

Towards Multi-domain Single Image Dehazing via Test-time Training

Huan Liu¹, Zijun Wu¹, Liangyan Li¹, Sadaf Salehkalaibar¹, Jun Chen¹, Keyan Wang²
¹McMaster University, ²Xidian University
 {liuh127, wuz146, lil61, salehkas, chenjun}@mcmaster.ca
 kywang@mail.xidian.edu.cn

Abstract

Recent years have witnessed significant progress in the area of single image dehazing, thanks to the employment of deep neural networks and diverse datasets. Most of the existing methods perform well when the training and testing are conducted on a single dataset. However, they are not able to handle different types of hazy images using a dehazing model trained on a particular dataset. One possible remedy is to perform training on multiple datasets jointly. However, we observe that this training strategy tends to compromise the model performance on individual datasets. Motivated by this observation, we propose a test-time training method which leverages a helper network to assist the dehazing model in better adapting to a domain of interest. Specifically, during the test time, the helper network evaluates the quality of the dehazing results, then directs the dehazing network to improve the quality by adjusting its parameters via self-supervision. Nevertheless, the inclusion of the helper network does not automatically ensure the desired performance improvement. For this reason, a meta-learning approach is employed to make the objectives of the dehazing and helper networks consistent with each other. We demonstrate the effectiveness of the proposed method by providing extensive supporting experiments.

1. Introduction

Single image dehazing is a classic but still active research topic in low-level computer vision, which aims to restore clean images from the degraded hazy counterparts. Recently, many deep learning approaches [5, 10, 14, 22, 25, 26, 31, 35, 45, 49, 50] have been proposed to address this problem by training a neural network to approximate the mapping from hazy images to haze-free ground truths. As more and more dehazing datasets have been released, such as RESIDE [23], O-Haze [3] and NH-Haze [2], these methods are able to demonstrate their outstanding ability in handling different haze patterns. However, one important issue is left behind for consideration, i.e., handling different types

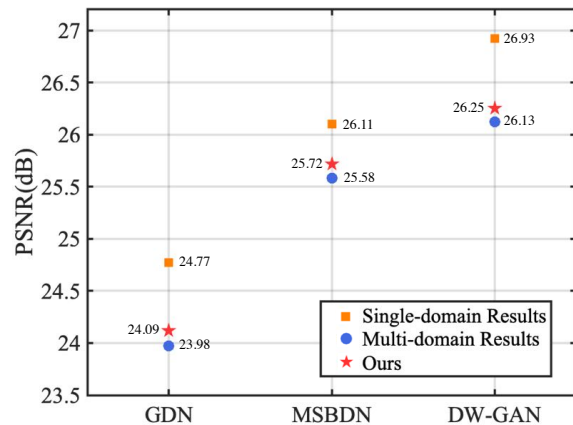


Figure 1: Average PSNR values of GDN [26], MSBDN [11] and DW-GAN [14] across four datasets. It can be observed that the dehazing methods perform better if the training and validation are conducted on a single dataset.

of hazy images by a single network. To be specific, current methods are usually trained on the training split of a particular dataset and tested on the corresponding testing split. For example, the test accuracy on RESIDE indoor test set [23] is obtained by validating a dehazing model trained on the RESIDE indoor training set. Such an evaluation strategy allows the neural network to focus on a specific domain but evades the important problem of learning a general model across datasets. A seemingly simple remedy is to train a single dehazing model on all available datasets jointly. Intuitively, with the increase of data, the network can benefit from considering more kinds of haze patterns, leading to boosted performance on every single dataset [1].

Somewhat surprisingly, we find that this naive solution actually compromises the dehazing performance on individual datasets. Indeed, it can be seen from Figure 1 that the dehazing models perform better when the training and testing are conducted on a single dataset (as opposed to all datasets combined). This unusual fact contradicts the common belief that the increase in data usually leads to improved performance. One possible explanation is that each

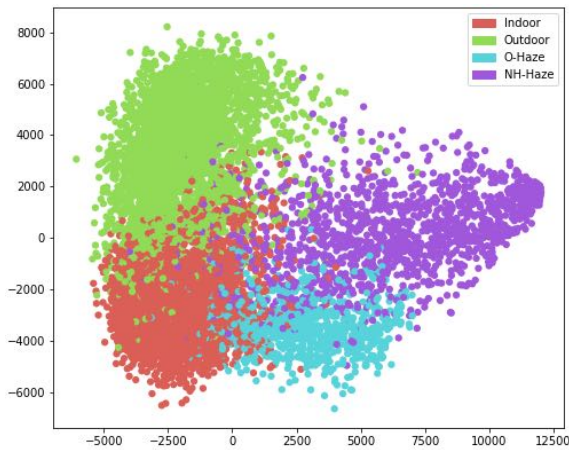


Figure 2: Visualization of image features using t-SNE [42]. Image features are obtained using a ResNet18 [16] pre-trained on ImageNet [9]. The fact that the features are clustered around four different centers shows clear discrepancies between the distributions of these datasets.

dataset has a specific distribution which might be significantly different from that of another dataset (see Figure 2). The representative examples from the four datasets under consideration are shown in Figure 3, where one can observe that both the haze pattern and background scene of the four datasets are significantly different from each other. In the RESIDE indoor and outdoor datasets, the haze pattern is homogeneous but the background scenes are vastly different (indoor vs. outdoor environments). In the O-Haze and NH-Haze datasets, background scenes are consistent (outdoor environments) but the haze patterns have remarkable distinctions. Against this backdrop, learning a general dehazing model on multiple datasets can be categorized as a multi-domain learning (MDL) problem.

In this paper, we propose a method that can enable a single dehazing model to cope with multiple domains. Here, each domain is formed by a dataset with a distinctive haze pattern and scene. Our goal is to find a model that can minimize the risk on the collection of domains for the dehazing task. Note that our problem definition is significantly different from the related field of domain adaptation and multi-task learning in the sense that the former aims to minimize the risk on a specific target domain while the latter performs optimization on a collection of tasks paired with a single domain. In principle, one can address the reformulated multi-domain dehazing problem by following the common practice in MDL. However, this requires designing sophisticated neural network structures with domain-specific modules, which is a highly non-trivial and cumbersome task in general.

