NTIRE 2024 Dense and Non-Homogeneous Dehazing Challenge Report

Codruta O. Ancuti Cosmin Ancuti Florin-Alexandru Vasluianu Radu Timofte Yidi Liu Xingbo Wang Yurui Zhu Gege Shi Xin Lu Xueyang Fu Zheng-jun Zha Wei Dong Han Zhou Ruiyi Wang Xiaohong Liu Guangtao Zhai Jun Chen Wei Song Yichang Gao Jiahao Xiong Hualiang Lin Xianger Li Dong Li Mohab Kishawy Ruibin Li Seyed Amirreza Mousavi Rana Rauf Yangyi Liu Huan Liu MingSheng Tu Kele Xu Jiawen Chen Qisheng Xu Tao Sun Jin Guo Ben Shao Tianli Liu Mohao Wu Xingzhuo Yan Minghan Fu Xiaozhong Luo Han Zhong Jiali Wei Chen Meng Lei Xu Bilel Benjdira Anas M. Ali Anis Koubaa Lehan Yang Xin Lin Lu Qi Jincen Song Xiaoqian Hu Linwei Tao Hongming Chen Xiang Chen Chuanlong Xie Zhao Zhang Junhu Wang Yanyan Wei Suiyi Zhao Shengeng Tang Sampada Malagi Amogh Joshi Nikhil Akalwadi Chaitra Desai Ramesh Ashok Tabib Uma Mudenagudi Wenjing Jiang Jagadeesh Kalyanshetti Vijayalaxmi Ashok Aralikatti Yashaswini P Nitish Upasi Dikshit Hegde Ujwala Patil Sujata C

Abstract

This study examines the results of the NTIRE 2024 Challenge on Dense and Non-Homogeneous Dehazing. Innovative methods were introduced and tested using a new image dataset named DNH-HAZE. The DNH-HAZE dataset comprises 50 pairs of authentic outdoor images showcasing dense and non-homogeneous haze alongside corresponding haze-free images of identical scenes. The haze was simulated using a professional setup designed to mirror real-world hazy conditions. The competition attracted 374 participants, with 16 teams presenting solutions for the final evaluation phase. The proposed solutions showed the leading edge of image dehazing technology.

1. Introduction

Haze is a naturally occurring phenomenon that can greatly reduce visibility in a scene with increasing distance, leading to reduced image quality. This atmospheric effect is triggered by the presence of small airborne particles that affect environmental properties. As a result, hazy scenes often exhibit low contrast, reduced saturation, altered colors, and increased noise.

The restoration of visual information from hazy images is crucial for various applications including aerial or ground surveillance, automatic traffic control, and autonomous driving. Consequently, image dehazing has garnered considerable interest over the past decade.

In recent years, there has been a significant surge of interest in image dehazing [1, 7, 9, 10, 15, 29, 32, 34, 37, 48, 49], driven by the necessity of restoring visual information from hazy images for diverse applications such as aerial or ground surveillance, automatic traffic control, and autonomous driving. More recently, deep learning architectures have been employed to address image dehazing [18, 38, 45, 51, 64].

A primary obstacle in objectively verifying and classifying dehazing algorithms stems from the absence of standardized test benchmarks. Assessing dehazing techniques is often complex due to the lack of a reference image or ground truth, as well as the absence of standardized algorithms for error detection and measurement. Despite the development of blind evaluation algorithms, their inconsistent results may be attributed to insufficient validation on real images.

Maintaining consistent lighting conditions and achieving pixel-by-pixel correspondence between reference and hazy images are critical challenges in gathering such am-

*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/2024/

biguous images. Consequently, the initial image dehazing datasets (e.g., D-HAZY [6]) were synthesized using information about scene depth and scene attenuation parameters.

However, a more effective approach involves capturing outdoor haze-free images first and then photographing the same scene with haze introduced using specialized equipment. The initial realistic image dehazing datasets were unveiled during the NTIRE 2018 image dehazing challenge [2]. O-HAZE [8] comprises 45 pairs of outdoor images, while I-HAZE [4] consists of 35 pairs of indoor images. The hazy scenes in O-HAZE and I-HAZE datasets are characterized by light and homogeneous haze. Similarly, DENSE-HAZE [3] includes dense (homogeneous) hazy images along with corresponding ground-truth images and was utilized in the NTIRE 2019 image dehazing challenge [16]. Conversely, the initial realistic non-homogeneous image dehazing datasets (NH-HAZE [11]) were employed for the NTIRE 2020 [12] and 2021 [13] image dehazing challenges.

The NTIRE 2024 image dehazing challenge marks progress in benchmarking single image dehazing techniques. It leverages the DNH-HAZE dataset, which comprises 50 hazy images (with dense and non-homogeneouse haze) along with their corresponding ground truth (haze-free) images depicting the same scenes. DNH-HAZE features real outdoor scenes with non-homogeneous haze produced using a professional haze setup. In the NTIRE 2024 image dehazing challenge, we conduct an objective evaluation by comparing the outcomes of competing methods against the reference images from the DNH-HAZE dataset.

This challenge is one of the NTIRE 2024 Workshop [1] associated challenges on: dense and non-homogeneous dehazing [5], night photography rendering [17], blind compressed image enhancement [59], shadow removal [50], efficient super resolution [44], image super resolution ($\times 4$) [21], light field image super-resolution [55], stereo image super-resolution [52], HR depth from images of specular and transparent surfaces [61], bracketing image restoration and enhancement [68], portrait quality assessment [20], quality assessment for AI-generated content [39], restore any image model (RAIM) in the wild [36], RAW image super-resolution [22], short-form UGC video quality assessment [35], low light enhancement [40], and RAW burst alignment and ISP challenge.

2. Image Dehazing Challenge

The objectives of the NTIRE 2024 challenge on dense and non-homogeneous 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 dense and non-homogeneous high resolution image dehazing (DNH-HAZE) dataset.

2.1. Dense and Non-homogeneous (DNH-HAZE) image dataset

The NTIRE 2024 image dehazing challenge was built on the extended version of the former NH-Haze [11] and HD-NH-HAZE [14] datasets. The DNH-HAZE dataset comprises 50 hazy images along with their corresponding ground truth (haze-free) images depicting the same scenes. This dataset features authentic outdoor settings with non-homogeneous haze created using professional haze-generating equipment. To introduce haze into the outdoor scenes, we utilized two professional haze machines capable of producing vapor particles with diameters typically ranging from 1 to 10 microns, similar to atmospheric haze particles. For image capture, we employed remotely controlled Sony A7 III cameras.

To ensure consistency in unaffected areas of haze across the image pairs, we manually adjusted and maintained constant camera parameters, including shutter speed, aperture (F-stop), ISO, and white balance, throughout consecutive recording sessions. Camera settings such as aperture, exposure, and ISO were determined using an external exposure meter (Sekonic), while white balance was calibrated using the medium gray card (18% gray) from the color checker.

The process of capturing each pair of images typically required approximately 20 to 30 minutes to complete.

2.2. Evaluation

For the NTIRE 2024 dehazing challenge, we hosted a Codalab competition. Participants were required to register on the Codalab platform to access the data and submit their produced results for evaluation according to the specified phases.

Quantitative measures for evaluation included the Peak Signal-to-Noise Ratio (PSNR) in decibels and the Structural Similarity Index (SSIM) computed between the inferred result and the ground truth image. Higher scores indicate better restoration fidelity to the ground truth image. Additionally, the LPIPS perceptual measure was used to assess the quality of the produced results.

The final ranking incorporated a Mean Opinion Score (MOS), derived from a user study organized by the challenge organizers, which considered feedback provided by teams during the final phase of the challenge.

2.3. Challenge Phases

1. **Development phase:** In this phase, the first 40 hazy images of the DNH-HAZE dataset were made public on the challenge platform [43], for the participants to use them in the development process of their solution.

