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 Jerome Chang
Xiyao Wang Jing Liu Yi Xu Xinjian Zhang Minyi Zhao Shuigeng Zhou
Tianyi Chen Jiahui Fu Wentao Jiang Chen Gao Si Liu 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 Haichuan Zhang Tiantong Guo Vishal Monga
Wenjin Yang Jin Lin Xiaotong Luo Guowen Huang Shuxin Chen Yanyun Qu
Kele Xu Lehan Yang Pengliang Sun Xuetong Niu Junjun Zheng
Xiaotong Ruan Yunfeng Wang Jiang Yang Zhipeng Luo Sai Wang Zhenyu Xu
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 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 Prajapati Vishal Chudasama Heena Patel Anjali Sarvaiya Kishor Upla
Kiran Raja Raghavendra Ramachandra Christoph Busch Hongyuan Jing
Zilong Huang Yiran Fu Haoqiang Wu Quanxing Zha Zhiwei Zhu Hejun Lv

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 [25, 55, 28, 56, 37, 9, 2, 43, 7, 10]. Recent methods using CNN [17, 51, 69, 44, 58] 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.

The first image datasets were synthesized and used information about scene depth and scene attenuation param-

*Codruta O. Ancuti (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/

eter. FRIDA [57] dataset designed for Advanced Driver Assistance Systems (ADAS) was developed using 66 synthetic ground images of various traffic scenes. D-HAZY [6] was generated using over 1400 real images and their known depth maps, by employing the Koschmieder's [36] 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 $NTIRE2018$ [3] image dehazing challenge. O-Haze [8] contains 45 outdoor images and the corresponding images affected by haze, and I-Haze [5] contains 35 indoor images and similar scenes affected by haze in a controlled way. Similarly, DENSE-HAZE [4] contains dense (homogeneous) hazy and ground-truth images and was employed by the NTIRE 2019 image dehazing challenge $NTIRE2019$ [14].

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 [11] dataset that was used in the NTIRE 2020 image dehazing challenge [12]. 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.

This challenge is one of the NTIRE 2021 associated challenges: nonhomogeneous dehazing [13], defocus deblurring using dual-pixel [1], depth guided image relighting [24], image deblurring [48], multi-modal aerial view imagery classification [40], learning the super-resolution space [46], quality enhancement of heavily compressed videos [65], video super-resolution [54], perceptual image quality assessment [27], burst super-resolution [15], high dynamic range [49].

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 [11] 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 [11] 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 (18$percent$ 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. 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 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. For the final ranking, results were analyzed in both fidelity an perceptual terms. Finally, the Mean Opinion Score (MOS) was used to differentiate between similar results in terms of both fidelity and perceptual metrics.

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. Section 4 offers a description for each of the solutions ranked in the final phase of the challenge.

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.

As you can observe in Table 1, The metric with the highest correlation to the Mean Opinion Score (MOS) is the PSNR, while LPIPS and SSIM can be used to differentiate similar results. However, the results corresponding to the top performing solutions in terms of perceptual metrics have, as expected, high SSSIM values and low LPIPS distance.

Architectures and main ideas

Excepting *Team BUUMASRC*, all the remaining solutions used end-to-end deep learning, employing GPU(s) for both training and inference. Table 1 can be used as reference point when comparing solutions complexity, as the inference time was provided for the majority of the ranked solutions.

Team BUUMASRC proposed an algorithm based on a light scattering model to estimate a dehazed image, using image level statistics and physical models for various mechanisms. Their algorithm offers a better estimation over the atmospheric light and participating media transmittance, based on the dark channel prior.

Ideas similar to ensemble learning were deployed by the majority of the top scoring teams, to reduce the level of variance produced by the limited amount of data. Many teams used branched structure to achieve a better restoration of the high-level details. Solutions developed in the adversarial framework were proposed, one of them being the *Adaptive Dehazing Network*, proposed by the challenge winner, *Team DWT dehaze*.

In terms of minimized objectives, the majority of the teams used the $L1$ loss, and a SSIM based loss, as those metrics were used for the public leaderboard provided on the challenge website. Perceptual losses employing pretrained feature extractors were also widely used by challenge participants. Similar to the last year challenge, they also deployed losses based on Fast Fourier Transform (FFT) or knowledge-transfer losses. Gradient or laplacian losses, or metrics defined in the *Lab* color space were successfully employed by some participants, to improve the quality of their results in the high-frequency domain.

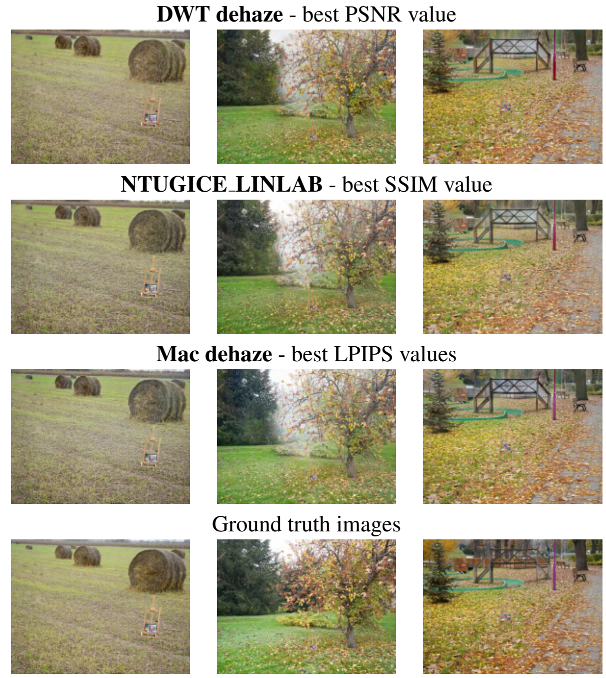


Figure 1: Visual results provided for best performing method on each of the metrics deployed. Best zoom-in on screen for a better view.

Some of the other ideas that will be encountered when reading the Section 4 are the usage of attention structures, where the majority of the teams came with existing methods or proposed novel designs, the multi-scale extracted features and residual learning. The *Trident Dehazing Network (TDN)*, the winner of the 2020 competition, served as one of the building blocks for many of the ranked solutions, and for some of the top scoring teams.

