# Object Visual Detection for Intelligent Vehicles

*Sirin Kumar Singh*[#1], *Sunil Vishwakarma*[*2]

[#1]*Department of Computer Science & Engineering, School of Engineering, Babu Banarasi Das University, Uttar Pradesh, India*

[*2]*Department of Computer Science & Engineering, School of Engineering, Babu Banarasi Das University, Uttar Pradesh, India*

***Abstract*— Object visual detection (OVD) aims to extract accurate real-time on-road environment information, which involves three phases detection of objects of interest, recognition of detected objects, and tracking of objects in motion. Here OpenCV tool provide the algorithm support for different object detection. Since recognition and tracking often rely on the results from detection, the ability to detect objects of interest effectively plays a crucial role in OVD. In this paper, we focus on three important classes of objects: traffic signs, cars, and cyclists. We propose to detect *all the three* important objects in a single learning-based detection framework (SLDF). The proposed framework consists of a dense feature extractor and detectors of three important classes. Once the dense features have been extracted, these features are shared with all detectors. The advantage of using one common framework is that the detection speed is much faster, since all dense features need only to be evaluated once in the testing phase. In contrast, most previous works have designed specific detectors using different features for each of these three classes. To enhance the feature robustness to noises and image deformations, we introduce spatially pooled features as a part of aggregated channel features. In order to further improve the generalization performance, we propose an object sub categorization method as a means of capturing the intra class variation of objects. We experimentally demonstrate the effectiveness and efficiency of the proposed framework in three detection applications: traffic sign detection, car detection, and cyclist detection. The proposed framework achieves the competitive performance with state-of-the-art approaches on several benchmark data sets.**

***Keywords*— Object visual detection (OVD), single learning-based detection framework (SLDF), traffic signs, OpenCV, cyclists.**

## I. INTRODUCTION

Object visual detection (OVD) is one of many fast-emerging areas in the intelligent transportation system. This field of research has been actively studied over the past decade. TSP involves three phases: detection, recognition and tracking of various objects of interest. Since recognition and tracking often rely on the results from detection, the ability to detect objects of interest effectively plays a crucial role in TSP. In this thesis, we focus on three important classes of objects: traffic signs, cars, and cyclists. a typical on-road traffic scene with the detected objects of interest and illustrates some positive examples from the three mentioned classes. Most previous methods have designed specific detectors using different features for each of these three classes. The approach we claim here differs from these existing approaches in that we propose a single learning based detection framework to detect all the three important classes of objects.

The proposed framework consists of a dense feature extractor and detectors of these three classes. Once the dense features have been extracted, these features are shared with all detectors. The advantage of using one common framework is that the detection speed is much faster, since all dense features need only to be evaluated once in the testing phase. Because of higher recognition exactness of optical stream technique, movement boundaries of moving articles are created which brings about abstaining from any covering of various moving items. The proposed calculation at first takes the video outlines as info individually gauges the normal stream vectors from them which brings about Optical stream vectors. Clamor sifting is done to eliminate the undesirable movement out of sight. At that point thresholding is done to accomplish double picture. There are some lopsided limits in edge picture which are corrected by morphological tasks. Associated parts are investigated to equitably fix the created white masses in paired picture. At long last, checking of moving item is finished with a case which demonstrates the movement of the articles exclusively. Optical stream strategy has been favored in light of its low intricacy and high precision [6].

For the most part, Object identification has applications in numerous regions of PC vision, including picture getting and video surveillance[1]. Well-informed spaces of article discovery incorporate face identification and passerby location. Great item identification framework decided the presence or nonappearance of articles in self-assertive scenes and be invariant to protest scaling and revolution, the camera see point and changes climate. Address discovery issue with various goals, which are characterized into two classifications: explicit and calculated. The previous includes discovery of known articles and letter includes the recognition of an item class or intrigued region. All article location frameworks use models either expressly or certainly and designate component indicators dependent on these item models. The theory arrangement and check segments fluctuate in their significance in various ways to deal with object identification. A few frameworks utilize just theory development and afterward select the article with most elevated coordinating as the right item. An article recognition

framework must choose right apparatuses and proper strategies for the preparing. In the choice of fitting techniques for a specific application must been considered by numerous variables. An article discovery framework discovers objects in reality from a picture of the world, utilizing object models which are known from the earlier. This cycle is shockingly intense. Since object detection (OD) [43][49] was given a role as an AI issue, the original OD techniques depended available created highlights and direct, max-edge classifiers. The best and agent technique in this age was the Deformable Parts Model (DPM) [13]. After the amazingly powerful work by Krizhevsky et al. in 2012 [14], profound learning (or profound neural organizations) has begun to overwhelm different issues in PC vision and OD was no exemption. The current age OD strategies are completely founded on profound realizing where both the hand-made highlights and direct classifiers of the original techniques have been supplanted by profound neural organizations.