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

2. **Validation phase:** Another set consisting of 5 images was made public to the participants. Using the validation server [43], 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 [43], they uploaded their predicted haze-free images for evaluation, thus being ranked in terms of PSNR and SSIM [56]. 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 NTIRE 2024 Dense and Non-Homogeneous Dehazing Challenge attracted 374 registered participants, with 16 teams submitting their results, solution descriptions, available codes, and team descriptions for the final phase of the challenge. The solutions from these teams are ranked in Table 1. Participants proposed innovative solutions that exhibited a high level of performance in terms of both reconstruction fidelity and perceptual properties.

As shown in Table 1, the metric most closely correlated with the user study results, quantified by the Mean Opinion Score (MOS), is the Peak Signal-to-Noise Ratio (PSNR). Perceptual property-based metrics such as LPIPS [65] and SSIM [56] were utilized to distinguish between similar results. Notably, the top-performing solutions based on perceptual metrics demonstrated consistent results in both reconstruction fidelity and perceptual properties, as anticipated.

4. Challenge Methods

4.1. USTC-Dehazers

The proposed method is based on a dual-branch model structure from previous years as the overall framework [41]. First, for the transfer learning branch, Flash InternImage [58] was employed,which incorporates Deformable Convolution v4 (DCNv4), demonstrates superior and more rapid long-range modeling capabilities, along with adaptive spatial aggregation. This improved speed and efficiency substantially enhance the network's dehazing capabilities.

Considering the high resolution (6000*4000) of the data for this challenge, for the fine-detail extraction branch, a lightweight model, Spatially-Adaptive Feature Modulation (SAFMN) [47] is utilized. This decision is motivated by SAFMN's superiority in feature fusion. By introducing selective attention mechanisms, features from different levels are dynamically fused, enhancing the model's perception of crucial information.

Another challenge of this task is posed by the scarcity of data samples, which often results in the model encountering overfitting issues. Although the problem is somewhat mitigated by the introduction of the two-branch architecture, it still constrains the further improvement of model performance. The method proposed in [57] is employed to introduce synthetic haze data, and a dynamic data enhancement strategy is proposed to control the ratio of synthetic data to real data. The above strategy effectively alleviates the dilemma of having few training samples.

Compared to the traditional VGG perceptual loss, EfficientVit-SAM [67] is introduced as a feature extractor to construct a novel enhanced perceptual loss. This loss reduces the output haze residue to a greater extent.

The overall architecture is depicted in Fig.1.

4.2. Dehazing_R

This team introduces DehazeDCT [28], which consists of two modules and the structure diagram is shown as Fig. 2.

The first part of the model is the **Dehazing module**, which aims to learn the color and texture mapping from hazy to clean image. This module has two branches, where the main branch is our proposed transformer-like architecture based on deformable convolution v4 [58]. This design has several advantages: (1) The convolution window of deformable convolution is smaller than large dense kernels such as 31×31, which significantly reduces the computational cost. (2) Compared to the fixed kernels applied in common CNNs, the deformable convolution dynamically learns flexible receptive field (long- or short-range) from training data. (3) The unnecessary softmax normalization common in traditional deformable convolution [53] is removed and significantly accelerates the forward speed. Besides, the frequency branch similar to [27, 69] is also adopted. The loss function utilized for optimizing of the dehaizing module is shown as Eq. 1:

$$L_{loss} = L_1 + \alpha L_{SSIM} + \beta L_{Percep} + \gamma L_{adv}, \quad (1)$$

where L_1, L_{SSIM} and L_{Percep} represent $L1$ loss, MS-SSIM loss, and perceptual loss, respectively. In addition, the authors adopt the discriminator in [70] to calculate adversarial loss (L_{adv}) for GAN training. α, β, and γ are hyper-parameters and are set to 0.4, 0.01, and 0.0005, respectively.

The second part of the proposed model is the **Refinement module**. The motivation of introducing the refinement module is to further recover the color and texture details and output realistic and visually pleasing results. In

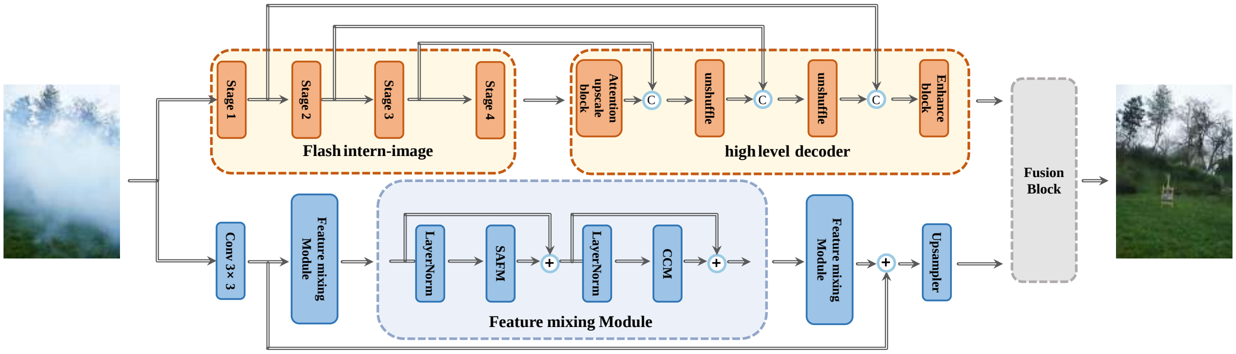

Figure 1. The network architecture of the solution proposed by team USTC-Dehazers.

Rank	Team	Username	PSNR↑	SSIM↑	LPIPS↓	MOS↑	Params.(M)	Runtime(s)	Device	Extra data
1	USTC-Dehazers	LYD	$22.943_{(1)}$	$0.729_{(1)}$	$0.352_{(7)}$	$6.315_{(1)}$	150.16	19	RTX4090	NTIRE '20, '21, '23
2	Dehazing_R	ZXCV, ylxb	$22.84_{(2)}$	$0.725_{(3)}$	$0.346_{(6)}$	$5.96_{(2)}$	256	3.42	A100	no
3	Team Woof	Studentns	$22.59_{(3)}$	$0.726_{(2)}$	$0.380_{(8)}$	$5.79_{(4)}$	521	12	RTX4040	NTIRE '20, '21, '23
4	ITB Dehaze	monday21, baixike, bkdlrb	$22.323_{(4)}$	$0.714_{(4)}$	$0.333_{(4)}$	$5.70_{(5)}$	110	9	2xTITAN RTX	NTIRE '20, '21, '23
5	TTWT	TTWT	$21.932_{(5)}$	$0.714_{(5)}$	$0.334_{(5)}$	$5.675_{(6)}$	n/a	n/a	n/a	n/a
6	DH-AISP	DH-AISP	$21.902_{(6)}$	$0.714_{(6)}$	$0.401_{(9)}$	$5.81_{(3)}$	1418	n/a	RTX 4080	n/a
7	BU-Dehaze	Xingzhuo Yan	$21.676_{(8)}$	$0.709_{(7)}$	$0.326_{(2)}$	$5.22_{(9)}$	13.4	9	6xGTX1080Ti	NTIRE '20, '21
8	RepD	LuoXz	$21.782_{(7)}$	$0.706_{(8)}$	$0.322_{(3)}$	$4.83_{(10)}$	41.32	10.62	RTX3090	no
9	PSU Team	Bilel Benjdira	$20.538_{(12)}$	$0.632_{(15)}$	$0.267_{(1)}$	$5.31_{(7)}$	161	100	A100	NTIRE '20, '21
10	xsourse	xsourse	$21.658_{(9)}$	$0.695_{(11)}$	$0.449_{(11)}$	$5.28_{(8)}$	2.387	0.98	A6000	NTIRE '21, '23, NH-HAZE
11	Pixel_warrior	chm	$20.747_{(11)}$	$0.701_{(9)}$	$0.413_{(10)}$	$3.39_{(12)}$	16	119	RTX3090	no
12	LVGroup_HFUT	wjh0610	$20.750_{(10)}$	$0.698_{(10)}$	$0.503_{(12)}$	$3.57_{(11)}$	5	0.5	A40	NH-Haze20, NH-Haze21
13	KLETech-CEVI_Lowlight_Hypnotise	Lowlight Hypnotise	$17.978_{(14)}$	$0.646_{(12)}$	$0.521_{(13)}$	$1.6_{(16)}$	20.5	0.9	RTX3090	no
14	Towards Sunshine	melody, jwj	$19.493_{(13)}$	$0.635_{(14)}$	$0.632_{(17)}$	$2.83_{(14)}$	60.57	0.41	A6000	D-HAZY
15	KLETech-CEVI_Dark_Knights	SampadaMalagi	$17.978_{(15)}$	$0.646_{(13)}$	$0.521_{(14)}$	$1.5_{(17)}$	14.5	1.8	RTX3090	no
16	HistoMask	achintya971251, lohith kandukuri	$14.475_{(16)}$	$0.494_{(17)}$	$0.628_{(16)}$	$1.7_{(15)}$	1.71	66.4	A40	no

Table 1. Quantitative results of the NTIRE 2024 Dense and 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.