4. Challenge Methods

4.1. DWT dehaze

Inspired by [42, 50, 38], 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 supervising the training process by a frequency domain loss [41], they proposed the idea of directly embedding the frequency domain knowledge into the de-

Participant		Results					Solution details					
		Fidelity		Perceptual quality								
Team	User	PSNR↑	SSIM↑	LPIPS_1 ↓	LPIPS_2 ↓	**MOS**↓	Runtime img.[s]	GPU/ CPU	extra data	ens.	deep learning framework	loss
Top perceptual quality solutions												
DWT dehaze	eason97	21.076_1	0.839_3	0.169	0.203	1	1.558	1080Ti	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	2	60	Tesla V100	NH-HAZE 20	8x	Pytorch	n/a
Mac dehaze	ken103	21.018_2	0.837_5	0.168	0.196	3	0.089	Tesla V100	NH-HAZE-20	-	Pytorch	$L1, L_{SSIM}, L_{GAN}, L_{perc}$
Bilibili AI & FDU	splinter23	20.598_5	0.823_{11}	0.182	0.212	4	0.64	1080Ti	NH-HAZE 20	8x	Pytorch	$L1, L_{FFT}, L_{BReLU}$
buaa_colab	buaa_colab	20.619_4	0.834_7	0.202	0.220	5	1.77	4x RTX2080Ti	-	8x	Pytorch	$L1, L_{Lab}, L_{Laplacian}, L_{KT}$
TJUVIPLab	WangYudong	20.537_6	0.835_6	0.183	0.205	6	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	7	1.4	4x Tesla V100	-	8x	Pytorch	$L1, L_{SSIM}, L_{VGG}$
iPAL-GridFFA	haichuan	19.567_{12}	0.839_2	0.178	0.194	8	1.01	RTX2080Ti	NH-HAZE		Pytorch	$L1$, smoothed $L1, L_{SSIM}, L_{GAN}$
VIPLab	Yangwj	19.675_{10}	0.824_{10}	0.173	0.203	9	0.042	1080Ti	NH-HAZE	-	Pytorch	$L1, L_{perc}$
Medium perceptual quality solutions												
debut_kele	debut kele	20.264_7	0.832_9	0.200	0.219	10	n/a	RTX2080Ti	-	-	Pytorch	$L1, L_{SSIM}, L_{std}$
alibaba-cipp	alibaba-cipp	20.231_8	0.802_{16}	0.178	0.220	10	9.0	8x Tesla V100	Place2, O-HAZE, DENSE-HAZE	-	Pytorch	n/a
DeepBlueAI	DeepBlueAI	18.970_{17}	0.816_{13}	0.197	0.210	10	1.0	4x Tesla V100	-	-	Pytorch	$L1, L_{perc.}, L2$
team_Dou	xiaodou	19.654_{11}	0.812_{14}	0.187	0.208	11	0.94	GPU	-	-	Pytorch	n/a
LDGLI	YiqunChen1999	19.522_{13}	0.838_4	0.192	0.207	11	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	11	n/a	GPU	NH-HAZE	n/a	Pytorch	n/a
VIP_UNIST	Eun-Sung	19.156_{15}	0.809_{15}	0.205	0.227	11	0.034	Titan RTX	I-HAZE O-HAZE DENSE-HAZE NH-HAZE	-	Pytorch	n/a
Low perceptual quality solutions												
SP-CET	Geethu	19.050_{16}	0.800_{17}	0.191	0.222	12	0.409	GPU	-	-	n/a	n/a
Dehaze_aicte	CHIPPYMMANU	18.302_{18}	0.733_{22}	0.295	0.309	12	1.0	GPU	-	-	Keras	n/a
HZLLC	BFZhang	18.043_{19}	0.742_{21}	0.313	0.295	12	0.018	RTX2060	-	-	Pytorch	$L2, L_{perc.}, L_{t.v.}$
WaveFull_XM	R0use	17.974_{20}	0.771_{20}	0.271	0.286	13	10.4	Titan Xp	-	-	Pytorch	n/a
SVNIT_NTNU_Team	kalpesh svnit	17.905_{21}	0.788_{18}	0.248	0.264	13	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	13	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	13	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 [72], 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 Mean Opinion Score (MOS) was added to determine the final ranking of the challenge. The results were split into three categories, with respect to their perceptual properties. Note that the perceptual differences can be rather subtle when comparing results from top scoring teams.

hazing network. They follow the U-Net [52] architecture to construct the first branch, as the wavelet net. It has a encoder that is linked to the decoder by massive skip connections. To meet the requirements for extracting frequency domain knowledge, they adopt five DWT downsampling modules and six convolutional downsampling layers to build the encoder. Then, the spatial and frequency representations are concatenated as the input of the downsampling process.

In the second branch, they use Res2Net [52] as encoder. Observing that the feature representations learned on a pretrained task can have positive impact on the target task [23, 67], they use the ImageNet [22] pretrained weights as initialization.

In the decoder module, they used pixel-shuffle layer for upsampling, which makes the size of the recovered feature maps to gradually increase to the original resolution. Channel and pixelwise 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 their work aims to balance the model behaviour, DW-GAN learning to generate low distortion and high perceptual quality images.

Therefore, they introduced the final loss blend function as stated in Equation 1, where $\alpha = 0.2, \beta = 0.001$ and $\gamma = 0.005$ are the weights for each of the loss functions. L_1 denotes L1 loss, L_{SSIM} represents MS-SSIM loss [59], $L_{perceptual}$ is perceptual loss [35] and, for the adversarial loss L_{adv}, they adopt the discriminator in [76].

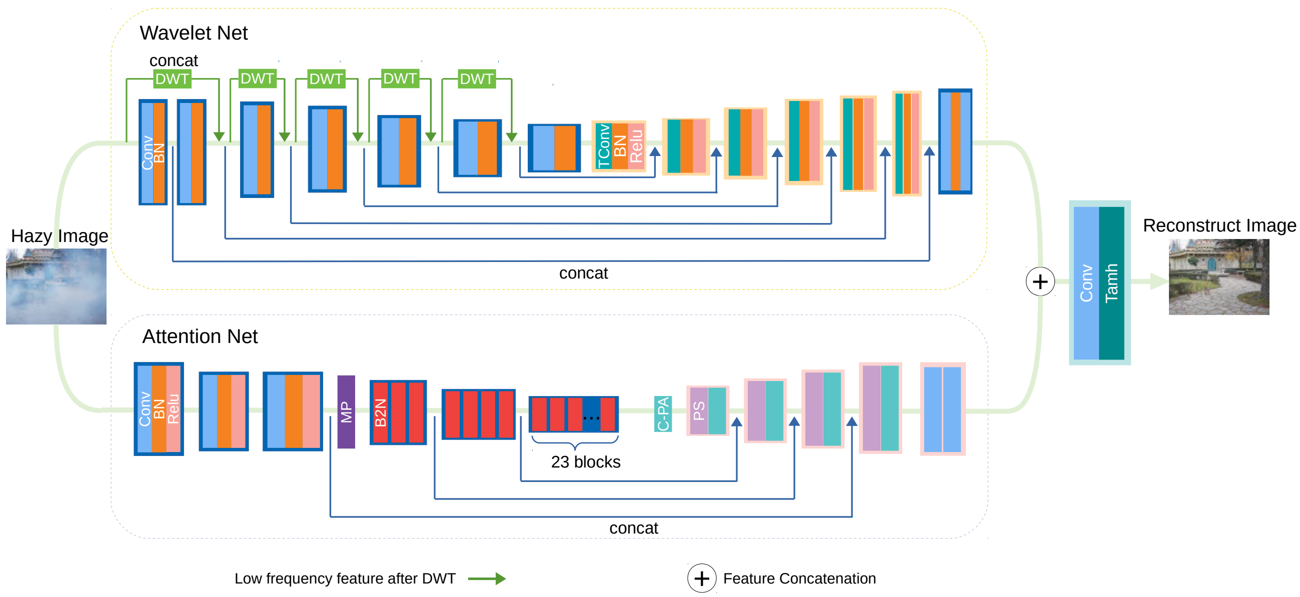
$$L_{total} = L_1 + \alpha L_{SSIM} + \beta L_{perceptual} + \gamma_4 L_{adv} \quad (1)$$

The overall network architecture is shown in Figure 2.

