Traffic Sign Detection and Recognition under Challenging Conditions

GROUP : 2

Shivam Raval, Om Thakkar, Jay Mohta, Anushree Rankawat, Himol Shah, Faraaz Kakiwala
School of Engineering and Applied Science, Ahmedabad University

***Abstract*—Robust and reliable traffic sign detection is necessary to bring autonomous vehicles onto our roads. Also, simultaneous traffic sign detection and recognition is certainly a very challenging task. Although traffic sign detection and recognition is a fairly archaic field of research, only a few works exist in literature that perform simultaneous detection and recognition using a dataset of realistic real world images. State of the art traffic sign detection algorithms in the literature successfully perform the task over existing databases that mostly lack realistic road conditions. In this paper, the focus has been culminated to detecting such traffic signs under challenging conditions. This paper makes use of a novel dataset that contains a variety of road conditions. The novelty of this paper lies in the fact that instead of focusing on improving the model architecture, we focus on pre-processing the video sequences by cleaning the various types of noises by using Deconvolution and Deblurring techniques, followed by detection and classification of traffic signs in the altered video sequence using Dense Optical Flow and well-known LeNet model architecture. Simulated results demonstrate the superiority of proposed approach.**

***Index Terms*—(Pending)**

I. Introduction

Traffic signs classification is one of the foremost important integral parts of autonomous vehicles and advanced driver assistance systems (ADAS) [1], [2], [3], [4], [5]. Most of the time driver misses traffic signs due to different obstacles and lack of attentiveness. Automating the process of classification of the traffic signs would help reducing accidents. A few challenges in automating the process are signs with the same general meaning, such as the various speed limits, have a common general appearance, leading to subsets of traffic signs that are very similar to each other. Illumination changes, partial occlusions, rotations, and weather conditions further increase the range of variations in visual appearance a classifier has to cope with.

Traditional computer vision and machine learning based methods were widely used for traffic signs classification [6], [7]. but those methods were soon replaced by deep learning based classifiers. Also, two benchmarks are widely used as performance metrics to evaluate the detection performance, namely PASCAL VOC [8] and ImageNet ILSVRC [9]. The limitations of these popular datasets is that they contain images where traffic signs occupies nearly 80% of the image and it becomes comparatively easier to detect and recognize them. However, in realistic scenarios, while trying to detect and recognize from CCTV cameras mounted on the roads, a typical traffic sign might be, for instance, 80 x 80 pixels, in a 2000 x 2000 pixel image, which is less than 0.2% of the image. Hence, it becomes inevitable that such a method that detects and recognizes such small, but significant objects in an image exists in the literature.

A lot of comprehensive research has been conducted since the launch of German traffic-sign detection and classification benchmark data [10, 11]. The research community has actively worked on the problem exploiting both the detection benchmarks (GTSDB) [11] and classification benchmark (GTSRB) [10] task. Current results demonstrate a perfect or nearly perfect performance, with 100% recall and precision for detection and 99.67% precision for classification. Nevertheless, the results have been obtained in a very constrained environment, in spite of it giving the appearance that the problem has been solved efficiently. As mentioned previously, the dataset consists of images where a majority of the image is contained by the traffic sign in the GTSRB classification benchmark. Also, in the GTSDB detection benchmark task, the algorithms must only detect traffic signs in one of 4 major categories. In addition, the algorithm must be so robust that it can easily filter out false positives while retaining true traffic signs.

Hence, in this paper, we make extensive use of datasets like CURE-TSD [cite] and Tencent 100k [cite], where the video sequences have incredibly obscure constraints making it difficult to detect and classify traffic signs. CURE-TSD dataset has an aggregate of 5733 video sequences and Tencent 100k dataset has 1000000 images. The detailed description of the datasets used has been elucidated in the subsequent sections. In addition to this, the novelty of this paper lies in the fact that we focus more on data pre-processing techniques rather than modifying the model architecture. Instead, we use standard model architectures like LeNet which results into a comparatively better accuracy due to efficient data preprocessing, even though it is not that deep and extensive architecture as compared to various state of the art architectures used for traffic sign detection and classification used previously. This approach requires detecting contours and using standard dense optical flow method to track the movement of the object across frames, then the tracked contours are saved and the task of recognition is performed using LeNet architecture.

