

NightHaze: Nighttime Image Dehazing via Self-Prior Learning

Beibei Lin^{1*}, Yeying Jin^{1*}, Wending Yan², Wei Ye², Yuan Yuan², Robby T. Tan¹

¹National University of Singapore

²Huawei International Pte Ltd

{beibei.lin, e0178303}@u.nus.edu, {yan.wending, yewei10, yuanyuan10}@huawei.com, robbytan@nus.edu.sg

Abstract

Masked autoencoder (MAE) shows that severe augmentation during training produces robust representations for high-level tasks. This paper brings the MAE-like framework to nighttime image enhancement, demonstrating that severe augmentation during training produces strong network priors that are resilient to real-world night haze degradations. We propose a novel nighttime image dehazing method with self-prior learning. Our main novelty lies in the design of severe augmentation, which allows our model to learn robust priors. Unlike MAE that uses masking, we leverage two key challenging factors of nighttime images as augmentation: light effects and noise. During training, we intentionally degrade clear images by blending them with light effects as well as by adding noise, and subsequently restore the clear images. This enables our model to learn clear background priors. By increasing the noise values to approach as high as the pixel intensity values of the glow and light effect blended images, our augmentation becomes severe, resulting in stronger priors. While our self-prior learning is considerably effective in suppressing glow and revealing details of background scenes, in some cases, there are still some undesired artifacts that remain, particularly in the forms of over-suppression. To address these artifacts, we propose a self-refinement module based on the semi-supervised teacher-student framework. Our NightHaze, especially our MAE-like self-prior learning, shows that models trained with severe augmentation effectively improve the visibility of input haze images, approaching the clarity of clear nighttime images. Extensive experiments demonstrate that our NightHaze achieves state-of-the-art performance, outperforming existing nighttime image dehazing methods by a substantial margin of 15.5% for MUSIQ and 23.5% for ClipIQA.

Introduction

Dehazing night images deals not only with haze but also with nighttime degradations such as low light, noise, glow, uneven light distribution, etc. The combination of these degradations and the absence of paired real data make the task significantly challenging.

Existing supervised and semi-supervised nighttime image dehazing methods (Yan, Tan, and Dai 2020; Jin et al. 2023;

Liu et al. 2022) use either synthetic data or both synthetic and real data in training. Unfortunately, as shown in Figure 1, their performance is unsatisfactory when evaluated on real-world nighttime haze images. The main reason is that there is a large domain gap between synthetic haze and real-world haze, and unpaired real data cannot provide pixel-level guidance. As a result, models trained on these data lead to inferior performance when applied to real haze images. Table 1 shows five non-reference image quality assessment (NF-IQA) scores of these methods on real night haze data. Noticeably, there are significant gaps between the NF-IQA scores of clear night images and those of restored results obtained from existing methods.

Masked autoencoders (MAE) (He et al. 2022) are known to yield strong representations for high-level tasks. The core design of MAE is to use severe augmentation for training. MAE masks 75% of the input image before reconstructing it. This paper aims to bring MAE-like strong representations to nighttime image dehazing. However, MAE’s masking strategies are designed to capture structural spatial priors by filling in the masked regions. In contrast, nighttime image dehazing aims to restore a clear background behind haze and nighttime degradations.

To obtain strong priors that are resilient to real-world night haze degradations, we design the severe augmentation for our learning process. Our main novelty lies in defining “severe” based on proper augmentation. Specifically, we first find two key challenging factors of nighttime images for augmentation: light effects and noise. During training, we intentionally degrade clear nighttime images by blending them with various light effects, including glow, and adding noise. As demonstrated in MAE (He et al. 2022), the severity of augmentation is key to yield strong feature representations or priors. Accordingly, we define “severe augmentation” as noise values approaching as high as the pixel intensity values blended with strong glow and other night light effects.

Based on our severe augmentation, we propose NightHaze, a simple yet effective nighttime image dehazing method with self-prior learning. Our main novelty, the design of severe augmentation, allows our network to learn the strong priors that are robust to real-world haze effects. Once the strong priors are obtained, our network can effectively enhance the visibility of nighttime haze images. Like MAE,

*These authors contributed equally.



Figure 1: Qualitative results from NightEnhance’23 (Jin et al. 2023), NightDeFog’20 (Yan, Tan, and Dai 2020), DiT’23 (Peebles and Xie 2023) and our method, on the real-world dataset. Ours not only suppress glow but also reveal the detailed textures of the night scenes, including those under low light and strong glow.

we apply severe augmentation and then train our network to restore the clear background. Table 1 shows that the NF-IQA scores of our self-prior learning are closer to the NF-IQA scores of clear night images.

While our self-prior learning effectively can enhance the visibility of night haze images, in some cases, artifacts can persist, typically in the form of over-suppression. To refine the output of our self-prior learning, we introduce a self-refinement module based on the teacher-student framework (Tarvainen and Valpola 2017). Unlike standard teacher-student frameworks that directly pass knowledge from the student to the teacher, we utilize NF-IQA scores to monitor the quality of knowledge from the student. This allows us to pass only high-quality information to the teacher, resulting in superior performance. Figure 1 shows our results in comparison to those of the state-of-the-art methods. The main contributions are as follows:

- We propose self-prior learning to bring MAE-like strong representations to nighttime image dehazing. Our main novelty lies in the design of severe augmentation, which enables our model to learn clear background priors.
- We propose self-refinement to refine our model by utilizing real-world nighttime haze images. Unlike the standard teacher-student framework, we introduce NF-IQA scores to assess the knowledge quality from our stu-

dent model’s learning. This approach avoids inaccurate knowledge transfer from our student to our teacher.

