

Traffic Adaptive Deep Learning based Fine Grained Vehicle Categorization in Cluttered Traffic Videos

Shobha B.S, Deepu.R

Department of Computer Science and Engineering
Maharaja Institute of Technology Mysore
Mysore, India

Abstract—Smart traffic management is being proposed for better management of traffic infrastructure and regulate traffic in smart cities. With surge of traffic density in many cities, smart traffic management becomes utmost necessity. Vehicle categorization, traffic density estimation and vehicle tracking are some of the important functionalities in smart traffic management. Vehicles must be categorized based on multiple levels like type, speed, direction of travel and vehicle attributes like color etc. for efficient tracking and traffic density estimation. Vehicle categorization becomes very challenging due to occlusions, cluttered backgrounds and traffic density variations. In this work, a traffic adaptive multi-level vehicle categorization using deep learning is proposed. The solution is designed to solve the problems in vehicle categorization in terms of occlusions, cluttered backgrounds.

Keywords—Vehicle categorization; deep learning; traffic density estimation; clutter

I. INTRODUCTION

Smart traffic management based on video feeds from traffic surveillance cameras is being proposed as a means for efficient traffic regulation at a lower cost compared to sensors based traffic management.

Smart traffic management aims to regulate the traffic conditions in peak hours, manage congestions, transport of emergency vehicles, detect and handle accidents/incidents in the road. Vehicle categorization is an important functionality in smart traffic management. Segmenting the vehicles and categorization of them based on multiple levels like vehicle type, speed, direction of travel, meta attributes like color, make etc. is important for localization and tracking of vehicles in smart traffic management.

Vehicle categorization becomes very challenging in presence of noisy cluttered background, environmental conditions (fog, rain, lighting and haze), shadow and occlusions etc. The problem becomes even more difficult in case of need for fine grained multi-level categorization like learning speed, direction and meta attributes of the vehicles from the video stream. But the applications of this multilevel categorization are innumerable in terms of regulation, description, indexing and tracking.

This work deals with this need and proposes a traffic adaptive deep learning multi-level vehicle categorization which can work in conditions of cluttered background and occlusions in the video. The approach is an integration of two different

mechanisms – one designed for low density traffic and another for high density traffic. A deep learning optimized topological active net segmentation is done to segment the vehicles in case of low density traffic. For high density traffic, convolutional neural network segmentation is done. After segmentation, features are extracted and mapping is done to learn various meta-attributes of the segmented vehicles.

II. LITERATURE SURVEY

Authors in [1] proposed an architecture called TPSVedet for categorization of vehicle to small, medium and large size. From the video, background model is constructed by averaging the frames over time. Background is subtracted from each frames and over the ROI region geometry based features are extracted. PCA is used for dimensionality reduction and the dimension reduced features are classified using machine learning methods like ANN, SVM and AdaBoost. The problem in this approach is that it can detect only one vehicle in the ROI region. Deep convolutional neural network along with joint fine tuning is used for vehicle classification in [2]. Deep residual network (ResNet) model is used for convolutional neural network. Drop out method is used for preventing the over fitting. The network is trained with images and their ground truth vehicle location marked as boxes. The trained model is able to localize the vehicle in the image and classify it one of 11 different types of vehicles. The method cannot work for multiple vehicles in the same image. Authors in [3] reviewed and compared the various methods for feature extraction, global representation and classification for automated vehicle classification. Most of the approaches were found to work for relatively static background and cannot work in presence of occlusions or changing lighting conditions. Geometrical feature based approaches were found to have higher misclassification rate. Texture-based approaches have high sensitivity and computational costs. Authors in [4] proposed a vehicle categorization approach based on geometric feature extraction. Geometric features like shape and size is extracted and passed to trained random forest model to classify the vehicle. This approach assumes fixed vehicle placement and it does not work for multiple vehicles in the image. Authors in [5] experimented with vehicle identification for cars with images taken in different viewpoint of “front (F)”, “rear (R)”, “side (S)”, “front-side (FS)”, “rear side (RS)”, and “All-View”. From their experiment, the convolutional neural network trained with Front side and Rear images were found to have higher accuracy compared to others. Front and Rear image had the information to detect the make for almost all the

