



Pixel-wise content attention learning for single-image deraining of autonomous vehicles

Yuande Jiang^{a,*}, Bing Zhu^b, Xiangmo Zhao^a, Weiwen Deng^c

^a Department of Information Engineering, Chang'an University, Xi'an 710064, China

^b Department of Automotive Engineering, Jilin University, Changchun 130022, China

^c Department of Transportation Science and Engineering, Beihang University, Beijing 100083, China



ARTICLE INFO

Keywords:

Single-image deraining
Content representation
Pixel-wise content attention
GAN
Closed field rain

ABSTRACT

Improving the performance of autonomous vehicles in adverse weather conditions is vital for the commercialization of such automated systems. Existing synthetic datasets for developing rain-tolerant vision are of limited value. To address this deficiency, a closed environment capable of simulating different degrees of rainfall is constructed. And a new *Closed Field Rain* dataset is collected in 36 testing cycles. Inspired by the idea that human can infer the content of rainy images directly without removing the raindrops. A new single-image deraining method is proposed, that does not require ground truth images. This method incorporates an image content estimation module applied to predict the scene content representation, and a pixel-wise content attention block used to evaluate the significance of each pixel. After that, an encoder-decoder network is applied to complete the image. On the other hand, it is almost impossible to obtain the ground truth of rainy images because of the dynamic characteristics of real traffic environment. Thus, the model is trained by employing PatchGAN, using a patch-based loss. Using common no-reference and feature point metrics as performance indicators, this paper conducts a comprehensive evaluation on both synthetic and real-world datasets including *Closed Field Rain* dataset. Results show the effectiveness of our model quantitatively and qualitatively.

1. Introduction

Autonomous vehicles have begun operation to demonstrate their capabilities in several cities, with some locales legalizing autonomous vehicles for use on public roads, as with the state of California and cities such as Amsterdam, Beijing and Berlin (Li et al., 2019; Scanlon et al., 2021). Reports from these early uses show the vehicles demonstrate excellent performance in normal traffic environments. However, autonomous vehicles rely heavily on sensors to perceive the surrounding environment, and performance degrades significantly in adverse weather such as rain, fog, and snow. The computational pipeline (localization, decision-making, path planning) that depends on sensor information has difficulty driving safety in such conditions, which is a critical hindrance to autonomous driving systems (Sharma et al., 2022). Rain is particularly common, so it is essential for autonomous vehicles to address the influence of rain.

Images captured by onboard cameras are often adversely affected by the rain streaks that degrade the sharpness of the images and reduce visibility. Degraded image quality further results in poor performance in

some functional modules that perform semantic segmentation, object detection, and object tracking. Deep neural networks have been widely applied in environment perception and achieve state-of-the-art results in some areas. There are two primary kinds of methods for improving the robustness of perception algorithms in rainy environment. The first method increases the data volume by adding rainy images, forming a mostly synthetic dataset (Ren et al., 2017; Cheng et al., 2018). Large amounts of data annotation are required, and the coupling between image preprocessing and object detection makes the perception task more complicated. The second method employs a network to improve the image quality by removing rain artifacts in a process called deraining. The pairs of clear and rainy images are usually necessary to optimize the neural network parameters within an end-to-end supervised deraining task. Although it is easy to obtain clear and rainy images in the real world, it is almost impossible to collect the ground truth (clear version) of rainy images because of the dynamic and stochastic characteristics of real traffic environments. Thus, end-to-end supervised learning is unsuitable when real-world data is used.

Deraining algorithms improve the image quality by removing rain

* Corresponding author.

E-mail addresses: yuande_jiang@163.com (Y. Jiang), zhubing@jlu.edu.cn (B. Zhu), xmzhao@chd.edu.cn (X. Zhao), kwdeng@188.com (W. Deng).

streaks. Existing methods are designed for two cases: video and single images. The lack of additional temporal information among different frames makes it more challenging to remove the rain layer from an individual image. To promote the development of environmental perception, several synthesized datasets have been created to train deep neural network models in an end-to-end manner. However, the reality gaps between synthetic and actual data cause performance degradation in practical applications. Further, the dataset should be as similar as possible to real data rather than built according to visual feeling. It is usually classified according to the rainfall intensity: light rain (less than 10 mm in 24 h), moderate rain (10–25 mm in 24 h), heavy rain (25–50 mm in 24 h), and torrential rain (50–100 mm in 24 h). Rainy images collected from different intensities are valuable in designing and verification of related algorithms. However, the randomness of weather and the dynamics of traffic agents make it difficult to obtain this kind of dataset. Closed testing fields designed for autonomous vehicle testing enable collecting large amounts of data with little background difference in different types of rainfall by establishing controllable rainfall simulation environments.

In this paper, we present our work on improving image quality in rainy conditions for autonomous vehicles. We establish a method of rainfall simulation and collect a large amount of data. We also propose an image deraining neural network using scene content representation of rainy images. The main contributions of this paper are as follows:

- A rainfall simulation scheme in the closed field is introduced. By adding targets and regulating the rainfall intensity, we create different traffic environments useful for autonomous vehicle testing and data collection. We collect the *Closed Field Rain* dataset from 36 predefined testing cycles.
- An image deraining algorithm is established using a generative adversarial network (GAN). Clear-rainy image pairs are no longer necessary using this adversarial optimization. We restore the clear image from a rainy one using the image content instead of modeling the rain characteristics as most current research does. By learning the image scene content significance using a pixel-wise attention module, the clear images can be obtained.
- The proposed algorithm is evaluated using both synthetic datasets and real datasets including our new *Closed Field Rain* dataset quantitatively and qualitatively. Because the real-world dataset lacks ground truth images, we verify performance using no-reference image quality and the image feature points metrics, as a complement to the commonly used reference image quality indicators.

The rest of this paper is organized as follows. Related research is presented in Section II. Section III introduces the rainfall simulation scheme and the design of testing cycles. Section IV presents the deraining neural network in detail. Section V presents our experimental results verifying the performance of our deraining method on both synthetic datasets and real-world datasets. Finally, our conclusions are presented in Section VI.

2. Related work

Cameras serve as primary sensors in autonomous driving systems and enable the vehicles to perceive surrounding vehicles, pedestrians, traffic signs, and so no. Object detection using digital image is a fundamental task in autonomous vehicles with a variety of detection algorithms proposed in recent years, including Faster R-CNN (Ren et al., 2017), YOLO (Redmon et al., 2016; Adarsh et al., 2020), and CenterNet (Zhou et al., 2019). Some open datasets have been put forward to advance the field of computer vision, as with ImageNet (Deng et al., 2009), KITTI (Geiger et al., 2013), BDD100K (Yu et al., 2020), and COCO (Lin et al., 2014). However, these datasets are biased towards clear weather environments. The quality of camera image often degrades in adverse weather conditions, degrading the detection

performance in turn. Thus, neural networks trained in normal weather are often incompetent in adverse weather. Expanding the coverage of training datasets by augmenting them with different weather information helps to improve the robustness of detection algorithms (Abdelraouf et al., 2022). Sakaridis et al. (2018) proposed the synthetic foggy cityscape dataset by adding synthetic fog layers to real-world images, and experiments demonstrated that semantic segmentation accuracy can be improved using this resulting dataset. To bridge the domain gap between synthetic datasets and the real world, a real-world dataset was proposed in (Ba et al. 2022). Kenk and Hassaballah (2020) proposed the DAWN dataset for vehicle detection in adverse weather conditions, including heavy fog, rain, snow, and sandstorms, but these datasets lack information about the weather intensity.

