NTIRE 2021 NonHomogeneous Dehazing Challenge Report

Codruta O. Ancuti Cosmin Ancuti Florin-Alexandru Vasluianu Radu Timofte
Minghan Fu Huan Liu Yankun Yu Jun Chen Keyan Wang Xiyao Wang
Jerome Chang Tianyi Chen Jiahui Fu Wentao Jiang Chen Gao Si Liu
Jing Liu Yi Xu Xinjian Zhang Minyi Zhao Shuigeng Zhou Yudong Wang
Jichang Guo Chongyi Li Qixin Yan Sida Zheng Syed Waqas Zamir
Aditya Arora Akshay Dudhane Salman Khan Munawar Hayat
Fahad Shahbaz Khan Ling Shao Wenjin Yang Jin Lin Xiaotong Luo
Guowen Huang Shuxin Chen Yanyun Qu Haichuan Zhang Tiantong Guo
Vishal Monga Kele Xu Lehan Yang Pengliang Sun Xuetong Niu Junjun Zheng
Xiaotong Ruan Yunfeng Wang Jiang Yangu Xiaochun Cao Jun Luo
Zhuoran Zheng Wenqi Ren Tao Wang Yiqun Chen Cong Leng Chenghua Li
Jian Cheng Chang-Sung Sung Jun-Cheng Chen Eunsung Jo Jae-Young Sim
Zhipeng Luo Sai Wang Zhenyu Xu Geethu M M Akhil K A Sreeni K G
Jeena R S Joseph Zacharias Chippy M Manu Zexi Huang Baofeng Zhang
Yiwen Zhang Jindong Li Mianjie Chen Quan Xiao Qingchao Su Lihua Han
Yanting Huang Kalpesh Prajapat Vishal Chudasama Heena Patel Anjali Sarvaiya
Kishor Upla Kiran Raja Raghavendra Ramachandra Christoph Busc
Jing Hongyuan Huang Zilong Fu Yiran Wu Haoqiang Zha Quanxing
Zhu Zhiwei Lv Hejun

Abstract

This work reviews the results of the NTIRE 2021 Challenge on Non-Homogeneous Dehazing. The proposed techniques and their results have been evaluated on a novel dataset that extends the NH-Haze datset. It consists of additional 35 pairs of real haze free and nonhomogeneous hazy images recorded outdoor. The nonhomogeneous haze has been introduced in the outdoor scenes by using a a professional setup that imitates the real conditions of haze scenes. 327 participants registered in the challenge and 23 teams competed in the final testing phase. The proposed solutions gauge the state-of-the-art in image dehazing.

1. Introduction

Haze is a natural process that affects image quality by drastically reducing visibility in the scene as distance increases. This atmospheric phenomenon is manifested in the presence of small particles in the air, which change significantly the properties of the environment. As a consequence, the hazy scenes are characterised by low contrast, low saturation, color change or additional 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 [20, 44, 22, 45, 31, 8, 1, 36, 6, 9]. Recent methods using CNN [14, 41, 56, 37, 47] have expanded the initial solutions built either on the physical model, or on improving the visual qualities of the image.

Despite of the large number of viable solutions, a significant current problem for the objective verification and classification of dehazing algorithms is the lack of standardized test benchmarks. In the absence of the reference image (ground truth), a common problem in the evaluation of the dehazing techniques is given by the fact that there are no standard algorithms for detecting and measuring errors. The blind evaluation algorithms developed so far do not always generate consistent results because they have also not been validated on real images.

*Codruta O. (codruta.ancuti@gmail.com), Cosmin Ancuti, Florin-Alexandru Vasluianu and Radu Timofte are the NTIRE 2021 challenge organizers. The other authors participated in the challenge.
Appendix.A contains the authors' team names and affiliations.
https://data.vision.ee.ethz.ch/cvl/ntire21/

The first image datasets were synthesized and used information about scene depth and scene attenuation parameter. FRIDA [46] dataset designed for Advanced Driver Assistance Systems (ADAS) was developed using 66 synthetic ground images of various traffic scenes. D-HAZY [5] was generated using over 1400 real images and their known depth maps, by employing the Koschmieder's [30] light propagation model.

An essential issue that makes extremely difficult to collect such hazy image, is the maintenance of lighting conditions, as well as the pixel-by-pixel correspondence between the reference and the hazy image.

Therefore, it is very complicated to record images with and without haze in the same lighting conditions and without changes in the scene. A feasible solution that is probably the most realistic one is to record haze-free natural images and then to record exactly the same scene with haze introduced in the scene by dedicated equipment. The first such image dehazing datasets were introduced at the $NTIRE2018$ [2] image dehazing challenge. O-Haze [7] contains 45 outdoor images and the corresponding images affected by haze, and I-Haze [4] contains 35 indoor images and similar scenes affected by haze in a controlled way. Similarly, DENSE-HAZE [3] contains dense (homogeneous) hazy and ground-truth images and was employed by the NTIRE 2019 image dehazing challenge $NTIRE2019$ [12].

The NTIRE 2021 image dehazing challenge represents a step forward in benchmarking single image dehazing. It is based on an extension of the NH-Haze [10] dataset that was used in the NTIRE 2020 image dehazing challenge [11]. The NH-Haze2 consists of 35 hazy images and their corresponding ground truth (haze-free) images of the same scene. NH-Haze2 contains real outdoor scenes with non-homogeneous haze generated using a professional haze setup. We perform an objective evaluation by comparing the restored output of the methods with the ground truth images of the dataset.

2. Image Dehazing Challenge

The objectives of the NTIRE 2021 challenge on non-homogeneous image dehazing are: (i) to gauge and push the state-of- the-art in image dehazing; (ii) to compare and promote the sota solutions; and (iii) to promote the non-homogeneous image dehazing dataset (NH-Haze [10] and its extension used in this workshop).

2.1. Nonhomogeneous image dataset

The NTIRE 2021 image dehazing challenge was built on the extended version of the former NH-Haze [10] dataset. The NH-Haze2 consists of 35 hazy images and their corresponding ground truth (haze-free) images of the same scene. NH-Haze2 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 ($18percent$ gray) of the color checker. The process of recording a pair of images took about 20-30 minutes.

2.2. Evaluation

For the NTIRE 2021 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.

2.3. Challenge Phases

1. **Development phase:** In this phase, the first 25 images of the NH-Haze dataset were available on the challenge website. The participants used them in order to develop their proposed solutions.

2. **Validation phase:** Another set consisting of 5 images was made public on the challenge website. The participants used the images to validate their solutions, by submitting the produced results to the validation server.

3. **Testing phase:** The participants had access to the last set of 5 images. They used the images to do inference on their proposed solution. The produced results were uploaded to the testing server, along with the factsheet containing information about team contribution, team members, codes and testing phase results.

3. Challenge Results

The challenge registered 327 participants, and a number of 23 teams were ranked in the final phase. Each team had to prepare their submission consisting of codes, testing phase results and a factsheet containing identification information and a description of their proposed solution.

The values for the deployed metrics, computed for each submission, are given in the Table 1, while results characterized by the best value for each of the deployed metrics were given in the Figure 1.

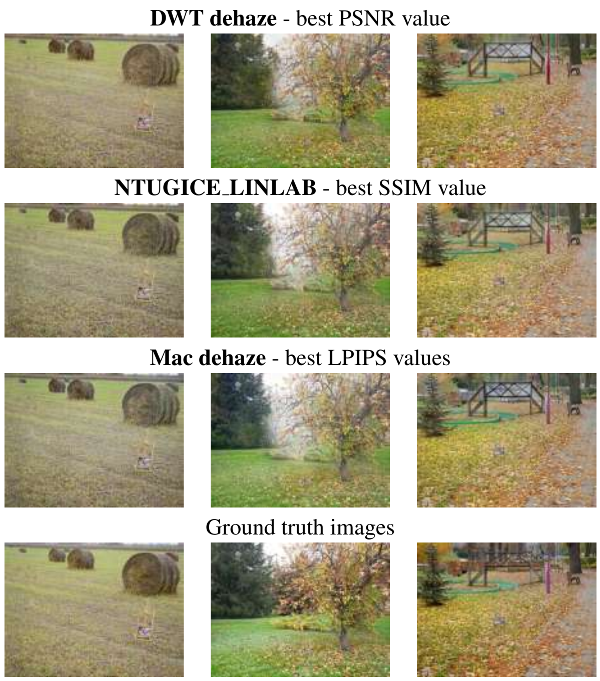


Figure 1: Visual results provided for best performing method on each of the metrics deployed. Best zoom-in on screen for a better view.

4. Challenge Methods

4.1. DWT dehaze

Inspired by [35, 40, 32], this team proposed a novel two-branch generative adversarial network, namely DW-GAN. The network structure is shown in Figure 2. For the first branch, unlike the other method that supervised the training process by a frequency domain loss [34], they proposed to directly embed frequency domain knowledge into the dehazing network. They follow the U-Net [42] to construct the first branch, as the wavelet net. It has a encoder, a decoder and massive skip connections. To meet the requirements for extracting frequency domain knowledge, they adopt five DWT down-sampling modules and six convolutional down-sampling layers to build our encoder. Then, the spatial and frequency representations are concatenated as the input of next down-sampling process. Besides, we employ massive skip connection from encoder to decoder at each feature scales.

In the second branch, they use Res2Net [42] as encoder. In observing that the feature representations learned on a pre-training task can have positive impact on the target task [19, 55], they use the ImageNet [18] pretrained weight as initialization.

In the decoder module, they used pixel-shuffle layer for up-sampling, which makes the size of feature maps gradually recovered to the original resolution. Channel and pixel-wise attention blocks are employed after each pixel-shuffle layer to identify the dynamic hazy patterns. Skip connections are added between encoder and decoder as shown in Figure 2.

Finally, they add a simple 7×7 convolution layer as fusion operation to map the features from two branch to clear images.

