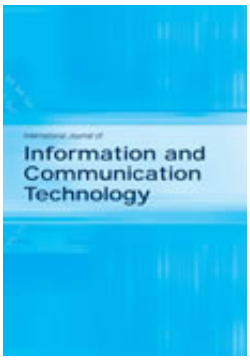


International Journal of Information and Communication Technology

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

A hybrid SE-YOLOv5 and GNN-SGCM approach for intelligent table tennis ball positioning and 3D recognition

Qizhou Gu

DOI: 10.1504/IJICT.2026.10075877

Article History:

Received:	06 August 2025
Last revised:	01 November 2025
Accepted:	18 November 2025
Published online:	04 February 2026

A hybrid SE-YOLOv5 and GNN-SGCM approach for intelligent table tennis ball positioning and 3D recognition

Qizhou Gu

Sports Department,
Nanjing Vocational Institute of Railway Technology,
Nanjing, 210000, China
Email: gqz_007@outlook.com

Abstract: At present, all the positioning methods for the trajectory of table tennis have limitations such as low accuracy and large deviation. Therefore, this study utilises the extrusion and excitation network to optimise YOLOv5, introduces the graph convolutional network to improve the hybrid algorithm of semi-global matching and census transformation, and combines the two to construct an intelligent positioning and recognition model for table tennis. The results show that the research model has an accuracy rate of 97.6%, a precision rate of 98.6%, a recall rate of 96.8%, and a specificity of 97.2%. The average error of the recall rate is 0.59%, and the overlap degree of trajectory positioning is 0.93. In conclusion, the research model not only ensures the reliability of the table tennis positioning and recognition results, but also improves the recognition efficiency and result quality, making significant contributions to the development of table tennis.

Keywords: YOLOv5; squeeze-and-excitation network; semi-global matching and census; SGCM; graph neural network; GNN; positioning and recognition; ping-pong balls.

Reference to this paper should be made as follows: Gu, Q. (2026) 'A hybrid SE-YOLOv5 and GNN-SGCM approach for intelligent table tennis ball positioning and 3D recognition', *Int. J. Information and Communication Technology*, Vol. 27, No. 5, pp.16–37.

Biographical notes: Qizhou Gu obtained his PhD in 2023 from Adamson University. Presently, he is working as an Associate Professor in the Sports Department, Nanjing Vocational Institute of Railway Technology. He mainly engaged in research on physical health promotion. He has published over ten academic papers.

1 Introduction

Ping-pong, as a global sport, not only promotes physical health and facilitates cultural exchange but also features characteristics such as high speed and strong spin (Wang et al., 2023). In sports competitions, the trajectory of a ping-pong ball is complex and varied, often leading to issues like motion blur and occlusion. Accurate localisation and recognition are essential for referees to make effective judgments on players' actions and the ball's trajectory. Therefore, research on localisation and recognition is of great

significance (Shi et al., 2024; Zaidi et al., 2023). Localisation and recognition aim to utilise advanced computer and image processing technologies to detect and analyse the position and state of an object, ultimately determining its complete motion trajectory. The core technologies involved include object detection, feature extraction, and motion tracking, which also constitute the major challenges in this field (Hu et al., 2023). As a result, numerous scholars worldwide have conducted research on this topic. For instance, the Tian team, in order to address the problem of existing algorithms being unable to precisely identify the movement position of human bodies, proposed a multidimensional information recognition algorithm for human targets based on CSI decomposition. This algorithm used linear discriminant analysis to achieve precise recognition of human body movement positions (Tian et al., 2023). To tackle the poor performance of existing methods in identifying faults and their positions in photovoltaic arrays, Sakthivel et al. (2023) proposed a PV fault experiment proposed method. By analysing module voltage, they could compute fault types and locations, showing that their approach could accurately identify fault locations and types (Sakthivel et al., 2023). To better recognise patients' activity trajectories at home, A. Leal-Junior's team introduced a sensor system based on feedforward neural networks. Through projection-reflection analysis, they detected patients' movements, with experimental results indicating that the system could accurately locate the patient's position and provide gait analysis (Leal-Junior et al., 2023). In response to the challenge of traditional methods being unable to perform localisation in DC grid systems, Rao and Jena (2023) proposed a fault detection scheme based on DC energy differences. This method used least squares to calculate the fault point's location and proved effective in distinguishing between internal and external faults, ensuring accurate localisation of the faults (Rao and Jena, 2023).

Although certain achievements have been made in the research of positioning and recognition at present, the traditional positioning and recognition methods generally have the following problems, which are difficult to meet the positioning requirements of small size and high speed of table tennis balls (Hu et al., 2024). One of the issues is the insufficient recognition accuracy. Traditional template matching algorithms are extremely sensitive to changes such as the posture of the ping-pong ball and the intensity of the light. In competition scenarios, matching deviations are prone to occur, resulting in inaccurate positioning results (Annaby and Fouda, 2023). Second, it has poor real-time performance. For instance, the computational complexity of the scale-invariant feature transformation algorithm is relatively high. In scenarios where table tennis balls move at high speed, it cannot quickly output positioning results, making it difficult to meet the demands of real-time refereeing and trajectory playback in competitions (Langford et al., 2023). Thirdly, it has weak adaptability to complex scenarios. In table tennis matches, situations such as blurred vision and chaotic scenes often occur. Existing algorithms are prone to losing targets in such scenarios, resulting in positioning interruption or increased errors. The YOLOv5 network, a target detection algorithm based on computer vision, is known for its speed and high accuracy and is widely used in fields such as object detection and image recognition (Rajamohanan and Latha, 2023). The semi-global matching and census (SGCM) transform algorithm encodes the local greyscale of motion images to generate specific representations, solving the issue of image matching for moving objects by judging the degree of matching between different representations (Liu et al., 2024). Furthermore, in order to further enhance the performance of the algorithm in table tennis positioning, the squeeze-and-excitation network (SE) and graph neural network (GNN) were introduced in the research to optimise and improve the YOLOv5

network and the SGCM algorithm respectively. By combining both, an intelligent model for ping-pong localisation and recognition is constructed, with the goal of solving the problems of low recognition accuracy and large deviation in traditional methods. The research aims to address the problems of low accuracy, large result deviation and weak adaptability to complex scenarios of traditional methods by constructing models, providing technical support for table tennis event adjudication, training analysis, etc.

The innovation of the research lies in the following three points. The first one is that, in view of the small target and high dynamic characteristics of table tennis, the channel attention mechanism of YOLOv5 is strengthened through the SE module to increase the weight of the network on key features such as the outline and movement trajectory of the table tennis. And the node association relationship of image features is constructed by using GNN, so that SGCM can still accurately match features when the local features of table tennis balls change. Secondly, a dual-module collaborative architecture for target detection and stereo matching was studied and constructed. With the help of the SE-YOLOv5 module, the real-time position of the table tennis ball was quickly located, providing precise initial coordinates for subsequent motion matching. Then, the three-dimensional scene diagram of table tennis movement is constructed with the GNN-SGCM module to achieve the continuous tracking of the movement trajectory. Thirdly, in response to the actual scene issues such as lighting changes, limb occlusion, and differences among multi-brand balls in table tennis matches, targeted optimisations were carried out during the model construction. For instance, the research enhances the model's robustness under different lighting conditions through the adaptive feature weighting of the SE module, and reduces the impact of limb occlusion on positioning by leveraging the global relationship modelling capability of GNN.

The study is divided into four main parts. The first part introduces the research background and relevant literature, analysing the current state of research on localisation and recognition. At the same time, study proposes a ping-pong localisation and recognition model based on the SE-YOLOv5 network and GNN-SGCM algorithm. The second part details the advantages of the SE-YOLOv5 and GNN-SGCM algorithms in target detection and image recognition, while also highlighting the benefits of the proposed model compared to others, and outlines its process for ping-pong localisation. The third part validates the performance of SE-YOLOv5 and GNN-SGCM through comparative experiments and evaluates the effect of the proposed model in practical applications. The fourth part discusses and summarises the experimental data and results, while also exploring potential future research directions.

