



Article **Progressive Hybrid-Modulated Network for Single Image Deraining**

Xiaoyuan Yu ¹, Guidong Zhang ², Fei Tan ¹, Fengguo Li ¹ and Wei Xie ^{3,*}

- ¹ School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou 510006, China
- ² School of Automation, Guangdong University of Technology, Guangzhou 510006, China
- ³ School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China
- * Correspondence: weixie@scut.edu.cn

Abstract: Rainy degeneration damages an image's visual effect and influences the performance of subsequent vision tasks. Various deep learning methods for single image deraining have been proposed, obtaining appropriate recovery results. Unfortunately, most existing methods ignore the interaction between rain-layer and rain-free components when extracting relevant features, leading to undesirable results. To break the above limitations, we propose a progressive hybrid-modulated network (PHMNet) for single image deraining based on the two-branch and coarse-to-fine framework. Specifically, a hybrid-modulated module (HMM) with a two-branch framework is proposed to blend and modulate the feature of rain-free layers and rain streaks. After cascading several HMMs in the coarsest reconstructed stage of the PHMNet, a multi-level refined module (MLRM) is adopted to refine the final deraining results in the refined reconstructed stage. By being trained using loss functions such as contrastive learning, the PHMNet can obtain satisfactory deraining results. Extended experiments on several datasets and downstream tasks demonstrate that our method performs favorably against state-of-the-art methods in quantitative evaluation and visual effects.

Keywords: image deraining; hybrid-modulation feature; two-branch framework

MSC: 68U10

1. Introduction

Image rain degeneration is hard to avoid when capturing images by cameras in rainy weather. The presence of rain is an undesirable factor. Due to the strong scattering and reflection of light, the rain streak would blur the background scene and change the color of objects and the content of the final images. Rainy degeneration damages the visual image quality, which always affects human visual perception and the performance of downstream vision tasks, such as object detection [1,2], image segmentation [3], and recognition [4].

Furthermore, although the performance of downstream vision tasks has been improved recently with deep learning technologies, heavy rain degradation would seriously reduce the processing accuracy of existing pre-trained models for downstream tasks. We take the object detection task as an example. The pre-trained model YOLOv5 is an optimized version based on YOLOv4 [5]. The model parameter provided from https://github.com/ultralytics/yolov5 can obtain an effective and similar performance for processing clear and light-rainy images. In contrast, the heavy-rainy images would lead to false detection, as shown in Figure 1. Therefore, how to restore the heavy-rainy images is a critical issue for visual image applications.



Citation: Yu, X.; Zhang, G.; Tan, F.; Li, F.; Xie, W. Progressive Hybrid-Modulated Network for Single Image Deraining. *Mathematics* **2023**, *11*, 691. https://doi.org/10.3390/ math11030691

Academic Editors: Romualdas Bausys and Jakub Nalepa

Received: 25 December 2022 Revised: 21 January 2023 Accepted: 28 January 2023 Published: 29 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).



Figure 1. The pre-training model YOLOv5 based on COCO dataset processes the (**a**) clear images and the (**b**) light-rainy images effectively. However, it would lead to missing detection or false detection when the model YOLOv5 processes the (**c**) heavy-rainy images.

The solution to the single image deraining (SID) issue is to restore the clean and rain-free background image from the rainy image. It is difficult because we not only need to remove the rain streaks but also need to recover the background of the rain-streak area simultaneously. Early researchers of the single-image deraining issue mainly focus on filter-based methods [6] and prior-based methods [7]. These methods perform well for light-rainy images but not for heavy-rain images. Meanwhile, most of the above methods have a high time consumption.

Recently, with the rapid development of deep learning techniques, many single image deraining methods have been proposed [8–10], mainly including sequential frameworks and multi-branch frameworks. Sequential frameworks [11,12] either directly learn to estimate rain-free components of maps from rain images or estimate rain traces after obtaining rain layer components and before obtaining rain-free images. For example, RESCAN [12], a well-known recursion-based SID method, sequentially extracts rain streaks from the input rain images. However, this framework has difficulty in recovering the details of rain-free images after subtracting rain streaks from the rainy images in a simple pixelwise manner, as shown in Figure 2c,d. On the other hand, the multi-branch framework [13,14] trains the maps to estimate both the rain-free component and the rain-layer component or other side information. For example, DerainCycleGAN [15] uses parallel networks and multiple networks to process both the rain layer and rain-free images. For the multi-branch framework, each component is estimated by an independent sub-network, and the interaction of different components is ignored during depth feature processing. Therefore, it is easy to cause the artificial ringing effect in deraining results, as shown in Figure 2e,f.

In particular, we observe that the rain streak layer has a serious blocking effect and interference with the background, especially for heavy-rain images. Meanwhile, part of the background area (sky, water, etc.) would affect the rendering of rain streaks. There-

fore, we assume that the rain-free component and rain-streak component interfere with each other, making it difficult to separate the rain-streak component from the rain-free component cleanly. As shown in Figure 2, no matter whether there are sequential frameworks [12,16] or multi-branch frameworks [13], the deraining results from these methods both display serious rain-streak traces, since they ignore the interaction between rain-streak and background.



Figure 2. Comparisons of deraining results on (**a**) Rainy image. (**b**) Ground-truth. (**c**,**d**) The results from sequential frameworks (RESCAN [12], MPRNet [16]). (**e**,**f**) The results from multi-stream frameworks (MSPFN [13], PCNet [17]). (**g**) Deraining result from our framework.

To break the above limitations, we propose a progressive hybrid-modulated network (PHMNet) for single image deraining within the two-branch and coarse-to-fine framework. More specifically, we design a hybrid-modulated module (HMM) to obtain rich features by indicating the interaction between the rain-free branch and the rain-streak branch. The HMM extracts the pure feature of each component progressively after blending and modulating the features of the rain-free image and rain streaks two times. Furthermore, a coarse-to-fine framework, including the coarsest reconstructed stage (CRS) and refine reconstructed stage (RRS), has been employed to construct the final deraining image with rich details. Finally, we employ the physical model to guide the generation of rain-free images and rain streaks from the CR stage while using a contrastive-learning manner to promote the quality of the final deraining image from the RR stage.

It is noted that the existing progressive coupled network (PCNet) [17] is proposed for image deraining by coupled representation modules (CRM) within a multi-stream framework. The features of the rain-free and rain-streak components are modulated only in the last part of each CRM by the joint features, respectively. Unlike the CRM, the proposed HMM adaptively modulates the feature of each component by just using one component. For instance, the previous rain-free feature is used to guide the next rain-streak feature and rain-free feature at the first stage of HMM, while the rain-streak information is used to exploit the next rain-free component and rain-streak feature at the second stage of HMM. Therefore, we alternately modulate features of the rain-free component and rain-streak component within the proposed HMM. Motivated by the spatial properties of rain streaks and objects, we introduce the spatial attention mechanism to represent further guided information of rain-free layers and rain streaks, which achieves better performance in image deraining (see Section 4.3).

