Time to Shine: Fine-Tuning Object Detection Models with Synthetic Adverse Weather Images

Thomas Rothmeier
University of Applied Sciences Ingolstadt
thomas.rothmeier@thi.de

Werner Huber
University of Applied Sciences Ingolstadt
werner.huber@thi.de

Alois C. Knoll
Technical University of Munich
k@tum.de

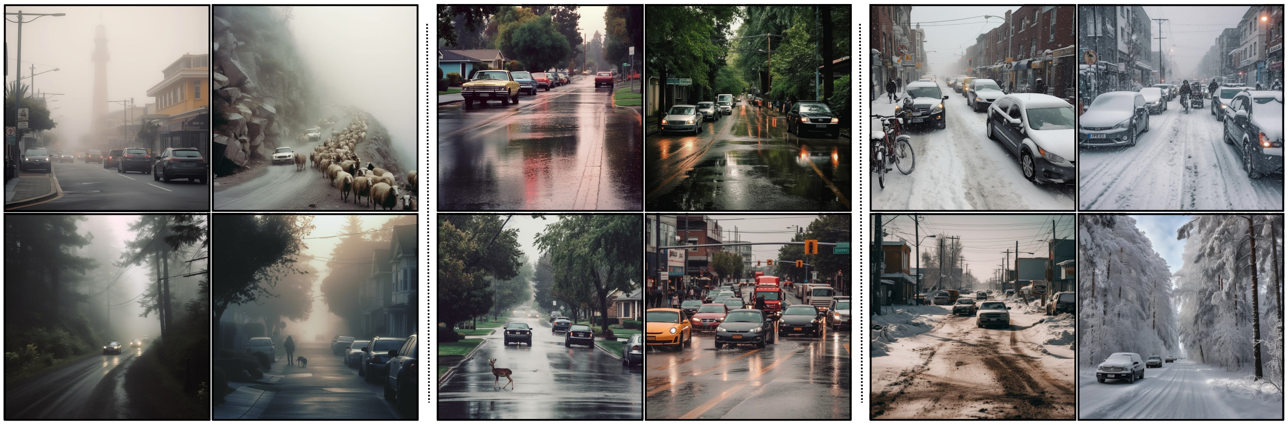

Figure 1. Images generated using the text-to-image model Midjourney. On the left, we present images depicting strong fog conditions. In the middle, we showcase images with intense rain conditions. On the right, we display images capturing the scenario of heavy snowfall.

Abstract

The detection of vehicles, pedestrians, and obstacles plays an important role in the decision-making process of autonomous vehicles. While existing methods achieve high detection accuracy under good environmental conditions, they often fail in adverse weather conditions due to limited visibility, blurred contours, and low contrast. These ”edge-case” scenarios are not well represented in existing datasets and are not handled properly by object detection algorithms. In our work, we propose a novel approach to synthesising photorealistic and highly diverse scenarios that can be used to fine-tune object detection algorithms in adverse weather conditions such as snow, fog, and rain. The approach uses the Midjourney text-to-image model to create accurate synthetic images of desired weather conditions. Our experiments show that training with our dataset significantly improves detection accuracy in harsh weather conditions. Our results are compared to baseline models and models fine-tuned on augmented clear weather images.

1. Introduction

The rapid progress in autonomous vehicle technology has brought us closer to a future of safe and efficient transportation. To achieve reliable autonomous driving, accurate object detection algorithms are essential, enabling vehicles to perceive and understand their surroundings. Creating high-quality training datasets plays a crucial role in the development of these algorithms, ensuring their ability to detect objects even under extreme environmental conditions.

Adverse weather conditions pose significant challenges for autonomous driving systems, as they can severely impact visibility and introduce dynamic environmental factors. Object detection algorithms must be robust and adaptable to effectively detect and recognise objects in these challenging scenarios. Current computer vision models rely on large-scale, diverse datasets [4–7, 9, 40, 46] to effectively learn robust features that enable detection in limited visibility conditions. Therefore, it is crucial to train algorithms using datasets that comprehensively capture the complexities and variations associated with adverse weather conditions. However, collecting these datasets can be expensive and time-consuming, especially for the adverse weather domain. A total of 1.4 million frames were collected in the dense project [4], yet only a minor amount of these frames include adverse conditions (2.3% in fog and 1.6% in rain). As a result, current large-scale image datasets used for training algorithms in the field of autonomous driving are heavily biased towards clear weather conditions (see Table 1).

The advancement of synthetic data generation techniques has revolutionised the field of computer vision, enabling the creation of highly diverse and photorealistic images at an unprecedented scale. Among these techniques, text-to-image models have emerged as powerful tools for generating high-quality images based on textual descriptions [31]. In the realm of autonomous vehicles, where training data plays a vital role in developing robust object detection algorithms, synthetic data generation has become a promising avenue for creating comprehensive and versatile datasets.

In this paper, we present a method to create camera image datasets to fine-tune object detection algorithms in a wide variety of different environmental conditions. By leveraging the generative AI Midjourney (MJ) [1], we were able to collect a diverse range of camera images that encompassed foggy, snowy, and rainy conditions, replicating the challenges faced by autonomous vehicles operating in adverse weather. The dataset includes camera images obtained from multiple viewpoints. This multi-perspective approach enables a comprehensive understanding of the objects present in the environment, facilitating accurate detection and localisation.

The integration of images synthesised in adverse weather domains into the training process offers promising results across multiple object detection models. Fine-tuning object detection models using the synthesised dataset allows the algorithms to adapt and learn from the complexities and variations present in adverse weather scenarios. By exposing the models to a wide range of adverse weather conditions, including fog, snow, and rain, they become more adept at detecting and recognising objects under challenging visual circumstances. In this work we show that the fine-tuned object detection models exhibit improved performance in adverse weather conditions compared to their counterparts trained solely on traditional datasets.

