# HardGAN: A Haze-Aware Representation Distillation GAN for Single Image Dehazing

Qili Deng*[2], Ziling Huang*†[1][0000−0003−3241−7911],
Chung-Chi Tsai[3][0000−0003−1792−9978], and Chia-Wen Lin[1][0000−0002−9097−2318]

[1] National Tsing Hua University Hsinchu 30013,
cwlin@ee.nthu.edu.tw, huangziling@gapp.nthu.edu.tw
[2] ByteDance AI Lab, Beijing
dengqili@bytedance.com
[3] Qualcomm Technologies, Inc., San Diego
chuntsai@qti.qualcomm.com

**Abstract.** In this paper, we present a Haze-Aware Representation Distillation Generative Adversarial Network (HardGAN) for single-image dehazing. Unlike previous studies that intend to model the transmission map and global atmospheric light jointly to restore a clear image, we approach this restoration problem by using a multi-scale structure neural network composed of our proposed haze-aware representation distillation layers. Moreover, we re-introduce to utilize the normalization layer skillfully instead of stacking with the convolutional layers directly as before to avoid useful information wash away, as claimed in many image quality enhancement studies. Extensive experiments on several synthetic benchmark datasets as well as the NTIRE 2020 real-world images show our proposed HardGAN performs favorably against the state-of-the-art methods in terms of PSNR, SSIM, LPIPS, and individual subjective evaluation.

**Keywords:** Image Dehazing, Generative Adversarial Network (GAN), Image Restoration, Deep Learning

## 1 Introduction

Haze often occurs when dusk and smoke particles accumulate in the air that absorbs and scatters the sunlight, resulting in noticeable visual quality degradation in object appearance and contrast. Thus, taking those low contrast input for many computer vision systems designed under the assumption of an ideal capture environment will impede its real performance. Hence, image dehazing becomes a prerequisite task for several important visual analysis tasks.

Image dehazing has been explored for many years. Many previous approaches [10, 11, 21, 35] depend on the formation of haze images by the following mathematical formulation [20]:

$$I(z) = J(z)t(z) + A(z)(1 - t(z)), \tag{1}$$

* indicates co-first author.
† indicates corresponding author.

where $I(z)$ is an observed hazy image, $J(z)$ is its haze-free version, $A(z)$ is the global atmospheric light intensity that depends on the unknown depth map, $t(z)$ is the transmission map, and $z$ is the pixel location. Thereby, the solution for haze-free image restoration is by estimating the transmission map and global atmospheric light intensity and then calculate the final result by Eq. (1). Though the approaches [10, 11, 21, 35] mentioned above show its effectiveness, many drawbacks remain. As we know, fitting the estimated transmission and the global atmospheric light intensity maps into Eq. (1) to obtain haze-free images might become problematic, since the formation of haze depends on several factors, e.g. , temperature, humidity, altitude. Therefore, the transmission map can hardly be described by a simple function, nonetheless to say, trying to approximate it in a complex natural environment.

With the success of data-driven approaches, many researchers proposed end-to-end CNN models [5, 19] for single image dehazing. To avoid washing away the essential spatial information, they discarded the normalization layer from the convolution layers [13, 30, 32]. However, lacking the normalization layer implies the networks will be shallower and hard to fit large-scale arbitrariness caused by haze. Furthermore, the gradient could vanish during training without the normalization layer even if skip-connection is implemented.

To addess these two issues jointly, we propose a Haze-Aware Representation Distillation GAN (HardGAN) to learn the mapping function between haze images and haze-free images directly. Different from previous works, we design a Haze-Aware Representation Distillation (HARD) module to incorporate the normalization layer into our work. The spatial information and atmospheric brightness are fused based on the haze-aware map due to different levels of haze concentration. The contribution of this paper is fourfold. First, We proposed a multi-scale network named HardGAN to learn style transfer mapping directly. Second, the instance normalization is introduced to image dehazing task skillfully, and we create Haze-Aware Representation Distillation (HARD) module to fuse global atmospheric brightness and local spatial structures attentively. Third, extensive experiments on existing datasets show more favorable results over state-of-the-art methods. We also provide comprehensive ablation studies to validate the importance and necessity of each Module. Lastly, we further apply our algorithm to nature dense non-homogeneous haze dataset. Our proposed Generative Adversarial Network can still accurately recover the unseen objects in those problematic cases.

## 2 Related Work

**Single image dehazing.** Single image dehazing with unknown transmission map and global atmospheric light is a challenging problem. In the past two decades, several methods are proposed to address this issue, and we categorize them into two types: prior-based method and learning-based method.

Previously, researchers utilized image statistics prior to compensate for information loss. For example, the albedo of the scene could be estimated as prior

knowledge based on [6]. Assuming local contrast of haze images were low, [29] proposed Markov Random Field to maximize color contrast. To calculate the transmission map more reliably, [11] discovered the dark channel to improve the quality of the dehaze image. More improvements for the dark channel were made on [33, 34]. [7, 3] introduced their algorithm separately based on the observation that small image patches typically exhibited a one-dimensional distribution in the RGB color space. In traditional prior-based methods, the assumption could hold only in some cases, not all which was restricted.

