Overcoming Degradation Imbalance for Consistent Image Dehazing

Pranjay Shyam
pranjay.shyam.psm@forvia.com
HyunJin Yoo
hyunjin.yoo@forvia.com
Faurecia IRYStec Inc.
Montreal, Canada

Abstract

MSE or MAE loss functions work under the premise of all pixels having equal contribution during optimization. However, natural haze degradations are non-homogeneous, resulting in varied haze distribution. This results in sub-optimal training of current state-of-the-art (SoTA) supervised learning-based image dehazing algorithms due to an imbalance in pixel contribution. The outcome is the poor recovery of areas affected by severe degradations as these are underrepresented vis-a-vis mild and moderate affected regions during training. To address this data imbalance and generate consistent, visually pleasing restored images, we identify strategies at data augmentation and loss computation stages to ensure degradation-balanced training for image dehazing. From a data augmentation perspective, we propose a peak-signal-to-noise ratio (PSNR) based patch sampling mechanism and an adversarial auto-augmentation mechanism to vary the degradation distribution intensity within training samples. Second, to reduce the bias introduced by increasing the proportion of accurately recovered pixels along the training cycle, we propose focal pixel loss that scales the contribution of individual pixels towards loss calculation with restoration accuracy as prior. We successfully integrate the proposed focal loss in current pixel and feature-based loss functions. Finally, to ensure perceptually pleasant and structurally accurate image restoration, we propose a dynamic version of contrastive regularization with dynamic boundary constraints to better constrain the latent representations. We demonstrate the proposed mechanisms' efficacy in improving the performance of SoTA image dehazing algorithms without modifying the underlying network architecture.

1 Introduction

Image dehazing is a widely researched area focusing on improving the perceptual quality of an image by recovering the affected regions. This application has found increasing utilization for its benefits towards both human and machine vision where the performance of downstream perception tasks such as object detection [9, 27, 34, 58], semantic segmentation [33, 50, 51, 56, 57], depth estimation [7] are shown to be improved. Current SoTA image dehazing algorithms can be categorized based on their working into data-driven [10, 22, 41, 47, 53, 54, 60, 68] or physics-based [17, 43, 64, 65, 70**?**]. In comparison, data-driven algorithms leverage a large corpus of paired training samples composed of hazy

Figure 1: We demonstrate (b) performance inconsistency and (c) efficient restoration by visual comparison on non-homogeneous hazy images from NH-Haze dataset [4] using (b) pretrained DIDH [54] and (c) retrained using the proposed approach. We include PSNR / SSIM metric results for quantitative evaluation.

and clear images and train an underlying algorithm to model the mapping between them. Alternatively, physics-based algorithms leverage the atmospheric scattering model [45] to estimate different constituents used to generate a hazy image. Specifically, a hazy image (I) can be represented as a combination of haze-free image (J), atmospheric light (A), and the medium transmission map (t) related by.

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

here transmission map represents the relationship between scene depth $(d(x))$ and atmospheric scattering coefficient (β). However, algorithms predicting different components of atmospheric scattering cannot account for non-homogeneous distribution, resulting in suboptimal restoration performance.

One issue plaguing the performance of SoTA dehazing algorithms is the inconsistent restoration of non-homogeneous hazy scenes, as demonstrated in Fig. 1. We conjecture this from biased training wherein the loss function is skewed by a higher quantum of accurately restored pixels out-weighing the inaccurately restored ones. This results in poor restoration of areas affected by extreme haze variations since these images are scarce in existing small-scale real-world datasets. Furthermore, this issue is exacerbated by the current training mechanisms wherein a random patch is sampled from the paired dataset and used for training. Since natural haze degradation is non-homogeneous, randomly generating training patches results in suboptimal training due to increased sampling of easy and moderately affected samples.

