A Comprehensive Survey on Image Dehazing Based on Deep Learning

Jie Gui[1] , **Xiaofeng Cong**[2] , **Yuan Cao**[3*] , **Wenqi Ren**[4] , **Jun Zhang**[2] , **Jing Zhang**[5] and **Dacheng Tao**[6]

[1]Southeast University
[2]AnHui University
[3]Ocean University of China
[4]Institute of Information Engineering, Chinese Academy of Sciences
[5]University of Sydney
[6]JD Explore Academy, JD.com, China
guijie@seu.edu.cn, cxf_svip@163.com, cy8661@ouc.edu.cn, renwenqi@iie.ac.cn,
wwwzhangjun@163.com, jing.zhang1@sydney.edu.au, dacheng.tao@gmail.com

Abstract

The presence of haze significantly reduces the quality of images. Researchers have designed a variety of algorithms for image dehazing (ID) to restore the quality of hazy images. However, there are few studies that summarize the deep learning (DL) based dehazing technologies. In this paper, we conduct a comprehensive survey on the recent proposed dehazing methods. Firstly, we summarize the commonly used datasets, loss functions and evaluation metrics. Secondly, we group the existing researches of ID into two major categories: supervised ID and unsupervised ID. The core ideas of various influential dehazing models are introduced. Finally, the open issues for future research on ID are pointed out.

1 Introduction

Images with clear visibility are required for computer vision tasks [Li *et al.*, 2017a]. However, due to the absorption by floating particles contained in the environment, images taken in hazy days often suffer from quality degradation. With the aim of restoring color and details from distorted images, it is necessary to design an effective dehazing algorithm.

Dehazing methods based on deep learning are mainly divided into supervised ID and unsupervised ID. There is not too much work about semi-supervised dehazing [Li *et al.*, 2020b]. In general, supervised methods require pairs of hazy images and haze-free images/transmission maps. The unsupervised methods often take unpaired images as the training set. There are three important factors in the research of supervised ID and unsupervised ID:

- Atmospheric scattering model (ASM) [McCartney, 1976; Nayar and Narasimhan, 1999; Narasimhan and Nayar, 2003] provides a theoretical basis for the research of image dehazing.

---

*Corresponding Author

Methods	ASM	Trans	Extra
DehazeNet [Cai *et al.*, 2016a]	✓	✓	
MSCNN [Ren *et al.*, 2016]	✓	✓	
AOD-Net [Li *et al.*, 2017a]	✓		
DCPDN [Zhang and Patel, 2018]	✓	✓	✓
SSMD-Net [Song *et al.*, 2019]	✓	✓	
BidNet [Pang *et al.*, 2020]	✓	✓	✓
FAMED-Net [Zhang and Tao, 2020]	✓		
FEED-Net [Zhang *et al.*, 2018]			
GFN [Ren *et al.*, 2018]			
Grid-Dehaze [Liu *et al.*, 2019]			✓
GCA-Net [Chen *et al.*, 2019]			
EPDN [Qu *et al.*, 2019]			✓
MSBDN [Dong *et al.*, 2020a]			
FFA-Net [Qin *et al.*, 2020]			
KDDN [Hong *et al.*, 2020]		✓	✓
DAID [Shao *et al.*, 2020]			
FD-GAN [Dong *et al.*, 2020b]			✓
cyc-D [Engin *et al.*, 2018]			✓
E-cyc [Liu *et al.*, 2020]	✓		✓
Deep-DCP [Golts *et al.*, 2020]	✓		
DHL [Cong *et al.*, 2020]			✓
ZID [Li *et al.*, 2020a]	✓		

Table 1: A taxonomy of dehazing methods and corresponding factors. From top to bottom: the first group is the supervised model based on ASM, the second group is the non-ASM-based supervised model, and the third group is the unsupervised model.

- Transmission map (Trans): In ASM, Trans is the part of light that is not scattered and reaches the camera. After measuring the depth information of the scene, Trans can be obtained by the ASM.

- Extra data: With the help of pre-trained models, dehazing algorithms have achieved better performances by minimizing the perceptual loss [Justin *et al.*, 2016].

Based on the above factors, researchers have designed many effective dehazing algorithms. For the convenience of discussion, we further divide the supervised dehazing algorithms into two types: ASM-based and non-ASM-based. Details can be found in Table 1.

Goals of our survey. We aim to (i) summarize commonly used datasets and loss functions in day-time dehazing task;

(ii) provide a taxonomy for state-of-the-art DL-based dehazing algorithms and introduce the core techniques among various methods; (iii) point out open issues that would inspire further researches in ID tasks.

The differences between our survey and others. [Singh and Kumar, 2018; Li *et al.*, 2017b; Xu *et al.*, 2016] gave a detailed summary of the classic dehazing methods, including depth estimation, wavelet, enhancement, filtering, but they lack the detailed discussion of recent neural network models. [Parihar *et al.*, 2020] provides some investigations of supervised dehazing models, but it does not pay attention to the latest applications of unsupervised methods. [Banerjee and Chaudhuri, 2020] systematically introduced and grouped the existing nighttime ID methods [Zhang *et al.*, 2020], but seldom mentioned the dehazing methods for the day-time. Our survey gives a comprehensive survey on the supervised and unsupervised DL-based dehazing models for the day-time.

2 Related Work

In this section, we will introduce the atmospheric scattering model (ASM) [McCartney, 1976; Nayar and Narasimhan, 1999; Narasimhan and Nayar, 2003], widely used datasets, metrics for evaluation, and loss functions in the research of ID.

