Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 04, April-2019

Simultaneous Detection and Boundary Estimation of Traffic Signs

Jisha Elizabeth Shaji
PG Student in Electronics and Communication Engineering
Mount Zion College of Engineering
Kadammanitta

Hari S
Asst.Professor in Electronics and Communication Engineering
Mount Zion College of Engineering
Kadammanitta

***Abstract*—This paper presents an overview of simultaneous traffic sign detection and boundary estimation. It describes the characteristics and requirements and also difficulties behind the road sign detection and recognition of the road signs. It shows the convolutional nueral network technique used for the recognition and classification of the road signs. The paper introduces a traffic sign detection system that accurately estimates the exact boundary of traffic signs using 2D pose and shape class prediction problem by convolutional neural network (CNN).In navigation system, estimation of accurate boundary of traffic signs is important where road signs used as 3-D landmarks. In the recent previous traffic sign detection system based on CNN only provides bounding boxes. Here the system provides precise boundary of traffic sign which helps the detection and recognition of signs properly. The method used is end-to-end trainable and more robust to occlusion, blurred images and small targets than other boundary estimation method. The CNN-based traffic sign detection and recognition network method gives a frame rate higher than seven frames/second.It also provides highly accurate and robust traffic sign detection and boundary estimation results on a low power mobile platform.**

Keywords*— *Traffic sign detection, traffic sign recognition, and convolutional neural network.

INTRODUCTION

TRAFFIC sign detection is a major crisis in intelligent vehicles, traffic sign recognition provides critical information like directions and alerts in autonomous driving or driver assistance systems. Another application of traffic sign detection is to compliment the navigation systems of intelligent vehicles, by using traffic signs as distinct landmarks for mapping and localization. Contrary to natural landmarks with arbitrary appearance, traffic signs have standard appearances such as shapes, colors, and patterns defined in regulations [1]. This makes it efficient and robust to be detected and matched under any conditions, thus making it a preferable choice as landmarks for road map reconstruction. For reconstructing detected traffic signs to a 3D map, point-wise correspondences of boundary corners of the signs across multiple frames is used, and then 3D coordinates of the boundary corners are computed by triangulation using the camera pose and internal parameters of the camera. For accurate triangulation of 3D position, estimation of boundary of signs with pixel-level accuracy is required. Existing traffic sign detection systems do not have this capacity as they only estimate bounding boxes of traffic signs [2]. Pixel-wise prediction methods such as semantic image segmentation that is applied successfully for road scenes can replace boundary estimation. But, it requires time consuming algorithms that can severely harm the performance of real-time systems for vehicles. By using templates of traffic signs, we effectively utilizes prior information of target shapes. This enables robust boundary estimation of traffic signs that are unclear, but it’s difficult in pixel wise prediction such as contour estimation and segmentation.

The 2D pose of target signs are encoded as 8-dimensional vectors which is defined as coordinates of four vertices, and it can be accurately determined by CNN which predicts the scores of each shape label. Using the predicted 2D poses and shape labels, the boundary corners of a traffic sign are computed by projecting the boundary corners of a corresponding template image of the sign using the predicted pose.

With respect to input resolution that is 1280 × 720 ,our method achieves a detection rates which is higher than 0.88 mean average precision (mAP), and boundary estimation error less than 3 pixels.The projecting boundary corners (matrix-vector products) requires less computation time and most of the required computation is from the CNN forward propogation which can be accelerated by GPUs. Combining with our efforts to find a base network architecture that provides the best trade-off between accuracy and speed, our precise boundary detection system can run on mobile platforms with frame rates higher than 7 frames per second (FPS) with affordable traffic sign detection and boundary estimation accuracy

