Enhancing Visual Clarity in Satellite Imagery for Remote Sensing and Analysis

Dr. Rachna Somkunwar
Associate professor
Department of Computer Engineering
Dr. D. Y. Patil Institute of Technology, Pimpri, Pune
rachnasomkunwar12@gmail.com

Ankit Pal
Department of Computer Engineering
Dr. D. Y. Patil Institute of Technology, Pimpri, Pune

Gaurav Vibhute
Department of Computer Engineering
Dr. D. Y. Patil Institute of Technology, Pimpri, Pune

Harshit Godawat
Department of Computer Engineering
Dr. D. Y. Patil Institute of Technology, Pimpri, Pune

Kartik Gonde
Department of Computer Engineering
Dr. D. Y. Patil Institute of Technology, Pimpri, Pune

Abstract

Image dehazing is the process of improving image quality by reducing the effects of atmospheric haze, making images clearer and more detailed. This involves estimating haze parameters, enhancing the image by removing haze, and applying post-processing as needed. It finds applications in surveillance, remote sensing, autonomous vehicles, and photography, where visibility is crucial. Many dehazing algorithms assume uniform haze, which may not hold in real-world scenarios and depend on accurate haze parameter estimation. In extreme cases, dehazing methods may not fully restore image details, and their computational intensity can be limiting in resource-constrained applications. In this work we propose an end to end architecture for image dehazing consisting of enhanced multi-head attention transformer block using wavelet transformation and deformable gated attention as skip connection in encoder decoder architecture.

I. Introduction

Advancements in aerial imaging technology have greatly improved the quality of aerial pictures, and these improved images find applications in various fields like identifying buildings, assessing earthquake damage, and breaking down images into their components. However, the success of these applications heavily relies on having clear and high-quality aerial images. Aerial images are captured from a distance, which makes them vulnerable to problems like reduced visibility, changes in color and blurriness. These issues are often caused by factors like changes in the atmosphere, the presence of clouds, or fog. When visibility is reduced due to these factors, it becomes challenging to monitor and assess situations, especially in cases like disaster management. Therefore, there is a pressing need for a reliable method to enhance the visibility of aerial images, ensuring that they are clear and useful for various applications.

In the pursuit of this goal, prior studies have made notable contributions in the field. For instance, in reference [1], a correction technique was employed that leveraged the correlation between low and high-frequency colour bands. Additionally, Liu and colleagues [2] explored the use of the virtual cloud point method as a means to address haze-related issues, while Long et al. [3] harnessed the dark channel prior (DCP) concept initially introduced by He and his team [4] to effectively mitigate haze in natural scene images.

The rise of deep learning has significantly advanced research in aerial image restoration. Capitalizing on the generalization capabilities of convolutional neural networks (CNNs), researchers have introduced various methodologies, including conditional generative adversarial networks (cGAN), unsupervised learning [5], and channel refinement [6], to tackle aerial image dehazing. Meanwhile, transformers, which have gained substantial prominence for their ability to capture global dependencies in images, have spawned diverse architectures [7, 8] for tasks like deblurring, denoising, deraining, among others. However, it's worth noting that the application of transformers to address the challenge of haze degradation in aerial images remains an underexplored area.

Motivated to design robust deep learning model capable of dehazing aerial images for different applications we propose DerailNet based on transformers [9]. The main contributions of our work can be summarised as:

- An encoder-decoder architecture capable of processing aerial hazy images into dehazed aerial images.
- We proposed enhanced attention network capable of processing aerial hazy images in different frequency spectrum using wavelet transform.
- Lastly, we made use of enhanced skip connection network for rertaining the high frequency features which are prone to be lost in deep networks.

II. Literature Survey

Initially, efforts were focused on dehazing images using manually designed priors. In their work, He et al. introduced a fundamental haze-related prior to obtain coarse-level depth information for dehazing. However, this approach proved to be ineffective in sky regions and led to the presence of halo artifacts near complex edge structures[4,10,11]. Over the last decade, researchers have been working on using CNNs to calculate scene transmission maps and then applying atmospheric scattering models to restore haze-free images. Cai et al[12]. introduced a deep network for estimating the transmission map and then applied an atmospheric scattering model to reconstruct the haze-free image. Dong et al.[13] proposed a boosted decoder that employs only the reconstruction error, computed using ground truth data, as the supervision method to progressively obtain the haze-free image.