The contributions of this paper are summarized as follows:

- We propose the HMM with a new blending and modulating form for extracting the latent feature between the rain-free branch and rain-streaks branch. The features of the rain-free component and rain-streak component are alternately modulated in each HMM.
- To enhance the quality of the final deraining image, we further propose the PHMNet with the two-branch and coarse-to-fine framework by cascading serval HMMs in the coarsest reconstructed stage and employing the multi-level refined module (MLRM) in the refined reconstructed stage.

• Extensive experimental results for the proposed PHMNet have been reported, which has achieved better performance than existing state-of-the-art single image deraining methods.

The structure of this paper is established as follows. The related works about SID and feature modulation are reviewed in Section 2. Section 3 introduces the proposed PHMNet and how to solve single image deraining. In Section 4, experimental results about the proposed PHMNet and other state-of-the-art methods are demonstrated. Finally, Section 5 presents the conclusions of this paper.

2. Related Work

2.1. Single Image Deraining

The single image deraining task aims to restore a rain-free image from a rainy one, which is the basis for other downstream computer vision tasks, such as object detection [1,18], surveillance [19], and scene analysis [20,21]. High-quality images are key to further analyzing latent information and application. Therefore, to improve the quality of deraining images, the solution of removing rain from a single image has increasingly attracted researchers' attention [6,22–26]. The main attempt is using guided-filter methods [6,27] and prior-based methods [7,22] to solve the single-image deraining issue. They perform well for processing rainy images with fixed rain levels while failing with complex rain conditions.

Recently, deep learning-based methods have been used for many image restoration tasks, including image deraining, image deblurring [28], and image super-resolution [29]. For single image deraining, the existing deep learning-based methods can be simply divided into sequential frameworks and multi-stream frameworks.

2.1.1. Sequential Frameworks

The sequential frameworks for SID aim to handle the single task (either the estimated rain-free image or the estimated rain streak) from the input rainy image sequentially, where data flow linearly [30,31]. For instance, Fu et al. [11] proposed a deep-learning architecture for estimating rain-free images after extracting the detail by a priori imagedomain knowledge. ReMAEN [32] used the single-branch framework to extract rain streaks from the input rainy image and then obtained the rain-free image based on the physical model. However, it is hard to process the heavy-rain scenes. Further, Yasarla et al. [23,33] used different distortion-level information, including density, direction of rain streaks, and location quality, to extract the residual map of rain streaks and improve the quality of deraining results. Wang et al. [24] developed a quasi-sparse distribution to approximate the sparsity of rain streaks for training an image-deraining network. Further, attention mechanisms are used to improve the performance of image deraining in the single-branch framework [16,34]. For example, Chen et al. [25] proposed the multi-scale hourglass extraction block with an attention mechanism to improve the accuracy of extracted features during image deraining. Wang et al. [35] used self-attention for extracting the importance features after feature aggregation with different scales.

2.1.2. Multi-Stream Frameworks

The multi-stream framework trains the mapping on estimating the formulation components of the rain-image model, or other side information from the input rainy image simultaneously [13,15,36]. For example, Deng et al. [14] explored the single-image deraining method using the context aggregation network, which includes a two-branch framework for learning the rain streak and image detail, respectively. Further, some articles proposed the multi-branch framework for extracting the components of a new rain image physical model, such as a haze-like effect [37] and vapor effect [38]. Zhu et al. [39] used a physical model to remove rain streaks from the input rainy images, which consists of a rain streak network, rain-free network, and guide-learning network. Similarly, ref. [26] also proposed a two-stage progressive network based on the physical model for deraining. In particular, some articles [17] couple and blend different components in the multi-stream framework. Zhang et al. [40] presented a density-aware image deraining method using a multi-stream dense network. Similarly, ref. [41] proposed a hybrid block within the multi-stream framework for extracting the rain streaks more precisely.

2.2. Feature Modulation

Feature modulation can replace existing feature parameters by using information learned from additional conditions. In common, traditional feature normalization modules, batch normalization (BN) [42], instance normalization (IN) [43], and group normalization (GN) [44] are deep learning techniques for training networks effectively. In addition, several vision tasks, namely image style translation [45], visual question answering [46], and image super-resolution [47], introduce external conditional information into deep learning networks for obtaining a suitable optimized solution. In [48], latent features within the model are directly modulated by the external semantic information, which is used to generate super-resolution images. Liu et al. [49] used the high frequency of different-scale images for modulating the intermediate features and achieved better deblurring results. Inspired by the above thesis of feature modulation, the hybrid-modulated module is proposed for image deraining, which is used to blend and modulate one formula component by the attentive information from the other formula component.

3. The Proposed Method

3.1. Motivation

In general, the physical model for the rainy image is written as follows [50]:

$$l = J + R \tag{1}$$

where *I*, *J*, and *R* denote the rainy image, rain-free image, and rain streak, respectively.

Sequential frameworks either directly estimate the rain-free image *J* from rainy image *I*, or estimate the rain streak R^* before obtaining the *J* by using the formulation $J = I - R^*$, as shown in Figure 3a. On the other hand, as shown in Figure 3b, a multi-branch framework estimates the formulation components of model (1) or other side information simultaneously. In particular, the framework of PCNet [17] is a special form of multi-branch framework, as shown in Figure 3c. The refined features of J^{n+1} and R^{n+1} are explored after concatenating the features of J^n and R^n . However, different elements of R^n have different refinement effects on the J^{n+1} , while some elements even play the opposite role. Meanwhile, elements of J^n also have similar characteristics.

Therefore, we proposed a new framework to estimate the formulation components of model (1), as shown in Figure 3d. Here, the features J^{n+1} and R^{n+1} are refined after the previous features J^n and R^n are decoupled by using attention units, respectively. In this way, each component can be modulated by the appropriate information from the previous feature.



Figure 3. The basic flow of different frameworks. Here, $C(\cdot)$ denotes the concatenation operator and $AU(\cdot)$ denotes the attention unit. (a) Sequential framework. (b) Multi-branch framework. (c) PCNet [17]. (d) Our framework.

3.2. Overview

Figure 4 shows the overall structure of the proposed PHMNet that is based on a two-branch and coarse-to-fine framework for single image deraining. To obtain the hybrid-

modulated feature of two components and fuse different-level spatial information, we design two modules: (1) HMM with the attention mechanism is used to modulate the feature between the rain-free branch and the rain-streak branch. (2) MLRM is designed to improve the feature representation capability by fusing global and local information. Here, the process of the proposed framework is described in detail as follows.



Figure 4. The structure of the proposed PHMNet. The HMM blends and modulates the features of the rain-free image and rain streaks. The MLRM concatenates and refines the feature. L_p and L_{CL} denote the physical loss term and contrastive-learning loss term, respectively.

