IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received September 16, 2019, accepted September 30, 2019, date of publication October 3, 2019,
date of current version October 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2945313

Intelligent Recognition of Medical Motion Image Combining Convolutional Neural Network With Internet of Things

YUCHENG ZHOU[1] AND ZHIXIAN GAO[2]
[1]Department of Sports, Chongqing Jiaotong University, Chongqing 400074, China
[2]School of Electronic and Information Engineering, Henan Institute of Technology, Xinxiang 453003, China

Corresponding author: Yucheng Zhou (yczhou@cqjtu.edu.cn)

ABSTRACT For the small-scale motion in medical motion images, the traditional medical motion image intelligent recognition algorithm has low recognition accuracy, and requires a large amount of calculation statistics. There is no self-learning function, which seriously affects the accuracy and speed of medical motion image recognition. Therefore, in order to improve the accuracy of human body small-scale motion recognition in medical motion images and the computational efficiency of large-scale data sets, an intelligent recognition algorithm based on convolutional neural network for medical motion images is proposed. The algorithm first learns the dense trajectory features and depth features, and then further fuses the dense trajectory features with the deep learning features. Finally, the extreme learning machine is applied to the convolutional neural network, and the fused features are further trained as input information of the convolutional neural network, and the features from the bottom layer to the upper layer can be extracted step by step from the raw data of the pixel level. Simulation experiments show that the algorithm can effectively improve the recognition accuracy of small-scale motion in medical moving images and improve the speed of motion.

INDEX TERMS Internet of things, convolutional neural network, medical motion image, intelligent recognition.

I. INTRODUCTION

The intelligent recognition of medical motion images is a cross-disciplinary field of comprehensive medical imaging, mathematical modeling, and computer technology [1], [2]. In the era of big data in medical motion images, massive and complex medical motion image data brings new problems in two aspects: On the one hand, the medical motion image data to be processed has higher dimensionality, and requires a model with stronger learning adaptability; On the other hand, for small-scale motion in medical motion images, precise recognition [3]–[5] is required. The traditional intelligent recognition algorithms for medical moving images often fail to meet people's requirements. Therefore, in the era of medical big data, how to deal with high-dimensional data of medical moving images and better recognition of small-scale movements in medical moving images has become an academic world. And research hotspots in industry [6], [7].

Many traditional medical motion image recognition algorithms have been studied, including template matching [8], statistical recognition [9], and fuzzy sets [10] and so on. Traditional medical motion images contain rich colors, edges, etc., but due to complex background, variable light, occlusion, angle of view and other factors, the accuracy of motion recognition algorithms is not high. In literature [11], a motion recognition method based on depth image is designed. The algorithm projects the depth image in three projection planes, extracts Gabor features from the three projection maps, and uses these features to train the extreme learning machine classifier. The algorithm has higher computational efficiency, but it is not ideal for small-scale motion recognition. Zhou proposed a time-series deep confidence network that can complete online human motion image recognition [12]. This model solves the current deep confidence network model only. The problem of static images can be identified, but the

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

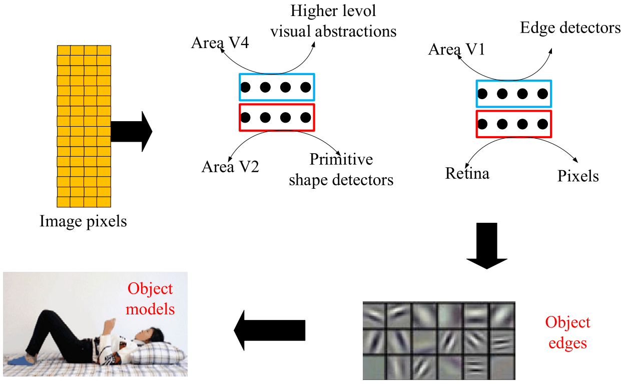

FIGURE 1. **Human brain vision mechanism.**

processing time of the model training process is long, which affects the application performance of the algorithm. With the development of convolutional neural networks, convolutional neural networks have become the preferred method to solve the task of extracting high-dimensional features. It can extract deep and abstract features in the data, and can capture long-range dependencies in the data in an efficient way, and can effectively analyze and identify medical moving images [13], [14]. In recent years, many research teams have tried to apply convolutional neural networks to medical motion image recognition. The study found [15], [16] that, convolutional neural networks are similar to the visual system of the human brain, and are a multi-layered, continuous abstract process. The original signal input to the brain is the lowest level of information. Some features are abstracted from this lower layer, and these features are taken as a new layer, and more abstract features are abstracted from this layer as a new layer. Iterations are repeated many times until the signal can be discerned. Figure 1 shows the process of human brain visual center observation.

Convolutional neural networks simulate the human brain visual center. By constructing a multi-layer network, the original input signals are continuously extracted from features until the features available to the classifier are abstracted. Fantoni *et al.* [17] first analyzed the medical image of the brain, and propose an image that can improve the state of the brain and thus recognize the image of the brain. Gao *et al.* [18] designed a learning structure for convolutional neural networks, which can classify the results of medical motion images in multiple categories. Madokoro *et al.* [19] proposed the use of both supervised and unsupervised learning algorithms to learn the characteristics of deep convolution images, so as to intelligently classify and identify medical moving images. In literature [20], feature extraction is performed by convolutional neural network, and generalized estimation equations are used to evaluate the recognition accuracy of medical moving images. Sumathi and Priya [21] proposed an automated convolutional neural network model that can analyze data from medical motion images. Liang *et al.* [22] used the tensor decomposition method to analyze the characteristic information of each patient's moving image, and then monitor the patient's indicators and physical health status. Hamedpour and Farnood Ahmadi [23] combined data mining methods with convolutional neural networks to analyze the data of medical moving images, and then mine features to intelligently identify medical moving images. In literature [24], the acceleration gesture trajectory map is established, and the trajectory features of unknown medical motion images are converted into low-dimensional sub-feature sequences, which are sent to the convolutional neural network for training, which improves the accuracy and time efficiency of medical motion images. Modeling human motion information in medical motion images through asymmetric system deviation, the algorithm introduces a gesture labeling mechanism to further improve the recognition performance of small-scale motion [25].

For the intelligent recognition of small-scale motion in medical motion images, the existing convolutional neural network model cannot analyze the relationship between the parts of the image and the nuances in the image, so it cannot describe the relationship between the small-scale motion and the whole image. The existing convolutional neural network model has low robustness and poor processing effect on high-dimensional image data, which limits the versatility and adaptability of the model and is difficult to implement in clinical applications.

In order to satisfy the intelligent recognition of high-dimensional data and small-scale motion at the same time, this paper designs an intelligent recognition algorithm based on convolutional neural network for medical motion images. The algorithm first uses the dense trajectory to extract the manual features. Secondly, the convolutional neural network is used to extract the convolutional neural network features of the motion information. Finally, the nuclear extreme learning machine is applied to the convolutional neural network, and the extracted manuals are applied. The feature and convolutional neural network features are merged and sent to the convolutional neural network for training. Compared with the classical convolutional neural network and the traditional medical motion image algorithm, the small-scale motion in the medical motion image can be better recognized.

Specifically, the technical contributions of our paper can be concluded as follows:

First: This paper presents an intelligent recognition algorithm for medical motion images based on convolutional neural networks. The model not only speeds up the convergence of the network, shortens the training time, but also significantly improves the small-scale motion in medical motion images.

Second: This paper applies the nuclear extreme learning machine to the convolutional neural network. By calculating the feature kernel of the manual feature and the convolutional neural network feature, and combining the two feature kernels to obtain a fusion feature kernel, the complementarity of the manual feature and the convolutional neural network feature can be utilized to describe the small human body in the medical moving image from different angles.

The rest of our paper was organized as follows. Theory of convolutional neural network was introduced in Section II. Section III described related theoretical derivation of the proposed system. Experimental results and analysis were discussed in detail in Section IV. Finally, Section V concluded the whole paper.

II. IDENTIFICATION BASIS OF MEDICAL MOTION IMAGES

A. THE BASIC PRINCIPLE OF MEDICAL MOTION IMAGE RECOGNITION

The recognition system based on medical moving images is generally divided into three parts [26]. The first part is the collection and acquisition of medical moving image information, and the second part is the processing and preprocessing of information. The three parts are the identification or classification process.

In a certain sense, for the medical target recognition task, the quality of the medical motion image feature extraction will play a vital role in the recognition result. And it will have a great influence on the calculation amount of the subsequent detection and recognition algorithm [27], [28]. The following are some common classifications of medical target recognition. The general medical target recognition method is mainly based on the following characteristics of the target:

1) Shape or structural features: The shape or structural features of the target can only be obtained when the medical moving target image has a high contrast and the target is within a certain distance. Generally, the shape or structural features of the target are based on Obtained by the binary image, these binary images are the image of the object region obtained after the image segmentation process, or the boundary of the object obtained by the edge extraction process. There are two ways to represent the shape or structure feature: one is a digital feature representation, and the other is a syntax language represented by a string and a graph.

2) Motion characteristics: By establishing the target motion model, the motion characteristics of the medical target can be obtained, and the moving target can be detected and identified, but the establishment of the target motion model is generally difficult. For the detection of medical sports targets, there are roughly two methods: feature recognition and motion-based recognition.

The feature recognition method consists of two steps:

a. First, extract features from two or more medical motion images at different moments, and establish correspondence;

b. The second is to calculate the structure and motion of the medical moving object based on the correspondence between these features. The advantage is that three-dimensional motion information can be acquired, and there is no limit to the target motion speed. The main difficulty lies in determining and extracting features. The motion-based recognition method is very different from it. It takes motion as the primary feature of the target, and generally adopts methods such as extracting optical flow field, inter-frame difference, and subtracting background.