The research paper by Zhao et al. [14] states a two-stage framework with weak supervision that incorporates unpaired adversarial learning. More recently, Jia et al. presented a meta-attention-based network designed for restoring hazy images. Liu et al. proposed a method centered on multi-branch feature extraction to integrate all pertinent information and reconstruct haze-free images. Additionally, Chen et al. proposed a method involving the generalization of a pre-trained network on synthetic data for adaptation to real-world images[15,16]. Li et al. introduced an unsupervised learning-based approach that incorporates compact multiscale feature attention and multi-frequency representations. It's worth noting that these techniques have primarily been employed for hazy images captured at ground level[17].

Many approaches have been put forward, specifically in the context of dehazing aerial images. Zhang et al. used a correction technique to link low and high color bands. Liu et al. employed the virtual cloud point method to remove haze. Long et al. used the DCP method by He et al. for hazy scene image enhancement. Guo et al. utilized residual learning with channel attention modules to achieve fast network convergence and improve channel correlation[1,2,4]. The model[18] focus on clearing cloudy areas using special attention that works from small to large areas to remove both clouds and haze. Grohnfeldt and the team introduced a cGAN, a kind of network, that combines SAR and multispectral data to get rid of clouds. Additionally, Huang and the team used both RGB and SAR data to create a network with dilated convolutions for a similar purpose[19]. Current dehazing methods have gotten better, but they often forget about the need to deal with both local and global factors at the same time. This inspired our work to make dehazing more effective by considering both of these factors for a stronger solution.

III. Proposed Architecture

<u>Encoders and Decoders:</u>

Transformers are the robust architecture which are being used in computer vision for different tasks like deblurring[20], denoising[21], deraining[22,23], etc. These transformers generally don't process keys and queries, which are prominently responsible for calculation of attention map. The queries are the ones using which attention is calculated on keys. So, by observing the prominent role of queries, we processed queries to obtain the clear features from hazy image and increased the performance of the transformer.

The queries are processed using wavelet transform, which is generally used in traditional image-processing algorithms for image denoising. The wavelet transform transforms the image from spatial domain to frequency domain using high pass and lowpass filters. It transforms the image to compute wavelet coefficients for low and high frequency features.

In hazy image it is necessary to maintain the integrity of the frequency of image features (low frequency and high frequency) while transforming it from encoders to decoders. Thus, while processing queries and keys dephthwise convolution on each

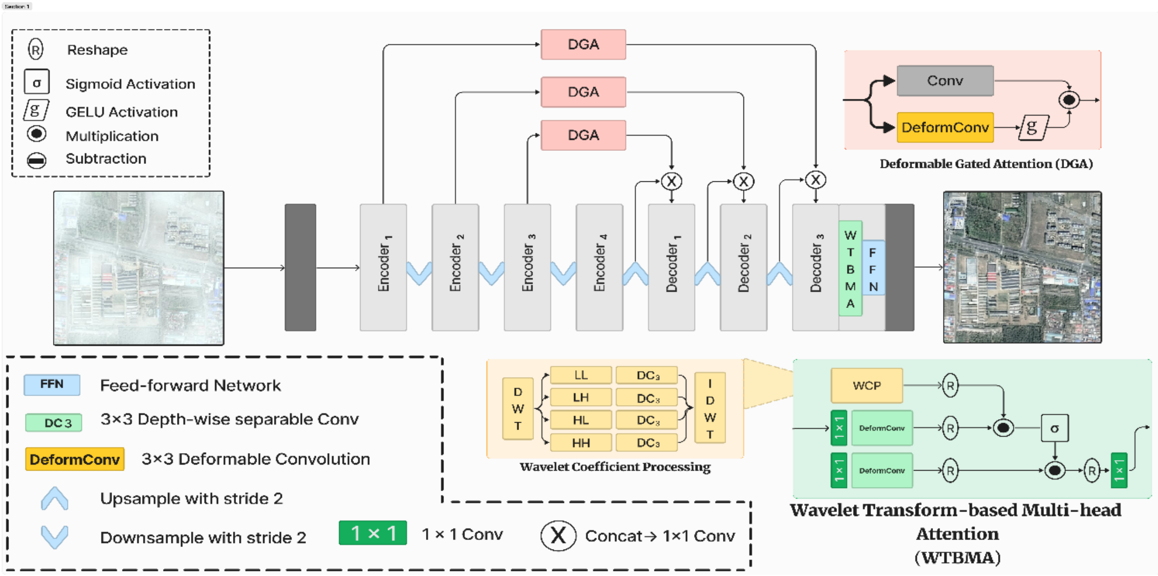

Figure 1 : Proposed Architecture

four coefficients (LL, LH, HH, HL) **(please refer fig [1] of proposed architecture)**. This process can be represented in equation as,

