NTIRE 2019 Image Dehazing Challenge Report

Codruta O. Ancuti Cosmin Ancuti Radu Timofte Luc Van Gool Lei Zhang
Ming-Hsuan Yang Tiantong Guo Xuelu Li Venkateswararao Cherukuri
Vishal Monga Hao Jiang Siyuan Yang Yan Liu Xiaochao Qu Pengfei Wan
Dongwon Park Se Young Chun Ming Hong Jinying Huang Yizi Chen
Shuxin Chen Bomin Wang Pablo Navarrete Michelini Hanwen Liu Dan Zhu
Jing Liu Sanchayan Santra Ranjan Mondal Bhabatosh Chanda Peter Morales
Tzofi Klinghoffer Le Manh Quan Yong-Guk Kim Xiao Liang Runde Li
Jinshan Pan Jinhui Tang Kuldeep Purohit Maitreya Suin A. N. Rajagopalan
Raimondo Schettini Simone Bianco Flavio Piccoli C. Cusano Luigi Celona
Sunhee Hwang Yu Seung Ma Hyeran Byun Subrahmanyam Murala
Akshay Dudhane Harsh Aulakh Tianxiang Zheng Tao Zhang Weining Qin
Runnan Zhou Shanhu Wang Jean-Philippe Tarel Chuansheng Wang Jiawei Wu

Abstract

This paper reviews the second NTIRE challenge on image dehazing (restoration of rich details in hazy image) with focus on proposed solutions and results. The training data consists from 55 hazy images (with dense haze generated in an indoor or outdoor environment) and their corresponding ground truth (haze-free) images of the same scene. The dense haze has been produced using a professional haze/fog generator that imitates the real conditions of haze scenes. The evaluation consists from the comparison of the dehazed images with the ground truth images. The dehazing process was learnable through provided pairs of haze-free and hazy train images. There were ~ 270 *registered participants and 23 teams competed in the final testing phase. They gauge the state-of-the-art in image dehazing.*

1. Introduction

Haze is a common atmospheric phenomenon produced by small floating particles that reduce the visibility of distant objects due to light scattering and attenuation. This results in a loss of local contrast for distant objects, in the addition of noise to the image, and in a selective attenuation of the light spectrum. Image dehazing is a challenging ill-posed problem that has drawn a significant attention in the last few years.

In the last decade a significant amount of literature focused on single image dehazing research. The performance of the top methods continuously improved [30, 32, 19, 5, 1, 22, 31, 2, 27, 11] showing that the field reaches maturity. Despite this growing interest, the field lacks standardized benchmarks to allow for evaluating objectively and quantitatively the performance of the existing dehazing techniques.

Basically, a major issue preventing further developments is related to the impossibility to reliably assess the dehazing performance of a given algorithm, due to the absence of reference haze-free images (ground-truth). A key problem in collecting pairs of hazy and haze-free ground-truth images lies in the need to capture both images with identical scene illumination.

In general the existing dehazing quality metrics are restricted to non-reference image quality metrics (NRIQA) [24]. For instance, the Fog Aware Density Evaluator (FADE) [13] estimates the visibility of a hazy/foggy scene from a single image without corresponding ground-truth. Unfortunately, due to the absence of the reference (haze-free) images in real-life scenarios, none of these approaches has been generally accepted by the dehazing community.

Recent works synthesize hazy images, using the optical model and known depth to synthesize the haze effect. For instance, FRIDA [33] dataset designed for Advanced Driver Assistance Systems (ADAS) is a synthetic image database with 66 computer graphics generated roads scenes.

C. O. Ancuti (codruta.ancuti@gmail.com, University Politehnica Timisoara), C. Ancuti, R. Timofte, L. Van Gool, L. Zhang and M.-H. Yang are the NTIRE 2019 organizers, while the other authors participated in the challenge.
Appendix A contains the authors' teams and affiliations.
NTIRE webpage: http://www.vision.ee.ethz.ch/ntire19/

D-HAZY [3] is a dataset of 1400+ images of real complex scenes has been derived from the *Middleburry* and the *NYU-Depth V2* datasets. The depth map associated to each high quality image has been used to yield synthesized hazy images based on the simplified optical model. Khoury *et al*. [20] introduce the CHIC (Color Hazy Image for Comparison) database, providing hazy and haze-free images in real scenes captured under controlled illumination. The dataset however only considers two indoor scenes, thereby failing to cover a large variation of textures and scene depth.

The first challenge for single image dehazing has been organized by NTIRE 2018 [4]. With the challenge a large number of dehazing solutions were introduced [4, 36, 25, 21, 15, 28, 16]. The NTIRE 2018 challenge used two novel datasets: I-HAZE [7] and O-HAZE [8]. The I-HAZE consists from 35 hazy images (with haze generated in a controlled indoor environment) and their corresponding ground truth (haze-free) images of the same scene. The O-HAZE dataset includes 45 hazy images and corresponding ground truth (haze-free) images. Both datasets allow for full-reference quality assessment of the dehazing results.

The NTIRE 2019 challenge represents a step forward in benchmarking single image dehazing. It uses a novel dehazing dataset, **Dense-Haze** [6], that consists from 55 hazy images with dense haze generated in indoor and outdoor environments; and their corresponding ground truth (haze-free) images of the same scene. The dense haze has been produced using a professional haze/fog generator that imitates the real conditions of haze scenes. The evaluation was performed by comparing the restored hazy images with the ground truth images.

2. NTIRE 2019 Challenge

The objectives of the NTIRE 2019 challenge on single image dehazing are: (i) to gauge and push the state-of-the-art in image dehazing; (ii) to compare different solutions; and (iii) to promote novel Dense-Haze datasets with real haze and ground truth haze-free images.

2.1. Dense-Haze dataset

Dense-Haze [6] dataset contains 33 outdoor scenes and 22 indoor scenes in presence or absence of dense haze. Our dataset allows to investigate the contribution of the haze over the scene visibility by analyzing the scene objects radiance starting from the camera proximity to a maximum distance of 20m.

For the indoor scenes, after carefully setting each scene, we first recorded the ground truth (haze-free image) and then immediately started introducing haze in the scene. We used two professional fog/haze machines (LSM1500 PRO 1500 W) to generate a dense vapor. These fog generators use cast or platen type aluminum heat exchangers, which causes evaporation of the water-based fog liquid. The generated particles (since are water droplets) have approximately the same diameter size of 1 - 10 microns as the atmospheric haze. Before shooting the hazy scene, we used a fan to obtain in a relatively short period of time a homogeneous haze distribution in the entire room (room kept isolated as much as possible by closing all the doors and windows). The entire process to generate haze took approximately 1 minute. Waiting approximately another 5-10 minutes, we obtained a homogeneous distribution of the haze. The distances between the camera and the target objects ranged form 3 to 10 meters. The recordings were performed during the daytime in relatively short intervals (20-30 minutes per scene recording) with natural lightning and when the light remained relatively constant (either smooth cloudy days or when the sun beams did not hit directly the room windows).

