Performance Evaluation of Object Detection Algorithms Under Adverse Weather Conditions

Thomas Rothmeier(✉) and Werner Huber

Ingolstadt University of Applied Sciences, Ingolstadt, Germany
{thomas.rothmeier,werner.huber}@thi.de

Abstract. Camera systems capture images from the surrounding environment and process these datastreams to detect and classify objects. However, these systems are prone to errors, often caused by adverse weather conditions such as fog. It is well known that fog has a negative effect on the camera's view and thus degrades sensor performance. This is caused by microscopic water droplets in the air, that scatter light, reduce contrast and blur the image. Object detection algorithms show severely worse performance and high uncertainty when exposed to fog. However, they need to work safe and reliable in all weather conditions to enable full autonomous driving in the future. This work focuses on the evaluation of several state-of-the-art object detectors in normal and foggy environmental conditions. It is shown that the detection performance deteriorates considerably when exposed to fog. Further, the results suggest that some algorithms are more robust towards fog than others.

Keywords: Object detection · Adverse weather · Autonomous driving

1 Introduction

One of the greatest challenges still remaining to enable fully autonomous vehicles, is the ability to drive safely and reliably even in low visibility conditions. Safety systems rely on data from surround sensors to correctly perceive their environment. Although most sensors perform well in good visibility conditions, their performance degrades extremely during adverse weather conditions such as rain, fog and snow. Hasirlioglu et al. showed that the effects of rain on camera, lidar and radar sensors degrade sensor performance [6,8]. Reway et al. evaluated camera-based object detection by simulating various environmental conditions [19].

Camera-based object detection is of high importance when it comes to vehicle safety. Currently, the camera is the only sensor that can reliably interpret a situation, due to its ability to recognize the semantic of an object. Early, accurate

This work is supported under the FH-Impuls program of the German Federal Ministry of Education and Research (BMBF) under Grant No. 13FH7I01IA.

A. L. Martins et al. (Eds.): INTSYS 2020, LNICST 364, pp. 211–222, 2021.
https://doi.org/10.1007/978-3-030-71454-3_13

and reliable detection results can prevent traffic accidents and thus save lifes. In the last few years huge progress has been made in the field of computer vision, mostly induced by the success of neural networks for object detection tasks. However, these novel algorithms are not yet able to completely mimic human perception. Furthermore, most of the algorithms have been trained under good weather conditions, which leads to problems and high uncertainty under poor visibility conditions.

Poor visibility can occur in fog, for example. Fog consists of microscopic water droplets in the air that scatter light and lead to reduced contrast, color saturation and less precision in contours and details. These effects grow with increasing distance to the object and lead to an overestimation of the distance to the object ahead [13,21].

In this work we focus on performance evaluation of camera-based object detection under foggy weather conditions. We want to quantify the extent to which the uncertainty of the predictions and the localisation error change with increasing distance to the object. Furthermore we aim to give a comparison of different object detectors in dense fog. The evaluation is based on data recorded in the adverse weather chamber of CARISSMA on which a car attrap moves from the camera sensor.

Outline. This paper is organised as follows: Sect. 2 gives an overview of related work towards object detection in adverse weather. Section 3 describes our experimental setup, data preparation and evaluation methods. Section 4 shows the results of the object detectors on our test scenario, while Sect. 5 summarizes our contribution and discusses limitations and future work.

2 Related Work

In recent years, neural networks, especially Convolutional Neural Networks (CNNs), have increasingly emerged as the standard for object detection and object recognition. There is a large variety of object detectors that use different features and sensor data as input. However, most of them share similar basic concepts that can be roughly divided into one-stage-detectors and two-stage-detectors. The following section will give a short overview of object detectors and shortly summarize their underlying concepts.

An example for a two-stage-detector is Faster R-CNN [18]. In the first stage a feature map is given to a Region Proposal Network (RPN) that proposes regions of interest that might contain an object. In the second stage the proposed regions of the RPN are then classified by another network layer. Since two network passes are required, these detectors are slower than networks that predict location and class in a single step.

