Learning to Remove Rain in Traffic Surveillance by using Synthetic Data

Chris H. Bahnsen[1], David Vázquez[2], Antonio M. López[3] and Thomas B. Moeslund[1]
[1]*Visual Analysis of People Laboratory, Aalborg University, Denmark*
[2]*Element AI, Spain*
[3]*Computer Vision Center, Universitat Autònoma de Barcelona, Spain*

Keywords: Rain Removal, Traffic Surveillance, Image Denoising.

Abstract: Rainfall is a problem in automated traffic surveillance. Rain streaks occlude the road users and degrade the overall visibility which in turn decrease object detection performance. One way of alleviating this is by artificially removing the rain from the images. This requires knowledge of corresponding rainy and rain-free images. Such images are often produced by overlaying synthetic rain on top of rain-free images. However, this method fails to incorporate the fact that rain fall in the entire three-dimensional volume of the scene. To overcome this, we introduce training data from the SYNTHIA virtual world that models rain streaks in the entirety of a scene. We train a conditional Generative Adversarial Network for rain removal and apply it on traffic surveillance images from SYNTHIA and the AAU RainSnow datasets. To measure the applicability of the rain-removed images in a traffic surveillance context, we run the YOLOv2 object detection algorithm on the original and rain-removed frames. The results on SYNTHIA show an 8% increase in detection accuracy compared to the original rain image. Interestingly, we find that high PSNR or SSIM scores do not imply good object detection performance.

1 INTRODUCTION

In computer vision-enabled traffic surveillance, one would hope for optimal conditions such as high visibility, few reflections, and good lighting conditions. This might be the case under daylight and overcast weather but is hardly representative of most real-life weather conditions. To name an example, the visibility of a scene might be impaired by the occurrence of precipitation such as rainfall and snowfall. The rain and snowfall are present in the images and videos as spatio-temporal streaks that might occlude foreground objects of interest. The accumulation of rain and snow streaks ultimately degrades the visibility of a scene (Shettle, 1990) which render faraway objects hard to distinguish from the background. These rain and snow streaks may even adhere to the camera lens as quasi-static rain drops that remain for several seconds, effectively blurring a region of the image. The above-mentioned properties of rain and snowfall have a detrimental effect on computer vision algorithms and the research community has therefore shown great interest to mitigate these effects. Since the first work by Hase et al. (1999), many subsequent authors have proposed algorithms with the aim of producing a realistic rain-removed image from a real-world rainy image.

When constructing an algorithm that artificially removes rain in an image or video, one would typically optimize for creating rain-removed images that resemble real-world images as much as possible. Typically, this is assessed by computing the Peak Signal-to-Noise-Ratio (PSNR) and the Structural Similarity Index (SSIM) (Wang et al., 2004) between the rain-removed image and the ground truth rain-free image. A high PSNR or SSIM score indicates that the source and target images are largely similar. The computation of these metrics, however, requires corresponding image pairs of rainy and rain-free images. For single-image rain removal, this requirement is usually met by overlaying artificial rain streaks on real-world images, typically by generating them in Adobe Photoshop or by using a collection of pre-rendered rain streaks (Garg and Nayar, 2006). A sample image is visible from the left part of Figure 1.

Although the individual rain streaks may look realistic, the visual impression of the artificially produced rain image is less pleasing. Because the generated rain streaks are layered on top of the rain-free image, all rain streaks appear to be in the immedi-

Bahnsen, C., Vázquez, D., López, A. and Moeslund, T.
Learning to Remove Rain in Traffic Surveillance by using Synthetic Data.
DOI: 10.5220/0007361301230130
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 123-130
ISBN: 978-989-758-354-4

