Deep Learning Based Image Recognition for 5G Smart IoT Applications

Jing Li
School of Education, Xizang Minzu University

Xinfang li (ningyuwen@nwu.edu.cn)
School of Education, Xizang Minzu University

Yuwen Ning
Information Technology Center, the Fourth Military Medical University

---

Research

Keywords: Deep Learning, Convolutional Neural Network (CNN), Internet of Things Image Recognition, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM)

Posted Date: November 19th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-108738/v1

Deep Learning Based Image Recognition for 5G Smart IoT Applications

Jing Li[1,a], Xinfang li[1,b*] and Yuwen Ning[2,c]

[1]*School of Education, Xizang Minzu University, Xianyang 712082, Shaanxi, China*

[2]*Information Technology Center, the Fourth Military Medical University, Xi'an 710032, Shaanxi, China*

[a]*jj2lj314@163.com* [b]*ningyuwen@nwu.edu.cn* [c]*ningyuwen2020@126.com*

**Corresponding author*

Abstract: *With the advent of the 5G era，the development of massive data learning algorithms and in-depth research on neural networks, deep learning methods are widely used in image recognition tasks. However, there is currently a lack of methods for identifying and classifying efficiently Internet of Things (IoT) images. This paper develops an IoT image recognition system based on deep learning, i.e., uses convolutional neural networks (CNN) to construct image recognition algorithms, and uses principal component analysis (PCA) and linear discriminant analysis (LDA) to extract image features, respectively. The effectiveness of the two PCA and LDA image recognition methods is verified through experiments. And when the image feature dimension is 25, the best image recognition effect can be obtained. The main classifier used for image recognition in the IoT is the support vector machine (SVM), and the SVM and CNN are trained by using the database of this paper. At the same time, the effectiveness of the two for image recognition is checked, and then the trained classifier is used for image recognition. It is found that a CNN and SVM-based secondary classification IoT image recognition method improves the accuracy of image recognition. The secondary classification method combines the characteristics of the SVM and CNN image recognition methods, and the accuracy of the image recognition method is verified to provide an effective improvement through experimental verification.*

Keywords: Deep Learning, Convolutional Neural Network (CNN), Internet of Things Image Recognition, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM)

1.Introduction

With the rapid development of Internet technology and the popularization of IoT imaging equipment, massive image data has accumulated on the network, and processing these image data has become a hot research topic nowadays. These images contain information on all aspects of human production and life, and have spawned a broad market for image recognition and application. Internet technology, especially the popularity of social media, has accelerated the transfer of these images in different fields and groups. Driven by the increasing informatizing

demand and advanced Internet of Things (IoT) technologies, 10's of IoT billions of devices that will not be intervened by people are envisioned connect to networks in 2020s. To address the emerging market, ITU initialized the solicitation of proposals of 5G standards in 2015 and identified massive machine type communication (mMTC) as one of the three main application scenarios with extreme or even contradictory performance requirements that candidate 5G standards have to address as. At the same time, the substantial increase in computer computing power, especially the parallel computing of the GPU, has accelerated the speed of image processing, allowing many difficult tasks to be completed within an acceptable time.

The IoT will deal with various problems based on big data in 5G, so its application is increasingly dependent on visual data and deep learning technology, so finding a suitable method to analyze the image data of IoT systems faces huge challenges [1]. In [2], the author proposed a medical sports rehabilitation image segmentation algorithm based on convolutional neural network hierarchical features. The algorithm accelerates the network convergence speed, shortens the training time, and greatly improves the accuracy of the medical sports rehabilitation image segmentation algorithm. In [3], the author designed an automatic recognition system for abnormal images of tourist attractions based on the Internet of Things vision. The system can accurately and effectively identify abnormal images of tourist attractions. In [4], the author proposes an efficient IoT PIM system that can calculate real image recognition applications. The proposed architecture is able to process 6 times more frames per second than the baseline, while increasing energy efficiency by 30 times. In [5], the author proposes a multi-mode deep learning (MMDL) method that integrates heterogeneous visual features by treating each type of visual feature as a way to optimize image recognition in the Internet of Things. In [6], the author implemented a real-time face recognition system in real-time through a high-level description language (such as python), and provided an innovative solution for automatic real-time face recognition for video. In [7], the author proposed a system that demonstrates that the Internet of Things can be applied in healthcare and daily life. The system is designed to help blind people to match clothing correctly, which is achieved using image and pattern recognition. In [8], the author proposed a stereo matching algorithm for intelligent video surveillance to optimize the accuracy of the surveillance system. It not only optimizes the performance of the smart home system, but also improves the safety factor. In [9, 10], the author experimentally verified the method of image recognition on different benchmark data sets, and concluded that the author's method can construct a network that supports multiple trade-offs between accuracy and computational cost. In [11], the author proposed a method of personal identity verification based on finger veins by exploring the direction and magnitude of competition from finger vein images. This method is superior to the method based on the latest directional coding (OC) and can be used for finger vein image enhancement. In [12, 13], the technique proposed by the author can achieve a high average correction rate of 94% and a calculation time of 44 frames per second, which can be effectively used for night vehicle detection and counting.

Despite so much research, at present, deep neural networks are still like a black box model, and it is difficult to understand the details. Therefore, most of the research on this model is empirical and experimentally driven. Although there are still many undetermined factors, so far deep neural networks are still the most effective and accurate mode for solving computer vision problems. How to apply deep learning and image recognition methods to analyze and mine useful information from these massive image data to solve practical problems in various fields is a

problem faced by universities, research institutes and technology companies. In addition, 5G IoT needs to provide accurate calculation for a large number of devices to support the real-time processing of massive data by terminal devices. Performance requirements of IoT devices in 5G massive IoT tend to be heterogeneous since they need to deliver increasingly diverse applications. Therefore, combined with the characteristics of 5G IoT, deep learning is applied image recognition.

Artificial intelligence (AI) will be widely used in 5G. And deep learning technologies are used to realize intelligent image recognition. Deep learning methods use multiple processing layers to learn the hierarchical representation of data, and have produced the latest technological achievements in 5G IoT, and various model designs and methods have been developed [14]. In [15], the author achieved the latest detection performance on the famous FDDB face detection benchmark evaluation. In [16], the author describes a hand-eye coordination method based on deep learning for grabbing robots from monocular images. Transmission experiments also show that data from different robots can be combined to learn a more reliable and effective mastering method. Deep learning has led to a paradigm shift in artificial intelligence, which has an increasing impact on chemical physics [17]. In [18], the author provides a segmentation model, which designs an image segmentation neural network based on a deep residual network, and uses guidance filters to extract buildings in remote sensing images. The proposed method shows excellent performance in extracting buildings from various objects in urban areas. In [19], the author first introduced IoT deep learning into the edge computing environment, and optimized the performance of IoT deep learning applications through edge computing. In [20], the author used an approximate message passing (LDAMP) network based on noise reduction. The neural network can learn the channel structure and estimate the channel from a large amount of training data. According to the analysis and simulation results, the LDAMP neural network is clearly superior to the latest algorithm based on compressed sensing. In [21], the authors combined a convolutional and circular architecture to train a deep network to predict the outcome of colorectal cancer based on images of tumor tissue samples. The most advanced deep learning technology can extract more prognostic information from the histomorphology of colorectal cancer [22]. The application of deep learning in robotics will lead to very specific problems and research problems, and computer vision and machine learning usually cannot solve these problems and research problems [23, 24]. In [25], the author constructed two deep learning models, DeepARG-SS and DeepARG-LS, which considered the difference matrix created using all new classes of antibiotic resistance genes (ARG). In [26], the author demonstrated the application of deep learning and convolutional autoencoder networks. The technology restores real-time 128×128 pixel video from a single-pixel camera sample at a rate of 2% at a compression rate of 2% [27].

This paper designs an 5G IoT image recognition system based on deep learning, which uses principal component analysis (PCA) and linear discriminant analysis (LDA) to extract IoT image features. The PCA method extracts image features through image transformation, and obtains relatively good experimental results. The LDA method uses the existing image information for image feature extraction. After projection, the image samples have a relatively large degree of dispersion, which is beneficial to the image recognition of the Internet of Things. In this paper, the classifier mainly used for image recognition in the Internet of Things is SVM. By using the database of this article to train SVM and CNN, it also checks the effectiveness of both for image

recognition, and then uses the trained classifier for image recognition.

The rest of the paper is organized as follows. Section 2 introduces the basics of IoT image recognition technology based on deep learning is presented in. Section 3 proposes the IoT image recognition system construction and scheme design in details. Simulation results are presented in Section 4 to justify the performance of our design, followed by Section 5 to conclude the paper.

2. IoT Image Recognition Technology Based on Deep Learning

Fig. 1 illustrates the architecture of the proposed system consisting of three functional layers: the IoT layer, the Fog layer, and the cloud layer. The different layers are integrated in order to support to the IoT image recognition.

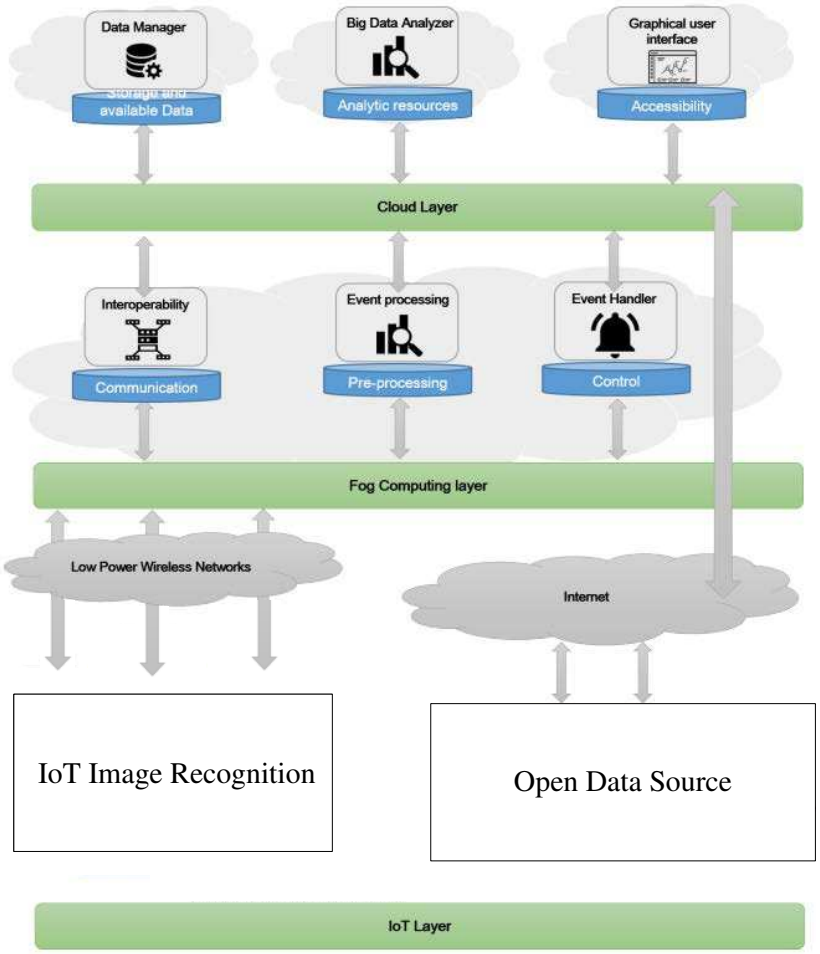


Figure 1 The high-level architecture of an IoT in Image Recognition

The IoT layer is the foundation of the entire system. It obtains data from different heterogeneous sources through different wireless networks. The fog layer enables the interoperability of heterogeneous data sources and the preprocessing and knowledge generation of heterogeneous data sources through fog computing methods. In the fog computing, a group of edge devices are placed between the sensing device and the cloud in order to extend the cloud resources at the edge of the network through the network to achieve the purpose of improving performance, storage, processing capabilities, etc., close to the terminal device. Due to the maturity and scalability of cloud computing, it is currently the preferred paradigm to undertake large storage, computationally intensive data processing and analysis tasks, because they allow

services to grow and shrink online without degradation, which greatly reduces the cost of smart devices burden.

