https://doi.org/10.2352/EI.2022.34.16.AVM-101

Spatial Precision and Recall Indices to assess the performance of Instance Segmentation algorithms

Mattis Brummel[1], Patrick Müller[1], Alexander Braun[1]
[1]University of Applied Sciences Düsseldorf, Düsseldorf, Germany

Abstract

Since it is essential for Computer Vision systems to reliably perform in safety-critical applications such as autonomous vehicles, there is a need to evaluate their robustness to naturally occurring image perturbations. More specifically, the performance of Computer Vision systems needs to be linked to the image quality, which hasn't received much research attention so far. In fact, aberrations of a camera system are always spatially variable over the Field of View, which may influence the performance of Computer Vision systems dependent on the degree of local aberrations. Therefore, the goal is to evaluate the performance of Computer Vision systems under effects of defocus by taking into account the spatial domain. Large-scale Autonomous Driving datasets are degraded by a parameterized optical model to simulate driving scenes under physically realistic effects of defocus. Using standard evaluation metrics, the Spatial Recall Index (SRI) and the new Spatial Precision Index (SPI), the performance of Computer Visions systems on these degraded datasets are compared with the optical performance of the applied optical model. A correlation could be observed between the spatially varying optical performance and the spatial performance of Instance Segmentation systems.

Introduction

The performance of Computer Vision (CV) algorithms is quantified with many different metrics like precision vs. recall and miss rate (MR) vs. false positives per image (FPPI) [2, 3]. In these metrics, the outputs of CV algorithms are binary classified into true positives (TPs) and false positives (FPs) by comparing them with labeled ground truth data based on accuracy measures with the intersection over union (IoU). For a certain score threshold defining how high the confidence score of the CV system's output must be in order to count as prediction, the set of FPs and TPs as well as actually positives (Ps) labeled in the dataset results in relative values such as precision and recall. As an example, recall is the ratio of correctly identified instances to the number of actually positive instances labeled in the dataset. Computing these relative values for all possible score thresholds results in the aforementioned curves, which in turn can be summarized into the average precision (AP) and log average miss rate (LAMR), respectively. Consequently, one important aspect all these metrics share is that they are all *aggregate*, meaning that each point in the curves is averaged over the whole dataset. In this process the information is lost *where* in the image the detection occurs, and what exact performance value that individual output of the CV algorithm contributes to the overall measure.

In our group we focus our research on the physical-realistic simulation of images for the training of algorithms used in an automotive context, for advanced driver assistance systems (ADAS) or autonomous driving (AD). Using realistic optical models such as [7, 6] or [5] we degrade given images – both from simulation and already recorded drive scenes – to investigate the effect that the different optical properties have on CV algorithms [9, 5]. Realistic optical properties vary over the field of view, they are not translation-invariant. This varying optical quality certainly influences the performance of the ML-based CV algorithms, but this influence is not visible in the established aggregate metrics.

We therefore introduced a novel metric *Spatial Recall Index* in [9] that spatially resolves the performance of object detection, i.e. it assigns a recall value to every pixel within the image size. This spatial resolution allows for a detailed look at the influence of the varying optical properties over the whole image size. Note that SRI is also an aggregate metric in the sense that all detected objects are compared to all ground truth (GT) objects, only with a spatial resolution of this information.

In this article we extend our previous work in two aspects. Firstly, we introduce the *Spatial Precision Index* analogous to the SRI. Secondly, we extend both metrics to work on instance segmentation algorithms with arbitrary form and shape of the objects, not only on object detection with rectangular object size. While the former is a straight-forward adaption, the latter requires a novel treatment of the intersection and union of the detected and the GT instances, in order to take the right pixels into account in this metric.

This work is structured as follows. After introducing the spatial evaluation metrics for instance segmentation algorithm, we describe our evaluation methods in detail. Here, we first present the system under test (SUT) and dataset as well as our considerations for their selection. Then, we briefly describe the spatially variant image degradation introduced in [9] before explaining the evaluation process both with standard metrics and the newly proposed spatial metrics. Finally, we present the results and conclude this work.

