Proceedings of the Fifth IIEEJ International Workshop
on Image Electronics and Visual Computing 2017
Da Nang, Vietnam, February 28- March 3, 2017

ROAD SIGN CLASSIFICATION USING SPATIAL PYRAMID CONVOLUTIONAL NEURAL NETWORK

Reza Fuad Rachmadi
Image Media Lab
Kumamoto Univ., Japan

Keiichi Uchimura
Image Media Lab
Kumamoto Univ., Japan

Gou Koutaki
Image Media Lab
Kumamoto Univ., Japan

ABSTRACT

The road sign classification system is a very important task for a lot of intelligent transportation system (ITS) applications, including autonomous car, and road/traffic flow mapping. In this paper, we proposed a spatial pyramid convolutional neural network classifier for road sign classification problem. To provide a detail image structure, two-level of spatial pyramid configuration is used as input to the classifier. The spatial pyramid CNN classifier is built on the top of AlexNet CNN architecture. Our spatial pyramid CNN classifier consists of five streams and each stream is built from convolution layer of AlexNet CNN architecture. At the end of the classifier, we use the fully-connected part of AlexNet to classify the final classification result. As a result, our experiments show that the average accuracy of the classifier is 99.6% and it outperforms our previous approach on Japan image road sign dataset.

1. INTRODUCTION

The road sign classification system is a very important task for a lot of applications, especially for ITS applications such as an autonomous vehicle or high-resolution map creation. There are several approaches to detect the road sign in the road including the new paradigm called V2I (Vehicle-to-Infrastructure) and image-based input system. In this paper, we propose the spatial pyramid CNN to tackle the road sign classification problem using image-based input. Our spatial pyramid CNN classifier is described in our previous paper [2], which proved very good to classify the social event detection in the static image.

2. PROPOSED SYSTEM

Our spatial pyramid CNN classifier consists of five different streams with input image resolution of 256x256 for one stream of level 0 spatial pyramid and 128x128 for another four streams. Each stream was built using the same configuration as first five convolutional layers of AlexNet CNN architecture [4]. At the end of the classifier, the concatenation layer and multi-layer perceptron are used for final classification results. The input for each stream in our spatial pyramid CNN is formed using two level spatial pyramid structure described in [5]. For the concatenation layer and multi-layer perceptron, we use fully connected configuration with 4096-4096-10 neurons configuration for fc6-fc7-fc8 layer. The dropout regularization method with drop ratio of 0.5 is used after layer fc6 and fc7 to prevent the overfitting problem. The complete diagram of our spatial pyramid CNN (SP-CNN) classifier can be viewed in Figure 1.

We use caffe framework [3] for the experiments with Stochastic Gradient Descent (SGD) as training algorithm. The training process is running for 50 epochs (around 19,000 iterations using mini batch of 256 examples) with the learning rate initialize to $\alpha = 0.01$ and decrease 10^{-1} for each 20 epochs. The momentum of 0.9 and weight decay of 0.0005 are used to accelerate the training process and the weights of the network are initialize using the weights from AlexNet CNN architecture [4] trained using ImageNet dataset. In the training process, the training dataset was cropped randomly into 227x227 resolution and with 50% mirroring probability before processed by the network.

3. RESULTS

For the training and testing experiments, we use our Japan road sign dataset which described in [1]. The dataset consists of 7,500 examples for each class with a total number of ten classes. Each class in the dataset divide into two splits, the first 5,000 examples are used for training process and the rest 2,500 examples are used for the testing process.

In the testing process, we use two different input variation of the road sign image, original image and the horizontal mirroring image. The final classification decision is an average output of that two input. Figure 2 shows the confusion matrix of our spatial pyramid CNN classifier using Japan road sign dataset with an average accuracy of 99.67%. From Figure 2, the accuracy of each class in the dataset is very high and some of them have 100% accuracy. Table 1 shows the comparison between our approach and another system described in our previous paper[1]. From Table 1, our spatial pyramid CNN classifier outperforms another method by 2% to 25%. The results show that the spatial pyramid configuration proved to be very effective for road sign classification problem.

