International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1S5, December 2019

Vehicle Classification and Detection using Deep Learning

V. Vijayaraghavan, M. Laavanya

Abstract: ***Intelligent transportation systems have acknowledged a ration of attention in the last decades. In this area vehicle classification and localization is the key task. In this task the biggest challenge is to discriminate the features of different vehicles. Further, vehicle classification and detection is a hard problem to identify and locate because wide variety of vehicles don't follow the lane discipline. In this article, to identify and locate, we have created a convolution neural network from scratch to classify and detect objects using a modern convolution neural network based on fast regions. In this work we have considered three types of vehicles like bus, car and bike for classification and detection. Our approach will use the entire image as input and create a bounding box with probability estimates of the feature classes as output. The results of the experiment have shown that the projected system can considerably improve the accuracy of the detection.***

Keywords: Convolutional neural network, Object detection, Deep learning, Image classification.

I. INTRODUCTION

People can easily identify and analyze things in the picture. Man's visual system is fast and precise, and can perform complex tasks, such as identifying many things and identifying obstacles with sensible thoughts. But in computer vision object recognition is one of the major challenge because, we shouldn't focus only on the classification of different images, we should also identify the location of things accurately in individual image. This bustle is called an object detection [1]. Object detection can provide valued information about the clear meaning of images and videos and is associated with numerous claims such as image classification [2], [3], human behavior analysis [4] and facial recognition [5]. In recent year's deep neural networks (DNN) have become a [6] powerful machine learning model. DNN show important differences with respect to traditional classification approaches. First they are profound architectures that have the ability to learn more complex models than surface models [7]. This Expressiveness and robust training algorithms allow powerful representations of objects without the need for manual design. However, large differences in types, poses, occlusions and lighting conditions make it tough to detect objects. Therefore, it attracts so much attention from researchers in this field [8], [9]. In this article, we show that algorithmic modification, which computes a deep network performance map, leads to a sophisticated and effective solution.

Manuscript published on 30 December 2019.
***** Correspondence Author (s)
V. Vijayaraghavan, Department of Electronics and Communication Engineering, Vignan's Foundation for Science, Technology and Research, Vadlamudi, Guntur, AP.vijayaraghavan123@gmail.com
M. Laavanya, Department of Electronics and Communication Engineering, Vignan's Foundation for Science, Technology and Research, Vadlamudi, Guntur, AP.laavanvijay@gmail.com

Our observation is that when the convolutional neural network extracts the features exactly, then the object is well detected by region based detectors.

Region Based Convolutional Neural Network

Region based convolution neural network (R-CNN) [10] demonstrates, the beauty of using a regional proposal [11], [12] with a neural network. R-CNN can learn more characters like textures, edges etc. One of the disadvantages is the huge overlap of the area in the image, which can lead to more calculations in the development of dig out the location of object. To resolve, fast R-CNN [13] follows the following steps: as a first step, it normalizes the complete image before sending it to CNN, then the fifth resolution level gets the properties of each sentence in the region. The next is the regression path of the restrictive framework to a neural network with regional classification according to a model with different tasks [13]. In [14] fast R-CNN uses sliding windows in the place of region proposals. In some cases, a single scale sliding window can be recast as a single convolutional level. But, fast R-CNN is time consuming, since it has to identify all region proposals. To overcome this problem, in 2015 Ross B. Girshick developed faster R-CNN [13]. Faster R-CNN directly trains and get the regions to promote the efficiency than fast R-CNN [15, 16, 17, 18]. The RPN component [9] is a fully connected detector that includes containment boxes for reference frames (anchors) of different sizes. Faster R-CNN significantly improves overall performance, especially with respect to speed detection.

II. METHODOLOGY

Image classification determines which objects in the image, such as a car or bicycle rail, while image localization provides a specific location for these objects by using restrictive fields. In order to classify the images, the convolution neural network had to recognize different objects, such as a car, bus and motorcycle. Hence image classification and localization can be defined as object detection.

Object detection = Image classification + Image localization

The workflow has 3 parts, first step is gathering the training data, second is training the model and the final one is predictions of new images. The stream of the scheme is exposed in below figure 1.

