TCL-Net: A Lightweight and Efficient Dehazing Network with Frequency-Domain Fusion and Multi-Angle Attention

Cheng Tang[1,2] and Wenqi Lou[1,2]⋆

[1] University of Science and Technology of China, Hefei, China
[2] Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou, China
tangcheng@mail.ustc.edu.cn, louwenqi@ustc.edu.cn

Abstract. Hazy images present a challenging ill-posed problem, suffering from information loss and color distortion. Current deep learning-based dehazing methods enhance performance by increasing network depth but incur substantial parameter overhead. Meanwhile, standard convolutional layers concentrate on low-frequency details, often overlooking high-frequency information, which hinders the effective utilization of prior information presented in blurred images. In this paper, we propose TCL-Net, a lightweight dehazing network which emphasizes on frequency-domain features. Our network first includes a sophisticated layer for extracting high-frequency and low-frequency information, specifically designed using Fast Vision Transformers for the original blurred images. Concurrently, we have designed a frequency-domain information fusion module that integrates high-frequency and low-frequency information with the characteristics of convolutional networks for subsequent convolutional layers. Furthermore, to better leverage spatial information of the original image, we introduce a multi-angle attention module. With the aforementioned design, our network achieves superior performance with a total parameter size of only 0.48MB, representing an order of magnitude reduction in parameters compared to other state-of-the-art lightweight networks.

Keywords: Image Dehazing · Lightweight Neural Network · Frequency-Domain Fusion · Multi-Angle Attention

1 Introduction

With the development of industrialization, haze has become a common atmospheric phenomenon. Hazy images suffer from information loss and color distortion. Image dehazing aims to remove these adverse effects to restore clear and sharp visuals. In the domain of dehazing, the atmospheric scattering model

⋆ Corresponding author

(ASM) stands as the most classical depiction of the formation of hazy images [27,28], which can be formally written as:

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

where $I(x)$ is observed as hazy image and $J(x)$ is clean image, A denotes the global atmospheric light, and $t(x)$ is the transmission map. According to Eq.(1), we can observe that the extent of hazy image restoration depends on two key parameters: the atmospheric light A and the value of the transmission map $t(x)$.

Therefore, in the early stages, researchers attempted to use convolutional neural networks (CNNs) [5,33,41,42] to accurately estimate the transmission map $t(x)$ and atmospheric light A values. These networks had relatively simple structures and processed limited information, leading to often poor dehazing performance. With the advancement of deep learning, many methods have been developed to map hazy images to clear ones [17,23,23,11,38,21]. Notable examples include FFA-Net [31], which proposes an end-to-end feature fusion attention network for single image dehazing, and RIDCP [39], which enhances dehazing performance and result adjustability through high-quality codebook priors and controllable feature matching operations. DCP [18] introduces a single image dehazing method based on the dark channel prior, estimating and restoring high-quality haze-free images and depth maps by analyzing the characteristics of outdoor haze-free images.

However, conventional convolutions predominantly focus on low-frequency information, paying insufficient attention to high-frequency details such as edges and contours [26]. This oversight leads to subpar dehazing results, especially in terms of preserving fine details. The aforementioned methods typically rely on traditional convolutional operations and tend to simply stack layers deeply. Despite this approach resulting in a significant increase in parameter count, it fails to effectively address the dehazing task. These networks have two main drawbacks: First, they use traditional convolutions that capture information from a single perspective, limiting their ability to comprehensively process the image; Second, they inadequately handle high-frequency information such as edges and contours, resulting in inferior dehazing performance. As shown in the Fig. 1, methods such as DCP, RIDCP, and FFA-Net, which do not adequately focus on high-frequency information in images, perform poorly in removing uneven dense fog compared to our network.

To address the two main drawbacks of existing dehazing models, we propose TCL-Net: A Lightweight and Efficient Dehazing Network with Frequency-Domain Fusion and Multi-Angle Attention. Firstly, to enhance the capability of traditional convolutional networks in capturing information, our model integrates a Multi-Angle Attention (MAA) module that combines horizontal, angular, central, and vertical differential convolution blocks in parallel. These blocks provide attention across various directions, capturing details from unique orientations and thus enriching the directional information available. Moreover, the implementation of parameter fusion techniques ensures that the parameter count remains controlled, even with the application of multidimensional convolutions.

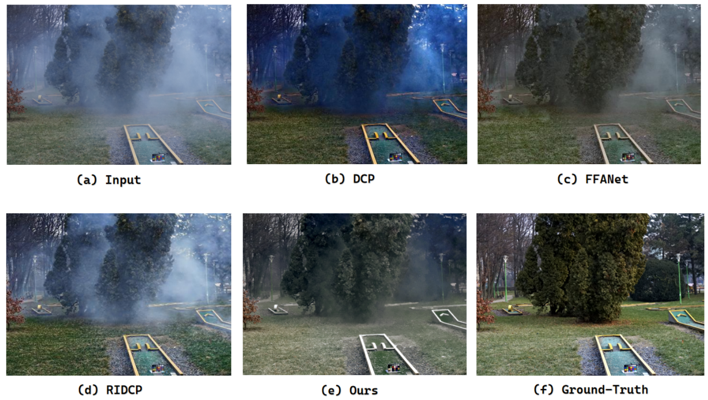

Fig. 1. Comparison of visual results on NH-HAZE dataset[2], (a) is the original hazy image, (b) is the result processed by DCP [18], (c) is the result processed by FFA-Net [31], (d) is the result processed by RIDCP [39], (e) is processed by ours, and (f) is the original clear image.(Color figure online)