In addition to image augmentation, some new algorithms also have been proposed to tackle the problem. Approaches fall into two major categories: restore the clear images by designing deraining and dehazing models, and design novel algorithms compatible with inclement weather. By bridging the domain gap between synthetic datasets and the real world, transfer learning has been studied in object detection in harsh conditions (Wang et al., 2022; Zhang et al., 2022). Porav et al. (2019) proposed a segmentation algorithm for images affected by adherent rain drops. Bijelic et al. (2020) established a single-shot model without access to real-world weather-distorted data by adaptively fusing multimodal features, and the prominent experimental results showed its effectiveness.

Various techniques have been proposed for removing rain streaks from images. Yang et al. (2020a) introduced a two-stage self-learning deraining network making use of the high correlation between adjacent frames in video sequences. Deng et al. (2021) proposed a rotational video deraining algorithm using nonconvex and nonsmooth algorithm to remove rain streaks in natural and stochastic scenes. Comparatively, removing rain streaks from individual images is more difficult since less information is available. Existing algorithms are commonly filter-based, prior-based, and learning-based methods (Wang et al., 2019). Ding et al. (2016) analyzed the physical properties of rain in images and proposed a guided L0 smoothing filter based on the prior knowledge that small rain streaks are brighter than its adjacent pixels. Maximum posterior has been applied in rain removal in another study (Mu et al., 2018), and a morphological component analysis-based deraining framework using dictionary learning and sparse coding was proposed (Fu et al., 2011). Taking the prior structures of rain streaks into consideration, Wang et al. (2021) designed a residual deraining network and another researcher used a guided filter to remove the raindrops. Shi et al. (2018) established a weighted median guided filter to preserve geometrical details. Another method employed a multiple-guided filter using the low-frequency parts of rainy images and high-frequency parts of clear images was proposed (Zheng et al., 2013). Guo et al. (2009) designed a pixel-wise dilation filter, where kernels for each pixel were learned from the datasets.

In the aspect of learning-based methods, deep neural network has been widely studied, and state-of-the-art performances are achieved with some datasets. Some methods use end-to-end supervised learning. Convolutional neural network has been applied to estimate the rain layer (Li & Gai, 2023; Yang et al., 2020b). Fu et al. (2017) designed a deep residual network-based method focusing on high-frequency detail during training. He and Patel (2018) proposed a novel density-aware, multiple streams, densely connected convolutional network to estimate the rain density that was used to guide the rain removal process. A recurrent memory unit network was proposed in (Zhang et al., 2021) to remove rain streaks from individual images. Hu et al. (2019) analyzed the visual effects of rain subjects on the scene depth and presented the RainCityscapes dataset. Graph convolutional network (GCN) was introduced to remove the rain streaks (Fu et al., 2021). GCN was used to learn the spatial coherence and channel correlation in (Hu, 2022). A novel spatial attentive network was proposed in (Wang et al., 2019) where the temporal priors and human supervision were incorporated to improve the image quality of real-world data. The rain layer was

parameterized as a latent variable vector in (Wang et al., 2021). Most of these algorithms have adopted an end-to-end training method and require abundant rainy-clear image pairs. Yet it remains nearly impossible to collect the ground truth of rainy images, which means that some well-designed neural networks are hard to train in real-world applications. Consequently, Wei et al. (2019) proposed a semi-supervised deraining paradigm that did not need clear images. Chen et al. (2022) proposed a method that consists of two cooperative branches. Generative adversarial network (GAN) methods, which are unsupervised, including generators and discriminators, have also been studied (Matsui & Ike-hara, 2021; Lin et al., 2020). By optimizing the two models together in a zero-sum game, GANs have achieved impressive results in some domains such as image-to-image translation.

With respect to autonomous driving, rain has the potential to cause traffic safety accidents. With exceedingly high requirements for driving safety, autonomous vehicles must be validated as safe before large-scale application. The influence of adverse weather is particularly challenging to the perception algorithm, and it is imperative to ensure the robustness of autonomous vehicles in all weather conditions. However, it is very difficult to test such vehicles in real-world conditions due to the inherent randomness, uncontrollability, and unpredictability of such environments. To tackle this problem, different solutions have been proposed, including software-in-the-loop (SIL) (Kang et al., 2019), hardware-in-the-loop (HIL) (Zhu et al., 2021), and vehicle-in-the-loop (VIL) (Tettamanti et al., 2018) methods. High-fidelity 3D environments can be established applying sophisticated graphic techniques using rendering engines in the simulation test. By adding a physical plant in the test pipeline, HIL provides an effective method for evaluating the performance of autonomous driving systems in real-time simulations. However, only simulated weather conditions are usable with these existing methods, and it is difficult to bridge the gap between software simulation and real-world conditions. Establishing a rainfall simulator in a closed field provides an efficient solution for autonomous vehicle testing. Because of the flexibility in rainfall level control and driving scenario construction, rainfall simulation has the potential to support testing autonomous vehicles thoroughly in different conditions. In this paper, we propose a rainfall simulation construction scheme and collect a large amount of data to compensate for the lack of rainfall levels in some open datasets. To improve the image quality of rainy images, a scene content information-based deraining algorithm is proposed in this paper.

Conclusions drawn from most existing studies can be summarized as follows. Autonomous driving in adverse conditions is difficult to implement in a real-world environment. Current publicly available rainy image datasets lack information about the rainfall severity. And most existing deraining methods address rain streak characteristics, rather than the scene content. This paper aims to design a reliable and flexible rainfall simulation and to provide a method for improving rainy image quality to improve perception algorithm performance in rainy conditions. Inspired by the fact that humans understand rainy images directly, we treat rainy images as combinations of scene content and raindrop disturbances to propose a single-image deraining algorithm based on the image content estimation in this paper. To enhance the image content and suppress the raindrop influence, a pixel-wise attention module is proposed.

3. Driving data collection

Based on current research into autonomous vehicle testing in closed fields, we design a rainfall simulation environment at CATARC Automotive Proving Ground Co., Ltd. and collect a new *Closed Field Rain* dataset.

3.1. Hardware platform

We built a high-fidelity and controllable rainfall simulation test

environment using the major components shown in Fig. 1(a). There are five hardware components: water reservoir, controller, pump, nozzle, and pipe. The water stored in the reservoir is pumped to the nozzle through the pipe, and the flow is adjusted by the controller. The rain level is determined by the diameter of raindrops, the rainfall intensity, and the rainfall uniformity.

To reflect natural phenomena, we design the height of the rainfall unit to be 5.75 m. Raindrops' terminal speed falls between 2 and 9 m/s depending on the height of the vehicle and initial raindrop velocity. The road was designed as a two-lane, one-way road, with each lane 3.75 m wide. The total length of the road is 100 m to match the detection range of the onboard sensors. Two 22 kw pumps are utilized in the scheme. Three different kinds of nozzles are applied to create moderate rain, heavy rain, and torrential rain using apertures of 2.5 mm, 3.5 mm, and 4.5 mm, respectively. The real environment is illustrated in Fig. 1(b).