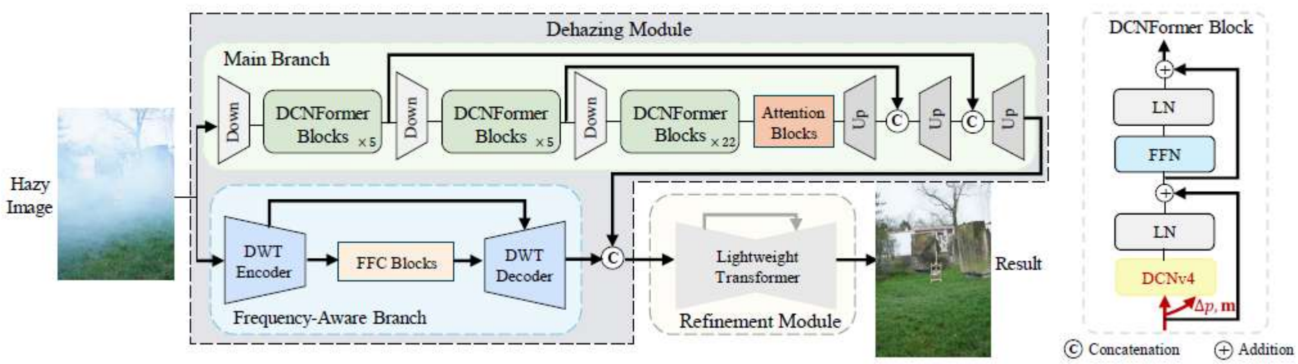

Figure 2. The network architecture of the solution proposed by team Dehazing_R.

this part, the authors adopt a lightweight transformer-based enhancement network based on Retinex theory [19]. This module only has a few parameters and is optimized using Eq. 2.

$$L_{loss} = L_1 + 0.4 * L_{SSIM} + 0.01 * L_{Percep}. \quad (2)$$

4.3. Team Woof

This work is based on ITB-Dehaze [41] and DWT-FFC-GAN [69] approaches. Previous approaches usually employ in the end a single convolution layer and Tanh function. However, this approach fails to achieve optimal imaging results because a single convolution layer is insufficient to effectively represent different image scales. To address this issue, this method introduces a multi-scale Attention head.

The decoder outputs features represented by Y, a tensor with dimensions $B \times C \times H \times W$. In this representation, B represents batchsize, C represents feature channels, and H and W denote the height and width of the input image, respectively. The multi-scale Attention head produces four feature sets, denoted as $[X1, X2, X3, X4]$, utilizing convolution kernels of varied sizes $[1, 3, 5, 7]$. These sets are subsequently fused through a modified SKFusion [46] process to create the aggregated feature X_{SK}. This modification involves replacing the original Squeeze and Excitation mechanism with the Efficient Channel Attention mechanism for enhanced feature integration, that is, Improved Selective Kernel Fusion in Fig. 3. Finally, the four feature sets $[X1, X2, X3, X4]$ and X_{SK} are concatenated to form the final feature set, which is then processed through a 7×7 convolution with padding of 3 and activated by a Tanh function.

4.4. ITB Dehaze

The authors propose a network with two branches(Fig. 4): one utilizing a pre-trained Swin Transformer V2 model [42] for feature extraction, and the other focusing on learning from NH-Haze datasets using lightweight data fitting with RCAN. The fusion of results from both branches generates haze-free output images, employing attention modules and skip-connections in a multi-phase fine-tuning strategy. Initially, the best model weights trained on NH-20, NH-21, and NH-23 was fine-tuned on the NH-24 dataset, resulting in performance enhancement. Further refinement was conducted on the top-performing models from this stage using additional datasets. Careful observation of results was necessary to balance issues of low-resolution and haze. Interestingly, increasing the number of training epochs did not consistently improve results. The focus remained on achieving thorough dehazing within limited time constraints, prioritizing overall effectiveness over high-resolution output. Also, the authors adopted Semi-Supervised Self-Supervised method [63] to overcome the shortage of denser hazy images for this year. This was used to generate pseudo labels for the unlabeled data and train the model on it again using a low learning rate and regularized setting for the model. We utilize the optimal configurations from the basic supervised baseline and adjust only the learning rate, weight decay, and weight of the newly introduced loss. Utilizing such a method introduces the availability of leveraging much more unlabeled data for training and improvement.

4.5. TTWT

The method is inspired from the work of [41]. Compared to the method proposed in [41], three main modifications are made:

1. A customized loss function:

$$L = L_{re} + \gamma_A * L_s + \gamma_B * L_{vgg} + \gamma_C * L_{smi} + \gamma_D * L_{adv} \quad (3)$$

where L_{re} is the smooth L_1 loss, L_s is used to maximize the value of SSIM, which is refereed as Multi-Scale Structure Similarity (MS-SSIM). where L_{vgg} represents the VGG loss, and L_{smi}, L_{adv} represent the similarity loss and the discriminator's loss respectively.

2. The authors explored multi-channel feature extraction. The advantage of a multi-channel feature extraction network lies in its ability to extract more comprehensive information from different channels and scales, which helps enhance the effectiveness of image dehazing.

3. The authors initialized an AdamW optimizer to optimize model parameters, with specific settings for the learning rate (0.00004), weight decay (1e-8), and beta parameters (0.9, 0.999). Additionally, they set up a cosine learning rate scheduler, gradually reducing the learning rate over the specified number of epochs (num_epochs) according to the cosine annealing schedule until reaching the minimum learning rate of 2.5e-6. This ensures more stable and effective training of deep learning models.

4.6. DH-AISP

The method is based on a detailed dual-branch model to process low-frequency and high-frequency images separately, so that the color and details of the image can be better processed. The network structure is shown in Fig. 5. For the global branch, the authors propose learning the mapping relationship between the original image and the hazy image at a low resolution, which has a larger receptive field compared to the original resolution and can recover information damaged by the haze through surrounding regions. Then, the results is sent to a global branch together with the input into the second branch, and learn the details differences caused by low-resolution images based on the global color and contours. Finally, the results of the two branches are blended to obtain the final result.