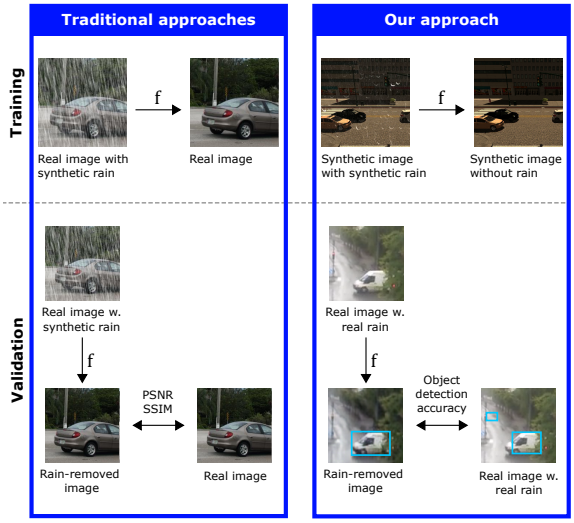


Figure 1: Proposed system-at-a-glance. As opposed to traditional methods, we use fully synthetic data for training rain removal algorithms. We validate on real-world images with real rain for which ground truth rain-free images do not exist, and we are thus unable to use traditional metrics such as SSIM or PSNR. Instead, we measure the accuracy of an object detection algorithm on both the original and rain-removed images. An effective rain removal algorithm should improve the visibility of foreground objects and thus increase the object detection accuracy.

ate foreground of the image. However, this fails to take into account that rain may fall in the entire three-dimensional volume of the scene and does not model the visibility degradation caused by the accumulation of rain drops.

The aim of this work is to create a single-image based rain removal algorithm that takes both the partial occlusions and the accumulation of rain into account. We accomplish this by introducing a new training dataset that consists of images from a purely synthetic, 3D-generated world. By using a computer-generated 3D world, we can simulate raindrops in the entirety of the scene and not just in front of the camera. This enables us to mimic the rain streaks, the accumulation of rain, and the adhering of rain drops to the virtual camera lens. The concept is illustrated in Figure 1.

Our contributions are the following:

1. To the best of our knowledge, we are the first to introduce fully synthetic training data for training and testing single-image based rain removal algorithms.
2. We train a rain removal algorithm using the data from above and compare with traditional approaches that use synthetic rain on top of real-world images. In order to assess the performance on real-world traffic surveillance images with real rain, we propose a new evaluation metric that assesses the performance of an object detection algorithm on the original and rain-removed frames. If effective, the rain-removed images should improve object detection performance.
3. The proposed evaluation metric is compared with the traditional PSNR and SSIM metrics to evaluate their usefulness in application-based rain removal.

2 RELATED WORK

The first single-image based rain removal algorithm was proposed by Fu et al. (2011) and treated rain removal as a dictionary learning problem where the challenge is to decide if image patches belong to the rain component, R, or the background component, B. Relying on the assumption that rain drops are high-frequency (HF) oscillations occurring on top of a low-frequency (LF) background image, the bilateral filter is applied to the input image to separate it into a HF and a LF component. The Morphological Component Analysis technique (Fadili et al., 2010) learns a dictionary of image patches from the HF image and rain streak patches are identified based on the assumption that they are brighter than other patches. The dictionary composition approach to rain removal was refined in subsequent works (Chen et al., 2014; Huang et al., 2014; Kang et al., 2012; Wang et al., 2017b).

An alternative approach was proposed by Chen and Hsu (2013) that treats the separation of the rain image R from the background image B as a matrix decomposition problem. It is assumed that B has low total variation and that R patches are linearly dependent. Based on these assumptions, the Inexact Augmented Lagrange Multiplier is used to solve the constrained matrix decomposition problem. Subsequent works on matrix decomposition (Jiang et al., 2017; Luo et al., 2015) have imposed additional requirements on B and R such as low rank, sparsity, or mutual exclusivity.

The Achilles heel of the mentioned dictionary component and matrix decomposition methods is that they solely rely on heuristically defined statistical properties to detect and remove the apparent rain. Real-world textures might not adhere to these statistical properties, however, and as a result, non-rain textures might be 'trapped' inside the rain component.

This problem is overcome by learning the appearance of rain streaks in an offline process that uses a collection of rain-free images overlaid with synthetic rain. Recent approaches use such images to train convolutional neural networks (CNNs) to remove rain from single images.

Fu et al. (2017a) combined a guided filter with a three-layer CNN to produce rain-free images. In the work to follow, the same author replaced the three-layer CNN with a much larger network containing residual connections (Fu et al., 2017b).

A CNN containing dilated convolutions were used in (Yang et al., 2017) whereas Liu et al. (2017) used a network based on the Inception V4 architecture (Szegedy et al., 2017). As opposed to the works of Fu et al. both methods operate directly on the input image and may as such capture rain drops that are not included in the filtered HF image.

