LFR-Net: Local feature residual network for single image dehazing☆

Xinjie Xiao, Zhiwei Li *, Wenle Ning, Nannan Zhang, Xudong Teng

College of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, 201620, Shanghai, China

ARTICLE INFO

ABSTRACT

Previous learning-based methods only employ clear images to train the dehazing network, but some useful information such as hazy images, media transmission maps and atmospheric light values in datasets were ignored. Here, we propose a local feature residual network (LFR-Net) for single image dehazing, which is aimed at improving the quality of dehazed images by fully utilizing the information in the training dataset. The backbone of LFR-Net is structured by feature residual block and adaptive feature fusion model. Furthermore, to preserve more details for the recovered clear images, we design an adaptive feature fusion model that adaptively fuses shallow and deep features at each scale of the encoder and decoder. Extended experiments show that the performance of our LFR-Net outperforms the state-of-the-art methods.

1. Introduction

In severe weather conditions such as haze, dust, and water vapor, ground-level visibility is reduced and the captured images are degraded due to the scattering of floating particles in the atmosphere. Images captured in hazy scenes are not suitable for high-level computer vision tasks [1–4]. Therefore, the work on single image dehazing is of positive significance for the development of computer vision.

The atmospheric scattering model (ASM) [5–7] is widely used for image dehazing by estimating the transmission and atmospheric light value in ASM. However, it is an ill-posed problem. Prior-based [8–11] methods are proposed by early researchers. For example, he et al. [8] proposed the dark channel prior theory, which estimates the medium transmission map based on the assumption that the intensity value of at least one channel of the clear image is close to zero, and a high quality dehazed image was restored. However, the assumption does not apply to images with sky regions, resulting in poor quality of the estimated medium transmission map.

Recently, many learning-based dehazing methods have been proposed, including the methods [12–15] for estimating the parameters of ASM, and the methods [16–19] for hazy-to-clear image transformation. These methods learn the transformation from a hazy image to a transmission map or clear image but do not work well with real-world images. For instance, Jiang et al. [20] proposed a residual spatial and channel attention network that adopts attention mechanism to estimate spatial and channel feature weights and restore dehazed image by using the method of hazy-to-clear image transformation. Meanwhile,

the prevailing training datasets [21] for image dehazing include hazy images, ground-truth, transmission maps and atmospheric light values. Previous learning-based image dehazing improved the performance of networks by increasing the layers of the dehazing network, which increases the resources for computation. However, the transmission maps and atmospheric light values contained in the dehazing training dataset are not used when dehazing networks are trained.

Most learning-based dehazing methods [14,22] usually have an encoder and decoder structure. The main features of the hazy image are extracted using the encoder, and a clear image is restored using a decoder. However, most network encoders have shallow layers in different scales, which can only extract the shallow features of the current scale, resulting in the loss of details in the clear image restored by the decoder. For example, the encoder of GFN [22] does not extract deep features at all scales, which results in loss of details in the dehazed results.

To address the above issues and obtain high quality dehazed images, we propose a notion of local feature residual network (LFR-Net) that contain a feature residual block and an adaptive feature fusion model, which is displayed in Section 3. In particular, we adopt feature residual (FR) block and adaptive feature fusion (AFF) model to extract and fuse the deeper semantic features. Feature residual block that enables the neural network to learn the deep semantic features of hazy images and learn significant features such as high-frequency information, and enhance the details of restored images. To prevent loss of shallow features, and texture features such as edges and contours of images

could not be preserved, we propose an adaptive feature fusion model to fuse the shallow feature and deep feature at each scale of encoder and decoder. Note that our adaptive feature fusion model gives adaptive fusion weights to the shallow and deep features of the encoder, which results in restored clear images with better texture features. Different from previous methods [13–17], we make the most of the data resources of the training dataset to optimize the dehazing ability of the network. In the training process of the proposed method, we designed haze loss to the loss function, which optimizes the network parameters by comparing the input hazy image and the haze-loss image that is generated by using the output of the neural network, the transmission map and atmospheric light value from the training dataset.

The following are the main contributions of this study:

- We propose a local feature residual dehazing network that aims to enhance the dehazed image quality by fully utilizing the information in the training dataset. The LFR-Net extracts deep features in the encoder and preserves the details of the restored clear images.
- The backbone of LFR-Net is constructed from feature residual block and adaptive feature fusion model. A feature residual block is designed to enable the dehazing network to learn deep semantic features of hazy images and learn useful features such as high-frequency information. To preserve and fuse shallow and deep features, we propose an adaptive feature fusion model that adaptively fuses shallow and deep features at each scale of the encoder and decoder.
- Haze loss is added to the loss function, and the hazy and clear images simultaneously supervise the training process of the network, which makes full use of data resources to optimize the dehazing ability of the network.

The following structure is used to compose this paper. The related work with image dehazing is described in Section 2. In Section 3, we systematically described the proposed approach and its various components. In Section 4, comparative experiments of LFR-Net with state-of-the-art (SOTA) methods and ablation experiments are carried out. Finally, in Section 5, we summarize this study and look forward to future research.

2. Related work

The prior-based dehazing methods and learning-based dehazing methods are two main types of existing single image dehazing methods.

2.1. Prior-based methods

The prior-based methods [8–11] use the theorem followed by a great deal of clear images to estimate the parameters of ASM. DCP [8] is proposed based on the presence of pixel points with intensity values close to zero in at least one channel in the clear image. Tan et al. [10] proposed two prior knowledge, the first one that clear images have higher contrast than hazy images and the second one that the atmospheric light value depends on the distance between the object and the camera, and further developed a cost function for single image dehazing based on the proposed prior knowledge with Markov random field. Berman et al. [11] proposed non-local prior based on the idea of haze-lines, specifically that the color of a haze-free image can be well approximated by hundreds of different colors and form color clusters in RGB space, which become haze-lines in RGB space due to the difference in transmission coefficients in hazy scenes, and then use the haze-lines to estimate the transmission map and recover a clear image. Although the prior-based methods dehazing work well, they would fail if the prior assumptions are not satisfied. For example, the DCP assumes that there exists at least one channel with intensity values close to zero for the image patches. However, this assumption is not satisfied for images with sky regions. DCP cannot preserve the detail of the sky for the hazy image with sky regions.

