CNN based learning: object classification on images from Aerial Photography

Jian-min Liu[1, 2*], Min-hua Yang[1]

[1]School of Geosciences and Info-Physics, Central South University, Changsha, 41000, China

[2]School of Information Science and Engineering, Hunan Institute of Humanities Science and Technology, Loudi, 417000, China

[*]Corresponding author's e-mail: liujianmin@csu.edu.cn

Received 6 July 2016, www.cmnt.lv

Abstract

These years, object recognition on remote sensing images with high resolution had boomed. We trained a multilayer convolutional neural network caffe based to classify the 79 thousand high-resolution and unlabeled optical remote sensing images via the Internet into the 4500 different classes. On the unlabeled test dataset, we obtained error rates of 19.7% which run really well than traditional machine learning techniques. With pre-trained model-aided, GTX750Ti GPU, Intel® Core™ i5-4590 processor, we sharply accelerated progress time. The results compared with the published ones, and good agreement is acquired.

Keywords:

CNN
remote sensing image
object recognition

1 Introduction

In planning, earth observation, coastal monitoring, terminal guidance of missile and many other civil fields, the demand for object recognition sharply increases. Over the past dozen years, scientists have worked on a variety of object classifiers with high efficiency and low cost, but high resolution also brought problems and challenges to the classification issues, the effect is just passable. [1, 2].

In recent years, deep learning technology gradually arisen, including but not limited to convolution neural network, boltzman machine, sparse autoencoder, etc, and applied to some fields including but not limited to speech recognition, image recognition, image retrieval, and other fields. Speech recognition is a classic application of neural network, in recent years, and deep learning has achieved remarkable results in speech recognition in recent years. Microsoft and Google have been gradually put deep learning's speech recognition algorithm into commercial applications. Microsoft's original speech recognition is based on the use of hybrid algorithm Gauss reduce the error rate of words to 16%. Bengio, Hinton and other scholars solved MNIST based classification problems, and breaking the old mark of 0.14% error rate set by traditional machine learning algorithm for many years [3, 4]. Krizhevsky had made significant progress on the ImageNet data set with a 15.3% error rate [5].

On the first hand, the planning image of city construction passed through stack autoencoder to generate specific activations, but on the other hand, status quo of city construction collected by law enforcement terminal also went through stack autoencoder to generate the corresponding features. The system compared the characteristics image of city construction with parity position and a different time. If the difference outpaced threshold, it triggered enforcement warning. Image change detection based on deep learning did not pay attention to non structural changes based on related image pixels, such as gray scale change, brightness shading, but focused on abstract the features of structure overall image, and obtained core attributes [6].

This paper presents object recognition from high resolution remote sensing images without tag based on convolutional Neural Network which consists of convolutional layer, Max-Pooling layer, and fully-connected layer. Therefore, the convolutional layer is an image convolution in essence of activations of the preceding layer, where the convolution filter called learned kernel which is feature of 16x16 or 8x8 in fact. The layer after fully-connected layer works with a final 319-way softmax. The results are compared with those traditional supervised machine learning ones. Conclusive, because GPU's memory available is scarce, network dimension is cramped. Our network spends about 129 hours to train on one GTX 750Ti 2GB GPUs.

2 Theoretical analysis

We can illustrate this graphically as follows:

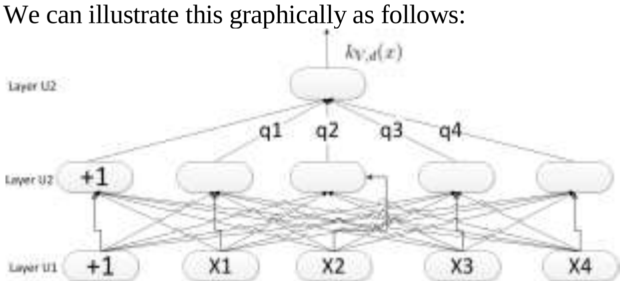


FIGURE 1 Feedforward computing from bottom to top

By executing a feedforward computing from bottom to top, we can obtain the activations for layers. We call this algorithm forward propagation. Figure 1 describe how the forward propagation works.

A convolutional neural network which consists of convolutional layer, max-pooling layer, and fully-connected layer. As a result, the convolutional layer is an image convolution in essence of features of the previous layer, where the convolution filter called learned kernel or weights which is feature of 16x16 in fact specify the convolution filter. The layer after fully-connected layer works with a final 319-way softmax. The results compared with those traditional supervised machine learning ones.

