Pedestrian and Vehicle Detection Using Night-Vision Camera through CNN on Indian Roads

Mayank Raj
Department of Mechanical Engineering
Galgotias University
Greater Noida, India
Email Id: - mail.mayankraj94@gmail.com

Dr. Swet Chandan
Department of Mechanical Engineering
Galgotias University
Greater Noida, India
Email Id: - swetchandan@gmail.com

***Abstract*— Low visibility is one of the leading causes of vehicle accidents on Indian roads. Presently there are not many research accessible to manage these sorts of circumstances. Performance of object detection algorithm has very less accuracy in case of night time because the intensity of luminance at night is very less with respect to daytime, even human eye unfit to foresee all objects at night. For better accuracy of detection, we have to use thermal vision camera which costs a lot. In the present work we proposed a Convolutional Neural Network (CNN) based modified Single Shot Multi-Box Detection (SSD) method to identify the pedestrian and vehicles at night time utilizing night-vision camera. We have tried the actualized calculation on tests from Delhi-NCR area which incorporates recordings of highways and road streets. We have implemented certain filters as pre-treatment of sample videos (29fps, 1080p) before implementing the algorithm to improve precision. With the help of our modified algorithm, we are able to detect vehicle and pedestrian using the night-vision camera in real-time. We have achieved an accuracy of 85.28% which is superior to any other algorithm and process in this field.**

Keywords*— SSD; Night-Vision; convolutional layer; ANN; vehicle detection; Pedestrian detection; CNN; real-time; multi-box; Filters;

I. Introduction

Recognition of vehicles and pedestrian are challenging even for human eye at night time. According to the report of ministry of road transport and highways [1], during night time driving is generally 3-4 times more hazardous as compared to daytime.

TABLE I. ROAD ACCIDENTS AT NIGHT AT DIFFERENT INTERVAL

Time	Number of Accidents
06:00 - 09:00 PM hrs	86,836
09:00 - 12:00 PM hrs	51,425
12:00 - 03:00 AM hrs	27,954
03:00 - 06:00 AM hrs	30,291
Total (Night Time)	196,506

(Source: http://pibphoto.nic.in/documents/rlink/2016/jun/p20166905.pdf)

Above table demonstrates that a number of the accident is very high at night time even the traffic is very less. If we take accidents to traffic ratio the statistics of night time is quite frightening. According to the report, pedestrian accidents are of 25% of all accidents and over half of their death takes place during night time. To increase the safety of the night time traffic some automotive industries introduces modern solutions. The fascinating solution in this field is a night vision system. It helps the driver by enhancing the driver nighttime observation ability. This system can reduce accidents of the vehicle as well as the pedestrian. However recent Uber nighttime accident that kills a pedestrian challenges the accuracy of innovation in the field of nighttime detection [2].

Our work currently focuses on the roads of Delhi- NCR region which solely accounts for 13,468 deaths in the year 2015 [1]. The detection of vehicle and pedestrian are very challenging in case of Indian roads as compare to developed countries roads. In case of Indian roads, there are many challenges such as inaccessibility of street lights, low visibility due to pollution, fog and smog, driver fatigue, unskilled drivers, unavailability of nighttime lane indicators, disgraceful behavior of pedestrian, wrong side overtaking and so on. So simply executing an algorithm is not sufficient, so we have to modify the algorithm using certain filters for better precision. As compared to works done till date, cost of using a thermal vision camera is very high. If we consider countries like India where everybody cannot afford these high-cost solutions so instead of using thermal vision camera we have to find a better alternative which is low cost but with high accuracy ie. Night-Vision Cameras. Getting 100% accuracy in the night time is still challenging. Consequently, all such factors must be considered to construct a precise nighttime vehicle and pedestrian detection method.