4.7. BU-Dehaze

The authors introduce a method that utilizes a two branch structure [60] for single image dehazing, as shown in Fig.6. Inspired by [30], the authors further develop a data preprocessing technique aimed at aligning the distribution of augmented data more closely with that o the target data. In the first branch of the model, the Swin Transformer V2 model [42] and self-attention mechanism, which has been pre-trained on the ImageNet dataset, is employed to extract pertinent multi-level features from hazy images. The choice of the Swin Transformer, proven to surpass traditional CNN-based architectures in performance, serves as the backbone for our feature extraction process. Embracing the concept of transfer learning, the authors initialize the Swin Transformer with the ImageNet pre-trained model, thereby enabling our system to leverage knowledge acquired from prior low-level tasks. In contrast, the second branch, built from scratch, employs the U-Net architecture, which complements the first by focusing exclusively

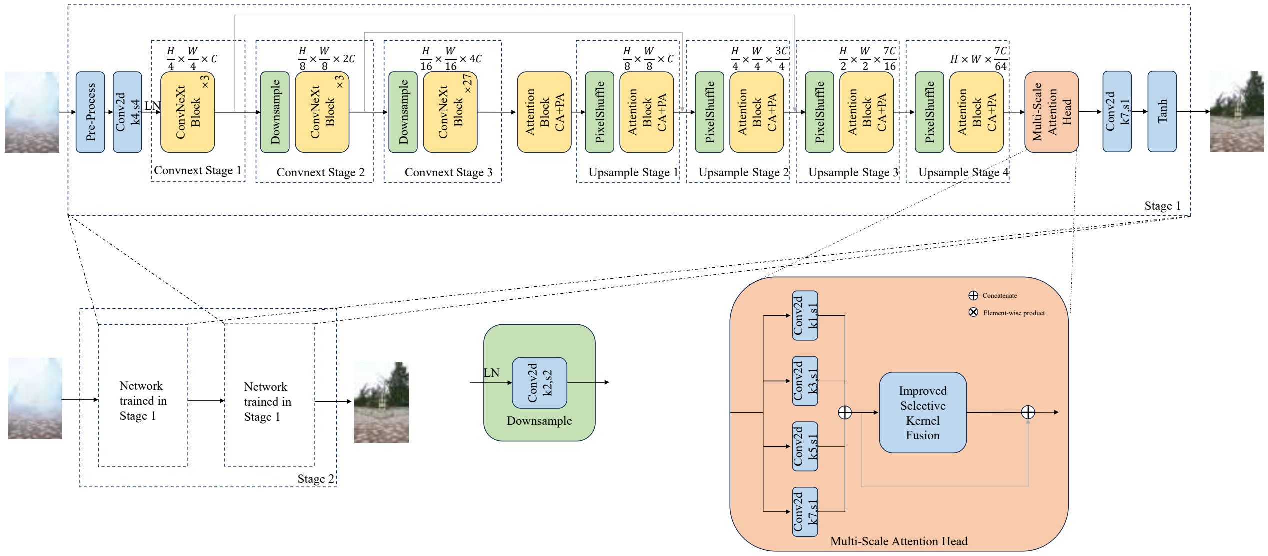

Figure 3. The network architecture of the solution proposed by team Team Woof.

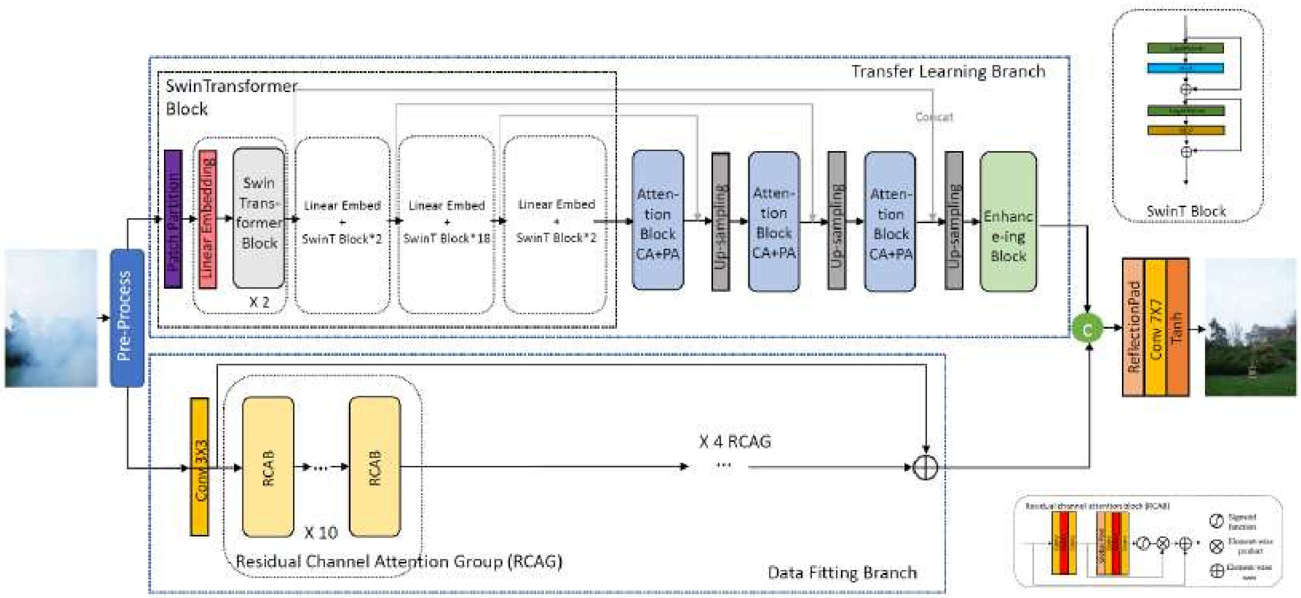

Figure 4. The network architecture of the solution proposed by team ITB Dehaze.

on the domain of target data. This synergistic approach allows for the specialized treatment of target data, enhancing the model's overall effectiveness in dehazing. The fusion tail of our model aggregates outputs from both branches, culminating in the production of clear, dehazed images.

4.8. RepD

The dehazing model named RepD (see Fig. 7) is built by combining a UniRepLKNet [26] subnet and an [66] RCAN subnet. Specifically, the authors utilized the UniRepLKNet subnet, where the concept of structural reparameterization is introduced to enhance feature extraction for dehazed images and improve inference speed. RCAN is employed as

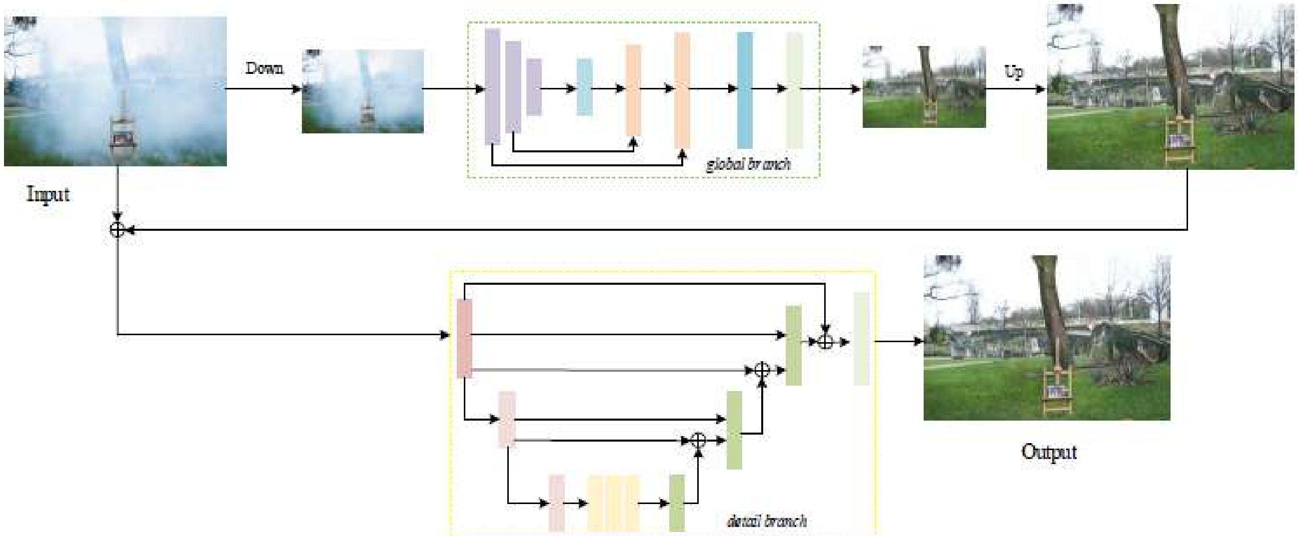

Figure 5. The network architecture of the solution proposed by team DH-AISP.

