

Semi-supervised Single-Image Dehazing Network via Disentangled Meta-Knowledge

Tongyao Jia, Jiafeng Li, *Member, IEEE*, Li Zhuo, *Member, IEEE*, Tianjian Yu

Abstract—Captured outdoor scene images are easily affected by haze. Most image dehazing methods have limited generalization capabilities for real-world hazy images owing to the complexities of real-world environments and domain gaps in the training datasets. This paper proposes a semi-supervised single-image dehazing network based on disentangled meta-knowledge. The symmetric and heterogeneous design of the disentangled network is conducive to the separation of the content and mask features of hazy images and these features are used as meta-knowledge to guide feature fusion in the dehazing network. Moreover, functions describing constant-color and disentangled-reconstruction-checking losses are designed to ensure the subjective qualities of the generated dehazed images. The results of extensive experiments conducted on synthetic datasets and real-world images indicate that the proposed algorithm outperforms state-of-the-art single-image dehazing algorithms. In addition, the algorithm effectively improves the performance of object-detection tasks. The source code is publicly available at <https://github.com/dehazing/DNDM>.

Index Terms—Single-image dehazing, disentangled representations, meta-learning, semi-supervised learning.

I. INTRODUCTION

IMAGES are one of the most important forms of information representation in multimedia, and the quality of images of outdoor scenes is often affected by weather conditions, such as rain, snow, and haze. Among these, haze severely reduces the visibility of the image, resulting in the loss of necessary information regarding the edges, colors, and details.

The purpose of single-image dehazing is to restore degraded input images obtained in hazy scenes to clean images, thus enhancing their subjective perception and improving the effectiveness of downstream tasks. Owing to the complexity and rapidly changing nature of outdoor imaging environments, single-image dehazing has received extensive attention in the fields of computer vision and graphics. Mathematically, the atmospheric scattering (AS) model [1] can be expressed as

$$I(z) = J(z)T(z) + A(z)(1 - T(z)), \quad (1)$$

where $I(z)$ is the observed hazy image at pixel z , which is considered to be generated by the reflection of a clear image

Manuscript received xxxx, xxxx. This work was supported by the Beijing Natural Science Foundation (No.L211017), General Program of Beijing Municipal Education Commission (No.KM202110005027), National Natural Science Foundation of China (No.61971016 and No.61701011), Beijing Municipal Education Commission Cooperation Beijing Natural Science Foundation (No.KZ201910005077). (Corresponding author: Jiafeng Li, Li Zhuo)

Tongyao Jia (E-mail: jty@emails.bjut.edu.cn), Jiafeng Li (E-mail: lijfeng@bjut.edu.cn) and Li Zhuo (E-mail: zhuoli@bjut.edu.cn) are with Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology, Beijing 100124, China.

Tianjian Yu (E-mail: yutianjian@csu.edu.cn) is with the School of Traffic and Transportation Engineering, Central South University, Changsha Hunan, China.

$J(z)$ and light scattered by the surrounding environment. This formula depends on several assumptions, such as the transmission map $T(z)$ and global atmospheric light $A(z)$. The atmospheric light scattering model is widely recognized and applied to image dehazing tasks.

As the atmospheric light scattering model is severely ill-conditioned, it is generally necessary to add some prior knowledge in the traditional processing task, such as the classic prior-based DCP [2] algorithm, CAP [3], NLD [4], whose generalization abilities are limited by the applicability of prior knowledge. Several researchers used polarization imaging to predict the global atmospheric light or transmission maps for image dehazing [5], [6]. However, the performance of this approach is restricted by parameter estimation accuracy.

Recently, some scholars attempted to combine the AS model with deep-learning networks, such as DehazeNet [7], AoDNet [8], and DCPDN [9]. To avoid the superposition errors of the above algorithms during model parameter estimation, some scholars directly employed end-to-end deep networks, such as EPDN [10], FFA [11], and MSBDN [12] to predict dehazed images. These deep-learning dehazing algorithms rely on large-scale supervised data for training. Although they achieve good performances on synthetic haze images, their generalization ability on real-world haze images is limited.

Several current approaches have adopted semi-supervised strategies to improve the generalization ability of networks for real-world haze images [13]–[16]. In the present study, an external meta-learning mechanism was used to effectively extract crucial dehazing information from real-world and synthetic images to improve the generalization performance of the network.

Accordingly, a semi-supervised dehazing network based on disentangled meta-knowledge (DNDM) is proposed, which learns single-image dehazing knowledge from synthetic and real-world images in training through a disentangled meta-network, and subsequently, the single-image dehazing knowledge is utilized to guide the dehazing network. The addition of meta-learning effectively improves the utilization and generalization ability of the backbone network and enhances the dehazing performance. The disentangled and reconstruction networks provide semi-supervised training support and disentangle the content and mask information of hazy images. In addition, some unique loss functions are used to obtain more realistic subjective results.

The main contributions of the present work are as follows:

1. A semi-supervised dehazing framework is proposed with a disentangled meta-network to effectively extract and fuse dehazing knowledge in synthetic and real-world datasets. In a disentangled network, the designed symmetric heterogeneous structure (SHS) of the content and mask branches effectively

improves the ability of the network to decompose haze images.

2. Meta-learning is combined with disentangled representations. The disentangle-meta-attention, which is generated by the current image, guides the feature fusion process at each receptive field level in the dehazing network. The addition of this external meta-knowledge effectively improves the utilization, robustness, and generalization ability of the proposed dehazing network.

3. To effectively enhance the details of the dehazed image and maintain its brightness, the disentangled reconstruction checking (DRC) loss L_{DRC} and constant-color loss L_{COL} are proposed. By incorporating these functions, the proposed approach outperforms state-of-the-art methods on real-world and synthetic datasets with satisfactory subjective and objective results.

The remainder of this paper is organized as follows. Section II discusses related work on single-image dehazing, disentangled representation learning, and meta-learning. Section III presents the structure of the proposed DNDM in detail. Section IV outlines the experimental setup and discusses the results of the performance analysis. Finally, Section V presents the conclusions and some possible avenues for future research.

II. RELATED WORK

This section discusses related work, including single-image dehazing, disentangled representation learning, and meta-learning.

A. Single-Image Dehazing

Single-image dehazing research is actively underway, and various methods have been proposed.

Supervised Methods: Early prior-based dehazing methods [2]–[6], [17], [18] focused on using the statistical characteristics of hazy images to estimate the transmission map and atmospheric light to restore clean images according to the degradation model given in (1). Although prior-based dehazing algorithms gave excellent dehazing results, their generalization ability remained limited because the assumed priors were invalid under certain conditions. Later, several methods incorporated deep convolutional neural networks (CNNs) to estimate the transmission map and atmospheric light [7]–[9], [19]. DehazeNet [7] removes haze directly by estimating a transmission map $t(x)$ from the hazy image using CNN. Li et al. [8] integrated the transmission map and atmospheric light into a parameter K and predicted it with a lightweight CNN for image dehazing. Zhang et al. [9] designed the DCPDN, which combines transmission maps, atmospheric light, and dehazed images to learn together by embedding AS models directly into the network. Song et al. [19] proposed a ranked CNN to obtain effective features that can be employed to train a haze density prediction model for image dehazing. However, these methods may lead to artifacts and color distortions in the output dehazed images with inaccurate estimates.

To solve this problem, some end-to-end methods [11], [12], [20]–[24] learned to directly restore clean images without estimating the transmission map and atmospheric light. Zhu et al. [20] proposed DehazeGAN, which uses an adversarial

composition network to simultaneously learn atmospheric light and transmission maps for end-to-end single-image dehazing. Qin et al. [11] developed an end-to-end feature fusion attention network that directly recovers haze-free images from hazy images. Dong et al. [12] proposed a multiscale boosted dehazing network with dense feature fusion based on the U-Net architecture. Hong et al. [21] proposed a dehazing network that distills image knowledge to perform dehazing with heterogeneous task imitation. Lin et al. [23] directly applied an attention mechanism designed a multiscale attention feature fusion network for end-to-end single-image dehazing. Tu et al. [24] designed the first MLP-based general U-Net backbone network for image dehazing by balancing the use of local and global operators. Owing to inherent limitations in collecting paired clear and hazy images, the results of existing dehazing methods may involve domain shifts for real-world scenes.

Semi-supervised and Unsupervised Methods: Recently, some dehazing methods have attempted to add unpaired real-world images to the training dataset to improve performance on real-world haze images. Li et al. [13] proposed a semi-supervised image dehazing algorithm containing both supervised and unsupervised learning branches. Shao et al. [14] developed a domain-adaptive dehazing algorithm that applies a bidirectional translation network to bridge the gap between synthetic and real-world domains. To improve the generalization performance of dehazing, Chen et al. [16] presented a principled synthetic-to-real dehazing (PSD) framework that exploits real-world hazy images to fine-tune a dehazing model backbone that is pre-trained on synthetic data. Cycle-Dehaze [25] was used for unsupervised single-image dehazing; it uses CycleGAN combined with cycle consistency loss and perceptual loss. Golts et al. [26] proposed an unsupervised dehazing method that minimizes the well-known dark channel prior energy function. Zhao et al. [27] proposed a weakly supervised refinement framework RefineDNet for single-image dehazing. In the first stage, the haze images are processed using a dark channel prior method and then refined by adversarial learning. Li et al. [28] designed a novel zero-shot image dehazing method in which the dehazed images sometimes included artifacts. However, the high computational complexity and lengthy processing times associated with these methods are significant drawbacks in real applications.

