This CVPR Workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

DehazeDCT: Towards Effective Non-Homogeneous Dehazing via Deformable Convolutional Transformer

Wei Dong[1] Han Zhou[1] Ruiyi Wang[2] Xiaohong Liu[2]* Guangtao Zhai[2] Jun Chen[1]
[1]McMaster University [2]Shanghai Jiao Tong University
wdong1745376@gmail.com, zhouh115@mcmaster.ca
{thomas25, xiaohongliu, zhaiguangtao}@sjtu.edu.cn, chenjun@mcmaster.ca

Figure 1. Test Result of our method on NTIRE 2024 Dense and Non-Homogeneous Dehazing Challenge [5]. Our **DehazeDCT** achieves the **second best** performance among 16 solutions and is capable to generate visually compelling outputs with vivid color and enhanced structure details.

Abstract

Image dehazing, a pivotal task in low-level vision, aims to restore the visibility and detail from hazy images. Many deep learning methods with powerful representation learning capability demonstrate advanced performance on non-homogeneous dehazing, however, these methods usually struggle with processing high-resolution images (e.g., 4000×6000*) due to their heavy computational demands. To address these challenges, we introduce an innovative non-homogeneous* ***Dehaz****ing* ***m****ethod via* ***D****eformable* ***C****onvolutional* ***T****ransformer-like architecture (****DehazeDCT****). Specifically, we first design a transformer-like network based on deformable convolution v4, which offers long-range dependency and adaptive spatial aggregation capabilities and demonstrates faster convergence and forward speed. Furthermore, we leverage a lightweight Retinex-inspired transformer to achieve color correction and structure refinement. Extensive experiment results and highly competitive performance of our method in NTIRE 2024 Dense and Non-Homogeneous Dehazing Challenge, ranking second among all 16 submissions, demonstrate the superior capability of our proposed method. The code is available:* https://github.com/movingforward100/Dehazing_R.

1. Introduction

Images captured in hazy conditions, whether naturally occurring or artificially synthesized, share similar properties

* Corresponding author

of low visibility, decreased contrast, and degraded structural details. These deteriorative characteristics severely impair the performance of various vision tasks, such as object recognition, tracking, and segmentation systems [12, 16, 32, 44], thereby hindering their application in hazy situations. This predicament urgently calls for dehazing methods that are effective across various scenarios.

Initiated upon the foundational principles of the atmospheric scattering model (ASM) [27], early endeavors in image restoration [19, 20, 46] have been primarily oriented towards delineating the correlation between hazy images and their haze-free counterparts. The ASM can be succinctly articulated as follows:

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

where I and J signify the hazy image and its clean counterpart; x and A represents the pixel location and the global atmosphere light; $t(x)$ denotes the transmission map , which is a function of the atmosphere scattering parameter β and the scene depth, articulated as $t(x) = e^{-\beta d(x)}$. This formulation posits the image dehazing to the precise estimation of the transmission map $t(x)$ and the global atmosphere light A. Notwithstanding, the ASM presupposes an idealized uniform haze distribution, a limitation rendering the model less effective in addressing non-homogeneous dehazing challenges.

Recent advancements in image dehazing have been significantly influenced by the application of deep learning techniques [6, 24, 25, 30, 41, 42, 48], a development prompted by their profound success in areas such as classification and object detection. Notably, compared to the previously dominant ASM framework, deep learning-based methods have demonstrated superior performance on removing the haze from images with complex and spatially varying haze distributions, emerging as the predominant approach for tackling non-homogeneous dehazing problems.

Recent research has predominantly focused on exploring robust and powerful representation learning mechanisms to delineate the mappings between hazy and haze-free images. These methods usually learn spatial and frequency representations simultaneously, integrate special architectures to increase the receptive field, or utilize large scale CNN networks pre-trained on large datasets to harness transfer learning benefits. For example, DWT-FFC [51] entails Discrete Wavelet Transform to capture spatial and spectral information effectively, employs Fast Fourier Convolution (FFC) [8, 34] to extend the receptive capacity, and harnesses the pre-trained ConvNext model to facilitate transfer learning; DehazeFormer [33] introduces a transformer-based architecture with a shifted window partitioning scheme based on reflection padding for dehazing.

Nonetheless, current methods encounter certain challenges that necessitate further exploration: **First**, traditional Convolutional Neural Networks (CNNs) often suffer from strict inductive bias, and CNN-based models frequently rely on sizable fixed dense kernels (e.g., 31×31) [11, 23] to facilitate robust representation learning. This strategy not only incurs significant computational loads but also lacks the capacity for adaptive spatial aggregation conditioned by the input. **Second**, while transformer-based architectures are capable to capture long-range dependencies and facilitate adaptive spatial aggregation, they are hampered by computational and memory inefficiencies. The inherent complexity of the self-attention mechanism, which scales quadratically with input resolution, precludes their application in high-resolution dehazing scenarios, exemplified the 4000×6000 resolution image in DNH-HAZE dataset [5].