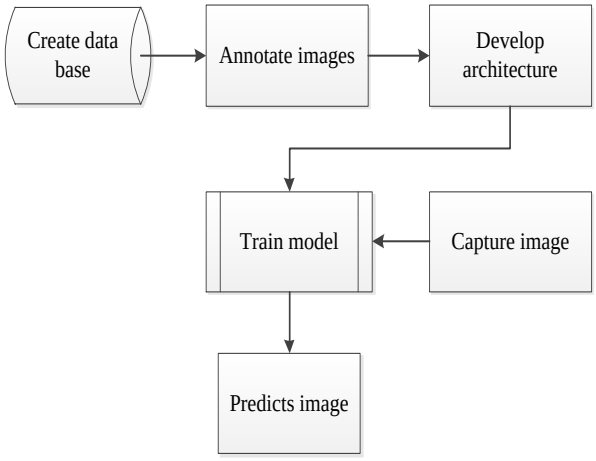

Figure 1. Work flow of object detection

Gather Training Data

For this task, camera is used to capture the data as close to the data that should be finally predicted. The data set collection has 1000 images per object. After the images are captured, the obtained set of images are resized and ground truth labelling is generated with location and labels of object of interest. But this process is a fairly intensive and time consuming task.

Training a Model using deep learning

The network design is based on the fastest R-CNN, since the convolution operation is performed only once for each image and a characteristic chart is spawned from it. Faster layers R-CNN has input, middle and last layers. The size of the input is the balance between the execution time and the number of spatial details that the detector has to decide. Intermediate levels are the main building blocks of the network, like convolution. ReLU sets and pools. These levels must be repeated to create a deeper network. The final CNN layers are usually a collection of fully connected, Softmax loss classification, and regression layers for image localization. In this document, CNN is developed from scratch and the non-linearity of Leaky-ReL U between fully coupled layers is added to progress the enactment of the detector. The developed network has 10 hidden layers, 588060 parameters and 27780 neurons. To train the object detector, the network structure of the "layers" will be transmitted through the "train Faster RCNN Object Detector" function. Once a network is developed, the network learns into a single processor system for a small set of data and for large set of data GPU is used. The GPU can be selected at run time in the training option.

Fast RCNN for image classification and localization

In Fast RCNN, we transmit the input image to CNN, which in turn creates maps of revolutionary objects. With these maps, regions of the proposal are extracted. We then use the RoI pool layer to convert all the proposed areas into a fixed size so that they can be transferred to a fully connected network. The RCNN fast approach is as follows

1. To take input image using a camera.
2. The input image is transferred to ConvNet, which returns the region of interest.
3. Apply the RoI pool level to the extracted areas.

Finally, these areas are transferred to a fully connected network, which classifies them, and also return bounding blocks, using both linear and softmax regression layers. The flow is displayed in Figure 2.

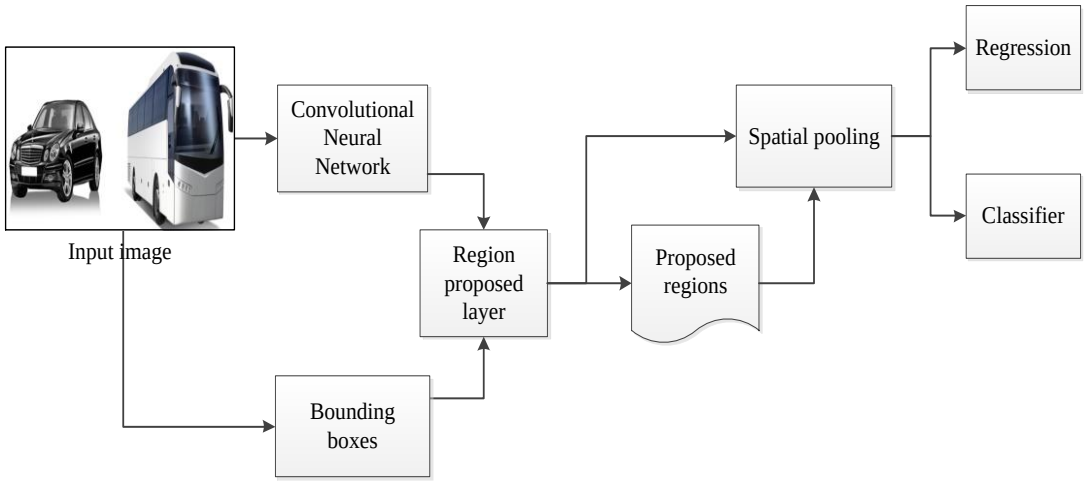

Figure 2. Fast R-CNN

III. EXPERIMENTAL RESULTS

The proposed method detects the objects by building convolutional neural network from base. The first level extracts the edges from the raw image and the second level extracts the shapes from the edge information and so on. The feature map of first level and second level convolutional layers are shown in figure 3a and 3b. The samples taken for training and the ground truth bounding boxes are shown in figure 4 for car, bike and person. The presented method is tested with another image which is not in the database and the prediction of car with bounding box is shown in figure 5. The execution step of the prediction is shown as below

Step 1: To train Region Proposal Network (RPN).