Secondly, to overcome the traditional networks' lack of focus on high-frequency details in images, we have developed a module based on Vision Transformers. This module separately and efficiently extracts high and low-frequency features from hazy images, facilitated by a fusion module that adaptively weights and merges these details. This method not only promotes comprehensive feature extraction but also maintains high performance, while preventing parameter explosion and significantly shortening the duration of training and inference cycles. This approach demonstrates a highly efficient processing capability, attributable to specific technical optimizations such as parallel processing and algorithmic enhancements.

In summary, this work makes the following contributions:

- **High- and Low-Frequency Pattern Extractor (HLFPE):** Our network includes a specialized layer that extracts both high-frequency and low-frequency information from hazy images. This layer leverages fast Vision Transformers to effectively capture the necessary features from the original blurred images.
- **Sophisticated High- and Low-Frequency Pattern Fusion (HLFPF):** We design a unique fusion module that integrates high-frequency and low-frequency information, enhancing the capability of subsequent convolutional layers. This module capitalizes on the complementary characteristics of different frequency domains to improve dehazing performance.
- **Multi-Angle Attention Module (MAA) :** To better utilize spatial information, we introduce a multi-angle attention module that enables the network to effectively capture features at various scales, further improving dehazing results.

2 Related Work

2.1 Single Image Dehazing

Image dehazing, which involves restoring clear images from hazy ones, has been a challenging and persistent ill-posed problem for a long time. Existing dehazing methods can be broadly categorized into two types: physics-based approaches and deep learning-based approaches.

Physics-based methods typically rely on the atmospheric scattering model (ASM) [27,28] or handcrafted priors. Handcrafted prior methods are pioneers in the image dehazing task, with well-known examples including the dark channel prior [18], color line prior [14], color attenuation prior [45], sparse gradient prior [6], maximum reflectance prior [43], and non-local prior [4]. However, handcrafted priors are mainly derived from the authors' empirical observations and can't accurately describe the formation and process of haze.

Unlike physics-based methods, deep learning-based dehazing models employ convolutional neural networks (CNNs) to learn image representations [5,42,24]. AOD-Net [21] enhances traditional dehazing quality by rewriting the ASM formula, simplifying the estimation of atmospheric light and transmission map into a single task. However, despite its lightweight nature, [21] has limited ability to handle image details. Traditional convolutional methods extract very limited information from the original images. To improve this issue, we designed the Multi-Angle Attention (MAA) module, which utilizes different convolutional kernels to extract high-frequency and low-frequency information from various directions in the image. By processing these features in parallel, our model enhances the extraction of original image characteristics while reducing model complexity.

2.2 Vision Transformer

The emergence of Transformers [36] in NLP introduced a non-recurrent architecture with an encoder-decoder framework and self-attention mechanisms [16,20], enabling parallel computation, reducing training time, and achieving state-of-the-art performance in machine translation [37,32]. Recently, vision Transformers (ViT) has outperformed almost all CNN-based models in high-level vision tasks. This success has led to the proposal of numerous modified architectures, positioning vision Transformers as a formidable challenge to the dominance of CNNs in high-level vision tasks [9,30,25,40,12,7,10].

However, ViT has not been fully exploited in the domain of low-level vision, especially in dehazing tasks. Pan et al. [29] introduced the HiLo attention mechanism to reduce the computational and memory demands of Vision Transformers, but their work is limited to image classification and object detection. We propose the High- and Low-Frequency Domain Extractor and Fusion module, which fully leverages the high- and low-frequency information in original hazy images through extraction and fusion. This approach achieves superior dehazing performance while maintaining low overheads.

3 Proposed Method

Overall Architecture. As shown in Fig. 2, our model consists of 4 parts: Feature Extractor Network (FEN), High- and Low-Frequency Pattern Extractor (HLFPE), High- and Low-Frequency Pattern Fusion (FLFPF), and Multi-Angle Attention (MAA). Our model is an end-to-end dehazing network with a straightforward and clear structure. Initially, the information extracted by the HLFPE model is fused by the HLFPF model and then processed through the FEN model. The outputs of these three modules are concatenated through a simple concatenation operation. Finally, a 1×1 convolution layer is applied to produce the final clear image. In the following sections, we will detail the composition and function of each module step by step.