cars. Deep learning based vehicle classification is proposed in [6]. A two layer convolution based CNN is trained with car images as input and vehicle type/color as output. The method can classify only a single image and time complexity for classification is also high. Also the approach has only 70% accuracy for color classification. Authors in [7] used texture characteristics in the headlight and grill area to classify the vehicle. Headlight and grill area is segmented from the vehicle falling in the ROI area and the GLCM texture features are extracted from it. Vehicles were then classified based on the similarity of GLCM features. Authors in [8] proposed a vehicle classification system based on side view profile of the vehicles. Side view images of vehicle are skeletonized and features such as joints and endpoints are extracted from it. The features are looked up for similarity against training image features to classify the vehicle. The accuracy of the proposed solution is very sensitive to occlusions. Authors in [9] used YOLOv3 deep learning network for vehicle detection in the images. Road surface area is split to two categories of remote area and proximal area. The vehicle object in the road area is segmented and classified using YOLOv3 to three vehicle category of bus, car and motorcycle. Though the solution can work for multiple vehicles in the image, it does not work for high density traffic. The solution assumes a larger gap between the vehicles. Author in [10] recognized vehicle logo using enhanced scale-invariant feature transform (SIFT)-based feature-matching scheme. Logo is segmented from the vehicle image applying phase congruency calculation. From the logo segment, SIFT features are extracted and matched against trained patterns to recognize the logo. Authors in [11] solved the problem of vehicle color recognition using BoW model. The ROI for identifying the dominant color is implicitly selected in this method. From the ROI region, local color features are extracted and classified using a multi-class SVM to recognize the color. Authors in [12] proposed a robust system for car make recognition from car front images even in presence of low contrast and compression based distortions. Car brand region is segmented and SIFT features are extracted from the brand region. The SIFT features are then matched to training images to recognize the car make. An unsupervised convolutional neural network is used for classification of vehicle type in [13] based on the vehicle frontal images. Sparse Laplacian filter learning is used to capture rich and discriminative information from the vehicle image. The output of the convolutional neural network is the probability of each type the vehicle belongs to. But the method can recognize only one vehicle in the image with no distortion in the frontal view. A simple convolutional neural network model is proposed in [14] for classifying six different vehicle types. Convolutional features are learnt from the low resolution input images. The convolutional features are classified with a fully connected standard network to probability of each class of vehicle. A two stage classification method for vehicle recognition is proposed in [15]. The classification uses both global and local features. An improved canny edge detection with smooth filtering is proposed to extract global features. Local features are extracted using Gabor wavelet. The vehicle is classified to small or large at first stage of classification using the global features and vehicle type is found in second stage classification using local features. Authors in [16] proposed a deformable model integrating both

detection and classification into one stage. A deformable part based model is trained using annotated vehicle image for classifying the vehicle. Vehicles are extracted from the traffic image and model alignment is done on extracted image crop. SVM classifier is trained to classify the model to the vehicle type. Regression analysis was used for vehicle classification in [17]. Foreground segments having vehicles are detected first using a wrapping method. Low level features are extracted from the foreground segments and cascaded regression approach is used to classify the vehicles. A stochastic multi class vehicle classification system based on Bayesian model is proposed in [18]. Low dimensional features of vehicle tail light are classified using a Bayesian network to four different vehicle types. Author in [19] used statistical random pixel distribution features acquired from low dimensional images to recognize the logo of the vehicle. Multiscale scanning algorithm is used to jointly detect and classify logos. Author in [20] used speeded up robust features (SURF) to recognize vehicle. SURF features are extracted from front and rear view of the vehicles. The features are then classified by multi class SVM to the type of the vehicle. Author in [21] have presented a detailed review on vehicle detection, recognition and tracking. Multi view methods for vehicle detection are also discussed.

III. FINE GRAINED TRAFFIC ADAPTIVE VEHICLE CATEGORIZATION

Most of existing solutions are based on drawing the bounding box around the foreground objects and classifying the vehicle type of them. But in case of dense traffic, occlusion makes the drawing of bounding box difficult and the accuracy of vehicle detection difficult. Cluttered background in terms of pedestrian movement, shadows etc. causes error in boundary box localization and due to this vehicle classification becomes erroneous. In this work, the vehicle classification is handled as a three stage process. In first stage, images in the video are pre-processed by removing the background, shadows and illumination artefacts. In the second stage, the image is split to two categories – Type 1 category where bounding box estimation would be easy and a Type 2 category where bounding box estimation is difficult due to clutters in the image. In third stage, for type 1 category images, deep learning convolutional neural network is used to generate bounding box for foreground vehicles and features extracted from the bounding box segment are used for fine grained vehicle classification. For type 2 category images, integration of deep learning with topological active net deformable model is used for efficient segmentation of the vehicles. The clutters are filtered out in this step and the features collected from the mesh are used for fine grained vehicle classification. The architecture of the proposed solution is given in Fig. 1. Each stage of the proposed solution is detailed below.

A. Preprocessing

In this stage, a background model is constructed based on analysis of the frames in the video. The goal of background modeling is to find a best estimate of background so that impact of shadows and sudden illumination changes on the foreground model is minimized.