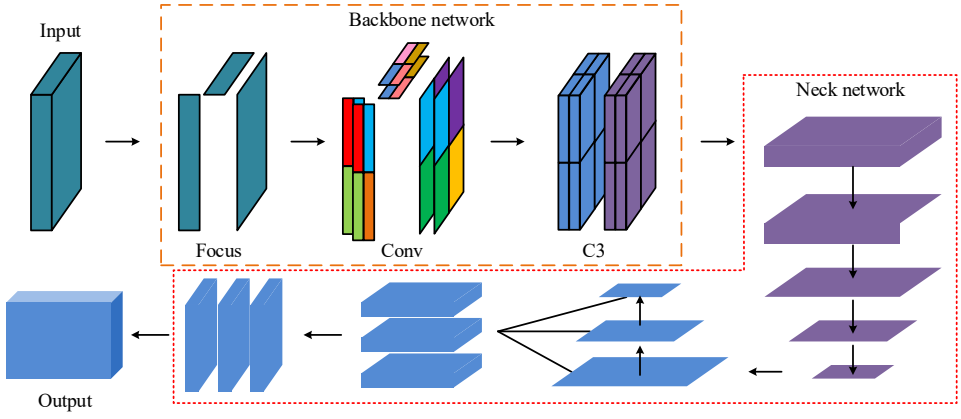
2 Methods and materials

2.1 Target detection algorithm design based on SE-optimised YOLOv5 network

Ping-pong localisation and recognition require both high accuracy and real-time performance, and during the actual process, the system must handle various motion states and the interference from complex backgrounds (Li and Wu, 2023). YOLOv5, as an optimised version of the YOLO network, possesses timely, efficient, and accurate object detection capabilities. With its well-designed network architecture and reasonable training strategies, YOLOv5 significantly enhances multi-scale object detection performance (Ghose et al., 2023; Saidani, 2023). During the iterative process, feature

reuse effectively reduces computational complexity, while various optimisation techniques are introduced during training to enhance the network's generalisation ability and convergence speed (Han et al., 2023). The specific structure of YOLOv5 is shown in Figure 1.

Figure 1 YOLOv5 structure diagram (see online version for colours)



As shown in Figure 1, YOLOv5 mainly consists of four parts: the input, backbone network, neck network, and output. When the input data enters the backbone network, it is first split by the Focus module. Then, the data undergo convolution stacking in the Conv module and feature maps are formed in the C3 module. Afterward, the data passes through the neck network, where the FPN-PAN structure further extracts local information from the feature map. Iteratively, multi-scale feature maps with feature information are generated, and they are combined and output as the final result. The convolution operation in the Conv module is shown in equation (1).

$$Y = N \cdot X_{M,N} \cdot W + b \tag{1}$$

In equation (1), Y represents the output feature map, X denotes the input feature map, M and N are the sizes of the convolution kernels, W represents the weights, b is the bias term, and N is the number of channels in the input feature map. The Conv module not only performs convolution on the input feature map but also alleviates gradient vanishing and prevents output overfitting through batch normalisation (Chitraningrum et al., 2024). The specific process of normalisation is shown in equation (2).

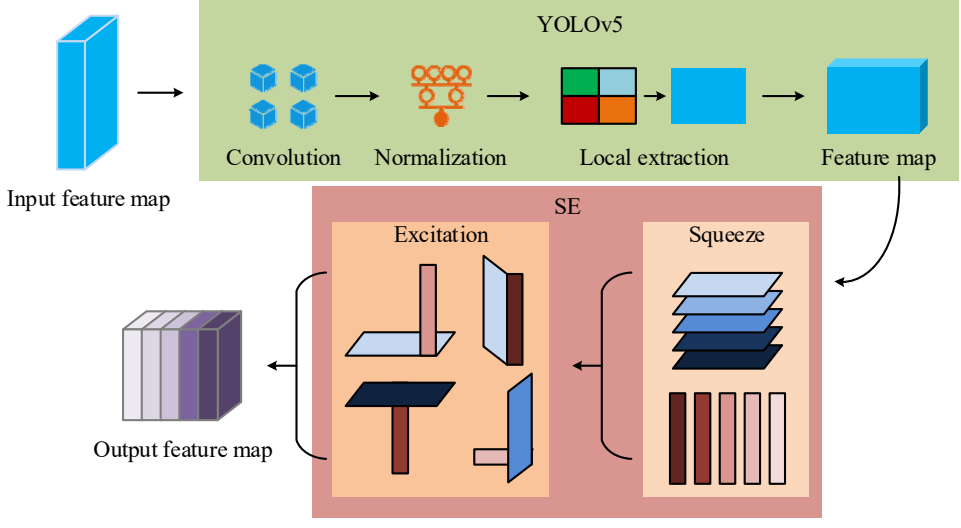
$$\begin{cases} x_i = \dfrac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}} \\ y_i = \gamma x_i + \beta \end{cases} \tag{2}$$

In equation (2), x_i and y_i denote the normalisation of the data in two branches, μ is the mean of the current data, σ^2 is the variance of the current data, ε is a constant for numerical stability, and γ and β are hyperparameters. Throughout the target data detection process, YOLOv5 uses its loss function to constrain data processing and prevent significant result deviations (Liang et al., 2023). The loss function is expressed in equation (3).

$$L = \lambda_{cls} L_{cls} + \lambda_{obj} L_{obj} \tag{3}$$

In equation (3), L_{cls} defines the classification loss function, λ_{cls} is its weight coefficient, L_{obj} represents the object confidence loss function, and λ_{obj} is its weight coefficient. However, a single YOLOv5 network faces issues such as weak motion blur handling and poor adaptability to complex environments when performing ping-pong localisation and recognition, requiring optimisation. SE, as an optimisation algorithm, can adjust the weights of the data channels, emphasising local feature information while suppressing unimportant features to improve the network's ability to extract feature maps (Phan et al., 2023; Liu et al., 2025). Therefore, the study introduces SE to optimise the YOLOv5 network, resulting in the SE-YOLOv5 hybrid algorithm, as shown in Figure 2.

Figure 2 SE-YOLOv5 hybrid algorithm flowchart Figure 1 YOLOv5 structure diagram (see online version for colours)



As shown in Figure 2, in the SE-YOLOv5 hybrid algorithm, SE primarily performs secondary processing on the feature maps generated by YOLOv5. It adjusts the network structure by weighting the channel attention of the feature maps and enables the network to learn the weighted features. Then, SE adaptively adjusts the response strength of the weighted features within the channels, ultimately enhancing the detail representation of the feature maps and outputting the results. The function for weighting channel attention is expressed in equation (4).

$$z_c = \frac{1}{H \times w} \sum_{m=1}^{H} \sum_{n=1}^{w} U_c(m, n) \quad c = 1, 2, \cdots, c \tag{4}$$

In equation (4), z defines the channel attention descriptor, $U_c(m, n)$ represents the value at position U in the c^{th} channel of feature map (m, n), and H and w are the weighting constants. Then, SE learns the weighted features across channels through a fully connected layer, generating weights within each channel. This process is shown in equation (5) (Ruifeng et al., 2024).

$$g = \mathrm{Re}\,LU\,(Kz + b') \tag{5}$$

In equation (5), K is the weight matrix, with values ranging from [0, 1], b' is the bias vector, and g defines the generated weights.

2.2 Binocular stereo matching algorithm based on GNN-improved SGCM

Although SE-YOLOv5 can quickly lock the two-dimensional position of the ping-pong ball in a single frame image, it helps solve the problems of insufficient accuracy and poor real-time performance of traditional algorithms in small target detection. However, the movement of table tennis has the characteristics of three-dimensional space and trajectory correlation. When using SE-YOLOv5 only for positioning and recognition, the following limitations exist. One is that the two-dimensional coordinates output cannot reflect the information of the sphere in three-dimensional space, which is likely to cause confusion in the positioning of table tennis balls at different distances on the court. Secondly, the single-frame detection results lack continuous correlation with the motion trajectory, making it difficult to handle the problem of target matching interruption in scenarios such as high-speed ball movement and athlete limb occlusion. Therefore, it is necessary to combine the stereo matching algorithm. SGCM, as a hybrid algorithm that combines SGM and Census transformations, possesses strong anti-interference ability and can better adapt to various complex scenarios. Its powerful feature matching ability improves computational efficiency to some extent, making it suitable for detecting the motion of small objects (Singh and Adhikari, 2025; Yao et al., 2025). The architecture of SGCM is shown in Figure 3.

