Multi-scale Feature Mapping Method Based on Clustering Convolution

Yun Jiang, Junping Ren, Jize Xiao, Hai Zhang and Xiufang Chen

College of Computer Science and Engineering, Northwest Normal university, Lanzhou 730070, China.
Email: jiangyun@nwnu.edu.cn

Abstract. In recent years, convolutional neural networks have attracted much attention due to their powerful image feature learning capabilities. However, in many cases, the network exhibits shortcomings such as difficulty in training and adjustment, and time-consuming training. This paper combines clustering and constrained Boltzmann machines to propose a multi-scale feature mapping method based on clustering convolution. Experiments were carried out on Cifar-10 and lung X-ray images. This method not only shortened the time but also improved the accuracy. Therefore, the proposed method is superior to the traditional feature extraction method.

1. Introduction

Lung cancer is one of the most malignant tumours with the highest morbidity and mortality worldwide. According to statistics, there were 228,190 new cases of lung cancer in the United States in 2013, ranking second in all malignant tumours after prostate cancer, and the number of deaths was 159,480, ranking first in mortality [1]. The incidence of lung cancer in China is even more serious. According to the data released by the National Office for the Prevention and Treatment of Lung Cancer, the incidence of lung cancer in the past 18 years is increasing year by year in China, with an average annual growth rate of 1.63% [2]. Most of the existing computer-aided diagnosis techniques use a deep learning method to construct a network, extract features of lung X-ray images, and finally classify them.

After the deep learning was put forward, the medical field has been greatly developed. Early researchers mainly used the pre-training method based on greedy choice to extract the abstract features of data layer-wise [3]. The Restricted Boltzmann Machine (RBM) proposed by Smolensky [4] can obtain random samples of the distribution obeyed by RBM through Gibbs sampling [5]. Roux and Bengio [6] theoretically proved that RBM can fit arbitrary discrete distributions when there are enough hidden elements. The Contrastive Divergence (CD) proposed by Hinton [7] speeds up the learning of RBM. In 2006, Hinton et al. [8] proposed the Deep Belief Nets (DBN) model based on RBM, an unsupervised layer-by-layer training algorithm, which greatly improved the modeling ability of the model. Adam Coates et al. [9] directly generated a convolution kernel which can be used for convolution feature extraction by extracting patch blocks from the image and using k-means clustering. This way, the convolution kernel can be obtained more quickly than CNN, and the abstract features of the image can be extracted. Recently, Sharma C et al. [10] proposed a new method of feeding the Radon transform of X-ray images to a back-propagation neural network trained by the Levenberg algorithm, which is called new method of artificial learning image diagnosis. In the same year, Dundar A et al. [11] used the convolutional k-means clustering method to further enhance the model's ability to extract abstract features and improve the classification accuracy of images. Subsequently,
Qiangchang Wang et al. [12] proposed a new multiscale rotation invariant (MRCNN) model using multi-scale rotation-invariant Gabor LBP images as input to CNN instead of the original image in traditional methods. Lakshmanaprabu S.[13] et al. constructed an optimal deep learning model for classification of lung cancer on CT images. The model uses the optimal depth neural network (ODNN) and linear discriminant analysis (LDA) to perform CT scans of lung images. The deep features extracted from a CT lung images and then dimensionality of feature is reduced using LDR to classify lung nodules. However, this way of directly clustering the dictionary on the patch block requires preprocessing (whitening, etc.) of the image, and due to the certainty of the patch block, the final generalization of the convolution kernel is not strong. Furthermore, in CNN, the training process for convolution kernels that are the decisive factor for feature extraction is very time consuming. The reason is that the generation of the convolution kernel is continuously trained by the random initial value through the back propagation algorithm. This process of depicting from nothing is a major factor in the time-consuming CNN training.

In this paper, a method based on cluster convolution for scale feature mapping (MFMCC) is proposed. This method is a process of forming a patch block from the original image, inputting the patch block to Gaussian Restricted Boltzmann Machines (GRBM) [14], and clustering the output to form a dictionary. In this paper, we performed experiments on Cifar-10 and lung X-ray images. On Cifar-10, we used a three-layer network with an accuracy of 83.53%. On the lung X-ray image, the accuracy reached 83.92%.

This paper is divided into five parts. The first part is the introduction. The second part is the related work. This part introduces the basic theory of clustering convolution and Gaussian restricted Boltzmann machine and the abstract feature clustering and reconstruction of patch block. The third part is the method proposed in this paper. The fourth part is the experiment on cifar-10 and lung X-ray image, which is analyzed in two aspects of accuracy and time. The fifth part summarizes the full text.