Contribution: This work makes a contribution by generating and annotating a synthetic adverse weather dataset using Midjourney. Our evaluation demonstrates the utility of synthetic adverse domain data for fine-tuning object detection models that were pre-trained on the coco dataset. We introduce image corruption models and evaluate their effectiveness for fine-tuning purposes. To ensure a comprehensive evaluation, we test the models on various publicly available benchmark datasets in adverse domains, as well as curate a dataset containing more intense weather conditions. Ablation studies are conducted to further validate our findings and strengthen the significance of our results.

Dataset	Clear	Fog	Snow	Rain
Waymo [40]	*	*	*	*
Argoverse [6]	*	*	*	*
Dense [4]	47.2	14.7	35.9	2.1
ACDC [38]	25.0	25.0	25.0	25.0
NuScenes [5]	80.6	0.0	0.0	19.4
Cityscapes [7]	100.0	0.0	0.0	0.0
KITTI [9]	100.0	0.0	0.0	0.0
BDD100K [46]	82.8	0.2	8.9	8.0
Ours	0.0	33.3	33.3	33.3

Table 1. Amount of adverse weather images in large-scale image datasets for autonomous driving in percent. '*' indicates that no emphasis was put into adverse weather labelling by the creators.

2. Related Work

Object Detection in Adverse Weather. Detecting objects in adverse weather conditions poses a significant challenge in the field of automated vehicles [11, 21, 22]. Over the past decade, extensive efforts have been dedicated to enhancing object detection algorithms for such scenarios [20, 21, 23, 42, 47] and to enhancing image quality by dehazing and de-raining [26, 44, 45]. However, a key limitation lies in the existing datasets, which predominantly exhibit a bias towards favourable weather conditions and lack the inclusion of severe disturbance factors like fog, snow, or rain. Although recent endeavours have aimed to address this imbalance by collecting more balanced datasets, their scale remains relatively small, hindering comprehensive analysis and evaluation [4, 37, 38]. Other approaches that has garnered attention is the fusion of multimodal sensor streams to enhance object detection capabilities [4]. By integrating information from diverse sensors, such as cameras, lidar and radar, a more holistic understanding of the surrounding environment can be achieved, potentially improving object detection. In our work we aim to improve camera-based detection algorithms by extending the underlying database with diverse synthetic data in a variety of adverse weather conditions. [8]

Adverse Weather Augmentation. Simulating adverse weather conditions on real images originally captured under clear weather conditions is a promising research direction. Within this area, researchers have explored two primary approaches. The first approach involves utilising models grounded in real-world physics to realistically simulate weather effects [30, 37, 42, 43]. The second approach focuses on domain adaptation techniques, which aim to shift the data distribution of clear weather images towards adverse weather conditions [14, 16, 24, 27–29, 35]. These techniques enable the adaptation of labeled clear weather data to challenging bad weather scenes, enhancing the performance of object detection algorithms in adverse weather conditions. While these approaches show potential, training models that can effectively simulate adverse weather conditions remains challenging due to the scarcity of weather-disturbed data. In our work we address the limitations of existing datasets by generating highly diverse artificial driving scenarios in varying weather conditions using Midjourney. Further, we apply simple augmentation techniques to render snowflakes, rain-streaks [21] and raindrops [29] on top of these images.

Synthetic Image Generation. The generation of fully synthetic images has been a topic of extensive research, particularly within the realm of Generative Adversarial Networks (GANs) [10]. Notably, recent advancements in text-to-image models, specifically those based on diffusion models [39], have demonstrated impressive results [13, 31]. Models such as DALL-E [32] and Midjourney [1] have shown exceptional capabilities in generating highly realistic images based on textual descriptions. Text-to-image models have also been utilised for improving semantic segmentation [3] and image in-painting and editing tasks [2, 15, 25]. In our work, we leverage the Midjourney text-to-image model to create real-world photorealistic driving scenarios specifically focused on adverse weather conditions. To the best of our knowledge, we are the first to utilise Midjourney in generating synthetic driving scenarios with the primary objective of enhancing object detection algorithms.

3. Synthetic Adverse Weather Dataset

In order to improve the performance of object detection models in adverse weather conditions, we have generated a diverse dataset using Midjourney consisting of more than 18.000 images. Our dataset encompasses a wide range of adverse weather scenarios, including fog, rain, and snow. A subset of 1538 images was annotated with precise object labels. It is to note, that only cars are labeled in the dataset, despite other objects (e.g. pedestrians, buses, traffic signs) are most often visible in the scene.

3.1. Dataset Creation

Midjourney. Midjourney [1] is an advanced generative AI model that excels in converting natural language prompts into highly realistic images. Leveraging the power of Midjourney, it becomes feasible to generate high-quality images from simple text-based descriptions. In the context of our research, we identified a scarcity of datasets that adequately represent edge-case scenarios in adverse domains. To address this gap, we used Midjourney to generate synthetic driving scenarios that closely resemble real-world conditions (see figure 1). In this work we utilise version 5 and version 5.1 of Midjourney. While specific implementation details are not disclosed by Midjourney, it is worth noting that earlier versions, such as version 4, were based on stable diffusion techniques. Consequently, it is reasonable to assume that some form of diffusion process is employed in generating the results. We decided against open source alternatives like Stable Diffusion [34], since at the time of writing the quality of generated adverse weather images was considerably worse here.

