

Received 18 September 2023, accepted 19 October 2023, date of publication 2 November 2023, date of current version 22 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3329575

## RESEARCH ARTICLE

# Design of an Intelligent Laboratory Facial Recognition System Based on Expression Keypoint Extraction

YING ZHOU<sup>1</sup>, YOUWANG LIANG<sup>2</sup>, AND PENG PENG TAN<sup>1,2</sup> 

<sup>1</sup>Infrastructure Construction Department, Nanjing Forestry University, Nanjing 210037, China

<sup>2</sup>Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China

Corresponding author: Pengpeng Tan (tpp041029@njfu.edu.cn)


**ABSTRACT** The number of clever facial recognition systems has been growing as artificial intelligence and robotics have advanced. However, due to the limited collection of biometric features by intelligent facial recognition systems compared to authentication methods such as iris and fingerprint, there are errors in the recognition process, low recognition accuracy, and low operational efficiency. To improve laboratory security and the efficiency of intelligent facial recognition systems, a study was conducted to extract facial feature information through facial expression key points, and a spatiotemporal graph convolutional network fused with attention mechanism was used to organize and match feature data. Finally, an improved facial recognition system was designed. A simulation experiment was conducted, and the results showed that the intelligent laboratory facial recognition system constructed by the research had a facial recognition accuracy of 89% in different datasets. Compared with the Line Hausdorff distance facial recognition method, the facial feature data was more concentrated, proving that the facial recognition system designed by the research has high recognition accuracy, wide application range, and strong application value. It helps to improve the optimization of facial recognition systems in the laboratory and improve the operational efficiency of the laboratory.

**INDEX TERMS** Facial recognition, face keypoints, GCNNs, self-attention mechanism, feature fusion.

## I. INTRODUCTION

Artificial intelligence technology has steadily grown to play a significant role in the advancement and development of human society today. As the level of informatization of human society continues to improve, the use of biometrics for identity verification has become the mainstream of research. Among them, iris recognition, fingerprint recognition and facial recognition have been applied to varying degrees in various fields of identity verification. However, although iris recognition is fast, the cost of recognition equipment is high, and it is easy to cause image distortion and affect the correct recognition rate. Fingerprint recognition, although less expensive, is less convenient and contact-based. facial recognition systems not only possess the speed of red

film recognition but are also cost-effective and contactless. Therefore, the usage of facial recognition for identity verification has become the central aspect of artificial intelligence and machine learning fields. Intelligent systems related to this have been implemented in building access control systems to enhance people's work and life convenience [1], [2], [3]. The feature capture of important facial expression cues can authenticate an object's identification by allowing it to be distinguished from that of other people. Facial recognition technology and access control systems for laboratories will increase the operational effectiveness of the lab while also ensuring its security and preventing the leakage of private data. However, the current generation of intelligent facial recognition systems lacks ideal accuracy and occasionally even fails to identify the user [4], [5]. To collect and match face feature data in this context, the study uses a spatio-temporal graph convolutional network that

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

incorporates an attention mechanism. The study then designs an improved facial recognition system, which is anticipated to increase the recognition accuracy of the intelligent face system and aid in increasing the effectiveness of lab operations.

Graph convolutional networks can help remove noise from graph signals and solve the feature information of aggregated neighboring points. They have wide applications in research fields such as video generation, mechanical state diagnosis, sample classification, and can also help solve the construction and performance prediction problems of facial recognition systems. P. Li. and his team proposed a meta-learning graph convolutional network model for video caption generation to address the issue that long and short-term memory encoders bridge the video semantic gap without taking the temporal relationship of video frames into account. This model treats video frames as nodes in a graph and then builds a temporal model of video nodes to capture semantic information through a multi-granularity lexical attention mechanism, ultimately improving caption quality. Dong [6]. and his team proposed a graph convolutional network model incorporating deep learning mechanisms to build an end-to-end keying framework and generate pixel features using deep learning mechanisms, and then refine the pixel features in the convolutional network structure to predict the results of alpha masking. The study's findings supported the model's predictions. The outcomes confirmed the model's accuracy in image processing, which is extremely beneficial in boosting the visual impression [7]. A transformer fault diagnosis model based on a graph convolutional network was created by Liao et al. to increase the accuracy of transformer fault diagnosis. The model uses an adjacency matrix to represent the direct similarity measure between the current sample and the unknown sample and a graph convolutional layer as a classifier to distinguish the non-linear relationship between dissolved gas and fault type to complete the model's training process. The results of the experiments demonstrate that the suggested diagnostic model is capable of achieving the diagnostic accuracy standards for transformers [8]. A system-level prediction framework based on geometric deep learning and graphical convolutional networks was proposed by Palazuelos et al. to analyse system-level modules for health diagnosis and prediction at different levels of reliability and to embed system learning features through graphical convolutional networks to achieve system degradation in order to address the issue that sensing technologies do not consider component interactions when dealing with complex systems [9]. Peng and his team members proposed a dynamic graph convolutional network model to address the issue of outlier influence in graph learning methods. This model combines graph learning and semi-supervised classification in a single framework, uses robust statistics to minimise the impact of outliers, and gives more weight to significant samples to maximise the benefits of graph learning [10].

Effective facial recognition systems have become a popular topic of study in the field of artificial intelligence technology

due to technological advancements. For the purpose of processing complex spectral data, Niu and the members of his team proposed a convolutional neural network framework for hyperspectral facial recognition with a channel attention mechanism, introducing second-order efficient channel attention as a variant to capture more details of face features. According to experimental findings, the facial recognition framework provides a diverse range of applications across multiple devices and high levels of accuracy and efficiency. Furthermore, it has the potential to save both system storage space and computing costs [11]. A deep learning model-based facial recognition system employing neural networks was suggested by Bajrami et al. to identify facial features through digital photography and analyse a large number of digital photographs. As per the experimental data, the facial recognition system can identify individuals by detecting their faces even in challenging situations. The results indicate that the system can also function at a lower cost [12]. To gather features of likeness and understand facial judgments despite inter-age differences, scholars such as Chen introduced an identity-based viewpoint triadic loss model for cross-age facial recognition. Cross-age facial recognition is ultimately attained by conducting triadic group selection at the identification level and utilizing a forward mining technique [13]. Using an attention development network model for multi-scale facial recognition in complex environments, proposed by Guo et al., the system's recognition area is focused on the face through a top-down attention mechanism, while a synaptic maintenance mechanism is employed to suppress background pixels, hence improving the accuracy of facial recognition. As the facial recognition model exhibits a 13% higher recognition accuracy compared to the bionic neural network, and requires less training time, it can be concluded that it holds a superior application value, based on the experimental results [14]. Wang and colleagues have proposed a linear classifier based on binary labels for facial recognition. The approach combines original and virtual samples and incorporates structure and low-rank constraints to preserve the intrinsic properties of the samples in order to mitigate the effect of environmental noise. The method also employs snake rolling to improve the performance of the linear classifier, enhancing facial recognition as a result [15].

Due to the fact that facial recognition mainly utilizes image recognition technology, the ability to extract and recognize image features also affects the performance of facial recognition. Lu et al. used two-dimensional images as input, encoded the input image data using 2DNPP, and then learned features from 2DNPP to complete image classification, which can enhance the algorithm's feature extraction ability [16]. Lu et al. applied the NMF method to image classification and introduced a low-rank structure to construct a non-negative factorization method. This method can learn the global data of the image and ensure the independence of the data, which improves the anti-interference ability of the image data [17].

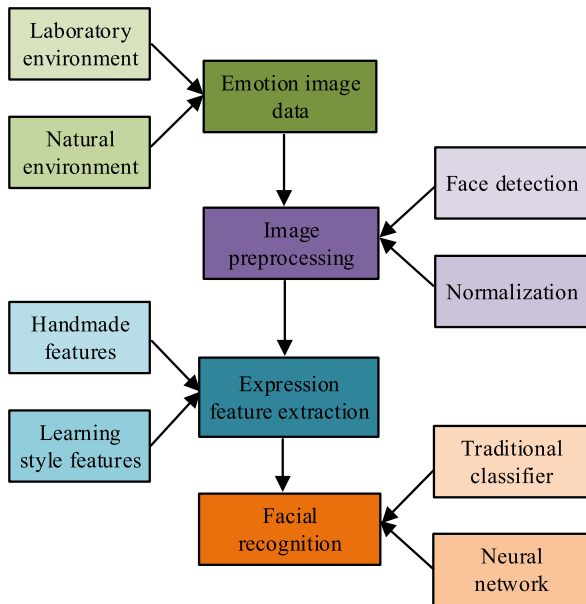


FIGURE 1. Overall process of facial recognition.