2.1 Theoretical Model of Convolutional Neural Network (CNN)

(1) Convolutional layer

Assuming that the number of filters is 100, the final weight parameter is 10,000. It can be seen that, compared with the fully connected structure, the convolutional layer reduces the amount of parameters from two aspects of local connection and weight sharing. The convolution process can be written as:

$$x^{(l+1)} = w^{l+1}x^{l} + b^{l+1} \tag{1}$$

where x^l is the output of layer l; $x^{(l+1)}$ is the output of layer $l+1$; w^{l+1} is the convolution kernel of $l+1$; b^{l+1} is the offset item of $l+1$. It is assumed that the input data and the convolution kernel have the same width and height, which are denoted by w and f, respectively, and the step size in both directions is s. The number of filled pixels at the boundary is p, and the space size of the output feature map is $\left(\frac{w-f+2p}{s}+1\right)\times\left(\frac{w-f+2p}{s}+1\right)$. For example, the input is 7×7 size, the convolution kernel is 3×3 size, the boundary is not filled, the output is 5×5 size when the step size is 1, and the output is 3×3 size when the step size is 2.

(2) Pooling layer

The convolutional layer controls the number of training parameters, but the size of the feature map generally does not change significantly, and the pooling layer is often used to reduce the size of the feature map. The pooling layer is inserted in the middle of the convolutional layer, which is modeled as a visual system for downsampling. The most commonly used pooling strategy is max pooling.

Assuming that the input data size of the pooling layer is $w_1 \times h_1 \times d_1$, the size of the sliding window is f×f, and the moving step size is s, the output feature map satisfies: $w_2 = \frac{w_1 - f}{s} + 1, \quad h_2 = \frac{h_1 - f}{s} + 1, \quad d_2 = d_1$. The most commonly used parameters are f=3, s=2, f=2, and s=2. The feature map output by the latter group of parameters is exactly half the size of the input.

(3) Loss function

Let the sample set contain N samples, denoted as $(X,Y) = (x_i, y_i), i \in [1,N]$, y_i is the true value of the i-th sample, $y_i = g(x_i)$ is the predicted value of the i-th sample, and g is the function representation of the model. Then the total loss function is:

$$L = \sum_{i=1}^{N} l(y_i, y_i) \tag{2}$$

In the formula, $l(y_i, y_i)$ is the value of the loss function between the true value y_i and the predicted value y_i. The Softmax classifier is most commonly used in classification tasks. It first

converts the model predictions into category probabilities, and then uses the cross-entropy loss function. Let the output of the model be $f \in R^k$, k represents the number of categories, f_j represents the output of the model corresponding to the j-th category, and $q \in R^k$ is the probability of the category corresponding to f, which is converted by the Softmax function of the following formula:

$$q_j = \frac{e^{f_j}}{\sum_j e^{f_j}} \tag{3}$$

The true class label is p, which is a one-hot vector [0,...,1,...,0], and the prediction result is recorded as q.

(4) Weight initialization

Suppose that a linear activation function f is used, which is derivable at 0. The single layer convolution is calculated as follows:

$$y = w_1 x_1 + \cdots + w_{n_i} x_{n_i} + b \tag{4}$$

In the formula, n_i- the input data dimension of layer i. When the mean value of the input and weight of the network is 0, the variance has the following relationship:

$$Var(w_i x_i) = E[w_i]^2 Var(x_i) + E[x_i]^2 Var(w_i) + Var(x_i)Var(w_i) = Var(x_i)Var(w_i)$$

In order to ensure that the input and output variances of the network are consistent, the weights need to satisfy $Var(w_i) = \frac{1}{n_i}$. Similarly, to ensure that the variances of the input and output gradients during back propagation are consistent, the weights also need to satisfy $Var(w_i) = \frac{1}{n_{i+1}}$. For measurement considerations, the final weight variance meets $Var(w_i) = \frac{2}{n_i + n_{i+1}}$. Therefore, the "Xavier" initialization uses the following uniform distribution:

$$W \sim U\left[-\frac{\sqrt{6}}{\sqrt{n_i + n_{i+1}}}, \frac{\sqrt{6}}{\sqrt{n_i + n_{i+1}}}\right] \tag{5}$$

Assuming that the second half of neurons after ReLU is 0, and the other half of neurons are activated, so in order to keep the variance of output and input unchanged, only need to satisfy $Var(w_i) = \frac{2}{n_i}$. The method uses Gaussian distribution to initialize the weights, and the mean is 0.

2.2 Image Preprocessing and Feature Extraction Methods

(1) Principal component analysis (PCA)

PCA is an image feature extraction method based on statistical feature learning. Principal component analysis is also called K-L transformation (Karhunen-Loeve transformation). The main principle is to use linear transformation to collect the features of the image to be recognized, and the transformed image can retain the effective information of the original image to the maximum extent. Suppose there is a matrix $X = [x_1, x_2, \ldots x_k]$ composed of K N-dimensional vectors, and the defined X covariance matrix is:

$$\Sigma_x = \frac{1}{k}\sum_{i=1}^{k}(x_i - \mu_x)*(x_i - \mu_x)^T \tag{6}$$

where μ_x is the mean vector of X: $\mu_x = \frac{1}{k}\sum_{i=1}^{k} x_i$. Among them, X represents an $N*K$-dimensional matrix, μ_x represents the average of K vectors, and is also an N-dimensional vector, so $\sum x$ represents an $N*N$ real symmetric matrix.

(2) Linear discriminant analysis (LDA)

Suppose there are two types of sample data, namely red data points and blue data points, which are two-dimensional data features. During the recognition process, these image data need to be projected onto a one-dimensional line. For the final projection result requirements, the projection of data points of the same color should be as close as possible, and the center distance between data points of different colors should be as large as possible, so as to achieve the best classification effect.

Similarities between LDA and PCA:

1) Both can perform dimensionality reduction on the image data to be recognized.

2) Both applied the idea of matrix feature decomposition to the process of image dimensionality reduction.

Differences between LDA and PCA:

1) PCA is an unsupervised image dimensionality reduction method, and LDA is a supervised image dimensionality reduction method.

2) In the projection process of LDA, a projection direction that can achieve the best image classification performance will be selected. However, PCA selects the direction with the largest variance as the projection direction of the sample point.

3) LDA can not only reduce the dimensionality of the image, but also can be used for image classification.

2.3 IoT Image Classification Method Based on Support Vector Machine (SVM)

Support vector machine SVM model principle: make the distance between all data points to be measured and the hyperplane greater than the specified value. The purpose is to enable all classification points to be located on both sides of the support vector after classification, providing a good basis for subsequent image recognition of the Internet of Things. The function expression of the model is defined as:

$$max\gamma = \frac{y*\left(w^T X + b\right)}{\| w\|_2} s.ty_i\left(W^T x + b\right) = \gamma' i \geq \gamma' (i = 1, 2, \ldots.m) \tag{7}$$

However, in the actual experimental process of image recognition, in order to make the calculation more convenient, the value of the function interval γ' is usually defined as 1, and the improved optimization function can be defined as:

$$max\frac{1}{\|w\|_2}s.ty_i\left(W^T x+b\right)\geq 1(i=1,2,\ldots..m) \tag{8}$$

It can be seen from the above function expression that under the constraint condition $y_{i^*}\left(w^T x+b\right)\geq 1(i=1,2,\cdots,m)$, the expression is maximized. Since the maximization of $\|w\|_2$ is equivalent to the minimization of $\frac{1}{2}\|w\|_2^2$. In this way, the optimization function of SVM is equivalent to:

$$min_2^1\|w\|_2^2 s.ty_i\left(W^T x+b\right)\geq 1(i=1,2,\ldots\ldots m) \tag{9}$$

From the above conclusions, we can see that the main principle of support vector machine SVM is, first, to define a suitable function. Then, through non-linear transformation, the low-dimensional space feature vector of the sample to be measured is mapped to the high-dimensional space, and the mapped samples can be linearly separated. It can be concluded from the classification principle of SVM that it is different from the traditional classification method. Instead, it determines the size by non-linear transformation and converts the samples to be identified to be linearly separable.

According to the specific recognition requirements, a multiple high-order neural network is constructed to achieve the optimal coverage of closed hypersurface measurement samples in high-dimensional space. The following functional expression is a mathematical model of higher-order neurons with double synaptic weights:

$$Y=f\left[\sum_{j=1}^{n}\left[\frac{w_j\left(x_j-w_j^{'}\right)}{\left|w_j\left(x_j-w_j^{'}\right)\right|}\right]^s\left|w_j\left(x_j-w_j^{'}\right)\right|^p-\theta^r\right] \tag{10}$$

In the above function expression, Y is the output of the neuron; f is the excitation function of the neuron, θr is the activation threshold of the neuron; xj is the j-th input data of the input terminal. The meanings expressed by wj and wj' are the direction weight and core weight of the j-th input terminal connected to the neuron; the S parameter determines the sign of the single expression, and P is the power parameter.