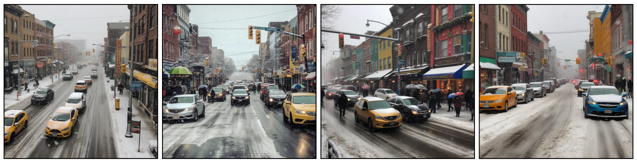

Figure 2. Multiple variations of an image depicting snowy weather conditions, all generated using the same text prompt.

Text prompts. To create driving scenarios using Midjourney, we relied on textual descriptions as input. For this purpose, we utilised the large language model ChatGPT in version 3.5. By engaging with ChatGPT, we were able to generate diverse and detailed textual descriptions of driving scenarios featuring cars in adverse weather domains. To ensure a wide range of descriptions, we obtained 300 textual descriptions for each weather domain. Our main focus was on generating images that prominently showcased the weather conditions and included one or multiple cars. To achieve this, we consistently initiated the sentences with the phrase 'A photo of cars ...' followed by 'in strong fog/rain/snow conditions.' We observed that starting a prompt with 'A photo of ...' led to more realistic results. Without this specific prompt start, the generated content sometimes leaned towards being more illustrative rather than resembling real-world footage. An example prompt we gave Midjourney to create a driving scene in snow looks like this (figure 2 shows four different variants of this text-prompt):

"A photo of cars driving on a downtown street in very strong snow conditions. The street has shops and restaurants on both sides and pedestrians walking with umbrellas. The cars are honking and the snow is melting on the pavement. A colourful image."

Selection and Annotation. To ensure a diverse dataset, each text prompt was used to generate 4 image outputs through Midjourney AI. To increase the variety of data for selection, we repeated the prompts up to 8 times, resulting in a multitude of outputs for similar scene descriptions. In total, we generated a dataset of 18.000 images, evenly distributed across the fog, snow, and rain domains. From this dataset, we carefully selected 513 (512 for rain) images for each weather domain, ensuring their qualitative value. Annotations in form of 2D bounding boxes were applied to all cars present in the images. However, due to adverse weather conditions, annotating cars accurately posed challenges. Moreover, smaller cars in the distance sometimes appeared deformed. We chose to annotate deformed objects as cars if they were still recognisable to a human observer. Objects that were not clearly identifiable as cars were not annotated.

Data Distribution. Table 2 presents the distribution of small, medium, and large size objects in the generated dataset, with an image size of 1024x1024 pixels. Small objects are defined as having an area less than 32^2 pixels, medium objects have an area between 32^2 and 96^2 pixels, and large objects have an area greater than 96^2 pixels. Comparing this distribution to other benchmark datasets in adverse domains, we observe that the distribution of small objects is lower in our dataset. This is most likely due to problems of Midjourney creating small objects that clearly resemble cars. We often observe slight deformations of these objects. For medium, and large size cars the distribution is consequently slightly higher. Across all datasets, the majority of objects are centered around medium size. Therefore, our dataset shows a slightly different object size distribution than the evaluated real-world datasets.

Data Variation. The diversity of images within the datasets is an important characteristic for quality. The proposed method would be limited in its usefulness if the images within each prompt show strong visual similarities. To measure image diversity, we employed the LPIPS score [48] similar to [14], which serves as a metric for image similarity. A higher LPIPS score suggests greater dissimilarity and distance between images, while a lower score indicates higher similarity. In our evaluation, we computed the LPIPS score between each pair of images within a domain (fog, rain, and snow) to assess diversity. We conducted a comparison of the LPIPS scores between our synthetic data and other real-world datasets. The LPIPS score for synthetic fog was found to be 0.731, while for rain it was 0.807, and for snow it was 0.773. In contrast, the averaged LPIPS scores for NuScenes [5], ACDC [38], and Cityscapes [9] were 0.603, indicating a higher level of diversity for the synthetic dataset. The Adverse Dataset, with an LPIPS score of 0.680, exhibited slightly greater diversity but still fell short of the diversity observed in synthetic data. Qualitatively, the Midjourney dataset displayed a broad range of scenes with diverse viewpoints and various adverse conditions.

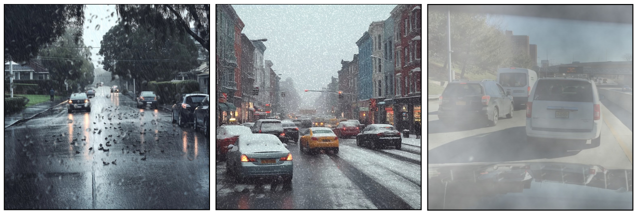

Figure 3. Images from the dataset with image corruptions. The left image showcases synthetic rain-streaks and raindrops. The middle image displays a synthetic snow overlay. The right image shows a clear image transformed with the synthetic fog model.

3.2. Image Corruptions

Controlling the visibility of rain-streaks, snowflakes, or raindrops in the generated output images using Midjourney proved to be challenging. To address this, we applied simple stochastic models to overlay these effects on the generated images. Specifically, we utilised the image corruption library proposed in [21] and a raindrop model described in [29]. These models allowed us to accurately introduce the desired weather effects onto the images, enhancing their realism and authenticity. We utilised the fog and raindrops models as proposed in their respective papers, making minor adjustments to the raindrop model to optimise it for GPU usage with a batch size of 4. Additionally, we made modifications to the snowflakes model by incorporating an alpha channel. As for the rain-streaks model, it was not available in the library, so we constructed it based on the snowflakes model by incorporating motion blur effects. All models were carefully parameterised by us to closely resemble real weather phenomena. Image corruptions have the capability to be layered on top of each other, allowing for the combination of different effects. For example, we can overlay a rain-streaks layer with a raindrops layer. Each image corruption is generated using a random seed, ensuring a wide range of diversity in each training epoch. This approach adds variability and realism to the generated images, enhancing the training process. Examples of the image corruptions can be seen in figure 3.