Although solutions exist to such sampling bias, encompassed in class-imbalanced learning literature, the focus is on improving long-tail performance for high-level perception tasks such as classification, segmentation, and detection. However, equivalent solutions for image restoration are nonexistent, which we primarily attribute to a need for more awareness of such issues. While prior works [55, 72] focused on developing region augmentation-based strategies to generate non-homogeneous training samples, these approaches have yet to be widely adopted due to region inconsistencies being present in the training patches. Specifically, prior works proposed a combination of CutMix [73] and MixUp [75] to generate varying intensities and localization of degradation within a training patch. Since such approaches generate training patches with varying degradation intensities without consideration of network performance towards degradation intensities, they need to address the sampling bias

effectively. Alternatively, another direction to avoid sampling bias is patch sampling wherein [61] proposed to utilize a CNN to determine the information quality within a patch for demosaicing. However, such approaches require training an additional CNN, which prolongs the training period by increasing the total number of optimization parameters.

With an emphasis on improving the perceptual quality of a haze-affected image where the recovered image is structurally consistent with high-fidelity chromatic features, we propose mechanisms at data augmentation and loss computation stages, keeping the underlying model architecture fixed. With this setup, we propose a PSNR-based adaptive patch sampling approach wherein given an initial minibatch (size > 2) triplet of clear (I_C), degraded (I_D), and restored (I_R) patches, we compute the PSNR of degraded and restored patches. Subsequently, we vary the frequency of training samples for the next mini-batches utilizing the restoration performance on the prior minibatch. This approach allows us to discard less informative samples without processing them through the restoration network, thereby avoiding the computation of gradients and allowing us to address the issue of degradation imbalance. However, a primary requirement for such an approach to be successful is the presence of a wide range of samples exhibiting diverse degradations within the training dataset, which is not available for image dehazing. Hence to ensure the existence of such non-homogeneous training samples, we integrate an adversarial auto-augmentation technique that learns a policy function to generate synthetic hard degradations that minimize network performance. Thus, combining synthetic degradation generation with PSNR sampling ensures the redressal of degradation imbalance within training data.

In addition, we highlight commonly used pixel-based loss functions, i.e., Mean-Square-Error (MSE) or Mean-Absolute-Error (MAE), to compute the mean for numerically representing the error of restored images. However, as extreme pixel inaccuracies (representative of poor restoration performance) are lower in proportion vis-a-vis accurately or moderately recovered pixels, naively computing mean over the complete image suppresses these extremities. Over the course of the complete training cycle, the quantum of accurately or moderately recovered pixels increases, thereby skewing the contribution of poorly affected restoration. This restricts accurate restoration in severely affected regions. To overcome this, we propose an adaptive weighting technique to reweigh pixel contribution during loss computation and remove bias caused by an uneven distribution of mild and moderate affected pixels over low-volume, extremely degraded pixels. As both pixel and feature-based losses have this vulnerability, this modification is a drop in replacement for current loss functions.

To ensure improved textural fidelity and structural consistency within restored images, perceptual loss [29] was proposed that minimized the L1 distance between multi-scale features of clear and restored images extracted by an ImageNet [16] pretrained VGG [59] network. Improving upon the design, [68] introduced contrastive regularization that utilized the feature triplet of clean, degraded, and restored images to constrain the latent space efficiently. Despite these advances, we highlight such approaches to result in fixed boundary constraints; thereby, these approaches cannot adjust the boundary constraints based on the changing training dynamics, restricting the perceptual quality of the restored image. Thus to ensure the incorporation of training dynamics within contrastive regulation, we propose a continual update mechanism of negative boundary constraint. Specifically, we generate pseudo-negative samples utilizing image blending operation to obtain images with varying degradation ranges. Such an approach generates hard-negative samples whose latent representation is close to a positive sample. We observe this to result in closer feature alignment between restored and clean images along the training cycle, thus improving the image quality of the restored image. In order to ensure effective utilization of hard-negative, wherein the

generated synthetic sample does not adversely affect the training, we propose a PSNR-based filtering mechanism wherein only negative samples close to restored PSNR are considered hard negatives during loss computation.

In summary, our contributions can be listed as

- We propose a PSNR-based adaptive patch sampling to ensure the utilization of informative training samples without computing their gradients.
- To overcome the lack of samples exhibiting diverse non-homogeneous degradations, we propose an adversarial auto-augmentation approach.
- To ensure the presence of pixel extremities within loss computation, we propose a focal loss function that reweighs pixel contribution based on degradation prior.
- To ensure compact feature space clustering and improved perceptual quality of restored images, we propose dynamic contrastive loss where negative boundary constraints are updated gradually as the training progresses.

