Hang, X.; Cao, D. (2022) Autonomous football exercise system based on convolutional neural network. Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte vol. 22 (85) pp. 231-249
DOI: https://doi.org/10.15366/rimcafd2022.85.015

ORIGINAL

Autonomous football exercise system based on convolutional neural network

Xiaochuan Hang[1] , Dan Cao[2*]
[1] Ministry of sports, Nanjing Medical University, NanJing 210000, JiangSu, China
[2]Physical Education Department, Shen Yang Medical College, Shenyang 110032, Liaoning, China
Email: caodan@symc.edu.cn

UNESCO Code / UNESCO Code:
Council of Europe classification / Council of Europe classification:

Recibido 30 de abril de 2020 **Received** April 30, 2020
Aceptado 26 de junio de 2020 **Accepted** June 26, 2020

Abstract

In order to build an intelligent autonomous football exercise system, this paper proposes a multi-level three-dimensional full convolutional network to segment the neuron football image according to the characteristics of neuron football image. In order to alleviate the problem that the segmentation prediction result generated by the network is biased to the background area and the foreground area is lost or only part of the foreground area is detected, the dice coefficient is introduced to calculate the overlap between the foreground class and the standard class and maximize this. For three-dimensional neuron football sports images, in order to detect three-dimensional breakpoints more conveniently, first analyze the two-dimensional slices. In addition, for two-dimensional neuron football motion image slices, two-dimensional high-curvature points are detected by using the covariance matrix eigenvalues of points on the curve segment and used as the initial screening of breakpoint candidate points. Finally, this paper applies the convolutional neural network to the autonomous football exercise system, researches the robot's action recognition and machine vision, builds an autonomous football exercise system based on the convolutional neural network, and analyzes and simulates its process. The research shows that the autonomous football exercise system based on convolutional neural network proposed in this paper has a certain degree of intelligence.

Keywords. Convolutional neural network, autonomous football, sports training, exercise system

1. INTRODUCTION

Football, as a sport with high popularity and wide participation in the world sports, has received extensive attention from all walks of life (XU et al., 2021). On the basis of a suitable development environment provided by the country and society, how to improve the level of skills and tactics has become an urgent problem for high-level football teams.

High-level football players are the backbone of the revitalization of football sports, and the importance of their tactical training is self-evident. The tactical training of high-level football players basically continues the uniform training program, that is, the high-level football players in the same queue are trained in a uniform way. In the training, the physical training method is relatively simple, and the technical and tactical training is only It is only temporarily guided by coaches on the field, and the proportion of tactical awareness training is even negligible. Under this training mode, it is difficult for high-level football players to achieve significant tactical improvement, and they cannot effectively improve their football tactics. This is for the revitalization of football in China. As for the development of high-level football players, there is an urgent need to improve the current tactical training mode.

How to cultivate and improve high-level football team tactical awareness has always been an important topic in football tactical training. In a football game, the use of tactics mainly depends on the opponent's understanding, analysis and judgment before the game, and then careful tactical arrangements. At the same time, it requires on-the-spot command during the game and the tactical cooperation of the players on the field. Football matches change rapidly. No matter how detailed the pre-match deployment and how timely and correct the on-site command is, it is impossible to accurately predict and arrange the deployment of every action and decision on the field before the match. Therefore, it is extremely important to rely on cultivating strong tactical awareness of the players on the field.

Based on the above analysis, this paper applies the convolutional neural network to the autonomous football exercise system to improve the autonomous exercise effect of football and provide corresponding theoretical references for the further development and promotion of football.

2. Related work

As an important branch of artificial intelligence, distributed artificial intelligence has become a research hotspot in recent years. Its research can be divided into: distributed problem solving (Szűcs & Tamás, 2018) and multi-agent systems (Gu et al., 2019). DPS focuses on information management, including task decomposition and distributed processing (Nasr et al., 2020). MAS takes human society as a reference target, and focuses on the study of collective intelligent behavior (Thành & Công, 2019). MAS is a self-organizing system composed of multiple Agem based on a certain coordination mechanism. The problem to be solved is the knowledge processing problem of distributed multi-agent communication and coordination in a real-time manner in a complex dynamic environment. Because MAS can reflect the

intelligence of human society better than DPS, and is more suitable for an open and dynamic environment, it has received more and more attention (Petrov et al., 2018). Robocup, the Robot World Cup football game, is a typical multi-agent system. Each robot player is composed of one agent, and the changes in the system are caused by the interaction between multiple agents. The purpose of hosting RoboCup robot soccer game is to promote the research and development of distributed artificial intelligence, intelligent robots and intelligent control technology (Hua et al., 2020). RoboCup provides a standard task to encourage researchers to make full use of various technologies and obtain better solutions. Compared with artificial intelligence problems such as computer chess, robot football games are more challenging. The participating robot soccer team consists of multiple robots that move quickly in a dynamic environment. The main feature is a distributed and real-time dynamic environment (Aso et al., 2021).

The classic method of building an Agent is to regard it as a special knowledge system, that is, to realize the representation and reasoning of the Agent through the method of symbolic artificial intelligence. This is the so-called deliberate agent. The biggest feature of deliberate agent is to regard the agent as a consciousness system. One of the purposes of Agent-based systems designed by people is to use them as intelligent agents for human individuals or social behaviors. Then, the Agent should be able to simulate or show the so-called conscious attitudes of the designer, such as beliefs, desires, intentions, goals, commitments, responsibilities, etc. (Mehta et al., 2017).

Literature (Liu et al., 2018) proposes to use belief (Belief), desire (Desire), intention (Intention) to represent Agent. Literature (Ershadi-Nasab et al., 2018) describes beliefs from a cognitive perspective, and believes that beliefs are an agent's estimation of the current world conditions and the possible behavior routes that may be taken to achieve a certain effect; Literature (Nie et al., 2018) describes desires from an emotional perspective, which considers desires It describes Agem's preferences for the future state of the world and possible behavioral routes; the literature (Nie et al., 2019) describes intentions from the perspective of intention, thinking that goals are a subset of desires, but there is no commitment to take specific actions, if one or Some goals are promised, and these goals are intentions. Literature (Zarkeshev & Csiszár, 2019) proposed a series of BDI logics to describe Agent consciousness, using three modal operators to describe beliefs, desires and intentions.

Due to the characteristics and limitations of symbolic artificial intelligence, such as the immaturity of the formal system of the deliberate agent, and the tools used to express the consciousness and attitude of the agent have not been finally unified, this has brought many unresolved and very difficult agents to the deliberate agent. Problems that are difficult or even impossible to solve, so researchers have proposed reactive agents (McNally et al., 2018). Literature (Díaz et al., 2021) believes that Agent should depend on perception and action, and thus proposes a "perception-action" model of agent's intelligent behavior. At this time, the Agent does not need knowledge,

representation, or reasoning. Agem can step by step just like humans. Evolution, Agem's behavior can only be manifested in the interaction between the real world and the surrounding environment. Literature (Bakshi et al., 2021) proposed a sub-premise structure, which is a hierarchical structure composed of behaviors (behaViors) used to complete the task. These structures compete with each other to gain control of the robot.

