
Bergische Universität Wuppertal

Fakultät für Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational Mathematics (IMACM)

Preprint BUW-IMACM 21/20

D. Brüggemann, R. Chan, H. Gottschalk and S. Bracke

# Software architecture for human-centered reliability assessment for neural networks in autonomous driving

July 2, 2021

http://www.imacm.uni-wuppertal.de

# Software architecture for human-centered reliability assessment for neural networks in autonomous driving

Journal Title
XX(X):1–7

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

**Dominik Brüggemann[1], Robin Chan[2], Hanno Gottschalk[2] and Stefan Bracke[1]**

**Abstract**

The computer vision task “semantic segmentation” forms a crucial building block in the interaction of several redundant systems. As metric to evaluate the performance and reliability of this perceptual function, one commonly considers the number of false negatives (FNs), i.e. counting instances that have been overlooked by the perception model. From a practitioner’s point of view, however, faulty detections need to be considered in a more differentiated way. For example in autonomous driving, detection errors of vulnerable road users (VRUs, e.g. pedestrians) far away from the probable travel path of the ego vehicle are not as safety-relevant as detection errors of VRUs on the path ahead. Moreover, standard evaluation approaches do not consistently specify how well the VRU instance must be covered by the perception model in order to consider the VRU to be found. In this work we therefore introduce a sophisticated evaluation framework that assesses semantic segmentation models for autonomous driving not only based on their classification and localization abilities but also on distance information of VRUs within a safety-relevant region of interest ahead of the ego vehicle. This allows distinguishing irrelevant FNs from potentially relevant FNs and thus provides more safety-aware metrics.

**Keywords**

semantic segmentation, autonomous driving, vulnerable road users, perception, evaluation framework, safety-aware metrics, false negative reduction

## Introduction

Semantic segmentation combines the computer vision tasks of object classification and localization, figure 1. Any pixel in a high resolution image is attributed a class from a pre-defined semantic space. As a metric for the performance of a segmentation algorithm, in most cases realized by deep convolutional neural networks (CNNs) (1; 2; 3), one often uses pixel based metrics on a test data set, which is then averaged both over pixels and over test samples. Pixel-wise quantities however do not always distinguish between different semantic classes, which generally is problematic if one semantic category of special importance is underrepresented. For this reason, metrics like the commonly-used intersection over union (IoU, also known as (4)) are computed per semantic category by comparing ground truth segmentation masks with predicted segmentation masks for a specific category and then averaged over the classes.

Although the IoU as performance metric combines important quantities like the numbers of true positive (TP), false positive (FP) and false negative (FN) class predictions, it is still agnostic with regard to the application context of the computer vision system. If such modern artificial intelligence (AI) methods are deployed as perceptive systems in a safety critical application like medical imaging, e.g. (5), or autonomous driving, e.g. (6), the measurement of performance necessarily has to take into account the application specific failure modes.

For the example of autonomous driving, errors in perception could either be irrelevant, like if a tree is confused with a lamppost, or fatal, if a pedestrian is overlooked due to a confusion with the street. Apart from the class specific asymmetry of importance of confusion events (7; 8), it is evident that a measurement of performance based on pixel coverage is insufficient and should be replaced by an instance based assessment, like overlooked vulnerable road users (VRUs). Information on instances is often part of the annotation in public domains test data sets, see e.g. (9; 10).

In this work, we therefore focus on pedestrians as VRU that are overlooked by a perception system. We present and implement methods that enable an instance based assessment and provide performance metrics as well as visualization tools for the usecase of semantic segmentation in autonomous driving. Our approach allows for further contextualization in two regards. We measure a segmentation CNN’s detection ability of VRUs in a reachable area depending on the ego-car’s velocity. Moreover, filtering via the degree of detection, or other geometric properties

[1]Chair for Reliability Engineering and Risk Analytics, IZMD, University of Wuppertal, Gaußstraße 20, Wuppertal, Germany.
[2]School of Mathematics and Natural Sciences, IZMD, University of Wuppertal, Gaußstraße 20, Wuppertal, GER.

