SSDC-GAN: Same Size Densely Connected GAN for Dehazing Network

Juan Wang[1,2], Chang Ding[1,2], Yonggang Ye[1,2], Minghu Wu[1,2(✉)], Zetao Zhang[1,2], Sheng Wang[1,2], Hao Yang[1,2], and Ye Cao[1,2]

[1] Hubei Energy Internet Engineering Technology Research Center, Hubei University of Technology, Wuhan 430068, China
wuxx1005@hbut.edu.cn

[2] Hubei Laboratory of Solar Energy Efficient Utilization and Energy Storage Operation Control, Hubei University of Technology, Wuhan 430068, China

Abstract. Haze weather negatively impacts the quality of external image collection and requires prompt resolution. However, most current deep learning image dehazing models struggle with restoring detail and color accuracy in real-world hazy images, hindering their practical application for high-quality image projects. To overcome this issue, we propose a novel connected mode (SSDC) for end-to-end dehazing that simplifies the problem to an image conversion task without relying on atmospheric scattering models or precise priors. The SSDC-GAN generator employs an encoder-decoder, same size densely connected architecture with residual blocks, and a depth discriminator to balance the relationship during training. Experimental results demonstrate that the proposed method performs favorably against state-of-the-art dehazing approaches on various benchmarks using real-world datasets O-HAZE and I-HAZE while preserving accurate contour and color information.

Keywords: Image Dehazing · Densely Connected · Generative Adversarial Network (GAN)

1 Introduction

Haze weather appears as a phenomenon where the atmosphere covers the scene. A large number of suspended particles will make an impact on collected images such as reflection, refraction and absorption of atmospheric light. The phenomenon will reduce the scene visibility, lead to blurred image details, low contrast, color distortion, and information loss. Interfere with the follow-up work of extracting useful information and solving practical problems in computer vision and other works. Inevitably, the problem of how to improve image quality and reduce the interference of photoelectric acquisition equipment in harsh environments must be solved. Dehazing algorithms can improve the accuracy and efficiency of computer vision tasks such as object detection [1], image classification [2] and semantic segmentation [3]. Surveillance cameras and other sensors capture images of pedestrians and cars, but hazy images reduce the quality of information needed for intelligent analysis technology.

K. Hemachandran et al. (Eds.): ICAID 2023, AHIS 9, pp. 327–337, 2023.
https://doi.org/10.2991/978-94-6463-222-4_35

In this paper, image dehazing problem has been converted into image-to-image translation problem. To improve the quality of image generation and preserve most of the image details information, inspired by DCPDN connection mode [4], we propose a novel connection architecture employed in generator, GAN.

In case of model collapses during training, G (generator) with high-quality conquered D (discriminator) because of sophisticated structure optimization, we deepen the network layer of discriminator and reduce the step size of the convolution kernel.

Our main contributions can be summarized as follows:

- We propose an end-to-end dehazing model with novel connection architecture in generator for retaining information of the images in SSDC-GAN. The same size densely connected method is considered from the perspective of image features, that greatly reduces network parameters, enhances image dehazing effects and alleviates the gradient vanishing problem.
- SSDC-GAN employs a generator which combines an encoder and decoder structure with embedded residual blocks in pre-trained Densenet-121 to better preserve the image details. We employ perceptual loss to generate more visually pleasing images. SSDC-GAN optimized a deep discriminator structure to deal with network complexity imbalance problem for non-haze images generation.

2 Proposed Method

2.1 Same Size Densely Connected Encoder-Decoder

We employ the GAN framework to convert the dehazing problem into a image conversion problem, and optimize the network architecture to retain features. To preserve the image and color features information as possible by modifying the network structure without estimating intermediate parameters. Generator is required to retain the image content and cover as much detail as possible. Recently works have revealed the dense connections possess the potentiality to promote feature extraction and utilization, especially for low-level vision tasks. For example, previous works [5, 6] demonstrate the excellent ability of densely connected encoder-decoder in retaining image feature information, including the color and target outline details.

