



# Image Classification Based on Improved Unsupervised Clustering Algorithm

Yichao Wang<sup>1</sup>(✉), Chunzhi Wang<sup>2</sup>, and Lingyu Yan<sup>2</sup>

<sup>1</sup> School of Computer Science and Technology, HanKou University, Wuhan, China  
echarwang@163.com

<sup>2</sup> School of Computer Science, Hubei University of Technology, Wuhan, China  
yanlingyu@hbut.edu.cn

**Abstract.** This paper proposes a k-means model based on density weighting, which is applied to the field of image classification and fused with deep neural network to train pseudo-labels. While clustering the learning features of the residual network, the network parameters are updated to achieve. The clustering performance of pseudo-labeled datasets is improved to solve the problem of scarcity of labeled data.

**Keywords:** K-means · Image Classification · Deep Neural networks · Clustering Algorithm

## 1 Introduction

The rapid development of information technology has brought a large amount of unlabeled data, many research subjects have shifted from all labeled data to only a small amount of labeled or unlabeled data, and research methods have shifted from supervised learning to unsupervised learning, which has more practical significance for image processing tasks. Clustering is an important unsupervised learning method that is widely used in machine learning and data mining. Clustering results depend on the clustering algorithm and the feature representation of the data. In the traditional supervised learning task, when the labeling data is insufficient, the generalization ability of the learning model is not strong and cannot meet the requirements of reality. In this technical context, deep clustering as a new unsupervised technology has been proposed, so as to achieve a new clustering method for large-scale end-to-end convolutional networks for a labelless unsupervised network model training, provided that there is a large amount of labeled data to prevent model overfitting, however, manual labeling data is a time-consuming, laborintensive, cost-consuming process, in order to better use existing unlabeled images, clustering and unsupervised learning has attracted great attention and interest from the academic community.

## 2 An Unsupervised Classification Model Based on Improved Kmeans Clustering Algorithm

### 2.1 Unsupervised Clustering Algorithm

The K-means algorithm is one of the most commonly used traditional clustering algorithms, which divides a given sample dataset into K user-specified classes. K sample data are randomly selected as the initial cluster center from N sample data, while for other data samples, they are assigned to the class with the highest or closest similarity based on their similarity or distance from the selected cluster center point. Then, calculate the average of the sample data in each class to update the cluster center point, repeating the process until the standard function J begins to converge.

$$J = \sum_{i=1}^k \sum_{j=1}^{n_k} (C_i - X_j)^2 \quad (1)$$

where: J denotes the sum of the average errors of the data sample objects in all classes,  $C_i$  denotes the cluster centers in the i-th class,  $X_j$  denotes the sample objects in the j-th class.

K-means algorithm steps:

Algorithm input: Sample dataset X,  $X = \{X_m\}_{m=1}^n$ , number of clusters K.

Algorithm output: Clustering represents set C,  $C = \{C_i\}_{i=1}^k$ .

Step1: From dataset X, arbitrarily select K sample data objects as the center of the initial cluster cluster.

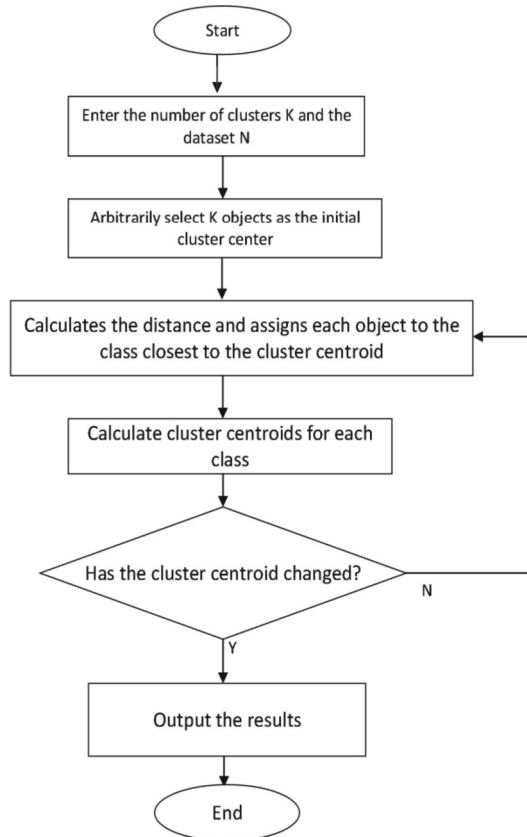
Step2: Calculate the distance from  $x_m$  of each sample in the sample dataset to the center point of the cluster  $c_i$  using the formula  $\text{dis}(x_m, c_i) = \sqrt{(x_m - c_i)^2}$ .