4. Experiments

In this section, we conduct an evaluation of multiple object detection models that have been fine-tuned on our synthesised dataset. To assess the performance, we utilise a range of benchmark datasets that encompass various levels of adverse weather conditions, ranging from light to strong. Our objective is to demonstrate the effectiveness of fine-tuning with synthetic data in improving the accuracy of object detection to a certain extent. Furthermore, we investigate the impact of the synthetic weather overlays on the

Dataset	Small	Medium	Large	Total	Average(S)	Average(M)	Average(L)
MJ Fog	337	2097	1187	3621	0.093	0.579	0.328
MJ Rain	409	2475	1976	4860	0.084	0.509	0.407
MJ Snow	585	2517	1696	4798	0.122	0.525	0.353
Baseline	64870	92279	57475	214624	0.282	0.454	0.263

Table 2. Analysis of the size distribution of cars within each weather domain. To assess this distribution, we compute the average occurrence of small, medium, and large cars across all domains and compare it to the distribution in the evaluated real-world datasets.

training process. We conduct several ablation studies to validate and analyse our findings. For all our results we report the Average Precision (AP) from 0.5 to 0.95 for the class car if not stated otherwise.

4.1. Benchmark Datasets

Adverse Dataset. This dataset was collected partly on the photo platform Flickr and shows more extreme weather conditions. It is based on [36] and was filled with additional images to obtain 200 images for each of the domains fog, rain and snow. It is to note that images with rainy weather conditions were added from the BDD100K dataset as it was difficult to gather enough valuable data for extreme rain conditions. The images are scaled and cropped to a size of 1024x1024px if possible, except for the ones from BDD100K.

ACDC Dataset. The Adverse Conditions Dataset (ACDC) [38] is a dataset to test perception methods on adverse visual conditions. For our evaluation we used the splits for rain, snow and fog conditions which consist of 500 images each.

NuScenes Dataset. NuScenes [5] is a dataset recorded in Boston and Singapore incorporating complex driving situations due to high traffic density in these cities. For the evaluation we took a subset of 5710 images in clear weather and a subset of 5422 images in rainy weather conditions. Since NuScenes only provides 3D bounding boxes, we converted them to 2D bounding boxes. These are not as tight-fitting as 2D bounding boxes, but are sufficient for our evaluation.

Cityscapes Dataset. Cityscapes [7] is a dataset collected in the streets of 50 different cities in clear weather conditions. For our evaluation we used 3475 images in total from the train and validation set. In addition we evaluated the Foggy Cityscapes Dataset [37] which augments fog in three different strengths to the original clear images.

BDD100K Dataset. The BDD100K dataset [46] consists of 100k annotated images with diverse scene types in varying weather and daytime conditions. For the evaluation we filtered all images recorded at daytime, dawn, dusk and overcast and grouped them into the weather conditions clear, rain and snow. The fog data was not used, as it is very weak and resembles more clear weather conditions. A subset of the images were used for the evaluation resulting in 4133 clear weather images, 3301 rainy images and 3794 snowy images.

4.2. Training Details

Object Detection Models. For the training four different object detection models are used: Faster R-CNN [33], SSD [19], RetinaNet [17] and FCOS [41]. All of these models were pre-trained on the COCO dataset [18]. The models use a ResNet FPN-50 backbone, except for SSD which uses a VGG16 backbone.

Training Parameters. In our training process, we adopt a common practice of removing the head (classification and regression) of all algorithms and replacing it with a new classification head dedicated to car detection. Throughout all our experiments, we utilise a batch size of 4 and train the models for 10 epochs. For optimising the model we use stochastic gradient descent with the learning rate set to 0.005, momentum 0.9 and weight decay of 0.0005. A random horizontal flip is applied to each image with a probability of 0.5.

Training Dataset. Our training dataset comprises a clear weather dataset, which includes 159 images collected from Flickr and 353 images from the BDD100K dataset captured under clear conditions. In our training process, we combine this dataset with additional extension datasets. Specifically, when training only with clear weather data we extend the training set by including another 512 clear images from the BDD100K dataset. When training with synthetic Midjourney images we extend the training data by 512 images from the respective domain. Thus we always have 1024 training images in total for a fair comparison. Additionally, we investigate the effects of augmenting images with additional corruptions. These augmentations are applied only to the extension data (e.g. to 512 clear images or 512 images from the synthetic data) to further diversify the training data (see section 3.2). For the detection results in section 4.3 we apply image overlays with a chance of 25% to the image. The strength of the intensity is chosen randomly from five different parameterisations.

4.3. Detection Results

Table 3 showcases the evaluation results across the benchmark datasets introduced in section 4.1, offering an average performance assessment within each weather domain. The resulting scores are the averaged AP across all datasets within the respective domains (e.g. all images with fog in the benchmark datasets). We establish a baseline using models pre-trained on the COCO dataset with all 80 classes. Although the baseline occasionally outperforms the fine-tuned models, we acknowledge that comparing it directly to our fine-tuning dataset would be unfair due to the vast size difference between the COCO dataset and our own. Therefore, our focus lies on the results obtained from training on clear weather and synthetic adverse weather data.

We find that models trained solely on clear weather data perform best on our clear weather test set. For almost all other scenarios, the models fine-tuned on the Midjourney dataset exhibit higher AP. In fog, our best model using Faster R-CNN achieves an improvement of 8.1% over the clear weather baseline. In snow, the AP is 5.2% better than the clear weather model. In the case of rain, we do not observe a significant performance gain in AP. This may be attributed to the rain scenarios often resembling clear weather conditions in the test set. However, our ablation studies in section 4.5 reveal a slight increase in AP on the adverse dataset with stronger rain conditions.

