Towards Domain Invariant Single Image Dehazing

Pranjay Shyam[1]**, Kuk-Jin Yoon** [2]**, Kyung-soo Kim** [1]

[1] Mechatronics, Systems and Control Lab
[2] Visual Intelligence Lab
Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST)
Daejeon, Republic of Korea, 34141
{pranjayshyam, kjyoon, kyungsookim}@kaist.ac.kr

Abstract

Presence of haze in images obscures underlying information, which is undesirable in applications requiring accurate environment information. To recover such an image, a dehazing algorithm should localize and recover affected regions while ensuring consistency between recovered and its neighboring regions. However owing to fixed receptive field of convolutional kernels and non uniform haze distribution, assuring consistency between regions is difficult. In this paper, we utilize an encoder-decoder based network architecture to perform the task of dehazing and integrate an spatially aware channel attention mechanism to enhance features of interest beyond the receptive field of traditional conventional kernels. To ensure performance consistency across diverse range of haze densities, we utilize greedy localized data augmentation mechanism. Synthetic datasets are typically used to ensure a large amount of paired training samples, however the methodology to generate such samples introduces a gap between them and real images while accounting for only uniform haze distribution and overlooking more realistic scenario of non-uniform haze distribution resulting in inferior dehazing performance when evaluated on real datasets. Despite this, the abundance of paired samples within synthetic datasets cannot be ignored. Thus to ensure performance consistency across diverse datasets, we train the proposed network within an adversarial prior-guided framework that relies on a generated image along with its low and high frequency components to determine if properties of dehazed images matches those of ground truth. We preform extensive experiments to validate the dehazing and domain invariance performance of proposed framework across diverse domains and report state-of-the-art (SoTA) results. The source code with pretrained models will be available at https://github.com/PS06/DIDH.

Introduction

Visibility degradations arising from environmental variations such as haze, smoke and fog affect image quality by concealing underlying information which is undesirable in applications where accurate surrounding information is necessary for safe operation such as autonomous vehicles, aerial robots and intelligent infrastructure. To overcome complications arising from deteriorations such as haze and fog, image dehazing has been extensively studied to recover clean image from its degraded version. Common approaches rely upon haze estimation using atmospheric scattering model (McCartney 1976; Narasimhan and Nayar 2000, 2002)(Eq. 1) to recover haze affected regions, that establishes a pixel-wise (x) relationship between ambient light intensity (A) and transmission matrix $t(x) = e^{-\beta d(x)}$ (representing fraction of light reaching camera sensor) by using scene depth $(d(x))$ and scattering coefficient (β), to generate a hazy image $(I(x))$ from a clean image $(J(x))$.

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

Traditional computer vision based dehazing algorithms relied upon handcrafted priors such as dark channel (He, Sun, and Tang 2010), color attenuation (Zhu, Mai, and Shao 2015), bi-channel (Jiang et al. 2017) and color lines (Fattal 2014; Berman, treibitz, and Avidan 2016) to estimate atmospheric light or transmission map to recover dehazed image, following atmospheric scattering model. However strong reliance on priors makes these methods vulnerable in scenarios when these priors don't hold. To avoid dependence on priors, different models leveraging the feature extraction capabilities of convolutional neural networks (CNNs) were proposed, following either atmospheric scattering model (He, Sun, and Tang 2010; Cai et al. 2016; Zhang and Patel 2018) or end-to-end approach (Ren et al. 2018; Li et al. 2017; Mei et al. 2018a; Chen et al. 2019; Engin, Genç, and Kemal Ekenel 2018) to estimate dehazed images.

Although learning based approaches represent current state-of-the-art (SoTA), they require large number of training samples accurately representing haze scenarios in different outdoor settings. Constructing such a large scale real dataset is both expensive and time consuming, thus atmospheric scattering model is used to generate synthetic haze corresponding to a clean image. However such approaches are limited in considering the effect of airborne particles on different wavelengths (Li et al. 2020) apart from presence of wind, resulting in difference between real and synthetic images in form of domain difference and haze distribution. This results in performance gap arising between real and synthetic haze trained models when evaluated upon either of the datasets.

To overcome the dual challenge of varying haze distribution and domain difference. In this paper, we propose a

framework for achieving domain and distribution invariant dehazing framework. We begin by focusing on achieving consistent performance irrespective of haze distribution and highlight the importance of localizing haze affected regions as a necessary step towards effective dehazing. To attain such characteristics, we concentrate upon data augmentation and architecture of the underlying CNN. Specifically, to generate non-uniform haze distribution on synthetic samples, we leverage greedy localized data augmentation proposed in (Shyam et al. 2021) that copies multiple patches of varying shapes and sizes from noisy image onto corresponding paired clean image to generate non-uniform noise patches. For our purpose, this approach results in generation of non-homogeneous haze. In order to accurately recover image regions affected by unknown haze distribution, we utilize an encoder-decoder framework built upon UNet (Ronneberger, Fischer, and Brox 2015), to aggregate and represent features across different scales into a higher order latent space which is subsequently be used by a decoder to reconstruct haze free image. To ensure information and color consistency between recovered patches with neighboring pixels we aggregate features from multiple scales using our proposed spatially aware channel attention mechanism and fuse these features into feature encoding obtained at encoder.

