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

NTIRE 2023 HR NonHomogeneous Dehazing Challenge Report

Codruta O. Ancuti Cosmin Ancuti Florin-Alexandru Vasluianu Radu Timofte
Han Zhou Wei Dong Yangyi Liu Jun Chen Yangyi Liu Huan Liu
Liangyan Li Zijun Wu Jun Chen Yubo Dong Yuyan Li Tian Qiu Yu He
Yonghong Lu Yinwei Wu Zhenxiang Jiang Songhua Liu Xingyi Yang
Yongcheng Jing Bilel Benjdira Anas M. Ali Anis Koubaa Hao-Hsiang Yang
I-Hsiang Chen Wei-Ting Chen Zhi-Kai Huang Yi-Chung Chen Chia-Hsuan Hsieh
Hua-En Chang Yuan-Chun Chiang Sy-Yen Kuo Yu Guo Yuan Gao
Ryan Wen Liu Yuxu Lu Jingxiang Qu Shengfeng He Wenqi Ren Trung Hoang
Haichuan Zhang Amirsaeed Yazdani Vishal Monga Lehan Yang Alex Jiahao Wu
Tiancheng Mai Xiaofeng Cong Xuemeng Yin Xuefei Yin Hazim Emad
Ahmed Abdallah Yahya Yasser Dalia Elshahat Esraa Elbaz Zhan Li
Wenqing Kuang Ziwei Luo Fredrik K. Gustafsson Zheng Zhao Jens Sjölund
Thomas B. Schön Zhao Zhang Yanyan Wei Junhu Wang Suiyi Zhao
Huan Zheng Jin Guo Yangfan Sun Tianli Liu Dejun Hao Kui Jiang
Anjali Sarvaiya Kalpesh Prajapati Ratnadeep Patra Pragnesh Barik
Chaitanya Rathod Kishor Upla Kiran Raja Raghavendra Ramachandra
Christoph Busch

Abstract

This study assesses the outcomes of the NTIRE 2023 Challenge on Non-Homogeneous Dehazing, wherein novel techniques were proposed and evaluated on new image dataset called HD-NH-HAZE. The HD-NH-HAZE dataset contains 50 high resolution pairs of real-life outdoor images featuring nonhomogeneous hazy images and corresponding haze-free images of the same scene. The nonhomogeneous haze was simulated using a professional setup that replicated real-world conditions of hazy scenarios. The competition had 246 participants and 17 teams submitted solutions for the final testing phase. The proposed solutions demonstrated the cutting-edge in image dehazing technology.

1. Introduction

Haze is a naturally occurring phenomenon that can significantly diminish visibility in a scene as the distance increases, resulting in poor image quality. This atmospheric process is caused by the presence of small particles in the air that alter the properties of the environment. Consequently, hazy scenes are typically characterized by low contrast, diminished saturation, altered colors, and increased noise.

Recovering visual information from hazy images is important for various applications, such as aerial or ground surveillance, automatic traffic control and automatic driving. Therefore, image dehazing has attracted significant interest in the last decade

In recent years, there has been significant interest in the field of image dehazing [1,6,8,9,26,33,38,42,57,58] due to the importance of recovering visual information from hazy images for various applications, including aerial or ground surveillance, automatic traffic control, and automatic driving. More recently, image dehazing has been tackled by various CNN architectures [17,43,53,60,72].

A major challenge in objectively verifying and classifying dehazing algorithms is the absence of standardized test benchmarks. The evaluation of dehazing techniques is often complicated by the absence of a reference image or ground truth, and the lack of standard algorithms for detecting and measuring errors. Additionally, blind evaluation algorithms have been developed, but their inconsistent results may be due to a lack of validation on real images.

Maintaining consistent lighting conditions and achieving

*Codruta O. Ancuti (codruta.ancuti@gmail.com), Cosmin Ancuti, Florin-Alexandru Vasluianu and Radu Timofte are the NTIRE 2023 challenge organizers. The other authors participated in the challenge.
Appendix.A contains the authors' team names and affiliations.
https://cvlai.net/ntire/2023/

pixel-by-pixel correspondence between the reference and hazy image are crucial factors that pose significant challenges in the collection of such unclear images. As a result, the first image dehazing datasets (e.g. D-HAZY [5]) were synthesized and used information about scene depth and scene attenuation parameters.

However, a better solution involves recording outdoor haze-free images, and subsequently capturing the same scene with haze introduced through specialized equipment. The first realistic image dehazing datasets were introduced at the $NTIRE2018$ [2] image dehazing challenge. O-HAZE [7] contains 45 outdoor pair of images and and I-HAZE [4] contains 35 indoor pair of images. The hazy scenes of O-HAZE and I-HAZE datsests are characterised by light and homogeneous haze. Similarly, DENSE-HAZE [3] contains dense (homogeneous) hazy and ground-truth images and was employed by the NTIRE 2019 image dehazing challenge $NTIRE2019$ [14]. On the other hand, the first realistic non-homogeneous image dehazing datasets (NH-HAZE [10]) were used for the NTIRE 2020 [11] and 2021 [12] image dehazing challenges.

The NTIRE 2023 image dehazing challenge represents a step forward in benchmarking single image dehazing. It is based on an HD-NH-HAZE dataset that consists of 50 HD hazy images and their corresponding ground truth (haze-free) images of the same scene. HD-NH-HAZE contains real outdoor scenes with non-homogeneous haze generated using a professional haze setup. In NTIRE 2023 image dehazing challenge, we perform an objective evaluation by comparing the results of the competing methods with the reference images of the HD-NH-HAZE dataset.

This challenge is one of the NTIRE 2023 Workshop [1] series of challenges on: night photography rendering [54], HR depth from images of specular and transparent surfaces [70], image denoising [40], video colorization [37], shadow removal [59], quality assessment of video enhancement [44], stereo super-resolution [61], light field image super-resolution [62], image super-resolution (×4) [74], 360° omnidirectional image and video super-resolution [18], lens-to-lens bokeh effect transformation [20], real-time 4K super-resolution [21], HR nonhomogenous dehazing [13], efficient super-resolution [39].

2. Image Dehazing Challenge

The objectives of the NTIRE 2023 challenge on nonhomogeneous image dehazing are: (i) to extend the study and increase the performance in terms of image dehazing; (ii) to compare and promote the state-of-the-art solutions; and (iii) to emphasize the availability of high quality datasets, such as the non-homogeneous high resolution image dehazing (HD-NH-HAZE) dataset.

[1] https://cvlai.net/ntire/2023/

2.1. Nonhomogeneous image dataset

The NTIRE 2023 image dehazing challenge was built on the extended version of the former NH-Haze [10] dataset. The HD-NH-HAZE consists of 50 hazy images and their corresponding ground truth (haze-free) images of the same scene. The dataset contains real outdoor scenes with non-homogeneous haze generated using a professional haze setup. To introduce haze in the outdoor scenes we employed two professional haze machines which generate vapor particles with diameter size (typically 1 - 10 microns) similar to the atmospheric haze particles. For recording images we used Sony A7 III cameras remotely controlled. To ensure consistency between the unaffected areas of the haze in the image pairs, the camera parameters (shutter-speed / exposure-time, the aperture / F-stop, the ISO and the white-balance settings) were adjusted manually and then kept unchanged between the two consecutive recording sessions. We set the camera parameters (aperture-exposure-ISO), using an external exposure meter (Sekonic) and for white balance we used the medium gray card (18 $\%$ gray) of the color checker. The process of recording a pair of images took about 20-30 minutes.

2.2. Evaluation

For the NTIRE 2023 dehazing challenge we set a Codalab competition. In order to access the data and submit produced results to the evaluation server, each participant had to register to the Codalab competition and follow the phases set.

The Peak Signal-to-Noise Ratio (dB) and the Structural Similarity index (SSIM) computed between the inferred result and the ground truth image are the quantitative measures. The higher the score is, the better the restoration fidelity to the ground truth image is. Additionally, the LPIPS perceptual measure was deployed, for assessing the quality of the produced results. The final ranking was done after introducing the Mean Opinion Score (MOS), as a result of an user study set by the challenge organizers, with the results provided by the teams in the final phase of the challenge.

2.3. Challenge Phases

1. **Development phase:** In this phase, the first 40 high resolution (4000×6000) images of the HD-NH-HAZE dataset were made public on the challenge platform [51], for the participants to use them in the development process of their solution.

2. **Validation phase:** Another set consisting of 5 images was made public to the participants. Using the validation server [51], without getting access to the ground-truth images, the participants were able to validate their solutions.

3. **Testing phase:** The test set, consisting of 5 images, was published on the challenge platform. Using the validation server [51], they uploaded their predicted haze-free images for evaluation, thus being ranked in terms of PSNR and SSIM [63]. Their best submission, along with the factsheet containing information about the proposed solution, team members and the software implementing the method, they prepared the final submission. For the final ranking, a user study was performed, and the Mean Opinion Score (MOS) was used to further evaluate the perceptual quality of the results produced by the ranked teams.

3. Challenge Results

The NTIRE2023 High Resolution NonHomogeneous Dehazing Challenge had a number of 246 registered participants, with 17 teams submitting their results, solution description, available codes and the team description for the final phase of the challenge. Solutions corresponding to those teams are ranked in the Table 1. The participants proposed novel solutions, characterized by a significant level of performance in terms of both reconstruction fidelity, and perceptual properties.

As you can observe in Table 1, the metric with the highest correlation to the user study results, quantified as the Mean Opinion Score (MOS) metric, is the PSNR. The perceptual properties based metrics, such as LPIPS [73] and SSIM [63], were used to differentiate similar results. However, the results corresponding to the top performing solutions in terms of perceptual metrics have, as expected, consistent results both in reconstruction fidelity and perceptual properties.

