International Journal of Engineering Applied Sciences and Technology, 2017
Vol. 2, Issue 9, ISSN No. 2455-2143, Pages 80-84
Published Online January 2018 in IJEAST (http://www.ijeast.com)

TRAFFIC SIGN RECOGNITION USING A MULTI-TASK CONVOLUTIONAL NEURAL NETWORK

Dr. S.V. Shinde
Department of IT
Pimpri Chinchwad college of Engineering
Pune, Maharashtra, India

Arshiya Sayyad
Department of IT
PCCOE

Uzma Shaikh
Department of IT
PCCOE

Abstract— **Traffic sign recognition has been studied for many years, most existing works are focused on the symbol-based traffic signs. Our project will propose a new data-driven system to recognize all categories of traffic signs, which include both symbol-based and text-based signs, in video sequences captured by a camera mounted on a car. The system consists of three stages, traffic sign regions of interest (ROIs) extraction, ROIs refinement and classification, and post-processing. Traffic sign ROIs from each frame are first extracted using maximally stable extremal regions on gray and normalized RGB channels. Then, they are refined and assigned to their detailed classes via any machine learning algorithm, which is trained with a large amount of data, including synthetic traffic signs and images labeled from street views. The post-processing finally combines the results in all frames to make a recognition decision and the output will be displayed to the driver on screen.**

Keywords— **Traffic sign detection, traffic sign classification, convolutional neural network, multi-task learning; *Traffic sign recognition; Driver support systems; Intelligent vehicles;***

I. INTRODUCTION

AUTOMATIC traffic sign detection and recognition is an important part of an advanced driver assistance system. Traffic symbols have several distinguishing features that may be used for their detection and identification. They are designed in specific colors and shapes, with the text or symbol in high contrast to the background. Because traffic signs are generally oriented up right and facing the camera, the amount of rotational and geometric distortion is limited. Information about traffic symbols, such as shape and color, can be used to place traffic symbols into specific groups; however, there are several factors that can hinder effective detection and recognition of traffic signs. These factors include variations in perspective, variations in illumination (including variations that are caused by changing light levels, twilight, fog, and shadowing), occlusion of signs, motion blur, and weatherworn deterioration of signs. Road scenes are also generally very cluttered and contain many strong geometric shapes that could easily be misclassified as road signs. Accuracy is a key consideration, because even one misclassified or undetected sign could have an adverse impact on the driver.

To recognize traffic signs in an image, most popular methods include two steps: Detection and Classification. There are a lot of researchers working on this challenging task with the already popular or specially designed vision algorithms. However, it is not easy to compare these methods since there did not exist a public available data set until the release of the German Traffic Sign Recognition Benchmark(GTSRB) [1] and German Traffic Sign Detection
Benchmark (GTSDB) [2] in 2011 and 2013 respectively. Since then, researchers can evaluate and compare their algorithms on the same benchmarks.

Nevertheless, there still exist some defects in the GTSDB and GTSRB: 1) they include only three categories of symbol-based traffic signs with regular shape and color which are relatively easy to detect and classify, while text-based traffic signs are more challenging; 2) the GTSDB only includes static images, but in real scenarios, continuous video captured by an in-vehicle camera is useful for detection and classification [3];
3) The final task of traffic sign recognition is to know the existing signs in a scene, but the two benchmarks separate it into two independent tasks with different datasets.

To alleviate these problems, we propose a new system to recognize existing traffic signs from video input and evaluate its performance on a new challenging data set with the following features: 1) it contains both the symbol-based and text-based traffic signs, up to seven categories, which is contrasted to the previous three symbol-based categories; 2) instead of a static image, each sample in the data set is a short video of 5 ~ 20 low quality frames which is captured from an in-vehicle camera. Some examples in the data set are shown in Fig. 1. It can be seen that symbol-based traffic signs have the same appearance with discrepancies in viewpoint, illumination, blur, background and so on, while text-based signs may have very different appearances even within the same class.

