Dimension-reduction fine-tuning method for image classification of few samples

Xueyang Zhang¹, Longlong Xiao, Yajun Yang and Shengdong Pan
Space Engineering University, Beijing, China, 101416
¹E-mail: zxy1135@qq.com

Abstract. The image classification problem of few samples is studied. The fine-tuning image classification method of deep convolutional neural network is analyzed, and it is pointed out that the image feature vector extracted by deep convolutional neural network can be projected into low-dimensional space, which increases the sample density. Based on this, a dimension-reduction fine-tuning method for image classification problem of few samples is proposed, which uses the principal component analysis algorithm to reduce the dimension of image features extracted by the pre-trained convolutional neural network, and then trains the classifier with the dimension-reduced feature vectors. This method improves the classification accuracy when the number of samples is very small, and is analyzed and verified by simulation.

1. Introduction

Image classification refers to the process of identifying an object in a picture with only one primary object. Since Alex proposed the AlexNet [1] of 8 layers in 2012 and won the ILSVRC 2012 image classification competition champion, deep convolutional neural networks have been widely used in image classification. A large number of practices [2-6] have proved that deep convolutional neural networks trained by massive data extract features with better robustness to morphological diversity, illumination variation diversity and background diversity than SIFT and HOG features of traditional computer vision.

Since the training of deep convolutional neural networks requires massive samples, the training results of deep convolutional neural networks are easily over-fitting [7] when the number of samples is small which are formed by the objects to be identified. In order to solve this problem, existing methods mostly adopt the "fine-tuning" transfer learning method [6, 8-10]. Transfer learning refers to adapting a well-trained model to a new problem by simple adjustment. It can be divided into four categories: sample-based transfer learning, parameter-based transfer learning, feature-based transfer learning, and relationship knowledge-based transfer learning [11]. In the field of image classification based on deep convolutional neural networks, the most widely used transfer learning method is the third category, and is generally called the “fine-tuning” method. But the existing "fine-tuning" classification method has poor recognition ability and low classification accuracy when the number of samples is very small.

Compressed sensing theory believes that high-dimensional data in reality has the characteristics of sparse low rank [12, 13], and the same is true for the image features extracted by convolutional neural networks. Each category of image has only a small dimension in the feature space. When the sample
number is too small, it is more susceptible to noise. Consider reducing the data dimension to increase the sample density and reduce the influence of noise.

Principal component analysis is a commonly used data dimension reduction method [14]. In this paper, a dimension-reduction fine-tuning method is proposed for image classification problem of few samples. This method uses the principal component analysis algorithm to reduce the dimension of image features extracted by the pre-trained convolutional neural network, and then trains the classifier with the dimension-reduced feature vectors. This method improves the classification accuracy when the number of samples is very small for increasing the sample density and reducing the influence of noise.

2. Fine-tuning image classification method of deep convolutional neural networks

The structure of deep convolutional neural networks currently used for image classification can be divided into two parts. The first part is composed of several convolutional layers and pooling layers for extracting image features, and the second part is fully connected layer and softmax layer for classification of features. These two parts finally output the probability that the image belongs to each category, as shown in Figure 1.

![Figure 1. Structure of deep convolutional neural network for image classification.](image)

For example, the widely used Inception v4 network was proposed by Google [15], whose predecessor—the Inception v1 network [16]—won the 2014 championship of the ILSVRC image classification competition. The most important feature of the network is to change the structure of the convolution neural network in series, that is, the structure of a convolution layer followed by a convolution layer. Instead, different convolution kernels are connected into an inception unit in parallel, and then inception units form the whole network.

