

Received 8 October 2023, accepted 19 November 2023, date of publication 21 November 2023,
date of current version 29 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3335618

SURVEY

Delving Deeper Into Image Dehazing: A Survey

GUOHOU LI^{1,2}, JIA LI^{1,2}, GONGCHAO CHEN^{1,2}, ZHIBIN WANG^{1,2}, SONGLIN JIN^{1,2},
CHANG DING³, AND WEIDONG ZHANG^{1,2}, (Member, IEEE)

¹School of Information Engineering, Henan Institute of Science and Technology, Xinxiang 543003, China

²Institute of Computer Applications, Henan Institute of Science and Technology, Xinxiang 543003, China

³School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin 541004, China

Corresponding author: Weidong Zhang (zwd@hist.edu.cn)

This work was supported in part by the Natural Science Foundation of Henan Province under Grant 232300420428, in part by the Key Specialized Research and Development Program of Science and Technology of Henan Province under Grant 232102210018 and Grant 232102210058, and in part by the Innovation Training Program for College Students of Henan Province under Grant 202310467015 and Grant 202310467031.

ABSTRACT Images captured under foggy or hazy weather conditions are affected by the scattering of atmospheric particles, resulting in decreased contrast and color variation, thereby limiting their practical applications. In recent years, deep learning methods showcase significant advancements in image dehazing. However, the complexity and degradation factors in hazy images challenge the generalization capacity of dehazing methods. This paper comprehensively reviews the recent developments in single-image dehazing techniques based on deep learning. From the perspectives of Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN), different models are introduced and classified into four categories: Encoder-Decoder, Multi-Module, Multi-Branch, and Dual-Generative Adversarial Networks. The robustness and effectiveness of deep learning models are analyzed by comparing their performance and model complexity on public datasets. Additionally, limitations of current benchmark datasets and evaluation metrics are identified, and unresolved issues and future research directions are discussed. Our efforts in this paper will serve as a comprehensive reference for future research and call for further development in deep learning-based image dehazing.

INDEX TERMS Image dehazing, deep learning, convolutional neural networks (CNNs), generative adversarial networks (GANs).

I. INTRODUCTION

The rapid advancement of computer vision has propelled image-dehazing technology to the forefront of research. It is applied extensively in civilian and military fields, encompassing object detection, traffic surveillance, remote sensing, and meteorological prediction. Such applications frequently necessitate the analysis of images or videos characterized by high quality. Nevertheless, unfavorable weather conditions like haze, fog, and rain can substantially deteriorate captured images. Dry particles such as dust and smoke and wet particles like water droplets and rainwater substantially contribute to this degradation process. Through scattering and absorption processes, these particles diminish the visibility, contrast, and color accuracy of the images,

consequently severely curtailing their efficacy in real-world situations. As a result, the imperative to mitigate the impact of haze on images and amplify their clarity has emerged.

Over the past decade, deep learning technology has significantly advanced and widely applied in diverse computer vision and image processing tasks [1], [2], [3], [4], [5]. These approaches significantly bolster high-level visual task performance, including tasks such as object detection [6], [7], [8], object recognition [9], [10], [11], and semantic segmentation [12], [13], [14]. Moreover, in low-level visual tasks like image super-resolution [15], [16], [17], denoising [18], [19], [20], [21], and enhancement [22], [23], [24], [25], [26], [27], deep learning has demonstrated notable performance advantages. However, even with the growing utilization of deep learning for image dehazing, the efficacy of deep learning-based methods in managing

The associate editor coordinating the review of this manuscript and approving it for publication was Yongqiang Cheng.

outdoor scenarios requires further refinement. Therefore, this paper provides a thorough overview and comparative analysis of recent advancements in deep learning-based models for image dehazing alongside a succinct summary of persisting challenges and unresolved matters. Additionally, we provide insights into possible directions for future research:

- We have comprehensively summarized the network architectures and training datasets used in deep learning-based image dehazing models. This paper provides a comprehensive analysis of such models, which could serve as a valuable resource for guiding future researchers in developing more robust and impactful deep learning models.
- We have conducted both qualitative and quantitative comparisons of deep learning-based image dehazing models. Through systematic experiments using diverse datasets, we thoroughly examined and evaluated the performance of existing deep learning models. Our investigation revealed biases in benchmark datasets, assessment metrics, and limitations of current deep models. These research findings are expected to offer valuable insights for guiding future studies in this domain.

The following is a summary of the information in this essay. The backdrop of picture dehazing is introduced in Section II, emphasizing dehazing models. Insights into current deep learning-based models and networks for picture dehazing are presented in Section III. Section IV covers evaluation metrics and datasets utilized, along with a quantitative and qualitative analysis of the experimental outcomes. Future research directions are suggested in Section V. The paper is concluded in Section VI.

II. BACKGROUND

The atmospheric scattering and dark channel prior models discussed in this section represent two frequently employed physical models in image dehazing. These models play a foundational role in data synthesis for training deep neural networks and their design. Additionally, they enhance our comprehension of haze formation by providing valuable insights into the underlying processes.

A. ATMOSPHERIC SCATTERING MODEL

Within scenes captured within scattering mediums, the incident light that effectively reaches the imaging sensor constitutes a minute fraction due to the prevalent absorption and scattering phenomena. These effects commonly lead to the creation of visually hazy images. Haze often introduces a blurred appearance to images, similar to foggy scenarios. Throughout the history of image analysis, the conventional atmospheric scattering model [28] has been the established approach for illuminating the intricacies of hazy image degradation. The atmospheric scattering model [28] can be

succinctly defined as follows:

$$\mathbf{I}_{\text{hazy}}^c(x) = \mathbf{J}_{\text{haze-free}}^c T_r(x) + A_t^c(1 - T_r(x)), \quad (1)$$

where, $\mathbf{I}_{\text{hazy}}^c$ represents the color channel, $\mathbf{I}_{\text{hazy}}^c$ corresponds to the captured hazy image, $\mathbf{J}_{\text{haze-free}}^c$ refers to the haze-free image, A_t^c represents the atmospheric light, T_r represents the transmission medium, and x denotes the pixel position. The transmission rate describes the portion of light that directly reaches the camera without scattering. The values of the transmission medium range from 0 to 1. Additionally, it is expressed as an exponential function of distance, dependent on two parameters: distance d and scattering coefficient β , as shown below:

$$T_r(x) = e^{-\beta d(x)}. \quad (2)$$

The next inversion technique can be used to get a haze-free image $\mathbf{J}_{\text{haze-free}}^c$:

$$\mathbf{J}_{\text{haze-free}}^c(x) = \frac{\mathbf{I}_{\text{hazy}}^c(x) - A_t^c}{T_r} + A_t^c, \quad (3)$$

single Image Dehazing (SID) is a complex problem since it necessitates estimating two crucial parameters: A_t^c and $\mathbf{J}_{\text{haze-free}}^c$ to produce a haze-free image. The estimate of important factors affects how well dehazing procedures work.

B. DARK CHANNEL PRIOR MODEL

Traditional atmospheric scattering models commonly employ a single global scattering coefficient to characterize atmospheric scattering. However, this approach cannot differentiate between the impacts of air scattering and depth-of-field on the image. Furthermore, it needs to capture localized fluctuations in atmospheric scattering. Consequently, the efficacy of conventional models in effectively mitigating haze, particularly within intricate scenes, may require enhancements.

To address these challenges, He et al. [29] proposed an improved atmospheric scattering model that integrates global and local scattering coefficients. The expression for this model is given by:

$$\mathbf{T}(x) = \mathbf{I}(x)t(x) + J(1 - t(x)), \quad (4)$$

where \mathbf{T} stands for the observed hazy image, \mathbf{I} for the dehazed image, J for the ambient light globally, and t for the medium transmittance.

The global scattering coefficient of this model accounts for the comprehensive atmospheric scattering impact, whereas the local scattering coefficient captures the localized scattering deviations. The local scattering coefficient is computed based on the pixel's spatial relationship with its neighboring pixels and is adjusted by a small constant to prevent division by zero. This model can replicate diverse atmospheric scattering phenomena more effectively, enhancing the dehazing outcome.

In contrast to conventional atmospheric scattering models, this approach incorporates local scattering coefficients to

provide a more nuanced representation of the intricacies of atmospheric scattering. This enables the model to account for spatial variations in atmospheric dispersion and more accurately differentiate between depth of field and the influence of atmospheric scattering on the image. Consequently, our model can adapt to the intricate variations in atmospheric distribution across different scenarios, improving image dehazing performance.

III. DEEP IMAGE DEHAZING ALGORITHM

In recent years, deep learning in image dehazing has surged due to its capacity for robust feature learning and extraction of contextual information. Two primary categories emerge within deep learning-based image dehazing: CNN-based models and GAN-based models. While GAN-based models concentrate on enhancing the perceived image quality, CNN models prioritize faithful reconstruction of the original image. However, this classification approach requires greater sophistication. Consequently, we adopt a more nuanced categorization method for network-based deep learning-based image dehazing models. In Figure 1, we depict the classification of deep networks, and in the subsequent sections, we systematically categorize each approach into distinct classes based on crucial factors accompanied by comprehensive information.

A. ENCODER-DECODER MODELS

A conventional, fully convolutional encoder-decoder architecture consists of two main components: an encoder and a decoder. The decoder employs the extracted features to reconstruct the target image, while the encoder is responsible for extracting features from the input image. Cross-layer connections are better employed to leverage multi-level feature information between the encoder and decoder. This encoder-decoder structure surpasses traditional CNN networks' feature extraction and expression capabilities, leading to improved network efficiency. As a result, it has gained broad application in various domains, including image dehazing.

