Dynamic Tactical Image Recognition and Analysis in Football Matches Using Convolutional Neural Networks

Qi Xie

School of Physical Education, Baoji University of Arts and Sciences, Baoji 721000, China

Corresponding Author Email: moray1981@163.com

ABSTRACT

With the increasing complexity of modern football tactics, how to intelligently and accurately analyze tactical changes in real-time during matches has become an important research direction. Traditional manual tactical analysis methods are inefficient and susceptible to subjective bias. Therefore, using computer vision and deep learning technologies for tactical image recognition and analysis in football matches has gradually become a research hotspot. Convolutional Neural Networks (CNNs), as a powerful image processing tool, have been widely applied in video analysis and player detection. However, multi-target motion prediction and tracking management in dynamic football match scenes still face significant challenges. Existing research mainly focuses on static image analysis or simple player tracking, but the high-frequency image updates, player interactions, and occlusion issues in football matches complicate multi-target tracking. While some deep learning-based methods for multi-target detection and tracking have made progress, challenges remain, such as handling high-density player targets and improving motion trajectory prediction accuracy. To address these shortcomings, this study proposes two core techniques based on CNNs: first, multi-target motion prediction, which accurately forecasts players' future positions based on historical motion data; second, multi-target tracking management, which uses deep learning to track and manage each player's movement trajectory in real-time. Through these two techniques, this research aims to improve the real-time and accuracy of tactical analysis in football matches, providing coaches and analysts with more scientific and efficient tactical decision-making support.

1. INTRODUCTION

With the increasing complexity and variability of tactics in modern football matches, the demand for real-time analysis of match dynamics by coaching teams is growing [1-4]. Traditional tactical analysis often relies on manual observation and recording, which is inefficient and difficult to cover the subtle changes in the match. In recent years, with the rapid development of computer vision and deep learning technologies, using optimized CNNs for tactical analysis in football matches has gradually become a research hotspot [5-9]. These technologies provide new ideas for automated tactical analysis, enabling real-time capture of movement trajectories and tactical changes on the field, thereby helping coaches and players better understand match dynamics and improve the accuracy of tactical execution.

In this context, dynamic tactical image recognition and analysis in football matches can not only provide real-time feedback for the match but also offer valuable information for post-match data analysis and tactical optimization. Therefore, how to effectively extract valuable tactical information from dynamic videos, especially how to process and analyze multi-target movement trajectories, has become a key research issue. By using CNNs for automatic image recognition and analysis, the efficiency of tactical analysis can be improved, and

subjective errors in manual analysis can be significantly reduced, advancing the depth and breadth of football tactical research.

However, existing research methods still have certain limitations. Many traditional tactical image analysis methods mainly focus on static images or limited motion trajectories, and research on multi-target motion prediction and tracking management in dynamic tactical images is still insufficient [10-14]. Although some deep learning-based multi-target detection and tracking methods have made significant progress in other fields, they still face numerous challenges in complex and dynamic scenarios like football matches [15-18], such as high-frequency image frame changes, rapid player interactions, occlusion issues, and the complexity of movement trajectories [19-23]. Therefore, accurately predicting target player movement and performing real-time tracking management in complex backgrounds remain key challenges in current research.

This paper addresses these issues through two main research tasks. First, it employs CNN-based multi-target motion prediction to anticipate player movements in football matches by analyzing historical trajectories and tactical changes. Second, it focuses on multi-target tracking management to ensure real-time and accurate tracking of players in dynamic scenes. These approaches aim to enhance the accuracy and

efficiency of tactical analysis while supporting future match prediction and tactical optimization.

2. MULTI-TARGET MOTION PREDICTION FOR DYNAMIC TACTICAL IMAGE FRAMES IN FOOTBALL MATCHES

2.1 Problem description

The positions of players in a football match constantly change on the field, influenced by various factors, including the players' own movement, interactions within and outside the field, and changes in the camera's viewpoint. Due to the high dynamics of football matches and complex tactical changes, a single motion model is difficult to adapt to all scenarios. Therefore, this paper proposes a motion prediction method that combines the Kalman filter with the Enhanced Correlation Coefficient (ECC) image registration method, aimed at overcoming this challenge. First, for predicting the movement trajectory of individual players, the Kalman filter can model each player's motion state and predict based on historical position data, addressing the rapid movement and position changes between players. As for the deviations between video frames caused by camera movement, the ECC image registration method is used to calculate the affine vector of consecutive frames, adjusting the differences between frames to reduce the impact of camera motion and ensure the accuracy of the prediction results.

Specifically, for tactical image analysis in football matches, player movement is not merely simple linear or rigid motion; it often includes complex non-rigid motions, such as acceleration, deceleration, and turning. These motion characteristics need to be processed through refined modeling and data association methods. With high-quality detection and high frame rates, data association methods based on Intersection over Union (IoU) can effectively achieve precise tracking of target players. However, under low frame rates or in complex scenarios, more accurate prediction and adjustment of the target player's movement position are required. By combining the Kalman filter for predicting non-rigid motion and ECC for aligning rigid motion, precise multi-target motion prediction can be achieved in dynamic football scenarios. Assume that the 8-dimensional motion state $[z_a,z_b,x,g,n_{za},n_{zb},n_x,n_g]$ is represented by t, where it includes the target player's center coordinates $[z_a,z_b]$, aspect ratio x, height g, and the rate of change of these four variables $[n_{za},n_{zb},n_x,n_g]$. The motion covariance is represented by W, the prior covariance by O, and the ECC model by WA. The model can be established using the following equation:

$$t_{s+1} = WA\left(Dt_t\right), O_{s+1} = DO_sD^S + W \tag{1}$$

Assume that the affine vector calculated by ECC is represented by Q, the affine vector of the static frame by E, the identity matrix by U, and the zero matrix by P. The intensity of the camera movement is defined by the following formula:

$$U_z = 1 - \frac{Q \times E}{\|Q\|_2 \times \|E\|_2}, E = [U;0] \tag{2}$$

Based on the above definitions, the Kalman filter can be adjusted by changing the state transition matrix. Assume that

the adjusted state transition matrix is represented by D_z, and the original time step of the Kalman filter is represented by f_s. Then, the state of the target player at frame $s+1$ is:

$$T_{s+1} = WA\left(D_zf_s\right), D_z = \begin{bmatrix} U\left(f_s+U_z\right) & U \\ 0 & U \end{bmatrix} \tag{3}$$

$$O_{s+1} = xD_zO_sD_z^S + W$$

After obtaining T_{s+1}, the vector $[z_a,z_b,x,g]$ of the position part can be further extracted, and the predicted bounding box of the target player in the dynamic tactical image frame of the football match can be calculated.