To alleviate the design burden, we propose a novel MDL approach for single image dehazing by helping a given de-



Figure 3: Representative examples from RESIDE indoor/outdoor, O-Haze and NH-Haze.

hazing network to adapt to a specific domain when it is needed. To achieve this, we propose a method to adjust the dehazing network during the testing phase. In this method, the parameters of the network are optimized using a self-supervised loss function which is basically provided by another entity called the *helper network*. This network is designed to learn diverse haze patterns using paired hazy and haze-free images (across multiple domains) and output a reconstructed version of the hazy image that is fed into it. At the test time, the helper network uses its knowledge to assess the quality of the output of the dehazing network, which is a dehazed image. In other words, this image together with its corresponding hazy counterpart are given to the helper network as its inputs. If the output of the helper network is close to the hazy image, then a small reconstruction loss is expected. However, if the dehazed image is defective, then a large reconstruction loss may be derived at the helper network. Considering the fact that the quality of dehazed image can be represented by the reconstruction loss of the helper network, we update the parameters of the dehazing network by minimizing this loss function.

Now a natural question arise: How to guarantee that the end-to-end performance of the dehazing network is eventually optimized by minimizing the reconstruction loss of the helper network. In order to ensure the consistency between the objectives of two networks, we adopt the meta-learning approach [8, 13, 41]. Here, the goal of meta-learning is to adjust the parameters of the dehazing network by minimizing the reconstruction loss of the helper network so that the dehazing output based on the adjusted parameters better matches the ground-truth haze-free image.

Our contributions can be summarized as follows. Firstly, we point out a largely unnoticed phenomenon in single image dehazing, namely, a model trained on multiple datasets

exhibits compromised performance on individual datasets. This leads to the formulation of designing a dehazing model for distribution-wise distinctive datasets as a MDL problem. Secondly, we put forward a solution to this problem by introducing a test-time training approach for better adapting the dehazing network to every single observation. Finally, we provide extensive experiments to demonstrate the effectiveness of our proposed method in addressing the multi-domain dehazing problem.

2. Related Works

2.1. Single Image Dehazing

Most of the conventional single image dehazing methods [4, 12, 15, 53] are based on the estimation of parameters in the atmospheric scattering model (ASM) using statistical priors. However, they are not robust in dealing with complex real scenes. Recently, there has been a significant progress in the single image dehazing by using the deep learning approach. Although, [5, 22, 29, 34, 40, 47] still rely on the ASM, they propose to adopt a neural network to first estimate the transmission and then restore it. Due to the limitations of the ASM which make it not to be an effective method in modeling complicated haze patterns, other works [6, 11, 17, 24, 26, 27, 31, 32, 35, 46, 51] are designed using the end-to-end deep neural networks to directly learn the mapping from hazy images to haze-free counterparts. Another line of works [7, 21, 37] mainly focus on enabling the deep learning system to deal with natural hazy images. For example, in [37], a model is trained on multiple synthetic domains and the performance is evaluated on a specific real dataset. Our work is different from [37] in the sense that the performance of our proposed network is verified over different domains.

2.2. Multi-domain Learning

Multi-domain learning (MDL) aims to enable a model with the ability to minimize the risk across multiple domains. Usually, the model parameters can be divided into two distinctive parts according to their functionalities. Specifically, while one part focuses on learning the shared representations across different domains, the other part learns the domain-specific mapping relations [19, 28, 33, 36, 48]. Recent works consider developing a general system without explicitly learning the cross-domain or domain-specific representations. For example, [38] proposes a single model-based method to address problems in medical imaging. It uses the meta-learning to dynamically estimate hyperparameters in the loss functions. Notice the difference between the meta objectives of [38] and ours, i.e., our meta objective is designed to learn the consistency across losses. [43] introduces a universal object detector consisting of a single network using domain attention modules. These

modules can activate the model parameters that are responsible for a particular domain. This approach still relies on a precise network design. However, our proposed approach is model-agnostic and can be used in a plug-and-play manner.

2.3. Meta-learning for Image Restoration

Meta-learning, also known as *learning to learn*, has attracted attention in the computer vision community, recently. Especially, the model-agnostic meta-learning (MAML) [13] is widely utilized in image restoration to improve the generalization ability of deep neural networks. For example, [30, 39] adopt MAML for super-resolution. The meta objective is to learn a model that can quickly adapt to novel scenes. [8] proposes to use the meta-auxiliary learning [41] for the test-time dynamic scene deblurring. Besides the obvious difference in the treated problems, our work offers two general insights regarding test-time training not present in [8]. 1) test-time training can be realized by building a helper network, detached from the main network (possibly off-the-shelf), to provide self-supervision during the test time. This idea is broadly applicable. It lifts the burden of jointly addressing the primary and auxiliary tasks in one framework and clears the way for wide adoption of test-time training. 2) The helper network should be designed by considering the special characteristics of the problem at hand (e.g., ASM is unique to dehazing) to maximize benefits of test-time training.

3. Methodology

Assume that we have a collection of N dehazing domains $\{D_i\}_{i=1}^N$ with M paired hazy and haze-free images $\{I_i, J_i\}_{i=1}^M$. We aim to train a dehazing model f_{θ_d} that is able to perform well on all domains. However, as mentioned before, we find through experiments that the model trained on a single domain can usually outperform the model trained on multiple domains.