II. Related Work

It is difficult to compare the published work on traffic sign recognition. Studies are based on different data and either consider the complete task chain of detection, classification and tracking or focus on the classification part only. Some articles concentrate on sub-classes of signs, for example on speed limit signs and digit recognition. In [12], present a holistic system covering all three processing steps. The classifier itself is claimed to operate with a correct classification rate of 94% on images from 23 classes. Training was conducted on 4000 traffic sign images featuring an unbalanced class frequency of 30-600 examples. The individual performance of the classification component is evaluated on a test set of 1700 samples. In [13], present a system for recognition of European and US speed limit signs. Their approach is based on single digit recognition using a neural network. Including detection and tracking, the proposed system obtains a performance of 89% for US and 90% for European speed limits, respectively, on 281 traffic signs. Individual classification results are not provided. Another traffic sign detection framework is presented, the overall system including detection and classification of 48 different signs achieves a performance of 85.3% while obtaining classification error rates below 9%. Also, apply multiple neural networks to classify different traffic signs. In order to choose the appropriate network, shape and color information from the detection stage is used. The authors only provide qualitative classification results. In [14], a classification performance of 95.5% is achieved using support vector machines. The database comprises approximately 36,000 Spanish traffic sign samples of 193 sign classes. However, it is not clear whether the training and test sets can be assumed to be independent, as the random split only took care of maintaining the distribution of traffic sign classes. To our knowledge, this database is not publicly available.

In colour thresholding approach morphological operation like connected component analysis was done for accurate location. Bahlmann et al[12] have used colour, shape, motion information and haar wavelet based features for detection, classification of the traffic sign. By using SVM based colour classification on a block of pixels Le et al [15] addressed the problems of weather variation. German Traffic Sign Recognition Benchmark (GTSRB) is one of the reliable datasets for testing and validating traffic sign classification and detection algorithms. In the competition of GTSRB, top-performing algorithm exceeds best human classification accuracy. By using committee of neural networks Ciresan et al achieved highest ever performance of 99.46%, which surpassed the best human performance of 98.84%. Their proposed committee composed of 25 networks each having 3 convolutional and 2 fully connected networks with traditional data augmentations and jittering. The main disadvantages of this committe are multiples networks, a huge number of parameters (around 90Millions) and dataset dependent handcrafted augmentations. Sermanet et al. proposed multiscale convolutional network [14] with 2 different features stages, which has achieved 98.31% accuracy in this dataset. In our previous work [3] a probabilistic latent semantic analysis based model was proposed, which was built upon traditional handcrafted features extraction methods. Also, other algorithm based on k-d trees and random forest achieved significant accuracy.

III. Datasets Used

A. CURE-TSD

Challenging Unreal and Real Environments for Traffic Sign Detection (CURE-TSD) is a very novel dataset. The video sequences in the dataset have been grouped into two classes: real data and unreal data. Real data correspond to processed versions of sequences acquired from real world. Unreal data corresponds to synthesized sequences generated in a virtual environment. There are 49 real sequences and 49 unreal sequences that do not include any specific challenge. The video sequences have been separated into 70% and 30% splits. Therefore, the dataset has 34 training videos and 15 test videos in both real and unreal sequences that are challenge-free. There are 300 frames in each video sequence. There are 49 challenge-free real video sequences processed with 12 different types of effects and 5 different challenge levels, which result in 2,989 (49*12*5+49) video sequences. Moreover, there are 49 synthesized video sequences processed with 11 different types of effects and 5 different challenge levels, which leads to 2,744 (49*11*5+49) video sequences. In total, there are 5,733 video sequences.