- Extensive experiments demonstrate that our Nighthaze achieves a significant performance improvement on real-world nighttime haze images. Our Nighthaze attains outstanding scores of 52.87 in MUSIQ and 76.88 in Clip-IQA, significantly exceeding existing nighttime image dehazing methods with a substantial margin of 15.5% for MUSIQ and 23.5% for Clip-IQA.

Related Work

Nighttime Image Dehazing Previous image dehazing methods (Tan 2008; Fattal 2008; He, Sun, and Tang 2011; Berman, Treibitz, and Avidan 2018) remove haze effects based on the haze model. Specifically, they estimate transmission maps and atmosphere lights to recover the clear background. However, nighttime haze images always suffer from noise. Thus, these methods are hard to estimate accurate transmission maps and atmosphere lights, resulting in inferior performance. Recent deep learning-based daytime image dehazing methods (Ye et al. 2022; Yu et al. 2022; Dong et al. 2020; Qin et al. 2020; Zheng et al. 2021; Wu et al. 2021; Guo et al. 2022; Song et al. 2023; Shao et al. 2020; Chen et al. 2021; Li et al. 2022, 2020, 2021; Huang et al. 2019; Liu et al. 2020; Zhao et al. 2021) propose differ-

ent networks to restore haze images. These methods can address nighttime haze images by training on synthetic nighttime dehazing datasets. However, these methods struggle to handle nighttime haze images as they often neglect complex nighttime degradation factors.

Existing optimization-based nighttime dehazing methods (Wang, Wang, and Liu 2022; Ancuti et al. 2016, 2020; Zhang, Cao, and Wang 2014; Zhang et al. 2017; Tang et al. 2021; Liu et al. 2022) address nighttime haze effects using the atmospheric scattering model, new imaging model, among others. However, these methods struggle to handle challenging hazy regions affected by significant noise and complex lighting conditions. The main issue is the difficulty in estimating atmospheric lights in these regions.

Existing learning-based nighttime dehazing methods (Zhang et al. 2020; Yan, Tan, and Dai 2020; Kuanar et al. 2022; Jin et al. 2023) propose CNNs for nighttime image dehazing. However, the synthetic datasets they used were constructed using less challenging augmentations, such as regions affected by light noise. Models trained using these datasets struggle to restore clear backgrounds in challenging real-world haze images, including those affected by significant noise and complex lighting conditions. Unsupervised methods (Yan, Tan, and Dai 2020; Jin et al. 2023), which are based on unpaired real-world data, cannot provide pixel-level guidance for training.

Image Quality Assessment To qualitatively evaluate the performance of real-world haze images without the corresponding ground truth, we introduce several image quality assessment methods, including MUSIQ (Ke et al. 2021), ManIQA (Yang et al. 2022), ClipIQA (Wang, Chan, and Loy 2023), HyperIQA (Su et al. 2020), and TRES (Golestaneh, Dadsetan, and Kitani 2022). MUSIQ (Ke et al. 2021) is built based on a transformer network and evaluate the image quality at multiple image scales. ManIQA (Yang et al. 2022) integrates the global and local information of images to assess the image quality. HyperIQA (Su et al. 2020) assesses in-the-wild image quality by using a self-adaptive hyper network architecture. TRES (Golestaneh, Dadsetan, and Kitani 2022) leverages transformers, relative ranking, and self-consistency losses to effectively evaluate the image quality. ClipIQA (Wang, Chan, and Loy 2023) introduces Contrastive Language-Image Pre-training (CLIP) models to evaluate the quality perception and abstract perception of images.

Proposed Method

The overview of our NightHaze is shown in Figure 2. It includes two components: self-prior learning and self-refinement. Our self-prior learning aims to bring the MAE-like framework to nighttime image enhancement, resulting in strong network priors that are resilient to real-world night haze effects. Our self-refinement utilizes unlabeled nighttime haze images to refine the dehazing ability of our network in the real scenes in a semi-supervised manner.

Self-Prior Learning

Our self-prior learning aims to learn strong network priors of clear backgrounds using severe augmentation. As shown in Figure 2, our training process is similar to the approach in MAE, where we severely augment clear images and force the network to recover their clear backgrounds. MAE employs masking strategies to augment the input images. This masking forces the network to learn the structural spatial priors of the world by predicting the contents in the masked regions. However, nighttime image dehazing is about restoring a clear background behind haze and nighttime degradations. Hence, masking is not a suitable augmentation for dehazing.

Nighttime haze degradations are mainly caused by glow light effects and noise. Due to this, we degrade clear images by blending them with light effects including glow and adding noise and then we restore the clear images, as demonstrated in Figure 3. The augmentation can be formulated as:

$$I = W_b * J + (1 - W_b) * L + \epsilon, \quad (1)$$

where I is the augmented image, J is the clear image, W_b is the blend weight map, L is the light map, ϵ is the noise. We explain the details of each term in Eq. (1) as follows.