In the following sections, we will provide detailed explanations of several methods: SEMI-CNN [30], EDN-GTM [31], DEA-Net [32], and MSBDN [33]. Each of these methods employs the encoder-decoder architecture to tackle the challenge of image dehazing.

1) SEMI-CNN

Li et al. [30] introduced an innovative approach within the encoder-decoder model. By exploring the correlation between synthetic and actual hazy images, SEMI [30] aims to eliminate various haze forms from images. SEMI [30] adopts a skip-connection encoder-decoder architecture. The encoder consists of three scales, each comprising three stacked residual blocks. This design resembles Nah et al.'s [34] work, wherein normalizing layers are not required for the residual blocks. Subsequently, Stride-Conv layers are used to downscale the feature maps by a ratio of 1/2.

Likewise, the decoder encompasses three scales, each containing three stacked residual blocks, and integrates Transposed-Conv layers for 2x upsampling. Non-linear ReLU layers [35] are applied following each convolutional layer, except for Conv24. Subsequently, the skip-connection feature maps are consolidated through summation. Moreover, residual learning captures the distinctions between blurred and sharp images.

Their groundbreaking research introduces a deep convolutional neural network (CNN) featuring supervised and unsupervised learning branches. These branches share network weights, where the supervised branch is trained using artificially blurred images, and the unsupervised branch is trained on actual hazy images. In the supervised branch, labeled losses, including mean square loss, perceptual loss, and adversarial loss, minimize the disparity between the predicted outputs and the actual data. In the unsupervised branch, constraints derived from apparent image attributes like the dark channel (DC) and image gradients (e.g., total variation (TV)) are introduced to prevent overfitting of the supervised branch to the training dataset. The entire network is trained on both synthetic data and real-world photographs.

2) EDN-GTM

A groundbreaking image dehazing network known as the Encoder-Decoder Network with Guided Transmission Map for Single Image Dehazing (EDN-GTM) has recently been introduced by Tran et al. [31]. In contrast to SEMI [30], EDN-GTM incorporates a guided transmission map to enhance dehazing performance significantly. Regarding network architecture, EDN-GTM draws inspiration from EPDN [36] and FD-GAN [37], both of which utilize the GAN framework for their dehazing methods. The core design of the generator network is based on U-Net [38], renowned for its effectiveness in encoder-decoder networks used in image restoration. U-Net consists of two paths: an expanding path (decoder) for feature synthesis and a contracting path (encoder) for feature extraction and analysis.

In contrast to the conventional U-Net architecture, EDN-GTM introduces three significant modifications to bolster U-Net's capabilities for dehazing: 1) incorporating a Spatial Pyramid Pooling (SPP) module within the bottleneck to augment the receptive field and extract essential contextual features; 2) substituting the ReLU activation function with the Swish activation function, given its consistent superiority over ReLU in deep networks; 3) appending a 3×3 convolutional layer before each downsampling and upsampling operation to expand the receptive field and capture more intricate high-level features from the input image. Concerning the discriminator's design, the encoder architecture of U-Net is adopted. This choice aims to equip the discriminator with equivalent aptitude in extracting and analyzing high-level features as the generator. This alignment fosters a competitive interplay between the two networks, enhancing performance.

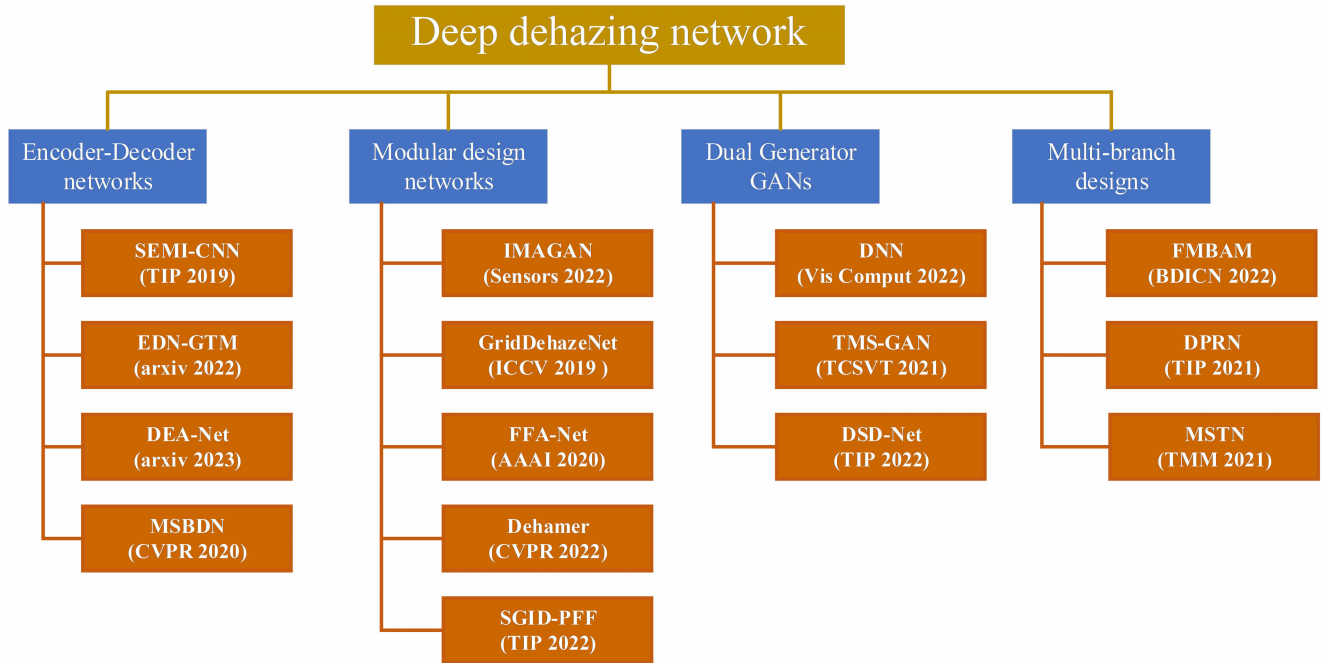


FIGURE 1. Dehazing Network Classification: Categorizing deep networks based on their essential aspects. References have been provided for each network to aid in further research.

3) DEA-NET

Recently, a novel approach known as DEA-Net (Single Image Dehazing Based on Detail-Enhanced Convolution and Content-Guided Attention) has been introduced by Chen and co-authors [32], gaining considerable attention. DEA-Net [32] employs feature modification to reconstruct high-quality haze-free images, distinguishing itself from SEMI [30] and EDN-GTM [31]. DEA-Net consists of three main components: an encoder, a feature transformation module, and a decoder. The central element of DEA-Net, the feature transformation module, employs stacked Detail Enhancement Attention Blocks (DEABs) to learn haze-free features. DEAB includes Detail Enhancement Convolution (DEConv) and Content Guidance Attention (CGA). In DEConv, they introduced differential convolution (DC) to address the dehazing problem and integrated local descriptors into regular convolutional layers. Compared to standard convolution, DEConv offers superior representational and generalization properties, and it can be seamlessly converted into standard convolution without additional parameters or processing resources. Each channel in CGA is associated with a specific Spatial Importance Map (SIM), enabling CGA to emphasize more relevant information in the encoder features. A fusion strategy based on CGA is devised to effectively merge low-level attributes from the encoder section with their corresponding high-level features.

4) MSBDN

Recently, Dong and co-authors [33] introduced a dehazing network named Multi-scale Boosted Dehazing Network

with Dense Feature Fusion (MSBDN), which has garnered significant attention. Unlike SEMI [30], EDN-GTM [31], and DEA-Net [32], MSBDN [33] adopts a U-Net architecture for dehazing purposes. The network integrates dense feature fusion techniques and operates based on enhancement and error feedback principles. The enhancement strategy [39], [40], [41], originally devised to progressively refine intermediate outcomes from previous iterations and applied in image denoising, is employed. Additionally, the error feedback mechanism, especially the backward projection technique [42], [43], [44], is utilized for super-resolution to restore missing details in the degradation process gradually.

B. MODULAR DESIGN

Modular or block-based architectures incorporate repeated structures, commonly called “blocks” or “modules,” to capture distinctive features. This architectural approach has succeeded in diverse computer vision and machine learning tasks. Illustrations of dehazing network architectures utilizing modular or block-based designs are presented below. Subsequently, various techniques are elaborated upon, encompassing IMGAN [45], GridDehazeNet [46], FFA-Net [47], Dehamer [48], and SGID-PFF [49].

1) IMGAN

Zhao et al. [45] introduced an Attention Encoder-Decoder Network with a Generative Adversarial Network (IMGAN) for remote sensing image dehazing, showcasing a design strategy centered around modularized network architectures. Within the IMGAN [45] framework, the generator network

comprises an encoder network and a decoder network. The construction of the generator network incorporates attention modules, distillation modules, CBlock modules, improvement modules, short skip connections, and local skip connections.

In the encoding phase, convolutional layers with ReLU activation functions are initially employed to extract local image features. Subsequently, an elevated level of elements is attained by integrating CBlock and attention modules. Downsampling convolutions lead to a reduction in feature map size. The initial convolutional layers and ReLU activations contribute to feature map downsizing via pooling operations. Local skip connections are introduced to integrate information and merge downscaled feature maps effectively. Moreover, the combination of feature maps is achieved by applying Cblock modules. Interestingly, introducing two extra Cblock modules and attention modules augments the acquisition of valuable supplementary features.

Attention, CBlock modules, and upsampling convolutions are employed to regenerate features during decoding. Furthermore, feature maps are subjected to upsampling using pixel shuffle. The two upsampled feature maps are concatenated (using local skip connections) through concatenation. The concatenated feature maps are subsequently merged using CBlock modules, and the features acquired at different scales are amalgamated using improvement modules. Convolutional layers, along with activation functions, are utilized to reduce the channel dimensions of the feature maps, ultimately converting them into an RGB image.

