IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 21, 2019, accepted October 27, 2019, date of publication November 1, 2019, date of current version November 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2950960

Segmentation Algorithm of Medical Exercise Rehabilitation Image Based on HFCNN and IoT

LIANG DING[1,2] AND HUIJIE REN[3]

[1]Department of Physical Education, Luliang University, Lüliang 033001, China
[2]Institute of Fitness and Rehabilitation, Luliang University, Lüliang 033001, China
[3]School of Sports Medicine, Dankook University, Cheonan-si 31116, South Korea

Corresponding author: Huijie Ren (renhuijie1990@126.com)

ABSTRACT Deep learning has achieved great success in the field of computer vision, and the precision in image classification and image detection has surpassed humans. Therefore, this paper combines deep learning and medical image segmentation, focusing on how to improve the accuracy and speed of segmentation algorithm of medical exercise rehabilitation image. Aiming at the shortcomings of traditional medical image recognition methods, a medical exercise rehabilitation image segmentation algorithm based on hierarchical features of convolutional neural networks is proposed, this paper calls it as hierarchical features of convolutional neural networks (HFCNN). The algorithm firstly samples the convolution output of multiple layers in the convolutional neural network to a unified scale and combines them to construct a hierarchical feature. This hierarchical feature combines the structural information of objects contained in the shallow layer of the network with the semantic information of objects contained in the deep layers of the network, so it has a strong ability to express. Secondly, the image can be segmented into multiple super pixels by the super pixel segmentation algorithm. The classifier is trained using the hierarchical features of the super pixel, and then the classification result is mapped back to the pixel. Finally, a fully connected conditional random field algorithm including one-potential potential energy and paired potential energy is constructed. The corresponding energy function is used to smooth the classification result of the pixel, and the regional consistency and continuity of the pixel mark are improved. Compared with many classical convolutional neural network algorithms, this algorithm not only accelerates the network convergence speed, shortens the training time, but also significantly improves the accuracy of segmentation algorithm of medical exercise rehabilitation image, showing good practical value.

INDEX TERMS Segmentation algorithm, IoT, medical exercise rehabilitation Image, deep learning.

I. INTRODUCTION

In recent years, with the vigorous development and application of image generation technologies such as magnetic resonance imaging, positron emission tomography, imaging spectroscopy, ultrasound imaging, X-ray, DSA imaging, etc., in more and more medical diagnosis Successful use of medical exercise rehabilitation image generation technology to improve the efficiency of doctors to diagnose and reduce the rate of misdiagnosis of doctors. Image generation technology can enable doctors to have a very intuitive understanding of the current state of the patient and the development trend of the disease, which will help the doctor to follow up the analysis and judgment of the case, and ultimately significantly improve the probability of the disease being cured. For medical diagnosis, disease treatment has great significance. With the wide application of the generation technology of medical exercise rehabilitation image, medical institutions need to process more and medical exercise rehabilitation images. It is not only time-consuming and labor-intensive, but because of the difference in doctor's level, because doctors manually process and analyze medical exercise rehabilitation images. The effects of physical fatigue can make the processed results unstable or even wrong. This makes medical institutions urgently need automated medical exercise rehabilitation image processing technology to help alleviate

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

the burden on doctors and to stably and effectively process medical moving images. With the accumulation of medical exercise rehabilitation images, the researchers have carried out detailed processing and analysis of massive medical exercise rehabilitation images, and learned the correlation between some diseases and their medical exercise rehabilitation images through algorithms, and achieved certain results in medical exercise rehabilitation image processing [1], [2]. A vital task in medical exercise rehabilitation image processing is medical image segmentation, and many other related tasks in medical exercise rehabilitation image processing require prior image segmentation. Therefore, segmentation algorithm of medical exercise rehabilitation image is a very important part of medical image processing.

Before the popularity of deep learning, many researchers were looking for features that could better express the essential properties of images. The better features proposed at the time were Gist [3], HoG [4], SIFT [5], SURF [6], etc. Although these features yield good results on some visual tasks, because each feature is constructed from some aspect of the image, it cannot accommodate all visual tasks. After deep learning, especially the emergence of convolutional neural networks, this method of auto-learning features using convolutional neural networks has driven the rapid development of many visual tasks, including segmentation algorithm of medical exercise rehabilitation image.

Since a single image feature is relatively one-sided and does not make full use of image information, some researchers have begun to extract more features under different conditions for more comprehensive representation of image information. However, the classifier usually needs to be input in the form of a single eigenvector, which requires merging multiple single eigenvectors to obtain a vector that effectively expresses the image features. On the other hand, based on the method of artificially designing image features, fewer features are extracted, and the classifier parameters are less. Although other information of medical exercise rehabilitation images, such as age, gender, and smoking history, can be combined with the experience of experts in related fields. Quickly deal with the problem to be solved, but the characteristics of the manual definition often vary from person to person, only work on some specific data or fields, and the generalization ability is poor. Moreover, in the current medical exercise rehabilitation image segmentation method based on deep learning, there are mainly two problems. The first problem is that the signal in the convolutional neural network is continuously attenuated from the shallow to deep propagation of the network, and the details of the objects contained in the convolution output are less and less. This problem is mainly caused by the presence of pooling and other down sampling in the CNN caused by the operation. The second problem is that the output of the general convolutional neural network is rather rough. On the one hand, the single classifier using softmax has limited classification ability, and does not consider the relationship between color and distance between pixels.

Aiming at the above problems, this paper proposes a segmentation algorithm of medical exercise rehabilitation image based on the hierarchical features of convolutional neural networks (HFCNN). The algorithm firstly samples the convolution output of multiple layers in the convolutional neural network to a unified scale and combines them to construct a hierarchical feature. Secondly, the image can be segmented into multiple super pixels by the super pixel segmentation algorithm. The classifier is trained using the hierarchical features of the super pixel, and then the classification result is mapped back to the pixel. Finally, a fully connected conditional random field algorithm including one-potential potential energy and paired potential energy is constructed. The corresponding energy function is used to smooth the classification result of the pixel, and the regional consistency and continuity of the pixel mark are improved. Compared with many classical convolutional neural network algorithms, this algorithm not only accelerates the network convergence speed, shortens the training time, but also significantly improves the accuracy of segmentation algorithm of medical exercise rehabilitation image, showing good practical value.

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

First: this paper analyzes the current research status of segmentation algorithms of medical exercise rehabilitation image, summarizes the commonly used algorithms for segmentation algorithm of medical exercise rehabilitation image, and studies the related algorithms of deep learning in segmentation algorithms of medical exercise rehabilitation image, which is a convolutional neural network and provides a reference value.

Second: This paper proposes a segmentation algorithm of medical exercise rehabilitation image based on the hierarchical features of convolutional neural networks. The algorithm not only accelerates the network convergence speed, shortens the training time, but also significantly improves the accuracy of segmentation algorithm of medical exercise rehabilitation image, showing good practical value.

The rest of this paper was organized as follows. Related work was introduced in Section. II. The segmentation algorithms of medical exercise rehabilitation image and deep learning algorithm were introduced in Section. III. Section. IV describes medical exercise rehabilitation image segmentation algorithm based on hierarchical features of convolutional neural networks. Experimental results and analysis were discussed in Detail in Section. V. Finally, Section. VI concluded the whole paper.

II. RELATED WORK

In solving the problem of segmentation algorithms of medical exercise rehabilitation image, researchers also use these artificially constructed features, and at the same time complete image semantic segmentation by means of Markov Random Fields (MRF), conditional random field or other probability map algorithms. In 2009, Shotton *et al.* [7] proposed a conditional random field algorithm that can integrate the background, texture, color, position and other information of objects in medical exercise rehabilitation images, because this algorithm combines the characteristics of multiple aspects of objects. Therefore, it has achieved good results in segmentation algorithms of medical exercise rehabilitation image. Noormohamadi *et al.* [8] proposed a layered random field algorithm, which can integrate image features on different layers, so that better segmentation and recognition results can be obtained. The ordinary conditional random field algorithm expresses the relationship between objects through the edge connection between adjacent pixels. Song *et al.* [9] proposed a fully connected condition random field; full connection condition exists in each pair of pixels in the random field algorithm. Edge-to-edge, this approach enhances the ability of the algorithm to capture relationships between images, thereby significantly improving the results of segmentation algorithms of medical exercise rehabilitation image. Although the probability map algorithm has achieved certain effects on segmentation algorithms of medical exercise rehabilitation image, the results of this method are still limited by the quality of features extracted from the image.

After deep learning, especially the emergence of convolutional neural networks, this method of auto-learning features using convolutional neural networks has driven the rapid development of many visual tasks, including segmentation algorithms of medical exercise rehabilitation image. Hoseini *et al.* [10] applied the study of deep learning algorithms to the segmentation of neural cell images. Xiao *et al.* [11] applied the deep learning algorithm to the benign and malignant discrimination of breast tumors. Lin *et al.* [12] used the deep learning algorithm to classify medical exercise rehabilitation image, and used the deep learning algorithm to classify medical exercise rehabilitation images. Girish *et al.* [13] used deep convolutional neural networks for segmentation of retinal vessels and performed well in retinal vessel segmentation. Liu *et al.* [14] researchers applied three-dimensional convolutional neural networks to the study of brain lesion segmentation, which has made great progress in clinical applications. Farabet *et al.* [15] first used a multi-scale convolutional neural network to extract dense features from original medical exercise rehabilitation images. This method yields features that include object texture, shape, and background information, so that objects can be well represented. The post-processing of the classification results of the convolutional neural network using the classical conditional random field further improves the pixel marking accuracy. Long *et al.* [16] used the Fully Convolutional Neural Network (FCN) to learn an algorithm from the original medical exercise rehabilitation image to the pixel marker map in an end-to-end manner. This full convolutional neural network implements pixel marking for the first time. End-to-end training learning, the final output of the network is the category of pixels. Because the output of the full convolutional neural network is the result of direct classification of the network, the spatial relationship between the object and the object and the pixel and the pixel is not taken into consideration, so the pixel marking result is rough. Chen *et al.* [17] used the fully connected conditional smoothing constraint on the output of the full convolutional neural network to improve the regional consistency and continuity of the pixel markers, thus improving the segmentation accuracy of medical moving images.