In order to address this issue other architectures were proposed that only require one network pass, so called one-stage-detectors. Well known examples are YOLOv3 [17], Single-Shot-Detector (SSD) [12] and RetinaNet [10]. They

differ from two-stage-detectors as they omit the region proposal stage and therefore predict bounding boxes and object classes in a single network run. This leads to faster inference times than with two-stage-detectors. SSD uses multiple feature maps at different scales, to predict bounding boxes and classify objects at different size. Similarily YOLOv3 uses a feature pyramid to extract features from 3 different scales. It is trimmed for fast inference time. RetinaNet introduces a new focal loss, that aims to tackle to problem of class imbalance encountered during training of dense detectors.

Although there is lots of progress in the field of object detection, adverse weather conditions still pose a huge problem. In [1,5,23] techniques for fog removal in images were proposed in order to recover scene contrasts and thus improve the overall image quality as a pre-processing step for object detectors.

Furthermore, approaches for the detection of fog in images are being researched. The only reliable information in images with fog is loss of contrast and blurring of the image. However, information about presence and density of fog could help to decrease uncertainty in object detection. Pavlic et al. proposed an approach to detect the presence of fog in images and classify it using Gabor Filters [14,15].

Volk et al. present a method to improve object detection algorithms by augmenting training data with synthetic rain variations [22]. Hnewa et al. tested YOLOv3 and Faster R-CNN under clear and rainy weather conditions and gave an overview of promising approaches to improve object detection under rainy weather conditions [9].

Reway et al. showed the drop of camera sensor performance in different daytime and weather conditions using a virtual simulation [19]. Therefore a real camera is placed in front of a high resolution monitor that films the simulated environment. Hasirlioglu et al. evaluated in [7] the performance of a camera sensor mounted inside a car with raindrops on the windshield. They measured, how these raindrops affect the performance of two different object detectors in between one wiping action. They showed that false detections increase proportionally with the amount of raindrops on the windshield. In [8] the performance of camera, radar and lidar sensors were assessed. Here, an adverse weather facility was used which is capable of simulating reproducible rain with different intensities.

Object detection does not solely rely on the camera sensor for detecting objects. An automated vehicle e.g. is equipped with a set of different sensors like camera, lidar and radar. These redudant sensor data can be used in fusion architectures for object detection. Pfeuffer et al. introduced a data fusion architecture based on deep neural networks that unifies the sensor streams to improve detection capabilities [16]. Bijelic et al. [2] collected a large set of data for camera and lidar and proposed a deep multimodal sensor fusion approach for improving object detection in bad weather.

In this work we focus on the performance evaluation of camera-based object detection algorithms in dense fog, as automotive camera sensors are cheap and are already widely used in existing cars. In particular, the contribution of our

work is to present a test method to measure the performance of object detectors in fog. Furthermore we contribute by evaluating several object detectors with the presented test method and show that the performance of different object detectors varies under the same environmental conditions.

3 Method and Materials

In this section we describe the experimental setup of our test scenario, the recorded dataset, the object detection algorithms and the evaluation metrics that were applied. We will speak of normal weather conditions when no fog is present.

3.1 Experimental Setup

In order to compare the performance of object detectors in different weather conditions we prepared a dynamic test scenario. Videos were recorded with a standard automotive camera with a resolution of 2 megapixels at a frame rate of 24 frames per second.

We recorded our test data in the indoor test facility of CARISSMA which is capable of simulating dense fog up to a human visual range of 20m. A standardized Euro NCAP Vehicle Target (EVT) [20] was positioned in front of the camera with a distance of 1m. The EVT is placed on a unmanned vehicle platform that is constantly moving away from the camera at a speed of 20 km/h over a distance of 50 m. Figure 1 shows the experimental setup with and without fog. The scenario mimics a highway scene with a car moving away from the camera sensor. The same scenario was recorded for five times under normal conditions and five times with dense fog. The EVT is the only object visible in the camera's field of view (FOV).

