Adversarial Auto-Augmentation for Data-Efficient Single Image Dehazing

Anonymous ICCV submission

Paper ID ****

Abstract

Supervised learning-based image dehazing algorithms are sensitive to degradation and training distribution, making them ill-suited for out-of-domain non-uniform restoration. We propose an adversarial auto-augmentation approach to address this limitation without explicitly collecting paired training data. Specifically, we generate images with a broad distribution representative of multiple domains by varying the degradation and color profiles achieved by leveraging new augmentation techniques, including mean-variance transfer, physically accurate atmospheric scattering model, and localized degradation generation. These techniques effectively account for non-homogeneous degradations, enhancing the robustness of the underlying degradation model. Apart from utilizing these synthetic negative images to train the underlying network, these also provide diverse image representations for enabling more effective contrastive regularization. In addition to the training modifications, we propose a frequency-based feature fusion mechanism that prioritizes semantic and structural information from the decoder and encoder. Finally, we incorporate depth and color attenuation priors to ensure perceptually pleasing and physically accurate restoration quality. To evaluate the efficacy of the proposed mechanism, we perform comprehensive experiments and obtain state-of-the-art (SoTA) results while achieving high fidelity and improving the performance of perception-based algorithms without fine tuning.

1. Introduction

Image dehazing has emerged as a widely researched field, driven by enhancing scene visibility and recovering regions affected by the haze. Such algorithms play a vital role in improving the perceptual fidelity of images [61, 60, 68, 25, 42, 56], benefiting human vision and facilitating high-level perception tasks such as semantic segmentation [53, 54, 57, 33], depth estimation [7] and object detection [28, 12, 34]. SoTA techniques in image dehazing have made significant strides by utilizing paired datasets to train restoration networks in an end-to-end mechanism. Despite their advancements, the performance of supervised learning algorithms relies heavily on specific training distributions, making them susceptible to poor performance when applied to images outside training distribution, as shown in Fig. 1. This limitation necessitates exploring more robust and adaptable approaches to overcome the sensitivity to training distributions and achieve better generalization capabilities.

Figure 1. Demonstration of dehazing on images (top) from training distribution and (bottom) outside training distribution using AECRNet [68] as the baseline on Images from NH-Haze2 [5] and ACDC [55] datasets, respectively.

Since such algorithms rely on large paired training datasets, which can be time-consuming and resource-intensive to gather, alternatively synthetic datasets have been employed to mitigate this limitation by leveraging the atmospheric scattering model [48] (ASM) to generate synthetic hazy images using clean images. This generates paired training data for the underlying neural network to learn the restoration process. The ASM establishes a relationship between the hazy image (I) and the haze-free image (J) using parameters such as atmospheric light $(A \in [0,1])$, scene depth $(d(x))$, and atmospheric scattering coefficient $(\beta \in [0,1])$, as outlined in Eq. 1. However, algorithms trained on synthetic datasets tend to exhibit poor generalization on real hazy images due to domain gaps arising from diverse levels of haze density and distribution [60].

$$I(x) = J(x) * t(x) + A(1 - t(x)) \;\; where \;\; t(x) = e^{-\beta * d(x)} \tag{1}$$

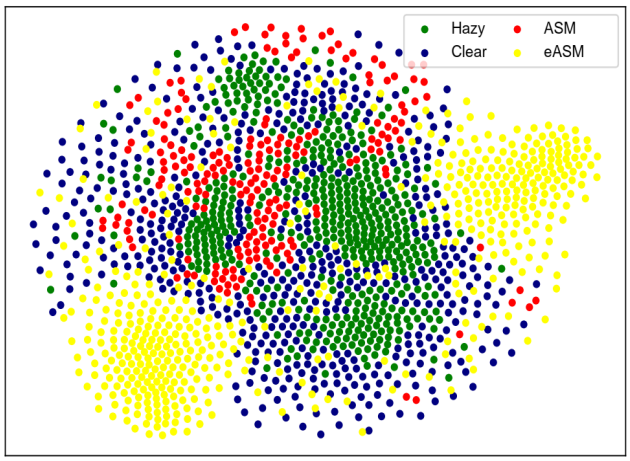

Figure 2. t-SNE plot of images captured in clear, hazy from NH-Dehaze [4] along with haze generated by atmospheric scattering model and proposed extension to qualitatively demonstrate non-overlapping and minimal overlap of feature distributions.

Specifically we attribute the performance gap to arise due to diverse chromatic differences in real-world datasets and the sub-optimal representation of the ASM in synthetic datasets. Addressing the former, real-world haze datasets can exhibit distinct tone curves from different cameras leading to a significant color distribution shift. As illustrated in the t-sne plot [63] shown in Fig. 2, these diverse distributions result in either non-or minimally overlapping feature spaces. This is counter-intuitive as one would expect a higher overlap in distribution, given the objective to model similar degradation characteristics. Hence, while combining multiple datasets improves domain-invariant performance compared to single-domain training, it may still yield sub-optimal results. For the latter part, i.e., synthetic dataset generation, we observe that prior approaches often adopt a simplified gray-scale uniform particle representation of atmospheric light (A) and scattering coefficient (β), overlooking the effects of different particle sizes on different wavelengths of light. We base this conclusion on observing a single-channel representation of atmospheric light with a fixed global value for the scattering coefficient. This is unrealistic in natural conditions, wherein particles of varying sizes can be non-homogeneously dispersed throughout natural scenes, introducing additional non-linearities in the haze generation process.