Because convolutional neural networks and probability map algorithms have their own advantages in solving the segmentation algorithm of medical exercise rehabilitation image, the combination of these two algorithms achieves a good segmentation effect. Zheng *et al.* [18] proposed a new network algorithm that combines the advantages of CNN and fully connected CRF. He organized the CRF into a circular neural network called CRF-RNN, and then added CRF-RNN to CNN to get a deep convolutional neural network. Through end-to-end training, this network can simultaneously have CNN's feature learning ability and CRF's smooth constraint ability, thus achieving the best results at that time. Maninis *et al.* [19] proposed a region-based network that first generates some rectangular candidate regions, then extracts the features of the regions, and finally uses SVM to classify the regions. Because the network extracts the local area features, it can express the objects in the area well, so finally the better regional classification results can be obtained. Some region-based methods [14], [20] use over-segmentation to produce regions of any shape that are small, non-overlapping, and then use CNN to extract region features. Because it is an area of arbitrary shape, this feature can better express irregular areas than rectangular areas. Generally, the regional-based segmentation algorithm of medical exercise rehabilitation image uses the convolutional neural network to extract regional features, and finally uses the classifier to classify the two parts separately. In order to unify these two parts, Fan *et al.* [21] proposed a region-based convolutional neural network algorithm that can end-to-end training, which can accept input from arbitrary shape regions and cover the background when extracting regional features, thereby obtaining a purer target area feature.

In recent years, CNN has been widely used in computer vision tasks. The success of convolutional neural networks is largely due to its strong feature learning ability. Compared with t1he previously constructed features, CNN learned features including image texture, shape, color and other information. Because richer features can more fully characterize the original image, CNN can be used to solve visual tasks.

III. MEDICAL EXERCISE REHABILITATION IMAGE SEGMENTATION ALGORITHM AND DEEP LEARNING ALGORITHM

Before introducing the HFCNN algorithm proposed in this paper, it is necessary to introduce the commonly used medical segmentation algorithm, the segmentation algorithm combined with deep learning and the current research status.

A. INTRODUCTION TO SEGMENTATION ALGORITHMS OF MEDICAL EXERCISE REHABILITATION IMAGE

The segmentation algorithms of medical exercise rehabilitation image are a hotspot in medical image research today, and it is a difficult point in this research field. The ultimate goal of segmentation is to extract specific tissue regions in medical exercise rehabilitation images for more intuitive analysis of the disease. In the analysis of medical exercise rehabilitation image, image segmentation plays an important role. For example, image-guided surgery, three-dimensional visualization, tumor radiotherapy, tissue structure analysis, and therapeutic evaluation require accurate segmentation of medical exercise rehabilitation images, or Image segmentation is based on research.

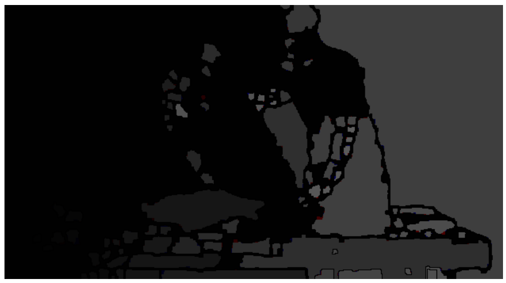
The segmentation algorithm of medical exercise rehabilitation image plays an important role in medical image analysis. In the future development, higher precision and faster computational efficiency [22] is the development direction of medical image segmentation. Although there is a lot of research in the field of segmentation algorithms of medical exercise rehabilitation image at home and abroad, and many good research results have been achieved, in view of the diversity, complexity, gray-scale uniformity of medical images and the accuracy of clinical medical image segmentation for medical images. With high requirements and other characteristics, today's segmentation algorithm does not meet the requirements of clinical medical use. Therefore, the segmentation algorithm of medical exercise rehabilitation image is still the research hotspot of medical image analysis. Although computer segmentation technology is limited in clinical diagnostic applications, with the development of technology, computer-aided diagnosis [23] has a long-term development prospect in application practice.

There are also many methods for the segmentation of medical exercise rehabilitation image, each of which has its own characteristics, such as region segmentation algorithm [24], watershed algorithm [25], cluster segmentation algorithm and threshold segmentation algorithm [26], and edge detection algorithm [27], deep learning algorithms and so on.

(1) The region segmentation algorithm is divided into a region splitting merge algorithm and a region-growing algorithm. The implementation ideas of the two algorithms are different. The core idea of regional growth is to gather pixels with similar properties together and form a region of pixels that are clustered together to achieve the purpose of segmentation. Another interpretation is to select a pixel from the target segmentation as the seed pixel, use this seed pixel as the starting point of the region growth, and then start to merge the pixels with the same or similar pixels to the surrounding, and merges the pixels. As a new seed pixel, it is cycled until the eligible pixels are included in the selected area, and the split task is completed. The region splitting and merging algorithm combines the obtained regions from different parts of an image by splitting. This method differs from the region-growing algorithm in that the region splits the image into many regions, and the regions are different in size and do not overlap, and then merge or split the regions to meet the segmentation requirements.

(2) The main application of the watershed algorithm is to extract approximately consistent objects from the background. In Figure 1(1), height information is added to each pixel, and the result image after segmentation is shown in Figure 1(2).

(1) Original image



(2) Result image

FIGURE 1. Example of medical image segmentation based on watershed algorithm.

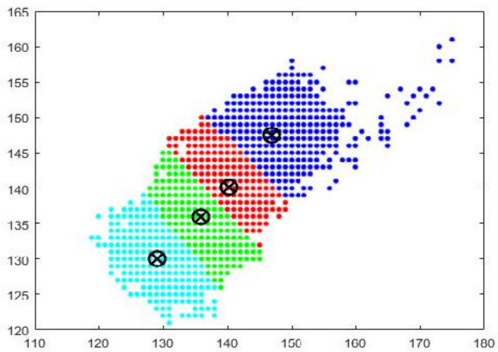
3) Clustering algorithm, the commonly used clustering algorithms in medical image segmentation are K-Means and fuzzy C-means (FCM). FCM is an iteration based on the principle of the lowest two-squares. Optimize the algorithm and prove that the algorithm converges to an extreme value. The FCM algorithm to obtain fuzzy classification of the data set uses the iterative optimization objective function, and the convergence of the algorithm is very good. Figure 2 is an effect diagram of a medical exercise rehabilitation image segmented by a K-Means algorithm.

(4) Threshold segmentation algorithm, which is a very simple and effective image segmentation algorithm. The threshold segmentation algorithm divides the gray level of the gray image to be divided into several parts with one or several thresholds, and the gray levels of the same kind of pixels are similar. The threshold selection formula is as follows:

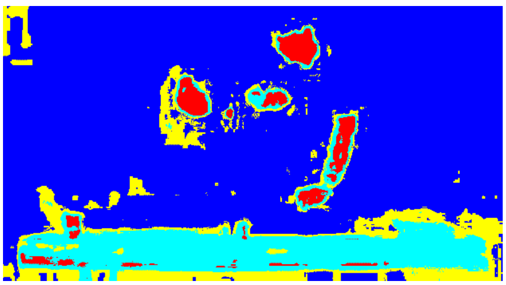
$$T = T[x, y, f(x, y), q(x, y)] \tag{1}$$

where f(x, y) is the gray value at the pixel (x, y) of the image, and q(x, y) is some local feature property in the field near the change point. That is, T is generally a function of points (x, y) and q(x, y).

(1) Original image



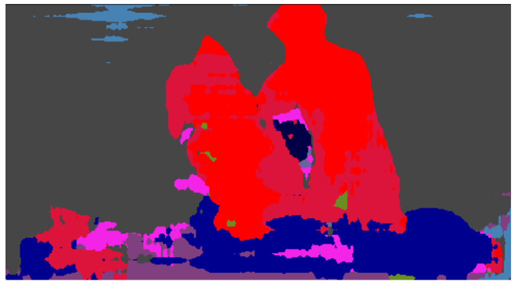
(2) Pixel point clustering result



(3) Result image

FIGURE 2. **Example of medical image segmentation based on K-Means algorithm.**

(1) Original image



(2) Result image

FIGURE 3. **Example of medical image segmentation based on deep learning algorithm.**

(5) Edge detection algorithm, which is based on the gray value of the pixel on the edge of the image is different from other places, the gray value of the pixel on the edge will change significantly, and the image segmentation is realized by detecting the determined uniform region edge. The edge detection algorithm is divided into parallel edge detection algorithm [28] and serial edge detection algorithm [29] according to different processing methods of pixels.

(6) Deep learning algorithm, deep learning algorithm in the field of medical image segmentation has just emerged in recent years, but its development process is very rapid, and it has achieved very good results in breast image segmentation and segmentation of brain lesions, and some results have been clinically applied. Figure 3 is an example of medical image segmentation based on deep learning algorithm

B. INTRODUCTION TO DEEP LEARNING ALGORITHMS

Artificial Intelligence is a rapidly emerging discipline and has become one of the fastest-growing technologies [30]. With the rapid development of computer technology, the news that various computers have defeated humans in certain fields has also been fully the widespread concern of the people in society. Advances in storage technology and the development of big data have made computers very powerful with the support of large amounts of data, and can even defeat humans in some areas. Machine learning is a branch of artificial intelligence. It is an algorithm that uses the learned rules to predict unknown data by automatically analyzing the given classified or labeled data information and obtaining the regularity.