Spatial Precision and Recall Indices

In a dataset with K labeled ground truth instances of a certain category from which N instances are correctly detected by the instance segmentation system, the recall value is defined as

$$\text{Recall} = \frac{\sum_n^N \text{TP}_n}{\sum_n^N \text{TP}_n + \sum_m^M \text{FN}_m} = \frac{\sum_n^N \text{TP}_n}{\sum_k^K \text{P}_k}, \tag{1}$$

where TP and FN denote true positives and false negatives, respectively, and P refers to labeled ground truth instances (i.e. positives) in the dataset. Moreover, with J wrong segmentation out-

puts, the precision value is defined as

$$\text{Precision} = \frac{\sum_n^N \text{TP}_n}{\sum_n^N \text{TP}_n + \sum_j^J \text{FP}_j}. \tag{2}$$

where FP denotes false positives outputted by the system. Simply put, the recall value penalizes FNs and thus ground truth instances that aren't detected by the instance segmentation system, whereas the precision value penalizes FP instances that are incorrectly or inaccurately outputted by the system.

Equation 1 and 2 are the basis for the spatial metrics to assess the performance of instance segmentation systems proposed in this paper. Given a dataset with equally sized images of height h and width w, the $h \times w$-sized spatial recall index (SRI) is defined for object detection in [9], and accordingly works on rectangular bounding boxes of the detected objects. It assesses the performance spatially by assigning a local recall value to each pixel (x, y) in $h \times w$. We now adapt the SRI metric to instance segmentation with arbitrary object forms by defining the $h \times w$-sized SRI as

$$\begin{aligned} \text{SRI} = &\left[\sum_{n=1}^{N} \begin{cases} 1 & (x,y) \in [\text{TP}_n \cap \text{P}_n] \\ 0 & else \end{cases}\right] \\ &\oslash \left[\sum_{k=1}^{K} \begin{cases} 1 & (x,y) \in \text{P}_k \\ 0 & else \end{cases}\right], \end{aligned} \tag{3}$$

where $(x, y) \in [\text{TP}_n \cap \text{P}_n]$ denotes all pixels that belong to the intersection area of the n-th TP mask and its corresponding ground truth mask, and $(x, y) \in \text{P}_k$ indicate all pixels belonging to the k-th ground truth mask. The sum over all N TPs and K Ps results in the $h \times w$-sized numerator and denominator, respectively, and the element-wise division denoted by $\oslash$ yields the SRI (see figure 1, upper row).

Moreover, in addition to extending the SRI to instance segementation, we take the logical next step to use precision as well, and propose a $h \times w$-sized spatial precision index (SPI) defined as

$$\begin{aligned} \text{SPI} = &\left[\sum_{n=1}^{N} \begin{cases} 1 & (x,y) \in [\text{TP}_n \cap \text{P}_n] \\ 0 & else \end{cases}\right] \\ &\oslash \left[\sum_{n=1}^{N} \begin{cases} 1 & (x,y) \in \text{TP}_n \\ 0 & else \end{cases} + \sum_{j=1}^{J} \begin{cases} 1 & (x,y) \in \text{FP}_j \\ 0 & else \end{cases}\right], \end{aligned} \tag{4}$$

where the denominator represents all TP and FP instances, from which the pixels (x, y) in $h \times w$ that belong to the respective instance mask are incremented by 1. Just like the numerators of the numerical precision and recall values, the numerators of the SRI and SPI are equal. For the sake of clarity, figure 1 shows a schematic formula of both spatial performance metrics.

