International Journal of Pattern Recognition
and Artificial Intelligence
Vol. 34, No. 3 (2020) 2051004 (21 pages)
© World Scientific Publishing Company
DOI: 10.1142/S0218001420510040

Classification of Fashion Article Images Based on Improved Random Forest and VGG-IE Algorithm

Jian Liu[*], Yuchen Zheng[†], Ke Dong[‡], Haitong Yu[§], Jianjun Zhou[¶],
Ye Jiang[‖], Zhaoneng Jiang[**], Sujie Guo[††] and Rui Ding[‡‡]

School of Computer Science and Information Engineering
School of Artificial Intelligence
Hefei University of Technology, Hefei 230009
[*] *jianliu@hfut.edu.cn*
[†] *819546483@qq.com*
[‡] *coliapaston@163.com*
[§] *hyu2666@qq.com*
[¶] *1326632773@qq.com*
[‖] *jiangye@hfut.edu.cn*
[**] *jiangzhaoneng@hfut.edu.cn*
[††] *809378709@qq.com*
[‡‡] *dingrui.hfut@gmail.com*

Received 9 January 2019
Accepted 2 April 2019
Published 31 July 2019

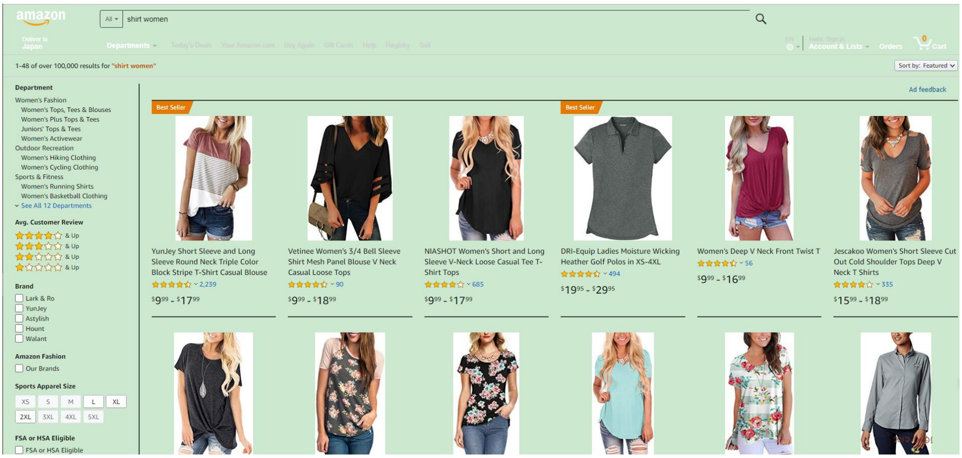
In classification of fashion article images based on e-commerce image recommendation system, the classification accuracy and computation time cannot meet the actual requirements. Herein, for the first time to our knowledge, we present two diverse image recognition approaches for classification of fashion article images called random-forest method based on genetic algorithm (GA-RF) and Visual Geometry Group-Image Enhancement algorithm (VGG-IE) to solve classification accuracy and computation time problem. In GA-RF, the number of segmentation times and the decision trees are the key factors affecting the classification results. Improved genetic algorithm is introduced into the parameter optimization of forests to determine the optimal combination of the two parameters with minimal manual intervention. Finally, we propose six different Deep Neural Network architectures, including VGG-IE, to improve classification accuracy. The VGG-IE algorithm uses batch normalization and seven kinds training-data augmentation for ease and promotion of learning process. We investigate the effectiveness of the proposed method using Fashion-MNIST dataset and 70 000 pictures, Experimental results demonstrate that, in comparison with the state-of-the-art algorithms for 10 categories of image recognition, our VGG algorithm has the shortest computational time when it satisfies certain classification accuracy. VGG-IE approach has the highest classification accuracy.

Keywords: Image recommendation; deep learning; random forest; VGG; Fashion-MNIST.

[**] Corresponding author.

1. Introduction

With the improvement of computer processing speed, parallel processing of GPU computers and support of large-scale clustering technology in recent years, the training process that takes months in computer vision and image recognition algorithms can be shortened by days or even hours. Computer vision algorithms based on deep learning are gradually applied in medical image diagnosis, agricultural and forestry industry image analysis, security video monitoring, unmanned driving assistance and other industries. Among them, the image recognition application and research of electronic commerce has become a research hotspot in recent years, Price Waterhouse Coopers (PWC) released the Total Retail Survey 2017, survey of 24 000 consumers in 29 countries around the world, the results showed that 40% of consumers prefer to buy online clothing and footwear products, 52% of consumers prefer online search fashion commodity information; In terms of clothing purchase, 72% of Chinese consumers prefer online shopping, with the highest proportion.[16,23] Relying on the prosperity of e-commerce, it is a research topic of great application value to automatically recommend clothes to users through intelligent algorithms and accurately match the types of clothes users really expect to buy. The previous text-based image recommendation system manually annotated the image and then used text retrieval technology to find the image that met the requirements (as shown in Fig. 1) according to the keywords entered by the customer. Merchants or e-commerce platforms have to spend a lot of manpower, material resources and financial resources to adjust keywords in case of a search, and customers have to spend a lot of time and energy on irrelevant or inaccurate search. The recommendation method based on image content is to intelligently analyze the images provided by customers or shopping malls, and make use of color, texture, shape, PHOG, SIFT, CEDD and other more intelligent and robust recommendation methods with different


All ▾ shirt women

1-48 of over 100,000 results for "shirt women"

Sort by: Featured

Fig. 1. The search result of shopping website (Image retrieval based on keywords).

characteristics.[20] Amazon A9 launched a service for users to search commodity images, and they are committed to applying this service to various image matching services in reality.[12] In Google, Like's main business is image search engine, which started with facial recognition technology. The company set up the technical goal of fashion search engine at the beginning of its establishment. The main recommendation mode is to allow users to enter the description of products or search the image information of the products selected by users to find the products with the most similar appearance or description for recommendation. As for these novel image-based recommendation systems, Alibaba and other research and development departments have begun to study this aspect, but progress is slow. For example, the mobile APP terminal launched the function of "pailitao."[21] Users can upload photos of any desired products to the cloud and compare and analyze the features of the products in the massive database to find the products with the highest similarity to the images uploaded by users for recommendation, but the disadvantages are low accuracy and long time. In the context of Internet + era, artificial intelligence has been deeply involved in various fields of work and life, but the desynchronization between academia and industry is very obvious. Therefore, it is of great significance to study and design algorithms and systems to meet the needs of online clothing recommendation market. Fast and accurate image retrieval system requires fast and accurate image classification algorithm of clothing or shoes. In addition, the electronic mall database of their own clothing image big data in a short time to complete accurate classification is also an urgent problem to solve. Therefore, the main contribution of this paper is in the clothing image fast and accurate classification algorithm of clothing or shoes. In addition, the electronic mall database of their own clothing image big data in a short time to complete accurate classification is also an urgent problem to solve. Therefore, the main contribution of this paper is the fast and accurate classification algorithm of clothing image and shoe image.