Conditional Generative Adversarial Networks. The recent advent of generative adversarial networks (GANs) that are conditioned on the input image have made major breakthroughs in image-to-image translation (Isola et al., 2017). A conditional GAN may be used to transfer an image in a specific domain to a corresponding image in another domain, for instance from rainy to rain-free images. Zhang et al. (2017a) modified the Pix2Pix framework by Isola et al. (2017) by including the perceptual loss function by Johnson et al. (2016) and trained the conditional GAN on corresponding image pairs with and without synthetic rain.

3 RAIN REMOVAL USING ENTIRELY SYNTHETIC DATA

In this section, we will describe our proposed rain removal framework. Like most other authors of rain removal algorithms, we want our network to be able to remove rain from real-world images with real rain, including the effects of rain streak occlusion and rain streak accumulation. The occluding effects of rain streaks might be modelled by imposing synthetic rain on real-world images but this approach cannot capture the effects arising from the accumulation of rain. In order to capture these effects, we propose to use fully synthetic training data generated from a computer-generated 3D world.

More specifically, we use renderings from the SYNTHIA virtual world (Ros et al., 2016) that capture four different road intersections as seen from an infrastructure-side traffic camera. The virtual world enables us to render two instances of the same sequence with the only difference that rain is falling in one instance but not in the other. Samples from the four sequences are shown in Figure 2. In total, the four sequences comprise of 9572 frames.

A benefit of the SYNTHIA virtual world is that it enables the generation of corresponding segmented images that may be used directly as ground truth for semantic segmentation and object detection purposes. Footage from SYNTHIA has previously been used to successfully transfer images from summer to winter (Hoffman et al., 2017) or to transfer from SYNTHIA to the real-world Cityscapes dataset (Zhang et al., 2017b). Based on these works, we therefore find it reasonable to learn the translation from rain images to no-rain images with the use of SYNTHIA.

Inspired by the recent work in image-to-image translation and domain adaption (Hoffman et al., 2017; Shrivastava et al., 2017), we use the conditional GAN architecture as the backbone of our rain removal framework.

3.1 Training the Conditional GAN-network

As point of departure, we take the rain removal algorithm from Zhang et al. (2017a), which consists of a conditional GAN-network, denoted as IDCGAN. We compare the IDCGAN with the state-of-the-art image-to-image translation framework Pix2PixHD (Wang et al., 2017a).

The discriminator of the IDCGAN-network uses a five-layer convolutional structure similar to the original Pix2Pix-network (Isola et al., 2017) whereas the generator uses a fully convolutional network with skip connections, the U-net. The generator architecture is different from the Pix2Pix-network in two ways:

1. The depth of the U-net is down from eight to six convolutional layers.
2. The skip-connections are adding the tensors instead of concatenating (joining) them.

The Pix2PixHD network is an improved version of Pix2Pix that enables the generation of more realistic, high-resolution images.

As training set, we use the aforementioned SYNTHIA dataset with 9572 corresponding image pairs. As representative of a dataset with real images and synthetic rain, the 700 training images from Zhang et al. (2017a) are used. The IDCGAN and Pix2PixHD networks are trained separately with the images of Zhang et al. and the SYNTHIA training images. Furthermore, we use a combination of the two datasets to train the Pix2PixHD network. An overview of the resulting five trained networks is found in Table 1. In order to make the training feasible on a 11 GB GPU, the training images are scaled down to a maximum resolution of 720 x 480 pixels. Otherwise, we use the default parameter settings for training the networks.