The loss functions adopted in this work aims to balance the learning of DW-GAN to generate low distortion and high perceptual quality images. Therefore, they introduced the final loss blend function as follows:

$$L_{total} = L_1 + \alpha L_{SSIM} + \beta L_{perceptual} + \gamma_4 L_{adv} \quad (1)$$

where $\alpha = 0.2, \beta = 0.001$ and $\gamma = 0.005$ are the hyper-paramters weighting for each loss functions. L_1 denotes L1 loss, L_{SSIM} represents MSSSIM loss [48], $L_{perceptual}$ is perceptual loss [29]. We adopt the discriminator in [63] to form the adversarial loss L_{adv}.

The overall network architecture is shown in Figure 2.

4.2. Mac dehaze

Mac dehaze team proposes a simple baseline for Non-homogeneous dehazing via transfer learning as a two-branch neural network to deal with the above mentioned problems. The structure diagram of the network is shown in Figure 3.

The first branch, namely transfer learning sub-net, is built upon a ImageNet [17] pretrained Res2Net[21] inspired by [51]. It aims to extract robust global representations from input images with pre-trained weights. To achieve this, instead of skip connecting all resolution features from encoder to decoder, they omitted the skip connection of full resolution features. It physically ensures that the fine details of input images would not be preserved and thus forces the network to focus more on extracting robust global representations. As a result, the ImageNet pretrained branch can help address the problem of lacking training data. Besides, in favor of the strong mapping capability of residual channel attention network (RCAN) [60], they chose the domain fitting sub-net using RCAN as second branch.

The domain fitting branch has five residual groups, and each group has ten residual blocks. Unlike the original network setting [60] that down-samples the input images at

Participant		Results				Solution details					
		Fidelity		Perceptual quality		Runtime	GPU/	extra		deep learning	loss
Team	User	PSNR↑	SSIM↑	$LPIPS_1\downarrow$	$LPIPS_2\downarrow$	img.[s]	CPU	data	ens.	framework	
Top perceptual quality solutions											
DWT dehaze	eason97	21.076_1	0.839_3	0.169	0.203	1.558	1080Ti	NH-HAZE-20	-	Pytorch	$L1, L_{SSIM}, L_{GAN}, L_{perc}$
Mac dehaze	ken103	21.018_2	0.837_5	0.168	0.196	2.03	Tesla V100	NH-HAZE-20	-	Pytorch	$L1, L_{SSIM}, L_{GAN}, L_{perc}$
NTUGICE_LINLAB	Jerome_Chang	20.898_3	0.844_1	0.175	0.194	60	Tesla V100	NH-HAZE 20	8x	Pytorch	n/a
buaa_colab	buaa_colab	20.619_4	0.834_7	0.202	0.220	2.52	3x RTX2080Ti	-	8x	Pytorch	$L1, L_{Lab}, L_{Laplacian}, L_{KT}$
Bilibili AI & FDU	splinter23	20.598_5	0.823_{11}	0.182	0.212	0.64	1080Ti	NH-HAZE 20	8x	Pytorch	$L1, L_{FFT}, L_{BReLU}$
TJUVIPLab	WangYudong	20.537_6	0.835_6	0.183	0.205	8.0	RTX3090	NH-HAZE O-HAZE DENSE-HAZE	8x	Pytorch	$L1, L_{SSIM}, L_{Gradient}, L_{perc}$
TeamInception	swz30	20.013_9	0.832_8	0.177	0.205	1.4	4x Tesla V100	-	8x	Pytorch	$L1, L_{SSIM}, L_{VGG}$
VIPLab	Yangwj	19.675_{10}	0.824_{10}	0.173	0.203	0.042	1080Ti	NH-HAZE	-	Pytorch	$L1, L_{perc}$
iPAL-GridFFA	haichuan	19.567_{12}	0.839_2	0.178	0.194	1.01	RTX2080Ti	NH-HAZE		Pytorch	$L1$, smoothed $L1$, L_{SSIM}, L_{GAN}
Medium perceptual quality solutions											
debut_kele	debut kele	20.264_7	0.832_9	0.200	0.219	n/a	RTX2080Ti	-	-	Pytorch	$L1, L_{SSIM}, L_{std}$
alibaba-cipp	alibaba-cipp	20.231_8	0.802_{16}	0.178	0.220	9.0	8x Tesla V100	Place2, O-HAZE, DENSE-HAZE	-	Pytorch	n/a
team_Dou	xiaodou	19.654_{11}	0.812_{14}	0.187	0.208	0.94	GPU	-	-	Pytorch	n/a
LDGLI	YiqunChen1999	19.522_{13}	0.838_4	0.192	0.207	1.13	RTX3090	NH-HAZE, hand-designed	-	n/a	$L2, L_{SSIM}$
NTUDS-LINLAB	ChangSung	19.288_{14}	0.817_{12}	0.220	0.234	n/a	GPU	NH-HAZE	n/a	Pytorch	n/a
VIP_UNIST	Eun-Sung	19.156_{15}	0.809_{15}	0.205	0.227	0.034	RTX2080Ti	I-HAZE O-HAZE DENSE-HAZE NH-HAZE	-	Pytorch	n/a
DeepBlueAI	DeepBlueAI	18.970_{17}	0.816_{13}	0.197	0.210	1.0	4x Tesla V100	-	-	Pytorch	$L1, L_{perc.}, L2$
Low perceptual quality solutions											
SP-CET	Geethu	19.050_{16}	0.800_{17}	0.191	0.222	0.409	GPU	-	-	n/a	n/a
Dehaze_aicte	CHIPPYMMANU	18.302_{18}	0.733_{22}	0.295	0.309	1.0	GPU	-	-	Keras	n/a
HZLLC	BFZhang	18.043_{19}	0.742_{21}	0.313	0.295	0.018	RTX2060	-	-	Pytorch	$L2, L_{perc.}, L_{t.v.}$
WaveFull_XM	R0use	17.974_{20}	0.771_{20}	0.271	0.286	10.4	Titan Xp	-	-	Pytorch	n/a
SVNIT_NTNU_Team	kalpesh svnit	17.905_{21}	0.788_{18}	0.248	0.264	24.0	Quadro P5000	O-HAZE, I-HAZE, Dense-HAZE	-	Pytorch	$L1$
CVML_Lab	vishalchudasama	17.657_{22}	0.783_{19}	0.247	0.260	1.2	Titan X Pascal	O-HAZE, I-HAZE, Dense-HAZE	-	Tensorflow	$L1$
BUUMASRC	BUUMASRC.	12.006_{23}	0.623_{23}	0.467	0.445	445.37	CPU	O-HAZE	-	Matlab	n/a
no processing	*baseline*	10.936	0.565	0.588	0.489	0.0					

Table 1: NTIRE 2021 NonHomogeneous Dehazing Challenge preliminary results in terms of PSNR, SSIM, LPIPS [59], on the NH-Haze test data. For LPIPS, both Alex-net($LPIPS_1$) and VGG16 ($LPIPS_2$) pretrained model were used as feature extractors.

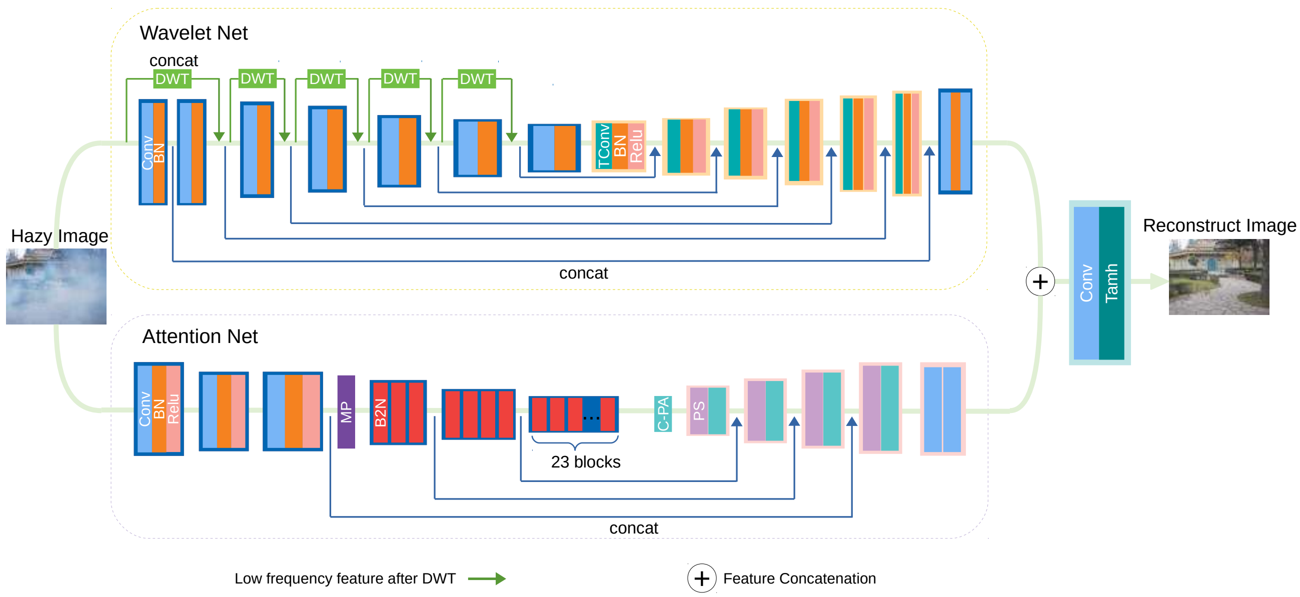
the front of entire network, the second branch always maintains the original resolution of the inputs and avoids using any down-sampling operation. This adjusting avoids losing of fine-detailed features. Since the sub-network is training from scratch and built with full-resolution purpose, it would fit on the current data and perform well on the specific training image domain. The final output of the entire network is produced by our fusion layer. Specifically, the fusion layer takes the concatenation of features from two branches and then maps the features to clear outputs. Moreover, adversarial loss is proved to be effective in helping restore photo-realistic images [32]. Especially for the small-scaled dataset, the pixel-wised loss function usually fails to provide sufficient supervision signals to train a network for recovering photo-realistic details. Therefore, they implemented the adversarial loss with the discriminator in [63]. The overall loss function is a linear combination of smooth L1 loss L_{l1}, MS-SSIM loss[49] L_{SSIM}, perceptual loss[29] $L_{perceptual}$, and adversarial loss L_{adv}, as shown in Equation 2.