3.2. Experiment cycle and data collection

To collect enough data for different driving conditions, a fake car and a dummy are used as traffic targets to create different scenes. The image data is captured in conditions of moderate rain, heavy rain, and torrential rain, with the distance between the camera and target set to be 5 m, 10 m, 15 m, 20 m, 25 m, and 30 m in different testing cycles. Thus, driving data from 36 testing cycles are obtained to construct our *Closed Field Rain* dataset. As presented in Fig. 2(a), our test vehicle is equipped with several sensors including a 32-line Lidar, a radar and two cameras. Camera 1 is a monocular camera that provides images with a resolution of 1200×720 pixels at 20fps. Camera 2 is a ZED binocular camera with a resolution of 2560×720 pixels. The real rain simulation scene and traffic targets are presented in Fig. 2(b).

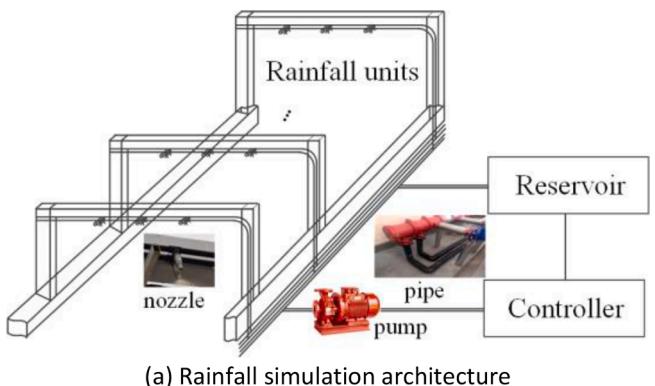


Fig. 1. Rainfall simulation scheme. Some rainfall units are applied to establish the rainfall simulation condition.



Fig. 2. Data-acquisition platform. By installing sensors on the vehicle, driving data can be collected. A fake car and a dummy are used to construct the driving scenes.

4. Image deraining using pixel-wise content attention

To improve the image quality captured in rain, a single-image deraining network is proposed in this paper. And better performance may be achieved by taking the temporal information from the image sequence into consideration in the future research. This section will detail the algorithm in the followings.

4.1. Neural network architecture

Previous studies usually treated the rainy image as either a combination of a clear image, a rain layer, and a fog layer, or as a disturbance addressable with filter-based methods. We treat the rainy image as a

combination of clear content and rainy noise. Thus, image deraining is the task of restoring image content from a degraded rainy image. **Fig. 3** shows the overall architecture of our proposed network. According to the GAN design, there are two parts of the algorithm: a generator and a discriminator. The generator accepts the rainy image as input and restores the clear image as output, and the discriminator evaluates the output as a clear image or a rainy image. Through this interaction, both the generator and discriminator are optimized. Although these modules already exist in some studies, the deraining framework of estimating the pixel-wise significance is novel.

The goal of deraining can be expressed as estimating a clear image I from rainy image O , which is an ill-posed problem. Most existing studies determine rain streaks using end-to-end training, but pay little attention to the content of the rainy image. Humans can eliminate the rain disturbances and understand the scene information from the semantic content instead of analyzing each pixel. Inspired by this idea, we propose the deraining method illustrated in **Fig. 3**, where a content estimation module, a pixel-wise attention module, an image inpainting module, and a patch discriminator are included. Parameters of these modules are optimized simultaneously in an adversarial manner. The content estimation module captures the scene content representation M_{con} of the rainy image. The pixel-wise attention module that produces the pixel-wise attention map $M_{atten} \in [0, 1]$ is used to estimate the probability that each pixel contains scene information or a raindrop. The image inpainting module performs image completion using a U-Net structure (Ronneberger et al., 2015). The coarse image is obtained via the element-wise multiplication of M_{con} and M_{atten} , and the concatenation with the rainy image as input of the inpainting module. The output M_{res} of the image inpainting module contains the representation of the fog layer and the impaired parts.

4.2. Content estimation block

The content estimation block extracts the scene content representation of the input image. Huang et al. (2018) assumed the target domain and source domain datasets are composed of a content code and a style code in the study of image-to-image translation. Inspired by this research, we adopt an encoder-decoder architecture to extract the content representation of rainy images. To adapt the encoder for different levels of rainfall, a shared block is used for different rainy image domains. The encoder E_{stru} consists of convolutional layers and residual blocks to embed the scene content features. The decoder D_{stru} reconstructs the content map. Several convolutional layers are employed to restore the M_{con} with the same resolution as the input rainy image.

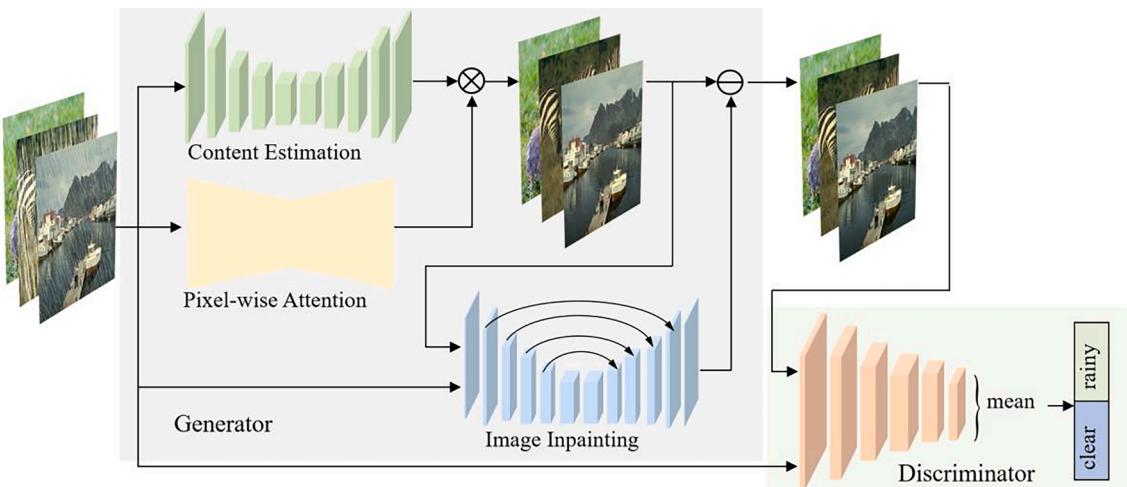


Fig. 3. Schematic illustration of the proposed model. The generator consists of a content estimation module, a pixel-wise attention learning module and an image inpainting module. The discriminator is composed of several convolutional layers.

4.3. Pixel-Wise attention block

To eliminate the disturbances of irrelevant feature during scene content extraction, we introduce a pixel-wise attention block to suppress the effects of raindrops. Just like the content estimation module, the significance of each pixel is estimated using an encoder-decoder manner from the input rainy images. We attempt to construct the network using different architectures, that is, the dilated residual network and Transformer-based model. In the PWSA-Res (residual network-based scene attention network), dilated residual convolutions are utilized in both the encoder and decoder, and a sigmoid layer is added to achieve an attention map M_{atten} with the same resolution as the rainy image.

In the PWRA-Tran (Transformer-based scene attention network) implementation, the encoder uses a Transformer algorithm based on the work of Vision Transformer (Dosovitskiy et al., 2010). The decoder is a convolutional network. As presented in Fig. 4, the PWRA-Tran consists of four components to convert the learned feature to the shape of M_{atten} : feature embedding, position embedding, Transformer encoder, and convolutional decoder.

