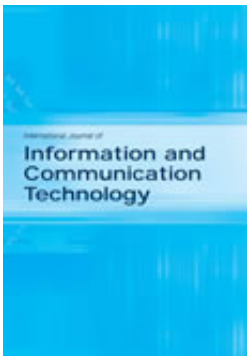


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Abnormal swimming behaviour detection and multi-object tracking based on YOLOv7 and DeepSORT

Ting Li

DOI: 10.1504/IJICT.2025.10075292

Article History:

Received:	31 July 2025
Last revised:	21 October 2025
Accepted:	30 October 2025
Published online:	15 January 2026

Abnormal swimming behaviour detection and multi-object tracking based on YOLOv7 and DeepSORT

Ting Li

Sports Department,
University of Shanghai for Science and Technology,
Shanghai, 200093, China
Email: litty136@outlook.com

Abstract: With the development of society and the improvement of people's living standards, swimming and fitness have gradually become important activities in daily life. To improve the safety management of swimming facilities, this study proposes a YOLOv7 and DeepSORT algorithm for abnormal swimming behaviour detection and multi-object tracking. This method first uses YOLOv7 for object detection, and then continuously tracks the detected targets through the DeepSORT algorithm. To optimise feature extraction for small targets, this study utilises spatial pyramid deformable convolution module and non-local attention module attention mechanism for improvement. In addition, to further improve tracking accuracy, DeepSORT introduces distance intersection and union ratio. The results showed that the improved object detection accuracy, recall and F1-value reached 94.56%, 93.89%, and 95.08%, respectively. The accuracy of multi-object tracking on the training and testing sets reached 88.56 and 90.54, with an improved accuracy value of 89.42. In addition, the detection rate of the research method exceeded 86% in crowded scenes and 91% in sparse scenes, with a minimum false alarm rate of only 1.2 times per hour. The constructed method can identify and track abnormal swimming behaviour, providing technical support for pool safety management.

Keywords: YOLOv7; DeepSORT; behaviour detection; multi-object tracking; swim.

Reference to this paper should be made as follows: Li, T. (2025) 'Abnormal swimming behaviour detection and multi-object tracking based on YOLOv7 and DeepSORT', *Int. J. Information and Communication Technology*, Vol. 26, No. 52, pp.20–40.

Biographical notes: Ting Li earned her Master's in Physical Education and Training from Shanghai Jiao Tong University in 2013. She currently serves as a Lecturer at the University of Shanghai for Science and Technology. Her research interests include information technology in sports training, school physical education, curriculum reform, and related areas. She has published over ten papers in various journals, including EI-indexed core publications. Her areas of interest also encompass motor pattern recognition and image processing.

1 Introduction

The progress of society has promoted the increasing use of public swimming pools and water parks, and the issue of swimming pool safety has received much attention. Therefore, standardised management of swimming pools and water parks, especially in preventing dangerous situations, has become increasingly important. At present, the monitoring of abnormal human behaviour mostly relies on traditional manual monitoring methods, but these methods have problems such as low efficiency and poor accuracy (Pramanik et al., 2021). Therefore, an efficient technological means is required to assist in the safety management of swimming pools. With the advancement of computer vision (CV) technology, deep learning (DL)-based multi-object tracking technology provides new possibilities for solving this problem. In complex multi-person environments, multi-object tracking technology faces challenges such as missed detections, false detections, and mismatched tracking due to the similarity in appearance features between targets and the tendency for crossing and occlusion during multi-object tracking (Preethi and Mamatha, 2023). Currently, object detection and multi-object tracking have become important research directions in CV and have made outstanding progress. In object detection, the you only look once (YOLO) series algorithms have gained widespread attention since their real-time object detection capacities. Among them, the YOLOv7 structure optimises the learning and convergence efficiency of the model while reducing computational complexity, thus demonstrating excellent performance in multiple tasks (Abba Haruna et al., 2022). Meanwhile, the improved multi-object tracking algorithm is based on the conventional simple online and real-time tracking (SORT) algorithm by introducing deep features for improvement (Lin et al., 2023; Guo et al., 2023). Therefore, based on this background, the study innovatively combines the SORT with a deep association metric (DeepSORT) algorithm with YOLOv7, aiming to enhance the intelligence level of swimming pool safety management.

2 Related works

The DeepSORT algorithm has shown excellent performance in target tracking tasks and has therefore attracted the attention of numerous researchers. Pereira et al. (2022) put forward a DeepSORT algorithm built on data association improvement to enhance the multi-object tracking performance of assisted mobile robots in navigation tasks. This algorithm solved the linear allocation problem by generating a cost matrix and evaluated target tracking using Euclidean distance and bounding box metrics. This method effectively improved the performance of robots in multi-object tracking. Tu et al. (2022) proposed an improved DeepSort method to effectively identify pig behaviour and track collective pigs in complex farm environments. This method combined two object detectors, Yolox-S and YOLOv5s, to detect and classify four pigs' behaviours, and improved the stability of pig behaviour tracking by optimising DeepSort. This method could stably track pig behaviour under commercial conditions. Jie et al. (2021) put forth an improved DeepSort ship detection and tracking approach to improve the accuracy of ship monitoring. This method optimised the initial values of anchor boxes using K-means clustering algorithm and modified the output classifier to a single-softmax-classifier. This method could effectively handle interference including occlusion and camera movement.

The YOLOv7 is crucial in target tracking tasks. Zhu et al. (2024b) developed a method built on an improved YOLOv7 detection model to effectively detect trees affected by pine wilt disease. This method utilised high-resolution helicopter images and DL models, combined with artificial intelligence attention mechanism technology to improve detection accuracy. The precision of this method was as high as 0.92, which could effectively automatically detect diseased trees in the forest. Dewi et al. (2023) designed a DL method grounded on the YOLOv7 to improve the accuracy of hand recognition in crowded environments. This method adopted CNN object recognition algorithm and conducted a concise analysis of YOLOv7 and YOLOv7x models, finding excellent performance in precision and stability of hand recognition. Wu et al. (2022) suggested a tea oil tree fruit (TOTF) detection method built on YOLOv7 and multiple data augmentation methods to improve the harvesting efficiency of TOTFs. The author established a DA-YOLOv7 model, and collected images of TOTFs to establish experimental sets. This method had good generalisation ability in complex conditions with superior detection performance.

In summary, although significant achievements have been made in research based on DeepSORT algorithm and YOLOv7, the combination of the two and their application to swimming behaviour recognition is still relatively rare. Therefore, this study combines YOLOv7 and DeepSORT algorithms for abnormal swimming behaviour detection (ASBD) and multi-object tracking, to lift the precision of swimming behaviour monitoring and enhance safety.

3 Abnormal swimming behaviour detection based on YOLOv7 and DeepSORT

This study is based on YOLOv7 and DeepSORT algorithms, combined with target detection and multi-object tracking, to achieve the localisation and behaviour tracking of swimmers in swimming scenes. YOLOv7 optimises small target feature extraction through SPD convolution module and adopts non-local attention module (NAM) to improve detection accuracy. DeepSORT improves tracking accuracy by introducing distance intersection over union (DIoU) and detects abnormal swimming behaviour by analysing unstable trajectories.

3.1 Object detection based on improved YOLOv7

In today's society, health fields such as fitness and swimming are gradually receiving more attention from people. Therefore, real-time monitoring of abnormal swimming behaviour has become one of the important means of current safety management. Object detection, as a core task in CV, has been broadly applied in various fields including security monitoring, intelligent transportation, and autonomous driving. However, these applications often require precise recognition and localisation of moving targets in dynamic environments, especially underwater environments, where there are special factors such as lighting changes and wave interference (Wang et al., 2023; Charles, 2023). Therefore, traditional target detection algorithms still face significant challenges in precisely locating underwater environments. The YOLO series algorithms have achieved significant results because of their high efficiency, real-time performance, and strong object detection capabilities (Rao and Kumar, 2025; Gong et al., 2024). YOLOv7

performs well in practical applications due to its superior detection accuracy, speed, and real-time processing capabilities (Hasanvand et al., 2023).

