AN INTROSPECTIVE LEARNING STRATEGY FOR REMOTE SENSING SCENE CLASSIFICATION

Jingran Su[1], *Qi Wang*[1], *Shangdong Chen*[2], *Xuelong Li*[1*]

[1] School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China.

[2] School of Information Science and Technology, Northwest University, Xi'an 710127, Shaanxi, P. R. China.

ABSTRACT

In this paper, a novel introspective learning strategy for remote sensing scene classification is proposed. Through this strategy, the neural network used for classification can introspectively generate negative samples. In most training deep neural networks, negative samples are rarely noticed. We are the first to actively introduce negative samples into the remote sensing scene classification tasks. The goal of this paper is to analyze the effect of introspective negative samples on remote sensing scene classification tasks. Experiments demonstrate that the introduction of negative samples in training can effectively improve the classification accuracy and robustness. In addition, we found that our method can effectively against invalid remote sensing images.

Index Terms— Scene classification, Remote sensing, Deep learning, Negative samples, Introspective strategie

1. INTRODUCTION

Recently, more and more people are paying attention to the scene classification problem of remote sensing images because of its wide range of application[1, 2]. A key step, encountered in almost all computer vision problems, is the design and extraction of features. In the early years, substantial efforts have been put to design handcrafted features, and there are a lot of techniques for calculating visual features, such as color histograms, scale-invariant feature transform (SIFT) and bag-of-visual-words (BoVW). Although these techniques for extracting features have yielded exciting results in many domains, these techniques cannot be directly applied in the remote sensing domain[3], due to the complexity and particularity of remote sensing images. However, the abovementioned feature extraction method cannot meet our requirements, for the task of scene classification of remote sensing images.

*Corresponding Author.

In the past few years, deep learning has developed rapidly and is now the preferred solution for dealing with visual problems. It replaces previous handcrafted features in many areas because it can learn many advanced features[4]. Given deep learning's powerful ability to process data, it is now widely used in various fields, including remote sensing. Deep learning based methods have achieved the best results in more and more remote sensing applications. In the scene classification task concerned in this paper, many methods based on deep learning are proposed and and achieved good results., such as Fusion by Addition[5], D-CNN[6], AttNet[7], etc. But the datasets used in these previous work did not contain all the image categories that the real world might encounter. When encountering an image of a category other than the datasets, the networks proposed by the previous works will output a wrong result. The method proposed in this paper will effectively solve this problem.

Many remote sensing image datasets are relatively small. It's a conventional strategy, in training neural network, to use data enhancement, which refers to the use of some means to expand the number of positive samples. However, the requirement of using negative samples is rarely noticed. Recent studies have shown that negative samples have a significant impact on the improvement of accuracy in the classification task. Inspired by [8], in which self-generated negative samples are utilized, we proposes an introspective learning strategy to deal with remote sensing scene classification tasks. In this strategy, the classification network also has the ability to generate. Specifically, our method iteratively expands the number of the negative sample during training, compared to the normal classification network. Also, we expand a new category in the dataset by using the negative samples of all categories. Therefore, the K-class classification task becomes a $(K+1)$-class classification task.

To conclude our introduction, we propose an introspective learning strategy for remote sensing scene classification. The contributions of this work can be summarized as follows:

(1) We introduced introspective strategies to enable the classification network to have the ability to generate negative samples at the same time.

(2) Experiments prove that the introduction of negative samples in training can effectively improve the classification accuracy.

(3) Using negative samples as a new category can effectively cope with images of categories not included in the dataset.

2. OUR METHOD

The whole framework of our method is illustrated by Fig. 1, which consists of two parts. The red arrow from left to right indicates the classification parts, and the blue arrow from right to left indicates the introspection part. The two parts are constantly alternating.

We define $\boldsymbol{x}$ as the input vector and $y \in \{1, ..., K\}$ as its label. Below we will discuss the binary classification problem first, in which $y \in \{-1, +1\}$. $p\left(y|\boldsymbol{x}\right)$ is the probability that $\boldsymbol{x}$ belongs to class y. $p\left(\boldsymbol{x}|y=+1\right)$ corresponds to the distribution of positive samples, and $p\left(\boldsymbol{x}|y=-1\right)$ corresponds to the distribution of negative samples. Under the Bayes rule:

$$p(\boldsymbol{x}|y=+1) = \frac{p(y=+1|\boldsymbol{x})p(y=-1)}{p(y=-1|\boldsymbol{x})p(y=+1)}p(\boldsymbol{x}|y=-1), \tag{1}$$

when assuming equal priors $p(y=-1) = p(y=+1)$, and $p(y=-1|\boldsymbol{x}) + p(y=+1|\boldsymbol{x}) = 1$:

$$p(\boldsymbol{x}|y=+1) = \frac{p(y=+1|\boldsymbol{x})}{1-p(y=+1|\boldsymbol{x})}p(\boldsymbol{x}|y=-1). \tag{2}$$

The relationship between positive and negative samples can be seen from Eq. (2).

