Let it Snow: On the Synthesis of Adverse Weather Image Data

Thomas Rothmeier[1] and Werner Huber[1]

Abstract— Camera systems of automated vehicles capture images from the surrounding environment and process these datastreams with algorithms to detect and classify objects. A lot of research has been devoted to improve object detection algorithms in order to provide highly accurate detection results in real time. However, these algorithms show a strong drop in performance as soon as they are exposed to adverse weather. Poor weather conditions such as rain, fog or snow lead to a reduction in visibility and thus objects are more difficult to recognize or not visible at all. This leads to a high degree of uncertainty for an automotive camera system. To enable automated driving, camera systems must be able to cope with adverse weather and the associated high uncertainty. Including more weather image data when training the algorithms can improve object detection in bad visibility conditions. However, weather image data is difficult to collect in reality and thus only available to a limited extent. In this work, we evaluate the possibility of using Generative Adversarial Networks to create synthetic weather image data. For this purpose, we compare the generated images of different network architectures trained on a diverse weather dataset collected from Flickr. The resulting data is evaluated qualitatively and quantitatively with respect to its realism and suggests that our approach is capable of generating realistic weather images.

I. INTRODUCTION

Automated vehicles need to perceive and analyse their surrounding environment permanently with high frame rates to even detect subtle environmental changes. For this purpose, they are equipped with camera sensors and corresponding computer vision algorithms in addition to other sensors. While camera sensor technology and algorithms are constantly improving, they are mostly designed for clear weather images and videos. They do not take into account adverse weather like rain, fog and snow [21]. However, it is not possible for an automated vehicle to simply "turn off" the bad weather. When exposed to these bad visibility conditions the performance of object detectors show a large decrease [24]. Thus, computer vision system for automated vehicles need to work reliably also in adverse weather conditions.

Atmospheric conditions such as rain, fog and snow are induced by small to microscopic particles in the air that scatter and absorb light rays and thus alter the appearance of a scene severely. The resulting images show often drastically degraded visibility, contrast and color fidelity, which makes it difficult to recognize and distinguish objects. This affects not only human drivers, but also the computer vision system of an automated vehicle. Lots of effort was put into image

[1]The authors are with the Center of Automotive Research on Integrated Safety Systems and Measurement Area (CARISSMA), Technische Hochschule Ingolstadt, Germany Email: thomas.rothmeier@thi.de and werner.huber@thi.de
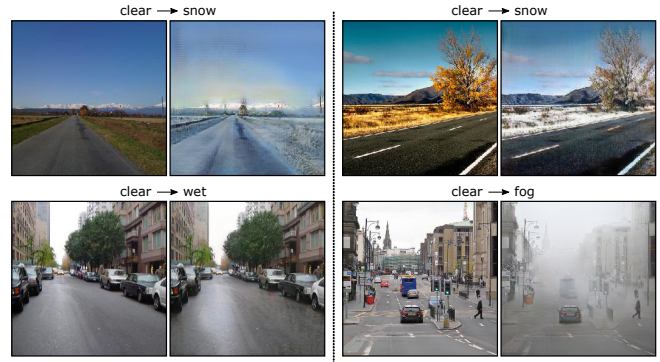
Fig. 1. Results of the weather domain adaptation models for different environmental conditions using CycleGAN.

dehazing [8], [22], [33] and deraining [4], [34], [37] to improve scene visibility. Yet, restoring the original scene from a single image remains a challenging task.

During the last few years, remarkable progress has been made in clear weather object detection and semantic scene understanding. These advances can be attributed to the use of AI algorithms in computer vision, mainly based on convolutional neural networks (CNNs). In general, large annotated datasets serve as training input for neural networks. Existing training datasets have mostly been recorded under clear weather conditions, which is why there are significant drops in the performance of the algorithms in adverse weather conditions. The lack of training data for adverse weather conditions can be attributed to the difficulties involved in recording it, as weather is not controllable and reproducible in reality. Further, recordings of critical traffic scenarios are already a rarity in clear weather, but not available in various weather conditions.

To address this problem, we turn away from the traditional approach of collecting data in real world. We also avoid using synthetic data captured in a fully virtual environment, as the gap between reality and simulation is often too large and models naively trained on synthetic data do not typically generalize to real images. Arising from the fact that large scale annotated datasets with clear weather images are available, we propose a method to enhance existing clear weather images with synthetic weather effects (see Fig. 1). For this purpose we use Generative Adversarial Networks (GANs) in an unsupervised setting where no paired input images are needed. Paired input images for identical scenes (e.g. clear and snow-covered scene) are seldom available in nature and thus very difficult to collect. Our primary contribution is to train and evaluate different neural network architectures

on the weather domain adaptation task and compare their outcomes qualitatively and quantitatively. Our domain adaptation models focus especially on visible influences on the scene like snow-covered roads, wet roads or the appearance of fog. We do not consider the influences of snowfall and raindrops. We also contribute by collecting adverse weather datasets online and at the CARISSMA Institute in Ingolstadt.