$$LL, LH, HH, HL = DWT(X_{in})$$

, where LL and LH are low frequency features obtained from low pass filter, similarly HH and HL are high frequency features obtained from high pass filter. DWT stands for discrete wavelet transform of input feature maps I_{in}.

The deformable convolution[24] helps in capturing the global features of the image more efficiently and maintain the geospatial adaptability. This can be summarised using following equation,

$$HH' = \phi(HH)$$

$$HL' = \phi(HL)$$

$$LH' = \phi(LH)$$

$$LL' = \phi(LL)$$

, where ϕ is the depthwise convolution and LL', LH', HL' and HH' are the convolved approximate, horizontal, vertical, and diagonal coefficients, respectively.

These convolved coefficients are then inversely transformed using inverse discrete wavelet transform to obtain feature maps in spatial domain.

The processed queries are used to compute attention map which is normalised using softmax function along the row. The attention map is then used by value tensor to calculate overall attention.

$$Attention(X_{in}) = \sigma(\frac{QK^T}{\sqrt{d}})V$$

, where Q and K are processed keys and queries, V is the value tensor, σ is the softmax function and V is the value tensor.

Finally, the processed input feature map is element-wisely added with the input feature map through skip connection, which can be summarised as,

$$\hat{X} = X_{in} + Attention(X_{in})$$

, where X_hat is the final processed output feature map.

<u>Deformable gated attention:</u>

In this work we are using encoder-decoder architecture in which the the feature maps are downsampled in encoder levels and the upsampled in decoder levels to obtain the original spatial dimension of input image. Because of these upsample and downsample operations the high frequency features of aerial image (like vehicles, buildings, etc) might be lost, so to retain these features skip connections are used to pass the features of certain encoder level to corresponding decoder level. In skip connections we are making use of deformable gated attention to further process

Figure 2: Qualitative comparison result on existing state of the art deep learning models on State1K Dataset

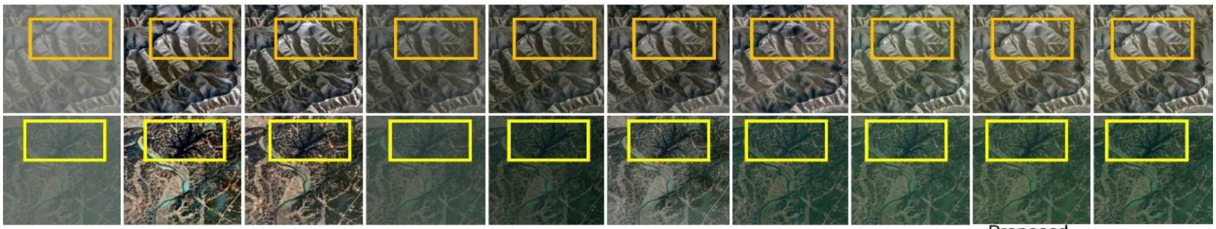

Figure 3: Qualitative comparison result on existing state of the art deep learning models on Rice Dataset

these high frequency features more efficiently. The whole computation can be represented as,

$$Dec_i = Conv(Enc_i) \cdot GELU(\phi(Enc_i))$$

where Conv is depthwise-convolution with kernel size = 3, GELU is the gelu activation function and Enc_i and Dec_i are the encoder and decoder, respectively of i^{th} level.

IV. Result Comparison

've methodically compared the outcomes of our proposed approach with established state-of-the-art methods to thoroughly examine its superior visual quality. We've provided a qualitative assessment of results for the Sate1K dataset in Figure 2, and a comparative analysis for the RICE dataset in Figure 3. In the domain of image restoration tasks, particularly de-hazing, it's well-known that there's a notable disparity between synthetic and real-world data. Consequently, we've evaluated the effectiveness of our method using genuine aerial images, and these authentic results have been showed in fig [2] and fig [3].