3. Application of multi-level 3D full convolutional network based on improved V-Net in football image recognition

V-Net is a kind of fully convolutional network, and its network structure is similar to U-Net, as shown in Figure 1. The main innovation is that the research object of V-Net is three-dimensional medical football sports images, and it directly uses the three-dimensional convolution kernel to extract the features of football sports images. At the same time, it handles the imbalance between the foreground area and the background area very well, and uses an end-to-end processing method (Colyer et al., 2018).

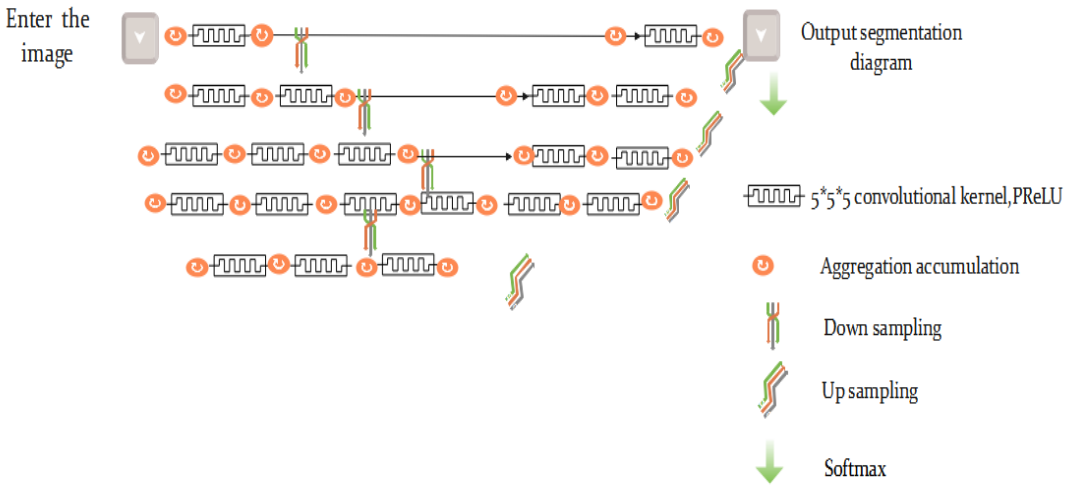


Figure 1. V-Net structure diagram

In the network, the convolution operation is often used to extract the features in the sample data, and an appropriate step size is used to reduce the resolution. The compressed path on the left is encoded, and the expanded path on the right is decoded to ensure that the input and output sizes are consistent. In the network, the convolution uses proper padding (Padding), and the PReLu nonlinear activation function is applied after the convolution operation.

Each level of the encoder on the left is composed of multiple convolutions, and the resolution of each layer is not consistent. At the same time, the input and output of each layer are added. In this stage, a 5×5×5 convolution kernel is used, and a downsampling with a step size of 2 and a convolution kernel of 2×2×2 is used between each level to reduce the resolution, so that the obtained The size of the feature map is halved, and the number of channels in the feature map is doubled. The replacement of pooling operation is beneficial to the subsequent network layer while reducing

the size of the input signal and expanding the range of the feature receptive field (Sárándi et al., 2020).

Similar to the basic concept of V-Net, the architecture model of the multi-level 3D full convolutional network is shown in Figure 2, and the convolution kernel size and input size of each layer are shown in Table 1. There are also two paths in the network, which are composed of an encoder (left) and a decoder (right). In the left path of the neural network, down-sampling is performed at the end of each layer to reduce the size of the input signal and increase the receptive field of the subsequent network layer calculation features. In the path on the right, the features extracted from the same level on the left are combined to collect more information and details. The integration of features at different levels helps to learn enough football sports image features. Similar to the left path, upsampling is used to increase the size of the input football motion image to maintain the same size as the left horizontal layer. The last layer uses a full convolution with a convolution kernel size of I×I×I, which can produce an output of the same size as the input, and then the feature map calculated by this layer is converted into the probability segmentation of the foreground and background area through the Softmax function (Azhand et al., 2021).

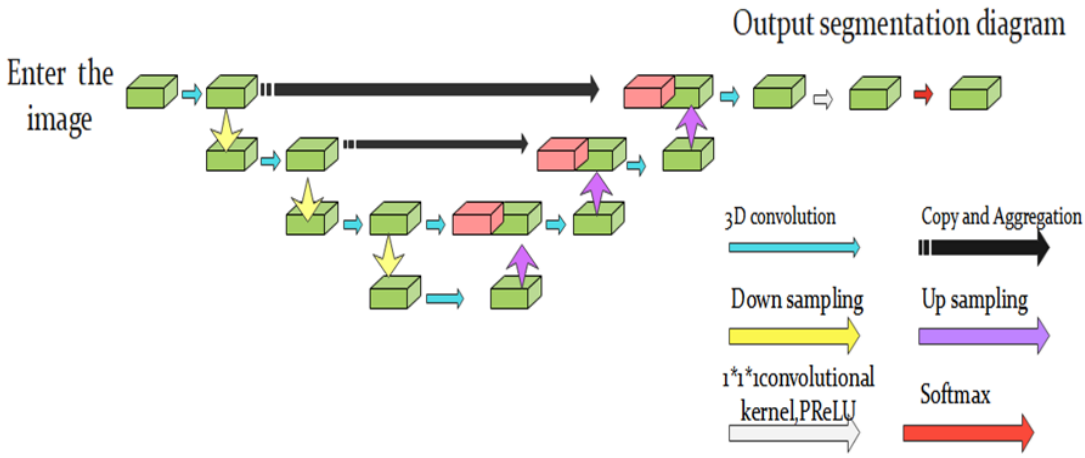


Figure 2. Structure diagram of multi-level three-dimensional full convolutional network

Table 1. Convolution kernel size and input size in the network

Number of network layers	Input size	Type	Convolution kernel size
First layer	64×64×32	3D convolution Downsampling and upsampling	7×7×3(32) 2×2×2(32)
Second layer	32×32×16	3D convolution Downsampling and upsampling	7×7×3(64) 2×2×2(64)
Third layer	16×16×8	3D convolution Downsampling and upsampling	7×7×3(128) 2×2×2(128)
Fourth layer	8×8×4	3D convolution	7×7×3(32)

There is a layered structure in the Inception model, and the initial design is shown in Figure 3. The first layer is connected to the input, and after

the lower layer (the layer close to the input), different sizes of convolution kernels are used to process different sizes of receptive fields, and finally features of different scales are merged. The output convolution kernels are connected together to form the input of the next stage. Among them, the use of convolution kernel sizes of 1×1, 3×3, and 5×5 is to facilitate feature fusion, and adding a pooling operation in each stage is conducive to the realization of the Inception model (Xu & Tasaka, 2020).