2 Related Works

2.1 Class Imbalanced Learning

Common real-world datasets exhibit class imbalance wherein the frequency of some classes is higher than others. This skews the training mechanism, as common classes would be sampled much higher than rare classes. Thereby resulting in higher performance for common classes vis-a-vis rare classes. Current solutions focus on high-level perception tasks such as image recognition [24, 42, 77], object detection [46], semantic segmentation [25] and instance segmentation [19, 53, 66]. These solutions can be categorized either into (1) data sampling mechanisms [49, 74] that down-sampling standard classes or over-sampling rare classes or (2) classifier balancing using cost-sensitive loss mechanism [15, 37, 52, 79], gradient re-balancing [26, 62] or margin adjustment [6, 31]. However, approaches devised for perception algorithms cannot be directly used for image restoration due to incompatible problem space. Unlike perception tasks that generate high-level attributes such as bounding boxes, object classification, and pixel/instance segmentation, image restoration is tasked to recover degraded pixels accurately.

2.2 Consistent Image Restoration

Multiple approaches have been proposed to ensure consistent performance of different image restoration on images with different degradation intensities, focusing on data augmentation, patch sampling, and adversarial augmentation. Data augmentation techniques [55, 65, 72] proposed to generate training samples with non-uniform degradations to ensure the underlying image restoration algorithms can localize and restore the affected regions. Alternatively, patch sampling techniques [61] were proposed that identified useful training patches within an image via a learning-based approach, thereby removing the samples with low information content. In addition, learnable augmentation techniques [14, 36, 69, 78] have also been proposed, albeit for high-level perception tasks that generate a set of data augmentation policies to maximize the error of the underlying perception algorithms. This results in a min-max optimization cycle, ensuring the trained perception algorithms are robust towards different domains, scenes, and degradations. Despite their promise of ensuring robust performance, such approaches have yet to be studied for image restoration algorithms that could benefit from these characteristics. Hence in this paper, we explore the effect of utilizing learnable adversarial augmentation approaches toward improving the performance of image restoration algorithms. To ensure the compatibility of problem space, we modify the search space to ensure compatibility with image restoration tasks.

2.3 Image Dehazing

Traditional single image dehazing algorithms were built upon the atmospheric scattering model (ASM) and utilized hand-crafted priors such as dark channel prior [23], non-local prior [8] or color attenuation [20, 80] to recover a haze affected image. However, such hand-crafted features cannot account for non-homogeneous haze distribution, thereby resulting in poor performance. With the availability of synthetic datasets built upon ASM, deep learning-based algorithms [17, 43, 64, 65, 70] were designed to estimate the atmospheric light as well as transmission map to obtain a dazed image. Alternatively, several works [10, 22, 41, 47, 53, 54, 60, 68] focus on leveraging the availability of large synthetic datasets to develop an end-to-end dehazing mechanism. While most of these works focus on improving the performance by modifying the underlying restoration network, some [12, 71] examine the benefits of including depth priors as a mechanism to improve dehazing performance.

2.4 Perceptual Losses for Improving Image Fidelity

In order to improve the fidelity of restored images, perceptual loss [29] was proposed that utilized a positive boundary constraint. This was extended by contrastive loss [68] wherein a negative boundary constraint was integrated and demonstrated an improved image restoration quality. The effectiveness of contrastive loss results from triplet construction, attracting the attention of works that focus upon different sampling mechanisms both within the dataset [30], synthetically generating them [32] or saving different properties of features [8]. Recently [68] proposed integrating contrastive loss to improve the perceptual quality of haze-removed images by ensuring closer distance between restored and clear images in the latent space. However, such an approach relies on a fixed negative boundary, which becomes less effective in constraining the latent representation as the training progresses. The underlying restoration network can remove global haze. Thus, to ensure the effective removal of localized haze, a dynamic negative boundary constraint is required that adapts to the network performance such that the latent space is constrained throughout the training process, thereby removing localized haze. Hence we propose a dynamic negative sample generation and sampling mechanism to incorporate the dynamic negative boundary constraint.

