METHODOLOGIES AND APPLICATION

Deep neural de-raining model based on dynamic fusion of multiple vision tasks

Yulong Fan[1] · Rong Chen[1] · Yang Li[1] · Tianlun Zhang[1]

Abstract

Image quality is relevant to the performance of computer vision applications. The interference of rain streaks often greatly depreciates the visual effect of images. It is a traditional and critical vision challenge to remove rain streaks from rainy images. In this paper, we introduce a deep connectionist screen blend model for single-image rain removal research. The novel deep structure is mainly composed of shortcut connections, and ends with sibling branches. The specific architecture is designed for joint optimization of heterogeneous but related tasks. In particular, a feature-level task is design to preserve object edges which tend to be lost in de-rained images. Moreover, a comprehensive image quality assessment is an additional vision task for further improvement on de-rained results. Instead of using rules of thumb, we propose an actionable method to dynamically assign appropriate weighting coefficients for all vision tasks we use. On the other hand, various factors such as haze also give rise to weak visual appeal of rainy images. To remove these adverse factors, we develop an image enhancement framework which enables the hyperparameters to be optimized in an adaptive way, and efficiently improves the perceived quality of de-rained results. The effectiveness of the proposed de-raining system has been verified by extensive experiments, and most results of our method are impressive. The source code and more de-rained results will be available online.

Keywords Deep neural network · Single-image de-raining · Screen blend model · Multi-task learning · Dynamic scheme · Evolutionary algorithm

✉ Rong Chen
rchen@dlmu.edu.cn

Yulong Fan
sumihui@dlmu.edu.cn

Yang Li
li_yang@dlmu.edu.cn

Tianlun Zhang
threekingdomst@163.com

[1] College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China

1 Introduction

In the past few years, an abundant literature devoted to the image recovery under bad weather (Zhao et al. 2015; He et al. 2011; Wang and Yuan 2017; Yang et al. 2019; Fu et al. 2017; Zhang and Patel 2018; Li et al. 2018b; Fu et al. 2019; Yang et al. 2017). Among these studies, the problem of rain removal has drawn a lot of attention (Fu et al. 2017; Zhang and Patel 2018; Li et al. 2018b; Fu et al. 2019; Yang et al. 2017; Li et al. 2018a). With rain streaks, the visibility of scene content tends to be drastically degraded. When suffering from images with visual quality decline, most outdoor vision systems, such as surveillance and autonomous navigation, fail to provide favorable performance. Therefore, it becomes critical to develop effective approaches for removing rain streaks from rainy images.

For improvement of perceived quality, the focus in de-raining researches is to decompose the rain-free background layer from a given rainy image. Obviously, this layer separation task is an inherently ill-posed problem as one rain-free layer can have connection with multifarious rainy scenarios.

Moreover, when rain streaks are denser, this connection will become more ambiguous since fine details of the background scene have little or no evidence in the corresponding rainy version. Therefore, in the absence of additional images or rich temporal information (Kim et al. 2015; Santhaseelan and Asari 2015; You et al. 2015), the rain removal on single image is an extremely challenging vision issue.

To make this issue well posed, numerous worthwhile explorations have been conducted for single-image de-raining purpose. On one hand, inspired by a statistical knowledge that the rainy patches have high absolute gradient, the most intuitive solution is to smooth the rain streaks through existing noise removal techniques such as linear smoothing (Kim et al. 2013; Zhang and Xiong 2009) and total variation regularization (Rudin et al. 1992). These simple methods can smooth away rain streaks in flat image regions, however, the details in background scene tend to be over smoothed or lost. On the other hand, an observed rainy image is characterized as a linear superimposition of the background layer and the rain streak layer. Based on this imaging model, many discriminative de-raining works use some conventional techniques (e.g., morphological component analysis Kang et al. 2012 and dictionary learning Li et al. 2016) to learn the distribution characteristics of rain streaks, and then distinguish object edges from spurious details caused by raindrops. These approaches can preserve background details to a greater degree, however, they often fail to detect and remove rain streaks because the heuristic cues and strong assumptions are less effective for some natural rainy situations, particularly in heavy rain, the rain streaks possibly have complicated statistical characteristics due to the various orientations, shapes, and densities. In addition, the rainy images modeled by linear superimposition can lose some crucial characteristics of real rainy scene, such as the appearance of internal reflections. And the linear model is sensitive to the intensity of illumination. When suffering from intense light, the methods based on this linear model tend to confuse the rain streaks and white edges in background. In addition, most de-raining methods ignore the atmospheric veils caused by rain streaks, leading to unfavorable visual quality of de-rained results.

Instead of predefined assumptions and priors, we propose a novel rain removal model based on deep learning method, which can adaptively learn the structural and contextual information in a data-driven manner. In recent years, deep learning has achieved significant successes in wide-ranging vision tasks (He et al. 2017; Girshick 2015; Simonyan and Zisserman 2015; Cai et al. 2016; Krizhevsky et al. 2012). Yet appealing to de-raining tasks, there are several crucial issues in developing such a connectionist de-raining system. First, in-appropriate model will make deep network under-fit the observation, a simple linear model usually leads to a paucity of common properties of rain streaks. To prop-

erly formulate the rain image, we instead propose to use a nonlinear model rendering the appearance of natural rainy scenes more faithfully. So the rain removal model can be robust for complicated rainy situations, such as the rain streak accumulation. Second, the learning of different but related tasks can improve the generalization ability of deep neural network. Thus, we develop the de-raining model within a multi-task learning framework. The follow-up problem is to associate different tasks with appropriate loss weights, which play an important role in model optimization. To address this problem, we extend evolutionary algorithm for adaptive weight assignment. In addition, sometimes a veiling effect may survive in de-rained results since accumulated raindrops in distant scenes lead to blurry vision in a manner similar to haze (Yang et al. 2017). Based on these key problems, our focus is to conduct some further investigations on degrading the visual effect of rain streaks. Figure 1 shows the proposed deep screen blend network that is particularly suitable to de-raining tasks. There are three main reasons. First, the screen blend model is a robust nonlinear composite representation for rainy image. Driven by this model, the proposed network can effectively learn much more features from rain observation. Second, at the end of the deep network, one branch structure is designed specifically for decomposing a rainy image into the corresponding rain layer (i.e., the *Rain-streaks*) and the background layer (i.e., the *Derain image*). Based on this structure, the rain extraction and removal can be learned in a mutually reinforced way. Last, based on the perceptual information, one edge-aware regularization (i.e., *L-perceptual*) is proposed for the detail preserving, which can avoid the over-smoothed results in de-raining task. Since water-droplets often create haze in rainy weather, haze removal is necessary to ensure higher visual quality. To this end, we further restore the rain removal results using a self-adaptive enhancement method. More concretely, our work can be concluded as follows.

(1) In this work, a novel tasks-constrained de-raining network with sibling branches is formulated to jointly learn the distributive characteristics of rain layer and background scene. Instead of the linear imaging model in existing neural rain removal methods, the objective function of the proposed deep network is developed on screen blend model, which is a more robust nonlinear system to model rainy conditions.

(2) Aside from pixel-wise tasks, two feature-level tasks are conducted in this work. In the first task, we employ a set of isotropic image gradient operators as the filter kernels to construct a perceptual loss model, which facilitates the edge preservation for de-rained results. It is the first endeavor of such technique for de-raining work. On the other hand, motivated by salient image features (i.e., luminance, contrast, and structure), we introduce another
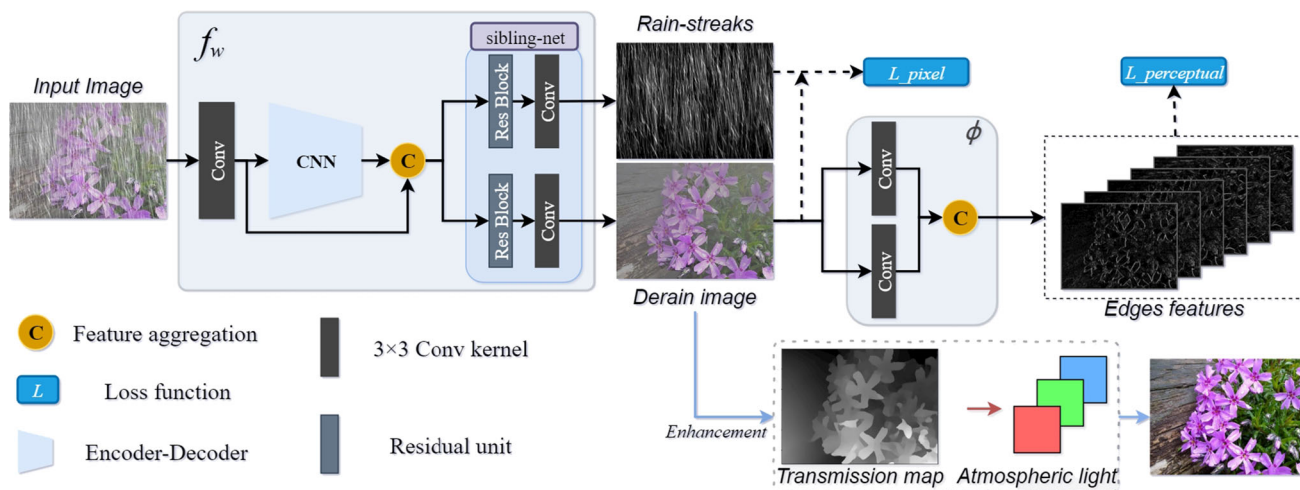
Fig. 1 Flowchart of entire system. The light blue area is the proposed de-raining model denoted as f_w. And the perceptual network model is ϕ. Arrows represent the main data flow of feed-forward (color figure online)

