Journal of Scientific & Industrial Research
Vol. 82, March 2023, pp. 328-337
DOI: 10.56042/jsir.v82i03.71760

A Fast-Dehazing Technique using Generative Adversarial Network model for Illumination Adjustment in Hazy Videos

T M Praneeth Naidu* & P Chandra Sekhar

Department of ECE, University College of Engineering, Osmania University, Hyderabad 500 007, India

Received 13 August 2022; revised 20 September 2022; accepted 06 October 2022

Haze significantly lowers the quality of the photos and videos that are taken. This might potentially be dangerous in addition to having an impact on the monitoring equipment' dependability. Recent years have seen an increase in issues brought on by foggy settings, necessitating the development of real-time dehazing techniques. Intelligent vision systems, such as surveillance and monitoring systems, rely fundamentally on the characteristics of the input pictures having a significant impact on the accuracy of the object detection. This paper presents a fast video dehazing technique using Generative Adversarial Network (GAN) model. The haze in the input video is estimated using depth in the scene extracted using a pre trained monocular depth ResNet model. Based on the amount of haze, an appropriate model is selected which is trained for specific haze conditions. The novelty of the proposed work is that the generator model is kept simple to get faster results in real-time. The discriminator is kept complex to make the generator more efficient. The traditional loss function is replaced with Visual Geometry Group (VGG) feature loss for better dehazing. The proposed model produced better results when compared to existing models. The Peak Signal to Noise Ratio (PSNR) obtained for most of the frames is above 32. The execution time is less than 60 milli seconds which makes the proposed model suited for video dehazing.

Keywords: Depth estimation, Discriminator model, Generative adversarial networks, Generator model, ResNet

Introduction

In recent times, single and multifunctional cameras primarily known as visual sensors have been utilized to capture outdoor images and videos in wide range of applications across the globe. In this scenario, the low visual quality is the primary obstacle to completely automating operations in visual sensors for various industrial and commercial video analysis. This is mainly due to the Haze, which is one of the general visual distortions that degrades visual quality. The dry atmospheric particles such as dewdrops, aerosols, water molecules etc. that float close to the earth's surface constitute haze. Also, the particles in the atmosphere absorb and scatter the light passing through, thus images and videos captured outside are considerably degraded, exhibiting diminished brightness, distortions, and a pale colour surface. The effect of haze in natural images is shown in Fig. 1.

The haze model is defined by Eq. (1). Let the input image $I(x)$be

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

where, $I(x)$ represents image with haze, $J(x)$ represents an image which is clear, $t(x)$ represents the transmittance of the scene, A is the global environmental illumination.

The scene transmittance is defined by Eq. (2)

$$t(x) = e^{\beta \times (d(x))} \quad \ldots (2)$$

where, β represents the coefficient of scattering and $d(x)$ represents the depth of the scene.

Unlike the other opaque things, for instance, haze hinders the human sight and it might be a problem for machine vision systems. As a result, a significant number of researchers have been interested in dehazing system.[1,2] Dehazing methods are used in

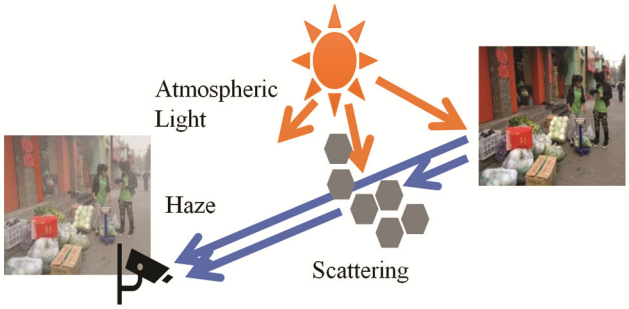

Fig. 1 — Haze in natural images

*Author for Correspondence
E-mail: praneethtm@gmail.com

different applications, including target identification, intelligent transportation systems, underwater imaging and remote sensing, etc.[3–5] Consequently, the creation of dehazing models has become a problem of utmost interest and relevance.

Although great progress has been achieved in the dehazing of single images, researchers suggest that video dehazing algorithms may produce superior results by using the temporal redundancy of adjacent frames. Predicting images from a video series requires building an internal model of the image's development and, to some extent, its content and dynamics. Pixel-space video prediction may be used for unsupervised feature learning. Due to the difficulties of gathering datasets for hazy videos, most of the research is carried out on image dehazing. Collecting real-time video datasets for deep learning models is difficult but worthwhile.[6] Moreover, because of the recent developments in the fields of image processing and computer vision, researchers have started using deep learning models for video dehazing problems.[7] It is crucial to highlight that considerable developments in the mainstream of deep learning have paved the way for significant improvements in vision hindered by hazz.[8]

