Studies on image recognition based on VAE and AAE

Haobin Zhou*, Bin Li and Qinglei Zhou
School of Zhengzhou University, Henan, China

*Corresponding author: iebinli@zzu.edu.cn

Abstract. Abstract: Image recognition is applied to all aspects of life, such as brush face authentication and defect detection. Traditional convolutional neural network trains network through a large number of data sets. However, in practice, it takes a lot of time and effort to collect training data. In the case of less training data and larger verification data set, the accuracy is relatively low. Therefore, this paper proposes two methods this method can improve the accuracy and loss function fluctuation of the case with less training data and more verification sets. The lenet5 convolutional network structure is used as the main body to form a new convolutional neural network by combining variational self coding and counter self coding. The experimental results show that in the cifar-10 data set and cifar-100 data set, the combination of variational self coding and combined counter self coding can increase by up to 4 percentage points compared with the traditional convolutional neural network, which is effective in the case of insufficient training data.

1. Introduction
With the rapid development of big data and artificial intelligence, Hinton et al. set off a research boom in deep learning [2]. Deep learning has been widely applied in image recognition, face recognition [11, 12], and optimization of industrial quality inspection, which has become a research hotspot in the field of scientific research. Convolutional neural networks have also developed rapidly from shallow models to deep models, from horizontal expansion to vertical and horizontal structures. Compared with traditional methods, deep learning can automatically learn features from a large amount of data and has a strong feature expression ability. However, in the fields of personal information acquisition, military, medical, etc., due to personal privacy, confidentiality mechanisms, high data acquisition costs, and time-consuming and laborious collection and labeling of data, the small amount of training data has become one of the reasons that restrict the performance of the model.

At present, the difficulties of data collection and the low amount of data lead to low accuracy of image recognition have become an urgent problem to be solved. After years of development, solutions have been continually proposed, such as adding pictures through preprocessing methods, random position cropping, random horizontal and vertical flips, random angle rotations, etc., or adding pictures through generative models such as cyclegan [3], or using an adaptive space regularization edge stack self-coder method proposed by Feng et al. to solve the classification of hyperspectral images [4].

This paper proposes two algorithms for this problem-convolutional variational autoencoding and convolutional adversarial autoencoding. This model is more effective in image feature extraction, has higher classification accuracy than traditional methods, and has smaller fluctuations in the loss function.
2. Related theories and methods

2.1. Convolutional Neural Network
Convolutional neural networks were first proposed by Fukushima [8]. In deep learning, convolutional neural networks are mostly used in the field of computer vision. The structure of the convolutional neural network model has five main components, which are input layer, convolution layer, pooling layer, fully connected layer and output layer [14].

Input layer. Generally, many images are used as the input layer.

Convolutional layer. It is composed of many convolution kernels, and performs dot product operation on the image area of the input layer and the convolution kernel. The convolution kernel performs dot product operations on all areas of the sliding image [13].

Activation function. Since the model is linear and lacks non-linear ability, the activation function is used to improve the expression and approximation ability of the model.

Pooling layer. As the output of the upper layer contains unimportant information, it will reduce the processing efficiency, the unimportant information is removed through the pooling layer and only the important data is retained.

Fully connected layer. It is composed of multiple neural networks and is responsible for outputting category probabilities. The subscript of the neuron with the highest probability is the predicted classification result. Finally, the softmax multi-classifier is used to scale the probability of each category to between 0 and 1, and the probability of all categories is added up and equal to 1. The softmax calculation formula is as follows:

\[
p(i) = \frac{\exp(\theta_i^T x)}{\sum_{k=1}^{K} \exp(\theta_k^T x)}
\]

2.2. Variational auto-encoding (VAE)
Variational auto-encoder (VAE) is an important type of generative model, which was proposed by DiederikP. Kingma and Max Welling in 2013 [1], he introduced on the basis of self-encoding Latent variable Z, through which Z enters the decoder to generate an image. VAE is a model that combines deep learning and statistics.

The encoder hopes to build a hidden layer variable Z through the decoder to generate the target data X model, in other words, they assume that the hidden layer variables obey some common distribution (such as normal distribution or uniform distribution), and then hope to train a Model X=g(Z). This model can map the probability distribution to the original image, and enter the encoder to generate corresponding noise for each type of image. Different types of noise have different noise distributions. Different types of noise distributions pass through the decoder to generate corresponding pictures, that is, the purpose of both is to transform distributions. The structure diagram of variational self-coding network is shown in Figure 1.