vision task to measure the feature difference between the reference image and the restored result. Moreover, instead, the optimization process of our model is to reach the minimum Pseudo-Huber Loss (PHL) (Charbonnier et al. 1997) rather than minimizing the Mean Square Error (MSE) which is widely deployed in existing de-raining works.

(3) Without the need of prior knowledge about weighting factors, we show that these vision tasks mentioned above can be optimized in a self-adaptive manner. In addition, to improve the image quality of de-rained result, we propose an adaptive method to assign appropriate hyperparameters for a post-processing framework of haze removal. In this work, these adaptive methods are implemented by solving multi-objective optimization problems through a population-based genetic algorithm (Qu et al. 2012).

The rest of this paper is organized as follows. In Sect. 2, we review some basic concepts related to our principal models. In Sect. 3, we present details of the proposed de-raining model. In Sect. 4, comprehensive experiments are presented. At last, the conclusion and future work are discussed in Sect. 5.

2 Background

In computer vision community, numerous researches have devoted attention to weaken the undesirable effect of rain streaks on vision quality (Narasimhan and Nayar 2003). Traditionally, image de-raining is considered as a noise filtering problem (Khmag et al. 2019), in which the distributive characteristics of rain streaks are different from the background scene in rainy images. Kang et al. (2012) devised an early rain

removal model based on image decomposition. From their views, the rain streaks can be considered as a part of the high-frequency information in rainy images, their solution thus is to use a bilateral filter to separate out the high-frequency component from a single rainy image, and use dictionary learning to decompose the high-frequency component into rainy parts and rain-free parts. On the other hand, in Luo et al. (2015), dictionary learning is utilized to provide discriminative sparse codes to approximate the local regions of rain and background layer. Apart from dictionary learning (Luo et al. 2015), low-rank prior structure also has been applied in several rain removal models (Chen and Hsu 2013). However, Li et al. (2016) argued that the de-raining methods based on dictionary learning and low-rank structure tend to result in a lot of remaining rain streaks or over smoothed details in background. To effectively improve the de-raining performance, they developed layer priors on each patch of rain and background. The Gaussian mixture model (GMM) (Li et al. 2016) as prior is employed to, respectively, learn the characteristics of rain and background in a patch-wise way. The superiority of GMM-based priors has been shown in boosting the overall visibility under complex rainy conditions.

From the discussion above, one can easily note that a challenging aspect of rain removal research is the constant struggle for extracting characteristics from rainy and clean layers. In this case, deep learning is an appealing solution to acquire rich features in a data-driven manner (Chen and Liu 2017). Recently, several researches therefore focused on connectionist de-raining model (Fu et al. 2017; Yang et al. 2017; Zhang and Patel 2018). In Fu et al. (2017), rain streak also is treated as a type of high-frequency component (Kang et al. 2012), so a well-designed convolutional neural network (CNN Krizhevsky et al. 2012) is employed to decompose rain

streaks on the high-pass detail layer, then the sum of rain-free detail layer and low-pass rainy image is produced as the final de-rained result. Typically, the high-frequency components are sparse matrices which can speed the convergence of deep learnings.

Meanwhile, Yang et al. (2017) conducted a rain removal research by using mask layer and cascaded CNN which are the prevalent techniques in semantic segmentation (He et al. 2017). This is the first work using rain masks to locate the rain streaks. And the cascaded CNN successively performs rain detection, estimation, and removal. In a follow-up study, the authors further proposed an enhanced version (Yang et al. 2019), in which an extra detail preserving step is introduced. The latest connectionist de-raining models were proposed by Zhang and Patel (2018). In their work, the de-raining task is performed on pixel-level, feature-level, and symbol-level information from rainy images. This is the first time that generative adversarial network (Goodfellow et al. 2014), densely connected network (Johnson et al. 2016), and classification learning play important roles in rain removal models. As a result, this work presents the strong feature representation ability of deep learnings, and performs significant improvement over other state-of-the-art de-raining models.

Recently, Fu et al. (2019) proposed a lightweight pyramid network for rain removal on single image. They introduced the mature Gaussian–Laplacian image pyramid decomposition technology into deep learning, which greatly simplifies the complexity of deep neural model. Meanwhile, Li et al. (2018b) combined the deep convolutional and recurrent neural network for single image rain removal. They used the dilated convolution to enlarge the receptive field of deep network, and removed the overlapping rain streaks by using recurrent network. Thus, this work models de-raining task as a temporal problem with multiple stages. In addition, several GAN-based de-raining methods have been proposed. For example, Xiang et al. (2019) trained a GAN model for which the supervision from ground truth is imposed on different layers of the generator network, and achieved good results. Matsui and Ikehara (2020) proposed a GAN-based de-raining network trained with mixture of two rain image composite models, these enables their proposed network robust enough to handle a variety of actual rain.

3 The proposed de-raining system

The proposed de-raining neural model is a fully convolutional network composed of homogenous kernels, each with the size of 3 by 3. The first 16 shareable convolutional layers are designed for representative learning; the following two sibling branches are generative networks, in which the characteristics learning for rain streaks and background scenes are simultaneously conducted. Other visual models such as
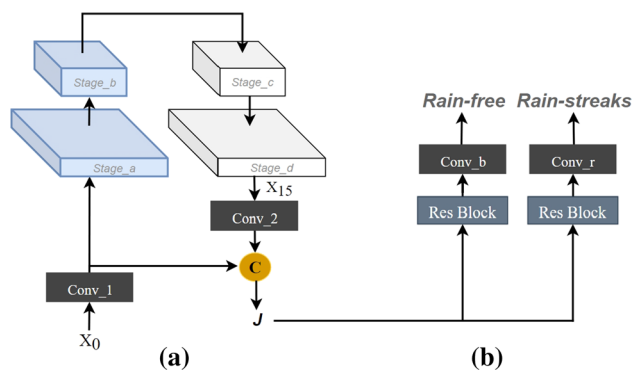

Fig. 2 An illustration of de-raining network. **a** Is the representative sub-network, **b** shows two sibling networks

perceptual loss are incorporated with the main de-raining network.