Apart from distribution gaps, the performance gap is also caused when using fixed synthetic datasets that only capture specific configurations of haze. This creates bias since complex training samples are not sufficiently represented and sampled. Consequently, the neural network's performance becomes bound by the limited diversity of haze samples during training. To address these issues, we propose an adversarial auto-augmentation pipeline that generates synthetic haze samples, varying their difficulty based on the restoration capability of the underlying network. This approach enriches the training data, enabling the network to handle a broader range of challenging scenarios and enhancing its generalization capacity. To ensure robustness to multiple tone curves from real-world datasets, we utilize a large data bank encompassing mean and variance information, facilitating the transfer of color profiles from multiple cameras without significant computational overhead. Extending the atmospheric scattering model (**e**xtended-ASM), we account for non-homogeneous particle distribution effects across wavelengths in the visible spectrum, integrating it into the auto-augmentation pipeline. This mechanism generates training images and provides an alternative view for computing contrastive loss, thereby improving perceptual fidelity. Finally, we propose depth and color attenuation-based loss optimizations to guarantee physical accuracy and color consistency in the restored images. By comprehensively addressing these challenges, our proposed approach advances image dehazing by bridging performance gaps, capturing diverse haze conditions, and improving perceptual quality and physical accuracy. We summarize our contributions as,

- We identify performance gaps caused by fixed synthetic datasets and real-world hazy datasets leading to sub-optimal optimization.
- We propose the integration of a data-bank for tone-transfer of real-world hazy images to ensure robustness towards camera sensors.
- We propose an adversarial auto-augmentation pipeline that generates synthetic haze samples of varying difficulty based on the restoration capability of the network by expanding the ASM.
- We utilize depth and color attenuation-based loss optimizations to guarantee physical accuracy and color consistency in the restored images.

2. Related Works

2.1. Single Image Dehazing

Early works focusing on image dehazing leveraged empirical statistics to construct prior-based dehazing mechanisms such as dark channel prior [26], non-local prior [6], color attenuation prior [78] and contrast maximization [62]. However, such simplistic approaches were found to be inadequate in representing complex haze models and subsequently restoring them. To alleviate the performance limitation, the atmospheric scattering model [48] was proposed, and different works [21, 70, 65, 44, 64] focused on estimating its components to restore a haze-affected image. However, such methods faced challenges due to error accumulation when estimating different properties. Recently learning-based approaches were developed [61, 15, 25, 42, 49, 56] that model image dehazing as an

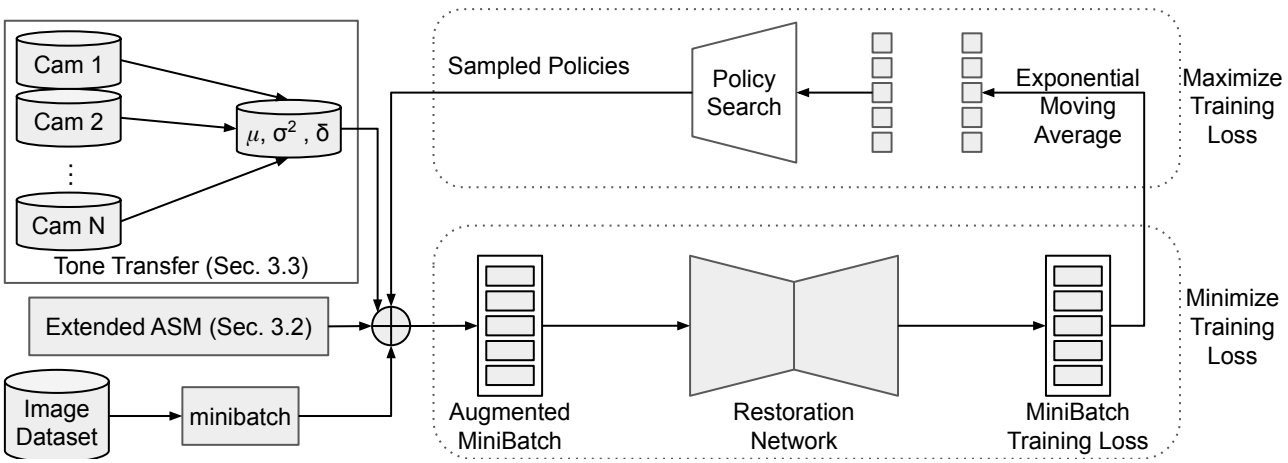

Figure 3. Overview of the proposed Adversarial Auto-Augmentation Approach for training robust single image dehazing algorithms.

end-to-end objective wherein an underlying neural network learned the representation space between a given pair of hazy and clean images. While these works focused on designing robust learning architectures, another line of work [30, 68] focused on improving the perceptual quality without making any architectural changes. Specifically, [68] demonstrated that it is possible to create compact dehazing algorithms while constraining the latent space to generate perceptually pleasant images using contrastive loss [11]. However, the reliance on a large amount of paired training data resulted in the network performing consistently when the distribution of the test set and training set matched. To overcome this limitation, an alternative mechanism leveraging unpaired-image translation [22], combining real and synthetic datasets [60], semi-supervised image translation [77] with additional focus on ensuring consistent performance on the synthetic dataset as well [71, 56].

Despite these efforts, we observe a significant performance gap when the distribution of the test set differs from the training set. Specifically, haze distribution and tone curves differ for synthetic and real datasets. Combining these datasets naively can result in multiple partially overlapping distributions, as observed in Fig. 2. Thus for an underlying restoration network to be domain invariant, the network should be able to generalize for all distributions, which is not guaranteed when using a fixed configuration synthetic dataset.

2.2. Data Efficient Training

In order to increase the diversity of training samples, data augmentation techniques are proposed that synthetically increase dataset size, given a fixed dataset. This approach has demonstrated effectiveness for both low [59, 8] and high-level vision tasks [74, 23]. These techniques can be categorized either as transformation-based (Flipping, Cropping), Region-based (Cut-Mix [72], Cut-Out [20], Copy-Blend [59]) and Color-based (Brightness, Contrast, Jitter). However, such simplistic approaches do not consider the model performance and are randomly applied. Thus these approaches cannot restrict training dataset bias to flow into the trained network resulting in sub-optimal performance. To overcome this limitation, composite augmentations [37, 69, 76, 18, 38] that can improve the performance and generalization of the network automatically were proposed. The key idea is to generate synthetic samples with varying difficulty levels via augmentation policies, taking into account the performance of the underlying network. These augmentation policies are generated by a separate policy network jointly trained with the target network in an adversarial framework [24]. The objective for the target network is to minimize the loss, while for the policy network is to design augmentation policies that maximize network loss. These approaches have shown promising results in high-level vision tasks such as classification [18, 46, 19], detection [13, 19] and segmentation [43] but are relatively unexplored for image restoration. We attribute this to the change in problem scope, wherein for high-level vision tasks, the objective is to ensure robust performance in changes in object appearance, lighting conditions, and scene complexity. However, for low-level tasks, the objective is to recover image details and improve perceptual quality. Hence for each restoration task, tailored augmentation policies are required. Nevertheless, given a physically accurate degradation model, the haze density and distribution can be adjusted using such an approach to ensure a robust and data-efficient image dehazing.