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch RMSE	Base Learning Rate
1	1	00:00:00	100.00%	0.65	1.0000e-05
1	50	00:00:11	100.00%	0.29	1.0000e-05
2	100	00:00:22	100.00%	0.65	1.0000e-05
2	150	00:00:32	100.00%	0.54	1.0000e-05
3	200	00:00:42	100.00%	0.66	1.0000e-05
3	250	00:00:53	100.00%	0.67	1.0000e-05
4	300	00:01:03	100.00%	0.13	1.0000e-05
4	350	00:01:13	100.00%	0.62	1.0000e-05
5	400	00:01:23	100.00%	0.26	1.0000e-05
5	450	00:01:33	100.00%	0.49	1.0000e-05
6	500	00:01:43	100.00%	0.99	1.0000e-05
6	550	00:01:54	100.00%	0.78	1.0000e-05
7	600	00:02:04	100.00%	0.72	1.0000e-05
7	650	00:02:14	100.00%	0.53	1.0000e-05
7	693	00:02:23	100.00%	0.71	1.0000e-05

Step 2: To train faster region convolution neural network.

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch RMSE	Base Learning Rate
1	1	00:00:00	100.00%	0.59	1.0000e-05
1	50	00:00:09	100.00%	0.66	1.0000e-05
2	100	00:00:19	100.00%	0.62	1.0000e-05
2	150	00:00:28	100.00%	1.59	1.0000e-05
3	200	00:00:37	100.00%	0.92	1.0000e-05
3	250	00:00:47	100.00%	0.83	1.0000e-05
4	300	00:00:56	100.00%	0.30	1.0000e-05
4	350	00:01:06	100.00%	1.20	1.0000e-05
5	400	00:01:15	100.00%	1.35	1.0000e-05
5	450	00:01:24	100.00%	0.87	1.0000e-05
6	500	00:01:33	100.00%	0.91	1.0000e-05
6	550	00:01:43	100.00%	1.02	1.0000e-05
7	600	00:01:52	100.00%	0.33	1.0000e-05
7	650	00:02:01	100.00%	0.43	1.0000e-05
7	686	00:02:08	100.00%	1.05	1.0000e-05

Step 3: To retrain region proposal network.

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch RMSE	Base Learning Rate
1	1	00:00:00	100.00%	0.88	1.0000e-05
1	50	00:00:03	100.00%	1.01	1.0000e-05
2	100	00:00:07	100.00%	0.88	1.0000e-05
2	150	00:00:10	100.00%	0.62	1.0000e-05
3	200	00:00:13	100.00%	0.62	1.0000e-05
3	250	00:00:17	100.00%	0.27	1.0000e-05
4	300	00:00:20	100.00%	1.18	1.0000e-05
4	350	00:00:25	100.00%	0.48	1.0000e-05
5	400	00:00:28	100.00%	0.57	1.0000e-05
5	450	00:00:31	100.00%	0.49	1.0000e-05
6	500	00:00:35	100.00%	0.78	1.0000e-05
6	550	00:00:38	100.00%	1.63	1.0000e-05
7	600	00:00:42	100.00%	1.43	1.0000e-05
7	650	00:00:45	100.00%	0.37	1.0000e-05
7	693	00:00:48	100.00%	0.87	1.0000e-05

Step 4: To retrain faster region convolution neural network.

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch RMSE	Base Learning Rate
1	1	00:00:00	100.00%	0.44	1.0000e-05
1	50	00:00:02	100.00%	0.67	1.0000e-05
2	100	00:00:05	100.00%	0.56	1.0000e-05
2	150	00:00:07	100.00%	0.83	1.0000e-05
3	200	00:00:10	100.00%	0.47	1.0000e-05
3	250	00:00:12	100.00%	0.21	1.0000e-05
4	300	00:00:15	100.00%	0.65	1.0000e-05
4	350	00:00:18	100.00%	0.14	1.0000e-05
5	400	00:00:21	100.00%	0.76	1.0000e-05
5	450	00:00:24	100.00%	1.21	1.0000e-05
6	500	00:00:26	100.00%	1.07	1.0000e-05
6	550	00:00:28	100.00%	0.55	1.0000e-05
7	600	00:00:31	100.00%	1.04	1.0000e-05
7	650	00:00:33	100.00%	0.50	1.0000e-05
7	686	00:00:35	100.00%	0.75	1.0000e-05

Table 1 displays, the performance of our network. Obviously our network is superior to other procedures. We have achieved a substantial improvement in terms of precision 10% compared to the faster modern R-CNN. Clearly, a person detection accuracy is lesser than that of a car and bicycle, since deep learning recognition procedures are not very convenient for small objects. Our network demonstrates robust detection capabilities for automobiles with a wide variety of scales, particularly for motor bike. It can be used for intelligent transport systems in real time. Therefore, our network achieves better accuracy than faster R-CNN.