3.1 Representative sub-network

The construction of the representative network involves two feature hierarchy modules, as shown in Fig. 2a, in which the proportionally scaled feature maps are successively outputted, and these layers with the same size of maps are defined as one hierarchy level (i.e., a stage) of a module, each stage is composed of several residual block (He et al. 2016), each is

$$f(X_i) = W^{i+1} * g(X_i) + X_i,$$
$$g(X_i) = \xi\left(W^i * X_i\right), \tag{1}$$

where X_i denotes input of the ith layer, W is the convolutional kernel with the size of 3×3, $*$ is a convolutional operation, and ξ is the batch normalization followed by the activation function ReLU (Nair and Hinton 2010). Here, the first module is a downsampling in-network with a fixed scaling step of 2. The top stage has stride of 4 pixels with respect to the input image, and outputs the highest-level semantic features in the bottom-up pathway. On the top-down way, the second module uses fractionally strided convolutions to recover the resolution. Despite being semantically stronger, the features become spatially coarser from top to down. The coarsest outputs of the Stage$_d$ directly lead to block artifact in the desired results. To remove the aliasing effect caused by upsampling, we add a single convolution layer at the end of Stage$_d$. In addition, to remain more accurate locations, we enhance the features output by Conv_2 via a lateral connection from the Conv_1 as follows:

$$J = [g(X_0), g(X_{15})], \tag{2}$$

where $[\cdot]$ represents concatenation operation, X_0 denotes the input rainy image, and X_{15} are the features output by Stage$_d$.
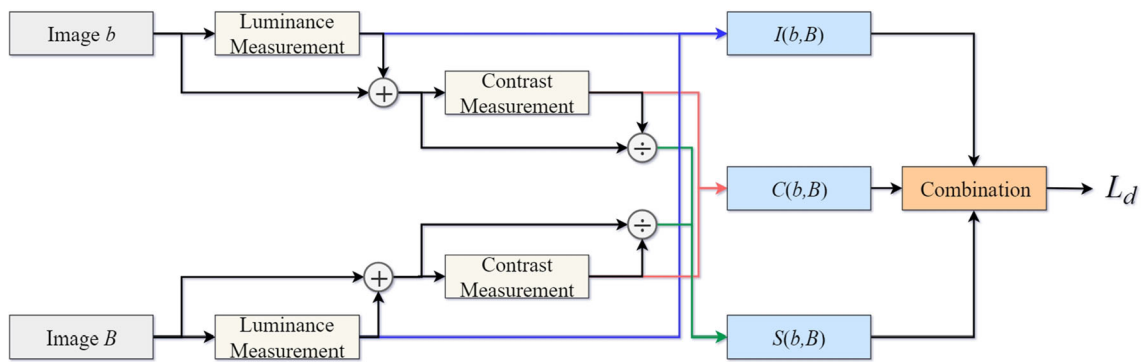
Fig. 3 Diagram of the saliency feature comparisons between b and B

In doing so, the features with the same shape but different semantics are merged as the final features of the representative model, i.e., J.

3.2 Generative sub-networks

Figure 2b illustrates the generative structure which is constructed by two building blocks. The first one learns the high-frequency characteristics of rainy layers R, and outputs the sparse approximation r. Meanwhile, the other one restores the rain-free results b by acquiring the representation of background scenes B. We formulate these two sibling models in what follows

$$L_b = \ell(b, B), \quad L_r = \ell(r, R),$$
$$g(f(J)) = \begin{cases} r, & \text{rain branch} \\ b, & \text{norain branch} \end{cases} \quad (3)$$

where ℓ is the empirical critical (EC); to ensure that all pixels have a reasonable influence on the final output, we use a scaled bilateral rectified linear unit (BReLU) (Cai et al. 2016) at the sibling layers to compute θ : $x \to \min(\max(x, 0), 255)$. Obviously, L_b is a difficult task because B tends to have complex information, while L_r is easy to reach its minimization due to the sparsity of rain-streaks. For this reason, we design the branch construction to utilize the easy task L_r to facilitate the optimization of complicated issue L_b via the following constraint

$$L_o = \ell(o, R + B - RB),$$
$$o = 1 - (1 - r)(1 - b) = r + b - rb. \quad (4)$$

Here, o is the reconstruction of observed rainy image O, which is expressed via a nonlinear system called screen blend model (SBM). Different from linear additive composite model (i.e., $O = B + R$), SBM not only reflects the overlapping effect of raindrops, but also the transparency effect

in most natural rain environment. As a result, the SBM is so robust that b behaves toward B when r is perfect. All the learning criterions mentioned are pixel-wise tasks, the next focus is on developing feature-aware vision tasks for further improvement.

Some assessments, e.g., structural similarity (SSIM Wang et al. 2004), emphasize the sensitivity of visual system to diverse vision signals which correlate well with the subjective fidelity ratings. Motivated by this, we introduce a comprehensive task to learn the statistical distributions from images. As illustrated in Fig. 3, this task comprises three image attribute measures: luminance (I), Contrast (C), and Structure (S), which can be represented as follows:

$$I(b, B) = \frac{2\mu_b\mu_B + \epsilon_1}{\mu_b^2 + \mu_B^2 + \epsilon_1},$$
$$C(b, B) = \frac{2\sigma_b\sigma_B + \epsilon_2}{\sigma_b^2 + \sigma_B^2 + \epsilon_2}, \quad (5)$$
$$S(b, B) = \frac{\sigma_{bB} + \epsilon_3}{\sigma_b\sigma_B + \epsilon_3},$$

where μ_b and μ_B are the mean signal intensities, σ_b and σ_B are the standard deviation of signal samples, σ_{bB} refers to the covariance. The nonnegative constants $\epsilon_1, \epsilon_2, \epsilon_3$ are included to avoid division by zero, and we set $\epsilon_2 = 2\epsilon_3$. Then, the comprehensive task can be defined as:

$$L_d = 1 - I(b, B) \cdot C(b, B) \cdot S(b, B)$$
$$= 1 - \frac{(2\mu_b\mu_B + \varepsilon_1)(2\sigma_{bB} + \varepsilon_2)}{(\mu_b^2 + \mu_B^2 + \varepsilon_1)(\sigma_b^2 + \sigma_B^2 + \varepsilon_2)}. \quad (6)$$

(a) *raw*   **(b)** *filter_v*   **(c)** *filter_h*   **(d)** *magnitude*
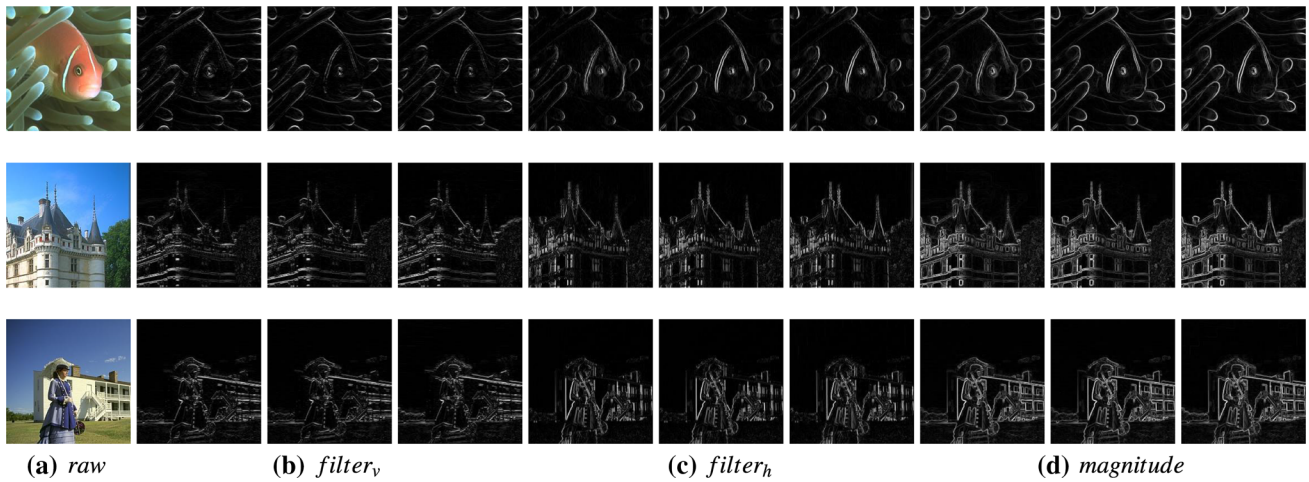
Fig. 4 Saliency maps of perceptual information. **a** Denotes the input image. From **b** to **d**, the first 3 maps in each row are the edge features with respect to vertical direction. The 3 middle maps are edge features of horizontal direction. The last 3 maps are the combinations of vertical and horizontal features. For visualization, the results are squared before square root (sqrt)

Algorithm 1 Dynamic-Weighting Scheme for loss weights.

Input: Loss $L(\theta)$, Population size Pop, Maximum iteration G, Mutation factor P_m, Crossover probability P_{cr},

Output: The loss weight λ^i for each auxiliary task with optimal solution under constraints.

Step 1. Initialize: The Initial Population is $\mathbf{P}_0 = \{\mathbf{P}_0^1,...,\mathbf{P}_0^{Pop}\}$, where $\mathbf{P}_0^i = \{\lambda_0^{i,r}, \lambda_0^{i,o}, \lambda_0^{i,p}, \lambda_0^{i,d}\}$. Each solution $\lambda_0^{i,n}$ in $\mathbf{P}_0^i$ corresponds to a loss L_n obtained from de-raining net. The initial value of ith individual is generated as $\mathbf{P}_0^i = \mathbf{P}_{min} + rand(0,1) \cdot (\mathbf{P}_{max} - \mathbf{P}_{min})$.

