Let's Get Dirty: GAN Based Data Augmentation for Camera Lens Soiling Detection in Autonomous Driving

Michal Uřičář[1*], Ganesh Sistu[2], Hazem Rashed[2], Antonín Vobecký[3,2], Varun Ravi Kumar[2], Pavel Křížek[1*], Fabian Bürger[2] and Senthil Yogamani[2]
[1]Independent Researcher [2]Valeo [3]CTU in Prague
uricar.michal@gmail.com antonin.vobecky@cvut.cz firstname.lastname@valeo.com

Abstract

Wide-angle fisheye cameras are commonly used in automated driving for parking and low-speed navigation tasks. Four of such cameras form a surround-view system that provides a complete and detailed view of the vehicle. These cameras are directly exposed to harsh environmental settings and can get soiled very easily by mud, dust, water, frost. Soiling on the camera lens can severely degrade the visual perception algorithms, and a camera cleaning system triggered by a soiling detection algorithm is increasingly being deployed. While adverse weather conditions, such as rain, are getting attention recently, there is only limited work on general soiling. The main reason is the difficulty in collecting a diverse dataset as it is a relatively rare event.

We propose a novel GAN based algorithm for generating unseen patterns of soiled images. Additionally, the proposed method automatically provides the corresponding soiling masks eliminating the manual annotation cost. Augmentation of the generated soiled images for training improves the accuracy of soiling detection tasks significantly by 18% demonstrating its usefulness. The manually annotated soiling dataset and the generated augmentation dataset will be made public. We demonstrate the generalization of our fisheye trained GAN model on the Cityscapes dataset. We provide an empirical evaluation of the degradation of the semantic segmentation algorithm with the soiled data.

1. Introduction

Level 5 autonomous driving ([10]) stands out as a challenging goal of a large part of the computer vision and machine learning community. Due to this problem's difficulty, a combination of various sensors is necessary to build a safe and robust system. Even just considering cameras, a

*Most of this work was done while Michal Uřičář and Pavel Křížek were employed by Valeo.

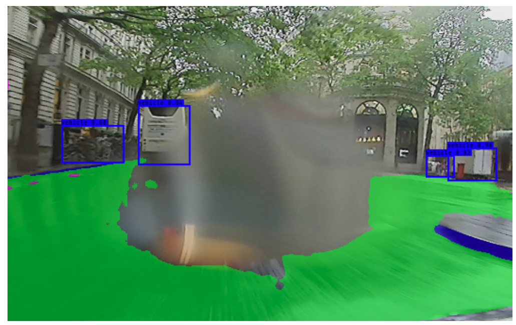


Figure 1: The example of a semi-transparent soiling in form of a water drop on the camera lens. The detection of the bus behind the water drop works still well, while the road segmentation (green) is highly degraded in the soiled region. In this scenario, a soiling detection algorithm is used to trigger a camera cleaning system which restores the lens hardware.

combination of a narrow field of view (FOV) long-range sensing and wide FOV short-range sensing is necessary. The latter is achieved by fisheye cameras, which provide complete near-field sensing, up to 10m around the vehicle. One use-case where it is necessary is the fish-bone parking [9], where ultrasonic sensors and narrow FOV cameras are not sufficient. Without exaggeration, these fisheye cameras are becoming *de facto* standard in low-speed maneuvering applications, like parking and traffic jams assist [16]. Enormous progress can be noticed in typical image processing tasks, such as semantic segmentation or object detection [23, 24, 25], which is mainly attributed to the prevailing success of deep learning. However, there are other less "popular" problems slowly getting into attention, which have to be solved as well for the ultimate goal of the full Level 5 autonomy.

One of these problems is the reliability of the sensory signal, which in the case of surround-view cameras means, *inter alia*, the ability to detect soiling on the camera lens. Failure to recognize severe weather conditions leading to a deterioration of the image quality to such a level that any further image processing is unreliable. Figure 2 shows

Figure 2: Automotive surround view cameras are exposed to harsh environmental setup. Left: camera lens covered by mud. Middle: image produced by the soiled camera from the left picture. Right: camera lens soiled during heavy rain.

how the surround-view camera can get soiled and the corresponding image output, as well as an example of images taken during a heavy rain. It usually happens when the tire is splashing mud or water from the road or due to wind depositing dust on the lens. Figure 1 shows an example of the strong impact of a significant water drop on the camera lens for object detection and semantic segmentation tasks.

In this work, we focus on soiling caused by a variety of unwanted particles reposing on the camera lens. The source of these particles is mostly mud, dirt, water, or foam created by a detergent. Based on the state of aggregation, such soiling can be either static (e.g., highly viscous mud tends to dry up very quickly, so it does not change its position on the output image over time) or dynamic (mostly water, and foam). Because of that, the acquisition of suitable data for training machine learning models or merely testing the effect on existing classification models is quite tedious. Human annotation is very time demanding, costly, and not very reliable since the precise labeling of soiling on a single image can sometimes be very challenging, e.g., due to the unclear boundary. We want to emphasize that the soiling detection task is necessary for an autonomous driving system as it is used to trigger a camera cleaning system that restores the lens [29]. It is complementary to building segmentation or object detection models robust to soiling without an explicit soiling detection step. Our contributions include:

- A proposition of a baseline pipeline for an opaque soiling generation, based on CycleGAN [34] and soiling segmentation learned from weak labels.
- A novel DirtyGAN network, which is an end-to-end generalization of the baseline pipeline.
- Public release of an artificial soiling dataset as a companion to the recently published *WoodScape Dataset* [33], coined *Dirty WoodScape Dataset*, to encourage further research in this area.
- An empirical evaluation of the degradation of semantic segmentation algorithms on soiled images.