4. Challenge Methods

4.1. DWT-FFC-GAN

They proposed a novel two-branch generative adversarial network [75] for high-resolution non-homogenous dehazing. The structure diagram is shown as Figure 1. The first branch is the DWT-FFC frequency branch, which aims to learn the feature mapping between hazy images and clear images. The encoder of this branch consists of three DWT downsampling blocks, which are used to detect both high-frequency and low-frequency features [28], and three FFC residual blocks, which allow for a wide receptive field by simultaneously leveraging spectral and spatial information and make the reconstructed image more realistic and perceptual [56]. The second branch, prior knowledge branch, is designed to bring additional information from large dataset image classification task to the current challenge. They use the first three stages of ConvNeXt pretrained on ImageNet 1K dataset to build the encoder due to its outstanding performance on the classification tasks [32, 47]. Then, for the decoder module, several upsampling layers are employed and each upsanpling layer contains pixel-shuffle layers block and a attention module. Specifically, pixel-shuffle blocks are introduced to decrease the computational burden and make the size of the feature maps gradually recover to the original resolution, attention blocks enable our model to identify the dynamic hazy patterns. Finally, final recovered clear images are generated by combining the outputs of each branch via a simple and effective fusion.

4.2. ITB Dehaze

The proposed method [45] follows [28, 69] to use a two-branch structure. Inspired by [28, 41, 69], they further develop a data pre-processing technique to shift the distribution of augmented data towards that of target data. This data-centric design shows significant improvements on the challenge results. In the first branch of the model, they adopt the Swin Transformer V2 model [46] that is pre-trained on ImageNet [22] to extract pretinent multi-level features of hazy images. The Swin Transformer is proved to be more powerful than traditional CNN based architectures, for example on COCO object detection and instance segmentation tasks. Therefore, they choose it to serve as the backbone for feature extraction. Adopting the idea of transfer learning, they use the ImageNet pre-trained model to initialize the Swin Transformer, which enables the system to leverage the knowledge learned in previous low-level tasks. Trained from scratch, the second branch complements the first one by exclusively working on the domain of target data. With no down-sampling and up-sampling operations, this branch operates in the full-resolution mode, thus extracts features distinct from the ones obtained by the first branch. The fusion tail aggregates the outputs from both branches and produces dehazed images.

4.3. [Mask]

They present a novel approach which involves a two-phase procedure. The first phase, NonHomogeneous Dehazing, is based on proximal gradient descent (PGD) [35] deep unfolding, incorporating a residual degradation learning strategy in the gradient descent step and a MixS^2 Transformer [23] Network as the proximal mapping module. In the second phase, Super-Resolution Refinement, they propose the MixS^2SR network to use the results generated in the first phase for super-resolution and refinement, which comprises multiple MixS^2 Blocks and a PixelShuffle module. This approach enables to effectively remove nonhomogeneous haze in high resolution images, while also improving the quality of the images through super-resolution refinement.

Rank	Team	Username	PSNR↑	SSIM↑	LPIPS↓	MOS↑	Params.(M)	Runtime(s)	Device	Extra data
1	DWT-FFC GAN	zhouh115	$22.87_{(2)}$	$0.71_{(2)}$	$0.346_{(2)}$)	$8.07_{(1)}$	373	23.3	RTX2080 Ti	NTIRE20, 21
2	ITB Dehaze	lillian	$22.96_{(1)}$	$0.71_{(1)}$	$0.345_{(1)}$	$7.85_{(3)}$	110	9.0	2×TitanXP	NTIRE20,21
3	[Mask]	ShawnDong98	$22.18_{(4)}$	$0.7_{(5)}$	$0.401_{(5)}$	$7.125_{(6)}$	9.31	0.125	A100	No
4	NUSRICQ DEHAZING	Yinwei_Wu	$21.97_{(8)}$	$0.68_{(9)}$	$0.38_{(3)}$	$7.9_{(2)}$	25.58	3.01	4× RTX3090	NH-HAZE, O-HAZE, IHAZE, DENSE-HAZE
5	PSU TEAM	Anas	$22.27_{(3)}$	$0.7_{(4)}$	$0.439_{(10)}$	$7.4_{(5)}$	5.0	480	RTX8000	No
6	NTU607-dehaze	HaoqiangYang	$22.11_{(5)}$	$0.71_{(3)}$	$0.442_{(11)}$	$7.1_{(7)}$	101.5	52.1	V100	NTIRE18, 19, 20
7	MIPCer	YuanGao	$21.75_{(10)}$	$0.7_{(6)}$	$0.404_{(6)}$	$6.95_{(8)}$	2.39	3.79	A100	No
8	iPAL-LightDehaze	lightdehaze	$22.01_{(7)}$	$0.70_{(7)}$	$0.384_{(4)}$	$5.35_{(13)}$	2.6	0.72	TitanXP	No
9	Xsourse	xsourse	$22.09_{(6)}$	$0.65_{(13)}$	$0.556_{(16)}$	$7.65_{(4)}$	n/a	37.0	A4000	No
10	Xiaofeng Cong	CongXiaofeng	$21.86_{(9)}$	$0.67_{(11)}$	$0.492_{(12)}$	$6.9_{(9)}$	24.7	10.0	V100	No
11	MengFeiHome	XuefeiYin	$21.00_{(12)}$	$0.68_{(10)}$	$0.411_{(8)}$	$5.15_{(14)}$	5	0.005	RTX3060	No
12	CANT HAZE	HazimEmad	$20.95_{(14)}$	$0.69_{(8)}$	$0.415_{(9)}$	$5.15_{(14)}$	80	0.15	T4/K80	No
13	DBCDN	ManHing	$21.62_{(11)}$	$0.66_{(12)}$	$0.501_{(13)}$	$5.85_{(11)}$	n/a	14.04	RTX3090	No
14	IR-SDE	ir-sde	$20.83_{(16)}$	$0.61_{(16)}$	$0.406_{(7)}$	$6.3_{(10)}$	78	5.0	A100	No
15	LVGroup HFUT	yan.wei	$20.97_{(13)}$	$0.65_{(14)}$	$0.521_{(14)}$	$5.625_{(12)}$	5.0	0.005	A40	NTIRE20,21
16	DH-AISP	DH-AISP	$20.91_{(15)}$	$0.62_{(15)}$	$0.521_{(15)}$	$4.7_{(16)}$	n/a	n/a	n/a	no
17	SVNIT NTNU	AnjaliSarvaiya	$16.32_{(17)}$	$0.47_{(17)}$	$0.815_{(17)}$	$0.75_{(17)}$	n/a	n/a	n/a	no

Table 1. Quantitative results of the NTIRE 2023 Non Homogeneous Dehazing Challenge . 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.

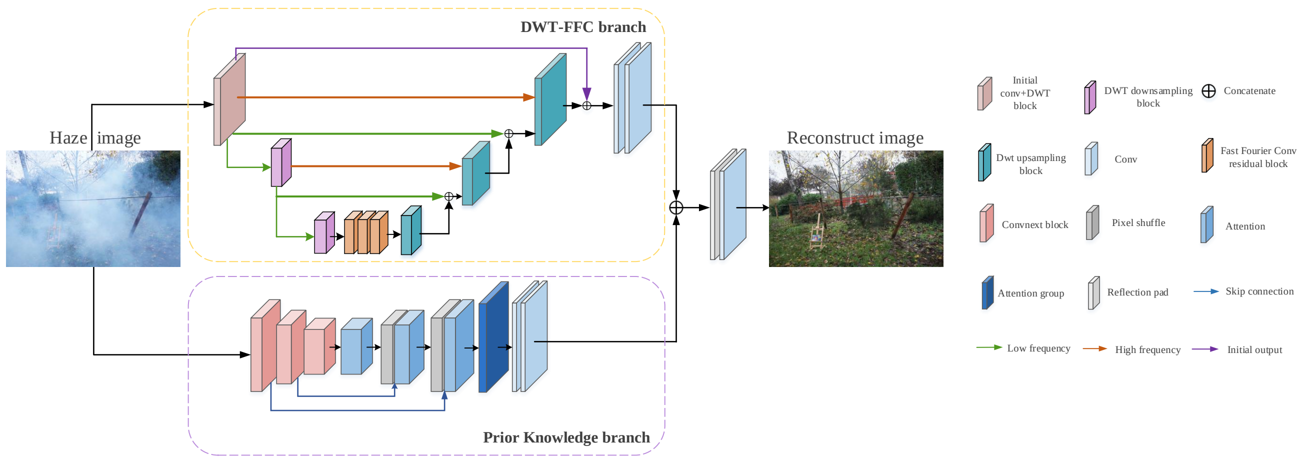

Figure 1. The network architecture of the solution proposed by team DWT-FFC-GAN.

4.4. NUSRICQ DEHAZING

In the first stage this approach down-sample the 4000×6000 resolution images to low resolution and input the low resolution into the WeatherDiffusion dehazing module [50]. After that, they use a series of DehazeFormer [55] with pixel shuffle as dehazing finetuning module and super resolution module to gradually increase the resolution of the images and finally reach the size of the original high resolution input. The whole processing method is shown in Figure 4.

4.5. PSU TEAM

The team designed a framework [16] called SGLC (Streamlined Global and Local Features Combinator) for handling High-Resolution hazy images. The framework SGLC is designed to scale the ability of any popular Dehazing model to work on High-Resolution images without resizing them or cropping them. To ensure this scalability, they used two patching methods: Grid patching and Window patching. They used the Uformer model as an example. The Grid patching was conceived to solve the problem of conservation of the global features during the training. SGLC is composed of two blocks. The first block is the Global Features Generator (GFG), which uses the Grid patching and a Dehazing Model (DM) trained on this specific patch type to generate the first dehazed image. This first dehazed image is supposed to contain robust global features taken from the whole scene's peculiarities. The second block is the Local Features Enhancer (LFE) which enhances the quality of the First dehazed images using local features by the intermediate of an Enhancer Model (EM) trained on Window patches of the data. They used Self-Supervised Learning before training the DM and EM models to increase their understanding of the data. In addition, they designed a customized loss function to help them learn the High-Frequency information in the data. Also, they used a smoothing algorithm (Multiple Overlapping Patches Smoother (MOPS)) to improve the quality of the final dehazed image. SGLC architecture is presented in Figure 5.