2.1 Atmosphere Scattering Model

The formula of ASM is as following:

$$I(x) = J(x)t(x) + A(1 - t(x)), \qquad (1)$$

where x represents the position of pixels and A stands for the global atmospheric light. $I(x)$ denotes the hazy scene and $J(x)$ is the haze-free image that the dehazing method expects to obtain. The $t(x)$ means the medium Trans:

$$t(x) = e^{-\beta d(x)}, \qquad (2)$$

where β and $d(x)$ represent the atmosphere scattering parameter and the depth of the scene, respectively. Meanwhile, it means that the Trans $t(x)$ can be obtained through the depth map $d(x)$, which is very critical for the synthesis of the dataset. If we can estimate $t(x)$ and A, the formula of the dehazing image $J(x)$ is

$$J(x) = \frac{I(x) - A(1 - t(x))}{t(x)}. \qquad (3)$$

This ASM is widely used in the research of image dehazing [Cai *et al.*, 2016a] and video dehazing [Li *et al.*, 2018a; Cai *et al.*, 2016b].

2.2 Datasets for Image Dehazing Tasks

In order to ensure the reliability of the data, the researchers mainly use two strategies to obtain paired hazy and haze-free images. One is synthesized based on the dataset containing depth annotation, such as the D-Hazy [Ancuti *et al.*, 2016] that synthesized on the basis of NYU-Depth [Silberman *et al.*, 2012] and Middlebury [Scharstein *et al.*, 2014]. Similar datasets include HazeRD [Zhang *et al.*, 2017] and RESIDE [Li *et al.*, 2019]. The other kind of dataset does not

Dataset	Type	Number	I/O
D-Hazy [Ancuti *et al.*, 2016]	Syn	1400+	I
HazeRD [Zhang *et al.*, 2017]	Syn	15	O
I-HAZE [Ancuti *et al.*, 2018c]	HG	35	I
O-HAZE [Ancuti *et al.*, 2018a]	HG	45	O
RESIDE [Li *et al.*, 2019]	S&R	10000+	I&O
Dense-HAZE [Ancuti *et al.*, 2019]	HG	33	O
NH-HAZE [Ancuti *et al.*, 2020a]	HG	55	O
MRFID [Liu *et al.*, 2020]	Real	200	O
BeDDE [Zhao *et al.*, 2020]	Real	200+	O

Table 2: Datasets for ID tasks. Syn means synthetic hazy images. HG stands for the hazy images generated from haze generator. Real represents real world scenes. S&R denote Syn&Real. I/O denote indoor/outdoor.

contain depth information, such as I-HAZE [Ancuti *et al.*, 2018c], O-HAZE [Ancuti *et al.*, 2018a], Dense-HAZE [Ancuti *et al.*, 2019], NH-HAZE [Ancuti *et al.*, 2020a], which are produced by a professional haze generator. Based on these four datasets, researchers have proposed a variety of effective dehazing algorithms in the New Trends in Image Restoration and Enhance (NTIRE 2018-2020) challenge [Ancuti *et al.*, 2018b; C.Ancuti *et al.*, 2019; Ancuti *et al.*, 2020b].

However, there is an inherent difference between synthetic/generated data and real data, which is called domain shift. The recently proposed MRFID [Liu *et al.*, 2020] and BeDDE [Zhao *et al.*, 2020] contain hazy and haze-free images collected in the real world, but its insufficient data makes it unsuitable for large-capacity dehazing models. We list the attributes of various datasets in Table 2.

2.3 Image Quality Metrics

The two most common full-reference metrics in dehazing research are Peak Signal to Noise Ratio (PSNR) and structural similarity index (SSIM) [Wang *et al.*, 2004]. Some dehazing researchers use CIEDE2000 [Sharma *et al.*, 2005] to evaluate the color difference between the dehazed image and the ground-truth clean image.

2.4 Loss Function

For convenience, this section introduces the commonly used loss functions of supervised and unsupervised dehazing models. The widely used pixel-wise loss functions for dehazing research are L1 loss and L2 loss, which require fewer operations in the optimization process. As a metric of the dehazing model, SSIM is also used as a loss function in the optimization process [Dong *et al.*, 2020b]:

$$L_{SSIM}(P) = \frac{1}{N} \sum_{p \in P} (1 - SSIM(p)). \qquad (4)$$

However, neither L1/L2 loss nor SSIM loss considers the special attributes of the human visual system in the process of perceptual evaluation. Recent researches use pre-trained classification neural networks to calculate the perceptual loss [Engin *et al.*, 2018; Cong *et al.*, 2020] in the feature space $\psi(\cdot)$:

$$L_{per}(x, y) = \sum_{i=1}^{n} ||\psi(x) - \psi(y)||^2. \qquad (5)$$
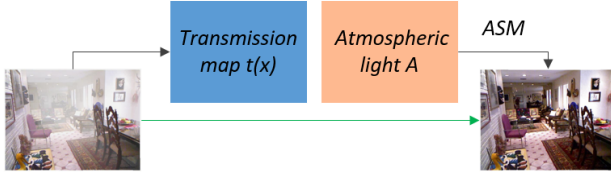
Figure 1: The workflow of the supervised dehazing method. Black arrows indicate ASM-based method and the green arrow stands for methods that do not require ASM.

Perceptual loss provides a way to use external data to assist the training of dehazing models as shown in Table 1, but we think it still needs further research.

3 Supervised Dehazing Algorithms

According to whether the ASM is used or not, the supervised dehazing methods can be divided into two groups as shown in Figure 1. The first group is to combine the ASM with the convolutional neural network to estimate the parameters for the dehazing process. This physical-model-based dehazing algorithm usually needs to utilize the depth information and the airlight value (black arrows in Figure 1). The second way is to directly learn the mapping from hazy to haze-free images (green arrow in Figure 1), which usually requires pairs of hazy and haze-free images.

