Classification of Gastric Slices based on Deep Learning and Sparse Representation

Bo Liu[1], Ming Zhang[1], Tongyu Guo[2], Yuanzhi Cheng[1*]

1. Harbin Institute of Technology, School of Computer and Technology, Harbin 150001
E-mail: 17s103145@ hit.edu.cn

2. Northeastern University, College of Information Science and Engineering, Shenyang 110004
E-mail: 13555921383@163.com

Abstract: In this paper, a classification method based on deep learning and sparse representation is proposed for gastric slice images. First of all, the convolution features of images are extracted through the convolutional neural network. Then an overcomplete learning dictionary of convolution features can be gained by K-SVD algorithm. Convolution features can obtain their sparse representation through decomposition on the overcomplete dictionary. Finally, a linear SVM classifier is used to classify the sparse representation of convolution features instead of the commonly used nonlinear classifiers, so as to achieve the purpose of classifying gastric slices. Experiments show that using linear SVM reduces the computation cost of classifier and achieves better classification effect than convolutional neural network with sigmoid output layer or SVM with RBF kernel. Therefore, the method proposed in this paper has better classification effect and lower computation cost.

Key Words: Convolutional Neural Network, Sparse Decomposition, Support Vector Machine, Gastric Slice Image

1 INTRODUCTION

Gastric cancer is one of the most common digestive tract malignant tumors in the world. Japan, South Korea and China are high incidence of gastric cancer in Asia. There are about 400 thousand new cases in China each year, accounting for 42% of the total number of cases in the world[1]. With the development of artificial intelligence, especially deep learning, people pay more and more attention to computer-aided diagnosis, where some progress has been made in the study of gastric cancer slice images. By extracting and classifying the features of the gastric slices, computer-aided diagnosis system can judge the normal and abnormal of the gastric and help doctors to make a diagnosis. In essence, the above process is to divide gastric slice images into two categories, cancer and non-cancer.

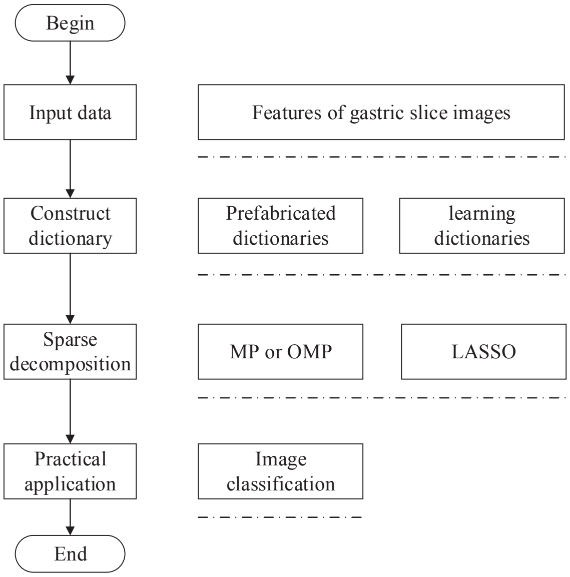


Figure 1 A flowchart of image classification based on sparse representation

The classification results have a great relationship with features extraction of images and performance of classifier. In previous work, there are two kinds of manual bottom feature extraction method: one is interest points detection, the other is dense extraction [3]. Interest points detection algorithm selects obvious feature pixels, edges, corner points or blocks by some criterion, which generally has the geometric invariance and small computation overhead, such as the Harris corner detection, Features from Accelerated Segment Test (FAST), Laplacian of Gaussian (LoG), etc. Dense extraction method extracts a large number of local features from the fixed step length and scale. Although a large number of local features have higher redundancy, the feature information is more abundant. This method will achieve better result compared with the feature extraction method based on interest points. The common local features include Scale-Invariant Feature Transform (SIFT), Histogram of Oriented Gradient (HOG), Local Binary (LBP), etc. However, a popular view in recent years is that using the low level feature descriptor as the first step of visual information processing often loses useful information too early. Directly learning feature descriptions related to task from image pixels is more effective than manual features. Some experiments also show that the

