River Ship Monitoring Based on Improved Deep-Sort Algorithm

Yan Zhai
Internship and Training Center, Zhengzhou Vocational College of Finance and Taxation, Zhengzhou, 450000, China
E-mail: rocky_zhai@163.com

Keywords: ship monitoring, object detection, target tracking, simple online real-time tracking

Received: March 11, 2024

As the economy develops rapidly, waterway transportation has gradually become an important part of the logistics industry. A model was built to improve the low detection and tracking accuracy of ship objects. First, the dilated convolution was introduced into the YOLOv3. A prediction scale of 104×104 and L2 regularization were introduced to detect small objects. A target detecting model using improved YOLOv3 was constructed. Then the improved YOLOv3 was used as the detector for the deep simple online real-time tracking algorithm. The D-IoU distance was introduced into the cascaded matching loss to build a ship tracking model based on the improved tracking algorithm. These results confirmed that the improved YOLOv3 had an accuracy of 6345, a detecting time of 21.3 seconds, a recall rate of 93.25%, a missing alarm rate of 6.76%, and an average precision of 92.53%. The proposed object detection model performed the best in terms of detecting accuracy, missing and false alarm rates, and average precision indicators, with values of 87.48%, 5.14%, 12.51%, and 94.35%, respectively. The proposed ship tracking model had the highest recall rate of 64.7% and a multi-target tracking accuracy of 61.8%. This study confirms that the proposed object detection and tracking models have good performance and contribute to the intelligent development of the waterway transportation industry.

Povzetek: Model za nadzor rečnih ladij uporablja izboljšan algoritem Deep-SORT z uvedbo dilatacijske konvolucije in L2 regularizacije v YOLOv3, kar povečuje natančnost zaznavanja in sledenja ladij.

1 Introduction

With the continuous deepening of economic globalization, the shipping industry also develops rapidly, but the increase of ships also poses serious challenges to river management [1]. Ship monitoring is an important task in river management, which not only ensures the normal operation of river transportation, but also ensures the safety of navigation. Therefore, the driving behavior of the drivers can be monitored, and navigation hazards caused by non-standard driving and unauthorized departure from the post can be avoided. The object detection and tracking are important methods for ship monitoring. Object detection is a key direction in the image processing, with the task of identifying all interested targets in the image. Target tracking refers to continuously tracking the position and shape information of targets in video sequences and updating the status of targets in real-time [2, 3]. The traditional video ship target monitoring method relies on manual searching and discrimination by the human eye. The target monitoring method has low efficiency and high cost due to the limited energy of the human body. As intelligent information processing technology develops, deep learning has extensive application in object detection and tracking, improving detection accuracy while saving labor costs [4]. However, the background in river ship monitoring videos is often complex and faces the challenge of small object detection, which affects the accuracy of object detection and tracking technologies. The detecting performance of existing research algorithms is needed for improvement [5]. In this context, ship tracking models are built based on an improved YOLOv3 object detection model and an improved Deep Simple Online Real-time Tracking (Deep-Sort) algorithm. There are two main innovations in this study. Firstly, dilated convolution is introduced into the backbone network of YOLOv3, and a prediction scale of 104×104 and L2 regularization is introduced to detect small objects. Secondly, this improved YOLOv3 will be regarded as the detector of Deep-Sort, and the D-IoU distance is introduced into the loss of cascade matching. The main structure of the study includes four parts. Firstly, an analysis is conducted on the current research. Secondly, an object detection model based on improved YOLOv3 and a ship tracking model based on improved Deep-Sort are built. Then an analysis of the application effectiveness of the proposed model is conducted. Finally, there is the conclusion of the entire study.

2 Related works

Ship monitoring is an important part in ensuring the safe operation of ships. Potential safety hazards can be identified and resolved in a timely manner by monitoring

the machinery, equipment, and electrical systems of ships, ensuring the safety and reliability of ships. Tsoumpris and Theotokatos developed a method for monitoring the autonomous ships using dynamic Bayesian networks and rule-based energy management strategies. They captured performance indicators while considering the reliability of the entire system and its components. These results confirmed that the proposed means heightened the ship monitoring capability of hybrid power plants [6]. Capezza et al. stated that the rapid development of data collection technology on modern ships led to data rich. The functional regression control charts addressed the issue of whether the observed CO2 emission profile was as expected given covariate values [7]. There are problems such as low detecting accuracy, displaying delay, and computational blockage in ship detecting in surveillance video. Therefore, Zheng et al. optimized the anchor box algorithm in YOLOv5 based on the characteristics of ship targets, and t-SNE was used to reduce and visualize dataset. These results confirmed that this method improved ship detecting accuracy and speed [8]. Wang et al. proposed a model using SSD framework that detected different feature parameters in response to the feature detection-based ship recognition technology in the maritime. These results confirmed that the proposed model had good compatibility and performed well in efficiency and recognition accuracy, with certain theoretical value and application prospects [9]. Kim et al. addressed the safety and reliability issues associated with autonomous and remote control of ships. The safety challenges of automatic ship operations in a hybrid navigation environment and several methods were studied to reduce safety risks. Potential practical and research interests in ship navigation were also discussed in the future [10]. Wang et al. developed a means assisted frictional electric intelligent pad system for monitoring crew members, which not only obtained crew information but also did not need to consider privacy issues in video shooting. The comprehensive monitoring of crew and cargo was achieved, and the ability and efficiency were improved to handle emergency situations [11].

Deep-Sort is a multi-target tracking algorithm based on object detection, and the quality of the object detection algorithm will affect its tracking performance. Meemesis et al. proposed a real-time multi-target tracking framework based on the improved Deep-Sort algorithm, which was combined with the YOLO detection method to address the low accuracy in tracking multiple objects. These results confirmed that the proposed improved Deep-Sort algorithm was effective, and the multi-target tracking framework had good performance [12]. Chang et al. proposed an abnormal behavior detection model with pedestrian detection and tracking, combining YOLOv3 and Deep-Sort, to improve the behavior recognition and detection of cameras. They used a network to predict abnormal behavior. This helped to satisfy the needs of real-time monitoring systems. These results confirmed that the proposed method had good recognition accuracy [13]. Mathias et al. proposed an adaptive Deep-Sort and YOLOv3 detecting and tracking scheme to address the difficulty of tracking and recognizing underwater objects caused by light refraction. This scheme could be used for tracking and recognition of underwater objects that were occluded. These results confirmed that the proposed scheme had good application effects in occlusion object detection tasks from different perspectives [14]. Zou et al. proposed a multi-target tracking model using an improved YOLOv3 as the detector for Deep-Sort to address the tracking livestock behavior and health status in livestock farming. The backbone of YOLOv3 was replaced by MobileNetV2 to improve the detecting speed. These results confirmed that the proposed model had high detection accuracy and performance [15]. Rishika et al. addressed the low accuracy in detecting and counting intelligent vehicles in the highway management. A Deep-Sort model based on YOLO-V4 was used to detect and track vehicles in real-time from video sequences and designed a vision-based vehicle detection and counting system. These results confirmed that the proposed method had certain feasibility and effectiveness [16]. Sahoo et al. proposed an optimization model that combined a region-based Convolutional Neural Network (CNN) with a detector and Deep-Sort to predict social distance in public places, addressing the personal loyalty monitoring toward social distance norms. These results confirmed that the proposed model had good distance detection performance [17]. The summary of relevant literature is shown in Table 1.