$$L = \gamma_1 L_{l1} + \gamma_2 L_{SSIM} + \gamma_3 L_{perceptual} + \gamma_4 L_{adv} \quad (2)$$

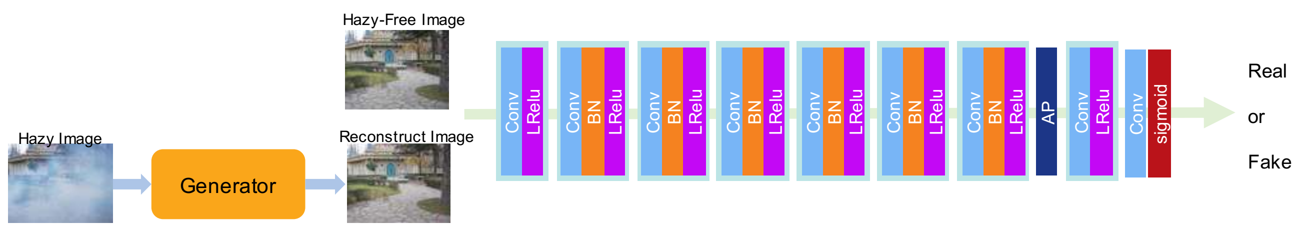
4.3. NTUGICE LINLAB

To cope with the property of nonhomogeneous haze, they proposed Adaptive Dehazing Network (ADN). This is a two-branch dehazing network which aims to adaptively dealing with region covered by thin haze or heavy haze. As shown in the Figure 5, ADN consists of two branches, the Primary Branch and the Enhanced Branch. The Primary Branch manage the region covered by thin haze. On the other hand, the Enhanced Branch will focus on making up for the region Primary Branch doesn't dehaze well, which is mostly severely contaminated area. Besides, to blend the output of the two branches, we design the Weight Map Generator which will output a two-channel weight map which is for output of the two branches respectively. Afterwards, the output of the two branches will be element-wise multiplied with their corresponding weight map and, by adding them, the final result will be produced.

The model of encoder and decoder of the primary branch is based on the Perceptual Pyramid Deep Network [57]. Both branches share the same encoder, but they own their individual decoders. The difference between the decoders is that we replace the normal convolution kernel with the dilated convolution kernel attempting to enlarge receptive field of the enhanced decoder, which enables the Enhanced Branch to gain ability to deal with heavy haze. Moreover, they added some attention modules such as CBAM[50]. This combines spatial attention and channel attention to let both decoders pay attention on the more important features extracted by the encoder.



(a) Generator



(b) Discriminator

Figure 2: The network structure of the proposed method. The generator is a two-branch network, which consists of Wavelet Net and Attention Net. The same color used in the rectangles denotes the same operation. 'Conv', 'BN', 'TConv', 'MP', 'PS', 'AP', 'LReLu' denotes convolution, batch normalization, transpose-convolution, max-pooling, pixel-shuffle, average-pooling, and leakyReLu. 'B2N', 'C-PA' and 'DWT' denote bottle2neck, channel and pixel-wise attention, and discrete wavelet transform modules respectively.

4.4. buaa_colab

Their contribution is the modified version of Knowledge Transfer Network [51], namely, the Super Resolution Knowledge Transfer Dehazing Network (SRKTDN). As is shown in Figure 6, the network we used contains two portions: main network and teacher network, and the main network consists of dehaze network and super-resolution network.

The dehaze network uses Res2Net101 as encoder, and PixelShuffle for the upsampling operation. The network uses attention mechanisms to restore the haze-free image, as the channel attention blocks and pixel attention blocks [40].

They used a teacher network to generate low-level feature map. The teacher network is trained by ground truth pairs of the dataset, in order to capture the necessary information for image restoring. Different from Knowledge Transfer, the structure of teacher network they used is different from dehaze network. While the dehaze network uses Res2Net101 as encoder to ensure capability of haze removal, the teacher network uses ResNet18 as the encoder, in order to improve the generalization ability and to reduce training time and GPU memory consumption.

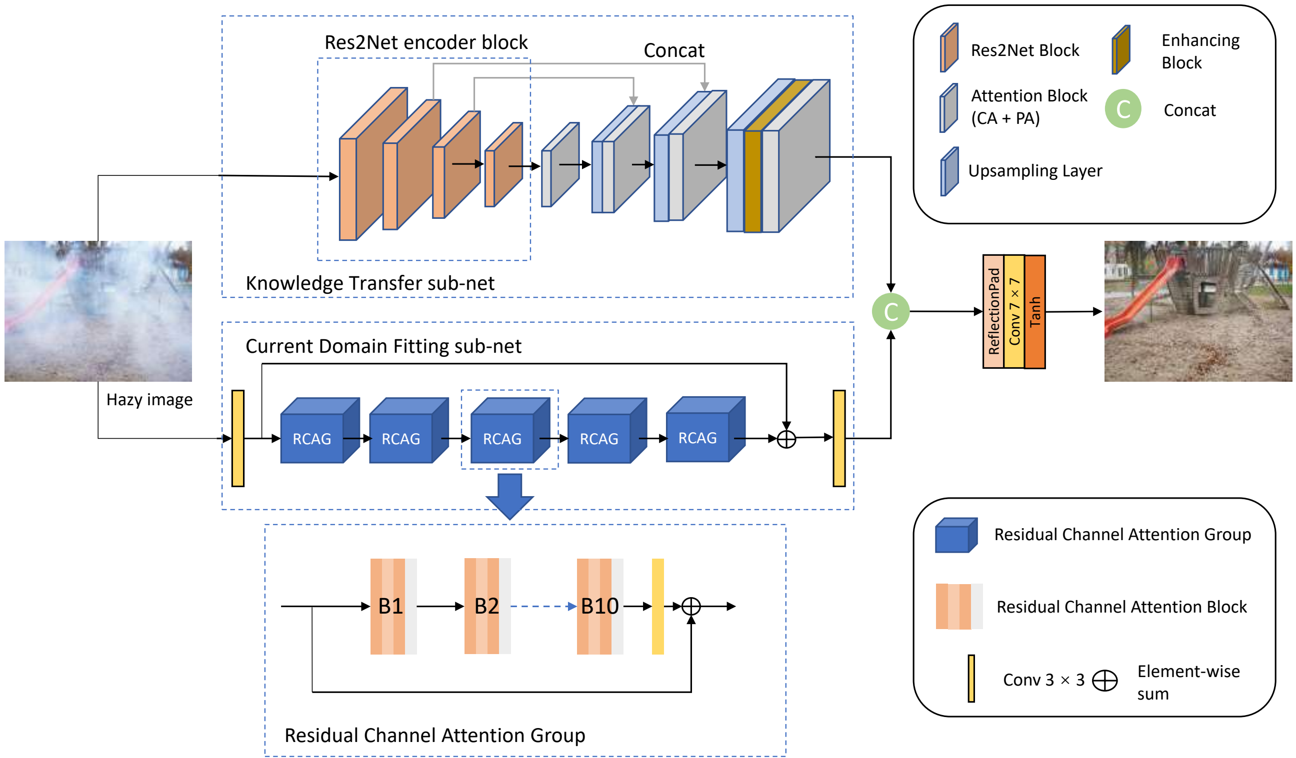


Figure 3: Illustration of Mac dehaze model architecture.

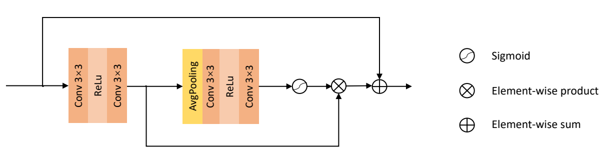


Figure 4: Residual Channel Attention Block

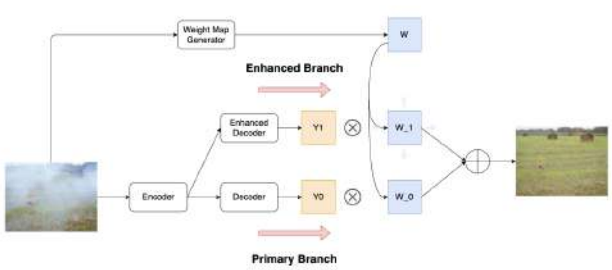


Figure 5: Information flow along NTUGICE-LINLAB proposed model.

Meanwhile, inspired by TDN[34], they used a super-resolution network to enhance detail restoration. The super-resolution network uses three Wide Activation Block to capture details.

The training objective used is a blending of L1 loss, Laplacian loss, Lab-color space L2 loss and the Knowledge Transfer loss.

L1 loss is calculated as states in Equation 3, where I and J refer to the hazy image and ground-truth haze free image, respectively, and $M(\cdot)$ stands for the main network.