In summary, although many scholars have made improvements to facial recognition systems from multiple perspectives and aspects, ordinary facial recognition systems have lower accuracy due to factors such as lighting and facial expression changes in the real facial recognition environment. At the same time, facial capture technology is mainly used for dynamic facial recognition, resulting in high recognition costs. In order to improve the facial recognition system's facial recognition accuracy in various environments and lower the system maintenance cost, the study adopts a spatio-temporal graph convolutional network incorporating attention mechanism for feature data collation and matching.

## II. INTELLIGENT LABORATORY FACIAL RECOGNITION SYSTEM BUILD

### A. MODEL DESIGN OF GCNNs BASED ON EXPRESSION KEYPOINT EXTRACTION

The extraction of key points that represent face features is necessary for the systematic intelligent recognition of faces. The system is then able to recognise face information by locating the distribution of key points of facial emotions [18], [19]. The method of identifying a given facial expression image sample is known as facial recognition. The face picture sample will be subjected to the following procedure: face detection using expression keypoints; cropping the face region; extracting face keypoint features from the region; and finally, classifying and identifying faces using the retrieved face features. The amount of various face keypoints distinguishes the recognition techniques for composite faces from those for standard basic faces, and Figure 1 illustrates the overall facial recognition process built for the study.

Recognising a face from a sample of a picture of a facial expression is called facial recognition. The face picture

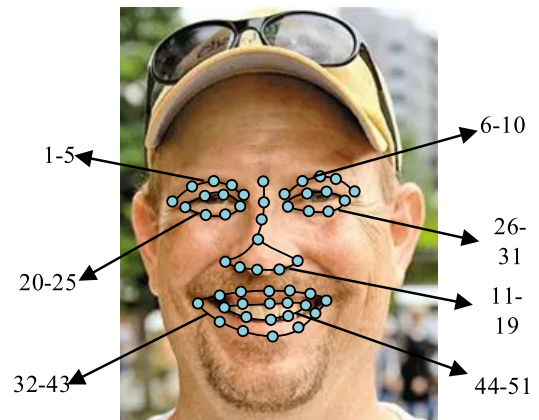


FIGURE 2. Facial expression key point number information.

sample will first be subjected to face detection using expression keypoints, after which the face region will be clipped and face keypoint features will be retrieved. Finally, the extracted face features will be utilised to categorise and recognise faces. Figure 1 depicts the general flow of facial recognition built for the study. Firstly, face image data depicting various emotions were collected from both natural and laboratory environments. The collected data was then normalized and preprocessed before extracting stylistic features present in the images via feature selection criteria according to various styles. Finally, a neural network classifier was utilized to perform facial recognition based on the extracted features. The amount of various face keypoints distinguishes the recognition methods for composite faces from those for typical basic faces. The Dlib library's detection algorithm, when compared to other intelligent detection algorithms, has the advantages of quick location information recognition and high expression keypoint identification accuracy [20], [21]. The set of face expression key points per frame is represented in equation (1) after the study eliminates some key points from the Dlib library's face expression key point extraction method that have little bearing on facial expressions, streamlines the extraction method, and numbers the remaining key points.

$$LM = \{LM_{t,i} | 1 \leq t \leq S, 1 \leq i \leq N\} \quad (1)$$

In equation (1),  $LM_{t,i}$  represents the coordinates of the expression keypoint numbered  $i$  in the recognized face in frame  $t$ ,  $N$  refers to the number of expression keypoints in each frame of the recognized object, and  $S$  refers to the frame length of the face expression sequence. The analysis of the face expression keypoint location and numbering information based on the Dlib library is shown in Figure 2.

Figure 2 shows a randomly selected face in the network with the key point numbering information extracted based on that face. As shown in Figure 1, the study streamlined the expression keypoints into 51 points, with the points representing the face at the eyebrows numbered 1-10, the points representing the bridge of the nose numbered 11-19,

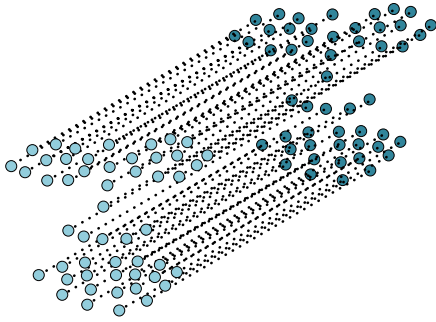


FIGURE 3. Connection method of adjacent frame facial key points.

the points representing the human eyes numbered 20-31 and the points representing the mouth numbered 32-51. The study makes full use of Graph Convolution Neural Networks (GCNNs) to extract temporal change information of face keypoints for dynamic facial recognition, and constructs spatio-temporal graphs based on face sequences. In order to build the spatio-temporal graph, edge sets must be used between the key points of the same frame in the face spatio-temporal sequence [22], [23]. The way in which edge sets are built will have an impact on the intrinsic relationship between the key points of the face and the direction in which the different frames move, which will significantly reduce the accuracy of the intelligent facial recognition system. Therefore, the study mainly adopts the three methods of face muscle distribution position, face organ geometry and full connection to form the edge set between the key points of the same frame of the face image, and the edge set between the key points of the same frame is represented as shown in equation (2) [24].

$$E_S = \{v_{t,i}v_{t,j} | (i, j) \subset H; i, j \in [1, N]\} \quad (2)$$

In equation (2),  $E_S$  refers to the set of normal edges between key points of faces in the same frame in the spatio-temporal map sequence, where  $N$  denotes the number of key points of faces at the moment of  $t$  in the spatio-temporal map sequence,  $H$  is the set consisting of key point numbers under different construction methods, and  $v_{t,i}v_{t,j}$  is a binarized variable, when key points are connected to key points, i.e.  $(i, j) \subset H$ ,  $v_{t,i}v_{t,j} = 1$ , otherwise  $v_{t,i}v_{t,j} = 0$ . Information on facial images needs to be propagated, requiring connections to be made between keypoints of adjacent frames. Technical term abbreviations will be explained when first used. Equation (3) demonstrates the process of connecting identically numbered feature points between successive frames to identify the edge set.

$$E_T = \{v_{t,i}v_{(t+1),j} | (i, j) \in [1, N]\} \quad (3)$$

In equation (3),  $E_T$  denotes the edge set of adjacent frame keypoints,  $v_{t,i}v_{(t+1),j}$  refers to the binarization variable, which is 1 when  $i$  is not the same as  $j$  and 0 otherwise, and the adjacent frame face keypoints are connected in the way shown in Figure 3.

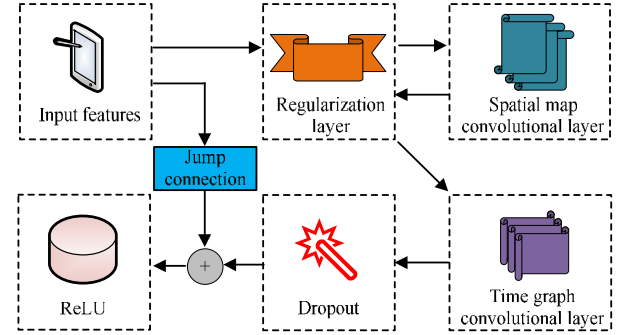


FIGURE 4. GCNNs structure.

In Figure 3, the light blue dot on the left indicates the key point of the first frame, the dark blue dot on the right indicates the face key point of the next frame, and the connected line segments in the middle are the edges connected by the same numbered key points. The GCNNs can dynamically analyse and integrate the key point information of face expressions through the model-specific modules, and can fuse the information between different frame nodes, and the structure of the spatio-temporal graphical convolutional network constructed in the study is shown in Figure 4.