Table 1: Summary of relevant literature

Method	Performance metrics	Key findings	Insufficient
Tsoumpris et al. [6]	Component reliability, engine speed	Proved the usefulness of expanding ship monitoring functions	Lack of practical application experiments
Capezza et al. [7]	Carbon dioxide emissions	Can be used for automatic tracking mode and trend	Do not directly allow real-time feedback control
Zheng et al. [8]	Accuracy and detection speed	Can achieve more accurate target frame positioning and improve target detection accuracy	The network structure is relatively complex

Wang et al. [9]	Calculating time, processing frame rate, and recognition accuracy	An important component of intelligent ship automatic recognition edge platform	The actual application effect has not been verified, and the stability is poor
Kim et al. [10]	Security challenges	Security challenges increase with the improvement of ship automation level	Increased complexity
Wang et al. [11]	Efficiency	Comprehensive monitoring of crew and cargo has been achieved	Increased complexity
Meimetis et al. [12]	Detection accuracy	The improved Deep SORT and YOLO detection methods have good detection performance	Increased complexity
Chang et al. [13]	Recognition rate	Can meet the needs of real-time monitoring	Increased complexity
Mathias et al. [14]	Efficiency and accuracy	Capable of underwater tracking and recognition in complex scenarios	Increased complexity
Zou et al. [15]	Average accuracy, identity switching	Implemented adaptive learning of multi-scale features of objects	Identity switching reduced
Rishika et al. [16]	Accuracy	Real time detection and tracking of vehicle video sequences have been achieved	The model calculation takes a long time
Sahoo et al. [17]	Accuracy, total loss, and training time	Effectively monitoring social distance	Increased complexity

In summary, although the ship detection technologies and Deep-Sort are studied widely, the complex background and concentrated target distribution of ship images pose significant challenges for ship detection and tracking. Therefore, the study will utilize the modified YOLOv3 and Deep-Sort algorithms to build object detection and tracking models for effective monitoring of river ships.

3 Object detection and tracking model construction for river ships

Due to the complex and diverse environment and the unique shooting perspective, traditional object detection and tracking algorithms are no longer sufficient for ship object detection in complex backgrounds. An object detection model based on improved YOLOv3 and a ship tracking model based on improved Deep-Sort are built to achieve precise detection and tracking of river ships.

3.1 Construction of an improved YOLOv3-based object detection model

The YOLO algorithms use the DarkNet model as the feature extraction network for object detection tasks, which can obtain rich features by extracting multi-level target information. The YOLOv3 network can predict multiple bounding box and category probabilities simultaneously on the entire image through a single forward operation. YOLOv3 has good detection accuracy and real-time performance, so YOLOv3 is used as the head of the object detection controller [18]. The YOLOv3 network forms a backbone network, DarkNet-53, with stronger feature extraction ability after borrowing from residual networks. DarkNet-53 alleviates the consumption of computing memory while deepening the network layers, improves the generalization ability of the networks, and accelerates the convergence speed of the training model. Multi-scale features are introduced using the Feature Pyramid Network (FPN) to detect small objects [19]. Figure 1 shows the network structure of FPN.

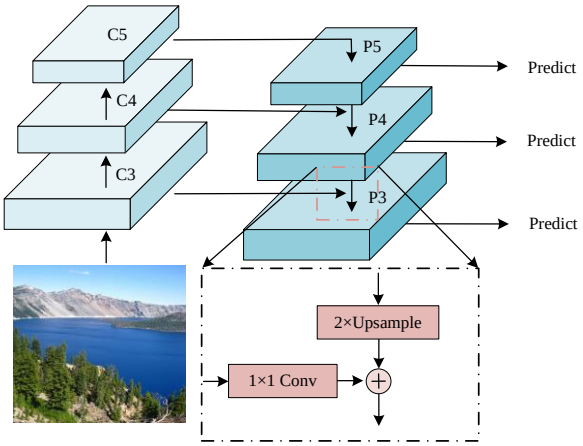


Figure1: The network structure of FPN

YOLOv3 will form a fixed number of predicted bounding boxes on the feature map and perform position regression on the generated bounding boxes. The bounding boxes prediction is represented by formula (1).

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_w = p_w e^{t_w} \\ b_h = p_h e^{t_h} \end{cases} \quad (1)$$

In formula (1), b_x, b_y, b_w, and b_h represent the border coordinate values. t_x and t_y are the positions from the center of the target to the upper left corner of the current grid. c_x and c_y refer to the quantity of grids that are unlike the midpoint of the prediction box to the up left corner. p_w and p_h mean the preset width and height of the anchor box. t_w and t_h represent the edge length of the predicted box. It is necessary to calculate each prediction box's confidence and set a threshold to avoid duplicate prediction bounding box, discarding prediction boxes with confidence levels outside the threshold. The confidence level is expressed using formula (2).

$$C_{conf} = \Pr(class_i | object) \times \Pr(object) \times IOU_{pred}^{truth} = \Pr(class) \times IOU_{pred}^{truth} \quad (2)$$

$\Pr(class_i | object)$ represents the probability of predicting C conditional categories within each grid cell in formula (2). $\Pr(object) \times IOU_{pred}^{truth}$ refers to the confidence when there is a target within the box. IOU_{pred}^{truth} means the intersection and union ratio between the predicted box area and the actual box area. A non-maximum suppression algorithm is used to determine the overlap of the remaining bounding box and to remove redundant predicted bounding box. Therefore, the predicted boxes having high reliability are retained as object detection boxes. The predicted bounding box consists of three loss functions, represented by formula (3).