Step 2. Evolution: Find optimal solutions

1. **for** $g = 1$ **to** G **do**
2.   **for** $i = 1$ **to** Pop **do**
3. $Mutation(\mathbf{P}_g^i, P_m)$:
4. $(h,j,k) = random\_choice(\{1,...,Pop\} - \{i\})$;   // Randomly select three indices.
5. $\mathbf{P}_{g+1}^{i'} = \mathbf{P}_g^h + P_m \cdot \left(\mathbf{P}_g^j - \mathbf{P}_g^k\right)$;   //Generate a new candidate solution.
6. $Crossover(\mathbf{P}_g^i, \mathbf{P}_{g+1}^{i'}, P_{cr})$:
7.   **for** $n \in \{r,o,p,d\}$ **do**
8.   **if** $rand(0,1) < P_{cr}$ **then**
9. $\lambda_{g+1}^{i',n} = \lambda_{g+1}^{i',n}$
10.   **else**
11. $\lambda_{g+1}^{i',n} = \lambda_g^{i,n}$
12.   **end if**
13.   **end for**
14. $Selection(\mathbf{P}_g^i, \mathbf{P}_{g+1}^{i'}, L(\theta), F)$:
15.   **if** $F(\mathbf{P}_g^i, L(\theta)) \geq F(\mathbf{P}_{g+1}^{i'}, L(\theta))$ **then**   // $F(\mathbf{P}_g^i, L(\theta)) \stackrel{\Delta}{=} \left(L_b - \sum_{n \in \{r,o,d,p\}} \lambda_g^{i,n} L_n\right)^2$
                                     $s.t.$ $\lambda_g^{i,r} L_r = \lambda_g^{i,o} L_o = \lambda_g^{i,d} L_d = \lambda_g^{i,p} L_p$
16. $\mathbf{P}_{g+1}^i = \mathbf{P}_{g+1}^{i'}$
17.   **else**
18. $\mathbf{P}_{g+1}^i = \mathbf{P}_g^i$
19.   **end if**
20.   **end for**
21. **end for**

Step 3. Get optimal solution: $\mathbf{P}_{best} = \arg\min_i(F(\mathbf{P}_G^i, L(\theta)))$

3.3 Perceptual model

Johnson et al. (2016) proposed the early perceptual network which is the VGG net (Simonyan and Zisserman 2015)

pre-trained on ImageNet dataset (Deng et al. 2009). These features obtained from hidden layers are considered as the perceptual information. It is a popular approach to main-

tain the structure of image. However, Li et al. also argued that the perceptual network only emphasize specific features which is helpful for object recognition since it is a discriminative model trained on a classification dataset with finite categories. Besides, we find another limitation of this model with respect to the information loss caused by pooling layer. For de-raining work, it is not necessary to capture the deep features with rich semantics, we thus propose a flat model as the alternative way of deeply trained structure. Concretely, our model unifies two types of discrete differentiation kernels which can be defined as

$$\text{filter}_h = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}, \text{filter}_v = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix}, \quad (7)$$

where filter_h estimates gradient along the horizontal direction, and the other produces measurement of the gradient along the vertical direction (refer to Fig. 4 for some examples). According to the two basic SF operators, the width of perceptual model can be extended through enriching the set of kernels S. Then, the perceptual loss can be represented as

$$L_p = \frac{1}{3|S|} \sum_c \sum_\omega^{|S|} \ell \left(\text{filter}_\omega * b^c, \text{filter}_\omega * B^c \right), \quad (8)$$

where $c \in \{\text{red, green, blue}\}$ denotes the channel of color image. Technically, these kernels are convolved with b and B to calculate approximations of the derivatives with respect to different orientations, then the average gradient distance is used to define the loss of details in the restored image.

3.4 Adaptive optimization

From what has been discussed above, the optimization model in our rain removal work is

$$L(\theta) = \sum_{i \in \Omega} \lambda^i L_i, \quad (9)$$

where $\Omega = \{b, r, o, d, p\}$, θ is the set of network parameters, λ^i is the loss weight of the ith task. For each i, ℓ can be defined by PHL whose generic form is

$$\ell(\hat{y}, y) = \frac{\delta^2 \left(\sqrt{1 + ((y - \hat{y})/\delta)^2} - 1 \right)}{N}, \quad (10)$$

where N is the amount of information. δ is a predefined threshold of residual, and when difference between restored signal $\hat{y}$ and reference signal y approaches toward a large value, the steepness of PHL approximates δ. Obviously, PHL combines the best properties of MSE and L1 absolute loss,

thus it is not only strongly convex at the vicinity of optimal points, but also less sensitive to extreme values.

One can note that the proposal de-raining model falls under the umbrella of multi-task learning (MTL), in which L_b is the main optimization objective, others are side tasks, and L is a differentiable joint function with respect to θ. The assignment of balancing coefficients is a non-trivial problem because these tasks in MTL have different learning difficulties and convergence rates in different iteration of training process (Yin and Liu Feb. 2018). For appropriate loss weights, we develop a dynamic-weighting method based on differential evolution (DE) (Qu et al. 2012) (refer to Algorithm 1 for details), in which the weight of main task L_b is 1, namely $\lambda^b = 1$. Meanwhile, the dynamic-weighting algorithm learns to allocate loss weights to overall auxiliary tasks, i.e., λ^n, $n \in \{o, r, d, p\}$. The ith candidate solution in the gth generation is $(\lambda_g^{i,o}, \lambda_g^{i,r}, \lambda_g^{i,d}, \lambda_g^{i,p})$, denoted as $\mathbf{P}_g^i$. And the solutions of initial population are randomly sampled from a finite instance space. Here, F denotes the fitness function in each evaluation phase. To overcome catastrophic forgetting (Kirkpatrick et al. 2017) during training process, we develop F as a constrained optimization model which avoids to halt any tasks, then three operations (i.e., mutation, crossover, and selection) successively evolve populations with random probabilities.

The last part of our de-raining system is the image enhancement approach. Based on the atmospheric scattering model (McCartney 1976), the imaging model to represent the formation of a rainy image can be extended as

$$O^c = t \left(B^c + R - B^c R \right) + (1 - t) A^c, \quad (11)$$

where t is the medium transmission, A is the atmospheric light. Here, we attempt to use an existing method (Gao et al. 2019) for estimations of t and A. However, the setting of hyperparameters in this method is also non-trivial. Thus, we use DE to develop an adaptive framework for image enhancement. The optimization objectives are regularization parameter λ and exponent β, the former is used to balance the data term and the gradient constraint, the latter determines the levels of sensitivity to gradients of the minimum channel, and the fitness function is Contrast Enhancement Image Quality (CEIQ) (Jia et al. 2018).

4 Experiments

In this section, comprehensive experiments are performed on synthetic and real rainy images. The optimization method used in our model is stochastic gradient descent (SGD) with momentum=0.9, and we set the initial learning rate as 1e-2, dividing it by 10 at the 20th epoch, and terminate training at the 40th epoch. The implementation of the proposed de-

Table 1 Influence of δ on model performance

Task	EC	δ	SSIM	VIF
Pixel	PHL	1	0.5459	0.1692
		2	0.8957	0.5128
		3	**0.9365**	**0.6380**
		4	0.9344	0.6234
		5	0.9341	0.6039
		10	0.9343	0.6229
		50	0.9321	0.6204
		100	0.9262	0.6085

Testing on Rain100L images
Bold values mean the maximum values

Table 2 Average SSIM/VIF values on Rain100L images

Task	EC	SSIM	VIF
Pixel	MSE	0.9211	0.6239
Pixel	PHL	0.9365	0.6380
Pixel $+ L_p$	PHL	0.9397	0.6429
Pixel+VGG	PHL	0.9391	0.6337
Pixel $+ L_p + L_d$	PHL	**0.9412**	**0.6435**
Fixed	PHL	0.9277	0.5907

Bold values mean the maximum values

Table 3 Average SSIM/VIF values on Rain100H images

Task	EC	SSIM	VIF
Pixel	MSE	0.8138	0.3881
Pixel	PHL	0.8279	0.3945
Pixel $+ L_p$	PHL	0.8402	0.3986
Pixel+VGG	PHL	0.8394	0.3979
Pixel $+ L_p + L_d$	PHL	**0.8521**	**0.4040**
Fixed	PHL	0.8478	0.3970

Bold values mean the maximum values

raining model is conducted on Python3.5, TensorFlow1.8, GeForce GTX TITAN with 12GB RAM.