As can be seen in Figure 4, GCNNs consist of four parts: the temporal graph convolution layer, the spatial graph convolution layer, the regularisation layer and the Dropout. In order to extract the original location information of the expression nodes, the spatio-temporal map needs to go through the regularisation layer to normalise the features of the nodes before they can be input to the spatial and temporal convolution layers. The information between nodes of the same frame is fused when the input data is located in the spatial map convolution layer. And between nodes of different frames when the input data is located in the temporal map convolution layer, and the spatio-temporal map is compressed in the temporal dimension. Then, input the spatiotemporal map into the Dropout layer. The Dropout layer can randomly ignore 50% of neural units during the training process, thereby reducing the dependency relationship between neural units and ultimately improving the model's generalization ability [25]. Also, to solve to some extent the overfitting problem caused by sparse datasets and the convergence problem of face keypoints caused by GCNNs, jump connections were introduced and finally a ReLU non-linear transformation was performed. The set of expression keypoints on the image of frame  $k$  is assumed to be set to  $V_k$ , and the relationship between the nodes between frames is set to  $E_S(k) = \{v_{t,i}, v_{t,j} | t = k, (i, j) \subset H; i, j \in [1, N]\}$ . In GCNNs, the face feature maps obtained from the spatio-temporal maps are all two-dimensional, and the size of the convolution kernel is assumed to be  $K \times K$ . The values located at the  $x$  position on the specific path of the output feature map are calculated as shown in equation (4).

$$f_{out}(x) = \sum_{h=1}^K \sum_{w=1}^K f_{in}(p(x, h, w)) \cdot (h, w) \quad (4)$$



In equation (4),  $f_{out}$  denotes the output feature map,  $f_{in}$  denotes the input feature map,  $K$  is the convolution kernel size,  $p$  is the probability,  $w$  is the convolution kernel and  $h$  is the node feature. Compared with 2D image data arranged according to rules, the number of neighbours of each node in spatio-temporal graph data will change, and the calculation method of traditional graph convolution cannot be applied directly. Therefore, the sampling and weighting functions in the graph convolution operation need to be improved according to the way the convolution layer is calculated. The process of calculating the neighbour points of the sampling function in the spatio-temporal graph convolution improved by the study is shown in equation (5).

$$N(v_{t,i}) = \{v_{t,j} | d(v_{t,i}, v_{t,j}) \leq D\} \quad (5)$$

In equation (5),  $N(v_{t,i})$  is the set of neighbouring nodes of node  $v_{t,i}$ ,  $d(v_{t,i}, v_{t,j})$  is the shortest path between  $v_{t,i}$  and  $v_{t,j}$ , and  $D$  denotes the distance where the usual value takes 1. The face features represented by the expression keypoints in the region of size  $K \times K$  obtained after the convolution operation all correspond to a  $c$ -dimensional weight vector in the weight function. And in the graph convolution operation, the neighbouring nodes  $v_{t,i}$  are divided into  $M$  mutually exclusive subsets according to their distances from the neighbouring face keypoints [26]. After defining a mapping  $l_{t,i} : N(v_{t,i}) \rightarrow \{0, \dots, M-1\}$  to give labels to the sampled node  $v_{t,i}$  and its neighbouring nodes attributes, and then derive the corresponding  $c$  dimensional weight vector from the weight function according to the labels, the equation for the graph convolution at this point is shown in equation (6).

$$f_{out}(v_{t,i}) = \sum_{v_{t,j} \in N(v_{t,i})} \frac{1}{Z_{t,i}(v_{t,j})} f_{in}(v_{t,j}) \cdot w(l_{t,i}(v_{t,j})) \quad (6)$$

In equation (6),  $Z_{t,i}$  denotes the number of nodes whose neighbours have the same label, and  $l$  denotes the node label. According to the uniform division strategy, a consistent weight vector  $w_1$  exists for each node and neighbouring point of the face expression key point, at which point the graph convolution is implemented as shown in equation (7).

$$f_{out} = \Lambda^{-\frac{1}{2}} (A + 1) \Lambda^{-\frac{1}{2}} f_{in} w_1 \quad (7)$$

In equation (7),  $A$  denotes the adjacency matrix,  $\Lambda$  denotes the degree matrix of the adjacency matrix  $A$ , and  $w_1$  denotes the weight vector. Information on facial images needs to be propagated, requiring connections to be made between keypoints of adjacent frames. Technical term abbreviations will be explained when first used. Equation (3) demonstrates the process of connecting identically numbered feature points between successive frames to identify the edge set.

$$f_{out} = \Lambda^{-\frac{1}{2}} (A + 1) \Lambda^{-\frac{1}{2}} f_{in} W \quad (8)$$

The weight matrix  $W$  is designated in equation (8). The face keypoints at the organs play a variety of roles throughout the change of distinct faces. For instance, the expression

keypoints at the nose are less variable than those at the mouth and have a less impact on the task of classifying faces. For the purpose of to enable the convolutional layer of each spatial map to learn more discriminative features, a trainable weight matrix  $M$  was created. The learning method is shown in equation (9), and it involves learning the relevance of each key point separately using the network model.

$$f_{out} = \Lambda^{-\frac{1}{2}} (A + 1) \Lambda^{-\frac{1}{2}} \odot M f_{in} W \quad (9)$$

$\odot$  is the product of corresponding nodes in equation (9). The study interconnects the same number of nodes between neighboring frames when building the spatio-temporal graph in order to let information move not only between nodes of the same frame but also between nodes of other frames. At this point, equation (10), which represents the set of nodes that node  $v_{t,i}$  neighbors, is displayed.

$$N(v_{t,i}) = \left\{ v_{q,i} | |q - t| \leq \left\lfloor \frac{m}{2} \right\rfloor \right\} \quad (10)$$

In equation (10),  $m$  refers to the temporal distance interval of the same numbered nodes, which usually takes odd values,  $t$  is the processing time and  $q$  denotes the feature vector. The arrangement of neighbours in the time dimension is ordered, so the set of neighbours can be divided into  $m$  subsets according to different weight vectors, at which point the labels of the nodes in the set of neighbours are shown in equation (11).

$$l_{t,i}(v_{q,i}) = l + q - t + \left\lfloor \frac{m}{2} \right\rfloor \quad (11)$$

Corresponding to the dimension  $C \times W \times H$  of the feature map after the convolutional layer of the usual image, the output feature dimension of the facial expression key points after the convolutional layer of the spatial map is set to  $C_{out} \times T \times N$ , the size of the convolutional layer is  $[m \times 1]$ , and the step size value is set to  $s$ . The facial key points perform the node convolution calculation of the following frame after the node convolution calculation of the  $m$  frame is finished, skipping the  $s$  frame along the time dimension, and continuing until all nodes with the same number in the same time series have finished the position convolution calculation. The output feature map dimension of the time map convolution is then obtained as  $C_{out} \times (T/s) \times 51$  by completing the convolution at the remaining number of nodes.

## B. IMPROVING GCNNS FOR FACIAL RECOGNITION SYSTEM CONSTRUCTION

After the construction of the spatio-temporal graph for detecting key points of face expressions, the study proposed a dynamic facial recognition model based on the spatio-temporal graph convolution mechanism and spatio-temporal graph convolution based on the Spatiotemporal Graph Convolution Neural Networks (SGCNNs) model [22], [23], [24]. The self-noticing mechanism calculates the similarity between nodes separately based on the features learned by each node. By fusing the location information of key points of the face, additional information can be provided to the convolution layer of the image, making the

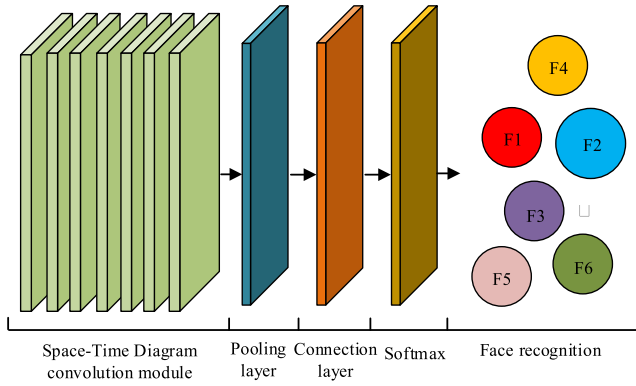


FIGURE 5. Structure diagram of STGCN-SA.

extracted face feature information more discriminative and more suitable for facial recognition tasks. Therefore, the study introduces a self-attentive mechanism based on spatial graph convolution to calculate the correlation between nodes in different spatio-temporal sequences, and obtains more feature information of face key points by weighted summation, and proposes Spatial-Temporal Graph Convolutional Networks With Self-Attention (STGCN). STGCN-SA mainly consists of seven self-attentive spatio-temporal graph convolutional modules, a global pooling layer and a softmax classifier, and the structure of the basic model of STGCN-SA is shown in Fig. 5.

