Joint Raindrop and Haze Removal from a Single Image

Yina Guo, *Member, IEEE,* Jianguo Chen, Xiaowen Ren, Anhong Wang and Wenwu Wang, *Senior Member, IEEE*

***Abstract*—In a recent study, it was shown that, with adversarial training of an attentive generative network, it is possible to convert a raindrop degraded image into a relatively clean one. However, in real world, raindrop appearance is not only formed by individual raindrops, but also by the distant raindrops accumulation and the atmospheric veiling, namely haze. Current methods are limited in extracting accurate features from a raindrop degraded image with background scene, the blurred raindrop regions, and the haze. In this paper, we propose a new model for an image corrupted by the raindrops and the haze, and introduce an integrated multi-task algorithm to address the joint raindrop and haze removal (JRHR) problem by combining an improved estimate of the atmospheric light, a modified transmission map, a generative adversarial network (GAN) and an optimized visual attention network. The proposed algorithm can extract more accurate features for both sky and non-sky regions. Experimental evaluation has been conducted to show that the proposed algorithm significantly outperforms state-of-the-art algorithms on both synthetic and real-world images in terms of both qualitative and quantitative measures.**

***Index Terms*—Raindrop removal, haze removal, generative adversarial network, visual attention.**

I. Introduction

RESTORING a windshield or lens image corrupted by raindrops is beneficial to various computer vision applications, such as autonomous driving [1] and video surveillance [2], [3]. Unlike the removal of rain streaks, the shape of the raindrops is similar to a fish-eye lens, which leads to the raindrop regions being formed by light reflected from a wider environment [4]. Thus the images degraded by raindrops have three types of visibility degradation caused by individual raindrops, distant raindrop accumulation and the atmospheric veiling which are visually similar to the haze, and the blurred appearance of the raindrop regions due to the focus of the camera on the background scene, respectively.

Many studies have been conducted to address the problem of raindrop removal from a single image, of which the two main approaches are model-based raindrop removal [5], [6], [7], [8] and deep learning based raindrop removal [9], [10], [1] respectively. The latter approach has received increasing interest recently and is the focus of this paper.

A raindrop degraded image is often modeled by the additive combination of background images and the effect of the raindrops, such as the recent work in [1]. However, a raindrop degraded image not only contains a background image and the effects of the raindrops, but also includes the haze effects. In addition, an image degraded by the raindrops is often accompanied by the blurred raindrop regions caused by the autofocus of the cameras. Therefore, enhancing an image corrupted by raindrops would require the removal of the haze effect, along with the removal of raindrops.

Existing methods, however, are designed either only for raindrop removal, such as the model-based approaches [5], [6], [7], and the deep learning based approaches [9], [10], [1], or only for haze removal, such as the model based approaches [11], [12], [13], and the deep learning based approaches [14], [15].

These methods can achieve relatively good performance in removing the targeted type of distortion (i.e. raindrop or haze) from a single image, but are ineffective in removing both types of distortions. To our knowledge, there is no existing study for removing the blurred raindrop regions and the haze effect simultaneously.

The aim of this paper is to convert an image corrupted by raindrops and haze into a clean one by removing them simultaneously. A potential approach to this problem is to cascade a raindrop removal method with a haze removal method, which, however, may be limited by the following challenges. For example, blurring artifacts are often introduced at the edges of the processed image with a typical haze removal (or raindrop removal) algorithm, which may lead to inaccurate estimation of the parameters of the model if the raindrop removal (or the haze removal) step is followed in the cascaded setting. In addition, existing haze removal methods, e.g. [11], [14], [15] are ineffective in removing the dense haze effects of an image corrupted by dense haze and raindrops. In order to address these technical challenges, a joint raindrop and haze removal (JRHR) problem is considered and our contribution is two-fold:

1) JRHR model: A new model of the JRHR problem is proposed in order to recover an image corrupted by raindrops and haze by detecting and removing the effects of the raindrops and the haze simultaneously.
2) JRHR algorithm: Based on the JRHR model, an inte-

This work was supported by National Natural Science Foundation of China under Grant 61301250, Key Research and Development Project of Shanxi Province under Grant 201803D421035, Natural Science Foundation for Young Scientists of Shanxi Province under Grant 201901D211313, Research Project Supported by Shanxi Scholarship Council of China under Grant HGKY2019080.

Yina Guo, Jianguo Chen, Xiaowen Ren and Anhong Wang are with the School of Electronic Information Engineering, Taiyuan University of Science and Technology, Taiyuan 030024, China. Yina Guo and Jianguo Chen contributed equally to this work (e-mail: zulibest@163.com, chenjg198@outlook.com).

Wenwu Wang is with the Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, Surrey GU2 7XH, U.K. (Corresponding to: w.wang@surrey.ac.uk).

grated multi-task algorithm is proposed by combining an improved estimate of the atmospheric light, a modified transmission map, a generative adversarial network (GAN) and an optimized visual attention network.

a) Firstly, an improved estimate of atmospheric light is presented by considering the medium brightness case, to mitigate certain artifacts, such as blocking effects, halo and gradient reversal artifacts, and to produce a smooth transmission map.
b) Secondly, with the estimated value of the atmospheric light, the transmission map is re-derived for the sky and non-sky regions, respectively, to facilitate the removal of haze at different levels.
c) Thirdly, an attentive GAN is presented by combining a GAN network with an optimized visual attention network to recover the background image from an image corrupted by raindrops and haze. We present a new loss function for the optimized visual attention network where a penalty term is introduced to improve the clarity of the raindrop regions in the attention maps, to improve its generalization performance by preventing it from overfitting, and also to relax the value range of the network parameters in order to reduce potential biases in their estimates.

The paper is organized as follows. Section II describes the related work. Section III formulates a new mathematical model for the JRHR problem. Section IV presents our proposed algorithm for the problem of JRHR. Section V shows numerical results. Section VI concludes the paper and draws potential future research directions.

II. Related Work

In the field of computer vision, there is an increasing interest in the problem of raindrop removal over the past decades [16], [17], [18], [19], [3], [20], [21]. Unlike the image recovery of the rain streaks, there are relatively few papers in recovering the raindrop degraded image [22], [23], [24], [25], [9], [26], [27], [1]. According to the required input amount of the images, the raindrop removal methods can be mainly divided into multi-image (or video) based methods and single image based methods. Multi-image (or video) based methods are mainly used for dynamic scenes, which include other moving objects apart from the raindrops coupled with possible movement of lens. For the video sequences with small amount of raindrops, the corrupted image can be enhanced by directly averaging the video frames, if the effect of the raindrops on the pixel is only in a few frames. Single image based methods are mainly used for static scenes, where no lens and other clear movement cases are involved.

Multi-image (or video) based methods: Kurihata et al. [22] proposed a raindrop detecting method by using video sequences. The shape of raindrops is learned by using principal component analysis (PCA). However, the number of raindrops that needs to be learned cannot be determined automatically for transparent raindrops with various shapes. Roser and Geiger [23] proposed a raindrop shape model based on cubic Bezier curves and a method to compare a synthetic raindrop with a raindrop patch. The raindrops are assumed to be a sphere section or an inclined sphere section. Later Roser et al. [24] presented a novel raindrop shape model for the detection of view-disturbing, adherent raindrops on inclined surfaces. The synthetic raindrop is assumed to be an oblique spherical section. Wu et al. [25] presented a machine learning based approach to detect and remove raindrops on windshield by analyzing the color, texture and shape characteristics of raindrops in images. The raindrops are assumed to be circular in each image frame under light and moderate rainy conditions. However, these assumptions cannot handle the situation for covering the windshield completely. Webster and Breckon [26] proposed two novel extensions for raindrop detection in video imagery: the use of additional shape priors in the classification model and the incorporation of scene context for all features used in the secondary stage of raindrop verification. You et al. [27] introduced a motion based method for detecting and removing raindrops in video, based on the observation that the motion of raindrop pixels is slower than that of non-raindrop pixels, and the temporal change of intensity of raindrop pixels is smaller than that of non-raindrop pixels. These methods can remove raindrops in multiple images, whereas they cannot be applied directly to a single image.

Single image based methods: Eigen et al. [9] presented a post-capture image processing solution that can remove localized raindrop and dirt artifacts from a single image. The key idea is to collect a dataset of clean/corrupted image pairs to train a convolutional neural network. The method works for relatively sparse and small droplets as well as dirt but is not effective for large and dense raindrops, since it assumes that the raindrops are separate and opaque small regions. Qian et al. [1] proposed a single-image based raindrop removal method by using a GAN with an attention map. The novelty is to insert an attention map into both generative network and discriminative network. This method focuses on the raindrop regions of a raindrop degraded image, but does not consider the haze effects caused by the distant raindrop accumulation and the atmospheric veiling.