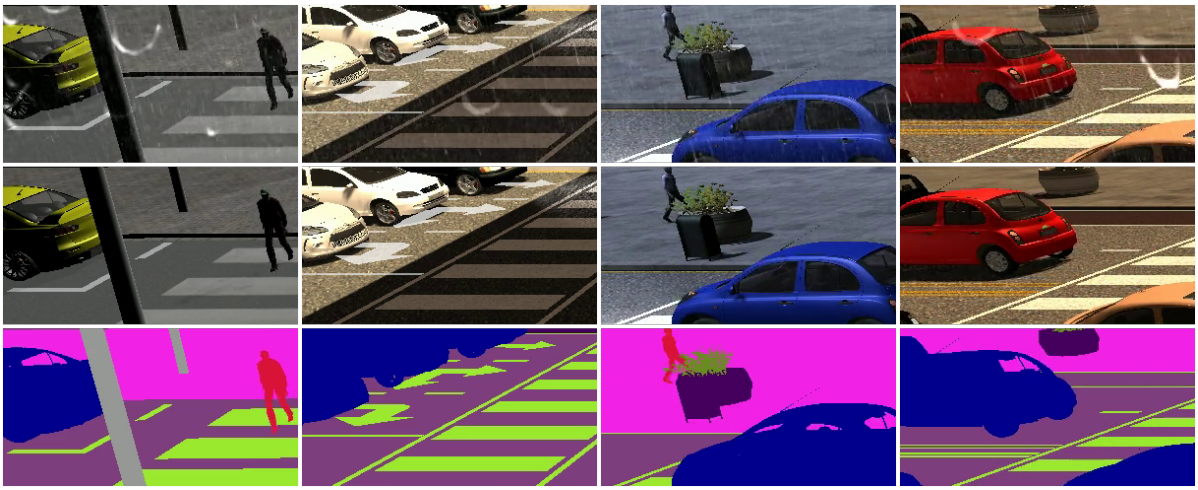


Figure 2: Synthetic images generated from SYNTHIA at four different locations in the virtual world. From top to bottom: rain image, no-rain image, ground truth segmented image. The images are cropped for viewing.

Table 1: Overview of the trained conditional GANs for rain removal. The training of IDCGAN-Real-Syn is equivalent to the original work of Zhang et al. (2017a).

Name	Training data
IDCGAN-Real-Syn	Real images with synthetic rain
IDCGAN-Syn-Syn	SYNTHIA rain images, SYNTHIA no-rain images
Pix2PixHD-Real-Syn	Real images with synthetic rain
Pix2PixHD-Syn-Syn	SYNTHIA rain images, SYNTHIA no-rain images
Pix2PixHD-Combined	SYNTHIA rain images, SYNTHIA no-rain images + real images with synthetic rain

4 ASSESSING THE RAIN REMOVAL QUALITY

As mentioned in the introduction, the classical approach of measuring the quality of the rain-removed image is to apply a rain removal algorithm on a rain-free image with overlaid synthetic rain and calculate the PSNR and the SSIM between the resulting image and the corresponding rain free-image. In a traffic surveillance context, it appears that the overlaid synthetic rain hardly resembles real-world rain. As such, there is no guarantee that a rain removal algorithm receiving high PSNR and SSIM scores on synthetic rain will translate well to real-world rain in a traffic surveillance image.

We therefore propose a new evaluation metric that measures the ability of an object detection algorithm to detect objects in the original and the rain-removed frames. An effective rain removal algorithm should be able to create a rain-removed image that resembles a true rain-free image. This means that the occlusion and visibility degradation originating from the rain streaks should be largely eliminated, creating an image in which objects are easier to detect. Instead of requiring the overlay of synthetic rain on rain-free images, this metric requires the annotation of bounding boxes around objects of interest. We find such sequences in the AAU RainSnow dataset[1] that contains 2200 annotated frames in a traffic surveillance context, taken from seven different traffic intersections. The dataset features a variety of challenging conditions such as rain, snow, low light, and reflections. Details about the dataset is found in the paper by Bahnsen and Moeslund (2018).

As object detection benchmark, we choose the state-of-the-art You Only Look Once algorithm (YOLOv2) (Redmon and Farhadi, 2017). YOLOv2 is chosen due to good detection performance and superior speed which is especially important in real-time traffic surveillance. The improvement in detection performance is assessed by:

1. Running pre-trained YOLOv2 on the original, rainy images of the SYNTHIA dataset.
2. Removing rain with the networks listed in Table 1 and running pre-trained YOLOv2 on the rain-removed images.

[1]https://www.kaggle.com/aalborguniversity/aau-rainsnow/

3. Measuring the detection accuracy of 1) and 2) by using the COCO API (Lin et al., 2014) and calculating the relative difference.

We also measure the improvement in detection performance on the AAU RainSnow dataset by following the above steps, replacing SYNTHIA with AAU RainSnow.

Because we have access to the ground-truth rain-free images on SYNTHIA, it is possible to calculate the PSNR and SSIM ratios between the rain-free images and the rain-removed images. This enables us to assess the dependency between the widely used PSNR and SSIM scores and the relative improvement in detection accuracy. On AAU RainSnow, it is not possible to compute the dependency as ground truth rain-free images do not exist for real-life precipitation.