3. IoT Image Recognition System Construction and Scheme Design

3.1 IoT Image Data Set

(1) CUB-200-2011 data set

The full name of this data set is Caltech-UCSD Birds-200-2011, which is collected by California Institute of Technology and is an extended data set of CUB-200, which is used for bird species classification. The number of bird categories is 200, and the total number of pictures is 11,788, of which 5,994 are for training and 5,794 are for testing. Each image contains a target frame label, 15 part labels and 312 attributes. The structure of all categories is finally divided into four layers, the total number of nodes is 264, of which 200 are leaf nodes, corresponding to the

original class label, and the number of parent nodes is 64, corresponding to the increased taxonomy of higher-level categories.

(2) CIFAR-100 data set

The data set is mainly something common in life. The data set contains 60,000 pictures, divided into 100 categories, the size of the image is 32×32, each category contains 600 pictures, 500 of which are training data, and the rest are used for testing. These 100 categories are combined into 20 superclasses, so each picture contains two class labels: a coarse class label and a fine class label, which correspond to the superclass and the original category, respectively. The correlation between these categories is relatively weak, such as furniture including beds, chairs, tables, etc. This is based on the concept of humans to summarize the categories, so different categories under the same super category are not strong in common.

(3) Insect data set

There are two sources of pictures for this data set: images taken with a microscope in the laboratory, and images taken with a mobile phone to add a magnifying glass to the port scene. A white or other color paper is put on the bottom when shooting, so the background of the image is relatively simple. Due to the large differences in the appearance of insects at various angles, the insects were photographed from three angles: back, side, and abdomen. The final data includes 7 insects, 35 genera and 14748 pictures. The division of training test samples is to randomly select 80% of each genus as the training set and 20% as the test set.

(4) CUHK03 data set

The data set is taken from 5 different pairs of perspectives and contains images of 1,467 pedestrians. The total number of pictures is 13,164. The data set contains two types: manually cropped and automatically detected human images. Manual cropping is the manual cutting of pedestrian pictures from the video, automatic detection is the method of pedestrian detection, and the quality is slightly worse than the manual cropping version.

3.2 Experimental Environment and Evaluation Criteria

The experimental environment of this paper builds a programming environment based on Google's deep learning framework TensorFlow on a Windows 7 system. The specific software and hardware environment configurations are shown in the following table:

Table 1 Experimental configuration

Lab environment		Configuration
Hardware environment	CPU	Intel(R) Core(TM)i7-7700K 4.20GHz
	GPU	NVIDI A GeForce GTX 1080 Ti 4G
Software Environment	RAM	64G
	System	Windows 7 64
	Programming environment	Tensorflow

In addition, in the training process of the model proposed in this paper, the Batchsize is set to 64, the learning rate is set to 1e-4, the exponential decay rate of moment estimation beta1 is set to 0.90.9, and beta2 is set to 0.999. For the data set used in this article, we use two criteria, TestAcc, to evaluate the effect of the experiment. The definition of test accuracy is:

$$TestAcc = \frac{TestImagesCurrently}{TestImagesAll} \tag{11}$$

In the formula, TestImagesCurrently represents the number of images tested correctly, TestImagesAll represents the total number of test images, and the corresponding average test accuracy rate is:

$$AvgAcc = \frac{1}{n}\sum_{i=1}^{n} TestAcc(i) \tag{12}$$

The definition of training accuracy is:

$$TrainAcc = \frac{TrainImagesCurrently}{TrainImagesAll} \tag{13}$$

In the formula, TrainImagesCurrently represents the number of correctly trained images, TrainImagesAll represents the total number of training images, and the average training accuracy rate is:

$$AvgAcc = \frac{1}{n}\sum_{i=1}^{n} TrainAcc(i) \tag{14}$$

3.3 IoT Image Recognition Scheme Design

In order to more effectively recognize the Internet of Things images, the image recognition scheme designed in this paper includes: image acquisition, image preprocessing, feature extraction and classification recognition are divided into four parts.

(1) Image acquisition. The images used in this experiment were collected from the dataset images in this article.

(2) Image preprocessing. Each type of item in the collected IoT images is manually segmented to complete the establishment of the experimental sample library. First, select the product category of the experiment, segment the image samples of each category from different angles, and finally scale it to a uniform size.

(3) Image feature extraction. By using principal component analysis and linear discriminant analysis, the features of the IoT image are extracted, and the acquired features are used for image recognition in the next step.

(4) Classification and identification. A variety of classifiers and CNNs are used to train the classifiers and CNNs by using different classification decision criteria and samples of IoT images. After the classifier training is completed, the extracted features are input for image recognition, and the classification result of the sample to be recognized is obtained. The block diagram of the Internet of Things recognition process is shown in Figure 2.

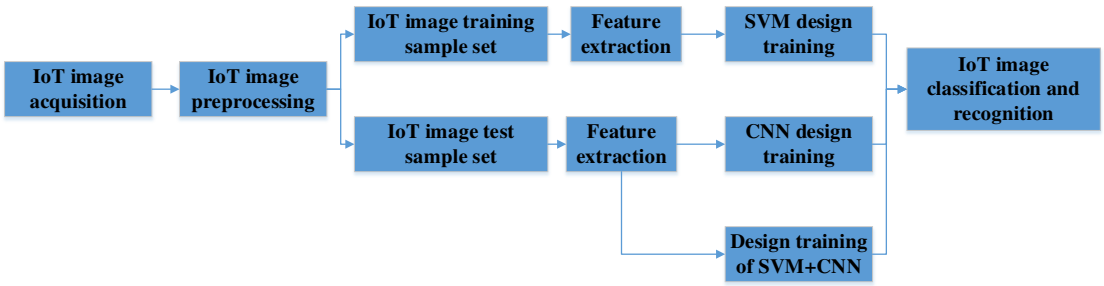


Figure 2 Block diagram of IoT image recognition process

3.4 Design and Implementation of IoT Image Recognition System Based on Deep Learning

The image recognition system designed in this paper mainly includes: image acquisition module, image transmission module, image preprocessing module and recognition module.

First, the sample image is collected to complete the sample image collection process. Next, the preprocessed image is preprocessed, and then the sample image is recognized by the secondary classification image recognition method. The main purpose of this experiment is to complete the recognition of the image. The main recognition process is as follows:

(1) First, perform feature extraction on the original image, obtain the feature subspace and feature vector of each sample, and use the collected image features to train the SVM and CNN respectively. Obtain the relevant classification parameters of the classifier and the trained classifier for the final image recognition.

(2) Read in the image to be recognized and perform image preprocessing on the sample image.

(3) Use the secondary classification image recognition method to perform image recognition on the pre-processed sample images, obtain the image recognition results and perform correlation analysis. The image recognition process of the Internet of Things is shown in Figure 3:

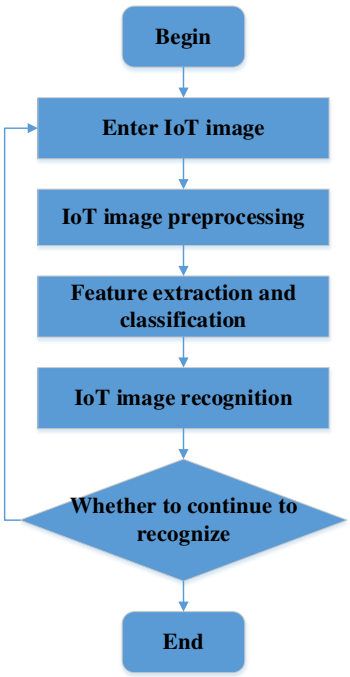


Figure 3 Image recognition flowchart of the Internet of Things image recognition system

4.Results and Discussions

4.1 Influence of Different Parameters of Convolutional Neural Network on the Image Recognition Rate of the Internet of Things

(1) The impact of transformation times on the experimental results

In the CUB-200-2011 and CIFAR-100 databases, the recognition rates corresponding to different transformation times are used (Figure 4), where k (k ∈ [1, 4]) represent variables. It can be clearly seen in the figure that as the number of transformations increases, the recognition rate decreases, and the recognition rate is the highest at random times, because the increase in the number of transformations increases the degree of disturbance and the feature

points will become more inconspicuous. The secondary transformation will increase the degree of difference between different categories, which is more conducive to machine identification.

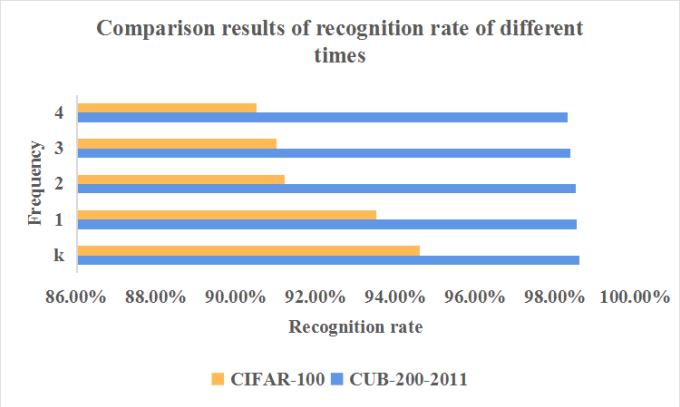


Figure 4 Comparison results of recognition rate of different times

For convolutional neural networks, the output of the previous layer will be the input of the next layer, so the clarity of the feature map of the previous layer will affect the final recognition rate. Experiments have shown that the number of transformations has an effect on the final recognition rate. The random number k is the best recognition effect, and this transformation will be used in future experiments.

(2) The effect of different convolution kernels on the network recognition rate

In the experiment, 2×2, 4×4, 6×6, and 8×8 convolution kernels are selected. The neural network operates on the four sizes of convolution kernels and observes the change of recognition rate. Figure 5 shows the convolution kernels. A graph of the relationship between size and recognition rate.

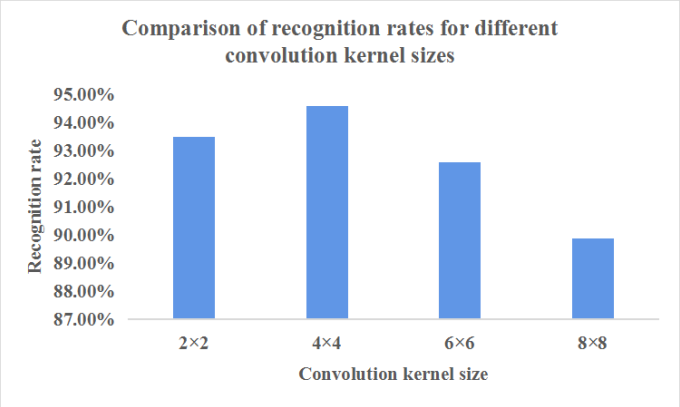


Figure 5 Comparison of recognition rates of different convolution kernel sizes

It can be clearly seen from Figure 4 that as the convolution kernel increases, the feature map will become more and more blurred, because the number of connections between adjacent layers is more when there are fewer convolution kernels than when the convolution kernel is larger. Therefore, the smaller the convolution kernel, the richer the extracted features, but the higher the computational complexity of the network. When the size of the convolution kernel is 4×4, the recognition rate is relatively high, and the amount of network calculation is moderate. Later in this paper, the 4×4 convolution kernel is selected as the neural network parameter.