In this section, we present one possible solution to address this problem. Firstly, we train the dehazing network f_{θ_d} using all image pairs from N domains by minimizing the following commonly used loss function:

$$\mathcal{L}_{dehaze}(\hat{J}, J) = \mathcal{L}_{smooth}(\hat{J}, J) + \gamma \mathcal{L}_{Per}(\hat{J}, J), \quad (1)$$

where \hat{J} and J represent the dehazed and haze-free images, respectively, and \mathcal{L}_{smooth} and \mathcal{L}_{Per} represent smooth L_1 and perceptual losses [18], respectively. The parameter γ is used to get a weighted combination of the two loss functions. Note that, the dehazing network can be any existing well-designed model.

Secondly, we will develop a helper network g_{θ_h} to learn the haze patterns in the following Section 3.1. It is basically employed to determine the quality of the dehazed image \hat{J}

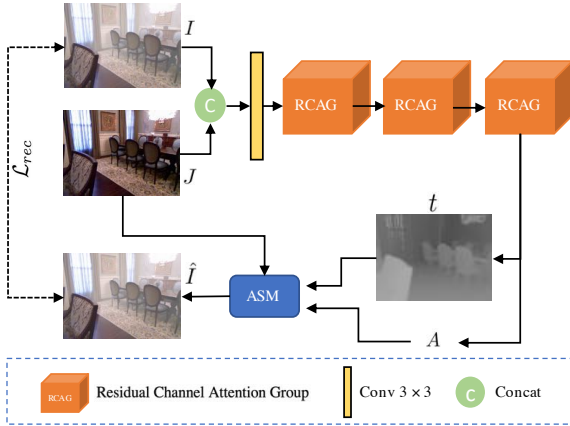


Figure 4: The helper network for learning the haze patterns. It consists of 3 stacked residual channel attention groups adopted from RCAN [52].

which is generated by the dehazing network. That is, if \hat{J} is close to J , then a small reconstruction loss is expected at the helper network. During the test time, the dehazing network can update its parameters by minimizing the reconstruction loss of the helper network as will be discussed in Section 3.2. Although, this method helps the dehazing network to get an improved dehazing loss on the particular hazy inputs, however, an improved performance is not generally guaranteed. To address this problem, we finally propose a meta-learning approach in the following Section 3.3 to associate the dehazing and reconstruction losses with each other. Once the meta-training is complete, our setup is able to conduct the test time training which enables the dehazing network in producing clearer dehazing results in a self-supervised manner.

3.1. Learning the Haze Patterns

As discussed before, given a dehazed image \hat{J} , our goal is to effectively determine its quality and guide the dehazing network to produce a clearer counterpart. To achieve this, we build a helper network to explicitly learn the haze patterns across multiple domains and use this pre-learned knowledge to determine the quality of the dehazed images. The network structure is shown in Figure 4, where it takes paired hazy and haze-free images as the inputs and it outputs two key components of a haze pattern, i.e., the transmission $t(x)$ and global atmospheric light A . Then, we use the modified atmospheric scattering model (ASM) to reconstruct the hazy image by the following:

$$\hat{I}(x) = J(x)t(x) + A(1 - t(x)), \quad (2)$$

where $\hat{I}(x)$ and $J(x)$ are the reconstructed hazy and haze-free images at pixel x , respectively. Note that in the conventional ASM, $t(x)$ is defined as $t(x) = e^{-\beta d(x)}$, where $d(x)$ denotes the scene depth and the parameter β is a con-

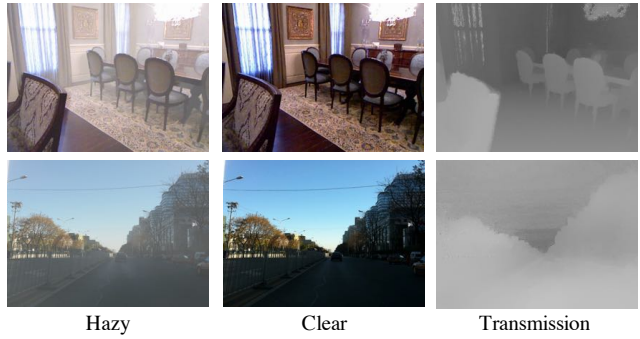


Figure 5: Illustrations of learned transmission map.

stant which controls the thickness of haze. However, in our proposed modification to this model, the parameter β is no longer assumed to be a constant. As shown in Figure 4, the transmission $t(x)$ is derived through a neural network. Using this method, it is possible to derive the transmission even when the haze is non-homogeneous. An illustration of learned transmission is shown in Figure 5. It is also worth emphasizing that one purpose of adopting the ASM model for the hazy image reconstruction is to avoid a trivial solution that the neural network can directly paste the input hazy image to the output side, without processing the hazy and haze-free images together.

Since the neural network and ASM model are fully differentiable, we can optimize the network on the combined domain using the loss (see Figure 4):

$$\mathcal{L}_{rec}(\hat{I}, I) = \mathcal{L}_{smooth}(\hat{I}, I) + \gamma \mathcal{L}_{Perc}(\hat{I}, I), \quad (3)$$

where I and \hat{I} denote the hazy and reconstructed hazy images, respectively.

Once the helper network is trained to converge, it is able to reconstruct the hazy image by jointly employing hazy and haze-free images.

3.2. Dehazing Using Haze Reconstruction

During the test time, the dehazing network outputs a dehazed image \hat{J} which might be different from the ground truth haze-free image J . So, feeding \hat{J} and I (the hazy image) to the helper network as the inputs, it outputs a haze pattern which can result in a defective reconstructed hazy image I^+ . And therefore, the corresponding loss $\mathcal{L}_{rec}(I^+, I)$ is larger than $\mathcal{L}_{rec}(\hat{I}, I)$ where \hat{I} is the output of the helper network when I and J (the hazy and haze-free images) are fed into it.