5 EXPERIMENTAL RESULTS

We have experimented with several hyper-parameter settings for YOLOv2 and found the best results by setting the detection threshold, hierarchical threshold, and the non-maximum suppression threshold to 0.1, 0.1, and 0.3, respectively. As detection metric, we use average precision (AP) over intersection-over-union (IOU) ratios from .5 to .95 with intervals of .05, denoted as AP[.5:.05:.95], and average precision at IOU=0.5, denoted as AP[.5].

5.1 Removing Rain From SYNTHIA Training Data

We start by measuring the ability to remove rain from the SYNTHIA data. This is a peculiar case as the Syn-Syn networks have seen the data during the training phase and we are thus unable to judge whether these networks generalize well. It does, however, give the opportunity to assess the feasibility of the rain removal algorithms in a best-case scenario and relate SSIM/PSNR scores and object detection performance, reported in Table 2. The relationship between PSNR scores and object detection performance is plotted in Figure 4.

The detection results of Table 2 show that only two rain removal algorithms, IDCGAN-Real-Syn and Pix2PixHD-Syn-Syn, improve detection performance compared to the original rain images, but neither of the two algorithms come close to the detection performance of the ground truth no-rain images. This is remarkable given the fact that Pix2PixHD-Syn-Syn has seen these images during training.

Interestingly, the SSIM and the PSNR scores of the two rain removal algorithms show little correspondence with the detection results. The IDCGAN-Real-Syn network is receiving the lowest SSIM score but shows good detection performance whereas the IDCGAN-Syn-Syn network is receiving the highest SSIM score but fails to consistently improve the detection results. If we relate to Figure 4, is is difficult to establish a consistent relationship between PSNR and detection results across the evaluated rain removal algorithms.

Example images from the SYNTHIA data are shown in Figure 3. The networks trained solely on the SYNTHIA data are able to remove the majority of rain from the image with IDCGAN-Syn-Syn leaving the best visual impression, whereas the Pix2PixHD-Syn-Syn network suffers from checkerboard artifacts in the reconstructed textures.

5.2 Removing Rain from AAU RainSnow Sequences

Sample images from running the rain removal algorithms on the AAU RainSnow dataset are shown in Figure 5 whereas detection results from running YOLOv2 on the rain-removed images are found in Table 3. The detection results show marginal improvements on the networks trained on real images with synthetic rain (Real-Syn and Combined), whereas networks trained on only synthetic data (Syn-Syn) deteriorate the detection results. If we look at the visual examples from Figure 5, the rain-removed images of IDCGAN-Syn-Syn have strange artifacts and do not seem to lie within the domain of visual images, whereas the images from the Pix2PixHD-Syn-Syn network appear to lie closer to the visual domain. The latter network even attempts to remove the rain drops from the lower image in Figure 5 and removes both the large rain streak and the reflections from the cars from the top image.

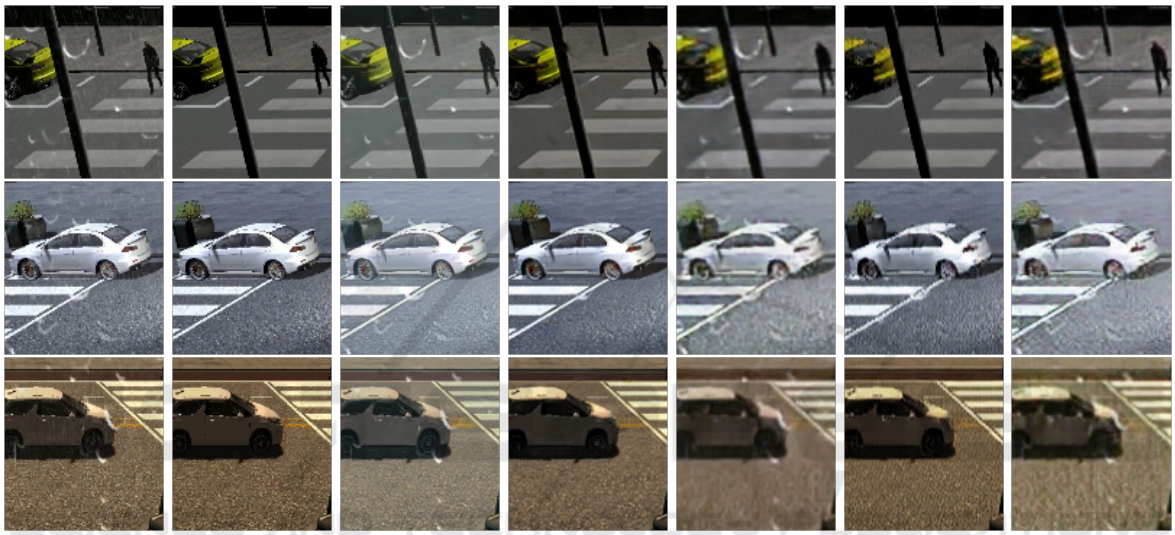
In general, the visual results also reveal plenty of room for improvement for all rain removal algorithms. As an example, the rain streaks on the top image and the rain drops on the lower image are not efficiently removed by any algorithm.

