are limited by poor illumination, low contrast, high noise, fish deformation, frequent occlusion, and dynamic backgrounds [14], [17]–[19]. Acoustic-based sensors and hydroacoustic methods, which are non-invasive, are beneficial for monitoring fish in turbid water environments and overnight, but their high hardware cost limits their popularity in intensive aquaculture settings [20]–[24]. Biosensors can provide valuable information on fish physiology and behaviour, but their invasive nature and the need for individual fish tagging can be challenging in large-scale aquaculture operations and may potentially harm the fish [25].

A systematic review of literature from 2000 to 2023 reveals significant growth in research related to these technologies. Annual publications have increased from around 200 in the early 2000s to over 2,000 in recent years. Computer vision techniques dominate the field, accounting for about 40% of the publications, followed by acoustic monitoring methods and behavioural analysis approaches, each representing about a quarter of the research outputs. The geographical distribution

of research and methodological preferences show interesting patterns. Asia, particularly China and Japan, leads in publications, contributing nearly half of the global research outputs, mostly on computer vision techniques in tracking and counting, driven by the prevalence of large-scale intensive aquaculture operations and robust technological infrastructure. Europe, contributing about 30% of publications, emerges as a leader in acoustic monitoring and behavioural analysis techniques. This trend aligns with the region's stringent welfare regulations and emphasis on environmental sustainability in aquaculture practices. North America, while producing fewer publications, pioneers the integration of machine learning with various monitoring methods, leveraging its strong artificial intelligence (AI) research base and focus on automation. However, this global research landscape also reveals significant gaps. Developing regions, including parts of Africa, Asia, and Latin America, show limited adoption of advanced monitoring technologies. This gap can be attributed to the high initial costs, the lack of technical expertise, and the predominance of small-scale farming in these areas. Oceania, despite its significant aquaculture industry, is underrepresented in research outputs, possibly due to a smaller research community and a focus on practical applications over academic publications. Eastern Europe and Central Asia also show fewer publications on novel monitoring technologies, likely due to the later development of intensive aquaculture in these regions and limited research funding.

To address these limitations, our paper takes a holistic approach by systematically surveying a range of technologies, including vision-based sensors, acoustic-based sensors, biosensors, and hydroacoustic methods. We consider fish tracking, counting, and behaviour analysis as interconnected components of a unified task in digital aquaculture. This comprehensive perspective facilitates a more integrated discussion of these tasks while identifying technology gaps in the current literature. Furthermore, we explore the potential of emerging technologies such as multimodal data fusion, deep learning, and multi-task learning models that can simultaneously handle multiple aspects of fish monitoring. We also discuss the integration of large language models (LLM) and their potential to enhance understanding and decision-making in digital aquaculture. By adopting this multifaceted approach, our review aims to provide a more complete understanding of the challenges and opportunities in digital aquaculture, paving the way for developing more efficient, accurate, and integrated fish monitoring systems. The insights provided in this review have the potential to guide future research directions, inform policy decisions, and ultimately contribute to the sustainable growth of the aquaculture industry.

The remainder of this article is structured as follows: Section **II** explores the advancements in fish tracking techniques, while Section **III** discusses the various methods and applications of fish counting. Section **IV** discusses the behaviour analysis of fish, and Section **V** introduces the potential of multimodal fusion in aquaculture. Section **VI** presents an overview of relevant public datasets. In Section **VII**, we examine the challenges faced by the aquaculture industry and discuss future development trends. Finally, Section **VIII** summarises the key

findings and conclusions presented in this paper.

II. FISH TRACKING

Vision-based multi-target tracking methods are increasingly used in fish behaviour analysis. However, fish tracking is challenging because of the small differences between individuals, complex environments, and variations in plankton, shapes, angles, and scales of swimming fish [26]. Fish tracking can be categorized into two-dimensional (2D) and three-dimensional (3D) tracking based on the swimming environment [27]. 2D tracking is used in shallow water containers, where fish swimming appears close to a 2D planar motion and is represented using (x, y) coordinates. Still, it can only analyse a part of the fish's behaviour. In contrast, 3D tracking considers depth information and is represented using (x, y, z) coordinates, enabling the analysis of spatial movement in natural environments.

In addition to vision-based tracking, acoustic techniques such as the Acoustic Tag System (ATS) are also used for fish tracking. ATS involves attaching acoustic tags to fish, emitting unique acoustic signals detected by hydrophone receivers. The position of the tagged fish can be estimated using the time difference of arrival of the acoustic signals at multiple receivers, allowing for 3D tracking of fish movement in natural habitats [28]. This section mainly analyzes the relevant literature on fish tracking methods based on visual technology (as shown in Table **I**) and acoustic techniques in recent years and provides a systematic summary.

A. Fish tracking based on 2-dimensional visual information

Fish tracking methods can be broadly categorized into three main approaches: classical algorithms, kernel correlation filter-based algorithms, and deep learning-based tracking algorithms [15]. Each category encompasses various techniques with their strengths and limitations, which will be explored in more detail in the following subsections.

1) *Fish tracking based on classical algorithms*: Classical algorithms have been widely used to address the challenges of fish tracking in complex underwater environments, such as rapid posture changes, occlusion, overlap, and poor image quality. The Tracking-Learning-Detection (TLD) algorithm, which updates salient features and target model parameters through online learning, has shown promise in providing stable tracking [41]. However, its median-flow tracker may fail when fish change their swimming posture rapidly. However, an adaptive scale mean-shift (ASMS) algorithm, utilizing fish shape and colour features, can handle posture changes, uneven illumination, and complex backgrounds [42].

Preserving individual fish identities during occlusion and overlap remains a significant challenge. Techniques that extract head shape or body geometry features have been explored [43], [44], but their effectiveness may be limited by the rapid movement and intense geometry of fish bodies [18]. Adaptive thresholding algorithms, which estimate each pixel's threshold based on its adjacent region, have shown promise in segmenting individual fish in binarized images [45]. The global nearest neighbour algorithm with fish posture as a

TABLE I: Summary of different methods in fish tracking.

Study Site	Maximum Fish Amounts	Points	Detection Methods	Tracking Methods	Tracking Metrics	Advantages	Limitation	References
Tank	13	Fish head	YOLOv2	Kalman Filter	CIR ¹ CTR	High frame rate not necessary	Larger fish quantities increase identification losses	[29]
Coast	-	Center of the fish head	YOLOv4	Kalman Filter	MOTA ²	Real-time tracking with high accuracy	Accuracy affected by different sea areas	[30]
Tank	5	Fish head and center of the fish body	Background subtraction	Kalman Filter	CIR CTR	Accurate, fast, and computational inexpensive	Fail to predict the motion state of rapidly transitioning	[31]
Tank	20	Fish head and fish body	Background subtraction	Kalman Filter	CIR CTR	Smoother resulting trajectory	Lower frame speeds lead to more track breaks and higher misidentification	[32]
Tank	5	Centroid	Background subtraction	Kalman Filter	CIR CTR	Enhances tracking performance under occlusion conditions	Abnormal water quality leads to an increased chance of fish body overlap	[33]
Tank	25	Fish head	DOH ³	CNN	Recall	Corrects trajectory errors, fills gaps, and evaluates credibility	Easy affected by floating objects, ripple reflections, fish sharp turns	[34]
Tank	10	Head feature point and central feature point	Background subtraction	Feature point matching	Precision Recall	Two-feature point model reduces tracking difficulty	Only traces a few objects for a very short process	[27]
Glass Aquarium	5	The head and tail of fish	Adaptive thresholding algorithm	GNN ⁴	Tracking errors	Accurate tracking by pose constraint, even at high speed	Unable to handle fish occlusion or attaching	[35]
Fringing Reef, Red Sea	4	Fish's body	Fast-RCNN, Inception V2	Linking consecutive frames	3D detection rate	Cost-effective, automated 2D track Reconstruction	Small groups of fish studied	[36]
Tank	50	Head	ResNet-101	Mahalanobis distance and cosine similarity	MOTA IDF1 ⁵	Performance well under multiple negative factors	Bad performance of long-term tracking	[37]
Tank Pond	50	Body	Transformer	Hungarian algorithm	MOTA IDF1	Accommodates individuals with significant appearance variations.	Limitations in accurate ID matching at high stocking densities (over 50 fish)	[38]
Tank	8	Head	LSTM	Kalman Filter	Precision Recall	Cross-view more robust in high densities	Multi -view map matching is difficult, and the calculation amount is large	[39]
Tank	49	Head	DoH CNN	Iterative tracking strategy	Precision Recall	Tracking individuals exhibiting frequent occlusions	Requires individuals to have at least one body part that remains robust	[40]

¹ CTR (correct tracking rate), CIR (correct identification reason)² MOTA (multiple objects tracking accuracy) (as shown in formula 1)³ DOH (determinant of Hessian)⁴ GNN (global nearest neighbour)⁵ IDF1 (identification-score, as shown in formula 3)

tracking constraint has been used to track small numbers of zebrafish [35], but it lacks individual recognition ability, leading to track exchanges during overlap or occlusion. The Toxld algorithm addresses this issue using intensity histograms and Hu-moments to link trajectory fragments and preserve individual fish identities [46]. However, with this method, the error increases with the number of fish.

To deal with poor image quality, retinex (MSR) based enhancement algorithms combined with object detection have been used to improve fish detection in unclear underwater images [47]–[49]. Kalman filters may not always be optimal for underwater fish tracking due to the presence of non-Gaussian noise and complex environments [50]. Mean offset technology, which models fish probability density based on colour histograms, can fail when the background colour closely resembles the fish colour distribution. Tracking algorithms based on covariance matrices of pixel-based feature sets, incorporate spatial and statistical characteristics, making them more suitable for tracking fish in challenging underwater environments [51], [52].

2) *Fish tracking based on Kalman filters*: The Kalman filter is an efficient autoregressive filter, which can estimate the state of a dynamic system from sensory measurements taken from an environment with uncertainties posed by noise and interferences [53]. However, when applied alone to complex tracking scenarios, it faces challenges in handling occlusions and maintaining accurate tracking in high-density situations [32]. To address these limitations, researchers explored combining Kalman filtering with advanced object detection methods. You Only Look Once (YOLO) [54] has emerged as a leading detection method due to its real-time performance and accuracy in identifying fish within video frames. Building upon these foundations, the Simple Online and Realtime Tracking (SORT) algorithm [33] integrates YOLO-based detection results with Kalman filtering for motion prediction and employs the Hungarian algorithm for data association [55]. This integration enables high-speed tracking (260 FPS) by focusing solely on bounding box information, making it ideal for applications requiring fast and efficient tracking. However, the performance of SORT is limited by its reliance on pure detection results without considering object appearance features, making it vulnerable to tracking failures during occlusions or complex scenarios [56].

To address these limitations, DeepSORT was developed by incorporating a convolutional neural network (CNN) model trained on large-scale datasets [57]. This advancement significantly improved tracking robustness to target loss and occlusions, making it particularly suitable for aquaculture applications [29]. For instance, Mathias et al. [54] successfully demonstrated this capability by implementing a hybrid adaptive DeepSORT with YOLOv3 [58] for underwater object tracking. However, challenges arise when fish undergo rapid body shape changes during fast turns, leading to blurry and difficult-to-track images [31]. To mitigate this issue, shorter exposure times and boundary boxes of variable size can be used, with the boundary boxes being estimated according to the motion state. Additionally, frame rate optimization has proven crucial for Kalman filter-based tracking performance:

higher rates facilitate more linear fish motion predictions, while lower rates increase misidentification risks (as shown in Fig. 1).

Recent advancements in tracking algorithms have introduced several innovative approaches to address challenges in occlusions, rapid movements, and varying environmental conditions. One major category focuses on feature enhancement strategies: StrongSORT improved upon DeepSORT by incorporating stronger appearance features through a more sophisticated feature extractor based on ResNet-50 backbone networks and enhanced motion cues with Kalman filter refinements [59]. Its integration with GN-YOLOv5 [60] further optimized both performance and processing speed through better feature extraction and more accurate object detection [61]. Another category advances association strategies: BoT-SORT innovated tracking through a redesigned association mechanism that combines cosine distance metrics (comparing new detections' features with stored tracklet features) for appearance-matching and Camera Motion Compensation (CMC) for handling camera movements [62]. Unlike StrongSORT which primarily relies on appearance features, BoT-SORT introduced a hybrid association strategy that balances appearance similarity with motion consistency. This dual approach has proven particularly effective in challenging scenarios like aquaculture, where combined with YOLOv8 [63], it successfully tracks fast-moving fish despite rapid movements and appearance changes by better leveraging both motion and appearance information [64].

The third category introduces observation-centric (OC) approaches, where OC-SORT and Deep OC-SORT target dense population scenarios by prioritizing actual detection results over predictions, unlike traditional trackers that rely heavily on predictive models [65], [66]. OC-SORT prioritizes high-quality observations and introduces an observation-centric Kalman filter (OCKF) for more reliable state estimation, along with a new track recovery strategy and adaptive track confidence mechanism [65]. Deep OC-SORT further enhances this approach by incorporating a deep association module for better feature matching, an enhanced track rebirth mechanism, and more sophisticated appearance modelling [66]. Building on these approaches, ByteTrack pioneered a detection utilization strategy that processes both high and low-confidence detections through separate association pipelines [67]. This differs from previous approaches that typically discarded low-confidence detections, enabling ByteTrack to maintain tracking continuity even under challenging conditions. Its success in aquaculture applications [68], [69] demonstrates robust performance under varying lighting conditions while effectively reducing object loss and trajectory fragmentation, with its flexible architecture supporting various model sizes for different application scenarios [70].

The evolution of tracking algorithms from SORT to recent approaches like StrongSORT, BoT-SORT, OC-SORT, and ByteTrack offers diverse solutions for different tracking scenarios. For practical applications, StrongSORT is suitable when strong feature extraction is crucial, BoT-SORT works well with moving cameras and appearance changes, while OC-SORT and ByteTrack are particularly effective in crowded scenes. Future research could focus on combining these

strengths, such as integrating robust feature extraction with observation-centric approaches, while also addressing computational efficiency to enable real-time tracking in resource-constrained environments.

3) *Fish tracking based on deep learning*: Deep learning has revolutionized fish tracking approaches, introducing various methodologies to address the complex challenges in aquaculture environments. These approaches can be broadly categorized into three main streams: (1) identity-based methods like idTracker [71] that focus on extracting unique fingerprint features for individual fish recognition, (2) detection-tracking frameworks that combine CNN-based detection with sophisticated tracking mechanisms, and (3) advanced tracking architectures, such as Siamese networks that focus on similarity learning between frames and transformers that leverage self-attention mechanisms. While these methods have shown significant improvements over traditional approaches, they each face distinct challenges in handling occlusions, maintaining tracking stability in dense populations, and adapting to varying environmental conditions [41], [72].

Notable identity-based tracking algorithms such as idTracker [71] and its upgraded version, idtracker.ai [73], have demonstrated success in individual fish tracking. idTracker pioneers the concept of extracting unique visual fingerprints for each animal, utilizing image-based features to maintain individual identities throughout videos. The upgraded version, idTracker.ai, enhances this approach by incorporating deep learning techniques. These algorithms extract unique fingerprint features from each animal and identify targets throughout the video, enabling automated tracking of untagged animals within groups. However, their application is often limited to controlled environments where fish movement is restricted to prevent overlapping, making them less suitable for real-world 3D tracking scenarios.

Detection-based tracking frameworks have achieved significant advancements by combining CNN-based detection methods with complementary tracking techniques. These approaches integrate with other techniques, such as head detection, motion state prediction, and SVM-based verification [40], showing improved robustness compared to idTracker in scenarios with higher fish density and increased occlusion frequency [34]. Despite the progress made in controlled laboratory environments, real-world marine environments pose additional challenges for fish tracking, such as light fluctuations and waves. To address these issues, researchers have developed methods like the real-time multi-class fish (RMCF) stock statistics method, which uses YOLOv4 [74] as the backbone network and adopts a parallel two-branch structure based on deep learning for detecting fish species, tracking, and counting fish [30]. Although these methods have shown promising results in complex marine environments, their recognition accuracy may vary in different sea areas due to differences in colour cast and contrast, necessitating the retraining of the network weight coefficients.

Advanced tracking architectures in fish tracking have evolved along two main directions: Siamese networks and transformer-based methods [75], [76]. Siamese network trackers have gained attention in recent years due to their ex-

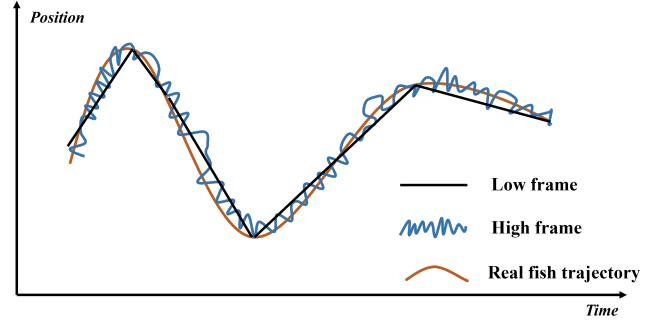


Fig. 1: Fish trajectory under different frame rates (This figure was adapted from [11]).

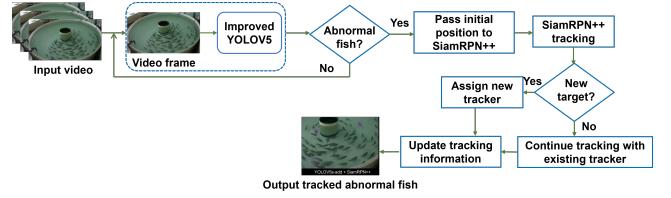


Fig. 2: Fish tracking method based on YoloV5 and SiamRPN++ (This figure was reproduced from [78]).

ceptional tracking speed and high accuracy. The introduction of advanced algorithms, such as SiamRPN++ (as shown in Fig. 2), has further demonstrated the performance of Siamese networks, surpassing the performance of the tracking algorithms based on correlation filters [77], [78]. Although there are currently few articles on Siamese networks specifically for fish tracking, this approach has the potential to be used in the field.

The emergence of transformer-based tracking methods has revolutionized object tracking, including fish tracking in aquaculture environments. Initially proposed for natural language processing tasks, transformers have been successfully adapted for computer vision tasks, including object detection and tracking [79]. Transformer-based trackers, such as TransTrack [80] and STARK [81], have demonstrated state-of-the-art performance on various tracking benchmarks. In the context of fish tracking, transformer-based methods have shown promising results. For instance, Li et al. [38] proposed a transformer-based multiple fish tracking model (TFMFT) to address the issue of instance loss in aquaculture ponds with complex background disturbances. This approach represents a shift from previous tracking-by-detection (TBD) methods [41] that relied on separate detection and embedding paradigms. Instead, the end-to-end Transformer architecture enables unified processing for both training and tracking. Building upon this work, Liu et al. [82] introduced FishTrack, an online multi-fish tracking model, that incorporates a specially designed encoder to encode historical position information of fish and automatically update spatiotemporal information in an autoregressive manner. This approach allows for the fusion of spatiotemporal information of fish targets without manual feature selection, demonstrating robustness even under long-term occlusion conditions.

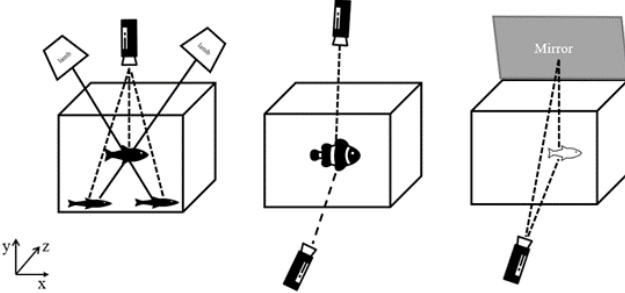


Fig. 3: Three methods to measure the 3-D position of a fish in an aquarium (This figure was adapted from [11]).

Despite these advancements, significant challenges remain across different tracking approaches. Identity-based methods like idTracker.ai [73] still face limitations in handling complex 3D environments and high-density scenarios. Detection-tracking frameworks struggle with environmental variations and often require retraining for different conditions, while Siamese networks, although promising, need further validation in real-world aquaculture settings. Transformer-based methods face computational complexity issues and require substantial training data, which can be scarce in aquaculture settings. They may also struggle with extended tracking sequences and lack inherent temporal modelling capabilities. Specific challenges include achieving accurate ID matching at high stocking densities and maintaining tracking stability during long-term monitoring. Furthermore, the potential for overfitting when the training data is limited or not diverse enough is a particular concern in aquaculture, where environmental conditions can vary widely. For example, TFMFT faces limitations in achieving accurate identity (ID) matching at high stocking densities and experiences some tracking losses during long-term monitoring.

Future research should focus on further adapting transformer architectures to address specific challenges, such as developing efficient training strategies to handle the limited availability of annotated fish-tracking datasets, improving performance in high-density scenarios, and enhancing long-term tracking stability. Additionally, efforts should be directed towards developing more efficient and lightweight models, improving their ability to handle long-term tracking and occlusions, and creating strategies to effectively train these models with limited data.

B. Fish tracking based on 3-dimensional visual information

3-D tracking methods offer advantages over 2-D tracking algorithms, as they can be used to study the behaviour of social animals and effectively address occlusion problems. However, 3-D tracking also presents significant challenges due to the large number of fish, similar individual appearance, occlusion, and uncertainty of stereo matching.

Two main types of 3-D tracking methods have been developed: “shadow” and “stereo” methods (as shown in Fig. 3). The “shadow” method, which requires only one camera, uses the shadow of the fish projected onto the substrate as a second view of the shoal. By calculating the 2-D positions of

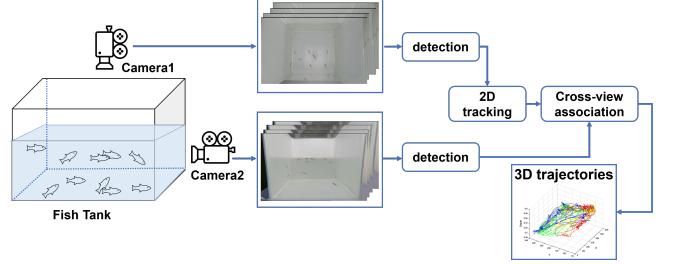


Fig. 4: Three-dimensional trajectory of multiple fish in water tank via multi video tracking (This figure was inspired by [86] with minor changes).

the fish and its shadow, the 3-D position of the fish can be obtained through triangulation. However, this method becomes increasingly difficult as the number of fish increases and shadows may be obscured, as it requires detecting each fish and its corresponding shadow [83].

Stereoscopic methods use multiple cameras, a camera and a mirror to capture simultaneous images at different angles [11]. Some researchers have developed platforms that use a single camera and mirror to obtain 3-D coordinates of fish [84], [85]. These methods calculate the centre coordinates of fish and combine the association of mirror view and direct view for tracking, addressing the problem of target loss caused by occlusion. However, they require high-precision equipment and may suffer from correspondence deviations due to the pixel centres of real and virtual fish not being at the same point. Moreover, these methods are not suitable for actual production environments.

In theory, two cameras are sufficient for stereo imaging. 3-D tracking with two cameras involves obtaining the 2-D motion trajectory from the top view with a larger viewing angle and then performing 3-D matching of the top view tracking results with the feature points in the side view to obtain movement of the object in 3-D space (as shown in Fig. 4) [86]. To track many objects, three or more high-speed cameras usually are needed to capture synchronous videos to mitigate the ambiguities and errors in distinguishing the objects. For example, Wang et al. [87] determined the location of fisheye under the top and side views using mixed Gaussian and Gabor models, respectively, and then obtained their 3-D motion trajectories by associating the top-view tracking results with the trajectories of two side views [39]. However, this method is limited due to the difficulty in distinguishing the characteristics from the eye area of fish. Furthermore, analyzing fish movement behaviour in three views requires the installation of complex equipment and the association and stereo matching between the views [88].