International Journal of Engineering Applied Sciences and Technology, 2017
Vol. 2, Issue 9, ISSN No. 2455-2143, Pages 80-84
Published Online January 2018 in IJEAST (http://www.ijeast.com)

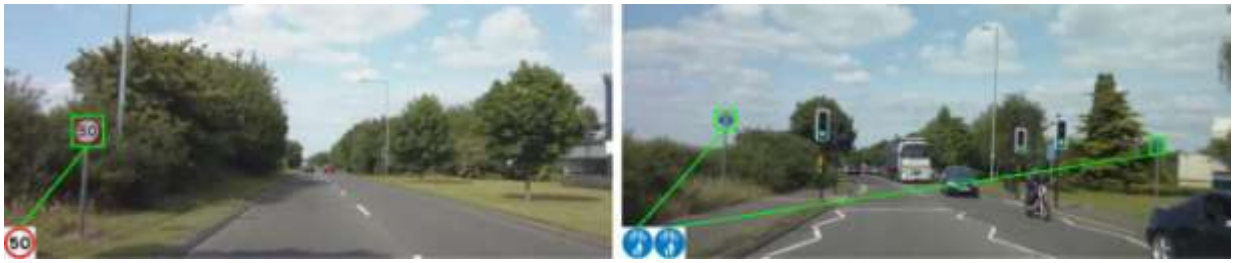


Fig. 1.Examples of the proposed road sign detection

Fig. 2. Pipeline of the proposed traffic sign recognition system

Our traffic sign recognition system consists of three stages: traffic sign regions of interest (ROIs) extraction, ROIs refinement and classification, and post-processing. First, for each frame in the video, traffic sign ROIs are detected with Maximally Stable Extremal Regions (MSERs) [4] on multi-channel images. Then, to refine and classify the ROIs, a multi-task Convolutional Neural Network (CNN) is proposed. Specifically, the ROIs are first fed to a binary classification layer, and only the positive ones are further classified with a deep multi-class classification network. The network is trained end-to-end with a large number of data, which consists of training data, synthetic signs and images labeled from street view. Finally, recognition results from each frame are fused to get the final results of the video. Such a system pipeline is illustrated in Fig. 2.

The main contributions of this paper are as follows: 1) while many existing works are focused on symbol-based traffic signs, we process all categories, including both symbol-based and text-based ones; 2) a new multi-task CNN which treats the tasks of ROIs refinement and classification jointly is proposed and recognition system;3)to address the problem of relatively small amount of labeled traffic signs, two kinds of data acquisition methods, street view images and synthetic images, are combined to obtain large number of labeled samples with low cost; 4) our system achieves the best result in a newly released, challenging data set.

II. BACKGROUND

A. Traffic Sign Detection

The images obtained by the camera are often of poor quality due to the sophisticated environment conditions. Low-level image pre-processing can be used to enhance the traffic sign regions of the captured images, which makes it easier for subsequent tasks. The most common way is transforming images into a new color space where the signs are more distinct. Many color spaces have been used, such as HSI [5], improved HLS [6] and normalized color space [7], [8]

Another pre-processing method is using machine learning to learn the color space mapping from data. Reference [9] proposed a color probability model which can enhance the main color of the signs while suppressing the background regions. In [8] and [10], an SVM classifier was trained to map each pixel in color images to a gray value which has high response in sign regions.

In the early stage of object detection, it was popular to use threshold based methods [5], [6]. In [7], different threshold based segmentation methods were compared. These kinds of methods are not robust in complex environment with unpredictable lighting conditions.

Recently, machine learning based object detection is becoming dominant in the research community. On the traffic sign detection, there are sliding window based methods and region of interest (ROI) based methods.

Another approach is to first extract regions of interest (ROIs) and then filter out non-object ROIs with a classifier. Compared to the sliding window based methods, it reduces the computational time and does not need to tune the parameters of sliding window. An important consideration of this method is the recall rate of target objects among the extracted ROIs. It is expected to have as high recall rate as possible while keeping the number of ROIs as low as possible. Given that traffic signs are designed with a large part of uniform region, MSERs have been proved to be very effective in extracting such ROIs [8], [9]. In a coarse sliding window method was used to extract ROIs. Template matching was also used for ROIs extraction in [10]. Filtering out non-sign objects from ROIs can be treated as a classification task. SVM classifier with HOG features is the most popular framework due to its excellent performance [8], [9]. Some other methods like Convolutional Neural Network (CNN) [10], Extreme Learning Machine were also used.