5.3 Domain Transfer Results

We find that the IDCGAN and Pix2PixHD networks behave inconsistently when tested on sequences that are dissimilar from their training data. On SYNTHIA, IDCGAN-Real-Syn improves the detection results, whereas Pix2PixHD-Real-Syn fails to do so, even if

Table 2: Results on the SYNTHIA dataset. Absolute values of detection performance are reported for the original rain image in italics, whereas other YOLOv2 results are relative to this baseline, shown in percentages. The original no-rain images are used as reference for computing SSIM and PSNR scores. Correlation denotes the Pearson product-moment correlation coefficient between the SSIM or PSNR values and the YOLOv2 detection performance (AP[.5:.05:.95]).

			YOLOv2		Correlation	
Rain removal method	SSIM	PSNR	AP[.5:.05:.95]	AP[.5]	SSIM/AP	PSNR/AP
Original rain image	-	-	*.025*	*.072*		
Original no-rain image	-	-	38.4	23.6		
IDCGAN-Real-Syn	.610	65.3	8.78	2.02	.306	.163
IDCGAN-Syn-Syn	.873	80.8	1.51	-7.15	-.118	-.011
Pix2PixHD-Real-Syn	.646	69.3	-32.0	-36.2	.069	.457
Pix2PixHD-Syn-Syn	.767	75.9	8.07	7.35	.329	.184
Pix2PixHD-Combined	.640	70.4	-32.7	-34.5	.385	.432



(a) Original rain image (b) Original no-rain image (c) IDCGAN-Real-Syn (d) IDCGAN-Synth-Syn (e) Pix2PixHD-Real-Syn (f) Pix2PixHD-Synth-Syn (g) Pix2PixHD-Combined

Figure 3: Rain-removal results on the SYNTHIA dataset. Each column represents the results of a rain removal algorithm on the original rain image.

Table 3: Detection results on the AAU RainSnow dataset. Absolute values of detection performance are reported for the original rain image in italics, whereas other YOLOv2 results are relative to this baseline, shown in percentages.

	YOLOv2	
Rain removal method	AP[.5:.05:.95]	AP[.5]
Original rain image	*.034*	*.070*
IDCGAN-Real-Syn	1.17	0.54
IDCGAN-Syn-Syn	-47.9	-42.9
Pix2PixHD-Real-Syn	-1.08	3.87
Pix2PixHD-Syn-Syn	-14.5	-5.05
Pix2PixHD-Combined	-2.43	3.19

providing a higher SSIM score. The results are reversed on AAU RainSnow sequences, with IDCGAN-Syn-Syn providing much worse detection results that Pix2PixHD-Syn-Syn. No obvious explanation of this behaviour exists and further experiments are needed in order to find and understand the most suitable network for domain transfer in rain removal.

6 CONCLUSIONS

We have investigated the use of fully synthetic data from the SYNTHIA virtual world to train a GAN-based, single-image rain removal algorithm. Using the fully synthetic data, we find that there is a considerable gap between detection scores on the rain-removed images from the best-performing rain removal algorithm and detection scores on the ground truth images with no rain. Furthermore, we found no consistent correlation between SSIM or PSNR scores and detection performance, questioning the usefulness of these metrics for application-based rain removal.