$$\begin{cases} loss(object)_{YOLOv3} = loss_1 - loss_2 - loss_3 \\ loss_1 = \lambda_{coord} \sum_{i=0}^{S \times S} \sum_{j=0}^{B} I_{ij}^{obj} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\ + \lambda_{coord} \sum_{i=0}^{S \times S} \sum_{j=0}^{B} I_{ij}^{obj} (2 - w_i \times h_i) \left[(w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2 \right] \\ loss_2 = \lambda_{obj} \sum_{i=0}^{S \times S} \sum_{j=0}^{B} I_{ij}^{obj} \left[\hat{C}_i \log(C_i) + (1 - \hat{C}_i) \log(1 - \hat{C}_i) \right] \\ + \lambda_{noobj} \sum_{i=0}^{S \times S} \sum_{j=0}^{B} I_{ij}^{noobj} \left[\hat{C}_i \log(C_i) + (1 - \hat{C}_i) \log(1 - \hat{C}_i) \right] \\ loss_3 = \sum_{i=0}^{S \times S} \sum_{j=0}^{B} I_{ij}^{obj} \sum_{c \in classes} \left[\hat{p}_i(c) \log(p_i(c)) + (1 - \hat{p}_i(c)) \log(1 - p_i(c)) \right] \end{cases} \quad (3)$$

In formula (3), $loss_1$ represents the loss of the predicted bounding box. $loss_2$ means the loss of predictive confidence. $loss_3$ refers to the loss of predicted categories. λ is the weight lost. $\wedge$ refers to true real value. x_i, y_i, w_i, and h_i represent the i th bounding box's four coordinates. C_i means the confidence level of the i th bounding box. $p_i(c)$ is the i th bounding box's class probability. However, YOLOv3 has poor sensitivity to small targets. Therefore, a dilated convolution is added to DarkNet-53 to expand the receptive field of the image. This modified network is called DC-DarkNet-53 [20]. CNN can perform convolution operations on images and automatically extract feature information of targets, providing rich detailed features for subsequent object detection and tracking. As the operation process repeats, the feature resolution of the input target will continuously decrease, and the information channels will increase accordingly. Dilated convolution has emerged to ensure that the output feature map contains more detailed target information [21]. Hollow convolution can increase receptive fields or domains by inserting voids into the standard convolution kernel. Hollow convolution can improve the model's performance, especially in tasks that handle large-sized inputs or require consideration of remote pixel relationships. Figure 2 shows the specific structure.

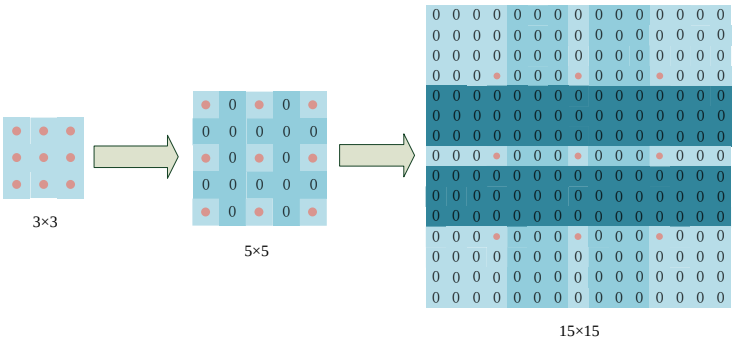


Figure 2: Structure diagram of dilated convolution

In addition, a prediction scale of 104×104 was introduced to address the poor real-time detection of YOLOv3 for small and medium objects. These mathematical model parameters fitted after network training are generally small. As the training samples increase, the previously reasonable sample distribution may be disrupted. L2 regularization can enhance the network's anti-interference ability by using smaller parameter weights, as expressed by formula (4).

$$J = J_0 + \lambda \sum_{\omega} \omega^2 \tag{4}$$

In formula (4), J_0 refers to an original loss function. λ refers to a regularization coefficient. When θ is solved, the loss function in linear regression is represented by formula (5).

$$\begin{cases} J(\theta) = \dfrac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) \\ h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \ldots + \theta_n x_n \end{cases} \tag{5}$$

Formula (6) can be obtained by using a gradient descending means to make the whole loss function reduced.

$$\begin{cases} \dfrac{\partial}{\partial \theta_j} J(\theta) = \dfrac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) x_j^{(i)} \\ \theta_j := \theta_j - a \dfrac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) x_j^{(i)} \\ \theta_j : L2 = \theta_j (1 - a \dfrac{\lambda}{m}) - a \dfrac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) x_j^{(i)} \\ Loss_{L2} = Loss + \dfrac{\lambda}{2N} \sum_{\omega} \omega^2 \end{cases} \tag{6}$$

In formula (6), $\theta_j :$ represents the original loss function θ_j 's iterative equation. $\theta_j : L2$ refers to an iterative equation after L2 regularization. $(1 - a\frac{\lambda}{m})$ is a penalty. $Loss_{L2}$ means an improved losing function. $\frac{\lambda}{2N} \sum_{\omega} \omega^2$ represents the L2 regularization term. Figure 3 shows the improved YOLOv3.

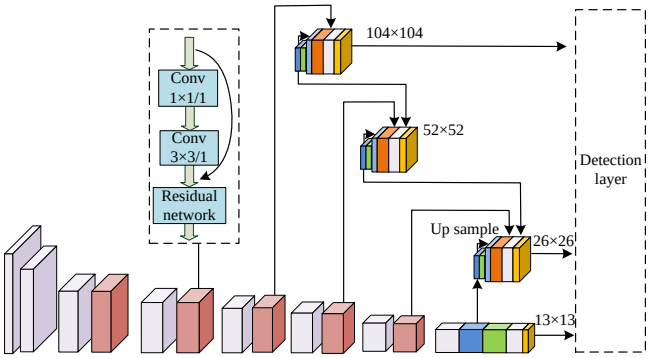


Figure 3: Improved YOLOv3 model

3.2 Construction of a ship tracking model based on improved Deep-Sort

Further target tracking can be carried out after implementing the object detection of the ship. Target tracking refers to continuously tracking the target in subsequent frames after the target is specified in the first frame of the video sequence. That is, boundary boxes are used to calibrate the target and achieve target localization and scale estimation. Deep-Sort can track ship targets.

The detection part of Deep-Sort utilizes the Faster R-CNN algorithm, which belongs to two-stage object detection method. Although the detection accuracy of Deep-Sort is high, its speed is slow. Therefore, the study uses the designed improved YOLOv3 detecting algorithm as the detector to modify Deep-Sort. The basic principle of Simple Online Real-time Tracking (SORT) is based on the object detection algorithm, using Kalman filtering for prediction and matching using Hungarian. Deep-Sort is a modified SORT that incorporates appearance information on top of the SORT algorithm. Figure 4 shows the ship target tracking based on Deep-Sort.

