2019 3rd Scientific Conference on Mechatronics Engineering and Computer Science (SCMC 2019)

Image Classification Based on Sparse Coding Improvement and Sparse Depth Learning Model

Yunpeng Chen

School of Computing, the University of Leeds, University of Leeds, Leeds LS2 9JT, UK, England

Keywords: Sparse Coding; Deep Learning Model; Image Classification

Abstract: For the traditional image sparse representation algorithm, it only pays attention to the feature extraction rate and ignores the multi-scale information of the image, and for the problem that the image sparse representation process is subject to noise interference and low system robustness, sparse coding improvement and deep learning are proposed. In view of the instability of the underlying pixel features of the image and the environmental impact, which is impossible to fully reflect on the semantic information and other issues, it is proposed to combine the multi-scale coefficient representation with the deep learning network through the deep network to create a learning framework and the experimental results shows that this algorithm can improve on the image recognition performance and improve the system robustness.

1. Introduction

In the process of the popularization of high-definition equipment, the image information collected by people gradually shows the characteristics of its complexity, high dimensionality and diversity, but the original data is also interfered by rotation, fouling, illumination, etc. The traditional image recognition algorithm has the actual needs that cannot be met. Deep learning refers to the use of human brain tissue mechanism to learn, to achieve an effective, abstract high-level representation of the combination of underlying features because the underlying pixel features are unstable and vulnerable to the environmental impacts [1]. Dictionary learning is combined with sparse representation in order to achieve a multi-scale dictionary training through the extracted high-level features, thus achieving the accurate image recognition.

2. Classification Algorithm Based on Sparse Representation

The process of representing the image classification algorithm is based on the sparse which is extracting the image data y in the training set, and using it as a test sample, calculating the sparse dictionary and the sparse code x mapping in the base matrix A through the test sample y, and then analyzing the coefficient in x giving the final image classification of the test sample. FIG. 1 is a feature extraction method based on sparse code.

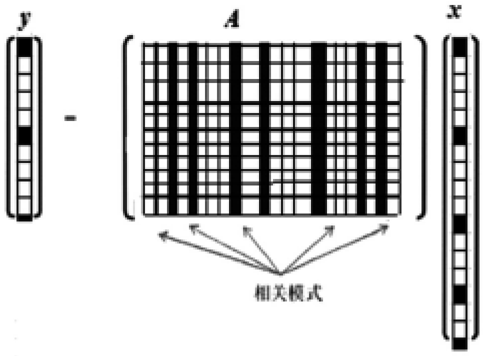

Figure 1 Feature Extraction Method Based on Coefficient Representation

The flow of the sparse representation algorithm:

(1) Input the training sample matrix A=[A_1, A_2, A_3, ..., A_i], $A \subseteq R^{m*n}$ (**m** and **n** refer to the

  DOI: 10.25236/scmc.2019.095

number of vector elements, **i** refers to the number of sample categories), $y \in R^m$ and input the test sample;

(2) Normalizing the elements of column A;

(3) Because A refers to the underdetermined matrix, and the problem of solving the L1 norm minimization changes toward the problem of convex optimization, and the sparse solution of underdetermined matrix is obtained. If the data is noisy, then the sparse solution $x_1 = \arg\min \|x\|_1$ makes Ax=y satisfy, from $\|Ax - y\|_2 \le \varepsilon$

(4) Calculate the residual $r_i(y) = \|y - A\delta_i(x_1)\|_2 (i = 1,2,...,n)$ between the test sample data and the sample, in which δ_i refers to the feature function.

(5) The output is defined as arg min $r_i(y)$[2].

3. Image Classification Based on Sparse Deep Learning Model

3.1 Unsupervised Rbm Coding

In this paper, the multi-layer constrained Boltzmann machine (RBM) is used to realize the coding learning with the SIFT feature library to generate a discriminative and representative visual dictionary.

Firstly, the SIFT feature is extracted from the training image set; Secondly, the SIFT feature library is used to code by RBM, and the CD fast algorithm **(Contrastive Divergence)** is used to train the RBM in order to obtain the hidden layer feature. Then, the adjacent hidden layer feature is made into the input of the next layer so that the output dictionary can be derived, because the RBM has a visible layer and a hidden layer, but the same neurons in RBM are not connected to each other making the learning process more simple [3].

In the process of training, the RBM hidden layer and the explicit layer are associated with the conditional probability distribution, the conditional and hidden layer conditional probabilities are:

$$p(z_j \mid x) = sigmoid(b_j + \sum_{i=1}^{I} w_{ij} x_i)$$

$$p(x_j \mid x) = sigmoid(c_i + \sum_{i=1}^{I} w_{ij} z_j)$$

Where $\mathbf{x_i}$ refers to the feature layer, $\mathbf{z_j}$ refers to the coding layer, that is, the display layer and the hidden layer; $\mathbf{w_{ij}}$ refers to the connection weight coefficient between the feature layer $\mathbf{x_i}$ and the coding layer $\mathbf{z_j}$. For a given set of input layers and coding layers **(x, z)** for RBM, the energy function is calculated as:

$$E(x,z) = -1gp(x,z) = -\sum_{i=1}^{I}\sum_{j=1}^{J} x_i w_{ij} z_j - \sum_{i=1}^{I} c_i x_i - \sum_{j=1}^{J} b_j z_j$$

A joint probability distribution function of **(x, z)** can be derived based on the energy function:

$$p(x,z) = \frac{e^{-E(x,z)}}{\sum_{x,z} e^{-E(x,z)}}$$

In this way, the probability distribution of the characteristic input nodes of the joint distribution can be obtained, which is:

$$p(x)=\frac{\sum_{x,z}e^{-E(x,z)}}{\sum_{x,z}e^{-E(x,z)}}$$

The main purpose of the RBM network training process is to maximize p(x), and use the CD algorithm to achieve a fast learning of RBM, thereby promoting the parameter convergence, thus obtaining the weight $\mathbf{w_{ij}}$ update amount:

$$\Delta w_{ij}=\varepsilon(<x_i z_j>_{data}-<x_i z_j>_{\mathrm{mod}el})$$

Which ε refers to the learning speed, the CD algorithm can be used to update parameters until parameter convergence, and thus obtain the initial visual dictionary [4].