This work is supported by National Nature Science Foundation under Grant ×××××××

978-1-5386-1243-9/18/$31.00 ©2018 IEEE

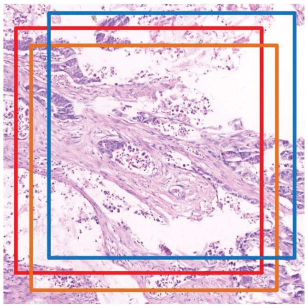


Figure 2 Cutting a picture of a specified size from an image randomly

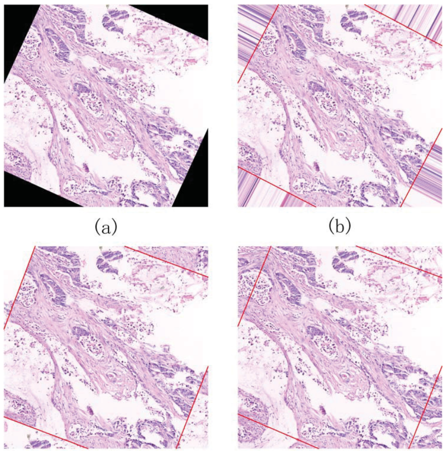
using of deep learning such as the convolutional neural network (CNN) to extract features is better than the traditional manual features in many cases [4]. As an end to end feature extraction method, CNN has been used more and more widely.

After extracting the features of the images, we need to classify them by classifier, which commonly use sigmoid neuron, support vector machine (SVM) or others. The sigmoid neuron is often used as the output layer for binary classification problem, while softmax layer is often used in multi-classification problem. However, the performance of these two classifiers are always lower than SVM classifier. SVM based on the maximum boundary are one of the most widely used classifiers, especially the one with kernel method. Although the kernel method improves the performance, it increases the amount of computation as well. The classification of high-dimensional nonlinear data often results in huge computation cost.

To reduce the amount of computation of the classifier at the same time to improve classification effect, we use sparse representation of convolution features rather than kernel method. Sparse representation of original features can get the representation of features in higher dimensions. It can not only obtain the essential structure of the features, but also increase the interpretability of the model. Because sparse representation is nonlinear representation, only using linear SVM classifier can get better classification results, which reduce the requirement of classifiers in computation [5]. In this paper, we propose a method for classification of gastric slices based on deep learning and sparse representation. Firstly, we use CNN to extract the convolution features of images. Then we get an overcomplete dictionary of these features through K-SVD algorithm. We extract the sparse representation of features from overcomplete dictionary by sparse decomposition. Finally, we classify the sparse representation of features by linear SVM classifier, so as to classify gastric slice images.

2 IMAGE FEATURE EXTRACTION BY CONVOLUTIONAL NEURAL NETWORK

As discussed above, CNN has been widely applied as an end to end feature extraction method in image classification, segmentation and recognition tasks. Specifically, the basic structure of CNN is made up of input layer, convolution layer, pool layer (also called sampling layer), fully connected layer and output layer.



(a) (b)

(c) (d)

Figure 3 Using different methods to fill empty parts of the images

2.1 Image Preprocessing

When the amount of data is small but the size of the CNN is large, overfitting happens usually. In order to suppress overfitting, two methods of data enhancement are mainly used here. One is random cutting and the other is geometric transformation. For the method of random cutting, a 1792 *1792 sub-image is cut randomly from a 2048 * 2048 original image, as shown in the red box in Figure 2. This method is a little similar to [6].

Repeating random cutting N times, the number of images for training is N+1 times of the original. Figure 2 is the case of N=3. Although random cutting can increase the number of training data, the cut data still have a high correlation. Therefore, we need to rotate, translate, flip the sub-images or do other geometric transformations. These methods can reduce correlation as well as increase the number of images for training.

When the original image is rotated and translated, there will often be empty parts which are necessary to be filled. The common ways of filling are constant filling, nearest value filling, wrapping filling and mirror image filling. Figure 3 is the result of different filling modes when the original image is rotated.

In Figure 3, image (a) is filled with constant value; image (b) is filled by nearest neighbor's value; image (c) is filled by wrapping; image (d) is filled with mirror image. Observing the four images, it's obvious that constant filling method directly loses the information which is rotated out of the original image, and nearest way adds the texture information that doesn't exist on the original image. Although the way of wrapping supplements image by original image, the connection part is meaningless. However, the method of mirror image filling can solve above problems. It supplements the information of image and make connection parts meaningful so mirror image is used to fill empty parts. Finally, images are resized to 512 * 512 * 3 as the input of neural network.