2.2 Network architecture

For the multi-target motion prediction problem in dynamic tactical image frames of football matches, this study designs a network architecture based on Faster-RCNN. By integrating CNNs and a motion prediction module, the network achieves precise location prediction for multiple football players. Specifically, the network architecture combines ResNet50 and the Feature Pyramid Network (FPN) as the backbone to fully utilize feature maps at different scales, thus improving the accuracy of target player detection and motion trajectory prediction, as shown in Figure 1. On this basis, the network further introduces three functional modules: the regression head, the classification head, and the ReID head. The regression head refines the target player's bounding box and outputs the precise player position. The classification head determines whether the image region is a target player or background, ensuring the accuracy of detection. The ReID head extracts the appearance feature vector of each target player for effective association and identification across consecutive frames. This structure not only effectively handles multiple target players in dynamic scenes but also enables long-term tracking and association between target players.
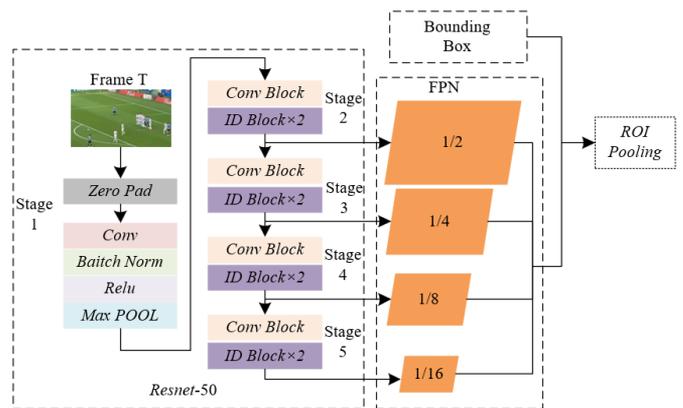

Figure 1. Basic backbone network structure of the constructed model

2.3 Training process

In the training process of the multi-target motion prediction algorithm for dynamic tactical image frames in football matches, this study simulates a real multi-target tracking task. By combining target player detection and motion prediction models, the training data's representativeness and diversity are enhanced. Unlike traditional methods that train based solely on detection tasks, this study adopts a strategy that generates

supplementary training samples by predicting the target player's position through a motion model. During the training phase, N consecutive frames are randomly selected from the training dataset as input. The tracking state of the target player is initialized using the true label from the first frame, including the player's position, velocity, and historical ROI feature set. Due to the rapid movement and occlusion of target players in football matches, a single target player detection model cannot handle all scenarios. Therefore, it is necessary to combine the motion model to predict the target player's position, ensuring the continuity and stability of tracking. Specifically, the target player's position is initialized using a Kalman filter, and historical ROI features are input into the ReID head to obtain the appearance features of each target player. The specific process flow is shown in Figure 2.
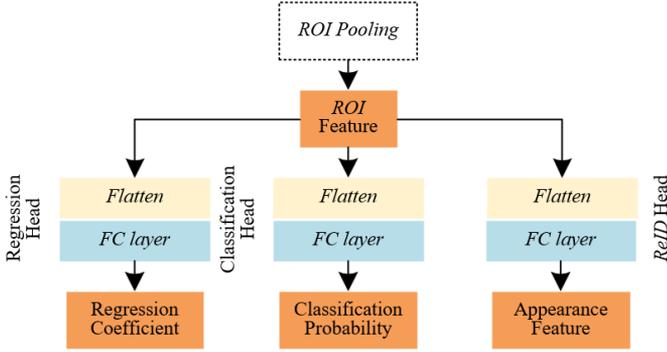

Figure 2. Process flow of ROI features input into the ReID Head

When predicting the next frame of dynamic tactical image frames in the football match, the losses for three parts are computed using the ground truth (GT) label: the loss from the Region Proposal Network (RPN) denoted as M^{mpz}_{EOV} and M^{zmt}_{EOV}; the loss from the regression model denoted as $M^{mpz}_{R\_H}$ and $M^{zmt}_{C\_H}$; and the loss used to train the ReID head denoted as M_{ME}. Assuming the factors controlling the influence of different sub-losses on the target player function are represented by η_1, η_2, and η_3, the overall target player loss function for the constructed model is as follows:

$$LOSS = \eta_1 \left(M^{mpz}_{EOV} + M^{zmt}_{EOV} \right)$$
$$+ \eta_1 \left(M^{mpz}_{R\_H} + M^{zmt}_{C\_H} \right) + \eta_3 M_{ME} \tag{4}$$

In multi-target motion prediction for football matches, considering the fast movement and frequent occlusions of players on the field, the RPN loss function needs to ensure that the network can adapt to the mutual influences between target players and the rapid positional changes of target players. To meet this demand, in addition to using traditional IoU threshold-based positive and negative sample judgments, the RPN loss function needs to account for the specificity of motion prediction. For instance, in some fast-moving scenarios, the predicted position of the target player may differ significantly from the real position. In such cases, introducing supplementary samples based on the motion model can enhance the RPN's adaptability in these complex scenarios. Additionally, since the motion trajectories of target players in football matches are highly nonlinear, the model must not only correctly identify the bounding box of the target player but also predict the potential position of the target player in future frames.