In the dynamic-weighting scheme, Pop is the size of population, which is a multiple of the size of solutions. A smaller Pop corresponds to the low diversity that often causes the evolutionary algorithm to stagnate at local optimum solution. However, a bigger Pop tends to result in a higher computational complexity. For a compromise between solution diversity and computational cost, we set Pop as 5 times as the size of solution. The maximum iteration G affects the evolutionary algorithm in the similar manner as that of Pop. Thus, we set G as 100 for a suitable trade-off. The mutation factor p_m and crossover probability p_{cr} affect the retrieval efficiency. To be more stable, p_m and p_{cr} usually are limited in the interval (0, 1). Generally, bigger p_m and p_{cr} may enlarge the retrieval range, for a better convergence rate, we

thus set smaller values to p_m and p_{cr}, i.e., $p_m = 0.5$ and $p_{cr} = 0.3$.

4.1 Data preparation

Since the pairs of rain and rain-free images from natural scenes are not massively available, the training instances are generated by synthesizing rainy images based on screen blend model. More specifically, 1800 rain-free images as ground truth data are collected from BSD300 (Martin et al. 2001). There are 1800 rain layers. To augment the size of training dataset, 10 pairs of image patches are cropped from each pair of synthetic rain image and corresponding ground truth image. As a result, there are 18,000 pairs of instances in our training dataset.

4.2 Results on synthetic rainy images

In this experimental part, we will investigate where the improvement of performance comes from. For objectivity, we conduct ablation study on two public benchmark datasets (Yang et al. 2017). To quantitatively assess these different settings, two evaluation criteria (i.e., SSIM Wang et al. 2004 and VIF Sheikh and Bovik 2004) are employed to measure the difference between de-rained result and corresponding ground truth image. A SSIM close to 1.0 indicates that the performance is perfect, and the higher VIF is, the better image reconstruction does. Here, the basic model is based on pixel-level tasks (denoted as Pixel). The results are shown in Tables 2 and 3.

The first comparison focuses on training criterion, as shown in the first two rows of Tables 2 and 3, and the basic model with PHL has better performance in terms of SSIM and VIF. These quantitative results demonstrate that PHL provides a robust regression finding the relation between observed and predicted images. According to Eq. (10), one can note that δ has a significant impact on the steepness of PHL. A smaller δ tends to cause vanishing gradient, on the contrary, a larger δ often results in the oscillation in optimization. For a suitable threshold of residual, we conduct the grid search on a set of discrete values. As reported in Table 1, when $\delta = 3$, the method based on PHL can get better results in terms of SSIM and VIF. Therefore, in the following experiment, δ is fixed as 3. Then, we analyze the suitability of feature-level tasks. From the second, third, and fifth lines of Tables 2 and 3, the comparison results indicate that the feature-level tasks we design are capable of assisting with the improvement of rain removal results. Also, one can note that results on the first dataset (i.e., Rain 100L) tend to be higher, this reason is that all rainy layers in this dataset are composed of slight rain streaks which are not difficult to be handled. When presented with heavy rain streaks in the second dataset (i.e., Rain100H), the feature-level tasks

Fig. 5 Prefix symbol '**a**' denotes the original rainy images, '**b**'–'**f**' are the results of Ours, DDN, LPNet, RESCAN, JORDER-E, respectively
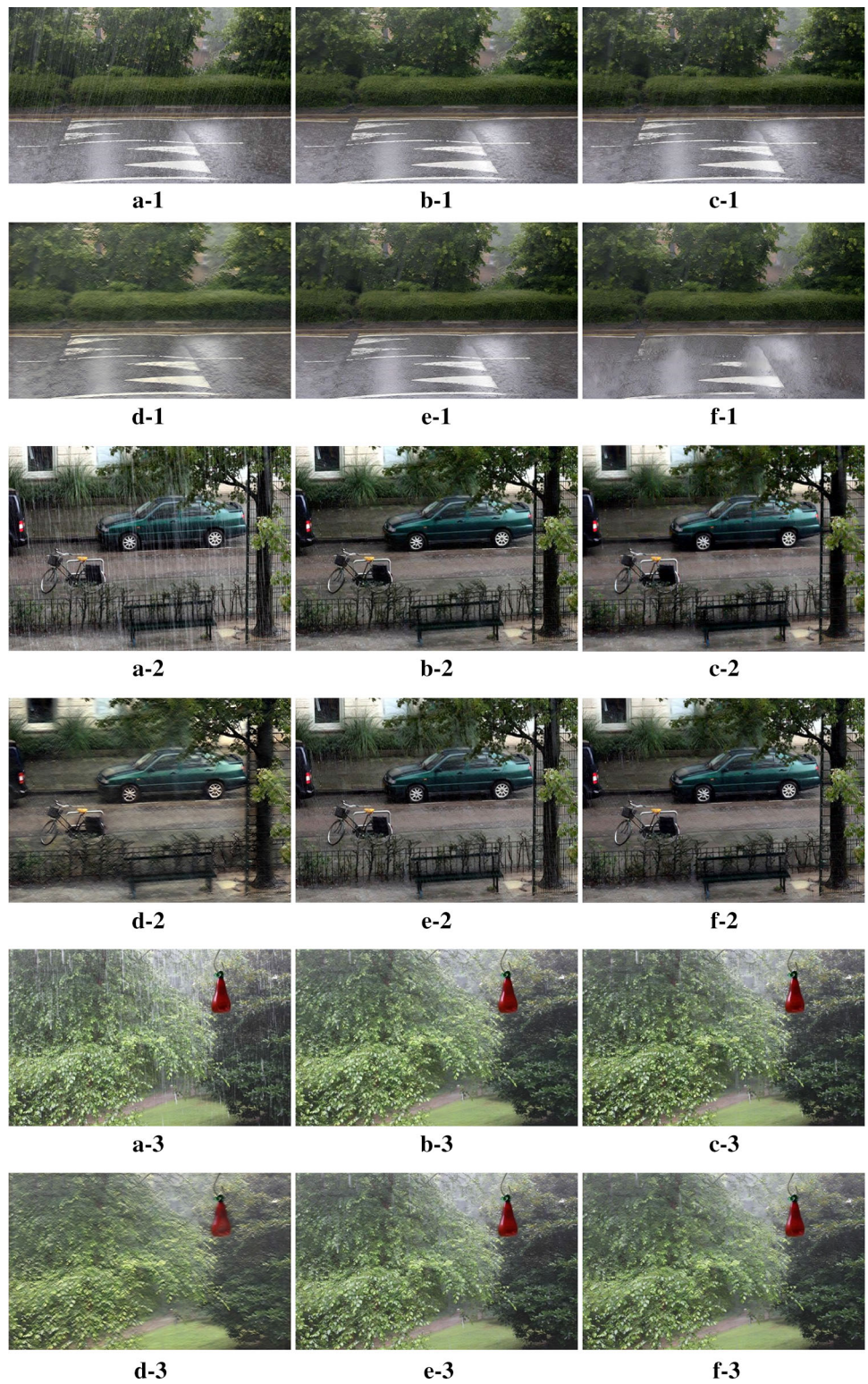
Table 4 Quantitative comparison among our method and state-of-the-art methods

Models	Rain100H		Rain100L	
	SSIM	VIF	SSIM	VIF
RESCAN Li et al. (2018b)	0.8055	**0.4057**	0.9401	0.6591
DDN Fu et al. (2017)	0.7492	0.3427	0.9097	0.5727
JORDER-E Yang et al. (2019)	0.8005	0.4033	0.9396	**0.6610**
LPNet Fu et al. (2019)	0.7719	0.3422	0.9248	0.6121
Ours	**0.8521**	0.4040	**0.9412**	0.6435

