A Comprehensive Survey on Image Dehazing Based on Deep Learning

Jie Gui, Xiaofeng Cong, Yuan Cao, Wenqi Ren, Jun Zhang, Jing Zhang and Dacheng Tao

Southeast University, AnHui University, Ocean University of China, Institute of Information Engineering, Chinese Academy of Sciences, University of Sydney, JD Explore Academy, JD.com, China

guijie@seu.edu.cn, cxf_svip@163.com, cy8661@ouc.edu.cn, renwenqi@iae.ac.cn, wwwzhangju@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.

Based on these 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) [McCurney, 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:

\[ I(x) = J(x)t(x) + A(1-t(x)), \]  

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:

\[ t(x) = e^{-\beta d(x)}, \]  

where \( \beta \) 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

\[ J(x) = \frac{I(x) - A(1-t(x))}{t(x)}. \]  

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 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.

### Table 2: Datasets for ID tasks.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Type</th>
<th>Number</th>
<th>I/O</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-Hazy</td>
<td>Syn</td>
<td>1400+</td>
<td>I</td>
</tr>
<tr>
<td>HazeRD</td>
<td>Syn</td>
<td>15</td>
<td>O</td>
</tr>
<tr>
<td>I-HAZE</td>
<td>HG</td>
<td>35</td>
<td>I</td>
</tr>
<tr>
<td>O-HAZE</td>
<td>HG</td>
<td>45</td>
<td>O</td>
</tr>
<tr>
<td>RESIDE</td>
<td>S&amp;R</td>
<td>10000+</td>
<td>I&amp;O</td>
</tr>
<tr>
<td>Dense-HAZE</td>
<td>HG</td>
<td>33</td>
<td>O</td>
</tr>
<tr>
<td>NH-HAZE</td>
<td>HG</td>
<td>55</td>
<td>O</td>
</tr>
<tr>
<td>MRVID</td>
<td>Real</td>
<td>200</td>
<td>O</td>
</tr>
<tr>
<td>BeDDE</td>
<td>Real</td>
<td>200+</td>
<td>O</td>
</tr>
</tbody>
</table>


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]:

\[ L_{SSIM}(P) = \frac{1}{N} \sum_{p \in P} (1 - SSIM(p)). \]  

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 \( \psi(\cdot) \):

\[ L_{Per}(x, y) = \sum_{i=1}^{n} ||\psi(x) - \psi(y)||^2. \] ;
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., 2016] as an example, the L2 loss is used to estimate the Trans as following:

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

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 $A$:

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

[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 uses 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:

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

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 reformulated ASM. According to the original ASM that shows in (1), the dehazing image $J(x)$ is

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

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

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

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

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

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 BidNet [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 \( F_r \in R_{B \times C \times H \times W} \), \( 1 \times 1 \) convolutions \((W_g^r, W_p^r, W_r^r)\) are used to get corresponding embeddings. The loss function of the Trans estimation process is composed of the left and right views:

\[
L_t = ||\hat{t}_l-t_l||_2 + ||\hat{t}_r-t_r||_2 + ||\hat{t}_l-t_r||_2 + ||\hat{t}_r-t_l||_2,
\]

where \( \hat{t}_l/\hat{t}_r \) denotes the estimated left/right Trans, \( \hat{t}_l/\hat{t}_r \) 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:

\[
J = C_{wb} \circ I_{wb} + C_{ce} \circ I_{ce} + C_{gc} \circ I_{gc},
\]

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:

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

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:

\[
CA_c = \delta(Conv(\rho(Conv(g_c)))),
\]

where \( \delta \) and \( \rho \) 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:

\[
PA = \delta(Conv(\rho(Conv(F^c)))).
\]

The Grid-DehazeNet [Li 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. FNA-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:

\[
L_A = \min_G \max_D \mathbb{E}_{(G, D_k)} \left[ \ell_A(G, D_k) \right].
\]

[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:

\[
L_T = ||J - T(J)||_1.
\]

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

\[
L_S = L_r + \lambda_p L_p + \lambda_{rm} L_{rm},
\]

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. \( \lambda_p \) and \( \lambda_{rm} \) are the weight factors. KDDN explored the process of migrating knowledge from T-network to S-network for dehazing.
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 \( G_{S \rightarrow R} \) and \( G_{R \rightarrow S} \) 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 follows:

\[
\begin{align*}
L_{\text{img}}^{G_{\text{ori}}} (X_R, \{X_S, D_S\}, D_R^{img}, G_{S \rightarrow R}) &= E_{x_s \sim X_S, d_s \sim D_s} [D_R^{img}(G_{S \rightarrow R}(x_s, d_s))] \\
&+ E_{x_r \sim X_R} [D_R^{img}(x_r) - 1].
\end{align*}
\]

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:

\[
L_{\text{cycle-p}} = ||\psi(x) - \psi(F(G(x)))||_2^2 + ||\psi(y) - \psi(G(F(y)))||_2^2.
\]

The overall loss function of Cycle-Dehaze is composed of cyclic perceptual consistency loss \( L_{\text{cycle-p}} \) and CycleGAN’s [Zhu et al., 2017] loss function \( L_{\text{cycle-GAN}} \). 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 disentangle 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 \( X_{\text{ori}} = \{x_{\text{ori}}^{(1)}, x_{\text{ori}}^{(2)}, \ldots, x_{\text{ori}}^{(m)}\} \) with corresponding labels \( L_{\text{ori}} = \{l_{\text{ori}}^{(1)}, l_{\text{ori}}^{(2)}, \ldots, l_{\text{ori}}^{(m)}\} \). Through (22) and (23), classification loss and adversarial loss in the source domain can be obtained

\[
L_{\text{cls}}^{\text{ori}} = \frac{1}{m} \sum_{i=1}^{m} -log D_{\text{cls}}(l_{\text{ori}}^{(i)} | x_{\text{ori}}^{(i)}),
\]

\[
L_{\text{dis}}^{\text{ori}} = \frac{1}{m} \sum_{i=1}^{m} -log D_{\text{dis}}(x_{\text{ori}}^{(i)}).
\]

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

\[
X_{\text{tar}} = \text{MST}(X_{\text{ori}}, L_{\text{tar}}).
\]
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 \( \Omega(x) \):

\[
J_{dark}(x) = \min_{c \in \{r,g,b\}} \min_{y \in \Omega(x)} (J^c(y)). \tag{25}
\]

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

\[
\bar{t}(x) = 1 - \omega \cdot \min_{c} \min_{y \in \Omega(x)} \left( \frac{I^c(y)}{A^c} \right). \tag{26}
\]

The energy function can be expressed as following:

\[
E(t_\theta, I) = \sum_{n=1}^{N} \sum_{k=1}^{K} W(I) \odot (T^k_\theta - T^k_\theta)^2 + \lambda \sum_{n=1}^{N} (t_\theta - \bar{t}(I))^2. \tag{27}
\]

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 \( f_\theta(\cdot) \) (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:

\[
L = L_{Rec} + L_A + L_J + L_{Reg}. \tag{28}
\]

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

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2242021R10097 and 202113037.
References