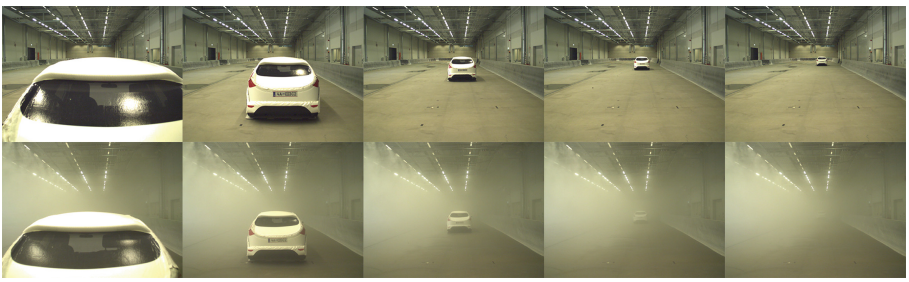

Fig. 1. Image sequence of the EVT moving away from the camera sensor. The sequence on top shows the test setup under normal conditions. The sequence on bottom shows the test setup with dense fog. Between the shown images from left to right are 55 frames each.

3.2 Dataset Preparation

As a dataset we used a set of ten videos recorded in the indoor test facility of CARISSMA. Half of the videos were recorded in normal weather conditions, while the other half was recorded with fog. All videos were edited to match the point in time when the EVT starts and stops. From the edited videos, we took 220 frames per video, where the first frame is always the point in time when the vehicle begins to move. In total, we considered 2200 frames, 1100 for normal and foggy conditions each. Each frame was hand labeled with a bounding box enclosing the EVT. The size of each frame was downscaled to a resolution of 800×600 pixels.

3.3 Object Detection

For the evaluation of the object detection algorithms under normal and foggy environmental conditions we chose four object detection algorithms: Faster R-CNN, SSD, YOLOv3 and RetinaNet. These algorithms are all capable of detecting objects in real time and with high accuracy. Each of them uses a pre-trained weight file trained on the COCO dataset. The COCO training split contains 118.000 images for training with 80 different object categories [11].

We have chosen different variants for each algorithm, which differ in training time and training image size. This gives us a better insight into how the algorithms behave and how training images and training time affect performance. For SSD we chose SSD-300 and SSD-512 and for YOLOv3 the variants YOLOv3-416 and YOLOv3-spp. They only differ in training image size. For the algorithms RetinaNet and Faster R-CNN, we have selected two variants that differ in terms of training time marked as 1x and 3x. The 3x variants were trained three times as long as the 1x variants.

Each object detector was executed on each image from the data set and predicted a confidence score, a class and an associated bounding box. The detection threshold was set to 0.1. This means that each result with a predicted confidence score of greater than the value of 0.1 was saved to a file for further processing.

3.4 Evaluation Metrics

In order to evaluate the object detectors against each other we considered several object detection metrics from the PASCAL VOC Challenge [4]. We considered the Intersection over Union (IoU), also known as the Jaccard similarity coefficient. It is defined by

$$IoU = \frac{area(a \cap b)}{area(a \cup b)} \tag{1}$$

where $a \cap b$ denotes the intersection of the predicted box with the ground truth box and $a \cup b$ the union of their bounding boxes. IoU is a measure for the accuracy of an object detector's predicted bounding boxes.

Furthermore we plot a Precision-Recall-Curve for each algorithm. It is a plot of precision and recall for ranked confidence scores. Precision is a measure for the accuracy of an object detector, whereas recall measures the the amount of returned results, also called sensitivity. Precision and Recall are defined by

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{2}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{3}$$

where a True Positive (TP) is defined as a correct detection with $IoU > t$. A False Positive (FP) is either a wrong detection or a detection with $IoU < t$. The IoU threshold value t is defined as the value above which we consider the IoU of a bounding box to be sufficiently correct. A False Negative (FN) is a ground truth that was not detected due to a low confidence score.