This paper studies the apparel and footwear data in shopping of e-commerce. In the research, the Passion MNIST dataset selected in this paper[9] provides about 70 000 positive images of different commodities, which are respectively from short sleeves, t-shirts, trousers, dresses, sports shoes and other 10 categories. In order to realize the fast and accurate classification and recognition of clothing image and shoe image, this paper proposes the improved random forest classification algorithms based on genetic algorithm (GA-RF algorithm) and the deep convolution neural network algorithm (VGG-IE algorithm) based on image enhancement. Two algorithms are proposed for computing time and classification accuracy. However, there is a contradiction between classification accuracy to meet the task requirements in different scenarios. The experiment uses the real data of Zalando, a German Fashion e-commerce Internet company, to build the Fashion-MNIST standard dataset to qualitatively and quantitatively evaluate the algorithm performance from the perspectives of computing time and classification accuracy.

2. Algorithm Principle and Application

2.1. *Traditional random forest classification*

Random forest is constructed by multiple decision trees, each tree obtains the classification result by contributing its own different classifier.[22] In the statistics of random forest classification algorithm based on random Bootstrap sample selection, Bootstrap refers to any test or measure that relies on random sampling with put back. This technique uses random sampling to estimate the sampling distribution, especially for small samples. The basic process of random forest classification is as follows:

(i) Construct the original training set, in which there are N trees and M times of segmentation, and the training set is used for tree construction;
(ii) Establish a random subset by means of random sampling with put back, so as to generate training for the random forest with N trees;
(iii) Before selecting variables (features) for each nonleaf node (internal node), random forest algorithm randomly selects a certain number of features from all features, uses them as segmentation features of the current decision tree, and selects the best segmentation node. The number of variables attempted in each partition is expressed as $Mtry(Mtry \leq M)j$;
(iv) Without pruning, the growth of trees can be maximized;
(v) The generated trees are combined to form a random forest. Where each tree votes for the classification of its decision, the output of the random forest classifier is determined by a majority vote of the tree.
(vi) Figure 2 shows an effective method to establish a decision tree based on Gini purity criterion.

Assuming that set 5 contains K-type eigenvalues and each type of eigenvalue generates a child node, Gini(i) calculates the formulate (1) for Gini coefficient of node i:

$$\mathrm{Gini}(i) = 1 - \sum_{j=1}^{h} \left[p\left(\frac{j}{i}\right) \right]^2, \tag{1}$$

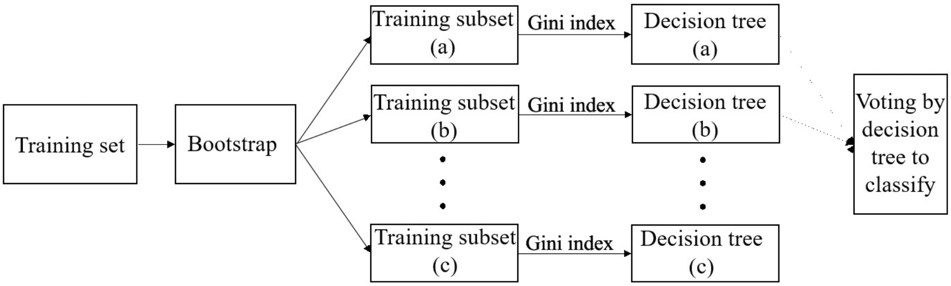


Fig. 2. The procedure of decision tree based Gini impurity.

where h is the node type $p(j/i)$ is the relative frequency of j type on node i. If the node is a single category, the impurity should be zero, which indicates that maximum entropy has been reached or the best classification has been used. $\mathrm{Gini}(i)$ should be at its maximum value when a node is equally allocated across all classes, which means that this is the worst classification used for splitting.

The calculation formula of the splitting exponent of set S is

$$\mathrm{Gini}_{\mathrm{split}}(S) = \sum_{i=1}^{r} \frac{s_i}{s} \mathrm{Gini}(i), \tag{2}$$

where $\mathrm{Gini}_{\mathrm{split}}$ is the split index of set S, R is the data type in set S, s_i is the number of node i, and s is the total number of set S.

The algorithm traverses all variables and nodes, and Gini stops splitting at lowest level, which means splitting is maintained until the end nodes have only a few nonsingle categories or until the nodes classified are completely single categories. Finally, each randomly generated tree is used to classify all image pixels. As shown in formula (3), the type of each pixel is classified by comprehensive voting for all classifications of each tree, that is,

$$c = \mathrm{argmax}_C \left\{ \sum_{k=1}^{N} I[h(x, \theta_k) = c] \right\}, \tag{3}$$

where c is the category, C is the collection of categories, N is the number of decision trees, $I(A)$ is an indicator function. When condition A is true, the value of $I(A)$ is 1; when condition A is false, the value of $I(A)$ is zero. $h(x, \theta_k)$ is the decision tree classifier; x is the training set, θ_k is a random, independent and identically distributed eigenvector.

2.2. *Random forest algorithm for base-stem genetic algorithms (GA-RF)*

It is necessary to select important trees from too many trees. By optimizing the key parameters of random forest algorithm, the classification accuracy can be improved within the acceptable efficiency range. When the number of training samples is determined, two important factors affect the accuracy of decision making: the number of decision trees n and the number of segmentation times m.

When the random forest is generated, a decision tree will be generated for each selected training set, and constant proportion variables will be randomly selected from all variables as the split variable set of the decision tree. When commodity images are classified by random forest algorithm, the segmentation frequency m will affect the intensity of decision tree and the correlation between decision trees. When the value of m is small, a single decision tree is weak, which often reduces the ability of classifier based on random forest. On the contrary, classifiers with low correlation between decision trees have strong functions, but they also face problems such as large computation amount and long computation time. Therefore, choosing the right m is the key to improve the classification accuracy. The number of decision trees (n) determines the accuracy of voting and random forest. In Ref. 5, it is proved that when the number of trees increases, the generalization error tends to converge and overfitting is avoided. However, Gislason pointed out that the actual experiment showed that the increase in the number of trees does not necessarily improve the classification accuracy, and the error of the test set will converge to the asymptotic value.[8] On the contrary, two parameter sets (number of segmentation m, number of trees n) can guarantee the accuracy, and the number of trees does not necessarily need to be a certain maximum. Because of the complexity of these two random processes, the constant increase in the number of trees will only increase the precision slightly. Therefore, this paper needs to determine the appropriate number of trees within the acceptable efficiency range to obtain the highest classification accuracy.

