Valeo
SMART TECHNOLOGY FOR SMARTER CARS

Online Out-of-Domain Detection for Automated Driving

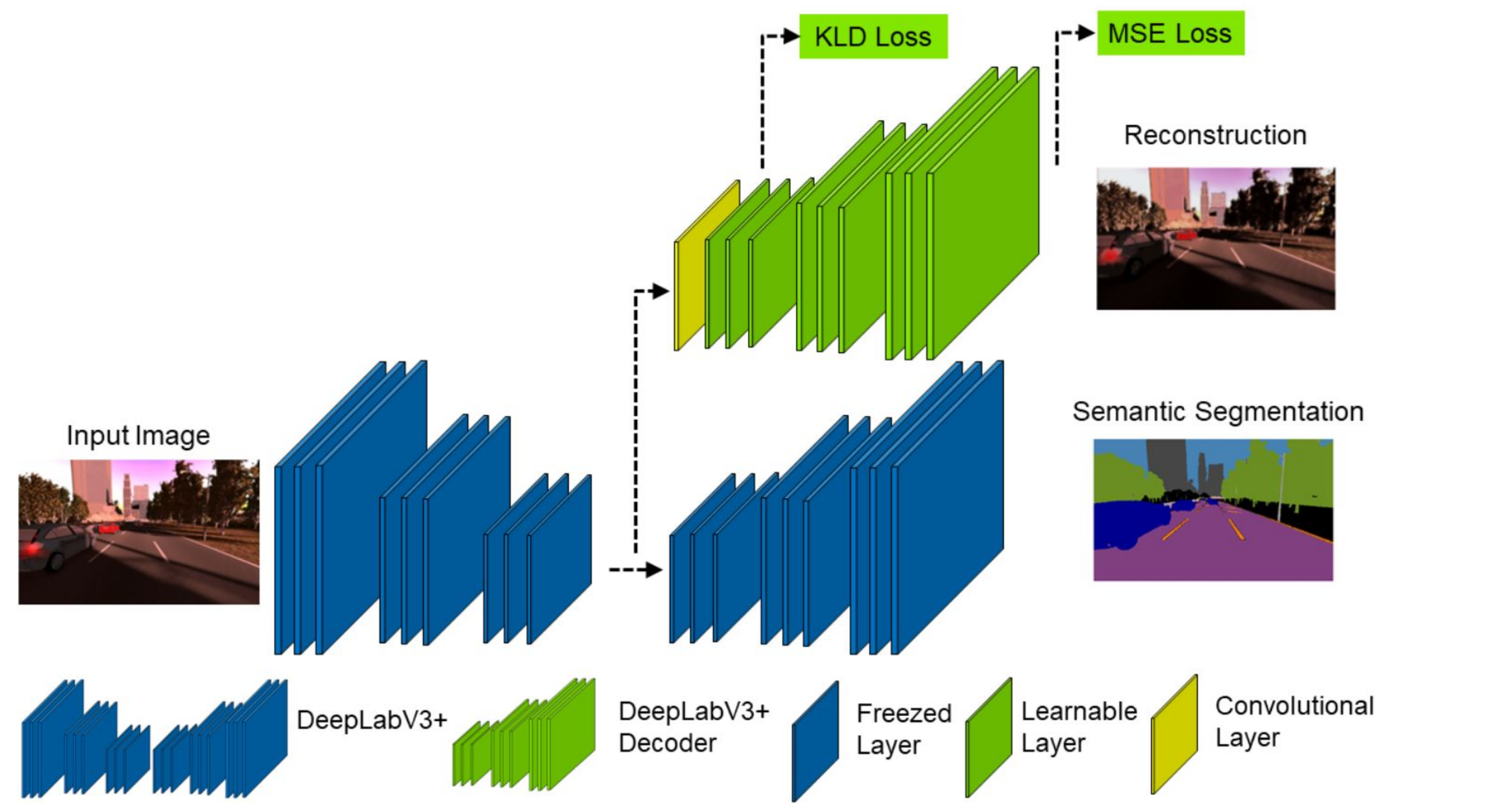
TECHNISCHE UNIVERSITÄT ILMENAU

Timo Sämann and Horst Michael Groß

Abstract

Ensuring safety in automated driving is a major challenge for the automotive industry. Special attention is paid to artificial intelligence, in particular to Deep Neural Networks (DNNs), which is considered a key technology in the realization of highly automated driving. DNNs learn from training data, which means that they only achieve good accuracy within the underlying data distribution of the training data. When leaving the training domain, a distributional shift is caused, which can lead to a drastic reduction of accuracy. In this work, we present a proof of concept for a safety mechanism that can detect the leaving of the domain online, i.e. at runtime. In