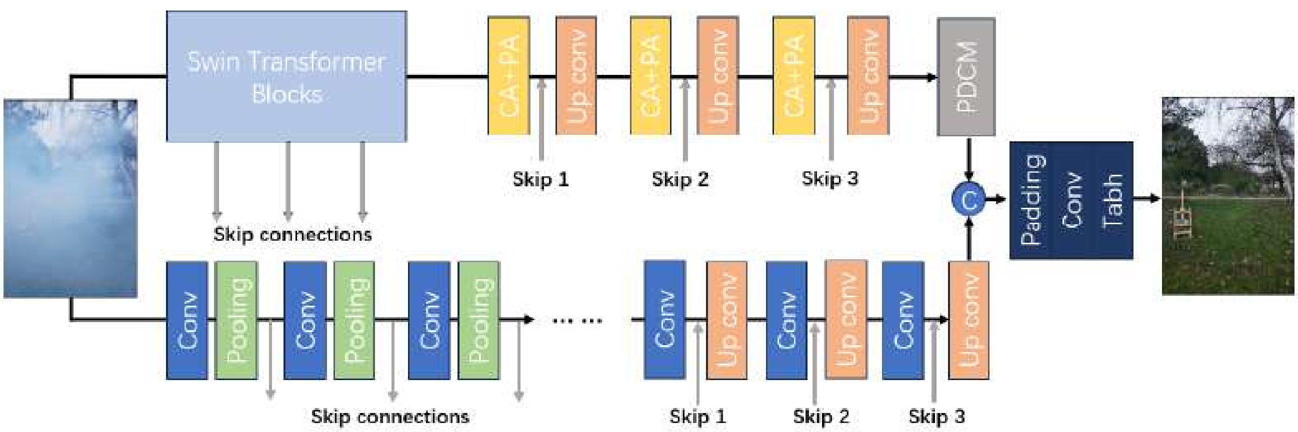

Figure 6. The network architecture of the solution proposed by team BU-Dehaze.

the other subnet to preserve fine details in the recovered images without downsampling or upsampling in resolution.

4.9. PSU Team

The authors developed a novel dual-phase architecture, named High-Resolution Image Dehazing and Enhancement (HRDE), designed to tackle this issue comprehensively (see Fig. 8). The first phase of HRDE employs a latent diffusion model, specifically tailored for extracting haze from high-resolution images, serving as the cornerstone for haze removal. However, diffusion models often suffer from distortion and noise exacerbation due to extended training periods and interactions with latent spaces. To address these drawbacks, the subsequent phase of HRDE utilizes a Laplacian Generative Adversarial Network (GAN), engineered to enhance image quality and correct distortions introduced in the first phase. This multi-layer Laplacian GAN approach leverages Laplace pyramids to exploit multi-scale image representations, thereby effectively preserving and enhancing structural details. This strategy ensures a better recovery of fine details without introducing noise or artifacts.

4.10. xsourse

The proposed method is based on method introduced in [31]. Due to the instability of GAN network training, the authors added a standard deviation loss with a coefficient of 0.01 to the original implementation of multiple losses including L1 loss, SSIM loss, etc. The introduction of standard deviation loss can effectively supervise the color distribution, making the GAN training process more stable and

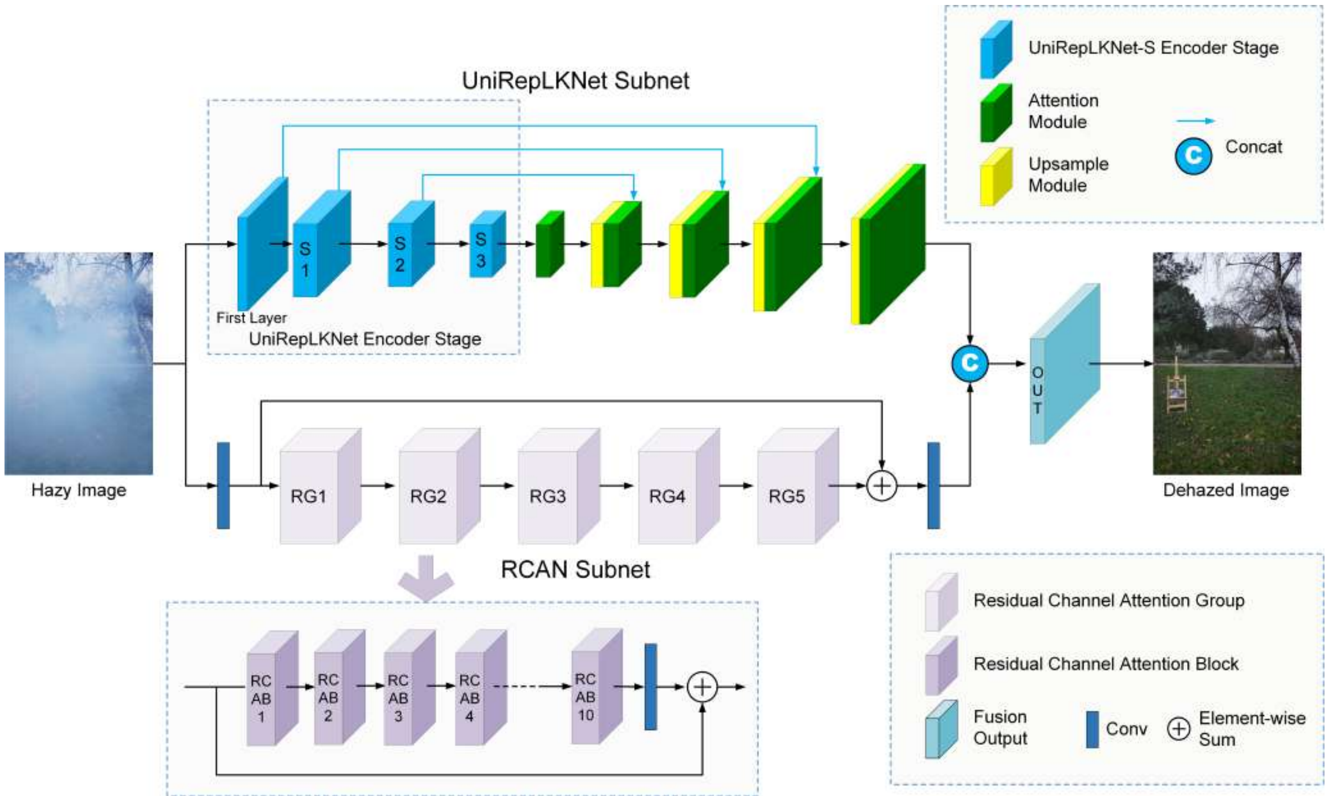

Figure 7. The network architecture of the solution proposed by team RepD.

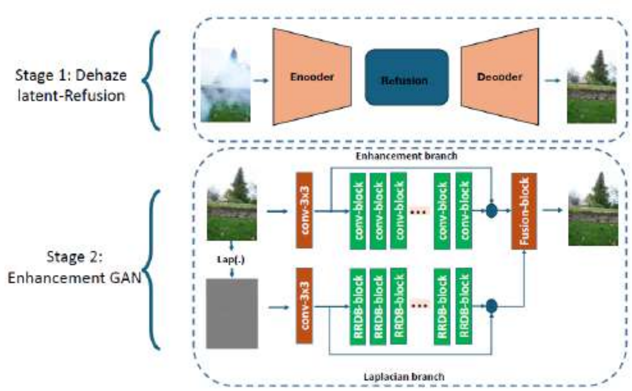

Figure 8. The network architecture of the solution proposed by team PSU Team.

converging faster. At the same time, the supervision of standard deviation on pixels can make the restored picture more vivid in color and closer to the original ground truth.

4.11. Pixel warrior

The team proposes a Transformer for Image Dehazing (TID) as shown in Fig 9. The transformer module includes both global feature extraction and local feature extraction modules. In response to the problem of insufficiently clear features in local regions, the team has designed a global feature extraction module to capture overall information. At the same time, they have also designed a local feature extraction module to obtain more refined local features.

4.12. LVGroup HFUT

The proposed architecture is strategically based on the U-Net framework. To equip the model with a multi-dimensional perspective, the team generates outputs from foggy images across various sizes. This approach not only enriches the model's input data but also ensures that the decoded outputs closely mirror the grundtruth of the images by presenting restored results in multiple sizes. Regarding the loss function, the authors meticulously compute both L1 and FFT losses for outputs at each size.