In addition to the commonly encoder and decoder connections, we propose a novel same size connection approach. SSDC perform different pooling operations on the output images in individually generator layer, including pooling layers with filters of sizes 2, 4, 8 and 16 to dwindle the feature map, and insert the reduced image into the network layer of the corresponding size.

SSDC-GAN employs the Denseblock and Transblock in pre-trained densenet-121 to extract image features in encoder, and make pooling operations with different filter sizes. At the end of each Denseblock & Transblock layer there are $2\times$ times downscaling for feature map continues to shrink. In the process of gradually shrinking the feature map, the network extracts the output image of each layer and projects to the decoder layers of the corresponding size. At the same time, the network merges one or more pooling operations of the corresponding size layers and different filters.

SSDC-GAN sets feature map transmission as follows: IMid, ln denotes the input feature graph of the nth layer of the decoder, OMie, ln denotes the output feature map

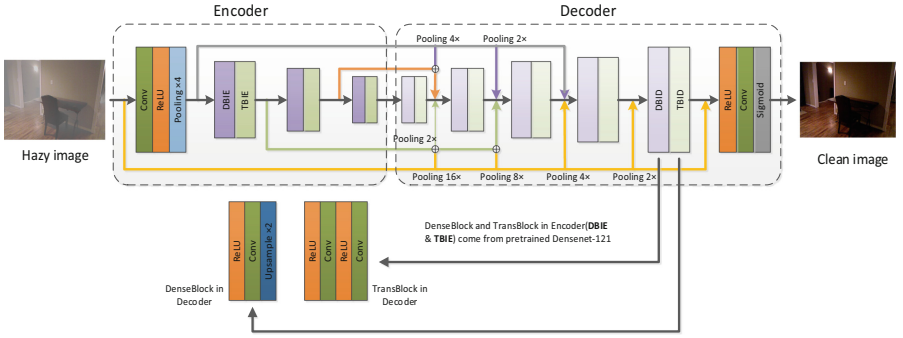

Fig. 1. Generator.

of the nth layer of the encoder, and the following formula demonstrates the input of the second layer in the decoder network:

$$IM^{i}_{d,l2} = IM^{i}_{e,l1}/16 + IM^{i}_{e,l2}/4 + IM^{i}_{e,l3}/2 + IM^{i}_{e,l4} + OM^{i}_{d,l1} \tag{1}$$

In the Eq. (1), the first three variables on the right equation are the input characteristic images of the first, second and third layers of the encoder for unequal pooling, and the output of the first layer in the decoder is attached. The input of the third layer in decoder is:

$$IM^{i}_{d,l3} = IM^{i}_{e,l1}/8 + IM^{i}_{e,l2}/2 + IM^{i}_{e,l3} + OM^{i}_{d,l2} \tag{2}$$

As shown in Fig. 1, the encoder comprises three DenseBlocks & TransBlock (in Encoder) modules, and convolution layer + ReLU with pooling layer (filter size 4). In the decoder module, there are 5 DenseBlocks & TransBlock (in Decoder) modules utilized to upscale the feature maps and restore the resolution.

2.2 Depth Discriminator

We compared the capabilities of the original generator and SSDC-GAN generator using a newly designed depth discriminator (Fig. 2). By deepening the network layers structure of the discriminator, we increased its feature extraction, analysis, and judgment ability, necessitating the rival generator to possess equal image generation ability. Compared to the original discriminator, we added more layers to improve its image information collection from G (generator) and better compare image features of real datasets for improved image authenticity determination. D competes with G to achieve Nash balance. To balance network complexity, we increased the stride of the convolution kernel to 2, reducing the number of parameters per layer and improving processing efficiency.

2.3 Loss Function

We employ the L1 loss function to estimate the pixel loss between the prediction and the target image. Pixel-wise loss is utilized to estimate image differences in detail, and

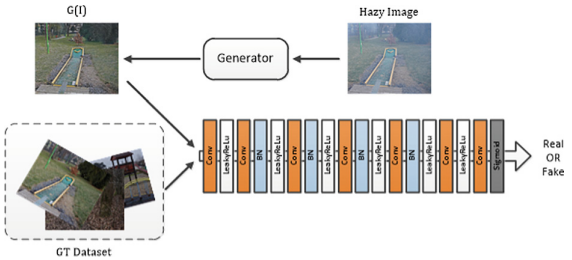

Fig. 2. Depth discriminator, by increasing the step size and layer depth.

actively fed back to the network for regulation.

$$L_1 = \sum_{i=1}^{N} \|G(I_i) - J_i\|_1 \tag{3}$$

Where Ii denotes the input hazy image, taking the output of generator G(Ii) and ground truth Ji into estimation. N denotes the total sample in dataset images. It measures the distortion and fidelity between the ground truth image and the dehazed images.