4.2 IoT Image Classification and Recognition Results and Analysis

In order to be able to more accurately identify the Internet of Things images and improve the accuracy of image recognition, on the basis of the original image recognition method, a

convolutional neural network is added as a secondary classifier to further improve the recognition efficiency of the image recognition algorithm. The recognition result using the convolutional neural network as the secondary classifier is shown in Figure 6.

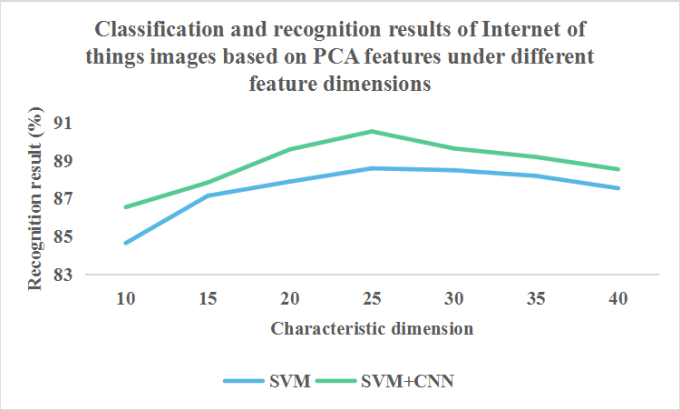


Figure 6 PCA-based image classification and recognition results under different feature dimensions

It can be seen from the figure that the image recognition effect by using the convolutional neural network CNN as the secondary classifier has been further improved compared to using a classifier alone. Using PCA as the method of image feature extraction, the recognition rate of a single SVM is 88.60%, and the recognition rate of the secondary classifier combined by convolutional neural network and SVM reaches 90.55%.

At the same time, the image recognition rate of the two image recognition methods changes with the increase of feature dimension basically consistent. It can be seen from the line chart that when the feature dimension of the image increases from 10 to 25 dimensions, the recognition rate of the image also increases. After more than 25 dimensions, the feature dimension continues to increase, and the recognition rate of the image begins to decrease slightly. Therefore, when the image feature dimension is 25, the best image recognition effect can be obtained. Of course, the linear discriminant analysis method can also be used instead of the principal component analysis method to extract the HSI spatial features of the color image, and then two secondary classification methods are used for image recognition. The recognition results using the secondary classification image recognition method are shown in Figure 7.

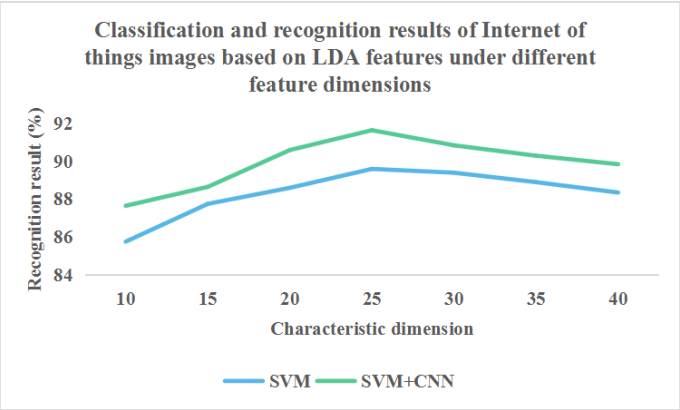


Figure 7 LDA-based image classification and recognition results under different feature dimensions

From the experimental results, it can be concluded that compared with the use of PCA features, the LDA feature is used as the feature vector and the secondary classifier has a higher

recognition rate for the image recognition of the Internet of Things, and the recognition effect is improved to a certain extent. The secondary classifier formed by the convolutional neural network and SVM is better than the single SVM, and the image recognition rate is increased by 1%-2%. The secondary classification image recognition method achieves the best recognition effect when the image feature dimension is 25 dimensions. In order to recognize the recognition rate of different image classification methods under different feature dimensions, as shown in Figure 8, the highest image recognition results of all image recognition methods under different feature dimensions are listed.

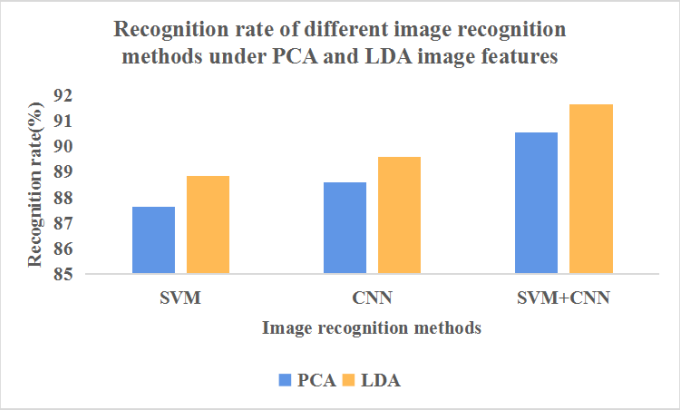


Figure 8 Recognition rate of different image recognition methods under PCA and LDA image features

As can be seen from the figure, among the three methods combined with different feature extraction methods and two secondary classifiers, the secondary classifier composed of convolutional neural network and SVM has a better recognition effect. Under both characteristics, the image recognition rate is higher than that of a single SVM or a single CNN. On the other hand, using the same classification method for image recognition, the correct recognition efficiency of LDA features is higher than that of PCA features. Therefore, the LDA method is used to extract the color features of the HSI space and the secondary classification method composed of a convolutional neural network and SVM is used to classify the images and obtain the best image recognition effect. Therefore, this secondary classification method is selected as the method of identifying the Internet of Things images.

4.3 Analysis of Recognition Results of IoT Images on Different Data Sets

In order to verify the image recognition results of the method proposed in this paper on different types of data sets, we added a comparison experiment of IoT image recognition results on the three data sets provided in this paper. The final result comparison is shown in Figure 9.

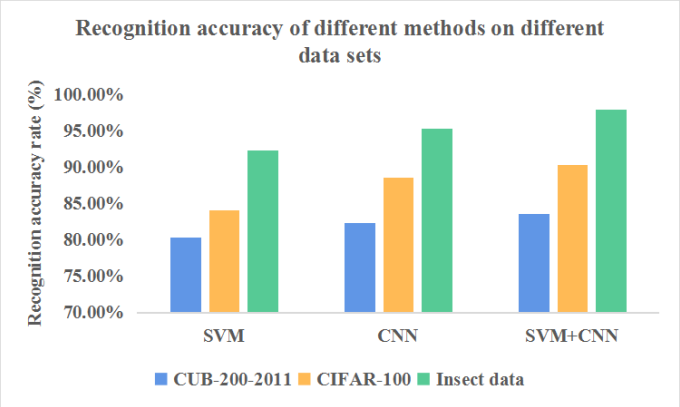
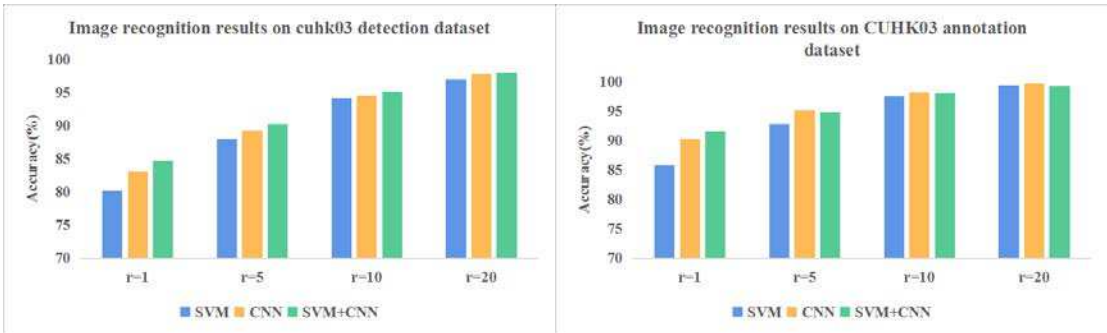


Figure 9 Recognition accuracy of different methods on different data sets

It can be seen from the experimental results that the SVM+CNN method proposed in this paper exceeds the single CNN method by 1.29% (83.56% vs 82.27%) on the CUB-200-2011 data set and 2.64% (97.97% vs 95.33%) on insect data. This shows that the method in this paper can effectively classify and recognize different types of data sets in image recognition. For a batch of fixed-size input images, the lower-level sub-classifier obtains a smaller number of pictures on average, so the randomness of the samples during the training process is strong, resulting in insufficient training and thus poor SVM performance. On the CIFAR-100 data set, the effect of SVM is worse than that of CNN, which is consistent with the insect data set and CUB-200-2011, indicating that SVM needs to play a role when the commonality between categories is large. Although CIFAR-100 is organized in a two-level structure, the differences between categories under a superclass are large, and the association between categories is loose. This shows that it is impossible to improve the classification accuracy by simply merging the categories. Combining the experimental results, it can be concluded that SVM+CNN can identify different types of data sets with higher accuracy.

According to the commonly used data distribution method, for the CUHK03 data, we will use the image of 1367 pedestrians as the training set and the image of 100 pedestrians as the test set. Randomly train the data five times, divide the test set, and randomly select a picture for each pedestrian in the test set to form a candidate set, and calculate the rank-k value. Take the average result of 20 random selections as the result of the current division, and then average the result of five divisions as the final value. The CMC curve on the CUHK03 detection data and the labeled data set is shown in Figure 10.



(a) CUHK03 detection data set (b) CUHK03 annotation data set

Figure 10 CUHK03 data set image recognition results

It can be seen from the figure that the methods in this paper exceed the single method in the data set of labeling and detection. In the annotation data set, the accuracy of rank-1 of CNN is 85.77%, which is more than the accuracy of 90.28% of the SVM method. In the detection data, the accuracy of rank-1 of CNN is 83.11%, which also exceeds 80.22% of SVM. The rank-1 value of SVM+CNN on the detection data set is 84.69%, and the accuracy rates of rank-10 and rank-20 are 95.20% and 98.02%, respectively. In detecting and labeling data, SVM+CNN is slightly better than CNN, and the accuracy of rank -1 is about 1.34%. This shows the effectiveness of the component network, but this advantage gradually decreases, and even worse when the recall value is larger. This is mainly because when the recall value is larger, the similarity of the pictures of the same individual is low, and the addition of fine local features will increase the interference of the remaining individual pictures.