Light Map L We manually capture atmosphere light maps from real haze images. When regions in a night image suffer from dense haze effects (where the signal is less than the noise), they can be regarded as atmospheric lights. Hence, we collect these regions from real haze images as light maps. Given an input image $J_i \in \mathbb{R}^{C_i \times H_i \times W_i}$, where C_i is the number of channels and (H_i, W_i) is image size, we first sample a light map and resize it to $\mathbb{R}^{C_i \times H_i \times W_i}$. Moreover, to simulate glow effects, we use Gaussian kernels to enhance the brightness of several random regions in the light maps. Figure 4 shows our augmented light maps.

Blending Weight Map W_b As the goal of our blending weight W_b is to mix a clear image and light maps, the range of the blending weight W_b is $(0, 1)$. To obtain the blending weight W_b , we first sample a value $t \in \mathbb{R}^1$ from a uniform distribution in $[T_{Low}, T_{High}]$, where both T_{Low} and T_{High} are lower than 1. We then expand $t \in \mathbb{R}^1$ to a map $W_b \in \mathbb{R}^{C_i \times H_i \times W_i}$. Moreover, we randomly adjust some regions' blending weights. Figure 5 visualizes several blending weight maps. Our blending weight map is inspired by transmission maps of the haze formulation. However, transmission maps rely on the depth of images for their generation. Obtaining accurate depth maps in nighttime images is a challenging problem. A straightforward approach is to use monocular depth estimates, but this may introduce errors. Unlike transmission maps, we randomly generate uneven blending weights for an image. This approach enables us to simply and effectively augment clear images.

Noise ϵ We take Gaussian noise as the noise ϵ , which can be formulated as $W_\epsilon * \mathcal{N}(0, 1)$, where W_ϵ is the weight of the Gaussian noise. Theoretically, $\mathcal{N}(0, 1)$ ranges from negative infinity to positive infinity, with about approximately 99.7% of the values fall within the range of -3 to 3 . We remove the rest 0.3% values and the negative values of $\mathcal{N}(0, 1)$. Thus, the range of our noise term ϵ is $(0, 3 * W_\epsilon)$.

Severe Augmentation MAE demonstrates that severe augmentation is key to obtaining strong representations. Mo-

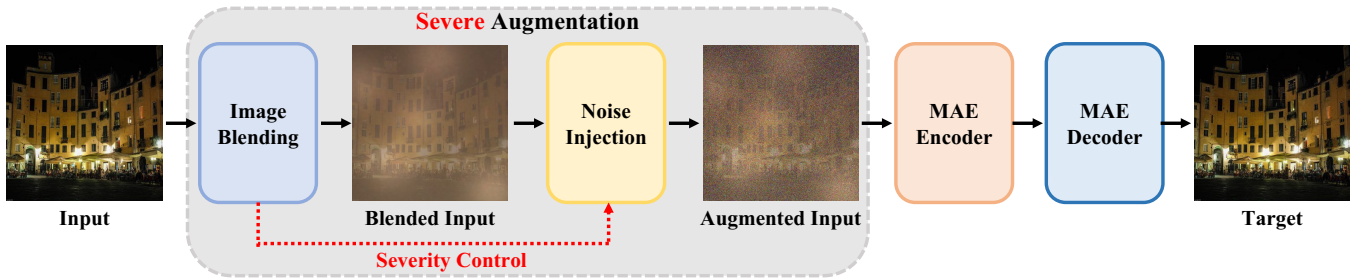


Figure 2: Overview of our self-prior learning. This approach aims to produce strong priors that are resilient to real-world night haze degradations. Given a clear nighttime image, we blend clear images with various maps collected from real-world haze images and add noise to these blended images. The augmented input is then fed into an Encoder-Decoder framework to recover the clear background. The loss is a simple L1 loss between the reconstructed image and the clear input. The key to self-prior learning is the severity of the augmentation. By increasing the noise values to approach as high as the pixel intensity values of the glow and light effect blended images, our augmentation becomes severe, resulting in stronger priors.

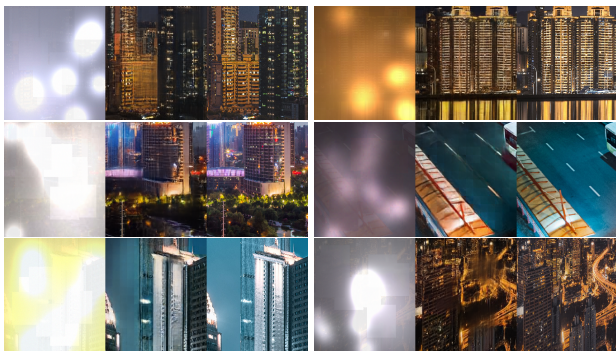


Figure 3: Example results on validation sets. For each triplet, we show the augmented image (left), our restoration (middle), and the ground-truth (right).

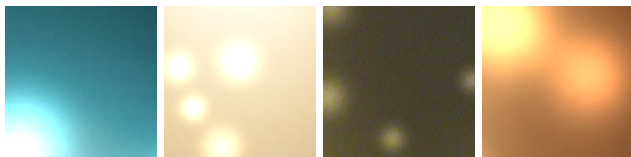


Figure 4: Visualization of different light maps that contain glow effects. We use Gaussian kernels to simulate glow effects. By adjusting the parameters of kernels, we can control the number, size, and brightness of glow regions.

tivated by this, we define “severe augmentation” as ϵ approaching $W_b * J$ according to Eq. (1). The main reason is that the level of the noise ϵ affects the difficulty of nighttime image restoration. To be more specific, we use min-max normalization to normalize the clear image and thus the range of J is $(0, 1)$. After adding light effects, the range of the term $W_b * J$ becomes to $(0, T_{\text{High}})$. Since its range depend on the value of T_{High} , we further define W_ϵ as $W_n * T_{\text{High}}$ and thus the range of ϵ becomes to $(0, 3 * W_n * T_{\text{High}})$. We empirically find that when $W_n = 0.1$ and $\epsilon \in [0, 0.3 * T_{\text{High}}]$, our augmentation is difficult enough.