In this paper, we build a new model of an image corrupted by the raindrops and the haze in view of the mixed effects of the background scene, the blurred raindrop regions, and the haze. An integrated multi-task algorithm by combining an improved haze removal method, a GAN network and an optimized visual attention network is used to detect and remove the raindrops and the haze.

III. Mathematical Model

The aim here is to recover an image corrupted by raindrops and haze by detecting and removing the effects of the raindrops and the haze. In this section, a mathematical model is presented for the JRHR problem.

Recently, a generalized rain model that depicts rain location and rain intensity separately [3] is expressed as follows:

$$\mathbf{O} = \mathbf{B} + \mathbf{S} \circ \mathbf{R}, \tag{1}$$

where $\mathbf{O} \in \mathbb{N}^{N \times M}$ is the input image corrupted by rain, $\mathbf{B} \in \mathbb{N}^{N \times M}$ is the background layer, $\mathbf{S} \in \mathbb{N}^{N \times M}$ is the

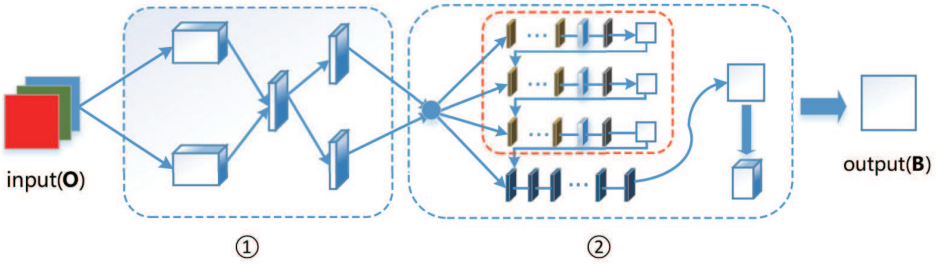

Fig. 1. The architecture of our multi-task joint raindrop and haze removal (JRHR) algorithm, including the determination of parameters ① and the joint haze and raindrop removal network ②.

rain layer, and a region-dependent variable $\mathbf{R} \in \mathbb{N}^{N\times M}$ to indicate the locations of individually visible rain, where $\circ$ means element-wise multiplication. Here, elements in $\mathbf{R}$ are binary values, where 1 indicates rain regions and 0 indicates non-rain regions. The model allows to detect rain regions first, and then to operate differently on rain and non-rain regions, preserving background details. However, in (1), $\mathbf{R}$ only considers the locations of rain regions and non-rain regions, without considering the haze effects [11], [28].

To overcome the drawback, a new model of the JRHR problem is proposed, where we aim to recover the background layer $\mathbf{B}$ from an image $\mathbf{O}$ corrupted by raindrops and haze.

$$\mathbf{O} = (\mathbf{B} + (\mathbf{I} - \mathbf{L}) \circ \mathbf{R}) \circ \mathbf{t} + A(\mathbf{I} - \mathbf{t}), \tag{2}$$

where $\mathbf{O} \in \mathbb{N}^{N\times M}$ is the input image corrupted by raindrops and haze, $\mathbf{B} \in \mathbb{N}^{N\times M}$ is the background layer, $A \in \mathbb{R}$ indicates the global atmospheric light, and $\mathbf{t} \in \mathbb{N}^{N\times M}$ denotes the transmission map. $\mathbf{R} = \sum_{i=1}^{r} \widetilde{\mathbf{R}}_i$ is the rain layer, where $\mathbf{R} \in \mathbb{N}^{N\times M}$ and each $\widetilde{\mathbf{R}}_i \in \mathbb{N}^{N\times M}$ is a layer of raindrops, i is the index of the raindrop layers, and r is the maximum number of raindrop layers. $\mathbf{I} \in \mathbb{N}^{N\times M}$ is an unit matrix (all-ones matrix), $(\mathbf{I} - \mathbf{L})$ indicates the locations of individually visible raindrops, and $\circ$ denotes the Hadamard product. Here, elements in $\mathbf{L}$ are binary values, where 0 indicates raindrop regions and 1 indicates non-raindrop regions.

In model (2), our goal is to recover the background layer $\mathbf{B}$ from an input image $\mathbf{O}$. Thus $\mathbf{B}$ can be expressed as

$$\mathbf{B} = (\mathbf{O} - A(\mathbf{I} - \mathbf{t})) \oslash \mathbf{t} - (\mathbf{I} - \mathbf{L}) \circ \sum_{i=1}^{r} \widetilde{\mathbf{R}}_i, \tag{3}$$

where $\oslash$ denotes the Hadamard division.

In real life, the raindrops are transparent and the haze is semi-transparent, and the camera is usually focused on the background scene. Moreover, the shape of the raindrops is similar to a fish-eye lens, and therefore the raindrop regions of the images are formed by light reflected from a wider environment. As a result, the imagery inside a raindrop region is mostly blurred, and transparent parts of the raindrop regions contain some information about the background. Based on (2), we can generate synthetic images that resemble natural images better than those generated by (1). Thus, we can use these images to train our network, and perform raindrop removal and haze removal, which provides convenience for model training.

IV. Joint Raindrop and Haze Removal Algorithm

In this section, we present an integrated multi-task algorithm in a two-step solution where joint raindrop and haze removal (JRHR) is performed to solve the problem in (3), as shown in Fig. 1. The first step is to determine the parameters of the global atmospheric light A and the transmission map $\mathbf{t}$. The second step is to recover the background image $\mathbf{B}$ from the degraded image $\mathbf{O}$.

According to (3), given the input image $\mathbf{O}$, our goal is to estimate the background layer $\mathbf{B}$. The JRHR problem can be described by

$$\arg\min_{\mathbf{B},\mathbf{R},\mathbf{L}} \|(\mathbf{O} - A(\mathbf{I} - \mathbf{t})) \oslash \mathbf{t} - \mathbf{B} - (\mathbf{I} - \mathbf{L}) \circ \mathbf{R}\|_2^2, \tag{4}$$

where A is the global atmospheric light parameter, $\mathbf{t}$ is the transmission map, $\mathbf{R}$ denotes the raindrop layer, $(\mathbf{I} - \mathbf{L})$ indicates the locations of individually visible raindrops, where $\mathbf{I}$ denotes an unit matrix (all-ones matrix), $\mathbf{L}$ denotes the binary values, $\circ$ denotes the Hadamard product, and $\oslash$ denotes the Hadamard division. Here, the elements in $\mathbf{L}$ are in binary values, where 0 indicates raindrop regions and 1 indicates non-raindrop regions. To reduce algorithmic complexity and training time, we fix the parameters A and $\mathbf{t}$ by estimating them directly from the input image, but learn the parameters $\mathbf{L}$ and $\mathbf{R}$ via a learning algorithm using some training data, as detailed in our experiments.

A. Determination of parameters

In real life, pictures are often taken in natural light or lamplight. For an image, the region with bright illuminations is called sky region, and the region with low illuminations is called non-sky region. Even in low-light or blowing sand environments, such as underground or the driving place of the mine, an image has both sky and non-sky regions.

As shown in Fig. 1 and Fig. 2, for the purpose of estimating the background layer $\mathbf{B}$, we need to find the global atmospheric light $A \in \mathbb{R}$ and the transmission map $\mathbf{t} = \{\mathbf{t}_1, \mathbf{t}_2\} \in \mathbb{N}^{N\times M}$ according to the sky region and non-sky region, respectively.

The method in [11] is based only on non-sky region. However, even in low-light environments, the transmission map estimated by the non-sky region is not smooth, but containing blocking artifacts. Different from [11], the methods in [12] and [13] take account of the non-sky region and the sky region together, which are effective in reducing halo and gradient reversal artifacts. However, the atmospheric light

A of these methods is determined only with the maximum values of the light channel image and the minimum values of the dark channel image, and the transmission map $\mathbf{t}$ is not evaluated according to sky region and non-sky region, respectively. Therefore, the performance of these methods is limited with different thickness of haze. In order to solve the above problems, we take the medium brightness case in full consideration, as well as present an improved atmospheric light and the corresponding transmission map.

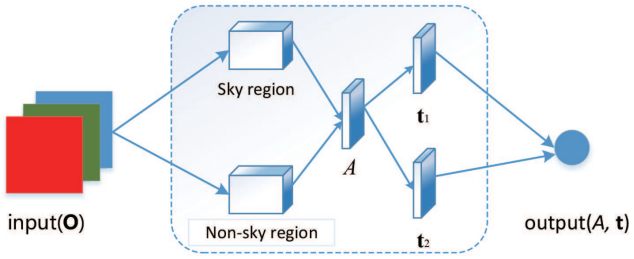

Fig. 2. The architecture of the determination of parameters. The global atmospheric light A and the transmission map $\mathbf{t} = \{\mathbf{t}_1, \mathbf{t}_2\}$ need to be determined according to sky and non-sky regions.