In this paper section I contains the introduction, section II contains the literature review details, section III contains the details about feature extraction, section IV contains the classification details, section V shows architecture details, VI describe the result and section VII provide conclusion of this paper.

## II. LITERATURE REVIEW

Pictures are the blend of pixels which are spread around on the window in an ordinary example and that each point in a pixel has a power esteem that contains a picture. Individuals can watch the picture by numerous qualities of it for distinguishing the article in picture. For machine, a picture is a two dimensional cluster of pixel powers. So methods are formulated to accomplish this objective of item identification. Numerous quantities of procedures has been proposed for object discovery in writing. Numerous investigates examine the issue of item discovery explicitly human location and its use for function arrangement and different undertakings. Here, study is limited to idea of identifying objects those are moving regarding the foundation.

There were numerous calculations proposed for the above errands which are recorded underneath:

- Frame differencing approach
- Viola Jones calculation
- Skin shading demonstrating

In a picture a particular limit that isolates two homogenous districts is taken as an edge. Edge differencing [7] and Edge Detection [49] calculation [8] deducts the two successive casings dependent on these edges. In the event that the distinction comes out to be non-zero qualities, it is viewed as moving. Yet, it has a few constraints that during catching the video because of the development in air or some other source may cause the unsettling influence in the situation of the camera coming about into the bogus location of the immobile articles [7]. The Viola-Jones calculation [9] utilizes Haar-like highlights that are scalar item between the picture and some Haar-like formats. In spite of the fact that it very well may be prepared to recognize an assortment of item classes, it was spurred fundamentally by the issue of face location [10]. Be that as it may, it has a few constraints like the locator is best just on frontal pictures of countenances and it is delicate to lighting conditions. The primer strides in skin identification [11] are the portrayal of picture pixels in shading spaces, appropriate conveyance of skin and non-skin pixels, and after that skin tone [10] displaying. As per skin colors circulation attributes on shading space, skin shading pixels can be identified rapidly with skin shading model. In any case, it has evident detriment like skin tone additionally changes starting with one individual then onto the next having a place with various ethnic gatherings and from people across various regions.

de la Escalera et. al. 2003, This paper manages object acknowledgment in outside conditions. In this kind of conditions, lighting conditions can't be controlled and anticipated, articles can be somewhat impeded, and their position and direction isn't known from the earlier. The picked sort of items is traffic or street signs, because of their helpfulness for sign upkeep, stock in roadways and urban communities, Driver Support Systems and Intelligent Autonomous Vehicles. A hereditary calculation is utilized for the discovery step, permitting an invariance localisation to changes in position, scale, revolution, climate conditions, fractional impediment, and the presence of different objects of a similar tone. A neural organization accomplishes the order. The worldwide framework perceives the traffic sign as well as gives data about its condition or state.

Alberto Broggi, et. al., 2008, [20] Autonomous driving in complex metropolitan conditions, including traffic combine, four-ways quit, overwhelming, and so forth, requires an exceptionally wide reach sensorial capacities, both in point and separation. This review paper presents a dream framework, intended to help converging into traffic on two-ways crossing points, and ready to give a long location separation (over 100m) for approaching vehicles. The framework is made of two high goal wide point

cameras, every one looking horizontally (70 degrees) with deference of the moving course, playing out a particular foundation deduction based method, alongside following and speed assessment. The framework works when the vehicle is halted at convergences, and is set off by the elevated level vehicle director. The framework has been created and tried on the Oshkosh Team's vehicle TerraMaxTM , one of the 11 robots admitted to the DARPA Urban Challenge 2007 Final Event.

Vamsi K. Vegamoor et. al. 2019, [29] This paper shows significant interest as of late in the advancement of associated and independent vehicles (CAVs). Programmed vehicle following ability is key for CAVs; in this article, we give an audit of the basic issues in the longitudinal control plan for programmed vehicle following frameworks (AVFS) utilized by CAVs. This explanatory audit varies from others in giving a survey of fundamental philosophies for plan of AVFS and the effect of AVFS on traffc portability and wellbeing.