II. Related Work

Vision-based object detection picking up a great amount of prominence in the area of advanced driver assistance system (ADAS) and self-driving cars in the recent years. Several amounts of research based on the object detection are presented to date. Object detection is the topic of research for quite a while. Designers and engineers had started working on it since 1958 but the most significant result comes in the form of Viola-Jones Algorithm in 2001 [3]. This algorithm had supported the measure of research in this field as it acts as a

ISBN: 978-1-5386-4119-4/18/$31.00 ©2018 IEEE

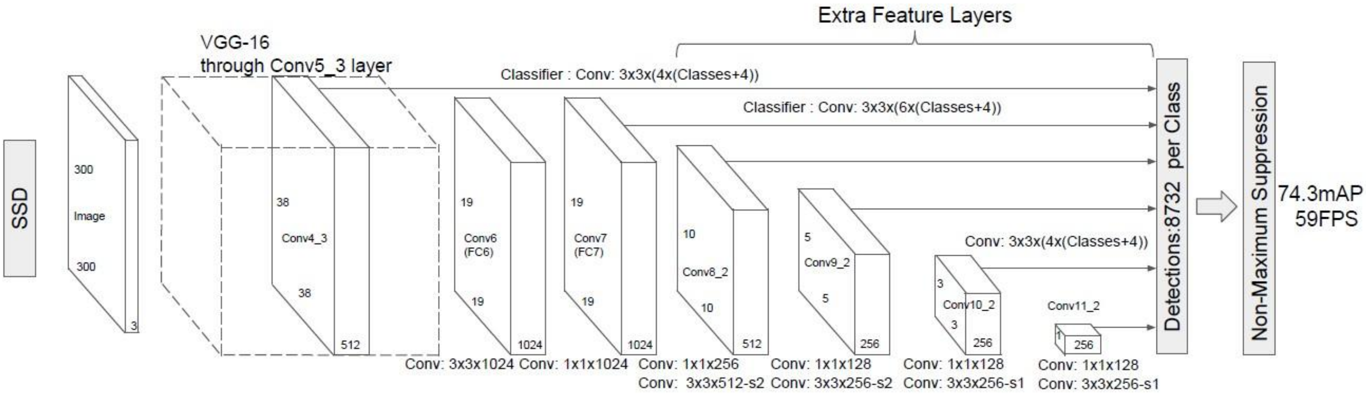

Fig. 2. SSD Method Architecture (Source: https://arxiv.org/pdf/1512.02325.pdf)

the basic framework of object detection. Rapid transformation in this field carried out with the introduction of deep learning. After that convolution net (CN) based Alex-Net is introduced which become so popular that it wins Large Scale Visual Recognition Competition (ILSVRC). CN is utilized as image feature mapping and it depends on ReLU and GU implementation and based on old Le-Net along with data augmentation [4]. G. Ross introduced Region Based ConvNet (RCNN) that is a combination of heuristic region proposal method. Alex-Net and support vector machine (SVM) model is then taken into consideration and trained to classify the object [5]. ILSVRC 2013 winner is ZF Net, which is normally an Alex-Net with slight modification [6]. For scalable and high-quality object detection Multi-box method is used which is not only an object recognition but a CN based object proposal method [7]. GoogLeNet (inception) is the winner of ILSVRC 2014 in which instead of using traditional convolution and max-pooling, it uses inception modules [8]. Some changes take place in the field of object detection when you only look once (YOLO) algorithm is innovated which is a modification of multi-box method. It reshaped multi-box from object proposal solution to an object recognition method substituting a soft-max layer [9]. Single Shot Multi-box Detection is invented in 2016 which can detect everything in a single shot, it focuses on the image once it does not have to go back to the image. It does not require the object proposal and does have to use many convolutional layers which lessens the time and computational cost [10]

In the night time, the attributes of a vehicle are not obvious. From 2002 to 2009 Night time detection is completely based on features of two taillights of the cars [11]-[18]. At first, it has to ignore non-vehicle object such as traffic light and then utilizes the characteristics to red-colored taillight to identify the vehicle. According to the brightness, vehicles are labeled. This method has some drawbacks as in case of a broken taillight, improper weather conditions it is difficult to detect the vehicles. Authors move their concentration towards radar, lidar, thermal vision camera for better accuracy. These are perfect for premium segments cars such as Cadillac, BMW, Audi, Mercedes and so on. But in case of a country like India where a normal person cannot afford these cars, so we have to move to low-cost solutions that are of high accuracy and stable detection in real-time which are Night-Vision camera. They further act as best ADAS and an important component of autonomous cars.