Based on the YOLOv7 for target detection of pool personnel, the paper introduces the NAM mechanism into its structure for improvement. Figure 1 shows the framework of YOLOv7.

Figure 1 Structure of YOLOv7 (see online version for colours)

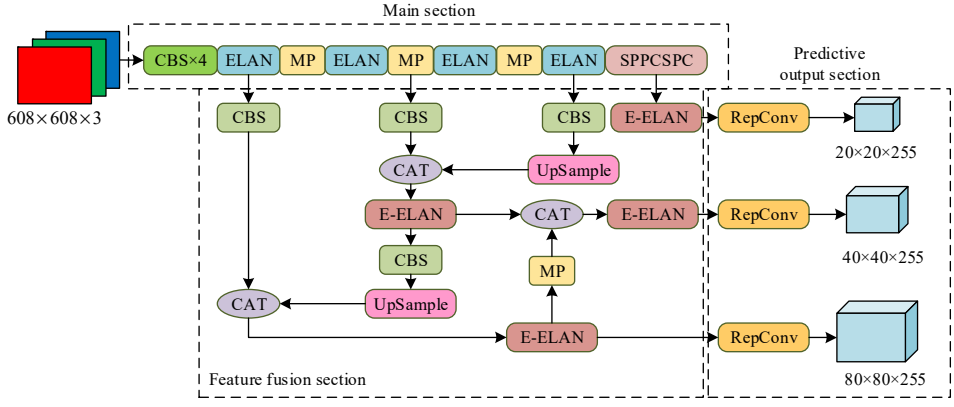
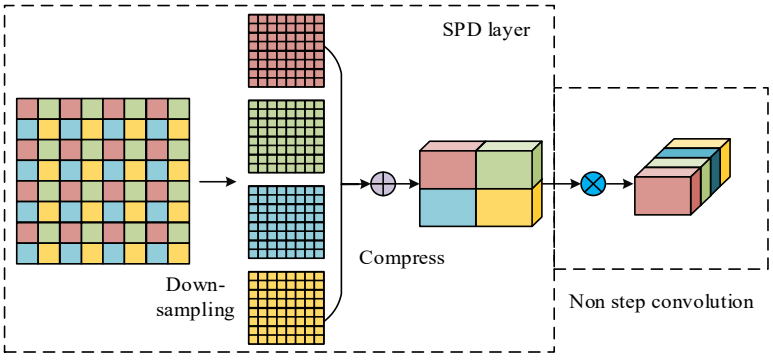


Figure 2 SPD convolutional structure (see online version for colours)



In Figure 1, the structure of YOLOv7 can be divided into the backbone, feature fusion, and prediction output. The main task of the backbone is to extract features from the input image. This section adopts an efficient layer aggregation network (ELAN) structure, which gradually extracts feature information at different levels through the processing of multiple convolutional layers. The ELAN structure can effectively improve the efficiency of feature extraction and accelerate the convergence speed of the network by optimising the gradient transfer path while reducing computational complexity. The feature fusion part utilises a multi-level feature fusion mechanism to efficiently combine features from different layers, thereby optimising adaptability to complex scenes. The task of the prediction output section is to perform final object detection and classification based on the extracted features. This part locates the target by generating bounding boxes and provides category and confidence scores for each target to complete the target detection task. However, in the pool environment, human targets are usually small and easily

mixed with the underwater background. Therefore, in response to the difficulty of extracting small target features, this study utilises the spatial pyramid deformable convolution (SPD) module to optimise the YOLOv7 structure. The SPD convolution structure is shown in Figure 2.

In Figure 2, in the ELAN structure of YOLOv7 backbone, the strip convolution and pooling layer used for downsampling and feature compression in the original ELAN structure were replaced with SPD convolution modules. The ELAN structure serves as the core feature extraction unit of YOLOv7 backbone, gradually extracting features at different levels through multiple convolutional layers; after replacement, the SPD module first compresses the spatial dimension of the feature map and remaps the multi pixel values to the depth dimension to preserve the details of small targets. Then, it adjusts the number of channels through stride free convolution, ultimately enhancing the backbone's feature capture ability for small-sized human targets in swimming pool scenes. The SPD structure mainly consists of one SPD layer and one non-step length convolutional layer. The function of the SPD layer compresses the spatial-dimension of the feature map and remaps the values of multiple pixels to the depth dimension to preserve more detailed information. Next, non-step length convolutional layers are used to further process these features, thereby reducing or increasing the number of channels. This study improves the model's perception capacity for small targets by using SPD convolution structure instead of the original stride convolution and pooling layer in YOLOv7 structure. The SPD feature transformation operation is shown in equation (1).

$$X' = \left(\frac{S}{scale}, \frac{S}{scale}, scale^2 C_1 \right) \tag{1}$$

In equation (1), X' is the data after SPD feature transformation. S is feature map's spatial-size. $Scale$ is the scaling factor. C_1 is the filter for SPD feature transformation. The non-stride convolution transformation operation is shown in equation (2).

$$X'' = \left(\frac{S}{scale}, \frac{S}{scale}, C_2 \right) \tag{2}$$

In equation (2), X'' is the data after non-stride convolution transformation. C_2 is a filter that further transforms non-step convolution (Zhu et al., 2024a). In order to further optimise the detection accuracy of YOLOv7 algorithm for small targets, the study introduces a NAM mechanism into its structure for improvement. The connection method of NAM mechanism in YOLOv7 is to concatenate NAM modules after the feature map layer formed by multi-scale feature fusion. Figure 3 shows the NAM mechanism.

In Figure 3, NAM mainly includes two parts: channel attention submodule (CAS) and spatial attention submodule (SAS) (Wu et al., 2025). The former utilises scaling factors in batch normalisation to process features in order to adaptively focus on differences between channels and accurately capture key features of each channel. The latter focuses on spatial regions in the image through normalisation operations, ensuring that the model can focus on important areas in the image, further enhancing its ability to detect small targets. The mathematical expression for batch normalisation is given by equation (3).

$$B_{out} = \gamma \left(\frac{B_{in} - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \right) + \beta \tag{3}$$

In equation (3), B_{out} and B_{in} are batch data for output and input. μ_B is the average of the input batch. σ_B^2 is the variance of the input batch. ε is a small constant. γ and β are learnable scaling and offset factors (Bui et al., 2024). The expression of the CAS is shown in equation (4).

$$M_C = \sigma\left(W_\gamma\left(BN\left(F_1\right)\right)\right) \tag{4}$$

In equation (4), M_C and F_1 are the output and input features of the CAS. σ is the sigmoid function, and W_γ is the learnable scaling convolution weight. BN is a batch normalisation operation (Shao et al., 2023). The formula for the SAS is given by equation (5).