our experiments with the Synthia data set we can show that a 100 % correct detection of whether the input data is inside or outside the domain is achieved. The ability to detect when the vehicle leaves the domain can be an important requirement for certification.

Approach

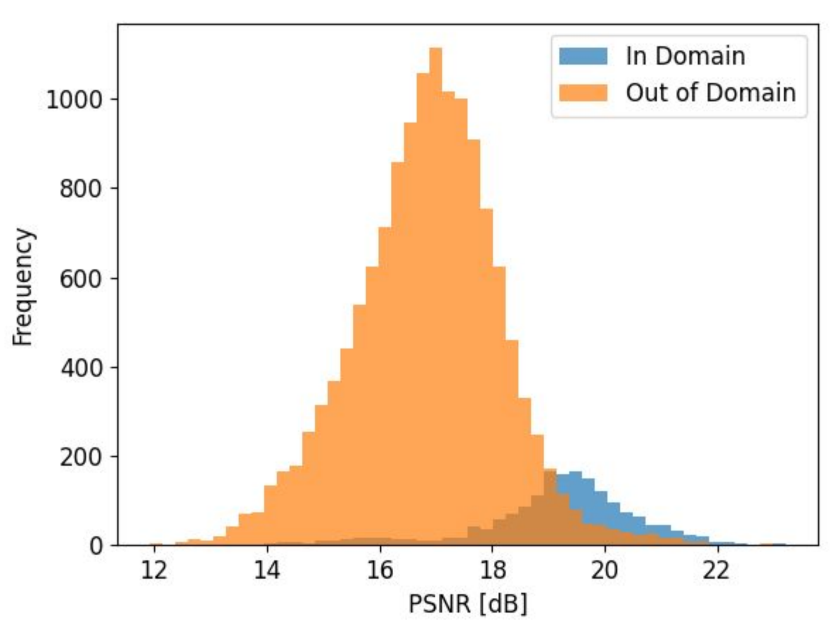


Dataset

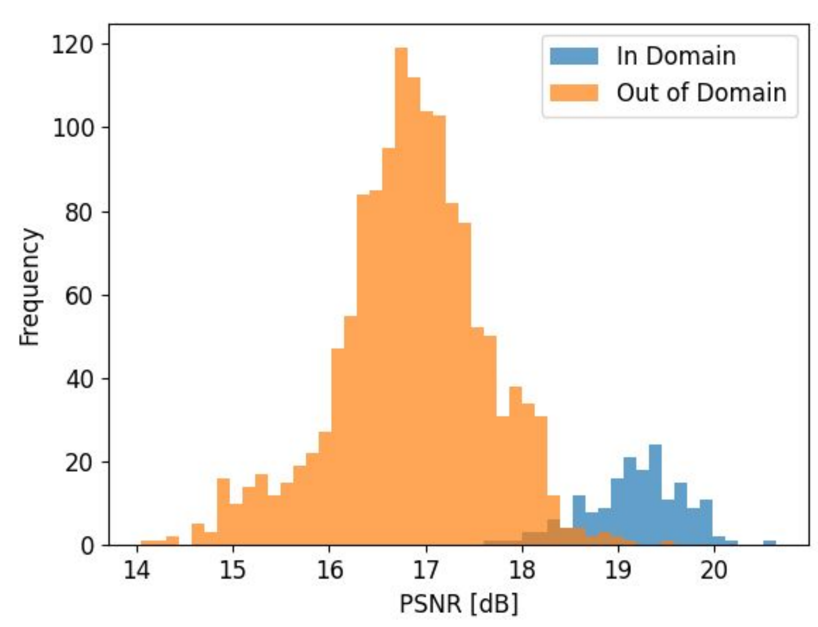
To carry out the experiments we used the Synthia Video Sequences. It contains a large number of domains such as dawn, fog, rain, winter, summer, spring, fall, night etc. We have used 4 sequences (SEQ), 2 SEQ with Highway and 2 SEQ with New York ish. Each sequence is further divided into the domains mentioned above. We used sequence 1 and 2 for training and 5 and 6 for testing. The resolution of the Synthia data is 768, 1280 pixels. For a simple overview of the data split used, we list it in the following:

- Training data, 2392 images: Dawn (SEQ1+2).
- Test data (in-domain), 1775 images: Dawn (SEQ5+6).
- Test data (out-of-domain), 13921 images: Winter (SEQ5+6), sunset (SEQ5+6), summer (SEQ5+6), spring (SEQ5+6), night (SEQ5+6), fog (SEQ5+6), rain (SEQ5), rainnight (SEQ5), softrain (SEQ5), winternight (SEQ5+6).

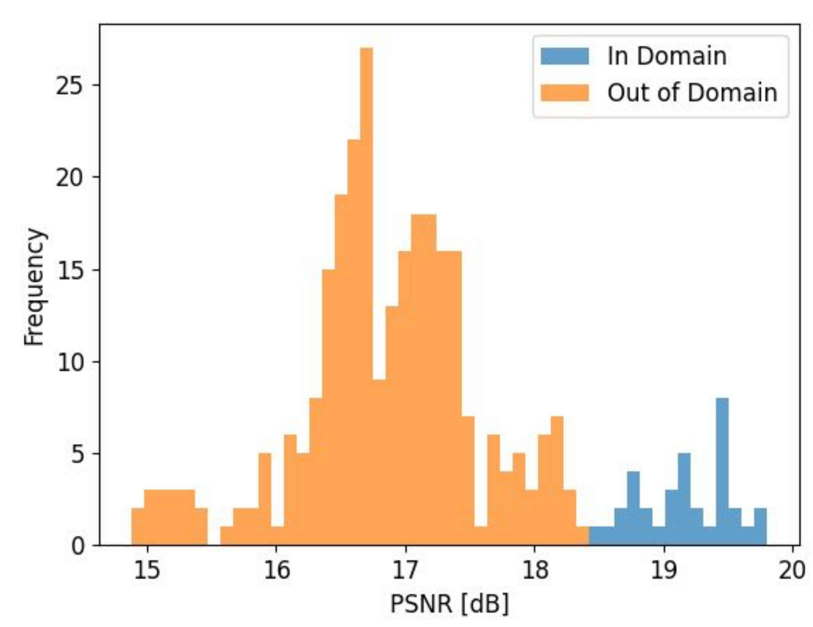
Results



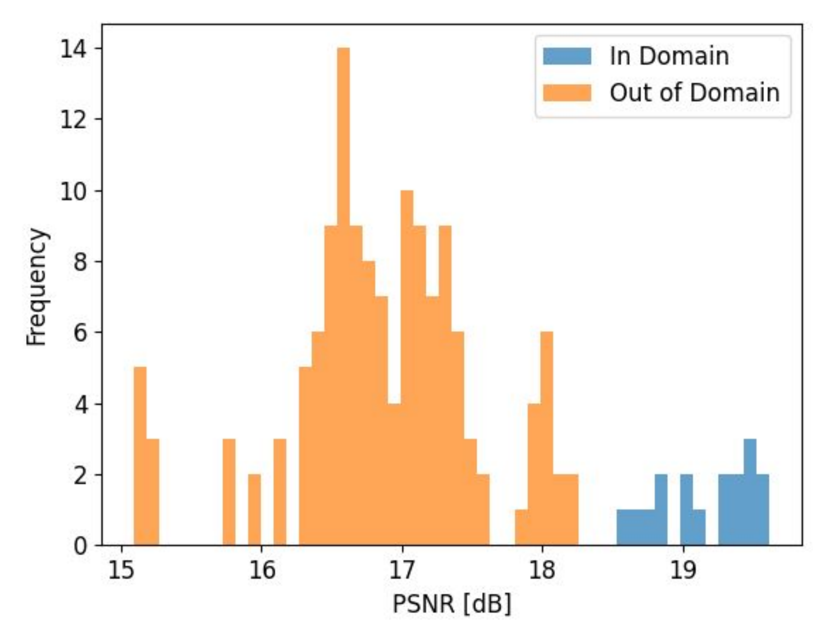
(a) τ = 1



(b) τ = 10



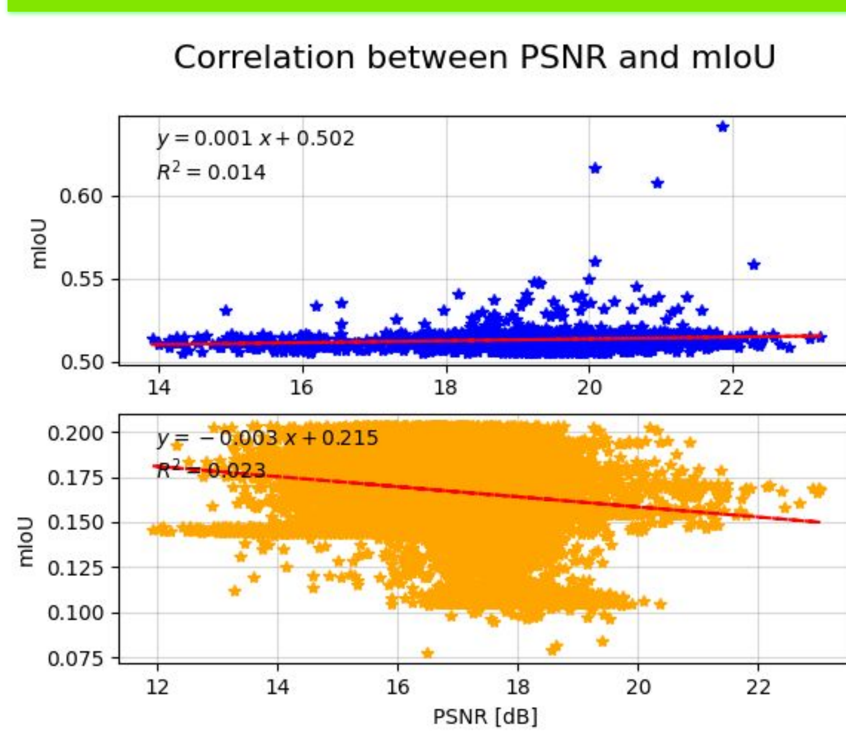
(c) τ = 50



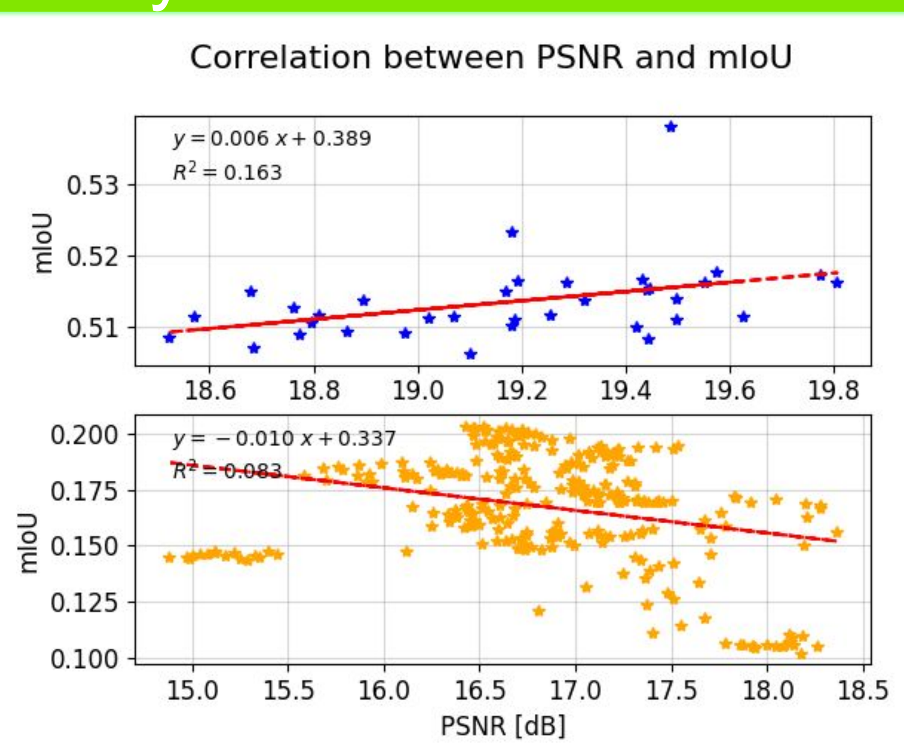
(d) τ = 100

The PSNR for in-domain and out-of-domain data was measured. The division of the data into in-domain and out-of-domain can be read in the dataset section. The result is shown in the form of a histogram. The x-axis represents the PSNR and is divided into 50 bins. The y-axis represents the frequency, i.e. the number of images evaluated. For (a) the evaluation shows that the in-domain data are mainly between 18 and 22 dB distributed. The out-of-domain data are mainly between 13 and 20 dB. Between 18 and 20 dB a clear overlap between in-domain and out-of-domain data can be seen. Thus, a clear separation between in and out-of-domain data is not easily possible in a single image analysis. For this reason a new value τ has been introduced, which indicates how