2.2 The Architecture of Convolutional Neural Network

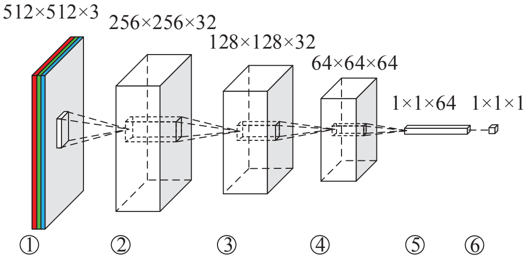


Figure 4 Convolutional neural network architecture

There are many popular neural network architectures, such as VGG or ResNET, but these network architectures are trained on massive data which can’t be directly applied to the classification of medical images. There are two main reasons for this problem. First of all, these models are trained with large scale data and often have large entropy capacity. Medical images often don’t have such a large scale data set. If we use medical images to train such a large model directly, the model will have serious overfitting. In addition, although we can use medical images to train large CNN by transfer learning, the effect is often not good enough. Because the correlation between source domain data and target domain data is too low, it will make the effect of the transfer learning worse. It may even cause phenomenon of negative transfer[7]. To classify gastric slice images, a CNN architecture, as shown in Figure 4, is designed.

As shown in Figure 4, the first layer is an input layer in which the size of input image is 512 * 512 * 3. Each cube in the second level to the fourth level represents the result after convolution and pooling. The size of the convolution core is 3 * 3, step length is 1. The size of the maximum pooling is 2 * 2, so the output data of each dimension is half of the upper level. The fifth layer is a convolution core of 64 * 1 * 1 through which we can obtain convolution features. The sixth layer is a sigmoid neuron.

By training the above neural network, we can get an image classifier based on CNN and sigmoid neuron. We can directly extract output of the fifth level as image features for other classifiers as well.

3 SPARSE REPRESENTATION FOR IMAGE CLASSIFIER

In recent years, neuroscience studies has shown that only a small portion of the corresponding neurons in the human brain will be active in a single signal stimulus. Sparse representation not only provides a simple representation of the redundant information, but also makes the upper sensing nerve obtain the most essential feature of the stimulus signal [8]. It can obtain the essential representation of the convolution feature of the image as well as reduce the computational cost of the classifier.

3.1 Sparse Representation of Convolution Features

A flowchart of image classification based on sparse representation is shown in Figure 1. First, an overcomplete dictionary is learned for the convolution features of gastric slice images. Then, the convolution features are decomposed on overcomplete dictionary to obtain their sparse representation. The sparse representation of convolution features are shown in formula (1).

$$y = D \cdot x \tag{1}$$

where D denotes overcomplete dictionary of convolution features, y denotes convolution features. The x is the sparse representation of convolution features under the overcomplete dictionary D.

It’s obvious that there are two core steps in sparse decomposition of convolution features, one is how to learn an overcomplete dictionary, the other is how to decompose the convolution features on it. For the convenience of narration, the sparse decomposition method will be introduced first.

3.2 Sparse Decomposition Method

When overcomplete dictionary has been learned, the sparse representation of the signal can be found through solving the problem in formula (2).

$$\begin{aligned} &\min_{x} \quad \|x\|_0 \\ &s.t. \quad y = Dx \end{aligned} \tag{2}$$

In formula (2), $\|x\|_0$ is the number of nonzero terms in vector x. However, under this condition, it is a NP hard problem to find a sparse representation from a random overcomplete dictionary. In order to solve this problem, it can be converted to an equivalent problem in formula (3).

$$\begin{aligned} &\min_{x} \quad \|x\|_1 \\ &s.t. \quad y = Dx \end{aligned} \tag{3}$$

This change converts the nature of the problem, turning the original NP hard problem into a new problem which can be solved by linear programming. Although the speed of solving this problem is improved obviously, using the method of linear programming in actual still has a large amount of computation. More commonly used methods are matching pursuit algorithm (MP) and orthogonal matching pursuit algorithm (OMP) which are based on greedy algorithm. This kind of algorithm searches for atoms closest to the input signal in overcomplete dictionary every time, until the error between reconstructed signal and original signal satisfies a certain threshold. In fact, the problem of solving optimal solution is converted to the problem of solving suboptimal solution.

The sparse decomposition problem can also be transformed to the LASSO problem, as shown in formula (4).

$$J(\mathrm{w}) = \frac{1}{2}\|y - Dx\|^2 + \lambda \sum_i |x_i| \tag{4}$$

Although the analytical solution can’t be obtained, the numerical solution can be quickly obtained which is good characteristics in application.

3.3 Construction of Overcomplete Dictionary

As shown in Figure 1, overcomplete dictionaries can be roughly divided into two categories: prefabricated dictionaries and learning dictionaries. DCT overcomplete dictionary is a prefabricated dictionary commonly used in sparse decomposition. It is obtained by making the DCT complete dictionary more fine frequency sampling and frequency adjustment.

