

I Can See Clearly Now : Image Restoration via De-Raining

Horia Porav, Tom Bruls and Paul Newman

Abstract—We present a method for improving segmentation tasks on images affected by adherent rain drops and streaks. We introduce a novel stereo dataset recorded using a system that allows one lens to be affected by real water droplets while keeping the other lens clear. We train a denoising generator using this dataset and show that it is effective at removing the effect of real water droplets, in the context of image reconstruction and road marking segmentation. To further test our de-noising approach, we describe a method of adding computer-generated adherent water droplets and streaks to any images, and use this technique as a proxy to demonstrate the effectiveness of our model in the context of general semantic segmentation. We benchmark our results using the CamVid road marking segmentation dataset, Cityscapes semantic segmentation datasets and our own real-rain dataset, and show significant improvement on all tasks.

I. INTRODUCTION

If we want machines to work outdoors and see while doing so, they have to work in the rain. When rain and lenses interact, computer vision becomes harder - wild local distortions of the image appear which dramatically impede image understanding tasks. However the distortions are not noise, they are structured, the light field is simply bent and attenuated, and accordingly can be modelled and reversed.

In this work we develop a filter which as a pre-processing step removes the effect of raindrops on lenses. Several tasks are affected by the presence of adherent water droplets on camera lenses or enclosures, such as semantic segmentation [1], localisation using segmentation [2], [3] or road marking segmentation [4]. In this paper we choose to use segmentation as an example task by which to test the effectiveness of our method. Many approaches so far have reached for multi-modal data [5], domain adaptation [6], [7] or training on synthetic data [8], however this can become awkward as:

- 1) Acquiring rainy images is time-consuming, expensive or impossible for many tasks or setups, especially in the case of supervised training, where ground truth data is needed.
- 2) Training, domain-adapting or fine-tuning each individual task with augmented data is intractable.

We take a different approach and build a system as an image preprocessor, the output of which is a cleaned, de-rained image that improves the performance of many tasks performed on the image.

We begin by creating a bespoke real-world small baseline stereo dataset where one lens is affected by real water droplets and the other is kept dry. The methodology and apparatus for doing so is presented in section IV-A. Using this dataset, we train a de-raining generator and show that it



Fig. 1. We learn a de-noising generator that can remove noise and artefacts induced by the presence of adherent rain droplets and streaks. On the top left, input images that are affected by real rain drops. On the top right, the cleaned, de-rained images. On the bottom left, input images that are affected by computer-generated rain drops. On the bottom right, the cleaned, de-rained images.

is able to both drastically improve the visual quality of images and restore performance on road marking segmentation tasks.

Secondly, we describe a way of efficiently adding computer-generated adherent rain droplets and adherent streaks to any image using GPU shaders. This system is presented in section III-A. As the Cityscapes dataset provides a good groundtruth for segmentation but does not contain images with significant rain on the lens, we modify it using this technique and use it as a proxy to study the effects of rain on general semantic segmentation. Additionally, we create a synthetic rain dataset by adding computer-generated rain drops to a full Oxford RobotCar dataset [9] and to the CamVid [10] dataset.

Our main contributions include:

- a de-raining model that produces state of the art results;
- using computer-generated water drops as a proxy to study the effects of rain on segmentation for datasets that provide a ground truth but do not normally contain rainy images; and
- a real-world very-narrow-baseline stereo dataset with rainy & clear images covering a wide array of dynamic scenes.

Our aim is to show that pre-processing the image leads to better performance as compared to training, retraining or fine-tuning a task-specific model with rain-augmented data.

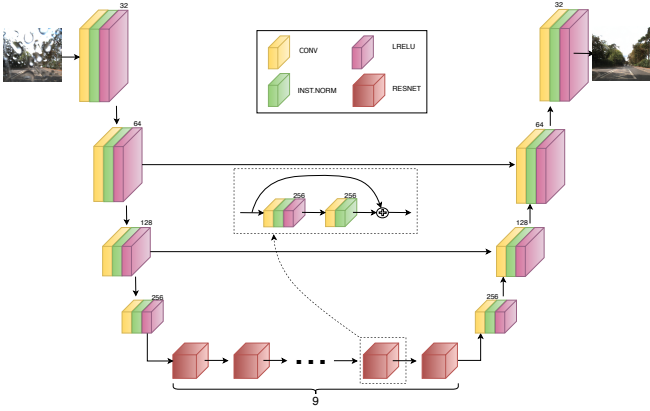


Fig. 2. The internal architecture of our generator. We motivate the addition of additive skip connections by observing that much of the structure of the input image should be kept, along with illumination levels and fine details.

