Desoiling Dataset: Restoring Soiled Areas on Automotive Fisheye Cameras

Michal Uřičář[1], Jan Uličný[3], Ganesh Sistu[2], Hazem Rashed[4], Pavel Křížek[1], David Hurych[1], Antonín Vobecký[1] and Senthil Yogamani[2]

[1]Valeo R&D Prague, Czech Republic [2]Valeo Vision Systems, Ireland [3]Valeo Vision Systems, Germany
[4]Valeo GEEDS, Egypt

firstname.lastname@valeo.com

Abstract

Surround-view cameras became an integral part of autonomous driving setup. Being directly exposed to harsh environmental settings, they can get soiled easily. When cameras get soiled, the degradation of performance is usually more dramatic compared to other sensors. Having this on mind, we decided to design a dataset for measuring the performance degradation as well as to help constructing classifiers for soiling detection, or for trying to restore the soiled images, so we can increase the performance of the off-the-shelf classifiers. The proposed dataset contains 40+ *approximately* 1 *minute long video sequences with paired image information of both clean and soiled nature. The dataset will be released as a companion to our recently published dataset* [14] *to encourage further research in this area. We constructed a CycleGAN architecture to produce de-soiled images and demonstrate 5% improvement in road detection and 3% improvement in detection of lanes and curbs.*

1. Introduction

The advances in autonomous driving show that combination of sensors is a necessary step to achieve difficult safety and reliability standards. Surround view cameras are becoming *de facto* standard in autonomous parking, where they significantly contribute to ultrasonic sensors by resolving difficult scenarios, such as fishbone parking or detecting a free parking spot outlined only by ground markings, which is completely not resolvable by using ultrasonic sensors solely. Some influential people even believe that cameras could replace expensive sensors like Lidars.

However, the surround view cameras are directly exposed to the environment, which can, sometimes, be very harsh. In certain conditions, e.g. heavy rain, snow or offroad driving, the surround view cameras can get soiled quite easily. When this happens, all processing based on the imagery acquired by these cameras is put in danger. The performance usually degrades dramatically, depending on the severity of the soiling. While there already exists studies for dealing with rain and water drops on the cameras [6, 12, 4, 7, 10, 13], not much was done with respect to other possible soiling sources.

Our contribution is two-fold. Firstly, we release a new dataset which contains both pairiness and temporal information for dealing with several soiling categories, as well as their mixture. Secondly, we propose a baseline image restoration method and perform extensive experimental evaluation, showing that even this baseline can help in leveraging the performance of the classical semantic segmentation classifier used in autonomous driving tasks.

The paper is organized as follows. In Section 2, we briefly introduce the problematics of the soiling on cameras in automotive area. Section 3 describes the proposed dataset design and acquisition. In Section 4, we formulate the baseline solution for the restoration of soiled images and summarizes the results we obtained. Finally, Section 5 concludes the paper.

2. Soiling on Automotive cameras

Automotive cameras are exposed directly to contaminants including mud, dust, rain, fog and snow. They get deposited on the lens as illustrated in Figure 1 and can cause severe degradation of quality of computer vision algorithms. Soiling of cameras on handheld consumer devices occurs commonly but it can be cleaned easily and does not pose a safety risk. For autonomous driving systems, it is essential to reliably detect soiling on the lens and notify the driver. Convolutional Neural Networks (CNN) and some classical geometric algorithms like optical flow are global operators and even a small soiled region could cause a severe degradation. Thus even when soiling is detected in a localized region, partial availability of the algorithms in other clean parts could be less reliable.

There is very little literature on this topic and it requires more attention to enable robust self driving. Soiled areas can be classified as opaque (mud, dust, snow) and transpar-

