Research on the Application of Virtual Technology-based Posture Detection Device in Swimming Teaching

Hongming Guo [1], Jingang Fan[2, 3*]
College of Physical Education, Putian University, Putian, 351100, China[1]
Department of Physical Education, Sichuan International Studies University, Chongqing, 400031, China[2]
Graduate School, JoséRizal University, Manila, 1552, Philippines[3]

***Abstract*—With the socio-economic development, the national demand for playing leisure sports has increased, and swimming is one of the popular choices. To help swimming beginners understand the correct swimming posture more quickly and directly, hybrid neural network algorithms based on sliding window detection and deep residual networks are designed in this study, and two corresponding virtual image classification models of swimmer's posture are designed based on these algorithms. In order to reduce the noise of the input data and reduce the cost of data collection, the virtual reality technology is used to convert the swimmer's swimming pose image into the image model in the virtual reality space as the input data of the algorithm. The performance test experimental results show that the classification accuracy of the swimmer pose recognition models based on PTP-CNN algorithm and SW-CNN algorithm designed in this research are 97.48% and 96.72% respectively on the test set, which are much higher than other comparison models, and the model built based on PTP-CNN algorithm has the fastest computation speed. The results of this research can be applied to assist participants in swimming pose recognition in teaching beginner swimmers.**

Keywords*—Virtual reality; pose recognition; swimming; neural network; sliding window

I. Introduction

Swimming has both leisure and physical exercise attributes, making it a better form of exercise for adults with a certain financial base [1]. In fact, the middle class population in China has grown significantly in the last decade, which has directly led to an increase in the number of people learning to swim in their spare time. However, it is difficult for beginners to distinguish between accurate and incorrect common swimming positions, but the ability to accurately judge and master scientific swimming positions is important to learn to swim and reduce the risk of personal safety during exercise [2]. In the past, beginners often need to spend a lot of time and energy on learning to identify and practice various correct swimming postures, which is inefficient, and may even make learners lose interest in swimming [3]. At the same time, the emergence of high and new technologies, including virtual reality and artificial intelligence, provides new ideas for automating traditional physical education teaching tasks. These technologies can free teachers from repetitive low-end teaching tasks, and the teaching efficiency and error rate are far lower than those of the teaching system based on traditional machine learning algorithms [4]. Aiming at this problem, this research tries to use virtual reality technology and artificial intelligence technology to design a model that can intelligently judge whether the swimmer's posture in the swimming image is correct, so as to provide timely and accurate feedback for beginners to learn the project. It is assumed that it may play a role in shortening the learning cycle of swimming learners and reducing the learning difficulty, this is also the contribution and value of this paper to the field of swimming teaching.

*Corresponding Author.

II. Related Work

A large number of related studies have been conducted by experts in different fields about virtual reality technology, bit-posture detection technology, and smart sports teaching involved in this research. Elbamby M S et al. discussed the challenges faced in achieving ultra-reliable and low-latency virtual reality by addressing the shortcomings of virtual reality technology in terms of high throughput, low latency, and reliable communication, thus finding that the use of millimeter wave communication, edge computing, and Intelligent networks with active caching techniques can solve the above problems of virtual reality technology applications to a certain extent [5]. Chen M et al. studied the resource management problem of networks with wireless virtual reality users communicating over small cell networks and proposed a virtual reality model based on multi-attribute utility theory, and simulation test results showed that the model has faster convergence than the traditional model and provides lower latency of virtual reality services [6]. Thies J et al. proposed an image processing algorithm based on virtual reality technology to build a more realistic virtual meeting environment in video teleconferencing [7]. Du et al. found that virtual reality technology has been gradually and widely used in the architecture, engineering, construction, and facilities management industries in recent years, because it can improve the workflow of teams by enhancing the consensus of team members. However, the current virtual reality application suffers from the difficulty of generating a virtual reality dialogue environment from input data. Therefore, the research team designed a synchronous communication system based on virtual reality technology, and the test results showed that the system has higher utilization value and communication efficiency than traditional virtual reality communication systems [8].

Cga B et al. found that obtaining accurate information about students' posture during learning is important for assessing students' learning status and improving teaching methods, but the presence of various occlusion elements in the educational environment makes it more difficult to carry out current student posture detection. So this study proposes a new posture detection method based on a single-stage object detector, which uses an adaptive fusion mechanism to learn complementary. This method uses an adaptive fusion mechanism to learn complementary spatial features to make the feature extractor more discriminative. Experimental results show that the student pose detection capability of the method is significantly better than other single-stage target detection methods on a real classroom pose dataset [9]. Lin G et al. found that fruit detection is necessary for automatic guava harvesting under real outdoor conditions, and that the branch-related pose of the fruit is also essential to guide the robot to approach and separate the target fruit without colliding with its parent branch. In order to perform automatic, collision-free harvesting, this research team designed a fruit detection and pose estimation method based on sensors with four channels of red, green, blue, and depth data images from cell phones. Quantitative experimental results show that the method has an accuracy and recall of 0.983 and 0.948 for guava fruit detection, respectively, which can be applied to improve the control of automatic guava harvesting machinery [10]. Tang L's research team found that it is difficult to detect the bit pose of small targets under poor imaging conditions such as severe occlusion and low resolution, and proposed a bit pose detection algorithm that combines merged region of interest pools and local static. Therefore, a pose detection algorithm combining merged region of interest pool and local static learning is proposed. Experimental results show that the method outperforms existing methods for pose recognition [11]. Zhou L et al. found that existing fabrication methods cannot produce flexible sensors that match the shape of soft robots. Therefore, a new 3D printed multi-functional inductive flexible stretchable liquid metal sensor was designed using the posture detection algorithm, and the test results showed that this sensor has a higher perception of deformation than the conventional sensors, which also means that the posture detection algorithm improves the manufacturing accuracy of the sensor components [12].