Unlike prior-based work, learning-based methods can learn the image prior automatically by large-scale datasets. [4, 24] proposed trainable end-to-end systems for transmission map estimation. However, in dehazing task, both transmission map and global atmospheric light should be considered. [16] leveraged a linear transformation to encode the transmission map and the atmospheric light into one variable. [35] introduced a new edge-preserving densely connected encoder-decoder structure with multi-level pyramid pooling module for estimating the transmission map. [5] adopted a multi-level gated network and smoothed dilation technique to restore high-quality haze-free images.

**Generative Adversarial Network (GAN).** Recently, we had witnessed the power of generative adversarial learning network in image-to-image translation field. [37] defined a class of image editing operations, and constrained their output image to lie on that learned manifold at all time. [31] could generate 2048x1024 visually appealing results with a unique adversarial loss, as well as multi-scale generator and discriminator architectures. [22] made a huge success in semantic image synthesis by proposing a spatially-adaptive normalization for modulating the activations in normalization layers through a spatially-adaptive, learned affine transformation. [27] attracted much haze-aware last year by constructing a pyramid of fully convolutional GANs, each responsible for determining the patch distribution at a different scale of the image.

## 3 Haze-Aware Representation Distillation GAN

Traditional dehazing methods resort to estimating the transmission map and global atmospheric light density in Eq. (1) based on certain prior information. However, the density of haze can be influenced by various factors, such as temperature, altitude, and humidity, making the formation of haze at individual spatial locations space-variant and non-homogeneous. As a result, haze usually cannot be accurately characterized by just a single transmission map. Therefore, to effectively tackle the spatial variance of haze, instead of learning the transmission map and atmospheric light density, our work focus on learning and distilling the global and spatial features for representing the underlying haze-free image using a GAN guided by non-homogeneous haze conditions. Given an input hazy image $X$, our goal is to restore a haze-free image from $X$. To capture the global properties (e.g., atmospheric light) of each object and the local structurese), we propose a generator to capture useful information at different scales. We then propose a Haze-Aware Representation Distillation (HARD) module to distill and

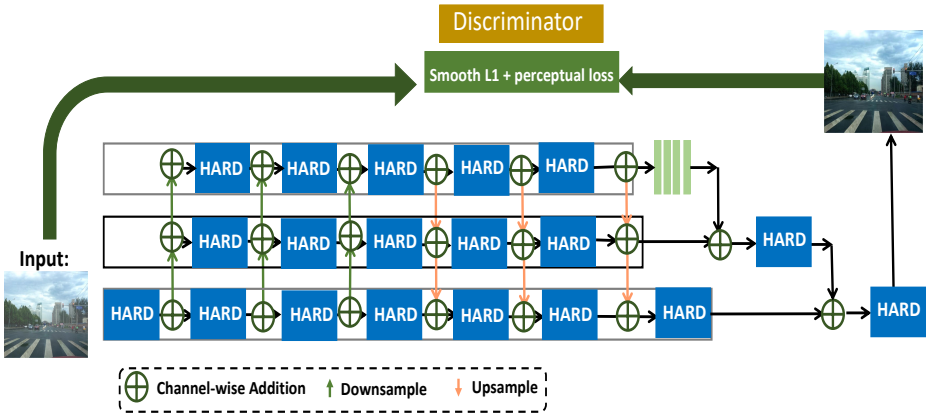


**Fig. 1.** Total framework of our HardGAN. The images are feed into the generator. We downsample the feature map at first to learn spatial information and upsample feature maps so as to learn details from different scales. The whole procedure is similar to climbing a stair. After obtaining haze-free images, we feed them into discriminator with full-scale haze images to discriminate our generated haze-free images real or fake.

combine the spatial features and atmospheric brightness adaptively. To ensure the visual realisticness of dehazed images, multi-scale patch-GAN discriminators [14] are utilized to discriminate real images from fake ones. Our framework is shown in Fig. 1.

### 3.1 Haze-Aware Representation Distillation GAN (HardGAN)

As illustrated in Fig. 1, the generator of HardGAN consists of three layers from coarse to fine: the first (coarsest) layer involving five Haze-Aware Representation Distillation (HARD) modules, the second (medium) with six HARDs, and the third (finest) with eight HARDs. Given an input hazy image $X$ and its target haze-free image $Y$, let $x_m^n$ and $y_m^n$ denote the input and output of the $n$-th HARD in the $m$-th layer (denoted $G_m^n$). The inputs of the second and first layers are $X\downarrow$ and $X\downarrow\downarrow$, respectively, where $\downarrow$ represents downsampling.

The generator at each scale starts from the finest scale and sequentially passes the extracted features up to the coarsest (1/4) scale, as formulated in Eq. (2) and Eq. (3):

$$x_2^n = ADD(y_3^{n-1}\downarrow, y_2^{n-1}) \tag{2}$$

$$x_1^n = ADD((y_3^{n-1}\downarrow)\downarrow, y_2^{n-1}\downarrow, y_1^{n-1}) \tag{3}$$

Subsequently, the multi-scale features are passed backward from the coarsest to the finest scale and finally fused at the finest scale to reconstruct the haze-free image as expressed by Eq. (4) and Eq. (5) for layers 2 and 1 respectively.

