NTIRE 2024 Image Shadow Removal Challenge Report

Florin-Alexandru Vasluianu[†] Tim Seizinger[†] Zhuyun Zhou[†] Zongwei Wu[†]
Cailian Chen[†] Radu Timofte[†] Wei Dong Han Zhou Yuqiong Tian Jun Chen
Xueyang Fu Xin Lu Yurui Zhu Xi Wang Dong Li Jie Xiao
Yunpeng Zhang Zheng-Jun Zha Zhao Zhang Suiyi Zhao Bo Wang Yan Luo
Yanyan Wei Zhihao Zhao Long Sun Tingting Yang Jinshan Pan Jiangxin Dong
Jinhui Tang Bilel Benjdira Mohammed Nassif Anis Koubaa Ahmed Elhayek
Anas M. Ali Kyotaro Tokoro Kento Kawai Kaname Yokoyama Takuya Seno
Yuki Kondo Norimichi Ukita Chenghua Li Bo Yang Zhiqi Wu Gao Chen
Yihan Yu Sixiang Chen Kai Zhang Tian Ye Wenbin Zou Yunlong Lin
Zhaohu Xing Jinbin Bai Wenhao Chai Lei Zhu Ritik Maheshwari
Rakshank Verma Rahul Tekchandani Praful Hambarde Satya Narayan Tazi
Santosh Kumar Vipparthi Subrahmanyam Murala Jaeho Lee Seongwan Kim
Sharif S M A Nodirkhuja Khujaev Roman Tsoy Fan Gao Weidan Yan
Wenze Shao Dengyin Zhang Bin Chen Siqi Zhang Yanxin Qian Yuanbin Chen
Yuanbo Zhou Tong Tong Rongfeng Wei Ruiqi Sun Yue Liu Nikhil Akalwadi
Amogh Joshi Sampada Malagi Chaitra Desai Ramesh Ashok Tabib
Uma Mudenagudi Ali Murtaza Uswah Khairuddin Ahmad 'Athif Mohd Faudzi
Adinath Dukre Vivek Deshmukh Shruti S. Phutke Ashutosh Kulkarni
Santosh Kumar Vipparthi Anil Gonde Subrahmanyam Murala Arun karthik K
Manasa N Shri Hari Priya Wei Hao Xingzhuo Yan Minghan Fu

Abstract

This work reviews the results of the NTIRE 2024 Challenge on Shadow Removal. Building on the last year edition, the current challenge was organized in two tracks, with a track focused on increased fidelity reconstruction, and a separate ranking for high performing perceptual quality solutions. Track 1 (fidelity) had 214 registered participants, with 17 teams submitting in the final phase, while Track 2 (perceptual) registered 185 participants, resulting in 18 final phase submissions. Both tracks were based on data from the WSRD dataset, simulating interactions between self-shadows and cast shadows, with a large variety of represented objects, textures, and materials. Improved image alignment enabled increased fidelity reconstruction, with restored frames mostly indistinguishable from the references images for top performing solutions.

1. Introduction

Shadow Removal enjoyed significant attention in the research community. The shadow formation model is, by its nature, describing a very complex phenomenon. The shadow intensity and shape is determined by a multitude of factors, starting with the properties of the light source, continuing with the geometry and characteristics of the shadow casting light occluder, then being finally determined by the properties of the surface upon which the shadow is being cast. Such a complex system requires extensive study, with shadow removal and/or detection remaining an active research field.

Early works are naturally based on the physical properties of a determined shadow formation model, characteristic to a (usually restrained) set of conditions, under which measurements can be successful. A category of shadow removal solutions, based on classical image processing pipelines, aimed at successfully recovering the shadow free image by

[†] Florin-Alexandru Vasluianu, Tim Seizinger, Zhuyun Zhou, Zongwei Wu, Cailian Chen, and Radu Timofte are the NTIRE 2024 challenge organizers. The other authors participated in the challenge.

Appendix.A contains the authors' team names and affiliations.
https://cvlai.net/ntire/2024

transferring local statistics from the image segment not affected by shadows, to the affected image area [19, 20]. This is a particularly difficult strategy, since it involves shadow detection (which itself is an equally difficult problem) as a sub-task. However, solutions based on the localization information regarding the shadow affected segment remain a popular design choice, with various algorithms [57, 65, 71], acknowledging the importance of a reliable shadow detection mask.

Later, hardware and software development enabled the growth of the deep-learning solutions, supported by the introduction of large scale image databases for shadow removal and/or detection [23, 31, 53, 61, 66]. This enabled a large group of deep learning solutions solving shadow removal as a particular image restoration sub-task, being developed either in the fully-supervised training framework [25, 40, 41, 53, 61, 66, 76], or using weakly supervised or self-supervised settings [18, 31, 33, 60].

Lately, the increased popularity diffusion models [30] appeared as backbone in various solutions for image shadow removal [26, 34]. Naturally, being characterized by top notch performance in terms of perceptual quality, this group of recently introduced solutions represent a reference in the field, on the currently established benchmarks, such as ISTD [66], ISTD+ [41, 66], or SRD [53].

All the aforementioned solutions represented steps forward in the shadow removal field, but, unfortunately, the studied shadow formation model is limited at the subset of scenarios represented at the data level. While LRSS [23] is a large collection of images representing various soft shadows, both SRD [53] and ISTD [66] are focusing on hard shadows, but targeting a simplified scenario. The acquisition setup is based on the natural light as the only light source, and a light occluder object. This object does not appear in the acquired image, but is the one casting a shadow in the photographed scene. Removing the light occluder enables the acquisition of the reference image. The usage of natural light as the only light source is necessary such that the color consistency between the input and the reference image is achieved (up to some degree), in the shadow free segment. A color correction method was proposed in [41], resulting in the (adjusted) ISTD+ and SRD+ dataset variations. This eliminates the color inconsistencies between images from the same pair, but various semantic inconsistencies still remain [33, 60].

However, the main limitation of this setup comes in the types of surfaces the shadow can be cast on. These surfaces have to be flat, to avoid any self-shadows appearing in the reference images. However, self-shadows are the most common type of natural shadows, so studying them is crucial for real domain applications, outside scenario-specific prototypes.

To circumvent this, in WSRD [61], a softbox-based laboratory setup was used to acquire images covering the interactions between self-shadows cast by rough surfaces and complex geometries, and outside cast shadows, characteristic to a fixed directional light. A set of softbox lights were used for near-optimal lighting distribution, acquiring the reference image with minimal self shadows. Recently, the setup was extended [63], with multiple directional lights, thus dropping the color consistency in the shadow free segment. This extends the study to a broader task, introducing an image database for the connected lighting normalization problem.

2. WSRD+

The WSRD [61] data used in the NTIRE 2023 Image Shadow Removal Challenge [62] represented a step ahead in studying more complex shadow interactions, with increased complexity surfaces and represented contents. However, a component of the setup was represented by the moving objects serving as shadow casting surfaces. This introduced a certain pixel-alignment inconsistency, that proved to be challenging as for the last year feedback offered by the participating teams.

In this edition, we performed a preliminary image alignment based on a homography estimation step, increasing the consistency between input and reference images by 1.5-2dB in terms of PSNR. This enabled increased quality submissions, with improved quality restored images, showing both in the quantified reconstruction fidelity or image quality properties. The improved alignment allowed the participating teams to focus on the shadow removal algorithms, with the previous edition winners investing significant effort in high-performing supervision schemes [62].

Moreover, given the limitations of the Codalab [52] server used for validation, 25 samples from the testing split were removed, keeping the most challenging samples, without modifying the distribution in terms of represented scenarios, contents, or crucially, the training unseen objects or surfaces. This resulted in a final submission of 75 images, with the same resolution as they were released in WSRD [1].

This challenge is one of the NTIRE 2024 Workshop [2] associated challenges on: dense and non-homogeneous dehazing [1], night photography rendering [2], blind compressed image enhancement [72], shadow removal [64], efficient super resolution [54], image super resolution (×4) [8], light field image super-resolution [69], stereo image super-resolution [68], HR depth from images of specular and transparent surfaces [73], bracketing image restoration and enhancement [77], portrait quality assessment [4], quality assessment for AI-generated content [48], restore any image model (RAIM) in the wild [46], RAW image

[1] The testing and validation splits are publicly available under https://github.com/fvasluianu97/WSRD-DNSR

[2] https://cvlai.net/ntire/2024/

super-resolution [12], short-form UGC video quality assessment [45], low light enhancement [49], and RAW burst alignment and ISP challenge.

3. Evaluation

This edition of the challenge was organized as a double track competition, with different ranking criteria for both tracks. As follows, we list all the criteria used for evaluation.

1. The reconstruction fidelity in terms of PSNR values;
2. The Structured Similarity Index (SSIM) [70] score;
3. The LPIPS [75] distance between the restorations and the ground-truth images. We used the ImageNet pretrained AlexNet [37] for LPIPS feature extraction;
4. The Mean Opinion Score (MOS) for the submitted predictions;
5. An efficiency metric based on the number of parameters reported by each team.

For the Fidelity Track, we considered the first three criteria with equal weighting. The fifth was used only to differentiate between solutions characterized by similar performance, favouring lower complexity methods. For the Perceptual Track, the main metric considered was the Mean Opinion Score (MOS), as a result of an user study performed after the Final Phase deadline.

4. Challenge Phases

1. **Development phase:** In this phase, the participants were granted access to the task description, alongside with a set of 1000 image pairs to train their models;
2. **Validation phase:** To validate their solutions, the participants got a set of 100 images as the input images from the validation split of WSRD+. The ground-truth images were not made public, but a Codalab validation server [52] was set up, comparing the submission images uploaded by each user to the private reference images.
3. **Final phase:** A set of 75 input images as the testing split of WSRD+ was sent to the challenge participants. They could validate the solutions using the Codalab server. Any test-set fine-tuning was limited through a low number of submissions per user. Finally, a submission template was provided, with instructions for the final submission preparation, each team providing a method description, the corresponding codes, and information regarding the team members and their affiliation, alongside a set of 75 restored images, corresponding to the testing split inputs.

5. User Study

The primary criteria for ranking the solutions submitted to the Perceptual Track of the challenge was the Mean Opinion Score (MOS). This was determined as the result of an user study, in which the images were analyzed by imaging experts, including professional photographers. The grading system is set such that the input image corresponds to a grade of 3. Grades 1 and 2 are kept for situations in which the algorithm degrades the input image, without improvement in terms of shadows. For algorithms successful in terms of output restored images, the range 3-10 was used, with increments of 0.5.

The user study was performed on a subset of challenging samples from the WSRD+ test split, with the index i in the subset $i \in \{2, 4, 11, 14, 38, 39, 53, 54, 57, 69\}$ given the 0-based indexing on the testing split samples. The analyzed images were chosen given the complexity of the shadow scenario, with contents both seen and unseen during training.

6. Challenge Results

The challenge ended with 17 valid submissions for the Track 1, and 18 submissions for Track 2. A number of three solutions (Teams IIM_TTI, AiRiA_Vision and NuNu) optimized their solutions specifically for perceptual quality, participating in the second track only. Section 7 provides details about each of the solutions ranked in the final phase, for both of the tracks.

Table 1 provides the quantitative evaluation of the submitted results for Track 1 (fidelity). Particular rankings along each metric evaluated are provided as subscripts. The solutions achieve a significant performance level, in terms of reconstruction fidelity and quantified perceptual quality, with the provided evaluations being backed by the results provided for visual comparison (see Figure 1).

Table 2 provides the ranking for the second track of the challenge, with the performed user study as the main criteria for evaluation. The first six solutions produced results characterized by similar MOS, correlating with the properties of their outputs, provided for visual evaluation in Figure 2. The top performers of both tracks submitted solutions characterized by a low degree of ghosting artefacts, correctly restored colors and textures, with solutions handling well the most complex shadow scenarios represented at the WSRD+ test split level.

7. Challenge Methods

7.1. LUMOS

Team LUMOS proposes a novel two-stage approach called HirFormer (Dynamic High-Resolution Transformer for Large-Scale Image Shadow Removal) to enhance the

Rank	Team	Username	PSNR↑	SSIM↑	LPIPS↓	Params.(M)	Runtime(s)	Device	Extra data
1	LUMOS	USTC_608	$24.78_{(2)}$	$0.832_{(2)}$	$0.110_{(4)}$	23	6.00	3060	NTIRE'23 [62]
2	Shadow_R	ylxb, ZXCV, tmmdh	$24.58_{(3)}$	$0.832_{(1)}$	$0.098_{(2)}$	376	2.55	RTX2080Ti	No
3	ShadowTech_Innovators	Wangxingbo, USTCX, xlu	$24.81_{(1)}$	$0.832_{(3)}$	$0.111_{(5)}$	26	3.15	3090, A40	NTIRE'23 [62]
4	LVGroup_HFUT	yan.wei	$24.35_{(4)}$	$0.823_{(6)}$	$0.082_{(1)}$	17	3.46	4090	No
5	USTC_ShadowTitan	YuruiZhu, Wangxingbo	$24.04_{(5)}$	$0.827_{(4)}$	$0.104_{(3)}$	83	6 hours	A40	NTIRE'23 [62]
6	GGBond	ZHZhao, Outsiders	$23.87_{(6)}$	$0.824_{(5)}$	$0.127_{(6)}$	8.895	4.50	A6000	No
7	simpleCrew	SimpleCrew	$22.46_{(7)}$	$0.804_{(7)}$	$0.149_{(10)}$	21.8	0.97	A6000	No
8	LSCM-HK	breezewrf	$22.32_{(9)}$	$0.780_{(11)}$	$0.131_{(7)}$	31.82	0.25	A5000	No
9	unicorns	unicorns, OptDev	$22.37_{(8)}$	$0.782_{(10)}$	$0.167_{(14)}$	5.25	0.30	A100	No
10	HKUST-VIP_Lab_01	HKUST-VIP Lab 01	$22.28_{(10)}$	$0.788_{(8)}$	$0.135_{(9)}$	94	2.30	RTX3090	No
11	KLETech-CEVI_ShadowFighters	Niksx	$21.86_{(13)}$	$0.786_{(9)}$	$0.158_{(12)}$	20.2	0.31	RTX3090	No
12	ataza	ataza	$21.17_{(14)}$	$0.778_{(12)}$	$0.151_{(11)}$	0.433	0.05	RTX3070	No
13	PSU-Team	mnaseif	$22.22_{(11)}$	$0.731_{(16)}$	$0.132_{(8)}$	85	36.00	A100	No
14	CVPR_IITRPR	VivekD	$20.14_{(15)}$	$0.765_{(13)}$	$0.160_{(13)}$	0.0036	0.000003	TitanXP	No
15	NJUPT-IOT	weidanyan, WayneFan	$22.19_{(12)}$	$0.732_{(15)}$	$0.263_{(17)}$	20.5	2.1	RTX3090	No
16	TrioTechies	Ak1306, Shri-Hari-Priya, Manasa	$18.47_{(17)}$	$0.751_{(14)}$	$0.230_{(16)}$	5	2.80	RTX3050Ti	No
17	FBU-ISR	Haowei0421	$18.48_{(16)}$	$0.541_{(17)}$	$0.169_{(15)}$	13.4	0.21	RTX3090	No

Table 1. Quantitative evaluations for the submissions corresponding to the *Track 1 (fidelity)* of the NTIRE 2024 Image Shadow Removal Challenge on the *WSRD+* test split. The report uses the naming convention $n_{(m)}$, where n is the value of the metric evaluated and (m) is the rank in the list of submissions sorted along the metric axis.