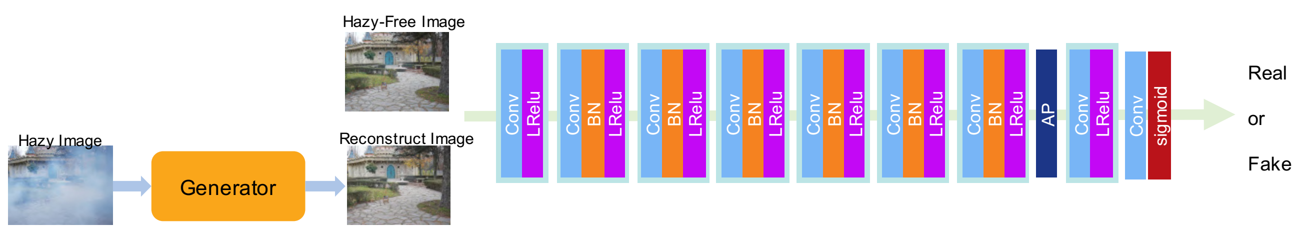
4.2. NTUGICE LINLAB

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 process the region covered by thin haze or heavy haze. As shown in the Figure 3, ADN consists of two branches, the Primary Branch and the Enhanced Branch. While the Primary Branch manage the region covered by thin haze, 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, they design the Weight Map Generator. This has the role of generating a two-channel weight map used to blend the outputs of the two branches. So, the output of each branch will



(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.

be element-wise multiplied with their corresponding weight map and the final result will be produced as the sum of the weighted outputs.

The model of the encoder-decoder structure of the primary branch is based on the Perceptual Pyramid Deep Network [70]. Both branches share the same encoder, but they own their individual decoders. The difference between the decoders is that normal convolution was replaced by the dilated convolution kernel, attempting to enlarge receptive field of the enhanced decoder. This enables the Enhanced Branch to gain ability to deal with heavy haze. Moreover, some attention modules such as CBAM[61] were added, combining spatial attention and channel attention to let the decoders concentrate the training procedure on the most important features extracted by the encoder.

4.3. Mac dehaze

Mac dehaze team proposes a two-branch neural network for non-nomogeneous dehazing via ensemble learning to deal with the above mentioned problems. The structure diagram of the network is shown in Figure 4.

The first branch, namely the transfer learning sub-net, is built upon a ImageNet [21] pretrained Res2Net[26] [62]. It aims to extract robust global representations from input

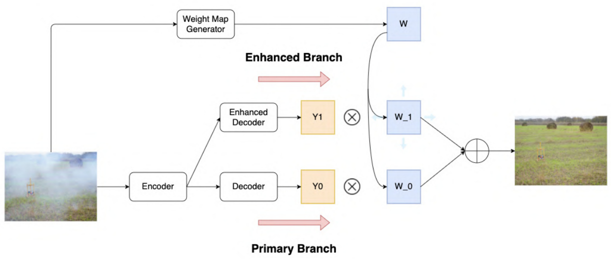


Figure 3: Information flow along NTUGICE-LINLAB proposed model.

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. This 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 this, in favor of the strong mapping capability of residual channel attention network (RCAN) [73], they designed the current data fitting sub-net using RCAN as second branch. The current data fitting branch has five residual groups, and each group has ten residual blocks. Unlike the original network setting [73] that does the downsampling of the input images, the second branch always maintains the original resolution of the inputs and avoids using any downsampling operation. This adjustment avoids the loss of fine detailed features. Since the sub-network is trained 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 a fusion layer. Specifically, the fusion layer takes the concatenation of features from the branches and then maps the features to clear outputs.

Moreover, adversarial loss is proved to be effective in helping restore photo-realistic images [38]. Especially for the small-scaled dataset, the pixelwise loss function usually fails to provide sufficient control to supervise the network training for recovering the photo-realistic details.

Therefore, they implemented the adversarial loss with the discriminator in [76]. The overall loss function is a linear combination of smooth L1 loss L_{l1}, MS-SSIM loss[60] L_{SSIM}, perceptual loss[35] $L_{perceptual}$, and adversarial loss L_{adv}, as shown in Equation 2.

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

4.4. Bilibili AI & FDU

They use the Trident Dehazing Network[41] proposed in NTIRE2020 NH-Dehazing challenge as their model. The architecture is depicted in the Figure 6. 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), as the training procedure continues.

4.5. buaa_colab

Their contribution is the modified version of Knowledge Transfer Network [62], namely, the Super Resolution Knowledge Transfer Dehazing Network (SRKTDN). As is shown in Figure 7, the network described contains two main components, the main network and teacher network. The main network consists of a dehaze network and a super-resolution network.

The dehaze network uses Res2Net101 as encoder, and PixelShuffle for the upsampling operation. The network uses an attention mechanism combining channel attention blocks and pixel attention blocks to restore the haze-free image [50].

They used a teacher network to generate low-level feature maps. The teacher network is trained by ground truth pairs of the dataset, in order to capture the necessary information for image restoring. Compared to the Knowledge Transfer Network, there are structural differences between the teacher network and the dehaze network. While the dehaze network uses Res2Net101 as encoder to ensure capability of haze removal, the teacher network uses ResNet18 as the encoder. This will further enhance the generalization ability and reduce the training time and GPU memory consumption.

Meanwhile, inspired by TDN[41], 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.

$$L_1 = |J - M(I)|_1 \quad (3)$$

Laplacian loss uses Laplacian pyramid representation of the image and calculates L1 loss for 5 levels[16]. $L^j(\cdot)$ in 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.