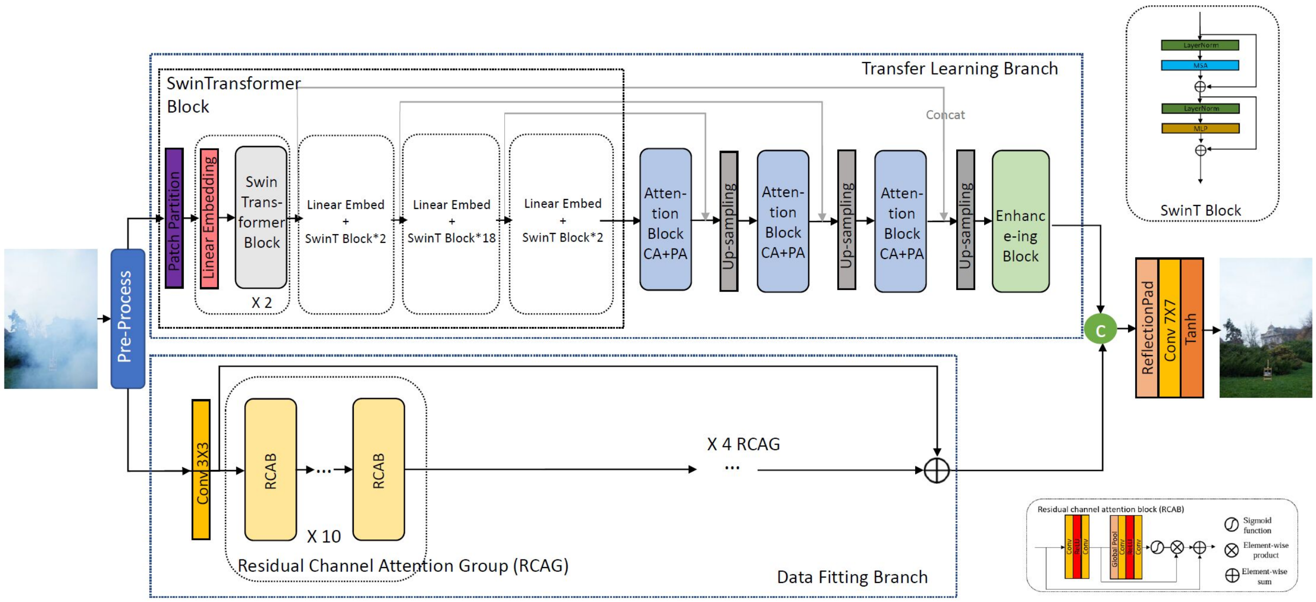

Figure 2. The network architecture of the solution proposed by team ITBdehaze.

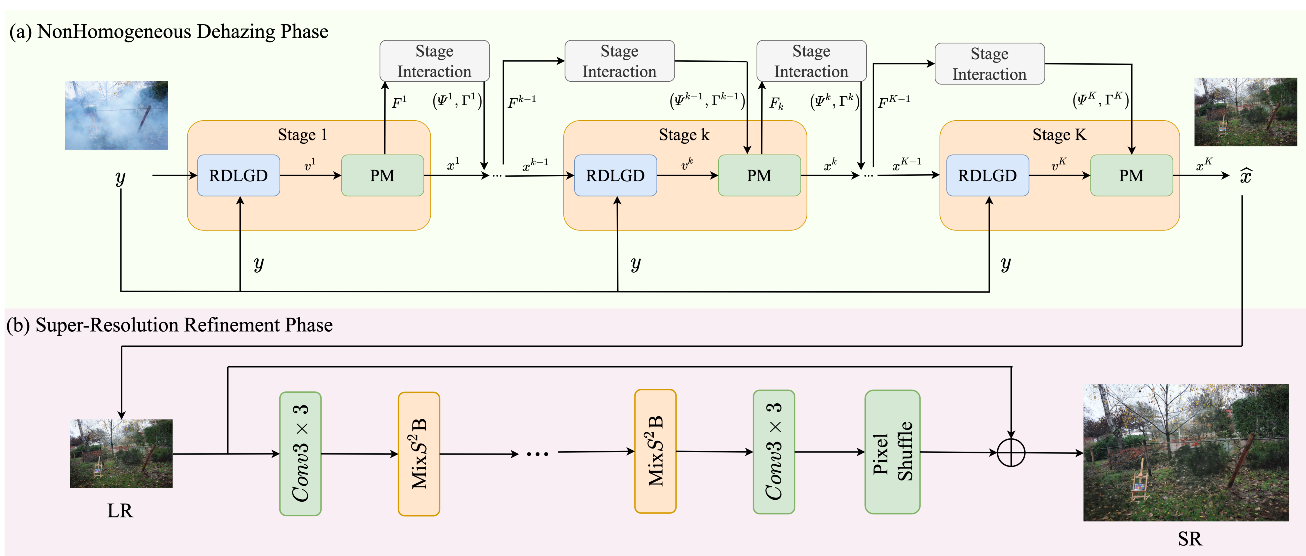

Figure 3. The network architecture of the solution proposed by team [Mask].

4.6. NTU607-dehaze

The overall method and experimental results are published in [65]. This method applies the semantic information to optimize the network and the team design two semantic-based loss functions. The first loss is the semantic corresponding loss L_{sem} that helps the dehazed images have an identical semantic representation to that of the ground truth, and it is written as:

$$L_{sem} = |S(I) - S(J)| \tag{1}$$

where $|\cdot|$ is the absolute value. $I \in R^{W\times H\times 3}$ and $J \in R^{W\times H\times 3}$ are the dehazed image and the ground truth. S is the semantic segmentation model and $S(.) \in R^{W\times H\times n}$ is the semantic segmentation map with n classes. This loss makes the model handle the dense hazy regions. The second loss is the semantic color tone consistency loss L_{sc} that can adjust the color tones based on separated classes. This loss function is written as:

$$L_{sc} = \sum_{i=1}^{n} |C_{I,i} - C_{J,i}| \tag{2}$$

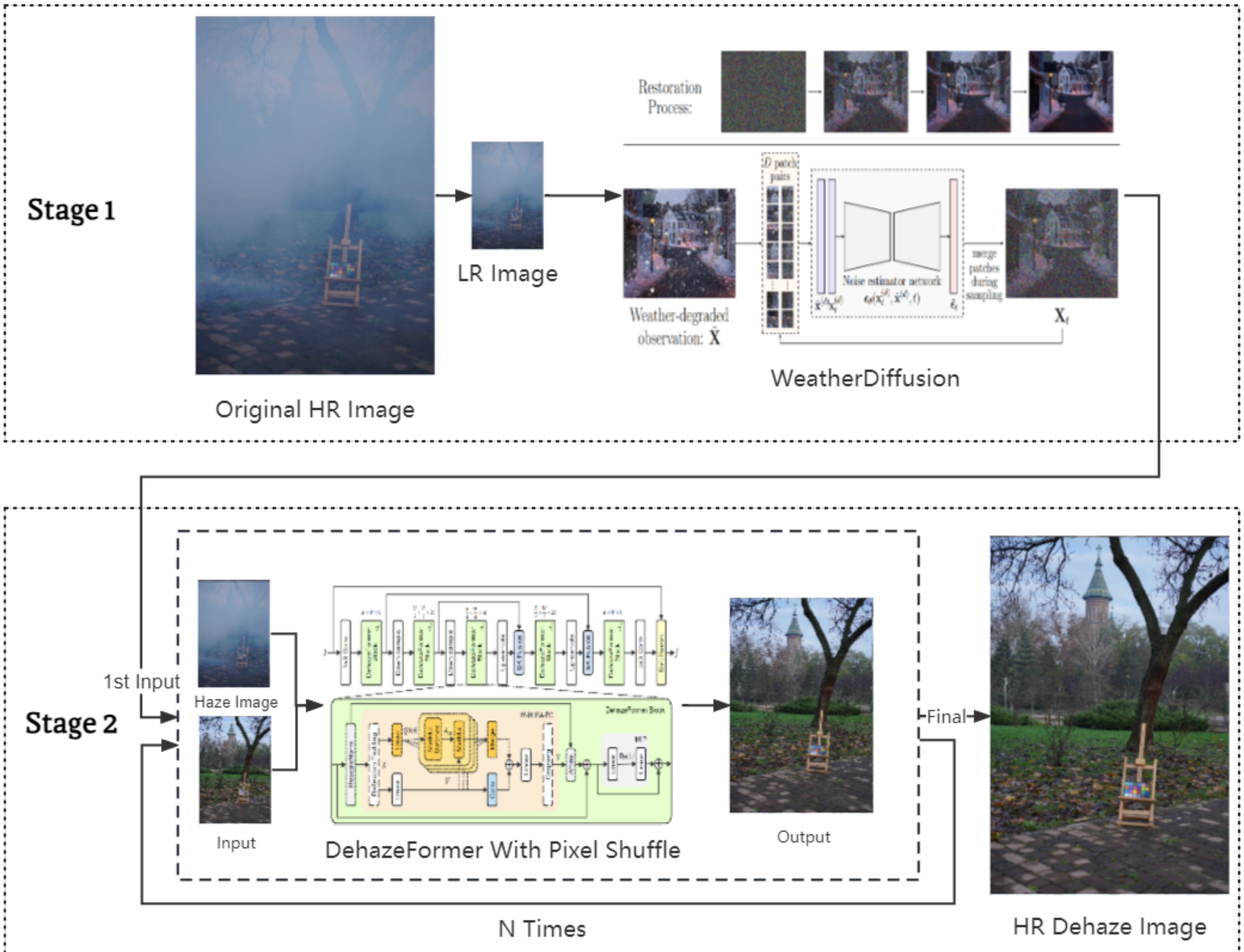

Figure 4. The network architecture of the solution proposed by team NUSRICQ.

where $C_{I,i}$ means the average color tone of the i th class of I. Specifically, $C_{I,i}$ can be written as:

$$C_{I,i} = \frac{1}{n} \sum_{S(x) \in i} I(x) \tag{3}$$

The architecture for non-homogeneous dehazing is presented in fig:NTU607. This network is based on DW-GAN [27] and contains two branches: Discrete wavelet transform (DWT) branch [68] and Res2Net [29, 66] branch. Based on its structure, they replace 3×3 convolutions with 5×5 convolutions in the whole model.