Perceptual loss compares the feature obtained by convolution of the real images with the feature obtained by the convolution of the generated picture, so that high-level information, such as the content of the images, is close to the global structure. Perceptual loss will improve the similarity of the entire image frame, and adjust the images information through functions in general.

$$L_P = \frac{1}{CHW} \sum_{c=1}^{C} \sum_{h=1}^{H} \sum_{w=1}^{W} \left| \varphi_{c,h,w}(I_d) - \varphi_{c,h,w}(I_g) \right| \tag{4}$$

The symbol φ represents the feature extracted from the pre-trained VGG16 model. C, W and H denotes the height, width, and image channel of the feature map in the j-th layer of the backbone network, where I_d denotes dehazed image and ground truth image is I_g.

We employ the adversarial loss of GAN. The generator is initialized to convert the hazy image into a haze-free image, and discriminator is employed to distinguish whether the image is real or fake. GAN is trained through a straightforward formula in Eq. (5), which can be expressed as:

$$L_A = \frac{1}{N} \sum_{i=1}^{N} \log(1 - D(I_i, \tilde{J}_i)) \tag{5}$$

Adversarial loss is a distance measure of each layer in the GAN. Where the $\tilde{J}_i$ denotes the processed output of generator, and I_i is the ground truth image. Optimizer rectify the training direction after estimation.

We combine Pixel-wise loss, which is used to make the dehazing image closer to the ground truth, perceptual loss that perfects the balance between the advanced information of images, adversarial loss is employed to advance restore photo-realistic images, and set weights for each one to adjust parameters dynamically. The parameters are α_1, α_2 and α_3.

$$L_T = \alpha_1 L_1 + \alpha_2 L_P + \alpha_3 L_A \tag{6}$$

3 Experiment

3.1 Datasets and Implementation

Recently, there are various algorithms for image dehazing, whereas, since pair real-world haze maps and haze-free images are laborious to collect on a massive scale for network training, most of the existent datasets for dehazing are artificially synthesized, such as, RESIDE [7], NYU [8], Haze-RD [9], KITTI [10], and Artificial haze, such as D-HAZY [11], O-HAZE [12] and I-HAZE [13]. Datasets can be divided into different categories from indoor, outdoor, haze density and synthesized or not, there are real world haze maps dataset BeDDE [14] as well.

Weight parameters $\alpha 1$, $\alpha 2$ and $\alpha 3$ are set to 1, 0.5 and 0.1, the most critical pixel-level loss parameter is 1, visual loss whose importance is lower than the former determines image balance, and adversarial loss in GAN set to 0.1 as a rule of thumb. Due to the limitations of the experimental facility memory and the large model of some previous meth-od, this type of methods cannot process original size image.

We select appropriate images from the dataset and resize them to $1600 \times 1200 \times 3$. 7 of the 45 images in the O-HAZE dataset are randomly selected as the testset, while 5 of 25 pieces in the I-HAZE dataset were selected as test sets. The remaining images are employed as the training set. Each image is divided into 16 pieces with the size of 400 * 300 to enlarge the trainset. We employ Adam optimizer with initial learning rate of 1-e4 for both generator and discriminator, the whole training epoch is 3,000 times, on the frame-work Pytorch with a Nvidia TITAN GPU.

3.2 Metric

Some quantitative measurement methods are employed to automatically evaluate the image quality objectively, so as to acquire the parameters reflecting the quality on the degree of loss as the evaluation result.

Utilizing images in the same scene on sunny days as the evaluation reference is the ideal objective metric, and estimating the degree of distortion between the dehazing result and the references.