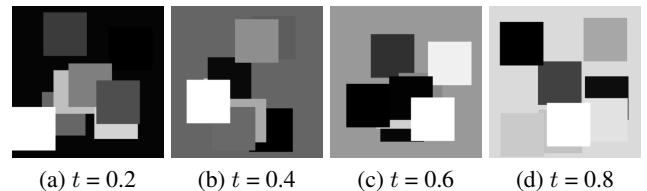


Figure 5: Visualization of blending weight maps, where the black and white regions represent low and high blending values. By adjusting the values of blending weights and the noise term, we can control the severity of our augmentation.

In light maps, we use Gaussian kernels to increase the brightness of several random regions. These regions can overlap, resulting in uneven light distribution in the augmented light maps. Moreover, due to the overlapping brightness enhancements, the glow effects can become oversaturated. When blending clear images with these oversaturated light effects, their signals can dominate, making restoration challenging.

Implementation Given a clear nighttime image, we first sample a light map L from our dataset and randomly enhance the brightness of several regions. Second, we sample a value t from a uniform distribution to generate a uniform map. We then randomly adjust the values of several regions in this uniform map to create the blending weight map W_b . Third, we generate a significant Gaussian noise map ϵ . Finally, we augment the input clear image using Eq. (1), as illustrated in Figure 3 which shows several augmented inputs. During the training stage, we calculate the L1 loss between the reconstructed and original clear images.

Self-Refinement

Although our self-prior learning effectively handles haze effects, it occasionally results in artifacts, particularly in the form of over-suppression. To address this issue, we propose a teacher-student framework (Lin et al. 2024; Chen et al. 2025) to further refine our network using real haze images. The network parameter from our self-prior learn-

Non-reference metrics	MUSIQ	TRES	HyperIQA	CLIPQA	MANIQA	Contrast
Statistics from clear images	53.91	78.69	0.6819	0.6834	0.4319	50.96
Statistics from haze images	42.12	57.94	0.5186	0.5267	0.2556	22.21
Backbone with syn-training	44.25	60.84	0.5903	0.5449	0.2665	24.08
NightDeFog (Yan, Tan, and Dai 2020)	45.76	65.83	0.6133	0.5146	0.3318	57.60
NightVDM (Liu et al. 2022)	40.90	53.87	0.5241	0.4656	0.2332	27.47
Restormer (Zamir et al. 2022)	39.13	52.97	0.4956	0.4763	0.2428	24.58
Uformer (Wang et al. 2022)	39.14	52.60	0.4902	0.4427	0.2420	25.03
WeatherDiffusion (Özdenizci and Legenstein 2023)	38.98	52.33	0.4868	0.5044	0.2338	24.04
NightEnhance (Jin et al. 2023)	44.33	60.07	0.5379	0.4736	0.2354	24.57
DiT (Peebles and Xie 2023)	41.02	55.54	0.5259	0.5059	0.2513	23.94
Ours	52.87	76.88	0.6747	0.6360	0.4009	44.35

Table 1: Quantitative comparison on the RealNightHaze dataset. For non-reference metrics, higher is better. In terms of image contrast, the better it is when the contrast closely matches the statistical value of a clear image. NightDeFog, NightVDM and NightEnhance are nighttime image dehazing methods, while the rest methods are image restoration backbones. “syn-training” means training our backbone on synthetic datasets.

ing is used to initialize a teacher model w_T and a student model w_S . The purpose of the teacher model is to generate the pseudo ground truth based on confidence score, while the student model is trained on high-confidence predictions and their corresponding inputs. By gradually learning from high-confidence predictions, our network is allowed to address more challenging regions.

Given a real-world hazy dataset $\mathbf{D}_{\text{uh}} = \{\mathbf{x}_i^{\text{uh}}\}_{i=1}^{N_{\text{uh}}}$, where \mathbf{x}_i^{uh} is the i -th haze image and N_{uh} is the number of haze images, we utilize the teacher model w_T to generate predictions \mathbf{y}_i^{uh} and confidence maps \mathbf{u}_i^{uh} . Specifically, we first sample overlapping regions from \mathbf{x}_i^{uh} and input the set of overlapping regions into our teacher model w_T to generate predictions. Then, we calculate the mean values and variance of overlapping regions as predictions and confidence maps. Subsequently, we use a threshold $\mathbf{v}_{\text{thr}}^1$ to convert the confidence map \mathbf{u}_i^{uh} to a binary mask \mathbf{m}_i^{uh} .

