Image Dehazing Transformer with Transmission-Aware 3D Position Embedding

Chunle Guo[1] Qixin Yan[2] Saeed Anwar[3] Runmin Cong[4] Wenqi Ren[5] Chongyi Li[6*]

[1] TMCC, CS, Nankai University [2] Tianjin University [3] Australian National University

[4] Beijing Jiaotong University [5] Sun Yat-sen University [6] S-Lab, Nanyang Technological University

guochunle@nankai.edu.cn qxyan@tju.edu.cn saeedanwarcse@gmail.com

rmcong@bjtu.edu.cn rwq.renwenqi@gmail.com lichongyi25@gmail.com

https://li-chongyi.github.io/Proj_DeHamer.html

This supplementary material provides the detailed network structure and parameters of our method and additional experimental results.

- We provide the detailed structures and parameters of our Transformer module, CNN encoder module, and CNN decoder module in Figure 1, Figure 2, and Figure 3, respectively.

- We provide the comparison with winner methods(TND [6] and DW-GAN [3]) of NTIRE 2020/2021 image dehazing challenge in Table 1.

- We provide the comparisons with the other two state-of-the-art methods (CL [10] and DA [9]) in Table 2.

- We provide the comparisons of inference time with state-of-the-art methods (FFANet [8], TND [6], and DW-GAN [3]) for processing an image of size 1600×1200×3 using an NVIDIA 1080Ti GPU in Table 3.

- We provide more visual comparison results in Figure 4, Figure 5, Figure 6, and Figure 7.

Table 1. **Quantitative comparisons with the winner methods of NTIRE 2020/2021 on SOTS, Dense-Haze, and NH-Haze datasets.** Our method achieves the highest PSNR and SSIM scores on the SOTS dataset and the highest PSNR score on the Dense-Haze dataset when compared with the 2020/2021 champion solutions TDN [6] and DW-GAN [3].

Methods	SOTS		Dense-Haze		NH-Haze	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
TDN ('20)	34.59	0.9754	15.50	0.5081	20.44	0.6683
DW-GAN ('21)	<u>35.94</u>	<u>0.9860</u>	<u>16.49</u>	**0.5911**	**21.51**	**0.7111**
DeHamer (Ours)	**36.63**	**0.9881**	**16.62**	<u>0.5602</u>	<u>20.66</u>	<u>0.6844</u>

*Chongyi Li (lichongyi25@gmail.com) is the corresponding author.

Table 2. **Quantitative comparisons with two additional state-of-the-art methods (CL [10] and DA [9]) on SOTS, Dense-Haze, and NH-Haze datasets.** Our method achieves the highest PSRN and SSIM scores on the Dense-Haze dataset and highest PSNR score on the NH-Haze dataset, and the second-highest scores across other datasets.

Methods	SOTS		Dense-Haze		NH-Haze	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
CL ('21)	**37.17**	**0.9901**	<u>15.80</u>	0.4660	<u>19.88</u>	**0.7173**
DA ('20)	25.30	0.9423	12.56	<u>0.4994</u>	13.77	0.5349
DeHamer (Ours)	<u>36.63</u>	<u>0.9881</u>	**16.62**	**0.5602**	**20.66**	<u>0.6844</u>

Table 3. **Inference time comparisons with state-of-the-art methods (FFANet [8], TND [6], and DW-GAN [3]).** Compared with these methods, our method achieves the fastest inference speed.

Methods	FFANet	TDN	DW-GAN	DeHamer (Ours)
Inference Time (s)	2.62	0.63	0.48	**0.41**

References

[1] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao. An end-to-end system for single image haze removal. *TIP*, 25(11):5187–5198, 2016. 4, 5, 6, 7

[2] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and M. H. Yang. Multi-scale boosted dehazing network with dense feature fusion. In *CVPR*, pages 2157–2167, 2020. 4, 5, 6, 7

[3] M. Fu, H. Liu, Y. Yu, J. Chen, and K. Wang. Dw-gan: A discrete wavelet transform gan for nonhomogeneous dehazing. In *CVPRW*, pages 203–212, 2021. 1

[4] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *TPAMI*, 33(12):2341–2353, 2011. 4, 5, 6, 7

[5] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. Aod-net: All-in-one dehazing network. In *ICCV*, pages 4770–4778, 2017. 4, 5, 6, 7