We define $p_i(y|\boldsymbol{x})$ represents the classification result of the network calculation in the i-th iteration. Obviously, in an ideal situation, $\lim_{i\to\infty} p_i(y|\boldsymbol{x}) = p(y|\boldsymbol{x})$. However, in many cases, $\lim_{i\to\infty} p_i(y|\boldsymbol{x})$ cannot approximate the true probability $p(y|\boldsymbol{x})$ for various reasons. One reason that is often encountered is that there are not enough samples in the training set, which means that $\boldsymbol{x} \sim p(\boldsymbol{x}|y=+1)$ have few samples. Many previous attempts have been to enhance the accuracy of the classifier by extending the positive samples of the dataset through data enhancement. But in this article, we try to sample negative samples $\boldsymbol{x} \sim p(\boldsymbol{x}|y=-1)$ by introspection to improve the classifier.

Now suppose $S_i = \{(\boldsymbol{x}_j, y_j), j = 1..n\}$ is the dataset used for training and $p_i^-(\boldsymbol{x})$ is $p(\boldsymbol{x}|y=-1)$ obtained by introspection and $S_i^- = \{\boldsymbol{x}_j | \boldsymbol{x}_j \in p_i^-(\boldsymbol{x})\}$ is the sample set obtained by sampling $p_i^-(\boldsymbol{x})$. At each time i, we do the following calculation:

$$p_i^-(\boldsymbol{x}) = \frac{1}{Z_i}\frac{p_i(y=+1|\boldsymbol{x})}{q_i(y=-1|\boldsymbol{x})}p_{i-1}^-(\boldsymbol{x}), \tag{3}$$

$$S_i^- = \{(\boldsymbol{x}_j, y_j=-1), j = 1..l, \boldsymbol{x}_j \in p_i^-(\boldsymbol{x})\}, \tag{4}$$

$$S_i = S_{i-1} \cup S_i^-. \tag{5}$$

Then training the classifier on S_i. Where

$$Z_i = \int \frac{p_i(y=+1|\boldsymbol{x})}{p_i(y=-1|\boldsymbol{x})}p_{i-1}^-(\boldsymbol{x})d\boldsymbol{x}, \tag{6}$$

S_0 is the original dataset. In the experiment we make $p_0^-(\boldsymbol{x})$ obey a Gaussian distribution.

Eq. (3) and Eq. (4) represent the process of introspection, which is the process of the blue arrow in Fig.1.

2.1. Classification Part

The classification part is the same as training a normal neural network on S_i. The parameters (weights) of the entire neural network are recorded as W_i. W_i is learned by the gradient descent method to minimize the cross-entropy loss. We use the loss function as follows:

$$L(W_i) = -\sum_{(\boldsymbol{x}_j, y_j)\in S_i} \log p_i(y_j|\boldsymbol{x}_j; W_i). \tag{7}$$

In our framework, the classification network can be divided into feature extraction layer and the last full connected layer. In the experiment we tried multiple network structures, such as VGG-16[9], ResNet[10] and, etc., as our feature extraction layer. Specific experimental results in the Sec. 3.

2.2. Introspection Part

The process of introspection is to use the ability of the current classifier to reverse the distribution of negative samples. Update Eq. 3 to:

$$p_i^-(\boldsymbol{x}) = \frac{1}{Z_i}\frac{p_i(y=+1|\boldsymbol{x}; W_i)}{q_i(y=-1|\boldsymbol{x}; W_i)}p_{i-1}^-(\boldsymbol{x}). \tag{8}$$

Similarly, we use the gradient descent method to minimize the cross entropy loss to find negative samples $\boldsymbol{x} \sim p_i^-(\boldsymbol{x})$. Now we introduce a new set $\boldsymbol{X}_i = \{\boldsymbol{x}_j, j = 1..l, \boldsymbol{x}_j \in p_{i-1}^-(\boldsymbol{x})\}$. Here the loss function is defined as:

$$L(\boldsymbol{X}_i) = -\sum_{\boldsymbol{x}_j\in\boldsymbol{X}_i} \log p_i(y=+1|\boldsymbol{x}_j; W_i). \tag{9}$$

At this point, we fixed the previously learned network parameters (weights) W_i , but instead optimized the input set $\boldsymbol{X}_i$. When the loss error is less than a certain threshold α, we can think that $\boldsymbol{X}_i$ is S_i^- in the Eq. (4). Obviously, the datas contained in $\boldsymbol{X}_0$ are some vectors obeying the Gaussian distribution.