3 Proposed Methodologies

3.1 PSNR based Patch Sampling

Existing learning-based image restoration techniques sample small patches from a degraded-clear pair for training to minimize memory requirements associated with using large-resolution images. As these patches are randomly sampled from the paired training dataset, a mechanism is lacking to select patches that contain suitable information for the restoration network, as showcased in Fig. 2. This allows for uneven distribution of samples that can be accurately, moderately, or poorly restored, with the distribution getting more skewed between poorly and accurately restored samples as the training progresses. Considering the hard samples to be representative of the long tail and characterized by lower sampling frequency during the training cycle, we

Figure 2: Motivation for sampling training patches to ensure increased presence of (top) informative samples (13.07 / 0.62) while suppressing (bottom) less informative samples (16.46 / 0.79).

conjecture a feedback mechanism to help alleviate this sampling bias. However, prior hard negative sampling approaches [18, 21, 48] are designed for high-level perception tasks and require the training patches to be passed through the underlying neural network. This results in gradient computation for patches that contain no-relevant details.

Thus to sample training patches without gradient computation, we propose a PSNR-based sampling mechanism that dynamically adjusts the sampling frequency of the training patches based on the PSNR of input and restored patches of the prior minibatch. We begin with an initial minibatch $(t \geq 2)$ to obtain the triplet of restored, clean and degraded patches and compute the degraded and restored PSNR. Given multiple degraded and restored PSNR, we fit a polynomial curve $(fn(.))$ representative of the network performance. This polynomial curve is updated after each training step to account for training dynamics which would vary the network performance. Thus we can estimate the restoration performance of a training sample $(\widehat{P_{output}}^{t+1})$ using only input PSNR $(P_{input}{}^{t+1})$. Based on estimated performance, easy samples are identified and replaced with more complex samples during data processing. We highlight that the additional computation of PSNR during data processing does not result in any observable computational overhead as the such operation is scaled using GPUs, reducing the overall computational time. A detailed sampling mechanism is included in Appendix A of the supplementary.

3.2 Adversarial Image Augmentation

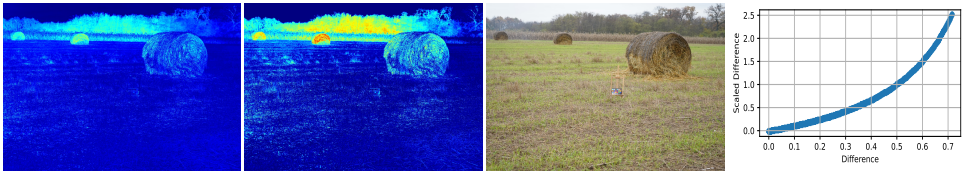
As there is an absence of a large corpus of paired images containing a wide variety of non-homogeneous haze distribution, we examine the utilization of synthetic haze generation as an alternative and mix the synthetic haze with real hazy images during training. With this intent, we examine the adversarial auto-augmentation approach [78] wherein data augmentation policies are dynamically generated to regularize the underlying CNN. While initially devised for high-level vision tasks such as classification and detection, this mechanism can be retrofitted for image restoration tasks, specifically image dehazing. We utilize handcrafted region-based data augmentation approaches to ensure non-uniform haze in training samples [55, 72]. However static nature of such augmentations, in the form of fixed crop size, hinders optimal training by failing to link network performance with degradation intensity and localization.

Adversarial Auto-Augment is designed as a min-max optimization framework wherein multiple sampled data-augmentation policies augment each minibatch. Subsequently, a target network is trained to minimize original and augmented minibatch loss. Following this, training loss is computed as a reward signal for each data augmentation policy. This reward signal is used to train the policy search network to maximize the training loss by generating adversarial augmentation policies.

Based on this framework, to ensure the compatibility of adversarial auto-augmentation for image dehazing tasks, we modify the search space operations while keeping the basic structure intact following [13]. An augmentation policy comprises 5 sub-policies, with each sub-policy composed of two sequential processes with a single control parameter to account for the magnitude of the operation. We generate 25 sub-policies randomly sampled and applied on each image within a minibatch. We provide additional details on different operations used in Appendix C of Supplementary.