Removing rain on real-world traffic surveillance

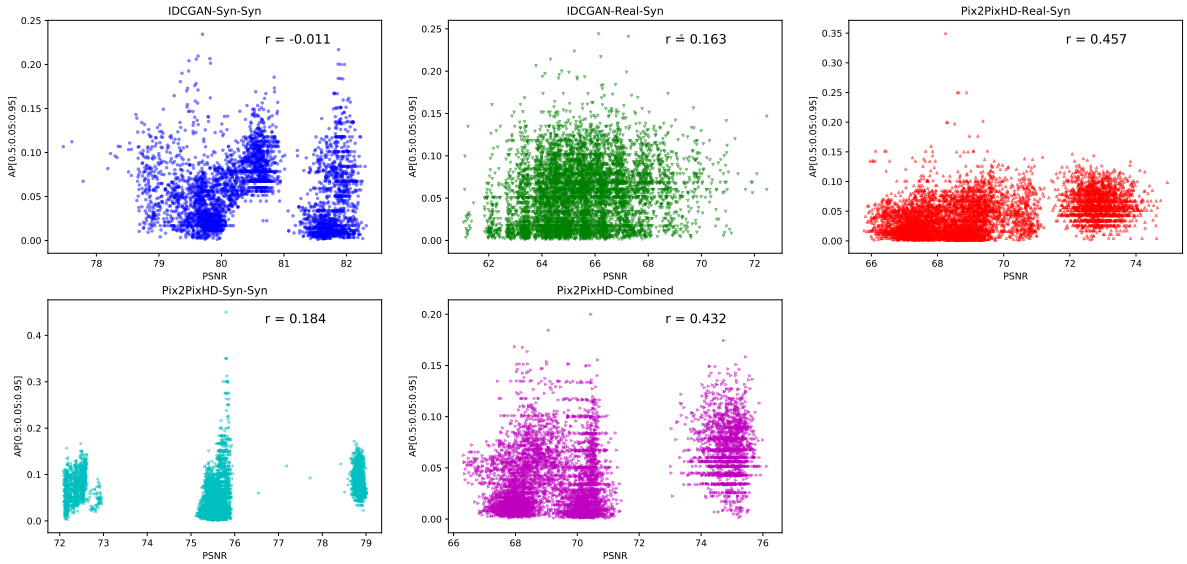


Figure 4: Relationship between the improvement in detection performance of the rain-removed images relative to the original rain images and the similarity score between the rain-removed images and the no-rain images. The r value indicates the Pearson product-moment correlation coefficient. Results are shown for PSNR but results for SSIM show the same pattern.

(a) Original rain image (b) IDCGAN-Real-Syn (c) IDCGAN-Synth-Syn (d) Pix2PixHD-Real-Syn (e) Pix2PixHD-Synth-Syn (f) Pix2PixHD-Combined

Figure 5: Rain-removal results on AAU RainSnow dataset. Each column represents the results of a rain removal algorithm on the original rain image.

imagery is hard and the evaluated rain-removal only results in marginal improvements in detection performance, if any. Using fully synthetic data for training allows the removal of some rain streaks that were not captured by networks trained with only synthetic rain. There exists, however, a domain gap between the synthetic data and the real-world sequences. Future work should address this by including more diverse synthetic data and more variability in real-world synthetic rain. One could also investigate the use of recurrent neural networks to incorporate temporal information from the SYNTHIA dataset. The influence of rain removal algorithms on detection performance in traffic surveillance has been studied by Bahnsen and Moeslund (2018) who observed a slight decrease in instance segmentation performance on the AAU RainSnow dataset. Future work could investigate if it might be beneficial to skip the rain removal step altogether and use the synthetic rain images to improve the training of object classification algorithms instead.

REFERENCES

Bahnsen, C. H. and Moeslund, T. B. (2018). Rain removal in traffic surveillance: Does it matter? *IEEE Trans on Intelligent Transportation Systems*, pages 1–18.

Chen, D.-Y., Chen, C.-C., and Kang, L.-W. (2014). Visual depth guided color image rain streaks removal using sparse coding. *IEEE Trans on Circuits and Systems for Video Technology*, 24(8):1430–1455.

Chen, Y.-L. and Hsu, C.-T. (2013). A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In *IEEE Int Conf on Computer Vision, Proc of the*, pages 1968–1975.

Fadili, M. J., Starck, J.-L., Bobin, J., and Moudden, Y. (2010). Image decomposition and separation using sparse representations: an overview. *Proc of the IEEE*, 98(6):983–994.