To tackle these challenges, we introduce **DehazeDCT**, a novel non-homogeneous **Dehaz**ing m**e**thod via **D**eformable **C**onvolutional **T**ransformer architecture. This model is comprised of two primary components: a dehazing module and a refinement module. In the **Dehazing** module, we engineer transformer-like dehazing branch that incorporates multiple DCNFormer blocks. Diverging from traditional Transformer architecture, our DCNFormer blocks utilize Deformable Convolution v4 [43] in lieu of the standard self-attention mechanism, thus ensuring our model benefits from both long-range dependency and adaptive spatial aggregation capabilities. Furthermore, by removing the redundant operation (softmax normalization in spatial aggregation) in traditional DCN [36], our DCNFormer architecture demonstrates faster convergence and forward speed. Besides, we integrate a frequency-aware branch to facilitate the acquisition of frequency representations. In the **Refinement** module, we enact a lightweight Retinex-inspired transformer for color correction and structure refinement. With these dedicated designs, our **DehazeDCT** is capable to achieve color and detail consistency, essential for visually compelling results as shown in Fig. 1.

Our mian contributions are as follows:

◇ We propose an effective non-homogeneous dehazing method based on deformable convolutional transformer, followed by a Retinex-based transformer network for color and detail refinement.

◇ We design a DCNv4 based transformer-like network for dehazing, which offers long-range dependency and adaptive spatial aggregation capabilities and demonstrates faster convergence and forward speed.

◇ The effectiveness and generalization of our proposed method are verified by comprehensive experiment results on several benchmark datasets and the visually compelling performance in the NTIRE 2024 Dense and Non-Homogeneous Dehazing Challenge.

2. Related Work

Frequency Based Image Restoration. Similar to the spatial domain, the frequency domain encapsulates abundant

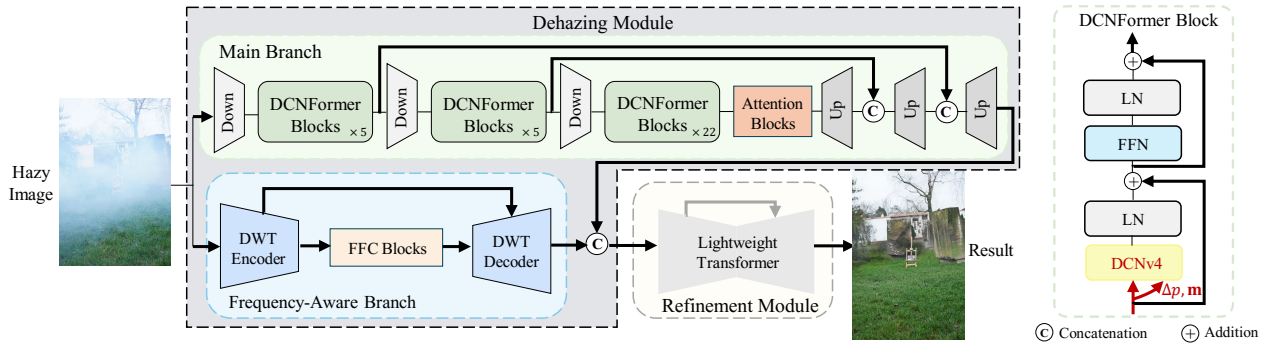

Figure 2. The overall architecture of our proposed model. In the Dehazing module, we introduce a transformer-like dehazing branch based on deformable convolution (DCNv4 [43]). In each DCNFormer block, DCNv4 is utilized to calculate the offset (Δp) and modulation scalar ($\mathbf{m}$). Besides, the frequency-aware branch proposed in [51] is also adopted as an auxiliary branch. In the Refinement module, we leverage a lightweight retinex-inspired transformer network to further reduce the color deviation and enhance texture details.

information of images. In recent years, growing research attention has been drawn to take advantage of frequency information for image restoration [9, 18, 35, 45, 47]. Specifically, Yoo *et al*. [47] estimate the DCT coefficient for each frequency band for JPEG compression artifact removal. Yang *et al*. [45] propose a wavelet U-Net where the downsampling and up-sampling are replaced by discrete wavelet transform and its inverse operation for better reconstruction of edges and colors. Jiang *et al*. [18] introduce a novel focal frequency loss considering both amplitude and phase information and a dynamic spectrum weighting to adaptively guide existing models for frequency domain reconstruction. Cui *et al*. [9] design a multi-branch dynamic selective frequency module (MDSF) which dynamically decomposes features into several frequency bands and uses channel-wise attention to highlight the useful frequency for image restoration.