The rest of the paper is organized as follows. Section 2 covers the related work and contrasts the soiling scenarios with adverse weather conditions. In Section 3, we give a detailed description of the proposed algorithms. Section 4 provides an evaluation of the quality of the generated images and quantifies the improvement of the soiling detection algorithm by adding generated images to training. Also, it describes the empirical evaluation of the degradation of semantic segmentation in the presence of soiling. Finally, Section 5 concludes the paper.

2. Related Work

As the visual perception modules for autonomous driving are becoming more mature, there is much recent effort to make them more robust to adverse weather conditions. It can be seen by the popular CVPR workshop "Vision for all seasons"[1], which focuses on the performance of computer vision algorithms during adverse weather conditions for autonomous driving. However, there is very little work on the related but different lens soiling problem. The two problems are similar in how they degrade image quality and can severely affect visual perception performance. Yet there are substantial differences.

The first significant difference is that soiling on the lens can be removed by a camera cleaning system that either sprays water or uses a more sophisticated ultrasonic hardware [29]. Secondly, there is temporal consistency for soiling where mud or water droplets remain static typically or sometimes have low-frequency dynamics of moving water droplets in contrast to higher variability in adverse weather scenes. Thus this temporal structure can be exploited further for soiling scenarios. Finally, soiling can cause more severe degradation as opaque mud soiling can completely block the camera.

We focus on the generic soiling detection task in this paper. Even disregarding camera cleaning, soiling detection is still needed to increase the uncertainty of vision algorithms in the degraded areas. The task of soiling detection on camera lenses in autonomous driving is shortly described in [27]. The authors present a sort of a proof of concept idea on how Generative Adversarial Networks (GANs) [7] could

[1] https://vision4allseasons.net/

be applied for dealing with the insufficient data problem in terms of an advanced data augmentation. In the same paper, the authors also outline another potential usage of GANs in the autonomous driving area. A more formal introduction to the soling detection and categorization is provided in [29]. The problem is formalized as a multilabel classification task and also discusses the applications of soiling detection, including the camera cleaning. Uřičář et al. [28] provided a desoiling dataset benchmark.

The rest of this section provides an overview of the commonly used GAN based image-to-image translation framework, which we use in the proposed method. Due to the limited work on soiling, we also review the closely related area of adverse weather scenarios.

2.1. GAN based Image-to-Image Translation

In recent years, the task of artificial image generation is dominated by GANs [7], showing a great ability to synthesize realistic-looking images in [18]. The Image-to-Image translation is a part of graphics and computer vision that aims to learn a mapping between a source domain X and a target domain Y with the use of paired data. In [11], the authors present a method using GANs to tackle the problem of image-to-image translation using paired data. However, obtaining such paired data can be difficult and sometimes even impossible. Therefore, an unsupervised version, without any examples of corresponding pairs, is even more critical and challenging.

This problem is tackled by CycleGAN [34] with the use of two mappings $G\colon X \to Y$ and $F\colon Y \to X$. Since these mappings are highly under-constrained, they propose to use a *cycle consistency loss* to enforce $F\left(G\left(X\right)\right) \approx X$. Even though it is not emphasized that much in the CycleGAN paper, the authors in their implementation use also *identity losses*, $G(Y) \approx Y$ and $F(X) \approx X$, which improve the results significantly.

2.2. Rainy Scenes

Rainy scenes are slightly related to water soiling situation. Because this degradation is semi-transparent and some background information is still visible, it is common to use an image restoration algorithm that improves the image's quality. The work of [15] provides a comprehensive analysis of this topic. The authors of [32] address the problem of rain removal from videos by a two-stage recurrent network. The rain-free image is estimated from the single rain frame at the first stage. This initial estimate serves as guidance along with previously recovered clean frames to help to obtain a more accurate, clean frame at the second stage.

In [21], the authors propose a progressive recurrent deraining network by repeatedly unfolding a shallow ResNet with a recurrent layer. In [30], a dataset of $\approx$ 29.5k rain/rain-free image pairs are constructed, and a SPatial Attentive Network (SPANet) is proposed to remove rain streaks in a local-to-global manner. Porav et al. [19] presented a method that improves the segmentation tasks on images affected by rain. They also introduced a dataset of clear-soiled image pairs, which is used to train a denoising generator that removes the effect of real water drops. Li et al. [14] proposed a two stage algorithm incorporating depth-guided GAN for heavy rain image restoration. Quan et al. [20] use double attention mechanism CNN for raindrop removal.

2.3. Dehazing

Another type of image quality degradation is caused by the presence of aerosols (e.g., mist, fog, fumes, dust, ...) in the environment surrounding the car. Due to the light scattering caused by these aerosol particles, the resulting image tends to have faint colors and looks hazy, which can inherently also impact the further image processing.

