ORIGINAL ARTICLE

Joint dual-teacher distillation and unsupervised fusion for unpaired real-world image dehazing

Yingxu Qiao[1] · Xiyan Zhan[2] · Fen Luo[2] · Zhanqiang Huo[2]

Abstract

Existing learning-based dehazing algorithms struggle to deal with real world hazy images for lack of paired clean data. Moreover, most dehazing methods require significant computation and memory. To address the above problems, we propose a joint dual-teacher knowledge distillation and unsupervised fusion framework for single image dehazing in this paper. First, considering the complex degradation factors in real-world hazy images, two synthetic-to-real dehazing networks are explored to generate two preliminary dehazing results with the heterogeneous distillation strategy. Second, to get more qualified ground truth, an unsupervised adversarial fusion network is proposed to refine the preliminary outputs of teachers with unpaired clean images. In particular, the unpaired clean images are enhanced to deal with the dim artifacts. Furthermore, to alleviate the structure distortion in the unsupervised adversarial training, we constructed an intermediate image to constrain the output of the fusion network. Finally, considering the memory storage and computation overhead, an end-to-end lightweight student network is trained to learn the mapping from the original hazy image to the output of the fusion network. Experimental results demonstrate that the proposed method achieves state-of-the-art performance on real-world hazy images in terms of no-reference image quality assessment and the parameters.

Keywords Generalize · Lightweight · Real-world hazy images · Dual-teacher · Fusion

Introduction

Images captured in hazy weather often have blurred edges, reduced contrast, and shifted color due to the scattering effect of floating practices in the environment. The obvious image degradation caused by haze can significantly affect the performance of computer vision systems, such as object detection [1–3], image segmentation [4], and classification [5]. Single image dehazing aims to recover the haze-free image from the degraded input. As a long-standing research problem in the vision community, it has attracted more and more attention in the computer vision and graphics community.

Mathematically, the haze process can be described with the physical scattering model, which is formulated as follows.

$$X (\mu) = Y (\mu) T (\mu) + A (1 - T (\mu)) \tag{1}$$

where $X (\mu)$ and $Y (\mu)$ indicate the hazy image and the corresponding haze-free version. $T (\mu)$ and A denote the transmission map and the global atmosphere light. Specifically, the transmission map $T (\mu) = e^{-\beta d(\mu)}$ can be expressed with the depth map $d (\mu)$ and the medium extinction coefficient β that reflects the haze density.

Given a hazy image X, recovering its haze-free version Y is a challenging ill-posed problem. Existing methods always estimate the transmission map $T (\mu)$ and the global atmosphere light A with various priors, such as the color attenuation prior [6] and the dark channel prior [7]. Unfortunately, statistical priors do not always hold for hazy images of the real world, leading to limited performance in a complex

✉ Zhanqiang Huo
  hzq@hpu.edu.cn

  Yingxu Qiao
  qiaoyingxu@hpu.edu.cn

  Xiyan Zhan
  zhanxiyan@home.hpu.edu.cn

  Fen Luo
  luofenjsj@hpu.edu.cn

[1] School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454003, China

[2] School of Software, Henan Polytechnic University, Jiaozuo 454000, China

practical scene [8]. Furthermore, these prior-based methods always suffer from nonconvex optimization problems with high computation overhead [9].

With the advent of deep neural networks [10], learning-based methods have achieved excellent performance. Supervised methods either utilize the neural network to estimate global air lights and transmission maps [11, 12], or generate clean images with the large-scale paired dataset [13–16]. However, most synthetic datasets are constructed based on the assumptions that spatially invariant attenuation coefficients, overcast lighting and wavelength-independent [17]. These assumptions do not always hold in many real scenes [18]. The dehazing networks trained with the synthetic dataset cannot perform well in real scenes due to the domain discrepancy between the synthetic and real domains. Since it is difficult to collect large mounts of perfectly aligned paired real data, some unsupervised methods [19–21] try to translate hazy images and unpaired clean images with the cycle constraint through generative adversarial learning (GAN). Nevertheless, these unsupervised GAN-based algorithms rely heavily on selection of the training data due to the unstable training process and model collapse. Recently, some researchers have taken this issue up with synthetic-to-real strategies. Shao et al. [8] proposed the first domain adaptation algorithm to reduce the domain gap by translating images from one domain to another domain. Since there are no ground truths for the real world hazy images, Wu et al. [22] synthesized several types of hazy images and proposed a real image dehazing framework with codebook priors (RIDCP). Taking into account complex degradations in real scenes, Qiao et al. [23] proposed an algorithm with the ensemble network and the intermediate domain module (ENID) to reduce the large distribution discrepancy between the synthetic and the real domains. However, due to the lack of paired clean real data, existing synthetic-to-real algorithms struggle to process the complex degradations in real-world hazy images, especially in the distant regions, such as the sky areas. Furthermore, these methods focus on improving generalization in real scenes without considering memory storage, which may bring difficulty in resource-limited environments. For example, the domain adaptation dehazing method (DAD) has 54.59 million parameters [8].

To improve real-world image dehazing performance with computational cost, we propose a joint dual teacher knowledge distillation and unsupervised fusion framework for single image dehazing in this paper. First, considering that there is no ground truth for real-world hazy images, two synthetic-to-realistic networks RIDCP [22] and ENID [23] are explored as teachers to generate two preliminary dehazing results with a specially designed strategy in the first stage. Specifically, the real-world hazy image is used as input of one teacher with the usual distillation strategy. Furthermore, considering severe information degeneration, the hazy image

is also enhanced to feed into the other teacher to generate another hazy result. Second, to obtain more qualified ground truth, an unsupervised fusion network is proposed to refine the outputs of two selected teachers via adversarial learning in the second stage. To deal with the dim artifacts, unpaired clear images are further enhanced with the color-preserving adaptive histogram equalization(CPAHE) [24] to improve the image quality of dark regions. To alleviate the structure distortion in the unsupervised adversarial training, the result of the fusion network is constrained with the intermediate image, which is obtained with adaptive similarity maps according to the chrominance information and the gradient modulus of the LMN color space. Finally, considering the memory storage and computation overhead, an end-to-end lightweight student network is trained to learn the mapping from the original hazy image to the output of the fusion network.

Compared with state-of-the-art dehazing methods, the proposed method can effectively recover haze-free images with vivid color. As shown in Fig. 1, only the sky regions of the proposed method are blue and visually pleasing. The contributions of the proposed method are summarized as follows:

- We propose a joint dual-teacher distillation and unsupervised fusion framework for unpaired real-world hazy images. Considering that there are no ground truth for real-world hazy images, two synthetic-to-real dehazing networks are explored to generate two preliminary dehazed results with different distillation strategies.
- We propose an unsupervised fusion scheme with a single generative adversarial network to refine preliminary dehazing results of two teachers. Unpaired clean images are enhanced to overcome the dim artifacts. Furthermore, to alleviate the structure distortion in the unsupervised adversarial training, we constructed an intermediate image to constrain the output of the fusion network.
- Comprehensive experiments demonstrate that the proposed method achieves state-of-the-art performance on the real-world hazy images, in terms of no-reference image quality assessment and the parameters.