International Journal of Engineering Applied Sciences and Technology, 2017
Vol. 2, Issue 9, ISSN No. 2455-2143, Pages 80-84
Published Online January 2018 in IJEAST (http://www.ijeast.com)

B. Traffic Sign Classification

Traditional methods for classification include feature extraction and classifier training. Some combinations reported in literature include a cascade of SVM classifiers with HOG features [8], K-d trees and Random Forests with Distance Trans-forms and HOG features, MLP(Multi-Layer Perceptron) with radial histogram features, ANN (Artificial Neutral Network) with RIBP (Rotation Invariant Binary Pattern) based features, SVM with LIPID (local image permutation interval descriptor) etc. In dense SIFT features, HOG features and LBP features were first extracted, then they were encoded through locality-constraint linear coding (LLC) and the resulting codes were pooled by spatial pyramid pooling(SPM).

The three different feature representations were concatenated as the final features of a traffic sign, and a linear SVM was used as the classifier. In general, it is very laborious and difficult to design a good feature.

Convolutional Neural Network (CNN) which can be trained without the need of hand-designed features is popular now days. In Multi-column CNNs which train multiple CNNs with different weight initialization or data pre-processing, were proposed to classify traffic signs. It won the first place in GTSRB competition in 2011. In two stage features in CNN, i.e. local and global features, were fused to recognize traffic signs. This method got the 2nd-best accuracy in the same competition. In a modified version of cross entropy loss was used in training CNN, obtaining an even better result than . Although CNN has shown its excellent performance in image classification, how to design a good network architecture and train a workable model are still challenging tasks.

To handle the geometry variations of traffic signs, data augmentation was used to enlarge the training data set.

Fig. 3. Raw color image and four channels used to extract MSERs. (a) Original color image. (b) Gray and normalized R,G,B channel.

Another method is to eliminate the geometry variations. In the traffic signs were first classified into several super classes, for each of which perspective adjustment was performed with a specially designed method. Then, the adjusted signs were classified into their detailed classes. Recently, Spatial Transformer Network (SPN) was proposed in which can explicitly learn geometry parameters of transformation, and be robust to the geometry variations of input images. It was shown in [9] and [10] that SPN could achieve state-of-the-art result on GTSRB without the need of complicated tricks used in previous works.

III. TRAFFIC SIGN DETECTION AND RECOGNITION SYSTEM

A. Overview of the System

The proposed system consists of the following two main stages: detection and recognition. The complete set of road signs used in our training data and recognized by the system is shown in Fig. 2. Candidates for traffic symbols are detected as MSERs, as described by Matas et al. [1]. MSERs are regions that maintain their shape when the image is threshold at several levels. This method of detection was selected due to its robustness to variations in contrast and lighting conditions. Rather than detecting candidates for road signs by border colour, the algorithm detects candidates based on the background colour of the sign, because these backgrounds persist within the MSER process. Our proposed method, as described in detail in the following section, is broadly illustrated in Fig. 3.

B. Detection of Road Signs as MSERs For the detection of traffic symbols with white background, MSERs are found for a grayscale image. Each frame is binaries at a number of different threshold levels, and the connected components at each level are found. The connected components that maintain their shape through several threshold levels are selected as MSERs. Fig. 4 shows different thresholds for an example image with the connected components coloured. It is shown that the connected component that represents the circular road symbol maintains its shape through several threshold levels. This helps ensure robustness to both variations in lighting and contrast. Several features of the detected connected component regions are used to further reduce the number of candidates. These features are width, height, aspect ratio, region perimeter and area, and bounding-box perimeter and area. Removing the connected components that do not match the requirements helps speed up the process and improve accuracy.