In summary, with the local performance indices SRI and SPI, each pixel (x, y) in $h \times w$ is assigned a recall and precision value, respectively. In a spatial and pixel-wise manner, the former penalizes FNs, while the latter penalizes FPs. Since only the intersection area of TP instances and corresponding ground truth instances are incremented in the numerator, both metrics additionally penalize small inaccuracies in TP instances. This is visualized in figure 1, where the red intersection area incremented by 1 in the numerator is smaller than the involved ground truth and TP instances in the denominator of the SRI and SPI, respectively.

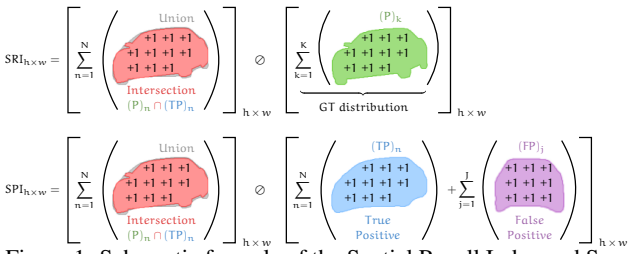


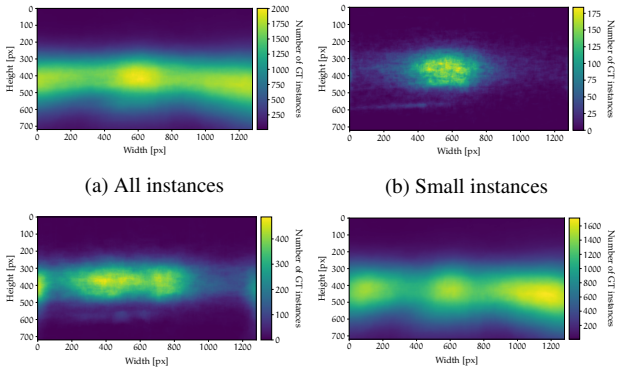
Figure 1: Schematic formula of the Spatial Recall Index and Spatial Precision Index for Instance Segmentation.

Methods

Our general approach is to first take a large-scale instance segmentation dataset with well distributed objects, then degrade its images using an optical model that simulates a real lens with spatially varying optical performance, and finally evaluate the performance with the newly proposed SRI and SPI metrics on both the original and the degraded datasets. By parameterizing the optical model, we create several degraded datasets with different effects of defocus, which we refer to as test cases of a defocus study. For each test case, we compute the difference between the spatial performance on the original dataset and the spatial performance on the degraded dataset by subtracting the baseline results from the results under effects of defocus. Lastly, we compare the spatial performance differences denoted by $\text{SPI}_{\texttt{drop}}$ and $\text{SRI}_{\texttt{drop}}$, respectively, with the optical performance of the parameterized lens model.

Selection of dataset and Computer Vision system

Considering that spatially well distributed instances are the most important aspect for a statistically reliable SRI and SPI computation, we base our evaluation on the Berkeley Deep Drive (BDD100K) dataset [13]. Its distinct feature is the diversity of scenes ranging from city streets over tunnels, residential areas and parking lots to gas stations in different times of day and various cities throughout the US [13]. It contains 8k labeled images with instance segmentation annotations, on which we evaluate the performance for the category "car". The aforementioned scene diversity contributes to the desired distribution of car instances. Figure 2 shows the distribution for all instances as well as different instance area ranges.



(a) All instances (b) Small instances

(c) Medium-sized instances (d) Large instances

Figure 2: Ground truth distribution of car instances for (a) all instances and (b-d) different instance area ranges defined as small $= [0, 32^2]$px, medium $= [32^2, 96^2]$px and large $= [96^2, \infty]$px.

For the performance evaluation to be representative of the performance on unseen data, the SUT must have been trained on a dataset other than BDD100K. Moreover, achieving a good baseline performance on the challenging BDD100K dataset requires the SUT to generalize well, which requires a high-quality dataset for training. We choose the Cascade Mask R-CNN [1] from Meta AI's Detectron2 repository trained on the COCO [8] training and validation set, which achieves a solid baseline performance on the COCO 2017 test set [12].