In summary, it can be found that virtual reality and posture detection techniques have been increasingly applied to industrial and educational fields, and some economically valuable and educational results have been achieved. However, studies combining virtual reality and posture detection technology to build a swimming stance recognition model with high recognition accuracy are still rare, so it is really necessary to use virtual reality technology to reduce the noise of training images and improve the data scaling effect when building this teaching aid model, while the posture detection work can help swimming beginners to learn swimming posture better, which is the reason for conducting this study.

III. Design of Swimming Position Detection Algorithm for Swimming Scene

A. Design of Bit Pose Detection Algorithm Incorporating Sliding Window Detection and Convolutional Neural Network

The dataset used for training and testing the pose detection algorithm in this study is a virtual model image of a swimming scene and its swimmer's swimming pose constructed by virtual reality technology. Before designing the pose detection algorithm, it is necessary to design the pre-processing process of the dataset [13-15]. The preprocessing of virtual image of swimmer's pose mainly includes image cropping, grasping rectangle frame extraction, image normalization and homogenization [16]. In order to reduce the content of irrelevant information in the image and the computational effort of the algorithm, the virtual model image needs to be cropped to a uniform size (640 480 size is chosen here) [17]. The purpose of homogenization and normalization of the virtual model images is to eliminate the negative impact of different inter-feature scales in the image data on the training results. The range of RGB (Red Green Blue) channel values in the image data is [0,255], and the normalization operation can be completed according to equation (1) to compress the range of values to [-1,1].

$$I_{i,j,k}^{(n)'} = \frac{2I_{i,j,k}^{(n)} - 255}{255} \quad (1)$$

$I_{i,j,k}^{(n)}$ represents the pixel values of row, column and channel numbers i, j and k positions in the n image before normalization. $I_{i,j,k}^{(n)'}$ represents the pixel values of row, column and channel numbers i, j and k positions in the n image after normalization. When constructing a virtual model of the swimmer using virtual technology, the original physical image is a depth map, and the depth value is used to describe the distance between the camera and the photographed object. But the index is mainly distributed in the middle part of the image, which needs to be mapped and processed according to the camera desktop distance and the range distance, see equation (2).

$$d = \begin{cases} 1.4m, d < 0.8m \\ 0.8m, d > 1.4m \\ d, d \in [0.8m, 1.4m] \end{cases} \quad (2)$$

d is the depth value after the mapping process, and m is the ratio of the camera range distance to the camera desktop distance. Then use equations (3) and (4) to normalize and homogenize d.

$$d_i = \frac{d_i - d_{\min}}{d_{\max} - d_{\min}} \quad (3)$$

$$d_i = d_i - \frac{\sum_{i=0}^{n} d_i}{n} \quad (4)$$

In equation (3), d_i is the depth value of the i row, $d_{\min}$ and $d_{\max}$ are the minimum and maximum values of the depth value of the i row, and n is the number of pixels of the i row. This time, the sliding window method is chosen to generate the rectangle to be selected. However, the number of selected rectangles generated by the image is too large, which makes the algorithm computationally inefficient, so it is necessary to filter out some of the rectangles. First of all, the background image can be removed using the background difference method, and then the image is divided into a grid of specified distances, and the detection frame containing the smallest object is set as the detection range of the sliding window, and the center point of the rectangular frame can only be in the intersection of the restricted range [18]. According to the size statistics of the dataset, it is need to set the pixel point taking steps for the rotation angle θ, length h, and width w of the grasping rectangle, and the range of values which are shown in equations (5), (6), and (7).

$$\theta = 0° : 15° : 180° \quad (5)$$

$$h = 10 : 10 : 90 \quad (6)$$

$$w = \min(10, h-50) : 10 : \max(90, h+50) \quad (7)$$

Equation (5) represents the interval θ in the range $0°$ to $180°$ taken by $15°$, and the meanings of equation (6) and equation (7) can be analogously derived in turn. Since convolutional neural network can generalize and extract features in images with high recognition accuracy, this study chooses it to build a swimming pose detection model based on the mature AlexNet, and the core hierarchy in AlexNet is designed below. In general, if there is an input image of size $n \times n \times n_c$ and its convolutional kernel size is $f \times f \times n_c$. n_c and f are the number of image channels and convolutional kernel size respectively, and the convolutional step parameter is s, then the size of the output value of the convolutional layer can be obtained by equation (8).

$$((n+2p-f)/s+1) \times ((n+2p-f)/s+1) \times m_c \quad (8)$$

p and m_c represent the size of the edge filling layer and the total number of convolutional kernels used, respectively. The activation functions of the convolutional layers are used to perform nonlinear mapping of the output elements of each layer. The commonly used activation functions are relu function, elu function, tanh function, and sigmoid function. However, for neural networks with more layers, using the tanh function and sigmoid function as the activation function may cause the gradient disappearance problem and reduce the overall convergence speed of the algorithm, so the relu function and elu function are more widely used. The function graphs of the above four activation functions are shown in Fig. 1.