2. Related Work

2.1. Cluster Convolution Model

Coates et al. [15] found that clustering methods can replace neural networks as a fast feature learning method. Classical \( k \)-means algorithm [16] minimizes the distance between data points and their cluster centers. \( D \) and \( x^{(i)} \) can be mapped into a vector form to minimize reconstruction errors, as shown in (1):

\[
\min_{\Theta} \sum_{i}^{n} \| D^{(i)} - x^{(i)} \|^2
\]

where \( D \in R^{mxk} \) is the clustering center, \( n \) is the data dimension, \( k \) is the number of center points, and \( x^{(i)} \in R^m, i = 1, ..., m \) is the data sample.

Coates et al. [17] used the improved version of \( k \)-means algorithm, Spherical \( k \)-means, by adding the mapping of to (2), Spherical \( k \)-means, by adding the mapping of to (2). Alternately trained \( D \) and \( s \) to achieve the goal of quickly training larger cluster centers. In order to achieve the goal of obtaining larger cluster centers more quickly, the algorithm is trained on \( D \) and \( s \) alternately.

\[
s^{(i)} = \begin{cases} \frac{D^{(i)T} x^{(i)}}{||D^{(i)}||}, & \text{if } j = \text{argmax} \frac{D^{(i)T} x^{(i)}}{||D^{(i)}||_2}, \forall_i, \\ 0, & \text{else} \end{cases}
\]

For image data, a large number of small patch blocks are randomly cut out from the unmarked image. These patch blocks are used as training samples, where \( n \) is the patch block dimension, \( m \) is the number of samples. First, the sample is preprocessed, including data normalization, whitening, etc. The clustering center of the patch block is obtained by formula (4) iteratively from the preprocessed data.

\[
D := XS^T + D
\]

\[
D^{(j)} := D^{(j)} / ||D^{(j)}||_2, \forall j
\]

\[
(\text{3})
\]

\[
(\text{4})
\]
Using the cluster center as a dictionary, the original image can be convolved through a dictionary to obtain the convolution characteristics of the image. They are shown in (5), where is a convolution operation.

\[ x_j^{\text{feature(i)}} = D_j \ast x^{\text{image(i)}}, j = 1, \ldots, k \quad (5) \]

For example, feature mapping of M Image X with \(28 \times 28\) pixels. Firstly, \(m\) patches of \(4 \times 4\) pixel size are randomly extracted from the image, and \(k\) clustering centers \(D \in \mathbb{R}^{m \times k}\) are obtained by clustering \(m\) patches. The characteristic \(x_j^{\text{feature(i)}}, j = 1, \ldots, k\) representation of each image can be obtained by (5). This method can be used to replace the convolutional neural network to learn the abstract features of the image.

2.2. Structure and Training Algorithm of Gaussian Restricted Boltzmann Machine

RBM [18] is an artificial neural network with a two-layer structure, generally a binary unit is used as its neuron node. However, Hinton [14] points out that, for natural image processing, the binary unit has poor performance in representing features. The original visible unit can be replaced by the Gauss linear unit with independent noise. For the GRBM model, the corresponding energy function is defined as following:

\[ E(v, h) = -\sum_{j \in \text{visible}} \frac{(v_j - a_j)^2}{2\sigma_v^2} - \sum_{j \in \text{hidden}} b_j h_j - \sum_{i,j} v_j w_{ij} h_j \quad (6) \]

where \(\sigma_v\) is the standard deviation of the Gaussian noise for visible unit \(i\).

When the state of the visible unit is determined, the activation state of each hidden unit is conditionally independent. The activation probability of the \(j\)th hidden unit is defined as following:

\[ P(h_j = 1|v, \theta) = \sigma(b_j + \sum_{i} v_i w_{ij}) \quad (7) \]

When the state of the hidden unit is determined, the activation states of the visible units are also independent of each other. The probability of the \(i\)th visible unit is as follow:

\[ P(v_i | h, \theta) = N(a_i + \sigma \sum_j w_{ij} h_j, \sigma^2) \quad (8) \]

\[ \sigma(x) = \frac{1}{1 + \exp(-x)} \quad (9) \]

here, \(\sigma(x)\) represents the sigmoid activation function, \(N(\mu, \sigma^2)\) represents the Gaussian function with the mean \(\mu\) and the variance \(\sigma^2\). \(\theta = [w, a, b]\) is the training parameter, where \(\omega\) is the weight matrix, \(a, b\) are the offsets of the visible layer and the hidden layer, respectively. The Contrastive Divergence (CD) algorithm proposed by Hinton [7] in 2002 is the standard algorithm for training RBM. The algorithm first obtains the activation state of the hidden unit through the visible layer, and then reconstructs the visible layer using the hidden layer, thereby obtaining the reconstruction of the input sample.