Alongside a spatial attention module, IMGAN's attention architecture incorporates a multi-scale attention module composed of three convolutional layers, a downsampling operation, and an upsampling operation. This configuration allows the network to assign weights to individual pixels, accentuating features with higher information content. By normalizing the pixels by their corresponding weights, the network can emphasize the acquisition of vital information. The multi-scale operation enables the network to encompass diverse receptive fields, enhancing attention precision. The attention module encompasses three branches: a direct path that retains the original input feature map, a secondary pathway dedicated to calculating pixel weights through channel pooling, concatenation, convolution, and sigmoid activation, and a tertiary pathway focused on extracting supplementary features. The final output feature map is synthesized by integrating, summing, and amalgamating the outputs from these three branches using a CBlock module.

2) GRIDDEHAZENET

In a recent research endeavor by Liu et al. [46], they introduced the Attention-based Multi-scale Network for Image Dehazing, or GridDehazeNet, as a comprehensive end-to-end trainable solution. This network architecture comprises three modules: preprocessing, backbone, and post-processing. This

design stands in clear contrast to the single-module approach of IMGAN [45]. In the preprocessing module, they integrate a convolutional layer without an activation function and a residual dense block (RDB) [50]. The RDB generates 16 feature maps that serve as the input for learning. This strategy aims to cultivate diverse valuable features within the learning input, effectively surmounting the limitations inherent in manual preprocessing techniques.

The backbone module utilizes attention-based multi-scale estimation to manage the learning input derived from the preprocessing module effectively. This module represents a refined version of GridNet [51], initially tailored for semantic segmentation tasks. It adeptly incorporates attention-based multi-scale estimation techniques, crucial for addressing the limitations commonly observed in conventional multi-scale strategies. The configuration of this backbone module resembles a grid network with three rows and six columns.

Each Residual Dense Block (RDB) incorporates five distinct convolutional layers. The first four layers are pivotal in increasing the number of feature maps. In contrast, the final layer employs channel attention to seamlessly integrate these feature maps with the input originating from the RDB block. The RDB block's growth rate is explicitly set to 16. While the architectural design of the blocks used for upsampling and downsampling remains consistent, utilizing distinct convolutional layers facilitates efficient alteration of the feature map dimensions.

GridDehazeNet employs the ReLU activation function for all its convolutional layers, except for the initial convolutional layer in the preprocessing module and the 1×1 convolutional layer within each RDB block. The feature maps are allocated across three scales: 16, 32, and 64, respectively. This distribution is designed to strike a balanced trade-off between output size and computational complexity. Artifacts are frequently observed in the output images from the backbone module. To mitigate this issue, a post-processing module is symmetrically integrated with the preprocessing module. This augmentation is devised to enhance the quality of the dehazed images.

3) FFA-NET

Xu et al. [47] introduced a dehazing network called the Feature Fusion Attention Network for Single Image Dehazing (FFA-Net). In contrast to IMGAN [45] and GridDehazeNet [46], FFA-Net [47] utilizes an end-to-end feature fusion attention network to restore haze-free images efficiently. FFA-Net's architecture includes several components: a convolutional layer, three group structures, a concatenate module, a CA module, a PA module, and two convolutional layers. Each group contains nineteen core block structures consisting of a convolutional layer, a ReLU layer, another convolutional layer, a CA module, and a PA module.

The FFA-Net introduces a distinctive Feature Attention (FA) module that amalgamates Channel Attention (CA) and

Pixel Attention (PA) techniques. This module effectively addresses the challenge of varying channel weights and the non-uniform haze distribution. Additionally, this module demonstrates adaptability in processing diverse input types and enhances the representational capacity of CNNs by selectively attending to particular features and pixels.

The fundamental block structure comprises Feature Attention (FA) and Local Residual Learning (LRL), that empowers the core network structure to prioritize essential information by circumventing less significant details like narrow haze regions or low frequencies via multiple local residual connections. Moreover, the Feature Fusion Attention (FFA) structure dynamically learns feature weights, prioritizing crucial attributes based on attention mechanisms. Furthermore, this structure transfers shallow-level information to more profound levels while preserving its integrity.

4) DEHAMER

Guo et al. [48] recently presented a novel dehazing network named “Image Dehazing Transformer with Transmission-Aware 3D Position Embedding” (Dehamer). In contrast to IMGAN [45], GridDehazeNet [46], and FFA-Net [47], Dehamer [48] innovatively combines CNN and Transformer [34] techniques for single-image dehazing. Dehamer introduces innovation by developing a method to modulate CNN features using a modulation matrix derived from Transformer features. This innovation combines Transformer’s global context modeling prowess with CNN’s local representation capability. Furthermore, Dehamer introduces an innovative 3D position embedding module enriched with transmission awareness, integrating haze density-related prior information into the Transformer framework.

To commence, Dehamer utilizes the transmission-aware 3D position embedding module to furnish preceding insights into haze density to the Transformer module after inputting a hazy image. Subsequently, the network independently employs the Transformer and CNN encoder modules to extract global and local features. The feature modulation module anticipates the modulation matrix (comprising coefficient and bias matrices) using Transformer features as conditional data. This modulation matrix is applied to shift and amplify suitable CNN encoder features, thus enhancing the encoder’s ability to replicate broad patterns within confined regions. Ultimately, the CNN decoder module progressively enhances resolution while adjusting for the hierarchical nature of Transformer features and CNN encoder characteristics, producing a sharp image that accentuates local details.

5) SGID-PFF

The novel strategy of Self-guided image Dehazing Using Progressive Feature Fusion (SGID-PFF), introduced by Bai et al. [49], presents a fresh perspective on image dehazing. SGID-PFF [49] leverages inherent cues in the input hazy

image to steer the dehazing procedure. This unique approach distinguishes itself from IMGAN [45], GridDehazeNet [46], FFA-Net [47], and Dehamer [48]. The structure of SGID-PFF encompasses three core components: a deep pre-dehazing module, a progressive feature fusion module, and an image restoration module. Inside the SGID-PFF framework, the deep pre-dehazing module generates a reference image from the input hazy image, showcasing evident structural disparities. The progressive feature fusion module methodically amalgamates the characteristics of this reference image with those of the hazy image, thus extracting extra guiding details. The last phase encompasses the image restoration module, which employs this guidance to reconstruct an optically lucid image.

Examining the inherent features of clear images to facilitate transmission mapping estimation underscores the importance of traditional manual prior-based methods, such as those rooted in the dark channel prior [29]. In contrast, the deep pre-dehazing technique employs the hazy image as a reference to generate an intermediate outcome with enhanced clarity, which is subsequently employed for variable estimation. They derive the reference image from the blurred image $\mathbf{J}(x)$ using the following formula:

$$\mathbf{I}(x) = \mathbf{J}(x) \frac{1}{\mathcal{N}_r(\mathbf{J}(x))} + 1 - \frac{1}{\mathcal{N}_r(\mathbf{J}(x))}, \quad (5)$$

where $\mathbf{I}(x)$ stands for the created reference image and \mathcal{N}_r for the deep pre-de hazer network, which accepts input from the hazy image $\mathbf{J}(x)$. The resulting reference image may have some residual haze since Equation (5) bases the dehazing procedure on an approximation of ambient light, but reference image $\mathbf{I}(x)$ has distinct structures. The usage of $\mathbf{I}(x)$ makes variable estimation easier and improves dehazing outcomes.

C. DUAL GENERATOR GANS

Incorporating dual-generator Generative Adversarial Networks (GANs) into image dehazing entails using multiple generators to forecast enhanced images. This strategy commonly involves employing a single discriminator with two generators or two discriminators with two generators. The core purpose is to promote the exchange of features between the generators or to feed the prediction of one generator into another. Various examples of dual-generator GANs have been developed. In the subsequent sections, we will concentrate on introducing the DDN [52], TMS-GAN [53], and DSD-Net [54] models.

1) DDN

Depth-aware image dehazing (DDN) is a dual-generator model proposed by Yang et al. [52]. Beyond its core objectives of mitigating hazy artifacts and restoring image fidelity, DDN also tackles depth map estimation. The approach in DDN [52] introduces a depth-aware mechanism for predicting depth maps, leveraging the structure of

traditional generative adversarial networks. This depth-aware technique contributes to the denoising process within a unified framework. DDN seamlessly integrates into existing CNN-based dehazing methods and effectively disentangles image content from varying degrees of haze by incorporating depth-related characteristics within the dehazing network.

The DDN architecture consists of four key components: the Depth Estimation Generator, Depth Estimation Discriminator, Dehaze Generator, and Dehaze Discriminator. The Dehaze Generator focuses on generating haze-free images, while the Depth Estimation Generator aims to predict depth maps. In order to integrate depth information into the dehazing process, the Dehaze Generator incorporates depth characteristics from the Depth Estimation Generator. The Depth Estimation Generator and the Dehaze Generator utilize the U-net structure [38], augmented with three additional refinement blocks. Discrimination between real and fake samples is facilitated by the Dehaze Discriminator and the Depth Estimation Discriminator [55]. The U-net structure comprises nine encoder layers and nine decoder levels, and each layer employs the LeakyReLU activation function.

2) TMS-GAN

A novel approach for single image dehazing, named Twofold Multi-Scale Generative Adversarial Network (TMS-GAN), was introduced by Wang et al. [53]. Unlike DDN [52], TMS-GAN employs a dual-scale productive negative network strategy for dehazing. The TMS-GAN architecture comprises two key components: a haze-generating GAN (HgGAN) and a haze-removal GAN (HrGAN). While HrGAN is designed to eliminate haze from synthesized and generated virtual data using supervised learning, HgGAN is responsible for introducing authentic haze characteristics into synthetic images.