II. RELATED WORK

Generative Adversarial Networks (GANs) have shown impressive results regarding image generation [5], [13]. More recent work applies the idea of GANs to tasks like image inpainting [23], text-to-image translation [35], super resolution [32] and video synthesis [30], [31]. The key to GANs success is an adversarial loss that forces the generated image to be indistinguishable from a real photo. A generative model tries to generate images that resemble real images, while a discriminative model distinguishes between real and synthetic ones. This loss has shown to be particularly useful in image generation tasks. We utilize several GAN based architectures for generating synthetic weather images.

Image-to-Image Translation aims to learn the mapping between a source image domain and a target image domain. Therefore a dataset of input-output examples is used to learn a translation function based on CNNs. In this work we use the pix2pix [12] framework, which employs a conditional GAN [20] to learn a mapping from source to target domain. However, paired input-output examples are needed for this approach, which makes it often difficult to apply to real world samples.

Unpaired Image-to-Image Translation is an approach to find a relationship between two data domains without the need for tuples of corresponding images. More recently, several approaches tackling the unsupervised setting have been proposed. CoGAN [19] achieves a common representation across domains by enforcing a weight-sharing constraint. CycleGAN [39], DiscoGAN [14] and UNIT [18] leverage a cycle consistency loss to strengthen semantic consistency between source and target image. To enforce more diversity in the output space MUNIT [10] and DRIT [15] assume that the image representation can be decomposed into a content code and a style code. In this work we utilize several state-of-the-art network architectures for unpaired image-to-image translation to generate adverse weather image data and compare the results against each other.

Synthetic Visual Data is data that was not recorded in reality, but generated by simulation methods. The enormous progress in the field of computer vision can be attributed to large scale annotated image datasets [2], [3]. At present, however, it is impossible to collect and annotate data for every new problem that requires training data. Often it is simply too difficult and time consuming to acquire visual grasping data. Thus, learning with synthetic data is gaining

more and more attention recently, as synthetic data is cheap and large amounts of data can be easily collected in no time. In [1], [27], [28] synthetic datasets are utilized to successfully improve real world robotics and object detection tasks. Yet, models trained purely on synthetic data often fail to generalize to real world scenarios. In this work we evaluate existing GAN architectures for synthetic weather image generation and examine their usefulness in reducing the reality gap between real and generated images.

Weather Effect Simulation is a method to enhance existing data with the influences of weather related phenomena. Sakaridis et al. proposed a method to enhance existing image data with fog by applying an optical model for fog [25]. A GAN based approach to weather simulation was shown in [17] with focus on photographic style transfer. A physics based approach for simulating rain effects for camera, radar and lidar was introduced in [7]. Volk et al. generated training data with rain variations by simulating rainstreaks with brightness reduction and raindrops on the windshield to improve on object detection [29]. In this work, on the other hand, the focus is on the visible influences of the weather on the environment, such as snow-covered or wet roads.

III. UNPAIRED IMAGE-TO-IMAGE TRANSLATION FOR ADVERSE WEATHER IMAGE GENERATION

As mentioned at the beginning, collecting images of adverse weather scenes in reality is often difficult, because the associated weather conditions cannot be reproduced at the push of a button. To tackle this problem we propose an approach to enhance existing images of clear weather scenes with adverse weather effects and let them look as though they were recorded in reality. First, we will look into challenges of the weather domain adaptation task and then give a short overview of unpaired image-to-image translation algorithms.

A. Weather Domain Adaptation

Referring to [26] environmental conditions can be divided into four categories: weather (wind, rain, snow), weather-induced roadway conditions (standing water, snow-covered road), particulate matter (fog, smoke) and illumination (lighting, daytime). Weather and weather-related effects occur in many forms and lead to visible changes in the scene. To realistically simulate weather effects, it is not enough to simply add raindrops or snowflakes to the image. A winter day, for example, leads to snow deposits on roads, vehicles and buildings. A heavy rainfall, on the other hand, leaves water on the road, which can lead to wet roads and reflective surfaces. Camera sensor perform in general very poorly in such environmental conditions. Furthermore, changes in environmental conditions lead to changes in vegetation. E.g. in a summer to winter change trees lose their foliage and plants wither. Also the appearance of the sky often changes due to weather changes. Fog is considered easiest, as it does not lead to weather-induced roadway conditions. It is currently impossible to map these complex processes of

weather change manually. However, to be sufficient as a dataset for object recognition training, the generated data must correspond to reality to a high degree. Therefore, in this work we try to learn the correlations on existing data and transfer them to new data.