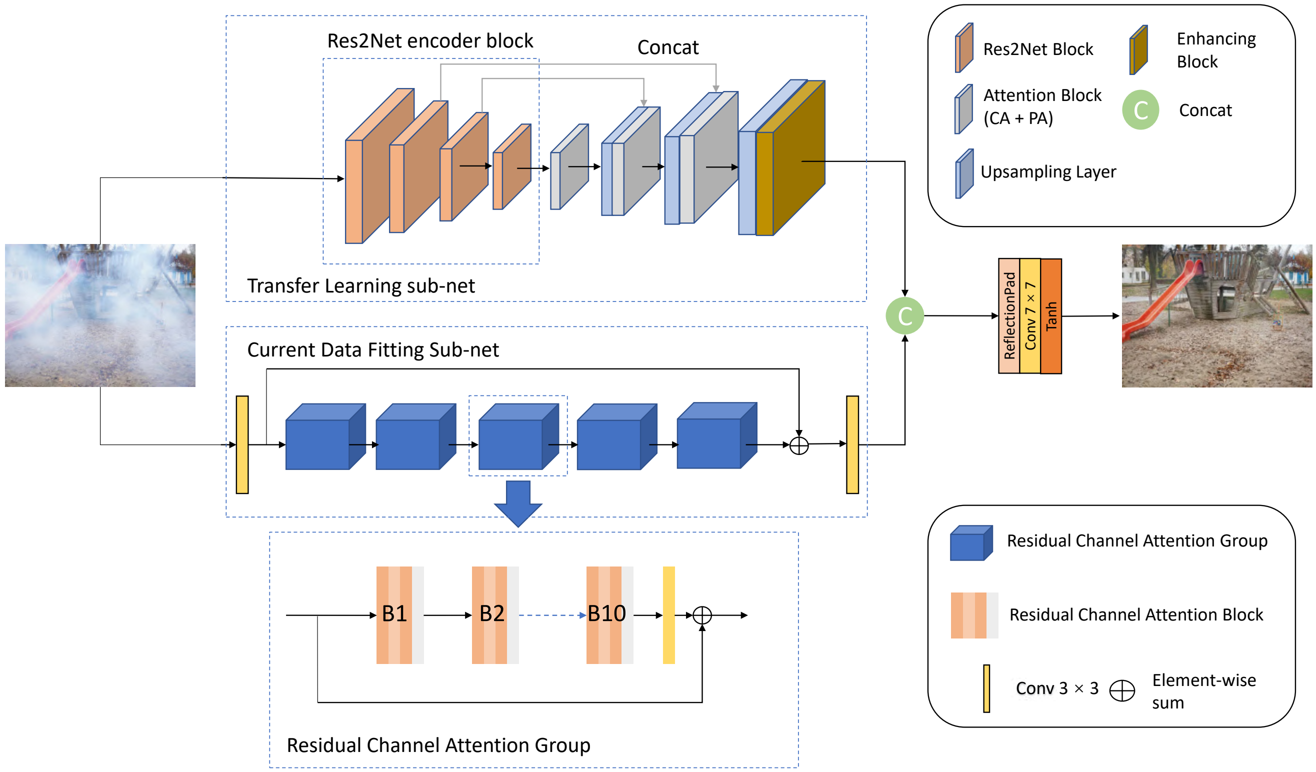


Figure 4: Illustration of Mac dehaze model architecture.

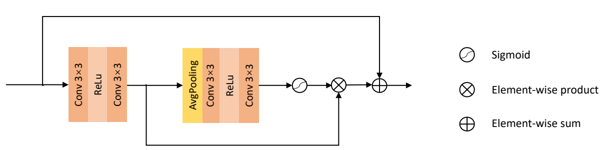


Figure 5: Residual Channel Attention Block

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

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.

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

Similar to the method proposed in [62], the Knowledge Transfer loss is L1 loss between feature map of dehaze network and the output of the teacher network. Knowledge Transfer loss helps the Res2Net101 encoder to imitate

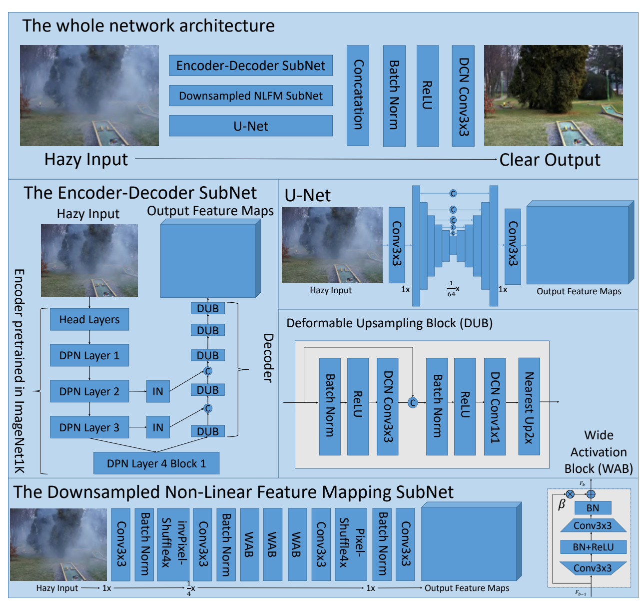


Figure 6: The network architecture of the solution proposed by team Bilibili AI & FDU

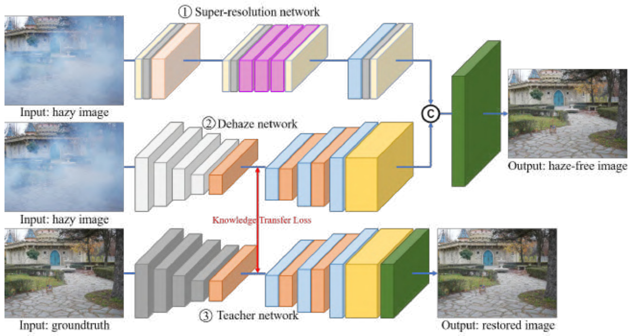


Figure 7: Architecture of buaa_colab proposed architecture.

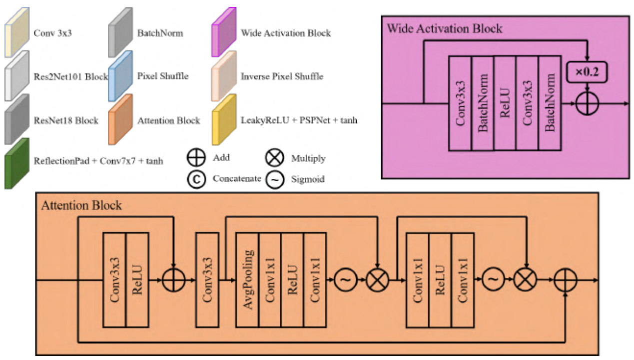


Figure 8: Details over the attention modules used by buaa_colab team.

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.

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

4.6. TJUVIPLab

TJU_VIPLab team proposed a CNN-based Multi-task Collaboration Dehazing Network (MCDNet) to directly learn the mapping between the nonhomogeneous haze image and haze-free clear image. MCDNet consists of three sub-nets inspired by [41] and a Channel Attention-Spacial Attention(CASA) Module inspired by [63]. The overall structure of MCDNet is shown in Figure 9. The Simple U-Net is used to obtain a preliminary haze-free image, the Encode-Decode sub-Net(EDN) is used to extract features and get basic dehazing feature maps, 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 the usage of available information to improve the perceptual properties of the produced image.

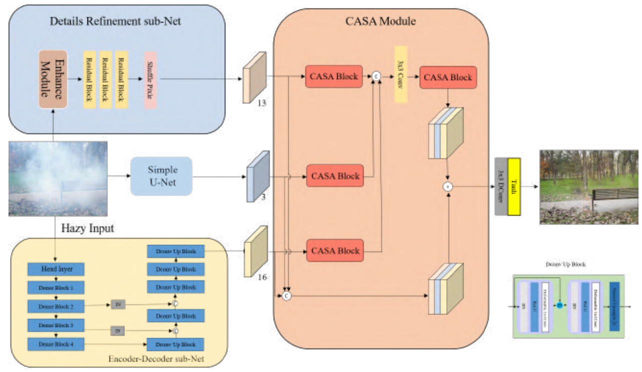


Figure 9: An overview of the proposed MCDNet architecture.