Literature Survey

To improve vision quality and to reduce haze, histogram equalization and gamma correction technologies are being developed. These essential techniques did not take hazy thickness into consideration, and hence they were inefficient. A variety of algorithms have been used to handle this issue of image dehazing, however, these algorithms cannot process videos. Dehazing a video is a significant job that has been the subject of substantial study. A video sequence has several temporal changes. A shifting scene video has varying scene depth, and frame-by-frame dehazing generates visual flickers. There should be no spatial discrepancies in the repaired video. Inner-frame discontinuities may result from native video dehazing. Video data is far bigger than image data, and there are no effective solutions for dealing with several frames at once. Further, the video dehazing is critical for outdoor monitoring, and its primary hurdles are spatial-temporal coherence and computing speed.[9–11] Adapting an algorithm for dehazing images to work on video is a difficult challenge. The approach for dehazing videos has to be able to effectively change a significant proportion of a video's pixels and frames. Dealing with the temporal incoherence of the frame in the video sequence is an essential step that must be taken. In spite of this, there has only been a little amount of work done to investigate video dehazing. To overcome these challenges, it is necessary and important to understand more about video dehazing methods.

An effective real-time video dehazing method has been proposed using the transmission learning and spatially-temporally coherent regularization conserves temporal coherence and conquers visual objects directly without any interactive inputs or additional information.[12] A novel encoder-decoder based deep learning technique for video dehazing is demonstrated that uses temporal coherence of successive transmission maps.[13] Further, a progressive video dehazing fuses alignment and restoration method using deep learning stages nearby frames without optical flow estimation and a refinement network to increase the network's dehazing performance.[14] A quick and accurate video dehazing approach is proposed to decrease computing complexity while maintaining the impact of video dehazing.[15] A video dehazing network based on the temporal consistency of subsequent video frames mostly utilized for object recognition in foggy environments is proposed.[8,16]

An optical stream technique is used to increase the haze/fog transmission outline, which demands a high computational quality and difficulty for continuing real-time applications. This study helps to highlight the spatial-transient concordance, which is emphasized by the use of the spatial-transient concordance strategy.[17] A solution for real-time computer vision applications capable of dehazing videos using a quick and effective technique that may decrease the amount of processing complexity need was simulated. The suggested approach was first developed and validated for single photos, and it was subsequently expanded for use with real-time videos.[4] DehazeNet, is one deep learning-based technique uses a Convolutional Neural Network (CNN) that accepts hazy photos as input and outputs transmission maps.[18] Also, a dehazing model using deep learning and global atmosphere light, transmission maps, and the atmospheric scattering model to dehaze the physical hazing processes is presented.[19]

An efficient dehazing algorithm should dehaze the frame very quickly. Minimum required speed should be less than 60 milli seconds. The existing Video

Dehazing models use non-deep learning architectures to achieve the required speed. Thereby losing the efficiency obtained when using deep learning models. The process of image-to-image translation changes images from one domain to another. One such deep learning algorithm that performs image-to-image translation is Generative Adversarial Networks (GANs).

Generative Adversarial Networks (GANs)

Ian Goodfellow of Google Brain scientists introduced the GANs in 2014 as a novel generating model.[20] GANs can be used for learning in both semi-supervised and unsupervised conditions. They are able to do this via the implicit modelling of high-dimensional data distributions. GANs are constructed from two distinct models, which are referred to as a generator and a discriminator, respectively. The first is the generator, whose role is to generate images from random input. The second is the discriminator, which classifies the output of generator images into real or fake.

It is usual practice to realize these two models via the use of neural networks; however, they might be realized through the application of any kind of discrete system that transfers information from one space to the other. GANs are now the subject of much research because to their vast application potential, such as image, video and language processing, etc. The evolution of video surveillance and processing enables researchers to use machine learning in innovative ways. As an effective solution to the issue of video processing in practical systems, the video dehazing system has drawn the interest of researchers in this area.[21] The structure of GAN is shown in Fig. 2.

A hazy video is periodically coherent and the most effective video dehazing techniques employ information from adjoining frames to predict transmission maps from the video sequence. The introduction of GAN as a novel learning model has overcome several drawbacks in the existing image transformation techniques and the accuracy of the prediction also increases when trained with an adequate number of images. GANs excel at tough tasks such as generating realistic image.[22] The GAN is trained to predict video frames which are clear and classified by the discriminator. Except for the final frame, all of the frames are real images. The benefit is that the discriminator may efficiently utilize temporal dimension information while also assisting in making the created frame consistent with all preceding frames.[23] Further, the Multistage GAN (MD-GAN) concept involves predicting video frames from a given first frame picture. In the first step it generates a time-lapse video with realistic material and the second step optimizes the first stage's findings, adding dynamic motion information to boost realism. Also, this model can generate realistic 128 × 128 resolution time-lapse video using 32 frames.[24] In recent years, generative algorithms have grown in popularity and relevance due to their wide variety of real-world applications. Machine learning algorithms are classified into two types: discriminative algorithms and generative algorithms. GANs are a kind of algorithm that belongs to the generative algorithm group. The phrase "generative models" refers to approaches that explicitly or implicitly reflect input and result distributions.[24] However, many deep learning methods based on GAN have been applied to image dehazing in the past few years. It is important to study the recent developments in the GAN based deep learning methods for image and video dehazing.