Determining the appropriate parameter set (segmentation frequency m, number of trees n) is the key to the classification results, especially when the number of noises becomes too large. You can use past experience or walk through all the parameter combinations to select the parameters of algorithm. Bryman tested the error rate of the algorithm, where m is the number of inputs from 1 to Int$(\log 2M + 1)$ in the segmentation number m, namely the total number of features), and concluded that when the number of trees in the forest increases, the generalization error of the forest converges to a limit. When the number of trees approaches infinity, the computation quantity increases sharply, but it does not help the improvement of accuracy. Gislason points out that when a point has reached error convergence and the total precision set based on the square root of the input number is close to the most accurate result, the random forest model is no longer sensitive to the segmentation frequency m. But this is not ideal for all cases. In general, there is no general solution for setting the optimal parameters of the random forest classifier in different datasets. In this paper, genetic algorithm is adopted to optimize two parameters of random forest (number of trees n and number of segmentation m).

As shown in Fig. 3, genetic algorithm can be introduced into the parameter optimization of random forest to determine the optimal combination of two parameters with the minimum manual intervention. The OOB classification accuracy estimated in the random forest is usually used to evaluate the error rate of the random forest, so the OOB can also be used as the fitness function of the genetic algorithm. The final goal of image classification based on random forest is to evaluate the classification accuracy of the test set, while the goal of parameter optimization is to obtain higher classification accuracy. Therefore, the fitness function of genetic algorithm is another feasible method to evaluate the overall classification using test sets.

In order to ensure the efficiency of the algorithm, the computation is reduced by setting the reasonable maximum parameter through programming experience. However, a certain number of initial communities can be generated within the range

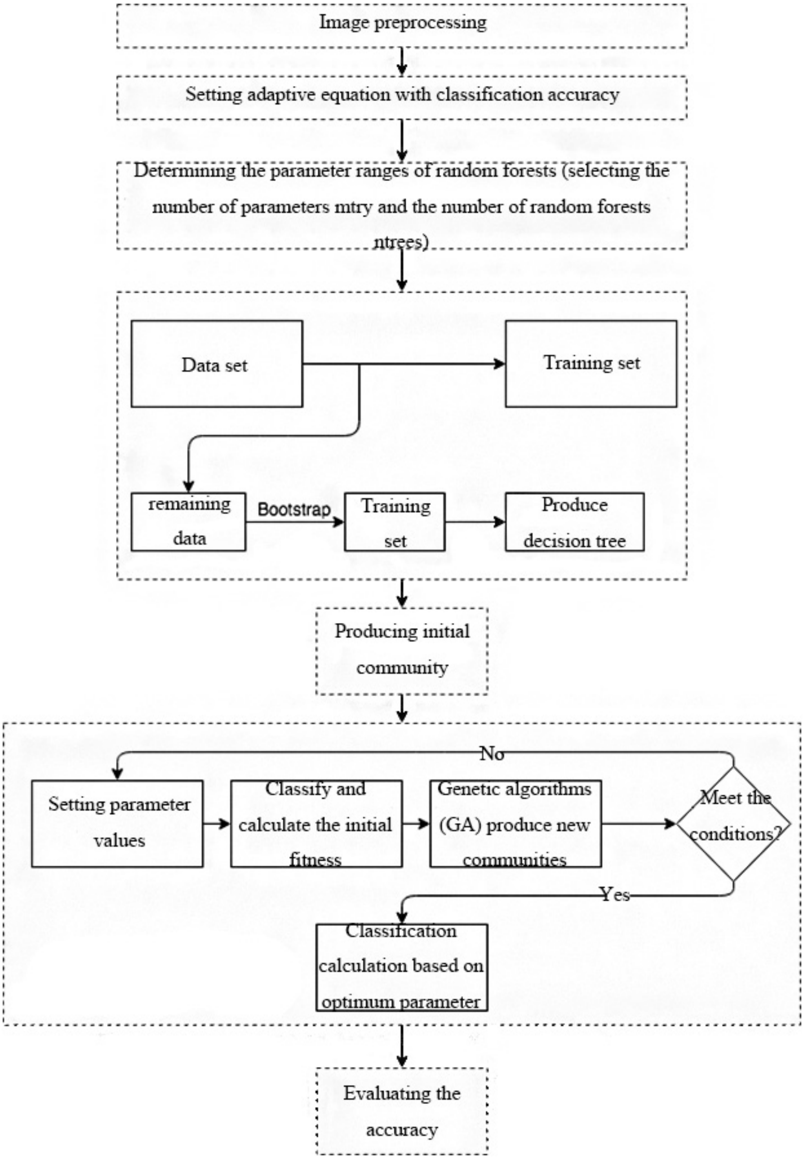


Image preprocessing
Setting adaptive equation with classification accuracy
Determining the parameter ranges of random forests (selecting the number of parameters mtry and the number of random forests ntrees)
Data set
Training set
remaining data
Bootstrap
Training set
Produce decision tree
Producing initial community
No
Setting parameter values
Classify and calculate the initial fitness
Genetic algorithms (GA) produce new communities
Meet the conditions?
Yes
Classification calculation based on optimum parameter
Evaluating the accuracy

Fig. 3. The process of random forest classifier based on genetic algorithm.

of partition frequency m and number of trees n. Then the fitness of the initial community was calculated.

If the result does not meet the end condition of the algorithm, the new community can be generated by selecting crossover function and mutation function until the terminal condition is satisfied, so as to output the optimal solution of parameter combination.

Genetic algorithms use simple coding techniques to represent complex structures, and through the genetic manipulation of reproduction, crossover and mutation generate alternative solution sets. Guided search is carried out through the selection mechanism of survival of the fittest.[3,4,8] GARF uses random forest model to classify samples, and uses permutation method in genetic algorithm to determine the boundary value of feature screening as the basis for the final determination of feature variables. This paper proposes the following steps for sample classification:

(i) Select and normalize meaningful indicators for sample classification. The normalization results are kept within the range of 0. The normalized index is used as the classification;
(ii) Take samples in each training set as training samples of random forest, and randomly select 30% of all samples as test samples (commodity types can be regarded as training samples);
(iii) Use genetic algorithm to optimize initial parameters and establish a clothing commodity classification model based on random forest;
(iv) Use the established model to classify the entire test set and evaluate the classification accuracy with the test results.

When using genetic algorithm for parameter optimization, this paper first finds the optimal number of trees n, and then uses the traditional grid search method to continuously find the combination of the optimal minimum segmentation number and maximum depth through violence calculation. Then the number of trees n and the number of segmentation m are searched by genetic algorithm.