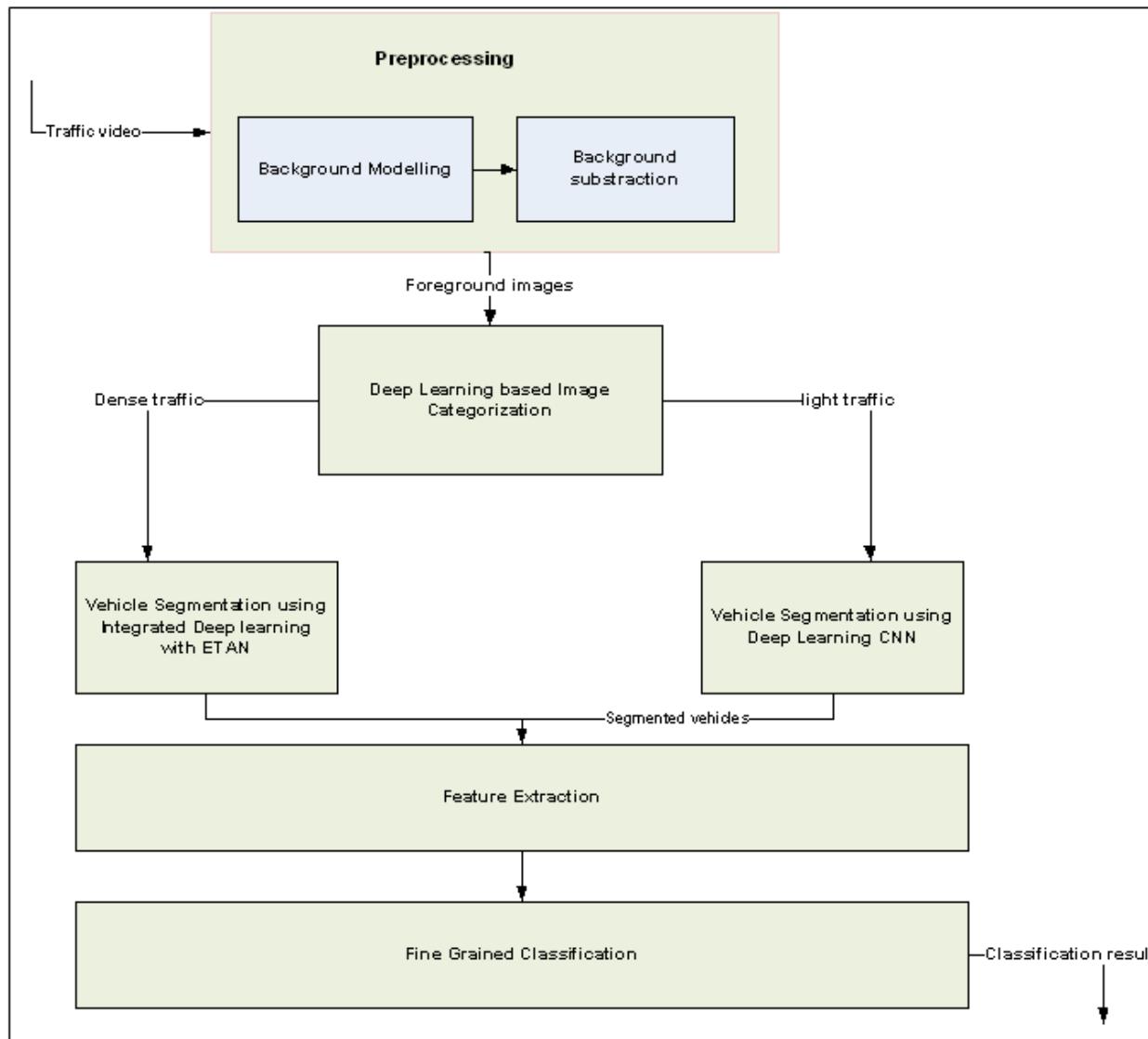


Fig. 1. Proposed Architecture.

This work proposes an adaptive background modeling where the background is first initialized by analysis of few frames and then it is continuously updated for every time foreground is extracted from subsequent frames. This is different from previous approaches of creating a fixed background model by analyzing all the frames. The advantage in this type of adaptive modeling is that unimportant backgrounds do not appear for long period of time and they disappear in subsequent background models.

The initial background model is initialized by taking the pixel values of the first frame and then the model is subsequently updated by calculating the pixel value for each pixel in the background model as

$$p_k(x) = \tilde{p}_{k-1}(x) + \frac{1}{G_k \sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \left(\frac{x-x_k}{\sigma}\right)^2\right) \quad (1)$$

Where, x_k is the pixel value observed at the k^{th} frame.

Where N is the number of frames so far considered.

The value of \tilde{p}_k is calculated as

$$\tilde{p}_k(x) = \frac{p_k(x)}{\sum_0^N p_k(x)} \quad (2)$$

Where N is the number of frames so far considered.

G_k is the gain parameter which control the learning rate of background modeling. The prior information slowly disappears and new information is learnt slowly by increasing the G_k value. The model learning can be made fast by decreasing the G_k value. G_k is made adaptive using sigmoid function so that learning is fast initially and slow later on and finally settles to constant after processing many frames.

$$G_k = Gain \times \frac{2}{1 + \exp(-\frac{cnt - \beta}{\gamma})} \quad (3)$$

The values of $Gain, \beta$ controls the inflection point of the sinusoid function and gradient is controlled by γ . cnt is the continuously increasing value proportional to the frames.