**Corresponding author:**
Dominik Brüggemann, Chair for Reliability Engineering and Risk Analytics, IZMD, University of Wuppertal, Germany.
Email: dbrueggemann@uni-wuppertal.de

**(a)** Synthetic input image **(b)** Corresponding ground truth

**Figure 1.** An example input image (a) from the dataset we used in our experiments and its corresponding ground truth semantic segmentation mask (b). Our focus lies at detecting pedestrian instances, which are labeled with red color in the segmentation. Images provided by: Mackevision Medien Design GmbH

of ground truth instances, is feasible. This matters, as downstream software modules tend to filter the raw semantic predictions of CNN prior to constructing the environmental model and determining the driving policy. Our software permits to asses the effects of filtering on the instance and zone based performance of detection.

The capabilities are demonstrated on a synthetic data set produced within the collaborative research project "KI - Absicherung - Safe AI for Automated Driving"*. The dataset consists of high-resolution street scene images with pixel-wise object class annotation as well as depth information.

The paper is organized as follows: We first briefly introduce the field of semantic segmentation in general, including available datasets and common performance metrics, and the neural network we employed for our experiments. Thereafter, we describe our approach and software architecture to provide contextualized performance information of the network. We provide results as well as visualizations for our testcase and finally give our conclusions along with a short outlook on our future software development.

## Deep neural networks for perception and test datasets

In this section, we give a brief introduction to the topic "semantic segmentation by CNNs". Furthermore, for our experiments we employ a state-of-the-art segmentation network, which we introduce as well. We then conclude this section with an overview of the dataset on which we report our results.

Semantic segmentation is the computer vision task of classifying each pixel in an image to a pre-defined class. Typically, the state-of-the-art in this field approaches this task by employing deep (convolutional) neural networks (CNNs), which can be understood as visual models extracting hierarchies of features. Those models are one of the basic components of an AI-based perception system for autonomous driving. By capturing the environment via cameras and feeding the produced images through a CNN, the ego-car's perception system is able to gain a visual understanding of the present scene.

One prominent test suite in this regard is the Cityscapes benchmark (9), which focuses on large-scale semantic segmentation of urban street scenes. Early success on this task was achieved by fully convolutional networks (11), allowing an efficient end-to-end training and inference of images of arbitrary size. Later, the introduction of "atrous" convolution operators, first employed in the *Deeplab* model (12), significantly impacted the segmentation quality by effectively enlarging the field of vision. Such models, therefore, are able to capture larger visual context and form the basis of many modern network architectures (2; 13). As the trend moved towards deeper and deeper network designs, the training of such models grew ever more difficult. The training process was then substantially eased by the integration of residual modules (3; 14; 15), which are part of nearly every recent network for semantic segmentation.

We, in this work, perform our experiments with a DeeplabV3+ model with a WideResnet backbone (13), that is pretrained by NVIDIA and publicly available. At the time of writing, this network is among the best performing models on the Cityscapes benchmark, reaching a mean intersection over union (mIoU) score of 83.5%. The mIoU as performance metric is commonly-used in other semantic segmentation benchmarks as well, such as KITTI (16) or Mapillary (17), and it measures the classification quality per pixel over an entire test dataset. Although further maximizing this global metric is important, particularly research-wise, it does not necessarily improve the overall system performance or guarantee safety as the mIoU evaluates methods independently of a scene's context. This drawback also concerns an AI-based perception system for autonomous driving, where usually other sensor data is additionally available.

In our proposed evaluation framework for semantic segmentation models, we take contextual additional information into account, such as the ego-car's velocity and the distance

*https://www.ki-absicherung-projekt.de/

of captured objects, in order evaluate the detection ability of vulnerable road users. Therefore, we aim at quantifying the deployability of camera-based perception models to real-world applications. For testing purposes, we demonstrate our results on a propriety synthetic dataset, provided by the "KI-Absicherung" consortium.

At the time of writing, the dataset contains 196 sequences with a total of 118,082 frames that were produced by the companies Mackevision Medien Design GmbH and BIT Technology Solutions GmbH. Special characteristics of the synthetically generated data include: different road surfaces, different lighting conditions (time of day, weather...), different road infrastructure and architecture, diverse vegetation, reflections, widely dispersed distribution of distance to VRUs, high frame rates in dynamic sequences (more than 10 fps), occlusions (various occluding objects) and groups of people.