Image Restoration with Deformable Convolution. To overcome the limitation of traditional CNNs in modeling large and unknown transformations, deformable convolution [10] is proposed. By adding offsets to the regular sampling positions, deformable convolution allows flexible form of the sampling grid. Deformable convolutional networks have achieved remarkable performance on various high-level vision tasks [36, 43, 52]. Recently, several works have devoted to apply deformable convolution for image restoration. Wang *et al*. [37] design a pyramid, cascading and deformable (PCD) alignment module, where deformable convolution is applied for feature-level frame alignment, for video restoration. Wu *et al*. [42] deploy two deformable convolutional layers after the deep layers of the autoencoder-like network for image dehazing. Wang *et al*. [38] propose an angular deformable alignment module (ADAM) where the deformable convolution is used to align features by their corresponding offsets.

Transformer Based Image Dehazing. Due to the flexibility of capturing global dependencies, transformers have succeeded in various computer vision tasks. For instance, Uformer [40], SwinIR [21] and Restormer [49] achieve remarkable performance on many image restoration tasks. Recently, multiple methods have applied transformers to single image dehazing. Based on Swin Transformer [26], Song *et al*. [33] propose DehazeFormer with modified normalization layer, activation function and spatial information aggregation method. Guo *et al*. [17] combine transformer and CNN for image dehazing through transmission-aware 3D position embedding and feature modulation. Qiu *et al*. [29] apply Taylor expansion to approximate the conventional softmax attention in transformer, which achieved linear complexity while retained the flexibility. These works demonstrate state-of-the-art performance on various dehazing benchmarks and inspired us to incorporate transformer into our model.

3. Methods

The key contribution of our work is to design a novel dehazing module based on DCNv4 [43] (Deformable Convolution v4) and adopt a lightweight transformer for color and detail enhancement (Refinement module). The overview of our framework is provided as Fig. 2, where the training process can be divided into two stages. In stage I, we optimize the Dehazing module to achieve preliminary dehazing (Sec. 3.1, Sec. 3.2). In stage II, the Refinement module is incorporated into the optimization process for detailed refinement. (Sec. 3.3).

3.1. DCNv4 based Transformer for Dehazing

Inspired by impressive ability of capturing long-range dependencies and facilitating adaptive spatial aggregation, we design a transformer-like branch, together with a frequency-aware branch similar to [13, 51], for effective dehazing, as shown in Fig. 2. Specifically, an hazy input $\mathbf{I} \in \mathbb{R}^{W \times H \times 3}$ is encoded into $\mathbf{F}_i \in \mathbb{R}^{\frac{W}{4*2^i} \times \frac{W}{4*2^i} \times d_i}$ by i downsampling

Methods	NH-HAZE [1]		NH-HAZE2 [3]		HD-NH-HAZE [4]		HD-NH-HAZE2 [5]	
	PSNR ↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
FFA [28]	19.50	0.644	20.56	0.811	20.23	0.710	20.14	0.707
TDN [22]	20.73	0.673	20.44	0.801	20.06	0.713	19.88	0.700
AECR-Net [42]	19.88	0.717	20.75	0.831	20.34	0.731	20.26	0.724
DWT-FFC [51]	22.64	0.730	22.82	0.874	22.20	0.746	21.58	0.738
DehazeFormer [33]	20.47	0.731	21.07	0.825	20.89	0.728	20.29	0.718
DehazeDCT (Ours)	**22.78**	**0.734**	**22.86**	**0.877**	**22.36**	**0.752**	**21.73**	**0.743**

Table 1. Quantitative comparisons between our proposed DehazeDCT and SOTA methods. Our proposed method achieves superior performance in terms of PSNR and SSIM across four datasets. These numbers are obtained from their original paper or training with their released code. [Key: **Best**, Second Best, ↑ (↓): Larger (smaller) values leads to better performance, HD-NH-HAZE2: Official dataset for NTIRE 2024 Dense and Non-Homogeneous Dehazing Challenge]

operations, where W, H, and d_i represent the image width, image height, the dimension of latent features. After each downsampling process, several DCNFormer blocks, which share similar architecture with common transformer blocks, are adopted for representation learning. However, instead of the global attention mechanism of transformers, the core operator of our DCNFormer is the deformable convolution v4, which is achieved by removing the softmax normalization operation in DCNv3 [36].