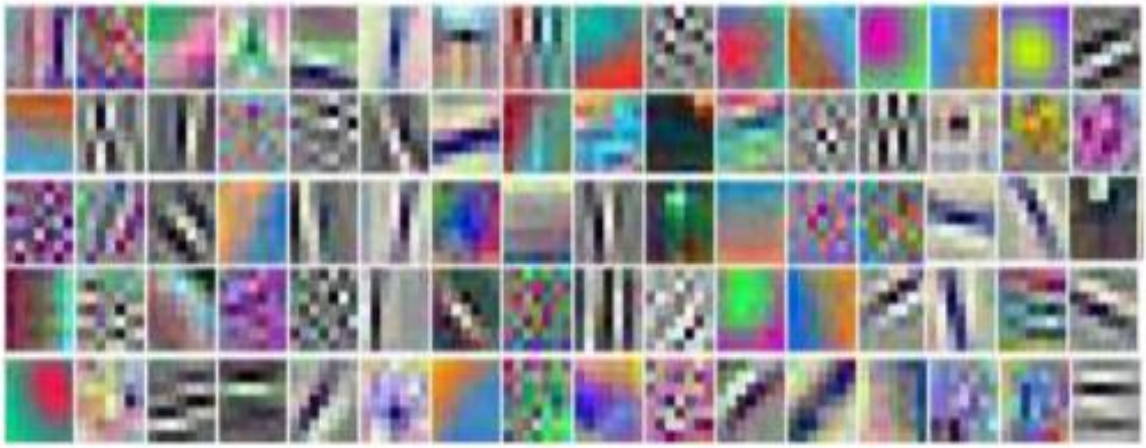

Figure 3. a) First convolutional layer

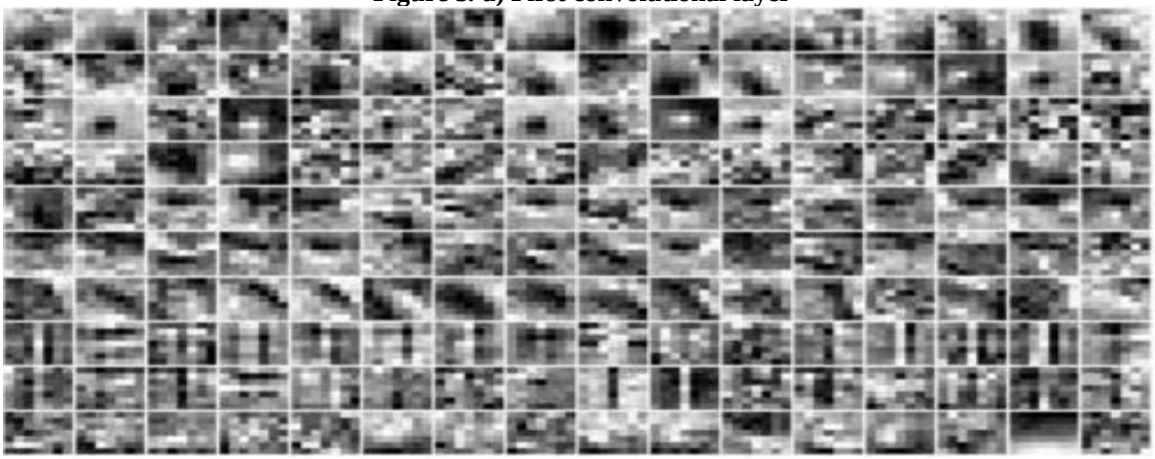

Figure 4. b) Second convolutional Layer

Figure 4. Samples of training image and ground truth bounding boxes a) car b) bike c) person

Figure 5. Predicted bounding box of car

Table 1. Accuracy of the projected scheme compared with Fast R-CNN

Class	Fast R-CNN	Proposed Method
BUS	0.85	0.9
CAR	0.8	0.87
Motor Bike	0.76	0.85

IV. CONCLUSION

We have developed a completely innovative convolutional neural network, that is simple but accurate and efficient. In object detection framework the convolutional features gathered from our system is better than state-of-art image classification network. Our method achieves accuracy by exchanging the flexibility characteristics with a faster R-CNN, both during training and during testing. But our model hasn't considered the noise while the image is being captured [19,20,21]. In future the noise will be consider as a pre-processing step. The proposed model performed well without noise, providing accurate prediction of some test images. Although it is accurate, but that it is not 100% accurate. We hope that our system will benefit from progress in this area.