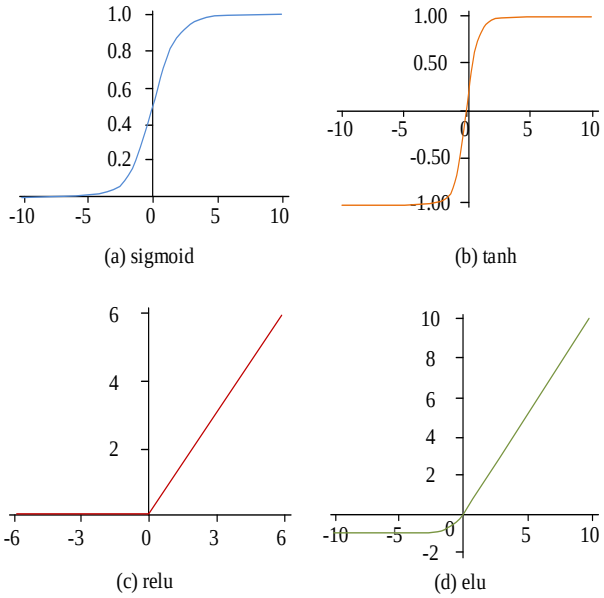

Fig. 1. Function curve of common activation functions

The role of pooling layer in AlexNet is to extract the high-latitude features in the images and reduce the size of the dataset. Depending on the processing method, there are two types of pooling layers, average pooling and maximum pooling. The maximum pooling method is chosen to construct the pooling layer in AlexNet in this study, and the calculation method of maximum pooling is shown in Fig. 2.

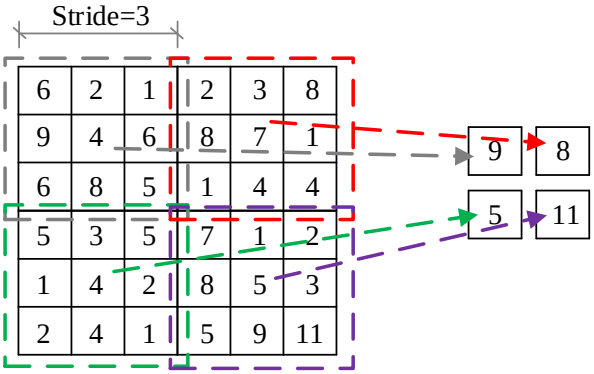

Fig. 2. Schematic diagram of the calculation rules for maximum pooling

According to the content above, based on the simplified AlexNet, a neural network is designed to judge whether the swimmer's swimming posture in the generated rectangular box is accurate. The input is the preprocessed RGD (Red Green Depth) three channel data, and the computational flow of the algorithm is shown in Fig. 3.

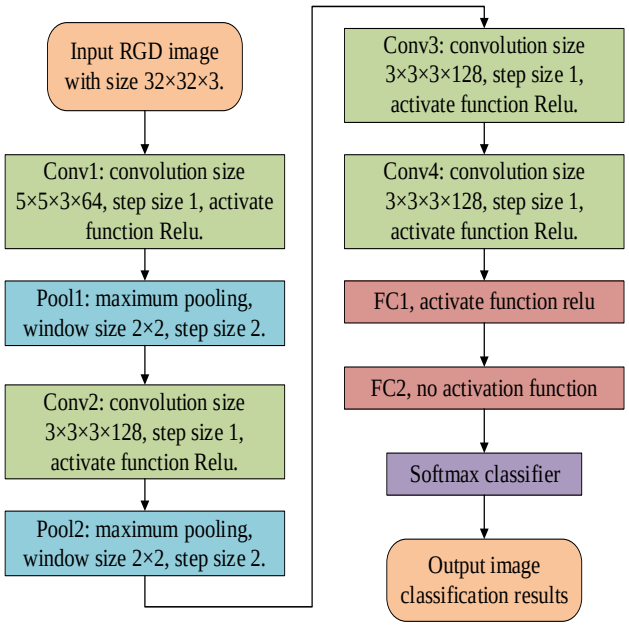

Fig. 3. Flow chart of the bit pose detection algorithm combining sliding window detection and convolutional neural network

As shown in Fig. 3, the algorithm contains four convolutional layers, two maximum pooling layers and two fully connected layers. The size of the convolutional kernel is 5×5, and the output data size is $16\times16\times64$ after the maximum pooling process, and then the output data size is $8\times8\times128$ after three convolutional layers with the convolutional kernel size of 3×3 and the maximum pooling layer. The size of the output data is adjusted to 2. Finally, the data is input to the softmax classifier for binary classification, and the judgment result of whether the object pose of the rectangular box in the input image data is correct is output.