Given each latent feature $\mathbf{F}_i$ and current pixel p_0, the principle of DCNv3 operation is described as:

$$\mathbf{y}(p_0) = \sum_{g=1}^{G}\sum_{k=1}^{K} \mathbf{w}_g \mathbf{m}_{gk} \mathbf{x}_g(p_0 + p_k + \Delta p_{gk}), \quad (2)$$

where K represents the total number of sampling points, k enumerates the sampling point and G denotes the total number of aggregation groups. For the g-th group, $\mathbf{w}_g \in \mathbb{R}^{d_i \times d_i'}$ denotes the location-irrelevant projection weights of the group, where $d_i' = d_i/G$ represents the group dimension. $\mathbf{m}_{gk} \in \mathbb{R}$ denotes the modulation scalar of the k-th sampling point in the g-th group, normalized by the softmax function along the dimension K. $\mathbf{x}_g$ represents the sliced input feature map. Δp_{gk} is the offset corresponding to the grid sampling location p_k in the g-the group.

Unlike utilizing softmax operation to normalize the modulation scalar $\mathbf{m}_{gk}$, we don't incorporate any normalization functions in our DCNFormer in order to achieve unbounded dynamic weights for $\mathbf{m}_{gk}$, which contributes to the significantly faster converge and forward speed compared to DCNV3 [36], common convolutions, and attention blocks in Transformers. Moreover, the core operator in our DCNFormer only adopts a 3×3 kernel to learn long-range dependencies, which is easier to be optimized compared to large kernels [11, 23]. Please note despite the comparatively large parameters of our dehazing module, our DCNFormer blocks allow our model to achieve high-resolution dehazing efficiently, without the need for special design of high-resolution images (e.g., 4000×6000).

3.2. Loss Function

The loss function utilized for optimizing our Dehaizng module is:

$$L_{loss} = L_1 + \alpha L_{SSIM} + \beta L_{Percep} + \gamma L_{adv}, \quad (3)$$

where L_1, L_{SSIM} and L_{Percep} represent $L1$ loss, MS-SSIM loss [51], and perceptual loss [31], respectively. In addition, we adopt the discriminator in [15] to calculate adversarial loss (L_{adv}). α, β, and γ are hyper-parameters and are set to 0.4, 0.01, and 0.0005, respectively.

3.3. Transformer based Refinement

Based our Dehazing module, which concentrates on removing haze from degraded images, we further incorporate a lightweight transformer similar to [7, 14] for refinement. Based on Retinex theory, given the input and its mean values for each pixel along the channel dimension, our Refinement module first predicts the lit-up image and light-up feature and then restore the color and detail corruption. Therefore, the our final dehazed image can be obtained by:

$$\mathbf{I}_{dehazed} = \boldsymbol{\phi}(\boldsymbol{\theta}(\mathbf{I}_{hazy}), mean(\boldsymbol{\theta}(\mathbf{I}_{hazy}))), \quad (4)$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ represent our Dehazing module and Refinement module, respectively. Notably, our Refinement module only has 1.61 M parameters, and we remove the adversarial loss in Eq. 3 to optimize the Refinement module $\boldsymbol{\phi}$.

4. Experiments

4.1. Experiment Settings

Datasets. We qualitatively and quantitatively evaluate our proposed method on four real-world datasets: NH-HAZE [1], NH-HAZE2 [3], HD-NH-HAZE [4] and DNH-HAZE2 [5] datasets. NH-HAZE dataset composes of 55 pairs of 1200×1600 hazy and corresponding clean images. We use the official testing data for evaluation while the remaining are utilized for training. NH-HAZE2 dataset consists of 25 training pairs, 5 validation pairs and 5 testing

Figure 3. Visual comparisons on NH-HAZE [1] dataset. Compared to other models, our method exhibits higher color fidelity and effective dehazing, yielding compelling results.

pairs with resolution 1200×1600. As the ground truth for validation and testing data isn't publicly accessible, we utilize the first 20 training pairs for training and use the rest as testing samples. HD-NH-HAZE and DNH-HAZE datasets compose of 40 training pairs, 5 validation pairs and 5 testing pairs of 4000×6000 images. As we can only access the ground truth of training pairs, we again evaluate on the last 5 training pairs and use the rest for training.

Implementation Details. We implement the training using Pytorch 1.11.0 on an NVIDIA RTX 4090 GPU. To augment the limited training data, images pairs are randomly cropped into patches of size 384×384, then possibly rotated at 90, 180, or 270 degrees, vertically or horizontally flipped. The training process of our proposed method composes of two stages. In stage I, we neglect the Refinement module. The model and discriminator are updated by Adam optimizer with decay factors $\beta_1 = 0.9$ and $\beta_2 = 0.999$ for a total of 5,000 epochs. The initial learning rate is set to be 1×10^{-4} and the learning rate decays by half at 1,500, 3,000, and 4,000 epochs. In stage II, we train our the Refinement module for 500 epochs and optimize our whole model for 500 epoch with the fixed the learning rate of 1×10^{-5}.