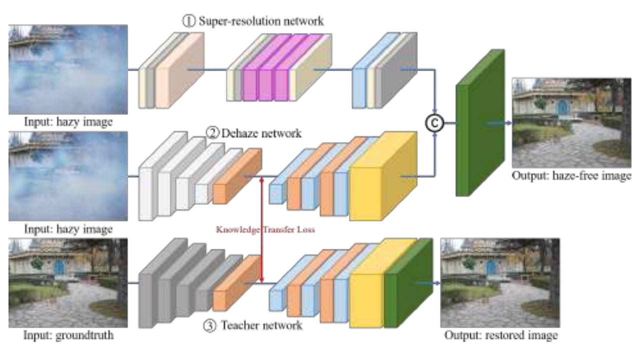


Figure 6: Architecture of buaa_colab proposed architecture.

$$L_1 = |J - M(I)|_1 \tag{3}$$

Laplacian loss uses Laplacian pyramid representation of the image and calculates L1 loss for 5 levels[13]. $L^j(\cdot)$ in

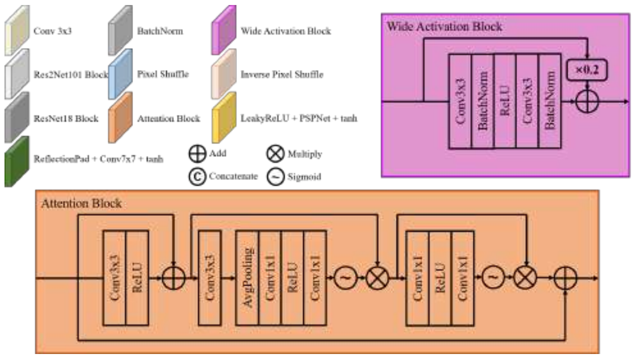


Figure 7: Details over the attention modules used by buaa_colab team.

the Equation 4 is the j-th level of the Laplacian pyramid representation. Laplacian loss focuses on edge of the image and prevent the output from being blurry.

$$L_{lap} = \sum_{j=1}^{5} 2^{2j} |L^j(J) - L^j(M(I))| \tag{4}$$

L2 loss of Lab color space is used to refine color of the output image. Different from L1 loss, L2 loss pay more attention to pixels that have a relatively high deviation from the ground-truth image. Besides, unlike RGB color space, Lab color space is designed to resemble human vision. $\mathrm{Lab}(\cdot)$ in the Equation 5 refer to the RGB-to-Lab transformation.

$$L_{Lab} = |\mathrm{Lab}(J) - \mathrm{Lab}(M(I))|_2 \tag{5}$$

Identical to the method proposed in [51], Knowledge Transfer loss is L1 loss between feature map of dehaze network and the one of teacher network. Knowledge Transfer loss helps the Res2Net101 encoder to imitate the teacher's output, hence learning information of haze removal. In the Equation 6, I' and J' refer to the output feature map of dehaze network encoder and teacher network encoder respectively.

$$L_{KT} = |J' - I'|_1 \tag{6}$$

The total loss is calculated using the Equation 7.

$$L = 1 \times L_1 + 0.3 \times L_{lap} + 0.5 \times L_{Lab} + 1 \times L_{KT} \tag{7}$$

4.5. Bilibili AI & FDU

They use the Trident Dehazing Network[34] proposed in NTIRE2020 NH-Dehazing challenge as their model. The architecture is depicted in the Figure 8. Different from the proposed paper, they are training their model using the image pairs with a small size (256×256), in the early phase of the training procedure. Then the resolution will be progressively increased to a higher dimension (384×384).

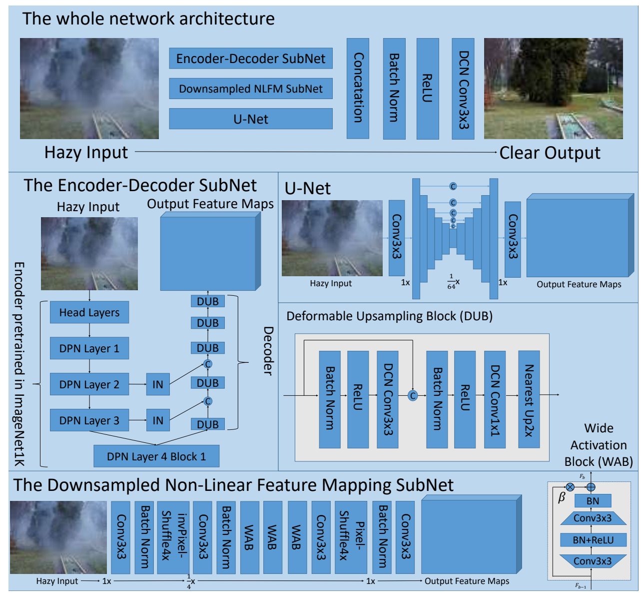


Figure 8: The network architecture of the solution proposed by team Bilibili AI & FDU

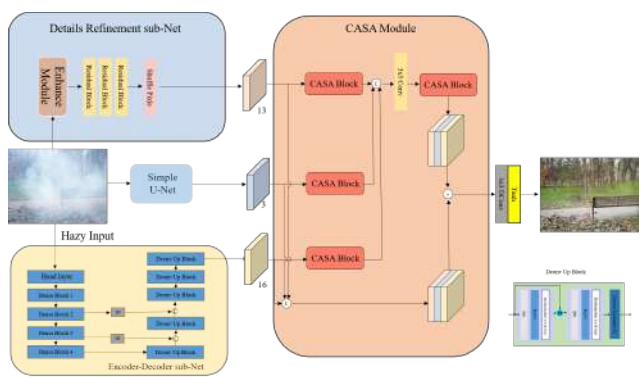


Figure 9: An overview of the proposed MCDNet architecture.

4.6. TJUVIPLab

TJU_VIPLab team proposed a CNN-based Multi-task Collaboration Dehazing Network(MCDNet) to directly learn the mapping relations between nonhomogeneous haze image and haze-free clear image. MCDNet consists of three sub-nets inspired by [34] and a Channel Attention-Spacial Attention(CASA) Module inspired by [52]. The overall structure of MCDNet is shown in Figure 9, the Simple U-Net is used to obtain basic haze-free image, the Encode-Decode sub-Net(EDN) is used to extract features and get basic dehazing feature map, and the Detail Refinement sub-Net(DRN) is used to get high frequency details of the haze free image features. CASA Module is used to enhance useful information to improve the realistic of image color.

The architecture of simple U-Net is shown in Figure 10, which is a light encode-decode structure. There are 6 downsampling/upsampling blocks, using 4×4 convolution (transposed convolution), with stride$=2$, and suitable padding, to match the dimensions of the input image.

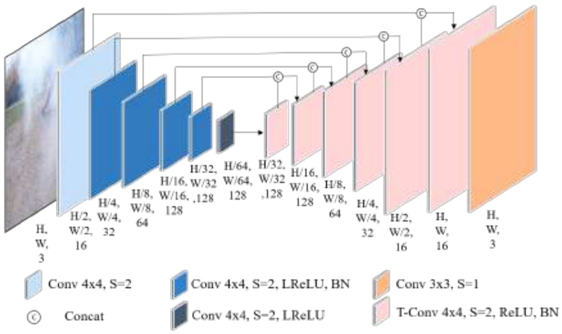


Figure 10: Structure of Simple U-Net

DenseNet-101 pretrained in ImageNet is as the backbone of EDN's encoder part. Same as [34], the decoder is composed of five Deformable Convolution Upsampling (*DConv Up*) blocks, as shown in right side of Figure 9. The *DConv Up* block consists of 2 deformable convolution blocks. The input feature is first fed into a residual-3×3 DConv block, then fed into a 1×1 *DConv* block, and finally go through an 2x nearest-upsampling layer to obtain the upsampled feature. The deepest two block used skip connection from the output of denseblock34 respectively. Meanwhile, EDN use trainable instance normalization for skip connections.

DRN first uses 2x downsampled enhancing model(EM) to capture multi-scale detailed feature maps, then fed into 3 residual block. Pixel Shuffle layer is a 2x upsampling way, which is used to change the feature maps from $H\times W\times4C$ to $2H\times2W\times C$, where H, W, C is the height, width and channel of feature map. As shown in Figure 11, EM obtain 4x, 8x, 16x, 32x downsampling feature and do 2x, 4x, 8x, 16x upsampling respectively. The feature maps concatenate with 2x downsampling and fed into 3×3 convolution layer.

CASA Module contains 4 CASA-block, which is shown at the bottom of Figure 9. Three sub-Net outputs first fed into 3 CASA-blocks respectively, then concatenate together and fed into the next CASA-block, which can further enhance useful information. Finally, the CASA Module output add with concatenate feature map of the output of three sub-Net, and fed into 3×3 DConv layer. The last layer uses $Tanh$ as activation function, which sets the output between -1 and 1.

4.7. Team Inception

They present an architecture, named MPRNet, that is based on a recent work [54]. As illustrated in Fig. 12, MPRNet consists of two stages to progressively restore images.

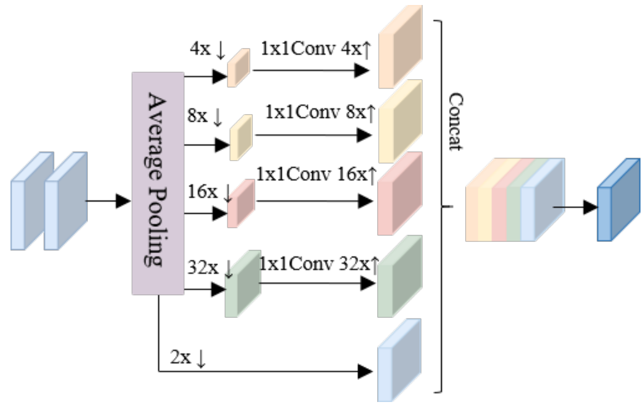


Figure 11: Structure of Enhance Module

In the first stage they employ three encoder-decoder subnetworks that independently operate on the red, green and blue channels of the hazy input image. It is based on the observation that each channel is affected by the haze differently. For instance, the density of haze in the blue channel is much higher than in the red channel. Therefore, the solution proposes different parameters allocation per channel, with respect to the haze density. For the output of each encoder-decoder subnetworks, they deployed a supervised attention module (SAM) [54]. The schematic diagram of SAM is shown in Figure 13.