In our analysis, we explored the combination of data from different adverse domains by merging the 512 clear images and all 1538 synthetic images to a new dataset (MJ_All). Additionally, we established the Clear(2x) dataset, comprising 2048 clear weather images sourced mostly from the BDD100K dataset for a fair comparison. The MJ_All dataset yielded an overall performance boost and enhanced robustness against various weather conditions.

Surprisingly, our evaluation suggests that the utilisation of synthetic overlays, such as snowflakes, raindrops, and rainstreaks, did not yield significant benefits. It is worth noting that simple augmentation models have the potential to occlude portions of the image, leading to invalid bounding boxes and confusion during the fine-tuning process. Section 4.5 presents further results that emphasise the advantages of synthetic training data over simplistic image overlays.

Our findings indicate that training object detection models with fully synthetic images from adverse domains can indeed enhance their robustness.

4.4. Dataset Split

During our evaluation, we also considered the proportion of clear and adverse weather data used for training and illustrate the results in Figure 4. We extend the clear weather training dataset from section 4.2 by another 512 clear weather images to have 1536 in total in order to match the synthetic dataset in size. The amount of synthetic images is gradually increased, starting with only clear weather images and gradually replacing them with 1538 adverse weather images in increments of 20%. For instance, an amount of 40% corresponds to 60% clear weather data and 40% adverse weather data. The data was randomly split according to the specified proportions. In the clear weather domain, we observe higher AP when fine-tuning only with clear weather data, with a gradual decrease as synthetic adverse domain data is added. For fog and snow, we see an increase in AP up to a training split of 80% synthetic data. All models experience a significant drop in performance when exposed solely to synthetic adverse domain data. It is to note that this evaluation is based only on Faster R-CNN.

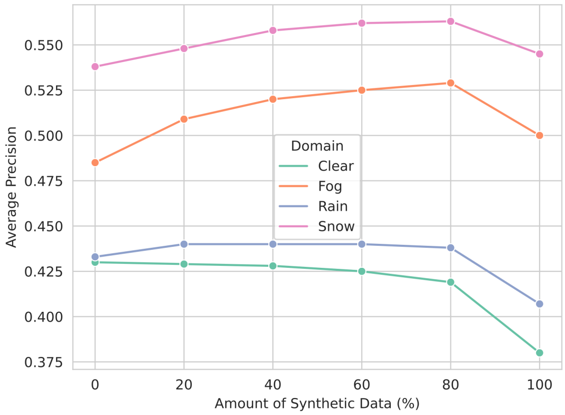

Figure 4. Evaluation results of different data splits between clear and adverse domain data for Faster R-CNN. Splits were gradually increased from adverse weather images to clear weather images in increments of 20%.

4.5. Ablation Studies

In this section, we show deeper insights into the benefits of the generated dataset for enhancing object detection in adverse weather conditions. We aim to examine various aspects and gain a comprehensive understanding of the dataset's effectiveness in improving detection performance.

Foggy Cityscapes

We assess the robustness of the algorithms on the cityscapes and foggy cityscapes dataset, where the fog intensity varies from clear to strong fog. While the models fine-tuned on clear weather data exhibit the best performance on cityscapes clear images, our evaluation in Figure 5 reveals that all algorithms display enhanced robustness towards fog when trained with synthetic overlays or synthetic fog images. Notably, algorithms fine-tuned solely on clear weather deteriorate more rapidly compared to their counterparts trained specifically in fog conditions. We observe an overall higher robustness towards fog when training with synthetic fog images.

	Faster R-CNN				**FCOS**				**RetinaNet**				**SSD**			
Trainset	Clear	Fog	Rain	Snow	Clear	Fog	Rain	Snow	Clear	Fog	Rain	Snow	Clear	Fog	Rain	Snow
Baseline	0.396	0.500	0.404	0.519	0.359	0.448	0.364	0.462	0.363	0.450	0.361	0.456	0.207	0.218	0.193	0.239
Clear	0.430	0.492	0.430	0.537	0.355	0.410	0.356	0.419	0.340	0.383	0.336	0.396	0.222	0.224	0.208	0.248
Clear(F)	0.430	0.513	0.437	0.541	0.349	0.414	0.351	0.419	0.344	0.399	0.345	0.406	0.217	0.235	0.209	0.243
Clear(S)	**0.431**	0.490	0.433	0.538	0.357	0.410	0.358	0.420	0.339	0.381	0.338	0.401	0.224	0.224	0.210	0.245
Clear(R)	0.429	0.489	0.433	0.537	0.357	0.410	0.358	0.420	0.337	0.379	0.336	0.397	0.222	0.225	0.208	0.243
Clear(D)	0.430	0.491	0.435	0.542	0.254	0.259	0.237	0.268	0.342	0.383	0.337	0.398	0.222	0.224	0.212	0.247
Clear(R+D)	0.428	0.486	0.434	0.539	0.354	0.409	0.356	0.416	0.340	0.391	0.341	0.404	0.222	0.223	0.213	0.248
MJ_Fog	0.423	**0.532**	0.436	0.555	0.343	0.421	0.352	0.421	0.331	0.410	0.332	0.403	0.211	0.237	0.198	0.238
MJ_Snow	0.422	0.503	0.432	0.561	0.338	0.408	0.341	0.420	0.333	0.408	0.333	0.419	0.212	0.230	0.198	0.258
MJ_Snow(S)	0.421	0.502	0.432	0.560	0.337	0.410	0.343	0.426	0.333	0.407	0.333	0.419	0.211	0.228	0.197	0.257
MJ_Rain	0.419	0.497	0.434	0.556	0.345	0.417	0.355	0.429	0.334	0.398	0.336	0.414	0.210	0.218	0.204	0.243
MJ_Rain(R)	0.418	0.496	0.435	0.556	0.328	0.399	0.339	0.408	0.332	0.398	0.334	0.413	0.212	0.224	0.203	0.246
MJ_Rain(D)	0.419	0.499	0.435	0.558	0.337	0.409	0.349	0.421	0.334	0.399	0.335	0.412	0.208	0.223	0.203	0.246
MJ_Rain(R+D)	0.421	0.500	0.437	0.559	0.337	0.409	0.346	0.417	0.334	0.400	0.338	0.413	0.210	0.224	0.204	0.247
Clear(2x)	0.429	0.489	0.435	0.539	**0.370**	0.426	**0.374**	0.440	**0.357**	0.398	**0.355**	0.415	**0.239**	0.243	**0.229**	0.266
MJ_All	0.419	0.524	0.437	0.561	0.355	0.445	0.366	0.455	0.342	**0.433**	0.346	**0.437**	0.217	**0.249**	0.213	0.270
MJ_All(A)	0.420	0.525	**0.438**	**0.565**	0.358	**0.446**	0.370	**0.458**	0.343	**0.433**	0.349	**0.437**	0.220	0.248	0.215	**0.274**

