Signal Processing: Image Communication 99 (2021) 116460

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Ensemble single image deraining network via progressive structural boosting constraints☆

Long Peng [a], Aiwen Jiang [a,*], Haoran Wei [b], Bo Liu [c], Mingwen Wang [a]

[a] School of Computer and Information Engineering, Jiangxi Normal University, Nanchang, China
[b] Department of Electrical and Computer Engineering, University of Texas at Dallas, Richardson, TX, USA
[c] Department of Computer Science and Software Engineering, Auburn University, Auburn, AL, USA

ARTICLE INFO

Keywords:
Single image derain
Boosting
Ensemble derain network
Rain removal

ABSTRACT

Image deraining is an extensively researched topic in low-level computer vision community. During the past, sufficient works have been proposed to address this problem. Though great improvements these methods have achieved, no derain network can confidently declare that it can solve rain removal problem perfectly. Single complex model may lead to overfitting, while simple model is too weak to achieve clear result. Therefore, in this paper, inspired from classic boosting idea, we have proposed an effective ensemble derain framework to aggregate multiple simple weak drain models to obtain a strong derain model. Cascade structural weighting-map are computed for adaptively emphasizing the quality of local derain regions. Struct-absolution losses are proposed to account for pixel-wise and local region-wise differences, and to facilitate embedding boosting idea into network training. The comprehensive experiments on public derain datasets and high-level vision tasks validate that our proposed model which just utilizes three generally weak derain subnets can achieve much better performance than compared state-of-the-art methods. Our ensemble framework has enough capacity that any state-of-the-art DL-based models can be taken as sub-modules to solve rain removal of multiple types within a single framework.

1. Introduction

Rainy day is one of the most common weather in our daily life. Images captured on rainy condition often suffer poor visual qualities with rain masks. Under different conditions, the rain-masks may appear different styles, such as linear-shape sparse rain-streaks, randomly scattered raindrops and accumulated hazy rain-mist. The image degradation caused by rain-masks brings great challenge to downstream high-level vision tasks, such as video surveillance, object recognition, auto-driving, etc. Therefore, it is important to develop novel and effective rain removal algorithm to restore image's clear content. In recent years, with increasing demands of industrial applications, image derain becomes a very hot research topic in computer vision communities.

During the past, many state-of-the-art methods have been proposed to solve the image derain problems. Though great improvements have been achieved, there are several deficiencies that need paying much attention on. On one hand, many derain methods were criticized that they were specifically designed for only one rain-mask, especially for rain-streak. Assuming degradation too much on one rain case may oversimplify the derain problem. On another hand, since image derain is a challenge ill-posed image restoration problem, in most cases, there does not exist a single derain model that is perfect for rain removal. Single complex derain model may lead to overfitting, resulting content details smoothed, while simple general model may be too weak to remove rain clearly. Therefore, some ensemble strategies are needed to combine methods for different rain-masks or methods considering different rain information aspects, taking their complementary advantages.

In this paper, we follow the ensemble learning strategy, and propose an ensemble single image derain network inspired from classic adaboost [1] idea. As we know, adaboost algorithm adaptively gives much emphasis on wrongly classified samples. For the derain problem, we take pixel-wise value and region's structural consistence as "sample labels". Therefore, we design cascade weighting map computation modules, and construct progressive structural boosting constraints on pre-trained weak derain subnets. Auxiliary struct-absolute losses (SALoss) are calculated on each subnet for fine-tuning. Then coarsely derained results from each "weak" subnet are aggregated and refined to generate final optimal clear output. We have performed comprehensive experiments on two public derain datasets to evaluate the effectiveness our ensemble framework. Though our initial motivation is to validate the boosting thought on derain nets, the experiment results amazingly

☆ This work was supported by the National Natural Science Foundation of China under Grant No. 61966018 and 61876074.
* Corresponding author.
E-mail address: jiangaiwen@jxnu.edu.cn (A. Jiang).

https://doi.org/10.1016/j.image.2021.116460
Received 14 October 2020; Received in revised form 23 April 2021; Accepted 23 August 2021
Available online 4 September 2021

demonstrate that our proposed ensemble model just aggregating three weak derain subnets can achieve more superior performance than state-of-the-art methods compared in this paper.

The contributions of this paper are summarized into three folds.

- We have successfully constructed an effective ensemble single deraining network via progressive structural boosting constraints. The ensemble framework has enough capacity that any state-of-the-art derain networks can be taken as its sub-modules to potentially solve rain removal of multiple types within a single framework.
- We have proposed a struct-absolute loss to consider both pixel-wise and local region-wise differences for network training. As auxiliary loss for fine-tuning, it can benefit for embedding boosting into network training.
- We have experimented our proposed ensemble network on public synthesized derain datasets and real-world rainy images. The experiment results demonstrate that our ensemble boosting derain network can achieve much better performance than many state-of-the-art methods, though we just only utilize three pre-trained "weak" derain models as subnets. On high-level vision tasks, our experimented model can also benefit high-level vision algorithms for reaching almost the same results on clear ground-truth.

In the following sections, we will briefly review some related works in Section 2. Then in Section 3, we will describe our ensemble boosting derain framework in details. In Section 4, we will comprehensively demonstrate our experiments. Finally, in Section 5, we give our conclusion and discus the future work.

2. Related work

Image deraining is an extensively researched topic in the low-level computer vision community. During the past, sufficient works have been proposed to address this problem. Since our proposed ensemble network belongs to deep neural network routine, in this section, we focus on reviewing deep learning based single image derain methods. Much more comprehensive analysis on image derain methods can be referred in recent survey works [2–4].

As we know, the era of deep-learning based deraining started from 2017. According to the existing rain models in literature, there are mainly three kinds of rain-masks, which are linear-shape rain-streaks, scattered raindrops [5] and accumulated rain-mist [6]. They are modeled as a linear superimposition on clear image, blurring the image's content and reducing image's visibility.

However, most of existing deraining works assume much more on rain-streaks. Fu et al. [7] proposed a DerainNet to learn nonlinear mapping between clear and rainy details. Further, they proposed a deep detail network (DDN) [8] to learn negative residual details for rain removal. In their latest work, they proposed to use domain-specific knowledge to simplify the learning process [9].