The output features from the first stage are concatenated and passed as input to the final stage. This stage act as a refinement stage and outputs the final dehazed image. To train the proposed network, they use L1 loss at the first stage, and the loss function stated in Equation 8 for the final stage.

$$\mathcal{L}_f = \alpha\mathcal{L}_1(\hat{\mathbf{y}},\mathbf{y}) + \beta\mathcal{L}_{\text{MS-SSIM}}(\hat{\mathbf{y}},\mathbf{y}) + \gamma\mathcal{L}_{\text{VGG}}(\hat{\mathbf{y}},\mathbf{y}) \quad (8)$$

The first term (L1 loss) and second term (multi-scale structural similarity measure) computes differences between the network's output and the ground truth directly at the pixel-level. The last term of the loss function compares the deep feature representations of the output and ground-truth images extracted with the VGG network pre-trained on the ImageNet dataset. In Equation 9, the formula of this loss function is given, where N is the number of pixels in the image and $\phi(.)$ is the transformation after the *conv2* layer of the VGG net.

$$\mathcal{L}_{VGG}(\hat{\mathbf{y}},\mathbf{y}) = \frac{1}{N} \parallel \phi(\hat{\mathbf{y}}) - \phi(\mathbf{y}) \parallel_2^2, \quad (9)$$

4.8. VIPLab

Densenet network has a wide range of applications in many fields due to its dense connection characteristics, we

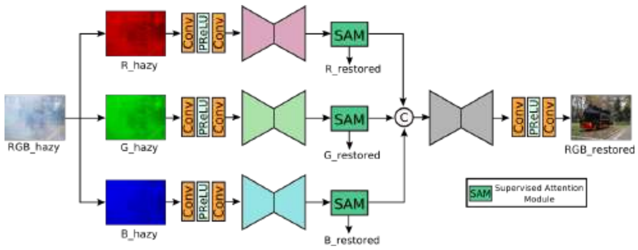


Figure 12: Overall framework of MPRNet.

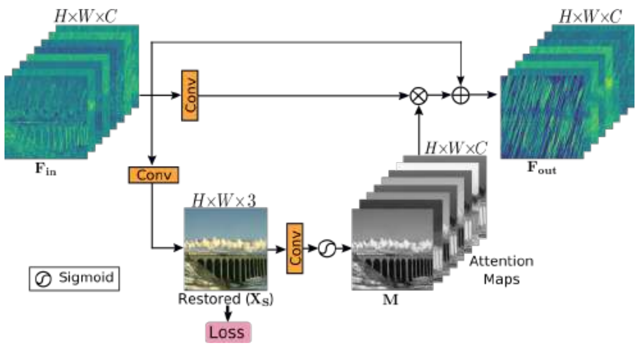


Figure 13: Supervised attention module (SAM).

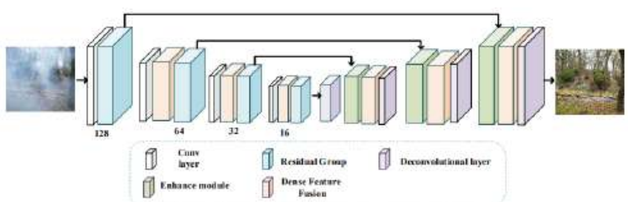


Figure 14: VIPLab proposed architecture.

use it as the backbone network for dehazing. The boosting algorithm operates the refinement

process on the strengthened image, based on the previously estimated image. The algorithm has been shown to improve the Signal-to-Noise Ratio (SNR) under the axiom that the denoising method obtains better results in terms of SNR on the images of the same scene but less noise. For image dehazing, the Enhance strategy can be formulated similarly as

$$\hat{j}^{n+1} = g\left(I + \hat{j}^{n}\right) - \hat{j}^{n} \tag{10}$$

where $\hat{J}^n$denotes the estimated image at the n-th iteration, $g()$ is the dehazing approach, and $I + \hat{J}^n$ represents the strengthened image using the hazy input I. They show that the boosting method can facilitate image dehazing performance in terms of Portion of Haze (PoH) under a similar axiom as that for denoising.

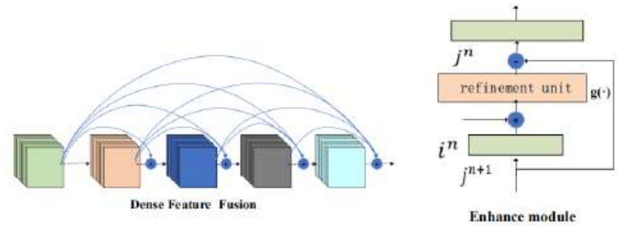


Figure 15: Dense Feature Fusion

4.9. iPAL-GridFFA

The team designed an end-to-end GAN Network for non-homogeneous haze removal which consists of a generator network, a group structure , and a discriminator. For the generator part, they chose a 3×6 Grid network with Feature Fusion Attention. The generator network is an enhanced network of GridDehazeNet [38].

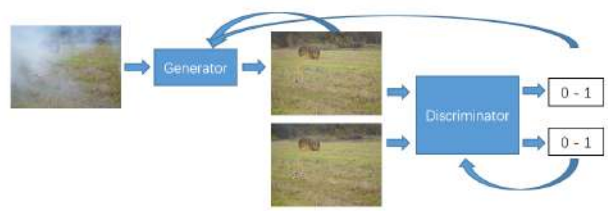


Figure 16: Solution proposed by iPAL-GridFFA

The Group Structure combines 15 Basic Block structures which conclude the *Pixel attention* [40] and *Channel attention* [23], with skip connections for each of the modules. For the discriminator architecture, they use a similar idea to *Patch GAN*[61], using the discriminator score for the image as the average score over the set of disjoint image patches that can be fed to the discriminator for each training image.

Besides the adversarial loss, they use SSIM loss function as well as Smooth L1 loss and L1 loss. Moreover, the cosine annealing [25] mechanism is used for the adjustment of the learning rate.

Figure 17 provides a detailed illustration of Group Structure. Local residual learning allows the region with a thin haze to be bypassed through multiple local residual connections. While *Channel Attention* concerns that different channel features have different weighted information, the *Pixel Attention* makes the network pay more attention to informative features.

They opted for a simple network with the building block made of a convolution layer, a Batch Normalization layers, and using ReLU as the activation function. The network contains 3 building blocks in serial, where the first two blocks are attached to a Max Pooling operation.

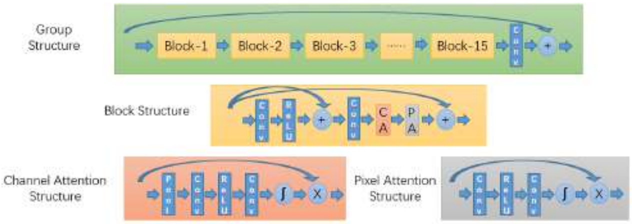


Figure 17: The architecture of the group structure proposed by iPAL-GridFFA.

4.10. debut_kele

They prooposed a deep learning architecture, similar to [39] that estimates physical parameters in the haze model. Compared to it, they experiemnt with different data augmentation strategies, a custom loss function, and the Stochastic Weight Averaging optimization [28]. Their network uses a shared DenseNet encoder and four parallel distinct decoders to jointly estimate the scene information. Moreover, the channel attention mechanism is utilized to generate different feature maps and a novel Dilation Inception module at the direct decoder to generate missing features at densely-hazed regions using non-local features. A final blend consisting of $L1$, L_{SSIM} and $L_{std.}$ is used, where $L_{std.}$ is used to suppress extreme values throughout the image.

4.11. alibaba-cipp

They adopt the GAN framework, which is widely known to be able to do image restoration. The generator consists of two stages to progressively dehaze the input hazy images. In the first stage, they use a residual-inresidual dense block (RRDB)[53] as the basic module to generate the coarse dehazed image. The second stage, we use a encoder-decoder architecture to refine the coarse image. In order to combine information from different receptive fields, they deployed a multi-patch transformer structure between the encoder and the decoder, to guide the network to refine the result. The proposed solution is illustrated in Figure 18.

4.12. Team Dou

Team Dou proposed an improvement over the work published in [54], based on muti-scale features extraction. Principles as attention mechanisms, residual learning, feature fusion and hybrid delated convolution are combined in an architecture illustrated in Figure 19.

4.13. LDGLI

The architecture is illustrated in the Figure 20. They used a pre-trained ResNeSt [58] model to extract the features at 5 levels, and employ the proposed NonHomogeneous Dehazing Block (NHDBlock) (see Figure 21) to remove the haze and recover the image, where the $\times 2$ is an upsampling operation which is done by a transposed convolution and a nonlinear activation. The NHDBlock, which mainly consists of a sequence of four NonHomogeneous Dehazing Units (NHDUint). Each of the proposed NHDUnit tries to augment the input feature I by utilizing the global feature G and local feature L, and produces output augmented feature O. They introduce the residual connection in NHDBlock to help preserve spatial details.

4.14. NTUDS-LINLAB

They proposed a U-Net architecture [43] (see Figure 22) dehazing model with multiscale dense feature based on dense blocks [26] and residual blocks [24]. Their Encoder module used densenet which was pretrained on image net. The most different between their model and U-Net is the re-designed skip connection. Aiming at utilizing lower level feature map, they used a concatenation between the decoder feature map and the upsampled lower feature map.

