

Enhancing Drivable Area Detection: A FCNN and VGG16 Approach for Autonomous Vehicles and ADAS

Subhadip Chandra¹, Kamlesh Kumar Dubey², Gaurav Agarwal³ and Nabanita Chakraborty⁴

¹Institute of Engineering and Management, WB, India

²⁻³Invertis University, UP, India

⁴Meghnad Saha Institute of Technology, WB, India

Email: subhadipchandra@gmail.com, kkdubey4maths@gmail.com, gaurav.a1@invertis.org, nabanita.chak2007@gmail.com

Abstract— In recent days, multiple approaches are developed on autonomous vehicles as well as on Advance Driver Assistance System (ADAS) to minimize road accidents and to secure life. The drivable areas in many developing nations, as well as rural areas, are barely defined and maintained. Several existing approaches attempt to determine the drivable area by analyzing the lane mark and determining the drivable area. Traditional lane detection methods, on the other hand, can be difficult to conduct with a high detection rate in bad weather or in complex road situations. In this paper, we propose a novel method using Fully Convolutional Neural Network (FCNN) and Visual Geometry Group (VGG16) to detect drivable area. We describe a supervised learning strategy for simultaneously recognizing lane mark and drivable region. To begin, we applied a segmentation mask to the annotated image using FCNN and VGG16 to segment the drivable area. On KITTI datasets, we present extensive experimental results for drivable area detection.

Index Terms— Autonomous Vehicle, Road Detection, Advance Driver Assistance Systems, Fully Convolutional Neural Network

I. INTRODUCTION

As per World Health Organization, every year in this world, the lives of approximately 1.3 million people cut-short as a result of a road traffic crash [1]. Between 20-50 million people suffer from non-fatal injuries with many incurring a disability as a result of their injury. And the major reason of these accidents is due to the lack of attentiveness of driver. To minimize this unfortunate scenario from the world, by providing Advance Driver Assistance System (ADAS) comes into account. To implement ADAS or Autonomous Vehicle (AV), system must know where to drive. For doing so many technologies are available like camera sensor, LiDAR, GPS etc. Convolution neural networks (CNNs) have recently been used to the field of computer vision, due to the fast development of graphics processing unit (GPU) technology. CNNs may be trained to learn representative features from photos using many datasets that contain vast amounts of image materials released online, such as ImageNet [2] and KITTI [3].

Using computer vision to detect road surfaces has been the subject of numerous studies. The majority of those ideas centered on enhancing road surface detection from a single image and detecting ground surface with hardware sensors. These methods, however, have a long latency and are insufficient for providing real-time notification to drivers.

In this article authors employed KITTI dataset to detect the drivable region for on-road drivers in this work. Because of its faster detection speed and higher accuracy, also utilized the FCNN model. Researchers confined their task to finding out drivable area of any road using FCNN and VGG 16 in this research work.

II. REVIEW OF PREVIOUS WORK

Over the years many studies have been conducted on the road surface. One of the notable works is done in [4] where the researchers used computer vision and data-driven method to detect the road surface. Authors of [5] used Mask R-CNN to detect inshore battleships and merchant ships. They claimed that their approach is more robust than other approaches. They used Soft Non-Maximum Suppression (Soft-NMS) to improve the detection of the framework. The ship segmentation is also done in their model which shows a good performance. Experiments on multiple objects detection show that [6] improved faster R-CNN exhibits better performance by eradicating some shortcomings of original faster R-CNN.

Prior-knowledge based on feature-based techniques [7, 8] uses known aspects of an object to determine its location, skeleton, or outlines, such as geometric symmetry, shape, and color. However, lanes marking geometric elements are not always present or consistent; they may be obscured by front vehicles, parked cars along the roadside, shadows cast by surrounding vehicles, or surface defects.

A lot of deep network architecture [9, 10] have been developed in the literature to address the challenge of identifying lane mark and drivable region. The R-CNN structure was proposed by Girshick et al. [11], which is an extension of the Faster R-CNN structure [12]. In [13], the v-disparity approach is one of the most used algorithms for determining the driveable area using stereo vision. However, it misidentifies the drivable area as the vegetation along county highways. In order to solve the issue of treating the vegetation along the roadside as a drivable area, this research suggests the drivable area detection method. Compared to the earlier v-disparity method, the new method is roughly 6.86% more accurate, according to the experimental data. In [14], Special rows in the image are chosen for this study. Then, in these rows, the boundaries of the driveable area are found. Thus, there is a huge reduction in computing expenses. The speed of the process reaches +300 frames per second when this model is tested on the Berkley Deep Drive dataset, which is quicker than previous approaches.