3.3 Degradation Balancing via Focal Pixel Loss

Commonly used objective functions, such as L1 or MSE, compute the loss by equally weighting each pixel, thereby being skewed by the intensity of the majority pixel difference.



(a) Difference (b) Weighted (c) Clear Image (d) Histogram

Figure 3: Visualization of (a) pixel difference and (b) weighted fixed difference computed using the absolute difference between clean (c) and hazy (Fig. 2) images. Here blue and red denote minimum and maximum difference intensities, respectively. (d) Highlights the scaling curve where each pixel is scaled to ensure balanced contribution irrespective of the quantum of the pixel.

This results in the loss of extreme peaks within the restored image as these functions compute the mean over all pixels within a training patch. To avoid such degradation imbalance, we propose focal pixel loss (FPL) that addresses class imbalance at the pixel level by scaling the pixel difference distribution. Low quantum high magnitude differences contribute more than high quantum log magnitude differences. Specifically, given a restored and clear image pair, we first compute the pixel difference, and the weight matrix is (W) is computed for each pixel (x,y) as,

$$\begin{aligned} W(x,y) &= |I_R(x,y) - I_C(x,y)|^{\alpha} \quad and \quad W(x,y) \in [0,1] \\ FPL &= \frac{1}{w \times h} \sum_{x=0}^{w-1} \sum_{y=0}^{h-1} W(x,y) \bigotimes |I_R(x,y) - I_c(x,y)| \end{aligned} \tag{2}$$

Here $\bigotimes$ represents Hadamard product for an image of width (w) and height (h), and α is the scaling factor and set to 1 during our experiments. Furthermore, we highlight that L1 loss can be replaced with MSE. As this matrix is calculated for each image independently, it has a dynamic nature. Thus, it can adjust the weights so that high magnitude is always amplified and low magnitude differences are suppressed.

We provide a visual demonstration of the proposed mechanism on a typical difference map in Fig. 3(a) and its reweighed version in Fig. 3(b). We also plot the histogram of the difference map to provide an intuitive understanding of the working of the proposed mechanism. Thus in Fig. 3(b), we can observe that severely affected regions are amplified such that their contribution towards loss computation is higher vis-a-vis moderately affected regions.

3.4 Dynamic Contrastive Loss

Recently contrastive loss was utilized within image restoration to improve the perceptual quality of the restored image. However, the typical contrastive loss is constructed to minimize the distance between an anchor and positive samples while maximizing the distance between the anchor and negative samples. This is achieved using the following optimization,

$$CR(I_C, I_R, I_D) = \sum_{i=1}^{N} \omega_i \frac{|\phi_i(I_C) - \phi(I_R)|}{|\phi_i(I_D), \phi(I_R)|} \tag{3}$$

Here $CR(.)$ refers to contrastive regularization, ϕ_i refers to features extracted from i^{th} layer using ImageNet [16] pre-trained VGG-19 [59] network, and ω_i refers to weight coefficient similar to ones used in perceptual loss [29] i.e., $\left(\frac{1}{32}, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, 1\right)$. Similar to perceptual loss, N is set to 5. In addition, we utilize a multiscale approach wherein we resize the image

by $0.5\times$ and $2\times$ using bicubic interpolation and compute the mean of all scales ($\in [0.5, 1, 2]$) as the final loss.

However, a drawback of CR is that the boundary constraints defining the positive and negative samples are fixed. This results in poor enforcement of latent alignment during the training process, wherein the restoration quality is improving, thereby turning the input hazy image into a soft negative. Thus to ensure effective latent representation alignment, it is desired to have negative samples that exhibit properties of hard negatives. To achieve this, we propose a dynamic contrastive loss, wherein we generate synthetic negative samples by blending the input clear image with the hazy image using a hyperparameter $\gamma(\in [0.05, 0.95]$. This generates a range of negative samples for which we subsequently compute the PSNR values. Finally, given the PSNR of the restored image, we identify the hard negative sample, which has a similar PSNR, and utilize it for loss computation. We demonstrate the efficacy of such a mechanism by visualizing the latent space as the training progresses and include the results in Appendix D of Supplementary.