The image come from natural has its own characteristics, that is to say, the statistical properties that are part of the image and the other part is the same. More appropriate interpretation is that when randomly selected from a large size images a small piece, such as sample 16x16, and learned some from this small sample characteristics, we can learn by the 16x16 sample to feature detectors, which is applied to any part of the image. This also means that the features we study in this part which can also be used in another part, so for all of the locations on this image, we can use the same features. In particular, we can learn by the 16x16 sample characteristics with the big size image convolution. Thereby contributing to the big size image of any location to get a different activation feature.

A specific example is given below: assuming that you have make a 384x384 image convolution using 16x16 sample characteristics, and then let them go through sparse autoencoder and trained to get the feature activation values. In this case, obviously we can get 400 sets, which each contains 369x369 characteristics of convolution.

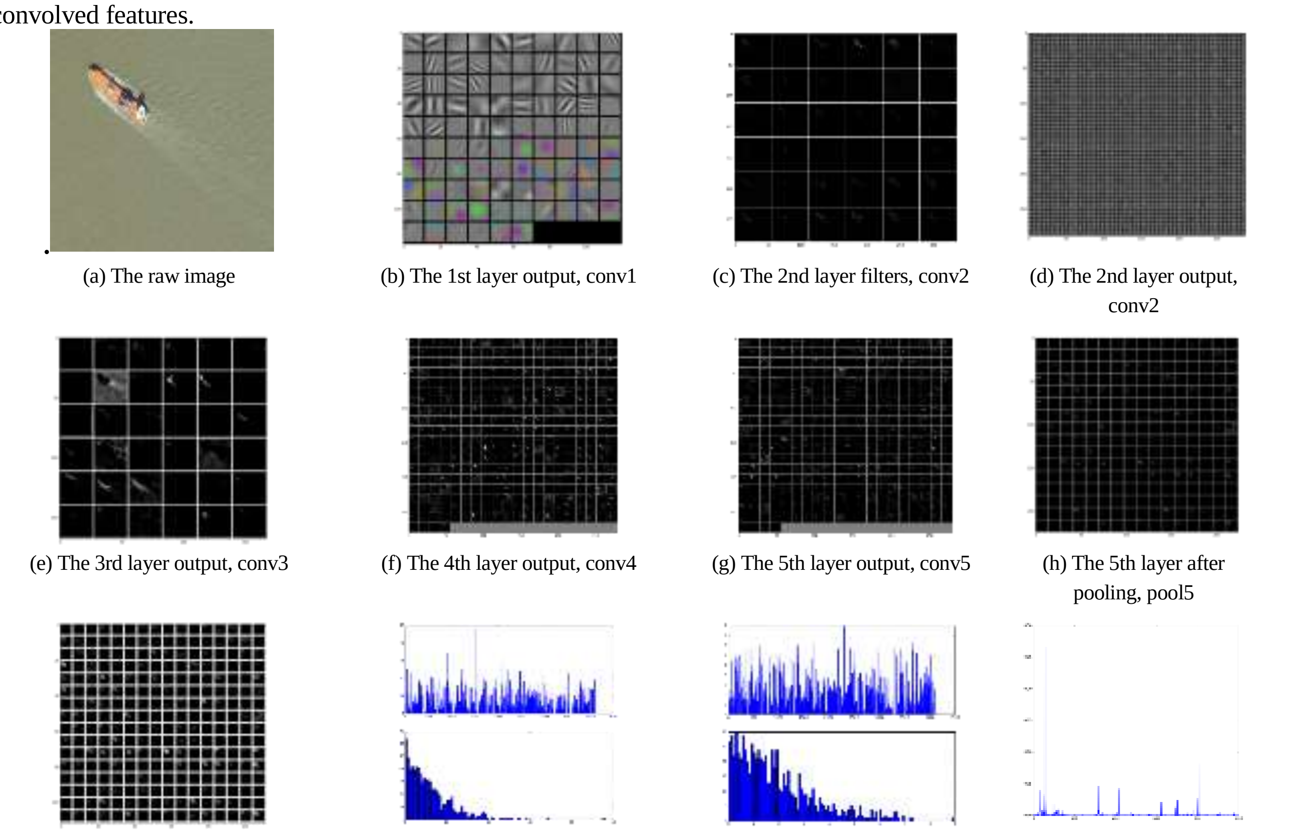
It is convolved that the input of convolution layer come from the previous layer's activations with learned kernel which is feature of 16x16 in fact. We run it through pre-trained sparse autoencoder to compute the activations, i.e., convolved features.

$$x_i^{\,u} = g(d_i^{\,u} + \sum_{i\in N_i} x_j^{\,u-1} * h_{ji}^{\,u}) \tag{1}$$

Concern to the images with 400x400 pixels, and then we obtain each convolution results order of magnitude to square of $(400 - 16 + 1)$ or 148225. Given these results and consider 319 features each result, softmax layer will face the characteristics of the astronomical level will fail soon. Before inputing all the extracted features logistic or softmax layer, we need to significantly reduce the amount of calculation. In order to solve this problem, max pooling layer will take overall average operation of image. On the last layer, we choose softmax model which generalizes logistic model in multi-class classification and recognition, we obtain a working model of softmax which come into play in multi-class classification.

