Performance Limiting Factors of Deep Neural Networks for Pedestrian Detection

Yasin Bayzidi, Alen Smajic, Jan David Schneider, Fabian Hüger, Ruby Moritz, Alois Knoll

Technical University of Munich

Introduction

Deep Neural Networks (DNN) for perception in automated driving have been extensively studied, while achieving strong results in detection performance on pre-annotated test sets. However, there has been a gap in the literature on a systematic analysis of DNN behavior to investigate the factors contributing to their misbehavior. As part of DNN safety, we propose to both analyze DNN behavior in challenging scenarios as well as the respective factors that actually contribute to their misbehavior. Although some of such factors have been studied individually, there is not a thorough study to compare all together in a systematic manner to unveil the impact of each factor leading to DNN failures.

Figure 1: Examples of different factors contributing to the possible performance drop of the DNNs. The top row examples are from the KI-Absicherung dataset which include distance, wetness, fog and contrast factors. The bottom row examples are from the CityPersons dataset, which include brightness, occlusion, crowdedness and distance factors.

Method

Performance Limiting Factor (PLF): A factor is considered performance limiting, if the presence of the respective factor in the input data causes significant drop in detection accuracy of the DNNs, such as precision and f1 score for image level and recall for object level factors. However, one should notice that the performance of a DNNs might be low in specific PLF levels due the under-representation of the data in such levels, that would cause the DNNs not to converge well for such data, or over-fit to other levels that are more frequent in the training dataset. Therefore, we consider a factor as a PLF, if there exists a correlation of such a factor with the performance of the DNNs regardless of the frequency of such a factor in the training dataset.

Studied Factors

Image Intensity

1- Edge Strength
2- Boundary Edge Strength(ours)
3- Background Edge Strength(ours)
4- Contrast
5- Contrast to Background (ours)
6- Brightness
7- Foreground Brightness
8- Object Entropy

Meta Annotations

9- Lens Flare Intensity
10- Vignette Intensity
11- Fog Intensity
12- Daytime Type
13- Sky Type
14- Wetness Type
15- Occlusion Ratio
16- Truncated
17- Distance

Geometrical Properties

18- Crowdedness (ours)
19- Bounding Box Height
20- Bounding Box Aspect Ratio
21- Visible Instance Pixels

Experiments

Studied DNNs:
2D Object detectsion: 1) FasterRCNN, 2) FCOS, 3) RetineNet and 4) SSD 300
Semantic Instance Segmentation: 5) MaskRCNN
Key-point Detection: 6) KeypointRCNN
Datsets: 1) CityPersons & 2) KI-Absicherung

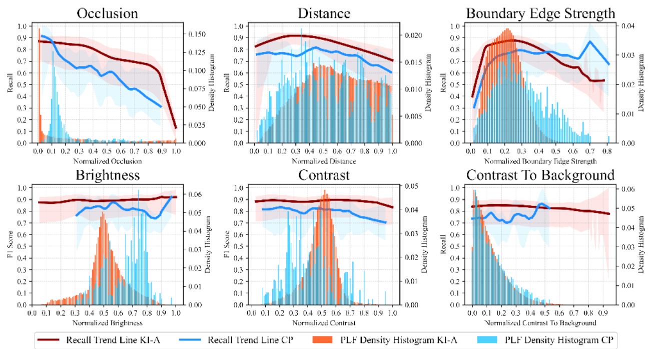

Figure 2: The correlation results of the three of the factors to be considered as PLF (top row) and three factors to not be considered as PLF (bottom row) averaged over all the six trained models on KI-Absicherung dataset and the five models trained on the CityPersons dataset. The x-axis represents the normalized values of the factors, and the y-axis the recall. The lines represent the local regression correlation of each factor and the according local performance. The histogram is calculated upon the frequency of each factor value in the respective dataset.

Results & Conclusions

The correlation coefficient results of all the studied factors are illustrated in the Figure 3. As the number of total factors extracted for the CityPersons dataset are lower than the KI-Absicherung dataset, a direct comparison of all the factors among the two datasets was not possible. Firstly, one can observe that there are various factors that have a very similar correlation coefficient for both of the datasets, e.g. occlusion, entropy, contrast, etc., while there are other factors that have a bigger difference, such as contrast to background, foreground brightness and edge strength. However, as the correlation coefficient is not the only metric to assess for the existence of a PLF, we also utilized a qualitative inspection of the graphs such as the ones illustrated in the Figure 2 to evaluate the factors for PLFs. Based on that, one can observe that some factors such as boundary edge strength can be considered as PLF despite their low correlation coefficient. This is only observable by such a qualitative inspection, where one can realize that there exists a strong correlation with the respective factor while having a general low correlation coefficient. Consequently, we selected the **occlusion, distance, boundary edge strength (ours), background edge strength (ours), height, crowdedness (ours), edge strength, fog intensity**, and wetness types as PLF, and disregarded all the others factors as non-PLFs. Based on this analysis, occlusion had the most significant effect among the selected metrics as PLF. However, we do not suggest to immediately reject the second group of factors before enhancing the training datasets with more samples of the respective factor to equalize the histogram distribution and re-assessing their effect on DNNs performance. This could possibly lead to alleviation of the effect of some of the factors that stem from dataset bias.

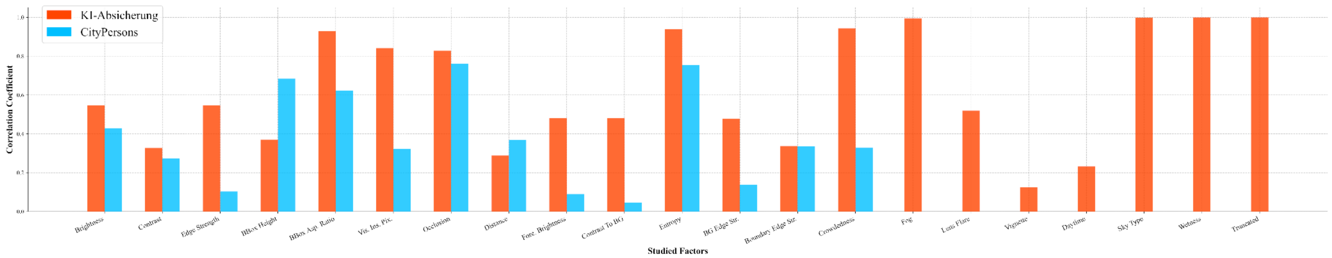

Figure 3: The results of the correlation coefficients calculated between the studied factors and the models recall values, averaged over all the trained models. X-axis include all the studied factors, while y-axis the according correlation coefficient values. The factors from left to right include brightness, contrast, edge strength, bounding box height, bounding box aspect ratio, visible instance pixels, occlusion ratio, distance, foreground brightness, contrast to background, object entropy, background edge strength, boundary edge strength, object crowdedness, fog intensity, lens flare intensity, daytime, sky type, wetness, and truncated. Orange bars represent the results from the KI-Absicherung dataset, and the blue bars represent the results from the CityPersons dataset.

Corresponding Author:
Yasin Bayzidi
yasin.bayzidi@volkswagen.de

Volkswagen AG

Technical University of Munich

KI ABSICHERUNG
Safe AI for Automated Driving