Related works

Prior-based methods

The classical dehazing algorithms mainly analyze the degradation mechanism of hazy images, and then estimate physical parameters such as transmission maps and atmospheric light based on certain image priors. The clear images are restored from estimated parameters and hazy images based on atmospheric scattering models. He et al. [7] proposed the famous
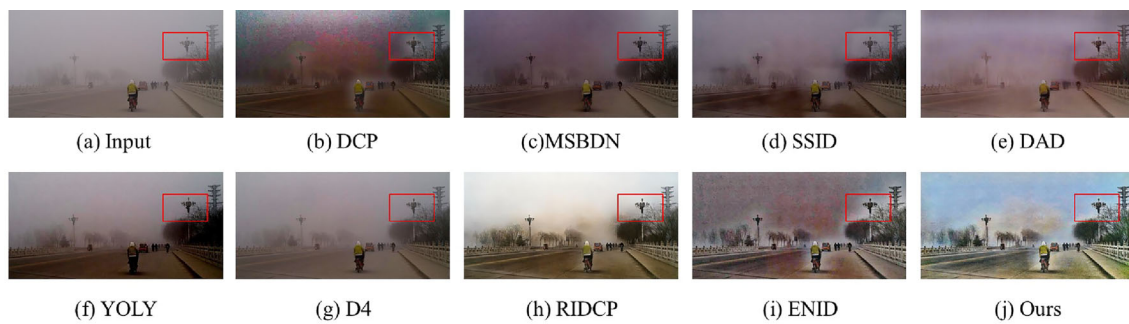
Fig. 1 Visual comparison of dehazing results for real-world haze images. Only sky regions of the proposed method are blue and visually pleasing

dark channel prior by statistically analyzing the channel information of clear images, which meant that clear images have at least one channel with lower brightness except for the sky area. The dehazing algorithm with the dark channel prior can cause obvious color distortion and blocky defects in the sky area of the image. Many subsequent algorithms improved the dark channel prior algorithm from different aspects. Dai et al. [25] combined Retinex theory to optimize the atmospheric scattering model, and decomposed the clear images into a reflection map and a spatially varying illumination map. BorKar and Mukherjee et al. [26] applied the dark channel prior to the Y channel of the YCbCr color space and estimated the transmission map in the RGB color space. Finally, the brightness information of the dehazed image was adjusted to avoid the dim artifact in some regions. Although these methods improved the dehazed results to some extent, they sometimes led to obvious color distortion for the varicolored hazy images. In addition to the dark channel prior, many algorithms proposed different types of priors. Kaur et al. [27] first estimated the transmission map and atmospheric light based on the gradient channel prior and optimized the transmission map using L0-guided filtering. Although it partially corrected for the problem of color distortion, the hyper-parameter selection was quite cumbersome. To improve the efficiency, Liu et al. [28] proposed a rank one prior for the statistical analysis of images, which greatly improved the efficiency in practice. Ju et al. [29] proposed a region line prior based on the relationship between the corresponding regions of foggy images and clear images. Based on this novel prior, the dehazing task was transformed into a two-dimensional optimization function. This algorithm alleviated oversaturation to some extent in the dehazed results. Liu et al. [24] transformed the dehazing procedure into a task of improving image visibility and contrast. They performed contrast and low-light image enhancement to obtain two intermediate images and then fused the intermediate images to obtain the corresponding dehazed results. Although prior-based algorithms achieved good dehazed results, they sometimes led to obvious artifacts and some issues when priors were not held in realistic scenes.

Learning-based methods

Recently, deep learning-based algorithms have made great progress in many fields [30–32], which can be roughly categorized into three groups based on the characteristics of the datasets: supervised methods, semi-supervised methods, and unsupervised dehazing methods. Supervised dehazing methods mainly utilize paired hazy/clear images, transmission maps, and atmospheric light in the synthesized dataset to train the neural network. Some algorithms first used neural networks to estimate the corresponding parameters such as the transmission map and atmospheric light, and then haze-free images were recovered through atmospheric scattering models. Li et al. [33] proposed the first deep learning-based dehazing algorithm. To avoid the cumulative error of estimating the transmission map and atmospheric light parameters separately, many dehazing algorithms transformed the atmospheric scattering model, converted the parameters of the transmission map and atmospheric light into an intermediate variable, and then obtained the clear dehazing results with the transformed model. To further reduce errors caused by parameter estimation, more algorithms learned the mapping of hazy images to clear images with improved network structures, loss functions and other aspects [14, 34–36]. Park et al. [37] proposed the learning-based fusion dehazing network to improve performance on outdoor hazy images. Due to the domain discrepancy, the dehazing network trained with synthetic datasets cannot perform well in real scenes. Many semi-supervised dehazing algorithms [8, 9, 22, 38, 39] were proposed one after another. Li et al. [9] proposed a semisupervised dehazing framework by exploring the dark channel before constraining the real-world hazy images. Furthermore, An et al. [38] proposed two subnets to estimate the atmospheric light and transmission map in semisupervised dehazing methods. To bridge the domain gap between the synthetic and the real domains, Shao et al. [8] introduced image translation modules and depth information into the dehazing algorithm, which improved generalization in the real world. Using several physical priors in a loss committee, Chen et al. [39] fine-tuned the pretrained dehazing model

with three priors to improve generalization performance. Recently, Wu et al. [22] proposed the novel codebook priors to improve the dehazing performance on real-world hazy images. In addition, some unsupervised dehazing algorithms [19, 40, 41] also achieved great performance on real-world hazy images. Yang et al. [19] proposed a self-augmented image dehazing algorithm by decomposing the transmission map into depth and density in the CycleGAN framework [40]. Li et al. [42] decomposed the single hazy image into the atmospheric light layer, transmission map layer, and the scene radiance layer, then the clear image was recovered in a self-supervised manner. The loss based on the color attenuation prior was proposed to constrain the dehazed results.

Knowledge distillation

Knowledge distillation attempts to transfer knowledge from a complicated teacher model to a light student model. As one branch of transfer learning, knowledge distillation has been widely used in many fields such as detection [43] and style transfer [44]. Common approaches can be divided into distillation on logits, the bounding box, the output, and feature maps. Wang et al. [43] analyzed the inconsistency of the target results and proposed an adaptive sample assignment to improve the quality of the pseudo-boxes. To improve the performance of semi-supervised shadow detection, Chen et al. [45] proposed a mean teacher model by exploring their complementary information and unlabeled images. Tang et al. [46] proposed a simple distillation framework with a data ensemble strategy to generate plenty of soft labels as targets, which significantly improved the performance of the student in the semi-supervised objection. Motivated by contrastive learning and mean-teacher self-training, Cao et al. [47] proposed a contrastive framework to enhance beneficial learning signals without requiring labels in the target domain. Deng et al. [48] proposed a harmonious framework to improve object detection by exploring the consistency of the localization scores and the classification. Recently, Hong et al. [49] proposed the first framework for knowledge distillation on paired synthetic hazy images with a heterogeneous task learning mechanism, an image reconstruction network was treated as the teacher to assist intermediate feature learning in training the student for synthetic image dehazing.

Unlike dehazing methods in synthetic paired datasets [37, 49], the proposed method mainly deals with the issue of unpaired real-world hazy images. Furthermore, unlike the single teacher in [49], the proposed method tries to explore the advantages of dual teachers with the unsupervised fusion strategy.

Method

Framework

As shown in Fig. 2, the overall framework of the proposed dehazing method consists of two teachers, a light student network and the unsupervised fusion network. The teachers generate the initial results with the fixed parameters during the training process. Two initial dehazed images are fused to generate the final result with the GAN-based unsupervised fusion network; then the student is trained to recover a hazy-free image from its hazy counterpart under the supervision of the final result. Only the student is used in the test practices.