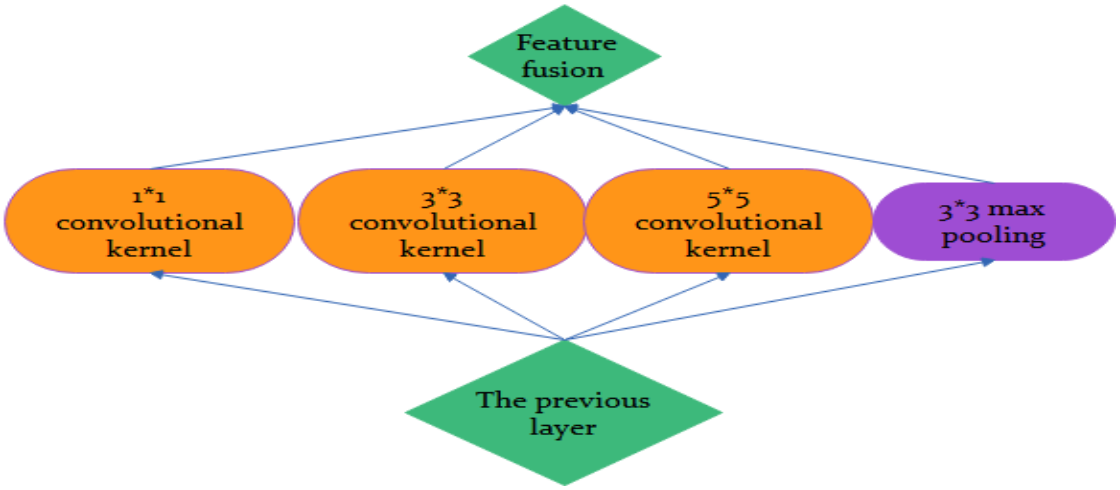


Figure 3. Initial Inception model

However, the stacking of Inception models on top of each other will result in different output data. When obtaining high-level abstract features, the spatial aggregation ability will be reduced. This indicates that the ratio of 3×3 and 5×5 convolution kernels should increase with the increase of the number of layers. When the output of the pooling layer is merged with the output of the convolutional layer, it will lead to an increase in the number of outputs, which will easily lead to a sharp increase in the amount of calculation of the network module. Therefore, a 1×1 convolution kernel is used for dimensionality reduction, and the Inception model for dimensionality reduction is shown in Figure 4.

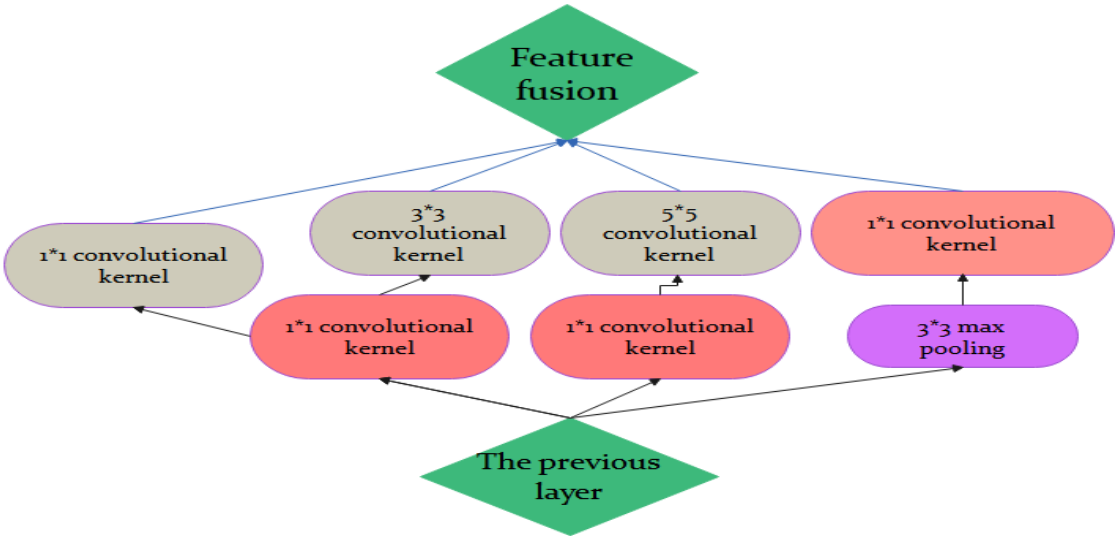


Figure 4. Inception model for dimensionality reduction

The Inception model is introduced into the down-sampling of the multi-level 3D full convolutional network to process neuron football images of different sizes. The multiple convolution kernels of different sizes used in the Inception model ensure that the information hidden on different scales can be processed at the same time. As shown in Figure 5, in each Inception model, different convolution kernels are used, including 3×3×3 convolution kernels and 5×5×5 convolution kernels to fit different football sports image size and reduce dimensions (Li et al., 2020).

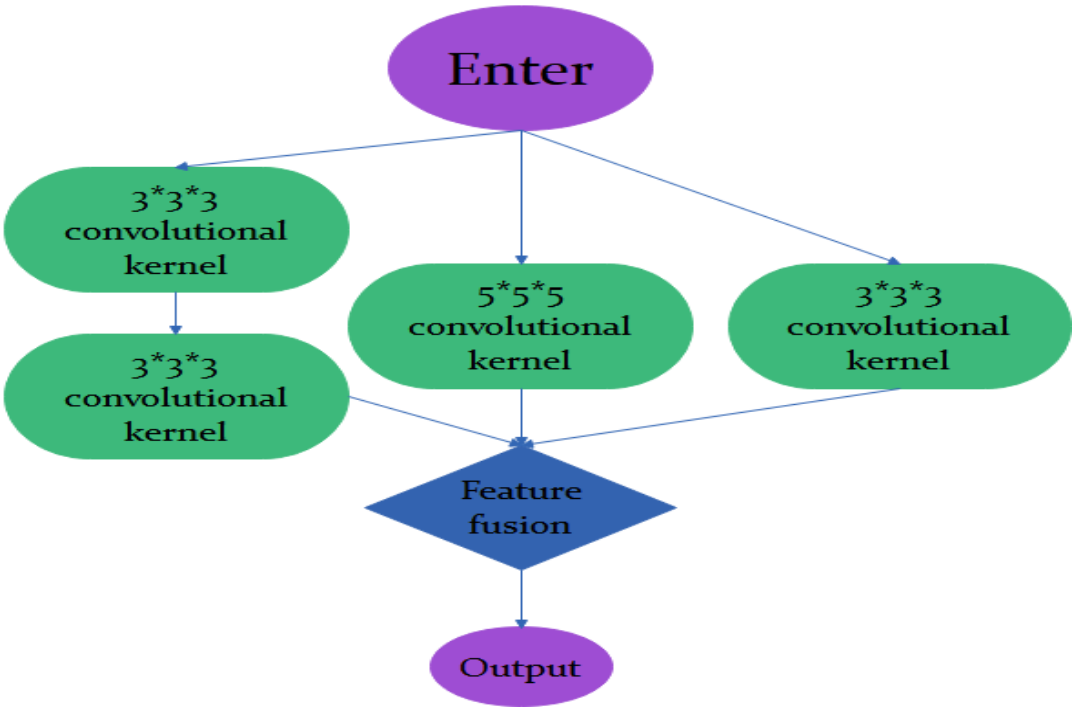


Figure 5. Inception model in a multi-level three-dimensional full convolutional network

This can also significantly increase the number of convolution kernels at each stage. In the subsequent stages, the network will not lose control due to increased computational complexity.