Motivated by observations of (Wang et al. 2020) that highlight domain gap arising due to sensitivities of CNN towards high frequency (HF) components within an image and the ability of adversarial training to incorporate domain invariant properties by focusing upon generalizable patterns within samples. We propose prior based dual discriminators, that use low frequency (LF) and high frequency components along with corresponding dehazed image to determine similarity between recovered image and ground truth. We thus summarize our contributions as

- Propose an end-to-end dehazing algorithm, directly recovering images affected by unknown haze distribution using spatially aware channel attention mechanism within CNN architecture to ensure feature enhancement and consistency between recovered and neighboring pixels.

- Integrate a local augmentation technique, to ensure the network learns to identify and recover haze affected regions in real and synthetic images.

- Perform exhaustive experiments to highlight performance inconsistencies between networks trained on synthetic and real datasets and attribute this to weak modeling of haze.

- To ensure consistent performance across synthetic and real datasets, prior based adversarial training mechanism is introduced that leverages LF and HF components within an image to ensure retention of color and structural properties within recovered image.

Related Works

Single Image Dehazing : Image dehazing algorithms can be categorized into model based and end-to-end. Model based algorithms utilize atmospheric scattering model to recover haze affected images either using prior or learning based approach. Prior based approaches such as (He, Sun, and Tang 2010) proposed using dark channel prior on the premise that pixel value within a color channel is close to zero in regions affected by haze, using which transmission map and atmospheric light within an image can be estimated. Other approaches (Zhu, Mai, and Shao 2015; Fattal 2014; Berman, treibitz, and Avidan 2016) devise priors such as color attenuation and color lines for estimating transmission map. Since prior based methods are sensitive towards environment variations, learning based approaches are utilized to leverage feature extraction capabilities of CNNs to estimate different components of atmospheric scattering model. Specifically, (Lu et al. 2016) used CNNs to estimate atmospheric light, (Cai et al. 2016) estimates transmission and (Yang et al. 2017; Li et al. 2018a) estimates both transmission and atmospheric light to recover regions affected by haze.

Recently learning based approaches have shown considerable performance improvement for recovering haze affected regions in an end-to-end manner. (Ren et al. 2018) proposed an encoder-decoder formulation to encode features from the hazy images which are then extrapolated by a decoder to reconstruct haze free images. (Qin et al. 2020; Mei et al. 2018a; Li et al. 2018a; Liu et al. 2019b) followed a similar approach by performing modifications in CNN network and loss functions. On contrary (Qu et al. 2019) proposed the task of dehazing as image-to-image translation and used a modified variant of Pix2Pix (Isola et al. 2017). To reduce reliance on paired dataset, (Engin, Genç, and Kemal Ekenel 2018) modified CycleGAN (Zhu et al. 2017) formulation for the task of dehazing. However these approaches are sensitive towards domain changes between synthetic and real datasets. To overcome this (Shao et al. 2020) proposed a domain adaptation mechanism to translate image from one domain to another thereby aiming to achieve the best of image translation and dehazing. In order generate more visually pleasing dehazed images, (Dong et al. 2020) proposed a fusion of frequency priors with image in adversarial learning framework. However, unlike prior approaches we emphasize on disentangling frequency information (LF and HF) from an image to extract multiple priors and independently learn association between LF (color) and HF (edge) components for retaining color and structural consistency, beyond the traditional loss functions.

Domain Invariance : Feature extraction capabilities of CNNs leads to their SoTA performance on various tasks. However performance inconsistencies arise when domain gap exists between test and train sets. To overcome such scenarios domain adaptation is proposed to perform either feature level (Tsai et al. 2018; Tzeng et al. 2017) or pixel level adaptation (Shrivastava et al. 2017; Bousmalis et al. 2017; Dundar et al. 2018). Feature level adaptation minimizes maximum mean discrepancy (Long et al. 2015) between source and target domain, while pixel level adaptation focuses upon image-to-image translation or style transfer to increase data in source or target. However reliance on target dataset makes this approach less favorable. An extension, domain generalization, focuses on techniques that provide consistent performance across unknown domains by
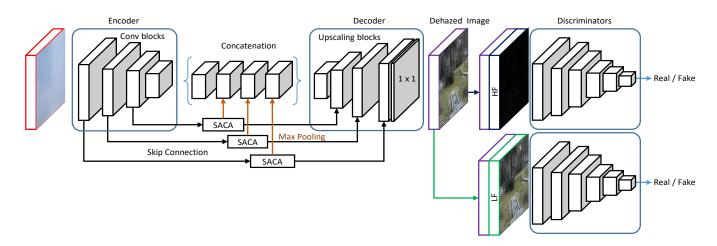
Figure 1: A simplified overview of complete framework covering dehazing and adversarial model

emphasizing on the task using stylization techniques (Somavarapu, Ma, and Kira 2020; Matsuura and Harada 2020), semantic features (Dou et al. 2019) or adversarial training (Li et al. 2018b; Agarwal, Chen, and Nguyen 2020). Using image stylization or semantic features is not possible since the input images are already affected by haze thereby obscuring underlying information. Thus we focus upon utilizing adversarial learning to achieve domain invariant performance as this method doesnt modify input image. We further lend support from observations by (Wang et al. 2020) on importance of LF and HF components within an image, where LF component capture structure and color information while the HF components capture edge information and effectively learning both results in achieving domain invariant performance.

Achieving Domain Invariant Dehazing

The overall structure of the proposed framework comprises of two parts namely greedy data augmentation (Fig. 2) and adversarial training framework (Fig. 1) comprising the dehazing and discriminator network.