3. Methodology

3.1. Frequency based Feature Fusion

We highlight data-driven image dehazing algorithms to be constructed in a UNet [52] based manner with an encoder-decoder architecture and skip connections to facilitate edge information transfer. The encoder extracts fea-

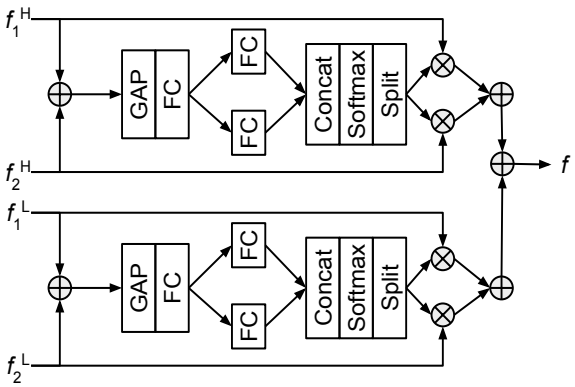

Figure 4. Overview of the proposed Frequency based Feature Fusion mechanism.

tures rich in semantic information, emphasizing high-level semantic details as the encoder progresses with increasing receptive field size. On the other hand, the decoder performs the upsampling operation to reconstruct the output image from the encoded features. The decoder also has access to fine-grained details and local information from the prior encoder block, which is available via skip connections. To fuse these complementary features, we introduce a gating mechanism that selectively adjusts the information concatenation between the skip connection and decoder block to reduce computational complexity. Distinct from previous works [61, 73, 36] that focused on enhancing feature representation within the fusion layer, we investigate the fusion operation from a frequency domain perspective. Our motivation stems from the observation that the frequency spectrum effectively demarcates low and high-frequency components, enabling selective sub-band information recovery. To realize this, we construct a low-pass filter using batch normalization, convolutional filters, and average pooling operations, as inspired by [**?**]. Dynamically adapting the cut-off frequency, we divide the input feature map into groups of four, facilitating the derivation of the corresponding high-pass filter by subtracting the low-pass filter from the identity kernel. To illustrate the efficacy of our proposed frequency-based feature fusion, we provide a visual illustration in Fig. 4.

3.2. extended Atmospheric Scattering Model

In the context of atmospheric scattering modeling, current approaches commonly assume a uniform scattering coefficient and a gray-world scenario, treating particles' impact on all visible spectrum wavelengths equally across the red, green, and blue channels within an image. However, this assumption overlooks the non-uniform nature of particle distribution and the varying impacts on different wavelengths in real-world settings. This discrepancy between the assumed and actual conditions creates limitations in synthetic datasets, contributing to the performance gap observed between real and synthetic datasets in image dehazing. To address this issue, we propose breaking these constraints by introducing a non-uniform scattering coefficient distribution. We sample the scattering coefficient from a Gaussian distribution, resulting in a scattering coefficient tensor of dimensions ($\beta \in R^{H\times W\times 1}$) for an image with height (H) and width (W). Furthermore, we expand the atmospheric light from a single coefficient to a multi-channel representation to simulate the implications of different particle sizes. Each channel is independently sampled from a Gaussian distribution, yielding an atmospheric light tensor of dimensions ($A \in R^{H\times W\times 3}$). To ensure a wide variety of degradation landscapes, we employ a randomized mean and variance process for both the scattering coefficient and atmospheric light sampling. This approach allows us to capture diverse degradation scenarios and generate qualitative samples showcasing the extended atmospheric scattering model. By introducing non-uniformity in the scattering coefficient and accounting for multiple particle sizes in the atmospheric light, our approach expands the scope of degradations covered by synthetic datasets with qualitative samples provided in Fig. 5.

3.3. Tone-Transfer

Images captured by different cameras under the same illumination conditions exhibit variations due to differences in imaging sensor properties and image signal processing pipelines. These disparities manifest as imperceptible variations in noise models and result in noticeable color discrepancies that create a distribution shift. Consequently, dehazing algorithms experience performance inconsistencies when applied to images captured by different imaging sensors. To address this challenge and enhance the robustness of the underlying restoration network to different tone curves associated with distinct cameras, we propose a tone transfer mechanism utilizing mean(μ)-variance(σ^2) transfer. Specifically, we utilize multiple datasets to compute the average mean and variance of well-illuminated images and their respective deviations (δ), which are stored in a data bank. We leverage this data bank to sample different combinations of mean and variance values, enabling the conversion of the distribution of a clean image to another camera's distribution following [51]. This tone transfer mechanism allows us to align the color characteristics of the input image with the target camera's profile, enhancing the fidelity of the dehazed output. We include qualitative samples in Fig. 6.

3.4. Adversarial-Auto Augmentation

Current data augmentation approaches for image restoration typically rely on randomized intensity selection, which may lead to sub-optimal training of restoration algorithms due to the absence of a feedback mechanism to control

Figure 5. Samples generated using proposed extended atmospheric scattering model using clean images from RESIDE [35] dataset.

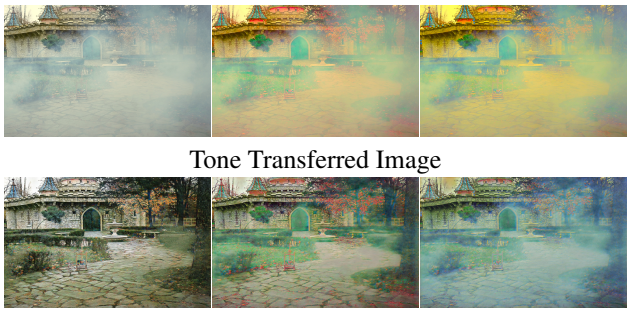

Figure 6. Samples generated using tone-transfer mechanism for a given clean image from NH-Haze [4] dataset in Fig. 7 and corresponding dehazing results from pretrained AECRNet [68].

degradation strength. To address this limitation, we propose a novel approach that enables dynamic degradation intensity adjustment based on the restoration network's performance. While such techniques have been extensively studied for high-level perception tasks, their applicability to low-level vision tasks still needs to be explored. In line with our proposed approach, we build upon [76] and replace the augmentation functions with proposed extended-ASM, tone-transfer mechanisms and introduce localized degradation using the copy-blend [59] mechanism. This reduced search space of image operations includes rotation, flipping, copy-blend, tone-transfer, and synthetic haze generation. We use a 10-part uniformly distributed magnitude of operation for each augmentation. Hence the search space of policy for each epoch is $(6 \times 10)^10 \approx 6.04 \times 10^{17}$ possibilities.