The proposed model undergoes a significant enhancement in the encoding and decoding phases. They replace conventional residual blocks with the more sophisticated Residual Dense Block (RDB) modules [54], substantially improving the model's capacity for nuanced feature extraction. Furthermore, to foster a richer interplay of features across different scales, the authors integrate the SKFF module [62], thereby facilitating a more effective synthesis of information from various scales.

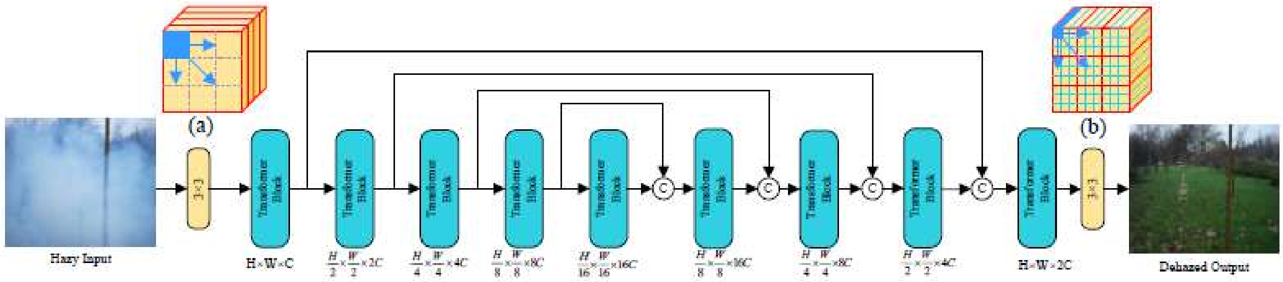

Figure 9. The network architecture of the solution proposed by team Pixel warrior.

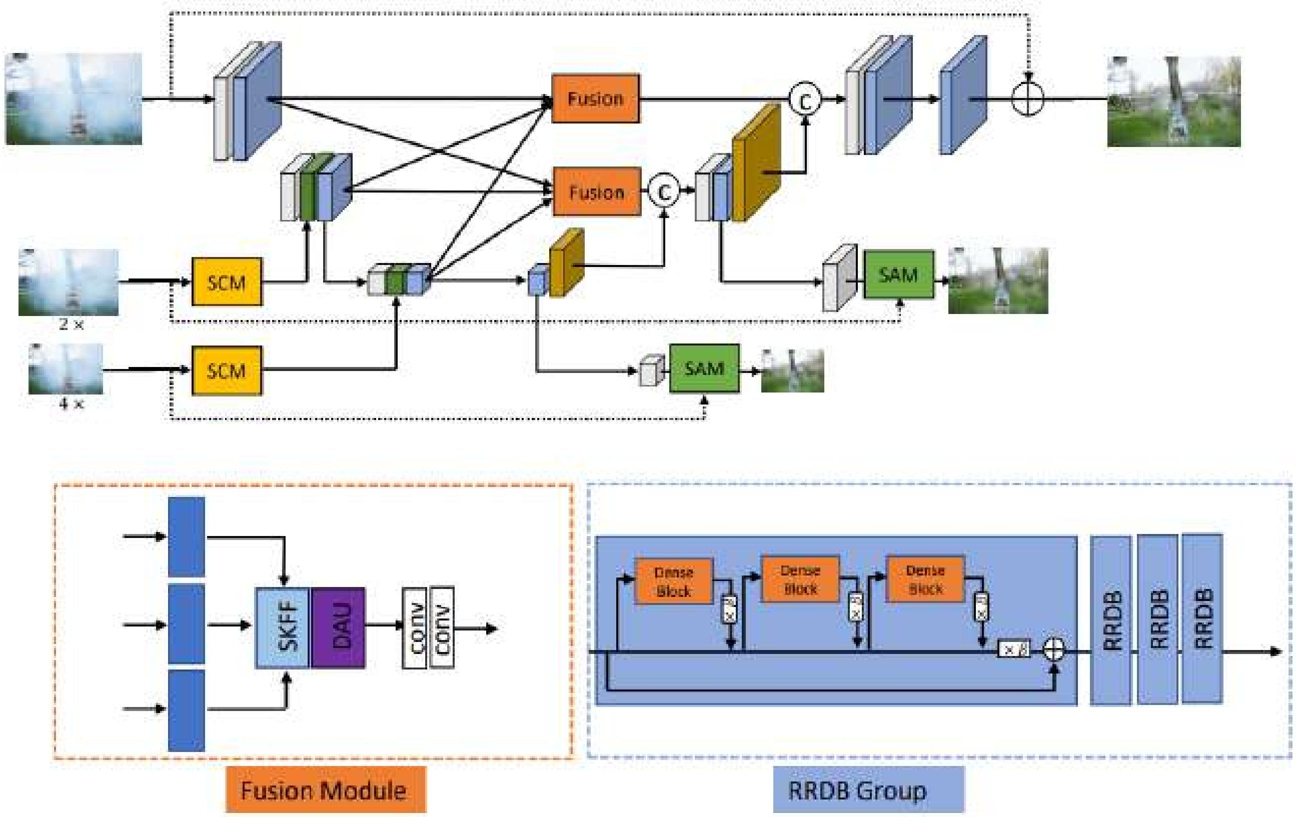

Figure 10. The network architecture of the solution proposed by team LVGroup HFUT.

4.13. KLETech-CEVI Lowlight Hypnotise

The proposed MFNN framework includes three main modules: the hierarchical spatio-contextual (HSC) feature encoder, Global-Local Spatio-Contextual (GLSC) block, and hierarchical spatio-contextual (HSC) decoder, as shown in Fig. 11. Typically, image restoration/enhancement networks employ feature scaling for varying the sizes of the receptive fields. The varying receptive fields facilitate learning of local-to-global variances in the features. With this motivation, this method learns contextual information from multi-scale features while preserving high resolution spatial details. This is achieved via a hierarchical style encoder-decoder network with residual blocks as the backbone for learning. The proposed MFNN optimize the learning of MFNN with the proposed $\mathcal{L}_{MFNN}$ and is given as,

$$\mathcal{L}_{MFNN} = (\alpha * L_1) + (\beta * \mathcal{L}_{VGG}) + (\gamma * \mathcal{L}_{MSSSIM}) \quad (4)$$

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 [23–25, 33].

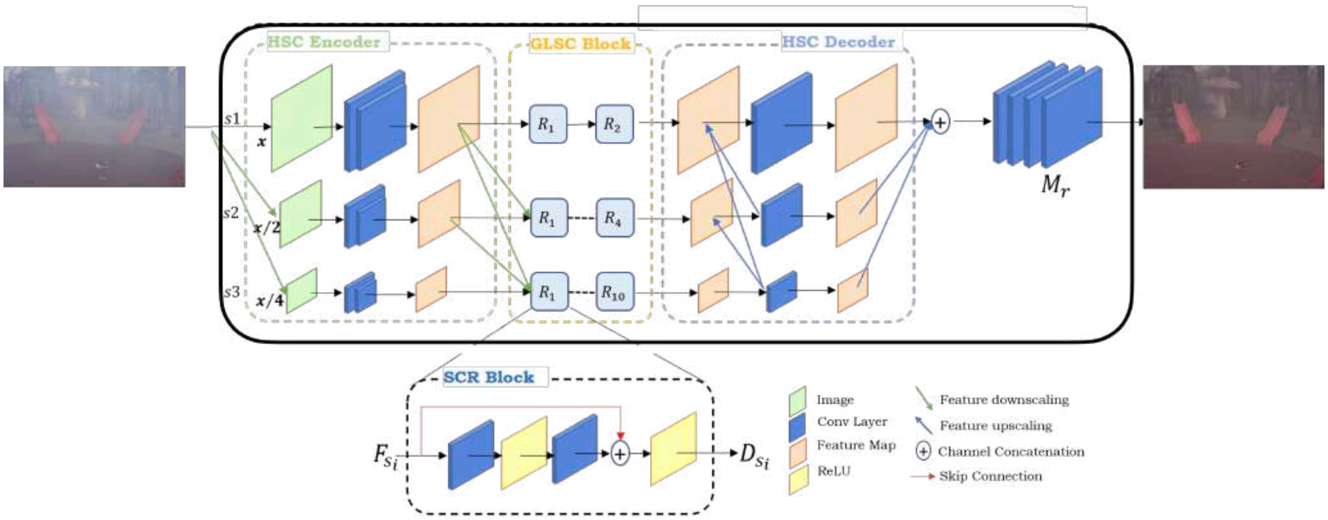

Figure 11. The network architecture of the solution proposed by team KLETech-CEVI Lowlight Hypnotise.