B. *Design of Swimming Pose Detection Algorithm based on Deep Residual Network*

The pose detection algorithm designed in this study, which combines sliding window detection and convolutional neural network (SW-CNN), has many layers of networks and complex processing, so it may have a problem of slow detection speed in actual use, which can't meet the timeliness requirements of its application in swimming teaching. Therefore, in order to further improve the recognition accuracy of virtual swimming stance images, a virtual model recognition algorithm of swimmer's stance that integrates end-to-end idea and convolutional neural network (PTP-CNN) is proposed. Unlike the SW-CNN algorithm, the former trained neural network can directly reflect the RGB image and its corresponding standard swimming pose mapping relationship, so the input data of the PTP-CNN algorithm in the training phase is the image and the corresponding standard swimming pose information in the figure. In addition, since the PTP-CNN algorithm removes the sliding window detection link, and the learning ability of the neural network with the requirement of diversity of input image data is increased. Therefore, migration learning and data augmentation, deep residual network methods are needed to learn swimming pose features.

Due to the inefficiency of constructing virtual models based on physical images of swimmers' swimming stance and the complexity of computational processing, the data set that can be used for training the algorithm is often small in number and cannot well meet the training requirements of the algorithm, and overfitting may even occur. Therefore, before designing the PTP-CNN algorithm, it is necessary to expand the dataset and increase the size of the dataset appropriately through data augmentation techniques to make up for the lack of data and the problem that the data diversity and feature information are relatively single, so that the neural network can learn to master more important image features. In this study, considering the type of task of the algorithm, we choose to use data augmentation methods that do not change the features themselves but only the pixel distribution to process the dataset by flipping transformation, rotation transformation, translation transformation, contrast transformation, color transformation, sharpening, adding noise disturbance, adding blurred information, etc. The program selects one or more measures to process the original dataset in a random way.

Even with data augmentation, the total number of samples in the dataset is still relatively limited, which makes it difficult and inefficient to obtain more data. And using the existing dataset directly for training may lead to overfitting of the model. In addition, training the whole neural network algorithm from scratch will consume a lot of computer and time resources. Therefore, the PTP-CNN algorithm is built using the migration learning approach. There are three common applications of transfer learning, as a feature extractor for the algorithm to be trained, as a pre-trained model for the algorithm to be trained, and as an aid for fine-tuning the model parameters. The most central step in the algorithm training here is to improve the image feature extraction capability of the algorithm, therefore, the convolutional neural network trained with Imagenet dataset is selected as the image feature extractor in PTP-CNN algorithm. Then the fully connected layer is set after the convolutional neural network, and the output of the last layer is changed to the position and stance parameters that measure the swimmer's stance. That is, the parameters of the layers within the migrated convolutional neural network remain unchanged, and only the parameters of the fully connected layer behind are fine-tuned to achieve migrated learning.

When the neural network has too many levels, its recognition accuracy will be degraded, and some scholars have proposed the use of central normalization and normalized initialization methods, which can solve the problem to some extent, but the effectiveness of these methods is also poor if the neural network has too many levels, and the residual neural network is more effective in dealing with the problem. In this study, a residual network module constructed by constant mapping is introduced into the ResNet network of PTP-CNN algorithm to solve the degradation problem of deep convolutional neural networks.

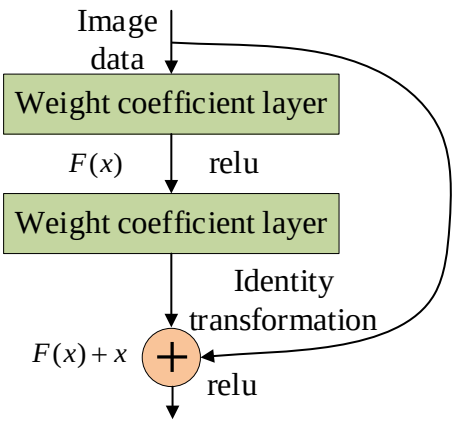

Fig. 4. Structure of residual neural network based on constant mapping

As shown in Fig. 4, the output of the residual module is obtained by superimposing the output of multiple convolutional layers and then performing ReLU processing, as shown in Equations (9) and (10).

$$F = W_2 \times relu(W_1 x) \quad (9)$$

$$y = F(x, W_1) + x \quad (10)$$

In equation (10) $F(x, W_1)$ is the residual mapping function constructed based on the input x, residual matching weights W_1, W_2, and y is the output data after the resultant residual processing. Here ResNet-50 containing 50 computational layers is used as the basis for forming the PTP-CNN algorithm, and the image data size of the input network needs to be adjusted to 224×224, then the convolutional kernel and input feature data of each core level in the network are shown in Table I.