The HgGAN generator comprises two main modules: the Multi-Attention Progressive Fusion Module (MAPFM) and the Spatial Residual Feature Aggregation Module (SRFAM). To enhance training stability, HgGAN incorporates a residual generation strategy based on a re-formulated atmospheric scattering model, drawing from the principles of stability and effectiveness found in residual learning [10]. The generation process is detailed as follows:

$$\mathbf{T}(x) = \mathbf{N}(x) - \mathbf{I}(x). \quad (6)$$

The expression for the composite variable $\mathbf{I}(x)$ is as follows:

$$\mathbf{I}(x) = \frac{(\mathbf{T}(x) - A)(1 - t(x))}{t(x)}, \quad (7)$$

in this methodology, the haze residual denoted as $\mathbf{I}(x)$ is obtained by combining the haze image, transmission map, and atmospheric light. Following this, a residual learning procedure is executed by subtracting the predicted haze residual map denoted as $\mathbf{I}_F(x)$ from the corresponding clear image represented as $\mathbf{N}(x)$. This process facilitates

the generation of both natural and synthetic blurry images denoted as $\mathbf{T}_F(x)$. The equations are as follows:

$$\mathbf{T}_F(x) = \mathbf{N}(x) - \mathbf{I}_F(x) = \mathbf{N}(x) - G_{Hg}(\mathbf{T}_S(x)), \quad (8)$$

where the symbol G_{Hg} denotes the generator of HgGAN, while symbol $\mathbf{T}_S(x)$ refers to the synthesized blurry image.

3) DSD-NET

Li et al. recently introduced a dehazing methodology titled “Dual-Scale Single Image Dehazing by Neural Augmentation” (DSD-Net) [54]. DSD-Net uniquely amalgamates model-based techniques with data-driven methods for image dehazing, setting it apart from the approaches of DDN [52] and TMS-GAN [53]. The approach begins with estimating the transmission map and atmospheric light using a model-based strategy. Subsequently, these estimations are refined using a dual-scale generative adversarial network (GAN). The DSD-Net approach stands out for its swift convergence and neural augmentation, a trait not commonly observed in conventional data-driven techniques. The reconstructed haze-free images are achieved using the Koschmieder rule, the estimated transmission map, and the ambient light.

A set of four distinct loss functions is employed to train the dual-scale GAN. Two single-scale loss functions are utilized: the gradient loss between the recovered and original images and the extremal channel loss [56]. The dual-scale L1 and adversarial loss functions [57] are also incorporated into the training process. While adversarial loss functions generally produce images with sharper details compared to L1 and L2 loss functions, utilizing a model-based approach effectively retains the visual quality characteristics of the images. Due to the restoration process involving the Koschmieder law, the PSNR and SSIM values might exhibit relatively lower values.

D. MULTI-BRANCH DESIGNS

Incorporating multi-branch designs in network architectures aims to process distinct inputs through separate branches or capture various facets of the same input across multiple hierarchical levels. Below, we will delve into the details of the FMBAM [58], DPRN [59], and MSTN [60] networks, all of which employ this strategy to enhance their performance and capabilities.

1) FMBAM

The Fusion of Multi-Branch and Attention Mechanisms (FMBAM) is a single-image dehazing network proposed by Yu et al. [58]. This end-to-end dehazing network integrates attention-based feature fusion with transfer learning within a multi-branch network architecture. The FMBAM network consists of two attention-based feature fusion networks and a transfer learning network designed to work together for image dehazing.

The FMBAM employs an Attention-based Feature Fusion (AFF) module that operates through multiple branches. This

module integrates channel and pixel attention mechanisms to extract weighted information from various channels and pixel-level details.

In the leading network, the input image is denoted as \mathbf{I}_{in} , and the transfer learning subnetwork is labeled as T . The resulting output from this subnetwork is denoted as F_1 , as calculated using Equation (9). The attention-based feature fusion subnetwork is referred to as J , and it produces outputs F_2 and F_3 through Equations (10) and (11), respectively. The equations are presented below:

$$F_1 = T(\mathbf{I}_{in}), \quad (9)$$

$$F_2 = J_1(\mathbf{I}_{in}), \quad (10)$$

$$F_3 = J_2(\mathbf{I}_{in}). \quad (11)$$

The different-level feature maps generated by the multi-branch network are merged and then fed into the tail convolutional module for reconstruction. Ultimately, a clear image \mathbf{GT}_{image} is constructed.

$$\mathbf{GT}_{image} = \text{tail}[F_1, F_1 \oplus F_2], \quad (12)$$

where $[\cdot]$ represents the concatenation operation of feature maps, and symbol \oplus denotes the addition of elements.

2) DPRN

The dual-path recurrent network (DPRN), a single-image dehazing method, was introduced by Zhang et al. [59]. In contrast to FMBAM [58], DPRN [59] tackles the dehazing challenge by simultaneously restoring the primary content and image features using a dual-path recurrence strategy. DPRN comprises four distinct blocks: an image reconstruction block, a dual-path block designed with parallel interaction capabilities, and a feature extraction block.

The dual-path block is a crucial element of DPRN, featuring two parallel branches dedicated to capturing the hazy image's fundamental content and intricate details. Each branch comprises convolutional LSTM blocks and convolutional layers. This dual-path structure facilitates the dynamic fusion of intermediate essential information and image details, promoting mutual enhancement between the two branches.

The clear image \mathbf{J} is reconstructed by the dual-path block using Equation (13), which is derived from the atmospheric scattering model:

$$\mathbf{J}[z] = \frac{\mathbf{I}[z]}{t[z]} + \frac{A[z]}{T[z]}, \quad (13)$$

there are two parts to the translucent image \mathbf{J} . The transmission map t divides the blurry image \mathbf{I} into two halves. The following equation is derived from the presumption that the parameters \mathbf{I} and t about the basic content \mathbf{J} determine the specifics of DPRN: \mathbf{J} .

$$\mathbf{J} = F_1(\mathbf{I}, t) + F_2(A, T) = J_1 + J_2, \quad (14)$$

the restoration functions $F_1(\mathbf{I}, t)$ and $F_2(A, T)$ are utilized to recover the two components presented in Equation (14).

In contrast to previous dehazing methods that treat Equation (14) as a unified restoration problem, it is observed that, in the case of a hazy image, the effects of haze on these two components, namely the fundamental content and image details, vary. Therefore, a homogeneous yet distinctive approach is employed to address these components, ensuring accurate restoration of the clear image. Additionally, the approximations of J_1 and J_2 are obtained through the use of two functions, mitigating the need for precise estimation of the transmission map and atmospheric light. Consequently, the Dual-Path Restoration Network (DPRN) decomposes the single-image dehazing task into two distinct restoration problems, denoted as $F_1(\mathbf{I}, t)$ and $F_2(A, T)$.

To determine suitable functions for constructing $F_1(\mathbf{I}, t)$ and $F_2(A, T)$, DPRN employs an Infinite Impulse Response (IIR) model, wherein J_1 is approximated as:

$$\begin{aligned} J_1[z] &\approx \frac{\mathbf{I}[z]}{t[0]} - \frac{\sum_{m=1}^M t[m] J_1[z-m]}{t[0]} \\ &\approx \frac{\mathbf{I}[z]}{t[0]} - \sum_{m=1}^M \frac{t[m]}{t[0]} \left(\frac{I[z-m]}{t[0]} - \frac{\sum_{l=1}^M t[l] J_1[z-m-l]}{t[0]} \right) \\ &\approx \frac{\mathbf{I}[z]}{t[0]} - \sum_{m=1}^M \frac{t[m] I[z-m]}{t[0]^2} \\ &\quad + \sum_{m=1, l=1}^{M, M} \frac{t[m] t[l] J_1[z-m-l]}{t[0]^2} \\ &\approx \dots, \end{aligned} \quad (15)$$

where \mathbf{I} represents the hazy image, t corresponds to the transmission map, M indicates the magnitude of t , and a similar methodology can be employed to obtain J_2 in Equation (14). Expanding on this analysis, we propose a dual-path module grounded in the Infinite Impulse Response (IIR) model, where the dual-path block comprises two parallel branches. One branch, denoted as Path J_1 , is dedicated to the recovery of J_1 , while the other branch, labeled as Path J_2 , focuses on the restoration of J_2 . Each branch leverages a recursive neural network to approximate the IIR model.

3) MSTN

Yi et al. recently introduced an innovative dehazing framework named the Efficient and Accurate Multi-scale Topological Network for Single-Image Dehazing (MSTN) [60]. Unlike the strategies employed by FMBAM [58] and DPRN [59], MSTN [60] addresses the challenge of image dehazing through the utilization of a specially designed Multi-scale Feature Fusion Module (MFFM) in conjunction with an Adaptive Feature Selection Module (AFSM). This strategic amalgamation enables the selective and effective integration of features from various scales. Consequently, MSTN excels at restoring clear information from hazy images, showcasing its distinct approach compared to prior methods such as FMBAM [58] and DPRN [59].

The MSTN employs a multi-branch network structure with i rows and j columns, where i represents the network's depth and j signifies the model's scale. Each branch is tasked with extracting visual characteristics at distinct scales. The architecture incorporates multiple MFFMs for feature selection and fusion and a residual block (RB [10]) for feature extraction. Skip connections are established between adjacent branches to account for the isolated nature of components and the absence of interaction among multi-scale features. This design enhances the network's capacity to recover image information by effectively selecting and fusing elements of varying sizes.

