Deep Neural Network Algorithms for Heavy Duty Truck Applications

FORD OTOSAN

Sabancı Üniversitesi — FACULTY OF ENGINEERING AND NATURAL SCIENCES

Students
Oğuzhan İlter
Asil Örgen

Faculty Member
Mustafa Ünel

Company Advisors
Berzah Ozan
Eren Aydemir

ABSTRACT

Self-driving technology aims to enable vehicles to safely operate without human drivers. Autonomous vehicles consist of various modules such as perception, localization, cognition and motion control. In the project, we have focused on visual perception to partition images into multiple segments/regions and classify them semantically, i.e. semantic segmentation. The aim is to improve the performance of the currently available segmentation methods for off-road areas and to build a real-time system which can provide satisfactory segmentation performance in unstructured environments. Segmented images reveal roads and obstacles which can be used by other systems to operate the vehicle autonomously in challenging terrains. In the scope of this project, we implemented and tested various Convolutional Neural Network (CNN) algorithms and quantified their accuracies on off-road areas. To improve the segmentation accuracy, we extended the publicly available Freiburg Forest Dataset [1] with new images provided by Ford Otosan. Results demonstrate that the BiSeNet algorithm [4] with superpixel post-processing achieves 89 % accuracy on test images.

INTRODUCTION

- According to American Trucking Associations (ATA), the number of qualified truck drivers is significantly dropped due to harsh working conditions. In addition, trucks carry 71% of all the freight tonnage in the U.S. [2].
- Autonomous vehicles provide increased safety by preventing human-based accidents and reduce labor especially in hard working conditions as in construction sites.
- Autonomous driving is deeply researched by both academia and industry-leading companies such as Ford, Tesla and Google.
- In this project, we have been working on a perception module to classify and locate road, sky, side of the road, non-drivable area and obstacles for cognition and motion control algorithms that generate a safe path for the autonomous system.
- Recent developments show that Convolutional Neural Network (CNN) architectures provide better semantic segmentation results in challenging environments.
- To satisfy the project requirements, our objectives are:
    - Design and development of CNN algorithms for segmenting off-road areas
    - Dataset generation for training and testing the developed algorithms in various scenarios
    - Testing CNN algorithms gradually, first on the constructed validation image set, and then on the images taken in real time

PROJECT DETAILS

The most promising CNN structures implemented in the project are AdapNet [3], BiSeNet [4], Deeplabv3+ [5], and SegNet [6], and they are tested on Freiburg Forest Dataset. Among these algorithms, the BiSeNet algorithm was selected because of its high accuracy and real-time performance. A superpixel post-processing method is utilized to increase classification accuracy of the BiSeNet algorithm.