B. Disentangled Representation Learning

Motivated by the layer disentanglement, Li et al. [29] proposed a self-supervised model called YOLY for real-world single-image dehazing. The model comprises three joint sub-networks to separate the observed hazy image into a scene radiance layer, transmission map layer, and atmospheric light layer. The three layers further compose the hazy image in a self-supervised manner. Lu et al. [30] designed an unsupervised single-image deblurring method based on disentangled representations. Du et al. [31] learned invariant representations via disentangling representations and adversarial domain adaption for image denoising. Zou et al. [32] proposed a unified framework for single superimposed image separation.

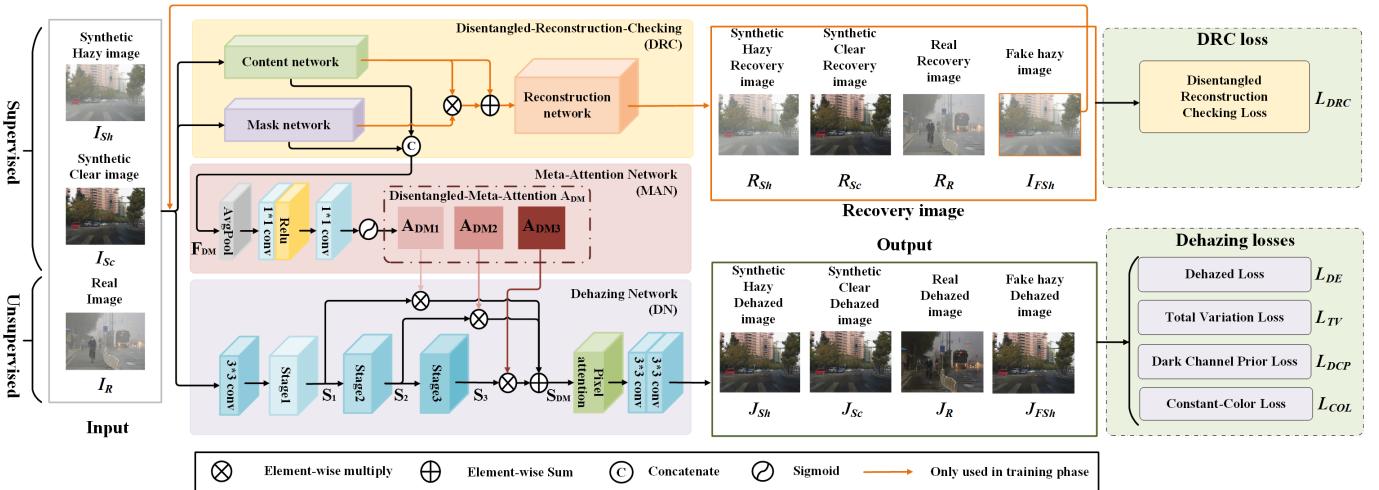


Fig. 1: Overall structure of the proposed dehazing framework. The content and mask branches of the disentangled meta-network learn meta-knowledge to generate disentangled-meta-attention to guide feature fusion in the dehazing network. The results of the disentangled meta-network and dehazing are used to calculate the DRC loss and dehazing losses, respectively.

Ye et al. [33] attempted to jointly learn real rain generation and a removal procedure within a unified disentangled image translation framework. Li et al. [34] explored an unsupervised single-image dehazing approach based on disentangled representations without paired training images. Liu et al. [15] presented a disentangled-consistency mean teacher network (DMT-Net) to boost single-image dehazing by leveraging feature-disentangled learning and unlabeled real-world images. All the above methods achieved good performances with disentangled and reconstruction networks, demonstrating the effectiveness of the networks for low-level image processing.

C. Meta-Learning

Shen et al. [35] built a neural style transfer via meta-networks to attain any style transfer. The meta-network accepts the style image and generates the corresponding image transformation network parameters through a single feed-forward propagation. Soh et al. [36] used meta-learning to find a general initial parameter suitable for internal learning that achieves considerable image enhancement results with only one gradient update. Chi et al. [37] used self-supervised meta-auxiliary learning to improve deblurring performance. In contrast to the above methods, this paper proposes a disentanglement network with meta-learning to learn the crucial dehazing knowledge in synthetic and real-world datasets to guide feature fusion of the backbone dehazing network.

III. PROPOSED DNDM SCHEME

Most deep-learning single-image dehazing algorithms use synthetic datasets for training, which seriously restricts the performances of such methods on real-world haze images. To improve the robustness of network models in practical application scenarios, the proposed DNDM is based on a combination of real-world and synthetic images for semi-supervised training. The entire network structure is shown in Fig. 1 and principally comprises three components: DRC network, meta-attention network (MAN), and dehazing network (DN).

In the testing phase, hazy images I are input into the content and mask networks for feature decoupling to obtain content and mask features, which are concatenated to F_{DM} as meta-knowledge and employed to generate disentangled-meta-attention $ADM_{1,2,3}$ through the MAN. Subsequently, the DN with disentangled-meta-attention is used to learn haze-free images J from hazy inputs I .

In the training phase, the input image group consists of a pair of synthetic images (I_{Sh} , I_{Sc}) and real-world images (I_R). As shown in Fig. 1, they are input into the DRC separately to obtain the content and mask features as the corresponding meta-knowledge F_{DM} . Moreover, the content and mask features of each image are fused and reconstructed, obtaining the recovery images R_{Sh} , R_{Sc} , R_R with reconstruction network and providing constraints; this is detailed in section III-C. In addition, we cross-reconstruct the content of synthetic clear image I_{Sc} and the mask features of the synthetic hazy image I_{Sh} into a fake hazy image I_{FSh} , which is then input into the network for decoupling, to further enhance the decoupling capability of the network. Each input image (include I_{FSh}) passes through the DN with $ADM_{1,2,3}$ to obtain the output dehazed images J_{Sh} , J_{Sc} , J_R , and J_{FSh} . Among these, $ADM_{1,2,3}$ is generated by the MAN based on its meta-knowledge F_{DM} . Furthermore, the DRC loss and the dehazing losses are calculated with the results of the DRC network and the DN, respectively, and the total loss is employed to optimize the entire network uniformly.

The following sections describe the proposed network structure and its operations in detail.

A. DRC and MAN

As shown in Fig. 2, the DRC network consists of content, mask, and reconstruction networks. The front and back ends of the content and mask network both contain a 3×3 convolutional layer, with the middle composed of eight residual network blocks. The reconstruction network, which contains eight residual blocks and four convolutional layers, fuses the

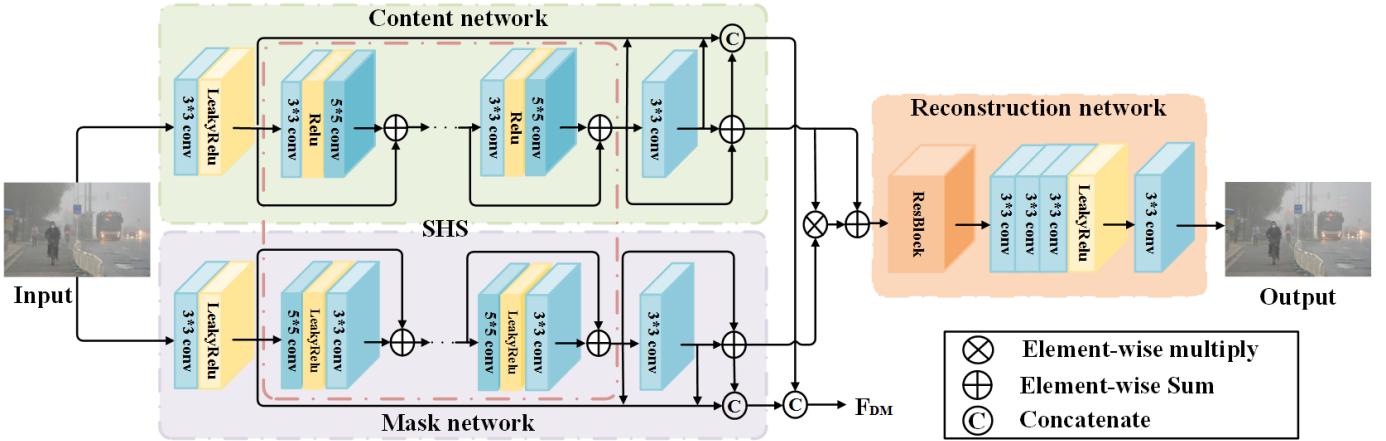


Fig. 2: Structure of the disentangled-reconstruction network. The content and mask network comprises two convolutional layers at the head and tail and eight residual blocks in the middle. The reconstruction network constitutes eight residual blocks and four layers of convolution.

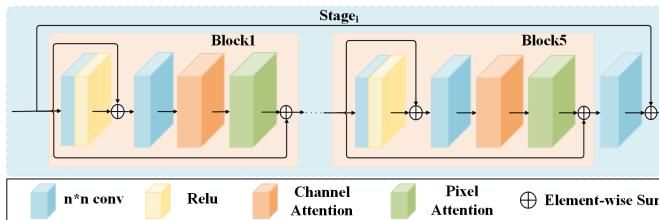


Fig. 3: Structure of $stage_i$ in the dehazing network. Each stage consists of five blocks, and the size n of each convolutional kernel in these blocks is equal to $(2i - 1) \times (2i - 1)$.

last layer features of the content and mask networks to restore the corresponding image.

