Supplementary Materials for "Curricular Contrastive Regularization for Physics-aware Single Image Dehazing"

Yu Zheng[1], Jiahui Zhan[1], Shengfeng He[2], Junyu Dong[1], Yong Du[1*]
[1] College of Computer Science and Technology, Ocean University of China
[2] School of Computing and Information Systems, Singapore Management University

1. Model Analysis

Block Number. Our C^{2}PNet includes 3 groups of FA blocks, with the proposed PDU embedded in each block. To investigate the effect of different numbers of blocks per group, we summarize the corresponding quantitative results on SOTS-indoor in Table 1. The results show that even with a lightweight structure (6 blocks per group), our proposed C^{2}PNet can achieve competent performance. Specifically, our method can improve the PSNR by 0.71dB and save 0.20M parameters compared to AECRNet, and improve the PSNR by 1.49dB and save 2.05M (nearly 0.5$\times$) parameters compared to FFANet. Additionally, as the number of blocks increases, the performance of our network further improves. Considering the parameter-performance trade-off, we adopt the setting of 19 blocks per group, which has a much lower number of parameters than MAXIM-2S (nearly 0.5$\times$) and DeHamer (nearly 0.05$\times$), with PSNR improvements of 4.45dB and 5.93dB, respectively.

Table 1. Test on different numbers of blocks per group.

Method	#Blocks	PSNR	SSIM	#Params
C^{2}PNet	6	37.88	0.9926	2.41M
	12	40.41	0.9945	4.62M
	18	42.33	0.9952	6.82M
	19	42.56	0.9954	7.17M
	20	42.69	0.9954	7.53M
	24	43.37	0.9957	9.02M
FFANet	19	36.39	0.9886	4.46M
AECRNet	-	37.17	0.9901	2.61M
MAXIM-2S	-	38.11	0.9908	14.1M
DeHamer	-	36.63	0.9881	132.45M

Negative Sample Setting. The principle behind selecting the existing methods used to generate the non-easy negatives is that their performance distribution, *e.g.*, PSNR, should be as comprehensive as possible. This consideration arises from a trade-off between the compactness of the guidance provided by the negatives and the effectiveness of the resultant force generated by C^{2}R, as also implied by Fig.1 of the main paper. Specifically, in the case of 7 negatives, we choose the results from DCP (16.62 PSNR), AODNet(19.06 PSNR), DehazeNet (21.14 PSNR), GCANet (30.06 PSNR), FFANet (36.39 PSNR), AECRNet (37.17 PSNR), and FFANet with our C^{2}R (39.24 PSNR). Note that we can obtain more negatives by deploying our C^{2}R on other methods.

Following this principle, we first examine the impact of different numbers of negatives and report the quantitative results on SOTS-indoor in Table 2. Similar to the findings in non-consensual CR, we observe that using more negatives leads to better dehazing performance. Specifically, our C^{2}R obtains a PSNR improvement of 1.28dB against non-consensual CR (*i.e.*, 42.60dB vs.41.32dB) in the case of 10 negatives which is the optimal setting reported in non-consensual CR. However, we find that the performance only slightly increases when the number of negatives changes from 7 to 10 in our method. Therefore, we set the number of negatives in our method to 7, as it can achieve a good trade-off between computational cost and training speed while still providing effective guidance for dehazing.

We further explore the impact of collecting consensual negatives using another approach, *i.e.*, generating the non-easy negatives by simply interpolating between the hazy image (the easy negative) and the ground-truth image (the positive) based on PSNR. We use a lightweight structure, *i.e.*, 6 blocks per group in the model, and list the quantitative results on SOTS-indoor in Table 4. The results show that using simple interpolation to generate the negatives severely decreased the performance. As discussed in the main paper, we believe this is because the interpolated image lacks post-dehazing priors. Therefore, such negatives cannot provide purposive cues to guide the model to learn from worse pat-

*Corresponding author (csyongdu@ouc.edu.cn).

Table 2. The effect of the number of negative samples.

#Neg.	1	2	4	7	10
PSNR	41.12	42.08	42.47	42.56	42.60
SSIM	0.9947	0.9952	0.9954	0.9954	0.9955

terns. By contrast, different patterns of haze residues in the recoveries produced by existing dehazers carry the information that tells the model what haze is hard to remove, thus aiding in regularizing the objective of the model. Although the over-dehazing or other related problems that may affect the stability of consensual CR exist in those non-easy samples, they can be alleviated by our curriculum learning strategy that fixes and maximizes the weight of the easy negative. In this way, the solution space would be almost unbiased, as the easy sample shares the same semantics with the positive (except for the haze) as well as embedding the original haze distribution. We also show an example of visual comparison in Fig. 1. Note that to generate each negative, we keep the average PSNR of the interpolated images basically the same as that of the recoveries by the selected dehazer.

