A Method for Identity Feature Recognition in Wireless Visual Sensing Networks Based on Convolutional Neural Networks

Chenyang Li(✉) and Zhiyu Huang

Shenyang Institute of Technology, Shenyang 113122, China
liklu22@163.com

Abstract. Due to the problems of low recognition accuracy and long recognition time in traditional wireless visual sensing network identity feature recognition methods, a convolutional neural network-based wireless visual senscto the operation results, the global threshold method is used to obtain the binary image sequence and perform morphological processing. Based on the processing results, Extract target regions from video image sequences of wireless visual sensing networks, detect human targets, and construct a Softmax classifier using convolutional neural networks to classify human targets in video image sequences of wireless visual sensing networks, in order to identify identity features. The simulation results show that the proposed method has high accuracy and short recognition time for identity feature recognition in wireless visual sensing networks.

Keywords: Convolutional Neural Network · Wireless Visual Sensing Network · Identity Feature Recognition · Image Sequence · Mean Method

1 Introduction

The wireless vision sensor network is composed of multiple wireless nodes integrated with miniature vision sensors, which can transmit the acquired visual perception information to the Sink node (Sink node) through the cooperative way of multiple nodes, and then send it to the application server for subsequent processing and analysis. Wireless vision sensor networks not only have the advantages of traditional wireless sensor networks, such as self-organization, self-healing, flexible configuration, fast coverage and low-cost deployment, but also have the characteristics of the traditional vision application system with rich information, which can support a wider range of intelligent applications, such as traffic monitoring and traffic statistics, assisted living, public behavior analysis and modeling, and virtual reality. However, the resources of unlicensed frequency band suitable for wireless multi-hop transmission of visual information are limited, which limits the scale and performance of the network. Using wireless technology to access idle authorized frequency bands is one of the feasible ways to enhance the performance of wireless visual sensing networks and achieve large-scale applications[1]. The main challenge of wireless visual sensing networks is the randomness of available spectra

L. Yun et al. (Eds.): ADHIP 2023, LNICST 550, pp. 182–198, 2024.
https://doi.org/10.1007/978-3-031-50552-2_12

or channels. Due to the opportunistic access of wireless nodes to the idle authorized spectrum, their underlying link transmission capacity is dynamically changing. Under the traditional design principles of layered network protocols, the transmission of upper level visual information does not adaptively match changes in the transmission capacity of the lower level, thus unable to fully utilize the benefits brought by radio. This requires the use of cross layer design methods to enable upper layer applications to perceive potential transmission opportunities on lower layer links in real-time, in order to enhance the end-to-end service quality of visual information [2].

With the rapid development of information technology, wireless visual sensor network technology has been widely used in finance, e-commerce and other fields. The information of wireless visual sensor network is increasing rapidly, and the security of sensor network information has become a key issue in this field. How to effectively increase the security performance of wireless visual sensor network has become an urgent problem to be solved, and identity feature recognition is a necessary prerequisite to ensure the security of visual sensor network. How to effectively identify the user's identity and protect the security of visual sensor network information has been widely paid attention by experts and scholars in related fields. At the same time, there are also some good adaptive identity feature recognition algorithms [3]. Literature [4] proposed that the difference of footstep induced structural vibration signals in the walking process was used to identify personnel. Based on the energy threshold method, footstep events and non-footstep events were detected. A total of 16 footstep characteristic parameters of a single footstep event for different test personnel were compared and analyzed in the time domain and frequency domain. It is found that the parameter difference under different feature combinations can be used as the basis for identity recognition. In order to verify the effectiveness of the method, support vector machine (SVM) was used as a classification tool. With a test population of 10 people and 500 data samples, 16 foot feature parameters were selected with an average recognition rate of 79.21%. The Pearson correlation coefficient method was used to screen out 10 unrelated foot feature parameters with an average recognition rate of 91%, which was 11.79% higher than the average recognition rate using 16 foot feature parameters, We compared the impact of classification tools on the average recognition rate of 10 selected foot feature parameters under different SVM kernel functions, and found that the highest average recognition rate was 96% under linear kernel functions. The results indicate that an effective combination of foot feature parameters is suitable for identity recognition in small samples. However, the accuracy of the above methods for identity feature recognition is relatively low, resulting in poor recognition performance. Literature [5] uses BGN semi homomorphic encryption algorithm and Shamir secret sharing to design a threshold identity scheme based on biometric identification, which mainly uses BGN homomorphic encryption algorithm on bilinear pairs for data protection, uses a third-party authentication center for secret segmentation, and the server authenticates the user's identity in the ciphertext state to achieve threshold identity authentication. Literature [6] proposes an online detection and automatic identification technology for network video monitoring devices. Stateless scanning technology is used to carry out online detection of network terminal devices, extract BANNER and HTML page information from HTTP header information returned from specific ports of terminal devices, and construct Web identity features of