(i) Determination of the atmospheric light A

Considering the medium brightness case, an improved atmospheric light is presented as follows:

$$A = pL_{med} + (1-p)D_{dmax}. \tag{5}$$

Here, $p = \frac{k}{K}$, K and k are the number of all pixels and the number of light pixels within the image, respectively. $L_{med} = \text{median}(O^{light}(x))$ is present to denote the median of the light channel image constructed by several $O^{light}(x)$ with the change of the pixel x. $\Omega(x)$ denotes a patch centered at the pixel x. c represents one of R, G, B channels and $\mathbf{O}^c$ means a c channel in $\Omega(x)$ of the input image $\mathbf{O}$. $O^{light}(x) = \max\limits_{y\in\Omega(x)}[\max\limits_{c\in\{R,G,B\}}(O^c(y))]$ represents a light channel in $\Omega(x)$ that contains the maximum R, G, B values of each pixel, namely sky region. $D_{dmax} = \max(O^{dark}(x))$ represents the maximum of the dark channel image constructed by several $O^{dark}(x)$ with the change of the pixel x. $O^{dark}(x) = \min\limits_{y\in\Omega(x)}[\min\limits_{c\in\{R,G,B\}}(O^c(y))]$ represents a dark channel that contains the minimum R, G, B values of each pixel, namely non-sky region.

The new method for estimating the atmospheric light has an advantage in mitigating certain artifacts, such as blocking effects and halo and gradient reversal artifacts, and thus resulting in more smooth estimation of the transmission map, as compared with the methods in [11], [12] and [13].

(ii) Determination of the transmission map $\mathbf{t}$

According to (2), the model of the JRHR problem can be transformed into (6).

$$\frac{\mathbf{O}}{A} = \frac{\mathbf{B} + (\mathbf{I}-\mathbf{L})\circ\mathbf{R}}{A} \circ \mathbf{t} + \mathbf{I} - \mathbf{t}. \tag{6}$$

Based on the improved estimate of the atmospheric light A, and the transformation model in (6), the transmission map $\mathbf{t}$ can be re-derived for the sky and non-sky regions, respectively.

For non-sky region ($O(x) < A$), the two minimum filtering operations are performed on both sides of (6).

$$\begin{aligned}&\min_{y\in\Omega(x)}[\min_{c\in\{R,G,B\}}\frac{O^c(y)}{A}]\\ &= \min_{y\in\Omega(x)}[\min_{c\in\{R,G,B\}}\frac{B^c(y)+(1-L(y))\circ R^c(y)}{A}]t(x) + 1 - t(x),\end{aligned} \tag{7}$$

where $O(x)$ means the maximum R, G, B value of the pixel x, $t(x)$ means the transmission map of the pixel x and A is the atmospheric light. $B^c(y)$ represents a color channel in $\Omega(x)$ of the background layer, $R^c(y)$ represents a color channel in $\Omega(x)$ of the rain layer, and $L(y)$ is a binary value in $\Omega(x)$ which indicates the location of the raindrop.

When $\min\limits_{y\in\Omega(x)}[\min\limits_{c\in\{R,G,B\}} B^c(y)+(1-L(y))\circ R^c(y)]$ is close to 0, the transmission map $t(x)$ is expressed as

$$t(x) = 1 - \omega\frac{\min\limits_{y\in\Omega(x)}[\min\limits_{c\in\{R,G,B\}} O^c(y)]}{A}, \tag{8}$$

where ω is a constant parameter with a value between [0,1] to make the image looks more natural.

For sky region ($O(x) \geq A$), two maximum filtering operations are performed on both sides of (6).

$$\begin{aligned}&\max_{y\in\Omega(x)}[\max_{c\in\{R,G,B\}}\frac{O^c(y)}{A}]\\ &= \max_{y\in\Omega(x)}[\max_{c\in\{R,G,B\}}\frac{B^c(y)+(1-L(y))\circ R^c(y)}{A}]t(x) + 1 - t(x).\end{aligned} \tag{9}$$

When $\max\limits_{y\in\Omega(x)}[\max\limits_{c\in\{R,G,B\}} B^c(y)+(1-M(y))\circ R^c(y)]$ is close to 1, the transmission map $t(x)$ is expressed as

$$t(x) = 1 - \omega\frac{1 - \max\limits_{y\in\Omega(x)}[\max\limits_{c\in\{R,G,B\}} O^c(y)]}{1-A}. \tag{10}$$

So, a modified transmission map $\mathbf{t}$ is presented as follows:

$$t(x) = \begin{cases} 1 - \omega\frac{\min\limits_{y\in\Omega(x)}[\min\limits_{c\in\{R,G,B\}} O^c(y)]}{A}, & O(x) < A \\ 1 - \omega\frac{1-\max\limits_{y\in\Omega(x)}[\max\limits_{c\in\{R,G,B\}} O^c(y)]}{1-A}. & O(x) \geq A \end{cases} \tag{11}$$

The new transmission map may provide significant benefits in two aspects. First, it defines the transmission map according to the ranges of the new atmospheric light more clearly. Second, it offers better chances in removing haze at different levels, as compared with [11], [12] and [13].

B. Recovery of the background image $\mathbf{B}$

In order to reconstruct $\mathbf{B}$, the maximum posteriori estimation is considered as

$$\arg\min_{\mathbf{B},\mathbf{R},\mathbf{L}} \|(\mathbf{O} - A(\mathbf{I}-\mathbf{t})) \oslash \mathbf{t} - \mathbf{B} - (\mathbf{I}-\mathbf{L})\circ\mathbf{R}\|_2^2. \tag{12}$$

The global atmospheric light A and the transmission map $\mathbf{t}$ obtained from (5) and (11) are combined in (12) to remove the haze effects of the degraded image $\mathbf{O}$. An attentive GAN is

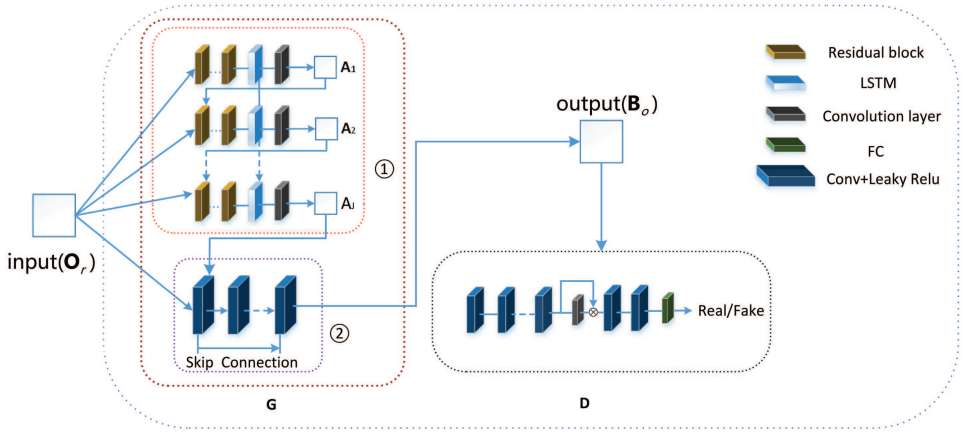

Fig. 3. The architecture of the joint haze and raindrop removal network. **G** represents the generative network which includes an optimized visual attention network ① and an autoencoder network ②. **D** represents the discriminative network. $\mathbf{A}_1, \mathbf{A}_2$, and $\mathbf{A}_J$ are the initial attention map, the second attention map and the J^{th} attention map produced by ①, respectively.

established to detect and remove the raindrops of the degraded image **O** by combining a GAN network and an optimized visual attention network.

1) Generative network: As shown in Fig. 3, the generative network consists of two sub-networks: an optimized visual attention network and an autoencoder network. Pairs of images with raindrops and without raindrops in the same background scene are used to train the generative network.

(i) Optimized visual attention network

The purpose of the optimized visual attention network is to find the raindrop regions of the degraded image **O**, which needs to get attention from the autoencoder network.

Each recurrent block at each iteration comprises of five layers of ResNet for extracting features from the input image and the previous block, as well as a convolutional long short-term memory (Cov-LSTM) unit with the convolutional layers for generating the attention maps [1].

In [1], the visual attention network can help to find raindrop regions of the input image that need to be attended. However, it may have potential negative effects in two aspects. First, in haze removal, typically, blurring artifacts may be introduced at the edges of the processed image which may lead to inaccurate estimation of the parameters of the model if the raindrop removal step is followed in the joint haze and raindrop removal setting. Second, it is prone to over-fitting with less data, and may limit the value range of the network parameters and introduce potential biases in their estimates.