The similarity between the deep learning network and the traditional neural network is that both have similar hierarchical structure, both of which include multi-layer networks such as input layer, hidden layer and output layer, only between adjacent layers. It is connected and there is no connection between each layer. The difference is that traditional neural networks generally have only two to three layers of neural networks, with limited parameters and computational units, limited ability to represent complex functions, and limited learning capabilities. Deep learning networks have more layers; each Layers are capable of extracting abstract features and being able to introduce more algorithms that are efficient. The structure of the deep learning network is similar to the structure of the human brain.

Common algorithms for deep learning algorithms include auto-encoder [31], sparse coding [32], Restricted Boltzmann Machine [33], and deep confidence network [34], convolutional neural networks, etc. Deep learning algorithms require more layers to obtain more abstract feature representations. The core idea is to link many layers together so that information can be transferred layer-by-layer, thus enabling different levels of feature extraction of input information.

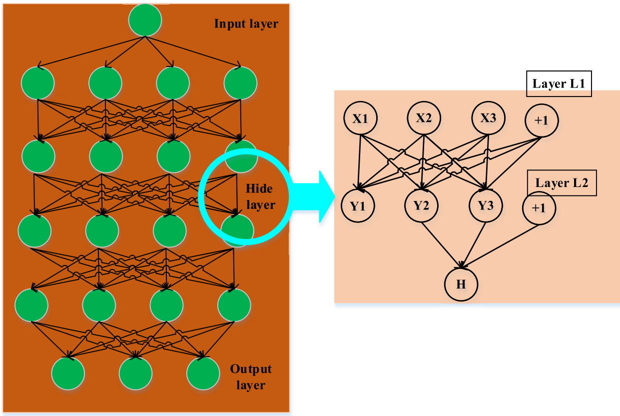


FIGURE 4. The multi-layer learning algorithm of the deep learning algorithm.

Figure 4 shows the multi-layer learning algorithm of the deep learning algorithm.

The deep learning architecture is composed of a multi-layer neural network. Therefore, deep learning is also called deep convolutional neural network. There is no limit to the number of layers in the network. The algorithm can extract features from a large amount of input data and can visualize it. In the 1960s, Hubel and Wiesel [35] studied the neuronal response of the cat's cortex in response to stimuli and discovered a unique network structure of the brain that effectively communicates information and proposes a volume. The concept of the neural network has been introduced into the deep learning algorithm through the continuous efforts and in-depth research of researchers. Through layer-to-layer convolution operations, features of different levels of the image are extracted from shallow to deep. Using the neural network training process, the parameters of the convolution kernel are automatically adjusted, so that the convolution kernel parameters most suitable for image feature extraction are generated unsupervised, and the effective features are extracted.

C. CONVOLUTIONAL NEURAL NETWORKS

The CNN network is a kind of artificial neural network [36]. On the basis of the original multi-layer neural network structure, all nodes of each layer will be calculated forward according to a certain weight and become the input of nodes of the next layer. Each weight connection line is different from each other, and the value of the next layer node is related to all nodes in the previous layer. CNN is designed to recognize two-dimensional images, and its special network structure is highly invariant to translation, scaling, tilting, or other forms of deformation. CNN main components: Convolutional Layer, Pooling Layer and Full Connection Layer. The CNN network has the structural characteristics of weight sharing. As shown in Figure 5, this structure is similar to the biological neural network, which can effectively reduce the complexity of the network algorithm and reduce the number of weights.

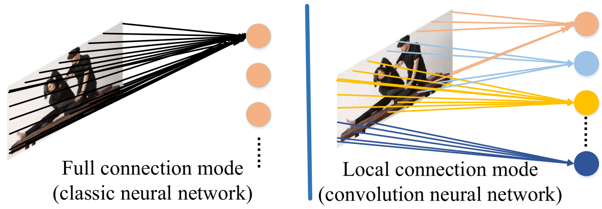


FIGURE 5. Convolutional neural network weight sharing algorithm.

When the input data is a multi-dimensional image, this paper directly uses the original image as input instead of operating as before. In addition, this paper avoided the complicated feature extraction and data reconstruction process in the traditional neural network, which greatly reduces work force input and time consumption. The CNN algorithm makes the multi-layer network structure training at the same time. In the CNN algorithm, the original image data is used as the input data of the hierarchical structure. The information is transmitted to the lower layer by layer through the connection between the layers, and the filtering is performed at each layer. The device can significantly mark the characteristics of the observed data for use in the next layer of learning.

In addition to the convolutional layer and the pooling layer in the CNN network, there are several layers of fully connected layers at the end of the network structure, which can transform the feature map generated by the convolutional layer into a set of eigenvectors of a specified length. Then perform a probability calculation. The AlexNet [37] network is a classic network in the CNN network structure. This network structure is used for image classification and regression. After training, after inputting the image in the network, a classification probability is obtained, such as ImageNet. After the data algorithm is trained in the AlexNet network, it will get a 1000-dimensional vector that is probabilistically calculated to the probability that the input image belongs to a certain classification.

IV. MEDICAL EXERCISE REHABILITATION IMAGE SEGMENTATION ALGORITHM BASED ON HIERARCHICAL FEATURES OF CONVOLUTIONAL NEURAL NETWORKS

In the segmentation algorithm of medical exercise rehabilitation image based on deep learning, the output of the fully connected layer in the CNN is generally taken as the extracted feature, and then the classifier is trained to classify. However, since the output of the fully connected layer is obtained by nonlinear transformation and dimensional reduction of the image multiple times, and contains more image semantic information, this high-level semantic information is useful for classifying the entire image. However, segmentation algorithm of medical exercise rehabilitation image needs to determine the position information of the contour of the object in the image and the category information of the object, so the use of high-level semantic information has certain limitations. In the hierarchical structure of CNN, since the shallow layer of the network contains only a small number of nonlinear transformation operations, the convolution output contains more structured information of the image, which depicts the contour, shape and size of the object in the image. Features that help determine the position of an object. Therefore, this paper proposes the HFCNN algorithm, which samples the convolution output of multiple levels in the convolutional neural network to a uniform size and combines them to construct a hierarchical feature. This hierarchical feature contains the structural information of the lower layer of the network. High-level semantic information, so it can retain the contour information of the object while having semantic information, and has strong expressive ability.

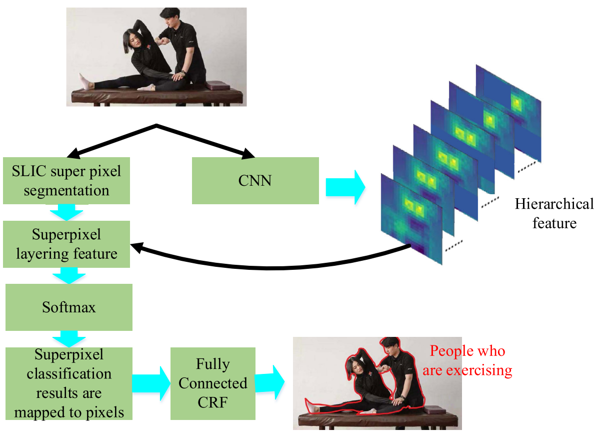


FIGURE 6. The HFCNN algorithm flow chart of this paper.

Although the goal of segmentation algorithm of medical exercise rehabilitation image is to mark each pixel in the image, there are two problems in directly using pixels as the basic unit. The first problem is some pixel noise will affect the result, and the second problem is high-resolution images. The number of pixels is very large, and there are problems in efficiency. Therefore, this paper uses SLIC [38] (Simple Linear Iterative Clustering, SLIC) to perform super pixel segmentation on the original image of medical motion, using the obtained super pixel as the basic unit. The use of super pixels can avoid the influence of individual noise pixels on the classification result and improve the anti-interference ability. In addition, the number of super pixels is much smaller than the number of pixels in the image, so the efficiency of classification can be improved. Finally, the super pixel classification result is mapped to the pixel to obtain the final pixel labeling result. The basic flow of the algorithm proposed in this paper is shown in Figure 6. The HFCNN algorithm first uses the CNN network to extract the hierarchical features. Then the super pixel segmentation method is used for further classification. Finally, using the fully connected conditional random field, the classification efficiency can be improved.

A. HIERARCHICAL FEATURE EXTRACTION BASED ON CONVOLUTIONAL NEURAL NETWORK ALGORITHM

CNN is a hierarchical structure, which is generally connected in series by an input layer, a convolution layer, a pooling layer, a fully connected layer, and an output layer. The output of the previous layer can be used as the input to the next layer, constantly abstracting. The convolutional neural network structure used in this chapter is shown in Table 1.

TABLE 1. Convolutional neural network structure.