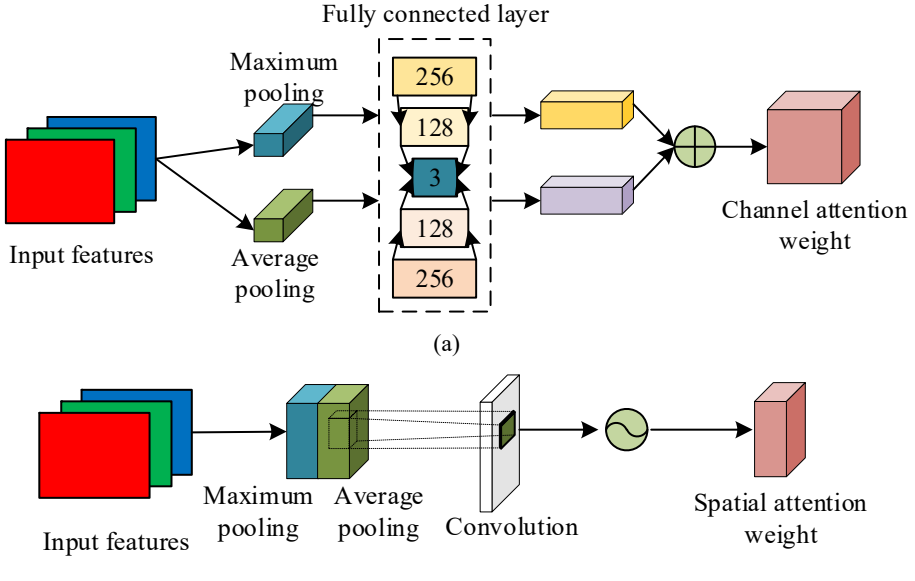
$$M_S = \sigma\left(W_\gamma\left(BN\left(F_2\right)\right)\right) \tag{5}$$

In equation (5), M_S and F_2 are the output and input feature maps of the SAS. The loss function formula of NAM is shown in equation (6).

$$L_{NAM} = \sum_{x,y} \varsigma(f(x,\omega), y) + p\sum g(\lambda) \tag{6}$$

In equation (6), L_{NAM} is the loss function of NAM. $f(x)$ denotes the output predicted by the model. y is the target label, and ω is the network weight. ς is the loss between the model output and the real label. λ is the regularisation parameter. p is the penalty value. g is the paradigm penalty function (Cheng et al., 2023; Jaihuni et al., 2023).

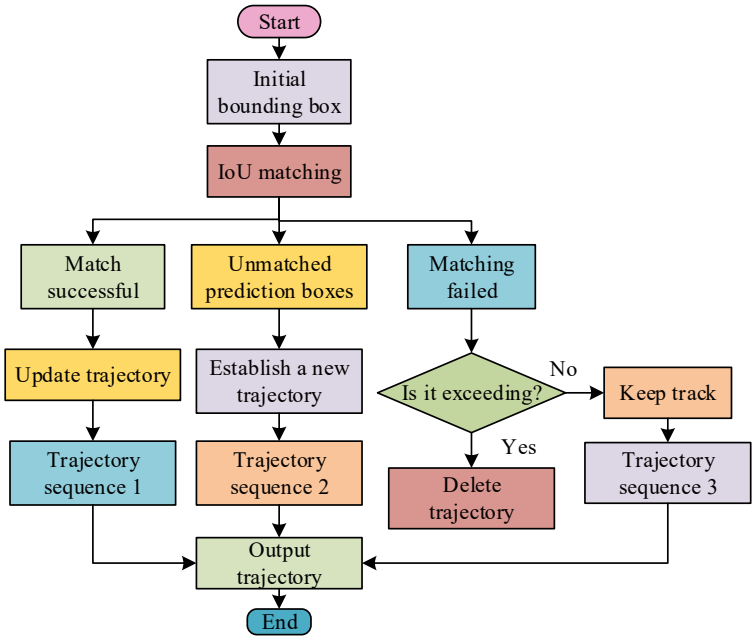
Figure 3 NAM mechanismi, (a) channel attention submodule (b) spatial attention submodule (see online version for colours)



3.2 Multi-object tracking based on improved DeepSORT

On the basis of object detection, to achieve real-time monitoring and analysis of abnormal swimming behaviour, it is also necessary to combine multi-object tracking. multi-object tracking technology refers to the method of continuously tracking a captured target by assigning a unique identifier (ID) to analyse its multi-object tracking speed, direction changes, and position movements, and predict its multi-object tracking trajectory. However, target tracking faces challenges including target occlusion and disappearance. The DeepSORT algorithm utilises Kalman filter (KF) and Hungarian algorithm to extract and match features of object detection boxes, effectively solving problems such as object occlusion, intersection, and environmental interference (Feng et al., 2024). Especially in multiplayer swimming scenarios, DeepSORT can assign independent IDs to each swimmer and update their multi-object tracking trajectories in real-time, ensuring accurate multi-object tracking. The DeepSORT multi-object tracking algorithm process is shown in Figure 4.

Figure 4 DeepSORT multi-object tracking algorithm process (see online version for colours)



In Figure 4, the improved YOLOv7 algorithm detects swimmers in each frame of the image and assigns them initial bounding boxes. The detected targets are then input into DeepSORT for matching through intersection over union (IoU). According to the matching situation, it can be divided into three types: successful matching, unmatched predicted boxes, and failed matching. If the match is successful, it will update the trajectory and form a trajectory sequence. Unmatched prediction boxes will create new trajectories. If the matching fails, it will determine the maximum lost frame. If the maximum tolerance for lost frames is exceeded, the trajectory will be deleted; otherwise, the trajectory will be retained and updated. This process ensures stable recognition of each swimmer's movement trajectory, thereby avoiding tracking errors caused by

occlusion or overlap. In the process of target localisation and tracking based on YOLOv7 and DeepSORT, it is possible to timely detect abnormal behaviours of swimmers such as stagnation, deviation from the lane, or rapid swimming. The KF formula is shown in equation (7).

$$x_t = F \cdot x_{t-1} + B \cdot u_t \tag{7}$$

In equation (7), x_t is the state of the target at time t. F and B are the state transition matrix and control input matrix. u_t is the control input at t (Sheng et al., 2024). The measuring residuals is shown in equation (8).

$$y_t = z_t - H_t \cdot x_{t|t-1} \tag{8}$$

In equation (8), y_t represents the measurement residual at time t, which is the actual observed value at that time. z_t means the predicted observation value at t. H_t denotes the observation matrix at t. $x_{t|t-1}$ is the current state predicted built on the previous time. The K_t formula for Kalman gain at t is shown in equation (9).

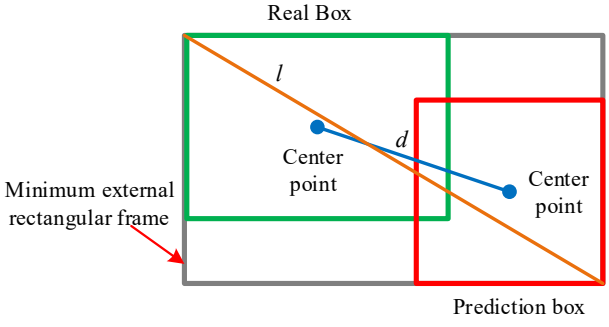
$$K_t = P_{t|t-1} \cdot H_t^T \cdot (S_t)^{-1} \tag{9}$$

In equation (9), $P_{t|t-1}$ means the covariance matrix of the prediction error, which represents the credibility of the state estimation during the model prediction at time $t - 1$. T is the transpose, and S_t is the innovation covariance at t. The formula for updating the status is given by equation (10).

$$x_{t|t} = x_{t|t-1} + K_t \cdot y_t \tag{10}$$

In equation (10), $x_{t|t}$ and $x_{t|t-1}$ are the updated state estimates and predicted state estimates at t. By updating the state estimation, the Kalman gain corrects the predicted state to be closer to the actual observed value, thereby improving the estimation's accuracy. DeepSORT uses IoU in the target matching stage to evaluate the degree of overlap between predicted and detected boxes, but IoU only considers the overlap between boxes and ignores the distance between their centre points (Liu et al., 2024). Therefore, when two boxes do not overlap, it may lead to tracking number errors (Djarah et al., 2024). Based on this limitation, this study utilises DIoU to improve the DeepSORT algorithm. Figure 5 shows the diagram of DIoU.