Fattal presents in [6] a method for single image dehazing, based on a refined image synthesis model and a depth estimation. Berman et al. [2], on the other hand, propose a solution, which is not based on local priors and builds on the assumption that a dehazed image can be approximated by a few hundred distinct colors which form tight clusters in the RGB color space. Ki et al. [12] propose fully end-to-end learning-based boundary equilibrium GANs to perform an ultra-high-resolution single image dehazing. Yan et al. [31] propose a semi-supervised learning using a mixture of real data without ground truth and synthetic data.

3. Artificial Soiling Generation

The task of single image soiling annotation on the fisheye cameras is quite tedious. We make use of polygonal annotation, which is a compromise of annotation speed and quality. However, even this kind of polygonal annotation is sometimes tough to interpret even by human annotators. It is particularly true for the soiling boundary, which is usually very fuzzy.

However, an even bigger problem is how to obtain the soiling data. In our setup, we apply a random pattern of soiling on the camera lens using a particular soiling source. Then drivers ride the car for a while and repeat the process several times. It has many limitations: Firstly, it is very inconvenient to record data for all probable scenarios (e.g., driving through the city, rural area, highway). Secondly, it is not possible to measure the real impact of soiling on the images, because we need a clean version of the same images for a fair comparison.

All these limitations motivate us to use the synthetic soiling data. In the following sections, the proposed soiling generation algorithms are described.

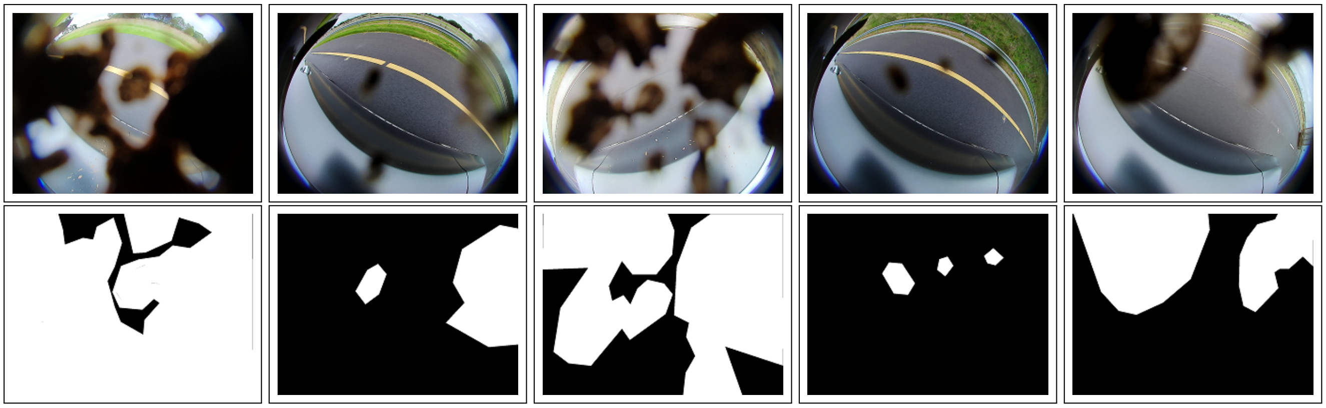


Figure 3: Several examples from the *WoodScape Dataset*. Top: RGB images from the fisheye camera. Bottom: corresponding human made annotations. White polygons represent soiling masks, while the background is marked by black color.

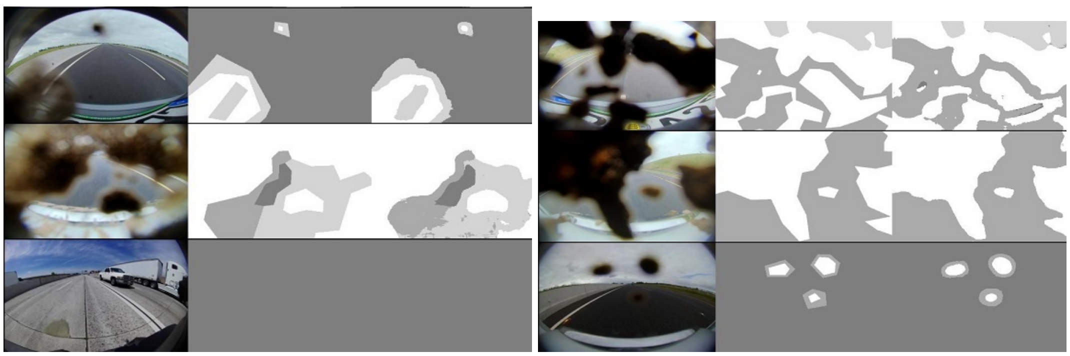


Figure 4: A few examples of the segmentation network results. In the left column are original soiled images, in the middle are the coarse annotations, and on the right column are the masks obtained by the segmentation network. The clean area is marked in dark gray, the opaque area is white, and the semi-transparent region light gray.