Image degradation

To show the dependence of the instance segmentation performance on the spatially variant image quality we use the optical model from [9]. The parameterizable model includes effects such as chromatic aberration, astigmatism and space-variant blur as present in real lens objectives. The model is available for different defocus parameters $Z_\Delta \in \{-1.25, \ldots, 0, \ldots, +1.25\}$. Figure 3 demonstrates the effect of the spatially variant model for different defocus parameters on a scene from [13]. The upper row (b-e) displays the edge location and the lower row (f-i) the central location as drawn in figure 3 a. The nominal position (d,h) adds the least amount of blur, but still shows a difference compared to the original scene (b,f). This difference is most visible in (b) as for the nominal position the amount of blur increases gradually towards the edges of images. The remaining image sections exhibit greater blurring caused by the larger defocus parameters. However, according to the respective local point spread functions (PSFs), either the edges of the image (c) are significantly blurrier or the middle areas (i). Note that this is a main characteristic of space-variant PSF models. Therefore, it is expected that, besides scene related difficulties in instance segmentation, according to the optics model, either objects located at the edges or objects at middle areas are harder to precisely segment.

(a) ROIs

(b) Original (c) $Z_\Delta = -1.25$ (d) $Z_\Delta = 0$ (e) $Z_\Delta = +1.25$

(f) Original (g) $Z_\Delta = -1.25$ (h) $Z_\Delta = 0$ (i) $Z_\Delta = +1.25$

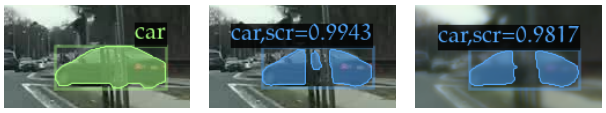
Figure 3: RoIs in image (a) at edge (b-e) and central (f-i) locations before and after degradation with $Z_\Delta \in \{-1.25, 0, +1.25\}$. Note that for the sake of clarity we trimmed from the original 1280×720-sized BDD100k image in (a) portions of 50px at the edges as well as 350px and 100px at the top and the bottom, respectively.

Standard performance evaluation

We first evaluate the overall performance using standard metrics. This involves measuring the accuracy of each instance segmentation output based on the IoU. Following usual benchmarks, an IoU threshold of 0.5 is used to assign TP and FP labels to each instance segmentation output [4]. Simply put, an output is assigned a TP label if it matches a GT instance with respect to the IoU threshold, whereas all outputs that don't match any GT instance are assigned FP labels. If one instance segmentation output matches multiple GTs, only the match with the highest IoU counts. Conversely, if one GT matches multiple instance segmentation outputs, only the match with highest scored segmentation output counts (i.e. outputs are matched in descending order by scores) [3]. Finally, by varying the score threshold from high to low, the number of TPs, FPs and Ps are used to compute the precision and recall (see equation 2 and 1) or MR ($= 1 -$ precision) and FPPI for each score threshold by suppressing all outputs that fall below it. Figure 4 shows exemplary two GT instances at central and edge locations, respectively, where the IoUs with corresponding outputs of the CV algorithm drop after image degradation. The IoUs of all instances exceed the IoU threshold of 0.5. Consequently, apart from the fact that the outputs show also a drop in the score through degradation, all standard metrics treat these instances as TPs, at least for positions in the curves that represent score thresholds below the outputted confidence scores.