Besides the proposed semantic loss functions L_{sem} and L_{sc}, we also train the network with three extra loss functions: Chaboinner loss L_{cha} [15], wavelet SSIM loss L_{W-SSIM} [67, 68] and perceptual loss L_{Per} [36]. The overall loss function is written as:

$$L_{Total} = L_{cha} + 1.1 L_{W-SSIM} + 0.1 L_{Per} + 0.1 L_{sem} + 0.1 L_{sc} \tag{4}$$

4.7. MIPCer

This team proposed a novel self-paced semi-curricular attention network (termed SCANet) [31] for non-homogeneous image dehazing, which focuses on the enhancement of haze-occluded regions. The network structure is shown in Figure 7. Their method consists of an attention generator network (AGN) and a scene reconstruction network (SRN). The AGN is composed of multiple dual-attention basic units (DAUs) to generate attention feature maps, while the SRN is an encoder-decoder network to reconstruct haze-free images. Specifically, to better distinguish between non-uniform haze and clear images, they use the luminance differences of images

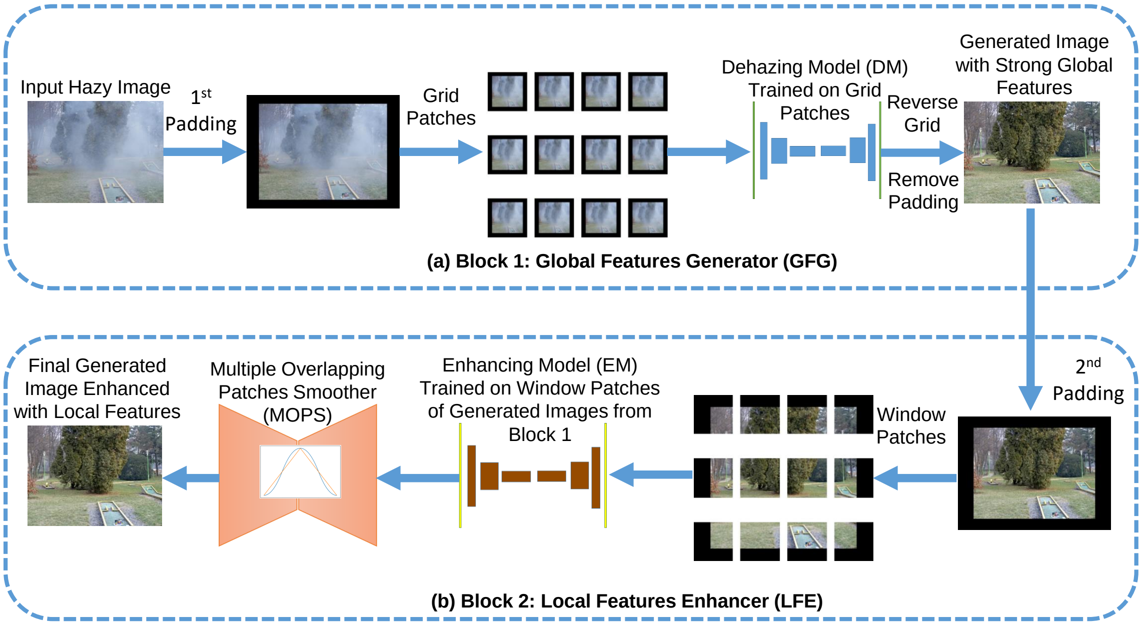

Figure 5. The network architecture of the solution proposed by team PSU TEAM.

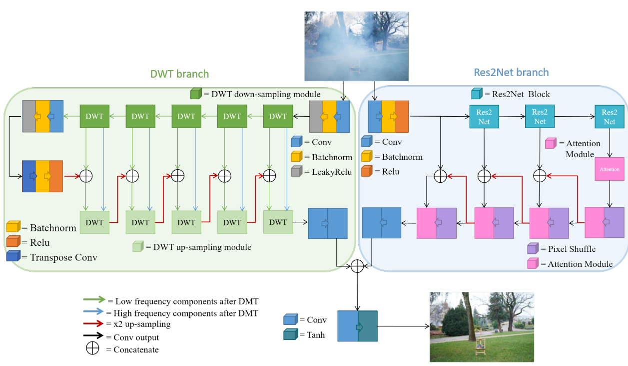

Figure 6. The network architecture of the solution proposed by team NTU607-dehaze.

to restrict the attention map. Note that multi-objective prediction tasks (i.e., simultaneously obtain haze-free feature map and attention map) tend to increase learning ambiguity. Inspired by [24], they introduce a self-paced semi-curricular learning strategy to supervise the attention map in the early stage of training, which enables the network to achieve faster convergence ability.

For the self-paced semi-curricular learning, the attention map M_g generated by AGN and the ground truth M_{GT} are fused to generate the final attention map M. Let λ be the trade-off parameter, M can be expressed mathematically as

$$M = \lambda \cdot M_g + (1 - \lambda) \cdot M_{GT}. \tag{5}$$

In particular, the trade-off parameter can be dynamically adjusted through the smooth L1 loss $\mathcal{L}^a_{sl1}$ of the attention map, i.e.,

$$\lambda = \begin{cases} 0, & \text{if } \mathcal{L}^a_{sl1} > 0.1, \\ \frac{\mathcal{L}^a_{sl1} - 0.1}{0.1 - 0.05}, & \text{if } 0.1 \geq \mathcal{L}^a_{sl1} > 0.05, \\ 1, & \text{if } \mathcal{L}^a_{sl1} \leq 0.05. \end{cases} \tag{6}$$

Equation 6 is used to adjust the specific gravity of M_g and M_{GT}. In the initial stage, M mainly consists of M_{GT} to alleviate the learning ambiguity due to the large value of $\mathcal{L}^a_{sl1}$. As $\mathcal{L}^a_{sl1}$ decreases, the proportion of the attention map M_g generated by the network will continue to increase. When $\mathcal{L}^a_{sl1}$ is less than 0.05, M will only consist of M_g. Meanwhile, they only adopt the semi-curricular learning strategy in the first 25% epochs to avoid the model's over-reliance on M_{GT}.

To obtain more satisfactory results, this network adopts five loss functions, i.e.,

$$\mathcal{L}_{\text{joint}} = \gamma_1 \mathcal{L}_{sl1} + \gamma_2 \mathcal{L}^a_{sl1} + \gamma_3 \mathcal{L}_p + \gamma_4 \mathcal{L}_{\text{MS-SSIM}} + \gamma_5 \mathcal{L}_a, \tag{7}$$

where $\mathcal{L}_{\text{joint}}$, $\mathcal{L}_{sl1}$, $\mathcal{L}^a_{sl1}$, $\mathcal{L}_p$, $\mathcal{L}_{\text{MS-SSIM}}$, and $\mathcal{L}_a$ are the total loss, smooth l1 loss [30] for the clear image, smooth l1 loss

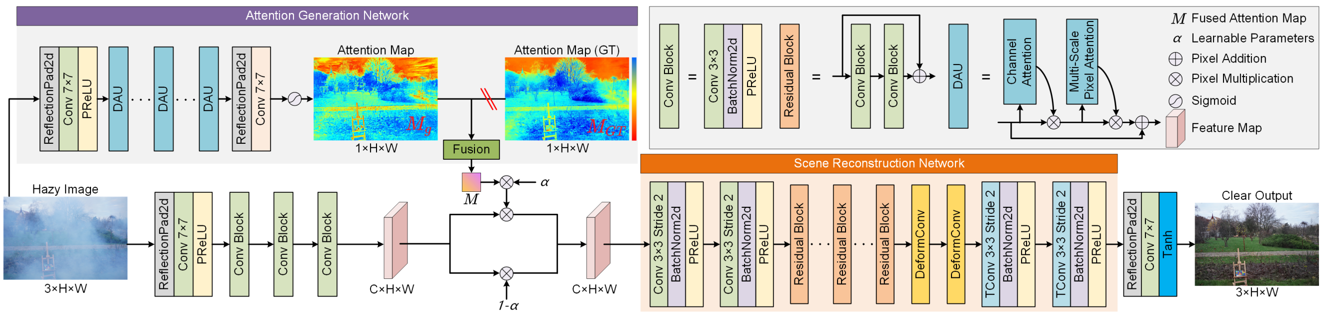

Figure 7. The network architecture of the solution proposed by team MIPCer.

for attention map, perception loss [36], multi-scale structure similarity loss [64], and adversarial loss [77], respectively. The penalty coefficients are set as $\lambda_1 = 1.0$, $\lambda_2 = 0.3$, $\lambda_3 = 0.01$, $\lambda_4 = 0.5$, $\lambda_5 = 0.0005$.

4.8. iPAL-LightDehaze

A two-stage lightweight neural network for non-homogeneous haze removal is designed. The proposed method named **TransER** [34] includes two models corresponding to two stages which are a lightweight **Trans**Conv Fusion Dehaze (TFD) network and a tiny **E**nsemble **R**econstruction (LER) network. The proposed TransER method is shown in Figure 8. In the first model (TFD), the network has one encoder and three distinct decoders to jointly estimate the scene information. Inspired by [55], they designed new TransConv Dehaze blocks in the encoder and decoders. We employ the feature attention module (FAM) [52] to extract parallelly the feature maps with the Vision Transformer method . They combined the local information from FAM and global information from Vision Transformer via Selective Kernel Fusion module [55]. The objective of two encoders is to estimate the ambient light (*A*) and the transmission map (*t(x)*) of the mathematical equation of the haze problem which is described:

$$I(x) = J(x) \times t(x) + A \times (1 - t(x)) \tag{8}$$

where $I(x)$ and $J(x)$ are the haze and haze-free image respectively, $t(x)$ is transmission map, while A denotes the atmospheric light. Therefore, our TFD model can generate haze-free images by inverse haze model as follows:

$$\hat{J}_{At}(x) = \frac{I(x) - \hat{A} \times (1 - \hat{t}(x))}{\hat{t}(x) + \epsilon} \tag{9}$$

where $\hat{J}_{At}(x)$ is the reconstructed parameter-based haze-free image (they define $P(x) = \hat{J}_{At}(x)$) in the Figure 8. $\hat{A}$ and $\hat{t}(x)$ denote the outputs of Decoder Atmospheric Light *A* and Decoder Transmission Map *t(x)*. ϵ is a very small value that helps to avoid division by zero and x is pixel location. The remaining decoder is Decoder Direct Haze-free *J(x)* whose job is to recover the haze-free image directly from the input hazy image (denoted by $S(x)$ in Figure 8). By running the experiments in different homogeneous and nonhomogeneous datasets , *S(x)* has better performance than *P(x)* in regions with dense haze, while for regions with shallow haze, *P(x)* performs better. As a result, they designed a new lightweight ensemble reconstruction network (LER) to produce the final haze-free image. In the second stage, LER is targeted to reconstruct both estimates pseudo direct haze-free *S(x)* which has good performance in dense haze scenes and parameter-based haze-free *P(x)* performing well in shallow haze regions. The LER model includes two encoders and only one decoder. Two encoders extract the feature maps from two pseudo input images (*S(x)* and *P(x)*), these features are added adaptively by AFA that can help to preserve the information. In particular, the output of the feature extraction stage is f_{out}, and the formula is as follow:

$$f_{out} = \text{AFA}[f(S(x)), f(P(x))] \tag{10}$$
$$= \sigma(\theta) \times f(S(x)) + (1 - \sigma(\theta)) \times f(P(x)) \tag{11}$$

where σ is $Sigmoid$ activation function, θ is a learnable factor, $f(.)$ denotes the output of Encoder, AFA is adaptively feature addition. The one decoder will recover to get the final haze-free image. They employed the gate convolution block from [55] to design both two encoders and a decoder. To make the LER model generate more natural clean images, they designed an additional teacher network to transfer knowledge from the intermediate feature. The teacher network which has exactly the same architecture as LER is trained as a simple task to reconstruct from clean image to clean image. They train the teacher network to assist in training the dehazing network by providing prior knowledge of distillation loss.