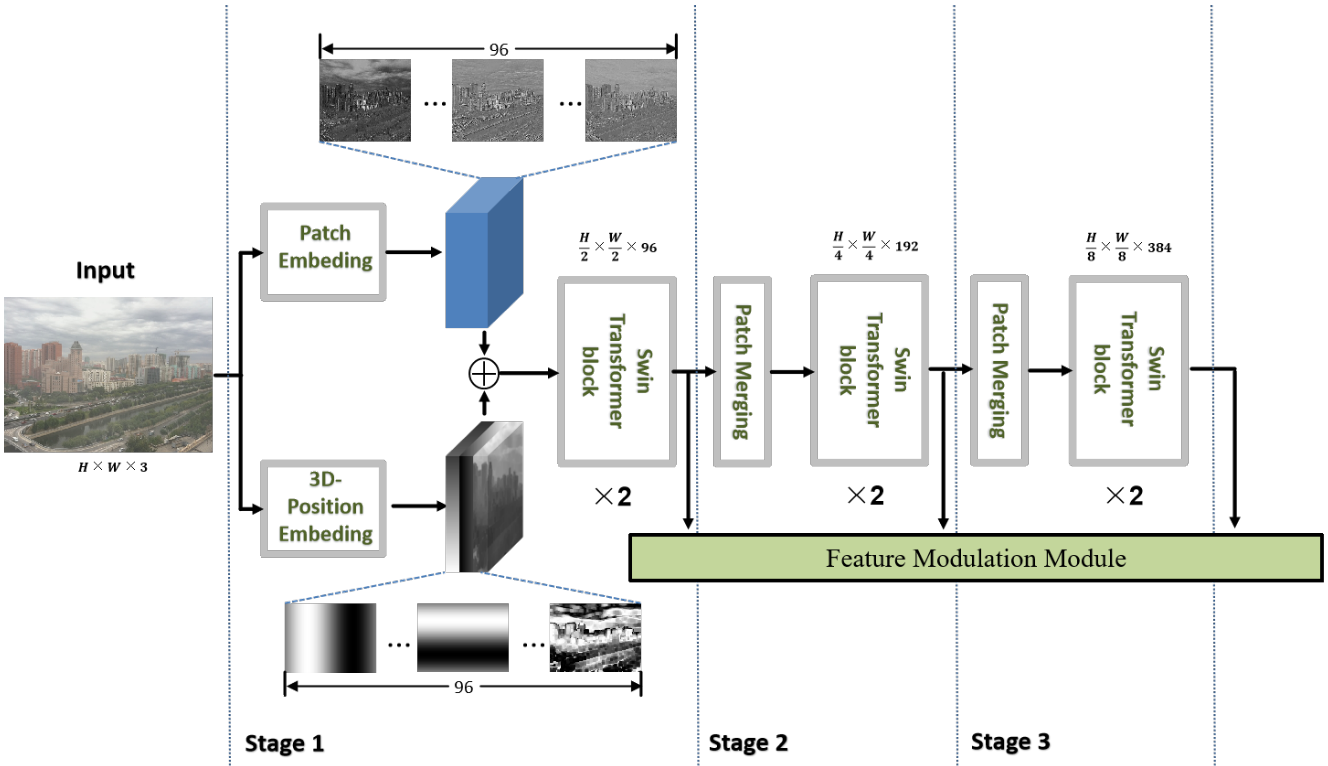


Figure 1. **Detailed network structure and parameters of the proposed Transformer module.**