The real-world hazy image X is fed to the student and one teacher to obtain the dehazed results Y and Y_2. Taking into account the complex degradation in real scenes, the real hazy image X is enhanced with the CPAHE [24] and the Homo_filter, respectively. The enhanced images are further fused according to the realness and adaptive similarity map to the real hazy image X, and the fused result X_{en} is then fed to the other teacher network to generate the dehazed image Y_1. To make full use of the different dehazed results, Y_1 and Y_2 are further refined through the fusion network with unsupervised adversarial training. The global discriminator D_G and the local discriminator D_L try to discern the fused result Y_{en} with the enhanced image Z_{en}.

Specifically, the unpaired clear image Z is further enhanced to restore the details of the dark regions, which is helpful to alleviate dim artifacts in single image dehazing. To avoid structure distortion, the result of the fusion network is constrained by the intermediate image Y_3, which is obtained with adaptive similarity maps according to the chrominance information and the gradient modulus of the LMN color space [50]. Finally, we propose a compact student network to learn the mapping function from the real-world hazy image X to the corresponding haze-free image Y supervised by Y_{en}, which is the output of the unsupervised fusion network. The dehazed result Y is also constrained with the dark channel loss, which can mitigate the haze residue in the dehazing process.

Fusion and enhancement

In the unsupervised fusion network, two preliminary dehazing results Y_1 and Y_2 are mixed with the adversarial learning strategy. To avoid the structure distortion during the unsupervised adversarial training, the intermediate image Y_3 is obtained with adaptive similarity maps according to the chrominance information and gradient modulus of the LMN color space [50].
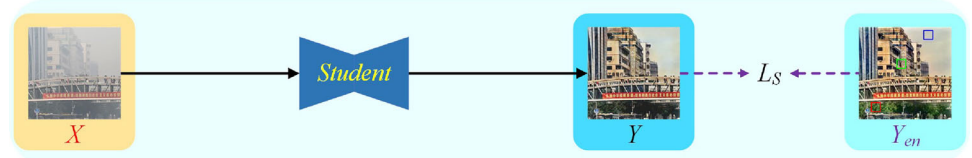
Fig. 2 The overall architecture of the proposed joint dual-teacher distillation and unsupervised fusion framework. Only the lightweight student is used in the test
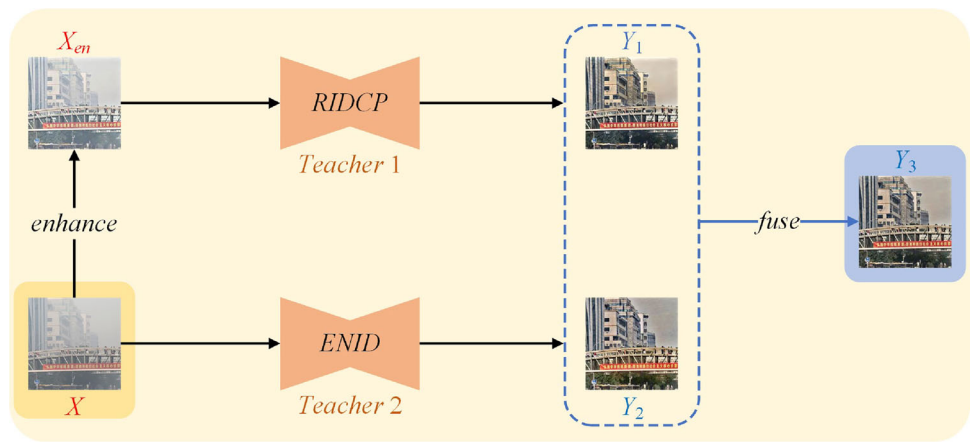
(a) Generate preliminary results in the first stage.
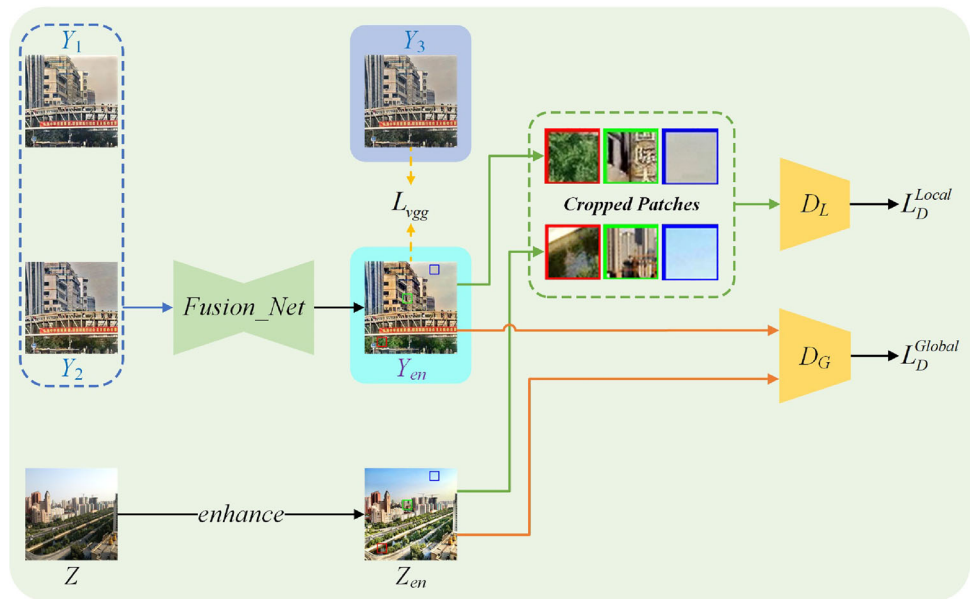
(b) Refine the preliminary results in the second stage.

(c) An end-to-end lightweight student network is trained in the third stage.

To obtain the chrominance information, we first calculated the M and N channels from the RGB color space as follows.

$$M = 0.114 \cdot B + 0.04 \cdot G + 0.30 \cdot R. \tag{2}$$

$$N = 0.17 \cdot B - 0.60 \cdot G + 0.34 \cdot R. \tag{3}$$

Then, the similarity map according tho the the chrominance information is computed as:

$$S^C(y) = \frac{2M_1(y) \cdot M_2(y) + 130}{M_1^2(y) + M_2^2(y) + 130} \cdot \frac{2N_1(y) \cdot N_2(y) + 130}{N_1^2(y) + N_2^2(y) + 130} \tag{4}$$

where M_1 and N_1 stand for M and N channels of Y_1, M_2 and N_2 related to M and N channels of Y_2.

As for the similarity related to the gradient information can be calculated as follows:

$$S^G(y) = \frac{2G_1(y) \cdot G_2(y) + 160}{G_1^2(y) + G_2^2(y) + 160} \tag{5}$$

We can calculate the overall similarity maps between the dehazed image and the original image as follows:

$$S^{CG}(y) = \left[S^C(y)\right]^{0.4} \cdot S^G(y) \tag{6}$$

The weight maps of Y_1 and Y_2 can be calculated with the Softmax function, which defined as follows:

$$\begin{bmatrix} W_1(y) \\ W_2(y) \end{bmatrix} = soft \max \left(\begin{bmatrix} S_1(y) \\ S_2(y) \end{bmatrix} \right) \tag{7}$$

The image Y_3 can be obtained by combining the preliminary dehazed results Y_1 and Y_2 with their weight maps as follows:

$$Y_3 = W_1(y) \cdot Y_1 + W_2(y) \cdot Y_2 \tag{8}$$

Y_3 is used as an intermediate image to constrain the output of the unsupervised fusion network.