4.9. Xsourse

The proposed is based on DW-GAN [28]. Although the dataset in this competition has an image resolution of 6000

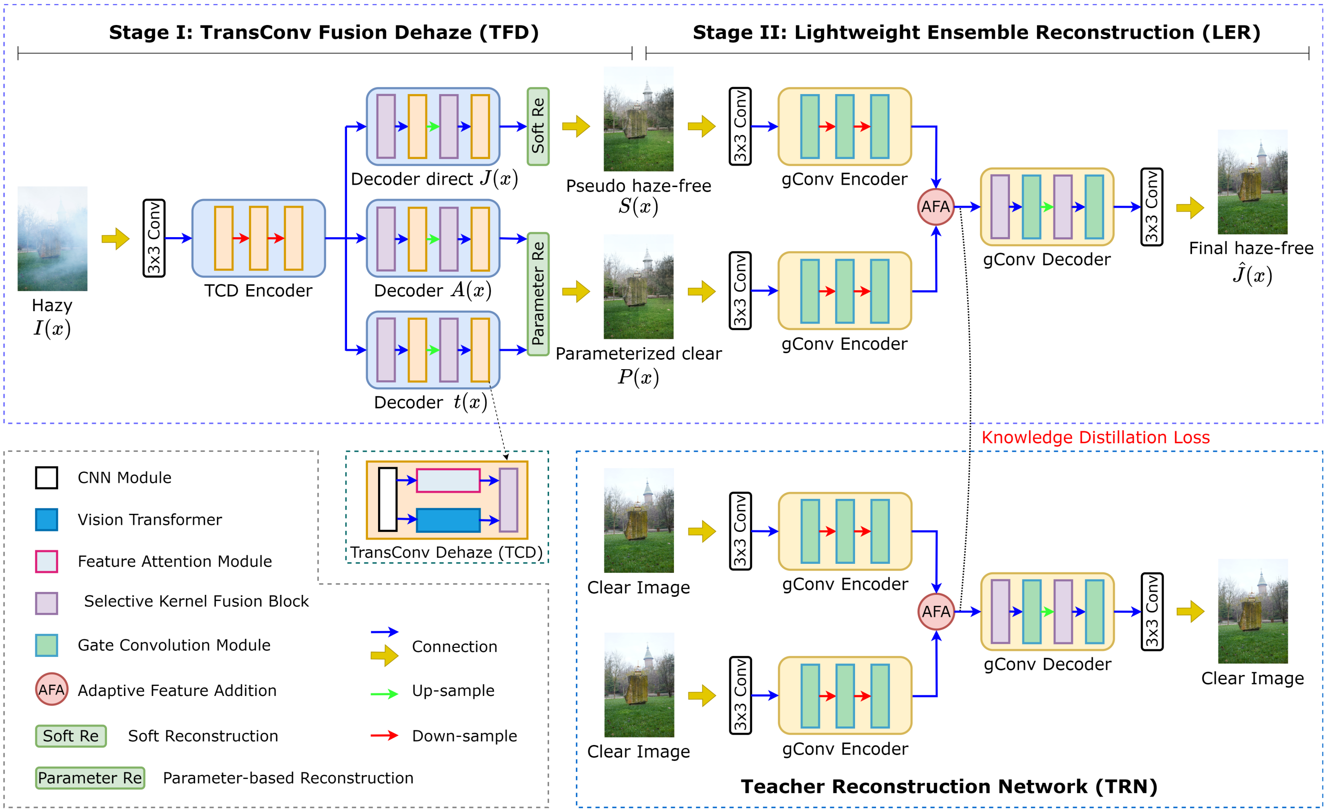

Figure 8. The network architecture of the solution proposed by team iPAL-LightDehaze.

x 4000, considering the limitations of GPU memory, they still resized the images to 1600 x 1200 as input images for training. During the training process, their baseline used random crop 256 x 256 size images as input. Their method use MixUp as an augmentation. They set the MixUp parameter alpha to 0.2. Within a batch, they performed linear weighted interpolation on random hazy images and ground truth image pairs to achieve dataset augmentation effects. They also added an additional standard deviation loss to train the model to reduce the very high or low values.

4.10. Xiaofeng Cong

The proposed dehazing network is composed of stacked transformer blocks and convolutional blocks [55]. The forward calculation includes two downsampling and two upsampling processes, which can be regarded as an encoding-decoding structure. The training process uses conventional MSE and SSIM losses.

4.11. MengFeiHome

This method uses DWGAN for non-homogeneous haze removal, which includes two processing branches: the discrete wavelet transform branch and the information branch, and a discriminator [28] The DWT branch is constructed by Unet, and each feature scale includes an encoder, a decoder, and a large number of skip links. For the information branch, they used ImageNet pretrains Res2Net. The discriminator is to reduce artifacts by using smooth L1 loss function, perceptual loss function, adversarial loss function, and total loss function. The DWT branch can preserve more image texture details, while the information branch can prevent overfitting and make the network ability significantly improve.

4.12. CANT HAZE

The method is built on the NAFNET [19] and Adaptive White Balancing (AWB).

4.13. DBCDN

The proposed method is based on an architecture that uses an isomorphic dual-branch network as its backbone, which effectively combines complementary information for various computer vision tasks. Further, a fusion module and a super resolution (SR) module are designed to boost the performance of image dehazing. Unlike most existing dual-branch networks, which adopt two heterogeneous branches, this network incorporates two completely identical branches [28]and distinguishes them only by an L1 loss

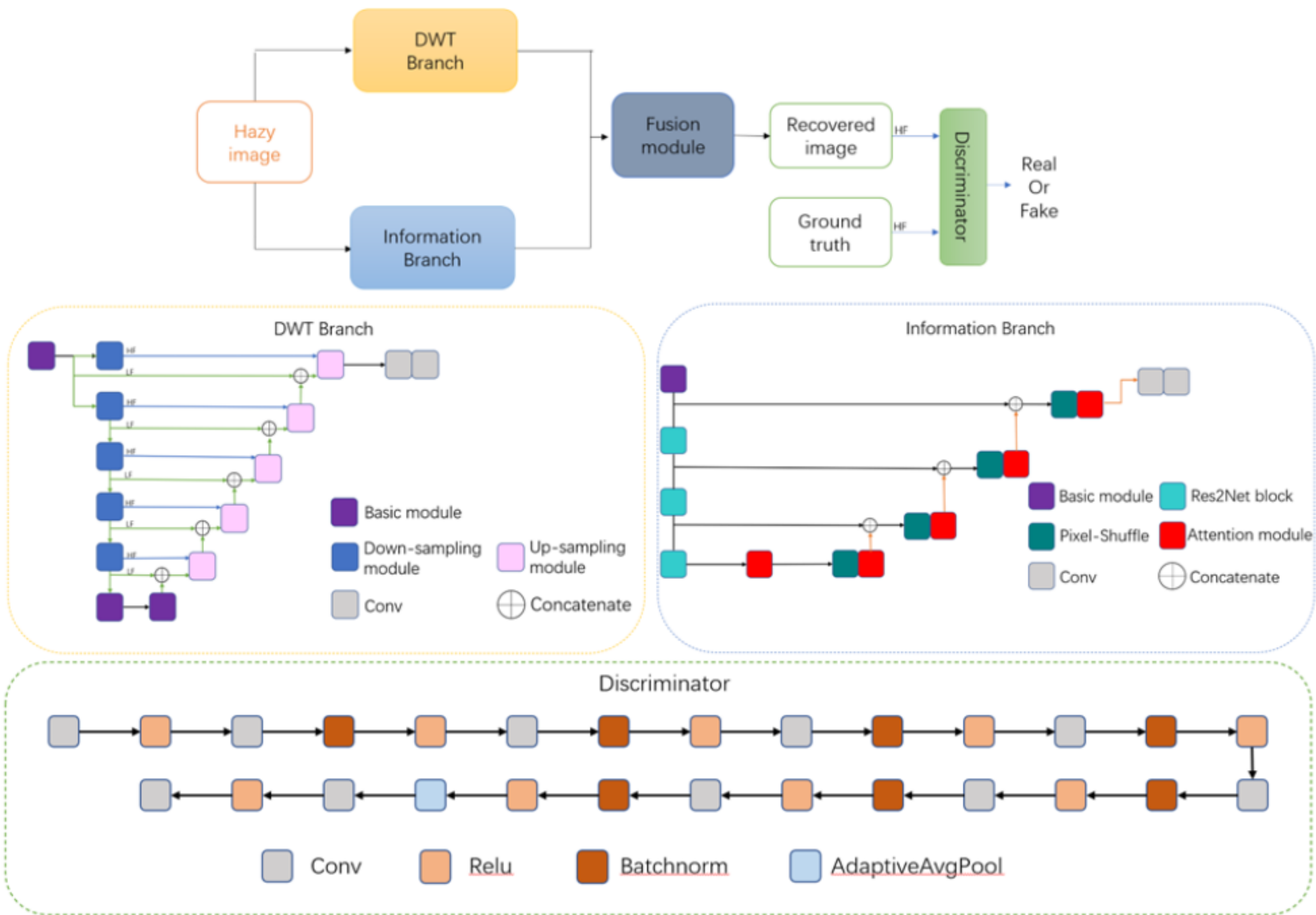

Figure 9. The network architecture of the solution proposed by team MengFeiHome.

classifier. In addition, considering the efficiency, large input hazy images are zoomed out to save the computing resources, and an SR module [71] is attached, which compensates for details lost during the downsampling process.