The critical advantage of auto-adversarial augmentation lies in its ability to generate diverse and realistic training samples, thereby enhancing the robustness and generalization of the underlying model. By exposing the model to a wide range of challenging samples, it can effectively learn to handle various variations and improve its performance on unseen data. In constructing the policy search network, we follow the adversarial auto-augmentation approach [76], leveraging its effectiveness in generating augmented samples and optimizing the training process for image restoration.

3.5. Physical and Perceptual Accuracy

Our approach incorporates depth and color consistency losses to ensure physically accurate and visually pleasing restored images. We leverage the zero-shot depth estimation network MiDAS [50] (ϕ) to generate depth maps for clear images (I_C) in both synthetic and real datasets. Furthermore, the restored images (I_R) are passed through the depth estimation network to obtain corresponding depth. Subsequently, depth loss is computed using mean squared error (Eq. 2) between restored and clear images. This ensures that the restored images are physically accurate compared to their clean counterpart.

$$L_D = ||\phi(I_R) - \phi(I_C)|| \tag{2}$$

We employ contrastive loss and color consistency loss to enhance perceptual accuracy. In the case of contrastive loss, positive and negative samples directly influence wherein image fidelity improves in the presence of mini-batch hard positives and negatives. Herein, images generated by the adversarial auto-augmentation can serve as hard negative synthetic samples apart from being used for training the restoration network without complex sampling mechanisms. Furthermore, we employ a color attenuation loss (L_C) (Eq. 3) that considers saturation (S) and luminance (L) differences between the restored and clean images. This loss ensures that the colors in the restored images are consistent and visually appealing. We extract luminance and saturation values by converting the clean and restored image into LAB color space and extract *L* and *A* channels respectively. By leveraging saturation and luminance as measures of color differences, we encourage the network to generate images that maintain coherent color distributions while preserving the overall brightness and vividness.

$$L_C = ||S(I_R) - S(I_C)|| + ||L(I_R) - L(I_C)|| \tag{3}$$

4. Experimental Evaluation

4.1. Datasets and Evaluation Metrics

For our experiments, we create a training set constructed from real-world paired dehazing datasets such as IHaze [2], OHaze [3], NH-Haze [4], NH-Haze2[5] and Dense-haze [1] datasets resulting in a total of 120 image pairs. We further create the tone-transfer data bank from well-illuminated images selected manually from outdoor datasets such as SID [9], ELD [67], GOPRO [47], Cityscapes [17], ACDC [55], IS2R [58] resulting in a total of 30 camera tone-maps. Furthermore, we also utilize clear daylight images from ACDC, GOPRO, and Cityscapes datasets for synthetic datasets, as these contain diverse object densities. For quantitative evaluation, we utilize pixel and feature-based metrics such as PSNR, [66], LPIPS [75], and NIQE [45]. Furthermore, since we generate relative depth using MiDAS, we utilize root mean squared error (RSME) to compute the accuracy of depth estimation on the restored image, following [50]. For our baseline network, we choose AECR [68] and DIDH [60] due to their high performance to compute ratio.

4.2. Implementation Details

We use a single layer LSTM [27] based RNN controller to construct the policy network. The hidden size of the LSTM layer is set to 100, with the projection size set to 32. The restoration network is used as the target network, and the loss function is composed of L1, contrastive, and proposed Depth and Color Attenuation Loss following,

$$L = L1 + 0.1 * L_{Contra} + 1.0 * L_D + 1.0 * L_C \quad (4)$$

We train the proposed pipeline with one RTX 4090 GPU using 256 x 256 crop size, ADAM [32] optimizer with β_1 and β_2 set to 0.9 and 0.99. Furthermore we set the controller learning rate to $3.5e^{-4}$ and target learning rate to $2e^{-4}$. Finally, an entropy penalty of $1e^{-5}$ is applied to the controller weights to avoid unexpected rapid convergence.

4.3. Comparison with SoTA

We summarize the quantitative performance of the SoTA algorithms such as DuRN-US [41], GridDehazenet [40], FFA-Net [49], TridentNet [39], DA-Dehaze [56], DIDH [60], AECR-Net [68], DeHamer [16], D4 [71], FogRemoval [29], DEANet [14], DehazeFormer-B [61] on NH-Haze and Dense-Haze in Tab. 2 with qualitative results in Fig. 7. Furthermore, we include qualitative results on foggy images from the ACDC dataset in Fig. 9. We observe that the proposed mechanism of including depth as auxiliary optimization loss results in more physically accurate restoration, compared to prior works D4 and DeHamer that consider depth information during data augmentation or position embedding, respectively. The qualitative and quantitative results show that the proposed training mechanism improves performance. Specifically, we observe a performance improvement across all datasets for both AECRNet [68] and DIDH [60]. We also observe current SoTA algorithms to result in reduced depth estimation performance, highlighting these algorithms to be inaccurate in their ability to generate physically accurate restoration. In comparison, when included in training, AECRNet [68] and DIDH [60], the proposed modifications result in higher depth accuracy in restored images.