Inspired by this, we perform test-time training on the dehazing network to minimize $\mathcal{L}_{rec}(I^+, I)$. Specifically, we update the dehazing network in few steps using $\mathcal{L}_{rec}(I^+, I)$ by the following:

$$\hat{\theta}_d \leftarrow \theta_d - \lambda_1 \nabla_{\theta_d} \mathcal{L}_{Rec}(f_{\theta_h}(f_{\theta_d}(I), I), I), \quad (4)$$

where λ_1 denotes the learning rate. Here, $\hat{\theta}_d$ is the updated weights of the dehazing model according to the reconstruction loss. Note that, the test-time training of the dehazing network is purely self-supervised by using the hazy image I , i.e., it does not require any manual labeling. Ideally, by minimizing $\mathcal{L}_{rec}(I^+, I)$, it can be expected that the image \hat{J} produced by the updated dehazing network gets closer to the ground truth J over time. Therefore, we can finally produce improved dehazed images using the updated $\hat{\theta}_d$.

Despite the idea of developing the helper network to determine if the output dehazing results are clean counterparts of input hazy images, one might have a question that “is it true that the dehazing network can always benefit from the supervision of the reconstruction loss?” Unfortunately, we will show in our ablation studies at Section 4.5 that minimizing $\mathcal{L}_{rec}(I^+, I)$ is not always equivalent to minimizing $\mathcal{L}_{dehaze}(\hat{J}, J)$, which means that even if sometimes the dehazing network’s produced output \hat{J} steps far away from the ground truth J , it may be adopted by the helper network to reconstruct a hazy image \hat{I} which is closer to I . The problem might be that the two losses are not consistent with each other and they lack explicit connections.

3.3. Learning Meta-objective

Inspired by the recent meta-learning approach [8, 41], where the test-time training is conducted via an auxiliary loss, we are further motivated to propose a meta-learning method across models. The goal of the meta-training is to learn the dehazing model parameters so that the dehazing loss is spontaneously minimized by optimizing the parameters based on the reconstruction loss.

Before the meta-training, we pre-train the dehazing and helper networks (θ_d and θ_h). They are independently trained by Eq. (1) and Eq. (3), respectively. Given a paired training data (I_i, J_i) , we update the dehazing network using the reconstruction loss as follows:

$$\hat{\theta}_d = \theta_d - \lambda_1 \nabla_{\theta_d} \mathcal{L}_{rec}(f_{\theta_h}(f_{\theta_d}(I_i), I_i), I_i) \quad (5)$$

Intuitively, this update enables the dehazing network to produce results that can be adopted by the helper network to get an improved reconstructed hazy image. Considering the fact that we intend to use $\hat{\theta}_d$ to minimize the dehazing loss, we update the dehazing network by encouraging the performance of dehazing network to be maximized if the helper network can employ the dehazed image to get an improved reconstructed hazy image.

To that end, our meta objective is formally defined as:

$$\arg \min_{\theta_d} \mathcal{L}_{dehaze}(f_{\hat{\theta}_d}(I_i), J_i). \quad (6)$$

Note that the dehazing loss is computed using the dehazed image $f_{\hat{\theta}_d}(I_i)$ produced by updated dehazing net-

Algorithm 1: Meta Training

Require: Pre-trained networks: $f_{\theta_d}, g_{\theta_h}$
Require: Learning rates λ_1 and λ_2
Output : Meta-learned model parameter θ_d

while not converge do

 Sample a batch of training data in $\{I_i, J_i\}_{i=1}^M$;

for each I_i **do**

 Compute updated parameters $\hat{\theta}_d$:

$\hat{\theta}_d = \theta_d - \lambda_1 \nabla_{\theta_d} \mathcal{L}_{rec}(g_{\theta_h}(f_{\theta_d}(I_i), I_i), I_i)$

Update:

$\theta_d \leftarrow \theta_d - \lambda_2 \nabla_{\theta_d} \mathcal{L}_{dehaze}(f_{\hat{\theta}_d}(I_i), J_i)$

work $f_{\hat{\theta}_d}$, while the optimization is performed on θ_d . Eq. (6) can be achieved using the gradient descent as follows:

$$\theta_d \leftarrow \theta_d - \lambda_2 \nabla_{\theta_d} \mathcal{L}_{dehaze}(f_{\hat{\theta}_d}(I_i), J_i) \quad (7)$$

where λ_2 denotes the learning rate. The overall meta-learning procedure is summarized in the Algorithm. 1.

Finally, after the meta-training, the updated dehazing and helper networks are ready to use. We can follow the procedure in Section 3.2 to conduct the test-time training.

4. Experimental Results

4.1. Datasets and Evaluation Metrics

Our experiments are conducted on widely used dehazing datasets, including O-Haze [3], NH-Haze [2] and RESIDE indoor/outdoor [23]. O-haze contains 40 image pairs, where the first 35 pairs are used for the training and the rest 5 pairs are adopted for the testing. NH-Haze consists of two variants that are released in 2020 and 2021. We form our NH-Haze dataset by combining both of them. For NH-Haze 2020, we adopt the official train, test split. As the validation and test data of NH-Haze 2021 is not released publicly, we take the first 22 pairs for the training and the other 3 pairs for the testing. Finally, our NH-Haze has a total of 67 training pairs and 8 testing pairs. RESIDE dataset is a benchmark for single image dehazing. We follow DADN [37] to form the training set by selecting 3000 indoor pairs and 3000 outdoor pairs and cropping them to the size of 256×256 . For the testing, we adopt the Synthetic Objective Testing Set (SOTS) of RESIDE. The quantitative evaluation metrics used in this paper are PSNR and SSIM [44]. Besides, we also conduct out-of-domain validation on a real-world dehazing dataset introduced in [12].