Figure 5 DIoU schematic diagram (see online version for colours)



In Figure 5, the green and red boxed represent the real and predicted boxed, and the grey box is the smallest bounding box that can contain both red and green. d represents the Euclidean distance between the centre points of green and red, and l represents the diagonal length of the smallest bounding rectangle containing those two boxes. The calculation for DIoU is given by equation (11).

$$DIoU = IoU - \frac{d^2}{l^2} \tag{11}$$

When the distance between boxes is far, the matching score of DIoU will decrease, so that the matching degree between targets can be more accurately determined based on the distance. The standardised result of DIoU is shown in equation (12).

$$L_{DIoU} = 1 - DIoU \tag{12}$$

In equation (12), L_{DIoU} is the standardised result of DIoU. The Hungarian matching algorithm can associate the detection target of the current frame with the tracking target of the previous frame by minimising the cost matrix, thereby establishing the trajectory of the target (Mokeddem et al., 2023). Each element in the cost matrix represents the matching cost between the two targets, with the goal of finding the optimal matching solution that minimises the total cost (Mathias et al., 2022). Minimise the total cost as shown in equation (13).

$$C_{\min} = \min \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} \cdot x_{ij} \tag{13}$$

In equation (13), $C_{\min}$ represents minimising the total cost. c_{ij} is the matching cost between tracking target i and detecting target j. x_{ij} represents the binary decision variable for whether to match the tracking target i with the detection target j. m and n are the total number of tracked and detected targets. The constraint conditions for tracking the target are shown in equation (14).

$$\sum_{i=1}^{m} x_{ij} \leq 1 \tag{14}$$

Equation (14) constrains each tracking target i to match at most one detection target. The constraint conditions for detecting the target are shown in equation (15).

$$\sum_{j=1}^{n} x_{ij} \leq 1 \tag{15}$$

Equation (15) constrains each detection target j to match at most one detection target. Research on using multi-object tracking accuracy (MOTA) to measure the comprehensive impact of various errors on overall tracking performance during the tracking process. The range of MOTA values is [–∞, 1], and the closer the value is to 1, the higher the tracking accuracy. The MOTA formula defined according to the MOTChallenge standard is shown in equation (16).

$$\text{MOTA} = 1 - \frac{FP + FN + IDsw}{GT} \tag{16}$$

In equation (16), *FP* represents the number of false positives, *FN* represents the number of missed detections, *IDsw* represents the number of ID switches, and *GT* represents the total number of real targets in the entire video sequence. Multi-object tracking precision (MOTP) is used to evaluate the position matching accuracy between the tracking box and the real target box, reflecting the positioning accuracy of the tracking box on the target. The value range is [0, 1], and the closer the value is to 1, the higher the overlap between the tracking box and the real target and the more accurate the positioning. The formula for MOTP is shown in equation (17).

$$\text{MOTP} = \frac{\sum IoU_{ij}}{Nm} \tag{17}$$

In equation (17), *Nm* represents the total number of successful matches between the tracking box and the real box in the entire video sequence, and IoU_{ij} represents the intersection and union ratio between the i^{th} tracking box and the j^{th} real box.

4 Verification of ASBD based on YOLOv7 and DeepSORT

This study first established an experimental environment and validated the model's performance built on the improved YOLOv7. Subsequently, the multi-object tracking performance based on the improved DeepSORT was validated, and the effectiveness in practical applications was finally evaluated.

4.1 Construction of experimental environment

To verify the performance of the ASBD system based on YOLOv7 and DeepSORT, this study uses Ubuntu 16.04.7 LTS operating system for construction. The hardware configuration includes Intel Core i7-9700 CPU, NVIDIA GeForce GTX 1080Ti GPU, and 32 GB of memory. In addition, the DL framework used is PyTorch 1.8.0, paired with CUDA 11.1 version, Python 3.7, the hardware storage uses a 1 TB solid-state drive model Samsung 860 EVO, the learning rate is 0.01, and the batch size is 64. The dataset used in the study was sourced from surveillance videos of three indoor public swimming pools to ensure scene diversity. The camera's field of view is mainly above the swimming pool, which accounts for 80% of the total video, and some are side views, which account for 20% of the total video, covering multi angle monitoring scenes. The video capture period includes peak pool hours from 18:00 to 21:00, 4 hours per day, and off peak hours from 9:00 to 12:00, 3 hours per day, with a total capture time of 140 hours. All videos are recorded at a uniform frame rate of 25 FPS. In the preprocessing stage, each video is segmented into static image frames at 1-frame intervals, retaining all frames to avoid losing motion information, resulting in a total of 1.26 million frames. Adjust the resolution of all frames uniformly to 1,280 × 720 pixels and use Gaussian filtering to eliminate water surface reflection noise. After filtering and preprocessing, the dataset is divided into a training set and a testing set in a 3:7 ratio, with the training set containing 378,000 frames and the testing set containing 882,000 frames. The dataset is manually annotated using the open-source annotation tool VGG image annotation tool, which supports simultaneous annotation of target bounding boxes and behaviour categories. When annotating, a rectangular bounding box is annotated for each swimmer

in the frame to determine the target position, and a unique temporary ID is assigned to ensure consistency in the annotation of the same target in consecutive frames. To ensure consistency in annotation, 10% of frames are randomly selected for trial annotation, and the annotation results are reviewed and corrected for deviations by a senior computer vision engineer. Subsequently, formal annotation will be carried out in batches, with 50,000 frames per batch. After each batch of annotation is completed, cross review will be conducted among annotation personnel. Table 1 shows the specific experimental configuration.

Table 1 Experimental settings

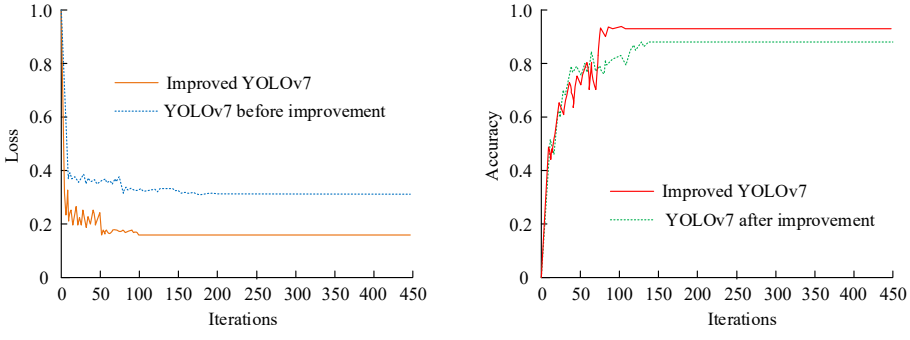
Environment	*Configuration*
Operating system	Ubuntu 16.04.7 LTS
CPU	Intel Core i7-9700
GPU	NVIDIA GeForce GTX 1080Ti
Memory	32 GB
Storage	1 TB solid state drive (SSD), model: Samsung 860 EVO
DL framework	PyTorch 1.8.0
CUDA version	11.1
Programming language	Python 3.7

4.2 Performance verification of object detection based on improved YOLOv7

To verify, the performance of the improved YOLOv7 was compared with the YOLOv7 before improvement, as displayed in Figure 6. In Figure 6(a), as the iteration approached 100 times, the improved YOLOv7 has converged, with a convergence loss value (CLV) of 0.16. The YOLOv7 before improvement only converged after nearly 180 iterations, with a CLV of 0.32. The CLV has decreased by 50% compared to before the improvement. In Figure 6(b), the object detection accuracy of YOLOv7 before and after improvement rapidly increased in the early stages of iteration, with final convergence values of 0.93 and 0.88. The accuracy has increased by 5.68% compared to before the improvement. In summary, the improved YOLOv7 exhibited significant advantages in both convergence velocity and detection accuracy, verifying the effectiveness of the improved method.