3) Gray-scale distribution characteristics: On the basis of the original medical moving object image, the law of gray level change of the surface of the object is analyzed, and the gray-scale distribution features such as the texture feature of the object can be obtained.

However, we know that there is no such feature that can completely describe the characteristics of an object, and most of these features do not reflect the essential characteristics of the target, especially when the posture of the target changes or the surrounding environment changes. And change. This also puts higher requirements on the feature extraction of medical motion images.

B. INTELLIGENT RECOGNITION OF MEDICAL MOVING IMAGES BASED ON TEMPLATE MATCHING

The template matching method is to draw a typical standard template as a recognition standard for each category of the medical moving image to be recognized, and then compare the category of the medical moving image to be identified with the standard template to make a judgment. This comparison is achieved by correlating a standard medical motion image with the input medical motion image in a classifier. According to the knowledge of random signal analysis, the main peak occurs at the time of autocorrelation. The recognition image is matched with the standard medical motion image, and a threshold is used as a decision rule to achieve correlation matching recognition.

If f(x, y) is used to represent the input image that matches the medical motion image of a known template, where x, y represent image coordinates. F(x, y) represents a preset standard template sequence, and Q(x, y) represents the correlation. After comparing the output, the template matching method can be represented. Let the random variable be represented by x1, x2, y1, y2, and the output of the classifier is

$$\Phi\,(x_1-x_2, y_1-y_2) = \int\int f(x,y)F[(x+(x_1-x_2), y+(y_1-y_2))]dxdy \tag{1}$$

When x1= x2, y1= y2,

$$\Phi\,(0,0) = \int\int f(x,y)F(x,y)dxdy \tag{2}$$

That is, the autocorrelation function of the input information, and there are Φ(0,0) ≥ Φ(x,y), so that Φ(x,y) appears as the main peak at Φ(0,0), which may appear elsewhere. The secondary peaks can be identified by appropriate thresholds as long as they are not equal to the main peak. This template matching method can be performed either in the time domain or in the frequency domain, either in the entire image or in the same Find the medical motion image with the same direction and shape on the small template in the large image, which is to record the position of the target found in the large image.

The advantage of this method is that under certain conditions, its error probability and rejection rate are the smallest, which is suitable for places with less noise such as pattern matching method and text matching method. The disadvantage is that the storage system has high storage requirements and the calculation amount is large when identifying. Similarly, because it is sensitive to noise, it does not apply to text images.

C. INTELLIGENT RECOGNITION OF MEDICAL MOTION IMAGES BASED ON STATISTICS

The input medical motion image information is a digital signal that it can be recognized by the computer. The purpose of medical motion image processing is to filter the collected medical motion image, which is beneficial to extract feature information from the original signal that can reflect the essence of medical motion image; finally, identify the classification and output the recognition result.

A large number of medical motion image samples of various categories are required in determining the discriminant function, and the identification parameters, that is, the statistical learning process, are set by calculating statistics of the sample features.

Assuming that N features have been extracted after processing the medical moving image, and the medical moving image sample set is classified into m categories, the pattern identifying the medical moving image can be regarded as the vector X of the N-bit vector space, i.e. $X = [x_1, x_2, \ldots, x_N]^T$.

The category of medical motion images is $w_1, w_2, \ldots, w_m$. The content that is identified is to determine which class of the medical motion image X belongs to w_m. Among them, the medical motion image pattern recognition has two key problems: one is the feature extraction method and the feature selection method; the other is the discriminant function selection, such as: the mode has $w_1, w_2, \ldots, w_m$, a total of m categories, then there are $D_1(X), D_2(X), \ldots, D_m(X)$, a total of m Discriminant function. If X belongs to the i-th class, then there is

$$D_i(X) > D_j(X) \quad (j = 1, 2, \ldots m; j \neq i) \tag{3}$$

When $D_i(X) = D_j(X)$, it means that X belongs to both w_i and w_j. If the discriminant function fails, other features must be considered.

D. INTELLIGENT RECOGNITION OF MEDICAL MOTION IMAGE BASED ON FUZZY

Because many concepts in the objective world are not certain, they are vague. The classical set theory only reflects the concept of certainty. Given a set, any element belongs to this set only or does not belong to this set. For example, a collection represents cleverness, so it is stupid to not belong to this collection. Obviously, such a clear separation method is unrealistic and unscientific.

Whether an element x in the classical set theory belongs to a set E can also be characterized by the set function $f_\varepsilon(x)$: if $f_\varepsilon(x) = 1$, then $x \in E$. If $f_\varepsilon(x) = 0$, then $x \notin E$. And the number 1 means belonging to this class. The number 0 means not belonging to this class. Specifically, a function f(x) with a value of [0, 1] is used to characterize the qualification and degree of "aging" by the value of this function. Therefore, a function f(x) represents a fuzzy set, and a function that takes only two values of 0 and 1 corresponds to a classical set. Therefore, fuzzy sets are the generalization of classical set theory. In the fuzzy concentration, various concepts such as calculations and relationships can also be introduced, thus forming a set of theories, and using this set of theories to solve the problems in medical motion image recognition.

Fuzzy medical motion image recognition is actually the method or idea of introducing fuzzy logic in pattern recognition. Fuzzy pattern recognition has achieved good application in statistical pattern recognition. The application of fuzzy theory in medical motion image recognition system mainly uses fuzzy theory to blur and fuzzy classification of medical motion image features.

The fuzzy feature is actually to divide a feature or a group of features of the medical motion image into multiple fuzzy variables according to a certain fuzzy rule, so that each fuzzy variable can express a part of the characteristics of the original feature, and then replace the original with these new fuzzy features. The feature performs pattern recognition. Although the number of fuzzy features is larger than the original feature, it may linearize the relationship between the classification result and the feature value, which simplifies the design of the classifier and improves the performance of the classifier.

Fuzzy classification is actually to divide the sample space into several subsets, and these subsets are replaced by the concept of fuzzy subsets, so as to obtain the fuzzy classification result, that is, the fuzzy result of the classification result. A sample in a fuzzy classification will no longer belong to a specific category, but belong to a certain category with different degrees of credibility. The advantages are: 1) the classification results can reflect the uncertainty in the classification process. Use the user to make decisions based on trustworthiness; 2) if multi-level classification is used, this can provide classification information for the lower classification.

E. INTELLIGENT RECOGNITION OF MEDICAL MOTION IMAGE BASED ON CONVOLUTIONAL NEURAL NETWORK

With the development of artificial neural network technology, a new recognition method for medical motion image recognition technology has emerged: a medical image recognition method based on convolutional neural network. To some extent, the identification method of convolutional neural networks is different from the identification algorithm of medical moving images mentioned above, but to a certain extent, there are many similarities with these methods.

Convolutional neural networks are a convolutional neural network algorithm based on traditional artificial neural networks and the first learning algorithm to successfully train multi-layer networks. The weight sharing of convolutional neural networks reduces system parameters, which improves the performance of the algorithm. As a convolutional neural network architecture, convolutional neural networks propose

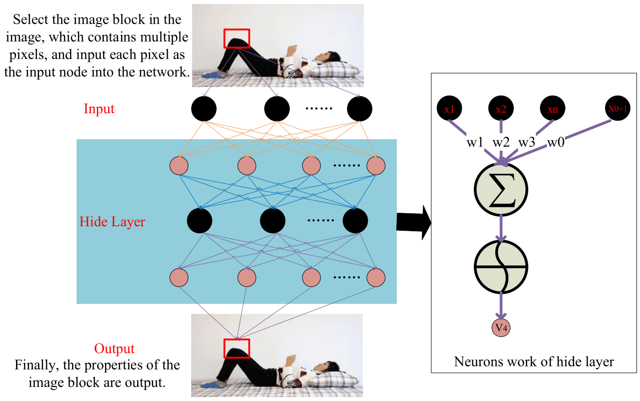
Select the image block in the image, which contains multiple pixels, and input each pixel as the input node into the network.
Input
Hide Layer
Output
Finally, the properties of the image block are output.
x1
x2
xn
X0=1
w1
w2
w3
w0
V4
Neurons work of hide layer

FIGURE 2. Convolutional neural network structure.

to reduce the preprocessing requirements for data. The input in convolutional neural networks is a small part of the original image, and the features of the previous layer are obtained by digital filtering or down-sampling [29], [30].

The convolutional neural network method can realize the recognition of medical moving images, which can deal with some environmental information is very complicated, the background knowledge is not clear, and the inference rules are not clear. The sample is allowed to have large defects and distortions, and the running speed is fast and the adaptive performance is obtained image with a higher resolution.

When a convolutional neural network is used for classification, it is first necessary to select various types of samples to train the network, and the number of each type of samples should be approximately equal [31]. The reason is that on the one hand, the network is prevented from being too sensitive to the category response of the sample after training, and is not sensitive to the category with a small number of samples. On the other hand, the training speed can be greatly improved, and the network can be prevented from falling into local minimum points.

Since the convolutional neural network does not have the ability of constant recognition, in order to make the network have the invariance of the rotation and expansion of the medical moving image, it is necessary to select samples of various possible situations as much as possible. For example, it is necessary to select representative samples such as different postures, different orientations, and different angles, so as to ensure a high recognition rate of the network [32].

The convolutional neural network is a directed acyclic network structure, which is formed by layering and interconnecting many perceptron; including input layer, hidden layer and output layer [33], [34]. The input layer directly accepts the data of the sample, passes through one or more hidden layers, and then propagates forward to the output layer. Figure 2 is a typical 3-layer neural network structure with an input layer, an implicit layer, and an output layer; each layer contains multiple neurons. Arrows represent the transfer of neuronal information between different levels.

It should be noted that the connection between neurons and neurons is not a simple weighting, but first it needs to be added to the offset, then weighted, and finally mapped by an excitation function.