To address these issues, a penalty term is introduced to the loss function of each recurrent block as follows:

$$\mathcal{L}_{ATT}(\{\mathbf{A}_{att}\}, \mathbf{L}) = \sum_{j=1}^{J} [\theta^{J-j} \mathcal{L}_{MSE}(\mathbf{A}_j, \mathbf{L}) + \lambda \|\mathbf{A}_j\|_2^2], \tag{13}$$

where j is the time step and $\mathbf{L}$ is defined in (2), $\mathbf{A}_j = ATT_j(\mathbf{F}_{j-1}, \mathbf{H}_{j-1}, \mathbf{C}_{j-1})$ represents the output attention map produced by the optimized visual attention network at time step j. The values of $\mathbf{A}_j$ become larger with the increase of iterations until the J^{th} iteration, which indicates the increase in confidence. ATT_j represents the optimized visual attention network at j. λ is a constant and set to 0.001. $\mathbf{F}_{j-1}$ is the concatenation of the input image and the attention map from the previous iteration. θ is a calibration factor. $\mathbf{C}_j = \mathbf{f}_j \circ \mathbf{C}_{j-1} + \mathbf{i}_j \circ \tanh(\mathbf{W}_{xc} * \mathbf{X}_j + \mathbf{W}_{hc} * \mathbf{H}_{j-1} + \mathbf{b}_c)$ encodes the cell state for the next LSTM unit. $\mathbf{H}_j = \mathbf{o}_j \circ \tanh(\mathbf{C}_j)$ describes the output features of the LSTM unit. Here $\mathbf{i}_j = \sigma[(\mathbf{W}_{xi} * \mathbf{X}_j + \mathbf{W}_{hi} * \mathbf{H}_{j-1} + \mathbf{b}_i)], \mathbf{f}_j = \sigma[(\mathbf{W}_{xf} * \mathbf{X}_j + \mathbf{W}_{hf} * \mathbf{H}_{j-1} + \mathbf{b}_f)], \mathbf{o}_j = \sigma[(\mathbf{W}_{xo} * \mathbf{X}_j + \mathbf{W}_{ho} * \mathbf{H}_{j-1} + \mathbf{b}_o)]$ are an input gate, a forget gate and an output gate of the convolutional LSTM unit, respectively. σ is the activation function of sigmoid. Operator $*$ and $\circ$ are used for the convolution and Hadamard product. $\mathbf{W}_x, \mathbf{W}_h$ and $\mathbf{b}$ are the weights and the biases of the linear relationship.

The new loss function may provide significant benefits in three aspects. First, it improves the clarity of the raindrop regions in the attention maps. Second, the generalization performance is improved by preventing it from over-fitting. Third, it relaxes the value range of the network parameters and reduces the potential biases in the estimates of the parameters.

(ii) Autoencoder network

The autoencoder network is used here to generate an image without raindrops. The input of the autoencoder network is the concatenation between the input image and the J^{th} attention map from the optimized visual attention network.

The architecture of the autoencoder network is shown in Fig. 3, which has sixteen conv-leakyrelu blocks and skip connections to prevent blurred outputs. In order to alleviate the neuron death, we use several conv-leakyrelu blocks instead of the conv-relu blocks.

The loss function of the autoencoder network includes two loss functions: the multi-scale loss and the perceptual loss. The multi-scale loss function of the autoencoder network is defined as (14), which can extract features with different scales [1].

$$\mathcal{L}_M(\{\mathbf{S}\}, \{\mathbf{T}\}) = \sum_{i=1}^{M} \phi_i \mathcal{L}_{MSE}(\mathbf{S}_i, \mathbf{T}_i), \tag{14}$$

where $\mathbf{S}_i$ and $\mathbf{T}_i$ represent the i^{th} output of the decoder layers and the ground truth which have the same scale. ϕ_i represents the i^{th} weight. The value of ϕ increases with the scales and is set typically between [0,1]. The outputs of the last first, third

and fifth layers are used whereas smaller layers are not used since the information is insignificant.

Based on the VGG, the perceptual loss function of the autoencoder network is defined as (15), that measures the global discrepancy between the features of the autoencoder s output and the corresponding ground-truth image can be learned from the training data [1].

$$\mathcal{L}_P(\mathbf{B},\mathbf{T}) = \sum_{i=1}^{M} \mathcal{L}_{MSE}(VGG(\mathbf{B}), VGG(\mathbf{T})), \tag{15}$$

where VGG is a pretrained CNN, and produces features from a given input image. $\mathbf{B} = G(\mathbf{O}_r)$ indicates the output image of the whole generative network. $\mathbf{T}$ is the ground-truth image without raindrops.

Therefore, the loss function of the generative network can be written as [1]:

$$\mathcal{L}_G = \mathcal{L}_{GAN}(\mathbf{B}) + \mathcal{L}_M(\{\mathbf{S}\},\{\mathbf{T}\}) + \mathcal{L}_P(\mathbf{B},\mathbf{T}) + \mathcal{L}_{ATT}(\{\mathbf{A}_{att}\},\mathbf{L}), \tag{16}$$

where $\mathcal{L}_{GAN}(\mathbf{B}) = \eta \log(1 - D(\mathbf{B}))$, D represent the process of producing an image by the discriminative network, and η is a constant and set to 0.01.

2) Discriminative network: As shown in Fig. 3, to differentiate candidates produced by the generator network from the true data distribution, the discriminative network aims to distinguish the regions degraded by the raindrops, which is constructed by seven convolution layers with the kernel of (3, 3), a fully connected layer of 1024 and each neuron with a sigmoid activation function [1].

The loss function of the discriminator network can be expressed as [1]:

$$\mathcal{L}_D = -\log(D(\mathbf{C})) - \log(1 - D(\mathbf{B})) + \gamma \mathcal{L}_{MAP}(\mathbf{B},\mathbf{C},\mathbf{A}_J). \tag{17}$$

Here, $\mathbf{C}$ is a sample image drawn from a pool of real and clean images, $\mathbf{A}_J$ denotes the J^{th} attention map, γ is the calibration factor. $\mathcal{L}_{MAP}(\mathbf{B},\mathbf{C},\mathbf{A}_J) = \mathcal{L}_{MSE}(D_{map}(\mathbf{B}),\mathbf{A}_J) + \mathcal{L}_{MSE}(D_{map}(\mathbf{C}),0)$ describes the loss between the features extracted from interior layers of the discriminator and the J^{th} attention map, where D_{map} represents the process of producing a 2D map by the discriminative network. 0 represents an attention map containing only 0 values and implies that no specific region needs to be attended.

The proposed algorithm is summarized in Algorithm 1.

Algorithm 1 Joint raindrop and haze removal algorithm

Input: The input image $\mathbf{O}$ corrupted by raindrops and the haze as given in (4).

Output: Recovery of $\mathbf{B}$.

1. *Global atmospheric light computation.*
$A = pL_{med} + (1-p)D_{dmax}$,
where $p, K, k, L_{med}, D_{dmax}$ are defined as in (5).
2. *Transmission map computation.*
$$t(x) = \begin{cases} 1 - \omega \frac{\min\limits_{y\in\Omega(x)} [\min\limits_{c\in\{R,G,B\}} O^c(y)]}{A}, & O(x) < A, \\ 1 - \omega \frac{1 - \max\limits_{y\in\Omega(x)} [\max\limits_{c\in\{R,G,B\}} O^c(y)]}{1-A}, & O(x) \geq A, \end{cases}$$
where $O(x), O^c(y), \omega, \Omega(x)$ are defined as in (5) and (7).
3. *Recovery of the background image* $\mathbf{B}$.
$\arg\min_{\mathbf{B},\mathbf{R},\mathbf{L}} \|(\mathbf{O} - A(\mathbf{I} - \mathbf{t})) \oslash \mathbf{t} - \mathbf{B} - (\mathbf{I} - \mathbf{L}) \circ \mathbf{R}\|_2^2$.
4. *Loss function of generative network.*
$\mathcal{L}_G = \mathcal{L}_{GAN}(\mathbf{B}) + \mathcal{L}_M(\{\mathbf{S}\},\{\mathbf{T}\}) + \mathcal{L}_P(\mathbf{B},\mathbf{T}) + \mathcal{L}_{ATT}(\{\mathbf{A}_{att}\},\mathbf{L})$,
where $\mathcal{L}_{GAN}, \mathcal{L}_M, \mathcal{L}_P, \mathcal{L}_{ATT}$ are defined as in (13), (14), (15) and (16).
5. *Loss function of discriminative network.*
$\mathcal{L}_D = -\log(D(\mathbf{C})) - \log(1 - D(\mathbf{B})) + \gamma \mathcal{L}_{MAP}(\mathbf{B},\mathbf{C},\mathbf{A}_J)$,
where $\mathcal{L}_{MAP}, \gamma, \mathbf{C}, \mathbf{A}_J$ are defined as in (17).