Copy-Blend for Localized Data Augmentation

In a realistic setting, haze can vary wildly across regions within an image (Fig. 2 (f)). However synthetic datasets while providing a large number of paired samples arent able to account for such non-homogeneous variations (Fig. 2 (d)), leading to inaccurate recovery in such scenarios (Fig. 5). A simplistic approach to overcome such a scenario would be to increase the dataset to account for these variations which is costly and time consuming. Thus to train the network towards such diverse scenarios, we leverage greedy localized data augmentation technique proposed in (Shyam et al. 2021) to generate small hazy patches of random shape and size within clean images and task the CNN to recover these affected images. This allows utilization of both real and synthetic datasets for homogeneous and non-homogeneous dehazing. Fig. 2 demonstrates the working of this augmentation mechanism, along with a visual comparison to syn-
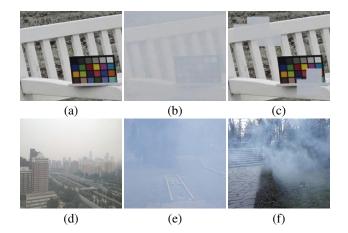

(a)   (b)   (c)
(d)   (e)   (f)

Figure 2: Given a patch of (a) a clean and (b) corresponding paired hazy image to generate (c) localized haze affected regions with maximum patch size set to 50×50. Samples showcasing (d) synthetic, (e) real homogeneous and (f) real non-homogeneous haze from SOTS-OUT, NTIRE-19 and NTIRE-20 datasets.

thetic, real and non-homogeneous haze examples.

Network Architecture

For recovering haze affected regions we build upon encoder-decoder architecture UNet. The encoder of the proposed dehazing network is tasked to represent noisy input image into a latent space using 4 convolution blocks (conv-blocks) that extract relevant features across different scales. Each conv-block represents 2 chains of convolutional filter of size 3×3, batch normalization layer and ReLU activation function. A max pool operation is performed after each convolution block to aggregate features while increasing the receptive field size for subsequent convolution blocks. In order to reconstruct noise-free image, 4 upscaling blocks, each comprising of pixel shuffle layer (Shi et al. 2016) followed by convolutional filter of size 3×3, batch normalization and ReLU activation and residual blocks are used. Features ob-

tained at each level of encoder are concatenated with features of corresponding level at the decoder end via long skip connections. This ensures presence of fine grained features extracted in early layers to be present in the noise-free image, helping to maintain boundary properties of objects present within image. The result from decoder is passed to a convolution layer with filter size 1×1.

Spatially Aware Channel Attention (SACA) : While such an encoder-decoder architecture tends to work on homogeneous haze distribution. In case of non homogeneous haze, affected regions might extend beyond receptive field of convolutional kernels resulting in weak representation being extracted along different scales of encoder. Thus dynamic adjustment of receptive field based on haze distribution for encompassing relevant features is required. For this we propose a spatially aware channel attention mechanism that comprises of non-local operation (Wang et al. 2018) followed by a channel attention mechanism. The non local operation allows in capturing long-range dependencies across spatial dimension, while channel attention mechanism filters important channels within feature map. Adding the channel attention mechanism helps in reducing the computational cost, thus allowing deployment of such blocks at different scales. The channel attention mechanism (Fig. 3) is constructed using 1×1 convolution, global average pooling and a softmax activation layer and works by amplifying relevant channels while suppressing irrelevant ones. In order to maximize the effect of spatially aware channel attention layer, we place them in long skip connections and justify such a design choice for ensuring long skip connections carry relevant local features by refining complete feature map corresponding to a particular scale without modifying features within conv-blocks in encoders. The resultant refined features from each SACA block is included into final embedding representation by performing a max pool operation (of varying size) to match feature map size and highlight that such a mechanism allows in enriching the feature space by concatenating additional features.

Frequency Prior based Discriminators for Adversarial Training

Domain gap between synthetic and real haze samples adversely affect performance of underlying dehazing algorithms (Tab. 2). Thus to obtain a domain invariant dehazing algorithm, we take advantage of frequency domain and
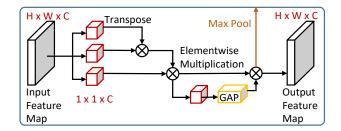

Figure 3: Structural overview of proposed Spatially Aware Channel Attention mechanism

propose a frequency prior based discriminator that relies on both high and low frequency components of an image to determine if a recovered image matches ground truth. The discriminator architecture comprises of 6 conv-blocks resulting in multi-dimensional output, similar to Patch-GAN (Isola et al. 2017), with a patch-size of 64. We utilize two independent discriminators with same architecture but using different frequency priors to obtain different set of weights. We base this design choice on two observations,

1. HF components cover edge information, while LF components cover structure and color information. In this context the intensity of LF components would be larger than HF components, which might lead to LF components gaining more importance in adversarial process.

2. (Wang et al. 2020) highlights during early optimization process LF components are learned owing to a steeper descent of loss surface.