3 Experimental results and analysis

In this Experiment, We classify the 79 thousand high-resolution and unlabeled optical remote sensing images via the Internet into the 4500 different classes by a multilayer pre-trained convolutional neural network caffe based. On the unlabeled test dataset, we obtained error rates of 19.7% which run really well than traditional machine learning techniques. With pre-trained model-aided, GTX750Ti GPU, Intel Core I5-4590 processor, we sharply accelerated progress time. The results compared with the published ones, and good agreement is acquired.



(a) The raw image (b) The 1st layer output, conv1 (c) The 2nd layer filters, conv2 (d) The 2nd layer output, conv2

(e) The 3rd layer output, conv3 (f) The 4th layer output, conv4 (g) The 5th layer output, conv5 (h) The 5th layer after pooling, pool5

(i) The 1st fully connected layer, fc6 (j) The 2nd fully connected layer, fc7 (k) The final probability output (l) The top 5 predicted labels

FIGURE 2 Experimental results

The process from the first layer to the seventh layer of the classification of an image is shown in Figure 2-(a-l). Figure 2-(a) show the input original image, Figure 2-(b) show the output of 1st layer called conv1 layer, Figure 2-(c) show the filters of 2nd layer called conv2 layer.

Figure 2-(d) show the output of 2nd layer called conv2 layer, Figure 2-(e) show the output of 3rd layer output called conv3 layer, Figure 2-(f) show the output of 4th layer called conv4 layer.

Figure 2-(g) show the output of 5th layer called conv5 layer, Figure 2-(h) show the 5th layer after pooling called pool5 layer, Figure 2-(i) show the 1st fully connected layer called fc6 layer.

Figure 2-(j) show the 2nd fully connected layer called fc7 layer, Figure 2-(k) show the output of final probability, and Figure 2-(l) show the top 5 predicted labels.

The top 1 predicted label is speedboat, the top 2 predicted label is steamer, the top 3 predicted label is water snake, ocean liner, the top 4 predicted label: leaf hopper, and the top 5 predicted label is terrapin. The predictions are in agreement with the original picture.

4 Conclusion

Convolutional Neural Network naturally learn a hierarchical representation within a recurrent network, thereby implementing a deep network with parameter sharing between the layers. As the application of the Convolutional Neural Network method with GPU aided calculation to object recognition on remote sensing images with high resolution, we can not only obtain visual and numerical results with a high accuracy, but also enjoy many advantages over other traditional methods. A number of experimental results were listed in the paper which proves the effectiveness of convolutional Neural Network with GPU aided calculation in the field of object recognition on remote sensing images with high resolution. As a next work we commit to use a large and very deep Convolutional Neural Network to detect objects belong to a lot of different classes.

References

[1] Foody G M, Mathur A 2004 A relative evaluation of multiclass image classification by support vector machines *IEEE Trans. Geosci. Remote Sens* **42**(6) 1335–43

[2] Chen Yushi 2014 Deep Learning-Based Classification of Hyperspectral Data *IEEE Journal of selected topics in applied earth observations and remote sensing* **7**(6)

[3] Hinton G E, Osindero S, Teh Y W 2006 A fast learning algorithm for deep belief nets *Neural Computation*

[4] Bengio Y, Mesnil G, Dauphin Y, Rifai S 2013 Better mixing via deep representations *Machine Learning*

[5] Krizhevsky A, Sutskever I, Hinton G E 2012 ImageNet classification with deep convolutional Neural Networks NIPS *Neural Information Processing Systems*

[6] Jianmin Liu, Fan Huang, Jun Dai 2015 Image recognition technology for remote sensing based on research and application Machine learning *Journal of Xi'an University of Arts & Science*

[7] Tallman D E, Wallace G G 1997 *Synth. Met.* 90 13

[8] Kroto H W, Fischer J E, Cox D E 1993 *The Fullerenes*

[9] MacDiarmid A G, Epstein A J 1991 in ed.W R Salaneck, D T Clark, E J Samuelson *Science and Applications of Conducting Polymers* Adam Hilger: Bristol 117

[10] Eaton D I 1975 *Porous glass support material*

AUTHORS

Liu Jian-min

Current position, grades: B.S. degree from the University of Hunan (UHN), the M.S. degree from the University of Xiamen (UX), pursuing the Ph.D. degree with the Department of GIS of the University of Central South (UCS)

Scientific interest: remote sensing, GIS and Machine learning