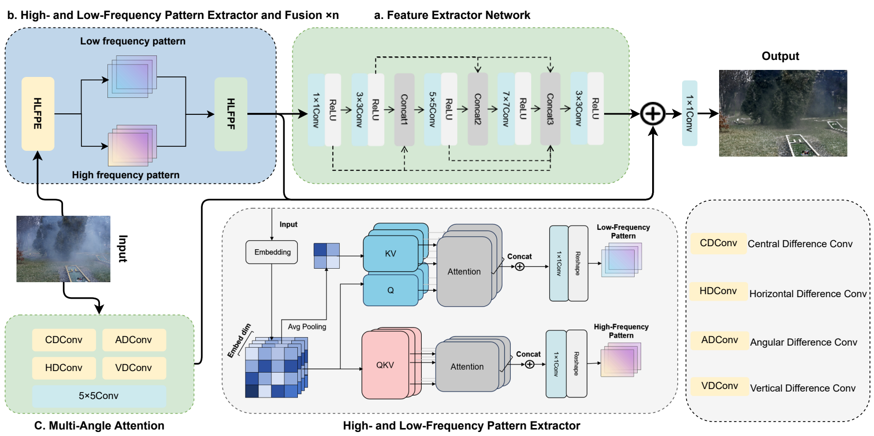

Fig. 2. Overview architecture of TCL-Net.

3.1 Feature Extractor Network

Previous research work relied on the independent computation of A and $t(x)$ based on the ASM, leading to significant restoration errors and inefficient runtime.

Eq. (1) can be reformulated as follow:

$$J(x) = \frac{1}{t(x)} I(x) - A\frac{1}{t(x)} + A \tag{2}$$

$$J(x) = K(x)I(x) - K(x) + c \tag{3}$$

where $K(x)$ can be defined as

$$K(x) = \frac{\frac{1}{t(x)}(I(x) - A) + (A - b)}{I(x) - 1} \tag{4}$$

Li et al. [21] obtained a deformed ASM by introducing the Eq. (4) in Eq. (1), so we can directly estimate the joint value $K(x)$ of transmission map $t(x)$ and atmospheric light A. By jointly estimating the transmission map and the global atmospheric light, the model simplifies by reducing the problem from estimating two variables to only one, making the network more efficient and effective. As shown in the Fig. 2 on the right, our Feature Extractor Network is constructed drawing inspiration from this.

3.2 High- and Low-Frequency Pattern Extractor

The Vision Transformers (ViT) model excels in establishing a global information interaction mechanism, but the computational complexity of global attention is considerably high, especially with long sequence inputs. Using traditional Vision Transformer methods to capture high-frequency details of objects can significantly degrade performance. As high frequencies encode local details, applying global attention to a feature map can be redundant and computationally expensive. Recent works Pan et al. [29] have introduced the HiLo attention mechanism to reduce the computational and memory demands of Vision Transformers. To improve model speed and information extraction capability, we propose a High- and Low-Frequency Pattern Extractor based on their work. To the best of our knowledge, this is the first use of ViT to independently extract high- and low-frequency patterns from hazy images.

As shown in the bottom part of Fig. 2, high-frequency pattern in images typically contains local details, such as edges and textures. And the upper section designed to capture the high-frequency patterns required for subsequent fusion. Initially, the input is embedded with positional encoding, followed by local self-attention windows (e.g., 2×2) to effectively capture high-frequency information. This method extracts fine details while reducing computational complexity.

The lower section is designed to capture low-frequency patterns. Initially, the embedded information undergoes pooling operations, which effectively capture global features of the image and reduce memory consumption, thereby enhancing the computational efficiency of the model. Unlike the high-frequency section, in the low-frequency section, the query (Q) originates from the original input feature map to maintain spatial resolution consistency, while the key (K) and value (V) are derived from the feature map after average pooling. This design allows the model to leverage global information in the low-frequency branch and local detail information in the high-frequency branch.

Experiments have demonstrated that HLFPE can effectively capture high- and low-frequency information from hazy images, thereby enhancing the network's dehazing performance.

3.3 High- and Low-Frequency Pattern Fusion

Previous dehazing work has demonstrated that integrating different information sources can enhance the network's ability to capture image information [15,13]. The simplest fusion method involves summing the elements extracted from high-and low-frequency information, which has been adopted in many prior methods [11,3]. However, simply summing the extracted high- and low-frequency features cannot effectively utilize the captured information. Therefore, we have specifically designed a High- and Low-Frequency Pattern Fusion module to merge high- and low-frequency information, enabling subsequent feature extraction modules to efficiently learn image priors.

As show in Fig. 3 in the process of amalgamating high-frequency and low-frequency information, an element-wise sum operation is initially carried out. Then, channel importance for both frequencies is calculated separately. Spatial significance is assessed with average and max pooling, and the sigmoid function determines the weights for their representational strength. A residual connection finally fuses the information, effectively leveraging the high- and low-frequency features to help the model focus on key areas more efficiently.

This module facilitates the precise fusion of high- and low-frequency features by weighting their respective contributions, leading to more informative representations. The efficacy of this approach is thoroughly validated in our ablation experiments.