These observations incline us to introduce two discriminators to avoid over reliance of one component over other, while optimizing the complete framework. Monitoring the optimization process ascertains that both LF and HF components are learned. To train the discriminators, for a given image, we first extract its high and low frequency components using laplacian and gaussian filters respectively (of filter size 3 and 7 respectively) and concatenate them with original image. To ensure standard pixel scale, we normalize the HF components before concatenation. Thus for a given pair of hazy I_N and its corresponding dehazed image I_R, the dehazing network estimates a dehazed image $G(I_N)$. Corresponding LF $(LF(.))$ and HF $(HF(.))$ components are extracted and concatenated, resulting in $([I_R, LF(I_R)], [I_R, HF(I_R)]$ and $([G(I_N), LF(G(I_N))], [G(I_N), HF(G(I_N))])$ prior based samples. These samples are used as inputs for corresponding low and high frequency discriminators, that would classify them as real or fake. Thus while training the discriminators would follow the min-max optimization cycle,

$$
\min_{G} \max_{D_{LF}, D_{RF}} \mathbb{E}_{I_R \sim \mathbb{R}_{real}} \{ log(D_{LF} [I_R, LF(I_R)]) \\
+ log(D_{LF} [I_R, HF(I_R)]) \}
$$

$$
+ \mathbb{E}_{G(I_N) \sim \mathbb{R}_{fake}} \{ log(1 - D_{LF}[G(I_N), LF(G(I_N))]) \\
+ log(1 - D_{HF}[G(I_N), HF(G(I_N))]) \}
\tag{2}
$$

Framework Optimization

To train the proposed framework, we follow standard GAN approach wherein the dehazing algorithm and discriminator are optimized alternatively. The optimization function for the dehazing algorithm is composed of L1, SSIM (Wang et al. 2004) and perceptual (Johnson, Alahi, and Fei-Fei 2016) losses along with dual adversarial loss.

| 7.09 / 0.33 | 16.76 / 0.55 | 18.18 / 0.62 | 14.07 / 0.43 | 16.39 / 0.60 | 19.07 / 0.58 |
| Input Image | DuRN-US | FFA-Net | Wavelet-UNet | SNDN | MSNet |

| 17.96 / 0.61 | 17.36 / 0.58 | 16.80 / 0.49 | 8.30 / 0.36 | 22.18 / 0.71 | |
| GDN | DA-Dehazing | PFFNet | Y-Net | Ours | Clean Image |

Figure 4: Performance of strong baseline version of different Dehazing algorithms when evaluated on NTIRE-19 dataset.

$$
\begin{aligned}
L_G(I_R, G(I_N)) = {} & L_1(I_R, G(I_N)) + SSIM(I_R, G(I_N)) \\
& + L_{VGG}(I_R, G(I_N)) \\
& + \lambda_1 * log(1 - D_{LF}[G(I_N), \ LF(G(I_N))]) \\
& + \lambda_2 * log(1 - D_{HF}[G(I_N), \ HF(G(I_N))])
\end{aligned}
\tag{3}
$$

where λ_1, λ_2 are loss balancing terms. In our experiments we set $\lambda_1 = \lambda_2 = 0.5$ to balance both LF and HF discriminators. We design the proposed framework in Pytorch 1.6. The input patch is set to square patches of size 512 normalized to $[0, 1]$. ADAM (Kingma and Ba 2014) is used as optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.9$ and learning rate of 0.0001 for dehazing and 0.0003 for discriminator networks respectively with a batch size of 4. Apart from the aforementioned greedy localized data augmentation (max patch size of 50×50), we also use random horizontal and vertical flipping as additional augmentation techniques. For our experiments we utilize a system equipped with Intel 8700-K CPU and 64GB RAM with Nvidia Titan V GPU.

Experimental Evaluations

Datasets and Evaluation Metrics : In order to evaluate performance of various algorithms across both synthetic and real datasets, exhibiting different haze distributions. We utilize real i.e. NTIRE-18 (Ancuti, Ancuti, and Timofte 2018), NTIRE-19 (Cai et al. 2019), NTIRE-20 (Yuan et al. 2020) and synthetic i.e. SOTS (Li et al. 2019) and Haze-RD (Zhang, Ding, and Sharma 2017) datasets and summarize their properties such as resolution, haze type and average PSNR and SSIM of hazy images in Tab. 2. For evaluating the NTIRE-20 dataset, we first create a subset from training sample and utilize the remaining dataset for training. Furthermore to compare the performance of different SoTA algorithms we utilize Peak-Signal-to-Noise Ratio (PSNR) and Structural Similarity Metric (SSIM) as evaluation metrics. We utilize open-source CNN based dehazing methods

such as DuRN-US (Liu et al. 2019b), FFA-Net (Qin et al. 2020), Wavelet-UNet (Yang and Fu 2019), SNDN (Chen et al. 2020), MSNet (Joyies 2020), GridDehazenet (GDN) (Liu et al. 2019a), PFFNet (Mei et al. 2018b), AtJ-DH+ (Guo et al. 2019), KTDN (Ancuti et al. 2020), TridentNet (Liu et al. 2020), Enhanced-Pix2Pix (Qu et al. 2019) and DA-Dehaze (Shao et al. 2020).

Individual vs Aggregated Dataset for Strong Baseline : Synthetic datasets provide access to extremely diverse characteristics such as scene setting, differing camera properties and illumination conditions, covered in large amount of paired datasets, making them indispensable despite their flaws in modeling haze. We begin by determining performance of algorithms when trained and evaluated on datasets having same distribution while summarizing the results in Tab. 1. We observe dehazing algorithms to perform well on test sets following similar distribution to training dataset, but their performance drops drastically when tested on datasets outside the training distribution, even for algorithms trained on synthetic samples. However, we observe that compared to previous methods, the performance drop of proposed approach is not substantial and we attribute this to utilizing frequency priors while training.