To capture haze-free and dense hazy images, we used a setup that includes a tripod and a Sony A5000 camera that was remotely controlled (Sony RM-VPR1). We acquired JPG and ARW (RAW) with 24 bit depth. The cameras were set on manual mode and we kept the camera still (on a tripod) over the entire shooting session of the scene. The camera was calibrated in haze-free scene, and then we kept the same parameters for the hazy scene. For each scene, the camera settings was calibrated by manually adjusting the aperture (F-stop), shutter-speed (exposure-time), ISO speed and the white-balance. Setting the three parameters aperture-exposure-ISO was realized using both the built-in light-meter of the camera and an external exponometer Sekonic. For the white-balance we used the gray-card, targeting a middle gray (18% gray). The calibration process was straight-forward, since it just required to set the white-balance in manual mode and to place the gray-card in front of the subject. In practice, we placed the gray-card in the center of the scene, two meters away from the camera.

For outdoor scenes, a crucial problem in collecting such images is represented by capturing pixel-level images for each scene with and without haze under identical conditions, using the same camera settings, viewpoint, etc. Besides assuring that the scene is static, the scene components keep do not change their spatial position during the recording (quite challenging for natural scenes due to numerous factors), the most challenging issue is to preserve the outdoor scene illumination. As a result, we recorded the outdoor scenes only during cloudy days, in the morning or in the sunset. Additionally, another important constraint was given by the influence of the wind. In order to limit fast spreading of the haze in the scene we could record images only when the wind speed was below 2-3 km/h. This con-

---

http://vision.middlebury.edu/stereo/data/scenes2014/

http://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html

http://www.vision.ee.ethz.ch/ntire18/i-haze/

http://www.vision.ee.ethz.ch/ntire18/o-haze/

straint was hard to meet, it is a main reason for the 8 weeks duration required by the recording of the 33 outdoor scenes from **Dense-Haze**.

To yield hazy outdoor scenes, the haze was spread using a similar setup used to generating indoor hazy scenes. To obtain a dense and homogeneous haze layer in the scene, we employed for 2-3 minutes both haze machines powered by a portable 2800 Watt generator, and waited for another 2-3 minutes. Additionally, all the recorded scenes contain a color checker to allow for the post-processing of the recorded images. We used a classical Macbeth color checker with the size 11 by 8.25 inches with 24 squares of painted samples (4×6 grid).

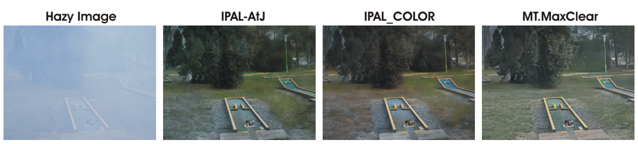

Figure 1. Qualitative dehazing results: from left to right initial hazy image and the results yielded by the methods of iPAL-Atj, iPAL-COLOR and MT.MaxClear.

2.2. Dense-Haze Challenge

For the NTIRE 2019 dehazing challenge we created a Codalab competition. To access the data and submit their dehazed image results to the CodaLab evaluation server each participant had to register.

Challenge phases *(1) Development (training) phase:* the participants got train data (hazy and haze-free images) (45 sets of images); *(2) Validation phase:* the participants received 5 additional sets of images and had the opportunity to test their solutions on the hazy validation images and to receive immediate feedback by uploading their results to the server. A validation leaderboard is available; *(3) Final evaluation (test) phase:* the participants got the hazy test images (5 additional set of images) and had to submit both their dehazed images and a description of their methods before the challenge deadline. One week later the final results were made available to the participants.

Evaluation protocol The Peak Signal-to-Noise Ratio (PSNR) measured in decibel (dB) and the Structural Similarity index (SSIM) [35] computed between an image result and the ground truth are the quantitative measures. The higher the score is the better the restoration fidelity to the ground truth image is.

3. Challenge Results

From more than 270 registered participants, 23 teams entered in the final phase and submitted results, codes/executables, and factsheets. Table 1 reports the final scoring results of the challenge and Table 2 shows the runtimes and the major details for each entry. Section 4 describes briefly the methods for each team while in the Appendix A are the team members and affiliations.

Team	User (+entry)	PSNR	SSIM
iPAL-AtJ	moonriverLucy	**20.258**	**0.657**
iPAL-COLOR	DH-IRCNN_123_CEDH	19.923	0.653
MT.MaxClear	ucenter52	19.469	0.652
BMIPL-UNIST-DW-1	Sprite+Ours	18.842	0.633
xddqm	Untitled Folder	18.521	0.640
ECNU	emmm+dpn_best	17.826	0.617
MOMOCV	meshpop	17.177	0.564
BMIPL-UNIST-DW-2	BMIPL-PDW+Hazing	16.857	0.610
BOE-IOT-AIBD	BOE-IOT-AIBD	16.780	0.612
MAHA@IIT	akshay.aad16	16.472	0.548
FastNet	tzofi+submission	16.371	0.569
IVL1	IVL+submission16	16.194	0.601
ecsuiplab1	san_santra+up_4	16.152	0.564
IPCV IITM	maitreya_ipcv+final	16.126	0.595
shh	sunhee+res	16.055	0.562
IVL2	IVL+submission17	15.801	0.600
ecsuiplab2	ranjanisi+ranjan	15.969	0.539
Alex_SDU	wang_cheng	15.936	0.557
hcilab	hcilab+final	15.122	0.580
IMag	dxllx+t88	14.928	0.555
XZSYS	ChuanshengWang	14.338	0.491
Vintage	jptarel+simple1	14.021	0.529

Table 1. NTIRE 2019 Challenge dehazing results and final rankings on DENSE-HAZE test data.

Architectures and main ideas Most of the proposed methods (excepting the one introduced by Vintage) use end-to-end deep learning strategies and employ the GPU(s) for both training and testing. The Vintage uses an unsupervised technique, Retinex, on the inverted intensities of a hazy image. However, the CPU-based technique is the lowest ranking in terms of PSNR and SSIM performances.

Restoration fidelity In PSNR and SSIM terms the iPAL-AtJ, iPAL-COLOR and MT.MaxClear are the winner teams of NTIRE 2019 dehazing challenge. iPAL-AtJ achieved more than 20dB for PSNR and 0.65 for SSIM.

Team	Runtime [s]	Platform	CPU/GPU (at runtime)
iPAL-AtJ	0.083	pytorch 1.0	NVIDIA TITAN Xp GPU 12GB
iPAL-COLOR	0.077	pytorch 1.0	NVIDIA TITAN Xp GPU 12GB
MT.MaxClear	80	pytorch 0.4.1	
BMIPL-UNIST-DW-1	0.75	pytorch 1.0	GPU >10GB
xddqm			
ECNU		pytorch 0.4.1	4 GTX 1080Ti
MOMOCV	0.020	pytorch 0.4	NVIDIA V100, NVIDIA 1080Ti
BMIPL-UNIST-DW-2	1.01	pytorch 1.0	GPU >10GB
BOE-IOT-AIBD	6	pytorch 1.0	NVIDIA Titan X GPU 12GB
MAHA@IIT	0.95	Tensorflow	NVIDIA GTX 1080 8 GB
FastNet	0.03	PytorchC++	NVIDIA Titan RTX
IVL1	0.00065	pytorch	NVIDIA Titan X Pascal GPU
ecsuiplab1	6	Tensorflow	8x Nvidia GeFroce GTX Titan Xp
IPCV IITM	1	pytorch	NVIDIA Titan X GPU
shh	1.53	pytorch	NVIDIA GTX 1080Ti
IVL2	0.03	pytorch	NVIDIA Titan X Pascal GPU
ecsuiplab2	12	Tensorflow	8x Nvidia GeFroce GTX Titan Xp
Alex_SDU			
hcilab	0.8	pytorch	NVIDIA 1080 GTX
IMag	2.4	pytorch	Nvidia GTX1080ti
Vintage	0.7	matlab	CPU