![Figure 1. Structure diagram of the variational self-coding network](image)
2.3. Adversarial auto-encoding (AAE)
Adversarial Auto-Encoding (AAE) [6] is a model jointly proposed by University of Toronto, Google Brain and OpenAI that uses auto-encoding for adversarial learning. It consists of three parts, an encoder, a discriminator, and a decoder.

\( x \) represents the image data to be input. We input it into the encoder, let the encoder encode it, and it can generate a variable \( z \) (here it is assumed that the variable obeys the probability distribution \( a(z) \)), and then the decoder will try to decode this latent variable \( z \), make it generate a picture again. The loss function is a loss function used by the ordinary encoder, the model here is just an ordinary autoencoder. The important part is that an adversarial network of encoders that act as generator \( G \) and lower discriminator \( D \), the generator and the discriminator are no longer linked to image data, but a one-dimensional vector \( z \). Through continuous training, the discriminator discriminates that the input vector comes from real data (subject to \( p(z) \) probability distribution) or fake data generated by the generator (subject to a predefined \( q(z) \) probability distribution) [9]. Since \( p(z) \) here can be a probability distribution that can be generated by any user, the entire adversarial learning process can actually be thought of as adjusting the encoder to continuously make the probability distribution \( q(z) \) of the data generated close to our predefined \( p(z) \) after the model training is completed, the user can sample a type of distribution, and then use the decoder to generate a new image data. The structure diagram of the confrontation network is shown in Figure 2.

![Figure 2. Structure of anti-self-coding](image)

3. Models and methods

3.1. Basic idea of the model
This paper first builds a convolutional neural network model. Because Lenet5 is a pioneering work in recognition of convolutional networks, it has laid the foundation for the subsequent development of convolutional networks and has a certain representativeness. So, based on the lenet5 [4] model structure, a new convolutional neural network model is constructed. In the convolutional variational autoencoder and the anti-autoencoder model, the structure of the encoder part is the same as that of the convolutional neural network, and the structure of the decoder part performs deconvolution operation based on the encoder structure. The data set is trained to retain the structure and weight of the encoder; under this structure, a fully connected layer is added and trained to obtain the final model. The classifier used is the multi-class regression model Softmax.

3.2. Convolutional neural network combined with VAE

3.2.1. Pretreatment. In the input layer, the image is routinely preprocessed, and the input image is of a uniform size of 32×32. After being rotated, flipped, and standardized, the mean variance is adjusted to 0.5, so that it can accelerate its convergence effect and speed during training.
### 3.2.2. Encoder and decoder.

The image enters the encoder after preprocessing. The encoder structure is composed of two sets of convolution kernels, activation functions, and pooling layers. The convolution kernel is 3×3 in size to extract higher-level features from the input image.

The activation function uses the RELU activation function to perform nonlinear transformation on the convolved data, which is characterized by fast convergence. The calculation formula is as follows:

$$f(x) = \max(o, x)$$

(2)

The pooling layer uses maximum pooling, and the encoder outputs two parts m (mean) and s (variance), where E is randomly generated and obeys a standard normal distribution, and z is a latent variable. The decoder is composed of deconvolution and activation functions, and the latent variables are restored into images through the decoder.

The loss function consists of two parts (bce_loss, kld_loss). bce_loss is the binary_cross_entropy loss, which is used to measure the pixel error between the original image and the generated image. The calculation formula is as follows.

$$\text{BCEloss} = -(y \times \log(y') + (1 - y) \log(1 - y'))$$

(3)

In the formula: y is the true value and y' is the estimated value.

kld is KL-divergence (KL divergence), which is used to measure the difference between the distribution of latent variables and the distribution of user sampling, the calculation formula is below:

$$\text{kldloss} = -0.5 \sum_{i=1}^{n} (1 + s - m^2 - e^i)$$

(4)

In the formula: s is the variance, m is the mean, e is the user-defined sampling distribution, and the overall loss is the sum of the binary cross-entropy and KL. The closer the overall loss is to 0, the more similar the generated image is to the original image. The structure of convolutional variational self-encoding is shown in Figure 4, the specific structure of Encoder is shown in Figure 4(b), and the specific structure of Decoder is shown in Figure 4(a).