2.3. *Traditional convolutional neural*

Artificial neural network classification is a very common method to solve the problem of image pattern recognition. Neural network is a mathematical model based on interconnected artificial neurons, which is similar to biological neural network. Neurons are organized in layers, and connections are made between neurons that only come from adjacent layers. The input low-level eigenvectors are put into the first layer and converted into high-level eigenvectors. The number of neurons in the output layer is equal to the number of classes in the classification. Therefore, the output vector is the vector representing the probability that the input vector belongs to the corresponding class.

The weighted adder is implemented by the artificial neuron, and its output is described in the formula (4):

$$a_j^i = \sigma\left(\sum_k a_k^{i-1} w_k^{ij}\right), \tag{4}$$

where a_j^i is the first neuron, w_k^{ij} represents the weight of synaptic connections between the j neuron and the $k-1$ neuron. In a wide range of regressing applications, logical functions are used as activation functions. It's worth noting that a single ratification neuron performs the logistic regression function. The training process based on the minimization of sensory function and gradient method is called reverse propagation. In the classification problem, the most commonly used cost function is to cross entropy:

$$H(p,q) = -\sum_i Y(i)\mathrm{log}y(i). \tag{5}$$

when training the neural network with multiple layers (called the depth network), the Sigmoid activation function may become invalid due to the vanishing gradient problem. Therefore, this paper uses ELU function, formula (6) as the activation function:

$$\mathrm{ELU}(x) = \begin{cases} \exp(x) - 1, & x \leq 0 \\ x, & x > 0. \end{cases} \tag{6}$$

Convolution neural network classification is the most prominent pattern recognition method in the field of computer vision in recent years. Unlike traditional neural networks, which work with one-dimensional eigenvectors, convolutional neuron networks work with two-dimensional eigenvectors and process them with convolution layer. Each convolution layer consists of A set of trainable filters and calculates the points between this filters and A to obtain the activation diagram. These filters are also known as kernels and are allowed to detect the same characteristics at different locations.

In this paper, different convolutional neural networks are constructed to solve the problems of commodity classification and recommendation and optimize them.

At the beginning of the experiment based on VGG and imagine enhancement, this paper attempts to establish a four-layer convolutional neural network. The Conv2D layer and Max-pooling layer are used in this paper, and the full-connection layer is used to classify the images at the end. For the middle layer, Relu activation function is used. In the Dense layer, this paper uses SoftMax activation functions to complete the classification work behind many nerve layers.

In the CNN-1 model, the first is a Lambda layer. In this paper, the data of the training set is processed first — each pixel is divided by the standard deviation of the pixel. In convolution neural network, normalization can compensate every weight equalization. The image is then extracted by a 3 by 3 convolution kernel with Relu activation function.

In this paper, it can be seen from Fig. 4 that the effect of the original picture after the convolution of various layers, and the similar characteristics of various types of shoes have been abstractly displayed. The data output from the convolutional layer are processed by Max pooling of 2 * 2 and thermalization layer. Samples are taken from the image to extract part of the data so as to reduce the over-fitting degree of network training parameters and model. Max pooling selects the maximum value in the Polling window as the sampling value, and Mean pooling is not used, because

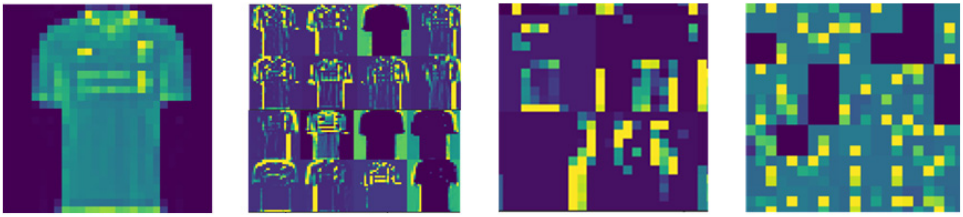


Fig. 4. The original drawing and convoluted images of short sleeves (a) the original drawing (b) the image after one layer of convolution (c) the image after two layer of convolution (d) the image after three layer of convolution.

Max pooling usually performs better.[18] Second, the Dropout layers was added in this paper. Dropout is a technique that can be randomly selected to ignore or "erase" connections between neurons during the training process. The effect of Dropout is that it can make the network less sensitive to the specific weight of neurons. This paper assigns Dropout a value of 2.5%, which indicates that 25% of the neuron connections in training are randomly ignored. Therefore, the network is better generalized to avoid over-fitting data. Finally, the flatten layer and the entire connection layer were added to change the shape of the data to satisfy the predicted output dimension of the flatten layer normalized, the output dimension is equal to the number of categories. In this dataset, the output dimension is 10, and the output value of each dimension is the probability that the test picture belongs to this category.

The first implementation of CNN-1 model structure. The architecture is composed of a Lambda, a convolution layer, a Dropout layer and a flatten layer, which are two completely connected layers (see Table 1 for the specific content). Network depth is related to the data volume.[17] Excessively deep neural networks and scarce data may produce over-fitting models. On the other hand, shallow networks with big data will not provide enough precision. Therefore, it is important to balance network depth and data volume.

When the network is trained in the framework of Table 1, the classification accuracy reaches above 0.9. However, for a training set with 60 000 data and a test with 10 000 data, this paper wants to test whether more layers perform better. This paper, therefore, decided to increase the number of convolution layer, after several failed attempts, the paper constructs the CNN-3 model structure (specific content see Table 2), the main difference is contains three layer convolution of image processing, each convolution layer using the 3*3 convolution kernels, each biggest soft used a 2*2 filter.

Table 1. The structure of CNN-1 method.

Layer	Output Shape	Number of Parameters
Conv2D	(26, 26, 64)	640
Max_Pooling2D	(13, 13, 64)	0
Dense	256	2769152
Dense	10	2570

Table 2. The structure of CNN-3 method.

Layer	Output Shape	Number of Parameters
Conv2d	(26, 26, 32)	320
Max_pooling2D	(13, 13, 32)	0
Conv2d	(13, 13, 64)	18496
Max_pooling2D	(5, 5, 64)	0
Conv2d	(3, 3, 128)	73856
Fully connected-1024	1024	1180672
Fully connected-10	10	10250

In the training of CNN-1 model and CNN-3 mode, 256 batch sizes, namely the number of samples for a training, are used in this paper to find an optimal balance between memory efficiency and memory capacity. Epocho was not enough to pass a complete dataset across a neural network once and back again after 50 epochs, and this paper needed to train the network to pass a complete dataset. Deep Convolution Neural Network Algorithm Combined with Image Enhancement.