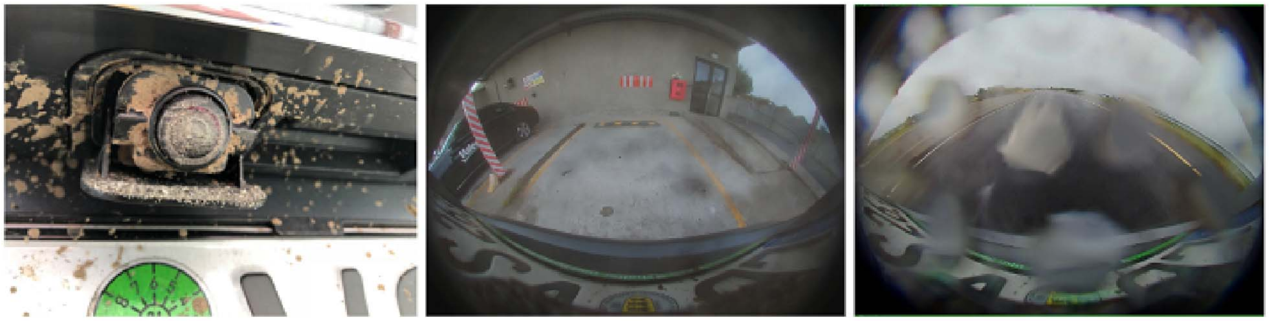


Figure 1: From left to right: a) soiled camera lens mounted to the car body; b) the image quality of the soiled camera from the previous image; c) an example of image soiled by a heavy rain.

ent (water). Transparent soiling in particular can be challenging to detect because of partial visibility of the background. Some advanced systems trigger a cleaning system using a water spray or air blower based on the soiling detection. Transparent soiling was addressed in recent work of Porav et al. [6] where a stereo camera was used in conjunction with a dripping water supply to simulate rain drops on camera lens. The authors also propose a de-raining algorithm using CNN.

There are three main ways to deal with the situation of soiled lens. The best case is when soiling is detected by an algorithm and then a cleaning system is triggered. But cleaning systems are currently uncommon due to their additional cost and maintenance requirements. The second way is to design algorithms which are robust to these scenarios by implicitly handling them in their model, Sakaridis et al. [8] proposed a robust algorithm for semantic segmentation which can deal with foggy scenes. However, opaque soiling will be challenging to be dealt in this manner. The third approach is to run a separate image restoration algorithm to improve the quality of the image. Some recent examples of the restoration algorithms in automotive scenarios are de-raining [6, 4, 7, 10, 13, 12], de-fogging [8] and de-hazing [3]. This would mainly help alleviate partial soiling. Restoration algorithms can be either single image based or video based. The latter is more computationally expensive but it can leverage visibility of soiling occluded regions over time.