BiSeNet Architecture

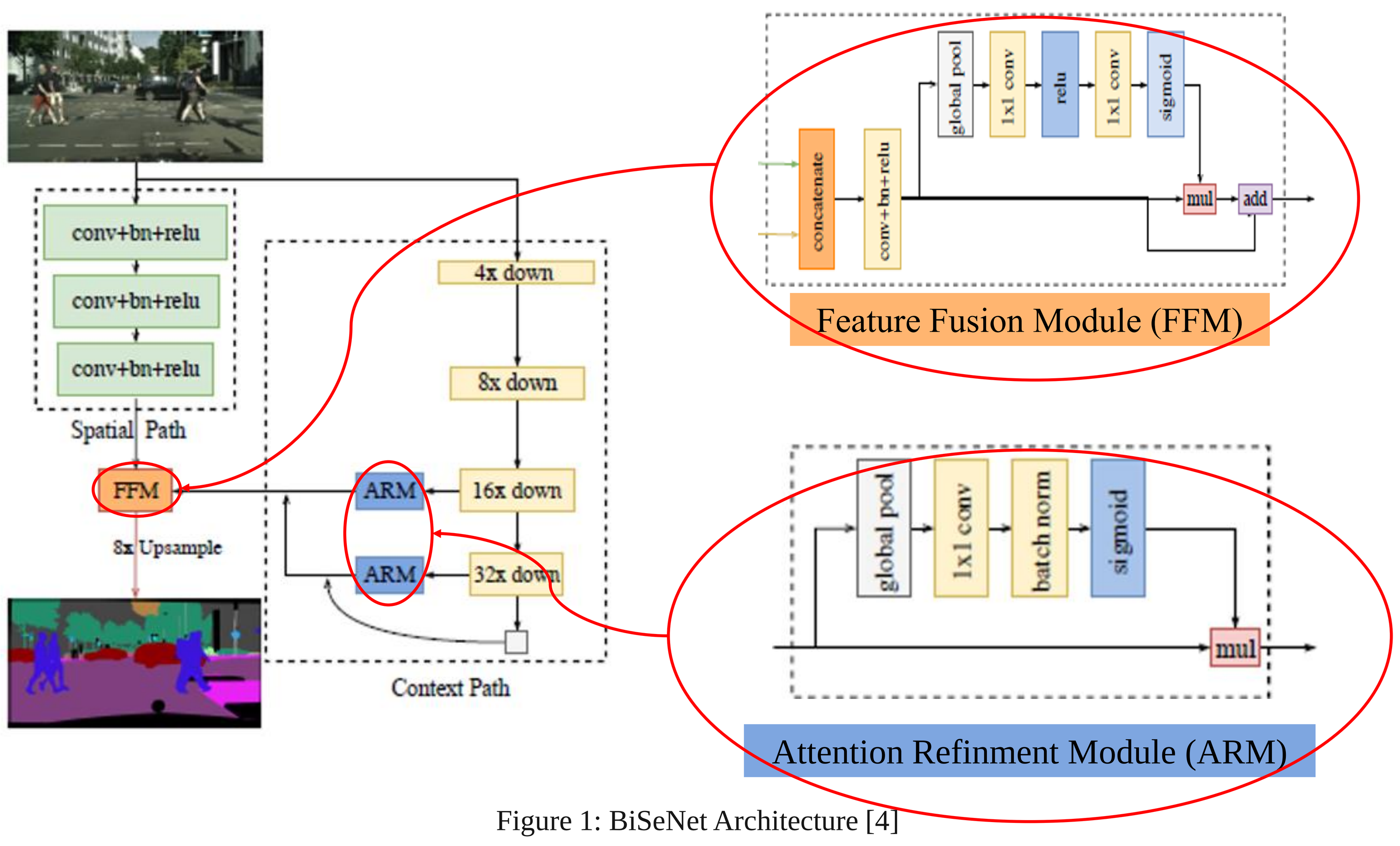


Figure 1: BiSeNet Architecture [4]

- Three dilated convolution layers with stride rate 2, and filter size 3x3 together with ReLU activation function and batch normalization are used to extract spatial features in the spatial path.
- Spatial information is conserved by applying padding in spatial path to keep input size of the layers the same.

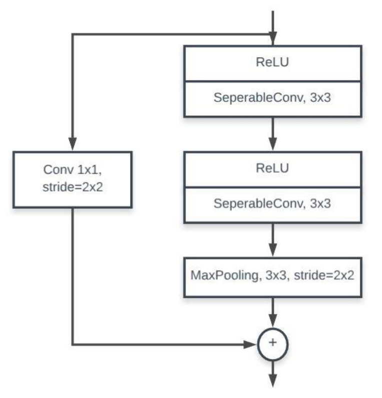


Figure 2: Xception [7] block

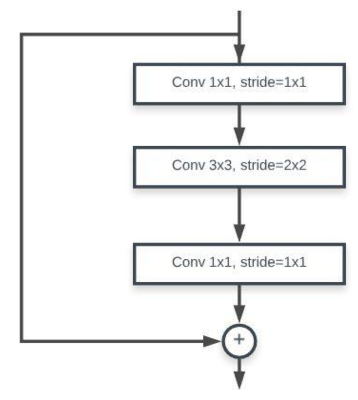


Figure 3: ResNet [8] block

- Context path can be designed using either Xception block (Figure 2) or Resnet block (Figure 3) that provide down sampling together with trainable parameters.
- The network is designed with these two blocks that are aligned consecutively to obtain required down sampling ratio.