Additionally, we calculate the Average Precision (AP) for every object detector which is the estimated area under curve of a Precision-Recall-Curve. It is defined by

$$AP = \sum_{n=0} (r_{n+1} - r_n) p_{interp}(r_{n+1}) \tag{4}$$

The interpolated precision $p_{interp}(r_{n+1})$ is defined by taking the maximum precision at each recall level r, where the corresponding recall value is greater than r_{n+1}. It is defined by

$$p_{interp}(r_{n+1}) = \max_{\tilde{r}:\tilde{r} \geq r_{n+1}} p(\tilde{r}) \tag{5}$$

4 Results and Discussion

In this section we present the results of the object detection algorithms on our recorded dataset. First, we will investigate how increasing distance affects detection results and the predicted bounding boxes. Then we look at precision and recall, plot a Precision-Recall-Curve and calculate the AP for each algorithm. It is to note, that we do not consider inference time in our evaluation.

4.1 Detection over Time

Confidence Score. For the evaluation of the detection capabilities we run each object detector on every frame in our dataset and save the predicted confidence scores, classes and bounding boxes. The evaluation results of all algorithms under normal and foggy environment conditions can be seen in Fig. 2. It shows the confidence scores of the respective algorithms on the left side and the IoU on the right side. For each value in the graph the average of the respective five recordings was taken.

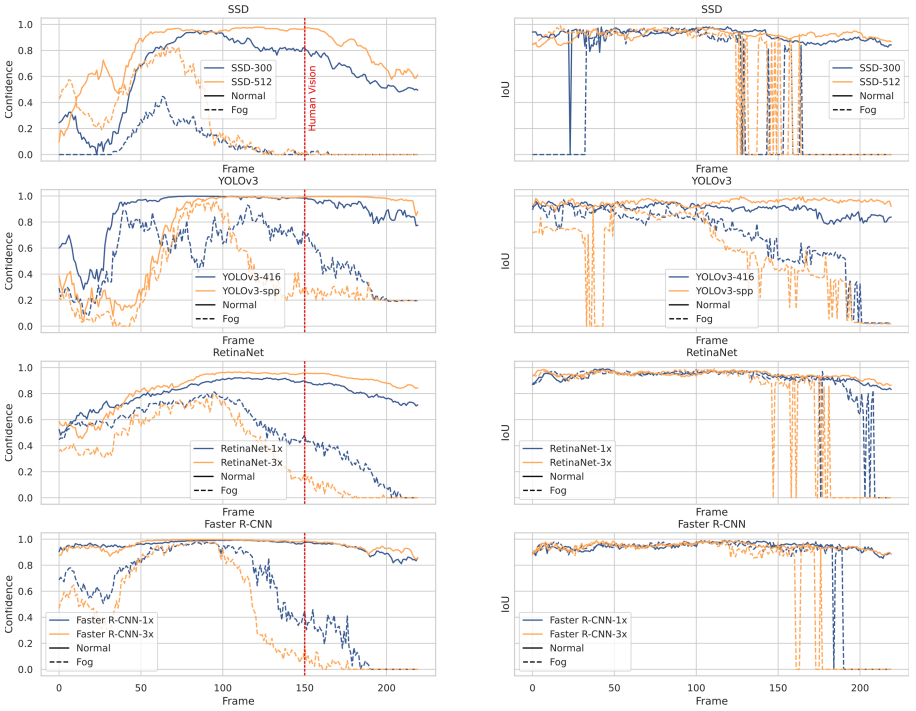

Fig. 2. Evaluation results of object detection for the dynamic scenario. The plots on the left show the confidence score for each frame and algorithm. The plots on the right show the IoU for each frame and algorithm. The values of confidence and IoU are averaged over five recordings each. (Color figure online)

The distance to the EVT increase with the number of frames. It should be noted that the frames do not accurately reflect the actual distance driven, as the unmanned vehicle platform may have minimal inaccuracies in the trajectory and acceleration.

Regarding Fig. 2 it is clearly visible that the object detectors show high performance under normal conditions. For most frames the confidence scores are higher than for the respective scenario with fog. We found that almost all algorithms - except of Faster R-CNN - have problems to detect the EVT within the first 30 frames, when the EVT is not completely visible in the camera's FOV. This is indicated by a low confidence score at the beginning and consequently higher uncertainty about the object class. This effect increases further with the presence of fog and leads to missed detections for the case of SSD-300.