4.2. Implementation details

We first pre-train the selected dehazing networks and our helper network on the combined dataset, which consists of the aforementioned O-Haze, NH-Haze and RE-

Methods		Indoor	Outdoor	O-Haze	NH-Haze	# Params	Runtime (s)
GDN [26]	Single-domain	27.79/0.953	28.93/0.972	23.23/0.808	19.14/0.710	3.84M	0.028
	Multi-domain	26.67/0.951	27.18/0.962	23.13/0.747	18.93/0.716	0.96M	0.028
	Ours	26.83/0.952	27.26/0.961	23.21/0.747	19.04/0.716	1.34M	1.043
MSBDN [11]	Single-domain	28.89/0.956	30.76/0.977	24.95/0.824	19.82/0.747	12.56M	0.153
	Multi-domain	28.53/0.961	30.31/0.973	23.97/0.764	19.51/0.725	3.14M	0.153
	Ours	28.68/0.961	30.42/0.974	24.14/0.766	19.62/0.726	3.52M	1.828
DW-GAN [14]	Single-domain	29.65/0.963	31.75/0.978	24.50/0.793	21.83/0.769	206.04M	0.076
	Multi-domain	28.84/0.941	31.21/0.974	24.02/0.789	20.44/0.763	51.51M	0.076
	Ours	28.95/0.942	31.39/0.974	24.13/0.789	20.53/0.762	51.89M	1.621

Table 1: Quantitative comparison of the dehazing results on multiple datasets using different training schemes. The term “single-domain” denotes that the method is trained on a single dataset and evaluated on the relative one; the term “multi-domain” represents that the network is trained using the combined dataset; the term “ours” denotes the results adopting the proposed test-time training. Accuracies are presented in the form of PSNR/SSIM.

SIDE indoor/outdoor datasets. The initial learning rate is set to 10^{-4} for the training of all networks except the Grid-DehazeNet [26], which is set to 10^{-3} . During the meta-training, the learning rates λ_1 and λ_2 in Eq. (5) and (7) are fixed to be 1.25×10^{-5} and 2.5×10^{-5} , respectively. The Adam optimizer [20] is used in both pre-training and meta-training with the default values of $\beta_1 = 0.9$ and $\beta_2 = 0.99$. In the test-time training phase, we perform 5 gradient updates on each hazy image and report the final accuracy. All our experiments are conducted on Nvidia V100 GPUs.

4.3. Degradation from Multi-domain Learning

As we denoted, a neural network trained on multiple domains is usually suboptimal when tested on each individual domain. Here, we provide quantitative results to investigate this phenomenon. Our experiments are conducted using three popular learning-based methods, i.e., GDN [26], MSBDN [11] and DW-GAN [14]. To be specific, we first implement the single-domain training and testing, where a dehazing network is trained on a single dataset and tested on the relative one. Then, we take the same network to conduct the multi-domain learning, where the dehazing network is trained on the combined dataset that consists of all datasets introduced in Section 4.1. The results of the single-domain learning and multi-domain learning are shown in Table 1 denoted by “single-domain” and “multi-domain”. It can be observed that both PSNR and SSIM of the multi-domain learned method are smaller than that of the single-domain learned. This fact indicates that simply collecting data for the dataset augmentation is not always useful.

4.4. Test-time Training on the State-of-the-art

To illustrate the effectiveness of our proposed method, we conduct quantitative and qualitative experiments using GDN [26], MSBDN [11] and DW-GAN [14] as the dehazing network. The experiments aim to show that our helper network can be helpful in boosting the performance of the existing approaches. Note that, our method is employed for the dehazing networks trained on multiple domains.

Quantitative Improvements: Table 1 summarizes the PSNR and SSIM measures on all four datasets. The performance of a dehazing network using our proposed test-time training is reported under the term “ours”. Thanks to different structures of the three networks, we can observe variations in the improvement of PSNR. For example, in the indoor testing set, our method can improve the PSNR of multi-domain learned GDN by 0.16dB, and improvements can be observed for MSBDN and DW-GAN, where the PSNR is increased by 0.15dB and 0.11dB, respectively. It can be easily checked throughout the table that our proposed test-time training can always improve the performance of the network trained on multiple domains. This further indicates the judicious model-agnostic property of our proposed test-time training method.

Qualitative Improvements: Figure 6 presents the dehazing results on O-HAZE and NH-HAZE. Here, we unfold the test-training process to provide a better understanding of our method. There are multiple problems shown in the initial results that can be fixed by conducting the proposed test-time training. In the first and second rows of Figure 6, we can notice that severe artifacts are added to the sky region of the dehazed images. Surprisingly, these artifacts can be removed gradually by few gradient updates. In the third and last rows, the results before updates are still hazy. However, our method is able to remove the haze from the initial results. Finally, other instances of color distortion are shown in the fourth and fifth rows, where after 5 updates, the dehazing network can produce more elegant images.

Algorithm Efficiency: Here, we investigate the efficiency of the proposed method in terms of the number of parameters. The last column of Table 1 reports the number of parameters that is required to dehaze on four domains. Thanks to the lightweight design of our helper network, by integrating our method with the current dehazing networks, the total number of parameters is still comparable to that of a single dehazing network. Especially when our method uses DW-GAN, there is a negligible increase in the number of parameters, however the PSNR is boosted by an aver-