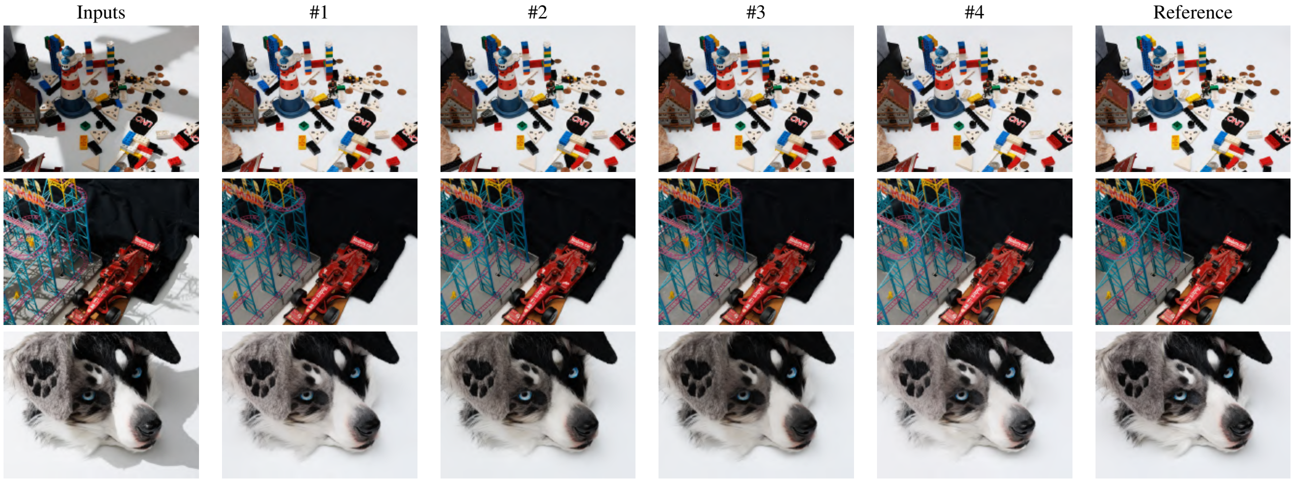


Figure 1. Equivalent samples from the WSRD+ test split, for Team LUMOS, the winner of Track 1 (fidelity) vs. the solutions ranked second, third and fourth. Note the increased intensity shadows, the complex shapes produced by fine structures, and the increased complexity backgrounds.

Rank	Team	Username	PSNR↑	SSIM↑	LPIPS↓	MOS↑	Params.(M)	Runtime(s)	Device	Extra data
1	Shadow_R	ylxb, ZXCV, tmmdh	$24.58_{(3)}$	$0.832_{(1)}$	$0.098_{(4)}$	$7.750_{(1)}$	376	2.55	RTX2080Ti	No
2	LVGroup_HFUT	yan.wei	$24.23_{(4)}$	$0.822_{(5)}$	$0.082_{(1)}$	$7.519_{(2)}$	17	3.46	4090	No
3	USTC_ShadowTitan	YuruiZhu, Wangxingbo	$24.04_{(5)}$	$0.827_{(4)}$	$0.104_{(5)}$	$7.444_{(3)}$	83	6 hours	A40	NTIRE'23 [62]
4	ShadowTech_Innovators	Wangxingbo, USTCX, xlu	$24.81_{(1)}$	$0.832_{(3)}$	$0.111_{(7)}$	$7.438_{(4)}$	26	3.15	3090, A40	NTIRE'23 [62]
5	GGBond	ZHZhao, Outsiders	$23.05_{(6)}$	$0.809_{(6)}$	$0.089_{(2)}$	$7.400_{(5)}$	8.895	4.50	A6000	No
6	PSU-Team	mnaseif	$22.22_{(12)}$	$0.731_{(16)}$	$0.132_{(9)}$	$7.400_{(6)}$	85	36.00	A100	No
7	LUMOS	USTC_608	$24.78_{(2)}$	$0.832_{(2)}$	$0.110_{(6)}$	$7.163_{(7)}$	23	6.00	3060	NTIRE'23 [62]
8	IIM_TTI	placerkyo, Yuki-11	$22.96_{(7)}$	$0.806_{(7)}$	$0.093_{(3)}$	$7.160_{(8)}$	415	16.50	V100/RTX	NTIRE'23 [62], ImageNet
9	AiRiA_Vision	yangbo	$21.90_{(15)}$	$0.689_{(18)}$	$0.238_{(17)}$	$6.825_{(9)}$	5	27.00	A100	No
10	HKUST-VIP_Lab_01	HKUST-VIP Lab 01	$22.28_{(11)}$	$0.788_{(10)}$	$0.135_{(10)}$	$6.619_{(10)}$	94	2.30	RTX3090	No
11	simpleCrew	SimpleCrew	$22.46_{(8)}$	$0.804_{(8)}$	$0.148_{(11)}$	$6.438_{(11)}$	21.8	0.97	A6000	No
12	unicorns	unicorns, OptDev	$22.37_{(9)}$	$0.782_{(11)}$	$0.167_{(15)}$	$6.388_{(12)}$	5.25	0.30	A100	No
13	NJUPT-IOT	weidanyan, WayneFan	$21.80_{(16)}$	$0.727_{(17)}$	$0.267_{(18)}$	$6.069_{(13)}$	20.5	3.33	RTX3090	No
14	NuNu	Nunu	$21.93_{(14)}$	$0.737_{(15)}$	$0.232_{(16)}$	$6.044_{(14)}$	9.3	8.00	4090	No
15	LSCM-HK	breezewrf	$22.32_{(10)}$	$0.779_{(12)}$	$0.131_{(8)}$	$5.950_{(15)}$	31.82	0.25	A5000	No
16	KLETech-CEVI_ShadowFighters	Niksx	$21.98_{(13)}$	$0.794_{(9)}$	$0.157_{(13)}$	$5.681_{(16)}$	20.2	0.31	RTX3090	No
17	ataza	ataza	$21.17_{(17)}$	$0.778_{(13)}$	$0.151_{(12)}$	$4.813_{(17)}$	0.433	0.05	RTX3070	No
18	CVPR_IITRPR	VivekD	$20.14_{(18)}$	$0.765_{(14)}$	$0.160_{(14)}$	$3.750_{(18)}$	0.003	0.000003	TitanXP	No

Table 2. Track 2 (perceptual). Quantitative results of the challenge final submissions on the *WSRD+* test split. Using naming convention $n_{(m)}$, where n is the value of the metric evaluated and (m) is the rank in the list of submissions sorted by the evaluated metric value.

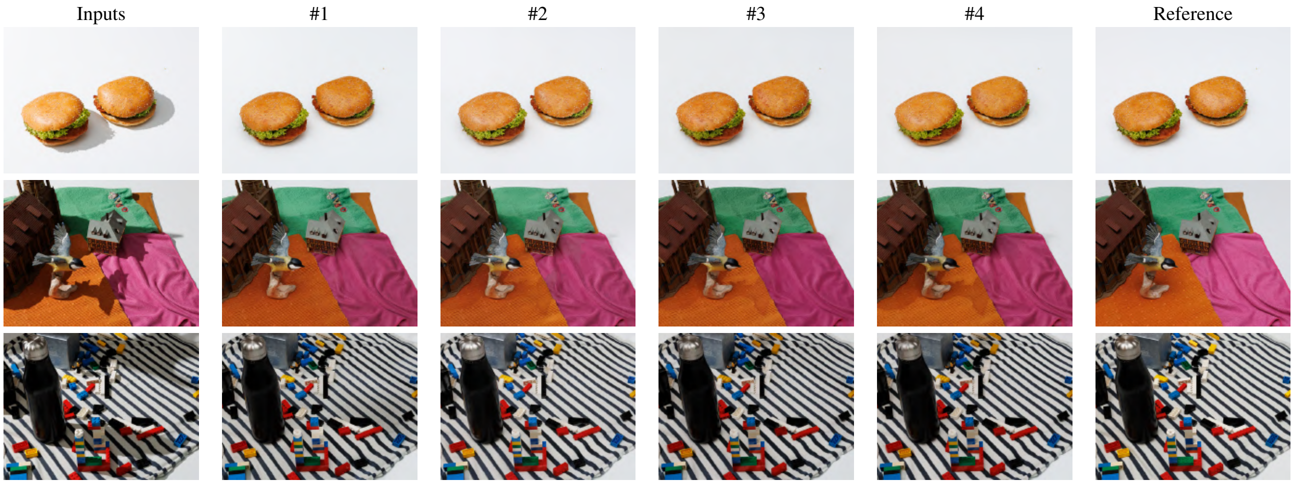


Figure 2. The winner of Track 2 (perceptual), Shadow_R, against other top ranked solutions, on the testing split of the WSRD+ dataset.

restoration quality specifically for high-resolution shadow images. Due to the strong contextual information extraction capability of Transformers in modeling long sequences, they utilize the Vision Transformer framework consistent with [27] to implement image restoration for the unevenly distributed shadow removal task. For high-resolution images in this competition, they propose a dynamic high-resolution approach to assist Transformers in efficiently handling large 2K-sized images. To address the issue of poor performance in handling dark shadows using a single-stage network for shadow removal, they draw inspiration from the Coarse-to-Fine image restoration concept in [81] and introduce the refinement of restored images using the NAFNet network from [6]. This refinement step helps to eliminate bad cases caused by the Transformer block effect. For specific modeling details, please refer to Figure 3.

7.2. Shadow_R

Team Shadow_R proposes a novel mask-free **Shadow** Removal and **Refine**ment (**ShadowRefiner**) model for the shadow removal task. The overall frame is illustrated in Figure 4. The Shadow Removal module is designed to remove shadow from the input via spatial and frequency representation learning, while the Refinement module is proposed to further improve color consistency and enhance texture details for outputs of the Shadow Removal module.

For the Shadow Removal module, this team designs a ConvNext-based U-Net architecture with 7 × 7 depth-wise convolution in each resolution [50]. In addition, the frequency branch network proposed in [80] is adopted. Therefore, the Shadow Removal module is capable of simultaneously leveraging spectral and spatial information for shadow-affected and shadow-free image mappings.

A preliminary experiment of theirs shows that the result of the Shadow Removal module exhibits pixel misalignment with the original clean image in terms of details and local consistency. To this end, the Shadow_R team introduces a transformer-based enhancement and refinement module. Specifically, in the Refinement module, with an input $\mathbf{I} \in \mathbb{R}^{H\times W\times 3}$, the encoder hierarchically reduces the spatial size, while expanding channel capacity, and finally generates latent features $\mathbf{z} \in \mathbb{R}^{\frac{H}{8}\times\frac{W}{8}\times d}$, where d represents the dimension for the latent space. The decoder takes $\mathbf{z}$ as input and progressively recovers the image. On each resolution, the encoder and decoder incorporate several Fast-Fourier Attention based Transformer (FFAT) blocks, where a new attention mechanism (Fast-Fourier Attention, FFA) different from common attention operations in transformers [3, 74] is designed by Team Shadow_R.

7.3. ShadowTech_Innovators

The shadow removal network proposed by ShadowTech_Innovators is primarily comprised of modules such as Gated Linear Units, SimpleGate, and Simplified Channel Attention, inspired by references [6, 32, 81]. Figure 5 is an overview of the framework. The integration of these intricate components results in a robust baseline capable of effectively addressing high-resolution shadow removal tasks.

7.4. LVGroup_HFUT

Shadow removal refers to the process of detecting and eliminating shadows from digital images or videos, while preserving the original texture and color of the area affected by the shadow. However, the shadow removal results usually contain severe boundary artifacts and remaining shadow patterns, which may be due to the lack of strong perceptual constraints. Therefore, this team chose NAFNet [6] with LPIPS perception constraint as the network architecture. The input image will undergo processing through two phases: training and inference. During the training

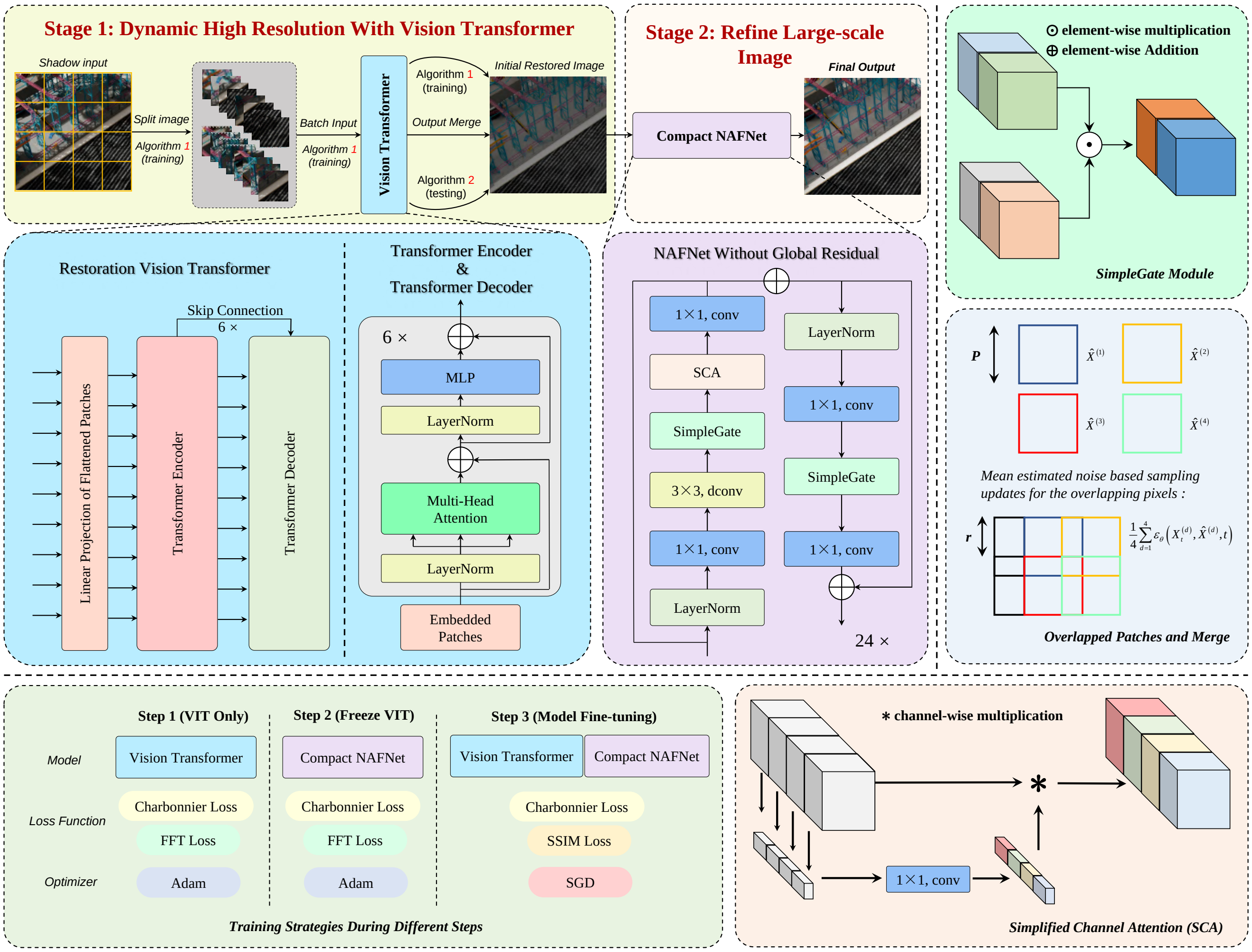


Figure 3. The framework diagram of HirFormer for high resolution image shadow removal, proposed by Team LUMOS. The shadow images undergo the initial restoration process using the Dynamic High Resolution Vision Transformer, yielding preliminary restoration results. Subsequently, the Initial Restored Image is further refined for large-scale images by passing it through a compact NAFNet network, ultimately producing the final clean image.

phase, the input image is fed into the NAFNet network, and it is constrained using three loss functions: LPIPS loss with weight 0.5, FFT loss with weight 0.1, and L1 loss with weight 1. In the inference phase, the process begins by combining a total of n equally spaced checkpoints obtained in the training phase. Then, the input image sequentially passes through these checkpoints, resulting in n outputs. Finally, the maximum ensemble [62] method is applied to process these n outputs, resulting in the ultimate output. There remains the problem of determining the approximate epoch at which the model is well fitted, and based on this, determining the checkpoints at equal intervals. This team uses the user study tool [78] to compare the performance of earlier models, and determine the time point P at which the model is just fit, and based on this, determine the n equally spaced checkpoints between the time point P and the time point when training ends. The entire architecture is shown in Figure 6.