Method	PSNR / SSIM	NIQE / LPIPS	RMSE
DuRN-US [41]	13.63 / 0.57	3.51 / 0.64	3.54
GridDehazenet [40]	12.96 / 0.50	4.16 / 0.84	4.15
FFA-Net [49]	14.01 / 0.56	3.71 / 0.87	3.47
TridentNet [39]	16.48 / 0.54	5.36 / 1.57	3.05
DA-Dehaze [56]	13.98 / 0.37	4.01 / 0.77	3.91
DIDH [60]	19.47 / 0.75	2.58 / 0.65	3.84
AECR-Net [68]	15.80 / 0.46	2.94 / 1.09	3.78
DeHamer [16]	16.62 / 0.56	3.55 / 0.94	3.99
D4 [71]	13.12 / 0.53	2.87 / 1.07	3.56
FogRemoval [29]	16.67 / 0.50	3.57 / 1.02	3.64
DEANet [14]	12.01 / 0.32	4.69 / 1.38	4.11
DehazeFormer-B [61]	11.68 / 0.32	3.58 / 1.14	4.96
Ours (DIDH)	19.93 / 0.71	2.55 / 0.63	2.45
Ours (AECRNet)	17.10 / 0.57	2.34 / 0.58	2.32

Table 1. Quantitative Evaluation of SoTA Image Dehazing algorithms on Dense-haze datasets

Method	PSNR / SSIM	NIQE / LPIPS	RMSE
DuRN-US [41]	15.27 / 0.50	4.21 / 1.42	5.89
GridDehazenet [40]	15.32 / 0.60	3.06 / 0.65	4.88
FFA-Net [49]	18.11 / 0.66	2.94 / 0.67	4.16
TridentNet [39]	21.41 / 0.71	3.66 / 1.20	2.69
DA-Dehaze [56]	11.42 / 0.31	3.91 / 1.13	4.55
DIDH [60]	21.17 / 0.78	2.98 / 0.56	3.64
AECR-Net [68]	20.68 / 0.82	3.48 / 0.84	3.72
DeHamer [16]	19.18 / 0.79	3.20 / 0.92	3.47
D4 [71]	12.65 / 0.37	4.92 / 0.90	5.17
FogRemoval [29]	20.99 / 0.61	3.45 / 0.65	4.58
DEANet [14]	10.98 / 0.25	2.99 / 1.40	4.42
DehazeFormer-B [61]	12.84 / 0.35	3.22 / 0.46	4.19
Ours (DIDH)	21.44 / 0.79	2.57 / 0.49	2.07
Ours (AECRNet)	21.70 / 0.68	3.08 / 0.34	1.98

Table 2. Quantitative Evaluation of SoTA Image Dehazing algorithms on Dense-haze and NH-Hazedatasets. We measure PSNR, SSIM, LPIPS using [31] and NIQE using [10].

4.4. Ablation Studies

For our ablation, we consider AECRNet [68] as the reference network evaluated on the Dense-haze dataset when trained using RESIDE [35] dataset. This setting allows us to evaluate the generalization between synthetic hazy samples vis-a-vis the proposed extended atmospheric scattering model. Finally, we summarize performance results for different scenarios in Tab. 3.

Based on empirical results, the proposed Frequency-based Feature Fusion (FFF) demonstrates a significant im-

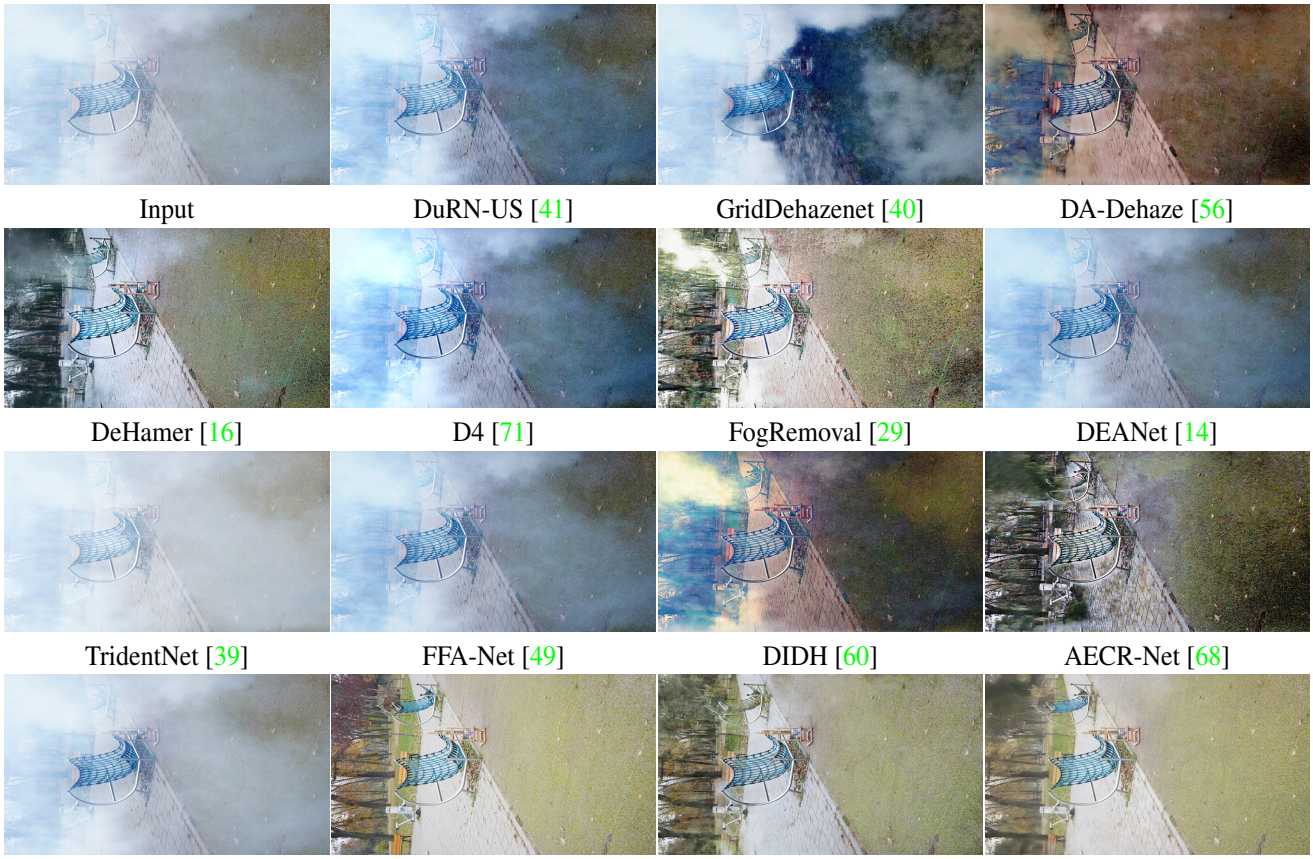

Figure 7. Impact of weather variation (fog) on the performance of SoTA segmentation network. In comparison, we demonstrate performance improvement powered by latent representation alignment and data augmentation, as this work proposes.