Table 2. Reported runtimes per image on DENSE-HAZE test data and details from the factsheets.

https://competitions.codalab.org

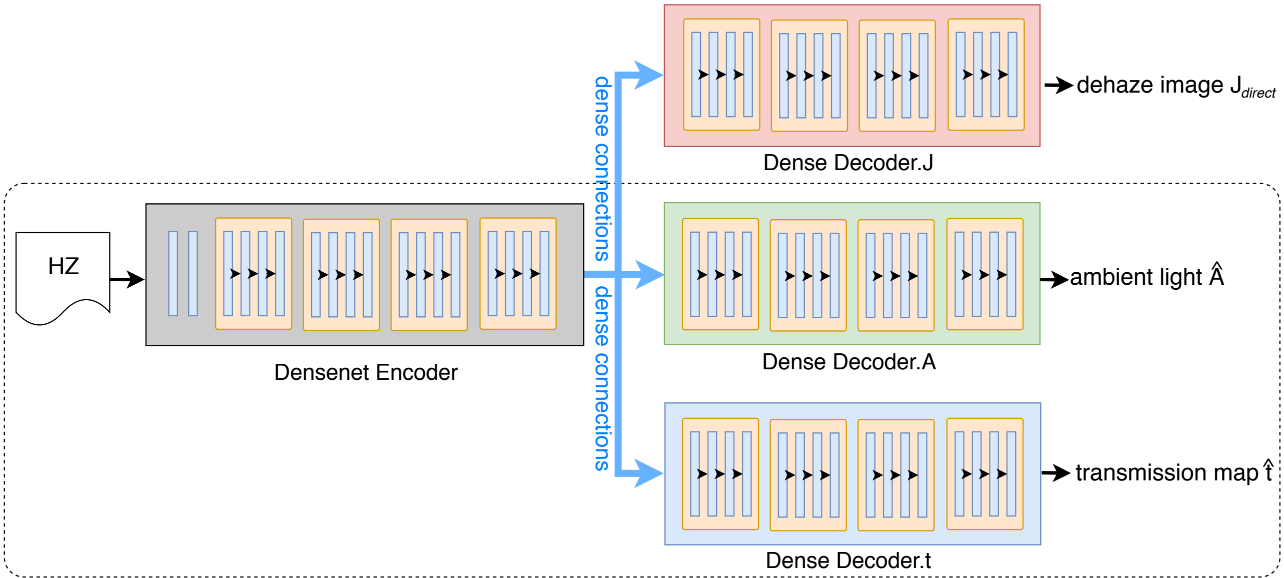

Figure 2. The proposed At-DH (within dash-lines) and AtJ-DH network (iPAL-AtJ Team). In At-DH, a shared encoder is employed to extract image features that are subsequently exploited by the two decoders (Decoder.A and Decoder.t) to jointly estimate the physical model parameters. AtJ-DH employs an additional decoder, *i.e*. Decoder.J which estimates the haze-free image.

Runtime / efficiency IVL1 solution is the most efficient. It runs in 0.16 ms on Intel I7-7700 CPU @ 3.60GHz, 16GB DDR4 RAM 2400MHz, NVIDIA Titan X Pascal GPU with 3840 CUDA cores but their results in terms of PSNR are around 16dB.

Train data Besides **Dense-Haze** dataset, I-HAZE [7] with 35 indoor set of images and O-HAZE [8] with 45 set of outdoor images were used by several competitors that in general found the amount of data sufficient for training their model, especially after data augmentation [34] (by operations such as flipping, rotation, scaling).

Conclusions By analyzing the challenge methods and their results we can draw several conclusions. (i) The proposed solutions have a degree of novelty and go beyond the published state-of-the-art methods. (ii) In general the best solutions performed the best for both measures (PSNR and SSIM). (iii) The evaluation based on SSIM is questionable since there is only a small variation of the SSIM results.

4. Challenge Methods and Teams

4.1. iPAL-AtJ

iPAL-AtJ Team introduces a dense CNN that particularly focuses on the haze generation physical model [18]. Many recent dehazing methods have addressed this challenge by designing deep networks that estimate physical parameters in the haze model, i.e. ambient light (A) and transmission map (t). The authors developed two novel network architectures to further this line of investigation. The first model, denoted as At-DH, designs a shared DenseNet based encoder and two distinct DensetNet based decoders to jointly estimate the scene information *viz.* A and t respectively.

To incorporate more structural information in the learning process and overcome the difficulty of estimating A and t in dense haze, the authors developed an extension of At-DH called the AtJ-DH network, which adds one more DenseNet based decoder to jointly recreate the haze-free image J along with A and t. The knowledge of training dehazed/clean (ground truth) images is exploited by a custom regularization term that further enhances the estimates of model parameters A and t in AtJ-DH.

The encoder in the both models share the same parameters which are obtained by pre-training a full autoencoder structure with the goal of reconstructing the haze-free images by using the hazy images as the input. After the pre-training, the encoder is used for both models as a common feature extractor. The At-DH and AtJ-DH network architectures are illustrated in Fig.2.

4.2. iPAL-COLOR

iPAL-COLOR team observed that usually the color information is harder to recover from the images with dense haze compared to the images with less haze. Thus the authors proposed work addressed these challenges by developing a network structure that comprises of: a common DenseNet based feature encoder whose output branches into three distinct DensetNet based decoders to yield estimates of the R, G, and B color channels of the image [17]. A subsequent refinement block further enhances the final synthesized RGB/color image by joint processing of these color channels. Inspired by its structure, the proposed network is called the One-To-Three Color Enhancement Dehazing

(123-CEDH) network. To ensure the recovery of physically meaningful and high quality color channels, the main network loss function is further regularized by a multi-scale structural similarity index term as well as a term that enhances color contrast. To take the inter-color channel information into consideration, the authors further use CNN layers at different scales to fine-tune the color information and generate the final dehazed image. The proposed 123-CEDH model is shown in Fig. 3.

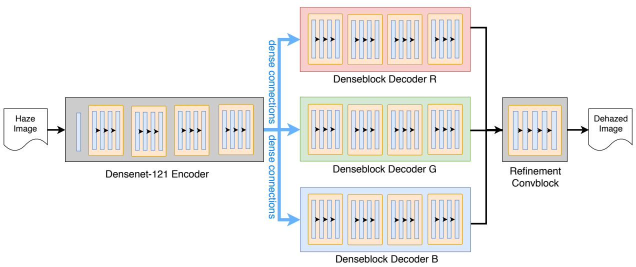

Figure 3. The proposed 123-CEDH network structure. There is one dense encoder to extract general image features, and three decoders to recover R, G, and B color channels. The final refinement block exploits the inter-color channel information for further enhancement.