Although the network structure of Inception v4 is very complex, it can be divided into two parts as a whole. Inception v4 network normalizes the picture to 299×299×3, and finally obtains a 1536-dimensional feature vector \( x \in \mathbb{R}^{1536} \) through convolutional layers and pooling layers. Then, through a fully connected layer, the relative magnitude of the probability that the picture belongs to each category is obtained, and the probability distribution is computed by the softmax layer. Finally the probability that the picture belongs to different categories is output. For example, if the total number of categories is 5, the coefficient matrix of the fully connected layer is \( A \in \mathbb{R}^{5x1536} \). Let the deviation be \( b \), then the final output is Softmax (\( Ax+b \)).

A large number of practices [2-6] have proved that deep convolutional neural networks trained by massive data extract features with better robustness to morphological diversity, illumination variation diversity and background diversity than SIFT and HOG features of traditional computer vision.

Usually, the network before the fully connected layer is called the bottleneck layer [2]. For the image classification problem of few samples, the deep convolutional neural network can be trained through the sample-rich public data set, and then the bottleneck layer parameters can be retained, only retraining the coefficients of the fully connected layer. In this way, the parameters that need to be retrained are greatly reduced.

Compressed sensing theory believes that high-dimensional data in reality has the characteristics of sparse low rank, and the same is true for the feature space obtained by convolutional neural networks. Each category of image has only a small dimension in the feature space. For example, the Inception v4 network obtains a feature space of 1536 dimensions. Mathematically, it takes 1536 samples to determine each row of the fully connected layer coefficient matrix \( A \). However, in practice, when using the transfer learning method for image classification, only 100 samples are needed. The ideal result can be obtained, which is because of the sparsity of the data.
When the sample number is too small, it is more susceptible to noise when determining the fully connected layer coefficient matrix $A$. Considering the dimension reduction of the data, extracting the main components, thereby increasing the sample density and reducing the influence of noise, based on this idea, a transfer learning method of dimension reduction is proposed.

3. Dimension-reduction fine-tuning method
This paper proposes a dimension-reduction fine-tuning method, which has the following steps:

- Train deep convolutional neural networks with sample-rich public data sets;
- The parameters of the bottleneck layer are retained, and use the bottleneck layer to calculate the feature vectors of the samples;
- Reduce the dimension of the feature vectors to extract the main components;
- Train the parameters of the fully connected layer with the dimension-reduced feature vectors to obtain the classifier.

The principal component analysis (PCA) algorithm is used to reduce the dimension of the feature vectors to extract the main components.

Principal component analysis is a widely used method for data dimension reduction, which can effectively reduce the data dimension, and extract the main components. Let the feature vectors of the $n$ samples be $\{x_1, x_2, \ldots, x_n\}$ and each feature vector consist of $p$ features, $z_1, z_2, \ldots, z_p$. This gives a feature vector matrix $X$ with the size of $p \times n$:

$$
X = \begin{bmatrix}
    x_1 & x_2 & \cdots & x_n \\
    z_1 & z_2 & \cdots & z_n \\
    z_2 & z_2 & \cdots & z_n \\
    \vdots & \vdots & \ddots & \vdots \\
    z_p & z_p & \cdots & z_p
\end{bmatrix}
$$

The steps of dimension reduction by principal component analysis are as follows:

1. Centralize all the feature vectors of samples
   $$
x_i \leftarrow x_i - \frac{1}{n} \sum_{i=1}^{n} x_i
$$

2. Calculate the covariance matrix of the centralized feature vectors $XX^T$, whose size is $p \times p$;
3. Perform eigenvalue decomposition on $XX^T$ to obtain eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ and their corresponding eigenvectors $u_1, u_2, \ldots, u_p$;
4. The eigenvectors $u_1, u_2, \ldots, u_k$ corresponding to the largest $k$ eigenvalues give the projection matrix $U = (u_1, u_2, \ldots, u_k)^T$, where $U$ is $k \times p$, obviously $k \leq p$;
5. Multiply $X$ by $U$ to get $UX$, which is $k \times n$, thus completing the dimension reduction of the data.

$k$ is determined by the percentage of principal component $\tau$:

$$
k = \min \left( \frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i} \geq \tau \right)
$$

The principal component analysis with a percentage of principal component $\tau$ is defined as PCA($\tau$).