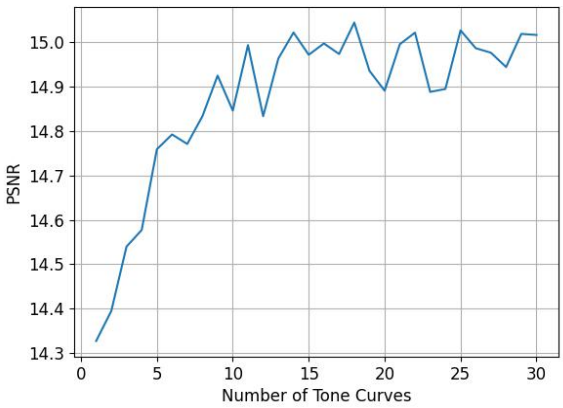

Figure 8. Overview of utilizing different tone-curves when trained with configuration (i) presented in Tab. 3.

provement of +0.51 dB in both pixel and perceptual performance of the restoration network (cf. (b)). To further investigate the implications of using the enhanced Atmospheric Scattering Model (ASM) compared to the traditional ASM, we evaluate the performance using clear images from the RESIDE dataset and generate synthetic samples on the fly without further modifications (cf. Figures (c, d)). The results show a substantial increase in performance, further boosted when the proposed Frequency-based Feature Fusion is utilized. This confirms the efficacy of the eASM in modeling haze, surpassing prior approaches (cf. (d)).

To enable a feedback mechanism that adjusts the difficulty of training samples based on restoration performance, we train the complete framework using Automatic Adversarial Augmentation (AAA). This leads to a performance improvement of +0.35 dB (cf. (e)). We expand the scope of utilization of adversarial images to contrastive loss during training results in an additional boost of +0.38 dB (cf. (f)). While the restoration quality improves, depth estimation accuracy remains limited. This highlights the trade-off between perceptually pleasant images and physical accuracy. To address this, we sequentially integrate depth and color attenuation losses to improve the physical accuracy of the restored images while ensuring color consistency.

Expanding our evaluation, we explore the influence of tone transfer and identify the optimal number of tone curves for achieving peak performance (Fig. 8). By linearly in-

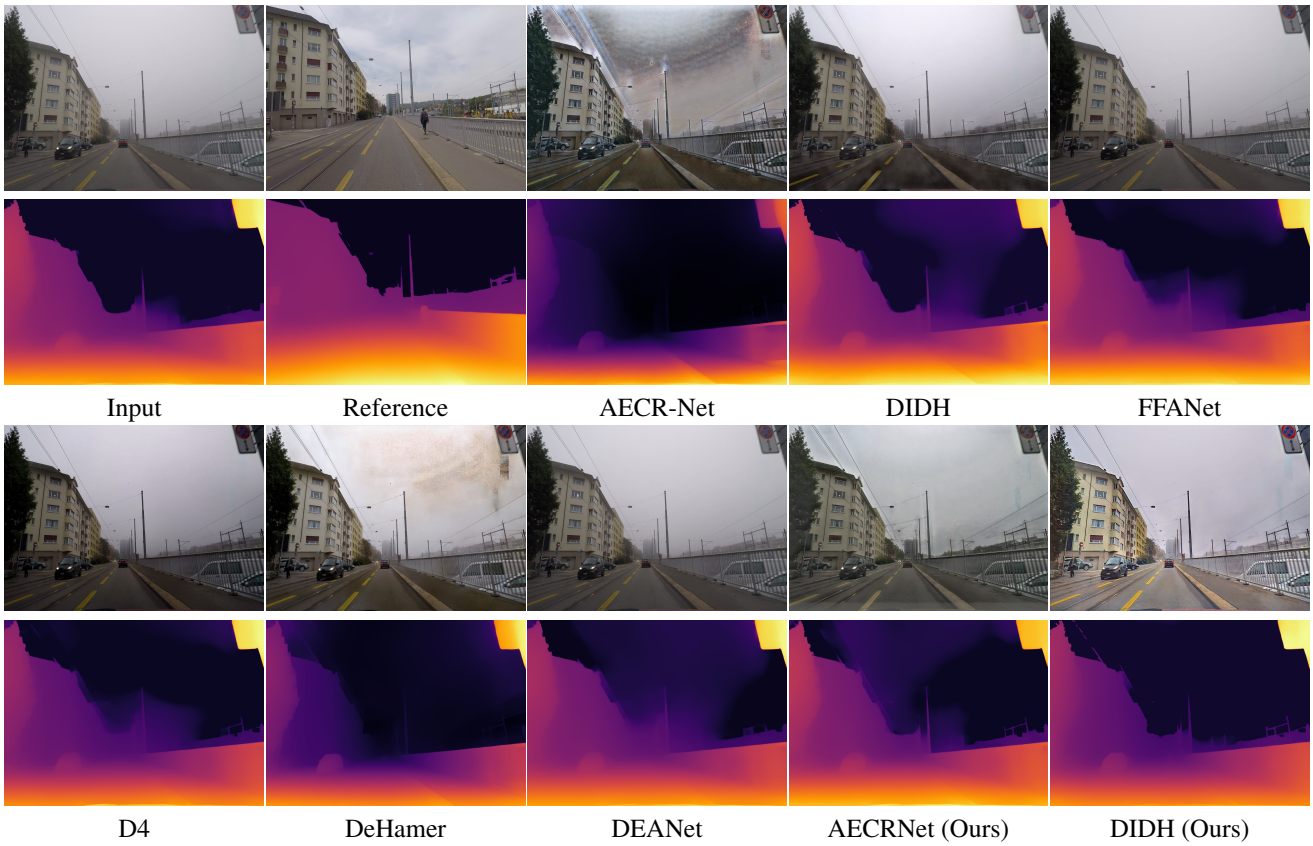

Figure 9. Impact of weather variation (fog) on the performance of SoTA segmentation network. In comparison, we demonstrate performance improvement powered by latent representation alignment and data augmentation, as this work proposes.

creasing the number of tone curves for tone transfer, we observe consistent performance improvements up to 12 tone curves, with minor improvements as the number increases. However, we caution that the optimal number of tone-transfer mechanisms should not be considered a hard threshold, as the learning capacity of the underlying restoration algorithm may also influence it. Identifying the optimal number of tone curves for a given restoration mechanism remains an open area for future research.