The content and mask information of hazy images are anisotropic, whereas different network structures have a guiding role in extracting different attribute features and facilitating the decoupling of content and mask features. Based on this, we design a symmetric heterogeneous structure (SHS) from three aspects—the size of the convolution kernel, the activation function, and the feature fusion method—to better distinguish the content and mask information of the image. Both the content and mask branches have eight residual structures, whereas the size of the kernels in the residual network—the content network (C35) and mask network (M53)—are 3×3 and 5×5 , 5×5 and 3×3 , respectively. In addition, Relu and LeakyReLU activation functions are used for the residual network blocks of content and mask networks, respectively. C35 combined with ReLU facilitates preserving the local detail information of the image, while M53 combined with LeakyReLU can extract the degradation information from coarse to fine. Related work [38] showed that the feature representation capability and reconstruction results can be improved by adjusting the feature fusion method. Conventionally, the content and mask features are equally concatenated or multiplied for image reconstruction, which cannot distinguish between them. However, content features are more important for reconstructing images and should not be treated equally.

Thus, we expect content features to contain more information and occupy a higher weight in the reconstruction process. Therefore, we chose the product of the content and mask and the sum of the content as the input to the reconstruction network. In this manner, the SHS guides content features to contain more detailed information and have higher involvement in the reconstruction process, whereas the mask features focus more on degenerate information.

As shown in Fig. 1, the input images pass through the DRC network to obtain multi-level content and mask features. Then, the content and mask features are concatenated to F_{DM} as the input of MAN to realize the disentangled-meta-attention mechanism. The formula is as follows.

$$A_{DM} = \text{sigmoid}(\text{Conv1Relu}(\text{Conv1GAP}(F_{DM}))), \quad (2)$$

where A_{DM} represents the disentangled-meta-attention, the features are processed in MAN through a global average pooling layer, two 1×1 convolutional layers, and a sigmoid layer sequentially.

B. Dehazing Network

As shown in Fig. 1, each input image is subjected to the DN with disentangled-meta-attention to get the final dehazed result. In the DN, the input images are preprocessed by a 3×3 convolutional layer and then passed through three stages.

The detailed structure of $stage_i$ in the DN is shown in Fig. 3. Each stage of the proposed dehazing model consists of five blocks and one convolutional layer. One of the blocks contains two convolutional layers, a channel attention (CA) module, and a pixel attention (PA) module. The convolutional layers in $stage_i$, with $(2i - 1) \times (2i - 1)$ convolution kernel size, extract features of different receptive fields. Continuous channel and pixel attention in the DN provide additional flexibility for processing different types of information.

Then, the feature maps S1, S2, and S3 from the three stages are weighted according to the disentangled-meta-attention A_{DM} that is calculated by MAN, as shown in Eq. (3). This external meta-knowledge enables the DN to adapt itself

Algorithm 1 Disentangled meta-knowledge-based semi-supervised dehazing scheme

1. **Input:** Input of this scheme

- Train: Paired haze image I_{Sh} and ground truth J_{Sc} from ITS and OTS datasets, real-world image I_R ;
- Test: Haze images I from synthetic or real-world.

2. **Output:** Output of this scheme

- Train: DNDM;
- Implement: Haze-free image \hat{J} ;

3. **Initialization:** Initialize the proposed DNDM with given hyper-parameters

- epoch = 20;
- optimizer = Adam, learning-rate = 0.0001;
- λ_1 to λ_6 are set to 10, 1, 0.001, 2*1e-7, 0.01 and 0.01 in sequence;

4. **Training:** Train with ITS, OTS, and real-world haze images for i = 1 to epoch

```

    for j = 1 to number of images in training datasets
        Random choose Input ( $I_{Sh}, I_{Sc}, I_R$ ) and center crop
        to  $128 \times 128$ ;
        • Image decoupling and reconstruct by DRC
            Obtain  $C_{I_{Sh}}, M_{I_{Sh}}, R_{Sh}$  by DRC( $I_{Sh}$ );
            Obtain  $C_{I_{Sc}}, M_{I_{Sc}}, R_{Sc}$  by DRC( $I_{Sc}$ );
            Obtain  $C_{I_R}, M_{I_R}, R_R$  by DRC( $I_R$ );
            Cross-reconstruct  $I_{FSH}$  by Reconstruction
            network( $C_{I_{Sc}}, M_{I_{Sh}}$ );
            Obtain  $C_{I_{FSH}}, M_{I_{FSH}}$  by DRC( $I_{FSH}$ );
        • Image dehazing by DN with meta-knowledge
            Obtain  $J_{Sh}$  by DN( $I_{Sh}$ , MAN( $C_{I_{Sh}}, M_{I_{Sh}}$ ));
            Obtain  $J_{Sc}$  by DN( $I_{Sc}$ , MAN( $C_{I_{Sc}}, M_{I_{Sc}}$ ));
            Obtain  $J_R$  by DN( $I_R$ , MAN( $C_{I_R}, M_{I_R}$ ));
            Obtain  $J_{FSH}$  by DN( $I_{FSH}$ , MAN( $C_{I_{FSH}}, M_{I_{FSH}}$ ));
        • Optimizer (DNDM) with loss Eq. (4);
    end
end
Output: DNDM;
```

5. **Implementation:** Implementation on synthetic/real-world haze image

- Input: a haze image I with arbitrary resolution;
- $\hat{J} = \text{DNDM}(I)$;
- Output: Haze-free image \hat{J} ;

according to the content and mask information of the input image.

$$S_{DM} = A_{DM1} \otimes S_1 + A_{DM2} \otimes S_2 + A_{DM3} \otimes S_3. \quad (3)$$

Subsequently, the PA module is used to select and optimize features. Finally, the dehazed image is output after two convolutional layers.

C. Loss Function

In the proposed method, the semi-supervised training of a single-image dehazing network is implemented using a disentangled-meta-attention mechanism. In forward propagation, a pair of synthetic images (I_{Sh}, I_{Sc}) and real-world images (I_R) are input. Each image is passed through the

DNDM network to obtain the content and mask features, and the dehazed and recovery images for loss calculation. Dehazed loss (L_{DE}) is only used with supervised data, while the remaining losses ($L_{DRC}, L_{COL}, L_{TV}, L_{DCP}$) are applied to both supervised synthetic and unsupervised real-world images. The details of the loss function are given in this section. Because the supervised and unsupervised branches share the same architecture and weights, the total loss function computed by both parts is used to update the weights. The overall DNDM scheme is presented in Algorithm 1.

The overall loss function of the network training is formulated as

$$L_{train} = \lambda_1 L_{DE} + \lambda_2 L_{DRC} + \lambda_3 L_{COL} + \lambda_4 L_{TV} + \lambda_5 L_{DCP}. \quad (4)$$

Dehazed loss. Dehazed loss L_{DE} , widely employed to constrain supervised parts in low-visual tasks, contains smooth L1 loss and perceptual loss L_p in the ratio 1:0.04. The perceptual loss L_p employs the VGG16 pre-trained on ImageNet as the loss network and extracts features from the last layer of the first four stages.

$$\begin{aligned} L_{DE} = & \|J_{Sh}, I_{Sc}\|_1 + 0.04 * L_p(J_{Sh}, I_{Sc}) + \\ & \|J_{Sc}, I_{Sc}\|_1 + 0.04 * L_p(J_{Sc}, I_{Sc}) + \\ & \|J_{FSH}, I_{Sc}\|_1 + 0.04 * L_p(J_{FSH}, I_{Sc}). \end{aligned} \quad (5)$$

DRC loss. The DRC network calculates the DRC loss function L_{DRC} . In addition to the widely used reconstruction loss, we adopt the content and mask branches in the design of L_{Con} and L_{Mask} . The formula is as follows:

$$L_{DRC} = L_{Con} + L_{Mask} + L_{Re}. \quad (6)$$

Here, L_{Con} calculated by the content branch is exploited to maintain the consistency of the image content, and L_{Mask} calculated by the mask branch is used to constrain image degradation.

To further improve the decoupling ability of the network, a fake hazy image I_{FSH} is cross-reconstructed from the content of the clear image $C_{I_{Sc}}$ and a mask of the hazy image $M_{I_{Sh}}$ using the reconstruction network, and then input into the DNDM network to obtain the corresponding dehazing image as well as content and mask features to calculate the loss function. As shown in Eq. (7), L_{Con} calculates the distance between the content of two images. A pair of synthetic images ($C_{I_{Sh}}, C_{I_{Sc}}$), fake hazy image and original synthetic clear image ($C_{I_{FSH}}, C_{I_{Sc}}$), dehazed image and clear image ($C_{J_{Sh}}, C_{I_{Sc}}$), and real dehazed image and real image (C_{J_R}, C_{I_R}) are included. L_{Mask} calculates the distances between the masks of two images and consists of the dehazed image and clear image ($M_{J_{Sh}}, M_{I_{Sc}}$) as well as the fake hazy image and original synthetic haze image ($M_{I_{FSH}}, M_{I_{Sh}}$), as shown in Eq. (8).

$$\begin{aligned} L_{Con} = & \|C_{I_{Sh}} - C_{I_{Sc}}\|_1 + \|C_{J_{Sh}} - C_{I_{Sh}}\|_1 \\ & + \|C_{I_{FSH}} - C_{I_{Sc}}\|_1 + \|C_{J_R} - C_{I_R}\|_1, \end{aligned} \quad (7)$$

$$L_{Mask} = \|M_{J_{Sh}} - M_{I_{Sc}}\|_1 + \|M_{I_{FSH}} - M_{I_{Sh}}\|_1, \quad (8)$$

where C and M represent the feature maps of the content and mask networks, respectively. The distances are calculated using the smooth L1.