Network structure

To improve the generalization on the real-world hazy image, two synthetic-to-realistic dehazing networks RIDCP [22] and ENID [23] are selected as teachers. The RIDCP encoder consists of some residual layers and four down-sampling layers followed by four residual transformer blocks. As for the decoder, it has a symmetric structure with the encoder. Specifically, the deformable convolution is introduced to align different features. In addition, we adopt the multi-scale boosted module dehazing network proposed in [23] as the second teacher, which consists of an encoder module, a boosted decoder module, and a dense feature fusion module. Different from the original architecture proposed in [13], the proposed encoder module contains dual learning branches with different parameters to deal with the distribution discrepancy between the synthetic hazy dataset and the realistic dataset. The Strengthen-Operate-Subtract strengthening strategy [13] is incorporated into the decoder to recover haze-free images. To remedy the missing spatial information caused by the down-sampling operation in U-Net, the dense feature fusion module exploits features from nonadjacent levels with the back-projection technique.

To generate a more visually pleasing result, we adopt a U-Net generator as the fusion network to fuse Y_1 and Y_2 with the adversarial learning strategy [51]. The original clean image Z is enhanced with CPAHE [24] to improve image quality and overcome the dim artifact. The global discriminator and the local discriminator try to discern whether the fused result Y_{en} is similar to the enhanced images Z_{en}.

Unlike the teachers, a lightweight dehazing network with an effective feature attention block is proposed as the student. The student first employs one regular convolutional layer and four down-sampling operations to quarter the input hazy image in side length. Each down-sampling module contains one 3*3 convolution layer with stride 1 and two convolution layers with stride 2 following a ReLU per layer. The down-sampled features are fed to one attention module in the bottleneck layer, and then four upsampling modules and one regular convolution layer are employed to restore the haze-free image. Furthermore, instead of the common skip connection with concatenation or addition, the adaptive mixup fusion strategy [52] is introduced in the student to exchange information between the shallow features and high-level features.

To enhance the feature extraction ability, we also introduce the attention mechanism in the bottleneck layer of the student. As shown in Fig. 3, the feature maps F_{in} in the low-resolution space are grouped with horizontal kernels and vertical kernels inspired by coordinate attention [53]. To promote information exchange between channels, the channels of the pooled feature maps F_h and F_v are squeezed from C to C/r, then F_h and F_v perform element-wise product operation to embed position information. Finally, the channels are changed to C with a Conv-Shuffle-Sigmoid block, and the feature map F_{out} can be generated with the attention map to emphasize the regions of interest.

Loss function

In the proposed framework, all the parameters of two teachers are fixed. The GAN-based unsupervised fusion network and the student are updated with different loss functions in the training process.

Loss function for the fusion network. Inspired by the unsupervised image-to-image translation, a U-Net is applied as a generator to fuse initial dehazed images of two teachers into the refined result with adversarial loss and structure-preserving loss. The loss function for the GAN-based unsupervised fusion network is defined as follows:

$$L_F = L_{adv} + \lambda_{vgg} L_{vgg} \tag{9}$$
$$\begin{aligned} L_{adv} = &\, E_{Y_{en} \sim P_{fake}}[(D_G(Y_{en}, Z_{en}) - 1)^2] \\ &+ E_{Z_{en} \sim P_{real}}[D_G(Z_{en}, Y_{en})^2] \\ &+ E_{Y_{en-patch} \sim P_{fake}}[(D_L(Y_{en-patch}) - 1)^2] \end{aligned} \tag{10}$$

where L_{adv} stands for the adversarial losses, Y_{en} and Z_{en} relate to examples sampled from the output distribution P_{fake} and haze-free distributions P_{real}. L_{adv} is employed to make the output results similar to the haze-free images.

In addition to the adversarial loss, the perceptual loss function L_{vgg} is also utilized to preserve the structure of the fused result. The perceptual loss with the pre-trained VGG19 net-

Fig. 3 The overall architecture of the proposed joint fusion and dual teachers knowledge distillation framework. Only the lightweight student is used in test
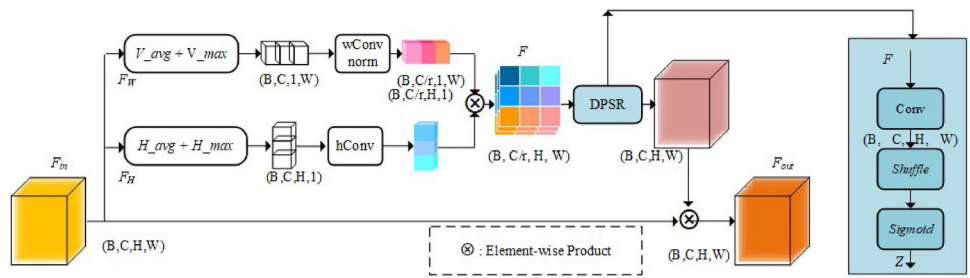
work is defined as follows:

$$
\begin{aligned}
L_{vgg} &= \frac{1}{WH} \sum_{m=1}^{W} \sum_{n=1}^{H} (\varphi_{i,j}(Y_{en}) - \varphi_{i,j}(Y_3))^2 \\
&+ \frac{1}{WH} \sum_{m=1}^{W} \sum_{n=1}^{H} (\varphi_{i,j}(Y_{en-patches}) \\
&- \varphi_{i,j}(Y_{3-patches}))^2
\end{aligned}
\tag{11}
$$

where $\varphi_{i,j}(\cdot)$ represents the feature maps of the pre-trained VGG19 network.

In addition, a global discriminator D_G and a local discriminator D_L are also used to discern the outputs Y_{en} and Z_{en} in the unsupervised adversarial training process. The loss functions for discriminators are defined as follows:

$$
\begin{aligned}
L_D^{Global} &= E_{Y \sim P_{fake}}[D_G(Y_{en}, Z_{en})^2] \\
&+ E_{Z_{en} \sim P_{real}}[(D_G(Z_{en}, Y) - 1)^2]
\end{aligned}
\tag{12}
$$

$$
\begin{aligned}
L_D^{Local} &= E_{Y \sim P_{real}}[(D_L(Z_{en-patches}) - 1)^2] \\
&+ E_{Y \sim P_{fake}}[(D_L(Y_{en-patches})^2]
\end{aligned}
\tag{13}
$$

where $Z_{en-patches}$ and $Y_{en-patches}$ used in the training of local discriminator stand for patches randomly sampled from the Z_{en} and Y_{en}.

Through adversarial training, the output Y_{en} of the fusion network is similar to the high-quality image Z_{en}. We then adopt Y_{en} as the corresponding haze-free image to train the student with the input of the original hazy image X.

Loss function for the student. The full loss function L_s for the student is formulated as:

$$
L_S = \lambda_{pix} L_{pix} + \lambda_{per} L_{per} + \lambda_{dcp} L_{dcp} + \lambda_{tv} L_{tv}
\tag{14}
$$

where L_{pix} stands for the reconstruction loss, L_{per} relates to the perceptual loss with the VGG19 networks, L_{dcp} stands for the dark channel loss, and the L_{tv} is the TV (total variation) loss.

To recover the haze-free image, the mean absolute error is adopted to ensure the result Y close to the refined result Y_{en}.

The pixel loss is defined as:

$$
L_{pix} = \frac{1}{N} \sum_{i-1}^{N} \| Y_{en} - Y \|_2
\tag{15}
$$

where N stands for the number of hazy images, Y and Y_{en} represent the dehazing results of the student and the refined result.

The perceptual loss is explored with the pre-trained VGG19 network, which is defined as follows:

$$
L_{per} = \frac{1}{WH} \sum_{m=1}^{W} \sum_{n=1}^{H} (\varphi_{i,j}(Y_{en}) - \varphi_{i,j}(Y))^2
\tag{16}
$$

where $\varphi_{i,j}(Y)$ and $\varphi_{i,j}(Y_{en})$ stand for the feature maps of Y and Y_{en} in VGG19 network.