A common approach to achieve domain invariant performance is to increase the dataset size by accumulating data from different sources. Following this we aggregate the aforementioned datasets and evaluate them individual sub test sets with results summarized in Tab. 3. We conclude this approach to aid in achieving peak performance for all algorithms on real datasets such as NTIRE-19 and NTIRE-20. We further corroborate that all algorithms including ours benefit from increased dataset size on account of merging synthetic and real datasets. In this scenario the proposed model outperforms the top algorithm DuRN-US by 3.78 db, 0.19 and 6.84db and 0.21 on NTIRE-20 and NTIRE-19 datasets, thus showcasing improved preservation of structural properties while improving PSNR of the recovered image.

Table 1:

	Trained on Synthetic (RESIDE-Indoor) dataset				Trained on Synthetic (RESIDE-Outdoor) dataset			
	SOTS-IN	SOTS-Out	NTIRE-19	NTIRE-20	SOTS-IN	SOTS-Out	NTIRE-19	NTIRE-20
DuRN-US	32.12 / **0.98**	19.55 / 0.83	10.81 / <u>0.51</u>	<u>11.27</u> / <u>0.51</u>	15.95 / 0.76	19.41 / 0.81	11.04 / 0.51	11.73 / 0.46
FFA-Net	<u>36.36</u> / **0.98**	20.05 / <u>0.84</u>	10.97 / 0.42	10.70 / 0.44	18.96 / 0.86	30.88 / 0.93	09.64 / 0.50	10.90 / 0.48
Wavelet-UNet	20.02 / 0.75	17.75 / 0.67	<u>11.48</u> / 0.47	10.88 / 0.36	16.26 / 0.73	21.95 / 0.76	10.36 / 0.49	11.05 / 0.43
MSNet	32.04 / **0.98**	<u>20.70</u> / **0.86**	09.90 / <u>0.51</u>	11.16 / <u>0.51</u>	21.75 / 0.88	29.80 / 0.93	09.56 / 0.48	11.35 / 0.51
SNDN	24.68 / 0.91	16.02 / 0.69	10.13 / 0.45	10.64 / 0.43	25.30 / 0.91	24.31 / 0.88	11.74 / 0.49	11.95 / 0.52
GDN	32.14 / **0.98**	16.22 / 0.76	09.50 / 0.49	09.01 / 0.40	20.99 / 0.89	29.18 / 0.93	10.16 / 0.50	11.23 / 0.49
PFFNet	26.58 / <u>0.92</u>	14.63 / 0.65	11.38 / <u>0.51</u>	11.14 / 0.43	20.32 / 0.85	27.65 / 0.91	10.75 / 0.50	11.55 / 0.52
Enhc. Pix2Pix	25.06 / 0.92	-	-	-	-	22.57 / 0.86	-	-
Ours	**38.91** / **0.98**	**25.75** / <u>0.84</u>	**16.21** / **0.78**	**16.28** / **0.67**	26.90 / 0.76	30.40 / 0.94	13.36 / 0.52	12.68 / 0.52
	Trained on Real (NTIRE-19) dataset				Trained on Real (NTIRE-20) dataset			
	SOTS-IN	SOTS-Out	NTIRE-19	NTIRE-20	SOTS-IN	SOTS-Out	NTIRE-19	NTIRE-20
DuRN-US	11.44 / <u>0.59</u>	13.05 / 0.61	13.63 / <u>0.57</u>	12.97 / 0.52	09.43 / 0.63	11.92 / 0.66	11.63 / <u>0.52</u>	15.27 / 0.50
FFA-Net	12.16 / 0.55	14.36 / 0.59	14.01 / 0.56	14.71 / <u>0.57</u>	09.96 / 0.63	<u>14.88</u> / <u>0.75</u>	12.43 / <u>0.52</u>	18.11 / 0.66
Wavelet-UNet	13.57 / 0.41	13.05 / 0.44	12.85 / 0.39	12.08 / 0.24	<u>12.04</u> / 0.32	13.85 / 0.41	11.46 / 0.28	12.08 / 0.21
MSNet	13.33 / 0.55	13.85 / 0.56	13.32 / 0.53	12.63 / 0.32	09.16 / 0.51	10.66 / 0.56	12.04 / 0.50	14.06 / 0.50
SNDN	12.56 / **0.66**	14.11 / <u>0.70</u>	13.54 / 0.54	<u>14.93</u> / 0.51	12.03 / <u>0.67</u>	14.14 / 0.73	11.73 / <u>0.52</u>	13.93 / 0.52
GDN	14.57 / <u>0.59</u>	13.47 / 0.60	12.96 / 0.50	12.07 / 0.32	11.60 / 0.58	12.75 / 0.72	<u>13.39</u> / <u>0.52</u>	15.32 / 0.60
PFFNet	13.51 / 0.50	<u>14.57</u> / 0.53	13.29 / 0.52	12.99 / 0.31	08.82 / 0.47	12.00 / 0.53	11.54 / 0.49	14.50 / 0.36
AtJ-DH+	-	-	<u>17.18</u> / 0.53	-	-	-	-	-
KTDN	-	-	-	-	-	-	-	20.85 / 0.69
TridentNet	-	-	-	-	-	-	-	**21.41** / <u>0.71</u>
Ours	**19.28** / **0.66**	**18.17** / **0.87**	**19.47** / **0.75**	**20.33** / **0.77**	**19.53** / **0.71**	**18.69** / **0.79**	**17.24** / **0.66**	<u>21.17</u> / **0.78**

Table 1: Model Performance when trained independently on synthetic and real datasets. Boldface and Underlined values represent best and second best results for each unique independent training dataset. - Indicates missing values not reported in Original Paper

