Retraction Notice

The Editor-in-Chief and the publisher have retracted this article, which was submitted as part of a guest-edited special section. An investigation uncovered evidence of systematic manipulation of the publication process, including compromised peer review. The Editor and publisher no longer have confidence in the results and conclusions of the article.

HC, QG, AL, and HZ either did not respond directly or could not be reached.

Deep graph convolutional network-based high-performance detection method for spectral domain gesture image stream

Hong Chen,[a,b] Qingjia Geng,[b] Aiyong Liu,[b] and Hongdong Zhao[a,c,*]

[a]Hebei University of Technology, School of Electronic and Information Engineering, Tianjin, China

[b]Hebei Normal University of Science and Technology, School of Mathematics and Information Technology, Qinhuangdao, China

[c]Science and Technology on Electro-Optical Information Security Control Laboratory, Tianjin, China

Abstract. The use of vision-based high-performance detection technology has become an innovative technical methodology. Gestures can express more actions and even emotions, which is more in line with the design idea of large-scale integrated human–computer interaction software, and then assists the development of emotion recognition and other fields. The emerging deep graph convolutional network can capture the interdependence between instances, and then infer the complete information of the image based on specific features. The two-dimensional discrete wavelet and multi-resolution analysis are used to replace the traditional Fourier transform to realize the convolution operation, which improves the accuracy of the generated graph data. This work studies the spectral clustering method of Graph Wavelet Neural Network and adds the local correlation preserving support vector machine as a classifier. This classifier has a simplified structure compared with the cascade classifier and can achieve faster and stable classification results. On the test set, the average accuracy of the algorithm is 93.40%, the recall rate is 96.27%, and the average detection time per frame is 359 ms. © *2022 SPIE and IS&T* [DOI: 10.1117/1.JEI.32.2.021603]

Keywords: graph convolutional network; high-performance detection; spectral domain images; gesture recognition; support vector machine.

Paper 220529SS received May 27, 2022; accepted for publication Aug. 22, 2022; published online Sep. 19, 2022; retracted Jun. 24, 2023.

Retracted

1 Introduction

Innovations in human–computer interaction methods should avoid the following problems, and its input information collection method should be more in line with human daily behavior habits and can provide humans with more convenience and better security. By capturing the signals emitted by human behaviors, computers can automatically interpret people's meanings and complete corresponding tasks.[1–3] If the traditional human–computer interaction model is machine-centered, then the new generation of human–computer interaction model is user-centered. It emphasizes the use of voice recognition, gesture recognition, and other technologies to enable electronic devices to actively observe the commands conveyed by users in order to provide users with more considerate service.[4–6]

The use of vision-based recognition technology for human–computer interaction will become a new method in the future. Although the current new types of human–computer interaction devices are more natural and intuitive in operation than traditional devices, they are not user-friendly and free compared with direct control of electronic devices with gestures. People are eager to be able to get rid of the shackles of the equipment, from looking for equipment to use, to equipment looking for human body to identify, i.e., from people relying on equipment to

*Address all correspondence to Hongdong Zhao, zhaohd@hebut.edu.cn

1017-9909/2022/$28.00 © 2022 SPIE and IS&T

the direction of equipment accommodating users.[7,8] Using the camera for gesture recognition can make people get rid of the limitations of the device because from the user's point of view, there is no actual contact with any device, i.e., the user does not need to deliberately search for the existence of the device, and can control the computer through their own gestures. At the same time, gestures can express more actions and even emotions, which is more in line with the design ideas of large-scale integrated human–computer interaction software in the future, and then assists the development of emotion recognition and other fields. It can be seen that after nearly half a century of development, human–computer interaction is moving toward the stage of using a camera for gesture recognition to control a computer.

If humans want to use gestures to complete human–computer interaction tasks, the first problem to be solved is how to accurately estimate the three-dimensional posture of the human hand. However, camera occlusion, image noise, large changes in viewpoint, and high complexity of hand joints are all problems hindering the development and advancement of related technologies. To improve the accuracy of hand posture estimation, it is necessary to establish the topological structure of the human hand joints so that the computer has a certain reasoning ability. The emerging graph neural network (GNN) has been developed to solve the above problems. It can capture the interdependence between instances and then infer the complete information of the image based on specific features. This enables the computer to not only have gesture recognition functions but also gesture prediction functions, which can effectively improve the accuracy of detection to a certain extent, thereby improving the efficiency of human–computer interaction. Therefore, the application of GNN and convolutional neural networks to the process of gesture recognition has great research significance.

William and Craig[9] used the image to be recognized and the background image without detection target to subtract to achieve the segmentation of gestures. This method has a good recognition effect, but it is limited by the camera's acquisition angle of view. Stergiopoulou and Papamarkos[10,11] proposed an algorithm that uses self-learning neuron networks to automatically update the shape parameters of the hand. The algorithm can accurately distinguish between the fingers and the palm and locate the center of mass. Gesture recognition under different viewing angles is very helpful. Even if the hand tilt changes, it will not affect the recognition result. However, the data storage and calculation amount of the algorithm is very large, so the detection speed needs to be further improved.

The basic principle of traditional pattern recognition is to form a "cluster" of similar samples in the pattern space, and then combine it with a classifier for classification and recognition. The main method is to express the image through mathematical statistical model, and then to recognize the image through image matching. With the rapid development of machine learning and deep learning, CNN has gradually been applied to machine vision, natural language processing, speech recognition, video analysis, and other fields.[12–14] More and more companies and researchers use deep learning to discuss and research image classification.[15]

Shah[16] proposed an iterative (IDLM) deep learning model, which can automatically layer faces and objects in images, learn recognition and representation, and first learn low-level translation through convolutional layer merging (Peak Constrained Least Squares) invariant features, and then use artificial neural networks (ANNs) to identify the nonlinear features of the input image set in the form of hierarchical iterative learning, targeting YouTube celebrities, Honda/UCSD, CMU Mobo, and people in the ETH-80 (object) data set for face and object recognition tasks. Thus, the proposed technology has been extensively evaluated.

Convolutional neural network is a new field developed on the basis of artificial neural network[17] and an important branch of machine learning.[18] With the improvement of big data deep models and hardware equipment, deep learning technology has also been developed by leaps and bounds,[19] which has promoted the advancement of computer vision, speech recognition, natural language processing, and other fields, and the success of these models is very dependent on the success of these models. However, in practical applications, due to the high cost and time consuming of collecting data sets and training models, limited data sets often lead to overfitting of the training model.[20] To verify the generalization ability of the model, the architecture based on the residual neural network can reduce the occurrence of the above problems. There are two fitting methods. The first method is to propose a cross combination abstention, i.e., reduce the size of the convolution kernel and reduce the fitting. The convolution kernel method

Retracted

of times reduces the training parameters by fitting, and the data of the cross-combination method makes the accuracy of kaggle on the cat and dog verification data set reach 95.37%, and the 30 types of engineering practice verification data set reach 90.31%; The second is based on the residual loop to propose the finetune residual neural network method to improve the model accuracy. The finetune residual neural network verifies the model accuracy on the kaggle dog and cat data sets, reaching 99.37%. The accuracy of the 30 types of engineering practice verification data set is 99.30%. At the same time, the residual accuracy of the finetune method also reaches 99.61%.