Yang et al. [10,11] proposed a joint rain detection and removal network (JORDER). Heavy rains, overlapping rain streaks and rain-mist were handled by predicting binary rain mask. A recurrent framework was then taken to remove rain streaks and clear up rain accumulation progressively. Similarly, Li et al. [12] proposed a recurrent squeeze-and-excitation context aggregation net (RESCAN) for image derain. Yang et al. [13] proposed to embed hierarchical representation of wavelet transform into recurrent rain removal process. In the latest, Ren et al. [14] proposed a progressive recurrent network (PReNet) for removing rain-streaks and provided a strong baseline for future studies on image deraining.

Zhang et al. [15] hold a viewpoint that derained image should be indistinguishable from its clear ground-truth. As a result, they proposed image deraining conditional generative adversarial network (ID-CGAN). Further, they proposed a density-aware multi-stream densely connected network (DID-MDN) [16] for joint rain density estimation and deraining. Similarly, to account for rain density information, Li et al. [17] proposed a two-stage derain network to address the gap between physical rain model and real-world rains. In their model, physics-related components, i.e. rain streaks, transmission map and atmospheric light were estimated. Then, a depth-guided GAN was proposed to compensate for lost details and to suppress introduced artifacts.

Through incorporating temporal priors and human supervision, Wang et al. [18] proposed a spatial attentive network (SPANet) to identify and remove rain-streaks in a local-to-global spatial attentive manner. As a concurrent work, Hu et al. [19]formulated a rain imaging model collectively with rain-streaks and fog. They learned depth-attentional features (DAFNet) for single image rain removal.

In the latest, Jiang et al. [20] proposed a multi-scale progressive fusion network (MSPFN) to exploit the correlated information of rain streaks across scales for single image deraining. Wang et al. [21] proposed a rain convolutional dictionary network (RCDNet), attempting to combine advantages of conventional model-driven prior-based and data-driven deep learning-based methodologies. Yasarla et al. [22] proposed a Gaussian Process-based semi-supervised learning framework through incorporating unlabeled real-world images into the training process for better generalization.

Besides rain-streaks, raindrops adhered to camera lens can also severely degrade visibility of the captured image. However, the degradation caused by raindrops are different from the one caused by rain-streak and rain-mist. Eigen et al. [23] attempted to tackle single-image raindrop removal through a three-layer convolution network. It could handle relatively sparse and small raindrops as well as dirt, but failed to produce clean results for large and dense raindrops. Qian et al. [24] proposed an attentive GAN (AttGAN) to make use of contextual information surrounding raindrop areas for restoring raindrop regions.

To above all, though great improvements these methods have achieved, there still exist many deficiencies that need to be solved, such as the commonly criticized issue that monotonous single rain-mask is only considered. Therefore, some ensemble strategies are needed to find optimal combination for solution.

3. Materials and methods

The architecture of our network is demonstrated in Fig. 1. The whole network is composed of two stages, which are *ensemble derain boosting stage* and *refined derain stage*.

At the ensemble derain stage, we follow the idea of boosting to specifically design a cascade structure to organize available rain-mask removing algorithms. At the refined derain stage, we comprehensively reutilize multiple coarsely derain results to finely learn final clear image. In the following subsections, we will describe them in details.

3.1. Ensemble derain boosting stage

The network structure at ensemble derain boosting stage is shown in Fig. 2.

The motivation of the boosting stage is that we assume that none of a single deraining method can solve all rain-mask removal cases perfectly. For example, some deraining methods can solve rain-streak, while some others can only solve rain-drops, and so on. Therefore, we aim to propose a general ensemble framework that can be extended to remove rain-masks of multiple types within the same network.

We firstly pre-train several derain networks, denoted as $I^k_{derain} = f_k\left(I_{original}\right), k = \{1,2,3\}$) are available. Then we combine these networks in a cascade way. Each of the networks progressively performs derain on result from its preceding network.

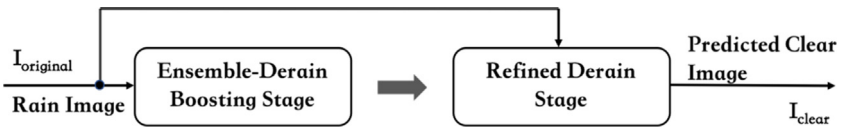

Fig. 1. The architecture of the proposed Ensemble Single Image Derain Network.

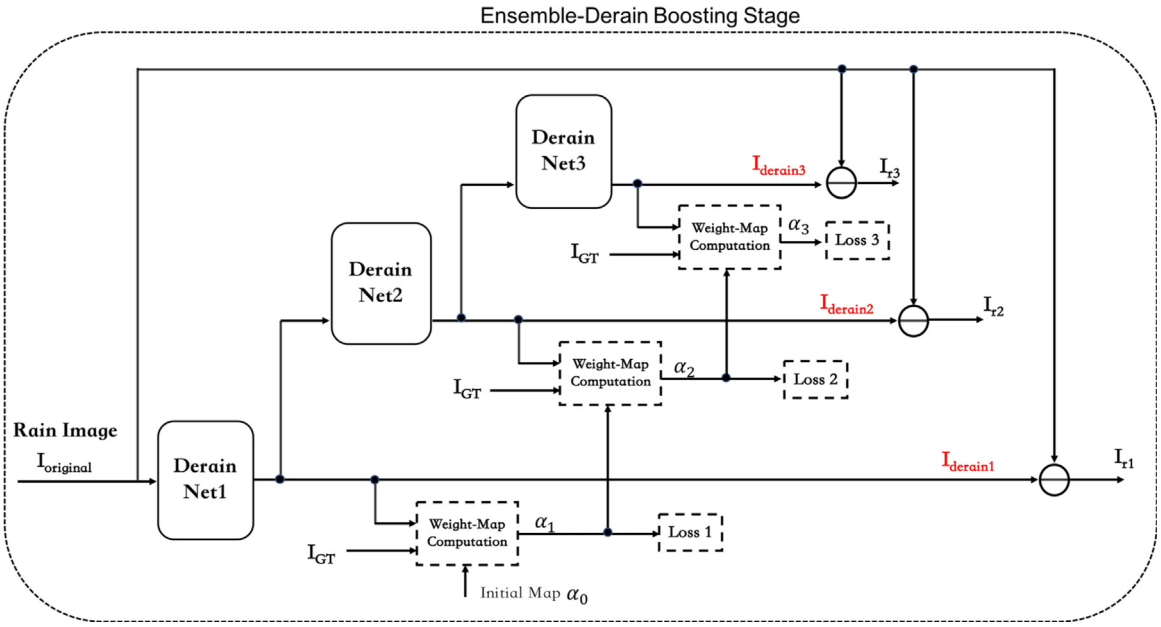

Fig. 2. The network structure at Ensemble-Derain Boosting Stage.