Step3: Find the minimum distance from each data object  $x_m$  to the cluster center  $c_i$   $\min\_dis(x_m, c_i)$ , and classify the data object  $x_m$  into the same class as  $c_i$ , that  $C_i = \{x_m : \text{dis}(X_m - C_i) < \text{dis}(X_m - C_j)\}$ .

Step4: Calculates the mean of objects in the same class, updates the cluster center.

Step5: Repeat steps Step2-Step4 until all cluster centers no longer change or the maximum number of runs is reached.

The flow chart of the K-means algorithm is as follows (Fig. 1):



**Fig. 1.** K-means algorithm flowchart

## 2.2 Canopy Algorithm

The Canopy algorithm is an unsupervised preclustering algorithm introduced by Andrew McCallum, Kamal Nigam, and Lyle Ungar in 2000 [1], and is often used as a preprocessing step for the K-means algorithm. As shown in Fig. 2, the Canopy algorithm sets two distance thresholds  $T_1$  and  $T_2$ , randomly selects the initial cluster center, and calculates the Euclidean distance between the sample and the initial center. Classify samples into corresponding clusters based on thresholds. Finally, the clustered dataset is divided into  $n$  clusters. The clustering of the dataset is completed by taking the cluster number and cluster center of the canopy algorithm as the input parameters of the K-means algorithm.

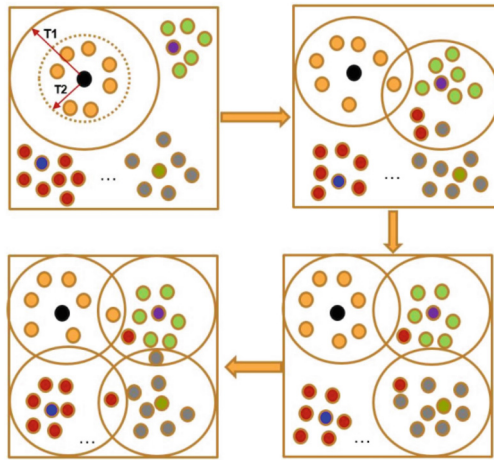
The steps of the Canopy algorithm are as follows:

Step 1: Given a dataset and quantify it, then set thresholds  $T1$  and  $T2$  ( $T1 > T2$ ).

Step 2: Randomly select a data sample point  $S$  from the data set  $D$ , and calculate the Euclidean distance  $d$  between the remaining data sample points in the data set  $D$  and the sample point  $S$  respectively. If there is  $d < T1$ , the data samples that meet the conditions will be Points are added to the current Canopy layer.

Step 3: Then compare the distance  $d$  with  $T2$ . If there is a condition  $d < T2$ , the sample points that satisfy the condition will be deleted from the data set  $D$ , so they will not be added to other Canopy layers.

Step 4: Repeat steps 2 and 3 until dataset  $D$  is empty.



**Fig. 2.** Canopy algorithm schematic

In the classic Canopy algorithm, the threshold is randomly selected in the algorithm, which has a great impact on the clustering results. In this section, the maximum weight product method is proposed to determine the optimal number of clusters, which reduces the instability caused by randomness and improves the clustering accuracy. The maximum weight product method is shown in Fig. 3:

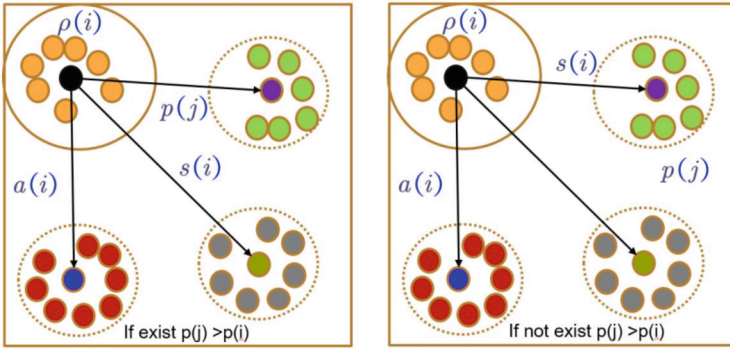


Fig. 3. Maximum weight

$\rho(i)$  denotes the density value of the sample element  $i$  in the data set  $D$ ,  $s_i$  denotes the cluster distance, the schematic diagram of obtaining the maximum weight of the best cluster center is shown in Fig. 4:

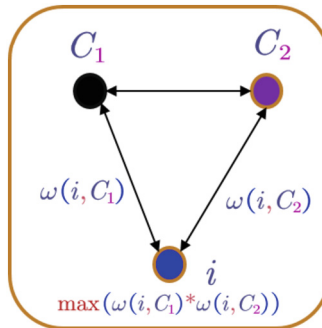


Fig. 4. Schematic diagram of the maximum weights of the largest cluster centers

### 2.3 Improved K-Means Algorithm Based on Density-Weighted Canopy Algorithm

This paper proposes an improved K-means algorithm based on the idea of density weighting. This paper proposes a density weighting method to solve this problem. In addition, the number of clusters and the initial cluster centers obtained by the density-weighted Canopy algorithm are used as input parameters of the k-means algorithm to complete the calculation of the clusters of the dataset. The improved algorithm flow is as follows (Fig. 5):

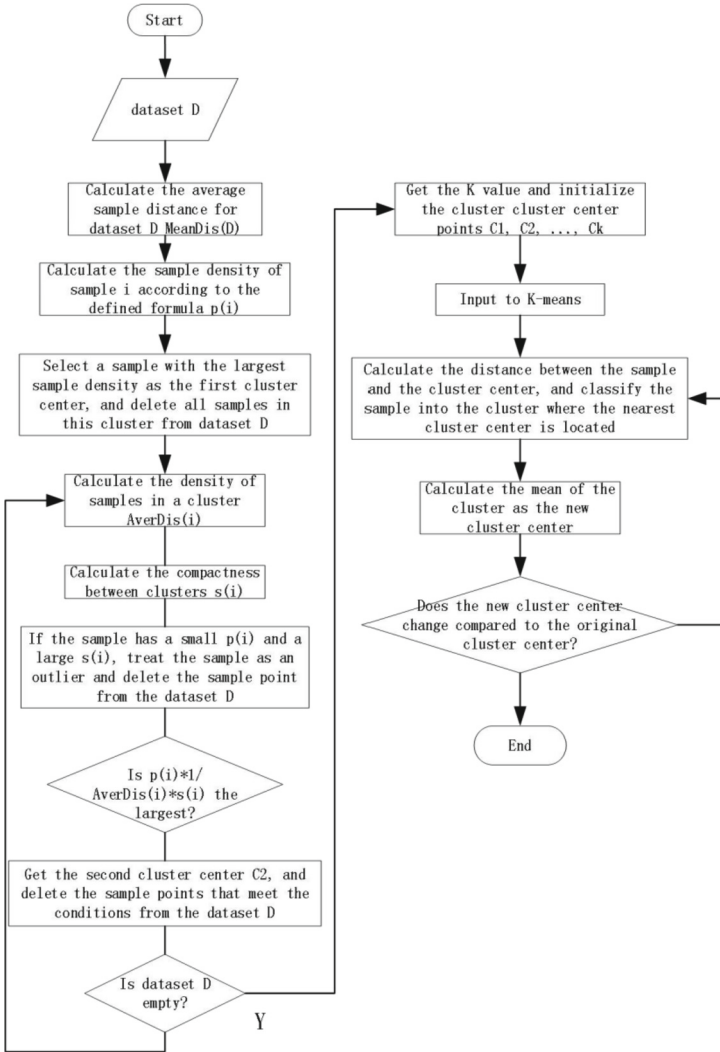


Fig. 5. Improved algorithm flowchart

### 3 Image Classification Combined with Residual Network Resnet Models

#### 3.1 Unsupervised Classification Model Design

For the traditional deep convolutional neural network,  $f_{\theta^*}$  denotes the mapping of the residual network from the original dataset image to a specific dimensional vector space, where  $\theta$  is corresponding parameter set, map this parameter set to the image of the ImageNet dataset, you can get the feature vector of the image information in the dataset for characterization learning, for the N images in the training set  $X = \{x_1, x_1, x_1, \dots, x_n\}$ ,