4.2. Comparisons with SOTA Methods

Compared Methods and Evaluation Metrics. In this section, we undertake a comprehensive evaluation of our method by quantitatively and qualitatively comparing it with current SOTA methods for dehazing. These benchmark models include the winner solution of NTIRE 2020 Non-Homogeneous Dehazing Challenge (TDN [2, 22]), the champion method in NTIRE 2023 HR Non-Homogeneous Dehazing Challenge (DWT-FFC [51]), as well as FFA [28], AECR-Net [42], and the recently proposed transformer-based dehazing approach (DehazeFormer [33]). Besides, we utilize two full-reference metrics for quantitative evaluation: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM [39]), which measures the pixel-level accuracy and the structural similarity of dehazed results.

Quantitative Comparisons. Tab. 1 presents the performance comparisons between our DehazeDCT and other methods across various datasets. As illustrated in Tab. 1, our DehazeDCT consistently demonstrates superior performance in terms of both PSNR and SSIM across all four datasets, with 21.73 dB PSNR and 0.743 SSIM on DNH-HAZE dataset. In particular, DehazeDCT surpasses the second best method by an average of 0.13 dB PSNR and 0.005 SSIM, underscoring the impressive capability of our proposed method.

Qualitative Comparisons. The visual comparisons of the dehazing results from Fig. 3, 4, and 5 demonstrate that our method achieves superior dehazing effects, capable of pro-

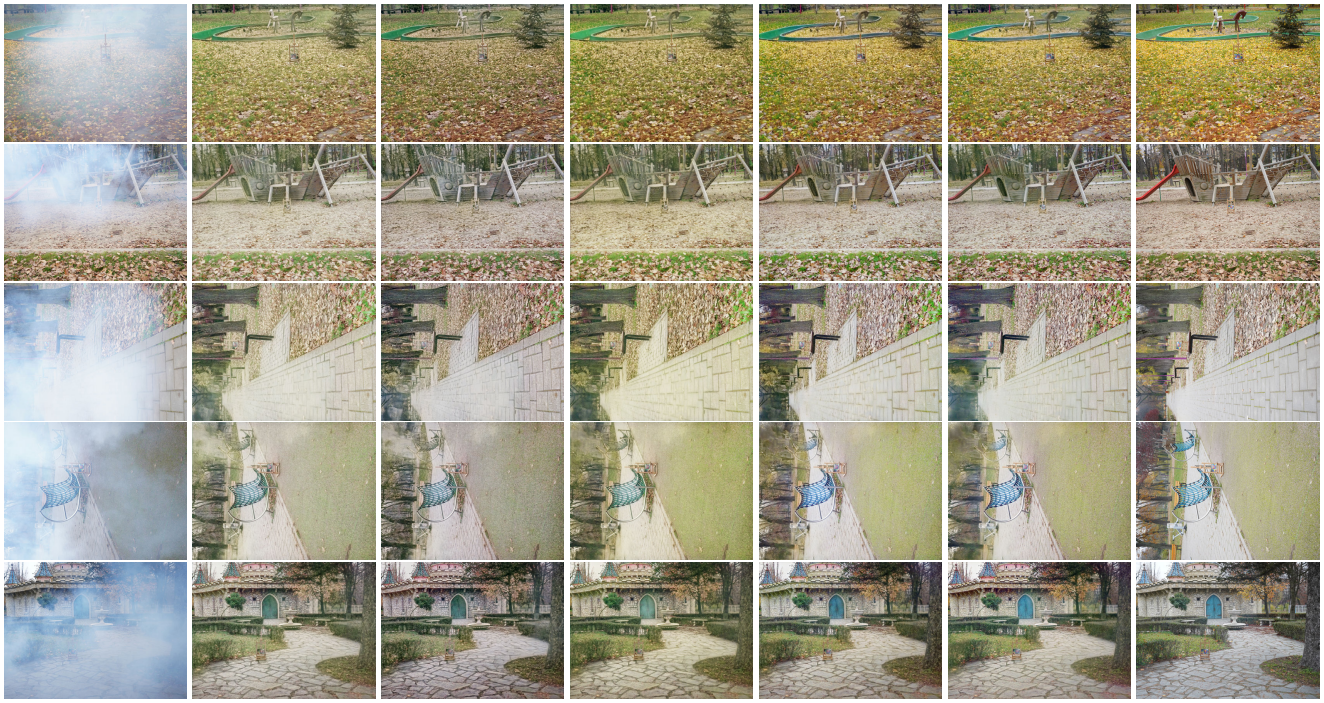

Figure 4. Visual experiment results on NH-HAZE [3] dataset. Obviously, our method demonstrates superior performance on color preservation and detail maintaining, further enhancing the overall quality of the output.