TABLE I. CONVOLUTIONAL KERNEL AND INPUT FEATURE PARAMETERS AT THE CORE LEVEL IN THE RESNET-50 NEURAL NETWORK

Layer Name Type	Output Data Size	Number of computing layers	Convolution kernel or pooling window size	Number of convolution kernels
C1	112×112	1	7×7	64
P1	112×112	1	3×3 (maximum pooling)	/
C2_X	56×56	3	1×1, 3×3, 1×1	64, 64, 256
C3_X	28×28	4	1×1, 3×3, 1×1	128, 128, 512
C4_X	14×14	6	1×1, 3×3, 1×1	256, 256, 1024
C5_X	7×7	3	1×1, 3×3, 1×1	512, 512, 2048
P5	1×1	4 (average pooling, fully connected, softmax)	3×3	/

After completing the above preparatory work, we started to formally design the virtual model recognition model of swimmer's stance based on PTP-CNN algorithm, the input image in this model needs to be pre-processed and data augmented, the feature extractor in the model is ResNet-50 convolutional neural network trained by ImageNet dataset, followed by docking pressure leveling layer, deletion layer, SoftMax, and finally, the overall structure of the model is shown in Fig. 5.

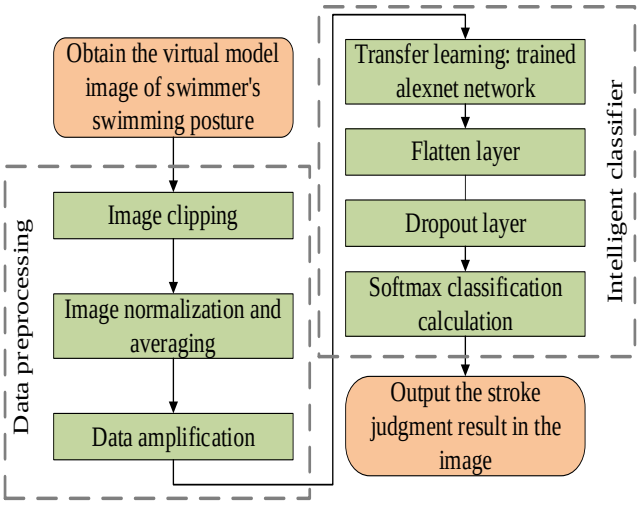

Fig. 5. Structure of virtual image recognition model of swimmer's position pose based on PTP-CNN algorithm

As shown in Fig. 5, the last layer of the standard ResNet network is the SoftMax classifier. But, in order to downscale the dimensionality of the data by one dimension for classification, a flattening layer needs to be added before, and a deletion layer is subsequently added to prevent overfitting of the model. Finally, the PTP-CNN algorithm is trained by using the cross-entropy loss function $loss(x)$, which is given in equation (11).

$$loss(x) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \ln y_{ip} + (1-y_i)\ln(1-y_{ip})] \quad (11)$$

In equation (11), y_i and y_{ip} are the label value and algorithm prediction of the image i, respectively.

IV. PERFORMANCE VERIFICATION EXPERIMENT OF SWIMMING POSTURE DETECTION ALGORITHM

A. Model Parameter Setting and Performance Verification Experimental Design

The experiments were designed to verify the recognition performance of the two virtual image recognition models of swimmer's stance designed in this study, and to build a comparative classification model based on the common SVM (Support Vector Machine) algorithm and AlexNet neural network. The data set used for model training is the images of different swimming stances purchased by the research team from a third-party data agency in China, and the images are labeled as "correct stance" and "wrong stance". The data set

contains 1843 images, and after data expansion, it reaches 15442 images, which are divided into training set and test set (containing 4632 images) according to the ratio of 7:3. The model based on SW-CNN algorithm also uses cross-entropy loss function and Adam optimization method, and the learning rate is determined to be 0.001 after multiple debugging, and the exponential decay rates of β_1 and β_2 for first-order and second-order moment estimation are 0.9 and 0.998, respectively, and the batch size is 64. The model based on PTP-CNN algorithm is also optimized using Adam, and the learning rate is determined to be. The exponential decay rates of β_1 and β_2 are 0.88 and 0.999 for first-order and second-order moment estimation, respectively, and the batch size is set to 32. The common evaluation indexes of image classification accuracy are Accuracy, Precision, Recall, and F1 value, which are calculated in equations (12) to (15).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{14}$$

$$F1 = \frac{2 * Precision * \mathrm{Re}call}{\left(Precision + \mathrm{Re}call\right)} \tag{15}$$

In Eqs. (12)~(15), TP represents the number of samples that are actually positive classes are predicted to be positive classes, referred to as the true positive number, and so on TN, FP, FN are the true negative number, false positive number, and false negative number, respectively. Considering the importance of identifying correct and incorrect postures in swimming teaching, the accuracy rate was chosen as the performance evaluation index of each model.

B. Analysis of Experimental Results

After completing all the performance test experiments, the Excel software and SPSS23.0 software were used to count the changes of the loss function values during the training process for the virtual image recognition models of swimmer's position pose constructed based on SW-CNN algorithm and PTP-CNN algorithm designed in this study, and the comparison models constructed based on SVM algorithm and AlexNet algorithm, as shown in Fig. 6.