3.1 Dehazing Based on the ASM

The dehazing algorithm based on deep learning and ASM needs to estimate two parameters: (1) Trans $t(x)$, (2) atmospheric light value A. A variety of deep learning strategies have been successfully applied to the process of parameter estimation, such as multi-scale [Ren *et al.*, 2016], pyramid network [Zhang and Patel, 2018], monocular estimation [Pang *et al.*, 2020]. According to the different types of input data, we divide the supervised model based on ASM into monocular dehazing and binocular dehazing.

Monocular Dehazing

For simplicity, we further divide the monocular dehazing model into three types.

The **first category** uses a deep neural network to estimate the $t(x)$ and uses prior assumptions to estimate A. Strictly, this kind of method is not an end-to-end dehazing model. Taking DehazeNet [Cai *et al.*, 2016a] as an example, the L2 loss is used to estimate the Trans as following:

$$L(t_i(x), t_i^*(x)) = \frac{1}{m} \sum_{i=1}^{m} ||t_i(x) - t_i^*(x)||^2, \quad (6)$$

where m is the number of data in the training set. Another important parameter A in the ASM is estimated using 0.1% darkest pixels of $t(x)$. Then, the final dehazed result $J(x)$ is jointly determined by $t(x)$ and A:

$$J(x) = \frac{I(x) - A}{max\{0.1, t(x)\}} + A. \quad (7)$$

[Ren *et al.*, 2020] also uses a trainable end-to-end approach to design a dehazing network via multi-scale convolutional neural networks with holistic edges. Compared with non-data-driven estimation methods [Zhu *et al.*, 2015; He *et al.*, 2010], these CNN based methods can obtain sharp-edged Trans.

The **second category** uses multiple neural networks to jointly estimate the $t(x)$ and A. Densely connected pyramid dehazing network (DCPDN) [Zhang and Patel, 2018] was proposed to jointly learn the $t(x)$ and A. In particular, DCPDN used a U-shape network to predict a 2D atmospheric light map with the same size as the image to be processed. The optimization process of joint discriminator is as following:

$$\min_{G_t, G_d} \max_{D_j} E_{I \sim p_{data(I)}}[log(1 - D_j(G_t(I)))]+$$
$$E_{I \sim P_{data(I)}}[log(1 - D_j(G_d(I)))]+ \quad (8)$$
$$E_{t, J \sim p_{data(t, J)}}[log D_j(t, J)].$$

DCPDN is an end-to-end physical dehazing model by fully embedding the ASM into the overall optimization framework.

The **third category** is to re-formulate the ASM, so as to avoid the cumulative error caused by the estimation of $t(x)$ and A. AOD-Net [Li *et al.*, 2017a] integrates the process of parameter estimation into a simple framework through the re-formulated ASM. According to the original ASM that shows in (1), the dehazing image $J(x)$ is

$$J(x) = \frac{1}{t(x)}I(x) - A\frac{1}{t(x)} + A. \quad (9)$$

In order to avoid the estimation of $t(x)$ in the ASM, AOD-Net proposed the $K(x)$, an intermediate parameter without physical meaning:

$$K(x) = \frac{\frac{1}{t(x)}(I(x) - A) + (A - b)}{I(x) - 1}. \quad (10)$$

Next, the acquisition of the clear image $J(x)$ only requires a single parameter $K(x)$:

$$J(x) = K(x)I(x) - K(x) + 1. \quad (11)$$

The experiment of AOD-Net has proved that reliable dehazing performance can still be obtained without separate estimation of $t(x)$ and A. Similar ideas are adopted by FAMED-Net [Zhu *et al.*, 2018; Zhang and Tao, 2020; Dong and Pan, 2020]. Since the collection of the $t(x)$ of an image is a very time-consuming and expensive work, exploring new ways of using ASM is a hot topic that can effectively simplify the dehazing process.

Binocular Dehazing

The dehazing methods based on binocular images usually use left and right views as joint input, rather than a monocular image. [Song *et al.*, 2019] proposed a multi-task network (SSMD-Net) that can join estimate clear images and disparity with the help of stereo hazy image pairs. SSMD-Net has proved that stereo matching and ID tasks can be synergistically formulated by incorporating the depth from $t(x)$ of the scene into the matching task. The limitation of

SSID-Net is that it can only estimate the left haze-free images. Different from SSMD-Net, the recent proposed Bid-Net [Pang *et al.*, 2020] can produce binocular dehazed image pairs. When estimating the stereo Trans, BidNet designed the Stereo Transformation Module (STM) to transform the depth information. For left input $F_l \in R_{B \times C \times H \times W}$ and right input $F_r \in R_{B \times C \times H \times W}$, 1×1 convolutions (W_θ^l, W_ψ^r and W_γ^r) are used to get corresponding embeddings. The loss function of the Trans estimation process is composed of the left and right views:

$$L_t = ||\hat{t}_l - t_l||_2^2 + ||\hat{t}_r - t_r||_2^2 + ||\hat{t}_{rl} - t_l||_2^2 + ||\hat{t}_{rr} - t_r||_2^2, \quad (12)$$

where $\hat{t}_l / \hat{t}_r$ denotes the estimated left/right Trans, $\hat{t}_{rl} / \hat{t}_{rr}$ means the estimated refined left/right Trans, and t_l / t_r stands for the ground truth left/right Trans. [Song *et al.*, 2020] combined the attention mechanism with binocular tasks and adapted the model to real-world images through self-supervised distillation. Although the binocular dehazing method has reliable performances, it further enhances the dependence of the dehazing model on datasets.