	Size	Pad	Stride
Conv1	[11,11,3,64]	[0,0,0,0]	[4,4]
Relu1	--	--	--
Norm1	--	--	--
Max-pool1	[3,3]	[0,1,0,1]	[2,2]
Conv2	[5,5,64,256]	[2,2,2,2]	[1,1]
Relu3	--	--	--
Norm4	--	--	--
Max-pool2	[3,3]	[0,1,0,1]	[2,2]
Conv3	[3,3,256,256]	[1,1,1,1]	[1,1]
Relu3	--	--	--
Conv4	[3,3,256,256]	[1,1,1,1]	[1,1]
Relu4	--	--	--
Conv5	[3,3,256,256]	[1,1,1,1]	[1,1]
Relu5	--	--	--
Max-pool5	[3,3]	[0,1,0,1]	[2,2]
Fc6	[6,6,256,4096]	[0,0,0,0]	[1,1]
Relu6	--	--	--
Fc7	[1,1,4096,4096]	[0,0,0,0]	[1,1]
Relu7	--	--	--
Fc8	[1,1,4096,1000]	[0,0,0,0]	[1,1]
Softmax	--	--	--

Since the input image is a color image, the number of channels in the first convolutional layer is three, and convolution is performed using 64 convolution kernels of size 11×11. The excitation function used is Relu, and then normalized. Get 64 feature maps. After the maximum pooling of 3×3, the dimension of the feature graph is reduced, and the maximum pooling can guarantee the rotation invariance of the convolutional neural network. Then, after multiple convolutions and pooling operations, the high-level feature representation is obtained, and then the full-connection is used to aggregate the features and then perform softmax classification.

In the various tasks of computer vision, the feature representation of the image has a great influence on the result, so constructing a strong image feature representation is the key to ensure better segmentation results. Image semantic segmentation is a fusion of two tasks: image segmentation and image segmentation. Image segmentation needs to divide different objects, that is, determine the specific location of the object. Image classification needs to judge the category of the object. Image semantic segmentation needs to use the extracted image features to complete segmentation and classification at the same time. Therefore, in order to obtain better semantic segmentation results, it is necessary to ensure that the extracted image features can be segmented and classified well.

The hierarchical structure of CNN provides the possibility to extract strong feature representations. The shallow convolution output of the network contains many low-level, structured information such as edges, corners, points, and outlines of objects. The convolution output of the upper layer

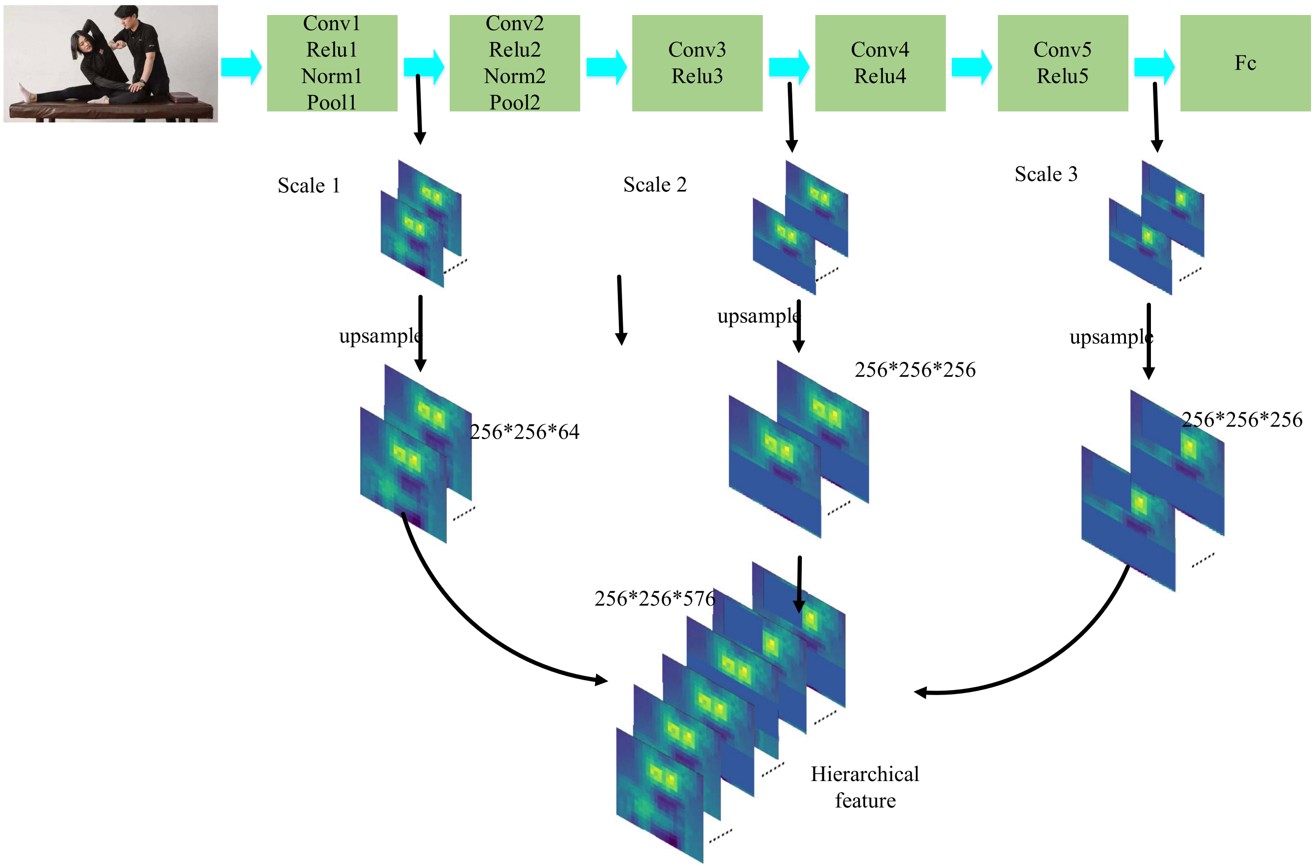


FIGURE 7. **Hierarchical feature extraction.**

of the network is the low-level feature that is continuously obtained by convolution and pooling. It is an abstraction of low-level features and therefore contains a lot of semantic information. Because the object information contained in the convolution output of different layers of the network is different, different levels of convolution output can be combined to construct a hierarchical image feature representation. In this paper, CNN is used to extract multiple levels of features, as shown in Figure 7.

In theory, the more levels of convolution output used, the stronger the ability to express hierarchical features. However, due to the sliding window mechanism in CNN, there is redundancy in the convolution output of some layers. In addition, the more feature maps used, the more calculations are required, which reduces efficiency. Therefore, in view of the above reasons and specific experiments, the convolution output of conv1 (64), conv3 (256), and conv5 (256) is used to construct a hierarchical feature, because the feature map size of each convolution layer is inconsistent, so it is used. The nearest neighbor interpolation method samples the feature map of each level to the original image size, and finally combines the up-sampled feature maps to obtain hierarchical features.

B. SUPER PIXEL SEGMENTATION AND CLASSIFICATION

Before super pixels are proposed, image segmentation is often performed using pixels as a unit. However, this method does not consider the spatial positional relationship between pixels, and efficiency is another problem. A super pixel is an irregular area composed of adjacent pixels, which are similar in texture, color, brightness, and the like. By super pixel segmentation of the original image, a plurality of super pixel regions aggregated by pixels of similar features can be obtained. When super pixel is used as the basic unit for segmentation, the influence of noise points on the segmentation result can be reduced. In addition, because the number of super pixels is relatively small, spatial position information between super pixels can be incorporated into the segmentation process to improve segmentation effect.

The main idea of the graph-based super pixel segmentation method is to regard the pixels in the image as nodes in the weighted undirected graph. The edges in the graph represent the relationship of adjacent pixels, and the weight of the edges can be similar to the pixels. Alternatively, the degree of difference is expressed, and finally the pixels are divided by using different rules to complete the segmentation.

SLIC is a super pixel segmentation algorithm based on gradient descent. The algorithm has the characteristics of low time complexity, fast segmentation speed and approximate super pixel shape. SLIC uses a K-means-like approach to continually divide pixels into nearest seed points to generate super pixels. The seed point is initialized at the beginning of the algorithm. The number of seed points is the number of pre-segmented super pixels. Then, the distance between each pixel and its surrounding seed points is calculated separately. The similarity between the pixel and the seed point is measured by the distance., the pixel that is most similar to the seed point is given the same mark as the seed point. By iterating, the mark of the pixel is constantly updated, and the iteration ends when the mark of the pixel no longer changes, and finally the pixel with the same mark constitutes the super pixel. Wherein, the coordinate distance information and the color distance information between the pixels are incorporated when calculating the distance between the pixel and the seed point. The specific calculation formula is as follows.

$$d_{lab} = \sqrt{(l_i - l_j)^2 + (a_i - a_j)^2 + (b_i - b_j)^2} \tag{2}$$

$$d_{xy} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{3}$$

$$D_i = d_{lab} + (m/s)d_{xy} \tag{4}$$

The variable d_{lab} is the color distance between pixels, d_{xy} is the coordinate distance between pixels, and D_i is the final distance between pixels, representing the similarity between pixels. The variable s is the distance between the seed points. Assuming that there are N pixels in total, and the number of divisions of the set super pixel is K, the distance between the seed points is approximately $\sqrt{N/K}$. The variable m is the balance factor used to control the color distance and the coordinate distance in the final distance. The smaller the distance between pixels is, the higher the similarity between pixels is. The super pixel segmentation effect of the SLIC algorithm is shown in Figure 8.

It can be seen from Figure 8 that the original image is divided into adjacent regions, each of which is approximately equal in size, and the colors and textures of the pixels in each region are substantially the same. Super pixels can be used to represent pixels by grouping similar pixels onto one super pixel.