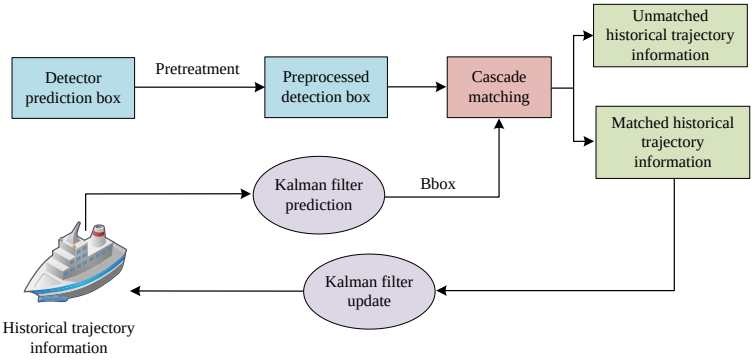


Figure 4: Ship target tracking flowchart of Deep-Sort

In target tracking, cascade matching is a crucial step. Deep-Sort considers the correlation between target feature information and motion information to achieve the pairing of preprocessed detection boxes and bbox. Deep-Sort constructs Mahalanobis distance and cosine distance to represent the matching cost between the preprocessed detection box and bbox. The calculation of Mahalanobis distance correlation is represented by formula (7).

$$d^1(i,j)=(d_j-b_i)^T S_i^{-1}(d_j-b_i) \qquad (7)$$

In formula (7), d_j refers to coordinate vector of the j th preprocessed detecting box. b_i refers to the target position predicted by the i th tracker. S_i means the covariance matrix of d_j and b_i. However, the matching degree of the Mahalanobis distance metric is not precise enough, which can easily lead to ID jumps. Therefore, Deep-Sort also introduces the cosine distance of feature vectors within the matching box. Meanwhile, CNN is used to extract target features within the box. At the same time, a 128-dimensional vector is output to represent the features of the target within the box. The minimum cosine distance between the last 100 successfully associated feature sets R_i of the i th tracker and the feature vectors of the current j th detection result is represented by formula (8).

$$d^2(i,j)-\min\left\{1-r_j^T r_k^i \middle| r_k^i \in R\right\} \qquad (8)$$

In formula (8), r_j represents the feature vector corresponding to the j th detecting box input by the current tracker. r_k^i refers to the k th feature vector in the feature set corresponding to the i th tracker. Formula (9) can be obtained by weighted averaging the Mahalanobis distance and cosine distance.

$$c_{i,j}=\rho d^1(i,j)+(1-\rho)d^2(i,j) \qquad (9)$$

In formula (9), ρ represents the weight coefficient. Finally, the linear weighting of the two distances is used as a measure of the matching loss between the two boxes. Hungarian is utilized to match the detection box and trajectory prediction box output [22]. In the object detection, the distance between the two boxes is usually constructed by combining the center coordinates of the predicted box with width and height as a whole. The distance loss function is used to calculate the predicted box and annotated box in reference ship detection to improve Deep-Sort. D-IoU distance is introduced in the loss of cascading matching. Figure 5 is a schematic diagram of IoU calculation and D-IoU distance.

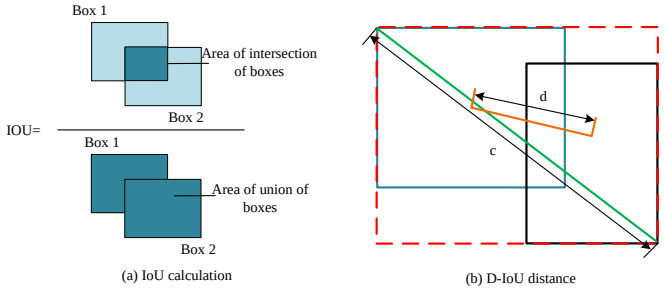


Figure 5: Schematic diagram of IoU calculation and D-IoU distance

D-IoU uses the intersection union ratio of two rectangular boxes to represent the overlap of the two boxes, expressed by formula (10).

$$\begin{cases} DIoU = IoU - \rho^2(b_1, b_2)/c^2 \\ IoU = \left|B_1 \cap B_2\right| / \left|B_1 \cup B_2\right| \end{cases} \tag{10}$$

In formula (10), b_1 and b_2 represent the center coordinates of the B_1 and B_2 boxes, respectively. ρ^2 refers to the Euclidean distance. c means the diagonal length of the minimum bounding rectangle between B_1 and B_2. These matched D-IoU loss and final weighted loss are represented by formula (11).

$$\begin{cases} d^3(i,j) = 1 - DIoU(d_j, b_i) \\ c_{i,j} = \rho_1 d^1(i,j) + \rho_2 d^2(i,j) + \rho_3 d^3(i,j) \end{cases} \tag{11}$$

In formula (11), d_j represents the predicted box of the j th preprocessed detection box. b_i refers to the prediction box of the i th tracker for the target. ρ is the weight, and $\rho_1 + \rho_2 + \rho_3 = 1$. The ship object detection and tracker can obtain the complete trajectory of ship movement in the video. Therefore, the direction of ship movement can be determined, and the ship flow in the river channel can be calculated during a fixed time period.

4 Effectiveness analysis of ship object detection and tracking models

An object detection model based on improved YOLOv3 and a ship tracking model based on improved Deep-Sort were built to effectively monitor ships in river channels. However, their practical application effects still needed further verification. The research mainly analyzed from two aspects. Firstly, the detecting performance of an object detection model based on the improved YOLOv3 was analyzed. Then the effectiveness of the ship tracking model based on the improved Deep-Sort was verified.

4.1 Effectiveness analysis of object detection models

A MyShip dataset containing 68054 ship targets was used to verify the improvement effect of DC-DarkNet-53 on YOLOv3. The weight attenuation value was 0.0001, the initial learning rate was 0.001, and the momentum was 0.9. The Faster R-CNN with ResNet-101 and VGG-16 backbone networks and the traditional DarkNet-53 model were compared. The comparison of the correct detection and the detection time is presented in Figure 6. Among the four models, DC-DarkNet-53 had the highest positive detection, with 6345, followed by DarkNet-53 with 62759 correct detection, and VGG-16 performed the worst with 38493 correct detections. In addition, the detection time of the proposed improvement YOLOv3 was 21.3 seconds, which was slightly higher than the 20.6 seconds of DarkNet-53, but still within an acceptable range. These results confirmed that dilated convolution improved the YOLOv3 positively, and the improved network had high detection accuracy and efficiency.