3.2.1. Coarsest Reconstructed Stage (CRS)

Given a rainy image I_{rain} , we obtain the coarse results of rain-free and rain streaks in the CRS. Specifically, we first extract the preliminary feature of the rain-free layer and rain-streak layer by using 1×1 convolution, respectively. Then, the rain-free feature f_J^0 and rain-streak feature f_R^0 are fed into the cascaded several hybrid-modulated modules (HMMs) for obtaining the rain-free and rain-streak feature representation. Here, the rain-free and rain-streak features are modulated in each HMM. Finally, the outputs of the CR stage include the coarse rain-free image J^c and rain streak R^* after using the 1×1 convolution and long-skip connection. The above process is presented as follow:

$$f_{J}^{0} = Conv_{1}^{J}(I_{rain})$$

$$f_{R}^{0} = Conv_{1}^{R}(I_{rain})$$

$$J^{c} = Conv_{3}^{J}(H((f_{J}^{0}, f_{R}^{0})) + f_{J}^{0})$$

$$R^{*} = Conv_{3}^{R}(H((f_{I}^{0}, f_{R}^{0})))$$
(2)

where $Conv_i^{\#}$ denote the convolution layer with filler $i \times i$ in the $\# \in \{BJ, BR\}$ branch; namely, BJ denotes the rain-free branch, and BR denotes the rain-streak branch. $H(\cdot)$ denotes the cascaded several HMMs.

3.2.2. Refined Reconstructed Stage (RRS)

Furthermore, we produce the refined deraining result in the RRS by using the results from the CR stage. We first get the fused feature f_{IR} by using two 3 × 3 convolution layers after concatenating the coarse rain-free image J^c and rain streak R^* . Further, the fused feature f_{IR} is fed into the multi-level refined module (MLRM) and generates the refined

feature f_J^* . Finally, we use a convolution layer and a tanh function to obtain the refined deraining image. The above process is presented as follows:

$$f_{IR} = Conv(J^{c}, R^{*})$$

$$f_{J}^{*} = F_{M}(f_{JR})$$

$$J^{*} = T(Conv(f_{I}^{*}))$$
(3)

where $Conv(\cdot)$ and $T(\cdot)$ denote the convolution layers and tanh function, respectively. $F_M(\cdot)$ denotes the MLRM.

3.3. Hybrid-Modulated Module

As in the previous introduction, the features of the rain-streak branch and rain-free branch affect each other while preserving the critical features in each branch. Therefore, we propose the HMM to modulate the information for preserving features.

As shown in Figure 5, taking the (t - 1)-th HMM as an example, the main principle of HMM is described as follows. For each HMM, the rain-free attention mask is first generated from the initial rain-free feature and is used to modulate and obtain the primary representation of rain-free content and rain streaks (f_I^{mid} and f_R^{mid}), respectively. Then, the rain attention mask is extracted from primary representation f_R^{mid} , which is then used to modulate the features f_I^t and f_R^t . Through this strategy, the HMM can encode the blending relations and obtain the refined representation of the rain-free content and rain streaks. Therefore, the final output of the modulated feature of the two components is formulated as follows:

$$f_{J}^{mid} = ConvR_{1}^{J}(\alpha_{1}M^{J} \otimes f_{J}^{t-1})$$

$$f_{J}^{t} = ConvR_{2}^{J}(f_{J}^{mid} \otimes \beta_{1}M^{R}) + f_{J}^{t-1}$$

$$f_{R}^{mid} = ConvR_{1}^{R}(\alpha_{2}M^{J} \otimes f_{R}^{t-1})$$

$$f_{R}^{t} = ConvR_{2}^{R}(f_{R}^{mid} \otimes \beta_{2}M^{R}) + f_{R}^{t-1}$$
(4)

where f_J^{t-1} and f_R^{t-1} are feature maps from the rain-free branch and rain-streak branch in the (t-1)-th HMM, respectively. M^J and M^R are attention maps based on the image feature f_J^{t-1} and primary representation f_R^{mid} , respectively. Meanwhile, $ConvR_i^J$, $ConvR_i^R$, i = 1, 2denote the convolution layer with ReLU operator in the rain-free branch and rain-streak branch, respectively. α_i , β_i , i = 1, 2 are the learnable parameters to fuse the features from different branches based on the attention maps, while we set $\sum_i \alpha_i = 1$ and $\sum_i \beta_i = 1$. After cascading several HMMs, we can effectively and progressively learn the latent feature of the rain-free layer and rain streaks, as shown in Figure 5.

Note that the important feature of one branch is preserved by employing the spatial attention module. Then, these features are used to modulate the features of two branches simultaneously. Here, we introduce the spatial attention unit (SAU) to extract the spatial attention map of features [51]. As shown in Figure 5, the input feature map f_J^{t-1} of the first SAU is used to extract the attention map M^J with size $1 \times H \times W$:

$$M^{J} = S(F_{sa}(Cat(GMP(f_{I}^{t-1}), GAP(f_{I}^{t-1}))))$$
(5)

where $F_{sa}(\cdot)$ denotes the convolution layers in the SAU and $S(\cdot)$ denotes the sigmoid function. $Cat(\cdot)$ denotes the concatenation. The attention map M^{J} explores the spatial positions that are essential for recovery of the deraining features and the preservation of the raining feature. Similarly, the input feature map of the second SAU is the middle feature map f_{R}^{mid} for extracting the attention map M^{R} with size $1 \times H \times W$:

$$M^{R} = S(F_{sa}(Cat(GMP(f_{R}^{mid}), GAP(f_{R}^{mid}))))$$
(6)

It should be noted that the attention map M^R guides the deraining branch for producing the more distinguished feature representation and preserving the critical spatial information, as shown in Figure 6. During training, the model achieves better performance with SAU than with other attention ways (Section 4.3).



Figure 5. The architecture of the proposed hybrid-modulated module (HMM). Here, the spatial attention unit (SAU) is an example of the attention unit (AU).



Figure 6. The feature visualization of the rain-free layer (the first row) and rain streak (the second row) from (**a**) *1st* HMM, (**b**) *4th* HMM, (**c**) *7th* HMM, (**d**) output of CR stage.

3.4. Multi-Level Refined Module

Combining multi-level features has been widely employed in several vision tasks [39,52]. Inspired by the previous work [39] that employs the Unet-likely framework with multi-scale residual block for restoring deraining results, we use a single multi-level refined module to refine the final deraining image after obtaining the coarse deraining result and rain streaks from the CR stage. As shown in Figure 4, we first extract the fused feature representation after concatenating the coarse results from the CR stage. Then, we introduce four-level pool-upsample operators to obtain the global context and local structure information with different scales from the fused feature. Each pool-upsample operator has a pooling layer and an upsampling layer. Here, the size of four pooling layers is set as 2, 4, 8, and 16, respectively. Each upsampling layer is used to process pooling features like the size of the fused feature. Finally, the refined deraining result is estimated after concatenating all pool-upsampling features and fused features.

3.5. Total Loss Function

The proposed PHMNet is optimized to obtain a deraining image by using the total loss function as follows:

$$L = L_{Re} + L_{SSIM} + \lambda_p L_p + \lambda_{CL} L_{CL}$$
⁽⁷⁾

where L_{Re} denotes the reconstruction loss term for the coarse deraining result and the refined result, L_{SSIM} denotes the SSIM loss term [39], L_p denotes the physical model loss term, and L_{CL} denotes the contrastive-learning loss term. λ_p and λ_{CL} are hyper-parameters for balancing the importance of L_p and L_{CL} , respectively.

