 DOI: 10.3217/978-3-85125-663-5-48

Semantic Image Segmentation using Convolutional Neural Nets for Lawn Mower Robots

Armin Pointinger[1] and Gerald Zauner[2]

Abstract— **Robots are becoming more and more part of our daily lives. They take on different tasks to make our everyday life easier. In order to be able to fulfill these tasks expediently, high demands are placed on the robots with regard to their abilities. Accordingly, lawn mower robots are also expected to achieve a perfect mowing result and ease of handling. To do this, the robot must be able to find its way around and be able to react appropriately.**

I. INTRODUCTION

Within a few years, semantic image segmentation has become a key task in image processing. This rapid progress already allows a paradigm shift in many areas with regard to the solution approach of many problems. Thus, it is obvious to use semantic image segmentation for autonomous lawnmowers. The resulting benefits are good orientation abilities in previously unseen environment, optimal path planning and the reduction of danger to humans and animals. Compared with conventional lawnmower robots, whose navigation usually relies on a perimeter wire, this could be an alternative in future.

This master's thesis deals with the comparison of different network architectures for semantic segmentation with respect to their suitability for use in autonomous lawn mowers. Sufficient segmentation accuracy and real-time capability are used as criteria for this.

II. APPROACH

By using different network architectures their advantages and disadvantages are evaluated. TensorFlow was used as framework and the implementation of the models is available on the following link on GitHub. [2] Extensive data material is essential for Deep Learning. No data set was available for this task and had to be created first. In order to keep the effort manageable, a two-class problem was assumed. One class is represented by the lawn and one class by the environment. In order to achieve good results, different hyper parameters are optimized manually and automatically during training. Using these results, performance tests were done on different hardware platforms.

A. Dataset

A data set was created which was used for this problem. The images were taken with an RGB camera from the perspective of a lawn mower robot at a distance of 26 cm from the ground. The original resolution of 4032 x 3024 pixels was reduced to 1280 x 704 pixels. Mainly private gardens represent the 25 different locations, which were also selected with regard to different lawn conditions. For each of these individually acquired images the ground truth was created as a label and a data set was generated, which divided the image data as follows.

- 860 Training images
- 89 Validation images
- 91 Test images

B. Training

Among other things, the models used have been selected in a way of providing a comparison of new, older, more complex and simpler architectures. In order to achieve good training results, the hyper parameters need to be adjusted. Therefore empirical attempts were made. The respective models require different hyper parameters for good results. The training process was done with a NVIDIA GeForce GTX 1080. No algorithms for data augmentation were used.

III. EXPERIMENTAL RESULTS

ResNet-101 was used as base model so that the test results could be compared as closely as possible. Table I shows used models and the related per class mean IoU (intersection over union). The result shows that all models provide very similar results in terms of accuracy. The average run time refers to one frame and was calculated with a NVIDIA GeForce GTX1080.

TABLE I
TEST RESULTS

Model	Mean IoU	Average Run Time
DeepLabV3+ [3]	0.966	0.073 sec
BiSeNet [6]	0.955	0.088 sec
DenseASPP [5]	0.952	0.037 sec
GCN [4]	0.959	0.098 sec

The aim is to obtain a system with real-time capability. This requires appropriate computing power. Table II lists three different hardware systems with their achieved inference speed. This measurement was performed with the DeepLabV3+ model. In order to make inference also work with the Jetson TX2, the image resolution was reduced to 640 x 352 pixels. This resolution was used for all three systems. It was observed that the mean IoU (intersection over union) is slightly lower than with the resolution of 1280 x 704 pixels, which was used for training. Since only temporal aspects are

This work was supported by the FH Oberösterreich Forschungs & Entwicklungs GmbH and the Ginzinger Electronic Systems GmbH.