2.2. Learning-based methods

There are two classes of learning-based approaches: the first category [12–15] uses neural networks to approximate the parameters of ASM needed for image dehazing. The second category [16–20] does not rely on ASM and learns the transformation rules of hazy-to-clear images directly. Cai et al. [12] proposed DehazeNet, which links neural networks with prior knowledge to estimate the medium transmission map for restoring clear images. Li et al. [13] make a mathematical transformation of ASM to represent the medium transmission map (t) and atmospheric light (A) by a parameter and propose an AOD-Net to estimate this parameter for hazy image dehazing. Wu et al. [23] proposed ACER-Net that is a hazy-to-clear image transformation dehazing method, and used contrast learning to train the neural network and obtained excellent dehazed results.

Previous learning-based dehazing methods ignore t and A contained in the training dataset. Therefore, we propose a local feature residual network for single image dehazing, which makes full use of this information from the dataset to update the parameters of the neural network.

3. Proposed method

Encoder–decoder structure [24–26] and residual learning [27–29] improve the performance of convolution neural networks on image processing. Specifically, the encoder–decoder structure of the network has a larger perceptual field and the residual learning avoids the network performance degradation as the network depth increases. Inspired by residual learning and encoder–decoder structure, we propose a local feature residual network that combines the advantages of residual learning and encoder–decoder structure. Different from previous methods [24,27], we add max pooling to residual block for more concern about semantic features, and design an adaptive feature fusion model to replace the skip connection of the encoder–decoder structure. The architecture of our LFR-Net is shown in Fig. 1. We adopt the five scales of down-sampling and up-sampling of U-Net [24]. From the first layer to the fifth layer, the number of channels per layer is 64, 128, 256, 512 and 1024 respectively. After extracting deep features by four times downsampling, four times up-sampling and adaptive feature fusion model are used to fuse deep features with shallow features to obtain high quality dehazed images. Double-Conv and FR block are the basic blocks of LFR-Net. Double-Conv is used to extract shallow features at the current scale, but the shallow features only concern texture features, which is not enough to recover a clear image. Therefore, we add the FR block with residual learning properties so that the LFR-Net network can obtain deeper semantic features of the current scale feature map, and thus recover and retain more image details. The Double-Conv contains two convolutional layers, BN and RELU for extracting the important features at the current scale. To extract and fuse the shallow and deep features at each scale, we design two network modules: (1) Feature residual block can extract deep semantic features at each scale. (2) Adaptive feature fusion model can adaptively fuse shallow and deep features at each scale.

3.1. Feature residual block

Most image dehazing networks with encoder and decoder [14,22] extract only shallow features at each scale without capturing the deeper features of the image, resulting in the loss of detailed features in the restored clear images. Our feature residual block (see Fig. 2) can provide deep semantic features, allowing the decoder to generate clear images with more details. The FR block concatenates shallow features when extracting deep features, which can ensure that the training performance of the network will not degrade as the network level increases. The FR block is used to extract deep features for each downsampled layer. Maximum pooling is concerned with the texture
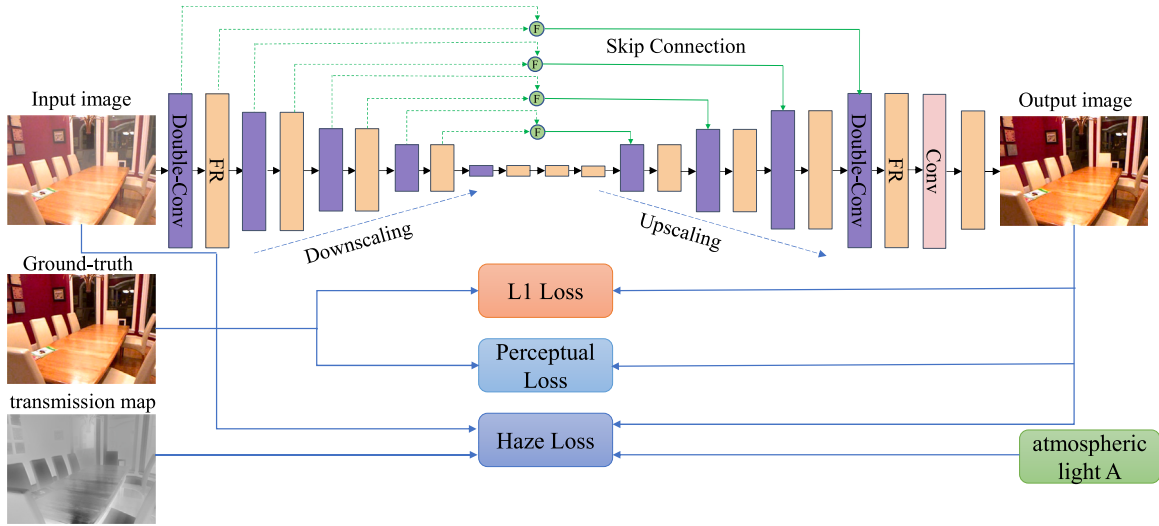
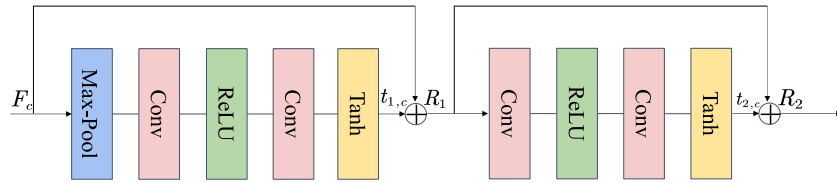**Fig. 1.** The Local Feature Residual Network (LFR-Net) architecture.