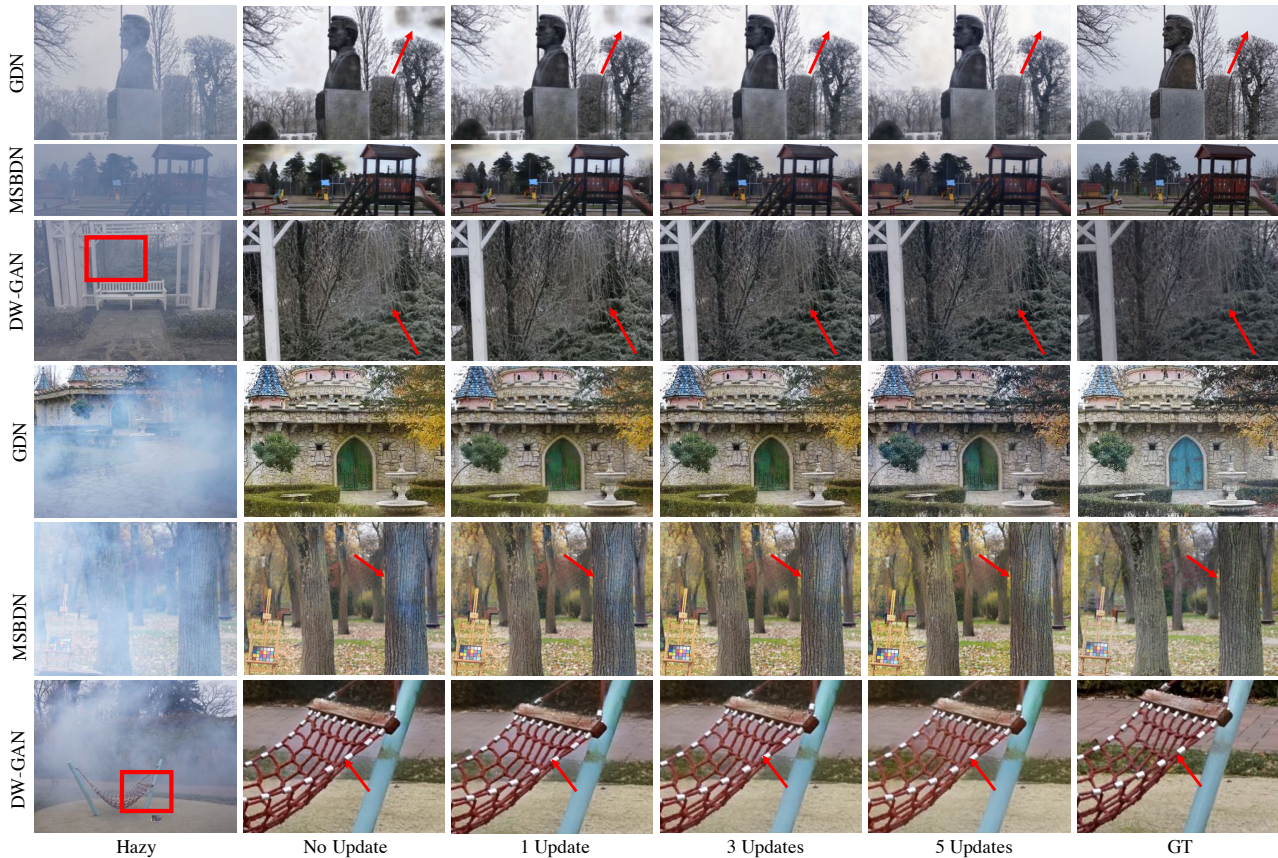


Figure 6: Qualitative results on O-Haze and NH-Haze datasets. The image samples of the first three rows are from O-Haze, and the others are from NH-Haze. After few gradient updates, the proposed test-time training can improve the image quality.

age of 0.12dB. Moreover, considering the fact that we have four domains, if a model is separately trained on each domain, the deployment of four models is extremely memory-consuming as four collections of parameters need to be stored; this fact can be verified by observing the total number parameters for “signal-domain” in Table 1. However, we also find that using test-time training is slower than its one-shot inference counterpart. This issue can be alleviated via a more efficient implementation of test-time training.

4.5. Ablation Studies

All our ablation studies are conducted using GDN [26] (baseline). In order to illustrate the effectiveness of the meta-learning approach, we then introduce three experimental setting to reveal this fact. They are presented as the following: (a) **GDN+Helper**: where the helper network is directly used to provide the test-time supervision; (b) **GDN+Helper+joint-training**: the training of GDN simultaneously employs both Eq. (1) and Eq. (3) as the loss function, i.e., the GDN is optimized in a manner such that both dehazing and hazy reconstruction losses are minimized; (c) **GDN+Helper+meta-learning**: where we use our proposed method. The quantitative results of the three methods are

Methods	Indoor	Outdoor	O-Haze	NH-Haze
GDN	26.67/0.951	27.18/0.962	23.13/0.747	18.93/0.716
(a)	26.69/0.950	27.16/0.961	22.89/0.743	18.65/0.712
(b)	26.71/0.951	27.19/0.962	22.94/0.745	18.71/0.713
(c)	26.83/0.952	27.26/0.961	23.21/0.747	19.04/0.716

Table 2: Ablation studies on our method. Numbers are presented in the form of PSNR/SSIM. (a), (b) and (c) denote three different methods introduced in Section 4.5.

presented in Table 2. It can be observed that without meta-training to associate the objectives of dehazing and helper networks, the helper network cannot generally assist the dehazing network to further converge on an unseen image.

4.6. Out-of-domain Validation on Real Scenes

For the above experiments, we assume that the hazy and haze-free images are from the same domain, while ignoring the out-of-domain (OOD) problem. Here, we take the real data [12] as an example to validate the domain generalization ability of our proposed method. To conduct comparison, we choose four state-of-the-art single image dehazing algorithms, i.e., DCP [15], DehazeNet [5], DADN [37] and AECR-NET [46]. Note that, the training of these methods