In this paper, the reconstruction loss is defined by using L1-norm [53] for measuring the difference among the ground-truth image J_{gt} , the coarse result J^c , and the refined result J^* , respectively:

$$L_{Re} = \lambda_c \|J^c - J_{gt}\|_1 + \lambda_r \|J^* - J_{gt}\|_1$$
(8)

where λ_c and λ_r are hype-parameters.

Then, we use the physical model loss term [39] to guide the CR stage for extracting the rain-free image and rain streaks. According to the physical model (1), the physical model loss term is set by L1-norm as follows:

$$L_p = \|J^c + R^* - I\|_1 \tag{9}$$

Contrastive learning [54–57] aims to learn a representation for pulling close to the positive sample and pulling far with negative samples, which has been widely employed in the computer vision field. It is noted that the existing contrastive-learning method for image dehazing [58] only constructs the latent feature representation space based on a pre-trained VGG. To construct a more accurate latent-feature space for characterizing rainy and clear images, we construct the latent-feature representation space based on the pre-trained classification network that is used to recognize rainy or clear images. Here the pre-trained classification network is constructed according to the Inception Network [59]. Therefore, we construct the contrastive-learning loss term based on [54,55] as follows:

$$L_{CL} = \frac{D_c(F_{PRnet}(J_{gt}), F_{PRnet}(J^*))}{D_c(F_{PRnet}(I_{rain}), F_{PRnet}(J^*))}$$
(10)

where F_{PRnet} denotes the pre-trained rain-classifier network used to extract the representation feature from the ground-truth image J_{gt} , the refined result J^* , and the rainy image I_{rain} , respectively. $D_c(x_1, x_2)$ denotes the cosine distance between x_1 and y_1 . Here, the refined deraining image J^* is set as the anchor. The ground-truth image J_{gt} is set as the positive sample, while the input raining image I_{rain} is set as the negative sample.

4. Experimental Results

4.1. Experimental-Setting Details

Comparison Methods and Evaluation Metrics. In this section, several single-image deraining algorithms, consisting of sequential methods (ReMAEN [32], RESCAN [12], and MPRNet [16]) and multi-branch methods (DID-MDN [40], MSPFN [13], and PCNet [17]), are compared with the proposed PHMNet. All comparison methods are trained with their public codes and setting parameters performed by the same dataset. Furthermore, three quality assessment metrics (peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and learned perceptual image patch similarity (LPIPS) [60]) are used to evaluate the experimental results for displaying the quality of deraining results. Here, the PSNR is used to measure the degree of image distortion. The higher the PSNR score, the less image distortion. The metric SSIM is based on three comparative measures between the restored image and ground truth, including brightness, contrast, and structure. The higher the SSIM score, the higher the image similarly. Meanwhile, the metric LPIPS is based on the

perceptual loss between the restored image and the ground truth. The smaller the LPIPS score, the better the image quality.

Datasets. We construct the training dataset from dataset [53] and the DID-MDN dataset [40]. Our training dataset here includes 15,600 image pairs with clear and rainy images. For the testing set, we choose the test samples from [53], including 200 heavy-rain images, named *Rain200H*. Meanwhile, the common synthetic rainy dataset from DID-MDN [40] with 1200 rainy images, named *Test1200*, is also used to evaluate the deraining methods. Furthermore, we choose the real data *Real300* [61] to evaluate the proposed PHM-Net. In particular, an unpaired rain/clear dataset is collected from Gopro [62], FLIR [63], and the above training dataset, then used to train the pre-trained rain-classifier network about the CL loss. The unpaired dataset includes 15,560 rainy images and 15,786 clear images. After using data augmentation with this dataset, we train the pre-trained rain-classifier network, which is used to recognize the rainy or clear image. Finally, based on this pre-training classification network, we construct the latent feature-representation model for the CL loss.

Training Details. The overall PHMNet is trained in an end-to-end manner based on the PyTorch framework. To enhance the trained performance of PHMNet, we crop the image patches of size 256×256 as the new training sample. During 200 training epochs, we use Adam as the optimizing solver with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Furthermore, the learning rate is set as 0.001 initially while changing by the cosine annealing strategy [64]. Our network is performed on four RTX-2080 GPUs. The batch size is set as 12, which is set based on the validation of the experimental results. In the baseline, the number of HMMs is set as 8 in the CR stage. The parameters λ_p and λ_{CL} of function (7) are set as 2 and 0.001, respectively, while the parameters λ_c and λ_r of function (8) are set as 1 and 1, respectively.

4.2. Comparison with Methods

Synthetic Datasets. We evaluate all test images from *Rain200H* and *Test1200* on the compared methods and use quantitative measures (PSNR, SSIM, and LPIPS) to measure the reconstructed results, respectively. Thus, the PSNR, SSIM, and LPIPS performances are reported in Table 1. It is noted that the proposed PHMNet outperforms other compared methods on the dataset *Rain200H* and dataset *Test1200*. On the dataset *Rain200H*, with heavy rain distributions, our method achieves improvement about 0.53 dB on PSNR, 0.0087 on SSIM, and 0.0124 on LPIPS compared to the sequential framework ReMAEN [32] with the second top performance. Similarly, on the other dataset *Test1200*, with multiple types of rainstreak distribution, the proposed method also gets the best results. Furthermore, compared to the existing best multi-stream framework (PCNet [17]), the proposed method has a large improvement of about 1.31 dB on PSNR, 0.0342 on SSIM, and 0.0128 on LPIPS with dataset *Test1200*, further verifying the effectiveness of its feature modulated mechanism between the rain-streak layer and the rain-free layer.

Table 1. Metric values (PSNR/SSIM/LPIPS) of the proposed PHMNet and the compared methods on the synthetic datasets. (\uparrow) indicates higher scores, better image quality. (\downarrow) indicates smaller scores, better image quality. Best results are marked in bold.

Metric	PSNR (†)	SSIM (†)	LPIPS (↓)
Data		Rain200H/Test1200	
ReMAEN [32]	27.43/28.18	0.8660/0.8738	0.1723/0.1456
RESCAN [12]	26.63/31.20	0.8355/0.9026	0.2161/0.1405
MPRNet [16]	27.30/31.04	0.8614/0.8861	0.1805/0.1529
DID-MDN [40]	-/27.93	-/0.8662	-/0.1692
MSPFN [13]	23.88/22.57	0.7974/0.7809	0.1891/0.1859
PCNet [17]	25.77/30.19	0.8436/0.8792	0.2103/0.1426
PHMNet	27.96/31.50	0.8747/0.9134	0.1647/0.1298

Figures 7 and 8 show several visual deraining results of different methods from the test dataset. For heavy rainy scenes, the proposed method performs favorably against other deraining methods, including the sequential frameworks and multi-stream frameworks, as shown in Figure 7. Although the RESCAN [12] gets high evaluation values for deraining results, there are blurry edges that affect the visual quality (Figure 7c). As zoomed into the cloudy regions from the second and fourth rows in Figure 7, the proposed PHMNet removes the rain streak in the sky and enhances the edge details of the cloud. Compared to other deraining methods, the proposed method successfully removes most rain streaks, enhances visibility, and lights up details in edge regions. On the other hand, for scenes with more diverse rainfall distributions, as shown in Figure 8, due to the feature hybrid-modulation mechanism, the proposed method outperforms other multi-branch framework methods, such as DID-MDN [40], MSPFN [13], and PCNet [17] on dataset Test1200. Specifically, as shown in the first, third, and fourth rows of Figure 8, the texture and details of the building structure damaged by rain streaks can be clearly reconstructed by the proposed PHMNet, while other methods are affected by blur and artificial ringing. Similar situations can be found by comparing the restored sky regions of the second row of Figure 7, the hurdle regions of the fifth row of Figure 8, and the body regions of the sixth row of Figure 7.