The model may be described as follows, where \mathbf{I}_{hazy} and $\mathbf{I}_{\text{clear}}$ represent the input hazy image and the reconstructed haze-free image:

$$\mathbf{I}_{\text{clear}} = F_{\text{MSTN}}(\mathbf{I}_{\text{hazy}}), \quad (16)$$

where $F_{\text{MSTN}}(\cdot)$ stands for MSTN. The multi-branch network (MSTN) described above has a topology of i rows and j columns, where each row denotes the network depth and each column represents a distinct model scale. $i = 0$ and $j = 0$ are the definitions of the first row and the first column, respectively. As a result, each RB's or MFFM's output $R_{i,j}$ may be characterized as:

$$R_{i,j} = \begin{cases} F_{i,j}(\mathbf{I}_{\text{hazy}}), & \text{when } i = 0, j = 0 \\ F_{i,j}(R_{i,j-1}), & \text{when } i = 0, j > 0 \\ M_{i,j}(R_{i-1,j}, R_{i-1,j+1}), & \text{when } i > 0, \end{cases} \quad (17)$$

where $F_{i,j}(\cdot)$ and $M_{i,j}(\cdot)$ stand in for the RB and MFFM operations in the i -th row and j -th column, respectively. Meanwhile, R' on the other hand, is the outcome of the downsampling procedures.

$$R' = R \downarrow_2. \quad (18)$$

IV. EXPERIMENTAL SETTINGS

A. COMMONLY USED DATASETS FOR IMAGE DEHAZING

With the growing research focus on deep learning-based image dehazing, relevant dehazing datasets are continuously emerging, significantly propelling advancements in this field. Table 1 overviews the eight most commonly used dehazing datasets, presenting their release date, data scale, and image sample generation methodology.

Examining the table reveals that, except for the “RESIDE” (REalistic Single Image DEhazing) and “MDID” (Multi-Degraded Image Dataset) datasets, the rest of the datasets are generally small in scale. For example, the “HazeRD” dataset [61] comprises merely 15 clear images and 75 synthesized images showcasing varying degrees of haze. Conversely, the “I-HAZE” dataset [62] covers 35 pairs of indoor photos, including both hazy and corresponding haze-free images. The foggy images within this dataset were captured using a specialized haze machine under authentic overcast conditions. In contrast, the “O-HAZE” dataset [63] comprises 45 pairs

of outdoor hazy and clear images, while both the “Dense-HAZE” [64] and “NH-HAZE” [65] datasets contain 55 pairs of outdoor gray and clear images each.

Due to the limitations inherent in synthesized image dehazing datasets, including inaccuracies in formation models, challenging assumptions, limited image diversity, and specificity to specific scenes, this section is dedicated to exploring real-world image dehazing datasets.

1) I-HAZE

The 35 pairs of indoor photos in the I-HAZE dataset are divided into hazy images and reference real-world clear images. This dataset uses a professional haze machine to produce real hazy photos, in contrast to the bulk of dehazing datasets that are currently accessible. Additionally, each scene includes a MacBeth color chart to make color calibration easier and enhance the evaluation of dehazing models. Notably, the clear and murky photographs are taken under carefully regulated lighting circumstances, guaranteeing consistency.

2) O-HAZE

Guangzhou University has made available the image dehazing dataset known as O-HAZE (Outdoor Hazy). This dataset contains 45 pairs of outdoor scene photos, including versions with and without haze. The haze-free photographs were taken under identical conditions as the original landscapes. At the same time, the authors created hazy images by adding fog to real-world scenes using a cold smoke machine. Based on imaging distance and depth, the O-HAZE dataset is split into four groups, including 11 pairs of photos. With a resolution of 72 pixels per inch and a size of 1080×1920 , these photos are offered in PNG format.

3) DENSE-HAZE

Dense-HAZE is an image-dehazing dataset offered by the ISA Laboratory that focuses on outdoor settings. This collection has 55 pairs of high-resolution photos, both with and without haze. The photos were taken outside in various locales, including cities, rural areas, steep terrain, and large bodies of water. Several foggy settings were created using the Schrader formula, and the hazy models were reproduced using image synthesis techniques. Each set of two images—one with haze, the other without—is 1920×1080 pixels in size and is offered in JPEG format. Annotation data from the dataset additionally contains transmission distance, image ID, and atmospheric light intensity.

4) NH-HAZE

Peking University has made available the image dehazing dataset NH-HAZE (Natural Hazy), centered on natural settings. The 55 pairs of hazy and haze-free photos in this collection were taken in various weather conditions and outdoor locations. The optical thickness of the haze layer is

TABLE 1. Dehazing datasets: including HazeRD [61], I-HAZE [62], O-HAZE [63], Dense-HAZE [64], NH-HAZE [65], RESIDE [66], MDID [67] and BeDDE [68].

Dataset	Year	Scale	Generation Method	Dataset Link
HazeRD [61]	2017	75	Model synthesis	https://labsites.rochester.edu/gsharma/research/computer-vision/hazerd/
I-HAZE [62]	2018	35	Real	https://data.vision.ee.ethz.ch/cvl/ntire18/i-haze/
O-HAZE [63]	2018	45	Real	https://data.vision.ee.ethz.ch/cvl/ntire18/o-haze/
Dense-HAZE [64]	2019	55	Real	https://data.vision.ee.ethz.ch/cvl/ntire19/dense-haze/
NH-HAZE [65]	2020	55	Real	https://data.vision.ee.ethz.ch/cvl/ntire20/nh-haze/
RESIDE [66]	2019	87125	Model synthesis/Real	https://sites.google.com/view/reside-dehaze-datasets/reside-standard?authuser=3D0
MDID [67]	2020	30346	Model synthesis	-
BeDDE [68]	2020	208	Real	https://github.com/xiaofeng94/BeDDE-for-defogging

**FIGURE 2.** Representative Images: Three sample images from the RESIDE [66], O-HAZY [63], NH-HAZE [65], I-HAZY [62], and DENSE-HAZY [64] datasets are shown to demonstrate the diversity of dehazed images.

calculated and then applied to the original photos to create the hazy images in the NH-HAZE dataset. The haze-free photos, however, match the hazy images. Each set of two images—one with haze, the other without—is 1920×1080 pixels in size and is offered in JPEG format.

5) RESIDE

Li et al. [66] introduced the RESIDE dataset in 2019. The training set of this dataset includes a total of 13,990 and 14,427 image pairs extracted from the indoor training set (ITS) and outdoor training set (OTS), respectively. Moreover, the dataset encompasses the Synthetic Objective Testing Set (SOTS), Real-world Task-driven Testing Set (RTTS), Unannotated Real-world Hazy Images (URHI), and Hybrid Subjective Testing Set (HSTS). These datasets serve as standard benchmarks for comparing quantitative and qualitative models. The URHI contains over 4,000 hazy images, whereas the SOTS comprises 500 pairs of gray images captured in indoor and outdoor settings. With 4,322 tagged hazy images meant for item recognition tasks, the RTTS offers substantial content. On the other hand, the HSTS dataset comprises ten artificially generated images and ten real photographs for subjective evaluation purposes.

6) BEDDE

Two hundred eight pairs of real photos make up the BeDDE (Benchmark Dataset for Dehazing Evaluation) dataset. A naturally blurry image and a perfectly matched, crisp reference image make up each pair. These unique images were gathered from 23 different cities. The publication of this dataset has given researchers a crucial database on which to study genuine hazy image dehazing. The scale of this dataset still has to be increased, however.

B. EVALUATION METRICS

Two types of assessment are commonly used for evaluating single-image dehazing: automatic evaluation metrics and the Human Visual System (HVS). In the context of automated evaluation, a set of ten metrics is often employed, out of which five are tailored explicitly for image dehazing: PSNR, MSE, SSIM [69], BRISQUE [70], NIQE [71], and MetalQA [72]. The most representative of the full-reference assessment measures are Peak Signal Noise Ratio (PSNR) and Structural Similarity (SSIM). Furthermore, there are additional measures designed for image dehazing, including VI [68], RI [68], and DHQI [73]. In the subsequent sections, we provide an overview of these criteria and thoroughly

examine their limitations, offering a comprehensive discussion. Additionally, we emphasize the significance of human visual judgment in this context.

C. AUTOMATIC EVALUATION METRICS

1) MSE AND PSNR

Mean Squared Error (MSE), a signal metric that measures how similar or distorted two signals are, serves as the starting point for the discussion. Comparing an original signal to a signal recovered from distortion or contamination is typical. The MSE between two signals may be written mathematically as:

$$MSE = \frac{1}{I} \sum_{i=1}^I (A_i - B_i)^2, \quad (19)$$

when A and B stand for two signals related to images, and A_i and B_i stand for the pixels at the i th position. I stands for the entire amount of pixels, similarly. Additionally, the assessment measure known as Peak Signal-to-Noise Ratio (PSNR) in the context of image processing literature is derived from MSE and may be mathematically represented as:

$$PSNR = 10 \log_{10} \frac{T^2}{MSE}, \quad (20)$$

where T stands for the image's dynamic range of pixel intensities (255). Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) use has several appealing characteristics, including (1) simplicity, (2) validity as distance measurements for all norms, (3) straightforward physical interpretation, and (4) efficacy as optimization measures. However, these measurements presuppose that the connections between the original signal, the distorted and original signals, and the error signal's signs have no bearing on the signal's fidelity. Sadly, none of these presumptions are accurate when visually evaluating how healthy images are perceived [74]. We examine different strategies to get around these restrictions in the following section.