TABLE I: Comparison results of state-of-the-art dehazing methods on the synthetic objective testing set. Red, blue, and green indicate the best, second-best, and third-best performances, respectively.

Type	Method	SOTS-Outdoor		HSTS		HazeRD	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Supervised	DehazeNet [7]	24.75	0.9269	24.48	0.9153	15.87	0.7876
	AoDNet [8]	22.71	0.9112	20.55	0.8973	16.85	0.7925
	EPDN [10]	22.32	0.8683	21.33	0.9046	15.60	0.7492
Unsupervised	DCP [2]	19.14	0.8605	14.84	0.7609	15.74	0.7964
	ZID [28]	13.52	0.6338	22.65	0.9011	12.81	0.4911
	RefineDNet [27]	20.61	0.8798	20.98	0.9070	17.39	0.8553
Semi-supervised	USID-Net [34]	23.89	0.9190	25.76	0.9230	16.25	0.7762
	SSL [13]	25.36	0.9212	24.59	0.9497	17.34	0.8029
	DA [14]	26.72	0.9226	28.74	0.9407	16.54	0.7636
	DMT-Net [15]	22.61	0.9327	22.21	0.9222	17.31	0.7915
Ours	PSD [16]	15.15	0.7325	14.69	0.7354	13.64	0.6937
	DNDM	27.10	0.9555	29.48	0.9653	17.06	0.8278

The reconstruction loss function L_{Re} is used to calculate the distance between the recovered images and original image. Moreover, the distance between the fake hazy image I_{FSH} and synthetic haze image I_{Sh} is calculated. The distance metric also adopts a smooth L1 loss and perception loss L_p in the ratio 1:0.04.

$$\begin{aligned} L_{Re} = & \|R_{Sh}, I_{Sh}\|_1 + 0.04 * L_p(R_{Sh}, I_{Sh}) + \\ & \|R_{Sc}, I_{Sc}\|_1 + 0.04 * L_p(R_{Sc}, I_{Sc}) + \\ & \|R_R, I_R\|_1 + 0.04 * L_p(R_R, I_R) + \\ & \|I_{FSH}, I_{Sh}\|_1 + 0.04 * L_p(I_{FSH}, I_{Sh}). \end{aligned} \quad (9)$$

Constant-color loss. A constant-color loss L_{COL} is introduced to constrain the dehazing image based on its color correlation with the original image. It includes L_{CAP} and L_{Lab} to consider the color and brightness of the image, respectively, with a coefficient ratio of 1:0.1. L_{CAP} was proposed by Li et al. [29] to constrain the brightness (V) and saturation (S) information of the dehazed image.

The theory of optical imaging systems assumes that atmospheric light as a system is independent of varying light inputs and outputs. In this study, a brightness constant loss function L_{Lab} is designed to constrain the stability of the DN in terms of image luminosity. The L_{Lab} function is shown in Eq. (10).

$$\begin{aligned} L_{Lab} = & \|L_{J_{Sh}}, L_{I_{Sc}}\|_1 + \|L_{J_{FSH}}, L_{I_{Sc}}\|_1 \\ & + \lambda_6 * \|L_{J_R}, L_{I_R}\|_1, \end{aligned} \quad (10)$$

where L represents the brightness information of the image and is determined by the luminosity channel of the Lab color space. The brightness distances between the dehazed synthetic image J_{Sh} and corresponding clear image I_{Sc} , dehazed image of the fake hazy image J_{FSH} and real clear image I_{Sc} , and dehazed real haze image J_R and original haze image I_R are calculated in the proposed method.

TV loss. The unsupervised total variation [39] loss L_{TV} is widely applied to low-level vision tasks to preserve the structure and detail information, as shown in Eq. (11).

$$L_{TV} = \|\nabla J_{Sh}\|_1 + \|\nabla J_R\|_1. \quad (11)$$

DCP loss. He et al. [2] proposed the dark channel prior and showed that one of the three RGB color channels of each pixel in the clear image is zero or close to zero. Based on this, Li et al. [13] reported the effectiveness of the dark channel loss L_{DCP} as an unsupervised dehazing loss, as shown in Eq. (12).

$$L_{DCP} = DCP(J_R) + DCP(J_{Sh}) + DCP(J_{Sc}) + DCP(J_{FSH}). \quad (12)$$

IV. EXPERIMENTS

In this section, the experimental setup used to verify the efficacy of the proposed method is detailed, and experimental results on real and synthetic datasets are presented. Furthermore, ablation studies were conducted to evaluate the influences of the network structure, loss function, and training data design on the results.

A. Datasets

In this study, we trained the network on both synthetic and real-world datasets. We manually collected 13,990 real-world hazy images and 14,427 clear images from the Internet to create unsupervised datasets; these data are publicly available at <https://github.com/dehazing/DNDM>. The indoor (ITS) and outdoor (OTS) synthetic datasets in the REalistic Single-Image DEhazing (RESIDE) dataset [40] were selected as the supervised datasets, where the indoor and outdoor datasets contained 13,990 and 72,135 pairs of images, respectively. To balance the number of synthetic and real-world images, 14,427 pairs of outdoor synthetic data were used in the training, where we randomly selected seven hazy images for each clear image in OTS.

The synthetic objective testing set (SOTS) in RESIDE provided 500 pairs of outdoor haze images, the hybrid subjective testing set (HSTS) contains 10 synthetic outdoor hazy images generated in the same manner as SOTS, and the real-world task-driven testing set (RTTS) consisted of 4,322 labeled haze images. These datasets were used to evaluate the performance of the proposed method. Furthermore, the HazeRD dataset [41] was used for testing and contained 15 clear images and

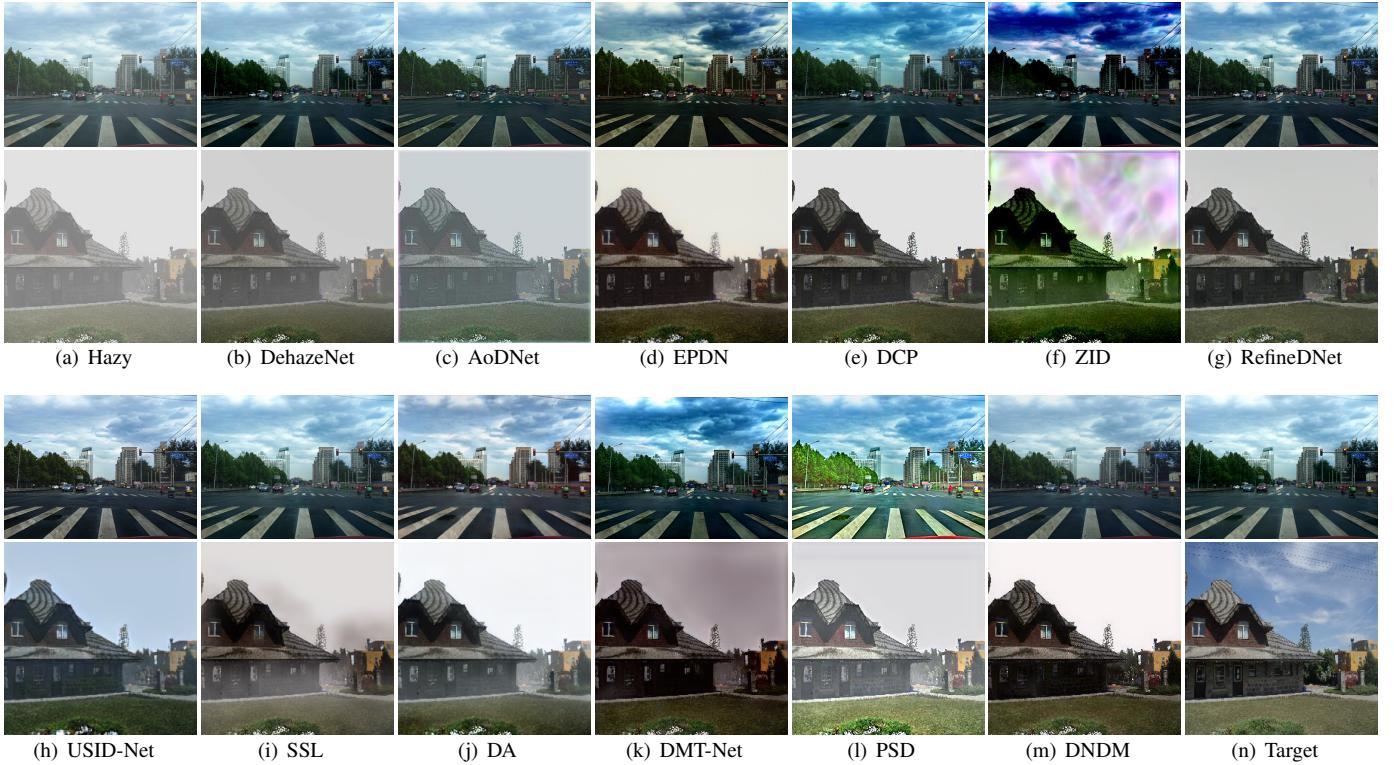


Fig. 4: Visual comparisons of synthetic haze images in SOTS [40] and HazeRD [41]. (a)–(n): input image, results produced by DehazeNet [7], AoDNet [8], EPN [10], DCP [2], ZID [28], RefineDNet [27], USID-Net [34], SSL [13], DA [14], PSD [16], DNDM, and Target.