we hope to find a parameter  $\theta^*$ , so that mapping  $f_{\theta^*}$  produces better visual general features, each image  $x_n$  is associated with a label  $y_n$  in  $\{0, 1\}^k$ , and then the parameterized classifier  $g_w$  predicts which of the image rate belongs to the correct label based on the visual feature  $f_{\theta}(x_n)$ , so the loss function at this time can be denotes by (2) and (3):

$$L = \frac{1}{N} \min_{\theta, w} \delta(g_w(f_{\theta}(x_n)), y_n) \quad (2)$$

$$\delta = -\frac{1}{N} \sum_{n=1}^N \log(p_n, I_n) \quad (3)$$

$p_n$  denotes the prediction probability that the sample belongs to each class,  $I_n$  denotes the true class of the sample data. The unsupervised classification model is optimized based on the loss function in the process of minimizing network training, and the accuracy of the model is inversely proportional to the size of the loss function.

Based on the improved k-means algorithm, the feature  $f_{\theta}(x_n)$  generated by the residual network is used as the input of the clustering algorithm, and the generated feature matrix vector is subjected to dimension reduction processing. Finally, the clustering algorithm divides them into k categories according to the corresponding geometric criteria. Formula (4) minimizes it, and jointly learns the cluster center matrix and the clustering result of each image.

$$p = \min_{c \in R^{d \times k}} \frac{1}{N} \sum_{n=1}^N \min_{y_n \in \{0, 1\}^k} \|f_{\theta}(x_n) - C y_n\|_2^2 \quad (4)$$

The results of clustering are used as pseudo-labels to optimize the cluster loss function, and the classifier parameters and mapping parameters are learned together to achieve the ultimate goal of updating the network parameters.

## 4 Simulation

### 4.1 Experimental Dataset

The datasets used in the experiments are cifar-10, ImageNet and Pascal VOC2007, of which the first two are datasets for image classification tasks, and the Pascal-VOC dataset is a target detection dataset.

### 4.2 Evaluation Indicators

This paper uses Accuracy (ACC) and Normalized Mutual Information (NMI) to measure the suitability of clustering results for unsupervised classification. If the total number of data sets is  $N$ , the real label mapped by each data is  $h_i$ , and the class label obtained by the unsupervised model is  $g_i$ , then the function  $\text{map}(g_i)$  that maps the class label obtained by unsupervised learning to the real label can be obtained, and the accuracy rate The ACC formula is as follows:

$$\text{ACC} = \frac{\sum_{i=1}^n \delta(h_i, \text{map}(g_i))}{N} \quad (5)$$

$\delta$  is a mapping association function that calculates the matching degree of  $h_i$  and  $map(g_i)$ , and its function expression is:

$$\delta = \begin{cases} 0, & h_i = map(g_i) \\ 1, & h_i \neq map(g_i) \end{cases} \quad (6)$$

In addition, this paper measures the information shared between two different assignments A and B between the same data sample by normalized mutual information (NMI), which is defined as formula (7):

$$NMI(A; B) = \frac{I(A; B)}{\sqrt{H(A)H(B)}} \quad (7)$$

I denotes mutual information and H denotes entropy, this performance measure can be applied to any cluster assignment between clusters or ground truth labels. The value of NMI varies continuously between 0 and 1. If the two clusters A and B are completely independent and identically distributed, it means that the  $NMI = 0$ . If the similarity between the two clusters is higher, it means that the value of the NMI is larger, but it is always less than 1.