Specifically, for each anchor point, we assign labels based on the IoU between the anchor point and the target player's bounding box. Positive anchor boxes are of two types: (1) the one with the highest IoU with a specific target player's label bounding box; (2) the ones with IoU greater than 0.7 with any target player's label bounding box. Negative samples are those anchor boxes whose IoU with all target player's bounding boxes is below 0.3. As with the standard operation of Faster-RCNN, any anchor point that is neither a positive sample nor a negative sample is ignored and not included in the loss calculation. Let the index of the anchor point be represented by u, the prediction probability of the anchor box containing the target player be represented by O_u, the vector of the four parameterized coordinates of the predicted bounding box be represented by s_u, and the coordinates of the ground truth box associated with the positive anchor be represented by s^*_u. The log loss between the target player and non-target player classes is represented by M_{zmt}, and the loss function expression is as follows:
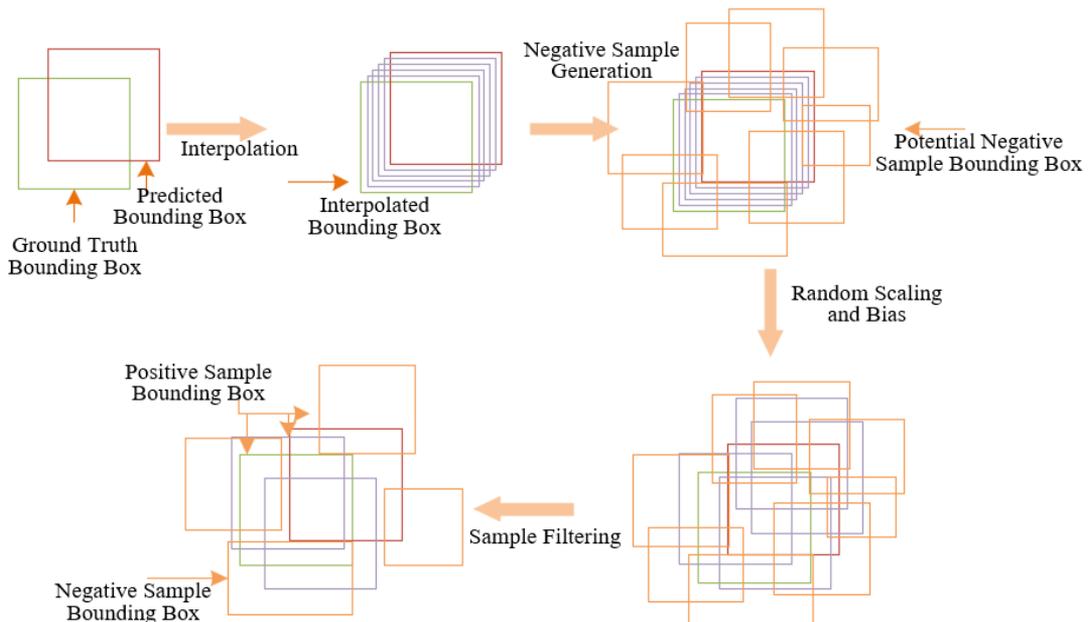

Figure 3. Illustration of the supplementary sample generation process

$$M_{EOV}^{zmt} = \frac{1}{V} \sum_u M_{zmt} \left(O_u, O_u^* \right) \tag{5}$$

$$M_{EOV}^{zmt} = \frac{1}{V_{RE}} \sum_u O_u^* M_{RE} \left(s_u, s_u^* \right) \tag{6}$$

Considering the rapid movement and potential occlusion of target players in football matches, the regression model's loss function needs to not only accurately predict the target player's position but also handle the complexity related to movement. Therefore, we construct training samples by combining the regions proposed by RPN and the predicted bounding boxes generated by the motion prediction module. During training, we first use the Ninter interpolation method to generate interpolated bounding boxes, which are created in such a way that they include both the true target player positions and are close to the motion prediction results. This allows us to generate more positive samples, especially when the number of target players is small, avoiding the problem of insufficient negative samples leading to data imbalance. The generation of supplementary samples proceeds through four steps: bounding box interpolation, negative sample generation, random scaling and shifting, and sample filtering. The process flow is illustrated in Figure 3. Each potential positive sample bounding box is trained alongside multiple negative samples to ensure that the model learns effective target player localization under various conditions.

To further enhance the regression model's performance, the regression loss function needs to account for both the balance between positive and negative samples and the accuracy of target player prediction. In dynamic football match scenes, players' movement trajectories are usually nonlinear and highly time-varying, requiring the regression model to not only consider the current frame's target player position but also predict the target player's future position. Therefore, in the loss function calculation, in addition to the traditional regression error, an error metric between the motion prediction results and the actual position is also introduced. Specifically, for each positive sample, the regression network predicts the target player's position and calculates the deviation between the predicted bounding box and the true target player. If the IoU value between the predicted box and the true target player box is below a certain threshold, it is considered a regression failure, and the error is penalized through the loss function. Additionally, the regression model must handle interactions and occlusions between target players, especially in cases of overlapping or alternating occlusions, ensuring that the model maintains stable prediction performance in a multi-target environment. Let the number of samples be represented by V_{SAM}, the number of positive samples be represented by V_{P_S}, and the probability of the category of target players (football players or background) within the sample bounding box be represented by O_u. The loss function expression is as follows:

$$M_{Z\_H}^{zmt} = \frac{1}{V_{SAM} \sum_u M_{zmt} \left(O_u, O_u^* \right)} \tag{7}$$

$$M_{R\_H}^{zmt} = \frac{1}{V_{P\_S} \sum_u O_u^* M_{RE} \left(s_u, s_u^* \right)} \tag{8}$$

In the multi-target motion prediction of dynamic tactical image frames in football matches, the design and training of the ReID head are crucial for accurately identifying and distinguishing players. To improve the stability and consistency of the model in recognizing the appearance features of the same player across different frames, this study adopts a metric loss function to optimize the extraction of appearance features in the ReID task. During training, by simulating the tracking process of target players, the appearance feature vectors are obtained from the true position of the target player in each frame, and these features are used to enhance the similarity of appearance features of the same target player across different time frames, thus strengthening their temporal correlation. This approach effectively reduces the potential issue of inconsistency in appearance features caused by insufficient random sampling in the training data. If the appearance features lack sufficient temporal correlation, the model is prone to identity switching during inference, meaning it could incorrectly identify different players as the same player or assign the same player to different identities, leading to a decrease in tracking stability and accuracy.

To further enhance the performance of the ReID head, this study adopts the idea of metric loss and makes certain adjustments. Specifically, in addition to using position data, the proposed scheme in this chapter computes the metric loss solely based on the appearance feature vector of the players, without involving position information. This is because, in a football match, the position data may become unreliable due to rapid player movements, positional changes, and occlusions. Therefore, this study focuses on judging identity consistency through the appearance features of the players. By extracting the appearance features of the target player from the current frame and comparing them with the appearance features from historical frames, the model can further encode the target player's appearance feature sequence using a Bidirectional Recurrent Neural Network (Bi-RNN). These continuous feature sequences are processed by two independent Bi-RNNs, producing phase-specific hidden layer representations, which are then converted into a soft assignment matrix through a fully connected layer. This ensures that the appearance features of each target player maintain temporal consistency, reducing the risk of misidentification. Finally, the soft assignment matrix generated through a sigmoid activation function optimizes the accurate maintenance of player identities, improving the stability and accuracy of tracking each player in multi-target motion prediction.