Dataset Name	PSNR / SSIM	Resolution	Type
NTIRE-18	14.60 / 0.67	4177 × 3134	Real
NTIRE-19	9.11 / 0.49	1600 × 1200	Real
NTIRE-20	10.42 / 0.46	1600 × 1200	Real
SOTS-IN	11.97 / 0.69	620 × 460	Synthetic
SOTS-OUT	15.92 / 0.81	550 × 478	Synthetic
HazeRD	14.60 / 0.67	3492 × 2558	Synthetic

Table 2: Properties of Different Datasets

Algorithm	SOTS-IN	NTIRE-19	NTIRE-20
DuRN-US	**26.8** / **0.95**	<u>15.96</u> / <u>0.61</u>	<u>19.88</u> / <u>0.69</u>
FFA-Net	19.15 / 0.85	14.06 / 0.54	15.93 / 0.59
Wavelet	15.46 / 0.65	12.23 / 0.51	12.66 / 0.42
MSNet	24.38 / 0.90	14.65 / 0.59	15.17 / 0.63
SNDN	22.94 / 0.88	14.66 / 0.59	18.65 / 0.67
GDN	22.87 / <u>0.91</u>	13.97 / 0.56	17.02 / 0.68
PFFNet	23.39 / 0.87	13.50 / 0.48	14.77 / 0.57
Ours	<u>25.39</u> / 0.80	**22.80** / **0.82**	**23.66** / **0.88**

Table 3: Algorithm Performance when trained on aggregated dataset. Boldface and Underlined values represent best and second best results.

The performance boost on real dataset comes with a reduced performance on synthetic dataset. However when considering the broader perspective of deployment of these algorithms in real scenarios, such a performance trade-off between real and synthetic datasets is reasonable provided these methods retain their performance, when deployed in another domain. To evaluate this scenario, we refer to algorithms trained on aggregated dataset as being representative of strong baseline for further evaluation.

Performance on datasets outside training distribution : To ascertain that higher performance of algorithms when trained on aggregated dataset ensures performance retention to unknown domains, we use Haze-RD and NTIRE-18 datasets for blind evaluation of strong baseline. We summarize numerical results in Tab. 4 and visual results in Fig. 5. While the proposed framework retains its performance on NTIRE-18 dataset, performance of all algorithms on Haze-RD drops significantly. However performance drop in terms of PSNR is not substantial on NTIRE-18 for both proposed

framework as well as strong baselines. Upon a visual examination of dehazed images, we observe that while performance in PSNR terms is mostly retained, prior works couldn't dehaze the image completely with some regions still affected by haze. Furthermore the structural properties of recovered objects is not retained. On contrary, the proposed framework was not only able to remove haze but also preserve color and structural properties of underlying objects to a substantial degree. Thereby demonstrating the effectiveness of proposed framework in unknown domains.

Ablation Studies : We examine the effects of different strategies proposed in this paper for improving performance using NTIRE-19 and SOTS-IN datasets. The numerical results for different experiments are summarized in Tab. 5. We begin by evaluating a simple UNet architecture that acts as
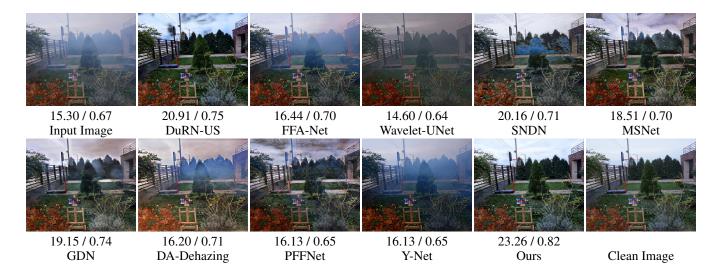
15.30 / 0.67	20.91 / 0.75	16.44 / 0.70	14.60 / 0.64	20.16 / 0.71	18.51 / 0.70
Input Image	DuRN-US	FFA-Net	Wavelet-UNet	SNDN	MSNet
19.15 / 0.74	16.20 / 0.71	16.13 / 0.65	16.13 / 0.65	23.26 / 0.82	
GDN	DA-Dehazing	PFFNet	Y-Net	Ours	Clean Image

Figure 5: Blind Evaluation of different Dehazing algorithms on NTIRE-18 dataset

Algorithm	Haze-RD	NTIRE-18
DuRN-US	15.26 / **0.83**	_18.85_ / 0.71
FFA-Net	15.77 / **0.83**	16.16 / 0.64
Wavelet	_16.30_ / _0.82_	16.01 / 0.70
MSNet	14.42 / 0.80	17.42 / 0.66
SNDN	16.05 / _0.82_	18.08 / 0.70
GDN	14.58 / 0.81	18.26 / _0.74_
PFFNet	15.20 / 0.77	15.66 / 0.58
DA-Dehaze	16.21 / 0.78	16.28 / 0.67
Ours	**21.42** / 0.81	**24.14** / **0.80**

Table 4: Performance on datasets outside training distribution

nator owing to weak supervision.