3.1. Soiling Generation Baseline Pipeline

The core of our baseline pipeline is formed by a CycleGAN [34] network, which we train to perform the image-to-image translation from clean images to their soiled counterparts. The main problem of the CycleGAN method is that it modifies the whole image. For our desired application, this can lead to undesired artifacts in the generated images. Besides that, the generated synthetic soiling patterns are often relatively realistic. Note that due to GPU memory requirements and time constraints, our CycleGAN training uses rescaled images ($\frac{1}{4}$ of both width and height).

Next, we train a soiling semantic segmentation network, $\mathcal{M}$, using the weak polygonal annotation of soiling (see Figure 3 for several examples). Even though the annotation is quite coarse, the segmentation network $\mathcal{M}$ is able to fit the soiling patterns more precisely. See Figure 4 for a few examples of $\mathcal{M}$ outputs in comparison to the original annotations. We use *WoodScape Dataset* [33] for training the soiling segmentation network. Last but not least, we train a super-resolution network $\mathcal{U}$, which we use to transform the GAN generated image to the original image resolution (i.e., up-scaling of $4\times$ factor).

The idea of a baseline data generation algorithm is described in Figure 5. We take the generator transforming a clean image to the soiled image ($\mathcal{G}_{\mathrm{C2S}}$) and apply it to the clean image I. It gives us an image with a random soiling pattern I_{s}. Next, we obtain the soiling mask m. This is achieved by applying the semantic segmentation network on the generated soiled image followed by a Gaussian smoothing filter γ: $m = \gamma\left(\mathcal{M}\left(I_{\mathrm{s}}\right)\right)$. The resulting soiling mask m is an image with values in range $[0, 1]$, where 0 means background, and 1 means soiling. The intermediate values can be understood as semi-transparent soiling. We apply the Gaussian smoothing filter because it mimics the physical nature of the soiling phenomenon where the edges of the soiling patterns are typically semi-transparent, due to photon scattering. Finally, the artificially soiled version of the original image $\hat{I}$ is a composition of the original image I and the soiling pattern I_{s} via the estimated mask m:

$$\hat{I} = \left(1 - \mathcal{U}\left(m\right)\right) \cdot I + \mathcal{U}\left(m\right) \cdot \mathcal{U}(I_{\mathrm{s}}) \,. \tag{1}$$

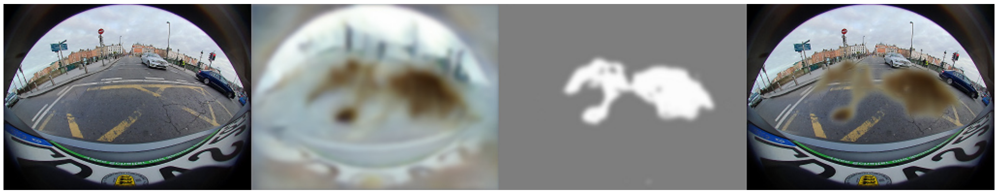


Figure 5: The soiling generation baseline pipeline. From left to right: an image into which we would like to paint a random soiling pattern; CycleGAN generated "soiled" version; blurred mask of the segmented soiling from the generated image; the resulting artificially "soiled" image obtained by convex combination of the original image and the generated soiling via the segmented mask.

Note, it is possible to use arbitrary images for the final composition, once we obtain I_s and m. The mask m obtained by the semantic segmentation network $\mathcal{M}$ serves as an automatic annotation of the soiling in the generated image.

This simple pipeline has certain limitations. The biggest one is that it cannot be expected to work smoothly for soiling types caused by water (e.g., raindrops, snow) in this specific formulation. One option for dealing with this issue is to follow the approach of [1], where the authors model the reflection patterns of the water drops using the whole image and apply filters and transformations consequently. The other option is to formulate a CyleGAN-like approach, which can cope with changing only those parts of the image that correspond to the soiling pattern and keep the rest unchanged. We formulate this approach in the following section.

3.2. DirtyGAN

The problem of applying CycleGAN in artificial soiling synthesis is twofold. Firstly, CycleGAN does not constrain the image generation to any specific regions and instead regenerates the whole image, affecting all pixels. In the case of the artificial soiling generation, this is highly undesirable. For the investigation of the soiling impact on the further image processing, the background (i.e., regions of the image not affected by the soiling) must remain untouched. Secondly, the generation branch "clean" $\rightarrow$ "soiled" is ill-posed, as there is no visual clue for where the soiling should be produced. There are infinitely many patterns that could be created. Furthermore, there is also no control over the soiling pattern production process.

The first problem can be addressed, e.g., by InstaGAN [17]. However, in such a case, the second problem becomes even a more significant issue. We decided to guide the pattern generation process via a Variational AutoEncoder (VAE) [5] and modify the CycleGAN algorithm so that it applies only on the masked regions of the source and target domain images. We coin the proposed network DirtyGAN.

We use the weak polygonal soiling annotations from the *WoodScape Dataset* for training the VAE. The main idea of using VAE for the soiling patterns generation is as follows. By using the encoder of the trained VAE, we can obtain the projection of an actual sample from the dataset to a lower-dimensional representation. If we select two samples $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ that are close on the soiling pattern manifold, we can obtain a novel sample $\boldsymbol{z}$ by taking their convex combination.

$$\boldsymbol{z} = \alpha \boldsymbol{z}_1 + (1-\alpha)\boldsymbol{z}_2 \ , \tag{2}$$

where $\alpha \in [0, 1]$. Then, we take this intermediate representation $\boldsymbol{z}$ and apply the trained decoder from VAE to reconstruct the corresponding soiling pattern. In Figure 6, we depict several examples of this intermediate soiling pattern reconstruction.