devices through rough set attribute reduction. The cosine distance is used to calculate the similarity between the Web identity features of online devices and the sample of the known device signature database to realize the detection and identification of online devices. However, it takes a long time for the above two methods to recognize the identity features, resulting in low efficiency.

In response to the problems existing in the above methods, this paper proposes a wireless visual sensing network identity feature recognition method based on convolutional neural networks. Firstly, the architecture of wireless visual sensor networks is analyzed. Then, the background image of the application scene is processed by background subtraction to realize the target detection in the video image sequence. Using convolutional neural network, the Softmax classifier is constructed to classify the human body in the video image sequence of the wireless visual sensor network, and finally realize the recognition of identity features. Simulation experiments have verified that this method can quickly and accurately recognize the identity features of wireless visual sensing networks, laying a certain foundation for the safe operation of wireless visual sensing networks.

2 Wireless Visual Sensor Network Identity Feature Recognition Method

2.1 Analysis of Wireless Vision Sensor Network Architecture

The wireless vision sensor network system preloads part of AI computing power to the edge computing unit through the network architecture of the cloud side, and then completes a certain degree of accurate and lossless video data selection in the sensing front end through the completion of massive unstructured video data. This network architecture scheme can not only effectively reduce the transmission pressure of network bandwidth. It will also save system storage and computing resources, improve the real-time response speed and analysis accuracy of the system, reduce the system delay, and achieve efficient and timely response [7].

The security cloud edge end architecture scheme is shown in Fig. 1. The image and video data are collected through the camera, and then the target face frame in the photo is detected through local edge computing and features are extracted. The recognition results are fed back to the camera for output through feature matching with the cloud end.

2.2 Target Detection in Video Image Sequence of Wireless Visual Sensor Networks

The application scenario of the wireless visual sensing network identity feature recognition method proposed in this article is mainly video surveillance systems. Video surveillance systems have the characteristics of simple background and high real-time requirements. The original data of the dataset used in this article is also videos with relatively simple background, and the background image has not changed much, whether it is the experimental environment or the actual application environment of this article, The

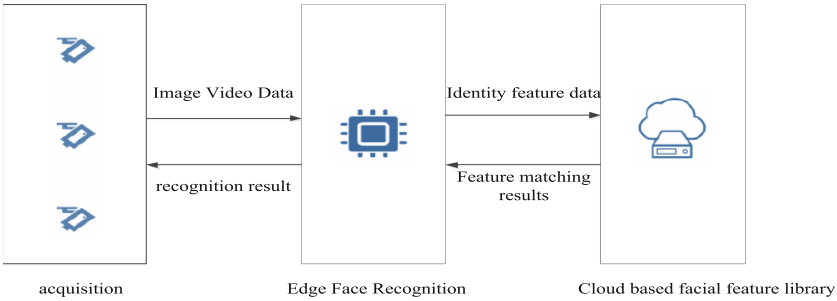


Fig. 1. Security Cloud Edge Architecture

background subtraction method is a good choice due to its performance advantages [8]. Therefore, this article chooses the background subtraction method as the basic algorithm for extracting human targets. For each image in the dataset, a rectangular box is used to mark the area where the human target is located and extract it.