Occlusion remains one of the main challenges in 3D fish tracking, as it is in other Multiple Object Tracking (MOT) tasks. However, the frequency of occlusion has not been adequately measured in the current literature, with the complexity indicator of the datasets used in existing studies typically being the number of fish rather than an assessment of fish occlusion events. For instance, a demo video in [27] shows only 4 occlusion events within 15 seconds for a group of 10

fish. Current system evaluations assess parameters such as ID swaps, fragments, precision, and recall for the generated 2-D and 3D tracks without describing how these indicators are calculated. The lack of uniform indicators makes it difficult to fairly compare the methods presented in various studies. Furthermore, most of the literature does not provide open-source code and annotated data, limiting the reproducibility of the results. A recent study by [37], [89] introduced a standard MOT evaluation framework for fish tracking, providing a good model for multi-target fish tracking. A unified evaluation standard should be introduced to ensure the fairness of fish tracking comparisons and facilitate progress in this field.

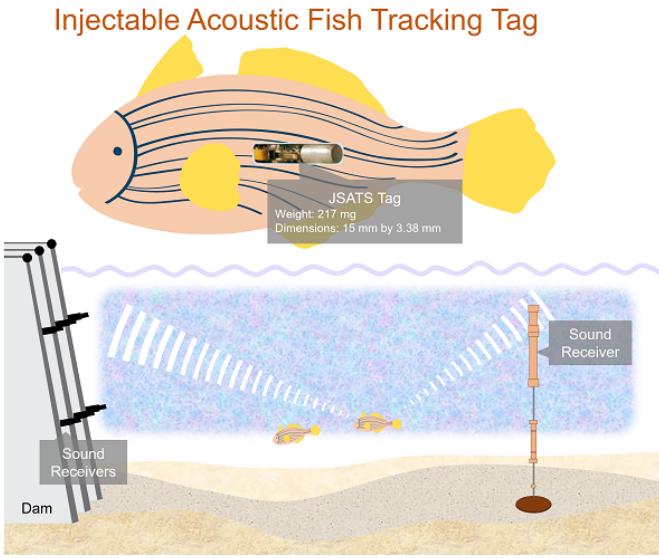


Fig. 5: Fish tracking method based on an acoustic tag (This figure was reproduced from [90] with minor changes).

C. Fish tracking based on acoustic tag system

The Acoustic Tag System (ATS), a passive acoustic monitoring technology, has become an important means of monitoring fish trajectories and studying fish behaviour [91]. Unlike vision-based tracking methods, which rely on clear water conditions and sufficient lighting, ATS can provide reliable tracking data in challenging underwater environments, such as turbid waters or low-light conditions [92]. The appropriate type and parameters of the acoustic tag (also called an acoustic signal transmitter) are selected according to the size of the fish and the research period (as shown in Fig. 5) [93]. The applications of ATS are diverse, including fish resource abundance assessment, swimming pattern analysis, habitat evaluation, spawning site identification, survival rate estimation, and behavioural differentiation [94], [95]. Kolarevic et al. [96] demonstrated the use of acoustic acceleration transmitter tags for monitoring Atlantic salmon swimming activity in recirculating aquaculture systems (RAS), while Fore et al. [97] found it feasible to monitor fish depth in real-time in commercial fish farms using acoustic telemetry.

Acoustic tag systems have been used to monitor fish movement and trajectories, obtain the three-dimensional coordinates

of the fish in real-time, and perform related data analysis and application. A notable example of an ATS application involves 682 subyearling Chinook salmon tagged with injectable acoustic transmitters in the Snake River, Washington State [90]. This study generated over 5 million clean detections and 403,900 3-D trajectories, helping to assign fish to passage routes through a dam and expanding understanding of near-dam fish behaviour [90]. Leclercq et al. [98] and Munoz et al. [99] evaluated the feasibility of using passive acoustic telemetry to monitor the welfare of fish in sea cages on an industrial scale, and the experimental results showed that the acoustic tracking system proved to be an effective tool for fish behaviour analysis. However, the application of acoustic tags in commercial aquaculture faces significant challenges. Macaulay et al. [100] systematically reviewed studies using tags to monitor farmed fish behaviour and found concerning results regarding tag-associated mortality. Mortality was 10 times higher in marine caged fish compared to tanks. Moreover, mortality in marine caged fish increased significantly in longer trials, ranging from 4% in single-day trials to 36% after 100 days.

Despite these challenges, ATS offers unique advantages in monitoring fish movement and obtaining real-time three-dimensional trajectory coordinates [101]. It excels in in-situ observation and simplifies data processing over vision-based methods. However, the technology's reliance on tags attached to fish can potentially impact on fish health and survival [102], [103], necessitating careful monitoring and prompt data analysis. Future research should focus on improving tag design, developing standardized reporting protocols, investigating mortality causes in marine caged fish, and exploring less invasive alternatives. By addressing these challenges, ATS could become a more viable tool for understanding and improving fish welfare and productivity in aquaculture, though its use should be limited currently and results interpreted cautiously in commercial settings.

D. Tracking evaluation metrics

Multi-target tracking evaluation indices directly reflect an algorithm's tracking ability. The MOT Challenge official multi-objective tracking evaluation indicators [104] provide a standardized framework for performance assessment. Key metrics include Multiple Object Tracking Accuracy (MOTA) and Multiple Object Tracking Precision (MOTP).

The MOTA combines three sources of errors to evaluate a tracker's performance, as follows:

$$MOTA = 1 - \frac{\sum_t (FN_t + FP_t + ID_{SW_t})}{\sum_t GT_t} \quad (1)$$

where FN_t , FP_t , ID_{SW_t} , and GT_t represent the number of false negatives, false positives, identity switches, and ground truth targets in frame t , respectively.

The MOTP is used to measure misalignment between annotated and predicted object locations, defined as:

$$MOTP = \frac{\sum_{i,t} d_t^i}{\sum_t c_t} \quad (2)$$

where d_t is the distance between the localization of objects in the ground truth and the detection output c_t is the total matches made between ground truth and the detection output.

Identification-Score (IDF_1) comprehensively considers Identification Precision (IDP) and Identification Recall (IDR) rate:

$$IDF_1 = \frac{TP}{TP + 0.5FP + 0.5FN} \quad (3)$$

where True Positive (TP), False Positive (FP), and False Negative (FN) involved in IDF_1 all consider ID, so the indicator is more sensitive to the accuracy of ID information.

To better capture the specific challenges of tracking fish populations, some literature has introduced additional metrics, such as Correct Tracking Ratio (CTR) and Correct Identification Ratio (CIR). CTR measures the percentage of correctly tracked frames for individual fish as follows,

$$CTR = \frac{\sum(\text{NumberOfCorrectFramesOfSingleFish})}{\text{NumberOfFish} \times \text{NumberOfFrames}} \quad (4)$$

while CIR represents the probability of correctly identifying all fish after an occlusion event:

$$CIR = \frac{\text{TimesThatAllFishGetCorrectIdentityAfterOcclusion}}{\text{NumberOfOcclusionsEvents}} \quad (5)$$

In addition to those metrics, tracking speed is another important factor to consider when evaluating fish-tracking algorithms, especially for real-time applications. Some common metrics for measuring tracking speed include frames per second (FPS) and processing time per frame. FPS indicates the number of frames a tracking algorithm can process in one second while processing time per frame measures the average time taken to process a single frame. Higher FPS and lower processing time per frame are desirable for efficient and real-time tracking performance. These metrics offer a valuable foundation for evaluating fish tracking performance, comprehensively assessing various errors, fish-specific challenges, and tracking speed. However, they may not always capture the full complexity of fish-tracking scenarios and can be limited by the lack of widespread adoption and the need for detailed ground-truth annotations. There is a potential scope to develop more specialized metrics and evaluation protocols by considering the specific requirements and challenges of fish tracking applications.

E. Comparative analysis of tracking methods

2D visual methods are ideal for controlled environments with clear water, offering a balance between cost-effectiveness

and non-invasiveness. However, they struggle in turbid or deep waters and occlusions in dense populations. 3D visual tracking provides more comprehensive data but requires more sophisticated equipment and setup, making it more suitable for research-intensive or high-value production environments. Acoustic tag methods excel in challenging water conditions but involve more invasive procedures that may impact fish welfare (as shown in Table II). The choice of a method often depends on the specific aquaculture environment. For example, shallow freshwater ponds might benefit more from 2D visual systems, while deep-sea cages may require acoustic or advanced 3D visual setups.

Regional preferences for tracking methods vary significantly. In Asia, 2D and 3D visual tracking is used predominantly, especially in countries like China and Japan [113], [114]. This preference is driven by the prevalence of intensive ponds, tank aquaculture, and strong technological infrastructure. A balanced approach is used in Europe, with significant adoption of all three methods. In North America, the integration of advanced AI with tracking systems is often employed, particularly by combining visual and acoustic tag methods for comprehensive monitoring [93]. In Oceania, acoustic tracking is favoured due to the prevalence of open-water aquaculture, especially in tuna farming [115]. Environmental conditions, regulatory frameworks, economic factors, and available technological infrastructure shape the trends for adoption. Developing regions in Africa and parts of Asia often lack access to advanced tracking technologies, highlighting the need for cost-effective solutions for small-scale farmers.

Animal welfare is an increasingly important consideration for the selection and development of tracking methods [116]. Visual methods generally cause minimal direct stress but may alter behaviour due to changes in lighting or the presence of equipment [117]. Acoustic tagging can cause initial stress during the tagging procedure and may affect long-term behaviour due to the presence of the tag [110], [112]. All methods can potentially cause stress if they require changes to the fish's environment or handling of the fish. Research on less invasive techniques includes the development of smaller, lighter acoustic tags to minimize the impact on fish movement and behaviour [3], advancements in computer vision and AI to improve tracking accuracy without the need for physical tags [118], and exploration of environmental Deoxyribonucleic Acid (DNA) or eDNA (Genetic material collected from environmental samples) methods for non-invasive population monitoring [119]. Balancing data collection needs with fish welfare involves implementing rotation systems in tagging

TABLE II: Comparison of fish tracking methods

Method	Advantages	Disadvantages	Welfare Implications	Suitable Environments
2D Visual [37], [105] [49], [78]	Non-invasive Cost-effective Real-time tracking	Limited to surface or shallow waters Affected by water turbidity Occlusion issues in dense populations	Minimal stress No physical interaction	Clear water ponds Shallow tanks Surface monitoring in sea cages
3D Visual [89], [106] [107], [108]	Accurate spatial tracking Works in deeper waters Better for dense populations	Higher computational requirements More complex setup Costly equipment	Minimal stress Potential light disturbance	Deep tanks Sea cages Research facilities
Acoustic Tags [109], [110] [111], [112]	Works in turbid waters Long-range tracking Effective in open waters	Invasive (if using tags) Potential signal interference Higher initial cost	Stress from tagging procedure Potential behavioral changes	Open sea farms Large lakes Deep water environments

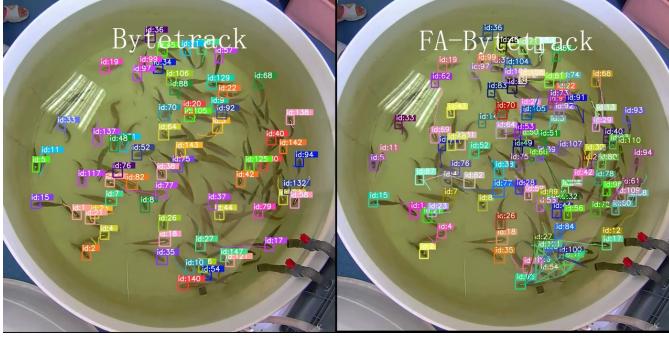


Fig. 6: Comparison of ByteTrack and FA-Bytetrack in controlled environment (This figure was reproduced from [69]).

studies to reduce the duration of tag attachment for individual fish, using hybrid methods to reduce the reliance on any single, potentially stressful technique, and developing guidelines for the ethical use of tracking technologies in aquaculture, considering factors such as tag size relative to fish size, duration of studies, and recovery periods [119], [120]. The aquaculture industry is trending towards methods that minimize stress and interference with natural behaviours while providing effective management data. This shift is driven by both ethical concerns and the recognition that improved welfare contributes to better product quality and farm productivity.

F. Real-world applications of fish tracking

Real-world applications of fish tracking primarily span two domains: commercial aquaculture monitoring in controlled tank environments and marine ecological research in natural underwater settings. While aquaculture applications focus on fish behaviour and health monitoring through overhead cameras, underwater tracking faces more complex challenges including varying visibility, coral occlusions, and unpredictable fish movements. A recent study by Zhao et al. [69] demonstrates the potential of advanced tracking and behaviour analysis techniques in aquaculture settings. The researchers developed a method combining an improved ByteTrack algorithm “Fish amendments (FA)-ByteTrack” with a spatiotemporal graph convolutional network (ST-GCN) to assess fish appetite in a controlled environment. In their experiment, they used a circular tank (1.5 m × 1.5 m × 1 m) with 120 large-mouth bass, each weighing 400-500 grams. This setup aimed to simulate the complex conditions of real aquaculture environments, including fish overlapping and occlusions. The system used a high-resolution camera (1920x1080 pixels) positioned 1.6 m above the water surface, recording at 60 frames per second. This method not only effectively solves the problems of bubble masking, intraclass variation and cross-occlusion but also achieves high-precision appetite assessment. Compared with that of ByteTrack, the average Multiple Object Tracking Accuracy (MOTA) of FA-ByteTrack increased by 21.3%. Moreover, the tool achieved 98.47% accuracy in appetite assessment (as shown in Fig. 6).

However, it is worth noting that while this study represents a significant advancement, it was conducted in a controlled

environment that does not fully reflect the challenges of real-world aquaculture settings. In commercial aquaculture operations, factors such as larger tank sizes, higher fish densities, varying water conditions, and different species could significantly impact the performance of such systems. Moreover, the study relied on a pre-trained YOLOv8 detector, which may need a large amount of high-quality labelled data for training detectors. Labelling them is time-consuming and costly.

The FISHTRAC dataset [105] exemplifies the challenges of real-world fish tracking in underwater environments, as shown in Fig. 7. Unlike controlled aquarium settings, natural underwater environments present unique challenges. Fish exhibit unpredictable movements, and rapidly changing appearances, and are frequently occluded by coral or other marine life. Moreover, fish often actively attempt to evade or hide from divers, adding another layer of complexity to the tracking task. A recent work by Mandel et al. [105] proposed a detection confidence-driven multi-object tracking to recover reliable tracks from unreliable detections. The study relied on a limited dataset of only 1800 images labelled as “fish” from the Google Open Images Dataset [121], many of which were not from real underwater environments. The limited scale of this dataset resulted in a detector with mediocre performance, highlighting the real-world challenges of applying deep learning methods to specialized domains like underwater fish tracking.

Despite these limitations, the Robust Confidence Tracking (RCT) algorithm [105] showed promising results. The key strength of RCT lies in its ability to fuse low-confidence detections with a motion model and single object tracker, maintaining consistent tracks even when high-confidence detections are sparse. This capability is especially valuable in challenging underwater conditions where high-quality detections are often unavailable. The study highlights the potential of RCT for practical applications in marine biology research and ecological monitoring, particularly in scenarios with scarce labelled data and challenging detection conditions. Advanced tracking algorithms based on Bayesian filters [122], such as particle filters [123] and Bernoulli filters [124], offer additional tools for trajectory smoothing and handling outliers or mis-detection. These methods, widely used in audio-visual speaker tracking, offer the potential to be applied in underwater environments. Future research directions include developing online versions of these algorithms for edge computing, improving performance in high-density scenarios through Bayesian incorporation of appearance information, and exploring adaptive approaches that adjust tracker behaviour based on detection quality. These advancements aim to bridge the gap between controlled studies and the complex realities of commercial aquaculture, paving the way for more robust and practical fish tracking and behaviour analysis systems.

III. FISH COUNTING

A. Fish counting methods based on sensor technology

Sensor-based counting devices are usually divided into resistance counters and infrared counters. Infrared counters detect infrared signals, which are electromagnetic waves with wavelengths between 760 nm and 1 mm [125]. Counting based

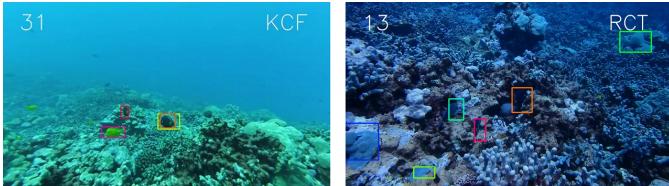


Fig. 7: FISHTRAC dataset in underwater environments (This figure was reproduced from [105]).

on infrared counters requires a tunnel structure to limit the movement of the fish. When a fish passes between the infrared transmitter and the receiver, the counting is completed [126], [127]. Although infrared sensors can count in small areas of space, their performance is affected by water depth and turbidity. At a depth of 17.9 centimetres (cm) in pure water, the intensity of the infrared light drops to 50% [128], and the presence of suspended particles can further degrade the performance of infrared counting at high turbidity levels [129]. In addition to environmental factors, the accuracy of infrared sensor counting devices is susceptible to the pass rate of fish, often resulting in an underestimation of the number of fish [130], [131]. This may be due to the slow swimming of some fish, the confusion when two or more fish enter the scanner unit simultaneously, and the reluctance of some fish to leave the device after entering the light tunnel, resulting in repeated scanning. Despite these limitations, infrared light can work in the dark, and the accuracy of counting can be improved by subsequent software algorithms, such as multiple object tracking (MOT) algorithms, which can solve false counting from multiple targets [129], [132], [133].

Resistivity counters, another type of sensor-based counting method, work by detecting changes in resistance when a fish passes between two electrodes [134], [135]. Like infrared sensor counting devices, electronic resistivity counters require the fish to pass through a specific tunnel and have similar disadvantages, such as repeating counts when a fish swims multiple times in the channel and missing counts when the number of fish is large [136]. However, electronic resistivity counters are suitable for limited lighting and long, narrow river channels while detecting non-destructively and without requiring specific lighting conditions [137].

Although both infrared and resistivity fish counters have limitations and may underestimate fish pass rates, they offer valuable tools for non-invasive fish counting in various environments. Future research should focus on developing and improving these technologies to enhance their accuracy and reliability. Potential avenues for improvement include modifying resistivity counters and exploring alternative sensors [138]. Researchers can provide valuable tools for effective fishery management and conservation efforts by addressing the current challenges and refining these sensor-based counting methods.

B. Fish counting methods based on computer vision technology

Accurate fish biomass assessment is crucial for optimizing management strategies and reducing feeding costs in the

aquaculture industry [139], [140]. Computer vision-based fish counting has gained prominence among various methods due to its non-invasive nature, low cost, and high efficiency [141], [142]. However, the complexity of underwater environments, including varying light conditions, backgrounds, and fish swimming patterns, poses challenges for accurate fish counting [143]–[145].

The current computer vision-based fish counting aquaculture methods focus on two main categories: image-based counting and video-based counting, as shown in Table. III. By bridging the gap between laboratory-based experiments and real-world applications, computer vision-based fish counting can become a useful tool for sustainable aquaculture management. The subsections provide more details about each category, the advancements, challenges, and future directions in the field.

1) *Image-based fish counting method:* Image-based fish counting methods can be broadly categorized into two main approaches: detecting-based methods, which aim to detect all fish in a region, and density-based methods [146], [147], which estimate the number of fish by analyzing the distribution of fish schools [148], [149].

Early studies focused on detecting-based methods, which rely heavily on the accuracy of fish image segmentation from the background [161]. These methods, such as artificial neural networks (BPNN) [162], showed potential for automatic fish counting in scenarios with a limited number of fish. However, they often struggled with complex adhesions in fish images and overlapping fish [156], [163]. To address the challenges of overlapping fish, adaptive segmentation algorithms were developed to extract the geometric features of fish [158]. Combined with machine learning models like LS-SVM, these algorithms showed improved counting accuracy compared to BPNN models, particularly in scenarios with similar fish sizes and low stocking densities. However, the performance of these models degrades when faced with high fish densities and changing geometric shapes due to fish overlap [164]. Further advancements in fish image segmentation were made by introducing more general adaptive thresholding methods and skeleton extraction-based methods to handle overlapping fish [165]. While these methods performed well under controlled laboratory conditions, their accuracy diminished in real-world aquaculture environments, where factors such as high fish school density, poor visibility, and insufficient light posed significant challenges [157].

Efforts to mitigate issues related to light, noise, and feature recognition led to the development of segmentation methods that combined local normalized filters and iterative selection thresholds [155]. Although these methods demonstrated high performance in correcting non-uniform lighting, reducing noise, and identifying features, the unique challenges posed by aquaculture settings, such as fish shadows caused by water refraction and continuous movement of shoals, continued to affect segmentation accuracy and limit the effectiveness of traditional computer vision methods for fish counting [156].

Introducing deep learning techniques has opened new avenues for fish counting in aquaculture. With the increasing availability of fish datasets, deep learning models have been

TABLE III: Different counting methods based on computer vision.

Study Cite	Fish amount	Dataset	Model	Count points	Evaluation index	Results	Advantages	Limitations	References
Tank	100	786	MAN ¹	Center	Accuracy	97.12%	Better generalization ability	larger error for areas with high fish density	[150]
Tank	-	4000	DG-LR ²	Fish-Connected Area	R ²³	96.07%	No need to detect every fish	No complex environments	[6]
Net Cage	214	1501	Hybrid Neural Network	Center points	Accuracy	95.06%	Improves model performance without losing resolution	Does not describe the distribution of fish school gathering and dispersing Repeating detection and wrong detection in high contrast areas	[151]
Cage	62	200	RCNN ⁴	Bounding Box	Accuracy	92.4%	Reduce count errors due to repeating detections	No detailed analysis of the number of fish in the system per unit time	[152]
Counter	1000	1500	Background Subtraction Kalman filter	Blob	Average precision	97.47%	Automatic counting, low cost	Threshold adapts to different numbers of fish	[153]
Containers	600	4000	CNN	Contours	Accuracy	99.17%	Pure white background, no noise	[154]	
Dishpan	100	-	Local Normalization Filter	Pixel Area	Accuracy F-measure	99.8% 98.83%	automated system.	Small sample size	[155]
Aquarium	350	1000	Background Subtraction	Contours	Accuracy	95.57%	Portable, low cost	Need a fixed size of fish and a certain area	[156]
Net Cage	250	1000	PTV ⁵	Centroid	Detection rate	90%	Potential application or industrial aquaculture	Affected by background noise sensitivity	[157]
Aquarium	100	600	LS-SVM ⁶	Skeleton	Accuracy	98.73%	Good generalization	Assume that the size of fish is similar	[158]
Aquarium	300	3200	MSENet ⁷	Centroid	MAE ⁸	3.33	Lightweight and low computation costs	limited to a scene with a fixed viewpoint	[159]
Long Channel	300	1318	YOLOv5-Nano	Bounding Box	Average precision	96.4%	Solves the problem of missing fish fry	Occlusion still causes some fish to be incorrectly detected	[160]

¹ MAN (Multi-modules and attention mechanism)² DG-LR (Image density grading and local regression)³ R2 (Coefficient of determination)⁴ RCNN (Region-based convolutional neural network)⁵ PTV (Particle tracking velocimetry)⁶ LS-SVM (Least squares support vector machine)⁷ MSENet (A lightweight network based on SENet)⁸ MAE (Mean absolute error)

applied to this domain, offering strong adaptability and easy transformation without requiring complex feature extraction work [152], [166]. Convolutional neural networks (CNNs) have achieved high accuracy in detecting and counting fish of different sizes by adjusting different thresholds [154].