Table 3. Results of the evaluation of fine-tuned object detection models on the examined benchmark datasets. The datasets are clustered in the respective domains clear, fog, rain and snow. We report the averaged AP over the respective domains. When indicated the datasets were augmented with synthetic overlays (F: Fog, S: Snowflakes, R: Rain-streaks, D: Raindrops, A: Snowflakes, rain-streaks and raindrops).

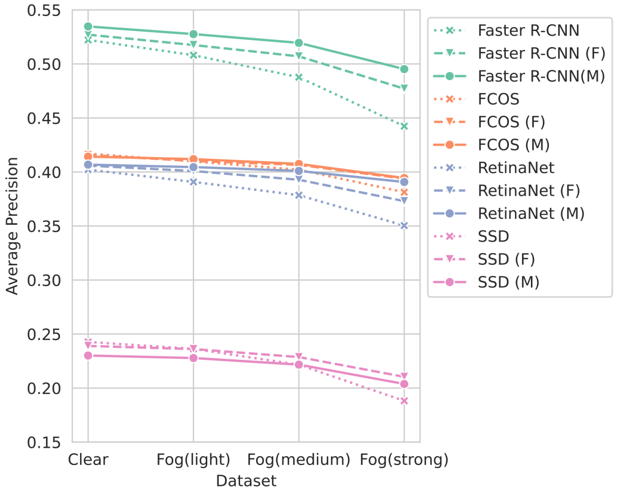

Figure 5. The evaluation results of all object detection models tested on Cityscapes and Foggy Cityscapes datasets, including three different fog intensities, are presented. The algorithms were fine-tuned using three different methods: clear images only, clear images with a fog overlay (F), and synthetic fog images generated by Midjourney (M). We observe that our synthetic data consistently demonstrates improved robustness to fog in almost all cases.

Image Corruptions

The focus of this evaluation is to analyse the impact of synthetic overlays when fine-tuning. To accomplish this, we introduce synthetic rain-streaks, snowflakes, and raindrops to the images and perform evaluations across all domains. The results are illustrated in Figure 6. Our findings indicate that the overlays have minimal effects on the AP of the detection models. It is possible that the test datasets lack sufficient data containing snowflakes, raindrops or streaks. It might be necessary to increase the intensity of the overlays to observe a more substantial contribution to the model's robustness. Nevertheless, these results indicate that the inclusion of synthetic adverse weather data may offer potential benefits compared to mere image overlays.

Dataset Correlation

To mitigate the potential influence of object size correlation between training and test datasets on the observed increase in AP, we conducted an investigation into the relationship between object size and AP. The results of this analysis are presented in Table 4. To this end, we compare the pearson correlation of AP to the distribution of object sizes (small, medium, large) of the MJ training set. Our findings indicate that there is no discernible correlation between the size distribution derived from the MJ dataset and the resulting AP values. By examining the relationship between object size and AP, we can ascertain that the observed improvements in AP are not solely driven by the size distribution between training and test data.

Extreme Adverse Weather

As we intended to create strong adverse conditions in the MJ dataset, we aimed to assess the performance on chal-

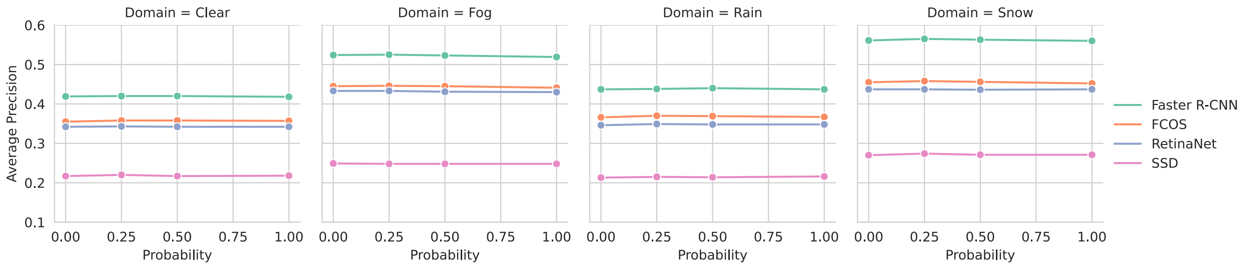

Figure 6. The evaluation presents the results of systematically adding weather corruptions to images. The corruptions are applied with probabilities of 0.0, 0.25, 0.5, and 1.0 during the fine-tuning of object detection models. The MJ_All(A) dataset is used for fine-tuning.