We benchmark our de-raining model on the following tasks:

- Road marking segmentation and image restoration on a real-world small baseline stereo dataset where one lens is affected by real water droplets and the other is kept dry and clear.
- Image reconstruction on the real-world dataset of [11].
- Road marking segmentation and image restoration on CamVid [10] and RobotCar [9] imagery with computer-generated droplets added.
- Semantic segmentation on Cityscapes [1] imagery with computer-generated droplets added.

The quantitative and qualitative results are presented in section V.

II. RELATED WORK

Generally speaking, the quality of an image can be affected in two ways by bad weather conditions. Firstly, contaminants in the atmosphere, such as falling rain, fog, smog or snow will hinder visibility or partially occlude a scene but do not significantly distort the image. Secondly, adherent contaminants such as water droplets, which stick to transparent surfaces or lenses, tend to heavily distort the image, essentially acting as a secondary lens with various degrees of blurring. Several techniques are employed to clean the first type of images, such as those used by [12], [13], [14], [15], [16], however these techniques cannot be used to restore images affected by adherent rain, as the optics involved differ significantly from those of atmospheric droplets. The remainder of this section outlines some of the techniques used to tackle the effects of adherent rain droplets and adherent streaks.

Rain Modelling and Simulation: In the context of computer vision, several studies have attempted to model the structure and optical properties of adherent water droplets. The authors of RIGSEC [17], [18] model raindrops first as sections of a sphere and later account for the effect of gravity using 2D Bezier curves, and confirm experimentally that a physically correct droplet shape can be computed using this method. [19] additionally study and model the dark band around the edges of adherent drops, and show that a simplified model is enough to correctly undistort the image on the surface of the droplet.

We base our simple synthetic droplet model on the works of [17], [18] and [19], by storing proto-droplet normal maps

which are subsequently warped and combined at run time using an approach similar to meta-balls [20].

Additionally, several small datasets have been created to benchmark the accuracy of de-raining techniques. In [21], water is sprayed on a glass pane fitted in front of a camera, but no ground truth is provided due to temporal illumination and scene changes. A video sequence where the lens is affected by real rain droplets is also provided, again without ground truth. The authors of [22] again use a glass pane sprayed with water to study the performance of their droplet detection and removal pipeline, but only offer ground truth for the position of the droplets. The first attempt to provide accurate ground truth is made by [11], in which images of static scenes are captured both with and without a glass pane sprayed with water in front of the camera. This process is, however, very difficult to scale to the number of images required by modern deep-learning approaches. To our best knowledge, we are the first to record a real-world large dataset of sequential dynamic scenes with an accurate, clear ground truth and a large variation in raindrop type and size.

Raindrop Detection and Removal: In [17] and [18], raindrops are detected by attempting to match a template of a synthetic raindrop at locations where the presence of a real drop is hypothesized. This approach breaks down when the shape of the real droplets differs significantly from that of the template. The authors of [22] take a different approach by observing that the motion inside droplets is between 1/30 and 1/20 slower than that in the scene. They use this information to detect raindrops and then attempt to restore the image by using a combination between image inpainting and recovering data from within the distorted image formed on the droplet. Both techniques use multi-frame information for image reconstruction, and are not applicable to single-images.

Multi-camera and pan-tilt setups are exploited by [23], [24], [25] and [26]. These techniques use disparities to detect droplets and subsequently attempt to replace the affected regions in one lens with information from the other lens. This approach does not work on single images and assumes that the same regions are not covered by rain in both frames.

Convolutional neural networks were used by [21] to restore images affected by dirt and rain. They use a simple 3-layer architecture, each with 512 units, which works well on small drops but breaks down with much larger contaminants. A much larger Generative Adversarial Network (GAN) model [27] is used by [11], along with attention [28]. They leverage their static dataset to provide a ground truth for the droplet attention mask and train a recurrent model that outputs a heatmap of the location of the droplets. This heatmap is then concatenated with the input image and run through the GAN. They produce state-of-the-art results and made their dataset publicly available, which has allowed us to directly compare our method with theirs.

III. LEARNING TO CLEAN IMAGES

A. Computer-Generated Synthetic Rain

We base our simple synthetic droplet model on the works of [17], [18] and [19], generate the locations of raindrops using a simple statistical approach, model the interactions between raindrops using metaballs [20] and implement its rendering efficiently using GPU shaders.

