Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 47 (2026), No. 7, pp. 2837–2845
DOI : 10.47974/JIOS-2406

Wavelet -assisted efficient Swin Transformer network for image dehazing

Rahul Vishnoi †
Alka Verma *
Vibhor Kumar Bhardwaj §
Department of Electronics & Communication (EC) Engineering
College of Engineering
Teerthanker Mahaveer University
Moradabad 244001
Uttar Pradesh
India

Abstract

Image dehazing increases the clarity of the visual aspect of a driver using a vehicle; when surveilling; and during consumer photographic use. Dehazing using CNN and the physics-based method has no method of understanding the entire global image scene; and both dehazing methods fail to perform in non-homogeneous haze setup, and finally transformer-based methods require either very large scale datasets or lots of processing power to accomplish their tasks. This research presents a unified framework for single-image dehazing named Physically Guided Wavelet Swin Network (PhyWave-Swin). PhyWave-Swin combines the necessary components of physical interpretability, multifrequency analysis and attention driven global modelling to advance this area of work. PhyWave-Swin was developed using the RESIDE dataset and underwent evaluation for both synthetic and real-world test data. The resulting performance metrics for PhyWave-Swin are 32.84/0.946 PSNR/SSIM respectively for SOTS-Indoor and 30.29/0.939 PSNR/SSIM respectively for SOTS-Outdoor test data; and PhyWave-Swin produces considerably better perceptual image quality versus other dehazing work that produces NIQE=3.65 and BRISQUE=33.9 in real-world scenarios. Therefore; a combination of physics, wavelet frequency representation and transformer-based global modelling provide the PhyWave-Swin with a very effective solution to single-image dehazing.

Subject Classification: *68T07, 68T45, 68U10, 94A08.*

Keywords: *Image dehazing, Convolutional neural network, Physically guided normalization, Deep image restoration.*

† *E-mail:* **rahulv.engineering@tmu.ac.in**
* *E-mail:* **dralka.engineering@tmu.ac.in (Corresponding Author)**
§ *E-mail:* **vibhor.engineering@tmu.ac.in**

1. Introduction

One of the most important steps in the processing of images from autonomous vehicles, surveillance systems, and consumer photography is to return the image clarity of those images that have been degraded by haze. When haze is removed from a haze-contaminated image, the resulting pixel similarity and statistical distributions in the image are restored so that the interpretation of that image can be completed with greater accuracy [1]. Many methods have been developed for de-hazing images. These methods include traditional convolutional neural networks (CNN) [2] developing models based upon the physics of transmission through the atmosphere [3], or more recently, using transformers [4]. All of those methods have some difficulty when it comes to processing images that have significant haze conditions. Traditional CNNs fail to represent the global characteristics of haze due to their restricted receptive fields and lack of global reasoning. Physics-based models, on the other hand, struggle to represent a complex image due to the restrictions placed on that image because of the assumption of uniformity in the distribution of haze and assumptions of a simplified atmosphere — both of which are not necessarily true. Transformers such as vision transformers (ViT) correct the limitations of CNNs in modeling from a global perspective [5], but they typically require very large and computationally intensive data to train and still may have difficulties representing the fine details of the pixels and texture information. Finally, both graph-based dehazing models [6] and classical dehazing models [7] appear to limit their effectiveness in dealing with complex haze conditions.

Current methods [8, 9, 10] typically address these aspects separately leading to suboptimal results where local details or global context are inadequately recovered. Moreover, many models [11, 12, 13] suffer from overfitting because of a lack of physical interpretability which makes them less reliable in practice. Recently, an unified architecture called the Wavelet-Based Physically Guided Normalization Dehazing Network (WBPGNDN) was proposed [14], which combined the strengths of physics-based constraints with advanced deep learning modules. This method introduces a physically guided normalization [15] and feature-domain wavelets to restore pixel similarity and capture the multi-scale context, but it sometimes oversimplifies the global structure to preserve semantic content. Although this method produces promising results, it lacks explicit global modelling and does not fully capture long-range dependencies. Therefore, the authors presented an more enhanced model by integrating Swin Transformer blocks per wavelet sub-band to compensate missing global information. The proposed model PhyWave-Swin enables an efficient long-range dependency capture while preserve the textures in image. The fusion of physics frequency and attention yields a more superior dehazing fidelity, especially in large haze regions and real-world scenes.