To understand the full breadth of the problem, it is indispensable to remember that a scene can take on any number of weather manifestations. For example, one can distinguish between light and heavy snow or rain, a snowy road but no snowfall, and so on. For an algorithm for weather domain adaptation, it is therefore important to be able to capture weather in its various forms and intensities. Optimally, this would be possible through a targeted parameterisation by a human. Furthermore, weather effects can be mixed, for example a snowy road with the presence of fog. Adding weather effects to existing images is a challenging task due to the wide variability of weather and the specific content-aware changes that need to be made to the source image. In this work we focus on generating realistic weather images for rain, fog and snow. We do not delve deeper into parameterisation or the mixing of weather effects, as this would go beyond the scope of this paper.

B. Unpaired Image-to-Image Translation

The general goal of unpaired image-to-image translation is to learn a mapping between two domains X and Y given a set of training samples $\{x_i\}_{i=1}^N$ where $x_i \in X$ and $\{y_j\}_{j=1}^M$ where $y_j \in Y$. The mapping function can have different underlying assumptions.

In case of CycleGAN [39] two mappings $G : X \to Y$ and $F : Y \to X$ are defined, where G and F are neural networks. Additionally two adversarial discriminators D_X and D_Y are defined, where D_X tries to distinguish between real images $\{x\}$ and fake images $\{F(y)\}$. The same applies for D_Y vice versa. An adversarial loss [5] is used as objective in order to match the distribution of synthesized images to the data distribution in the target domain. Moreover a cycle consistency loss ensures semantic consistency between input image and generated images.

The authors of UNIT [18] assume that a pair of images (x_1, y_1) can be mapped to the same latent code z in a shared-latent space Z. Based on this code z, both images can be reconstructed. Therefore two encoding functions E_1 and E_2 that map images to latent codes z and two generator functions G_1 and G_2 that map latent codes to images are defined. A function that maps from X to Y can then be represented by $G_2(E_1(x_1))$ and from Y to X by $G_1(E_2(x_2))$. Similar to CycleGAN an adversarial loss and a cycle consistency constraint are defined as necessary objectives.

The MUNIT framework [10] extends this thought and proposes a partially shared latent space. This means that an image x_i consists of a shared latent code z shared between two domains and an additional style latent code $s_i \in S_i$ that is specific to each domain. The separation of content and style in the MUNIT framework enables many-to-many mappings.

While sharing similar ideas of loss functions the underlying implementation and architecture for each network is different. For a more in-depth explanation of the concepts, we refer the reader to the respective publications. In the following chapter we aim to utilize the proposed GAN architectures to generate images with realistic weather effects using images recorded in clear weather conditions as input.

IV. EXPERIMENTS

We collected different datasets in order to use them as training data for a variety of GANs. The trained GANs are capable of translating an input image into another weather domain. The results are evaluated using image quality metrics and human judgement.

A. Evaluation Metrics

Peak Signal-to-Noise ratio (PSNR) is a widely used image quality metric that serves as a performance indicator for subjective image quality. A high PSNR indicates good image quality [11].

Structural Similarity Index (SSIM) is a measurement tool that is based on luminance, structure and contrast of an image to better suit the human visual system. A high value suggests high similarity [38].

Frechet Inception Distance (FID) was first introduced in [9] and is a metric that calculates the distance between feature vectors for real and generated images. It summarizes how similar two datasets are in terms of statistics based on computer vision features that are calculated by an inception v3 model. A lower scores correlates with higher image quality.

Learned Perceptual Image Patch Similarity (LPIPS) measures the distance between two images using feature vectors of deep neural networks. A higher value indicates more diversity between two images, whereas a lower value indicates more similarity [36].

Human Judgement Human judgement is the gold standard in evaluating image quality as evaluation metrics often lack expressiveness. Human qualitative judgment in this work is conducted by the authors.

B. Datasets

Flickr Weather Image Dataset. Weather image datasets are hardly available in the public domain. This is largely due to the lack of reproducability of weather image data in reality. To conduct our experiment, we collected adverse weather image data from flickr and sorted it by weather domains. The weather domains considered are clear, fog, snow and rain (mainly wet roads due to the difficulty of collecting images with visible rainstreaks). The collected data set mainly shows traffic scenarios and is divided into images with objects (e.g. cars, trucks, persons) and without objects (only road and environment visible). This results in a total of 6 unpaired data sets: *clear-snow, clear-snow-objs, clear-fog, clear-fog-objs, clear-wet, clear-wet-objs*. Table I gives an overview of the amount of images used. Since the data was collected on flickr with images from all over the planet, it has a high

diversity.