4.14. IR-SDE

The method leverages diffusion models for realistic image restoration [49]. Specifically, they use IR-SDE [48] as the base diffusion framework, which can naturally transform the high-quality image to its degraded counterpart, without caring how complicated the degradation is (even for real-world degradation). As shown in Figure 11, IR-SDE is a mean-reverting SDE in which the forward process is defined as:

$$dx = \theta_t\,(\mu - x)dt + \sigma_t dw, \tag{12}$$

where θ_t and σ_t are time-dependent positive parameters that characterize the speed of the mean-reversion and the stochastic volatility, respectively. Since it is an Ito SDE, it can be derived a reverse-time SDE:

$$dx = \big[\theta_t\,(\mu - x) - \sigma_t^2\,\nabla_x \log p_t(x)\big]dt + \sigma_t d\hat{w}. \tag{13}$$

At test time, the only unknown part is the score $\nabla_x \log p_t(x)$ of the marginal distribution at time t. As other diffusion-based models, they employ a CNN network to estimate the score backward from the low-quality image to the high-quality image.

Unlike other L_1 loss normally trained networks which usually produce smooth/blurry results, the proposed Reffusion aims to achieve a highly competitive perceptual performance as well as the distortion scores (PSNR). In addition, they further improve the results by updating the score-network from U-Net to NAFNet [19], which is more efficient and also has a good performance compared with recent Transformers. To adaptively insert the scalar time into the network, they construct a simple multi-layer perceptron to learn two pairs of scale-shift parameters and apply them to the features with affine transforms. Such a network leads to better learning of score function conditioned on current state x_t, original low-quality image x_t, and time t.

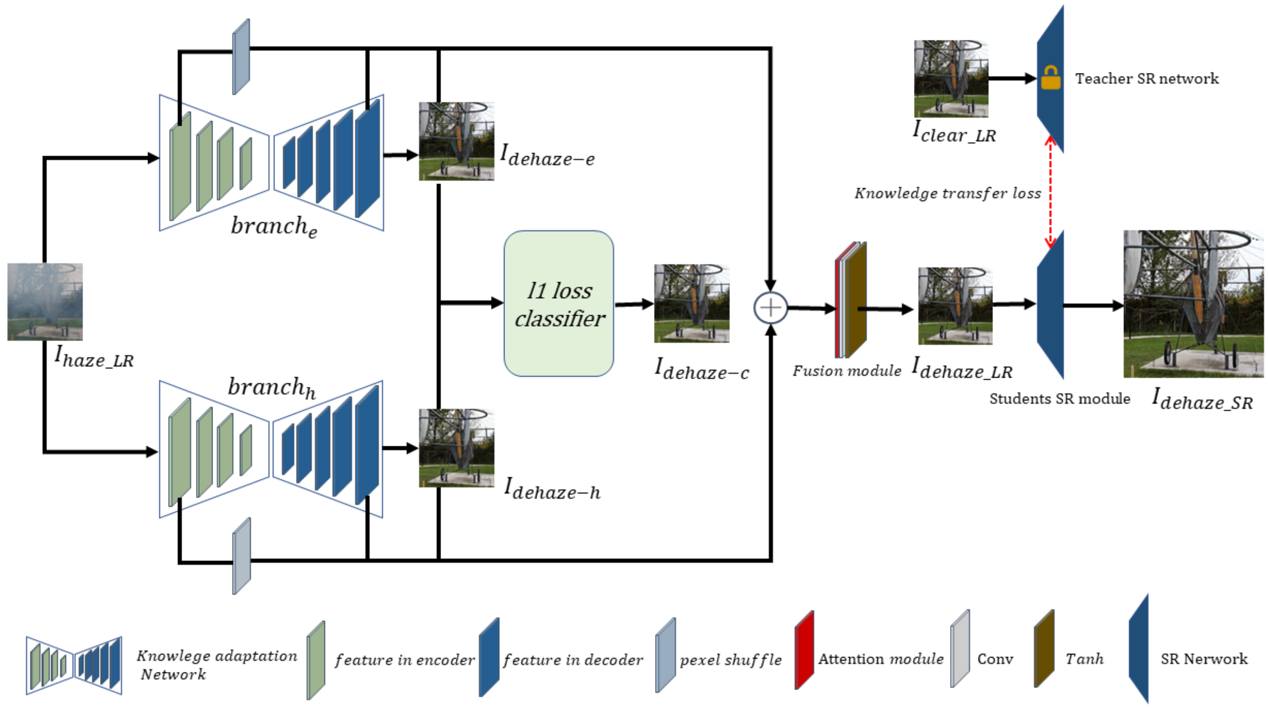

Figure 10. The network architecture of the solution proposed by team DBCDN.

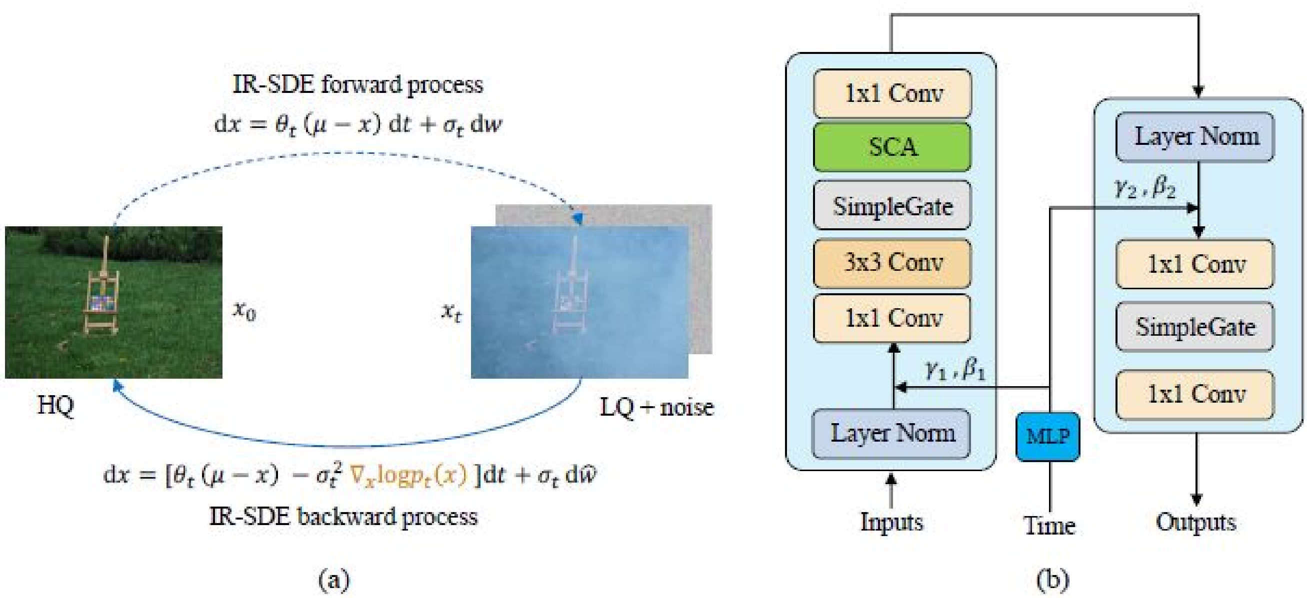

Figure 11. The network architecture of the solution proposed by team IR-SDE.

4.15. LVGroup

This team introduces a novel two-branch dehazing network that can extract multi-scale features and image details, which is of great favor to dehazing. The overall network is shown in Figure 12. Specifically, the upper branch is inspired by [25] and we mainly extract the multi-scale features to dehaze, which is achieved by the multi-scale input and the cross-scale fusion modules. Besides, the designed bottom branch is inspired by [71], which is used to learn the detailed information from hazy images to enhance the coarse dehazed images obtained by the upper branch. Finally, they concatenate their features to obtain the final haze-free images.

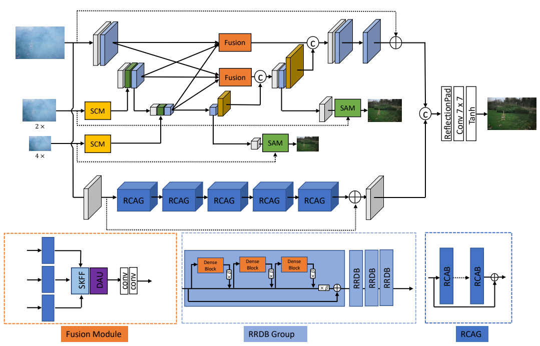

Figure 12. The network architecture of the solution proposed by team LVGroup.

4.16. DH-AISP

The method introduces an elaborated dual-branch model to process low and high frequency image respectively, so that the color and details of the image can be better processed. The network structure is shown in Figure 13.

For the first branch, they propose an end-to-end model to learn the projection relationship between haze data and Low frequency data. They follow the UNet++ [76] architecture to construct the first branch. Compared with Unet architecture, Unet++ enjoys better capability to aggregate the multi-scale features to reconstruct more robust results. Meanwhile, the joint constraints of L1 and VGG loss are used to encourage more realistic contents generation. For the second branch, they adopted a Unet to learn high-frequency information of images. In order to obtain a larger receptive field, they have used dilated convolution. Finally, they add the results of the two branches to get the final result.

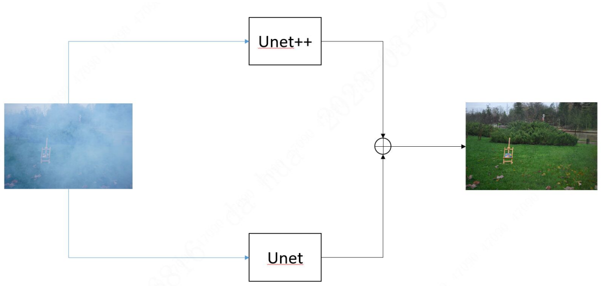

Figure 13. The network architecture of the solution proposed by team DH-AISP.