many images are combined into a sequence before a single PSNR value is determined, i.e. the average PSNR over the sequence with the length τ is determined. τ is 1 for the first histogram (a), 10 for the second (b), 50 for the third (c) and 100 for the fourth (d). This averaging of the values narrows the variance of the individual domains, since outliers at the edge of the spectrum are smoothed out. The higher τ the lower the variance. A complete separation of in- and out-of-domain data is already possible with τ = 50. Since the data was generated with a frequency of 5 Hz, this means that at run time every 10 seconds a reliable estimation can be made whether the input images are in or out-of-domain. At a higher frame rate, a reliable estimation can probably be made after a shorter time.

Correlation Between Reconstruction Error and DNN Accuracy



(a) τ = 1



(b) τ = 50

This section describes the investigations on the correlation between the reconstruction error (PSNR) and the DNN accuracy (mean Intersection over Union, mIoU). It is assumed that for a low reconstruction error (high PSNR) a high DNN accuracy (high mIoU) is achieved and vice versa. For the datasets in- and out-of-domain the PSNR and the mIoU were calculated per image. In both subplots the x-axis represents the PSNR in dB and the y-axis represents the mIoU. Each image is represented by a star. The upper subplot with the blue stars describes the in-domain data and the lower one with the yellow stars the out-of-domain data.

References

Simon Burton, Lydia Gauerhof, Bibhuti Bhusan Sethy, Ibrahim Habli, and Richard Hawkins. Confidence arguments for evidence of performance in machine learning for highly automated driving functions. In International Conference on Computer Safety, Reliability, and Security, pages 365–377. Springer, 2019.

Timo Sämann, Peter Schlicht, and Fabian Hüger. Strategy to increase the safety of a dnn-based perception for had systems. arXiv preprint arXiv:2002.08935, 2020.

Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV), pages 801–818, 2018.

Li Wang, Dong Li, Yousong Zhu, Lu Tian, and Yi Shan. Dual super-resolution learning for semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3774–3783, 2020.

Jonas Lohdefink, Justin Fehrling, Marvin Klingner, Fabian Huger, Peter Schlicht, Nico M Schmidt, and Tim Fingscheidt. Self-supervised domain mismatch estimation for autonomous perception. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 334–335, 2020.

German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M. Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.