In the figure, $x_1 \sim x_n$ are the input signals from the upper layer of network neurons, so that W_{ij} represents the weight from the upper layer neuron j to the layer of neurons i, and x_0 is offset. Expressing the input as X and the weight as W, then the relationship between the output of neuron i and the input can be expressed as:

$$net_i = XW \tag{4}$$

$$y_i = f(net_i) = f(XW) \tag{5}$$

Traditional neural network algorithms generally first design an algorithm for feature extraction, such as common HOG features, LBP features, and SIFT features, and then pass the extracted features to a trainable classifier to train the classifier. Finally, Import test samples into the classifier for classification. In this mode, since the extracted features are generally small, a fully connected multi-layer network can be designed as a classifier.

If the feature extraction is directly performed by the traditional neural network, the features of the hidden layer must not be too small, and the pixels of the incoming medical moving image tend to be relatively large. Suppose the input layer data has 1000 neurons, and the first hidden layer has 100 neurons. Since the neural network is fully connected, there are only more than 100,000 connection weights between the two layers. Even if the training speed is not considered, if there are not enough training samples, these parameters are difficult to fit the network; secondly, the fully connected network learns each sample, and the incoming of each sample will affect the parameter update, and the incoming There is often a great similarity between the data. For example, in medical motion image recognition, the difference between the two images may be only a small azimuth movement, and the fully connected network cannot capture this information, and cannot be trained according to the sample pair. The process is optimized, time consuming and laborious. In the convolutional neural network, the system makes full use of the local characteristics of the data through weight sharing and feature down-sampling to alleviate these problems.

III. INTELLIGENT RECOGNITION OF MEDICAL MOTION IMAGE BASED ON CONVOLUTIONAL NEURAL NETWORK

Although the intelligent recognition method of medical motion images mentioned above can solve the practical problems in the field of medical motion image recognition to a certain extent, these recognition methods require a large amount of calculation statistics, the recognition speed is slow, the discriminant function is difficult to select, and there is no self-learning function, seriously affected its application in the field of medical motion image recognition. Therefore, this paper proposes an intelligent recognition algorithm based on convolutional neural network for medical motion images, which not only improves the recognition rate of medical moving images, but also improves the recognition speed.

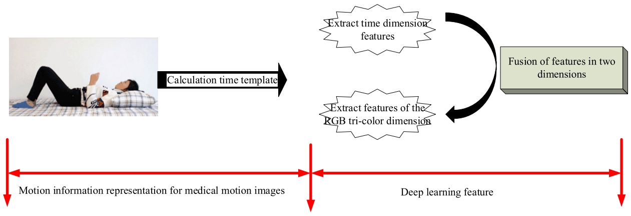
Calculation time template
Extract time dimension features
Extract features of the RGB tri-color dimension
Fusion of features in two dimensions
Motion information representation for medical motion images
Deep learning feature

FIGURE 3. Convolutional neural network process based on motion information.

A. CONVOLUTIONAL NEURAL NETWORK CHARACTERISTICS BASED ON MOTION INFORMATION

In this paper, a motion information representation scheme is proposed for the intelligent recognition of medical motion images, which emphasizes the saliency of different time domain motion information, thus improving the discriminability of small-scale motion in medical motion image sequences. The overall architecture of the module is shown in Figure 3.

The time template can extract the entire motion sequence of a medical moving image frame, so the motion recognition in this paper uses a time template. The calculation method of the time template is to calculate the weighted harmonic value of the medical moving image information, and calculate the motion information between the frames by using the difference between the image frames. The calculation formula of the time template TT is as follows:

$$T = (1/255)\sum w_i \bullet m(i) \tag{6}$$

where: n represents the number of frames, m(i) represents the motion information of the i-th frame, and wi represents the weight value of the i-th frame (set to the gray value), and 1/255 is to control the range of weights to [0, 255].

Transforming equation (6) to obtain the following formula

$$T = \sum (1/255) w_i \bullet m(i) \tag{7}$$

By replacing wi/255 of equation (7) with a fuzzy membership function μ(i), the following equation can be obtained.

$$T = \sum \mu(i) \bullet m(i) \tag{8}$$

where μ(i) represents a fuzzy membership function.

It can be seen from equation (8) that wi determines the saliency of the motion picture information assigned to the i-th frame in the time template. This mechanism can enhance the time domain motion in the time template by selecting the appropriate fuzzy membership function μ(i). Figure 4 is a graph of four membership functions, and four membership functions are set to $\mu 1$ to $\mu 4$, which are defined as equations (9) to (12), respectively. Where, the variable i indicates that the four membership functions are set to $\mu 1$ to $\mu 4$. And n represents the number of frames.

$$\mu_1(i) = 1, \quad \forall i \in [0, n] \tag{9}$$

$$\mu_2(i) = \frac{i}{n}, \quad \forall i \in [0, n] \tag{10}$$

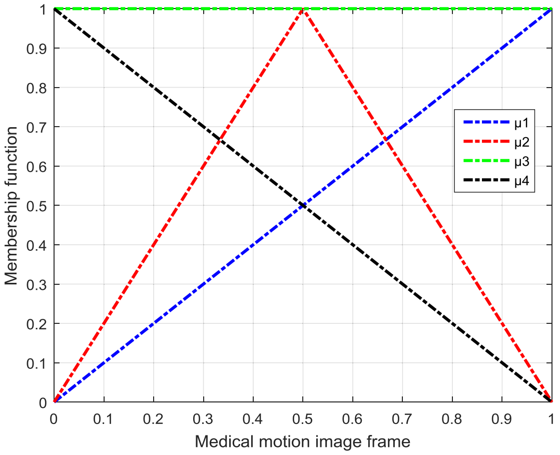

FIGURE 4. Fuzzy membership function graphics.

$$\mu_3(i) = 1 - \frac{i}{n}, \quad \forall i \in [0, n] \tag{11}$$

$$\mu_3(i) = \begin{cases} \frac{2i}{n}, & 0 < i \le \frac{n}{2} \\ 2 - \frac{2i}{n}, & \frac{n}{2} < i \le n \end{cases} \tag{12}$$

It can be observed from Figure 4 that $\mu 1$ calculates a motion energy image and $\mu 2$ calculates a motion history image. Since $\mu 1$ is a constant function, the motion energy image assigns equal weights to motion information for all time domains. And $\mu 2$ is a linear increasing function, so motion history image assigns the highest degree of saliency to the most recent medical motion image sequence. And $\mu 3$ is a linear decreasing function, so $\mu 3$ assigns the lowest saliency to the most recent medical motion image sequence. And $\mu 4$ assign the highest degree of saliency to the medical motion image sequence in the middle region of the time domain. Finally, the functions $\mu 2$, $\mu 3$, and $\mu 4$ emphasize the beginning, end, and middle of the time domain, respectively.

The convolutional neural network is used to learn the characteristics of medical moving images. The time template of the medical moving image sequence is input into the convolutional neural network to learn the feature set of the medical moving image recognition. This paper uses the CNN architecture of Figure 5 to extract convolutional neural network features.

The structure diagram of the convolutional neural network model proposed in this paper is shown in Figure 5. The model consists of three convolutional layers (C1, C2, and C3), three pooling layers (S1, S2, S3) and one full. The connection layer (F1) is composed.

The input to the network is a medical motion image containing R channel information, G channel information, and B channel information for each image block.

The output layer is the deep learning feature of the output.

Input layer: The input is a sample image of 640 × 480 pixels.

Convolutional layer C1: There are 4 convolution kernels with a size of 3 × 3, the step size is 2, and a convolution

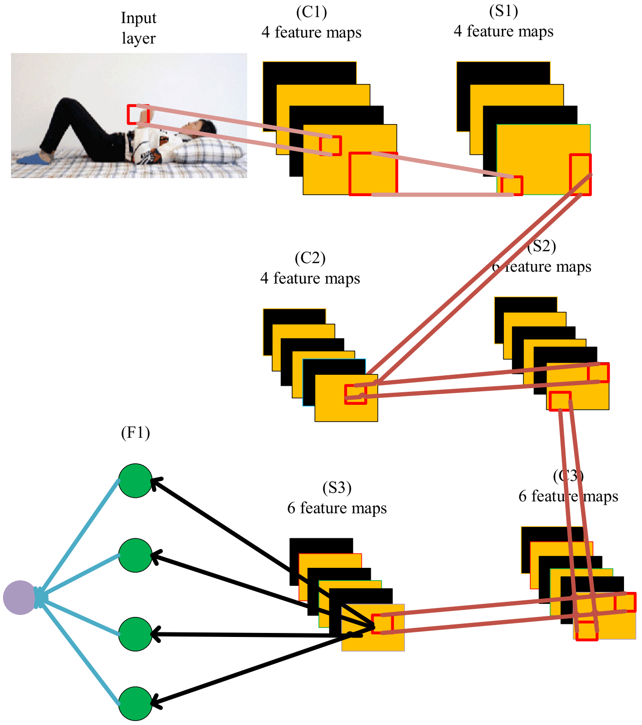

FIGURE 5. Convolutional neural network structure diagram used in this paper.

kernel will get a characteristic map, so the layer consists of 4 feature map maps.

The pooling layer S1:S1 adopts the average pooling strategy, which is obtained after the C1 layer is down-sampled. The size of the pooled area in S1 is 3 × 3, and the step size is 2. This layer consists of four feature maps with a size of 16 × 16.

Convolutional layer C2: There are six 4 × 4 convolution kernels with a step size of 1. After convolution, six feature maps with a size of 53 × 53 will be obtained.

Pooling layer S2: S2 adopts the average pooling strategy, which is obtained after the C2 layer is down-sampled. The pooled area has a size of 3 × 3 and a step size of 2, and consists of six feature maps with a size of 27 × 27.