![Figure 3. Convolutional Variational Self Coding Structure](image)

#### 3.2.3. Parameter migration.

After many trainings, the encoder's ability to extract image features has been continuously improved, retaining its encoder structure and weights, as the initialization
parameters of the convolutional layer of the convolutional neural network, and the fully connected layer on the splicing as the specific image classification output, after the Softmax function, output the probability of each category and the probability value range is [0,1], where the subscript of the maximum value is the category of the picture. When the model is trained, the learning rate of the fully connected layer of the convolutional neural network is fine-tuned, and then the neural network is classified and trained on this basis.

![Diagram of parameter migration](image)

**Figure 4. Parameter migration**

### 3.3. Convolutional neural network combined with counter-self-encoding

Since the variational auto-encoding is to directly generate the picture and the mean square error of the original picture without confrontation, the generated picture is fuzzy [7], so a discriminator is added for confrontation learning. The preprocessing method is the same as the above. Improved on the basis of VAE, the top row is still a variational self-encoding structure, the latent variable \( z \) reconstructs the image, and the bottom row adds a second network as a discriminator. After training, it can be judged whether the sample comes from a potential variable or the user's own sampling distribution, the discriminator is composed of multiple linear layers, and the final output variable in the interval \([0,1]\) indicates whether the latent variable is sampled from the user's sampling distribution. The encoder is regarded as a generator, the user's sampling distribution is regarded as true, and the output of the generator is false. The generator and the discriminator play against each other to improve the ability of the generator.

The loss function consists of two parts: one is the loss function of variational self-encoding, which is responsible for updating the parameters of the encoder and decoder, and the other is the loss function of the adversarial network which is responsible for updating the parameters of the discriminator. The calculation formula of the loss function of the adversarial network discriminator is as follows:

\[
loss = \log D(x) + \log(1 - D(z))
\]  

(5)

In the formula: \( x \) represents user sampling, and \( z \) is the potential variable output by the generator, that is, the encoder. \( D \) is the discriminator. It is hoped that the data sampled by the user enters the discriminator as true, and the latent variable \( z \) from the variational self-encoding enters the discriminator as false. The calculation formula of the encoder loss function is as follows:

\[
loss = \log(D(z)) + BCEloss + Kldloss
\]  

(6)

Among them, the latent variable output by the encoder passes through the discriminator, hoping to fool the discriminator to make its output true, so that the latent variable constrains the original distribution to approximate the user acquisition distribution, and the loss function of the variational
self-encoding itself makes it have a distribution Transformation mapping ability. After continuous training to improve the representation ability of the encoder, the specific operation steps are as follows:

Step 1: Extract a batch of samples from the data set, and initialize the encoder, decoder, and discriminator.

Step 2: Update the parameters of the encoder and decoder using the errors of the original image and the generated image.

Step 3: Input the latent variable z generated by the encoder into the discriminator, and input the data sampled by the user into the discriminator, so that the discriminator can distinguish whether the data comes from the user sample or the latent variable, calculate the loss function, and reverse update the discriminator and the encoder's parameter weights.

Step 4: Repeat the operation continuously to make the distribution of latent variables generated by the encoder consistent with the distribution of user sampling.

Step 5: Finally, the encoder parameters are transferred to the convolutional layer of the classification model as the initialization parameters, and the model is classified and trained.

4. Experiment and Analysis

4.1. Experimental data

The experimental data uses the cifar-10 data set and the cifar-100 data set provided by the pytorch framework. The cifar-10 data set has a total of 10 categories, including 50,000 images in the training set and 10,000 images in the test set, and the test set contained 10,000 images. The size of the images was 32×32 RGB color images. The data set of 50,000 images was randomly divided with a ratio of 1:9, training set 5000, verification set 45000, and test set 10000. The specific data is shown in Table 1.

The cifar-100 data set has 2 sets of labels, one of which has 100 classes called fine labels, and the other has 20 classes called coarse labels or super classes. In the experiment, coarse labels are used. The training set has 50,000 tests. Set 10,000, divide 50,000 images into 1:4, training set 10,000, validation set 40,000. The specific data is shown in Table 2.