As shown in Figure 3, SGCM is divided into two parts: SGM and census. First, SGM matches image features with actual objects by optimising cost paths in multiple directions to find the best matching results, while capturing long-range correlations in the image to improve result quality. Then, Census performs a mapping transformation on the best matching results, turning abstract image features into clear greyscale values, which are then integrated and output. The process of image feature matching by SGM is shown in equation (6).

$$C = \alpha \times \left(C_l\left(x', y'\right), C_r\left(x' - d, y'\right)\right) \tag{6}$$

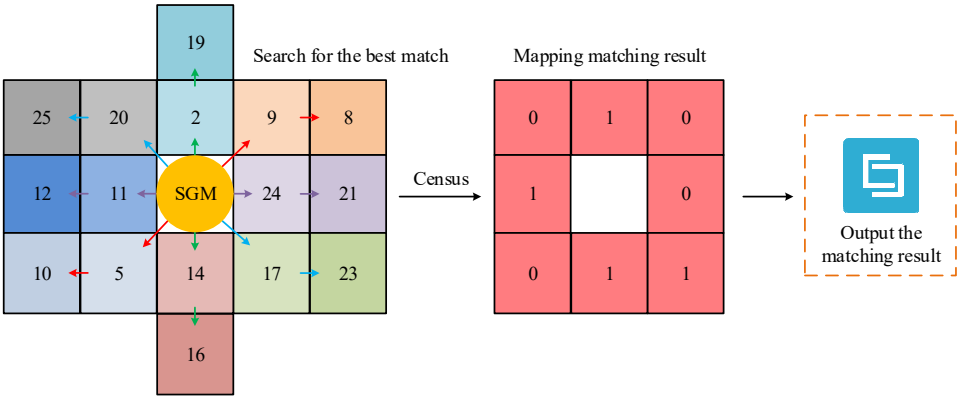
In equation (6), α defines the Hamming distance used to optimise the path for finding the optimal match, $C_l(x', y')$ represents the left image feature at pixel (x', y'), and $C_r(x' - d, y')$ is the right image feature with disparity. The specific expression of the Hamming distance is shown in equation (7).

$$\alpha(A, B) = \sum_{i'}^{N'} \left(A_{i'} \oplus B_{i'}\right) \tag{7}$$

In equation (7), A and B represent two binary numbers, $A_{i'}$ and $B_{i'}$ are their i'^{th} bit digits, N' is the number of binary digits, and $\oplus$ represents the operator. Optimising the path for finding the best match using the Hamming distance significantly improves the convergence speed, as shown in equation (8).

$$S = \sum_{r'=1}^{R} L(r', d) \tag{8}$$

Figure 3 SGCM structure diagram Figure 1 YOLOv5 structure diagram (see online version for colours)



In equation (8), R represents the total number of directions, and $L(r', d)$ defines the disparity between different directions. Census transformation, as a local image feature description method, reduces complexity by mapping feature images to greyscale values, which in turn improves matching accuracy (Oddo et al., 2024). The process of census transformation is shown in equation (9).

$$C'_{m',n'} = \sum_{(o,p)\in(m',n')} a_{o,p} \times 2^{e} \tag{9}$$

In equation (9), (m', n') represents the centre pixel position of the image feature, (o, p) denotes the pixel positions of other image features in the neighbourhood, and a defines the matching result. However, a single SGCM has weak adaptability in special situations. For example, when the ping-pong ball rotates rapidly, the texture and features of the image change continuously, affecting the matching results. Therefore, further optimisation methods are needed. GNN is a neural network designed for processing image structural data. The core idea is to pass information between the nodes of image features to learn the representations of each node. This approach allows GNN to capture the complex relationships and dependencies between nodes (Preethi and Mamatha, 2023; Kunal et al., 2023). Therefore, the study introduces GNN to improve SGCM, resulting in the GNN-SGCM hybrid algorithm. The specific process is shown in Figure 4.

As shown in Figure 4, in the GNN-SGCM hybrid algorithm, SGCM preprocesses the data first, and then GNN transforms the image processed by SGCM into a specific structure. Based on the spatial distance, structural similarity, colour, and other information between structures, GNN constructs a related model. The model then learns the feature representations of each structure, ultimately generating results with all the prominent features of the structure. The process of transforming the image into a graph structure by GNN is shown in equation (10).

$$h_{j'}^{0} = \left[d_{j'}, g'_{j'} \right] \tag{10}$$

In equation (10), $h_{j'}^{0}$ defines the initial graph structure, $d_{j'}$ represents the disparity value of the j'^{th} image, and $g'_{j'}$ is the greyscale value of the j'^{th} pixel. When constructing the related model, GNN constantly updates the model parameters by passing information

between the graph structures. The information transmission function is shown in equation (11).

$$q_t = \sum_{v \in N^*} f\left(h^{t-1}\right) \tag{11}$$

In equation (11), f represents the nonlinear function, h^{t-1} is the graph structure at time $t-1$, q_t is the graph structure information transmitted at time t, v is the node in the graph structure, and N^* is the feature set of the neighbouring nodes. After receiving the information, each node updates its features according to its state. The mathematical formula for this update is shown in equation (12).

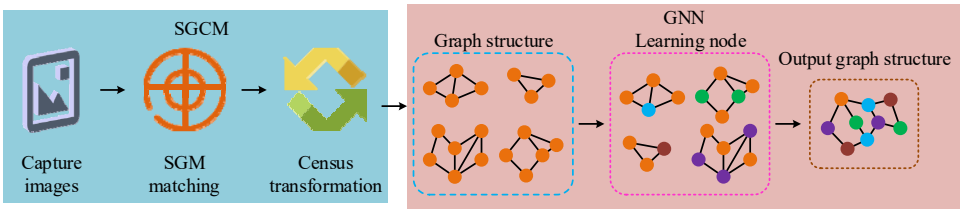
$$h^t = \eta\left(v^t \times W' \cdot h^{t-1}\right) \tag{12}$$

In equation (12), W' defines the weight matrix of the learning process, with values ranging from [0, 1], and η represents the activation function. By continuously updating the node features, a final image feature is obtained, which can optimise the disparity values in the process of converting the image to a graph structure. The optimisation process is shown in equation (13).

$$d^* = \delta\left(h^{T'}\right) \tag{13}$$

In equation (13), δ represents the specific regression function, d^* is the disparity value optimised using the final node features, and T' is the number of iterations.

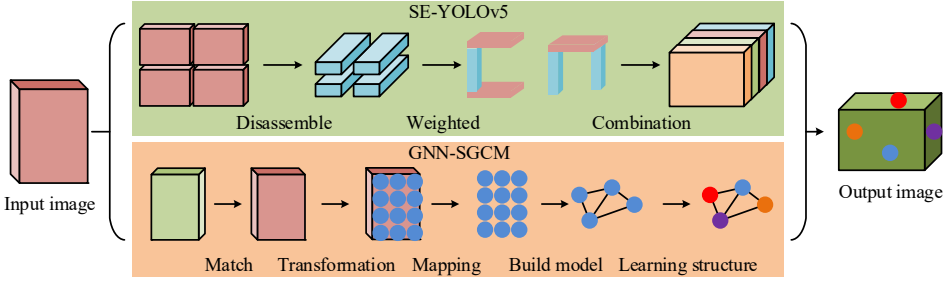
Figure 4 GNN-SGCM hybrid algorithm flowchart (see online version for colours)



2.3 *Construction of ping-pong localisation and recognition model combining SE-YOLOv5 and GNN-SGCM*

Compared to the traditional YOLOv5 network, SE-YOLOv5 not only adapts the importance of feature channels through the SE module, but also retains the fast detection characteristics of YOLOv5, meeting the real-time recognition requirements. At the same time, it demonstrates strong generalisation ability in various conditions and complex dynamic environments. On the other hand, GNN-SGCM also has unique advantages in localisation and recognition. By transforming image features into a graph structure, GNN-SGCM uses its powerful relationship modelling ability to capture the spatial relationships between objects and environmental elements. Even in the presence of interference and cluttered backgrounds, it can accurately infer the object's position using the scene graph information for precise localisation. Based on this, the study combines SE-YOLOv5 and GNN-SGCM to achieve a more comprehensive localisation and recognition of ping-pong balls. The structure of the hybrid algorithm is shown in Figure 5.

Figure 5 Diagram of the hybrid algorithm combining SE-YOLOv5 and GNN-SGCM (see online version for colours)



As shown in Figure 5, the hybrid algorithm performs localisation and recognition by processing the input image through two operation units, SE-YOLOv5 and GNN-SGCM. SE-YOLOv5 performs operations such as splitting, weighting, and combining the image, leveraging its efficient feature extraction and fast object detection capabilities to achieve high-quality recognition results. Meanwhile, GNN-SGCM uses its powerful relationship modelling ability and strong adaptability to complex scenes to perform operations such as matching, mapping, and learning, yielding positioning results with prominent feature information. The image processing by SE-YOLOv5 is assisted by the activation function, as expressed in equation (14).

$$f_{sy} = x^* \cdot \frac{1}{1+e^{-x^*}} \tag{14}$$

In equation (14), f_{sy} defines the activation function of the SE-YOLOv5 network, specifically the Sigmoid function, and x^* represents the input image. In the GNN-SGCM process of transforming the image into a graph structure, the final disparity of each image feature needs to be calculated. The specific calculation process is shown in equation (15).