Convolutional layer C3: It has six 5 × 5 convolution kernels with a step size of 1, and consists of six feature maps with a size of 23 × 23.

Pooling layer S3: S3 adopts the average pooling strategy, which is obtained after the C3 layer is down-sampled.

Output layer: This layer is a fully connected layer that pulls the features of the output layer output into a straight line to act on the neural network. Figure 6 is the connection process:

Here fx is a digital filter, bx is the offset, Cx is the feature map of the convolutional layer; wx + 1 is the weight of the down-sampling, bx + 1 is the corresponding weight, and the down-sampling layer Sx + 1 is obtained.

Here, as long as the input layer size and the local receptive field size are determined, the size of the C layer is also determined. The S layer is a down-sampling layer whose purpose is to change a plurality of pixels of the C layer into one pixel.

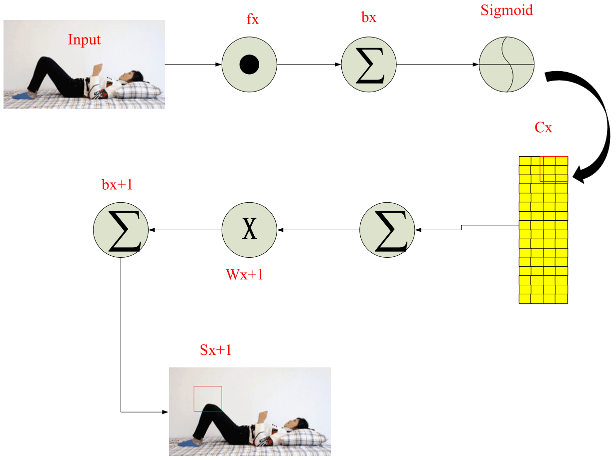

FIGURE 6. Convolutional neural network connection process.

The C layer is used as a convolutional layer for feature extraction. First, each neuron is connected to a small piece of the upper layer. Second, move the receptive field and map the new receptive field to another neuron on the C layer. Then, using a sigmoid function the process has displacement invariance. The C layer can extract features from the bottom layer to the upper layer step by step from the raw data of the pixel level. Therefore, using the C layer as the convolution layer of feature extraction can improve the accuracy of feature extraction.

The loss function is used to evaluate the degree of inconsistency between the predicted value of the model -f(x) and the true value -y. The smaller the loss function, the better the robustness of the representative model, and the loss function can guide the model learning. This paper chooses cross entropy as the loss function. Cross Entropy is used to evaluate the difference between the probability distribution and the real distribution obtained by the current training. Reducing the cross entropy loss is to improve the prediction accuracy of the model. Its discrete function form:

$$H(p,q) = -\sum_{x} p(x)\log(q(x)) \tag{13}$$

Among them, the variable p(x) is the probability of real distribution, and the variable q(x) is the probability estimate calculated by the model through the data. The cross entropy characterizes the distance between two probability distributions, or it can be said that it is difficult to express the probability distribution p(x) by the probability distribution q(x), p(x) represents the correct answer, q(x) represents the predicted value, and the smaller the cross entropy, the closer the distribution of the two probabilities is. The loss function at this time is:

$$loss = \frac{1}{2m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{14}$$

Among them, the variable y_i represents the expected output, the variable $\hat{y}_i$ represents the original actual output, and the variable m represents m samples.

B. CONVOLUTIONAL NEURAL NETWORK OF NUCLEAR EXTREME LEARNING MACHINE

The different features are nucleated and can contain features of different dimensions of the medical motion image. Therefore, this paper combines the manual feature kernel with the convolutional neural network feature kernel of meaning motion information, and uses the L2 general number to calculate the linear kernel. A linear kernel matrix can be defined as

$$K(x_i, x_j) = h(x_i)h^T(x_j) \tag{15}$$

where: K(xi, xj) is the (i, j) element of K. $h(x_i)$ and $h^T(x_j)$ indicate the type of kernel function used. The fused feature kernel is calculated by calculating the kernel matrix average of different feature sources. After the feature kernel fusion, three cores are obtained: convolutional neural network feature kernel, manual feature kernel, and fusion feature kernel. Then, the kernel extreme learning machine is used to calculate the prediction scores of different feature kernels.

It is assumed that the prediction score can be combined with the convolutional neural network to calculate the final action classification of the medical motion image sequence. Therefore, the three output score vectors in this paper are merged into one score vector s. Suppose {si,ti},i= 1, . . . ,n denote training data set, where n is the number of training samples, si∈R3q is the prediction score, q is the total number of action classifications, and ti∈R3q is the real action classification . Considering the output of the first layer as the input layer second feature, this paper uses the L1 general number to regularize the output of the first layer. After that, we obtain a feature vector for each medical motion image, and calculate it in the second layer. Because the radial basis function kernel is better than the linear kernel for the L1 norm feature, the radial basis function kernel is used, and the radial basis function kernel is defined as

$$K(s_i, s_j) = e^{-\frac{||s_i - s_j||^2}{\sigma^2}} \tag{16}$$

where: si and sj are the predicted scores of medical motion images i and j, respectively. Note that $K(s_i, s_j)$ is the (i, j) element of kernel. And the variable σ represents the variance of the predicted scores of i and j for all medically motion images.

C. IMAGE RECOGNITION FRAMEWORK BASED ON CONVOLUTIONAL NEURAL NETWORK

For the manual feature, the IDT descriptor of the medical moving image is first extracted, and the IDT descriptor includes: a trajectory line, an HOG (describe static feature), an HOF (pixel absolute motion feature), and an MBH (pixel relative motion feature). Then, the Fisher vector is used to encode the IDT feature. Because the Fisher feature needs to encode both the first-order statistic and the second-order statistic between the medical motion image descriptor and the Gaussian mixture model, the amount of data is large. Mechanism for reducing feature descriptor dimensions: Principal component analysis processing of descriptors and setting the number K of Gaussian elements to 256 [35]. A GMM (Gaussian Mixture Model, GMM) model is trained by randomly sampling a subset of 256 000 features from the training set. Finally, each descriptor type of each medical motion image is represented as a Fisher vector of 2 × D × K dimensions, where D represents the descriptor dimension after the principal component analysis process. The Fisher vector is processed using the L2 norm. Finally, the kernel matrix K_h of the manual features is calculated as the average of the descriptor kernel matrix.

$$K_h = \frac{1}{n_d} \sum K_i \tag{17}$$

where: n_d is set to 4, indicating that there are 4 different descriptors, namely trajectory line, HOG (describe static feature), HOF (pixel absolute motion feature), MBH (pixel relative motion feature). K_i represents the kernel matrix of the i-th descriptor.

For convolutional neural network features, this paper designs convolutional neural network features based on motion information and convolutional neural networks. The convolutional neural network features are organized into a 4096-dimensional medical motion image descriptor, and L2 norm processing is used for the descriptor. Then, the linear kernel of the convolutional neural network feature is calculated, and the kernel matrix Kd of the convolutional neural network feature is established. In the nuclear fusion process, the result of nuclear fusion is obtained by calculating the average value of the convolutional neural network feature Kd and the manual feature Kh: K = (Kd + Kh)/2.

Finally, the convolutional neural network of this paper uses three feature checks to input the medical motion images into action classification. The three feature kernels are: manual feature kernel, convolutional neural network feature kernel and fusion kernel. Figure 7 is a block diagram of a convolutional neural network. Firstly, the manual features are extracted by dense trajectory. Secondly, the deep learning feature of motion information is extracted by convolutional neural network. Finally, the nuclear extreme learning machine is applied to the convolutional neural network, and the extracted manual features are extracted. The deep learning features are fused and fed into the convolutional neural network for training.

Therefore, this paper combines the features extracted by manual features and convolutional neural networks, and can complement the information of the two, and finally calculate the action characteristics of the unified medical motion image. Because the extreme learning machine has strong generalization ability, applying the extreme learning machine to the convolutional neural network can make the convolutional neural network have better generalization ability and robustness.

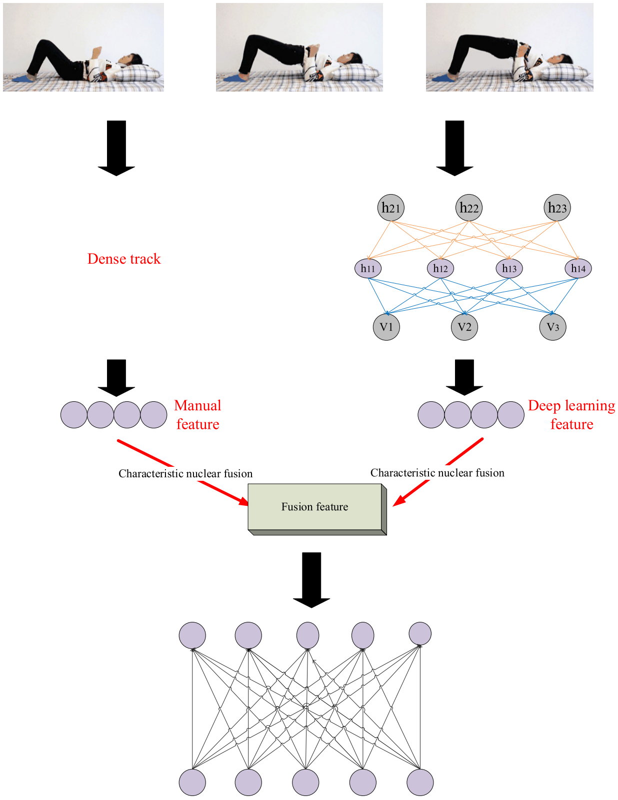

FIGURE 7. Applying the extreme learning machine to the convolutional neural network framework.