In RBM, the "Reconstruction Error" indicates the difference between the original data and the data reconstructed by the RBM. It is usually evaluated using L1 norm or L2 norm. The reconstruction error can reflect the likelihood of RBM on the sample data to some extent [14]. Let \(v_{\text{data}}\) represent raw data and \(v_{\text{recon}}\) represent data reconstructed by RBM, the reconstruction error can be expressed as:

\[ \text{error} = \left\| v_{\text{data}} - v_{\text{recon}} \right\|_2 \quad (10) \]


### 2.3. Abstract Feature Clustering and Reconstruction of Patch Blocks

When using $k$-means to learn feature representations, the raw data needs to be whitened. Although whitening can eliminate the correlation between patch blocks and make the edges of the image more prominent, it can not improve the expressive ability of the abstract features of the image. Whitening can not make the patch block which is finally used to map image features have stronger generalization ability, either. On the other hand, although the size of the patch block image is small, the number of the patch block image cannot be ignored. In order to obtain a wider range of image features, a large number of patch blocks need to be clustered. It is particularly important to further reduce the feature dimension for clustering.

In view of the above considerations, this paper further improves the method of learning feature representation using $k$-means. The abstract feature mapping and the patch block clustering center reconstruction mechanism are employed, so as to obtain more information of the abstract feature relations between the patch blocks of the same category, and to improve the generalization ability of the dictionary.

![Figure 1. MFMCC model proposed in this paper](image)

### 3. Improved Method for Multi-scale Feature Mapping Based on Clustering Convolution

In this paper, a multi-scale feature mapping method based on clustering convolution (MFMCC) is proposed. This method extracts patch blocks of different scales, obtains dictionaries of different scales, and maps the features of images through multi-scale dictionaries. This method improves the classification accuracy of the mapped features. That is to say, when patch blocks are extracted, patch blocks of different scales are trained separately, and finally dictionaries of different scales can be obtained. Using dictionaries of different scales to map image features can compensate for the negative effect of single scale dictionaries as filters on classification results. The structure of multi-scale feature mapping network based on clustering convolution and restricted Boltzmann machine is shown in figure 1.
MFMCC needs to extract patch blocks of different scales in the original image and train different RBMs separately. For patches of different scales, abstract features are extracted and clustered, and the cluster center is reconstructed to obtain a dictionary group. Different scales of the dictionary can map convolution features of different scales. Finally, the multi-scale convolution clustering features of the image are obtained by splicing the convolution features of multiple scales. Compared with single-scale method, multi-scale method needs to extract patch blocks of different scales and train them separately. On the other hand, it needs to merge the feature mapping results of different scales. The algorithm is shown in Table 1.

Table 1. Multi-scale feature mapping algorithm based on clustering convolution and restricted Boltzmann machine

| Input: Sample $X(M$ images with size $N\times N)$ |
| Output: classification accuracy |
| Initialization: normalize $X$ and extract some data from it as a training sample (denoted as $X_{train}$), the rest are used as test samples (recorded as $X_{test}$); a GRBM is randomly initialized. |
| Step one: |
| for $t=1,2,...,n$ |
| Randomly extract $m_t$ patch blocks of size $p_t \times p_t$ from all $X_{train}$, and convert it into $m_t$ vectors $X_{patch}(i), i=1,2,...,m_t$ |
| Using $X_{patch}$ as a training sample, training the first $t$ GRBMs with a learning rate of $\eta_t$ and an iteration of $e_t$. |
| Enter $X_{patch}$ into the visible layer of the first $t$ GRBMs, the feature vector $y_{patch}$ is mapped by GRBM is extracted from the hidden layer by equation (6). |
| Find $k_t$ cluster centers $c(i), i=1,2,...,k_t$ of $y_{patch}$ from equations (2~4). |
| The cluster center $c_t$ is input into the hidden layer of the $t$-th GRBM, and the abstract feature is reconstructed by the equation (7) to obtain the reconstructed cluster center. |
| Convert each reconstructed cluster center to a dictionary block of $p_t \times p_t$ to get $D_t = \{d_1, ..., d_{k_t}, t, d \in R^{p_t \times p_t} \}$ |
| Using $D_t$ as the convolution kernel, calculate $x_t = R(X_{train} * D_t)$, where $R(x)$ is the ReLu (Rectified Linear Units) activation function and $*$ is the convolution operation. |
| end for |
| Step two: |
| The $x_t(i=1,2,...,n)$ is pooled and spliced into a set of vectors, which is the mapped feature vector. |

4. Experiment Analysis

4.1. Datasets
CIFAR-10: CIFAR-10[19] data set consists of 60,000 color images with a size of 32×32, a total of 10 classes, 6000 images per class, 50,000 images as training samples and 10,000 images as test samples.