4.17. SVNIT NTNU

In this method the hazy image is applied to the input of the network and it is passed to extract salient features from it. The hazy image is first passed through a series of RES and RCA blocks then via series of RCA and RPCA blocks and finally through a series of RCPA and RPCSA blocks, aim is to gradually extract finer details from the hazy image. The architecture uses the Exponential Linear Unit (ELU) activation function to improve learning performance at each layer in efficient manner. A new and core element of the proposed architecture is the partially densely connected design of ResBlock that preserves the high frequency details of the hazy image by retaining salient features . The kernel sizes (i.e., 3×3) adopted in ResBlock recover details distributed at local and global regions.

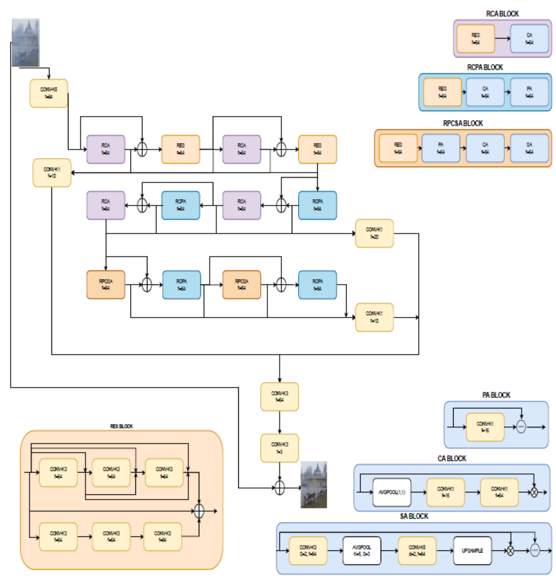

Figure 14. The network architecture of the solution proposed by team SVNIT NTNU.

5. Conclusion

In NTIRE 2023 Image Dehazing Challenge 246 participants took part, and during the final phase, 17 teams were ranked. These teams explored various architectures and introduced innovative solutions, surpassing the current results. The designs demonstrated in previous years were effectively integrated as fundamental components, presenting immense possibilities for advancement.

The final ranking was based on the Mean Opinion Score produced by our user study. The ranking was primarily influenced by the accuracy of the recovered images since it had the strongest correlation with user feedback on the presented outcomes.

Acknowledgments This work was partially supported by the Humboldt Foundation. Part of this work was supported by a grant of the Ministry of Research, Innovation and Digitization, CCCDI - UEFISCDI, project number PN-III-P2-2.1-PED-2021-1572, within PNCDI III. Part of this work was supported by European Marie Skvodowska-Curie Individual Fellowships H2020-MSCA-IF-2019. We thank the NTIRE 2023 workshop and challenges sponsors: Sony Interactive Entertainment, Meta Reality Labs, ModelScope, ETH Zürich (Computer Vision Lab) and University of Würzburg (Computer Vision Lab).

A. Teams and Affiliations

NTIRE 2023 Team

Title:
NTIRE 2023 Challenge on NonHomogeneous Dehazing
Members:
Codruta O. Ancuti[1] *(codruta.ancuti@gmail.com)*, Cosmin Ancuti[1,2], Florin-Alexandru Vasluianu[3], Radu Timofte[3,4]
Affiliations:
[1] ETcTI, Universitatea Politehnica Timisoara, Romania
[2] ICTEAM, UCL, Belgium
[3] Computer Vision Lab, University of Wuerzburg, Germany
[4] Computer Vision Lab, ETH Zurich, Switzerland

DWT-FFC GAN

Title:
Breaking Through the Haze: An Advanced Non-Homogeneous Dehazing Method based on Fast Fourier Convolution and ConvNeXt
Members:
Han Zhou[1] *(zhouh115@mcmaster.ca)*, Wei Dong[2], Yangyi Liu[1], Jun Chen[1]
Affiliations:
[1] Department of Electrical and Computer Engineering, McMaster University
[2] Department of Electrical and Computer Engineering, University of Alberta

ITB Dehaze

Title:
A Data-Centric Solution to NonHomogeneous Dehazing via Vision Transformer.
Members:
Yangyi Liu (liu5@mcmaster.ca), Huan Liu, Liangyan Li, Zijun Wu, Jun Chen
Affiliations:
McMaster University

[Mask]

Title:
Residual Degradation Learning Unfolding Framework with Mixing Priors across Spectral and Spatial for NonHomogeneous Dehazing
Members:
Yubo Dong (ybdong@stu.xidian.edu.cn), Yuyan Li, Tian Qiu, Yu He, Yonghong Lu
Affiliations:
Xidian University

NUSRICQ DEHAZING

Title:
Two-Stage Dehazing for High Resolution Non-Homogeneous Image
Members:
Yinwei Wu[1] *(wyw3276027@gmail.com)*, Zhenxiang Jiang[1], Songhua Liu[2], Xingyi Yang[2], Yongcheng Jing[3]
Affiliations:
[1] National University of Singapore (Chongqing) Research Institute
[2] National University of Singapore
[3] University of Sydney

PSU TEAM

Title:
Streamlined Global and Local Features Combinator (SGLC) for High Resolution Image Dehazing
Members:
Bilel Benjdira (bbenjdira@psu.edu.sa), Anas M. Ali, Anis Koubaa
Affiliations:
Robotics and Internet-of-Things Laboratory, Prince Sultan University, Riyadh 12435, Saudi Arabia

NTU607 dehaze

Title:
Semantic guidance learning for non-homogeneous dehazing
Members:
Hao-Hsiang Yang[2] *(islike8399@gmail.com)*, I-Hsiang Chen[2], Wei-Ting Chen[1], Zhi-Kai Huang[2], Yi-Chung Chen[3], Chia-Hsuan Hsieh[3], Hua-En Chang[2], Yuan-Chun Chiang[2], Sy-Yen Kuo[2]
Affiliations:
[1]Graduate Institute of Electronics Engineering, National Taiwan University, Taiwan
[2]Department of Electrical Engineering, National Taiwan University, Taiwan

[3]Graduate Institute of Communication Engineering, National Taiwan University, Taiwan

MIPCer

Title:
SCANet: Self-Paced Semi-Curricular Attention Network for Non-Homogeneous Image Dehazing
Members:
Yu Guo[1] *(yuguo@whut.edu.cn)*, Yuan Gao[1], Ryan Wen Liu[1], Yuxu Lu[1], Jingxiang Qu[1], Shengfeng He[2], Wenqi Ren[3]
Affiliations:
[1] Wuhan University of Technology, China
[2] Singapore Management University, Singapore
[3] Sun Yat-sen University, China

iPAL-LightDehaze

Title:
Integration of Lightweight TransConv Fusion Networks and Knowledge Distillation for Non-homogeneous Dehazing
Members:
Trung Hoang (tvh5567@psu.edu), Haichuan Zhang, Amirsaeed Yazdani, Vishal Monga
Affiliations:
Electrical Engineering Department - Pennsylvania State University

Xsourse

Title:
Ensemble for Quality-Aware High Resolution NonHomogeneous Dehazing
Members:
Lehan Yang (lehan.yang@sydney.edu.au), Alex Jiahao Wu, Tiancheng Mai
Affiliations:
The University of Sydney

Xiaofeng Cong

Title:
A Dehazing Method based on Visual Transformer.
Members:
Xiaofeng Cong (cxf_svip@163.com)
Affiliations: Southeast University

MengFeiHome

Title:
GAN Networks with Two Branch Dehazing
Members:
Xuemeng Yin[1] , Xuefei Yin[2] (22722112@bjtu.edu.cn)
Affiliations:
[1]University of California, Los Angeles
[2]Beijing Jiaotong University

CANT HAZE

Title:
NAFNET with Canny edge detection and adaptive white balancing
Members:
Yahya Yasser (Yahyayasser837@gmail.com), Ahmed Abdallah, Hazim Emad, Dalia Elshahat, Esraa Elbaz
Affiliations:
Mansoura Univeristy

DBCDN

Title:
Dual-branch complementary dehazing network
Members:
Zhan Li (lizhan@jnu.edu.cn) , Wenqing Kuang
Affiliations:
College of Information Science and Technology, Jinan University

IR-SDE

Title:
Refusion: Enabling Large-Size Realistic Image Restoration with Latent-Space Diffusion Models [49]
Members:
Ziwei Luo (ziwei.luo@it.uu.se), Fredrik K. Gustafsson, Zheng Zhao, Jens Sjölund, Thomas B. Schön
Affiliations:
Department of Information Technology, Uppsala University

LVGroup HFUT

Title:
A Two-branch Multi-scale Dehazing Network
Members:
Zhao Zhang (cszzhang@gmail.com), Yanyan Wei, Junhu Wang, Suiyi Zhao, Huan Zheng
Affiliations:
Hefei University of Technology

DH-AISP

Title:
Members:
Jin Guo (gjguo-jingj@gmail.com), Jin Guo, Yangfan Sun, Tianli Liu, Dejun Hao, Kui Jiang

Affiliations:
Zhejiang Dahua Technology

SVNIT NTNU

Title:
Non-Homogeneous Image Dehazing using Layered Multi Attention Network

Members:
Anjali Sarvaiya[1] *(anjali.sarvaiya.as@gmail.com)*, Kalpesh Prajapati[1], Ratnadeep Patra[1], Pragnesh Barik[1], Chaitanya Rathod[1], Kishor Upla[1], Kiran Raja[2], Raghavendra Ramachandra[2], Christoph Busch[2]
Affiliations:
[1]Sardar Vallabhbhai National Institute of Technology, India
[2]Norwegian University of Science and Technology, Norway

References

[1] C.O. Ancuti and C. Ancuti. Single image dehazing by multi-scale fusion. *IEEE Transactions on Image Processing*, 22(8):3271–3282, 2013. 1

[2] C. Ancuti, C.O. Ancuti, R. Timofte, L. Van Gool, and L. Zhang et al. NTIRE 2018 challenge on image dehazing: Methods and results. *IEEE CVPR, NTIRE Workshop*, 2018. 2

[3] C. Ancuti, C. O. Ancuti, M. Sbert, and R. Timofte. DENSE-HAZE: A benchmark for image dehazing with dense-haze and haze-free images. *IEEE ICIP*, 2019. 2

[4] C. Ancuti, C. O. Ancuti, R. Timofte, and C. De Vleeschouwer. I-HAZE: a dehazing benchmark with real hazy and haze-free indoor images. *International Conference on Advanced Concepts for Intelligent Vision Systems*, 2018. 2