The benefit of using sampling from the learned VAE is that we could even use it to create animated masks, e.g., to mimic dynamic soiling effects, such as water drops in heavy rain or to be able to investigate the impact of dynamic soiling in general. After training the VAE, we limit CycleGAN to be applied only on the masked regions corresponding to the generated mask for the "clean" $\rightarrow$ "soiled" translation or the mask obtained by the soiling semantic segmentation mask $\mathcal{M}$. We use a similar composition as in the baseline presented in Section 3.1. In Figure 7, we depict the whole DirtyGAN scheme.

3.3. Dirty Datasets

We have used the baseline pipeline to generate artificial soiling on our recently published *WoodScape Dataset* [33] with 10k images, which comes with semantic segmentation annotation. It makes it a suitable candidate for the soiling generation since we can merge the provided annotation with the soiling mask and measure the direct impact on classification models. Our generated data with updated annotation will be released as a *WoodScape Dataset* companion under the name of *Dirty WoodScape Dataset*[2].

In Figure 8, we show several generated examples together with their automatically obtained annotations. Our method of soiling generation is not limited to fisheye images

[2] https://github.com/valeoai/WoodScape

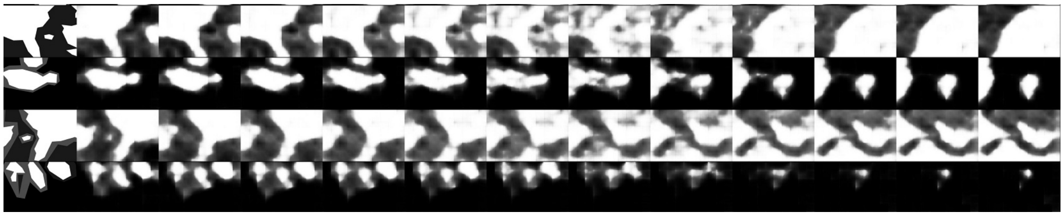


Figure 6: Variational AutoEncoder and the walk on the soiling manifold. The leftmost column depicts the original soiling pattern (the annotation for some particular soiled image). The next column is the reconstructed version by applying the whole VAE. The next 10 columns represent the transition (the walk on the soiling manifold) from the leftmost image to the rightmost one, which represents another example.

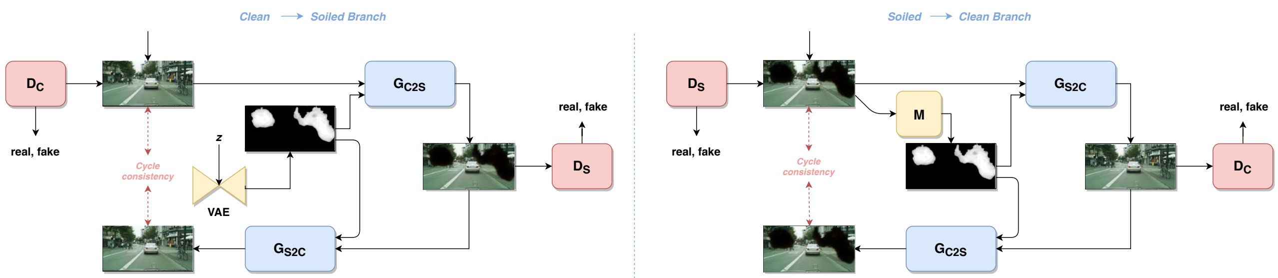


Figure 7: The DirtyGAN scheme. The original CycleGAN scheme is enhanced by VAE for novel soiling pattern generation in "clean" $\rightarrow$ "soiled" branch and by mask estimation in the opposite branch. Please see the supplementary material for a higher resolution version of this image.

only. Since we would like to support standard benchmarking as well, we also release a Dirty Cityscapes dataset. It is, as the name suggests, based on the Cityscapes [4] dataset.

4. Experimental Evaluation

4.1. Data Augmentation for Soiling Detection Task

A primary purpose of generating soiling is to mimic the real soiling scenarios. Firstly, we trained a soiling segmentation network on real soiling data only (8k images). We tested this model performance on real soiling test data and achieved 73.95% accuracy. Then, we added generated soiling data (8k images) to our training catalog. As the generated soiling replicates the real soiling patterns, the network's performance increased to 91.71% on real soiling test data, forming a 17.76% increase in accuracy without the need for costly annotations and real-time soiling scene captures (see the second and fourth row of Table 1).

We observe a significant reduction in accuracy when the network is trained on artificial images solely. It can be attributed to a limited ability to capture the entire data diversity to support the original soiled data. We also conduct a simple ablation study with standard data augmentation techniques (flipping, contrast changes) to match the size of training data with generated samples and observe a much lesser improvement in accuracy (third row), compared to the scenario with augmentation by generated synthetic samples.

The classifier used in the experiment uses ResNet50 [8] for an encoder and FCN8 [22] for a decoder. The binary cross-entropy was used as a loss function with the ADAM optimizer [13] for training with a learning rate of 1×10^{-4}. The image resolution was 640×480 pixels.