Fu, X., Huang, J., Ding, X., Liao, Y., and Paisley, J. (2017a). Clearing the skies: A deep network architecture for single-image rain removal. *Image Processing, IEEE Trans on*, 26(6):2944–2956.

Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., and Paisley, J. (2017b). Removing rain from single images via a deep detail network. In *Computer Vision and Pattern Recognition, IEEE Conf on*.

Fu, Y.-H., Kang, L.-W., Lin, C.-W., and Hsu, C.-T. (2011). Single-frame-based rain removal via image decomposition. In *Acoustics, Speech and Signal Processing, IEEE International Conf on*, pages 1453–1456. IEEE.

Garg, K. and Nayar, S. K. (2006). Photorealistic rendering of rain streaks. In *Graphics, ACM Trans on*, volume 25, pages 996–1002. ACM.

Hase, H., Miyake, K., and Yoneda, M. (1999). Real-time snowfall noise elimination. In *Image Processing, International Conf on*, volume 2, pages 406–409. IEEE.

Hoffman, J., Tzeng, E., Park, T., Zhu, J.-Y., Isola, P., Saenko, K., Efros, A. A., and Darrell, T. (2017). Cycada: Cycle-consistent adversarial domain adaptation. *arXiv preprint arXiv:1711.03213*.

Huang, D.-A., Kang, L.-W., Wang, Y.-C. F., and Lin, C.-W. (2014). Self-learning based image decomposition with applications to single image denoising. *Multimedia, IEEE Trans on*, 16(1):83–93.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *arXiv preprint*.

Jiang, T.-X., Huang, T.-Z., Zhao, X.-L., Deng, L.-J., and Wang, Y. (2017). A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors. In *Computer Vision and Pattern Recognition, IEEE Conf on*.

Johnson, J., Alahi, A., and Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In *European Conf on Computer Vision*, pages 694–711. Springer.

Kang, L.-W., Lin, C.-W., and Fu, Y.-H. (2012). Automatic single-image-based rain streaks removal via image decomposition. *Image Processing, IEEE Trans on*, 21(4):1742–1755.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European Conf on Computer Vision*, pages 740–755. Springer.

Liu, Y.-F., Jaw, D.-W., Huang, S.-C., and Hwang, J.-N. (2017). Desnownet: Context-aware deep network for snow removal. *arXiv preprint arXiv:1708.04512*.

Luo, Y., Xu, Y., and Ji, H. (2015). Removing rain from a single image via discriminative sparse coding. In *Computer Vision and Pattern Recognition, IEEE Conf on*, pages 3397–3405.

Redmon, J. and Farhadi, A. (2017). Yolo9000: better, faster, stronger. *arXiv preprint*.

Ros, G., Sellart, L., Materzynska, J., Vazquez, D., and Lopez, A. M. (2016). The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *Proceedings of the IEEE Conf on Computer Vision and Pattern Recognition*, pages 3234–3243.

Shettle, E. P. (1990). Models of aerosols, clouds, and precipitation for atmospheric propagation studies. In *In AGARD, Atmospheric Propagation in the UV, Visible, IR, and MM-Wave Region and Related Systems Aspects 14 p (SEE N90-21907 15-32)*.

Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W., and Webb, R. (2017). Learning from simulated and unsupervised images through adversarial training. In *CVPR*, volume 2, page 5.

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, pages 4278–4284.

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., and Catanzaro, B. (2017a). High-resolution image synthesis and semantic manipulation with conditional gans. *arXiv preprint arXiv:1711.11585*.

Wang, Y., Liu, S., Chen, C., and Zeng, B. (2017b). A hierarchical approach for rain or snow removing in a single color image. *Image Processing, IEEE Trans on*, 26(8):3936–3950.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE transa on*, 13(4):600–612.

Yang, W., Tan, R. T., Feng, J., Liu, J., Guo, Z., and Yan, S. (2017). Deep joint rain detection and removal from a single image. In *Computer Vision and Pattern Recognition, IEEE Conf on*.

Zhang, H., Sindagi, V., and Patel, V. M. (2017a). Image de-raining using a conditional generative adversarial network. *Computer Vision and Pattern Recognition, IEEE Conf on*.

Zhang, Y., David, P., and Gong, B. (2017b). Curriculum domain adaptation for semantic segmentation of urban scenes. In *The IEEE Int Conf on Computer Vision (ICCV)*, volume 2, page 6.