ShadowRefiner: Towards Mask-free Shadow Removal via Fast Fourier Transformer

Wei Dong[1] Han Zhou[1*] Yuqiong Tian[1] Jingke Sun[1] Xiaohong Liu[2] Guangtao Zhai[2] Jun Chen[1*]
[1]McMaster University [2]Shanghai Jiao Tong University
wdong1745376@gmail.com, {zhouh115, tiany86, sun409}@mcmaster.ca
{xiaohongliu, zhaiguangtao}@sjtu.edu.cn, chenjun@mcmaster.ca

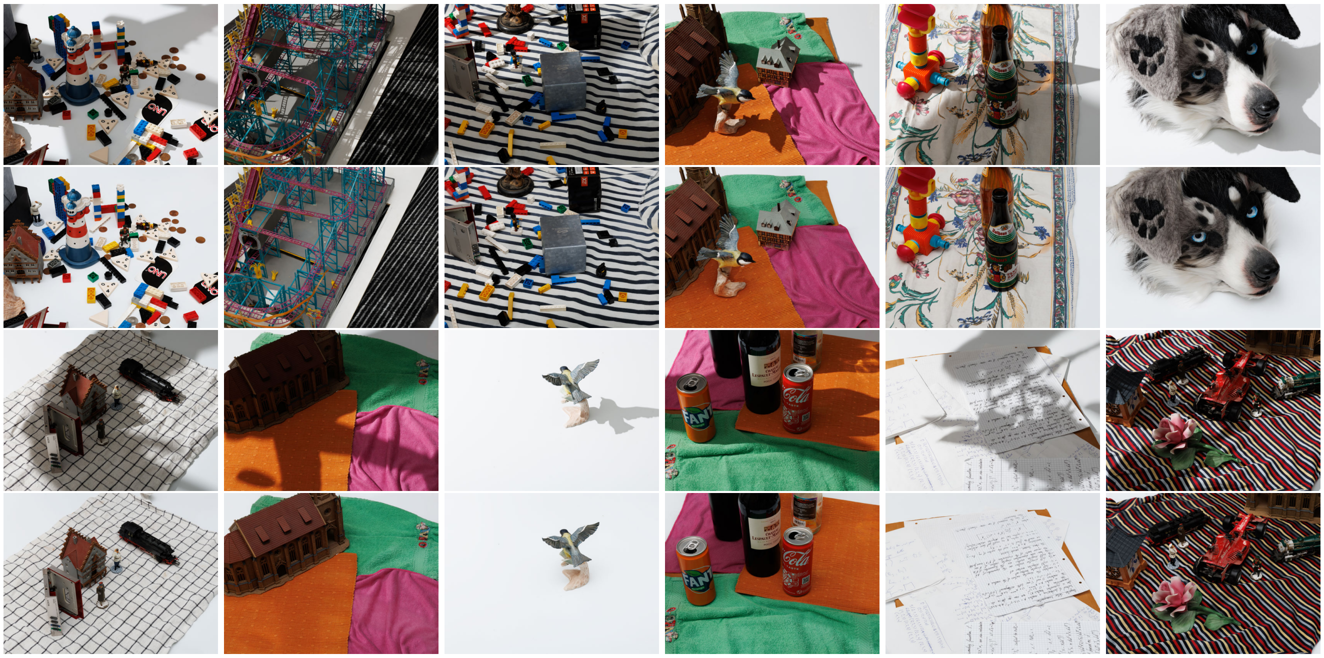

Figure 1. Our test results on NTIRE 2024 Image Shadow Removal Challenge [38]. Our proposed method is the **champion** in the **Perceptual Track** and achieves the second best performance in the Fidelity Track.

Abstract

*Shadow-affected images often exhibit pronounced spatial discrepancies in color and illumination, consequently degrading various vision applications including object detection and segmentation systems. To effectively eliminate shadows in real-world images while preserving intricate details and producing visually compelling outcomes, we introduce a mask-free **Shadow** Removal and **Refine**ment network (**ShadowRefiner**) via Fast Fourier Transformer. Specifically, the Shadow Removal module in our method aims to establish effective mappings between shadow-affected and shadow-free images via spatial and frequency representation learning. To mitigate the pixel misalignment and further improve the image quality, we propose a novel Fast-Fourier Attention based Transformer (FFAT) architecture, where an innovative attention mechanism is designed for meticulous refinement. Our method wins the championship in the Perceptual Track and achieves the second best performance in the Fidelity Track of NTIRE 2024 Image Shadow Removal Challenge. Besides, comprehensive experiment result also demonstrate the compelling effectiveness of our proposed method. The code is publicly available:* https://github.com/movingforward100/Shadow_R.

1. Introduction

Shadow-affected images typically emerge under scenarios where the light source is either partially or completely blocked, leading to spatial variations in color and illumina-

[*] Han Zhou and Jun Chen are corresponding authors

tion distortions. The objective of shadow removal is to enhance the visibility within shadow regions and achieve illumination consistency across both shadow and non-shadow areas, whilst preserving the integrity of naturalistic details. Such enhancement is pivotal for improving the performance of a plethora of downstream applications such as object detection, tracking, and segmentation systems [7, 15, 34, 44].

Numerous traditional methodologies proposed for image shadow removal are predominantly designed around physics-based illumination models [12, 17]. Despite their theoretical underpinnings, these approaches generally exhibit limited effectiveness of removing shadows from real-world, shadow-affected images. This limitation largely stems from the difficulty in establishing an accurate physical correlation between shadow areas and their unblemished counterparts, rendering these traditional techniques less effective in practical scenarios.