The proposed architecture is based on PyTorch 2.2.1 and an NVIDIA 4090 with 24G memory. This team set 2000 epochs for training with batch size 6, using AdamW with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ for optimization. The initial learning rate was set at 0.001, and cosine annealing was used for the adjustment of the learning rate. For data augmentation, a random crop to 512×512 and then a horizontal flip with probability 0.5 is performed.

For the fidelity track, a total of 10 equally spaced checkpoints from the training phase are used in the maximum ensemble method; their amount is doubled to 20 for the perceptual track.

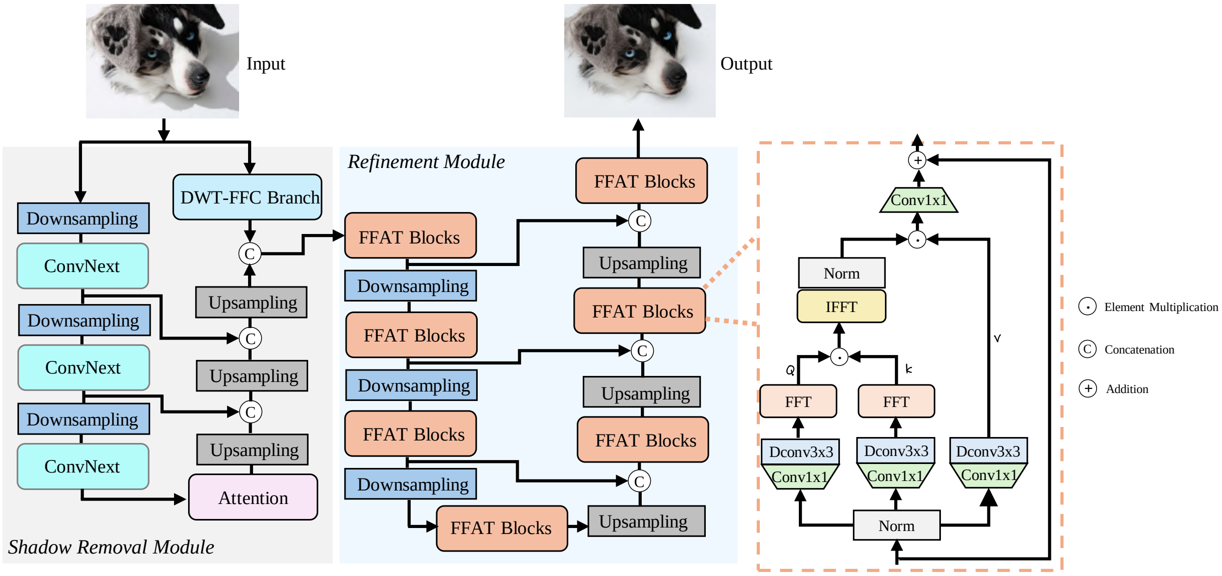


Figure 4. The overall architecture proposed by Team Shadow_R. In the Shadow Removal module, besides the frequency branch proposed in [80], team Shadow_R designs a ConvNext-based U-Net architecture with 7 × 7 depth-wise convolution on each resolution. In the Refinement module, team Shadow_R designs a new attention mechanism (Fast-Fourier Attention, FFA) different from common attention operations in transformers to further enhance texture details.

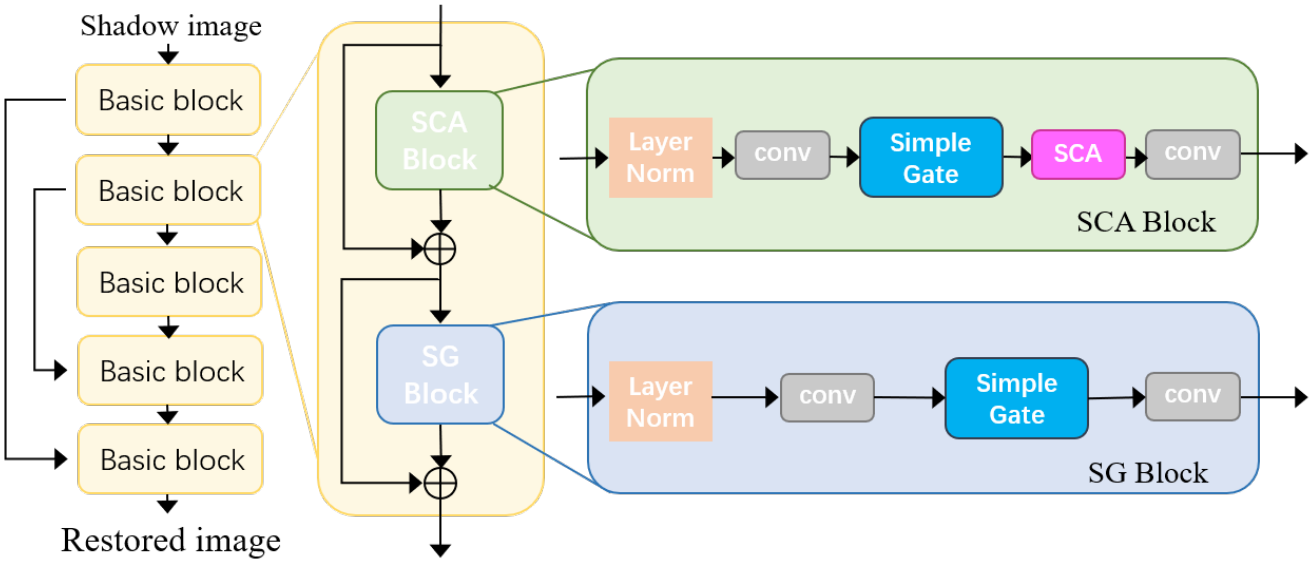


Figure 5. The Global Residual-Free Unet architecture proposed by Team ShadowTech_Innovators.

7.5. USTC_ShadowTitan

The method proposed by USTC_ShadowTitan adopts a two-stage strategy for shadow removal, as shown in Figure 7, focusing on GPU's memory efficiency and large image resolution. Initially, a conventional restoration model (NAFNet) is applied for preliminary shadow removal. Subsequently, leveraging WeatherDiffusion, the result is refined while addressing memory constraints and enhancing image quality. This approach allows them to achieve competitive shadow removal performance while optimizing memory usage and preserving image resolution.

7.6. GGBond

As shown in Figure 8, given a blurred image $I \in \mathbb{R}^{3\times H\times W}$, the method proposed by GGBond first employs a convolution as feature extraction module to extract shallow features $F_0 \in \mathbb{R}^{C\times H\times W}$ from the image, where H and W denote spatial resolution and C denotes the number of channels. The shallow features are then fed into an encoder-

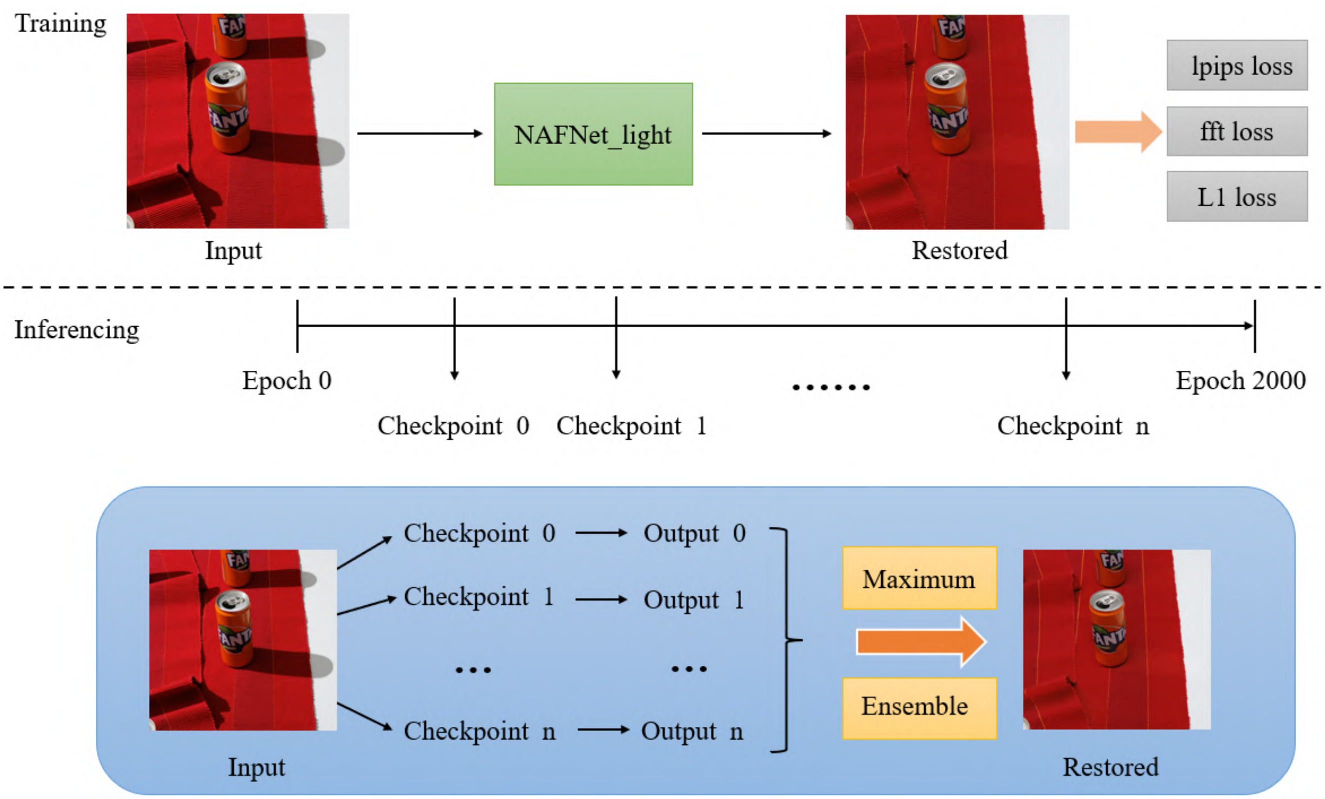


Figure 6. The architecture of the model ensemble strategy used by Team LVGroup_HFUT.

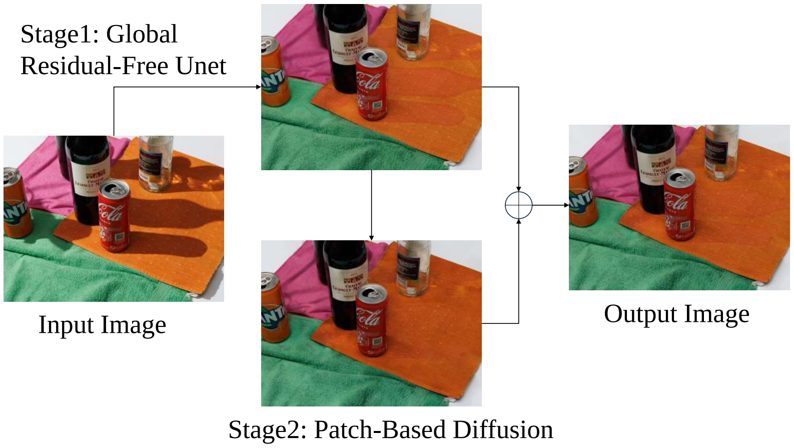
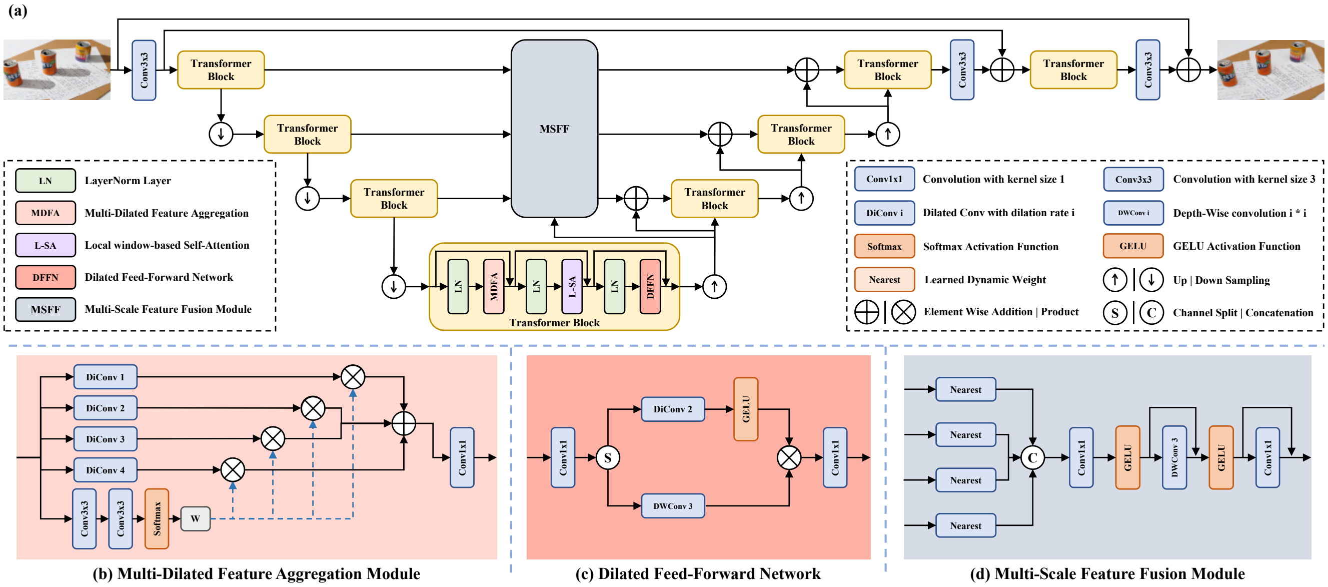