A proto-raindrop is created using a simple refractive model that assumes a pinhole camera. The refraction angle is encoded following a scheme similar to normal mapping [29] by using a 2D look-up table represented by the RED and GREEN channels of a texture T , with the thickness of the drop encoded in the BLUE channel of the same texture. This texture T is then masked using an alpha layer that allows blending of the water drops with the background image and other drops, as shown in Figure 3a. With the drop acting as a simple lens, the coordinate (x_r, y_r) of the world point that is rendered at the location (u, v) on the surface of a drop is given by the following simplified distortion model:

$$x_r = u + (R * B) \quad (1)$$

$$y_r = v + (G * B). \quad (2)$$

Each image location (u, v) has a probability P_r of becoming the center of a proto-raindrop whose dimensions are scaled along the horizontal and vertical directions by a tuple of random values S_x and S_y . For each timestep, the center of a droplet may undergo a slip of D_x pixels along the horizontal and D_y pixels along the vertical direction as a function of the droplet diameter d :

$$D_x, D_y = \begin{cases} 0, 0 & d \leq 4mm \\ x \sim \mathcal{N}(0, 3), P_d * 5 & d > 4mm, \end{cases}$$

where P_d represents the probability of slip along the vertical direction and x denotes the random deviation of the slip along the horizontal direction. We empirically choose a maximum of 5 pixels of vertical displacement.

For each timestep, droplets that are close to each other are merged using the metaballs approach [20], as shown in Figure 3b. By default, each texture location $T(u, v)$ that does not fall under a droplet encodes a normal that is perpendicular to the background image. Finally, the image is sampled using the normal map defined by the texture T to produce a result similar to the one in the top-left corner of Fig 1. Using this technique we have created three synthetic rain datasets:

- synthetic rain added to CamVid, complete with road marking ground truth;
- synthetic rain added to Cityscapes, complete with semantic segmentation ground truth; and
- synthetic rain added to the dry images from our stereo dataset, complete with road marking ground truth.



Fig. 3. Metaballs.

B. The de-raining network

The de-raining network architecture is based on Pix2PixHD [30]. The architecture is shown in Fig. 2. We employ 4 down-convolutional layers with stride 2, followed by 9 ResNet [31] blocks and 4 up-convolutional layers. We

motivate the addition of skip connections by observing that most of the structure of the input image should be kept, along with illumination levels and fine details.

To promote better generalization and inpainting, we refrain from using any direct pixel-wise loss and instead use a combination of adversarial, perceptual, and multi-scale discriminator feature losses. The discriminator architecture is a CNN with 5 layers, similar to PatchGAN [32]. We present the full structure of the losses in the next section.

C. Losses

Similar to [33], we apply an adversarial loss through a discriminator on the output of the generator. This loss is formulated as:

$$\mathcal{L}_{adv} = (D(G(I_{rainy})) - 1)^2. \quad (3)$$

The discriminator is trained to minimize the following loss:

$$\mathcal{L}_{disc} = (D(I_{clear}) - 1)^2 + (D(I_{de-rained}))^2, \quad (4)$$

where $I_{de-rained}$ is sampled from a pool of previously de-rained images.

The perceptual loss [34] is applied between the label and reconstructed image:

$$\mathcal{L}_{perc} = \sum_{i=1}^{n_{VGG}} \frac{1}{w_i^{perc}} \|VGG(I_{clear})_i - VGG(G(I_{rainy}))_i\|_1, \quad (5)$$

where n_{VGG} represents the number of VGG layers that are used to compute the loss and $w_i^{perc} = 2^{(n_{VGG}-i)}$ weighs the importance of each layer.

Additionally, a multi-scale discriminator feature loss [30] is applied between the label and reconstructed image:

$$\mathcal{L}_{msadv} = \sum_{i=1}^{n_{ADV}} \frac{1}{w_i^{adv}} \|D(I_{clear})_i - D(G(I_{rainy}))_i\|_1, \quad (6)$$

where n_{ADV} represents the number of discriminator layers that are used to compute the loss and $w_i^{adv} = 2^{(n_{ADV}-i)}$ weighs the importance of each layer.