Fig. 2. Feature Residual Block.

information in the image. The neural network removes haze and re-dundant information through multiple times maximum pooling. First, we use global maximum pooling to reduce information redundancy and extract the main features.

$$p_c = M\left(F_c\right),\tag{1}$$

where $M\left(\cdot\right)$ denotes the maximum pooling operation. Then, to obtain deep semantic features, p_c pass through two convolutional layers, RELU and Tanh activation functions. To improve the efficiency of training we selected RELU and Tanh activation functions instead of others. And skip connection is used to ensure that the network has a strong dehazing ability.

$$t_{1,c} = T\left(conv\left(RELU\left(conv\left(p_c\right)\right)\right)\right),\tag{2}$$

$$R_1 = t_{1,c} + F_c,\tag{3}$$

where $T\left(\cdot\right)$ and $t_{1,c}$ denote the Tanh activation function and its output, respectively. To fully enhance the ability of the network to extract deep features, features are passed through a residual block that extracts deep features at each scale.

$$t_{2,c} = T\left(conv\left(RELU\left(conv\left(R_1\right)\right)\right)\right),\tag{4}$$

$$R_2 = t_{2,c} + R_1.\tag{5}$$

Our Feature Residual Block ensured that the network performance would not degrade with the increase of depth, and the deep features of each scale are extracted and preserved, so that the restored clear image has more details.

3.2. Adaptive feature fusion

Shallow networks tend to concern with the texture features of the image (e.g. edges and contours) [30]. However, with deeper network

layers, the network has a larger receptive field and is more concerned with the semantic features of the images. To solve this problem, pre-vious work [24] proposed fusing shallow and deep features using skip connection(see Fig. 3). Due to our deeper network, the shallow features are lost in the upsampling and downsampling layers if they do not employ the skip connection, resulting in poor performance of our restored clear images on features such as edges. In our LFR-Net, the Double-Conv and FR block contain shallow features and deep features at the current scale, respectively. If we adopt a regular skip connection, the dehazed image would be lost the shallow features of the encoder. Therefore, we design an adaptive feature fusion module to enhance the ability of shallow feature preservation. The feature maps of Double-Conv and FR blocks are fused using adaptive weights, and then the fused features are concatenated to the downsampled deep features. The result of the adaptive feature fusion can be expressed as:

$$m_{\uparrow i} = AFF\left(m_{\downarrow i_1},\ m_{\downarrow i_2}, m_{\uparrow i+1}\right)$$
$$= \sigma\left(\theta_i\right) \cdot m_{\downarrow i_1} + \left(1 - \sigma\left(\theta_i\right)\right) \cdot m_{\downarrow i_2} + m_{\uparrow i+1},\tag{6}$$

$$m_{\uparrow} = AFF\left(m_{\downarrow 1_1},\ m_{\downarrow 1_2}, m_{\uparrow 1}\right)$$
$$= \sigma\left(\theta_1\right) \cdot m_{\downarrow 1_1} + \left(1 - \sigma\left(\theta_1\right)\right) \cdot m_{\downarrow 1_2} + m_{\uparrow 1}.\tag{7}$$

where $m_{\downarrow i_1}$ and $m_{\downarrow i_2}$ are the feature maps of the Double-Conv and FR block of the ith downsampling layer, respectively. $m_{\uparrow i}$ is the feature map of the ith upsampling layer, $m_{\uparrow}$ is the final output. $\sigma\left(\theta_i\right)$, $i = 1, 2, 3, 4$ is the learnable weight for fusing feature maps of the Double-Conv and FR block of the ith downsampling layer. θ_i and σ activation functions jointly determine the learnable weights.

3.3. A loss function for LFR-Net

The loss function of LFR-Net consists of three components, which are L1, perceptual loss and haze loss. Lim et al. [31] indicate that L1 loss and perceptual loss perform well in image restoration tasks. Therefore, we choose L1 loss and perceptual loss to optimize LFR-Net.
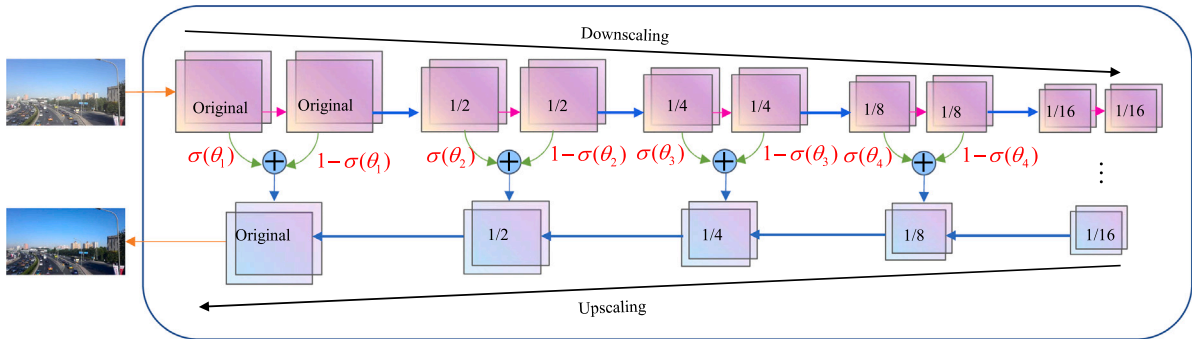
Fig. 3. Adaptive feature fusion model.