The process of human object extraction is shown in Fig. 2, which can be divided into seven steps, namely image sequence acquisition, background modeling, background difference operation, binarization, open operation, connected domain analysis and target region extraction. Among them, open operation and connected domain analysis can be selected according to the actual situation.

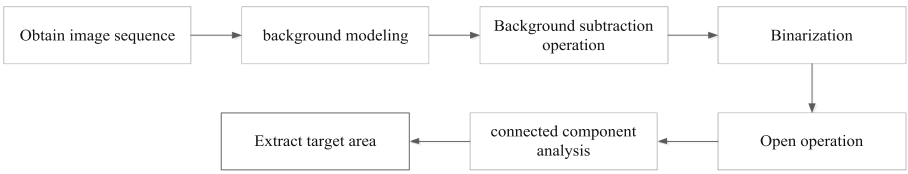


Fig. 2. Process of background difference method

(1) Obtain image sequence.

The data obtained from surveillance videos are all in video format, and the data set adopted in this paper is also in video format. The image sequence corresponding to each video can be obtained by extracting video frames from the wireless visual sensor network. For the background video provided in the data set, frame by frame extraction is adopted in this paper to ensure that higher quality background images can be obtained in the process of background modeling. For the video of wireless vision sensor network containing human objects, extraction is carried out according to the number of 5 frames per second, so that the image sequence can contain the main gestures in the walking process. In terms of quantity, it can also meet the training needs of convolutional neural network, and at the same time avoid excessive redundant data leading to increased computation and reduced training efficiency [9].

(2) Background modeling.

After obtaining a wireless visual sensing network video image sequence through video, the desired human target area is the foreground, while other areas of the image

are processed as backgrounds. The purpose of background modeling is to obtain a background image. Only by obtaining an accurate background image can the desired human target be obtained through background difference calculation. Therefore, the quality of the background image directly determines whether the foreground target can be accurately extracted. At present, there are four commonly used background modeling methods: Mean method, Median method, Single Gaussian model and Mixed Gaussian model [10].

In this paper, according to the actual situation, we choose the Mean method for background modeling. The process of the mean method is: take n consecutive frames of images in an image sequence, calculate the average gray value of the corresponding pixel, and the average value is used as the final gray value of the pixel at the same position in the background image. The reason for choosing this method is that the quality of the background modeling averaging method can meet the requirements of the algorithm, the calculation speed is fast, and it is more in line with the real-time requirements of the video surveillance system.

Taking a background video from the CASIA (Chinese Academy of Sciences Institute of Automation) database Dataset B, Dataset SURF and Dataset CBD as an example, this article uses the mean method to calculate the background modeling process as follows:

After obtaining the image sequence of the background video frame by frame, grayscale each frame to obtain the image sequence P_i and i as the frame numbers. Calculate the average grayscale values of all corresponding points in the image sequence P_i, and establish the background image. The calculation formula is:

$$B(x, y) = \frac{1}{n} \sum_{i=1}^{n} P_i(x, y) \tag{1}$$

where, $B(x, y)$ is the background image, n is the total number of images in image sequence P_i, $P_i(x, y)$ is the grayscale image of frame i, and i is the frame number.

Through calculation, a background image can be obtained for the image sequence of each background video, that is, each person has a grayscale background image in each perspective, and the same background image is used for different dressing or walking posture under the same perspective.

(3) Background subtraction operation

For an image sequence containing human targets, first convert each image in the wireless visual sensing network video image sequence into a grayscale image to obtain image sequence G_i, and then perform a difference operation with the corresponding background image of the image sequence to obtain a new image F_i. The calculation formula is:

$$F_i(x, y) = |G_i(x, y) - B(x, y)| \tag{2}$$

where, F_i is the video image sequence of wireless visual sensor network obtained after background difference calculation, G_i is the video image sequence of grayscale wireless visual sensor network, and B is the background image.

As shown in Fig. 3, this figure is a new image obtained by background error calculation. It can be seen that the perfect foreground target cannot be obtained only by simple background error calculation.