3.2 Dehazing without the ASM

Some works focus on the design of dehazing models that do not rely on the estimation of $t(x)$ and A in the ASM. For the direct learning of clear images, researchers have explored a variety of effective deep learning methods, such as attention mechanism [Qin *et al.*, 2020], knowledge distillation [Hong *et al.*, 2020], domain adaptation [Shao *et al.*, 2020]. Here, we introduce their core ideas and contributions.

Multi-input Fusion
Without using the atmospheric scattering model, GFN [Ren *et al.*, 2018] found two phenomena that affect the dehazing process: (i) the color of the image is sometimes distorted under the influence of atmospheric light, (ii) the visibility of the scene is reduced due to the scattering and attenuation. Based on these two observations, GFN proposed the idea of fusing three inputs: white balanced I_{wb}, contrast enhanced I_{ce}, and gamma corrected I_{gc}. The clear image J is determined by the combination of three inputs:

$$J = C_{wb} \circ I_{wb} + C_{ce} \circ I_{ce} + C_{gc} \circ I_{gc}, \quad (13)$$

where C_{wb}, C_{ce}, and C_{gc} denote the confidence maps for gating the three inputs, and $\circ$ means element-wise multiplication. The strategy of multi-input fusion provides a new idea of using hazy inputs, which is different from data enhancement (rotation, cropping, etc.).

Attention for Dehazing
Through theoretical analysis, the FFA-Net [Qin *et al.*, 2020] proposed that channel-wise and pixel-wise features should not be treated equally when using CNN for feature extraction. Based on this observation, FFA-Net designed a feature attention (FA) module that includes the channel attention and pixel attention. First, global average pooling H_p is used to obtain global spatial information:

$$g_c = H_p(F_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i, j), \quad (14)$$

where $X_c(i, j)$ stands for the c-th channel at position (i, j). Then, the channel attention output CA_c is obtained as following:

$$CA_c = \delta(Conv(\rho(Conv(g_c)))), \quad (15)$$

where δ and ρ represent sigmoid and ReLU, respectively. The purpose of pixel attention PA is to pay more attention to areas such as thick-hazed pixels and high-frequency:

$$PA = \delta(Conv(\rho(Conv(F^*)))). \quad (16)$$

The Grid-DehazeNet [Liu *et al.*, 2019] contains three sub-processes: pre-processing, backbone, and post-processing. The backbone module uses attention-based multi-scale estimation to learn features of the output of pre-processing. FFA-Net and Grid-DehazeNet prove that the attention mechanism has a reliable ability in the process of feature extraction and image restoration, which brings an effective strategy for dehazing research that does not use the ASM.

Generative Adversarial Networks (GAN) for Dehazing
The recent proposed enhanced pix2pix dehazing network (EPDN) [Qu *et al.*, 2019] is based on the visual perception global-first theory. It contains three parts: the multi-resolution generator, the multi-scale discriminator, and the enhancer. Different from the basic GAN [Goodfellow *et al.*, 2014], the adversarial loss of EPDN includes the optimization process of two discriminators.

$$L_A = \min_{\widetilde{G}} [\max_{D_1, D_2} \sum_{k=1,2} \ell_A(\widetilde{G}, D_k)]. \quad (17)$$

[Li *et al.*, 2018b] also utilized the high-quality image generation capabilities of GAN. However, the training process of GAN may be unstable, and the adversarial loss is not easy to converge. Regarding the direct application of GAN in the field of image dehazing, more research is needed.

Knowledge Distillation for Dehazing
Knowledge distillation technology has been widely used in the research of classification and segmentation and it has been proven that knowledge transfer can be achieved. However, the teacher and student network of classification and segmentation often deal with the same task, which is not suitable for dehazing. The KDDN [Hong *et al.*, 2020] set the teacher network (T) as the image reconstruction task, and made the student network (S) imitate this process. Combined with other solutions, KDDN bridges the gap between image dehazing and knowledge distillation. The unsupervised loss is used to constrain the features of each layer in the teacher network for the haze-free image:

$$L_T = ||J - T(J)||_1. \quad (18)$$

With the aim of domain transfer, KDDN combines three loss functions to train the S-network:

$$L_S = L_r + \lambda_p L_p + \lambda_{rm} L_{rm}, \quad (19)$$

where L_r and L_p denote the reconstruction loss and the perceptual loss, respectively. L_{rm} is the representation mimicking loss which is designed for mimicking the representation of the haze-free domain. λ_p and λ_{rm} are the weight factors. KDDN explored the process of migrating knowledge from T-network to S-network for dehazing.
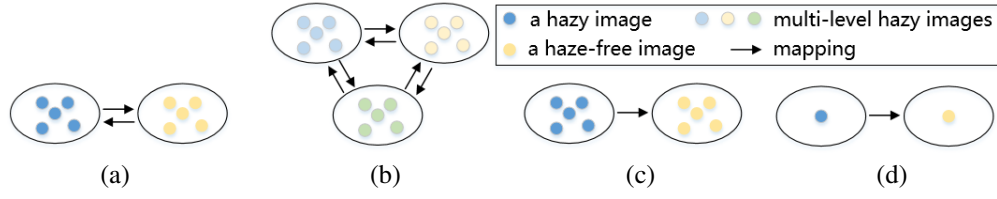
Figure 2: Four representative unsupervised dehazing models. (a) learn the mapping between hazy and haze-free domains [Engin *et al.*, 2018], (b) translation among multiple domains [Cong *et al.*, 2020], (c) estimate haze-free images from hazy images [Golts *et al.*, 2020], (d) predict haze-free image from a single hazy image [Li *et al.*, 2020a].