III. PROPOSED METHODOLOGY

FCNN is a network that does not contain any “Dense” layers (as in traditional CNNs) instead it contains 1x1 convolutions that perform the task of fully connected layers (Dense layers). A fully convolutional neural network (FCNN) is more feasible for pixel-to-pixel semantic segmentation. Architecture of FCNN is represented in figure 1.

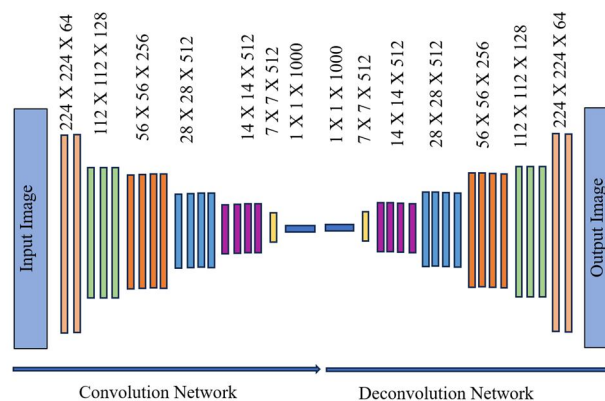


Figure. 1. Architecture of Fully Convolutional Neural Network

It has proven to be capable of working successfully without the usage of any further machinery. FCNN, unlike the classic CNN, which uses the Fully Connected layers after the Convolutional layers in the network, FCNN can take input of the arbitrary size and has a de-convolutional layer or up sample operations. This results in the output image size being exactly the same as the raw input image size. The working model of obtaining segmented image has been presented in figure 2.

In the case of VGG16, it has a uniform architecture. It has millions of parameters that can be used for convenience. It follows this arrangement of convolution and max pool layers consistently throughout the whole architecture. Figure 3 depicts, at the end it has 2 FC (fully connected layers) followed by a SoftMax for output. It has pooling layers which are used to reduce the dimensions of the feature maps. Thus, it reduces the number of parameters to learn and the amount of computation performed in the network. This makes the model more robust to variations in the position of the features in the input image. In figure 3, architecture of VGG 16 is depicted.

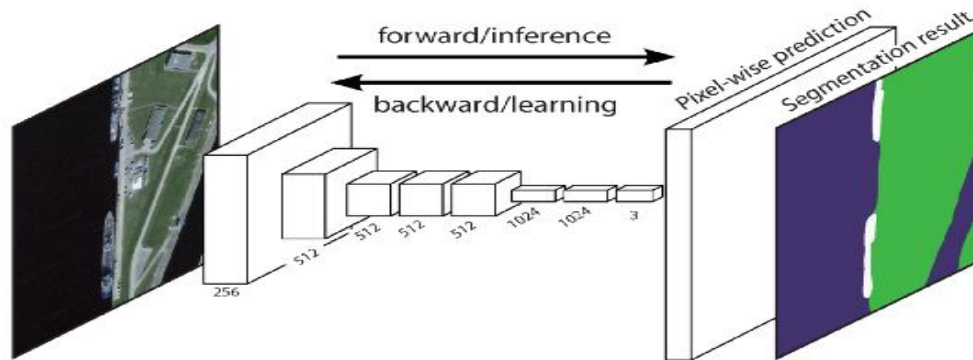


Figure 2. Working model of Segmented image

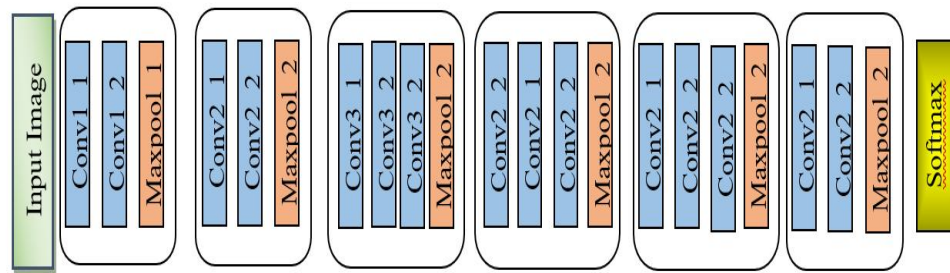


Figure 3. Architecture of Visual Geometry Group (VGG 16)

It is basically a Convolutional Neural Network which is extensively used in image segmentation projects such as visual object recognition research. The name VGG is an abbreviation of Visual Geometry Group, that is, a group of researchers who worked on visual geometry. And the '16' is simply because it contains 16 layers. One major reason to use VGG16 for this work is its accuracy. It showed 92.7% accuracy in top-5 ImageNet, which is an improvement were made on VGG16 by AlexNet. Due to its simplicity, VGG16 is used in deep learning image classification algorithms.