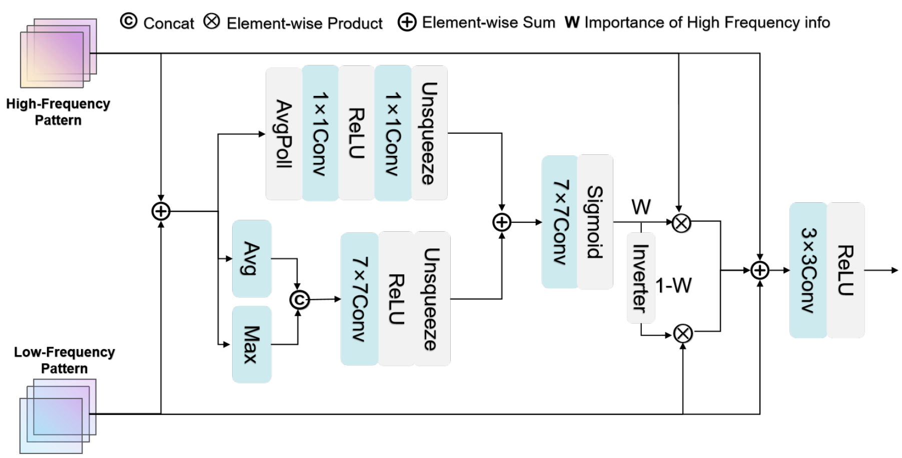

Fig. 3. The high- and low-frequency pattern fusion model (HLFPF).

3.4 Multi-Angle Attention

A single image contains various significant details, such as textures, edges, and corner points, which are crucial for image interpretation. In addition to extracting high and low-frequency information, we have incorporated a Multi-Angle Attention (MAA) module. This module is equipped with a set of convolutional blocks operating in parallel, comprising Horizontal Difference Convolution, Angular Difference Convolution, Central Difference Convolution, and Vertical Difference Convolution.

The differential convolutions in various directions can effectively capture high-frequency details in different orientations, thus offering a richer set of directional information. Simultaneously, the introduction of differential convolutions also enhances the extraction of global contextual information, thereby providing a comprehensive supplement to the detailed information captured by the previously processed high and low-frequency features.

3.5 Loss Function

Incorporating the methodology detailed by Andrey et al. [19], we employ a color loss function to fine-tune the visual accuracy of our results, thereby enhancing fidelity to the original image characteristics. The color loss function is formally defined as:

$$L_{color}(J, GT) = ||J_b - GT_b||_2^2 \tag{5}$$

where J_b and GT_b are the blurred images of J and GT, respectively:

$$J_b(i,j) = \sum_{k,l} J(i+k, j+l) * G(k,l) \tag{6}$$

The 2D Gaussian blur operator is given by:

$$G(k,l) = Aexp(-\frac{(k-u_J)^2}{2\sigma_J} - \frac{(l-u_{GT})^2}{2\sigma_{GT}}) \tag{7}$$

where $A = 0.053$, $u_{J,GT} = 0$, and $\sigma_{J,GT} = 3$.
Our TCL-Net is trained by minimizing the pixel-wise difference between the predicted clean image J and the corresponding ground truth GT. In our implementation, we use a combination of mean squared error (MSE) and color loss functions:

$$L_{GT,J} = \alpha * ||J - GT||_2 + (1-\alpha) * L_{color}(J, GT) \tag{8}$$

In our proposed loss function, the parameter α plays a key role in balancing the contributions of different components. To determine the optimal value of α, we conducted a series of experiments, systematically varying α across a range of values from 0 to 1 with a step size of 0.05 and assessing the model's performance on a validation dataset, specifically evaluating the maximum PSNR and SSIM values on the SOTS_indoor dataset [22]. Our experiments revealed that setting $\alpha = 0.8$ consistently yielded the best performance. This value strikes an effective balance between the loss components, significantly improving the model's accuracy.

4 Experiments

4.1 Datasets

We conducted experiments by training and evaluating our model on both synthetic RESIDE [22] and real-world datasets: DenseHaze [1] and NH-Haze [2].

Additionally, we compared our model's performance with the latest state-of-the-art models.

Synthetic Datasets. We utilized the RESIDE dataset, proposed by Li et al. [22], to evaluate the performance of our model in synthetic scenarios. Specifically, we employed the ITS (Indoor Training Set) and OTS (Outdoor Training Set) subsets for training, while the SOTS (Synthetic Objective Test Set) was designated for evaluation. The ITS dataset comprises 1,399 clean images paired with 13,990 synthetic hazy counterparts, and the OTS dataset contains 2,061 clean images alongside 72,135 synthetic hazy images. For testing, the SOTS dataset includes 500 indoor and 500 outdoor images.

Real-Word Datasets. To evaluate our model's performance in challenging real-world scenarios, we utilized two real-world datasets: DenseHaze [1] and NH-HAZE [2]. DenseHaze comprises densely and uniformly hazy scenes, whereas NH-HAZE consists of non-uniform hazy scenes. Each dataset contains 55 images, and we split the datasets into training, validation, and testing sets with a ratio of 8:1:1.

4.2 Implementation Details

Our architecture is implemented using the PyTorch framework and deployed on an Nvidia A100 GPU. We employ the Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.9999$, and the initial learning rate is set to 1×10^{-4}. The learning rate is dynamically adjusted using a cosine annealing schedule. The batch size is set to 16, and the patch size is 256×256. All experiments under the same conditions to ensure fairness. The total parameter count for our TCL-Net is 0.48M.