As we know, the idea of boosting in classification topic is to construct strong classifier from multiple weak classifiers through adaptive weighting wrongly-classified samples. In deep network, since parameters learning are greatly affected by the designation of training losses, we propose to embody the boosting thought through cascade weighting-map computation.

Specifically, we design a weight-map computation module based on SSIM (Structural SIMilarity) map. The $\mathtt{SSIMMap}_k$ is computed on intermediate deraining result I^k_{derain} with sliding window size 11×11 and gaussian smoothing operation. The process is the same as [25]. Then a cascade weighting map α_k is computed as Eq. (1). The initial weight α_0 is set all 1's matrix.

$$\alpha_k = \alpha_{k-1} \otimes \mathtt{SSIMMap}_k \tag{1}$$

where, $\otimes$ denote elementary multiplication. The $\mathtt{SSIMMap}_k$ has the same spatial resolution size as image I^k_{derain}. Each element of $\mathtt{SSIMMap}_k(i,j) \in [0,1]$ at position (i,j) represents the local SSIM value of image I^k_{derain}. It spatially reflects the vision qualities of a derain image. Therefore, the α_k weight-map recursively considers all $l^{th}(l \le k)$ intermediate derain results.

During training, we design auxiliary struct-absolute loss for each derain subnetwork f_k as Eq. (2).

$$loss_k = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left|I_{gt} - I^k_{derain}\right| \otimes (1-\alpha_k) \tag{2}$$

where, I_{gt} represents the ground-truth clear image. N and M are image's width and height respectively. The auxiliary losses encourage each subnetwork to learn parameters specifically for local "hard" derain regions, with more emphasis on regions that cannot be dealt well by its preceding networks.

Though we wish each subnetwork f_k could remove rain-mask well, in practice, the derain results are not perfect as we expected.

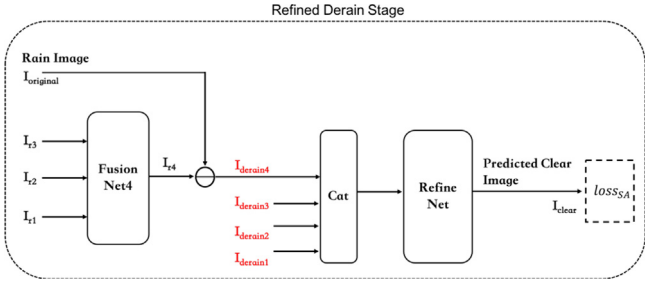

Fig. 3. The network structure at Refined Derain Stage.

The residuals over-removed or under-removed often contain important and non-negligible contents. Therefore, we propose to reutilize these sub-optimal residuals at the following *refined derain stage*. These sub-optimal residuals are computed simply as Eq. (3).

$$I_{rk} = I_{original} - I^k_{derain}, k = \{1,2,3\} \tag{3}$$

3.2. Refined derain stage

The network's structure at refined derain stage is shown in Fig. 3.

Sub-optimal residuals I_{rk} are first composited to learn a synthetical residual I_{r4} through a fusion net g. The learned residual I_{r4} is expected to preserve compositional rain-masks better. Then we can reversely obtain a synthetical derain result I^4_{derain}, as shown in Eq. (4).

$$I_{r4} = g\left(I_{r1}, I_{r2}, I_{r3}\right), \qquad I^4_{derain} = I_{original} - I_{r4} \tag{4}$$

Finally, we concatenate all available derain results I^k_{derain} in channel-wise and employ a refine network Φ to learn final optimal clear

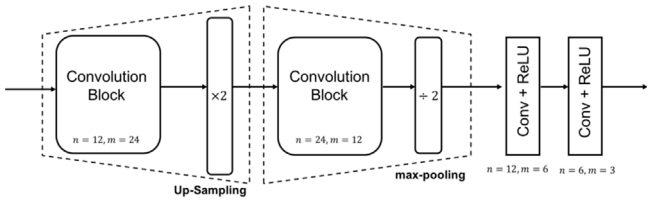

Fig. 4. The structure of refine-net.

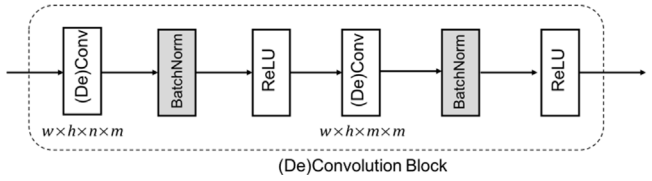

Fig. 5. The structure of Convolution Block.

Table 1
The performance of three derain subnetworks on **Rain1200**.

	Res-Net	U-Net	W-Net
PSNR	24.53	26.40	29.70
SSIM	0.75	0.75	0.78

output. The process is illustrated as Eq. (5).

$$\hat{I}_{clear} = \Phi\left([I^1_{derain}, I^2_{derain}, I^3_{derain}, I^4_{derain}]\right) \tag{5}$$

Herein, the fusion net g is composed of two convolution blocks, each of which consists of a convolution layer followed by a relu activation. The refine network Φ is composed of two convolution blocks followed by two convolution layers. The structural details of refine network is illustrated in Fig. 4.

In our framework, the structure of convolution block is specifically designed as shown in Fig. 5. n and m are the input channel and output channel sizes respectively.

During training, for the final optimal derain result $\hat{I}_{clear}$, we compute the training loss as Eq. (6). The loss computation in Eq. (6) is similar to Eq. (2). We specific name it as **SALoss**.

$$loss_{\mathrm{SA}} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left|I_{gt} - \hat{I}_{clear}\right| \otimes \left(1 - \mathtt{SSIMMap}_{\hat{I}_{clear}}\right) \tag{6}$$

where the $\mathtt{SSIMMap}_{\hat{I}_{clear}}$ computes SSIM map on predicted clear output $\hat{I}_{clear}$.