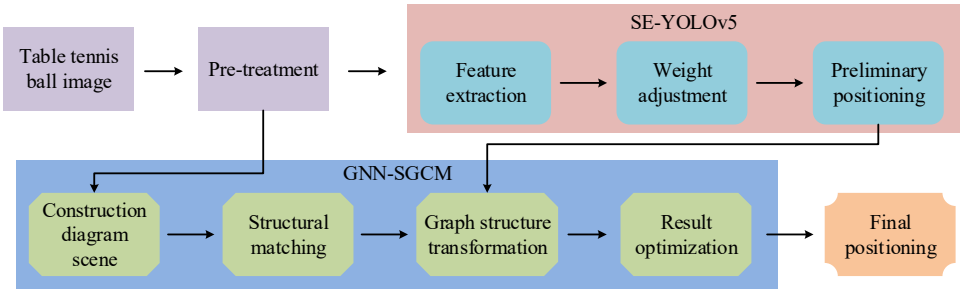
$$D\left(m^*, n^*\right) = \arg\min_d S^*\left(m^*, n^*, d^*\right) \tag{15}$$

In equation (15), $D(m^*, n^*)$ represents the final disparity value of pixel (m^*, n^*) after transforming into a graph structure, and $S^*(m^*, n^*, d^*)$ is the accumulated cost during the transformation process. Compared to the individual SE-YOLOv5 and GNN-SGCM, the hybrid algorithm combines efficient feature extraction and fast object detection capabilities, while also considering dynamic and complex environments, facilitating the rapid localisation of ping-pong balls. Based on this, the study constructs an intelligent model suitable for ping-pong ball localisation and recognition, as shown in the localisation process diagram in Figure 6.

As shown in Figure 6, when the proposed model performs ping-pong ball localisation and recognition, the first step is to preprocess the captured image, including resizing and data deduplication. Then, the backbone network of the YOLOv5 network is used for feature extraction, and the SE module enhances image features while suppressing irrelevant interference. Meanwhile, GNN-SGCM generates preliminary boundary locations and confidence scores, and constructs a graph scene based on image elements. Subsequently, SE-YOLOv5 generates initial localisation results, which are input into the GNN-SGCM operation unit to match the image features and transform them into a

specific graph structure. Finally, the model performs relationship reasoning between the ping-pong ball and surrounding objects, the environment, etc., to optimise the preliminary localisation results by excluding errors that do not fit the logic or scene, ultimately outputting precise ping-pong ball localisation results.

Figure 6 Flowchart of the proposed model for ping-pong ball positioning and recognition (see online version for colours)

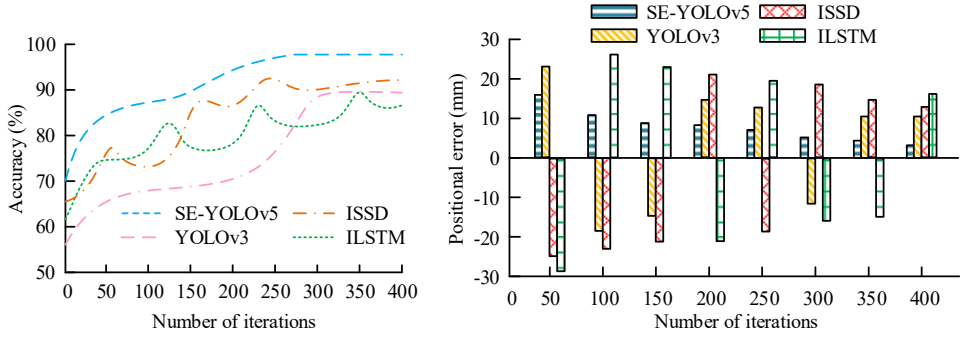


3 Results

3.1 Performance validation of SE-YOLOv5 for ping-pong motion path detection

To verify the performance of SE-YOLOv5, the study compared it with the improved single shot multibox detector (ISSD) algorithm, improved long short-term memory (ILSTM), and YOLOv3 network. And four different line types and ICONS are adopted for distinction. The software and hardware framework of the experimental environment and the key training parameters of the algorithm are shown in Table 1.

Figure 7 Comparison of localisation accuracy and position error, (a) comparison of the accuracy rates of each algorithm in locating table tennis balls (b) comparison of the position errors of each algorithm in locating table tennis balls (see online version for colours)



(a) (b)

The experimental dataset was rotated by OpenTTGames, which includes over 38,000 table tennis training samples, covering ping-pong ball position coordinates, semantic segmentation templates such as figures and scoreboards, and other labelled content,

making it suitable for ball detection tasks. The study first conducted comparison experiments on the accuracy and position error of ping-pong ball localisation by the four algorithms. The results are shown in Figure 7.

Table 1 Information table of software and hardware configuration and algorithm hyperparameters

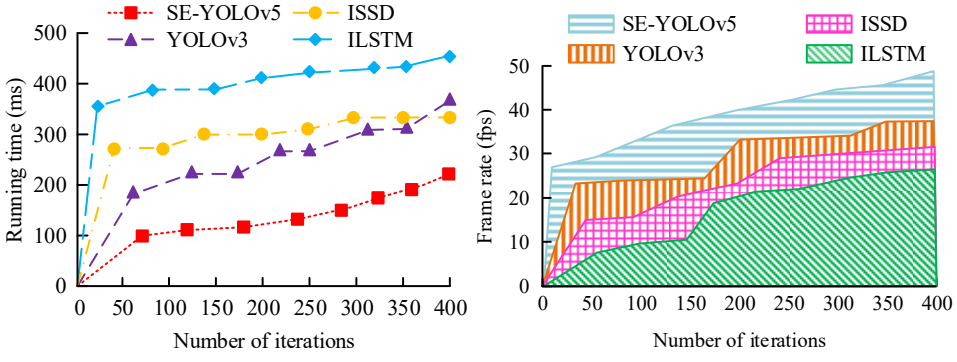
Category	*Specific information*
Hardware configuration	CPU: Intel Core i7-14600KF
	GPU: NVIDIA RTX 5060Ti 16 GB
	Internal memory: 64 GB DDR5 6,400 MHz
	Storage: 2 TB SSD and 2 TB Mechanical hard disk
	Operating system: Windows 11 Professional
Software framework	Deep learning framework: PyTorch 2.1.0
	Image processing library: OpenCV 4.8.1
	Data annotation tool: LabelImg 1.8.6
	Visualisation tool: Matplotlib 3.8.2
	Programming language: Python 3.8
Algorithm hyperparameters	Learning rate:0.001
	Batch size: 16
	Optimiser: AdamW
	Epochs:300
	Loss function: CIoU loss

As shown in Figure 7(a), SE-YOLOv5 achieved the highest localisation accuracy of 97.6% at the 268th iteration, which is significantly higher than ISSD at 91.2%, YOLOv3 at 88.4%, and ILSTM at 85.9%. Its average accuracy was 90.3%, also outperforming the average accuracy of the three comparison algorithms. Additionally, SE-YOLOv5's accuracy curve rises quickly with only two increases in value, without the sharp fluctuations seen in ISSD and ILSTM. Although YOLOv3's accuracy also fluctuates relatively little, the increase is slow and much lower than SE-YOLOv5's accuracy. From Figure 7(b), it can be observed that SE-YOLOv5's localisation results are consistently on the same side as the actual position, and the error rapidly decreases as the number of iterations increases. Its minimum position error was 2.7 mm, significantly lower than YOLOv3's 9.8 mm, ISSD's 12.2 mm, and ILSTM's 15.3 mm, indicating that SE-YOLOv5's localisation results are more aligned with the actual position of the ping-pong ball. The study then compared the localisation efficiency, and the experimental results are shown in Figure 8.

In Figure 8(a), SE-YOLOv5 required a maximum time of only 207 ms for ping-pong ball localisation, significantly lower than ISSD at 317 ms, YOLOv3 at 338 ms, and ILSTM at 446ms. This indicates that SE-YOLOv5 has a well-structured design that efficiently utilises computational resources, and the algorithm can provide timely position information for ping-pong balls in scenarios with high real-time requirements. As shown in Figure 8(b), SE-YOLOv5 achieved a maximum frame rate of 48FPS for ping-pong ball localisation, which is significantly higher than YOLOv3 at 32FPS, ISSD at 29FPS, and ILSTM at 25FPS. Furthermore, SE-YOLOv5's frame rate showed the smallest fluctuation and followed an approximately linear relationship. Overall, SE-YOLOv5 not

only demonstrated higher accuracy and lower error in ping-pong ball localisation but also showed strong efficiency, enabling smooth and timely tracking of the ball's movement.