Fig. 3. Image obtained by background subtraction method

(4) binarization

Binarization is a method often used in the process of image processing. Through binarization, an image can be divided into two areas that are either black or white, and the part of interest is usually set as white. The method of setting a threshold for the whole image is called the global threshold method. The image can also be divided into multiple areas, and each area sets a threshold, which is called the local threshold method.

In this paper, the global threshold method is used to set a global threshold T. If the gray value of a pixel is greater than T, the color of the pixel is set to white. If the gray value of a pixel is less than T, the color of the pixel is set to black, and the binary image sequence R_i is obtained. The image more accurately selects the area where the foreground target is located and displays it as white. The calculation formula is:

$$R_i(x, y) = \begin{cases} 1F_i(x, y) \geq T \\ 0F_i(x, y) < T \end{cases} \tag{3}$$

After many experiments, for Dataset B, Dataset SURF and Dataset CBD of CASIA database, threshold T was set to 40, which had the most ideal effect. Figure 4 shows the image after binarization. It can be seen that some noises in the figure need further processing, and some details of the human body are also missing.

(5) Open operation

After binarization processing, the image still cannot fully meet the requirements, because there will be foreground empty points and background noise points, especially the background noise points have a great impact on the subsequent work of this paper, so it is necessary to carry out morphological processing on the binary image obtained before.

Fig. 4. Binary Diagram

Morphological processing of image is to improve the quality of binary image by logical operation between structural elements and binary image. The two most basic operations of morphological processing are corrosion and expansion. Corrosion can eliminate noise points smaller than structural elements, and expansion can fill the holes in the target, but both corrosion and expansion will obviously change the area of the target region.

The combination of open and closed operations through corrosion and expansion solves this problem. In this paper, the binary image is opened, that is, the erosion operation is carried out first, and then the expansion operation is carried out. The main purpose is to remove some small noises in the binary image.

First, the corrosion operation is carried out, and the set obtained by etching the binary image R_i through the structural element S is the set of the origin position of S when the structural element S is completely included in the binary image R_i. The corrosion calculation is as follows:

$$D_i = R_i \odot S \tag{4}$$

where, D_i is the image obtained after corrosion operation, R_i is the binary graph, and S is the structural element.

After the expansion operation, the set obtained by the expansion of the binary image D_i through the structural element S is the set of the origin position of S when the displacement of S' intersects with at least one non-zero element in the binary image D_i. The expansion calculation formula is:

$$E_i = D_i \oplus S \tag{5}$$

In the equation, E_i represents the image after expansion operation.

Figure 5 (a) shows the image after corrosion operation on Fig. 4, and Fig. 5 (b) shows the image after expansion operation.

(6) Connected domain analysis.

The open operation can remove the noise with relatively small area, but the noise with relatively large area can be removed by connected domain analysis. Connected domain refers to the area formed by the connection of points with adjacent positions and equal pixel values in the image, and connected domain analysis refers to marking

(a) Corrosion

(b) Expansion

Fig. 5. Corrosion and Expansion of Binary Graph

the white area in the binary image so that each single connected area has a unique mark, so that geometric parameters such as contour, centroid and external rectangle of these blocks can be obtained further. In this paper, pixel labeling method, which is widely used in connected domain analysis, is used.

Scan all pixels in the binary graph and assign a unique label to the set of pixels located in the same connected region. During the scanning process, there may be multiple small connected regions assigned different labels, but these small connected regions belong to a larger connected region. Write these small connected region labels into equivalent pairs and record their equality relationship.

Merges equal connected domains into one connected domain and assigns a new tag. The area of each connected domain is calculated, and the connected domain whose area is less than a certain threshold is regarded as noise for removal. Through many experiments, it is found that for the data set used in this paper, the effect of setting the threshold as 125 pixels is ideal.

The image obtained after connected domain processing is shown in Fig. 6. It can be seen that the noise with a relatively large area is also removed. Although some details are lost in the foreground area, it can meet the need of extracting the human body target in the video image sequence of wireless visual sensor network.

Fig. 6. Connected Domain Analysis

(7) Extract target area.