For the latent space, we choose the middle feature layer VGG-16 with fixed and pre-trained weights [18]. The formula for the L1 loss and perceptual loss is

$$Loss1 = \|J - LFR(I)\| + \beta_1 \|V_i(J) - V_i(LFR(I))\|^2, \tag{8}$$

Most dehazing methods use the output of the neural network and ground-truth to construct a loss function that optimizes the parameters of the neural network to generate better dehazed results. However, the input hazy images, medium transmission maps, atmospheric light values are ignored in the process of updating the parameters. Therefore, we propose a haze loss that utilizes the input hazy image and the haze-loss image to supervise the training process of the neural network. And the haze-loss image is generated by using the output of the proposed LFR-Net, medium transmission map and atmospheric light value. The haze loss can be formulated as

$$O = LFR(I) \times t + A \times (1 - t), \tag{9}$$

$$Loss2 = \|V_i(I) - V_i(O)\|^2, \tag{10}$$

The final loss function can be expressed as

$$Loss = \|J - LFR(I)\| + \beta_1 \|V_i(J) - V_i(LFR(I))\|^2 + \\ \beta_2 \|V_i(I) - V_i(O)\|^2. \tag{11}$$

where J and I denote the ground-truth and the input hazy image, $LFR(\cdot)$ denotes the output of LFR-Net, β_1 and β_2 denote the regularization coefficients, O denotes the haze-loss image, and V_i denotes the latent space features extracted from VGG-16.

4. Experiments

In this section, the dehazing ability of the LFR-Net is evaluated. First, the experiment settings are introduced. Then, We organize comparison experiments with seven SOTA methods to prove better the proposed method. Finally, we exploit ablation experiments for demonstrating the efficacy of the architecture of our LFR-Net.

4.1. Experiment settings

This section describes the implementation details and the dataset used in the experiments.

4.1.1. Implementation details
We trained our LFR-Net using ITS and OTS in RESIDE [21], respectively. In detail, the medium transmission map, atmospheric light, depth map, and dehazed image generate a haze-loss image using ASM. We use SOTS, and HSTS in RESIDE and I-HAZE [34], and O-HAZE [35] as our test sets and compare the LFR-Net with the nine popular methods (e.g., DCP [8], DehazeNet [12], AOD-Net [13], EPDN [16], TCN [14], IDE [32], RefineDNet [33], DDAP [36] and USID-Net [37]). PSNR and

Table 1
Quantitative comparison between the LFR-Net and SOTA dehazing methods on SOTS, and the best result is presented in bold.

Methods	Indoor		Outdoor	
	PSNR	SSIM	PSNR	SSIM
DCP	18.9795	0.8456	16.7640	0.8215
IDE	14.4178	0.7579	15.6784	0.7622
DehazeNet	21.3312	0.8758	23.0305	0.8927
AOD-Net	18.7131	0.8449	19.6345	0.8914
EPDN	25.1676	0.9254	20.2775	0.8383
RefineDNet	20.2994	0.8357	20.8647	**0.9119**
TCN	16.0179	0.7704	19.9535	0.8858
DDAP	17.9366	0.8295	17.0658	0.8404
USID-Net	17.4533	0.7699	23.8627	0.8886
Ours	**27.9772**	**0.9446**	**24.4735**	0.8954

SSIM [38] are applied to evaluate the haze removal ability of the proposed LFR-Net. The comparison experiments for this work were implemented on a PC with Quadro RTX 2080Ti. The proposed LFR-Net was implemented on Pytorch version 1.9.1 from a Windows 10 system using Python 3.6.13. We adopt Adam optimizer [39] to train our LFR-Net. Furthermore, we applied the following learning rate decay strategy

$$Lr = 0.5 \times (1 + \cos(setp \times \pi \div N)) \times initLr. \tag{12}$$

where $setp$, N and $initLr$ are the number of current training iterations, the total number (100000) of pre-set training iterations, and the initial learning rate whose value is 0.0001, respectively.

4.2. Comparisons on SOTS dataset

In this subsection, the proposed method will be compared quantitatively and qualitatively with the seven methods on the SOTS dataset.

Fig. 4 shows a qualitative comparison between our LFR-Net with the SOTA seven approaches on the SOTS dataset. We have randomly selected four images from SOTS for comparison. We observe that the dehazed results of DCP [8], AOD-Net [13] and RefineDNet [33] are dim compared to ground-truth (P2 in Fig. 4(b), (e), (g)). IDE [32], EPDN [16], and TCN [14] show color distortion in the dehazed results (P4 in Fig. 4(c), (f), (h)). Besides, DCP, IDE, DehazeNet, AOD-Net, EPDN and TCN are not perfect for the sky region, for example, the P4 sky region in Fig. 4(b)–(f) and (h) show different degrees of color distortion. In the experiment, the proposed LFR-Net obtains dehazed results that are more similar to ground-truth, preserving more detail in shadow areas and avoiding the phenomenon of dehazed results being dark and color distortion (P2 and P4 in Fig. 4(i)). Thanks to FR block and AFF model to extract and fuse deep features, and haze loss being introduced, our LFR-Net takes full advantage of the information in the dataset and achieves more satisfying results in comparison experiments.