4.3 Evalutation Metrics and Comparisons Methods

Evalutation Metrics. In the field of image dehazing, PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index Measure) are distortion-based metrics and are two crucial evaluation metrics used to assess the effectiveness of dehazing algorithms. PSNR evaluates image quality by calculating the ratio between the maximum possible pixel value of the original image and the difference between the original and processed images. Generally, a higher PSNR value indicates lower image distortion and better image quality. SSIM evaluates the visual quality of the image by considering human visual perception. Unlike PSNR, which primarily focuses on pixel value differences, SSIM takes into account the characteristics of human visual systems and assesses image quality by simulating human visual system principles. In this experiment, we employed both of these common evaluation metrics.

Comparsion Methods. We compare our method with state-of-the-art (SOTA) methods both qualitatively and quantitatively. The comparison includes three classic models (DCP [18], DehazeNet [5], AOD-Net [21]) and eight SOTA models (FFA-Net [31], LDN [35], RDN [44], AECR-Net [38], Dehamer [17], RIDCP [39], DehazeFormer-B [34], DEA-Net [8]), all of which have been published in recent top-tier journals and conferences.

Table 1. Quantitative comparisons on different datasets for different dehazing methods. Best values and second-best values for each metric are color-coded.

Method	NH-Haze		SOTS_outdoor		SOTS_indoor		Dense_Haze	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
DCP	14.91	0.6742	19.14	0.8605	16.61	0.8546	14.15	0.5521
DehazeNet	16.62	0.5241	27.75	0.9269	19.82	0.9269	13.84	0.4252
AOD-Net	15.53	0.6332	24.14	0.9198	20.51	0.8162	14.51	0.4883
FFA-Net	19.87	0.6913	36.39	0.9886	36.39	0.9886	14.39	0.4524
LDN	20.95	0.7961	—	—	21.27	0.8321	18.27	0.6031
RDN	12.37	0.5392	24.12	0.9151	—	—	12.15	0.4262
AECR-Net	18.51	0.6561	—	—	33.34	0.9824	15.86	0.4663
Dehamer	20.66	0.6821	36.63	0.9881	36.73	0.9891	16.62	0.5632
RIDCP	12.27	0.5014	18.36	0.7526	18.36	0.7526	8.09	0.4173
DehazeFormer	17.37	0.7256	31.46	0.9864	33.58	—	—	—
DEA-Net	19.55	0.6645	—	—	39.16	0.9921	—	—
Ours	**21.45**	**0.7261**	**36.78**	**0.9913**	**37.85**	**0.9924**	**18.34**	**0.6121**

4.4 Performance Analysis

Visual Analysis of the Model's High-Frequency Information Processing Capabilities. To validate the effectiveness of our model in capturing high-frequency components, we visualized the frequency response of various algorithms after 1,000 iterations using the Discrete Fourier Transform (DFT). A greater concentration of high-frequency energy in the DFT magnitude spectra indicates a superior capacity to preserve fine details. As shown in Fig. 4, our model outperforms previous SOTA methods in high-frequency retention, demonstrating its distinct advantages. The results highlight that our model is adept at preserving a greater amount of high-frequency details, reinforcing its robustness in effectively managing critical image information.

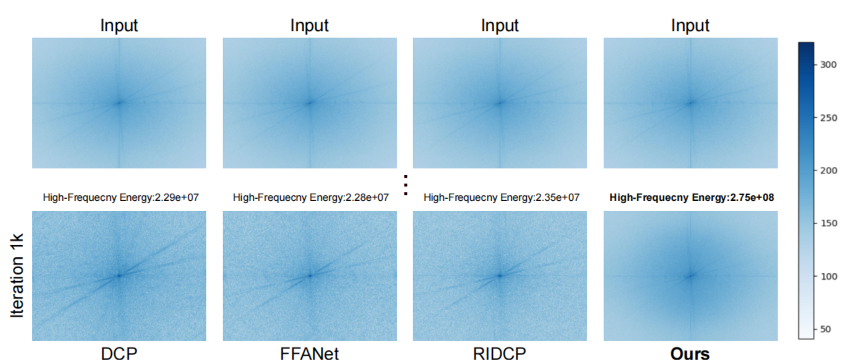

Fig. 4. The DFT results of filtered features.

Comparision with SOTAs on synthetic hazy images. Table 1 presents a quantitative comparison of our method and various approaches on the indoor and outdoor of SOTS datasets. Our approach achieves optimal performance in SSIM on both datasets and performs competitively in PSNR. Specifically, our model achieves the best performance on the SOTS_outdoor dataset, with a PSNR value of 36.78 dB and SSIM of 0.9913. Compared to Dehamer [17], our method improves PSNR by 0.15 dB and SSIM by 0.0032 on the SOTS_outdoor datasets. Moreover, our model achieved the second-highest PSNR and the highest SSIM on the SOTS_indoor dataset.