5.Conclusions

This paper designed an IoT image recognition system based on deep learning, which used PCA and LDA to extract IoT image features. The PCA method extracted image features through image transformation, and obtained relatively good experimental results. The LDA method used existing image information for image feature extraction. After projection, the image samples had a relatively large degree of dispersion, which was beneficial to the image recognition of the IoT. Through experiments, this paper verified the effectiveness of the two image recognition methods PCA and LDA, and when the image feature dimension was 25, the best image recognition effect could be obtained. In this paper, the classifier mainly used for image recognition in the IoT was SVM. By using the database to train SVM and CNN, at the same time, the effectiveness of the two for image recognition was tested, and then the trained classifier was used for image recognition. It was found that a CNN and SVM-based secondary classification IoT image recognition method improved the accuracy of image recognition. After experimental verification, the secondary classification method combined the characteristics of SVM and CNN image recognition methods, and the accuracy of the image recognition method had been mentioned as an effective improvement.

Acknowledgements

This work was supported by Young Learners Cultivation Project of Xizang Minzu University "Research on the Learning Differences of College Students in Tibet in the Era of Smart Education" (19MDX05), National Education Science "13th Five-year Plan" Project "Research on Differentiated Instruction Strategies for College Students in Tibet under the Background of Smart Education" (FMB180673).

Authors' Contributions

The work presented here was carried out in collaboration between all authors. Jing Li, Xinfang li and Yuwen Ning defined the research theme. Jing Li, and Xinfang li designed methods and experiments, carried out the experiments, interpreted the results, and wrote the paper. Yuwen Ning co-designed methods and made important revisions to the manuscript. Co-designed experiments and co-worked on analysis. All authors have contributed to, seen, and approved the manuscript.

Funding

This work was supported by Young Learners Cultivation Project of Xizang Minzu University "Research on the Learning Differences of College Students in Tibet in the Era of Smart Education" (19MDX05),National Education Science "13th Five-year Plan" Project "Research on Differentiated Instruction Strategies for College Students in Tibet under the Background of Smart Education" (FMB180673).

Availability of Data and Materials

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Competing Interests

The authors declare that they have no competing interests.

References

1) Deng, Z., Cao, Y., Zhou, X., Yi, Y., & You, I. (2020). Toward efficient image recognition in sensor-based iot: a weight initialization optimizing method for cnn based on rgb influence proportion. Sensors, 20(10), 2866.
2) Ding, L., & Ren, H. (2019). Segmentation algorithm of medical exercise rehabilitation image based on hfcnn and iot. IEEE Access,7(99), 160829-160844.
3) Hexiang, L. I. (2017). Design of iot vision based image automatic identification system for tourist attraction abnormalities. Modern Electronics Technique, 40(4), 124-131.
4) Santos, P. C., De Lima, J. P. C., De Moura, R. F., Ahmed, H., Alves, M. A. Z., & Beck, A. C. S., et al. (2019). A technologically agnostic framework for cyber-physical and iot processing-in-memory-based systems simulation. Microprocessors and microsystems, 69(SEP.), 101-111.
5) Zhong, F., Chen, Z., Ning, Z., Min, G., & Hu, Y. (2018). Heterogeneous visual features integration for image recognition optimization in internet of things. Journal of computational ence, 28(SEP.), 466-475.
6) Thilagavathi, B., & Suthendran, K. (2018). Boosting based implementation of biometric authentication in iot. Journal of Cyber Security & Mobility, 7(1), 131-144.
7) Srinivasa, K. G., Sowmya, B. J., Shikhar, A., Utkarsha, R., & Singh, A. (2018). Data analytics assisted internet of things towards building intelligent healthcare monitoring systems: iot for healthcare. Journal of Organizational and End User Computing, 30(4), 83-103.
8) Yang, A., Zhang, C., Chen, Y., Zhuansun, Y., & Liu, H. (2020). Security and privacy of smart home systems based on the internet of things and stereo matching algorithms. IEEE Internet of Things Journal, 7(4), 2521-2530.
9) Leroux, S., Bohez, S., De Coninck, E., Van Molle, P., Vankeirsbilck, B., & Verbelen, T., et al. (2019). Multi-fidelity deep neural networks for adaptive inference in the internet of multimedia things. Future generation computer systems, 97(AUG.), 355-360.
10) Liu, D., Yang, C., Li, S., Chen, X., Ren, J., & Liu, R., et al. (2019). Fitcnn: a cloud-assisted and low-cost framework for updating cnns on iot devices. Future generation computer systems, 91(FEB.), 277-289.
11) Lu, Y., Wu, S., Fang, Z., Xiong, N., Yoon, S., & Park, D. S. (2017). Exploring finger vein based personal authentication for secure iot. Future Generation Computer Systems, 77(dec.), 149-160.
12) Hsia, C. H., Yen, S. C., & Jang, J. H. (2019). An intelligent iot-based vision system for nighttime vehicle detection and energy saving. Sensors & Materials, 31(6(1)), 1803-1814.
13) Chen, K. X., & Wu, X. J. (2018). Component spd matrices: a lower-dimensional discriminative data descriptor for image set classification. Computational Visual Media, 4(03), 51-58.
14) Tom, Y., Devamanyu, H., Soujanya, P., & Erik, C. (2018). Recent trends in deep learning based natural language processing [review article]. IEEE Computational Intelligence Magazine, 13(3), 55-75.
15) Sun, X., Wu, P., & Hoi, S. C. H. (2018). Face detection using deep learning: an improved faster rcnn approach. Neurocomputing, 299(JUL.19), 42-50.
16) Levine, S., Pastor, P., Krizhevsky, A., Ibarz, J., & Quillen, D. (2018). Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. The International journal of robotics research, 37(4-5), 421-436.
17) Schütt, K. T., Sauceda, H. E., Kindermans, P. J., Tkatchenko, A., & Müller, K.-R. (2018). Schnet –

a deep learning architecture for molecules and materials. Journal of Chemical Physics, 148(24), 241722.

18) Yongyang, X., Liang, W., Zhong, X., & Zhanlong, C.(2018). Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. Remote Sensing, 10(144), 1-18.
19) Li, H., Ota, K., & Dong, M. (2018). Learning iot in edge: deep learning for the internet of things with edge computing. IEEE Network, 32(1), 96-101.
20) He, H., Wen, C. K., Jin, S., & Li, G. Y.(2018). Deep learning-based channel estimation for beamspace mmwave massive mimo systems. Wireless Communications Letters, IEEE, 7(5), 852-855.
21) Bychkov, D., Linder, N., Turkki, R., Nordling, S., Kovanen, P. E., & Verrill, C., et al. (2018). Deep learning based tissue analysis predicts outcome in colorectal cancer. entific Reports, 8(1), 3395.
22) Jing, Y., Bian, Y., Hu, Z., Wang, L., & Xie, X. Q. S. (2018). Deep learning for drug design: an artificial intelligence paradigm for drug discovery in the big data era. Aaps Journal, 20(3), 58.
23) Niko Sünderhauf, Brock, O., Scheirer, W., Hadsell, R., Fox, D., & Jürgen Leitner, et al. (2018). The limits and potentials of deep learning for robotics. The International journal of robotics research, 37(4-5), 405-420.
24) Jia, F., Lei, Y., Guo, L., Lin, J., & Xing, S. (2018). A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines. Neurocomputing, 272(jan.10), 619-628.
25) Arango-Argoty, G., Garner, E., Pruden, A., Heath, L. S., Vikesland, P., & Zhang, L. (2018). Deeparg: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome, 6(1), 23.
26) Higham, C. F., Murray-Smith, R., Padgett, M. J., & Edgar, M. P. (2018). Deep learning for real-time single-pixel video. Rep, 8(1), 2369.
27) Wang, P., Li, W., Ogunbona, P., Wan, J., & Escalera, S. (2018). Rgb-d-based human motion recognition with deep learning: a survey. Computer vision and image understanding, 171(JUN.), 118-139.

Author Information

Jing Li was born in Anyang, Henan, P.R. China, in 1983. She received the master's degree from Beijing Jiaotong University, P.R. China. Now,she works in School of Education, Xizang Minzu University. Her research interests include Education Informationalization, Educational Design.
E-mail: jj2lj314@163.com

Xinfang Li was born in Jining, Shandong, P.R. China, in 1982. He received the master's degree from Northeast Normal University, P.R. China. Now, he works in School of Education, Xizang Minzu University. His research interests include E-learning, Education Informationalization.
E-mail:lidongdong1982@126.com

Yuwen Ning was born in Luoyang, Henan, P.R. China, in 1984. He received the doctor's degree from The Fourth Military Medical University, P.R. China. Now, he works in Information Technology Center of the Fourth Military Medical University. His research interests include Information Technology, Medical Education, Artificial Intelligence.
E-mail:ningyuwen@126.com

Figure:

Figures

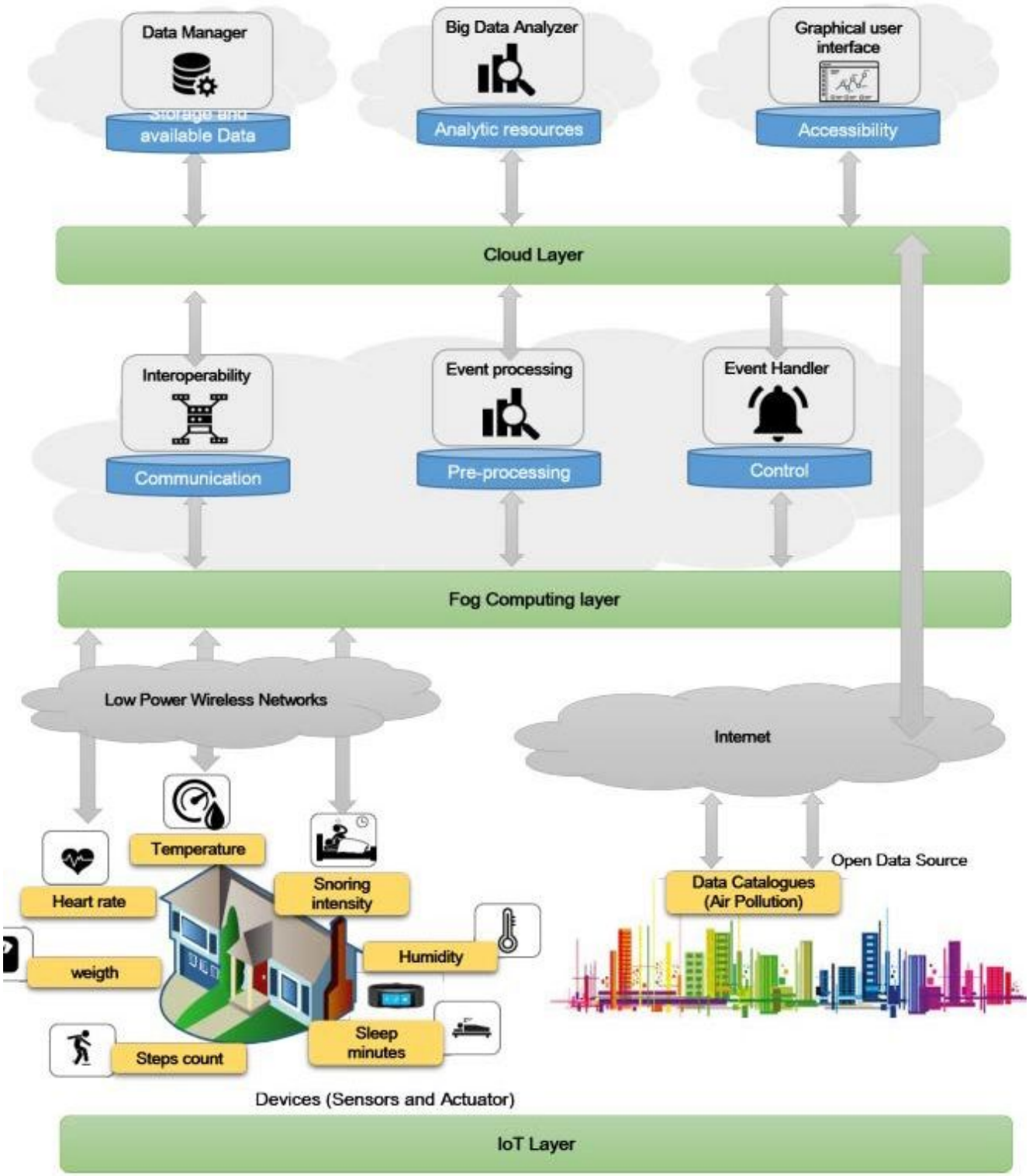


Figure 1

The high-level architecture of an IoT in Image Recognition

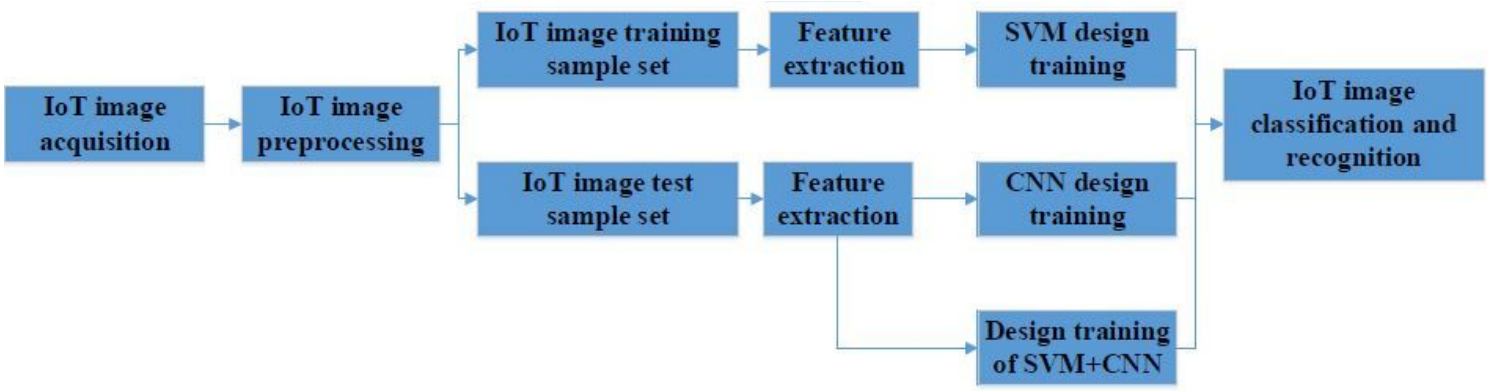


Figure 2

Block diagram of IoT image recognition process

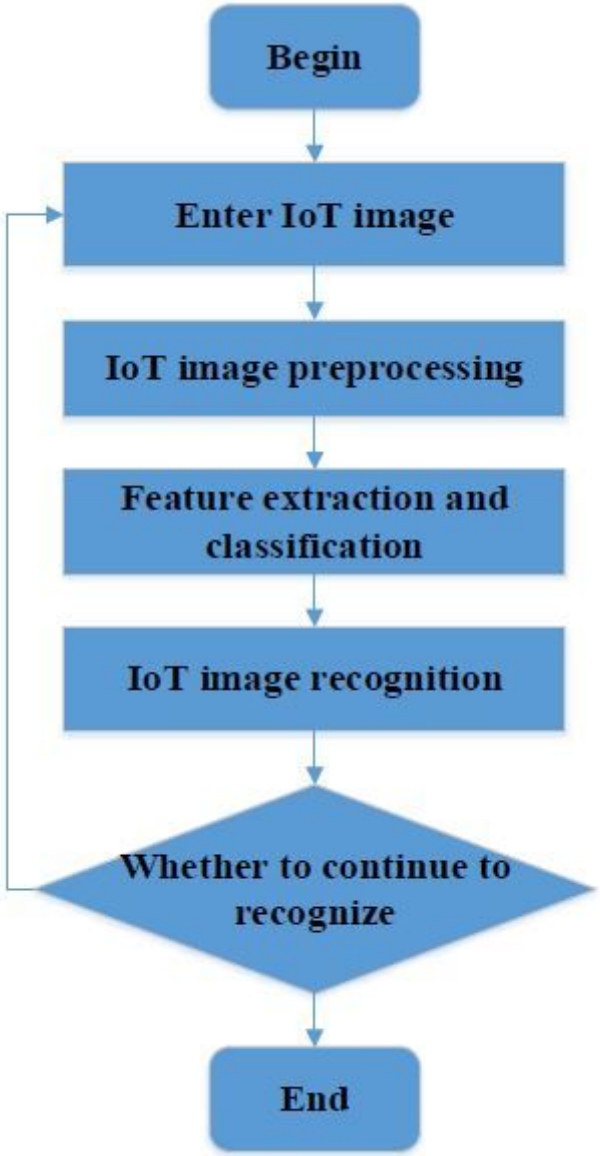


Figure 3

Image recognition flowchart of the Internet of Things image recognition system

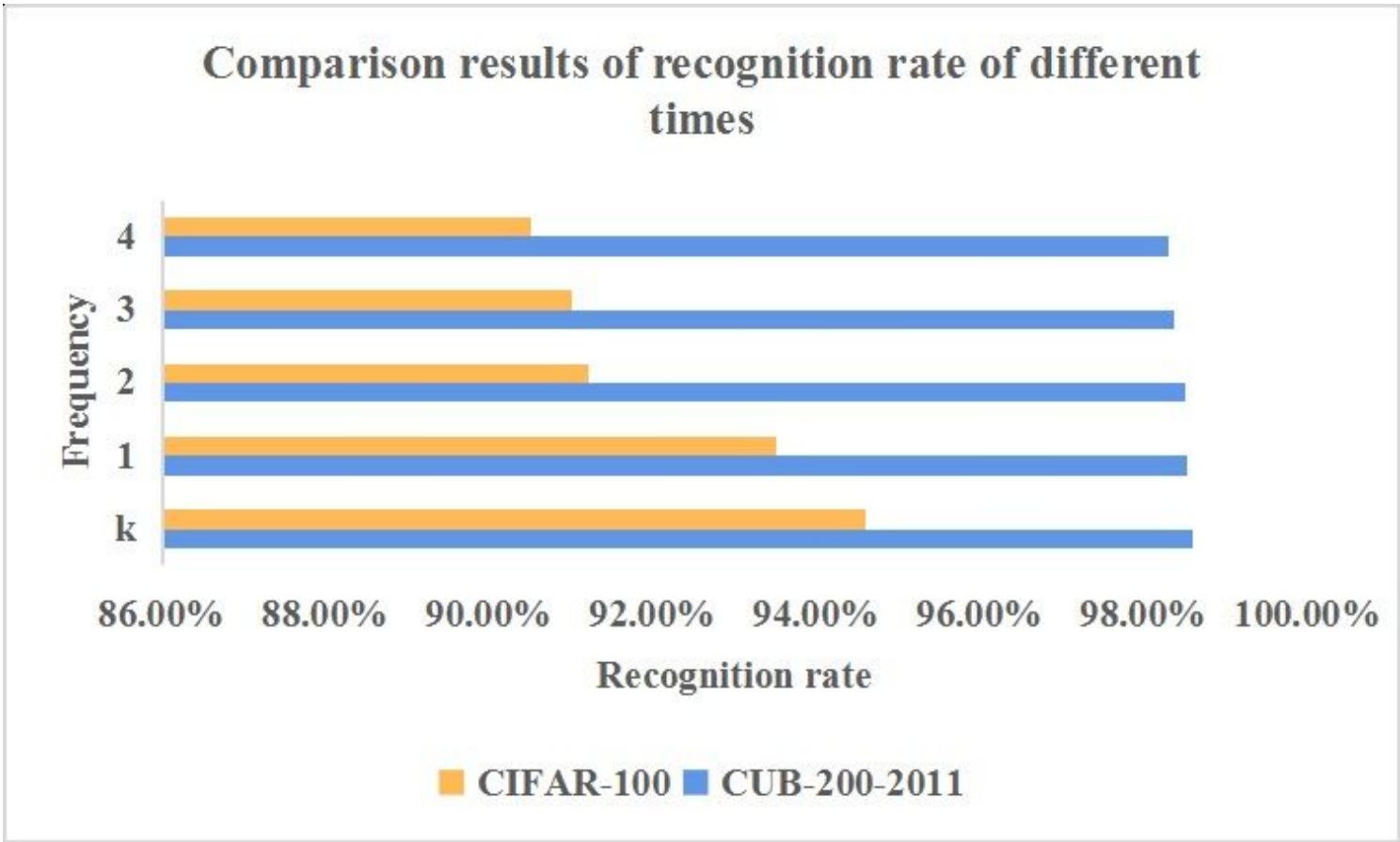


Figure 4

Comparison results of recognition rate of different times