I. RELATED WORK

Recently, the great advanced work on object detection have been achieved by CNN. Besides the discriminating power of CNN on object category classification, the detection networks shows the capability of accurate localization by regression on object location. Two different architectures of detection networks are currently being developed: direct detection and region proposal based detection. In direct detection, predictions on the position (by regression) and class (by classification) of target objects are directly obtained from convolution feature layers, resulting in relatively faster run time. On the other hand, region proposal based method first generates a number of candidate regions regardless of their classes, and then performs prediction on object position and class for each candidate region. By performing regression and classification twice in different stages of the network pipeline, the region proposal based methods pursue more accurate detection, with relatively slower run time than direct detection methods. For the case of traffic sign detection for autonomous driving, direct detection method is adequate due to the latency of detection under limited computational resources. Although most of the recent CNN object detection methods provide accurate bounding box and class label prediction, further processes should follow to obtain precise object boundaries from the predicted bounding boxes. To resolve this issue, boundaries of traffic signs are simultaneously obtained as segmentation masks by Over Feat-style convolutional neural network trained on multiple tasks comprising bounding box detection, mask segmentation, and sign category classification [3]. However, the prediction of pixel-wise segmentation [4] masks requires intensive computation which results in very slow speed of the network. On the other hand, we propose boundary estimation method which does not require pixel-wise segmentation and thus enables fast detection speed.

II. PROPOSED SYSTEM

The overall procedure of the proposed method is illustrated in the figure.

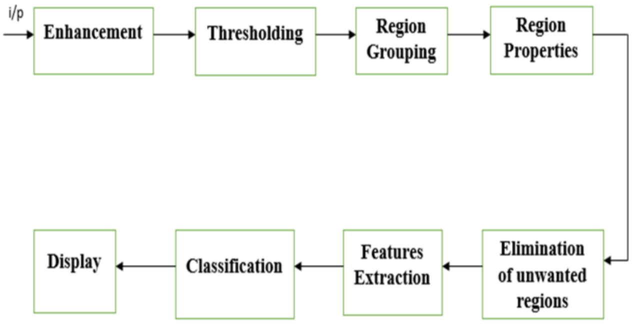


Fig 1: Block Diagram

In this work, we create a CNN block where predictions are directly preformed across multiple feature levels. The main difference of our network with the other detection networks is its prediction of output: instead of predicting bounding box of traffic sign,the network performs estimation of accurate boundary of corresponding traffic signs. When an input image (traffic sign) is detected while a car is moving,it is enhanced [5] .

The process of adjusting digital images is image enhancement therefore the results are more suitable for display. The main Objective of image enhancement is process the image (e.g. contrast improvement, image sharpening) so that it is better suited for identifying key features.After that it undergoes thresholding, Image thresholding is a simple, yet effective, way of partitioning an image into a foreground and background. The image analysis technique is a type of image segmentation that converts grayscale images into binary images. Image thresholding is mostly needed in images with high levels of contrast. Common image thresholding algorithms include histogram and multi-level thresholding.It then undergoes to a process of region grouping and region properties. Grouping is a process of extracting and representing information from an image is to group pixels together into regions of similarity. Image regions is also called objects. It can be contiguous or discontiguous. A region in an image can have properties, such as an area, center of mass, orientation, and bounding box.After that the image undergoes for feature extractionfeature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant.Feature extraction is a dimensionality reduction process, where an initial set of raw variables is reduced to particular manageable groups (features) for processing. When the input data to an algorithm is too large to be processed and it is suspected to be redundant, then it can be transformed into a reduced set of features. The initial features subset is called feature selection. The features that are selected contains the required information from the input data, so that the desired task can be performed by using this reduced representation of information instead of the complete initial data.Finally the image is classified with their features after passing through the convolutional layer [6].

Convolution neural network algorithm is a multilayer perceptron that is the special design for identification of two-dimensional image information. Always has more layers that is input layer, convolution layer, sample layer and output layer. A 2-D convolutional layer performs sliding convolutional filters to the input. The layer convolves the input vertically and horizontally by moving the filters and then computes the dot product of the weights and the input, and then adds a bias term.