D. PARAMETER REDUCTION AND WEIGHT SHARING OF CONVOLUTIONAL NEURAL NETWORKS

In the convolutional neural network model, neurons can be divided into two categories, one is the S element used for feature extraction, and the other is the anti-deformation C element. There are two important parameters in the S element, namely the threshold parameter and Receptive field; Receptive field is how much space is extracted from the input layer as input, and the threshold parameter controls the degree of response of the output to the input. Similarly, a convolutional neural network is a multi-layer network structure, each layer of which is actually composed of multiple feature maps, each of which represents a feature; there are many independent nerves on each feature map. Correspondingly, the network layer of the convolutional neural network is divided into a convolutional layer and a down-sampling layer, also called down-sampling or subsampling; a non-linear mapping between network layers, from the convolutional layer to the down-sampling layer is a down-sampling. The process from the down-sampling layer to the convolutional layer is a convolutional filtering process.

Since the weights on the mapping surface are shared, that is to say, the weights of each neuron are the same, the whole network parameters are greatly reduced and the complexity is reduced; the network adopts a combination of feature extraction and down-sampling, and the sub-sampling is used to obtain a local average. This structure makes the network highly resistant to distortion.

The biggest advantage of convolutional neural networks is the reduction of network parameters, i.e. parameter reduction and weight sharing, through receptive fields and shared weights. This approach makes training faster and requires fewer samples for training. This method of simulating the human brain visual center naturally has better algorithmic efficiency.

The system parameters can be greatly reduced by the regional receptive field. The so-called regional receptive field means that the network is not fully connected, and each neuron is only connected to a small area of the upper layer, so that the parameters between the layers are greatly reduced. If the resolution of the training sample image is 1000×1000, the number of nodes in the convolutional layer of the network is 106, and the total number of weights for you, the full connection is 1012. However, similar to the process of people observing things, not every neuron needs to remember all the information, but only needs to record a small area, and then combine the information recorded by these different neurons at the upper level. The convolutional neural network simulates this principle, which greatly reduces the weight of the network that needs to be trained. If the size of the local receptive field is 10x10, then the number of neurons required for the hidden layer of the convolutional neural network is less than 106, and each neuron is only connected to a region of 100 pixels, so the number of total weight required is less than 10^8, reduced by 4 orders of magnitude, saving time and effort.

Of course, the general convolutional neural network extracts a variety of features, so each feature map contains 10^8 weights, and the final parameters are also approximately Nx10^8, where N is the number of feature maps.

The convolutional neural network transforms the input of the corresponding receptive field into a neuron of the convolutional layer by convolving the digital filter by weighting the pixels in the receptive field, and adding an offset to the volume. In the stack, it is equivalent to input to an excitation function, which is also called a convolution kernel. The process of convolution is that the receptive field moves one pixel at a time, thus obtaining a new receptive field, which is then mapped to another neuron in the convolutional layer by an excitation function. Move the field so much until it covers the entire input layer.

In this case, each local receptive field is mapped to a neuron in the convolutional layer through a convolution kernel. The original pixel is 100×100, the receptive field size is 10×10, and the moving step size is 1, then the number of final receptive fields is 991^2. If each receptive field corresponds to a convolution kernel, each convolution kernel is a 10×10 to 1 mapping, so each convolution kernel has 100 parameters, so the parameter is 10^8. This is not a small amount.

In order to further reduce the number of weights, the convolutional neural network adopts a network mode of weight sharing. Weight sharing means that for the same feature map, the convolution kernel of each receptive field is the same. Then for the above model, the same convolution kernel is

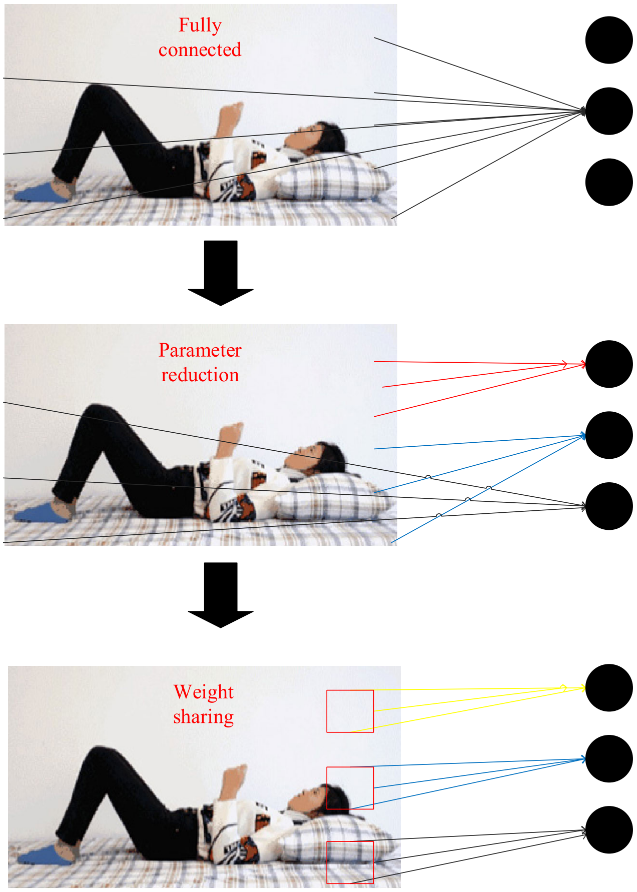

FIGURE 8. Schematic diagram of convolutional neural network parameter reduction and weight sharing.

used, only 100 parameters are needed. This is the meaning of weight sharing. In real-world applications, such feature extraction is not enough for the entire system, because only one feature is extracted, and the single feature is not enough for the system to learn and classify. In order to improve the performance of the algorithm, the convolutional neural network uses multiple filters. By using this filter to convolve the input image, a feature map can be obtained. Each filter corresponds to a different convolution kernel, because the convolution kernel is different. The extracted features are different. If the system designs 100 filters, then the extracted 100 features are enough for the system to complete learning and classification. The weight of the system becomes 10,000, which is reduced by 4.

The convolutional network combines the three local structures of local receptive field, weight sharing and down-sampling to make the system have a certain degree of displacement, scale and deformation invariance. Figure 8 depicts the parameter reduction and weight sharing of a convolutional neural network.

Using a random retreating neural network structure to fine-tune network parameters can effectively alleviate over-fitting and improve recognition rate. The parameters that are fine-tuned by the random repellent neural network are more stable than the ordinary back-propagation algorithm, and more useful features can be extracted.

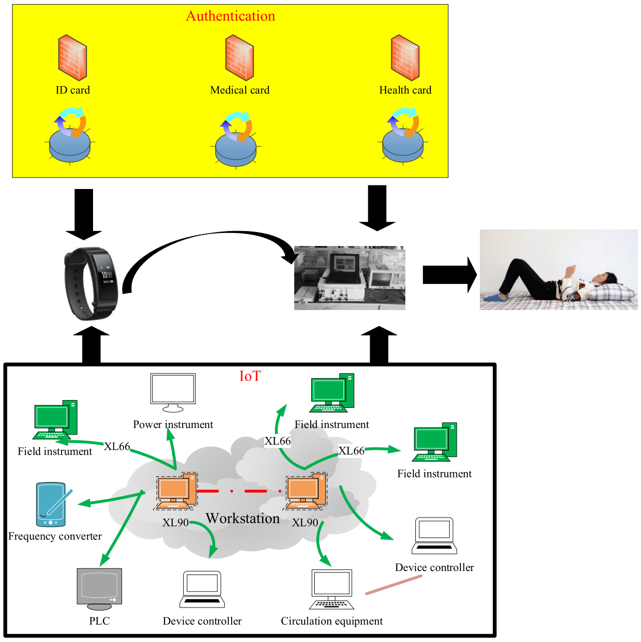

FIGURE 9. The designing diagram of an acquisition device based on the Internet of things.

IV. EXPERIMENTS AND RESULTS

A. DATABASE DESCRIPTION

The rise of the Internet of Things, especially with the manufacturing industry, has fostered the emerging manufacturing model of manufacturing IoT. In the manufacturing environment, the deployed sensors generate a large amount of real-time data streams containing information about the patient's physical state, which needs to be analyzed and extracted from the hidden information and knowledge real-time monitoring and precise control of patients. This paper uses the Internet of Things technology to design a collection device based on Internet of Things technology and medical moving images. Figure 9 is a design diagram of an acquisition device based on the Internet of Things.

B. IMAGE PREPROCESSING

Image down-sampling removes a portion of the pixels from the neighborhood of the image, leaving another portion of the pixels. The amount of down-sampled picture information is reduced, and the picture may be partially distorted, but there is no loss of information for the entire global image.

The three most common methods of image down-sampling are bilinear interpolation, nearest neighbor interpolation, and bicubic interpolation. All three methods have their own advantages and disadvantages. For this paper, down-sampling is only a means. The purpose of down-sampling here is not to preserve the original picture information, but to eliminate the redundant information of the input picture to the greatest extent. Alleviate overfitting while reducing training time.

FIGURE 10. Comparison of down-sampled medical motion images.

Therefore, the nearest neighbor interpolation algorithm is used here. The biggest advantage of the most adjacent interpolation algorithm is that the operation speed is fast, but the picture loss is large. The down-sampling process is to select the gray value of the closest point among the several pixels of the neighborhood of the point to be sampled as a result, and the method is mainly for the two-dimensional image. The nearest neighbor interpolation algorithm has a small amount of calculation and the algorithm is simple, so the operation speed is fast. Figure 10 is a comparison of down-sampled medical motion images. After down-sampling, the image size is scaled proportionally.