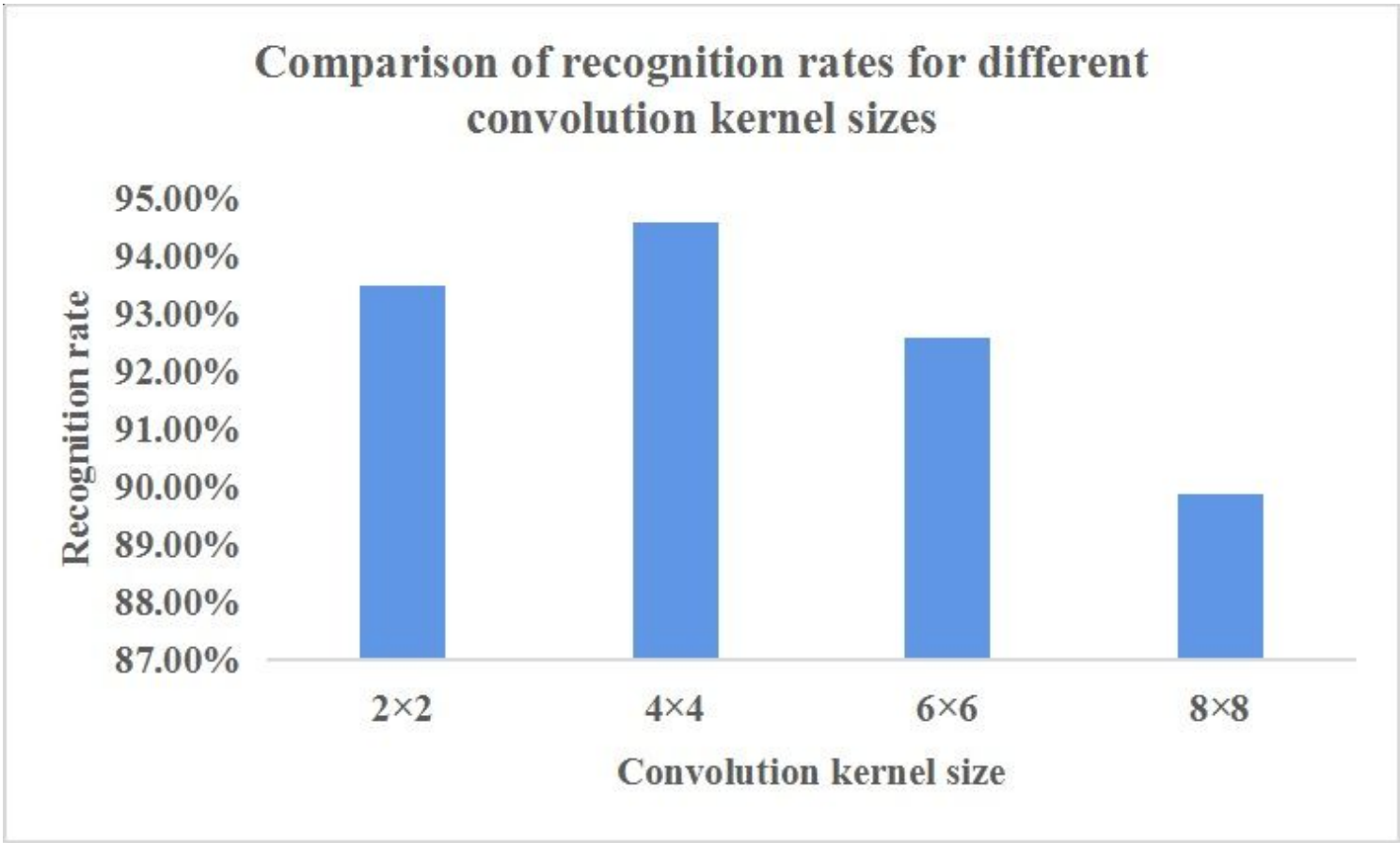


Figure 5

Comparison of recognition rates of different convolution kernel sizes

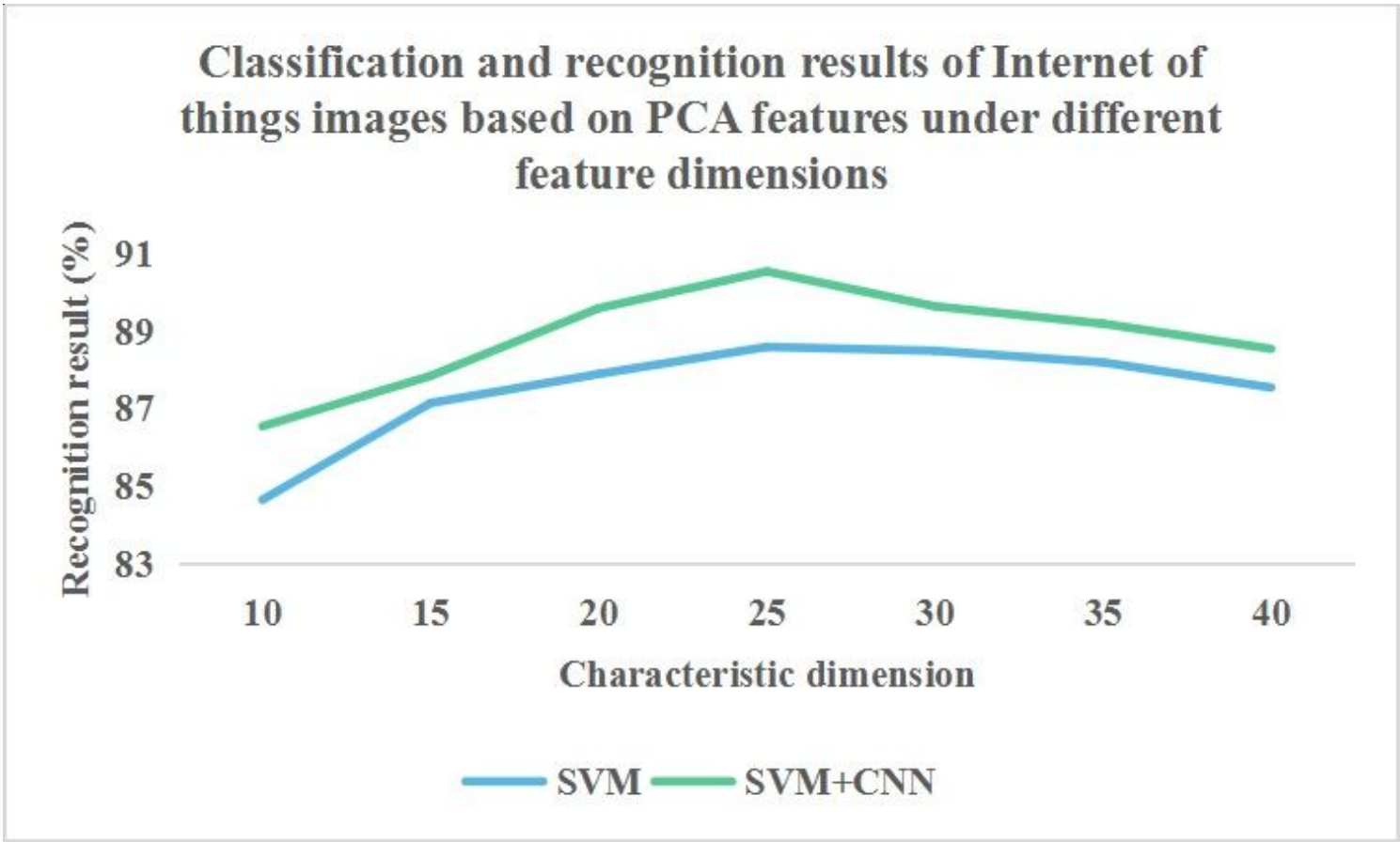


Figure 6

PCA-based image classification and recognition results under different feature dimensions

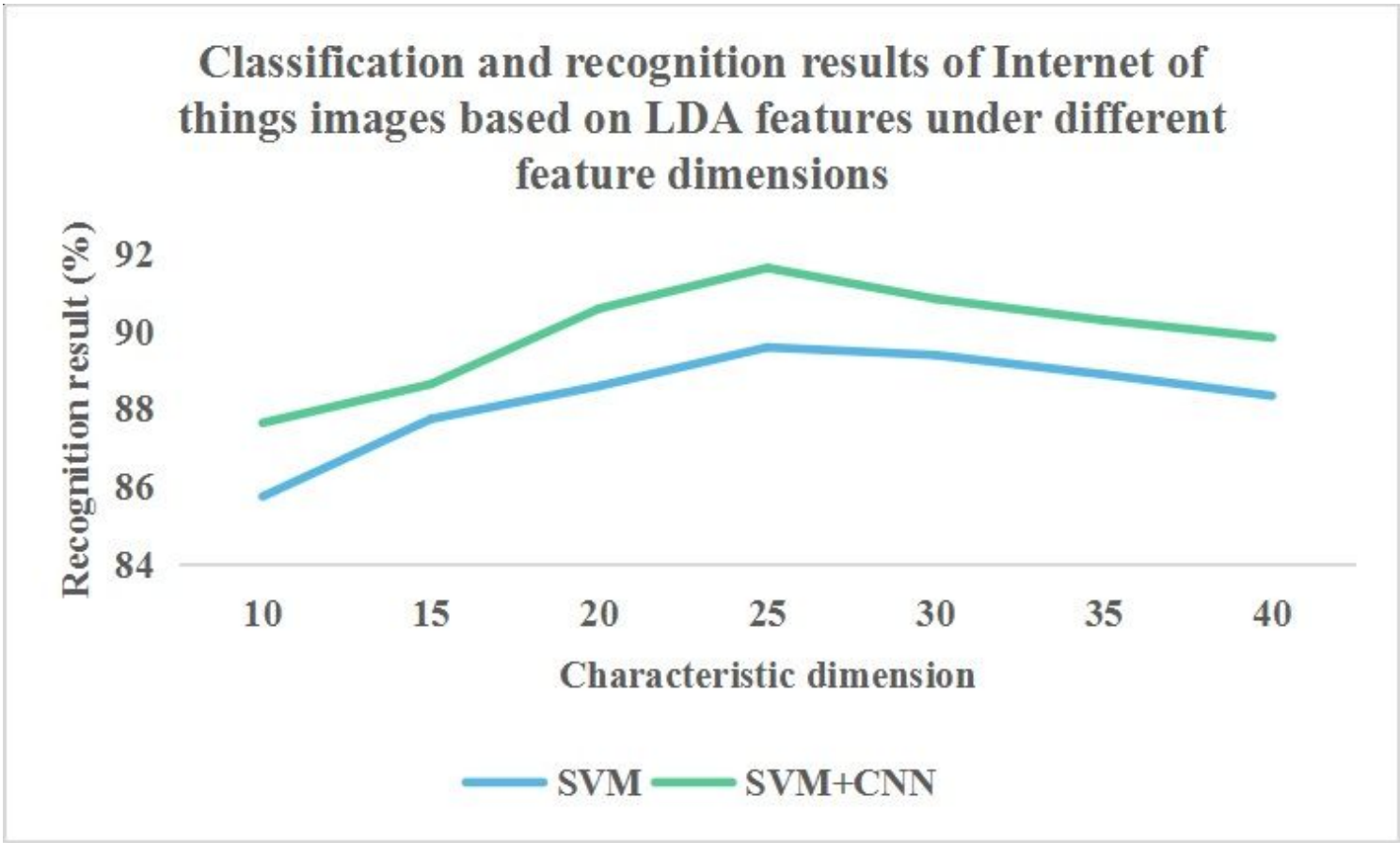


Figure 7

LDA-based image classification and recognition results under different feature dimensions