Table 1. Statistical Results of Experimental Data

| classes | training sets | validation sets | validation sets |
|---------|---------------|----------------|-----------------|
| plane   | 505           | 4495           | 1000            |
| car     | 519           | 4481           | 1000            |
| bird    | 506           | 4494           | 1000            |
| cat     | 482           | 4518           | 1000            |
| deer    | 524           | 4476           | 1000            |
| dog     | 479           | 4521           | 1000            |
| frog    | 490           | 4510           | 1000            |
| horse   | 493           | 4507           | 1000            |
| ship    | 497           | 4503           | 1000            |
| truck   | 505           | 4495           | 1000            |
| total   | 5000          | 45000          | 10000           |
Table 2. Statistical Results of Experimental Data

| classes                          | training sets | validation sets | validation sets |
|---------------------------------|--------------|----------------|----------------|
| aquatic                         | 527          | 1973           | 500            |
| fish                            | 502          | 1998           | 500            |
| flowers                         | 454          | 2046           | 500            |
| food containers                 | 469          | 2031           | 500            |
| fruit and vegetables            | 500          | 2000           | 500            |
| household electrical devices    | 519          | 1981           | 500            |
| household furniture             | 527          | 1973           | 500            |
| insects                         | 527          | 1973           | 500            |
| large carnivores                | 462          | 2038           | 500            |
| large man-made outdoor things   | 514          | 1986           | 500            |
| large natural outdoor scenes    | 482          | 2018           | 500            |
| Large omnivores and herbivores  | 499          | 2001           | 500            |
| medium-sized mammals            | 517          | 1983           | 500            |
| non-insect invertebrates        | 503          | 1997           | 500            |
| people                          | 505          | 1995           | 500            |
| reptiles                        | 514          | 1986           | 500            |
| small mammals                   | 497          | 2003           | 500            |
| trees                           | 486          | 2014           | 500            |
| vehicles 1                      | 503          | 1997           | 500            |
| vehicles 2                      | 493          | 2007           | 500            |
| total                           | 10000        | 40000          | 10000          |

4.2. Analysis of experimental results
Convolutional neural network model and combined variational auto-encoding and combined adversarial auto-encoding are performed on the three methods of convolutional neural network model with less training set data to verify their effectiveness. Experimental environment consists of window10+pycharm+python3.6+pytorch.

4.2.1. Comparison based on cifar-10 data set. The traditional convolutional neural network optimizer uses Adam optimizer, the learning rate is 0.0001, the loss function uses CrossEntropyLoss (cross entropy loss function), and the number of training is 60 times.

Variational auto-encoding model and anti-auto-encoding training are divided into two steps. The first step is pre-training: variational auto-encoding is pre-trained on the training set. The training times are 400 times. The optimizer uses Adam optimizer with a learning rate of 0.0001. The original image and the generated image have the smallest loss function, and the loss function is KL divergence plus binary cross entropy.

The second stage is fine-tuning: retain part of the weight structure of the trained variational autoencoder. The encoder structure is consistent with the convolutional layer structure of the convolutional network, and the parameters are migrated to the convolutional layer as the initial test parameters. Select the multi-class cross entropy loss function, the optimizer is Adam, the learning rate of the convolutional layer is 0.0001, the learning rate of the fully connected layer is 0.0000001, and the number of training times is 60. The result of the validation set is shown in Figure 5.
It can be seen from Figure 5 that when the number of training times of the traditional convolutional neural network is less than 10 times, the model is constantly learning image features, the loss function drops rapidly, and the accuracy rate rises rapidly. After more than 10 times, the loss function begins to fluctuate. The verification set loss function begins to rise, over-fitting occurs, and the classification accuracy of the validation set is not high. Therefore, the traditional convolutional neural network has limitations under the condition of insufficient training data, and the convolutional neural network combined with the variational autoencoding training, the number of training There are fluctuations around 15 training times, but the size of the fluctuations is far less than that of traditional convolutional neural networks, and the minimum value of the loss function is also lower than the traditional one, and the accuracy is higher than the traditional one, so it is more effective than the traditional one. The result of training combined with the adversarial network is smaller and more stable than the combined variational auto-encoding loss function, and the value is lower. In terms of accuracy, when training is more than 30 times, the classification effect of combined adversarial self-coding is better than that of variational self-coding. Therefore, the adversarial model is more effective than the variational model.