In Figure 5, F1-F6 refer to the different faces recognized, and the output set of the self-attention mechanism consists of a new set of node features. To convert the features of each node to higher-level features before calculating the attention coefficients, the study multiplies all nodes in the spatio-temporal graph by a trainable weight matrix, and then uses the attention mechanism to calculate the attention coefficients between the nodes, calculated as shown in equation (12).

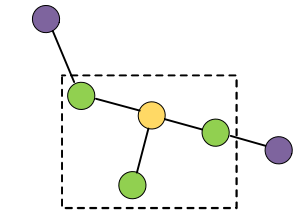
$$e_{i,j} = a \left( W_2 \vec{h}_i, W_2 \vec{h}_j \right) \quad (12)$$

In equation (12),  $h = \{ \vec{h}_1, \vec{h}_2, \dots, \vec{h}_N \}$  denotes the set composed of the features of each node,  $e_{i,j}$  refers to the direct attention coefficients of the nodes, and  $a$  is the weight vector. The study immediately follows the normalisation of all attention coefficients of node  $i$  using the Softmax function, and the correlation of node  $j$  to node  $i$  is calculated as shown in equation (13).

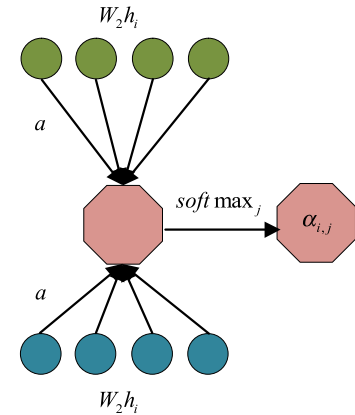
$$\alpha_{i,j} = \text{softmax}_j(e_{i,j}) = \frac{\exp(e_{i,j})}{\sum_{k=1}^N \exp(e_{i,k})} \quad (13)$$

In equation (13),  $\alpha$  refers to the importance parameter and the attention factor is calculated as shown in equation (14).

$$e_{i,j} = \vec{a}^T \left[ W_2 \vec{h}_i \parallel W_2 \vec{h}_j \right] \quad (14)$$



(a) Node division strategy diagram



(b) Self attention mechanism architecture

FIGURE 6. Neighborhood partitioning and attention mechanism structure.

In equation (14),  $\vec{a}$  denotes the parameterised weight vector,  $T$  is the total number of vectors and the normalised attention factor is shown in equation (15).

$$\alpha_{i,j} = \frac{\exp \left( \vec{a}_2 W_2 \vec{h}_i \right)}{\sum_{k=1}^N \exp \left( \vec{a}_2 W_2 \vec{h}_k \right)} \quad (15)$$

In equation (15), the weight vector normalisation process is followed by a loss of node location information for all weight parameters, thus losing relevance to the node before. In order to retain the key point location information, the study first uses the activation function to perform a non-linear transformation of the attention coefficients before normalisation, with the slope of the function set to 0.2, at which point the importance of node  $i$  to node  $j$  is calculated as shown in equation (16).

$$\alpha_{i,j} = \frac{\exp \left( \text{Leaky Re LU} \left( \vec{a}^T \left[ W_2 \vec{h}_i \parallel W_2 \vec{h}_j \right] \right) \right)}{\sum_{k=1}^N \exp \left( \text{Leaky Re LU} \left( \vec{a}^T \left[ W_2 \vec{h}_i \parallel W_2 \vec{h}_j \right] \right) \right)} \quad (16)$$

In equation (16),  $\parallel$  denotes the connection operator, at which point the structure of the node-division region and the self-attentive mechanism is shown in Figure 4.

At a distance of 1, the green node is next to the yellow node, which is the current node. The adjacency matrix represents the connection relationship between the key points, which can clearly represent the neighbours of each node, and the

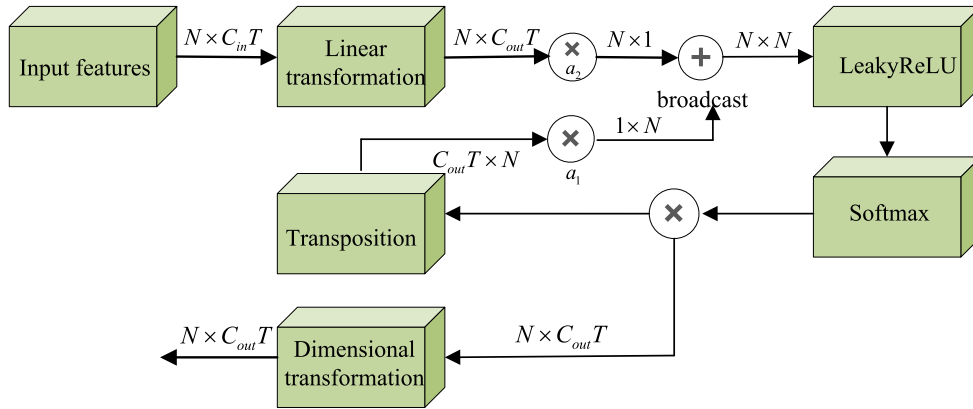


FIGURE 7. Flow chart of self attention mechanism.

normalised attention coefficients in Figure 6(a) are obtained in Figure 6(b). The current node and its neighbours are delineated as a subset with the same label. The final output of each node is obtained by weighting and adding the attention coefficients using the node features. In equation (17), the output characteristics of node  $i$  are displayed.

$$\vec{h}'_i = \sigma \left( \sum_{j=1}^N \alpha_{i,j} W_2 \vec{h}_j \right) \quad (17)$$

In equation (17),  $\sigma$  is the non-linear transformation function and  $\vec{h}'_i$  denotes the output characteristics of node  $i$ .

$$f(x) = \begin{cases} \alpha (e^x - 1), & x < 0 \\ x, & x \geq 0 \end{cases} \quad (18)$$

In an effort to expedite the training of the model, the study fixed the value of the parameter  $\alpha$  to 1 and the distribution of the output of the activation function to a zero-mean one. This self-attentive process was expanded to a multi-headed attention mechanism to add more information to the learned features [27]. That is, as stated in equation (19), characteristics are extracted and connected using a number of independent self-attentive methods.

$$\vec{h}'_i = \left\|_{k=1}^K \sigma \left( \sum_{j=1}^N \alpha_{i,j}^k W_2^k \vec{h}_j \right) \right. \quad (19)$$

Transform equation (19) by replacing the continuous operation with the averaging operation as shown in equation (20).

$$\vec{h}'_i = \sigma \left( \frac{1}{K} \sum_{k=1}^K \sum_{j=1}^N \alpha_{i,j}^k W_2^k \vec{h}_j \right) \quad (20)$$

Since the spatio-temporal graphs in the study have a temporal dimension, SGCNNs need to be improved by assuming that there are  $N$  nodes in each frame, the number of temporal dimensions is  $T$ , the node feature dimension in the input feature map of the graph attention layer is  $C_{in}$ , and the node feature dimension of the output is  $C_{out}$ . To streamline the

computational process, the study treats nodes of the same sequence as one node in the order of the temporal dimensions, and the specific implementation flow is shown in Figure 8.