Sensor images with up to three different sensor configurations (different contrast or brightness sensitivity) are available for the generated frames. In addition, a pixel-level annotation of the euclidean distance outgoing from the ego-car is provided. Also, many other annotation formats are available, including semantic instance segmentation, body part segmentation and 2D- and 3D-bounding boxes.

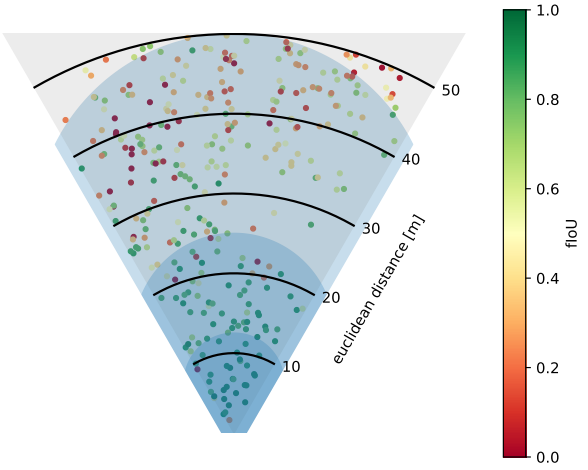


**Figure 2.** The blue shaded areas mark the reachable area for emergency braking (dark blue), normal braking (lighter blue) and delayed normal braking (light blue) at a speed of 50 km/h or less. The areas are also referred to as priority areas 1 (dark blue), 2 (lighter blue) and 3 (light blue). The different areas are shown in a representation that uses euclidean distance instead of longitudinal and lateral distance as seen in figure 6. The points represent the instances of the distance-filtered dataset. The colour of the dots represents the fIoU of the associated instances.

## Human-centered performance evaluation for CNNs

In this section, we first explain why our evaluation and the effort of the project revolve around the road user "pedestrian". We then discuss how to distinguish between relevant and non-relevant instances of this class of vulnerable road users. Furthermore, we present the prerequisites and assumptions the evaluations are based on and motivate, why the information from different types of sensors must be available evaluate the performance of CNNs with respect to VRU safety. We introduce the performance measure fIoU and give a brief overview of the preparation of the data for the evaluations of the ensuing section.

### *Vulnerable road users*

The project "KI-Absicherung" aims to make safety of AI-based function modules for highly automated driving verifiable. The project results form the foundation to build an industry consensus for a general safeguarding strategy of AI functions. As the primary research objective, the detection of pedestrians was selected within the project, as from a safety point of view, the relevance of VRUs detection is rated particularly high as pedestrians form the largest group of fatalities in street accidents that are not motor vehicle occupants*. Furthermore, the task of detecting pedestrians as VRU is an example for perception via mutliple sensors. In the following, pedestrians and VRUs are used interchangeably, as other VRU classes like cyclists are not represented in our data.

### *Relevant and non-relevant VRUs*

One objective of our framework is to refine the evaluation of CNNs by distinguishing between relevant and non-relevant false negative (FN) VRUs, i.e. pedestrians that have been overlooked. Note that we divide all VRUs into relevant and non-relevant, but focus specifically on the false-negative instances (i.e. the overlooked instances). As non-relevant false negative we consider those overlooked instances that are out of the zone that could be reached during a braking process. This means that no hazardous street scenario arises from this false negative, unless it is also overlooked on later image frames. In order to make this distinction, sensor information as well as information on the ego-car's velocity are used. An exemplary filtering of the relevant VRUs is shown in the section "test results".

In the evaluation we consider three different zones, which can bee seen in figure 2. We have chosen three zones, which are listed in ascending size and **descending prioritization** in the following:

1) The smallest zone has a longitudinal length of 12.5m. The length of the zone is based on the braking distance during a emergency braking at a speed of 50km/h or less. All points in this zone can theoretically be reached during such emergency braking. VRUs in this zone have the highest priority. In the following we will refer to this zone as **priority area 1**.
2) The second largest zone has a longitudinal length of 25m, which corresponds to the braking distance for a normal braking of 50 km/h speed. The zone thus includes all points that can be reached during a normal braking of 50 km/h or less. VRUs in this zone have the second highest priority. Note that the entire first zone is contained in the second zone. In the following we will refer to this zone as **priority area 2**.