$$x_2^n = ADD(y_1^{n-1}\uparrow, y_2^{n-1}) \tag{4}$$

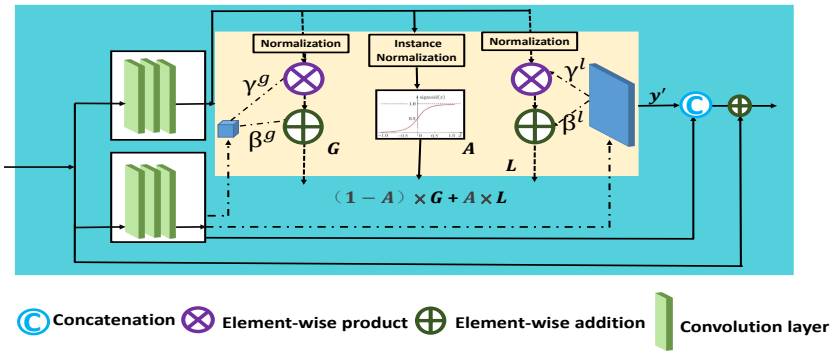


**Fig. 2. Haze-Aware Representation Distillation (HARD) Module.** HARD is composed of two branches. The second branch is used to learn spatial information $\gamma_g$, $\beta_g$ and global atmospheric light information $\gamma_l$, $\beta_l$, then feed them into the first branch to form intermediate results $y'$. After channel attention, the final result of HARD is produced.

$$x_3^n = ADD((y_1^{n-1}\uparrow)\uparrow, y_2^{n-1}\uparrow, y_3^{n-1}) \tag{5}$$

where $ADD(\cdot)$ denote channel-wise addition, and $\uparrow$ represents upsampling.

### 3.2 Haze-Aware Feature Distillation (HARD) Module

Existing dehazing networks usually discard normalization layers and introduce skip connections in convolutional layers to avoid losing local spatial structures in representation learning. However, the normalization layer is helpful as it can avoid gradient vanishing so that the number of layers can be increased, thereby increasing the network capacity of representation learning as well. Therefore, in this work we propose adding normalization layers back again. Furthermore, since haze can be space-variant when the density of haze is high, distilling features that can well capture local structures becomes crucial. We therefore propose using two instance normalization layers to distill global atmospheric light intensity and local spatial structures and then adaptively fuse them together. To this end, we introduce a haze-aware attention map to estimate the density of haze at individual locations and propose a haze-guided adaptive feature distillation and fusion approach.

Previous works [19, 5] feed input haze images into a deep network to learn spatial information by stacking a number of convolutional layers simply. Nevertheless, the more the number of stacked convolutional layers, the higher the possibility of gradient vanishing, thereby significantly limiting the representing power of learned features. To tackle this problem, we introduce the Haze-Aware Representation Distillation (HARD) modules each having the same structure. Let $x_i$ denote the $i$-th feature map of the input, where $C_i$, $H_i$ and $W_i$ stand for the image channel, height, and width, respectively. We aim to fuse the spatial information and atmospheric brightness together based on a learned haze-aware

map. To this end, we combine SPADE [22] that can well preserve spatial information and adaIN [12] that can restore targeted atmospheric brightness together to produce haze-free images.

Each HARD module contains two branches, as shown in Fig. 2. Instead of computing atmospheric brightness such as the mean and standard deviation from training samples directly like [30], we learn it automatically. The second branch is used to combine the spatial information and atmospheric brightness together. It contains three sub-branches for haze-aware map generation, global atmospheric brightness estimation, and spatial information insertion.

Because haze in the real world is always in an irregular pattern and it obscures objects resulting in low contrast images, restoring image contrast selectively is a key task in image dehazing. To this end, we encode the atmospheric brightness as a linear model of the input in a $1 \times 1 \times 2$ matrix for each channel, represented as $\gamma_i^g$ and $\beta_i^g$. The atmospheric brightness control function is defined as follows:

$$G_i = \gamma_i^g \frac{x - \mu}{\sigma} + \beta_i^g \tag{6}$$

where $\mu$ and $\sigma$ are the mean and standard deviation of input $x$.

Similarly, we use an $H \times W \times 2$ matrix to encode the pixel-wise spatial information for each channel, represented as $\gamma_i^l$ and $\beta_i^l$. The spatial information preserving function is defined in Eq. (7).

$$L_i = \gamma_i^l \frac{x - \mu}{\sigma} + \beta_i^l \tag{7}$$

To fuse atmospheric brightness and spatial information adaptively, the output feature maps are fed into an Instance Normalization followed by a Sigmoid layer to produce the haze-aware map $A$ for each channel, where $A_i$ represents haze-aware map for the $i$-th channel. This approach ensures our model changes their focus when encountering irregular type haze.

After obtaining these three features, we consider to fuse them together to produce the output by

$$y_i = (1 - A_i) \otimes G_i + A_i \otimes L_i \tag{8}$$

where $\otimes$ denotes element-wise product.