2. Theoretical Concept and Proposed Model

Haze plays a pivotal role in dehazing of image due to a physical formation process which involves atmospheric light and depth-dependent transmission. To achieve stability and restoration such physical cues are implemented into learning-based models The proposed Physically Guided Normalization (PGN) normalization adds this physical insight straight into the process of feature normalization. The equation 1 is built on the atmospheric scattering model.

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{1}$$

Haze is responsible for creating in image features additive bias and contrast attenuation which results in shifting of mean and reduction of variance in deep feature space. Therefore, PGN performs prior-guided normalization to achieve Haze free information. Equation 2 shows how features normalization is achieved.

$$\hat{F} = \gamma(\pi)\frac{F - \mu(F)}{\sigma(F) + \varepsilon} + \beta(\pi) \tag{2}$$

where $\gamma(\pi)$ and $\beta(\pi)$ are prior-guided parameters which dynamically recompense the effects of transmission and airlight.

Figure 1 shows the unified dehazing framework. The framework synergistically addresses two key challenges. These include long-range dependency modelling and haze-induced pixel dissimilarity. The proposed method follows an encoder-wavelet-transformer-normalization-decoder model called the PhyWave-Swin. The proposed method follows the following steps to dehaze an image. It begins with extracting shallow features from the hazy input, then decomposes them into multi-frequency sub-bands using the wavelet transform, applies Swin Transformer blocks independently to each sub-band for long-range modelling, adaptively fuses cross-band information, and finally restores pixel similarity using a physics-guided normalization layer before decoding to the clean output. PGN generates stable results with convergence being consistent and produces numerically stable outputs and consistent convergence throughout various datasets.

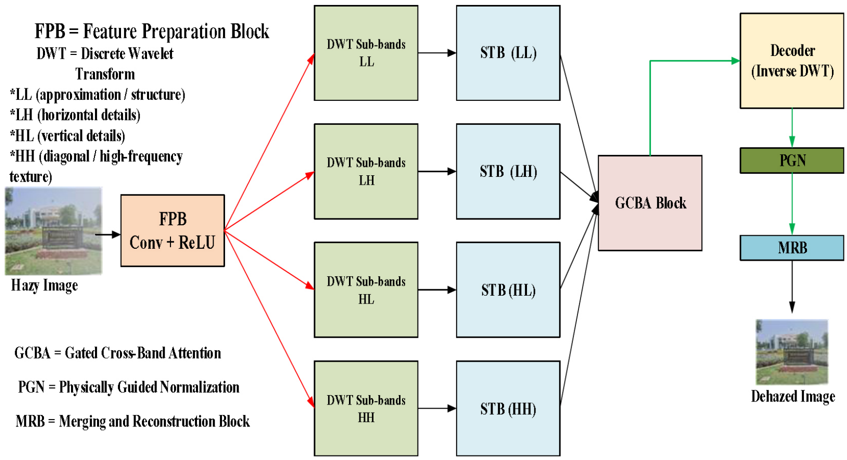


Figure 1

Simplified Architecture of the proposed PhyWave-Swin.

3. Results and Discussions

The model was trained on the RESIDE-Standard indoor dataset (13,990 hazy-clean pairs) and evaluated on synthetic (SOTS-Indoor/ Outdoor) as well as on real-world datasets (I-Haze, O-Haze, D-Hazy, and NH-Haze 2.0) [16, 17, 18, 19] to ensure its performance across diverse haze conditions. With PyTorch on an RTX 3090 GPU using a 4-layer encoder-decoder with Haar wavelets and Swin-based attention fusion the experiment was performed. Training used AdamW optimization with cosine learning rate decay and multi-term loss balancing for stable convergence and efficient dehazing.