Density-based methods, which estimate the number of fish by mapping input images to corresponding density maps, have also shown promise in fish counting applications. These methods provide additional information about the spatial distribution of fish, which can be valuable for various purposes [167]. Hybrid neural network models, such as those combining

MCNN and DCNN architectures, have been proposed to improve fish counting accuracy, outperforming traditional CNNs and MCNNs [142], [150].

Despite the advancements made in fish counting methods, several challenges remain. Density-based methods are sensitive to the degree of occlusion, with higher fish densities leading to greater errors. Moreover, variations in water quality, light conditions, camera angle, water depth, and surface refraction can cause significant differences in the appearance of fish across different farming environments, affecting the accuracy and generalization ability of the counting models. To

address these challenges, future research should create more comprehensive and diverse datasets that capture the variability encountered in real-world aquaculture settings. Efforts should also be directed towards improving counting accuracy, model generalization ability in high-density areas, and maintaining accuracy under different pond conditions.

2) *Video-based fish counting method*: Video-based counting methods offer a more comprehensive approach to enumerating fish than static image-based techniques, leveraging temporal information and movement patterns [168]. These methods can be broadly categorized into frame-by-frame analysis, tracking-based counting, and segmentation-based counting [9].

The frame-by-frame analysis based methods extend the image-based techniques to video sequences, applying detection and counting algorithms to each frame. Zhao et al. [33] proposed a method using background subtraction and object detection to count fish in video streams, achieving higher accuracy than single-image based counting methods due to the ability to average counts over multiple frames. Like this work, Pai et al. [169] used the YOLOv5 model [60] to detect and count fish in video streams and get similar results. This approach is relatively simple to implement and can leverage existing image-based algorithms. However, it does not fully use temporal information, may struggle with rapid fish movements between frames, and is prone to double-counting or missing fish due to occlusions.

Tracking-based counting methods follow individual fish across frames, offering insights into fish movement patterns and behaviours. In addition, the tracking-based counting methods can deal with the temporary occlusions, and reduce double-counting errors [30]. Albuquerque et al. [153] and Zhou et al. [170] developed a system combining object detection with multiple object tracking to count fish in dynamic videos for large-scale counting in practical production. This method uses segmentation and association, and can effectively solve the problem of mis-detection caused by adhesive fingerlings in mechanical counters (as shown in Fig. 8). However, in practical applications, these algorithms often struggle to achieve a balance between counting accuracy and processing speed, limiting their applicability in real-time commercial settings. Addressing this challenge, Zhang et al. [160] proposed a novel combination of the SORT algorithm [33] and YOLOv5-Nano [171] for fry tracking and counting. This innovative approach achieves accurate dynamic counting with even faster running speeds, representing a significant step towards real-time fish counting in aquaculture environments. Integrating lightweight deep learning models with efficient tracking algorithms showcases a promising direction for future research, potentially enabling more widespread adoption of video-based counting methods in commercial aquaculture operations.

Segmentation-based counting methods focus on isolating fish from the background in each frame and then counting the segmented objects. Instance segmentation methods have emerged as a promising solution for effectively segmenting individual fish in dense scenes, and combining these methods with advanced detectors has the potential to achieve real-time performance [172]. Wang et al. [173] propose a lightweight instance segmentation model based on YOLOv8 [63], realizing

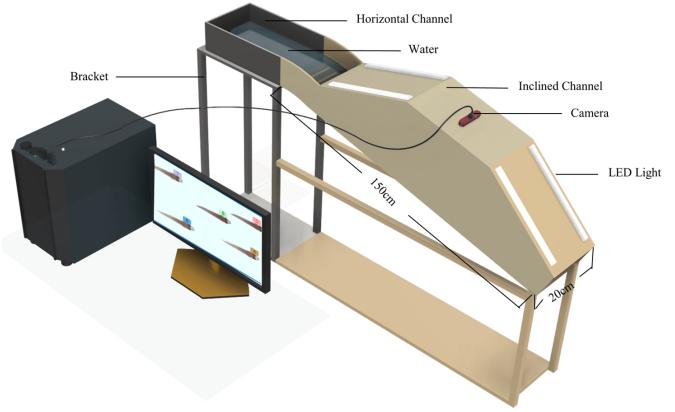


Fig. 8: Overview of the fry counting system. Fish is poured into a channel with water flow, LED lights provide stable illumination. An industrial camera records the fry swimming process and transmits it to a computer (or edge device), and the device counts the number of fry passed in real-time through the designed method.

high-density and multi-target fish segmentation and counting in an occluded environment. This approach demonstrates the potential of segmentation-based methods in handling complex aquaculture scenarios. However, the segmentation performance is notably diminished in scenarios of significant occlusions, presenting a significant challenge for dense fish populations. To address this limitation, recent advancements in computer vision, such as the Segment Anything Model (SAM) [174], offer new possibilities. The SAM model is pre-trained on a large-scale dataset and exhibits exceptional robustness in segmentation tasks across various domains. Although no literature is found for fish counting based on SAM, its excellent segmentation capabilities hold promise for future applications in aquaculture. Looking ahead, integrating the robust segmentation abilities of SAM with specialized fish counting algorithms could potentially overcome the limitations of current methods, particularly in highly occluded environments. It could be a promising direction to adapt SAM for fish-specific segmentation tasks and optimise its real-time performance to meet the demands of commercial aquaculture operations.

C. Fish counting methods based on acoustic technology

Acoustic technology for fish counting can be divided into two main categories: acoustic imaging and hydroacoustic methods. While underwater visible imaging suffers from limitations due to light attenuation caused by water absorption and scattering, resulting in blurred images and reduced image quantity as shooting distance increases, acoustic-based counting methods offer a viable alternative. Sound waves can travel far through water without significant attenuation, making them suitable for situations where visual counting is inappropriate or ineffective.

1) *Acoustic imaging methods*: Multi-beam imaging sonar such as Dual-frequency Identification Sonar (DIDSON) and Adaptive Resolution Imaging Sonar (ARIS) are normally used to monitor migratory fish in rivers [175]. These systems

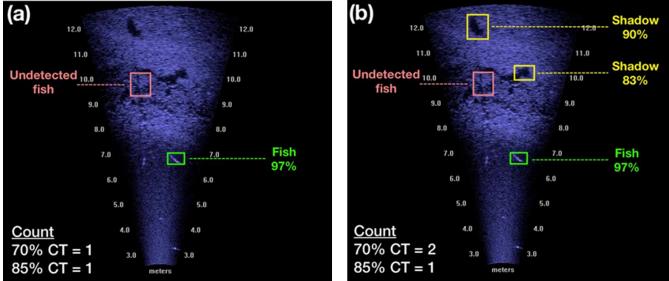


Fig. 9: Example of DIDSON image counting (This figure was reproduced from [183] with minor changes).

produce high-resolution underwater sonar video output without the need for underwater light, allowing for fish counting and measuring directly from the footage, even in turbid waters and overnight [176].

DIDSON [177] is a multi-beam sonar system frequently used to acquire underwater acoustic images for fish identification and counting (as shown in Fig. 9). As DIDSON uses sound instead of light, it is not affected by water turbidity and can collect data during both day and night [178], [179]. However, studies have shown that manual counting of DIDSON data can be time-consuming and prone to errors, with large deviations between operators [180], [181]. This may be because Echoview repeatedly calculated at nearly stationary horizontal positions within the DIDSON field of view [182].

To reduce the time and cost of DIDSON data processing, various subsampling methods can be employed, with automation-assisted subsampling being the best method to reduce the cost of estimating migratory fish populations in rivers [22]. Multi-beam echogram processing software, such as Echoview or DIDSON Control and Display software, can partially perform fish detection and counting functions [184], [185]. Echoview uses a Component Object Model (COM) interface that allows users to build customized pre-processing and post-processing scripting modules, streamlining the processing method and providing the ability to refine fish counting using various fish detection parameters [186], [187]. However, the echograms of the video-like data files generated by DIDSON require manual counting, which is tedious, time-consuming, and can produce large errors for large datasets [188]. Semi-automatic post-processing of imaging sonar data is possible using existing software (e.g., Echoview Software Pty Ltd., Hobart, Australia) [184], [189], but the process still requires manual calibration for non-fish target noise, which is cumbersome and inefficient. Furthermore, post-processing software can be very expensive, limiting its accessibility for many researchers and practitioners.

Digital image-processing technology offers an inexpensive and rapid alternative that has been successfully applied in various scientific fields. Several studies have focused on the automatic processing of fish targets in imaging sonar data. For example, K-nearest neighbour background subtraction with DeepSort target tracking to track and count fish automatically [190] and GPNet, a novel encoder-decoder network with global attention and point supervision, to boost sonar image-

based fish counting accuracy [191].

The new generation of acoustic cameras includes ARIS (Sound Metrics Corp, WA, USA), which operates at higher frequencies compared to DIDSON, offering greater flexibility and improved image resolution [192], [193]. A comparison of fish monitoring data based on the ARIS sonar system and the GoPro camera showed that the detection rate of the sonar-based system was 62.6% (compared to the amount captured by the net), exceeding the 45.4% of the camera-based system [194].

While sonar imaging counting methods are powerful tools for gathering fish abundance estimates in difficult-to-observe, structurally complex, chaotic, and dark environments, they can still be disturbed by various types of underwater noise. Additionally, sonar imaging equipment is relatively expensive and requires professional personnel to conduct analysis, making it more suitable for investigating fish abundance in ocean fishing and river ports [195], [196].

These recent advancements in digital image-processing techniques showcase the growing interest in developing efficient and accurate methods for automatic fish tracking and counting in sonar data. By leveraging the power of deep learning and computer vision algorithms, these approaches aim to overcome the limitations of manual processing and provide more reliable and scalable solutions for aquaculture monitoring and management. However, while these methods show promising results, they still face challenges such as dealing with occlusions, varying fish densities, and the need for large annotated datasets for training. Future research should address these limitations and develop more robust and generalizable algorithms that can be easily adapted to different sonar imaging systems and underwater environments.

Table IV summarizes the advantages and disadvantages of sonar imaging counting methods in aquaculture and their practical applications. Despite the limitations, acoustic counting methods remain valuable for monitoring fish populations in challenging underwater environments where visual counting methods may be impractical or ineffective.

2) *Hydroacoustic methods*: Acoustic echo-sounding is one of the most popular methods for estimating fish abundance due to their simplicity and non-invasive nature [198]. These methods rely on the physical characteristics of the target and the water medium. When an echo sounder's transducer emits an acoustic wave, it spreads through the water and encounters the target object. Due to the difference in acoustic impedance between the object and the water medium, the object scatters the incident acoustic wave, and a portion of it is backscattered to the transducer, known as the echo signal [199], [200].

The target's depth can be measured according to the interval between the acoustic emission and the reception of the target's echo. By analyzing the strength and structure of the echo signal, the intensity, number, and distribution of the target can be estimated. The Echo Integration method is one of the main methods for underwater acoustic assessment of fish stocks. It calculates the number of fish by dividing the integral value of the echo intensity of fish in the sampling unit area by the ultrasonic reflectance of individual fish (target intensity, TS). Several studies have used the echo integration technique to

TABLE IV: A comparison of different methods based on acoustic imaging.

Site	Technology	Software	MHz	Metrics	Results	Advantages	Limitations	References
River	ARIS	Echoview	1.1MHz frequency	Accuracy	84%	Distinguishes downstream moving fish from other objects	Results vary among operators	[20]
Lagoon	ARIS	Sound Metrics	1.8MHz	R^2	0.99	Consistent results with manual	The results varied greatly among different operators	[180]
River	ARIS	ARIS software Fish	1.8MHz	F1-scores	75%	Faster and no post-processing	Underestimates total fish count	[197]
Reservoir	ARIS	KNN background subtraction and DeepSort	1.8MHz	Accuracy	73%	Automatic calibration saves data processing time	Unable to identify fish in bottom background, long processing time	[190]
River	ARIS	ARISfish	3.0MHz	Detection Rate	62.6%	Counts fish > 100 mm in night and turbid conditions	May not detect small fish	[194]
River	DIDSON	Echoview 6.0	1.8MHz	Accuracy	83.7%	Avoids manual counting errors and biases	Time-consuming calculations	[22]
River	DIDSON	Sound Metrics	1.8MHz	F1 scores	79%	Performs well using direct, shadow, and combined detections	Low fish densities in each image	[183]
Reservoir	DIDSON	NN-EKF2; Echoview	1.8MHz	Error compared with the manual detection results	Less than 5%	Less calculation and easy to implement Estimates	Inaccurate when targets overlap	[196]
River	DIDSON	Sound Metrics; Echoview	1.2MHz	Accuracy	90% (upstream) 41% (downstream)	potamodromous fish passage in large lakes Not limited by surface disturbances or turbidity	High processing times and costs	[182]
River	DIDSON	Manual counting	1.8MHz	Average Percent Error (APE)	5.4%	Better acoustic target identification and resolution	Shadowing from passing fish	[181]
River	DIDSON	Hand-counter	1.8MHz	Coefficient of Variation (CV)	9.63%		Data loss on small fish in highly turbulent environments	[195]

estimate the number of fish based on the backscattering echoes observed with an echo sounder [201], [202].

Although the sound intensity reflected by a shoal is related to the number of fish [203], the use of echo sounders in fish tanks and cages presents several challenges [204]. Reverberation in a cage can occur due to the echo of an acoustic signal from the boundary, necessitating the removal of the cage boundary signal during counting [205]. Another issue with acoustic estimation of fish populations is shadow utility, which is needed to compensate for the attenuation of echo strength when dense shoals are in focus [200]. To investigate the possibility of using commercial echo sounders for real-time fish counting in offshore cages, a study by [206] employed an echosounder and echo-integration technology. The experimental results showed that the proposed method could achieve more than 90% estimation accuracy [207], indicating its reliability for future fish management decisions.

Despite the increasing use of underwater echo sounders in fishery research, their application is subject to interference from various factors, such as differences in instrument performance, the blind area of the echo sounder itself, external environmental factors, and the evasive behaviour of fish in re-

sponse to survey ships and sound waves [208], [209]. Furthermore, echosounders are expensive and technically demanding, making them unsuitable for factory aquaculture needs. Future research should focus on reducing the cost of instruments or developing alternative instruments suitable for promotion to meet the actual needs of aquaculture.

D. Comparative analysis and trends in fish counting methods

Fish counting methods in aquaculture can be broadly categorized into sensor-based, computer vision-based, and acoustic-based approaches, each with distinct advantages and limitations. Sensor-based counting, including infrared and resistivity counters, offers real-time data but may be limited by water conditions and fish density [9]. Infrared counters are non-invasive but struggle in turbid waters, while resistivity counters can operate in various conditions but may require fish to pass through specific channels [13], [132]. Computer vision-based counting is divided into image-based and video-based methods. Image-based counting is cost-effective and can handle large datasets but may struggle with overlapping fish [153]. Video-based counting offers more dynamic analysis, and better handling of movement and overlaps, but requires

more computational power [169], [170]. Acoustic-based counting excels in challenging environments, including acoustic imaging and hydroacoustic methods [3]. Acoustic imaging (e.g., DIDSON) works well in turbid waters and at night but can be costly [179], [182]. Hydroacoustic methods are effective for quantifying fish density and biomass at large scales but may lack species-specific accuracy [199], [205].

Analysis of research literature reveals that the selection of fish tracking methods is closely tied to aquaculture environments and infrastructure types. In intensive pond and tank systems, computer vision-based methods have been widely studied and implemented [210], demonstrating particular success in controlled environments. Research in large-scale marine farming has explored various approaches, including acoustic methods for deep-water applications [197]. Recent studies have increasingly focused on AI-integrated tracking systems [22], while research in open-water aquaculture settings has predominantly investigated acoustic tracking methods due to their effectiveness in larger volumes [211]. The adoption trends are shaped by environmental conditions, species being counted, regulatory requirements, and economic factors. Due to cost constraints, developing regions often rely on manual counting or basic sensor technology. Welfare considerations are increasingly important in counting method selection. Non-invasive methods like computer vision and acoustic imaging are preferred to minimize fish stress. Ongoing research focuses on improving accuracy while reducing the impact on fish behaviour.

E. Real-world applications in fish counting

Fish counting is a critical aspect of aquaculture management, essential for controlling production volume, feed usage, and regulatory reporting. Yanmar Marine Systems Co., Ltd. (YMS) developed a real-world application (as shown in Fig. 10) of automated fish counting to address the challenges faced in tuna farming¹. Traditionally, farm fish counting relied on labour-intensive and potentially inaccurate methods such as visual counting from camera footage or subtractive counting of dead fish. To overcome these limitations, YMS developed an automated fish counting system aimed at achieving 98% or higher accuracy in real-time tuna counting. This system integrates advanced image recognition technologies and consists of underwater cameras, a dedicated image processing computer, and specialized software. Key features include real-time counting using simultaneous imaging and count recognition, detection of environmental disturbances, and remote adjustment for optimal imaging conditions. The system is beneficial when introducing young Pacific bluefin tuna into nets, during the division of fish into multiple nets during growth phases, and for daily inventory management and feed usage monitoring. In testing, it achieved the target of 98% or higher accuracy, leading to its commercial release in December 2020. This automated system significantly reduces the workload on farm workers while maintaining high accuracy, demonstrating the practical application of computer vision and image processing

¹https://www.yanmar.com/global/about/technology/vision2/fish_counting_system.html

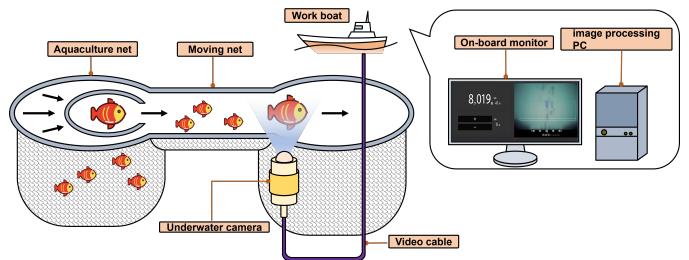


Fig. 10: Example of the counting method developed by YMS.

techniques in commercial aquaculture settings and potentially revolutionizing farm management practices.

Another typical example is a large commercial tilapia hatchery that implemented an advanced dynamic fry counting system using 32 cm wide channels (as shown in Fig. 8), high-resolution cameras, and a central server running YOLOv5-Nano [171] detection model and an improved SORT algorithm [160]. The system uses several innovative approaches to address common challenges such as occlusion, high density, and adhered fry. The improved SORT algorithm maintained consistent tracking of individual fry across frames, even in high-density situations. The detection model was trained on diverse datasets including edge cases to handle adhered fry. The optimized channel design, with careful calibration of width and inclination, encouraged fry separation and reduced occlusion. Temporal information from multiple frames was utilized to resolve ambiguities caused by occlusion or high density. After six months, the incubator reported significant improvements: fry count accuracy increased to within 3-4% (from 20% variance), counting time was reduced by 85%, overfeeding decreased by 12%, and fry survival rates improved by 3%. The implementation led to a 15% reduction in operational costs, a 10% increase in production efficiency, a 70% decrease in counting-related labour costs, and a 20% reduction in lost sales. Improved filtration and model fine-tuning addressed challenges introduced by turbid water and varying fry strains. This application demonstrates the potential of using advanced computer vision and tracking algorithms to enhance commercial aquaculture operations.

IV. FISH SCHOOL BEHAVIOUR ANALYSIS

Fish behaviour, a direct result of the living environment and growth state, includes both normal (i.e. feeding behaviour, swimming behaviour, reproduction behaviour, gathering behaviour) and abnormal behaviours (i.e. disease behaviour, hypoxia behaviour, cannibalism behaviour) [212]–[214]. Poor water quality and management in aquaculture can cause fish stress behaviour [215], leading to immune suppression, slow growth, and reduced productivity and welfare [216]. Traditional fish behaviour analysis, relying on human observers, is often unreliable, time-consuming, and labour-intensive [4], [5]. Accurate estimation of fish behaviour is crucial for optimizing resource use, controlling water quality, and improving fish welfare and economic benefits [217]. The following subsections will explore the latest advancements in fish behaviour analysis,

providing insights into the current state of the art and potential future directions for research and application in this field.

A. Fish school behaviour analysis based on computer vision

1) *Fish feeding behavior*: In intensive aquaculture, feeding is the main expenditure [218], and feeding optimization is crucial for improving efficiency and reducing costs [219]. Traditional feeding methods based on farmers' experience are limited by low efficiency and high labour intensity, and they cannot accurately address the problems of overfeeding or underfeeding [220]. The intensity and amplitude of changes in fish behaviour can directly reflect fish appetite. Computer vision technology can effectively quantify fish feeding behaviour, optimize feeding strategies, and reduce feeding costs.

Many researchers have used traditional methods, such as background subtraction and optical flow, to extract target features for determining feeding indices [221]. While these methods can accurately capture fish feeding behaviour, they require complex foreground segmentation processes that may decrease computational efficiency and are easily affected by water surface fluctuations and reflective areas [24]. With its advantages of automatic feature extraction and large-capacity modelling, deep learning has been widely used in aquaculture [222].

Existing approaches mainly use digital cameras to capture the corresponding images as input and characterize the fish behaviour with discrete feeding intensity (e.g., "None", "Weak", "Medium" and "Strong" [223]–[226]) as a classification problem, modelled by CNNs. However, fish-feeding behaviour is a dynamic and continuous process. Single images are insufficient to capture the context of fish feeding intensity [222], [227]. As an alternative, video-based methods have been proposed to exploit spatial and temporal visual information for fish feeding intensity assessment (FFIA), which offers rich context for capturing fish feeding behaviour. Raw RGB videos were converted into optical flow image sequences and fed into a 3D CNN to evaluate fish feeding intensity, achieving a very high accuracy [228], [229].

While recent advancements in computer vision and deep learning have shown promise in analyzing fish feeding be-

haviour, some limitations still need to be addressed. One major challenge is the discrepancy between the ideal environments in which fish-feeding datasets are collected and the real-world conditions found in aquaculture settings. Factors such as water turbidity, fluctuating light levels, and variable camera angles can significantly impact the performance of these models when deployed in real-world farms.

Another limitation is the computational complexity of video-based models, which often require substantial computational resources, making them difficult to deploy on resource-constrained devices commonly used in aquaculture. The large size of these models can also hinder their real-time performance, which is crucial for timely decision-making in aquaculture management. Furthermore, the limited generalizability of current models to new fish species is a significant challenge. Many existing models are trained on species-specific datasets, and their performance often drops significantly when applied to new or unseen species due to differences in morphological features, colour patterns, and behavioural characteristics.

To address these limitations, future research should focus on developing more robust, adaptable, and species-agnostic models that can effectively handle the variability encountered in real aquaculture environments. This may involve collecting more diverse and representative datasets, exploring domain adaptation, transfer learning, and few-shot learning techniques, and optimizing models for efficient inference on edge devices.

2) *Hypoxia behavior*: Hypoxia, a common issue in aquaculture systems, can significantly impact fish mortality and lead to substantial production losses [230]. Fish exhibit various behavioural responses to hypoxic conditions, such as changes in ventilatory frequency (VF), swimming activity, surface respiration, and vertical habitat [231]–[234]. To provide early warning of hypoxia in aquaculture, it is essential to evaluate the specific behavioural responses of fish when oxygen levels in the water drop sharply.