Domain Adaptation for Dehazing

Due to the existence of domain shift, the generalization ability of the dehazing model trained on synthetic data cannot be guaranteed when it is applied to real world hazy images. The DAID [Shao *et al.*, 2020] proposed a domain adaptation method, which was realized by the joint training of an image translation module and two image dehazing modules. The image translation modules $G_{S \to R}$ and $G_{R \to S}$ are responsible for the translation between the synthetic domain and the real domain. With the help of an image-level discriminator D_R^{img} and a feature-level discriminator D_R^{feat}, the adversarial loss of the conversion process is calculated as following:

$$L_{Gan}^{img}(X_R, (X_S, D_S), D_R^{img}, G_{S \to R})$$
$$= E_{x_s \frown X_S, d_s \frown D_S}[D_R^{img}(G_{S \to R}(x_s, d_s))] \quad (20)$$
$$+ E_{x_r \frown X_R}[D_R^{img}(x_r) - 1].$$

The domain adaptation method designed by DAID makes full use of the synthetic data for the dehazing task. However, the training of the dehazing network is a combination of unsupervised and supervised processes. The domain adaptation dehazing with the completely unsupervised training process is a topic worthy of further research.

4 Unsupervised Dehazing Algorithms

Although supervised dehazing algorithms have achieved excellent performances, they are highly dependent on paired data. In outdoor scenes that include grass, sky or water, it is difficult to ensure the correspondence between the content of the scene with and without haze. Therefore, the unsupervised dehazing methods are more suitable for real-world dehazing task, which can effectively reduce the cost of data acquisition. In this section, we introduce four types of representative unsupervised dehazing models as shown in Figure 2.

4.1 Two Domains

In order to avoid the usage of paired data, the Cycle-Dehaze network [Engin *et al.*, 2018] was proposed to learn the bidirectional mapping functions between the hazy domain and haze-free domain. The generator G is responsible for learning the mapping from hazy images to haze-free images, and the training purpose of generator F is to obtain the mapping in the opposite direction. Randomly sampling x and y from the hazy image domain X and the haze-free image domain

Y, the Cycle-Dehaze uses the (denoted as $\psi(\cdot)$) to calculate cyclic perceptual consistency loss:

$$L_{cyc-p} = ||\psi(x) - \psi(F(G(x)))||_2^2 + \qquad (21)$$
$$||\psi(y) - \psi(G(F(y)))||_2^2.$$

The overall loss function of Cycle-Dehaze is composed of cyclic perceptual consistency loss L_{cyc-p} and Cycle-GAN's [Zhu *et al.*, 2017] loss function $L_{cyc-GAN}$. Cycle-Dehaze provides a bridge between ID task and image-to-image translation within two domains as shown in Figure 2 (a). Similar methods [Huang *et al.*, 2019; Liu *et al.*, 2020] based on CycleGAN have been proposed, which adopted a CNN-based atmospheric degradation model to synthesize hazy images. In addition, DDN [Yang *et al.*, 2018] used physical-model-based disentanglement and reconstruction mechanism for unsupervised training. First, the distorted image is disentangled into three factors: the clear image, $t(x)$, and A, by three generators. Then, the three factors are combined to reconstruct the hazy input by the ASM.

4.2 Multiple Domains

Most deep learning dehazing methods generally treat hazy images of different levels as the same image domain, without considering the domain difference between multiple haze densities. The discrete haze level dehazing network (DHL-Dehaze) [Cong *et al.*, 2020] processed multiple haze level domains from the perspective of multi-domain image-to-image translation as shown in Figure 2 (b). First, sample $X_{ori} = \{x_{ori}^{(1)}, x_{ori}^{(2)}, \ldots, x_{ori}^{(m)}\}$ with corresponding labels $L_{ori} = \{l_{ori}^{(1)}, l_{ori}^{(2)}, \ldots, l_{ori}^{(m)}\}$. Through (22) and (23), classification loss and adversarial loss in the source domain can be obtained

$$L_{cls}^{ori} = \frac{1}{m} \sum_{i=1}^{m} -logD_{cls}(l_{ori}^{(i)} \mid x_{ori}^{(i)}), \qquad (22)$$

$$L_{dis}^{ori} = \frac{1}{m} \sum_{i=1}^{m} -logD_{dis}(x_{ori}^{(i)}). \qquad (23)$$

The generation of the target domain images require $L_{tar} = \{l_{tar}^{(1)}, l_{tar}^{(2)}, \ldots, l_{tar}^{(m)}\}$ and the images of the source domain. DHL-Dehaze designed a multi-scale generator (MST) to produce high-quality images:

$$X_{tar} = MST(X_{ori}, L_{tar}). \qquad (24)$$

Then, the classification and adversarial loss of the target domain image can be obtained by using the similar process as (22) and (23). Through the adversarial training process between the source domain and the target domain, the hazy image can achieve the translation of haze level without changing the content of the scene. Furthermore, DHL-Dehaze studied the dehazing result of the same model under different haze densities.