Recently, learning-based approaches have emerged as a formidable mainstream within the domain of shadow removal, capitalizing on the substantial modeling capability inherent in deep learning frameworks [23, 24, 41, 46]. These methodologies can be bifurcated into mask-based [31] and mask-free shadow removal strategies, contingent upon their dependence on shadow masks for guidance. Compared to the latter, mask-based shadow removal strategies not only employ pairs of shadow-affected and shadow-free images but also integrate the location information of shadow regions, either provided by benchmark datasets or generated through pre-trained mask prediction models, as learning guidance. The introduction of precise shadow location enables these models to concentrate on unraveling the complex mappings between shadow regions and their clean counterparts, thereby achieving exceptional performance on shadow removal.

Nonetheless, the dependency on mask information unveils critical challenges to mask-based methods. **Firstly**, the acquisition of precise shadow masks is notably challenging [27]. Accurate shadow masks provided in public datasets are typically obtained under simply scenarios, such as a single person standing in a wide-open square. In contrast, for complex scenes, as exemplified by the WSRD and WSRD+ datasets [36], annotating images with appropriate masks or employing pre-trained models to predict accurate masks proves to be impractical. **Secondly**, the absence of precise shadow masks markedly undermines the performance of mask-based models, substantially hindering their applicability to complex real-world data.

Latest advancements in mask-free shadow removal methodologies often leverage generative strategies [48] to learn the mappings between shadow-affected and shadow-free images. However, the adoption of frequency domain analysis remains largely under-explored within the sphere of shadow removal studies. Notably, several innovative works that integrate spatial and frequency representations has shown promising results in the broader field of image restoration, such as dehazing and low-light image enhancement task [19, 51], suggesting a potential direction for future exploration for shadow removal.

In this paper, we introduce a novel mask-free model that integrates spatial and frequency domain representations for image shadow removal, which achieves compelling performance on NTIRE 2024 Image Shadow Removal Challenge [38], as illustrated in Fig. 1. Specifically, we propose a **Shadow** Removal and **Refine**ment architecture, termed **ShadowRefiner**, with two specific modules: Shadow Removal module and Refinement module. In the **Shadow Removal** module, we design a shadow removal U-Net [32] branch with the backbone of ConvNext [28] blocks. Besides, a frequency branch similar to [51] equipped with high-frequency, low-frequency representations, and large receptive field, is also leveraged in our Shadow Removal module. Preliminary experiment results, however, indicate obvious pixel misalignment between the output of Shadow Removal module and the ground truth, manifesting as pronounced detail deterioration and compromised color consistency. To this end, we introduce a Fast-Fourier Attention based Transformer (FFAT) as the **Refinement** module, distinguished by its innovative attention mechanism, which significantly enhances the model's capacity to remove shadows while producing results that are simultaneously high in fidelity and visually appealing.

Our contributions are three-folds:

◇ We introduce an innovative mask-free shadow removal approach that initially clear shadow via spatial and frequency representation learning and further refined by our proposed frequency attention based transformer architecture .

◇ To mitigate the pixel misalignment, we introduce a Fast Fourier Transformer network endowed with a novel frequency attention mechanism, achieving superior performance on recovering texture details and maintaining color consistency.

◇ Extensive experiments across multiple shadow removal benchmarks, as well as the highly competitive outcomes achieved in the NTIRE 2024 Image Shadow Removal Challenge (ranking first and second in the Perceptual Track and Fidelity Track, respectively), underscore the remarkable performance of our proposed **ShadowRefiner** model.

2. Related Work

Mask-based Image Shadow Removal. Mask-based methods can be categorized into two main types: feature-based and deep learning-based. Feature-based methods rely on techniques such as color constancy [50], texture analysis and edge detection [2, 43]. Gryka *et al.* [14] first intro-

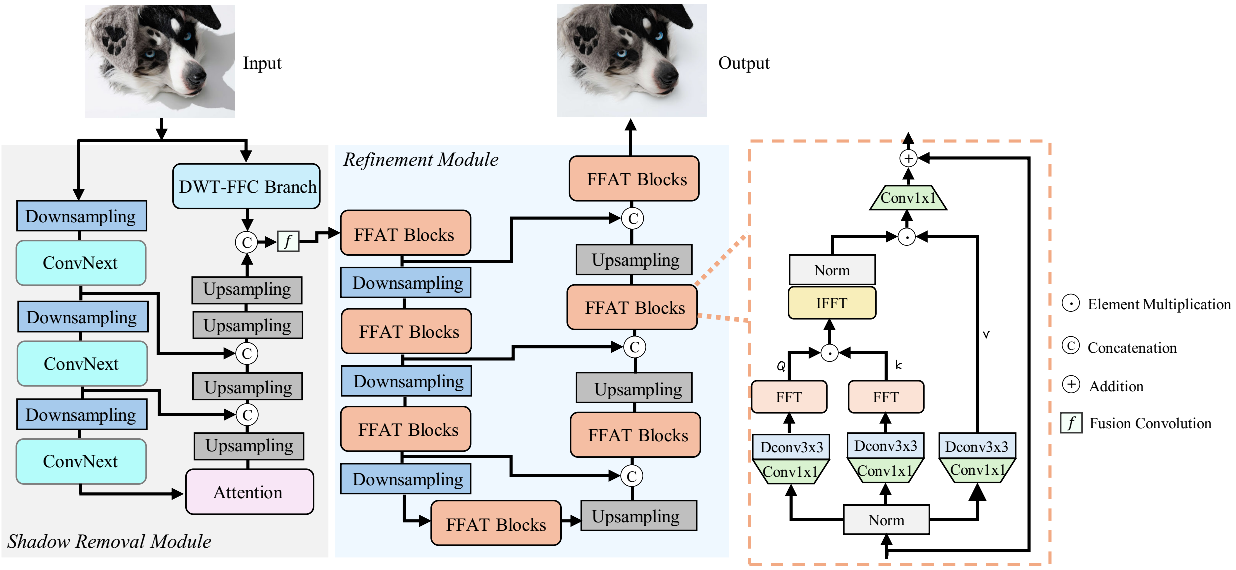

Figure 2. The overall architecture of our model. In the Shadow Removal module, besides the DWT-FFC branch proposed in [1, 51], we design a ConvNext-based U-Net architecture with 7 × 7 depth-wise convolution in each resolution. In the Refinement module, we design a new attention mechanism (Fast-Foruier Attention, FFA) different from common attention operations in transformers to further enhance texture details.