3. Overview of the dataset

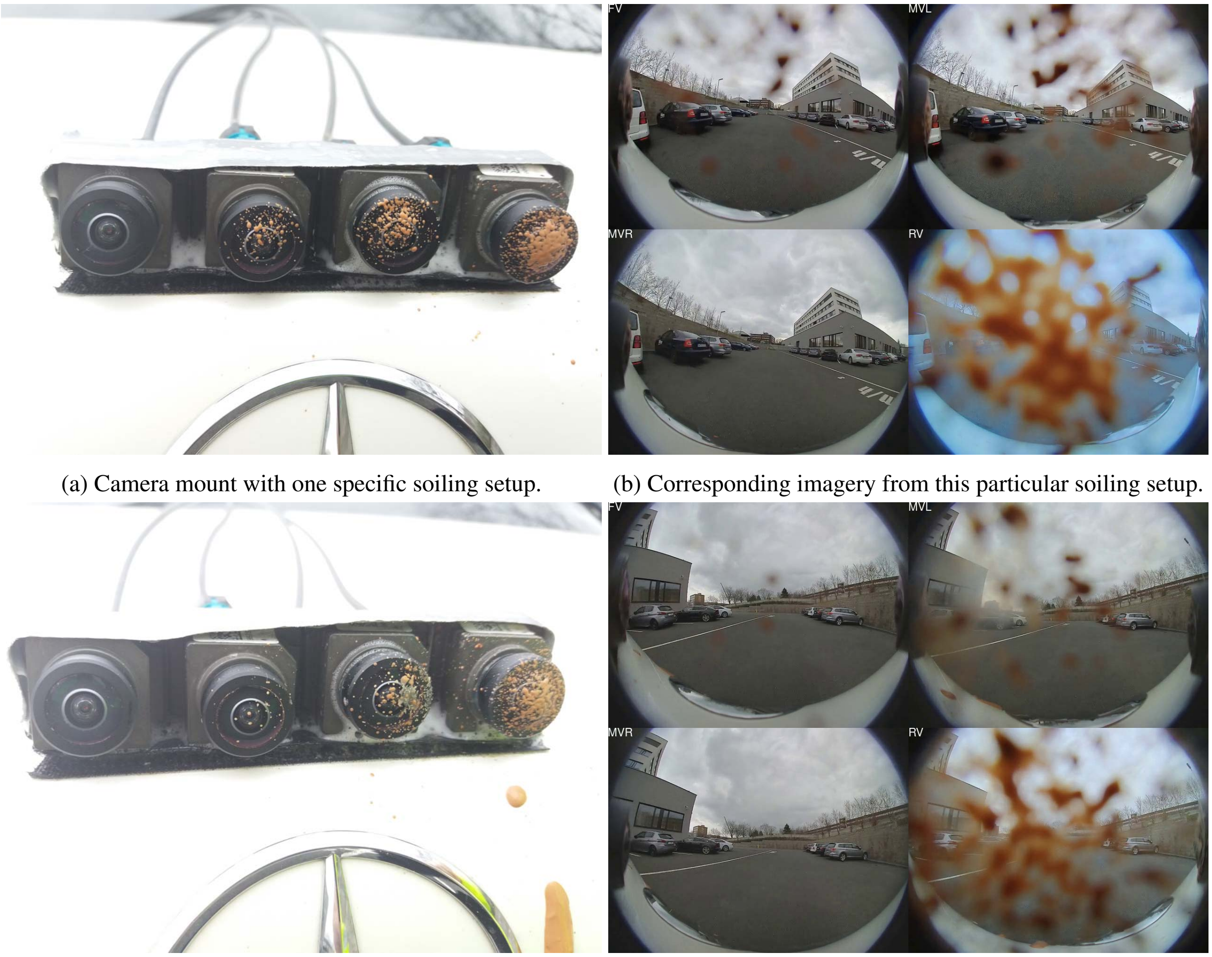
The main goal of the proposed dataset is a restoration of soiled images. The dataset is formed by 40+ video captures, each of them is approximately 1 minute long and contain low speed maneuvering of the car in a close proximity of a parking place. Part of the scenario is also parking between parked cars. Each capture consist of image data from a setup of 4 cameras, that are positioned on the car trunk in a row, one camera next to each other. One camera is always kept clean, while the rest 3 cameras are manually soiled in some way (see Figures 2a and 2c).

3.1. Dataset Acquisition

The data were collected on a small test track of our facility. The test track speed limit is 20 kmph and the data are collected within this speed limit. The test track is located around one building and outlined by a fence and foliage. Some parts of the test track are also reserved as parking lots and there are some line markings on the road as well.

During the acquisition, we were using only one vehicle, with the same camera mount position. The 4 cameras were lined up one next to each other and fastened on the vehicle trunk by a hook-and-loop fastener. The 40+, approximately 1 minute long image sequences were obtained in 3 recording sessions, which were conducted each on a different day with slightly different weather conditions. While one camera was always kept clean, the remaining 3 cameras were manually soiled by a different type of soiling (e.g., ceramic mud of different consistency, ISO mud, muddy water, water or foam from formed by a cleaning agent). For applying the soiling, we used either a toothbrush by which we sprayed randomly the camera hood, or an aerosol spray which we used to spray water drops of different size. In Figure 2, we show both the camera mount and alignment (Figs 2a and 2c) as well as the corresponding imagery from this camera setup (Figs 2b and 2d). Thanks to our setup, it is possible to used both pairiness (clean and soiled image with a small shift in camera position) and temporal information (consecutive frames from the video streams). We believe this is beneficial not only for the task of image restoration, but also for soiling detection and other admissible tasks.

We used similar driving scenario for all sequences. It consisted of a short stay at a starting spot (a place where we were applying the soiling on cameras). Then a short drive around the testing track, parking between parked cars in a reverse motion and then again a short drive through the testing track back to the original position. The driving scenario covers typical classes used for semantic segmentation in au-



(a) Camera mount with one specific soiling setup.

(b) Corresponding imagery from this particular soiling setup.

(c) Camera mount with another specific soiling setup.

(d) Corresponding imagery from this particular soiling setup.