### 3.3 Network Training

We train our proposed architecture step by step in a coarse-to-fine manner. The training loss for HardGAN is comprised of an adversarial loss term, a smooth L1 loss term, and a perceptual loss term [15], as formulated below:

$$\mathcal{L} = \lambda_{adv}\mathcal{L}_{adv} + \lambda_{L_1}\mathcal{L}_1 + \lambda_{per}\mathcal{L}_{per} \tag{9}$$

**Adversarial Loss.** We use the WGAN-GP loss [9] to increase the training stability coupled with a patch Discriminator to classify each of the overlapping patches of its input as real or fake, so we define adversarial loss in Eq. (10):

$$\mathcal{L}_{adv}(G, D) = E[D(y)] - E[D(G(x))] + \lambda E[(|\nabla D(\alpha x - (1 - \alpha G(x)))| - 1)^2] \tag{10}$$

**Smooth L1 Loss.** We employ the smooth L1 loss to measure the difference between a dehazed image and its ground-truth image quantitatively. Compared with L2 loss, the smooth L1 loss, as expressed in Eq. (11), can prevent potential gradient explosion [8].

$$\mathcal{L}_1 = \frac{1}{N}\sum_{y=1}^{N}\sum_{i=1}^{3}\alpha(\hat{Y}_i(z) - Y_i(z)) \tag{11}$$

where $\hat{Y}_i(z)$ and $Y_i(z)$ denote the intensity of the $i$-th channel of pixel $z$ in the reconstructed haze-free image and in the ground truth, respectively, $N$ denote the total number of pixels. and $\alpha$ is specified in Eq. (12).

$$\alpha(e) = \begin{cases} 0.5e^2, & \text{if } |e| < 1 \\ |e| - 0.5, & \text{otherwise} \end{cases} \tag{12}$$

**Perceptual Loss** Instead of encouraging an output dehazed image $y$ to be exactly the same as its ground-truth $y_t$ in the pixel domain, the perceptual loss aims to encourage it to have similar a feature representation in the backbone network (e.g., VGG19 pre-trained on imagenet [25] in this work). The perceptual loss is defined as follows:

$$\mathcal{L}_{per} = \sum_{j=1}^{3}\frac{1}{C_j H_j W_j}||\phi_j(y) - \phi_j(y_t)|| \tag{13}$$

where $H_j$, $W_j$, and $C_j$ denote the height, width, and image channel of the feature map in the $j$-th layer of the backbone network, $\phi_j$ is the activation of the $j$-th layer.

## 4 Experiments

We first conduct our experiments on two public synthetic datasets to validate the effectiveness of the proposed HardGAN. Furthermore, we apply our algorithm to a dense non-homogeneous haze image dataset to demonstrate its generality. We also conduct an ablation study to justify the use of the core modules of HardGAN. The source code cab be found in our Github site.

### 4.1 Datasets

It is time-consuming to collect real-world hazy images and their haze-free counterparts at various locations, which poses a challenge to collect a large-scale useful dataset for data-driven dehazing methods. To address this problem, a few synthetic datasets have been proposed, in which haze images are generated from

https://github.com/huangzilingcv/HardGAN

haze-free images based on the atmosphere scattering model in Eq. (1) via a random and proper choice of the scattering coefficient and the atmospheric light intensity. In this work, we utilize the synthetic RESIDE dataset proposed in [17] to train and test HardGAN. RESIDE contains synthetic hazy images in both indoor and outdoor scenarios. In its Indoor Training Set (ITS), 13,990 hazy indoor images are generated from 1,399 haze-free images with $\beta \in [0.6, 1.8]$ based on Eq. (1) with $t(z) = e^{-\beta d(z)}$ and $A \in [0.7, 1.0]$, where the depth maps $d(z)$ of images are obtained from the NYU Depth V2 [28] and Middlebury Stereo [26] datasets. The Synthetic Objective Testing Set (SOTS) with 500 indoor and 500 outdoor hazy images are produced in the same way. For Outdoor Training Set (OTS), 296,695 hazy outdoor images are generated from 8,477 haze-free images with $\beta \in [0.04, 0.2]$ and $A \in [0.8, 1.0]$, for which the depth maps of outdoor images are obtained based on [18]. Moreover, for evaluations on real-world images, we use the SOTS real-world dataset containing Internet-collected indoor and outdoor hazy images without haze-free ground-truths.

### 4.2 Experiment Settings

To train HardGAN, we follow the training manner in [27] from coarse to fine: training the 1/4-scale generator at first, then the 1/2-scale generator, and finally the full-scale generator. The full-scale RGB input images are with a resolution of $240 \times 240$. For the indoor dataset, each scale is trained for 120 epochs using the Adam optimizer with an initial learning rate of 0.001, which is then halved every 3 epochs. For the outdoor dataset, since the synthetic haze is lighter, the number of epochs for each scale is reduced to 18, while the setting for Adam optimizer is the same as above. Our experiments are carried out on two NVIDIA GeForce GTX 1080Ti with a batch size of 24 separately. In the following experiments, we set $\lambda_1 = 1.2$, $\lambda_{per} = 0.04$, and $\lambda_{adv} = 0.05$, respectively.

### 4.3 Synthetic Hazy Images

We compare the performance of HardGAN with several state-of-the-art data-driven methods including AODNet [16], DehazeNet [4], GCANet [5], GridDehazeNet [19], and FFANet [23] on synthetic hazy images qualitatively and quantitatively. For a fair comparison, all methods are trained in the same way with HardGAN and then evaluated on RESIDE and SOTS. For the quantitative comparison, we use three objective quality metrics: Peak Signal to Noise Ratio (PSNR), Structural SIMilarity index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) [36]. Given a dehazed image and its ground-truth, PSNR and SSIM measure their average pixel-wise and structural fidelity/similarity (i.e., the higher, the better), whereas LPIPS measures their perceptual discrepancy (the lower, the better).

