symmetry

MDPI

Article

Semi-RainGAN: A Semisupervised Coarse-to-Fine Guided Generative Adversarial Network for Mixture of Rain Removal

Rongwei Yu, Ni Shu *, Peihao Zhang and Yizhan Li

Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan 430070, China; roewe.yu@whu.edu.cn (R.Y.); peihaozhang@whu.edu.cn (P.Z.); downsivan@whu.edu.cn (Y.L.)
* Correspondence: niishu@whu.edu.cn

Abstract: Images taken in various real-world scenarios meet the symmetrical goal of simultaneously removing foreground rain-induced occlusions and restoring the background details. This inspires us to remember the principle of symmetry; real-world rain is a mixture of rain streaks and rainy haze and degrades the visual quality of the background. Current efforts formulate image rain streak removal and rainy haze removal as separate models, which disrupts the symmetrical characteristics of real-world rain and background, leading to significant performance degradation. To achieve this symmetrical balance, we propose a novel semisupervised coarse-to-fine guided generative adversarial network (Semi-RainGAN) for the mixture of rain removal. Beyond existing wisdom, Semi-RainGAN is a joint learning paradigm of the mixture of rain removal and attention and depth estimation. Additionally, it introduces a coarse-to-fine guidance mechanism that effectively fuses estimated image, attention, and depth features. This mechanism enables us to achieve symmetrically high-quality rain removal while preserving fine-grained details. To bridge the gap between synthetic and real-world rain, Semi-RainGAN makes full use of unpaired real-world rainy and clean images, enhancing its generalization to real-world scenarios. Extensive experiments on both synthetic and real-world rain datasets demonstrate clear visual and numerical improvements of Semi-RainGAN over sixteen state-of-the-art models.

Keywords: mixture of rain removal; semisupervised learning; coarse-to-fine guidance mechanism

check for updates

Citation: Yu, R.; Shu, N.; Zhang, P.; Li, Y. Semi-RainGAN: A Semisupervised Coarse-to-Fine Guided Generative Adversarial Network for Mixture of Rain Removal. *Symmetry* **2023**, *15*, 1832. https://doi.org/10.3390/sym15101832

Academic Editor: Alice Miller

Received: 1 September 2023
Revised: 20 September 2023
Accepted: 23 September 2023
Published: 27 September 2023

1. Introduction

Rain is one of the most common weather phenomena that often brings a series of visibility impairments, including blurred background scenes, occlusion of perceived objects, and distortion of image color. These impairments inevitably hamper the performance of various computer vision tasks, such as self-driving [1], traffic surveillance [2], and road sign recognition [3]. Therefore, rain removal from images has emerged as a crucial task in the computer vision community.

Rain removal from images aims to restore a high-quality rain-free image from its intricate entanglement with rain. To resolve such an ill-posed and challenging problem, early image rain removal wisdom employs a range of hand-crafted priors, such as sparse coding [4], the Gaussian mixture model [5], and low-rank representation [6,7], to separate the rain layer from the background layer. However, these prior-based approaches have limited ability in terms of rain streak representation when dealing with complex rainy scenes, such as rain streaks with various densities, directions, and sizes.

Recently, convolutional neural networks (CNNs) have enabled significant advancements in the task of removing rain from images [8–10], owing to their powerful representation capabilities. Despite the potential of these deep-learning-based approaches, they often face degraded performance when dealing with real-world rainy scenes, mainly due

to three factors. First, it is observed that real-world rain is a mixture of rain streaks and rainy haze. Mathematically, a rainy image ($O(x)$) at pixel x can be modeled as

$$O(x) = B(x)(1 - R(x) - A(x)) + R(x) + \alpha A(x) \quad (1)$$

where $B(x)$, $R(x)$, and $A(x)$ denote the clean background image, rain streak layer, and rainy haze layer, respectively, and α signifies the global atmospheric light. However, most existing learning-based approaches tend to disregard the mixture of rain by focusing only on either rain streak removal or rainy haze removal (as shown in Figure 1b–e,g). Second, these approaches employ no guidance mechanisms or coarse guidance mechanisms to focus on local rain regions with limited receptive fields, leading to suboptimal rain removal performance and loss of image details (as shown in Figure 1b–g). Third, models are only trained on paired synthetic data, which cannot adequately capture the intricate characteristics of real-world rain. The distribution shift between synthetic and real-world rainy images leads to poor generalization of real-world rain removal (as shown in Figure 1b,c,f,g).


(a) Rainy image
(b) MSPFN
(c) MPRNet
(d) Syn2Real
(e) JRGR
(f) DGNL-Net
(g) SPA-Net
(h) Ours

Figure 1. Results of removal of rain a real-world rainy image. (**a**) Real-world rainy image; (**b**,**c**) the supervised learning results of MSPFN [11] and MPRNet [12]; (**d**,**e**) the semisupervised learning results of Syn2Real [13] and JRGR [14]; (**f**) the depth-guided results of DGNL-Net [15]; (**g**) the attention-guided results of SPA-Net [8]; (**h**) the results of our Semi-RainGAN. Notably, Semi-RainGAN excels in preserving background details and effectively removing the mixture of rain.

To deal with these issues, we propose a novel semisupervised coarse-to-fine guided generative adversarial network (Semi-RainGAN) for the removal of the mixture of rain, which achieves a symmetrical balance between rain removal and detail restoration. We observe that recent approaches attempt to generate promising results with the guidance of an attention map [8,11] or depth map [15,16]. The former, focusing on diverse characteristics of rain streaks (e.g., shapes, directions, densities, etc.), fails to remove thick rainy haze (as shown in Figure 1g), whereas the latter, focusing on different rainy haze densities based on corresponding scene depths, struggles to remove multiscale rain streaks (as shown in Figure 1f). This motivates us to raise an intriguing question: **Can we combine both a coarse attention-guided manner and a depth-guided manner to build a coarse-to-fine guidance mechanism?** To answer this question, Semi-RainGAN employs three subnetworks to jointly predict the attention and depth maps, then effectively fuse the predicted image, attention, and depth features to remove the mixture of rain in a coarse-to-fine guided manner. Furthermore, it leverages both synthetic and real-world rainy images to capture the intricate properties of real-world rain, thus improving the generalization ability in real-world rainy scenes. Extensive experiments demonstrate

that Semi-RainGAN outperforms sixteen state-of-the-art models on both synthetic and real-world rainy images.

The main contributions of this paper can be summarized as follows:

- We propose a novel semisupervised coarse-to-fine guided generative adversarial network, dubbed Semi-RainGAN, to remove the mixture of rain. Semi-RainGAN leverages both synthetic (paired) and real-world (unpaired) rainy images for training, boosting the generalization ability on real-world rainy images.
- We propose two parallel subnetworks, i.e., a multiscale attention prediction network (MAPN) to fully exploit complementary multiscale information for attention map prediction and a global depth prediction network (GDPN) for accurate depth map prediction. These predicted attention and depth maps guide Semi-RainGAN to remove entangled rain streaks and rainy haze.
- We propose a coarse-to-fine guided rain removal network (CFRN) to integrate the predicted image features with estimated depth and attention features. This subnetwork is connected with the first two subnetworks in a cascaded way and provides sufficient and robust feature fusion to generate high-quality derained images.

2. Related Work

2.1. Rain Streaks Removal

The removal of rain from images has drawn considerable attention within the low-level computer vision community. Several schemes have been proposed to address this problem [7,11,13,17], including rain removal methods for videos [18,19] and single-image rain removal methods [20,21]. Comparatively, single-image rain removal is more challenging than video rain removal due to the lack of temporal information. We mainly focus on single-image rain removal in this work.

Traditional Methods: Early approaches [4,6,7,22,23] exploit various hand-crafted priors for single-image rain removal based on low-level image statistics. Kang et al. [22] first proposed the problem of single-image rain removal. They decomposed the rainy image into low- and high-frequency layers and performed dictionary learning and sparse coding on the high-frequency layer to eliminate the rain streaks. However, this may inadvertently eliminate certain portions of the non-rain elements that share similar gradient directions with the rain component. Chen [6] introduced a single-objective function designed for decomposition of the background and rain streak layers. While the concept of formulating the problem as an objective function is compelling, it is worth noting that the applied constraints may not be robust enough to fully address the issue. Later, Luo et al. [4] introduced a non-linear screen blend model to model rain images and adopted discriminative sparse coding to separate the rain layer. Unlike the above methods, Li et al. [5] employed the Gaussian mixture model to effectively capture the distribution of rain streaks and distinguish them from the background layer. However, these approaches utilize uniform model types to describe both the background and rain streak components, leading to a requirement for external data to train distinct dictionaries or Gaussian mixture models (GMMs) for these individual layers. More recently, Gu et al. [23] combined analysis of sparse representation and synthesis of sparse representation to remove more rain streaks and maintain details in the background layer. Despite their successes, these prior-based methods are limited in their ability to model complex rain streak characteristics, rendering them ineffective in handling intricate rainy scenes.