Image processing algorithms have been proposed to quantify the hypoxia behaviour of fish in aquariums [235]. However, these methods often rely on complex foreground segmentation processes, which can decrease computational efficiency and are easily affected by water surface fluctuations. Deep learning

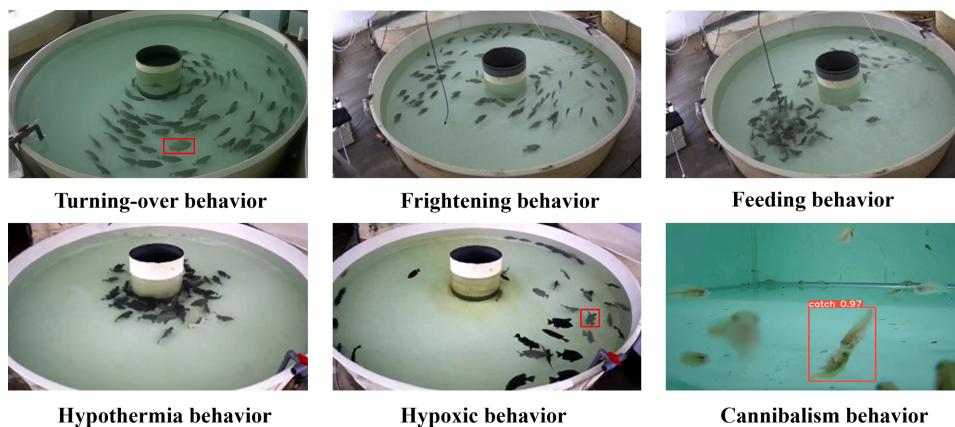


Fig. 11: Abnormal behaviors: "Turning-over behavior", "Frightening behavior", "Feeding behavior", "Hypothermia behavior", "Hypoxic behavior", "Cannibalism behavior".

methods, such as YOLO object detection, have emerged as powerful tools for transforming and upgrading fish farming practices by quickly detecting fish behaviour with high accuracy [218].

Despite the progress made in recognizing fish hypoxia behaviour, most experiments have been conducted under laboratory conditions, which may not accurately reflect the challenges encountered in actual production systems. Factors such as water turbidity, uneven illumination, and high fish density can make it more difficult to identify individual fish and their specific behaviours in real-world settings. Furthermore, inducing hypoxia through human intervention in laboratory experiments can compromise animal welfare and cause irreversible damage to fish health.

To address these limitations, future research should focus on developing more robust and adaptable methods for detecting fish hypoxia behaviour in real-world aquaculture systems. Moreover, integrating multiple data sources, such as water quality sensors and video monitoring systems, could provide a more comprehensive understanding of fish behaviour and enable early detection of hypoxia-related issues. By combining advanced computer vision techniques with domain expertise in aquaculture and fish physiology, researchers can develop more effective and practical solutions for monitoring and managing fish health in real-world settings.

3) *Other abnormal behavior*: Abnormal fish behaviours, such as aggression, fear, stress, illness, parasitic infection, and cannibalism, can have significant impacts on aquaculture production (as shown in Fig. 11), fish welfare, and population balance [214], [236]–[238]. While less common than feeding and hypoxia behaviours, these abnormalities still play a crucial role in aquaculture warning operations. Detecting and localizing abnormal behaviours, particularly those occurring within small groups or individuals, remains challenging in computer vision. To address this challenge, researchers have adapted techniques from human behaviour analysis, such as motion-effect maps and deep learning algorithms, to detect, localize, and recognize abnormal fish behaviours in intensive aquaculture systems [239], [240]. These methods have shown promising results in identifying specific behaviours and evaluating various health and environmental factors. However, further research is needed to investigate the complex interplay between local and global abnormal behaviours and develop robust, multi-target tracking systems that operate efficiently in real-world aquaculture settings.

Monitoring and protecting fish during critical life events, such as spawning aggregations, is essential for maintaining population balance and preventing overfishing [241], [242]. Computer vision techniques, including stereoscopic video analysis and 3D neural networks, have been employed to quantify fish reproductive behaviour and classify complex behaviours [243], [244], providing valuable tools for baseline studies and long-term monitoring.

While computer vision and image processing technologies offer economical and effective means for monitoring abnormal fish behaviour, the relative scarcity of abnormal behaviour data has hindered in-depth research. Most existing studies have been conducted in controlled laboratory environments, which

may not accurately represent the complex factors in real-world aquaculture settings [245], [246]. Overcoming the challenges posed by complex water environments, uneven lighting, large numbers of individuals, and intricate fish movements are crucial for developing robust and reliable abnormal behaviour monitoring and tracking systems [78].

B. Fish behaviour analysis based on multi-object tracking

Visual-based monitoring systems for detecting abnormal fish behaviour often rely on known scenes and predefined movement models, which can be subjective and lack adaptability to different environments [247]. Advanced tracking methods, however, offer more robust and adaptable solutions for analysing fish behaviour in diverse aquaculture settings [227].

Tracking technologies enable researchers to obtain comprehensive data on fish movement, including average, maximum, and minimum speeds, acceleration, average collision frequency, and trajectory changes. For example, studies have tracked zebrafish using YOLOv2 [248] and Kalman filters, obtaining movement trajectories that showed significantly faster swimming, greater agitation, and agglomeration in the centre of the aquarium during feeding periods [29]. Similarly, semi-automatic in situ tracking systems have been developed to reconstruct synchronized 3D movement trajectories of individual reef fish in social groups, analysing their behaviour when capturing plankton prey [36].

Recent advancements in tracking algorithms have expanded the capabilities of fish behaviour analysis. Huang et al. [249] developed an early warning system for detecting nocardiosis in largemouth bass (*Micropterus salmoides*) based on YOLOv8 [63] and ByteTrack [67]. The system quantifies fish velocity and turning angles through trajectory analysis, enabling effective early disease prevention. Similarly, Zhao et al. [250] employed YOLOv8+ByteTrack multi-target tracking to monitor zebrafish behaviour under various environmental stressors, extracting key movement features (e.g. speeds, average collision frequency, and trajectory changes) to identify exposure to different pollutants rapidly. Xiao et al. [251] enhanced the tracking performance of largemouth bass (*Micropterus salmoides*) by combining an improved YOLOv8 [63] with ByteTrack, incorporating trajectory confidence information. This approach improved tracking accuracy and enabled more detailed analysis of fish swimming patterns. By examining the relationship between swimming speed, swimming time, and spatial coordinates, the researchers were able to quantify swimming ability in terms of both speed and endurance.

Despite the advancements in tracking algorithms, their application to abnormal fish behaviour analysis in aquaculture remains limited. This scarcity can be attributed to several factors: limited availability of open datasets on abnormal fish behaviour, the rare occurrence of such behaviours making data acquisition challenging, the complexity of fish trajectories containing multi-dimensional information (position, speed, direction), and difficulty in defining ‘abnormal’ trajectories which may encompass multiple aspects [252], [253].

To address these challenges and advance aquaculture, future research should focus on developing specialized datasets for

abnormal fish behaviour in aquaculture settings, adapting advanced tracking algorithms to the unique characteristics of fish and aquatic environments, integrating domain knowledge from aquaculture experts to better define and detect abnormal behaviours, and drawing inspiration from successful applications of anomaly detection in other fields, such as crowd and vehicle monitoring. By focusing on these areas, researchers can create more powerful and flexible models for identifying and comprehending abnormal fish behaviour in various aquaculture contexts, leveraging the full potential of advanced tracking technologies.

C. Fish behaviour analysis based on passive acoustic monitoring

Passive acoustic monitoring (PAM) has emerged as a non-invasive and increasingly accessible remote sensing technology for monitoring underwater environments [254], [255]. With approximately 1,000 out of the 35,000 known fish species confirmed to produce sounds underwater [256], [257], PAM offers a unique opportunity to analyze fish behaviour through the sounds they generate. An audio example of fish abnormal behaviour is shown in Fig. 12.

Fish can produce a series of sounds during feeding, and the frequency spectrum of these sounds can be used to analyse their feeding behaviour. For example, turbots generate feeding sounds that vary with food intake intensity, ranging from 15 to 20 dB in the frequency range 7–10 kHz [258]. Similarly, feeding sounds produced by various fish species, such as rainbow trout (0.02–25 kHz) [259], Japanese minnow (1–10 kHz) [260], Atlantic horse mackerel (1.6–4 kHz) [261], yellowtail (4–6 kHz) [262], have comparable frequency ranges.

The Fish feeding behaviour analysis based on audio was initially proposed by [16], [263], where the audio signal is first transformed into log Mel spectrograms and then fed into a CNN-based model for FFIA. Subsequent work [264], [265] have further demonstrated the feasibility of using audio as input for FFIA. Audio-based methods offer advantages such as energy efficiency and lower computational costs compared to vision-based methods [266], [267]. However, audio-based models have lower classification performance than video-based FFIA due to their inability to capture full visual information and sensitivity to environmental noise [268]. Moreover, rapidly swimming predatory fish, such as brown and rainbow trout, often combine forward swimming with feeding, accompanied by splashing sounds and strong tail patting [217]. The rapid pellet capture by these species superimposes feeding sounds, and pellet impacts pose a challenge in obtaining accurate feeding sound data.

To overcome these challenges, future research should focus on developing advanced signal processing techniques to separate feeding sounds from ambient noise and other interfering sounds. Additionally, exploring the integration of audio and visual data could help improve the overall classification performance and robustness of fish behaviour analysis systems.

D. Fish behaviour analysis based on biosensor technology

Biosensor technology has shown great potential in collecting individual animal information, such as individual trajectory,

acceleration, velocity, respiration frequencies, heartbeat frequency, and tail beat frequency [270], [271]. In recent years, accelerometers have been increasingly used in marine biology research to study the feeding behaviour of aquatic animals.

The feeding behaviour of most fish leads to characteristic changes in acceleration that differ from their normal movement patterns [272]. These characteristic changes in acceleration can be effectively used to distinguish feeding behaviour patterns from other behaviour patterns [273]. For example, in [274], accelerometer tags were used to investigate the feeding behaviour of Atlantic cod (*Gadus morhua*) in the wild. The authors found that the accelerometer data could accurately identify feeding events and provide insights into the foraging ecology of this species. Similarly, a study by [275] used a combination of accelerometers and gyroscopes to analyse the feeding behaviour of captive yellowtail kingfish (*Seriola lalandi*). The authors demonstrated that the sensor data could be used to classify different types of feeding behaviour, such as biting, chewing, and swallowing, with high accuracy.

In addition to feeding behaviour, biosensors have been used to study other aspects of fish behaviour, such as swimming activity and energy expenditure. For instance, in [276], accelerometers were used to investigate the swimming behaviour and energy expenditure of wild Atlantic salmon (*Salmo salar*) while migrating to spawning grounds. The authors found that the accelerometer data provided valuable insights into the swimming performance and energy costs of this species in natural conditions.

However, using biosensors in fish behaviour analysis also presents some challenges and concerns. Biosensors are typically surgically attached or implanted into the fish's body, which can lead to the direct death of the fish or cause behavioural changes that may affect the results of experiments. Moreover, this method may cause irreversible harm to the fish and compromise animal welfare. To address these issues, researchers should focus on developing minimally invasive or non-invasive biosensor technologies that can be safely attached to or removed from fish without causing undue stress or harm. Furthermore, ethical considerations should be prioritized when using biosensor technology in fish behaviour analysis.

Despite these challenges, biosensor technology offers a promising approach to studying fish behaviour at the individual level, providing valuable insights into the feeding ecology, swimming performance, and energy expenditure of various fish species. By combining biosensor data with other monitoring techniques, such as passive acoustic monitoring and vision-based methods, researchers can develop a more comprehensive understanding of fish behaviour in both captive and wild settings. As biosensor technology continues to advance, it is essential to balance the potential benefits of these tools with the need to ensure the welfare and ethical treatment of the fish being studied.

E. Comparative analysis and trends in fish school behaviour analysis

Fish school behaviour analysis in aquaculture focuses on understanding and monitoring fish activities, stress responses,

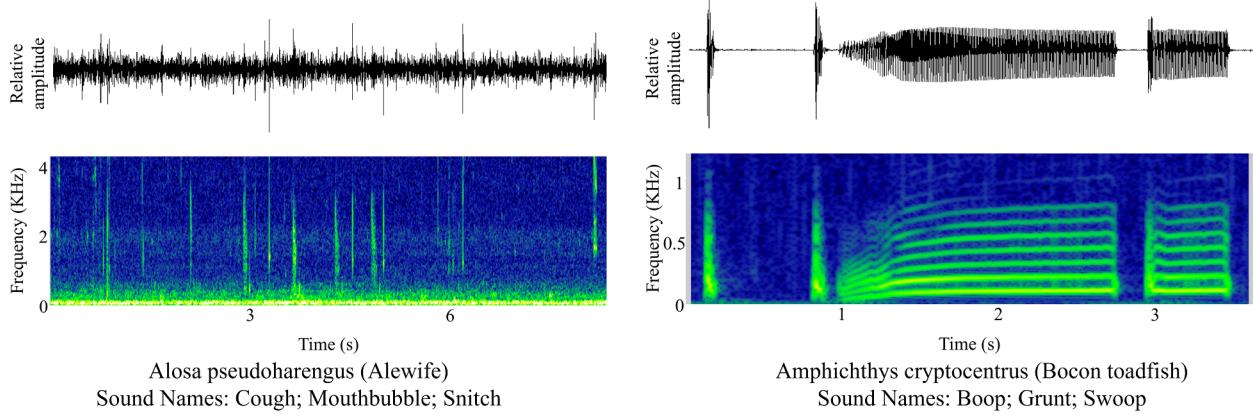


Fig. 12: The audio spectrum of different fish abnormal behaviours (This figure was reproduced from [269]).

and interactions within the aquaculture environment. This field employs various technologies, each with distinct capabilities and applications. Computer vision-based analysis excels in observing feeding patterns, detecting abnormal behaviours, and identifying signs of stress or disease [277]. Advanced algorithms can recognize specific behaviour patterns such as aggressive interactions, schooling dynamics, and responses to environmental changes. While being non-invasive, this method's effectiveness can be limited by water clarity and lighting conditions [26]. Multi-object tracking extends computer vision capabilities by following individual fish within schools over time. This approach provides insights into social hierarchies, individual growth rates, and how behaviour patterns spread through a population. It requires sophisticated algorithms and high computational power but offers unparalleled detail in understanding group dynamics [4], [11].

Passive acoustic monitoring analyses sounds produced by fish to infer behaviour states and stress levels [254]. This method is particularly valuable for nocturnal species or in turbid environments where visual observation is challenging. It can detect feeding activity, spawning events, and even some types of distress calls, offering a unique perspective on fish behaviour [278]. Biosensor technology, including implantable tags and external sensors, provides detailed physiological data that can be correlated with behaviour [274]. These sensors can measure parameters like heart rate, muscle activity, and hormone levels, offering insights into stress responses, energy expenditure during different activities, and even feeding intensity [279]. However, the invasive nature of some biosensors raises welfare concerns and may itself affect behaviour. The choice of behaviour analysis method often depends on research goals, species-specific characteristics, and environmental conditions. For example, salmon farms might prioritize feeding behaviour analysis to optimize feed conversion ratios, while tilapia farms might focus more on aggression and territoriality behaviours.

Regional trends in adoption vary. Asia, with its diverse aquaculture species and systems, shows broad adoption across methods [280], [281]. Europe emphasizes non-invasive techniques due to strict welfare regulations [282]. North America leads in integrating multiple methods for comprehensive behaviour analysis. Welfare considerations are increasingly

central to method selection [283], [284]. There is a growing preference for non-invasive techniques that can provide detailed behavioural data without causing stress. This has spurred research into improving the resolution and accuracy of passive monitoring methods. The trend in fish behaviour analysis is moving towards holistic approaches that combine multiple methods to create a comprehensive picture of fish welfare and behaviour. For instance, combining computer vision with passive acoustics can provide round-the-clock monitoring that captures both visual and auditory behavioural cues [263]. Real-world applications have demonstrated the value of behaviour analysis in improving aquaculture management. Farms using advanced behaviour monitoring have reported earlier detection of disease outbreaks, more efficient feeding practices, and improved overall fish welfare, leading to better growth rates and product quality [285], [286]. As aquaculture intensifies and faces new challenges like climate change impacts, behaviour analysis will play an increasingly crucial role in maintaining fish health, optimizing production, and ensuring sustainable practices.

F. Real-world applications in fish behaviour analysis

Fish behaviour analysis in real-world applications has emerged as a powerful tool for monitoring aquaculture environmental conditions and fish welfare. By tracking and analysing changes in swimming patterns, spatial distribution, and movement characteristics, these systems can serve as early warning indicators for water quality issues and stress responses, enabling timely interventions in commercial aquaculture operations. Xu et al. [287] demonstrate how fish behaviour analysis can be used to monitor water quality by studying responses to ammonia nitrogen stress. Their system combines an improved YOLOv8 model (enhanced with multi-head self-attention and Wise Intersection over Union (Wise-IoU)) with dual-view cameras for 3D tracking. By integrating advanced tracking techniques including Kalman filtering and Kernelized Correlation Filters [288], the system accurately monitors multiple fish simultaneously. The study analysed behavioural patterns of different species (sturgeon, bass, and crucian) under various ammonia nitrogen stress levels, exam-

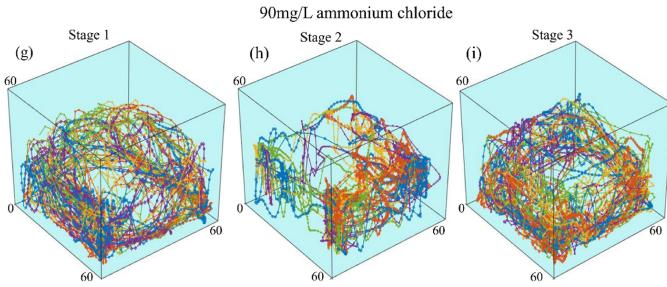


Fig. 13: Behavioral trajectories of sturgeon under ammonia nitrogen stress. The five colours of the behavioural trajectory represent five fish (This figure was reproduced from [287]).

ining their trajectories, exercise volumes, spatial distribution, and movement velocities.

Fig. 13 shows that sturgeons swim around the aquarium under normal conditions while they swim up and down at a certain position in the aquarium under ammonia nitrogen stress [287]. Meanwhile, the phenomenon of sturgeon swimming around the aquarium under ammonia nitrogen stress has decreased. Key findings revealed significant changes in fish swimming patterns, spatial distribution, and velocity under ammonia stress, with species-specific responses to different concentrations. The method successfully differentiated normal behaviour from stress-induced abnormalities, demonstrating its potential for early detection of water quality issues in aquaculture settings. However, challenges such as turbid water and varying light conditions required additional solutions like adaptive image preprocessing techniques and regular maintenance protocols.

V. MULTIMODAL FUSION IN AQUACULTURE

Multimodal data fusion has been used in aquaculture where data from various sources are integrated to understand fish behaviour, health, and environmental conditions. Recent advancements in other fields, such as deep learning, offer promising applications for aquaculture, particularly in fish tracking, counting, and behaviour analysis.

1) *Multimodal fusion in fish tracking:* For fish tracking, state-of-the-art vision-language models show significant potential. CLIP2Video [289] and AudioCLIP [290] extend the capabilities of their predecessors in image-text and audio-text matching by leveraging the contrastive language image pretraining (CLIP) model [291], and could be adapted for multimodal fish detection. For instance, the AudioCLIP architecture [290] could be modified to correlate visual fish features with sonar data, potentially improving detection accuracy in turbid waters commonly encountered in aquaculture settings. However, the significant differences between underwater RGB images and sonar images present unique challenges in multimodal tracking for aquaculture. Many existing multimodal tracking methods assume alignment between different modal spaces, using a prediction head to estimate the same target box across modalities. This assumption often fails in RGB-Sonar (RGB-S) tracking, where the representation of the target can vary significantly between the two imaging types [292],

[293]. Furthermore, the spatial misalignment of multimodal features complicates the integration of information from these diverse sources. A critical challenge in multimodal fusion for aquaculture is data synchronization. Different sensors often have varying sampling rates and latencies. For instance, acoustic sensors might collect data at several kilohertz (kHz), while visual sensors typically capture frames at 30-60 Hz, and environmental sensors (like temperature or pH sensors) might only take measurements every few seconds or minutes.

Developing robust synchronization methods, such as timestamp alignment and interpolation techniques, is crucial for accurately fusing these diverse data streams in real-time aquaculture monitoring systems. To address these challenges, researchers have developed specialized approaches for underwater tracking. Li et al. [294] introduced a novel spatial cross-attention method that effectively bridges the gap between underwater RGB and sonar modalities. Their approach, which builds upon the Cross-Attention method [76], enables more nuanced interactive fusion between the different modalities. By accounting for the unique characteristics of each imaging type, this method achieves state-of-the-art performance in underwater tracking, demonstrating its potential for improving fish monitoring in complex aquaculture environments. The introduction of large-scale datasets like WebUOT-1M [295], which includes language prompts for video sequences, opens up new possibilities for underwater vision-language tracking. This could be particularly useful for developing more robust fish tracking systems in complex aquaculture environments, such as sea cages with varying light conditions and water clarity. However, a significant research gap remains in bridging vision-language models with sonar data for aquaculture applications, particularly in developing frameworks that can effectively leverage textual descriptions for enhanced underwater object detection and tracking. Future research should focus on developing more sophisticated multimodal fusion techniques to improve tracking accuracy and robustness in complex underwater environments.

2) *Multimodal fusion in fish counting:* In the realm of fish counting, recent developments in crowd counting using vision-language models offer promising approaches. CrowdCLIP [296] leverages pre-trained vision-language models to address the challenge of costly manual labelling, especially in dense scenes. This approach could be adapted for counting fish in crowded aquaculture tanks or ponds, where traditional methods often fail. CLIP-Count [297] presents an innovative end-to-end pipeline for estimating density maps of open-vocabulary objects with text guidance in a zero-shot manner. Its patch-text contrastive loss and hierarchical patch-text interaction module could be adapted to generate high-quality density maps for fish in various aquaculture scenarios, potentially improving counting accuracy in complex underwater environments with overlapping fish. Furthermore, VLCounter [298] explores the implicit association of semantic-patch embeddings of CLIP, showing improved performance in zero-shot object counting. This approach could be useful in aquaculture for counting different fish species or identifying specific fish features without extensive labelled datasets, a common challenge in diverse aquaculture operations. However, adapting

these vision-language models to underwater environments presents unique challenges, particularly in handling water turbidity and varying lighting conditions. Future research should focus on developing specialized vision-language architectures that can better handle the specific challenges of underwater counting tasks while maintaining the benefits of zero-shot learning capabilities.

3) *Multimodal fusion in fish behaviour analysis*: In fish behaviour analysis, multi-sensor approaches combined with advanced algorithms offer new insights. The Multimodal Foundation Model (M3AE) [299], designed for general-purpose multimodal understanding, could be adapted to integrate visual, acoustic, and environmental data for comprehensive fish behaviour monitoring. This integration could enable early detection of stress or disease by correlating subtle behavioural changes with environmental parameters. Cui et al. [263] introduced a novel unified mixed-modality-based method for fish feeding behaviour assessment, capable of processing audio, visual, or audio-visual modalities. This approach demonstrates the potential of multimodal fusion to provide more robust and accurate assessments of fish behaviour than single-modality based methods. Similarly, Du et al. [300] showed that fusing information from audio, video, and imaging sonar modalities resulted in higher accuracy than using a single modality in assessing fish behaviour. While these studies demonstrate promising results, there remains a need to develop more sophisticated behavioural pattern recognition algorithms that can identify complex fish behaviours across different species and environmental conditions. Future research should explore integrating biological knowledge with multimodal deep learning approaches to better understand and interpret fish behavioural patterns in various aquaculture scenarios.

4) *Challenges and solutions in multimodal fusion for aquaculture*: Noise management is particularly challenging in underwater environments. Each modality (visual, acoustic) is subject to different types of noise. For example, visual data may be affected by turbidity and light scattering, while acoustic data can be distorted by ambient noise from equipment or other marine life. Advanced filtering techniques, such as Bernoulli filters [124] and differentiable particle filtering [123], need to be developed specifically for aquaculture applications to ensure reliable data fusion in these complex environments [301]. The relative importance and reliability of different data sources can vary based on environmental conditions and the specific behaviour being analyzed. Developing dynamic weighting schemes that adjust the contribution of each modality based on real-time quality assessments is essential [302]. For instance, acoustic data might be weighted more heavily in turbid conditions than visual data, while in clear water, the reverse might be true. Machine learning techniques, such as attention mechanisms in transformer-based models, offer promising approaches for adaptive weighting in multimodal fusion systems.