Specifically, let the appearance feature of the target player in the current frame extracted from the target player's labeled bounding box be represented by d_{xoo}. The historical appearance features of the target player and the average historical appearance feature computed by the ReID head are represented by d_{H_A}. Let the cosine distance between X and Y be denoted as $COS(X, Y)$. The following equation gives the calculation of the feature distance between the current frame's target player appearance feature and the saved average historical appearance feature:

$$DIS_{xoo} = \frac{1}{2} \left(1 - COS \left(d_{xoo}, d_{H\_A} \right) \right) \tag{9}$$

In multi-target motion prediction for dynamic tactical image frames in football matches, the metric loss function of the ReID head mainly consists of the Multi-Target Tracking Accuracy (*FMA*) and Multi-Target Tracking Precision (*FMP*). These two indicators help optimize the model's target tracking

performance in complex match environments. *FMA* evaluates the overall quality of the tracking results, taking into account false positives, missed detections, and identity switches. *FMP* measures the tracking accuracy of the target, calculating the error between the predicted position of the target in the current frame and the actual position. Let the approximate representations of false positives, missed detections, and identity switches be $\tilde{DO}$, $\tilde{DV}$, and $\widetilde{UFT}$, respectively. The matching target players are denoted by L, the binary assignment matrix corresponding to the distance matrix F_{D-A} is represented by Y^{SO}, and the weight factor controlling the proportion is represented by ε_{ME}. The following formula gives the calculation:

$$FMA = 1 - \frac{\tilde{DO} + \tilde{DV} + \varepsilon_{ME}\widetilde{UFT}}{L} \qquad (10)$$

$$FMP = 1 - \frac{\left\| F_{D\_A} \otimes Y^{SO} \right\|_1}{\left\| Y^{SO} \right\|_0} \qquad (11)$$

To effectively compute these metrics, this study constructs matrices Z^S and Z^Z by adding a row or column to the soft assignment matrix, respectively filling it with a threshold σ and performing the softmax operation, further refining the matching relationship between targets.

$$\tilde{DO} = \sum_{V_{TA}} Z^e_{V_{TA}, V_{TA}+1}, \tilde{DV} = \sum_{V_{TA}} Z^z_{V_{TA}, V_{TA}+1} \qquad (12)$$

In the calculation of *FMA* and *FMP*, the method in this paper avoids saving the binary assignment matrix of each frame, instead directly utilizing the target matching relationship between the previous frame and the current frame to simplify the calculation of *IDS*. The core strategy here is that by only considering the targets that exist in both the current frame and the previous frame, the computational complexity is reduced, and real-time performance is improved. Especially in football matches, player position changes and rapid movements may lead to occlusions and mismatches. Therefore, by directly calculating the target matching status for each frame, the temporal consistency and accuracy of the ReID head are ensured. In the specific implementation, this study uses the calculation formulas for false positives, missed detections, and identity switches, further optimizing the matching of target appearance features and positions during the target tracking process, thereby improving overall tracking accuracy.

When calculating the distance matrix F_{D-A}, this paper only considers the target players that exist in both the current frame and the previous frame of the dynamic tactical image of the football match, and the order of the target players in F_{D-A} and X^-_{D-A} correspond to each other. Let the L_1 normalization of the flattened matrix be denoted by $\|\bullet\|_1$. The matrix of size $V_{TA} \times V_{TA}$, with diagonal elements being 0 and other elements being 1, is denoted by U^-_{VT}. The simplified calculation formula is:

$$\widetilde{UFT} = \left\| Z^z_{1:V_{TA}, 1:V_{TA}} \otimes \bar{U}_{V_{TA}} \right\|_1 \qquad (13)$$

Finally, the weight factor η_{ME} for controlling the sub-loss ratio is represented, and the metric loss is assumed as follows:

$$M_{ME} = (1 - FMA) + \eta_{ME}(1 - FMP) \qquad (14)$$

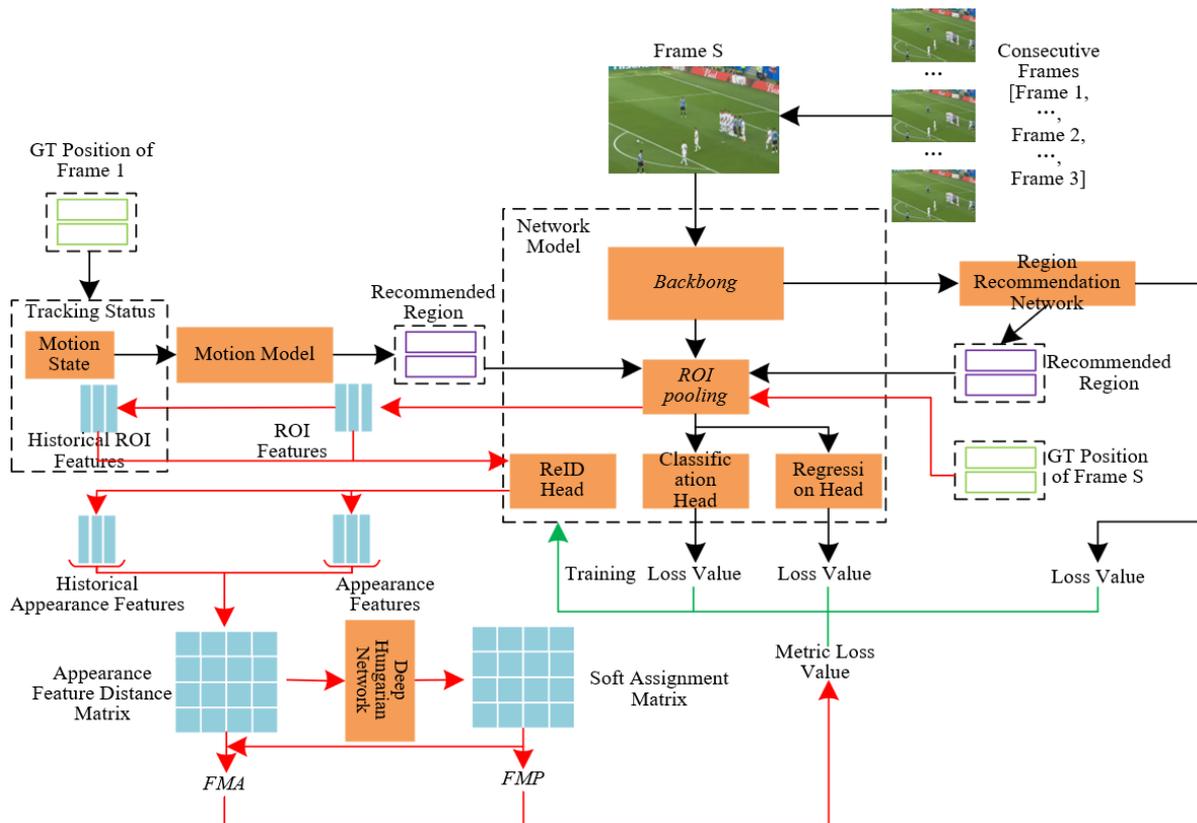The overall training strategy of the proposed network model is shown in Figure 4.