To solve the problem of human motion recognition based on the hidden Markov model, Wang et al.[21] took the lead in solving the problem of feature space dimensionality reduction. First, he described how to derive different features from the captured human actions based on the markers, and defined a total of 29 features and a total of 702 dimensions to describe human actions. Then he proposed a strategy for systematically exploring the possible subspaces of these features as meaningful low-dimensional feature vectors. He uses a data set of 353 actions to evaluate his method, which is divided into 23 different types of whole-body actions. The experimental results show that the low-dimensional feature space is sufficient to realize the recognition of high-dimensional motion, and only four dimensions are used, and an accuracy of 94.76% can be achieved on the data set, which is equivalent to the feature vector considering many features. In recent years, human motion recognition based on smartphones has received more and more attention in many fields such as mobile health, health tracking, and pervasive computing. However, changes in the orientation and position of the mobile phone can easily affect the performance of motion recognition. Most of the existing work focuses on one or two aspects of the above-mentioned problems, or trains different models for different mobile phone positions and orientations. Wang et al.[22] proposed a universal framework for human action recognition based on smartphones, which can effectively distinguish six daily actions regardless of the location and orientation of the device. He designed a set of more powerful and effective functions to solve the problem of detection performance degradation caused by different cell phone positions, cell phone orientations, and users. In the experiment, he validated the method using data sets collected by three volunteers on Android smartphones. Experimental results show that the proposed feature extraction algorithm is better than most existing algorithms. Gurbuz and Amin[23] proposed a new human motion recognition framework based on hidden Markov model, i.e., a probabilistic sequence of mixed events, which can identify unmarked motion in the video. First, the center of the moving object is used to effectively extract the motion trajectory; second, the sequence is constructed from trajectories representing different human behaviors; finally, an improved particle swarm optimization algorithm with inertial weight is introduced to identify human behavior. He evaluated the proposed method on the UCF human behavior data set, with an accuracy rate of 76.67%. The results of comparative experiments show that the recognition effect of the new method is relatively ideal.

2 Methodology

From the perspective of the development history of GNN, due to the fact that GNN cannot fully express the feature information of the edges, and the redundancy of node features leads to slow update of node parameters, the researchers proposed a stronger learning ability. Graph Convolutional Neural Network (GCN), after 10 years of updates so far, has also developed many types of models, the more typical two are the GCN in the spectral domain and the GCN in the spatial domain.

The graph convolutional neural network based on the spectral domain evolved on the basis of the graph convolutional neural network in the spatial domain, and from the aspect of the algorithm design process, the GCN in the spectral domain is a special case of the GCN in the spatial domain. Therefore, after clarifying the network structure and algorithm principle of the spatial domain GCN, it is helpful to understand the mechanism of the spectral domain GCN, and then apply it to the scene of gesture recognition. In addition to the theoretical system mentioned above, the design process of spectral domain GCN also uses the following knowledge content, which will be introduced in detail.

Retracted

2.1 *GWNN Model*

2.1.1 *Spectral method*

With the help of the convolution theorem, data can be multiplied in the spectral domain space, and then the inverse wavelet transform can be used to implement the graph convolution process. The entire feature extraction process has undergone the original space-spectrum space-original space change. The purpose of this is to map the graph data to the spectral domain space for convolution operation to complete the feature extraction of the graph data.

Because the structure of the graph does not satisfy translation invariance, it is impossible to directly define convolution in the spatial domain. Therefore, it is necessary to transform the signal into the frequency domain, implement the convolution operation in the frequency domain, and then convert it back to the spatial domain. This is the complete spectrum method.

The spatial method is to directly convolve the graph data in the spatial domain. Such a simple operation obviously has certain calculation errors, which cannot be eliminated. But the main problem faced by the space domain method is the neighborhood problem. Because the neighbor nodes of each node are inconsistent in size, it is impossible to define a neighborhood of the same size. Therefore, parameter sharing is not yet possible, but the solution is that one node is in the neighbor node. The weighted average of the above, so many subsequent methods are aimed at solving the problem of parameter sharing.

The calculation formula of the convolution theorem is as follows:

$$F(f^*g) = F(f) \cdot F(g). \tag{1}$$

Inverse wavelet transform

$$f^*g = F^{-1}(F(f) \cdot F(g)). \tag{2}$$

Based on wavelet transform and inverse transform, the convolution theorem, the graph convolution operator is obtained

$$x *_G y = U((U^T x) \odot (U^T y)). \tag{3}$$

However, the existing spectral methods still have certain limitations: e.g., they rely on the Eigen-decomposition of the Laplace transform, are computationally expensive, and are not local.

There are some excellent spectral methods that also overcome the above difficulties to a certain extent:

1. Chebyshev network (ChebyNet): The core content of the convolution operation—the parameters of the convolution kernel can be adjusted in real time, so that arbitrary changes of the convolution kernel can be realized, and the local convolution function can be realized at the same time. The network model also reduces the complexity of parameters and calculations;
2. GraphHeat: While analyzing the spectrum method, it focuses on the potential role of the filter. Since ChebyNet and the first-order graph convolutional neural network (GCN-1) are both similar to a high-pass filter, this leads to the problem of smooth prior inconsistency when the two types of networks perform semisupervised learning tasks. Based on this, what GraphHeat needs to do is a low-pass filter, and its convolution kernel uses thermal kernel functions for parameter tuning;
3. Graph convolutional neural network based on personalized pagerank (PPNP): PPNP uses two practical tools when performing functions: the feature propagation method based on pagerank and the decoupling dimension transformation method, which will first make dimensional changes to a small number of network layers, and then no longer perform parameter learning during and after training, which greatly reduces the time for the network to learn features;
4. Concise first-order graph convolutional neural network (SGC): SGC is different from the previous GCN. It has made bold changes in principle to make it have better performance in

Retracted

certain fields. The network completely abandons the nonlinear parameter exchange between different layer structures and connects the characteristics of multiple noncontact layers into one layer, which simplifies the network structure to a certain extent and reduces the complexity of network calculations.

There is also a classic method, which is proposed by approximating the parameters on the basis of the aforementioned Chebyshev network. It is mainly aimed at the simplified calculation of the first-order approximation, which will not be repeated here.