As show in Fig. 5, DCP, DehazeNet, and LDN perform poorly in restoring background colors during dehazing. In the third row, it is noticeable that the walls exhibit significant pitting and color distortion after dehazing with these methods. In contrast, AOD and FFA-Net handle the walls better but still suffer from dim color issues. Dehamer and our results are visually very similar, with only minor differences in some fine detail contours. Our model, by fully leveraging the high- and low-frequency prior information of the original hazy images, achieves more natural detail processing compared to other models.

On the SOTS_indoor dataset, from a quantitative analysis perspective, our results are only slightly inferior to DEA-Net, with a PSNR lower by 1.31 dB but an SSIM higher by 0.0003 compared to the highest value. As shown in Fig. 6, DCP, DehazeNet, and LDN exhibit color distortion and dim color issues, while AOD, FFA-Net, and Deharmer perform relatively well. Although our quantitative metrics are lower than DehazeNet's, our visual performance excels in high-frequency contour details of some objects. It is important to note that the SOTS dataset consists of synthetically sparse fog, where many existing methods have already demonstrated high performance and visually pleasing dehazing results.

Comparsion with SOTAs on real-world hazy images. In the realm of practical applications, effectively mitigating dense and irregular haze remains a persistently challenging issue. Existing approaches often struggle to achieve optimal results due to constraints imposed by limited datasets. Our model capitalizes on the inherent high and low-frequency priors present in original images, yielding remarkable visual dehazing outcomes. Moreover, in quantitative analysis, our model attains state-of-the-art performance levels. Specifically, on the NH-HAZE dataset, our PSNR exceeds the best-performing model by 0.5 dB, while on the DenseHaze dataset, our PSNR surpasses the highest model by 0.07 dB.

From Fig. 1, it is evident that DCP and RIDCP exhibit limited dehazing capabilities for irregular dense fog, with DCP showing significant color distortion issues. Compared to FFANet, our approach achieves higher fidelity in restoring input hazy images and excels in fine-detail processing, approaching GroundTruth results closely.

Model Complexity and Inference Time. As show in Fig. 7, we present the parameter counts and FLOPs for processing a single image on the SOTS dataset for different models. Notably, the values for DEA-Net are slightly higher than

Fig. 5. Dehazed results of Ours and Other models on SOTS_outdoor dataset.

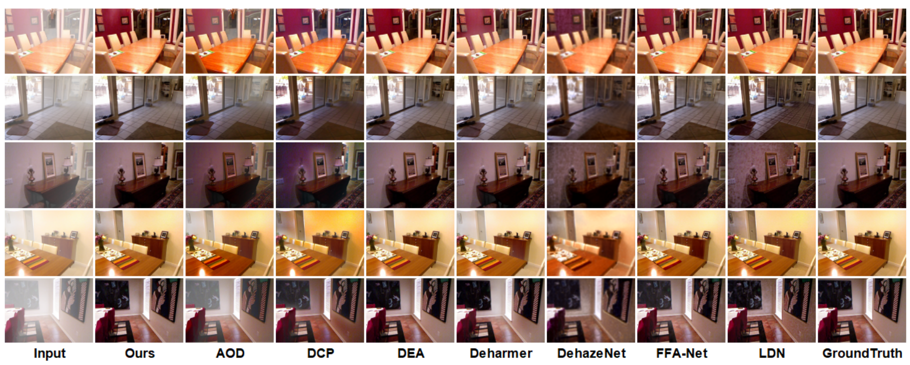

Fig. 6. Dehazed results of Ours and Other models on SOTS_indoor dataset.

those reported in its original paper due to differences in deployment environments and the PyTorch version used. We have made efforts to account for these practical considerations in our comparisons. It is worth noting that although DehazeNet, AOD-Net, and LDN have relatively small parameter counts, their actual effectiveness and quantitative performance are suboptimal. Compared to the most SOTA model DEA-Net, our model slightly outperforms it on the NH-HAZE dataset while performing slightly worse on the SOTS dataset, which can be negligible in the performance. In our experiment, our model has only 16.9% of the parameter count, 56.8% of the computational cost, and 38.8% of the inference time compared to DEA-Net.

4.5 Ablation Studies

In this section, we conducted several ablation experiments to evaluate the proposed method. These experiments include the following ablation models:

- **(a) Baseline:** Only the Feature Extractor Network.
- **(b) No-HLFPEF:** The model without the High- and Low-Frequency Pattern Extractor and Fusion (HLFPEF) module.
- **(c) No-Concat-Fusion:** The model using simple concatenation for high- and low-frequency information fusion after the High- and Low-Frequency Pattern Extractor.
- **(d) No-MAA:** The model without the Multi-Angle Attention (MAA) module.
- **(e) Ours:** The final configuration of our proposed method.

These models, along with our complete model, were trained using the same configuration. Table 2 summarizes the performance of these models. It is evident that our HLFPEF, HLFPF, and MAA modules contribute to enhancing the model's performance.