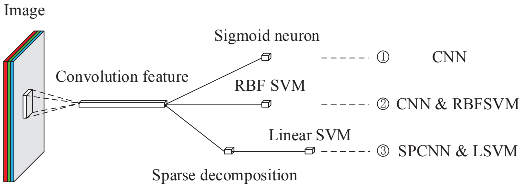


Figure 5 Diagrammatic sketch of experimental design

In learning dictionaries, they can be divided into structural dictionaries and unstructured dictionaries. In this paper, we use K-SVD algorithm to get a learning dictionary. In K-SVD, the sparsity is used as the constraint target and the optimal fitting of the signal is obtained as formula (5).

$$\begin{aligned}&\min_{D,\{x_i\}_{i=1}^{M}} \sum_{i=1}^{M} \left\| y_i - Dx_i \right\|_2 \\ &s.t.\quad \left\| x_i \right\|_0 \le k_0,\quad 1 \le i \le M\end{aligned} \tag{5}$$

In the K-SVD algorithm, the column d_{j0} can be updated by multiplying its coefficients while all the other columns remain unchanged. We isolate the d_{j0} related coefficients and rewrite in formula (6).

$$\begin{aligned}\left\| Y - DX \right\|_F^2 &= \left\| Y - \sum_{j=1}^{m} d_j x_j^T \right\|_F^2 \\ &= \left\| \left(Y - \sum_{j \ne j_0} d_j x_j^T \right) - d_{j0} x_{j0}^T \right\|_F^2\end{aligned} \tag{6}$$

In formula (6), x_j^T stands for line j of X. The goal of the update step is d_{j0} and x_j^T. The items of the parentheses are used as a known error matrix, as shown in formula (7), which has been calculated in advance.

$$E_{j0} = Y - \sum_{j \ne j_0} d_j x_j^T \tag{7}$$

The optimal d_{j0} and x_{j0}^T for the minimization of formula (7) are the approximation of the rank of 1 of the E_{j0}, which can be obtained by the SVD algorithm. But this usually produces a dense x_{j0}^T, which means that it increases the number of nonzero terms in the X representation. To minimize the known error matrix, a subset of the E_{j0} column should be taken out in order that all the expressed potential is invariable. The column of this subset corresponds to the sample set using the signal of the j_0 atom, so these columns are non-zero in the line x_{j0}^T. Therefore, we only allow the non-zero coefficients in the x_{j0}^T to change which can keep the potential. The above process is iterated. When the reconfiguration error satisfies the threshold requirement, the iteration is stopped and the overcomplete dictionary after learning is obtained.

Both of the above two methods can be used to get the overcomplete dictionary for sparse decomposing. The way to construct DCT overcomplete dictionary is simple and universal, but can't adjust itself according to the images. Although the training process of learning dictionary by K-SVD is slow, the trained dictionary is more targeted which makes the result of sparse decomposition better.

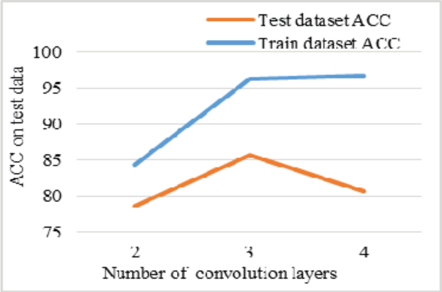


Figure 6 Influence of the number of convolution layers on the classification results.

Table 1 Image classification results of different methods

Method	Acc	Recall	Precision	F1
CNN	86.4%	89.3%	93.5%	91.3%
CNN & RBFSVM	89.2%	89.7%	93.8%	91.7%
SPCNN & LSVM	95.0%	90.2%	94.0%	92.1%

4 EXPERIMENTAL TEST

4.1 Data Set and Experimental Platform

The gastric slice images were collected from the "BOT AI Challenge of Pathological Section Identification" and renamed BOT data for short[9]. The slices were stained with conventional HE, the magnification was 20 * 20. The size of the each image was 2048 * 2048 pixels. Among them, there are 560 cancer images and 140 non-cancer images. We selected 448 cancer images and 112 non-cancer images as training set. Then we selected the rest 112 cancer images and 28 non-cancer images as test set. The training set was enhanced to 2688 positive and negative samples by random cutting and geometric transformations.

The CNN used in this experiment is set up and trained by Keras2.0.4[10]. The K-SVD algorithm to learn overcomplete dictionary and sparse decomposition is carried out by spams-python-v2.6.1 [11]. The servers has Intel Xeon (R) CPU and its memory is 64 GB. GPU is Quadro K6000 with 12GB.

