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
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 NTIRE 2018 [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 datasets 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. 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.

2.1. 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 [65], 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-homogeneous 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 upsampling 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 preinent 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. Without 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$^2$SR 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.
Table 1. Quantitative results of the NTIRE 2023 Non Homogeneous Dehazing Challenge. Using naming convention n1(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.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Team</th>
<th>Username</th>
<th>PSNR↑</th>
<th>SSIM↑</th>
<th>LPIPS↓</th>
<th>MOS↑</th>
<th>Params (M)</th>
<th>Runtime(s)</th>
<th>Device</th>
<th>Extra data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DWT-FFC-GAN</td>
<td>zhoub115</td>
<td>22.57(1)</td>
<td>0.71(1)</td>
<td>0.346(1)</td>
<td>9.07(1)</td>
<td>73</td>
<td>23.3</td>
<td>RTX2080 Ti</td>
<td>NTIRE20.21</td>
</tr>
<tr>
<td>2</td>
<td>ITB Dehaze</td>
<td>lillian</td>
<td>22.96(1)</td>
<td>0.71(1)</td>
<td>0.345(1)</td>
<td>7.85(1)</td>
<td>110</td>
<td>9.0</td>
<td>2×TitanXP</td>
<td>NTIRE20.21</td>
</tr>
<tr>
<td>3</td>
<td>[Mask]</td>
<td>Shawoong98</td>
<td>22.18(5)</td>
<td>0.71(5)</td>
<td>0.401(5)</td>
<td>7.12(5)</td>
<td>9.31</td>
<td>0.125</td>
<td>A100</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>NUSRICQ DEHAZING</td>
<td>Yinwei_Wu</td>
<td>21.97(2)</td>
<td>0.69(2)</td>
<td>0.38(2)</td>
<td>7.7(2)</td>
<td>4.25</td>
<td>3.0</td>
<td>4×RTX3090</td>
<td>NH-HAZE, O-HAZE, RAIZE, DENSE-HAZE</td>
</tr>
<tr>
<td>5</td>
<td>PSU TEAM</td>
<td>Anas</td>
<td>22.27(3)</td>
<td>0.7(3)</td>
<td>0.430(10)</td>
<td>7.4(3)</td>
<td>5.0</td>
<td>480</td>
<td>RTX8000</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>NTU607-dehaze</td>
<td>HaoqiangYang</td>
<td>22.11(6)</td>
<td>0.72(6)</td>
<td>0.442(11)</td>
<td>7.1(2)</td>
<td>101.5</td>
<td>52.1</td>
<td>V100</td>
<td>NTIRE18, 19, 20</td>
</tr>
<tr>
<td>7</td>
<td>MIPCer</td>
<td>YuanGao</td>
<td>21.73(7)</td>
<td>0.7(7)</td>
<td>0.406(4)</td>
<td>6.95(6)</td>
<td>2.39</td>
<td>3.79</td>
<td>A100</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>iPAL-LightDehaze</td>
<td>lightdehaze</td>
<td>22.01(10)</td>
<td>0.7(10)</td>
<td>0.384(6)</td>
<td>5.35(10)</td>
<td>2.6</td>
<td>0.72</td>
<td>TitanXP</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>Xsource</td>
<td>Xsource</td>
<td>22.09(8)</td>
<td>0.67(8)</td>
<td>0.556(6)</td>
<td>7.6(5)</td>
<td>na</td>
<td>37.0</td>
<td>A4000</td>
<td>No</td>
</tr>
<tr>
<td>10</td>
<td>Xiaofeng Cong</td>
<td>CongXiaofeng</td>
<td>21.86(9)</td>
<td>0.67(9)</td>
<td>0.492(12)</td>
<td>6.9(9)</td>
<td>24.7</td>
<td>10.0</td>
<td>V100</td>
<td>No</td>
</tr>
<tr>
<td>11</td>
<td>MengFedHome</td>
<td>XuefeiYin</td>
<td>21.08(13)</td>
<td>0.69(13)</td>
<td>0.411(8)</td>
<td>5.15(14)</td>
<td>5</td>
<td>0.005</td>
<td>RTX3060</td>
<td>No</td>
</tr>
<tr>
<td>12</td>
<td>CANT HAZE</td>
<td>HazimElmad</td>
<td>20.95(4)</td>
<td>0.69(4)</td>
<td>0.415(9)</td>
<td>5.15(4)</td>
<td>80</td>
<td>0.15</td>
<td>T4/K80</td>
<td>No</td>
</tr>
<tr>
<td>13</td>
<td>DCBDN</td>
<td>Maofling</td>
<td>21.62(11)</td>
<td>0.69(12)</td>
<td>0.503(13)</td>
<td>5.86(11)</td>
<td>na</td>
<td>14.04</td>
<td>RTX3090</td>
<td>No</td>
</tr>
<tr>
<td>14</td>
<td>IR-SDE</td>
<td>r-sde</td>
<td>20.83(12)</td>
<td>0.61(16)</td>
<td>0.406(7)</td>
<td>6.3(12)</td>
<td>78</td>
<td>5.0</td>
<td>A100</td>
<td>No</td>
</tr>
<tr>
<td>15</td>
<td>LVGroup HFUT</td>
<td>yun.wei</td>
<td>20.97(11)</td>
<td>0.64(14)</td>
<td>0.521(14)</td>
<td>5.62(12)</td>
<td>5.0</td>
<td>0.005</td>
<td>A40</td>
<td>NTIRE20.21</td>
</tr>
<tr>
<td>16</td>
<td>BH-AISP</td>
<td>BH-AISP</td>
<td>20.91(15)</td>
<td>0.62(15)</td>
<td>0.521(15)</td>
<td>4.7(15)</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>no</td>
</tr>
<tr>
<td>17</td>
<td>SVNIT NTNU</td>
<td>AnjaliSarvaiya</td>
<td>16.82(17)</td>
<td>0.47(17)</td>
<td>0.415(17)</td>
<td>0.72(17)</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>no</td>
</tr>
</tbody>
</table>