Deep-Learning-based Methods: Driven by the surge of deep neural networks, several deep-learning-based methods [10,14,21,24] have been proposed for single-image rain removal. Fu et al. [25] first introduced a deep learning network for rain removal, which was further improved by reducing the mapping range [9]. Many rain removal approaches have since been proposed in an attempt to devise advanced networks to improve rain removal performance, such as dense blocks [26,27], residual blocks [11,15], recurrent networks [28], and generative adversarial networks [29]. However, these approaches heavily rely on paired rainy/clean images, which limits their ability to generalize to real-world rain removal.

To rectify this weakness, Wei et al. [30] first proposed a semisupervised transfer learning framework to solve the problem of single-image rain removal. They employed a likelihood term imposed on the Gaussian mixture model and Kullback–Leibler divergence to optimize the unsupervised learning branch. More semisupervised approaches [13,14,21,31] have since been proposed for real-world removal of rain from images. Specifically, Syn2Real [13] utilizes a Gaussian process to model the intermediate latent space of rain and creates pseudolabels to supervise the unlabeled data. Huang et al. [21] presented a self-supervised memory module and a self-training mechanism to improve the semisupervised image deraining. Wei et al. [31] proposed a semisupervised single-image rain removal network based on CycleGAN [32], which has impressive generalization power in real scenes. Despite the strong performance exhibited by these semisupervised approaches in specific real-world rainy images, they tend to focus on removing rain streaks while neglecting the generated rainy haze, leading to degraded performance when confronted with real-world rainy images containing a mixture of rain.

2.2. Rain Streaks and Rainy Haze Removal

Efforts towards rainy haze removal [33,34] were originally separate from rain streak removal until Li et al. [35] observed that rain accumulation can produce haze effects and result in deeper scenes appearing increasingly blurry. To remove both rain streaks and rainy haze, they decomposed the rainy image into high-frequency and low-frequency components, which allows for the estimation of rain streaks, transmission maps, and atmospheric light. They also used a depth-guided GAN to recover the background details. Additional approaches [15,16,36,37] have been proposed to address the mixture of rain. For instance, Wang et al. [36] observed that rain streaks and rainy haze are intricately connected, while current rain image generation models fails to accurately model this property. Thus, they rethought rain image formation by formulating both rain streaks and rainy haze as a transmission medium to better eliminate the mixture of rain. Hu et al. [15,16] developed a rain imaging process based on the visual effects of rain in relation to scene depth and presented a depth-guided network to generate a rain-free image. Later, MBA-RainGAN [37], a multibranch attention generative adversarial network, was proposed to remove entangled rain streaks, rainy haze, and raindrops. However, these methods are only trained on synthetic rainy images and fall short in predicting an accurate depth map, thereby limiting their performance on real-world rainy images.

3. Proposed Method

Beyond previous wisdom, we propose an effective semisupervised coarse-to-fine guided generative adversarial network (Semi-RainGAN) for removal of the mixture of rain, as illustrated in Figure 2. Semi-RainGAN consists of a supervised branch and an unsupervised branch, following [31,38], both of which share the weights during training. Semi-RainGAN comprises three generators (G_L, G_U, and G'_U) and two discriminators (D_L and D_U). G_L and G_U are responsible for mapping synthetic and real-world rainy images to rain-free images, respectively. Conversely, G'_U is used to reconstruct rainy images from rain-free images. To make the final derained images more realistic, we use two discriminators (D_L and D_U) to ensure that the generated images have feature distributions similar to those of the target domain. Since the structures of G_L and D_L are identical to those of G_U and G'_U and D_U, respectively, we only present the structures of G_L and D_L (as shown in Figure 3 and Figure 4, respectively).

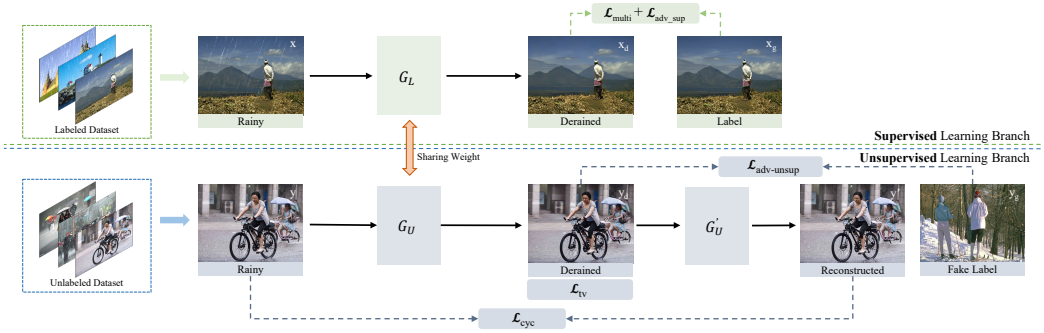


Figure 2. The pipeline of Semi-RainGAN. Semi-RainGAN consists of a supervised branch and an unsupervised branch. Concretely, x and y are synthetic and real-world rainy images, respectively; x_d and y_d are derained images from G_L and G_U, respectively; x_g is the corresponding ground truth of x; y' is a reconstructed rainy image from the generator (G'_U); and y_g is an unpaired rain-free image randomly selected from synthetic datasets as the fake label of y.

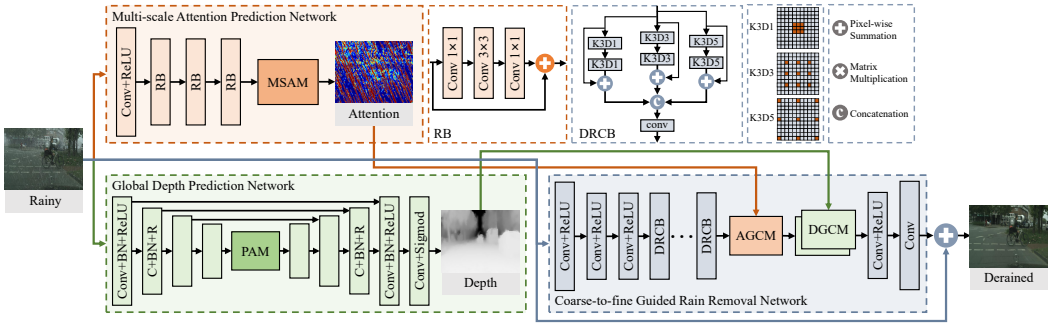


Figure 3. The overall architecture of the generator. It consists of three subnetworks for removal of rain from images, i.e., a multiscale attention prediction network (MAPN), global depth prediction network (GDPN), and coarse-to-fine guided rain removal network (CFRN).

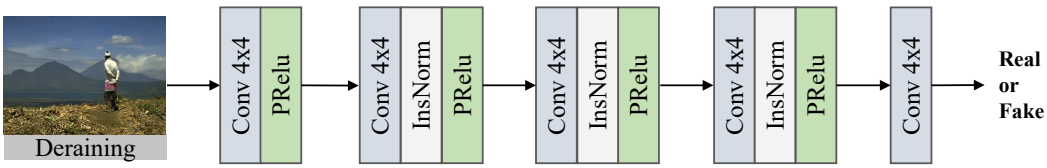


Figure 4. Details of the discriminator.

3.1. Generator

The generator takes in the rainy image and produces a clear image using three subnetworks: (1) a multiscale attention prediction network (MAPN) for attention map estimation, (2) a global depth prediction network (GDPN) for depth map estimation, and (3) a coarse-to-fine guided rain removal network (CFRN), which reconstructs a clean and rain-free image generated using the predicted attention map and depth map as guidance. Figure 3 provides the comprehensive architecture of the generator.

3.1.1. Multiscale Attention Prediction Network

In real-world rainy scenarios, multiscale rain streaks are randomly distributed across the image. However, existing approaches usually neglect this scale-specific knowledge, failing to capture the correlations of rain across different scales. To tackle this issue, we construct a multiscale attention prediction network (MAPN) to estimate attention maps.

Our developed multiscale spatial attention module (MSAM) allows MAPN to excavate the inherent multiscale correlation of rain streaks. As illustrated in Figure 3, the MAPN first leverages three standard residual blocks (RB) [39] to extract features. Then, the extracted features are passed through the MSAM to obtain the final attention map, which successfully models the distribution of multiscale rain.

Multiscale spatial attention module. Capturing multiscale complementary rain information contributes to more powerful feature representations, which can ameliorate the performance of single-image rain removal. In light of this, we incorporate pyramid pooling into our MSAM, which is commonly applied to acquire relationships from multiscale features across many computer vision tasks [40].

As depicted in Figure 5, we start by adopting an average pooling operation to the input, which downsamples it into four feature maps of different scales. To model rain characteristics across various scales, we incorporate a spatial attention module (SAM) to capture long-range contextual information from the entire feature map. Specifically, we integrate a two-round, four-directional IRNN into our SAM. Figure 6 shows the process of gathering global contextual information using a two-round, four-directional IRNN. Specifically, the first IRNN creates an intermediate feature map that summarizes the spatial contexts along the four principal directions. The second IRNN then produces a global feature map by collecting non-local contextual information. $h_{i,j}$ denotes the feature located at pixel (i, j), and the IRNN operation to the right (as well as for the other directions) at (i, j) can be computed as:

$$h_{i,j} = max(\alpha_{right} h_{i,j-1} + h_{i,j}, 0) \tag{2}$$

where α_{right} is the weight parameter in the recurrent convolution layer for the right direction. α_{right} and weights for the other directions are initialized as a learned identity matrix. We repeat the IRNN operation n times each round, where n corresponds to the width of the input feature map. Moreover, we observe that the low-resolution map may lose the high-level information in the pyramid pooling module. To address this, we introduce an additional branch to explore the complementary spatial information.