4.2 Experimental Design

In this part, 3 groups of experiments were carried out, as shown in Figure 5. The first group was named CNN, the second group was named CNN & RBFSVM and the third group was named SPCNN & LSVM. This kind of naming is the same as in Table 1.

In the first group, we used the neural network structure which was shown in Figure 4 for training and classification. We could get the convolution features of the images when the classification result of sigmoid output layer has been gained. In this experiment, we constantly adjusted the structure of the CNN, especially the number of convolution layers. By adjusting these hyperparameters, the most suitable architecture of CNN was found for this task. In the second group, convolution features were put into the SVM classifier with RBF kernel to classify gastric slice images. In the third group, we obtained the overcomplete dictionary of convolution features by K-SVD algorithm, and then extracted the sparse representation of convolution features by sparse decomposition on overcomplete dictionary. Put the sparse representation into SVM classifier with linear kernel, and gastric slice images can be classified excellently.

4.3 Image Classification Results

In the first group, we compared the effects of the network layers' number on the neural network classification results. The experimental results are shown in Figure 6. It shows that when the number of convolution layers is less than 3, the accuracy of test set increases with the increasing of the accuracy of training set. When it is larger than 3, overfitting phenomenon is serious. When using BOT data set and enhancing data by our method, 3 convolution layers should be used to obtain the best classification effect.

After determining the number of the convolution layers, the number of neurons and hyperparameters of L_2 regularization term are adjusted to suppress the overfitting. The classification results of test data in 3 groups of experiments are shown in Table 1. It shows that the accuracy of classification results obtained through CNN and sigmoid output layer is 86.4%. Extracting features by CNN and classifying it by SVM classifier with RBF kernel, the accuracy of test data is 89.2%. The features extracted from CNN is firstly learned by K-SVD to gain overcomplete dictionary, then sparse decomposition is performed. Only using linear kernel SVM classifier, the accuracy of test data is 95.0%. On Recall, Precision and F1, the third group is the best as well.

In above three groups of experiment, the best result is obtained based on the deep learning and sparse representation. It has not only the best classification effect, but also a lower amount of computation for classifier. It is a kind of classification algorithm with high performance and real time.

5 CONCLUSION

In this paper, an image classification method based on deep learning and sparse representation is proposed for the classification of gastric slice images. The convolution features of the images are extracted by CNN. An overcomplete learning dictionary of these features can be then learned through K-SVD algorithm. The convolution features are decomposed on overcomplete dictionary to get their sparse representation. Finally, the best classification result can be obtained by only using linear SVM, not nonlinear. This method reduces the computational complexity of the classifier while obtaining the best classification effect, which may make the classification method work in real time.

REFERENCES

[1] Zou Wenbin, Li Zhaoshen. Research Progress on the Incidence and Mortality of Gastric Cancer in China [J]. Chinese Journal of Practical Internal Medicine, 2014, 34 (04): 408-415.

[2] Yang Xiaolan, Qiang Yan, Zhao Juanjuan, Du Xiaoping, Zhao Wenting. Based on Medical Signs and Convolution Neural Network, CT Image Hash Retrieval of Pulmonary Nodules, [J/OL]. Intelligent System Journal, 2017, (06): 1-9 (2017-11-09).

[3] Huang, Cage, Ren Weiqiang, Tan Tieniu. Computer Image Classification and Object Detection Algorithms [J]. Journal of computer science, 2014, 37 (06): 1225-1240.

[4] Xu J, Luo X, Wang G, et al. A Deep Convolutional Neural Network for Segmenting and Classifying Epithelial and Stromal Regions in Histopathological Images[J]. Neurocomputing, 2016, 191: 214-223.

[5] Yang J, Yu K, Gong Y, et al. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification[C]//Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 1794-1801.

[6] Krizhevsky, Alex, Sutskever, Ilya, Hinton, Geoffrey E. ImageNet Classification with Deep Convolutional Neural Networks [J]. Communications of the Acm, 2012, 60(2):2012.

[7] Pan S J, Yang Q. A Survey on Transfer Learning [J]. IEEE Transactions on knowledge and data engineering, 2010, 22(10): 1345-1359.

[8] M. P. Yong, S. Yamane. Sparse Population Coding of Faces in the Inferotemporal Cortex [J]. Science, 1992, 256(1): 1327-1330.

[9] Pathological Section Identification of AI Challenge [EB/OL]. http://www.datadreams.org/race-data-3.html. 2017

[10] Chollet, François, et al. Keras [EB/OL]. https://github.com/fchollet/keras. 2017

[11] Julien, Mairal. SPArse Modeling Software [EB/OL]. http://spams-devel.gforge.inria.fr/. 2017