CARISSMA Weather Image Dataset. This set of image data was recorded in the weather facility of CARISSMA Research Institute in Ingolstadt. The institute has the possibility to reproducibly simulate rain and fog under controlled environmental conditions over a length of 50m in a closed indoor laboratory. We recorded a car attrap that was placed on a platform which moves away from the camera sensor over a distance of 50m in different environmental conditions (clear, fog, rain). Using the recorded data two unpaired datasets were created: *clear-fog-indoor, clear-rain-indoor*. As it is possible to record similar images with and without the influence of weather effects in the indoor laboratory we also created two datasets with paired images where we matched corresponding images: *clear-fog-indoor-paired, clear-rain-indoor-paired*. 295 images were collected for the fog-indoor dataset and 281 images for the rain-indoor dataset.

All images in the flickr and CARISSMA datasets are center cropped and scaled to a size of 256x256 pixels.

TABLE I

OVERVIEW OF THE RESPECTIVE AMOUNT OF IMAGES COLLECTED FOR EACH DATASET.

	Clear	Wet	Fog	Snow
No Objects	1076	395	492	1042
Objects	827	667	485	789

C. Baselines

Network Details. Many GAN network architectures have recently appeared that share similar ideas. Not all of them have been evaluated, but for the sake of completeness, they are listed in parentheses to the respective similar architectures. For our evaluation we use CycleGAN [39] (DiscoGAN [14]), UNIT [18] and MUNIT [10] (DRIT [16]), all of which learn mappings between two domains in an unsupervised manner. For the paired dataset we additionaly evaluate the pix2pix network [12].

Training Details. We trained each CycleGAN and pix2pix network for 200 epochs on the respective datasets with a batch size of 1. The learning rate is decayed to zero linearly after 100 epochs. UNIT and MUNIT were trained for 300k iterations on the flickr dataset and 100k iterations on the CARISSMA weather image dataset with a batch size of 1. In case of CycleGAN we also varied the identity loss between 0.0, 0.1 and 1.0. The identity loss is used in the original paper to preserve the color of input images. The original code provided on GitHub by the respective authors is used.

D. Qualitative Evaluation

Flickr Weather Image Dataset. For the qualitative evaluation of the trained models we use images unseen by the algorithm and transform them into the respective weather domains. Fig. 2 shows the results of the models trained on the flickr weather image dataset. Looking at the results

obtained, we found that our trained models are capable of generating realistic images for each weather domain. We have noticed that UNIT and MUNIT have problems in maintaining structures with the necessary sharpness and generate in general blurrier results. CycleGAN on the other hand generates images that are visually more appealing and often look as though they were sampled from the target domain. Looking at all transformations, fog seems to be the simplest domain. This may be explained by the fact that less content-aware changes need to be performed in the fog domain. In the snow domain, the added snow looks sometimes artificial and misses 3-dimensional structure. For CycleGAN(0.0) we often observed that the model simply inverses parts of the image in the snow domain. In the case of wet roads, we believe it is the most difficult domain, as reflections and mirror effects have to be added to the image. Yet the results are better than initially expected and the trained models often add a reflective surface to the road. For CycleGAN with identity values of 0.0 and 0.1, we often found that cars in the target domain were colored differently than in the source domain. This behavior can be prevented by setting the identity loss to 1.0. It can be said that the CycleGAN(1.0) generally provides the best quality results from a human visual perspective. Moreover, we found that some domain adapatation was harder than the other and did not deliver appropriate results. Fig. 3 provides an overview of some poor results.

CARISSMA Weather Image Dataset. Fig. 4 shows the results of the trained models on the CARISSMA weather image dataset. Here we see that UNIT and MUNIT also have difficulties to generate sharp results. The CycleGAN results tend to be closer towards the reference image in terms of image similarity. As we were able to collect paired data in this scenario we also evaluated a pix2pix model on the dataset. It can be observed, that the vehicle in the image is blurred and has a lower image quality compared to CycleGAN. This is the result of paired images that are not completely identical in terms of their semantics. This fact underlines the problem of matching images with different environmental conditions. Image quality is generally good on this dataset, partly because it has very little diversity compared to the flickr weather dataset and thus is easier to learn for a model. In addition we also compared the generated results with physics-based fog and rain models developed in [6]. Regarding realism of simulated weather effects our trained models show higher perceptual realism compared to the physics-based models.