Figure 7 shows the time dimension of the face input feature map as  $N \times C_{in} T$ , the weight coefficients as  $a_1$  and  $a_2$ , and the broadcast method as the plus sign (plus). The input feature map is linearly transformed to give  $N \times C_{out} T$ . Summing the weight vectors gives the exact number of rows and columns of the time dimension as  $N \times N$ . This number is then used to create the time matrix, which is dimensionally transformed by Softmax. Once the weight vectors are combined, the number of rows and columns in the time dimension is determined. This value is employed to generate the time matrix. The nodes that correspond to the rows and columns of the matrices are summed sequentially in order to obtain the ultimate convolution computation when the dimensions of the time and space matrices coincide, as illustrated in equation (21).

$$[1 \ 2 \ 3] + \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix} \quad (21)$$

It is crucial to merge these two characteristics because node feature updates in the spatio-temporal graph depend on both the global features derived from the spatial graph convolution layer and the distinctive features extracted from the samples using the attention mechanism. The study introduces weight vectors into each module and trains the expressions along with the parameters of the facial recognition model to adaptively generate the important attention mechanism with self-output features, taking into account the varying importance of the output features of the spatial graph convolution layer and the attention mechanism in different modules. Equation (22) displays the output characteristics of the spatial graph's convolutional layer with self-attention enhancement.

$$f_{out}^{(l)} = f_g^{(l)} + \alpha^{(l)} \cdot f_a^{(l)} \quad (22)$$

In equation (22),  $f_g^{(l)}$  denotes the output feature label of the spatial graph convolution layer,  $f_a^{(l)}$  denotes the output feature

label of the self-attentive mechanism, and  $\alpha^{(l)}$  is the label weight. the STGCN-SA model achieves good dynamic facial recognition performance by tracking the coordinate changes of key points in the face expression sequence. However, there are frequently a lot of unfavourable aspects in actual applications, such as lighting conditions and variations in face keypoint detection techniques, which leads to inaccurate tagging of face expression keypoints and negatively impacts the model's facial recognition performance [28]. Therefore, in addition to precisely identifying the important points of faces, compensation must be done using various techniques in order to increase the network model's identification accuracy. The study used a convolutional neural network model to extract spatial texture information from peak frames of dynamic expression sequences and an adaptive spatial map convolutional network model to extract temporal change information from spatio-temporal maps. In order to improve the recognition accuracy of STGCN-SA, it is necessary to integrate these two different types of information together. Depending on the information fusion method, it can be divided into feature layer fusion and decision layer fusion. Feature layer fusion belongs to the intermediate layer fusion algorithm in the field of fusion. In the field of image recognition, by fusing different classes and scales of features of image samples, more robust and effective face features can be obtained to improve the representation of features and obtain better recognition results. Decision-level fusion is a widely used approach in the area of early information fusion and is considered advanced fusion. This method initially acquires classification vector data for each feature according to the distinct feature information of the sample. Subsequently, categorisation information from each characteristic is combined by employing a correlation algorithm, culminating in the delivery of the final classification outcome. This is a simple and effective method that assigns different weights to each feature classifier and obtains the final expressed category by weighted summation. The final category vector is shown in Equation (23).

$$l = w_0 \cdot E + w_1 \cdot F \tag{23}$$

In equation (23),  $w_0$  and  $w_1$  denote the weights of the output vectors of the network model,  $E$  is the output vector of the spatio-temporal map convolutional network with fused attention mechanism, and  $F$  denotes the output vector of the convolutional neural network. The study makes use of feature layer fusion and decision layer fusion for dynamic extraction of spatio-temporal sequence information respectively, which in turn helps the STGCN-SA system model to achieve the function of facial recognition.

### III. FACIAL RECOGNITION MODEL PERFORMANCE TESTING

#### A. GCNNs MODEL SIMULATION TEST

The study uses face datasets for dynamic face recognition effect analysis to validate the performance of GCNNs models. Currently, the main public datasets for facial recognition are

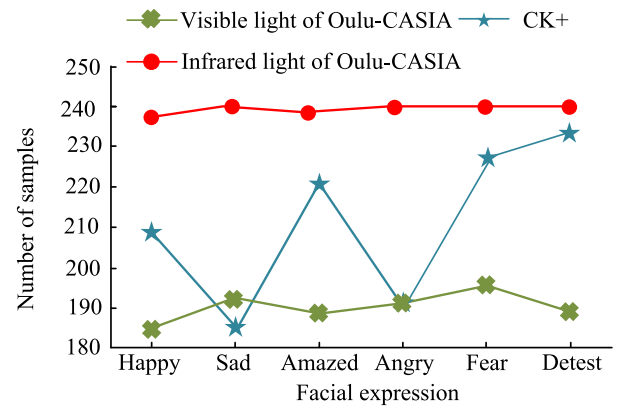


FIGURE 8. Facial expression distribution at CK+ and Oulu-CASIA.

TABLE 1. Dataset experimental parameters.

Data set	CK+	Oulu-CASIA
Learning rate	0.1	0.01
Batch size	100	256
Periodization	300	300
Regularization coefficient	0.00001	0.00001

AFEW, CK+, Oulu-CASIA and MMI dynamic face datasets. There are more side face images in the MMI dataset, which cannot help the model to accurately locate expression keypoints. In addition, most of the samples in the AFEW dataset are clips from movies, which again cannot provide valid face images suitable for face keypoint capture. Therefore, the study used the CK+ and Oulu-CASIA datasets to verify the validity of the method. To facilitate the differentiation of the experimental information, the six expressions of happiness, sadness, surprise, anger, fear and disgust were considered as different faces, and the experimental results are shown in Figure 8.

The distribution of face expressions in the Oulu-CASIA dataset is relatively even, with the distribution of faces in IR and visible light being around 240 and 190 respectively, while sad and angry expressions in CK+ are slightly lower than other face expressions, but in general they are not too far apart, being around 180-230. The number of frames of key point sequences in each dataset was studied and set to 16. The parameters of the network model were randomly initialised during the training process and the parameters of the network model were optimised by Adam's algorithm using the cross-entropy function as the loss function. The model parameters that minimise the loss function are saved during the training of the GCNNs model. Finally, in the testing phase, the test set was fed into the saved facial recognition model to obtain the recognition accuracy of the model. the parameters such as learning rate, batch size and training period in the CK+ and Oulu CASIA datasets were set as shown in Table 1.

The test procedure designed for the study mainly used data from the dynamic face datasets in CK+ and Oulu CASIA. However, due to the relatively small total number of samples



**TABLE 2. Classification performance of GCNNs model.**

Data set	CK+			Oulu-CASIA		
	Facial muscle distribution	Facial geometric structure	Full connected	Facial muscle distribution	Facial geometric structure	Full connected
Parameter quantity (Million)	1.1539	1.1539	1.1539	1.1537	1.1537	1.1537
Accuracy(%)	92.36	92.67	93.89	77.71	76.58	78.69

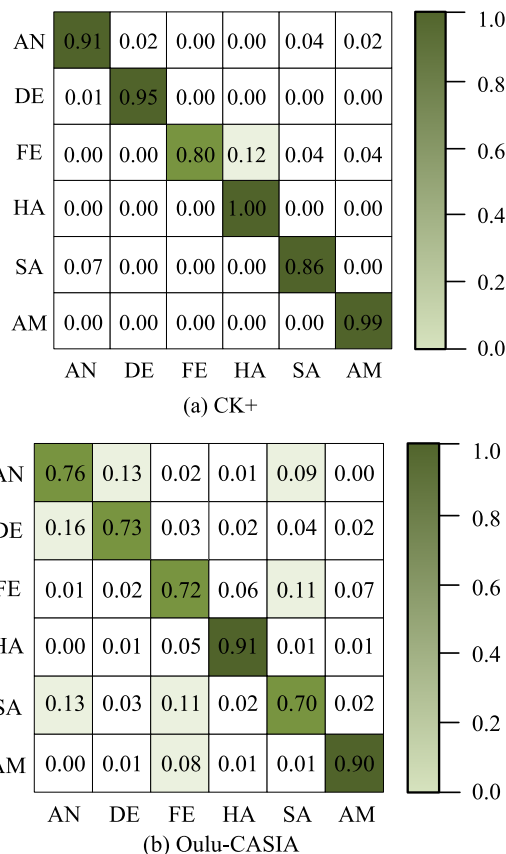
in CK+ and Oulu CASIA, it is necessary to use dataset partitioning methods to improve the estimation accuracy of the algorithm. Ten fold cross validation method refers to the method of dividing a dataset into ten parts for cross validation, which can help select the best model parameters and accurately evaluate the performance of the model. Therefore, the study used the tenfold cross validation method to evenly divide the data into 10 parts, with 9 parts as the training set and the remaining 1 part as the test set, and the average accuracy was taken as the final result. Table 2 shows the classification performance of the GCNNs model for spatio-temporal graphs constructed in different ways in the CK+ and Oulu CASIA dynamic representation datasets.