*Source: Federal Statistical Office of Germany 2020

3) The largest zone has a longitudinal length of 50m, which corresponds to twice the braking distance in a normal braking of 50 km/h speed. VRUs in this zone have the third highest priority. Note that the first and second zone are contained in the third zone. In the following we will refer to this zone as **priority area 3**.

To reiterate the significance of the different zones: In the case of false negatives of VRUs in **priority area 1**, a possible collision can no longer be prevented by emergency braking. In this case, an accident can only be avoided by circumventing the VRU instance. In the case of false negatives of VRUs in **priority area 2**, an accident can be avoided by hazardous braking, and in the case of false negatives in **priority area 3**, by ordinary braking.

## Experimental setup

For our evaluation, we choose a subset of the project dataset that was provided by BIT Technology Solutions GmbH. This subset consists of 50 images and contains individual scenes that are largely independent of each other. We consider the following parameters, which all can be derived from the available virtual sensors, to determine the relevance of a false negative:

- Speed: The speed has a significant influence on the braking distance and thus on the area that can be reached while braking.
- Road surface: The surface characteristics of the road have an influence on steering movements or braking distance.
- Weather conditions: Wet or snow-covered roads can significantly increase the braking distance.
- Trajectory: The potential driving trajectory of the ego-car determines the size of the reachable area.
- Movement of the VRU: The likeliness, whether a VRU crosses the ego-car's driving trajectory, depends on both the direction and movement speed of of the VRU.

In our experiments, we consider the speed as main parameter, which is set to a maximum of 50 km/h. Furthermore, we assume that the road surface is ordinary and dry asphalt, since all our test images depict urban street scenes. For the sake of simplicity, we for now omit the trajectories of the ego-car and VRU.

An exemplary modeling of the possible reachable zones for braking distances of 12.5 m (red, corresponding to normal braking at ∼35 km/h) and 25 m (blue, corresponding to normal braking at 50 km/h) can be viewed in figure 3. It can be seen that at lower speeds, additional sections of the area can be reached, as a smaller curve radius has to be taken into account. Since only an upper limit is given for the speed, the reachable area is extended to the area outlined in green. Please note that the modeling shown is not based on scientific research, nor does it cover areas that could be reached after a collision with another road object. Instead, the displayed zones function as placeholders, which shall be replaced by scientifically proven reachable areas at a later stage.

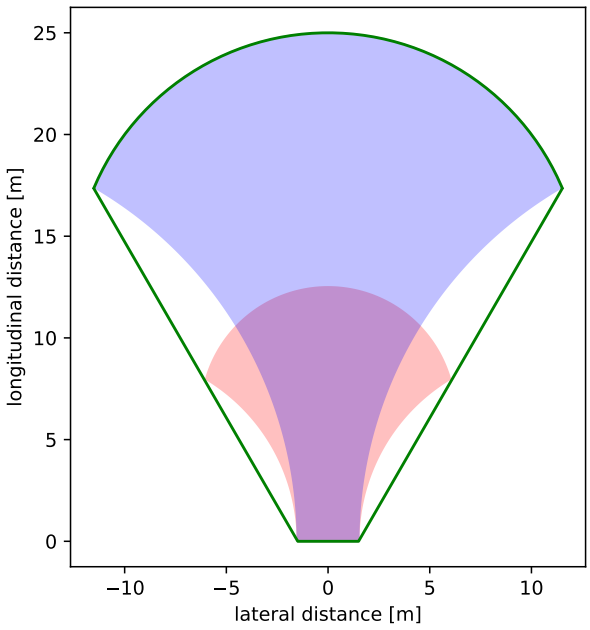


**Figure 3.** Reachable area for normal braking at 35 km/h (red) and 50 km/h (blue) ego-car velocity, respectively, when the ego-car is located at (0,0). Note, that the reachable area at **exactly** 35 km/h does not completely cover the reachable area at **exactly** 50 km/h. Therefore, under the assumption that the ego-car's velocity is any **below** 50 km/h, the reachable area for normal braking then is depicted as the convex hull of reachable areas at all velocities up to 50 km/h (green contours).