Because the hierarchical feature is obtained by up-sampling multiple convolutional layer outputs to the original image size, it can be represented as a 3-dimensional array $H \in R^{N\times H\times W}$, where the variable N is the number of feature maps, H and W is the height and width of the feature map, respectively. Each feature map can be viewed as a feature representation of the image, so the features of one pixel in the image can be represented as $P \in R^N$. A super pixel is a representative of similar pixels adjacent to some positions in an image, and all pixel feature values in the super pixel can be averaged to obtain a super pixel feature representation $F_P \in R^N$. Thus, the features of all the super pixels included in the entire image can be expressed as $F \in R^{M\times N}$, where


(1) Original image


(2) Result image

FIGURE 8. **The super pixel segmentation effect of the SLIC algorithm.**

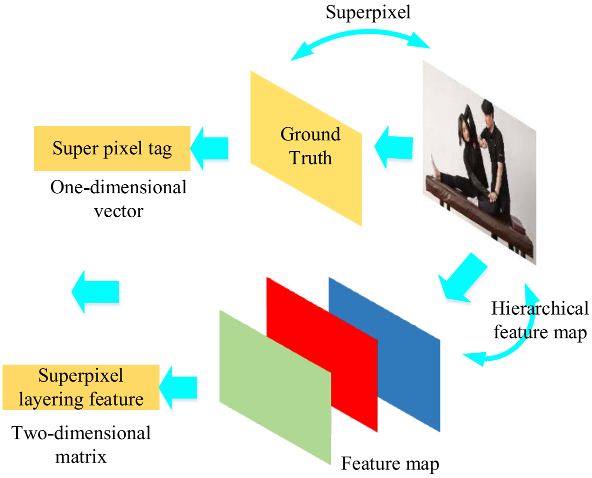


FIGURE 9. **Extracting super pixel features.**

M represents the number of super pixel blocks included in the image, and N represents the feature dimension, that is, the number of feature maps. The feature extraction process of super pixel is shown in Figure 9.

The feature representation of the super pixel is obtained, and the softmax classifier can be trained. The loss function is:

$$J(\theta)=-\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} l\{y^{(i)}=j\}\log\frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}}\right]+\frac{\lambda}{2}\sum_{i=1}^{k}\sum_{j=0}^{n}\theta_{ij}^2 \tag{5}$$

The variable θ_j is the weight vector that requires training to learn, and n is its corresponding dimension. The variable m is the number of samples in the training set, and k is the number of categories in all samples in the training set. The formula

$l\{y^{(i)} = j\}$ is the indicator function. When $y^{(i)} = j$, the result is 1, otherwise it is 0. In order to avoid overfitting, a regular term is added, and λ is a regularization parameter. After the training is completed, the weight vector can be obtained. The probability distribution of super pixel $S_i, i \in \{1, \ldots, t\}$ can be expressed as $d_{Si,a}$. The specific calculation process is as follows:

$$y_j = \theta_j^T x^{(i)} \tag{6}$$

$$d_{S_{i,a}} = e^{y_{j,a}} / \sum\nolimits_{b \in classes} e^{y_{j,b}} \tag{7}$$

j∈{1, …, k}, k is the number of categories. The variable a indicates the category to which the super pixel belongs, so that the probability that each super pixel belongs to each class is obtained. In order to obtain the probability distribution of the pixels, the probability that the super pixel belongs to each class can be mapped onto the pixels it contains. The specific mapping relationship is as follows:

$$U_{PI}^C = U_{Sj(\varphi(i)=j)}^C \tag{8}$$

The variable U_{Sj}^C is the probability that super pixel S_j belongs to category C, and U_{PI}^C is the probability that pixels in super pixels belong to category C. The formula $(\varphi(i) = j)$ means that the category of the super pixel is mapped to the pixel it contains. After mapping, you can get the probability distribution of pixels belonging to each class.

C. FULLY CONNECTED CONDITIONAL RANDOM FIELD ALGORITHM CONSTRUCTION

The most basic conditional random field algorithm includes a unitary potential energy defined on a pixel or image area and a pair of potential energy defined between adjacent pixels or adjacent areas. Because the paired potential energy is only built on adjacent pixels or image regions, it can well model the spatial positional relationship between adjacent pixels or adjacent regions, but model the distance between pixels or regions that are far apart. Poor ability causes excessive smoothing at the boundary between different objects in the image. Therefore, if the ability to express paired potential energy to long-distance spatial relationships between pixels or image regions can be enhanced, the problem that different object boundaries are excessively smoothed can be alleviated.

The maximum difference between a fully connected conditional random field and the basic conditions described above is that the paired potential energy is established on all pairs of pixels in the image. As shown in Figure 10, a connection is established between the pixels in the fully connected conditional random field. Because the relationship between all pairs of pixels is considered, a good segmentation effect can be obtained.

The energy function of the fully connected conditional random field is as follows:

$$E(x) = \sum_i \psi_u(x_i) + \sum_{i<j} \psi_p(x_i, x_j) \tag{9}$$

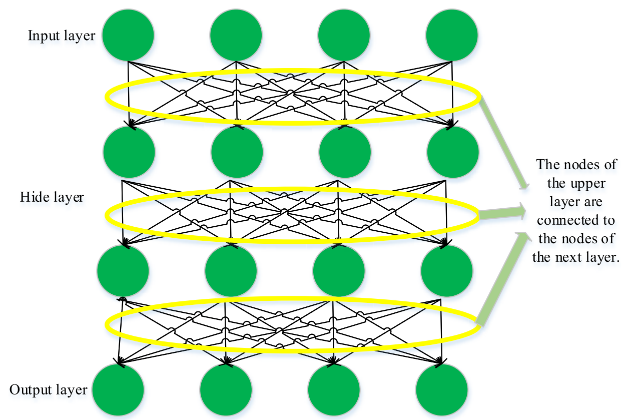


FIGURE 10. **Fully connected conditional random field algorithm.**

The variable x_i and x_j represent the pixels in the image and the unitary potential energy $\psi_u(x_i)$ are calculated by the classifier to obtain the class probability distribution of each pixel.

$$\psi_u(x_i) = -\log(P(x_i)) \tag{10}$$

The variable $P(x_i)$ represents the probability distribution of the pixel x_i. The variable $\psi_p(x_i, x_j)$ represents the paired potential energy between pixels. The specific form is as follows:

$$\psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^{k} w^{(m)} k^{(m)}(f_i, f_j) \tag{11}$$

The variable $\mu(x_i, x_j)$ is a label interchangeability function. The formula $\mu(x_i, x_j) = [x_i \neq x_j]$ is used to penalize the case where adjacent similar pixels are marked as different labels. The variable $k^{(m)}(f_i, f_j)$ represents a Gaussian kernel function, and $w^{(m)}$ represents a weight of a Gaussian kernel function. The variable f_i and f_j denotes the feature vector of the pixel i and j in the feature space. The specific Gaussian kernel function is as follows:

$$k(f_i, f_j) = w^{(1)} e^{-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}} + w^{(2)} e^{-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}} \tag{12}$$

Two kernel functions are used here. The position and color information of the pixel is used in the first kernel function. The variable p_i and p_j represent the coordinate positions of pixels i and j, and the variable I_i and I_j represent the color values of pixels i and j. The variable θ_α and θ_β control the degree of similarity between the pixels and the degree of color similarity, respectively. In the second kernel function, only the position information of the pixel is used, which is used to remove some isolated regions and play a smoothing role. The algorithm transforms the message passing process into filtering in the feature space. With the kernel function, the time complexity of message passing can be reduced from the second power to the first power, which further improves the efficiency of the algorithm.

V. EXPERIMENTS AND RESULTS

A. MEDICAL EXERCISE REHABILITATION IMAGE ACQUISITION DEVICE BASED ON INTERNET OF THINGS

Internet of Things (IoT) is an extended and extended network based on the Internet. It combines various information sensing devices with the Internet to form a huge network that can be realized at any time, any place, people, machines, Interconnection of objects [39]. From the perspective of communication objects and processes, the basic characteristic of the Internet of Things is the information interaction between objects and objects and people is the core of the Internet of Things. The basic characteristics of the Internet of Things can be summarized as overall perception, reliable transmission, and intelligent processing. Overall Perception—you can use the sensing devices such as radio frequency identification, QR code, and smart sensors to sense various types of information acquired. Reliable transmission - through the integration of the Internet and wireless networks, the information of objects can be transmitted in real time and accurately for information exchange and sharing. Intelligent Processing—uses various intelligent technologies to analyze and process the perceived and transmitted data and information to realize intelligent monitoring and control.

With the rapid development of IoT-related industries and the widespread application of intelligent mobile terminals, mobile terminals are playing an increasingly important role in the rich Internet of Things applications. The premise of IoT application is the collection of massive information, and the combination of multi-source information collection and intelligent mobile devices allows the Internet of Things to be further integrated into people's production and life. In this paper, IoT technology is used to design a collection device based on Internet of Things technology, and medical images are collected as a research database. Figure 11 is a design diagram of an acquisition device based on the Internet of Things.

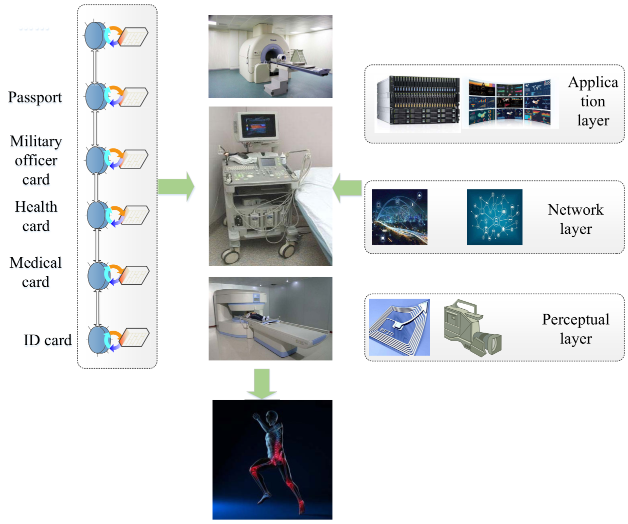


FIGURE 11. IoT-based acquisition device design.