During the training stage, we use the noise and haze augmentation to degrade the haze images \mathbf{x}_i^{uh} . The degraded haze images $\hat{\mathbf{x}}_i^{\text{uh}}$ are then fed into our student model for training. The loss function can be formulated as:

$$\text{Loss} = \frac{1}{N} \sum_{i=1}^N |\hat{\mathbf{y}}_i^{\text{uh}} - \mathbf{y}_i^{\text{uh}}| * \mathbf{m}_i^{\text{uh}}, \quad (2)$$

where N is the number of samples in the training batch and $\hat{\mathbf{y}}_i^{\text{uh}}$ is the output of the degraded haze images $\hat{\mathbf{x}}_i^{\text{uh}}$. Once the losses are obtained, traditional teacher-student frameworks use back-propagation to update the student model w_S . Then they use the Exponential Moving Average (EMA) to update the teacher model. However, the student model is trained using augmented real-world haze inputs. Since the augmentation is random, the updates of the student model in some training steps may be wrong and thus affect the performance of the teacher model by EMA. To avoid this, we introduce a non-reference metric to monitor each update of the student model, which can be formulated as:

$$\text{Score} = F_{\text{IQA}}(F(w_s^{t+1}, \mathbf{x}_i^{\text{uh}})) + F_{\text{IQA}}(F(w_s^t, \mathbf{x}_i^{\text{uh}})), \quad (3)$$

where $F_{\text{IQA}}(\cdot)$ represents non-reference image assessment metrics, $F(w_s^{t+1}, \mathbf{x}_i^{\text{uh}})$ and $F(w_s^t, \mathbf{x}_i^{\text{uh}})$ are \mathbf{x}_i^{uh} 's outputs generated using the student weights w_s^{t+1} and w_s^t . w_s^t is the current weight of the student model, while w_s^{t+1} is the updated weight. If $\text{Score} > \mathbf{v}_{\text{thr}}^2$, the student model will accept this update and pass the knowledge to the teacher model. Otherwise, the student model will not accept this update.

Experiments

Due to the lack of ground truth, previous methods typically use synthetic datasets for quantitative evaluation and real-world datasets for qualitative evaluation. However, the quantitative evaluation on synthetic datasets cannot effectively represent the real-world dehazing capabilities. Motivated by this, we introduce a new benchmark to quantitatively evaluate the dehazing performance on real-world nighttime haze datasets. Specifically, we use five non-reference image quality assessments, including MUSIQ (Ke et al. 2021), ManIQA (Yang et al. 2022), ClipIQA (Wang, Chan, and Loy 2023), HyperIQA (Su et al. 2020), and TRES (Golestaneh, Dadsetan, and Kitani 2022), along with image contrast, to evaluate the performance of existing methods.

Our real-world nighttime haze datasets, *RealNightHaze*, is collected from internet and existing nighttime dehazing methods (Zhang et al. 2020; Jin et al. 2023), and consists of 440 real-world nighttime haze images.

Implementation Details

Our NightHaze includes two parts: self-prior learning and self-refinement. We adopt the MAE network (He et al. 2022) as our backbone. In our self-prior learning, the size of input images is 224×224 . During the training stage, we use Adam as our optimizer and set the initial learning rate to $1.5e-4$. The total training steps and the training batch size

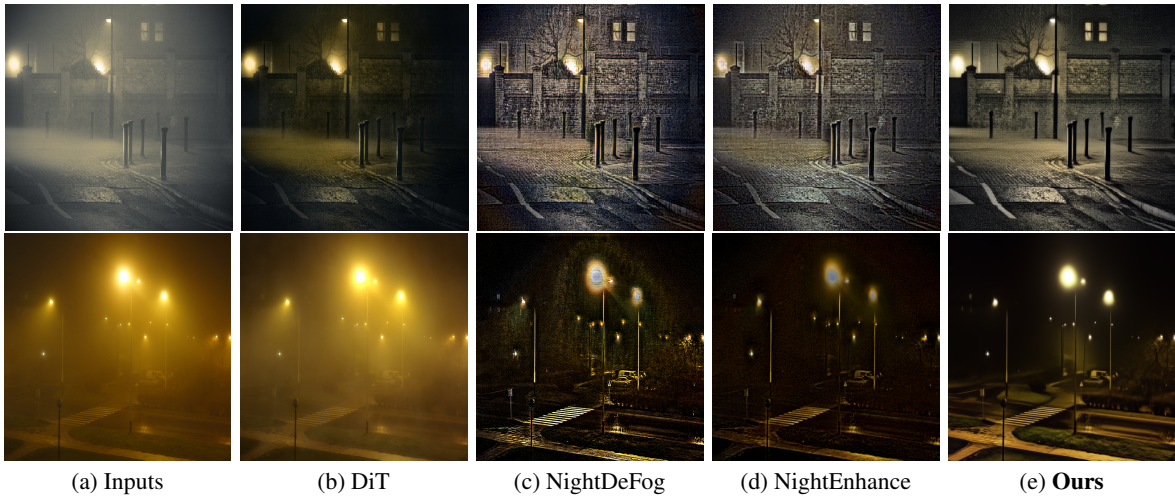


Figure 6: Qualitative results from NightEnhance’23 (Jin et al. 2023), NightDeFog’20 (Yan, Tan, and Dai 2020), DiT’23 (Peebles and Xie 2023), and our method, on the real-world dataset. Our method not only handles glow effects but also reveals detailed textures in low-light regions, such as walls and streets. Zoom in for better visualisation.

are set to 20,000 and 128, respectively. Our self-prior learning consists of three parts: the light map, the blending weight map, and noise.

Light Map At each training step, we randomly select an atmospheric light map for every clear image. Then, we select various regions and amplify their brightness using Gaussian kernels. The number of selected regions is between 2 and 10, and the size of Gaussian kernels ranges from 15 to 80.