For safety-critical applications such as automotive systems, a reasonable choice of a score threshold or sensitivity for deployment is often based on an upper limit of acceptable FPPI independent of the instance density in the dataset [3]. Therefore, we base our overall evaluation on the standard MR vs. FPPI metric, from which we extract operating points for the subsequent spatial analyses. Analogous to [11], we standardize all operating points based on the baseline curve rather than adjusting the corresponding score thresholds to FP rates on degraded data. This process mimics a realistic scenario where the system is deployed with a specific choice of a score threshold based on the evaluation on a test set and confronted with different defocus conditions through lens aberrations afterwards. Pezzementi et al. [10] refer to operating points up to an average of 10^{-1} FPs per image as reasonable or realistic for deployment of automotive applications. For the spatial evaluation, we use FPPI $= 0.1$ and further shift the operating point for test purposes to rather unrealistic scenarios with an

(a) GT (edge location) (b) baseline (c) $Z_\Delta = -1.25$



(d) GT (central location) (e) baseline (f) $Z_\Delta = +1.25$

Figure 4: Ground truth instances on edge (a) and central (d) locations as well as examples of instance segmentation outputs where the IoU drops after image degradation. The IoU for (a) drops from 0.94 in (b) to 0.62 in (c) and the IoU for (d) drops from 0.7 in (e) to 0.65 in (f).

average of 1 and 10 FPs per image to cover more positions of the MR vs. FPPI curve.

Spatial performance evaluation

Given an operating point extracted from the MR vs. FPPI curve, we have a set of matches between TPs and Ps, as well as FPs and FNs to compute the SRI and SPI for each test case. For a statistically reliable spatial evaluation, however, it is important to take the ground truth distribution into consideration. Figure 2 indicates that the distribution of cars throughout the dataset changes for different instance sizes. To prevent the spatial evaluation results from being dependent on instance sizes, it is therefore crucial to restrict the analyses to a sufficiently small instance area range. We partitioned the instances into three area ranges as shown in 2 b-d, from which we consider the distribution of medium-sized instances to be best suited for the spatial evaluation. Based on the COCO API, all ground truth and FP instances beyond the area range are ignored [8].

For each test case of the aforementioned defocus study, we compute the SRI and SPI with the set of instances extracted for specific operating points within the applied area range of $[32^2, 96^2]$px. Then, we compute the spatial performance drop through the image degradation for each test case by

$$\mathrm{SRI}_{\mathrm{drop}} = \mathrm{SRI}_{\mathrm{base}} - \mathrm{SRI}_{Z_\Delta}\,, \tag{5}$$

$$\mathrm{SPI}_{\mathrm{drop}} = \mathrm{SPI}_{\mathrm{base}} - \mathrm{SPI}_{Z_\Delta}\,, \tag{6}$$

where the minuend refers to the baseline result and the subtrahend indicates the results for the test case under effects of defocus simulated by the optical model with the respective defocus offset Z_Δ. Finally, in the result section, we compare both the $\mathrm{SRI}_{\mathrm{drop}}$ and $\mathrm{SPI}_{\mathrm{drop}}$ with the optical performance of the optical model.

Results

The overall instance segmentation performance is evaluated for the baseline and all test cases of the defocus study by taking into account all car instances in the dataset. Figure 5 shows the resulting MR vs. FPPI curve. Similar to [11], we mark three score thresholds that lead to an FPPI value of 0.1 (rectangle), 1 (triangle), and 10 (circle) in the baseline curve, representing the operating points for subsequent analyses. As expected, the performance drops only slightly through moderate defocus conditions simulated by the optical model in nominal position, while a larger performance drop is observed after parameterization with $Z_\Delta = \pm 1.25$. The marker indicate that a reduced image quality doesn't necessarily lead to an overall increase in the average number of FPs per image. Instead of generating more FPs, the system produces fewer outputs and misses more ground truth instances when the image quality worsens. The latter is reflected by an increased MR at the operating points, especially for $Z_\Delta = \pm 1.25$.

Before evaluating the performance spatially with the SRI and SPI, we show the precision vs. recall curve for medium-sized car instances in figure 6. We place the marker based on the score thresholds that lead to FPPI $= \{0.1, 1.0, 10.0\}$ in the overall evaluation shown in figure 5. In fact, the marker in figure 6 represent exactly the overall precision and recall values that result from the set of instances that are considered in the SRI and SPI computation with equation 3 and 4, respectively.