duce a learning-based method using a supervised regression algorithm to automatically remove umbra and penumbra shadows. Wang *et al*. [40] propose ST-CGAN, an end-to-end framework that integrates shadow detection and removal using two stacked Conditional Generative Adversarial Networks (CGANs). Bao *et al*. [3] propose S2Net that emphasizes semantic guidance and refinement for image integrity. This method uses shadow masks to guide shadow removal, with semantic-guided blocks transferring data from non-shadow to shadow areas, effectively eliminating shadows while preserving clean regions. He *et al*. [18] design Mask-ShadowNet, which ensures global illumination consistency through Masked Adaptive Instance Normalization (MAdaIN) and adaptively refines features using aligner modules. Additionally, Fu *et al*. [11] introduce FusionNet, which generates fusion weight maps to eliminate shadow traces further using a boundary-aware RefineNet. However, these methods heavily rely on the accuracy of the input shadow masks. The complexity and variability of real-world scenarios could pose challenges in generating precise shadow masks, potentially affecting the performance of these methods in practical applications.

Mask-free Image Shadow Removal. Fan *et al*. [9] introduce an end-to-end deep convolutional neural network consisting of an encoder-decoder network for predicting the shadow scale factor and a small refinement network for enhancing edge details. Chen *et al*. [5] design a CANet which utilizes a two-stage process for shadow removal, employing a Contextual Patch Matching (CPM) module to identify matching pairs between shadow and non-shadow patches and a Contextual Feature Transfer (CFT) mechanism to transfer contextual information, effectively eliminating shadow influence. Vasluianu *et al*. [37] introduce Ambient Lighting Normalization (ALN) to improve image restoration under complex lighting and propose IFBlend that enhances images by maximizing Image-Frequency joint entropy without relying on shadow localization. Liu *et al*. [25] propose a shadow-aware decomposition network to separate illumination and reflectance layers, followed by a bilateral correction network for lighting adjustment and texture restoration.

Transformer-based Image Restoration. Transformer based networks, usually adopt self-attention mechanisms to understand the relationships between different components and demonstrate high superiority in handling long dependencies, have shown state-of-the-art performance on image restoration. SwinIR [22], a famous backbone for image restoration, is designed based on several residual Swin Transformer [26] blocks. With the backbone of Vision Transformer [39], DehazeFormer [35] is proposed for dehazing task. Recently, a lightweight transformer architecture [4] is proposed for low-light image enhancement based on Retinex theory.

3. Methods

3.1. ConvNext based U-Net for Shadow Removal

In order to achieve satisfactory shadow removal performance, powerful deep-learning networks are pivotal to extracting important features from shadow-affected images

Methods	**Mask-free**	ISTD [40]			ISTD+ [21]			WSRD+ [36]		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
DHAN [6]	No	24.86	0.919	0.0535	27.88	0.917	0.0529	22.39	0.796	0.1049
BMNet [52]	No	29.02	0.923	0.0529	31.85	0.932	0.0432	24.75	0.816	0.0948
FusionNet [11]	No	25.84	0.712	0.3196	27.61	0.725	0.3123	21.66	0.752	0.1227
SADC [45]	No	29.22	0.928	0.0403	—	—	—	—	—	—
ShadowFormer [16]	No	30.47	0.928	0.0418	32.78	0.934	0.0385	25.44	0.820	0.0898
DCShadowNet [20]	Yes	24.02	0.677	0.4423	25.50	0.694	0.4237	21.62	0.593	0.4744
Refusion [29]	Yes	25.13	0.871	0.0571	26.28.	0.887	0.0437	22.32	0.738	0.0937
ShadowRefiner (Ours)	Yes	28.75	0.916	0.0521	31.03	0.928	0.0426	26.04	0.827	0.0854

Table 1. Quantitative comparisons with SOTA methods. Our ShadowRefiner significantly outperforms other mask-free methods across three benchmarks. Compared to mask-based methods, our ShadowRefiner achieves comparable or even better performance (WSRD+ dataset). [Key: Best performance among mask-free models, Best performance among mask-based methods]

and modeling the mapping from shadow-affected and clean images. In this work, we introduce a ConvNext-based U-Net architecture, where multi-scale ConvNext blocks function as strong encoders for robust latent feature learning.

As shown in Fig. 2, the ConvNext-based U-Net serves as the primary component in the Shadow Removal module, and the DWT-FFC branch [8, 51] is incorporated as an auxiliary branch. The contributions of each branch is provided in Sec. 4.4.

Specifically, our ConvNext-based U-Net encompasses three downsampling layers and each downsampling operation is followed by several ConvNext blocks to feature extraction. Given a latent feature $\mathbf{F}_{in}$, the ConvNext block first adopts a 7×7 depthwise convolution, which functions similar to the self-attention mechanism in Transformers. Then the Layer Normalization (LN) is applied before two 1×1 convolutions, which are equivalent to MLP block in Transformer. Besides, only one GELU function is leveraged between two 1×1 convolutions inspired by the fact that Transformer MLP block incorporates only one activation function.

For the decoding process, the latent feature is aggregated using one attention block [30] and several upsampling operations are utilized to recover the latent feature to their original resolution. In each upsampling layer, there are one pixel-shuffle operation and one attention block. Besides, encoder features are transferred to the decoding process via skip-connection.

In Stage I, we only optimizing the Shadow Removal module and the training objective is shown as below:

$$L_{loss} = L_1 + \alpha L_{SSIM} + \beta L_{Percep} + \gamma L_{adv} \tag{1}$$

where L_1, L_{SSIM} and L_{Percep} represent the $L1$ loss, MS-SSIM loss [51], and perceptual loss [33], respectively. In addition, we leverage the discriminator proposed in [13] (not provided in Fig. 2) to calculate the adversarial loss (L_{adv}). α, β, and γ are set to 0.2, 0.01, and 0.0005 for the optimization.