4.14. Towards Sunshine

The authors introduced a dehazing network based on self-classification guidance. Specifically, the network consists of two parts, the dehazing subnetwork and the classification subnetwork. As shown in Fig. 12, in the training process, the dehazing sub-network is first trained, and after it has good dehazing ability, the clean image, foggy image and dehazing output image of the training data are used to perform the three-classification task. The total loss of the network is:

$$Loss_{total} = loss_{dehazing} + \lambda * loss_{cls} \tag{5}$$

where $loss_{dehazing}$ is the loss of the dehazing subnetwork. $loss_{cls}$ is the loss of the classification subnetwork, and λ is the weight parameter which was setting to 0.05. The co-evolution of the two networks and the backpropagation process of the classification sub-network can promote the improvement of the performance of the dehazing sub-network. The classification subnetwork selected in this competition is the Alexnet network pre-trained on the ImageNet dataset , and the classification subnetwork can be matched with any dehazing subnetwork, and the dehazing subnetwork selected in this competition is an end-to-end feature fusion attention network.

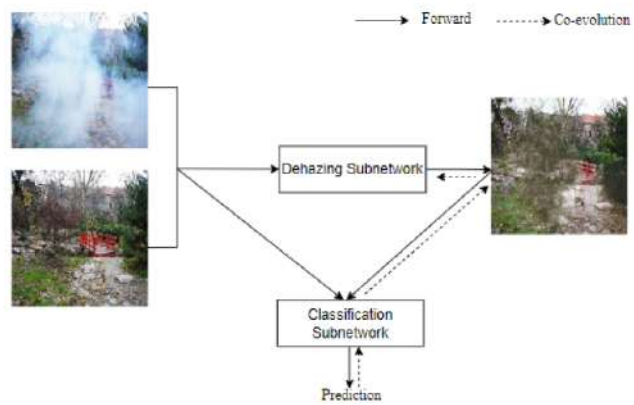

Figure 12. The network architecture of the solution proposed by team Towards Sunshine.

4.15. KLETech-CEVI Dark Knights

The proposed MFNN framework includes three main modules: the hierarchical spatio-contextual (HSC) feature encoder, Global-Local Spatio-Contextual (GLSC) block, and hierarchical spatio-contextual (HSC) decoder, as shown in Fig. 11. Typically, image restoration/enhancement networks employ feature scaling for varying the sizes of the receptive fields. The varying receptive fields facilitate learning of local-to-global variances in the features. With this motivation, this method learns contextual information from multi-scale features while preserving high resolution spatial details. This is achieved via a hierarchical style encoder-decoder network with residual blocks as the backbone for learning. We optimize the learning of MFNN with the proposed $\mathcal{L}_{MFNN}$ and is given as,

$$\mathcal{L}_{MFNN} = (\alpha * \mathcal{L}_{Charbonnier}) + (\beta * \mathcal{L}_{FFTLoss}) \tag{6}$$

where, α, and β, are the weights. We experimentally set the weights to $\alpha = 0.5$, and $\beta = 0.7$, $\mathcal{L}_{MFNN}$ is a weighted combination of three distinct losses. $L_{Charbonnier}$ loss to minimize error at pixel level, FFTLoss to efficiently restore contextual information between the ground-truth image and the output image. The aim here is to minimize the weighted

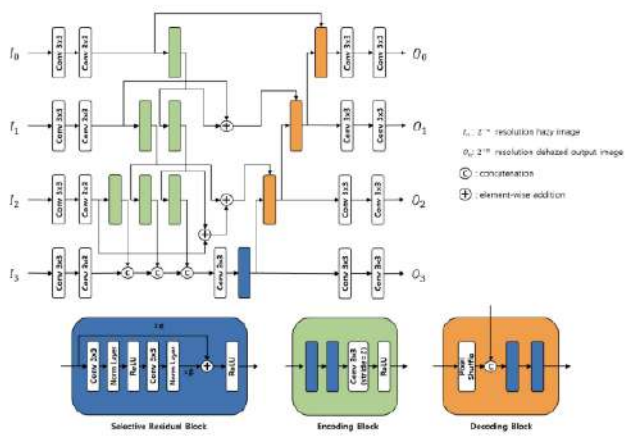

Figure 13. The network architecture of the solution proposed by team HistoMask.

combinational loss $\mathcal{L}_{MFNN}$ given as,

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N} \parallel MFNN(x_i - y_i) \parallel_{\mathcal{L}_{MFNN}} \qquad (7)$$

where, θ denotes the learnable parameters of the proposed framework, N is the total number of training pairs, x and y are the input and output image respectively, and $MFNN(\cdot)$ is our proposed framework for haze removal in images.

4.16. HistoMask

The method (see Fig. 13) contains 2 stages:
MASK : Each image is processed at 4 scales like a typical U-net with encoder and decoder convolution blocks, hence time complexity: $O(n^2k^2)$. The space complexity is $O(n^2)$, since the biggest contribution is storing the 512x512 image and the rest are continually down-sampled versions of the same.
Histogram Matching and Gamma Correction : The global histogram is computed and stored. The histogram of the image is computed taking a time complexity of $O(n^2)$ and the matching is performed needing a complexity of $O(n^2)$. The gamma correction again takes a time complexity of $O(n^2)$. The space complexity is once again $O(n^2)$ for both histogram matching and gamma correction.

5. Conclusion

During the NTIRE 2024 Image Dehazing Challenge, 374 participants engaged in the competition, leading to the final phase where 16 teams were distinguished for their achievements. These teams innovatively employed diverse architectures and methods, surpassing prior benchmarks and leveraging past design concepts as foundational elements, showcasing significant potential for progress.

The final team rankings were determined by the Mean Opinion Score derived from our user study. The accuracy of the image recovery played a pivotal role in these rankings, exhibiting the highest correlation with user feedback on the results presented.

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 2024 sponsors: Meta Reality Labs, OPPO, KuaiShou, Huawei and University of Würzburg (Computer Vision Lab).

A. Teams and Affiliations

NTIRE 2024 Team

Title:
NTIRE 2024 Challenge on Dense and Non-Homogeneous Dehazing
Members:
Codruta O. Ancuti[1] *(codruta.ancuti@gmail.com)*, Cosmin Ancuti[1,2], Florin-Alexandru Vasluianu[3], Radu Timofte[3]
Affiliations:
[1] ETcTI, Universitatea Politehnica Timisoara, Romania
[2] ICTEAM, UCL, Belgium
[3] Computer Vision Lab, University of Wuerzburg, Germany

USTC-Dehazers

Title:
Coarse-to-Fine Hybrid Network for Dense and Nonhomogeneous Dehaze
Members:
Yidi Liu[1] *(liuyidi2023@mail.ustc.edu.cn)*, Xingbo Wang[1], Yurui Zhu[2], Gege Shi[1], Xin Lu[1] , Xueyang Fu[1], Zheng-jun Zha[1]
Affiliations:
[1] Department of Automation, School of Information Science and Technology, University of Science and Technology of China, Hefei, China.
[2] School of Cyber Science and Technology, University of Science and Technology of China, Hefei, China.