After background subtraction, binarization, open operation, and connected domain analysis, the background noise of the color image containing the target character is eliminated, resulting in a binary image containing the complete target, where the white area is the target area. If the white part is used as the selection area to extract the target, as shown in Fig. 7 (a), there will be some loss in the details, especially in the facial details, which will have a certain negative impact on the accuracy of the recognition results. Therefore, the method adopted in this article is to calculate the outer rectangle of the target character area through the outline of the white area, and use the internal area contained in the rectangle as the selection area of the target character, Through this selection area, a rectangular area can be captured from the original color image, which includes both the human target and some background images, as shown in Fig. 7 (b). The rectangular area is used as the final human target to be extracted.

2.3 Recognition Model Based on Convolutional Neural Network

Convolutional neural networks (CNN) have evolved from traditional neural networks, with the main difference being that the feature extractor of convolutional neural networks is mainly composed of convolutional feature extractors, while traditional neural networks are mainly composed of fully connected layers. The structure of CNN network used in this paper includes three main parts: convolution layer, activation layer and pooling layer. Among them, the convolutional layer is one of the core components in CNN, which extracts the local features of the input image through the convolution operation. The convolutional layer consists of multiple kernels, each of which multiplies element-wise with a local region of the input image and sums the results to obtain an element of the output feature map. The activation layer introduces nonlinear transformation to increase the expression ability of the network. Commonly used activation functions include ReLU

(a) Human targets

(b) Rectangular area

Fig. 7. Human target area

(Rectified Linear Unit), sigmoid, and tanh. The activation layer usually follows the convolutional layer and performs element-wise activation function computation on the output of the convolutional layer. The pooling layer reduces the size of the feature map by downsampling operation while retaining important feature information, and maps the pooling results to the output layer to realize the identification of identity features. In this paper, convolutional neural networks are used to construct Softmax classifier to classify human body targets in video image sequences of wireless visual sensor networks, so as to identify identity features.

2.3.1 Convolutional Layer

In the traditional fully connected network, each node will connect all nodes in the upper layer, which will lead to excessive parameters, difficulty in model training, and overfitting. In convolutional neural networks, the convolutional layer implements local

connections, which can effectively reduce the number of model parameters. Each neuron in the convolution layer is only connected to the local Receptive field of the previous layer. The size of the local Receptive field depends on the size of the convolutional nucleus. In CNN convolution, two-dimensional convolution is defined as follows:

$$s(i,j) = (X \times W)(i,j) \tag{6}$$

It can be seen from the above equation that the convolution result of the human object image in the video of wireless visual sensor network is the result of multiplying the local region of the image and the elements of each position of the convolution kernel matrix, and then adding. It is assumed that the size of the feature mapping image input at layer l is $W \times H \times D$, where W and H are the width and height of the human object identity feature image in the two-dimensional wireless visual sensor network video, and D determines the depth of the feature image. Parameter P, the number of 0 elements filled around the image, is used to adjust the size of the output identity feature map.If the size of the two-dimensional convolution kernel is $W_c \times H_c$, the number of output channels set is k, and the step size is S, then the size of the feature map obtained after convolution is $W' \times H' \times D'$. We can control the size of the feature image after convolution by using the convolution kernel size, step size S, and zero padding P. For example, if the convolution kernel size is set to 3, P is set to 1, and step size is set to 1, the feature image remains the same size after passing through the convolution layer. If the size of the convolution kernel is set to 2, P is set to 0, and the step size is set to 2, the feature image will be reduced to 1/4 of the original to achieve the effect of Downsampling.

2.3.2 Activation Layer

Convolutional layers are essentially linear, but for sample data that needs to be learned, their distribution may not necessarily be linearly separable. In order to learn the nonlinear part, an activation layer is usually connected behind the convolutional layer. The activation function in the activation layer transforms the data nonlinearly, which is the key for neural networks to solve nonlinear problems. Several common activation function are sigmoid, tanh, Re LU.