PSNR (Peak Signal to Noise Ratio) is employed to measure the ratio between the maximum possible value of the signal and the distortion noise power that affects the quality of signal. PSNR is based on the error between the corresponding pixels, and the most widely used objective image evaluation metric.

SSIM [15] (Structural Similarity) is the realization of the structural similarity theory, the structural similarity index defines structural information from the perspective of image composition as being independent of brightness and contrast, reflecting the properties of the object structure in the scene, and modeling distortion into three aspects: brightness, contrast and structure.

LPIPS [16] (Learned Perceptual Image Patch Similarity) permits the features extracted from the network structure of the model can be measured to obtain judgments that are more consistent with human perception. Feature difference between the real sample and the generated sample in the model has been analyzed along with the difference estimated in each channel by L2. Ultimately, LPIPS makes a weight summation for all channels.

3.3 Result Evaluation

We draw a comparison (Fig. 3) (Table 1) between SSDC-GAN and mainstream methods on the dataset O-HAZE since it matches the features of real-world hazy maps. We employ an improved DCP method, which remains haze portion and color deviation at the junction of the sky and objects. In view of too many comparison methods are employed, paper demonstrates the evaluation metrics for DehazeNet and AOD-Net merely. EPDN [17] employs GAN framework and utilizes image translation for images processing, which shows an average increase of 16% in all metrics, but visual contrast reduced and the images towards darker. GCANet [18] has color deviation in dehazed images and FFA-Net remains extensive haze on the whole image. DMPHN [19] is mainly utilized for non-homogeneous haze images, which brings 6% and 15% improvement on PSNR and LPIPS compared with the first three methods, but SSIM has reduced, meanwhile, DMPHN gets haze residues at the junction of objects with predominant haze has been removed. MSBDN adopts the U-net [20] and appends the SOS module enhancement strategy into decoder for haze removal. Table 1 shows that MSBDN's dehazing records are higher than the previous algorithm in two metrics, but there is low image cleanliness, residual haze and color deviation.

The dehazed results generated by our algorithm in Table 1, we observe the model with same size connection gets result closer to GT images due to the preservation of the image color information and more visually faithful to the ground-truth. Although SSDC-GAN improves PSNR by 0.15% only, it improves by 38% on LPIPS compare to the second algorithm on O-HAZE, which means that we have reduced the feature difference between real and generated samples in the images.

Fig. 3. Comparison of the different methods on O-HAZE dataset (From left to right are Hazy, DCP, EPDN, GCANet, FFA-Net, DMPHN, MSBDN, **Our model**, GT).

Table 1. Quantitative comparisons with different methods on public datasets O-HAZE.

Methods\Metric	PSNR	SSIM	LPIPS	Runtime
DCP	15.9665	0.6496	0.384	28.68
DehazeNet	15.5421	0.6938	0.328	16.46
AOD-Net	15.6184	0.6374	0.363	3.98
EPDN	18.3350	0.7405	0.301	3.02
GCANet	18.5462	0.7456	0.292	5.84
DMPHN	19.7813	0.7171	0.273	1.34
FFA-Net	18.8655	0.7492	0.316	6.77
MSBDN	20.0987	0.7465	0.254	1.64
SSDC-GAN	20.1304	0.7387	0.184	4.47
GT	$+\infty$	1.00	0.00	$\times$

Fig. 4. Comparison of the different methods on I-HAZE dataset (From left to right are Hazy, DCP, AOD-Net, EPDN, FFA-Net, DMPHN, FFA-Net, **Our model**, GT).

For I-HAZE, Fig. 4 and Table 2 demonstrate SSDC-GAN achieves the effect closest to GT in subjective perception, and the highest index result in metrics. During training processing, the original size dataset images cannot run on the device since excessive memory requirements, and our lightweight model is capable of handling larger size images and achieving superior performance.

3.4 Ablation Experiments

To demonstrate the effectiveness and superiority of the connection principle, we conduct experiments under minor quantity of training epochs to evade other factors' affecting. Experimental results compare the performance of the two connected structures: (a) Hazy;

Table 2. Quantitative comparisons with different methods on public datasets I-HAZE.