Figure 7. The Overall Framework describing the multi-stage solution proposed by Team USTC_ShadowTitan.

decoder-based network to obtain features $F_d \in \mathbb{R}^{C\times H\times W}$. The encoder-decoder network consists of multiple levels, each comprising several Transformer blocks. Between adjacent levels, the resolution of features is halved while doubling the number of channels, or the resolution is doubled while halving the channels. For feature down-sampling and up-sampling, convolution and transposed convolution operations are used. To fully leverage the features of different scales, GGBond proposes a multi-scale feature fusion module, aiming to provide improved guidance for feature reconstruction during decoder stages. Then, to fuse shallow and deep features, a convolution operation with kernel-size 1×1 is used, aiming to capture structural and content information. After that, several Transformer blocks are employed for further enhancement. Finally, the networks proposed by team GGBond uses a convolution layer to process the en-


(a)
Conv3x3
Transformer Block
MSFF
LN
MDFA
L-SA
DFFN
Transformer Block
LN LayerNorm Layer
MDFA Multi-Dilated Feature Aggregation
L-SA Local window-based Self-Attention
DFFN Dilated Feed-Forward Network
MSFF Multi-Scale Feature Fusion Module
Conv1x1 Convolution with kernel size 1
Conv3x3 Convolution with kernel size 3
DiConv i Dilated Conv with dilation rate i
DWConv i Depth-Wise convolution i * i
Softmax Softmax Activation Function
GELU GELU Activation Function
Nearest Learned Dynamic Weight
Up | Down Sampling
Element Wise Addition | Product
S C Channel Split | Concatenation
DiConv 1
DiConv 2
DiConv 3
DiConv 4
Conv3x3
Conv3x3
Softmax
W
Conv1x1
Conv1x1
S
DiConv 2
GELU
DWConv 3
Conv1x1
Nearest
Nearest
Nearest
Nearest
C
Conv1x1
GELU
DWConv 3
GELU
Conv1x1

(b) Multi-Dilated Feature Aggregation Module

(c) Dilated Feed-Forward Network

(d) Multi-Scale Feature Fusion Module

Figure 8. (a) The overall architecture of our proposed MDFormer. The proposed encoder-decoder network consists of multiple levels, each composed of several Transformer blocks. Moreover, an MSFF module is employed to fully leverage features of multiple scales. Within each Transformer block, an MDFA module is used to extract and aggregate non-local information, and then an L-SA module is employed to compute the correlations between pixels. Finally, a DFFN is applied to further extract and fuse local and non-local features. (b) The proposed MDFA module. (c) The proposed DFFN module. (d) The proposed MSFF module.

hanced features to generate residual image $I_R \in \mathbb{R}^{3\times H\times W}$. By adding the residual image to the input blur image, a restored image $\hat{I} \in \mathbb{R}^{3\times H\times W}$ is obtained.

7.7. IIM_TTI

Deterministic models for shadow removal [62] may produce blurred images. Although recent probabilistic models, such as diffusion models [30, 58], solve this problem, their computational cost is high. In addition to the computational cost, even if the diffusion model is conditioned by any input image for image enhancement and restoration, the diffusion model may not maintain global color consistency, as validated in [11]. Based on these arguments, the method proposed by IIM_TTI first obtains an initial shadow-removal image using the deterministic method [36]. This initial image is then fed into a later step of the diffusion model (Figure 9) to resolve the aforementioned disadvantages of the diffusion model (i.e. high computational cost and poor color consistency). This idea has previously proven to be useful for Burst Super-Resolution [38]. An overview of their method is given in Figure 9.

Deterministic model: Since the deterministic method [36] requires precomputed shadow masks, SASMA [36] is used.

Diffusion model: The diffusion model of IIM_TTI is based on [17] and has been modified as follows.

Since for Refusion [51], the winner of the NTIRE 2023 Image Shadow Removal Challenge, the noise removal network is not based on a vanilla U-Net [56] but on NAFNet [6], the same is used. General diffusion models start from random noise; therefore, their computational cost and training difficulty are high. To address this issue, IIM_TTI uses the ground-truth image (i.e., w/o shadows) and the output image of the deterministic shadow removal model [36] as inputs for a later step of the diffusion model. Each noise removal network in the diffusion model is conditioned by the input image (i.e., w/ shadows) and its shadow mask image estimated by [36] or [9]. Following [67], the encoder of this conditioning module is based on U-Net. Encoded features are fed into all layers of the NAFNet-based noise removal network. This conditioning is done in all steps in the diffusion model, improving the PSNR and LPIPS of the output images. The noise scheduler is not the linear, but the sigmoid scheduler, as its effectiveness is validated in [38]. The best number of steps (i.e., 10-steps) in their diffusion model was found empirically.

7.8. simpleCrew

The approach of simpleCrew centers on a dual-domain strip attention mechanism, which efficiently captures spatial and frequency information essential for effective restoration. The spatial strip attention unit aggregates information from adjacent pixels within the same row or column, guided by learned attention weights derived from a simple convolutional branch. This enables precise local information gathering, which is crucial for shadow removal. Additionally, their frequency strip attention unit [42] addresses

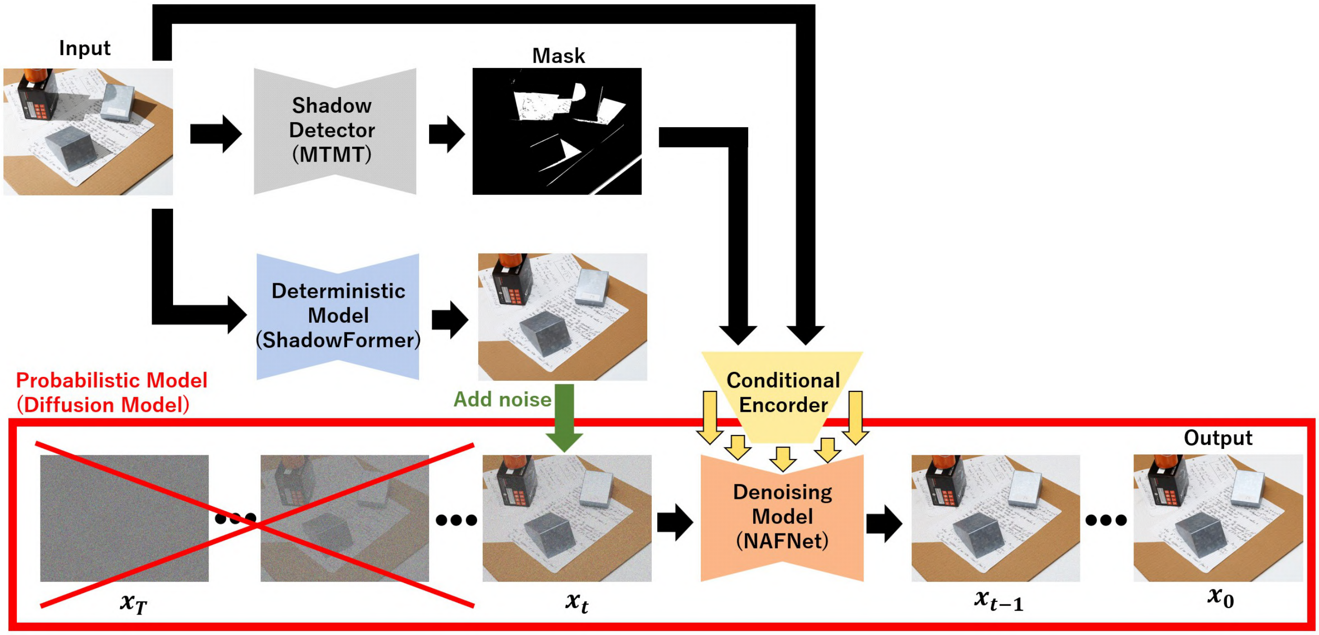


Figure 9. The overall architecture of IIM_TTI model. The proposed model fuses the deterministic model and the probabilistic model. First, shadow affected image input to the deterministic model and the shadow detector. The output image of the deterministic model is added noise and fed into a later step of the diffusion model (the probabilistic model) as x_t. Then the diffusion model follows an inverse process to output an image that is free of shadow and reproduces great detail. The denoising Model in the diffusion model is conditioned by the shadow affected image and the output of the shadow detector.

frequency gaps between sharp and degraded image pairs by separating features into different frequency components and modulating them with lightweight attention weights. This frequency-based refinement ensures accurate restoration, which is particularly beneficial for shadow removal tasks. To handle shadows of varying sizes, Team simpleCrew employs different strip sizes for group-wise feature aggregation. This enables their model to tackle shadows across diverse image contexts effectively. The proposed model architecture integrates ResNet-18 as the feature extractor backbone [28], capitalizing on its powerful representation learning capabilities. Furthermore, they improve performance by introducing cross-attention between the encoder output and feature extractor output, facilitating better integration of features for more accurate shadow removal. The overall framework is illustrated in Figure 10.

7.9. LSCM-HK

The backbone of LSCM-HK's solution is based on NAFNet [6] and Fast Fourier Convolution (FFC) Blocks. Compared to its original implementation, more NAF Blocks are used in the encoding and decoding stages: 2, 2, 4 and 8 blocks in the encoder and 2 blocks in each decoder layer, respectively, while the bottom bridge has 12 NAF Blocks. The FFC [10] is built on the principles of the channel-wise Fast Fourier Transform (FFT), which enables efficient processing of spatial information in images. By incorporating FFT into the network architecture, the FFC allows for a receptive field that covers the entire image, capturing more detail in high-resolution tasks. Following experiments by LSCM-HK, the FFC block is inserted into the middle bottom stage. Lastly, LSCM-HK notes that a RELU activation function is the last operation for generating the restoration image, this is to prevent unexpected noise from being generated. The overall architecture is shown in Figure 11.

7.10. AiRiA_Vision

In response to the challenges posed by variations in lighting and the diverse complexity of shadow depth and color resulting from random background factors, Team AiRiA_Vision proposes a shadow removal method based on diffusion, segmentation, and super-resolution models [44], as shown in Figure 12. This method integrates the powerful generative ability of the diffusion model [55, 79] with the semantic understanding of segmentation methods.

To ensure that the input for the shadow removal model covers the entire image, AiRiA_Vision team first downsamples the image by a factor of 2, followed by shadow removal and super-resolution operations, yielding initial results. Additionally, to retain more details of the image, the

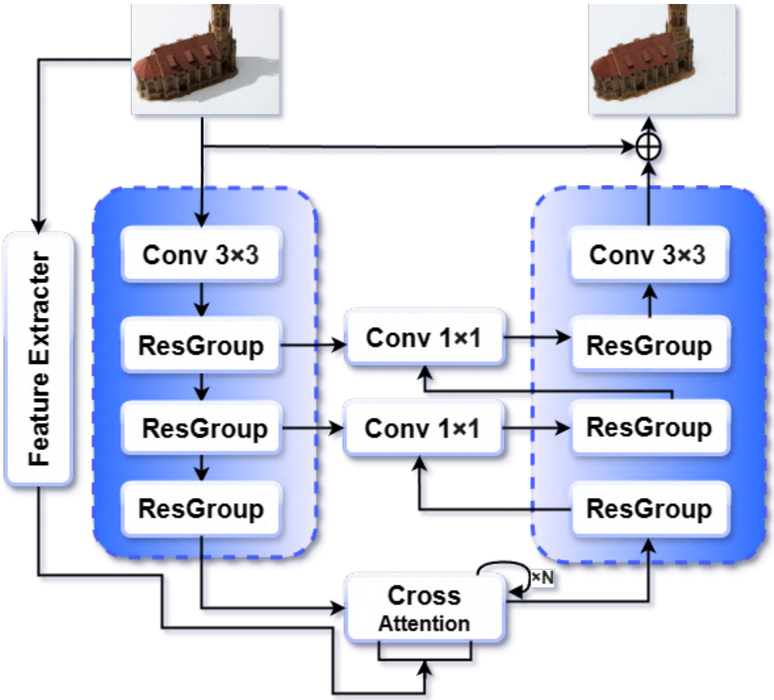


Figure 10. Backbone-Assisted De-shadowing with Advanced Attention Mechanism

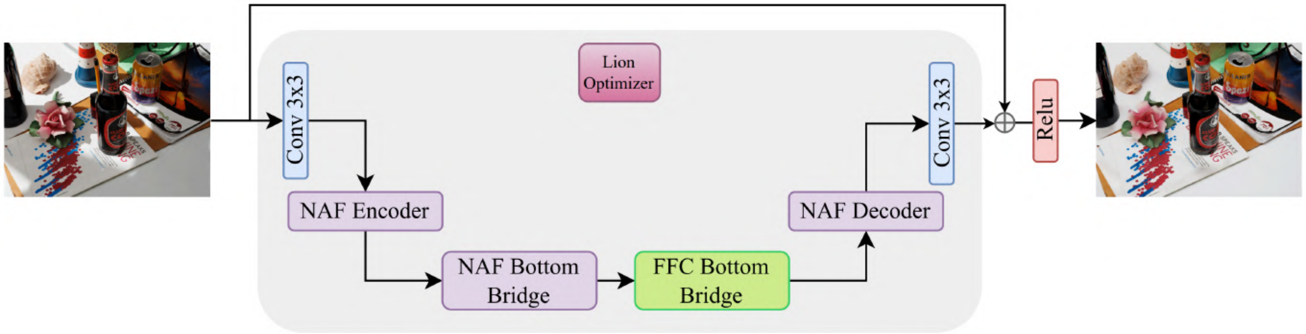
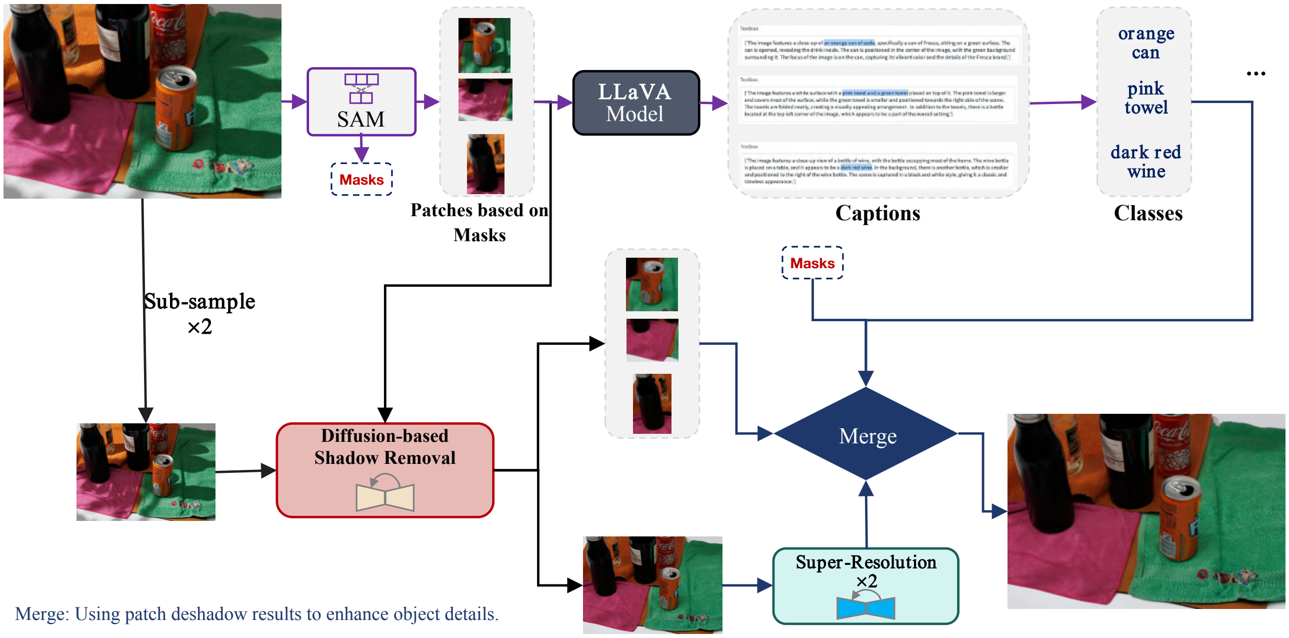