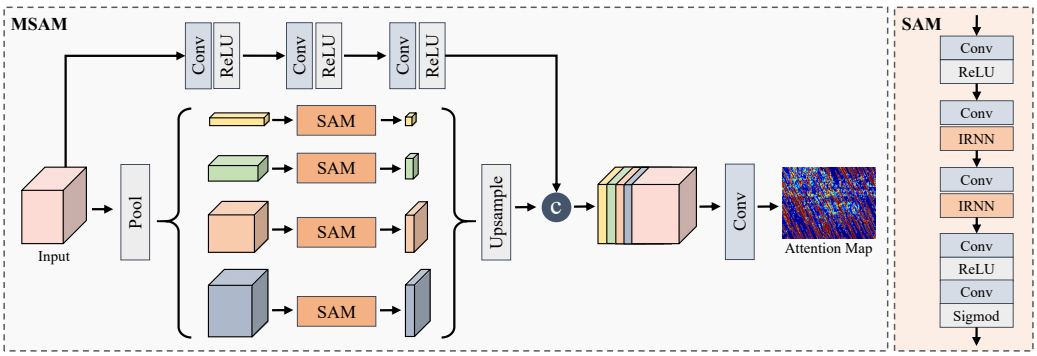


Figure 5. The structure of MSAM. We first apply an average pooling operation to downsample the input into multiscale representations. Then, we use four spatial attention modules (SAM) to generate multiscale attention maps, followed by an upsampling operation to obtain the same size as the input. Another branch leverages the convolution layers and non-linear functions to extract complementary features. We fuse the multiscale attention maps and complementary features through a concatenation layer, and the fused feature representation is fed into a convolution layer to obtain the final attention map. The SAM consists of five convolution layers, two rectified linear units, a sigmoid function, and two identity matrix recurrent neural networks for attention regression.

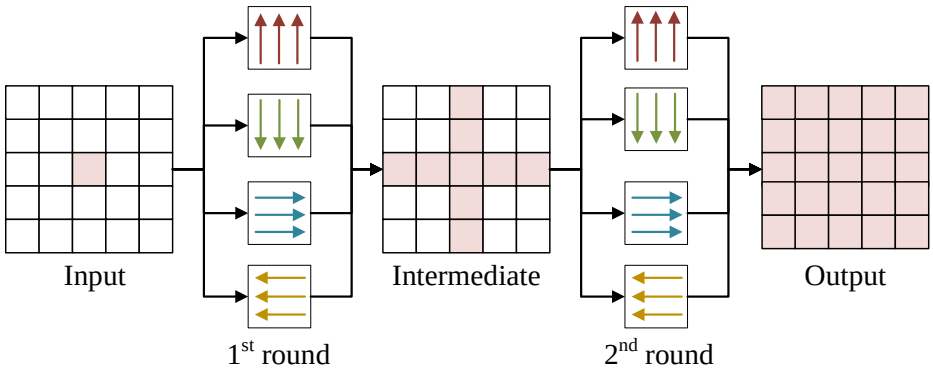


Figure 6. The process of gathering global contextual information with a two-round, four-directional IRNN. In the first round, we adopt four-directional recurrent convolutional layers for each position in the input feature map to capture horizontal and vertical neighborhood information. In the second round, we acquire the global contextual information of the input feature map by repeating the previous operations.

3.1.2. Global Depth Prediction Network

Images captured in real-world rainy scenes are often obscured by a mixture of rain streaks and rainy haze. The transformation from rain streaks into rainy haze depends on the scene depth. Based on this fact, it is reasonable to provide additional depth estimation to guide the rain removal model to remove the mixture of rain. To this end, we propose a global depth prediction network (GDPN) for depth map estimation. We incorporate a position attention module [41] (PAM) into the GDPN to aggregate the global contextual information. This module enhances the representation capability of the predicted depth map because it focuses more on informative pixels, such as thick hazy regions.

As depicted in Figure 3, GDPN is an encoder–decoder network with skip connections designed to generate a depth map. Concretely, GDPN consists of eight blocks for feature extraction, each including a convolutional layer, a batch normalization, and a rectified linear unit. We then input the extracted features into a convolutional layer and a sigmoid function layer to generate the final depth map. Moreover, we adopt a PAM between the encoder and decoder to aggregate features at each position, which helps to regress a more accurate depth map. Figure 7 shows the detailed structure of the position attention module. It takes the feature (F_E) extracted from the encoder as an input and feeds it into the convolutional layers to generate the query (Q), the key (K), and the value (V), where $Q, K, V \in \mathbb{R}^{H \times W \times C}$. The output of the position attention module can be computed by the following formula:

$$P_{out} = \alpha SV + F_E \tag{3}$$

$$S = softmax(QK^T) \tag{4}$$

where $S \in \mathbb{R}^{(H \times W) \times (H \times W)}$ is the spatial attention map, which measures the similarity between any two positions of the feature (F_E); α is a weight parameter that gradually increases from 0 to a higher value during the learning process; and $P_{out} \in \mathbb{R}^{H \times W \times C}$ is the weighted sum of the spatial attention map and the original features, thereby selectively aggregating contexts across all positions to obtain an accurate depth prediction. With the help of the PAM, our GDPN provides a more accurate depth map compared to DGNL-Net [15], as shown in Figure 8.

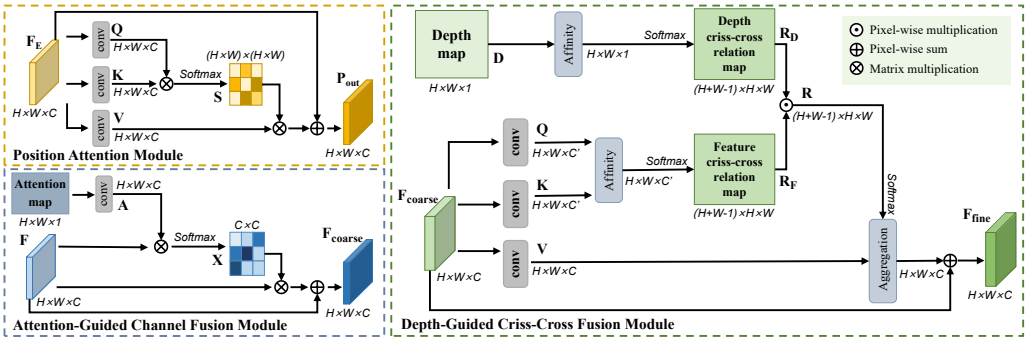
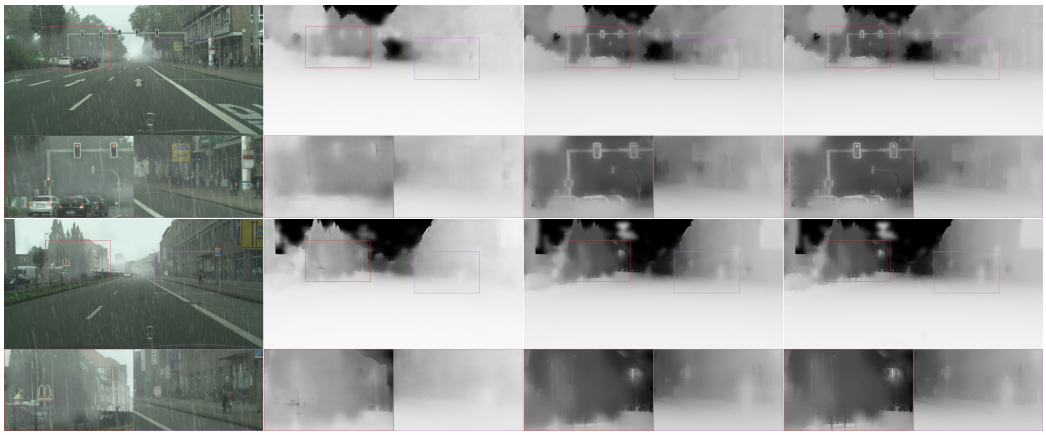


Figure 7. The detailed structures of the position attention module (PAM), attention-guided channel fusion module (AGCM), and depth-guided crisscross fusion module (DGCM).



(a) Rainy image (b) DGNL-Net (c) Ours (d) GT

Figure 8. Visualization of the estimated depth maps from DGNL-Net [15], Semi-RainGAN, and ground truth.

3.1.3. Coarse-to-Fine Guided Rain Removal Network

In order to enhance the quality of rain removal by effectively merging image features with predicted attention and depth, we introduce a coarse-to-fine guided rain removal network (CFRN) as shown in Figure 3. Unlike previous approaches, CFRN combines two coarse guidance mechanisms (attention-guided and depth-guided) to generate clean, rain-free images in a coarse-to-fine manner.