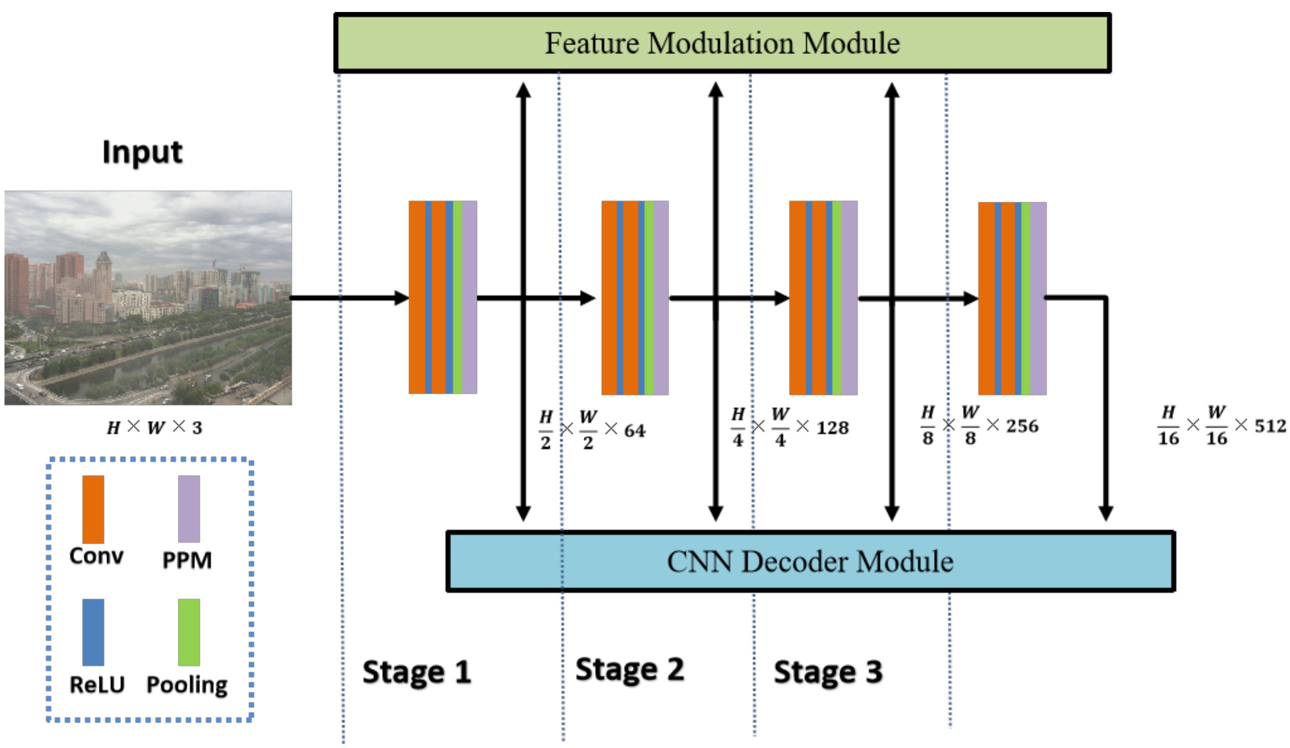


Figure 2. **Detailed network structure and parameters of the proposed CNN encoder module.**