Blending Weight Map We initially select a blending weight t from a uniform distribution $(T_{\text{low}}, T_{\text{high}})$ and expand it into a map. Then, we randomly adjust the blending weights of several regions. T_{low} and T_{high} are set to 0.001 and 0.1, respectively. We set the number of the selected regions to 8. The size of each region is 64×64 and the range of the changed value is a uniform distribution $(0, 0.04)$.

Noise During the training stage, we use Gaussian noise to augment our input images. We set W_n to 0.1. Thus, the value range of our ϵ becomes to $(0, 0.03)$.

Self-Refinement For a haze image of size 256×256 , we first sample overlapping regions within the input haze image. The image size of the overlapping regions is 224×224 and the stride of the overlap sampling is 4. Once the pseudo ground truths are obtained, we randomly crop 224×224 regions from both the haze images and the pseudo ground truths for training our student model. We use Adam as our optimizer, with the initial learning rate set to $2e-5$. We set the training step and the training batch size to 10,000 and 16, respectively. The EMA weight is set to 0.9999. The thresholds v_{thr}^1 and v_{thr}^2 are set to 0.005 and 0.

Quantitative Evaluation on RealNightHaze

We compare our NightHaze method on RealNightHaze with state-of-the-art (SOTA) methods, including NightEnhance (Jin et al. 2023), NightDeFog (Yan, Tan, and Dai 2020), NightVDM (Liu et al. 2022), Uformer (Wang et al. 2022), Restormer (Zamir et al. 2022), WeatherDiffu-

sion (Özdenizci and Legenstein 2023) and DiT (Peebles and Xie 2023). Table 1 shows that our method achieves a significant performance improvement in all non-reference metrics. For example, the MUSIQ and TRES of NightDeFog are 45.76 and 65.83, respectively. Our method achieves 52.87 of MUSIQ and 76.88 of TRES, which outperforms NightDeFog by 7.11 and 11.05, respectively. Furthermore, it can be observed that our image quality scores are closer to those of clean night images. This is because we propose self-prior learning, which brings MAE-like strong representations to nighttime image enhancement. By utilizing severe augmentation, we can generate strong priors that demonstrate robustness against real-world haze effects, thereby leading to superior performance.

Additionally, we compare our method with other powerful image restoration backbones, including Uformer (Wang et al. 2022), Restormer (Zamir et al. 2022), WeatherDiffusion (Özdenizci and Legenstein 2023) and DiT (Peebles and Xie 2023). The experimental results are shown in Table 1. It can be observed that existing image restoration methods trained using synthetic datasets struggle to restore haze images. The main reason is that their learning strategies, based on synthetic datasets, struggle to develop strong priors. This results in inferior performance.

Qualitative Evaluation on RealNightHaze

Figure 6 presents a qualitative comparison on RealNightHaze. It is evident that existing nighttime image dehazing methods cannot effectively recover a clean background in challenging areas, such as those with low-light and glow effects. The main reason is that there is a large domain gap between their synthetic datasets and nighttime haze degradation. In contrast, our NightHaze method not only removes nighttime haze effects but also enhances the visibility of images. This is because our self-prior learning effectively forces our network to learn strong priors. Similar to MAE,

Non-reference metrics	MUSIQ	TRES	HyperIqa	CLIQQA	MANIQA	Contrast
Statistics from clear images	53.91	78.69	0.6819	0.6834	0.4319	50.96
Statistics from haze images	42.12	57.94	0.5186	0.5267	0.2556	22.21
Backbone with syn-training	44.25	60.84	0.5903	0.5449	0.2665	24.08
Backbone with 0% severe	44.67	62.18	0.5773	0.5364	0.2930	23.40
Backbone with 25% severe	48.64	67.96	0.6253	0.5715	0.3361	28.73
Backbone with 50% severe	49.12	68.74	0.6286	0.5750	0.3417	29.80
Backbone with 75% severe	49.31	68.92	0.6289	0.5891	0.3420	31.15
Backbone with 100% severe	52.87	76.88	0.6747	0.6360	0.4009	44.35

Table 2: Ablation studies on the RealNightHaze dataset. For non-reference metrics, higher is better. In terms of image contrast, the better it is when the contrast closely matches the statistical value of a clear image. “xx% severe” means xx% of the training data was severely augmented, while the remaining data was non-severely augmented. “syn-training” means training our backbone on synthetic datasets.

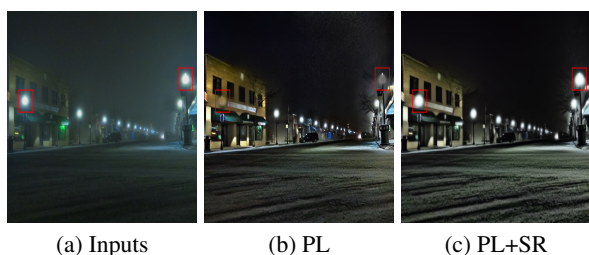


Figure 7: Ablation studies on real-world haze images. ‘PL’ and ‘SR’ are self-prior learning and self-refinement, respectively. (a) Input (b) Results obtained using our self-prior learning. (c) Results obtained using our self-prior learning and self-refinement. Zoom in for better visualisation.

our self-prior learning is also built on augmentation. In our case, we utilize two key challenging aspects of nighttime images as augmentation factors: light effects and noise. Following the same principle as MAE, we define and employ severe augmentation to obtain a strong prior, which significantly enhances performance. Furthermore, we employ a self-refinement module to further refine our model through learning from real haze images. To be more specific, our teacher model generates high-confidence predictions from real haze images to guide the student model’s learning. This method enables our model to gradually adapt to real-world high-confidence predictions, thereby addressing more challenging scenarios.