However, several challenges limit the direct application of these methods in aquaculture. Limited labelled data is a significant hurdle, as aquaculture-specific datasets are scarce compared to general computer vision tasks. Environmental variability in underwater conditions, such as changing turbid-

ity, lighting, and temperature, can significantly impact sensor reliability and model performance. Additionally, species-specific behaviours require specialized models, as different fish species exhibit unique patterns [303]. The dynamic nature of fish movement and the three-dimensional space of aquatic environments further complicate the application of these technologies. Several approaches have shown promise in addressing these challenges. Few-shot learning techniques could help the models adapt to new species or environments with limited data [304], crucial for the diverse range of fish species in aquaculture. Self-supervised learning methods [305], [306] could leverage large amounts of unlabeled aquaculture data to pre-train robust models, potentially overcoming the scarcity of labelled datasets. Domain adaptation techniques [307] could help transfer knowledge from data-rich domains to aquaculture-specific tasks, bridging the gap between general computer vision tasks and the unique requirements of underwater environments.

Future research should focus on developing specialized datasets that capture the diversity of aquaculture environments and species. This could include multi-sensor data collected across different water conditions, fish life stages, and farming systems. Adapting advanced algorithms to underwater environments is crucial, particularly in handling variable lighting, water turbidity, and three-dimensional fish movement. Creating robust, generalizable models for diverse aquaculture scenarios will be key to the widespread adoption of these technologies. Integrating multiple data sources and advanced processing techniques can revolutionize fish monitoring, enhance production efficiency, and improve animal welfare in aquaculture operations. However, it is important to note that applying these technologies in real-world aquaculture settings may face practical challenges such as cost, maintenance of underwater sensors, and the need for real-time processing capabilities. As research progresses, critically evaluating the practicality and cost-effectiveness of these advanced multimodal approaches in various aquaculture settings will be essential.

VI. PUBLIC DATASET

High-quality public datasets are crucial for developing and evaluating various methods including deep learning methods for fish detection, tracking, and behaviour analysis. Despite growing efforts, public datasets on underwater fish scenes are scarce, particularly in realistic aquaculture environments. This limitation has led many researchers to conduct analyses and behavioural studies under ideal or controlled conditions, which may not fully represent the challenges of real-world aquaculture settings. Table V summarizes the available public fish datasets, highlighting their diverse characteristics and limitations.

The Fish4Knowledge dataset [308], captured in Taiwan waters from 2010 to 2013, offers a variety of marine scenes but suffers from low resolution (i.e. 320×240 pixels), limiting its applicability to current research needs. SeaCLEF2016 [309] builds on this with diverse species coverage and high-quality annotations, enhancing automated marine life identification capabilities. However, both datasets have limitations, including

TABLE V: Summary of the various fish datasets.

Dataset	No. of videos/image	Resolution	Number of labeled data	Tasks	Reference
Fish4-Knowledge	700,000 videos with 10 min each clip Training set: 20 videos and 20,000 images, Test set: 73 videos	320 × 240 640 × 480 320 × 240	- 9,000	Classification, Detection and Tracking Classification, Counting	[308] [309]
SeaCLEF 2016					
NCFM	16,915 images (3,777 training, 13,138 testing) 30 videos	1920 × 1080	10000	Detection, classification and counting	[310]
Sonar image counting dataset	sequence with 537 images Training Set: 54052 images, Test Set: 32400 images	360 × 360	-	Counting	[311]
3D-ZeF20		2704 × 1520	86,452	Tracking	[89]
FishTracker23	Approaching 1 million boxes and frames 1,500 underwater videos (total 10.5 hours)	1920 × 1080 1280 × 720	850,000	Tracking	[312]
WebUOT-1M	and 408 categories	-	1,100,000	Tracking	[295]
Automated Fish Tracking	189 videos of varying durations (1- 30 seconds)	1920 × 1080	8,700	Detection, Tracking	[313]
DeepFish	39,766 images	1920 × 1080	3200	Segmentation, counting and Classification	[314]
FISHTRAC	14 videos	1920 × 1080	3,449	Tracking and detection	[105]
BrackishMOT	98 videos each lasting about 1 minute	2704 × 1520		Tracking	[315]
CFC	527215 SONA images over 100k SONA images	288 × 624 to 1086 × 2125	515,933	Detection, Tracking and Counting	[316]
Mullet Schools Dataset		320 × 576	500	Detection Counting	[306]
Fish Sounds	115 different fish sound clips	64kbps	-	Behaviour analysis	[317]
AV-FFIA	27000 video and sound clips	1086 × 2125 256kbps	All	Feeding Behaviour analysis	[263]

geographic specificity to Taiwanese waters, fixed camera positions that may miss certain behaviours, and low video quality issues due to underwater conditions. The DeepFish dataset [314] significantly contributes to underwater fish monitoring, containing 39,766 images from 20 locations in Australian marine environments. It offers annotations for multiple tasks including classification, counting, localization, and polygonal annotations for precise fish shape delineation. While being versatile, DeepFish initially lacks segmentation masks, limiting its use in instance segmentation tasks. Its diverse environmental coverage, from clear to turbid waters, enhances its utility for developing robust algorithms. However, the dataset's focus on still images and specific marine ecosystems may limit its direct applicability to some aquaculture scenarios or dynamic behaviour studies.

The WebUOT-1M dataset [295] represents a significant advancement, offering a large-scale solution with 1.1 million frames across 1,500 underwater videos and 408 diverse target categories. Its inclusion of language prompts for each video facilitates multi-modal research and enables the development of more generalized and multi-modality underwater object-

tracking models. However, the dataset's web-sourced nature means it may include some biases in image quality, object representation, and environmental conditions. While this diversity can be beneficial for creating robust models, it may not always accurately represent real-world underwater scenarios encountered in specific research or industrial applications. For specialized research, datasets like 3DZeF20 [89] and AV-FFIA [263] offer unique perspectives. 3DZeF20 provides high-precision 3D data for zebrafish in laboratory settings, invaluable for detailed behaviour analysis but limited in ecological applicability. AV-FFIA innovatively combines hydrophone and video data for fish feeding behaviour analysis, offering deep insights into specific patterns. While all of them are invaluable for specific research questions, their applicability to broad aquaculture scenarios may be limited due to their focus on controlled environments or single species.

FishTrack23 [312] and FISHTRAC [105] represent the current state-of-the-art in real-world underwater object detection and tracking. These datasets address the complexities of natural marine environments, providing comprehensive annotations for multiple tasks. They excel in developing practical

monitoring tools but may still face challenges with image quality consistency and global species representation. The BrackishMOT dataset [315] uniquely addresses tracking in turbid environments, which is crucial for many aquaculture settings. The Caltech Fish Counting (CFC) dataset [316] significantly contributes to underwater computer vision and marine ecology and is specifically designed to address the challenges of fish counting in complex underwater environments. Its diverse scenes capture different lighting conditions, water turbidity levels, and fish densities, making it an excellent resource for developing robust fish detection and counting algorithms. The dataset's strength lies in its real-world applicability, presenting researchers with the genuine challenges faced in marine population surveys, such as occlusions, varying fish sizes, and species diversity. However, users should be aware that while the CFC dataset offers a comprehensive view of coral reef fish populations, it may have limitations in representing global marine ecosystems. The Fishsounds dataset [317], while offering audio data for various fish species, lacks sufficient data for comprehensive behaviour analysis. The Mullet Schools dataset [306] provides large-scale sonar imagery for fish counting but is species-specific, limiting its broad applicability in aquaculture research.

These datasets highlight several limitations in current public resources for aquaculture research. Many focus on specific conditions or species, not capturing the full range of aquaculture environments and diversity. Most are unimodal, lacking the integration of visual, acoustic, and environmental data crucial for comprehensive aquaculture monitoring. Varying annotation standards across datasets complicate cross-dataset evaluations. We suggest creating multi-modal datasets that integrate visual, acoustic, and environmental data specific to aquaculture settings to address these limitations and better serve future research needs. Existing datasets should be expanded to include a wider range of species and environmental conditions relevant to global aquaculture practices. Developing standardized annotation protocols would ensure consistency across aquaculture-focused datasets. Establishing benchmark datasets specifically designed for aquaculture applications, covering various farming systems and environmental conditions, would greatly benefit the field.

VII. CHALLENGES AND FUTURE PERSPECTIVES

Fish tracking, counting, and behaviour analysis play a crucial role in the intelligent development of aquaculture production. While computer vision technology is currently a popular method for these tasks, it faces several challenges due to the unique characteristics of aquaculture environments, such as high fish density, complex water backgrounds, and irregular fish movement. These factors can lead to interference between multiple targets, false detections, missed counts, and tracking failures.

Acoustic methods offer an alternative approach that enables automatic and rapid fish counting and tracking in low-light and turbid water conditions. However, underwater noises, high equipment costs, and the need for professional expertise make acoustic methods more suitable for large-scale operations like marine fishing rather than factory or pond farming

environments. To further increase the level of intelligence in aquaculture, we predict several different trends for future development:

1) **Massively available datasets:** The wide application of intelligent technology in aquaculture, especially the success of deep learning algorithms in image processing [19], has highlighted the need for large labelled datasets. Although available datasets are gradually increasing, most are limited to identifying and detecting fish species. Open data on fish tracking, counting, and behaviour analysis is scarce. Passive acoustic monitoring is also gaining popularity for underwater listening [255], [318] and public sound data of underwater fish (e.g., Fishsound) are emerging. However, the sample size of these datasets has not yet reached critical mass. In the future, developing an international platform for sharing images and acoustic data will be essential to promote sustainable aquaculture development.

2) **Edge computing and real-time processing:** As aquaculture operations become more technologically advanced, there is an increasing need for real-time data processing and decision-making at the edge. Developing efficient algorithms and hardware solutions that can perform complex tasks like fish tracking, counting, and behaviour analysis on-site, without relying on cloud computing, will be crucial. This approach can reduce latency, improve data privacy, and enable faster responses to changing conditions in aquaculture environments. Future research should optimise existing algorithms for edge devices, develop specialized hardware for aquaculture applications, and create integrated systems that seamlessly combine multiple monitoring and analysis tasks in real-time.

3) **On-device machine learning:** Most current fish tracking, counting, and behavioural analysis models are performed in the cloud or on high-performance GPUs. However, many aquaculture tasks require real-time responses, such as fish feeding and abnormal behaviour detection. Cloud-based models may struggle to guarantee this real-time performance, and many devices in remote and harsh aquaculture environments may not have consistent internet connectivity. On-device models can greatly reduce exercise pressure and make devices more intelligent, providing users with a better experience. However, terminal devices are often limited in processing power, power consumption, cost, and volume. Future developments could focus on reducing the complexity of computing and storage by optimizing neural network algorithms or compressing network models using techniques like knowledge distillation to enable their deployment on device chips.

4) **Integration of fish tracking, counting, and behaviour analysis:** Most research addresses fish tracking, counting, and behaviour analysis as separate tasks. However, these tasks are often interconnected in real-world aquaculture scenarios and must be performed continuously in the same environment. Developing a joint model that can handle all three tasks simultaneously would be more memory-efficient and suitable for practical applications in aquaculture. A joint model would leverage the shared features and information among the tasks, reducing redundancy and improving overall performance. For example, accurate fish tracking can provide valuable counting and behaviour analysis information. In contrast, behaviour

analysis can help identify and resolve tracking challenges such as occlusions and interactions between fish. More comprehensive and efficient systems could be developed for monitoring and managing aquaculture farms by integrating these tasks into a single framework. This approach would also reduce the computational resources required, making it more feasible to deploy such systems in real-world settings. Future research should focus on developing novel architectures and training strategies that can effectively combine fish tracking, counting, and behaviour analysis tasks.

5) Integration of large language models (LLMs) and artificial general intelligence (AGI): Recent advancements in LLMs and AGI have the potential to revolutionize fish tracking, counting, and behaviour analysis. LLMs, such as GPT-4 [319] and LLaMA [320], can be fine-tuned on aquaculture-specific datasets to generate accurate descriptions and analyses of fish behaviour from textual data. AGI systems, like DeepMind's Gato [321], which can perform a wide range of tasks using a single model, could be adapted to integrate multiple modalities (e.g., vision, acoustics, and text) for comprehensive fish monitoring and management. By leveraging the power of LLMs and AGI, aquaculture researchers and practitioners can develop more intelligent and adaptable systems for understanding and optimizing fish welfare and production.

VIII. CONCLUSIONS

This survey provides a comprehensive analysis of the current state of digital technologies in aquaculture, including vision-based sensors, acoustic-based sensors, and biosensors, for fish tracking, counting, and behaviour analysis. These technologies offer valuable tools for optimizing production efficiency, fish welfare, and resource management in aquaculture. However, each technology has its limitations, such as the sensitivity of vision-based sensors to environmental conditions, the high cost and complexity of acoustic-based sensors, and the potential invasiveness of biosensors. Despite the advancements in these technologies, significant challenges remain, including the scarcity of comprehensive fish datasets, the lack of unified evaluation standards, and the need for more robust and adaptable systems that can handle the complexities of real-world aquaculture environments. To address these challenges and drive progress in the field, future research should focus on developing diverse and representative datasets, establishing standardized evaluation frameworks, and exploring integrating multiple technologies to create more comprehensive and reliable monitoring systems. Emerging technologies such as multimodal data fusion, deep learning, and edge computing present exciting opportunities for advancing digital aquaculture. By leveraging these technologies, more accurate, efficient, and practical solutions can be developed for fish tracking, counting, and behaviour analysis, ultimately contributing to the sustainable growth and development of the aquaculture industry. As the field progresses, it is crucial to consider these technologies' ethical implications and environmental impact. Developing solutions that are not only technologically advanced but also sustainable and respectful of animal welfare will be paramount for the future of aquaculture.

IX. AUTHOR CONTRIBUTIONS

Meng Cui: Conceptualization; writing – original draft. **Xubo Liu:** Investigation; validation; methodology; data curation. **Haohe Liu:** Validation; writing – original draft; data curation; formal analysis. **Jinzheng Zhao:** Validation; writing – review and editing. **Daoliang Li:** Funding acquisition; validation; project administration. **Wenwu Wang:** Funding acquisition; validation; project administration; resources; writing-review and editing.

X. ACKNOWLEDGMENT

This work was supported by the Research and demonstration of digital cage integrated monitoring system based on underwater robot [China grant 2022YFE0107100], Digital Fishery Cross-Innovative Talent Training Program of the China Scholarship Council (DF-Project) and a Research Scholarship from the China Scholarship Council (202006350248).

XI. DATA AVAILABILITY STATEMENT

Since this is a review paper, there is no data available. All information can be found in the cited references.

XII. CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest.

REFERENCES

- [1] T. Clavelle, S. E. Lester, R. Gentry, and H. E. Froehlich, "Interactions and management for the future of marine aquaculture and capture fisheries," *Fish and Fisheries*, vol. 20, no. 2, pp. 368–388, 2019.
- [2] A. G. Tacon, "Trends in global aquaculture and aquafeed production: 2000–2017," *Reviews in Fisheries Science & Aquaculture*, vol. 28, no. 1, pp. 43–56, 2020.
- [3] D. Li, Z. Du, Q. Wang, J. Wang, and L. Du, "Recent advances in acoustic technology for aquaculture: A review," *Reviews in Aquaculture*, vol. 16, no. 1, pp. 357–381, 2024.
- [4] D. An, J. Huang, and Y. Wei, "A survey of fish behaviour quantification indexes and methods in aquaculture," *Reviews in Aquaculture*, vol. 13, no. 4, pp. 2169–2189, 2021.
- [5] S. Duarte, L. Reig, and J. Oca, "Measurement of sole activity by digital image analysis," *Aquacultural Engineering*, vol. 41, no. 1, pp. 22–27, 2009.
- [6] L. Zhang, W. Li, C. Liu, X. Zhou, and Q. Duan, "Automatic fish counting method using image density grading and local regression," *Computers and Electronics in Agriculture*, vol. 179, p. 105844, 2020.
- [7] C. Zhou, D. Xu, K. Lin, C. Sun, and X. Yang, "Intelligent feeding control methods in aquaculture with an emphasis on fish: a review," *Reviews in Aquaculture*, vol. 10, no. 4, pp. 975–993, 2018.
- [8] D. V. Politikos, D. Kleftogiannis, K. Tsiaras, and K. A. Rose, "Movcufish: A data mining tool for discovering fish movement patterns from individual-based models," *Limnology and Oceanography: Methods*, vol. 19, no. 4, pp. 267–279, 2021.
- [9] D. Li, Z. Miao, F. Peng, L. Wang, Y. Hao, Z. Wang, T. Chen, H. Li, and Y. Zheng, "Automatic counting methods in aquaculture: A review," *Journal of the World Aquaculture Society*, vol. 52, no. 2, pp. 269–283, 2021.
- [10] V. Puig-Pons, P. Munoz-Benavent, V. Espinosa, G. Andreu-Garcia, J. M. Valiente-Gonzalez, V. D. Estruch, P. Ordóñez, I. Pérez-Arjona, V. Atienza, B. Melich, *et al.*, "Automatic bluefin tuna (*thunnus thynnus*) biomass estimation during transfers using acoustic and computer vision techniques," *Aquacultural Engineering*, vol. 85, pp. 22–31, 2019.
- [11] J. Delcourt, M. Denoel, M. Ylieff, and P. Poncin, "Video multitracking of fish behaviour: a synthesis and future perspectives," *Fish and Fisheries*, vol. 14, no. 2, pp. 186–204, 2013.
- [12] C. Xia, L. Fu, Z. Liu, H. Liu, L. Chen, and Y. Liu, "Aquatic toxic analysis by monitoring fish behavior using computer vision: A recent progress," *Journal of toxicology*, vol. 2018, 2018.

[13] L. Yang, Y. Liu, H. Yu, X. Fang, L. Song, D. Li, and Y. Chen, "Computer vision models in intelligent aquaculture with emphasis on fish detection and behavior analysis: a review," *Archives of Computational Methods in Engineering*, vol. 28, pp. 2785–2816, 2021.

[14] D. Li and L. Du, "Recent advances of deep learning algorithms for aquacultural machine vision systems with emphasis on fish," *Artificial Intelligence Review*, pp. 1–40, 2022.

[15] Y. Mei, B. Sun, D. Li, H. Yu, H. Qin, H. Liu, N. Yan, and Y. Chen, "Recent advances of target tracking applications in aquaculture with emphasis on fish," *Computers and Electronics in Agriculture*, vol. 201, p. 107335, 2022.

[16] M. Cui, X. Liu, J. Zhao, J. Sun, G. Lian, T. Chen, M. D. Plumbley, D. Li, and W. Wang, "Fish feeding intensity assessment in aquaculture: A new audio dataset AFFIA3K and a deep learning algorithm," in *2022 IEEE 32nd International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, IEEE, 2022.

[17] L. Zhang, J. Wang, and Q. Duan, "Estimation for fish mass using image analysis and neural network," *Computers and Electronics in Agriculture*, vol. 173, p. 105439, 2020.

[18] R. Soltanzadeh, B. Hardy, R. D. Mcleod, and M. R. Friesen, "A prototype system for real-time monitoring of arctic char in indoor aquaculture operations: Possibilities & challenges," *IEEE Access*, vol. 8, pp. 180815–180824, 2020.

[19] X. Yang, S. Zhang, J. Liu, Q. Gao, S. Dong, and C. Zhou, "Deep learning for smart fish farming: applications, opportunities and challenges," *Reviews in Aquaculture*, vol. 13, no. 1, pp. 66–90, 2021.

[20] J. Helminen and T. Linnansaari, "Object and behavior differentiation for improved automated counts of migrating river fish using imaging sonar data," *Fisheries Research*, vol. 237, p. 105883, 2021.

[21] F. Capoccioni, C. Leone, D. Pulcini, M. Cecchetti, A. Rossi, and E. Cicotti, "Fish movements and schooling behavior across the tidal channel in a mediterranean coastal lagoon: An automated approach using acoustic imaging," *Fisheries Research*, vol. 219, p. 105318, 2019.

[22] M. R. Eggleston, S. W. Milne, M. Ramsay, and K. P. Kowalski, "Improved fish counting method accurately quantifies high-density fish movement in dual-frequency identification sonar data files from a coastal wetland environment," *North American Journal of Fisheries Management*, vol. 40, no. 4, pp. 883–892, 2020.

[23] S. F. Colborne, D. W. Hondorp, C. M. Holbrook, M. R. Lowe, J. C. Boase, J. A. Chiotti, T. C. Wills, E. F. Roseman, and C. C. Krueger, "Sequence analysis and acoustic tracking of individual lake sturgeon identify multiple patterns of river-lake habitat use," *Ecosphere*, vol. 10, no. 12, p. e02983, 2019.

[24] C. Zhou, K. Lin, D. Xu, L. Chen, Q. Guo, C. Sun, and X. Yang, "Near infrared computer vision and neuro-fuzzy model-based feeding decision system for fish in aquaculture," *Computers and Electronics in Agriculture*, vol. 146, pp. 114–124, 2018.

[25] J. Kolarevic, J. Calduch-Giner, A. M. Espmark, T. Evensen, J. Sosa, and J. Perez-Sanchez, "A novel miniaturized biosensor for monitoring atlantic salmon swimming activity and respiratory frequency," *Animals*, vol. 11, no. 8, p. 2403, 2021.

[26] S. Shreesha, M. P. MM, U. Verma, and R. M. Pai, "Computer vision based fish tracking and behaviour detection system," in *2020 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, pp. 252–257, IEEE, 2020.

[27] Z. M. Qian and Y. Q. Chen, "Feature point based 3d tracking of multiple fish from multi-view images," *PloS One*, vol. 12, no. 6, p. e0180254, 2017.

[28] H. Wu, M. Murata, H. Matsumoto, H. Ohnuki, and H. Endo, "Integrated biosensor system for monitoring and visualizing fish stress response," *Sensors & Materials*, vol. 32, 2020.

[29] M. d. O. Barreiros, D. d. O. Dantas, L. C. d. O. Silva, S. Ribeiro, and A. K. Barros, "Zebrafish tracking using YOLOv2 and Kalman filter," *Scientific Reports*, vol. 11, no. 1, p. 3219, 2021.

[30] T. Liu, P. Li, H. Liu, X. Deng, H. Liu, and F. Zhai, "Multi-class fish stock statistics technology based on object classification and tracking algorithm," *Ecological Informatics*, vol. 63, p. 101240, 2021.

[31] Z. Wang, C. Xia, and J. Lee, "Parallel fish school tracking based on multiple appearance feature detection," *Sensors*, vol. 21, no. 10, p. 3476, 2021.

[32] S. H. Wang, X. E. Cheng, Z. M. Qian, Y. Liu, and Y. Q. Chen, "Automated planar tracking the waving bodies of multiple zebrafish swimming in shallow water," *PloS One*, vol. 11, no. 4, p. e0154714, 2016.

[33] X. Zhao, S. Yan, and Q. Gao, "An algorithm for tracking multiple fish based on biological water quality monitoring," *IEEE Access*, vol. 7, pp. 15018–15026, 2019.

[34] Z. Xu and X. E. Cheng, "Zebrafish tracking using convolutional neural networks," *Scientific Reports*, vol. 7, no. 1, p. 42815, 2017.

[35] C. Xia, T. S. Chon, Y. Liu, J. Chi, and J. Lee, "Posture tracking of multiple individual fish for behavioral monitoring with visual sensors," *Ecological Informatics*, vol. 36, pp. 190–198, 2016.