We approach the detection of traffic symbols with red or blue backgrounds in a slightly different manner. Rather than detecting MSERs for a grayscale image, the frame is first transformed from red–green–blue (RGB) into a "normalized red/blue" image ΩRB such that, for each pixel of the original image, values are found for the ratio of the blue channel to the sum of all channels and the ratio of the red channel to the sum of all channels. The greater of these two values is used as the pixel value of the normalized red/blue image.

Although MSER offers a robust form of detection for traffic signs in complex scenes, it can be computationally expensive. Therefore, to increase the speed, we threshold only at an appropriate range of values rather than at every possible value, which is the norm in the original MSER [1]. Fig. 7 shows the number of used thresholds plotted against the processing time and accuracy of detection. The thresholds were evenly spaced between the values 70 and 190, because the MSERs that represent road signs generally appear within this range. The number of thresholds selected was 24, which, in this example, corresponds to 94.3% accuracy and 50.1-ms processing time.

International Journal of Engineering Applied Sciences and Technology, 2017
Vol. 2, Issue 9, ISSN No. 2455-2143, Pages 80-84
Published Online January 2018 in IJEAST (http://www.ijeast.com)

B. Multi-Task CNN for ROIs Refinement and Classification

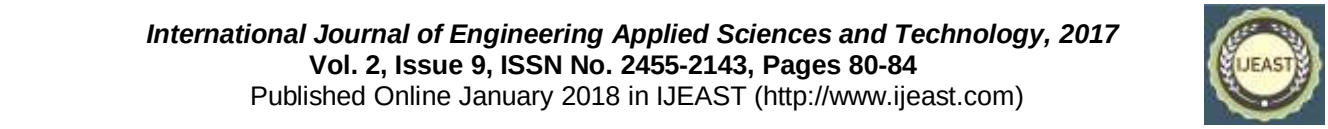
After the traffic sign ROIs extraction, traffic signs and a large number of backgrounds are obtained. The tasks of this stage are to filter out backgrounds and determine the detailed classes of the remaining ROIs, namely ROIs refinement and classification, respectively. Traditionally, most related works tackled the two tasks separately. For ROIs refinement, the widely used method is using SVM classifier with HOG features. For the task of classification, CNN is the mainstream method, which has been proved to be an excellent model in computer vision. It can be trained end-to-end from the image data, and no manually designed features are needed any more. In this paper, we propose a new CNN architecture which unifies the two tasks. We call this architecture as multi-task CNN. There are two important issues in the CNN-based method, structure of the network and large amount of training data. In this subsection we will describe the structure of proposed CNN, and the method for acquiring enough training data will be introduced in the next section.

There are two decision layers in the proposed multi-task CNN. One is called binary classification layer for distinguishing backgrounds and traffic signs, and the other is called multi-class traffic sign classification layer. They correspond to the tasks of ROI refinement and classification respectively in the traditional methods. Here, the binary classification layer aims to fast eliminate most background ROIs and allows some hard backgrounds to pass, which will be removed by the multi-class classification layer based on deeper features. During the training and testing stages, all ROIs are first fed to the binary classification layer, and only the positive ROIs are fed to the next part of the network to obtain the detailed classes. In the training stage, loss from both decision layers are used to jointly optimize the network.

The basic structure of the multi-task CNN is shown in Fig.
4. The *conv (k,m)* means a convolutional layer with kernel size $k \times k$, and output channel number m. Appropriate padding precedes all convolutional operations so as to keep the width and height of input channel, and the stride is 1 in all operations. The *relu* denotes the rectified linear unit (ReLU) layer. The *maxpooling(k)* means max pooling layer with kernel size $k \times k$ and stride k.

To design a good network, it is important to consider the depth of the network. In this paper, we define and compare four network structures with different depths and convolutional kernel sizes. They have the similar basic structure as shown in Fig. 4. Specifically, the input of network is color image with size of 48 × 48. The node number of two decision layers are 2 and 73, which represent background and traffic sign in the first case and the number of sign classes and an additional background class in the second case. ReLU layer is added after each

International Journal of Engineering Applied Sciences and Technology, 2017
Vol. 2, Issue 9, ISSN No. 2455-2143, Pages 80-84
Published Online January 2018 in IJEAST (http://www.ijeast.com)



full connected layer except the final decision layers. Dropout layer with a probability of 0.5 is added after the two full connected layers which connect from pooling layers. The detailed structures of the four models are listed in Table I. In the shallow models, all convolutional layers are with big filters, while the deep models split big filters into a few number of small filters with fixed size of 3×3 . Each deep or shallow model includes 2 or 3 max pooling layers, so there are four models in total.