The sigmod activation function has two main disadvantages: (1) It is easy to be saturated. When the input is very large or very small, the gradient is approximately 0, which will lead to the gradient dispersion in the back propagation. (2) The non zero mean output of sigmoid will affect the output of the back layer, thereby affecting the update of the gradient: if the inputs of the back layer neurons are all positive, the local gradient obtained is positive. During the backpropagation process, the parameters will always update in the positive direction, and vice versa, the parameters will always update in the negative direction, resulting in slower convergence speed.

tanh activation function compresses the output to the range of -1 ~ 1, and its output basically follows the mean value distribution of 0, but the problem of gradient saturation still exists in tanh function.

The Re LU activation function does not require exponential calculation, and its computational complexity is relatively low. It only inhibits negative input values, and its output is sparse. Re LU solves the problem of gradient saturation in sigmod function. In this paper, Re LU is selected as the activation function, and its expression is as follows:

$$f(x) = \max(x, 0) \tag{7}$$

2.3.3 Pooling Layer

In the deep convolutional neural network, a huge amount of parameters will be generated with the deepening of the network, and the hardware conditions limit the infinite increase of parameters in the convolutional neural network, which requires the control of the number of parameters in the network. In addition, the image is static, and the same feature may apply to different areas in the image, which indicates that there must be redundancy in the original parameters. The essence of the pooling layer is to aggregate statistics of features of different locations and compress the input wireless visual sensor network identity feature map, that is, to conduct downsampling of the feature map.Pooling layer can effectively reduce the parameters required by subsequent layers, reduce the possibility of overfitting, and make the convolutional neural network translation invariant, that is, when the pixels in the feature map have a small displacement in the neighborhood, the pooling layer can keep the output unchanged, which enhances the robustness of the network. The commonly used pooling layer downsampling methods include mean pooling and max pooling. Pooling of regional average values can preserve the characteristics of the overall data and highlight the background information. Pooling the maximum value of a region can better preserve the features on the texture. The pooling methods used in this article are regional maximum pooling and mean pooling.

2.3.4 Softmax Classifier

Softmax is a multinomial logic regression model. Logistic regression model belongs to log-linear model and is also a probabilistic model, which is generally used for binary classification problems. Multinomial logistic regression model is the extension of logistic regression and can be used to solve multi-classification problems. The calculation process of Softmax is as follows:

$$P(Y|x) = \frac{\exp(w_k \times x)}{k = 1 \sum^{K-1} \exp(w_k \times x)} \tag{8}$$

In the equation, w is the desired parameter model, x is the input vector, and $P(Y|x)$ is the probability value of predicting the category of x as category Y.

Input human target identity features from wireless visual sensing network videos as samples into Softmax classifier to obtain classification results for identifying wireless visual sensing network identity features:

$$\theta_t = -\eta \times g_t \tag{9}$$

where, η is the global learning rate initially set, and g_t is the gradient.

3 Experimental Analysis

3.1 Preparation for Experiment

In order to verify the effectiveness of the wireless visual sensing network identity feature recognition method based on convolutional neural networks proposed in this article in practical applications, the accuracy and recognition time of the wireless visual sensing network identity feature recognition were selected as experimental indicators, and the methods of reference [4] and reference [5] were used as comparative methods for experimental testing.

This article also used GPU for acceleration during the experimental process. The GPU brand model is NVIDIA Tesla K40c, which has 2880 Cuda cores and 12G graphics memory. Its parallel computing and storage capabilities can well meet the needs of convolutional neural network training. Using M3001 robot module camera, capturing image size 1920 × 1080, which can set parameters such as automatic exposure, automatic white balance, color correction, brightness, contrast, saturation, sharpness, etc. It supports a variety of protocols such as common TCP/IP, ICMP, HTTP, FTP, DHCP, DNS, DDNS, RTP, RTSP, etc. It can access the network through RJ45 10M/100M adaptive Ethernet port, and transmit the photos taken to the local processing unit. The camera and processor are shown in Fig. 8.