Table 1 shows a quantitative comparison between our LFR-Net and the nine methods. Our LFR-Net outperformed all SOTA methods on the SOTS indoor dataset. The PSNR and SSIM are 27.977 and 0.9446, which
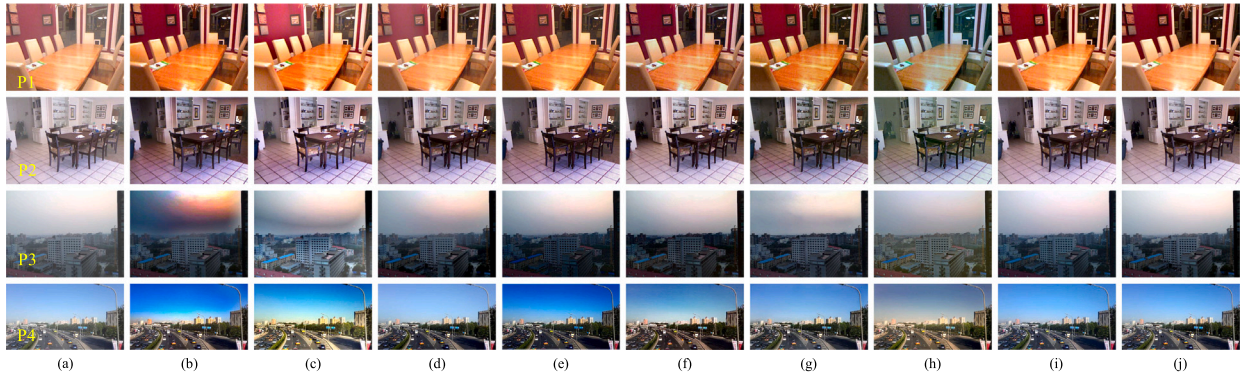
Fig. 4. Qualitative comparison between the LFR-Net and well-known methods on SOTS. (a) Hazy images, (b) DCP [8], (c) IDE [32], (d) DehazeNet [12], (e) AOD-Net [13], (f) EPDN [16], (g) RefineDNet [33], (h) TCN [14], (i) Proposed method, (j) Ground-truth.

Fig. 5. LFR-Net and well-known methods were compared qualitatively on synthetic images of the HSTS datasets. (a) Hazy images, (b) DCP [8], (c) IDE [32], (d) DehazeNet [12], (e) AOD-Net [13], (f) EPDN [16], (g) RefineDNet [33], (h) TCN [14], (i) Proposed method, (j) Ground-truth.
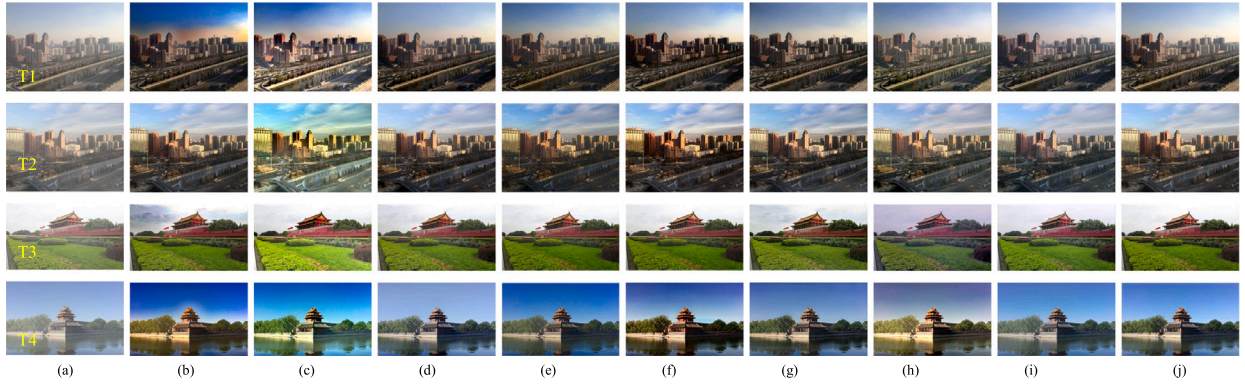
Table 2
Quantitative comparison between LFR-Net and the SOTA dehazing method on images, and the best result is presented in bold.

IMAGES	Metrics	DCP	IDE	DehazeNet	AOD-Net	EPDN	RefineDNet	TCN	Ours
T1	PSNR	13.057	14.834	22.565	18.605	18.472	19.248	22.719	**27.935**
	SSIM	0.8562	0.7956	0.9136	0.8903	0.8738	0.9241	0.9016	**0.9359**
T2	PSNR	17.973	15.677	28.453	20.430	23.591	24.327	25.670	**28.513**
	SSIM	0.8858	0.8082	0.9238	0.8867	0.9051	0.9325	0.9025	**0.9504**
T3	PSNR	20.023	15.762	24.6898	25.768	25.388	24.342	18.658	**26.494**
	SSIM	0.8845	0.7965	0.8568	0.8693	0.8794	0.8890	0.8112	**0.8932**
T4	PSNR	16.642	18.869	21.911	18.148	19.6516	22.177	18.434	**25.341**
	SSIM	0.8555	0.8640	0.9278	0.8648	0.8685	**0.9429**	0.8773	0.9056

exceed the second best method by 2.81 and 0.0192, respectively. In addition, our LFR-Net obtains the best and second best PSNR and SSIM results for SOTS outdoor images, respectively.

4.3. Comparisons on HSTS datasets

In this subsection, we did comparison experiments on the synthetic test set and the real-world test set of HSTS, respectively. Since the real-world images do not have ground-truth, we only did quantitative comparison experiments on the synthetic test set.

We randomly selected four images from the synthetic test set of HSTS for quantitative and qualitative analysis. The prior knowledge of DCP is not universally applicable, color distortion exists in the sky region in the dehazed results (T1, T2 and T3 in Fig. 5(b)). The IDE performs well on dehazing, but due to its overexposure, the dehazed images suffer from serious color distortion (T1, T2, T3, and T4 in Fig. 5(c)). DehazeNet and AOD-Net do not work well for shadow areas, for instance, T1 and T2 in Fig. 5(d) and (e), the shaded areas of DehazeNet and AOD-Net are too dim compared to the ground-truth,

and details of the shaded areas are lost. The dehazed results of EPDN and TCN show color distortion, where the dehazed results of EPDN and TCN (T1, T2 and T4 in Fig. 5(f) and (h)) show yellowish color and cool color respectively, and the dehazed results of TCN also show distortion in the sky region. RefineDNet fuses the dehazed results of the neural network and DCP, resulting in color distortion in the sky region of the dehazed result. Our LFR-Net performs better in the color restoration, for example, in T2, the cars and trees in the shaded area can be observed (T2 in Fig. 5(i)). In T4, the proposed method is the best for the recovery of the sky region (T4 in Fig. 5(i)). Table 2 shows the quantitative comparison between the proposed method and the seven SOTA methods. Our LFR-Net takes special treatment of deep-level features and outperforms all methods except T4 on PSNR and SSIM.