B. GTSRB

German Traffic Sign Recognition Benchmark is a multi-class, single-image classification which has more than 40 classes having more than 50,000 images in total. The dataset has reliable ground-truth data due to semi-automatic annotation. The dataset also ensures that the physical traffic sign instances are unique within the dataset i.e., each real world traffic sign only occurs once. Discussing about the images and their format, every image just contains one traffic sign each. The images contain a border of at least 10% around the actual traffic sign, which accounts for a minimum of 5 pixels. This allows for edge-based approaches for detection and recognition. The image sizes vary from 15*15 to 250*250 pixels but the images are not necessarily square because it ensures there is no bias to the image size or resolution. The annotations are provided in a separate CSV file and include the height and width of the bounding box, along with the the coordinates of the top-left and bottom-right corner of traffic sign bounding box.

C. Tencent 100k

Tencent 100k is a large traffic sign benchmark from 100000 Tencent Street View panoramas, going beyond state of the art benchmarks. Tencent Street Views cover about 300 Chinese cities and the road networks linking them. The original panoramas were captured by 6 SLR cameras and then stitched together. Images were captured both from vehicles and shoulder-mounted equipments, at intervals of about 10m. The

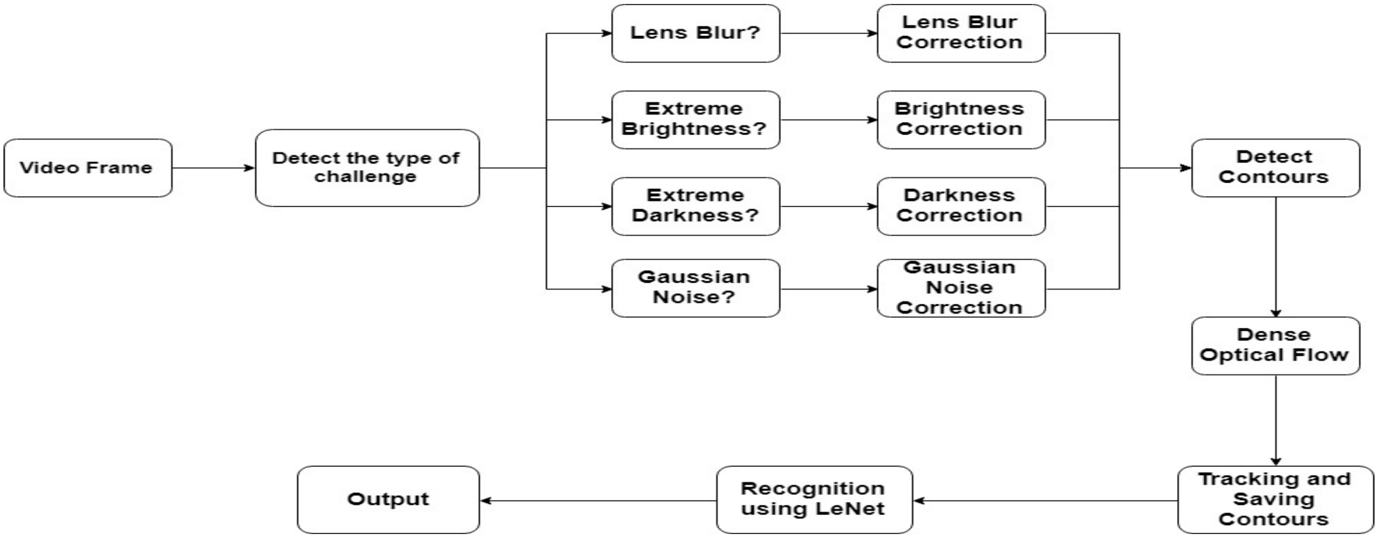

Fig. 1: Block Diagram: Overall Proposed Approach

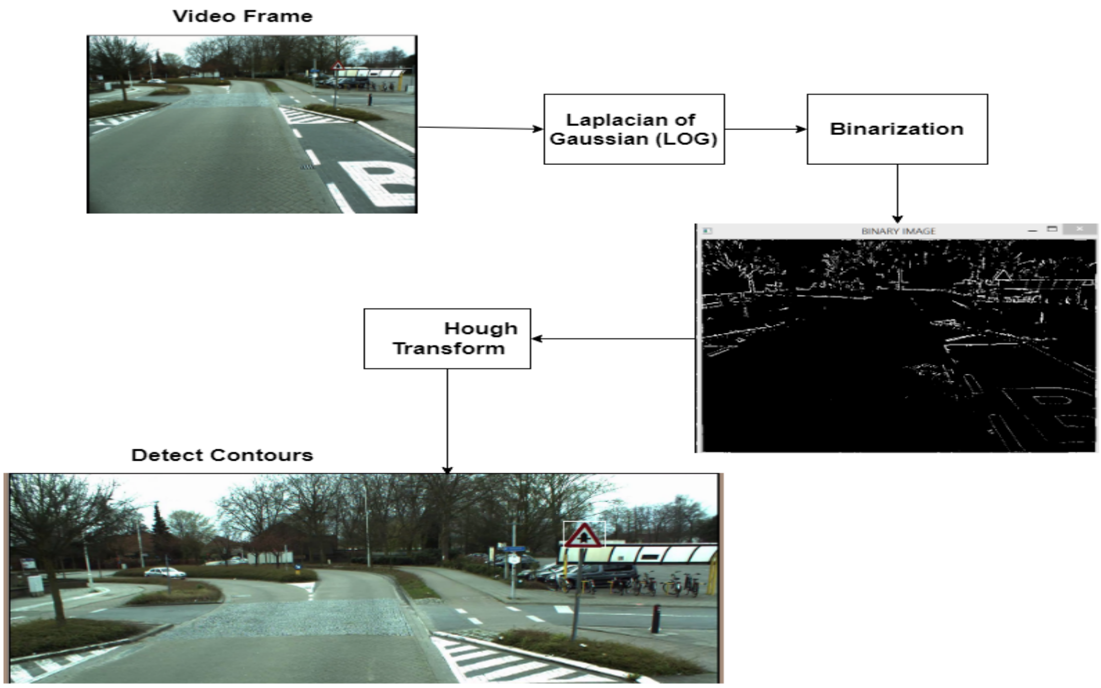

Fig. 2: Block Diagram: Detecting Contours

dataset contains 100000 images containing 30000 traffic sign instances. These images cover large variations in illuminations and weather conditions. Each traffic sign is annotated with a class label, its bounding box and pixel mask. The annotation of data has been done manually. Also, the images in this benchmark have resolution 2048 x 2048.

IV. Proposed Approach

A block diagram of the proposed approach has been given in Fig.1. As shown in the diagram, a few preprocessing steps have been undertaken in order to detect and resolve the commonly encountered issues challenges in traffic sign detection and classification. Initially, the type of challenge among the 4 assumed challenges is detected. The 4 assumed challenges are: Lens Blur, Extreme Brightness, Extreme Darkness and Gaussian Noise. In order to detect these challenging conditions is quite challenging in the first place. We employ the following approach in order to detect the type of challenge:

- **Lens Blur**: We take a single channel of an image and convolve it with a 3 x 3 Laplacian Kernel. Then, we obtain variance of the response. If the variance falls below