V. Numerical Experiments

In this section, we conduct numerical simulations to demonstrate the competitive performance of the proposed multi-task JRHR algorithm.

Experimental Data. We use the following two kinds of images for the experiment:

[*Synthetic images*] A dataset RH of 1619 images is composed of two parts, including the dataset captured by Qian et al. [1] and 500 clean/corrupted pairs of images captured by us. We use Nikon D5300 to capture various background scenes which include the raindrops and the haze. The thickness of the glass slabs is 3 mm. In order to minimize the reflective effect of the glass, the distance between the glass slabs and the camera lens has been set between 2 to 8 cm to generate the diverse raindrop images. Fig. 4 shows some examples of the dataset RH.

Fig. 4. Samples of the dataset. Top: The images corrupted by raindrops and haze. Bottom: The corresponding ground-truth images.

[*Real-world images*] Different from the synthetic images, the real-world images without ground truth are selected from Google and Baidu search engines, and captured from several surveillance cameras without movement of lens.

Our algorithm is compared with the state-of-the-art algorithms on these two kinds of images. The dataset for training our network is the dataset RH of 1619 images. The testing images considered for the synthetic image simulations are randomly picked from the dataset RH. The testing images considered for the real-world image simulations are selected from Google and Baidu search engines, and captured from several surveillance cameras without movement of lens.

Baseline methods. We compare some versions of our JRHR algorithm: A1 (removing two layers of ResNet), A2 (increasing two layers of ResNet), B1 (removing two convolution layers of the generative network), B2 (increasing two convolution layers of the generative network), C1 (removing two attention maps), C2 (increasing two attention maps), D (changing loss function to MSE), JRHR (full version of our JRHR algorithm)

with the state-of-the-art algorithms: Feature Fusion Attention Network (FFA)[1] [14], All-in-One Dehazing Network (AOD)[2] [15], Single Image Haze Removal Using Dark Channel Prior (DCP)[3] [11], CNN-based raindrop removal method (CNN)[4] [9], conditional adversarial networks (Pix2Pix)[5] [29], and attentive generative adversarial network (AGAN)[6] [1].

All our algorithms are trained from scratch. Other methods come from online available resources kindly provided by the authors. For evaluations on synthesized images, we train the model with the corresponding training data from scratch, without any fine-tuning. CNN-based raindrop removal method [9] is implemented in MATLAB. The facilities that were used to perform the experiments include AMD Ryzen 7 2700 3.2 GHz CPU, NVIDIA GeForce RTX 2080Ti Graphics Card and 14.9 GB memory. The results are given in Table I. The results show that the proposed JRHR algorithm has better performance in mean values of SSIM and PSNR than the DCP, AOD, FFA, CNN, Pix2Pix and AGAN algorithms. The SSIMs and PSNRs of the DCP, AOD, FFA, CNN, Pix2Pix and AGAN algorithms are less than 0.87 and 26 dB. The computational complexity of both methods in terms of run time was also approximately calculated. Our proposed algorithms in GPU are capable of dealing with a 480×640 image corrupted by raindrops and haze in less than 3s.

TABLE I
THE PERFORMANCE AND TIME COMPLEXITY OF OUR JRHR ALGORITHM COMPARED WITH STATE-OF-THE-ART METHODS.

Algorithm	SSIM	PSNR (dB)	Running time (sec)
DCP [11]	0.7342	19.71	0.47
AOD [15]	0.7840	21.79	2.23
FFA [14]	0.8042	23.88	2.13
CNN [9]	0.8114	22.45	5.32
P2P [29]	0.7458	18.24	1.58
AGAN [1]	0.8640	25.32	2.15
A1	0.8412	24.38	1.56
A2	0.9114	27.64	2.83
B1	0.8571	23.93	1.65
B2	0.9128	28.18	2.35
C1	0.8167	20.59	2.14
C2	0.9128	27.33	2.67
D	0.8735	27.59	3.18
JRHR	**0.9131**	**28.31**	2.11

For the experiments on synthetic images, the performance of the proposed algorithm can be evaluated by Structure Similarity Index (SSIM) [30] and Peak Signal-to-Noise Ratio (PSNR) [31], [32]. For the experiments on real-world images, the performance of the proposed algorithm can be evaluated by blind image quality index (BIQI) [33] and Blind referenceless image spatial quality evaluator (BRISQUE) [34].

[1]https://github.com/zhilin007/FFA-Net
[2]https://github.com/weber0522bb/AODnet-by-pytorch
[3]https://github.com/He-Zhang/image_dehaze
[4]https://cs.nyu.edu/∼deigen/rain/
[5]https://github.com/phillipi/pix2pix
[6]https://github.com/MaybeShewill-CV/attentive-gan-derainnet

Image quality assessment (IQA) can be achieved using subjective or objective methods. The real-world images used in our experiments do not have the ground truth that we can compare with. For subjective IQA, we can only use single-stimulus methods, which depend mainly on the way in which the viewers rate their opinions based on their perceptions of image quality. One way to ensure the reliability of the results is to get experienced personnel to rate their opinions based on their perceptions, and the other way is to recruit a large number of viewers to rate their opinions based on their perceptions of image quality. These opinions are afterwards mapped onto numerical values. This method is costly and time consuming. Therefore, we consider objective IQA, which is a no-reference (NR) method for assessing the quality of the enhanced image obtained from the real-world images without ground truth. Blind image quality index (BIQI) and blind/referenceless image spatial quality evaluator (BRISQUE) are commonly used NR methods which are based on natural scene statistic (NSS), and evaluated on the LIVE IQA database [35]. Once trained, the BIQI and BRISQUE methods do not require any knowledge of the distortions introduced, and can be extended to any number of distortions. Therefore, it is economically cheaper and more efficient to obtain the perceptual scores [33], [34]. In addition, such metrics have been shown to be highly correlated with the subjective IQA [33] and therefore they can be used as an alternative to subjective IQA when the resources for performing perceptual tests are limited.

The analysis of variance (ANOVA) based statistical significance evaluation [36] of the proposed JRHR algorithm as compared with the baseline methods is also given in Section V. ANOVA is a statistical hypothesis testing heavily used in the analysis of experimental data (e.g., in image and speech processing), which is a relatively robust procedure with respect to violations of the normality assumption, and has lower probability of introducing Type I errors (false positives) compared to T-Tests. According to the number of factors considered in the tests, ANOVA includes one-way ANOVA, two-way ANOVA, and multi-way ANOVA.

In this paper, considering the single-factor results (e.g. SSIM, PSNR, BIQI, or BRISQUE), we use one-way ANOVA based statistical significance evaluation (using the F distribution) [36] on the means results obtained by the proposed method as compared with the baseline methods, for both the synthetic images and the real-world images.

A. JRHR for the synthetic images

In the first set of simulations, we evaluate the restoration performance of the proposed JRHR algorithm described in Algorithm 1. The synthetic images considered for the first simulation are randomly picked from an image dataset RH.

For the improved haze removal method, the parameter ω in the transmission map is a value between [0,1]. For the optimized visual attention network, the total number of iterations J of the attention maps, the calibration factor θ and the parameter λ in the penalty term are set to be 4, 0.8 and 0.001, respectively. $\mathbf{F}_0$ is the input image concatenated with the initial attention map $\mathbf{A}_1$ with the values of 0.5. For

the autoencoder network, in the multi-scale loss function, the output sizes of the last first, third, and fifth layers are 1/4, 1/2 and 1 of the original size, and ϕ are set to 0.6, 0.8, 1.0, respectively. In the loss function of the generative network, η is a constant and set to 0.01. In the loss function of the discriminative network, γ is a calibration factor and set to 0.05.

Fig. 5 shows the restoration results of the JRHR algorithm on the images corrupted by raindrops and haze. According to the density of the raindrops and haze, the proposed JRHR algorithm is successful in removing the majority of haze and raindrops, and recovering background images.

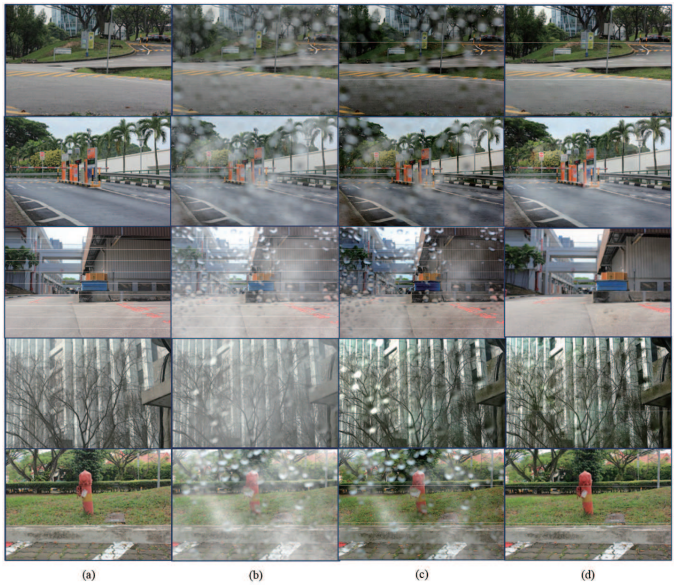

Fig. 5. The restoration results of the JRHR algorithm: (a) Ground-truth images (b) Input images corrupted by raindrops and haze (c) De-hazed images (d) Restored background images.