III. METHODOLOGY

We are using a Convolutional Neural Network Based Single Shot Multi-box detection algorithm for detection of vehicle and pedestrian. Firstly pre-treatment of the video will be carried out before implementing the algorithm. Single shot multi-box detection is the method based on a pre-trained model for object detection. It is a based on a deep neural network that does not resample pixels and it is very precise than other methods [14]. Have a glance at Fig.1, we can easily observe that cars are running on the road. The challenge is here how the algorithm will detect these cars. According to the basics of Convolutional Neural Network (CNN) [15] we only look for an object in an image, but with SSD we detect the exact location of object class and put a rectangle to interpret the result. SSD algorithm helps us to save time and computational cost by bypassing the object proposal method.

Fig. 1. A sample of Night-Vision image

For better understanding take Fig.1 as a reference image, so using basics we have to look for object proposals. We have noticed through our naked eye that only two portion of the whole image contains cars. If we start searching for a car in the whole image by pixels by pixels then it takes a lot of time and lot of costs and thus algorithm becomes slow.

So why we prefer SSD over other algorithms, it’s because it makes the process much easier that we can do all object class detection in one shot [10]. This algorithm refers the image once and does not come back to it after that. It does not have to run many convolutional layers which reduces the computational costs and time. Refer Fig.2, here we notice that this fig contains many boxes. All these boxes go through the

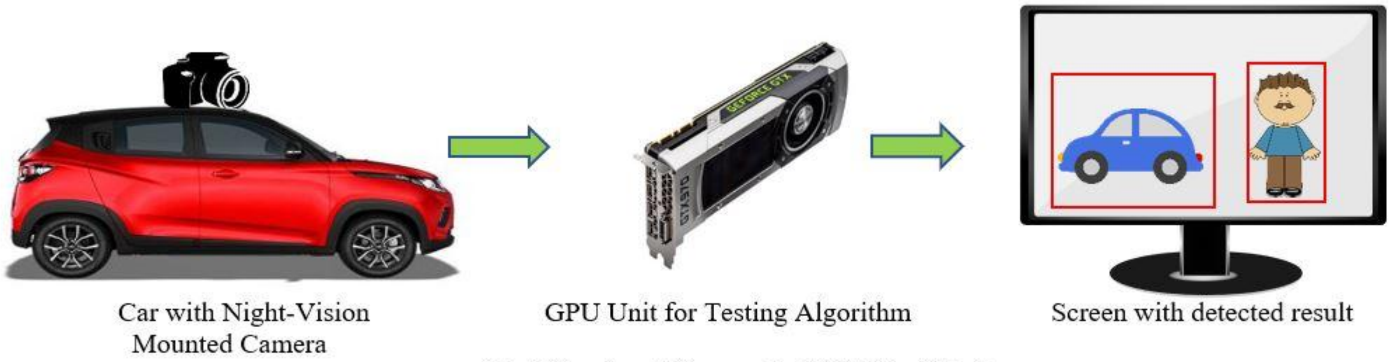

Fig. 3. Experimental Framework of Night-Time Detection

network at the same time and also remembers which boxes to deal with. The fascinating side of SSD Algorithm is that there are many convolutions to reduce the image size which starts with 300 x 300 pixels then 38, 19, 10, 5, 3, 1 and the detection happens in one huge convolution. We have divided methodology in some parts.

A. Experimental Setup

Night-Time detection equipment is embedded in premium segments of cars. Companies like Cadillac, BMW, Audi, Mercedes are manufacturing cars with night-vision supportable equipment. Cost of the equipment and also the cost of the cars are very high. So, in this paper, we assemble a very low-cost Night-Vision system which is very affordable in case of Indian consumers. Our experimental setup contains Canon DSLR camera fitted with Astro-Scope Night Vision Module with lens aperture: F/5, Focal length: 52 mm, ISO: 800 and with no flash. For testing the algorithm, we have taken CUDA enabled NVIDIA GTX 1080 as our graphics processing unit (GPU). And finally, a normal HP touchscreen monitor to interpret the result. The whole setup can be easily assembled in any vehicle of any class with ease.

B. Pretreatment of Video Sample

Live feed of video sample is continuously transferred from the camera to GPU. SSD algorithm can easily detect the sample of video during the daytime. But in case of night-time where illuminance is very low. Even human unable to detect each and every object. If we use the night sample directly with any pre-treatment and without Night-Vision Camera accuracy of detection is very low which leads to very less computational efficiency.