For every new frame, background model is updated and then foreground is extracted by subtraction. After obtaining the foreground, shadow and sudden illumination changes are removed and made suitable for further processing. Shadows are detected in HSV color space since it has the information of hue, saturation and brightness. The presence of shadow for a pixel x is calculated as

$$S(x) = \begin{cases} 1, & \text{if } \alpha \leq \frac{I_K^V(x)}{B_K^V(x)} \leq \beta^{I_K^S(x)-B_K^S(x)} \leq \\ & \tau_S \wedge |(I_K^H(x) - B_K^H(x))| \leq \tau_H \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

I_K^V is the brightness value of current frame and B_K^V is the brightness value of the background frame. I_K^S is the saturation value of current frame and B_K^S is the saturation value of the background frame. I_K^H is the hue value of the current frame and B_K^H is the hue value of the background frame. τ_S and τ_H are the thresholds of hue and saturation. The parameters of α and β are usually between 0 and 1. α is related to the brightness and β is related to light intensity. After the shadow mask is constructed, the pixel value for the places where 1 is set in the mask is brightened. By this way shadows are removed.

For sudden brightness or darkness, fast adaptation is needed in the background model. This is done by initializing the cnt value as below.

$$cnt = \frac{\beta}{2}, \forall |mean(intensity(x)) - mv_k| > T \quad (5)$$

$$mv_k = (G_k - 1) \times \frac{mv_{k-1}}{G_k} + mean(intensity(x))/G_k \quad (6)$$

Where T is the threshold to initialize cnt .

B. Image Categorization

The preprocessed image must be categorized to two types based on the complexity in arriving at boundary box for foreground vehicles.

Type 1 - bounding box estimation would be easy.

Type 2 - bounding box estimation is difficult.

A convolution neural network is trained with traffic images [23] of various vehicle densities and their labels (type 1 or type 2). The five layer convolutional neural network is trained with configuration detailed in Table I.

The trained convolutional neural network is used to classify the preprocessed image to type 1 or type 2.

TABLE I. CONVOLUTIONAL NEURAL NETWORK

Input	200 * 200 resized image
Convolution layers	1 st and 2 nd layers
Pooling layers	2
Activation function	ReLU
Dropout	1 dropout
Output	2

After resizing the image to 200*200 pixels, the resized image is passed to first convolutional layer. The output from the convolutional layer is passed to max pooling layer to reduce the dimension of features. This process is repeated for all convolutional layers. Over fitting is avoided by adding a dropout in the 4th layer. Classification is done at last layer using Softmax function. The output layer has two neurons each corresponding to one class – one for type 1 and one for type 2. The network will finally output the coordinates, confidence, and category of the object.

C. Vehicle Segmentation

This work proposes two segmentation methods. For type 1 image segmentation, YOLO deep learning network is used for segmenting the vehicles. For type 2 image segmentation, extended topological active mesh net is applied.

Type 1 Segmentation

YOLO v3 network is used for segmentation of type 1 images. YOLOv3 algorithm uses convolutional neural network adopting Darknet-53 network structure to extract features. The input image is split to equal size grids. Presence of object is detected at each grid by YOLOv3. Final bounding around the object is drawn by connecting the neighborhood grids containing the object. A novel part of YOLOv3 is that due to use of direct learning of residuals, training is simplified and detection accuracy increases.

The segmentation flow using YOLOv3 algorithm is shown in Fig. 2.

The final output of the YOLOv3 algorithm is the coordinates of the detected vehicles and the vehicle category in terms of car, bus, truck and motorcycle. From the coordinates bounding box is drawn on the input image and segmentation is done on the bounding boxes to give the vehicle images and their classified type. The vehicle image is passed to fine grained classification to extract further features like color and make.

Type 2 Segmentation

YOLOv3 algorithm alone cannot work for the case of higher density of vehicles and occlusions in the Type 2 images. This work augments YOLOv3 with deformable model based solution using extended topological active net segmentation for type 2 images.

The image is segmented using YOLOv3 as type1 and vehicle bounding box is obtained.

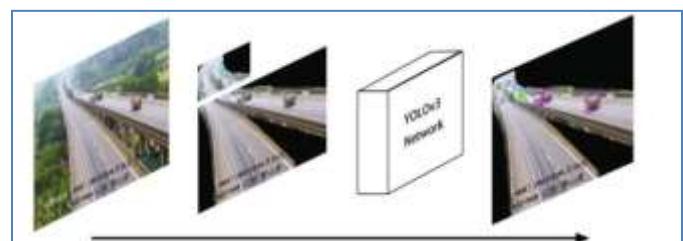


Fig. 2. Type 1 Segmentation.