4 Experimental Evaluation

4.1 Datasets and Evaluation Metrics

For examining the effect of the proposed mechanism on SoTA, we use pixel- and feature-based metrics such as PSNR, SSIM [67], LPIPS [76], and NIQE [44] that are consistent across image restoration literature. Furthermore, for evaluating the performance of image dehazing algorithms, we use an aggregated training set comprising 30 images from IHaze [1], OHaze [2], NH-Haze [4], and Dense-haze [3] datasets. In addition, we utilize 10000 paired synthetic images from Reside [35] dataset. We utilize 5 images from NH-Haze [4], and Dense-haze [3] datasets to evaluate the performance.

We consider AECR [58], DIDH [24], and FFANet [47] as baseline algorithms upon which we integrate proposed mechanisms. For training purposes, we replace the pixel and feature-based losses with proposed Focal Pixel Loss and Dynamic Contrastive Loss and include the Auto-Augmentation based Patch Sampling into the data loader. It should be noted that the proposed loss functions are drop-in replacements and do not alter the architecture of the underlying restoration network.

4.2 Comparison with SoTA

We summarize the performance results on both real-world non-homogeneous dehazing datasets and synthetic datasets in Tab. 1. We would redirect readers to Appendix B in the Supplementary for an exhaustive qualitative and quantitative comparison. From Tab. 1, we observe a wide performance gap between Synthetic and Real datasets, i.e., SOTS-IN/OUT vs. Dense-haze and NH-Haze irrespective of the restoration algorithm, thereby demonstrating synthetic datasets to lack the representative power for modeling non-homogeneous haze distributions. Nevertheless, when integrating the proposed modifications to retrain FFA-Net, DIDH, and AECR-Net, we observe a consistent performance increase across all datasets, i.e., 1.96 dB, 0.46 dB, and 1.3 dB for FFANet, DIDH, and AECR-Net respectively for Dense-haze. Furthermore, similar improvements are observed for NH-Haze, with an increased magnitude of improvement for SOTS-IN and SOTS-OUT. However, since we observe a noticeable improvement in PSNR and perceptual quality (NIQE and LPIPS score from Appendix B) and not in structural similarity (SSIM), we conclude the performance improvement arises from the restoration of high-frequency features, i.e., the edge details within the restored images. This is validated by qualitative evaluation of the restored images included in Appendix D of Supplementary. We thus establish that the proposed mechanisms can further push the

restoration capability of current image dehazing algorithms and include the necessary ablations to identify individual contribution of different proposed mechanisms in Appendix C of Supplementary.

	Dense-haze PSNR (↑) / SSIM (↑)	NH-Haze PSNR (↑) / SSIM (↑)	SOTS-IN PSNR (↑) / SSIM (↑)	SOTS-OUT PSNR (↑) / SSIM (↑)
DuRN-US [41]	13.63 / 0.57	15.27 / 0.50	32.68 / 0.97	33.04 / 0.97
GridDehazenet [32]	12.96 / 0.50	15.32 / 0.60	32.16 / 0.98	30.86 / 0.98
FFA-Net [31]	14.01 / 0.56	18.11 / 0.66	36.39 / 0.98	33.57 / 0.98
TridentNet [33]	16.48 / 0.54	21.41 / 0.71	34.87 / 0.98	33.42 / 0.98
DA-Dehaze [5]	13.98 / 0.37	11.42 / 0.31	27.76 / 0.93	25.43 / 0.85
DIDH [6]	19.47 / 0.75	21.17 / 0.78	38.91 / 0.98	30.40 / 0.94
AECR-Net [43]	15.80 / 0.46	20.68 / 0.82	37.17 / 0.99	33.51 / 0.89
DeHamer [14]	16.62 / 0.56	19.18 / 0.79	36.63 / 0.98	35.18 / 0.98
D4 [40]	13.12 / 0.53	12.65 / 0.37	25.42 / 0.93	25.83 / 0.95
FogRemoval [42]	16.67 / 0.50	20.99 / 0.61	28.68 / 0.87	27.97 / 0.86
DEANet [21]	12.01 / 0.32	10.98 / 0.25	41.31 / 0.99	36.59 / 0.98
DehazeFormer-B [34]	11.68 / 0.32	12.84 / 0.35	37.84 / 0.99	34.95 / 0.98
Integrating Proposed Modifications				
Ours (FFA-Net)	15.97 (+1.96) / 0.64 (+0.08)	20.08 (+1.97) / 0.61 (-0.05)	38.19 (+1.80) / 0.95 (-0.03)	37.71 (+4.14) / 0.95 (-0.03)
Ours (DIDH)	19.93 (+0.46) / 0.71 (-0.04)	21.44 (+0.27) / 0.79 (+0.01)	40.11 (+1.2) / 0.96 (-0.02)	34.19 (+3.79) / 0.95 (+0.01)
Ours (AECR-Net)	17.10 (+1.3) / 0.57 (+0.11)	21.70 (+1.02) / 0.68 (-0.14)	40.67 (+3.5) / 0.99 (0)	36.73 (+3.22) / 0.98 (+0.09)