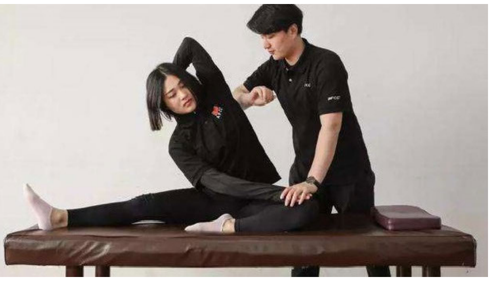
B. IMAGE PREPROCESSING AND DATA ENHANCEMENT

In order to improve the performance of the convolutional neural network, preprocessing is a necessary step in building the data set.

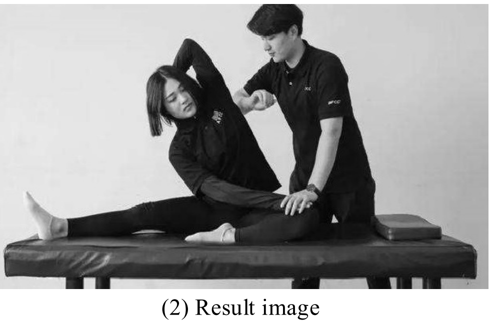
Due to the different acquisition methods, different magnifying glass magnifications, different presentation modes, and the inconsistent types of medical exercise rehabilitation images collected, the medical exercise rehabilitation images in the data set are different in shape, clear and dark, and large. There is a large difference in imaging between the two, which will affect the segmentation algorithm of medical exercise rehabilitation image, so it is necessary to pre-process it before segmentation. The following are the detailed steps of the medical exercise rehabilitation image preprocessing done in this paper.

1) GRAYSCALE

Most of the data sets are grayscale pictures, and a few are color pictures. In order to facilitate the subsequent steps and improve the network processing speed, it is necessary to perform grayscale processing on the color pictures, and convert them into grayscale pictures, such as Figure 12 shows.



(1) Original image



(2) Result image

FIGURE 12. Comparison of before and after grayscale.

2) STANDARDIZATION

In order to make the pixels in the image obey the standard normal distribution with a mean of zero and a variance of one, calculate the mean and standard deviation of the gray values in the data set, and then subtract the average of the gray values of all the images in the data set. The value is divided by its standard deviation. In the test, the picture to be tested also needs to subtract the average value of the gray value corresponding to the training process, and then divide by the standard deviation of the gray value corresponding to the training process. The standardized calculation formula is shown in equation (13).

$$I(x) = (x - \mu)/\sigma \tag{13}$$

In the above formula, x is the gray value of the image pixel, μ is the average value of the gray value, and σ is the standard deviation of the gray value.

3) HISTOGRAM EQUALIZATION

Due to the small image contrast and background contrast in the dataset, it may make the algorithm difficult to distinguish between the nucleus and the background. Therefore, the dataset is subjected to histogram equalization preprocessing to improve the contrast. By using a certain way to change the distribution of the gray histogram of the image from a relatively concentrated to a relatively uniform distribution, it can significantly increase the contrast of the image, so that the content contained in the image is clearer, which is beneficial to the subsequent deal with. As shown in Figure 13, (a) is the picture before the histogram equalization process, and it can be seen that the contrast between the cell and the background is poor, and (b) is the picture after the histogram equalization, the gray of the cell and the background. Significant differences in degrees help the network extract more features.

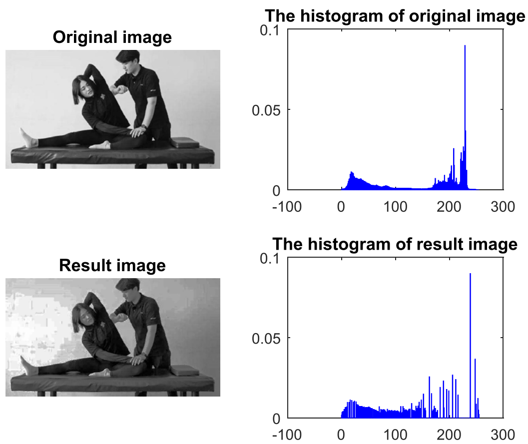


FIGURE 13. **Comparison of before and after histogram equalization.**

4) FILTER SMOOTHING

In the process of image acquisition, it is often interfered and polluted by various noises, which results in the decrease of signal to noise ratio and the edges of the cells and the background become blurred. In order to improve the signal-to-noise ratio of the image and reduce the influence of noise on the image, usually preprocessing that requires filtering of the image. In this paper, the most commonly used Gaussian smoothing filter is used to preprocess the image. The calculation formula of the filtering is shown in equation (14).

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-n-1)^2+(y-n-1)^2}{2\sigma^2}} \tag{14}$$

where the variable σ is the standard deviation of the gray value and the variable n is the dimension of the Gaussian convolution kernel.

5) ZCA WHITENING

The whitening process can reduce the redundancy of the picture information. There is a certain relationship between adjacent pixel blocks in the original image. After the image is whitened, the pixels have the same variance and the correlation is significantly reduced. ZCA whitening and PCA whitening are the two most common methods of whitening. This article will use the former to preprocess the image.

The best way to generalize the convolutional neural network algorithm is to use more data for training. Training the CNN algorithm in this case will make the network converge slowly and the recognition effect is relatively poor. The solution to this problem is data enhancement, which prevents the algorithm from overfitting and helps to build a simpler and better-generalized robust algorithm. This article uses the following methods to amplify the data set:

(1) Mirror surface symmetry

The image is respectively symmetrical, left, and right symmetrical, the corresponding label image is also transformed accordingly, and the number of images will be expanded to four times;

(2) Translation

The image is shifted from −60 pixels in the up and down and left and right directions to a pixel of 10 pixels in steps of 10 pixels, and the corresponding label image is also transformed accordingly.

(3) Rotation

Rotate the image 90 degrees clockwise, 2×90 degrees, 3×90 degrees clockwise, and rotate the corresponding label image accordingly.

(4) Zoom

The scaling of the image is changed from 0.9 to 1.1 in steps of 0.05, and the corresponding annotation image is transformed accordingly.

(5) Miscut transformation

Separate the x or y coordinates of all the pixels, and translate the corresponding y or x coordinates proportionally to the vertical distance of the pixel to the x and y axes, and correspondingly mark the corresponding images transformation

(6) Random cropping

The image is cropped using a random image difference method, and the corresponding label image is transformed accordingly.

(7) Adjust the contrast

In the HSV color space of the image, keep the hue H constant, and use the exponential operation (exponential factor starts from −0.6 to 0.6, and steps in 0.2) to change the saturation S and brightness V of each pixel in the image, and the corresponding label image is also transformed accordingly.

(8) Noise disturbance

The RGB values of each pixel in the image are randomly perturbed by Gaussian noise or salt and pepper noise, and the corresponding annotation images are transformed accordingly.

(9) Color conversion

Principal component analysis is performed on the three-color channels of RGB by PCA, then Gaussian perturbation is performed on several extracted principal components, and the corresponding annotation images are transformed accordingly.

(10) The image is blurred

Convolution with different templates produces a blurred image, and the corresponding labeled image is transformed accordingly.

C. SELECTION OF NETWORK LAYERS

In the process of using the deep learning algorithm, since the number of hidden layer nodes of the CNN algorithm and the SAE algorithm is random, and the offsets and weights in the hidden layer are randomly generated, adjusting the number of hidden layer nodes affects the training result. In theory, the more hidden nodes, the better the algorithm can fit the training data. However, too many nodes will result in increased training time. The selection of the number of hidden layer nodes requires a comprehensive consideration of the computational complexity and the degree of information retention.

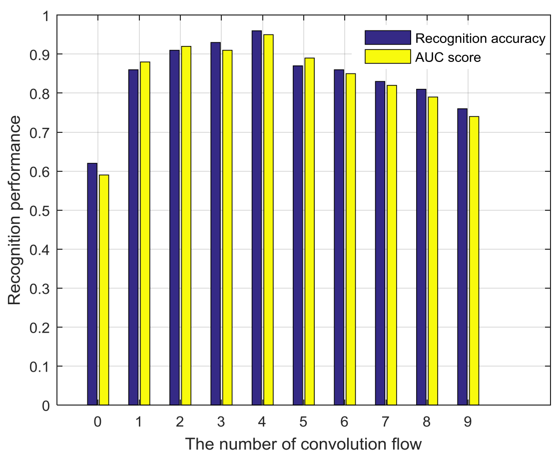


FIGURE 14. Recognition accuracy rate and AUC score with the number of convolution flow curve.

In order to select the appropriate number of network layers, a layer-by-layer increase in the form of a convolution stream is used. A convolution stream includes a convolution layer and a pooling layer, which are recursively incremented from one convolution stream. The experimental results are shown in Figure 14. As can be seen from Figure 14, as the number of convolutional streams increases, the recognition accuracy and AUC values will gradually increase and then decline.

TABLE 2. Experimental results of different convolutional layer combinations.

Method	MPA	MCA
Conv1+Conv3+Conv5+CRF	0.821	0.712
Conv1+Conv4+Conv5+CRF	0.815	0.701
Conv1+Conv3+Conv5	0.794	0.697
Conv3+Conv4+Conv5	0.789	0.691
Conv1+Conv5	0.763	0.617
Conv1+Conv3	0.751	0.655
Conv3+Conv5	0.743	0.653

When the number of convolution streams is five, the recognition effect is the best. After selecting the number of convolution streams, select three fully connected layers for classification. The network algorithm trains iterations 5000 times, and saves the trained algorithm every 1000 iterations. The initial learning rate is set to 1.0×10^{-3}, and the learning rate attenuation mode is Poly, power = 0.5.