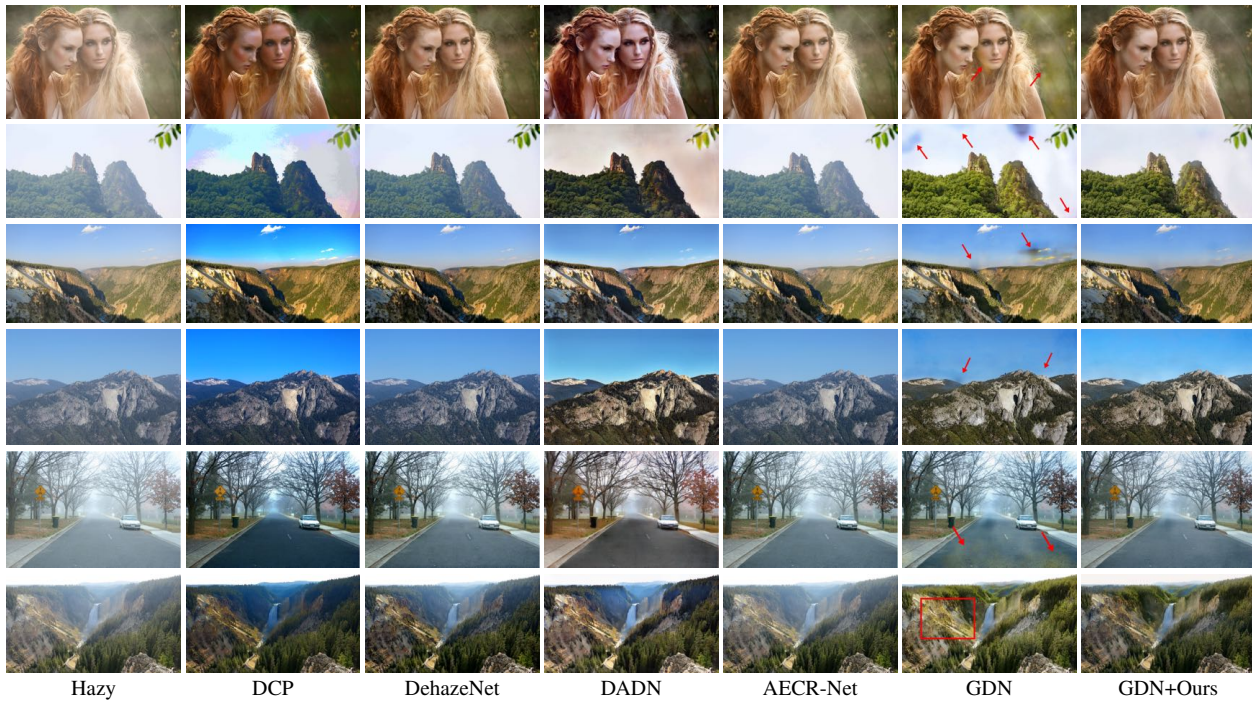


Figure 7: Comparison of the dehazing methods on real hazy images from [12].

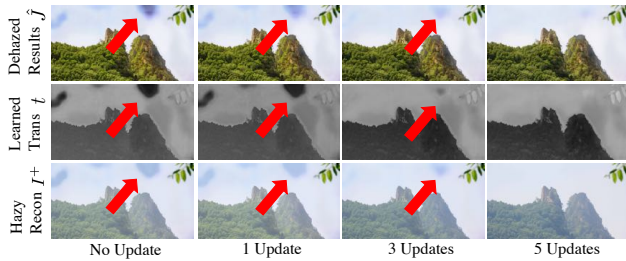


Figure 8: Breakdown of dehazing on a real hazy image.

follows a common setting to use RESIDE indoor and outdoor datasets. In addition, GDN [26] is trained using the combined dataset as mentioned in Section 4.1.

There are two notes that should be mentioned, see Figure 7. First, since GDN is trained on the combined dataset (including O-HAZE and NH-Haze datasets), its dehazed images are usually cleaner and visually pleasing. For example, the third row shows a mountain covered by haze. Here, GDN completely removes the haze from the mountain, while others cannot remove the haze effectively and suffer from color distortion. Despite this success, GDN also generates severe artifacts in the sky region. This reminds us that although multi-domain learning is beneficial in some places but it still needs further research to be employed for removing haze from all kinds of scenes. Second, comparing our outputs (GDN+Ours) with those of vanilla GDN, it can be easily observed that the artifacts are gone and our dehazed images look more natural. Another example is in the last row, where GDN paints the mountains to be yellow, while our result presents a more natural color.

Besides, Figure 8 gives a breakdown of dehazing on a real image that is presented in the second row of Figure 7. We can observe that the reconstructed hazy image reveals the potential issues with the dehazed image. By minimizing the hazy reconstruction loss at test time, the dehazed results and hazy reconstructions are improved simultaneously.

5. Conclusion

In this paper, we reveal a critical problem that has not been considered in single image dehazing, that is, a dehazing network trained on multiple domains can perform worse than that trained only on a single domain. Based on this observation, we formulate the problem into a multi-domain learning setup, where a single model should be designed carefully to perform well on multiple domains. To address this issue, we propose a helper network to provide self-supervision to the dehazing network and improve its performance during the test time. A meta-learning approach has also been introduced to handle the problem that the supervision signal from the helper network cannot always help the dehazing network to gain an improved performance. Extensive experiments and analyses strongly support both our observation and the proposed method.

Limitation and future works. Due to the resource constraint, our current experiments are conducted using four representative datasets and three popular dehazing networks. It is preferable to have a more comprehensive evaluation of our method by considering a larger collection of datasets and dehazing networks.