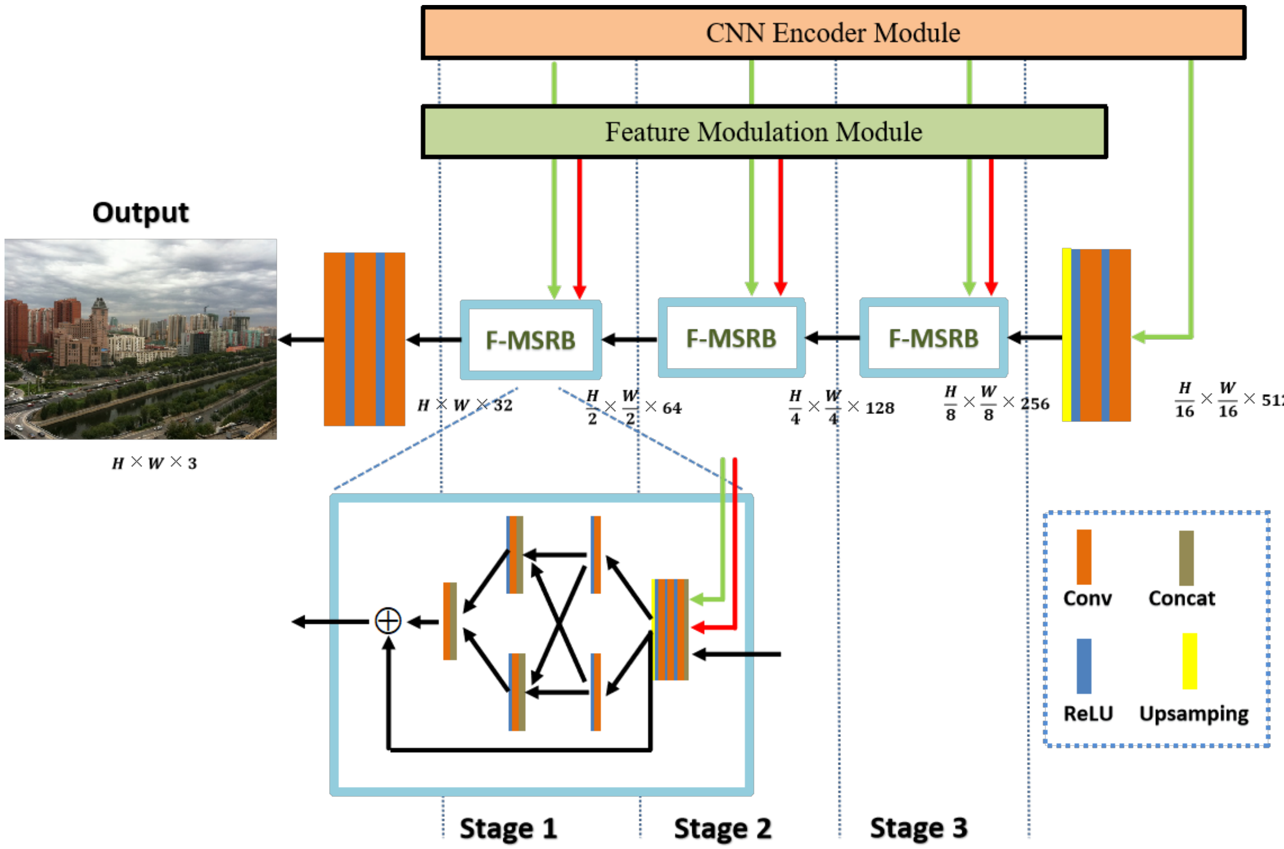


Figure 3. **Detailed network structure and parameters of the proposed CNN decoder module.**

[6] J. Liu, H. Wu, Y. Xieand Y. Qu, and L. Ma. Trident dehazing network. In *CVPRW*, pages 430–431, 2020. 1

[7] X. Liu, Y. Ma, Z. Shi, and J. Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *CVPR*, pages 7314–7323, 2019. 4, 5, 6, 7

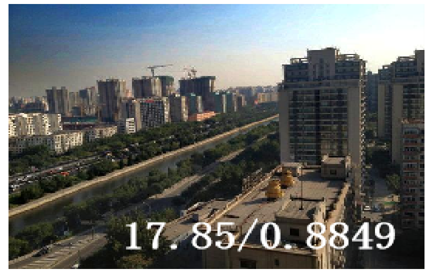
[8] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia. FFA-Net: Feature fusion attention network for single image dehazing. In *AAAI*, pages 11908–11915, 2020. 1, 4, 5, 6, 7

[9] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang. Domain adaptation for image dehazing. In *CVPR*, pages 2808–2817, 2020. 1

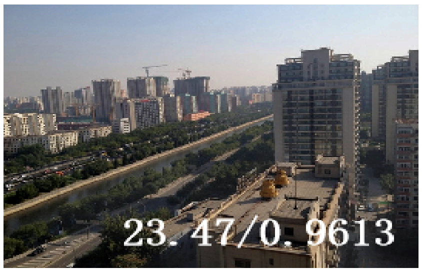
[10] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma. Contrastive learning for compact single image dehazing. In *CVPR*, pages 10551–10560, 2021. 1

[11] Z. Zheng, W. Ren, X. Cao, X. Hu, T. Wang, F. Song, and X. Jia. Ultra-high-definition image dehazing via multi-guided bilateral learning. In *CVPR*, pages 16185–16194, 2021. 4, 5, 6, 7

(a) input

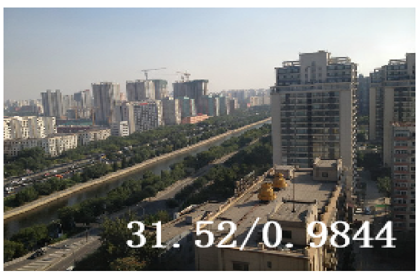


(b) DCP [4]



(c) DehazeNet [1]

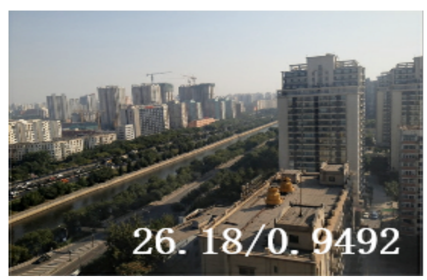
(d) AODNet [5]



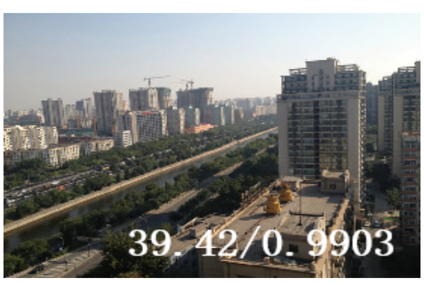
(e) GridDehazeNet [7]

(f) FFANet [8]

(g) MSBDN [2]



(h) UHD [11]



(i) DeHamer (Ours)

(j) GT

Figure 4. **Visual comparisons on a synthetic hazy image sampled from SOTS-Outdoor testing set.** Zoom in for best view.