A mesh is placed over the entire preprocessed image. The mesh portions lying only with the vehicle bounding box found by YOLOv3 is kept and rest of the mesh is removed. Over this image, extended topological active net segmentation is done to arrive at accurate vehicle boundaries. To achieve this, the links in the mesh can be categorized to one of following.

- 1) Links completely within the object.
- 2) Links at boundary of object and background.
- 3) Links at the background.

The links at the boundary must be removed, so that the remaining links represents the object. To speed the process of removing of the links, the links must be first classified. To speed up this classification a Naïve Bayesian classifier is built.

For a link following five features are extracted and they are classified to one of three labels defined above using a trained Naïve Bayes classifier. Following are the features extracted from each link.

- 1) Local minima of a link (f1).
- 2) Thickness of a probable edge (f2).
- 3) *LoG* value around the starting node (f3).
- 4) *LoG* value around the ending node (f4).
- 5) Difference in dominant color around the two endpoint of the link (f5).

Local minima of the link (AB) are calculated using link features shown in Fig. 3 as follows.

With D as the middle of AB, the horizontal direction HH' is split to equal spaced points along the span of the next neighbor link. If the intensity distribution along S₀, S₁, S₂... are monotonically increasing, the difference between the initial and final sampling point is taken as candidate feature for the direction of DH'. In case S₀, S₁, S₂... are not monotonically increasing, the candidate feature value for the direction of DH' is taken as 0. Similarly the feature values along DH, DV, DH' is taken and the maximum of these values is taken as local minima of the link.

Thickness of a probable edge is calculated as follows:

For each of axis(DH', DH, DV, DV') the maximum range of monotonically increasing or decreasing value of the sampled points is taken and the minimum of these four values is the indication of thickness of probable edge at the boundary.

LoG is calculated with Laplacian of a Gaussian filter of size 5X5 around the starting node A and the ending node B. It indicates the presence of edges near the nodes A and B. It is calculated as

$$LoG(x, y) = -\frac{1}{\pi\sigma^2} \left(1 - \frac{x^2+y^2}{2\sigma^2}\right) e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (7)$$

For all links in the mesh obtained after YOLOv3, link are categorized using Naïve Bayes classifier and all the links at the boundary are removed.

Each mesh obtained after link refinement is processed for occlusion and clutter removal as follows. Following geometric features are extracted from the mesh.

- Length (l)
- Width (w)
- Area (a)

The extracted features are compared against the known clutters like humans trespassing for similarity matching. In case of similarity, the mesh is not passed to next step for vehicle classification. If not similar, an image with mesh part alone is created and passed to YOLOv3 to arrive at the vehicle category to one of following types- car, bus, truck and motorcycle. The flow of this procedure is given in Fig. 4. Due to this refinement of object boundaries using ETAN segmentation, the clutters affecting the accuracy of vehicle classification are removed in the images.

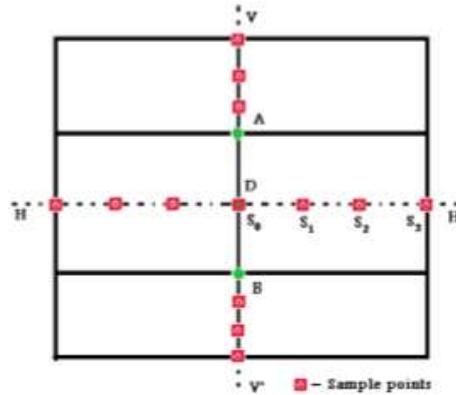


Fig. 3. Link Features.

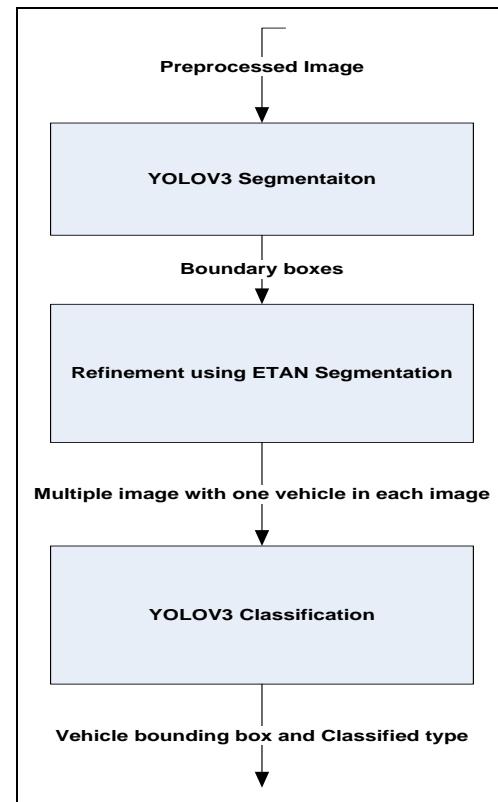


Fig. 4. Type 2 Segmentation Flow.