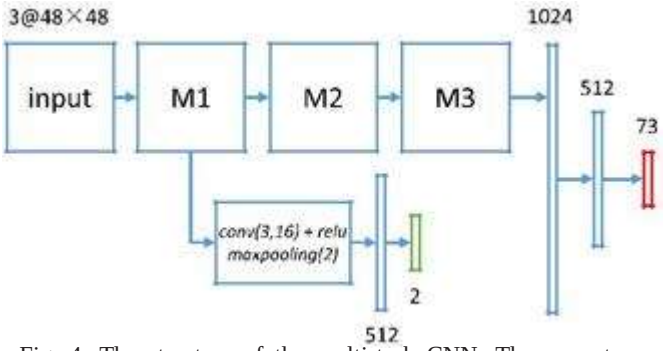


Fig. 4. The structure of the multi-task CNN. There are two decision layers, one for binary classification (green layer) and another for multi-class classification (red layer).

C. Post Processing

Each extracted traffic sign ROI is fed to the above multi-task CNN to get the classification result. This operation is applied to each frame of a test video. In nearby frames of the video, the recognition results may be slightly different due to their different appearance in different frames. For this reason, it is necessary and important to fuse results from all frames to get the final recognition result in a short video.

IV. CONCLUSION

We have proposed a novel real-time system for the automatic detection and recognition of traffic symbols. Candidate regions are detected as MSERs. This detection method is significantly insensitive to variations in illumination and lighting conditions. Traffic symbols are recognized using HOG features and a cascade of linear SVM classifiers. A method for the synthetic generation of training data has been proposed, which allows large data sets to be generated from template images, removing the need for hand labeled data sets. Our system can identify signs from the whole range of ideographic traffic symbols , which form the basis of our training data. The system retains a high accuracy at a variety of vehicle speeds.

V. REFERENCES

[1] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German traffic sign recognition benchmark: A multi-class classification competition," in Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2011, pp. 1453–1460.

[2] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, "Detection of traffic signs in real-world images: The German traffic sign detection benchmark," in Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN), Aug. 2013, pp. 1–8.

[3] A. Bolovinou, C. Kotsiourou, and A. Amditis, "Dynamic road scene classification: Combining motion with a visual vocabulary model," in Proc. 16th Int. Conf. Inf. Fusion (FUSION), Jul. 2013, pp. 1151–1158.

[4] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," Image Vis. Comput., vol. 22, no. 10, pp. 761–767, 2004.

[5] A. de la Escalera, J. M. Armingol, and M. Mata, "Traffic sign recognition and analysis for intelligent vehicles," Image Vis. Comput., vol. 21, no. 3, pp. 247–258, 2003.

[6] H. Fleyeh, "Color detection and segmentation for road and traffic signs," in Proc. IEEE Conf. Cybern. Intell. Syst., vol. 2. Dec. 2004, pp. 809–814.

[7] W. Ritter, F. Stein, and R. Janssen, "Traffic sign recognition using colour information," Math. Comput. Model., vol. 22, nos. 4–7, pp. 149–161, 1995.

[8] J. Greenhalgh and M. Mirmehdi, "Real-time detection and recognition of road traffic signs," IEEE Trans. Intell. Transp. Syst., vol. 13, no. 4, pp. 1498–1506, Dec. 2012.

[9] Y. Yang, H. Luo, H. Xu, and F. Wu, "Towards real-time traffic sign detection and classification," IEEE Trans. Intell. Transp. Syst., vol. 17, no. 7, pp. 2022–2031, Jul. 2016.

[10] Y. Wu, Y. Liu, J. Li, H. Liu, and X. Hu, "Traffic sign detection based on convolutional neural networks," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Aug. 2013, pp. 1–7.