X-ray image of the lung: This image uses the lung X-ray image acquired by the National Medical Library biomedical image search engine as an experimental data set, containing 2 classes of 800 color images, 480 images belonging to normal images, and 320 disease images. The section image is used for testing and the size is 512×512; part of the image is shown in Figure 2.
4.2. Experiment on CIFAR-10 and Lung X-ray Image

The patch blocks we extracted on CIFAR-10 are $3 \times 3$, $5 \times 5$, and $7 \times 7$, each of which is 100,000, which are used to train GRBM. After full training, abstract features of three different patch blocks are obtained. The abstract features are clustered and the cluster center is input into the hidden layer of the GRBM, and the visible layers are reconstructed to obtain the dictionaries $D_{3\times3}$, $D_{5\times5}$ and $D_{7\times7}$. The dictionary is shown in Figure 3.

We use the trained dictionary to directly perform the convolution feature mapping to generate the classifier, and then classify it on the test set with the trained classifier. Table 2 shows the accuracy of classification of CIFAR-10 images by different methods.

| Different methods | Classification |
|-------------------|----------------|
| Original cluster convolution feature mapping(1 layer) [6] | 78.30% |
| Original clustering convolution feature mapping method (2 layers) | 81.2% |
| Original clustering convolution feature mapping method (3 layers) | 82.0% |
| MFMCC (1 layer) | 80.32% |
| MFMCC (2 layers) | 82.84% |
| MFMCC (3 layers) | 83.53% |
| Convolutional Deep Belief Network (Conv.DBN) [20] | 78.90% |
| Deep Convolutional Neural Network (Deep CNN) [21] | 80.49% |

It can be seen from Table 2 that the 2-layer network of MFMCC can achieve the effect of the traditional method, and the accuracy of the three-layer network is 83.53%.

On the lung X-ray image, we set the first layer, and the second and third layers have convolution kernels of 64, 128, and 32, respectively. Figure 4 shows the comparison of the $4 \times 4$ patch blocks generated by the first two layers of the network with the reconstruction and the visualization of the
The final patch block used as the filter. Our training and testing was ran on GPU NVIDIA GeForce GT 740, with 4GB memory.

In this experiment, experiments were carried out on the lung X-ray images using the proposed method and the original cluster convolution method. Table 3 shows the classification accuracy on the lung X-ray image and the time taken to classify each picture.

Combine the above two experiments. Compared with the traditional method, the MFMCC has improved classification accuracy. The reason is that the MFMCC mainly uses patch blocks instead of the original image to extract features. A large number of patch blocks make the extracted features more adequate. GRBM has powerful feature extraction functions. The output features are clustered, and the clustering produces representative features, which reduces the redundancy between the features, so that the generated filter can produce a more accurate classifier. In addition, in terms of time, the network is faster, because the patch block has a very low dimension and fewer parameters. Although GRBM and clustering may slow down the training process of the overall network, it can save some time compared with the traditional process of repeatedly searching in the network.

![Figure 4](image1.png)

**Figure 4.** Figure 4(a) and Figure 4(b) gives the results of patch visualization in Layer 1 network, and Figure 4(c) and Figure 4(d) gives the results of patch visualization in Layer 2 network. With Figure 4(a) and Figure 4(c) show the reconstruction comparison between the partially segmented patches and the RBM feature mapping layer. In these figures, the odd lines show the patch blocks extracted from the original image, and the even lines show the reconstruction results. By comparing Figure 4(a) and Figure 4(c) it can be found that the RBM reconstruction layer can generalize and adjust the original randomly extracted patches, which is embodied in the enhancement of the edge features of patches when the edge features of patches are not obvious. Figure 4(b) and Figure 4(d) show the final visualization results of the filters. Using these filters as the convolution kernels of the convolution feature mapping layer, the image feature mapping can be obtained. One of the main advantages of in-depth learning in image feature mapping is to extract more abstract features layer-wise. The edge features of the image are shown in Figure 4(b), mainly in the color edge information of the skeleton and muscle in the lung X-ray image. In the second layer of the network shown in Figure 4(d), the results of the filter visualization are more complex and abstract than those of the first layer, showing that the deeper network tends to represent more abstract features.
Table 3. Comparison of classification accuracy of different methods for lung X-ray images

| Different methods | Classification accuracy | Time(s) |
|-------------------|-------------------------|---------|
| Original cluster convolution feature mapping (1 layer) [6] | 82.24% | 0.535 |
| Convolutional Deep Belief Network (Conv. DBN) [20] | 77.08% | 0.581 |
| Deep Convolutional Neural Network (Deep CNN) [21] | 78.76% | 0.796 |
| MFMCC | 83.92% | 0.429 |

Table 3 shows that compared with the traditional method, the accuracy of the proposed method reaches 83.92%, and the time is shortened to 0.429 seconds. Our network is superior to traditional networks in terms of accuracy and time in lung X-ray images.