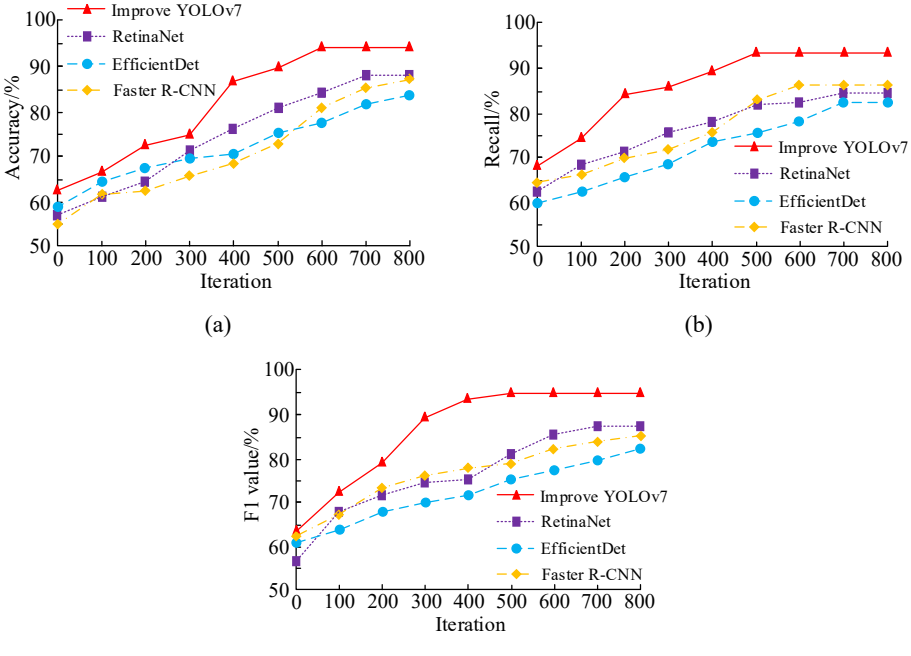
To further validate the performance of the improved YOLOv7, a comparative analysis was conducted with other advanced algorithms. Other algorithms included retina network (RetinaNet), efficient detection (EfficientDet) and faster region CNN (Faster R-CNN) (Zidani et al., 2024; Guo et al., 2024; Bhosle and Musande, 2023). Figure 7 shows a comparison of object detection performance between different algorithms. In Figure 7(a), the improved YOLOv7 had an accuracy rate of up to 94.56%, which was 6.19%, 10.47%, and 6.64% higher than the accuracy rates of RetinaNet, EfficientDet, and Faster R-CNN of 88.37%, 84.09%, and 87.92%. In Figure 7(b), the improved YOLOv7 achieved a recall rate of 93.89%, which was 8.86%, 11.42%, and 7.53% higher than the 85.03%, 82.47%, and 86.36% of the other three methods. In Figure 7(c), the F1 value of improved YOLOv7 was 95.08%, which was 7.47%, 12.52%, and 9.90% higher than the 87.61%, 82.56%, and 85.18% of others. Improving the YOLOv7 has obvious advantages in object detection performance.

Figure 6 Loss values of YOLOv7 algorithm before and after improvement, (a) comparison of convergence loss values on the training set (b) comparison of object detection accuracy on the training set (see online version for colours)



(a) (b)

Figure 7 Comparison of object detection performance of different algorithms, (a) object detection accuracy between different algorithms (b) target detection recall rates for different algorithms (c) F1 object detection with different algorithms (see online version for colours)

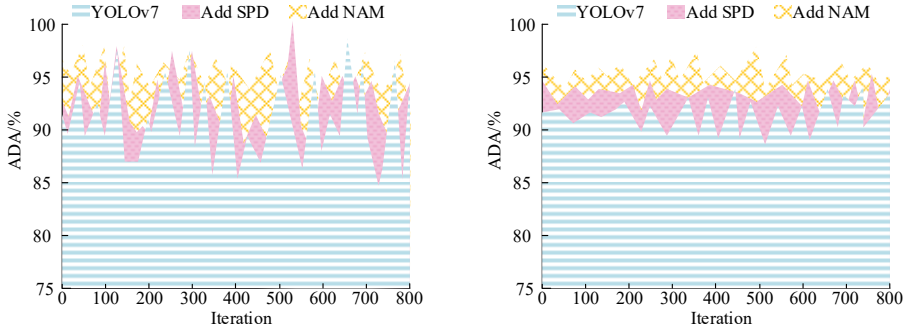


(a) (b)

(c)

To verify the role of each module in the improved YOLOv7, the ablation experiment in Figure 8 is conducted. In Figure 8(a), on the training set, the average detection accuracy (ADA) of the original YOLOv7 is 91.56%, which is improved to 93.64% after adding the SPD module, and further improves to 95.27% after adding the NAM mechanism. In Figure 8(b), on the test set, the ADA of YOLOv7 is 91.73%, which increases to 93.47% after adding SPD, and reaches 95.32% after adding NAM. The addition of each improved module enhances the object detection performance of the algorithm.

Figure 8 Ablation experiment, (a) results in the training set (b) results in the test set (see online version for colours)



4.3 *Performance verification of multi-object tracking based on improved DeepSORT*

To validate the performance of the improved DeepSORT in multi-object tracking, it is compared with other tracking algorithms, including fair multi-object tracking (Fair MOT), deep multi-object tracking (Deep MOT), and the original DeepSORT. The comparison metrics used are multi-object tracking accuracy (multi-object trackingA) and multi-object tracking precision (multi-object trackingP). Figure 9 displays the multi-object tracking performance of different algorithms. In Figure 9(a), on the training set, the multi-object trackingA of the improved DeepSORT reaches 88.56, while the multi-object trackingA of Fair MOT, Deep MOT, and the original DeepSORT are 68.51, 66.7, and 79.84. In addition, the improved multi-object trackingP of DeepSORT is 89.42, which is 11.35%, 8.78%, and 12.11% higher than the multi-object trackingP values of 79.27, 81.56, and 78.59 of other algorithms. In Figure 9(b), on the test set, the multi-object trackingA and multi-object trackingP of the improved DeepSORT algorithm are 90.54 and 91.27, while the multi-object trackingA and multi-object trackingP of other algorithms do not exceed 84. This indicates that the improved DeepSORT algorithm has excellent performance in multi-object tracking tasks.

To further validate the performance of the improved DeepSORT algorithm, the multi-object tracking evaluation metrics of different tracking algorithms are compared, as shown in Table 2. In both datasets, the improved DeepSORT outperforms the compared algorithm in all metrics. In the training set, the ID switching frequency of the improved DeepSORT is only three times, while other algorithms have more than 20 times. The tracking failure rate of improved DeepSORT is the lowest, only 1.56%, while the failure

rates of other algorithms are all higher than 4.8%. From the perspective of trajectory integrity, the improved DeepSORT achieves 92.40%, while other algorithms do not exceed 90%. The improved DeepSORT has a processing speed of 37 FPS, which still has a slight advantage. This indicates that the improved DeepSORT exhibits higher integrity in multi-object tracking tasks.

Figure 9 Multi-object tracking performance of different algorithms, (a) MOT performance on the training set (b) MOT performance on the test set (see online version for colours)