The complete generator objective \mathcal{L}_{gen} becomes:

$$\mathcal{L}_{gen} = \lambda_{adv} * \mathcal{L}_{adv} + \lambda_{perc} * \mathcal{L}_{perc} + \lambda_{msadv} * \mathcal{L}_{msadv}. \quad (7)$$

Each λ term is a hyperparameter that weights the importance of each term of the loss equation. We wish to estimate the generator G and discriminator D functions such that:

$$G, D = \arg \min_{G, D} \mathcal{L}_{gen} + \mathcal{L}_{disc}. \quad (8)$$

In the following section we describe how the network is trained to minimise the above losses.

IV. EXPERIMENTAL SETUP

A. Stereo rain dataset

In this section we present the hardware used to record our narrow-baseline stereo dataset that allows one lens to be affected by real water droplets while keeping the other lens clear. The camera setup is shown in Figure 8. A 3D-printed bi-partite chamber is sandwiched between two acrylic clear panels and placed in front of the two lenses, with the left-hand section of the chamber being kept dry at all times, while



Fig. 4. CamVid road marking segmentation results. From left to right: rainy input image, segmentation result on rainy image, derained input image, segmentation result on derained image.

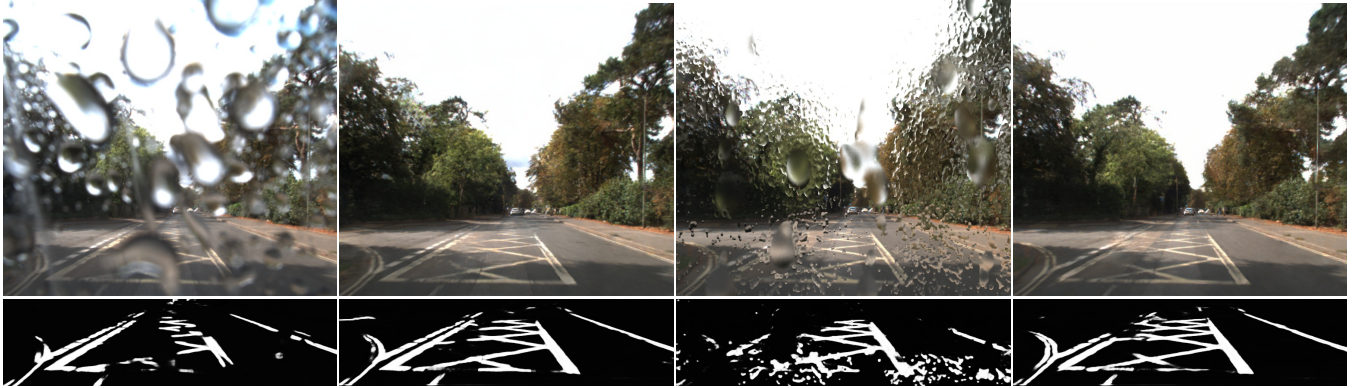


Fig. 5. RobotCar road marking segmentation results. First column shows a RobotCar(R) real rain image and segmentation result. Second column shows the derained real rain image and segmentation result. Third column shows a RobotCar(S) computer-generated rain image and segmentation result. Fourth column shows the derained computer-generated rain image and segmentation result.