Real-world Dataset. To demonstrate the effect of the proposed method, several experiments by different methods on real-world dataset *Real300* are displayed in Figure 9. This dataset includes more challenging rainy images with unknown degradation, which have different real-world scenes, including daytime and nighttime. Due to the lack of the corresponding ground-truth clear images, the generated deraining images are only evaluated qualitatively. Compared with other methods, the proposed PHMNet can remove the rain streaks and obtain acceptable deraining results, as shown in Figure 9i. Our method not only removes the rain streaks and retains the shape's edges (Figure 9i) but also obtains the suitable rain-streak image (Figure 9j). Most other methods cannot process this rainy image well, retaining tiny rain streaks. Specifically, for daytime scences, as shown in the first, second, and fourth rows of Figure 9, the proposed PHMNet can remove the rain streak and reconstruct the details of the background, including branches, trunks, and leaves of a tree. For black backgrounds or nighttime scences, the deraining result from the proposed method has fewer rain streaks than that of results from other methods. In particular, the proposed method performs better than other methods, generating a clear result with few rainy streaks in the vehicle-light area, as shown in the results in the fifth row of Figure 9.



Figure 7. The de-raining results of the synthetic dataset *Rain200H* by comparing SOTA methods. (a) Rainy image. (b) Ground-truth. (c–g) Results from RESCAN [12], MPRNet [16], PCNet [17], ReMAEN [32], MSPFN [13]. (h) Deraining result from CR stage of our method. (i) Deraining result from RR stage of our method. (j) Rain streak by our method. Please zoom in to see clearly.



Figure 8. The visualization results by comparing SOTA methods from the synthetic dataset *Test1200*. (a) Rainy image. (b) Ground-truth. (c–g) Results from RESCAN [12], MPRNet [16], PCNet [17], ReMAEN [32], DID-MDN [40]. (h) Deraining result from CR stage of our method. (i) Deraining result from RR stage of our method. Please zoom in on the red region to see clearly.



Figure 9. The visualization results by comparing SOTA methods from the real-world dataset *Real300.* (a) Rain image. (b–g) Results from RESCAN [12], MPRNet [16], PCNet [17], ReMAEN [32], MSPFN [13], DID-MDN [40]. (h) Deraining result from CR stage of our method. (i) Deraining result from RR stage of our method. (j) Rain streak by our method. Please zoom in to see clearly.

Two Downstream Tasks. We use YOLOv5, an optimized version based on YOLOv4 [5], to verify the image quality restored by different deraining methods in the object detection tasks. The model parameter is provided from https://github.com/ultralytics/yolov5,

accessed on 15 July 2022. Meanwhile, the MTCNN [65] is also used to demonstrate the image quality restored by different deraining methods in the face detection tasks. For object detection tasks, the proposed method provides restored images with higher quality, which leads to less misdetection. For instance, since the restored images from MPRNet [16] and PCNet [17] have noticeable rainy streaks, all cars are recognized as trains or trucks, as shown in the second row of Figure 10c,d. Furthermore, as shown in the third row of Figure 10, the deraining image resulted from the proposed method can provide higher quality information to detect the details of two buses, while only one bus can be detected in the other deraining results.

For the face detection task, MTCNN [65] processes the images recovered by our method with stable detection results, as shown in Figure 11g. Although the MPRNet [16] can yeild the deraining results for detecting all facial regions, its deraining results contain tiny rain streaks. Meanwhile, we observe that the recovered images from some deraining methods, such as RESCAN [12] and DID-MDN [40], are not able to detect the correct face, while the MTCNN [65] can detect the face from the corresponding rainy image. Therefore, it is essential to construct a deraining method that can provide highly visual images and does not affect the performance of downstream tasks.

Limitations. Since the training datasets do not include the rainy type of raindrops, the proposed method can not remove the drop-like effect well, as shown in the second row of Figure 9j. On the other hand, the loss function (9) is constructed by the physical model (1). However, if the low-quality image has haze, water vapor, and other degradation factors, the proposed model is hard to trained by using the loss function (9), which would lead to unpleasant results. As shown in the second row of Figure 11g, the area of the chair of the deraining result is unclear due to the water vapor. Therefore, future work is planned to develop the new physical model including different degradation factors.



Figure 10. The visualization results of object detection by YOLOv5 for images from different deraining methods. (a) Rain image. (b–f) Tesults from RESCAN [12], MPRNet [16], PCNet [17], ReMAEN [32], DID-MDN [40]. (g) Object detection result from our method. Please zoom in to see clearly and note the red arrow.





Figure 11. The visualization results of face detection by MTCNN [65] for images from different deraining methods. (a) Rain image. (b–f) Results from RESCAN [12], MPRNet [16], PCNet [17], ReMAEN [32], DID-MDN [40]. (g) Object detection result from our method. Please zoom in to see clearly.

4.3. Ablation Study

To demonstrate the contribution of various parts in PHMNet qualitatively and quantitatively, some ablation experiments trained on the *Rain200H* dataset are shown as follows. First, we explore the modulated mechanism of PHMNet between feature maps of rainstreak and rain-free layers. Further experiments are shown to evaluate why spatial attention is implemented in the HMM. Then, the effectiveness of each loss term is evaluated. Meanwhile, an ablation study of hyperparameters of a loss function has been provided. Finally, an additional study on finding appropriate latent-feature space to construct contrastive learning loss is presented.

Effect of Modulated Mechanism. Here, a baseline is constructed by removing the feature modulation (FM) of the proposed HMMs. This means that both the rain-free branch and the rain-streak branch of the baseline are constructed by several residual modules. Figure 12 shows that PHMNet (HMM with FM) can achieve rapid feature modulation and extraction, leading to more meaningful features. As shown in Figure 12b,c PHMNet's third module (the third row of Figure 12) extracts the outline feature of the object approximately, while the third module of baseline (the first row) has not obtained effective features. Meanwhile, PHMNet's rain-streak branch extracts more effective spatial features of the rain streak, as shown in the fourth row of Figure 12b,c. Furthermore, Table 2 shows that the proposed PHMNet performs better than the baseline. In particular, the baseline results in a large decrease of 2.06 dB on PSNR, 0.0077 on SSIM, and 0.0055 on LPIPS with dataset *Rain200H*, validating the decision to use the feature modulation between the rain-free branch and rain-streak branch. Thus, the proposed HMM with FM can extract more valuable features, leading to better deraining results.

Table 2. Metric values (PSNR/SSIM/LPIPS) of the proposed method (HMM with the FM) and baseline (HMM without the FM).