Figure 4. The training strategy of the proposed network model

2.4 Multi-target re-identification module

In the multi-target motion prediction algorithm for dynamic tactical image frames in football matches, the construction of the multi-target re-identification module is crucial because players frequently change positions during the match and may encounter occlusion, which poses significant challenges to the stability and accuracy of target tracking. To effectively address these issues, the algorithm constructs a simple yet efficient football player re-identification module through the appearance features extracted from the ReID head. The core idea of this module is to re-identify the target in case of tracking loss, utilizing the appearance features. Specifically, the algorithm saves the lost tracking information in fixed image frames, avoiding a complete loss of the target when it disappears. Whenever tracking is lost, the appearance feature distance between the lost track and the newly detected target is calculated, and re-identification is performed based on a set threshold. In this way, even if the target briefly disappears or is occluded by another target, re-identification can effectively restore the tracking information, thereby maintaining multi-target tracking stability.

To further improve re-identification accuracy and avoid erroneous re-identification, this study uses the IoU intersection-over-union threshold of the bounding box to filter out mismatched targets. This means that re-identification will only be performed if the IoU of the bounding boxes of the lost track and the newly detected target exceeds the set threshold. This strategy effectively reduces mismatches caused by targets being close in position or having similar appearance, thus improving the algorithm's robustness in complex dynamic environments. At the same time, the motion model continues to be applied to the lost tracking targets to supplement their motion information, further improving re-identification accuracy and real-time performance.

3. MULTI-TARGET TRACKING MANAGEMENT IN DYNAMIC TACTICAL IMAGE FRAMES OF FOOTBALL MATCHES

In the multi-target tracking management of dynamic tactical image frames in football matches, it is crucial to ensure that the movement trajectories of players can be accurately tracked in each frame and that the system can handle rapid movements, occlusions, and interactions between multiple target players. To achieve this, the tracking management process in this study is divided into five key steps:

(1) Tracking initialization
Tracking initialization is the first step in multi-target tracking management. In this phase, the first frame of the video is selected, and the filtered detection results are treated as the tracking positions for each target player, typically represented by the player's bounding box $o_s=[a,b,q,g]$, where a and b are the center coordinates of the bounding box, and q and g are the width and height, respectively. These positions are input into the model's ReID head to extract the appearance features of each target player. Based on the initial positions and appearance features of the target players, the tracking state of each player is constructed, including historical positions g_{os} and historical appearance features g_{ds}. This information is stored as a tracking set $T=\{t_s\}$, where s represents the frame index of the target player, and the number of target players is S. Through this process, the model can correctly identify and track these target players in subsequent frames.

(2) Motion prediction
Motion prediction is the second step in tracking management. In this phase, based on the historical positions g_{os} of the target players, a motion model is used to predict the position of each target player in the current frame. To accommodate the rapid movement characteristics of players in football matches, this study employs the Kalman filter and the ECC model for motion prediction. The Kalman filter provides state estimation based on the historical positions of the target players, while the ECC model corrects the trajectories of the target players, ensuring accurate prediction under high-speed and complex scenarios. Through motion prediction, the model obtains the predicted position o^{PE}_s of the target player in the current frame, providing an initial location for further bounding box refinement and data association.

(3) Bounding box refinement
Bounding box refinement is the third step and can be seen as a fine adjustment of the target player's position. Although motion prediction provides an initial location for the target player, dynamic changes and complex backgrounds in football matches often introduce errors in the predicted position. To reduce these errors, the predicted position o^{PE}_s is input into the regression head, which outputs the refinement coefficients $[s_a,s_b,s_q,s_g]$, representing the correction amounts for the target player's position. Using these regression coefficients, the model refines the predicted position, resulting in a more accurate position o^{RE}_s for the target player. This process is similar to the detection correction step in single-target player tracking but is crucial in accurately determining each target player's position in complex environments.

(4) Data association
Data association is the fourth step, aimed at matching the target players with the detections by computing the IoU between the predicted positions of each target player and the detection results in the current frame. Specifically, the larger the IoU value, the more overlap there is between the target player and the detection, and the higher the probability of a match. Using a greedy matching algorithm, the model selects the target player and detection with the highest IoU for association and updates the refined position o^{RE}_s of the matched target player to its tracking position in the current frame. In this process, the ReID head extracts the appearance features of the matched target player again and adds them to the historical appearance feature set of the target player. For target players that do not match any detection, they are marked as lost targets and will be handled by the subsequent re-identification module for re-identification and tracking recovery.

(5) New trajectory addition
The final step is the addition of new trajectories, which primarily handles new detection target players in the current frame. If any detections in the current frame do not match any existing target players, they are treated as new target players and initialized as new tracking trajectories. This process is similar to the initialization step. First, the bounding box of the new detection target player is used to extract appearance features through the ReID head, and then it is added to the tracking set as a new tracking state. In the subsequent tracking process, the new target player will be continuously tracked and associated with other target players to ensure accurate tracking of all players. Through this process, the system can effectively handle dynamic situations in football matches, ensuring that new target players or player switches can be quickly and

accurately initialized and tracked.