The graph convolution form popularized by the above method is as follows:

$$Z = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}X\Theta. \tag{4}$$

This is also the approximate convolution formula on the graph obtained by wavelet transform.

In this way, you can directly use the neural network for parameter training

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right). \tag{5}$$

Summarize the spectrum method:

First, ChebyNet and GCN-1 both focus on the parameterization process of the convolution kernel, while GraphHeat focuses on the transformation of the low-pass filter. Research shows that the transformation effect is good.

Second, from the perspective of spatial methods, ChebyNet always insists on the Laplacian matrix polynomial as the aggregation function, which makes the Chebyshev network always stable when extracting features.

In summary, Graph Wavelet Neural Network (GWNN), as an excellent spectral method, has stable output and outstanding advantages. All of the above spectral methods can be regarded as the link between the spectral method and the spatial method.

The following sections mainly introduce two GWNN classification methods: supervised classification and semisupervised classification. Supervised classification will introduce in detail the standard support vector machine (SVM) and its three popularization forms: minimized variance SVM (MCVSVM), locally minimized variance SVM (MCLPV_SVM), and local correlation-preserving SVM (LCPSVM). Semisupervised classification will introduce semisupervised learning methods based on graphs, and present existing graph-based Laplacian SVMs (Lap-SVM) and Laplacian least squares (Lap-RLS) models.

2.1.2 *Supervised support vector machine*

For the supervised binary classification problem, first, give the general equation of the training set

$$T = \{(x_1, y_1), \cdots, (x_l, y_l)\} \in (R^n \times Y)^l, \tag{6}$$

where $x_i \in R^n$, $y_l \in Y = \{-1, 1\}$, $i = 1, \ldots, l$. In the training set, suppose there are m_1 positive points with a label of 1, these positive points form a matrix $X_+ \in R^{m_1 \times n}$, and the output corresponding to the positive points is denoted as $Y_+ \in R^{m_1}$. There are m_2 negative points with the label -1, these negative points form a matrix $X_- \in R^{m_2 \times n}$, and the output corresponding to the negative points is denoted as $Y_- \in R^{m_2}$. The goal of this paper is to determine whether the label of any new point x is 1 or -1.

1. Support vector machine

The standard support vector classification machine, namely C-support vector classification machine, is a learning method based on statistical theory. It seeks a pair of parallel hyperplanes between the positive and negative types of points and separates the two types of training points and maximizes the interval between the hyperplanes. As shown in Fig. 1, assuming that the two types of points can be strictly linearly separated, two supporting hyperplanes can be constructed

Retracted

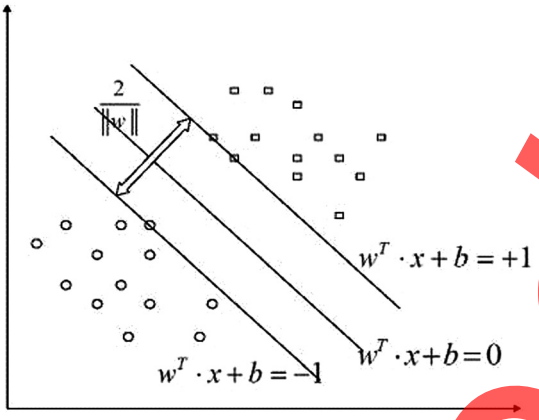

Fig. 1 Linearly separable SVM.

$$\begin{cases} w^T x + b = 1 \\ w^T x + b = -1. \end{cases} \tag{7}$$

Separate the positive and negative points respectively, as shown in Fig. 1.
Therefore, the goal of the SVM is to find the following optimal hyperplane:

$$f(x) = w^T x + b = 0. \tag{8}$$

Here, $w \in R^n$, $b \in R$.
So the optimization model of linear separable SVM is

$$\min_{wb} \frac{1}{2}\|w\|^2 \quad \text{s.t. } y_i((x_i^T \cdot w) + b) \geq 1, \ i = 1, \cdots, l. \tag{9}$$

When the positive and negative points cannot be strictly linearly separable, consider introducing slack variables. After introducing slack variables, this kind of method is called soft interval SVM. This kind of method will separate the training points as much as possible while allowing a small number of points that are not correctly divided. The model of the soft interval SVM is as follows:

$$\min_{wb} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} \xi_i \text{s.t. } y_i((x_i^T \cdot w) + b) \geq 1 - \xi_i, \ \xi_i \geq 0, \ i = 1, \cdots, l, \tag{10}$$

where $\xi = (\xi_1, \cdots \xi_l)^T$ is the slack variable, and $C > 0$ is the penalty parameter.

To solve this optimization problem, it is first required to solve its dual problem. Then find the solution of w_b, so that the final classification hyperplane is obtained.

The algorithm of the soft interval SVM is given below:

1. Given a training set $T = \{(x_1, y_1), \cdots, (x_l, y_l)\} \in (R^n \times Y)^l$, here $x_i \in R^n$, $y_i \in Y = \{-1,1\}$, $i = 1, \ldots, l$;
2. Given the value of penalty parameter C;
3. Constructing and solving dual problems;

$$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j (x_i \cdot x_j)\alpha_i \alpha_j - \sum_{j=1}^{l} \alpha_j \quad \text{S.t. } \sum_{i=1}^{l} y_i \alpha_i = 0 \quad 0 \leq \alpha_i \leq C, \ i = 1, \cdots, l$$

Get the solution of α, where α is the Lagrangian multiplier vector.

4. Calculate $w = \sum_{i=1}^{l} \alpha_i y_i x_i$, select α_i in the interval $(0, C)$ from the α components, and calculate $b = y_i - \sum_{i=1}^{l} y_i \alpha_i (x_i \cdot x_j)$;

Retracted

5. Construct the differentiation hyperplane $w^T x + b = 0$, from which the decision function $f(x) = \mathrm{sgn}(g(x))$ is obtained, where $g(x) = w^T x + b$.