| classes   | convolutional neural network | variational self-coding model | adversarial self-coding model |
|-----------|------------------------------|--------------------------------|-------------------------------|
| PLANE     | 63%                          | 69%                            | 71%                           |
| CAR       | 72%                          | 74%                            | 74%                           |
| BIRD      | 51%                          | 51%                            | 55%                           |
| CAT       | 43%                          | 45%                            | 48%                           |
| DEER      | 55%                          | 57%                            | 59%                           |
| DOG       | 52%                          | 55%                            | 56%                           |
| FROG      | 76%                          | 75%                            | 75%                           |
| HORSE     | 69%                          | 71%                            | 74%                           |
| SHIP      | 75%                          | 73%                            | 76%                           |
| TRUCK     | 71%                          | 73%                            | 74%                           |
| AVERAGE   | 63.22%                       | 64.95%                         | 66.61%                        |

Testing on the test set shows that the variational autoencoding model has a higher accuracy rate than the traditional convolutional network in 7 of the 10 categories on the cifar-10 data set, two categories are lower than the traditional, and one category is the same. Among the 10 categories of anti-autoencoding, only FROG is one category lower than the variational auto-encoding network. Two categories are the same as the variational auto-encoding. The other categories are higher than the variational auto-encoding. The average accuracy of the anti-autoencoding is averagely accurate. The highest rate is 66.61%, which is about 3% higher than traditional convolution and about 2% higher than variational auto-encoding. Based on the above comparison, it can be concluded that the adversarial model is more suitable for classification in the case of insufficient training set.
4.2.2. Comparison based on cifar-100 data set. The training is also divided into two parts, one is pre-training to train the autoencoder, the other is to transfer the generator parameters to full connection, fine-tuning the full connection learning rate, the optimizer uses adam, the convolutional layer learning rate is 0.0001, and the number of training is 400 times. The learning rate of the fully connected layer is 0.0000001, the loss function adopts multi-class cross entropy, training 60 times, and the result of the verification set is shown in Figure 6.

![Image](http://example.com/image.jpg)

(a) Change in loss  
(b) Changes in accuracy

**Figure 6.** Comparison of Experimental Results

It can be seen from the above figure that due to insufficient training samples. The loss function of the traditional convolutional network starts to rise after training 10 times. The over-fitting phenomenon is obvious, while the variational model and the adversarial model are relatively stable. The adversarial model is more excellent and accurate. In terms of rate, the traditional convolutional network has the lowest accuracy. The adversarial model surpassed the other two at 20 iterations. The adversarial network is better in terms of loss function and accuracy.

| classes                        | convolutional neural network | variational self-coding model | adversarial self-coding model |
|-------------------------------|-----------------------------|------------------------------|------------------------------|
| aquatic                       | 35%                         | 34%                          | 35%                          |
| fish                          | 46%                         | 44%                          | 42%                          |
| flowers                       | 65%                         | 68%                          | 66%                          |
| food containers               | 49%                         | 42%                          | 48%                          |
| fruit and vegetables          | 42%                         | 57%                          | 53%                          |
| household electrical devices  | 29%                         | 32%                          | 40%                          |
| household furniture           | 47%                         | 56%                          | 55%                          |
| insects                       | 43%                         | 49%                          | 37%                          |
| large carnivores              | 21%                         | 30%                          | 39%                          |
| large man-made outdoor things | 57%                         | 58%                          | 68%                          |
| large natural outdoor scenes  | 62%                         | 63%                          | 57%                          |
| Large omnivores and herbivores| 30%                         | 46%                          | 35%                          |
| medium-sized mammals          | 44%                         | 32%                          | 34%                          |
| non-insect invertebrates      | 26%                         | 30%                          | 40%                          |
| people                        | 49%                         | 53%                          | 58%                          |
| reptiles                      | 24%                         | 32%                          | 22%                          |
| small mammals                 | 29%                         | 33%                          | 37%                          |
| trees                         | 75%                         | 75%                          | 79%                          |
| vehicles 1                    | 49%                         | 47%                          | 46%                          |
| vehicles 2                    | 42%                         | 49%                          | 49%                          |
| average                       | 43.62%                      | 46.91%                       | 47.34%                       |
Testing on the test set shows that the variational auto-encoding model has a higher accuracy rate than the traditional convolutional network in 15 of the 20 categories on the cifar-100 data set, four categories are lower than the traditional, and one category is the same. The 20 categories and 8 categories in the anti-autoencoding are lower than the variational auto-encoding network. One category is the same as the variational auto-encoding, and other categories are higher than the variational auto-encoding. The average accuracy of the anti-autoencoding is the highest, which achieved 47.43%, and is about 4% higher than traditional convolution and about 1% higher than variational autoencoding. Based on the above comparison, it can be concluded that the adversarial model is more suitable for classification in the case of insufficient training set.