In the field of deep learning, it is easier for the network to learn the features and information of the part with a larger proportion of data samples during the training and learning process, and it is easy to ignore the features and information of the part with a smaller proportion.

In the V-Net network, this paper considers that the proportion of the target area in the medical football sports image is small. In order to alleviate the problem that the segmentation prediction result generated by the network is biased to the background area and the foreground area is lost or only part of the foreground area is detected, this paper introduces the Dice Coefficient to calculate the overlap between the foreground class and the standard class and maximize this. The Dice coefficient is shown in equation (1).

$$D = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \quad (1)$$

Among them, p_i represents the probability of the predicted class, g_i represents the probability of the standard class, and N represents the total number of pixels. We differentiate the predicted j-th pixel to obtain the gradient as in formula (2).

$$\frac{\partial D}{\partial p_j}=\left[2\frac{g_i\left(\sum_i^N p_i^2+\sum_i^N g_i^2\right)-2p_j\left(\sum_i^N p_i g_i\right)}{\left(\sum_i^N p_i^2+\sum_i^N g_i^2\right)^2}\right] \quad (2)$$

Combined with the objective function of the Dice coefficient, it is possible to achieve a better balance effect and final result without assigning loss weights to different types of samples. However, due to the square term in the denominator of the derivative, a larger gradient may be obtained in the calculation, which affects the stability of training.

For the research object of the three-dimensional neuron football motion image stack in this paper, because the background area occupies a much larger proportion than the foreground area, the number of foreground pixels and background pixels is obviously unbalanced. Therefore, if the foreground class and the background class are treated equally in the loss function calculation, the network will mainly capture the characteristics of the background area and ignore the information of the foreground area. In order to solve this unbalanced problem, a weighted cross-entropy loss function is used in the training process of the network, which is defined as shown in equation (3).

$$L(\theta)=-\alpha\sum_{i=Y_+}\log\left(h_\theta(y_i)\right)-\beta\sum_{i=Y_-}\log\left(1-h_\theta(y_i)\right) \quad (3)$$

Among them, Y_+ and Y_- represent foreground pixels and background pixels, respectively, and α and β are used as weights to balance the uneven number of foreground and background categories. $h_\theta(y_i)$ represents the probability map generated by executing the Sigmoid function, as shown in equation (4).

$$h_\theta(y_i)=\frac{1}{1+e^{-\theta^T y_i}} \quad (4)$$

For the three-dimensional neuron football motion image. In order to detect three-dimensional breakpoints more conveniently, we first analyze the two-dimensional slices. For the research object at this stage, that is, the three-dimensional neuron segment, detecting the breakpoint of the three-dimensional neuron is to detect the terminal point of the neuron segment. Therefore, it is consistent with the detection object of the multi-scale ray emission model. For the two-dimensional neuron football motion image slice, the two-dimensional high curvature points are detected by the eigenvalues of the covariance matrix of the points on the curve segment, and they are used as the initial screening of the breakpoint candidate points.

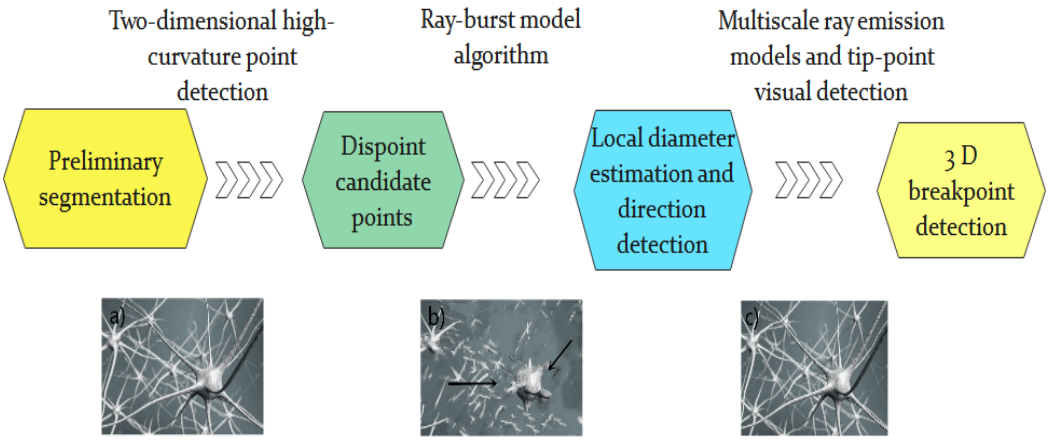
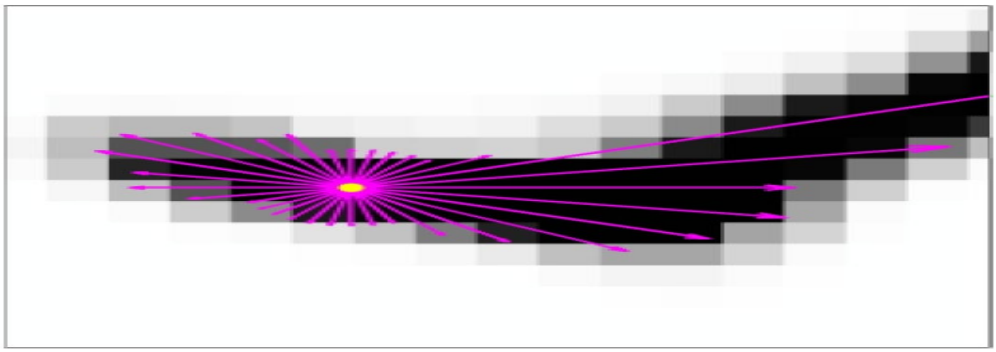
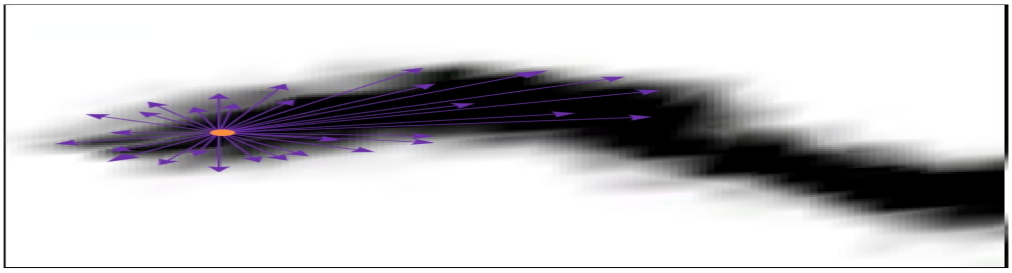


Figure 6. Flow chart of neuron breakpoint detection

As shown in Figure 7, (a) is a two-dimensional ray burst model, and (b) is a three-dimensional ray burst model. The two-dimensional ray burst model is relatively simple. It mainly rotates a single ray in equal angle increments in the plane to form a two-dimensional sampling core. In the three-dimensional ray burst model, the regular polyhedron is subdivided recursively to generate a three-dimensional sampling kernel, and try to make the end points of the sampling kernel evenly distributed on the unit sphere. In theory, the larger the number of sampling cores, the more accurate the result will be, but it will also result in a decrease in the running speed of the algorithm. Therefore, the researcher needs to comprehensively consider the accuracy and speed, and select the appropriate number of sampling cores (rays).