References

- [1] Md Zahangir Alom, Tarek M Taha, Chris Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Mahmudul Hasan, Brian C Van Essen, Abdul AS Awwal, and Vijayan K Asari. A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8(3):292, 2019. **1**
- [2] Codruta O. Ancuti, Cosmin Ancuti, and Radu Timofte. NH-HAZE: an image dehazing benchmark with non-homogeneous hazy and haze-free images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, IEEE CVPR 2020, 2020. **1, 5**
- [3] 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 *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 867–8678, 2018. **1, 5**
- [4] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1674–1682, 2016. **3**
- [5] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016. **1, 3, 7**
- [6] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua. Gated context aggregation network for image dehazing and deraining. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1375–1383, 2019. **3**
- [7] Zeyuan Chen, Yangchao Wang, Yang Yang, and Dong Liu. Psd: Principled synthetic-to-real dehazing guided by physical priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7180–7189, June 2021. **3**
- [8] Zhixiang Chi, Yang Wang, Yuanhao Yu, and Jin 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*, pages 9137–9146, 2021. **2, 3, 5**
- [9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. **2**
- [10] Qili Deng, Ziling Huang, Chung-Chi Tsai, and Chia-Wen Lin. Hardgan: A haze-aware representation distillation gan for single image dehazing. In *European Conference on Computer Vision*, pages 722–738. Springer, 2020. **1**
- [11] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2157–2167, 2020. **1, 3, 6**
- [12] Raanan Fattal. Dehazing using color-lines. *ACM transactions on graphics (TOG)*, 34(1):1–14, 2014. **3, 5, 7, 8**
- [13] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pages 1126–1135. PMLR, 2017. **2, 3**
- [14] Minghan Fu, Huan Liu, Yankun Yu, Jun Chen, and Keyan Wang. Dw-gan: A discrete wavelet transform gan for nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 203–212, 2021. **1, 6**
- [15] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010. **3, 7**
- [16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. **2**
- [17] Ming Hong, Yuan Xie, Cuihua Li, and Yanyun Qu. Distilling image dehazing with heterogeneous task imitation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3462–3471, 2020. **3**
- [18] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution, 2016. **3**
- [19] Mahesh Joshi, Mark Dredze, William Cohen, and Carolyn Rose. Multi-domain learning: when do domains matter? In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1302–1312, 2012. **3**
- [20] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. **6**
- [21] Boyun Li, Yuanbiao Gou, Jerry Zitao Liu, Hongyuan Zhu, Joey Tianyi Zhou, and Xi Peng. Zero-shot image dehazing. *IEEE Transactions on Image Processing*, 29:8457–8466, 2020. **3**
- [22] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. An all-in-one network for dehazing and beyond. *ArXiv*, abs/1707.06543, 2017. **1, 3**
- [23] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018. **1, 5**
- [24] Runde Li, Jinshan Pan, Zechao Li, and Jinhui Tang. Single image dehazing via conditional generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8202–8211, 2018. **3**
- [25] Jing Liu, Haiyan Wu, Yuan Xie, Yanyun Qu, and Lizhuang Ma. Trident dehazing network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. **1**
- [26] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Grid-dehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7314–7323, 2019. **1, 3, 6, 7, 8**
- [27] Zheng Liu, Botao Xiao, Muhammad Alrabeiah, Keyan Wang, and Jun Chen. Generic model-agnostic convolutional neural network for single image dehazing. *arXiv preprint arXiv:1810.02862*, 2018. **3**

- [28] Hyeonseob Nam and Bohyung Han. Learning multi-domain convolutional neural networks for visual tracking. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4293–4302, 2016. 3
- [29] Y. Pang, J. Nie, J. Xie, J. Han, and X. Li. Bidnet: Binocular image dehazing without explicit disparity estimation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5930–5939, Los Alamitos, CA, USA, jun 2020. IEEE Computer Society. 3
- [30] Seobin Park, Jinsu Yoo, Donghyeon Cho, Jiwon Kim, and Tae Hyun Kim. Fast adaptation to super-resolution networks via meta-learning. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16*, pages 754–769. Springer, 2020. 3
- [31] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07):11908–11915, Apr. 2020. 1, 3
- [32] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8152–8160, 2019. 3
- [33] Sylvestre-Alvise Rebuffi, Hakan Bilen, and Andrea Vedaldi. Learning multiple visual domains with residual adapters. *arXiv preprint iclararXiv:1705.08045*, 2017. 3
- [34] Wenqi Ren, Sibao Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *ECCV*, 2016. 3
- [35] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3253–3261, 2018. 1, 3
- [36] Alice Schoenauer-Sebag, Louise Heinrich, Marc Schoenauer, Michele Sebag, Lani F Wu, and Steve J Altschuler. Multi-domain adversarial learning. *arXiv preprint arXiv:1903.09239*, 2019. 3
- [37] Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain adaptation for image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2808–2817, 2020. 3, 5, 7
- [38] Anthony Sicilia, Xingchen Zhao, Davneet S Minhas, Erin E O’Connor, Howard J Aizenstein, William E Klunk, Dana L Tudorascu, and Seong Jae Hwang. Multi-domain learning by meta-learning: Taking optimal steps in multi-domain loss landscapes by inner-loop learning. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 650–654. IEEE, 2021. 3
- [39] Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho. Meta-transfer learning for zero-shot super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3516–3525, 2020. 3
- [40] Taeyong Song, Youngjung Kim, Changjae Oh, and Kwanghoon Sohn. Deep network for simultaneous stereo matching and dehazing. In *BMVC*, page 5, 2018. 3
- [41] Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei Efros, and Moritz Hardt. Test-time training with self-supervision for generalization under distribution shifts. In *International Conference on Machine Learning*, pages 9229–9248. PMLR, 2020. 2, 3, 5
- [42] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008. 2
- [43] Xudong Wang, Zhaowei Cai, Dashan Gao, and Nuno Vasconcelos. Towards universal object detection by domain attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7289–7298, 2019. 3
- [44] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 5
- [45] Haiyan Wu, Jing Liu, Yuan Xie, Yanyun Qu, and Lizhuang Ma. Knowledge transfer dehazing network for nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 1
- [46] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, 2021. 3, 7
- [47] Dong Yang and Jian Sun. Proximal dehaze-net: A prior learning-based deep network for single image dehazing. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 702–717, 2018. 3
- [48] Yongxin Yang and Timothy M Hospedales. A unified perspective on multi-domain and multi-task learning. *arXiv preprint arXiv:1412.7489*, 2014. 3
- [49] Yankun Yu, Huan Liu, Minghan Fu, Jun Chen, Xiyao Wang, and Keyan Wang. A two-branch neural network for non-homogeneous dehazing via ensemble learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 193–202, 2021. 1
- [50] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3194–3203, 2018. 1
- [51] Jing Zhang and Dacheng Tao. Famed-net: A fast and accurate multi-scale end-to-end dehazing network. *IEEE Transactions on Image Processing*, 29:72–84, 2020. 3
- [52] Yulun Zhang, Kungpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 4
- [53] Qingsong Zhu, Jiaming Mai, and Ling Shao. Single image dehazing using color attenuation prior. In *BMVC*. Citeseer, 2014. 3