E. Quantitative Evaluation

The results of the quantitative evaluation for the flickr weather image dataset are shown in Table II. The FID is calculated between the respective real and synthetized image data (e.g. fog ↔ synthetic-fog, snow ↔ synthetic-snow). CycleGAN(1.0) achieves the best results on almost all datasets except for the *clear-wet* dataset. In general all CycleGAN models have a low FID score indicating good
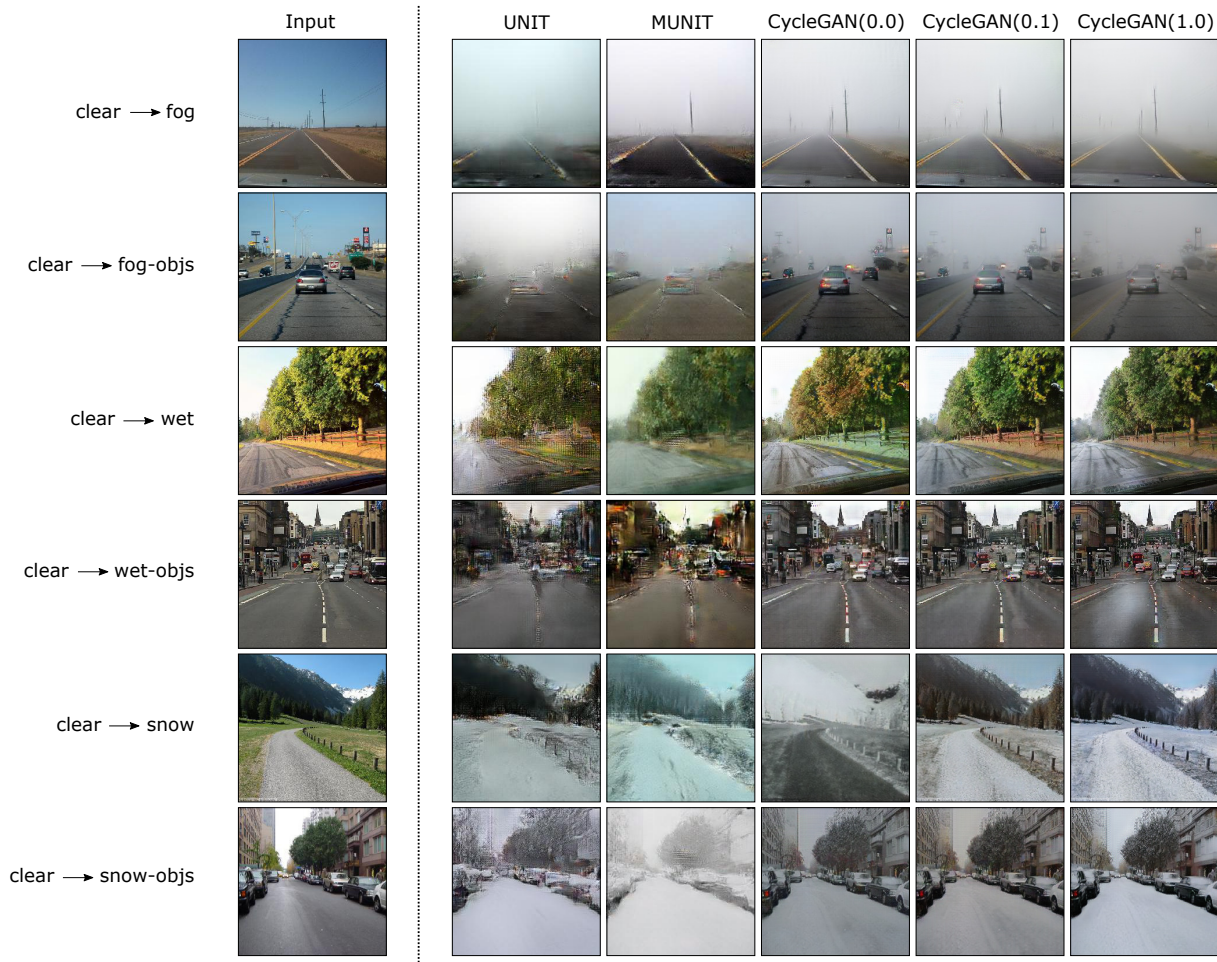
Fig. 2. Qualitative evaluation of the GAN models trained on the flickr weather image dataset. An input image is given to the trained GAN model and transformed into the respective weather domain. The CycleGAN model is evaluated with varying identity losses. UNIT and MUNIT tend to produce blurry results while CycleGAN generates visually more appealing images in the respective weather domains.

image quality.