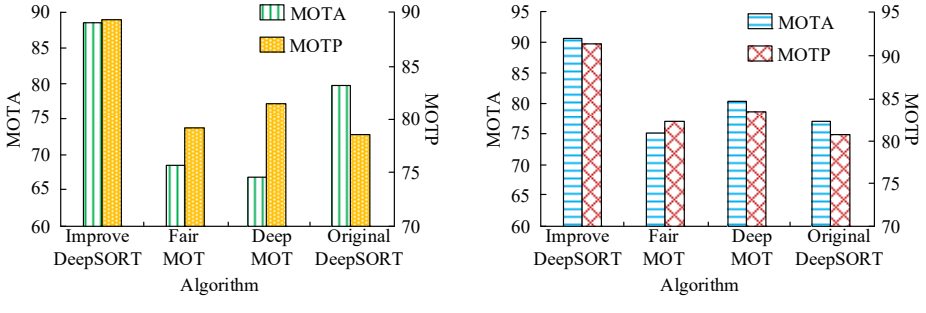
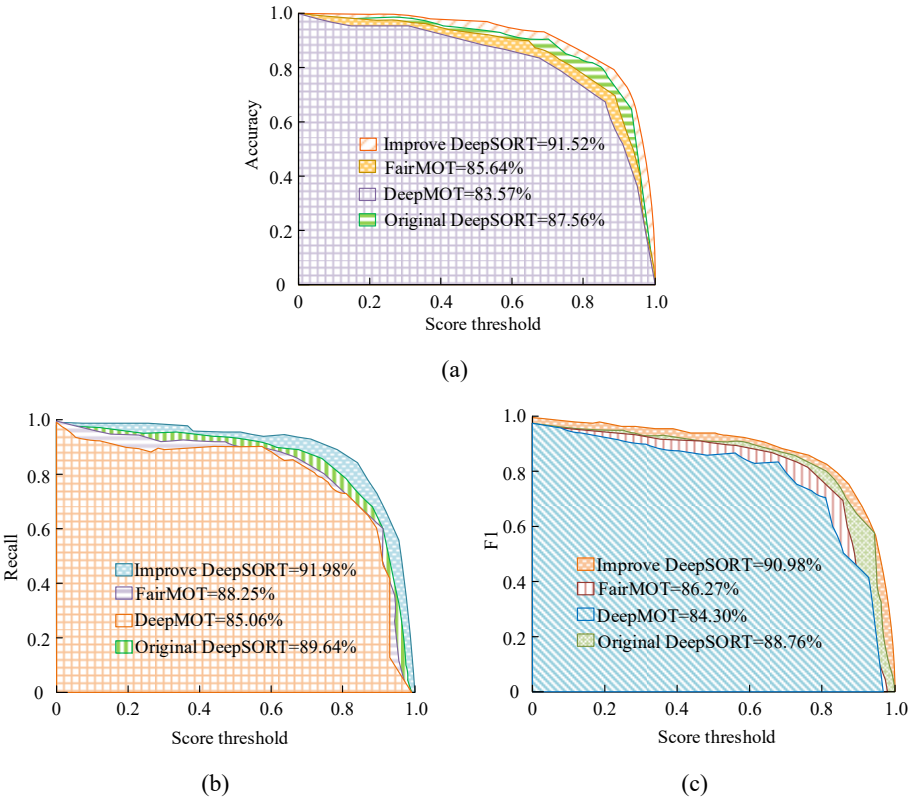


Table 2 Multi-object tracking evaluation metrics for different tracking algorithms

Dataset	*Algorithm*	*ID switches*	*Tracking failure rate (%)*	*Trajectory completeness (%)*	*Processing speed (FPS)*
Training set	Improved DeepSORT	3	1.56	92.40	37
	Fair multi-object tracking	25	5.20	88.78	40
	Deep multi-object tracking	30	6.06	87.11	45
	Original DeepSORT	20	4.83	89.36	42
Test set	Improved DeepSORT	5	1.82	91.29	35
	Fair multi-object tracking	28	5.52	87.91	39
	Deep multi-object tracking	32	6.34	86.52	38
	Original DeepSORT	22	5.09	88.64	41

To verify the detection Acc and completeness of the improved DeepSORT in target tracking, the detection performance of different algorithms is compared, as exhibited in Figure 10. In Figure 10(a), the detection Acc of the improved DeepSORT in object tracking is as high as 91.52%, which is 5.88%, 7.95%, and 3.88% higher than the 85.64%, 83.57%, and 87.64% of Fair MOT, Deep MOT, and the original DeepSORT. In Figure 10(b), the recall rate of the improved DeepSORT reaches 91.98%, which is 3.73%, 6.92%, and 2.34% higher than the compared algorithms of 88.25%, 85.06%, and 89.64%. In Figure 10(c), the improved DeepSORT's F1 value is 90.98%, which is 4.71%, 6.68%, and 2.22% higher than Fair MOT's 86.27%, Deep MOT's 84.30%, and the original DeepSORT's 88.76%. Therefore, improving the DeepSORT algorithm has shown excellent performance in target tracking tasks.

Figure 10 Detection performance of different tracking algorithms, (a) comparison of detection accuracy of different algorithms (b) comparison of detection recall of different algorithms (c) comparison of detection F1-values of different algorithms (see online version for colours)



4.4 Practical application verification

To verify the effectiveness of the ASBD system based on YOLOv7 and DeepSORT in practical applications, this study selects monitoring videos from actual swimming pools for testing. The video is shot from different angles, including top view, side view, etc., and covers different scene densities, such as sparse and crowded environments. This study sets four types of abnormal behaviours: stationary for more than 30 seconds, high-frequency swinging of the arm, staying in restricted areas, and floating on the back without moving. In addition, to further validate the model's cross scene generalisation ability, the study added a dataset of outdoor public swimming pools with outdoor swimming pools, differentiated lighting, and water quality for comparative analysis. The verification results of actual application scenarios are shown in Table 3. The inference speed of the system in different scenarios is below 30 FPS, indicating that its performance in real-time meets practical application requirements. Mostly lost (ML) metric reflects the proportion of successful target tracking by the system. The ML value is highest in the top-down sparse scene, reaching 95.20%, and lowest in the sideways crowded scene, reaching 87.45%. The mostly tracked (MT) metric represents the

proportion of tracked targets lost by the system. The minimum MT is only 2.11%, although it exceeds 5% in crowded scenes, the mixed perspective reduces it to 3.71%. In all test scenarios, the anomaly detection rate shows a high level, exceeding 86% in crowded scenarios, 91% in sparse scenarios, and 90% in mixed view scenarios. The lowest false alarm rate is only 1.2 times/hour, and the highest in crowded scenarios is only 3.5 times/hour. From the perspective of outdoor scenes, in strong light outdoor scenes, the anomaly detection rate reaches 89.23%, with a false alarm rate of only 3.8 times per hour. In the scenario of moderate turbid water quality outdoors, the abnormal detection rate is 85.76% and the false alarm rate is 4.2 times/hour, but it still meets practical needs. Comparing the results indoors and outdoors, it can be seen that the model performance only fluctuates by 3.2–5.8%, proving that it still has strong adaptability under different lighting, water quality, and viewing angle conditions, effectively solving the problem of insufficient generalisation caused by single scene data. This indicates that the ASBD system based on YOLOv7 and DeepSORT has demonstrated good performance in practical applications.

Table 3 Actual application scenario verification results

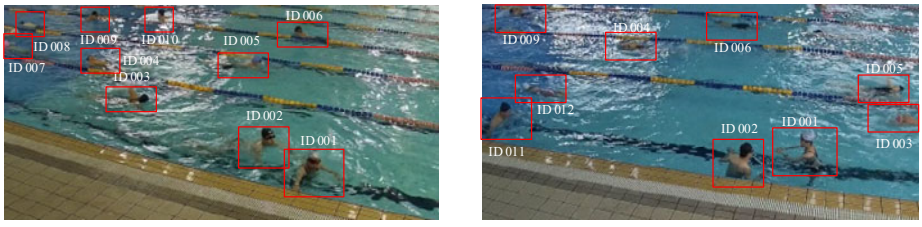
Scene		*Inference speed (FPS)*	*MT (%)*	*ML (%)*	*Abnormal detection rate (%)*	*False alarm rate (times/hour)*
Indoor	Top view – sparse	28.5	95.20	2.11	93.88	1.2
	Top view – crowded	22.3	90.71	5.32	87.54	2.8
	Side view – crowded	18.6	87.45	5.50	86.76	3.5
	Side view – sparse	24.1	88.94	2.83	91.71	3.1
	Mixed perspective	19.8	93.62	3.71	90.68	3.3
Outdoor pool	Strong light, clear water	25.6	89.57	4.12	89.23	3.8
	Overcast, slightly turbid water	22.8	88.13	4.76	87.51	4.0
	Dusk, clear water	20.3	86.92	5.01	86.84	4.1
	Noon, moderately turbid water	18.5	85.36	5.28	85.76	4.2

To more intuitively verify the effectiveness of this method in reality, the paper randomly selects the 200th and 500th frames from the surveillance video for tracking performance comparison, as shown in Figure 11. In the video frames, the swimmers are successfully labelled with IDs, and the target ID number from the previous frame 200 is maintained at frame 500, proving the continuous tracking of the target. In addition, newly appearing targets in the image have been correctly assigned new IDs. This indicates that the method can achieve stable target tracking performance and has good tracking consistency when dealing with dynamic changes.