	Faster R-CNN	FCOS	RetinaNet	SSD
Train(S)	-0.003	-0.0593	0.022	0.057
Train(M)	0.050	-0.010	-0.006	0.030
Train(L)	-0.044	0.020	-0.005	-0.056

Table 4. Results for the correlation between AP and object size (small, medium, large) in the MJ training set. The results provide a quantitative measure of the correlation between these variables. The closer the number is to 0, the weaker the correlation.

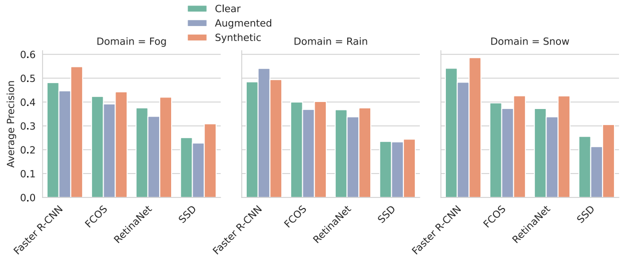

Figure 7. Evaluation results of detection models trained on clear, augmented or synthetic weather images. The results were evaluated on the Adverse Dataset. Each plot illustrates the outcomes of fine-tuning and evaluating images in a specific weather domain.

lenging test data. Figure 7 shows the results on the Adverse Dataset when the algorithms were fine-tuned on synthetic images specific to their respective weather domains. Each domain is compared to a model trained on clear weather images and to a model fine-tuned on data with standard augmentation techniques. We chose the AugMix policy [12] as it showed highest AP in our experiments. To ensure stability amidst variations caused by random transformations, we averaged results from 5 separate runs for AugMix. We observe enhanced robustness towards all weather conditions when fine-tuning on synthetic adverse weather data, compared to fine-tuning solely on clear weather images or augmented images. Faster R-CNN demonstrates a significant improvement of 13.9% in fog, and an 8.2% improvement in snow compared to clear weather fine-tuning. The improvements in rain conditions are only minor.

5. Discussion

Our work underscores the importance of synthetic data generation techniques to enrich existing datasets with adverse weather scenarios. This approach proves beneficial in fine-tuning object detection models designed to operate effectively in adverse weather conditions, enhancing the safety and reliability of autonomous driving.

Through extensive experimentation and evaluation, we provide empirical evidence that supports the effectiveness of our technique. The fine-tuned detection models exhibit improved performance in adverse weather compared to models trained solely on traditional datasets. Further, our approach offers unlimited possibilities for generating edge-case scenarios that are challenging or ethically difficult to produce in reality. Our method presents an alternative to real-world data collection.

It is important to acknowledge the limitations of our approach. The generated images may not always adhere to semantic correctness, featuring anomalies such as cars driving in false directions, deformed objects, or misplaced objects. Thus the synthesised data may have limited value outside the realm of computer vision. The single class annotations may reduce result validity and more studies are required to prove the effectiveness with multi-class data. Moreover, annotating synthetic data still requires effort, although it is likely less time-consuming than real world collection.

Overall, our research demonstrates the potential benefits and challenges associated with incorporating synthetic adverse weather data, paving the way for advancements in the field of autonomous driving and computer vision.

Acknowledgment

This work is supported and published within the AI-SEE project. AI-SEE is a co-labelled PENTA and EURIPIDES project endorsed by EUREKA. National funding authorities: Austrian Research Promotion Agency (FFG), Business Finland, Federal Ministry of Education and Research (BMBF), Canadian National Research Council Industrial Research Assistance Program (NRC-IRAP).

References

[1] Midjourney. https://www.midjourney.com. 2, 3

[2] Omri Avrahami, Dani Lischinski, and Ohad Fried. Blended Diffusion for Text-driven Editing of Natural Images. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, June 2022. IEEE. 3

[3] Dmitry Baranchuk, Andrey Voynov, Ivan Rubachev, Valentin Khrulkov, and Artem Babenko. Label-Efficient Semantic Segmentation with Diffusion Models. In *International Conference on Learning Representations*, Jan. 2022. 3

[4] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide. Seeing Through Fog Without Seeing Fog: Deep Multimodal Sensor Fusion in Unseen Adverse Weather. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 2

[5] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuScenes: A Multimodal Dataset for Autonomous Driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 2, 4, 5

[6] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, and James Hays. Argoverse: 3D Tracking and Forecasting With Rich Maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 2

[7] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 2, 5

[8] Quanfu Fan, Lisa Brown, and John Smith. A closer look at Faster R-CNN for vehicle detection. In *2016 IEEE Intelligent Vehicles Symposium (IV)*, June 2016. 2

[9] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012. 2, 4

[10] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. 3

[11] Sinan Hasirlioglu and Andreas Riener. Challenges in Object Detection Under Rainy Weather Conditions. In Joao Carlos Ferreira, Ana Lúcia Martins, and Vitor Monteiro, editors, *Intelligent Transport Systems, From Research and Development to the Market Uptake*, volume 267. Springer International Publishing, Cham, 2019. 2

[12] Dan Hendrycks*, Norman Mu*, Ekin Dogus Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshminarayanan. AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty. In *International Conference on Learning Representations*, Sept. 2019. 8

[13] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*, volume 33. Curran Associates, Inc., 2020. 3

[14] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal Unsupervised Image-to-Image Translation. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, volume 11207. Springer International Publishing, Cham, 2018. 3, 4

[15] Bahjat Kawar, Shiran Zada, Oran Lang, Omer Tov, Huiwen Chang, Tali Dekel, Inbar Mosseri, and Michal Irani. Imagic: Text-Based Real Image Editing With Diffusion Models. 3