4.3 Unsupervised Loss

The cyclic consistency loss used in the unsupervised training of cyc-Dehaze [Engin *et al.*, 2018] and DHL-Dehaze [Cong *et al.*, 2020] allows the model to learn the mapping for ID task without using paired data. Further, Deep-DCP [Golts *et al.*, 2020] proposed to use only hazy images as input to update the parameters of the model, whose training process is to minimize the dark channel prior (DCP) [He *et al.*, 2010] energy function. The dark channel is defined by image patch $\Omega(x)$:

$$J^{dark}(x) = \min_{c \in \{r,g,b\}} (\min_{y \in \Omega(x)} (J^c(y))). \qquad (25)$$

Then, the following formula can be used to obtain the predicted Trans:

$$\widetilde{t}(x) = 1 - \omega \cdot \min_c (\min_{y \in \Omega(x)} (\frac{I^c(y)}{A^c})). \qquad (26)$$

The energy function can be expressed as following:

$$E(t_\theta, I) = \sum_{n=1}^{N} \sum_{k=1}^{K} W(I) \odot (T_\theta^I - T_\theta^J)^2 + \lambda \sum_{n=1}^{N} (t_\theta - \widetilde{t}(I))^2. \qquad (27)$$

The DCP loss designed by Deep-DCP can easily cooperate with different CNN architectures. The exploration of unsupervised loss is a valuable research direction.

4.4 Zero-shot Dehazing

Since unsupervised models always require sufficient data for training, zero-shot image dehazing (ZID) [Li *et al.*, 2020a] tries to minimize the dependence on the number of images. ZID has the characteristics of zero-shot while being unsupervised, that is, it only needs to learn and infer about a single hazy image. ZID contains three sub-networks $f_J(\cdot)$ (J-Net), $f_T(\cdot)$ (T-Net), and $f_A(\cdot)$ (A-Net), which are used to estimate the dehazed image, Trans, and atmospheric light, respectively. The overall loss function is as following:

$$L = L_{Rec} + L_A + L_J + L_{Reg}. \qquad (28)$$

The purpose of L_{Rec} loss is to disentangle the hazy image $I(x)$ through the reconstruction process. The acquisition of atmospheric light A requires disentangled atmospheric light $f_A(x)$ and Kullback-Leibler (KL) divergence. For J-Net, the unsupervised channel loss L_J is used to decompose the clear image $J(x)$. The limitation of ZID is that the inference speed for a single hazy image is relatively slow, because it requires separate training for each input.

5 Challenges and Opportunities

The current dehazing method has achieved good performances. Here, we attempt to highlight the problems that still exist and the open issues that would inspire the future research.

- **Problems with current ASM**: Although the current ASM has proven to be effective, it is only an approximation of the process of haze formation. In addition, [Ju *et al.*, 2021] discovered that ASM-based dehazing methods often lead to very dim results due to intrinsic limitation, and proposed enhanced ASM (EASM). In order to design a better dehazing method, a more complete study of the ASM is worthwhile.

- **Shift between the real domain and synthetic domain**: (i) Due to the difficulty of collecting real-world data, training models on synthetic data and effectively migrating to the real world is an important topic. (ii) It is a challenging research to explore more realistic haze synthesis algorithms. (iii) Some works (MRFID/BeDDE) have tried to construct real-world datasets containing pairs of hazy and haze-free images, but their data volume is only 200+. Large and reliable datasets with paired real-world data are in urgent need.

- **Computational efficiency and new metrics**: Since dehazing models are often used as pre-processing modules for high-level computer vision tasks, we must find a balance between the number of parameters, inference time, and quantitative performance. [Zhang and Tao, 2020] briefly discussed the computational efficiency of 14 dehazing methods. Future research can try to design a new evaluation metric that can comprehensively consider the running speed, model size, and dehazing quality.

- **External data and pre-trained models**: Many studies have shown that both supervised and unsupervised dehazing methods can improve the quality of dehazed images with the help of perceptual loss. However, there is no detailed discussion and proof of how external data and pre-trained models improve the performance of models. For example, do different types of pre-training models have different effects on performances?

- **The haze density that the dehazing algorithm can handle**: Most papers have discussed the model's effect on light and medium haze. Intuitively, the higher the density, the lower the quality of the dehazed image obtained by the dehazing model.

6 Conclusion

This paper provides a comprehensive survey of supervised and unsupervised image dehazing algorithms. We introduced the dependency relationship between the design of the dehazing model and specific data types. Finally, we pointed out the issues worthy of attention in the future of dehazing research.

Acknowledgments

References

[Parihar *et al.*, 2020] A. Parihar, Y. Gupta, Y. Singodia, V. Singh, and K. Singh. A comparative study of image dehazing algorithms. In *5th International Conference on Communication and Electronics Systems*, pages 766–771, 2020.

[Ancuti *et al.*, 2016] C. Ancuti, C. Ancuti, and V. De. D-hazy: A dataset to evaluate quantitatively dehazing algorithms. In *IEEE International Conference on Image Processing*, pages 2226–2230. IEEE, 2016.

[Ancuti *et al.*, 2018a] C. Ancuti, C. Ancuti, R. Timofte, and V. De. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 754–762, 2018.

[Ancuti *et al.*, 2018b] C. Ancuti, C. Ancuti, R. Timofte, L. Van Gool, L. Zhang, M. Yang, V. M. Patel, H. Zhang, and V. A. Sindagi. Ntire 2018 challenge on image dehazing: Methods and results. In *Computer Vision and Pattern Recognition Workshops*, pages 1004–1010, 2018.

[Ancuti *et al.*, 2018c] C. Ancuti, C. Ancuti, Radu Timofte, and Christophe De Vleeschouwer. I-haze: a dehazing benchmark with real hazy and haze-free indoor images. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 620–631. Springer, 2018.

[Ancuti *et al.*, 2019] C. O. Ancuti, C. Ancuti, M. Sbert, and R. Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *IEEE International Conference on Image Processing*, pages 1014–1018, 2019.

[Ancuti *et al.*, 2020a] C. Ancuti, C. Ancuti, and T. Radu. NH-HAZE: an image dehazing benchmark with non-homogeneous hazy and haze-free images. In *Computer Vision and Pattern Recognition Workshops*, IEEE CVPR 2020, 2020.

