The Signal Processing and Recognition of Street View Images by CNNs and Softmax

Liu Jian-min[1,2], Yang Min-hua[1]
[1] School of Geosciences and Info-Physics, Central South University, Hunan, 410000, China
[2] School of Information, Hunan Institute of Humanities Science and Technology, Hunan, 417000, China
e-mail: liujianmin@csu.edu.cn

***Abstract*—In this paper, we focused on signal processing and proposed a recognition model of street view images based on convolutional neural networks (CNNs) and softmax. On the one hand, this paper processed signal and visualized signal processing activation values of each layer of with CNNs on street view images detection, on the second hand, this paper analyzed and compared them layer by layer. For independent common scene of city street scene image, the kappa coefficient of this method was better than some existing methods.**

Keywords-signal processing; recognition; street view; CNNs

I. INTRODUCTION

By virtual roaming of various cities, tens of thousands of sensors located in the street and street view cars with a mobile sensor platform, street view service is provided by Google Earth and Baidu Map which has provided panoramic views from positions along many roads in China and the world, in other words took 360-degree pictures of high value targets as those sensors can.

This means that Gigabytes received per second and requires large capacity devices for storing and efficiently methods for fast signal processing, analyzing, automatic extraction their features, recognition, data mining and visualizing. In the past, it takes a lot of manpower and time to finish the above task by classic machine learning methods.

A framework for automatic scale selection (Lindeberg T, 1998) is proposed to analysis complex image under unknown environments. It processes the input image, adapts the corresponding scales and focuses autonomously on interesting scale levels and interesting structures in image [1].

An edge detector which integrates generalized type-2 fuzzy logic operator and sobel operator (Claudia I. Gonzalez, Patricia Melin, Juan R. Castro, Olivia Mendoza, Oscar Castillo, 2016) obtains promising results for testing with synthetic images. To reduce the computational cost, a type-1 fuzzy opertaor is implemented after sobel method, followed by an interval type-2 fuzzy operator and a generalized type-2 fuzzy operator [2-4].

The deep belief networks (Hinton and Salakhutdinov, 2006) with sparse variant optimization (Honglak Lee, Chaitanya Ekanadham, Andrew Y. Ng, 2007) detect the edge of image, and was like Gabor filter known to visual area V1, and trustily simulated primary properties of visual area V2. These more higher-order and obscure features matched the essence of subject well [5-7].

A deep convolutional neural network (Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, 2012) which has 650,000 neurons and consists of convolutional layers, max-pooling layers, fully-connected layers, softmax layers, classifies 1.2 million images into the 1000 different classes, and achieves 37.5% error rate which is considerably better than traditional method [8].

However, little attention has been done on signal processing and recognition by self-learning hierarchical features extracting of street view images. These previous results are inconclusive for street view images. This paper aims to provide more studies, in addition, also discuss increase the data-set. Further studies are still essential.

II. MODEL DESIGNING

When training CNNs without pre-training, this article initialized the weights of the 4 convolutional layers and the 3 fully connected layers by using the random initialisation procedure, in the meantime, initialized the biases of each layers with zero.

$$C(V,d)=\sum_{j=1}^{h}\left\|x_j-\hat{x}_j\right\|_2^2+\frac{\lambda}{4}(\|V\|_F^2+\|V'\|_F^2) \quad (1)$$

The process of signal processing with convolution is given below, and for every 684x684 image, the authors used convolution operator of 32x32 kernel-size.

$$x_i^u=g(\sum_{i\in N_i}x_j^{u-1}*h_{ji}^u+d_i^u) \quad (2)$$

Concern to the images with 684x684 pixels, and then the authors got each result by signal processing with convolution order of magnitude to square of 426409. Before importing all the extracted features into classifier layer, it was necessary to reduce the dimension of the signal by max pooling layer.

And then, the authors chose soft-max model which has an outstanding performance on multi-class classification and recognition. Usually, it is helpful to provide an output map for each layer.