Anjan Gudigar, et. al., 2016, [28] Obviously, Intelligent Transport System (ITS) has advanced colossally the entirety of its way. The center of ITS are identification and acknowledgment of traffic sign, which are assigned to satisfy wellbeing and solace needs of driver. This paper gives a basic survey on three significant strides in Automatic Traffic Sign Detection and Recognition(ATSDR) framework i.e., division, identification and acknowledgment with regards to vision based driver help framework. Likewise, it centers around various exploratory arrangements of picture obtaining framework. Further, conversation on conceivable future exploration challenges is made to make ATSDR more proficient, which inturn produce a wide scope of chances for the scientists to do the point by point investigation of ATSDR and to join the future angles in their examination.

Ichikawa, et. Al., 2018,[30] A programmed driving framework incorporates an electronic control gadget arranged to : recognize a driving activity input sum during a programmed driving control for a vehicle ; decide if the driver can begin manual driving during the programmed driving control for the vehicle ; yield a sign for performing changing from programmed heading to the manual driving dependent on a consequence of a correlation between the driving activity input sum and a driving exchanging edge that is a limit for the changing from the programmed heading to the manual driving ; set the driving changing edge to a first driving exchanging edge when it is resolved that the driver can begin the manual driving ; and set the driving changing edge to a subsequent driving exchanging edge surpassing the first driv ing exchanging edge when it is resolved that the driver can't begin the manual driving.

Adam Coates, et. al.,2011, [22] While vector quantization (VQ) has been applied generally to create highlights for visual acknowledgment issues, much late work has zeroed in on more impressive techniques. Specifically, scanty coding has developed as a solid option in contrast to customary VQ approaches and has been appeared to accomplish reliably better on benchmark datasets. The two methodologies can be part into a preparation stage, where the framework learns a word reference of premise capacities, and an encoding stage, where the word reference is utilized to separate highlights from new sources of info. In this work, we examine the purposes behind the accomplishment of inadequate coding over VQ by decoupling these stages, permitting us to isolate out the commitments of preparing and encoding in a controlled manner. Through broad trials on CIFAR, NORB and Caltech 101 datasets, we think about a few preparing and encoding plans, including meager coding and a type of VQ with a delicate edge actuation work. Our outcomes show not just that we can utilize quick VQ calculations for preparing, yet that we can similarly too utilize haphazardly picked models from the preparation set. As opposed to spend assets on preparing, we discover it is more essential to pick a decent encoder—which can frequently be a basic feed forward non-linearity. Our outcomes remember best in class execution for both CIFAR and NORB.

Arturo de la Escalera, et. al., 1997, [23] A dream based vehicle direction framework for street vehicles can have three fundamental jobs: 1) street location; 2) hindrance discovery; and 3) sign acknowledgment. The initial two have been read for a long time and with numerous great outcomes, however traffic sign acknowledgment is a less-examined field. Traffic signs furnish drivers with truly significant data about the street, so as to make driving more secure and simpler. We feel that traffic signs must assume similar part for self-ruling vehicles. They are intended to be effectively perceived by human drivers mostly in light of the fact that their shading and shapes are altogether different from indigenous habitats. The calculation portrayed in this paper exploits these highlights. It has two fundamental parts. The first, for the discovery, utilizes shading thresholding to portion the picture and shape examination to recognize the signs. The subsequent one, for the grouping, utilizes a neural organization. A few outcomes from normal scenes are appeared. Then again, the calculation is legitimate to distinguish different sorts of imprints that would advise the versatile robot to play out some errand at that place.

Shivani Agarwal, et. Al., 2002,[24] We present a methodology for figuring out how to distinguish objects in still dark pictures, that depends on a scanty, part-based portrayal of articles. Avocabulary of data rich item parts is consequently built from a bunch of test pictures of the article class of revenue. Pictures are then spoken to utilizing parts from this jargon, alongside spatial relations saw among them. In view of this portrayal, an element productive learning calculation is utilized to figure out how to distinguish occasions of the article class. The structure created can be applied to any object with recognizable parts in a

generally fixed spatial design. We report investigates pictures of side perspectives on vehicles. Our examinations show that the technique accomplishes high identification exactness on a troublesome test set of true pictures, and is profoundly hearty to incomplete impediment and foundation variety. Likewise, we examine and offer answers for a few methodological issues that are huge for the examination network to have the option to assess object location approaches.

Timo Ahonen, et.al., 2004, [25] In this work, we present a novel way to deal with face acknowledgment which considers both shape and surface data to speak to confront pictures. The face territory is initial separated into little areas from which Local Binary Pattern (LBP) histograms are removed and connected into a solitary, spatially upgraded include histogram proficiently speaking to the face picture. The acknowledgment is performed utilizing a closest neighbor classifier in the processed component space with Chi square as a disparity measure. Broad investigations obviously show the predominance of the proposed plot over completely thought about strategies (PCA, Bayesian Intra/extrapersonal Classifier and Elastic Bunch Graph Matching) on FERET tests which incorporate testing the vigor of the strategy against various outward appearances, lighting and maturing of the subjects. Notwithstanding its proficiency, the effortlessness of the proposed strategy takes into account quick element extraction.