Ref.	FFF	ASM	AAA	Loss	PSNR / SSIM	NIQE / LPIPS	RMSE
(a)		ASM			8.59 / 0.11	5.83 / 1.84	12.58
(b)	✓				9.10 / 0.17	5.83 / 1.80	12.51
(c)		eASM			10.29 / 0.26	4.29 / 1.19	8.42
(d)	✓	eASM			10.55 / 0.27	4.25 / 0.98	6.84
(e)	✓	eASM	✓		10.90 / 0.32	3.98 / 0.87	6.01
(f)	✓	eASM	✓	L_{Contra}	11.28 / 0.40	3.59 / 0.82	5.99
(g)	✓	eASM	✓	L_D	11.03 / 0.38	3.99 / 0.85	2.91
(h)	✓	eASM	✓	L_C	11.89 / 0.42	3.29 / 0.81	5.49
(i)	✓	eASM	✓	L_{Contra}, L_D, L_C	14.89 / 0.52	3.04 / 0.84	2.45

Table 3. Quantitative Evaluation of SoTA Image Dehazing algorithms on Dense-haze datasets.

5. Conclusion

In this work, we propose insights to improve image dehazing techniques. Our contributions include a frequency-based feature fusion mechanism that combines low and high-frequency details, preserving semantic and edge information. Additionally, we introduce an extended atmospheric scattering model that accurately represents diverse haze degradations by considering non-homogeneous particle distribution and its impact on different wavelengths. To enhance generalization, we incorporate a tone-transfer mechanism capturing various camera properties. Integration into an automatic adversarial augmentation pipeline enables dynamic adjustment of degradation intensity based on network performance. We ensure physically accurate and visually pleasing results through depth and color attenuation losses. Synthetic adversarial images are leveraged within the contrastive loss framework to improve restoration quality. Extensive experiments validate our proposed mechanisms, yielding significant improvements in image dehazing. Our approach bridges the gap between physical accuracy and perceptual fidelity, contributing to advancements in the field.

References

[1] Codruta O Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *2019 IEEE international conference on image processing (ICIP)*, pages 1014–1018. IEEE, 2019. 6

[2] Codruta O. Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. I-haze: a dehazing benchmark with real hazy and haze-free indoor images. In *arXiv:1804.05091v1*, 2018. 6

[3] Codruta O. Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *IEEE Conference on Computer Vision and Pattern Recognition, NTIRE Workshop*, NTIRE CVPR'18, 2018. 6

[4] Codruta O Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, and Radu Timofte. Ntire 2020 challenge on nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 490–491, 2020. 2, 5, 6

[5] Codruta O Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, and Radu Timofte. Ntire 2021 nonhomogeneous dehazing challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 627–646, 2021. 1, 6

[6] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1674–1682, 2016. 2

[7] Laurent Caraffa and Jean-Philippe Tarel. Stereo reconstruction and contrast restoration in daytime fog. In *Computer Vision–ACCV 2012: 11th Asian Conference on Computer Vision, Daejeon, Korea, November 5-9, 2012, Revised Selected Papers, Part IV 11*, pages 13–25. Springer, 2013. 1

[8] Chia-Ming Chang and Tsung-Nan Lin. Damix: A density-aware mixup augmentation for single image dehazing under domain shift. *arXiv preprint arXiv:2109.12544*, 2021. 3

[9] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018. 6

[10] Chaofeng Chen and Jiadi Mo. [Online]. Available: https://github.com/chaofengc/IQA-PyTorch, 2022. 6

[11] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020. 3

[12] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster r-cnn for object detection in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3339–3348, 2018. 1

[13] Yukang Chen, Yanwei Li, Tao Kong, Lu Qi, Ruihang Chu, Lei Li, and Jiaya Jia. Scale-aware automatic augmentation for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9563–9572, 2021. 3

[14] Zixuan Chen, Zewei He, and Zhe-Ming Lu. Dea-net: Single image dehazing based on detail-enhanced convolution and content-guided attention. *arXiv preprint arXiv:2301.04805*, 2023. 6, 7

[15] Zeyuan Chen, Yangchao Wang, Yang Yang, and Dong Liu. Psd: Principled synthetic-to-real dehazing guided by physical priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7180–7189, 2021. 2

[16] Saeed Anwar Runmin Cong Wenqi Ren Chongyi Li Chun-Le Guo, Qixin Yan. Image dehazing transformer with transmission-aware 3d position embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 6, 7

[17] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 6

[18] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 113–123, 2019. 3

[19] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 702–703, 2020. 3

[20] Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*, 2017. 3

[21] Jiangxin Dong and Jinshan Pan. Physics-based feature dehazing networks. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXX 16*, pages 188–204. Springer, 2020. 2

[22] Deniz Engin, Anil Genç, and Hazim Kemal Ekenel. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 825–833, 2018. 3

[23] Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D Cubuk, Quoc V Le, and Barret Zoph. Simple copy-paste is a strong data augmentation method for instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2918–2928, 2021. 3

[24] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 3

[25] Chun-Le Guo, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, and Chongyi Li. Image dehazing transformer with transmission-aware 3d position embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5812–5820, 2022. 1, 2

[26] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010. 2

[27] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997. 6

[28] Shih-Chia Huang, Trung-Hieu Le, and Da-Wei Jaw. Dsnet: Joint semantic learning for object detection in inclement weather conditions. *IEEE transactions on pattern analysis and machine intelligence*, 43(8):2623–2633, 2020. 1

[29] Yeying Jin, Wending Yan, Wenhan Yang, and Robby T Tan. Structure representation network and uncertainty feedback learning for dense non-uniform fog removal. In *Computer Vision–ACCV 2022: 16th Asian Conference on Computer Vision, Macao, China, December 4–8, 2022, Proceedings, Part III*, pages 155–172. Springer, 2023. 6, 7

[30] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 694–711. Springer, 2016. 3

[31] Sergey Kastryulin, Jamil Zakirov, Denis Prokopenko, and Dmitry V. Dylov. Pytorch image quality: Metrics for image quality assessment, 2022. 6

[32] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[33] Sohyun Lee, Taeyoung Son, and Suha Kwak. Fifo: Learning fog-invariant features for foggy scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1