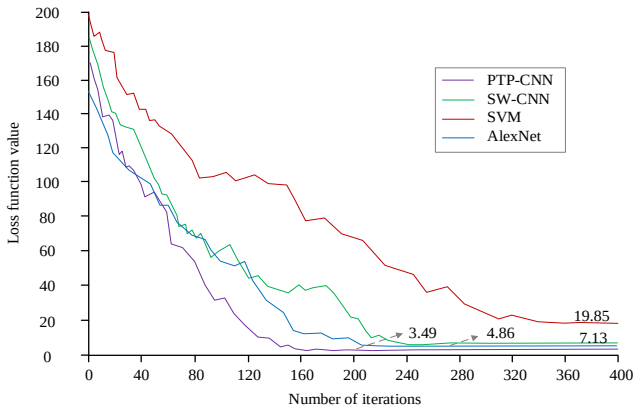

Fig. 6. Loss function value change curve of virtual image recognition model for each swimmer's position pose during training

In Fig. 6, the horizontal axis represents the number of iterations of the core algorithm in each model, the vertical axis represents the loss function value of each algorithm, and different colors represent different virtual image recognition models of swimmer's position pose. As we can see in Fig. 6, the model built based on PTP-CNN algorithm has the fastest convergence speed during the training process and the smallest loss function value after convergence, while the model built based on SVM algorithm has the slowest convergence speed and the largest loss function value after convergence. Specifically, when all the models finished training (i.e., the loss function values converged), the loss function values of the virtual image recognition models constructed based on the PTP-CNN algorithm, AlexNet algorithm, SW-CNN algorithm, and SVM algorithm were 3.49, 4.86, 7.13, and 19.85, respectively. The detection accuracy of each model on the test set is analyzed below. The statistical results are shown in Fig. 7.

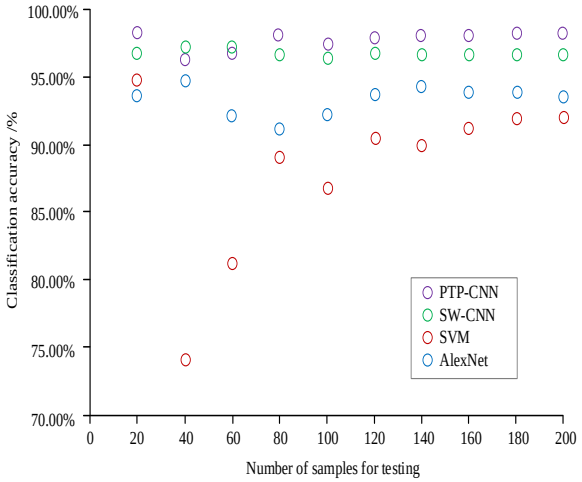

Fig. 7. Recognition accuracy of virtual image recognition models for each swimmer's position pose on the test set

The horizontal axis in Fig. 7 represents the number of samples used to test each model, and the vertical axis represents the classification accuracy of each model on image data in %, with two valid digits retained. Since the accuracy of each model changes slightly after the number of samples exceeds 200, only the test results before the number of test samples is not higher than 200 are retained here. It can be seen from Fig. 7 that the classification accuracy fluctuation degree of the model changes with the number of test samples. The descending sorting results are SW-CNN, PTP-CNN, AlexNet and SVM. However, from the perspective of the overall classification accuracy, the descending sorting results are PTP-CNN, SW-CNN, AlexNet and SVM. When the number of test samples is 200, the classification accuracy of each algorithm is 97.48%, 96.72%, 93.60% and 91.39%, respectively. The image recognition speed of each trained model with different test samples is then analyzed and shown in Fig. 8.

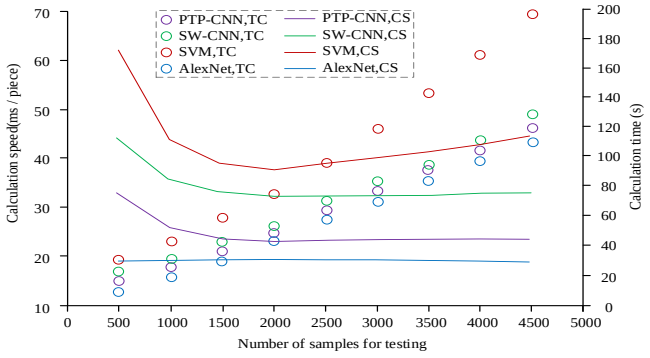

Fig. 8. Computational speed of the virtual image recognition model for each swimmer's position pose on different number of test samples

In Fig. 8, the horizontal axis is the number of samples used to test the model, and the test step size is 500 samples in line with the scale. The circle and "TC" represent the computation time, and the data line and "CS" represent the computation speed. As can be seen in Fig. 8, when the training samples are small, the computational speed of each model is high because the algorithm takes a relatively constant time to start. From the model perspective, the model based on the SVM algorithm has the slowest computation speed due to the high computational complexity of its own algorithm, while the computation speed of the other three models remains roughly stable as the number of samples grows. When the number of test samples is 4500, the computation time of the models built based on PTP-CNN algorithm, AlexNet algorithm, SW-CNN algorithm and SVM algorithm is 24 ms/sheet, 19 ms/sheet, 34 ms/sheet and 45 ms/sheet, respectively. Finally, the computer memory occupancy of each algorithm during operation is analyzed in Fig. 9. It is need to note that five parallel experiments were conducted for each experimental scheme to ensure the reliability of the test results.