Bold values mean the maximum values

Fig. 6 Prefix symbol 'a' denotes the original rainy images, 'b'–'f' are the results of Ours, DDN, LPNet, RESCAN, JORDER-E, respectively
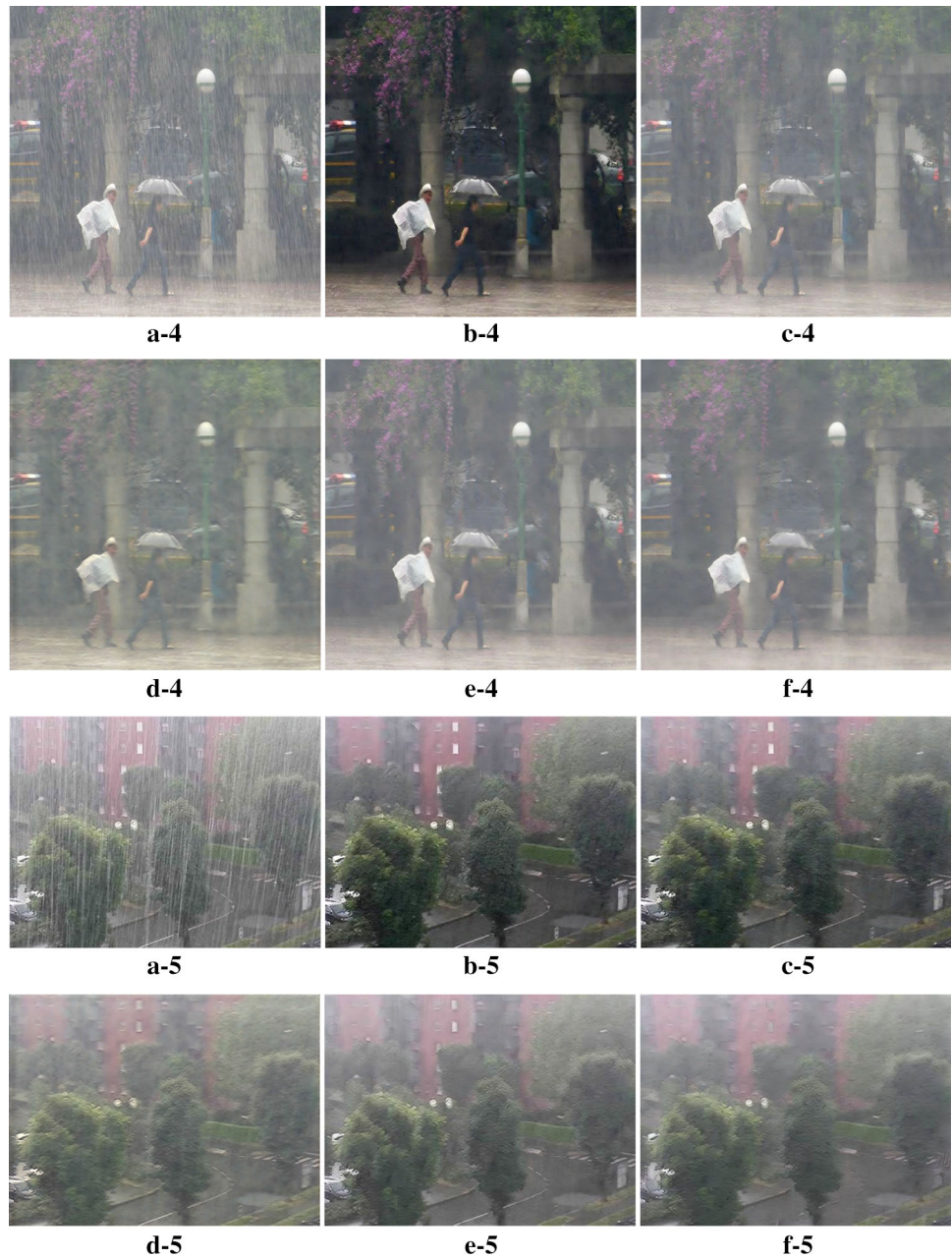
Fig. 7 Prefix symbol 'a' denotes the original rainy images, 'b'–'f' are the results of Ours, DDN, LPNet, RESCAN, JORDER-E, respectively
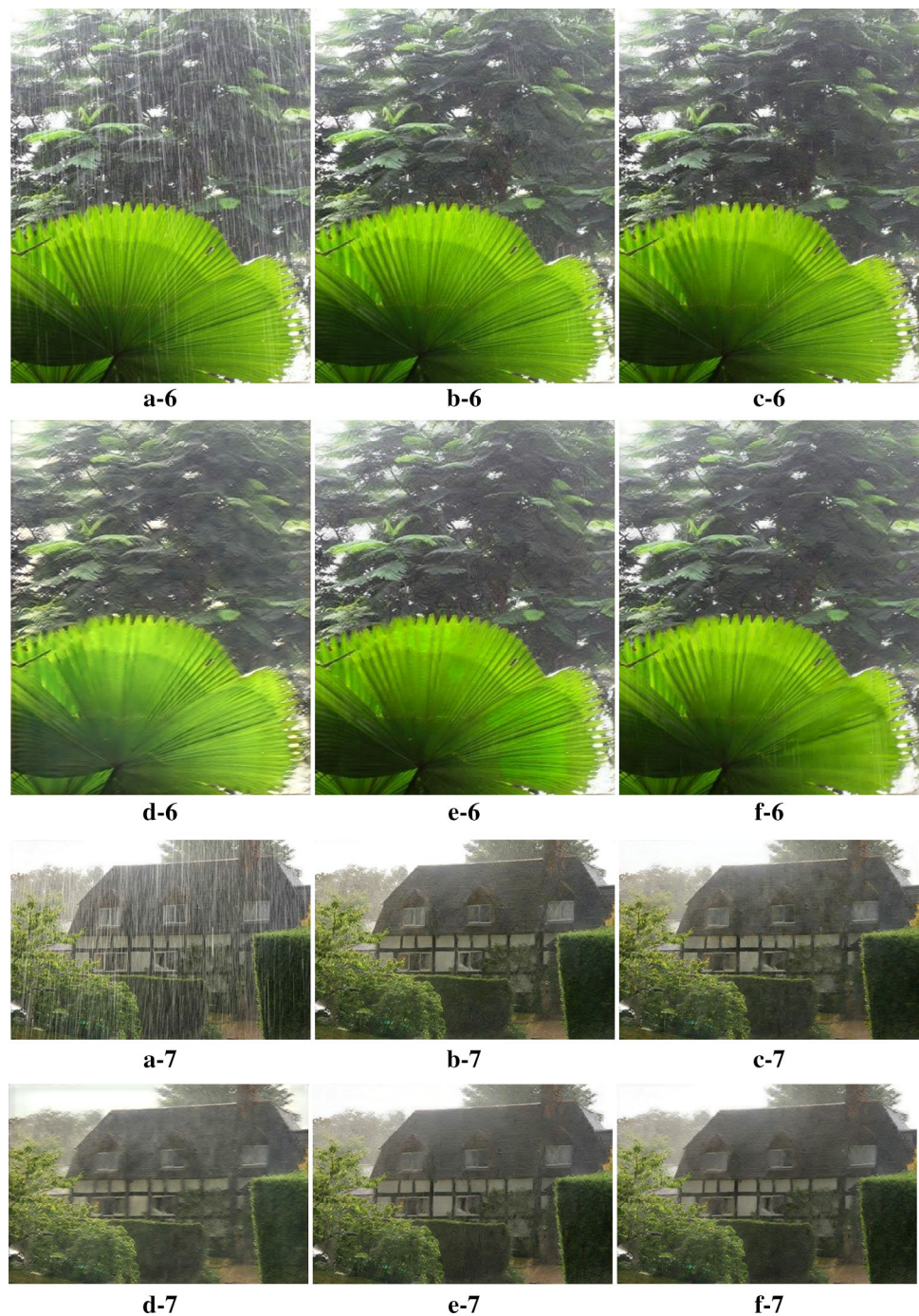
play crucial role for improvement. Other interesting comparisons also are presented in the two tables. On one hand, the proposed perceptual model gets competitive results compared with VGG net, this demonstrates that simple gradient operators are good enough for image reconstruction, and can replace the complex pre-trained neural model. On the other hand, in the last two rows in these tables, 'Fixed' denotes the model with fixed weighting factors, and these results

demonstrate that the dynamic optimization method gets better generalization ability on evaluation datasets.

In addition, the proposed method (namely Pixel $+ L_p + L_d$) is also compared with other state-of-the-art methods, namely DDN (Fu et al. 2017), RESCAN (Li et al. 2018b), LPNet (Fu et al. 2019), and JORDER-E (Yang et al. 2019). From the results reported in Table 4, one can note that the proposed method has the best SSIM and competitive VIF. These advantages can be mainly attributed to three reasons.