2) SSIM

The Structural Similarity (SSIM) index, first introduced by Wang and Bovik [75] and developed in [76] and [77], is another extensively used measurement. The SSIM index considers two patches, denoted as x and y , which are extracted from different locations of the images to be compared. SSIM incorporates three components: (1) the luminance similarity $a(x, y)$, (2) the contrast similarity $b(x, y)$, and (3) the local structure similarity $c(x, y)$. As pointed out in [78], The local SSIM value is created by adding these similarities using simple statistical computations and is written as follows:

$$\begin{aligned} SSIM &= a(x, y) \cdot b(x, y) \cdot c(x, y), \\ &= \left(\frac{2\mu_x \mu_y + D_1}{\mu_x^2 \mu_y^2 + D_1} \right) \left(\frac{2\sigma_x \sigma_y + D_2}{\sigma_x^2 \sigma_y^2 + D_2} \right) \left(\frac{\sigma_{xy} + D_3}{\sigma_x + \sigma_y + D_3} \right), \end{aligned} \quad (21)$$

where μ_x and μ_y are the means of image patches x and y , while σ_x and σ_y are their standard deviations. Similarly, σ_{xy} is the cross-correlation between the mean-removed image patches. Constants D_1 , D_2 , and D_3 stabilize these terms to avoid division by values close to zero.

3) BRISQUE

BRISQUE primarily aims to extract the image's MSCN (Mean Subtracted Contrast Normalized) coefficients. The properties of the extended Gaussian distribution that are impacted by the distortions are then captured by fitting these MSCN coefficients into the distribution. In order to acquire the assessment result for image quality, Support Vector Regression (SVR) is utilized to construct the regression from the retrieved features.

4) NIQE

The quality assessment of distorted images is determined using a simple distance metric between the model statistics and the statistics of the distorted image. This metric relies on the Natural Scene Statistics (NSS) model within the spatial domain.

5) META IQA

By leveraging the concept of meta-learning, shared prior knowledge among various distortions is learned to enhance the generalization capability of quality assessment models.

6) VI AND RI

The Visibility Index (VI) and the Realness Index (RI), two full-reference metrics explicitly designed for assessing the quality of dehazed images, were introduced by Zhao et al. [68] using the collected BeDDE dataset. The VI measures image quality by evaluating the similarity between an idea and its corresponding reference. Meanwhile, the RI assesses the realism of dehazed images by measuring the similarity between the dehazed image and the connection in feature space.

7) DHQI

Three sets of essential characteristics—haze removal, structure preservation, and over-enhancement—are extracted from the dehazed results and combined for assessment.

D. HUMAN VISUAL SYSTEM

Human assessments are used to assess anticipated image quality to include perceptual measures when actual reference data is lacking. These human evaluations can be acquired through crowdsourcing or by enlisting competition experts. Nevertheless, these methods have yet to demonstrate a clear advantage over mathematical ones. The reasons listed below explain why mathematical metrics are still appealing. They are independent of various people and observational settings and are often simple to compute and cost little to compute.

Furthermore, the role of viewing conditions in influencing human perception of image quality has been recognized.

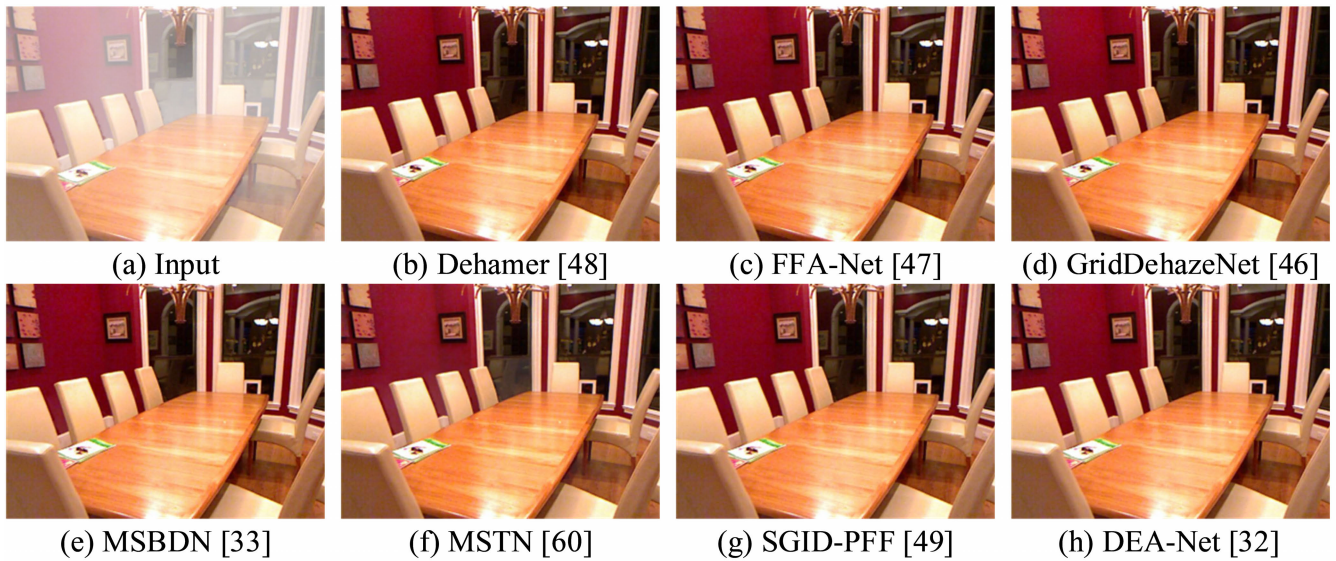


FIGURE 3. Below is a visual comparison of a synthesized hazy image from the SOTS-Indoor testing set. Please zoom in for the best view.

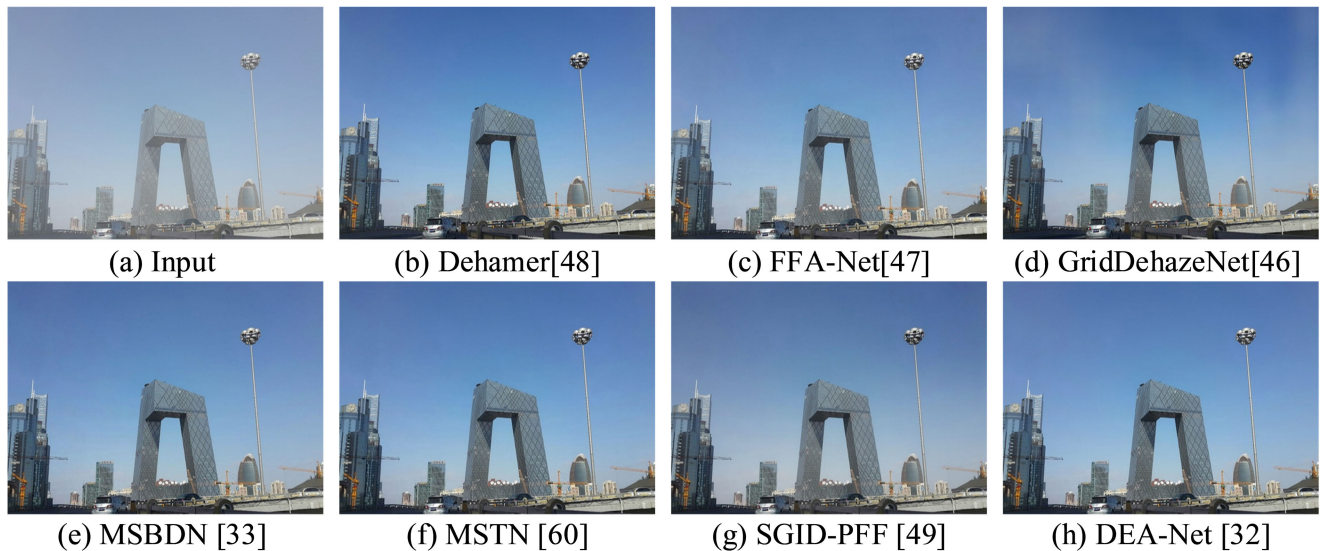


FIGURE 4. Below is a visual comparison of a synthesized hazy image from the SOTS-Outdoor testing set. Please zoom in for the best view.

However, methods that depend on specific observation conditions may produce disparate estimation results when multiple observation conditions are present, leading to inconvenience. Additionally, such methods can be user-specific, requiring individuals to calculate the observation conditions and provide outputs to the measurement system. In contrast, observation-independent methods compute a single metric, offering a general image quality assessment. Furthermore, volunteer experience has a significant influence on how people see. Volunteers with knowledge of degradation and artifact effects and the appearance when these are not adequately corrected can offer more reliable subjective ratings.

E. BENCHMARK RESULTS

In order to provide a more transparent demonstration of the performance of various methods, this section presents objective metrics and subjective visual comparisons of representative dehazing models at different stages. The SOTS subset, which comprises 500 indoor and 500 outdoor hazy photos, is the subset from which we chose the ITS and OTS subsets of the RESIDE dataset and assessed them. Table 2 compares the PSNR and SSIM measures of seven everyday deep learning dehazing models on the RESIDE dataset. The top results are shown in red. The output of each model on synthetic indoor and outdoor hazy photos is shown in Figures 3 and 4, respectively. Early

TABLE 2. Comparison of objective indicators of image dehazing representative models.

RESIDE	Methods	DEA-Net	Dehamer	GridDehazeNet	MSBDN	MSTN	SGID-PFF	FFA-Net
	code link	https://github.com/cecet3350/DEA-Net	https://github.com/Li-Chongyi/Dehamer	https://github.com/proteus1991/GridDehazeNet	https://github.com/BookerDeWitt/MSBDN-DFF	https://github.com/Joyies/MSTN	https://github.com/csbhr/SGID-PFF	https://github.com/zhilin007/FFA-Net
SOTS-I	PSNR SSIM	41.16 0.9937	36.59 0.9879	32.13 0.9831	32.61 0.9799	35.64 0.9894	38.31 0.9860	36.35 0.9884
SOTS-O	PSNR SSIM	36.14 0.9882	35.07 0.9849	30.75 0.9811	34.71 0.9849	32.91 0.9831	30.16 0.9803	33.57 0.9837

versions, such as GridDehazeNet and MSBDN, exhibit worse dehazing performance due to the imaging model's limited expressiveness. However, recent approaches like DEA-Net and Dehamer have dramatically improved objective metrics thanks to developments in deep learning techniques.