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

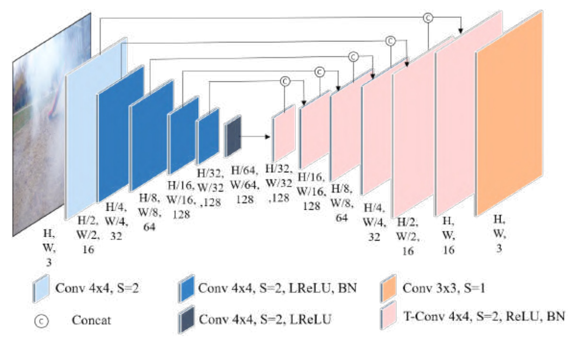


Figure 10: Structure of Simple U-Net

DenseNet-101 pretrained on the ImageNet is the backbone of EDN's encoder part. Same as [41], 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 $2\times$ nearest-upsampling layer to obtain the upsampled feature. The deepest two blocks used skip connection from the output of the third and the fourth dense-block, respectively. Moreover, EDN uses trainable instance normalization for skip connections.

DRN starts with two downsampled enhancing models (EM) to capture multi-scale detailed feature maps. Then, their output is fed into three residual blocks. The Pixel Shuffle layer implements a $2\times$ upsampling operation, which is used to change the feature maps from $H \times W \times 4C$ to $2H \times 2W \times C$, where H, W, C are the height, width and the number of channels of feature map. As shown in Figure 11, EM obtains $4\times$, $8\times$, $16\times$, $32\times$ downsampling features, and performs $2\times$, $4\times$, $8\times$, $16\times$ upsampling respectively. The feature maps are concatenated with $2\times$ downsampling and fed into 3×3 convolution layer.

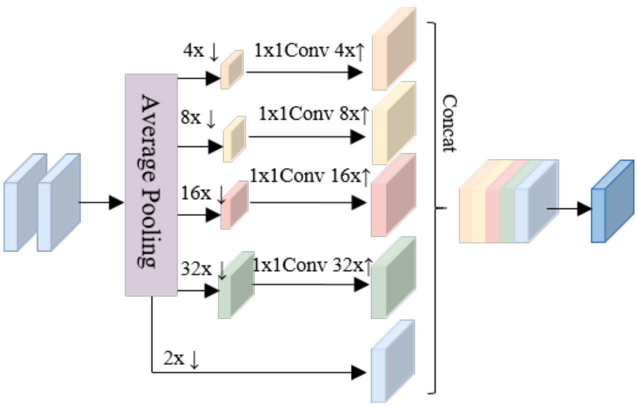


Figure 11: Structure of Enhance Module

CASA Module contains four CASA-blocks, which is shown at the bottom of Figure 9. Three Sub-Net outputs are fed into three CASA-blocks respectively. Then, their outputs are concatenated and used as input of the next CASA-block, which can further enhance useful information. The CASA Module output adds with concatenated feature map consisting of the output of three Sub-Nets. Finally, this feature map is fed into a 3×3 DConv layer. This layer uses $Tanh$ as activation function, which normalizes the output in the $[-1, 1]$ interval.

4.7. Team Inception

They present an architecture, named MPRNet, that is based on a recent work [66]. As illustrated in Fig. 12, MPRNet consists of two stages to progressively restore images. 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) [66]. 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.

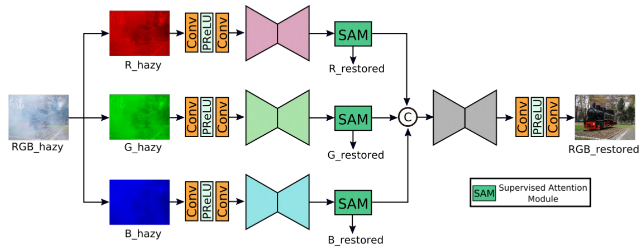


Figure 12: Overall framework of MPRNet.

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

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.

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

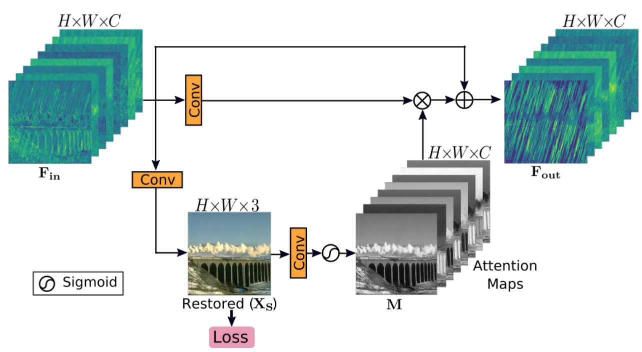


Figure 13: Supervised attention module (SAM).

4.8. 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 architecture, they chose a 3×6 Grid network with Feature Fusion Attention. The generator network is an enhanced network of GridDehazeNet [45].

The Group Structure combines 15 Basic Block structures which conclude the *Pixel attention* [50] and *Channel attention* [29], with skip connections for each of the modules.

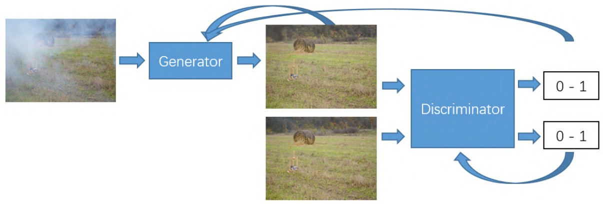


Figure 14: Solution proposed by iPAL-GridFFA

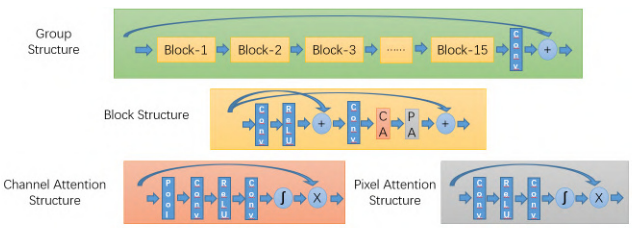


Figure 15: The architecture of the group structure proposed by iPAL-GridFFA.

For the discriminator architecture, they use a similar idea to *Patch GAN*[74], 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 [31] mechanism is used for the adjustment of the learning rate.

Figure 15 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 three building blocks in serial, where the first two blocks are attached to a Max Pooling operation.

4.9. VIPLab

Densenet network has a wide range of applications in many fields due to its dense connection characteristics, and so, this team used 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.

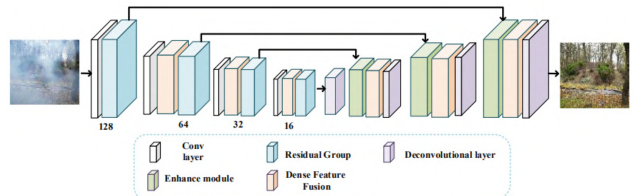


Figure 16: VIPLab proposed architecture.

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.

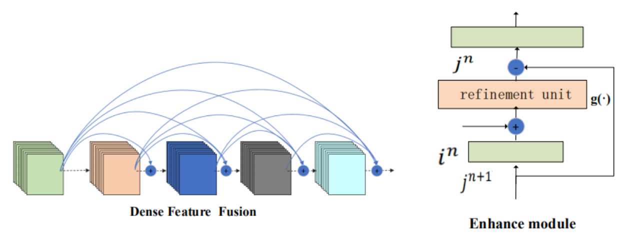


Figure 17: Dense Feature Fusion

4.10. debut_kele

They proposed a deep learning architecture, similar to [47], that estimates physical parameters in the haze model. Compared to it, they experiment with different data augmentation strategies, a custom loss function, and the Stochastic Weight Averaging optimization [34]. 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. A novel Dilation Inception module at the direct decoder is used to generate additional features at densely-hazed regions using the non-local features principle.