The rainy image is denoted as $x \in R^{3 \times H \times W}$, where H and W are the height and width of the image, respectively. The image is split into some patches $[x_1, x_2, \dots, x_N]$, where $x_i \in R^{3 \times L_h \times L_w}$, and (L_h, L_w) is the patch size, and the number of patches is $N = H \times W / (L_h \times L_w)$. Each patch is projected into a new feature space and fed into the encoder combined with position encodings obtained from a learnable position embedding method as $y_0 = [f_1, f_2, \dots, f_N]$. Applying a multi-head self-attention module, the computation can be conducted as given in Eqs. (1)–(3) (Chen et al., 2021):

$$q_i = k_i = v_i = LN(y_{i-1}) \quad (1)$$

$$y'_i = MSA(q_i, k_i, v_i) + y_{i-1} \quad (2)$$

$$y_i = FFN(LN(y'_i)) + y'_i \quad (3)$$

The output of the Transformer encoder is denoted as y_l , where l is the number of layers. LN is layer normalization; MSA is the multi-head self-attention module; FFN represents the feed forward network.

In the design of decoder, a 3-layer network of 1×1 conv + sync batch norm + 1×1 conv is used to project the encoder's features to RGB channel. Then we bilinearly upsample the output to the same resolution of the rainy image. Finally, a sigmoid layer is added at the end of the convolutional decoder. The resulting attention map M_{atten} is used to weight the content map to obtain the coarse image M_{img} :

$$M_{img} = M_{con} \otimes M_{atten} \quad (4)$$

where \otimes denotes element-wise multiplication.

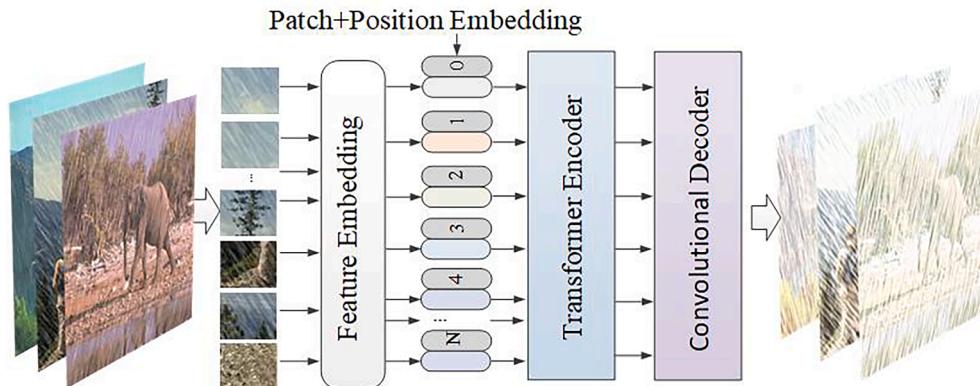


Fig. 4. PWRA-Tran framework. The rainy image is split into a predetermined number of patches, and each patch and its position are linearly embedded. The encoder is designed using the Transformer network. The output of the encoder is processed using a convolutional network based decoder to restore the same resolution as the rainy image.

4.4. Image inpainting block

To obtain the fog layer and restore the coarse image, a U-Net structure is introduced, as present in Fig. 3. It is an encoder-decoder structure with five Conv-BN-Relu blocks in both encoder module and decoder module, and a skip connection concatenates different features. The input of the image inpainting block is the raw rainy image and the coarse image M_{img} . The output M_{res} is predicted with the same shape as the rainy image.

4.5. Discriminative training

GANs have achieved impressive results in some applications, with its adversarial learning method reducing the reliance on ground truth samples. In real traffic scenes, the captured rainy images have no ground truth due to the changing scene and randomness. Motived by current research into GANs using global and local discriminators, this paper uses PatchGAN (Isola et al., 2016) for training the deraining network. A convolutional network is used as the discriminator as presented in Fig. 3, with the output being a 2D-shape tensor with 1 channel. In the discriminator, six strided convolutions with kernel size 4 and stride 2 are stacked to evaluate the output image. The resulting evaluation matrix is used as the assessment result. Each element focuses on a different region of the image. The average of the evaluation matrix is applied as a loss function of the generative module:

$$L_{G1} = \frac{\lambda_g}{H_{dis}W_{dis}} \sum_{j=1}^{H_{dis}} \sum_{i=1}^{W_{dis}} (\log(-D(P))) \quad (5)$$

where P is the output image of deraining network, $D(P)$ denotes the output of the discriminative network, λ_g is a constant, and $[H_{dis}, W_{dis}]$ is the dimension of discriminator output.

Some synthetic datasets have been proposed for deraining purpose, where the ground truth T is provided for evaluation. To utilize this kind of dataset, a perceptual loss in the training of generative network is added in this paper, which provides a semantic discrepancy. By using a pre-trained VGG16 model, the loss can be computed as:

$$L_{G2} = L_1(VGG(P), VGG(T)) \quad (6)$$

Thus, the loss of deraining network can be obtained as:

$$L_G = L_{G2} + L_{G1} \quad (7)$$

The discriminative network is used to discriminate whether P is true or fake. The whole loss function of the discriminator can be written as:

$$L_D = -\frac{\lambda_1}{H_{dis}W_{dis}} \sum_{j=1}^{H_{dis}} \sum_{i=1}^{W_{dis}} ((\log(1 - D(P))) - \frac{\lambda_2}{H_{dis}W_{dis}} \sum_{j=1}^{H_{dis}} \sum_{i=1}^{W_{dis}} (\log(D(x)))) \quad (8)$$

where $D(x)$ is the evaluation of rainy image, and λ_1 and λ_2 are constants.

5. Experiments

To validate the algorithm, experiments are conducted using Rain1400, Rain100H (Yang et al., 2017), and real-world datasets, and the dataset collected in section III.

5.1. Results on synthetic datasets

Several open synthetic benchmark datasets have been proposed: Rain12 (Li, 2016), Rain100L (Yang et al., 2019), Rain100H and Rain1400. Rain12 has only 12 image pairs that are not usable as training sample. Rain100L consists of 200 training image pairs and 100 for test pairs. Rain100H has 1800 image pairs for training and 100 for evaluation, which is more challenging. Rain1400 has 14 types of rain streaks and consists of 14,000 rainy images and 1000 clear images. To comprehensively validate and evaluate our algorithm, some experiments are carried out using Rain100H and Rain1400, both of which are challenging datasets. Several commonly used metrics are used, including the peak signal to noise ratio (PSNR) and structural similarity (SSIM). Both PSNR and SSIM are in the range of [0, 1], with higher values indicating better results. If two images are the same, PSNR and SSIM are both 1. We take the clear image as ground truth to compute the PSNR and SSIM values. Quantitative evaluations are presented to compare with some existing methods.

The network architecture presented in Fig. 3 is implemented using Pytorch, and Adam optimization is performed with $\beta_1 = 0.9$ and $\beta_2 =$

0.99 is applied. We use an Nvidia RTX3080 for the training. To visually display the deraining performance, we have included several results from Rain100H and Rain1400 in Fig. 5, which help to provide qualitative evidences for the proposed model. As presented above, M_{atten} has the same resolution as the input rainy image (has RGB channels). The attention visualization is obtained by converting the pixel-wise attention value ([0, 1]) to an RGB value in the range of 0–255 by multiplying by 255 (255* M_{atten}). As shown in Fig. 5, the attention maps focus on the rain streaks region, which helps to remove rain streaks from the rainy images and restore the image contents.