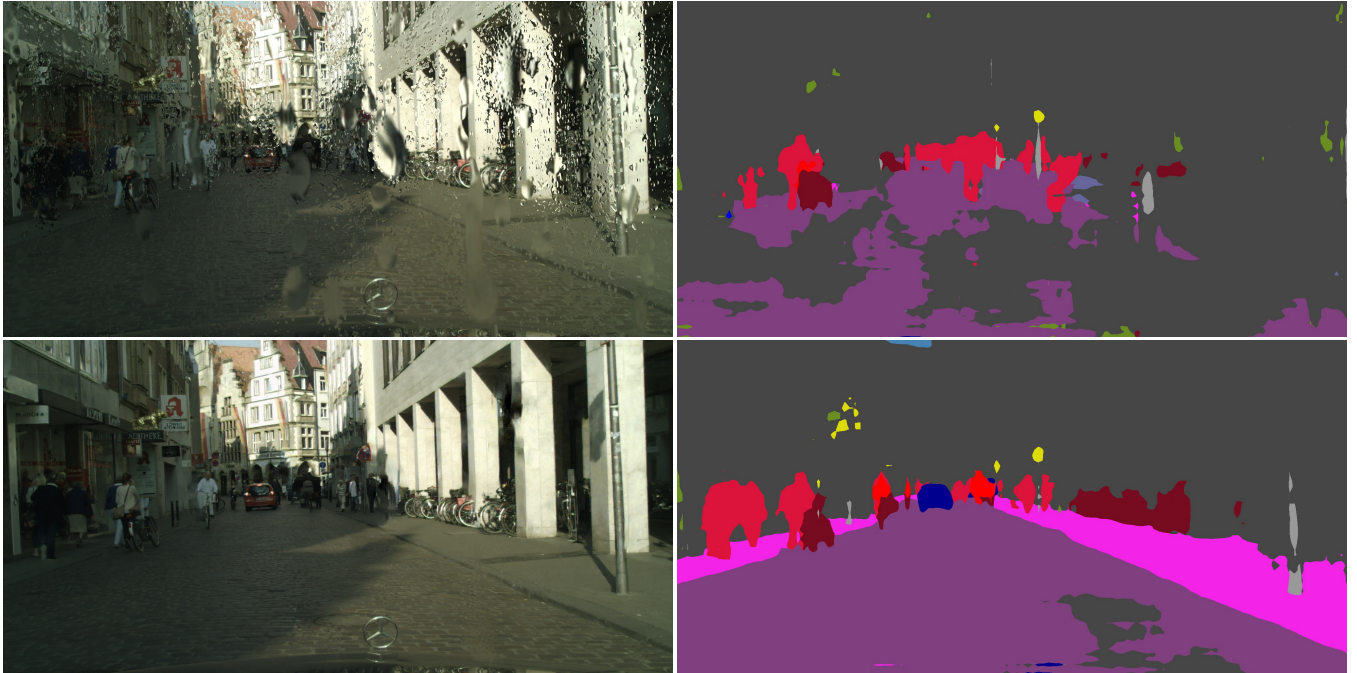


Fig. 6. Cityscapes semantic segmentation results. The first row shows a rainy image on the left and its corresponding semantic segmentation on the right. The second row shows the derained image on the left and its corresponding semantic segmentation on the right.



Fig. 7. An example from our stereo dataset. The image on the left is produced by the left lens, which is affected by water drops. The image in the middle is produced by the dry right hand lens. The image on the right is the road marking segmentation ground truth.

the right-hand section is sprayed with water droplets using an internal nozzle fitted at the top of the chamber. The angle of this chamber with respect to the axes of the cameras can be modified to simulate a slanted windscreen or enclosure, and the distance from the lenses can be increased or decreased accordingly to replicate different levels of focus or blur on the droplets.

The nozzle spans the entire width of the right chamber and is capable of producing water droplets with a diameter between 1mm and 8mm, as well as streaks of water. This variability is achieved by modulating the water pressure using a number of pulse width modulation regimes. The water is drained from the bottom of the chamber and is returned to a storage tank for recirculation. The cameras used are Point Grey Grasshopper 2 with 4.5 mm F/1.4 lenses, a baseline of 29mm and automatic synchronisation. The system is fully portable and the water is completely contained within the circuit formed by the right chamber, pump and tank.

We have collected approximately 50000 pairs of images by driving in and around the city of Oxford. The image pairs are undistorted, cropped and aligned. We have selected 4818 image pairs to form a training, validation and testing dataset. From the testing partition, we have created ground truth road marking segmentations for 500 images. An example from our dataset is shown in Figure 7.

Compared to the painstakingly-collected dataset of [11], our setup is a set-and-forget approach: once the stereo camera has been mounted on a vehicle, it is trivial to collect large amounts of well-synchronised and well-aligned pairs of images.

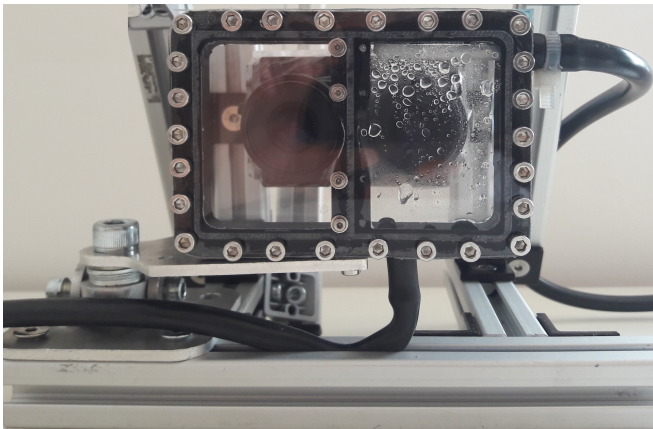


Fig. 8. Our small-baseline stereo camera setup. A bi-partite chamber with acrylic clear panels is placed in front of the lenses, with the left-hand section being kept dry at all times, while the right-hand section is sprayed with water droplets using an internal nozzle.