As depicted in Figure 3, the first two convolutional layers are used to reduce the resolution of the feature map and increase its number of channels. Then, CFRN employs five dilated residual concatenation blocks (DRCBs) to acquire long-range contextual information. Each DRCB consists of three branches, each of which contains a dilated residual block with dilation rates of 1, 3, and 5, respectively. The outputs of these branches are concatenated, and a 1×1 convolution layer is used to reduce the feature dimensions. Moreover, we present an attention-guided channel fusion module (AGCM) and a depth-guided crisscross fusion module (DGCM) to integrate the attention map and depth map, respectively. Finally, the last two convolution layers upsample the feature map to the size of the input image and obtain the final rain-free image.

Attention-guided channel fusion module. To handle multiscale rain streaks, we propose an attention-guided channel fusion module (AGCM) to fuse the feature map with the estimated attention map, which contains multiscale rain characteristic information. With the help of AGCM, we can improve the discriminative learning ability of the model via typical scale-specific knowledge. As depicted in Figure 7, AGCM takes the attention map and feature map ($F \in \mathbb{R}^{H \times W \times C}$) as inputs. We first input the attention map into a convolutional layer to generate $A \in \mathbb{R}^{H \times W \times C}$. We compute the fused features using the following formula:

$$F_{coarse} = \beta FX + F \tag{5}$$

$$X = softmax(A^T F) \tag{6}$$

where $X \in \mathbb{R}^{C \times C}$ computes the correlation between the attention map and the feature map. The weight (β) gradually learns from 0 to balance the original feature and the fused feature. The output ($F_{coarse} \in \mathbb{R}^{H \times W \times C}$) successfully integrates the attention map into the feature map.

Depth-guided crisscross fusion module. Although previous studies [8] have verified the efficacy of the attention-guided manner for rain streak removal, they failed to handle the mixture of rain, especially thick rainy haze. Therefore, in accordance with the transformation process from rain streaks to rainy haze, we designed a depth-guided crisscross fusion module (DGCM) to further fuse the coarse fused image features with the depth map for better rain removal. Motivated by [42], we extend the crisscross attention module to our DGCM, which globally correlates the depth map and coarse fused feature map for enhanced fusion.

The detailed structure of the DGCM is illustrated in Figure 7. Concretely, the DGCM takes the depth map ($D \in \mathbb{R}^{H \times W \times 1}$) and feature map ($F_{coarse} \in \mathbb{R}^{H \times W \times C}$) as inputs. We adopt the affinity operation and softmax normalization to generate the relation map ($R_D \in \mathbb{R}^{(H+W-1) \times (H \times W)}$) between each pixel in the same row or column pixel over the depth map (D). Meanwhile, we measure the feature relation map between each position in the same row or column one. Specifically, we apply two 1×1 convolution layers to the input (F) to reduce the computation and memory overhead. Then, we obtain two feature maps (Q and K, where $Q, K \in \mathbb{R}^{H \times W \times C'}$ and C' is less than C). For each position (m) in the spatial dimension of Q, we obtain a vector ($Q_m \in \mathbb{R}^{C'}$, where $m \in \{1 \ldots H \times W\}$). Correspondingly, by extracting feature vectors from K that are in the same row or column as location m, we can obtain the set $\Theta_m \in \mathbb{R}^{(H+W-1) \times C'}$. Θ_m consists of the vector $\Theta_{i,m} \in \mathbb{R}^{C'}$, where $i \in \{1 \ldots H + W - 1\}$. For any $i \in \{1 \ldots H + W - 1\}$, we compute the degree of correlation between features Q_m and $\Theta_{i,m}$ as $R_{F_{i,m}}$ via the affinity operation and softmax normalization:

$$R_{F_{i,m}} = softmax(Q_m \Theta_{i,m}^T) \tag{7}$$

where $R_{F_{i,m}} \in R_F$, and $R_F \in \mathbb{R}^{(H+W-1) \times (H \times W)}$. Then, we utilize element-wise multiplication and a softmax layer to model the inter-relationship R between the spatial locations of R_F and R_D.

$$R = softmax(R_F \odot R_D) \tag{8}$$

where $\odot$ denotes pixel-wise multiplication and $R \in \mathbb{R}^{(H+W-1) \times (H \times W)}$.

In another branch, a 1×1 convolution layer is used on F_{coarse} to generate $V \in \mathbb{R}^{W \times H \times C}$ for feature adaptation. For each position (m) in the spatial dimension of V, we can obtain a vector ($V_m \in \mathbb{R}^C$) and a set ($\Phi_m \in \mathbb{R}^{(H+W-1) \times C}$, where $m \in \{1 \ldots H \times W\}$). Set Φ_m consists of feature vectors ($\Phi_{i,m}$) in the same row or column as position m. Then, we employ an aggregation operation on R and Φ_m to collect contextual information, which is added to the original coarse feature (F_{coarse}) to obtain the final fine fused features (F_{fine}).

$$F_{fine}^m = \sum_{i=0}^{H+W-1} R_{i,m} \Phi_{i,m} + F_{coarse}^m \tag{9}$$

where $R_{i,m}$ is a scalar value at channel i and position m in R, and F_{fine}^m and F_{coarse}^m are feature vectors at position m in $F_{fine} \in \mathbb{R}^{H \times W \times C}$ and $F_{coarse} \in \mathbb{R}^{H \times W \times C}$, respectively. Since the single crisscross module can only obtain information in horizontal and vertical directions, we simply repeat two depth-guided crisscross fusion modules to capture the dense contextual information from all pixels. The second DGCM takes the depth map ($D \in \mathbb{R}^{H \times W \times 1}$) and $F_{fine} \in \mathbb{R}^{H \times W \times C}$ as inputs.

3.2. Discriminators

The discriminators [43], namely D_L and D_U, serve to distinguish whether the accepted images are real rain-free images or faked. As illustrated in Figure 4, the discriminator comprises five convolution layers, three instance normalization layers, and four parametric rectified linear units, which guide the generator to generate more realistic rain-free images.

3.3. Comprehensive Loss Function

We undertake a comprehensive consideration of both supervised and unsupervised loss functions to enhance the performance of the mixture of rain removal. The overall loss function comprises multitask loss, adversarial loss, cycle consistency loss, and total variation loss, which can be expressed as follows:

$$L_{total} = \lambda_1 L_{multi} + \lambda_2 L_{adv\_sup} + \lambda_3 L_{adv\_unsup} + \lambda_4 L_{cyc} + \lambda_5 L_{tv} \tag{10}$$

Multitask loss. The multitask loss function is used to jointly optimize the process of rain removal and depth prediction. Specifically, we minimize the $\ell 1$ distance between the derained image and depth map and their corresponding ground truth. This can be computed as follows:

$$L_{multi} = \left\| G_L(x) - x_g \right\|_1 + \left\| D(x) - d_g \right\|_1 \tag{11}$$

where x refers to synthetic rainy images, $G_L(x)$ denotes the derained images generated by the generator (G_L), and $D(x)$ represents the depth maps predicted by the GDPN. Furthermore, x_g and d_g are ground-truth rain-free images and ground truth depth maps, respectively.

Adversarial loss. We develop two discriminators to enhance the clarity and realism of the final derained images ($G_L(x)$ and $G_U(y)$). In this work, a least squares GAN (LSGAN) [44] is utilized to compute the adversarial loss, which can be expressed as:

$$L_{adv\_sup}(G_L) = E_{G_L(x) \sim P_{fake}}[(D_L(G_L(x)) - 1)^2] \tag{12}$$

$$L_{adv\_sup}(D_L) = E_{x_g \sim P_{real}}[(D_L(x_g) - 1)^2] + E_{G_L(x) \sim P_{fake}}[(D_L(G_L(x)))^2] \tag{13}$$

$$L_{adv\_unsup}(G_U) = E_{G_U(y) \sim P_{fake}}[(D_U(G_U(y)) - 1)^2] \tag{14}$$

$$L_{adv\_unsup}(D_U) = E_{y_g \sim P_{real}}[(D_U(y_g) - 1)^2] + E_{G_U(y) \sim P_{fake}}[(D_U(G_U(y)))^2] \tag{15}$$

where x and x_g represent synthetic rainy images and corresponding ground truth, respectively, and y refers to real-world rainy images. In the unsupervised branch, the real-world rainy images have no corresponding rain-free images as labels for training. Thus, following the protocol described in [31], we select clean images from synthetic datasets as fake labels to constrain the training of the unsupervised branch.

Cycle consistency loss. The cycle consistency loss [32] aims to ensure that the feature distributions of derained images approximate those of the clean real-world images. We adopt it in the unsupervised branch to promote similarity between the reconstructed rainy images and the corresponding original rainy images:

$$L_{cyc} = \| G'_U(y_d) - y \|_1 \tag{16}$$

where y and y_d represent the original real-world rainy images and corresponding derained images, respectively, and $G'_U(y_d)$ denotes the reconstructed rainy images.

