Aerial Image Dehazing with Attentive Deformable Transformers

Ashutosh Kulkarni and Subrahmanyam Murala

Supplementary Material

Overview

The supplementary material contains:

- Conceptual Differences between Proposed Method and Existing Methods
- Loss Functions
- More Discussion about Ablation Study
- More Qualitative Results

1 Conceptual Differences between Proposed Method and Existing Methods

The conceptual differences between the networks of the existing methods and the proposed method are provided in Table S 1. The proposed method incorporates all the essential blocks required for effective aerial image dehazing.

2 Loss Functions

We train the proposed network using L_1 loss ($\mathbb{L}_1$), edge loss ($\mathbb{L}_{Edge}$) and perceptual loss ($\mathbb{L}_p$). The detailed equations of these loss functions are:

2.1 L_1 Loss

$\mathbb{L}_1$ loss aims at reduction of the per pixel difference between the output image (I_O) and ground truth (I_{GT}), which can be formulated as:

$$\mathbb{L}_1 = \|I_O - I_{GT}\|_1 \tag{1}$$

2.2 Edge Loss

Edge loss ($\mathbb{L}_{Edge}$) aims at improving the edge details by calculating the discrepancies between edges of the output image and ground truth image. It can be represented as:

$$\mathbb{L}_{Edge} = \|\psi(I_O) - \psi(I_{GT})\|_1 \tag{2}$$

where, $\psi(\cdot)$ is Sobel operator.

2.3 Perceptual Loss

The perceptual loss ($\mathbb{L}_P$) aims at improving the perceptual quality of the output by calculating the differences between output and ground truth at various feature levels. The pretrained layers of VGG-16 [1] model are used as feature space for the loss calculation. The perceptual loss can be represented mathematically as:

$$\mathbb{L}_P = \left\|\Phi_{i\in(3,8,15)}(I_O) - \Phi_{i\in(3,8,15)}(I_{GT})\right\|_1 \tag{3}$$

where, $\Phi_i(\cdot)$ is i^{th} layer of VGG-16 model.

3 More Discussion about Ablation Study

The description regarding which blocks and loss functions are used in each network setting in ablation study is provided in Table S 2.

Method	**Publication**	**Local Dependency**	**Global Dependency**	**Geometric Adaptability**	**Edge Boosting Skip Connections**
SPA-GAN [7]	-	✓	×	×	×
Huang *et al.* [9]	WACV-20	✓	×	×	×
SkyGAN [10]	WACV-21	✓	×	×	×
UFormer [8]	CVPR-22	✓	✓	×	×
Proposed Method	-	✓	✓	✓	✓

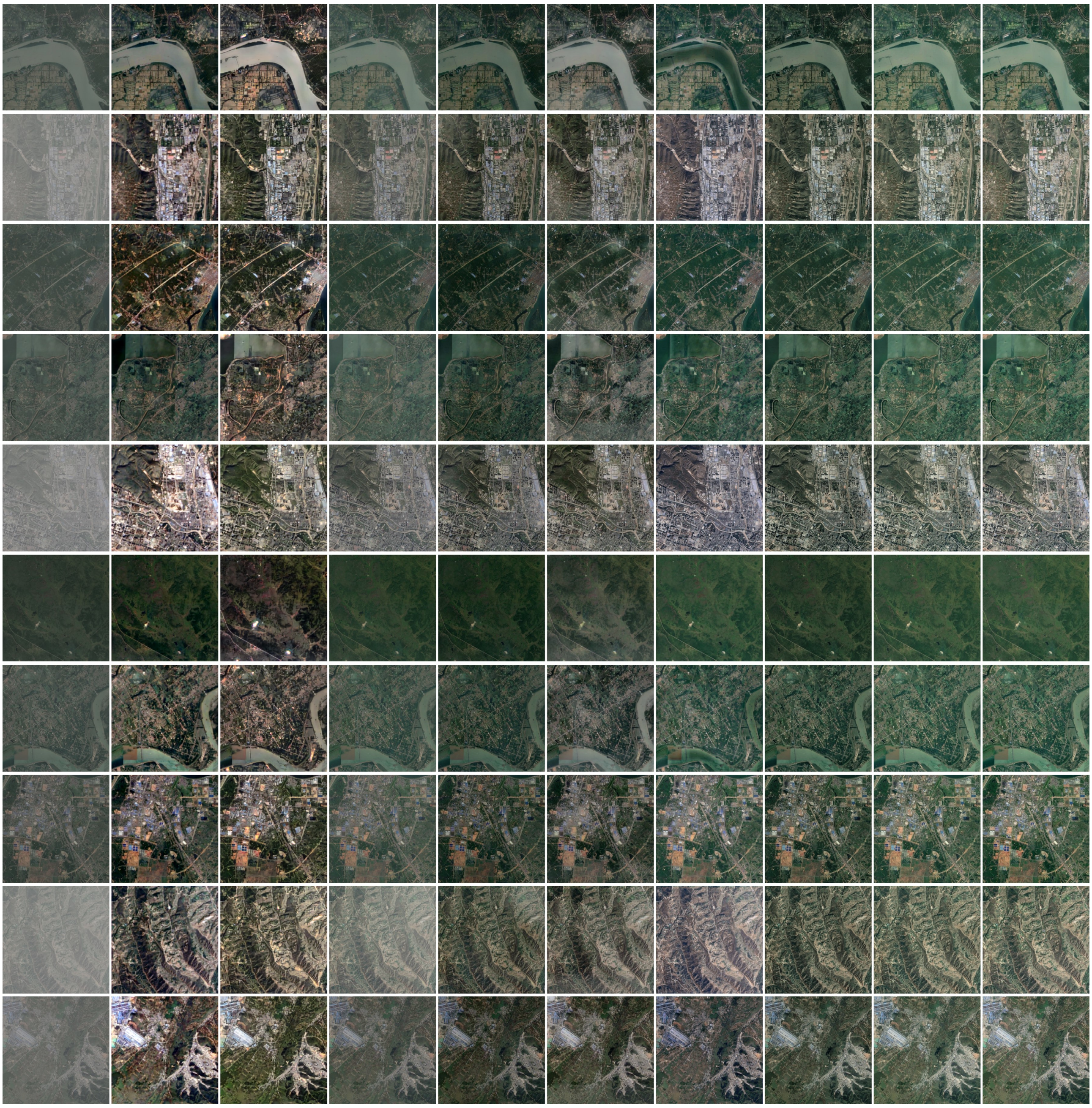
Table S 1: Conceptual comparison of the proposed method with existing state-of-the-art methods.

