Self-Supervised Online Training of FCNs for Free-Space Detection

Willem P. Sanberg Gijs Dubbelman Peter H.N. de With
VCA Research Group, Eindhoven University of Technology, the Netherlands
w.p.sanberg@tue.nl

Abstract

Recently, vision-based Advanced Driver Assist Systems have gained broad interest. In this work, we investigate free-space detection in front of the driving vehicle, for which we propose to employ a Fully Convolutional Network (FCN). We show that this FCN can be trained in a self-supervised *manner and achieve similar results compared to training on manually annotated data, thereby reducing the need for large manually annotated training sets. Additionally, our self-supervised training facilitates* online *training of the FCN instead of offline, so that the free-space analysis becomes highly adaptive to any traffic scene that the vehicle encounters. We have validated our algorithm using publicly available data and on a new challenging benchmark dataset, showing that the online training boosts performance with* 5% *over offline training.*

1. Introduction

In recent years, vehicles are becoming increasingly intelligent with so-called Advanced Driver Assistance Systems (ADAS). This development is expected to reduce traffic accidents, traffic congestion and fuel consumption, while increasing travelling comfort simultaneously. A key component in such an embedded real-time system is free-space detection: determining where the vehicle can and cannot drive in a cost-effective manner. Additionally, ADAS have to be flexible and robust, since traffic scenes come in a wide variety (urban/rural, highway/city-center), and with varying imaging conditions (weather, day/night). This work aims at showing the feasibility of self-supervised training and the benefits of online learning for free-space segmentation.

End-to-end learning of neural nets has become increasingly popular in computer vision systems, and is also proving to be successful within the field of Intelligent Vehicles. Currently, large neural nets (>10 layers) are being applied. In general, these nets require an exhaustive training process (multiple days on multiple GPUs) with many training data (millions of samples). In our research, we show the successful application of a small neural net with only 5 layers to free-space detection [1].

2. Method

The color-based segmentation algorithm used as a basis of our work is a Fully Convolutional Neural Network (FCN). For our research, we have relied on the net architecture and the CN24 library as described in [2]. The special feature of the proposed net is the spatial prior, that exploits the spatial bias which is naturally present in free-space segmentation in traffic scenes. Provided that the context (road detection) and data (images captured from within a vehicle [3]) are comparable to our research, we adopt the proposed network and training recommendations.

Self-supervised training requires an algorithm that generates weak training labels. To this end, we make use of a stereo camera and process disparity measurements with the Stixel World algorithm [4]. This is a probabilistic framework that segments traffic scenes into vertically stacked, rectangular patches that are labeled as either ground or obstacle. We rely on disparity for weak labeling, instead of on the color modality that is analyzed by the FCN. This increases the likelihood that the trained algorithm can correct unavoidable errors in the weak labels, instead of stepping into the same pitfalls.

For online training, we adopt the training strategy as presented in [5]: generate weak labels for several preceding frames to train or update a classifier for the current frame. A schematic overview of our experimental framework for free-space detection is shown in Figure 1. We train an FCN

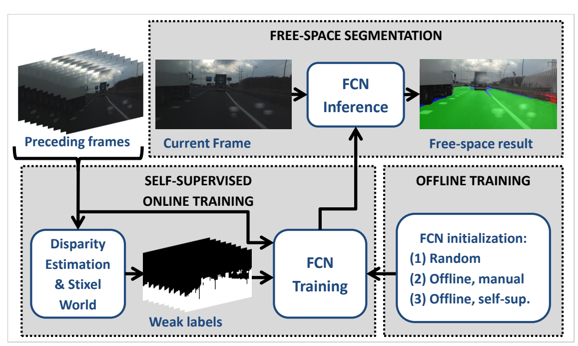

Figure 1. Schematic overview of our free-space detection method.