Regarding the object detectors under foggy conditions, it can be seen that the confidence scores start to decrease with increasing distance. The fog particles scatter light and lead to reduced contrast and unsharp contours. This leads to a degradation in the object detector's performance.

However, there are also measureable differences among the algorithms. While YOLOv3, Faster R-CNN and RetinaNet show decreasing confidence scores around frame 100, the SSD algorithm's score decreases already at frame 60. Also the SSD algorithm is unable to detect the EVT at all after frame 120. The rest of the algorithms are still capable of detecting the EVT until frame 190, although with a low confidence score near zero. YOLOv3 managed to detect the car in one video for all frames, therefore it does not drop below a value of 0.2 in the evaluation.

For the authors of this paper the EVT is still clearly visible as a car up to frame 150. From this point on it becomes more difficult to recognize the EVT, but it is still often recognizable until the last frame. The human vision boundary for fog is marked in Fig. 2 as red line.

Intersection over Union. In the previous section we evaluated the confidence scores over time. However, a high confidence value alone is not sufficient for object detection. Therefore we evaluate the quality and correctness of the predicted bounding boxes in this section. The results of the comparison of IoU can be seen in Fig. 2 on the right side.

In normal conditions the IoU for each algorithm is constantly above a value of 0.8. Hence, all algorithms can detect the location of the EVT with high accuracy in normal environmental conditions. We cannot even notice a decrease in performance with increasing distance.

For the fog environment we also see very accurate bounding box predictions. It is to note that an IoU value of 0 was chosen when there is no bounding box predicted for a frame. For SSD, RetinaNet and Faster R-CNN the IoU value stays above a value of 0.8 as long as there exists a prediction. If no detection result is available, this can be recognized by outliers in the curve that drop to a value of 0. We can note, that even in foggy conditions the bounding boxes are highly accurate, even when the size of the EVT becomes small.

For YOLOv3 algorithm we see a steady decrease in IoU starting at frame 100. It seems to have problems to detect the correct boundaries of the EVT when the image contours become blurred and contrast decreases.

4.2 Overall Detection Capabilities

In this section we analyze the overall detection capabilities of each algorithm in normal and foggy environmental conditions. The results of our analysis can be seen in Fig. 3. We calculate precision and recall and plot a Precision-Recall-Curve. As IoU threshold we have chosen a value of 0.5.

All of the object detectors have very high precision and recall in the normal environment. This can be seen from the high curve, which only begins to fall at a high recall value. However, the same object detectors show significant differences when tested on the foggy dataset. All algorithms show a high drop in performance compared to normal weather conditions.

Table 1 shows the AP for each detector under foggy and normal environmental conditions. The highest performing algorithm is RetinaNet-1x. It shows the