2.3. Multiple Classification

In order to clarify the principle, the above discussion is based on binary classification. But in the experiment of remote sensing scene classification, we are doing multi-classification. To achieve this, we extend the dataset containing K categories

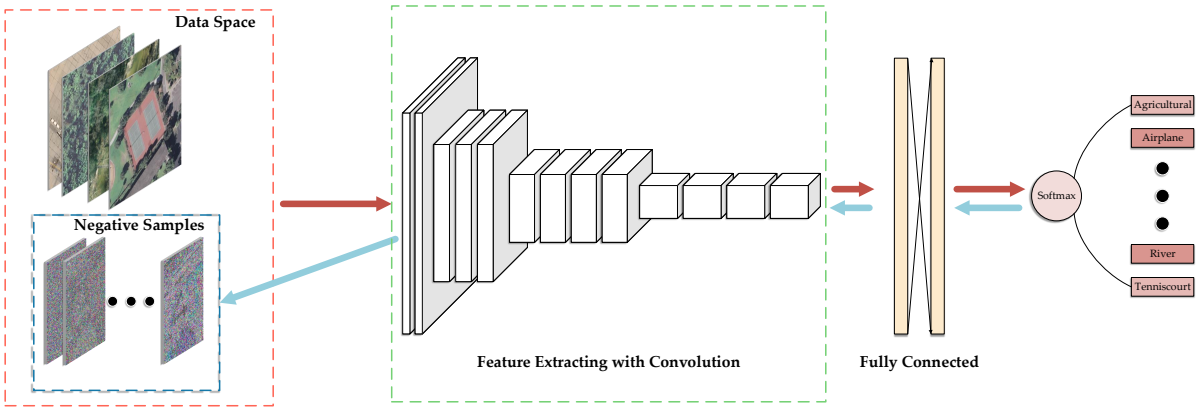

Fig. 1. Overview of the components of our proposed strategy. The red arrow from left to right indicates the training process of the common classifier, and the blue arrow from right to left indicates the process of finding the input data that minimizes the loss (training the input data), where the network parameters are frozen.

to the $K+1$ categories. Negative samples of all categories together form the $(K+1)$-th class. And we add the softmax layer after the fully connected layer of the network structure. The change in data distribution during the running of the algorithm can be described as Fig. 2.

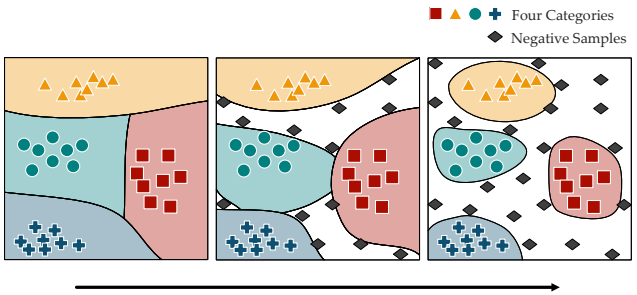

Fig. 2. An example (a four-class classification task) of classification boundaries change when generating negative samples.

3. EXPERIMENTS

3.1. Experimental Datasets

The UC Merced Land-Use (UCM) dataset[11] is one of the first ground truth datasets derived from a publicly available high resolution overhead image[7]. It contains 2100 images in 21 categories (each containing 100 images). Each is an RGB image with a resolution of $256 * 256$ pixels. Many methods of remote sensing classification use this dataset to test performance because it is challenging.

3.2. Experimental Details

In the experiment, we tried different networks as our feature extraction layer. As shown in Fig. 1., after the feature extraction layer is a two-layers fully connected layer and a softmax layer. The algorithm can be divided into two parts alternately: the classification part and the generation of the negative sample part. A classification part contains 20 epochs. In the introspection part, each X_i set contains 210 vectors, which correspond to 10 negative samples for each category. To reduce the time it takes to generate negative samples, we set the threshold α to 1.2. When testing on UCM Dataset we put 80% of the data into training, and the remaining 20% of the data as a test set. Meanwhile, we use overall accuracy and confusion matrix as evaluation methods for experimental results, and all implementations in this paper are based on PyTorch with four NVIDIA Titan X.