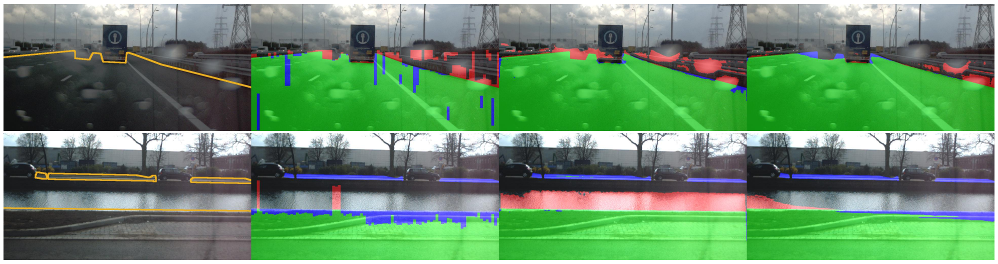

Figure 2. Qualitative results, left to right: input frames with ground truth, Stixel World baseline; our offline FCN result; our online FCN result. Coloring: true free-space (green), missed free-space (blue) and false free-space (red).

from scratch, or start with one of the offline pre-trained models and then tune the entire model with online data.

3. Validation

We perform the offline training of our FCN on two publicly available datasets (188 annotated frames with each 10 unlabeled preceding frames) [5]. We compensate for the low number of frames by performing a patch-wise training for the FCN. Our test set consists of 265 newly annotated frames[1] that were captured in a similar configuration.

For our offline-trained FCN, the quantitative results of supervised (manual labels) and self-supervised (labels from the disparity Stixel World) are nearly identical: F_{max} and AP [3] differ only 0.005. This confirms the feasibility of self-supervised learning, as relying on weak labels does not hamper the performance of our system. Our online-trained FCN outperforms the offline version (F_{max} and AP +5%) for all three initialization strategies. The advantage of online-tuned training is not in the quality of the final result, but crucial for the speed of convergence. More specifically, the models for which we apply tuning, outperform offline methods and the baseline already after 100 training iterations, whereas models trained from scratch need at least 500 iterations to match the offline FCN and more than 2,000 to exceed the Stixel World algorithm.

Figure 2 illustrates qualitative results of our experiments. The offline-trained FCN detects less false obstacles than the Stixel World baseline but tends to fail in rare, yet important cases (i.e. the canal). In contrast, our online-trained FCN outperforms both the Stixel World baseline and the offline training strategy.

Additionally, we have conducted three different experiments on the robustness of our online-training strategy. First, we have tested the drop in performance as a function of the delay between the frames on which the online training is performed and the frame under analysis. We have found that the segmentation score drops only about 2% for a delay of 2.5 seconds. Second, we have validated the influence of the number of FCN layers that are tuned online on the free-space detection result. Tuning only the final layer provides results within 1.5% of the full-tuning approach. Both experiments imply tradeoffs between online computation time and performance quality. Third, we misaligned the training frames with the test frame. This degraded the performance below that of the offline-trained models, also for the FCNs that were initialized with offline pre-trained nets. As the online FCNs outperform all other methods when their training sequence and test frame are aligned, this validates our claim that the online training is giving the system flexibility to adapt to new circumstances, and that it can be exploited beneficially in the context of free-space detection.

4. Conclusion

We have shown that Fully Convolutional Networks can be trained end-to-end in a self-supervised fashion in the context of free-space segmentation for ADAS. The segmentation results are similar to a conventional supervised strategy, thereby reducing the need for large amounts of manually labeled data. Furthermore, we have extended this result to show that it facilitates *online* training of a segmentation algorithm. Consequently, the free-space analysis is highly adaptive to any traffic scene that the vehicle encounters, which significantly improves the segmentation performance. In conclusion, we exploit the fact that our inherently adaptive approach can effectively be over-trained on a single traffic scene, which allows for a small FCN whose training converges fast enough to make real-time deployment feasible in the near future.

References

[1] Sanberg, W.P., Dubbelman, G., de With, P.H.N.: Free-Space Detection with Self-Supervised and Online Trained Fully Convolutional Networks. IS&T-EI 2017, (in press) 1

[2] Brust, C.A., Sickert, S., Simon, M., Rodner, E., Denzler, J.: Convolutional patch networks with spatial prior for road detection and urban scene understanding. VISIGRAPP 2015, 510-517 1

[3] Fritsch, J., Kuehnl, T., Geiger, A.: A new performance measure and evaluation benchmark for road detection algorithms. ITSC 2013 1, 2

[4] Pfeiffer, D.I.D.: The Stixel World. PhD thesis, Humboldt-Universität zu Berlin (2012) 1

[5] Sanberg, W.P., Dubbelman, G., de With, P.H.N.: Color-based free-space segmentation using online disparity-supervised learning. ITSC 2015, 906–912 1, 2

[1] available at http://www.willemsanberg.net.