TABLE II

FID CALCULATED BETWEEN REAL AND SYNTHETIC IMAGES ON THE
FLICKR WEATHER DATASET.

real ↔ synthetic	UNIT	MUNIT	CycleGAN (0.0)	CycleGAN (0.1)	CycleGAN (1.0)
fog	131.487	142.313	69.602	70.146	**63.528**
fog-objs	158.493	177.815	72.160	70.228	**69.116**
wet	209.408	107.617	63.030	**57.242**	58.899
wet-objs	148.942	125.996	61.921	58.549	**54.807**
snow	110.401	93.645	75.320	59.302	**57.459**
snow-objs	138.279	143.461	67.355	65.306	**62.923**

For the CARISSMA weather dataset we calculated the PSNR, SSIM, LPIPS and FID. These metrics are calculated between 9 images that were synthesized from clear weather images and 9 real images from the respective weather domain. For the *fog-indoor* dataset the CycleGAN(1.0) model scores best for each metric (see Table III). For the *rain-indoor* dataset UNIT scores best for SSIM, whereas PSNR, LPIPS and FID are best for CycleGAN(0.1) and Cycle-GAN(1.0) (see Table IV). The image quality metrics largely correlate with the qualitative evaluation. Yet, it should be noted that PSNR and SSIM as image quality metrics may not be suitable for assessing synthetic data. In addition, the significance of all metrics applied may be reduced by the small number of test images.

TABLE III

DIFFERENT IMAGE QUALITY METRICS CALCULATED BETWEEN REAL
AND SYNTHETIC FOG IMAGES ON THE CARISSMA WEATHER IMAGE
DATASET.

Fog-Indoor	PSNR	SSIM	LPIPS	FID
UNIT	16.939	0.254	0.303	198.028
MUNIT	16.172	0.222	0.301	227.698
pix2pix	19.701	0.190	0.308	125.821
CycleGAN(0.0)	22.424	0.241	0.244	102.650
CycleGAN(0.1)	22.556	0.267	0.229	96.793
CycleGAN(1.0)	**23.317**	**0.271**	**0.223**	**82.707**

V. CONCLUSION

In this paper we have demonstrated that GANs trained on adverse weather datasets can be utilized to generate realistically looking weather effects for the domains fog, rain
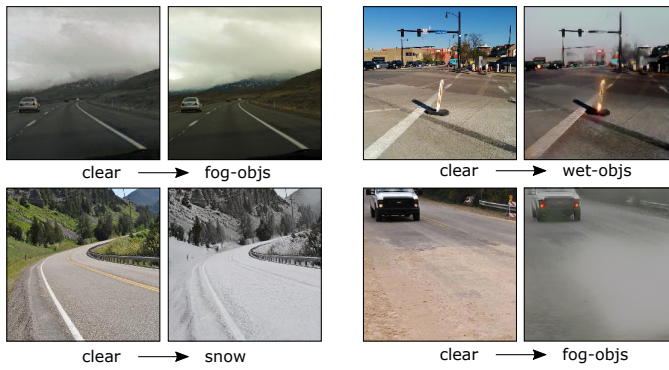
3304

Fig. 3. Some examples from the experiments where the weather domain adaptation showed poor results. From top-left to bottom-right: 1) No fog was added to the scenery, 2) Instead of a wet surface we get a dark surface and glowing objects, 3) The image appears to be grayscaled, 4) No fog was added to the scenery and the visible vehicles front light appears to be a rear light.

TABLE IV

DIFFERENT IMAGE QUALITY METRICS CALCULATED BETWEEN REAL AND SYNTHETIC RAIN IMAGES ON THE CARISSMA WEATHER IMAGE DATASET.

Rain-Indoor	PSNR	SSIM	LPIPS	FID
UNIT	17.571	**0.055**	0.527	236.306
MUNIT	17.669	0.042	0.499	255.162
pix2pix	17.522	0.027	0.500	331.679
CycleGAN(0.0)	17.625	0.022	0.460	122.322
CycleGAN(0.1)	**17.909**	0.037	0.462	117.186
CycleGAN(1.0)	17.898	0.044	**0.440**	**105.265**

and snow. We have compared different network architectures qualitatively and quantitatively and found that CycleGAN delivers the best results when trained on the proposed datasets. We also found that some transformations were easier than others. Scaling the images to a size of 256x256 for training may be problematic, as the loss of information due to image compression affects the quality of the simulated weather effects (e.g. texture of snow is often missing). We believe that a larger and higher quality training dataset will contribute significantly to the quality of the generated images. Yet, the usefulness of the data generated this way must be evaluated in a next step and applied to the training of object detection algorithms. To use this approach for training object detection algorithms for automated driving functions, it is important to establish a way to get more control over the generated results and more consistent results overall. Nevertheless, this work shows that it is possible to generate realistic weather image data with the help of neural networks and thus contributes towards safe automated driving in bad visibility conditions.