2. Local correlation preserving SVM (LCPSVM)

LCPSVM is a new supervised classification problem proposed by researchers on the basis of SVM, MCVSVM, and MCLPV_SVM. In LCPSVM, a local divergence matrix P_w is proposed to maintain the structural information of the data. The researcher obtained the local divergence matrix P_w of LCPSVM through rigorous theoretical proof, including the matrix S_w proposed by MCVSVM and the matrix Z_w proposed by MCLPV_SVM. Therefore, LCPSVM is theoretically superior to MCVSVM and MCLPV_SVM. The optimization problems of LCPSVM are as follows:

$$\min_{w,b} \frac{1}{2} w^T P_w w + C \sum_{i=1}^{l} \xi_i \text{s.t. } y_i((x_i^T \cdot w) + b) \geq 1 - \xi_i \quad \xi_i \geq 0, \ i = 1, \cdots, l$$

The matrix $P_w = P_+ + P_-$. Here, P_+ and P_- are the local divergence matrices of the data sets X_+ and X_-, respectively

$$P_+ = \sum_{i=1}^{X_*} \left(x_i - \sum_{j \in ne(x_i)} \frac{v_{ij}}{\sum_{j \in ne(x_i)} v_{ij}} x_j \right) \left(x_i - \sum_{j \in ne(x_i)} \frac{v_{ij}}{\sum_{j \in ne(x_i)} v_{ij}} x_j \right)^T$$

$$P_- = \sum_{i=1}^{|X|} \left(x_i - \sum_{j \in ne(x_i)} \frac{v_{ij}}{\sum_{J \in ne(x_i)} v_{ij}} x_j \right) \left(x_i - \sum_{j \in ne(x_i)} \frac{v_{ij}}{\sum_{j \in ne(x_i)} v_{ij}} x_j \right)^T,$$

where v_{ij} is the degree of similarity between two data points, and it is defined as follows:

$$v_{ij} = \begin{cases} \exp(-\|x_i - x_j\|_2^2 / 2\sigma^2) & x_i \in \mathrm{ne}(x_j) \text{ or } x_j \in \mathrm{ne}(x_i) \\ 0 & \text{otherwise} \end{cases}. \tag{11}$$

For each data point x_i, $\mathrm{ne}(x_i)$ represents the k nearest neighbors of the point x_i with the same label, $\sum_{j \in ne(x_i)} \frac{v_{ij}}{\sum_{j \in ne(x_i)} v_{ij}} x_j$ represents the local weighted mean. To solve the optimization problem, the Lagrangian function is introduced

$$L(w, b, \alpha, \beta, \xi) = \frac{1}{2} w^T S_w w + C \sum_{i=1}^{l} \xi_i - \sum_{i=1}^{l} \alpha_i [y_i (w^T x_i + b) - 1 + \xi_i] - \sum_{i=1}^{l} \beta_i \xi_i.$$

The optimality conditions are

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \Rightarrow w = P_w^{-1} \sum_{i=1}^{l} \alpha_i y_i x_i \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \Rightarrow C - \alpha_i - \beta_i = 0 \end{cases}.$$

The dual problem can be obtained through the above optimality conditions

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j (x_i^T P_w^{-1} x_j) \alpha_i \alpha_j - \sum_{j=1}^{l} \alpha_j \text{s.t. } \sum_{i=1}^{l} y_i \alpha_i = 0 \quad 0 \leq \alpha_i \leq C, \ i = 1, \cdots, l.$$

Retracted

The decision function can be obtained by solving the above dual problem

$$f(x) = \operatorname{sgn}\left(P_w^{-1}\sum_{i=1}^{l}\alpha_i y_i\langle x_i \cdot x\rangle + b\right). \tag{12}$$

To improve the classification effect of the LCPSVM model, it is also necessary to consider the distribution of the data in the penalty item. The improved model is as follows:

$$\min_{w,b}\frac{1}{2}w^T P_w w + C\sum_{i=1}^{l}\left(\xi_i + \sum_{j\in ne(x_i)} v_{ij}\xi_j\right)\text{s.t. } y_i((x_i^T \cdot w) + b) \geq 1 - \xi_i,$$
$$\xi_i \geq 0,\ i = 1, \cdots, l.$$

For the optimization problem, the dual problem is also solved, and the dual problem is as follows:

$$\min_{\alpha}\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j (x_i^T P_w^{-1} x_j)\alpha_i\alpha_j - \sum_{j=1}^{l}\alpha_j \quad \text{s.t. } \sum_{i=1}^{l} y_i\alpha_i = 00 \leq \alpha_i \leq C_i,\ i = 1, \cdots, l.$$

Among them, $C_i = C(1 + \sum_{j\in ne(x_i)} v_{ij})$, for a new point, use the decision function to determine its label $f(x) = \operatorname{sgn}(P_w^{-1}\sum_{i=1}^{l}\alpha_i y_i\langle x_i \cdot x\rangle + b)$.

2.1.3 *Semisupervised classification*

With the development of the economy and the progress of society, the penetration of computer technology in human's daily life has become more and more in-depth, and the ability of humans to collect and store data has been greatly improved. Of course, the speed of human production of information has also been greatly accelerated. For these unprecedented massive data, mining valuable information has become a hot topic.

In most cases, it is very difficult to exhaust the categories and labels of objects, and sometimes requires a lot of manpower and material input. To solve this problem, people have proposed a semi-supervised method, which can simultaneously use labeled points and unknown labeled points to achieve the expected classification effect.

1. A semisupervised classification framework based on graphs

For the semisupervised classification problem of n-dimensional data space, we give a labeled training set $X_L = \{(x_1, y_1), \cdots, (x_l, y_l)\}$ and an unlabeled training set $X_U = \{x_{l+1}, x_{l+2}, \cdots, x_{l+u}\}$, where $x_i \in R^n$, $1 \leq i \leq l+u$, $y_i \in \{-1, +1\}$, $1 \leq i \leq l$. The goal of semisupervised classification problem learning is to obtain a classifier to predict the label of unknown label points or the label of new points.

To solve the semisupervised classification problem, in 2006, a method of manifold rules was proposed: assuming that the distribution of data has a Riemannian manifold structure, the points with labels obey the P distribution, and the points without labels obey the edge distribution PX of P. The labels of two points close to each other on the PX distribution should be the same or similar.

The solution formula for manifold regularization term (MR term) is as follows:

$$\|f\|_{\mathcal{M}}^2 = \sum_{i,j}^{l+u} v_{ij}(f(x_i) - f(x_j))^2 = f^T L f. \tag{13}$$

Here $L = -D_V$ is the Laplacian graph matrix, D is a diagonal matrix whose diagonal elements $D_{ii} = \sum_{j=1}^{l+u} v_{ij}$, and the weight matrix V is passed k obtained by the nearest neighbor method

Retracted

$$v_{ij} = \begin{cases} \exp(-\|x_i - x_j\|_2^2/2\sigma^2) & x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i) \\ 0 & \text{otherwise} \end{cases}. \tag{14}$$

Through the kernel function and regenerating the Hilbert kernel space, the semisupervised learning framework can be expressed by the following equation:

$$f = \arg\min_{f \in \mathcal{H}_k} \sum_{i=1}^{l} V(f(x_i), y_i, f) + \gamma_{\Re}\|f\|_{\mathcal{H}}^2 + \gamma_M \|f\|_{\mathcal{M}}^2 \tag{15}$$

Here, f is the decision function, V is the loss function, $\gamma_{\Re}$ and γ_M are the weight parameters, $\|f\|^2$ represents the complexity of the regenerated Hilbert kernel space f, and $\|f\|_{\mathcal{M}}^2$ represents the complexity of f on the Riemannian manifold controlled by the manifold regular term.