Figure 8 Comparison of ping-pong ball localisation efficiency, (a) comparison of the time required for different algorithms to locate tennis balls (b) comparison of frame rates of tennis ball positioning results by different algorithms (see online version for colours)



(a) (b)

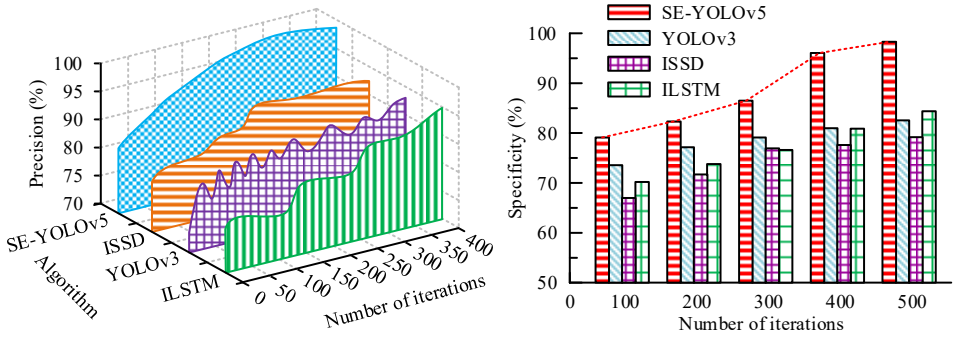
In addition, the study also analysed the practical application of SE-YOLOv5. The research selected the standard balls for international competitions and conducted actual tests in two real scenarios: amateur and professional venues. Among them, the lighting in the amateur competition venue was uneven, and there was also a situation where athletes' limbs were blocked. The study embedded the SE-YOLOv5 into the live event system for continuous detection of ten matches. The results showed that its average accuracy rate in positioning table tennis balls was 89.24%. When facing occluding scenes, its accuracy rate only dropped to 89.93%, and the frame rate was stable at 40–45 FPS. It can accurately output the trajectory of the ball to assist referees in judging balls that are out of bounds or on the edge. However, in professional training venues, there are scenes of strong light interference and multiple balls moving simultaneously, and the rotation intensity of the balls is high. The study integrated SE-YOLOv5 into the training analysis system to provide trajectory playback and landing point statistics for athletes. The measured results show that the accuracy rate of SE-YOLOv5 in distinguishing multiple balls is as high as 87.63%, and the positioning error in high-spin scenarios is less than 4.2 mm. It can accurately count the landing point distribution of table tennis balls to assist coaches in adjusting training plans.

3.2 Performance evaluation of GNN-SGCM for ping-pong stereo motion recognition

To verify the performance of GNN-SGCM, the study compared it with the YOLOv3 network, ISSD, and ILSTM. To ensure the accuracy and reliability of the experimental results, the experimental environment and specific parameters were kept unchanged, with only the dataset being switched to the PaddlePaddle-based temporal action localisation dataset. This dataset mainly involves videos of athletes performing the action of hitting a ping-pong ball, designed specifically for instantaneous action localisation. The study first

conducted comparison experiments on the precision and specificity of the four algorithms in recognising the movement of the ping-pong ball. The results are shown in Figure 9.

Figure 9 Comparison of recognition precision and specificity, (a) comparison of the accuracy rates of four algorithms in locating and recognising table tennis balls (b) comparison of the specificity of four algorithms in locating and recognising table tennis balls (see online version for colours)



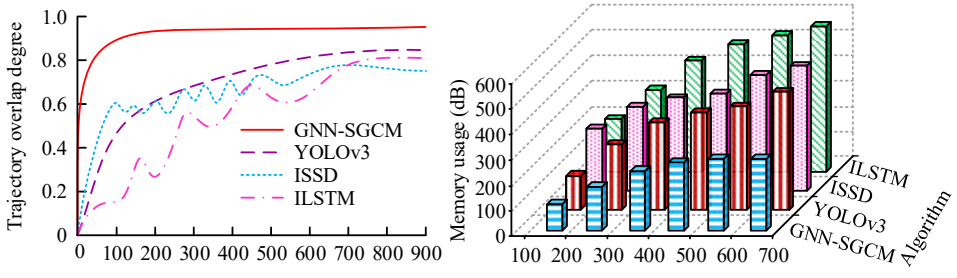
(a) (b)

As shown in Figure 9(a), the precision of GNN-SGCM in recognising the movement of the ping-pong ball reached its maximum value of 98.6% at the 293rd iteration, after which it stabilised and remained unchanged. This is significantly higher than ISSD's 92.3%, YOLOv3's 91.9%, and ILSTM's 95.1%. Furthermore, the precision curve of GNN-SGCM increased smoothly without any obvious fluctuations, whereas YOLOv3's precision curve experienced significant variations during the first 250 iterations, with the other two comparison algorithms also showing noticeable fluctuations. In Figure 9(b), the specificity of all four algorithms increased significantly during the first 300 iterations. However, after 300 iterations, only GNN-SGCM continued to show a significant increase in specificity, reaching a maximum of 97.2%, which is much higher than YOLOv3's 82.5%, ISSD's 78.6%, and ILSTM's 84.1%. This indicates that GNN-SGCM is better at distinguishing non-ping-pong ball objects when recognising the movement positions of objects. The study then compared the trajectory overlap and memory usage of each algorithm in recognising the position of the ping-pong ball, with the experimental results shown in Figure 10.

As shown in Figure 10(a), GNN-SGCM's trajectory overlap reached its maximum value of 0.93 quickly before the 200th iteration, which is significantly higher than YOLOv3's 0.84, ISSD's 0.77, and ILSTM's 0.81. Moreover, GNN-SGCM maintained a higher trajectory overlap than the other three comparison algorithms throughout the process and showed the fastest and smoothest increasing trend as the number of iterations increased, without frequent fluctuations. In Figure 10(b), GNN-SGCM's maximum memory usage during trajectory recognition was only 223 dB, which is significantly lower than YOLOv3's 346dB, ISSD's 403dB, and ILSTM's 518 dB. Additionally, its memory usage showed a nearly linear increase before 450 iterations, after which it gradually stabilised without any significant changes. In summary, using GNN-SGCM to track the movement trajectory of the ping-pong ball not only achieved high recall, specificity, and trajectory overlap but also demonstrated low computational resource requirements. This suggests that GNN-SGCM can effectively capture the spatial

information of the ping-pong ball and simulate its movement trajectory, ultimately completing the localisation and recognition.

Figure 10 Comparison of trajectory overlap and memory usage, (a) comparison of the overlap degree of the table tennis ball trajectories identifies by each algorithm (b) comparison of the memory usage of each algorithm for locating and identifying table tennis balls (see online version for colours)



(a) (b)

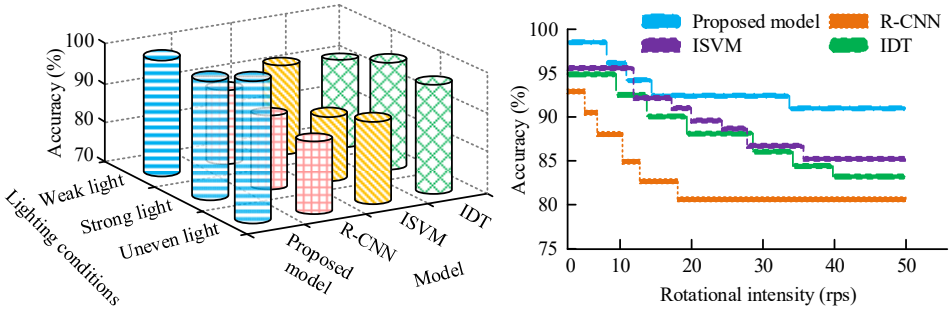
3.3 Application effect evaluation of ping-pong localisation and recognition model based on SE-YOLOv5 and GNN-SGCM

After verifying the performance of SE-YOLOv5 and GNN-SGCM, the study conducted field tests to prove the feasibility and superiority of the proposed model. The study selected the Hongshuangxi competition-grade three-star D40+ ping-pong ball as the experimental object and invited two professional table tennis players to cooperate in the experiment. First, the study compared the anti-interference ability of the proposed model with that of the improved support vector machine (ISVM) model, improved decision tree (IDT) model, and region convolutional neural network (R-CNN) model for ping-pong ball localisation recognition. The results are shown in Figure 11.