Method	Metric
baseline	25.89/0.8670/0.1702
our method	27.96/0.8747/0.1647



Figure 12. The first and second row show the visualization results of the rain-free layer and rain streak of the **baseline (HMM without FM)**, respectively. The third and fourth row display the visualization results of the rain-free layer and rain streak of the **proposed PHMNet (HMM with FM)**, respectively. The first column shows a rainy image. The results of other columns are from (**a**) *1st* module, (**b**) *3rd* module, (**c**) *5th* module, (**d**) *7th* module, (**e**) output of CR stage.

Effect of the Proposed HMM. As introduced in Section 3.3, the proposed HMM is designed by using two spatial attention units. Therefore, we compare the proposed HMM with several baseline modules to verify the effect of spatial attention units. It is noted that all baseline modules are designed by using the backbone of Section 3.3. The baseline modules are shown as follows.

- HMM-CC: This baseline module is designed by replacing all attention units of HMM with channel attention.
- HMM-CS: This baseline is designed by just using channel attention to replace the 1st attention unit of HMM with channel attention.
- HMM-SC: This baseline just replaces the 2nd attention unit of HMM with channel attention.

It is noted that each baseline module is used to replace the HMM of the proposed PHMNet and trained by the same training dataset and same loss terms as that of PHMNet. Table 3 shows the metric values (PSNR/SSIM/LPIPS) of the deraining network with each baseline module and the proposed HMM. Here, the baseline HMM-SC results in a large decrease of 1.30 dB on PSNR, 0.0099 on SSIM, and 0.0134 on LPIPS with dataset *Rain200H*, while the performance of the baseline HMM-SC decreases 1.14 dB on PSNR, 0.0025 on SSIM, and 0.0002 on LPIPS with dataset *Rain200H*. Furthermore, the baseline HMM-CC results in a large decrease of 0.65 dB on PSNR, 0.0172 on SSIM, and 0.0096 on LPIPS with dataset *Rain200H*. The results show that spatial information plays an important role in hybrid-modulating features within the HMMs, which leads to a better performance of the proposed method.

Method	Position		Metric		
	1st ¹	$2nd^{1}$	PSNR	SSIM	LPIPS
HMM-CC	CA ²	СА	27.31	0.8575	0.1743
HMM-CS	CA	SA ²	26.82	0.8722	0.1649
HMM-SC	SA	CA	26.66	0.8648	0.1781
HMM (our method)	SA	SA	27.96	0.8747	0.1647

Table 3. Metric values (PSNR/SSIM/LPIPS) of the proposed HMMs with different attention mechanisms. Our method uses the spatial attention in all attention units of HMM. Best results are marked in bold.

¹ 1st and 2nd denote the first and the second attention unit in HMM, respectively. ² CA and SA denote the channel attention and spatial attention, respectively.

Effect of Different Loss Terms. In order to prove the effectiveness of the proposed PHMNet loss function (7), several combinations of loss function terms are used to train PHMNet. As shown in Table 4, the combination of different loss terms leads to consistent improvement, and the combination of all these loss terms produces the best quantitative results. In particular, after adding the loss term L_{SSIM} to train a network with dataset *Rain200H*, the network results in an improvement of 0.0068 on SSIM and 0.0144 on LPIPS compared to that of a network trained by loss term L_{Re} . It demonstrates that the loss term L_{SSIM} enhances the quality of the deraining results on the SSIM metric. Furthermore, the network trained without the loss term L_p with dataset Rain200H results in a large decrease of 1.23 dB on PSNR, 0.0007 on SSIM, and 0.0007 on LPIPS. Thus, the loss term L_p combines the information of the input rain image and helps the proposed PHMNet obtain improvement quality of the deraining images. It is noted that the loss term L_{CL} also improves the visual effect of the deraining result. The usage of the loss term L_{CL} results in a large improvement of 1.1 dB on PSNR, 0.0008 on SSIM, and 0.0007 on LPIPS for dataset Rain200H. This benefits from using the latent feature to represent space to measure the similarity and difference among the rainy image, restored image, and ground truth.

Table 4. Metric values (PSNR/SSIM/LPIPS) of the proposed PHMNet trained by different loss terms with dataset *Rain200H*.

L_{Re}	L_p	L_{SSIM}	L_{CL}	Metirc
\checkmark				26.47/0.8586/0.1835
\checkmark	\checkmark			26.41/0.8604/0.1816
\checkmark		\checkmark		26.71/0.8654/0.1691
\checkmark	\checkmark	\checkmark		26.84/0.8667/0.1717
\checkmark		\checkmark	\checkmark	26.73/0.8740/0.1654
\checkmark	\checkmark	\checkmark	\checkmark	27.96/0.8747/0.1647

Effect of Different Hyperparameters. Taking the dataset *Rain200H* as an example, we explore the performance of the proposed method with different hyperparameters of loss function after using various image quality measurement methods (PSNR, SSIM, and LPIPS). Here, we focus mainly on two hyperparameters of function (7), including λ_p and λ_{CL} .

On the one hand, hyperparameter λ_p is employed mainly to balance the influence of physical model loss terms. Here, λ_p is set from 0.1 to 2. As shown in Table 5, the performance of the proposed PHMNet is improving, while the value of λ_p is increasing. It is noted that the proposed PHMNet has a very similar performance when λ_p is set to 1 or 2. Therefore, λ_p is set to 2 in this paper. On the other hand, hyperparameter λ_{CL} is used to constrain the importance of a contrastive-learning loss term. Here, λ_{CL} is set from 5×10^{-4} to 1×10^{-2} . As shown in Table 6, the proposed PHMNet can achieve the best performance when λ_{CL} is set to 1×10^{-3} . Furthermore, the performance of the proposed method would be reduced when λ_{CL} is set to a large term, such as 1×10^{-2} . Therefore, λ_{CL} is set to 2 in this paper.

Value of λ_p	Metric
0.1	26.90/0.8644/0.1725
0.5	27.12/0.8701/0.1698
1	27.88/0.8740/0.1645
2	27.96/0.8747/0.1647

Table 5. Metric values (PSNR/SSIM/LPIPS) of the proposed PHMNet trained with different λ_p hyperparameters. It is noted that other hyperparameters are fixed.

Table 6. Metric values (PSNR/SSIM/LPIPS) of the proposed PHMNet trained with different λ_{CL} hyperparameters. It is noted that other hyperparameters are fixed.

Value of λ_{CL}	Metric
0.01	27.23/0.8723/0.1747
0.005	27.83/0.8739/0.1646
0.001	27.96/0.8747/0.1647
0.0005	27.63/0.8729/0.1721

Rain Classifier Network for Contrastive-learning Loss. Existing image restoration methods [58] only construct the latent feature representation space based on a pre-trained VGG. In this paper, we construct the latent feature space as follows. Firstly, we construct a rain-classifier network based on inception network [59] for discriminating whether the input image is rainy. Then, we lose the last classified layer of the rain-classifier network and fix the other parameter as the latent feature-representation model after training the RC net. Finally, we construct the contrastive-learning loss as shown in Equation (10).