[16] Minjun Li, Haozhi Huang, Lin Ma, Wei Liu, Tong Zhang, and Yugang Jiang. Unsupervised Image-to-Image Translation with Stacked Cycle-Consistent Adversarial Networks. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, volume 11213. Springer International Publishing, Cham, 2018. 3

[17] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal Loss for Dense Object Detection. *arXiv:1708.02002 [cs]*, Feb. 2018. 5

[18] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft COCO: Common Objects in Context. *arXiv:1405.0312 [cs]*, Feb. 2015. 5

[19] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: Single Shot MultiBox Detector. *arXiv:1512.02325 [cs]*, 9905, 2016. 5

[20] Wenyu Liu, Gaofeng Ren, Runsheng Yu, Shi Guo, Jianke Zhu, and Lei Zhang. Image-Adaptive YOLO for Object Detection in Adverse Weather Conditions. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(2), June 2022. 2

[21] Claudio Michaelis, Benjamin Mitzkus, Robert Geirhos, Evgenia Rusak, Oliver Bringmann, Alexander S. Ecker, Matthias Bethge, and Wieland Brendel. Benchmarking Robustness in Object Detection: Autonomous Driving when Winter is Coming, Mar. 2020. 2, 3, 4

[22] Muhammad Jehanzeb Mirza, Cornelius Buerkle, Julio Jarquin, Michael Opitz, Fabian Oboril, Kay-Ulrich Scholl, and Horst Bischof. Robustness of Object Detectors in Degrading Weather Conditions. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, Sept. 2021. 2

[23] M. Jehanzeb Mirza, Jakub Micorek, Horst Possegger, and Horst Bischof. The Norm Must Go On: Dynamic Unsupervised Domain Adaptation by Normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 2

[24] Valentina Musat, Ivan Fursa, Paul Newman, Fabio Cuzzolin, and Andrew Bradley. Multi-weather city: Adverse weather stacking for autonomous driving. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, BC, Canada, Oct. 2021. IEEE. 3

[25] Alexander Quinn Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob Mcgrew, Ilya Sutskever, and Mark Chen. GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. In *Proceedings of the 39th International Conference on Machine Learning*. PMLR, June 2022. 3

[26] Ko Nishino, Louis Kratz, and Stephen Lombardi. Bayesian Defogging. *International Journal of Computer Vision*, 98(3), July 2012. 2

[27] Taesung Park, Jun-Yan Zhu, Oliver Wang, Jingwan Lu, Eli Shechtman, Alexei Efros, and Richard Zhang. Swapping Autoencoder for Deep Image Manipulation. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33. Curran Associates, Inc., 2020. 3

[28] Fabio Pizzati, Pietro Cerri, and Raoul de Charette. Model-Based Occlusion Disentanglement for Image-to-Image Translation. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, Lecture Notes in Computer Science, Cham, 2020. Springer International Publishing. 3

[29] Fabio Pizzati, Pietro Cerri, and Raoul de Charette. Physics-Informed Guided Disentanglement In generative networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 3, 4

[30] Horia Porav, Tom Bruls, and Paul Newman. I Can See Clearly Now: Image Restoration via De-Raining. In *2019 International Conference on Robotics and Automation (ICRA)*, Montreal, QC, Canada, May 2019. IEEE. 3

[31] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical Text-Conditional Image Generation with CLIP Latents, Apr. 2022. 2, 3

[32] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-Shot Text-to-Image Generation. In *Proceedings of the 38th International Conference on Machine Learning*. PMLR, July 2021. 3

[33] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv:1506.01497 [cs]*, Jan. 2016. 5

[34] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, June 2022. IEEE. 3

[35] Thomas Rothmeier and Werner Huber. Let it Snow: On the Synthesis of Adverse Weather Image Data. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, Sept. 2021. 3

[36] Thomas Rothmeier, Diogo Wachtel, Tetmar von Dem Bussche-Hünnefeld, and Werner Huber. I Had a Bad Day: Challenges of Object Detection in Bad Visibility Conditions. In *2023 IEEE Intelligent Vehicles Symposium (IV)*, June 2023. 5

[37] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic Foggy Scene Understanding with Synthetic Data. *International Journal of Computer Vision*, 126(9), Sept. 2018. 2, 3, 5

[38] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. ACDC: The Adverse Conditions Dataset With Correspondences for Semantic Driving Scene Understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021. 2, 4, 5

[39] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*. PMLR, June 2015. 3

[40] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in Perception for Autonomous Driving: Waymo Open Dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 2

[41] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. FCOS: Fully Convolutional One-Stage Object Detection. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), Oct. 2019. IEEE. 5

[42] Maxime Tremblay, Shirsendu Sukanta Halder, Raoul de Charette, and Jean-François Lalonde. Rain rendering for evaluating and improving robustness to bad weather. *International Journal of Computer Vision*, 129(2), Feb. 2021. 2, 3

[43] Alexander von Bernuth, Georg Volk, and Oliver Bringmann. Simulating Photo-realistic Snow and Fog on Existing Images for Enhanced CNN Training and Evaluation. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Auckland, New Zealand, Oct. 2019. IEEE. 3

[44] Y. Wang and C. Fan. Single Image Defogging by Multiscale Depth Fusion. *IEEE Transactions on Image Processing*, 23(11), Nov. 2014. 2

[45] Wenhan Yang, Robby T. Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep Joint Rain Detection and Removal from a Single Image. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 2

[46] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, June 2020. IEEE. 2, 5

[47] Dan Zhang, Jingjing Li, Lin Xiong, Lan Lin, Mao Ye, and Shangming Yang. Cycle-Consistent Domain Adaptive Faster RCNN. *IEEE Access*, 7, 2019. 2

[48] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, June 2018. IEEE. 4