Table 1. Results of different methods with the UCM Dataset

Methods	Accuracy (%)
Combing Scenarios I and II[12]	98.49
CNN-NN[13]	97.19
VGG-16[9]	96.19
ResNet[10]	96.43
AttNet[7]	99.05
GoogLeNet[9]	95.02
VGG-16 with Our method	**98.57**
ResNet with Our method	**98.57**
AttNet with Our method	**99.30**

3.3. Experimental Results

The comparison between our proposed method and some state-of-art methods is shown in Table. 1. As can be seen in this table, the proposed introspective strategy can effectively improve the accuracy of remote sensing scene classification task. The accuracy of VGG-16 has increased by 2.38%, by putting it into our proposed framework. Similarly, ResNet also has a 2.14% accuracy boost.

The confusion matrix of VGG-16 with our method using 80% training rates on UCM Dataset is shown in Fig. 3. We use 740 images from the 'Desert' and 'Mountain' classes in the Aerial Image Dataset[9] as the Invalid class here. It is obvious from the confusion matrix that all unrelated images have no effect on the accuracy of our classifier.

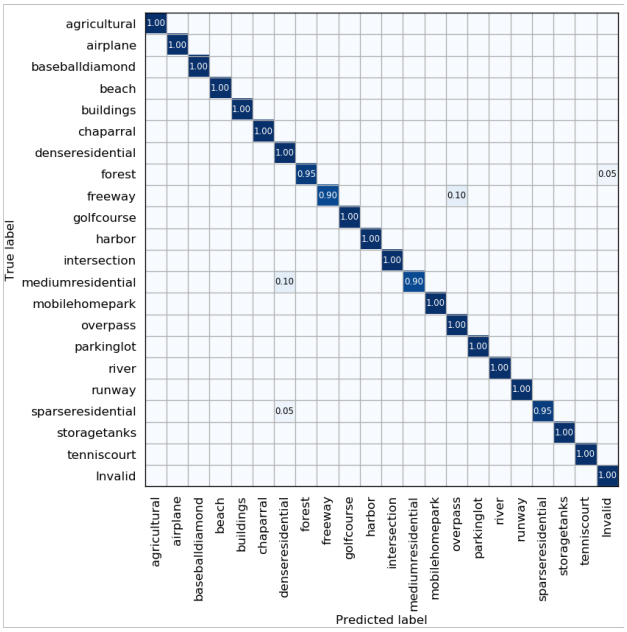

Fig. 3. The confusion matrix with UCM Dataset Under the training ratio of 80%.

4. CONCLUSION

This paper introduces a novel introspective learning strategy for remote sensing scene classification, focuses on the improvement of accuracy by negative samples. We fine-tune multiple networks to apply our proposed strategy. Experimental results on UCM Dataset show that our method can improve classification accuracy and effectively against invalid images.

5. ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China under Grant U1864204 and 61773316, Natural Science Foundation of Shaanxi Province under Grant 2018KJXX-024, and Project of Special Zone for National Defense Science and Technology Innovation.

6. REFERENCES

[1] Q. Wang, X. He, and X. Li, "Locality and structure regularized low rank representation for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–13, 2018.

[2] Q. Wang, S. Liu, J. Chanussot, and X. Li, "Scene classification with recurrent attention of vhr remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–13, 2018.

[3] Keiller Nogueira, Otávio AB Penatti, and Jefersson A dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognition*, vol. 61, pp. 539–556, 2017.

[4] J. Yu, C. Hong, Y. Rui, and D. Tao, "Multitask autoencoder model for recovering human poses," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 6, pp. 5060–5068, June 2018.

[5] Souleyman Chaib, Huan Liu, Yanfeng Gu, and Hongxun Yao, "Deep feature fusion for vhr remote sensing scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4775–4784, 2017.

[6] Gong Cheng, Ceyuan Yang, Xiwen Yao, Lei Guo, and Junwei Han, "When deep learning meets metric learning: remote sensing image scene classification via learning discriminative cnns," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 5, pp. 2811–2821, 2018.

[7] Shaoteng Liu, Qi Wang, and Xuelong Li, "Attention based network for remote sensing scene classification," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 4740–4743.

[8] Kwonjoon Lee, Weijian Xu, Fan Fan, and Zhuowen Tu, "Wasserstein introspective neural networks," in *The IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[9] Gui-Song Xia, Jingwen Hu, Fan Hu, Baoguang Shi, Xiang Bai, Yanfei Zhong, Liangpei Zhang, and Xiaoqiang Lu, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[11] Yi Yang and Shawn Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2010, pp. 270–279.

[12] Fan Hu, Gui-Song Xia, Jingwen Hu, and Liangpei Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sensing*, vol. 7, no. 11, pp. 14680–14707, 2015.

[13] Esam Othman, Yakoub Bazi, Naif Alajlan, Haikel Alhichri, and Farid Melgani, "Using convolutional features and a sparse autoencoder for land-use scene classification," *International Journal of Remote Sensing*, vol. 37, no. 10, pp. 2149–2167, 2016.