4.3. MT.MaxClear

MT.MaxClear team uses a modified HRNet [29] as their backbone network for its excellent ability in combining various features of different resolutions. The proposed network consists of 4 stages, each stage adds one additional stride-2 down-sampling branch compared to previous stage. Features of different resolutions are fused frequently between these stages. They add an additional PSP module after the HRnet module for obtaining better global information representation, which, based on some observations, is crucial in Image Dehaze Task. The authors used L1, L2, GAN and perceptual (VGG) loss to train the net-work. Apart from including training images from last year, the team also exploits extensively on various data augmentation methods, some of them are proved to be quite useful, including random input size, random patch cropping etc. The authors use a novel haze-generating method to further expand training data. For this purpose, a haze-generation network is trained, the input and target of which is simply reversed pairs of original training data - they use GT images as input and hazy inputs as target. This haze- generation network is then applied on the raw GT images to generate new training data.

4.4. BMIPL-UNIST-DW-1

The BMIPL-UNIST-DW-1 Team introduced a learning method based on a network that consists of several Encoders and Decoders. The Encoder is composed of Dilated Resnet-18. The Resnet-18 downscaling minimizes resolution loss up to 32 times and Dilated Resnet-18 downscaling up to 8 times. The authors also extend receptive field using Pyramid Pooling. The decoder part consists of Bilinear + convolution layer, and Skip connection part is set to add instead of concatenation. Also, we used Decoder Skip connection channel fixed to 256 channels. The architecture of the proposed method is shown in Fig. 4

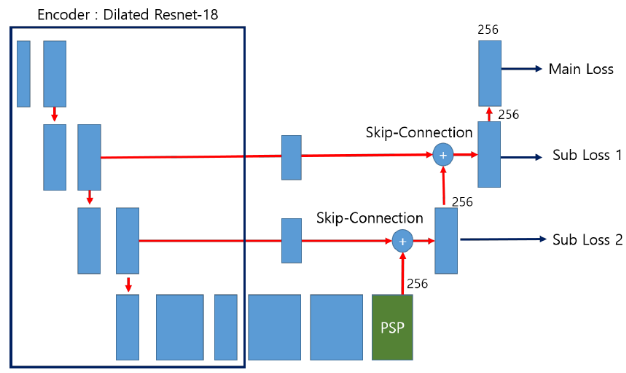

Figure 4. The proposed architecture of BMIPL-UNIST-DW-1 Team.

4.5. xddqm

The xddqm Team proposed a Multi-scale Cascade Generative Adversarial Network (MSCGAN) for image dehazing [12]. The generator of MSCGAN contains two subnets which have a similar Unet-shape architecture, all consists of two downsample residual block, nine residual block, two upsample residual block and a refine convolution layer. Differently, the first network (Net1) takes the downsampled haze images as inputs; the second one (Net2) takes the original haze images and features of Net1 before refine convolution layer as inputs (see Fig. 5).

Motivated by multi-scale discriminators used in Pix2PixHD and spectral normalized GANs for stable training, the authors developed a multi-scale spectral normalization discriminator structure for training the dehaze net. The discriminator uses 3 models, D1, D2 and D3, with a identical structure that trained at 3 different scales of the real and dehazed high resolution images,images downsample by 2× and 4×, respectively. They also adopt the spectral normalization technology to stabilize the training process of CCGAN.

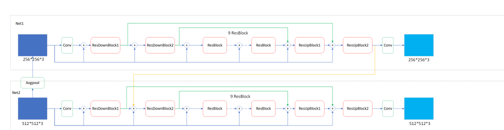

Figure 5. xddqm Team: architecture with 7 used convolutional layers. The digits denotes numbers of channels.

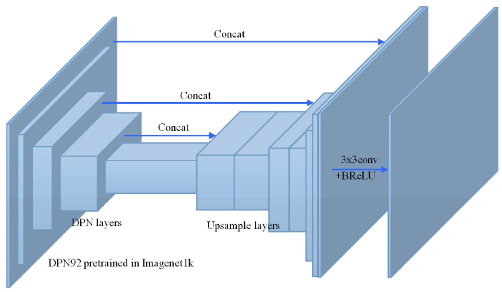

Figure 6. ECNU Team: the architecture of the proposed method.

4.6. ECNU

The ECNU Team proposes a network architecture that is an end-to-end Encoder-Decoder structure Network (EDN) to learn the Hazy-GT mapping, as shown in Fig. 6. The backbone of encoder we use in the competition is DPN92 pre-trained in ImageNet1K. Concat is used in 8x,16x,32x downsampled scales and 1x. The decoder is composed of several upsample layers. The upsample layers are composed of attention modules, dense block and upsample block, details are in Fig. 7. The depthwise conv layer in Fig. 7 provides each spatial attention map for each channel. The second ReLU layer in dense blocks is replaced by xUnit, a learnable nonlinear activation function with spatial connections. The channel attention layer in attention module is the same as SENet, composed by a global average pooling layer, two linear layers, a ReLU layer and a sigmoid layer in the tail of the attention module. In the tail of the network, the authors use a BReLU layer to limit the output between 0 and 1. To enhance the results, they add a U-net after EDN to adjustment the intensity of each pixel. The structure of U-net is the same as DCPDN. They train EDN first, and then joint learning the whole network.

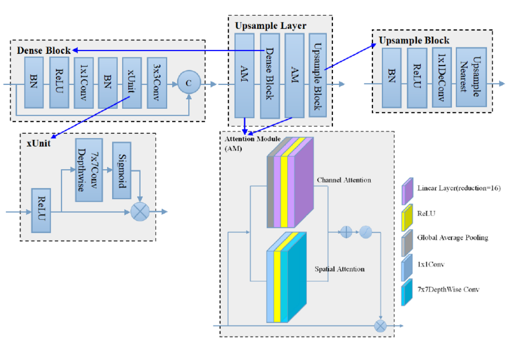

Figure 7. ECNU Team: modules in the proposed upsampling layers.

4.7. MOMOCV

The authors introduce a generative adversarial model, with some important modifications to solve the dehazing problem. They use a similar structure as CycleGAN [38] architecture, however fully utilize the paired characteristics with the content loss on both sides. To be specific, they compare real clear image (in domain B) with the dehazed image (generated by the forward network from the dense-hazy image in domain A), and similarly for the real hazy image (in domain A) and the hazed image from clear image (in domain B). They carefully select the criterion for those comparison. They use the VGG network to extract the semantic information of the images to be paired. The proper VGG layer and the weight relative to the other losses play important roles.

4.8. BMIPL-UNIST-DW-2

The BMIPL-UNIST-DW-2 introduces a CNN networks that consists of several Encoders and Decoders. The Encoder is composed of DenseNet-169. The DenseNet-169 downscaling minimizes resolution loss up to 32 times. The authors also extend receptive field using Pyramid Pooling. The decoder part consists of Bilinear + convolution layer, and Skip connection part is set to concatenation. Also, they used Decoder Skip connection channel (512, 256, 128, 64, 32, 16) channels.

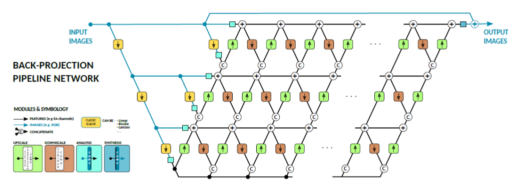

Figure 8. BOE-IOT-AIBD Team: the architecture of the proposed method.