Figure 11. The overall architecture proposed by LSCM-HK. The proposed model mainly consists of NAF Encoder, NAF Bottom Bridge, NAF Decoder and FFC Bottom Bridge. Skip connections between Encoder and Decoder are omitted in the visualization for simplicity. For the FFC Bottom Bridge, specifically, the ratio of global input feature and global output feature is set to 0.75.

proposed method uses the segmentation-map of the SAM model [35] to partition the image into 1024x1024 blocks. After shadow removal, these blocks are merged with the initial results based on masks. Furthermore, this method uses the LLaVa multimodal large language model [47] to further optimize shadow removal results based on analysis of the validation set. Specifically, for a given input image, the proposed method consists primarily of two branches. First, the shadow removal branch: the input image is initially downsampled by bilinear interpolation to achieve 2x downsampling. It is then fed into the shadow removal model, which is trained to eliminate shadows while preserving details in non-shadow areas. The output results are further enhanced using super-resolution techniques, such as HAT-SR [7], to restore the image resolution to that of the original input image.

Second, the semantic segmentation branch: the input image undergoes semantic segmentation using the SAM model [35], generating masks for all regions in the image. These mask image blocks corresponding to all masks are then fed into the shadow removal model to obtain shadow-removed image blocks. Compared to shadow removal after downsampling, these results retain more detail. The two results are then merged using masks to improve the PSNR



['The image features a close-up of an orange can of soda, specifically a can of Fresca, sitting on a green surface. The can is opened, revealing the drink inside. The can is positioned in the corner of the image, with the green background surrounding it. The focus of the image is on the can, capturing its vibrant color and the details of the Fresca brand.']

['The image features a white surface with a pink towel and a green towel placed on top of it. The pink towel is larger and covers most of the surface, while the green towel is smaller and positioned towards the right side of the scene. The towels are folded neatly, creating a visually appealing arrangement. In addition to the towels, there is a bottle located at the top left corner of the image, which appears to be a part of the overall setting.']

['The image features a close-up view of a bottle of wine, with the bottle occupying most of the frame. The wine bottle is placed on a table, and it appears to be a dark red wine. In the background, there is another bottle, which is smaller and positioned to the right of the wine bottle. The scene is captured in a black and white style, giving it a classic and timeless appearance.']

Figure 12. Framework of Team AiRiA_Vision

and perceptual metrics of the shadow removal output. Additionally, the mask-corresponding image blocks are also processed using the LLaVA model [47] to obtain semantic descriptions for each patch. After determining the class of objects corresponding to the mask area, these results are fused with the previous results to further enhance the method's performance.

7.11. unicorns

Team unicorns developed a transformer-based deep network to learn shadow-free image mapping from a shadow-corrupted image, as shown in Figure 13. Their architecture comprises two separate encoder-decoder blocks (EDB) and a multihead correlation block (MHCB) to produce plausible images. Team unicorns leverage illumination mapping from Retinex theory [39] to accelerate the reconstruction performance. The team examined the practicability of illumination mapping in generic image restoration techniques like shadow removal, incorporating Retinexformer [3] into the first half of their architecture. In shadow removal, Retinexformer can outperform well-known image restoration methods like Uformer, Restormer, etc. However, like other restoration models, it failed to address shadows in complex regions. To address this limitation, they proposed to utilize an MHCB, followed by another EDB. Using the correlated features with intermediate output in the second EDB they improve their restoration results. In addition to that, a perceptual loss, including luminance-chrominance guidance, is used to address color inconsistencies.

7.12. HKUST-VIP_Lab_01

Team HKUST-VIP_Lab_01 proposes a **D**egradation-**r**esidual **Diff**usion with **A**daptive **P**rompt for Real-world Shadow Removal (**APDrDiff**), visualized in Figure 14.

They shift the emphasis of restoration to the shadow regions, leveraging the generative capabilities of a diffusion model to learn the degradation residuals specific to the shadow regions. To improve the performance of the diffusion model, they utilize a routing mechanism that allows the network to adaptively select the optimal prompt to condition it.

7.13. KLETech-CEVI_ShadowFighters

The proposed MFNN framework includes three main modules: the Hierarchical Spatio-Contextual (HSC) feature encoder, Global-Local Spatio-Contextual (GLSC) block, and the Hierarchical Spatio-Contextual (HSC) decoder, as shown in Figure 15. Typically, image-shadow removal networks employ feature scaling to vary the sizes of the receptive fields. The varying receptive fields facilitate the learning of local-to-global variances in the features. Therefore, the solution proposed by KLETech-CEVI_ShadowFighters learns contextual information from multiscale features while preserving high-resolution spatial details via a hierarchical-style encoder-decoder network with residual blocks as the backbone [14]. The proposed MFNN optimize the learning of MFNN with the proposed

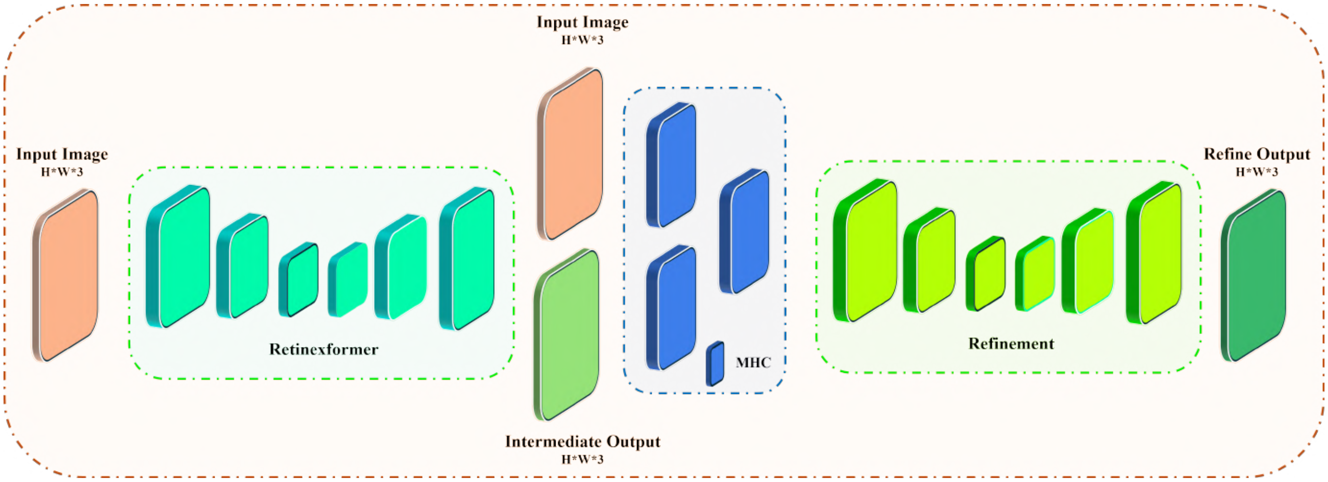


Figure 13. The overview of proposed DFormer for shadow removal.

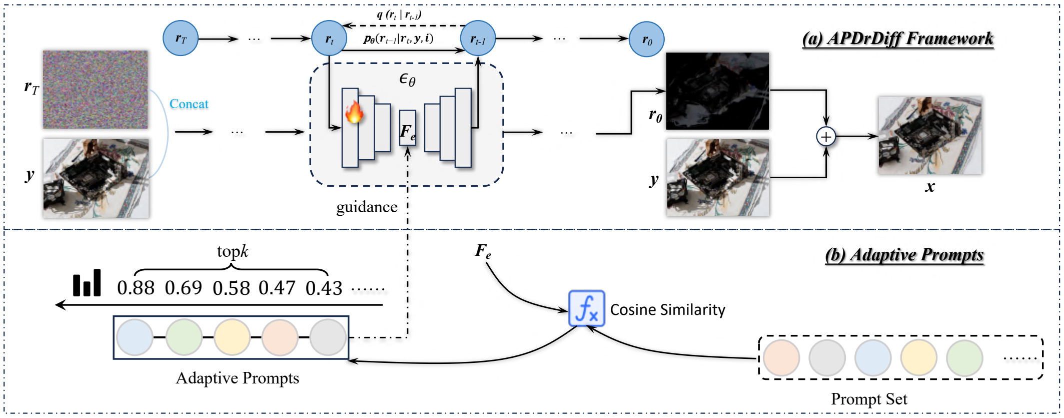


Figure 14. The overview of Team HKUST-VIP_Lab_01 proposed method for Real-world shadow removal.

$\mathcal{L}_{MFNN}$ and is given as,

$$\mathcal{L}_{MFNN} = (\alpha * L_1) + (\beta * \mathcal{L}_{VGG}) + (\gamma * \mathcal{L}_{MSSSIM}) \quad (1)$$

where, α, β, and γ are the weights. We experimentally set the weights to $\alpha = 0.5$, $\beta = 0.7$, and $\gamma = 0.5$. $\mathcal{L}_{MFNN}$ is a weighted combination of three distinct losses inspired from [13, 15, 16, 29].

7.14. ataza

Team ataza proposes a low-parameter deep learning model for removing shadows from images. Instead of processing images in the spatial domain, they utilize Discrete Wavelet Transform (DWT) to shift the image representation into the frequency domain. Specifically, they leverage the HAAR wavelet for this purpose. The model strikes a balance between performance and efficiency, delivering competitive results while consisting of only 433,494 parameters. Additionally, it can process a 1024x1024 image in just 0.14s and a 512x512 image in 0.04s.

Motivated by the observation that shadows predominantly affect the low-frequency components rather than the high-frequency components in an image, the team ataza employs Discrete Wavelet Transform (DWT) to partition the image into low- and high-frequency features in order to process them independently. This approach allows them to focus on recovering low-frequency information while preserving high-frequency information that might get lost in the deeper layers. Figure 16 shows the overall model architecture. Therefore, their model comprises two branches: a high-frequency branch and a low-frequency branch. Initially, a convolutional operation is applied to the input image to extract features and enhance its channels. The Discrete Wavelet Transform (DWT) block subsequently divides the image into high-frequency and low-frequency features, directing them to their respective branches. Within the low-frequency branch, three Low-Frequency blocks (LF block) process and extract low-frequency features. Following each LF Block, a DWT block once again splits the output, di-

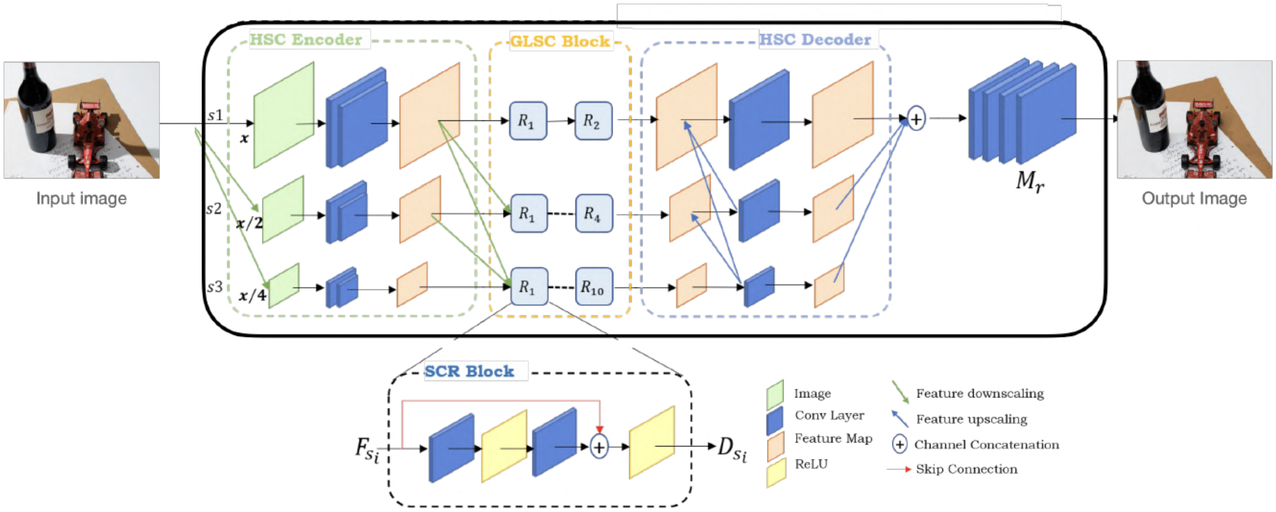


Figure 15. Overview of the proposed Multi-scale Feature FusionNet (MFNN). The encoder extracts features on three distinct scales, with information passed across hierarchies (green dashed box). Fine-grained global-local spatial and contextual information is learned through the GLSC block (orange dashed box). In the decoder, information exchange occurs in reverse hierarchies (blue dashed box).

recting the high-frequency features to the high-frequency branch. There, the previous high-frequency features are added to those received from the low-frequency branch, before being processed by the High-Frequency Block (HF Block). In total, four HF Blocks are employed. The outputs of the three LF blocks are concatenated, and the outputs of the last three high-frequency (HF) blocks are concatenated as well. Passing the features through the Discrete Wavelet Transform (DWT) block halves their spatial dimensions. Although the LF block upscales the features back to their original size, the high-frequency features in the HF branch remain half-sized. To restore them to their original size, bilinear interpolation is applied. Both sets of features undergo point convolution and pixel attention before being fused through addition. Subsequently, pixel attention, along with convolution, is employed to generate the final image.

The loss function used to train the model is a combination of L1 and a loss based on the frequency domain defined in Equation 2. There a Fast Fourier Transform is applied to both the target and generated image, and L1 loss is calculated between them. The combined loss function is given by Equation 3, with I being the reference image and $\hat{I}$ the corresponding restored image. The $FFT(.)$ function represents the power spectrun of the frequency representation of the input image.

$$FFT_{loss} = L1_{loss}(|FFT(\hat{I})|, |FFT(I)|) \tag{2}$$

$$Total_{loss} = 20 \times L1_{loss} + FFT_{loss} \tag{3}$$

7.15. PSU-Team

PSU team introduced *RRRN* Robust Refusion and Restormer network for Shadow Elimination with Self-supervised labeling, a two-stage model designed to address the challenging task of shadow removal. By integrating the strengths of diffusion, via Refusion [43], and transformer, via Restormer [74], architectures as shown in Figure 17, the team aims to balance the preservation of image details with the reduction of artifacts typically associated with the shadow removal process.