Table 1: Comparison of Soiling Segmentation model trained on generated and real soiled images. Accuracy is computed on a real test dataset with $2,000$ images.

ResNet50-FCN8 based Soiling Segmentation Model	**Accuracy [%] on real test 2K dataset (mIoU)**
Trained solely on generated images (8000 samples)	47.41
Trained solely on real images (8000 samples)	73.95
Trained on real & data augmentation (16000 samples)	78.20
Trained on real & generated images (16000 samples)	**91.71**

4.2. Artificial Soiling Quality

We performed a subjective visual study of the quality of the artificial soiling similarly as it is done in other GAN based image generation algorithms. We selected represen-

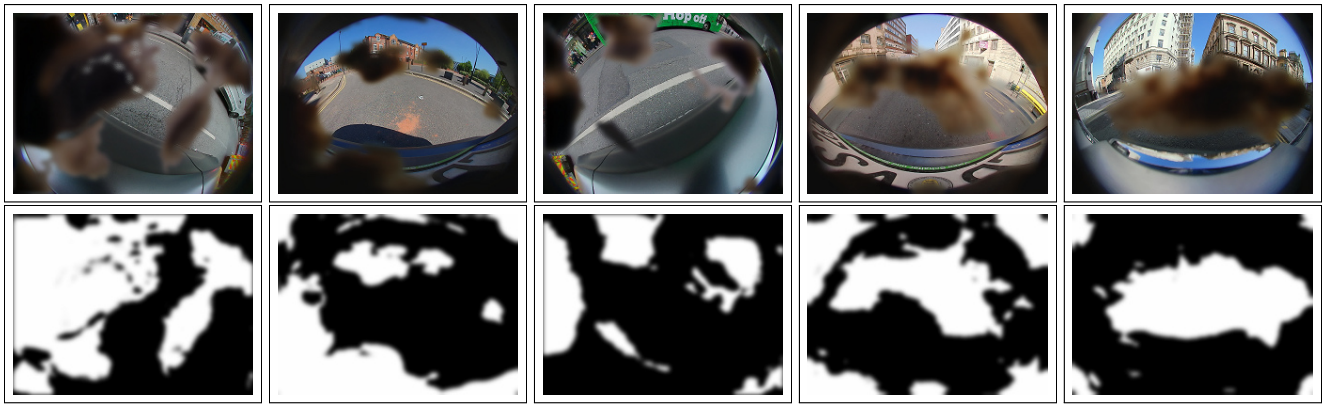


Figure 8: Examples of the generated images from the *DirtyWoodScape Dataset*, together with the generated annotations. Top: *WoodScape Dataset* RGB images with generated opaque soiling. Bottom: corresponding automatically generated annotations. White pixels represent the soiling, while black pixels represent the background.

tative human participants with a variable level of the problem understanding: ranging from absolutely no knowledge about the computer vision and surround-view camera imaging systems for autonomous driving to people working with the soiling data daily. The participants were asked to classify the presented images either as real ones or fakes. Then, we randomly showed images with the real soiling from the *WoodScape Dataset* [33] and with the artificially generated soiling. To make the task even more complicated, we used soiled images showing similar scenes as they occur in the *WoodScape Dataset*. Otherwise, the participants might eventually spot the differences in the data distribution as the real soiling data comes only from limited scenarios.

Table 2: Quantitative evaluation on our *WoodScape* dataset using DeepLabV3+. The accuracy measure is mIoU [%].

Train\Test	**Clean**	**Soiled**
Clean	56.6	34.8
Soiled	52.1	48.2

Table 3: Quantitative evaluation on Cityscapes using DeepLabV3+. The accuracy measure is mIoU [%].

Train\Test	**Clean**	**Soiled**
Clean	38.1	26.6
Soiled	35.5	38.0

The non-expert participants were not able to recognize real images from fakes. The expert participants were sometimes able to spot a difference in the soiling pattern because it was novel, and also they sometimes spotted small artifacts, e.g., blurry texture. However, expert participants were biased as they know the real data. In general, we can say that the image quality of the generated artificial soiling is satisfactory when judged by human inspection for 95% of scenarios.

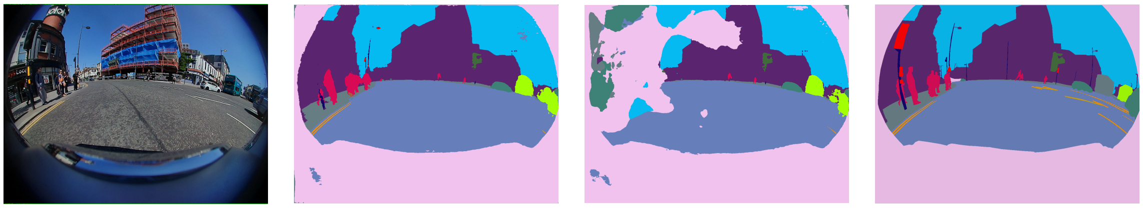
We chose a random subset of 5% of generated samples for quantitative comparison due to the manual annotation cost. We use the mIoU metric to compare between artificially generated annotation and manual annotation. We obtain a high accuracy of 93.2%. It is to be noted that manual annotation can be subjected, especially at boundaries, and sometimes it is worse than artificial annotation based on visual inspection.