5. Summary
This paper proposes a multi-scale feature mapping method based on clustering convolution (MFMCC). On the one hand, the patch block is used to train the RBM to obtain the abstract features of the patch block, the abstract features are clustered, and the clustering center is reconstructed by RBM to obtain a dictionary with more generalization ability. The other party extracts patches of different scales from the original samples, and finally obtains a multi-scale dictionary. We performed experiments on CIFAR-10 and lung X-ray images, respectively. Experiments show that the proposed method is superior to the traditional method in both accuracy and time. Finally, we analyse the experiment.

6. References
[1] Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013[J]. CA Cancer J Clin, 2013, 63(1):11-30.
[2] Wangqing Chen, et al. Estimation and epidemic trend of lung cancer death in China [J]. Chinese Journal of Lung Cancer, 2010.
[3] Hinton G E, Sejnowski T J. Learning and relearning in Boltzmann machines [A]. In: Rumelhart D E, McClelland J L (eds.). Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations. Cambridge, MA: MIT Press, 1986.
[4] Smolensky P. Information processing in dynamical systems: Foundations of harmony theory [A]. In: Rumelhart D E, McClelland J L (eds.). Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations. Cambridge, MA: MIT Press, 1986.
[5] Freund Y, Haussler D. Unsupervised learning of distributions on binary vectors using two layer network [R]. Santa Cruz: University of California, UCSC-CRL-94-25, 1994.
[6] Roux N L, Bengio Y. Representational power of restricted Boltzmann machines and deep belief networks [J]. Neural Computation, 2006, 20(6): 1631-1649.
[7] Hinton G E. Training products of experts by minimizing contrastive divergence [J]. Neural Computation, 2002, 14(8): 1771-1800.
[8] Hinton G E, Osindero S, Teh Y. A fast learning algorithm for deep belief net [J]. Neural Computation, 2006, 18(7): 1527-1554.
[9] Coates A, Ng A, Lee H. An analysis of single-layer networks in unsupervised feature learning[C]//Proceedings of the fourteenth international conference on artificial intelligence and statistics, 2011: 215-223.
[10] Sharma C, Basheer S, Francis A, et al. Radon transform processed neural network for lung X-ray image based diagnosis[C]// Convergence of Technology. IEEE, 2015.
[11] Dundar A, Jin J, et al. Convolutional Clustering for Unsupervised Learning [J]. Computer Science, 2015: 1143-1151.
[12] Wang Q, Zheng Y, et al. Multi-Scale Rotation-Invariant Convolutional Neural Networks for Lung Texture Classification [J]. IEEE Journal of Biomedical and Health Informatics, 2017: 1-1.
[13] Lakshmanaprabu S, K, et al. Optimal deep learning model for classification of lung cancer on CT images [J]. Future Generation Computer Systems, 2019, 92: 374–382.
[14] Hinton G. A practical guide to training restricted Boltzmann machines [J]. Momentum, 2010.
[15] Coates A, Ng A Y. Learning feature representations with $k$-means[M]/Neural networks: Tricks of the trade. Springer Berlin Heidelberg, 2012: 561-580.

[16] Han, Jiawei, Jian Pei, et al. Data mining: concepts and techniques. Elsevier, 2011.

[17] Coates A, Ng A, Lee H. An analysis of single-layer networks in unsupervised feature.

[18] Salakhutdinov R, Mnih A, Hinton G. Restricted Boltzmann machines for collaborative filtering[C]/Proceedings of the 24th international conference on Machine learning. ACM, 2007.

[19] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images[J]. 2009.

[20] Krizhevsky A, Hinton G. Convolutional deep belief networks on cifar-10[J]. Unpublished manuscript, 2010, 40: 7.

[21] Ciresan D C, Meier U, Masci J, et al. Flexible, high performance convolutional neural networks for image classification[C]/IJCAI Proceedings-International Joint Conference on Artificial Intelligence. 2011, 22(1): 1237.