(a) input (b) DCP [4]

(c) DehazeNet [1] (d) AODNet [5]

(e) GridDehazeNet [7] (f) FFANet [8]

(g) MSBDN [2] (h) UHD [11]

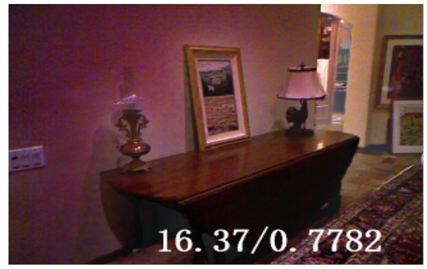
(i) DeHamer (Ours) (j) GT

Figure 5. **Visual comparisons on a synthetic hazy image sampled from SOTS-Outdoor testing set.** Zoom in for best view.

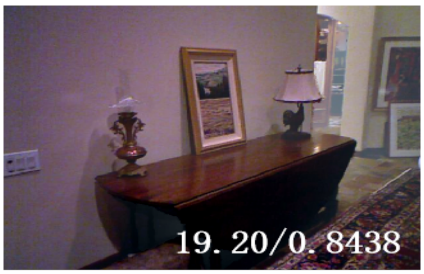
(a) input
(b) DCP [4]
(c) DehazeNet [1]
(d) AODNet [5]
(e) GridDehazeNet [7]
(f) FFANet [8]
(g) MSBDN [2]
(h) UHD [11]
(i) DeHamer (Ours)
(j) GT

Figure 6. **Visual comparisons on a synthetic hazy image sampled from SOTS-indoor testing set.** Zoom in for best view.

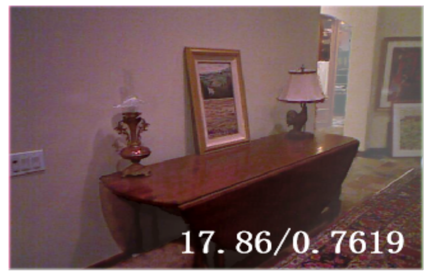
(a) input



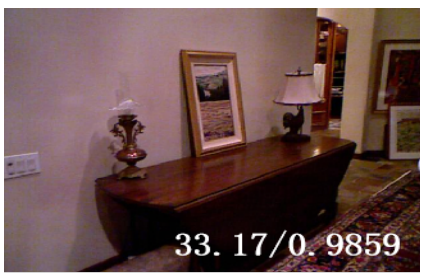
(b) DCP [4]



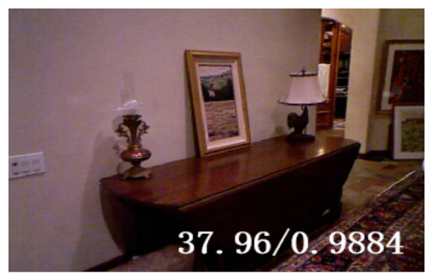
(c) DehazeNet [1]



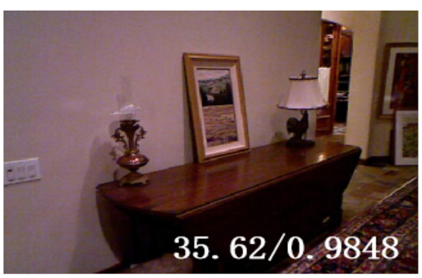
(d) AODNet [5]



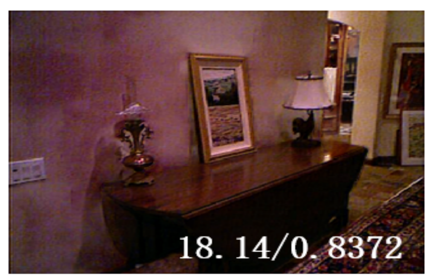
(e) GridDehazeNet [7]



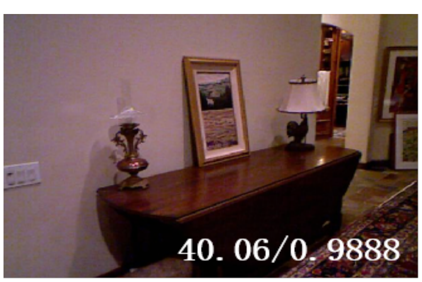
(f) FFANet [8]



(g) MSBDN [2]



(h) UHD [11]



(i) DeHamer (Ours)

(j) GT

Figure 7. **Visual comparisons on a synthetic hazy image sampled from SOTS-indoor testing set.** Zoom in for best view.