4.3. Degradation Effect on Semantic Segmentation

To demonstrate the impact of the soiling on the camera lens, we provide an empirical evaluation on semantic segmentation task. Pixel level classification is commonly used in autonomous driving applications. Our dataset consists of 10k pairs of images (clean and synthetically soiled using the proposed baseline framework). We split our data into training and testing by $80 : 20$ ratio according to the sampling strategy described in [26]. We trained two DeepLabV3+ [3] models on the clean and soiled images, respectively. We evaluate the performance separately on clean and soiled test data. Table 2 summarizes the obtained results.

A segmentation model trained on clean images records 56.6% mIoU on clean test data and 34.8% on soiled data, a performance drop of 21.8% compared to clean images. This significant drop shows that soiling can cause severe degradation to a standard visual perception task in autonomous driving.

A model trained on the synthetic soiling data shows a limited degradation to 16.1%. However, training on the soiled images shows a 4% accuracy degradation on clean test data compared to the baseline when evaluated on clean images. Note, that model trained on the synthetic data perform reasonably well on the clean data with a few percentage drop in mIoU. This is expected as the data contain additional class, which is treated as background/unknown (void class). Thus, the trained model used less portion of data to



(a) Clean Image (b) Train clean, test clean (c) Train clean, test soil (d) Ground Truth

(e) Soiled Image (f) Train soil, test clean (g) Train soil, test soil

Figure 9: Qualitative evaluation using clean vs soiled images on the semantic segmentation network DeepLabV3+.

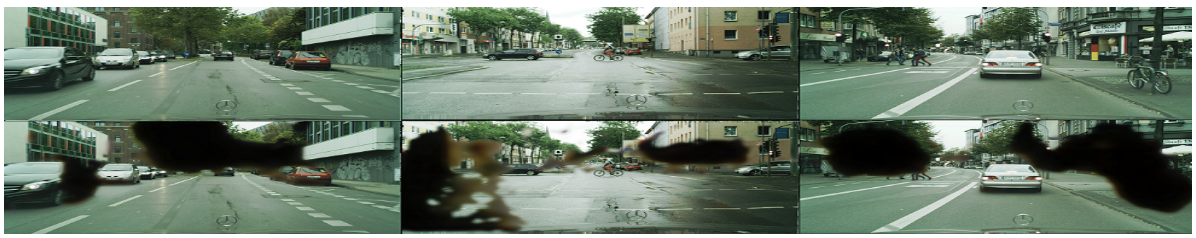


Figure 10: Examples of the generated images from the *Dirty Cityscapes Dataset*. Top: *Cityscapes Dataset* RGB images. Bottom: Generated soiled images corresponding to clean images.

train the remaining classes.

Figure 9 depicts a qualitative evaluation of the soiling impact on the segmentation task. The baseline model trained on clean images 9a is evaluated on soiled images in 9c showing a high level of degradation due to the soiling. Figure 9d shows the ground truth annotations, while 9f and 9g illustrate results of model trained on the soiled images while testing on the clean images in 9a and soiled images in 9e. In a realistic scenario, annotations are not available for the occluded region of soiled images. Using our GAN generated dataset, we use annotations in the soiled area to enable the model to interpolate segmentation classes in occluded soiled parts. Figure 9f shows the capability of segmentation networks to perform segmentation even behind the soiled area. However, it is less reliable compared to the clean baseline and sensitive to overfitting.

The same type of experiments were conducted using the Cityscapes dataset [4]. The results are presented in Table 3, which, as you can see, show a similar trend as in the WoodScape experiment.

5. Conclusions

In this paper, we proposed two algorithms for the generation of soiling on images from surround view fisheye cameras. The first algorithm is a pipeline built from several well-known blocks, such as CycleGAN [34], semantic segmentation of the generated soiling, and image composition. The second algorithm is a novel DirtyGAN network, which can generate similar results as the baseline pipeline in an end-to-end fashion. The possibility to generate random but realistic soiling patterns on camera images is an integral component in examining the degradation of other image processing methods. We provided an empirical evaluation of the performance degradation on several typical classification tasks common for autonomous driving scenarios. We demonstrate that our soiling model trained on fisheye images generalizes well on the Cityscapes dataset, enabling us to create dirty versions of public datasets. Last but not least, we release a public dataset as a companion to the recently published *WoodScape Dataset* [33], coined *Dirty WoodScape Dataset*, which can serve as a benchmark for measuring the degradation of the off-the-shelf classification algorithms.

Acknowledgements. Antonín Vobecký was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS18/205/OHK3/3T/37. All authors would like to thank Valeo for the support of publishing.