Based on the evaluation results of several rain removal algorithms in (Chen et al., 2020), we compare the performance of PWRA-Res and PWRA-Tran with SPANet, PReNet (Ren et al., 2019), MPRNet (Zamir et al., 2021), HINet (Chen et al., 2021), Restormer (Zamir et al., 2022), MARDNet (Wang et al., 2020), RCDNet (Wang et al., 2020), DualGCN (Fu et al., 2021), MPRNet (Zamir et al., 2021), and RLNet (Chen & Li,

Table 1
Quantitative results on synthetic datasets.

Model	Metrics on Rain1400		Metrics on Rain100H	
	PSNR	SSIM	PSNR	SSIM
SPANet	28.57	0.8913	25.11	0.8332
PReNet	30.73	0.9184	29.45	0.8980
MPRNet	31.72	0.9186	30.41	0.89
HINet	31.55	0.9157	30.65	0.894
RCDNet	31.26	0.9127	30.83	0.906
DualGCN	30.28	0.9105	29.25	0.894
MPRNet	29.81	0.9042	29.65	0.887
RLNet	30.12	0.9133	28.87	0.902
Restormer	32.35	0.9280	31.46	0.904
MARDNet	31.68	0.9215	30.19	0.9153
PWRA-Res	30.01	0.9016	29.607	0.8806
PWRA-Tran	31.50	0.9282	31.50	0.9091



Fig. 5. Performance of our method on several testing images in Rain100H and Rain1400. The first two rows are taken from Rain100H, and the remaining images are taken from Rain1400. The first column shows the rainy images. The second column presents a pixel-wise attention visualization obtained from the pixel-wise attention blocks. The third column shows the rain layer extracted by the network. The fourth column shows the output of our method. The fifth column shows the ground truth provided by the datasets.

2021). The quantitative metrics are presented in Table 1, with the proposed method achieving good performance. As can be seen from the table, both PWRA-Res and PWRA-Tran achieve high quantitative scores and indicating that our models remove rain streaks from rainy images well.

5.2. Experiments on real datasets

As depicted in section III, a large amount of data is collected in the closed testing field. About 31,000 images can be extracted from the dataset. To maintain the differences among training samples, we sample 5800 rainy images for algorithm evaluation captured from Camera 1. And 5000 images are used as training samples to tune the model parameters, with the remainder used as testing samples. And the model of PWRA-Tran is evaluated.

As illustrated in Fig. 6, the results of our algorithm from four testing cycles are presented. Case 1: torrential rain, and the distance between the camera and the fake car is 10 m. Case 2: torrential rain, and the distance between camera and the fake car is 15 m. Case 3: heavy rain, and the distance between camera and the fake car is 20 m. Case 4: torrential rain, and the distance between the camera and the fake car is 25 m.

Several image areas are marked with red rectangles in Fig. 6. The patches located at the bottom left of the images are partially enlarged detail of raindrops or fog. As presented, the image clarity of roadside grass was upgraded, and the raindrops were removed. Through statistical analysis, the proposed PWRA-Tran model is effective in removing the raindrops running at 17.5 FPS.

To provide further evidence of our proposed deraining method's abilities, we employ several no-reference image quality metrics to test the algorithm without requiring corresponding clear images. Both Brenner and Tenengrad gradients are applied to evaluate the image quality quantitatively. As illustrated in Fig. 7, the average image quality is improved by using our deraining method, with both the Brenner gradient and Tenengrad gradient increase.

Object detection is an important task for autonomous vehicle perception, but rain, snow, and fog reduce detection accuracy. Some existing studies have shown that the negative influence can be alleviated by rain, fog, and snow removal algorithms.

To quantitatively evaluate the algorithm performance, we use the number of image feature points as indicators. SIFT (Scale Invariant Feature Transform) and ORB (Oriented FAST and Rotated BRIEF) are two methods that have been widely used in autonomous driving systems

(Juan & Gwun, 2009; Rublee et al., 2011). Examples of extracted feature points are shown in Fig. 8.

To compare the number of SIFT and ORB feature points in different testing cycles, the feature points of images collected at different distances between camera and the object are computed, individually. As illustrated in Fig. 9, the average number of extracted feature points increases by 4.255%–36.328%. The number of ORB points increases from 3116 to 4248 in the cycle that the distance between the camera and the object is 30 m, and an increase of 36.328% is obtained.

Additional experiments are conducted using open real-world datasets. The datasets of Rain in Driving (RID, has 2495 images), Rain in Surveillance (RIS, has 2348 images), and SPA-Data (28500 for training and 1000 for testing) are used (Jiang et al., 2020; Li et al., 2019; Wang et al., 2019). We split the RID and RIS into three parts: 200 images for testing, and 200 images for validation, and the remaining for training. The average scores in RID and RIS are presented in Table 2. As can be seen, the presented deraining algorithm can improve the real-world images. Moreover, several representative results are shown in Fig. 10 for visual comparison. Rain streaks can be removed in the open datasets, which verifies the proposed method from another aspect.

5.3. Ablation study

To analyze the effectiveness of each module in the deraining algorithm, an ablation study is conducted. To fully verify different modules, three different variants are evaluated with average results shown in Table 3, where CET is the content estimation module, and INP denotes the image inpainting module. As can be seen, both CET + INP + PWRA-Res and CET + INP + PWRA-Tran achieve good deraining results, which illustrate that the pixel-wise attention block can improve the performance greatly. And Transformer model has better results compared with residual network.

6. Conclusions

To test autonomous vehicles and provide a method to analyze the influence of adverse weather on autonomous driving systems, we design and build a rainfall simulation environment in this paper. Using this environment, we construct the *Closed Field Rain* dataset with types of rainfall (moderate rain, heavy rain, and torrential rain) to aid the study of perception algorithms in rainy conditions. In addition, a pixel-wise content attention learning deraining network is proposed to determine scene content from rainy images. A pixel-wise attention block is



Fig. 6. Rain removal effect on Closed Field Rain Dataset. The clarities of rainy images are improved.

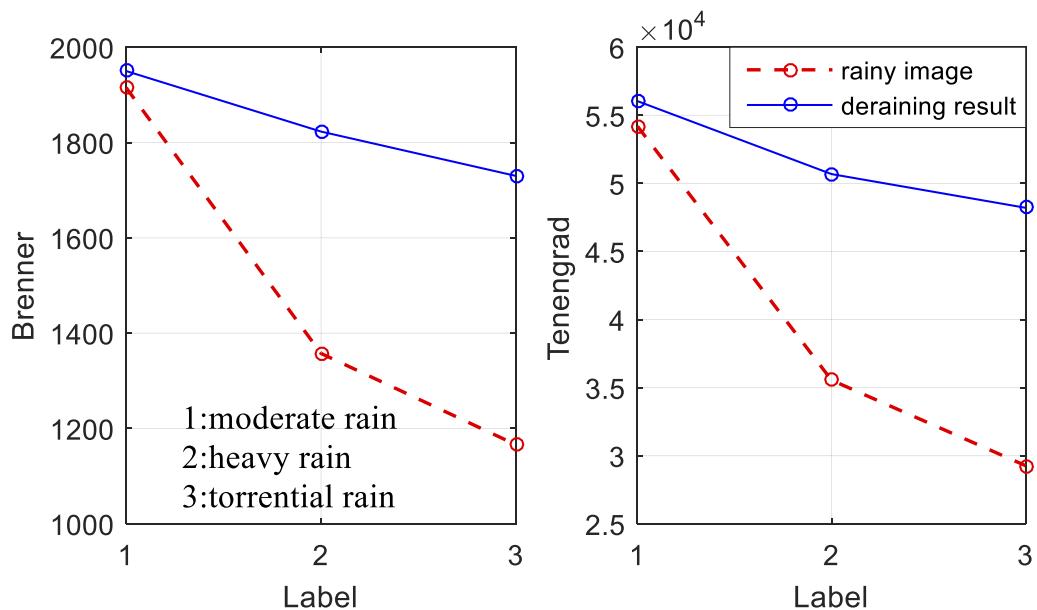


Fig. 7. No-reference metrics of the deraining performance. Both Brenner and Tenengrad gradients are improved by employing the proposed algorithm.