Configurations	PSNR↑	SSIM↑
w/o Refinement module	21.50	0.736
w/o Dehazing module	20.56	0.728
only main branch	21.44	0.730
only frequency branch	20.07	0.697

Table 2. The ablation result of our method on DNH-HAZE[5] dataset. Each component of our DehazeDCT contribute positively to our final dehazing performance.

L_1	L_{Percep}	L_{SSIM}	L_{adv}	PSNR	SSIM
✓	✓	✓	✓	21.50	0.736
✓	✓	✓		21.43	0.733
✓	✓			21.21	0.725
✓				21.17	0.720

Table 3. We conduct experiments to illustrate the rationality of loss function used for Stage I training (w/o Refinement module). Our adopted loss function help achieve optimal performance.

ducing colors that are more closely aligned with the ground truth, manifesting higher clarity, greater color fidelity, and more distinct details. In contrast, other methods present distinct shortcomings. For instance, the FFA method appears to struggle with color retention, resulting in a washed-out appearance. TDN and DehazeFormer leave the image with a slightly hazy residue that obscures finer details.

4.3. Ablation Study

To analyze the effectiveness of each component in our proposed method and justify the optimization objective utilized for training, we conduct extensive ablation experiments on DNH-HAZE Dataset. Since our method contains two separate modules and our dehazing module includes two branches, we adopt a break-down ablation to study the effectiveness of each module and each branch for dehazing.

Effectiveness of Refinement Module. To study the importance of refinement module, we remove this module from our model and Tab. 2 reports the quantitative performance of remaining architecture, which still demonstrate competitive performance compared to the current SOTA in Tab. 1. However, compared to our DehazeDCT, only utilizing the dehazing module suffers from obvious decrease in PSNR and SSIM, which vividly underscores the contributions of Refinement module in our proposed method. In addition, various combinations of loss functions are compared in Tab. 3, showing that the loss function we adopted in Stage I training is reasonable and effective.

Importance of Dehazing Module. In order to evaluate the importance of our proposed Dehazing module, based on our DehazeDCT, we remove the Dehazing module and directly leverage the transformer based Refinement module for dehazing and we obtain the result as Tab. 2 row 2. Compared to the result of our complete model reported in Tab. 1,

Figure 5. Visual Comparisons on HD-NH-HAZE dataset [4]. Our method exhibits superior haze removal, evidenced by more vivid colors and clearer details, especially in the foliage and background structures.

the significantly lower PSNR and SSIM values indicate that employing the Refinement module alone is insufficient for effective dehazing, underscoring the critical importance of the integration our Dehazing module.

Contributions of DCNFormer blocks and Frequency-Aware Branch. To further evaluate the effectiveness of each branch in the Dehaizng module, we separately adopt each branch without Refinement module to examine their contributions for dehazing. As shown in Tab. 2 row 3 and row 4, the inferior results of main branch and frequency-aware branch, compared to the complete Dehazing module, verify the beneficial contributions of each branch for dehazing and substantiate the rationality of our model's architecture. Furthermore, compared to the frequency-aware branch, our proposed DCNFormer blocks demonstrate more powerful representation learning and achieve

Figure 6. Our results on the validation set of NTIRE 2024 Dense and Non-Homogeneous Dehazing Challenge [5], achieving the best performance in terms of both PSNR and SSIM on the validation leaderboard.

Team	PSNR↑	SSIM↑	LPIPS↓	MOS↑	Average Rank	Fianl Rank↓
USTC-Dehazers	**22.94**	**0.7294**	0.3520	**6.315**	**2.25**	**1**
Dehazing_R (Ours)	22.84	0.7253	0.3466	5.96	3.25	2
Team Woof	22.60	0.7269	0.3809	5.79	4.25	3
ITB Dehaze	22.32	0.7149	0.3337	5.705	4.25	4
TTWT	21.93	0.7146	0.3345	5.675	5.25	5
DH-AISP	21.90	0.7144	0.4017	5.81	6	6
BU-Dehaze	21.68	0.7094	0.3267	5.22	6.5	7
RepD	21.78	0.7061	0.3328	4.83	7	8
PSU Team	20.54	0.6328	**0.2678**	5.31	8.75	9
xsourse	21.66	0.6955	0.4493	5.28	9.75	10

Table 4. Final ranking (top 10 teams) of NTIRE 2024 Dense and Non-Homogeneous Dehazing Challenge [5]. Our team (Dehazing_R) achieves the **second best** performance among all submitted solutions (16 submissions in total). [Key: **Best**, Second Best, ↑ (↓): The larger (smaller) represents the better performance].

more pleasant result.