4.9. BOE-IOT-AIBD

BOE-IOT-AIBD Team introduces a new deeplearning architecture with the aim to solve general image enhancement problems based on previous work in image super-resolution [23]. The proposed network design follows signal processing principles based on the Iterative BackProjection (IBP) algorithm. Compared to similar approaches, the authors propose a novel solution to make backprojections run in multiple resolutions by using a data pipeline workflow. The residual and sequential nature of our system provides convenient features similar to ResNets. First, there are straight connection from the output to every convolutional layer, providing gradient superhighways; and second, the sequential processing allows very large models to run efficiently without demanding excessive memory. Finally, a distinctive feature of their design is the use Instance Normalization layers (see Fig. 8). The authors found this to be very effective to make the network stable during training. At the expense of better results, the final architecture presents difficulties to be applied when image resolutions are very different in training and inference.

4.10. MAHA@IIT

The authors designed an end-to-end generative adversarial network (GAN) for single image haze removal [14]. The proposed network bypasses the intermediate stages and directly recovers the haze-free scene. Generator architecture of the proposed network is designed using novel residual inception (RI) module (see Fig. 9). The RI module comprises of dense connections within the multi-scale convolution layers which allows it to learn the integrated avors of the haze-related features. Also, the authors propose a dense residual module for discriminator network of RI-GAN. Further, to preserve the edge and the structural details in the recovered haze- free scene, structural consistency loss and edge consistency loss along with the L1 loss are incorporated in proposed RI-GAN.

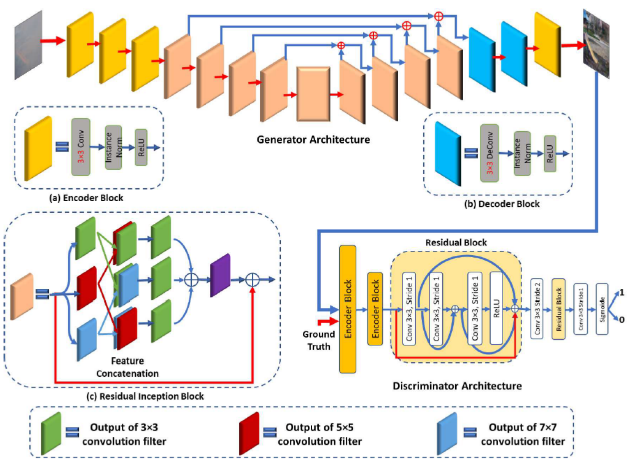

Figure 9. MAHA@IIT Team: overview of the proposed method.

4.11. FastNet

The proposed approach [26] utilizes a feed-forward convolutional neural network trained only on the competition's provided dataset. The network uses a pre-trained feature encoder-to-decoder convolutional architecture, feeding to dense convolutional refinement layers. Two approaches were attempted. One approach estimates airlight and transmission maps with two separate encoder-to-decoder networks that feed into the refinement layers. This model is named DualFastNet. The second approach uses a single encoder-to-decoder network that feeds into the refinement layers. This model is named FastNet50. The competition test results were generated from the DualFastNet architecture (see Fig. 10).

4.12. IVL1

The core of the proposed method [9] is the encoder-decoder model (shown in Fig. 11) inspired by the U-Net architecture. In order to reduce the generation of artifacts and

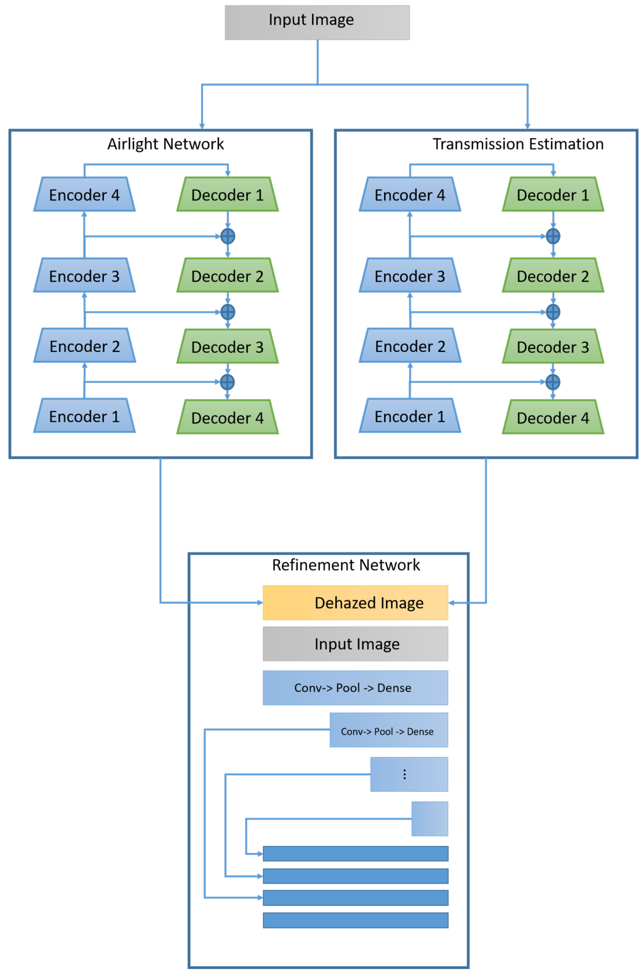

Figure 10. The network architecture of FastNet team.

gain sample independence, we respectively replace Convolution Transpose layers with Pixel Shuffle layers and remove all Batch Normalization layers.

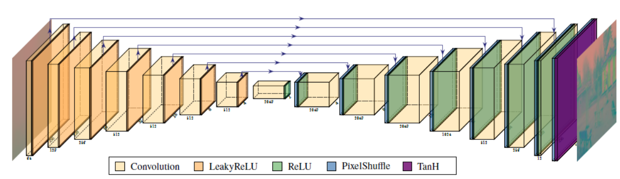

Figure 11. IVL1 Team: overview of the proposed method.

4.13. ecsuiplab1

The ecsuiplab1 team trained the dehazing network in adversarial manner. It consists of two networks: a dehazing network, and a discriminator network. The block structure of the dehazing network is given in Fig. 12. They applied residual blocks in two image scales for feature extraction. They have also taken noise (uniform random values between -1 and 1) as the input apart from the locally contrast stretched image. From each of the inputs, features are extracted separately. Then they are concatenated and utilized to generate the final dehazed output. The discriminator is a simple CNN with some conv, Instance norm and Leaky ReLU layers stacked together. It takes the generated or the real haze-free image along with the hazy image as input to decide whether the patches of the input are real or generated. The discriminator is trained to minimize the MSE loss of the predicted labels (real/fake). Whereas, the dehazing network is trained with the aim of minimizing the error of labelling the generated images as real by the discriminator along with DSSIM and MAE loss of the dehazed image from the ground-truth. The weight of these three terms is taken to be 1, 25 and 50. DSSIM is a measure of structural dissimilarity based on SSIM metric.