Considering this the research is a multi-classification problem. In order to evaluate the unsupervised classification model more objectively and fairly, it is necessary to use a unified parameter index to evaluate the model. Before introducing the evaluation index, first give the concept of confusion matrix as shown in the Table 1:

**Table 1.** Confusion matrix

Classification result Reality	True	False
True	TP	FN
False	FP	TN

Based on the confusion matrix, this paper aims at image classification. The evaluation indicators mainly use Precision and Average Precision (AP) to judge the accuracy of the experimental classification results, which are defined as follows:

$$P = \frac{TP}{TP + FP} \quad (8)$$

Precision (P) refers to the ratio of the number of correctly classified positive samples to the number of all classified positive samples, and its calculation formula is as follows:

$$mAP = \frac{1}{C} \sum_{q \in C} AP(q) \quad (9)$$



### 4.3 Experimental Results and Analysis

In this paper, the unsupervised pre-training model on ImageNet is transferred to the PascalVOC dataset, and multilabel classification is realized by fine-tuning. The following figure shows the schematic diagram of the bottom entropy classification result (Fig. 6).



**Fig. 6.** Bottom entropy classification result visualization

In this paper, the feature map visualization of the convolutional layers of Conv1 to Conv5 is carried out to verify that the improved algorithm can promote the feature extractor of the residual network. After each Conv\_x, the visualization operation of the feature map is carried out. This paper Select the first 12 feature maps for visualization. This article randomly selects an original image, the original image is shown in Fig. 7, and the effect image is shown in Figs. 8, 9, 10, 11 and 12.