Through these five steps, multi-target tracking management in dynamic tactical image frames of football matches can achieve precise tracking, ensuring that target players maintain stable tracking in high-speed and variable match scenarios. This method can not only handle rapidly moving players but also deal with complex backgrounds and occlusions, providing a solid data foundation for subsequent tactical analysis and motion prediction.

4. EXPERIMENTAL RESULTS AND ANALYSIS

From the data of the line chart on the left side of Figure 5, it can be seen that under the zone defense tactic, the cumulative number of defenders increases significantly over time, eventually approaching 400. In contrast, the number of defenders under the full-field pressing tactic remains relatively low and fluctuates less. The image on the right shows the tactical scenes, with different colored boxes marking different players, representing the results of tracking management. This indicates that in the zone defense tactic, the dynamic changes of players are more frequent, and the number of defenders gradually increases. This is because the tactic requires constant adjustments to defensive zones and player positioning based on the ball's location and the offensive team's changes. Therefore, multi-target motion prediction and tracking

management based on CNNs face greater challenges. On the other hand, the full-field pressing tactic is relatively stable, with little change in the number of defenders, making the difficulty of tracking and prediction relatively low. Hence, the proposed method shows different levels of adaptability to various tactical scenarios with varying complexity. It is more adept at handling relatively stable full-field pressing tactics, while there is still room for optimization in dealing with the more complex and dynamic zone defense tactics.

From the multiple comparison experiment data in Table 1, it can be seen that the proposed method outperforms significantly in multi-target tracking accuracy (48.9%), main target tracking rate (28.9%), and main target missed rate (23.5%). Compared to other methods, the CNN-based multi-target tracking and motion prediction method presented in this study shows a substantial improvement in both accuracy and robustness. For instance, compared to Gradient Boosting Tree (36.2%) and Regularized Particle Filter (37.5%), the multi-target tracking accuracy of the proposed method has improved by about 12-13 percentage points. At the same time, the false alarm count and false negative count are also reduced compared to other methods. The proposed method has 2895 false alarms and 22368 false negatives, which is significantly better than methods such as Optical Flow Method (false alarm count 7263, false negative count 25146), demonstrating the reliability and efficiency of the method in practical applications.
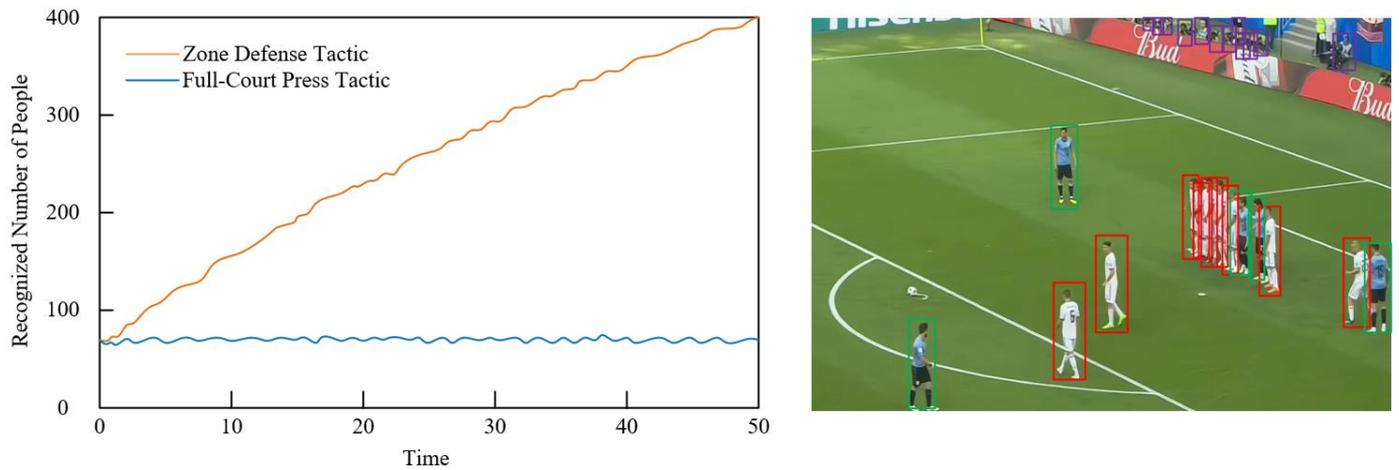

Figure 5. Comparison of test results on different tactics

Table 1. Comparison on public sports event video screenshot dataset

Method	Multi-Target Tracking Accuracy (%)	Multi-Target Tracking Precision (%)	Main Target Tracking Rate (%)	Main Target Missed Rate (%)	False Alarm Count	False Negative Count
Gradient Boosting Tree	36.2	71.2	12.9	41.2	6658	31526
Regularized Particle Filter	37.5	71.5	12.5	32.6	5326	32589
Extended Kalman Filter	37.6	72.6	16.5	42.8	4248	26598
Optical Flow Method	37.4	71.5	15.4	32.6	7263	25146
Bidirectional LSTM	43.2	71.2	17.9	25.6	6358	25632
Monocular Visual SLAM	45.6	75.6	17.8	26.7	4598	25318
Proposed Method	48.9	76.8	28.9	23.5	2895	22368

Table 2. Comparison on dataset provided by professional sports data company

Method	Multi-Target Tracking Accuracy (%)	Multi-Target Tracking Precision (%)	Main Target Tracking Rate (%)	Main Target Missed Rate (%)	False Alarm Count	False Negative Count
Gradient Boosting Tree	47.6	72.1	16.5	35.2	9136	83265
Regularized Particle Filter	47.5	72.9	16.5	37.9	5789	85468
Extended Kalman Filter	48.2	74.6	14.7	33.6	7236	83261
Optical Flow Method	53.6	77.8	18.9	35.6	3154	78956
Bidirectional LSTM	53.8	76.2	18.2	36.8	2896	77841
Monocular Visual SLAM	55.4	78.9	21.6	34.2	2236	75623
Proposed Method	58.9	78.8	23.5	31.2	2569	67895