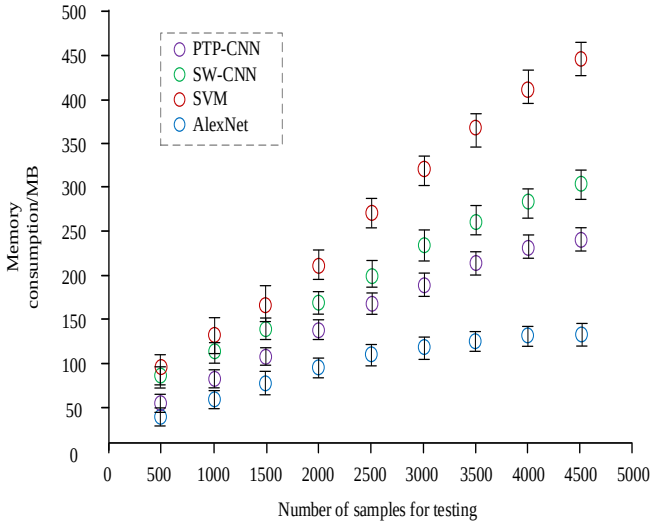

Fig. 9. Memory consumption of the virtual image recognition model computed for each swimmer's position pose

In Fig. 9, the horizontal axis is the number of test samples, the vertical axis is the computational memory consumption in MB, and the vertical coordinate of the center point of the circle in the figure is the mean computational memory consumption of multiple experiments, and the upper and lower limits represent the upper and lower deviation of memory consumption. As we can see in Fig. 9, the computational memory consumption of each model tends to increase as the number of test samples increases, but the memory consumption of the model built by the SVM algorithm grows the fastest, the memory consumption of the model built by the AlexNet algorithm grows the slowest, and the model built by the PTP-CNN algorithm grows the second. When the number of test samples is 4500, the memory consumption of the models constructed based on the PTP-CNN algorithm, AlexNet algorithm, SW-CNN algorithm, and SVM algorithm are 225MB, 105MB, 269MB, and 438MB, respectively.

V. Discussion

For the purpose of helping swimmers learn swimming postures by themselves, this research uses sliding window and depth residual network to build a hybrid neural network algorithm, and based on this algorithm, two corresponding virtual image classification models of swimmers' postures are designed. At the same time, in order to reduce the noise of the input data and reduce the cost of data collection, virtual reality technology is used to convert the swimming pose image of swimmers into an image model in the virtual reality space as the input data of the algorithm. The performance test experiment results show that the model based on PTP-CNN algorithm designed in this study is the first to complete convergence in the training phase, and the loss function value

after convergence is also significantly lower than all the comparison models, which is 3.49. The research results of X. Tong et al. also show that the convergence speed of the pose detection neural network incorporating sliding window technology in the training phase is significantly faster than that of the neural network before improvement [19]. From the perspective of the pose recognition accuracy index of the test set, the classification accuracy of the model built based on the PTP-CNN algorithm designed in this study is 97.48%, which is the highest among all the comparison models, while the classification accuracy of the recognition model built based on the SW-CNN, AlexNet and SVM algorithms is 96.72%, 93.60% and 91.39%, respectively. The research results of J. M. Liang et al. also show that the use of virtual reality technology to convert teaching material images into virtual reality models is conducive to improving the work progress of the teaching assistance system [20]. From the perspective of computing efficiency, when the number of test samples is 4500, the computing speed and memory consumption of the model based on PTP-CNN, SW-CNN, AlexNet and SVM algorithms are respectively 24 ms/piece, 34 ms/piece, 19 ms/piece, 45 ms/piece and 225 MB, 269MB, 105 MB, 438 MB. This is mainly because the deep residual network consumes more computing resources than the general convolutional neural network, slowing down the computing speed of the entire model.

VI. Conclusion

To solve the problem of swimming teaching, beginners' learning progress of swimming skills is affected due to insufficient knowledge of correct swimming posture. This study uses virtual reality technology to convert physical images of swimmers' swimming posture and position into models in virtual space, and constructs two hybrid convolutional neural network-based image classification models for virtual models of swimmers' position posture. The performance test results show that the model based on PTP-CNN algorithm designed in this study has the fastest convergence speed in the training phase, and the loss function value after convergence is the smallest than the comparison model, 3.49. The classification accuracy of the models based on PTP-CNN, SW-CNN, AlexNet and SVM algorithms on the test set is 97.48%, 96.72%, 93.60% and 91.39% respectively. When the number of test samples is 4500, the calculation speed and memory consumption are 24 ms/piece, 34 ms/piece, 19 ms/piece, 45 ms/piece and 225 MB, 269MB, 105 MB and 438 MB, respectively. The experimental data shows that the model based on PTP-CNN algorithm and SW-CNN algorithm designed in this study is higher than the common neural network algorithm and machine learning algorithm in the virtual image recognition accuracy of swimmer's swimming posture, but the former has faster computation speed and consumes less resource for computation. However, due to the limitation of experimental conditions, further classification and recognition of swimmer's swimming posture by joints were not carried out, which could help learners to use the recognition model in more cases, and this is also a point for improvement in the subsequent research.