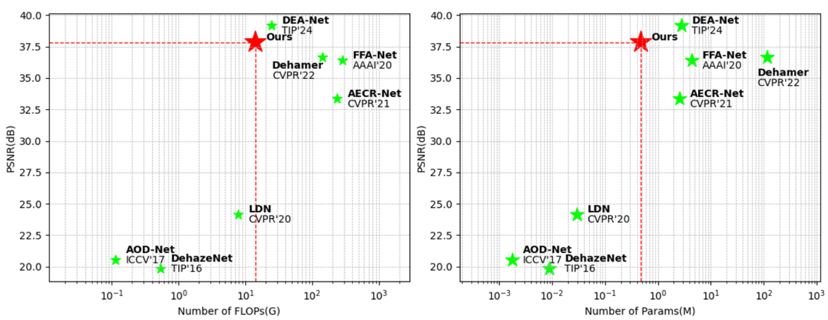

Fig. 7. Comparative analysis of model complexity for various methods. FLOPs are computed using an input size of 620×460, which corresponds to the dimensions of an image from the SOTS_indoor dataset. **Note that** the x-axis represents the logarithmic scale. Although DEA-Net appears close to our model, its parameter size is **300% larger** than ours.

Table 2. The results of ablated models on SOTS_outdoor dataset.

Metric	a	b	c	d	e
PSNR↑	25.14	28.76	31.32	32.25	36.78
SSIM↑	0.92	0.94	0.96	0.96	0.99

Fig. 8 illustrates the performance of different ablation models on the NH-HAZE dataset. It is evident that the model's ability to process high-frequency information in images is significantly diminished without the HLFPEF module.

Fig. 8. Results of different ablation models on NH-Haze dataset.

5 Conclusion

This paper introduces TCL-Net, a lightweight and efficient dehazing network employing frequency-domain fusion and multi-angle attention. We developed a High- and Low-Frequency Pattern Extractor based on a Fast Vision Transformer to effectively extract both high- and low-frequency information from hazy images. Our High- and Low-Frequency Pattern Fusion module enhances this extraction by leveraging prior image knowledge. Additionally, we introduced a Multi-Angle Attention module that incorporates various differential convolution blocks to capture high-frequency details from multiple directions, enriching directional information. Experimental results demonstrate that TCL-Net excels in detail processing, achieving state-of-the-art performance while maintaining a lightweight architecture, with only 16.9% of the parameter count, 56.8% of the computational cost, and 38.8% of the processing time compared to recent models. In future work, we aim to explore the generalization capabilities of TCL-Net across various tasks and investigate a more versatile architecture.

Acknowledgments. This work was supported in part by the National Key R&D Program of China under Grants 2022YFB4501600 and 2022YFB4501603, in part by the National Natural Science Foundation of China under Grants 62102383, 61976200,and 62172380.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ancuti, C.O., Ancuti, C., Sbert, M., Timofte, R.: Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In: 2019 IEEE international conference on image processing (ICIP). pp. 1014–1018. IEEE (2019)
2. Ancuti, C.O., Ancuti, C., Timofte, R.: Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 444–445 (2020)
3. Bai, H., Pan, J., Xiang, X., Tang, J.: Self-guided image dehazing using progressive feature fusion. IEEE Transactions on Image Processing **31**, 1217–1229 (2022)
4. Berman, D., Avidan, S., et al.: Non-local image dehazing. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1674–1682 (2016)
5. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. IEEE transactions on image processing **25**(11), 5187–5198 (2016)
6. Chen, C., Do, M.N., Wang, J.: Robust image and video dehazing with visual artifact suppression via gradient residual minimization. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. pp. 576–591. Springer (2016)
7. Chen, C.F.R., Fan, Q., Panda, R.: Crossvit: Cross-attention multi-scale vision transformer for image classification. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 357–366 (2021)
8. Chen, Z., He, Z., Lu, Z.M.: Dea-net: Single image dehazing based on detail-enhanced convolution and content-guided attention. IEEE Transactions on Image Processing (2024)
9. Choromanski, K., Likhosherstov, V., Dohan, D., Song, X., Gane, A., Sarlos, T., Hawkins, P., Davis, J., Mohiuddin, A., Kaiser, L., et al.: Rethinking attention with performers. arXiv preprint arXiv:2009.14794 (2020)
10. Chu, X., Tian, Z., Wang, Y., Zhang, B., Ren, H., Wei, X., Xia, H., Shen, C.: Twins: Revisiting the design of spatial attention in vision transformers. Advances in neural information processing systems **34**, 9355–9366 (2021)
11. Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., Yang, M.H.: Multi-scale boosted dehazing network with dense feature fusion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2157–2167 (2020)
12. Dong, X., Bao, J., Chen, D., Zhang, W., Yu, N., Yuan, L., Chen, D., Guo, B.: Cswin transformer: A general vision transformer backbone with cross-shaped windows. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12124–12134 (2022)
13. Fan, G., Hua, Z., Li, J.: Multi-scale depth information fusion network for image dehazing. Applied Intelligence **51**(10), 7262–7280 (2021)
14. Fattal, R.: Dehazing using color-lines. ACM transactions on graphics (TOG) **34**(1), 1–14 (2014)
15. Galdran, A., Vazquez-Corral, J., Pardo, D., Bertalmio, M.: Fusion-based variational image dehazing. IEEE Signal Processing Letters **24**(2), 151–155 (2016)
16. Gehring, J., Auli, M., Grangier, D., Yarats, D., Dauphin, Y.N.: Convolutional sequence to sequence learning. In: International conference on machine learning. pp. 1243–1252. PMLR (2017)
17. Guo, C.L., Yan, Q., Anwar, S., Cong, R., Ren, W., Li, C.: Image dehazing transformer with transmission-aware 3d position embedding. In: Proceedings of the