Methods\Metric	**PSNR↑**	**SSIM↑**	**LPIPS↓**
DCP	14.5188	0.7088	0.3044
AOD-Net	16.4720	0.7982	0.2818
EPDN	15.4007	0.7367	0.2956
GCANet	13.5035	0.6727	0.3032
DMPHN	16.4720	0.7642	0.3478
FFA-Net	16.8792	0.8124	0.2678
SSDC-GAN	**18.5607**	**0.8371**	**0.2318**
GT	$+\infty$	1.00	0.00

(b) Ordinary connection model training for 1000 epochs; (c) Same size dense connection model training for 1000 epochs; (d) Ground truth.

The results, analyzed through the whole information (Fig. 5, Table 3) and the details (Fig. 6), demonstrate that the network dehazing effect after employing the same size dense connection bring improvement of 0.88% PSNR and 5.76% LPIPS increase in numerical metrics, which is more consistent with the physical model, visually.

We tested multiple methods to remove haze from images and analyzed their effect on object details such as color and contours. Some methods (Fig. 7), like AOD and EPDN, made the overall color of the image appear more obscure and confused edge information in the background. DCP and DehazeNet were not effective for multi-detail images or eliminating residual haze. GCANet and FFA-Net did not completely remove haze at edges and junctions of objects. DMPHN enhanced contrast but resulted in limited color distortion. Our method produced visually more faithful images with sharper textures and

(a) (b) (c) (d)

Fig. 5. Processed images through different connection structures.

Table 3. Quantitative comparisons with different methods on public datasets.

	Situation(b)	Situation(c)	Situation(d)
PSNR	22.64	22.84	∞
SSIM	0.77	0.77	1
LPIPS	0.312	0.295	0.000

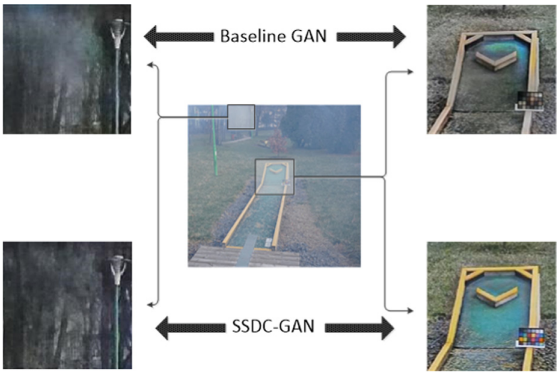

Fig. 6. The effect of connection method on image color saturation and object information recovery.

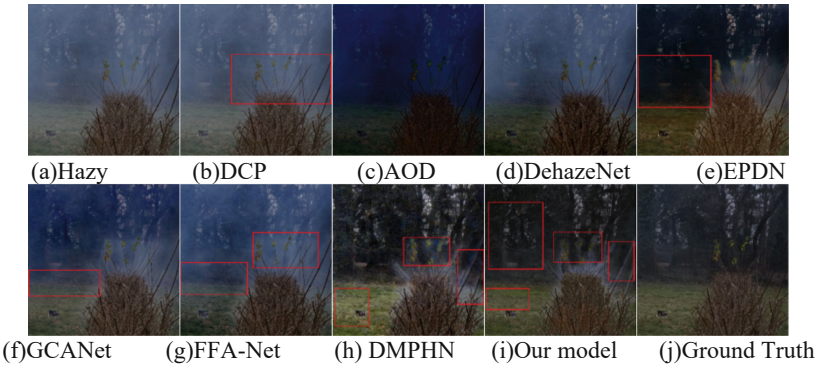

Fig. 7. Various dehazing method for complex background hazy images and object details. AOD is generally dark, DehazeNet and DCP are similar. There are no red check boxes in the above images.

better color fidelity, although some haze remained on branches due to a small number of iterations.

4 Conclusion

In this paper, we propose a dense connection method that transfers the image information from the encoder to the decoder to preserve the object details and color information of the dehazed images. Simultaneously, a depth discriminator is employed to balance the complexity between the generator and the discriminator, so as to maintain the steadiness in the training and restrain the mode collapse and vanishing gradient.