The final training losses are therefore calculated as Eq. (7).

$$LOSS = loss_{\mathrm{SA}} + \lambda\sum_{k=1}^{3} loss_k \tag{7}$$

where, λ is a super-parameters that weights the auxiliary losses. In this paper, we empirically set $\lambda = 0.01$.

4. Experiments

4.1. Experiment settings

For simplicity, Res-Net, U-Net and W-Net [26] are taken as sub-modules of our ensemble framework. Their architectures are illustrated in Fig. 6. However, we still would like to emphasize that our ensemble framework are with enough capacity that any SOTA derain networks can be taken as our sub-modules.

We train and test our derain networks on two publicly available synthesized datasets. The first one is **Rain800** from ID-CGAN dataset [15]. It contains about 700 training sets and 100 test sets. The second one

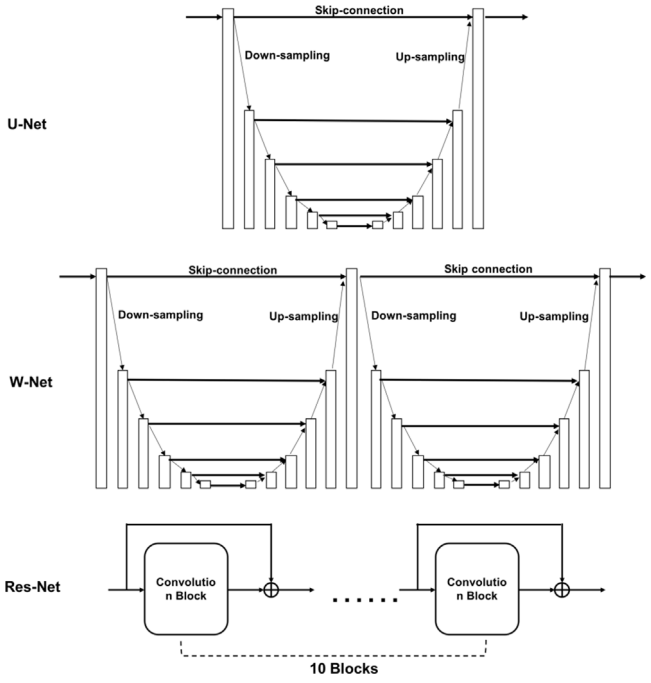

Fig. 6. The architectures of experimented Derain Sub-Networks.

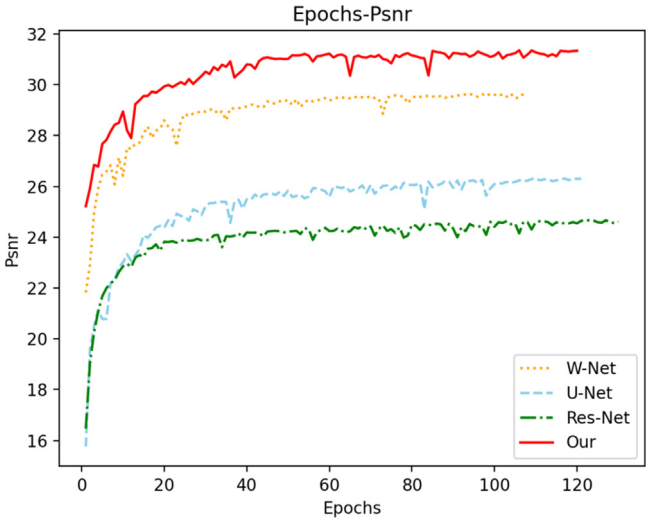

Fig. 7. Testing curves during training phase on **Rain1200**.

is **Rain1200** from DID-MDN dataset [16]. It contains about 12000 training images and 1200 test images.

For fair comparison, the splitting ways of both datasets follow strictly as the same as the original papers, so that all methods can be compared in the same experiment settings. Specifically, for **Rain800**, the training set consists of a total of 700 images, where 500 images are randomly chosen from the first 800 images in the UCID dataset and 200 images are randomly chosen from the BSD-500's training set. The test set consists of a total of 100 images, where 50 images are randomly chosen from the last 500 images in the UCID dataset and 50 images are randomly chosen from the test-set of the BSD-500 dataset. For **Rain1200**, the training set consists of 4000 images respectively with heavy, middle, light rain densities. The test set consists of 400 images respectively with heavy, middle, light rain densities.

Moreover, to evaluate the generalization of the proposed model, a real-world rainy scene dataset is also utilized for testing. The real-world