[5] C. Ancuti, C. O. Ancuti, and Christophe De Vleeschouwer. D-Hazy: A dataset to evaluate quantitatively dehazing algorithms. *IEEE ICIP*, 2016. 2

[6] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and A. C. Bovick. Day and night-time dehazing by local airlight estimation. In *IEEE Transactions on Image Processing*, 2020. 1

[7] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and R. Timofte. O-HAZE: a dehazing benchmark with real hazy and haze-free outdoor images. *IEEE CVPR, NTIRE Workshop*, 2018. 2

[8] C. O. Ancuti, C. Ancuti, C. Hermans, and P. Bekaert. A fast semi-inverse approach to detect and remove the haze from a single image. *ACCV*, 2010. 1

[9] C. O. Ancuti, C. Ancuti, M. Sbert, and C. De Vleeschouwer. Color channel compensation (3c): A fundamental preprocessing step for image enhancement. In *IEEE Transactions on Image Processing*, 2020. 1

[10] C. O. Ancuti, C. Ancuti, and R. Timofte. NH-HAZE: An image dehazing benchmark with nonhomogeneous hazy and haze-free images. *IEEE CVPR, NTIRE Workshop*, 2020. 2

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

[12] Codruta O Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, Radu Timofte, et al. NTIRE 2021 nonhomogeneous dehazing challenge report. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 2

[13] 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

[14] C. O. Ancuti, C.Ancuti, R. Timofte, L. Van Gool, and L. Zhang et al. NTIRE 2019 challenge on image dehazing: Methods and results. *IEEE CVPR, NTIRE Workshop*, 2019. 2

[15] Jonathan T Barron. A general and adaptive robust loss function. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 6

[16] Bilel Benjdira, Anas M. Ali, and Anis Koubaa. Streamlined global and local features combinator (sglc) for high resolution image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 4

[17] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 2016. 1

[18] Mingdeng Cao, Chong Mou, Fanghua Yu, Xintao Wang, Yinqiang Zheng, Jian Zhang, Chao Dong, Ying Shan, Gen Li, Radu Timofte, et al. NTIRE 2023 challenge on 360° omnidirectional image and video super-resolution: Datasets, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[19] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, pages 17–33. Springer, 2022. 9, 10

[20] Marcos V Conde, Manuel Kolmet, Tim Seizinger, Thomas E. Bishop, Radu Timofte, et al. Lens-to-lens bokeh effect transformation. NTIRE 2023 challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[21] Marcos V Conde, Eduard Zamfir, Radu Timofte, et al. Efficient deep models for real-time 4k image super-resolution. NTIRE 2023 benchmark and report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[22] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. 3

[23] Yubo Dong, Dahua Gao, Tian Qiu, Yuyan Li, Minxi Yang, and Guangming Shi. Residual degradation learning un-

folding framework with mixing priors across spectral and spatial for compressive spectral imaging. *arXiv preprint arXiv:2211.06891*, 2022. 3

[24] Yong Du, Junjie Deng, Yulong Zheng, Junyu Dong, and Shengfeng He. Dsdnet: Toward single image deraining with self-paced curricular dual stimulations. *Comput. Vision Image Understanding*, 230:103657, 2023. 7

[25] Cho Sung-Jin et al. Rethinking coarse-to-fine approach in single image deblurring. *in Proc. IEEE ICCV*, 2021. 11

[26] Raanan Fattal. Single image dehazing. *SIGGRAPH*, 2008. 1

[27] 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 (CVPR) Workshops*, pages 203–212, June 2021. 6

[28] Minghan Fu, Huan Liu, Yankun Yu, Jun Chen, and Keyan Wang. Dwgan:a discrete wavelet transform gan for non-homogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 3, 8, 9

[29] Shanghua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip HS Torr. Res2net: A new multi-scale backbone architecture. *IEEE transactions on pattern analysis and machine intelligence*, 2019. 6

[30] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015. 7

[31] Yu Guo, Yuan Gao, Wen Liu, Yuxu Lu, Jingxiang Qu, Shengfeng He, and Wenqi Ren. Scanet: Self-paced semi-curricular attention network for non-homogeneous image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 6

[32] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, and Piotr Dollar. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 3

[33] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *In IEEE CVPR*, 2009. 1

[34] Trung Hoang, Haichuan Zhang, Amirsaeed Yazdani, and Vishal Monga. Transer: Hybrid model and ensemble-based sequential learning for non-homogenous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 8

[35] Eunsung Jo and Jae-Young Sim. Multi-scale selective residual learning for nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 3

[36] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016. 6, 8

[37] Xiaoyang Kang, Xianhui Lin, Kai Zhang, Zheng Hui, Wangmeng Xiang, Jun-Yan He, Xiaoming Li, Peiran Ren, Xuansong Xie, Radu Timofte, et al. NTIRE 2023 video colorization challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[38] L. Kratz and K. Nishino. Factorizing scene albedo and depth from a single foggy image. *ICCV*, 2009. 1

[39] Yawei Li, Yulun Zhang, Luc Van Gool, Radu Timofte, et al. NTIRE 2023 challenge on efficient super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[40] Yawei Li, Yulun Zhang, Luc Van Gool, Radu Timofte, et al. NTIRE 2023 challenge on image denoising: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[41] Huan Liu, Zijun Wu, Liangyan Li, Sadaf Salehkalaibar, Jun Chen, and Keyan Wang. Towards multi-domain single image dehazing via test-time training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5831–5840, 2022. 3

[42] Q. Liu, X. Gao, L. He, and W. Lu. Single image dehazing with depth-aware non-local total variation regularization. *IEEE Trans. Image Proc.*, 2018. 1

[43] R. Liu, L. Ma, Y. Wang, and L. Zhang. Learning converged propagations with deep prior ensemble for image enhancement. *IEEE Trans. Image Proc.*, 2019. 1

[44] Xiaohong Liu, Xiongkuo Min, Wei Sun, Yulun Zhang, Kai Zhang, Radu Timofte, Guangtao Zhai, Yixuan Gao, Yuqin Cao, Tengchuan Kou, Yunlong Dong, Ziheng Jia, et al. NTIRE 2023 quality assessment of video enhancement challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[45] Yangyi Liu, Huan Liu, Li, Liangyan, Zijun Wu, and Jun Chen. A data-centric solution to nonhomogeneous dehazing via vision transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 3

[46] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, et al. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12009–12019, 2022. 3

[47] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 3

[48] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Image restoration with mean-reverting stochastic differential equations. *arXiv preprint arXiv:2301.11699*, 2023. 10

[49] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 10, 14

[50] Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 4

[51] 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

[52] 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*, volume 34, pages 11908–11915, 2020. 8

[53] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang. Single image dehazing via multi-scale convolutional neural networks. *Proc. European Conf. Computer Vision*, 2016. 1

[54] Alina Shutova, Egor Ershov, Georgy Perevozchikov, Ivan A Ermakov, Nikola Banic, Radu Timofte, Richard Collins, Maria Efimova, Arseniy Terekhin, et al. NTIRE 2023 challenge on night photography rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[55] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 2023. 4, 8, 9

[56] Roman Suvorov, Elizaveta Logacheva, Anastasia Remizova Anton Mashikhin, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions. In *In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022. 3

[57] Robby T. Tan. Visibility in bad weather from a single image. *In IEEE Conference on Computer Vision and Pattern Recognition*, 2008. 1

[58] J.-P. Tarel and N. Hautiere. Fast visibility restoration from a single color or gray level image. *In IEEE ICCV*, 2009. 1

[59] Florin-Alexandru Vasluianu, Tim Seizinger, Radu Timofte, et al. NTIRE 2023 image shadow removal challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[60] A. Wang, W. Wang, J. Liu, and N. Gu. AIPNet: Image-to-image single image dehazing with atmospheric illumination prior. *IEEE Trans. Image Proc.*, 2019. 1

[61] Longguang Wang, Yulan Guo, Yingqian Wang, Juncheng Li, Shuhang Gu, Radu Timofte, et al. NTIRE 2023 challenge on stereo image super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[62] Yingqian Wang, Longguang Wang, Zhengyu Liang, Jungang Yang, Radu Timofte, Yulan Guo, et al. NTIRE 2023 challenge on light field image super-resolution: Dataset, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[63] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *The IEEE Transactions on Image Processing (TIP)*, 13(4):600–612, 2004. 3

[64] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003. 8

[65] Hao-Hsiang Yang, Chang Chen, I-Hsiang, Chia-Hsuan Hsieh, Hua-En, Yuan-Chun Chiang, Yi-Chung Chen, Wei-Ting Huang, Zhi-Kai Chen, and Sy-Yen Kuo. Semantic guidance learning for high-resolution non-homogeneous dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023. 5

[66] Hao-Hsiang Yang, Wei-Ting Chen, and Sy-Yen Kuo. S3net: A single stream structure for depth guided image relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 276–283, 2021. 6

[67] Hao-Hsiang Yang, Chao-Han Huck Yang, and Yi-Chang James Tsai. Y-net: Multi-scale feature aggregation network with wavelet structure similarity loss function for single image dehazing. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020. 6

[68] Hao-Hsiang Yang, Chao-Han Huck Yang, and Yu-Chiang Frank Wang. Wavelet channel attention module with a fusion network for single image deraining. In *IEEE International Conference on Image Processing (ICIP)*, 2020. 6

[69] Yankun Yu, Huan Liu, Minghan Fu, Jun Chen, Xiyao Wang, and Keyan Wang. A two-branch neural network for non-homogeneous dehazing via ensemble learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 3

[70] Pierluigi Zama Ramirez, Fabio Tosi, Luigi Di Stefano, Radu Timofte, et al. NTIRE 2023 challenge on hr depth from images of specular and transparent surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[71] et al. Zhang, Yulun. Image super-resolution using very deep residual channel attention networks. *in Proc. ECCV*, 2018. 10, 12

[72] H. Zhang, V. Sindagi, and V. M. Patel. Multi-scale single image dehazing using perceptual pyramid deep network. *IEEE CVPR*, 2018. 1

[73] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *The European Conference on Computer Vision (ECCV)*, September 2018. 3

[74] Yulun Zhang, Kai Zhang, Zheng Chen, Yawei Li, Radu Timofte, et al. NTIRE 2023 challenge on image super-resolution (x4): Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[75] 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. 3

[76] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 39(6), 2019. 12

[77] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 8