Figure 2: The documentation of the dataset acquisition setup. Figs. 2a and 2c depicts the camera mount and show some examples of the soiling sprayed on the camera hood. Figs 2b and 2d show how this camera setup sees the world. It also roughly shows how the testing track looks like.

tonomous driving, such as building, other vehicles, ground line markings, foliage, and sparsely also pedestrians.

To be GDPR compliant, we blur out all the license plate numbers as well as faces which appear in the sequences. Since the dataset is aimed on the image restoration task, we believe this decision will not negatively impact any proposed solutions.

4. Soiling Restoration Baseline and Results

The main goal of our dataset is to give the research community an opportunity to explore what are the possibilities of image restoration for leveraging other processing algorithms degradation. We propose the following baseline method for soiling imagery restoration.

Since the shift in the physical position of cameras introduces non-affine perturbations of the images, we decided to use the CycleGAN [15] architecture, which should be able to deal with the non-aligned data. The CycleGAN scheme is depicted in Figure 3. It consists of a pair of generators and a pair of discriminators. For the soiling restoration purposes, we are interested only in a single generator, which takes the soiled data on its input and provides "de-soiled"/clean images on its output. However, due to the cycle-consistency, we need all four networks.

We trained the CycleGAN reckless to both the temporal and the pairiness information. We simply sampled $17,828$ images altogether (both clean and soiled) and created the following split: training set ($8,913$ images), validation set ($4,457$ images), and testing set ($4,458$ images). The training images were used to train the four networks in CycleGAN scheme. The validation images were used for displaying the training progress (which we used as the stopping criterion). The testing images were used for the experimental

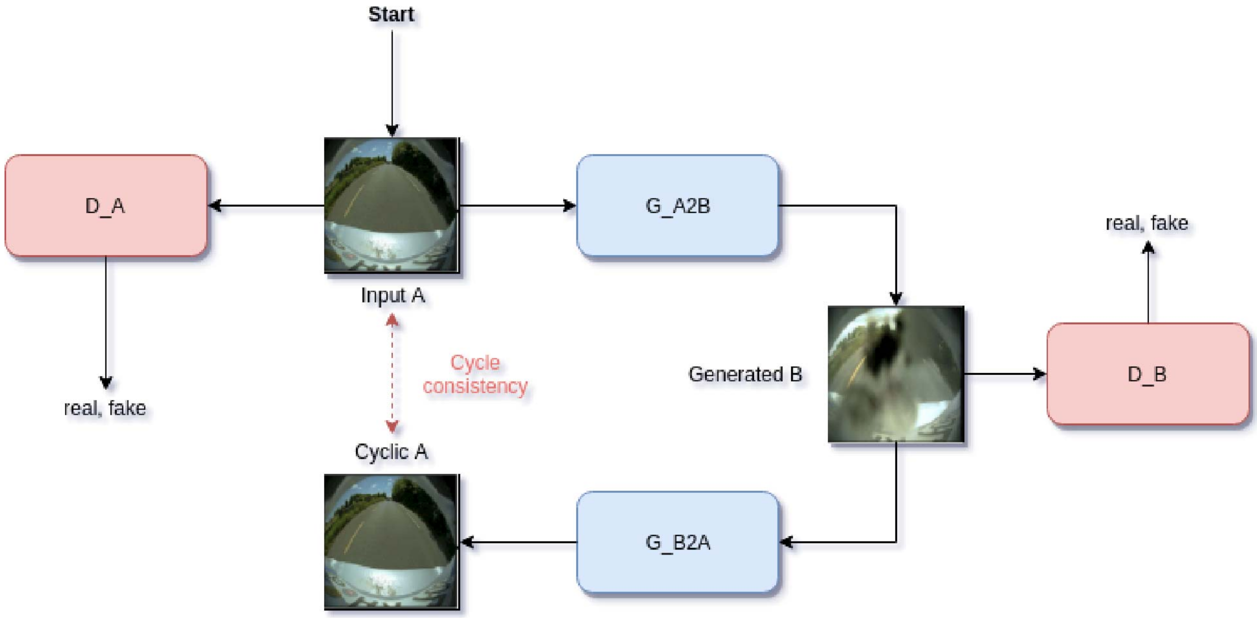


Figure 3: The CycleGAN [15] architecture.

evaluation.