Algorithm	SOTS-IN	NTIRE-19
Baseline	18.08 / 0.51	11.05 / 0.34
GLDA	21.53 / 0.74	15.32 / 0.53
SACA	22.07 / 0.82	16.70 / 0.59
MSFA	26.75 / 0.85	17.18 / 0.62
Simple Discriminator	28.52 / 0.89	17.72 / 0.69
HF prior	32.87 / 0.94	19.02 / 0.72
LF and HF prior	38.91 / 0.98	19.47 / 0.75

Table 5: Effect of Different Strategies on Network Performance

baseline model, upon which enhancements are performed. We observe using greedy local data augmentation technique (GLDA) significantly boosts performance of baseline model on both known and unknown datasets. We attribute this towards the ability of the network to focus explicitly on haze affected regions. To corroborate this observation, we progressively introduce SATA module with multi-scale feature aggregation (MSFA) and report continuous performance improvement both in terms of PSNR and SSIM, with SACA contributing more to SSIM whereas MSFA contributes towards improved PSNR. This validates the design choice to introduce these enhancements to preserve structural and feature properties respectively. While the PSNR and SSIM were considerably improved on NTIRE-19, the same wasnt observed for SOTS-IN dataset. Thus we examine the effect of using a HF prior based adversarial learning setup, which improves structural preservation across datasets but not the PSNR of recovered images. Subsequently we introduce an additional LF prior based discriminator and observe significant performance retention is achieved. This confirms our hypothesis of adding HF and LF prior based discriminators to preserve structural and color consistency within recovered images which wasnt possible when using a simple discrimi-

Conclusion

In this paper, we focused on dual challenge of domain and haze distributions that significantly reduce performance of dehazing models. We proposed a spatially aware channel attention mechanism integrated within a CNN for increasing the receptive field and utilized local data augmentation to simulate non-uniform haze regions. Subsequently, we trained proposed network within an adversarial framework that uses high and low frequency components as priors to determine whether a given image is real or fake. This is shown to improve performance retention in unknown domains. Extensive experiments demonstrate effectiveness of proposed mechanism on real and synthetic images.

Acknowledgments

References

Agarwal, C.; Chen, P.; and Nguyen, A. 2020. Intriguing generalization and simplicity of adversarially trained neural networks. *arXiv preprint arXiv:2006.09373* .

Ancuti, C.; Ancuti, C. O.; and Timofte, R. 2018. Ntire 2018 challenge on image dehazing: Methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 891–901.

Ancuti, C. O.; Ancuti, C.; Vasluianu, F.-A.; and Timofte, R. 2020. Ntire 2020 challenge on nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 490–491.

Berman, D.; treibitz, T.; and Avidan, S. 2016. Non-Local Image Dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Bousmalis, K.; Silberman, N.; Dohan, D.; Erhan, D.; and Krishnan, D. 2017. Unsupervised pixel-level domain adaptation with generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3722–3731.

Cai, B.; Xu, X.; Jia, K.; Qing, C.; and Tao, D. 2016. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing* 25(11): 5187–5198.

Cai, J.; Gu, S.; Timofte, R.; and Zhang, L. 2019. Ntire 2019 challenge on real image super-resolution: Methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 0–0.

Chen, D.; He, M.; Fan, Q.; Liao, J.; Zhang, L.; Hou, D.; Yuan, L.; and Hua, G. 2019. Gated context aggregation network for image dehazing and deraining. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1375–1383. IEEE.

Chen, Z.; Hu, Z.; Sheng, B.; Li, P.; Kim, J.; and Wu, E. 2020. Simplified non-locally dense network for single-image dehazing. *The Visual Computer* 1–12.

Dong, Y.; Liu, Y.; Zhang, H.; Chen, S.; and Qiao, Y. 2020. FD-GAN: Generative Adversarial Networks with Fusion-Discriminator for Single Image Dehazing. In *AAAI*, 10729–10736.

Dou, Q.; de Castro, D. C.; Kamnitsas, K.; and Glocker, B. 2019. Domain generalization via model-agnostic learning of semantic features. In *Advances in Neural Information Processing Systems*, 6450–6461.

Dundar, A.; Liu, M.-Y.; Wang, T.-C.; Zedlewski, J.; and Kautz, J. 2018. Domain stylization: A strong, simple baseline for synthetic to real image domain adaptation. *arXiv preprint arXiv:1807.09384* .

Engin, D.; Genç, A.; and Kemal Ekenel, H. 2018. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 825–833.

Fattal, R. 2014. Dehazing using color-lines. *ACM transactions on graphics (TOG)* 34(1): 1–14.

Guo, T.; Li, X.; Cherukuri, V.; and Monga, V. 2019. Dense scene information estimation network for dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 0–0.

He, K.; Sun, J.; and Tang, X. 2010. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence* 33(12): 2341–2353.

Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134.

Jiang, Y.; Sun, C.; Zhao, Y.; and Yang, L. 2017. Image dehazing using adaptive bi-channel priors on superpixels. *Computer Vision and Image Understanding* 165: 17–32.

Johnson, J.; Alahi, A.; and Fei-Fei, L. 2016. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, 694–711. Springer.

Joyies. 2020. MSNet: A Novel End-to-End Single Image Dehazing Network with Multiple Inter-Scale Dense Skip-connections. In *IET Image Processing*.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* .

Li, B.; Peng, X.; Wang, Z.; Xu, J.; and Feng, D. 2017. An all-in-one network for dehazing and beyond. *arXiv preprint arXiv:1707.06543* .

Li, B.; Ren, W.; Fu, D.; Tao, D.; Feng, D.; Zeng, W.; and Wang, Z. 2019. Benchmarking Single-Image Dehazing and Beyond. *IEEE Transactions on Image Processing* 28(1): 492–505.