**Fig. 7.** Original image

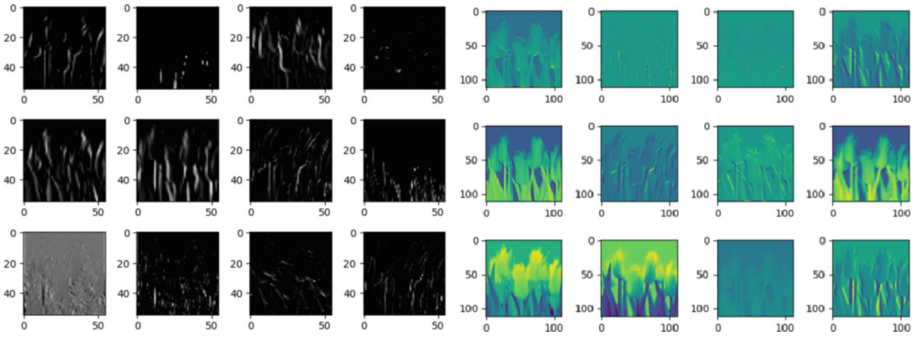


Fig. 8. Feature comparison map of Conv1\_x before and after improvement

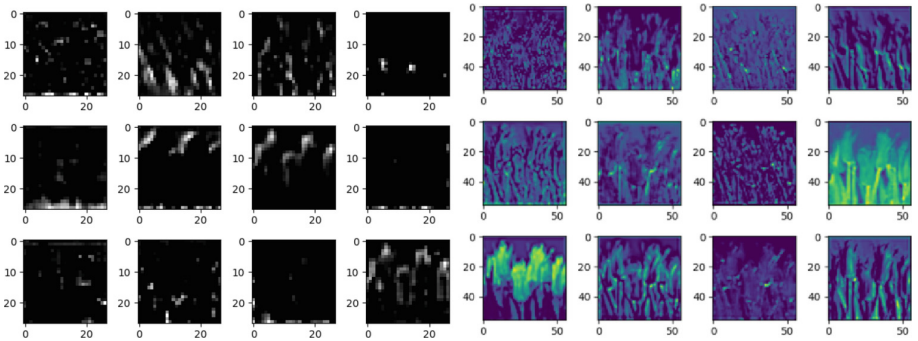


Fig. 9. Feature comparison map of Conv2\_x before and after improvement

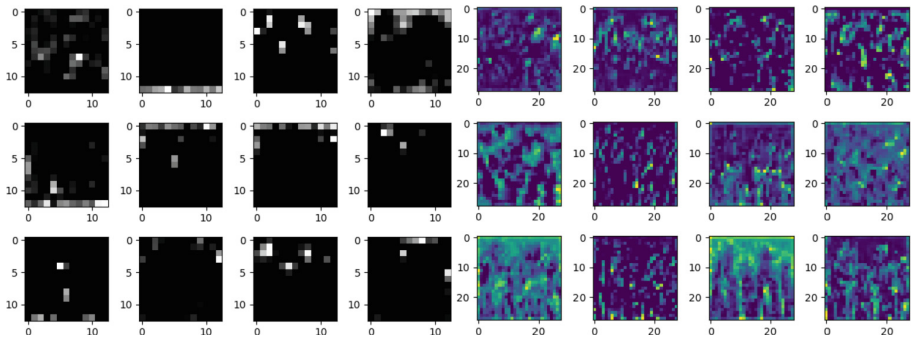
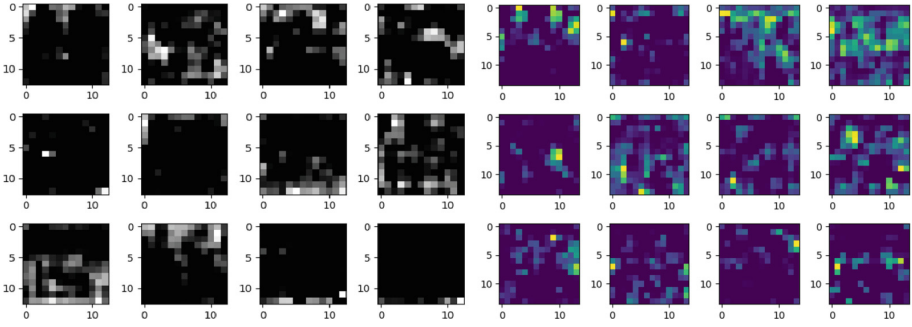
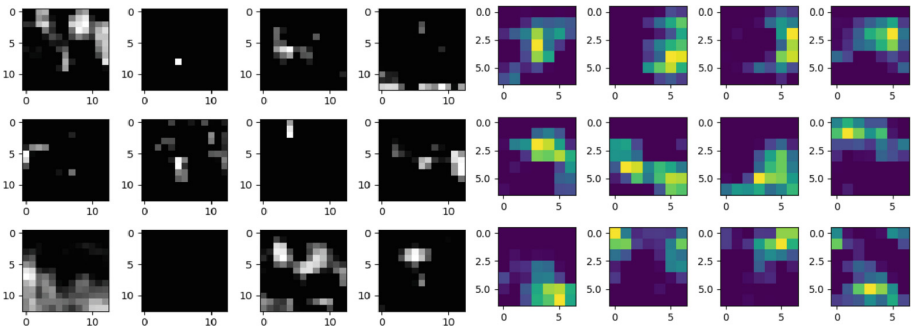


Fig. 10. Feature comparison map of Conv3\_x before and after improvement



**Fig. 11.** Feature comparison map of Conv4\_x before and after improvement



**Fig. 12.** Feature comparison map of Conv5\_x before and after improvement

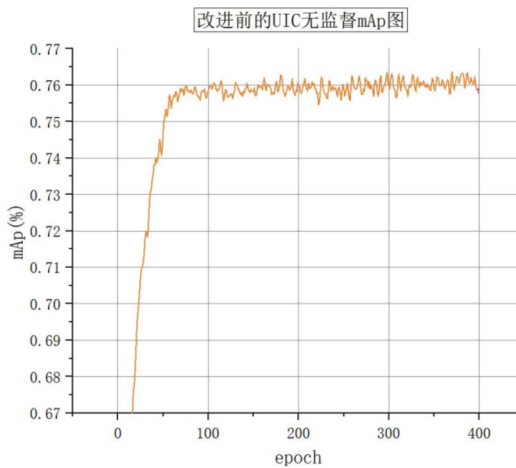
It can be seen from the figure that as the depth of the convolution layer deepens from Conv1\_x to Conv5\_x, the features of the image extracted by the convolution filter also become abstract. By comparing the five left and right pictures, we can see that the five pictures on the right show the ability of the convolutional layer to extract features from the image is obviously better than the ability of the convolutional filters of the five images on the left to extract features from the image. It can be seen that the improved k-means algorithm proposed in this paper is effective for the convolutional filters. To a certain positive effect, the ability to extract feature information is improved, so it also reflects that the algorithm can improve the prediction ability of classification.