## Merging different sensors for evaluation

In order to evaluate the relevance of false negatives, the information from several sensors must be combined. In particular we require the real image and some kind of depth information. In our evaluation, the depth information is available as pixel-wise depth mask, but sensor data such as LIDAR or radar could also be used to obtain this information. Based on the image and the depth information (which measures the euclidean distance to the camera that takes the image), the position of a VRU relative to the ego vehicle can be determined. Since the angle of view is known, the position of the VRU in the real image can be used to convert the euclidean distance into a longitudinal and a lateral distance.

## Fair component-wise IoU

The most-commonly used performance metric in semantic segmentation is the intersection over union (IoU (4)), measuring how well a CNN detects and localizes objects. As the name suggests, given one prediction with its corresponding target, this metric is computed by dividing the area of overlap by the area of union, see figure 4 for an illustrative example. The prediction and target type in our evaluation are the connected components of pixels, sharing the class label of VRU, in the prediction and ground truth segmentation masks, respectively. By means of the component-wise IoU, we determine whether one VRU instance has been overlooked or not.

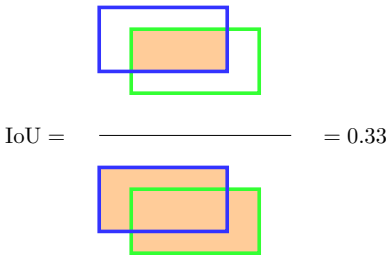


**Figure 4.** The ordinary component-wise IoU is computed by dividing the area of overlap (orange area top) by the area of union (orange area bottom). In this example, the prediction (blue rectangle) and target (green rectangle) are of size 2, whereas the overlap of these two is of size 1, yielding $\mathrm{IoU} = 0.33$.

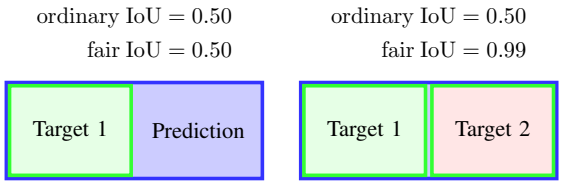


**Figure 5.** An illustrative comparison between the ordinary and fair component-wise IoU. In both examples the prediction (blue rectangle) is the same but different targets are covered (green rectangles). In this examples, the component to be evaluated is "target 1". Then, the fair component-wise IoU subtracts all correctly predicted regions from the union which are disjoint from target 1 (red area in the right figure). In this way, the fair component-wise IoU (fIoU) does not penalize the case when one prediction covers multiple targets.

However, standard semantic segmentation models do not provide predictions for single instances[†]. Therefore, we employ an adapted version of the component-wise IoU for our instance-based evaluation. We call this adjusted metric *fair component-wise IoU* (fIoU), which already has been introduced in (18; 19), see also figure 5 for an illustrative example. It is called fair since this metric does not penalize if one large prediction covers multiple instances, making this metric particularly suitable to perform an instanced-based evaluation of standard CNNs for semantic segmentation. We consider one VRU instance as detected if its component-wise IoU is greater than a chosen threshold, i.e. if $\mathrm{fIoU} > t, t \in [0, 1)$, and otherwise as overlooked.

### *Preparation of the DataFrame*

In order to prepare the evaluation, a structured data set is created by our framework. More precisely, a pandas DataFrame is created where each row corresponds to one VRU instance and contains the following information about that instance:

- *Fair component-wise IoU [%]*
- *Euclidean distance [m]* to the ego vehicle
- *Instance size [number of pixels]*
- *Longitudinal distance [m]* to the ego vehicle
- *Lateral distance [m]* to the ego vehicle

The generated DataFrame can then be filtered, limiting the evaluation to the relevant instances (or also relevant false negatives).