4.4. Performance of Our Method on NTIRE 2024 Dense and Non-Homogeneous Challenge

The challenge results are evaluated by PSNR, SSIM, Learned Perceptual Image Patch Similarity (LPIPS) [50] and Mean Opinion Score (MOS) via a user study [5]. The quantitative results of the top 10 teams are shown in Tab. 4 and our solution achieves the second best performance in total. Specifically, apart form the first solution, our method outperforms other solutions by a margin, which is quantitatively evidenced by an augmentation of 0.24 dB in PSNR and a notable elevation of 0.17 in MOS. Additionally, we provide our results of the official test set and validation set in Fig. 1 and Fig. 6, respectively. Obviously, it can be observed that the results generated by our model exhibit high fidelity and visual appeal. In dense hazy areas, there are no residual traces of haze and disjointed feeling with other regions. Additionally, our results feature well-defined details and vivid color, showing the superiority of our model.

5. Conclusion

In this paper, we introduce an effective non-homogeneous **Dehaz**ing **m**ethod based on **D**eformable **C**onvolutional **T**ransformer (**DehazeDCT**). Specifically, we design a transformer-like architecture for effective dehazing with the core operator of deformable convolution v4, which offers long-range dependency and adaptive spatial aggregation capabilities and demonstrates faster convergence and forward speed. Additionally, we use a streamlined transformer grounded in Retinex theory to further improve the color and structure details. Comprehensive experiment results demonstrate the effectiveness of our proposed method. Furthermore, our method achieves outstanding performance in NTIRE 2024 Dense and Non-Homogeneous Dehazing Challenge (**second best** performance among 16 solutions).

References

[1] Codruta O. Ancuti, Cosmin Ancuti, and Radu Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, 2020.

[2] Codruta O Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, Radu Timofte, et al. Ntire 2020 challenge on nonhomogeneous dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, 2020.

[3] Codruta O. Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, and Radu Timofte. Ntire 2021 nonhomogeneous dehazing challenge report. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, 2021.

[4] Codruta O. Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, Radu Timofte, et al. Ntire 2023 hr nonhomogeneous dehazing challenge report. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, 2023.

[5] Codruta O. Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, Radu Timofte, et al. Ntire 2024 dense and nonhomogeneous dehazing challenge report. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, 2024.

[6] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and DachengTao. An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing (TIP)*, 2016.

[7] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2023.

[8] Lu Chi, Borui Jiang, and Yadong Mu. Fast fourier convolution. *Advances in Neural Information Processing Systems*, 2020.

[9] Yuning Cui, Yi Tao, Zhenshan Bing, Wenqi Ren, Xinwei Gao, Xiaochun Cao, Kai Huang, and Alois Knoll. Selective frequency network for image restoration. In *The Eleventh International Conference on Learning Representations*, 2023.

[10] 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 (ICCV)*, 2017.

[11] Xiaohan Ding, Xiangyu Zhang, Jungong Han, and Guiguang Ding. Scaling up your kernels to 31x31: Revisiting large kernel design in cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11963–11975, 2022.

[12] Wei Dong, Han Zhou, and Dong Xu. A new sclera segmentation and vessels extraction method for sclera recognition. In *2018 10th International Conference on Communication Software and Networks (ICCSN)*, 2018.

[13] Wei Dong, Han Zhou, Yuqiong Tian, Jingke Sun, Xiaohong Liu, Guangtao Zhai, and Jun Chen. ShadowRefiner: Towards mask-free shadow removal via fast fourier transformer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, 2024.

[14] Kang Fu, Yicong Peng, Zicheng Zhang, Qihang Xu, Xiaohong Liu, Jia Wang, and Guangtao Zhai. Attentionlut: Attention fusion-based canonical polyadic lut for real-time image enhancement. *arXiv preprint arXiv:2401.01569*, 2024.

[15] Ian J. Goodfellow, Pouget-Abadie Jean, Mehdi Mirza, Bing Xu, Warde-Farley David, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Association for Computing Machinery*, 2020.

[16] Guan Guang, Xingang Wang, Wenqi Wu, Han Zhou, and Yuanyuan Wu. Real-time lane-vehicle detection and tracking system. In *Chinese Control and Decision Conference (CCDC)*, 2016.

[17] Chun-Le Guo, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, and Chongyi Li. Image dehazing transformer with transmission-aware 3d position embedding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[18] Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. Focal frequency loss for image reconstruction and synthesis. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2021.