Fig. 8 Prefix symbol '**a**' denotes the original rainy images, '**b**'–'**f**' are the results of Ours, DDN, LPNet, RESCAN, JORDER-E, respectively
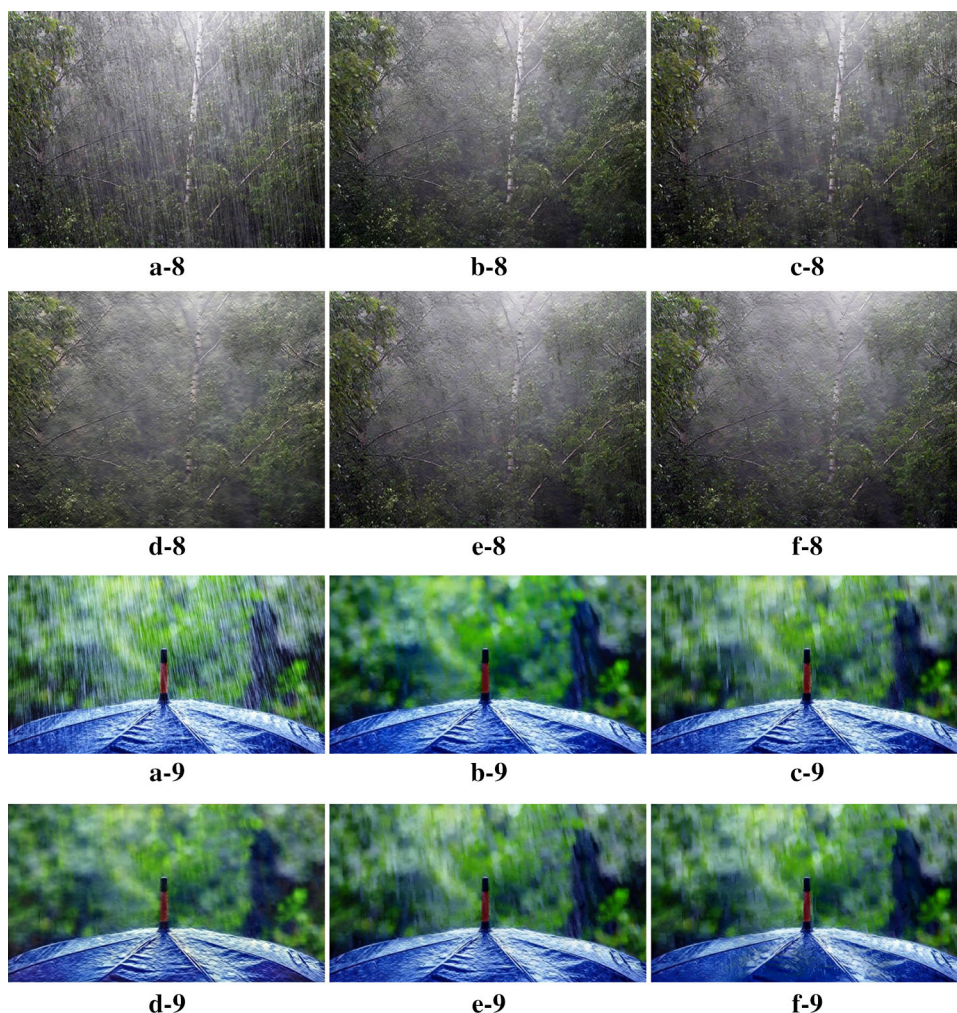

First, our method adopts the screen blend model which is more robust than the linear additive composite model used by other methods. Second, in the multitask learning, the side task L_d can help the deep network learn much more salient image features, such as the luminance, contrast, and structure, while the other auxiliary task L_p can effectively preserve the background details. The third reason lies in the dynamic fusion, and all rain removal tasks can be combined to achieve favorable performance.

4.3 Results on real-world rainy images

In this part, the proposed de-raining model without post-processing step is compared with several state-of-the-art rain removal models as mentioned above. The real-world rainy images are kindly provided by existing publications. These images cover different rainy situations in terms of the size, velocity, and angle of rain streaks, and contain rich details. Thus, the generalization capabilities of de-raining models can be effectively verified by tackling these complex cases.

The visual comparisons of different de-raining methods are shown in Figs. 5, 6, 7 and 8. Through visual inspection, one can note that all methods can effectively remove rain streaks in most cases. However, when confronting dense rain streaks, some methods tend to leave heterogeneous veiling effect (see Figs. 6 and 8), which degrades the visual quality. Instead, our method can avoid this unfavorable effect (see the b-4 in Fig. 6), because in the multitask learning, the proposed network can learn not only how to remove the rain streaks, also to enhance the visual effect of image. Besides, some methods may lead to over-smoothed results (such as f-1 and f-3 in Fig. 5), in which many details are lost, while our method can significantly maintain details and much more texture of background scene. This advantage mainly benefits from the $S(b, B)$ and the perceptual model based on SF operators. These comparisons demonstrate that our method is more robust to deal with various rainy conditions while preserving image details.

Fig. 9 Rainy images and the results. The first row of each group shows the full image, and the second row shows two zoomed-in regions. It is clear that the proposed method performs well in edge preserving

(a) Raw rainy image   **(b)** JORDER-R   **(c)** Ours
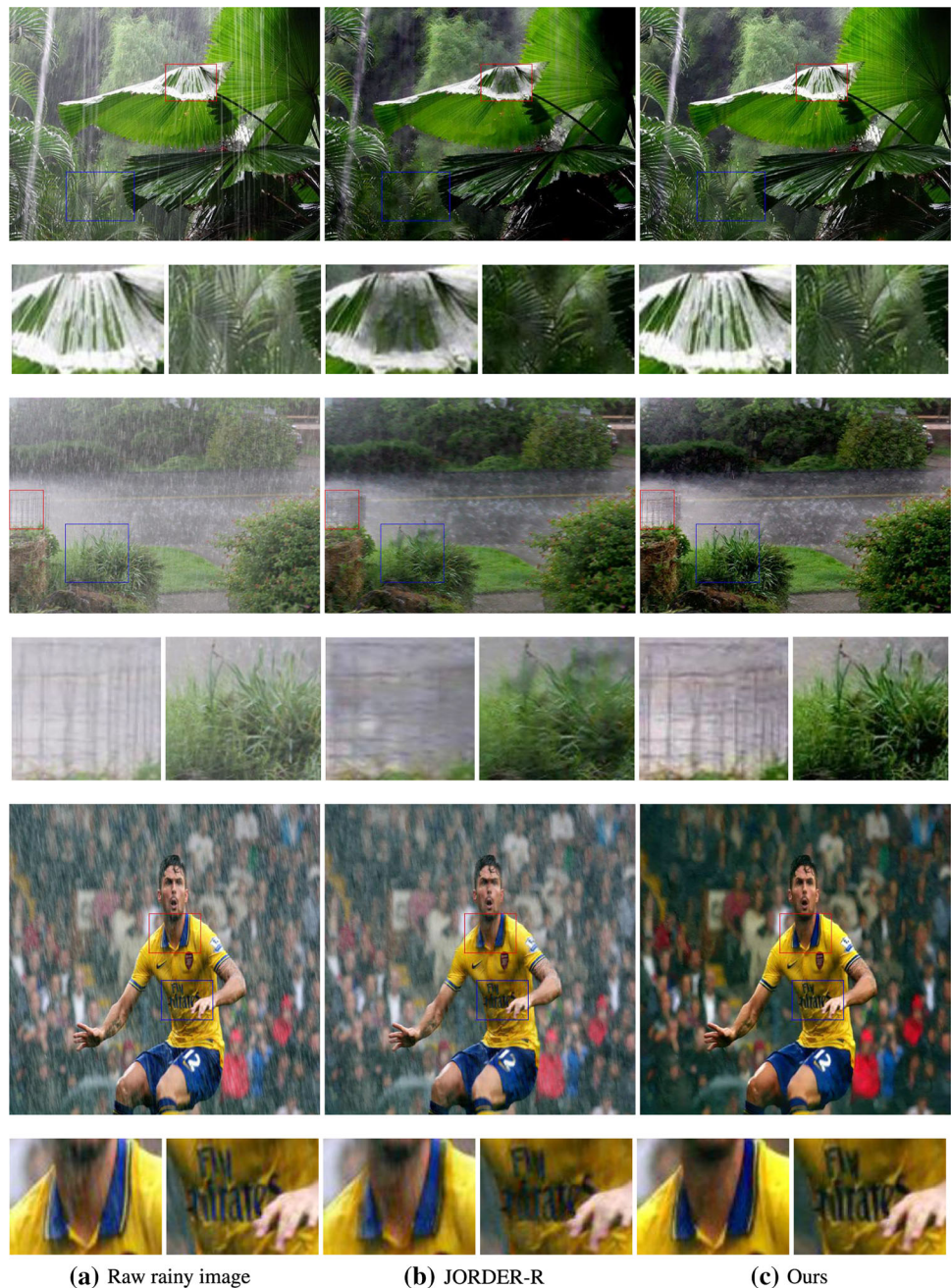
4.4 Impact of adaptive post-processing step

In this section, the effect of entire de-raining system including post-processing step will be presented. The state-of-the-art de-raining model denoted as JORDER-R (Yang et al. 2017) is compared with our system, because JORDER-R also embeds a de-hazing method He et al. (2011) into its de-raining framework. Figure 9 shows comparisons between JORDER-R and our system. Though observing the de-rained results, one can note that the surviving haze can be effectively removed from de-rained images. Compared with JORDER-R, the proposed

method tends to keep some details of rainy images, such as the lines of leaves, the outlines of bars and the stripes of clothes.