To further verify the role of various structures in ASBD based on YOLOv7 and DeepSORT, this study conducts ablation experiments, as shown in Table 4. Among them, '√' indicates that the module exists, while '/' means that the module does not exist. When using only the YOLOv7, the detected multi-object trackingA is 74.86, while with the addition of SPD convolution and NAM mechanism, the multi-object trackingA increases to 76.14 and 77.50. After further introducing the DeepSORT algorithm, multi-object trackingA is significantly improved, reaching 82.15. Finally, after introducing DIoU for improvement, multi-object trackingA is further increased to 87.42. From the change in ID

switching frequency, only YOLOv7 has a higher ID switching frequency, reaching 52 times. With the addition of SPD and NAM, the number of ID switches gradually decreased to 48 and 23 times. After introducing DeepSORT, the number of ID switches is further reduced to 11 times, while after improvement by DIoU, it is reduced to 3 times.

Figure 11 Comparison of tracking effects of surveillance videos, (a) frame 200 (b) frame 500 (see online version for colours)



(a) (b)

In order to accurately evaluate the independent contributions of each module, the study designed only SPD and only NAM modules as control observation groups. The results showed that in the experimental group using only SPD modules, MOTA increased by 1.28 compared to the original YOLOv7. The reduction of ID switching frequency by 4 times validates the independent optimisation effect of SPD module on small target feature extraction. In the experimental group using only the NAM module, MOTA improved by 1.06 compared to the original YOLOv7, and the number of ID switches decreased by 2 times, reflecting the independent improvement effect of the NAM module on object detection accuracy in complex backgrounds. By comparing the two, it can be seen that the contribution of SPD module in small target feature optimisation is slightly higher than that of NAM module, and the performance superposition effect is more significant when the two are combined, further verifying the rationality of module combination. The gradual addition of various modules not only improves the multi-object trackingA value but also effectively reduces the number of ID switches, verifying the contribution of each module to the detection performance of abnormal swimming behaviour.

Table 4 Ablation experiment

YOLOv7	*SPD convolution*	*NAM*	*DeepSORT*	*DIoU*	*Multi-object trackingA*	*ID switches*
√	/	/	/	/	74.86	52
√	√	/	/	/	76.14	48
√	/	√	/	/	75.92	50
√	√	√	/	/	77.50	23
√	√	√	√	/	82.15	11
√	√	√	√	√	87.42	3

Note: '√' indicates use of the module, '/' indicates not used.

5 Discussion

Currently, swimming has gradually become a part of people's daily fitness activities, and the demand for swimming safety management is also increasing. To improve the safety of swimmers, this study proposed an ASBD and multi-object tracking method based on YOLOv7 and DeepSORT. In the experiment, YOLOv7 demonstrated significant advantages in object detection. By introducing SPD convolution and NAM mechanism in YOLOv7, the detection accuracy of small targets has been successfully improved. The improved YOLOv7 achieved accuracy, recall, and F1 value of 94.56%, 93.89%, and 95.08%, which were significantly better than traditional object detection algorithms. Especially in complex swimming scenes, YOLOv7 could effectively identify and accurately locate multiple swimmers, especially in crowded or multi-target environments, where its performance was particularly outstanding. In addition, after adding NAM, the detection accuracy of YOLOv7 was further improved to 95.32%. This indicated that NAM enhanced the adaptability of the model in detecting complex backgrounds and small targets by weighting global and local information of the target. Pereira R's team also designed a multi-object tracking method, but this method did not specifically focus on small targets or target overlap issues, so it is not suitable for the dynamic and target dense environment of swimming pools.

From the perspective of multi-object tracking performance, the improved DeepSORT performed excellently in both multi-object trackingA and multi-object trackingP values. The multi-object trackingA in the two sets reached 88.56 and 90.54, and the multi-object trackingP value increased to 89.42, showing significant advantages compared to other algorithms. It indicated that the improved DeepSORT could reduce ID switching and tracking failures when tracking multiple targets in complex environments, providing more stable tracking performance. Especially in crowded scenarios, the improved DeepSORT could maintain a high multi-object trackingA value, demonstrating its robustness in complex environments. Meanwhile, in the training set, the improved DeepSORT had only 3 ID switching times, a tracking failure rate as low as 1.56%, and a trajectory integrity as high as 92.40%. These results validated that the method could effectively reduce target loss and misidentification. In contrast, Tu et al. (2022) proposed an improved DeepSORT method, mainly used for identifying pig behaviour in complex farm environments and tracking collective pigs. Although this method has achieved good results in animal behaviour tracking, its main focus is on animal behaviour tracking. This study focused on abnormal detection of swimming behaviour and drowning judgement, therefore the differences in targets and scenes resulted in different tracking accuracy requirements for the two. To cope with the challenges of small targets and complex scenes in swimming pools, YOLOv7, which performed better in accuracy and recall, was chosen for this study. Especially in multi-target swimming scenes, the interaction and multi-object tracking patterns between targets were more complex, and tracking difficulty was greater.

In summary, this study is based on YOLOv7 and DeepSORT's ASBD and multi-object tracking method, which can effectively improve the safety management level of swimming pools. Especially in multi-objective and complex environments, this method demonstrates excellent detection and tracking capabilities, providing effective technical support for the safety management of water sports venues such as swimming pools.

6 Conclusions

With the rise of people's health awareness, swimming has become a widely popular form of exercise. To improve the safety of swimming management, this study designed an ASBD and multi-object tracking grounded on YOLOv7 and DeepSORT. The improved YOLOv7 had an accuracy of 94.56%, a recall of 93.89%, and an F1 value of 95.08%, significantly better than traditional algorithms. In terms of multi-object tracking, the improved DeepSORT model outperformed the comparative algorithms in multiple performance metrics. In the training set, the improved DeepSORT only had 3 ID switching times, with the lowest tracking failure rate of 1.56%, while the trajectory integrity reached 92.40%. In terms of practical application, swimmers have successfully labelled their IDs and achieved continuous tracking of their targets. In summary, the ASBD and multi-object tracking method based on YOLOv7 and DeepSORT effectively achieve continuous tracking and abnormal detection of swimming behaviour. However, drowning assessment is not limited to just a few common abnormal behaviours. The focus of this study is on object detection and tracking, while there is still room for improvement in drowning detection of abnormal swimming behaviour. Therefore, future research can further expand the recognition ability of drowning judgement and improve the overall performance of the system.

Declarations

The author declares that he has no conflicts of interest.

References

Abba Haruna, A., Muhammad, L.J. and Abubakar, M. (2022) 'Novel thermal-aware green scheduling in grid environment', *Artificial Intelligence and Applications*, November, Vol. 1, No. 4, pp.244–251, DOI: 10.47852/bonviewAIA2202332.

Bhosle, K. and Musande, V. (2023) 'Evaluation of deep learning CNN model for recognition of Devanagari digit', *Artif. Intell. Appl.*, February, Vol. 1, No. 2, pp.114–118, DOI: 10.47852/bonviewAIA3202441.

Bui, T., Wang, G., Wei, G. and Zeng, Q. (2024) 'Vehicle multi-object detection and tracking algorithm based on improved you only look once 5s version and DeepSORT', *Appl. Sci.*, March, Vol. 14, No. 7, pp.2690–2698, DOI: 10.3390/app14072690.

Charles, D. (2023) 'The lead-lag relationship between international food prices, freight rates, and Trinidad and Tobago's food inflation: a support vector regression analysis', *Green and Low-Carbon Economy*, October, Vol. 1, No. 2, pp.94–103, DOI: 10.47852/bonviewGLCE3202797.

Cheng, X., Zhang, M., Lin, S., Zhou, K., Zhao, S. and Wang, H. (2023) 'Two-stream isolation forest based on deep features for hyperspectral anomaly detection', *IEEE Geosci. Remote Sens. Lett.*, May, Vol. 20, No. 1, pp.1–5, DOI: 10.1109/LGRS.2023.3271899.

Dewi, C., A.Chen, P.S. and Christanto, H.J. (2023) 'Deep learning for highly accurate hand recognition based on yolov7 model', *Big Data Cogn. Comput.*, March, Vol. 7, No. 1, pp.53–57, DOI: 10.3390/bdcc7010053.