Table 3. Comparison on specific scene dataset

Method	Multi-Target Tracking Accuracy (%)	Multi-Target Tracking Precision (%)	Main Target Tracking Rate (%)	Main Target Missed Rate (%)	False Alarm Count	False Negative Count
Gradient Boosting Tree	51.4	75.1	18.9	32.5	13265	246358
Regularized Particle Filter	51.8	75.3	22.3	31.2	24658	223157
Extended Kalman Filter	51.5	75.4	-	-	21305	230152
Optical Flow Method	52.3	77.9	18.9	35.6	12035	235684
Bidirectional LSTM	52.7	76.2	18.5	35.8	11245	231056
Monocular Visual SLAM	55.6	77.5	22.3	34.2	8795	223658
Proposed Method	61.2	77.8	25.6	28.9	12365	210356

Table 4. Ablation analysis results comparison

Target Player Motion Prediction	Bounding Box Refinement	Data Association	Multi-Target Tracking Accuracy (%)	False Alarm Count	False Negative Count
			43.26%	1526	11256
√			44.58%	1548	11458
√	√		52.31%	723	11895
√	√	√	51.26%	659	12356

Based on the comparison data in Table 2, the proposed method demonstrates superior performance in several key metrics, including multi-target tracking accuracy (58.9%), main target tracking rate (23.5%), and main target missed rate (31.2%). Compared to other methods, the proposed method leads in multi-target tracking accuracy, especially outperforming methods such as Optical Flow Method (53.6%), Bidirectional LSTM (53.8%), and Monocular Visual SLAM (55.4%). Furthermore, the main target missed rate (31.2%) of the proposed method is significantly lower than other methods (e.g., Extended Kalman Filter at 33.6%, Regularized Particle Filter at 37.9%), indicating better tracking stability. Meanwhile, the false alarm count (2569) and false negative count (67895) are reduced compared to other methods (e.g., Gradient Boosting Tree with 9136 false alarms and 83265 false negatives), further proving the efficiency and low false alarm capability of the proposed method.

According to the data in Table 3, the proposed method achieves the highest multi-target tracking accuracy (61.2%) and main target tracking rate (25.6%) on the specific scene dataset, showing a clear advantage over other methods.

Compared to Gradient Boosting Tree (51.4%), Regularized Particle Filter (51.8%), and other traditional methods, the proposed method improves multi-target tracking accuracy by approximately 9 percentage points. Furthermore, the proposed method also performs well in the main target missed rate (28.9%), significantly reducing the number of lost targets compared to methods such as Regularized Particle Filter (31.2%) and Bidirectional LSTM (35.8%). Although the false alarm count (12365) is slightly higher than other methods, such as Monocular Visual SLAM (8795), the proposed method has a significantly lower false negative count (210356), indicating its advantage in reducing false negatives and maintaining player trajectories more accurately.

These experimental results further confirm that the CNN-based multi-target motion prediction and tracking method can effectively improve multi-target tracking accuracy and stability in specific scenes, especially in complex and dynamic football match scenarios. By combining historical motion trajectories and tactical changes for prediction, the proposed method significantly reduces target loss, improves tracking accuracy, and decreases false negative counts, enhancing its

reliability in practical applications. Although the false alarm count is slightly higher, this result still demonstrates that the method can achieve more precise player tracking in real-time tracking systems, especially in fast-moving and complex scenarios, with broad application prospects.

From the ablation analysis data in Table 4, it can be seen that the integration of different modules significantly impacts the multi-target tracking performance. First, target player motion prediction alone (43.26% accuracy) provides an initial tracking result, but the accuracy is lower, and the false alarm count (1526) and false negative count (11256) are relatively high. After adding bounding box refinement, the accuracy increases to 44.58%, but both false alarms and false negatives increase, indicating that bounding box refinement improves accuracy while introducing additional computational and association complexity. When the data association module is added, the tracking accuracy significantly improves to 52.31%, and both the false alarm count and false negative count decrease, suggesting that data association effectively optimizes the tracking process. Finally, the combined method of motion prediction, bounding box refinement, and data association achieves an accuracy of 51.26%, which is slightly lower than the result using data association alone, but it maintains a relatively balanced false alarm and false negative count (659 and 12356), demonstrating the comprehensive advantage of the method across all aspects.

These ablation experiment results indicate that the collaborative effect of the three modules—motion prediction, bounding box refinement, and data association—is crucial in multi-target tracking. While motion prediction alone can provide basic target tracking, it lacks precise bounding boxes and data association support, leading to higher false alarms and false negatives. Bounding box refinement enhances the accuracy of target localization but introduces additional errors, especially in fast-moving scenes. The optimal result occurs when all three modules are combined, where, despite a slight decrease in accuracy compared to using data association alone, false alarms and false negatives are significantly reduced, showcasing the model's robustness in handling complex dynamic scenes.

5. CONCLUSION

In this paper, two major research tasks were conducted using CNNs in football match scenarios: first, multi-target motion prediction for dynamic tactical image frames, and second, multi-target tracking management. By combining historical motion trajectories and tactical changes, the proposed method not only predicts the player's motion state in the next moment but also tracks each player's motion trajectory in real-time and accurately within complex dynamic scenes. Experimental results demonstrate that the proposed method exhibits excellent performance on both professional sports datasets and specific scene datasets, significantly improving multi-target tracking accuracy and stability while reducing false alarms and false negatives. Ablation experiments further validate the collaborative effect of the three modules—motion prediction, bounding box refinement, and data association—demonstrating the robustness of the integrated solution in handling complex dynamic scenes.

Despite the significant achievements in multi-target tracking, there are still some limitations. For example, false alarms remain relatively high in some scenes, particularly in target-dense or rapidly changing scenarios, which may affect the overall tracking performance. Additionally, the current model still has room for improvement in handling highly complex and variable tactical changes, especially when dealing with extreme cases and anomalous motion trajectories. To address these limitations, future research can focus on the following directions for improvement and expansion: first, further optimize the balance between false alarms and false negatives by adopting more advanced bounding box refinement algorithms and data association methods to improve tracking accuracy and robustness; second, integrate other deep learning techniques, such as reinforcement learning and generative adversarial networks, to enhance the model's predictive ability and adaptability; and finally, expand the diversity and scale of the dataset, especially by incorporating more real-world match data and various tactical scenarios, to enhance the model's generalization ability and practical application effectiveness. Through these improvements, the overall performance of the multi-target tracking management system can be further enhanced, promoting its widespread application in more practical use cases.