Total variation loss. The total variation loss [45] is an $\ell 1$ regularization gradient prior that is employed in the unsupervised branch to maintain structures and details of the derained images:

$$L_{tv} = \|\nabla_h(G_U(y)) + \nabla_v(G_U(y)))\|_1 \tag{17}$$

where ∇_h and ∇_v refer to the horizontal and vertical differential operation matrices, respectively.

4. Experimental Results and Analysis

4.1. Experimental Settings

4.1.1. Datasets

We conducted extensive experiments on both synthetic and real-world datasets to evaluate our method.

Synthetic datasets: We consider three paired synthetic benchmark datasets: (1) the Rain200L dataset [46], which contains 1800 synthetic images for training and 200 synthetic images for testing, with a single type of rain streak; (2) the Rain200H dataset [46], which also includes 1800 training images and 200 testing images but with rain streaks that vary in orientation, scale, and shape; and (3) the RainCityscapes dataset [15], which comprises 9432 training images and 1188 testing images and leverages the camera parameters and scene depth information to synthesize rain and fog based on the Cityscapes dataset [47].

Real-world datasets: Furthermore, we constructed a new dataset, Mix200, by collecting real-world images that contain a mixture of rain from [29,30,46] and a Google search using the term "rain and haze image". Mix200 comprises 400 real-world rainy images divided into 200 training images and 200 testing images. This dataset helps our paradigm to learn the characteristics of a mixture of rain in real scenarios.

4.1.2. Training Details

Semi-RainGAN is implemented using the Pytorch framework and trained on an NVIDIA GeForce RTX 3090 GPU. To optimize the model, we utilize the Adam optimizer [48], with a momentum value of 0.9 and weight decay of 0. The learning rates of generators and discriminators are initially set to 5×10^{-4} and 1×10^{-5}, respectively. For the hyperparameters of the loss setting, we empirically set λ_1, λ_2, λ_3, λ_4, and λ_5 to 1.0, 0.5, 0.5, 1.0, and 0.1, respectively. In the experiments, we resize the image patch to 512×1024 to train the Semi-RainGAN on the RainCityscapes dataset with a batch size of four. For training on the Rain200L and Rain200H datasets, the image patch size is 256×256. Additionally, we resize the images from the Mix200 dataset to match the image sizes of the synthetic datasets.

4.1.3. Evaluation Metrics

We evaluate the rain removal results using quantitative and qualitative measures. For quantitative evaluation, we employ the SSIM (structural similarity index) and PSNR (peak signal-to-noise ratio) metrics, which are commonly used as criteria to assess image quality in rain removal tasks. In specific terms, SSIM is utilized to assess the similarity between corresponding images in relation to aspects such as illumination, structure, and contrast. Meanwhile, PSNR calculates the peak signal-to-noise ratio in decibels between two images. In general, higher SSIM and PSNR values signify better rain removal results. Qualitative evaluations are conducted based on the visual rain removal results, such as the degree of removal of rain streaks and rainy haze, the level of detail restoration, and the extent of color distortion.

4.2. Comparison with State-of-the-Art

4.2.1. Baselines

We compare our Semi-RainGAN with sixteen rain removal methods, including rain streak removal methods and rain streaks and rainy haze removal methods. The rain streak removal methods comprise two traditional methods (DSC [4] and GMM [5]), with eight supervised learning methods (UMRL [49], SPA-Net [8], PreNet [28], DCSFN [50], MSPFN [11], MPRNet [12], DerainRLNet [51], and CCN [52]) and three semisupervised learning methods (SIRR [30], Syn2Real [13], and JRGR [14]). The rain streak and rainy

haze removal methods comprise DAF-Net [16], DGNL-Net [15], and MBA-RainGAN [37]. Additionally, we compare our method with these rain streaks removal method combined with the FFA-Net dehazing method [53]. We take the images generated by these rain streaks removal methods as inputs and further use the pretrained FFA-Net for haze removal. To ensure fair comparisons, we retrain all the supervised methods on synthetic datasets, and semisupervised methods are trained on both synthetic and real-world datasets. Since the depth maps of Rain200L and Rain200H are not available, we follow [16] and assume a constant depth value of 0.5 on the whole image.

4.2.2. Results on the RainCityscapes Dataset

We compare Semi-RainGAN with other rain removal methods on the RainCityscapes testing set. As demonstrated in Table 1, Semi-RainGAN achieves the best performance in terms of both PSNR and SSIM. Notably, compared with the second-best rain streak removal results achieved using MPRNet [12], our method achieves 4.76 dB and 0.038 gains in terms of PSNR and SSIM, respectively. Compared with depth-guided DGNL-Net [15] and attention-guided SPA-Net [8], Semi-RainGAN obtains 1.61 dB and 12.92 dB PSNR gains, respectively. We also compare our method with the semisupervised SIRR method [30] and Syn2Real [13], achieving 5.08 dB and 5.16 dB gains, respectively. Furthermore, we compare our method with rain streak removal methods [11–14] combined with the FFA-Net dehazing method [53]; the results demonstrate that our method performs better in handling the mixture of rain. Overall, our method achieves superior performance over both supervised methods and semisupervised methods. Additionally, we compare the average running time of Semi-RainGAN against that of different models on an 512×1024 image patch. It is observed that Semi-RainGAN achieves promising rain removal results with a low time cost. Figure 9 shows visual comparisons on the RainCityscapes testing set. It is observed that most of the compared methods only focus on rain streaks, failing to remove the mixture of rain, distorting the details. Although DGNL-Net can handle both rain streaks and rainy haze, some large rain streaks or artifacts remain in the generated rain-free images. Comparatively, Semi-RainGAN obtains the cleanest rain-free images with rich details.

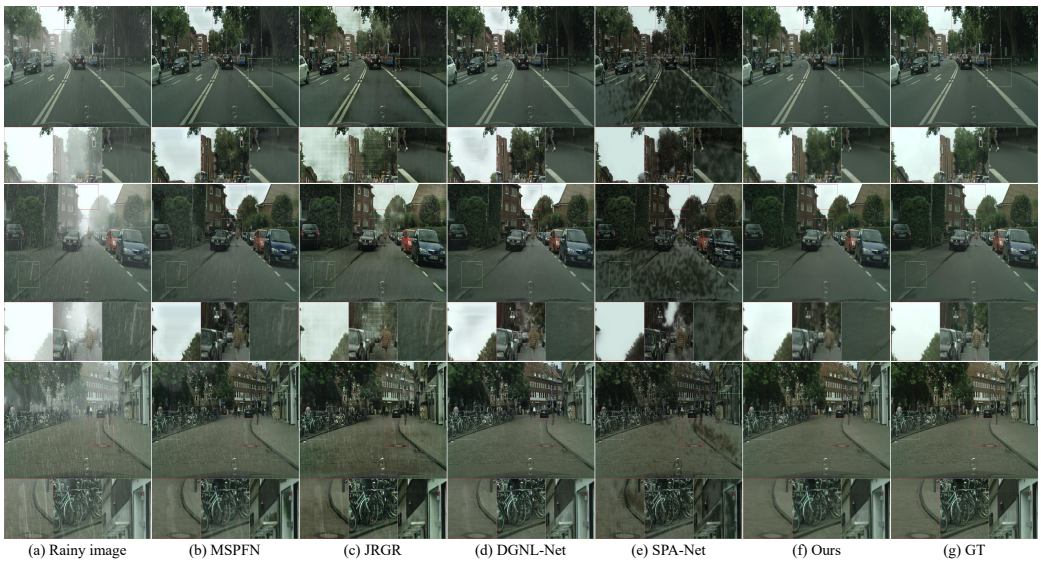


Figure 9. Results of the removal of rain removal from images on the RainCityscapes dataset. (**a**) Rainy images. Rain removal results of (**b**) MSPFN [11], (**c**) JRGR [14], (**d**) DGNL-Net [15], (**e**) SPA-Net [8], (**f**) our model, and (**g**) the ground truth.

Table 1. Quantitative results of different methods on the RainCityscapes, Rain200L, and Rain200H testing sets.