We also conducted qualitative comparison experiments on the real-world test set of HSTS. Fig. 6 shows the comparison results of LFR-Net and six SOTA methods on a real-world test set. As the sky region does not satisfy the dark channel prior, the dehazed result of DCP is enormous color distortion and dark (R1, R2 and R3 in Fig. 6(b)). The exposure properties of IDE enable excellent visualization of the dehazed
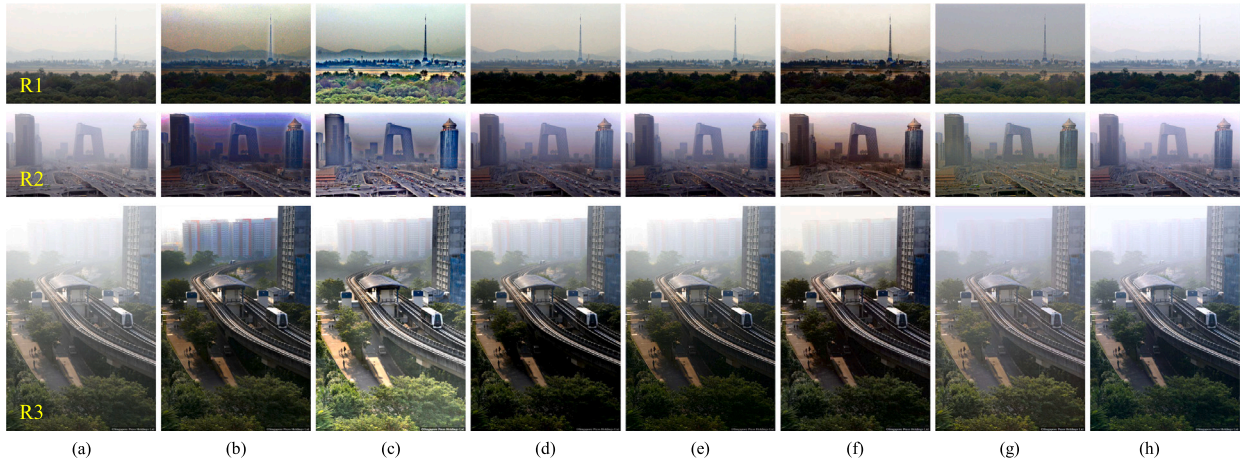
Fig. 6. The LFR-Net and well-known methods were compared qualitatively on the real-world images of HSTS. (a) Hazy images, (b) DCP [8], (c) IDE [32], (d) DehazeNet [12], (e) AOD-Net [13], (f) EPDN [16], (g) TCN [14], (h) Proposed method.
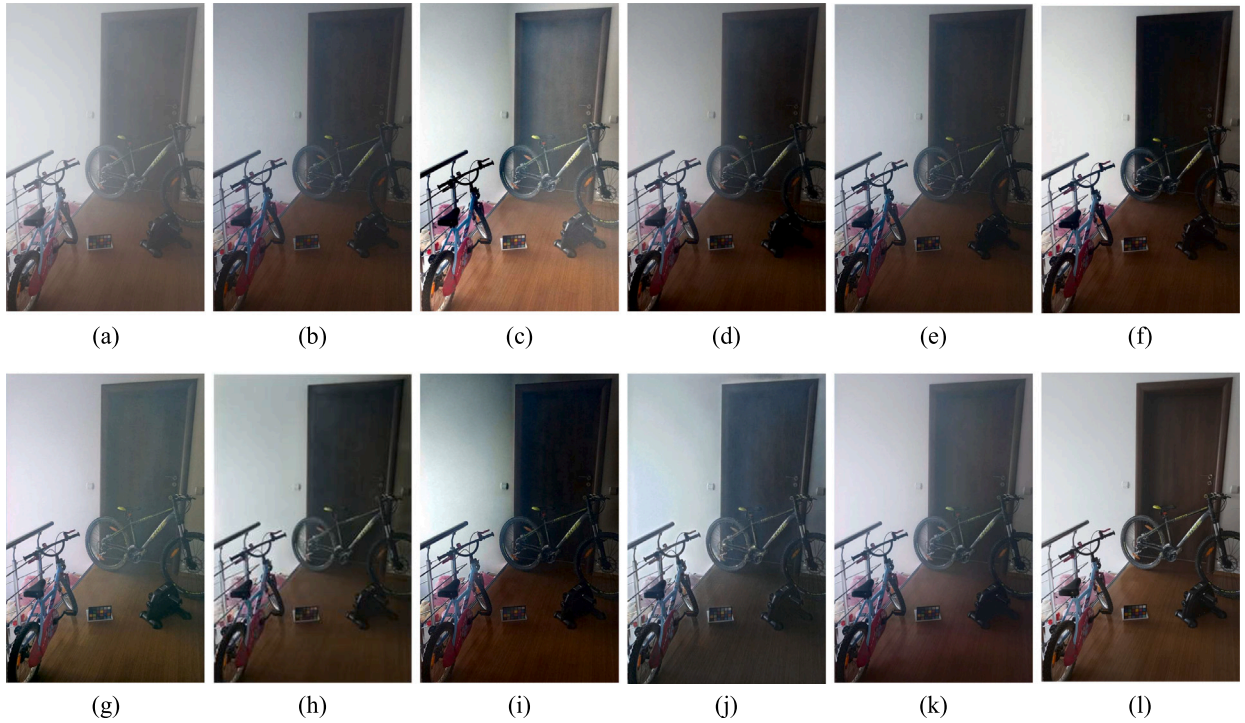