The minimized objective consists of a final blend 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 a sequence of two stages to progressively dehaze the input hazy images.

In the first stage, they use a residual-in-residual dense block (RRDB)[64] as the basic module to generate the coarse dehazed image. The second stage, they employ 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. 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. This method consists of randomly initializing two fog masks with a total area of 64×64 and merging them with the sub-images, in order to solve the problem of insufficient training for non-haze area/shallow haze area.

Figure 19 shows the effect of HRA on the dehazing results. The left and right columns are the compared results produced before and after using HRA. HRA effectively removes dense haze and maintains the original texture of the image, making the result clearer.

4.13. Team Dou

Team Dou proposed an improvement over the work published in [66], based on muti-scale features extraction. Principles as attention mechanisms, residual learning, feature fusion and hybrid dilated convolution are combined in an architecture illustrated in Figure 20.

4.14. LDGLI

The architecture is illustrated in the Figure 21. They used a pre-trained ResNeSt [71] model to extract the features at five different levels, and employed the proposed NonHomogeneous Dehazing Block (NHDBlock) (see Figure 22) to remove the haze and recover the image. The $2\times$ is an upsampling operation which is done by a transposed convolution and a nonlinear activation.

The NHDBlock, 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.

4.15. NTUDS-LINLAB

They proposed a U-Net architecture [53] (see Figure 23) dehazing model using multiscale dense features, based on dense blocks [32] and residual blocks [30]. Their Encoder module used Densenet which was pretrained on ImageNet dataset. One important difference between their model and U-Net is the re-designed skip connection. Aiming at utilizing lower level feature maps, they used a concatenation between the decoder feature map and the upsampled lower dimension feature map.

4.16. VIP_UNIST

They proposed an end-to-end dehazing method named Selective Residual Learning for Multi-scale Dehazing. Overall network architecture (see Figure 24) shows the multiscale inputs and outputs and the use of proposed selective residual blocks.

Firstly, 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.

Secondly, the selective residual block reduces unnecessary artifacts of the final outputs. The selective residual block is an operation that is similar to the residual block in the ResNet.

However, the final output $O(\mathbf{x})$ at the pixel location $\mathbf{x}$ is the activated weighted sum of the input feature $F(\mathbf{x})$ and the estimated residual feature $R(\mathbf{x})$, 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.

$$O(\mathbf{x}) = \sigma(\alpha F(\mathbf{x}) + \beta R(\mathbf{x})), \tag{11}$$

4.17. SP-CET

This method includes a multi-level CNN model called Deep Multi-patch Hierarchical Network(DMPHN) inspired by [20] and [68]. 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.

4.18. Dehaze_aicte

This team proposed the GANID method, tackling the image dehazing problem in the adversarial learning framework. Deep supervision [39] in UNet++ is used for the generator (see Figure 26), to create secondary output maps, which allows for models to be pruned, therefore, applying the model pruning process.

Deep supervision operates in two modes, namely the accurate mode and the fast mode. In the accurate mode, the averaged output is calculated from all output branches. In

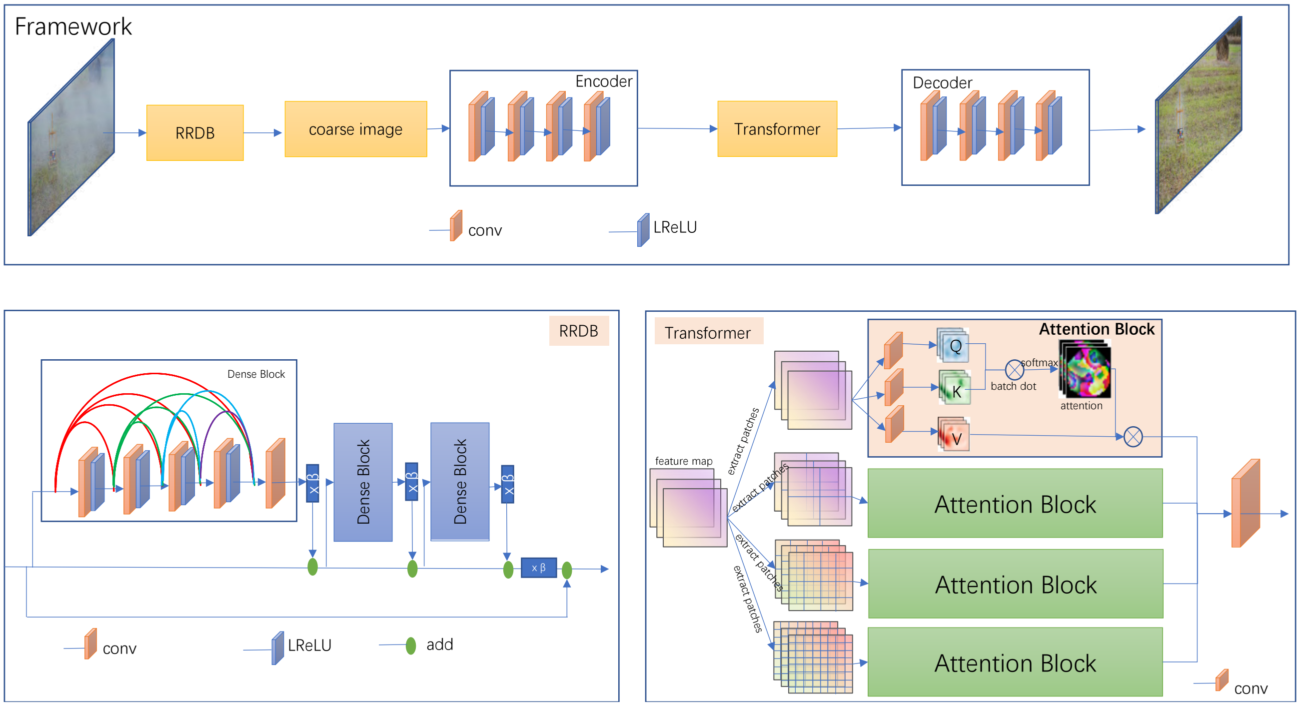


Figure 18: The framework proposed by Team alibaba-cipp.