The precision vs. recall curve for medium-sized instances confirms the aforementioned observation that the image quality

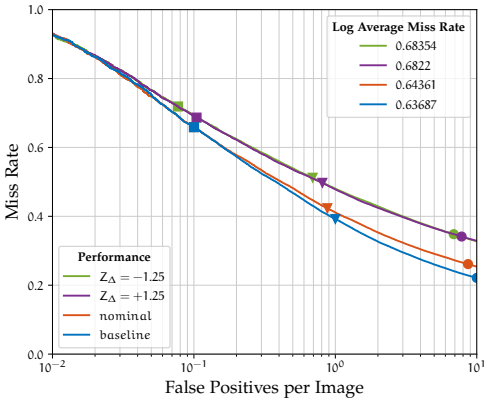


Figure 5: Miss Rate vs. FPPI evaluated on all car instances of the dataset. The marker indicate different operating points (i.e. score thresholds) at FPPI $= \{0.1, 1.0, 10.0\}$ standardized based on the baseline performance.

doesn't negatively influence the number of FPs generated by the system, especially for FPPI $= 0.1$ where the precision value is almost equal among all test cases. Only the test case with positive defocus offset shows a considerable drop in the precision value when shifting the score threshold towards unrealistic operating points. The recall is simply the complement of the MR, so the marker in figure 6 suggest analogous to the MR vs. FPPI curve in figure 5 that a worse image quality correlates with the number of ground truth instances missed by the system. Figure 6 also highlights the operating points where the largest performance differences between the baseline and the performance under effects of defocus with respect to the precision and recall values occur.

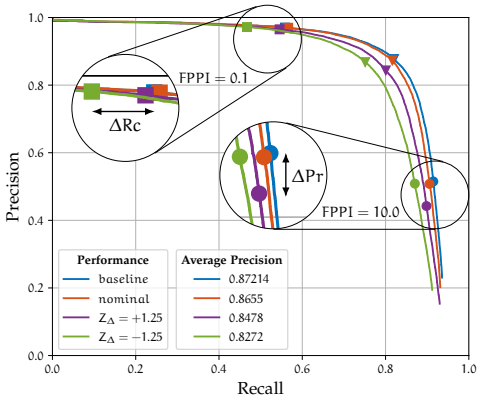
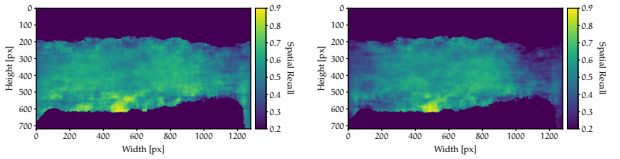


Figure 6: Precision vs. Recall evaluated on medium-sized instances of the dataset. Marker are set based on the operating points extracted from the Miss Rate vs. FPPI metric in figure 5.

How the performance drop is constructed is visualized in figure 7, where the baseline SRI is contrasted with the SRI under effects of defocus. Subtracting the latter from the former results in the $\mathrm{SRI}_{\mathrm{drop}}$ shown in figure 8 a. We show the spatial performance drop for these examples in comparison with the optical performance of the underlying lens model. Figure 8 a shows the $\mathrm{SRI}_{\mathrm{drop}}$ for $Z_\Delta = -1.25$ at FPPI $= 0.1$ and figure 8 c presents the $\mathrm{SPI}_{\mathrm{drop}}$ for $Z_\Delta = +1.25$ at FPPI $= 10$. The optical performance is shown alongside in figure 8 b and d, respectively.