From Table 2, we can see that on the ImageNet dataset, the features extracted by the network model in this paper have excellent performance values from Conv3\_x to Conv5\_x through the linear detection classifier., but as the convolutional layer deepens, this gap narrows.

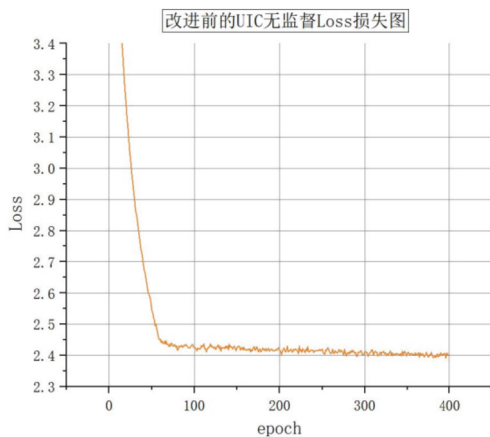
The unsupervised pre-training model on ImageNet is transferred to the PASCAL VOC2007 dataset, and multi-label classification is performed by fine-tuning. The pre-training parameters are set as follows: batchsize is 256, learning rate lr is 0.001, weight decay is set to 1, using 4 GPUs are used for pre-training, and the experimental results are as follows (Figs. 13 and 14):

**Table 2.** ImageNet’s linear detection evaluation table

Methods	ImageNet				
	Conv1_x	Conv2_x	Conv3_x	Conv4_x	Con5_x
ImageNet labels	19.3	36.3	44.2	48.3	50.5
Random	11.6	17.1	16.9	16.3	14.1
DeepCluster [2]	13.4	32.3	41.0	39.6	38.2
SelfLabel3kx1 [3]	---	---	43.0	44.7	40.9
SelfLabel3kx10 [3]	22.5	37.4	44.7	47.1	44.1
Contentxt [4]	15.3	35.2	43.5	45.6	46.7
BiGan [5]	16.2	23.3	30.2	31.7	29.6
Split-brain [6]	17.7	24.5	31.0	29.9	28.0
Jigsaw [7]	17.7	29.3	35.4	35.2	32.8
RotNet [8]	18.2	28.8	34.0	33.9	27.1
AND [9]	18.8	31.7	38.7	38.2	36.5
AET [10]	15.6	27.0	35.9	39.7	37.9
RetNet + retrieval [11]	19.3	35.0	44.0	43.6	42.4
UIC(ours)	12.8	34.3	41.6	41.5	45.3

**Fig. 13.** mAp accuracy map of the unsupervised model before improvement on the VOC dataset.

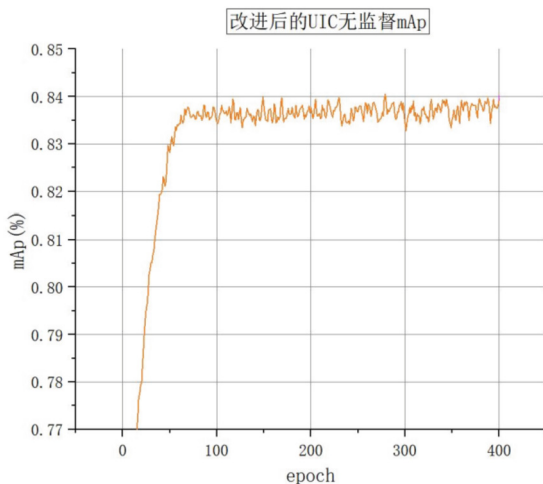
It can be seen from the figure that when the model is trained for 100 epochs, the model begins to converge. Driven by the pre-trained model, the average classification accuracy mAp of the unsupervised classification model on the PascalVOC2007 validation set