Table 1 report the quantitative results for the synthetic SOTS-Indoor and SOTS-Outdoor benchmarks, demonstrating the superior performance of the proposed PhyWave-Swin model computing PSNR (in decibels) and SSIM. Both above metrics along with ± standard deviation (Std) suggests that model is stable depicting consistency too.

Table 1

Quantitative Results of Synthetic Benchmarks (SOTS).

Method	SOTS-Indoor		SOTS-Outdoor	
	PSNR	SSIM	PSNR	SSIM
DehazeNet	21.14 ± 0.17	0.847 ± 0.002	22.19 ± 0.18	0.832 ± 0.006
AOD-Net	22.29 ± 0.21	0.862 ± 0.004	23.87 ± 0.17	0.861 ± 0.005
FFA-Net	30.23 ± 0.15	0.932 ± 0.003	28.42 ± 0.15	0.913 ± 0.006
MSBDN	31.08 ± 0.13	0.939 ± 0.002	29.31 ± 0.11	0.924 ± 0.002
GridDehazeNet	31.48 ± 0.14	0.942 ± 0.003	29.56 ± 0.12	0.926 ± 0.003
DehazeFormer-T	32.17 ± 0.11	0.946 ± 0.003	30.01 ± 0.14	0.930 ± 0.002
Uformer-T	32.39 ± 0.14	0.947 ± 0.002	30.24 ± 0.13	0.932 ± 0.003
ESTINet	31.82 ± 0.17	0.944 ± 0.003	29.87 ± 0.12	0.929 ± 0.002
WBPGNDN	32.67 ± 0.12	0.945 ± 0.003	30.21 ± 0.11	0.938 ± 0.002
Proposed work	**32.84 ± 0.09**	**0.946 ± 0.002**	**30.29 ± 0.10**	**0.939 ± 0.003**

Furthermore, on real-world datasets i.e., I-Haze, O-Haze, D-Hazy and NH-Haze, the proposed method also demonstrates a superior generalization capability though certain edge cases remain challenging. Table 2 summarizes the qualitative assessment results of the presented method.

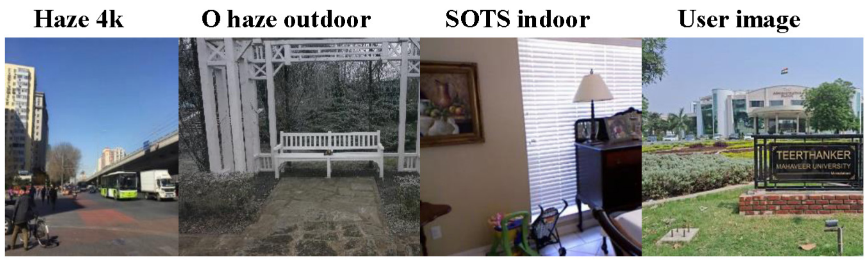
Table 2

Quantitative Results of Real-World Benchmarks.

Method	I-Haze (PSNR)	O-Haze (PSNR)	D-Hazy (NIQE)	NH-Haze (BRISQUE)
FFA-Net	20.31	19.87	4.32	38.70
MSBDN	21.05	20.43	4.18	37.20
DehazeFormer-T	21.42	20.81	3.97	35.80
Uformer-T	21.63	20.95	3.89	35.10
ESTINet	21.38	20.76	4.02	36.40
WBPGNDN	21.67	20.86	4.10	34.14
Proposed work	22.19	21.67	3.65	33.90