V. FUTURE AND EMERGING DIRECTIONS

Image dehazing is a well-established research area that has made significant strides recently, mainly because deep learning techniques have advanced quickly. Image dehazing has yet to be thoroughly studied in several dimensions, in contrast to other image improvement techniques like image super-resolution and underwater image enhancement; as a result, the area of image dehazing shelters many potential directions for future study. We will discuss several prospective research axes in the sections that follow.

- **“Domain Gap” Problem:** In image dehazing, the common practice is to acquire training samples through simulation, which results in a substantial “domain gap” between simulated and real-world images. Consequently, the performance of trained models degrades significantly or even becomes ineffective when applied to real-world images. Training image dehazing models using non-paired samples in a semi-supervised learning approach presents challenges such as training instability and inconsistent content. It is crucial to devise effective mechanisms and strategies to bridge this “domain gap genuinely,” representing a pivotal direction for future in-depth research in image dehazing.
- **Knowledge transfer:** Designing efficient knowledge transfer strategies to enhance the dehazing performance on real-world hazy images remains a research focus in the foreseeable future. Recently, new advancements in machine learning, such as meta-learning, domain generalization, domain adaptation, and zero-shot learning, have emerged. Applying these novel approaches to image dehazing warrants in-depth investigation and holds the potential for innovative research outcomes.
- **Model complexity:** Although some research efforts have attempted to design lightweight network architectures to meet the requirements of outdoor visual systems, the performance of these models could be improved. Exploring how to achieve a good trade-off between model processing performance and speed can be a fruitful area for further research.

- **Transformer architecture:** Transformer is currently an emerging research focus in computer vision, and various Transformer network architectures have emerged and been applied to various visual tasks such as semantic segmentation and object detection, achieving promising performance. In the future, considering the application of a Transformer in image dehazing can leverage its powerful capability to express internal image structures, thereby enhancing the dehazing performance.
- **Evaluation of dehazed image quality:** Evaluating the quality of dehazed images to effectively guide the design of dehazing models has always been a challenge in image dehazing. Some researchers have attempted to develop specific methods for evaluating the quality of dehazed images, but currently, there needs to be more authoritative and widely accepted evaluation criteria. It is worth exploring the design of evaluation criteria for dehazing models' performance in the future, considering specific downstream intelligent analysis tasks.
- **Enhancing the Effectiveness of Atmospheric Scattering Models:** The present atmospheric scattering model has demonstrated efficacy in elucidating the haze formation process in numerous dehazing methodologies. Nonetheless, inherent limitations within the atmospheric scattering model may lead to a perceptible blurring in dehazed images. Enhancements to the atmospheric scattering model are poised to exert a substantial influence on the dehazing efficacy of extant methods dependent on this model. Hence, there is merit in investigating a more precise model for the haze formation process.

VI. CONCLUSION

We conducted an exhaustive literature review on Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs) applied to image dehazing. Our review encompasses all pertinent deep learning techniques, including those accessible on arXiv. A comprehensive overview and evaluation of datasets suitable for model training and testing is also provided. We delved into the nuances and constraints of evaluation metrics, utilizing benchmark datasets for performance and visual comparisons to emphasize model robustness and complexity differences. Transitioning from CNNs to GANs, deep learning in image dehazing closely aligns with broader advancements. Prominent network topologies, such as encoder-decoder

networks and TMS-GAN, serve as foundational models with many contemporary variations. Notably, the divergence lies in training data, mainly including fuzzy photos. The development of specialized network architectures and loss functions, though addressing image dehazing, often yields unreliable and aesthetically unsatisfactory outcomes. Deep learning techniques are catching up with traditional methods, yet reliance on synthetic data limits generalization. The field exhibits significant room for improvement, indicating substantial prospects for advancing deep learning-based image dehazing.

REFERENCES

- [1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [2] P. Zhuang, "Retinex underwater image enhancement with multioriented gradient priors," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021, pp. 1709–1713.
- [3] P. Zhuang, J. Wu, F. Porikli, and C. Li, "Underwater image enhancement with hyper-Laplacian reflectance priors," *IEEE Trans. Image Process.*, vol. 31, pp. 5442–5455, 2022.
- [4] Y. Liu, Z. Yan, A. Wu, T. Ye, and Y. Li, "Nighttime image dehazing based on variational decomposition model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 639–648.
- [5] Z. Liang, W. Zhang, R. Ruan, P. Zhuang, and C. Li, "GIFM: An image restoration method with generalized image formation model for poor visible conditions," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4110616.
- [6] K. Li, L. Wu, Q. Qi, W. Liu, X. Gao, L. Zhou, and D. Song, "Beyond single reference for training: Underwater image enhancement via comparative learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 6, pp. 2561–2576, Jun. 2023.
- [7] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
- [8] Z. Zhao, Y. Zhou, B. Liu, J. He, J. Zhao, Y. Cai, J. Fan, X. Li, Z. Wang, Z. Lu, J. Wu, H. Qi, and Q. Dai, "Two-photon synthetic aperture microscopy for minimally invasive fast 3D imaging of native subcellular behaviors in deep tissue," *Cell*, vol. 186, no. 11, pp. 2475–2491, May 2023.
- [9] J. Wu, Y. Guo, C. Deng, A. Zhang, H. Qiao, Z. Lu, J. Xie, L. Fang, and Q. Dai, "An integrated imaging sensor for aberration-corrected 3D photography," *Nature*, vol. 612, no. 7938, pp. 62–71, Dec. 2022.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [11] J. Wu et al., "Iterative tomography with digital adaptive optics permits hour-long intravital observation of 3D subcellular dynamics at millisecond scale," *Cell*, vol. 184, no. 12, pp. 3318–3332, Jun. 2021.
- [12] Z. Huang, S. Yang, M. Zhou, Z. Li, Z. Gong, and Y. Chen, "Feature map distillation of thin nets for low-resolution object recognition," *IEEE Trans. Image Process.*, vol. 31, pp. 1364–1379, 2022.
- [13] X. Pan, J. Cheng, F. Hou, R. Lan, C. Lu, L. Li, Z. Feng, H. Wang, C. Liang, Z. Liu, X. Chen, C. Han, and Z. Liu, "SMILE: Cost-sensitive multi-task learning for nuclear segmentation and classification with imbalanced annotations," *Med. Image Anal.*, vol. 88, Aug. 2023, Art. no. 102867.
- [14] W. Ren, Y. Tang, Q. Sun, C. Zhao, and Q.-L. Han, "Visual semantic segmentation based on few/zero-shot learning: An overview," *IEEE/CAA J. Autom. Sinica*, early access, Mar. 28, 2023, doi: 10.1109/JAS.2023.123207.
- [15] B. Xie, S. Li, M. Li, C. H. Liu, G. Huang, and G. Wang, "SePiCo: Semantic-guided pixel contrast for domain adaptive semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 7, pp. 9004–9021, Jul. 2023.
- [16] C. Guo, C. Li, J. Guo, R. Cong, H. Fu, and P. Han, "Hierarchical features driven residual learning for depth map super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2545–2557, May 2019.
- [17] P. Wang, B. Bayram, and E. Sertel, "A comprehensive review on deep learning based remote sensing image super-resolution methods," *Earth-Sci. Rev.*, vol. 232, Sep. 2022, Art. no. 104110.
- [18] Y. Liu, Z. Yan, J. Tan, and Y. Li, "Multi-purpose oriented single nighttime image haze removal based on unified variational retinex model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 4, pp. 1643–1657, Apr. 2023.
- [19] Q. Jiang, Y. Gu, C. Li, R. Cong, and F. Shao, "Underwater image enhancement quality evaluation: Benchmark dataset and objective metric," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 5959–5974, Sep. 2022.
- [20] Y. Liu, Z. Yan, T. Ye, A. Wu, and Y. Li, "Single nighttime image dehazing based on unified variational decomposition model and multi-scale contrast enhancement," *Eng. Appl. Artif. Intell.*, vol. 116, Nov. 2022, Art. no. 105373.
- [21] X. Guo, Y. Yang, C. Wang, and J. Ma, "Image dehazing via enhancement, restoration, and fusion: A survey," *Inf. Fusion*, vols. 86–87, pp. 146–170, Oct. 2022.
- [22] X. Chen, J. Pan, K. Jiang, Y. Li, Y. Huang, C. Kong, L. Dai, and Z. Fan, "Unpaired deep image deraining using dual contrastive learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 2007–2016.
- [23] P. Zhuang and X. Ding, "Underwater image enhancement using an edge-preserving filtering retinex algorithm," *Multimedia Tools Appl.*, vol. 79, pp. 17257–17277, Jan. 2020.
- [24] Z. Liang, C. Li, S. Zhou, R. Feng, and C. C. Loy, "Iterative prompt learning for unsupervised backlit image enhancement," 2023, *arXiv:2303.17569*.
- [25] C. Li, C.-L. Guo, M. Zhou, Z. Liang, S. Zhou, R. Feng, and C. Change Loy, "Embedding Fourier for ultra-high-definition low-light image enhancement," 2023, *arXiv:2302.11831*.
- [26] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, and C. Li, "Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement," *IEEE Trans. Image Process.*, vol. 31, pp. 3997–4010, 2022.
- [27] W. Zhang, Y. Wang, and C. Li, "Underwater image enhancement by attenuated color channel correction and detail preserved contrast enhancement," *IEEE J. Ocean. Eng.*, vol. 47, no. 3, pp. 718–735, Jul. 2022.
- [28] H. Koschmieder, "Theorie der horizontalen sichtweite," *Beiträge Zur Physik Der Freien Atmosphäre*, pp. 33–53, 1924.
- [29] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
- [30] L. Li, Y. Dong, W. Ren, J. Pan, C. Gao, N. Sang, and M.-H. Yang, "Semi-supervised image dehazing," *IEEE Trans. Image Process.*, vol. 29, pp. 2766–2779, 2020.
- [31] L.-A. Tran, S. Moon, and D.-C. Park, "A novel encoder-decoder network with guided transmission map for single image dehazing," *Proc. Comput. Sci.*, vol. 204, pp. 682–689, 2022.
- [32] Z. Chen, Z. He, and Z.-M. Lu, "DEA-net: Single image dehazing based on detail-enhanced convolution and content-guided attention," 2023, *arXiv:2301.04805*.
- [33] 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 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2157–2167.
- [34] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3883–3891.
- [35] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," 2017, *arXiv:1710.05941*.
- [36] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced Pix2pix dehazing network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8160–8168.
- [37] Y. Dong, Y. Liu, H. Zhang, S. Chen, and Y. Qiao, "FD-GAN: Generative adversarial networks with fusion-discriminator for single image dehazing," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, 2020, pp. 10729–10736.
- [38] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Intervent. MICCAI 18th Int. Conf.*, Munich, Germany, Cham, Switzerland: Springer, Oct. 2015, pp. 234–241.
- [39] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 106–128, Jan. 2013.
- [40] M. Charest, M. Elad, and P. Milanfar, "A general iterative regularization framework for image denoising," in *Proc. 40th Annu. Conf. Inf. Sci. Syst.*, Mar. 2006, pp. 452–457.