(a) Two-dimensional ray burst model



（b） Three-dimensional ray burst model

Figure 7. Two-dimensional ray burst model and three-dimensional ray burst model

A two-dimensional Ray-Shooting model is proposed to detect the terminal points in a three-dimensional neuron football motion image. The main steps are shown in Figure 8. Firstly, the two-dimensional football image of each layer in the three-dimensional neuron football image is studied, and the high curvature points detected in the edge of the two-dimensional football image are used as the candidate points of the terminal points in the two-dimensional football image.

Subsequently, the ray emission model is used to screen the two-dimensional terminal points, so as to obtain the two-dimensional terminal points and use them as the candidate points of the terminal points in the three-dimensional neuron football motion image. The gray level changes of the three-dimensional terminal points in each layer of the two-dimensional football image slice have certain rules. That is, two-dimensional peripheral points can be detected on the same horizontal and vertical coordinates in adjacent slices, and their gray values show a decreasing trend to the background area.

Therefore, detecting whether there is a terminal point at the same position in the adjacent slices of each two-dimensional terminal point, and gathering the regular two-dimensional terminal points to complete the detection of the three-dimensional terminal point.

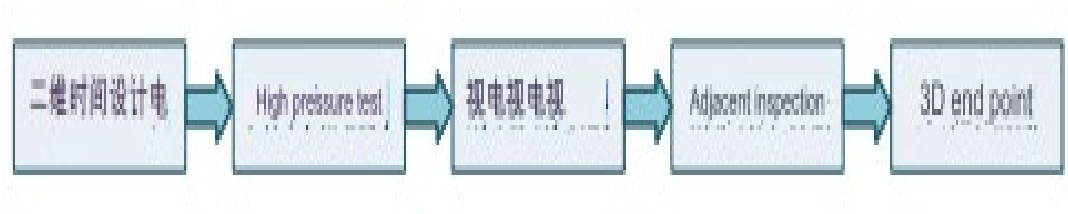
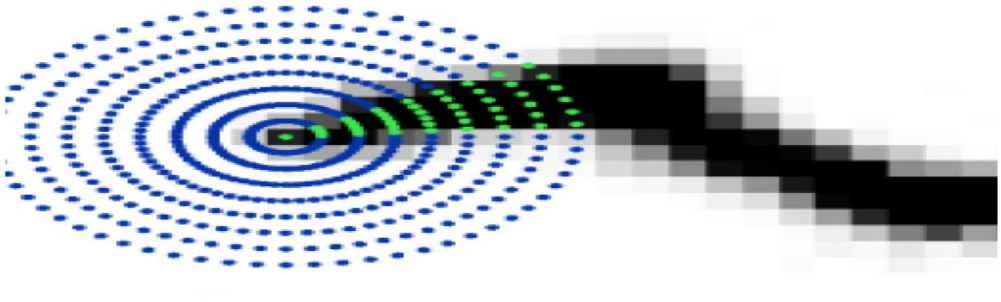


Figure 8. The flow chart of the ray emission model to detect the three-dimensional terminal point

The two-dimensional ray emission model starts from a certain node and emits multiple rays of equal length to the surroundings at equal angles. As shown in Figure 9, a ray with more than half of the pixels in the foreground area is called a foreground ray, as shown by the ray composed of purple dots in the figure.

In contrast, rays with more than half of the pixels in the background area are called background rays, as shown by the rays composed of orange dots in the figure. The ray emission model mainly uses the information provided by the foreground rays and the background rays to detect the peripheral points.

(a) Schematic diagram of foreground rays



(b) Background rays

Figure 9. Schematic diagram of foreground rays and background rays

For the point of interest P, we assume that there are M rays in the ray emission model, each ray has N points, and the ray length is N, then the average grayscale intensity A(i) of each ray is shown in equation (5) .

$$A(i)=\frac{2\sum_{j=1}^{N} I(i,j)}{N} \tag{5}$$

Among them, i=1:M, j=1:N, and I(i,j) represents the pixel value of the point on the ray. At this time, the maximum average gray-scale intensity M_I of the rays is shown in formula (6).

$$M_I=\max_i\left(A(i)\right) \tag{6}$$

We set the threshold T_0. If the maximum value M_I of the average gray intensity of the ray is not greater than the threshold T_0, the point of interest is regarded as the background point. that is, the pixel point (noise or non-front spot) located in the background area.

Conversely. if the maximum average gray-scale intensity M_I of the ray is greater than the threshold T_0, this paper sets a threshold $T=M_I\times R$ to determine whether the ray is a foreground ray, where $R\in(0,1)$. If the maximum value M, of the average gray intensity of the ray is greater than the threshold T, then the ray is a foreground ray. Otherwise, it is a background ray. At this time, the set Q and the set size n of the foreground rays are shown in equation (7).

$$\begin{cases} Q=\{i\mid A(i)>T, i=1:M\} \\ n=\#Q \end{cases} \tag{7}$$

From this, the maximum angle MA formed by the foreground ray of the point of interest P can be calculated, as shown in equation (8).

$$MA=\max_{p\in Q,q\in Q}\left(\arg\left(r_p,r_q\right)\right) \tag{8}$$

In the formula, $arg(r_p, r_q)$ is the angle formed by the foreground ray rp and the foreground ray rq.

In the ray emission model, for the point of interest P, the number of foreground rays n and the maximum angle MA formed by the foreground rays need to be within a certain range to be judged as a terminal point, as shown in equation (9).

$$P \in TP_2 if \begin{cases} T_1 < \frac{n}{M} < T_2 \\ MA < T_3 \end{cases} \quad (9)$$

Among them, TP_2 represents the peripheral point of the two-dimensional neuron football motion image, and T_1, T_2, and T_3 are all constant values set based on experimental experience.

In order to obtain more abundant features, the multi-scale ray emission model implements the ray emission model on L different scales $\{s^k\}_{k=1}^{L}$ of the two-dimensional high curvature point p. The ray length of each model is $N = s^k$. The scale range is automatically determined by the local diameter of the neuron d(p), as shown in equation (10).

$$\begin{cases} s^1 = d(p), d(p) > u \\ s^1 = u, d(p) \le u \\ s^{k+1} = s^k + \eta, k = 1, ..., L-1 \end{cases} \quad (10)$$

Among them, u is set to prevent the scale from being too small, and n is the step size. The multi-scale ray emission model is similar to the ray emission model. It mainly extracts the two features of the number of foreground rays and the maximum angle composed of foreground rays. The number sequence and maximum angle sequence of foreground rays are shown in equations (11) and (12) respectively.

$$F_1 = \{n_k\}_{k=1}^{L} \quad (11)$$

$$F_2 = \{MA_k\}_{k=1}^{L} \quad (12)$$

In the formula, n_k and n_k represent the number of foreground rays and the maximum angle in scale s^k, respectively. After that, this paper analyzes the two characteristic sequences of F_1 and F_2 to judge the two-dimensional terminal points. If the candidate point p is a terminal point, its feature sequences F_1 and F_2 need to meet the following conditions:

The ray emission model is implemented on the scale s^k, and r_c and r_d are used to represent the two rays in the maximum angle of the foreground ray. Then, the number of rays $\tilde{n}_k$ in the two rays is as shown in equation (13).

$$\tilde{n}_k = \frac{MA_k}{\theta} + 1 \qquad (13)$$

If the foreground ray is continuous, then $\tilde{n} = n$, otherwise, $\tilde{n} > n$ If p is a terminal point, the foreground ray should be continuous at every scale, that is $\tilde{n}_k = n_k, k = 1, \ldots, L-1$. We substitute formula (13) into it to obtain formula (14).

$$n_k = \theta\left(n_k - 1\right) \qquad (14)$$

Since 0 is a constant, there is a linear relationship between n_k and n_k at the tip.