[36] A. Engel, Y. Reuben, I. Kolesnikov, D. Churilov, R. Nathan, and A. Genin, "In situ three-dimensional video tracking of tagged individuals within site-attached social groups of coral-reef fish," *Limnology and Oceanography: Methods*, vol. 19, no. 9, pp. 579–588, 2021.

[37] W. Li, F. Li, and Z. Li, "CMFTNet: Multiple fish tracking based on counterpoised jointnet," *Computers and Electronics in Agriculture*, vol. 198, p. 107018, 2022.

[38] W. Li, Y. Liu, W. Wang, Z. Li, and J. Yue, "TFMFT: Transformer-based multiple fish tracking," *Computers and Electronics in Agriculture*, vol. 217, p. 108600, 2024.

[39] S. H. Wang, J. Zhao, X. Liu, Z. M. Qian, Y. Liu, and Y. Q. Chen, "3D tracking swimming fish school with learned kinematic model using LSTM network," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1068–1072, IEEE, 2017.

[40] S. H. Wang, J. W. Zhao, and Y. Q. Chen, "Robust tracking of fish schools using CNN for head identification," *Multimedia Tools and Applications*, vol. 76, pp. 23679–23697, 2017.

[41] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409–1422, 2011.

[42] J. Wang, M. Zhao, L. Zou, Y. Hu, X. Cheng, and X. Liu, "Fish tracking based on improved TLD algorithm in real-world underwater environment," *Marine Technology Society Journal*, vol. 53, no. 3, pp. 80–89, 2019.

[43] K. Terayama, K. Hongo, H. Habe, and M. Sakagami, "Appearance-based multiple fish tracking for collective motion analysis," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 361–365, IEEE, 2015.

[44] K. Terayama, H. Habe, and M. Sakagami, "Multiple fish tracking with an NACA airfoil model for collective behavior analysis," *IPSI Transactions on Computer Vision and Applications*, vol. 8, pp. 1–7, 2016.

[45] 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 Bioinformatics*, vol. 17, pp. 1–11, 2016.

[46] A. Rodriguez, H. Zhang, J. Klaminder, T. Brodin, and M. Andersson, "ToxID: an efficient algorithm to solve occlusions when tracking multiple animals," *Scientific Reports*, vol. 7, no. 1, p. 14774, 2017.

[47] Y. Wageeh, H. E. Mohamed, A. Fadl, O. Anas, N. ElMasry, A. Nabil, and A. Atia, "YOLO fish detection with Euclidean tracking in fish farms," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 5–12, 2021.

[48] O. Anas, Y. Wageeh, H. E. Mohamed, A. Fadl, N. ElMasry, A. Nabil, and A. Atia, "Detecting abnormal fish behavior using motion trajectories in ubiquitous environments," *Procedia Computer Science*, vol. 175, pp. 141–148, 2020.

[49] H. E. Mohamed, A. Fadl, O. Anas, Y. Wageeh, N. ElMasry, A. Nabil, and A. Atia, "MSR-YOLO: Method to enhance fish detection and tracking in fish farms," *Procedia Computer Science*, vol. 170, pp. 539–546, 2020.

[50] Z. M. Qian, X. E. Cheng, and Y. Q. Chen, "Automatically detect and track multiple fish swimming in shallow water with frequent occlusion," *PloS One*, vol. 9, no. 9, p. e106506, 2014.

[51] C. Spampinato, E. Beauxis Aussalet, S. Palazzo, C. Beyan, J. van Ossenbruggen, J. He, B. Boom, and X. Huang, "A rule-based event detection system for real-life underwater domain," *Machine Vision and Applications*, vol. 25, pp. 99–117, 2014.

[52] C. Spampinato, S. Palazzo, B. Boom, J. van Ossenbruggen, I. Kavasidis, R. Di Salvo, F. P. Lin, D. Giordano, L. Hardman, and R. B. Fisher, "Understanding fish behavior during typhoon events in real-life underwater environments," *Multimedia Tools and Applications*, vol. 70, pp. 199–236, 2014.

[53] S. Chen, "Kalman filter for robot vision: a survey," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 11, pp. 4409–4420, 2011.

[54] A. Mathias, S. Dhanalakshmi, and R. Kumar, "Occlusion aware underwater object tracking using hybrid adaptive deep SORT-YOLOv3 approach," *Multimedia Tools and Applications*, vol. 81, no. 30, pp. 44109–44121, 2022.

[55] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3464–3468, IEEE, 2016.

[56] R. Pereira, G. Carvalho, L. Garrote, and U. J. Nunes, "Sort and deep-SORT based multi-object tracking for mobile robotics: evaluation with new data association metrics," *Applied Sciences*, vol. 12, no. 3, p. 1319, 2022.

[57] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3645–3649, IEEE, 2017.

[58] A. Farhadi and J. Redmon, "Yolov3: An incremental improvement," in *Computer vision and pattern recognition*, vol. 1804, pp. 1–6, Springer Berlin/Heidelberg, Germany, 2018.

[59] Y. Du, Z. Zhao, Y. Song, Y. Zhao, F. Su, T. Gong, and H. Meng, "StrongSORT: Make deepSORT great again," *IEEE Transactions on Multimedia*, vol. 25, pp. 8725–8737, 2023.

[60] S. Tang, S. Zhang, and Y. Fang, "HIC-YOLOv5: Improved YOLOv5 for small object detection," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6614–6619, IEEE, 2024.

[61] X. Zhai, H. Wei, H. Wu, Q. Zhao, and M. Huang, "Multi-target tracking algorithm in aquaculture monitoring based on deep learning," *Ocean Engineering*, vol. 289, p. 116005, 2023.

[62] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "BoT-SORT: Robust associations multi-pedestrian tracking," *arXiv preprint arXiv:2206.14651*, 2022.

[63] M. Sohan, T. Sai Ram, R. Reddy, and C. Venkata, "A review on YOLOv8 and its advancements," in *International Conference on Data Intelligence and Cognitive Informatics*, pp. 529–545, Springer, 2024.

[64] B. Xing, M. Sun, Z. Liu, L. Guan, J. Han, C. Yan, and C. Han, "Sonar fish school detection and counting method based on improved YOLOv8 and BoT-SORT," *Journal of Marine Science and Engineering*, vol. 12, no. 6, p. 964, 2024.

[65] J. Cao, J. Pang, X. Weng, R. Khirodkar, and K. Kitani, "Observation-centric sort: Rethinking sort for robust multi-object tracking," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9686–9696, 2023.

[66] G. Maggiolino, A. Ahmad, J. Cao, and K. Kitani, "Deep oc-sort: Multi-pedestrian tracking by adaptive re-identification," in *2023 IEEE International Conference on Image Processing (ICIP)*, pp. 3025–3029, IEEE, 2023.

[67] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, "Bytetrack: Multi-object tracking by associating every detection box," in *European conference on computer vision*, pp. 1–21, Springer, 2022.

[68] S. Y. Alaba, J. H. Prior, C. Shah, M. Nabi, J. E. Ball, R. Moorhead, M. D. Campbell, F. Wallace, and M. D. Grossi, "Multifish tracking for marine biodiversity monitoring," in *Ocean Sensing and Monitoring XVI*, vol. 13061, pp. 106–112, SPIE, 2024.

[69] H. Zhao, H. Cui, K. Qu, J. Zhu, H. Li, Z. Cui, and Y. Wu, "A fish appetite assessment method based on improved bytetrack and spatiotemporal graph convolutional network," *Biosystems Engineering*, vol. 240, pp. 46–55, 2024.

[70] Z.-M. Qian, X. Chen, and H. Jiang, "Fish tracking based on yolo and bytetrack," in *2023 16th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1–5, IEEE, 2023.

[71] A. Perez-Escudero, J. Vicente-Page, R. C. Hinz, S. Arganda, and G. G. De Polavieja, "idTracker: tracking individuals in a group by automatic identification of unmarked animals," *Nature Methods*, vol. 11, no. 7, pp. 743–748, 2014.

[72] A. Bhateja, B. Lall, P. K. Kalra, and K. Chaudhary, "Suze: A hybrid approach for multi-fish detection and tracking," in *Global Oceans 2020: Singapore-US Gulf Coast*, pp. 1–5, IEEE, 2020.

[73] F. Romero-Ferrero, M. G. Bergomi, R. C. Hinz, F. J. Heras, and G. G. De Polavieja, "Idtracker. ai: tracking all individuals in small or large collectives of unmarked animals," *Nature Methods*, vol. 16, no. 2, pp. 179–182, 2019.

[74] R. Gai, N. Chen, and H. Yuan, "A detection algorithm for cherry fruits based on the improved YOLO-v4 model," *Neural Computing and Applications*, vol. 35, no. 19, pp. 13895–13906, 2023.

[75] D. Chicco, "Siamese neural networks: An overview," *Artificial Neural Networks*, pp. 73–94, 2021.

[76] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[77] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "SiamRPN++: Evolution of siamese visual tracking with very deep networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4282–4291, 2019.

[78] H. Wang, S. Zhang, S. Zhao, Q. Wang, D. Li, and R. Zhao, "Real-time detection and tracking of fish abnormal behavior based on improved YOLOv5 and SiamRPN++," *Computers and Electronics in Agriculture*, vol. 192, p. 106512, 2022.

[79] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European Conference on Computer Vision*, pp. 213–229, Springer, 2020.

[80] X. Chen, B. Yan, J. Zhu, D. Wang, X. Yang, and H. Lu, "Transformer tracking," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8126–8135, 2021.

[81] B. Yan, H. Peng, J. Fu, D. Wang, and H. Lu, "Learning spatio-temporal transformer for visual tracking," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10448–10457, 2021.

[82] Y. Liu, B. Li, X. Zhou, D. Li, and Q. Duan, "Fishtrack: Multi-object tracking method for fish using spatiotemporal information fusion," *Expert Systems with Applications*, vol. 238, p. 122194, 2024.

[83] M. G. B. Palconit, R. S. Concepcion II, J. D. Alejandrino, M. E. Pareja, V. J. D. Almero, A. A. Bandala, R. R. P. Vicerra, E. Sybingco, E. P. Dadios, and R. N. Naguib, "Three-dimensional stereo vision tracking of multiple free-swimming fish for low frame rate video," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 25, no. 5, pp. 639–646, 2021.

[84] J. Mao, G. Xiao, W. Sheng, Z. Qu, and Y. Liu, "Research on realizing the 3D occlusion tracking location method of fish's school target," *Neurocomputing*, vol. 214, pp. 61–79, 2016.

[85] G. Xiao, W. K. Fan, J. F. Mao, Z. B. Cheng, D. H. Zhong, and Y. Li, "Research of the fish tracking method with occlusion based on monocular stereo vision," in *2016 International Conference on Information System and Artificial Intelligence (ISAI)*, pp. 581–589, IEEE, 2016.

[86] X. E. Cheng, S. S. Du, H. Y. Li, J. F. Hu, and M. L. Chen, "Obtaining three-dimensional trajectory of multiple fish in water tank via video tracking," *Multimedia Tools and Applications*, vol. 77, pp. 24499–24519, 2018.

[87] S. H. Wang, X. Liu, J. Zhao, Y. Liu, and Y. Q. Chen, "3D tracking swimming fish school using a master view tracking first strategy," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 516–519, IEEE, 2016.

[88] X. Liu, Y. Yue, M. Shi, and Z. M. Qian, "3-D video tracking of multiple fish in water tank," *IEEE Access*, vol. 7, pp. 145049–145059, 2019.

[89] M. Pedersen, J. B. Haurum, S. H. Bengtson, and T. B. Moeslund, "3D-Zef: A 3D zebrafish tracking benchmark dataset," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2426–2436, 2020.

[90] J. Martinez, T. Fu, X. Li, H. Hou, J. Wang, M. B. Eppard, and Z. D. Deng, "A large dataset of detection and submeter-accurate 3-d trajectories of juvenile chinook salmon," *Scientific Data*, vol. 8, no. 1, p. 211, 2021.

[91] D. Li, Z. Du, Q. Wang, J. Wang, and L. Du, "Recent advances in acoustic technology for aquaculture: A review," *Reviews in Aquaculture*, vol. 16, no. 1, pp. 357–381, 2024.

[92] A. Pursche, C. Walsh, and M. Taylor, "Evaluation of a novel external tag-mount for acoustic tracking of small fish," *Fisheries Management and Ecology*, vol. 21, no. 2, pp. 169–172, 2014.

[93] J. K. Matley, N. V. Klinard, A. P. B. Martins, K. Aarestrup, E. Aspíllaga, S. J. Cooke, P. D. Cowley, M. R. Heupel, C. G. Lowe, S. K. Lowerre-Barbieri, *et al.*, "Global trends in aquatic animal tracking with acoustic telemetry," *Trends in Ecology & Evolution*, vol. 37, no. 1, pp. 79–94, 2022.

[94] E. Aspíllaga, R. Arlinghaus, M. Martorell-Barcelo, M. Barcelo-Serra, and J. Alos, "High-throughput tracking of social networks in marine fish populations," *Frontiers in Marine Science*, p. 794, 2021.

[95] R. J. Lennox, K. Aarestrup, S. J. Cooke, P. D. Cowley, Z. D. Deng, A. T. Fisk, R. G. Harcourt, M. Heupel, S. G. Hinch, K. N. Holland, *et al.*, "Envisioning the future of aquatic animal tracking: technology, science, and application," *BioScience*, vol. 67, no. 10, pp. 884–896, 2017.

[96] J. Kolarevic, O. Aas-Hansen, A. Espmark, G. Baeverfjord, B. F. Terjesen, and B. Damsgard, "The use of acoustic acceleration transmitter tags for monitoring of atlantic salmon swimming activity in recirculating aquaculture systems (ras)," *Aquacultural Engineering*, vol. 72, pp. 30–39, 2016.

[97] M. Fore, K. Frank, T. Dempster, J. A. Alfredsen, and E. Hoy, "Biomonitoring using tagged sentinel fish and acoustic telemetry in commercial salmon aquaculture: a feasibility study," *Aquacultural Engineering*, vol. 78, pp. 163–172, 2017.

[98] E. Leclercq, B. Zerafa, A. J. Brooker, A. Davie, and H. Migaud, "Application of passive-acoustic telemetry to explore the behaviour of

ballan wrasse (*labrus bergylta*) and lumpfish (*cyclopterus lumpus*) in commercial scottish salmon sea-pens," *Aquaculture*, vol. 495, pp. 1–12, 2018.

[99] L. Munoz, E. Aspíllaga, M. Palmer, J. L. Saraiva, and P. Arechavalalopez, "Acoustic telemetry: a tool to monitor fish swimming behavior in sea-cage aquaculture," *Frontiers in Marine Science*, vol. 7, p. 645, 2020.

[100] G. Macaulay, F. Warren-Myers, L. T. Barrett, F. Oppedal, M. Fore, and T. Dempster, "Tag use to monitor fish behaviour in aquaculture: a review of benefits, problems and solutions," *Reviews in Aquaculture*, vol. 13, no. 3, pp. 1565–1582, 2021.

[101] J. Macaulay, A. Kingston, A. Coram, M. Oswald, R. Swift, D. Gillespie, and S. Northridge, "Passive acoustic tracking of the three-dimensional movements and acoustic behaviour of toothed whales in close proximity to static nets," *Methods in Ecology and Evolution*, vol. 13, no. 6, pp. 1250–1264, 2022.

[102] N. V. Klinard and J. K. Matley, "Living until proven dead: addressing mortality in acoustic telemetry research," *Reviews in Fish Biology and Fisheries*, vol. 30, no. 3, pp. 485–499, 2020.

[103] D. V. Notte, R. J. Lennox, D. C. Hardie, and G. T. Crossin, "Application of machine learning and acoustic predation tags to classify migration fate of atlantic salmon smolts," *Oecologia*, vol. 198, no. 3, pp. 605–618, 2022.

[104] P. Dendorfer, H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid, S. Roth, K. Schindler, and L. Leal-Taix'e, "Motchallenge: A benchmark for single-camera multiple target tracking," *International Journal of Computer Vision*, vol. 129, no. 4, pp. 845–881, 2021.

[105] T. Mandel, M. Jimenez, E. Risley, T. Nammoto, R. Williams, M. Panoff, M. Ballesteros, and B. Suarez, "Detection confidence driven multi-object tracking to recover reliable tracks from unreliable detections," *Pattern Recognition*, vol. 135, p. 109107, 2023.

[106] Z. Wu, C. Wang, W. Zhang, G. Sun, W. Ke, and Z. Xiong, "Online 3d behavioral tracking of aquatic model organism with a dual-camera system," *Advanced Engineering Informatics*, vol. 61, p. 102481, 2024.

[107] S. Liu, X. Zheng, L. Han, X. Liu, J. Ren, F. Wang, Y. Liu, and Y. Lin, "FishMOT: A simple and effective method for fish tracking based on IoU matching," *Applied Engineering in Agriculture*, 2024.

[108] Y.-L. Fan, C.-H. Hsu, F.-R. Hsu, and L.-D. Liao, "Exploring the use of deep learning models for accurate tracking of 3d zebrafish trajectories," *Frontiers in Bioengineering and Biotechnology*, vol. 12, p. 1461264, 2024.

[109] A. L. Koeberle, W. Pearsall, B. E. Hammers, D. Mulhall, J. E. McKenna Jr, M. Chalupnicki, and S. A. Sethi, "Whole-lake acoustic telemetry to evaluate survival of stocked juvenile fish," *Scientific Reports*, vol. 13, no. 1, p. 18956, 2023.

[110] J. K. Matley, N. V. Klinard, A. B. Martins, A. Oakley-Cogan, C. Huveneers, C. S. Vandergoot, and A. T. Fisk, "TrackdAT, an acoustic telemetry metadata dataset to support aquatic animal tracking research," *Scientific Data*, vol. 11, no. 1, p. 143, 2024.

[111] R. J. Lennox, K. Aarestrup, J. Alós, R. Arlinghaus, E. Aspíllaga, M. G. Bertram, K. Birnie-Gauvin, T. Brodin, S. J. Cooke, L. S. Dahlmo, *et al.*, "Positioning aquatic animals with acoustic transmitters," *Methods in Ecology and Evolution*, vol. 14, no. 10, pp. 2514–2530, 2023.

[112] A. M. Kanigan, S. G. Hinch, A. G. Lotto, K. Szlachta, S. D. Johnston, and S. A. Lingard, "Acoustic telemetry tracking of coho salmon smolts released from a community-run hatchery into a marine inlet reveals low early ocean survival," *North American Journal of Fisheries Management*, vol. 44, no. 2, pp. 359–376, 2024.

[113] J. Luo, Z. Sun, L. Lu, Z. Xiong, L. Cui, and Z. Mao, "Rapid expansion of coastal aquaculture ponds in southeast asia: Patterns, drivers and impacts," *Journal of Environmental Management*, vol. 315, p. 115100, 2022.

[114] H. Kuroda, "History, current status, and future vision of particle-tracking simulation applied to marine biology, fisheries science, and ecological engineering around japan," *Fisheries Science*, vol. 89, no. 2, pp. 129–146, 2023.

[115] A. Looby, K. Cox, S. Bravo, R. Rountree, F. Juanes, L. K. Reynolds, and C. W. Martin, "A quantitative inventory of global soniferous fish diversity," *Reviews in Fish Biology and Fisheries*, vol. 32, no. 2, pp. 581–595, 2022.

[116] L. Zhang, W. Guo, C. Lv, M. Guo, M. Yang, Q. Fu, and X. Liu, "Advancements in artificial intelligence technology for improving animal welfare: Current applications and research progress," *Animal Research and One Health*, vol. 2, no. 1, pp. 93–109, 2024.

[117] L. Cai, N. E. McGuire, R. Hanlon, T. A. Mooney, and Y. Girdhar, "Semi-supervised visual tracking of marine animals using autonomous underwater vehicles," *International Journal of Computer Vision*, vol. 131, no. 6, pp. 1406–1427, 2023.

[118] Y. Liu, W. Li, X. Liu, Z. Li, and J. Yue, "Deep learning in multiple animal tracking: A survey," *Computers and Electronics in Agriculture*, vol. 224, p. 109161, 2024.

[119] P. Verhelst, R. Brys, S. J. Cooke, I. Pauwels, M. Rohtla, and J. Reubens, "Enhancing our understanding of fish movement ecology through interdisciplinary and cross-boundary research," *Reviews in Fish Biology and Fisheries*, vol. 33, no. 1, pp. 111–135, 2023.

[120] J. Matley, N. Klinard, S. Larocque, M. McLean, J. Brownscombe, G. Raby, V. Nguyen, and A. Barbosa Martins, "Making the most of aquatic animal tracking: a review of complementary methods to bolster acoustic telemetry," *Reviews in Fish Biology and Fisheries*, vol. 33, no. 1, pp. 35–54, 2023.

[121] A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, S. Popov, M. Mallochi, A. Kolesnikov, *et al.*, "The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale," *International journal of computer vision*, vol. 128, no. 7, pp. 1956–1981, 2020.

[122] J. Zhao, Y. Xu, X. Qian, D. Berghi, P. Wu, M. Cui, J. Sun, P. J. Jackson, and W. Wang, "Audio-visual speaker tracking: Progress, challenges, and future directions," *arXiv preprint arXiv:2310.14778*, 2023.

[123] X. Qian, A. Brutti, O. Lanz, M. Omologo, and A. Cavallaro, "Multi-speaker tracking from an audio-visual sensing device," *IEEE Transactions on Multimedia*, vol. 21, no. 10, pp. 2576–2588, 2019.

[124] J. Zhao, P. Wu, X. Liu, Y. Xu, L. Mihaylova, S. Godsill, and W. Wang, "Audio-visual tracking of multiple speakers via a pmbm filter," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5068–5072, IEEE, 2022.

[125] R. Ewing, M. Evenson, and E. Birks, "Infrared fish counter for measuring migration of juvenile salmonids," *The Progressive Fish-Culturist*, vol. 45, no. 1, pp. 53–55, 1983.

[126] S. Cadieux, F. Michaud, and F. Lalonde, "Intelligent system for automated fish sorting and counting," in *Proceedings. 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000)(Cat. No. 00CH37113)*, vol. 2, pp. 1279–1284, IEEE, 2000.

[127] F. Ferrero, J. Campo, M. Valledor, and M. Hernando, "Optical systems for the detection and recognition of fish in rivers," in *2014 IEEE 11th International Multi-Conference on Systems, Signals & Devices (SSD14)*, pp. 1–5, IEEE, 2014.

[128] J. Santos, P. Pinheiro, M. Ferreira, and J. Bochechas, "Monitoring fish passes using infrared beaming: a case study in an iberian river," *Journal of Applied Ichthyology*, vol. 24, no. 1, pp. 26–30, 2008.

[129] L. Baumgartner, M. Bettanin, J. McPherson, M. Jones, B. Zampatti, and K. Beyer, "Influence of turbidity and passage rate on the efficiency of an infrared counter to enumerate and measure riverine fish," *Journal of Applied Ichthyology*, vol. 28, no. 4, pp. 531–536, 2012.

[130] I. Klapp, O. Arad, L. Rosenfeld, A. Barki, B. Shaked, and B. Zion, "Ornamental fish counting by non-imaging optical system for real-time applications," *Computers and Electronics in Agriculture*, vol. 153, pp. 126–133, 2018.

[131] T. Shardlow and K. Hyatt, "Assessment of the counting accuracy of the vaki infrared counter on chum salmon," *North American Journal of Fisheries Management*, vol. 24, no. 1, pp. 249–252, 2004.

[132] D. Li, Y. Hao, and Y. Duan, "Nonintrusive methods for biomass estimation in aquaculture with emphasis on fish: a review," *Reviews in Aquaculture*, vol. 12, no. 3, pp. 1390–1411, 2020.