As can be observed from Table 2, the GCNNs model has the highest recognition accuracy of 93.89% and 78.69% for spatio-temporal maps formed by fully connected keypoints in the same frame, both in the CK+ dataset and the Oulu CASIA dataset, respectively. This indicates that fully connected is superior to construction methods derived from facial muscle distribution and methods based on facial organ geometry. By connecting each node to all the remaining nodes in the same frame, the information fusion process between nodes can be effectively accelerated, whereas the other two construction methods have difficulties in transmitting information between nodes. Furthermore, as can be seen from the table, the total number of parameters in the GCNNs model is only approximately 1.15 million in the CK+ and Oulu CASIA datasets. However, as the CK+ dataset has more expression categories than the Oulu CASIA dataset, the number of parameters in the network model differs. Six different expressions were fused with each other to represent six simple faces versus 30 mixed faces, where anger, disgust, fear, happiness, sadness and surprise were represented by AN, DE, FE, HA, SA and AM respectively, and used in the construction of the confusion matrix experiments, Figure 9 shows the confusion matrix of the GCNNs model under different datasets.

As can be seen from Figure 9, the GCNNs model has an average accuracy of 92% and 79% for simple facial recognition in the CK+ and Oulu CASIA datasets, which is good for facial recognition and can meet the requirements for facial recognition in a simple single face laboratory environment.

**B. STGCN-SA FACIAL RECOGNITION SYSTEM PERFORMANCE TEST**

To validate the recognition effect of STGCN-SA with different multiple faces in the same frame, the study conducted experiments on the CK+ and Oulu-CASIA dynamic



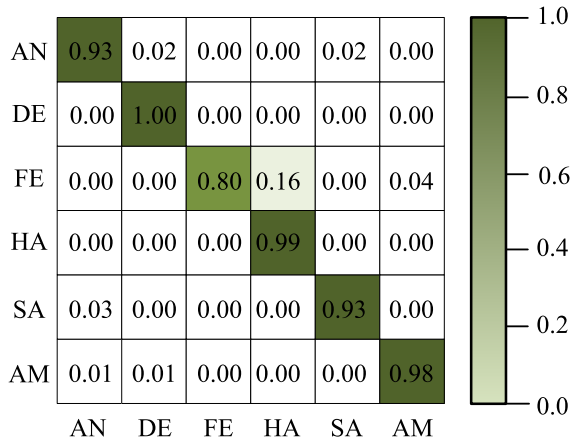
**FIGURE 9. Confusion matrix of GCNNs model.**

expression datasets, and the performance test results of the facial recognition system are shown in Table 3.

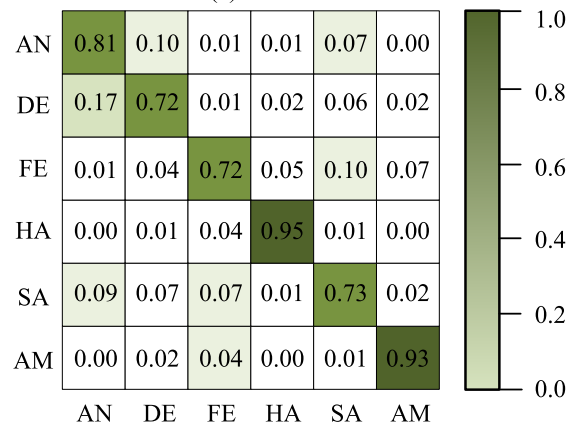
As can be seen from Table 3, the recognition accuracy of STGCN-SA in the experimental dataset was on average 94.04% versus 79.89%, respectively. Compared with the highest recognition accuracy of the traditional GCNNs model, STGCN-SA improved by 0.92% on the CK+ dataset and 2.29% on the Oulu-CASIA dataset, verifying the feasibility and validity. This is mainly because the key point distribution features are different between different faces, and the convolutional layer of the spatial graph in the spatio-temporal graph convolutional network can only integrate information with the fixed topology in the spatio-temporal graph. Therefore, by embedding a self-attentive mechanism, STGCN-SA can provide additional information to the spatial graph convolutional layer based on the unique node relationship of each expression sample, making the extracted features more

TABLE 3. STGCN-SA performance test result.

Data set	CK+				Oulu-CASIA			
Number of faces	1	2	3	4	1	2	3	4
Parameter quantity (Million)	1.3159	1.4778	1.6397	1.8016	1.3856	1.4775	1.6394	1.8014
Accuracy(%)	94.81	93.28	94.19	93.89	79.20	79.51	79.86	80.99



(a) CK+



(b) Oulu-CASIA

FIGURE 10. Confusion matrix of STGCN-SA model.

discriminative. For the CK+ and Oulu-CASIA datasets, the confusion matrix of the dynamic facial recognition model based on the self-attentive mechanism spatio-temporal graph convolutional network is shown in Figure 10.

As shown in Figure 10, the STGCN-SA model has an average accuracy of 94% and 81% for multi-facial recognition in the same frame in the CK+ and Oulu CASIA datasets, which is good for facial recognition and can meet the requirements for facial recognition in complex laboratory environments. The confusion matrix of the STGCN-SA model combining feature layer and decision layer data fusion under different datasets is shown in Figure 11.

As can be seen from Figure 11, the facial recognition accuracy of the STGCN-SA model is 96% versus 93% in the CK+ database and 90% versus 89% in the Oulu-CASIA

database after fusion of data in the feature or decision layers. The detection results of facial recognition models in different datasets have some variability. This is because the number of images detected varies between datasets, and the points of expression differ. The Oulu-CASIA dataset features faces ranging from 23 to 58 years of age, with 73.8% of the subjects being male. In contrast, the CK+ dataset, an image database for facial expressions coded at the CMU Pittsburgh AU, includes 2,105 digitised images from 182 adult subjects of various ethnicities. Despite the variability in detection, the STGCN-SA model still achieves an accuracy of over 89%. Indicating that the recognition effectiveness of the STGCN-SA model has been improved.

C. COMPARISON AND ANALYSIS OF STGCN-SA MODEL EFFECTS

The study chooses the main network to extract initialized traits in order to ascertain the optimal configuration for identifying facial expressions. Experiments then compare the STGCN-SA with the Line Hausdorff distance (LHD) facial recognition approach, demonstrating the superiority and effectiveness of the STGCN-SA. The experiments are judged by the accuracy of face expression recognition. According to the model structure of LHD and STGCN-SA, an image of size 100\*100 is crawled during the construction of the face spatio-temporal map, and the initialized feature dimension coming out of the GCNNs is set to 2000, and the feature dimension after the graph convolution network and attention mechanism is 128. Since the t-distributed stochastic neighborhood embedding (t-SNE) method can reduce the distribution gap between different points and make the spatial distribution of data points more uniform, thus clearly showing the effect of visualization and processing of high-dimensional data. Therefore, this study uses the t-SNE method to visualize the distribution of face features to verify the performance of the model. The samples used four face features, AN, FE, SA and DE, and the feature distribution of LHD and STGCN-SA in the dataset used for the study are shown in Figure 12.

Figures 12(a) and 12(c) show that the face features successfully recognised by the LHD model in the two experimentally selected datasets are more dispersed and are likely to cause face confusion during practical application, while Figures 12(b) and 12(d) show that the face features recognised by the STGCN-SA facial recognition system developed by the study are more concentrated and closer together in the feature space, demonstrating that the model has improved.