3.2 Adjustment of Sparseness and Selectivity:

Sparsity refers to the simple expression of image content using a small number of basis vectors, and most of the components in the sparse vector are zero, and only a small part is non-zero, which fully displays the intrinsic structure and essential properties of the image data; Selectivity is the measure of the response of a single visual word to the input feature vector, and is responsible for the extent to which the feature is expressed.

Whether the dictionary can accurately express the image connotation after feature encoding which plays an important role in the realization of the middle layer semantic expression of the image data, adding the regular term h(z) to the RBM target optimization function to adjust the objective function to:

$$\arg\min_{w,c,b}-\sum_{k=1}^{K}\lg(\sum_{z}p(x_k,z_j))+\lambda h(z)$$

Among them h(z) is the regular term weighting function, and the deep learning coding can improve the selectivity of the visual dictionary, so that the image representation vector can have a good sparsity.

$$h(z)=\sum_{j=1}^{J}\left\|\hat{p}-\frac{1}{K}\sum_{k=1}^{K}pjk\right\|$$

The first element in formula is the expected value of the average activation probability of K features for each word; The expected value of a word z_j for response probability of a single feature x_k can be labeled as p_{jk}.

3.3 Creating a Visual Vocabulary

The process of creating a visual vocabulary is:

(1) Randomly selecting images from training images and extracting SIFT features from those images $F=\{f_1, f_2, \ldots, f_N\}$;

(2) Creating visual Vocabulary by using the above algorithm.

3.4 Sparse Representation

After the visual vocabulary is created, the image sparse representation is implemented. The process is as follows:

(1)The image set **I** and the visual vocabulary **V**;

(2) Extracting the SIFT features and creating a feature vector set F={f_1, f_2, ..., f_M};

(3) Creating a sparse feature vector α_1 of each feature vector f_1 with a visual vocabulary V;

(4) Obtain all sparse feature vectors α={α_i}(i=1,2,...,M)

The image feature vectors after sparse representation refer to a high-dimensional number. In order to ensure the calculation efficiency and reducing the storage capacity, the pooling method is used to reduce the image representation dimension. The final feature vector P is calculated by the following formula:

$$P = \frac{1}{M}\sum_{i=1}^{M}\alpha$$

3.5 Image Classification Process

In the training process, the image SIFT features are first extracted and the feature vector set F={f_1,f_2,...,f_N} is created, N refers to the number of SIFT feature vectors, and then the sparse feature vector S and the visual vocabulary V are implemented. The steps are as follows:

(1) Input the sparse feature vector S;

(2) Training the first layer of RBM;

(3) Training the second layer RBM through the first training result;

(4) Repeat for the second and third steps by CD algorithm until all RBM training is completed;

(5) Fine-tuning the overall network parameters through the BP algorithm [5].

4. Simulation Experiment

In order to effectively verify the performance of the algorithm, the classification experiment was carried out based on UC Merced images set. The training images were initialized by randomly selecting 80% images, and the other images were verified by the algorithm classification results. Each image set was subjected to 10 experiments to obtain the final classification results.

Table 1 compares the classification accuracy of different methods in the image set. As can be seen from Table 1, the image classification based on the sparse encoding depth learning model can improve the image classification accuracy and enhance the distinguishability and discriminability of the extracted image features, which makes more easy to classify and recognize images[6].

Table 1 Comparison of Classification Accuracy of Different Methods in Image Sets

Method	BoVM	SPMK	SC+SVM	No Fisher	Article
Precision rate	75.2%	78.6%	84.5%	83.6%	85.2%

5. Conclusion

In the process of continuous development of multimedia technology, the world has also entered the era of big data, and the Multimedia data is highly valued by experts and scholars because it realizes the exchange of people's life information through image approach, and there are many useful information in the image on how to realize the image sorting and classification which is the main problem that people are faced with. Based on the problem of low image classification accuracy, this paper uses the improved sparse coding and sparse depth learning model to create a new image classification method, which can be used as an image classification tool to improve on the image classification accuracy.

References

[1] Zhu Jun, Zhao Jiewei, Dong Zhenyu. Hierarchical Feature Learning Image Classification Based on Fusion of Severe Information [J]. Journal of Computer Research and Development, 2014, 51(9): 1919-1928.

[2] ZHAO Yongwei, LI Ting, YAN Boyu. Image Classification Method Based on Deep Learning Coding Model [J]. Engineering Science and Technology, 2017, 49(1): 213-220.

[3] Zhang Rui, Wang Dong. Using the homotopy regularization method to establish a new sparse self-coding model for deep learning [J]. Journal of Shanxi University (Natural Science), 2015, 38(4): 652-656.

[4] Shang Hongtao. Hyperspectral remote sensing image classification method based on SSAE deep learning feature representation [J]. Computer Measurement & Control, 2018, 26(10): 297-300.

[5] Sun Yuyang. Review of Deep Learning and Its Research in Image Classification and Recognition [J]. Information Technology and Informatization, 2018(1):138-140.

[6] Ma Hongqiang, Ma Shiping, Xu Yuelei, et al. Image denoising based on improved stack sparse denoising self-encoder [J]. Computer Engineering and Applications, 2018, 54(4): 199-204.