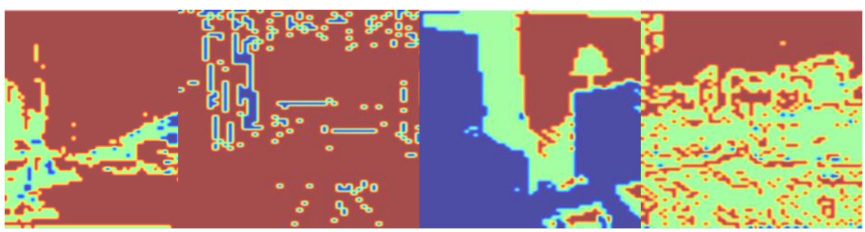
Qualitative assessment as depicted in Figure 2, was performed on the datasets proposed which focussed on identification failures cases. However severe failures were not reported, only small inconsistencies appeared on certain conditions which included severe dense haze, texture being low and regions being highly cluttered. Analysis of Attention map and ablation study predicts model stability and robustness.

Figure 2 Qualitative Assessment of the PhyWave-Swin Method for Haze Removal and Detail Recovery.

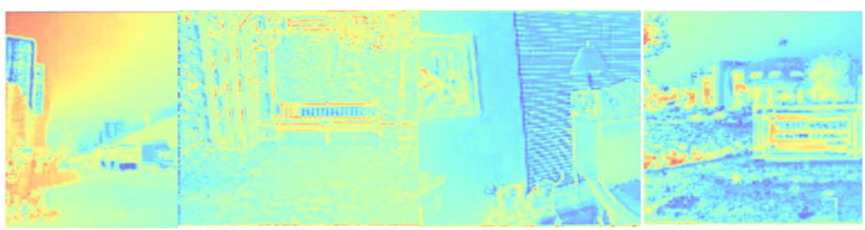


(a) Original Images

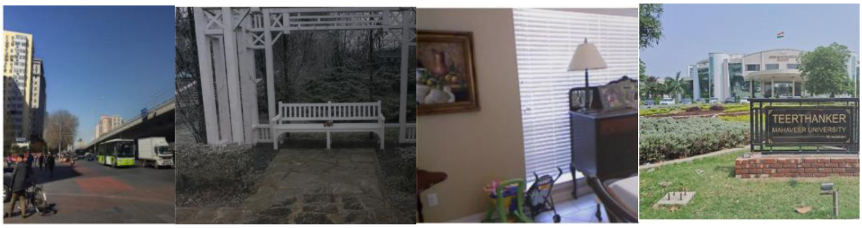
(b) Hazy Images



(c) Cross-attention map



(d) Attention maps



(e) Dehazed images

Figure 2

Qualitative Assessment of the PhyWave-Swin Method for Haze Removal and Detail Recovery.

Table 3 depicts the summary of the ablation study of the proposed model. A paired t-test was performed between the wavelet-only variant (+ Feature-Domain DWT, 30.17 dB) and the wavelet + Swin variant (+ Swin per Sub-band, 31.52 dB) on per-image PSNR values and it was presented that $p < 0.01$, resulting in increase of 1.35 dB which is statistically substantial. This predicts that the Swin transformer results in good performance boost when implemented to wavelet features.

Table 3

Summary of the Ablation Study

Variant	PSNR	SSIM	Δ PSNR
Baseline CNN (no wavelet, no Swin)	28.91	0.912	-
+ Feature-Domain DWT	30.17	0.926	+1.26
+ Swin per Sub-band	31.52	0.938	+1.35
+ Cross-Band Fusion (GCBA)	32.24	0.945	+0.72
+ Physically Guided Norm	33.05	0.951	+0.81
Proposed work	**33.05**	**0.951**	**-**

4. Conclusion

This study introduces PhyWave-Swin, a novel wavelet–transformer–physics integrated framework that effectively restore haze-degraded images. By combining multifrequency feature decomposition with Swin Transformer-based long-range modelling and physics-guided normalization, the proposed network achieves superior restoration fidelity and perceptual realism across both synthetic and real-world datasets and also demonstrate strong potential for deployment in practical vision systems including autonomous driving and outdoor imaging applications.