3.2. Fast-Fourier Attention Transformer based Refinement

Early experimental results suggest that our proposed ConvNext-based U-Net can effectively remove shadows, but distinct shadow contours remain, as illustrated in Fig. 6. In order to further refine image details and maintain color consistency, we introduce an efficient transformer architecture with a novel frequency attention mechanism, as shown in Fig. 2. Similar to [47], our Refinement module adopts a encoder-decoder architecture, and our proposed FFAT blocks are utilized at each resolution level in both encoding and decoding process.

Given a latent feature $\mathbf{F}$, FFAT block first utilize 1×1 point-wise convolution and 3×3 depth-wise convolution to generate three features: $\mathbf{Q}, \mathbf{K}, \mathbf{V}$. Instead of directly leveraging these features for attention calculation, we apply Fast Fourier Transform to $\mathbf{Q}, \mathbf{K}$ to calculate their frequency correlation $\mathbf{A}_F$ based on their frequency domain representations ($\mathcal{F}(\mathbf{Q})$, $\mathcal{F}(\mathbf{K})$) by:

$$\mathbf{A}_F = \mathcal{F}(\mathbf{Q})\mathcal{F}(\mathbf{K})', \tag{2}$$

where $\mathcal{F}(\cdot)$ represents the FFT process and $(\cdot)'$ denotes the transpose operation. Then, the spatial correlation between $\mathbf{Q}$ and $\mathbf{K}$ can be obtained by the inverse FFT operation ($\mathcal{F}^{-1}(\cdot)$) and a layer normalization ($\mathcal{LN}$). Finally, the aggregated feature $\mathbf{F}_A$ and the final output of our FFAT block $\mathbf{F}_{out}$ can be estimated as:

$$\begin{aligned} \mathbf{F}_A &= \mathcal{LN}(\mathcal{F}^{-1}(\mathbf{A}_F))\mathbf{V} \\ \mathbf{F}_{out} &= \mathrm{Conv}_{1\times1}(\mathbf{F}_A) + \mathbf{F}. \end{aligned} \tag{3}$$

Compared to the global attention mechanism in Transformer and other attention strategy [10], our FFAT blocks can not only effectively capture long-dependencies, but also demonstrate stronger representation learning with high efficiency. This advantage stems from the frequency domain feature learning and frequency attention calculation process. To train our FFAT-based Refinement module, we first

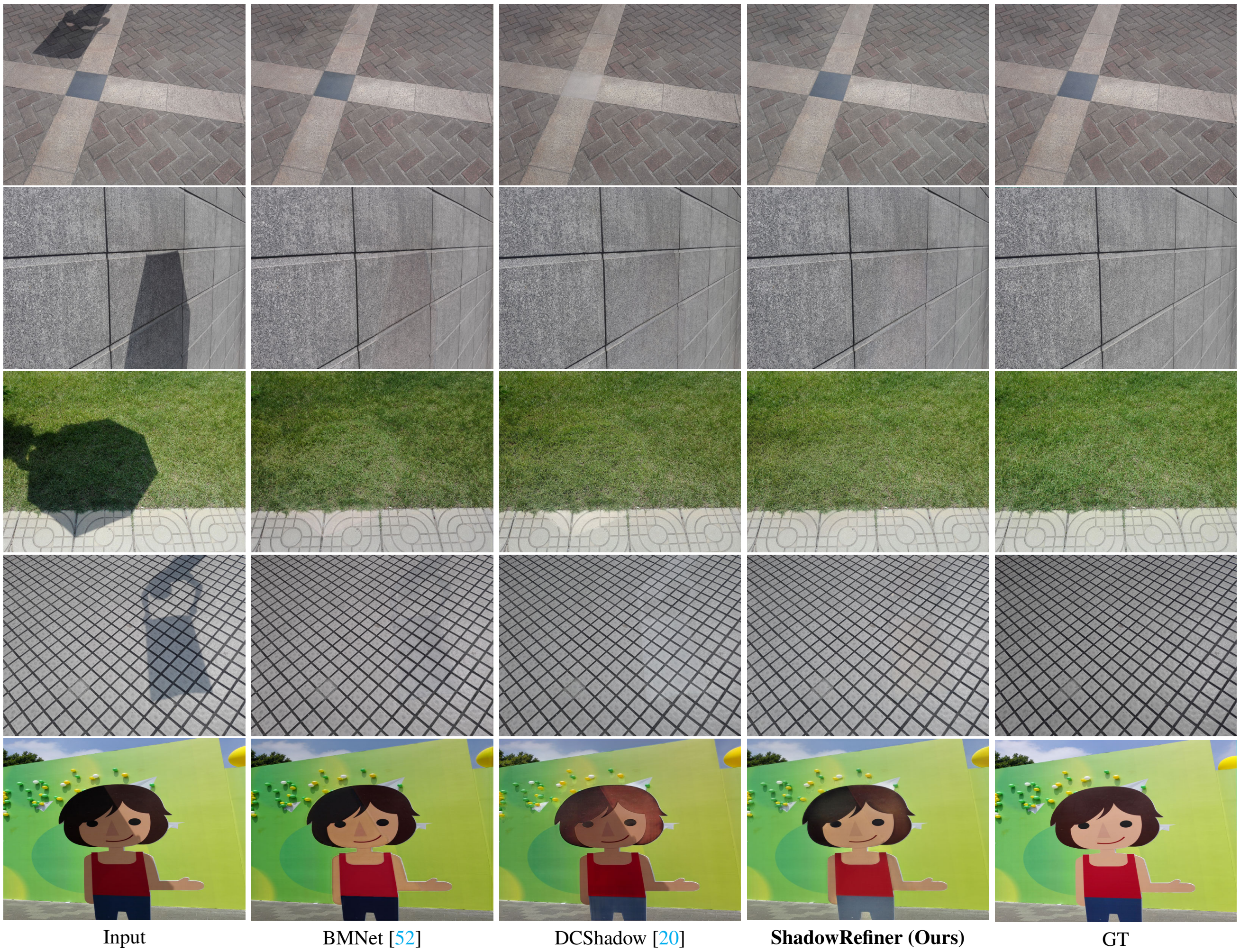

Figure 3. Visual comparisons on the ISTD dataset [40]. Compared to other methods, our ShadowRefiner successfully remove shadows without incorporating artifacts.

fix the Shadow Removal module and remove the adversarial loss in Eq. 1 for optimization.