Method	RainCityscapes		Rain200L		Rain200H		Time(s)
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
Rain Streak Removal							
DSC [4]	16.41	0.771	25.68	0.875	15.29	0.423	199.5
GMM [5]	18.39	0.819	27.16	0.898	14.54	0.548	600.4
UMRL [49]	27.97	0.912	31.24	0.954	27.27	0.898	1.349
SPA-Net [8]	20.90	0.862	31.59	0.965	23.85	0.852	**0.154**
PReNet [28]	26.83	0.910	36.76	0.980	28.08	0.887	0.262
DCSFN [50]	26.37	0.872	38.21	0.982	28.26	0.899	1.524
MSPFN [11]	25.51	0.903	32.98	0.969	27.38	0.869	3.863
MPRNet [12]	29.06	0.918	37.32	0.981	28.32	0.916	3.668
DerainRLNet [51]	27.39	0.881	37.38	0.980	28.87	0.895	0.925
CCN [52]	29.34	0.950	37.01	0.982	29.12	0.921	0.518
SIRR [30]	28.74	0.920	35.32	0.968	26.21	0.813	0.281
Syn2Real [13]	28.66	0.919	34.26	0.946	25.19	0.806	1.241
JRGR [14]	23.85	0.877	30.15	0.934	22.19	0.801	0.401
Rain Streak Removal + Rainy Haze Removal							
MSPFN [11] + FFA [53]	25.56	0.906	32.98	0.969	27.40	0.869	3.927
MPRNet [12] + FFA [53]	29.10	0.920	37.33	0.981	28.33	0.917	3.733
Syn2Real [13] + FFA [53]	28.72	0.922	34.27	0.947	25.21	0.807	1.304
JRGR [14] + FFA [53]	23.89	0.879	30.14	0.934	22.21	0.802	0.465
Rain Streak Removal + Rainy Haze Removal							
DAF-Net [16]	30.66	0.924	34.07	0.964	24.65	0.860	0.209
DGNL-Net [15]	32.21	0.936	36.42	0.979	27.79	0.886	0.332
MBA-RainGAN [37]	29.51	0.917	33.51	0.948	23.73	0.854	0.377
Ours	**33.82**	**0.956**	**38.41**	**0.985**	**29.17**	**0.917**	0.346

4.2.3. Results on the Rain200L and Rain200H Datasets

We also compare Semi-RainGAN with sixteen state-of-the-art methods on the Rain200L and Rain200H testing sets. As shown in Table 1, Semi-RainGAN still outperforms other approaches on these two synthetic datasets. Specifically, when compared with the JRGR semisupervised method [14], our method achieves 8.26 dB and 6.98 dB PSNR gains on the testing sets of Rain200L and Rain200H, respectively. Furthermore, Semi-RainGAN still achieves superior performance when compared with supervised approaches. For example, Semi-RainGAN obtains 1.99 dB and 1.38 dB PSNR improvements relative to DGNL-Net [15] on the testing sets of the Rain200L and Rain200H datasets, respectively. The qualitative results obtained on the testing sets of the Rain200L and Rain200H datasets are shown in Figure 10. We observe that other existing methods fail to handle large rain streaks, whereas Semi-RainGAN can remove rain streaks across various scales, better preserving texture details.

4.2.4. Results on Real-World Rainy Images

To further verify the effectiveness of Semi-RainGAN in handling real-world scenes, we compare our method with state-of-the-art methods on real-world rainy images from the testing set of the Mix200 dataset. Figure 11 shows the results on real-world rain images. Semi-RainGAN produces better visual effects than other methods. Rain streak removal approaches, including supervised approaches [8,12] and semisupervised approaches [13], can eliminate most small rain streaks but still leave large rain streaks and rainy haze in the real-world rainy image. Moreover, the results obtained using MPRNet [12] + FFA-Net [53] still contain abundant rainy haze. DGNL-Net [15] considers both rain streaks and rainy haze removal; however, this model still fails to remove large rain steaks and thick haze. Comparatively, Semi-RainGAN is capable of capturing the characteristics of the mixture of rain to remove rainy haze and multiscale rain streaks while better preserving the structure and details of the background. Figure 12 shows the results on real-world rainy images that contain only rain streaks. Our method still achieves the best results, which demonstrates

that Semi-RainGAN can effectively handle solely rain streaks in various real-world rainy scenes and effectively preserve image details.

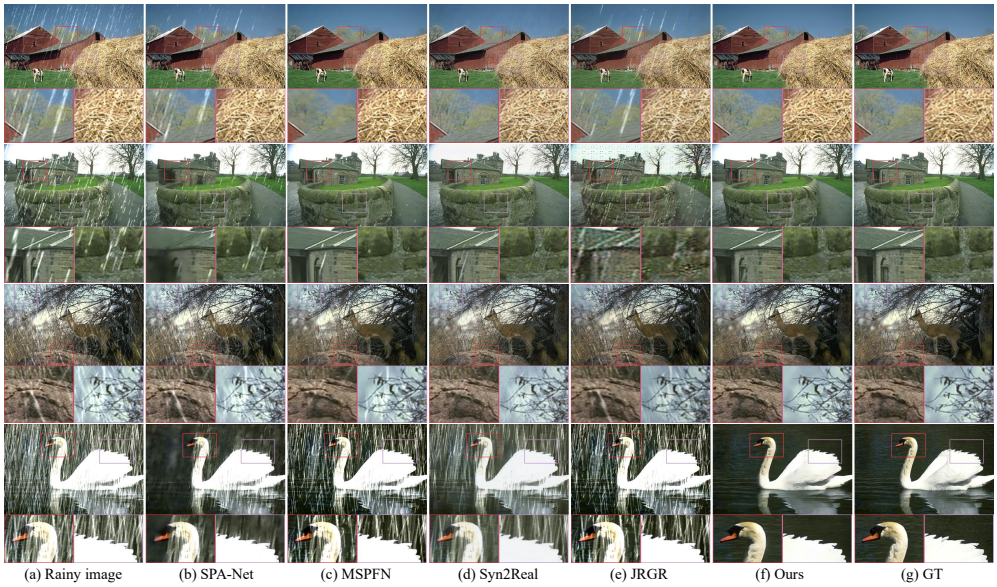


Figure 10. Qualitative results of rain removal from images in the testing sets of the Rain200L and Rain200H datasets. (**a**) Aainy images. Rain removal results of (**b**) SPA-Net [8], (**c**) MSPFN [11], (**d**) Syn2Real [13], (**e**) JRGR [14], (**f**) our method, and (**g**) the ground truth. Note that the first two rows are from the Rain200L dataset, and the last two rows are from the Rain200H dataset.

Figure 11. Qualitative results of rain removal from real-world rain images. (**a**) Rainy images. Rain removal results of (**b**) SPA-Net [8], (**c**) MPRNet [12], (**d**) MPRNet [12] + FFA-Net [53], (**e**) Syn2Real [13], (**f**) DGNL-Net [15], and (**g**) our method.

4.3. Ablation Study

4.3.1. Component Analysis

We conducted an ablation study on the testing set of RainCityscapes to evaluate the effectiveness of various components of Semi-RainGAN. These components are listed as follows:

- M1: A single rain removal network (baseline) is used for rain removal. It regresses the final rain-free images directly, without guidance from the depth map and attention map.
- M2: The attention prediction network is utilized to predict an attention map but without the multiscale attention module (MSAM). The attention-guided channel fusion module (AGCM) is substituted with a simple fusion operation that entails

matrix multiplication between the attention map and feature map, followed by the addition of the original feature map.

- M3: Only one SAM is added, and the output of the SAM is concatenated directly with another branch.
- M4: SAM is replaced with MSAM in the attention prediction network to construct the complete multiscale attention prediction network (MAPN).
- M5: The simple fusion operation is replaced with the AGCM.
- M6: The depth prediction network is added to forecast a depth map as guidance but without the position attention module (PAM). Two depth-guided crisscross fusion modules (DGCM) are substituted with a dot product operation.
- M7: The PAM is included in the depth prediction network to build the complete global depth prediction network (GDPN).
- M8: The dot product is replaced with two DGCMs.

Figure 12. Qualitative results of rain removal from real-world rainy images containing only rain streaks. (**a**) Rainy images. Rain removal results of (**b**) MSPFN [11], (**c**) Syn2Real [13], (**d**) DGNL-Net [15], and (**e**) our method.

As shown in Table 2, a single rain removal network can only make PSNR and SSIM reach 30.65 dB and 0.910, respectively. After adding the attention map, the PSNR and SSIM scores increase by 0.86 dB and 0.011, respectively. The results of M3 and M4 show that the incorporation of MSAM yields superior results compared to a single SAM, which demonstrates that four multiscale SAMs can further improve performance compared to only one SAM. Adding the AGCM also results in improvements in quantitative measures, which indicates its effectiveness. Furthermore, the PSNR and SSIM scores exhibit considerable improvements after the inclusion of the complete global depth prediction network, which proves that the depth map can provide fine guidance for better rain removal performance. Finally, the overall structure achieves the best performance, which illustrates the effectiveness of DGCM.

4.3.2. Loss Function Analysis

Apart from the analysis of different components of Semi-RainGAN, we adopt different loss function settings to verify their effectiveness. Table 3 shows the results of five loss function settings, from which we can conclude that both the PSNR and SSIM of Semi-RainGAN exhibit a gradual improvement as the loss functions are progressively integrated.

Table 2. Quantitative results of component analysis on the testing set of RainCityscapes. Note: w/o denotes "without", and M1–M8 denote the eight model settings.