Then, to verify the accuracy of the STGCN-SA facial recognition system, the study selected 10 facial images from

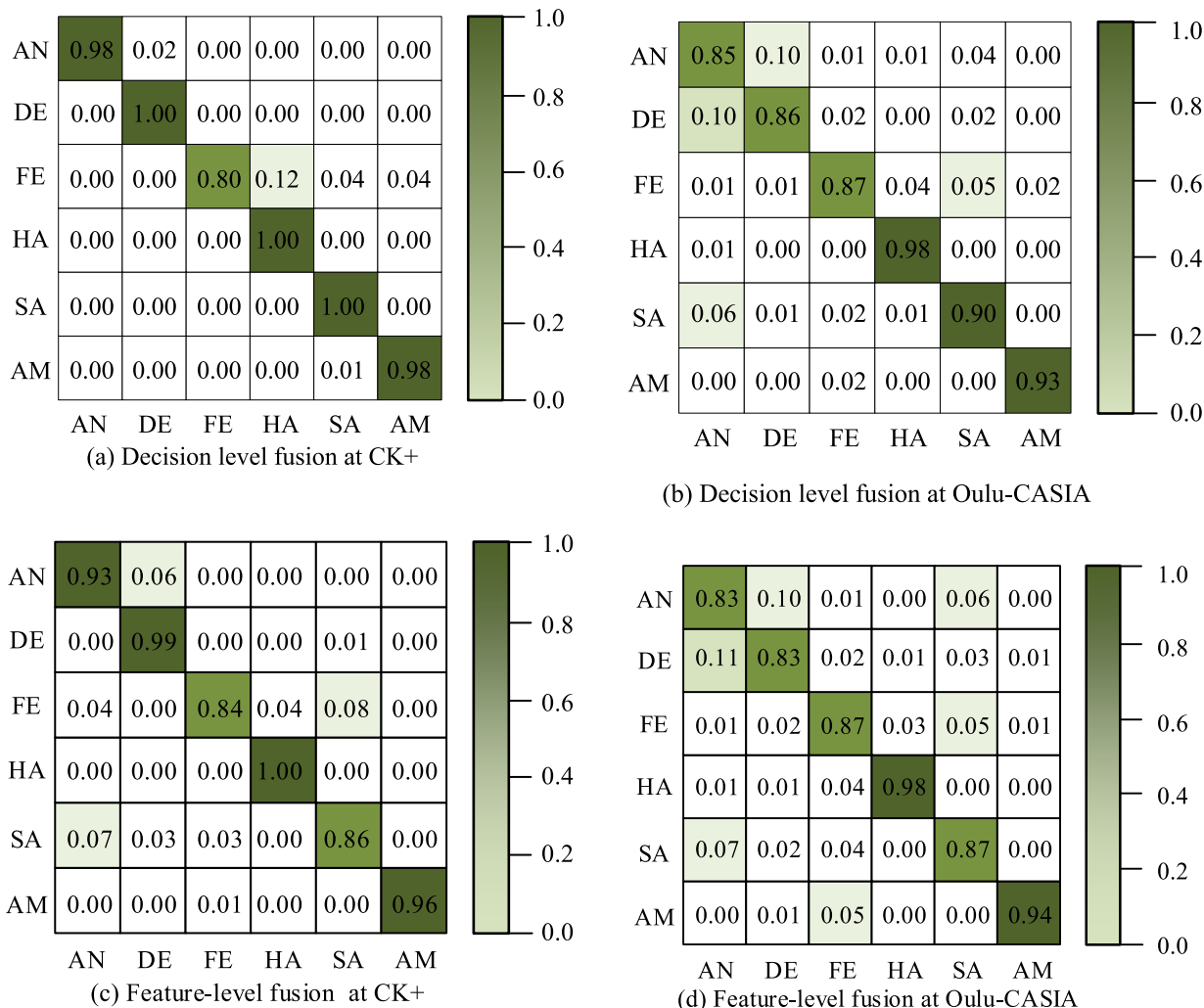


FIGURE 11. Data fusion confusion matrix of STGCN-SA model.

TABLE 4. STGCN-SA comparative experimental results.

Number of total faces	Number of faces to be recognized	Number of interference faces	Number of errors identified		STGCN-SA
			Method of reference [12]	Method of reference [14]	
40	30	10	0	0	0
50	30	20	0	0	0
60	30	30	0	0	0
70	50	20	0	6	0
70	60	10	0	6	0
80	60	20	0	5	0
80	70	10	1	8	0
90	70	20	3	20	0
90	80	10	3	6	0
100	80	20	5	16	1
100	90	10	6	17	0

the CK+dataset, with each candidate taking 10 images. Then, the facial recognition system designed in [12] and [14] was compared with the STGCN-SA facial recognition system. The experimental results are shown in Table 4.

From Table 4, it can be seen that the STGCN-SA facial recognition system has a higher accuracy in facial recognition compared to the facial recognition systems designed in [12] and [14]. It can accurately recognize the faces that need to

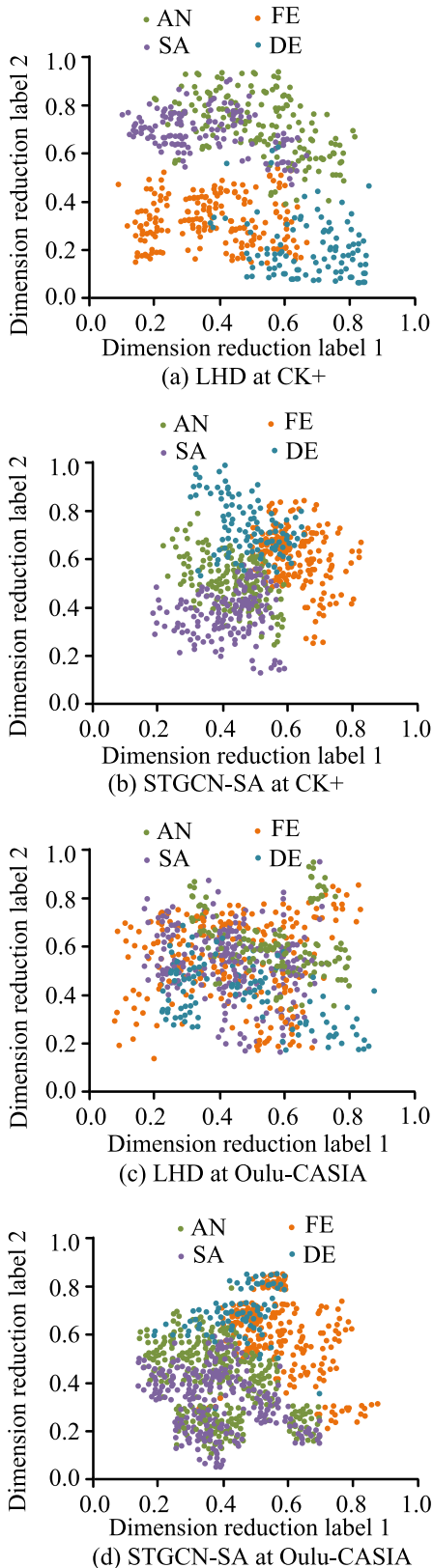


FIGURE 12. Facial feature distribution of LHD and STGCN-SA.

be recognized in the presence of interference terms, and has higher application value.

IV. CONCLUSION

To improve the performance of the laboratory facial recognition system, the study proposes an improved STGCN-SA facial recognition system using GCNNs to capture spatio-temporal sequences of key points of facial expressions and then incorporating attention mechanisms in the spatio-temporal graph, and improves the efficiency of the laboratory operation through data fusion between the feature and decision layers. The experimental results show that the facial recognition system designed for the study has a facial recognition accuracy of 92% in the simple environment of the CK+ database, 79% in the simple environment of the Oulu CASIA database, 94% in the complex environment of the CK+ database and 91% in the complex environment of the Oulu CASIA database. The STGCN-SA facial recognition system’s facial recognition accuracy has improved after data fusion, and the improvement is above 89% in various data sets, which can already match the requirements of facial recognition in most circumstances. This indicates that the face recognition technology in this study has a high recognition accuracy, and thus can be applied to fields that require high security performance. For example, applying the high-precision face recognition system to school access control, the facial information of school teachers and students is collected in advance and entered into the facial recognition system, and when the personnel enter the school gate, the facial recognition system can compare the detected facial features and recognize them according to the entered database. In order to reduce the risk of accidental entry of outsiders; applied to the bank access control, can also improve the security protection coefficient of the bank peripheral system. In addition, facial recognition can also be applied to network security payments, the payment system is set up as a facial recognition system, payment can be made by scanning the facial information of the payer to determine the payment security environment.

Although the facial recognition technology proposed by the research can already achieve a high accuracy rate, the technology is still limited to front-facing recognition of the human face, and accurate recognition is challenging when facial features are viewed from the side. Meanwhile, in academic research, the GCNN algorithm has been combined with the STGCN-SA technique to create a facial recognition model. This development advances the technology for extracting key facial features and improves the detection performance of the facial recognition system. In order to enhance the facial recognition system’s performance in recognizing facial features in various orientations, future research could consider incorporating additional information, like improving the system’s ability to identify side facial features. This study primarily extracted face information data utilizing the key domain of frontal facial expression.

REFERENCES

[1] J. Luo and X. Zhang, “Convolutional neural network based on attention mechanism and Bi-LSTM for bearing remaining life prediction,” *Appl. Intell.*, vol. 52, pp. 1076–1091, May 2022.