TABLE II: Comparison results of state-of-the-art dehazing methods on I-HAZE [42] and O-HAZE [43]. Red, blue, and green indicate the best, second-best, and third-best performances, respectively.

Method	I-HAZE		O-HAZE	
	PSNR	SSIM	PSNR	SSIM
DehazeNet [7]	16.3191	0.6382	15.4857	0.5116
AoDNet [8]	14.6558	0.6091	14.9796	0.5012
EPN [10]	15.4581	0.6571	16.1149	0.6027
DCP [2]	13.9493	0.6133	15.8626	0.6088
ZID [28]	13.8202	0.5517	14.1226	0.5141
RefineDNet [27]	16.0831	0.6903	17.0079	0.6373
USID-Net [34]	16.5149	0.6845	15.5944	0.5389
SSL [13]	15.8325	0.6872	15.9698	0.6390
DA [14]	16.9790	0.6956	14.8559	0.5497
DMT-Net [15]	14.3332	0.6038	15.1658	0.6049
PSD [16]	14.1788	0.6172	12.8758	0.5835
DNDM	17.1869	0.7044	16.3347	0.6405

75 corresponding synthetic hazy images, which were adjusted to sizes of 256×256 for testing owing to the processing limitations of the comparison method.

In contrast to most of the existing dehazing databases, hazy images in the indoor scenes database (denoted I-HAZE [42]) and the outdoor scenes database (denoted O-HAZE [43]) were generated using real-world haze produced by a professional haze machine. I-HAZE contains 35 image pairs of hazy and

corresponding haze-free (ground truth) indoor images. O-HAZE contains 45 different outdoor scenes depicting the same visual content recorded in haze-free (ground truth) and hazy conditions.

B. Implementation Details

The proposed model was optimized using the Adam Optimizer, with $\beta_1 = 0.9$ and $\beta_2 = 0.99$. This network was trained for 20 epochs, and the weight decay of the first three rounds was 0.0001, which was then set to 0.00001. The entire model was implemented using the PyTorch framework, and experiments were conducted using an NVIDIA GeForce RTX 2080Ti GPU. The images used for network training were center-cropped to 128×128 pixels. In the experiments, the hyper-parameters λ_1 to λ_6 were set to 10, 1, 0.001, 2×10^{-7} , 0.01, and 0.01, respectively, based on experience and experiments.

C. Evaluations on Synthetic and Real Datasets

The objective and subjective results of the proposed method and prior methods were compared on the synthetic dataset. Classical metrics such as peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), were used for performance evaluation. As observed from Table I, our method achieved the best results on SOTS and HSTS datasets and was competitive on the HazeRD dataset. Among them, RefineDNet exploits DCP and the AS model for image dehazing, which allows

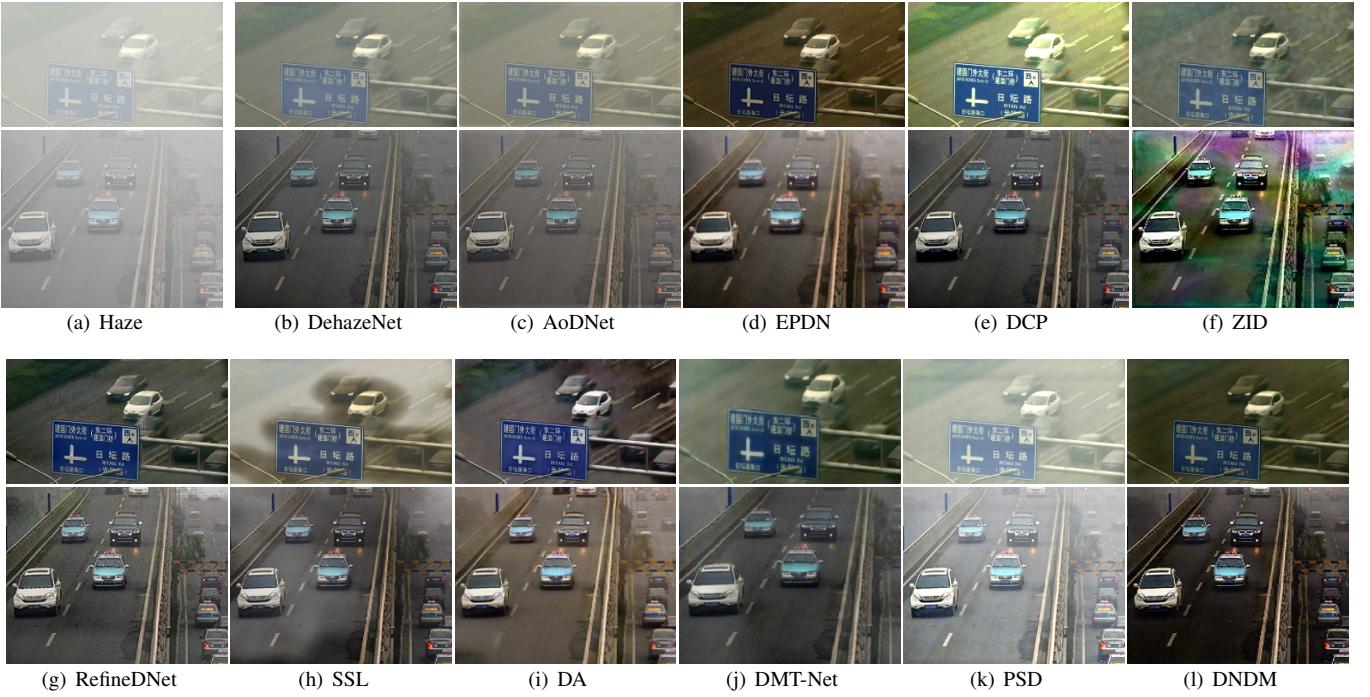


Fig. 5: Visual comparisons of real-world haze images. (a)–(l): input haze image, results by DehazeNet [7], AoDNet [8], EPDN [10], DCP [2], ZID [28], RefineDNet [27], SSL [13], DA [14], DMT-Net [15], PSD [16], and DNDM.

TABLE III: Object-detection results for the RTTS [40] dataset. Red, blue, and green indicate the best, second-best, and third-best performances, respectively.

Module	Person	Bicycle	Car	Motorbike	Bus	mAP
Haze	0.721	0.441	0.617	0.473	0.364	0.523
DehazeNet	0.722	0.424	0.622	0.485	0.375	0.526
AoDNet	0.720	0.448	0.629	0.498	0.376	0.534
EPDN	0.706	0.419	0.598	0.458	0.347	0.506
DCP	0.723	0.433	0.615	0.485	0.370	0.525
ZID	0.584	0.336	0.468	0.383	0.320	0.418
RefineDNet	0.726	0.459	0.625	0.499	0.373	0.536
USID-Net	0.707	0.469	0.620	0.506	0.377	0.536
SSL	0.722	0.440	0.625	0.500	0.379	0.533
DA	0.716	0.431	0.653	0.495	0.397	0.538
DMT-Net	0.690	0.417	0.608	0.442	0.348	0.501
PSD	0.684	0.398	0.612	0.470	0.368	0.506
DNDM	0.723	0.456	0.631	0.512	0.385	0.542

it to achieve excellent performance on the HazeRD dataset based on physically realistic parameters and AS model for image synthesis. However, this also limits its applicability to other datasets. In contrast, DNDM is model-independent and has better generalization capability. As shown in Fig. 4, the proposed algorithm performed well relative to state-of-the-art dehazing methods.

We evaluated the ability of the model to process real-world images. As shown in Fig. 5, most of the algorithms exhibited limitations with incomplete dehazing. These include DehazeNet [7], AoDNet [8], EPDN [10], DCP [2], SSL [13],

DMT-Net [15], and PSD [16]. In addition, some artifacts were observed in the dehazed images of DA [14], ZID [28] and RefineDNet [27], which affected the visual appearances of the images. The proposed method performed well and achieved better visual results. We also conducted objective metric comparisons on the real-world datasets I-HAZE and O-HAZE. Table II shows that our method achieved the best results on the I-HAZE dataset and was competitive on the O-HAZE dataset.

D. Task-driven Evaluation

Haze severely reduces the visibility of images in vision-based outdoor industry systems [44], [45]. For example, in monitoring systems, human faces are obscured by haze, resulting in insufficient visual and identity clues. This poses a challenge to face detection [46], [47] and recognition [48], [49] algorithms trained without hazy scenes. In addition, contour and detailed information of vehicles and pedestrians are lost in the haze, leading to false detection in autonomous driving systems [34], which seriously affects the stability of the systems. The proposed dehazing algorithm can effectively improve the visibility of images and recover identity cues of faces, as well as detailed pedestrian and vehicle information, thus improving the accuracy of subsequent detection and recognition tasks.