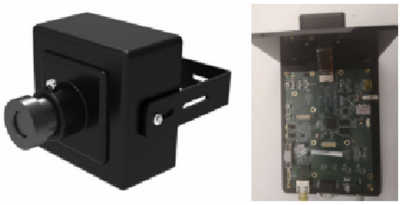


Fig. 8. Camera and processor

The Hiss 3516DV300 neural network processing chip used in the terminal computing unit in this paper NNIE, short for Neural Network Inference Engine, is a hardware unit specializing in accelerated processing of neural networks, especially deep learning convolutional neural networks in So C of Hesis Media. It supports most existing public networks. For example, classification networks such as Alex Net, VGG16, Res Net18 and Res Net50, detection networks such as Faster R-CNN, YOLO and SSD, and scene segmentation networks such as Seg Net and FCN.

3.2 Data Set Construction and Training

(1) Building a dataset.

Obtaining multi view video data of all objects to be identified through the reasonable arrangement of monitoring equipment. From a practical application perspective, it is reasonable to have no more than three views. Then, the video data is converted into an image sequence and target detection is performed using background subtraction to

extract the rectangular area where the human target is located as the dataset. If image data from three different perspectives is obtained, the dataset is divided into three subsets according to three different perspectives, and each subset is further divided into a training set, a validation set, a single perspective testing set, and a multi perspective testing set.
(2) Training model.

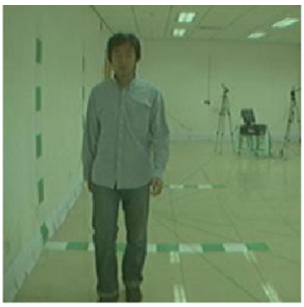
The multi-network identity model is trained and tested on the multi-view data set according to the training and testing process. Firstly, an appropriate convolutional neural network should be built as the subnet for training each perspective, and the training set and verification set of each subset should be input into the corresponding subnet for training, so as to obtain the model of each subnet. During the training process, each subnet should be adjusted independently, so that each subnet can obtain the best recognition result on its corresponding data set. The accuracy of each subnet is tested by the single view test set, and the weight of each subnet is calculated according to the weight calculation formula, and the weighted fusion identity model is obtained. The final identity model is tested through the multi-view test set to verify its comprehensive performance and further optimize the model.
(3) Identity recognition.

Input the multi view image of the object to be recognized into the corresponding subnet for recognition, obtain the recognition results of each subnet, and then calculate the final identity recognition result through weighting.

CASIA database, created by the Institute of Automation of the Chinese Academy of Sciences, consists of three data sets, of which Dataset A is a small-scale database with 20 people, each with three shooting angles (0°, 45°, 90°), a total of 240 image sequences, and the acquisition environment is outdoor. Select Dataset B, Dataset SURF and Dataset CBD of the CASIA database as the dataset for this article. Each subject has eleven perspectives, and the images from different perspectives are shown in Fig. 9.

From the different perspectives mentioned above, the methods of this article, reference [4], and reference [5] were used to compare and analyze the accuracy of identity feature recognition in wireless visual sensing networks. The comparison results are shown in Table 1.

According to Table 1, the accuracy of the method used in this paper for identity feature recognition in wireless visual sensing networks can reach up to 99.5%. The accuracy of the method used in reference [4] for identity feature recognition in wireless visual sensing networks can reach up to 84.1%. The accuracy of the method used in reference [4] for identity feature recognition in wireless visual sensing networks can reach up to 70.5%. The accuracy of the method used in this paper for identity feature recognition in wireless visual sensing networks is the highest, and the recognition effect is the best.



(a) 0 ° viewing angle

(b) 90 ° viewing angle

(c) 180 ° viewing angle

Fig. 9. Images from Different Perspectives

Using the methods of this article, reference [4], and reference [5], a comparative analysis was conducted on the time required for identity feature recognition in wireless visual sensing networks. The comparison results are shown in Table 2.

As can be seen from Table 2, the time used for the identification of wireless visual sensor network identity features by the method in this paper is within 6.2s, the time used for the identification of wireless visual sensor network identity features by the method in reference [4] is within 16.5s, and the time used for the identification of wireless visual sensor network identity features by the method in reference [4] is within 26.4s. The method presented in this paper has the shortest time and the highest recognition efficiency for wireless visual sensor network identity feature recognition.