Santosh K. Divvala et.al., 2012, [26] The Deformable Parts Model (DPM) has as of late developed as an extremely valuable and well-known apparatus for handling the intra-classification variety issue in object identification. In this paper, we sum up the vital experiences from our exact investigation of the significant components comprising this identifier. All the more explicitly, we study the connection between the function of deformable parts and the combination model segments inside this indicator, and comprehend their relative significance. To start with, we find that by expanding the quantity of parts, and exchanging the instatement venture from their perspective proportion, left-right flipping heuristics to appearance based bunching, extensive improvement in execution is acquired. In any case, more intriguingly, we saw that with these new segments, the part misshapenings would now be able to be killed, yet getting outcomes that are nearly comparable to the first DPM indicator.

Navneet Dalal, et. al., 2005,[27] We study the subject of capabilities for hearty visual item acknowledgment, receiving straight SVM based human identification as an experiment. In the wake of looking into existing edge and inclination based descriptors, we show tentatively that lattices of Histograms of Oriented Gradient (HOG) descriptors fundamentally beat existing capabilities for human identification. We study the impact of each phase of the calculation on execution, presuming that one-scale inclinations, one direction binning, generally coarse spatial binning, and top notch neighborhood contrast standardization in covering descriptor blocks are exceptionally significant for good outcomes. The new methodology gives close ideal division on the first MIT person on foot information base, so we present an additionally testing dataset containing more than 1800 commented on human pictures with a huge scope of posture varieties and foundations.

## III. METHODOLOGY

Most previous methods have designed specific detectors using different features for each of these three classes. The approach we claim here differs from these existing approaches in that we propose a single learning based detection framework to detect all the three important classes of objects. In order to further improve the generalization performance, we propose an object sub categorization method as a means of capturing the intra-class variation of objects.

### A. Generic Object Detection

Object detection is a challenging but important application in the computer vision community. It has achieved successful outcomes in many practical applications such as face detection and pedestrian detection. Complete survey of object detection can be found in. This section briefly reviews several generic object detection methods. These frameworks achieve excellent detection results on rigid object classes. However, for object classes with a large intra-class variation, their detection performance falls down dramatically. Recently, a new detection framework which uses aggregated channel features (ACF) and an AdaBoost classifier has been proposed in. This framework uses exhaustive sliding-window search to detect objects at multi-scales. It has been adapted successfully for many practical applications.

### B. Traffic Sign Detection

Many traffic sign detectors have been proposed over the last decade with newly created challenging benchmarks. Interested reader should see which provides a detailed analysis on the recent progress in the field of traffic sign detection. Most existing traffic sign detectors are appearance-based detectors. These detectors generally fall into one of four categories, namely, color-based approaches, shape-based approaches, texture-based approaches, and hybrid approaches. One standard benchmark for traffic sign detection is the German traffic sign detection benchmark (GTSDB) which collects three important categories of road signs (prohibitory, danger, and mandatory) from various traffic scenes. All traffic signs have been fully annotated with the rectangular regions of interest (ROIs). Researchers can conveniently compare their work based on this benchmark.

### C. CAR DETECTION

Many existing car detectors are vision based detectors. Interested reader should see which discusses different approaches for vehicle detection using mono, stereo, and other vision-sensors. We focus on vision-based car detectors using monocular information in this paper. These detectors can be divided into three categories: DPM-based approaches, sub categorization-based approaches and motion based approaches.

### D. CYCLIST DETECTION

Many existing cyclist detectors use pedestrian detection techniques since appearances of pedestrians are very similar to appearances of cyclists along the road. These detectors are mainly derived from the fixed camera-based approaches. Fixed camera-based approaches are designed for traffic monitoring using fixed cameras corner feature extraction, motion matching, and object classification are combined to detect pedestrians and cyclists simultaneously. In a stereo vision based approach is proposed for pedestrian and cyclist detection. It uses the shape features and matching criterion of partial Hausdorff distance to detect targets. The authors of propose a cyclist detector to detect two wheels of bicycles on road, but this approach is limited to detect crossing cyclists.