Let's look at the visual expression of activation value of each layer, including several convolution layers, pooling layer and fully connected layer different input map.

$$x_i^s=g(\sum_{j=1}^{M_{jm}}\beta_{ji}(h_j^s*x_j^{s-1}+a_i^s) \quad (3)$$

978-1-5090-3822-0/17/$31.00 ©2017 IEEE

We obtained an available model of soft-max. We optimized a multi-layer CNNs and obtained a pre-training classifier. A small labeled data and unlabeled big data sets are combined.

III. Experimental Results

In confirmation hereof, we built lab environment on ubuntu 16.04 LTS and NVIDIA TITAN X GPU with 3584 stream processing units. For we can training the corresponding model in advance, the time will be significantly reduced before classification.

In the experiment, the proportion of the labeled raw picture was adjusted from 0.1% to 1%, and the differences of experimental results were not obvious. The proportion of the labeled data of the experimental results is as follows: 1%. The Table I describe hit rate of there classifiers. Experimental results show that the method has a high degree of discrimination. Unsupervised learning of hierarchical representations processed of signal layer by layer and extracted the most informative features without supervision from these raw pictures, and further, gain the satisfactory classification results by carefully pre-trained recognizer based on little labeled data. The Table II show Kappa and confusion matrix based on CNNs.

For Figure 1-(a-1), the first identified object is fountain ((0.91769981). The first classification label probability was up to 91.77%. It is hard to avoid the conclusion that classification accuracy of the image of scenery theme was quite high.

For Figure 1-(a-2), the first identified object is ashcan, trash can, garbage can, was tebin, ash bin, dustbin, trash barrel, trash bin(0.77993715), the second identified object is mailbox, letter box (0.14689735). The sum of first and second classification label was probability up to 92.68%. The probability sum of first and second classification label was up to 92.68%.

For Figure 1-(a-3), the first identified object is park bench (0.98125941). The first classification label probability was up to 98.13%.

It also was easy to draw the following conclusions that although classification accuracy of the image of municipal facilities theme was high, but there were still some fluctuations

For Figure 1-(a-4), the first identified object is picket fence, paling (0.71303946), the second identified object is radiator (0.17295557), the third identified object is worm fence, snake fence, snake-rail fence, virginia fence (0.089055374).The probability sum of first, second and third classification label was up to 97.51%.

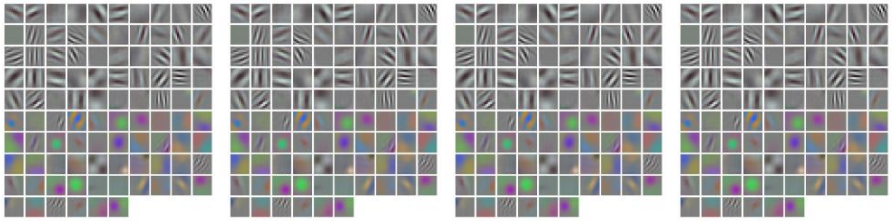
For Figure 2-(a-1), the first identified object is tabby, tabby cat (0.31243578), the second identified object is tiger cat (0.23797166), the third identified object is c Egyptian cat (0.12387238), the fourth identified object is red fox, vulpes (0.10075691), the fifth identified object is catamount'(0.070956819). The probability sum of first, second, third, fourth and fifth classification label only reached 84.6%.

For Figure 2-(a-2), the first identified object is passenger car, coach, carriage 0.72685152), the second identified object is trailer truck, tractor trailer, trucking rig, rig, articulated lorry, semi (0.14145052), the third identified object is recreational vehicle, RV, R.V. (0.042620871), the fourth identified object is minibus (0.025430871). The probability sum of first, second, third and fourth classification label only reached 84.6%.

For Figure 2-(a-3), the first identified object is cinema, movie theater, movie theatre, movie house, picture palace ((0.65424326), the second identified object is streetcar, tram, tramcar, trolley, trolley car ((0.11533506), the third identified object is palace (0.10086082), the fourth identified object is monastery (0.0890835). The probability sum of first, second, third and fourth classification label only reached 84.6%.