Fig. 3 shows the qualitative comparisons on some synthetic indoor and outdoor images of SOTS. Compared with the ground-truths, the dehazed outputs

The authors from Taiwan universities and ByteDance completed the experiments.

of AODNet, DehazeNet, and GCANet still contain a significant amount of haze. Moreover, we can observe severe color distortions in all of their outputs. In contrast, although GridDehazeNet and FFANet effectively mitigate the color distortion problem effectively, they cannot completely clean the haze in their outputs (see the roads and buildings in Fig. 3(e) and Fig. 3(f)). Besides, both GridDehazeNet and FFANet introduce unexpected noisy artifacts (see the wall and ceiling in Fig. 3(e) and Fig. 3(f)). Compared with these state-of-the-art data-driven methods, **HardGAN** produces the highest-fidelity dehazed results that also look perceptually close to the reference ground-truths.

Furthermore, Table 1 compares the quantitative dehazing results on the SOTA test dataset, showing that **HardGAN** outperforms all the previous dehazing methods in terms of PSNR, SSIM, and LPIPS.

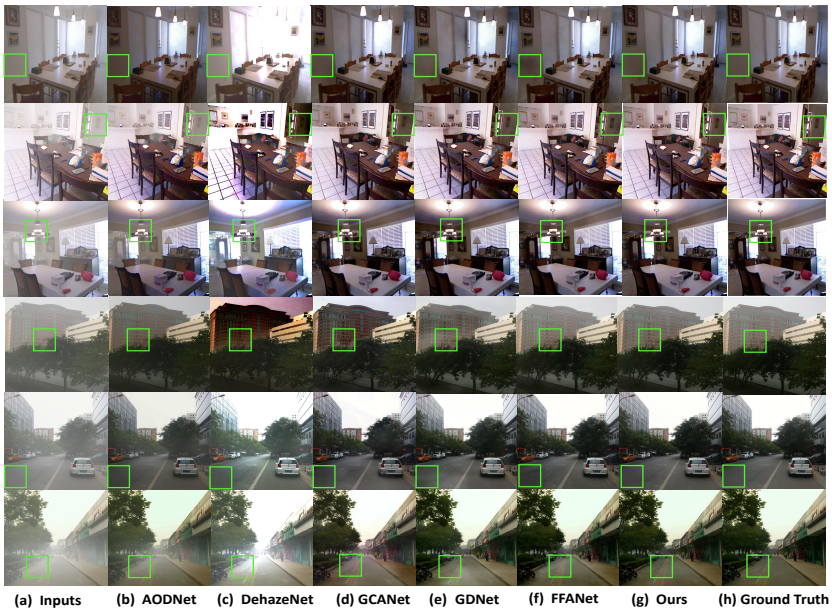


**Fig. 3.** Qualitative comparison of various dehazing methods on some indoor (the first three rows) and outdoor (the last three rows) synthetic hazy images of SOTS. Compared with the ground-truths in (h), we can observe a significant amount of haze and severe color distortions in the dehazed outputs of AODNet, DehazeNet and GCANet. In contrast, GridDehazeNet and FFANet effectively mitigate color distortions but still cannot fully clean the haze in their outputs (see the roads and buildings in (e) and (f)). Besides, both GridDehazeNet and FFANet also introduce unexpected noisy artifacts (see the wall and ceiling in (e) and (f)).

**Table 1.** Quantitative comparison of various dehazing methods on SOTS. HardGAN outperforms all previous dehazing methods in all metrics, where ↑ means the higher the better, and ↓ means the lower the better

| Method | Indoor | | | Outdoor | | |
|---|---|---|---|---|---|---|
| | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ |
| AODNet [16] | 20.5 | 0.8162 | 0.247 | 24.14 | 0.9198 | 0.085 |
| DehazeNet [4] | 19.82 | 0.8209 | 0.334 | 24.75 | 0.9296 | 0.127 |
| GCANet [5] | 30.23 | 0.9800 | 0.217 | - | - | 0.048 |
| GridDehazeNet [19] | 32.16 | 0.9836 | 0.209 | 30.86 | 0.9819 | 0.012 |
| FFANet [23] | 36.39 | 0.9556 | 0.209 | 33.57 | 0.9840 | 0.021 |
| HardGAN (Ours) | **36.56** | **0.9905** | **0.201** | **34.34** | **0.9871** | **0.010** |

### 4.4 Real-world Hazy Images

In this section, we test our methods on real-world datasets. For a fair comparison, all the compared models are trained on the SOTS outdoor training dataset. Because there is no ground-truth for real-world dataset, we conduct a user study to evaluate the subjective perceptual quality quantitatively.

Fig. 6 shows the qualitative comparisons on real-world images. Similar to Fig. 3, the outputs of AODNet, DehazeNet and GCANet again lead to severe color distortions (see the electric line, buildings and heaven in Fig. 6(c) and Fig. 6(e)). Although GridDehazeNet and FFANet effectively solve the color-distortion problem, they cannot fully clean the haze (see the trees and buildings in Fig. 6(e) and Fig. 6(f)). Besides, both GridDehazeNet and FFANet would introduce unexpected noisy(see building and heaven in Fig. 6(e) and Fig. 6(f)). Compared with previous state-of-the-art methods, we produce high quality results with the best perceptual quality.