Fig.4. Input & Output of normal camera (Night-Time)

In Fig.4 we can notice that accuracy of detection is very less as it can detect only one car which is not acceptable. Let us take the same sample from night vision camera and check its accuracy without pre-treatment of the video.

Fig.5. Input & Output from Night-Vision Camera

From Fig.5 we can notice that accuracy of detection using Night-Vision camera is slightly better than Normal cameras but still the accuracy is not up to the mark. For better accuracy, we have to apply certain filters.

1) Negative Filters:

In negative filters in which area with light appears darkest and the area in dark appears lightest. This feature helps the video sample to extract the area in darkest and discover every small bit of features. So cars hidden inside dark zone after using Night-Vision camera can easily be extracted out. The method used for sampling contains 9 bytes keyframe extraction and about 340 frames from the video [19].

Fig.6. Output video after Negative Filter (Reference to Fig.5)

From the Fig.6 we can notice that it after applying negative filter we have a very good detailing of object outline. It enhances the detection but we have to apply some more filters so that we can achieve a precise platform for enhanced detection. Every pixel in the gradient image measures the change in intensity.

2) Gradient Filter (Edge Extraction):

This filter brought changes in the intensity or color of an image. Image gradients are used for edge detection by extracting information from the image. Firstly we apply negative filter because we want to extract every bit of information from the image. From gradient image pixel with large gradient assume to be an edge pixel. Here we use log operator edge detection theory and canny edge detection operator. Feature map of this operator is given in Fig.7 [20].

-2	-4	-4	-4	-2
-4	0	8	0	-4
-4	8	24	8	-4
-4	0	8	0	-4
-2	-4	-4	-4	-2

Fig.7. Log Operator Feature Map

After applying this operator we get following results.

Fig.8. Output after Edge detection

After getting the edge detection result we can notice that edge of the car can be easily noticed even in slight dark condition [21]. We have tested our algorithm in output result we observed that with little color enhancement we can get a better output result for better object detection in the night time. The final step in pre-treatment of the video is color extraction in which we apply a color code FF000#. After applying we get an incredible output which enhances and provides a better platform to our algorithm to provide a precise result (Fig.9)

C. Multi-Box Concept

If we notice the Fig.10 we can observe that every car are in rectangular boxes. We take these boxes as ground truth and after that we are going to see how the algorithm works through its own boxes to detect and again build the same boxes. Ground truth can be defined as observed evidences or empirical evidences from inferred evidences. After applying

Fig.9 Final output of video after pre-treatment

The algorithm the boxes are made on the detected object are called inferred boxes. These boxes are the suggestion given by algorithm. We need ground truth to compare them.

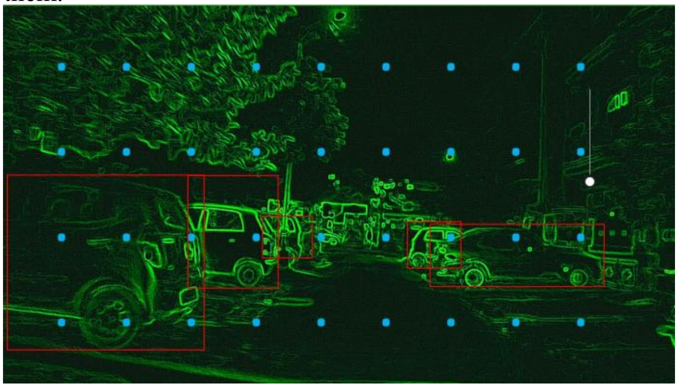

Fig.10 Image with segmentation

From Fig.10 we can observe that our algorithm will divide the image in to segments and then for each segment it will construct boxes. Then every single box is going to ask question for every class of object during testing. We can notice through Fig.11 boxes highlighted in red detect objects. We can notice from Fig.11 that some of the boxes detect very small features of the car, so we can conclude that it does not have to be full car as long as it detects features of car in the boxes. It will say as probability that 70-80% is the chance that in boxes there is a car. If it sees enough feature to say that