2.4. *Deep convolution neural network algorithm combined with image enhancement (VGG-IE)*

VGG model (Table 3) is a very deep convolutional neural network for large-scale image recognition, which is improved based on the Alex-Net model.[11] Compared with the CNN model, VGG has smaller convolution kernel (3*3 size) and deeper and wider layers. With fewer parameters, the small convolution kernel can increase the number of layers without worrying about the computation quantity explosion, thus achieving a higher classification accuracy.[10] In this model, Batch Normalization is added between each hidden layer to "preprocess" input to the next neural layer to reduce the displacement of internal covariables.

The input data of this paper is 28 * 28 images. This paper adjusts part of the architecture based on the size of the put image. The original VGG had five convolution layers, but only two pooling layers were used to simplify the data. Finally, this paper also uses three full connection layers and a Softmax activation function.

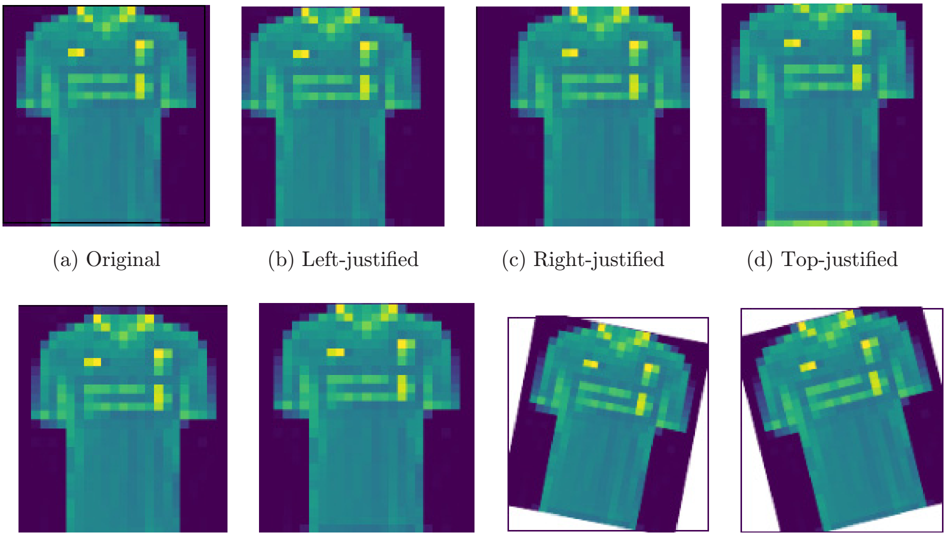
Table 3. The structure of VGG method.

Layer	Output Shape	Number of Parameters
Conv2D	(28, 28, 32)	320
Conv2D	(28, 28, 64)	18496
MaxPooling2D	(14, 14, 64)	0
Conv2D	(14, 14, 128)	73856
Conv2D	(14, 14, 256)	295168
Conv2D	(14, 14, 256)	590080
MaxPooling2D	(7, 7, 256)	0
Fully connected-512	512	6423040
Fully connected-512	512	262656
Fully connected-10	10	5130

Experimental results show that the pre-trained CNN is a very effective method for object image extraction. Alex-Net, which is preprocessed with data focal point containing a large amount of image data, is used for image feature extraction. Multiple database checks that the average error rate of CNN network without image data enhancement is higher than that of CNN network with image data enhancement in the test.[19]

Affine transformation and elastic distortion are widely used in object recognition in image enhancement. They are used to generate new samples from the original samples and expend the training set. Affine transformation is to transform images by generating simple distortions, such as translation, rotation, scaling and tilt. Elastic distortion is an image transformation that imitates stylistic changes in handwriting.

This paper also applies the enhancement methods such as rotation, tilt and elastic deformation that are also applied in the training set to increase the diversity of training images. Figure 5 shows an example of a pattern enhanced by the cosine function. The original pattern uses the cosine function to generate a new image with left adjustment, right adjustment, up adjustment, down adjustment, center adjustment and magnification. Figure 5 shows a new image that is distorted by rotation, tilt and elasticity. The training set enhanced by image can increase the generalization ability of the model and more adapt to the application scenarios in daily life.



(a) Original (b) Left-justified (c) Right-justified (d) Top-justified

(e) Buttom-justified (f) Center-justified (g) Rotation:-19° (h) Rotation:+19°

Fig. 5. Training-data augmentation of Fashion-MNIST object images. Original images are translated and justified to the left, right, top, or bottom by cosine function. Images are also center-justified, enlarged and rotated.

3. Experimental Process

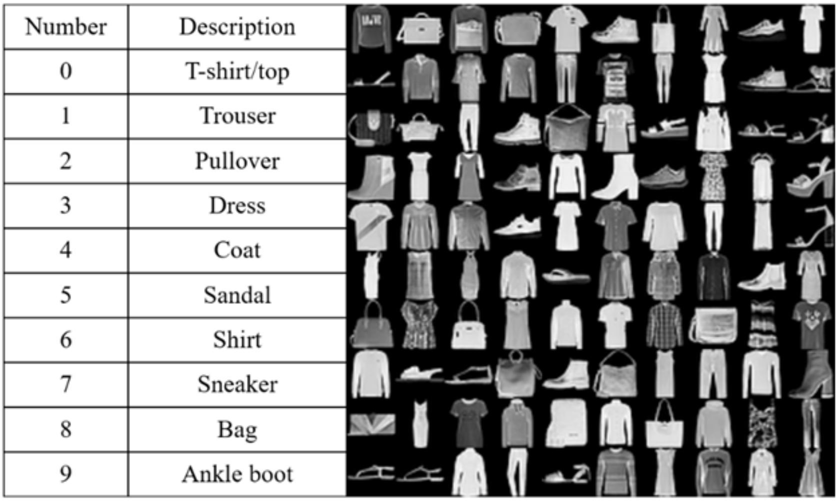
3.1. *Introduction to datasets*

Fashion-MNIST (Fig. 6) provides a total of 70 000 positive images of different products from 10 categories, which is an image dataset used to replace the handwritten digital set of MNIST and provided by the research department of Zalando (German Fashion e-commerce Internet Company). The size, format and training set test set division of Fashion-MINIST are completely consistent with the original MNIST. The training test data are divided into 28 * 28 gray images.[9]

In this paper, this dataset is divided into two sub-datasets, top and bottom, so as to study the comparison of top and bottom classification of different algorithms. Tops include: 0 short sleeved T-shirts, 2 pullovers, 4 coats and 6 shirts; the lower outfit includes: 5 sandals, 7 sneakers and 9 ankle boots. At the same time, this paper uses the image preprocessing software developed by the QT platform to perform convolution transformation on any size image uploaded by users and output a standard image of 28 * 28. This experiment adopts the TensorFlow machine learning framework under Windows. The hardware environment is GPU: Intel Core i7-4710HQ,GPU:NVIDIA GeForce GTX 970, and memory: 16 GB.