[Ancuti *et al.*, 2020b] C. Ancuti, C. Ancuti, F. Vasluianu, R. Timofte, J. Liu, H. Wu, Y. Xie, Y. Qu, L. Ma, Z. Huang, Q. Deng, and J. Chao. Ntire 2020 challenge on nonhomogeneous dehazing. In *Computer Vision and Pattern Recognition Workshops*, pages 2029–2044, 2020.

[Banerjee and Chaudhuri, 2020] S. Banerjee and S. Chaudhuri. Nighttime image-dehazing: A review and quantitative benchmarking. *Archives of Computational Methods in Engineering*, pages 1–33, 2020.

[Cai *et al.*, 2016a] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao. Dehazenet: An end-to-end system for single image haze removal. *TIP*, 25(11):5187–5198, 2016.

[Cai *et al.*, 2016b] B. Cai, X. Xu, and D. Tao. Real-time video dehazing based on spatio-temporal mrf. In *Pacific Rim Conference on Multimedia*, pages 315–325. Springer, 2016.

[C.Ancuti *et al.*, 2019] C.Ancuti, C. Ancuti, R. Timofte, L. Van Gool, L. Zhang, M. Yang, T. Guo, X. Li, V. Cherukuri, V. Monga, and H. Jiang. Ntire 2019 image dehazing challenge report. In *Computer Vision and Pattern Recognition Workshops*, pages 2241–2253, 2019.

[Chen *et al.*, 2019] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua. Gated context aggregation network for image dehazing and deraining. In *WACV*, pages 1375–1383, 2019.

[Cong *et al.*, 2020] X. Cong, J. Gui, K. Miao, J. Zhang, B. Wang, and P. Chen. Discrete haze level dehazing network. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 1828–1836, 2020.

[Dong and Pan, 2020] J. Dong and J. Pan. Physics-based feature dehazing networks. In *European Conference on Computer Vision*, pages 188–204, 2020.

[Dong *et al.*, 2020a] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and M. H. Yang. Multi-scale boosted dehazing network with dense feature fusion. In *Computer Vision and Pattern Recognition*, pages 2154–2164, 2020.

[Dong *et al.*, 2020b] Y. Dong, Y. Liu, H. Zhang, S. Chen, and Y. Qiao. Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(7):10729–10736, 2020.

[Engin *et al.*, 2018] D. Engin, A. Gen, and H. Kemal. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In *Computer Vision and Pattern Recognition Workshops*, pages 825–833, 2018.

[Golts *et al.*, 2020] A. Golts, D. Freedman, and M. Elad. Unsupervised single image dehazing using dark channel prior loss. *TIP*, 29:2692–2701, 2020.

[Goodfellow *et al.*, 2014] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *Advances in Neural Information Processing Systems*, 3:2672–2680, 2014.

[He *et al.*, 2010] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2010.

[Hong *et al.*, 2020] M. Hong, Y. Xie, C. Li, and Y. Qu. Distilling image dehazing with heterogeneous task imitation. In *Computer Vision and Pattern Recognition*, pages 3462–3471, June 2020.

[Huang *et al.*, 2019] L. Huang, J. Yin, B. Chen, and S. Ye. Towards unsupervised single image dehazing with deep learning. In *IEEE International Conference on Image Processing*, pages 2741–2745, 2019.

[Ju *et al.*, 2021] M. Ju, C. Ding, W. Ren, Y. Yang, D. Zhang, and Y. Guo. Ide: Image dehazing and exposure using an enhanced atmospheric scattering model. *TIP*, 30:2180–2192, 2021.

[Justin *et al.*, 2016] J. Justin, A. Alexandre, and F. Li. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711. Springer, 2016.

[Li *et al.*, 2017a] B. Li, X. Pengand Z. Wang, and J. Xuand D. Feng. Aod-net: All-in-one dehazing network. In *IEEE International Conference on Computer Vision*, pages 4780–4788, Los Alamitos, CA, USA, oct 2017. IEEE Computer Society.

[Li *et al.*, 2017b] Yu Li, Shaodi You, Michael S. Brown, and Robby T. Tan. Haze visibility enhancement: A survey and quantitative benchmarking. *Computer Vision and Image Understanding*, 165:1–16, 2017.

[Li *et al.*, 2018a] B. Li, X. Peng, Z. Wang, and J. Xu. End-to-end united video dehazing and detection. In *AAAI conference on artificial intelligence*, pages 7016–7023, 2018.

[Li *et al.*, 2018b] R. Li, J. Pan, Z. Li, and J. Tang. Single image dehazing via conditional generative adversarial network. In *Computer Vision and Pattern Recognition*, pages 8202–8211, 2018.

[Li *et al.*, 2019] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang. Benchmarking single-image dehazing and beyond. *TIP*, 28(1):492–505, 2019.

[Li *et al.*, 2020a] B. Li, Y. Gou, J. Liu, H. Zhu, J. Zhou, and X. Peng. Zero-shot image dehazing. *TIP*, 29:8457–8466, 2020.

[Li *et al.*, 2020b] L. Li, Y. Dong, W. Ren, J. Pan, C. Gao, N. Sang, and M. Yang. Semi-supervised image dehazing. *TIP*, 29:2766–2779, 2020.

[Liu *et al.*, 2019] X. Liu, Y. Ma, Z. Shi, and J. Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *IEEE/CVF International Conference on Computer Vision*, pages 7313–7322, 2019.