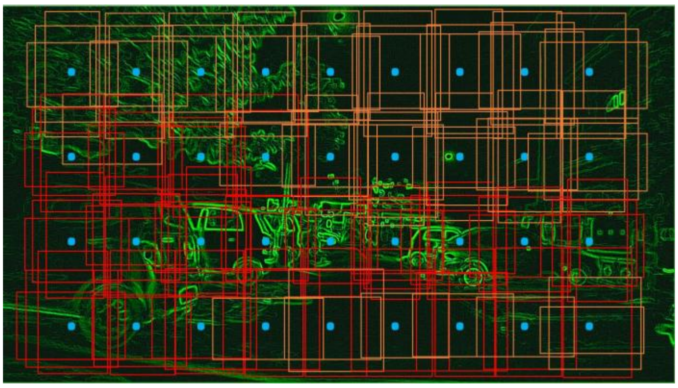

Fig.11 Region of interest- Red high-lighted Boxes (Image with boxes)

there is car inside the boxes. Every time boxes predict something we will have an output Number. If the boxes predict something then the neural network searches for features for prediction. Suppose in the boxes we want a car, but it predicts any other boxes that do not contain cars then an error has been generated. Error has been derived on the basis of ground truth. Then all errors are taken into consideration and can be reduced through backpropagation through the network and by updating the weights. It takes many iterations and training process goes until we have a negligible error.

D. *Predicting Object positions*

Focus on each box from the Fig.11 and accept it as a separate image and then for each box our algorithm will ask a question is there an object or not. We can notice from this figure that highlighted boxes overlap with the ground truth and then it is easy to say that it picks up sorts of features. But during algorithm training unhighlighted rectangle may pick up the object and highlighted rectangle unable to pick up the object. Then through the iterative process, they will know about the features. Through training two solutions to be noticed first thing is that each rectangle will now better predict the class of the object and identify if it has objects inside there and each box will use the ground truth to assess that. But the second thing accuracy of the final output rectangle that will predict the object has very less error and very high accuracy.

E. *Training and Running SSD*

Our whole algorithm works on the datasets and their training. These data sets help us to detect different classes of object.

1) Datasets: We will need training and test datasets with ground truth bounding boxes and assigned class labels to predict object. Up to 20 image class is trained in this datasets. Pascal VOC and COCO datasets are best for training.

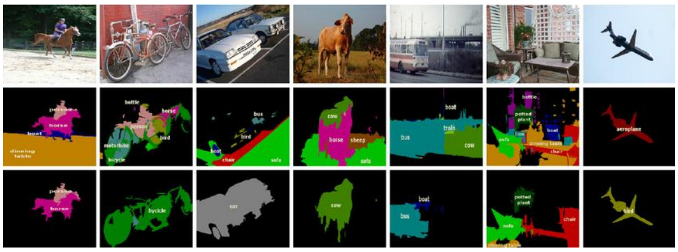

Fig.12. Image from Pascal VOC Datasets (Source: http://host.robots.ox.ac.uk/pascal/VOC/)

2) Default Bounding Box: It is needed to configure different sets of bounding boxes, of different scale and ratios for better precision of detection.

3) Feature Map: These are the output of convolutional layers and are a represent a most dominant feature of the image at a different magnitude.

4) Data Augmentation: Similar to other deep learning applications, data augmentation is very essential to teach the network to become more robust to different sizes of the object in the input.

IV. ARCHITECTURE DESIGN

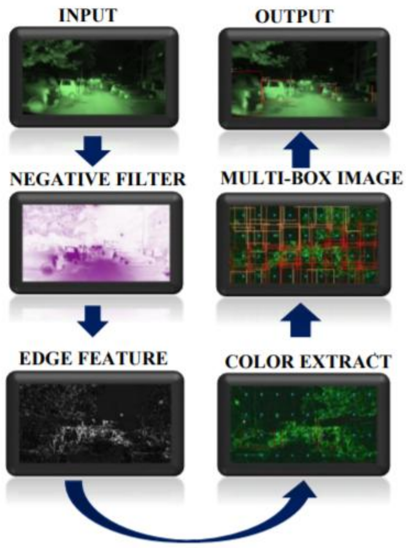

Fig.13. Architecture Design

In designing this work we have to go through four different and most important steps to get our desired output. As we can notice in Fig.13 that we have our input image (video) which first processed through a Negative filter to get our required output for the second step. In the second process, we have used Edge operator to detect all kind of edge that helps in detection objects at night easily, this step is then followed by color extraction step which makes the edges more vibrant and clear for ease of detection. Our input is now ready for the algorithm to start testing. In the final step, our third process input is tested via our SSD algorithm, so that we can get our final results.