The maximum angle n_k of the foreground ray in the scale s^k should be within a certain range, as shown in equation (15).

$$\min_{k \in L}\left(\left\{MA_k\right\}\right) < \tau \qquad (15)$$

Among them, τ is the angle threshold. Normally, because the smaller scale may lead to the maximum angle of the foreground ray, there is no specific requirement on the size of the maximum angle n_k of each foreground ray in the scale s^k. However, the minimum value of the maximum angle n_k of the foreground ray should be small enough to ensure the correctness of the distal point.

In order to avoid other disturbances that have nothing to do with the structure of neurons. Before generating the maximum intensity projection, we first dig out a local cube B centered on the candidate point p of the three-dimensional neuron terminal point. The pixels whose distance from the candidate point p of the three-dimensional terminal point is greater than Sl cannot be acquired and analyzed by the multi-scale ray emission model. Therefore, the side length h of the local cube B is as shown in equation (16).

$$h = 2 \times s^l + 1 \qquad (16)$$

Normally, the local cube B centered on the candidate point of the three-dimensional neuron terminal point generates two-dimensional maximum intensity projections in the XY, YZ, and ZX planes, which are called I_{XY}, I_{YZ}, and I_{ZX}, respectively. According to TVP, two-dimensional peripheral points are detected in these three planes respectively. If there are two-dimensional peripheral points in two or more of the three planes generated with the candidate point as the center, then the candidate point is judged to be a neuron Three-dimensional tip point.

The segmentation repair model based on the Heather matrix is the last step to repair the broken structure in the neuron segment. It mainly uses the eigenvalues of the Heather matrix to judge each point in the local area, and then completes the repair of the broken part in the neuron structure. In the neuron breakpoint detection part, the local diameter of the neuron breakpoint

has been calculated using the ray burst model and the direction of the neuron segment has been detected. According to the location of the breakpoint ρ and the local diameter d and the direction $\vec{o}$, this paper establishes a cylindrical area to analyze the information near the breakpoint. As shown in Figure 10, it takes the break point p (yellow dot) as the center to generate a circular plane with the diameter d (blue line segment) as the size and perpendicular to the direction $\vec{o}$ (red arrow). The plane moves along the direction C, thereby defining a cylindrical (gray) area with a length of L. The direction of the generatrix of this cylinder is consistent with the direction $\vec{o}$ of the neuron segment where the breakpoint is located. For each pixel in this cylindrical structure, this paper analyzes its tubular feature value and judges whether the pixel belongs to the foreground area.

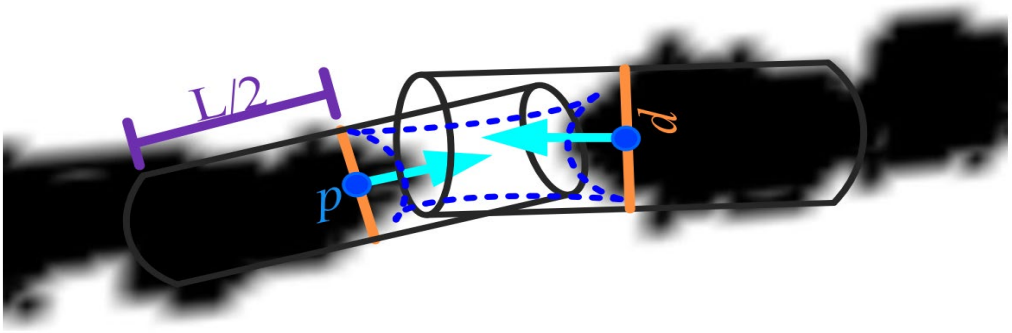


Figure 10. Schematic diagram of repair based on Heather matrix analysis

The tubular eigenvalues are obtained by analyzing the eigenvalues of the Heather matrix. For the pixel point in the cylindrical area formed by the breakpoint p as the center, this paper assumes its position is (x,y,z), then the grayscale intensity here is f(x,y,z), then Heather The matrix H is shown in equation (17).

$$H = \begin{bmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{bmatrix} \tag{17}$$

Among them, f_{ij} represents the mixed second derivative on i and j. Because of $f_{xy} = f_{yx}$, $f_{xz} = f_{zx}$, and $f_{yz} = f_{zy}$., the Heather matrix H is a symmetric matrix. The three eigenvalues λ_1, λ_2, and λ_3 of the Heather matrix H can be obtained by calculation. It is assumed that the relationship between these three is shown in equation (18).

$$|\lambda_1| \le |\lambda_2| \le |\lambda_3| \tag{18}$$

If the position (x, y, z) is in the tubular structure, it should conform to equation (19).

$$|\lambda_1| \approx 0, |\lambda_1| \ll |\lambda_2|, \lambda_2 \approx \lambda_3 \tag{19}$$

Because the neuron structure is stronger than the pixel intensity of the background area, so $\lambda_2 < 0.\lambda_3 < 0$.

Obviously, there needs to be a calculation rule to make the difference of the sum as large as possible. Therefore, a magnitude function $M(\lambda_1, \lambda_2, \lambda_3)$ is defined as shown in equation (10).

$$M(\lambda_1, \lambda_2, \lambda_3) = \left(|\lambda_1| - |\lambda_2|\right)^q \quad (20)$$

In the formula, q is set to 2 in the experiment. At the same time, the likelihood function $L(\lambda_1, \lambda_2, \lambda_3)$ is shown in equation (21).

$$L(\lambda_1, \lambda_2, \lambda_3) = \frac{\left(|\lambda_2| - |\lambda_1|\right)}{|\lambda_3|} \quad (21)$$

Therefore, the output $O(\lambda_1, \lambda_2, \lambda_3)$ of the tubular eigenvalue is designed as equation (22).

$$O(\lambda_1, \lambda_2, \lambda_3) = M(\lambda_1, \lambda_2, \lambda_3) \times L(\lambda_1, \lambda_2, \lambda_3) \quad (22)$$

When $\lambda_2 < 0$ and $\lambda_3 < 0$, the size of the output $O(\lambda_1, \lambda_2, \lambda_3)$ is shown in equation (23).

$$O(\lambda_1, \lambda_2, \lambda_3) = \frac{\left(|\lambda_2| - |\lambda_1|\right)^3}{|\lambda_3|} \quad (23)$$

In other cases, 0 will be output. For the output tubular eigenvalues, select the appropriate threshold s. If the tubular feature value output result $O(\lambda_1, \lambda_2, \lambda_3)$ obtained at the position (x, y, z) is greater than the threshold ε, it is considered that the place should be a foreground pixel. Otherwise, it is considered a background pixel.