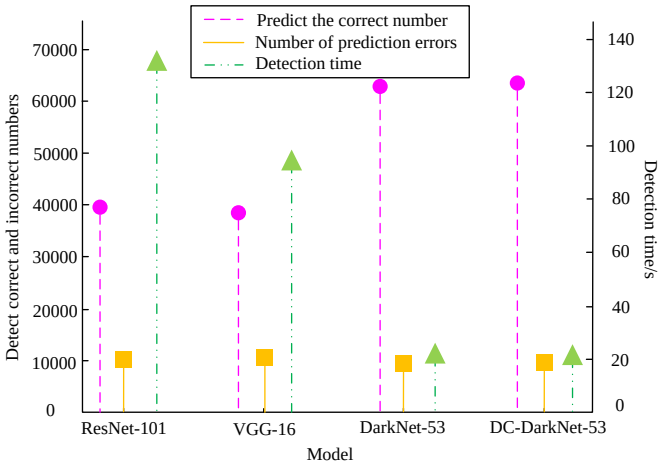


Figure 6: Comparison results of detection accuracy and detection time for four models

The recall, missing alarm rate, and Average Precision (AP) of the four models were compared to verify the detection performance of the proposed DC-DarkNet-53. From Figure 7 (a), among the four models, the recall rate of this study model was the highest at 93.25%, followed by DarkNet-53 at 92.21%, and VGG-16 was the worst at 56.55%. From Figure 7 (b), the missing alarm rate of the study model was the lowest, at 6.76%, which was 1.03% lower than DarkNet-53. From Figure 7 (c), the AP index of the research model was the highest, at 92.53%, which was 1.32% higher than DarkNet-53. These results confirmed that the proposed DC-DarkNet-53 had good detection performance, feasibility, and effectiveness.

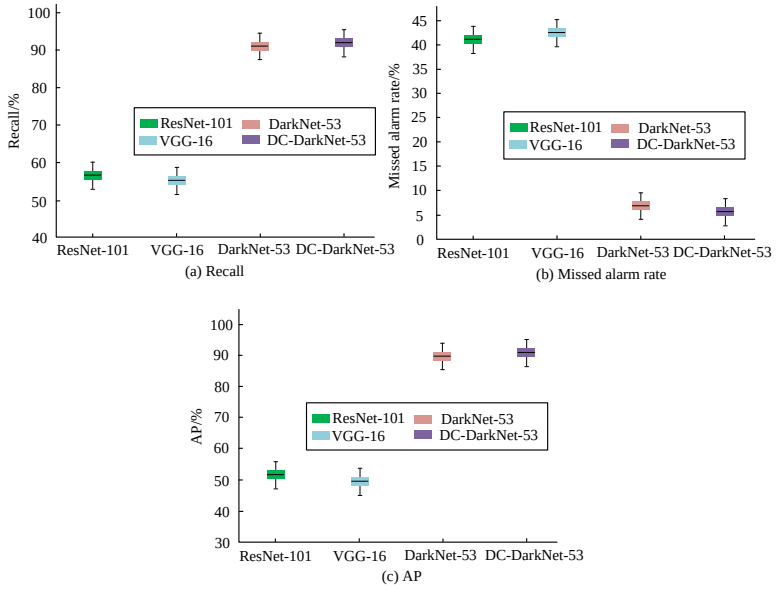


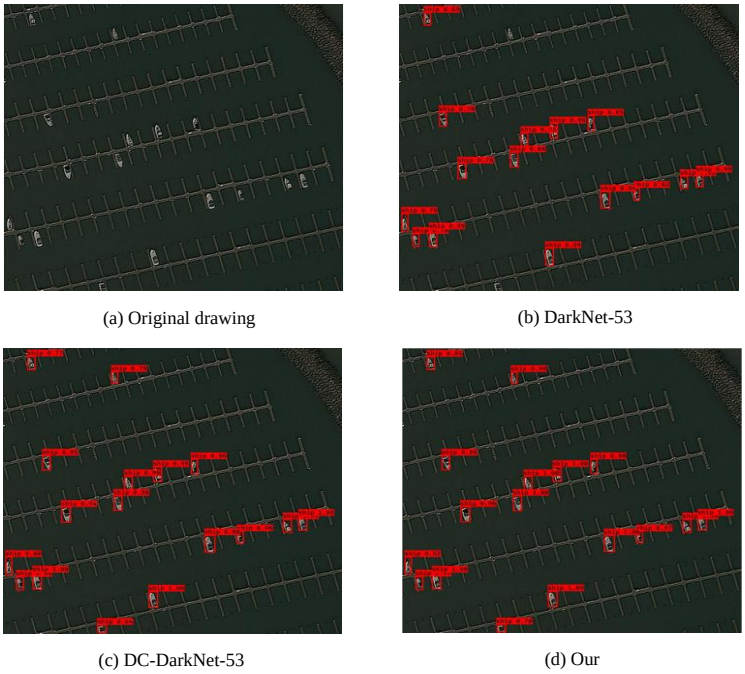
Figure 7: Comparison of detecting performance of four models

The maximum iteration was set to 50000, the model batch was set to 4 to verify the detecting performance of the object detection model using improved YOLOv3. These remaining conditions remained unchanged for experimentation. This model was compared with ResNet-101, VGG-16, DarkNet-53, and DC DarkNet-53 in Table 2. Among the five models, the research model showed the best performance in detection precision, missing alarm rate, false alarm rate, and AP, with values of 87.48%, 5.14%, 12.51%, and 94.35%, respectively. Although the training time was higher than DarkNet-53 and DC DarkNet-53, it was still within an acceptable range. These results confirmed that the object detection model based on improved YOLOv3 had good detecting performance.

Table 2: Comparison of detecting performance of five models

Model	Precision/%	Missing alarm rate/%	False alarm rate/%	Training time/h	AP/%
ResNet-101	79.58	42.09	20.42	-	53.37
VGG-16	78.51	43.44	21.48	-	51.29
DarkNet-53	86.95	7.79	13.05	6	91.31
DC-DarkNet-53	87.14	6.77	12.87	6	92.50
This research	87.48	5.14	12.51	8	94.35

DarkNet-53, DC-DarkNet-53, and the research model were used to detect ship images in MyShip to verify the practical application effect of the model. Figure 8 shows the final detecting results of the three models. There were 17 ships in the original sample by comparing Figures 8 (b), 8 (c), and 8 (d). The research model successfully detected 17 ships, while the DC-DarkNet-53 model only detected 15 ships. These results confirmed that the object detection model based on improved YOLOv3 had good practical application effects and detection accuracy.