Model	M1	M2	M3	M4	M5	M6	M7	M8
BL	✓	✓	✓	✓	✓	✓	✓	✓
ATT	w/o	✓	✓	✓	✓	✓	✓	✓
SAM	w/o	w/o	✓	✓	✓	✓	✓	✓
MSAM	w/o	w/o	w/o	✓	✓	✓	✓	✓
AGCM	w/o	w/o	w/o	w/o	✓	✓	✓	✓
DEPTH	w/o	w/o	w/o	w/o	w/o	✓	✓	✓
PAM	w/o	w/o	w/o	w/o	w/o	w/o	✓	✓
DGCM	w/o	w/o	w/o	w/o	w/o	w/o	w/o	✓
PSNR	30.65	31.51	31.54	31.77	31.94	32.87	33.38	33.82
SSIM	0.910	0.921	0.922	0.926	0.930	0.945	0.950	0.956

Table 3. Quantitative results of different loss function settings on the testing set of RainCityscapes. Note: w/o denotes "without", and L1–L5 denote the five loss function settings.

Loss	L1	L2	L3	L4	L5
L_{multi}	✓	✓	✓	✓	✓
L_{adv_sup}	w/o	✓	✓	✓	✓
L_{adv_unsup}	w/o	w/o	✓	✓	✓
L_{cyc}	w/o	w/o	w/o	✓	✓
L_{tv}	w/o	w/o	w/o	w/o	✓
PSNR	31.91	32.27	32.82	33.51	33.82
SSIM	0.939	0.942	0.947	0.951	0.956

4.3.3. Semisupervised Paradigm Analysis

We delve deeper into the effectiveness of the semi-supervised paradigm, that is, the capacity of our method to leverage unlabeled data of varying quantities for network training. Following the protocol proposed in [13,21], we denote the labeled synthetic training set from the Rain200H dataset as $\mathcal{D}_{\mathcal{L}}$ and the unlabeled real-world training set from the Mix200 dataset as $\mathcal{D}_{\mathcal{U}}$. Specifically, we conduct a series of experiments, where $\mathcal{D}_{\mathcal{L}}$ consists of 10%, 20%, 40%, 60%, and 100% paired synthetic images and the rest comes from $\mathcal{D}_{\mathcal{U}}$, which comprises the real-world rainy images without labels. Table 4 shows that our semisupervised deraining paradigm effectively utilizes additional unlabeled data to enhance the rain removal performance. Notably, our method employs only 60% labeled images and additional unlabeled images that can achieve comparable performance to that achieved using 100% labeled images. Compared to Syn2Real [13], our approach yields better quantitative results in both supervised and unsupervised settings. Figure 13 shows the visual results when using only 40% labeled rainy images and extra unpaired images, demonstrating that the inclusion of additional unlabeled images helps to improve the rain removal performance.

Table 4. Quantitative results of semisupervised learning analysis on the testing set of Rain200H. Note: "-" denotes that $\mathcal{D}_{\mathcal{U}}$ is 0 or the gain is 0.

$\mathcal{D}_{\mathcal{L}}$	Syn2Real						Ours					
	PSNR			SSIM			PSNR			SSIM		
	$\mathcal{D}_{\mathcal{L}}$	$\mathcal{D}_{\mathcal{L}}+\mathcal{D}_{\mathcal{U}}$	Gain	$\mathcal{D}_{\mathcal{L}}$	$\mathcal{D}_{\mathcal{L}}+\mathcal{D}_{\mathcal{U}}$	Gain	$\mathcal{D}_{\mathcal{L}}$	$\mathcal{D}_{\mathcal{L}}+\mathcal{D}_{\mathcal{U}}$	Gain	$\mathcal{D}_{\mathcal{L}}$	$\mathcal{D}_{\mathcal{L}}+\mathcal{D}_{\mathcal{U}}$	Gain
10%	22.89	23.51	0.62	0.740	0.759	0.019	26.28	27.01	0.73	0.851	0.874	0.023
20%	23.15	23.87	0.72	0.752	0.774	0.022	27.11	27.76	0.65	0.869	0.886	0.017
40%	23.80	24.59	0.79	0.770	0.791	0.021	27.73	28.36	0.63	0.885	0.903	0.018
60%	24.51	25.16	0.65	0.785	0.804	0.019	28.35	28.97	0.62	0.901	0.916	0.015
100%	25.19	-	-	0.806	-	-	29.17	-	-	0.917	-	-

(a) (b) (c) (d) (e)

Figure 13. Qualitative results on 40% labeled images from the Rain200H dataset. (**a**) Rainy images. Rain removal results of (**b**) Syn2Real on $\mathcal{D}_{\mathcal{L}}$, (**c**) Syn2Real on $\mathcal{D}_{\mathcal{L}} + \mathcal{D}_{\mathcal{U}}$, (**d**) ours method on $\mathcal{D}_{\mathcal{L}}$, and (**e**) our method on $\mathcal{D}_{\mathcal{L}} + \mathcal{D}_{\mathcal{U}}$.

4.4. Application

Images captured in rainy scenes inevitably suffer from poor visibility, which degrades the performance of high-level computer vision tasks, such as object detection [54] and semantic segmentation [40]. To assess the extent to which our approach contributes to improved object detection accuracy in rainy scenarios, we compare the accuracy of object detection in rainy images, images derained by DGNL-Net [15], and images derained by Semi-RainGAN. We adopt YOLOx [54] for object detection and retrain it on the training set of Cityscapes [47]. Specifically, we choose five categories of objects, i.e., person, bicycle, car, bus, and motorbike, for training and evaluation. Figure 14 depicts the qualitative results on the testing set of the RainCityscapes dataset. Moreover, we employ YOLOx for object detection in real rainy images, which we select from the Mix200 dataset. Figure 15 shows the qualitative results on real rainy images, which demonstrate that Semi-RainGAN can effectively achieve object detection in real-world rainy images.

(a) Rainy Image (b) DGNL-Net (c) Our

Figure 14. Object detection results on the testing set of the RainCityscapes dataset. Object detection results in (**a**) rainy images, (**b**) derained images from DGNL-Net [15], and (**c**) derained images from Semi-RainGAN.

(a) Rainy Image (b) DGNL-Net (c) Our

Figure 15. Object detection results on the testing set of the Mix200 dataset. Object detection results in (**a**) the rainy images, (**b**) derained images from DGNL-Net [15], and (**c**) derained images from Semi-RainGAN.

5. Discussion

The proposed Semi-RainGAN demonstrates promising performance in effectively removing the mixture of rain in real-world scenarios. Notably, it achieves a balance between the simultaneous removal of foreground rain streaks and the restoration of intricate background details. As discussed in the Experimental Results and Analysis section, the collaborative efforts of the multiscale attention prediction network (MAPN), the global depth prediction network (GDPN), and the coarse-to-fine guided rain removal network (CFRN) result in the production of high-quality rain-free images. This achievement is attributed to the effective fusion of image information, attention maps, and depth features. However, like any methodology, there are areas that warrant further investigation and improvement. When confronted with more challenging rainy scenarios, such as the interplay of three rain forms (rain streaks, raindrops, and rainy haze) or heavy rain, our approach may necessitate enhancement. In the future, we aim to expand our network to an all-in-one network to handle multiple adverse weather removal tasks. Furthermore, we are committed to integrating our approach with high-level computer vision techniques to advance task performance in adverse weather scenarios.

6. Conclusions

In this paper, we propose a novel semisupervised approach that leverages a coarse-to-fine guided generative adversarial network (Semi-RainGAN) for the removal of a mixture of rain. Semi-RainGAN takes advantage of both synthetic and real-world rainy images for training, which enables smooth generalization in real-world rainy scenes. One central question we addressed in this study is whether it is feasible to combine a coarse attention-guided approach and a depth-guided approach to establish a coarse-to-fine guidance mechanism. Our research unequivocally answers this question in the affirmative. The integration of these mechanisms not only boosts rain removal performance but also aligns with the symmetrical goal of preserving image details—a unique achievement in this domain. Semi-RainGAN comprises three pivotal subnetworks: a multiscale attention prediction network (MAPN), global depth prediction network (GDPN), and coarse-to-fine guided rain removal network (CFRN). These components work in concert to effectively fuse image information, attention maps, and depth features, culminating in the production of high-quality rain-free images. Extensive experiments demonstrate that Semi-RainGAN outperforms existing rain removal models in both synthetic and real-world rainy images. Despite the promising results achieved by our proposed Semi-RainGAN, there are several

areas for further research. In the future, we aspire to broaden the scope of our network to encompass a comprehensive, all-in-one framework capable of addressing multifaceted adverse weather removal challenges. Moreover, we will strive to incorporate our approach into high-level computer vision techniques, thereby propelling task performance within the domain of adverse weather scenarios.

Author Contributions: Conceptualization, N.S.; Data curation, P.Z. and Y.L.; Formal analysis, N.S.; Funding acquisition, R.Y.; Investigation, N.S.; Methodology, N.S.; Project administration, R.Y.; Resources, R.Y.; Software, N.S.; Supervision, R.Y.; Validation, N.S.; Visualization, P.Z. and Y.L.; Writing—original draft, N.S.; Writing—review and editing, P.Z. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (No.42071431); in part by the National Key Research and Development Program of China (No.2022YFB4500800 and No.2020YFB1805400).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The code used to generate the results and figures is available at https://github.com/ninishu/Semi-RainGAN (accessed on 23 September 2023). The data used to support the findings of this study are available at: https://www.icst.pku.edu.cn/struct/Projects/joint_rain_removal.html (Rain200L), https://www.icst.pku.edu.cn/struct/Projects/joint_rain_removal.html (Rain200H), https://www.cityscapes-dataset.com/downloads/ (RainCityscapes), and https://github.com/ninishu/Semi-RainGAN (Mix200).