Based on the quality of the generators on validation images and our timeline constraints, we stopped the training after 200 epochs. While the generator which takes the soiled images on its input and produces clean images on the output was working already quite reasonably, the other generator was not so convincing. It was able to introduce only "water"-like soiling in the clean images.

4.1. Results

In this sub-section, we present the results of our baseline networks. We make use of the proposed baseline network based on the CycleGAN [15] to generate the de-soiled images. Qualitative results are illustrated in Figure 4, they show reasonable restoration. We also obtained quantitative results using commonly used image similarity metric Structural Similarity Index (SSIM) [11] and the comparison of mean Intersection over Union (mIoU), which is commonly used to express semantic segmentation accuracy. The results are summarized in Table 1.

In case of the SSIM, the comparison is made between the soiled and de-soiled images using the clean image as a ground truth. We observe an improvement of 6% without any tuning of the algorithm.

To obtain more application oriented metrics, we measured the mIoU scores on semantic segmentation of the road, lanes and curbs classes. Due to the lack of segmentation ground truth on these images, we run the same network on clean images and use it as a ground truth. We observe an improvement of 5% for the road class and 3% improvement for lanes and curb classes. We used the encoder-decoder architecture of the semantic segmentation network with the ResNet-50 [2] encoder and the FCN8 [9] decoder. The network is pre-trained on ImageNet and then trained on our internal fisheye dataset [14].

Table 1: Comparison of accuracy metrics on soiled data vs desoiled data

Accuracy Metric	**Soiled data**	**Desoiled data**
Image Similarity (SSIM)	0.40	0.46
Semantic Segmentation		
Road (IoU)	0.51	0.56
Lanes (IoU)	0.74	0.77
Curbs (IoU)	0.87	0.90

Recent comprehensive benchmark on de-raining [5] concluded that no existing de-raining algorithm helps to improve object detection accuracy. This shows that this is a challenging problem and we have obtained encouraging results using our dataset with the baseline methods to encourage further research into this problem. However transparent soiling on the camera lens has better structure to be exploited than rainy scenes.

In addition to our baseline results, we extend our work by utilizing temporal information. Autonomous driving scenes are highly dynamic where there are strong motion clues due to ego-motion and due to moving objects where collision risk usually arises from moving objects. When there is an area in the FOV that is blocked by a soiled part of the cam-

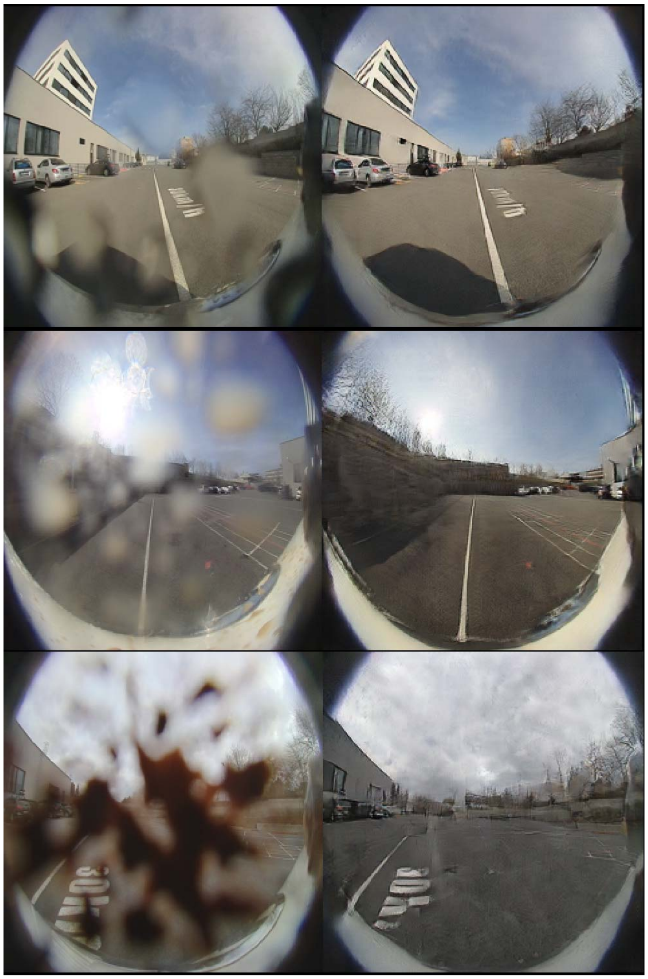


Figure 4: Qualitative results of restored images using our proposed CycleGAN architecture.