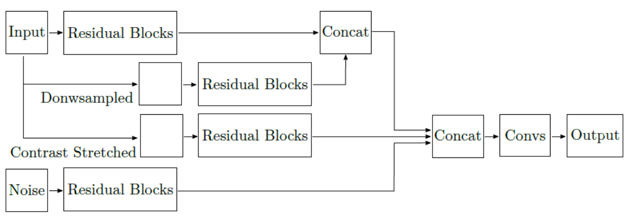

Figure 12. The network architecture of ecsuiplab1 team.

4.14. IPCV-IITM

The proposed network consists of a Deep dense-residual encoder-decoder structure with subpixel convolution and multi-level pyramid pooling module for efficiently estimating the dehazed image. The input image is first downsampled by a factor of 2 and fed to the encoder, which increases the receptive field of the network and reduces computational footprint. The encoder is made of densely connected modules. It helps to address the issue of vanishing gradients and feature propagation while substantially reducing the model complexity. The encoder has a similar structure to the network mentioned in [37]. The weights are initialized using pre-trained Dense-121 network. Each layer in a block receives feature maps from all earlier layers, which strengthens the information flow during forward and backward pass making training deeper networks easier. The final output is the residual between the ground-truth colored image and the input image. The Decoder accepts the features estimated by the encoder at various levels and processes them using residual blocks before increasing their spatial resolution through bi-linear up-sampling and convolution. The intermediate features with higher spatial resolution in the decoder are concatenated with the corresponding-sized encoder features. Finally, the decoder output is enhanced through multi-scale context aggregation through pooling and upsampling at 4 scales, before being fed to the final layer. To match the features to the input image resolution, the authors employ a subpixel-convolution block to upsample them by a factor of 2. They optimize the inference time of the network by performing computationally intensive operations on features at lower spatial resolution. This also reduces memory footprint while increasing the receptive field. They make the downsample-upsample operation end-to-end trainable which gives better performance compared to bilinear operations. The upsampled features are passed through a convolutional layer to construct the dehazed output (see Fig. 13).

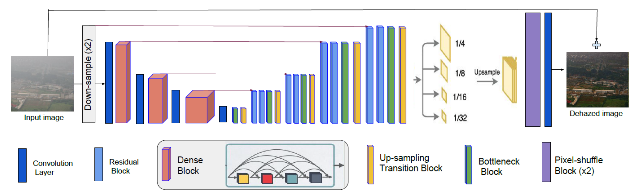

Figure 13. IPCV-IITM Team: overview of the proposed method.

4.15. shh

The authors introduce a the multi-task dehazing generative adversarial network that generates haze-free image and transmission image from the input hazy image. The network has two generators with thirteen shared convolutional networks and three convolutional networks for each task. The authors uses U-Net based generator as shown in Fig. 14. For the multi-task image generation, we removed two skip-connection layers. We also added four convolutional layers followed by 512-dimension features to achieve outperform results. To train the network we take three procedures as follows. First two parts are trained by NTIRE 2018 dataset to generate the pre-trained dehazing model.

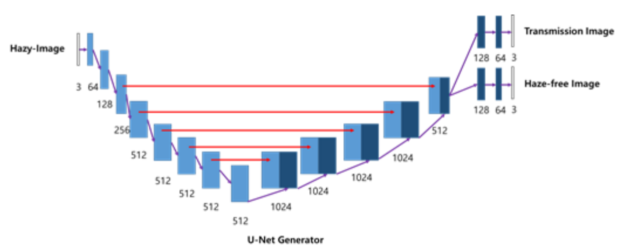

Figure 14. shh Team: overview of the proposed method.

4.16. IVL2

The authors propose a content-preserving method for image enhancement that estimates a global color transformation. The proposed method takes inspiration by the work of Bianco *et al.* [10], and is composed by two different neural networks: the first one performs global enhancement and is in charge of estimating the coefficients of a global color transformation, in the form of a continuous piece wise function, which is later applied to the input image; the second one performs local enhancement and estimates the best spatial filters to be applied to further improve the enhancement (see Fig. 15).

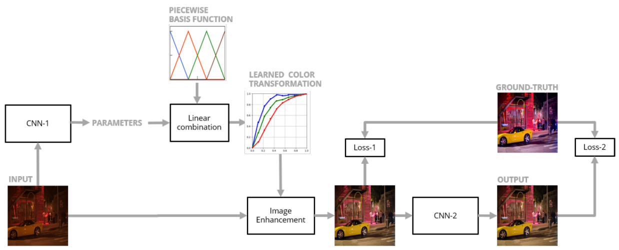

Figure 15. IVL2 Team: overview of the proposed method.

4.17. ecsuiplab2

The ecsuiplab2 team trained the dehazing network in adversarial manner. It consists of three networks: a dehazing network, a discriminator network and a refinement network. The block structure of the dehazing network is given in Fig. 16. There are four paths from the given input. In two of them features are extracted from local contrast stretched version of the input image with two different window sizes. Out of the remaining two paths one extracts global features and the other one extracts local features from the image. As the global features can help in computation of the local features, the global features are concatenated before computing the local features. The features computed from all four paths are concatenated and then used for generating the output. The discriminator is a simple classifier CNN with conv, instance norm and leaky relu blocks. It takes the generated or the real haze-free image along with the hazy image as input to decide whether the patches of the input are real or generated. The discriminator is trained to minimize the MSE loss of the predicted labels (real/fake). Whereas, the dehazing network is trained with the aim of minimizing the error of labelling the generated images as real by the discriminator along with DSSIM and and MAE loss of the dehazed image from the ground-truth. The refinement network has an architecture similar to UNet that tries to further refine the the result generated by the dehazing network.

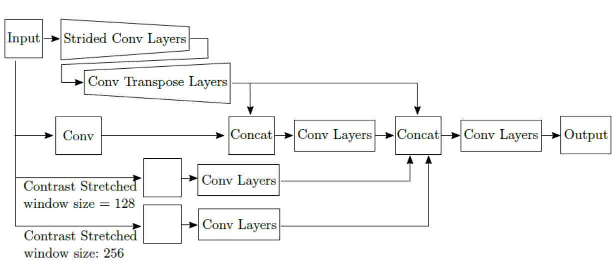

Figure 16. The network architecture of ecsuiplab2 team.

4.18. Alex-SDU

The method proposed by Alex-SDU team is a combination of GCANet and pyramid pooling. This kind of model have potential to improve dehazing performance. Moreover, we utilize channel attention mechanism and wide activation operation to boost the performance. In order to get optimal results the authors used a patch-based strategy for the inference process. The authors trained their model with 600×800 patch size for 100000 epochs. The learning rate was set to 0.001 and decayed by 0.5 times each 10000 epochs. The batch size was 8.

4.19. hcilab

The hcilab team customized the U-net with 16-Block Layers network for training and testing the images. The network consists of 8 blocks layers for encoding the images and 8 blocks for decoding the images. For encoding part each block contains: conv-conv-pool combination layers and for decoding part each block contains upsampling-conv-conv combination layers.

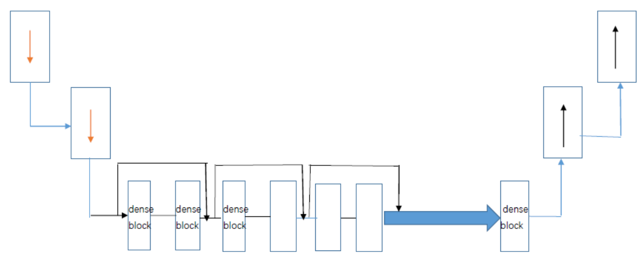

Figure 17. The network architecture of iMag team.