Dataset Generation

- Implemented networks were tested on Freiburg Forest Dataset. This dataset was extended with new images provided by Ford Otosan.
- Since the boundaries of regions in the image are not polygon, superpixel method was used to increase labeling the accuracy, precision and speed.
- We modified and used Segment Annotator [9] class definitions and member functions for labeling the dataset.

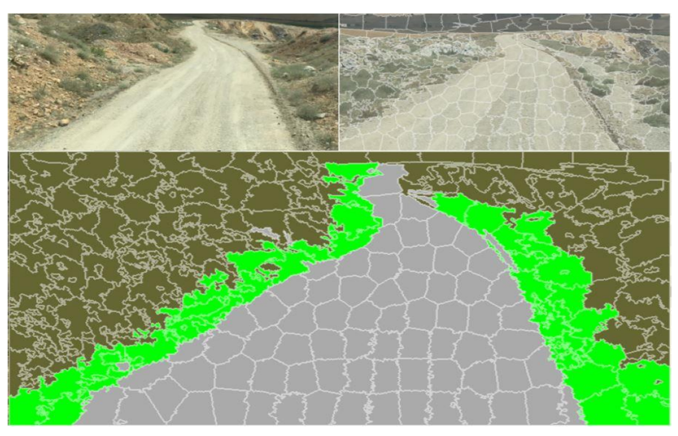


Figure 4: Superpixel demonstration

Post-processing Method