References

[1] Stefano Alletto, Casey Carlin, Luca Rigazio, Yasunori Ishii, and Sotaro Tsukizawa. Adherent raindrop removal with self-supervised attention maps and spatio-temporal generative adversarial networks. In *The IEEE International Conference on Computer Vision (ICCV) Workshops*, 2019. 5

[2] Dana Berman, Tali Treibitz, and Shai Avidan. Non-local image dehazing. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, 2016. 3

[3] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, 2018. 7

[4] 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. 6, 8

[5] Carl Doersch. Tutorial on variational autoencoders. In *CoRR*, 2016. 5

[6] Raanan Fattal. Single image dehazing. In *ACM Trans. Graph.*, 2008. 3

[7] 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*, 2014. 2, 3

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *Proc. of CVPR*, 2016. 6

[9] Markus Heimberger, Jonathan Horgan, Ciarán Hughes, John McDonald, and Senthil Yogamani. Computer vision in automated parking systems: Design, implementation and challenges. In *Image and Vision Computing*. Elsevier, 2017. 1

[10] S. A. E. International. Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems J3016, 2018. 1

[11] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, 2017. 3

[12] Sehwan Ki, Hyeonjun Sim, Jae-Seok Choi, Saehun Kim, and Munchurl Kim. Fully end-to-end learning based conditional boundary equilibrium GAN with receptive field sizes enlarged for single ultra-high resolution image dehazing. In *2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2018, Salt Lake City, UT, USA, June 18-22, 2018*, 2018. 3

[13] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In *Proc. of ICLR*, 2015. 6

[14] Ruoteng Li, Loong-Fah Cheong, and Robby T. Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 1633–1642. Computer Vision Foundation / IEEE, 2019. 3

[15] Siyuan Li, Iago Breno Araujo, Wenqi Ren, Zhangyang Wang, Eric K. Tokuda, Roberto Hirata Junior, Roberto Cesar-Junior, Jiawan Zhang, Xiaojie Guo, and Xiaochun Cao. Single image deraining: A comprehensive benchmark analysis. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3

[16] Pullaro Maddu, Wayne Doherty, Ganesh Sistu, Isabelle Leang, Michal Uricar, Sumanth Chennupati, Hazem Rashed, Jonathan Horgan, Ciaran Hughes, and Senthil Yogamani. Fisheyemultinet: Real-time multi-task learning architecture for surround-view automated parking system. In *Proceedings of the Irish Machine Vision and Image Processing Conference*, 2019. 1

[17] Sangwoo Mo, Minsu Cho, and Jinwoo Shin. Instagan: Instance-aware image-to-image translation. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, 2019. 5

[18] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, 2019. 3

[19] Horia Porav, Tom Bruls, and Paul Newman. I can see clearly now : Image restoration via de-raining. In *CoRR*, 2019. 3

[20] Yuhui Quan, Shijie Deng, Yixin Chen, and Hui Ji. Deep learning for seeing through window with raindrops. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 2463–2471. IEEE, 2019. 3

[21] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3

[22] Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017. 6

[23] Mennatullah Siam, Sara Elkerdawy, Martin Jagersand, and Senthil Yogamani. Deep semantic segmentation for automated driving: Taxonomy, roadmap and challenges. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017. 1

[24] Mennatullah Siam, Mostafa Gamal, Moemen Abdel-Razek, Senthil Yogamani, and Martin Jagersand. Rtseg: Real-time semantic segmentation comparative study. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018. 1

[25] Mennatullah Siam, Heba Mahgoub, Mohamed Zahran, Senthil Yogamani, Martin Jagersand, and Ahmad El-Sallab. MODNet: Motion and appearance based moving object detection network for autonomous driving. In *Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018. 1

[26] Michal Uřičář, David Hurych, Pavel Křížek, and Senthil Yogamani. Challenges in designing datasets and validation for autonomous driving. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and

Computer Graphics Theory and Applications - Volume 5: VISAPP,. SciTePress, 2019. 7

[27] Michal Uřičář, Pavel Křížek, David Hurych, Ibrahim Sobh, Senthil Yogamani, and Patrick Denny. Yes, we GAN: applying adversarial techniques for autonomous driving. In *Electronic Imaging*, 2019. 2

[28] Michal Uricár, Jan Ulicny, Ganesh Sistu, Hazem Rashed, Pavel Krizek, David Hurych, Antonin Vobecky, and Senthil Yogamani. Desoiling dataset: Restoring soiled areas on automotive fisheye cameras. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2019. 3

[29] Michal Uřičář, Pavel Křížek, Ganesh Sistu, and Senthil Yogamani. Soilingnet: Soiling detection on automotive surround-view cameras. In *2019 22nd International Conference on Intelligent Transportation Systems (ITSC)*, 2019. 2, 3

[30] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson W.H. Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3

[31] Wending Yan, Aashish Sharma, and Robby T. Tan. Optical flow in dense foggy scenes using semi-supervised learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 13256–13265. IEEE, 2020. 3

[32] Wenhan Yang, Jiaying Liu, and Jiashi Feng. Frame-consistent recurrent video deraining with dual-level flow. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3

[33] Senthil Yogamani, Ciarán Hughes, Jonathan Horgan, Ganesh Sistu, Padraig Varley, Derek O'Dea, Michal Uřičář, Stefan Milz, Martin Simon, Karl Amende, Christian Witt, Hazem Rashed, Sumanth Chennupati, Sanjaya Nayak, Saquib Mansoor, Xavier Perroton, and Patrick Perez. WoodScape: A multi-task, multi-camera fisheye dataset for autonomous driving. In *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 2, 4, 5, 7, 8

[34] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, 2017. 2, 3, 4, 8