We compare the proposed JRHR algorithm with six state-of-the-art algorithms as shown in Fig. 6. As observed, the proposed JRHR algorithm significantly outperforms DCP, AOD, FFA, CNN, Pix2Pix and AGAN algorithms with respect to the density of the raindrops and haze in removing raindrops and haze, enhancing the visibility and preserving details.

Table II shows the results of different algorithms. As observed, the SSIMs of the proposed JRHR algorithm are closer to 1 than the DCP, AOD, FFA, CNN, Pix2Pix, AGAN algorithms, and the PSNRs of the proposed algorithm are better than those of the baseline algorithms. The proposed JRHR algorithm achieves better results than the baseline algorithms in terms of both SSIM and PSNR.

To evaluate the statistical significance of the performance, we perform one-way ANOVA based F-test [36] for the SSIM and PSNR of the DCP, AOD, FFA, CNN, Pix2Pix, AGAN and the proposed JRHR algorithms in Table III. The average results of SSIM and PSNR for 800 synthetic images are also given in Table III. The p-value stands for the probability of a more extreme (positive or negative) result than what we actually achieved, given that the null hypothesis is true. F-value can be defined as the ratio of the variance of the group means to the mean of the within group variances. All the F-tests in this work have been carried out at 5 % significance level. If p-value is greater than 0.05 (5 % significance level), then the given results are statistically insignificant. It can be observed that the p-values of all the algorithms in Table III are smaller than 0.05, suggesting that the improvement given by the proposed JRHR algorithm as compared with the baseline methods is statistically significant.

TABLE II
PERFORMANCE COMPARISON OF THE DCP, AOD, FFA, CNN, PIX2PIX AND AGAN ALGORITHMS FOR THE SYNTHETIC IMAGES.

		Algorithm						
Image	Metric	DCP	AOD	FFA	CNN	P2P	AGAN	JRHR
Road	SSIM	0.7011	0.7480	0.7647	0.7998	0.7551	0.7679	**0.8843**
	PSNR (dB)	22.82	21.99	26.39	23.67	22.65	24.88	**27.87**
Entrance	SSIM	0.6478	0.6585	0.7151	0.6423	0.7068	0.8112	**0.8685**
	PSNR (dB)	19.88	24.98	25.10	18.86	20.56	25.21	**26.92**
Hill	SSIM	0.7820	0.8349	0.8499	0.9127	0.8708	0.9067	**0.9430**
	PSNR (dB)	25.15	26.17	26.40	26.81	25.87	27.60	**29.34**
Building	SSIM	0.7656	0.7528	0.7528	0.6972	0.7723	0.7465	**0.8550**
	PSNR (dB)	21.79	20.78	22.01	21.71	22.18	24.92	**28.38**
Safety	SSIM	0.8126	0.8070	0.8572	0.7654	0.7768	0.8423	**0.8743**
	PSNR (dB)	18.40	21.47	21.19	23.94	22.24	24.58	**27.56**

TABLE III
ANOVA BASED STATISTICAL SIGNIFICANCE EVALUATION OF THE PSNR AND SSIM FOR THE DCP, AOD, FFA, CNN, PIX2PIX, AGAN AND PROPOSED JRHR ALGORITHMS.

	JRHR (SSIM: 0.9131, PSNR: 28.31 dB)					
Algorithm	SSIM			PSNR		
	mean	F-value	p-value	mean (dB)	F-value	p-value
DCP	0.7342	46.22	1.4508e-11	19.71	54.26	2.7089e-13
AOD	0.7840	47.52	7.6227e-12	21.79	42.1	1.1281e-10
FFA	0.8042	25.26	5.5181e-07	23.88	26.43	3.0434e-07
CNN	0.8114	14.37	0.0002	22.45	14.82	0.0001
P2P	0.7458	97.93	1.7067e-22	18.24	84.87	8.9739e-20
AGAN	0.8640	12.57	0.0004	25.32	11.24	0.0008

In Fig. 7, the learned features with the raindrop region detection by the proposed JRHR algorithm are visualized in the testing stage. As observed, the learned features is mostly correlated to the raindrop regions and relevant structures, which demonstrates the necessity of employing raindrop region detection in the JRHR algorithm.

The proposed JRHR algorithm is compared with the haze removal algorithms (DCP, AOD and FFA) and the raindrop removal algorithms (CNN, Pix2Pix and AGAN) as shown in Fig. 8 and Fig. 9, respectively. It is observed that the proposed JRHR algorithm outperforms DCP, AOD, FFA, CNN, Pix2Pix and AGAN algorithms in removing the effects of the haze and the raindrops respectively.

In Table IV and V, the statistical significance evaluation of the performance achieved by the haze and raindrop removal algorithms, respectively. The average results of SSIM and PSNR of the haze removal algorithms are given in Table IV,

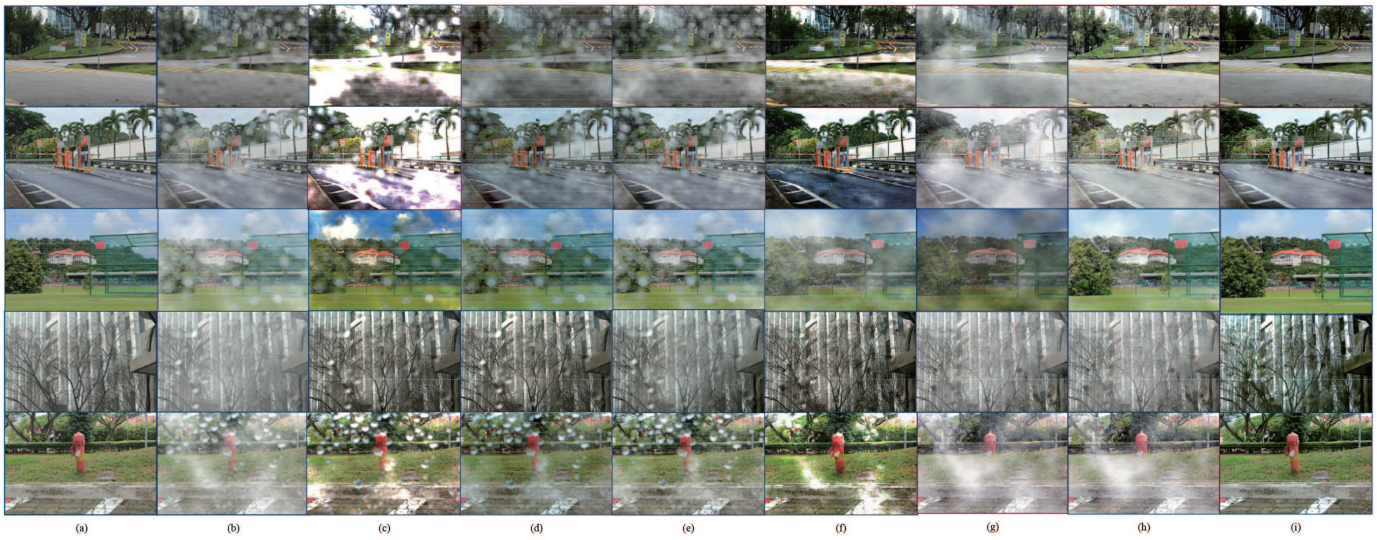

Fig. 6. The restoration results of different algorithms on synthesized images (a) Ground-truth images (b) Input images corrupted by raindrops and haze (c) DCP (d) AOD (e) FFA (f) CNN (g) Pix2Pix (h) AGAN (i) Proposed JRHR.

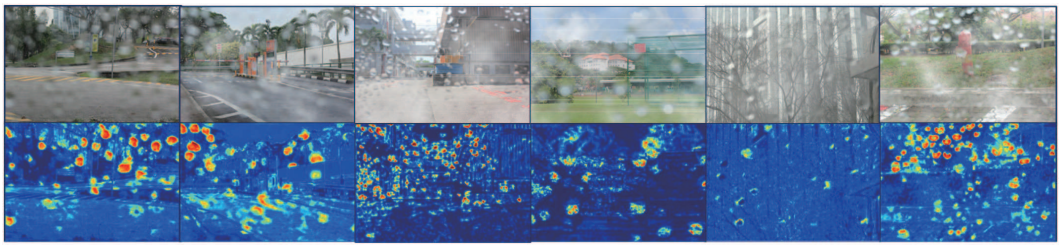

Fig. 7. The learned features with the raindrop region detection with respect to the density of the raindrops and haze. Top: Input images corrupted by raindrops and haze. Bottom: Detected raindrop regions.

