



HAL
open science

Driveable Area Detection Using Semantic Segmentation Deep Neural Network

P. Subhasree, P. Karthikeyan, R. Senthilnathan

► **To cite this version:**

P. Subhasree, P. Karthikeyan, R. Senthilnathan. Driveable Area Detection Using Semantic Segmentation Deep Neural Network. 3rd International Conference on Computational Intelligence in Data Science (ICCIDS), Feb 2020, Chennai, India. pp.222-230, 10.1007/978-3-030-63467-4_18 . hal-03434796

HAL Id: hal-03434796

<https://inria.hal.science/hal-03434796v1>

Submitted on 18 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Driveable Area Detection Using Semantic Segmentation Deep Neural Network

P. Subhasree^{1,*}, P. Karthikeyan¹, R. Senthilnathan²

¹Division of Mechatronics, Department of Production Technology, MIT Campus, Anna University, Chennai, India

²Department of Mechatronics Engineering, SRM Institute of Science and Technology, Kattankulathur, India

*subhasreesenthilnathan@gmail.com

Abstract. Autonomous vehicles use road images to detect roads, identify lanes, objects around the vehicle and other important pieces of information. This information retrieved from the road data helps in making appropriate driving decisions for autonomous vehicles. Road segmentation is such a technique that segments the road from the image. Many deep learning networks developed for semantic segmentation can be fine-tuned for road segmentation. The paper presents details of the segmentation of the driveable area from the road image using a semantic segmentation network. The semantic segmentation network used segments road into the driveable and alternate area separately. Driveable area and alternately driveable area on a road are semantically different, but it is a difficult computer vision task to differentiate between them since they are similar in texture, color, and other important features. However, due to the development of advanced Deep Convolutional Neural Networks and road datasets, the differentiation was possible. A result achieved in detecting the driveable area using a semantic segmentation network, DeepLab, on the Berkley Deep Drive dataset is reported.

Keywords: Road detection, Road segmentation, Driveable area detection, Semantic segmentation, Autonomous driving.

1 Introduction

Advanced Driver Assist Systems (ADAS) and autonomous driving technology have greatly contributed to the explosive growth of the field of deep learning which was fueled by the massive collection and availability of road datasets for public usage. Among the many computer vision tasks involved in ADAS and Autonomous driving, road detection has been considered one of the most important topics of research especially in the scope of level 4 and level 5 of driving automation [1]. The road perception algorithms are the first step for any subsequent path planner which aid in autonomous navigation. Remarkable progress has been achieved in road segmentation in the last decade both in feature-based and learning-based approaches. Automatic driving decisions collision-free on the segmentation roads from vehicles. Often high-level information about roads such as ego-lane detection [2], object-lane relationships [3], etc is required for successful cognitive actions ensuring collision-free navigation.

With the advent of deep neural networks and publicly available datasets that aid in explicit labelling of road related features such as drivable area, lane markings, etc there is a tremendous interest in extending the road detection problem to drivable area detection. In fact, much recent research had focused on the detection of flat areas which are considered to be drivable. Such high-level information allow self-driving cars to act very similarly to human drivers. The paper presents the details of the implementation of drivable area detection using a deep neural network developed for semantic segmentation. Drivable area detection involves precise segmentation of ego-lane i.e. the lane of travel of the vehicle on which the camera is mounted (ego-vehicle). The drivable area is segmented based on the availability of the lane markings. The dataset used for training in the research work presented is Berkeley Deep Drive (BDD) Dataset [4].

2 Related Work

Many researches and works have been going on to solve road segmentation problem since the development of deep neural networks. Even before the recent evolution of deep neural networks, road segmentation has been carried out using other computer vision algorithms, but the features learned by DCNNs were incomparably good. In a work, Road images were segmented into parking lanes, sidewalk and into a road using ariel and ground views of the image, with inputs from Camera, GPS, and IMU systems [5]. Caltagirone L et al, in their work, detected road using DCNN with data from camera and LIDAR [6]. They have applied FCN on a subset of the KITTI dataset for the experiment. In another work, a different approach was carried out to detect roads with minimal amounts of labeled data. The work used semi-supervised learning applied to inputs from camera and LIDAR data combined together. Yang X et.al extracted lane marking and detected road using a combination of Recurrent Neural Network and U-net to reduce the propagation error and improve the accuracy of detection [7]. A model based on the conditional random field, which takes in inputs from both camera and LIDAR as random variables, was experimented by Xiao L [8]. Various research has been carried out for road detection and lane marking. The introduction of the Berkeley DeepDrive dataset gave rise to the experimentations on drivable area related research