4.15. VIP_UNIST

They proposed an end-to-end dehazing method named Selective Residual Learning for Multi-scale Dehazing. Overall network architecture (see Figure 23) shows multi-scale inputs and outputs and the use of proposed selective residual blocks. First, adopting the multi-scale architecture in the method is an effective way to train model that can extract both high-level and low-level features. Second, the selective residual block reduces unnecessary artifacts of the final outputs. The selective residual block is an operation that is basically same as the residual block in the ResNet. However the final output $\mathbf{y}$ is the weighted sum of the skip connection($\mathbf{x}$) and the output of the last convolutional layer ($\mathcal{F}(\mathbf{x})$), which can be denoted as Equation 11. Since both the skip connection and the convolutional output are weighted, the block selectively takes the branches. Therefore, the artifacts that are crucial to the fidelity of the final outputs are alleviated.

$$\mathbf{y} = \alpha\mathbf{x} + \beta\mathcal{F}(\mathbf{x}) \qquad \alpha, \beta \in \mathrm{R} \tag{11}$$

4.16. DeepBlueAI

They used Trident Dehazing Network as the core subnetwork, and based on DMPHN, they designed a new network named Cascaded Multi-Path Dehazing Network (CMPDN). The team used a simple but effective data augmentation strategy named Hazing Reinforcement Augmentation (HRA). Compared with the traditional method, they perform additional data augmentation on the cropped subimages, that is, randomly initialize two fog masks with a total area of 64×64 and merge them with the sub-images, in order to solve the problem of insufficient training for nonhaze area/shallow haze area. Figure 24 shows the effect of

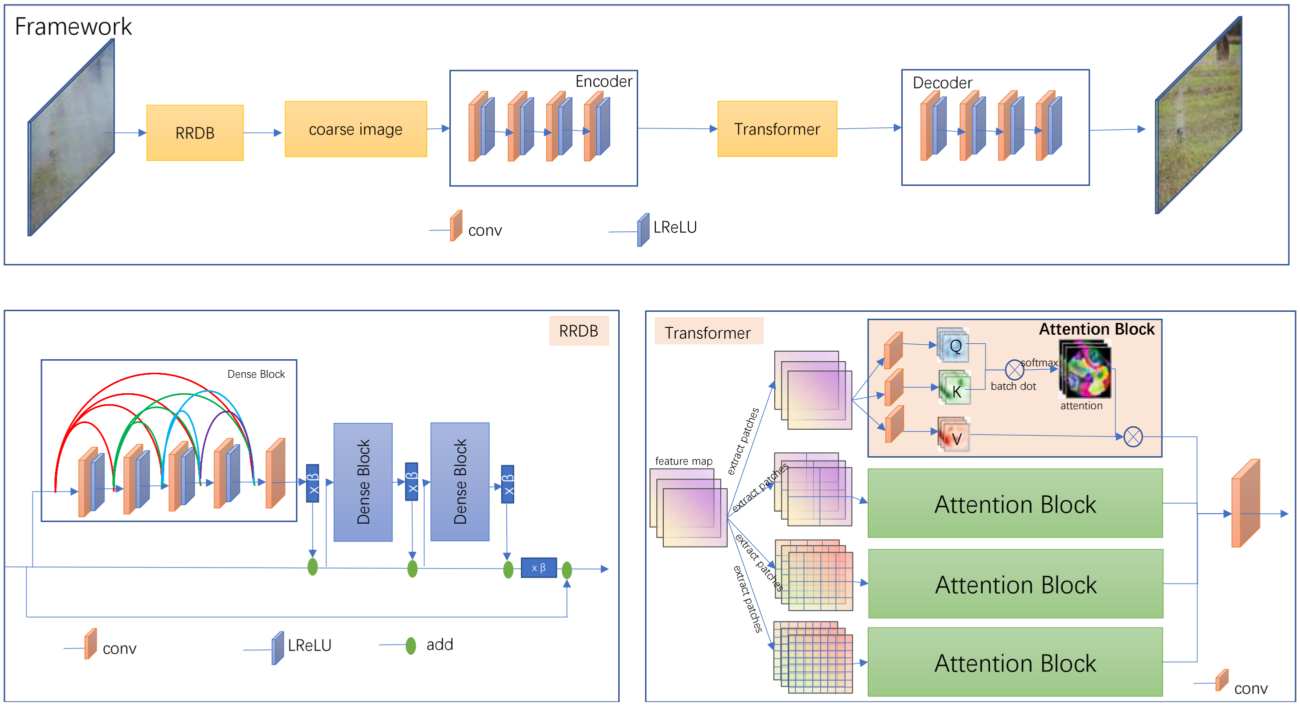


Figure 18: The framework proposed by Team alibaba-cipp.

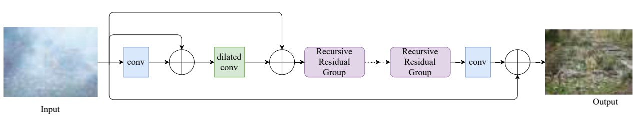


Figure 19: Architecture of the solution proposed by Team Dou.

HRA on the dehazing results. The left and right columns are the comparison results before and after using HRA. HRA effectively removes dense haze and maintains the original texture of the image, making the result clearer.

4.17. SP-CET

This method includes a multi-level CNN model called Deep Multi-patch Hierarchical Network(DMPHN). It uses multi patch hierarchy as input and exploits dehazing at different scales. Each level of the network consists of an encoder and a decoder. The overall architecture of the method is shown in Figure 25.

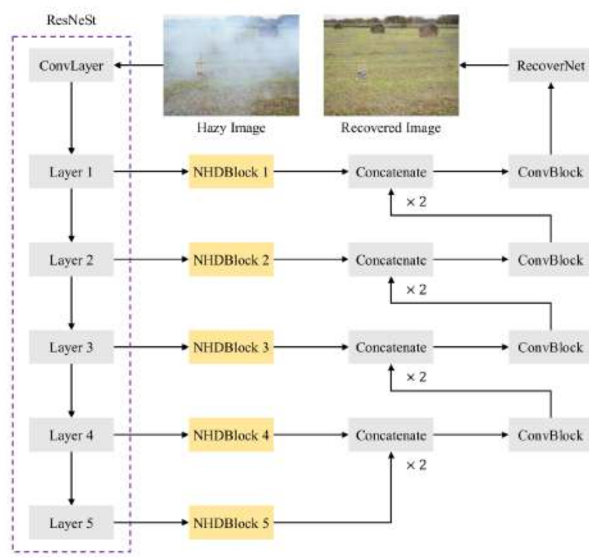


Figure 20: Architecture of the solution proposed by LDGLI team.

4.18. Dehaze_aicte

This team proposed the GANID method, tackling the image dehazing problem in the adversarial learning framework. Deep supervision [33] in UNet++ is used in generator (see Figure 26) to create secondary output maps, which allows for models to be pruned, called model pruning. Deep supervision operates in two modes, namely accurate mode and fast mode. In accurate mode averaged output is calculated from all output branches. In fast mode, one of the

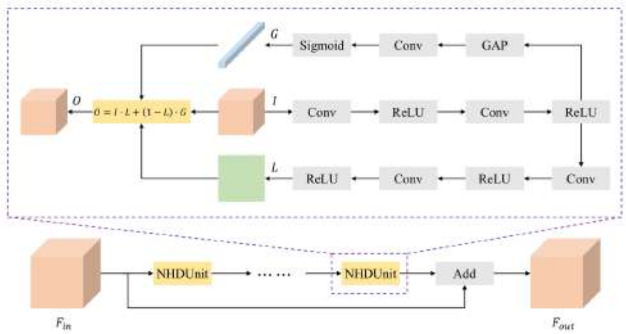


Figure 21: Schematic illustration of the NHDBlock used by LDGLI team.

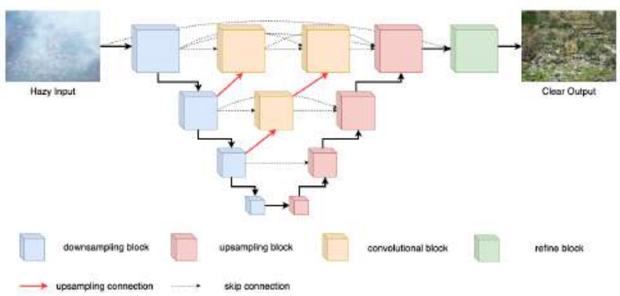


Figure 22: The architecture of the model used by NTUDS-LINLAB team.

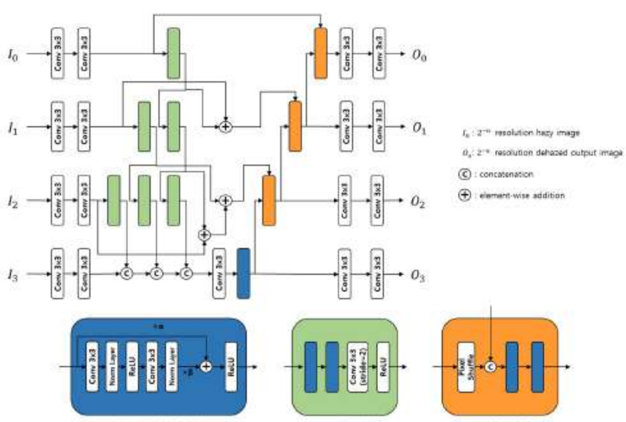


Figure 23: The architecture of the VIP_UNIST proposed method.

output branches is selected for the final response map. Fast node is also known as a pruned mode. Model pruning reduces the complexity of the network with some modest drop inaccuracy. The accurate model is used in the proposed method. Deep supervision means that all the responses from nodes $X^{k,l}$ with $k = 0$ and $l = 1, 2, 3, 4$ appended a 1×1 convolution along with k kernel followed by an activation

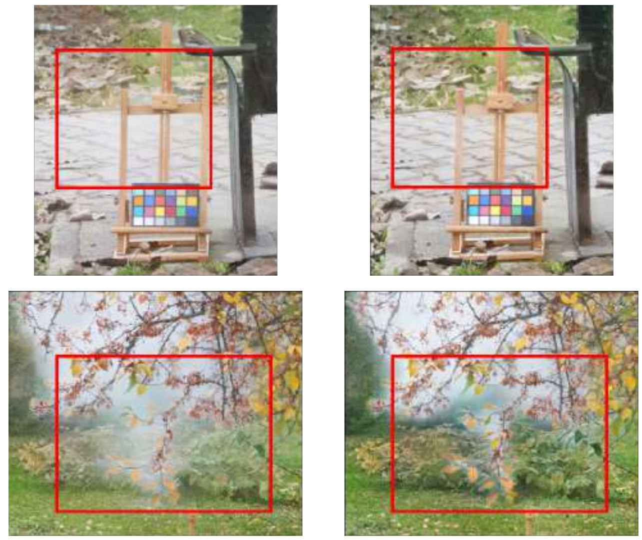


Figure 24: The effect of HRA on the dehazing results. The left and right columns are the comparison results before and after using HRA. HRA effectively removes dense haze and maintains the original texture of the image, making the result clearer.