Djarah, D., Benmakhlouf, A., Zidani, G. and Khettache, L. (2024) 'Online multi-object tracking with YOLOv9 and DeepSORT optimized by optical flow', *Eng. Technol. Appl. Sci. Res.*, December, Vol. 14, No. 6, pp.17922–17930, DOI: 10.48084/etasr.8770.

Feng, K., Huo, W., Xu, W., Li, M. and Li, T. (2024) 'CNA-DeepSORT algorithm for multi-object tracking', *Multimedia Tools Appl.*, January, Vol. 83, No. 2, pp.4731–4755, DOI: 10.1007/s11042-023-15813-z.

Gong, D., Zhao, S., Wang, S., Li, Y., Ye, Y., Huo, L. and Bai, Z. (2024) 'On-line detection method of salted egg yolks with impurities based on improved YOLOv7 combined with DeepSORT', *Foods*, August, Vol. 13, No. 16, pp.2562–2568, DOI: 10.3390/foods13162562.

Guo, D., Shuai, H. and Zhou, F. (2024) 'Multi-target vehicle tracking algorithm based on improved DeepSORT', *Sensors*, October, Vol. 24, No. 21, pp.7014–7018, DOI: 10.3390/s24217014.

Guo, T., He, L., Luo, F., Gong, X., Li, Y. and Zhang, L. (2023) 'Anomaly detection of hyperspectral image with hierarchical antinoise mutual-incoherence-induced low-rank representation', *IEEE Trans. Geosci. Remote Sens.*, April, Vol. 61, No. 1, pp.1–13, DOI: 10.1109/TGRS.2023.3269097.

Hasanvand, M., Nooshyar, M., Moharamkhani, E. and Selyari, A. (2023) 'Machine learning methodology for identifying vehicles using image processing', *AIA*, April, Vol. 1, No. 3, pp.170–178, https://doi.org/10.47852/bonviewAIA3202833.

Jaihuni, M., Gan, H., Tabler, T., Prado, M., Qi, H. and Zhao, Y. (2023) 'Broiler mobility assessment via a semi-supervised deep learning model and neo-deep sort algorithm', *Animals*, August, Vol. 13, No. 17, pp.2719–2726, DOI: 10.3390/ani13172719.

Jie, Y., Leonidas, L., Mumtaz, F. and Ali, M. (2021) 'Ship detection and tracking in inland waterways using improved YOLOv3 and Deep SORT', *Symmetry*, February, Vol. 13, No. 2, pp.308–311, DOI: 10.3390/sym13020308.

Lin, Y., Hu, W., Zheng, Z. and Xiong, J. (2023) 'Citrus identification and counting algorithm based on improved YOLOv5s and deepsort', *Agronomy*, June, Vol. 13, No. 7, pp.1674–1679, DOI: 10.3390/agronomy13071674.

Liu, Y., An, B., Chen, S. and Zhao, D. (2024) 'Multi-target detection and tracking of shallow marine organisms based on improved YOLO v5 and DeepSORT', *IET Image Process.*, April, Vol. 18, No. 9, pp.2273–2290, DOI: 10.1049/ipr2.13090.

Mathias, A., Dhanalakshmi, S. and Kumar, R. (2022) 'Occlusion aware underwater object tracking using hybrid adaptive deep SORT-YOLOv3 approach', *Multimedia Tools Appl.*, May, Vol. 81, No. 30, pp.44109–44121, DOI: 10.1007/s11042-022-13281-5.

Mokeddem, M.L., Belahcene, M. and Bourennane, S. (2023) 'COVID-19 risk reduction based YOLOv4-P6-FaceMask detector and DeepSORT tracker', *Multimedia Tools Appl.*, June, Vol. 82, No. 15, pp.23569–23593, DOI: 10.1007/s11042-022-14251-7.

Pereira, R., Carvalho, G., Garrote, L. and Nunes, U. (2022) 'Sort and deep-SORT based multi-object tracking for mobile robotics: evaluation with new data association metrics', *Appl. Sci.*, January, Vol. 12, No. 3, pp.1319–1324, DOI: 10.3390/app12031319.

Pramanik, A., Pal, S.K., Maiti, J. and Mitra, P. (2021) 'Granulated RCNN and multi-class deep sort for multi-object detection and tracking', *IEEE Trans. Emerg. Topics Comput. Intell.*, January, Vol. 6, No. 1, pp.171–181, DOI: 10.1109/TETCI.2020.3041019.

Preethi, P. and Mamatha, H.R. (2023) 'Region-based convolutional neural network for segmenting text in epigraphical images', *Artif. Intell. Appl.*, September, Vol. 1, No. 2, pp.119–127, DOI: 10.47852/bonviewAIA2202293.

Rao, M.K. and Kumar, P.A. (2025) 'Advanced object tracking in video surveillance systems with adaptive deep SORT enhancement', *Eng. Technol. Appl. Sci. Res.*, April, Vol. 15, No. 2, pp.20871–20877, DOI: 10.48084/etasr.9529.

Shao, X., Li, X., Yang, T., Yang, Y., Liu, S. and Yuan, Z. (2023) 'Underground personnel detection and tracking based on improved YOLOv5s and DeepSORT', *Coal Sci. Technol.*, May, Vol. 51, No. 10, pp.291–301, DOI: 10.13199/j.cnki.cst.2022-1933.

Sheng, W., Shen, J., Huang, Q., Liu, Z. and Ding, Z. (2024) 'Multi-objective pedestrian tracking method based on YOLOv8 and improved DeepSORT', *Math. Biosci. Eng.*, January, Vol. 21, No. 2, pp.1791–1805, DOI: 10.3934/mbe.2024077.

Tu, S., Zeng, Q., Liang, Y., Liu, X., Huang, L., Weng, S. and Huang, Q. (2022) 'Automated behavior recognition and tracking of group-housed pigs with an improved DeepSORT method', *Agriculture*, November, Vol. 12, No. 11, pp.1907–1912, DOI: 10.3390/agriculture12111907.

Wang, Y., Liu, Y., W. Feng and Zeng, S. (2023) 'Waste haven transfer and poverty-environment trap: evidence from EU', *Green Low-Carbon Econ.*, February, Vol. 1, No. 1, pp.41–49, DOI: 10.47852/bonviewGLCE3202668.

Wu, D., Jiang, S., Zhao, E., Liu, Y., Zhu, H., Wang, W. and Wang, R. (2022) 'Detection of *Camellia oleifera* fruit in complex scenes by using YOLOv7 and data augmentation', *Appl. Sci.*, November, Vol. 12, No. 22, pp.11318–11321, DOI: 10.3390/app122211318.

Wu, S., Wang, J., Wei, W., Ji, X., Yang, B., Chen, D., Lu, H. and L. Liu. (2025) 'On the study of joint YOLOv5-DeepSort detection and tracking algorithm for *Rhynchophorus ferrugineus*', *Insects*, February, Vol. 16, No. 2, pp.219–223, DOI: 10.3390/insects16020219.

Zhu, K., Dai, J. and Gu, Z. (2024a) 'Dynamic tracking method based on improved DeepSORT for electric vehicle', *World Electr. Vehicle J.*, August, Vol. 15, No. 8, pp.374–379, DOI: 10.3390/wevj15080374.

Zhu, X., Wang, R., Shi, W., Liu, X., Ren, Y., Xu, S. and Wang, X. (2024b) 'Detection of pine-wilt-disease-affected trees based on improved YOLO v7', *Forests*, April, Vol. 15, No. 4, pp.691–695, DOI: 10.3390/f15040691.

Zidani, G., Djarah, D., Benmakhlouf, A. and Khettache, L. (2024) 'Optimizing pedestrian tracking for robust perception with YOLOv8 and DeepSort', *Appl. Comput. Sci.*, March, Vol. 20, No. 1, pp.72–84, DOI: 10.35784/acs-2024-05.