- Superpixel post-processing aims to increase the smoothness of the predicted image by comparing it with the input image.
- The method recovers the lost spatial features by creating regions in the original image. These regions are arranged such that they cover pixels from a single class.
- By comparing pixel classes in a region and classes of adjacent regions in the predicted image, the algorithm determines mislabeled pixels and corrects them.

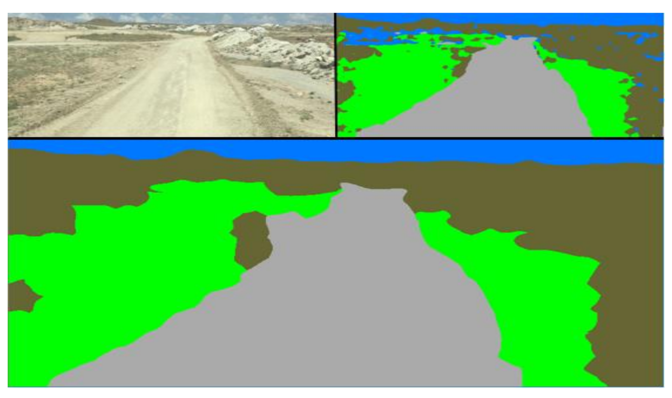
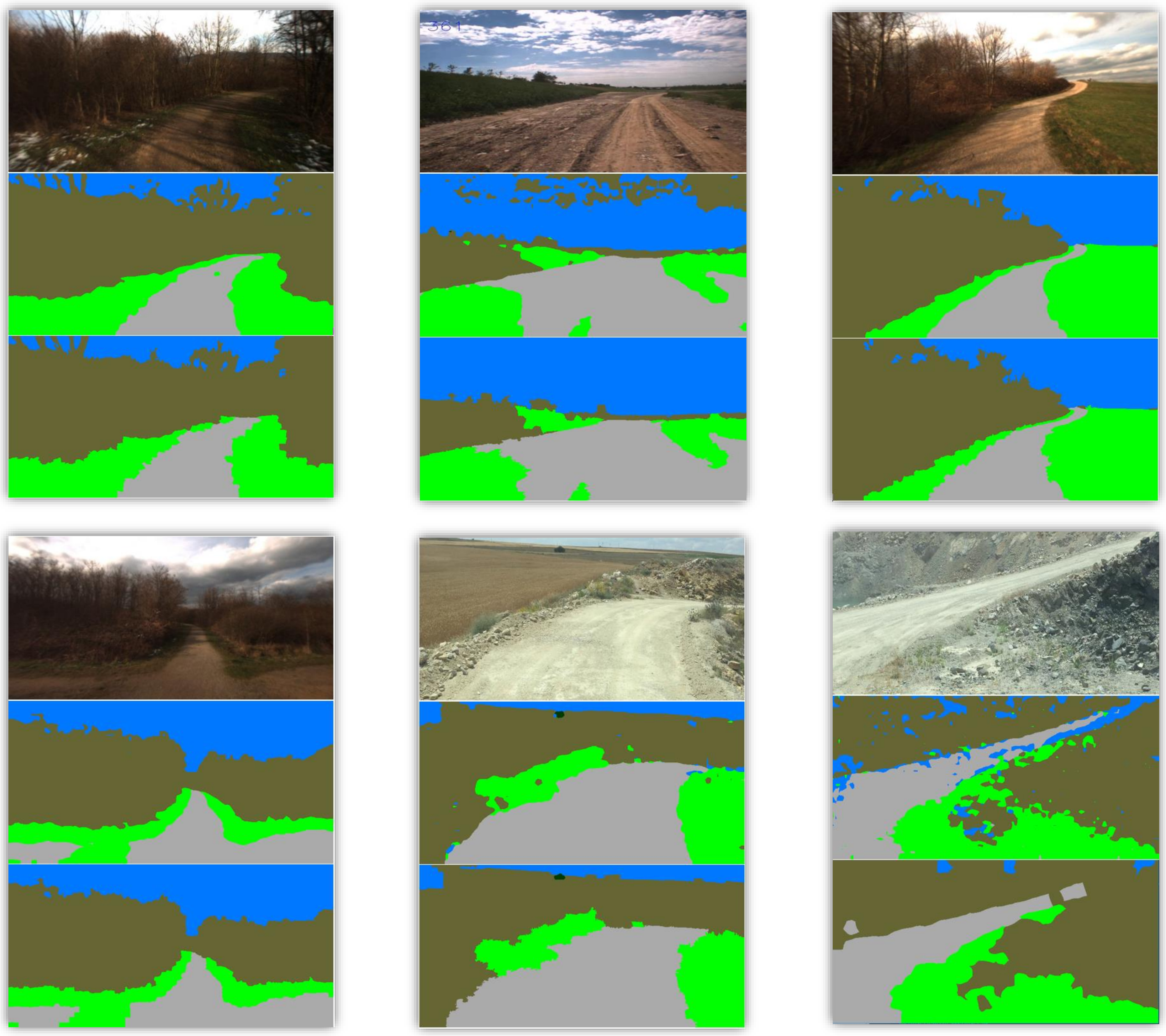