era, it is likely to be visible when the vehicle moves as it will be captured by an unsoiled part. On the other hand, if the vehicle is at standstill, moving objects that are not seen due to soiled part are likely to be seen in the upcoming time frames. In this case, a sequence of temporal images can be used to extrapolate the parts that have been hidden by soiled parts and therefore reconstruct the whole scene. We argue that leveraging temporal information is crucial for the de-soiling task. For that purpose, we define our reconstruction problem as an in-painting problem. We make use of the temporal information through optical flow images as demonstrated by [1]. We provide preliminary results as illustrated in Figure 5, where we show the benefit of utilizing the time information. The first column shows the masked input where the black mask represents the soiled part. The second column shows our restoration results on our fisheye dataset [14] and the third column shows the ground truth. The first 4 rows show results on our rear-view camera and the last 3 rows show results on the front-view fisheye images. It is shown that the scene is being restored based on temporal neighbors where the car in the middle is completely or partially masked out, and it was restored correctly due to being seen in the neighboring frames. These results motivate our future work which will extend our experiments to incorporate the time information as well.

5. Conclusions

In this paper, we discussed the problem of soiling on automotive cameras and motivated the possibility of restoration. We created a four-camera setup with varying levels of soiling where one of the images is clean and acts as ground truth. We will make this dataset comprising of $\tilde{1}8$k images public to encourage further exploration of soiling restoration problem which is a nascent area of research in autonomous driving. We construct a baseline using CycleGAN which demonstrates reasonable restoration both qualitatively and quantitatively. We also applied a recent video inpainting algorithm which produces better results than our baseline.

References

[1] Y.-L. Chang, Z. Y. Liu, K.-Y. Lee, and W. Hsu. Free-form video inpainting with 3d gated convolution and temporal patchgan. *In Proceedings of the International Conference on Computer Vision (ICCV)*, 2019. 5, 7

[2] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 4

[3] S. Ki, H. Sim, J.-S. Choi, S. Kim, and M. Kim. Fully end-to-end learning based conditional boundary equilibrium gan with receptive field sizes enlarged for single ultra-high resolution image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 817–824, 2018. 2

[4] S. Li, I. B. Araujo, W. Ren, Z. Wang, E. K. Tokuda, R. H. Junior, R. Cesar-Junior, J. Zhang, X. Guo, and X. Cao. Single image deraining: A comprehensive benchmark analysis. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1, 2

[5] S. Li, I. B. Araujo, W. Ren, Z. Wang, E. K. Tokuda, R. H. Junior, R. Cesar-Junior, J. Zhang, X. Guo, and X. Cao. Single image deraining: A comprehensive benchmark analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3838–3847, 2019. 4

[6] H. Porav, T. Bruls, and P. Newman. I can see clearly now : Image restoration via de-raining. *CoRR*, abs/1901.00893, 2019. 1, 2

[7] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng. Progressive image deraining networks: A better and simpler baseline. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1, 2

[8] C. Sakaridis, D. Dai, and L. Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, pages 1–20, 2018. 2

[9] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(4):640–651, 2017. 4

[10] T. Wang, X. Yang, K. Xu, S. Chen, Q. Zhang, and R. W. Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1, 2

[11] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 4

[12] W. Yang, J. Liu, and J. Feng. Frame-consistent recurrent video deraining with dual-level flow. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1, 2

[13] W. Yang, J. Liu, S. Yang, and Z. Guo. Scale-free single image deraining via visibility-enhanced recurrent wavelet learning. *IEEE Trans. Image Processing*, 28(6):2948–2961, 2019. 1, 2

[14] S. Yogamani, C. Hughes, J. Horgan, G. Sistu, P. Varley, D. O'Dea, M. Uřičář, S. Milz, M. Simon, K. Amende, C. Witt, H. Rashed, S. Chennupati, S. Nayak, S. Mansoor, X. Perroton, and P. Perez. WoodScape: A multi-task, multi-camera fisheye dataset for autonomous driving. *CoRR*, abs/1905.01489, 2019. To appear in ICCV 2019. 1, 4, 5