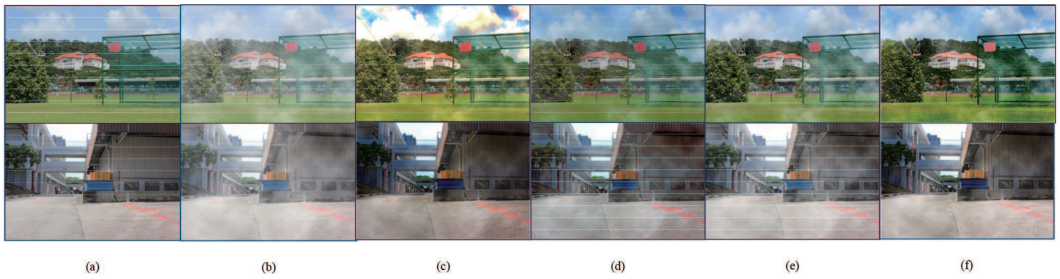

Fig. 8. The restoration results of different haze removal algorithms on synthesized images (a) Ground-truth images (b) Input images corrupted by haze (c) DCP (d) AOD (e) FFA (f) Proposed JRHR.

and the average results of SSIM and PSNR of the raindrop removal algorithms are given in Table V. All the p-values of the haze removal algorithms (DCP, AOD and FFA) and the raindrop removal algorithms (CNN, Pix2Pix and AGAN) are smaller than 0.05. This indicates that the proposed JRHR algorithm outperforms these compared algorithms in removing the effects of the raindrops and the haze, respectively.

B. JRHR for the real-world images

In the second set of simulations, we evaluate the restoration performance of the proposed JRHR algorithm for the real-world images without ground truth.

Fig. 10 demonstrates the restoration results of the JRHR algorithm on the real-world images corrupted by raindrops and haze, which shows the effectiveness of the JRHR algorithm in

TABLE IV
ANOVA BASED HAZE REMOVAL STATISTICAL SIGNIFICANCE EVALUATION OF THE PSNR AND SSIM FOR THE DCP, AOD, FFA AND PROPOSED JRHR ALGORITHMS.

Algorithm	JRHR (SSIM: 0.9300, PSNR: 30.70 dB)					
	SSIM			PSNR		
	mean	F-value	p-value	mean (dB)	F-value	p-value
DCP	0.9041	15.99	6.6483e-05	26.79	23.85	1.1381e-06
AOD	0.9252	4.27	0.0388	29.99	4.85	0.0277
FFA	0.9283	3.95	0.047	30.03	4.32	0.0379

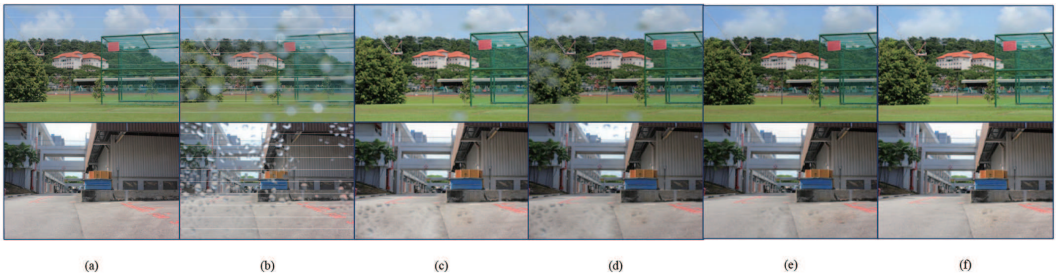

Fig. 9. The restoration results of different raindrop removal algorithms on synthesized images (a) Ground-truth images (b) Input images corrupted by raindrops (c) CNN (d) Pix2Pix (e) AGAN (f) Proposed JRHR.

TABLE V
ANOVA BASED RAINDROP REMOVAL STATISTICAL SIGNIFICANCE EVALUATION OF THE PSNR AND SSIM FOR THE CNN, PIX2PIX, AGAN AND PROPOSED JRHR ALGORITHMS.

	JRHR (SSIM: 0.9289, PSNR: 30.31 dB)					
Algorithm	SSIM			PSNR		
	mean	F-value	p-value	mean (dB)	F-value	p-value
CNN	0.9263	10.56	0.0012	27.22	10.04	0.0016
P2P	0.9039	173.35	8.6310e-38	25.70	104.83	6.3504e-24
AGAN	0.9246	4.67	0.0299	30.27	5.09	0.0242

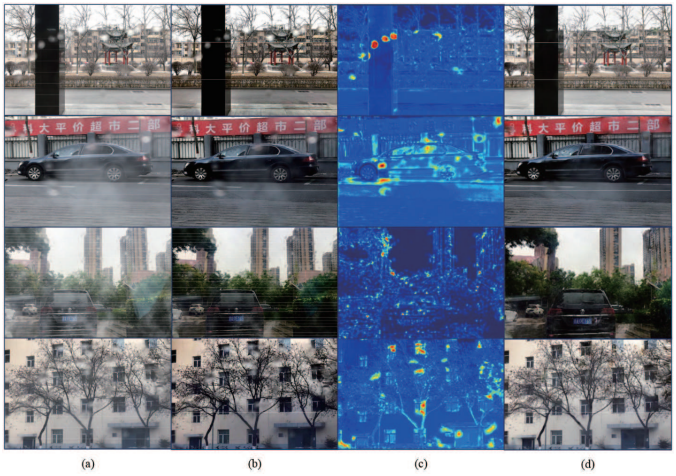

Fig. 10. The restoration results of the JRHR algorithm: (a) Input images corrupted by raindrops and haze (b) De-hazed images (c) Detected raindrop regions (d) Restored background images.

TABLE VI
PERFORMANCE COMPARISON OF THE DCP, AOD, FFA, CNN, PIX2PIX AND AGAN ALGORITHMS FOR THE REAL-WORLD IMAGES.

		Algorithm						
Image	Metric	DCP	AOD	FFA	CNN	P2P	AGAN	JRHR
Pavilion	BIQI (dB)	39.71	32.73	29.02	26.04	35.99	33.84	**25.94**
	BRISQUE (dB)	28.37	26.92	27.98	23.06	22.01	34.94	**21.84**
Car	BIQI (dB)	30.44	27.21	23.93	24.10	29.12	27.08	**21.52**
	BRISQUE (dB)	30.68	26.88	28.54	34.32	26.00	30.79	**25.99**
Skyscraper	BIQI (dB)	38.07	31.27	27.08	27.11	45.11	31.55	**21.71**
	BRISQUE (dB)	34.68	25.61	26.88	24.04	42.67	33.46	**17.72**
Building	BIQI (dB)	58.24	50.56	43.79	52.76	56.57	43.73	**41.63**
	BRISQUE (dB)	27.61	21.26	22.89	21.59	22.95	12.47	**12.07**

TABLE VII
ANOVA BASED STATISTICAL SIGNIFICANCE EVALUATION OF THE BIQI AND BRISQUE FOR THE REAL-WORLD IMAGES.

	JRHR (BIQI: 27.17 dB, BRISQUE: 18.38 dB)					
Algorithm	BIQI			BRISQUE		
	mean (dB)	F-value	p-value	mean (dB)	F-value	p-value
DCP	42.37	27.01	0.0001	32.79	24.38	0.0002
AOD	39.25	26.75	0.0001	28.43	25.64	0.0002
FFA	36.30	14.66	0.0018	25.67	12.93	0.0029
CNN	35.40	7.81	0.0143	24.47	11.33	0.0046
P2P	48.32	29.33	9.1005e-05	35.06	27.59	0.0001
AGAN	31.57	5.79	0.0305	20.19	6.19	0.026

recovering background images.

Considering no ground-truth images, we use BIQI and BRISQUE to evaluate the restoration performance on the basis of the blind image quality assessment [37]. Table VI shows that the BIQIs and BRISQUEs of the proposed JRHR algorithm are smaller than those of DCP, AOD, FFA, CNN, Pix2Pix and AGAN algorithms. As observed, the proposed JRHR algorithm outperforms these algorithms in terms of both BIQI and BRISQUE.

Table VII illustrates the statistical significance evaluation of the performance by performing one-way ANOVA based F-test [36] for the BIQI and BRISQUE of the DCP, AOD, FFA, CNN, Pix2Pix, AGAN and the proposed JRHR algorithms. All the F-tests in this work have been carried out at 5 % significance level and all the p-values in Table VII are smaller than 0.05, suggesting that the improvement by the proposed JRHR algorithm over the baseline algorithms is statistically significant for the real-world images.