D. Fine Grained Vehicle Classification

From the individual vehicle detected in earlier process, features are extracted for fine grained classification like color, make of the vehicle. From the vehicle image, following color features are extracted

- 1) RGB color histogram.
- 2) Hue histogram.
- 3) Color moment.

A linear SVM is trained with color features as the input and the vehicle color as the output. The color features extracted from the vehicle is passed as input to the linear SVM as in Fig. 5 to classify the color of the vehicle.

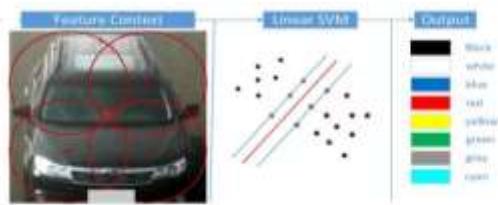


Fig. 5. Vehicle Color Classification.

Bag of SURF feature based make and model recognition approach proposed in [20] is executed on the vehicular image to classify the make and model.

IV. CONTRIBUTION OF THE PROPOSED SOLUTION

The proposed solution has following contributions.

- 1) A novel adaptive background model proposed in this work with variable gain control is able to remove occlusions like shadows, illuminations from foreground.
- 2) The vehicle segmentation model to be applied for better segmentation of vehicles is decided based on the density of vehicle distribution in the image. Compared to previous works of applying one particular segmentation model for all density, the proposed work uses different segmentation model based on density distribution.
- 3) Novel deformation model based segmentation is proposed to mitigate the effects of clutters and occlusion due to colliding vehicle.

V. RESULT

The performance of the proposed traffic adaptive fine grained vehicular classification is tested against MIO-TCD dataset [22]. The dataset has about 6 lakh images in 11 categories. The image of varying vehicle densities and occlusions are selected for testing. The performance of the proposed solution is compared against YOLO based classification method proposed in [9] and CNN based classification method proposed in [14].

The performance is compared for vehicular classification in terms of following parameters.

- 1) Precision
- 2) Recall
- 3) Accuracy

The performance is measured for a total of 1000 images from the dataset for four different categories of car, bus, truck and motorcycle.

Vehicle categorization accuracy for Bus in the proposed solution is 7.4% higher compared to [9] and 4.3% higher compared to [14]. The comparison of vehicle categorization for bus is shown in Fig. 6. The result of vehicle categorization performance for bus is given in Table II.

TABLE II. VEHICLE CATEGORIZATION PERFORMANCE FOR BUS

	[9]	[14]	Proposed
Accuracy	87	90	94
Precision	88	91	96
Recall	89	88	93

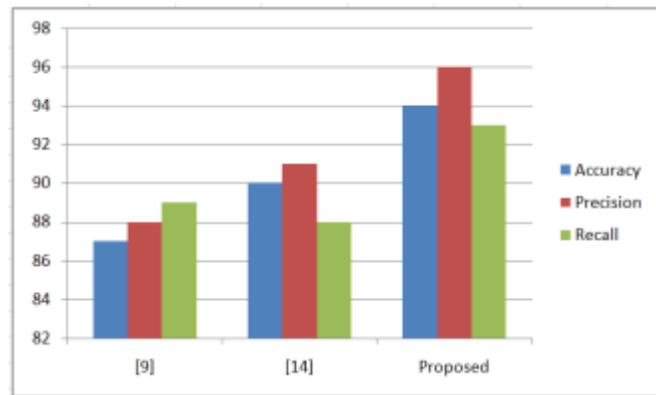


Fig. 6. Comparison of Vehicle Categorization for Bus.

Vehicle categorization accuracy for Truck in the proposed solution is 7.3% higher compared to [9] and 4.2% higher compared to [14] and is given Table III. The comparison of vehicle categorization for truck is shown in Fig. 7.

TABLE III. VEHICLE CATEGORIZATION PERFORMANCE FOR TRUCK

	[9]	[14]	Proposed
Accuracy	88	91	95
Precision	89	92	97
Recall	89	89	94

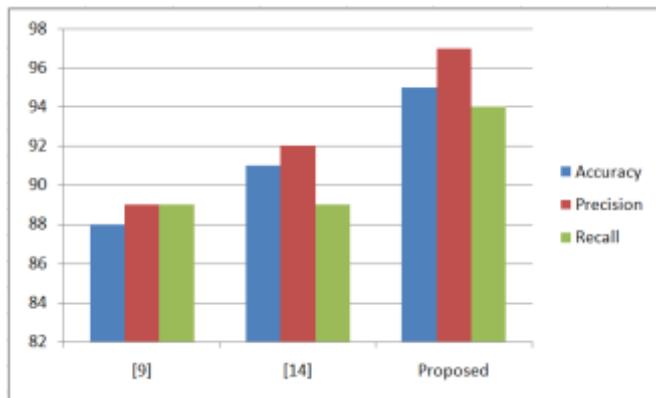


Fig. 7. Comparison of Vehicle Categorization for Truck.