Table 1: Quantitative Evaluation of SoTA Image Dehazing algorithms on Real (Dense-haze, NH-Haze) and Synthetic (SOTS-IN, SOTS-OUT) datasets. We demonstrate performance improvement on prior works by retraining them using proposed modifications. An exhaustive performance summary is included in Appendix B of Supplementary.

5 Conclusion

In this paper, we highlighted the non-uniform distribution of haze variations to suppress extremities during the optimization process. To overcome such limitations, we propose solutions from the point of degradation imbalance. Thus, we develop the solutions at the data augmentation and loss computation level. Specifically, we propose a PSNR-based patch sampling mechanism to identify degradation intensity and estimate restoration performance. Based on this information, we balance the training sample distribution to cover the entire degradation landscape. We also identified the limitation of real-world datasets to capture a large diversity of natural degradations. Thus we propose a synthetic adversarial degradation generation mechanism powered by automatic adversarial augmentation and integrate PSNR-based patch sampling to identify hard negative samples and increase its sampling frequency. From the loss computation perspective, we propose focal pixel loss to ensure large pixel differences are considered during loss computation and are not suppressed by significant low pixel differences. In addition, we also propose dynamic contrastive loss that aims to ensure a compact latent space alignment between clear and restored images by adaptively adjusting the negative boundary constraint. Finally, we demonstrated the proposed mechanisms' efficacy by retraining prior dehazing algorithms and reporting significant performance improvement without modification to the underlying architecture.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. *Advances in neural information processing systems*, 32, 2019.

[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.

[8] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *arXiv preprint arXiv:2006.09882*, 2020.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. Autoaugment: Learning augmentation policies from data. *arXiv preprint arXiv:1805.09501*, 2018.

[14] 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.

[15] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9268–9277, 2019.

[16] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[17] 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.

[18] Bi'an Du, Xiang Gao, Wei Hu, and Xin Li. Self-contrastive learning with hard negative sampling for self-supervised point cloud learning. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 3133–3142, 2021.

[19] Ruochen Fan, Ming-Ming Cheng, Qibin Hou, Tai-Jiang Mu, Jingdong Wang, and Shi-Min Hu. S4net: Single stage salient-instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6103–6112, 2019.

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

[21] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, 32(9):1627–1645, 2009.

[22] 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.

[23] 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.

[24] Youngkyu Hong, Seungju Han, Kwanghee Choi, Seokjun Seo, Beomsu Kim, and Buru Chang. Disentangling label distribution for long-tailed visual recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6626–6636, 2021.

[25] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9924–9935, 2022.

[26] Ting-I Hsieh, Esther Robb, Hwann-Tzong Chen, and Jia-Bin Huang. Droploss for long-tail instance segmentation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 1549–1557, 2021.

[27] 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.

[28] 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.

[29] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.

[30] Yannis Kalantidis, Mert Bulent Sariyildiz, Noe Pion, Philippe Weinzaepfel, and Diane Larlus. Hard negative mixing for contrastive learning. *arXiv preprint arXiv:2010.01028*, 2020.

[31] Byungju Kim and Junmo Kim. Adjusting decision boundary for class imbalanced learning. *IEEE Access*, 8:81674–81685, 2020.