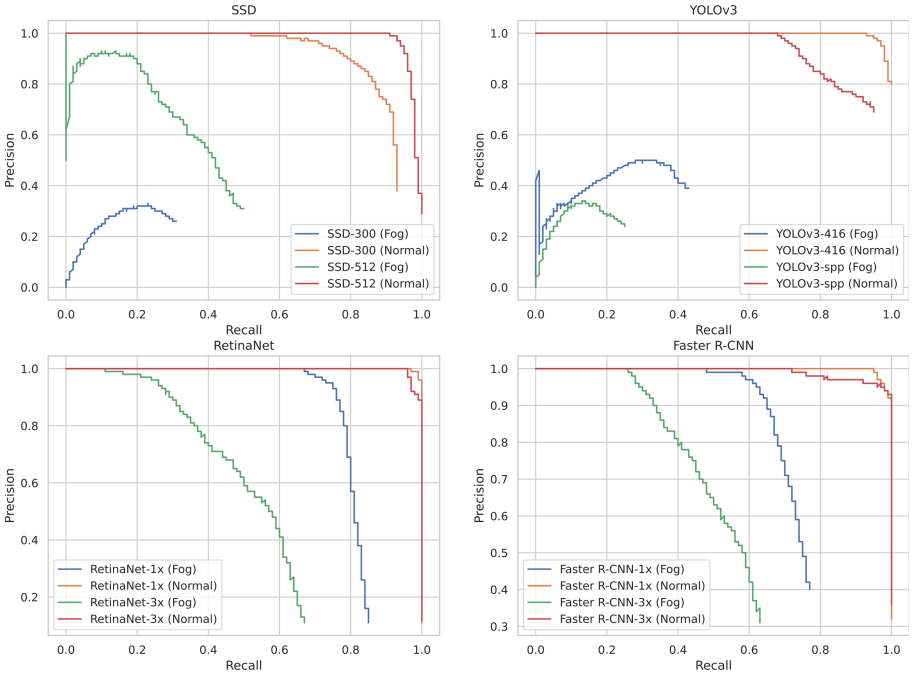

Fig. 3. Precision-Recall-Curves for the dynamic object detection scenario under normal and foggy environmental conditions.

best overall results on our dataset with an AP of 0.807 on the foggy dataset. It is only by a value of 0.192 worse than under normal weather conditions. SSD and YOLOv3 show a decline of more than 0.6 in AP compared to normal environmental conditions.

RetinaNet and Faster R-CNN are the most promising detectors. We also found that the models with less training cycles (1x) show higher AP than the corresponding models with more training cycles. RetinaNet-1x has an AP that is 0.292 higher than RetinaNet-3x. We see a similar behavior for Faster R-CNN. Those models could be less overfit to training data and thus be more unbiased towards environmental conditions.

For SSD we see that the model with larger training image size (SSD-512) performs better than the one with smaller image size (SSD-300). Here, the SSD-300 model seems to underfit as its performance is worse for the normal and the foggy case. YOLOv3 shows a contrary behaviour: The model with larger training image size (YOLOv3-spp) has worse performance than the model with smaller training image size (YOLOv3-416). Further research would be necessary at this point to clarify this behavior.

While we see very high AP for all models under normal conditions, there is still a large gap between the algorithms when exposed to bad visibility conditions. Our evaluation indicates, that different models are suited better to run in foggy

environmental conditions and that training time and training parameters also have an effect on the performance under adverse weather conditions.

Table 1. The AP for algorithms under normal and foggy conditions.

	AP (Normal)	AP (Fog)	Difference
SSD-300	0.893	0.080	0.813
SSD-512	0.983	0.361	0.622
YOLOv3-416	0.996	0.181	0.785
YOLOv3-spp	0.907	0.068	0.839
RetinaNet-1x	**0.999**	**0.807**	**0.192**
RetinaNet-3x	0.997	0.515	0.482
Faster R-CNN-1x	0.998	0.728	0.270
Faster R-CNN-3x	0.992	0.526	0.466

5 Conclusion

In this work we performed a dynamic indoor test under controllable environmental conditions. We recorded data with and without fog and evaluated various object detectors on it. The results were first analysed time-based and then examined for their general performance.

We showed that our tested object detection algorithms show high confidence and IoU in the scenario without fog. Even with increasing distance the overall performance is still high. In contrast, the object detectors in fog show strongly reduced performance. The uncertainty about the existence of the EVT raises with increasing distance to it.

However, while we see a general decline in performance, we have noticed that the algorithms also show strong differences among themselves when exposed to fog. We could show that RetinaNet and Faster R-CNN have a generally higher accuracy and sensitivity than the other tested algorithms. In addition, we found better results when the algorithms were trained for less time in the case of RetinaNet and Faster R-CNN. We explain this by the fact that the algorithms are less biased due to less training time and therefore also recognize blurred and low-contrast structures in the image as a vehicle. With these results we contribute towards testing object detection algorithms in adverse weather conditions.

The results of this work are limited in that we consider a very simple scenario with only one object. Furthermore, only the performance in dense fog is considered, other environmental influences like rain or snow are neglected. However, it should be noted that the evaluation method was deliberately kept simple in order to obtain representative results. New, and more complex scenarios can be designed and researched based on the test method and evaluation approach presented in this work.

Future work should adapt the training data to take weather conditions into account. In addition, the algorithms should be made more independent of weather domains. A first approach was already explored in the work of Chen et al., where they proposed a domain adaptive Faster R-CNN which is more robust towards different domains [3].