Since object detection is an essential component of vision-based outdoor industrial systems, this section evaluates the object-detection results of the RTTS dataset in a task-driven manner to demonstrate the effectiveness of DNDM in vision-based outdoor industrial systems. The RTTS dataset contained 4,322 real-world haze images. In this study, a pre-trained YOLO-v5l algorithm was used to detect the objects in the



Fig. 6: Visualization of the results of object detection in real-world images. (a): Haze may cause loss of detection capability. (b)-(l): The dehazing methods effectively improve the subjective visual effects and object-detection accuracy of the images.



Fig. 7: Visualization of the results of face detection on UFDD [50]. Top row: haze leads to face detection omission; Bottom row: the dehazing methods effectively improve the subjective visual effects and face detection accuracy of the images.

hazy images and dehazed images, respectively. The images in RTTS were rescaled to 0.3 times for object detection, which better meets the requirements of small object detection in actual vision-based outdoor industry systems and ensures consistency of testing platforms. Fig. 6 shows that the object-detection algorithm can be expected to fail with issues such as loss of detection and uncertain category recognition owing to haze in images. The dehazing approach adopted by DNDM effectively improved the subjective visual appearance and detection accuracy of processed images. Table III compares the proposed method and other dehazing methods in object detection in hazy images. The proposed method is competitive in terms of detection accuracy for each category and shows an advantage in mAP.

Furthermore, we verified the effectiveness of the proposed algorithm in face detection and recognition tasks. For haze images from UFDD [50], face detection was performed using the classical MTCNN [51] algorithm before and after haze removal, and the average precision (AP) increased from 58.8% to 60.3%. The visualization results of facial detection before and after dehazing by DNDM are illustrated in Fig. 7; notably, haze leads to face omission, and the proposed algorithm effectively improves the subjective visual effects and face detection accuracy of the images. In addition, the influence of haze on face recognition was verified on synthetic hazy face images based on the LFW dataset [52]. The face recognition accuracy of FaceNet [53] improved from 87.5% to 92.3% after the hazy face images were preprocessed using DNDM.

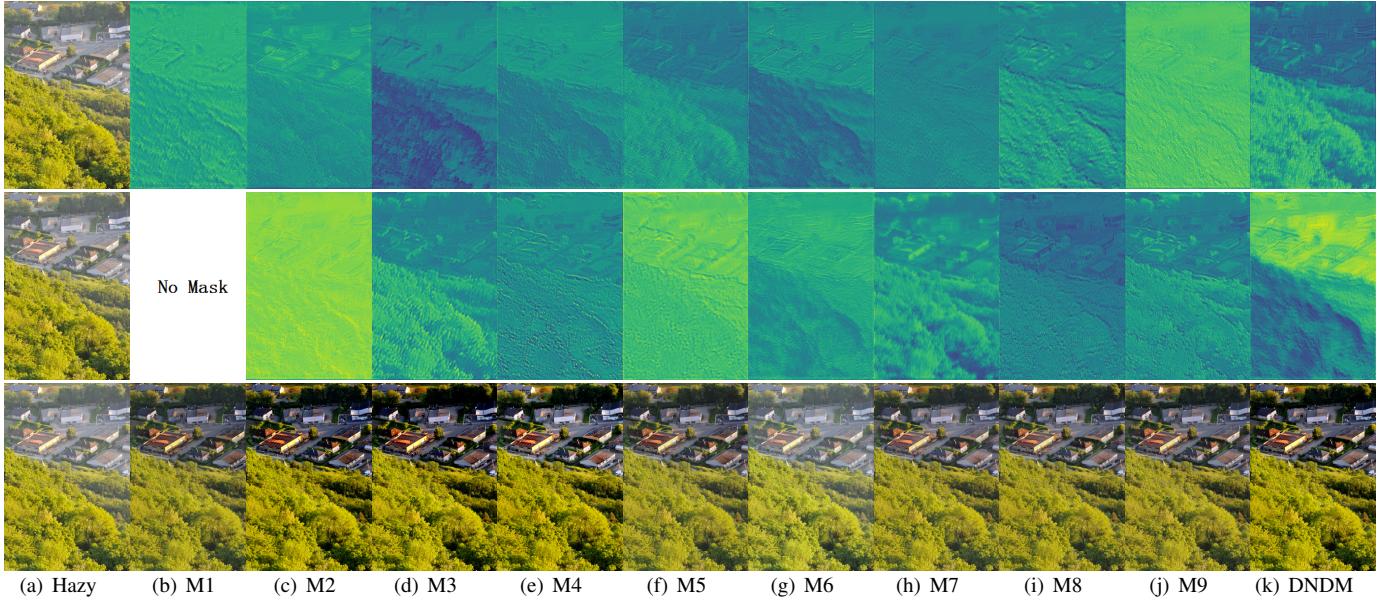


Fig. 8: Comparison of content/mask features and dehazed images visualization for different sizes of the convolution kernel, feature fusion methods, and activation functions. The first row is content features, the second row is mask features, and the last row is dehazed images.

TABLE IV: Number of parameters and runtimes of different dehazing methods. Red, blue, and green indicate the best, second-best, and third-best performances, respectively.

Method	Platform	Params(M)	Runtime(s)
DehazeNet	Matlab(CPU)	0.008	1.3537
AoDNet	PyTorch(GPU)	0.002	0.0022
EPDN	PyTorch(GPU)	17.379	0.0254
DCP	Matlab(CPU)	-	0.1319
ZID	PyTorch(GPU)	40.406	59.9900
RefineDNet	PyTorch(GPU)	65.795	0.4541
USID-Net	PyTorch(GPU)	3.770	0.0144
SSL	PyTorch(GPU)	9.231	0.0777
DA	PyTorch(GPU)	54.591	0.0350
DMT-Net	PyTorch(GPU)	51.790	0.1321
PSD	PyTorch(GPU)	6.205	0.1385
DNDM	PyTorch(GPU)	1.791	0.0214

The above comparison results prove that the proposed scheme has superior processing effects on synthetic and real-world haze images. In addition, we compared the average runtime of each method on 500 images of 512×512 pixels and calculated the parameters and runtimes. For fairness, all models were executed on a desktop PC with an NVIDIA GeForce RTX 2080Ti GPU. As shown in Table IV, compared with other methods, DNDM only lagged behind AoDNet and USID-Net but achieved more promising dehazing results.

In conclusion, DNDM can achieve pleasing dehazing results with low computational cost, and dehazing preprocessing can effectively improve object detection performance. The above observations verify the potential of meta-knowledge-driven image enhancement algorithms in subsequent applications. Besides, the meta-knowledge-driven approach can also be applied to other related fields, such as low-resolution face recognition

TABLE V: Ablation on different network structures in SHS. C35/M35 indicates that the convolutional kernels used in the residual network of the content/mask network were 3×3 and 5×5 , successively. Le/Re represent LeakyRelu/Relu, respectively.

Model	Convolution	Activate Function	Fusion methods	PSNR	SSIM
M1	C35 No mask	Re/Le	$C \otimes M \oplus C$	25.43	0.9451
M2	C35M35	Re/Le	$C \otimes M \oplus C$	25.48	0.9444
M3	C53M53	Re/Le	$C \otimes M \oplus C$	26.43	0.9521
M4	C53M35	Re/Le	$C \otimes M \oplus C$	26.80	0.9567
M5	C35M53	Re/Re	$C \otimes M \oplus C$	26.56	0.9538
M6	C35M53	Le/Le	$C \otimes M \oplus C$	26.45	0.9526
M7	C35M53	Le/Re	$C \otimes M \oplus C$	26.74	0.9551
M8	C35M53	Re/Le	Concatenate(C, M)	26.47	0.9522
M9	C35M53	Re/Le	$C \otimes M$	26.15	0.9524
DNDM	C35M53	Re/Le	$C \otimes M \oplus C$	27.10	0.9555

[54], [55]. The details and facial cues of low-resolution face images can be explicitly reconstructed by accumulating face image super-resolution meta-knowledge, thus improving face image quality and increasing recognition accuracy.

E. Ablation Study

Ablation experiments to demonstrate the effectiveness of the proposed network structure and loss function were performed and are described in this section.