As shown in Figure 11(a), the proposed model achieved the highest precision of 98.7% in locating the ping-pong ball under uneven lighting, which is significantly higher than R-CNN's maximum precision of 87.3%, ISVM's maximum precision of 89.8%, and IDT's maximum precision of 91.0%. Under weak and strong lighting conditions, the proposed model's precision in locating the ping-pong ball was 95.6% and 92.9%, respectively, with an average precision of 95.7%. These results were better than those of the three comparison models. Additionally, the precision difference in locating and recognising the ping-pong ball under different lighting conditions was small, with no obvious discrepancies. In Figure 11(b), it can be seen that when the rotation intensity was 33 rps, the precision of the proposed model in locating the ping-pong ball decreased to its minimum value of 90.7%, which is still noticeably higher than ISVM's 86.0%, IDT's 83.6%, and R-CNN's 81.9%. Furthermore, the proposed model showed the slowest decrease in precision as it tracked the ping-pong ball, maintaining a higher precision than the other three models throughout the experiment, indicating that the proposed model possesses strong anti-interference capability and higher accuracy when locating and recognising the ping-pong ball. The study then introduced the TTSwing dataset and compared the sensitivity of each model across the OpenTTGames, PaddlePaddle Action

Localisation, and TTSwing datasets to verify the generalisation ability of the proposed model. The results are shown in Table 2.

Figure 11 Comparison of anti-interference ability during localisation, (a) comparison of the accuracy of table tennis balls under different lighting conditions in four models (b) comparison of the accuracy of the four models under different rotational intensities (see online version for colours)



(a) (b)

Table 2 Comparison of sensitivity across different datasets

Dataset	*Number of experiments*	*Sensitivity/%*			
		Proposed model	*R-CNN*	*ISVM*	*IDT*
OpenTTGames	Experiment 5	92.7	85.6	74.7	90.2
	Experiment 10	93.6	89.4	73.5	92.4
	Experiment 15	95.3	88.3	79.2	89.8
	Experiment 20	97.6	89.5	80.4	91.5
	Experiment 25	97.9	90.7	83.1	90.6
Positioning of the flying paddle movement	Experiment 5	90.7	78.9	82.8	73.8
	Experiment 10	92.4	77.2	83.5	79.4
	Experiment 15	94.9	73.4	86.9	85.9
	Experiment 20	95.5	75.6	89.2	86.7
	Experiment 25	96.8	77.3	92.0	88.2
TTSwing	Experiment 5	91.7	82.0	88.4	83.4
	Experiment 10	92.6	85.6	89.5	86.9
	Experiment 15	94.2	86.2	91.3	84.8
	Experiment 20	96.0	90.1	90.8	83.3
	Experiment 25	97.1	87.5	92.7	88.0

As shown in Table 2, the maximum sensitivity of the proposed model in the OpenTTGames, PaddlePaddle Action Localisation, and TTSwing datasets was 97.9%, 96.8%, and 97.1%, respectively. The average sensitivities for these datasets were 95.42%, 94.06%, and 94.32%, all significantly higher than the maximum sensitivity of the three comparison models. Additionally, the sensitivity difference of the proposed model between the OpenTTGames and PaddlePaddle Action Localisation datasets was only 1.36%, the difference between OpenTTGames and TTSwing was 1.1%, and the

difference between PaddlePaddle Action Localisation and TTSwing was as small as 0.26%. In contrast, the sensitivity of the three comparison models varied greatly across the datasets. For instance, the ISVM model achieved a sensitivity of 92.7% in the TTSwing dataset but only 73.5% in the OpenTTGames dataset. These experimental data and comparison results clearly show that the proposed model has strong generalisation ability. The study then randomly added Gaussian white noise to the TTSwing dataset and compared the recall rates and localisation errors of the four models at different noise intensities to verify their robustness. The experimental results are shown in Figure 12.

Figure 12 Comparison of recall rate and localisation error under different noise intensities, (a) comparison of the recall rates of table tennis balls identified by different models under noise interference (b) comparison of positioning errors of table tennis balls by different models under noise interference (see online version for colours)

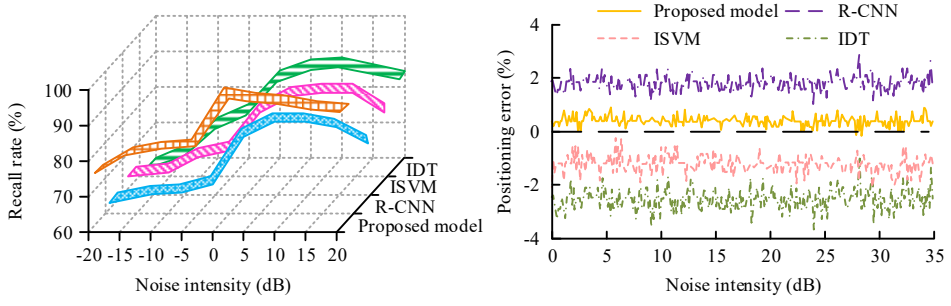
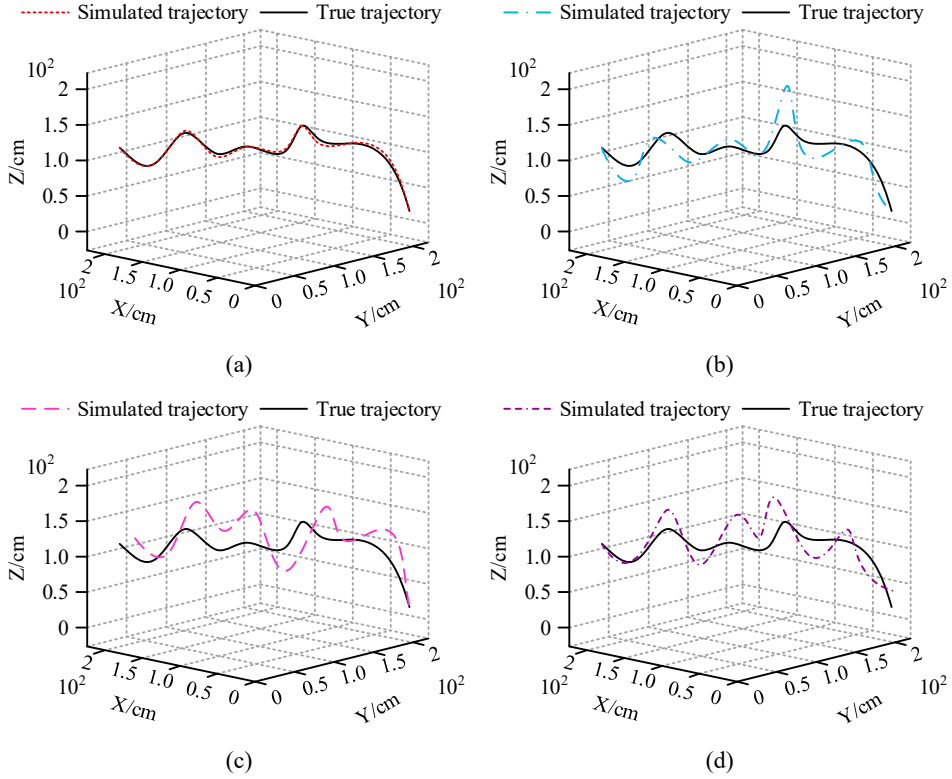


Table 3 Comparison of localisation coverage for tracking ping-pong ball motion trajectories

Dataset	*Types of table tennis*	*Positioning coverage rate (%)*			
		Proposed model	*R-CNN*	*ISVM*	*IDT*
OpenTTGames	DHS	92.7	82.7	65.4	86.4
	DoubleFish	96.1	83.6	69.8	81.8
	Butterfly	93.4	88.4	74.6	83.6
	STIGA	95.2	86.9	79.2	79.7
	YINHE	94.6	83.5	83.0	80.2
Positioning of the flying paddle movement	DHS	97.1	78.8	92.3	53.7
	DoubleFish	93.6	76.2	88.7	59.4
	Butterfly	94.8	75.1	90.5	63.2
	STIGA	96.0	79.6	87.4	68.1
	YINHE	94.2	80.9	89.6	74.9
TTSwing	DHS	91.9	83.4	67.8	83.4
	DoubleFish	93.7	86.1	62.5	86.7
	Butterfly	98.3	84.9	69.3	90.1
	STIGA	96.5	83.8	72.1	85.4
	YINHE	93.4	82.0	74.6	89.5

As shown in Figure 12(a), with noise intensities ranging from –20 dB to 0 dB, the recall rate of the proposed model was 76.9%, which is significantly higher than R-CNN's 62.3%, ISVM's 65.1%, and IDT's 60.4%. As the noise increased to 0dB, the recall rate of all four models experienced a sharp increase, with the proposed model reaching a maximum recall rate of 96.8%, much higher than the recall rates of the three comparison models. As the noise further increased, the recall rate of all models showed a decreasing trend, but the proposed model's recall rate decreased the least, maintaining a recall rate of 91.4% even at a 20 dB noise level. From Figure 12(b), it can be seen that the average localisation error of the proposed model was as low as 0.59%, the error calculation basis is the relative deviation between the recall rate output by the model and the manually labelled statistical recall rate in the real scenario, which is much lower than ISVM's 1.68%, R-CNN's 1.96%, and IDT's 3.07%. These data further demonstrate the robust performance of the proposed model. Next, the study compared the localisation coverage of the four models in tracking the motion trajectories of different brands of ping-pong balls across the three datasets. The results are shown in Table 3.