Table 7 shows that contrastive-learning loss based on the proposed rain classifier network improves the performance of the proposed PHMNet. As shown in Table 7, the usage of a rain classifier network for constructing contrastive-learning loss results in an excellent improvement of 1.57 dB on PSNR, 0.0085 on SSIM, and 0.0066 on LPIPS for dataset *Rain200H*. This means that the rain classifier network can provide stronger supervising information on rain streaks and background to train the proposed method, resulting in better image quality.

Table 7. Metric values (PSNR/SSIM/LPIPS) of the proposed PHMNet trained by the CL with different feature representation space.

Method	Metric
CL with VGG	26.39/0.8662/0.1713
our method	27.96/0.8747/0.1647

5. Conclusions

We propose a progressive hybrid-modulated network (PHMNet) for single image deraining within a two-branch and two-stage framework. Firstly, the hybrid-modulated module (HMM) with two branches is introduced to blend and modulate the feature of the rain-free image layer and the rain streak layer, which refine the hybrid feature of each component progressively. Then, we cascade several HMMs as the CRS and adopt the multilevel refined module (MLRM) as the refined reconstructed stage (RRS), which can improve the quality of the final deraining result. Finally, we employ the physical model to guide the generation of rain-free images and rain streaks in the CR stage while using a contrastivelearning manner to promote the quality of the final deraining image from the RR stage. The quantitative and qualitative analysis of the experimental results demonstrates that the proposed PHMNet performs favorably against state-of-the-art methods when processing both synthetic and real-world datasets. The current framework based on a single physical model has limitations that make it difficult to process a low-quality image with multiple degradation factors, such as haze, water vapor, and so on. Therefore, we plan to improve the physical model and address these issues in future work.

Author Contributions: Conceptualization, X.Y. and W.X.; methodology, X.Y.; software, X.Y. and F.T.; validation, X.Y., G.Z. and F.T.; formal analysis, F.L.; investigation, G.Z.; resources, G.Z.; data curation, F.L.; writing—original draft preparation, X.Y. and W.X.; writing—review and editing, X.Y., G.Z. and F.L.; visualization, X.Y. and W.X.; supervision, W.X.; project administration, X.Y.; funding acquisition, W.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key-Area Research and Development Program of Foshan City under Grant 2020001006812, the Shunde District Core Technology Research project under Grant 2030218000174.

Institutional Review Board Statement: The studies were not involving humans or animals.

Informed Consent Statement: The studies were not involving humans.

Data Availability Statement: The open source research datasets can be obtained from https://github.com/ZhangXinNan/RainDetectionAndRemoval, https://xueyangfu.github.io/projects/LPNet.html, and https://github.com/hezhangsprinter/DID-MDN.

Acknowledgments: The authors would like to thank previous researchers for their valuable paper and public code. Further, the authors would like to thank the editor and the anonymous reviewers for their valuable suggestions and comments.

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

References

- 1. Nejatishahidin, N.; Fayyazsanavi, P.; Kosecka, J. Object pose estimation using mid-level visual representations. *arXiv* 2022, arXiv:2203.01449.
- Qian, X.; Cheng, X.; Cheng, G.; Yao, X.; Jiang, L. Two-stream encoder GAN with progressive training for co-saliency detection. IEEE Signal Process. Lett. 2021, 28, 180–184. [CrossRef]
- Fu, J.; Liu, J.; Jiang, J.; Li, Y.; Bao, Y.; Lu, H. Scene Segmentation with Dual Relation-Aware Attention Network. *IEEE Trans. Neural Netw. Learn. Syst.* 2021, 32, 2547–2560. [CrossRef] [PubMed]
- Li, P.; Prieto, L.; Mery, D.; Flynn, P.J. On Low-Resolution Face Recognition in the Wild: Comparisons and New Techniques. *IEEE Trans. Inf. Forensics Secur.* 2019, 14, 2000–2012. [CrossRef]
- Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. Scaled-yolov4: Scaling cross stage partial network. In Proceedings of the IEEE/cvf Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13029–13038.
- Xu, J.; Zhao, W.; Liu, P.; Tang, X. Removing rain and snow in a single image using guided filter. In Proceedings of the 2012 IEEE International Conference on Computer Science and Automation Engineering, Zhangjiajie, China, 25–27 May 2012; Volume 2, pp. 304–307. [CrossRef]
- Sun, S.H.; Fan, S.P.; Wang, Y.C.F. Exploiting image structural similarity for single image rain removal. In Proceedings of the IEEE International Conference on Image Processing, Paris, France, 27–30 October 2014; pp. 4482–4486.
- 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, 16–17 June 2019; pp. 12262–12271.
- Chen, D.; He, M.; Fan, Q.; Liao, J.; Zhang, L.; Hou, D.; Yuan, L.; Hua, G. Gated Context Aggregation Network for Image Dehazing and Deraining. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision, Waikoloa Village, HI, USA, 7–11 January 2019; pp. 1375–1383.
- 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, 20–25 June 2021; pp. 7742–7751.
- 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/CVF Conference on Computer Vision and Pattern Recognition, Los Alamitos, CA, USA, 21–26 July 2017; pp. 1715–1723.
- 12. Li, X.; Wu, J.; Lin, Z.; Liu, H.; Zha, H. Recurrent Squeeze-and-Excitation Context Aggregation Net for Single Image Deraining. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018 ; pp. 262–277.
- Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Han, Z.; Lu, T.; Huang, B.; Jiang, J. Decomposition Makes Better Rain Removal: An Improved Attention-Guided Deraining Network. *IEEE Trans. Circuits Syst. Video Technol.* 2021, 31, 3981–3995. [CrossRef]
- 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. 14548–14557.