4. Experiments

4.1. Datasets and Implementation Details

Datasets. We evaluate our proposed method on WSRD+ [36], ISTD [40], and ISTD+ [21] datasets. WSRD+ dataset, as the enhanced version of the WSRD dataset with improved pixel-alignment, is used as the benchmark dataset for NTIRE 2024 Image Shadow Removal Challenge. This dataset consists of 1200 high-resolution image pairs. The training set, validation set, and test set are split in proportions of 10:1:1, and we train our model on the training set and evaluate the performance on the validation set. ISTD dataset contains 1870 image triplets obtained from 135 distinct scenarios, of which 1330 are assigned for training and the remaining 540 are for testing. ISTD+ dataset is a color-adjusted version of ISTD and it has the same number and structure as the ISTD dataset.

Implementation Details. One RTX 2080Ti GPU is used to execute the two-stage training of our method. For data augmentation, we implement random cropping of patches with dimensions of 384 × 384, combined with random rotations of 90, 180, or 270 degrees, as well as vertical and horizontal flipping. The Adam optimizer with the default hyper-parameters, where β_1 and β_2 are set to 0.9 and 0.999 respectively, is utilized for optimization. In Stage I, only Shadow Removal module is optimized and the learning rate is initially set to 1×10^{-4} and is gradually reduced to 6.25×10^{-6}. In Stage II, we adopt a constant learning of 1×10^{-5} to simultaneously update the Shadow Removal module and Refinement module.

Evaluation Metrics. To comprehensively evaluate the performance of various shadow removal methods, three metrics are adopted for quantitative comparison: The Peak Sig-

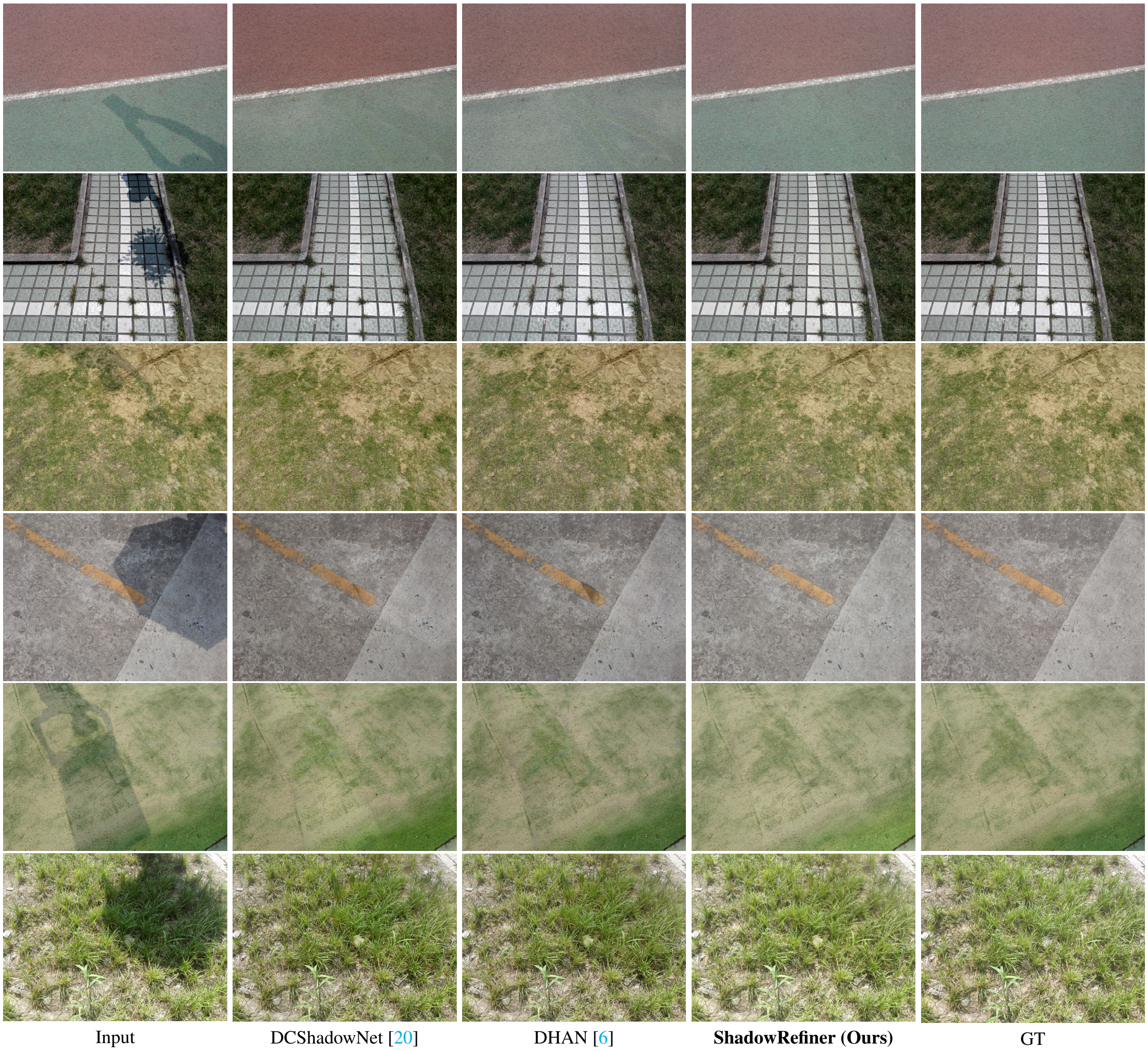

Figure 4. Visual comparisons on the ISTD+ dataset[21]. Obviously, our ShadowRefiner excels in maintaining color consistency and recovering structure details

nal to Noise Ratio (PSNR), the Structural Similarity Index (SSIM) [42], and Learned Perceptual Image Patch Similarity (LPIPS) [49] to assess the pixel-level accuracy and perceptual quality.