C. SELECTION OF EXPERIMENTAL PARAMETERS

In deep neural networks, adjusting the hyper-parameter combination is not an easy task because training deep neural networks is time consuming and requires multiple parameters to be configured. Among them, the famous Lenet-5 network, AlexNet and FCN networks all achieve the best network performance by adjusting the combination of hyper-parameters. This paper, like the Lenet-5 network, AlexNet and FCN networks, seeks optimal parameters for network performance by setting different combinations of parameters.

In order to test the influence of the initialization parameters on the training results, the standard deviation of the initialization parameters is set to 0.1, 0.2, and 0.5 respectively. The experimental results after 8 iterations are shown in Table 1. It can be seen from Table 1 that the smaller the standard deviation of the initialization parameters, the better the training results. When the standard deviation of the initialization parameters is too large, there may be cases of non-convergence. When the standard deviation of the initialization parameters is 0.1, the difference of the loss function is 0.368, which is twice the standard deviation of 0.2, and the training accuracy is also 0.909, which is 0. 2 hours of accuracy is high. When the standard deviation of the random number is 0.5, after the training, the loss function not only does not decrease, but increases, and the model does not converge. When the standard deviation of the initialization parameters is relatively small, not only the convergence is faster, but also the training effect is better.

In this experiment, under the same conditions, the size of the pooled core and the step size of the movement were modified, and 80 iterations were performed for each training. The parameters and results of the experiment are shown in Table 2. Changing the size of the pooled core and the step size of the movement has different effects on the implementation results.

TABLE 1. Effects of different initialized parameters standard deviation on training effect.

Standard deviation of random numbers	0.1	0.2	0.3
Initial loss function value	0.591	0.412	0.003
Final loss function value	0.205	0.222	0.006
Loss function difference	0.386	0.193	-0.003
Training accuracy	0.909	0.853	0.803
Test accuracy	0.903	0.812	0.763

TABLE 2. Effects of pooling size and moving step size on experiment.

Pool size	10x1	20x1	20x1
Moving step	2	1	2
Loss function	0.002	0.001	0.001
Training accuracy	0.963	0.976	0.955
Test accuracy	0.920	0.930	0.924
Poor training and test accuracy	0.043	0.046	0.031

The longer the pooling length is, the larger the moving step is, the less features are retained, and the shorter the training time is. The shorter the pooling length is, the smaller the moving step is, and the more features are retained, the longer the training time is. But the more features that are not retained, the better the classification will be. According to the data in Table 2, it can be seen that when the pooling length is 20 and the moving step is 1, the effect is best. Therefore, the author will adopt a pooled core with a size of 20 and a moving step size of 1.

Observing the last row in Table 2 is the difference between the training accuracy and the test accuracy, that is, the training accuracy minus the test accuracy. It can be found that there is a difference between the test accuracy and the training accuracy when the value of the loss function is close to 0.01. The larger difference, that is, the test accuracy is less than the training accuracy, it is likely that over-fitting has occurred. After the author joins Dropout, the model is retrained and evaluated. The specific results of the experiment are shown in Table 3. In Table 3, in addition to the difference between training accuracy, test accuracy, training accuracy and test accuracy, the percentage difference is also added.

It can be seen from Table 3 that when the Dropout process is added, when the number of trainings is 100, the value of the loss function is still large, because only half of the neurons participate in the training, so after adding the Dropout, Increase the number of training. Comparing the last line without the addition of Dropout and other experiments with Dropout, the loss function is reduced to around 0.001 under the same conditions, and the training accuracy is basically the same. The test accuracy after the Dropout process is

TABLE 3. Effect of Dropout on training result.

Whether to join Dropout	Yes	Yes	Yes	No
Number of training	100	200	500	80
Loss function	0.439	0.003	0.002	0.001
Training accuracy	0.811	0.816	0.958	0.955
Test accuracy	0.793	0.807	0.938	0.923
Poor training and test accuracy	0.018	0.009	0.020	0.032
Percentage difference	2.181	1.032	2.093	3.304

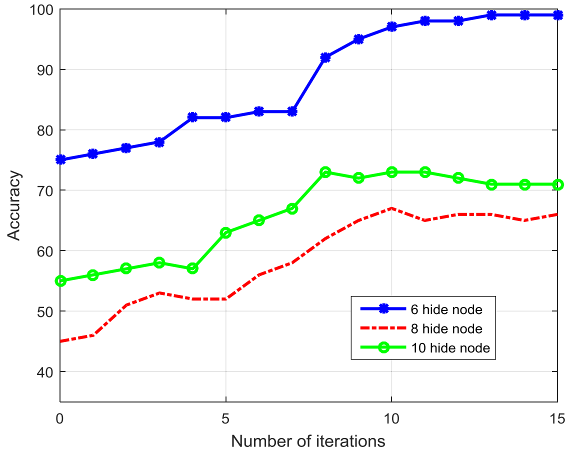

FIGURE 11. Different number of hidden layers of model test accuracy.

significantly higher than that without the Dropout. In order to increase the speed, the experiment used a pooling core of 20 × 1 and a moving step size of 2, so the training precision reached 0.958 at the end.

In this experiment, a 10-fold cross-validation of all training data is used to find the convolutional neural network structure of the optimal number of hidden layer nodes. First, a convolutional neural network with different numbers of hidden layers is established in this paper, and then repeated experiments are performed for each network structure. In the experiment, it is found that after the number of experiments exceeds 15 times, the average value of the accuracy will gradually stabilize, and there will be no large fluctuations, which means that the classification effect is better, as shown in Figure 11. As can be seen from Figure 11, the convolutional neural network with six hidden layers achieved the highest classification accuracy 13 times in all 15 experiments. In addition, the network average classification accuracy of the 8 hidden layers is 66.47%, and the average accuracy of the 10 hidden layer networks is 53.53%. Therefore, this paper selects six hidden layer convolutional neural network structures to establish an intelligent recognition algorithm for medical moving images.

D. PERFORMANCE COMPARISON OF TWO FUSION STRATEGIES

The current feature fusion strategy is mainly divided into two strategies, including early fusion and late fusion. The early fusion strategy performs the fusion of feature kernels before the classifier process. The latter fusion strategy first fuses the score vectors of each feature into a score vector, and then scores. The vector is again processed by the classifier, and Figure 12 and Figure 13 are the flow of the two fusion strategies. Evaluate the impact of different fusion strategies of the two features on the performance of the motion recognition algorithm.

The algorithm is used to carry out experiments on two fusion strategies. The results are shown in Figure 14. It can be seen from Figure 14 that the recognition accuracy of early fusion strategies is better than that of late fusion strategies.

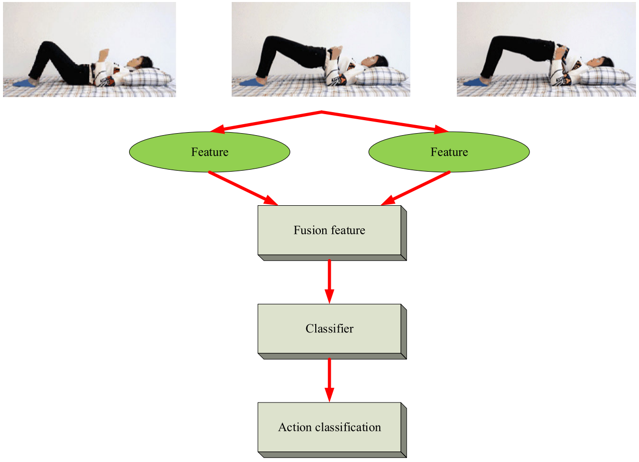

FIGURE 12. Early integration strategy.

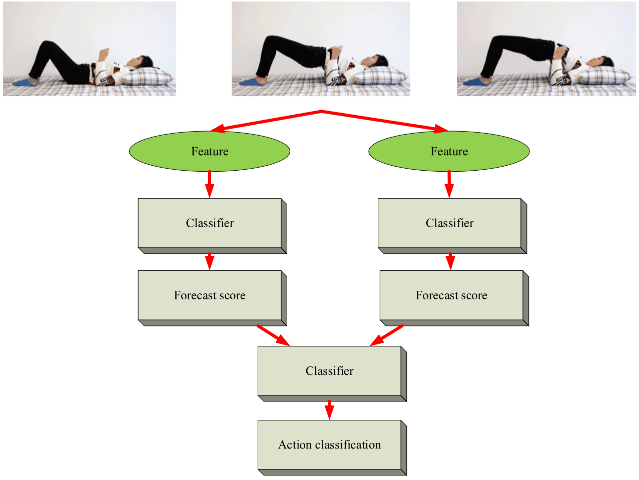

FIGURE 13. Late fusion strategy.

E. PERFORMANCE COMPARISON OF NETWORK MODELS

The performance of the intelligent recognition algorithm based on deep neural network for medical motion images proposed in this paper is shown in Figure 15. The abscissa indicates the number of iterations of the convolutional neural network model training, and the ordinate indicates the model in the training set. Corresponding accuracy performance on the test set.