For Figure 2-(a-4), the first identified object is street sign (0.81803691), the second identified object is digital clock (0.12396449). The probability sum of first and second classification label only reached 84.6%.

(a-1) (a-2) (a-3) (a-4)



(b-1) (b-2) (b-3) (b-4)

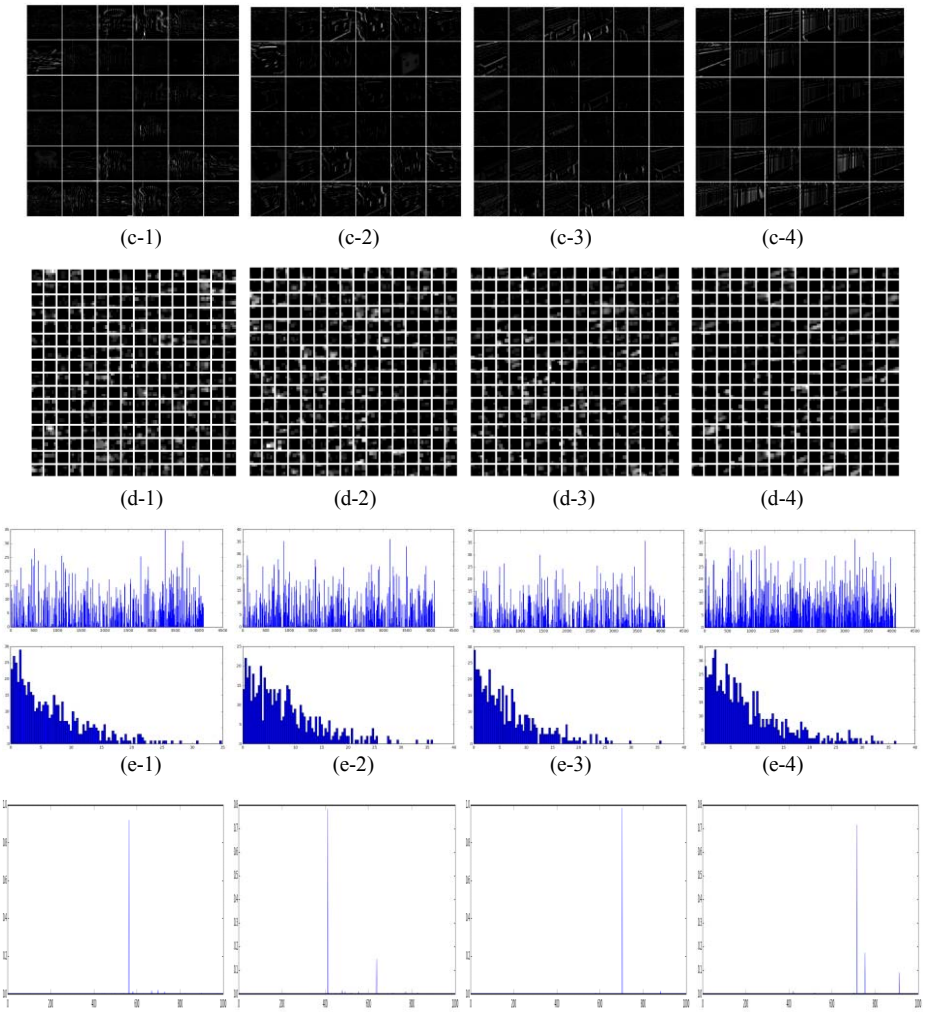
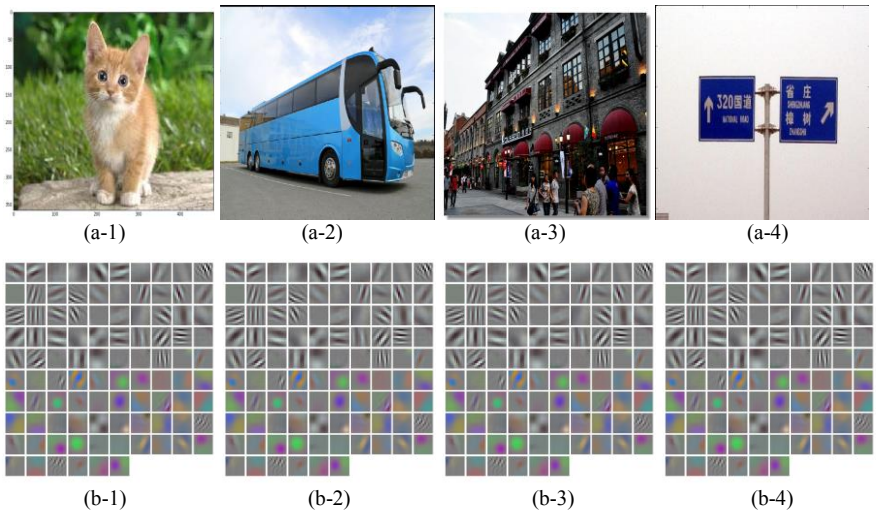