4. Autonomous football exercise system based on convolutional neural network

The chromosome is the expression of the team's strategy. A certain segment on the chromosome represents the strategy of a certain player. The combination of all such segments is enough to form the strategy of the entire team. The coding of chromosomes is closely related to the way the court is divided. According to the current state (offensive and defensive, whether to hold the ball), the strategy of a certain division on the court is mapped to the robot on the team chromosome to form a segment of the team chromosome, and all such segments are combined to form the complete chromosome of the team. The division method of the court and the division number of the right team are shown in Figure 11.

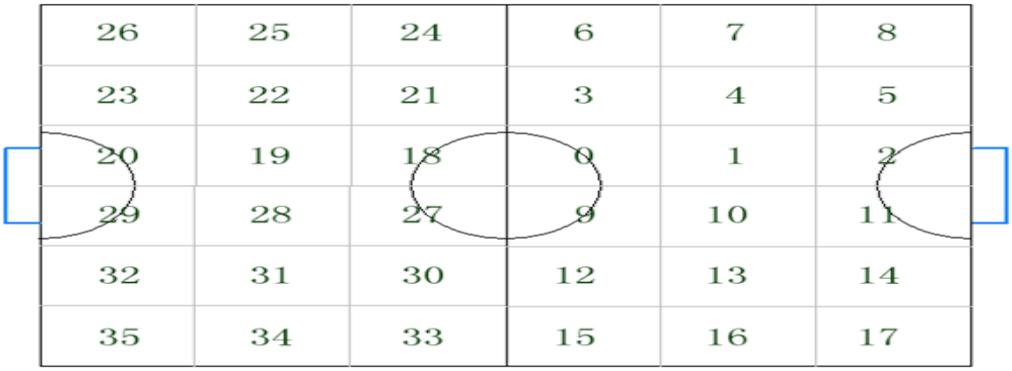


Figure 11. The way the court is zoned and the zone number of the team

Figure 12 shows the core process of strategy training. It should be noted that there are two modes of strategy training, namely competition and training mode. The game mode is just a game between two teams. According to the single-loop organization of the game, the end of the current round is the end of the single-loop game. The training mode is to train multiple teams (N ≥ 2) in accordance with the basic convolutional neural network process. As long as the training process is not interrupted, the training process will be carried out automatically according to the above process.