4.20. iMag

The iMag team's approach uses a condition gan network. They use Residual in Resid-ual Dense Network as the generator and multi-scale discriminator in Pix2PixHd as the discriminator. They also use spectral normalization in discriminator to stable our training. In Residual in Residual Dense Network, the authors downsample two times in encoder and upsample two times in decoder at last. They adopt a mix of GAN loss, L1 loss, perceptual loss and SSIM loss as the total losses (see Fig. 17).

4.21. Vintage

The proposed method is based on filtering the image to guess the value of the colored veil which can be due to fog, dust or smoke. Then the Koschmieder's model is reversed to obtain the restored image with fog, dust or smoke attenuated. A linear and gamma mapping is then performed to obtain a less dark image and a little smoothing is also performed to attenuate noise.

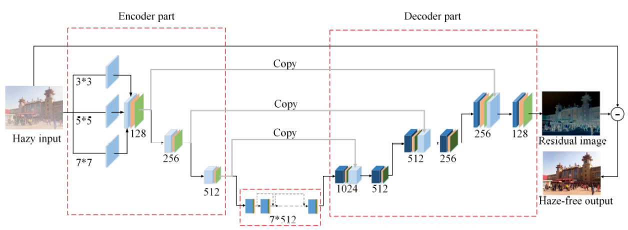

Figure 18. XZSYS Team: overview of the proposed method.

4.22. XZSYS

The authors introduce a deep deep symmetric residual network DSRNet consists of three parts, as illustrated in Fig. 18: an encoder part (left side), a nonlinear transformation part (middle part) and a decoder part (right side). The DSRNet is directly motivated to learn the haze residual images between the hazy images and the clear images. Based on this motivation, the authors adopt a multi-scale CNN to vastly enhance the performance by aggregating more context information in the first layer of the encoder part and then encode it into feature maps. Subsequently, for the sake of extract high-level features, they insert seven residual blocks which effectively improve the convergence and help enhance the performance of the the proposed DSRNet. The high-level feature maps is finally decoded back by decoder part to get the residual map. Finally, by subtract the input hazy image to the haze residual image, it is obtained the haze-free image. The authors also introduce a novel algorithm of feature extract and a useful nonlinear activation function, called Reverse Parametric Rectified Linear Unit (RPReLU).

Acknowledgments

We thank the NTIRE 2019 sponsors: Huawei Technologies Co. Ltd., NVIDIA Corp., Amazon.com, Inc., Samsung, MediaTek, Oppo and ETH Zurich (Computer Vision Lab). Part of this work was supported by research grant GNaC2018 - ARUT, no. 1361-01.02.2019, financed by Politehnica University of Timisoara. Part of this work was supported by 2020 European Union Research and Innovation Horizon 2020 under the grant agreement Marie Sklodowska-Curie No 712949 (TECNIOspring PLUS), as well as the Agency for the Competitiveness of the Company of the Generalitat de Catalunya - ACCIO: TECSPR17-1-0054.

A. Teams and affiliations

NTIRE2019 team

Title: NTIRE 2019 Challenge on Image Dehazing
Members: *Codruta O. Ancuti*[1] *(codruta.ancuti@gmail.com)*, Cosmin Ancuti [2,1], Radu Timofte [3], Luc Van Gool [3,4], Lei Zhang [5], Ming-Hsuan Yang [6]
Affiliations:
[1] ETcTI, University Politehnica Timisoara, Romania
[2] Institute of Informatics and Applications, University of Girona, Spain
[3] Computer Vision Lab, ETH Zurich, Switzerland
[4] ESAT, KU Leuven, Belgium
[5] Polytechnic University of Hong Kong, China
[6] University of California at Merced, US

iPAL-AtJ

Title: Dense AtJ Dehazing Network
Members: *Tiantong Guo (tong.renly@gmail.com), Xuelu Li, Venkateswararao Cherukuri, Vishal Monga*
Affiliation:
The Pennsylvania State University, School of Electrical Engineering and Computer Science, The Information Processing and Algorithms Laboratory, US

iPAL-COLOR

Title: 123 Dense Color Enhance Dehazing Network
Members: *Tiantong Guo (tong.renly@gmail.com), Xuelu Li, Venkateswararao Cherukuri, Vishal Monga*
Affiliation:
The Pennsylvania State University, School of Electrical Engineering and Computer Science, The Information Processing and Algorithms Laboratory, US

MT.MaxClear

Title: HR-DehazeNet
Members: *Hao Jiang (jh1@meitu.com), Siyuan Yang, Yan Liu, Xiaochao Qu, Pengfei Wan*
Affiliation:
MTlab (Meitu,Inc), Beijing, China

BMIPL-UNIST-DW-1

Title: Dilated ResNet based Multi-scale CNN for Single Image Dehazing Problem
Members: *Dongwon Park (dong1@unist.ac.kr), Se Young Chun*
Affiliation:
Ulsan National Inst. of Science and Technology (UNIST), South Korea

xddqm

Title: Multi-scale Cascade Generative Adversarial Network (MSCGAN)
Members: *Ming Hong (1184307675@qq.com), Jinying*

Huang, Yizi Chen, Shuxin CHen
Affiliation:
Xiamen university, China

BMIPL-UNIST-DW-2

Title: Densenet based CNN for Single Image Dehazing Problem
Members: *Dongwon Park (dong1@unist.ac.kr), Se Young Chun*
Affiliation:
Ulsan National Inst. of Science and Technology (UNIST), South Korea

Alex-SDU

Title:
Members: *Bomin Wang (201712354@mail.sdu.edu.cn)*
Affiliation:

BOE-IOT-AIBD

Title: BackProjection Pipelines
Members: *Pablo Navarrete Michelini (pnavarre@boe.com.cn), Hanwen Liu, Dan Zhu*
Affiliation:
BOE Technology Group Co., Ltd., China

ECNU

Title: Encoder-Decoder architecture for dehazing
Members: *Jing Liu (splinter02@163.com)*
Affiliation:
East China Normal University(ECNU), Multimedia and Computer Vision Lab, China

ecsuiplab1

Title: Con-Haze
Members: *Sanchayan Santra (sanchayan.santra@gmail.com),Ranjan Mondal, Bhabatosh Chanda*
Affiliation:
Indian Statistical Institute, Kolkata, India

ecsuiplab2

Title: MC-Dehaze
Members: *Sanchayan Santra (sanchayan.santra@gmail.com),Ranjan Mondal, Bhabatosh Chanda*
Affiliation:
Indian Statistical Institute, Kolkata, India

FastNet

Title: Feature Forwarding for Efficient Image Dehazing
Members: *Peter Morales (peter.morales@ll.mit.edu), Tzofi Klinghoffer*
Affiliation:
MIT Lincoln Laboratory, US

hcilab

Title: Unet For Single Image Dehazing with Small Training Dataset
Members: *Le Manh Quan (quanle-hci@gmail.com), Yong-Guk Kim*
Affiliation:
Sejong University, Republic of Korea