In addition to the pixel loss and the perceptual loss, the unsupervised dark channel loss and the total variation losses are also adopted to ensure that the recovered images have the characteristics of the clean images. Dark channel prior is widely used in the image dehazing task. Inspired by the dark channel prior, the dark channel loss is defined as:

$$
L_{dcp} = \frac{1}{N} \sum_{i=1}^{N} \|Dark(Y)\| = \min_{y \in Q(y)} \left[\min_{c \in \{r,g,b\}} Y^c(y) \right]
\tag{17}
$$

where y represents the coordinate of image patch Q, and $Y^c(y)$ relates one color channel of the dehazed image Y.

To preserve the sharp structure of the dehazed result Y, the total variation loss L_{tv} is also utilized with an L1 regularization gradient prior, which is expressed by:

$$
L_{tv} = \frac{1}{N} \sum_{i=1}^{N} (\| \nabla_h Y \|_1 + \| \nabla_v Y \|_1)
\tag{18}
$$

where ∇_h represents the horizontal differential operation matric, ∇_v stands for the vertical differential operation matric.

Experiments

Datasets and implementation details

We choose the Unannotated Real Hazy Images (URHI) subset from the Realistic Single Image Dehazing (RESIDE) dataset [17] as the real-world hazy images in training the student. URHI subset consists of 4807 real-world hazy images without ground truth. All images are crawled from the Internet with different haze density, covering complex scenarios and including various degradation factors, such as poor light condition. Furthermore, 3577 unpaired clear images in RefinedDet [50] are adopted as the haze-free domain in training the unsupervised GAN-based fusion network. RTTS subset is adopted as the test dataset to evaluate the dehazing performance of different dehazing methods. RTTS consists of 4322 unpaired real-world hazy images, which are annotated with 41,203 bounding boxes and five object categories.

All parameters of the teacher networks are fixed during the training process. The GAN-based fusion network is first trained 100 epochs with a learning rate of 1e-4. The learning rate is linearly decayed to 0 in the following 100 epochs. The student network is trained 120 epochs with Adam optimizer and a batch size of 32. The initial learning rate 1e-4 decays every 45 epochs with a rate of 0.75 until 5e-5. The loss weights are set as $\lambda_{vgg} = 1$, $\lambda_{pix} = 10$, $\lambda_{per} = 1$, $\lambda_{dcp} = 1$ and $\lambda_{tv} = 0.0001$. We carry out experiments with the PyTorch framework on two GEFORCE RTX 3090.

Comparison with state-of-the-art methods

We compared the proposed method with eight state-of-the-art dehazing methods quantitatively and qualitatively. Among these methods, the traditional method DCP [7] proposed the famous dark channel prior to recover the haze-free images. the supervised method [13] made great dehazing performance for the well-designed network structure. The semi-supervised methods are SSID [9], DAD [8], RIDCP [22] and ENID [23]. The zero-shot dehazing method YOLY [42] only needs one hazy image in the training stage. The unsupervised dehazing method D4 [19] tried to improve the dehazing performance with the generative adversarial network. Quantitative and qualitative experiments are conducted on the real hazy images, including visual quality, no-reference image quality evaluation, task-driven evaluation, and model complexity.

Visual Quality. We perform visual quality comparison on the widely used RTTS dataset, which crawled 4,322 real-world hazy images from the Internet, covering diverse haze densities with complex degradations in outdoor environments. The dehazed results of the prior-based method DCP, the supervised method MSBDN; the semi-supervised methods SSID,

DAD, RIDCP and ENID, the unsupervised methods YOLY, D4, and the proposed method are shown in Fig. 4. It can be observed that the classical prior-based method DCP suffers from obvious color distortion, such as the purple sky in the first example. The main reason is that the dark channel prior is not suitable in the sky area. Similarly, the YOLY also suffers from the color distortion issue. Although MSBDN, SSID, DAD, and D4 can deal with color distortion well, the distant regions in the dehazed images seem a little different from the original hazy images due to the dense fog, for example, the distant building in the first example. Recently, RIDCP and ENID improve the generalization on real-world hazy images, but there are still obvious haze residues in the dehazed results, such as the building in the first example and the trees in the second example. Especially, all the former-mentioned approaches cannot deal well with the sky area with dense haze, the sky regions in the recovered images are still hazy. In comparison, the proposed method can generate visually pleasing and clean results in terms of less haze residue, sharper edges, and brighter details.

Furthermore, most existing approaches cannot overcome dim artifacts in the dehazing process because of the duality between image dehazing and low-light image enhancement. The detailed information in the near region is lost in the dehazed results, such as the roads in Fig. 5. Although RIDCP and ENID can alleviate this problem to some extent, they cannot recover the details in low light regions, such as the trees in the second example of Fig. 5. Compared to existing methods, the proposed dehazing algorithm can generate the best perceptual haze-free images with high-quality details.

No-Reference Image Quality Assessment. Since there are no ground truths for the real world hazy images, the widely used fog-aware density assessor (FADE) index [54] is adopted as the metric of non-reference image quality assessment in the experiment. The lower value of the FADE index indicates better dehazing performance. Quantitative comparisons of dehazed results on the RTTS dataset are illustrated in Table 1. It can be seen that the proposed method achieves the best performance 0.590 in terms of the FADE index. Compared to the original value of 2.484, the proposed method achieves the gains with 73.8%. Especially, it outperforms the two teachers RIDCP and ENID approximately 37.5% and 9.2%, respectively.

Task-Driven Evaluation. Hazy images always suffer from blurred edges, reduced contrast, and limited visibility, which may severely deteriorate the performance of high-level vision tasks, such as object detection. Single image dehazing is often utilized as a pre-processing step to improve the detection accuracy. We compared the mean Average Precision (mAP) of different dehazing results with the YOLO [55] on the RTTS dataset, which are annotated with five classes and bounding boxes. All 4322 real-world hazy images are recovered before

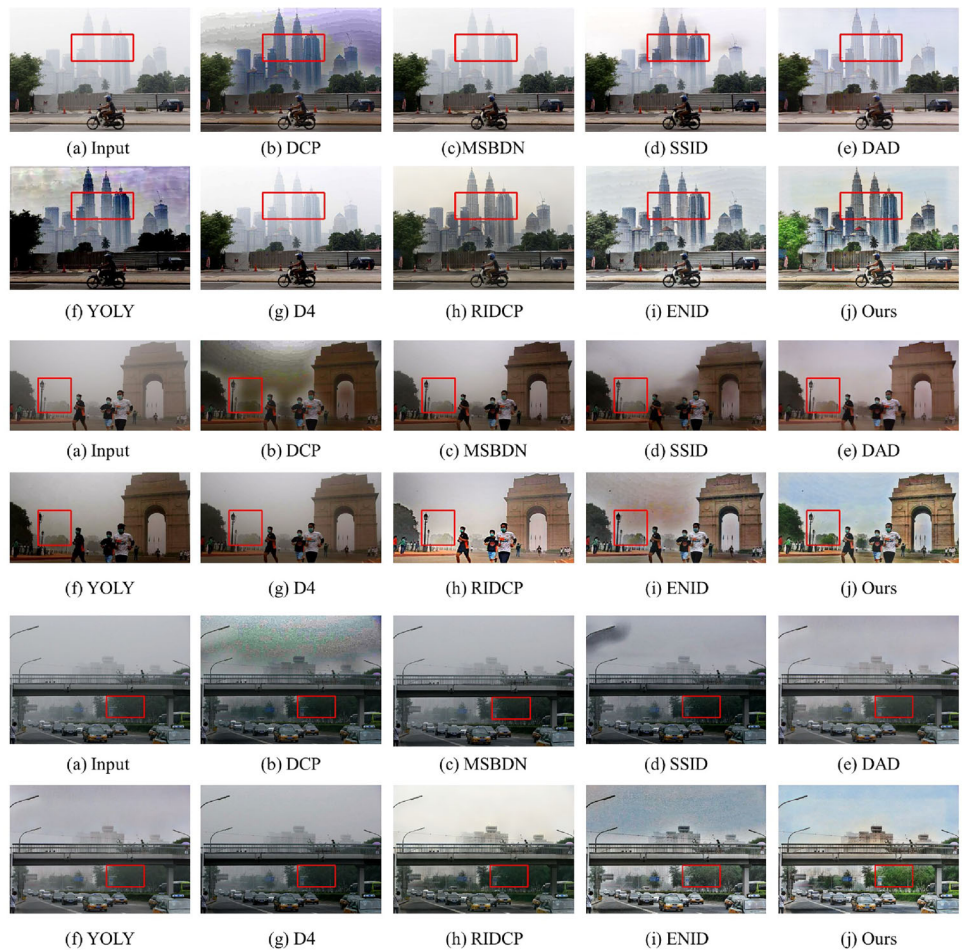**Fig. 4** Visual comparison on the real-world hazy images