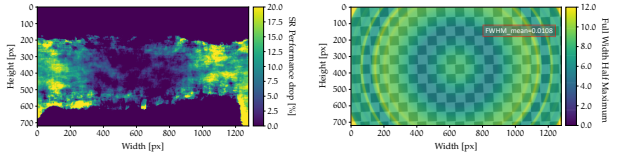
Note that for statistical purposes we compute the local per-



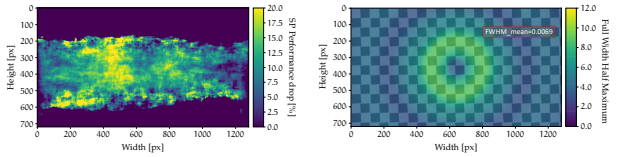
(a) SRI: Baseline (b) SRI: $Z_\Delta = -1.25$

Figure 7: Comparison of the baseline SRI and the SRI on the degraded dataset with defocus offset $Z_\Delta = -1.25$ for medium-sized car instances with an operating point standardized based on FPPI $= 0.1$ analogous to [9]. See resulting $\mathrm{SRI}_{\mathrm{drop}}$ in figure 8 a.

formance indices only for pixels that belong to at least 20 instances throughout the dataset. This results in the blue regions in the upper and lower parts of the heatmaps, where the number of medium-sized car instances is particularly low based on the ground truth distribution in figure 2 c.



(a) SRI drop: $Z_\Delta = -1.25$ (b) FWHM map: $Z_\Delta = -1.25$



(c) SPI drop: $Z_\Delta = +1.25$ (d) FWHM map: $Z_\Delta = +1.25$

Figure 8: Comparison of the SRI drop for FPPI $= 0.1$ and SPI drop for FPPI $= 10.0$ with the optical performance of the underlying lens model for defocus offsets $Z_\Delta = \pm 1.25$ as in [9].

The Full Width at Half Maximum (FWHM) maps visualize optical performance of the filtering, with higher values indicating a stronger blurring effect [9]. When comparing the FWHM maps with the spatial performance drop of the instance segmentation system, it can be seen that the axially symmetric behavior of the simulated lens with the respective defocus offsets is reflected both on the $\mathrm{SRI}_{\mathrm{drop}}$ and $\mathrm{SPI}_{\mathrm{drop}}$. The former is shown for the test case with defocus offset $Z_\Delta = -1.25$ where the drop in performance increases gradually towards the edges, similar to the optical performance of the images. The $\mathrm{SPI}_{\mathrm{drop}}$, on the other hand, is shown for the test case with defocus offset $Z_\Delta = +1.25$, where the largest performance drop occurs around the center of the images where the parameterized lens exhibits the worst optical performance.

Conclusion

In this work, we introduce with the SRI and SPI two metrics for the spatial performance assessment of instance segmentation systems. The former, SRI, was already introduced for object detection in [9] and is now adapted to instance segmentation. It assigns a recall value to each pixel of $h \times w$-sized images in the dataset. The latter, SPI, complements the SRI by assigning a local precision value to each pixel in $h \times w$.

Using the optical model from [9] with different parameterizations, we produce several datasets by degrading all labeled images of the BDD100K for instance segmentation. We evaluate the performance on these degraded datasets and contrast them with the baseline performance on the respective set of images in the original BDD100K dataset. The standard metrics show the expected effects on the overall performance through the image degradation with an increase in the LAMR and a decrease in the AP.

With the newly proposed spatial evaluation metrics, a correlation is observed between the spatial performance drop of an instance segmentation system and the spatially variant quality of images in the degraded datasets. The axially symmetric optical performances of the underlying, physically realistic, lens model with different defocus parameterizations (shown with FWHM maps) are comparable with the spatial performance drops evaluated with the SRI and SPI. The SRI shows a drop in performance caused by the degraded image quality even at a realistic operating point for deployment. This highlights the need for robustness tests of CV algorithms to naturally occurring lens aberrations by taking into account the spatial domain. For the SPI (and the numerical precision value), on the other hand, a drop in performance is only observed after lowering the score threshold to rather unrealistic operating points.