For creating the segmentation mask, we need to use annotations. If we already have the annotations with us, then we do not need any preprocessing of the images in our training dataset. It comes along with ready annotations and segmentation masks. Thereby we can directly head towards the training part.

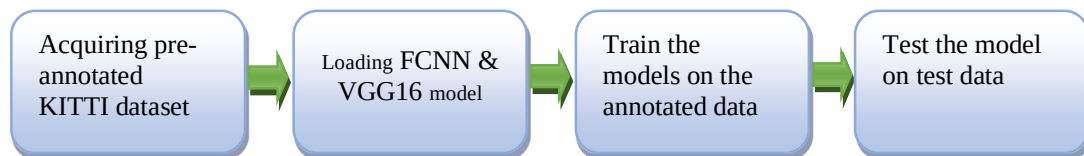


Figure 4. Work flow of the proposed model

The general structure of a semantic segmentation procedure can be conceptualized as an encoder network followed by a decoder network. In general, the encoder is a pre-trained classification network like VGG/ ResNet, followed by the decoder. The Work flow of the proposed model has been represented in figure 4.

In order to achieve dense classification, the encoder learns features (lower resolution) that are semantically projected onto the pixel space (higher resolution) by the decoder. In its original form, the Fully Convolutional Network (FCN) maps pixels to pixels without extracting the regions of interest. The FCN network pipeline, an extension of the classical CNN, is designed to make the classical CNN work on arbitrary sized images.

In contrast to CNNs, FCNs have only convolutional and pooling layers, which enable them to make predictions on inputs of any size. CNNs are restricted to accepting and producing labels on specific sized inputs due to their fully-connected layers which are fixed.

As a result of propagating the output feature maps through several alternated convolutional and pooling layers, the resolution of the feature maps can be reduced considerably. The direct predictions of FCN are usually in low resolution, resulting in relatively fuzzy object boundaries.

Here are the key features of the FCN architecture:

In order to perform semantic segmentation, the FCN transfers knowledge from the VGG16. A class presence heat map is produced using 1x1 convolution, following the conversion of VGG16's fully connected layers into convolutional layers. By transposing the convolutions (initialized with bilinear interpolation filters) to become higher resolution, the semantic feature maps are up sampled. Throughout the up-sampling process, coarser but higher resolution features are added from lower layers in VGG16, thus further reducing the up-sampling process. In the next convolution step, skip connections are introduced to enable the subsequent block to extract more abstract, class-specific features from the previously pooled features.

IV. RESULT

Semantic segmentation is one of the most popular areas where Fully Convolutional Neural Networks come into play. We have used the Tensor Flow module and different Python3 libraries (such as NumPy and SciPy). KITTI Road dataset was used to train the model. This dataset contains a very good number of pre-annotated images of roads. Pre-annotated means, the details of the drivable area (the pixel locations in the image etc.) are given. These are also known as weights, which are used to train the model.

FCNN mainly sends information from VGG16 to do semantic segmentation. The fully connected VGG16 layers form a fully convolutional neural network and in turn, create a low-resolution heat map. These low-resolution semantic heat maps are up sampled using bilinear interpolation filters. This is mainly performed by the lower layers of VGG16. The multiple pooling layers are used to refine the features extracted from each layer.

In this architecture, the pre-trained VGG16 is used as an encoder. The decoding is performed by the layers from layer 7. Padding is used in every layer. The last layer of VGG16 is a 1x1 convolution. By using up sampling from the 11th layer of FCNN, we get the actual image back.

First VGG16 is loaded into Tensor flow, and a GPU-enabled runtime is used to utilize the tensor cores of the same. The layers of FCNN are made in the next step. Here we apply 1x1 convolutions to the encoder and then add decoder layers for the up sampling. Next, we optimize our neural network by using cross entropy and the number of epochs as 10 and batch size as 72. We train the model using load_vgg, layers and optimize functions. Figure 5, shows overall process of getting drivable path using segmentation mask. Figure 5(a) represents the sample raw image of a road with off-road view, figure 5(b) showing the corresponding segmentation mask and figure 5(c) represents the respective annotation. The final result is represented in figure 5(d).



Figure 5. Drivable road segmentation a) Raw image b) Segmented Mask c) Annotation d) Corresponding result

Some sample raw images from training set and corresponding output set are presented in figure 6. Figure 7 shows the loss graph after optimize the neural network by using cross entropy and the number of epochs as 10 and batch size as 72.