The projection matrix $U$ and the original feature vector mean $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ need to be saved after dimension reduction. For the test sample feature vector $x_{test}$, $U(x_{test} - \bar{x})$ project it into low-dimensional space.
By projecting the samples into a low-dimensional space, the sample density can be increased and the influence of noise can be reduced. An example is given to verify the effect of the proposed dimension-reduction fine-tuning method.

4. Simulation and analysis

The proposed method is validated by the classical flower classification problem. The problem has an open data set (http://download.tensorflow.org/example_image/flower_photos.tgz), which has five kinds of flower sample images, with different background, different illumination and different view angles. Each picture is RGB color image, with different sizes. The sample information is shown in Table 1.

| Number of samples | Daisy | Dandelion | Roses | Sunflowers | Tulips |
|-------------------|-------|-----------|-------|------------|--------|
|                   | 633   | 898       | 641   | 699        | 799    |

The network used for transfer learning is the Google Inception V4 network. The parameters used in the bottleneck layers are trained by Google using the ImageNet dataset. The trained Inception V4 network can successfully classify 1000 types of objects with an error rate of less than 4%, which exceeds humans, but is not very effective for flower classification datasets. Usually, the bottleneck layers parameters are fixed by the transfer learning method, and the coefficients of the fully connected layer are retrained.

First, we examine the effect of transfer learning under different sample numbers. Randomly take 100 pictures from each class as the test set, and then randomly take 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, ..., 100, 200, 300, 400 pictures from each of the remaining samples as the train sets, using the transfer learning method to train the full connection layer coefficients, and calculating the classification accuracy on the test set by the network. Repeat 10 times, and obtain the average accuracy, as shown in Figure 2.

It can be seen that as the number of samples per class increases, the accuracy curve first rises rapidly and then tends to be flat. It is surprising that even if there is only one sample per class, the classification accuracy on the test set is 52.8%. When the number of samples per class reaches 10, the accuracy is rapidly increased to about 75%, which shows the dimension of each class in the feature space is very low. When the number of samples per class reaches 20, the accuracy is about 80%. After that, when the number of samples per class increases again, the accuracy increases very slowly.

For the number of samples per class is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, ..., 100, the dimension-reduction fine-tuning method proposed in the previous section is used. Using PCA (0.95) to reduce the dimension, the initial dimension of each sample feature vector is 1536, and the dimension after dimension reduction is shown in Figure 3.

The fully connected layer coefficients were trained using the dimension-reduction fine-tuning method, and the classification accuracy on the test set is shown in Figure 4 and compared with the previous result.

It can be seen that when the number of samples per class is very small, the classification accuracy of the dimension-reduction fine-tuning method is better than the original one. As the number of samples per class increases to 20, the accuracy begins to be smaller than the original result. This is
also consistent with expectations, that few samples are more susceptible to noise, and as the number of samples per class increases, the effects of noise are gradually reduced. At the same time, we can see that although the dimension of the feature vector is greatly reduced by dimension reduction, even if the number of samples increases, the accuracy is always near the original result, which shows each class is sparsely low-rank in the feature space and most of the useful information is retained after dimension reduction.

Figure 2. Accuracy of classification under different sample numbers per class of flowers by transfer learning.

Figure 3. Reduced dimension by PCA(0.95).

Figure 4. Classification accuracy by dimension-reduction fine-tuning method.
Figure 5. Classification accuracy by dimension-reduction fine-tuning method (1~20 samples).

Focus on the case of 1~20 samples per class, and get Figure 5. The statistics results show that dimension-reduction fine-tuning method can improve the classification accuracy by 1%~2%.