References

[1] W. Wang and X. Yuan, "Recent advances in image dehazing," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 3, pp. 410–436 (Jul. 2017).

[2] Y. Song, J. Li, X. Wang, and X. Chen, "Single image dehazing using ranking convolutional neural network," *IEEE Transactions on Multimedia*, vol. 20, no. 6, pp. 1548–1560 (Jun. 2018).

[3] J. Dong and J. Pan, "Physics-based feature dehazing networks," in *Computer Vision – ECCV 2020 Workshops*, Cham, Switzerland: Springer, pp. 188–204 (2020).

[4] Y. Wang, X. Yan, M. Wei, and J. Xiong, "USCFormer: Unified transformer with semantically contrastive learning for image dehazing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 10, pp. 11321–11333 (Oct. 2023).

[5] Y. Song, X. Du, H. Qian, and Z. He, "Vision transformers for single image dehazing," *IEEE Transactions on Image Processing*, vol. 32, pp. 1927–1941 (2023).

[6] Y. Yeh, L.-W. Kang, and C.-Y. Lin, "Haze effect removal from image via haze density estimation in optical model," *Optics Express*, vol. 21, no. 22, pp. 27127–27141 (2013).

[7] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353 (Dec. 2011).

[8] K. Ding and T. P. Chen, "Study on damage detection of bridge based on wavelet multiscale analysis," *Advanced Materials Research*, vol. 639–640, pp. 1010–1014 (2013).

[9] T. Tao, X. Guan, H. Zhou, and H. Xu, "MCADNet: A multi-scale cross-attention network for remote sensing image dehazing," *Mathematics*, vol. 12, no. 23, Art. no. 3650 (Nov. 2024).

[10] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," *arXiv preprint arXiv:2103.14030* (Mar. 2021).

[11] I. Mitov, B. Depaire, K. Vanhoof, and K. Ivanova, "Classifier PGN: Classification with high confidence rules," *Serdica Journal of Computing*, vol. 7, no. 2, pp. 143–164 (2013).

[12] G. Chen, Y. Jia, Y. Yin, S. Fu, D. Liu, and T. Wang, "Remote sensing image dehazing using a wavelet-based generative adversarial network," *Scientific Reports*, vol. 15, no. 1 (2025).

[13] S. Thakral and P. Manhas, "Image processing using different types of discrete wavelet transform," in *Proceedings of the International Conference on Advanced Computing, Communication and Informatics*, Singapore: Springer, pp. 499–507 (2018).

[14] S. Zhang, L. Shen, W. Ren, S. Wan, and X. Zhang, "Wavelet-based physically guided normalization network for real-time traffic dehazing," *Pattern Recognition*, vol. 172, Art. no. 112451 (Apr. 2026).

[15] L. Qi, H. Yang, X. Geng, and Y. Shi, "NormAUG: Normalization-guided augmentation for domain generalization," *IEEE Transactions on Image Processing*, vol. 33, pp. 1419–1431 (2024).

[16] C. Ancuti, C. O. Ancuti, R. Timofte, and C. De Vleeschouwer, "I-HAZE: A dehazing benchmark with real hazy and haze-free indoor images," in *Proc. 19th Int. Conf. Advanced Concepts for Intelligent Vision Systems (ACIVS)*, Lecture Notes in Computer Science, vol. 11182, Cham, Switzerland: Springer, pp. 620–631 (2018), doi: 10.1007/978-3-030-01449-0_52.

[17] C. Ancuti, C. O. Ancuti, and C. De Vleeschouwer, "D-HAZY: A dataset to evaluate quantitatively dehazing algorithms," in *Proc. IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, pp. 2226–2230 (Sep. 2016).

[18] C. O. Ancuti, R. Timofte, and C. Ancuti, "NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 444–453 (Jun. 2020).

[19] Z. Wang, X. Cun, J. Bao, and J. Liu, "Uformer: A general U-shaped transformer for image restoration," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17662–17672 (Jun. 2022).

Received December, 2025