**Table 2.** Quantitative comparison of various dehazing methods on NH-HAZE, where ↑ means the higher the better, and ↓ means the lower the better.

| Team Name | NH-HAZE | |
|---|---|---|
| | PSNR↑ | SSIM↑ |
| ECNU-Trident | 21.41 | $\mathbf{0.71}_{(1)}$ |
| ECNU-KT | 20.85 | 0.69 |
| NTU-Dehazing | 20.11 | 0.66 |
| VICLAB-DoNET | 19.70 | 0.68 |
| iPAL-NonLocal | 21.10 | 0.69 |
| VIP_UNIST | 18.77 | 0.54 |
| Team JJ | 19.49 | 0.66 |
| iPAL-END | 19.22 | 0.66 |
| Ours | **21.70** | $0.70_{(2)}$ |

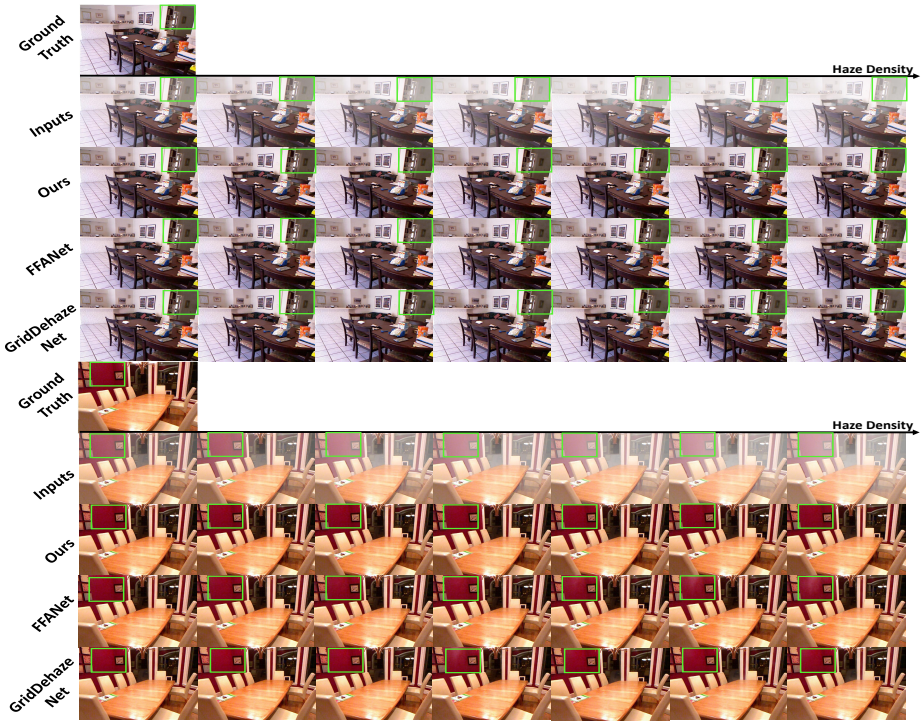


**Fig. 4.** Performance evaluation of HardGAN, FFANet, and GridDehazeNet on synthetic hazy images with different haze patterns from SOTS. In the first set of data in rows 1–5, HardGAN produces high-fidelity outcomes close to the ground-truth, regardless of the haze patterns. In contrast, the results produced by FFANet and GridDehazeNet are unstable (e.g., the wall in row 1–5), especially for heavy haze. Similar results can also be observed in the second set of data in rows 6–10 (see the red wall).


**Fig. 5.** Dehazed results of five images with dense non-homogeneous haze from the validation dataset of NTIRE2020 [1, 2]. The results show that HardGAN effectively removes most of the haze while uncovering clear scenes.

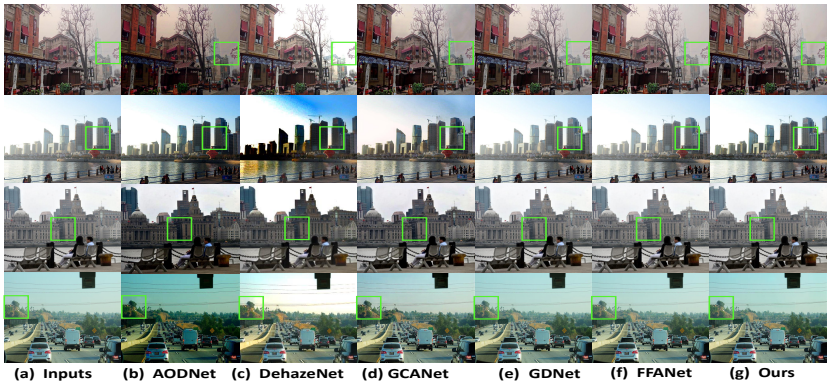


**Fig. 6.** Qualitative comparison of various dehazing methods on the SOTS Real-World dataset. of AODNet, DehazeNet and GCANet lead to unnatural colors (see the electric line, buildings and heaven in (c) and (e)). Although GridDehazeNet and FFANet avoid such color distortions, there is still light haze that remains unremoved in their outputs (see the trees and buildings in (e) and (f)). Besides, both GridDehazeNet and FFANet introduce undesired noise (see the building and heaven in (e) and (f)