1) Network Architecture

To better illustrate the rationality of the proposed network architecture, this section presents the results of ablation experiments conducted to evaluate the efficacy of various components—specifically, a) the SHS design of the content and mask branch networks in the disentangled network and b) the configuration of the network structure and training data.

a) The SHS

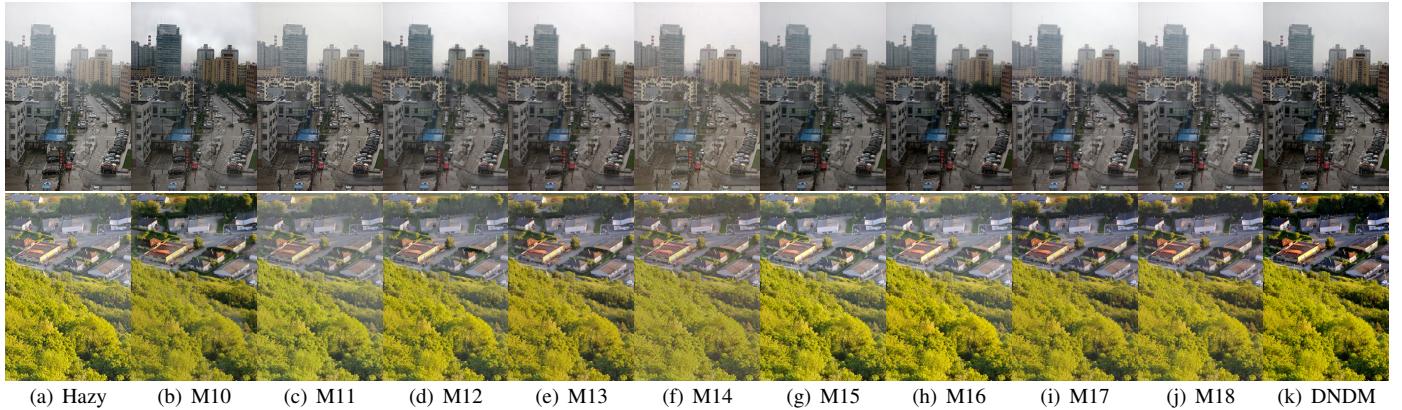


Fig. 9: Comparison of different network structure settings and training data. M10/M11 represents the DN without CA/PA. M12/M13 means that DNDM is trained without meta mechanism/real-world hazy images. The performance of DNDM with different loss functions. Experiments M14-M18 represent algorithm trained without L_{DE} , L_{DRC} , L_{COL} , L_{DCP} and L_{TV} , respectively.

TABLE VI: Ablation on different network structures. CA/PA represents channel/pixel attention in DN. Meta and Real mean that DNDM was trained with/without meta mechanism/real-world hazy images.

Model	CA	PA	Meta	Real	PSNR	SSIM
M10	✗	✓	✓	✓	25.17	0.9453
M11	✓	✗	✓	✓	23.10	0.9038
M12	✓	✓	✗	✓	25.01	0.9378
M13	✓	✓	✓	✗	26.18	0.9521
DNDM	✓	✓	✓	✓	27.10	0.9555

The SHS was designed with a focus on three aspects: the size of the convolution kernel, activation function, and feature fusion method.

The quantitative comparison results of the SHS ablation experiments are shown in Table V, where DNDM achieved the best performance. Further, the content/mask features visualization results and dehazed images are shown in Fig. 8. They prove that SHS effectively improves the ability of the disentangled network to decompose the images, further enhancing the performance of the algorithm. The mask network is designed for better learning of meta-knowledge through reconstruction and cross-reconstruction, and it can be observed from the ablation results of M1 that the absence of the mask network leads to different degrees of degradation in both subjective and objective aspects.

Kernel size of the convolution in SHS: The notation C35 in M2–M4 signifies that the convolutional kernels used in the residual network of the content network were 3×3 and 5×5 , successively. Similarly, M53 signifies that the convolutional kernels employed in the residual network of the mask network were 5×5 and 3×3 in turn. The visualization of content/mask features and dehazed images for M2–M4 and DNDM in Fig. 8 shows that the unique convolutional kernel size setting in SHS guides the content network to extract image details, while the mask network focuses more on hazy degradation.

Activation functions in SHS: The results of ablation experiments on the activation functions in SHS are given in M5–M7 of Fig. 8 and Table V, where Re/Le represent Relu/LeakyRelu. It can be seen that more detailed information can be preserved in the content features when Relu and LeakyRelu activation functions are employed in the content and mask networks, respectively.

Feature fusion methods: We also performed ablation experiments on the feature fusion methods of the content and mask features. The results can be observed in M8–DNDM in Fig. 8. When we fuse content and mask features using concatenation or multiplication alone, the mask features contain more information than the content features, which is not conducive for the network to decouple information. The content feature multiplies the mask and adds itself to better guide the content network to extract detailed information from the images and achieve the best dehazing performance.

b) Dehazing network

The comparison results of ablation experiments with different network structures and training data are shown in Fig. 9 and Table VI. It can be seen that both the channel and PA mechanisms in the dehazing network are effective in improving the performance of the algorithm (M10/M11/DNDM). Furthermore, DNDM improves the dehazing performance to a greater extent because the combination of CA and PA provides additional flexibility to the network.

In addition, the model with meta mechanism, trained with real-world hazy images, can address the domain gap between synthetic and real-world haze images (M12/M13/DNDM). By introducing decoupled meta-knowledge and real-world hazy images, the objective index PSNR on synthetic data is improved by 2.09/0.92 dB, respectively. Furthermore, DNDM removes haze and restores image details more effectively for real-world hazy images. The results showed that the extraction of meta-knowledge by decoupling the input images effectively improves the performance of the DN. Introducing real-world hazy data in the training phase enhances the meta-knowledge representation and effectively improves the robustness of the

TABLE VII: Ablation on different loss functions.

Module	L_{DE}	L_{DRC}	L_{COL}	L_{TV}	L_{DCP}	PSNR	SSIM
M14	✗	✓	✓	✓	✓	23.16	0.8269
M15	✓	✗	✓	✓	✓	24.75	0.9323
M16	✓	✓	✗	✓	✓	25.99	0.9487
M17	✓	✓	✓	✗	✓	26.77	0.9536
M18	✓	✓	✓	✓	✗	26.84	0.9575
DNDM	✓	✓	✓	✓	✓	27.10	0.9555

TABLE VIII: Ablation results for different hyper-parameter values of the loss weights.

Loss	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	PSNR	SSIM
L_{DE}	100 1	1	0.001	2*1e-7	0.01	0.01	26.14 24.25	0.9520 0.7790
L_{DRC}	10 0.1	10 0.001	0.001	2*1e-7	0.01	0.01	26.59 25.44	0.9561 0.9425
L_{COL}	10 1	1 0.0001	0.01 2*1e-7	0.01	0.01	0.01	26.31 26.43	0.9560 0.9550
L_{TV}	10 1	1 0.001	0.001 2*1e-6	0.01 2*1e-8	0.01	0.01	27.03 26.54	0.9577 0.9540
L_{DCP}	10 1	1 0.001	0.001 2*1e-7	0.1 0.001	0.01	0.01	23.84 26.51	0.7641 0.9523
L_{Lab}	10 1	1 0.001	0.001 2*1e-7	0.01	0.1 0.001	0.01	27.03 26.84	0.9551 0.9556
DNDM	10 1	1 0.001	0.001 2*1e-7	0.01	0.01	0.01	27.10	0.9555

dehazing algorithm in both synthetic and real-world domains.

2) Loss Functions

We removed one loss at a time from the overall loss function to verify its impact on the network. The quantitative results are shown in Table VII, where L_{DE} and L_{DRC} considerably impacted the network performance, which can be attributed to their ability to stabilize model training better than other loss functions. The subjective results of the experiments are seen in Fig. 9. DRC loss and constant-color loss functions were specially designed during training to optimize the dehazing network. Experiments M14, M15, and DNDM in Table VII and Fig. 9 show the necessity of adding L_{DE} and L_{DRC} . A comparison of experiments M16 and DNDM shows that the addition of L_{COL} allowed the model to retain the color information of the dehazed images. Experiments M17, M18, and DNDM in Table VII and Fig. 9 show that L_{TV} and L_{DCP} enable more details and better visual experience. The quantitative and qualitative results show that DNDM achieves superior dehazing performance owing to the combination of the abovementioned loss functions.

To find the balance between the loss functions, we conducted ablation experiments using different hyperparameters. There are many combinations of loss function hyper-parameter settings. In this study, we experimented with different weights in the upper and lower order of magnitude. Table VIII shows the results of the ablation experiments under different hyper-parameter settings of the loss functions. As can be observed from the results, the sensitivity of each weight to the final result is within a reasonable range and varies. The hyper-

parameter settings in the paper are all at a reasonable level, and slight adjustments to a particular parameter do not affect the overall performance drastically.

V. CONCLUSION AND FUTURE WORK

This paper proposes a semi-supervised dehazing algorithm that uses disentangled meta-knowledge. For the first time, the disentangled representations were used to realize meta-learning and a semi-supervised dehazing network design, which can further enhance the generalization ability of the network for real-world scenes. To better extract the effective guiding information of the haze image dehazing process, the disentangled network had a unique symmetrical and heterogeneous design and used the DRC loss to understand entanglement. Finally, the addition of constant-color loss ensured the color stability of the dehazed images and better preserved the image information. The results of extensive experiments conducted on synthetic and real-world haze images demonstrated that the proposed algorithm performed more favorably than state-of-the-art dehazing methods.

Although the proposed DNDM achieved favorable generalization capability in image dehazing, there is still potential for improvement as the network structure is designed manually. In addition, the dataset remains as one of the factors affecting the performance of image dehazing, and we will focus on collecting more real haze images and investigating their effects on the performances of semi-supervised and unsupervised methods. The proposed fusion method in SHS can guide the content features to retain more detailed information and is applicable to many types of degraded images. At the same time, the parameters of a symmetric heterogeneous network structure needs a special design for different types of degraded images. In future work, network structure search will be used to explore the applicability of different symmetric heterogeneous network structures to other types of degraded images.