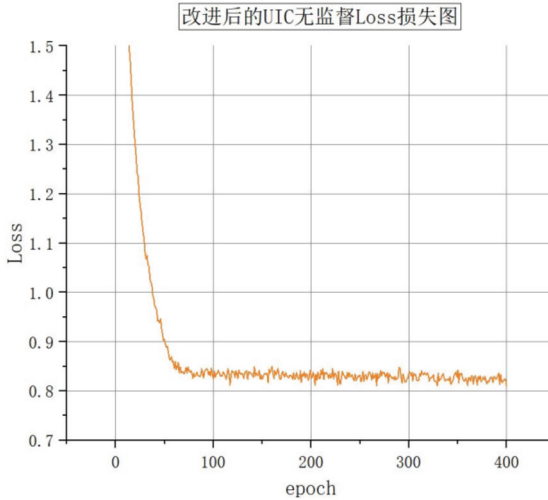
**Fig. 14.** Loss plot of the unsupervised model before improvement on the PASCAL VOC dataset

is 76.3%, and the unsupervised classification on the PascalVOC2007 training set. The training loss is close to 2.4.

The improved unsupervised clustering algorithm is applied to the unsupervised image classification model. The pre-trained model is used. The batchsize is set to 256, the learning rate  $lr$  is set to 0.001, the weight decay is set to 1, and the number of epoch training iterations is set to 400, the mAp obtained by the improved unsupervised classification model trained on the VOC2007 validation set and the loss loss map trained by the pre-training model are as follows (Figs. 15 and 16):



**Fig. 15.** mAp plot of the improved unsupervised model on the VOC dataset.



**Fig. 16.** Loss map of the improved unsupervised model on the VOC dataset

It can be seen from the figure that the model tends to converge after 100 epochs. Recently, under the iteration of 400 epochs, the final model classification accuracy mAP value is 83.9%, and the improved loss loss graph is also close to about 0.8.

## 5 Conclusion

Based on the unsupervised classification model, this paper improves the unsupervised clustering algorithm, and combines the residual network to obtain an improved unsupervised image classification model, which is trained on the ImageNet dataset without labels, and the learned feature representation Ability to transfer to the Pascal VOC dataset for multi-label classification, fine-tune based on the pre-trained model on the VOC dataset to verify the accuracy of the PASCAL VOC2007 validation set, mAP values of the unsupervised classification model before improvement and unsupervised classification after improvement The mAP values of the model on the validation set are 76.3% and 83.9%, respectively, indicating that the improved algorithm proposed in this paper has certain feasibility to improve the performance of the unsupervised classification model.

**Acknowledgment.** This work is funded by the National Natural Science Foundation of China under Grant No. 61772180, the Key R & D plan of Hubei Province(2020BAB012).

## References

1. McCallum, A., Nigam, K., Ungar, L.: Efficient clustering of high-dimensional data sets with application to reference matching. In: Proceeding of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 494–499 (2000). <https://doi.org/10.1145/347090.347123>

2. Caron, M., Bojanowski, P., Joulin, A., Douze, M.: Deep clustering for unsupervised learning of visual features (2018)
3. Asano, Y., Rupprecht, C., Vedaldi, A.: Selflabelling via simultaneous clustering and representation learning (2020)
4. Doersch, C., Gupta, A., Efros, A.: Unsupervised visual representation learning by context prediction (2015). <https://doi.org/10.1109/ICCV.2015.167>
5. Donahue, J., Krähenbühl, P., Darrell, T.: Adversarial feature learning (2016)
6. Zhang, R., Isola, P., Efros, A.: Split-brain autoencoders : unsupervised learning by cross-channel prediction, 645–654 (2017). <https://doi.org/10.1109/CVPR.2017.76>
7. Noroozi, M., Favaro, P.: Unsupervised learning of visual representations by solving Jigsaw puzzles (2016)
8. Gidaris, S., Singh, P., Komodakis, N.: Unsupervised representation learning by predicting image rotations (2018)
9. Huang, J., Dong, Q., Gong, S., Zhu, X.: Unsupervised deep learning by neighbourhood discovery (2019)
10. Zhang, L., Qi, G.-J., Wang, L., Luo, J.: AET vs AED: unsupervised representation learning by auto- encoding transformations rather than data, 2542-2550 (2019). <https://doi.org/10.1109/CVPR.2019.00265>
11. Feng, Z., Xu, C., Tao, D.: Self-supervised representation learning by rotation feature decoupling (2019)