(a) Input (b) DCP (c)MSBDN (d) SSID (e) DAD

(f) YOLY (g) D4 (h) RIDCP (i) ENID (j) Ours

(a) Input (b) DCP (c) MSBDN (d) SSID (e) DAD

(f) YOLY (g) D4 (h) RIDCP (i) ENID (j) Ours

(a) Input (b) DCP (c) MSBDN (d) SSID (e) DAD

(f) YOLY (g) D4 (h) RIDCP (i) ENID (j) Ours

Fig. 5 Visual comparison on the real-world hazy images in dim scenes

(a) Input (b) DCP (c) MSBDN (d) SSID (e) DAD

(f) YOLY (g) D4 (h) RIDCP (i) ENID (j) Ours

(a) Input (b) DCP (c) MSBDN (d) SSID (e) DAD

(f) YOLY (g) D4 (h) RIDCP (i) ENID (j) Ours
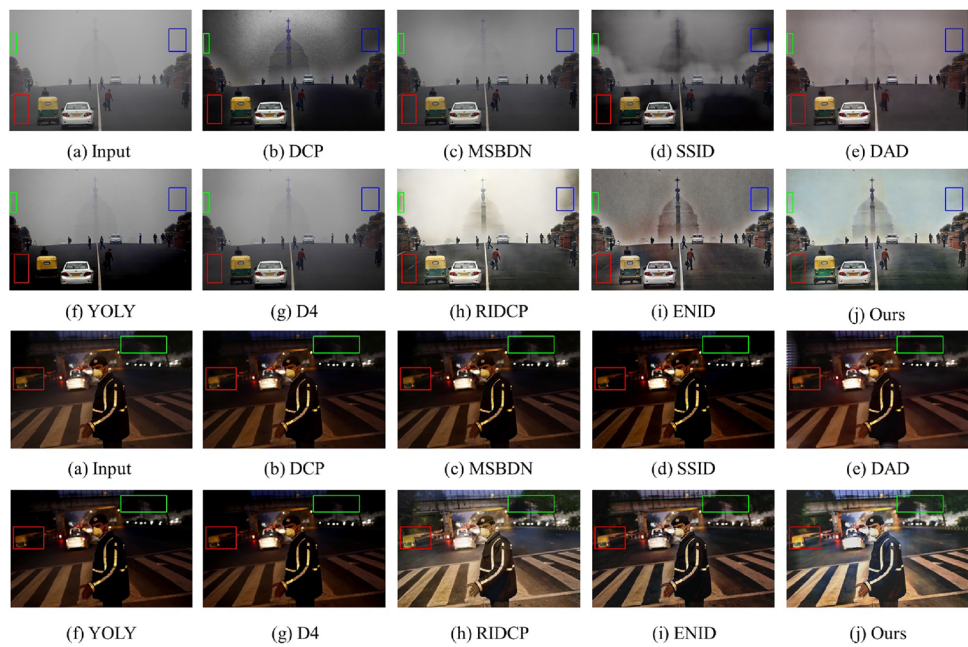
Table 1 Non-reference Evaluations on the RTTS Dataset

	Haze	MSBDN	SSID	DAD	YOLY	D4	RIDCP	ENID	Ours
FADE	2.484	1.343	0.867	1.130	1.035	1.358	0.944	0.650	**0.590**
Gain (%)	–	45.1	65.1	54.5	58.3	45.3	62.0	73.8	**76.3**

Table 2 Object Detection Result on the RTTS Dataset

	Haze	MSBDN	SSID	DAD	YOLY	D4	RIDCP	ENID	Ours
mAP (%)	63.32	65.16	56.98	65.02	55.12	64.59	64.67	**65.97**	65.23
Gain	–	+1.84	−6.34	+1.70	−8.10	+1.27	+1.35	**+2.65**	+1.91

Table 3 Parameters of different methods. Lower value indicates better results

	MSBDN	SSID	DAD	YOLY	D4	RIDCP	ENID	Ours
Params (M)	31.0	9.23	54.59	32.0	10.7	29.5	31.4	**6.5**

the detection task. As shown in Table 2, the proposed method comes second in the mAP index, and improves the value from 63.32 to 65.23, which further demonstrates that the proposed method can generate cleaner results and sharper structures.

The comparison on the parameters of the methods. In addition, it also outperforms the two complex teachers with a lightweight structure. As shown in Table 3, it can further reduce the parameters from 29.5 M and 31.4 M to 6.5 M. In total, the best dehazing performance and the least parameters demonstrate the superiority of the proposed method.

Ablation study

Effectiveness of different components. To evaluate the effectiveness of different components, a series of ablation experiments are conducted as follows. (1) W/O ENID: the teacher ENID is not adopted in the training process. (2) W/O RIDCP: the teacher RIDCP is not utilized in the training process. (3) W/O pre_en: the hazy images without enhancement are directly fed to the RIDCP dehazing network. (4) W / O GAN: The general adversarial network is not used to fuse the results of the two teachers. (5) W/O DCP: the dark channel loss is removed. (5) Ours. The dehazing results of algorithms with different components are shown in Fig. 6. It can be observed that the proposed method can generate the most visually pleasing result compared to the other methods, such as the tree area in the red box of the first example and the sky region in the green box of the second example. In Fig. 6b, c, dehazing methods without ENID or RIDCP may result in some haze residue in distant areas. In Fig. 6d, f, dehazing methods without the pre_en or DCP loss may cause lower contrast of the restored images. In particular, if adversarial training is removed in the second stage, the grass area and the sky region in the green box in Fig. 6e cannot be well recovered. Compared to other methods, the proposed method can

generate visually pleasing results, which demonstrates the effectiveness of different components.

Impact and selection of hyperparameters λ_{vgg} and λ_{dcp}. With the adversarial training strategy, an unsupervised fusion network used as generator is proposed to make the fused results similar to the unpaired clean images. In addition to adversarial loss, vgg loss with weight λ_{vgg} is introduced with the preliminary fused results Y_3. The value of the hyperparameters λ_{vgg} and the fused results are shown in Fig. 7. It can be observed that the outputs of the fusion network are almost the same as Y_3 when $\lambda_{vgg} >= 2$. The main reason is that large values of the weight λ_{vgg} tend to result in the model collapse in the adversarial training.