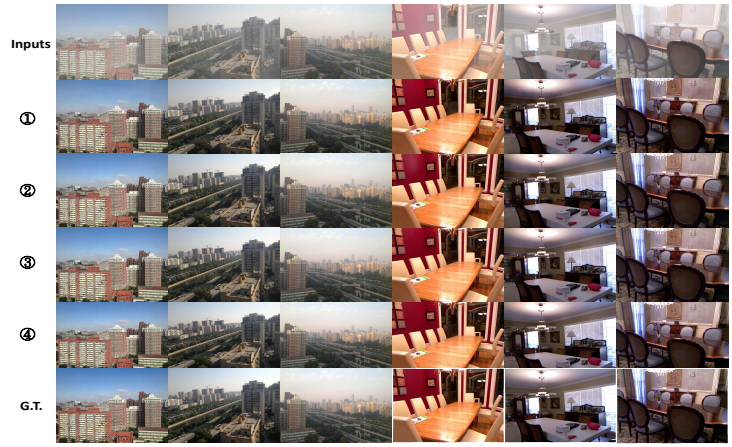


**Fig. 7.** Qualitative comparison for ablation study. Compared with the baseline (Variant ①), Variant ② better preserves local details, thanks to its local spatial structures preservation. Besides, the atmospheric brightness with Variant ③ is more consistent with the corresponding ground-truth than the baseline. Considering both local and global terms together, Variant ④ preserves more details and leads to more consistent atmospheric brightness with the ground-truth.

**Table 3.** Ablation study for the core components of HardGAN. Both Spatial Information Preserving and Atmospheric Brightness Control contribute to final result. Integrating them together can produce haze-free images with highest score.

| | **Settings** | | **Outdoor** | | | **Indoor** | | |
|---|---|---|---|---|---|---|---|---|
| Variant | $L$ | $G$ | PSNR | SSIM | LPIPS | PSNR | SSIM | LPIPS |
| ① | × | × | 30.78 | 0.9821 | 0.025 | 32.25 | 0.9838 | 0.220 |
| ② | ✓ | × | 33.54 | 0.9857 | 0.018 | 35.20 | 0.9878 | 0.222 |
| ③ | × | ✓ | 33.77 | 0.9858 | 0.014 | 35.48 | 0.9892 | 0.210 |
| ④ | ✓ | ✓ | **34.34** | **0.9871** | **0.010** | **36.56** | **0.9905** | **0.201** |

### 4.5 Results on Dense Non-homogeneous Hazy Images

Fig. 5 illustrates the results for five real-world test images with dense non-homogeneous haze from the validation dataset of NTIRE2020 Challenge [2, 1]. The results show that HardGAN successfully removes most haze while uncovering clear scenes. Specifically, since the training samples in [1] contain real-world scenes with similar objects (e.g., trees, grass, sculptures) to that in the test samples, the dehazed images in Fig. 5 present more natural scenes than that of real-world hazy images for which HardGAN is trained on SOTS synthetic outdoor dataset and there exist large differences between the training and testing samples. Following the protocol in [2, 1], Table. 2 shows HardGAN outperforms the others by a large margin in PSNR and achieves the second best SSIM.

### 4.6 Ablation Study

Table 3 shows our ablation study on the SOTS Indoor and Outdoor datasets, where $L$ stands for preserving local spatial structure and $G$ stands for controlling the global atmospheric brightness, respectively. Comparing Variant ② with Variant ①, we can find that adding the preservation of local spatial information is more effective than the baseline, since the spatial information is vital for single image dehazing. Similarly, global atmospheric light control (i.e., Variant ③) also effectively improve the performance. taking into account both the local spatial information and global atmospheric brightness (i.e., Variant ④) achieves the best performance. Fig. 7 shows the qualitative comparison for the ablation study. Variant ② better preserves local details than Variant ①, thanks to the local spatial information preservation. Besides, the atmospheric brightness with Variant ③ is more consistent with the corresponding ground-truth. Again, considering both local and global terms together ④ preserves more details and leads to more consistent atmospheric brightness with the ground-truth.

### 4.7 Network Stability

To further demonstrate the effectiveness of our method, we also conduct a network stability experiment. HardGAN dehazes the hazy versions of an image synthesized with different haze patterns, as shown in Fig. 4 **Inputs**. In the first set of data in rows 1–5, HardGAN produces high-fidelity outcomes close

to ground truth, regardless of the haze patterns. In contrast, In contrast, the results produced by FFANet and GridDehazeNet, however, are unstable (e.g., the wall in row 1–5 of Fig. 4), especially for heavy haze. Similar results can also be observed in the second set of data in rows 6–10 (see the red wall in Fig. 4), so our method is more robust than those methods.

### 4.8 Network Convergence Analysis

Fig. 8 shows loss curves to verify the necessity of HARD module. We train generator only (without discriminator). The experiment is performed on the dataset of NTIRE2020 [1, 2] Fig. 8(a) shows that the training loss curve decreases steadily and Fig. 8(b) shows the PSNR value with adaptive and instance normalization. increases with time. Note, the training loss value shown in Fig. 8(c) drops initially then stays at 0.15 afterwards, whereas, as shown in Fig. 8(d), the PSNR value without adaptive instance normalization and SPADE increases to 8.0 but stays stable after that. These phenomenons illustrates Instance normalization can help model converge while ensures good dehazing results.