[133] C. Haas, P. K. Thumser, M. Hellmair, T. J. Pilger, and M. Schletterer, "Monitoring of fish migration in fishways and rivers—the infrared fish counter 'riverwatcher' as a suitable tool for long-term monitoring," *Water*, vol. 16, no. 3, p. 477, 2024.

[134] W. Beaumont, C. Mills, and G. Williams, "Use of a microcomputer as an aid to identifying objects passing through a resistivity fish counter," *Aquaculture Research*, vol. 17, no. 3, pp. 213–226, 1986.

[135] D. Dunkley and W. Shearer, "An assessment of the performance of a resistivity fish counter," *Journal of Fish Biology*, vol. 20, no. 6, pp. 717–737, 1982.

[136] H. Forbes, G. Smith, A. Johnstone, and A. Stephen, "An assessment of the performance of the resistivity fish counter in the borland lift fish pass at dundreggan dam on the river moriston," *Fisheries Research Services Report No 02*, p. 13pp, 2000.

[137] M. Aprahamian, S. Nicholson, D. McCubbing, and I. Davidson, "The use of resistivity fish counters in fish stock assessment," *Stock Assessment in Inland Waters ed I. Cowx*, pp. 27–43, 1996.

[138] J. J. Sheppard and M. S. Bednarski, "Utility of single-channel electronic resistivity counters for monitoring river herring populations,"

North American Journal of Fisheries Management, vol. 35, no. 6, pp. 1144–1151, 2015.

[139] A. J. Cheal and M. J. Emslie, “Counts of coral reef fishes by an experienced observer are not biased by the number of target species,” *Journal of Fish Biology*, vol. 97, no. 4, pp. 1063–1071, 2020.

[140] M. P. Pais and H. N. Cabral, “Effect of underwater visual survey methodology on bias and precision of fish counts: a simulation approach,” *PeerJ*, vol. 6, p. e5378, 2018.

[141] M. Saberioon, A. Gholizadeh, P. Cisar, A. Pautsina, and J. Urban, “Application of machine vision systems in aquaculture with emphasis on fish: state-of-the-art and key issues,” *Reviews in Aquaculture*, vol. 9, no. 4, pp. 369–387, 2017.

[142] S. Zhang, X. Yang, Y. Wang, Z. Zhao, J. Liu, Y. Liu, C. Sun, and C. Zhou, “Automatic fish population counting by machine vision and a hybrid deep neural network model,” *Animals*, vol. 10, no. 2, p. 364, 2020.

[143] Y. Duan, L. H. Stien, A. Thorsen, O. Karlsen, N. Sandlund, D. Li, Z. Fu, and S. Meier, “An automatic counting system for transparent pelagic fish eggs based on computer vision,” *Aquacultural Engineering*, vol. 67, pp. 8–13, 2015.

[144] M. R. Shortis, M. Ravanbakhsh, F. Shafait, and A. Mian, “Progress in the automated identification, measurement, and counting of fish in underwater image sequences,” *Marine Technology Society Journal*, vol. 50, no. 1, pp. 4–16, 2016.

[145] J. Li, C. Xu, L. Jiang, Y. Xiao, L. Deng, and Z. Han, “Detection and analysis of behavior trajectory for sea cucumbers based on deep learning,” *IEEE Access*, vol. 8, pp. 18832–18840, 2019.

[146] C. Arteta, V. Lempitsky, J. A. Noble, and A. Zisserman, “Interactive object counting,” in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part III 13*, pp. 504–518, Springer, 2014.

[147] L. Fiaschi, U. Kothe, R. Nair, and F. A. Hamprecht, “Learning to count with regression forest and structured labels,” in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pp. 2685–2688, IEEE, 2012.

[148] H. Liu, X. Ma, Y. Yu, L. Wang, and L. Hao, “Application of deep learning-based object detection techniques in fish aquaculture: a review,” *Journal of Marine Science and Engineering*, vol. 11, no. 4, p. 867, 2023.

[149] A. Saleh, M. Sheaves, D. Jerry, and M. R. Azghadi, “Applications of deep learning in fish habitat monitoring: A tutorial and survey,” *Expert Systems with Applications*, p. 121841, 2023.

[150] X. Yu, Y. Wang, D. An, and Y. Wei, “Counting method for cultured fishes based on multi-modules and attention mechanism,” *Aquacultural Engineering*, vol. 96, p. 102215, 2022.

[151] S. Zhang, X. Yang, Y. Wang, Z. Zhao, J. Liu, Y. Liu, C. Sun, and C. Zhou, “Automatic fish population counting by machine vision and a hybrid deep neural network model,” *Animals*, vol. 10, no. 2, p. 364, 2020.

[152] G. Xu, Q. Chen, T. Yoshida, K. Teravama, Y. Mizukami, Q. Li, and D. Kitazawa, “Detection of bluefin tuna by cascade classifier and deep learning for monitoring fish resources,” in *Global Oceans 2020: Singapore–US Gulf Coast*, pp. 1–4, IEEE, 2020.

[153] P. L. F. Albuquerque, V. Garcia, A. d. S. O. Junior, T. Lewandowski, C. Detweiler, A. B. Gonçalves, C. S. Costa, M. H. Naka, and H. Pistori, “Automatic live fingerlings counting using computer vision,” *Computers and Electronics in Agriculture*, vol. 167, p. 105015, 2019.

[154] S. M. D. Lainez and D. B. Gonzales, “Automated fingerlings counting using convolutional neural network,” in *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, pp. 67–72, IEEE, 2019.

[155] L. Coronel, W. Badoy, and C. Namoco, “Identification of an efficient filtering-segmentation technique for automated counting of fish fingerlings,” *Int. Arab J. Inf. Technol.*, vol. 15, no. 4, pp. 708–714, 2018.

[156] J. M. Hernandez-Ontiveros, E. Inzunza-Gonzalez, E. E. Garcia-Guerrero, O. R. Lopez-Bonilla, S. O. Infante-Prieto, J. R. Cardenas-Valdez, and E. Tlelo-Cuautle, “Development and implementation of a fish counter by using an embedded system,” *Computers and Electronics in Agriculture*, vol. 145, pp. 53–62, 2018.

[157] S. Abe, T. Takagi, K. Takehara, N. Kimura, T. Hiraishi, K. Komeyama, S. Torisawa, and S. Asaumi, “How many fish in a tank? constructing an automated fish counting system by using pvt analysis,” in *Selected Papers from the 31st International Congress on High-Speed Imaging and Photonics*, vol. 10328, pp. 380–384, SPIE, 2017.

[158] L. Fan and Y. Liu, “Automate fry counting using computer vision and multi-class least squares support vector machine,” *Aquaculture*, vol. 380, pp. 91–98, 2013.

[159] W. Li, Q. Zhu, H. Zhang, Z. Xu, and Z. Li, “A lightweight network for portable fry counting devices,” *Applied Soft Computing*, vol. 136, p. 110140, 2023.

[160] H. Zhang, W. Li, Y. Qi, H. Liu, and Z. Li, “Dynamic fry counting based on multi-object tracking and one-stage detection,” *Computers and Electronics in Agriculture*, vol. 209, p. 107871, 2023.

[161] M. T. Tran, D. H. Kim, C. K. Kim, H. K. Kim, and S. B. Kim, “Determination of injury rate on fish surface based on fuzzy c-means clustering algorithm and 1 ab color space using zed stereo camera,” in *2018 15th International Conference on Ubiquitous Robots (UR)*, pp. 466–471, IEEE, 2018.

[162] P. F. Newbury, P. F. Culverhouse, and D. A. Pilgrim, “Automatic fish population counting by artificial neural network,” *Aquaculture*, vol. 133, no. 1, pp. 45–55, 1995.

[163] B. Al-Saaidah, W. Al-Nuaimy, M. R. Al-Hadidi, and I. Young, “Automatic counting system for zebrafish eggs using optical scanner,” in *2018 9th International Conference on Information and Communication Systems (ICICS)*, pp. 107–110, IEEE, 2018.

[164] I. Aliyu, K. J. Gana, A. A. Musa, M. A. Adegbeye, and C. G. Lim, “Incorporating recognition in catfish counting algorithm using artificial neural network and geometry,” *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 14, no. 12, pp. 4866–4888, 2020.

[165] J. Le and L. Xu, “An automated fish counting algorithm in aquaculture based on image processing,” in *2016 international forum on mechanical, control and automation (IFMCA 2016)*, pp. 358–366, Atlantis Press, 2017.

[166] A. B. Labao and P. C. Naval Jr, “Cascaded deep network systems with linked ensemble components for underwater fish detection in the wild,” *Ecological Informatics*, vol. 52, pp. 103–121, 2019.

[167] N. Liu, Y. Long, C. Zou, Q. Niu, L. Pan, and H. Wu, “Adcrowdnet: An attention-injective deformable convolutional network for crowd understanding,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3225–3234, 2019.

[168] F. Shafait, A. Mian, M. Shortis, B. Ghanem, P. F. Culverhouse, D. Edgington, D. Cline, M. Ravanbakhsh, J. Seager, and E. S. Harvey, “Fish identification from videos captured in uncontrolled underwater environments,” *ICES Journal of Marine Science*, vol. 73, no. 10, pp. 2737–2746, 2016.

[169] K. M. Pai, K. A. Shenoy, and M. M. Pai, “A computer vision based behavioral study and fish counting in a controlled environment,” *IEEE Access*, vol. 10, pp. 87778–87786, 2022.

[170] J. Zhou, D. Ji, J. Zhao, S. Zhu, Z. Peng, G. Lu, and Z. Ye, “A kinematic analysis-based on-line fingerlings counting method using low-frame-rate camera,” *Computers and Electronics in Agriculture*, vol. 199, p. 107193, 2022.

[171] H. Ghahremannezhad, H. Shi, and C. Liu, “Object detection in traffic videos: A survey,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 7, pp. 6780–6799, 2023.

[172] H. Liu, T. Liu, Y. Gu, P. Li, F. Zhai, H. Huang, and S. He, “A high-density fish school segmentation framework for biomass statistics in a deep-sea cage,” *Ecological Informatics*, vol. 64, p. 101367, 2021.

[173] G. Wang, J. Yu, W. Xu, A. Muhammad, and D. Li, “Automated fish counting system based on instance segmentation in aquaculture,” *Expert Systems with Applications*, p. 125318, 2024.

[174] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, *et al.*, “Segment anything,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4015–4026, 2023.

[175] F. Martignac, A. Daroux, J.-L. Bagliniere, D. Ombredane, and J. Guillard, “The use of acoustic cameras in shallow waters: new hydroacoustic tools for monitoring migratory fish population. a review of DIDSON technology,” *Fish and Fisheries*, vol. 16, no. 3, pp. 486–510, 2015.

[176] C. Lagasse, M. Bartel-Sawatzky, J. Nelitz, and Y. Xie, “Assessment of Adaptive Resolution Imaging Sonar (ARIS) for fish counting and measurements of fish length and swim speed in the lower Fraser river, year two: a final project report to the Southern Boundary Restoration and Enhancement Fund,” *Pacific Salmon Commission*, 2017.

[177] E. Belcher, W. Hanot, and J. Burch, “Dual-frequency identification sonar (DIDSON),” in *Proceedings of the 2002 International Symposium on Underwater Technology (Cat. No. 02EX556)*, pp. 187–192, IEEE, 2002.

[178] G. Cronkite, H. Enzenhofer, T. Ridley, J. Holmes, J. Lilja, and K. Benner, “Use of high-frequency imaging sonar to estimate adult sockeye salmon escapement in the Horsefly River, British Columbia,” *Canadian Technical Report of Fisheries and Aquatic Sciences*, vol. 2647, 2006.

[179] S. L. Maxwell and N. E. Gove, “The feasibility of estimating migrating salmon passage rates in turbid rivers using a dual frequency identifica-

tion sonar (DIDSON)," *Alaska Department of Fish and Game Regional Information Report*, no. 2A04-05, 2004.

[180] R. Lagarde, J. Peyre, E. Amilhat, M. Mercader, F. Prellwitz, G. Simon, and E. Faliex, "In situ evaluation of European eel counts and length estimates accuracy from an acoustic camera (ARIS)," *Knowledge & Management of Aquatic Ecosystems*, no. 421, p. 44, 2020.

[181] S. L. Maxwell and N. E. Gove, "Assessing a dual-frequency identification sonars' fish-counting accuracy, precision, and turbid river range capability," *The Journal of the Acoustical Society of America*, vol. 122, no. 6, pp. 3364–3377, 2007.

[182] I. C. Petreman, N. E. Jones, and S. W. Milne, "Observer bias and subsampling efficiencies for estimating the number of migrating fish in rivers using dual-frequency identification sonar (DIDSON)," *Fisheries Research*, vol. 155, pp. 160–167, 2014.

[183] R. Connolly, K. Jinks, A. Shand, M. Taylor, T. Gaston, A. Becker, and E. Jinks, "Out of the shadows: Automatic fish detection from acoustic cameras," *Aquatic Ecology*, vol. 57, no. 4, pp. 833–844, 2023.

[184] K. M. Boswell, M. P. Wilson, and J. H. Cowan Jr, "A semiautomated approach to estimating fish size, abundance, and behavior from dual-frequency identification sonar (DIDSON) data," *North American Journal of Fisheries Management*, vol. 28, no. 3, pp. 799–807, 2008.

[185] J. B. Hughes and J. E. Hightower, "Combining split-beam and dual-frequency identification sonars to estimate abundance of anadromous fishes in the Roanoke River, North Carolina," *North American Journal of Fisheries Management*, vol. 35, no. 2, pp. 229–240, 2015.

[186] A. Berghuis, *Performance of a single frequency split-beam hydroacoustic system: an innovative fish counting technology*. Arthur Rylah Institute for Environmental Research, 2008.

[187] J. Han, A. Asada, and M. Mizoguchi, "DIDSON-based acoustic counting method for juvenile ayu plecoglossus altivelis migrating upstream," *The Journal of the Marine Acoustics Society of Japan*, vol. 36, no. 4, pp. 250–257, 2009.

[188] E. Mora, S. Lindley, D. Erickson, and A. Klimley, "Estimating the riverine abundance of green sturgeon using a dual-frequency identification sonar," *North American Journal of Fisheries Management*, vol. 35, no. 3, pp. 557–566, 2015.

[189] M. Kang, "Semiautomated analysis of data from an imaging sonar for fish counting, sizing, and tracking in a post-processing application," *Fisheries and Aquatic Sciences*, vol. 14, no. 3, pp. 218–225, 2011.

[190] W. Shen, Z. Peng, and J. Zhang, "Identification and counting of fish targets using adaptive resolution imaging sonar," *Journal of Fish Biology*, vol. 104, no. 2, pp. 422–432, 2024.

[191] Y. Duan, S. Zhang, Y. Liu, J. Liu, D. An, and Y. Wei, "Boosting fish counting in sonar images with global attention and point supervision," *Engineering Applications of Artificial Intelligence*, vol. 126, p. 107093, 2023.

[192] R. E. Jones, R. A. Griffin, and R. K. Unsworth, "Adaptive Resolution Imaging Sonar (ARIS) as a tool for marine fish identification," *Fisheries Research*, vol. 243, p. 106092, 2021.

[193] S. Shahrestani, H. Bi, V. Lyubchich, and K. M. Boswell, "Detecting a nearshore fish parade using the adaptive resolution imaging sonar (ARIS): An automated procedure for data analysis," *Fisheries Research*, vol. 191, pp. 190–199, 2017.

[194] L. Egg, J. Pander, M. Mueller, and J. Geist, "Comparison of sonar, camera and net-based methods in detecting riverine fish movement patterns," *Marine and Freshwater Research*, vol. 69, no. 12, pp. 1905–1912, 2018.

[195] J. A. Holmes, G. M. Cronkite, H. J. Enzenhofer, and T. J. Mulligan, "Accuracy and precision of fish-count data from a "dual-frequency identification sonar"(DIDSON) imaging system," *ICES Journal of Marine Science*, vol. 63, no. 3, pp. 543–555, 2006.

[196] D. Jing, J. Han, X. Wang, G. Wang, J. Tong, W. Shen, and J. Zhang, "A method to estimate the abundance of fish based on dual-frequency identification sonar (DIDSON) imaging," *Fisheries Science*, vol. 83, pp. 685–697, 2017.

[197] A. Le Quinio, E. De Oliveira, A. Girard, J. Guillard, J.-M. Roussel, F. Zaoui, and F. Martignac, "Automatic detection, identification and counting of anguilliform fish using in situ acoustic camera data: Development of a cross-camera morphological analysis approach," *PLoS One*, vol. 18, no. 2, p. e0273588, 2023.

[198] J. Wanzenbock, T. Mehner, M. Schulz, H. Gassner, and I. J. Winfield, "Quality assurance of hydroacoustic surveys: the repeatability of fish-abundance and biomass estimates in lakes within and between hydroacoustic systems," *ICES Journal of Marine Science*, vol. 60, no. 3, pp. 486–492, 2003.

[199] D. C. Mesiar, D. M. Eggers, and D. M. Gaudet, "Development of techniques for the application of hydroacoustics to counting migratory fish in large rivers," *Rapports et Proces Verbaux des Reunions, Conseil International pour l'Exploration de la Mer*, vol. 189, pp. 223–232, 1990.

[200] X. Zhao and E. Ona, "Estimation and compensation models for the shadowing effect in dense fish aggregations," *ICES Journal of Marine Science*, vol. 60, no. 1, pp. 155–163, 2003.

[201] Y. Nishimori, K. Iida, M. Furusawa, Y. Tang, K. Tokuyama, S. Nagai, and Y. Nishiyama, "The development and evaluation of a three-dimensional, echo-integration method for estimating fish-school abundance," *ICES Journal of Marine Science*, vol. 66, no. 6, pp. 1037–1042, 2009.

[202] Y. Takao and M. Furusawa, "Dual-beam echo integration method for precise acoustic surveys," *ICES Journal of Marine Science*, vol. 53, no. 2, pp. 351–358, 1996.

[203] J. Simmonds and D. N. MacLennan, *Fisheries acoustics: theory and practice*. John Wiley & Sons, 2008.

[204] V. Espinosa, E. Soliveres, V. D. Estruch, J. Redondo, M. Ardid, J. Alba, E. Escuder, and M. Bou, "Acoustical monitoring of open mediterranean sea fish farms: Problems and strategies," in *EAA European Symposium on Hydroacoustics. Gandia. FAO*, pp. 1–75, 1994.

[205] S. G. Conti, P. Roux, C. Fauvel, B. D. Maurer, and D. A. Demer, "Acoustical monitoring of fish density, behavior, and growth rate in a tank," *Aquaculture*, vol. 251, no. 2-4, pp. 314–323, 2006.

[206] P. Sthapit, M. Kim, and K. Kim, "A method to accurately estimate fish abundance in offshore cages," *Applied Sciences*, vol. 10, no. 11, p. 3720, 2020.

[207] P. Sthapit, Y. Teekaraman, K. MinSeok, and K. Kim, "Algorithm to estimation fish population using echosounder in fish farming net," in *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 587–590, IEEE, 2019.

[208] A. Bjordal, J. Ejuell, T. Lindem, and A. Ferno, "Hydroacoustic monitoring and feeding control in cage rearing of Atlantic salmon (*Salmo salar L.*)," in *Fish Farming Technology*, pp. 203–208, CRC Press, 2020.

[209] M. Godlewski, M. Colon, L. Doroszczyk, B. Dlugoszewski, C. Verges, and J. Guillard, "Hydroacoustic measurements at two frequencies: 70 and 120 khz—consequences for fish stock estimation," *Fisheries Research*, vol. 96, no. 1, pp. 11–16, 2009.

[210] H. L. Aquino, R. S. Concepcion, E. P. Dadios, E. Sybingco, and A. A. Bandala, "Trend forecasting of computer vision application in aquaponic cropping systems industry," in *2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, pp. 1–6, IEEE, 2020.

[211] A. Heenan, I. D. Williams, T. Acoba, A. DesRochers, R. K. Kosaki, T. Kanemura, M. O. Nadon, and R. E. Brainard, "Long-term monitoring of coral reef fish assemblages in the western central pacific," *Scientific Data*, vol. 4, no. 1, pp. 1–12, 2017.

[212] R. Fotedar *et al.*, "Water quality, growth and stress responses of juvenile barramundi (*lates calcarifer* bloch), reared at four different densities in integrated recirculating aquaculture systems," *Aquaculture*, vol. 458, pp. 113–120, 2016.

[213] J. C. Marques, S. Lackner, R. Felix, and M. B. Orger, "Structure of the zebrafish locomotor repertoire revealed with unsupervised behavioral clustering," *Current Biology*, vol. 28, no. 2, pp. 181–195, 2018.

[214] P. J. Ashley, "Fish welfare: current issues in aquaculture," *Applied Animal Behaviour Science*, vol. 104, no. 3-4, pp. 199–235, 2007.

[215] D. Li, G. Wang, L. Du, Y. Zheng, and Z. Wang, "Recent advances in intelligent recognition methods for fish stress behavior," *Aquacultural Engineering*, p. 102222, 2021.

[216] G. Kawamura, T. U. Bagarinao, and L. L. Seng, "Fish behaviour and aquaculture," *Aquaculture Ecosystems: Adaptability and Sustainability*, pp. 68–106, 2015.

[217] J. Liu, F. Bienvenido, X. Yang, Z. Zhao, S. Feng, and C. Zhou, "Nonintrusive and automatic quantitative analysis methods for fish behaviour in aquaculture," *Aquaculture Research*, vol. 53, no. 8, pp. 2985–3000, 2022.

[218] X. Hu, Y. Liu, Z. Zhao, J. Liu, X. Yang, C. Sun, S. Chen, B. Li, and C. Zhou, "Real-time detection of uneaten feed pellets in underwater images for aquaculture using an improved YOLO-V4 network," *Computers and Electronics in Agriculture*, vol. 185, p. 106135, 2021.

[219] M. Sun, S. G. Hassan, and D. Li, "Models for estimating feed intake in aquaculture: A review," *Computers and Electronics in Agriculture*, vol. 127, pp. 425–438, 2016.

[220] D. Li, Z. Wang, S. Wu, Z. Miao, L. Du, and Y. Duan, "Automatic recognition methods of fish feeding behavior in aquaculture: a review," *Aquaculture*, vol. 528, p. 735508, 2020.

[221] C. Zhou, B. Zhang, K. Lin, D. Xu, C. Chen, X. Yang, and C. Sun, “Near-infrared imaging to quantify the feeding behavior of fish in aquaculture,” *Computers and Electronics in Agriculture*, vol. 135, pp. 233–241, 2017.

[222] S. Feng, X. Yang, Y. Liu, Z. Zhao, J. Liu, Y. Yan, and C. Zhou, “Fish feeding intensity quantification using machine vision and a lightweight 3D ResNet-GloRe network,” *Aquacultural Engineering*, vol. 98, p. 102244, 2022.

[223] N. Ubina, S.-C. Cheng, C.-C. Chang, and H.-Y. Chen, “Evaluating fish feeding intensity in aquaculture with convolutional neural networks,” *Aquacultural Engineering*, vol. 94, p. 102178, 2021.

[224] C. Zhou, D. Xu, L. Chen, S. Zhang, C. Sun, X. Yang, and Y. Wang, “Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision,” *Aquaculture*, vol. 507, pp. 457–465, 2019.

[225] Y. Zhang, C. Xu, R. Du, Q. Kong, D. Li, and C. Liu, “MSIF-MobileNetV3: An improved MobileNetV3 based on multi-scale information fusion for fish feeding behavior analysis,” *Aquacultural Engineering*, vol. 102, p. 102338, 2023.