V. EXPERIMENTATIONS AND RESULTS

In this paper our entire framework is executed on the Nvidia GTX 1080 Cuda enabled GPU based on Windows 10 platform along with Canon Astro-scope night-vision camera attached to the GPU through Network server (Netgear Router) installed in the car. The camera is fixed to the roof of the car using 3M adhesive tape. Resolution of captured video is 1080 x 1920 pixels in 29 frames per seconds (fps). Our algorithm works on Anaconda based python in which all the packages are installed before testing the algorithm.

Fig.13. Experimental Setup for Night-Time Detection

Some illustrative result of the experiment is demonstrated in the Fig.14, 15, 16, 17, 18, 19. In Fig.14 we can notice that car and motorbike can be easily detected in the same frame.

Fig.14. Car and motorbike detected in the frame

In Fig.15 we can see that Pedestrian and car effectively recognized in the edge.

Fig.15. Pedestrian and car detected at the same instance

In the Fig.16, 17 our accuracy of the modified algorithm can be easily seen as it detects that car which is barely seen through the human eye.

Fig.16. Car in fully dark is easily detected

Fig.17. Right side car which is barely visible is detected

In Fig.18 we can notice that different class of vehicles can be easily detected. In this figure car and bus is easily detected.

Fig.18. Different class of vehicle is detected (Car and Bus)

In case of fig.19, we can observe that auto-rickshaw cannot be detected in low light. Our algorithm is also trained to detect auto-rickshaw. But in this case, it is failed to detect it.

Fig.19. The car is detected but auto-rickshaw is not detected

TABLE I. VEHICLE AND PEDESTRIAN ACCURACY

S/N	Result Analysis of Object Class Accuracy		
	Class of Object	***Success Rate***	***Miss Rate***
1.	Vehicle Class (All)	83.26	16.74
2.	Pedestrian	87.54	12.46

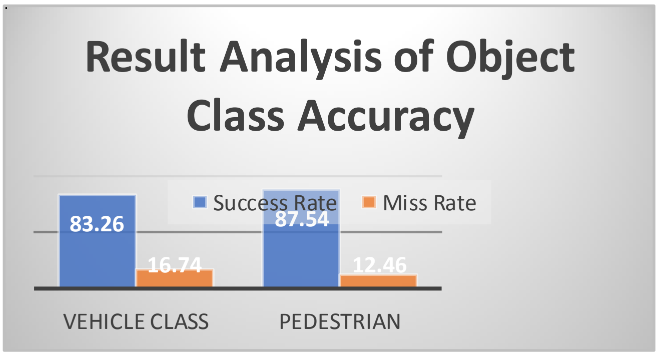

Fig.20. Graph showing our Result Accuracy

Our modified algorithm provides an average accuracy of 85.28%, which is very higher if we compare it with the accuracy of normal-vison camera used at night time ie. 63.28%.

VI. CONCLUSION

In this paper we have presented object detection algorithm with using certain filters for pre-processing the images. We have used the night-vision camera in low light conditions. Our system can effectively detect pedestrian and vehicle during night-time in roads of Delhi-NCR with an overall accuracy of 85.28% despite several challenges. Our main aim is to develop a system which offers low cost solutions for pedestrian and vehicle detection which can be further used in the developing countries and save millions of lives. In the future despite everything we still need to consider environmental factor to optimize system performance and make it more useful in the coming time.

ACKNOWLEDGMENT

This research is partially supported by Galgotias University, Greater Noida. We would like to show our sincere gratitude to the Hadelin de Ponteves (AI Entrepreneur) for sharing his pearls of wisdom with us during the course of this research, although any errors are our own and should not tarnish the reputations of these esteemed person.

REFERENCES

[1] Report by transport research wing, “Statistics of an accident in India (2015),” Ministry of Road Transport and Highway, Government of India.