Semantic segmentation has evolved since the evolution of CNNs had taken heights. First of such state-of-the-art architectures, was developed by Long J et.al [9]. Using the classification network's learned features, the model adapted the classification network into a fully convolutional network and fine-tuned it for semantic segmentation. Unlike [9], which has done end-to-end learning, the model proposed by Noh H et.al, which was built on the classification network VGG-16, integrated deep deconvolution network and proposals prediction [10]. R-CNN model developed by Girshick R et.al applied convolutions on each of the region proposals extracted from the image and classified the regions into labels [11]. This work gave rise to a new branch of research on combining region proposals and convolutions. Convolution and deconvolution on the image during segmentation, due to downsampling and pooling, misses the high-resolution information required for segmentation. In the

model developed by Lin G et.al, to retain the high-level spatial information, residual connections between the convolution and deconvolution layers were given like identity mapping [12]. An interesting work by Luc P combined the semantic segmentation network with an adversarial network. Using adversarial training, the inconsistencies between ground truth and the predicted map were detected and corrected [13]. Then came the set of models that concentrated on dilated convolutions, that helped in getting the local information along with the global information, by increasing the receptive field of the network [14,15,16,17]. One such model is DeepLab which combines deep convolutions, dilated convolutions at different range, and Conditional random fields. This model gained 79.7 percent mIOU on PASCAL VOC-2012 dataset.

3 Details of Dataset

To find a suitable dataset for the intended purpose of driveable area detection, the datasets available for road data were referred. PASCAL-VOC 2012 dataset[18], Microsoft COCO dataset [19], ADE20K dataset are some of the general semantic segmentation datasets. These datasets contain different objects of vast categories, thus helpful in training the network to identify any object present in the image. For some specific objects of interest to be identified by the network (e.g. road, car, signal, etc. for the autonomous driving scenario), the weights of the network can be fine-tuned by training it with objects specific dataset (e.g. Datasets with road objects). Some road-specific datasets available are Cityscapes dataset [20], KITTI road dataset [21], Apolloscape dataset [22], Oxford RobotCar Dataset [23], Berkley Deep Drive dataset [4]. Among these, the BDD dataset was chosen based on the merit of large scale annotations for driveable area segmentation. This information will be helpful in determining vehicle localization on the road with image data and in turn be helpful in decision making. Though few other datasets have annotations done for road segmentation, these were non-standard and very small in size. BDD dataset is promising in the scope of its attributes such as diverse nature, a sheer large number of annotations of images specifically made for driveable area detection. The dataset has images taken from a variety of geographical locations, environmental and weather conditions. The dataset has labels , “Directly Drivable Area”, Alternatively Drivable Area”. “Directly Drivable Area” is the area that the driver is currently driving on. The driver has priority of this area over the other cars. “Alternately Drivable Area” is the area that the driver is not currently driving on , but can drive on it by changing lane. The dataset contains 91626 instances of Directly drivable area and 88392 instances of Alternatively drivable area. The annotation is given as a mask image which contains pixel level labels for drivable area, alternative area and the background. The dataset is also annotated for lane markings, object detection and instance segmentation. For all the work reported in this paper, the BDD dataset images with annotations for driveable area detection alone are utilized. The driveable area annotations consist of a pair of images for each example. One image is the raw RGB image of resolution 1280×720 pixels and the other image is the annotation image of same size which is also an RGB image. The annotation image has the ego-lane pixels in red, other lanes marked

in blue and all other regions of the image in black. A sample image pair is shown in Fig. 1.



Fig. 1. A Sample Example of Annotation for Driveable Area Detection in BDD Dataset

4 Deep Learning Network

The work presented in the paper is about the implementation details of applying a popular semantic segmentation network for the purpose of driveable area detection. Though conceptually driveable map inferencing is similar to the general semantic segmentation problem, the challenges involved in the former is very high. This is mainly because the classification is between two regions which are exactly identical in terms of all visual properties. The only differentiating element is the lane marking which is assumed to be of contrasting the color of the driveable area (roads). Deep convolutional neural networks have been the choice for various classification tasks in the computer vision system mainly attributed to the built-in invariance to local image transformations. But this property doesn't meet the requirements of dense inferences such as semantic segmentation where abstraction of information in the image is not a requirement. Among the Deep CNN based architectures available for semantic segmentation in the current work reported in the paper, DeepLab [24] is identified to be suitable for the intended purpose of driveable area detection. This is mainly because of its key property such as the good resolution of the features, good ability to tackle the problem of objects' existence at multiple scales and high localization accuracy which, in a deep CNN framework is difficult to obtain due to the invariance property. DeepLab enjoys these merits by strategically incorporating significant changes in the architecture. Some of the important changes introduced in the DeepLab which not only has benefited the semantic segmentation problem but also to a higher level of complexities such as the driveable area detection problem as follows:

- Atrous convolution instead of convolution with downsampled filters: Generally CNNs would have a series of convolutional layers interleaved with pooling layer resulting in downsampling of filters. Atrous convolution does convolution after the last few pooling layers with upsampled filters instead of downsampled filters as in the case of regular convolution. This strategy basically improves the feature resolution which is very valuable in pixel-level inferences as in the case of driveable area detection. It may be noticed that the atrous convolution serves as a valuable replacement for the

deconvolutional layers which is commonly found in most semantic segmentation networks [25].

- Atrous spatial pyramid pooling: To address the problem of scale invariance, generally images of objects acquired at many scales are used. DeepLab utilizes a pyramid pooling strategy which is a computationally efficient way of ensuring scale invariance. The strategy involves the usage of multiple filters that have complementary fields of view. This ensures capturing of objects as well the context in multiple scales resulting in the scale invariance property.
- Fully-connected CRFs: The property of accurate localization is incorporated through the use of fully connected pairwise conditional random fields.

A high-level block diagram of the network architecture is shown in Fig. 2.

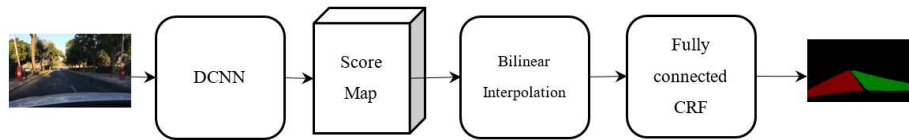


Fig. 2. Abstract Block Diagram of DeepLab Network

Beyond the specified advantages that DeepLab specifically addresses some important advantages that are relevant to the driveable detection problem from a practical stand-point are speed, accuracy, and simplicity of the network. Since the intended application is for a driving scenario real-time nature, accuracy and speed are primary concerns for any segment of algorithms. The detailed block diagram of the DeepLab network is presented in Fig. 3.

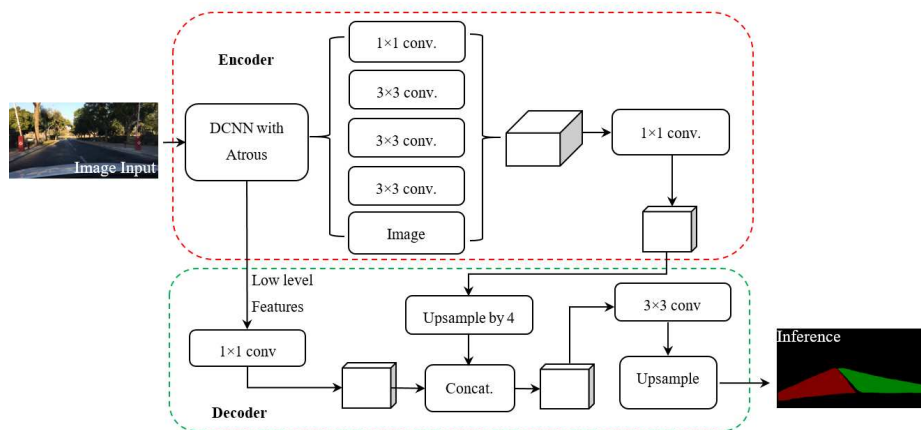


Fig. 3. Detailed Block Diagram of DeepLab Network

The implementation of DeepLab network for the driveable area detection is carried out with Tensorflow API. Tensorflow provides pre-trained models of DeepLab which

are meant for semantic image segmentation of a wide variety of classes. The network is pre-trained on PASCAL-VOC 2012 dataset. The weights of the pre-trained model which are available in a frozen inference graph is used before the training is initiated on the BDD driveable dataset. All the frozen inference graphs by default use an output stride of 8. The backbone of the DeepLab utilized in this work is Xception_65 [26]. The block diagram of a single block of a DeepLab network is presented in Fig. 4. This network is pre-trained on datasets such as ImageNet, MS-COCO part for VOC. As mentioned earlier all the training had been only for generic semantic image segmentation. To ensure robustness the network is initialized with pre-trained weights from PASCAL-VOC. The training is carried out in a phased manner to understand the convergence of the model The key specifications of the training are listed in Table 1.