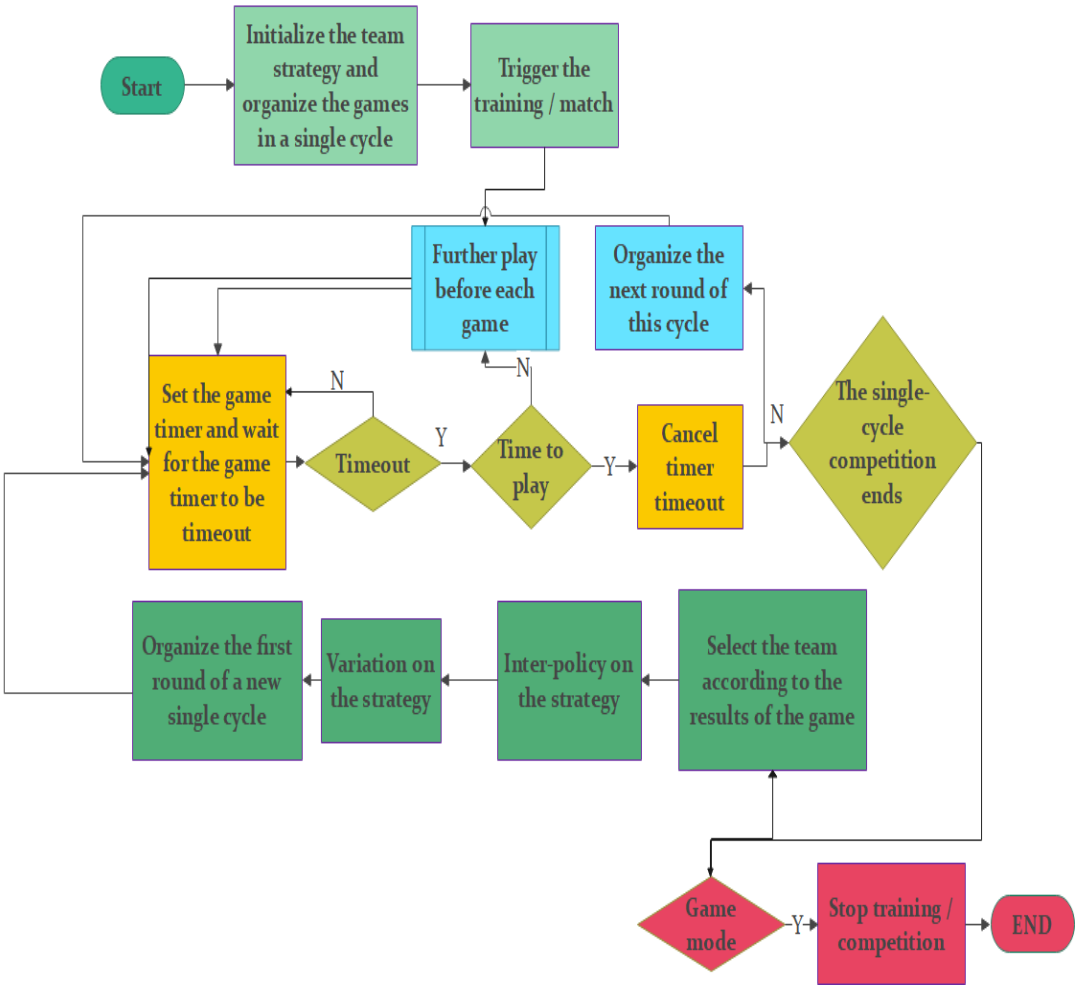


Figure 12. The core process of the strategy training algorithm

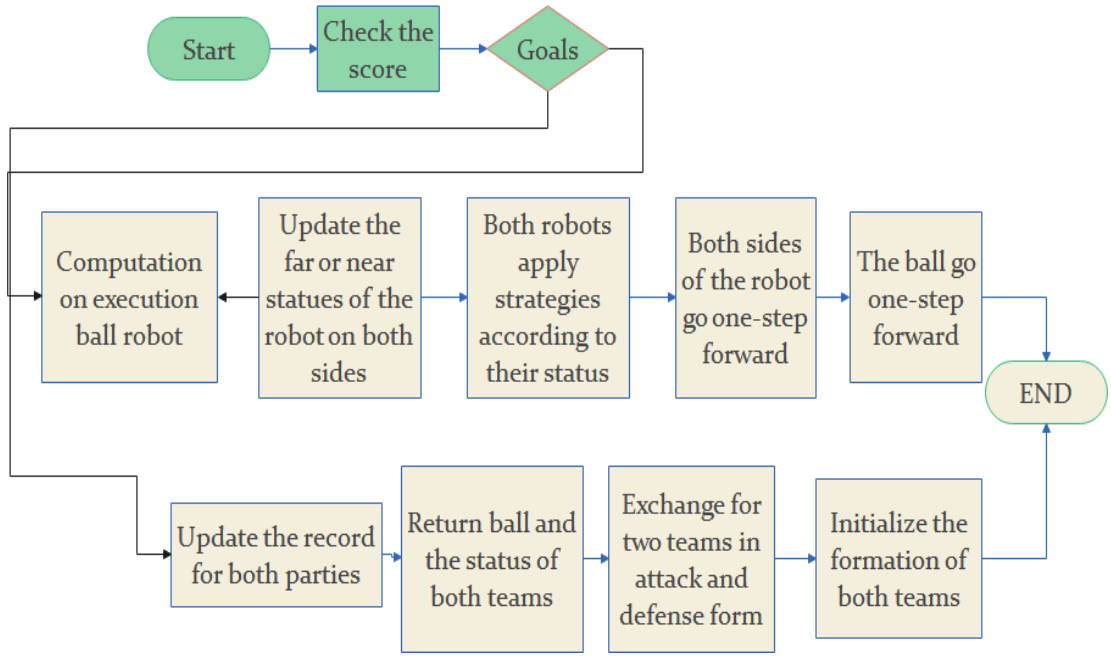


Figure 13. The process of step-by-step process of each game

5. Conclusion

With the development of computer technology, the theory and application research of multi-agent systems in distributed artificial intelligence has become a hot spot in artificial intelligence. The intelligent football exercise system is a concentrated embodiment of artificial intelligence and robotics. Its purpose is to promote the research of artificial intelligence and robotics, through the universal platform of football, to evaluate various theories, algorithms and the architecture of intelligent bodies. The cooperative control and decision-making in the machine system can be used for the assisted control of the unmanned combat platform group. In this paper, the convolutional neural network is applied to the autonomous football exercise system, the robot's action recognition and machine vision are studied, and the autonomous football exercise system based on the convolutional neural network is constructed, and its process is analyzed and simulated. The research results show that the autonomous football exercise system based on convolutional neural network proposed in this paper has a certain degree of intelligence.

References

Aso, K., Hwang, D.-H., & Koike, H. (2021). Portable 3D human pose estimation for human-human interaction using a chest-mounted fisheye camera. Proceedings of the Augmented Humans International Conference 2021,

Azhand, A., Rabe, S., Müller, S., Sattler, I., & Heimann-Steinert, A. (2021). Algorithm based on one monocular video delivers highly valid and reliable gait parameters. *Scientific Reports*, *11*(1), 14065.

Bakshi, A., Sheikh, D., Ansari, Y., Sharma, C., & Naik, H. (2021). Pose estimate based yoga instructor. *International Journal of Recent*

Advances in Multidisciplinary Topics, *2*(2), 70-73.

Colyer, S. L., Evans, M., Cosker, D. P., & Salo, A. I. (2018). A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system. *Sports medicine-open*, *4*(1), 1-15.

Díaz, R. G., Laamarti, F., & El Saddik, A. (2021). DTCoach: your digital twin coach on the edge during COVID-19 and beyond. *IEEE Instrumentation & Measurement Magazine*, *24*(6), 22-28.

Ershadi-Nasab, S., Noury, E., Kasaei, S., & Sanaei, E. (2018). Multiple human 3d pose estimation from multiview images. *Multimedia Tools and Applications*, *77*, 15573-15601.

Gu, R., Wang, G., Jiang, Z., & Hwang, J.-N. (2019). Multi-person hierarchical 3d pose estimation in natural videos. *IEEE Transactions on Circuits and Systems for Video Technology*, *30*(11), 4245-4257.

Hua, G., Li, L., & Liu, S. (2020). Multipath affinage stacked—hourglass networks for human pose estimation. *Frontiers of Computer Science*, *14*, 1-12.

Li, Z., Bao, J., Liu, T., & Jiacheng, W. (2020). Judging the normativity of PAF based on TFN and NAN. *Journal of Shanghai Jiaotong University (Science)*, *25*, 569-577.

Liu, S., Li, Y., & Hua, G. (2018). Human pose estimation in video via structured space learning and halfway temporal evaluation. *IEEE Transactions on Circuits and Systems for Video Technology*, *29*(7), 2029-2038.

McNally, W., Wong, A., & McPhee, J. (2018). Action recognition using deep convolutional neural networks and compressed spatio-temporal pose encodings. *Journal of Computational Vision and Imaging Systems*, *4*(1), 3-3.

Mehta, D., Sridhar, S., Sotnychenko, O., Rhodin, H., Shafiei, M., Seidel, H.-P., Xu, W., Casas, D., & Theobalt, C. (2017). Vnect: Real-time 3d human pose estimation with a single rgb camera. *Acm transactions on graphics (tog)*, *36*(4), 1-14.

Nasr, M., Ayman, H., Ebrahim, N., Osama, R., Mosaad, N., & Mounir, A. (2020). Realtime multi-person 2D pose estimation. *International Journal of Advanced Networking and Applications*, *11*(6), 4501-4508.

Nie, X., Feng, J., Xing, J., Xiao, S., & Yan, S. (2018). Hierarchical contextual refinement networks for human pose estimation. *IEEE Transactions on Image Processing*, *28*(2), 924-936.

Nie, Y., Lee, J., Yoon, S., & Park, D. S. (2019). A multi-stage convolution machine with scaling and dilation for human pose estimation. *KSII Transactions on Internet and Information Systems (TIIS)*, *13*(6), 3182-3198.

Petrov, I., Shakhuro, V., & Konushin, A. (2018). Deep probabilistic human pose estimation. *IET Computer Vision*, *12*(5), 578-585.

Sárándi, I., Linder, T., Arras, K. O., & Leibe, B. (2020). Metrabs: metric-scale truncation-robust heatmaps for absolute 3d human pose estimation. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *3*(1), 16-30.

Szűcs, G., & Tamás, B. (2018). Body part extraction and pose estimation

method in rowing videos. *Journal of computing and information technology*, *26*(1), 29-43.

Thành, N. T., & Công, P. T. (2019). An evaluation of pose estimation in video of traditional martial arts presentation. *Journal on Information Technologies & Communications*, *2019*(2), 114-126.

Xu, J., & Tasaka, K. (2020). keep your eye on the ball: detection of kicking motions in multi-view 4K soccer videos. *ITE Transactions on Media Technology and Applications*, *8*(2), 81-88.

XU, J., TASAKA, K., & YAMAGUCHI, M. (2021). Fast and accurate whole-body pose estimation in the wild and its applications. *ITE Transactions on Media Technology and Applications*, *9*(1), 63-70.

Zarkeshev, A., & Csiszár, C. (2019). Rescue method based on V2X communication and human pose estimation. *Periodica Polytechnica Civil Engineering*, *63*(4), 1139-1146.