REFERENCES

- [1] S. G. Narasimhan and S. K. Nayar, "Chromatic framework for vision in bad weather," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, vol. 1. IEEE, 2000, pp. 598–605.
- [2] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.
- [3] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE transactions on image processing*, vol. 24, no. 11, pp. 3522–3533, 2015.
- [4] D. Berman, S. Avidan *et al.*, "Non-local image dehazing," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1674–1682.
- [5] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Polarization-based vision through haze," *Applied optics*, vol. 42, no. 3, pp. 511–525, 2003.
- [6] L. Shen, Y. Zhao, Q. Peng, J. C.-W. Chan, and S. G. Kong, "An iterative image dehazing method with polarization," *IEEE Transactions on Multimedia*, vol. 21, no. 5, pp. 1093–1107, 2018.
- [7] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.
- [8] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4770–4778.
- [9] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3194–3203.

- [10] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced pix2pix dehazing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8160–8168.
- [11] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 11 908–11 915.
- [12] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2157–2167.
- [13] L. Li, Y. Dong, W. Ren, J. Pan, C. Gao, N. Sang, and M.-H. Yang, "Semi-supervised image dehazing," *IEEE Transactions on Image Processing*, vol. 29, pp. 2766–2779, 2019.
- [14] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang, "Domain adaptation for image dehazing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2808–2817.
- [15] Y. Liu, L. Zhu, S. Pei, H. Fu, J. Qin, Q. Zhang, L. Wan, and W. Feng, "From synthetic to real: Image dehazing collaborating with unlabeled real data," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 50–58.
- [16] Z. Chen, Y. Wang, Y. Yang, and D. Liu, "Psd: Principled synthetic-to-real dehazing guided by physical priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7180–7189.
- [17] D. Yang and J. Sun, "Proximal dehaze-net: A prior learning-based deep network for single image dehazing," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 702–717.
- [18] H.-M. Hu, H. Zhang, Z. Zhao, B. Li, and J. Zheng, "Adaptive single image dehazing using joint local-global illumination adjustment," *IEEE Transactions on Multimedia*, vol. 22, no. 6, pp. 1485–1495, 2019.
- [19] Y. Song, J. Li, X. Wang, and X. Chen, "Single image dehazing using ranking convolutional neural network," *IEEE Transactions on Multimedia*, vol. 20, no. 6, pp. 1548–1560, 2017.
- [20] H. Zhu, X. Peng, V. Chandrasekhar, L. Li, and J.-H. Lim, "Dehazegan: When image dehazing meets differential programming," in *IJCAI*, 2018, pp. 1234–1240.
- [21] M. Hong, Y. Xie, C. Li, and Y. Qu, "Distilling image dehazing with heterogeneous task imitation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3462–3471.
- [22] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma, "Contrastive learning for compact single image dehazing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10 551–10 560.
- [23] C. Lin, X. Rong, and X. Yu, "Msaff-net: Multiscale attention feature fusion networks for single image dehazing and beyond," *IEEE Transactions on Multimedia*, 2022.
- [24] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. Li, "Maxim: Multi-axis mlp for image processing," *arXiv preprint arXiv:2201.02973*, 2022.
- [25] D. Engin, A. Genc, and H. K. Ekenel, "Cycle-dehaze: Enhanced cyclegan for single image dehazing," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 938–9388.
- [26] A. Golts, D. Freedman, and M. Elad, "Unsupervised single image dehazing using dark channel prior loss," *IEEE Transactions on Image Processing*, vol. 29, pp. 2692–2701, 2019.
- [27] S. Zhao, L. Zhang, Y. Shen, and Y. Zhou, "Refinednet: a weakly supervised refinement framework for single image dehazing," *IEEE Transactions on Image Processing*, vol. 30, pp. 3391–3404, 2021.
- [28] B. Li, Y. Gou, J. Z. Liu, H. Zhu, J. T. Zhou, and X. Peng, "Zero-shot image dehazing," *IEEE Transactions on Image Processing*, vol. 29, pp. 8457–8466, 2020.
- [29] B. Li, Y. Gou, S. Gu, J. Z. Liu, J. T. Zhou, and X. Peng, "You only look yourself: Unsupervised and untrained single image dehazing neural network," *International Journal of Computer Vision*, pp. 1–14, 2021.
- [30] B. Lu, J.-C. Chen, and R. Chellappa, "Unsupervised domain-specific deblurring via disentangled representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 225–10 234.
- [31] W. Du, H. Chen, and H. Yang, "Learning invariant representation for unsupervised image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 483–14 492.
- [32] Z. Zou, S. Lei, T. Shi, Z. Shi, and J. Ye, "Deep adversarial decomposition: A unified framework for separating superimposed images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 806–12 816.
- [33] Y. Ye, Y. Chang, H. Zhou, and L. Yan, "Closing the loop: Joint rain generation and removal via disentangled image translation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2053–2062.
- [34] J. Li, Y. Li, L. Zhuo, L. Kuang, and T. Yu, "Usid-net: Unsupervised single image dehazing network via disentangled representations," *IEEE Transactions on Multimedia*, 2022.
- [35] F. Shen, S. Yan, and G. Zeng, "Neural style transfer via meta networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8061–8069.
- [36] J. W. Soh, S. Cho, and N. I. Cho, "Meta-transfer learning for zero-shot super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3516–3525.
- [37] Z. Chi, Y. Wang, Y. Yu, and J. Tang, "Test-time fast adaptation for dynamic scene deblurring via meta-auxiliary learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9137–9146.
- [38] X. Zhang, H. Dong, Z. Hu, W.-S. Lai, F. Wang, and M.-H. Yang, "Gated fusion network for degraded image super resolution," *International Journal of Computer Vision*, vol. 128, no. 6, pp. 1699–1721, 2020.
- [39] P. Rodriguez, "Total variation regularization algorithms for images corrupted with different noise models: a review," *Journal of Electrical and Computer Engineering*, vol. 2013, 2013.
- [40] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2018.
- [41] Y. Zhang, L. Ding, and G. Sharma, "Hazerd: an outdoor scene dataset and benchmark for single image dehazing," in *2017 IEEE international conference on image processing (ICIP)*. IEEE, 2017, pp. 3205–3209.
- [42] C. Ancuti, C. O. Ancuti, R. Timofte, and C. D. Vleeschouwer, "I-haze: a dehazing benchmark with real hazy and haze-free indoor images," in *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 2018, pp. 620–631.
- [43] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "O-haze: a dehazing benchmark with real hazy and haze-free outdoor images," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 754–762.
- [44] J. Li, L. Zhuo, H. Zhang, G. Li, and N. Xiong, "Effective data-driven technology for efficient vision-based outdoor industrial systems," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4344–4354, 2019.
- [45] T. Jia, J. Li, L. Zhuo, and G. Li, "Effective meta-attention dehazing networks for vision-based outdoor industrial systems," *IEEE Transactions on Industrial Informatics*, 2021.
- [46] S. Ge, J. Li, Q. Ye, and Z. Luo, "Detecting masked faces in the wild with lle-cnns," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2682–2690.
- [47] D. Zeng, H. Liu, F. Zhao, S. Ge, W. Shen, and Z. Zhang, "Proposal pyramid networks for fast face detection," *Information Sciences*, vol. 495, pp. 136–149, 2019.
- [48] S. Ge, C. Li, S. Zhao, and D. Zeng, "Occluded face recognition in the wild by identity-diversity inpainting," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 10, pp. 3387–3397, 2020.
- [49] H. Jiang and D. Zeng, "Explainable face recognition based on accurate facial compositions," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 1503–1512.
- [50] H. Nada, V. A. Sindagi, H. Zhang, and V. M. Patel, "Pushing the limits of unconstrained face detection: a challenge dataset and baseline results," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2018, pp. 1–10.
- [51] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE signal processing letters*, vol. 23, no. 10, pp. 1499–1503, 2016.
- [52] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition*, 2008.
- [53] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.
- [54] S. Ge, S. Zhao, C. Li, and J. Li, "Low-resolution face recognition in the wild via selective knowledge distillation," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 2051–2062, 2018.

- [55] S. Ge, S. Zhao, C. Li, Y. Zhang, and J. Li, "Efficient low-resolution face recognition via bridge distillation," *IEEE Transactions on Image Processing*, vol. 29, pp. 6898–6908, 2020.

Tongyao Jia is currently pursuing a Ph.D. with the Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology. Her research interest is in image restoration/enhancement.



Jiafeng Li received a PhD in pattern recognition and intelligence system from Beihang University, Beijing, China, in 2016. He was with the Department of Neurosurgery at the University of Pittsburgh, Pittsburgh, PA, USA, as a visiting scholar from 2014 to 2015. He is currently an assistant professor with the Faculty of Information Technology, Beijing University of Technology. He is also affiliated with the Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology. His research interests include image restoration/enhancement, object detection, and object tracking.



Li Zhuo received a PhD in pattern recognition and intelligent systems from Beijing University of Technology, Beijing, China, in 2004. She is currently a Professor with the Faculty of Information Technology, Beijing University of Technology. She is also with the Beijing Key Laboratory of Computational Intelligence and Intelligent Systems, Beijing University of Technology. She authored or coauthored more than 180 refereed journal and conference papers and has written six book chapters. Her current research interests include image/video processing, image recognition, and medical image processing.



Tianjian Yu received a PhD in traffic equipment and information engineering from Central South University (CSU), Changsha, China, in 2017. Since 2014, he has been a visiting PhD student with the University of Pittsburgh, Pittsburgh, PA, USA. Currently, he is a lecturer with Central South University. His research interests include fault diagnosis of traffic equipment, focusing on bearings and inverters, by means of machine learning.