We also demonstrate the practicability of our C^2R in two additional perceptual metrics used for distinguishing hard and ultra-hard negatives, *i.e.*, LPIPS (reference-based) and NIQE (distortion-based). The results are presented in Table 3, showing that using these two metrics can yield performances comparable to that of the original PSNR metric, and can even deliver performance gains in the real-world scene.

Table 3. A comparison of using diverse distinguishing metrics.

Distinguishing Metric	SOTS-indoor		NH-Haze2	
	PSNR	SSIM	PSNR	SSIM
PSNR	**42.56**	**0.9954**	21.19	0.8334
LPIPS	42.51	0.9954	**21.29**	0.8337
NIQE	42.54	0.9954	21.16	**0.8361**

Parameter Sensitivity for the CL strategy. To verify the effect of the hyperparameter γ in our CL strategy, we present the results in Table 5. When $\gamma = 0$, the hard and ultra-hard negatives are not treated differently, resulting in a poor performance against the other settings. This demonstrates the necessity of the proposed CL strategy. On the other hand, it can be observed that the best PSNR performance is 42.56dB when $\gamma = 0.25$.

2. Additional Visual results

To qualitatively evaluate the proposed PDU, we follow PFDN to use a SIFT descriptors based method [2] to map the features generated by the last FDU and our PDU in the base+C^2R network into a 3D color space. The visualizations are shown in Fig. 2. It can be seen that our PDU can produce more precise features especially in preserving the structural information (see the painting on the wall), compared to FDU. This suggests the superiority of PDU over

Table 4. A quantitative comparison using the negatives generated by simple interpolation.

Neg.	by dehazer	by interpolation
PSNR	37.88	19.45
SSIM	0.9926	0.8230

Table 5. Performance on SOTS-indoor concerning different values of the hyperparameter γ. $\gamma = 0$ means to average the contributions of different negative samples.

γ	0	0.2	0.25	0.3	0.4	0.5
PSNR	42.18	42.45	**42.56**	42.40	42.37	42.32
SSIM	0.9952	0.9954	**0.9954**	0.9953	0.9953	0.9953

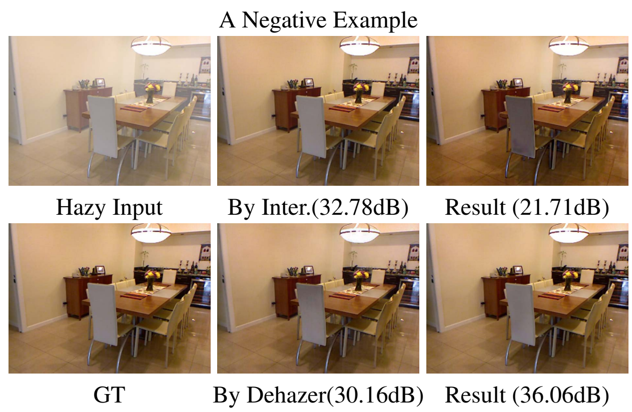


Figure 1. Visual results with negatives generated by interpolation or existing dehazers. (Zoom in for better view.)

Figure 2. Visualizations of the features from FDU and PDU. (Zoom in for better view.)

FDU in haze removal, mainly due to the disentangled estimation of the features corresponding to the atmospheric light and the transmission map.

To further investigate the generalization ability of the proposed C^2PNet, we conduct an additional experiment on the Fattal's dataset [1], in which the images are captured in real-world environments. The visual comparisons are provided in Fig. 3. Our method generates visually satisfying results with fewer haze residues or color distortions compared to all the other competitors, demonstrating its superiority and generalization ability in real-world scenes.

We also provide visual results regarding the generality of C^2R in Fig. 4-Fig. 7, respectively. Our C^2R consistently achieves the best visual results compared to other methods. For example, in Fig. 5, GCANet produces the result containing several gray spots and artifacts in the wall. Although the other competitors can remove some of them, they introduce color distortion problems. In contrast, our method generates the restoration that is closest to the ground truth. These results powerfully demonstrate the generality of our C^2R, which can boost the performance of various image dehazing approaches.