2. Laplace SVM (Lap-SVM)

Based on the theory of the above-semisupervised learning framework, Lap-SVM came out. According to the representation theory and kernel function technology, the goal of Lap-SVM is to find the classification hyperplane

$$f(x) = \sum_{i=1}^{l+u} \alpha_i K(x, x_i). \tag{16}$$

Here, $\alpha_i \in R$. By introducing the hinge loss function

$$V(x_i, y_i, f) = \max\{0, 1 - y_i f(x_i)\}.$$

The classification hyperplane can be obtained by the following quadratic planning:

$$\min_{f \in \mathcal{H}_k} \frac{1}{l}\sum_{i=1}^{l} \xi_i + \gamma_{\mathcal{H}}\|f\|_{\mathcal{H}}^2 + \gamma_{\mathcal{M}}\|f\|_{\mathcal{M}}^2 \quad \text{s.t. } y_i f(x_i) \geq 1 - \xi_i \quad \xi_i \geq 0,\ i = 1, \cdots, l.$$

Among them, $\xi_i (i = 1, \cdots l)$ is based on labeled slack variables. Substituting $f(x) = \sum_{i=1}^{l+u} \alpha_i K(x, x_i)$ into the above equation, we can get the following optimization model:

$$\min_{f \in \mathcal{H}_k} \frac{1}{l}\sum_{i=1}^{l} \xi_i + \gamma_{\mathcal{H}}\alpha^T K\alpha + \gamma_{\mathcal{M}}\alpha^T \text{KLK}\alpha$$

$$\text{s.t. } y_i\left(\sum_{j=1}^{l+u} \alpha_j K(x_i, x_j)\right) \geq 1 - \xi_i \quad \xi_i \geq 0,\ i = 1, \cdots, l.$$

To solve the above optimization problem, we introduce βi and ζi as multiplier vectors to construct the Lagrangian function

$$L(\alpha, \xi, b, \beta, \zeta) = \frac{1}{l}\sum_{i=1}^{l} \xi_i + \gamma_{Mt}\alpha^T K\alpha + \gamma_{\mathcal{M}}\alpha^T \text{KLK}\alpha$$

$$- \sum_{i=1}^{l} \beta_i \left(y_i \left(\sum_{j=1}^{l+u} \alpha_j K(x_i, x_j) - 1 + \xi_i \right) \right) - \sum_{i=1}^{l} \zeta_i \xi_i.$$

Retracted

The optimality conditions are

$$\begin{cases} \frac{\partial L}{\partial \alpha} = 0 \Rightarrow \alpha = (2\gamma_{\mathcal{H}} I + 2\gamma_{\mathcal{M}} LK)^{-1} J^T Y \beta \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{l} \beta_i y_i = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \Rightarrow \frac{1}{l} - \zeta_i - \beta_i = 0 \end{cases}.$$

A dual problem can be obtained through the above optimality conditions

$$\max_{\beta} \sum_{j=1}^{l} \beta_j - \frac{1}{2} \beta^T Q \beta \quad \text{s.t.} \sum_{i=1}^{l} y_i \beta_i = 0 \quad 0 \le \beta_i \le \frac{1}{l}, \ i = 1, \cdots, l.$$

Here, $Q = \mathrm{YJK}(2\gamma_{\mathcal{H}} I + 2\gamma_M \mathrm{LK})^{-1} J^T K$, $J = [1, 0]$ is the matrix of $l \times (l+u)$, and K is the identity matrix of $l \times l$, $Y = \mathrm{diag}(y_1, y_2, \cdots, y_l)$.

The final decision function can be obtained by solving the above dual problem.

3. Laplacian regularized least squares (Lap-RLS)

On the basis of the semisupervised learning framework proposed above, if the loss function is a quadratic loss, i.e.

$$V(x_i, y_i, f) = \min (y_i - f(x_i))^2. \tag{17}$$

Then the optimization problem becomes Lap-RLS, and the optimization problem becomes

$$\min_{f \in \mathcal{H}_k} \frac{1}{l} \sum_{i=1}^{l} (y_i - f(x_i))^2 + \gamma_{\mathcal{H}} \|f\|_{\mathcal{H}}^2 + \gamma_{\mathcal{M}} \|f\|_{\mathcal{M}}^2.$$

Substituting $f(x) = \sum_{i=1}^{l+u} \alpha_i K(x, x_i)$ into the above equation can be obtained

$$\min_{f \in \mathcal{H}_k} \frac{1}{l} \sum_{i=1}^{l} (y_i - f(x_i))^2 + \gamma_{\mathcal{H}} \alpha^T K \alpha + \gamma_{\mathcal{M}} \alpha^T \mathrm{KLK} \alpha$$

Because the above equation is an unconstrained optimization problem. Therefore, the solution of α can be obtained directly from the original problem. The following is the partial derivative of α:

$$\frac{1}{l} (Y - \mathrm{JK}\alpha)^T (-\mathrm{JK}) + (\gamma_{\mathcal{H}} K + \gamma_{\mathcal{M}} \mathrm{KLK}) \alpha = 0. \tag{18}$$

Among them, $\alpha = (\mathrm{JK} + \gamma_{\mathcal{H}} lI + \gamma_M \mathrm{LK})^{-1} Y$, $J = \mathrm{diag}(1, \cdots, 1, 0, \cdots, 0)$ is $(l+u) \times (l+u)$ diagonal matrix, the first l diagonal elements are 1, and the rest are 0, $Y = [y_1, \cdots, y_l, 0, \cdots, 0]$ is a $(1+u)$-dimensional vector. The final decision function can be obtained by solving the above problems.

2.1.4 *Basic model structure*

GWNN is a GCN network model based on the spectral domain. It uses graph signal theory and uses two-dimensional discrete wavelet transform as a bridge to successfully migrate CNN to the graph. Starting from the spectral domain, this method uses the Laplacian matrix to process data such as graphs with non-Euclidean structure. The whole mathematical derivation process is very clever, and it is worthy of in-depth study and research.

The model structure of GWNN is shown in Fig. 2.