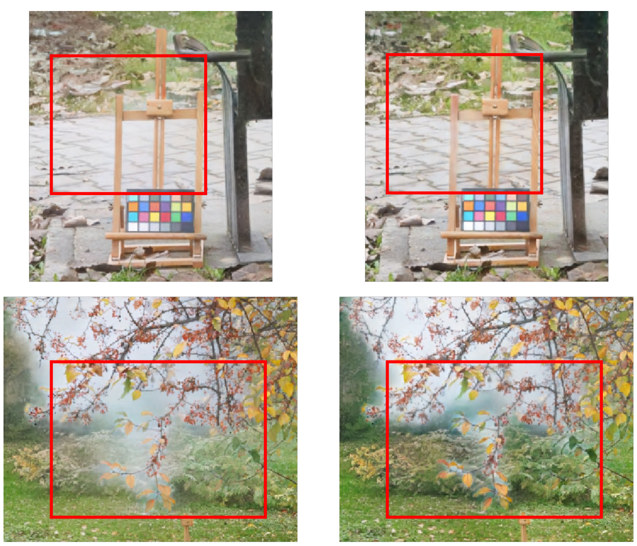


Figure 19: The effect of Hazing Reinforcement Algorithm.

the fast mode, one of the output branches is selected for the final response map. The fast mode 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

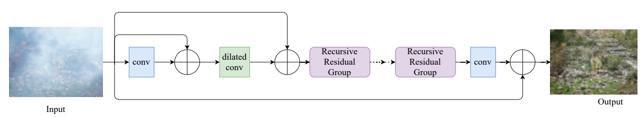


Figure 20: Architecture of the solution proposed by Team Dou.

$k = 0$ and $l = 1, 2, 3, 4$ are passed through a 1×1 convolution along with a k kernel, followed by an activation function (sigmoid). A detailed description of UNet++ is given [75]. Patch discriminator in the Conditional GAN [33] 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 completed.As to the bottom-top flow,the third output of

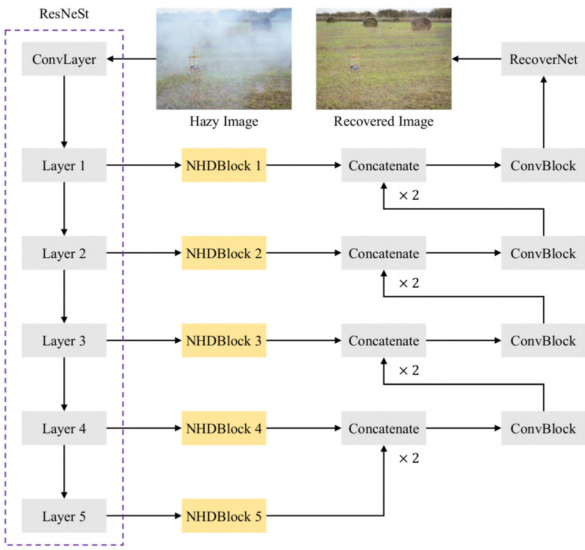


Figure 21: Architecture of the solution proposed by LDGLI team.

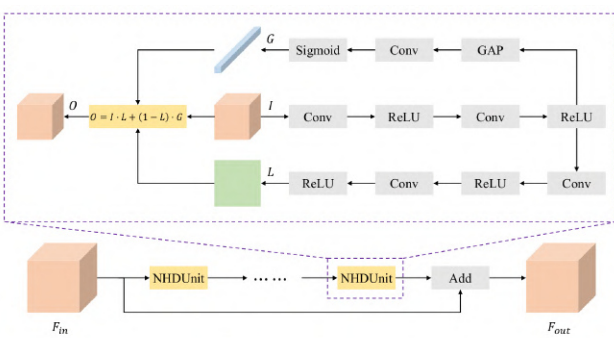


Figure 22: Schematic illustration of the NHDBlock used by LDGLI team.

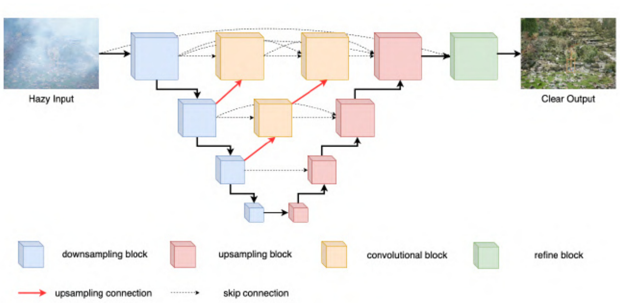


Figure 23: The architecture of the model used by NTUDS-LINLAB team.

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

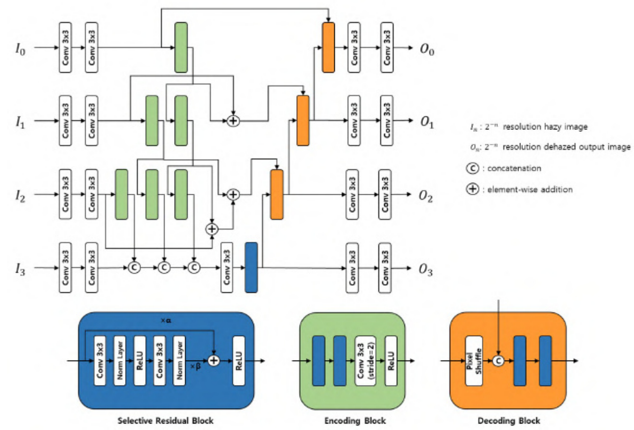


Figure 24: The architecture of the VIP_UNIST proposed method.

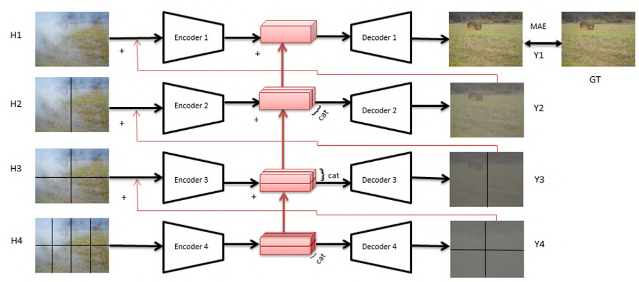


Figure 25: Architecture of the DMPHN model.

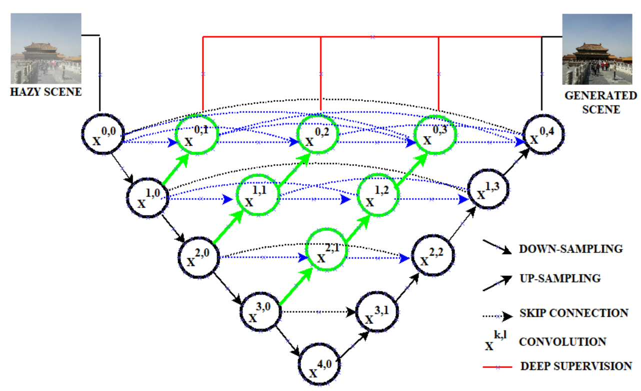


Figure 26: Generator of the GANID method.

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 [18] and the PAM model [19] to build a network implementing residual learn-

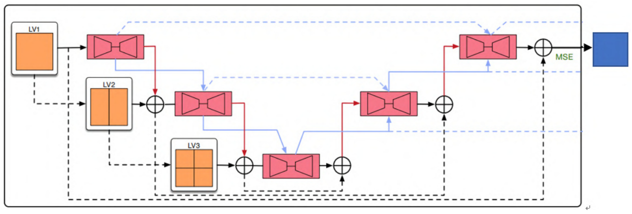


Figure 27: VMPHN model architecture.