4.2. Comparison with State-of-the-Art Methods

We compare our proposed method with several State-of-the-art (SOTA) algorithms. Specifically, several mask-free methods including Refusion [29], DCShadowNet [20] and recent proposed mask-based approaches including ShadowFormer [16], SADC [45] are adopted for comparison.

Quantitative Results. As documented in Tab. 1, our ShadowRefiner demonstrates superior performances compared to other mask-free methods across three datasets. On the ISTD dataset [40], there is a marked improvement of 3.62 dB in PSNR, a 0.045 increase in SSIM, and a 0.005 decline in LPIPS. On the ISTD+ dataset [21], we observe a 4.75 dB increase in PSNR and a 0.041 increase in SSIM. Besides, our ShadowRefiner yields a 3.72 dB increase in PSNR, a 0.089 increase in SSIM, and a 0.008 decline in LPIPS on the WSRD+ dataset [38]. Furthermore, compared to mask-based methods, which require shadow mask information and naturally outperform mask-based models, our mask-free ShadowRefiner is capable to generate com-

Team	PSNR↑	SSIM↑	LPIPS↓	MOS↑	Fianl Rank↓
Shadow_R (Ours)	24.578	**0.832**	0.098	**7.750**	**1**
LVGroup_HFUT	24,232	0.821	**0.082**	7.519	2
USTC_ShadowTitan	24.042	0.827	0.104	7.444	3
ShadowTech Innovators	**24.810**	0.832	0.111	7.438	4
GGBond	23.050	0.809	0.089	7.400	5
PSU Team	22.219	0.731	0.132	7.400	6
LUMOS	24.783	0.832	0.110	7.163	7
IIM_TTI	22.955	0.806	0.093	7.160	8
AiRiA_Vision	21.902	0.689	0.238	6.825	9
HKUST-VIP_Lab_01	22.284	0.788	0.135226084	6.619	10

Table 2. Final ranking (top 10 teams) of Perceptual Track in NTIRE 2024 Shadow Removal Challenge. Our solution achieves the best performance among all 19 submitted solutions. [Key: **Best**, Second Best, ↑ (↓): The larger (smaller) represents the better performance].

parable results on ISTD and ISTD+ datasets. Moreover, on WSRD+ dataset where precise mask image is unavailable, though a mask prediction method proposed in [6] is utilized to assist these mask-based models, our ShadowRefiner achieves superior performance than the best mask-based approach (ShadowFormer [16]), demonstrating that our ShadowRefiner can work effectively in complex scenarios.

Qualitative Comparisons. Visual comparisons on ISTD dataset, ISTD+ dataset and WSRD+ dataset are reported in Fig. 3, 4, and 5, respectively. Obviously, the results of our method closely match GT images in both color and detail preservation. For instance, in Fig. 3 row 3, our ShadowRefiner removes the shadow without introducing artifacts or discoloration, issues commonly seen with other methods. In Fig. 4 row 2, our method skillfully adjusts the shadow areas on the pavement, effectively lightening the shadows to blend with the sunlit parts without distorting the underlying patterns or hues. Moreover, for images featuring toys and colorful objects shown in Fig. 5, our ShadowRefiner restores the vivid colors and intricate patterns that are obscured by shadows in the input images.

4.3. Performance on NTIRE 2024 Shadow Removal Challenge (Fidelity and Perceptual Track)

According to the challenge report [38], our model is the **first place** of **Perceptual Track** and the **second place** of **Fidelity Track** with the highest SSIM (0.832) and Mean Opinion Scor (MOS, 7.750) and competitive PSNR (24.58), demonstrating advanced performance of our method on shadow removal task. We also report the result of our method for the official validation and test data used in NTIRE 2024 Shadow Removal Challenge as Fig. 1 and Fig. 5. We can see that our ShadowRefiner achieves notably satisfying shadow removal effect in pictures of different scenarios. For example, the shadow in the first image in Fig. 1 is completely eliminated, retaining the integrity of the scattered toy blocks in their original layout. Additionally, the color consistency between toys initially positioned in shadowed areas and those under direct illumination attests to the high fidelity in color reproduction. The restoration of the final image depicting a dog conveys a sense of uniform, soft lighting across the entire subject, further exemplifying ShadowRefiner's capability to enhance illumination homogeneity.

Configurations	PSNR↑	SSIM↑	LPIPS↓
ShadowRefiner (Ours)	**26.04**	**0.827**	**0.0854**
w/o Refinement module	25.54	0.816	0.0886
w/ Restormer module [47]	25.68	0.818	0.0878
only ConvNext-based U-Net	25.29	0.810	0.0914
only DWT-FFC branch	23.36	0.791	0.1016

Table 3. The ablation results on WRSD+ dataset. Each component in our ShadowRefiner help achieve competitive performance on shadow removal task, and our proposed Refinement module performs better than Restormer.

4.4. Ablation Study

In this section, we conduct several ablation experiments on WSRD+ dataset [36].

Importance of FFAT based Refinement module. To study the contribution of the Refinement module and our proposed FFAT blocks, we first remove the Refinement module to figure out its contribution for the outstanding performance of our method provided in Tab. 1. The quantitative result is reported in Tab. 3, which demonstrates removing Refinement module leads to obviously degraded performance. Then, we replace our Refinement module with Restormer module [47] and we notice a marked decrease in PSNR (0.36 dB ↓) and SSIM (0.009 ↓) and a discernible increase of LPIPS (0.0024 ↑). Visual comparisons provided in Fig. 6 also demonstrate our Refinement module can help generate more visually appealing results compared to the Restormer module. This ablation experiment underscore the importance of our proposed Refinement module for satisfactory shadow removal performance.