3.2. *Qualitative*

Figure 7 for part of the experimental results, shows the use of t-distributed stochastic neighbor embedding (t-SNE), the results of classification results VGG model visualization of each color corresponding Fashion-MNIST a category of MNIST dataset, which can visually distinguish Fashion-MNIST types of classification accuracy. t-SNE is a machine learning algorithm for dimensionality reduction. It is a nonlinear



Number	Description
0	T-shirt/top
1	Trouser
2	Pullover
3	Dress
4	Coat
5	Sandal
6	Shirt
7	Sneaker
8	Bag
9	Ankle boot

Fig. 6. The overview of the dataset.

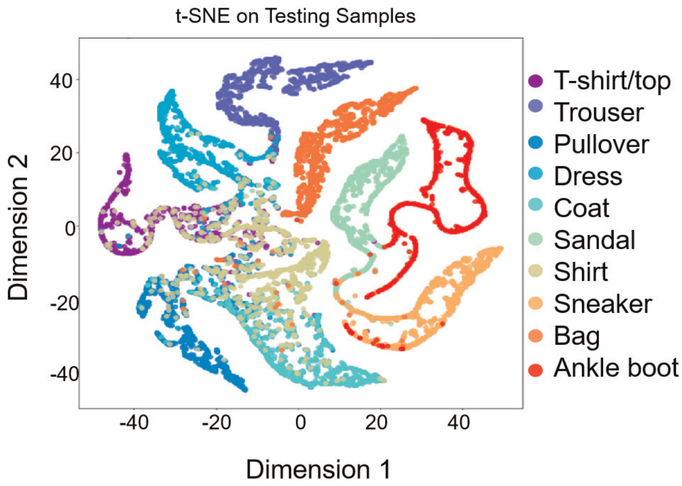


Fig. 7. The *t*-SNE visualization result of Fashion-MNIST datasets.

dimensionality reduction algorithm proposed by Laurens van der Maaten and Geoffrey Hinton in 2008, which is very suitable for the visualization of high-dimensional data from long dimension to 2 dimension or 3 dimension.[15]

Looking at the image, this paper found that the yellow–green points (category 6, shirts) were the most scattered in the figure, with a large degree of coincidence with purple, lake blue and mint green points, and these overlapping categories were all in the jacket dataset. This paper also found that the point coincidence of degree of several colors belonging to the lower assembly dataset was low, which meant that the lower assembly classification accuracy was higher, so this paper guessed that the lower assembly classification accuracy was higher than the upper assembly, which was indeed confirmed in the following experiments.

This paper also tested the performance of different experimental models in the top dataset and the bottom dataset: all the models have a good classification effect on the bottom dataset, while the classification effect on the top is poor. Almost all models are about 10% more accurate in the classification of the lower part than the upper part. CNN-1 model, for example, tops in the classification accuracy is lower than bottoms, while classification loss value is higher than the bottoms (specific content see attachment center 8).

The best VGG model results are as shown in Fig. 8, the confusion matrix, 0, 2 category, category 3, 4 and 6 category compared with other categories has low accuracy, these categories contains all tops dataset where, similar to the *t*-SNE visualization results, the classification accuracy of category 0 and category 6 is only 80%, and there are many types of error mode (see Table 4).

From the experimental results, it can be clearly seen in this paper that a large number of category 6 items are misclassified, and the predicted misclassification is also the top category (see Fig. 10 in the attachment for details).

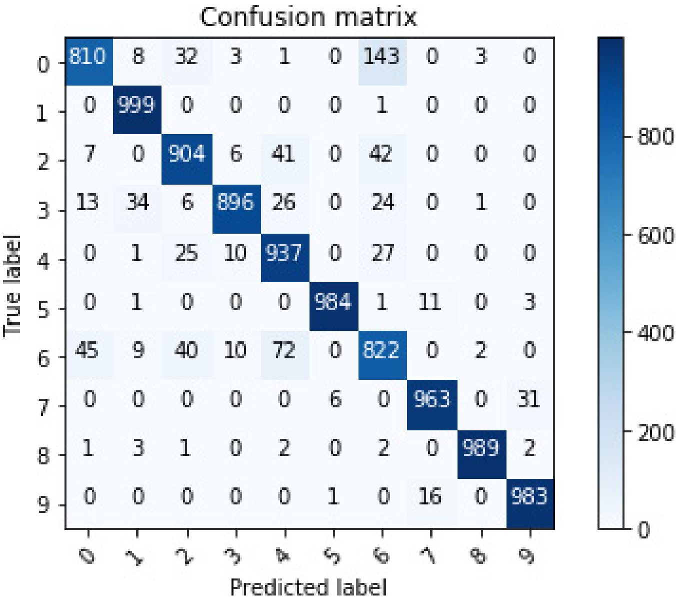


Fig. 8. The confusion matrix of VGG method.

Table 4. The confusion matrix and classification error of VGG method.

	Forecast										
Fact	0	1	2	3	4	5	6	7	8	9	Error rate
0 T-shirt/top	**810**	**8**	**32**	**3**	**1**	**0**	**143**	**0**	**3**	**0**	**19.00%**
1 Trouser	0	999	0	0	0	0	1	0	0	0	0.10%
2 Pullover	7	0	904	6	41	0	42	0	0	0	9.60%
3 Dress	13	34	6	896	26	0	24	0	1	0	10.40%
4 Coat	0	1	25	10	937	0	27	0	0	0	6.30%
5 Sandal	0	1	0	0	0	984	1	11	0	3	2.60%
6 Shirt	**45**	**9**	**40**	**10**	**72**	**0**	**822**	**0**	**2**	**0**	**17.80%**
7 Sneaker	0	0	0	0	0	6	0	973	0	31	2.70%
8 Bag	1	3	1	0	2	0	2	0	989	2	1.10%
9 Ankle boot	0	0	0	0	0	1	0	16	0	983	1.70%
Number of the error classification	66	56	104	29	142	7	240	27	6	36	**7.13%**

This paper expects to find the optimal combination of the number of trees n and the maximum depth m. According to Fig. 8, n is the number of the most focal trees in the random forest classifier. Through the genetic algorithm as shown in Fig. 2, this paper can obtain the optimized values of the number of trees and the maximum depth m.

In this paper, two groups of random forest classifiers, three groups of convolutional neural network models and three groups of convolutional neural network models enhanced by images proposed in this paper are compared with the k-nearest classifier (KNN) proposed in the literature.[22] It is proved that the scheme proposed in this paper can guarantee the running speed and improve the accuracy at the same time.

The algorithm proposed in this paper solves the problem that classification accuracy and classification rate are not compatible, and puts forward two schemes to solve the problem of e-commerce image retrieval.