5 Conclusion

A neural de-raining system based on screen blend model is presented in this paper. We take advantage of the synergy among main task and side tasks to improve the effect of rain removal model. Especially, we design a vision task

to preserve complex texture of background layer by using Sobel–Feldman Operator. And we also introduce another feature-level task with respect to salient image features that convey the crucial information for human visual system. To tackle with unknown balancing parameters among these tasks we use, a dynamic method is embedded into the learning process whose objective is to minimize the PHL. Due to the design of multiple tasks and adaptive optimization, we show that a simple neural network with sibling branches can achieve state-of-the-art de-rained results on most rainy scenes. Finally, an evolutionary-based post-processing framework for haze removal is utilized to further improve the visual quality of de-rained image. Comprehensive experiments have been conducted on synthetic and real rainy images, and the effectiveness of the proposed de-raining system is verified in terms of reference criteria and visual inspection. In the future work, we will research other appropriate tasks that can significantly boost the visual effect of rainy images.

Compliance with ethical standards

Conflict of interest Yulong Fan declares that he has no conflict of interest. Rong Chen declares that he has no conflict of interest. Yang Li declares that he has no conflict of interest. Tianlun Zhang declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Cai B, Xu X, Jia K, Qing C, Tao D (2016) Dehazenet: an end-to-end system for single image haze removal. IEEE Trans Image Process 25(11):5187–5198

Charbonnier P, Blancferaud L, Aubert G, Barlaud M (1997) Deterministic edge-preserving regularization in computed imaging. IEEE Trans Image Process 6(2):298–311

Chen J-C, Liu C-F (2017) Deep net architectures for visual-based clothing image recognition on large database. Soft Comput 21(11):2923–2939

Chen Y, Hsu C (2013) A generalized low-rank appearance model for spatio-temporally correlated rain streaks, pp 1968–1975

Deng J, Dong W, Socher R, Li L, Li K, Feifei L (2009) Imagenet: a large-scale hierarchical image database, pp 248–255

Fu X, Huang J, Zeng D, Huang Y, Ding X, Paisley J (2017) Removing rain from single images via a deep detail network, pp 3855–3863

Fu X, Liang B, Huang Y, Ding X, Paisley J (2019) Lightweight pyramid networks for image deraining. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2019.2926481

Gao Y, Hu H-M, Li B, Guo Q, Pu S (2019) Detail preserved single image dehazing algorithm based on airlight refinement. IEEE Trans Multimed 21(2):351–362

Girshick R (2015) Fast r-cnn

Goodfellow IJ, Pougetabadie J, Mirza M, Xu B, Wardefarley D, Ozair S, Courville AC, Bengio Y (2014) Generative adversarial nets, pp 2672–2680

He K, Sun J, Tang X (2011) Single image haze removal using dark channel prior. IEEE Trans Pattern Anal Mach Intell 33(12):2341–2353

He K, Gkioxari G, Dollar P, Girshick RB (2017) Mask r-cnn. In: International conference on computer vision, pp 2980–2988

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. Comput Vis Pattern Recognit. https://doi.org/10.1109/CVPR.2016.90

Jia Y, Jie L, Xin F (2018) No-reference quality assessment of contrast-distorted images using contrast enhancement

Johnson J, Alahi A, Feifei L (2016) Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision, pp 694–711

Kang LW, Lin CW, Fu YH (2012) Automatic single image based rain streaks removal via image decomposition. IEEE Trans Image Process 21(4):1742–1755

Khmag A, Ramli AR, Kamarudin N (2019) Clustering-based natural image denoising using dictionary learning approach in wavelet domain. Soft Comput 23:8013–8027. https://doi.org/10.1007/s00500-018-3438-9

Kim J-H, Lee C, Sim J-Y, Kim C-S (2013) Single-image deraining using an adaptive nonlocal means filter, pp 914–917

Kim JH, Sim JY, Kim C (2015) Video deraining and desnowing using temporal correlation and low-rank matrix completion. IEEE Trans Image Process 24(9):2658–2670

Kirkpatrick J, Pascanu R, Rabinowitz NC, Veness J, Desjardins G, Rusu AA, Milan K, Quan J, Ramalho T, Grabskabarwinska A et al (2017) Overcoming catastrophic forgetting in neural networks. Proc Natl Acad Sci USA 114(13):3521–3526

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105

Li Y, Tan RT, Guo X, Lu J, Brown MS (2016) Rain streak removal using layer priors, pp 2736–2744

Li X, Wu J, Lin Z, Liu H, Zha H (2018b) Recurrent squeeze-and-excitation context aggregation net for single image deraining. In: European conference on computer vision. Springer, pp 262–277

Li M, Xie Q, Zhao Q, Wei W, Gu S, Tao J, Meng D (2018a) Video rain streak removal by multiscale convolutional sparse coding, pp 6644–6653

Luo Y, Xu Y, Ji H (2015) Removing rain from a single image via discriminative sparse coding, pp 3397–3405

Martin DR, Fowlkes CC, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. vol 2, pp 416–423

Matsui T, Ikehara M (2020) Gan-based rain noise removal from single-image considering rain composite models. IEEE Access 8:40892–40900

McCartney EJ (1976) Optics of the atmosphere: scattering by molecules and particles. Wiley, New York

Nair V, Hinton GE (2010) Rectified linear units improve restricted Boltzmann machines, pp 807–814

Narasimhan SG, Nayar SK (2003) Contrast restoration of weather degraded images. IEEE Trans Pattern Anal Mach Intell 25(6):713–724

Qu B, Suganthan PN, Liang J (2012) Differential evolution with neighborhood mutation for multimodal optimization. IEEE Trans Evol Comput 16(5):601–614

Rudin LI, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. Physica D 60(1–4):259–268

Santhaseelan V, Asari VK (2015) Utilizing local phase information to remove rain from video. Int J Comput Vis 112(1):71–89

Sheikh HR, Bovik AC (2004) Image information and visual quality. In: 2004 IEEE international conference on acoustics, speech, and signal processing, vol 3(iii), p 709

Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: International conference on learning representations

Wang W, Yuan X (2017) Recent advances in image dehazing. IEEE/CAA J Autom Sinica 4(3):410–436

Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612

Xiang P, Wang L, Wu F, Cheng J, Zhou M (2019) Single-image de-raining with feature-supervised generative adversarial network. IEEE Signal Process Lett 26(5):650–654

Yang W, Tan RT, Feng J, Liu J, Yan S, Guo Z (2019) Joint rain detection and removal from a single image with contextualized deep networks. IEEE Trans Pattern Anal Mach Intell 42:1377–1393

Yang X, Li H, Fan Y, Chen R (2019) Single image haze removal via region detection network. IEEE Trans Multimed 21(10):2545–2560

Yang W, Tan RT, Feng J, Liu J, Guo Z, Yan S (2017) Deep joint rain detection and removal from a single image. Comput Vis Pattern Recognit. https://doi.org/10.1109/CVPR.2017.183

Yin X, Liu X (2018) Multi-task convolutional neural network for pose-invariant face recognition. IEEE Trans Image Process 27(2):964–975

You S, Tan RT, Kawakami R, Mukaigawa Y, Ikeuchi K (2015) Adherent raindrop modeling, detection and removal in video. IEEE Trans Pattern Anal Mach Intell 38:1721–1733

Zhang X, Xiong Y (2009) Impulse noise removal using directional difference based noise detector and adaptive weighted mean filter. IEEE Signal Process Lett 16(4):295–298

Zhang H, Patel VM (2018) Density-aware single image de-raining using a multi-stream dense network. Comput Vis Pattern Recognit. https://doi.org/10.1109/CVPR.2018.00079

Zhao H, Xiao C, Yu J, Xu X (2015) Single image fog removal based on local extrema. IEEE/CAA J Autom Sinica 2(2):158–165