### E. PROPOSED SOLUTION

We propose a single learning based detection framework (SLDF) to detect all the three important classes of objects. The proposed framework consists of a dense feature extractor and detectors of these three classes. Once the dense features have been extracted, these features are shared with all detectors. The advantage of using one common framework is that the detection speed is much faster, since all dense features need only to be evaluated once in the testing phase. The proposed framework introduces spatially pooled features as a part of aggregated channel features to enhance the feature robustness to noises and image deformations. In order to further improve the generalization performance, we propose an object sub categorization method as a means of capturing the intra-class variation of objects.

### F. Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) is one of the variants of neural networks used heavily in the field of Computer Vision. It derives its name from the type of hidden layers it consists of. The hidden layers of a CNN typically consist of convolutional layers, pooling layers, fully connected layers, and normalization layers. Here it simply means that instead of using the normal activation functions defined above, convolution and pooling functions are used as activation functions. To understand it in detail one needs to understand what convolution and pooling are. Both of these concepts are borrowed from the field of Computer Vision. Step used in CNN algorithm is:

- Step 1: Convolution Operation. ...
- Step 1(b): ReLU Layer. ...
- Step 2: Pooling. ...
- Step 3: Flattening. ...
- Step 4: Full Connection. ...
- Step 1 - Convolution Operation. ...
- Step 1(b): The Rectified Linear Unit (ReLU) ...
- Step 2 - Max Pooling.

### G. Region-based Convolutional Neural Networks(R-CNN)

R-CNN is a state-of-the-art visual object detection system that combines bottom-up region proposals with rich features computed by a convolutional neural network. At the time of its release, R-CNN improved the previous best detection performance on PASCAL VOC 2012 by 30% relative, going from 40.9% to 53.3% mean average precision. Unlike the previous best results, R-CNN achieves this performance without using contextual rescoring or an ensemble of feature types. To bypass the problem of selecting a huge number of regions, Ross Girshick et al. proposed a method where we use selective search to extract just 2000 regions from the image and he called them region proposals. Therefore, now, instead of trying to classify a huge number of regions, you can just work with 2000 regions.

R-CNN algorithms have truly been a game-changer for object detection tasks. There has suddenly been a spike in recent years in the amount of computer vision applications being created, and R-CNN is at the heart of most of them.

## IV. ARCHITECTURE

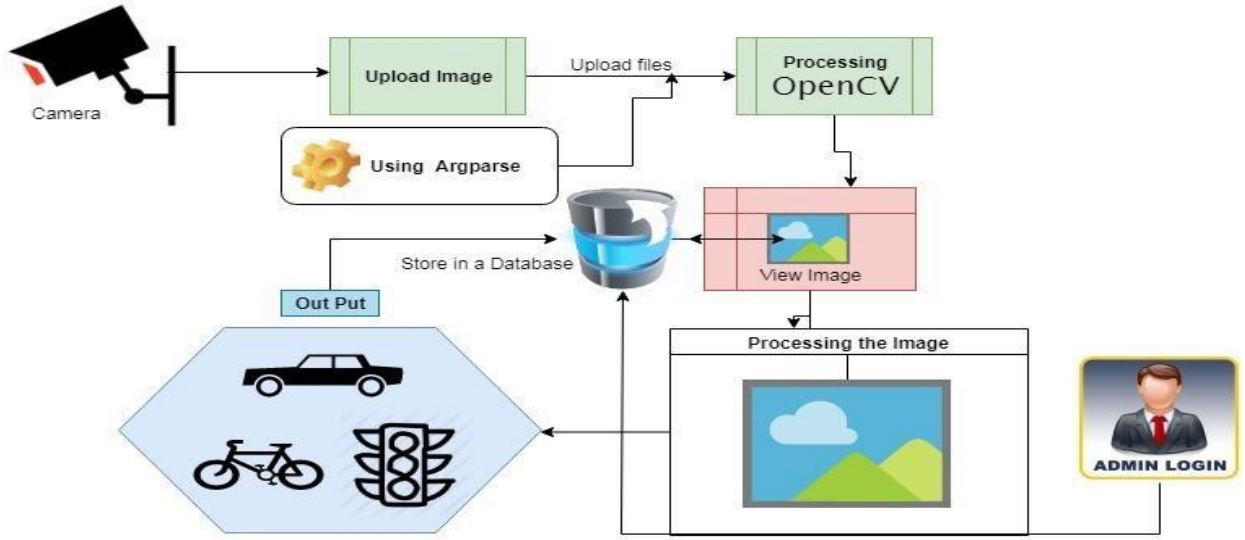


Figure 1: Architecture diagram

## V. RESULT

Object detection in computer vision. Object detection is the process of finding instances of real-world objects such as Car, bicycles, and Traffic sign in images or videos. Object detection algorithms typically use extracted features and learning algorithms to recognize instances of an object category. Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, cars, bicycles , Traffic sign) in digital images and videos.