D. SELECTION OF DIFFERENT CONVOLUTION COMBINATIONS

In this paper, the average pixel accuracy (MPA) and the average class precision (MCA) are used to quantitatively evaluate the pixel labeling results. The specific calculation methods of the two evaluation indicators are as follows:

$$MPA = \frac{pixel\_ok}{pixel\_total} \tag{15}$$

$$MCA = \frac{\sum\limits_{i \in classes}^{n} \frac{pixel\_ok[i]}{pixel\_total[i]}}{n} \tag{16}$$

In MPA calculations, $pixel\_ok$ represents the number of pixels in the test set where all images are classified correctly, and $pixel\_total$ represents the total number of pixels in all images in the test set, dividing by the average pixel precision. The average pixel accuracy characterizes the classification ability of the classification algorithm. In the calculation of MCA, $pixel\_ok[i]$ represents the number of pixels in which i-th object in the test set is classified correctly, and $pixel\_total[i]$ represents the total number of pixels occupied by the i-th object in the test set, and the average pixel precision of each type of object is obtained. After accumulating and then dividing by the number of categories, the average class precision is obtained. The average class precision characterizes the proportion of pixels that each type of object occupies correctly. The results of the various methods in the experiment are shown in Table 2.

In Table 2, Conv1+Conv3+Conv5+CRF represent the use of Conv1, Conv3, and Conv5 to build hierarchical features, and then use softmax classification, and finally post-process with fully connected CRF. Conv1+ Conv3+Conv5 mean that only hierarchical features are built by using Conv1, Conv3, and Conv5 and softmax. As can be seen from the above table, when using Conv1, Conv3, Conv5 to build hierarchical features and finally using CRF, the highest average pixel precision and average class precision are achieved. The table compares MPA and MCA when using other levels of convolution output to build hierarchical features by using CRF. The results in the table indicate the validity of using CRF and the convolution of three layers of Conv1, Conv3, and Conv5. The outputs are combined to give a better representation of the features.

The ROC curve visually reflects the classification performance of different CNN networks. AUC is the area under the curve of the ROC. The closer the AUC value is to one, the better the classification effect. Figure 15 is a ROC curve of different convolutional layer combinations, with the abscissa being 1-Specificity and the ordinate being Sensitivity.

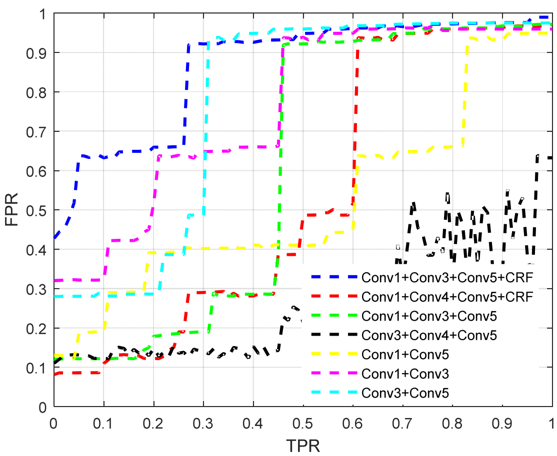


FIGURE 15. **ROC curves of different convolutional layer combinations.**

It can be seen from Figure 15 that the Conv1+Conv3+Conv5+CRF algorithm is closest to the point (0, 1), that is, the segmentation effect is the best. It can also be seen that although Conv1+Conv4+Conv5+CRF and Conv1+Conv3+Conv5 respectively have significant improvements in various indicators, the segmentation performance of the Conv1+Conv3+Conv5+CRF method is still not high.

E. PERFORMANCE COMPARISON OF NETWORK MODELS

The performance of the segmentation algorithm of medical exercise rehabilitation image based on the hierarchical feature of convolutional neural network in this paper is shown in Figure 16. The abscissa represents the iteration number of deep neural network algorithm training, and the ordinate representation algorithm is training corresponding accuracy performance on sets and test sets.

It can be seen from Figure 16 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 4000 iterations, the performance of the algorithm on the training set and the test set tends to be stable, indicating that the algorithm gradually converges.

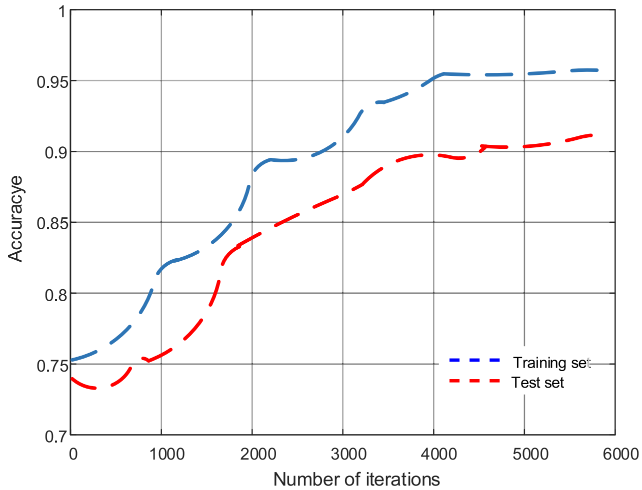


FIGURE 16. **Deep neural network training performance.**

In order to verify the superiority of the segmentation algorithm of medical exercise rehabilitation image based on the hierarchical feature of convolutional neural network, this paper uses the same data set to compare the performance of the proposed algorithm with the current popular algorithms and classical algorithms. The results are as shown in Table 3 and Figure 17.

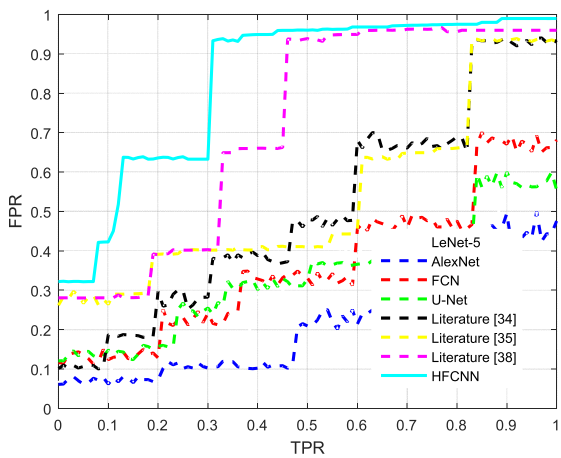


FIGURE 17. **Comparison of ROC curve with classical CNNs.**

As shown in Tables 3 and 4, the segmentation results obtained by the method of this paper are compared with those obtained in the literature [33], [34] and [37], and on the Accuracy, Sensitivity and Specificity indicators improved. Compared with U-Net, although the accuracy of this method is not much improved in Accuracy, Sensitivity and Specificity, in the segmentation speed, since this paper only needs to input the original image to be segmented directly into training. In the parameter algorithm, it takes only 31ms to divide each image on average, so the segmentation speed of this paper is obviously improved and the real-time performance is better. This paper also compares with the FCN

TABLE 3. Performance comparison between proposed method and others.

Model	Accuracy	Sensitivity	Specificity	AUC
LeNet-5	0.745	0.752	0.751	0.731
AlexNet	0.605	0.610	0.602	0.582
FCN	0.722	0.732	0.715	0.725
U-Net[40]	0.861	0.825	0.861	0.857
Literature [33]	0.891	0.897	0.856	0.885
Literature [34]	0.911	0.931	0.925	0.905
Literature [37]	0.721	0.732	0.787	0.712
HFCNN	0.935	0.951	0.921	0.926

TABLE 4. Time taken for a single image using the algorithm and comparison algorithm of this paper.

Model	Time（s）
LeNet-5	0.745
AlexNet	0.605
FCN	0.722
U-Net[40]	0.861
Literature [33]	0.891
Literature [34]	0.091
Literature [37]	0.121
HFCNN	0.031

used for image segmentation in deep learning. Because FCN divides the details of the image at the time of segmentation, the segmentation accuracy of this method is not satisfactory.

The traditional manual extraction of single features has a poor classification effect. In order to obtain better classification results, it is necessary to extract features in many aspects, but the artificial feature extraction is more complicated and the generalization ability is poor. The classification method based on deep learning can be simpler, more convenient, and has a strong self-learning ability, but the accuracy needs to be further improved. As can be seen from the above chart, compared with the classical convolutional neural network algorithm and the above references, the proposed algorithm has higher superiority in accuracy, sensitivity, specificity and AUC value, and can effectively improve medical moving images. The segmentation accuracy rate reduces the workload of doctors and reduces the rate of misdiagnosis and missed diagnosis of diseases.

VI. CONCLUSION

The great success of deep learning methods in the field of computer vision processing has led many researchers to apply it to the different tasks of medical exercise rehabilitation image processing. Deep learning algorithms play an important role in helping doctors make accurate and effective diagnosis, especially in image recognition and classification, image localization and detection, lesion segmentation, image registration and fusion, computer-aided diagnosis and microscopic image analysis. The segmentation algorithm of medical exercise rehabilitation image is a classification of pixel granularity that can determine the position and type of objects in an image relatively accurately. Aiming at the shortcomings of traditional medical image recognition methods, a medical exercise rehabilitation image segmentation algorithm based on hierarchical features of convolutional neural networks is proposed. In this paper, super pixel is used as the basic unit, and the image features are extracted by deep learning method to train the classifier. Finally, the classification result of the fully connected condition random field modified classifier is constructed to improve the consistency and continuity of the regional pixel markers. The research results show that the proposed method not only has significant improvement in segmentation accuracy, but also reduces the running time of the algorithm, reduces the workload of doctors, and reduces the rate of misdiagnosis and missed diagnosis of diseases.