B. Training

We used a network training regimen similar to [30]. For each iteration we first trained the discriminator on a clear image and a de-rained image from a previous iteration with the goal of minimizing $\mathcal{L}_{\text{disc}}$, and then trained the generator on rainy input images to minimize \mathcal{L}_{gen} . We used the Adam solver [35] with an initial learning rate set at 0.0002, a batch size of 1, $\lambda_{\text{adv}} = 1$, $\lambda_{\text{perc}} = 1$ and $\lambda_{\text{msadv}} = 1$.

C. Segmentation Tasks

We used the trained generator G to de-rain all of the rainy input images. To benchmark both the images with computer-generated water drops and the images with real water drops, in the context of road marking segmentation, we used the approach of [4] which trains a U-Net to segment road markings in a binary way. To benchmark the computer-generated water drop images in the context of semantic segmentation, we used DeepLab v3 [36] which has achieved state-of-the-art performance on the Cityscapes dataset.

The generator runs at approximately 1 Hz for images with a resolution of 1280×960 , and at approximately 3 Hz for images with a resolution of 640×480 on an Nvidia Titan X GPU.

V. RESULTS

We benchmark our results taking into consideration several metrics across several tasks, and also present results on the quality of the image reconstruction.

A. Quantitative results

Table I presents results for road marking segmentation, in the case of RobotCar with real water drops (R), RobotCar with computer-generated water drops (S) and CamVid with computer-generated water drops (S). Our baseline is represented by the performance of clear images tested on models that were trained using clear images (REFERENCE). For both RobotCar (R), Robotcar (S), and the CamVid (S) datasets, the results show a severely degraded performance when testing rainy images on models that were trained using clear images (RAINY). Retraining the road marking segmentation models with a dataset augmented with rainy images will lead to an improvement in performance (AUGM). However, de-raining the images using our method and testing them on a model trained using clear images (DERAINED) restores the performance of the segmentation to levels that are close to the baseline recorded on clear images. Figure 4 shows road marking segmentation results on CamVid, before and after deraining. Figure 5 shows road marking segmentation results on RobotCar(R)&(S), before and after deraining.

As expected, re-training the segmentation model with a dataset that is augmented with rainy images helps to improve performance, however using a specialised de-raining preprocessing step significantly outperforms this approach, even when tested on a model trained exclusively with clear images. This is the expected advantage of having a model dedicated, in its entirety, to a specific image-to-image mapping task (de-raining), which narrows the variety of images fed to the segmentation task.

Table II presents results for semantic segmentation on the Cityscapes dataset. We benchmark 4 different combinations of models and datasets:

- Cityscapes-clear images tested on a model trained using Cityscapes-clear images;
- Cityscapes-rainy images tested on a model trained using Cityscapes-clear images;
- Cityscapes-rainy images tested on a model trained using Cityscapes-clear and Cityscapes-rainy images; and

TABLE I
ROAD MARKING SEGMENTATION RESULTS

Dataset	REFERENCE(CLEAR)				RAINY				AUGM.				DERAINED			
	Prec.	Rec.	F1	IOU	Prec.	Rec.	F1	IOU	Prec.	Rec.	F1	IOU	Prec.	Rec.	F1	IOU
RobotCar(R)	0.627	0.918	0.734	0.594	0.512	0.628	0.550	0.396	0.486	0.807	0.593	0.434	0.603	0.841	0.689	0.544
RobotCar(S)	0.627	0.918	0.734	0.594	0.364	0.595	0.437	0.287	0.654	0.770	0.690	0.541	0.661	0.816	0.715	0.569
CamVid(S)	0.576	0.927	0.699	0.551	0.353	0.576	0.425	0.279	0.457	0.771	0.563	0.405	0.520	0.755	0.603	0.444

TABLE II
CITYSCAPES SEMANTIC SEGMENTATION RESULTS

Cityscapes Img. vs. Segm. Model	mIOU
CLEAR on CLEAR	0.692
RAINY on CLEAR	0.405
RAINY on AUGMENTED	0.611
DERAINED on CLEAR	0.651

TABLE III
RECONSTRUCTION RESULTS

Dataset	RAW		DERAINED	
	PSNR	SSIM	PSNR	SSIM
RobotCar-Rainy(R)	13.02	0.5574	22.82	0.8188
RobotCar-Rainy(S)	16.80	0.6134	25.17	0.8699
CamVid-Rainy(S)	16.89	0.6064	22.11	0.7524
Qian et al.[11](R)	24.09	0.8518	31.55	0.9020

- Cityscapes-derained(Cityscapes-rainy preprocessed using our deraining model) images tested on a model trained using Cityscapes-clear images.