Fig. 7. The LFR-Net and well-known methods were compared qualitatively on the real-world images of I-HAZE. (a) Hazy images, (b) DCP [8], (c) IDE [32], (d) DehazeNet [12], (e) AOD-Net [13], (f) EPDN [16], (g) RefineDNet [33], (h) TCN [14], (i) DDAP [36], (j) SUID-Net [37], (k) Proposed method, (l) Ground-truth.

images, but the dehazed images also show color distortion and do not work well in the sky region (R1 and R2 in Fig. 6(c)). DehazeNet and AOD-Net are effective in dehazing, but the dehazed image is dark, for example, the tree areas recovered by DehazeNet and AOD-Net are significantly darker and the details are not restored (R1 and R3 in Fig. 6(d) and (e)). EPDN successfully removed the haze, but the sky region of the dehazed images shows color distortion, and the trees in the images are dark (R1 and R2 in Fig. 6(f)). TCN avoids dark details while removing haze, but there is color distortion in dehazed results (R1 and R2 in Fig. 6(g)). Our LFR-Net extracts and recovers the deep features and also adds haze-image loss to the loss function. Therefore, our dehazed images not only remove haze but also have a high-quality

image, avoiding color distortion and restoring more details. We can observe that LFR-Net recovers more details of the tree area (R1 and R3 in Fig. 6(h)).

4.4. Comparisons on real-world datasets

We conducted comparison experiments on the I-HAZE [34] and O-HAZE [35] datasets to comprehensively evaluate the proposed method. From Table 3, LFR-Net achieved the best performance on PSNR and SSIM on the test dataset I-HAZE compared with SOTA methods. Compared with SOTA methods on the test dataset O-HAZE, the proposed method achieved the third-best results on PSNR.

Table 3

Quantitative comparison between LFR-NET and SOTA methods on the I-HAZE and O-HAZE datasets, and the best result is presented in bold.

Methods	I-HAZE		O-HAZE	
	PSNR	SSIM	PSNR	SSIM
DCP	15.259	0.6973	15.947	0.5583
IDE	14.702	0.6998	13.423	0.6219
DehazeNet	15.208	06746	15.391	0.5754
AOD-Net	14.732	0.6733	15.003	0.5187
EPDN	14.947	0.6869	**16.906**	0.6342
TCN	16.450	0.7139	14.704	0.6436
RefineDNet	15.781	0.7121	16.616	0.5904
DDAP	10.858	0.5417	12.713	0.5755
USID-Net	15.165	0.6209	15.594	**0.6446**
OURS	**16.510**	**0.7312**	16.407	0.6134

Table 4

Quantitative comparison between the LFR-Net and the SOTA dehazing method on SOTS.

Baseline	✓	✓	✓	✓
NetWork 1		✓	✓	✓
NetWork 2			✓	✓
LFR-Net				✓
PSNR	21.57	23.35	24.03	24.47

Additionally, We also conducted a qualitative comparison experiment. Fig. 7 show that DCP [8], DehazeNet [12], AOD-Net [13], and EPDN [16] generated darker image compared with the ground-truth images (Figs. 7(b) and 7(d)-7(f)). IDE [32], RefineDNet [33], TCN [14], DDAP [36], and USID-Net [37] show color distortion (Figs. 7(c) and 7(g)-7(j)). LFR-Net generates the Image, which is closer to the ground-truth style than IDE and has higher visibility than other SOTA methods (Fig. 7(k)).

4.5. Ablation study

To prove the structural superiority of LFR-Net, we carried out ablation study to assess the usefulness of each module, including feature residual block, adaptive feature fusion model and haze loss. We choose U-Net, L1 loss and perceptual loss as the baseline network. Based on the baseline network, each module is added in turn to perform the ablation experiment: NetWork 1: baseline+FR, feature residual block is added to each scale of U-Net to constitute NetWork 1. NetWork 2: baseline+FR+ AFF, based on Network1, we use the adaptive feature fusion model to fuse shallow features and deep features when the generator generates feature maps at different scales, respectively. LFR-Net: LFR-Net has the same network architecture as NetWork 2, and adds haze loss to the loss function to better optimize the parameters of the neural network. The test set used in the ablation study is outdoor images of SOTS, and we used PSNR as an evaluation metric.

The results of the ablation experiments are displayed in Table 4. By adding the feature residual block, adaptive feature fusion model and haze loss respectively to the baseline, the PSNR gradually improved and finally reached 24.47. Therefore, the effectiveness of the architecture of the LFR-Net is demonstrated.

5. Conclusions

This work proposed a local feature residual network for single image dehazing, which takes full advantage of the information, including the hazy images, media transmission maps and atmospheric light values, in the training dataset. Also, the feature residual block of the proposed LFR-Net is exploited to enhance the deep features extraction capacity of the encoder. The adaptive feature fusion model is used to fuse the shallow and deep features of the encoder, which is finally connected to the decoder through a skip connection to preserve more details in the restored image. A haze loss is introduced to the loss function

during the LFR-Net training process to update the parameters of the network utilizing haze-free images, hazy images, medium transmission maps, and atmospheric light values. Finally, we conducted extended experiments to demonstrate the superiority of LFR-Net from qualitative and quantitative aspects and organized ablation experiments to prove the usefulness of the LFR-Net structure. Moreover, most learning-based dehazing methods use synthetic datasets to train the weights, and poorly process real-world images. To overcome this problem, we will conduct unsupervised image dehazing research in the future.

CRediT authorship contribution statement

Xinjie Xiao: Proposed idea, Wrote the code and the manuscript. **Zhiwei Li:** Guided the writing manuscript, Provided language polish. **Wenle Ning:** Conducted experiments. **Nannan Zhang:** Conducted experiments. **Xudong Teng:** Provided guidance.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work.

Data availability

No data was used for the research described in the article.

References

[1] Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst 2015;28.

[2] Fujii K, Kawamoto K. Generative and self-supervised domain adaptation for one-stage object detection. Array 2021;11:100071.