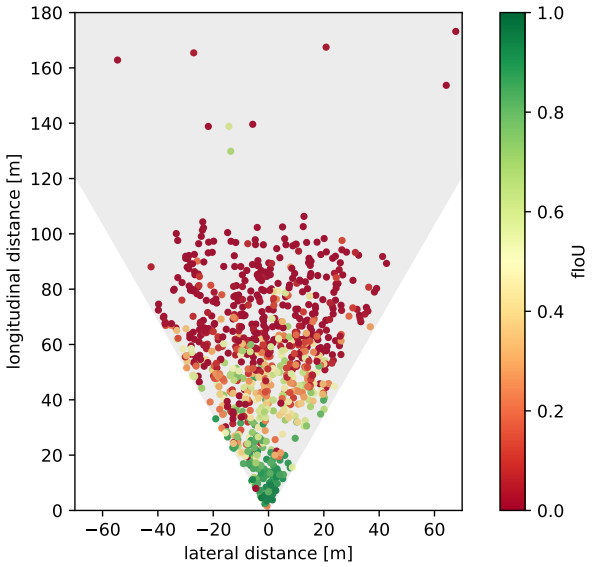


**Figure 6.** Representation of all VRU instances of the sequence under consideration. Each instance is represented by a dot. The colour of the dot indicates the fair component-wise IoU of the instance, ranging form red (fIoU = 0.0) over yellow (fIoU = 0.5) to green (fIoU = 1.0). The position of a dot is determined by the longitudinal and lateral distance of the instance to the ego vehicle. The grey area shows the angle of view.

## Test results

In this section, an exemplary analysis of a sequence will be shown. To this end, we first give some statistical details on the sequence to be evaluated. Subsequently, the filtering of the data set is considered step by step before the results of the analysis are presented.
There are a total of 893 VRU instances in our evaluation data set, as can be seen in figure 6. Most instances are not further than 110m away. However, a few outliers exist, being up to 185.95m away from the ego-car. In the figure, it can be observed that the quality of the detection decreases with increasing distance.

As can be seen in table 1, a large proportion of the false negatives are located on non-relevant areas (i.e. none of the three prioritized areas). Of the 893 instances, 305 instances (34.15%) are in the prio. 3 zone, 95 (10.64%) in the prio. 2 zone and 32 (3.58%) in the prio. 1 zone. The same statement holds true for the different IoU thresholds, whereby with decreasing IoU, the proportion of false negatives that occur in the prioritised areas decreases further. For example, 15 (6.15%) of the total 244 instances that were completely overlooked (i.e. fIoU = 0%) are located in priority area 3 and one (0.41%) is located in priority area 1 (and thereby in priority area 2 as well). Of the 659 false-negatives with fIoU $<$ 50%, 125 (18.97%) are located in priority area 3, twelve (1.82%) are located in priority area 2 and three (0.46%) are located in priority area 1. Of the 305 instances in priority area 3, 15 (4.92%) were completely missed and 125 (40.99%)

[†]The task of detecting and localizing each distinct object of interest appearing in an image is referred to as instance segmentation.

| fIoU | $= 0\%$ | $\leq 10\%$ | $\leq 20\%$ | $\leq 30\%$ | $\leq 40\%$ | $\leq 50\%$ | $\leq 60\%$ | $\leq 70\%$ | $\leq 80\%$ | $\leq 90\%$ | total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| #FN | 244 | 430 | 490 | 559 | 609 | 659 | 713 | 764 | 806 | 868 | 893 |
| #rel. FN (prio. 3) | 15 | 31 | 46 | 74 | 100 | 125 | 153 | 185 | 218 | 280 | 305 |
| #rel. FN (prio. 2) | 1 | 5 | 6 | 8 | 9 | 12 | 15 | 21 | 29 | 71 | 95 |
| #rel. FN (prio. 1) | 1 | 1 | 1 | 2 | 2 | 3 | 3 | 3 | 4 | 15 | 32 |

**Table 1.** The number of instances whose fIoU fall below the threshold value specified in the header is indicated. These numbers are given for the unfiltered dataset (in the row #FN) as well as the three priority areas. The last column also shows the total number of instances in the respective areas.

were detected with a fIoU of less than 50%. Of the 95 instances in priority area 2, one (1.05%) was completely missed and just twelve (12.63%) were detected with a fIoU of less than 50%. In summary, it can be observed that the majority of false negatives are not relevant, as they do not occur in any of the three prioritised areas.