Figure 13 Comparison of simulated ping-pong ball motion trajectories for each model, (a) generate the simulation results of table tennis movement trajectories using the proposed model (b) generate the simulation results of table tennis movement trajectories using the ISVM (c) generate the simulation results of table tennis movement trajectories using the R-CNN (d) generate the simulation results of table tennis movement trajectories using the IDT (see online version for colours)



As shown in Table 3, the proposed model achieved the maximum localisation coverage of 96.1%, 97.1%, and 98.3% in the three datasets, with average coverage rates of 94.4%, 95.1%, and 94.8%, all of which are noticeably better than the localisation coverage of the three comparison models. Furthermore, the proposed model showed no significant differences in the localisation coverage for different brands of ping-pong balls, such as Hongshuangxi, Double Fish, and Stiga, with the maximum difference being only 6.4%. This fully demonstrates the promising application prospects of the proposed model. Finally, to further demonstrate the superiority of the proposed model in tracking ping-pong ball motion trajectories, the study compared the simulated ping-pong ball motion trajectories generated by the four models. The results are shown in Figure 13.

From Figures 13(a) to 13(d), it can be seen that the motion trajectories simulated by the proposed model closely matched the actual trajectories. The ISVM model's trajectory also closely followed the actual path, with only slight deviations in the initial and middle stages. The R-CNN and IDT models performed poorly, with their simulated trajectories deviating significantly from the actual path and even shifting the starting point, making their results much less stable and reliable compared to the proposed model. In conclusion, the proposed model demonstrated strong robustness and generalisation ability in ping-pong ball localisation and recognition, and its simulated motion trajectories were the most accurate and closest to the real trajectories.

4 Discussion

To address the current issues of low accuracy and high errors in ping-pong ball localisation and recognition, the study introduced SE and GNN to optimise and improve the YOLOv5 network and SGCM algorithm, respectively. The performance of both models was then verified. In the performance verification experiment of SE-YOLOv5, the integrated algorithm achieved an impressive ping-pong ball localisation accuracy of 97.6%, with a landing error of only 2.7 mm. The computational efficiency was as low as 207 ms, and the recognition image quality reached 48 FPS, all of which were superior to the three comparison algorithms. Xu et al. (2024) proposed an improved YOLOv5 detection method to address the issue of false positives when detecting small objects. This method improved detection quality by optimising the backbone module of YOLOv5. Experimental results showed that this method effectively alleviated the issues of small target imbalance and poor target pixel quality (Xu et al., 2024). This approach, which involves improving the traditional network architecture to enhance performance, shares similarities with the results in this study. Furthermore, in the performance testing of GNN-SGCM, the results showed that its accuracy in matching ping-pong ball positions was 98.6%, with a specificity of 97.2%. This indicated that the integrated algorithm was more effective at recognising non-ping-pong ball movements. Additionally, the algorithm's trajectory tracking overlap for the ping-pong ball was 0.93, and the maximum memory usage during computation was only 223 dB. These results were similar to those of Ramezani's work, where he improved existing recognition methods for LiDAR positioning, breaking through the limitations of frame-to-frame retrieval. To address the issue of current recognition methods having poor adaptability to scenes with indistinct features, M. Ramezani and his team proposed a posture attention GNN, which achieved precise localisation by comparing key nodes between sequential and non-sequential subgraphs. Their method was found to be adaptable to various scenarios (Ramezani et al.,

2023). These experimental data and discussion results thoroughly verify the strong performance of SE-YOLOv5 and GNN-SGCM in ping-pong ball localisation and recognition.

In the field tests, the SE-YOLOv5 and GNN-SGCM-based intelligent ping-pong ball localisation and recognition model also performed exceptionally well. The proposed model achieved an average accuracy of 95.7% in tracking ping-pong ball movement under different lighting conditions. Even when the ball was rotating at high speeds, the minimum accuracy for ping-pong ball localisation was 90.7%. When tested with three different datasets, the average sensitivity of the model's recognition results was 95.42%, 94.06%, and 94.32%, all of which surpassed the three comparison models. Moreover, when Gaussian white noise was added to the dataset, the model achieved a maximum recall rate of 96.8% in tracking the ping-pong ball's trajectory, with a result error of only 0.59%. When tested with different datasets, the model achieved a maximum localisation coverage rate of 96.1%, 97.1%, and 98.3%, respectively. Additionally, the ping-pong ball movement trajectory generated by the model closely matched the actual motion path. These experimental results are similar to the final experimental results of Chiang's team in predicting the motion trajectory of a ping-pong ball. Chiang's team proposed a complex system based on a binocular vision system, odourless Kalman filter trajectory, and speed prediction system to improve the accuracy of ping-pong ball trajectory tracking by 25%, controlling the error range within 86 mm (Chiang et al., 2024). These experimental results fully demonstrate the feasibility and superiority of the proposed model.

The contributions of this study are primarily reflected in three aspects. First, by introducing SE and GNN to optimise and improve the traditional YOLOv5 network and SGCM, and combining the two into an integrated algorithm. Second, an innovative intelligent model for ping-pong ball movement trajectory localisation and recognition was built based on the integrated algorithm. Third, the proposed model was successfully applied to a real-world problem, and its feasibility and superiority were verified through field experiments comparing it with three traditional models. These three contributions not only provide new ideas and methods for research in related fields but also lay the foundation and guarantee for the practical localisation and recognition of ping-pong balls, contributing to the development of the ping-pong ball field.

5 Conclusions

To address the deficiencies of existing ping-pong ball localisation and recognition methods, the study optimised and improved the YOLOv5 network and SGCM algorithm using SE and GNN, respectively, and combined the two to form a novel algorithm. Based on this, an intelligent model for ping-pong ball movement trajectory localisation and recognition was developed. The results showed that the proposed model not only achieved rapid and accurate ping-pong ball localisation but also demonstrated strong robustness and generalisation ability, with excellent adaptability to various complex scenarios. Although the proposed model performed excellently in practical applications, however, the research has the following two deficiencies. The first is that the specific table tennis competition rules were not integrated into the model, and only the objective position and trajectory of the table tennis ball were located and identified. Second, the research did not subdivide the trajectory of table tennis by technical movements, which

may make it difficult to meet the analysis requirements for the deviation between movements and trajectories during training. Therefore, future research will explore from the following three aspects. The first is to quantify the specific parameters of table tennis related rules and construct a scenario-based decision-making module that matches the positioning results with the rules, which is conducive to enabling the model to directly output the penalty results. Second, the research needs to collect trajectory samples of various technical actions, so as to achieve the collaborative recognition of trajectory tracking and action classification by the model. Thirdly, it is necessary to introduce multi-object tracking algorithms to optimise the graph structure modelling of GNN, so as to enhance the stability of the model in scenarios such as multi-ball interaction and limb occlusion, and further expand the practical application value of the model.

Declarations

The author confirms that there are no relevant financial or non-financial competing interests to report.

References

Annaby, M. and Fouda, Y. (2023) 'Fast template matching and object detection techniques using φ-correlation and binary circuits', *MULTIMED. TOOLS APPL.*, June, Vol. 83, No. 3, pp.6469–6496, DOI:/10.1007/s11042-023-15564-x.

Chiang, H.T., Tseng, B.Y., Chen, J.L. and Hsieh, H.C. (2024) 'Trajectory analysis in UKF: predicting table tennis ball flight parameters', *IT PROF*., June, Vol. 26, No. 3, pp.65–72, DOI: 10.1109/MITP.2024.3380837.