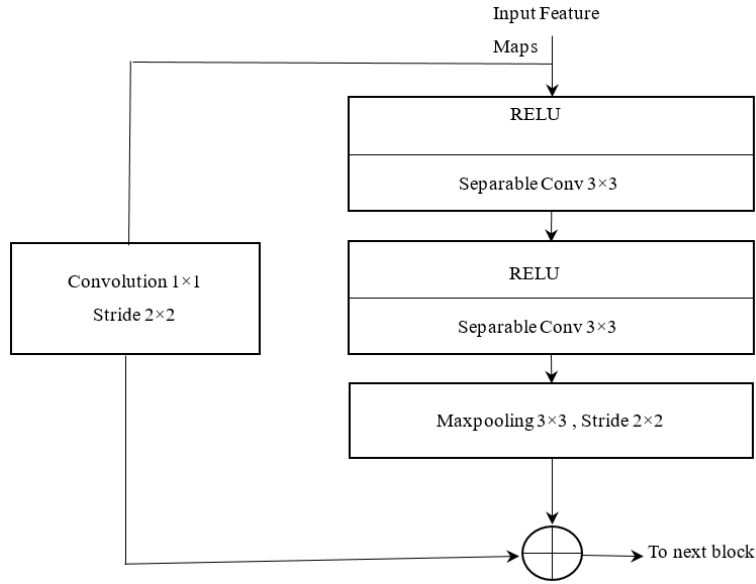


Fig. 4. Xception-65 Block at Entry Flow

Table 1. Key Training Specifications

Parameter	Specification
Number of iterations	2000
Atrous Rates	6, 12, 18
Output Stride	16
Decoder Output Stride	4
Train crop size	513
Train batch size	4
Fine tuning of batch normalization	Yes

5 Results and Conclusion

In the lines of most semantic segmentation benchmarking, mIoU (mean intersection of union) is used as the evaluation metric to understand the performance of the network. Other evaluation metrics are not presented within the scope of this paper. Sample images from the validation set and the corresponding inference images, containing three classes viz., driveable area, alternately drivable area, background, are presented in Fig. 5. The intersection of union is measured between the inference images and the ground truth in the dataset [19]. IoU for the first 48 images are presented in Fig. 6. The mean IoU are 0.9379145 and 0.93885483 for 3 lakhs and 4 lakhs steps respectively. A step here is (Total number of training images / Batch size)

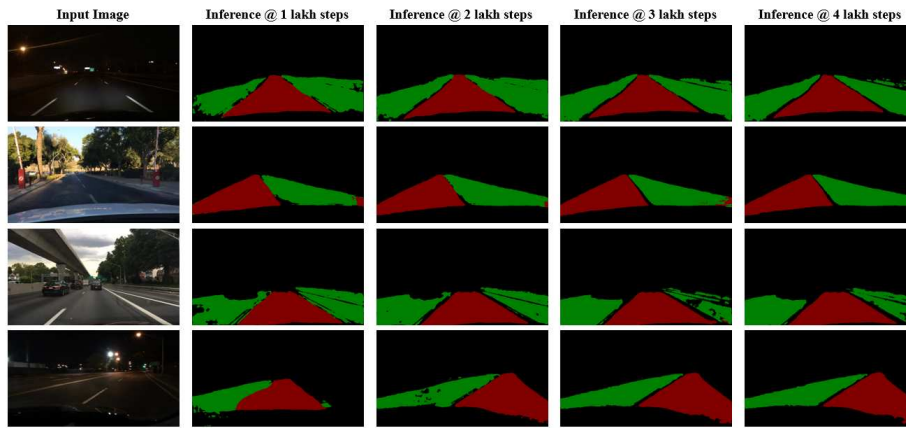


Fig. 5. Sample Segmentation Results

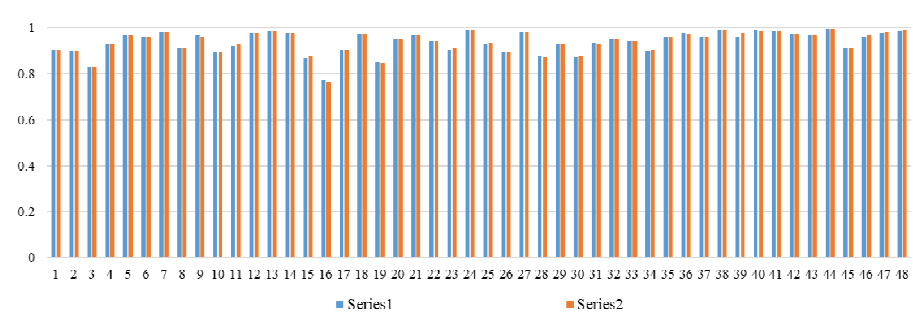


Fig. 6. IoU for first 48 Images from Validation Set

The IoU for driveable area detected using DeepLab semantic segmentation network is given in the paper. As part of the future work, driveable area detection would be carried out on state-of-the-art semantic segmentation networks and compared. And extraction of high level semantics useful for autonomous driving would be carried out using the detected driveable area.