5. Conclusion
Aiming at the problem of difficult data collection and low amount of training data resulting in low image recognition accuracy, this paper proposes two algorithm models of convolutional neural network combined with variational auto-encoding and combined with anti-auto-encoding. These two algorithm models form a new convolution neural network, which has good image recognition ability under the condition of less training data and large test data. The experimental results show that the accuracy of combined variational auto-encoding in the cifar-10 data set reaches 64.95%, the accuracy of combined adversarial auto-encoding reaches 66.61%, the accuracy of traditional convolutional networks in cifar-100 reaches 43.62%, the variational auto-encoding accuracy rate reached 46.91%, and the anti-self-encoding reached 47.34%. The anti-self-encoding model has more advantages in image classification than the variational self-encoding model. Although the accuracy rate has improved and the over-fitting phenomenon has been alleviated, it still exists. In the future, combined with facenet [10] model, the convolutional layer will be trained with the idea that the feature vectors of the same category are close to each other and the feature vectors of different categories are far away, so as to study whether the classification effect is effective when the training data is relatively small.

References
[1] Diederik P Kingma and Max Welling. Auto-encoding variational bayes arXiv preprintarXiv: 1312. 6114, 2013.
[2] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313 (5786): 504-507.
[3] Zhu Junyan, Park T, Isola P et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C] // Proc of IEEE International Conference on Computer Vision. 2017: 2223-2232.
[4] Chu Xiangyang, Liu Qiao, Ye ou. Research on face authentication algorithm combining LeNet-5 and Siamese neural network model [J/OL]. Computer Engineering and application: 1-8 [2020-06-27]. http://kns.cnki.net/kcms/datetail/11.2127.TP.20190917.1452.004.html
[5] Feng J, Liu L, Cao X, et al. Marginal Stacked Autoencoder With Adaptively-Spatial Regularization for Hyperspectral Image Classification [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(9): 3297-3311.
[6] MAKHZANI A, SHLENS J, JAITLY N, et al. Adversarial autoencoders [EB/OL]. [20180620]. https: // arxiv.org/ abs/1511.05644.
[7] Chen Lifang, Lu Guojun. Mnist handwriting digital image generation method based on the Fusion network of VAE and GAN Journal of Langfang Normal University (Natural Science Edition) 2019, 19 (02): 25-29.
[8] Fukushima K. Neocognitron: A hierarchical neural network capable of visual pattern recognition [J]. Neural Networks, 1988, 1 (2): 119-130.
[9] Wu Yuqian; Wang Junli; Yang Li; Review of research on cost-sensitive deep learning methods for Yu Miaomiao [j]. Computer science 2019, 46 (05): 1-12.
[10] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering [C]// Proceedings of the IEEE conference on computer vision and pattern
[11] Zhang Guoyun, Xiang Canqun, Luo Baitong. An improved CNN structure for face recognition [J]. Computer Engineering and Applications, 2017, 53 (17): 180-185.

[12] Zheng Yinhuan, Wang Beizhan, Wang Jiajun. Application of deep Convolution Neural Network in Face feature point detection [J]. Computer Engineering and Applications, 2019, 55 (04): 173-178.

[13] Li Zhou's continuous human motion recognition based on 3D CNN [D] Harbin: Harbin Institute of Technology 2018

[14] Wan shining. Research and implementation of face recognition based on Convolution Neural Network [D]. Chengdu: University of Electronic Science and Technology, 2016.

[15] Research Hu Zhongxu gesture recognition technology in human-computer interaction in virtual scenes [D] Wuhan: Huazhong University of Science and Technology, 2018