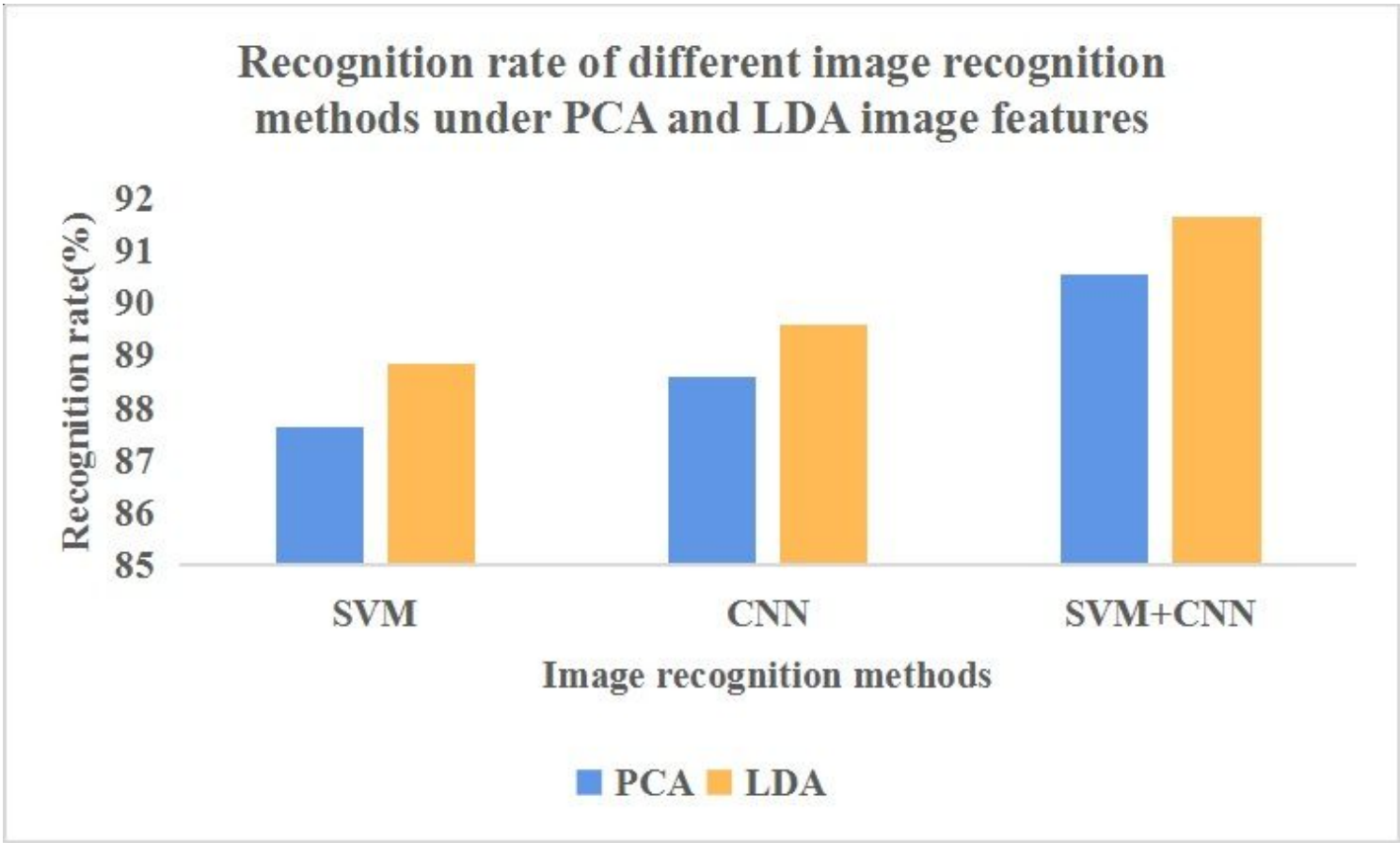


Figure 8

Recognition rate of different image recognition methods under PCA and LDA image features

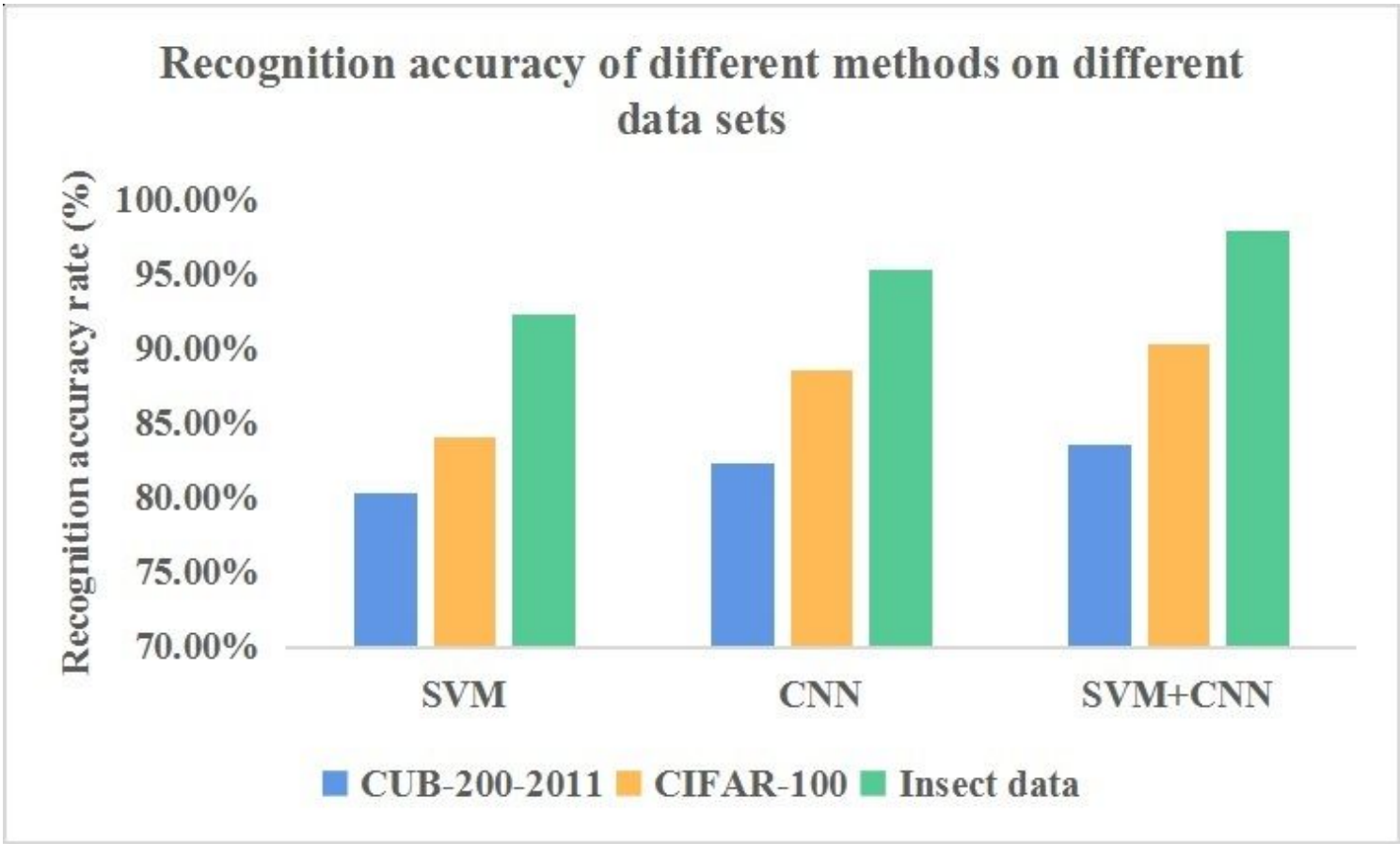
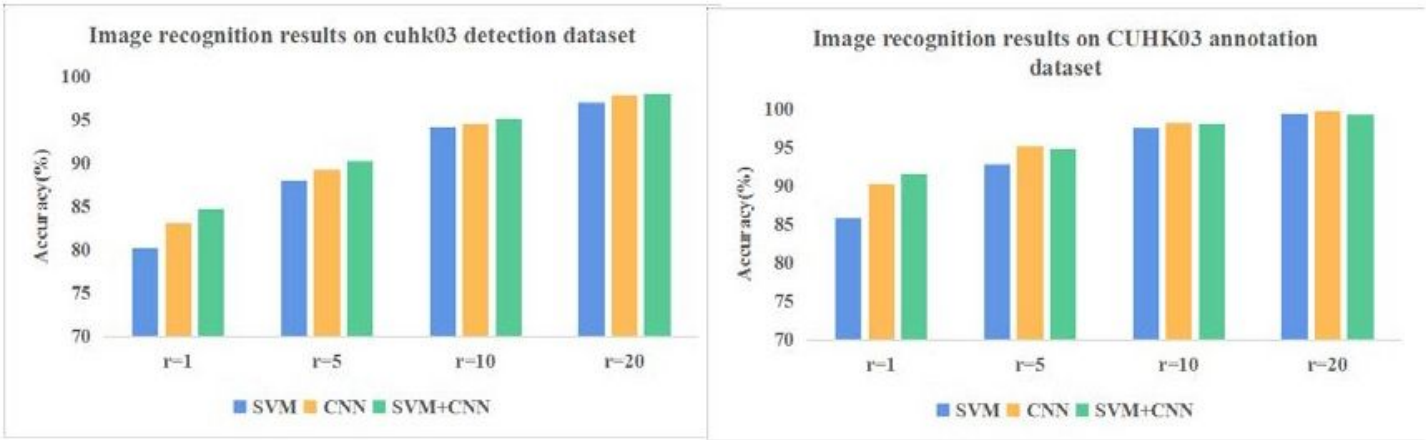


Figure 9

Recognition accuracy of different methods on different data sets



(a) CUHK03 detection data set (b) CUHK03 annotation data set

Figure 10

CUHK03 data set image recognition results