The convolutional layer is the important building block of a CNN. The parameters consist of a group of learnable filters or kernels, which have a small receptive field, but extended through a full depth of the input volume. Each filter is convolved across the width and height of the input volume and computes the dot product between the entries of the filter and the input during forward pass and producing a 2-dimensional activation map of that filter.After convolution it undergoes ReLu layer.It effectively removes negative values from an activation map by setting them to zero Another important concept of CNNs is pooling, which is a form of non-linear down-sampling. There are several non-linear functions to implement pooling among which max pooling is the most common. It divides the input image to a set of non-overlapping rectangles and for each such sub-region it maximize the outputs and then feed into fully connected layers.

After a series of convolution 2D poses and shape class probabilities are obtained by two separated convolutional layers which is pose regression layer and shape classification layer, combined with successive operations that convert the convolution outputs to the 2D pose values and class probabilities, respectively which is carried out through Softmax and Decoder. Finally, the obtained 2D poses and shape class probabilities is used to compute boundary corners.

III. RESULTS

The output of the detection stage is a group of objects that could be probable traffic road signs. This is forwarded to the recognizer for further evaluation, and then to the classifier decides whether the detected objects are either rejected objects or road signs, and in this case the classifier responds with a sign code.For a good recognizer, some parameters should be taken into consideration. Firstly, the recognizer should provide a well discriminative power and low computational cost. Secondly, it should be robust to orientation such as vertical or horizontal, the size, and the position of the traffic sign in the image. Thirdly, it should be robust to noise. Fourthly, for real time applications the recognition should be carried out quickly when a sign is detected. Furthermore, the classifier should have the capability to learn a large number of classes and as much a priori knowledge about road signs should be employed into the classifier design, as possible. Our method is more robust to partial occlusion, cluttered background, and close signs which cannot be directly handled by segmentation algorithm.The recognized output is displayed in the figure shown below.The proposed method provides detection frame rate higher than seven frames/second and high accuracy.The confusion matrix of recognition is also shown below

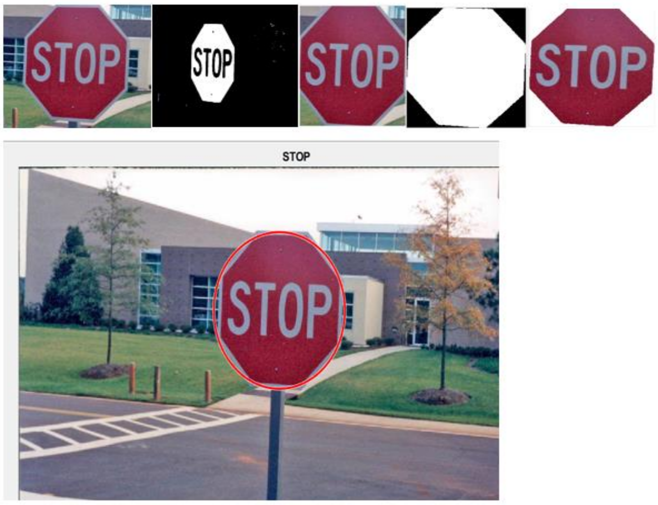
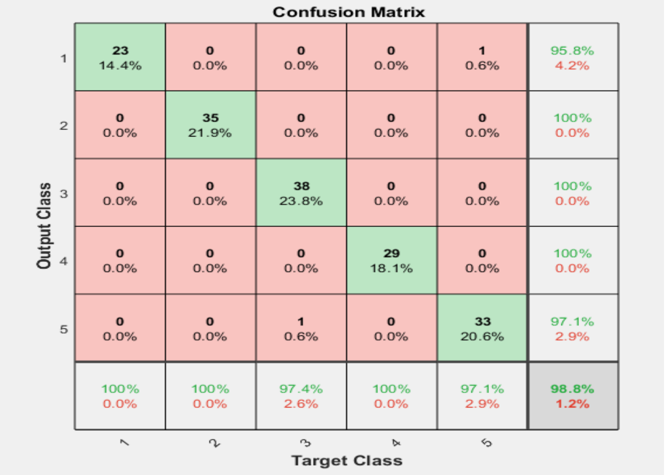