(a) Original drawing (b) DarkNet-53

(c) DC-DarkNet-53 (d) Our

Figure 8: Final detection results of three models

The study compared the proposed model with the standard YOLOv3 and deep sorting algorithms using computational time and resource utilization as indicators to investigate the computational efficiency of the proposed model. The results are shown in Figure 9. The calculation time of the proposed model was 23.3 seconds, slightly longer than the standard YOLOv3 and deep sorting algorithms, but still within an acceptable range. The resource utilization rate was 76.65%, slightly lower than the standard YOLOv3. The results indicated that the calculation time of the proposed model increased, but it was still within an acceptable range and performed better overall.

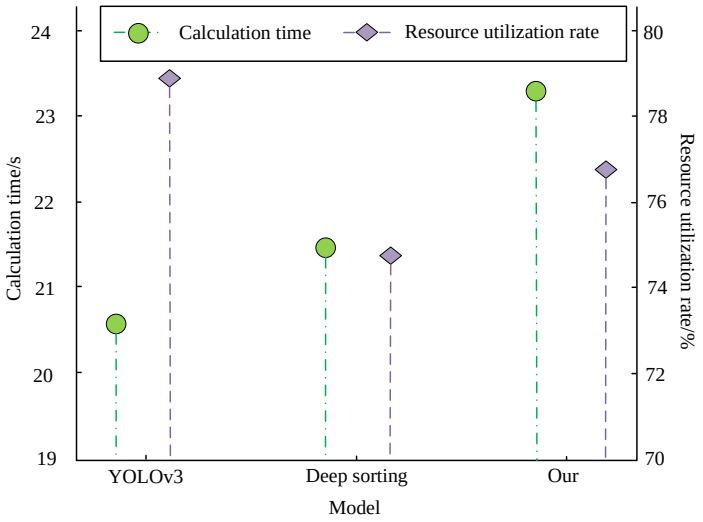


Figure 9: Comparison results of calculation time and resource utilization

4.2 Effectiveness analysis of ship tracking models

Experiments were conducted using the Ships in Satellite Imagery dataset to validate the modifying effect of the ship tracking method using the improved Deep-Sort. Recall rate, ID conversion, and Multi-Object Tracking Accuracy (MOTA) were used as indicators. The improved Deep-Sort was compared with traditional Deep-Sort, Deep-Sort with YOLOv3 detector, and Deep-Sort with improved YOLOv3 detector, denoted as Models 1, 2, and 3, respectively. From Figure 10 (a), the recall rate of the study model was the highest, at 64.7%. From Figure 10 (b), the research model also achieved good performance in ID conversion, with the lowest ID conversion of 804. From Figure 10 (c), the MOTA index of the research model was 61.8%, indicating good tracking accuracy. These results confirmed that the proposed ship tracking model based on improved Deep-Sort had better improvement effects compared to traditional Deep-Sort.

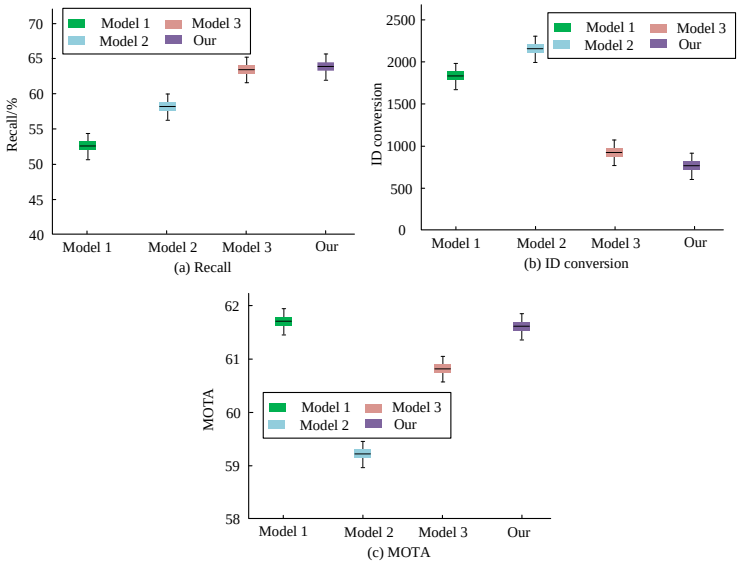


Figure 10: Tracking performance of four models

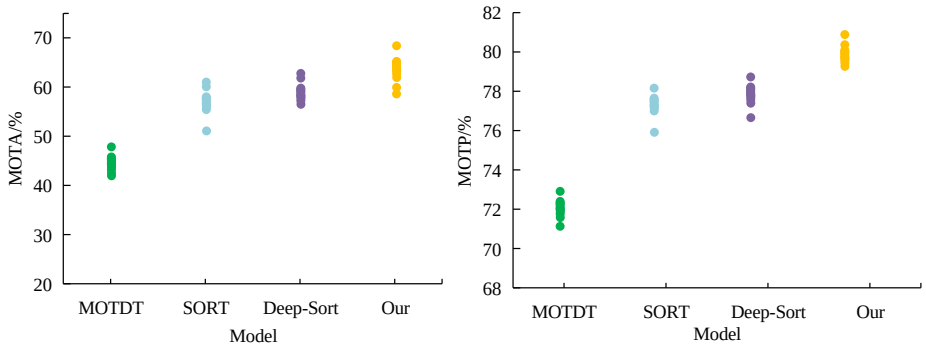
Experiments were conducted to further verify the detecting performance of the ship tracking model. MOTA, Multi-Object Tracking Precision (MOTP), total missing detection, and total false detection were used as indicators. The ship tracking model was compared with three algorithms: MOTDT, SORT, and Deep-Sort. Table 3 shows the comparison results. Among these four models, the MOTA and MOTP indicators of the research model were the highest, with 65.4% and 80.8%, respectively. The total missing detection and false detection were the lowest, at 53449 and 7964, respectively. These results confirmed that the proposed tracking model achieved

good performance and had certain feasibility and effectiveness.

Table 3: Comparison of tracking model effects

Model	MOTA/%	MOTP/%	Number of missing detection	Number of false positives
MOTDT	47.5	74.7	85433	9255
SORT	59.7	79.5	63246	8699
Deep-Sort	61.3	79.2	56559	12853
This research	65.4	80.8	53449	7964

Experiments were conducted using the USVInland dataset containing different weather conditions to verify the adaptability of the proposed model under different environmental conditions. Other conditions remained unchanged. The results of the MOTA and MOTP indicators for the four models are shown in Figure 11. From Figure 11 (a), the MOTA index of the proposed model was 65.2% in the USVInland dataset, which was higher than the comparison models. From Figure 11 (b), the MOTA index of the proposed model was 80.6%, which was still higher than the other three models. The results indicated that the ship tracking model based on the improved Deep-Sort algorithm had good tracking performance under different conditions.