Table 1. Comparison results of identification accuracy of wireless visual sensor networks /%

Number of experiments/times	Textual method	Method of reference [4]	Method of reference [5]
10	94.2	80.1	64.5
20	94.9	80.6	65.2
30	95.6	81.2	66.2
40	96.8	81.6	67.4
50	97.5	82.3	68.1
60	98.2	83.5	69.2
70	99.5	84.1	70.5

Table 2. Comparison results of identity feature recognition time in wireless visual sensor network /s

Number of experiments/times	Textual method	Method of reference [4]	Method of reference [5]
10	5.2	15.2	22.2
20	5.3	15.6	22.6
30	5.4	15.7	23.4
40	5.5	15.9	23.5
50	5.5	15.9	24.9
60	5.8	16.2	25.8
70	6.2	16.5	26.4

4 Conclusion

In recent years, with the gradual development and widespread application of wireless visual sensing networks, people's requirements for information security have become increasingly high. Identity verification, as one of the important means to ensure information security, can advantageously ensure that system users have corresponding application rights. Therefore, studying adaptive recognition algorithms for identity features is of great significance and has become a key research topic for relevant scholars, receiving increasingly widespread attention. As a key issue in the development of network security, identity recognition technology has received increasing attention from scholars. Commonly used identity recognition technologies mainly include face recognition, iris recognition, fingerprint recognition, and related algorithm research has achieved certain results.However, in the research of visual optimization identification, due to the influence of posture, light, expression and other factors, it is impossible to accurately identify the identity in the wireless visual sensor network under the uncontrollable environment. In traditional visual identity recognition, uncontrollable factors need to be transformed

into controllable and stable characteristic factors in an uncontrollable environment with relatively complex node distribution before identity recognition. The conversion process leads to long recognition time and low efficiency. In this paper, an identity feature recognition method based on convolutional neural networks is proposed for wireless visual sensor networks, and the experimental verification shows that the proposed method has good identification effect and high recognition efficiency. Future research should also focus on the use of identity feature recognition techniques for privacy protection and security. This paper explores how to fully consider the needs of user privacy and information security while ensuring high accuracy.

References

1. Zheng, Y.L., Burns, J.H., Wang, R.F., et al.: Identity recognition and the invasion of exotic plant. Flora Morphol. Distrib. Funct. Ecol. Plants **280**, 151828 (2021)
2. Yang, W.-H., Dai, D.-Q.: Two-dimensional maximum margin feature extraction for face recognition. IEEE Trans. Syst. Man Cybern. Part B (Cybern.) **39**(4), 1002–1012 (2009). https://doi.org/10.1109/TSMCB.2008.2010715
3. Dongbo, L.I., Huang, L.: Reweighted sparse principal component analysis algorithm and its application in face recognition. J. Comput. App. **40**(3), 717–722 (2020)
4. Hou, X., Li, R., Zhang, Y.: Personnel characteristics identification based on foot induced structural vibration. J. Vibr. Shock **41**(23), 241-248,292 (2002)
5. Yao, L., Guo, S., Yang, X.: Threshold identity authentication scheme based on biometrics. App. Res. Comput. **39**(4), 1224–1227 (2022)
6. Ding, W.: Network video surveillance equipment identification based on Web identity characteristics. J. Shenyang Univ. Technol. **42**(4), 427–431 (2020)
7. Zhao, D., Lu, Y., Liu, X., et al.: Design of emergency UAV network identity authentication protocol based on Beidou. MATEC Web Conf. **336**, 04004 (2021)
8. Yichao, Z., Ziwen, S.: Identity authentication for smart phones based on an optimized convolutional deep belief network. Laser Optoelect. Progr. **57**(8), 081009 (2020)
9. Tian, Z., Yan, B., Guo, Q., et al.: Feasibility of identity authentication for IoT based on blockchain. Proc. Comput. Sci. **174**, 328–332 (2020)
10. Liu, Y.N., Lv, S.Z., Xie, M., et al.: Dynamic anonymous identity authentication (DAIA) scheme for VANET. Int. J. Commun. Syst. **32**(5), e3892.1-e3892.13 (2019)