Fig 2: Detection and Recognition of sign with boundary

A confusion matrix is a specific table layout that gives information about the performance of an algorithm. Each row of the confusion matrix consist of the instances in a predicted class where each column represents the instances in an actual class.Here there are five target classes and the possibility getting the accurate class for each input is shown in the table as percent. The total efficiency achieved is 98.8%



Confusion Matrix

Output Class / Target Class

	1	2	3	4	5	
1	23 14.4%	0 0.0%	0 0.0%	0 0.0%	1 0.6%	95.8% 4.2%
2	0 0.0%	35 21.9%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	38 23.8%	0 0.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	29 18.1%	0 0.0%	100% 0.0%
5	0 0.0%	0 0.0%	1 0.6%	0 0.0%	33 20.6%	97.1% 2.9%
	100% 0.0%	100% 0.0%	97.4% 2.6%	100% 0.0%	97.1% 2.9%	98.8% 1.2%

Fig 3: Confusion Matrix

IV. CONCLUSIONS

In this paper, we proposed an efficient traffic sign detection and recognition method where locations of traffic signs are estimated together with their precise boundaries. To this end, we generalized the traffic sign templates with precise boundaries and high accuracy .To achieve practical detection speed,we explored the best-performing convolutional nueral network for both detection and recognition considering the characteristics of traffic signs. By using the templates of traffic signs, our method effectively utilizes strong prior information of target shapes to the drivers. This enables robust boundary estimation for traffic signs that are occluded or blurry and also detects the multiple signs. In addition, by optimizing the resolution of network input for the best trade-off between speed and accuracy, our detector can run with frame rate of 7 FPS on low-power mobile platforms.

The Future direction of our method is that we can adopt the latest architectures such as feature pyramid network and multi-scale training for better speed and accuracy. Finally, the proposed method can be applied not only to traffic sign but also to any other planar objects having standard shapes.

REFERENCES

[1] M. Liang, M. Yuan, X. Hu, J. Li, and H. Liu, “Traffic sign detection by ROI extraction and histogram features-based recognition,” in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Aug. 2013, pp. 1–8.

[2] M. Mathias, R. Timofte, R. Benenson, and L. Van Gool, “Traffic sign recognition—How far are we from the solution?” in *Proc. IEEE Int.Joint Conf. Neural Netw.*, Aug. 2013, pp. 1–8.

[3] P. Dollár, R. Appel, S. Belongie, and P. Perona, “Fast feature pyramids for object detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36,no. 8, pp. 1532–1545, Aug. 2014.

[4] C. Liu, F. Chang, and C. Liu, “Occlusion-robust traffic sign detection via cascaded colour cubic feature,” *IET Intell. Transp. Syst.*, vol. 10,no. 5, pp. 354–360, 2015

[5] A. Møgelmose, D. Liu, and M. M. Trivedi, “Detection of U.S. traffic signs,” *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3116–3125,Dec. 2015.

[6] 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.

[7] J. Uhrig, M. Cordts, U. Franke, and T. Brox. (2016). “Pixel-level encoding and depth layering for instance-level semantic labeling.” [Online]. Available: https://arxiv.org/abs/1604.05096

[8] G. Lin, C. Shen, A. van den Hengel, and I. Reid. (2016). “Exploring context with deep structured models for semantic segmentation.” [Online]. Available: https://arxiv.org/abs/1603.03183

[9] O. Dabeer et al., “An end-to-end system for crowdsourced 3D maps for autonomous vehicles: The mapping component,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Sep. 2017, pp. 634–641.

[10] A. Gudigar, C. Shreesha, U. Raghavendra, and U. R. Acharya, “Multiple thresholding and subspace based approach for detection and recognition of traffic sign,” Multimedia Tools Appl., vol. 76, no. 5, pp. 6937–6991, 2017