[226] L. Yang, H. Yu, Y. Cheng, S. Mei, Y. Duan, D. Li, and Y. Chen, “A dual attention network based on efficientNet-B2 for short-term fish school feeding behavior analysis in aquaculture,” *Computers and Electronics in Agriculture*, vol. 187, p. 106316, 2021.

[227] H. Maloy, A. Aamodt, and E. Misimi, “A spatio-temporal recurrent network for salmon feeding action recognition from underwater videos in aquaculture,” *Computers and Electronics in Agriculture*, vol. 167, p. 105087, 2019.

[228] J.-Y. Su, P.-H. Zhang, S.-Y. Cai, S.-C. Cheng, and C.-C. Chang, “Visual analysis of fish feeding intensity for smart feeding in aquaculture using deep learning,” in *International Workshop on Advanced Imaging Technology (IWAIT) 2020*, vol. 11515, pp. 94–99, SPIE, 2020.

[229] D. Wei, E. Bao, Y. Wen, S. Zhu, Z. Ye, and J. Zhao, “Behavioral spatial-temporal characteristics-based appetite assessment for fish school in recirculating aquaculture systems,” *Aquaculture*, vol. 545, p. 737215, 2021.

[230] W. McFarlane, K. Cubitt, H. Williams, D. Rowsell, R. Moccia, R. Gosine, and R. McKinley, “Can feeding status and stress level be assessed by analyzing patterns of muscle activity in free swimming rainbow trout (*oncorhynchus mykiss* walbaum)?,” *Aquaculture*, vol. 239, no. 1-4, pp. 467–484, 2004.

[231] K. Stierhoff, T. Targett, and P. Grecay, “Hypoxia tolerance of the mummichog: the role of access to the water surface,” *Journal of Fish Biology*, vol. 63, no. 3, pp. 580–592, 2003.

[232] J. C. Taylor and J. M. Miller, “Physiological performance of juvenile southern flounder, *paralichthys lethostigma* (jordan and gilbert, 1884), in chronic and episodic hypoxia,” *Journal of Experimental Marine Biology and Ecology*, vol. 258, no. 2, pp. 195–214, 2001.

[233] D. Israeli and E. Kimmel, “Monitoring the behavior of hypoxia-stressed *carassius auratus* using computer vision,” *Aquacultural Engineering*, vol. 15, no. 6, pp. 423–440, 1996.

[234] G. E. Nilsson, P. Rosen, and D. Johansson, “Anoxic depression of spontaneous locomotor activity in crucian carp quantified by a computerized imaging technique,” *Journal of Experimental Biology*, vol. 180, no. 1, pp. 153–162, 1993.

[235] G. Wang, A. Muhammad, C. Liu, L. Du, and D. Li, “Automatic recognition of fish behavior with a fusion of rgb and optical flow data based on deep learning,” *Animals*, vol. 11, no. 10, p. 2774, 2021.

[236] M. Frye, T. B. Egeland, J. T. Nordeide, and I. Folstad, “Cannibalism and protective behavior of eggs in Arctic charr (*Salvelinus alpinus*)” *Ecology and Evolution*, vol. 11, no. 21, pp. 14383–14391, 2021.

[237] R. Riesch, M. S. Araujo, S. Bumgarner, C. Filla, L. Pennafont, T. R. Goins, D. Lucion, A. M. Makowicz, R. A. Martin, S. Pirroni, *et al.*, “Resource competition explains rare cannibalism in the wild in livebearing fishes,” *Ecology and Evolution*, vol. 12, no. 5, p. e8872, 2022.

[238] M. L. Andersson, K. Hulthen, C. Blake, C. Bronmark, and P. A. Nilsson, “Linking behavioural type with cannibalism in Eurasian perch,” *Plos One*, vol. 16, no. 12, p. e0260938, 2021.

[239] J. Zhao, W. Bao, F. Zhang, S. Zhu, Y. Liu, H. Lu, M. Shen, and Z. Ye, “Modified motion influence map and recurrent neural network-based monitoring of the local unusual behaviors for fish school in intensive aquaculture,” *Aquaculture*, vol. 493, pp. 165–175, 2018.

[240] H. Wang, S. Zhang, S. Zhao, J. Lu, Y. Wang, D. Li, and R. Zhao, “Fast detection of cannibalism behavior of juvenile fish based on deep learning,” *Computers and Electronics in Agriculture*, vol. 198, p. 107033, 2022.

[241] B. Erisman, W. Heyman, S. Kobara, T. Ezer, S. Pittman, O. Aburto-Oropeza, and R. S. Nemeth, “Fish spawning aggregations: where well-placed management actions can yield big benefits for fisheries and conservation,” *Fish and Fisheries*, vol. 18, no. 1, pp. 128–144, 2017.

[242] Y. Sadovy and M. Domeier, “Are aggregation-fisheries sustainable? Reef fish fisheries as a case study,” *Coral Reefs*, vol. 24, pp. 254–262, 2005.

[243] E. Rastoin-Laplane, J. Goetze, E. S. Harvey, D. Acuna-Marrero, P. Ferrié, and P. Salinas-de Leon, “A diver operated stereo-video approach for characterizing reef fish spawning aggregations: The Galapagos Marine Reserve as case study,” *Estuarine, Coastal and Shelf Science*, vol. 243, p. 106629, 2020.

[244] 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, p. 101591, 2020.

[245] R. H. Piedrahita, “Reducing the potential environmental impact of tank aquaculture effluents through intensification and recirculation,” *Aquaculture*, vol. 226, no. 1-4, pp. 35–44, 2003.

[246] M. Verdegem, R. Bosma, and J. Verreth, “Reducing water use for animal production through aquaculture,” *Water Resources Development*, vol. 22, no. 1, pp. 101–113, 2006.

[247] F. A. Francisco, P. Nuhrenberg, and A. Jordan, “High-resolution, non-invasive animal tracking and reconstruction of local environment in aquatic ecosystems,” *Movement Ecology*, vol. 8, pp. 1–12, 2020.

[248] X. Wang, H. Li, X. Yue, and L. Meng, “A comprehensive survey on object detection yolo,” *Proceedings <http://ceur-ws.org> ISSN*, vol. 1613, p. 0073, 2023.

[249] Z. Huang, H. Zhao, Z. Cui, L. Wang, H. Li, K. Qu, and H. Cui, “Early warning system for nocardiosis in largemouth bass (*Micropterus salmoides*) based on multimodal information fusion,” *Computers and Electronics in Agriculture*, vol. 226, p. 109393, 2024.

[250] Z. Haixiang, C. Hongwu, Z. HUANG, W. Lei, C. Zhengguo, Q. Keming, *et al.*, “Application of a bytetrack-based, multi-target tracking algorithm for zebrafish toxicity-response behavior recognition,” *Progress in Fishery Sciences*, vol. 45, no. 2, pp. 136–149, 2024.

[251] P. Xiao, M. Chen, G. Feng, W. Zhai, Y. Zhao, and Y. Huang, “Yolo-AWD+ CBT: An efficient algorithm for *Micropterus salmoides* swimming behavior analysis based on multi-object tracking,” *Aquaculture*, p. 741554, 2024.

[252] S. Shreesha, M. M. Pai, U. Verma, and R. M. Pai, “Fish tracking and continual behavioral pattern clustering using novel *sillago sihama* vid (SSVid),” *IEEE Access*, vol. 11, pp. 29400–29416, 2023.

[253] C. Liu, Z. Wang, Y. Li, Z. Zhang, J. Li, C. Xu, R. Du, D. Li, and Q. Duan, “Research progress of computer vision technology in abnormal fish detection,” *Aquacultural Engineering*, p. 102350, 2023.

[254] L. Chapuis, B. Williams, T. A. Gordon, and S. D. Simpson, “Low-cost action cameras offer potential for widespread acoustic monitoring of marine ecosystems,” *Ecological Indicators*, vol. 129, p. 107957, 2021.

[255] M. J. Parsons, T.-H. Lin, T. A. Mooney, C. Erbe, F. Juanes, M. Lambers, S. Li, S. Linke, A. Looby, S. L. Nedelec, *et al.*, “Sounding the call for a global library of underwater biological sounds,” *Frontiers in Ecology and Evolution*, p. 39, 2022.

[256] R. FROESE, “Fishbase. world wide web electronic publication,” <http://www.fishbase.org>, 2009.

[257] A. N. Rice, S. C. Farina, A. J. Makowski, I. M. Kaatz, P. S. Lobel, W. E. Bemis, and A. H. Bass, “Evolutionary patterns in sound production across fishes,” *Ichthyology and Herpetology*, vol. 110, no. 1, pp. 1–12, 2022.

[258] J. Lagardere and R. Mallekh, “Feeding sounds of turbot (*scophthalmus maximus*) and their potential use in the control of food supply in aquaculture: I. spectrum analysis of the feeding sounds,” *Aquaculture*, vol. 189, no. 3-4, pp. 251–258, 2000.

[259] M. Phillips, “The feeding sounds of rainbow trout, *salmo gairdneri* richardson,” *Journal of Fish Biology*, vol. 35, no. 4, pp. 589–592, 1989.

[260] Y. Yamaguchi, “Spectrum analysis of sounds made by feeding fish in relation to their movement,” *Bull. Fac. Fish., Mie Univ.*, vol. 2, pp. 39–42, 1975.

[261] E. Shishkova, “Notes and investigations on sound produced by fishes,” *Tr. Vses. Inst. Ribn. Hozaist. Okeanograf.*, vol. 280, p. 294, 1958.

[262] A. Takemura, “The attraction effect of natural feeding sound in fish,” *Bull. Fac. Fish. Nagasaki Univ.*, vol. 63, pp. 1–4, 1988.

[263] M. Cui, X. Liu, H. Liu, Z. Du, T. Chen, G. Lian, D. Li, and W. Wang, “Multimodal fish feeding intensity assessment in aquaculture,” *arXiv preprint arXiv:2309.05058*, 2023.

[264] Z. Du, M. Cui, Q. Wang, X. Liu, X. Xu, Z. Bai, C. Sun, B. Wang, S. Wang, and D. Li, "Feeding intensity assessment of aquaculture fish using mel spectrogram and deep learning algorithms," *Aquacultural Engineering*, p. 102345, 2023.

[265] Y. Zeng, X. Yang, L. Pan, W. Zhu, D. Wang, Z. Zhao, J. Liu, C. Sun, and C. Zhou, "Fish school feeding behavior quantification using acoustic signal and improved swin transformer," *Computers and Electronics in Agriculture*, vol. 204, p. 107580, 2023.

[266] R. Gao, R. Feris, and K. Grauman, "Learning to separate object sounds by watching unlabeled video," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 35–53, 2018.

[267] R. Gao, T. H. Oh, K. Grauman, and L. Torresani, "Listen to look: Action recognition by previewing audio," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10457–10467, 2020.

[268] K. Choi, M. Kersner, J. Morton, and B. Chang, "Temporal knowledge distillation for on-device audio classification," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 486–490, IEEE, 2022.

[269] U. of Rhode Island, "Discovery of sound in the sea website." <https://dosits.org/>, 2012. Accessed: May 14, 2024.

[270] J. A. Martos-Sitcha, J. Sosa, D. Ramos-Valido, F. J. Bravo, C. Carmona-Duarte, H. L. Gomes, J. A. Calduth-Giner, E. Cabruja, A. Vega, M. A. Ferrer, *et al.*, "Ultra-low power sensor devices for monitoring physical activity and respiratory frequency in farmed fish," *Frontiers in Physiology*, vol. 10, p. 667, 2019.

[271] E. Rosell-Moll, M. Piazzon, J. Sosa, M. A. Ferrer, E. Cabruja, A. Vega, J. A. Calduth-Giner, A. Sitja-Bobadilla, M. Lozano, J. A. Montiel-Nelson, *et al.*, "Use of accelerometer technology for individual tracking of activity patterns, metabolic rates and welfare in farmed gilthead sea bream (*sparus aurata*) facing a wide range of stressors," *Aquaculture*, vol. 539, p. 736609, 2021.

[272] J. Horie, T. Sasakura, Y. Ina, Y. Mashino, H. Mitamura, K. Moriya, T. Noda, and N. Arai, "Development of a pinger for classification of feeding behavior of fish based on axis-free acceleration data," in *2016 Techno-Ocean (Techno-Ocean)*, pp. 268–271, IEEE, 2016.

[273] H. Tanoue, T. Komatsu, T. Tsujino, I. Suzuki, M. Watanabe, H. Goto, and N. Miyazaki, "Feeding events of Japanese lates lates japonicus detected by a high-speed video camera and three-axis micro-acceleration data-logger," *Fisheries Science*, vol. 78, no. 3, pp. 533–538, 2012.

[274] F. Broell, T. Noda, S. Wright, P. Domenici, J. F. Steffensen, J.-P. Auclair, and C. T. Taggart, "Accelerometer tags: detecting and identifying activities in fish and the effect of sampling frequency," *Journal of Experimental Biology*, vol. 216, no. 7, pp. 1255–1264, 2013.

[275] Y. Kawabata, T. Noda, Y. Nakashima, A. Nanami, T. Sato, T. Takebe, H. Mitamura, N. Arai, T. Yamaguchi, and K. Soyano, "Use of a gyroscope/accelerometer data logger to identify alternative feeding behaviours in fish," *Journal of Experimental Biology*, vol. 217, no. 18, pp. 3204–3208, 2014.

[276] K. Birnie-Gauvin, H. Flavio, M. L. Kristensen, S. Walton-Rabideau, S. J. Cooke, W. G. Willmore, A. Koed, and K. Aarestrup, "Cortisol predicts migration timing and success in both Atlantic salmon and sea trout kelts," *Scientific Reports*, vol. 9, no. 1, p. 2422, 2019.

[277] D. An, J. Hao, Y. Wei, Y. Wang, and X. Yu, "Application of computer vision in fish intelligent feeding system—a review," *Aquaculture Research*, vol. 52, no. 2, pp. 423–437, 2021.

[278] J. J. Luczkovich, D. A. Mann, and R. A. Rountree, "Passive acoustics as a tool in fisheries science," *Transactions of the American Fisheries Society*, vol. 137, no. 2, pp. 533–541, 2008.

[279] B. K. Behera, "Recent developments in biosensor technology for fisheries and aquaculture," *Biotechnological Tools in Fisheries and Aquatic Health Management*, pp. 1–14, 2023.

[280] S. Pangorn, "Issues and challenges in sustainable development of fisheries and aquaculture of the southeast asian region: Fishery management: Application of geographic information system (gis) and remote sensing (rs) for fisheries management," *The Southeast Asian State of Fisheries and Aquaculture 2022*, pp. 155–158, 2022.

[281] H. Eguiraun and I. Martinez, "Entropy and fractal techniques for monitoring fish behaviour and welfare in aquacultural precision fish farming—a review," *Entropy*, vol. 25, no. 4, p. 559, 2023.

[282] M. S. Roth, F. Wagner, T. Roessger, I. Kopecki, D. Powalla, and J. Stamm, "An experimental approach for the quantitative assessment of downstream swimming fish behavior," *River Research and Applications*, 2024.

[283] S. D. Kuiper, N. C. Coops, S. G. Hinch, and J. C. White, "Advances in remote sensing of freshwater fish habitat: A systematic review to identify current approaches, strengths and challenges," *Fish and Fisheries*, vol. 24, no. 5, pp. 829–847, 2023.

[284] J. Aguzzi, D. Chatzievangelou, J. Company, L. Thomsen, S. Marini, F. Bonofoglio, F. Juanes, R. Rountree, A. Berry, R. Chumbinho, *et al.*, "The potential of video imagery from worldwide cabled observatory networks to provide information supporting fish-stock and biodiversity assessment," *ICES Journal of Marine Science*, vol. 77, no. 7–8, pp. 2396–2410, 2020.

[285] M. Fore, K. Frank, T. Norton, E. Svendsen, J. A. Alfredsen, T. Dempster, H. Eguiraun, W. Watson, A. Stahl, L. M. Sunde, *et al.*, "Precision fish farming: A new framework to improve production in aquaculture," *Biosystems Engineering*, vol. 173, pp. 176–193, 2018.

[286] C. Wang, Z. Li, T. Wang, X. Xu, X. Zhang, and D. Li, "Intelligent fish farm—the future of aquaculture," *Aquaculture International*, pp. 1–31, 2021.

[287] W. Xu, C. Liu, G. Wang, Y. Zhao, J. Yu, A. Muhammad, and D. Li, "Behavioral response of fish under ammonia nitrogen stress based on machine vision," *Engineering Applications of Artificial Intelligence*, vol. 128, p. 107442, 2024.

[288] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2014.

[289] H. Fang, P. Xiong, L. Xu, and Y. Chen, "Clip2video: Mastering video-text retrieval via image clip," *arXiv preprint arXiv:2106.11097*, 2021.

[290] A. Guzhov, F. Ruae, J. Hees, and A. Dengel, "Audioclip: Extending clip to image, text and audio," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 976–980, IEEE, 2022.

[291] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, *et al.*, "Learning transferable visual models from natural language supervision," in *International Conference on Machine Learning*, pp. 8748–8763, PMLR, 2021.

[292] Z. Lu, T. Zhu, H. Zhou, L. Zhang, and C. Jia, "An image enhancement method for side-scan sonar images based on multi-stage repairing image fusion," *Electronics*, vol. 12, no. 17, p. 3553, 2023.

[293] M. Qadri, K. Zhang, A. Hinduja, M. Kaess, A. Pediredla, and C. A. Metzler, "Aoneus: A neural rendering framework for acoustic-optical sensor fusion," in *ACM SIGGRAPH 2024 Conference Papers*, pp. 1–12, 2024.

[294] Y. Li, B. Wang, J. Sun, X. Wu, and Y. Li, "RGB-Sona tracking benchmark and spatial cross-attention transformer tracker," *arXiv preprint arXiv:2406.07189*, 2024.

[295] C. Zhang, L. Liu, G. Huang, H. Wen, X. Zhou, and Y. Wang, "WebUOT-1M: Advancing deep underwater object tracking with a million-scale benchmark," *arXiv preprint arXiv:2405.19818*, 2024.

[296] D. Liang, J. Xie, Z. Zou, X. Ye, W. Xu, and X. Bai, "Crowdclip: Unsupervised crowd counting via vision-language model," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2893–2903, 2023.

[297] R. Jiang, L. Liu, and C. Chen, "Clip-count: Towards text-guided zero-shot object counting," in *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 4535–4545, 2023.

[298] S. Kang, W. Moon, E. Kim, and J.-P. Heo, "Vlcounter: Text-aware visual representation for zero-shot object counting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, pp. 2714–2722, 2024.

[299] X. Geng, H. Liu, L. Lee, D. Schuurmans, S. Levine, and P. Abbeel, "M3AE: Multimodal masked autoencoders learn transferable representations," *tech. rep.*, Technical Report, 2022.

[300] Z. Du, M. Cui, X. Xu, Z. Bai, J. Han, W. Li, J. Yang, X. Liu, C. Wang, and D. Li, "Harnessing multimodal data fusion to advance accurate identification of fish feeding intensity," *Biosystems Engineering*, vol. 246, pp. 135–149, 2024.

[301] S.-W. Yang, L.-H. Shen, H.-H. Shuai, and K.-T. Feng, "CMAF: Cross-modal augmentation via fusion for underwater acoustic image recognition," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 20, no. 5, pp. 1–25, 2024.

[302] H. Chen, Z. Wang, H. Qin, and X. Mu, "UAMFDet: Acoustic-optical fusion for underwater multi-modal object detection," *Journal of Field Robotics*, 2024.

[303] W. Li, Z. Du, X. Xu, Z. Bai, J. Han, M. Cui, and D. Li, "A review of aquaculture: From single modality analysis to multimodality fusion," *Computers and Electronics in Agriculture*, vol. 226, p. 109367, 2024.

[304] L. Jin and H. Liang, "Deep learning for underwater image recognition in small sample size situations," in *OCEANS 2017-Aberdeen*, pp. 1–4, IEEE, 2017.

- [305] A. Saleh, M. Sheaves, D. Jerry, and M. Rahimi Azghadi, "How to track and segment fish without human annotations: a self-supervised deep learning approach," *Pattern Analysis and Applications*, vol. 27, no. 1, p. 4, 2024.
- [306] P. Tarling, M. Cantor, A. Clapes, and S. Escalera, "Deep learning with self-supervision and uncertainty regularization to count fish in underwater images," *Plos One*, vol. 17, no. 5, p. e0267759, 2022.
- [307] T. Zhao, Z. Shen, H. Zou, P. Zhong, and Y. Chen, "Unsupervised adversarial domain adaptation based on interpolation image for fish detection in aquaculture," *Computers and Electronics in Agriculture*, vol. 198, p. 107004, 2022.
- [308] R. B. Fisher, K.-T. Shao, and Y.-H. Chen-Burger, "Fish4-knowledge website," <https://homepages.inf.ed.ac.uk/rbf/fish4knowledge/>, 2016. Accessed: May 14, 2024.
- [309] X. Li, M. Shang, J. Hao, and Z. Yang, "Seaclef2016 website." <https://www.imageclef.org/lifeclef/2016/sea>, 2016. Accessed: May 14, 2024.
- [310] FalkSchuetzenmeister, M. M. M. Risdal, suezpollock, and W. Kan, "The nature conservancy fisheries monitoring." <https://kaggle.com/competitions/the-nature-conservancy-fisheries-monitoring>, 2016. Accessed: May 14, 2024.
- [311] S. M. Corp, "Sound metrics website." <http://www.soundmetrics.com/>, 2016. Accessed: May 14, 2024.
- [312] M. Dawkins, J. Prior, B. Lewis, R. Faillettaz, T. Banez, M. Salvi, A. Rollo, J. Simon, M. Campbell, M. Lucero, *et al.*, "Fishtrack23: An ensemble underwater dataset for multi-object tracking," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 7167–7176, 2024.
- [313] S. Lopez-Marcano, E. L Jinks, C. A. Buelow, C. J. Brown, D. Wang, B. Kusy, E. M Ditria, and R. M. Connolly, "Automatic detection of fish and tracking of movement for ecology," *Ecology and Evolution*, vol. 11, no. 12, pp. 8254–8263, 2021.
- [314] A. Saleh, I. H. Laradji, D. A. Konovalov, M. Bradley, D. Vazquez, and M. Sheaves, "A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis," *Scientific Reports*, vol. 10, no. 1, p. 14671, 2020.
- [315] M. Pedersen, D. Lehotsky, I. Nikolov, and T. B. Moeslund, "Brackishmot: The brackish multi-object tracking dataset," in *Scandinavian Conference on Image Analysis*, pp. 17–33, Springer, 2023.
- [316] J. Kay, P. Kulis, S. Stathatos, S. Deng, E. Young, S. Beery, G. Van Horn, and P. Perona, "The caltech fish counting dataset: A benchmark for multiple-object tracking and counting," in *European Conference on Computer Vision (ECCV)*, 2022.
- [317] K. Kaschner, "Fishesounds website." <https://www.fishbase.se/physiology/SoundsList.php>, 2012. Accessed: May 14, 2024.
- [318] T.-H. Lin, T. Akamatsu, F. Sinniger, and S. Harii, "Exploring coral reef biodiversity via underwater soundscapes," *Biological Conservation*, vol. 253, p. 108901, 2021.
- [319] OpenAI, "Gpt-4." <https://openai.com/product/gpt-4>, 2023. Accessed: May 14, 2024.
- [320] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, V. Chaudhary, C. Musat, E. Shatokhinina, and M. Caron, "Llama: Open and efficient foundation language models," 2023. Accessed: May 14, 2024.
- [321] S. Reed, K. Zolna, E. Parisotto, S. G. Colmenarejo, A. Novikov, G. Barth-Maron, M. Gimenez, Y. Sulsky, J. Kay, J. T. Springenberg, *et al.*, "A generalist agent," *arXiv preprint arXiv:2205.06175*, 2022.