IEEE/CVF conference on computer vision and pattern recognition. pp. 5812–5820 (2022)
18. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE transactions on pattern analysis and machine intelligence **33**(12), 2341–2353 (2010)
19. Ignatov, A., Kobyshev, N., Timofte, R., Vanhoey, K., Van Gool, L.: Dslr-quality photos on mobile devices with deep convolutional networks. In: Proceedings of the IEEE international conference on computer vision. pp. 3277–3285 (2017)
20. Jozefowicz, R., Vinyals, O., Schuster, M., Shazeer, N., Wu, Y.: Exploring the limits of language modeling. arXiv preprint arXiv:1602.02410 (2016)
21. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: All-in-one dehazing network. In: Proceedings of the IEEE international conference on computer vision. pp. 4770–4778 (2017)
22. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing **28**(1), 492–505 (2018)
23. Liu, X., Ma, Y., Shi, Z., Chen, J.: Griddehazenet: Attention-based multi-scale network for image dehazing. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 7314–7323 (2019)
24. Liu, Y., Pan, J., Ren, J., Su, Z.: Learning deep priors for image dehazing. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 2492–2500 (2019)
25. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021)
26. Lou, W., Gong, L., Wang, C., Du, Z., Zhou, X.: Octcnn: A high throughput fpga accelerator for cnns using octave convolution algorithm. IEEE Transactions on Computers **71**(8), 1847–1859 (2021)
27. Narasimhan, S.G., Nayar, S.K.: Contrast restoration of weather degraded images. IEEE transactions on pattern analysis and machine intelligence **25**(6), 713–724 (2003)
28. Nayar, S.K., Narasimhan, S.G.: Vision in bad weather. In: Proceedings of the seventh IEEE international conference on computer vision. vol. 2, pp. 820–827. IEEE (1999)
29. Pan, Z., Cai, J., Zhuang, B.: Fast vision transformers with hilo attention. Advances in Neural Information Processing Systems **35**, 14541–14554 (2022)
30. Park, N., Kim, S.: How do vision transformers work? arXiv preprint arXiv:2202.06709 (2022)
31. Qin, X., Wang, Z., Bai, Y., Xie, X., Jia, H.: Ffa-net: Feature fusion attention network for single image dehazing. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 11908–11915 (2020)
32. Qin, Y., Lou, W., Wang, C., Gong, L., Zhou, X.: Enhancing long sequence input processing in fpga-based transformer accelerators through attention fusion. In: Proceedings of the Great Lakes Symposium on VLSI 2024. pp. 599–603 (2024)
33. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.H.: Single image dehazing via multi-scale convolutional neural networks. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. pp. 154–169. Springer (2016)
34. Song, Y., He, Z., Qian, H., Du, X.: Vision transformers for single image dehazing. IEEE Transactions on Image Processing **32**, 1927–1941 (2023)

35. Ullah, H., Muhammad, K., Irfan, M., Anwar, S., Sajjad, M., Imran, A.S., de Albuquerque, V.H.C.: Light-dehazenet: a novel lightweight cnn architecture for single image dehazing. IEEE transactions on image processing **30**, 8968–8982 (2021)
36. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
37. Wang, C., Gong, L., Li, X., Zhou, X.: A ubiquitous machine learning accelerator with automatic parallelization on fpga. IEEE Transactions on Parallel and Distributed Systems **31**(10), 2346–2359 (2020)
38. Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., Ma, L.: Contrastive learning for compact single image dehazing. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10551–10560 (2021)
39. Wu, R.Q., Duan, Z.P., Guo, C.L., Chai, Z., Li, C.: Ridcp: Revitalizing real image dehazing via high-quality codebook priors. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 22282–22291 (2023)
40. Yang, J., Li, C., Zhang, P., Dai, X., Xiao, B., Yuan, L., Gao, J.: Focal attention for long-range interactions in vision transformers. Advances in Neural Information Processing Systems **34**, 30008–30022 (2021)
41. Yang, X., Xu, Z., Luo, J.: Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
42. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3194–3203 (2018)
43. Zhang, J., Cao, Y., Fang, S., Kang, Y., Wen Chen, C.: Fast haze removal for nighttime image using maximum reflectance prior. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7418–7426 (2017)
44. Zhao, S., Zhang, L., Shen, Y., Zhou, Y.: Refinednet: A weakly supervised refinement framework for single image dehazing. IEEE Transactions on Image Processing **30**, 3391–3404 (2021)
45. Zhu, Q., Mai, J., Shao, L.: Single image dehazing using color attenuation prior. In: BMVC. vol. 4, pp. 1674–1682. Citeseer (2014)