[Liu *et al.*, 2020] W. Liu, X. Hou, J. Duan, and G. Qiu. End-to-end single image fog removal using enhanced cycle consistent adversarial networks. *TIP*, 29:7819–7833, 2020.

[McCartney, 1976] J. McCartney. Optics of the atmosphere: scattering by molecules and particles. *New York, John Wiley and Sons, Inc., 1976. 421 p.*, 1976.

[Narasimhan and Nayar, 2003] S. Narasimhan and S. Nayar. Contrast restoration of weather degraded images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):713–724, 2003.

[Nayar and Narasimhan, 1999] S. Nayar and S. Narasimhan. Vision in bad weather. In *IEEE International Conference on Computer Vision*, pages 820–827. IEEE, 1999.

[Pang *et al.*, 2020] Y. Pang, J. Nie, J. Xie, J. Han, and X. Li. Bidnet: Binocular image dehazing without explicit disparity estimation. In *IEEE conference on Computer Vision and Pattern Recognition*, pages 5930–5939, 2020.

[Qin *et al.*, 2020] Xu. Qin, Z. Wang, Y. Ba, X. Xie, and H. Jia. Ffanet: Feature fusion attention network for single image dehazing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(7):11908–11915, 2020.

[Qu *et al.*, 2019] Y. Qu, Y. Chen, J. Huang, and Y. Xie. Enhanced pix2pix dehazing network. In *Computer Vision and Pattern Recognition*, pages 8152–8160, 2019.

[Ren *et al.*, 2016] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M. Yang. Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision*, pages 154–169. Springer, 2016.

[Ren *et al.*, 2018] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M. Yang. Gated fusion network for single image dehazing. In *Computer Vision and Pattern Recognition*, pages 3253–3261, 2018.

[Ren *et al.*, 2020] W. Ren, J. Pan, H. Zhang, X. Cao, and M. Yang. Single image dehazing via multi-scale convolutional neural networks with holistic edges. *International Journal of Computer Vision*, 128:240–259, 2020.

[Scharstein *et al.*, 2014] D. Scharstein, H. Hirschmuller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *Pattern Recognition*, pages 31–42, 2014.

[Shao *et al.*, 2020] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang. Domain adaptation for image dehazing. In *Computer Vision and Pattern Recognition*, pages 2805–2814, 2020.

[Sharma *et al.*, 2005] G. Sharma, W. Wu, and E. Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application: Endorsed by Inter-Society Color Council, etc*, 30(1):21–30, 2005.

[Silberman *et al.*, 2012] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from rgbd images. In *European Conference on Computer Vision*, pages 746–760. Springer, 2012.

[Singh and Kumar, 2018] D. Singh and V. Kumar. A comprehensive review of computational dehazing techniques. *Archives of Computational Methods in Engineering*, 26(4):1395–1413, 2018.

[Song *et al.*, 2019] T. Song, Y. Kim, C. Oh, and K. Sohn. Deep network for simultaneous stereo matching and dehazing. In *29th British Machine Vision Conference*, pages 1–12, 2019.

[Song *et al.*, 2020] T. Song, Y. Kim, C. Oh, H. Jang, N. Ha, and K. Sohn. Simultaneous deep stereo matching and dehazing with feature attention. *International Journal of Computer Vision*, pages 1–19, 2020.

[Wang *et al.*, 2004] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *TIP*, 13(4):600–612, 2004.

[Xu *et al.*, 2016] Y. Xu, J. Wen, L. Fei, and Z. Zhang. Review of video and image defogging algorithms and related studies on image restoration and enhancement. *IEEE Access*, 4:165–188, 2016.

[Yang *et al.*, 2018] X. Yang, Z. Xu, and J. Luo. Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In *AAAI conference on artificial intelligence*, pages 7485–7492, 2018.

[Zhang and Patel, 2018] H. Zhang and V. M. Patel. Densely connected pyramid dehazing network. In *Computer Vision and Pattern Recognition*, pages 3194–3203, 2018.

[Zhang and Tao, 2020] J. Zhang and D. Tao. Famed-net: A fast and accurate multi-scale end-to-end dehazing network. *TIP*, 29:72–84, 2020.

[Zhang *et al.*, 2017] Y. Zhang, Li. Ding, and G. Sharma. Hazerd: An outdoor scene dataset and benchmark for single image dehazing. In *IEEE International Conference on Image Processing*, pages 3205–3209, 09 2017.

[Zhang *et al.*, 2018] S. Zhang, W. Ren, and J. Yao. Feed-net: Fully end-to-end dehazing. In *IEEE International Conference on Multimedia and Expo*, pages 1–6, 2018.

[Zhang *et al.*, 2020] J. Zhang, Y. Cao, Z. Zha, and D. Tao. Nighttime dehazing with a synthetic benchmark. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 2355–2363, 2020.

[Zhao *et al.*, 2020] S. Zhao, L. Zhang, S. Huang, Y. Shen, and S. Zhao. Dehazing evaluation: Real-world benchmark datasets, criteria, and baselines. *TIP*, 29:6947–6962, 2020.

[Zhu *et al.*, 2015] Q. Zhu, J. Mai, and L. Shao. A fast single image haze removal algorithm using color attenuation prior. *TIP*, 24(11):3522–3533, 2015.

[Zhu *et al.*, 2017] J. Zhu, T. Park, P. Isola, and A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *International Conference on Computer Vision*, pages 2223–2232, 2017.

[Zhu *et al.*, 2018] H. Zhu, X. Pengand V. Chandrasekhar, L. Li, and J. Lim. Dehazegan: When image dehazing meets differential programming. In *International Joint Conference on Artificial Intelligence*, pages 1234–1240, 2018.