ACKNOWLEDGMENT

REFERENCES

[1] Konstantinos Bousmalis, Nathan Silberman, David Dohan, Dumitru Erhan, and Dilip Krishnan. Unsupervised Pixel-Level Domain Adaptation with Generative Adversarial Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 95–104, Honolulu, HI, July 2017. IEEE.

[2] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, June 2009. ISSN: 1063-6919.

[3] Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010. Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 2 Publisher: Springer US.

[4] K. Garg and S. K. Nayar. Detection and removal of rain from videos. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pages I–I, June 2004. ISSN: 1063-6919.

[5] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. *arXiv:1406.2661 [cs, stat]*, June 2014. arXiv: 1406.2661.

[6] Sinan Hasirlioglu. *A Novel Method for Simulation-based Testing and Validation of Automotive Surround Sensors under Adverse Weather Conditions / submitted by Sinan Hasirlioglu*. PhD thesis, Universität Linz, 2020.

[7] Sinan Hasirlioglu and Andreas Riener. A General Approach for Simulating Rain Effects on Sensor Data in Real and Virtual Environments. *IEEE Transactions on Intelligent Vehicles*, pages 1–1, 2019. Conference Name: IEEE Transactions on Intelligent Vehicles.

[8] K. He, J. Sun, and X. Tang. Single Image Haze Removal Using Dark Channel Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, December 2011. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[9] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *arXiv:1706.08500 [cs, stat]*, January 2018. arXiv: 1706.08500.

[10] 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, pages 179–196. Springer International Publishing, Cham, 2018.

[11] Quan Huynh-Thu and Mohammed Ghanbari. The accuracy of PSNR in predicting video quality for different video scenes and frame rates. page 14.

[12] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. *arXiv:1611.07004 [cs]*, November 2016. arXiv: 1611.07004.

[13] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation. *arXiv:1710.10196 [cs, stat]*, October 2017. arXiv: 1710.10196.

[14] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to Discover Cross-Domain Relations with Generative Adversarial Networks. *arXiv:1703.05192 [cs]*, May 2017. arXiv: 1703.05192.

[15] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang. Diverse Image-to-Image Translation via Disentangled Representations. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, volume 11205, pages 36–52. Springer International Publishing, Cham, 2018.

[16] Hsin-Ying Lee, Hung-Yu Tseng, Qi Mao, Jia-Bin Huang, Yu-Ding Lu, Maneesh Singh, and Ming-Hsuan Yang. DRIT++: Diverse Image-to-Image Translation via Disentangled Representations. *arXiv:1905.01270 [cs]*, December 2019. arXiv: 1905.01270.

[17] Xuelong Li, Kai Kou, and Bin Zhao. Weather GAN: Multi-Domain Weather Translation Using Generative Adversarial Networks. *arXiv:2103.05422 [cs]*, March 2021. arXiv: 2103.05422.

[18] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised Image-to-Image Translation Networks. *arXiv:1703.00848 [cs]*, July 2018. arXiv: 1703.00848.
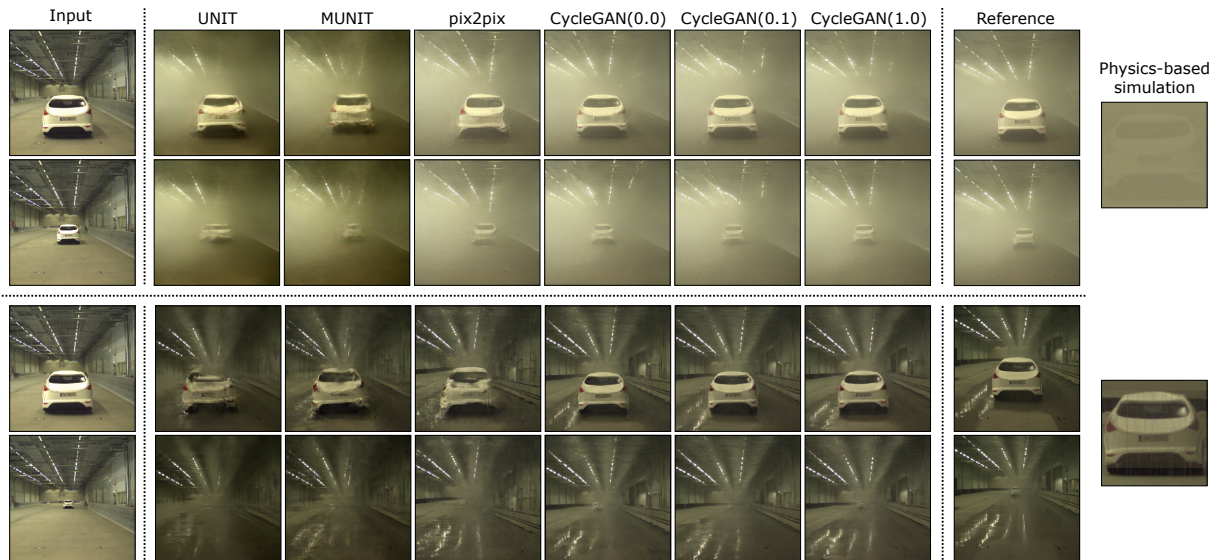
Fig. 4. Qualitative evaluation of the GAN models trained on the CARISSMA weather image dataset. An image is given to the GAN as input and then transformed into the respective weather domain. For this task we evaluated UNIT, MUNIT, pix2pix and CycleGAN with different identity losses. In addition, we visually compare our approach with a physics-based approach towards weather simulation.