REFERENCES

[1] P. Aparna and P. V. V. Kishore, "Biometric-based efficient medical image watermarking in E-healthcare application," *IET Image Process.*, vol. 13, no. 3, pp. 421–428, 2019.

[2] S. Queirós, P. Morais, D. Barbosa, J. C. Fonseca, J. L. Vilaça, and J. D'Hooge, "MITT: Medical image tracking toolbox," *IEEE Trans. Med. Imag.*, vol. 37, no. 11, pp. 2547–2557, Nov. 2018.

[3] B. Li, F. Sun, and Y. Zhang, "Building recognition using gist feature based on locality sensitive histograms of oriented gradients," *Pattern Recognit. Image Anal.*, vol. 29, no. 2, pp. 258–267, 2019.

[4] K. V. V. Kumar and P. V. V. Kishore, "Indian classical dance mudra classification using HOG features and SVM Classifier," *Int. J. Electr. Comput. Eng.*, vol. 7, no. 5, pp. 2537–2546, 2017.

[5] J. Sun, Y. Zhao, and S. Wang, "Improvement of SIFT feature matching algorithm based on image gradient information enhancement," *J. Jilin Univ.*, vol. 43, no. 4, pp. 335–342, 2018.

[6] H. C. Nguyen, S.-H. Liang, C.-H. Su, J.-H. Lin, C.-C. Chien, and S.-S. Chen, "Enzymatic production of biodiesel from insect fat using methyl acetate as an acyl acceptor: Optimization by using response surface methodology," *Energy Convers. Manage.*, vol. 158, pp. 168–175, Feb. 2018.

[7] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "TextonBoost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," *Int. J. Comput. Vis.*, vol. 81, no. 1, pp. 2–23, Dec. 2007.

[8] N. Noormohamadi, P. Adibi, and S. M. S. Ehsani, "Semantic image segmentation using an improved hierarchical graphical model," *IET Image Process.*, vol. 12, no. 11, pp. 1943–1950, 2018.

[9] X. Song, S. Jiang, and L. Herranz, "Multi-scale multi-feature context modeling for scene recognition in the semantic manifold," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2721–2735, Jun. 2017.

[10] F. Hoseini, A. Shahbahrami, and P. Bayat, "AdaptAhead optimization algorithm for learning deep CNN applied to MRI segmentation," *J. Digit. Imag.*, vol. 32, no. 1, pp. 105–115, 2019.

[11] T. Xiao, L. Lei, W. Qin, S. Yu, Z. Li, and L. Kai, "Comparison of transferred deep neural networks in ultrasonic breast masses discrimination," *Biomed Res. Int.*, vol. 2018, Jun. 2018, Art. no. 4605191.

[12] J. Lin, L. Zhao, S. Li, R. Ward, and Z. J. Wang, "Active-learning-incorporated deep transfer learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4048–4062, Nov. 2018.

[13] G. N. Girish, B. Thakur, S. R. Chowdhury, A. R. Kothari, and J. Rajan, "Segmentation of intra-retinal cysts from optical coherence tomography images using a fully convolutional neural network model," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 1, pp. 296–304, Jan. 2019.

[14] J. Liu, F. Chen, M. Zhu, X. Zhang, L. Zhang, H. Liao, and C. Pan, "A cascaded deep convolutional neural network for joint segmentation and genotype prediction of brainstem gliomas," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1943–1952, Sep. 2018.

[15] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.

[16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–649, Apr. 2017.

[17] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2017.

[18] S. Zheng, S. Jayasumana, V. Vineet, Z. Su, D. Du, C. Huang, P. H. S. Torr, and B. Romera-Paredes, "Conditional random fields as recurrent neural networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 179–189, doi: 10.1109/ICCV.2015.179.

[19] K.-K. Maninis, J. Pont-Tuset, L. Van Gool, and P. Arbeláez, "Convolutional oriented boundaries: From image segmentation to high-level tasks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 819–833, Apr. 2018.

[20] C. J. Moran, E. M. Rowell, and C. A. Seielstad, "A data-driven framework to identify and compare forest structure classes using LiDAR," *Remote Sens. Environ.*, vol. 211, pp. 154–166, Jun. 2018.

[21] F. Li, D. A. Clausi, L. Xu, and A. Wong, "ST-IRGS: A region-based self-training algorithm applied to hyperspectral image classification and segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 3–16, Jan. 2018.

[22] J. Shen, X. Dong, X. Jin, L. Shao, F. Porikli, and J. Peng, "Submodular function optimization for motion clustering and image segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2637–2649, Sep. 2019.

[23] M. R. Mohebian, H. R. Marateb, M. A. Mañanas, F. Mokarian, and M. Mansourian, "A hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (HPBCR) using optimized ensemble learning," *Comput. Struct. Biotechnol. J.*, vol. 15, pp. 75–85, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2001037016300666

[24] T. Nyholm, S. Svensson, S. Andersson, J. Jonsson, M. Sohlin, C. Gustafsson, E. Kjellén, K. Söderström, P. Albertsson, L. Blomqvist, B. Zackrisson, L. E. Olsson, and A. Gunnlaugsson, "MR and CT data with multiobserver delineations of organs in the pelvic area—Part of the Gold Atlas project," *Med. Phys.*, vol. 45, no. 3, pp. 1295–1300, 2018.

[25] M. A. Zengqiang, Y. Wang, and Z. Song, "The color tread image segmentation based on improved labeled watershed," *J. Graph.*, vol. 39, no. 1, pp. 36–42, 2018.

[26] W. Yan and Y. Gelan, "A new brain magnetic resonance imaging segmentation algorithm based on subtractive clustering and fuzzy C-means clustering," *J. Med. Imag. Health Inform.*, vol. 8, no. 3, pp. 602–608, 2018.

[27] M. Mittal, A. Verma, B. Kaur, M. Sharma, L. M. Goyal, T.-H. Kim, S. Roy, and I. Kaur, "An efficient edge detection approach to provide better edge connectivity for image analysis," *IEEE Access*, vol. 7, pp. 33240–33255, 2019.

[28] J. Cao, L. Chen, Y. Tian, and M. Wang, "Implementing a parallel image edge detection algorithm based on the Otsu-canny operator on the Hadoop platform," *Comput. Intell. Neurosci.*, vol. 2018, May 2018, Art. no. 3598284.

[29] F. Jamour, S. Skiadopoulos, and P. Kalnis, "Parallel algorithm for incremental betweenness centrality on large graphs," *IEEE Trans. Parallel Distrib. Syst.*, vol. 29, no. 3, pp. 659–672, Mar. 2018.

[30] S. Feldstein, "The road to digital unfreedom: How artificial intelligence is reshaping repression," *J. Democracy*, vol. 30, no. 1, pp. 40–52, 2019.

[31] M. S. Seyfioğlu, A. M. Özbayoglu, and S. Z. Gürbüz, "Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 4, pp. 1709–1723, Aug. 2018.

[32] W. Chu, H. Xue, and C. Yao, "Sparse coding guided spatiotemporal feature learning for abnormal event detection in large videos," *IEEE Trans. Multimedia*, vol. 21, no. 1, pp. 246–255, Jan. 2019.

[33] S. Feng and C. L. P. Chen, "A fuzzy restricted boltzmann machine: Novel learning algorithms based on the crisp possibilistic mean value of fuzzy numbers," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 1, pp. 117–130, Feb. 2018.

[34] N. Zhang, S. Ding, H. Liao, and W. Jia, "Multimodal correlation deep belief networks for multi-view classification," *Appl. Intell.*, vol. 49, no. 5, pp. 1925–1936, 2018.

[35] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, 1962.

[36] C.-L. Liu, W.-H. Hsaio, and Y.-C. Tu, "Time series classification with multivariate convolutional neural network," *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4788–4797, Jun. 2019.

[37] A. Rehman, S. Naz, M. I. Razak, and I. A. Hameed, "Automatic visual features for writer identification: A deep learning approach," *IEEE Access*, vol. 7, pp. 17149–17157, 2019.

[38] J. Cong, B. Wei, X. Xi, Y. Zheng, and Y. Yin, "Performance evaluation of simple linear iterative clustering algorithm on medical image processing," *Bio-Medical Mater. Eng.*, vol. 24, no. 6, pp. 3231–3238, 2014.

[39] A. Apsel, "A simple guide to low-power wireless technologies: Balancing the tradeoffs for the Internet of Things and medical applications," *IEEE Solid-State Circuits Mag.*, vol. 10, no. 4, pp. 16–23, Nov. 2018.

[40] M. Yuan, Z. Liu, and F. Wang, "Using the wide-range attention U-Net for road segmentation," *Remote Sens. Lett.*, vol. 10, no. 5, pp. 506–515, 2019.

LIANG DING was born in September 1989. He received the master's degree from the School of Physical Education, Shanxi Normal University, in 2015. He has been a Teaching Assistant with the Physical Education Department, Luliang University, since 2015. He has published more than ten academic articles. His current research interests include sports management and training.

HUIJIE REN was born in Shanxi, China, in 1990. He received the bachelor's degree from the Shanxi Normal University Sports Institute, in 2012, and the master's degree from the Postgraduate College, Shanxi Normal University. Since 2016, he attended the General College, Dankook University, South Korea. He has published a total of five articles in China and South Korea.