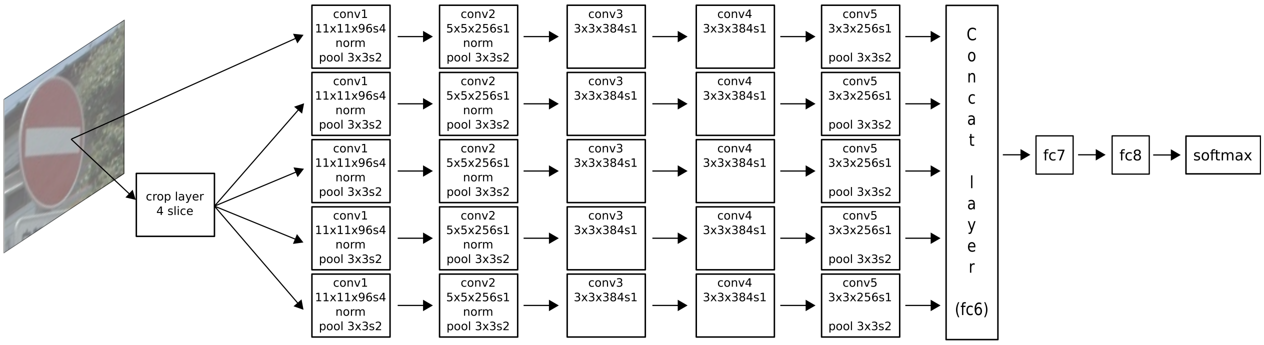

Figure 1: Our spatial pyramid convolutional neural network classifier with two level of pyramid configuration and five stream of convolutional layer.

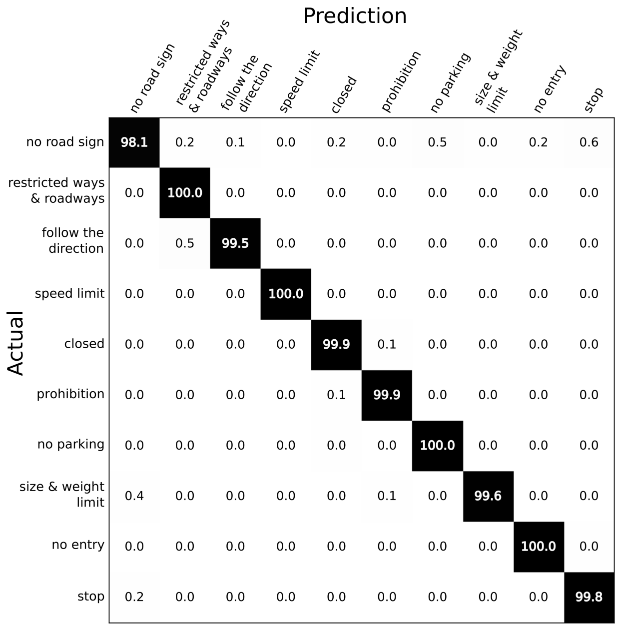

Figure 2: Confusion matrix of our spatial pyramid CNN using Japan road sign dataset.

4. CONCLUSION

We present a spatial pyramid CNN classifier consists of five convolutional streams with spatial pyramid configuration for the input of each stream. Based on the experiments, our spatial pyramid CNN proved very effective to classify road sign image with an average accuracy of 99.67% using Japan road sign dataset. Comparing with another approach of road sign classification system listed on [1], our approach outperforms another method by 2% to 25%.

Design the classifier based on more shallow CNN architecture may reduce the time execution while maintaining the average accuracy of the system. By combining our approach with the road sign detection system produce a complete road sign recognition system that can be used for a lot of applications, especially for intelligent transportation system applications.

Table 1: Comparison of our spatial pyramid CNN with another approach.

Method	Accuracy
BoF (1,000 codebooks) + SVM [1]	73.88%
NiN classifier [1]	79.74%
C-CNN [1]	97.94%
SP-CNN	**99.67%**

5. REFERENCES

[1] R. F. Rachmadi, Y. Komokata, K. Uchimura, G. Koutaki: "Road sign classification system using cascade convolutional neural network," International Journal of Innovative Computing, Informatics, and Control, Vol. 13, Issue 1, pp. 95-109, 2017.

[2] R. F. Rachmadi, K. Uchimura, G. Koutaki: "Spatial pyramid convolutional neural network for social event detection in static image," International Student Conference on Advanced Science and Technology (ICAST), 2016.

[3] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell: "Caffe: Convolutional Architecture for Fast Feature Embedding," 22nd ACM international conference on Multimedia, pp. 675-678, 2014.

[4] A. Krizhevsky, I. Sutskever, G.E. Hinton: "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems (NIPS), pp. 1097-1105, 2012.

[5] S. Lazebnik, C. Schmid, J. Ponce: "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," IEEE Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 2169-2178, 2006.