3.3. *Quantitative evaluation*

As shown in Fig. 9, in the random forest classification, this paper first found the optimal number of trees $n = 260$, and the method looks for the number of variable group trees n and the maximum depth m, and finds that the best combination is that n is equal to 110 and m is qual to 16. When the number of trees is equal to 260, the accuracy on the test set is 88.73%, while the accuracy on the test set of the optimal combination model found by genetic algorithm is 87.37%.

For comparison, Table 5 gives experimental data of 10 models including K nearest neighbor (KNN) and random forest (RF), as well as the classification accuracy of each model in the gold dataset, the upper dataset and the lower dataset, as well as the training time of each model in the gold data set.

The test results of all 10 models (including K nearest neighbor and random forest) are shown in Table 4, which lists the exact test values of convolutional neural network model and the results of these values under the combined model of data enhancement. The results showed that the best performing VGG model had the highest classification accuracy in the gold dataset, upper dataset and lower dataset (93.97%, 87.55% and 96.73%, respectively). Compared with other neural networks, it has an increase of 2–2.5%. The accuracy of the most basic CNN-1 model is 91.59%, which is almost the same as that of the CNN-3 model (91.63%). At the same time, experiment results showed that the accuracy of all models in the test set was improved by 0.2%

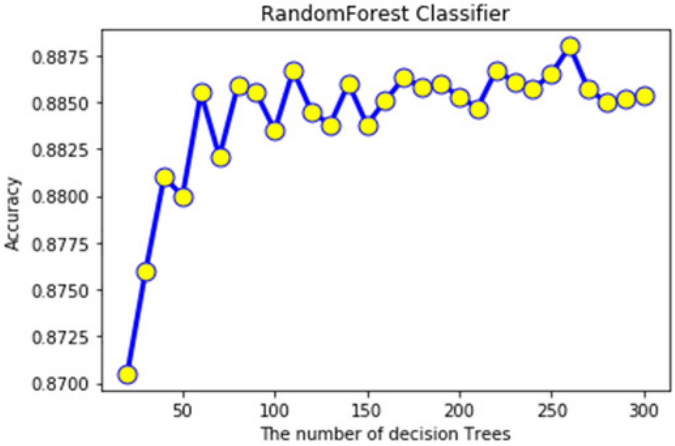


Fig. 9. The relation of the accuracy of the random forest classifier to the number of the tree.

Table 5. The classify accuracy of 10 methods in different datasets.

	The Accuracy of Full Dataset	The Training Time(s)	The Accuracy of Tops Dataset	The Accuracy of Bottoms' Dataset
KNN(Manhattan Distance, $K = 6$)	86.73%	1210	79.27%	92.92%
KNN(Euclidean Distance, $K = 7$)	86.14%	1169	77.91%	90.83%
Random Forest ($n = 260$)	88.67%	288	83.59%	94.63%
GA-RF ($n = 110$, $m = 16$)	87.42%	**73**	81.23%	92.22%
CNN-1	91.39%	2123	85.83%	94.67%
CNN-3	91.63%	2263	86.11%	94.17%
VGG	92.19%	404	85.38%	95.84%
CNN-1 + IE	91.81%	2201	86.40%	95.14%
CNN-3 + IE	91.98%	2398	85.77%	95.38%
VGG + IE	**93.97%**	503	**87.55%**	**96.73%**

to 1.5% after the combination of image generator. Among the 129 algorithms provided by Zalando Research with different algorithm frameworks and different parameters, the highest average classification accuracy is only 89.7%, below the algorithm (http://fashion-mnist.3-website.eu-central-1.amazonaws.com).

Compared with this paper that realize the traditional K nearest neighbor classifier and random forest, one of the best model is a tree of 260 the number of random forest model, its accuracy is 88.67% and the highest accuracy of VGG+ image enhancement model accuracy (93.97%) of lead by close to 6%.

In terms of training time, random forest has an absolute speed advantage. When the number of trees is 110 and the number of segmentation is 16, it only takes 73 s to train once. Although VGG network is deeper and wider for general neural networks and usually requires more training,[18] in order to obtain the highest accurate value in the CNN-1 and CNN-3 models, this paper continuously adjusted the model parameters (steps Per epoch = 600 and epochs = 50), which made them need more time to train. The VGG network (steps per epoch = 48000 and epochs = 10) used less time, but achieved greater accuracy and cost effectiveness.

4. Conclusion

This study proposes two algorithms to improve the accuracy and efficiency of image retrieval in the image recommendation system related to e-commerce search engines based on the Fashion-MNIST dataset. In this study, the existing random forest classifier (RF) and deep convolutional neural network (VGG) were optimized, respectively.

The random forest classifier (GA-RF) based on genetic algorithm used genetic algorithm to optimize the random forest classifier twice, and optimized the number of trees for the first time. The second time the optimal combination of the number of trees and the number of partitions is optimized. This algorithm optimizes two important parameters of the random forest classifier by genetic algorithm, and enhances the ability of the random forest classifier. In particular, in the second set of experimental data, the optimized random forest achieved the shortest training time (73 s).

The deep convolutional neural network (VGG-IE) combined with image enhancement uses the pre-trained VGG model for image extraction. The algorithm combines the improvement of recognition ability of neural network by image enhancement with the high efficiency and accuracy of deep convolutional neural network VGG. While the highest accuracy (93.97%) was obtained in the experimental group, the training time was only about half of K's nearest classifier and a quarter of the series model.

The results show that the two algorithms proposed in this paper either consume the shortest training time or achieve the highest. Of course, there are still some problems with the algorithm in this paper. For example, the content of the Fashion-MNIST dataset is relatively single, which cannot fully represent the existing e-commerce fashion related products. For the 10 algorithms, the classification accuracy of the upper dataset is lower than that of the lower one. How to further improve the classification accuracy of random forest classifier based on genetic algorithm still needs to be further studied and solved. In the future, robust algorithm[6,7] in complex clothing scenes with depth uncertainty should also be considered, which is expected to improve the classification accuracy in the context of big data.

Acknowledgments

This work was supported in part by the Fundamental Research Funds for the Central Universities (Grant Nos. JZ2018HGBZ0178, JZ2018HGBZ0177, JZ2018HGTA0219 and JZ2019HGBZ0149), the National Nature Science Foundation of China (No. 61501159), Natural Science Foundation of Anhui Province (1608085MF124), Maker laboratory of Hefei University of Technology (Grant Nos. 2018cxsys17 and 2018cxsys19).