Various state-of-the-art methods recently have been employed for comparison, and SSDC-GAN performs superior results compare to other approaches. To prove the superiority of the connection principle, we conduct a ablation experiment and demonstrate the two networks differ in their ability to recover image information under small amount

training epochs. Nevertheless, SSDC-GAN preserves unnecessary information at times. In the follow-up work, effectual attention mechanism will be applied to focus on retaining fundamental image information and eliminating unnecessary information.

Acknowledgements. National Natural Science Foundation of China (Grant/Award Number: 62006073).

Key Research and Development Plan of Hubei Province (Grant/Award Number: 2021BGD013).

References

1. M. S. Ye, S. J. Xu and T. Y. Cao, HVNet: Hybrid Voxel Network for LiDAR Based 3D Object Detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1631–1640.
2. D. Zoran, M. Chrzanowski, P. S. Huang, S. Gowal, A. Mott and P. Kohli, Towards Robust Im-age Classification Using Sequential Attention Models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9483–9492.
3. S. F. Huang, T. R. Hui, S. Liu, G. B. Li, Y. C. Wei, J. Z. Han, L. Q. Liu and B. Li, Referring Image Segmentation via Cross-Modal Progressive Comprehension. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10488–10497.
4. H. Zhang and V. M. Patel, Densely Connected Pyramid Dehazing Network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3194–3203.
5. Y. L. Zhang, Y. P. Tian, Y. Kong, B. N. Zhong and Y. Fu, Residual Dense Network for Image Super-Resolution. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2472–2481.
6. G. Huang, Z. Liu, L. V. D. Maaten and K. Q. Weinberger, Densely Connected Convolutional Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4700–4708.
7. B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng and Z. Wang, Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing (TIP)*, 28(1):492–505, 2019.
8. N. Silberman, D. Hoiem, P. Kohli and R. Fergus, Indoor segmentation and support inference from rgbd images. *In European Conference on Computer Vision*, 746–760. Springer.
9. Y. F. Zhang, D. Li and G. Sharma, Hazerd: an outdoor scene dataset and benchmark for single image dehazing. *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017.
10. A. Gaidon, Q. Wang, Y. Cabon and E. Vig, Virtual Worlds as Proxy for Multi-Object Tracking Analysis. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4340–4349.
11. C. Ancuti, C. O. Ancuti and C. D. Vleeschouwer, D-HAZY: a dataset to evaluate quantitatively dehazing algorithms. *ICIP*, pages 2226–2230, 2016.
12. C. O. Ancuti, C. Ancuti R. Timofte and C. D. Vleeschouwer, O-HAZE: A Dehazing Benchmark with Real Hazy and Haze-Free Outdoor Images. *CVPR* Workshops, pp. 754–762, 2018.
13. C. Ancuti, C. O. Ancuti, R. Timofte and C. D. Vleeschouwer, I-HAZE: A dehazing benchmark with real hazy and haze-free indoor images. *CVPR* Workshops, pp. 620–631.

14. S. Zhao, L. Zhang, S. Huang, Y. Shen and S. Zhao, Dehazing Evaluation: Real-World Benchmark Datasets, Criteria, and Baselines. *in IEEE Transactions on Image Processing*, vol. 29, pp. 6947–6962.
15. Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity. *in IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612.
16. R. Zhang, P. Isola, A. A. Efros, E. Shechtman and O. Wang, The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–595.
17. Y. Y. Qu, Y. Z. Chen, J. Y. Huang and Y. Xie, Enhanced Pix2pix Dehazing Network. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8160–8168.
18. D. Chen, M. M. He and Q. N. Fan, et al., Gated Context Aggregation Network for Image Dehazing and Deraining. *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa Village, HI, USA, pp. 1375–1383.
19. S. D. Das and S. Dutta, Fast Deep Multi-Patch Hierarchical Network for Nonhomogeneous Image Dehazing. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* Workshops, pp. 482–483.
20. Ronneberger O., Fischer P., Brox T, U-Net: Convolutional Networks for Biomedical Image Segmen-tation. Medical Image Computing and Computer-Assisted Intervention – MICCAI. 2015.