Ablation Studies

To verify the effectiveness of our key ideas, we conduct ablation studies on the RealNightHaze dataset. The experimental results are shown in Figure 7 and Table 2.

Analysis of Self-Prior Learning To verify the effectiveness of our self-prior learning, we use different ratios to mix severe and non-severe augmentation. The experimental results are shown in Table 2. It can be observed that our backbone with non-severe augmentation cannot sufficiently enhance the visibility of haze images. The MUSIQ and TRES scores

are 44.67 and 62.18, outperforming the original haze images’ scores by 2.55 and 4.24. In contrast, our backbone with severe augmentation achieves significant performance improvement. The main reason for this is that severe augmentation yields strong priors that are robust to real haze.

Analysis of Self-Refinement We conduct ablation studies to verify the effectiveness of our self-refinement. The experimental results are shown in Figure 7. Although our self-prior learning can effectively enhance the visibility of night haze images, it still leaves some artifacts caused by over-suppression. For example, the light sources in Figure 7(b) are over-suppressed. Figure 7(c) demonstrates that our self-refinement further refines our outputs.

Conclusion

This paper brings the MAE-like framework to nighttime image restoration and shows that severe augmentation during training yields strong priors that are robust to real-world night haze effects. We present NightHaze, a novel nighttime dehazing method with self-prior learning. Our self-prior learning is simple: we intentionally augment clear nighttime images by blending them with various light effects including glow and adding noise. Specifically, we blend clear images with various light maps containing artificial glow and add noise to these blended images. Inspired by MAE, which indicates that the severity of augmentation is key to obtaining strong priors, we define severe augmentation using our night-specific augmentations. We show that severe augmentation in training yields strong priors that are robust to real-world haze. Although our self-prior learning can effectively handle haze effects, it still leaves some artifacts caused by over-suppression. Our self-refinement is proposed to address this issue. It is built on the semi-supervised teacher-student framework. Unlike traditional teacher-student frameworks, we incorporate non-reference metrics to monitor each update of our student model. This approach allows us to prevent inaccurate knowledge transfer from the student to the teacher. Extensive experimental results demonstrate that our NightHaze achieves a significant performance improvement on real-world nighttime hazy datasets.

References

- Ancuti, C.; Ancuti, C. O.; De Vleeschouwer, C.; and Bovik, A. C. 2016. Night-time dehazing by fusion. In *2016 IEEE International Conference on Image Processing (ICIP)*, 2256–2260. IEEE.
- Ancuti, C.; Ancuti, C. O.; De Vleeschouwer, C.; and Bovik, A. C. 2020. Day and night-time dehazing by local airlight estimation. *IEEE Transactions on Image Processing*, 29: 6264–6275.
- Berman, D.; Treibitz, T.; and Avidan, S. 2018. Single image dehazing using haze-lines. *IEEE transactions on pattern analysis and machine intelligence*, 42(3): 720–734.
- Chen, T.; Lin, B.; Jin, Y.; Yan, W.; Ye, W.; Yuan, Y.; and Tan, R. T. 2025. Dual-Rain: Video Rain Removal Using Assertive and Gentle Teachers. In *ECCV*, 127–143. Cham: Springer Nature Switzerland.
- Chen, Z.; Wang, Y.; Yang, Y.; and Liu, D. 2021. PSD: Principled synthetic-to-real dehazing guided by physical priors. In *CVPR*, 7180–7189.
- Dong, H.; Pan, J.; Xiang, L.; Hu, Z.; Zhang, X.; Wang, F.; and Yang, M.-H. 2020. Multi-scale boosted dehazing network with dense feature fusion. In *CVPR*, 2157–2167.
- Fattal, R. 2008. Single image dehazing. *ACM transactions on graphics (TOG)*, 27(3): 1–9.
- Golestaneh, S. A.; Dadsetan, S.; and Kitani, K. M. 2022. No-reference image quality assessment via transformers, relative ranking, and self-consistency. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1220–1230.
- Guo, C.-L.; Yan, Q.; Anwar, S.; Cong, R.; Ren, W.; and Li, C. 2022. Image Dehazing Transformer with Transmission-Aware 3D Position Embedding. In *CVPR*, 5812–5820.
- He, K.; Chen, X.; Xie, S.; Li, Y.; Dollár, P.; and Girshick, R. 2022. Masked autoencoders are scalable vision learners. In *CVPR*, 16000–16009.
- He, K.; Sun, J.; and Tang, X. 2011. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12): 2341–2353.
- Huang, L.-Y.; Yin, J.-L.; Chen, B.-H.; and Ye, S.-Z. 2019. Towards Unsupervised Single Image Dehazing With Deep Learning. In *2019 IEEE International Conference on Image Processing (ICIP)*, 2741–2745. IEEE.
- Jin, Y.; Lin, B.; Yan, W.; Yuan, Y.; Ye, W.; and Tan, R. T. 2023. Enhancing visibility in nighttime haze images using guided apsf and gradient adaptive convolution. In *Proceedings of the 31st ACM International Conference on Multimedia*, 2446–2457.
- Ke, J.; Wang, Q.; Wang, Y.; Milanfar, P.; and Yang, F. 2021. Musiq: Multi-scale image quality transformer. In *ICCV*, 5148–5157.
- Kuanar, S.; Mahapatra, D.; Bilas, M.; and Rao, K. 2022. Multi-path dilated convolution network for haze and glow removal in nighttime images. *The Visual Computer*, 38(3): 1121–1134.
- Li, B.; Gou, Y.; Gu, S.; Liu, J. Z.; Zhou, J. T.; and Peng, X. 2021. You only look yourself: Unsupervised and untrained single image dehazing neural network. *International Journal of Computer Vision*, 129(5): 1754–1767.
- Li, B.; Gou, Y.; Liu, J. Z.; Zhu, H.; Zhou, J. T.; and Peng, X. 2020. Zero-shot image dehazing. *IEEE Transactions on Image Processing*, 29: 8457–8466.
- Li, Y.; Chang, Y.; Gao, Y.; Yu, C.; and Yan, L. 2022. Physically Disentangled Intra-and Inter-Domain Adaptation for Varicolored Haze Removal. In *CVPR*, 5841–5850.
- Lin, B.; Jin, Y.; Yan, W.; Ye, W.; Yuan, Y.; Zhang, S.; and Tan, R. T. 2024. NightRain: Nighttime Video Deraining via Adaptive-Rain-Removal and Adaptive-Correction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 3378–3385.
- Liu, W.; Hou, X.; Duan, J.; and Qiu, G. 2020. End-to-end single image fog removal using enhanced cycle consistent adversarial networks. *IEEE Transactions on Image Processing*, 29: 7819–7833.
- Liu, Y.; Yan, Z.; Wu, A.; Ye, T.; and Li, Y. 2022. Night-time Image Dehazing Based on Variational Decomposition Model. In *CVPR*, 640–649.
- Özdenizci, O.; and Legenstein, R. 2023. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Peebles, W.; and Xie, S. 2023. Scalable Diffusion Models with Transformers. In *ICCV*, 4195–4205.
- Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; and Jia, H. 2020. FFA-Net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 11908–11915.
- Shao, Y.; Li, L.; Ren, W.; Gao, C.; and Sang, N. 2020. Domain Adaptation for Image Dehazing. In *CVPR*, 2808–2817.
- Song, Y.; He, Z.; Qian, H.; and Du, X. 2023. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32: 1927–1941.
- Su, S.; Yan, Q.; Zhu, Y.; Zhang, C.; Ge, X.; Sun, J.; and Zhang, Y. 2020. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In *CVPR*, 3667–3676.
- Tan, R. T. 2008. Visibility in bad weather from a single image. In *CVPR*, 1–8. IEEE.
- Tang, Q.; Yang, J.; He, X.; Jia, W.; Zhang, Q.; and Liu, H. 2021. Nighttime image dehazing based on Retinex and dark channel prior using Taylor series expansion. *Computer Vision and Image Understanding*, 202: 103086.
- Tarvainen, A.; and Valpola, H. 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30.
- Wang, J.; Chan, K. C.; and Loy, C. C. 2023. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 2555–2563.