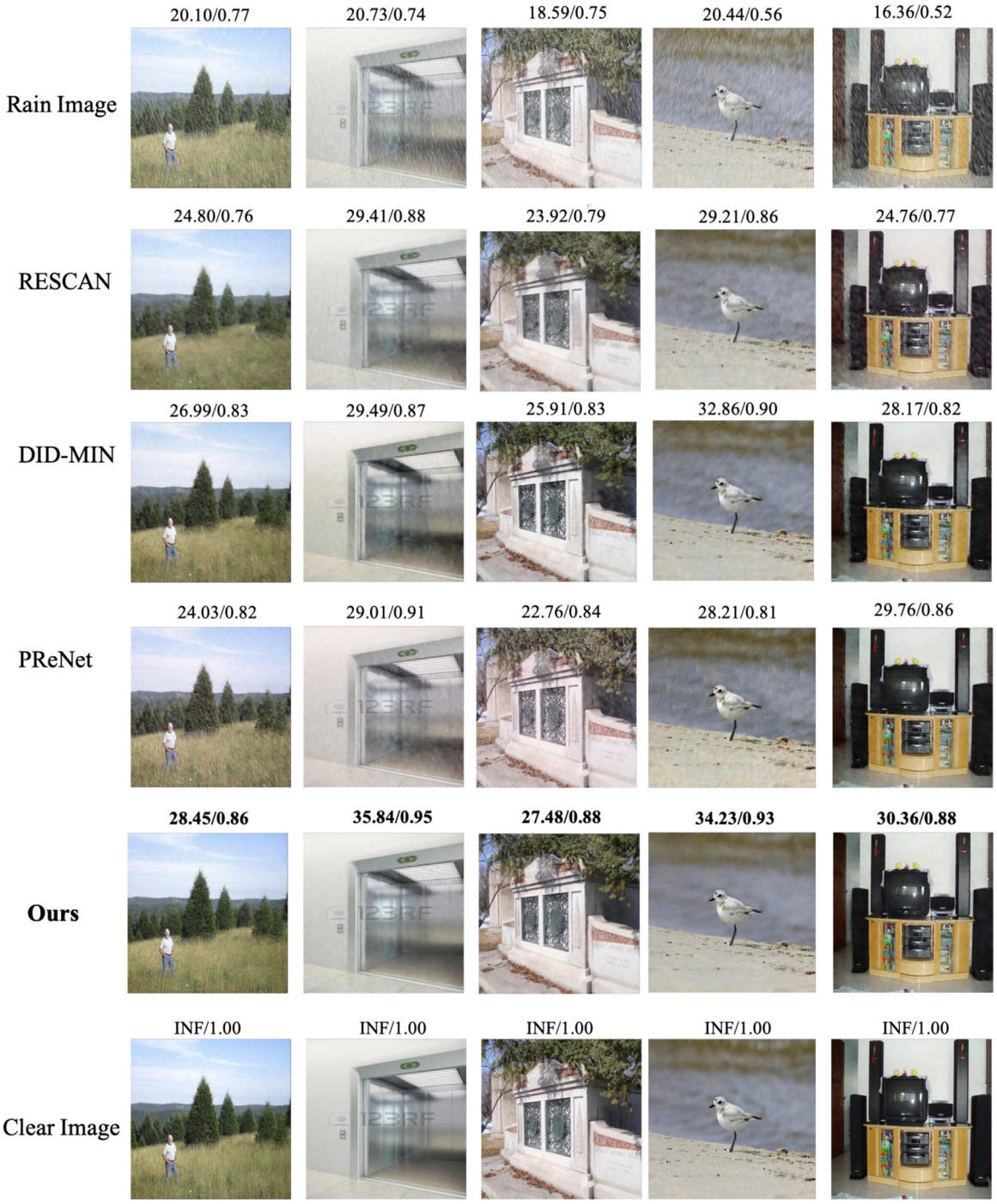

Fig. 8. Visual comparisons on synthesized data.

dataset is randomly collected by ourselves from Google website. 500 rainy images having similar resolution with our training images and good visual qualities are retained after data cleaning. PSNR and SSIM are employed as metrics to evaluate output image's quality from synthesized datasets. Non-reference image quality criterions like PIQE [27], NIQE [28] and Brisque [29] are employed to evaluate the quality of output on real-world rainy dataset.

Our networks are implemented by PyTorch on a piece of NVIDIA GTX 1080Ti. During training, we adopted batch size of 8. All convolution layers in our framework employ convolution kernels with spatial size of 3×3. Except special description, all convolution layers do not change the spatial size of input feature map. Adam optimizer is accepted to learn our network's parameters. Other related training details and network structural details are available in our public source codes at https://github.com/peylnog/model.

4.2. Comparisons with state-of-the-arts

We compare our proposed framework with several representative SOTA methods: DSC [30] and GMM [31] are traditional optimization-based methods; DerainNet [7], DDN [8], JORDER [11], JBO [32], DID-MDN [16], RESCAN [12], ID-CGAN [15], PReNet [14] and Syn2real [22] are popular deep network based methods.

We firstly pre-train the three previously mentioned derain sub-networks on **Rain1200**. The loss use for pre-training these sub-

Table 2
Quantitative performance comparisons on **Rain800**.

	PSNR	SSIM
Rain	21.16	0.65
DualCNN	21.22	0.82
GMM	22.27	0.74
RESCAN	24.09	0.84
DID-MDN	21.89	0.80
JORDER	21.09	0.75
DDN	23.23	0.82
PreNet	24.90	0.85
ID-CGAN	24.34	0.84
Sys2real	25.75	0.86
Ours	**25.81**	**0.86**

Table 3
Quantitative performance comparisons on **Rain1200**.

	PSNR	SSIM
Rain	21.15	0.78
DSC	21.44	0.79
GMM	22.75	0.84
DerainNet	22.07	0.84
RESCAN	29.10	0.88
JORDER	24.32	0.86
DID-MDN	27.95	**0.91**
JBO	23.05	0.85
PreNet	29.12	0.87
Sys2real	30.10	0.91
Ours	**31.31**	0.90

Table 4
Quantitative performance comparisons on real-world rain image.

	PIQE	NIQE	Brisque
PreNet	14.59	11.86	0.4988
RESCAN	14.90	10.72	**0.5142**
Ours	**14.41**	**10.47**	0.5079

networks is **MSELoss**. The initial learning rate is set to be 1e-4. The coarsely derain performances are shown in Table 1.

Then, we retrain our holistic ensemble framework with global learning rate set to be 1e-4, besides 5e-5 for fine-tuning pre-trained modules. In order to obtain sufficient training convergence, we train the networks for 500 epochs. The test curves during training phase on **Rain1200** dataset are drawn in Fig. 7.

The final performance comparisons are recorded in Tables 2 and 3.

From the experimental results, we can validate the effectiveness of our ensemble boosting network when compared with its independent subnetworks. Especially on SSIM index, our ensemble network boosts the performance more than 15.4% on **Rain1200**.

More examples for visual comparison are shown in Fig. 8. And Some SSIM Maps are visualized for more detailed illustration are shown in Fig. 9. From the visualization results, we can observe that our ensemble network can remove rain-mask with visual pleasing appearances.

We further perform comparisons on our collected real-world dataset. We directly test our ensemble network and other two top performance methods (PReNet and RESCAN) on real-world dataset to evaluate their generalization abilities. The performance comparisons are shown in Table 4. More visual comparisons on real rainy scenes are demonstrated in Fig. 10.

From the test results on real scenes, we can also validate that our ensemble network has better generalization ability.

In addition, We compare parameters size and processing time on some typical methods such as RESCAN, JORDER and Syn2Real. The experiment results are shown in Table 5. RESCAN has less parameters since many parameters are shared among blocks in its recurrent framework. However, its average processing time still costs much higher than ours. All these results demonstrate that our ensemble network has more competitive efficiency with moderate model size.