[19] Ming-Yu Liu and Oncel Tuzel. Coupled Generative Adversarial Networks. *arXiv:1606.07536 [cs]*, June 2016. arXiv: 1606.07536.

[20] Mehdi Mirza and Simon Osindero. Conditional Generative Adversarial Nets. *arXiv:1411.1784 [cs, stat]*, November 2014. arXiv: 1411.1784.

[21] Srinivasa G. Narasimhan and Shree K. Nayar. Vision and the atmosphere. In *ACM SIGGRAPH ASIA 2008 courses on - SIGGRAPH Asia '08*, pages 1–22, Singapore, 2008. ACM Press.

[22] Ko Nishino, Louis Kratz, and Stephen Lombardi. Bayesian Defogging. *International Journal of Computer Vision*, 98(3):263–278, July 2012.

[23] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros. Context Encoders: Feature Learning by Inpainting. *arXiv:1604.07379 [cs]*, November 2016. arXiv: 1604.07379.

[24] Thomas Rothmeier and Werner Huber. Performance Evaluation of Object Detection Algorithms Under Adverse Weather Conditions. In Ana Lúcia Martins, João C. Ferreira, Alexander Kocian, and Vera Costa, editors, *Intelligent Transport Systems, From Research and Development to the Market Uptake*, volume 364, pages 211–222. Springer International Publishing, Cham, 2021. Series Title: Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering.

[25] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic Foggy Scene Understanding with Synthetic Data. *International Journal of Computer Vision*, 126(9):973–992, September 2018.

[26] Eric Thorn, Shawn Kimmel, and Michelle Chaka. A framework for automated driving system testable cases and scenarios, 2018.

[27] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 23–30, September 2017. ISSN: 2153-0866.

[28] Jonathan Tremblay, Aayush Prakash, David Acuna, Mark Brophy, Varun Jampani, Cem Anil, Thang To, Eric Cameracci, Shaad Boochoon, and Stan Birchfield. Training Deep Networks With Synthetic Data: Bridging the Reality Gap by Domain Randomization. pages 969–977, 2018.

[29] Georg Volk, Stefan Muller, Alexander von Bernuth, Dennis Hospach, and Oliver Bringmann. Towards Robust CNN-based Object Detection through Augmentation with Synthetic Rain Variations. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 285–292, Auckland, New Zealand, October 2019. IEEE.

[30] Carl Vondrick, Hamed Pirsiavash, and Antonio Torralba. Generating Videos with Scene Dynamics. *arXiv:1609.02612 [cs]*, September 2016. arXiv: 1609.02612.

[31] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Guilin Liu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. Video-to-Video Synthesis. *arXiv:1808.06601 [cs]*, August 2018. arXiv: 1808.06601.

[32] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. *arXiv:1711.11585 [cs]*, November 2017. arXiv: 1711.11585.

[33] Y. Wang and C. Fan. Single Image Defogging by Multiscale Depth Fusion. *IEEE Transactions on Image Processing*, 23(11):4826–4837, November 2014. Conference Name: IEEE Transactions on Image Processing.

[34] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep Joint Rain Detection and Removal From a Single Image. page 10.

[35] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris Metaxas. StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. *arXiv:1612.03242 [cs, stat]*, August 2017. arXiv: 1612.03242.

[36] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. *arXiv:1801.03924 [cs]*, April 2018. arXiv: 1801.03924.

[37] X. Zhang, H. Li, Y. Qi, W. K. Leow, and T. K. Ng. Rain Removal in Video by Combining Temporal and Chromatic Properties. In *2006 IEEE International Conference on Multimedia and Expo*, pages 461–464, July 2006. ISSN: 1945-788X.

[38] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004. Conference Name: IEEE Transactions on Image Processing.

[39] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *arXiv:1703.10593 [cs]*, November 2018. arXiv: 1703.10593.