VI. CONCLUSION

The model and the algorithm for the problem of joint raindrop and haze removal (JRHR) have been investigated in this paper. Our contributions to this challenging problem are as follows:

Model: We form a new model of the JRHR problem to recover an image corrupted by raindrops and haze by detecting and removing the effects of the raindrops and the haze.

Algorithm: Based on the JRHR model, an integrated algorithm which combines an improved estimate of the atmospheric light, a modified transmission map, a GAN network and an

optimized visual attention network is presented as a solution to the JRHR problem.

Numerical experiments show that the proposed JRHR algorithm performs well in restoring the images corrupted by raindrops and haze. In the future, it is interesting to investigate how to incorporate an end to end optimization method into the JRHR algorithm. In the future, it is interesting to investigate how to incorporate an end to end optimization method into the JRHR algorithm. It is also tempting to consider blind source separation idea for restoring the images corrupted by raindrops and haze.

ACKNOWLEDGEMENT

The authors would like to thank the associate editor and anonymous reviewers for their constructive comments for improving this paper.

REFERENCES

[1] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2482–2491.

[2] N. Brewer and N. Liu, "Using the shape characteristics of rain to identify and remove rain from video," in *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. Springer, 2008, pp. 451–458.

[3] W. Yang, R. T. Tan, J. Feng, J. Liu, S. Yan, and Z. Guo, "Joint rain detection and removal from a single image with contextualized deep networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[4] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. Brown, "Single image rain streak separation using layer priors." *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, 2017.

[5] J. Xu, W. Zhao, P. Liu, and X. Tang, "An improved guidance image based method to remove rain and snow in a single image," *Computer and Information Science*, vol. 5, no. 3, p. 49, 2012.

[6] J.-H. Kim, C. Lee, J.-Y. Sim, and C.-S. Kim, "Single-image deraining using an adaptive nonlocal means filter," in *2013 IEEE International Conference on Image Processing*. IEEE, 2013, pp. 914–917.

[7] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, "Rain streak removal using layer priors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2736–2744.

[8] L. Zhu, C.-W. Fu, D. Lischinski, and P.-A. Heng, "Joint bi-layer optimization for single-image rain streak removal," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2526–2534.

[9] D. Eigen, D. Krishnan, and R. Fergus, "Restoring an image taken through a window covered with dirt or rain," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 633–640.

[10] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley, "Clearing the skies: A deep network architecture for single-image rain removal," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2944–2956, 2017.

[11] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.

[12] Y. Xu, X. Guo, H. Wang, F. Zhao, and L. Peng, "Single image haze removal using light and dark channel prior," in *2016 IEEE/CIC International Conference on Communications in China (ICCC)*. IEEE, 2016, pp. 1–6.

[13] D. Singh and V. Kumar, "Single image haze removal using integrated dark and bright channel prior," *Modern Physics Letters B*, vol. 32, no. 04, p. 1850051, 2018.

[14] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing," 2020.

[15] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017.

[16] K. Garg and S. K. Nayar, "Detection and removal of rain from videos," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1. IEEE, 2004, pp. I–I.

[17] ——, "Vision and rain," *International Journal of Computer Vision*, vol. 75, no. 1, pp. 3–27, 2007.

[18] Y.-H. Fu, L.-W. Kang, C.-W. Lin, and C.-T. Hsu, "Single-frame-based rain removal via image decomposition," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 1453–1456.

[19] Y.-L. Chen and C.-T. Hsu, "A generalized low-rank appearance model for spatio-temporally correlated rain streaks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1968–1975.

[20] X. Hu, C.-W. Fu, L. Zhu, and P.-A. Heng, "Depth-attentional features for single-image rain removal," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8022–8031.

[21] S. Li, W. Ren, J. Zhang, J. Yu, and X. Guo, "Single image rain removal via a deep decomposition–composition network," *Computer Vision and Image Understanding*, 2019.

[22] H. Kurihata, T. Takahashi, I. Ide, Y. Mekada, H. Murase, Y. Tamatsu, and T. Miyahara, "Rainy weather recognition from in-vehicle camera images for driver assistance," in *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.* IEEE, 2005, pp. 205–210.

[23] M. Roser and A. Geiger, "Video-based raindrop detection for improved image registration," in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*. IEEE, 2009, pp. 570–577.

[24] M. Roser, J. Kurz, and A. Geiger, "Realistic modeling of water droplets for monocular adherent raindrop recognition using bezier curves," in *Asian Conference on Computer Vision*. Springer, 2010, pp. 235–244.

[25] Q. Wu, W. Zhang, and B. V. Kumar, "Raindrop detection and removal using salient visual features," in *2012 19th IEEE International Conference on Image Processing*. IEEE, 2012, pp. 941–944.

[26] D. D. Webster and T. P. Breckon, "Improved raindrop detection using combined shape and saliency descriptors with scene context isolation," in *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 4376–4380.

[27] S. You, R. T. Tan, R. Kawakami, Y. Mukaigawa, and K. Ikeuchi, "Adherent raindrop modeling, detectionand removal in video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 9, pp. 1721–1733, 2015.

[28] A. Dudhane and S. Murala, "Ryf-net: Deep fusion network for single image haze removal," *IEEE Transactions on Image Processing*, vol. 29, pp. 628–640, 2019.

[29] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1125–1134.

[30] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli *et al.*, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[31] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electronics Letters*, vol. 44, no. 13, pp. 800–801, 2008.

[32] Y. Guo, X. Zhao, J. Li, A. Wang, and W. Wang, "Blind multiple input multiple output image phase retrieval," *IEEE Transactions on Industrial Electronics*, 2019.

[33] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 513–516, 2010.

[34] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Blind/referenceless image spatial quality evaluator," in *Proceedings of the IEEE Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2011, pp. 723–727.

[35] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[36] P. G. Hoel, *Workbook for Elementary Statistics*. J. Wiley, 1976.

[37] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1202–1213, 2017.

Yina Guo (M'16) received the B.Sc. degree in 2002 from China University of Mining and Technology, the M.E. degree in 2007, and the Ph.D. degree in 2014, all from Taiyuan University of Science and Technology, China.

She is currently a Professor in Signal Processing, at Taiyuan University of Science and Technology, China. She has authored/coauthored more than 40 papers and three books, and was granted eight patents and four software copyrights in China. Her research interests include blind source separation, biosignal processing and phase retrieval.

Prof. Guo was the recipient of several science and technology awards from Shanxi Province and holds grants from the National Natural Science Foundation of China and Shanxi Province. She is a Senior Member of China Institute of Communications.

Jianguo Chen received the B.E. degree in Communication Engineering from Huainan Normal University in 2017. He is a master student in the Department of Electronics and Information Engineering, Taiyuan University of Science and Technology, China.

His research interests include image processing and phase retrieval.

Xiaowen Ren received the B.E. degree in Electronic Science and Technology from Yuncheng University in 2016. He received the M.E. degree in 2019, in Electronics and Communications Engineering from Taiyuan University of Science and Technology, China.

He is currently an assistant researcher of Taiyuan Research Institute within China Coal Technology Engineering Group, China. His research interests include Image processing and electrical control.

Anhong Wang received the B.Sc. degree and the M.E. degree respectively in 1994 and 2002, all in Electronic Information Engineering from Taiyuan University of Science and Technology, China. She received the Ph.D. degree in 2009, in Information Science from Beijing Jiaotong University, China.

She is currently a Professor and the Director of Digital Media and Communication Institute, Taiyuan University of Science and Technology, China. She has authored/coauthored more than 90 papers in international journals and conferences. She is leading several research projects, including three funded by the National Science Foundations of China. Her research interests include image and video coding and transmission, compressed sensing, and secret image sharing.

Wenwu Wang (M'02-SM'11) received the B.Sc. degree in 1997, the M.E. degree in 2000, and the Ph.D. degree in 2002, all in Electrical Engineering from Harbin Engineering University, China.

He is currently a Professor in Signal Processing and Machine Learning, at University of Surrey, and a Co-Director of the Machine Audition Lab within the Centre for Vision Speech and Signal Processing. He has coauthored more than 250 publications. His current research interests include signal processing and machine learning.

Prof. Wang has been a Senior Area Editor since 2019 and an Associate Editor from 2014 to 2018 for IEEE Transactions on Signal Processing. He is an Associate Editor since 2020 for IEEE/ACM Transactions on Audio Speech and Language Processing, and an Associate Editor since 2019 for EURASIP Journal on Audio Speech and Music Processing. He is a Publication Co-Chair for the ICASSP 2019, Brighton, UK. He also serves as a Member since 2019 of the International Steering Committee of Latent Variable Analysis and Signal Separation.