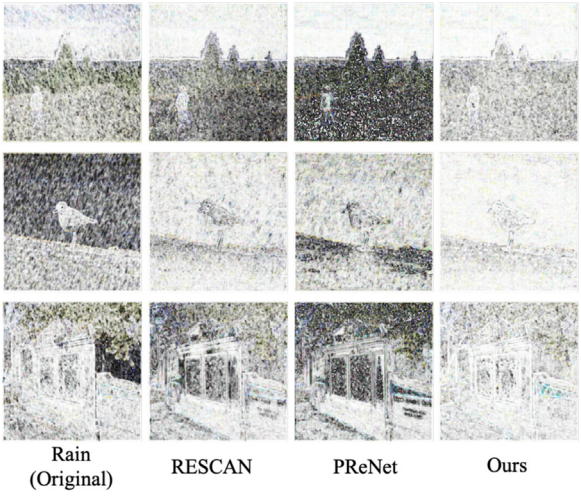

Fig. 9. SSIM Maps for some derained images. The value at each pixel position on SSIM Map represents the structure similarity quality of corresponding local region. The lighter the map pixel is, the better its local structure is preserved.

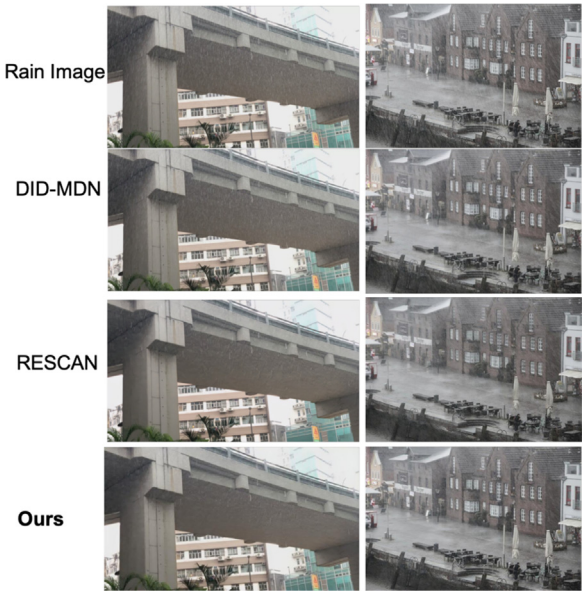

Fig. 10. Visual Comparisons on real scenes.

Table 5
Parameters size and required precessing for test. The processing time are averagely estimated on 1200 test images.

	Parameters	Time (s)
RESCAN	499,668	0.19
JORDER	4,169,024	2.3
Syn2real	2,618,294	0.21
Ours	1,344,837	0.08

4.3. Ablation studies

In order to validate the rationalities of our selections on loss computation, subnets' order, super-parameters, we perform several ablation studies. All the ablation studies are conducted on **Rain1200**.

4.3.1. The number of selected subnets

We have empirically experimented three cases of different subnet combinations in the ensemble framework. The combinations and

Table 6
Ablation study: different subnet combinations.

W-Net	U-Net	ResNet	DenseNet	PSNR	SSIM
✓	✓	–	–	30.13	0.88
✓	✓	✓	–	31.31	0.90
✓	✓	✓	✓	31.40	0.90

Table 7
Ablation study: The comparisons between MSELoss and SALoss.

	PSNR	SSIM
Our ensemble network with MSELoss	31.25	0.89
Our ensemble network with SALoss	**31.31**	**0.90**
W-Net with MSELoss	29.70	0.78
W-Net with SALoss	**29.74**	**0.79**

Table 8
Ablation study: different SALoss computations.

	PSNR	SSIM
$E(x)E(y)$	31.11	0.89
$E(xy)$	**31.31**	**0.90**

Table 9
Ablation study: The effects of subnets' order at boosting derain stage.

	PSNR	SSIM
Ascending	31.18	0.89
Descending	**31.31**	**0.90**

Table 10
Ablation study: The influence of super-parameter λ on **Rain1200**.

	PSNR	SSIM
$\lambda = 0.00$	30.87	0.89
$\lambda = 0.01$	**31.31**	**0.90**
$\lambda = 0.10$	30.75	0.89

their corresponding performances are compared in Table 6. Herein, besides previously mentioned W-Net, U-Net and ResNet, we pretrained DenseNet for derain. The coarsely derain performance of DenseNet is 27.55 for PSNR and 0.78 for SSIM.

From the ablation study, we observe the performance is improved greatly from a framework with 2 subnets to the one with 3 subnets. However, the improvement becomes negligibly when the number of subnet is increased to 4. Therefore, in our ensemble network, we suggest 3 subnets as we have done in consideration of the balance between model's complexity and performance.

4.3.2. Loss selection: MSELoss vs. SALoss

For neural network training, MSELoss is a widely adopted loss. Its computation is like as Eq. (8) shows.

$$loss_{\mathrm{MSE}} = \sqrt{\left| I_{gt} - \hat{I}_{clear} \right|^2} \tag{8}$$

The MSELoss globally emphasize much more on pixel differences. However, spatial structural differences are lost in its computation. In our SALoss computation in Eq. (6), both pixel-wise and local structure-wise differences are considered.

In order to validate the effectiveness of SALoss, an ablation study is performed by replacing SALoss with MSELoss during training. The testing results are shown in Table 7. From the results, though a little small the improvements are, the trend of improvements is consistent, which shows the effectiveness of SALoss.

4.3.3. SALoss computation: $E(xy)$ vs. $E(x)E(y)$

We have considered two kinds of computation processes for SALoss. One computation is like the way Eq. (6) does, performing element-wise multiplication before average. The other one is as following Eq. (9), doing global average before multiplication.

$$\begin{aligned} loss_{\mathrm{SA2}} = & \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \left| I_{gt} - \hat{I}_{clear} \right| \\ & \times \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \left(1 - \mathtt{SSIMMap}_{\hat{I}_{clear}} \right) \end{aligned} \tag{9}$$

We record the performances of models using different loss computation on Table 8, where "$E()$" represents average (expectation) operation.

The results in Table 8 validate that the loss computation "$E(xy)$" way in Eq. (6) is better.

4.3.4. The order of SubNets in ensemble framework

We perform an ablation study to discover the influences that the subnets' organization order at *boosting derain stage* pose on final derain performance. We organize subnets on *Ascending* order and *Descending* order according to their independent performances. The experiment results are shown in Table 9.

From the experiments, we can find that the better organization way for derain subnets f_k is in *descending* way according to their independent performances. This conclusion is also inline with the standard process that traditional boosting-related algorithm selects the best weak classifier according to their independent classification errors.