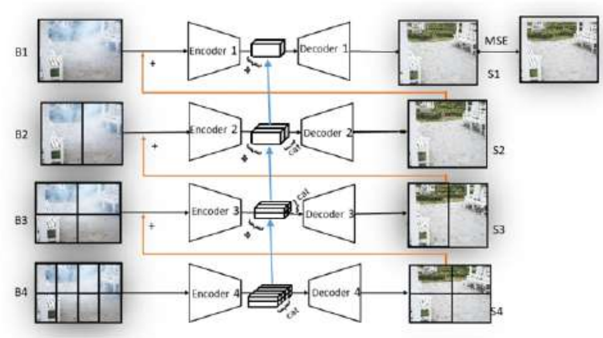


Figure 25: Architecture of the DMPHN model.

function (sigmoid). A detailed description of UNet++ is given [62]. Patch discriminator in the Conditional GAN [27] is used with some additional layers. Rather than using pixel-based comparison, a patch-based comparison is made in this model.

4.19. HZZLC

This team proposed a solution named VMPHN, using an end-to-end Multi-patch architecture. Figure 27 depicts the architecture of the proposed solution. The information flow is like a "V" shape.The level-1 patch is just an original image that is fed to the first Encoder-Decoder and its output is then added to the level-2 patch.The result of level-2 is the input of the second Encoder-Decoder,and the level-3 is the same condition.Now the top-bottom flow is

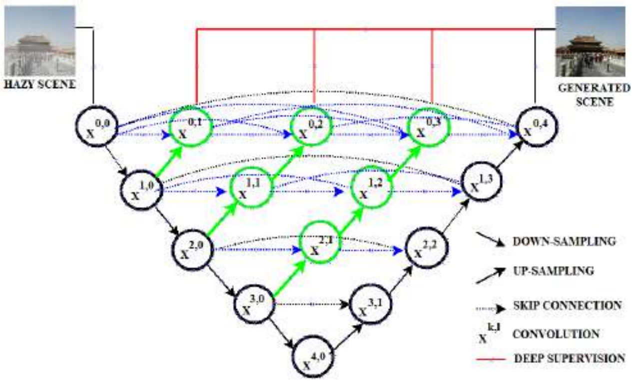


Figure 26: Generator of the GANID method.

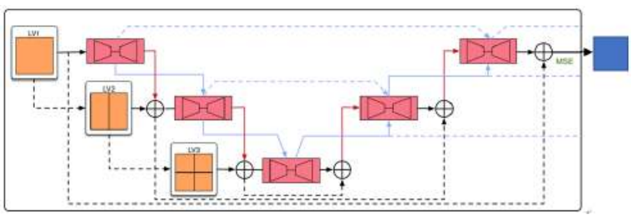


Figure 27: VMPHN model architecture.

completed.As to the bottom-top flow,the third output of the Encoder-Decoder is added with the input of Encoder-Decoder,the result is then feed to the forth Encoder-Decoder net.Finally,we get the fifth Encoder-Decoder's output and adopt the MSE loss, perception loss and total variation loss to get the dehazed images.

4.20. WaveFull-XM

This team combines the GCAN model [15] and the PAM model [16] to build a network implementing residual learning, in order to learn the features on non-homogeneous haze. Several PAM modules are added to different layers of the residual network, which enable the network to learn local information from both high-level semantics and low-level semantics. Finally, the features of different layers are fused as the ultimate features.

4.21. SVNIT_NTNU_Team

The proposed *MACNet* consists of a multiple attention based approach to tune with the given non-homogeneous haze image adaptively. The architecture of the solution is depicted in Figure 28. In order to deal with the non-homogeneous haze, the proposed network use channel attention, pixel attention and spatial attention which help the network in learning the statistical characteristics of haze image. The $L1$ loss function, between the hallucinated image and the ground truth result was used as the minimized objective.

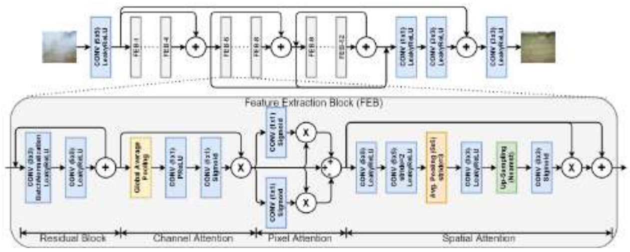


Figure 28: Architecture of the MACNet model.

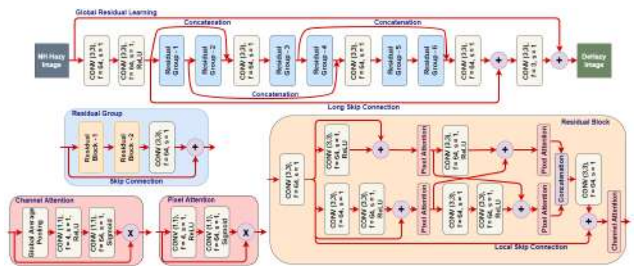


Figure 29: Architecture of the CVML proposed solution.

4.22. CVML

To tackle the non-homogeneous haze, they proposed a new appraoch called Depth-in-Residual Mulit-Path CNN for Non-Homogeneous DeHazing (i.e., DMCNN-DHaze) and the design of the same is depicted in the Figure 29. The proposed DMCNN-DHaze model consists of several residual groups (i.e., consisting depth-in-Residual blocks) where multi-path connections along with attention networks are utilized in order to remove the non-homogeneous haze and produce plausible solutions.

4.23. BUUMASRC

Their algorithm refines the estimation of the atmospheric ambient light and transmittance based on the original dark channel prior algorithm, thus get more effective estimate values, which significantly improve the dehazing effect. The flowchart diagram in the algorithm is represented in Figure 30.

The algorithm makes estimations over the ambient light and the atmospheric light using image level statistics. Those estimations are used to compute a color layer transmittance matrix, and then, this is used for the image dehazing procedure.

5. Conclusion

The challenge registered 327 participants, and 23 teams were ranked in the final phase. They experimented with various architectures and proposed several novel solutions, improving over the existing results. Designs presented in the past years were successfully deployed, showing them as

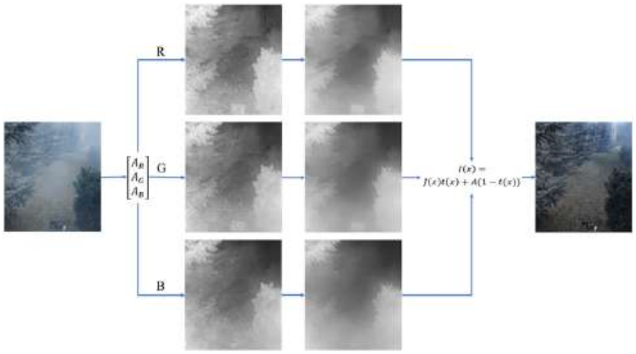


Figure 30: Flowchart diagram of the algorithm proposed by Team BUUMASRC

useful building blocks, with a lot of potential for improvement.

Acknowledgements

Part of this work was supported by the Romanian Ministry of Education and Research, CNCS UEFISCDI (PNCDI III), under Project PN-III-P1-1.1-TE-2019-1111. Part of this work was supported by European Marie Skłodowska-Curie Individual Fellowships H2020-MSCA-IF-2019.

A. Teams and Affiliations

NTIRE 2021 Team

Title:
NTIRE 2021 Challenge on NonHomogeneous 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, ETH Zurich, Switzerland

DWT-Dehaze

Title:
DW-GAN: A discret wavelet transform GAN for NonHomogenous Image dehazing
Members:
Minghan Fu[1] *(fum16@mcmaster.ca)*, Huan Liu[1], Yankun Yu[1], Jun Chen[1] and Keyan Wang[2]
Affiliations:
[1]McMaster University
[2]Xidian University

Mac Dehaze

Title:
A Simple Baseline for Non-homogeneous Dehazing via Transfer Learning
Members:
Yankun Yu[1] *(yuy142@mcmaster.ca)*, Huan Liu[1], Minghan Fu[1], Jun Chen[1], Xiyao Wang[2], Keyan Wang[3]
Affiliations:
[1] McMaster University
[2] Center for Research on Intelligent System and Engineering, Institute of Automation, University of Chinese Academy of Sciences
[3] Xidian University

NTUGICE_LINLAB

Title:
Adaptive Dehazing Network
Members:
Jerome Chang[1] *(llarry9272tw@gmail.com)*
Affiliations:
[1] Natioanal Taiwan University

buaa_colab

Title:
Super Resolution Knowledge Transfer Dehazing Network
Members:
Tianyi Chen[1] *(18374165@buaa.edu.cn)* Jiahui Fu[1], Wentao Jiang[1], Chen Gao[1], Si Liu[1]
Affiliations:
[1] Beihang University

Bilibili AI & FDU

Title:
Trident Dehazing Network
Members:
Jing Liu (liujing04@bilibili.com)[1]*, Yi Xu*[2] *(yxu17@fudan.edu.cn)*, Xinjian Zhang, Minyi Zhao, Shuigeng Zhou
Affiliations:
[1] Bilibili Inc.
[2] Fudan University

TJU_VIPLab

Title:
Multi-task Collaboration Dehazing Network
Members:
Yudong Wang (yudongwang@tju.edu.cn)[1], Jichang Guo[1], Chongyi Li[2], Qixin Yan[1] Sida Zheng[1]

Affiliations:
[1] School of Electrical and Information Engineering, Tianjin University, Tianjin, 300072, China
[2] School of Computer Science and Engineering, Nanyang Technological University (NTU), Singapore

TeamInception

Title:
Multi-Stage Progressive Image Restoration
Members:
Syed Waqas Zamir (waqas.zamir@inceptioniai.org)[1], Aditya Arora, Akshay Dudhane, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ling Shao
Affiliations:
[1] Inception Institute of Artificial Intelligence (IIAI)

VIPLab

Title:
Densenet multi-scale feature enhancement dehazing network
Members:
Wenjin Yang (836441517@qq.com)[1], Jin Lin, Xiaotong Luo, Guowen Huang, Shuxin Chen, Yanyun Qu
Affiliations:
[1] Department of Computer Science, Xiamen University

iPAL-GridFFA

Title:
GAN Networks with Grid Feature Fusion Attention
Members:
Haichuan Zhang (zhanghaichuan1998@gmail.com)[1], Tiantong Guo, Vishal Monga
Affiliations:
[1]The Pennsylvania State University, School of Electrical Engineering and Computer Science, The Information Processing and Algorithms Laboratory (iPAL)

debut_kele

Title:
Improved NonLocal Channel Attention for NonHomogeneous Image Dehazing
Members:
Kele Xu (kelele.xu@gmail.com)[1], Lehan Yang, Pengliang Sun, Xuetong Niu
Affiliations:
[1]Key Laboratory for Parallel and Distributed Processing, Changsha, China

alibaba-cipp

Title:
Progressive Image Dehazing with Transformers
Members:
Junjun Zheng[1] *(fangcheng.zjj@alibaba-inc.com)*, Xiaotong Ruan[1], Yunfeng Wang[1], Jiang Yangu[1]
Affiliations:
[1]Alibaba Inc.