- Wei, Y.; Zhang, Z.; Wang, Y.; Xu, M.; Yang, Y.; Yan, S.; Wang, M. DerainCycleGAN: Rain Attentive CycleGAN for Single Image Deraining and Rainmaking. *IEEE Trans. Image Process.* 2021, 30, 4788–4801. [CrossRef]
- 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, 20–25 June 2021; pp. 14816–14826.
- 17. Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Wang, Z.; Wang, X.; Jiang, J.; Lin, C.W. Rain-free and residue hand-in-hand: A progressive coupled network for real-time image deraining. *IEEE Trans. Image Process.* **2021**, *30*, 7404–7418. [CrossRef]
- Sultani, W.; Chen, C.; Shah, M. Real-world anomaly detection in surveillance videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6479–6488.
- 19. Peng, J.; Wang, Y.; Wang, H.; Zhang, Z.; Fu, X.; Wang, M. Unsupervised vehicle re-identification with progressive adaptation. *arXiv* 2020, arXiv:2006.11486.
- Javaheri, E.; Kumala, V.; Javaheri, A.; Rawassizadeh, R.; Lubritz, J.; Graf, B.; Rethmeier, M. Quantifying mechanical properties of automotive steels with deep learning based computer vision algorithms. *Metals* 2020, 10, 163. [CrossRef]
- Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 1998, 20, 1254–1259. [CrossRef]
- Mu, P.; Chen, J.; Liu, R.; Fan, X.; Luo, Z. Learning Bilevel Layer Priors for Single Image Rain Streaks Removal. *IEEE Sign. Process.* Lett. 2019, 26, 307–311. [CrossRef]
- 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, 16–17 June 2019; pp. 8397–8406.
- 24. Wang, Y.; Ma, C.; Zeng, B. Multi-Decoding Deraining Network and Quasi-Sparsity Based Training. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13375–13384.
- Chen, X.; Huang, Y.; Xu, L. Multi-Scale Hourglass Hierarchical Fusion Network for Single Image Deraining. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 872–879.
- 26. Yang, W.; Tan, R.T.; Feng, J.; Wang, S.; Cheng, B.; Liu, J. Recurrent Multi-Frame Deraining: Combining Physics Guidance and Adversarial Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, 44, 8569–8586. [CrossRef] [PubMed]
- Zheng, X.; Liao, Y.; Guo, W.; Fu, X.; Ding, X. Single-Image-Based Rain and Snow Removal Using Multi-guided Filter. In Proceedings of the Neural Information Processing, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 258–265.
- Zhang, H.; Dai, Y.; Li, H.; Koniusz, P. Deep Stacked Hierarchical Multi-Patch Network for Image Deblurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 5971–5979.
- 29. Fang, F.; Li, J.; Zeng, T. Soft-Edge Assisted Network for Single Image Super-Resolution. *IEEE Trans. Image Process* 2020, 29, 4656–4668. [CrossRef] [PubMed]
- Ren, D.; Shang, W.; Zhu, P.; Hu, Q.; Meng, D.; Zuo, W. Single image deraining using bilateral recurrent network. *IEEE Trans. Image Process.* 2020, 29, 6852–6863. [CrossRef]
- 31. Zhang, Y.; Zhang, J.; Huang, B.; Fang, Z. Single-image deraining via a Recurrent Memory Unit Network. *Knowl.-Based Syst.* 2021, 218, 106832. [CrossRef]
- Yang, Y.; Lu, H. Single Image Deraining using a Recurrent Multi-scale Aggregation and Enhancement Network. In Proceedings
 of the IEEE International Conference on Multimedia and Expo, Shanghai, China, 8–12 July 2019; pp. 1378–1383.
- Yasarla, R.; Patel, V.M. Confidence Measure Guided Single Image De-Raining. *IEEE Trans. Image Process* 2020, 29, 4544–4555. [CrossRef]
- Zheng, Y.; Yu, X.; Liu, M.; Zhang, S. Single-Image Deraining via Recurrent Residual Multiscale Networks. *IEEE Trans. Neural Netw. Learn. Syst.* 2022, 33, 1310–1323. [CrossRef] [PubMed]
- Wang, C.; Wu, Y.; Su, Z.; Chen, J. Joint Self-Attention and Scale-Aggregation for Self-Calibrated Deraining Network. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 2517–2525.
- Wang, G.; Sun, C.; Sowmya, A. Context-Enhanced Representation Learning for Single Image Deraining. Int. J. Comput. Vis. 2021, 129, 1650–1674. [CrossRef]
- 37. Wang, Y.; Gong, D.; Yang, J.; Shi, Q.; Hengel, A.V.D.; Xie, D.; Zeng, B. Deep Single Image Deraining via Modeling Haze-Like Effect. *IEEE Trans. Multimed.* 2021, 23, 2481–2492. [CrossRef]
- 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; pp. 367–382.
- Zhu, H.; Wang, C.; Zhang, Y.; Su, Z.; Zhao, G. Physical model guided deep image deraining. In Proceedings of the 2020 IEEE International Conference on Multimedia and Expo, London, UK, 6–10 July 2020; pp. 1–6.
- Zhang, H.; Patel, V.M. Density-aware single image de-raining using a multi-stream dense network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 695–704.
- Wei, Y.; Zhang, Z.; Zhang, H.; Hong, R.; Wang, M. A coarse-to-fine multi-stream hybrid deraining network for single image deraining. In Proceedings of the 2019 IEEE International Conference on Data Mining, Beijing, China, 8–11 November 2019; pp. 628–637.
- Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings
 of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.

- 43. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Instance normalization: The missing ingredient for fast stylization. *arXiv* 2016, arXiv:1607.08022.
- 44. Wu, Y.; He, K. Group normalization. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
- 45. Huang, X.; Belongie, S. Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1510–1519.
- Perez, E.; Strub, F.; De Vries, H.; Dumoulin, V.; Courville, A. Film: Visual reasoning with a general conditioning layer. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; pp. 3942–3951.
- 47. Hu, Y.; Li, J.; Huang, Y.; Gao, X. Channel-Wise and Spatial Feature Modulation Network for Single Image Super-Resolution. *IEEE Trans. Circuits Syst. Video Technol.* 2020, 30, 3911–3927. [CrossRef]
- Wang, X.; Yu, K.; Dong, C.; Change Loy, C. Recovering Realistic Texture in Image Super-Resolution by Deep Spatial Feature Transform. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 606–615.
- 49. Liu, Y.; Fang, F.; Wang, T.; Li, J.; Sheng, Y.; Zhang, G. Multi-scale Grid Network for Image Deblurring with High-frequency Guidance. *IEEE Trans. Multimed.* 2021, 24, 2890–2901. [CrossRef]
- 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]
- Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
- Zhang, H.; Patel, V.M. Densely connected pyramid dehazing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3194–3203.
- 53. Yang, W.; Tan, R.T.; Feng, J.; Guo, Z.; Yan, S.; Liu, J. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 1377–1393. [CrossRef] [PubMed]
- 54. Hjelm, R.D.; Fedorov, A.; Lavoie-Marchildon, S.; Grewal, K.; Bachman, P.; Trischler, A.; Bengio, Y. Learning deep representations by mutual information estimation and maximization. *arXiv* **2018**, arXiv:1808.06670.
- He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 9729–9738.
- Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual Event, 13–18 July 2020; pp. 1597–1607.
- Caron, M.; Misra, I.; Mairal, J.; Goyal, P.; Bojanowski, P.; Joulin, A. Unsupervised learning of visual features by contrasting cluster assignments. arXiv 2020, arXiv:2006.09882.
- Wu, H.; Qu, Y.; Lin, S.; Zhou, J.; Qiao, R.; Zhang, Z.; Xie, Y.; Ma, L. Contrastive Learning for Compact Single Image Dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10551–10560.
- Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA, 27–30 June 2016; pp. 2818–2826.
- Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 586–595.
- Fu, X.; Liang, B.; Huang, Y.; Ding, X.; Paisley, J. Lightweight pyramid networks for image deraining. *IEEE Trans. Neural Netw. Learn. Syst.* 2019, 31, 1794–1807. [CrossRef]
- Kupyn, O.; Budzan, V.; Mykhailych, M.; Mishkin, D.; Matas, J. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8183–8192.
- 63. Zhang, H.; Fromont, E.; Lefevre, S.; Avignon, B. Multispectral Fusion for Object Detection with Cyclic Fuse-and-Refine Blocks. In Proceedings of the 2020 IEEE International Conference on Image Processing, Nanjing, China, 3–5 July 2020; pp. 276–280.
- 64. Loshchilov, I.; Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. arXiv 2016, arXiv:1608.03983.
- 65. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.