References

1. G. Adomavicius and A. Tuzhilin, Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, *IEEE Trans. Knowl. Data Eng.* **17**(6) (2005) 734–749.
2. S. Bhatnagar, Classification of fashion article images using convolutional neural networks, *2017 Fourth Int. Conf. Image Information Processing* (Shimla, India, 2017), pp. 1–6.
3. L. Breiman, Bagging predictors, *Mach. Learn.* **24** (1996) 123–140.
4. L. Breiman, Random forests, *Mach. Learn.* **45** (2001) 5–32.
5. Z. Falin, Z. Tao and L. Kang, Application of random forest model based on Genetic Algorithm in screening of characteristic genes, *Chin. Health Stat.* **33**(4) (2016) 559–562.

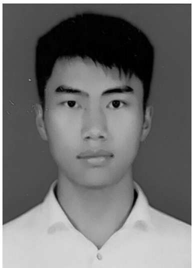
6. L. Gao *et al.*, Robust global sensitivity analysis under deep uncertainty via scenario analysis, *Environ. Model. Soft.* **76** (2016) 154–166.
7. L. Gao and B. A. Bryan, Incorporating deep uncertainty into the elementary effects method for robust global sensitivity analysis, *Ecol. Model.* **321** (2016) 1–9.
8. P. O. Gislason, J. A. Benediktsson and J. R. Sveinsson Random forests for land cover classification, *Pattern Recognit. Lett.* **27** (2006) 294–300.
9. X. Han, K. Rasul and R. Vollgraf, Fashion-MNIST: A Novel Image Dataset for Benchmarking Machine Learning Algorithms, arXiv:1708.07747.
10. A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Communications of the ACM* **60**(6) (2017) 84–90.
11. F. Lauer, C. Y. Suen and G. Bloch, A trainable feature extractor for handwritten digit recognition, *Pattern Recognit.* **40**(6) (2007) 1816–1824.
12. L. Lihui and C. Ming, Application of content-based image retrieval in E-commerce, *J. Jilin Normal Univ.* **33**(3) (2012) 86–89.
13. Z. Linting and S. Tao, First exploration of application of content-based image retrieval in E-commerce, *Shopping Malls Modern.* **33** (2007) 138–139.
14. J. Liu, K. Hao, Y. Ding, L. Gao and S. Yang, Multi-state self-learning template library updating approach for multi-camera human tracking in complex scenes, *Int. J. Pattern Recognit. Artif. Intell.* **31**(12) (2017) 1755016.
15. N. Pezzotti and L. van der Maaten, Approximated and user steerable tSNE for progressive visual analytics, *IEEE Trans. Visual. Comput. Graph.* **23**(7) (2017) 1739–1752.
16. D. Yong, 2017 total retail survey "content + data" wrestling China's retail industry, *Shanghai Business* **6** (2017) 24.
17. Y. Shima, Image augmentation for object image classification based on combination of pre-trained CNN and SVM, *J. Phys.* **1004** (2018) 1–8.
18. A. Shustanov and P. Yakimov, CNN design for real-time traffic sign recognition, *Procedia Eng.* **201** (2017) 718–725.
19. K. Simonyan and A. Zisserman Very deep convolutional networks for large-scale image recognition, *ICLR 2015* (San Diego, CA, USA, 2015), pp. 1–8.
20. D. Sorokina and E. Cantu-Paz, Amazon search: The joy of ranking products, in *Proc. 39th Int. ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR' 16)* (Pisa, Italy, 2016), pp. 459–460.
21. P.-N. Tan, M. Steinbach and V. Kumar, *Introduction to Data Mining* (People's Post and Telecommunications Press, Beijing, 2010), pp. 127–187.
22. L. Xinhai, Application of random forest model in classification and regression analysis, *J. Appl. Entomol.* **50**(4) (2013) 1190–1197.
23. Y. Yiwei, C. Yifu and D. Jian, Research on the current situation and problems of College Students' online shopping in Shanghai, *BUSINESS* **1** (2016) 286–287.

Jian Liu is currently a Lecturer at the School of Computer Science and Information Engineering, Hefei University of Technology, Anhui, China. He obtained the B.S. and Ph.D. degrees in Electrical Engineering from Donghua University, Shanghai, China in 2013 and 2017, respectively. From 2015 to 2016, he was a Visiting scholar at Commonwealth Scientific and Industrial Research Organisation, AU, Australia. He has published more than 15 technical papers. His scientific interests include machine vision, image processing, data science and NLP.

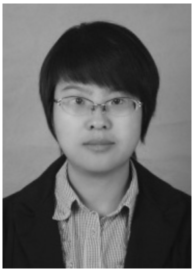
Yuchen Zheng is currently an Sophomore Student at the Department of Computer Science and Information Engineering, Hefei University of Technology (HFUT), Anhui, China. Her research interests include machine vision, machine learning, image processing and data science.

Ke Dong is currently an Sophomore Student at the Department of Computer Science and Information Engineering, Hefei University of Technology (HFUT), Anhui, China. Her research interests include machine vision, machine learning, image processing and data science.

Haitong Yu is currently pursuing M.S. of Software Management in Carnegie Mellon University, CA, USA. He obtained the B.S. in Computer Science from Hefei University of Technology, Anhui, China in 2018. From 2015 to 2018, as an exchange student, he obtained the B.S of Computer Science in George Mason University, VA, USA. His research fields include data science, NLP, and computer vision.



Jianjun Zhou is currently an Sophomore Student at the Department of Computer Science and Information Engineering, Hefei University of Technology (HFUT), Anhui, China. His research interests include machine vision, machine learning, image processing and data science.



Ye Jiang is currently a Lecturer at School of Computer Science and Information Engineering, Hefei University of Technology, Anhui, China. She obtained the B.S. and M.S. degrees in Communication Engineering from China University of Mining and Technology in 2010 and 2013, respectively, and her Ph.D. degree in Communication and Information System from Chinese Academy of Sciences in 2017. She has published more than seven technical papers. Her scientific interests include machine learning and intelligence information processing.

Zhaoneng Jiang is currently an Associated Professor at the Department of Computer Science and Information Engineering, Hefei University of Technology (HFUT). He received the B.S. degree in Physics from Huaiyin Normal College, Huai'an, Jiangsu, China, in 2007, and the Ph.D. degree in Electrical Engineering from Nanjing University of Science and Technology (NJUST), Nanjing, China, in 2013. He has published more than 60 technical papers. His research interests include computational electromagnetics, antennas and electromagnetic scattering and propagation, and electromagnetic modeling of microwave integrated circuits.

Sujie Guo is currently a junior at the Department of Computer Science and Information Engineering, Hefei University of Technology (HFUT), Anhui, China. His research interests include machine vision, machine learning, image processing, data science.

Rui Ding is currently a junior at the Department of Computer Science and Information Engineering, Hefei University of Technology (HFUT), Anhui, China. His research interests include machine vision, machine learning, image processing, data science.