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.
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:

$$L_{sem} = |S(I) - S(J)|$$  \hspace{1cm} (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:

$$L_{sc} = \sum_{i=1}^{n} |C_{I,i} - C_{J,i}|$$  \hspace{1cm} (2)
where $C_{l,i}$ means the average color tone of the $i$th class of $I$. Specifically, $C_{l,i}$ can be written as:

$$C_{l,i} = \frac{1}{n} \sum_{S(x) \in i} I(x)$$  \hspace{1cm} (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:

$$L_{Total} = L_{cha} + 1.1L_{W-SSIM} + 0.1L_{Per} + 0.1L_{sem} + 0.1L_{sc}$$  \hspace{1cm} (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.
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 $\lambda$ be the trade-off parameter, $M$ can be expressed mathematically as

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

In particular, the trade-off parameter can be dynamically adjusted through the smooth L1 loss $L_{s11}^\alpha$ of the attention map, i.e.,

$$\lambda = \begin{cases} 
0, & \text{if } L_{s11}^\alpha > 0.1, \\
\frac{L_{s11}^\alpha - 0.1}{0.1 - 0.05}, & \text{if } 0.1 \geq L_{s11}^\alpha > 0.05, \\
1, & \text{if } L_{s11}^\alpha \leq 0.05.
\end{cases} \quad (6)$$

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 $L_{s11}^\alpha$. As $L_{s11}^\alpha$ decreases, the proportion of the attention map $M_g$ generated by the network will continue to increase. When $L_{s11}^\alpha$ 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.,

$$L_{\text{joint}} = \gamma_1 L_{s11} + \gamma_2 L_{s11}^\alpha + \gamma_3 L_p + \gamma_4 L_{\text{MS-SSIM}} + \gamma_5 L_a, \quad (7)$$

where $L_{\text{joint}}$, $L_{s11}$, $L_{s11}^\alpha$, $L_p$, $L_{\text{MS-SSIM}}$, and $L_a$ are the total loss, smooth L1 loss [30] for the clear image, smooth L1 loss...
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 TransConv Fusion Dehaze (TFD) network and a tiny Ensemble Reconstruction (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 the 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:

$$I(x) = J(x) \times t(x) + A \times (1 - t(x)) \quad (8)$$

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 hazy model as follows:

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

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 $t(x)$. $\epsilon$ 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, $S(x)$ 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:

$$f_{out} = AFA[f(S(x)), f(P(x))]$$

$$= \sigma(\theta) \times f(S(x)) + (1 - \sigma(\theta)) \times f(P(x)) \quad (10)$$

where $\sigma$ is Sigmoid activation function, $\theta$ 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
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 up-sampling 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.
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 downampling 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 \(\theta_t\) and \(\sigma_t\) 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:

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

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 Refusion 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 \(t\).
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 de-
signed 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-P2-1.1-PED-2021-1572, within PNCDI III. Part of this work was supported by European Marie Skłodowska-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. Ancuti¹ (codruta.ancuti@gmail.com), Cosmin Ancuti¹², Florin-Alexandru Vasluianu³, Radu Timofte³⁴

Affiliations:
¹ ETCI, University of Politechnica 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 Zhou¹ (zhouh115@mcmaster.ca), Wei Dong², Yangyi Liu¹, Jun Chen¹

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

NUSRICQ DEHAZING

Title:
Two-Stage Dehazing for High Resolution Non-Homogeneous Image

Members:
Yinwei Wu¹ (ww3276027@gmail.com), Zhenxiang Jiang¹, Songhua Liu², Xingyi Yang³, Yongcheng Jing³

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 Yang² (islike8399@gmail.com), I-Hsiang Chen², Wei-Ting Chen¹, Zhi-Kai Huang², Yi-Chung Chen³, Chia-Hsuan Hsieh³, Hua-En Chang², Yuan-Chun Chiang², Sy-Yen Kuo²

Affiliations:
¹ Graduate Institute of Electronics Engineering, National Taiwan University, Taiwan
² Department of Electrical 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, Amir-saeed Yazdani, Vishal Monga

Affiliations: Electrical Engineering Department - Pennsylvania State University

Xsourse

Title: Ensemble for Quality-Aware High Resolution Non-Homogeneous 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 Elbazz

Affiliations: Mansoura University

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. Schöö

Affiliations: Department of Information Technology, Uppsala University

LVGroup HFUT

Title: A Two-branch Multi-scale Dehazing Network

Members: Zhao Zhang (czzzhang@gmail.com), Yanyan Wei, Junhu Wang, Suixi Zhao, Huan Zheng

Affiliations: Hefei University of Technology

DH-AISP

Title: A Dehazing Method based on Visual Transformer.

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


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


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


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


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


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