λ_{dcp} in existing algorithms is often set as a small value, for example, 10^{-5} in SSID [9], 10^{-3} in PSD [9]. In the proposed method, the λ_{dcp} is set to 1 in the training of the student network. We retrain the student network by varing hyperparameters λ_{dcp} while fixing the others. The loss curves with different λ_{dcp} are shown in Fig. 8. It can be observed that the dark channel loss can converge well when $\lambda_{dcp} = 1$.

Effectiveness of different teachers. To verify the effectiveness of the proposed framework, two other dehazing networks PSD [22] and FFANet [34] are selected as new teachers to replace RIDCP [22] and ENID [22]. The dehazing results of the algorithms with different teachers are shown in Fig. 9. It can be observed that the proposed method generates clean results in terms of less haze residue, for example, in the distant trees in the first example and the building area in the green box of the second example. In Fig. 9b, the dehazed framework with PSD and FFANet as teachers suffers from severe haze residues; the main reason is that FFANet trained with the synthetic dataset cannot generalize well in the real scenes due to the domain shift. Although the synthetic-to-realistic dehazing method PSD improves the performance on the real-world hazy images, it tends to cause color distortion in some regions of the dehazed results, such as the regions
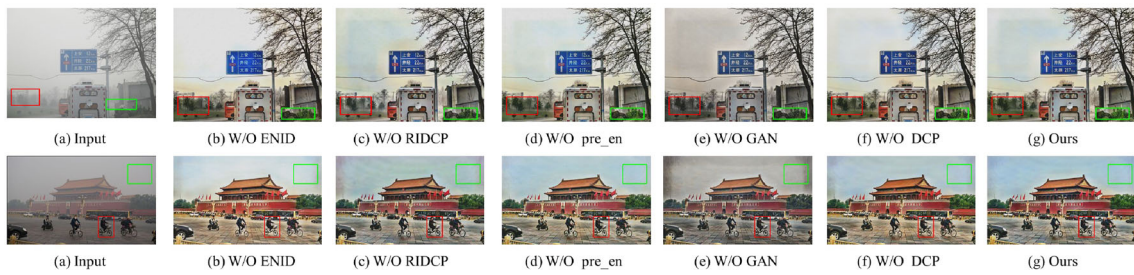
Fig. 6 Visual comparison on the real-world hazy images with different components

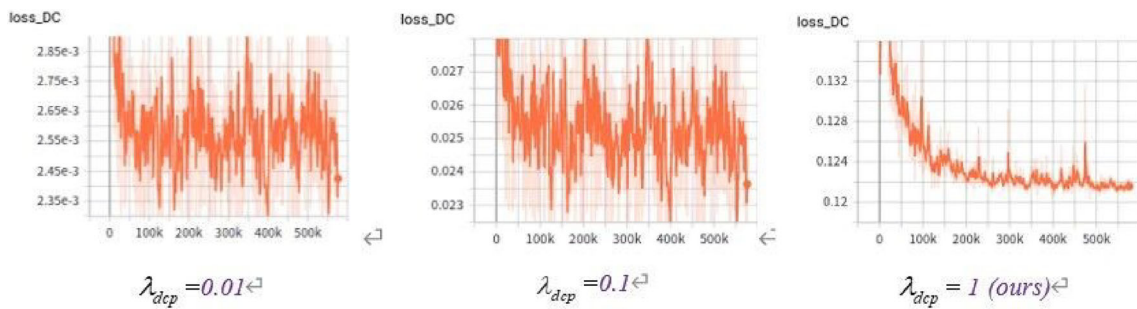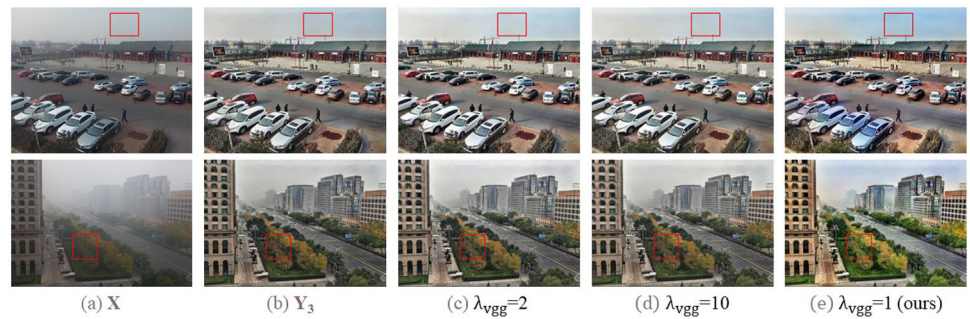**Fig. 7** Impact of the hyperparameter λ_{vgg}

Fig. 8 Selection of the hyperparameter λ_{dcp}

Fig. 9 Visual comparison with different teachers

in the red box of the second example. Compared to existing methods, the proposed dehazing algorithm can generate the best perceptual haze-free images with high-quality details.

Conclusion

In this paper, we propose a joint dual-teacher distillation and unsupervised fusion framework for unpaired real-world hazy images. An unsupervised fusion scheme for the dehazing results of two teachers is proposed with the generative adversarial network to assist the student in training. Specially, the unpaired clean images are further enhanced to overcome the dim artifacts that occurred in the dehazing task. We conduct comprehensive experiments to verify the superiority of the proposed method in terms of no-reference image quality assessment and the parameters. In the future, we will explore the unsupervised domain adaptation and the depth map to deal with dense haze removal, especially in the distant region.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of 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. The corresponding author states that there is no conflict of financial or non-financial interests. We would like to declare that the work described was original research that has not been published previously. It is not under consideration for publication elsewhere, in whole or in part. All the authors listed have approved the manuscript that is enclosed.

References

1. Li C, Zhou H, Liu Y, Yang C, Xie Y, Li Z, Zhu L (2023) Detection-friendly dehazing: object detection in real-world hazy scenes. IEEE Trans Pattern Anal Mach Intell 45:8284–8295

2. Liu W, Ren G, Yu R, Guo S, Zhu J, Zhang L (2021) Image-adaptive yolo for object detection in adverse weather conditions. In: AAAI Conference on Artificial Intelligence

3. Yang X, Mi MB, Yuan Y, Wang X, Tan RT (2022) Object detection in foggy scenes by embedding depth and reconstruction into domain adaptation. In: Asian Conference on Computer Vision

4. Lee S, Son T, Kwak S (2022) Fifo: Learning fog-invariant features for foggy scene segmentation. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 18889–18899

5. Yan L, Fan B, Liu H, Huo C, Xiang S, Pan C (2020) Triplet adversarial domain adaptation for pixel-level classification of vhr remote sensing images. IEEE Trans Geosci Remote Sens 58:3558–3573

6. Zhu Q, Mai J, Shao L (2014) Single image dehazing using color attenuation prior. In: British Machine Vision Conference. https://api.semanticscholar.org/CorpusID:264220633

7. He K, Sun J, Tang XJ (2011) Single image haze removal using dark channel prior. IEEE Trans Pattern Anal Mach Intell

8. Shao Y, Li L, Ren W, Gao C, Sang N (2020) Domain adaptation for image dehazing. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2805–2814

9. Li L, Dong Y, Ren W, Pan J-S, Gao C, Sang N, Yang M-H (2019) Semi-supervised image dehazing. IEEE Trans Image Process 29:2766–2779

10. Aggarwal AK (2020) Enhancement of gps position accuracy using machine vision and deep learning techniques. J Comput Sci 16:651–659