- [2] Y. Xu, H. Qin, J. Huang, and Y. Wang, "An end-to-end deep context gate convolutional visual odometry system based on lightweight attention mechanism," *Ind. Robot, Int. J. Robot. Res. Appl.*, vol. 49, no. 1, pp. 47–53, Jan. 2022.
- [3] Y. Tian, S. Sun, Z. Qi, Y. Liu, and Z. Wang, "Non-tumorous facial pigmentation classification based on multi-view convolutional neural network with attention mechanism," *Neurocomputing*, vol. 483, pp. 370–385, Apr. 2022.
- [4] Z. Zhang, X. Gong, and J. Chen, "Face recognition based on adaptive margin and diversity regularization constraints," *IET Image Process.*, vol. 15, no. 5, pp. 1105–1114, Apr. 2021.
- [5] T. Chen, T. Gao, S. Li, X. Zhang, J. Cao, D. Yao, and Y. Li, "A novel face recognition method based on fusion of LBP and HOG," *IET Image Process.*, vol. 15, no. 14, pp. 3559–3572, Dec. 2021.
- [6] P. Li, P. Zhang, and X. Xu, "Graph convolutional network meta-learning with multi-granularity POS guidance for video captioning," *Neurocomputing*, vol. 472, pp. 294–305, Feb. 2022.
- [7] L. Dong, Z. Liang, and Y. Wang, "Graph convolutional network-based image matting algorithm for computer vision applications," *IET Image Process.*, vol. 16, no. 10, pp. 2817–2825, Aug. 2022.
- [8] W. Liao, D. Yang, Y. Wang, and X. Ren, "Fault diagnosis of power transformers using graph convolutional network," *CSEE J. Power Energy Syst.*, vol. 7, no. 2, pp. 241–249, Mar. 2021.
- [9] R. T. Palazuelos and E. L. Drogue, "System-level prognostics and health management: A graph convolutional network-based framework," *Proc. Inst. Mech. Eng., O. J. Risk Rel.*, vol. 235, no. 1, pp. 120–135, 2021.
- [10] L. Peng, F. Kong, C. Liu, and P. Kuang, "Robust and dynamic graph convolutional network for multi-view data classification," *Comput. J.*, vol. 64, no. 7, pp. 1093–1103, Aug. 2021.
- [11] J.-Y. Niu, Z.-H. Xie, Y. Li, S.-J. Cheng, and J.-W. Fan, "Scale fusion light CNN for hyperspectral face recognition with knowledge distillation and attention mechanism," *Appl. Intell.*, vol. 52, no. 6, pp. 6181–6195, Apr. 2022.
- [12] X. Bajrami and B. Gashi, "Face recognition with raspberry pi using deep neural networks," *Int. J. Comput. Vis. Robot.*, vol. 12, no. 2, pp. 177–193, 2022.
- [13] X. Chen and H. Y. K. Lau, "The identity-level angular triplet loss for cross-age face recognition," *Int. J. Speech Technol.*, vol. 52, no. 6, pp. 6330–6339, Apr. 2022.
- [14] P. Guo, G. Du, L. Wei, H. Lu, S. Chen, C. Gao, Y. Chen, J. Li, and D. Luo, "Multiscale face recognition in cluttered backgrounds based on visual attention," *Neurocomputing*, vol. 469, pp. 65–80, Jan. 2022.
- [15] S. Wang, H. Ge, J. Yang, and S. Su, "Virtual samples based robust block-diagonal dictionary learning for face recognition," *Intell. Data Anal.*, vol. 25, no. 5, pp. 1273–1290, Sep. 2021.
- [16] Y. Lu, Z. Lai, X. Li, W. K. Wong, C. Yuan, and D. Zhang, "Low-rank 2-D neighborhood preserving projection for enhanced robust image representation," *IEEE Trans. Cybern.*, vol. 49, no. 5, pp. 1859–1872, May 2019.
- [17] Y. Lu, C. Yuan, W. Zhu, and X. Li, "Structurally incoherent low-rank nonnegative matrix factorization for image classification," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5248–5260, Nov. 2018.
- [18] J. Newton, "Trusting AI it's time to unravel the ethical and social nuances of using artificial intelligence for scientific research," *Chem. World*, vol. 19, no. 5, pp. 30–31, 2022.
- [19] B. Crowell, J. Cug, and K. F. Michalikova, "Smart wearable Internet of Medical Things technologies, artificial intelligence-based diagnostic algorithms, and real-time healthcare monitoring systems in COVID-19 detection and treatment," *Amer. J. Med. Res.*, vol. 9, no. 1, pp. 17–32, 2022.
- [20] S. A. Wagle and R. Hari Krishnan, "Prediction of tomato plant disease with meteorological condition and artificial intelligence," *ECS Trans.*, vol. 107, no. 1, pp. 20377–20384, Apr. 2022.
- [21] M. Eitzbach, "Artificial intelligence helps detect and enforce safety violations," *Police Chief*, vol. 89, no. 2, pp. 18–20, 2022.
- [22] J. T. Gruenwald, J. Nistor, and J. Tutar, "Powering our nuclear fleet with artificial intelligence," *Nucl. News*, vol. 65, no. 2, pp. 50–58, 2022.
- [23] X. Y. Wang, M. H. Cheng, J. Eaton, C. Hsieh, and S. F. Wu, "Fake node attacks on graph convolutional networks," *J. Comput. Cogn. Eng.*, vol. 1, no. 4, pp. 165–173, 2022.
- [24] Y. Yang and X. Song, "Research on face intelligent perception technology integrating deep learning under different illumination intensities," *J. Comput. Cognit. Eng.*, vol. 1, no. 1, pp. 32–36, 2022.
- [25] J. Choe, S. Lee, and H. Shim, "Attention-based dropout layer for weakly supervised single object localization and semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4256–4271, Dec. 2021.
- [26] N. Hajarolasvadi and H. Demirel, "Deep facial emotion recognition in video using eigenframes," *IET Image Process.*, vol. 14, no. 14, pp. 3536–3546, Dec. 2020.
- [27] S. Chatterjee, B. Singh, A. Diwan, Z. R. Lee, M. H. Engelhard, J. Terry, H. D. Tolley, N. B. Gallagher, and M. R. Linford, "A perspective on two chemometrics tools: PCA and MCR, and introduction of a new one: Pattern recognition entropy (PRE), as applied to XPS and ToF-SIMS depth profiles of organic and inorganic materials," *Appl. Surf. Sci.*, vol. 433, pp. 994–1017, Mar. 2018.
- [28] S. A. Khan, R. Shabbir, S. M. Malik, and U. Sehar, "Comparison of different face recognition algorithms," *Int. J. Innov. Sci. Eng. Technol.*, vol. 4, no. 7, pp. 219–223, 2016.



three articles. His research interests include smart campus construction, the intelligent management of weak current, and teaching building construction.

**YING ZHOU** was born in Jiangsu, China, in 1984. He received the B.S. degree in agriculture science and the M.S. degree in forest cultivation from Nanjing Forestry University, Jiangsu, in 2006 and 2019, respectively. From 2006 to 2013, he was a Weak Current Engineer with Jiangsu Dongda Jinzhi Intelligent Technology Company Ltd., Jiangsu. Since 2014, he has been a Staff Member of the Infrastructure Office, Nanjing Forestry University. He is the author of more than



**YOUWANG LIANG** was born in Guangxi, China, in 1979. He received the B.S. degree in agriculture science and the M.S. degree in forest cultivation from Nanjing Forestry University, Jiangsu, China, in 2003 and 2006, respectively. From 2006 to 2009, he was an Assistant Experimentalist, and he has been an Experimentalist, since 2014. He is the coauthor of one book, more than five articles. His research interests include the cultivation and utilization of non-wood forests and laboratory management technique.



interests include the cultivation and utilization of non-wood forests, plant tissue culture technique, and laboratory management technique.

**PENGPENG TAN** was born in Shandong, China, in 1982. She received the B.S. degree in agriculture science and the M.S. degree in forest cultivation from Nanjing Forestry University, Jiangsu, China, in 2006 and 2009, respectively. From 2009 to 2014, she was an Assistant Experimentalist. From 2015 to 2022, she was an Experimentalist. She has been a Senior Experimentalist, since 2022. She is the coauthor of one book, more than five articles. Her research