Figure 1. Experimental results: (Y∈ [1,2,3,4]). (a-Y) The input image (b-Y) The first layer output, conv1 (c-Y) The fifth layer after pooling, pool5 (d-Y) The first fully connected layer, fc6 (rectified) (e-Y) The final probability output, prob (f-Y) The top 5 predicted labels. (e-Y) The third layer output, conv3 (f-Y) The fourth layer output, conv4



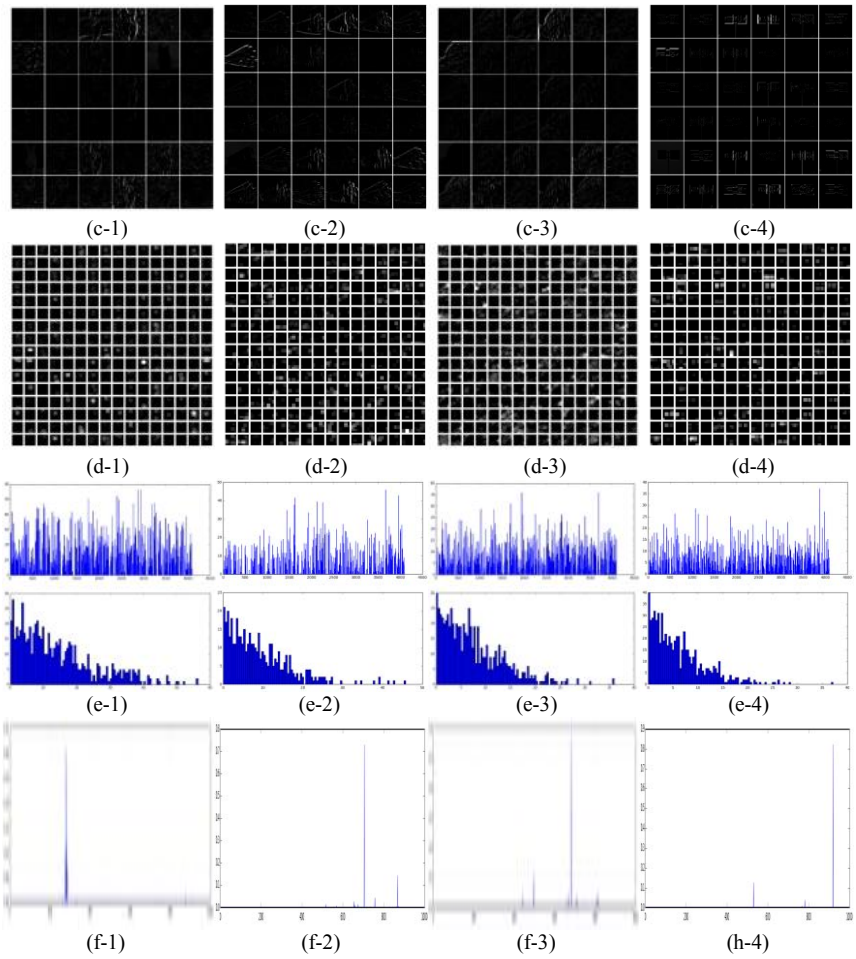


Figure 2. Experimental results: (Y∈[1,2,3,4]). (a-Y) The input image (b-Y)The first layer output, conv1 (c-Y) The fifth layer after pooling, pool5 (d-Y) The first fully connected layer, fc6 (rectified) (e-Y) The final probability output, prob (f-Y) The top 5 predicted labels. (e-Y) The third layer output, conv3 (f-Y) The fourth layer output, conv4