Similar to the case of road marking segmentation, we notice the same severe degradation of performance when testing with rainy images (RAINY on CLEAR) as compared to the baseline (CLEAR on CLEAR). Again, the performance of derained images tested on a model trained using clear images (DERAINED on CLEAR) is significantly better than the performance of rainy images tested on a model trained using a dataset augmented with rainy images (RAINY on AUGMENTED). Figure 6 shows semantic segmentation results on Cityscapes, before and after deraining.

B. Reconstruction results

Table III presents results on the quality of the image reconstruction using two widely used image-quality metrics, Peak signal-to-noise ratio (PSNR), and Structural similarity (SSIM). We benchmark our model on our real-world RobotCar-Rainy (R) dataset, RobotCar-Rainy with computer-generated rain (S), CamVid-Rainy with computer-generated rain (S), and on the dataset provided by [11]. The RAW column shows the quality of the rainy images, while the DERAINED column shows the quality of the de-rained images, all relative to their clear ground truth. We show that in all cases, de-raining the rain-affected images using our preprocessor significantly increases the quality of the images, as compared to the reference case where raw rainy images are used. Both the real-world rainy dataset images and the images with computer-generated rain are significantly more degraded than the rainy images provided by [11], as seen in column RAW.

Table IV presents reconstruction results on the reference rainy dataset provided by [11]. We show that we achieve state-of-the-art PSNR reconstruction results on images affected by real water drops and only slightly lower SSIM,

TABLE IV
RECONSTRUCTION QUALITY COMPARISON TO STATE OF THE ART

Model vs. Dataset	Dataset from [11]	
	PSNR	SSIM
Original	24.09	0.8518
Eigen13[21]	28.59	0.6726
Pix2Pix[37]	30.14	0.8299
Qian et al.(no att.)[11]	30.88	0.8670
Qian et al.(full att.)[11]	31.51	0.9213
Ours(no att.)	31.55	0.9020

while, in contrast to [11], not requiring an attention [28] mechanism, which simplifies and speeds up inference and training.

VI. CONCLUSIONS

We have presented a system that restores performance of images affected by adherent raindrops on important segmentation tasks. Our results show that road marking segmentation, an important task for autonomous driving systems, is severely affected by adherent rain and that performance can be restored by first running the images through a de-raining preprocessor. Similarly, we show the same reduction and restoration of performance in the case of semantic segmentation, a task that is important in many fields. Additionally, we produce state-of-the-art results in terms of the quality of image restoration, while being able to run in real time. Finally, our system processes the image streams outside of the segmentation pipeline, either offline or online, and hence can be used naturally as a front end to many existing systems. The dataset will be made available at <https://ciumonk.github.io/RobotCar-rainy/>, along with a video describing our results at <https://ciumonk.github.io/RobotCar-rainy/video.html>.

VII. FUTURE WORK

Future work may involve designing a mechanism for producing computer-generated rain that is indistinguishable from real rain in terms of its usefulness in training models that quantitatively rather than qualitatively improve performance on image-based tasks.

VIII. ACKNOWLEDGEMENTS

This work was supported by Oxford-Google DeepMind Graduate Scholarships and Programme Grant EP/M019918/1. The authors wish to thank Valentina Musat for labelling the road markings in our dataset.

REFERENCES

- [1] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3213–3223.
- [2] E. Stenborg, C. Toft, and L. Hammarstrand, "Long-term visual localization using semantically segmented images," *CoRR*, vol. abs/1801.05269, 2018.