4.3.5. The influence of super-parameter λ

The super-parameter λ weights the auxiliary losses $loss_k$ on SALoss. In order to find the influence of different weight selection on final performance, we experiment several typical cases. The results are shown in Table 10.

From the experimental results, we can find that when $\lambda = 0.10$, the trained ensemble network achieves the best performance. We therefore believe that the best λ is around 0.01, and set $\lambda = 0.01$ as our optimal super-parameter.

4.4. High-level computer vision tasks on derain image

Image derain, as a low-level vision task, aims to remove rain-mask and enhance the vision quality of rain image. However, its ultimate goal is to service for down-stream high-level computer vision tasks, such as object detection, semantic segmentation, instance segmentation in rainy environments.

We therefore employ the popular Lw-RefineNet [33] for semantic segmentation, YOLOv3[34] for object detection, and Mask-RCNN [35] for instance segmentation on derain images. We select and compare our model with several top-performance derain methods on **Rain1200**. The results on high-level vision tasks are illustrated in Fig. 11.

From the visual comparison results, it is not difficult to find that for both semantic and instance segmentation tasks, high-level vision algorithms achieve very close results on our derain images to "ground-truth" labels on clear image. For detection task, YOLOv3 detects more objects on our derain images than on other derain images. Therefore, our ensemble network can outperform most state-of-the-art methods through boosting several weak derain networks.

5. Conclusion

In this paper, we have proposed an effective ensemble single image derain framework inspired from classic boosting idea. Cascade structural weighting-map are computed for adaptively emphasizing the quality of local derain regions. Struct-absolution losses are proposed to account for pixel-wise and local region-wise differences, and to facilitate embedding boosting idea into network training. Finally,

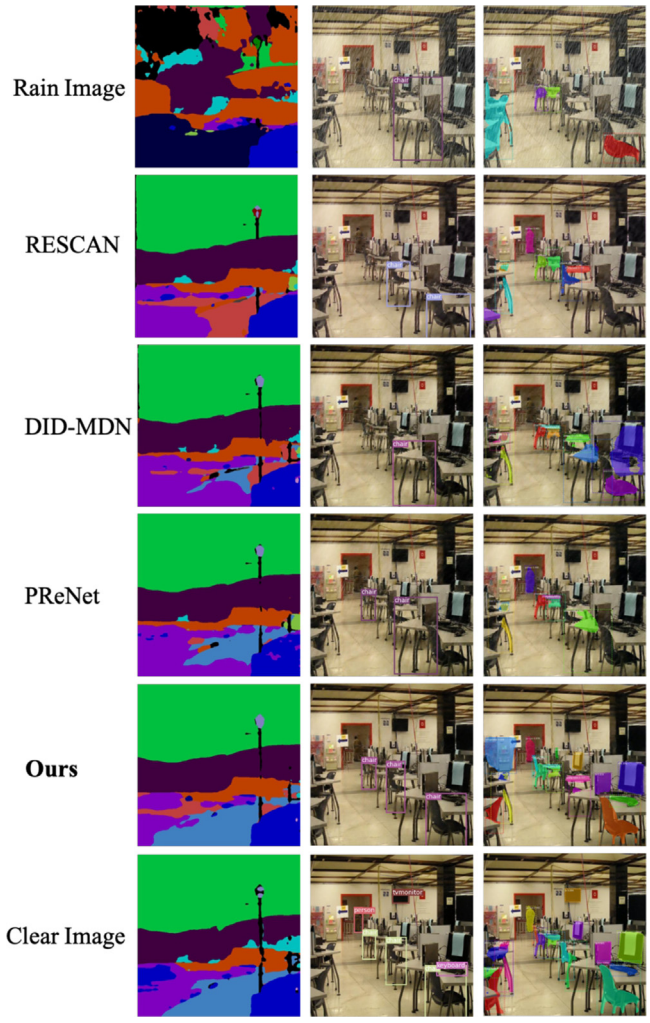

Fig. 11. Visual comparisons of rain-streak removal on High-level Computer Vision Tasks. The first column denotes semantic segmentation results of Lw-RefineNet. The second column shows the object detection results by YOLOv3. We use Mask R-CNN for instance segmentation and the results are shown in the last column.

coarsely derained results from each subnet are aggregated and refined to generate final optimal clear output. We have experimented our proposed network on public derain datasets and high-level vision tasks. The experiment results validate that our boosting idea on drain model's aggregation is effective. Our proposed network illustrated in this paper outperforms many state-of-the-art methods, through just utilizing three simple derain subnets.

In the future, we will further strength the explainability of the boosting derain process and give more comprehensive validation on high-level vision tasks.

CRediT authorship contribution statement

Long Peng: Design, Formal analysis, Interpretation of data for the work. **Aiwen Jiang:** Conceptualization, Design, Funding acquisition, Formal analysis, Interpretation of data for the work, Writing – original draft, Writing – review & editing. **Haoran Wei:** Formal analysis. **Bo Liu:** Formal analysis, Interpretation of data for the work. **Mingwen Wang:** Conceptualization, Funding acquisition, Formal analysis.