[19] Mingye Ju, Zhenfei Gu, and Dengyin Zhang. Single image haze removal based on the improved atmospheric scattering model. *Neurocomputing*, 2017.

[20] Yunan Li, Qiguang Miao, Jianfeng Song, Yining Quan, and Weisheng Li. Single image haze removal based on haze physical characteristics and adaptive sky region detection. *Neurocomputing*, 2016.

[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 International Conference on Computer Vision Workshops (ICCV Workshops)*, 2021.

[22] Jing Liu, Haiyan Wu, Yuan Xie, Yanyun Qu, and Lizhuang Ma. Trident dehazing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, 2020.

[23] Shiwei Liu, Tianlong Chen, Xiaohan Chen, Xuxi Chen, Qiao Xiao, Boqian Wu, Mykola Pechenizkiy, Decebal Mocanu, and Zhangyang Wang. More convnets in the 2020s: Scaling up kernels beyond 51x51 using sparsity. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2023.

[24] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention based multi-scale network for image dehazing. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019.

[25] Xiaohong Liu, Zhihao Shi, Zijun Wu, Jun Chen, and Guangtao Zhai. Griddehazenet+: An enhanced multi-scale network with intra-task knowledge transfer for single image dehazing. *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[26] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer:

Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2021.

[27] W. E. K. Middleton. *Vision through the atmosphere*. University of Toronto Press, 1952.

[28] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2020.

[29] Yuwei Qiu, Kaihao Zhang, Chenxi Wang, Wenhan Luo, Hongdong Li, and Zhi Jin. Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2023.

[30] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[31] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[32] Vishwanath A. Sindagi, Poojan Oza, Rajeev Yasarla, and Vishal M. Patel. Prior-based domain adaptive object detection for hazy and rainy conditions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[33] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. In *IEEE Transactions on Image Processing (TIP)*, 2023.

[34] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[35] Wenyi Wang, Jun Hu, Xiaohong Liu, Jiying Zhao, and Jianwen Chen. Single image super-resolution based on multiscale structure and nonlocal smoothing. *EURASIP Journal on Image and Video Processing*, 2021.

[36] Wenhai Wang, Jifeng Dai, Zhe Chen, Zhenhang Huang, Zhiqi Li, Xizhou Zhu, Xiaowei Hu, Tong Lu, Lewei Lu, Hongsheng Li, et al. Internimage: Exploring large-scale vision foundation models with deformable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[37] Xintao Wang, Kelvin C.K. Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, 2019.

[38] Yingqian Wang, Jungang Yang, Longguang Wang, Xinyi Ying, Tianhao Wu, Wei An, and Yulan Guo. Light field image super-resolution using deformable convolution. In *IEEE Transactions on Image Processing (TIP)*, 2021.

[39] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. In *IEEE Transactions on Image Processing (TIP)*, 2004.

[40] Zhendong Wang, Xiaodong Cun, Jianmin Bao, and Jianzhuang Liu. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[41] Haiyan Wu, Jing Liu, Yuan Xie, Yanyun Qu, and Lizhuang Ma. Knowledge transfer dehazing network for nonhomogeneous dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, 2020.

[42] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[43] Yuwen Xiong, Zhiqi Li, Yuntao Chen, Feng Wang, et al. Efficient deformable convnets: Rethinking dynamic and sparse operator for vision applications. *arXiv preprint arXiv:2401.06197*, 2024.

[44] Dong Xu, Wei Dong, and Han Zhou. Sclera recognition based on efficient sclera segmentation and significant vessel matching. In *The Computer Journal*, 2022.

[45] Hao-Hsiang Yang and Yanwei Fu. Wavelet u-net and the chromatic adaptation transform for single image dehazing. In *IEEE International Conference on Image Processing (ICIP)*, 2019.

[46] Xiangyu Yin, Xiaohong Liu, and Huan Liu. Fmsnet: Underwater image restoration by learning from a synthesized dataset. In *International Conference on Artificial Neural Networks (ICANN)*, 2019.

[47] Jaeyoung Yoo, Sang-ho Lee, and Nojun Kwak. Image restoration by estimating frequency distribution of local patches. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[48] Hu Yu, Naishan Zheng, Man Zhou, Jie Huang, Zeyu Xiao, and Feng Zhao. Frequency and spatial dual guidance for image dehazing. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022.

[49] 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 Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[50] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[51] Han Zhou, Wei Dong, Yangyi Liu, and Jun Chen. Breaking through the haze: An advanced non-homogeneous dehazing method based on fast fourier convolution and convnext. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, 2023.

[52] Xizhou Zhu, Han Hu, Stephen Lin, and Jifeng Dai. Deformable convnets v2: More deformable, better results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.