- [41] Y. Romano and M. Elad, "Boosting of image denoising algorithms," *SIAM J. Imag. Sci.*, vol. 8, no. 2, pp. 1187–1219, Jan. 2015.
- [42] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP, Graph. Models Image Process.*, vol. 53, no. 3, pp. 231–239, May 1991.
- [43] S. Dai, M. Han, Y. Wu, and Y. Gong, "Bilateral back-projection for single image super resolution," in *Proc. IEEE Multimedia Expo Int. Conf.*, Jul. 2007, pp. 1039–1042.
- [44] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1664–1673.
- [45] L. Zhao, Y. Zhang, and Y. Cui, "An attention encoder-decoder network based on generative adversarial network for remote sensing image dehazing," *IEEE Sensors J.*, vol. 22, no. 11, pp. 10890–10900, Jun. 2022.
- [46] X. Liu, Y. Ma, Z. Shi, and J. Chen, "GridDehazeNet: Attention-based multi-scale network for image dehazing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7313–7322.
- [47] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature fusion attention network for single image dehazing," *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 11908–11915.
- [48] C. Guo, Q. Yan, S. Anwar, R. Cong, W. Ren, and C. Li, "Image dehazing transformer with transmission-aware 3D position embedding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5812–5820.
- [49] H. Bai, J. Pan, X. Xiang, and J. Tang, "Self-guided image dehazing using progressive feature fusion," *IEEE Trans. Image Process.*, vol. 31, pp. 1217–1229, 2022.
- [50] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.
- [51] D. Fourure, R. Emonet, E. Fromont, D. Muselet, A. Tremeau, and C. Wolf, "Residual conv-deconv grid network for semantic segmentation," 2017, *arXiv:1707.07958*.
- [52] F. Yang and Q. Zhang, "Depth aware image dehazing," *Vis. Comput.*, vol. 38, no. 5, pp. 1579–1587, May 2022.
- [53] P. Wang, H. Zhu, H. Huang, H. Zhang, and N. Wang, "TMS-GAN: A twofold multi-scale generative adversarial network for single image dehazing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2760–2772, May 2022.
- [54] Z. Li, C. Zheng, H. Shu, and S. Wu, "Dual-scale single image dehazing via neural augmentation," *IEEE Trans. Image Process.*, vol. 31, pp. 6213–6223, 2022.
- [55] F. Yang, Z. Lu, G. Qiu, J. Lin, and Q. Zhang, "Capsule based image synthesis for interior design effect rendering," in *Proc. Comput. Vis. – ACCV 14th Asian Conf. Comput. Vis.*, Perth, WA, Australia, Cham, Switzerland: Springer, Dec. 2019, pp. 183–198.
- [56] Z. Li, C. Zheng, H. Shu, and S. Wu, "Model-based single image deep dehazing," 2021, *arXiv:2111.10943*.
- [57] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.
- [58] X. Yu and H. Yu, "Single image dehazing based on the fusion of multi-branch and attention mechanism," in *Proc. Int. Conf. Big Data, Inf. Comput. Netw. (BDICN)*, Jan. 2022, pp. 675–679.
- [59] X. Zhang, R. Jiang, T. Wang, and W. Luo, "Single image dehazing via dual-path recurrent network," *IEEE Trans. Image Process.*, vol. 30, pp. 5211–5222, 2021.
- [60] Q. Yi, J. Li, F. Fang, A. Jiang, and G. Zhang, "Efficient and accurate multi-scale topological network for single image dehazing," *IEEE Trans. Multimedia*, vol. 24, pp. 3114–3128, 2022.
- [61] Y. Zhang, L. Ding, and G. Sharma, "HazeRD: An outdoor scene dataset and benchmark for single image dehazing," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3205–3209.
- [62] C. Ancuti, C. O. Ancuti, R. Timofte, and C. De Vleeschouwer, "I-haze: A dehazing benchmark with real hazy and haze-free indoor images," in *Advanced Concepts for Intelligent Vision Systems*, Poitiers, France, Cham, Switzerland: Springer, Sep. 2018, pp. 620–631.
- [63] C. O. Ancuti, C. O. Ancuti, R. Timofte, and C. De Vleeschouwer, "O-HAZE: A dehazing benchmark with real hazy and haze-free outdoor images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 754–762.
- [64] 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 *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1014–1018.
- [65] C. O. Ancuti, C. Ancuti, and R. Timofte, "NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 444–445.
- [66] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.
- [67] K. Wang, L. Zhuo, J. Li, T. Jia, and J. Zhang, "Learning an enhancement convolutional neural network for multi-degraded images," *Sens. Imag.*, vol. 21, no. 1, pp. 1–15, Dec. 2020.
- [68] S. Zhao, L. Zhang, S. Huang, Y. Shen, and S. Zhao, "Dehazing evaluation: Real-world benchmark datasets, criteria, and baselines," *IEEE Trans. Image Process.*, vol. 29, pp. 6947–6962, 2020.
- [69] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. Morgan & Claypool Publishers, 2006.
- [70] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
- [71] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2012.
- [72] H. Zhu, L. Li, J. Wu, W. Dong, and G. Shi, "MetaQA: Deep meta-learning for no-reference image quality assessment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 14143–14152.
- [73] X. Min, G. Zhai, K. Gu, X. Yang, and X. Guan, "Objective quality evaluation of dehazed images," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 2879–2892, Aug. 2019.
- [74] Q. Wu, P. Wang, C. Shen, A. Dick, and A. Van Den Hengel, "Ask me anything: Free-form visual question answering based on knowledge from external sources," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4622–4630.
- [75] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6062–6071, Dec. 2015.
- [76] X. Ye, H. Xu, X. Ji, and R. Xu, "Underwater image enhancement using stacked generative adversarial networks," in *Proc. Pacific Rim Conf. Multimedia*. Cham, Switzerland: Springer, 2018, pp. 514–524.
- [77] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3929–3938.
- [78] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.



GUOHOU LI received the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in September 2011. He is currently a Professor with the School of Information Engineering, Henan Institute of Science and Technology, Xinxiang, China. His research interests include intelligent control and signal and information processing.



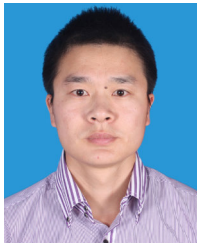
JIA LI is currently pursuing the master's degree with the School of Information Engineering, Henan Institute of Science and Technology. His research interests include image processing and deep learning.



GONGCHAO CHEN is currently pursuing the master's degree with the School of Information Engineering, Henan Institute of Science and Technology. His research interests include image processing and deep learning.



ZHIBIN WANG is currently pursuing the master's degree with the School of Information Engineering, Henan Institute of Science and Technology. His research interests include image processing and machine learning.



SONGLIN JIN received the M.S. degree in computer application technology from Central China Normal University, in 2010. He is currently a Lecturer with the Henan Institute of Science and Technology. His research interests include machine learning and hyperspectral image processing.



CHANG DING received the Ph.D. degree in information and communication engineering from Dalian Maritime University, Dalian, China, in 2018. He is currently a Lecturer and a Graduate Student's Monitor with the School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology. He has pursued his postdoctoral work with the China University of Mining and Technology. He has been a Visiting Professor with the National University of Singapore, since 2023. He has published more than 20 papers in the international and national journals. He is also a Reviewer for some international and national journals, such as *Measure Science and Technology*. His research interests include image enhancement, infrared target detection and recognition, industrial detection, and machine learning.



WEIDONG ZHANG (Member, IEEE) received the Ph.D. degree in information and communication engineering from the School of Information Science and Technology, Dalian Maritime University, Dalian, China, in June 2022. He is currently a Lecturer with the School of Information Engineering, Henan Institute of Science and Technology, Xinxiang, China. He has authored or coauthored more than 50 papers in some famous journals or conferences, including IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF OCEANIC ENGINEERING, and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and AAAI. His current research interests include image processing, computer vision, and machine learning.

...