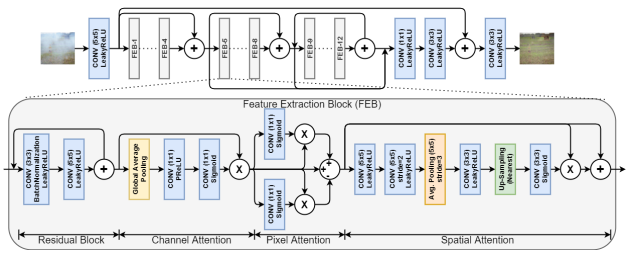


Figure 28: Architecture of the MACNet model.

ing, 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, able 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 uses channel attention, pixel attention and spatial attention, helping the network to learn the statistical characteristics of haze image. The $L1$ loss function, between the hallucinated image and the ground truth haze-free image was used as the minimized objective.

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

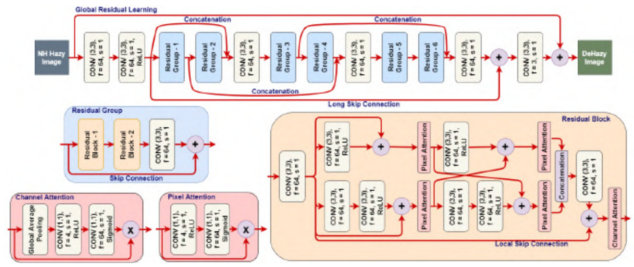


Figure 29: Architecture of the CVML proposed solution.

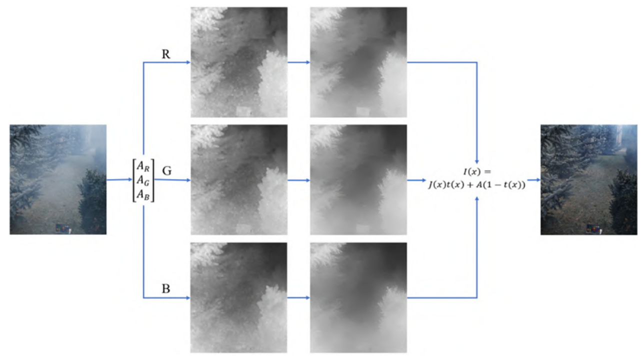


Figure 30: Flowchart diagram of the algorithm proposed by Team BUUMASRC

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

The final ranking was done with respect to the Mean Opinion Score resulting of our user study, and the solutions were split into three categories with respect to their perceptual properties. Finally, the ranking was highly influenced by the recovered images fidelity, as this had the highest correlation to the users feedback about the presented results.

Acknowledgements

We thank the NTIRE 2021 sponsors: Huawei, Facebook Reality Labs, Wright Brothers Institute, MediaTek, and ETH Zurich (Computer Vision Lab). 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

NTUGICE_LINLAB

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

Mac Dehaze

Title:
A Two-branch Neural Network for Non-homogeneous Dehazing via Ensemble 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

Bilibili AI & FDU

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

TJU_VIPLab

Title:
Multi-task Collaboration Dehazing Network
Members:
Yudong Wang[1] *(yudongwang@tju.edu.cn)*, 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[1] *(waqas.zamir@inceptioniai.org)*, Aditya Arora, Akshay Dudhane, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ling Shao
Affiliations:
[1] Inception Institute of Artificial Intelligence (IIAI)

iPAL-GridFFA

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

VIPLab

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

debut_kele

Title:
Improved NonLocal Channel Attention for NonHomogeneous Image Dehazing
Members:
Kele Xu[1] *(kelele.xu@gmail.com)*, 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 Yang[1]
Affiliations:
[1]Alibaba Inc.

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

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[1] *(chenyiqun2021@ia.ac.cn)*, Cong Leng[2], Chenghua Li[1], Jian Cheng[2]
Affiliations:
[1]CASIA
[2]AiRiA, CASIA

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

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, Jeena R S, Joseph Zacharias
Affiliations:
[1]CV Lab, Dept. of ECE, College of Engineering, Trivandrum

Dehaze_aicte

Title:
GANID: A Novel Generative Adversarial Network for Image Dehazing
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 Prajapati[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:
Hongyuan Jing[1]*(jqrhongyuan@buu.edu.cn)* Zilong Huang[1], Yiran Fu[1], Haoqiang Wu[1], Quanxing Zha[1], Zhiwei Zhu[1] and Hejun Lv[1]
Affiliations:
[1] College of Robotics, Beijing Union University

References

[1] Abdullah Abuolaim, Radu Timofte, Michael S Brown, et al. NTIRE 2021 challenge for defocus deblurring using dual-pixel images: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 2

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

[3] 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

[4] 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

[5] 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

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

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

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

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

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

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

[12] Codruta O. Ancuti, Cosmin Ancuti, 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

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

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

[15] Goutam Bhat, Martin Danelljan, Radu Timofte, et al. NTIRE 2021 challenge on burst super-resolution: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 2

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

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

[18] 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

[19] 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

[20] S. D. Das and S. Dutta. Fast deep multi-patch hierarchical network for nonhomogeneous image dehazing. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1994–2001, 2020. 11

[21] 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. 5

[22] 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. 4

[23] 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. 4

[24] Majed El Helou, Ruofan Zhou, Sabine Süsstrunk, Radu Timofte, et al. NTIRE 2021 depth guided image relighting challenge. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 2

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

[26] 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. 5

[27] Jinjin Gu, Haoming Cai, Chao Dong, Jimmy S. Ren, Yu Qiao, Shuhang Gu, Radu Timofte, et al. NTIRE 2021 challenge on perceptual image quality assessment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 2

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

[29] 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

[30] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 11

[31] 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. 10

[32] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks, 2018. 11

[33] 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

[34] 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

[35] 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. 4, 6

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

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

[38] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic 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, 6

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

[40] Jerrick Liu, Oliver Nina, Radu Timofte, et al. NTIRE 2021 multi-modal aerial view object classification challenge. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 2

[41] 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, 8

[42] 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

[43] 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

[44] 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

[45] 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

[46] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. NTIRE 2021 learning the super-resolution space challenge. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 2

[47] 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

[48] Seungjun Nah, Sanghyun Son, Suyoung Lee, Radu Timofte, Kyoung Mu Lee, et al. NTIRE 2021 challenge on image deblurring. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 2

[49] Eduardo Pérez-Pellitero, Sibi Catley-Chandar, Aleš Leonardis, Radu Timofte, et al. NTIRE 2021 challenge on high dynamic range imaging: Dataset, methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 2

[50] 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, 6, 9

[51] 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

[52] 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. 4

[53] 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. 11

[54] Sanghyun Son, Suyoung Lee, Seungjun Nah, Radu Timofte, Kyoung Mu Lee, et al. NTIRE 2021 challenge on video super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 2

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

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

[57] 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

[58] 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

[59] 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. 4

[60] 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. 6

[61] 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. 5

[62] 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. 5, 6, 7

[63] 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. 8

[64] 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. 11

[65] Ren Yang, Radu Timofte, et al. NTIRE 2021 challenge on quality enhancement of compressed video: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 2

[66] 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. 9, 11

[67] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014. 4

[68] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. *CoRR*, abs/1904.03468, 2019. 11

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

[70] 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. 5

[71] 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. 11

[72] 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

[73] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *The European Conference on Computer Vision (ECCV)*, September 2018. 6

[74] 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. 10

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

[76] 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. 4, 6