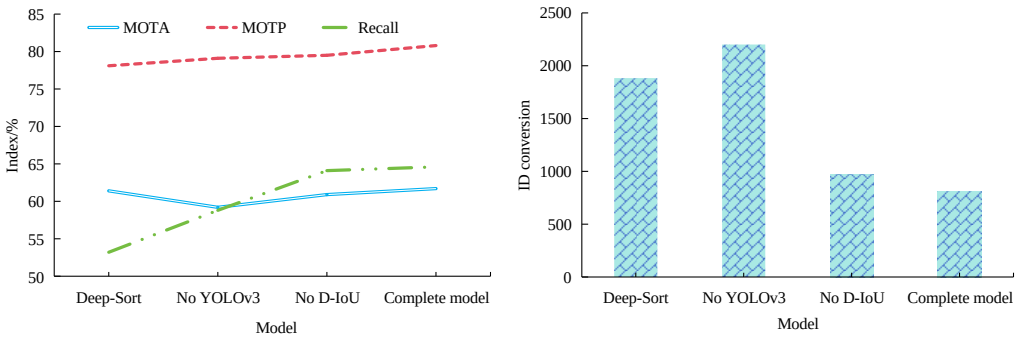


(a) Comparison results of MOTA indicators

(b) Comparison results of MOTP indicators

Figure 11: MOTA and MOTP indicators match the results

MOTA, MOTP, recall rate, and ID conversion were used as evaluation indicators to validate the feasibility of the proposed model, and the MyShip dataset was selected for ablation experiments. Figure 12 shows the outcomes of the ablation experiment. From Figure 12 (a), the complete ship tracking model performed the best for MOTA, MOTP, and recall, with values of 61.7%, 80.8%, and 64.6%, respectively. From Figure 12 (b), the ID conversion of the complete ship tracking model was the lowest, at 805. These results confirmed that the improved YOLOv3 was treated as a detector for the Deep-Sort, and D-IoU distance was introduced in the loss of cascade matching. The improved YOLOv3 effectively improved the ship tracking performance of the model and had certain feasibility and effectiveness.



(a) Comparison results of MOTA, MOTP, and recall

(b) Comparison results of ID conversion

Figure 12: Results of ablation experiment

5 Discussion

An object detection model based on the improved YOLOv3 algorithm and a ship tracking model based on the improved Deep-Sort algorithm were built to address the ship monitoring in river channels. The experimental results in the MyShip dataset showed that the proposed object detection model performed well with a detection accuracy of 6345, a recall rate of 93.25%, a missing alarm rate of 6.76%, and an AP index of 92.53%. The proposed model performed better than the Faster R-CNN with ResNet-101 and VGG-16 backbone networks and the traditional DarkNet-53 model. This is because dilated convolution can expand the receptive field of images, effectively improve the sensitivity of YOLOv3 algorithm to small targets, improving the detection accuracy and efficiency. The proposed target tracking model performed well in the Ship in Satellite Imagery dataset, with a recall rate of 64.7%, an ID conversion of 804, and a MOTA metric of 61.8%, demonstrating good tracking accuracy. The proposed model performed better than the traditional Deep-Sort algorithm, the Deep-Sort algorithm with YOLOv3 detector, and the Deep-Sort algorithm with improved YOLOv3 detector. This is because introducing D-IoU distance into the loss of cascade matching can obtain the complete trajectory of ship motion in the video, thereby determining the direction of ship and improving tracking accuracy.

This study conducted comparative experiments on actual ship images, demonstrating the better practical application of the proposed model compared with references [6] and [9]. The proposed model adopted the Deep-Sort algorithm for online real-time tracking, which had better real-time performance compared with reference [7]. The proposed target recognition model increased the computational complexity and time to a certain extent, which was similar to references [8], [10-14], and [16, 17]. Therefore, further methods such as introducing lightweight networks should be adopted to explore ways to improve computational efficiency while ensuring model recognition performance in the future. The proposed model performed better in ID conversion and better met the user needs in actual target tracking scenarios compared with reference [15].

The proposed object detection model demonstrated good performance in ship monitoring and tracking. This method can be applied to fields such as maritime rescue and road traffic monitoring. Therefore, rescue efficiency and road safety can be improved by identifying rescue targets and vehicles. However, object detection models for ships may be sensitive to morphological and texture features, which limits their applicability in scenarios other than river ship monitoring. In addition, there may be issues such as overlapping, occlusion, and target confusion among ships in densely populated situations. These issues may pose challenges for the model to accurately detect and track each ship target, affecting the practical application effect of the model. Therefore, target segmentation and recognition technology can be further combined to segment the target into separate parts in future research. Meanwhile, different sensor data can be combined to obtain more dimensional information to improve the accuracy of object detection and tracking of the model.

6 Conclusion

As the economy develops and intelligent information processing technology is continuously mature based on deep learning, ship monitoring technology is also moving towards intelligence and automation. An improved YOLOv3-based object detection model and an improved Deep-Sort-based ship tracking model were built to deal with the low accuracy of ship object detection and tracking. These results confirmed that the improved YOLOv3 had the highest positive detection, with 6345, followed by DarkNet-53 with 62759 correct detection, and VGG-16 had the worst performance with 38493 correct detections. The improved YOLOv3 had the highest recall rate of 93.25%, the lowest missing alarm rate of 6.76%, and the highest AP rate of 92.53%. The proposed object detection model performed the best in terms of detecting accuracy, missing and false alarm rates, and AP index, with values of 87.48%, 5.14%, 12.51%, and 94.35%, respectively. The proposed object detection model successfully detected all 17 ship targets in actual samples. The proposed ship tracking model had the highest recall rate of 64.7%, the lowest ID conversion rate of 804, and a multi-target tracking accuracy of 61.8%. In addition, the ship tracking performance could be effectively improved by using the improved YOLOv3 as the detector for the Deep-Sort and introducing D-IoU distance into the cascaded matching loss. In summary, the constructed model had certain feasibility and effectiveness. However, the data collected through research is still limited for the dissemination of object detection, which may affect the practical application effectiveness of the model. Therefore, more data should be collected in future research to validate the practical application effectiveness of the model.

7 Fundings

The research is supported by: Provincial level, Education Reform Project of Henan Provincial Department of Education, "Research on Comprehensive Interdisciplinary Practical Training in Finance and Economics Based on 'Double Innovation' in the Context of Free Trade Zones", (No. 2019SJGLX684).