References

1. Levinson, Jesse, et al. : Towards fully autonomous driving: Systems and algorithms. In: IEEE Intelligent Vehicles Symposium (IV) 2011. IEEE (2011).
2. Kim, Jiman, and Chanjong Park.: End-to-end ego lane estimation based on sequential transfer learning for self-driving cars. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.(2017)
3. Ohn-Bar, Eshed, and Mohan Manubhai Trivedi.: Are all objects equal? Deep spatio-temporal importance prediction in driving videos. Pattern Recognition 64 , 425-436 (2017)
4. Yu, Fisher, et al.: BDD100K- A Diverse Driving Video Database with Scalable Annotation Tooling. In: arXiv (2018).
5. Mátyus G, Wang S, Fidler S, Urtasun R. Hd maps.: Fine-grained road segmentation by parsing ground and aerial images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 3611-3619 (2016)
6. Caltagirone L, Bellone M, Svensson L, Wahde M.: LIDAR-camera fusion for road detection using fully convolutional neural networks. Robotics and Autonomous Systems. 111, 25-31. (2019)
7. Yang X, Li X, Ye Y, Lau RY, Zhang X, Huang X.: Road Detection and Centerline Extraction Via Deep Recurrent Convolutional Neural Network U-Net. IEEE Transactions on Geoscience and Remote Sensing. 57.9, 7209-20. (2019)
8. Xiao L, Wang R, Dai B, Fang Y, Liu D, Wu T.: Hybrid conditional random field based camera-LIDAR fusion for road detection. Information Sciences. 432, 543-58. (2018)
9. Long J, Shelhamer E, Darrell T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition . 3431-3440. (2015)
10. Noh H, Hong S, Han B.: Learning deconvolution network for semantic segmentation. In: Proceedings of the IEEE international conference on computer vision. 1520-1528. (2015)
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. 580-587. (2014)
12. Lin G, Milan A, Shen C, Reid I.: Refinene - Multi-path refinement networks for high-resolution semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 1925-1934. (2017)
13. Luc P, Couprie C, Chintala S, Verbeek J. :Semantic segmentation using adversarial networks. arXiv preprint arXiv:1611.08408. (2016)

14. Wang P, Chen P, Yuan Y, Liu D, Huang Z, Hou X, Cottrell G.: Understanding convolution for semantic segmentation. In: IEEE winter conference on applications of computer vision. 1451-1460. (2018)
15. Yu F, Koltun V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122. (2015)
16. Wei Y, Xiao H, Shi H, Jie Z, Feng J, Huang TS.: Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 7268-7277. (2018)
17. Zhao H, Qi X, Shen X, Shi J, Jia J.: Icnet for real-time semantic segmentation on high-resolution images. In: Proceedings of the European Conference on Computer Vision. 405-420. (2018)
18. Everingham M, Eslami SA, Van Gool L, Williams CK, Winn J, Zisserman A.: The pascal visual object classes challenge: A retrospective. International journal of computer vision. 111.1, 98-136. (2015)
19. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL.: Microsoft coco: Common objects in context. In: European conference on computer vision. Springer, Cham. 740-755. (2014)
20. Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S, Schiele B.: The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 3213-3223. (2016)
21. Geiger A, Lenz P, Stiller C, Urtasun R.: Vision meets robotics: The kitti dataset. The International Journal of Robotics Research. 32.11, 1231-7. (2013)
22. Huang X, Wang P, Cheng X, Zhou D, Geng Q, Yang R.: The apolloscape open dataset for autonomous driving and its application. arXiv preprint arXiv:1803.06184. (2018)
23. Maddern, Will, et al.: 1 year, 1000 km- The Oxford RobotCar dataset. The International Journal of Robotics Research. 36.1 , 3-15 (2017)
24. Chen, Liang-Chieh, et al.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. In: IEEE transactions on pattern analysis and machine intelligence 40.4, 834-848. (2018)
25. Chen LC, Papandreou G, Schroff F, Adam H.: Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587. (2017).
26. Chollet, François.: Xception: Deep learning with depthwise separable convolutions. arXiv preprint: 1610-02357(2017)