TABLE I. HIT RATE BETWEEN THERE CLASSIFIERS

Class	Animal	Vehicle	Classical architecture	Road signs	Scenery	Municipal facilities	Traffic facilities	*Other*
Sample quantity	4833	5108	4569	5774	3720	7795	5615	19908
Fuzzy C-means	67%	65.2%	59.52%	54.70%	74.10%	41.21%	42.31%	51.10%
CNN	81.13%	84.14%	85.34%	85.78%	86.83%	86.98%	83.70%	72.3%

TABLE II. KAPPA AND CONFUSION MATRIX BASED ON CNNS

Class	Animal	Vehicle	Classical architecture	Road signs	Scenery	Municipal facilities	Traffic facilities	*Other*	*Total*
Animal	3921	0	0	0	0	0	0	0	3921
Vehicle	0	4298	0	0	0	0	0	2019	6317
Classical architecture	0	0	3899	0	0	0	0	0	3899
Road signs	0	0	0	4953	0	0	0	0	4953
Scenery	0	0	0	0	3230	0	0	1277	4507
Municipal facilities	0	0	0	0	0	6780	0	0	6780

Traffic facilities	0	0	0	0	0	0	4700	2214	6914
Other	912	810	670	821	490	1015	915	14398	20031
Total	4833	5108	4569	5774	3720	7795	5615	19908	57322
Kappa	0.7618								

IV. CONCLUSION AND FURTHER WORK

As the application of the CNNs, the excellent results are shown in kappa coefficient based on the confusion matrix. Despite its preliminary character, this study can clearly indicate this model was fairy better than some existing methods like fuzzy c-means.

The Table I shows statistical data in classification accuracies. Compared with conventional classifiers, the designed classifiers based on CNNs both keep high classification accuracies and reduce the time and space complexities.

We can easily come to the following conclusions that CNNs worked went relatively well than the mainstream and traditional techniques, such as fuzzy c-means. As shown in Table II, the value of CNN's kappa co-efficient was 0.7618.

This article preliminarily focuses on signal processing and recognition by learning without supervision hierarchical features extracting of street view images based on with convolutional neural networks. Further studies on signal processing and recognition of street view images will be summarized in our next study.

ACKNOWLEDGMENT

Supported by the Key Research Projects of Hunan Provincial Department of Education (161142).

Supported by Philosophy and Social Science Foundation of Hunan Province (2014YBA224).

REFERENCES

[1] Lindeberg T. Feature Detection with Automatic Scale Selection[J]. International Journal of Computer Vision, 1998, 30(2):77-116.

[2] Gonzalez C I, Melin P, Castro J R, et al. Optimization of interval type-2 fuzzy systems for image edge detection[J]. Applied Soft Computing, 2014, 47:631–643.

[3] Gardiner B, Coleman S, Scotney B. Multiscale Edge Detection using a Finite Element Framework for Hexagonal Pixel-based Images.[J]. IEEE Transactions on Image Processing, 2016, 25(4):1849-1861.

[4] Gonzalez C I, Melin P, Castro J R, et al. An improved sobel edge detection method based on generalized type-2 fuzzy logic [J]. Soft Computing, 2016, 20(2):773-784.

[5] G. E. Hinton* and R. R. Salakhutdinov Reducing the Dimensionality of Data with Neural Networks SCIENCE VOL 313 28 JULY 2006

[6] Honglak Lee Chaitanya Ekanadham Andrew Y. Ng. Sparse deep belief net model for visual area V2[P].NIPS 2007.

[7] Geoffrey E. Hinton A Fast Learning Algorithm for Deep Belief Nets.Neural Computation 18, 1527–1554 (2006)

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pp. 1097–1105, 2012.