Contributions of ConvNext-based U-Net. Based on the

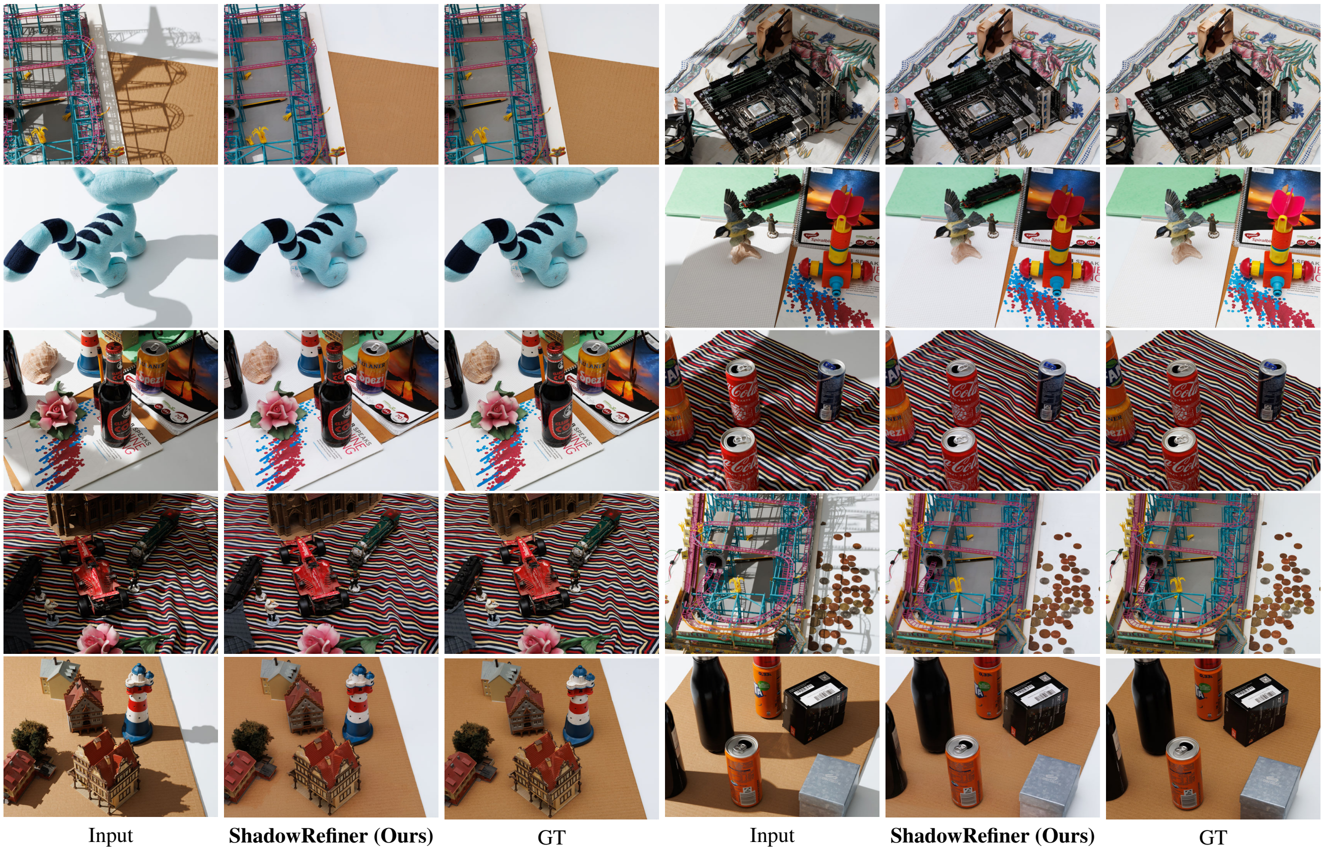

Figure 5. Our results on the validation set of WSRD+ dataset [36]. Our method performs well on color and detail recovery.

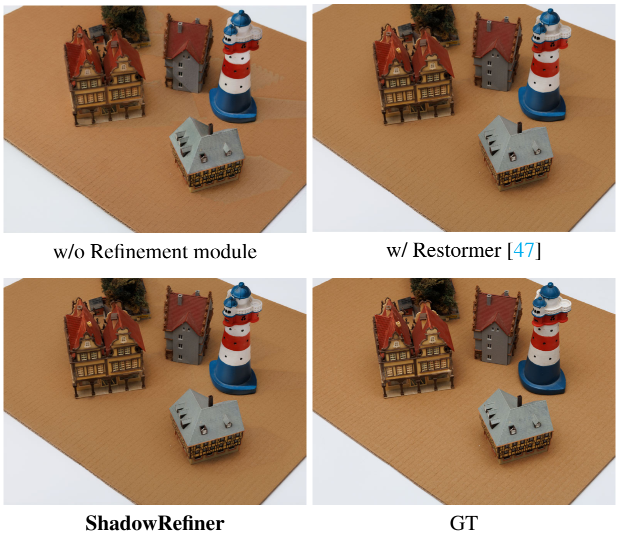

Figure 6. Visual ablation comparisons on the WSRD+ dataset.

configuration for Tab. 3 row 2, we implement several further adaptations to illustrate the effectiveness of ConvNext-based U-Net and the DWT-FFC branch leveraged in the Shadow Removal module. By separately comparing row 4 and row 5 to row 2 in Tab. 3, we can conclude that adopting a two-branch architecture in the Shadow Removal module help achieve more pleasant performance than single branch. Moreover, compared to the DWT-FFC branch, the enhanced performance of the ConvNext-based U-Net architecture highlights its predominant role within the Shadow Removal module.