[1]Armin Pointinger, University of Applied Sciences Upper Austria, 4600 Wels, Austria armin.pointinger@gmx.at

[2]Gerald Zauner, University of Applied Sciences Upper Austria, 4600 Wels, Austria gerald.zauner@fh-wels.at

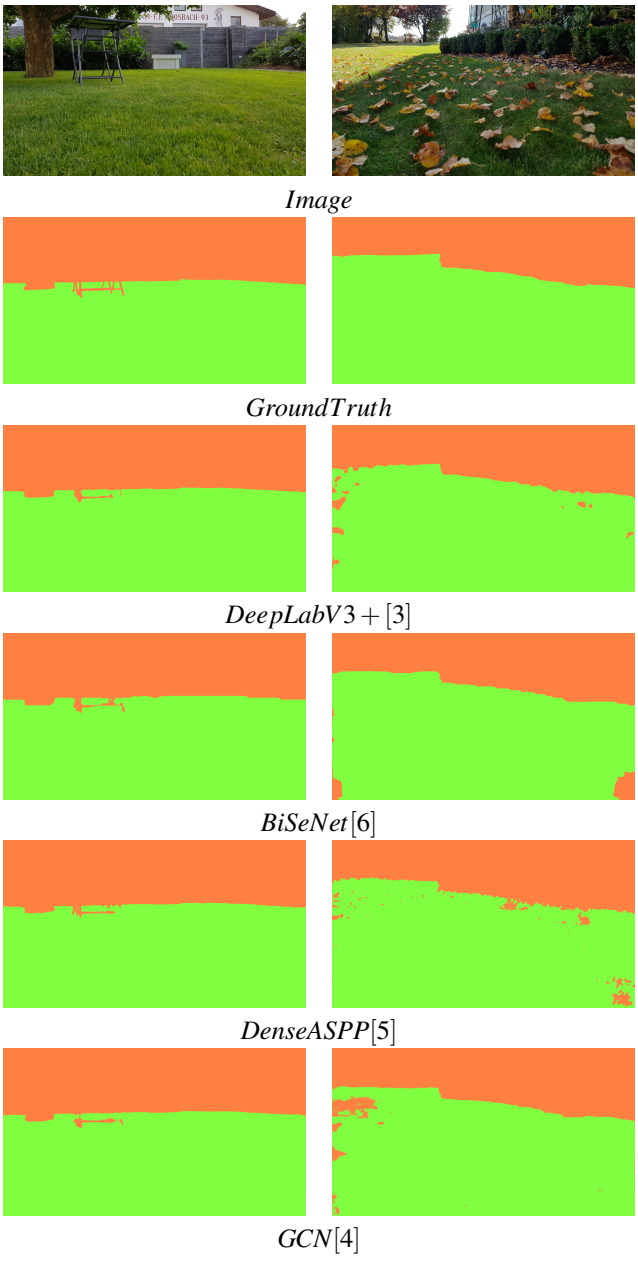
listed in table II, it should be noted that the labels generated by the Jetson TX2 did not contain useful information.

TABLE II
RUN TIME

Hardware	Average Run Time	Frames per Second
GeForce GTX1080	0.030 sec	33.3
Quadro P2000	0.085 sec	11.8
Jetson TX2	0.403 sec	2.48

Figure 1 shows two different test images, their ground truth and the predicted labels of different models. This Link can be used to watch a video about inference. [1]

Fig. 1. Test Images



IV. CONCLUSIONS

Good results were achieved with all tested models. In addition, the data set used is very small and should be extended for better results. In order to guarantee real-time capability, high resolution images require high computing power. The performance of a Jetson TX2 module is too low for this task, and the price for more powerful hardware is currently too high to compete with conventional consumer lawn mower robots. Nevertheless, semantic image segmentation provides lawn mower robots a good basis for terrain orientation and lawn recognition.

ACKNOWLEDGMENT

This work was supported by the FH Oberösterreich Forschungs & Entwicklungs GmbH and the Ginzinger Electronic Systems GmbH.

REFERENCES

[1] (2019) Inference video. [Online]. Available: https://youtu.be/GPCfSAO0TYc
[2] (2019) Semantic segmentation suite. [Online]. Available: https://github.com/GeorgeSeif/Semantic-Segmentation-Suite
[3] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," *ECCV*, 2018.
[4] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters - improve semantic segmentation by global convolutional network," 2017.
[5] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "Denseaspp for semantic segmentation in street scenes," *CVPR*, 2018.
[6] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Bisenet: Bilateral segmentation network for real-time semantic segmentation," *ECCV*, 2018.