Table 2: Time consumed by the algorithm for detecting object in images

| sn | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| result | yes | yes | yes | yes | yes |
| Time/sec | 6.1884 | 5.3134 | 5.7031 | 5.1045 | 5.8712 |
| Average/sec | 5.6361 | | | | |

Table 3: Time consumed by the algorithm for detecting object in videos

| sn | 1 | 2 | 3 |
|---|---|---|---|
| Number of frames | 706 | 812 | 950 |
| Single frame/time/ms | 6.2012 | 5.4219 | 5.1362 |
| total time | 4378.0273 | 4402.5625 | 4616.7293 |
| Average time | 4465.7730 | | |


Figure 2 : Bus and Person detection


Figure 3: Traffic Signal detection

Figure 4: Cycle detection

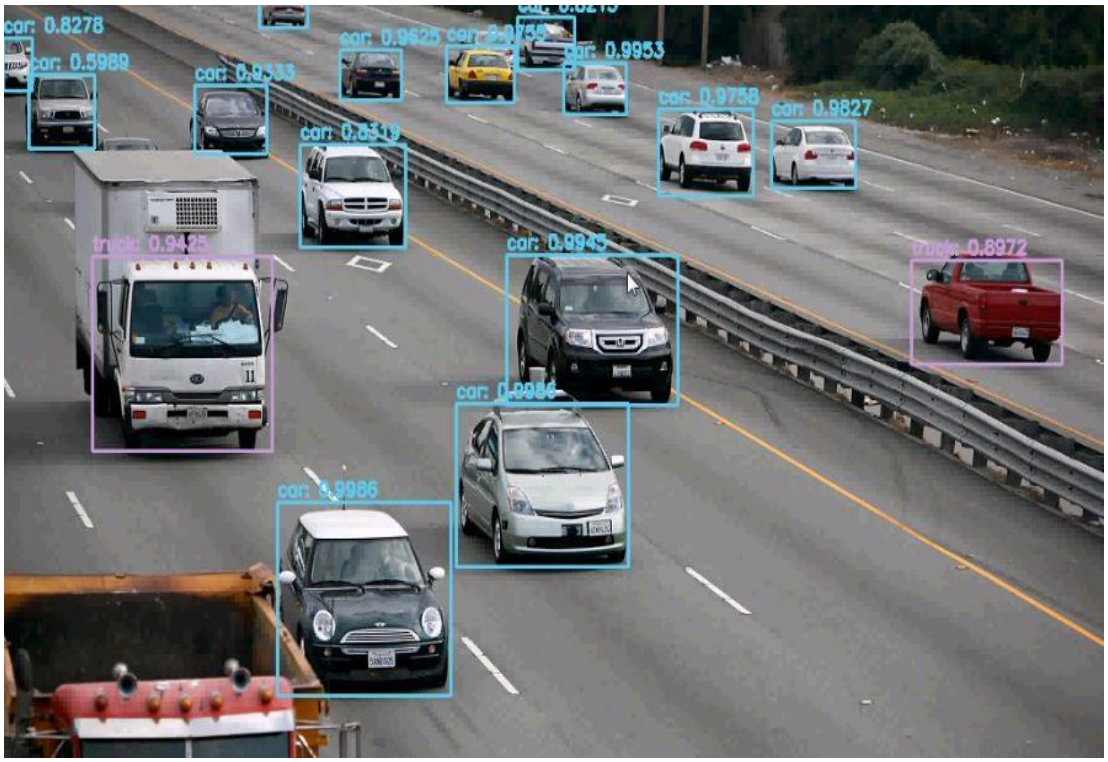


Figure 5: Car detection

## VI. CONCLUSION & FUTURE SCOPE

This paper includes a common detection framework for detecting three important classes of objects in traffic scenes. The proposed framework introduces spatially pooled features as a part of aggregated channel features to enhance the feature robustness and employs detectors of three important classes to detect multiple objects. The detection speed of the framework is fast since dense features need only to be evaluated once rather than individually for each detector. To remedy the weakness of the VJ framework for object classes with a large intra-class variation, we propose an object sub categorization method to improve the generalization performance by capturing the variation. We demonstrated that our detector achieves the competitive results with state-of-the-art detectors in traffic traffic sign detection, car detection, and cyclist detection. Future work could include that contextual information can be used to facilitate object detection in traffic scenes and convolutional neural network can be used to generate more discriminative feature representations.