[15] J. Zhu, T. Park, P. Isola, and A. 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*, pages 2242–2251, 2017. 3, 4

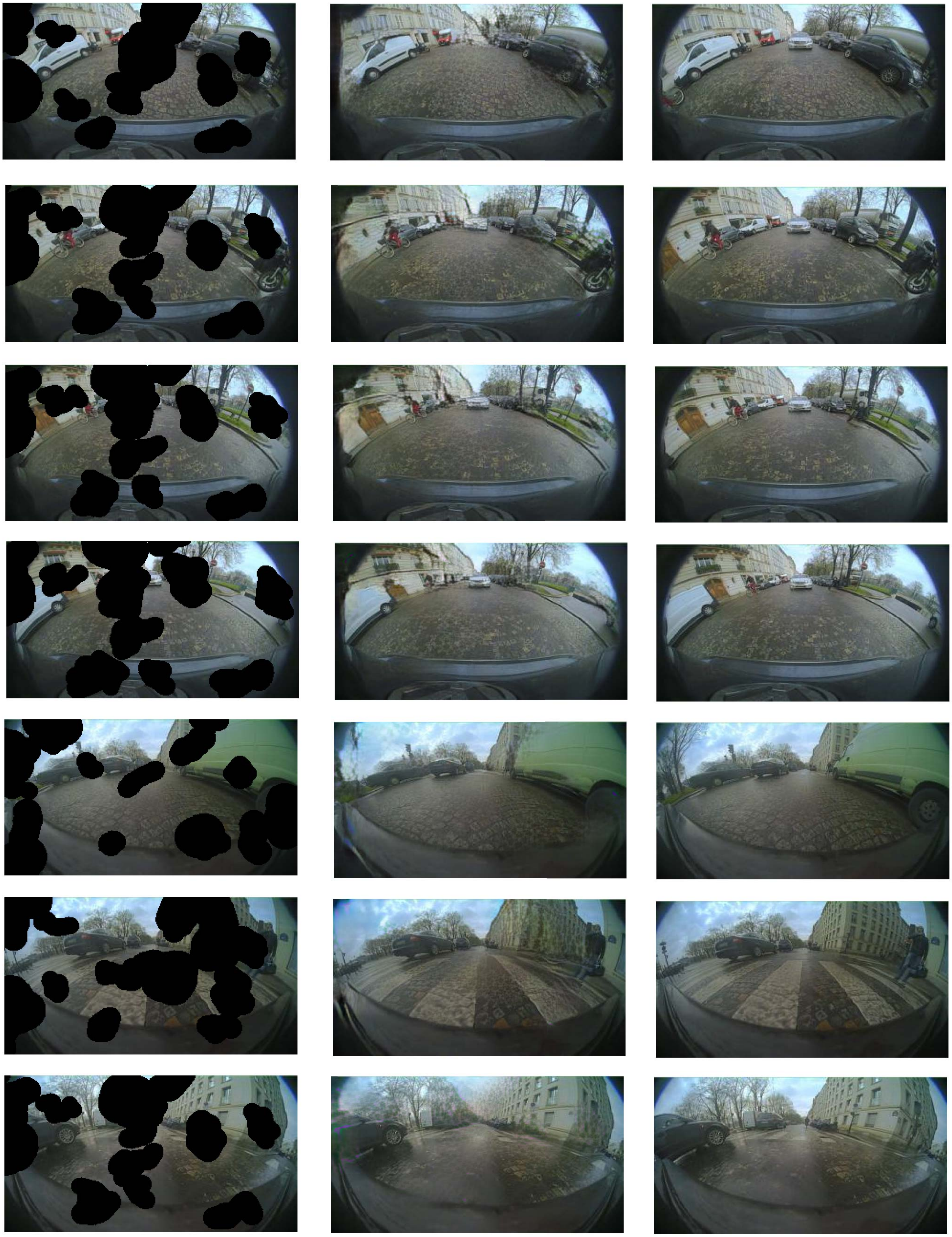


Figure 5: Qualitative results of restored images using Video Inpainting [1]. First column represents the masked images which simulate soiled camera frames. Second column shows the reconstruction results compared to ground truth in the third column.