## Conclusion and outlook

In this work, we presented a safety-aware evaluation framework for semantic segmentation models, accompanied with several visualization tools. In contrast to traditional performance metrics, that measure a model's classification capability independently of the application context, we take further available sensor data into account in order to provide a contextualized safety metric. More precisely, we focus on the semantic segmentation of street scenes and we additionally consider depth information to determine pedestrians as VRUs that are within reachable area of the ego-car. Besides being assessed by means of the detection quality, positional information of the VRUs allows to distinguish between relevant false negative VRUs (overlooked pedestrians) and non-relevant ones. We demonstrated our software's capabilities on a state-of-the-art semantic segmentation network and a synthetic dataset produced within the collaborative research project "KI -Absicherung - Safe AI for Automated Driving". The generated report provides greater insight in terms of deployability of segmentation models as perception system for autonomous driving.

We plan to extend the software functionalities to handle further perception modalities such as 2D or 3D bounding boxes. Moreover, other sensor data, like the exact ego car's velocity or the steering angle, is also planned to be included since this affects the relevant zones. Up to now, we only evaluated single images. As image data is often available in video sequences, we plan to add temporal consistency checks, since there is a major difference between failing to detect an instance once and failing to detect repeatedly in consecutive frames.

### Acknowledgements

This work is funded by the German Federal Ministry for Economic Affairs and Energy within the project "KI Absicherung – Safe AI for Automated Driving", grant no. 19A19005R. The authors would like to thank the consortium for the helpful cooperation.

### References

[1] Chollet F. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1251–1258.

[2] Chen LC, Zhu Y, Papandreou G et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*. pp. 801–818.

[3] Sandler M, Howard A, Zhu M et al. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 4510–4520.

[4] Jaccard P. The distribution of the flora in the alpine zone. 1. *New phytologist* 1912; 11(2): 37–50.

[5] Dong H, Yang G, Liu F et al. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In *annual conference on medical image understanding and analysis*. Springer, pp. 506–517.

[6] Chen C, Seff A, Kornhauser A et al. Deepdriving: Learning affordance for direct perception in autonomous driving. In *Proceedings of the IEEE international conference on computer vision*. pp. 2722–2730.

[7] Chan R, Rottmann M, Hüger F et al. Application of Maximum Likelihood Decision Rules for Handling Class Imbalance in Semantic Segmentation. In *The 30th European Safety and Reliability Conference (ESREL)*. pp. 3065–3072.

[8] Chan R, Rottmann M, Dardashti R et al. The ethical dilemma when (not) setting up cost-based decision rules in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. pp. 0–0.

[9] Cordts M, Omran M, Ramos S et al. The cityscapes dataset for semantic urban scene understanding. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[10] Geyer J, Kassahun Y, Mahmudi M et al. A2D2: Audi Autonomous Driving Dataset, 2020. URL https://www.a2d2.audi. 2004.06320.

[11] Shelhamer E, Long J and Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2017; 39(4): 640–651. DOI:10.1109/TPAMI.2016.2572683. URL https://doi.org/10.1109/TPAMI.2016.2572683.

[12] Chen LC, Papandreou G, Kokkinos I et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2018; 40(4): 834–848. DOI:10.1109/TPAMI.2017.2699184.

[13] Zhu Y, Sapra K, Reda FA et al. Improving semantic segmentation via video propagation and label relaxation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. URL https://nv-adlr.github.io/

publication/2018-Segmentation.

[14] He K, Zhang X, Ren S et al. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[15] Wu Z, Shen C and van den Hengel A. Wider or Deeper: Revisiting the ResNet Model for Visual Recognition. *Pattern Recognition* 2019; 90: 119–133. DOI:https://doi.org/10.1016/j.patcog.2019.01.006. URL https://www.sciencedirect.com/science/article/pii/S0031320319300135.

[16] Geiger A, Lenz P and Urtasun R. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

[17] Neuhold G, Ollmann T, Bulò SR et al. The mapillary vistas dataset for semantic understanding of street scenes. In *2017 IEEE International Conference on Computer Vision (ICCV)*. pp. 5000–5009. DOI:10.1109/ICCV.2017.534.

[18] Rottmann M, Colling P, Hack TP et al. Prediction Error Meta Classification in Semantic Segmentation: Detection via Aggregated Dispersion Measures of Softmax Probabilities. In *2020 IEEE International Joint Conference on Neural Networks (IJCNN)*.

[19] Chan R, Lis K, Uhlemeyer S et al. Segmentmeifyoucan: A benchmark for anomaly segmentation, 2021. 2104.14812.