Future work could include that contextual information can be used to facilitate object detection in traffic scenes and convolutional neural network can be used to generate more discriminative feature representations. We proposed a method for shape-based object detection using distance transforms which takes combined courses to fine approach in shape and parameter space as well. It works in real time environment with multiple detection objects in a single framework method.

## REFERENCES

[1] P. F. Felzenszwalb, R. B. Girshick, D. Mcallester, and D. Ramanan, “Object detection with discriminatively trained part-based models,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, p. 1627, 2010.
[2] K. K. Sung and T. Poggio, “Example-based learning for view-based human face detection,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 1, pp. 39–51, 2002.

[3] C. Wojek, P. Dollar, B. Schiele, and P. Perona, “Pedestrian detection: An evaluation of the state of the art,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 4, p. 743, 2012.

[4] H. Kobatake and Y. Yoshinaga, “Detection of spicules on mammogram based on skeleton analysis.” IEEE Trans. Med. Imag., vol. 15, no. 3, pp. 235–245, 1996.

[5] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” in ACM MM, 2014.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in NIPS, 2012.

[7] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, “Realtime multi-person 2d pose estimation using part affinity fields,” in CVPR, 2017.

[8] Z. Yang and R. Nevatia, “A multi-scale cascade fully convolutional network face detector,” in ICPR, 2016.

[9] C. Chen, A. Seff, A. L. Kornhauser, and J. Xiao, “Deepdriving: Learning affordance for direct perception in autonomous driving,” in ICCV, 2015.

[10] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, “Multi-view 3d object detection network for autonomous driving,” in CVPR, 2017.

[11] A. Dundar, J. Jin, B. Martini, and E. Culurciello, “Embedded streaming deep neural networks accelerator with applications,” IEEE Trans. Neural Netw. & Learning Syst., vol. 28, no. 7, pp. 1572–1583, 2017.

[12] R. J. Cintra, S. Duffner, C. Garcia, and A. Leite, “Low-complexity approximate convolutional neural networks,” IEEE Trans. Neural Netw. & Learning Syst., vol. PP, no. 99, pp. 1–12, 2018.

[13] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri, “Cost-sensitive learning of deep feature representations from imbalanced data.” IEEE Trans. Neural Netw. & Learning Syst., vol. PP, no. 99, pp. 1–15, 2017.

[14] A. Stuhlsatz, J. Lippel, and T. Zielke, "Feature extraction with deep neural networks by a generalized discriminant analysis." IEEE Trans. Neural Netw. & Learning Syst., vol. 23, no. 4, pp. 596–608, 2012.

[15] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in CVPR, 2014.

[16] R. Girshick, "Fast r-cnn," in ICCV, 2015.

[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in CVPR, 2016.

[18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards realtime object detection with region proposal networks," in NIPS, 2015, pp. 91–99.

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. of Comput. Vision, vol. 60, no. 2, pp. 91–110, 2004.

[20] Alberto Broggi, Andrea Cappalunga, Stefano Cattani and Paolo Zani, 2008, "Lateral Vehicles Detection Using Monocular High Resolution Cameras on TerraMax", IEEE Intelligent Vehicles Symposium Eindhoven University of Technology Eindhoven, The Netherlands, June 4-6, 2008.

[21] A. de la Escalera, J.Ma Armingol, M. Mata, 2003, "Traffic sign recognition and analysis for intelligent vehicles" Image and Vision Computing 21 (2003) 247–258.

[22] Adam Coates, Andrew Y. Ng, 2011, "The Importance of Encoding Versus Training with Sparse Coding and Vector Quantization" Appearing in Proceedings of the 28 th International Conference on Machine Learning, Bellevue, WA, USA, 2011.

[23] Arturo de la Escalera, Luis E. Moreno, Miguel Angel Salichs, and Jos´e Mar´ia Armingol, 1997," Road Traffic Sign Detection and Classification", IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, VOL 44, NO 6, DECEMBER 1997.

[24] Shivani Agarwal and Dan Roth, 2002, "Learning a Sparse Representation for Object Detection", A. Heyden et al. (Eds.): ECCV 2002, LNCS 2353, pp. 113–127, 2002. Springer-Verlag Berlin Heidelberg 2002.

[25] Timo Ahonen, Abdenour Hadid, and Matti Pietik¨ainen, 2004, "Face Recognition with Local Binary Patterns", ECCV 2004, LNCS 3021, pp. 469–481, 2004. Springer-Verlag Berlin Heidelberg 2004.