Wang, W.; Wang, A.; and Liu, C. 2022. Variational Single Nighttime Image Haze Removal With a Gray Haze-Line Prior. *IEEE Transactions on Image Processing*, 31: 1349–1363.

Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; and Li, H. 2022. Uformer: A general u-shaped transformer for image restoration. In *CVPR*.

Wu, H.; Qu, Y.; Lin, S.; Zhou, J.; Qiao, R.; Zhang, Z.; Xie, Y.; and Ma, L. 2021. Contrastive learning for compact single image dehazing. In *CVPR*, 10551–10560.

Yan, W.; Tan, R. T.; and Dai, D. 2020. Nighttime defogging using high-low frequency decomposition and grayscale-color networks. In *European Conference on Computer Vision*, 473–488. Springer.

Yang, S.; Wu, T.; Shi, S.; Lao, S.; Gong, Y.; Cao, M.; Wang, J.; and Yang, Y. 2022. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *CVPR*, 1191–1200.

Ye, T.; Zhang, Y.; Jiang, M.; Chen, L.; Liu, Y.; Chen, S.; and Chen, E. 2022. Perceiving and Modeling Density for Image Dehazing. In *Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX*, 130–145. Springer.

Yu, H.; Zheng, N.; Zhou, M.; Huang, J.; Xiao, Z.; and Zhao, F. 2022. Frequency and spatial dual guidance for image dehazing. In *Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX*, 181–198. Springer.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 5728–5739.

Zhang, J.; Cao, Y.; Fang, S.; Kang, Y.; and Wen Chen, C. 2017. Fast haze removal for nighttime image using maximum reflectance prior. In *CVPR*, 7418–7426.

Zhang, J.; Cao, Y.; and Wang, Z. 2014. Nighttime haze removal based on a new imaging model. 4557–4561. IEEE.

Zhang, J.; Cao, Y.; Zha, Z.-J.; and Tao, D. 2020. Nighttime dehazing with a synthetic benchmark. In *Proceedings of the 28th ACM International Conference on Multimedia*, 2355–2363.

Zhao, S.; Zhang, L.; Shen, Y.; and Zhou, Y. 2021. RefinedNet: A weakly supervised refinement framework for single image dehazing. *IEEE Transactions on Image Processing*, 30: 3391–3404.

Zheng, Z.; Ren, W.; Cao, X.; Hu, X.; Wang, T.; Song, F.; and Jia, X. 2021. Ultra-high-definition image dehazing via multi-guided bilateral learning. In *CVPR*, 16180–16189. IEEE.