[32] Minseon Kim, Jihoon Tack, and Sung Ju Hwang. Adversarial self-supervised contrastive learning. *arXiv preprint arXiv:2006.07589*, 2020.

[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.

[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.

[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.

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

[37] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[38] 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.

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

[40] 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.

[41] 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.

[42] Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, Jiayun Wang, Boqing Gong, and Stella X Yu. Large-scale long-tailed recognition in an open world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2537–2546, 2019.

[43] 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.

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

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

[46] Kemal Oksuz, Baris Can Cam, Sinan Kalkan, and Emre Akbas. Imbalance problems in object detection: A review. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3388–3415, 2020.

[47] 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.

[48] Joshua Robinson, Ching-Yao Chuang, Suvrit Sra, and Stefanie Jegelka. Contrastive learning with hard negative samples. *arXiv preprint arXiv:2010.04592*, 2020.

[49] José A Sáez, Julián Luengo, Jerzy Stefanowski, and Francisco Herrera. Smote–ipf: Addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering. *Information Sciences*, 291:184–203, 2015.

[50] 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.

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

[52] Dvir Samuel and Gal Chechik. Distributional robustness loss for long-tail learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9495–9504, 2021.

[53] 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.

[54] 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.

[55] Pranjay Shyam, Kuk-Jin Yoon, Kyung-Soo Kim, and Sandeep Singh Sengar. Evaluating region modification based data augmentation techniques for low level vision tasks. In *IJCAI 2021 AI4AD Workshop on Artificial Intelligence for Autonomous Driving*. IJCAI 2021 Workshop on Artificial Intelligence for Autonomous Driving, 2021.

[56] Pranjay Shyam, Antyanta Bangunharcana, Kuk-Jin Yoon, and Kyung-Soo Kim. Dgss: Domain generalized semantic segmentation using iterative style mining and latent representation alignment. *arXiv preprint arXiv:2202.13144*, 2022.

[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.

[58] Pranjay Shyam, Sandeep Singh Sengar, Kuk-Jin Yoon, and Kyung-Soo Kim. Lightweight hdr camera isp for robust perception in dynamic illumination conditions via fourier adversarial networks. *arXiv preprint arXiv:2204.01795*, 2022.

[59] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

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

[61] Shuyang Sun, Liang Chen, Gregory Slabaugh, and Philip Torr. Learning to sample the most useful training patches from images. *arXiv preprint arXiv:2011.12097*, 2020.

[62] Jingru Tan, Xin Lu, Gang Zhang, Changqing Yin, and Quanquan Li. Equalization loss v2: A new gradient balance approach for long-tailed object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1685–1694, 2021.

[63] Jiaqi Wang, Wenwei Zhang, Yuhang Zang, Yuhang Cao, Jiangmiao Pang, Tao Gong, Kai Chen, Ziwei Liu, Chen Change Loy, and Dahua Lin. Seesaw loss for long-tailed instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9695–9704, 2021.

[64] 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.

[65] 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.

[66] Tao Wang, Yu Li, Bingyi Kang, Junnan Li, Junhao Liew, Sheng Tang, Steven Hoi, and Jiashi Feng. The devil is in classification: A simple framework for long-tail instance segmentation. In *European conference on computer vision*, pages 728–744. Springer, 2020.

[67] 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.

[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.

[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.

[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.

[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.

[72] Jaejun Yoo, Namhyuk Ahn, and Kyung-Ah Sohn. Rethinking data augmentation for image super-resolution: A comprehensive analysis and a new strategy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8375–8384, 2020.

[73] 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.

[74] Yuhang Zang, Chen Huang, and Chen Change Loy. Fasa: Feature augmentation and sampling adaptation for long-tailed instance segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3457–3466, 2021.

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

[76] 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.

[77] Songyang Zhang, Zeming Li, Shipeng Yan, Xuming He, and Jian Sun. Distribution alignment: A unified framework for long-tail visual recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2361–2370, 2021.

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

[79] Yaochi Zhao, Fusheng Lin, Shiguang Liu, Zhuhua Hu, Hui Li, and Yong Bai. Constrained-focal-loss based deep learning for segmentation of spores. *IEEE Access*, 7:165029–165038, 2019.

[80] 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.