Acknowledgments: The authors wish to thank the Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Janai, J.; Güney, F.; Behl, A.; Geiger, A. Computer vision for autonomous vehicles: Problems, datasets and state of the art. In *Foundations and Trends® in Computer Graphics and Vision*; Adobe Research: San Francisco, CA, USA, 2020; Volume 12, pp. 1–308.
2. Buch, N.; Velastin, S.A.; Orwell, J. A review of computer vision techniques for the analysis of urban traffic. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 920–939. [CrossRef]
3. Zhu, Z.; Liang, D.; Zhang, S.; Huang, X.; Li, B.; Hu, S. Traffic-sign detection and classification in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2110–2118.
4. Luo, Y.; Xu, Y.; Ji, H. Removing rain from a single image via discriminative sparse coding. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 13–16 December 2015; pp. 3397–3405.
5. Li, Y.; Tan, R.T.; Guo, X.; Lu, J.; Brown, M.S. Rain streak removal using layer priors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2736–2744.
6. Chen, Y.L.; Hsu, C.T. A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; pp. 1968–1975.
7. Zhang, H.; Patel, V.M. Convolutional sparse and low-rank coding-based rain streak removal. In Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 24–31 March 2017; IEEE: New York, NY, USA, 2017; pp. 1259–1267.
8. Wang, T.; Yang, X.; Xu, K.; Chen, S.; Zhang, Q.; Lau, R.W. Spatial attentive single-image deraining with a high quality real rain dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 12270–12279.
9. Fu, X.; Huang, J.; Zeng, D.; Huang, Y.; Ding, X.; Paisley, J. Removing rain from single images via a deep detail network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3855–3863.
10. Deng, S.; Wei, M.; Wang, J.; Feng, Y.; Liang, L.; Xie, H.; Wang, F.L.; Wang, M. Detail-recovery image deraining via context aggregation networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 14560–14569.
11. Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Huang, B.; Luo, Y.; Ma, J.; Jiang, J. Multi-scale progressive fusion network for single image deraining. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 8346–8355.

12. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H.; Shao, L. Multi-stage progressive image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 10–25 June 2021; pp. 14821–14831.
13. Yasarla, R.; Sindagi, V.A.; Patel, V.M. Syn2Real transfer learning for image deraining using Gaussian processes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 2726–2736.
14. Ye, Y.; Chang, Y.; Zhou, H.; Yan, L. Closing the Loop: Joint Rain Generation and Removal via Disentangled Image Translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 10–25 June 2021; pp. 2053–2062.
15. Hu, X.; Zhu, L.; Wang, T.; Fu, C.W.; Heng, P.A. Single-image real-time rain removal based on depth-guided non-local features. *IEEE Trans. Image Process.* **2021**, *30*, 1759–1770. [CrossRef] [PubMed]
16. Hu, X.; Fu, C.W.; Zhu, L.; Heng, P.A. Depth-attentional features for single-image rain removal. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 8022–8031.
17. Bhutto, J.A.; Zhang, R.; Rahman, Z. Symmetric Enhancement of Visual Clarity through a Multi-Scale Dilated Residual Recurrent Network Approach for Image Deraining. *Symmetry* **2023**, *15*, 1571. [CrossRef]
18. Santhaseelan, V.; Asari, V.K. Utilizing local phase information to remove rain from video. *Int. J. Comput. Vis.* **2015**, *112*, 71–89. [CrossRef]
19. Liu, J.; Yang, W.; Yang, S.; Guo, Z. Erase or fill? Deep joint recurrent rain removal and reconstruction in videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3233–3242.
20. Zhu, L.; Fu, C.W.; Lischinski, D.; Heng, P.A. Joint bi-layer optimization for single-image rain streak removal. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 27–29 October 2017; pp. 2526–2534.
21. Huang, H.; Yu, A.; He, R. Memory oriented transfer learning for semi-supervised image deraining. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 10–25 June 2021; pp. 7732–7741.
22. Kang, L.W.; Lin, C.W.; Fu, Y.H. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Trans. Image Process.* **2011**, *21*, 1742–1755. [CrossRef] [PubMed]
23. Gu, S.; Meng, D.; Zuo, W.; Zhang, L. Joint convolutional analysis and synthesis sparse representation for single image layer separation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 27–29 October 2017; pp. 1708–1716.
24. Wei, M.; Shen, Y.; Wang, Y.; Xie, H.; Wang, F.L. RainDiffusion: When Unsupervised Learning Meets Diffusion Models for Real-world Image Deraining. *arXiv* **2023**, arXiv:2301.09430.
25. Fu, X.; Huang, J.; Ding, X.; Liao, Y.; Paisley, J. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Trans. Image Process.* **2017**, *26*, 2944–2956. [CrossRef] [PubMed]
26. Li, G.; He, X.; Zhang, W.; Chang, H.; Dong, L.; Lin, L. Non-locally enhanced encoder-decoder network for single image de-raining. In Proceedings of the 26th ACM International Conference on Multimedia, Seoul, Republic of Korea, 22–26 October 2018; pp. 1056–1064.
27. Wang, G.; Sun, C.; Sowmya, A. Erl-net: Entangled representation learning for single image de-raining. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 29 October–1 November 2019; pp. 5644–5652.
28. Ren, D.; Zuo, W.; Hu, Q.; Zhu, P.; Meng, D. Progressive image deraining networks: A better and simpler baseline. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3937–3946.
29. Zhang, H.; Sindagi, V.; Patel, V.M. Image de-raining using a conditional generative adversarial network. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 3943–3956. [CrossRef]
30. Wei, W.; Meng, D.; Zhao, Q.; Xu, Z.; Wu, Y. Semi-supervised transfer learning for image rain removal. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3877–3886.
31. Wei, Y.; Zhang, Z.; Wang, Y.; Zhang, H.; Zhao, M.; Xu, M.; Wang, M. Semi-deraingan: A new semi-supervised single image deraining. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July 2021; IEEE: New York, NY, USA, 2021; pp. 1–6.
32. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 27–29 October 2017; pp. 2223–2232.
33. Qu, Y.; Chen, Y.; Huang, J.; Xie, Y. Enhanced pix2pix dehazing network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 8160–8168.
34. Wang, Y.; Yan, X.; Guan, D.; Wei, M.; Chen, Y.; Zhang, X.P.; Li, J. Cycle-snspgan: Towards real-world image dehazing via cycle spectral normalized soft likelihood estimation patch gan. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 20368–20382. [CrossRef]
35. Li, R.; Cheong, L.F.; Tan, R.T. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1633–1642.
36. Wang, Y.; Song, Y.; Ma, C.; Zeng, B. Rethinking image deraining via rain streaks and vapors. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: New York, NY, USA, 2020; pp. 367–382.
37. Shen, Y.; Feng, Y.; Wang, W.; Liang, D.; Qin, J.; Xie, H.; Wei, M. MBA-RainGAN: A Multi-Branch Attention Generative Adversarial Network for Mixture of Rain Removal. In Proceedings of the ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 22–27 May 2022; IEEE: New York, NY, USA, 2022; pp. 3418–3422.

38. Li, L.; Dong, Y.; Ren, W.; Pan, J.; Gao, C.; Sang, N.; Yang, M.H. Semi-supervised image dehazing. *IEEE Trans. Image Process.* **2019**, *29*, 2766–2779. [CrossRef] [PubMed]
39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 July 2016; pp. 770–778.
40. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
41. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3146–3154.
42. Huang, Z.; Wang, X.; Huang, L.; Huang, C.; Wei, Y.; Liu, W. Ccnet: Criss-cross attention for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 29 October–1 November 2019; pp. 603–612.
43. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
44. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.; Wang, Z.; Paul Smolley, S. Least squares generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 27–29 October 2017; pp. 2794–2802.
45. Aly, H.A.; Dubois, E. Image up-sampling using total-variation regularization with a new observation model. *IEEE Trans. Image Process.* **2005**, *14*, 1647–1659. [CrossRef] [PubMed]
46. Yang, W.; Tan, R.T.; Feng, J.; Liu, J.; Guo, Z.; Yan, S. Deep joint rain detection and removal from a single image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1357–1366.
47. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3213–3223.
48. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
49. Yasarla, R.; Patel, V.M. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 8405–8414.
50. Wang, C.; Xing, X.; Wu, Y.; Su, Z.; Chen, J. Dcsfn: Deep cross-scale fusion network for single image rain removal. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 1643–1651.
51. Chen, C.; Li, H. Robust representation learning with feedback for single image deraining. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 10–25 June 2021; pp. 7742–7751.
52. Quan, R.; Yu, X.; Liang, Y.; Yang, Y. Removing raindrops and rain streaks in one go. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 10–25 June 2021; pp. 9147–9156.
53. Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; Jia, H. FFA-Net: Feature fusion attention network for single image dehazing. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 11908–11915.
54. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.