Evident from the prominently highlighted regions in these respective comparative illustrations is the impressive ability of our proposed method to effectively remove haze from aerial images. Simultaneously, it retains intricate textural intricacies, achieves a harmonious color balance, and elevates overall perceptual quality.

V. Conclusion

In our investigation, we have pioneered an innovative approach to enhance the quality of hazy aerial images. Our design involves a network that leverages efficient attention mechanisms to ensure the preservation of significant textures within the image. To further optimize our method, we have incorporated skip connections, which serve to enhance the edges in the image and facilitate the transmission of critical information from the image's surface to the deeper layers of our network. Comprehensive assessments of our technique were conducted across a variety of datasets, varying in both quantity and image quality. The results revealed that our approach delivers comparable performance to existing methodologies.

References:

1. Ying Zhang and Bert Guindon. Quantitative assessment of a haze suppression methodology for satellite imagery: Effect on land cover classification performance. IEEE Transactions on Geoscience and Remote Sensing, 41(5):1082–1089, 2003.
2. Changbing Liu, Jianbo Hu, Yu Lin, Shihong Wu, and Wei Huang. Haze detection, perfection and removal for high spatial resolution satellite imagery. International Journal of Remote Sensing, 32(23):8685–8697, 2011.
3. Jiao Long, Zhenwei Shi, Wei Tang, and Changshui Zhang. Single remote sensing image dehazing. IEEE Geoscience and Remote Sensing Letters, 11(1):59–63, 2013.
4. 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.
5. Aditya Mehta, Harsh Sinha, Murari Mandal, and Pratik Narang. Domain-aware unsupervised hyperspectral reconstruction for aerial image dehazing. In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pages 413–422, 2021.
6. Jianhua Guo, Jingyu Yang, Huanjing Yue, Hai Tan, Chunping Hou, and Kun Li. Rsdehazenet: Dehazing network with channel refinement for multispectral remote sensing images. IEEE Transactions on geoscience and remote sensing, 59(3):2535–2549, 2020.
7. Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)
8. Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 5728–5739, June 2022.
9. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.
10. Codruta O Ancuti, Cosmin Ancuti, Chris Hermans, and Philippe Bekaert. A fast semi-inverse approach to detect and remove the haze from a single image. In Asian Conference on Computer Vision, pages 501–514. Springer, 2010.
11. Raanan Fattal. Single image dehazing. ACM transactions on graphics (TOG), 27(3):1–9, 2008.
12. Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. IEEE Transactions on Image Processing, 25(11):5187–5198, 2016
13. Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 2157–2167, 2020.
14. Shiyu Zhao, Lin Zhang, Ying Shen, and Yicong Zhou. Refinednet: A weakly supervised refinement framework for single image dehazing. IEEE Transactions on Image Processing, 30:3391–3404, 2022
15. Tongyao Jia, Jiafeng Li, Li Zhuo, and Guoqiang Li. Effective meta-attention dehazing networks for vision-based outdoor industrial systems. IEEE Transactions on Industrial Informatics, 18(3):1511–1520, 2022
16. Ryan Wen Liu, Yu Guo, Yuxu Lu, Kwok Tai Chui, and Brij B. Gupta. Deep network-enabled haze visibility enhancement for visual iot-driven intelligent transportation systems. IEEE Transactions on Industrial Informatics, pages 1–1, 2022
17. Jiafeng Li, Yaopeng Li, Li Zhuo, Lingyan Kuang, and Tianjian Yu. Usid-net: Unsupervised single image dehazing network via disentangled representations. IEEE Transactions on Multimedia, pages 1–1, 2022.
18. Heng Pan. Cloud removal for remote sensing imagery via spatial attention

generative adversarial network. arXiv preprint arXiv:2009.13015, 2020

19. Claas Grohnfeldt, Michael Schmitt, and Xiaoxiang Zhu. A conditional generative adversarial network to fuse sar and multispectral optical data for cloud removal from sentinel2 images. In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, pages 1726–1729. IEEE, 2018.
20. Jie Xiao, Xueyang Fu, Aiping Liu, Feng Wu, and Zheng-Jun Zha. Image de-raining transformer. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022
21. Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 1833–1844, 2021.
22. Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 17683–17693, June 2022
23. Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5728– 5739, 2022
24. Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In Proceedings of the IEEE international conference on computer vision, pages 764–773, 2017.