- [3] J. L. Schönberger, M. Pollefeys, A. Geiger, and T. Sattler, "Semantic visual localization," *CoRR*, vol. abs/1712.05773, 2017.
- [4] T. Bruls, W. Maddern, A. A. Morye, and P. Newman, "Mark yourself: Road marking segmentation via weakly-supervised annotations from multimodal data," in *Robotics and Automation (ICRA), 2018 IEEE International Conference on*. IEEE, 2018, p. in press.
- [5] A. Valada, J. Vertens, A. Dhall, and W. Burgard, "Adapnet: Adaptive semantic segmentation in adverse environmental conditions," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 4644–4651.
- [6] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. V. Gool, "Domain adaptive faster R-CNN for object detection in the wild," *CoRR*, vol. abs/1803.03243, 2018.
- [7] M. Wulfmeier, A. Bewley, and I. Posner, "Addressing appearance change in outdoor robotics with adversarial domain adaptation," *CoRR*, vol. abs/1703.01461, 2017.
- [8] G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. M. Lopez, "The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [9] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 Year, 1000km: The Oxford RobotCar Dataset," *The International Journal of Robotics Research (IJRR)*, vol. 36, no. 1, pp. 3–15, 2017.
- [10] G. J. Brostow, J. Fauqueur, and R. Cipolla, "Semantic object classes in video: A high-definition ground truth database," *Pattern Recognition Letters*, vol. 30, no. 2, pp. 88–97, 2009.
- [11] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," *CoRR*, vol. abs/1711.10098, 2017.
- [12] J. Chen and L. Chau, "A rain pixel recovery algorithm for videos with highly dynamic scenes," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1097–1104, March 2014.
- [13] J. Kim, J. Sim, and C. Kim, "Stereo video deraining and desnowing based on spatiotemporal frame warping," in *2014 IEEE International Conference on Image Processing (ICIP)*, Oct 2014, pp. 5432–5436.
- [14] —, "Video deraining and desnowing using temporal correlation and low-rank matrix completion," *IEEE Transactions on Image Processing*, vol. 24, no. 9, pp. 2658–2670, Sept 2015.
- [15] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *European conference on computer vision*. Springer, 2016, pp. 154–169.
- [16] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley, "Clearing the skies: A deep network architecture for single-image rain removal," *CoRR*, vol. abs/1609.02087, 2016.
- [17] M. Roser and A. Geiger, "Video-based raindrop detection for improved image registration," in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, Sept 2009, pp. 570–577.
- [18] M. Roser, J. Kurz, and A. Geiger, "Realistic modeling of water droplets for monocular adherent raindrop recognition using bezier curves," in *Computer Vision – ACCV 2010 Workshops*, R. Koch and F. Huang, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 235–244.
- [19] S. You, R. T. Tan, R. Kawakami, Y. Mukaigawa, and K. Ikeuchi, "Waterdrop stereo," *CoRR*, vol. abs/1604.00730, 2016.
- [20] J. F. Blinn, "A generalization of algebraic surface drawing," *ACM Trans. Graph.*, vol. 1, no. 3, pp. 235–256, July 1982.
- [21] D. Eigen, D. Krishnan, and R. Fergus, "Restoring an image taken through a window covered with dirt or rain," in *2013 IEEE International Conference on Computer Vision*, Dec 2013, pp. 633–640.
- [22] S. You, R. T. Tan, R. Kawakami, Y. Mukaigawa, and K. Ikeuchi, "Adherent raindrop modeling, detection and removal in video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 9, pp. 1721–1733, Sept 2016.
- [23] A. Yamashita, M. Kuramoto, T. Kaneko, and K. T. Miura, "A virtual wiper - restoration of deteriorated images by using multiple cameras," in *IROS*. IEEE, 2003, pp. 3126–3131.
- [24] A. Yamashita, T. Kaneko, and K. T. Miura, "A virtual wiper-restoration of deteriorated images by using a pan-tilt camera," in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, vol. 5, April 2004, pp. 4724–4729 Vol.5.
- [25] A. Yamashita, Y. Tanaka, and T. Kaneko, "Removal of adherent waterdrops from images acquired with stereo camera," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Aug 2005, pp. 400–405.
- [26] M. Kuramoto, A. Yamashita, T. Kaneko, and K. T. Miura, "Removal of adherent waterdrops in images by using multiple cameras," in *MVA*, 2002, pp. 80–83.
- [27] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, ser. NIPS'14. Cambridge, MA, USA: MIT Press, 2014, pp. 2672–2680.
- [28] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," *CoRR*, vol. abs/1406.6247, 2014.
- [29] J. Cohen, M. Olano, and D. Manocha, "Appearance-preserving simplification," in *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '98. New York, NY, USA: ACM, 1998, pp. 115–122.
- [30] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on*. IEEE, 2018, pp. 1–13.
- [31] 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*, 2016, pp. 770–778.
- [32] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *European Conference on Computer Vision*. Springer, 2016, pp. 702–716.
- [33] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.
- [34] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*. Springer, 2016, pp. 694–711.
- [35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [36] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *ECCV*, 2018.
- [37] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *CoRR*, vol. abs/1611.07004, 2016.