Figure 5: Predicted (top right) and post-processed (bottom)

RESULTS



Non-drivable Area · Sky · Side of the Road · Trail (Road) · Tree · Obstacle

Figure 6: Real images (1st row), BiSeNet with Xception modules (2nd row), BiSeNet with Xception modules + Post-processing (3rd row)

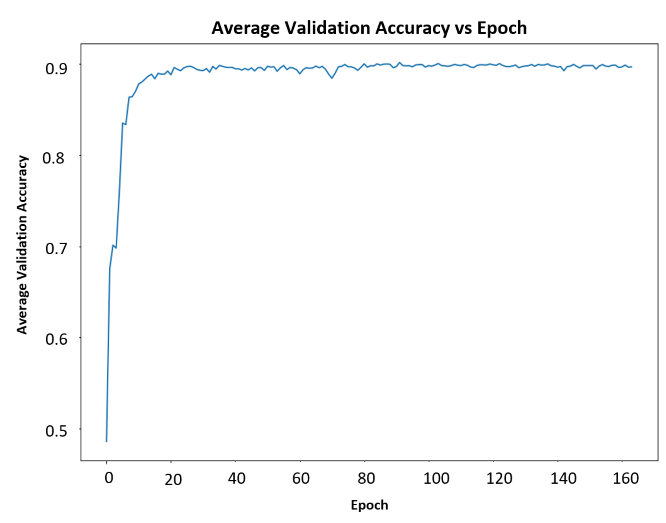


Figure 7: BiSeNet with Xception modules accuracy plot

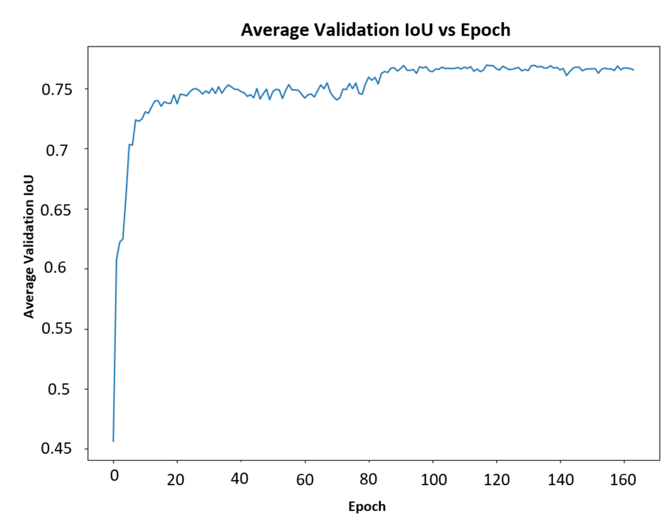


Figure 8: BiSeNet with Xception modules IoU plot

	BiSeNet with ResNet Modules	BiSeNet with Xception modules	BiSeNet + Xception + Post-processing
Accuracy (%)	87.9	88.3	89.3
IoU (%)	72.5	73.7	74.7

Table 1: Accuracy and Intersection over Union (IoU) metrics for different methods

DISCUSSION & CONCLUSION

- Publicly available Freiburg Forest Dataset was enriched with new images provided by Ford Otosan.
- As seen from above images, the BiSeNet algorithm achieves remarkable semantic segmentation performance.
- Abrupt color intensity changes might cause misclassification on large segments.
- Pretrained ResNet and Xception modules provide a sharp accuracy increase in the first 20 epochs.
- As seen from Table 1, the best results in terms of accuracy and Intersection over Union (IoU) metrics have been achieved with post-processing. BiSeNet + Xception + Post-processing achieves 89 % accuracy.
- Xception modules leverage the performance of the BiSeNet architecture compared to ResNet modules.

FUTURE WORKS

- Testing the algorithm in real environments where the real-time performance will be significantly affected by the test setup.
- Optimizing the accuracy and the real-time running performance
- CNN with superpixel pooling layer

REFERENCES

[1] Abhinav Valada et al. “Deep Multispectral Semantic Scene Understanding of Forested Environments using Multimodal Fusion". In: The 2016 International Symposium on Experimental Robotics (ISER 2016). Tokyo, Japan, Oct. 2016. url: http://ais.informatik.uni-freiburg.de/publications/papers/valada16iser.pdf.

[2] American Trucking Associations, “Reports, Trends & Statistics”, 2016. Available: https://www.trucking.org/News_and_Information_Reports_Driver_Shortage.aspx. [Accessed 2019]

[3] Abhinav Valada et al. “AdapNet: Adaptive semantic segmentation in adverse environmental conditions". In: 2017 IEEE International Conference on Robotics and Automation (ICRA) (2017), pp. 4644{4651.

[4] Changqian Yu et al. “BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation". In: CoRR abs/1808.00897 (2018). arXiv: 1808.00897. url: http://arxiv.org/abs/1808.00897.

[5] Liang-Chieh Chen et al. “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation". In: CoRR abs/1802.02611 (2018). arXiv: 1802.02611. url: http://arxiv.org/abs/1802.02611.

[6] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation". In: CoRR abs/1511.00561 (2015). arXiv: 1511.00561. url: http://arxiv.org/abs/1511.00561.

[7] Fran¸cois Chollet. “Xception: Deep Learning with Depthwise Separable Convolutions”. In:arXive-prints, arXiv:1610.02357 (Oct. 2016), arXiv:1610.02357. arXiv:1610.02357 [cs.CV].

[8] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In:arXiv e-prints, arXiv:1512.03385(Dec. 2015), arXiv:1512.03385. arXiv:1512.03385 [cs.CV].

[9] Pongsate Tangseng, Zhipeng Wu, and Kota Yamaguchi. “Looking at Outfit to Parse Clothing”. In:(Mar. 2017). arXiv:1703.01386v1[cs.CV]. url:http://arxiv.org/abs/1703.01386v1.