References

[1] A. Amro, V. Gkioulos, and S. Katsikas, “Communication architecture for autonomous passenger ship,” Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and

Reliability, vol. 237, no. 2, pp. 459-484, 2023. https://doi.org/10.1177/1748006X211002546

[2] R. Li, J. Wu, and L. Cao, "Ship target detection of unmanned surface vehicle base on efficientdet," Systems Science & Control Engineering, vol. 10, no. 1, pp. 264-271, 2022. https://doi.org/10.1080/21642583.2021.1990159

[3] C. Yan, and C. Wang, "Ship target detection in sar image based on selective coordinate attention," Acta Electonica Sinica, vol. 51, no. 9, pp. 2481-2491, 2023. https://doi.org/ 10.12263/DZXB.20211416

[4] T. Yao, R. Miao, W. Wang, Z. Li, J. Dong, Y. Gu, and X. Yan, "Synthetic damage effect assessment through evidential reasoning approach and neural fuzzy inference: Application in ship target," Chinese Journal of Aeronautics, vol. 35, no. 8, pp. 143-157, 2022. https://doi.org/10.1016/j.cja.2021.08.010

[5] Y. Liu, D. Jiang, C. Xu, Y. Sun, G. Jiang, B. Tao, X. Tong, M. Xu, G. Li, and J. Yun, "Deep learning based 3D target detection for indoor scenes," Applied Intelligence, vol. 53, no. 9, pp. 10218-10231, 2023. https://doi.org/10.1007/s10489-022-03888-4

[6] C. Tsoumpris, and G. Theotokatos, "Performance and reliability monitoring of ship hybrid power plants," Journal of ETA Maritime Science, vol. 10, no. 1, pp. 29-38, 2022. https://doi.org/10.4274/jems.2022.82621

[7] C. Capezza, F. Centofanti, A. Lepore, A. Menafoglio, B. Palumbo, and S. Vantini, "Functional regression control chart for monitoring ship CO2 emissions," Quality and Reliability Engineering International, vol. 38,no. 3, pp. 1519-1537, 2022. https://doi.org/10.1002/qre.2949

[8] J. C. Zheng, S. D. Sun, and S. J. Zhao, "Fast ship detection based on lightweight YOLOv5 network," IET Image Processing, vol. 16, no. 6, pp. 1585-1593, 2022. https://doi.org/10.1049/ipr2.12432

[9] X. Wang, J. Liu, X. Liu, Z. Liu, O. I. Khalaf, J. Ji, and O.Y. Quan, "Ship feature recognition methods for deep learning in complex marine environments," Complex & Intelligent Systems, vol. 8, no. 5, pp. 3881-3897, 2022. https://doi.org/10.1007/s40747-022-00683-z

[10] T. Kim, L. P. Perera, M. P. Sollid, B. M. Batalden, and A. K. Sydnes, "Safety challenges related to autonomous ships in mixed navigational environments," WMU Journal of Maritime Affairs, vol. 21, no. 2, pp. 141-159, 2022. https://doi.org/10.1007/s13437-022-00277-z

[11] Y. Wang, Z. Hu, J. Wang, X. Liu, Q. Shi, Y. Wang, L. Qiao, Y. Li, H. Yang, J. Liu, L. Zhou, Z. Yang, C. Lee, and M. Xu, "Deep learning-assisted triboelectric smart mats for personnel comprehensive monitoring toward maritime safety," ACS Applied Materials & Interfaces, vol. 14, no. 21, pp. 24832-24839, 2022. https://doi.org/10.1021/acsami.2c05734

[12] D. Meimetis, I. Daramouskas, I. Perikos, and I. Hatzilygeroudis, "Real-time multiple object tracking using deep learning methods," Neural Computing and Applications, vol. 35, no. 1, pp. 89-118, 2023. https://doi.org/10.1007/s00521-021-06391-y

[13] C. W. Chang, C. Y. Chang, and Y. Y. Lin, "A hybrid CNN and LSTM-based deep learning model for abnormal behavior detection," Multimedia Tools and Applications, vol. 81, no. 9, pp. 11825-11843, 2022. https://doi.org/10.1007/s11042-021-11887-9

[14] A. Mathias, D. Samiappan, 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. https://doi.org/10.1007/s11042-022-13281-5

[15] X. Zou, Z. Yin, Y. Li, F. Gong, Y. Bai, Z. Zhao, W. Zhang, Y. Qian, and M. Xiao, "Novel multiple object tracking method for yellow feather broilers in a flat breeding chamber based on improved YOLOv3 and deep SORT," International Journal of Agricultural and Biological Engineering, vol. 16, no. 5, pp. 44-55, 2023. https://doi.org/10.25165/j.ijabe.20231605.7836

[16] A. L. Rishika, C. Aishwarya, A. Sahithi, and M. Premchender, "Real-time vehicle detection and tracking using YOLO-based deep sort model: A computer vision application for traffic surveillance," Turkish Journal of Computer and Mathematics Education (TURCOMAT), 14(1): 255-264, 2023. https://doi.org/10.17762/turcomat.v14i1.13530

[17] S. K. Sahoo, G. Palai, B. R. Altahan, S. H. Ahannad, P. P. Priya, M. A. Hossain, and A. N. Z. Rashed, "An optimized deep learning approach for the prediction of social distance among individuals in public places during pandemic," New Generation Computing, vol. 41, no. 1, pp. 135-154, 2023. https://doi.org/10.1007/s00354-022-00202-1

[18] S. Pal, A. Roy, P. Shivakumara, and U. Pal, "Adapting a Swin transformer for license plate number and text detection in drone images," Artificial Intelligence and Applications, vol. 1, no. 3, pp. 145-154, 2023. https://doi.org/10.47852/bonviewAIA3202549

[19] X. Zhou, and L. Zhang, "SA-FPN: An effective feature pyramid network for crowded human detection," Applied Intelligence, vol. 52, no. 11, pp. 12556-12568, 2022. https://doi.org/10.1007/s10489-021-03121-8

[20] K. Wang, and M. Liu, "YOLOv3-MT: A YOLOv3 using multi-target tracking for vehicle visual detection," Applied Intelligence, vol. 52, no. 2, pp. 2070-2091, 2022. https://doi.org/10.1007/s10489-021-02491-3

[21] A. Chaudhuri, "Hierarchical modified fast R-CNN for object detection," Informatica, vol. 45, no. 7, pp. 67-81, 2021. https://doi.org/ 10.31449/inf.v45i7.3732

[22] M. Z. Alam, and A. Jamalipour, “Multi-agent drl-based hungarian algorithm (madrlha) for task offloading in multi-access edge computing internet of vehicles (iovs),” IEEE Transactions on Wireless Communications, vol. 21, no. 9, pp. 7641-7652, 2022. https://doi.org/10.1109/TWC.2022.3160099