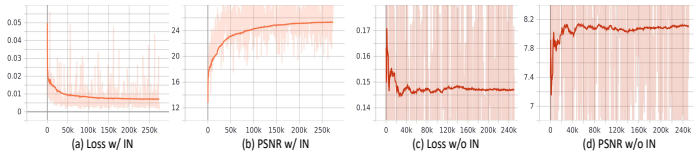


**Fig. 8.** Loss curves to verify the necessity of normalization layers (without discriminator). Here, (a) is the training loss curve with adaptive instance normalization and SPADE while (c) is the training loss curve without adaptive instance normalization and SPADE. It is clear instance normalization can help model converge.

## 5 Conclusion

We proposed a novel multi-scale image dehazing network. Instead of explicitly estimating the transmission map and atmospheric light intensity, our method adaptively fuses local spatial information and global atmospheric brightness together guided by the learned haze-aware maps for individual channels. Extensive experiments on synthetic and real-world hazy images demonstrate the effectiveness of our method. Besides images with homogeneous haze, our method can also do a good job for removing dense non-homogeneous haze in an image.

## Acknowledgments

This work was funded in part by Qualcomm through a Taiwan University Research Collaboration Project and in part by the Ministry of Science and Technology, Taiwan, under grant MOST 109-2634-F-007-013.

## References


1. Ancuti, C.O., Ancuti, C., Timofte, R.: Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 444–445 (2020)
2. Ancuti, C.O., Ancuti, C., Vasluianu, F.A., Timofte, R.: Ntire 2020 challenge on nonhomogeneous dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 490–491 (2020)
3. Berman, D., Avidan, S., et al.: Non-local image dehazing. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1674–1682 (2016)
4. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. IEEE Transactions on Image Processing **25**(11), 5187–5198 (2016)
5. Chen, D., He, M., Fan, Q., Liao, J., Zhang, L., Hou, D., Yuan, L., Hua, G.: Gated context aggregation network for image dehazing and deraining. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1375–1383. IEEE (2019)
6. Fattal, R.: Single image dehazing. ACM transactions on graphics (TOG) **27**(3), 1–9 (2008)
7. Fattal, R.: Dehazing using color-lines. ACM transactions on graphics (TOG) **34**(1), 1–14 (2014)
8. Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 1440–1448 (2015)
9. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of wasserstein gans. In: Advances in neural information processing systems. pp. 5767–5777 (2017)
10. Hautière, N., Tarel, J.P., Aubert, D.: Towards fog-free in-vehicle vision systems through contrast restoration. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8. IEEE (2007)
11. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE transactions on pattern analysis and machine intelligence **33**(12), 2341–2353 (2010)
12. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1501–1510 (2017)
13. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
14. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)
15. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision. pp. 694–711. Springer (2016)
16. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: An all-in-one network for dehazing and beyond. arXiv preprint arXiv:1707.06543 (2017)
17. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing **28**(1), 492–505 (2018)

18. Liu, F., Shen, C., Lin, G., Reid, I.: Learning depth from single monocular images using deep convolutional neural fields. IEEE transactions on pattern analysis and machine intelligence **38**(10), 2024–2039 (2015)
19. Liu, X., Ma, Y., Shi, Z., Chen, J.: Griddehazenet: Attention-based multi-scale network for image dehazing. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 7314–7323 (2019)
20. McCartney, E.J.: Optics of the atmosphere: scattering by molecules and particles. New York, John Wiley and Sons, Inc., 1976. 421 p. (1976)
21. Meng, G., Wang, Y., Duan, J., Xiang, S., Pan, C.: Efficient image dehazing with boundary constraint and contextual regularization. In: Proceedings of the IEEE international conference on computer vision. pp. 617–624 (2013)
22. Park, T., Liu, M.Y., Wang, T.C., Zhu, J.Y.: Semantic image synthesis with spatially-adaptive normalization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2337–2346 (2019)
23. Qin, X., Wang, Z., Bai, Y., Xie, X., Jia, H.: Ffa-net: Feature fusion attention network for single image dehazing. arXiv preprint arXiv:1911.07559 (2019)
24. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.H.: Single image dehazing via multi-scale convolutional neural networks. In: European conference on computer vision. pp. 154–169. Springer (2016)
25. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International journal of computer vision **115**(3), 211–252 (2015)
26. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. vol. 1, pp. I–I. IEEE (2003)
27. Shaham, T.R., Dekel, T., Michaeli, T.: Singan: Learning a generative model from a single natural image. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 4570–4580 (2019)
28. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from rgbd images. In: European conference on computer vision. pp. 746–760. Springer (2012)
29. Tan, R.T.: Visibility in bad weather from a single image. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8. IEEE (2008)
30. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022 (2016)
31. Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional gans. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8798–8807 (2018)
32. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C.: Esrgan: Enhanced super-resolution generative adversarial networks. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 0–0 (2018)
33. Xie, B., Guo, F., Cai, Z.: Improved single image dehazing using dark channel prior and multi-scale retinex. In: 2010 International Conference on Intelligent System Design and Engineering Application. vol. 1, pp. 848–851. IEEE (2010)
34. Xu, H., Guo, J., Liu, Q., Ye, L.: Fast image dehazing using improved dark channel prior. In: 2012 IEEE International Conference on Information Science and Technology. pp. 663–667. IEEE (2012)
35. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3194–3203 (2018)

36. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 586–595 (2018)
37. Zhu, J.Y., Krähenbühl, P., Shechtman, E., Efros, A.A.: Generative visual manipulation on the natural image manifold. In: European Conference on Computer Vision. pp. 597–613. Springer (2016)