Vehicle categorization accuracy for Car in the proposed solution is 10.63% higher compared to [9] and 9.57% higher compared to [14] and is given Table IV. The comparison of vehicle categorization for car is shown in Fig. 8.

TABLE IV. VEHICLE CATEGORIZATION PERFORMANCE FOR CAR

	[9]	[14]	Proposed
Accuracy	84	85	94
Precision	83	86	96
Recall	84	85	94

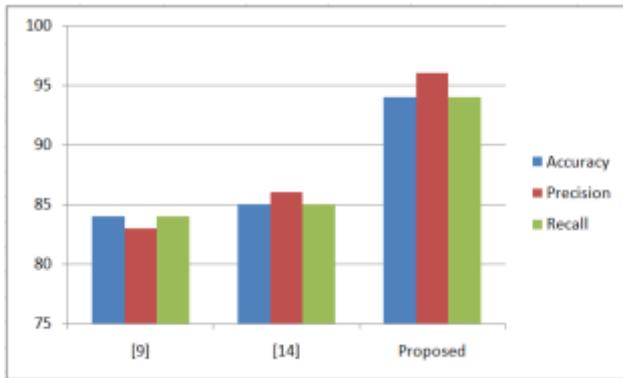


Fig. 8. Comparison of Vehicle Categorization for Car.

Vehicle categorization accuracy for Motorcycle in the proposed solution is 14.13% higher compared to [9] and 11.95% higher compared to [14] and is given Table V. The comparison of vehicle categorization for car is shown in Fig. 9.

TABLE V. VEHICLE CATEGORIZATION PERFORMANCE FOR MOTORCYCLE

	[9]	[14]	Proposed
Accuracy	79	81	92
Precision	79	81	95
Recall	78	78	91

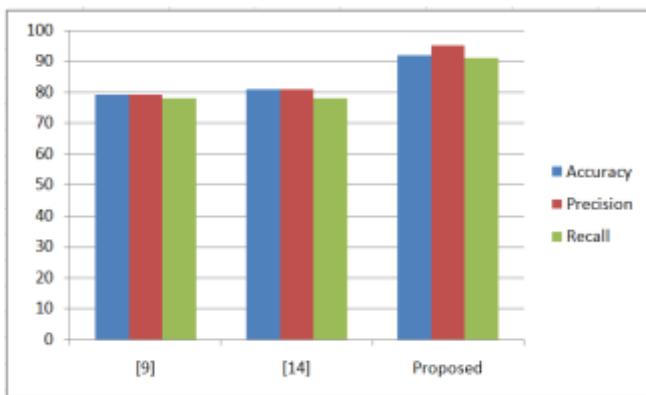


Fig. 9. Comparison of Vehicle Categorization for Motorcycle.

The vehicle categorization accuracy is higher in the proposed solution due to better segmentation of vehicles even in presence of occlusions and clutters. ETAN segmentation is able to accurately detect vehicle boundaries. Due to accurate segmentation of vehicles, the accuracy of deep learning

classification has also increased. Due to this reason, the proposed solution performed better than other deep learning classification methods. The results also prove a consistent performance of proposed solution for all types of vehicles. Even for small vehicles, the classification accuracy is higher in proposed solution due to accurate segmentation with ETAN.

VI. CONCLUSION

A traffic adaptive fine grained vehicular classification using deep learning is proposed in this work. The proposed solution is able to solve the problems of shadow, clutter etc. Adaptive background modeling with shadow elimination and dynamic contrast elimination is done as preprocessing. The preprocessed image is classified into two types based on the complexity in arriving at bounding boxes around the vehicles. YOLOV3 deep learning model and its integration with extended topological active net segmentation is followed for vehicle segmentation and vehicle classification. The proposed solution has on average 7.5% more accuracy compared to existing solutions in terms of vehicle classification. Tracking the classified vehicles in successive frames in the video can be considered as the part of future work.