Two-Stage Process: Initially Refusion removes shadows, but also introduces artifacts. Restormer is then used to refining the image, removing the artifacts introduced by Refusion and therefore increasing image fidelity (e.g PSNR and SSIM).

Self-supervised Data Augmentation Technique: To enhance the model's training, PSU team implement self-supervised labeling techniques, utilizing both internal and external data augmentation methods: Internally, PSU team generate pseudo-shadow masks by subtracting shadow-free images from their shadow-affected counterparts, enriching the dataset with diverse training samples. This process is represented as:

$$M = |I_s - I_{sf}| \tag{4}$$

where M is the shadow mask, I_s is the shadow-affected image, and I_{sf} is the shadow-free image. Furthermore, the PSU team simulate shadows by applying shadow-effect stickers to shadow-free images, further expanding the training dataset's diversity:

$$I_{sa} = I_{sf} + S \tag{5}$$

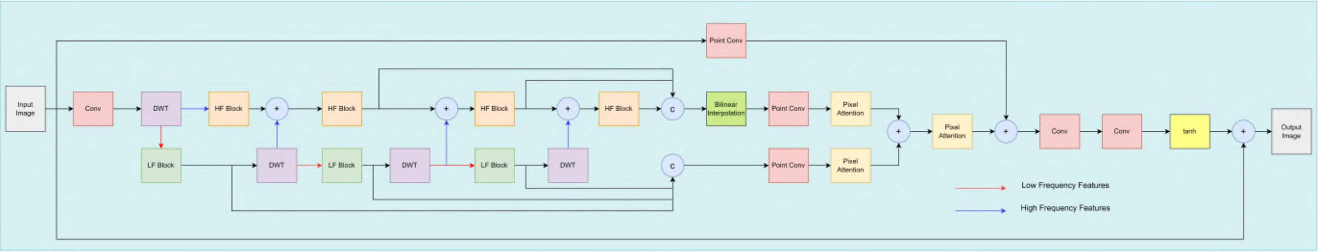


Figure 16. The overall model architecture of the solution proposed by Team ataza.

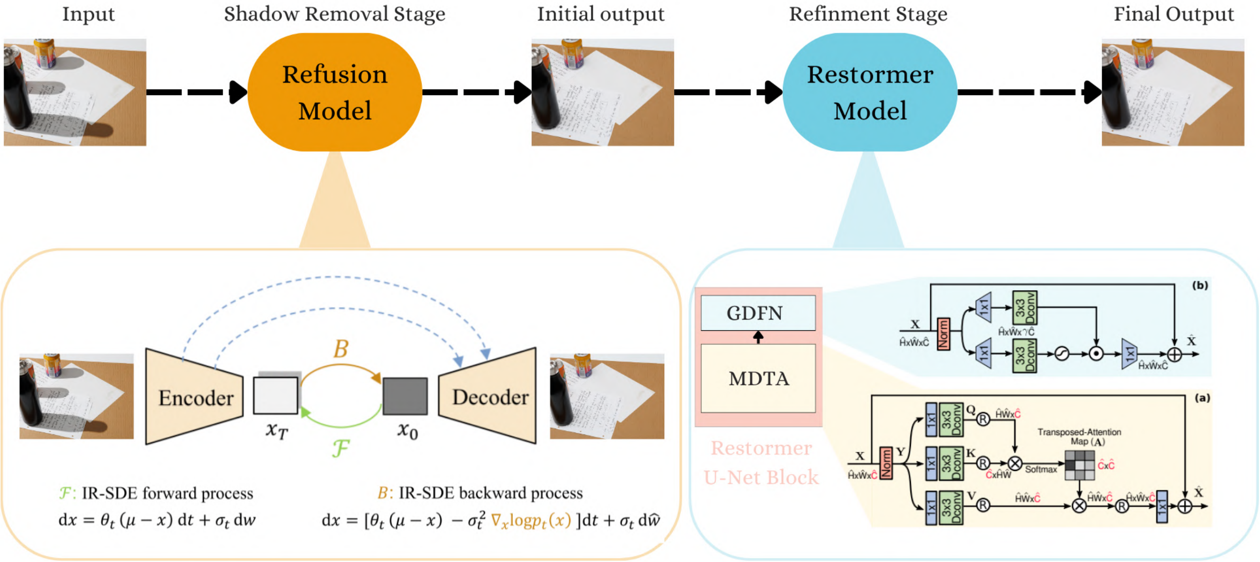

Input
Shadow Removal Stage
Initial output
Refinment Stage
Final Output
Refusion Model
Restormer Model
Encoder
Decoder
B
x_T
x_0
$\mathcal{F}$
$\mathcal{F}$: IR-SDE forward process
$dx = \theta_t (\mu - x)\, dt + \sigma_t\, dw$
B: IR-SDE backward process
$dx = [\theta_t (\mu - x) - \sigma_t^2 \nabla_x \log p_t(x)]dt + \sigma_t\, d\hat{w}$
GDFN
MDTA
Restormer U-Net Block
(b)
(a)
Norm
1x1
3x3 Dconv
Transposed-Attention Map (A)
Softmax
Q
K
V
X
Y
$\hat{X}$

Figure 17. The overall architecture proposed by PSU Team. In the Shadow Removal module, the two-stage model combines the perceptual strengths of Refusion with the structural precision of Restormer.

In this equation, I_{sa} represents the shadow-augmented image, I_{sf} is the original shadow-free image, and S is the shadow sticker effect.

Adoption of Focal Frequency Loss: Focal Frequency Loss [5] is employed to guide the model towards focusing on challenging frequency components, ensuring that the shadow-removed images preserve high fidelity. This objective is particularly useful for refining the output of the Restormer stage, and is defined as:

$$\mathcal{L}_{FFL} = \sum_{i=1}^{N} \left(1 - \frac{|\hat{F}_i - F_i|}{\max(|\hat{F}_i|, |F_i|)} \right)^{\gamma}, \tag{6}$$

where $\hat{F}_i$ and F_i are the predicted and true frequency components of the image, respectively, and γ is a focusing parameter.

7.16. CVPR_IITRPR

Team CVPR_IITRPR has proposed a computationally efficient and lightweight network for image shadow removal with a very small number of parameters (0.003M). The schematic of the proposed network is shown in Figure 18. It consists of the Global-Local Feature Extraction (GLFE) block and a Simple Feed Forward Network (FFN). The GLFE Block extracts relevant information using two NAFNet [6] blocks: a vanilla baseline block to extract global information and a Half residual Convolutional branch [59] block for local feature retrieval. This extracted information is further traversed through the feed forward network to assist with the image shadow removal task.

7.17. NJUPT-IOT

Team NJUPT-IOT has fine-tuned a recently open sourced baseline model called *MambaIR* [24]. In recent years, the Selective-State Space Model (SSM) and its improved variants have shown great potential in solving the long-term correlation dependencies problem of sequential data, and have gradually been used in the field of image processing, which also provides a way to solve the problem of shadow removal tasks. *MambaIR* improves vanilla Mamba by introducing local enhancement and channel attention.

And this work has verified the ability of the advanced Selective State-Space Model to enhance the long-term cor-

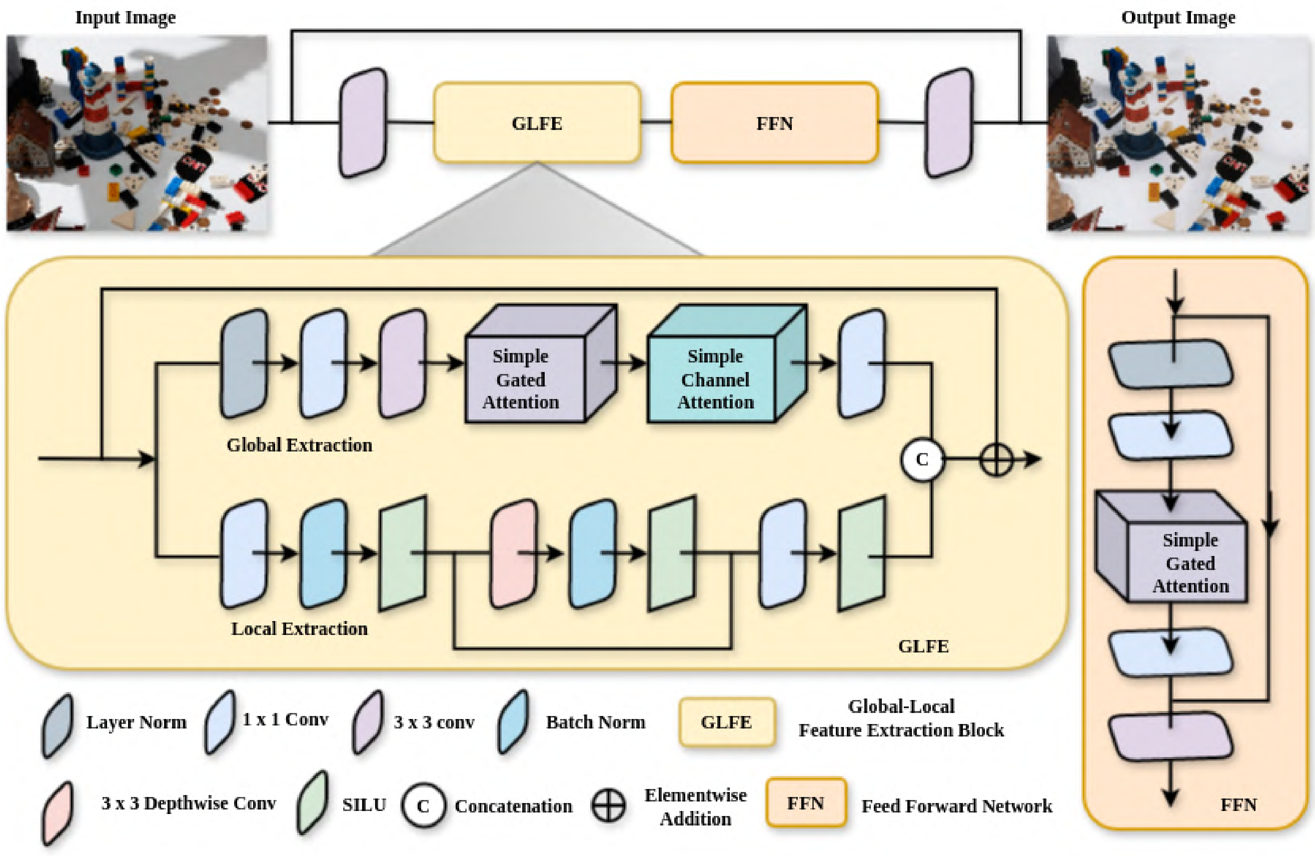


Figure 18. Architectural details of the proposed approach for shadow removal by team CVPR_IITRPR.

relation extraction of feature sequences, as well as the effectiveness of processing image shadow removal tasks through training and inference experiments.

7.18. NuNu

The approach of team NuNu has three main parts, as depicted in Figure 19.

(1) **Shadow mask acquisition.** Obtain the shadow mask by calculating the difference between the real shadow image and the shadow-free image of the training set and use this as training data to train the BDRAR.

(2) **Train pipline.** In order to speed up the training process, the image is downsampled three times, then BDRAR predicts the shadow area in the image. This tuple of images is then used to train Shadowformer [25] on the WSRD+ dataset.

(3) **Infer pipeline.** For the infer phase, BDRAR is used to extract the shadow mask. Two additional versions of the image and mask are created by flipping horizontally and vertically.

ShadowFormer then removes the shadows from all three images, the ultimate result is obtained by obtaining the maximum pixel value of the corresponding pixel position. Finally, SwinIR is used to restore the image to its original resolution.

7.19. TrioTechies

The ultimate aim of the method proposed by TrioTechies is to handle high-resolution shadow removal directly via a frequency-aware network. For this, they propose Frequency-Aware Shadow Erasing Net (FSENet), a transformer-based model that works in the frequency domain. The solution proposed by Team TrioTechies is illustrated in Figure 20. The input image is divided into different low- and high-frequency components using the Laplacian Pyramid (LP) for separate processing. The two main modules present in the architecture are:

Low-Frequency Deshading Module: This module uses several transformer-based blocks to reduce the color and illumination distortions of the shadow image. It contains components such as the Dimension-Aware Transformer (DAT), Tri-layer Attention Alignment (TAA) block, and Deeper Feature Extraction (DFE) block.

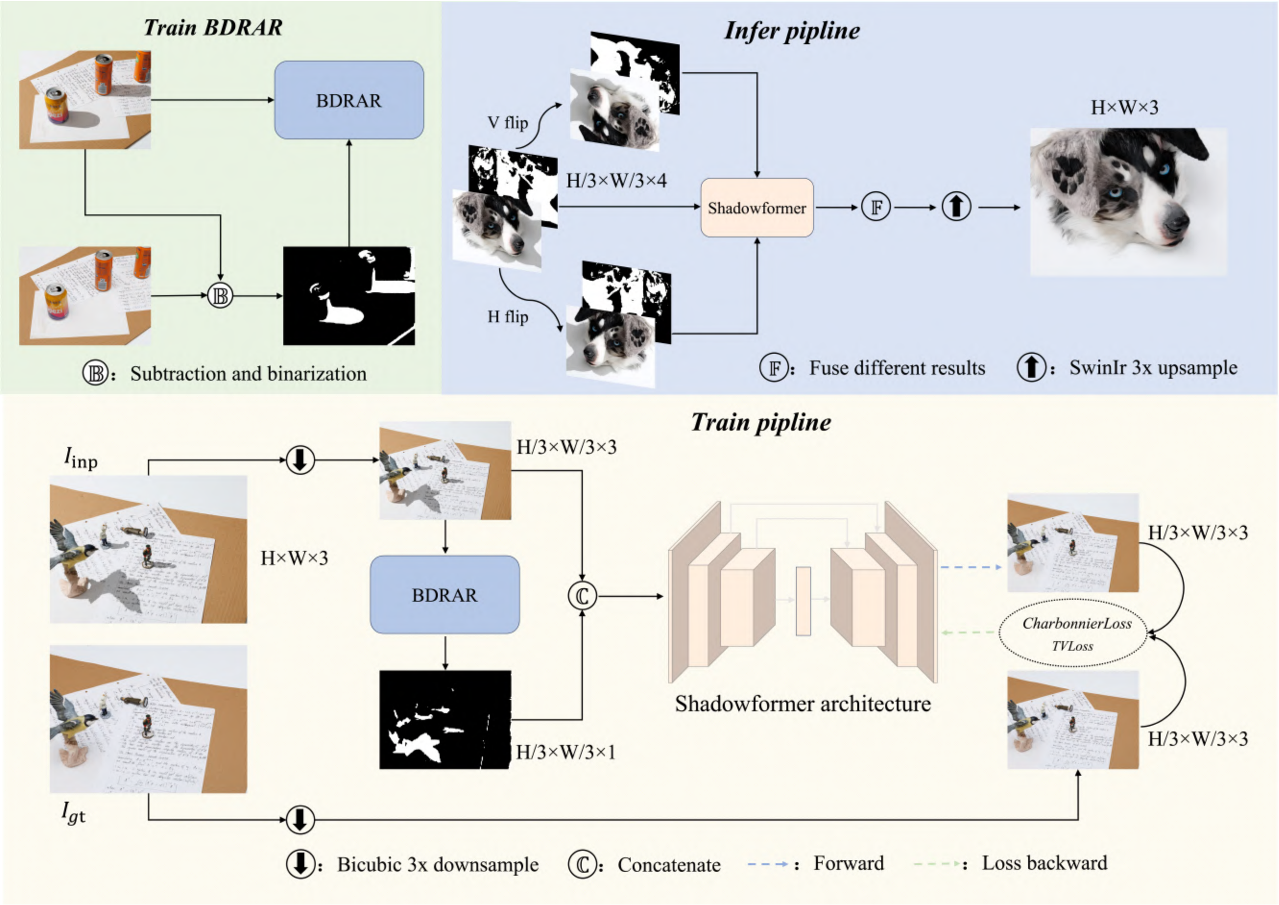


Figure 19. Overview of team NuNu's approach.