Chitraningrum, N., Banowati, L., Herdiana, D., Mulyati, B., Sakti, I., Fudholi, A. and Andria, A. (2024) 'A comparison study of corn leaf disease detection based on deep learning YOLO-v5 and YOLO-v8', *J. ENG. TECHNOL. SCI.*, October, Vol. 56, No. 1, pp.61–70, DOI: 10.5614/j.eng.technol.sci.2024.56.1.5.

Ghose, P., Ghose, A., Sadhukhan, D., Pal, S. and Mitra, M. (2023) 'Improved polyp detection from colonoscopy images using finetuned YOLO-v5', *MULTIMED. TOOLS APPL.*, October, Vol. 83, No. 14, pp.42929–42954, DOI:/10.1007/s11042-023-17138-3.

Han, J., Cao, R., Brighente, A. and Conti, M. (2023) 'Light-YOLOv5: a lightweight drone detector for resource-constrained cameras', *IEEE Internet Things*, November, Vol. 11, No. 6, pp.11046–11057, DOI: 10.1109/JIOT.2023.3329221.

Hu, B., Tian, H., Ni, W., Fan, S., Ni, W. and Hossain, E. (2024) 'Multipath identification, user localization, and environment mapping in radio SLAM', *IEEE T. COMMUN.*, April, Vol. 72, No. 10, pp.6457–6473, DOI: 10.1109/TCOMM.2024.3393977.

Hu, M., Zhu, X., Wang, H., Cao, S., Liu, C. and Song, Q. (2023) 'STDFormer: spatial-temporal motion transformer for multiple object tracking', *IEEE T. CIRC. SYST. VID.*, April, Vol. 33, No. 11, pp.6571–6594, DOI: 10.1109/TCSVT.2023.3263884.

Kunal, K., Dhar, T., Madhusudan, M., Poojary, J., Sharma, A.K. and Xu, W.B. (2023) 'GNN-based hierarchical annotation for analog circuits', *IEEE T. COMPUT. AID D.*, January, Vol. 42, No. 9, pp.2801–2814, DOI: 10.1109/TCAD.2023.3236269.

Langford, S.V., Dryahina, K. and Španěl, P. (2023) 'Robust automated SIFT-MS quantitation of volatile compounds in air using a multicomponent gas standard', *J. Am. Soc. Mass Spectrom*., November, Vol. 34, No. 12, pp.2630–2645, DOI:/10.1021/jasms.3c00312.

Leal-Junior, A., Avellar, L., Blanc, W., Frizera, A. and Marques, C. (2023) 'Opto-electronic smart home: heterogeneous optical sensors approaches and artificial intelligence for novel paradigms in remote monitoring', *IEEE Internet Things*, October, Vol. 11, No. 6, pp.9587–9598, DOI: 10.1109/JIOT.2023.3323481.

Li, Y. and Wu, L. (2023) 'A wireless self-powered sensor network based on dual-model convolutional neural network algorithm for tennis sports', *IEEE SENS. J.*, March, Vol. 23, No. 18, pp.20745–20755, DOI: 10.1109/JSEN.2023.3255226.

Liang, X., Cheng, W., Zhang, C., Wang, L., Yan, X. and Chen, Q. (2023) 'YOLOD: a task decoupled network based on YOLOv5', *IEEE T Consum. Electr.*, May, Vol. 69, No. 4, pp.775–785, DOI: 10.1109/TCE.2023.3278264.

Liu, S., Liu, M., Wu, Y., Li, Z. and Xiao, Y. (2025) 'A multiscale model based on squeeze-and-excitation network for classifying obstacles in front of vehicles in autonomous driving', *IEEE Internet Things*, January, Vol. 12, No. 10, pp.14219–14228, DOI: 10.1109/JIOT.2025.3526226.

Liu, Y., Zhou, X., Wei, C. and Xu, Q. (2024) 'Sparse-view spectral CT reconstruction and material decomposition based on multi-channel SGM', *IEEE T. MED. IMAGING*, June, Vol. 43, No. 10, pp.3425–3435, DOI: 10.1109/TMI.2024.3413085.

Oddo, M.I.B., Kobourov, S. and Munzner, T. (2024) 'The census-stub graph invariant descriptor', *IEEE T VIS. COMPUT. GR.*, December, Vol. 31, No. 3, pp.1945–1961, DOI: 10.1109/TVCG.2024.3513275.

Phan, D.T., Ta, Q.B., Ly, C.D., Nguyen, C.H., Park, S. and Choi, J. (2023) 'Smart low level laser therapy system for automatic facial dermatological disorder diagnosis', *IEEE J. BIOMED. HEALTH*, January, Vol. 27, No. 3, pp.1546–1557, DOI: 10.1109/JBHI.2023.3237875.

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

Rajamohanan, R. and Latha, C.B. (2023) 'An optimized YOLO v5 model for tomato leaf disease classification with field dataset', *ENG. TECHNOL. APPL. SCI.*, September, Vol. 13, No. 6, pp.12033–12038, DOI:/10.48084/etasr.6377.

Ramezani, M., Wang, L., Knights, J., Li, Z., Pounds, P. and Moghadam, P. (2023) 'Pose-graph attentional graph neural network for lidar place recognition', *IEEE ROBOT AUTOM. LET.*, December, Vol. 9, No. 2, pp.1182–1189, DOI: 10.1109/LRA.2023.3341766.

Rao, G.K. and Jena, P. (2023) 'A novel fault identification and localization scheme for bipolar DC microgrid', *IEEE T. IND. INFORM.*, March, Vol. 19, No. 12, pp.11752–11764, DOI: 10.1109/TII.2023.3252409.

Ruifeng, D., Yuanlin, Z., Haiyan, Z., Xinze, L., Peng, C. and Yonghui, L. (2024) 'A high resolution convolutional neural network with squeeze and excitation module for automatic modulation classification', *China Commun.*, November, Vol. 21, No. 10, pp.1–16, DOI: 10.23919/JCC.ja.2022-0270.

Saidani, T. (2023) 'Deep learning approach: YOLOv5-based custom object detection', *ENG, TECHNOL. APPL. SCI.*, October, Vol. 13, No. 6, pp.12158–12163, DOI: 10.48084/etasr.6397.

Sakthivel, S.S., Arunachalam, V. and Jagatheesan, K. (2023) 'Detection, classification, and location of open-circuit and short-circuit faults in solar photovoltaic array: an approach using single sensor', *IEEE J. Photovolt.*, August, Vol. 13, No. 6, pp.986–990, DOI: 10.1109/JPHOTOV.2023.3304113.

Shi, Z., Jia, Y.T., Shi, G.Q., Zhang, K.X., Ji, L.M. and Wang, D.H. (2024) 'Design of motor skill recognition and hierarchical evaluation system for table tennis players', *IEEE SENS. J.*, January, Vol. 24, No. 4, pp.5303–5315, DOI: 10.1109/JSEN.2023.3346880.

Singh, N. and Adhikari, M. (2025) 'FedTune-SGM: a Stackelberg-driven personalized federated learning strategy for edge networks', *IEEE T. PARALL. DISTR.*, February, Vol. 36, No. 4, pp.791–802, DOI: 10.1109/TPDS.2025.3543368.

Tian, Y., Chen, C., Zhang, Q., Li, Y., Li, S. and Ding, X. (2023) 'Multidimensional information recognition algorithm based on CSI decomposition', *IEEE Internet Things*, January, Vol. 10, No. 10, pp.9234–9248, DOI: 10.1109/JIOT.2023.3234054.

Wang, Y., Luo, Y.G., Zhang, H.B., Zhang, W., Dong, K. and He, Q.Y. (2023) 'A table-tennis robot control strategy for returning high-speed spinning ball', *IEEE-ASME T. MECH.*, November, Vol. 29, No. 3, pp.2115–2124, DOI: 10.1109/TMECH.2023.3316165.

Xu, B., Gao, B. and Li, Y. (2024) 'Improved small object detection algorithm based on YOLOv5', *IEEE INTELL. SYST.*, May, Vol. 39, No. 5, pp.57–65, DOI: 10.1109/MIS.2024.3399053.

Yao, Y., Ishikawa, R. and Oishi, T. (2025) 'Stereo-LiDAR fusion by semi-global matching with discrete disparity-matching cost and semidensification', *IEEE ROBOT AUTOM. LET.*, March, Vol. 10, No. 5, pp.4548–4555, DOI: 10.1109/LRA.2025.3552236.

Zaidi, Z., Martin, D., N. Belles, D., Zakharov, V., Krishna, A. and Lee, K.M. (2023) 'Athletic mobile manipulator system for robotic wheelchair tennis', *IEEE ROBOT AUTOM. LET.*, March, Vol. 8, No. 4, pp.2245–2252, DOI: 10.1109/LRA.2023.3249401.