Team Dou

Title:
Improved Learning Enriched Features for Real Image Dehazing
Members:
Xiaochun Cao[1] *(caoxiaochun@iie.ac.cn)*, Jun Luo, Zhuoran Zheng, Wenqi Ren, Tao Wang
Affiliations:
[1]State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing

LDGLI

Title:
Learning to Dehaze from Global and Local Information
Members:
Yiqun Chen (chenyiqun2021@ia.ac.cn), Cong Leng, Chenghua Li, Jian Cheng

NTUDS-LINLAB

Title:
Dehazing U-Net with Multiscale Dense Feature
Members:
Chang-Sung Sung[1] *(r09946014@ntu.edu.tw)*, Jun-Cheng Chen
Affiliations:
[1]National Taiwan University Data Science Degree Program

VIP_UNIST

Title:
Multi-scale Selective Residual Learning for Single Image Dehazing
Members:
Eunsung Jo[1] *(esjo93@unist.ac.kr)*, Jae-Young Sim
Affiliations:
[1]Ulsan National Institute of Science and Technology

DeepBlueAI

Title:
Cascaded Multi-Path Dehazing Network (CMPDN)
Members:
Zhipeng Luo[1] *(luozp@deepblueai.com)*, Sai Wang[1] (wangs@deepblueai.com), Zhenyu Xu[1] (xuzy@deepblueai.com)
Affiliations:
[1]DeepBlue Technology Shanghai Co.,Ltd

SP-CET

Title:
Image Dehazing using Deep Multi-patch Hierarchical Network
Members:
Geethu M M[1] *(geethumm94@gmail.com)*, Akhil K A, Dr. Sreeni K G, Prof. Jeena R S, Dr. Joseph Zacharias
Affiliations:
[1]CV Lab, Dept. of ECE, College of Engineering, Trivandrum

Dehaze_aicte

Title:
Image Dehazing using Deep Multi-patch Hierarchical Network
Members:
Chippy M Manu[1] *(1995chippy@gmail.com)*, Sreeni K G
Affiliations:
[1]AICTE University

HZZLC

Title:
V-shape Multi-patch Hierarchical Network for Nonhomogeneous Image Dehazing
Members:
Zexi Huang[1] *(1248274624@qq.com)*, Baofeng Zhang,Yiwen Zhang,Jindong Li,Mianjie Chen
Affiliations:
[1]School of Electronic and Information Engineering,South China University of Techonology

WaveFull-XM

Title:
GCAN+PAM
Members:
Quan Xiao[1] *(wavegroup167@126.com)*, Qingchao Su, Lihua Han, Yanting Huang
Affiliations:
[1]Shenzhen Wave Kingdom Co., Ltd

SVNIT_NTNU_Team

Title:
Multi-Attention based Convolutional Neural Network (MACNet) for Non-Homogeneous Dehazing
Members:
Kalpesh Prajapat[1] *(kalpesh.jp89@gmail.com)*, Vishal Chudasama, Heena Patel, Anjali Sarvaiya, Kishor Upla, Kiran Raja, Raghavendra Ramachandra, Christoph Busch
Affiliations:
[1]Sardar Vallabhbhai National Institute of Technology, India

CVML

Title:
Depth-in-Residual Multi-Path Convolutional Neural Network for Non-Homogeneous Dehazing (DMCNN-DHaze)
Members:
Vishal Chudasama[1] *(vishalchudasdama2188@gmail.com)*, Kalpesh Prajapati, Heena Patel, Anjali Sarvaiya, Kishor Upla, Kiran Raja, Raghavendra Ramachandra, Christoph Busch
Affiliations:
[1]Sardar Vallabhbhai National Institute of Technology, India

BUUMASRC

Title:
Fog removal algorithm based on regional similarity optimization
Members:
Jing Hongyuan[1] *(jqrhongyuan@buu.edu.cn)*, Huang Zilong, Fu Yiran, Wu Haoqiang, Zha Quanxing,Zhu Zhiwei, Lv Hejun

References

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

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

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

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

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

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

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

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

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

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

[11] Codruta O. Ancuti, Cosmin Ancuti, and Radu Timofte. NH-HAZE: An image dehazing benchmark with nonhomogeneous hazy and haze-free images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2020. 2

[12] 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

[13] Piotr Bojanowski, Armand Joulin, David Lopez-Paz, and Arthur Szlam. Optimizing the latent space of generative networks. *arXiv preprint arXiv:1707.05776*, 2017. 6

[14] 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

[15] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, and Gang Hua. Gated context aggregation network for image dehazing and deraining. In *2019 IEEE winter conference on applications of computer vision (WACV)*, pages 1375–1383. IEEE, 2019. 13

[16] X. Chen, C. Fu, Y. Zhao, F. Zheng, J. Song, R. Ji, and Y. Yang. Salience-guided cascaded suppression network for person re-identification. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3297–3307, 2020. 13

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

[18] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 3

[19] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In *International conference on machine learning*, pages 647–655. PMLR, 2014. 3

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

[21] S. H. Gao, M. M. Cheng, K. Zhao, X. Y. Zhang, M. H. Yang, and P. Torr. Res2net: A new multi-scale backbone architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(2):652–662, 2021. 3

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

[23] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010. 9

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 10

[25] Tong He, Zhi Zhang, Hang Zhang, Zhongyue Zhang, Junyuan Xie, and Mu Li. Bag of tricks for image classification with convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 558–567, 2019. 9

[26] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks, 2018. 10

[27] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 12

[28] Pavel Izmailov, Dmitrii Podoprikhin, Timur Garipov, Dmitry Vetrov, and Andrew Gordon Wilson. Averaging weights leads to wider optima and better generalization. *arXiv preprint arXiv:1803.05407*, 2018. 10

[29] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016. 3, 4

[30] H. Koschmieder. Theorie der horizontalen sichtweite. In *Beitrage zur Physik der freien Atmosphare*, 1924. 2

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

[32] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photorealistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 3, 4

[33] Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, and Zhuowen Tu. Deeply-supervised nets, 2014. 11

[34] Jing Liu, Haiyan Wu, Yuan Xie, Yanyun Qu, and Lizhuang Ma. Trident dehazing network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 430–431, 2020. 3, 6, 7, 8

[35] Pengju Liu, Hongzhi Zhang, Kai Zhang, Liang Lin, and Wangmeng Zuo. Multi-level wavelet-cnn for image restoration. In *Proceedings of the IEEE conference on computer*

vision and pattern recognition workshops, pages 773–782, 2018. 3

[36] 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

[37] 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

[38] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7314–7323, 2019. 9

[39] K. Metwaly, X. Li, T. Guo, and V. Monga. Nonlocal channel attention for nonhomogeneous image dehazing. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1842–1851, 2020. 10

[40] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11908–11915, 2020. 3, 5, 9

[41] 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

[42] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 3

[43] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 10

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

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

[46] J.-P. Tarel, N. Hautière, L. Caraffa, A. Cord, H. Halmaoui, and D. Gruyer. Vision enhancement in homogeneous and heterogeneous fog. *IEEE Intelligent Transportation Systems Magazine*, 2012. 2

[47] 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

[48] 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

[49] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, volume 2, pages 1398–1402 Vol.2, 2003. 4

[50] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. CBAM: convolutional block attention module. In *The European Conference on Computer Vision (ECCV)*, September 2018. 4

[51] Haiyan Wu, Jing Liu, Yuan Xie, Yanyun Qu, and Lizhuang Ma. Knowledge transfer dehazing network for nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 3, 5, 7

[52] H. Wu, J. Liu, Y. Xie, Y. Qu, and L. Ma. Knowledge transfer dehazing network for nonhomogeneous dehazing. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1975–1983, 2020. 7

[53] T. Xie, X. Yang, Y. Jia, C. Zhu, and X. LI. Adaptive densely connected single image super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3432–3440, 2019. 10

[54] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, 2021. 8, 10

[55] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014. 3

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

[57] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Multi-scale single image dehazing using perceptual pyramid deep network. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 902–911, 2018. 4

[58] Hang Zhang, Chongruo Wu, Zhongyue Zhang, Yi Zhu, Zhi Zhang, Haibin Lin, Yue Sun, Tong He, Jonas Muller, R. Manmatha, Mu Li, and Alexander Smola. Resnest: Split-attention networks. *arXiv preprint arXiv:2004.08955*, 2020. 10

[59] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 4

[60] 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

[61] Yajie Zhao, Weikai Chen, Jun Xing, Xiaoming Li, Zach Bessinger, Fuchang Liu, Wangmeng Zuo, and Ruigang Yang. Identity preserving face completion for large ocular region occlusion. *arXiv preprint arXiv:1807.08772*, 2018. 9

[62] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation, 2018. 12

[63] 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, 4