CLASSIFICATION OF MOVING VEHICLES BASED ON
CONVOLUTIONAL NEURAL NETWORK

Hosung Park1, Eun Som Jeon2, Minkyu Lim3, Donghyun Lee4, Unsang Park5 and Ji-hwan Kim*6
1, 3, 4, 5, 6 Sogang University, Korea
(E-mail: {hosungpark, lmkhi, redizard, unsangpark, kimjihwan}@sogang.ac.kr)
2 Dongguk University, Korea
(E-mail: jeunsom@dgu.edu)

ABSTRACT - This paper suggests a method of applying convolutional neural network (CNN) to real-time moving-vehicle image dataset, and experiments its performances. Compared to Support vector machine (SVM), the CNN led to a 13.59% increase in performance.

Keywords: Neural network, Convolutional neural network, Vehicle classification

1. INTRODUCTION

The use of computer vision to classify vehicles is widespread in applications such as closed-circuit television (CCTV) and vehicle sensors. In particular, the real-time detection of moving vehicles is used in vehicle event data recorders and for tracking criminals. This study proposes vehicle-image classification based on convolutional neural network (CNN), which is a type of deep neural network (DNN). Section 2 explains the CNN and the vehicle dataset. Section 3 explains the classification using the CNN. The conclusion in Section 4 provides a discussion of the experimental results and future work.

2. RELATED WORK

2.1. Convolutional neural network

A CNN is a process for properly extracting features from a given image by using a particular filter. A key process is to generate the most appropriate feature map, for which the CNN is the learning process.

Figure 1 Convolution processing

Figure 1 depicts the convolution and pooling layers; the subsampling layer on the right corresponds to the pooling layer. As shown in Fig. 1, a filter shape of a certain size is created to extract features from a designated part of the image. In this case, the number of parameters to be considered for a given image is reduced, which allows for effective feature extraction. However, constructing a layer with such a filter shape produces more features with similar values because of the extent of the overlapping parts on the image. To overcome this drawback, a subsampling method—called the pooling layer—is applied to

* Corresponding author: Ji-Hwan Kim
reduce the size of the feature map. A CNN comprises a convolution layer and a pooling layer; the CNN referred to in this study uses such a combination, formed by a deep neural net [1].

2.2. Audiovisual vehicle dataset

An audiovisual vehicle (AVV) dataset [2] was used for experiments of this study, consisting of 961 vehicle sample images and vehicle-movement sounds. This study used the sample images only, dividing them into five classes: bikes, buses, sedans, trucks, and vans.

3. VEHICLE CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK

In order to implement CNN, augmentation is involved for each of the 961 pieces of data. For the augmentation, this study implemented transformation methods: scaling (85%), brightness transformation (5%), gamma-value transformation (5%), rotation (±15°), smoothing (7%), and sharpening (7%). In total, 64,821 further sample images were obtained, which were then normalized and resized to 100×41. The dataset was divided into training data (60%), validation data (20%), and test data (20%). In the experiment, we conducted three convolutional processing and a fully connected layer with 800 hidden nodes and five output layers. The filter size for the convolution and pooling layer was 5×5 pixels and 2×2 pixels, respectively. The best CNN performance was obtained for a learning rate of 0.01. The numbers of kernels (feature maps) were set as 40, 80, and 160 per convolution layer. Previous research [2] used a support vector machine (SVM) to categorize images and the results of classification (images of only cars from the original data) showed an accuracy of approximately 80.39%. The proposed method using a DNN was intended to be more accurate than [2]. The results of doing so showed that the application of the CNN led to an accuracy of 93.98% for the test dataset. Compared to [2], the performance was enhanced by 13.59%, proving that application of the CNN results in better performance than that of the SVM.

4. CONCLUSION

This study proposed a method of applying a CNN to a real-time moving-vehicle image dataset, and confirmed its performance through experiments. Applying the CNN led to a 13.59% increase in performance over that of an SVM. Follow-up studies are planned to visualize various measurements (e.g., accuracy for each class) by developing a confusion matrix module, and to increase the recognition rate by performing image preprocessing at the front-end.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NO. NRF-2014R1A1A1002197).

REFERENCES