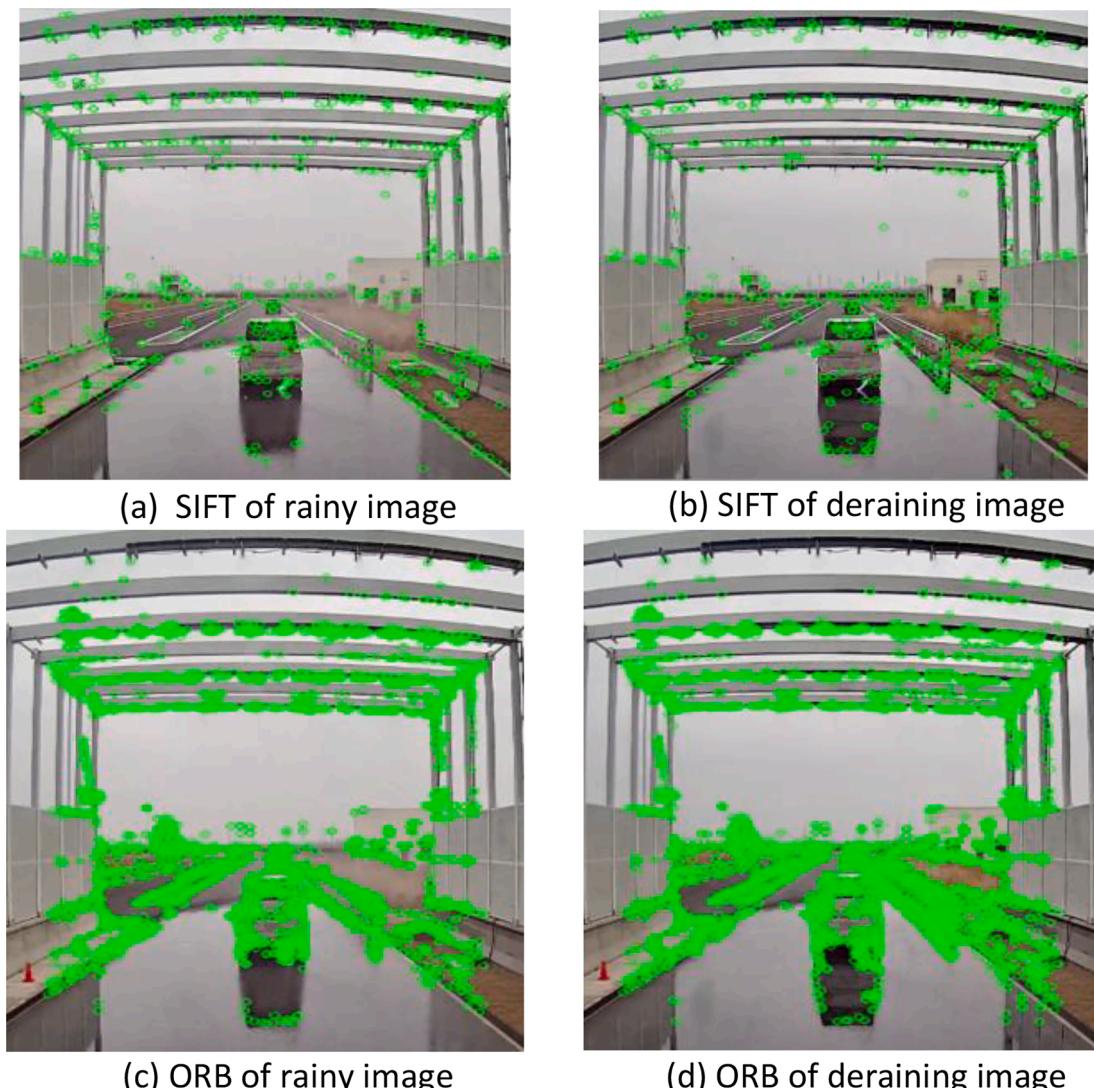


Fig. 8. Comparisons of the rainy image and the deraining result.

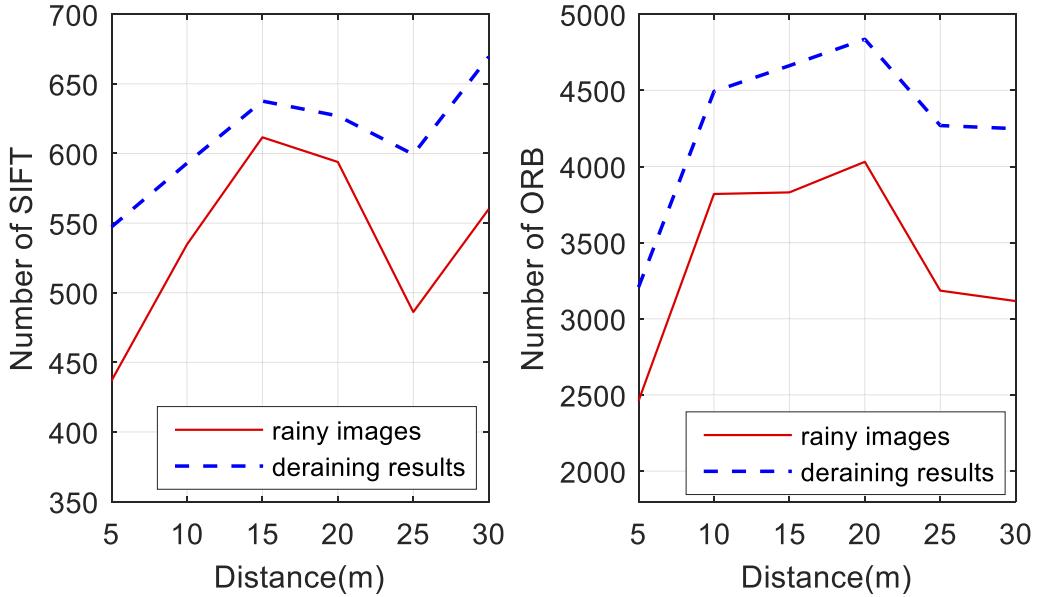


Fig. 9. The number of SIFT and ORB feature points detected from the rainy image and the output of our deraining algorithm.

Table 2
Average results in RID/RIS/SPA-Data.