[26] Santosh K. Divvala, Alexei A. Efros, and Martial Hebert, 2012, "How Important Are "Deformable Parts" in the Deformable Parts Model?", ECCV 2012 Ws/Demos, Part III, LNCS 7585, pp. 31–40, 2012. Springer-Verlag Berlin Heidelberg 2012.

[27] Navneet Dalal, Bill Triggs, 2005, "Histograms of Oriented Gradients for Human Detection", International Conference on Computer Vision & Pattern Recognition (CVPR '05), Jun 2005, San Diego, United States. pp.886–893, 10.1109/CVPR.2005.177. inria-00548512

[28] Anjan Gudigar, Shreesha Chokkadi & Raghavendra U, 2016, "A review on automatic detection and recognition of traffic sign", Multimedia Tools and Applications volume 75, pages333–364(2016)

[29] Vamsi K. Vegamoor, Swaroop Darbha* and Kumbakonam R. Rajagopal, 2019, "A Review of Automatic Vehicle Following Systems", . Indian Inst. Sci.|VOL 99:4|567–587 December 2019|journal.iisc.ernet.in.

[30] Ichikawa, 2018, "Automatic Driving System", Sep . 4 , 2018, US 10 , 067 , 505 B2.

[31] Y. Aoyagi, T. Asakura, A study on traffic sign recognition in scene image using genetic algorithms and neural networks, 22nd International Conference on Industrial Electronics, Control, and Instrumentation, IEEE August (1996).

[32] G. Adorni, V. Da´ndrea, G. Destri, M. Mordoni, Shape searching in real world images: a CNN-based approach, Fourth Workshop on Cellular Neural Networks and their Applications, IEEE June (1996).

[33] G. Adorni, M. Mordonini, A. Poggi, Autonomous agents coordination through traffic signals and rules, Conference on Intelligent Transportation Systems, IEEE November (1997).

[34] P. Arnoul, M. Viala, J.P. Guerin, M. Mergy, Traffic signs localization for highways inventory from a video camera on board a moving collection van, Intelligent Vehicles Symposium, IEEE September (1996).

[35] H. Austerirmeier, U. Bu¨ker, B. Merstching, S. Zimmermann, Analysis of traffic scenes using the hierarchical structure code, International Workshop on Structural and Syntactic Pattern Recognition August (1992).

[36] Md Amirul Islam, Mrigank Rochan, Neil DB Bruce, and Yang Wang. Gated feedback refinement network for dense image labeling. CVPR, pages 3751–3759, 2017.

[37] Zhaowei Cai, Quanfu Fan, Rogerio S Feris, and Nuno Vasconcelos. A unified multi-scale deep convolutional neural network for fast object detection. ECCV, pages 354–370, 2016.

[38] Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: Delving into high quality object detection. CVPR, pages 6154–6162, 2018.

[39] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. ECCV, 2018.

[40] Franc¸ois Chollet. Xception: Deep learning with depthwise separable convolutions. CVPR, pages 1610–02357, 2017.

[41] Stefan Elfwing, Eiji Uchibe, and Kenji Doya. Sigmoidweighted linear units for neural network function approximation in reinforcement learning. Neural Networks, 107:3–11, 2018.

[42] Mark Everingham, S. M. Ali Eslami, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. International Journal of Computer Vision, 2015.

[43] Golnaz Ghiasi, Tsung-Yi Lin, Ruoming Pang, and Quoc V. Le. Nas-fpn: Learning scalable feature pyramid architecture for object detection. CVPR, 2019.

[44] Ross Girshick. Fast r-cnn. ICCV, 2015.

[45] Kaiming He, Ross Girshick, and Piotr Dollar. Rethinking ´ imagenet pre-training. ICCV, 2019.

[46] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Gir- ´ shick. Mask r-cnn. ICCV, pages 2980–2988, 2017.

[47] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. CVPR, pages 770–778, 2016.

[48] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. Searching for mobilenetv3. ICCV, 2019.

[49] Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, et al. Speed/accuracy trade-offs for modern convolutional object detectors. CVPR, 2017.

[50] Seung-Wook Kim, Hyong-Keun Kook, Jee-Young Sun, Mun-Cheon Kang, and Sung-Jea Ko. Parallel feature pyramid network for object detection. ECCV, 2018.

[51] Alexander Kirillov, Ross Girshick, Kaiming He, and Piotr Dollar. Panoptic feature pyramid networks. ´ CVPR, 2019. 6 [17] Tao Kong, Fuchun Sun, Chuanqi Tan, Huaping Liu, and Wenbing Huang. Deep feature pyramid reconfiguration for object detection. ECCV, 2018.