REFERENCES

- [1] Kul, Seda&Eken, Süleyman&Sayar, Ahmet. (2017). Distributed and collaborative real-time vehicle detection and classification over the video streams. International Journal of Advanced Robotic Systems. 14. 172988141772078. 10.1177/172988141772078.
- [2] H. Jung, M. Choi, J. Jung, J. Lee, S. Kwon and W. Y. Jung, "ResNet-Based Vehicle Classification and Localization in Traffic Surveillance Systems," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, 2017, pp. 934-940, doi: 10.1109/CVPRW.2017.129.
- [3] Boukerche, Azzedine& Siddiqui, Abdul &Mammeri, Abdelhamid. (2017). Automated Vehicle Detection and Classification: Models, Methods, and Techniques. ACM Computing Surveys. 50. 1-39. 10.1145/3107614.
- [4] Xiao, Wen &Vallet, Bruno & Schindler, Konrad &Paparoditis, Nicolas. (2016). Street-side vehicle detection, classification and change detection using mobile laser scanning data. ISPRS Journal of Photogrammetry and Remote Sensing. 114. 166-178. 10.1016/j.isprsjprs.2016.02.007.
- [5] L. Yang, P. Luo, C. C. Loy and X. Tang, "A large-scale car dataset for fine-grained categorization and verification," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 3973-3981, doi: 10.1109/CVPR.2015.7299023.
- [6] W. Maungmai and C. Nuthong, "Vehicle Classification with Deep Learning," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019, pp. 294-298, doi: 10.1109/ICCS.2019.8821689.
- [7] Nam, Y., Nam, Y. Vehicle classification based on images from visible light and thermal cameras. J. Image Video Proc. 2018, 5 (2018). <https://doi.org/10.1186/s13640-018-0245-2>.
- [8] Sarikan, Selim&Ozbayoglu, Murat &Zilci, Oguzhan. (2017). Automated Vehicle Classification with Image Processing and Computational Intelligence. Procedia Computer Science. 114. 515-522. 10.1016/j.procs.2017.09.022.
- [9] Song, H., Liang, H., Li, H. et al. Vision-based vehicle detection and counting system using deep learning in highway scenes. Eur. Transp. Res. Rev. 11, 51 (2019). <https://doi.org/10.1186/s12544-019-0390-4>.
- [10] Psyllos, Apostolos&Anagnostopoulos, Christos-Nikolaos &Kayafas, Eleftherios. (2010). Vehicle Logo Recognition Using a SIFT-Based Enhanced Matching Scheme. Intelligent Transportation Systems, IEEE Transactions on. 11. 322 - 328. 10.1109/TITS.2010.2042714.
- [11] P. Chen, X. Bai, and W. Liu, "Vehicle Color Recognition on Urban Road by Feature Context.", IEEE Transactions on Intelligent Transportation Systems, 2014.

- [12] Paweł Badura and Maria Skotnicka. 2014. Automatic car make recognition in low-quality images. In Information Technologies in Biomedicine, Volume 3, EwaPietka, Jacek Kawa, and WojciechWieclawek (Eds.). Advances in Intelligent Systems and Computing, Vol. 283. Springer International Publishing, 235–246. DOI:http://dx.doi.org/10.1007/978-3-319-06593-9_21.
- [13] Z. Dong, Y. Wu, M. Pei, and Y. Jia. 2015. Vehicle type classification using a semi supervised convolutional neural network. IEEE Trans. Intell. Transport. Syst. 99 (2015), 1–10.
- [14] Roecker, Max & Costa, Yandre& Almeida, Joao & Matsushita, Gustavo. (2018). Automatic Vehicle type Classification with Convolutional Neural Networks. 1-5. 10.1109/IWSSIP.2018.8439406.
- [15] W. Sun, X. Zhang, S. Shi, J. He, and Y. Jin, “Vehicle type recognition combining global and local features via two-stage classification,” Mathematical Problems in Engineering, vol. 2017, November 2017.
- [16] S. Bai, Z. Liu, and C. Yao, “Classify vehicles in traffic scene images with deformable part-based models,” Machine Vision and Applications, pp. 1–11, November 2017.
- [17] M. Liang, X. Huang, C. Chen, X. Chen and A. Tokuta, "Counting and Classification of Highway Vehicles by Regression Analysis," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2878-2888, Oct. 2015, doi: 10.1109/TITS.2015.2424917.
- [18] M. Kafai and B. Bhanu. 2012. Dynamic bayesian networks for vehicle classification in video. IEEE Trans. Industr. Info. 8, 1 (Feb 2012), 100–109.
- [19] Haoyu Peng, Xun Wang, Huiyan Wang, and Wenwu Yang. 2015. Recognition of low-resolution logos in vehicle images based on statistical random sparse distribution. IEEE Trans. Intell. Transport. Syst. 16, 2 (April 2015), 681–691. DOI:<http://dx.doi.org/10.1109/TITS.2014.2336675>.
- [20] A. J. Siddiqui, A. Mammeri, and A. Boukerche. 2016. Real-time vehicle make and model recognition based on a bag of SURF features. IEEE Trans. Intell. Transport. Syst. 17, 11 (Nov 2016), 3205–3219. DOI:<http://dx.doi.org/10.1109/TITS.2016.2545640>.
- [21] Shobha B S and Deepu R, “A Review on Video Based Vehicle Detection, Recognition and Tracking”, 3rd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2018.
- [22] <https://www.kaggle.com/yash88600/miotcd-dataset-50000-imagesclassification>.
- [23] <https://www.aicitychallenge.org/2021-data-and-evaluation/>.