REFERENCES

[1] Wang, S., Pu, Z., Pan, Y., Liu, B., Ma, H., Yi, J. (2024). Long-term and short-term opponent intention inference for football multi-player policy learning. IEEE Transactions on Cognitive and Developmental Systems, 16(6): 2055-2069. https://doi.org/10.1109/TCDS.2024.3404061

[2] Wang, J., Chen, J. (2022). Design and research of dynamic evolution system in football tactics under computational intelligence. Mathematical Problems in Engineering, 2022(1): 3772236. https://doi.org/10.1155/2022/3772236

[3] Moura, F.A., van Emmerik, R.E., Santana, J.E., Martins, L.E.B., Barros, R.M.L.D., Cunha, S.A. (2016). Coordination analysis of players' distribution in football using cross-correlation and vector coding techniques. Journal of Sports Sciences, 34(24): 2224-2232. https://doi.org/10.1080/02640414.2016.1173222

[4] Fang, L., Wei, Q., Xu, C.J. (2021). Technical and tactical command decision algorithm of football matches based on big data and neural network. Scientific Programming, 2021(1): 5544071. https://doi.org/10.1155/2021/5544071

[5] Itoh, M., Chua, L.O. (2007). Advanced image processing cellular neural networks. International Journal of Bifurcation and Chaos, 17(4): 1109-1150. https://doi.org/10.1142/S0218127407017896

[6] Cuevas, E., Díaz-Cortes, M.A., Mezura-Montes, E. (2019). Corner detection of intensity images with cellular neural networks (CNN) and evolutionary techniques. Neurocomputing, 347: 82-93. https://doi.org/10.1016/j.neucom.2019.03.014

[7] Ilesanmi, A.E., Ilesanmi, T.O. (2021). Methods for image denoising using convolutional neural network: A review. Complex & Intelligent Systems, 7(5): 2179-2198. https://doi.org/10.1007/s40747-021-00428-4

[8] Paul, A., Bhoumik, S., Chaki, N. (2021). SSNET: An improved deep hybrid network for hyperspectral image classification. Neural Computing and Applications, 33: 1575-1585. https://doi.org/10.1007/s00521-020-05069-1

[9] Frei, M., Kruis, F.E. (2021). FibeR-CNN: Expanding Mask R-CNN to improve image-based fiber analysis. Powder Technology, 377: 974-991. https://doi.org/10.1016/j.powtec.2020.08.034

[10] Wei, Y., Chen, Z., Zhao, C., Chen, X., He, J., Zhang, C. (2023). A three-stage multi-objective heterogeneous integrated model with decomposition-reconstruction mechanism and adaptive segmentation error correction method for ship motion multi-step prediction. Advanced Engineering Informatics, 56: 101954. https://doi.org/10.1016/j.aei.2023.101954

[11] Wei, Y., Chen, Z., Zhao, C., Chen, X. (2023). Deterministic ship roll forecasting model based on multi-objective data fusion and multi-layer error correction. Applied Soft Computing, 132: 109915. https://doi.org/10.1016/j.asoc.2022.109915

[12] Wei, Y., Chen, Z., Zhao, C., Tu, Y., Chen, X., Yang, R. (2022). An ensemble multi-step forecasting model for ship roll motion under different external conditions: A case study on the South China Sea. Measurement, 201: 111679. https://doi.org/10.1016/j.measurement.2022.111679

[13] Gregory, U., Ren, L. (2019). Intent prediction of multi-axial ankle motion using limited EMG signals. Frontiers in Bioengineering and Biotechnology, 7: 335. https://doi.org/10.3389/fbioe.2019.00335

[14] Zhong, J., Ye, C., Cao, W., Wang, H. (2024). Parallel multi-stage rectification networks for 3D skeleton-based motion prediction. Scientific Reports, 14(1): 26058. https://doi.org/10.1038/s41598-024-75782-7

[15] Li, M.S., Chen, M.J., Yeh, C.H., Tai, K.H. (2015). Performance improvement of multi-view video coding based on geometric prediction and human visual system. International Journal of Imaging Systems and Technology, 25(1): 41-49. https://doi.org/10.1002/ima.22119

[16] Zhang, L., Long, Z., Cai, J., Luo, F., Fang, J., Wang, M. Y. (2015). Multi-objective optimization design of a connection frame in macro–micro motion platform. Applied Soft Computing, 32, 369-382. https://doi.org/10.1016/j.asoc.2015.03.044

[17] Wang, Z., Yan, Y., Zeng, X., Li, R., Cui, W., Liang, Y., Fan, D. (2024). Joint multi-objective optimization based on multitask and multi-fidelity Gaussian processes for flapping foil. Ocean Engineering, 294: 116862. https://doi.org/10.1016/j.oceaneng.2024.116862

[18] Shi, W., Guo, Z., Chen, M., Li, S., Hu, J., Dai, Z. (2025). Multi-step prediction of ship heave motion using transformer-enhanced multi-scale CNN. Measurement, 242: 115787. https://doi.org/10.1016/j.measurement.2024.115787

[19] Alexopoulos, K., Mavrikios, D., Pappas, M., Ntelis, E., Chryssolouris, G. (2007). Multi-criteria upper-body human motion adaptation. International Journal of Computer Integrated Manufacturing, 20(1): 57-70. https://doi.org/10.1080/09511920500233749

[20] An, Y., Wu, J., Cui, Y., Hu, H. (2023). Multi-object tracking based on a novel feature image with multi-modal information. IEEE Transactions on Vehicular Technology, 72(8): 9909-9921. https://doi.org/10.1109/TVT.2023.3259999

[21] Vo, G., Zakharov, D., Park, C. (2021). Data association algorithm for large-scale multi-object tracking with complex interactions. Journal of Electronic Imaging, 30(6): 063021-063021. https://doi.org/10.1117/1.JEI.30.6.063021

[22] Bryant, D.S., Vo, B. T., Vo, B.N., Jones, B.A. (2018). A generalized labeled multi-Bernoulli filter with object spawning. IEEE Transactions on Signal Processing, 66(23): 6177-6189. https://doi.org/10.1109/TSP.2018.2872856

[23] Dang, Z., Sun, X., Sun, B., Guo, R., Li, C. (2024). OMCTrack: Integrating occlusion perception and motion compensation for UAV multi-object tracking. Drones, 8(9): 480. https://doi.org/10.3390/drones8090480