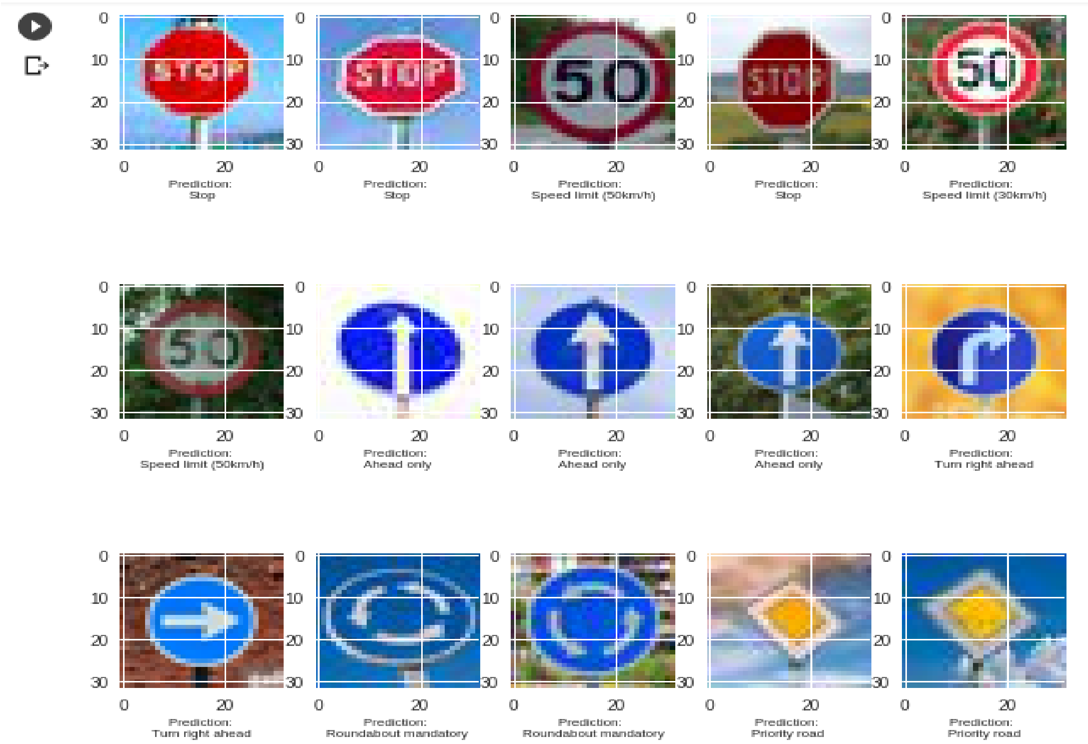

Fig. 3: Validation Results

a pre-defined threshold, then the image is considered blurry; otherwise, the image is not blurry. If variance is low, it implies that there is a tiny spread of responses, indicating there are very little edges in the image and vice versa.

- **Extra Bright/Dark**: To detect whether an image is extra bright or dark we extract the RGB information from the image and averaging the three values gives the measure of brightness or darkness.
- **Gaussian Noise**: To detect whether Gaussian noise is present in an image we look at the histogram of the image, if the noise is present, the histogram follows a curve like Gaussian distribution.

Now, after detecting the type of challenge we detect contours in the video frames. We use the technique of Generalized Hough Transform in order to detect the contours. A block diagram for the same has been given for better understanding. As shown in Fig.3, we apply Laplacian of Gaussian (LoG) on the video frame and then binarize the image in order to detect edges. Then, we use Hough transform in order to detect shapes like circle, rectangle and triangle.

After detecting Hough Transform, we use dense optical flow (based on Gunner Farnebach approach) in order to track the detected contours. Further, we use SVM binary classifier in order to classify whether the detected contour contains a traffic sign. This completes the detection task.

For classification of traffic sign, we use a preexisting model architecture namely LeNet architecture. The specifications of the architecture are as follows:

- Number of Layers: 7
  - Layer 1: Convolutional (30 5x5 filters)
  - Layer 2: Convolutional (200 5x5 filters)
  - Layer 3: Fully connected (2200 depth)
  - Layer 4: Fully connected (1000 depth)
  - Layer 5: Fully connected (500 depth)
  - Layer 6: Fully connected (120 depth)
  - Layer 7: Output Layer (43 Traffic Sign Classes)
- Activation function: ReLU
- Dropout: 0.5 throughout fully connected layers
- L2 regularization of 1e-6 is also applied.

V. Results

Based on the proposed approach elucidated above, we use a set of 15 traffic signs from validation set and test the images. The obtained results have been demonstrated here.

As evident from the results, the proposed approach predicts majority of the traffic signs correctly. The model has been trained on GPU using Google CoLaboratory. The training accuracy obtained is around 99.8% and the testing accuracy turns out to be around 95.8%. These accuracy stats compete with state of the art results without using any extensive or exhaustive resources or deep networks which might consume a lot of resources and computation power.