[3] Wei H, Huang Y. Online multiple object tracking using spatial pyramid pooling hashing and image retrieval for autonomous driving. Machines 2022;10(8):668.

[4] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection. IEEE Trans Pattern Anal Mach Intell 2011;34(7):1409–22.

[5] McCartney EJ. Optics of the atmosphere: scattering by molecules and particles. New York; 1976.

[6] Narasimhan SG, Nayar SK. Contrast restoration of weather degraded images. IEEE Trans Pattern Anal Mach Intell 2003;25(6):713–24.

[7] Liu Z, He Y, Wang C, Song R. Analysis of the influence of foggy weather environment on the detection effect of machine vision obstacles. Sensors 2020;20(2):349.

[8] He K, Sun J, Tang X. Single image haze removal using dark channel prior. IEEE Trans Pattern Anal Mach Intell 2010;33(12):2341–53.

[9] Zhu Q, Mai J, Shao L. A fast single image haze removal algorithm using color attenuation prior. IEEE Trans Image Process 2015;24(11):3522–33.

[10] Tan RT. Visibility in bad weather from a single image. In: 2008 IEEE conference on computer vision and pattern recognition. IEEE; 2008, p. 1–8.

[11] Berman D, Avidan S, et al. Non-local image dehazing. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, p. 1674–82.

[12] Cai B, Xu X, Jia K, Qing C, Tao D. Dehazenet: An end-to-end system for single image haze removal. IEEE Trans Image Process 2016;25(11):5187–98.

[13] Li B, Peng X, Wang Z, Xu J, Feng D. Aod-net: All-in-one dehazing network. In: Proceedings of the IEEE international conference on computer vision. 2017, p. 4770–8.

[14] Shin J, Park H, Paik J. Region-based dehazing via dual-supervised triple-convolutional network. IEEE Trans Multimed 2021;24:245–60.

[15] Liang T, Li Z, Ren Y, Mao Q, Zhou W. A progressive single-image dehazing network with feedback mechanism. IEEE Access 2021;9:158091–7.

[16] Qu Y, Chen Y, Huang J, Xie Y. Enhanced pix2pix dehazing network. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019, p. 8160–8.

[17] Chen D, He M, Fan Q, Liao J, Zhang L, Hou D, et al. Gated context aggregation network for image dehazing and deraining. In: 2019 IEEE winter conference on applications of computer vision. IEEE; 2019, p. 1375–83.

[18] Qin X, Wang Z, Bai Y, Xie X, Jia H. FFA-Net: Feature fusion attention network for single image dehazing. In: Proceedings of the AAAI conference on artificial intelligence. Vol. 34. No. 07. 2020, p. 11908–15.

[19] Han W, Zhu H, Qi C, Li J, Zhang D. High-resolution representations network for single image dehazing. Sensors 2022;22(6):2257.

[20] Jiang X, Zhao C, Zhu M, Hao Z, Gao W. Residual spatial and channel attention networks for single image dehazing. Sensors 2021;21(23):7922.

[21] Li B, Ren W, Fu D, Tao D, Feng D, Zeng W, et al. Benchmarking single-image dehazing and beyond. IEEE Trans Image Process 2018;28(1):492–505.

[22] Ren W, Ma L, Zhang J, Pan J, Cao X, Liu W, et al. Gated fusion network for single image dehazing. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, p. 3253–61.

[23] Wu H, Qu Y, Lin S, Zhou J, Qiao R, Zhang Z, et al. Contrastive learning for compact single image dehazing. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021, p. 10551–60.

[24] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer; 2015, p. 234–41.

[25] Yamanaka J, Kuwashima S, Kurita T. Fast and accurate image super resolution by deep CNN with skip connection and network in network. In: International conference on neural information processing. Springer; 2017, p. 217–25.

[26] Zhang K, Li Y, Zuo W, Zhang L, Van Gool L, Timofte R. Plug-and-play image restoration with deep denoiser prior. IEEE Trans Pattern Anal Mach Intell 2021.

[27] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, p. 770–8.

[28] Wang X, Li Z, Shan H, Tian Z, Ren Y, Zhou W. Fastderainnet: A deep learning algorithm for single image deraining. IEEE Access 2020;8:127622–30.

[29] Fu X, Liang B, Huang Y, Ding X, Paisley J. Lightweight pyramid networks for image deraining. IEEE Trans Neural Netw Learn Syst 2019;31(6):1794–807.

[30] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: European conference on computer vision. Springer; 2014, p. 818–33.

[31] Lim B, Son S, Kim H, Nah S, Mu Lee K. Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017, p. 136–44.

[32] Ju M, Ding C, Ren W, Yang Y, Zhang D, Guo YJ. IDE: Image dehazing and exposure using an enhanced atmospheric scattering model. IEEE Trans Image Process 2021;30:2180–92.

[33] Zhao S, Zhang L, Shen Y, Zhou Y. RefineDNet: A weakly supervised refinement framework for single image dehazing. IEEE Trans Image Process 2021;30:3391–404.

[34] Ancuti C, Ancuti CO, Timofte R, Vleeschouwer CD. I-HAZE: a dehazing benchmark with real hazy and haze-free indoor images. In: International conference on advanced concepts for intelligent vision systems. Springer; 2018, p. 620–31.

[35] Ancuti CO, Ancuti C, Timofte R, De Vleeschouwer C. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018, p. 754–62.

[36] Li Z, Shu H, Zheng C. Multi-scale single image dehazing using Laplacian and Gaussian pyramids. IEEE Trans Image Process 2021;30:9270–9.

[37] Li J, Li Y, Zhuo L, Kuang L, Yu T. USID-Net: Unsupervised single image dehazing network via disentangled representations. IEEE Trans Multimed 2022.

[38] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004;13(4):600–12.

[39] Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014, arXiv preprint arXiv:1412.6980.