Acknowledgment

This research is supported by: Sichuan International Studies University used the teaching reform project (No. JY2296285) of 2022: college students used PDCA to overcome sports acquired helplessness; Research project of Sichuan International Studies University, Investigation and analysis of college sports coaches and athletes' Sense of Gain in the new era (No. sisu2018074).

References

[1] Y. Zhang, S. Mi, J. Wu, and X. Geng. "Simultaneous 3D hand detection and pose estimation using single depth images," Pattern Recognition Letters, vol. 140, no. 9, pp. 43-48, 2020.

[2] R. Jin, J. Jiang, Y. Qi, D. Lin and T. Song. "Drone detection and pose estimation using relational graph networks," Sensors, vol. 19, no. 6, pp. 1479-1498, 2019.

[3] Hu Z, Xing Y, Lv C, et al. "Deep convolutional neural network-based bernoulli heatmap for head pose estimation," Neurocomputing, vol. 436, no. 5, pp. 198-209, 2021.

[4] Py A, Jy A, Gl B, et al. "Graph neural network for 6D object pose estimation," Knowledge-Based Systems, vol. 218, no. 4, pp. 106839.1-106839.9, 2021.

[5] M. S. Elbamby, C. Perfecto, M. Bennis and K. Doppler. "Towards low-latency and ultra-reliable virtual reality," IEEE Network, vol. 32, no. 2, pp. 78-84, 2018.

[6] M. Chen, W. Saad and C. Yin. "Virtual reality over wireless networks: quality-of-service model and learning-based resource management," IEEE Transactions on Communications, vol. 66, no. 11, pp. 5621-5635, 2018.

[7] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, M. Nießner. "FaceVR: Real-time facial reenactment and eye gaze control in virtual reality," ACM Transactions on Graphics, vol. 37, no. 2, pp. 1-15, 2018.

[8] J. Du, Z. Zou, Y. Shi, D. Zhao. "Zero latency: real-time synchronization of BIM data in virtual reality for collaborative decision-making," Automation in Construction, vol. 85, pp. 51-64, 2018.

[9] C. Cao, S. Ye, H. Tian and Y. Yan. "Multi-scale single-stage pose detection with adaptive sample training in the classroom scene," Knowledge-Based Systems, vol. 222, no. 6, pp. 107008-107013, 2021.

[10] G. Lin, Y. Tang, X. Zou, J. Xiong and J. Li. "Guava detection and pose estimation using a low-cost RGB-D sensor in the field," Sensors, vol. 19, no. 2, 2019.

[11] L. Tang, C. Gao, X. Chen and Y. Zhao. "Pose detection in complex classroom environment based on improved Faster R-CNN," IET Image Processing, vol. 13, no. 3, pp. 451-457, 2019.

[12] L. Zhou, Q. Gao and J. F. Zhan. "3D printed wearable sensors with liquid metals for the pose detection of snakelike soft robots," ACS Applied Materials & Interfaces, vol. 10, no. 27, pp. 23208-23217, 2018.

[13] X. Li, Z. Fan, Y. Liu and Q. Dai. "3D pose detection of closely interactive humans using multi-view cameras," Sensors, vol. 19, no. 12, pp. 2831-2836. 2019.

[14] Q. Gao, J. Liu, Z. Ju and X. Zhang. "Dual-hand detection for human-robot interaction by a parallel network based on hand detection and body pose estimation," IEEE Transactions on Industrial Electronics, vol. 66, no. 12, pp. 9663-9672, 2019.

[15] Hu W and Guan Y. "Landmark-free head pose estimation using fusion inception deep neural network," Journal of Electronic Imaging, vol. 29, no. 4, pp. 43030.1-43030.11, 2020.

[16] C. Chen, T. Wang, D. Li and J. Hong. "Repetitive assembly action recognition based on object detection and pose estimation – ScienceDirect," Journal of Manufacturing Systems, vol. 55, no. 5, pp. 325-333, 2020.

[17] S. Qi, S. Li and J. Zhang. "Designing a teaching assistant system for physical education using web technology," Mobile Information Systems, no. 6, pp. 1-11, 2021.

[18] L. Zhao and Y. Zhao. “The construction of the fusion and symbiosis path of infant sports development based on intelligent environment,” Mathematical Problems in Engineering, 2021, 2021(3):1-9.

[19] X. Tong, R. Li, L. Ge, L. Zhao, K. Wang. “A new edge patch with rotation invariance for object detection and pose estimation,” Sensors vol. 20, no. 3, pp. 1-17, 2020.

[20] J. M. Liang, W. C. Su, Y. L. Chen, S. L. Wu, J. J. Chen. “Smart interactive education system based on wearable devices,” Sensors, vol. 19, no. 15, pp. 3260-3267, 2019.