Declaration of competing interest

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

References

[1] Yoav Freund, Robert E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. System Sci. 51 (1) (1997) 119–139.
[2] S. Li, I.B. Araujo, W. Ren, Z. Wang, E.K. Tokuda, R.H. Junior, R. Cesar-Junior, J. Zhang, X. Guo, X. Cao, Single image deraining: A comprehensive benchmark analysis, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 3833–3842.
[3] Hong Wang, Yichen Wu, Minghan Li, Qian Zhao, Deyu Meng, A survey on rain removal from video and single image, 2019, arXiv:1909.08326.
[4] Wenhan Yang, Tan Robby T., Shiqi Wang, Yuming Fang, Jiaying Liu, Single image deraining: From model-based to data-driven and beyond, 2019, arXiv:1912.07150.
[5] S. You, R.T. Tan, R. Kawakami, Y. Mukaigawa, K. Ikeuchi, Adherent raindrop modeling, detectionand removal in video, IEEE Trans. Pattern Anal. Mach. Intell. 38 (9) (2016) 1721–1733.
[6] Zheng Zhang, Wu Liu, Huadong Ma, Xinchen Liu, Going clear from misty rain in dark channel guided network, in: IJCAI Workshop on AI for Internet of Things, 2017.
[7] Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, John Paisley, Clearing the skies: A deep network architecture for single-image rain removal, IEEE Trans. Image Process. 26 (6) (2017) 2944–2956.
[8] Xueyang Fu, Jiabin Huang, Delu Zeng, Huang Yue, Xinghao Ding, John Paisley, Removing rain from single images via a deep detail network, in: Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2019.
[9] Xueyang Fu, Borong Liang, Yue Huang, Xinghao Ding, John Paisley, Lightweight pyramid networks for image deraining, IEEE Trans. Neural Netw. Learn. Syst. 31 (6) (2020) 1794–1807.
[10] Wenhan Yang, Tan Robby T., Jiashi Feng, Jiaying Liu, Zongming Guo, Shuicheng Yan, Deep joint rain detection and removal from a single image, in: Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017.
[11] Wenhan Yang, Tan Robby T., Jiashi Feng, Jiaying Liu, Shuicheng Yan, Zongming Guo, Joint rain detection and removal from a single image with contextualized deep networks, IEEE Trans. Pattern Anal. Mach. Intell. (2019) http://dx.doi.org/10.1109/TPAMI.2019.2895793.
[12] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, Hongbin Zha, Recurrent squeeze-and-excitation context aggregation net for single image deraining, in: European Conference on Computer Vision, Springer, 2018, pp. 262–277.
[13] Wenhan Yang, Jiaying Liu, Shuai Yang, Zongming Guo, Scale-free single image deraining via visibility-enhanced recurrent wavelet learning, IEEE Trans. Image Process. 28 (6) (2019) 2948–2961, http://dx.doi.org/10.1109/TIP.2019.2892685.
[14] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, Deyu Meng, Progressive image deraining networks: A better and simpler baseline, in: Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019.
[15] He Zhang, Vishwanath Sindagi, Vishal Patel, Image de-raining using a conditional generative adversarial network, IEEE Trans. Circuits Syst. Video Technol. (2019) http://dx.doi.org/10.1109/TCSVT.2019.2920407.
[16] He Zhang, Vishal M. Patel, Density-aware single image de-raining using a multi-stream dense network, in: Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA.
[17] Ruoteng Li, Loong-Fah Cheong, Tan Robby T., Heavy rain image restoration: Integrating physics model and conditional adversarial learning, in: Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019.
[18] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Rynson Lau, Spatial attentive single-image deraining with a high quality real rain dataset, in: Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA.
[19] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, Pheng-Ann Heng, Depth-attentional features for single-image rain removal, in: IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019.
[20] Kui Jiang, Zhongyuan Wang, Peng Yi, Baojin Huang, Yimin Luo, Jiayi Ma, Junjun Jiang, Multi-Scale Progressive Fusion Network for Single Image Deraining, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
[21] Hong Wang, Qi Xie, Qian Zhao, Deyu Meng, A model-driven deep neural network for single image rain removal, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[22] Rajeev Yasarla, Vishwanath A. Sindagi, Vishal M. Patel, Syn2Real transfer learning for image deraining using Gaussian processes, in: The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
[23] D. Eigen, D. Krishnan, R. Fergus, Restoring an Image Taken through a Window Covered with Dirt or Rain, in: 2013 IEEE International Conference on Computer Vision, 2013, pp. 633–640.
[24] R. Qian, R.T. Tan, W. Yang, J. Su, J. Liu, Attentive generative adversarial network for raindrop removal from A single image, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 2482–2491.
[25] Z. Wang, Q. Li, Information content weighting for perceptual image quality assessment, IEEE Trans. Image Process. 20 (5) (2011) 1185–1198.
[26] L. Peng, A. Jiang, Q. Yi, M. Wang, Cumulative rain density sensing network for single image derain, IEEE Signal Process. Lett. 27 (2020) 406–410.
[27] N. Venkatanath, D. Praneeth, Bh. Maruthi Chandrasekhar, S.S. Channappayya, S.S. Medasani, Blind image quality evaluation using perception based features, in: 2015 twenty first National Conference on Communications (NCC), 2015, pp. 1–6.
[28] A. Mittal, R. Soundararajan, A.C. Bovik, Making a "completely blind" image quality analyzer, IEEE Signal Process. Lett. 20 (3) (2013) 209–212.
[29] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.
[30] Y. Luo, Y. Xu, H. Ji, Removing rain from a single image via discriminative sparse coding, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 3397–3405.
[31] Y. Li, R.T. Tan, X. Guo, J. Lu, M.S. Brown, Rain Streak Removal Using Layer Priors, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2736–2744.
[32] L. Zhu, C. Fu, D. Lischinski, P. Heng, Joint bi-layer optimization for single-image rain streak removal, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2545–2553.
[33] Ian Reid Vladimir Nekrasov, Chunhua Shen, Light-weight RefineNet for real-time semantic segmentation, in: British Conference on Machine Vision, 2018.
[34] Joseph Redmon, Ali Farhadi, Yolov3: An incremental improvement, 2018, arXiv: 1804.02767.
[35] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2) (2020) 386–397.

Long Peng is an undergraduate student in School of Computer and Information Engineering, Jiangxi Normal University. His research interest is image enhancement.

Aiwen Jiang is full professor in School of Computer and Information Engineering, Jiangxi Normal University. He received his PhD. degree from Institute of Automation, Chinese Academy of Sciences, China, in 2010. He received his bachelor degree from Nanjing University of Post and Telecommunication, China, in 2005. His research interests are computer vision, machine learning.

Haoran Wei is a PhD candidate in the Department of Electrical and Computer Engineering at the University of Texas at Dallas, Richardson, TX. He received his BE degree in communication engineering from Shanghai Normal University, Shanghai, China, in 2014, and the MS degree in communication and information system from Shanghai Normal University, Shanghai, China, in 2017. His research interests include real-time image processing, video processing, speech processing, and machine learning.

Bo Liu is current a tenure-track assistant professor in College of Computer Science and Software Engineering at Auburn University, Auburn. He received his Ph.D. degree in computer science from University of Massachusetts Amherst, Massachusetts, in 2015. He is the recipient of the UAI-2015 Facebook Best Student Paper Award. His primary research area covers machine learning, deep learning, stochastic optimization.

Mingwen Wang is full professor and director in School of Computer and Information Engineering, Jiangxi Normal University. He received his PhD degree from Shanghai Jiaotong university, China, in 2001. His research interests are machine learning.