In fact, assuming that the feature vectors form a linear subspace in the feature space and the dimension is $k$ for the $i$-th class of samples, then only $k$ linearly independent feature vectors $w_1, w_2, \ldots, w_k$ of the $i$-th class need to be found, and the fully connected layer coefficient matrix $A$ is trained, s.t.

$$Aw_j = e_i, 1 \leq j \leq k$$  \hspace{1cm} (4)

where $e_i$ is the unit vector in which only the $i$-th component is 1 and the remaining components are 0.

Then for any sample belonging to the $i$-th class, its feature vector $w$ can be written as the linear combination of $w_1, w_2, \ldots, w_k$, i.e.

$$w = \sum_{j=1}^{k} \mu_j w_j$$  \hspace{1cm} (5)

Then

$$Aw = A\sum_{j=1}^{k} \mu_j w_j = \sum_{j=1}^{k} \mu_j Aw_j = \eta e_i, \eta = \sum_{j=1}^{k} \mu_j$$  \hspace{1cm} (6)

Thus the sample is correctly classified.

5. Conclusions

In this paper, a dimension-reduction fine-tuning method is proposed for image classification problem of few samples based on the sparse low-rank feature of high-dimensional data. This method uses the principal component analysis algorithm to reduce the dimension of image features extracted by the pre-trained convolutional neural network, and then trains the classifier with the dimension-reduced feature vectors. This method improves the classification accuracy when the number of samples is very small, and is analyzed and verified by simulation.

References

[1] Krizhevsky A, Sutskever I and Hinton G E 2012 ImageNet classification with deep convolutional neural networks Neural information processing systems pp 1097-105

[2] Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E and Darrell T 2014 DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition International
conference on machine learning 647-55

[3] Pittaras N, Markatopoulou F, Mezaris V and Patras I 2017 Comparison of Fine-Tuning and Extension Strategies for Deep Convolutional Neural Networks International Conference on Multimedia Modeling, pp 102-14

[4] Razavian A S, Azizpour H, Sullivan J and Carlsson S 2014 CNN Features Off-the-Shelf: An Astounding Baseline for Recognition computer vision and pattern recognition 512-9

[5] Tajbakhsh N, Shin J Y, Gurudu S R, Hurst R T, Kendall C B, Gotway M B and Liang J 2016 Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? IEEE Transactions on Medical Imaging 35 1299-312

[6] Yanai K and Kawano Y 2015 Food image recognition using deep convolutional network with pre-training and fine-tuning. In: international conference on multimedia and expo pp 1-6

[7] Srivastava N, Hinton G E, Krizhevsky A, Sutskever I and Salakhutdinov R 2014 Dropout: a simple way to prevent neural networks from overfitting Journal of Machine Learning Research 15 1929-58

[8] Pittaras N, Markatopoulou F, Mezaris V and Patras I 2017 Comparison of Fine-Tuning and Extension Strategies for Deep Convolutional Neural Networks conference on multimedia modeling pp 102-14

[9] Radenovic F, Tolias G and Chum O 2016 CNN Image Retrieval Learns from BoW: Unsupervised Fine-Tuning with Hard Examples european conference on computer vision 3-20

[10] Shin H, Roth H R, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D J and Summers R M 2016 Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning IEEE Transactions on Medical Imaging 35 1285-98

[11] Pan S J and Yang Q 2010 A Survey on Transfer Learning IEEE Transactions on Knowledge and Data Engineering 22 1345-59

[12] Candes E J and Wakin M B 2008 An Introduction To Compressive Sampling IEEE Signal Processing Magazine 25 21-30

[13] Yang S, Zhang L, He L and Wen Y 2019 Sparse Low-Rank Component-Based Representation for Face Recognition With Low-Quality Images IEEE Transactions on Information Forensics and Security 14 251-61

[14] Turk M and Pentland A 1991 Eigenfaces for recognition Journal of Cognitive Neuroscience 3 71-86

[15] Szegedy C, Ioffe S, Vanhoucke V and Alemi A A 2016 Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning national conference on artificial intelligence 4278-84

[16] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S E, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A 2015 Going deeper with convolutions computer vision and pattern recognition 1-9