5. Conclusion

In this paper, we propose **ShadowRefiner**, a novel mask-free model for shadow removal task. Specifically, the Shadow Removal module with ConvNext-based U-Net is firstly introduced to extract spatial and frequency representations and effectively learning the mapping between shadow-affected and clean images. Then, we design a novel transformer module based on Fast Fourier Attention Transformer to enhance color and structure consistency. Extensive experiments demonstrate ShadowRefiner significantly outperforms the current mask-free methods and its capacity is comparable to mask-based shadow removal approaches. Furthermore, our method wins the **championship** in the Perceptual Track and **ranks second** in the Fidelity Track of NTIRE 2024 Image Shadow Removal Challenge.

References

[1] Codruta O Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, Radu Timofte, et al. NTIRE 2023 challenge on nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023.

[2] Eli Arbel and Hagit Hel-Or. Shadow removal using intensity surfaces and texture anchor points. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011.

[3] Qiqi Bao, Yunmeng Liu, Bowen Gang, Wenming Yang, and Qingmin Liao. S2net: Shadow mask-based semantic-aware network for single-image shadow removal. In *IEEE Transactions on Consumer Electronics*, 2022.

[4] 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.

[5] Zipei Chen, Chengjiang Long, Ling Zhang, and Chunxia Xiao. Canet: A context-aware network for shadow removal. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2021.

[6] Xiaodong Cun, Chi-Man Pun, and Cheng Shi. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2020.

[7] 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.

[8] Wei Dong, Han Zhou, Ruiyi Wang, Xiaohong Liu, Guangtao Zhai, and Jun Chen. DehazeDCT: Towards effective non-homogeneous dehazing via deformable convolutional transformer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, 2024.

[9] Hui Fan, Meng Han, and Jinjiang Li. Image shadow removal using end-to-end deep convolutional neural networks. In *Applied Sciences*, 2019.

[10] Kang Fu, Peng Yicong amd Zhang Zicheng, 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.

[11] Lan Fu, Changqing Zhou, Qing Guo, Felix Juefei-Xu, Hongkai Yu, Wei Feng, Yang Liu, and Song Wang. Auto-exposure fusion for single-image shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[12] Han Gong and Darren Cosker. Interactive removal and ground truth for difficult shadow scenes. In *JOSA*, 2016.

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

[14] Maciej Gryka, Michael Terry, and Gabriel J Brostow. Learning to remove soft shadows. In *ACM Trans. Graph.*, 2015.

[15] 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.

[16] Lanqing Guo, Siyu Huang, Ding Liu, Hao Cheng, and Bihan Wen. Shadowformer: Global context helps image shadow removal. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2023.

[17] Ruiqi Guo, Qiqyun Dai, and Derek Hoiem. Paired regions for shadow detection and removal. In *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2012.

[18] Shengfeng He, Bing Peng, Junyu Dong, and Yong Du. Mask-shadownet: Toward shadow removal via masked adaptive instance normalization. In *IEEE Signal Processing Letters*, 2021.

[19] Hai Jiang, Ao Luo, Songchen Han, Haoqiang Fan, and Shuaicheng Liu. Low-light image enhancement with wavelet-based diffusion models. In *Siggraph Asia*, 2023.

[20] Yeying Jin, Aashish Sharma, and Robby T. Tan. Dc-shadownet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2021.

[21] Hieu Le and Dimitris Samaras. Shadow removal via shadow image decomposition. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019.

[22] 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.

[23] 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.

[24] 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.

[25] Yuhao Liu, Zhanghan Ke, Ke Xu, Fang Liu, Zhenwei Wang, and Rynson W. H. Lau. Recasting regional lighting for shadow removal. *arXiv preprint arXiv:2402.00341*, 2024.

[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 Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[27] Zhihao Liu, Hui Yin, Xinyi Wu, Zhenyao Wu, Yang Mi, and Song Wang. From shadow generation to shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[28] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[29] Ziwei Luo, Fredrik K. Gustafsson, Zheng Zhao, et al. Refusion: Enabling large-size realistic image restoration with

latent-space diffusion models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, 2023.

[30] 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.

[31] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson W. H. Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[32] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, 2015.

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

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

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

[36] Florin-Alexandru Vasluianu, Tim Seizinger, and Radu Timofte. Wsrd: A novel benchmark for high resolution image shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, 2023.

[37] Florin-Alexandru Vasluianu, Tim Seizinger, Zongwei Wu, Rakesh Ranjan, and Radu Timofte. Towards image ambient lighting normalization. In *arXiv preprint arXiv:2403.18730*, 2024.

[38] Florin-Alexandru Vasluianu, Tim Seizinger, Zhuyun Zhou, Zongwei Wu, Cailian Chen, Radu Timofte, et al. NTIRE 2024 image shadow removal challenge report. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, 2024.

[39] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, and Łukasz Kaiserand Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[40] Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[41] 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.

[42] 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.

[43] Minghu Wu, Rui Chen, and Ying Tong. Shadow elimination algorithm using color and texture features. In *Computational Intelligence and Neuroscience*, 2020.

[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] Yimin Xu, Mingbao Lin, Hong Yang, Fei Chao, and Rongrong Ji. Shadow-aware dynamic convolution for shadow removal. In *Pattern Recognition*, 2024.

[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] 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.

[48] Ling Zhang, Chengjiang Long, Xiaolong Zhang, and Chunxia Xiao. Ris-gan: Explore residual and illumination with generative adversarial networks for shadow removal. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2020.

[49] 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.

[50] Yunfeng Zhao, Chris Elliott, Huiyu Zhou, and Karen Rafferty. Pixel-wise illumination correction algorithms for relative color constancy under the spectral domain. In *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 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] Yurui Zhu, Zeyu Xiao, Yanchi Fang, Xueyang Fu, Zhiwei Xiong, and Zheng-Jun Zha. Efficient model-driven network for shadow removal. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2022.