[34] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. End-to-end united video dehazing and detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. 1

[35] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019. 5, 6

[36] Xiang Li, Wenhai Wang, Xiaolin Hu, and Jian Yang. Selective kernel networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 510–519, 2019. 4

[37] Sungbin Lim, Ildoo Kim, Taesup Kim, Chiheon Kim, and Sungwoong Kim. Fast autoaugment. *Advances in Neural Information Processing Systems*, 32, 2019. 3

[38] Chen Lin, Minghao Guo, Chuming Li, Xin Yuan, Wei Wu, Junjie Yan, Dahua Lin, and Wanli Ouyang. Online hyper-parameter learning for auto-augmentation strategy. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6579–6588, 2019. 3

[39] Jing Liu, Haiyan Wu, Yuan Xie, Yanyun Qu, and Lizhuang Ma. Trident dehazing network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 430–431, 2020. 6, 7

[40] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *ICCV*, pages 7314–7323, 2019. 6, 7

[41] Xing Liu, Masanori Suganuma, Zhun Sun, and Takayuki Okatani. Dual residual networks leveraging the potential of paired operations for image restoration. In *Proc. Conference on Computer Vision and Pattern Recognition*, pages 7007–7016, 2019. 6, 7

[42] Ye Liu, Lei Zhu, Shunda Pei, Huazhu Fu, Jing Qin, Qing Zhang, Liang Wan, and Wei Feng. From synthetic to real: Image dehazing collaborating with unlabeled real data. In *Proceedings of the 29th ACM international conference on multimedia*, pages 50–58, 2021. 1, 2

[43] Junyan Lyu, Yiqi Zhang, Yijin Huang, Li Lin, Pujin Cheng, and Xiaoying Tang. Aadg: automatic augmentation for domain generalization on retinal image segmentation. *IEEE Transactions on Medical Imaging*, 41(12):3699–3711, 2022. 3

[44] Gaofeng Meng, Ying Wang, Jiangyong Duan, Shiming Xiang, and Chunhong Pan. Efficient image dehazing with boundary constraint and contextual regularization. In *Proceedings of the IEEE international conference on computer vision*, pages 617–624, 2013. 2

[45] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 6

[46] Samuel G Müller and Frank Hutter. Trivialaugment: Tuning-free yet state-of-the-art data augmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 774–782, 2021. 3

[47] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3883–3891, 2017. 6

[48] Srinivasa G Narasimhan and Shree K Nayar. Vision and the atmosphere. *International journal of computer vision*, 48(3):233, 2002. 1, 2

[49] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 11908–11915, 2020. 2, 6, 7

[50] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. *ICCV*, 2021. 5, 6

[51] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *IEEE Computer graphics and applications*, 21(5):34–41, 2001. 4

[52] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 3

[53] Christos Sakaridis, Dengxin Dai, Simon Hecker, and Luc Van Gool. Model adaptation with synthetic and real data for semantic dense foggy scene understanding. In *Proceedings of the european conference on computer vision (ECCV)*, pages 687–704, 2018. 1

[54] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126:973–992, 2018. 1

[55] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10765–10775, 2021. 1, 6

[56] Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain adaptation for image dehazing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2808–2817, 2020. 1, 2, 3, 6, 7

[57] Pranjay Shyam, Kyung-Soo Kim, and Kuk-Jin Yoon. Giqe: Generic image quality enhancement via nth order iterative degradation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2077–2087, 2022. 1

[58] Pranjay Shyam, Sumit Mishra, Kuk-Jin Yoon, and Kyung-Soo Kim. Infra sim-to-real: An efficient baseline and dataset for infrastructure based online object detection and tracking using domain adaptation. In *2022 IEEE Intelligent Vehicles Symposium (IV)*, pages 1393–1399. IEEE, 2022. 6

[59] Pranjay Shyam, Sandeep Singh Sengar, Kuk-Jin Yoon, and Kyung-Soo Kim. Evaluating copy-blend augmentation for low level vision tasks. *arXiv preprint arXiv:2103.05889*, 2021. 3, 5

[60] Pranjay Shyam, Kuk-Jin Yoon, and Kyung-Soo Kim. Towards domain invariant single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 9657–9665, 2021. 1, 3, 6, 7

[61] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023. 1, 2, 4, 6, 7

[62] Robby T Tan. Visibility in bad weather from a single image. In *2008 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2008. 2

[63] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008. 2

[64] Jinbao Wang, Ke Lu, Jian Xue, Ning He, and Ling Shao. Single image dehazing based on the physical model and msrcr algorithm. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(9):2190–2199, 2017. 2

[65] Jin-Bao Wang, Ning He, Lu-Lu Zhang, and Ke Lu. Single image dehazing with a physical model and dark channel prior. *Neurocomputing*, 149:718–728, 2015. 2

[66] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6

[67] Kaixuan Wei, Ying Fu, Jiaolong Yang, and Hua Huang. A physics-based noise formation model for extreme low-light raw denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2758–2767, 2020. 6

[68] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, 2021. 1, 3, 5, 6, 7

[69] Dong Yang, Holger Roth, Ziyue Xu, Fausto Milletari, Ling Zhang, and Daguang Xu. Searching learning strategy with reinforcement learning for 3d medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 3–11. Springer, 2019. 3

[70] Xitong Yang, Zheng Xu, and Jiebo Luo. Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018. 2

[71] Yang Yang, Chaoyue Wang, Risheng Liu, Lin Zhang, Xiaojie Guo, and Dacheng Tao. Self-augmented unpaired image dehazing via density and depth decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2037–2046, 2022. 3, 6, 7

[72] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6023–6032, 2019. 3

[73] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 492–511. Springer, 2020. 4

[74] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017. 3

[75] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 6

[76] Xinyu Zhang, Qiang Wang, Jian Zhang, and Zhao Zhong. Adversarial autoaugment. *arXiv preprint arXiv:1912.11188*, 2019. 3, 5

[77] Hongyuan Zhu, Xi Peng, Vijay Chandrasekhar, Liyuan Li, and Joo-Hwee Lim. Dehazegan: When image dehazing meets differential programming. In *IJCAI*, pages 1234–1240, 2018. 3

[78] Qingsong Zhu, Jiaming Mai, and Ling Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE transactions on image processing*, 24(11):3522–3533, 2015. 2