Indicator	Brenner	Tenengrad	SIFT	ORB
Rainy images	117.56/ 190.5/450.2	189.5/3.07e3/ 0.89e3	856.1/ 441.6/239	3438.7/ 1232.7/460
Output images	138.42/ 215.3/493.9	215.3/4.62e3/ 1.18e4	1091.2/ 616.4/320	4208.8/ 1637.3/489

established to suppress the information of raindrop and enhance the response of scene content by learning a pixel-wise attention map. And two kinds of block, PWSA-Res and PWSA-Tran are presented. A U-Net based image inpainting module is added to predict the fog layer and the incomplete information. Finally, we conduct experiments using synthetic and real datasets including *Closed Field Rain* dataset to

demonstrate our method's performance compared to others. It is demonstrated that PWSA-Tran performs well overall and comparable performance can be achieved. Nonetheless, it has some limitations. Performance degradation of this method may appear in some different meteorological conditions (e.g., foggy and snowy environments). As future work, we plan to study enhancement methods applicable in some

Table 3
Ablation study. Three variants of the algorithm are evaluated.

method	Brenner	Tenengrad	SIFT	ORB
Original rainy image	1.479e3	3.967e4	537.199	3.407e3
CET + INP	1.548e3	4.269e4	547.653	3.549e3
CET + INP + PWA-Res	1.815e3	5.046e4	604.153	4.092e3
CET + INP + PWA-Tran	1.834e3	5.165e4	612.272	4.286e3

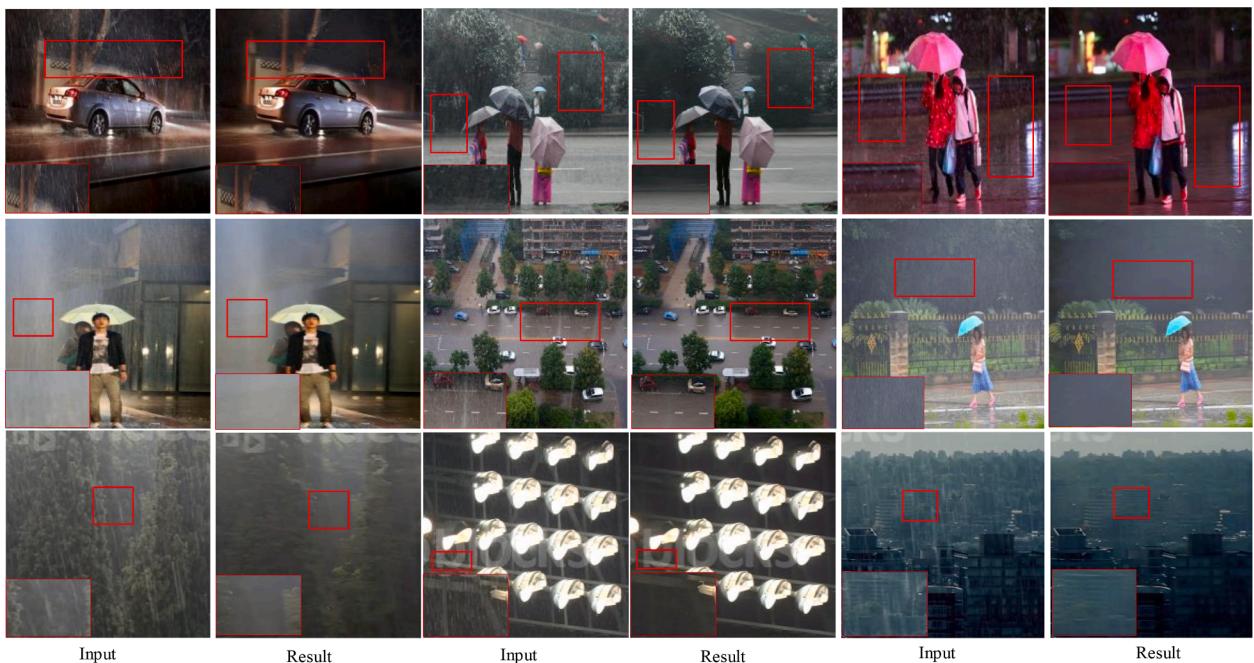


Fig. 10. Rain removal performance of our method on real-world images.

different adverse environments taking the temporal feature obtained from image sequence into consideration. And delve into the performance bounds analysis to ensure driving safety. In addition, we plan to deploy our deraining method in an autonomous vehicle, so that it can be tested in the real-world environment.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work is supported by the National Natural Science Foundation for Regional Innovation and Development (U22A20247), the Natural Science Basis Research Program of Shaanxi Province (2023-JC-QN-0393), and the National “111” Project of China’s Higher Education (B14043).

Data availability

Data will be made available on request.