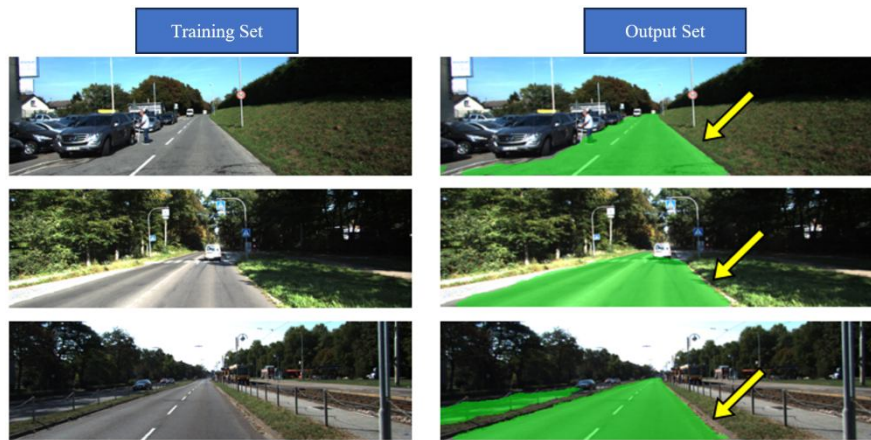


Figure 6. Sample test result

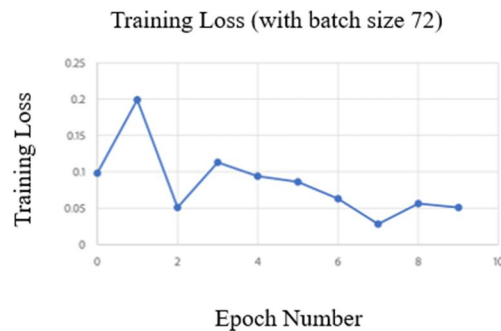


Figure 7. Training Loss graph

V. CONCLUSION

The semantic segmentation technique by the FCN and VGG 16 is used in this research to propose a novel instance segmentation approach for road scene interpretation to find out drivable area. To acquire the precise contour and subsequently collect object classification results to segment the road region thoroughly, an FCN and VGG 16 are used to classify each pixel in the image. Thorough testing shows that the proposed methods not only considerably increase the level of drivable area extracted, but also improve the integrity and accuracy of edge contour detection for each object. Future research will focus on developing a deep instance segmentation architecture to improve multi-object recognition and tracking, segmentation accuracy, quantitative analysis, and comparison of other detection techniques.

REFERENCES

- [1] <https://www.who.int/>
- [2] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition pp. 248-255, June, 2009
- [3] Geiger, A., Lenz, P., & Urtasun, R. Are we ready for autonomous driving? the kitti vision benchmark suite. In 2012 IEEE conference on computer vision and pattern recognition (pp. 3354-3361), June, 2012
- [4] Varadharajan, S., Jose, S., Sharma, K., Wander, L., & Mertz, C. Vision for road inspection. In IEEE Winter Conference on Applications of Computer Vision pp. 115-122, March, 2014
- [5] Nie, S., Jiang, Z., Zhang, H., Cai, B., & Yao, Y. Inshore ship detection based on mask R-CNN. In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium pp. 693-696, July, 2018.
- [6] Zuo, Z., Yu, K., Zhou, Q., Wang, X., & Li, T. Traffic signs detection based on faster r-cnn. In 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW) pp. 286-288, June, 2017

- [7] Yim, Y. U., & Oh, S. Y. Three-feature based automatic lane detection algorithm (TFALDA) for autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 4(4), 219-225, 2003.
- [8] Parajuli, A., Celenk, M., & Riley, H. B. Robust lane detection in shadows and low illumination conditions using local gradient features. *Open Journal of Applied Sciences*, 3(01), 68, 2013.
- [9] Gurghian, A., Koduri, T., Bailur, S. V., Carey, K. J., & Murali, V. N. Deeplanes: End-to-end Lane position estimation using deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* pp. 38-45, 2016.
- [10] Liu, X., & Deng, Z. Segmentation of drivable road using deep fully convolutional residual network with pyramid pooling. *Cognitive Computation*, 10(2), 272-281, 2018.
- [11] Girshick, R., Donahue, J., Darrell, T., & Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* pp. 580-587, 2014.
- [12] Ren, S., He, K., Girshick, R., & Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.
- [13] S. Lee and B. Moon, "Drivable Area Detection Method Capable of Distinguishing Vegetation Area on Country Road," in *2018 International SoC Design Conference (ISOCC)*, Daegu, Korea (South), 2018, pp. 80-81, doi: 10.1109/ISOCC.2018.8649902
- [14] H. Asgarian, A. Amirkhani and S. B. Shokouhi, "Fast Drivable Area Detection for Autonomous Driving with Deep Learning," in *5th International Conference on Pattern Recognition and Image Analysis (IPRIA)*, Kashan, Iran, 2021, pp. 1-6, doi: 10.1109/IPRIA53572.2021.9483535.