REFERENCE

1. PF Felzenszwalb, RB Girshick, D Mcallester, and D Ramanan, "Object detection with discriminatively trained part-based models," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, No. 9, pp. 1627-1645, 2010.
2. Y Jia, E Shelhamer, J Donahue, S Karayev, J Long, R Girshick, S Guadarrama, and T Darrell, "Caffe: Convolutional architecture for fast feature embedding," Proceedings of the 22nd ACM international conference on Multimedia, Orlando, Florida, USA, 03-07 November, 2014, pp 675-678.
3. A Krizhevsky, I Sutskever, and GE Hinton, "Imagenet classification with deep convolutional neural networks," Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, 03-06 December, 2012, Vol. 1, pp. 1097-1105.
4. Z Cao, T Simon, S-E Wei, and Y Sheikh, "Real time multi-person 2D pose estimation using part affinity fields," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21-26 July 2017.
5. Z Yang, and R Nevatia, "A multi-scale cascade fully convolutional network face detector," Proceedings of the 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4-8 December 2016.
6. GE Hinton, and RR Salakhutdinov, "Reducing the dimensionality of data with neural networks", Science Journal, Vol. 313, No. 5786, pp. 504-507, 2006.
7. Y Bengio, Learning deep architectures for AI. Journal Foundations and Trends in Machine Learning, Vol. 2, No. 1, pp. 1-127, January 2009.
8. J Redmon, S Divvala, R Girshick, and A Farhadi, "You only look once: Unified, real-time object detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27-30 June 2016.
9. S Ren, K He, R Girshick, and J Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 39, No. 6, pp. 1137-1149, 2017.
10. R Girshick, J Donahue, T Darrell, and J Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, 23-28 June 2014, pp. 580-587.
11. JRR Uijlings, KEA Van De Sande, T Gevers, and AWM Smeulders, "Selective search for object recognition", International Journal of Computer Vision, Vol. 104, No. 2, pp. 154-171, 2013.
12. CL Zitnick, and P Dollár, "Edge boxes: Locating object proposals from edges", Proceedings of the European Conference on Computer Vision, 391-405, 2014.
13. R Girshick, "Fast R-CNN", Proceedings of the IEEE International Conference on Computer Vision (ICCV), Washington, DC, USA, 07-13 December 2015, pp. 1440-1448.
14. K Lenc, and A Vedaldi, "R-CNN minus R", Computer Vision and Pattern Recognition, pp. 1-12, 2015.
15. H Jiang, and EL Miller, "Face detection with the faster R-CNN", Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA, 30 May-3 June 2017, pp. 650-657.
16. YH Byeon, and KC Kwak, "A performance comparison of pedestrian detection using faster RCNN and ACF", Proceedings of the 2017 6th IIAI International Congress on Advanced Applied Informatics, IIAI-AAI 2017, Hamamatsu, Japan, 9-13 July 2017, pp. 858-863.
17. X Zhao, W Li, Y Zhang, TA Gulliver, S Chang, and Z Feng, "A faster RCNN-based pedestrian detection system", Proceedings of the IEEE Vehicular Technology Conference, Montreal, QC, Canada, 18-21 September 2016.
18. MC Roh, and JY Lee, "Refining faster-RCNN for accurate object detection", Proceedings of the 15th IAPR International Conference on Machine Vision Applications, Nagoya, Japan, 8-12 May 2017, pp. 514-517.
19. M Laavanya, and V Vijayaraghavan, "A sub-band adaptive visushrink in wavelet domain for image denoising", International Journal of Recent Technology and Engineering, Vol. 7, No. 5S4, pp. 289-291, 2019.
20. M Laavanya, and M Karthikeyan, "Dual tree complex wavelet transform incorporating SVD and bilateral filter for image denoising", International Journal of Biomedical Engineering and Technology (Inderscience), Vol. 26, No. 3-4, pp. 266-278, 2018.
21. V Vijayaraghavan, M Laavanya, and M Karthikeyan, "Real oriented 2-D dual tree wavelet transform with non-local means filter for image denoising", Journal of Electrical Engineering, Vol. 17, No.2, pp. 106-111, 2017.