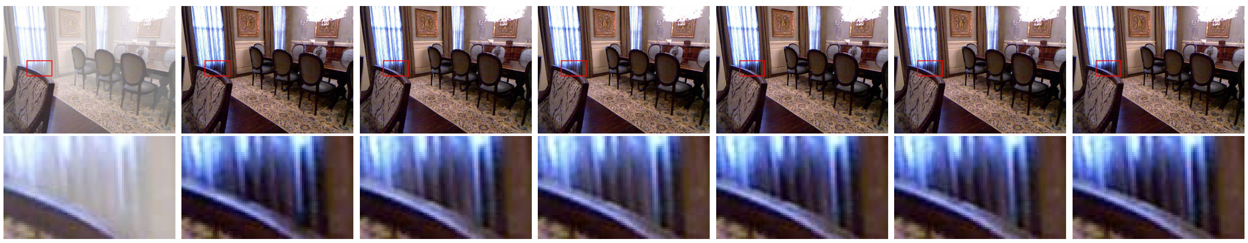
References

[1] Raanan Fattal. Dehazing using color-lines. *ACM TOG*, 34(1):1–14, 2014. 3

[2] Ce Liu, Jenny Yuen, and Antonio Torralba. Sift flow: Dense correspondence across scenes and its applications. *IEEE TPAMI*, 33(5):978–994, 2010. 2

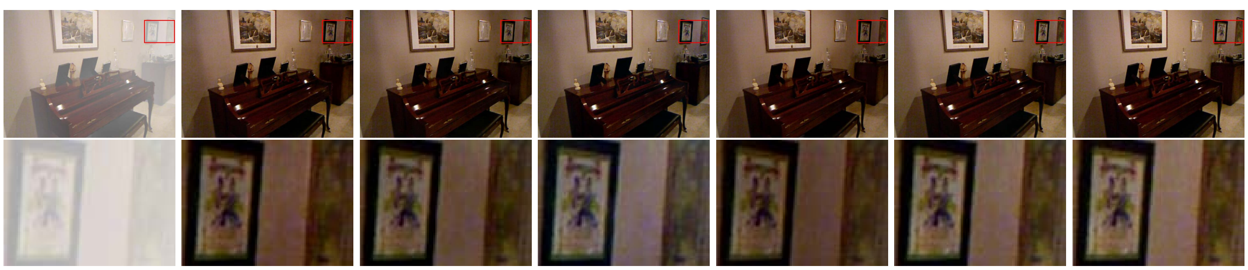
Haze AODNet GDN FFANet AECRNet MAXIM DeHamer Ours

Figure 3. Visual comparisons on the Fattal's dataset. (Zoom in for better view.)



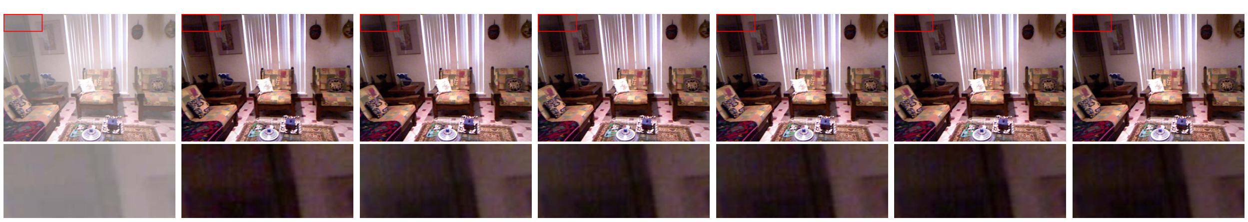
(a) Hazy Input (b) FFA-Net (c) CR:✓ +Space:non-consensual +CL:× (d) CR:✓ +Space:consensual +CL:× (e) CR:✓ +Space:consensual +CL:self-paced (f) CR:✓ +Space:consensual +CL:ours (g) GT

Figure 4. Visual comparisons for different contrastive regularizations on FFA-Net. (Zoom in for better view.)



(a) Hazy Input (b) GCANet (c) CR:✓ +Space:non-consensual +CL:× (d) CR:✓ +Space:consensual +CL:× (e) CR:✓ +Space:consensual +CL:self-paced (f) CR:✓ +Space:consensual +CL:ours (g) GT

Figure 5. Visual comparisons for different contrastive regularizations on GCANet. (Zoom in for better view.)



(a) Hazy Input (b) GDN (c) CR:✓ +Space:non-consensual +CL:× (d) CR:✓ +Space:consensual +CL:× (e) CR:✓ +Space:consensual +CL:self-paced (f) CR:✓ +Space:consensual +CL:ours (g) GT

Figure 6. Visual comparisons for different contrastive regularizations on GDN. (Zoom in for better view.)

(a) Hazy Input (b) MSBDN (c) CR:✓ +Space:non-consensual +CL:× (d) CR:✓ +Space:consensual +CL:× (e) CR:✓ +Space:consensual +CL:self-paced (f) CR:✓ +Space:consensual +CL:ours (g) GT

Figure 7. Visual comparisons for different contrastive regularizations on MSBDN. (Zoom in for better view.)