Compared with graph-SAGE, GWNN has three improvements. First, the two-dimensional discrete wavelet transform is used instead of the traditional Fourier transform, which makes the convolution operation more suitable for feature extraction of graph data; second, the spectral clustering method is added, so that the features extracted by the model have better aggregation

Retracted

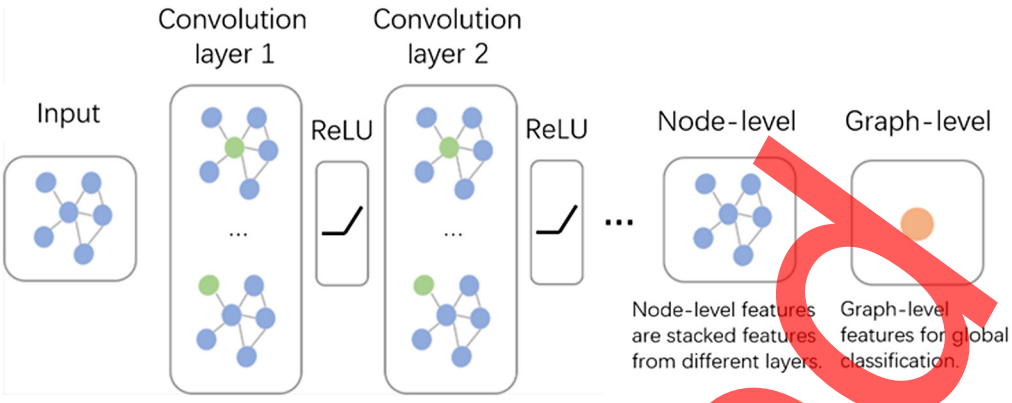

Fig. 2 GWNN network model.

characteristics, and lay a good foundation for the subsequent classification process; third, LCPSVM is used as a classifier to improve the perception ability of the classifier for features. It can not only do classification and recognition, but also automatically analyze the distribution of feature data, so that the classifier can achieve better and more stable classification effect.

The graph convolution operation consists of four parts: (1) input graph signal; (2) feature extraction; (3) graph classification; and (4) output signal, output the label of each node, to achieve the entire classification task.

Considering that the dimension of the input feature is not a single dimension, the activation function in the network uses a nonlinear activation function to enhance the nonlinear fitting ability of the network. In this way, the new node features after aggregation are output, and the dimensionality reduction operation is completed.

Although this model implements convolution operations on the graph and uses neighbor node characteristics to characterize the central node characteristics, it has achieved a qualitative leap in GCN and laid the cornerstone for the development of spectral-domain GCN. But it also has shortcomings. First, when updating the central node feature, the node feature of the upper layer used is not necessarily the neighbor node of the central node; second, the GCN of the spectral domain is based on the Laplacian matrix. The adjacency matrix with neighbor information is used, so it does not have locality, or the locality is not strong enough; finally, the computational complexity of this method is $O(n^2)$, and the calculation requires high hardware cost and time-consuming.

At present, the third-generation GCN considers applying the spectral domain graph convolutional neural network in actual semisupervised scenarios. In these scenarios, the labels of some nodes are unknown, and the task of classifying nodes needs to be completed. To make the graph convolutional neural network have a good classification effect and better local connection characteristics, Chebyshev polynomials of second order are gradually applied in the feature extractor of GCN.

Another improvement point of GWNN is to further improve the computational efficiency of the network. It is not difficult to see that GCN based on spectral domain has certain limitations. First of all, its network structure cannot be too large or deep, and GCN in the spectral domain is not suitable for modeling on large-scale graph data. Since the Laplacian matrix of all nodes needs to be used in the forward propagation process, the graph structure data faced in the actual industry is basically tens of millions of nodes. The Laplacian matrix input of the node is very unrealistic, and the calculation cost is very high or even impossible to run at all. And in related research, the researchers only used two-layer neural network. When the number of layers is increased, the performance of GCN based on the spectral domain is not significantly improved, and the two-layer neural network cannot take advantage of deep learning. Therefore, the inability to construct more hidden layers is also one of the limitations of this type of method. Whether to construct a larger and deeper GNN is a thorny issue commonly faced by current scholars. Finally, this type of method cannot process the data of the directed graph structure, because when applying the spectrum theory derivation, the symmetry of the Laplacian matrix must be ensured to perform the spectrum decomposition and complete the entire mathematical derivation process.

Retracted

Fig. 3 Four gestures.

2.2 *Data Acquisition*

A series of influencing factors such as skin color, angle, light, and background are taken into consideration when collecting gesture images.

This article chooses to collect four international general gestures, including OK, PEACE, PUNCH, and STOP. It is assumed that the system does not detect any gestures, then it will output NOTHING. Examples of the four gestures are as follows in Fig. 3:

In addition, this article also selects one image type as a negative sample, namely arm occlusion, as shown in Fig. 4.

This article uses a 1080P HD nondistortion USB camera to collect image information, the image size is 480×640, the output format is RGB three-channel color image, and a protection device is designed for it. The overall structure is shown in Fig. 5. This type of camera can automatically fill light to shoot in the case of insufficient light.

Fig. 4 Comparison picture.

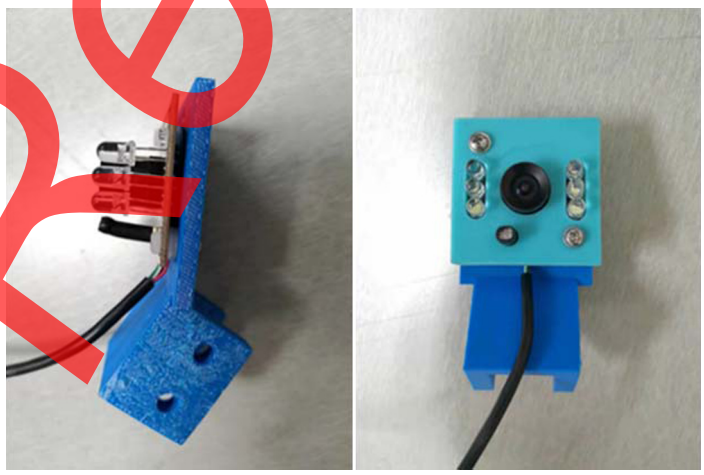

Fig. 5 Image acquisition device.