Li, C.; Guo, J.; Porikli, F.; Fu, H.; and Pang, Y. 2018a. A cascaded convolutional neural network for single image dehazing. *IEEE Access* 6: 24877–24887.

Li, H.; Jialin Pan, S.; Wang, S.; and Kot, A. C. 2018b. Domain generalization with adversarial feature learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5400–5409.

Li, R.; Zhang, X.; You, S.; and Li, Y. 2020. Learning to Dehaze From Realistic Scene with A Fast Physics Based Dehazing Network. *arXiv preprint arXiv:2004.08554* .

Liu, J.; Wu, H.; Xie, Y.; Qu, Y.; and Ma, L. 2020. Trident dehazing network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 430–431.

Liu, X.; Ma, Y.; Shi, Z.; and Chen, J. 2019a. GridDe-hazeNet: Attention-Based Multi-Scale Network for Image Dehazing. In *ICCV*.

Liu, X.; Suganuma, M.; Sun, Z.; and Okatani, T. 2019b. Dual Residual Networks Leveraging the Potential of Paired Operations for Image Restoration. In *Proc. Conference on Computer Vision and Pattern Recognition*, 7007–7016.

Long, M.; Cao, Y.; Wang, J.; and Jordan, M. 2015. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, 97–105.

Lu, H.; Li, Y.; Nakashima, S.; and Serikawa, S. 2016. Single image dehazing through improved atmospheric light estimation. *Multimedia Tools and Applications* 75(24): 17081–17096.

Matsuura, T.; and Harada, T. 2020. Domain Generalization Using a Mixture of Multiple Latent Domains. In *AAAI*, 11749–11756.

McCartney, E. J. 1976. Optics of the atmosphere: scattering by molecules and particles. *New York* .

Mei, K.; Jiang, A.; Li, J.; and Wang, M. 2018a. Progressive feature fusion network for realistic image dehazing. In *Asian conference on computer vision*, 203–215. Springer.

Mei, K.; Jiang, A.; Li, J.; and Wang, M. 2018b. Progressive Feature Fusion Network for Realistic Image Dehazing. In *Asian Conference on Computer Vision (ACCV)*.

Narasimhan, S. G.; and Nayar, S. K. 2000. Chromatic framework for vision in bad weather. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, volume 1, 598–605. IEEE.

Narasimhan, S. G.; and Nayar, S. K. 2002. Vision and the atmosphere. *International journal of computer vision* 48(3): 233–254.

Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; and Jia, H. 2020. FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. In *AAAI*, 11908–11915.

Qu, Y.; Chen, Y.; Huang, J.; and Xie, Y. 2019. Enhanced pix2pix dehazing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8160–8168.

Ren, W.; Ma, L.; Zhang, J.; Pan, J.; Cao, X.; Liu, W.; and Yang, M.-H. 2018. Gated fusion network for single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3253–3261.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.

Shao, Y.; Li, L.; Ren, W.; Gao, C.; and Sang, N. 2020. Domain Adaptation for Image Dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2808–2817.

Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A. P.; Bishop, R.; Rueckert, D.; and Wang, Z. 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1874–1883.

Shrivastava, A.; Pfister, T.; Tuzel, O.; Susskind, J.; Wang, W.; and Webb, R. 2017. Learning from simulated and unsupervised images through adversarial training. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2107–2116.

Shyam, P.; Sengar, S. S.; Yoon, K.-J.; and Kim, K.-S. 2021. Evaluating COPY-BLEND Augmentation for Low Level Vision Tasks. *arXiv preprint arXiv:2103.05889* .

Somavarapu, N.; Ma, C.-Y.; and Kira, Z. 2020. Frustratingly Simple Domain Generalization via Image Stylization. *arXiv preprint arXiv:2006.11207* .

Tsai, Y.-H.; Hung, W.-C.; Schulter, S.; Sohn, K.; Yang, M.-H.; and Chandraker, M. 2018. Learning to adapt structured output space for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7472–7481.

Tzeng, E.; Hoffman, J.; Saenko, K.; and Darrell, T. 2017. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7167–7176.

Wang, H.; Wu, X.; Huang, Z.; and Xing, E. P. 2020. High-frequency Component Helps Explain the Generalization of Convolutional Neural Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8684–8694.

Wang, X.; Girshick, R.; Gupta, A.; and He, K. 2018. Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7794–7803.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13(4): 600–612.

Yang, H.; Pan, J.; Yan, Q.; Sun, W.; Ren, J.; and Tai, Y.-W. 2017. Image dehazing using bilinear composition loss function. *arXiv preprint arXiv:1710.00279* .

Yang, H.-H.; and Fu, Y. 2019. Wavelet u-net and the chromatic adaptation transform for single image dehazing. In *2019 IEEE International Conference on Image Processing (ICIP)*, 2736–2740. IEEE.

Yuan, S.; Timofte, R.; Leonardis, A.; and Slabaugh, G. 2020. Ntire 2020 challenge on image demoireing: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 460–461.

Zhang, H.; and Patel, V. M. 2018. Densely connected pyramid dehazing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3194–3203.

Zhang, Y.; Ding, L.; and Sharma, G. 2017. Hazerd: an outdoor scene dataset and benchmark for single image dehazing. In *2017 IEEE international conference on image processing (ICIP)*, 3205–3209. IEEE.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*.

Zhu, Q.; Mai, J.; and Shao, L. 2015. A fast single image haze removal algorithm using color attenuation prior. *IEEE transactions on image processing* 24(11): 3522–3533.