Dehazing_R

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

Members:
Wei Dong[1] (wdong1745376@gmail.com), Han Zhou[1], Ruiyi Wang[2], Xiaohong Liu[2], Guangtao Zhai[2], Jun Chen[1]
Affiliations:
[1] McMaster University, Canada
[2] Shanghai Jiao Tong University, China

Team Woof

Title:
Two Stage Training and Multi-Scale Attention Head For Dehazing
Members:
Wei Song (sonwe@mail2.gdut.edu.cn), Yichang Gao, Jiahao Xiong, Hualiang Lin, Xianger Li, Dong Li
Affiliations:
Guangdong University of Technology, China

ITB Dehaze

Title:
A Fine-Tuned Self-supervised End-to-end Solution to NonHomogeneous Dense Dehazing via Transformer.
Members:
Mohab Kishawy (kishawym@mcmaster.ca), Ruibin Li, Seyed Amirreza Mousavi, Rana Rauf, Yangyi Liu, Huan Liu, Jun Chen
Affiliations:
McMaster University

TTWT

Title:
Improved NonLocal Channel Attention for NonHomogeneous Image Dehazing.
Members:
MingSheng Tu[1] (1779380803@qq.com), Kele Xu[2], JiaWen Chen[1], Qisheng Xu[2], Tao Sun[2]
Affiliations:
[1] Data Engineering and Visualization Laboratory of Chongqing University of Posts and Telecommunications, China
[2] National University of Defense Technology, China

DH-AISP

Title:
Members:
Jin Guo(gjguojingj@gmail.com), Ben Shao, Tianli Liu, Mohao Wu
Affiliations:
Zhejiang Dahua Technology Co.,Ltd.

BU-Dehaze

Title:
Transformer-Based Dense and Non-Homogeneous Dehazing Network
Members:
Xingzhuo Yan[1] (ayx1sgh@bosch.com), Minghan Fu[2]
Affiliations:
[1] Bosch Investment Ltd.
[2] University of Saskatchewan

RepD

Title:
Dehazing Method based on UniRepLKNet and RCAN
Members:
Xiaozhong Luo[1,2] (repd2024@outlook.com), Han Zhong[2], Jiali Wei[2], Chen Meng, Lei Xu
Affiliations:
[1] Beijing Police College, Beijing 102202, China
[2] People's Public Security University of China, Beijing 100038, China

PSU Team

Title:

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

xsourse

Title:
Color Aware Attention Network for Ensemble Enhancement Dehazing
Members:
Lehan Yang[1] (lehan.yang@sydney.edu.au), Xin Lin[2], Lu Qi[3], Jincen Song[1], Xiaoqian Hu[4], Linwei Tao[1]
Affiliations:
[1] University of Sydney
[2] Sichuan University
[3] University of California Merced
[4] University of New South Wales

Pixel-warrior

Title:
Transformer for Image Dehazing (TID)
Members:

Hongming Chen[1] *(chenhongming@stu.sau.edu.cn)*, Xiang Chen[2], Chuanlong Xie[1]
Affiliations:
[1] Shenyang Aerospace University
[2] Nanjing University of Science and Technology

LVGroup-HFUT

Title:
DMIMO-Net Dense multi-input multi-output U-net
Members:
Zhao Zhang (cszzhang@gmail.com), Junhu Wang, Yanyan Wei, Suiyi Zhao, Shengeng Tang
Affiliations:
Hefei University of Technology (HFUT)

KLETech-CEVI Lowlight Hypnotise

Title:
Multi-scale Feature FusionNet (MFNN).
Members:
Sampada Malagi[1,3] (sampadamalagi12@gmail.com), Amogh Joshi[1], Nikhil Akalwadi[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

Towards Sunshine

Title:
SCG-FFA: Self-classification guided feature fusion network for image dehazing.
Members:
Wenjing Jiang (wjjiang hbu@163.com.)
Affiliations:
College of Electronic and Information Engineering, Hebei University, Baoding 071002,China

KLETech-CEVI Dark Knights

Title:
Multi-scale Feature FusionNet (MFNN).
Members:
Jagadeesh Kalyanshetti[1,3] (01fe21bcs004@kletech.ac.in),
Vijayalaxmi Ashok Aralikatti[1,3],
Palani Yashaswini[1,2],
Nitish Upasi[1],
Dikshit Hegde[1],
Ujwala Patil[1,2],
Meena S M[1,3]
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

HistoMask

Title:
HistoMask
Members:
Kandukuri Lohith Saradhi(lohith21110257@snuchennai.edu.in), Achintya Lakshmanan
Affiliations:
Shiv Nadar University Chennai

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] 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
[6] C. Ancuti, C. O. Ancuti, and Christophe De Vleeschouwer. D-Hazy: A dataset to evaluate quantitatively dehazing algorithms. *IEEE ICIP*, 2016. 2
[7] 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
[8] 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
[9] 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

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

[11] 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

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

[13] 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

[14] 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

[15] C. O. Ancuti, C. Ancuti, and C. De Vleeschouwer. Effective local airlight estimation for image dehazing. In *In IEEE ICIP*, 2018. 1

[16] 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

[17] 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

[18] 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

[19] Y. 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 Conference on Computer Vision and Pattern Recognition*, 2023. 4

[20] 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

[21] 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

[22] 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. 2

[23] 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. 9

[24] 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. 9

[25] 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. 9

[26] X. Ding, Y. Zhang, Y. Ge, S. Zhao, L. Song, X. Yue, and Y. Shan. Unireplknet: A universal perception large-kernel convnet for audio, video, point cloud, time-series and image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 6

[27] Wei Dong, Han Zhou, Yuqiong Tian, Jingke Sun, Xiaohong Liu, Guangtao Zhai, and Jun Chen. ShadowRefiner: Towards mask-free shadow removal via fast fourier transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 3

[28] Wei Dong, Han Zhou, Ruiyi Wang, Xiaohong Liu, Guangtao Zhai, and Jun Chen. DehazeDCT: Towards effective non-homogeneous dehazing via deformable convolutional transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 3

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

[30] 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. 5

[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. 7

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

[33] 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. 9

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

[35] 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. 2

[36] 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

[37] 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

[38] 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

[39] 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

[40] 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. 2

[41] 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, 4, 5

[42] 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. 5

[43] 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

[44] 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

[45] 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

[46] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 2023. 5

[47] Long Sun, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. Spatially-adaptive feature modulation for efficient image super-resolution. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 13144–13153, 2023. 3

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

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

[50] 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

[51] 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

[52] 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

[53] W. Wang, Jifeng Dai, Zhe Chen, Zhenhang Huang, Zhiqi Li, Xizhou Zhu, Xiaowei Hu, Tong Lu, Lewei Lu, Hongsheng Li, Xiaogang Wang, and Yu Qiao. Internimage: Exploring large-scale vision foundation models with deformable convolutions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 3

[54] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018. 8

[55] 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

[56] 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

[57] Rui-Qi Wu, Zheng-Peng Duan, Chun-Le Guo, Zhi Chai, and Chongyi Li. Ridcp: Revitalizing real image dehazing via high-quality codebook priors. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22282–22291, 2023. 3

[58] Yuwen Xiong, Zhiqi Li, Yuntao Chen, Feng Wang, Xizhou Zhu, Jiapeng Luo, Wenhai Wang, Tong Lu, Hongsheng Li, Yu Qiao, Lewei Lu, Jie Zhou, and Jifeng Dai. Efficient deformable convnets: Rethinking dynamic and sparse operator for vision applications. 2024. 3

[59] Ren Yang, Radu Timofte, et al. NTIRE 2024 challenge on blind enhancement of compressed image: Methods and re-

sults. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 2

[60] 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. 5

[61] 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

[62] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. *arXiv preprint arXiv:2003.06792*, 2020. 8

[63] Xiaohua Zhai, Avital Oliver, Alexander Kolesnikov, and Lucas Beyer. S4l: Self-supervised semi-supervised learning. *arXiv preprint arXiv:1901.09005*, 2019. 5

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

[65] 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

[66] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 286–301, 2018. 6

[67] Zhuoyang Zhang, Han Cai, and Song Han. Efficientvit-sam: Accelerated segment anything model without performance loss. *arXiv preprint arXiv:2402.05008*, 2024. 3

[68] 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

[69] 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, 4

[70] 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. 3