11. 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 25:5187–5198

12. Zhang H, Patel VM (2018) Densely connected pyramid dehazing network. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3194–3203

13. Dong H, Pan J-s, Xiang L, Hu Z, Zhang X, Wang F, Yang M-H (2020) Multi-scale boosted dehazing network with dense feature fusion. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2154–2164

14. Wu H, Qu Y, Lin S, Zhou JJ, Qiao R, Zhang Z, Xie Y, Ma L (2021) Contrastive learning for compact single image dehazing. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10546–10555

15. Zhang X, Wang T, Luo W, Huang P (2020) Multi-level fusion and attention-guided cnn for image dehazing. IEEE Trans Circ Syst Video Technol 31:4162–4173

16. Qu Y, Chen Y, Huang J, Xie Y (2019) Enhanced pix2pix dehazing network. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 8152–8160

17. Li B, Ren W, Fu D, Tao D, Feng D, Zeng W, Wang Z (2017) Benchmarking single-image dehazing and beyond. IEEE Trans Image Process 28:492–505

18. Peng Y-T, Lu Z, Cheng F-C, Zheng Y, Huang S-C (2020) Image haze removal using airlight white correction, local light filter, and aerial perspective prior. IEEE Trans Circ Syst Video Technol 30:1385–1395

19. Yang Y, Wang C, Liu R, Zhang L, Guo X, Tao D (2023) Self-augmented unpaired image dehazing via density and depth decomposition. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2027–2036

20. Dudhane A, Murala S (2019) Cdnet: Single image de-hazing using unpaired adversarial training. 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), 1147–1155

21. Jin Y, Gao G, Liu Q, Wang Y (2020) Unsupervised conditional disentangle network for image dehazing. 2020 IEEE International Conference on Image Processing (ICIP), 963–967

22. Wu R, Duan Z-P, Guo C, Chai Z, Li C (2023) Ridcp: Revitalizing real image dehazing via high-quality codebook priors. 2022

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 22282–2229

23. University HP (2022) Physics-based Constraint Ensemble Dehazing Algorithm With the Intermediant Domain (in Chinese), Cn115546049a edn. 454000 2001 Century Avenue, High-tech Zone, Jiaozuo City, Henan Province

24. Liu X, Li H, Zhu C (2022) Joint contrast enhancement and exposure fusion for real-world image dehazing. IEEE Trans Multimedia 24:3934–3946

25. Dai C, Lin M-x, Wu X, Zhang D (2020) Single hazy image restoration using robust atmospheric scattering model. Signal Process 166:2

26. Borkar K, Mukherjee S (2020) Single image dehazing by approximating and eliminating the additional airlight component. Neurocomputing 400:294–308

27. Kaur M, Singh D, Kumar V, Sun K (2020) Color image dehazing using gradient channel prior and guided l0 filter. Inf Sci 521:326–342

28. Liu J, Liu RW, Sun J, Zeng T (2021) Rank-one prior: Toward real-time scene recovery. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 14797–14805

29. Ju M, Ding C, Guo CA, Ren W, Tao D (2021) Idrlp: Image dehazing using region line prior. IEEE Trans Image Process 30:9043–9057

30. Maini DS, Aggarwal DAK (2018) Camera position estimation using 2d image dataset. https://api.semanticscholar.org/CorpusID:225091809

31. Amandeep Kaur APSC, Aggarwal AK (2022) Dynamic deep genomics sequence encoder for managed file transfer. IETE J Res 0(0):1–13. https://doi.org/10.1080/03772063.2022.2060869

32. Kaur A, Chauhan APS, Aggarwal AK (2019) Machine learning based comparative analysis of methods for enhancer prediction in genomic data. 2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT), 142–145

33. Li B, Peng X, Wang Z, Xu J, Feng D (2017) Aod-net: All-in-one dehazing network. 2017 IEEE International Conference on Computer Vision (ICCV), 4780–4788

34. Qin X, Wang Z, Bai Y, Xie X, Jia H (2019) Ffa-net: Feature fusion attention network for single image dehazing. arXiv:1911.07559

35. Sahu G, Seal A, Bhattacharjee D, Nasipuri M, Brida P, Krejcar O (2022) Trends and prospects of techniques for haze removal from degraded images: a survey. IEEE Trans Emerg Top Comput Intell 6:762–782

36. Singh M, Laxmi V, Faruki P (2022) Visibility enhancement and dehazing: research contribution challenges and direction. Comput Sci Rev 44:100473

37. Park J, Han DK, Ko H (2020) Fusion of heterogeneous adversarial networks for single image dehazing. IEEE Trans Image Process 29:4721–4732

38. An S, Huang X, Wang L, Wang L, Zheng Z (2021) Semi-supervised image dehazing network. Vis Comput 38:2041–2055

39. Chen Z, Wang Y, Yang Y, Liu D (2021) Psd: Principled synthetic-to-real dehazing guided by physical priors. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7176–7185

40. Zhu J-Y, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. 2017 IEEE International Conference on Computer Vision (ICCV), 2242–2251

41. Li B, Gou Y, Liu J, Zhu H, Zhou JT, Peng X (2020) Zero-shot image dehazing. IEEE Trans Image Process 29:8457–8466

42. Li B, Gou Y, Gu S, Liu J, Zhou JT, Peng X (2020) You only look yourself: Unsupervised and untrained single image dehazing neural network. Int J Comput Vis 129:1754–1767

43. Wang X, Yang X, Zhang S, Li Y, Feng L, Fang S, Lyu C, Chen K, Zhang W (2022) Consistent-teacher: Towards reducing inconsistent pseudo-targets in semi-supervised object detection. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3240–3249

44. Qiao Y, Cui J, Huang F, Liu H, Bao C, Li X (2021) Efficient style-corpus constrained learning for photorealistic style transfer. IEEE Trans Image Process 30:3154–3166

45. Chen Z, Zhu L, Wan L, Wang S, Feng W, Heng P-A (2020) A multi-task mean teacher for semi-supervised shadow detection. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5610–5619

46. Tang Y, Chen W, Luo Y, Zhang Y (2021) Humble teachers teach better students for semi-supervised object detection. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3131–3140

47. Cao S, Joshi D, Gui L, Wang Y-X (2023) Contrastive mean teacher for domain adaptive object detectors. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 23839–23848

48. Deng J, Xu D, Li W, Duan L (2023) Harmonious teacher for cross-domain object detection. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 23829–23838

49. Hong M, Xie Y, Li C, Qu Y (2020) Distilling image dehazing with heterogeneous task imitation. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3459–3468

50. Zhao S, Zhang L, Shen Y, Zhou Y (2021) Refinednet: a weakly supervised refinement framework for single image dehazing. IEEE Trans Image Process 30:3391–3404

51. Jiang Y, Gong X, Liu D, Cheng Y, Fang C, Shen X, Yang J, Zhou P, Wang Z (2019) Enlightengan: Deep light enhancement without paired supervision. IEEE Trans Image Process 30:2340–2349

52. Zhang H, Cissé M, Dauphin Y, Lopez-Paz D (2017) mixup: Beyond empirical risk minimization. arXiv:1710.09412

53. Hou Q, Zhou D, Feng J (2021) Coordinate attention for efficient mobile network design. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 13708–13717

54. Choi LK, You J, Bovik AC (2015) Referenceless prediction of perceptual fog density and perceptual image defogging. IEEE Trans Image Process 24:3888–3901

55. Redmon J, Farhadi A (2018) Yolov3: An incremental improvement. arXiv:1804.02767