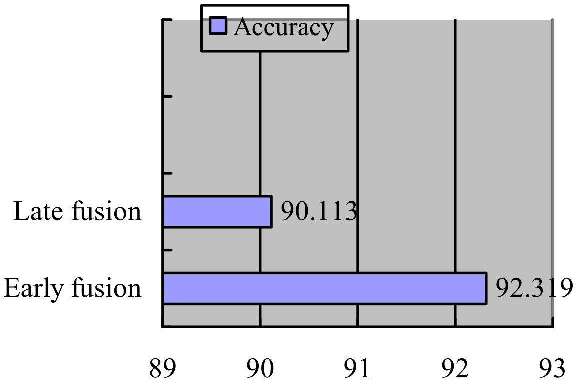

FIGURE 14. **The recognition accuracy of the algorithm in different fusion strategies.**

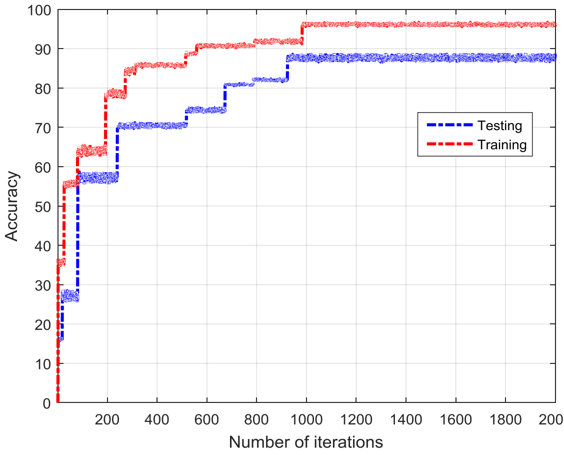

FIGURE 15. **Deep neural network training and testing performance.**

It can be seen from Figure 15 that the performance of the algorithm on the training set is significantly better than the test set, because the algorithm is prone to over-fitting on the training set, and the test set can prevent the system from over-fitting and improve the generalization ability of the system. The effect of the algorithm after 1000 iterations, the performance of the algorithm on the training set and the test set tends to be stable, indicating that the algorithm gradually converges.

In order to evaluate the recognition performance of the algorithm, the algorithm is compared with other medical moving image recognition algorithms. The motion recognition rate results and ROC curves of different algorithms are shown in Figure 16 and Figure 17. It can be seen that the recognition rates of CNN-T, SVM and this algorithm are better than those of CNN, IDT and C3D, CNN-T, Both SVM and this algorithm belong to the multi-feature fusion recognition algorithm, while CNN, IDT and C3D are single feature recognition algorithms. It can be concluded that the recognition performance of multi-feature fusion is better than the recognition performance of single feature. In addition, the recognition rate of this algorithm is slightly better than the two algorithms of CNN-T and SVM. This algorithm is similar to the SVM algorithm. The main difference is that this algorithm designs a motion information mechanism based on event template, which can effectively improve the precision of motion recognition.

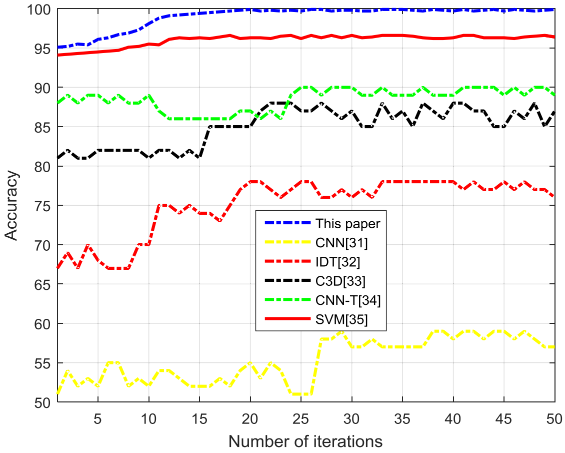

FIGURE 16. **Motion recognition rate results of different medical motion image recognition algorithms.**

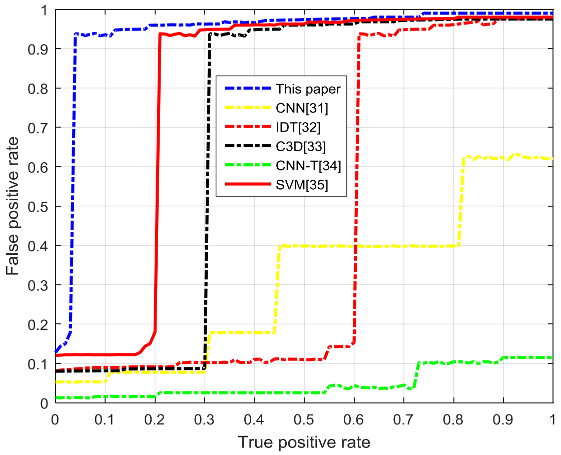

FIGURE 17. **The ROC curve of different medical motion image recognition algorithms.**

In order to evaluate the recognition performance of the algorithm for small-scale motion, before extracting the feature, 64 × 48 frames are selected as the time template of the learning feature of the convolutional neural network, and Table 4 is the recognition accuracy of the four membership functions for the different dimensional features. As can be seen from the table, in general, the accuracy of the RGB color feature is better than the time feature, because for a small amount of human motion, the action and the human body have more overlapping phenomena. In this case, the color feature is discriminated more sexual.

According to the results of Table 4, the effect of the $\mu1$ function is poor, and the three membership functions of $\mu2$, $\mu3$, and $\mu4$ emphasize the saliency of the early, middle, and late time domains, respectively. The three membership functions $\mu2$, $\mu3$ and $\mu4$ are superimposed and merged,

TABLE 4. Accuracy of recognition of different dimensional features by four membership functions.

Membership function	μ1	μ2	μ3	μ4
Time dimension	35.213	44.333	44.213	44.331
Color dimension	60.881	64.083	61.931	66.831
Fusion feature	59.278	61.582	62.934	68.856

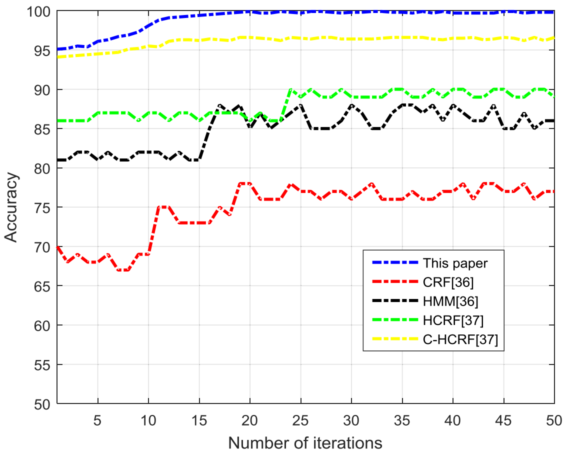

FIGURE 18. Recognition accuracy result of small amplitude motion recognition algorithm.

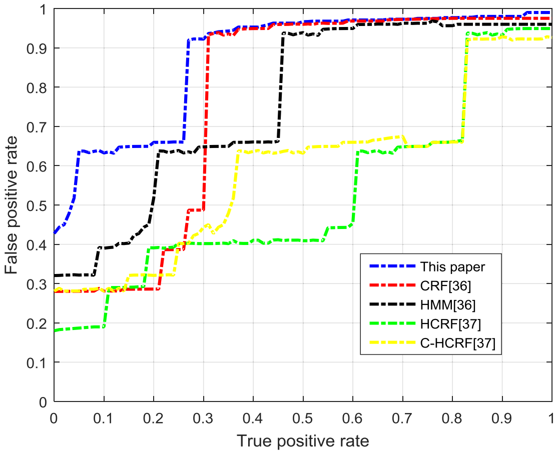

FIGURE 19. The ROC curve of small amplitude motion recognition algorithm.

and the recognition accuracy of the proposed algorithm is retested using the fused membership function. The algorithm is compared with the four motion recognition algorithms supporting small amplitude gesture recognition mentioned in literature [36] and literature [37], and the recognition result and ROC curve are shown in Figure 18 and Figure 19. It can be seen from the results that the algorithm combines motion information and RGB three-color features, and shows a high recognition accuracy for small-scale motion, which is significantly better than the first three algorithms. The C-HCRF algorithm extracts medical motion images. The multi-view feature of the sequence can effectively analyze the multi-level information of the medical moving image, and also achieves a very high recognition accuracy, which is close to the algorithm.

V. CONCLUSION

For the intelligent recognition of small-scale motion in medical motion images, the existing convolutional neural network model cannot analyze the relationship between the parts of the image and the nuances in the image, so it cannot describe the relationship between the small-scale motion and the whole image. The existing convolutional neural network model has low robustness and poor processing effect on high-dimensional image data, which limits the versatility and adaptability of the model and is difficult to implement in clinical applications.

In view of the above problems, this paper proposes an intelligent recognition algorithm based on convolutional neural network for medical motion images. Firstly, the manual features are extracted. Secondly, the convolutional neural network is used to extract the motion information of the medical moving images. Finally, the nuclear extreme learning machine is applied to the convolutional neural network, and the previously extracted manual features and convolutions are learned. Features are fused to fully characterize the motion characteristics of medical motion image sequences. The algorithm can not only assign different saliency to different time domains of motion information, but also input the time template of medical motion image sequence into the convolutional neural network to learn the feature set of human motion recognition. Moreover, the manual features and convolutional neural network features of the algorithm are complementary, and the human motion information of medical moving images is described from different angles. The experimental results of the simulation experiment verify the effectiveness of the algorithm.

REFERENCES

[1] M. Gao, J. Jiang, G. Zou, V. John, and Z. Liu, "RGB-D-based object recognition using multimodal convolutional neural networks: A survey," *IEEE Access*, vol. 7, pp. 43110–43136, 2019.

[2] J. H. Tan, Y. Hagiwara, W. Pang, I. Lim, S. L. Oh, M. Adam, R. S. Tan, M. Chen, and U. R. Acharya, "Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals," *Comput. Biol. Med.*, vol. 94, pp. 19–26, Mar. 2018.

[3] G. Chen, D. Xiang, and B. Zhang, "Automatic pathological lung segmentation in low-dose CT image using eigenspace sparse shape composition," *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1736-1749, Jul. 2019.

[4] L. Resnik, F. Acluche, M. Borgia, G. Latlief, and S. Phillips, "EMG pattern recognition control of the DEKA arm: Impact on user ratings of satisfaction and usability," *IEEE J. Eng. Health Med.*, vol. 7, 2019, Art. no. 2100113.