Retracted

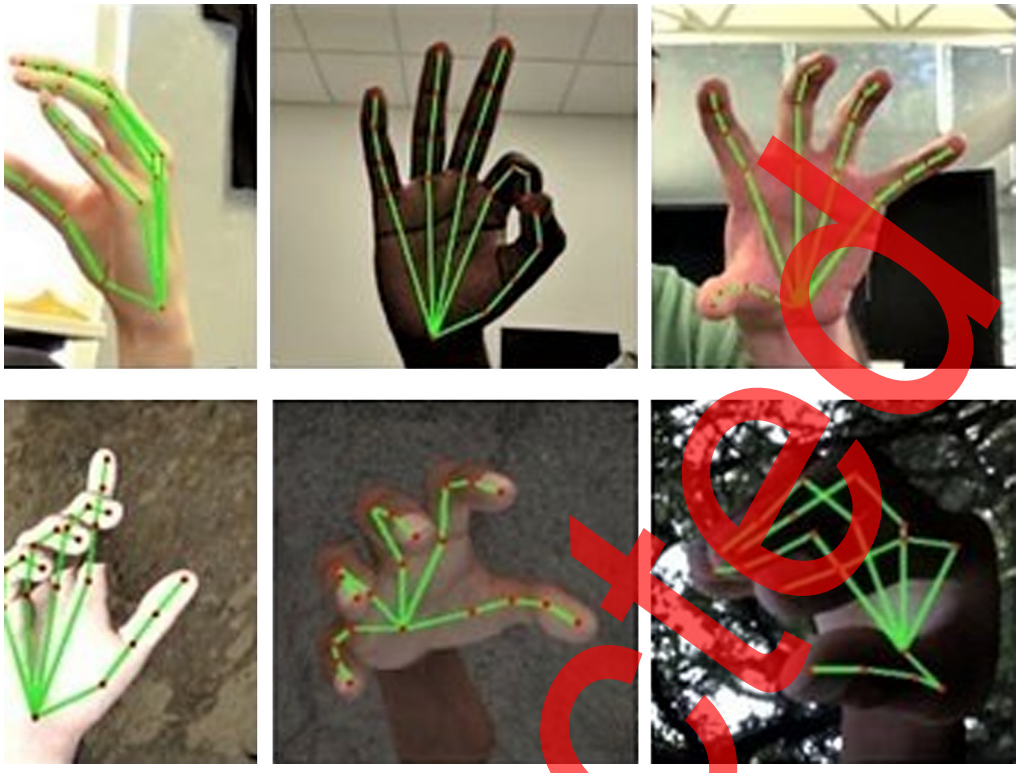

Fig. 6 Hand topology.

2.3 *Hand Topology Analysis*

Since GWNN can only identify the graph data input into the network, it is still necessary to analyze the topological structure of the human hand and establish a homogeneous graph model. The structure analysis of the human hand is shown in Fig. 6.

In Fig. 6, each joint of the hand is regarded as a node, and the stored features of the node are bending and unbending. This is the standard hand topology. Since the above analysis and processing of human hand joints will make the algorithm complex and processing time slow, we simplify it to the following structure.

In Fig. 7, each finger of the human hand can be seen as directly connected to the palm of the hand, i.e., straight line AE. When a finger is bent, the finger can be divided into two segments, namely line segment AC and line segment DE. The bending of the finger represents the bending between AC and DE. GWNN can identify the line segments AC and DE, which is equivalent to identifying the area corresponding to the hand, and then further determine whether there is a bend between AC and DE, and input the node and edge information into the homogenous graph model.

2.4 *Data Set Production*

The data set used for training GWNN is the graph data corresponding to the image data. The graph data set is composed of hand homogenous maps extracted from hand images. The hand homogeneity map contains all the topological structure information of the hand in the image, and it is saved in the form of a matrix.

Therefore, the production of the data set includes four steps:

1. Hand image collection
2. Image preprocessing
3. Establish a homogenous map model of the hand
4. Uniform data size and format

Retracted

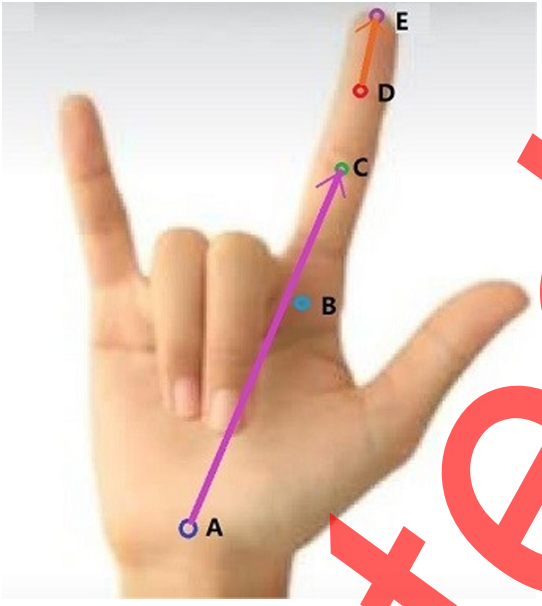

Fig. 7 Hand structure.

Of course, the above process is the production step of the positive sample (including gestures) data set. For the production of the negative sample (not including gestures) data set, the third step is not required.

The positive sample includes 4015 hand pictures, of which 2810 are in the training set, 803 are in the verification set, and 402 are in the test set. The ratio of training set, verification set, and test set is 7:2:1. The negative sample includes 1005 nonhuman images.

2.5 *Model Training*

The training content of this article mainly includes two core parts. The first part is the training of the graph convolutional neural network, which is the GWNN model. Training it can allow the entire detection system to correctly identify and extract the homogenous graph model information of the hand. The second part is the training of LCPSVM, which can allow the classifier to correctly determine the meaning of gestures expressed by the homogenous graph model, and then realize static gesture recognition.

3 Results

The algorithm usability evaluation standard in this article uses the accuracy values AC and AUC of the algorithm on the verification set.

The entire detection system has undergone 14 rounds of iterative training, and the training accuracy and loss are shown in Figs. 8 and 9.

As you can see in Fig. 9, after 14 parameter iterations, the accuracy of the model on the training set and the validation set has improved, exceeding 99%; at the same time, the loss of the model on the training set and validation set is also both are <0.02. This result is ideal.

After training, the specific performance of the algorithm on the validation set is shown in Table 1.

The evaluation standard of algorithm detection effect in this article adopts the average accuracy value AP and recall rate RC of the algorithm on the test set.

The specific performance of the algorithm on the test set is shown in Table 2.

The performance of the model on the test set is relatively good. The loss of accuracy may be due to the lack of matching features. The recall rate proves the correctness of the research ideas in this article. The near real-time detection effect can also reduce the false detection rate to a certain extent.