High-Frequency Restoration Module: This module involves several dilated convolution-based texture recovery modules to restore the refined details of the learned features. It contains components such as the Texture Recovery Module (TRM) and Spatial Pooling Pyramid (SPP).

DAT block: Each DAT block is built via sequential height attention and width attention to learn the feature individually in different aspects, and a convolution-based layer for local feature processing. For an image $I \in R^{H\times W\times 3}$, a low frequency component $L_3 \in R^{\frac{H}{4}\times\frac{W}{4}\times 3}$ is fed to the DAT block that outputs features F_1, F_2, F_3 belonging to the space $R^{\frac{H}{4}\times\frac{W}{4}\times C}$.

TAA block: The output features of the DAT block are fed to the TAA block upon concatenation for feature mixing. Concatenated features $F_in \in R^{N\times\frac{H}{4}\times\frac{W}{4}\times C}$ (where N=3 in this case) are reshaped into $\hat{F_in} \in R^{\frac{H}{4}\times\frac{W}{4}\times 3C}$ which is subjected to a set of convolutions to obtain the output feature F4.

DFE block: After obtaining the feature F4 , DFE blocks are used to learn multi-scale features. A DFE block begins with a LayerNorm, followed by two convolutions and a deformable convolution.

TRM block: After the pixel-wise multiplication of the contours learned and the features in high frequency restoration, the TRM block contains a series of dilated convolutions which have a complexity of *O(d)* where d denotes one dot product, and attentive aggregation nodes which are computationally lightweight.

SPP block: The SPP block follwing the TRM block facilitates the remixing of multi-context features. It has a complexity of $O(r \cdot s^2)$.

7.20. FBU-ISR

The solution developed by the FBU-ISR team is illustrated in Figure 21. To address pixel misalignment in the training data, the team implemented an initial pixel offset correction technique. NAFNet [6] was implemented with a depth of 5, to improve dense encoding within the model, two blocks were stacked after each up/down-conv.

Loss Function. In alignment with the methodologies delineated in prior research [21,22], FBU-ISR team devises and implements a hybrid loss function as detailed below:

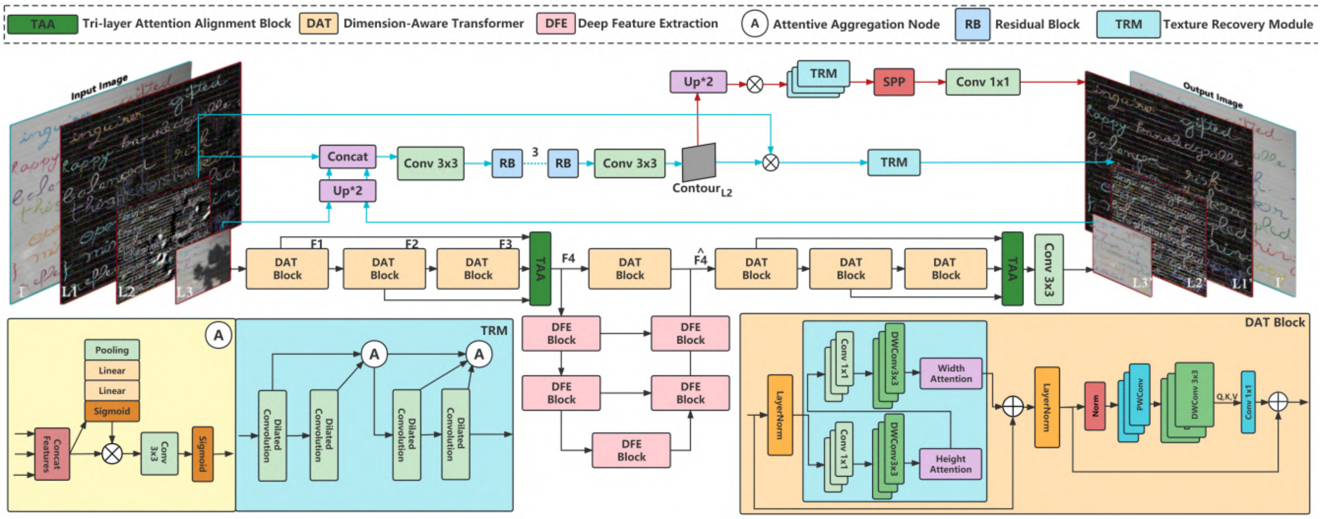


Figure 20. A graphical representation of the solution proposed by Team TrioTechies.

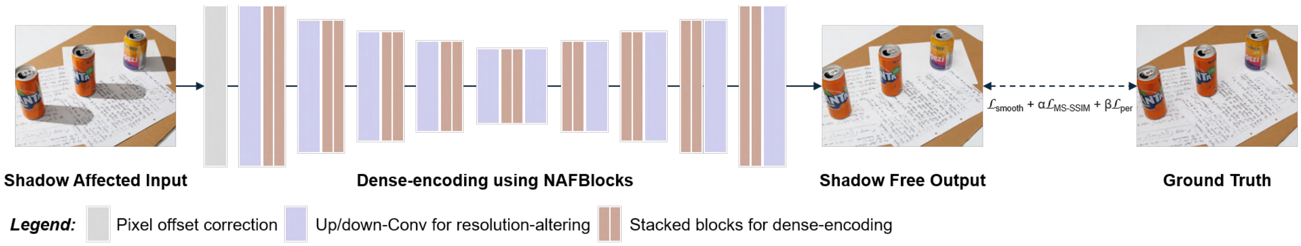

$\mathcal{L}_{\text{smooth}} + \alpha\mathcal{L}_{\text{MS-SSIM}} + \beta\mathcal{L}_{\text{per}}$
Shadow Affected Input
Dense-encoding using NAFBlocks
Shadow Free Output
Ground Truth
Legend:
Pixel offset correction
Up/down-Conv for resolution-altering
Stacked blocks for dense-encoding

Figure 21. Details of the FBU-ISR team's architectural approach to shadow removal.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{smooth}} + \alpha\mathcal{L}_{\text{MS-SSIM}} + \beta\mathcal{L}_{\text{per}} \tag{7}$$

where $\alpha = 0.1$ and $\beta = 0.01$ are the hyperparameters that weight each loss component.

8. Conclusion

The NTIRE 2024 Image Shadow Removal challenge builds on the success of the previous edition, enjoying significant attention from the computer vision community. The challenge was split between the competitors optimizing for fidelity and perceptual quality optimized solutions, with a separate track designed for solutions excelling in image generation. Both tracks received a high number of submissions, with the analyzed solutions achieving a significant quantified performance level. This shows in the quantified PSNR, SSIM and LPIPS characterizing the restored images, which were the criteria for ranking the first track of the challenge. An user study was conducted, as the primary way of ranking the solutions submitted for the second track. Many challenge participants provided insightful feedback, with future ideas for the following editions of the challenge.

Acknowledgements

This work was partially supported by the Humboldt Foundation. We thank the NTIRE 2024 sponsors: Meta Reality Labs, OPPO, KuaiShou, Huawei and University of Würzburg (Computer Vision Lab).

A. Teams and Affiliations

NTIRE 2023 Team

Title:
NTIRE 2024 Challenge on Shadow Removal
Members:
Florin-Alexandru Vasluianu[1], Tim Seizinger[1], Zhuyun Zhou[1], Zongwei Wu[1], Cailian Chen[2], Radu Timofte[1]
Affiliations:
[1] Computer Vision Lab, IFI & CAIDAS, University of Würzburg
[2] Shanghai Jiaotong University, China

Shadow_R Team

Title:
ShadowRefiner: Towards Mask-free Shadow Removal via Fast Fourier Transformer
Members:
Wei Dong[1], Han Zhou[1], Yuqiong Tian[1], Jun Chen[1]
Affiliations:
[1] Department of Electrical and Computer Engineering, McMaster University, Canada

LUMOS

Title:
HirFormer: Dynamic High Resolution Transformer for Large-Scale Image Shadow Removal
Members:
Xin Lu[1], Yurui Zhu[1], Xi Wang[1], Dong Li[1], Jie Xiao[1], Yunpeng Zhang[1], Xueyang Fu[1], Zheng-Jun Zha[1]
Affiliations:
[1] University of Science and Technology of China, Hefei, China

LVGroup_HFUT

Title:
Exploring the Application of NAFNet in Shadow Removal
Members:
Zhao Zhang[1], Suiyi Zhao[1], Bo Wang[1], Yan Luo[1], Yanyan Wei[1]
Affiliations:
[1] Hefei University of Technology, China

ShadowTech Innovators

Title:
Shadow Removal via Global Residual-free Unet and Shadow generation
Members:
Dong Li[1], Yurui Zhu[1], Xi Wang[1], Jie Xiao[1], Xin Lu[1], Yunpeng Zhang[1], Xueyang Fu[1], Zheng-Jun Zha[1]

Affiliations:
[1] University of Science and Technology of China, Hefei, China

USTC_ShadowTitan

Title:
Shaffusion: Diffusion Based Two-Stage Refinement for High Resolution Image Shadow Removal
Members:
Yunpeng Zhang[1], Xi Wang[1], Jie Xiao[1], Dong Li[1], Yurui Zhu[1], Xin Lu[1], Xueyang Fu[1], Zheng-Jun Zha[1]
Affiliations:
[1] University of Science and Technology of China, Hefei, China

GGBond

Title:
Lightweight Multi-Dilated Transformer for Shadow Removal
Members:
Zhihao Zhao[1], Long Sun[1], Tingting Yang[1], Jinshan Pan[1], Jiangxin Dong[1], Jinhui Tang[1]
Affiliations:
[1] Nanjing University of Science and Technology, Nanjing, Jiangsu Province, China

PSU-Team

Title:
RRRN: Robust Refusion and Restormer network for Shadow Elimination with Self-supervision labelling
Members:
Bilel Benjdira[1], Mohammed Nassif[12], Anis Koubaa[1], Ahmed Elhayek[2], Anas M. Ali[1]
Affiliations:
[1] Robotics and Internet-of-Things Laboratory, Prince Sultan University, Riyadh 12435, Saudi Arabia
[2] Artificial Intelligence Department, Prince Muqrin University, Medinah 41311, Saudi Arabia

IIM_TTI

Title:
A Hybrid Approach to Shadow Removal: Fusing Deterministic and Probabilistic Models
Members:
Kyotaro Tokoro[1], Kento Kawai[1], Kaname Yokoyama[1], Takuya Seno[1], Yuki Kondo[1], Norimichi Ukita[1]
Affiliations:
[1] Intelligent Information Media laboratory, Toyota Technological Institute, Japan

AiRiA_Vision

Title:
Shadow Removal based on Diffusion, Segmentation and Super-resolution Models
Members:
Chenghua Li[1], Bo Yang[1], Zhiqi Wu[1], Gao Chen[1], Yihan Yu[2]
Affiliations:
[1] Nanjing Artificial Intelligence Research of IA (AiRiA)

[2] High School Affiliated to Nanjing Normal University Jiangning Campus

HKUST-VIP_Lab_01

Title:
Degradation-residual Diffusion with Adaptive Prompts for Real-world Shadow Removal
Members:
Sixiang Chen[1], Kai Zhang[1], Tian Ye[1], Wenbin Zou[2], Yunlong Lin[3], Zhaohu Xing[1], Jinbin Bai[4], Wenhao Chai[5], Lei Zhu[1]
Affiliations:
[1] Hong Kong University of Science and Technology (Guangzhou)
[2] South China University of Technology
[3] Xiamen University
[4] National University of Singapore
[5] University of Washington

simpleCrew

Title:
Backbone-Assisted De-shadowing with Advanced Attention Mechanism (BADAAM) Net
Members:
Ritik Maheshwari[1], Rakshank Verma[1], Rahul Tekchandani[1], Praful Hambarde[2], Satya Narayan Tazi[1], Santosh Kumar Vipparthi[2], Subrahmanyam Murala[3]
Affiliations:
[1] GEC Ajmer
[2] CVPR Lab IIT Ropar
[3] SCSS Trinity College Dublin

unicorns

Title:
DoubleFormer (DFormer)
Members:
Jaeho Lee[1], Seongwan Kim[1], Sharif S M A[1], Nodirkhuja Khujaev[1], Roman Tsoy[1]
Affiliations:
[1] Opt-AI

NJUPT-IOT

Title:
Finetuned MambaIR
Members:
Fan Gao[1], Weidan Yan[1], Wenze Shao[1], Dengyin Zhang[1]
Affiliations:
[1] Nanjing University of Posts and Telecommunications, China

NuNu

Title:
A Simple Pipline for Image Shadow Removal
Members:
Bin Chen[1], Siqi Zhang[2], Yanxin Qian[2], Yuanbin Chen[1], Yuanbo Zhou[1], Tong Tong[1]
Affiliations:
[1] Fuzhou University, China
[2] Xiamen University, China

LSCM-HK

Title:
Fast Fourier Convolution Enhanced NAFNet for High Resolution Shadow Removal
Members:
Rongfeng Wei[1], Ruiqi Sun[1], Yue Liu[2]
Affiliations:
[1] Logistics and Supply Chain MultiTech R&D Centre, University of Hong Kong
[2] Sun Yat-sen University

KLETech-CEVI ShadowFighters

Title: ShadowNet: Hierarchical Network for Image Shadow Removal
Multi-scale Feature FusionNet
Members:
Nikhil Akalwadi[1,3], Amogh Joshi[1], Sampada Malagi[1,3], Chaitra Desai[1,3], Ramesh Ashok Tabib[1,2], Uma Mudenagudi[1,2]
Affiliations:
[1] Center of Excellence in Visual Intelligence (CEVI), KLE Technological University, Hubballi, Karnataka, INDIA
[2] School of Electronics and Communication Engineering, KLE Technological University, Hubballi, Karnataka, INDIA
[3] School of Computer Science and Engineering, KLE Technological University, Hubballi, Karnataka, INDIA

ataza

Title:
An Efficient Frequency Guided Image Enhancement Network for Low Light Image Enhancement and Shadow Removal
Members:
Ali Murtaza[1,2], Uswah Khairuddin[1,2], Ahmad ’Athif Mohd Faudzi[2]
Affiliations:

[1] Malaysia-Japan International Institute of Technology (MJIIT), University Teknologi Malaysia, Kuala Lumpur, , Malaysia
[2] Center for Artificial Intelligence and Robotics (CAIRO), Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

CVPR_IITRPR

Title:
A lightweight Architecture with Global-Local Feature Extraction for Image Shadow Removal
Members:
Adinath Dukre[1], Vivek Deshmukh[1], Shruti S. Phutke[2], Ashutosh Kulkarni[2], Santosh Kumar Vipparthi[2], Anil Gonde[1], Subrahmanyam Murala[3]
Affiliations:
[1] Shri Guru Gobind Singhji Institute Of Engineering and Technology, Nanded, India
[2] Computer Vision and Pattern Recognition Lab, Indian Institute of Technology Ropar, India
[3] CVPR Lab, School of Computer Science and Statistics, Trinity College Dublin, Ireland

TrioTechies

Title:
High Resolution Shadow Removal via Frequency-Aware Shadow Erasing Net
Members:
Arun karthik K[1], Manasa N[1], Shri Hari Priya[1]
Affiliations:
[1] School of Engineering, Shiv Nadar University, Chennai, India

FBU-ISR

Title:
Removing Shadows from Images Using an Autoencoder
Members:
Wei Hao [1], Xingzhuo Yan[2], Minghan Fu[3]
Affiliations:
[1] Fortinet, Inc.
[2] Bosch Investment Ltd.
[3] University of Saskatchewan.