References

- Abdelraouf, A., Abdel-Aty, M., & Wu, Y. (2022). Using vision transformers for spatial-context-aware rain and road surface condition detection on freeways. *IEEE Transactions on Intelligent Transportation Systems*.
- Adarsh, P., Rathi, P., & Kumar, M. (2020). YOLO v3-Tiny: Object Detection and Recognition using one stage improved model[C]//2020. In *6th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 687–694).
- Ba Y, Zhang H, Yang E, et al. Not Just Streaks: Towards Ground Truth for Single Image Deraining[C]//Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII. Cham: Springer Nature Switzerland, 2022: 723–740.
- Bijelic M, Gruber T, Mannan F, et al. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 11682–11692.
- Chen C, Li H. Robust representation learning with feedback for single image deraining [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 7742–7751.
- Chen H, Wang Y, Guo T, et al. Pre-trained image processing transformer[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 12299–12310.
- Chen L, Lu X, Zhang J, et al. HiNet: Half instance normalization network for image restoration[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 182–192.
- Chen X, Pan J, Jiang K, et al. Unpaired deep image deraining using dual contrastive learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 2017–2026.
- Cheng, G., Wang, Z., & Zheng, J. Y. (2018). Modeling weather and illuminations in driving views based on big-video mining. *IEEE Transactions on Intelligent Vehicles*, 3 (4), 522–533.
- Chen X, Huang Y, Xu L. Multi-scale Attentive Residual Dense Network for Single Image Rain Removal[C]//Proceedings of the Asian Conference on Computer Vision. 2020.
- Deng, J., Dong, W., Socher, R., et al. (2009). Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. IEEE, 248–255.
- Deng, L., Xu, G., Zhu, H., et al. (2021). RoDeRain: rotational video derain via nonconvex and nonsmooth optimization. *Mobile Networks and Applications*, 26(1), 57–66.
- Ding, X., Chen, L., Zheng, X., et al. (2016). Single image rain and snow removal via guided L0 smoothing filter. *Multimedia Tools & Applications*, 75(5), 2697–2712.
- Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.
- Fu X, Huang J, Zeng D, et al. Removing rain from single images via a deep detail network [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 3855–3863.
- Fu X, Qi Q, Zha Z J, et al. Rain streak removal via dual graph convolutional network[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(2): 1352–1360.
- Fu Y H, Kang L W, Lin C W, et al. Single-frame-based rain removal via image decomposition[C]//2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2011: 1453–1456.
- Geiger, A., Lenz, P., Stiller, C., et al. (2013). Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11), 1231–1237.
- Guo Q, Sun J, Juefei-Xu F, et al. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining. arXiv preprint arXiv:2009.09238, 2020.
- He Z, Patel V M. Density-Aware Single Image De-raining Using a Multi-stream Dense Network[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.
- Hu B. (2022). Single image deraining using contrastive perceptual regularization. *IET Image Processing*, 16(10), 2759–2768.
- Hu X, Fu C W, Zhu L, et al. Depth-attentional features for single-image rain removal[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 8022–8031.
- Huang X, Liu M Y, Belongie S, et al. Multimodal unsupervised image-to-image translation [C]// Proceedings of the European conference on computer vision (ECCV). 2018: 172–189.
- Isola P, Zhu J Y, Zhou T, et al. Image-to-Image Translation with Conditional Adversarial Networks[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2016.
- Jiang K, Wang Z, Yi P, et al. Multi-scale progressive fusion network for single image deraining[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 8346–8355.
- Juan, L., & Gwin, O. (2009). A comparison of sift, pca-sift and surf. *International Journal of Image Processing (IJIP)*, 3(4), 143–152.
- Kang, Y., Yin, H., & Berger, C. (2019). Test your self-driving algorithm: An overview of publicly available driving datasets and virtual testing environments. *IEEE Transactions on Intelligent Vehicles*, 4(2), 171–185.
- Li, P., & Gai, S. (2023). Single image deraining using multi-scales context information and attention network. *Journal of Visual Communication and Image Representation*, 90, Article 103695.
- Lin, C. T., Huang, S. W., Wu, Y. Y., et al. (2020). GAN-based day-to-night image style transfer for nighttime vehicle detection. *IEEE Transactions on Intelligent Transportation Systems*, 22(2), 951–963.
- Kenk, M. A., & Hassaballah, M. (2020). DAWN: vehicle detection in adverse weather nature dataset. *arXiv preprint arXiv:2008.05402*.
- Li S, Araujo I B, Ren W, et al. Single image deraining: A comprehensive benchmark analysis[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 3838–3847.
- Li, W., Pan, C. W., Zhang, R., et al. (2019). AADS: Augmented autonomous driving simulation using data-driven algorithms. *Science Robotics*, 4(28).
- Li Y, Tan R T, Guo X, et al. Rain streak removal using layer priors[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2736–2744.
- Lin T Y, Mairi M, Belongie S, et al. Microsoft coco: Common objects in context[C]// European conference on computer vision. Springer, Cham, 2014: 740–755.
- Matsui T, Ikebara M. Gan-based rain noise removal from single-image considering rain composite models[C]//2020 28th European Signal Processing Conference (EUSIPCO). IEEE, 2021: 665–669.
- Mu, P., Chen, J., Liu, R., et al. (2018). Learning bilevel layer priors for single image rain streaks removal. *IEEE Signal Processing Letters*, 26(2), 307–311.
- Porav, H., Bruls, T., & Newman, P. (2019). I can see clearly now: Image restoration via de-raining. *2019 International Conference on Robotics and Automation (ICRA)*. IEEE.
- Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779–788.
- Ren D, Zuo W, Hu Q, et al. Progressive image deraining networks: A better and simpler baseline[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 3937–3946.
- Ren, S., He, K., Girshick, R., et al. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 39(6), 1137–1149.
- Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234–241.
- Rublee, E., Rabaud, V., Konolige, K., et al. (2011). ORB: An efficient alternative to SIFT or SURF[C]//2011 International conference on computer vision. IEEE, 2564–2571.
- Sakaridis, C., Dai, D., & Van Gool, L. (2018). Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126(9), 973–992.
- Scanlon, J. M., Kusano, K. D., Daniel, T., et al. (2021). Waymo simulated driving behavior in reconstructed fatal crashes within an autonomous vehicle operating domain. *Accident Analysis & Prevention*, 163, Article 106454.
- Sharma, T., Debaque, B., Duclos, N., et al. (2022). Deep learning-based object detection and scene perception under bad weather conditions. *Electronics*, 11(4), 563.
- Shi, Z., Li, Y., Zhang, C., et al. (2018). Weighted median guided filtering method for single image rain removal. *Eurasip Journal on Image & Video Processing*, 2018(1), 35.
- Tettamanti T, Szalai M, Vass S, et al. Vehicle-in-the-loop test environment for autonomous driving with microscopic traffic simulation[C]//2018 IEEE International Conference on Vehicular Electronics and Safety (ICVES). IEEE, 2018: 1–6.
- Wang, G., Liu, J., Lo, W., et al. (2020). Learning multiple instance deep quality representation for robust object tracking. *Future Generation Computer Systems*, 113, 298–303.
- Wang, H., Wu, Y., Xie, Q., et al. (2021). Structural residual learning for single image rain removal. *Knowledge-Based SLSTMsystems*, 213, Article 106595.

- Wang H, Yue Z, Xie Q, et al. From rain generation to rain removal[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 14791-14801.
- Wang Y, Ma C, Liu J. Removing Rain Streaks via Task Transfer Learning. arXiv preprint arXiv:2208.13133, 2022.
- Wei W, Meng D, Zhao Q, et al. Semi-supervised transfer learning for image rain removal [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 3877-3886.
- Yang W, Tan R T, Wang S, et al. (2020a). Self-learning video rain streak removal: When cyclic consistency meets temporal correspondence[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 1720-1729.
- Wang H, Xie Q, Zhao Q, et al. A model-driven deep neural network for single image rain removal[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 3103-3112.
- Yang W, Tan R T, Feng J, et al. Deep joint rain detection and removal from a single image [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1357-1366.
- Yang, W., Tan, R. T., Feng, J., et al. (2019). Joint rain detection and removal from a single image with contextualized deep networks. *IEEE transactions on pattern analysis and machine intelligence*, 42(6), 1377-1393.
- Wang T, Yang X, Xu K, et al. Spatial attentive single-image deraining with a high quality real rain dataset[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 12270-12279.
- Yu F, Chen H, Wang X, et al. Bdd100k: A diverse driving dataset for heterogeneous multitask learning[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 2636-2645.
- Zamir S W, Arora A, Khan S, et al. Multi-stage progressive image restoration[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 14821-14831.
- Zamir S W, Arora A, Khan S, et al. Restormer: Efficient transformer for high-resolution image restoration[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 5728-5739.
- Yang, W., Tan, R. T., Wang, S., et al. (2020b). Single Image Deraining: From Model-Based to Data-Driven and Beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99, pp. 1-1.
- Zhang, X. X., Lu, X. Y., & Peng, L. (2022). A complementary and precise vehicle detection approach in RGB-T images via semi-supervised transfer learning and decision-level fusion. *International Journal of Remote Sensing*, 43(1), 196-214.
- Zhang, Y., Zhang, J., Huang, B., et al. (2021). Single-image deraining via a recurrent memory unit network. *Knowledge-Based Systems*, 218, Article 106832.
- Zheng, X., Liao, Y., Guo, W., et al. (2013). Single-image-based rain and snow removal using multi-guided filter[C]// International Conference on Neural Information Processing (pp. 258-265). Berlin, Heidelberg: Springer.
- Zhou X, Wang D, Krähenbühl P. Objects as points. arXiv preprint arXiv:1904.07850, 2019.
- Zhu, B., Sun, Y., Zhao, J., et al. (2021). Millimeter-Wave Radar in-the-Loop Testing for Intelligent Vehicles. *IEEE Transactions on Intelligent Transportation Systems*.