Setting	**Attention Type**	**Offset Extraction Type**	**Skip Connection Type**	**Loss Functions**
Network 1	Vanilla Attention [11]	Spatially Attentive	Edge Boosting Skip Connection	$\mathbb{L}_1 + \mathbb{L}_{Edge} + \mathbb{L}_P$
Network 2	DW-MSA [8]	Spatially Attentive	Edge Boosting Skip Connection	$\mathbb{L}_1 + \mathbb{L}_{Edge} + \mathbb{L}_P$
Network 3	Deformable MSA	Deformable [12]	Edge Boosting Skip Connection	$\mathbb{L}_1 + \mathbb{L}_{Edge} + \mathbb{L}_P$
Network 4	Deformable MSA	Modulated Deformable [13]	Edge Boosting Skip Connection	$\mathbb{L}_1 + \mathbb{L}_{Edge} + \mathbb{L}_P$
Network 5	Deformable MSA	Spatially Attentive	No Skip Connection	$\mathbb{L}_1 + \mathbb{L}_{Edge} + \mathbb{L}_P$
Network 6	Deformable MSA	Spatially Attentive	Regular Skip Connection	$\mathbb{L}_1 + \mathbb{L}_{Edge} + \mathbb{L}_P$
Network 7	Deformable MSA	Spatially Attentive	Edge Boosting Skip Connection	$\mathbb{L}_1$
Network 8	Deformable MSA	Spatially Attentive	Edge Boosting Skip Connection	$\mathbb{L}_1 + \mathbb{L}_{Edge}$
Proposed	Deformable MSA	Spatially Attentive	Edge Boosting Skip Connection	$\mathbb{L}_1 + \mathbb{L}_{Edge} + \mathbb{L}_P$

Table S 2: Details regarding modules and losses used in each network setting *(DW-MSA: DepthWise convolution based Multi-head Self Attention)*.

4 More Qualitative Results

We provide more qualitative results on RICE dataset for aerial haze removal in Figure S 1. We compare the qualitative results with state-of-the-art methods GCANet [2], USID [3], MSBDN [4], TSDNet [5], RDNet [6], SPA-GAN [7] and UFormer [8]. As seen from the results, the proposed method is able to maintain more spatial content and color balance.



Input GCANet [2] USID [3] MSBDN [4] TSDNet [5] RDNet [6] SPA-GAN [7] UFormer [8] Ours Ground Truth

Figure S 1: Qualitative results comparison of the proposed method with existing state-of-the-art methods on RICE dataset for aerial haze removal.

References

[1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[2] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua, "Gated context aggregation network for image dehazing and deraining," in *2019 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2019, pp. 1375–1383.

[3] J. Li, Y. Li, L. Zhuo, L. Kuang, and T. Yu, "Usid-net: Unsupervised single image dehazing network via disentangled representations," *IEEE Transactions on Multimedia*, pp. 1–1, 2022.

[4] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2157–2167.

[5] R. W. Liu, Y. Guo, Y. Lu, K. T. Chui, and B. B. Gupta, "Deep network-enabled haze visibility enhancement for visual iot-driven intelligent transportation systems," *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2022.

[6] S. Zhao, L. Zhang, Y. Shen, and Y. Zhou, "Refinednet: A weakly supervised refinement framework for single image dehazing," *IEEE Transactions on Image Processing*, vol. 30, pp. 3391–3404, 2021.

[7] H. Pan, "Cloud removal for remote sensing imagery via spatial attention generative adversarial network," *arXiv preprint arXiv:2009.13015*, 2020.

[8] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general u-shaped transformer for image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 17 683–17 693.

[9] B. Huang, L. Zhi, C. Yang, F. Sun, and Y. Song, "Single satellite optical imagery dehazing using sar image prior based on conditional generative adversarial networks," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2020, pp. 1806–1813.

[10] A. Mehta, H. Sinha, M. Mandal, and P. Narang, "Domain-aware unsupervised hyperspectral reconstruction for aerial image dehazing," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2021, pp. 413–422.

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[12] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 764–773.

[13] X. Zhu, H. Hu, S. Lin, and J. Dai, "Deformable convnets v2: More deformable, better results," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 9308–9316.