Retracted

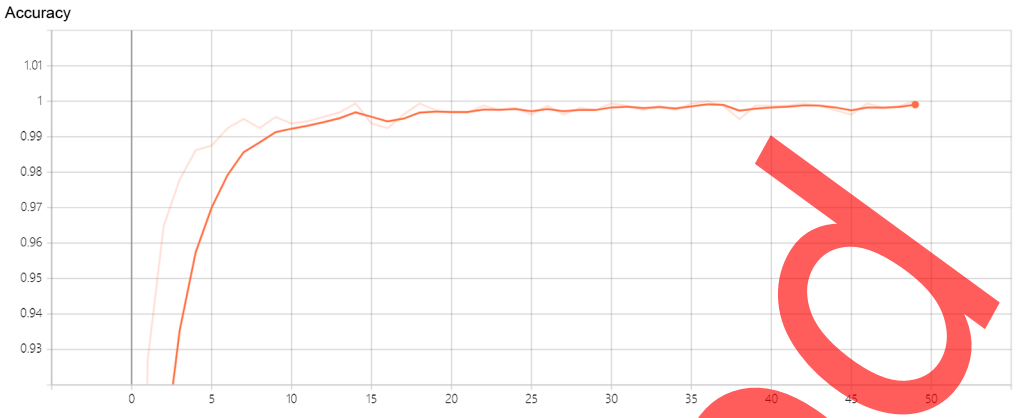

Fig. 8 Accuracy change of model training.

Fig. 9 Loss change of model training.

Table 1 Algorithm usability evaluation.

	AC (%)	AUC	Image processing time per frame/ms
Calculated value	99.79	0.673	353

Table 2 Evaluation of algorithm detection effect.

	AP (%)	RC (%)	Image processing time per frame/ms
Calculated value	93.4	96.27	359

4 Conclusions

After comprehensively considering the clustering effect and time consumption, the spectral clustering method is used as the feature clustering method of GWNN, which is suitable for complex backgrounds hand feature clustering under the following; wavelet transform includes continuous wavelet transform and discrete wavelet transform, which helps to improve the accuracy of convolution to extract features, so this type of wavelet transform is used to replace Fourier transform.

Retracted

This article elaborates on the relevant knowledge of GWNN network, including spectral method, classification method, and model structure. The classification method includes two types, namely supervised SVM and semisupervised classification method. Since the problem optimization effect of LCPSVM is the best, so LCPSVM is used as the feature matcher of GWNN for classification.

This article introduces the production process of the data set and the GWNN training and recognition detection process. By calculating the performance of the model on the verification set, it is proved that the model is available; by analyzing the performance of the model on the test set, it shows that the detection effect of the model is good, and the average detection accuracy is 93.4%.

Acknowledgments

The study was supported by "Tianjin Science and Technology Project (Grant No. 21YDTPJC00050)" and "Science and Technology on Electro-optical Information Security Control Laboratory Project (Grant No. 2021JCJQLB055008)."

References

1. R. Pech et al., "Link prediction via matrix completion," *Europhys. Lett.* **117**(3), 38002 (2016).
2. C. Yang et al., "Network representation learning with rich text information," in *Twenty-Fourth Int. Joint Conf. Artif. Intell.* (2015).
3. L. Tang and H. Liu, "Relational learning via latent social dimensions," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discov. and Data Mining*, ACM (2009).
4. W. J. Zheng et al., "M-GWNN: multi-granularity graph wavelet neural networks for semi-supervised node classification," *Neurocomputing* **453**, 524–5351 (2021).
5. S. T. Yong et al., "Convolutional neural network with spatial pyramid pooling for hand gesture recognition," *Neural Comput. Appl.* **33**, 5339–42 (2020).
6. Y. S. Tan, K. M. Lim, and C. P. Lee, "Hand gesture recognition via enhanced densely connected convolutional neural network," *Expert Syst. Appl.* **175**(90), 114797 (2021).
7. J. X. Wang, T. T. Liu, and X. Wang, "Human hand gesture recognition with convolutional neural networks for K-12 double-teachers instruction mode classroom," *Infrared Phys. Technol.* **111**, 103464 (2020).
8. Y. N. Xing, G. Di Caterina, and J. Soraghan, "A new spiking convolutional recurrent neural network (SCRNN) with applications to event-based hand gesture recognition," *Front. Neurosci.* **14**, 590164 (2020).
9. T. F. William and D. W. Craig, "Television control by hand gesture," in *1995 IEEE Int. Workshop Autom. Face and Gesture Recognit.* (2015).
10. E. Stergiopoulou and N. Papamarkos, "A new technique for hand gesture recognition," in *Proc. 2006 Int. Conf. Image Process., Atlanta (USA)*, pp. 2600–2657 (2016).
11. E. Stergiopoulou and N. Papamarkos, "Hand gesture recognition via a new self-organized neural network," *Lect. Notes Comput. Sci.* **3773**, 891–904 (2015).
12. E. Cambria and B. White, "Jumping NLP curves: a review of natural language processing research," *Comput. Intell. Mag. IEEE* **9**(2), 48–57 (2014).
13. P. Dighe, A. Asaei, and H. Bourlard, "Sparse modeling of neural network posterior probabilities for exemplar-based speech recognition," *Speech Commun.* **76**, 230–244 (2015).
14. R. Errattahi, A. El Hannani, and H. Ouahmane, "Automatic speech recognition errors detection and correction: a review," *Procedia Comput. Sci.* **128**, 32–37 (2018).
15. T. H. Hu et al., "Advantages and application prospects of deep learning in image recognition and bone age assessment," *J. Forensic Med.* **6**, 629–634 (2017).
16. S. Shah, "Iterative deep learning for image set based face and object recognition," *Neurocomputing* **174**, 866–874 (2015).

Retracted

17. N. Shah, P. Chaudhari, and K. Varghese, "Runtime programmable and memory bandwidth optimized FPGA-based coprocessor for deep convolutional neural network," *IEEE Trans. Neural Networks Learn. Syst.* **99**, 1–13 (2018).
18. J. T. Fei et al., "Multi-input convolutional neural network based on gradient," *Opto-Electron. Eng.* **42**(3), 33–38 (2015).
19. W. Wang, G. Chen, and H. B. Chen, "Deep learning at scale and at ease," *ACM Trans. Multimedia Comput. Commun. Appl.* **12**, 69 (2016).
20. R. Liu and D. F. Gillies, "Overfitting in linear feature extraction for classification of high-dimensional image data," *Pattern Recognit.* **53**(204), 73–86 (2015).
21. P. C. Wang, W. Q. Li, and P. Ogunbona, "RGB-D-based human motion recognition with deep learning: a survey," *Comput. Vision Image Understanding* **171**(1), 118–139 (2018).
22. P. Wang, H. Y. Liu, and L. H. Wang, "Deep learning-based human motion recognition for predictive context-aware human-robot collaboration," *CIRP Ann.* **67**(1), 17–20 (2018).
23. S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: promising applications for indoor monitoring," *IEEE Signal Process. Mag.* **36**(4), 16–28 (2019).

Hong Chen received her master's degree from Changchun University of Science and Technology, China. She currently works at the School of Mathematics and Information Technology, Hebei Normal University of Science and Technology. Her research interests include image information processing and artificial intelligence.

Biographies of the other authors are not available.

Retracted