We restrict the spatial evaluation to instances of a certain area range with relatively equally distributed instances in order to reduce the dependency of the SRI and SPI on instance sizes. However, apart from our general suggestion to including the spatial domain into the performance assessment of CV algorithms, removing the dependency of the SRI and SPI on instance sizes entirely should be focus of future work.

References

[1] Zhaowei Cai and Nuno Vasconcelos. “Cascade R-CNN: High Quality Object Detection and Instance Segmentation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.5 (May 2021). Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1483–1498.

[2] Jesse Davis and Mark Goadrich. “The Relationship Between Precision-Recall and ROC Curves”. In: vol. 06. June 2006.

[3] Piotr Dollar et al. “Pedestrian Detection: An Evaluation of the State of the Art”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.4 (Apr. 2012). Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 743–761.

[4] Mark Everingham et al. “The Pascal Visual Object Classes (VOC) Challenge”. en. In: *International Journal of Computer Vision* 88.2 (June 2010), pp. 303–338.

[5] Christian Krebs, Patrick Müller, and Alexander Braun. “Impact of Windshield Optical Aberrations on Visual Range Camera Based Classification Tasks Performed by CNNs”. In: *London Imaging Meeting* 2021.1 (Sept. 20, 2021), pp. 83–87.

[6] Matthias Lehmann et al. “Modeling realistic optical aberrations to reuse existing drive scene recordings for au-

tonomous driving validation". In: *Journal of Electronic Imaging* 28.01 (2019), p. 1.

[7] Matthias Lehmann et al. "Resolution and accuracy of nonlinear regression of point spread function with artificial neural networks". In: *Optical Engineering* 58.04 (2019), p. 1.

[8] Tsung-Yi Lin et al. "Microsoft COCO: Common Objects in Context". en. In: *Computer Vision – ECCV 2014*. Ed. by David Fleet et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2014, pp. 740–755.

[9] Patrick Müller, Mattis Brummel, and Alexander Braun. "Spatial recall index for machine learning algorithms". In: *London Imaging Meeting* 2021.1 (Sept. 2021), pp. 58–62.

[10] Zachary Pezzementi et al. "Comparing Apples and Oranges: Off-Road Pedestrian Detection on the NREC Agricultural Person-Detection Dataset". In: (July 2017).

[11] Zachary Pezzementi et al. "Putting Image Manipulations in Context: Robustness Testing for Safe Perception". en. In: *2018 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. Philadelphia, PA: IEEE, Aug. 2018, pp. 1–8.

[12] Yuxin Wu et al. *Detectron2*. `https://github.com/facebookresearch/detectron2`. 2019.

[13] Fisher Yu et al. "BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. ISSN: 2575-7075. June 2020, pp. 2633–2642.

Author Biography

Mattis Brummel is a Master student in Electrical Engineering and Information Technology at the University of Applied Sciences in Düsseldorf. He received his Bachelor's degree from FH Bielefeld. His research interests cover Deep Learning and Computer Vision.

Patrick Müller received his B.Eng. in 2016 and his M.Sc. in 2018. His Master's thesis examined the influence of a Point Spread Function Model to Digital Image Processing algorithms. He is currently pursuing his PhD with a focus on the application of optical models to digital images, their validation, performance and correlation with the performance of Computer Vision algorithms.

Alexander Braun received his diploma in physics with a focus on laser fluorescent spectroscopy from the University of Göttingen in 2001. His PhD research in quantum optics was carried out at the University of Hamburg, resulting in a Doctorate from the University of Siegen in 2007. He started working as an optical designer for camera-based ADAS with the company Kostal, and became a Professor of Physics at the University of Applied Sciences in Düsseldorf in 2013, where he now researches optical metrology and optical models for simulation in the context of autonomous driving. He's member of DPG, SPIE and IS&T, participating in norming efforts at IEEE (P2020) and VDI (FA 8.13), and currently serves on the advisory board for the AutoSens conference.