IMag

Title: Unet For Single Image Dehazing with Small Training Dataset
Members: *Xiao Liang (947884819@qq.com), Runde Li, Jinshan Pan, Jinhui Tang*
Affiliation:
Nanjing University of Science and Technology, China

IPCV-IITM

Title: Efficient Image Dehazing Using Subpixel Dense Residual Network
Members: *Kuldeep Purohit (kuldeeppurohit3@gmail.com), Maitreya Suin, A. N. Rajagopalan*
Affiliation:
Indian Institute of Technology Madras, India

IVL1

Title: Content-preserving tone adjustment for image enhancement
Members: *Raimondo Schettini[1] (schettini@disco.unimib.it), Simone Bianco[1], Flavio Piccoli[1], C. Cusano[2]*
Affiliation:
[1]Department of Informatics, Systems and Communication, University of Milano-Bicocca
[2]Department of Electrical, Computer and Biomedical Engineering, University of Pavia

IVL2

Title: High-Resolution Single Image Dehazing using Encoder-Decoder Architecture
Members: *Raimondo Schettini (schettini@disco.unimib.it), Simone Bianco, Luigi Celona, Flavio Piccoli*
Affiliation:
Department of Informatics, Systems and Communication, University of Milano-Bicocca

shh

Title: Multi-Task Dehazing Generative Adversarial Network
Members: *Sunhee Hwang[1] (sunny16@yonsei.ac.kr), Yu Seung Ma[2], Hyeran Byun[1]*
Affiliation:
[1] Yonsei University, Seoul, South Korea
[2] Electronics and Telecommunications Research Institute, Daejeon, South Korea

MAHA@IIT

Title: RI-GAN: An End-to-End Network for Single Image Haze Removal
Members: *Subrahmanyam Murala (2017eez0001@iitrpr.ac.in), Akshay Dudhane, Harsh Aulakh*
Affiliation:
Indian Institute of Technology Ropar, India

MOMOCV

Title: Content and Cycle Generative Adversary Network for Image Dehazing
Members: *Zheng Tianxiang (zheng.tianxiang@immomo.com), Tao Zhang, Weining Qin, Runnan Zhou, Shanhu Wang*
Affiliation:
Beijing Momo Technology Co., Ltd., China

Vintage

Title: Single Image Desmoking
Members: *Jean-Philippe Tarel (Jean-Philippe.Tarel@ifsttar.fr)*
Affiliation:
IFSTTAR, France

XZSYS

Title: Deep deep symmetric residual network
Members: *Chuansheng Wang[1] (1851607769@qq.com), Jiawei Wu[2]*
Affiliation:
[1]Harbin University of Science and Technology Harbin, Heilongjiang province, China
[2]Minjiang University Fuzhou, Fujian Province, China

References

[1] C. 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, A. Bovik, and C. De Vleeschouwer. Night time dehazing by fusion. *IEEE ICIP*, 2016. 1

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

[4] C. Ancuti, C. O. Ancuti, R. Timofte, et al. Ntire 2018 challenge on image dehazing: Methods and results. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018. 2

[5] 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

[6] C. O. Ancuti, C. Ancuti, M. Sbert, and R. Timofte. Dense haze: A benchmark for image dehazing with dense-haze and haze-free images. In *arXiv:1904.02904*, 2019. 2

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

[8] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer. O-HAZE: a dehazing benchmark with real hazy and haze-free outdoor images. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018. 2, 4

[9] S. Bianco, L. Celona, F. Piccoli, and R. Schettini. High-resolution single image dehazing using encoder-decoder architecture. In *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019. 7

[10] S. Bianco, C. Cusano, F. Piccoli, and R. Schettini. Learning parametric functions for color image enhancement. In *Computational Color Imaging Workshop*, 2019. 8

[11] 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

[12] S. Chen, Y. Chen, Y. Qu, J. Huang, and M. Hong. Multi-scale adaptive dehazing network. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019. 5

[13] L. K. Choi, J. You, and A. C. Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. *In IEEE Trans. on Image Processing*, 2015. 1

[14] A. Dudhane, H. S. Aulakh, and S. Murala. Ri-gan: An end-to-end network for single image haze removal. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019. 7

[15] D. Engin, A. Genc, and H. Ekenel. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018. 2

[16] A. Galdran, A. Alvarez-Gila, A. Bria, J. Vazquez-Corral, and M. Bertalmío. On the duality between retinex and image dehazing. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018. 2

[17] T. Guo, V. Cherukuri, and V. Monga. Dense '123' color enhancement dehazing network. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019. 4

[18] T. Guo, X. Li, V. Cherukuri, and V. Monga. Dense scene information estimation network for dehazing. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019. 4

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

[20] J. E. Khoury(B), J.-B. Thomas, and A. Mansouri. A color image database for haze model and dehazing methods evaluation. *ICISP*, 2016. 2

[21] S. Ki, H. Sim, S. Y. Kim, J.-S. Choi, S. Kim, and M. Kim. Fully end-to-end learning based conditional boundary equilibrium gan with receptive field sizes enlarged for single ultra-high resolution image dehazing. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018. 2

[22] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan. Efficient image dehazing with boundary constraint and contextual regularization. *In IEEE Int. Conf. on Computer Vision*, 2013. 1

[23] P. N. Michelini, H. Liu, and D. Zhu. Multigrid backprojection super–resolution and deep filter visualization. In *Proceedings of the Thirty–Third AAAI Conference on Artificial Intelligence (AAAI 2019)*. AAAI, 2019. 6

[24] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *In IEEE Trans. on Image Processing*, 2012. 1

[25] R. Mondal, S. Santra, and B. Chanda. Image dehazing by joint estimation of transmittance and airlight using bi-directional consistency loss minimized FCN. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018. 2

[26] P. Morales, T. Klinghoffer, and S. J. Lee. Feature forwarding for efficient single image dehazing. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019. 7

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

[28] H. Sim, S. Ki, S. Y. Kim, J.-S. Choi, S. Kim, and M. Kim. High-resolution image dehazing with respect to training losses and receptive field sizes. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018. 2

[29] K. Sun, B. Xiao, D. Liu, and J. Wang. Deep high-resolution representation learning for human pose estimation. In *arXiv:1902.09212*, 2019. 5

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

[31] K. Tang, J. Yang, and J. Wang. Investigating haze-relevant features in a learning framework for image dehazing. *In IEEE Conference on Computer Vision and Pattern Recognition*, 2014. 1

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

[33] J.-P. Tarel, N. Hautire, L. Caraffa, A. Cord, H. Halmaoui, and D. Gruyer. Vision enhancement in homogeneous and heterogeneous fog. *IEEE Intelligent Transportation Systems Magazine*, 2012. 1

[34] R. Timofte, R. Rothe, and L. Van Gool. Seven ways to improve example-based single image super resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 4

[35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004. 3

[36] H. Zhang, V. Sindagi, and V. Patel. Multi-scale single image dehazing using perceptual pyramid deep network. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018. 2

[37] H. Zhang, V. Sindagi, and V. M. Patel. Multi-scale single image dehazing using perceptual pyramid deep network. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018. 8

[38] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE International Conference on Computer Vision (ICCV)*, 2017. 6