VI. Observations and Conclusions

In this paper, we have proposed an approach where the use of extensive and exhaustive resources and computation power becomes unnecessary, due to superior data preprocessing techniques used. Also, the superiority of the results can be attributed to the use of novel and wild datasets such as CURE-TSD and Tencent 100k which challenge the state of art GTSRB Benchmark. In this paper, a lot of non-deep learning strategies have been implemented by attempting to outperform the state of art results without using unconventional approach of deep learning techniques. As a part of future work, one can work on incorporating different challenge types and attempt to make the algorithm robust against various difficulty levels.

VII. Acknowledgement

The authors would like to express their gratitude to Dr. Mehul Raval for supporting this work. Also, the authors are also immensely grateful to Mr. Vandit Gajjar for providing extensive help during the course of this work.

References

[1] Haloi, M., Jayagopi, D. B. A robust lane detection and departure warning system. In 2015 IEEE Intelligent Vehicles Symposium (IV) (pp. 126-131). (2015).

[2] McCall, J. C., Trivedi, M. M. Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation. IEEE transactions on intelligent transportation systems, 7(1), 20-37. (2006).

[3] Haloi, M., Jayagopi, D. B. Vehicle Local Position Estimation System. arXiv preprint arXiv:1503.06648. (2015).

[4] Haloi, M., Jayagopi, D. B. Characterizing driving behavior using automatic visual analysis. In Proceedings of the 6th IBM Collaborative Academia Research Exchange Conference (I-CARE) on I-CARE 2014 (pp. 1-4). ACM. 2014.

[5] Braunagel, C., Kasneci, E., Stolzmann, W., Rosenstiel, W. Driver-activity recognition in the context of conditionally autonomous driving. In 2015 IEEE 18th International Conference on Intelligent Transportation Systems (pp. 1652-1657). IEEE. (2015).

[6] Le, T. T., Tran, S. T., Mita, S., Nguyen, T. D. Real time traffic sign detection using color and shape-based features. In Intelligent Information and Database Systems (pp. 268-278). Springer Berlin Heidelberg (2010)

[7] Haloi, M. A novel pLSA based Traffic Signs Classification System. arXiv preprint arXiv:1503.06643 (2015).

[8] M. Everingham, L. Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vision*, 88(2):303-338, June 2010.

[9] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, pages 1-42, April 2015.

[10] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel. The german traffic sign recognition benchmark: a multi-class classification competition. In Neural Networks (IJCNN), The 2011 International Joint Conference on, pages 1453-1460. IEEE, 2011.

[11] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. Neural Networks, 2012.

[12] Bahlmann, C., Zhu, Y., Ramesh, V., Pellkofer, M., Koehler, T. (2005). A system for traffic sign detection, tracking, and recognition using color, shape, and motion information. In Proceedings of the IEEE intelligent vehicles symposium (pp. 255-260). IEEE Press.

[13] Moutarde, F., Bargeton, A., Herbin, A., Chanussot, A. (2007). Robust on-vehicle real-time visual detection of American and European speed limit signs with a modular traffic signs recognition system. In Proceedings of the IEEE intelligent vehicles symposium (pp. 1122-1126). IEEE Press.

[14] Maldonado Bascon, S., Acevedo Rodriguez, J., Lafuente Arroyo, S., Caballero, A., Lopez-Ferreras, F. (2010). An optimization on pictogram identification for the road-sign recognition task using SVMs. Computer Vision and Image Understanding, 114(3), 373-383

[15] Le, T. T., Tran, S. T., Mita, S., Nguyen, T. D. Real time traffic sign detection using color and shape-based features. In Intelligent Information and Database Systems (pp. 268-278). Springer Berlin Heidelberg (2010).

[16] B. Huval, T. Wang, S. Tandon, J. Kiske, W. Song, J. Pazhayampallil, M. Andriluka, P. Rajpurkar, T. Migimat- su, R. Cheng-Yue, F. Mujica, A. Coates, and A. Y. Ng. An empirical evaluation of deep learning on highway driving. CoRR, abs/1504.01716, 2015.