References

1. Berman, D., Treibitz, T., Avidan, S.: Non-local image dehazing. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1674–1682. IEEE, Las Vegas, June 2016
2. Bijelic, M., et al.: Seeing through fog without seeing fog: deep multimodal sensor fusion in unseen adverse weather. arXiv:1902.08913 [cs], February 2020
3. Chen, Y., Li, W., Sakaridis, C., Dai, D., Van Gool, L.: Domain adaptive faster R-CNN for object detection in the wild. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3339–3348. IEEE, Salt Lake City, June 2018
4. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The Pascal Visual Object Classes (VOC) challenge. Int. J. Comput. Vis. **88**(2), 303–338 (2010)
5. Fattal, R.: Single image dehazing. ACM Trans. Graph. **27**(3), 1–9 (2008)
6. Hasirlioglu, S., Kamann, A., Doric, I., Brandmeier, T.: Test methodology for rain influence on automotive surround sensors. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), pp. 2242–2247, November 2016
7. Hasirlioglu, S., Reway, F., Klingenberg, T., Riener, A., Huber, W.: Raindrops on the windshield: performance assessment of camera-based object detection. In: 2019 IEEE International Conference on Vehicular Electronics and Safety (ICVES), pp. 1–7, September 2019
8. Hasirlioglu, S., Riener, A.: Challenges in object detection under rainy weather conditions. In: Ferreira, J.C., Martins, A.L., Monteiro, V. (eds.) INTSYS 2018. LNICST, vol. 267, pp. 53–65. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-14757-0_5
9. Hnewa, M., Radha, H.: Object detection under rainy conditions for autonomous vehicles (2020)
10. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollér, P.: Focal Loss for Dense Object Detection. arXiv:1708.02002 [cs], February 2018
11. Lin, T.Y., et al.: Microsoft COCO: Common Objects in Context. arXiv:1405.0312 [cs], February 2015
12. Liu, W., et al.: SSD: Single Shot MultiBox Detector, vol. 9905, pp. 21–37. arXiv:1512.02325 [cs] (2016)
13. Oakley, J., Satherley, B.: Improving image quality in poor visibility conditions using a physical model for contrast degradation. IEEE Trans. Image Process. **7**(2), 167–179 (1998)
14. Pavlić, M., Belzner, H., Rigoll, G., Ilić, S.: Image based fog detection in vehicles. In: 2012 IEEE Intelligent Vehicles Symposium, pp. 1132–1137, June 2012
15. Pavlic, M., Rigoll, G., Ilic, S.: Classification of images in fog and fog-free scenes for use in vehicles. In: 2013 IEEE Intelligent Vehicles Symposium (IV), pp. 481–486, June 2013

16. Pfeuffer, A., Dietmayer, K.: Optimal Sensor Data Fusion Architecture for Object Detection in Adverse Weather Conditions. arXiv:1807.02323 [cs], July 2018
17. Redmon, J., Farhadi, A.: YOLOv3: An Incremental Improvement. arXiv:1804.02767 [cs], April 2018
18. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv:1506.01497 [cs], January 2016
19. Reway, F., Huber, W., Ribeiro, E.P.: Test methodology for vision-based ADAS algorithms with an automotive camera-in-the-loop. In: 2018 IEEE International Conference on Vehicular Electronics and Safety (ICVES), pp. 1–7. IEEE, Madrid, September 2018
20. Sandner, V.: Development of a test target for AEB systems, p. 7 (2013)
21. Schechner, Y., Narasimhan, S., Nayar, S.: Instant dehazing of images using polarization. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, vol. 1, p. I, December 2001
22. Volk, G., Müller, S., von Bernuth, A., Hospach, D., Bringmann, O.: Towards robust CNN-based object detection through augmentation with synthetic rain variations. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp. 285–292 (2019)
23. Xu, Z., Liu, X., Chen, X.: Fog removal from video sequences using contrast limited adaptive histogram equalization. In: 2009 International Conference on Computational Intelligence and Software Engineering, pp. 1–4 (2009)