References

[1] Cosmin Ancuti, Codruta O Ancuti, Florin-Alexandru Vasluianu, Radu Timofte, et al. NTIRE 2024 dense and non-homogeneous dehazing challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 2

[2] Nikola Banić, Egor Ershov, Artyom Panshin, Oleg Karasev, Sergey Korchagin, Shepelev Lev, Alexandr Startsev, Daniil Vladimirov, Ekaterina Zaychenkova, Dmitrii R Iarchuk, Maria Efimova, Radu Timofte, Arseniy Terekhin, et al. NTIRE 2024 challenge on night photography rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 2

[3] 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/CVF International Conference on Computer Vision*, pages 12504–12513, 2023. 5, 12

[4] Nicolas Chahine, Marcos V. Conde, Sira Ferradans, Radu Timofte, et al. Deep portrait quality assessment. a NTIRE 2024 challenge survey. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 2

[5] Jianhui Chang, Zhongnian Li, Yifan Jiang, Yinqiang Zheng, and Yoichi Sato. Focal frequency loss for image reconstruction and synthesis, 2020. 15

[6] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. *arXiv preprint arXiv:2204.04676*, 2022. 5, 9, 10, 15, 17

[7] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22367–22377, 2023. 11

[8] Zheng Chen, Zongwei WU, Eduard Sebastian Zamfir, Kai Zhang, Yulun Zhang, Radu Timofte, Xiaokang Yang, et al. NTIRE 2024 challenge on image super-resolution (×4): Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 2

[9] Zhihao Chen, Lei Zhu, Liang Wan, Song Wang, Wei Feng, and Pheng-Ann Heng. A multi-task mean teacher for semi-supervised shadow detection. In *CVPR*, 2020. 9

[10] Lu Chi, Borui Jiang, and Yadong Mu. Fast fourier convolution. *Advances in Neural Information Processing Systems*, 33:4479–4488, 2020. 10

[11] Jooyoung Choi, Jungbeom Lee, Chaehun Shin, Sungwon Kim, Hyunwoo Kim, and Sungroh Yoon. Perception prioritized training of diffusion models. In *CVPR*, 2022. 9

[12] Marcos V. Conde, Florin-Alexandru Vasluianu, Radu Timofte, et al. Deep raw image super-resolution. a NTIRE 2024 challenge survey. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 3

[13] Chaitra Desai, Nikhil Akalwadi, Amogh Joshi, Sampada Malagi, Chinmayee Mandi, Ramesh Ashok Tabib, Ujwala Patil, and Uma Mudenagudi. Lightnet: Generative model for enhancement of low-light images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2231–2240, 2023. 13

[14] Chaitra Desai, Sujay Benur, Ujwala Patil, and Uma Mudenagudi. Rsuigm: Realistic synthetic underwater image generation with image formation model. *ACM Trans. Multimedia Comput. Commun. Appl.*, apr 2024. Just Accepted. 12

[15] Chaitra Desai, Sujay Benur, Ramesh Ashok Tabib, Ujwala Patil, and Uma Mudenagudi. Depthcue: Restoration of underwater images using monocular depth as a clue. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, pages 196–205, January 2023. 13

[16] Chaitra Desai, Badduri Sai Sudheer Reddy, Ramesh Ashok Tabib, Ujwala Patil, and Uma Mudenagudi. Aquagan: Restoration of underwater images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 296–304, June 2022. 13

[17] Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat gans on image synthesis. In *NeurIPS*, 2021. 9

[18] Bin Ding, Chengjiang Long, Ling Zhang, and Chunxia Xiao. Argan: Attentive recurrent generative adversarial network for shadow detection and removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10213–10222, 2019. 2

[19] Graham D Finlayson, Mark S Drew, and Cheng Lu. Entropy minimization for shadow removal. *International Journal of Computer Vision*, 85(1):35–57, 2009. 2

[20] Graham D Finlayson, Steven D Hordley, and Mark S Drew. Removing shadows from images. In *European conference on computer vision*, pages 823–836. Springer, 2002. 2

[21] Minghan Fu, Yanhua Duan, Zhaoping Cheng, Wenjian Qin, Ying Wang, Dong Liang, and Zhanli Hu. Total-body low-dose ct image denoising using a prior knowledge transfer technique with a contrastive regularization mechanism. *Medical Physics*, 50(5):2971–2984, 2023. 17

[22] Minghan Fu, Huan Liu, Yankun Yu, Jun Chen, and Keyan Wang. Dw-gan: A discrete wavelet transform gan for nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 203–212, 2021. 17

[23] Maciej Gryka, Michael Terry, and Gabriel J. Brostow. Learning to remove soft shadows. *ACM Transactions on Graphics*, 2015. 2

[24] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. *arXiv preprint arXiv:2402.15648*, 2024. 15

[25] Lanqing Guo, Siyu Huang, Ding Liu, Hao Cheng, and Bihan Wen. Shadowformer: Global context helps image shadow removal. *arXiv preprint arXiv:2302.01650*, 2023. 2, 16

[26] Lanqing Guo, Chong Wang, Wenhan Yang, Siyu Huang, Yufei Wang, Hanspeter Pfister, and Bihan Wen. Shadowdiffusion: When degradation prior meets diffusion model for shadow removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14049–14058, 2023. 2

[27] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners, 2021. 5

[28] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 10

[29] Dikshit Hegde, Tejas Anvekar, Ramesh Ashok Tabib, and Uma Mudengudi. Da-ae: Disparity-alleviation auto-encoder towards categorization of heritage images for aggrandized 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5093–5100, 2022. 13

[30] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 2, 9

[31] Xiaowei Hu, Yitong Jiang, Chi-Wing Fu, and Pheng-Ann Heng. Mask-ShadowGAN: Learning to remove shadows from unpaired data. In *ICCV*, 2019. 2

[32] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 5

[33] 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/CVF International Conference on Computer Vision*, pages 5027–5036, 2021. 2

[34] Yeying Jin, Wei Ye, Wenhan Yang, Yuan Yuan, and Robby T Tan. Des3: Adaptive attention-driven self and soft shadow removal using vit similarity. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 2634–2642, 2024. 2

[35] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. *arXiv preprint arXiv:2304.02643*, 2023. 11

[36] Yuki Kondo, Riku Miyata, Fuma Yasue, Taito Naruki, and Norimichi Ukita. NTIRE 2023 Image Shadow Removal Challenge Technical Report: Team IIM_TTI. *arXiv*, 2024. 9

[37] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, may 2017. 3

[38] Tokoro Kyotaro, Akita Kazutoshi, and Norimichi Ukita. Burst super-resolution with diffusion models for improving perceptual quality. *IJCNN*, 2024. 9

[39] Edwin H Land and John J McCann. Lightness and retinex theory. *Josa*, 61(1):1–11, 1971. 12

[40] Hieu Le and Dimitris Samaras. Shadow removal via shadow image decomposition. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019. 2

[41] Hieu Le and Dimitris Samaras. From shadow segmentation to shadow removal. In *The IEEE European Conference on Computer Vision (ECCV)*, August 2020. 2

[42] Chongyi Li, Chun-Le Guo, Man Zhou, Zhexin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Embedding fourier for ultra-high-definition low-light image enhancement. *arXiv preprint arXiv:2302.11831*, 2023. 9

[43] Chunyuan Li and Others. Refusion: Enhancing perceptual quality with diffusion models for shadow removal. *arXiv preprint arXiv:2304.08291*, 2023. 14

[44] Chenghua Li, Bo Yang, Zhiqi Wu, Gao Chen, and Yihan Yu. Shadow removal based on diffusion, segmentation and

super-resolution models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2024. 10

[45] Xin Li, Kun Yuan, Yajing Pei, Yiting Lu, Ming Sun, Chao Zhou, Zhibo Chen, Radu Timofte, et al. NTIRE 2024 challenge on short-form UGC video quality assessment: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 3

[46] Jie Liang, Qiaosi Yi, Shuaizheng Liu, Lingchen Sun, Rongyuan Wu, Xindong Zhang, Hui Zeng, Radu Timofte, Lei Zhang, et al. NTIRE 2024 restore any image model (RAIM) in the wild challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 2

[47] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36, 2024. 11, 12

[48] Xiaohong Liu, Xiongkuo Min, Guangtao Zhai, Chunyi Li, Tengchuan Kou, Wei Sun, Haoning Wu, Yixuan Gao, Yuqin Cao, Zicheng Zhang, Xiele Wu, Radu Timofte, et al. NTIRE 2024 quality assessment of AI-generated content challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 2

[49] Xiaoning Liu, Zongwei WU, Ao Li, Florin-Alexandru Vasluianu, Yulun Zhang, Shuhang Gu, Le Zhang, Ce Zhu, Radu Timofte, et al. NTIRE 2024 challenge on low light image enhancement: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 3

[50] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Xie Saining. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 5

[51] Ziwei Luo, Fredrik K. Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B. Schön. Refusion: Enabling large-size realistic image restoration with latent-space diffusion models. In *CVPR*, 2023. 9

[52] Adrien Pavao, Isabelle Guyon, Anne-Catherine Letournel, Xavier Baró, Hugo Escalante, Sergio Escalera, Tyler Thomas, and Zhen Xu. Codalab competitions: An open source platform to organize scientific challenges. *Technical report*, 2022. 2, 3

[53] L. Qu, J. Tian, S. He, Y. Tang, and R. W. H. Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2308–2316, July 2017. 2

[54] Bin Ren, Yawei Li, Nancy Mehta, Radu Timofte, et al. The ninth NTIRE 2024 efficient super-resolution challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 2

[55] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, June 2022. 10

[56] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015. 9

[57] Yael Shor and Dani Lischinski. The shadow meets the mask: Pyramid-based shadow removal. *Computer Graphics Forum*, 27(2):577–586, Apr. 2008. 2

[58] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *ICML*, 2015. 9

[59] Zixuan Su, Jingjing Chen, Lei Pang, Chong-Wah Ngo, and Yu-Gang Jiang. Adaptive split-fusion transformer. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1169–1174. IEEE, 2023. 15

[60] Florin-Alexandru Vasluianu, Andrés Romero, Luc Van Gool, and Radu Timofte. Shadow removal with paired and unpaired learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 826–835, 2021. 2

[61] Florin-Alexandru Vasluianu, Tim Seizinger, and Radu Timofte. Wsrd: A novel benchmark for high resolution image shadow removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 2

[62] Florin-Alexandru Vasluianu, Tim Seizinger, Radu Timofte, Shuhao Cui, Junshi Huang, Shuman Tian, Mingyuan Fan, Jiaqi Zhang, Li Zhu, Xiaoming Wei, et al. Ntire 2023 image shadow removal challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1788–1807, 2023. 2, 4, 6, 9

[63] Florin-Alexandru Vasluianu, Tim Seizinger, Zongwei Wu, Rakesh Ranjan, and Radu Timofte. Towards image ambient lighting normalization. *arXiv preprint arXiv:2403.18730*, 2024. 2

[64] 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/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 2

[65] T. F. Y. Vicente, M. Hoai, and D. Samaras. Leave-one-out kernel optimization for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(3):682–695, March 2018. 2

[66] 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*, pages 1788–1797, 2018. 2

[67] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin C. K. Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *CoRR*, 2023. 9

[68] Longguang Wang, Yulan Guo, Juncheng Li, Hongda Liu, Yang Zhao, Yingqian Wang, Zhi Jin, Shuhang Gu, Radu Timofte, et al. NTIRE 2024 challenge on stereo image super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 2

[69] Yingqian Wang, Zhengyu Liang, Qianyu Chen, Longguang Wang, Jungang Yang, Radu Timofte, Yulan Guo, et al.

NTIRE 2024 challenge on light field image super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 2

[70] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 3

[71] Tai-Pang Wu, Chi-Keung Tang, Michael S. Brown, and Heung-Yeung Shum. Natural shadow matting. *ACM Trans. Graph.*, 26(2):8–es, June 2007. 2

[72] Ren Yang, Radu Timofte, et al. NTIRE 2024 challenge on blind enhancement of compressed image: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 2

[73] Pierluigi Zama Ramirez, Fabio Tosi, Luigi Di Stefano, Radu Timofte, Alex Costanzino, Matteo Poggi, et al. NTIRE 2024 challenge on HR depth from images of specular and transparent surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 2

[74] Waqas Zamir, Syed, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Khan, Shahbaz, and Ming-Hsuang Yan. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 5, 14

[75] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 3

[76] Xiao Feng Zhang, Chao Chen Gu, and Shan Ying Zhu. Spaformer: Transformer image shadow detection and removal via spatial attention. *arXiv e-prints*, pages arXiv–2206, 2022. 2

[77] Zhilu Zhang, Shuohao Zhang, Renlong Wu, Wangmeng Zuo, Radu Timofte, et al. NTIRE 2024 challenge on bracketing image restoration and enhancement: Datasets, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 2

[78] Suiyi Zhao. User study tool. https://github.com/suiyizhao/UserStudyTool, 2023. 6

[79] Chuanxia Zheng, Tat-Jen Cham, Jianfei Cai, and Dinh Phung. Bridging global context interactions for high-fidelity image completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11512–11522, June 2022. 10

[80] 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/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 5, 7

[81] Yurui Zhu, Xi Wang, Xueyang Fu, and Xiaowei Hu. Enhanced coarse-to-fine network fornbsp;image restoration fromnbsp;under-display cameras. In *Computer Vision – ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part V*, page 130–146, Berlin, Heidelberg, 2023. Springer-Verlag. 5