[5] X. Y. Zhou, G. Z. Yang, and S. L. Lee, "A real-time and registration-free framework for dynamic shape instantiation," *Med. Image Anal.*, vol. 44, pp. 86–97, Dec. 2017.

[6] R. D. Farid, J. Azizi, M. S. Allahyari, C. A. Damalas, and H. Sadeghpour, "Marketing mix for the promotion of biological control among small-scale paddy farmers," *Int. J. Pest Manage.*, vol. 65, no. 1, pp. 59–65, 2018.

[7] R. Funatsu, T. Kajiyama, T. Yasue, K. Kikuchi, K. Tomioka, T. Nakamura, H. Okamoto, E. Miyashita, and H. Shimamoto, "8K 240-Hz full-resolution high-speed camera and slow-motion replay server systems," *SMPTE Motion Imag. J.*, vol. 128, no. 3, pp. 44–49, Apr. 2019.

[8] I. A. Matveev and V. P. Novik, ''Method of optimal circular path for iris template matching,'' *Pattern Recognit. Image Anal.*, vol. 29, no. 1, pp. 42–50, Jan. 2019.

[9] V. E. Antsiperov, ''Object identification on low-count images by means of maximum-likelihood descriptors of precedents,'' *Pattern Recognit. Image Anal.*, vol. 29, no. 1, pp. 21–34, Jan. 2019.

[10] R. Słowiński, *Fuzzy Sets in Decision Analysis, Operations Research and Statistics*, vol. 123, no. 3. New York, NY, USA: Social Science Electronic, 2018, pp. 407–409.

[11] C. Peng, S. L. Lo, J. Huang, and A. C. Tsoi, ''Human action segmentation based on a streaming uniform entropy slice method,'' *IEEE Access*, vol. 6, pp. 16958–16971, 2018.

[12] L. Zhou, C. Zhang, Z. Wang, Y. Wang, and Z. Lu, ''Hierarchical palmprint feature extraction and recognition based on multi-wavelets and complex network,'' *IET Image Process.*, vol. 12, no. 6, pp. 985–992, Jun. 2018.

[13] G. Yang, Y. Bao, and Z. Liu, ''Localization and recognition of pests in tea plantation based on image saliency analysis and convolutional neural network,'' *Trans. Chin. Soc. Agricult. Eng.*, vol. 33, no. 6, pp. 156–162, 2017.

[14] M. Long, C. J. Ouyang, H. Liu, and Q. Fu, ''Image recognition of camellia oleifera diseases based on convolutional neural network & transfer learning,'' *Trans. Chin. Soc. Agricult. Eng.*, vol. 34, no. 18, pp. 194–201, 2018.

[15] K. L. Campbell and L. K. Tyler, ''Language-related domain-specific and domain-general systems in the human brain,'' *Current Opinion Behav. Sci.*, vol. 21, pp. 132–137, Jun. 2018.

[16] C. Du, C. Du, L. Huang, and H. He, ''Reconstructing perceived images from human brain activities with Bayesian deep multiview learning,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 8, pp. 2310–2323, Aug. 2019.

[17] E. R. Fantoni, A. Chalkidou, J. T. O'Brien, G. Farrar, and A. Hammers, ''A systematic review and aggregated analysis on the impact of amyloid pet brain imaging on the diagnosis, diagnostic confidence, and management of patients being evaluated for alzheimer's disease,'' *J. Alzheimers Disease*, vol. 63, no. 2, pp. 783–796, 2018.

[18] M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H. C. Shin, H. Roth, G. Z. Papadakis, A. Depeursinge, R. M. Summers, Z. Xu, and D. J. Mollura, ''Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks,'' *Comput. Methods Biomech. Biomed. Eng. Imag. Vis.*, vol. 6, no. 1, pp. 1–6, 2015.

[19] H. Madokoro, A. Yamanashi, and K. Sato, ''Unsupervised semantic indoor scene classification for robot vision based on context of features using Gist and HSV-SIFT,'' *Pattern Recognit. Phys.*, vol. 1, no. 1, pp. 93–103, Aug. 2013.

[20] L. S. Richman, C. Garcia, N. Bouchard, P. R. Muskin, and A. L. Dzierba, ''Evaluation of a symptom-triggered protocol for alcohol withdrawal for use in the emergency department, general medical wards, and intensive care unit,'' *J. Psychiatric Pract.*, vol. 25, no. 1, pp. 63–70, Jan. 2019.

[21] P. C. Sumathi and N. Priya, ''Analysis of an Automatic Text Content Extraction Approach in Noisy Video Images,'' *Int. J. Comput. Appl.*, vol. 69, no. 4, pp. 6–13, 2013.

[22] G. Liang, H. Hong, and W. Xie, ''Combining convolutional neural network with recursive neural network for blood cell image classification,'' *IEEE Access*, vol. 6, p. 36188- 36197, 2018.

[23] A. Hamedpour and F. F. Ahmadi, ''Recognition and tracking of moving objects in the images captured by UAV intelligently in Earth observation operations,'' *Arabian J. Geosci.*, vol. 11, no. 23, p. 738, Dec. 2018.

[24] A. G. Patwardhan, R. M. Havey, N. D. Wharton, P. P. Tsitsopoulos, P. Newman, G. Carandang, and L. I. Voronov, ''Asymmetric motion distribution between components of a mobile-core lumbar disc prosthesis: An explanation of unequal wear distribution in explanted CHARITÉ polyethylene cores,'' *J. Bone Joint Surg.*, vol. 94, no. 9, pp. 846–854, May 2015.

[25] R. O. Panicker, K. S. Kalmady, and J. Rajan, ''Automatic detection of tuberculosis bacilli from microscopic sputum smear images using deep learning methods,'' *Biocybern. Biomed. Eng.*, vol. 38, no. 3, pp. 691–699, 2018.

[26] L. Sun, J. Xu, and Y. Yin, ''Principal component-based feature selection for tumor classification,'' *Bio-Med. Mater. Eng.*, vol. 26, no. 1, pp. S2011–S2017, 2015.

[27] R. Bao, Z. Xue, X. Zhang, H. Su, and P. Du, ''Classification merged with clustering and context for hyperspectral imagery,'' *Geomatics Inf. Sci. Wuhan Univ.*, vol. 42, no. 7, pp. 890–896, 2017.

[28] H. Jiang, J. P. Wachs, and B. S. Duerstock, ''Integrated vision-based system for efficient, semi-automated control of a robotic manipulator,'' *Int. J. Intell. Comput. Cybern.*, vol. 7, no. 3, pp. 253–266, Aug. 2014.

[29] L. Iliadis and C. Jane, ''Special issue: Engineering applications of neural networks,'' *Neural Comput. Appl.*, vol. 28, no. 6, pp. 1–3, 2017.

[30] R. Paolucci, F. Gatti, and M. Infantino, ''Broadband ground motions from 3D physics-based numerical simulations using artificial neural networks,'' *Bull. Seismolog. Soc. Amer.*, vol. 108, no. 3A, pp. 1272–1286, Feb. 2018.

[31] N.-H. Lin, T.-L. Lin, X. Wang, W. T. Kao, H. W. Tseng, S. L. Chen, Y. S. Chiou, J. F. Villaverde, and Y. F. Kuo, ''Teeth detection algorithm and teeth condition classification based on convolutional neural networks for dental panoramic radiographs,'' *J. Med. Imag. Health Inform.*, vol. 8, no. 3, pp. 507–515, 2018.

[32] P. G. D. Feulner, J. Schwarzer, and M. P. Haesler, J. I. Meier, and O. Seehausen, ''A dense linkage map of lake victoria cichlids improved the pundamilia genome assembly and revealed a major QTL for sex-determination,'' *G3*, vol. 8, no. 7, pp. 2411–2420, Jul. 2018.

[33] M. Chaa, Z. Akhtar, and A. Attia, ''3D palmprint recognition using unsupervised convolutional deep learning network and SVM classifier,'' *IET Image Process.*, vol. 13, no. 5, pp. 736–745, Apr. 2019.

[34] Z.-F. Xie, Y.-C. Guo, S.-H. Zhang, W.-J. Zhang, and L.-Z. Ma, ''Multi-exposure motion estimation based on deep convolutional networks,'' *J. Comput. Sci. Technol.*, vol. 33, no. 3, pp. 487–501, May 2018.

[35] J. Jiang, Z. Wen, and M. Zhao, ''Series arc detection and complex load recognition based on principal component analysis and support vector machine,'' *IEEE Access*, vol. 7, pp. 47221–47229, 2019.

[36] A. Mezari and I. Maglogiannis, ''An easily customized gesture recognizer for assisted living using commodity mobile devices,'' *J. Healthcare Eng.*, vol. 2018, Jul. 2018, Art. no. 3180652.

[37] J. Han, Z. Zhang, G. Keren, and B. Schuller, ''Emotion recognition in speech with latent discriminative representations learning,'' *Acta Acustica United Acustica*, vol. 104, no. 5, pp. 737–740, 2018.

YUCHENG ZHOU was born in Liaoning, China, in 1979. He received the bachelor's degree from Shenyang Sport University, in 2004, and the master's degree from Beijing Sport University, in 2011. From 2004 to 2008, he was with the University of Science and Technology, Liaoning. He is currently with Chongqing Jiaotong University. He has published five articles, one of which has been indexed by SCI. His research interests include sports medicine and the IoT.

ZHIXIAN GAO was born in Henan, China, in 1981. She received the bachelor's degree from Henan University, in 2004, the master's degree from Henan Normal University, in 2011, and the Ph.D. degree from Bioengineering College, Chongqing University, in 2019. She is currently a Lecturer with the Henan Institute of Technology. She has published a total of four articles. Her research interests include sports medicine and the IoT.