[2] Daisuke Wakabayashi, “Self-driving car kills a pedestrian in Arizona,” The New-York Times.

[3] Paul Viola & Michael Jones. “Rapid Object Detection using a Boosted Cascade of Simple Features,” Conference on Computer Vision and pattern recognition, 2001.

[4] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E.Hinton. “Imagenet classification with deep convolutional neural networks”.

[5] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” Cornell University Library, 11 Nov 2013.

[6] Pierre Sermanet, David Eigen, Xiang Zhang, Michael Mathieu, Rob Fergus, Yann LeCun, “OverFeat: Integrated Recognition, Localization, and Detection using Convolutional Networks,” Cornell University Library, 21 Dec 2013.

[7] Christian Szegedy, Scott Reed, Dumitru Erhan, “Scalable High-Quality Object Detection,” Cornell Univerity Library, 9 dec 2015.

[8] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, ” Going Deeper with Convolutions,” Cornell University library, 17 Sep 2014.

[9] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi,” You Only Look Once: Unified, Real-Time Object Detection,” Cornell University Library, 8 June 2015.

[10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," in European Conference on Computer Vision(ECCV), 2016.

[11] Y. L. Chen, C. T. Lin, C. J. Fan, C. M. Hsieh, and B. F. Wu, “Vision-based Nighttime Vehicle Detection and Range Estimation for Driver Assistance,” in Proc. IEEE Int’l Conf. on Systems, Man, and Cybernetics, 2008, pp. 2988-2993.

[12] Y. L. Chen, Y. H. Chen, C. J. Chen, and B. F. Wu, “Nighttime Vehicle Detection for Driver Assistance and Autonomous Vehicles,” in Proc. 18th Int’l Conf. on Pattern Recognition, 2006, pp. 687-690.

[13] M. C. Lu, C. P. Tsai, M. C. Chen, Y. Y. Lu, W. Y. Wang, and C. C. Hsu, “A Practical Nighttime Vehicle Distance Alarm System,” in Proc. IEEE Int’l Conf. on System, Man, and Cybernetics, 2008, pp. 3000-3005.

[14] Y. L. Chen, B. F. Wu, and C. J. Fan, “Real-time Vision-based Multiple Vehicle Detection and Tracking for Nighttime Traffic Surveillance,” in Proc. IEEE Int’l Conf. on System, Man, and Cybernetics, 2009, pp. 3352-3358.

[15] C. C. Wang, S. S. Huang, and L. C. Fu, “Driver Assistance System for Lane Detection and Vehicle Recognition with Night Vision,” in Proc. IEEE Int’l Conf. on Intelligent Robots and Systems, Aug. 2005, pp. 3530-3535.

[16] S. Naqumo, H. Hasegawa, and N. Okamoto, “Extraction of Forward Vehicles by Front-mounted Camera using Brightness Information,” in Proc. IEEE Int’l Conf. on Electrical and Computer Engineering, May. 2003, pp 1243-1246.

[17] S. Y. Kim, S. Y. Oh, J. K. Kang, Y. W. Ryu, K. Kim, S. C. Park, and K. H. Park, “Front and Rear Vehicle Detection and Tracking in the Day and Night Times using Vision and Sonar Sensor Fusion,” in Proc. IEEE Int’l Conf. on Intelligent Robots and Systems, Aug. 2005, pp 2173-2178.

[18] M. Y. Chern and P. C. Hou, “The Lane Recognition and Vehicle Detection at Night for a Camera-assisted Car on Highway,” in Proc. IEEE Int’l Conf. on Robotics and Automation, Vol. 2, 14-19, Sep. 2003, pp. 2110-2115.

[19] Hanung Adi Nugroho, Denny Hardiyanto, and Teguh Bharata Adji, “Negative Content Filtering for Video Application,” 7th International Conference on Information Technology and Electrical Engineering (ICITEE), 2015.

[20] Yaan Lecun, Leon Bottou, Yogusha Begio, Patrik Haffner,” Gradient based learning applied to document recognition,” IEEE, 1998.

[21] You-yi Zheng, Ji-lai Rao and Lei Wu, “Edge Detection Methods in Digital Image Processing,” The 5th International Conference on Computer Science & Education Hefei, China 2010.