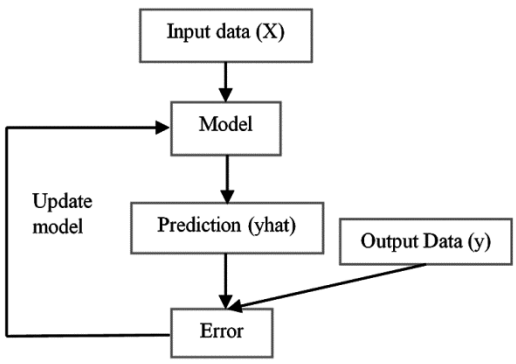

Fig. 2 — Structure of GANs

GANs have been successfully applied to photographs, and the resulting images have a resolution of 1024 × 1024. It is also conceivable to apply GANs in the video domain due to the fact that a video is just a series of pictures. However, the fact that a video consists of many pictures rather than a single image is the primary obstacle in the process of synthesizing films. A video is a kind of multimodal data since it combines many types of information, such as speed, movement, still images, and audio. In addition to these factors temporal dependence also contributes to the difficulty of producing videos. Despite the fact that working with video in GANs may be more difficult than working with pictures, several research papers have already used GANs to video datasets.[25,26]

Experimental Considerations and Methodology

Proposed GAN model for Video Dehazing

The proposed video dehazing algorithm using GAN model is developed to be fast and efficient. The model is build keeping in mind the following points:

a) The Generator Model Should be Simple So That it Can Operate Very Fast: In a GAN model, the Generator is trained to generate fake samples which look like the real images. In this application, the generator is trained to generate haze free images from a hazy image. While testing the real-time images/video frames, the trained generator model is used and not the discriminator. For this reason, in the proposed work, a simple discriminator model is used rather than a complex one to reduce the runtime complexity and time.

b) The Discriminator Model Should Be Complex: As discussed above, because the generator model is kept simple, the discriminator model is made very complex so that the proposed model is able to generate haze free images much faster. As the discriminator is used only to train the generator, it does not affect the speed of the executing during testing.

c) Train Separate Models for Varying Hazy Conditions: In the proposed approach, as the generator model is kept simple, there are changes that the trained model cannot handle different hazy weather conditions all at once. Hence, in the proposed approach, by using the scattering coefficient, images with different hazy conditions are created and different models are saved for each haze condition. This way, the less complex generator would be able to work effectively during runtime.

d) Estimate the Haze in the Input Video Using Monocular Depth Estimation: As there are several models generated during training, the input video frame must first be assessed to check the amount of haze. This is performed using a haze estimation model. The haze in the input video is estimated using a ResNet model trained for monocular depth estimation. More the haze in the input video frame, lesser the depth. This principle is utilized to estimate the haze.

e) Select the Appropriate Model To Dehaze the Input Video Feed: Once the haze in the proposed model is estimated, the appropriate trained model can be used to perform dehazing.

This section presents the details of image haze estimation model followed by the generator and discriminator models used in the proposed technique. The next section presents the proposed GAN model architecture.

Input Frame Haze Estimation

The first stage in deducing scene geometry from 2D photos is depth estimate. Monocular depth estimation from a single input RGB image is obtained by forecasting the depth information of each pixel. The depth in the image is inversely proportional to the amount of haze present in the scene. More the haze in the input scene, lesser the depth that can be extracted from the scene. A trained depth estimation ResNet Model is used to estimate the haze in the input frame by estimating the depth.[27] Based on the estimated depth, the haze coefficient in the input frame is approximated. An output image obtained by the trained ResNet model is shown in Fig. 3.

Generator Model

The generator model (Fig. 4) is kept simple to work fast in real-time video dehazing. It consists of an input layer, a convolution layer, PReLu layer, Batch normalization layer, element wise sum and residual blocks. These layers are described briefly here:

- ***Input Layer:*** The input layer collects the frames as input. The size of the input frames is 224 × 224.
- ***Convolutional Layer:*** The convolution layer contains a series of filters or kernels that can be learned. They have small receptive fields. Each convolution kernel has parameters such as kernel size, padding, and stride. Internal product operations are performed from the image's top left corner to extract high-level image features. This layer is primarily used for feature extraction from the input data.

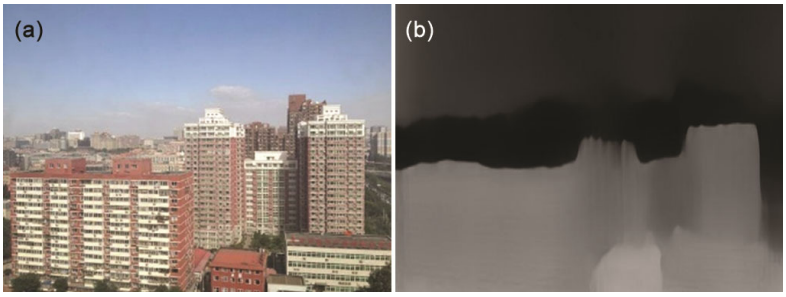

Fig. 3 — Monocular depth estimation: (a) Input Image (b) Depth Estimation output

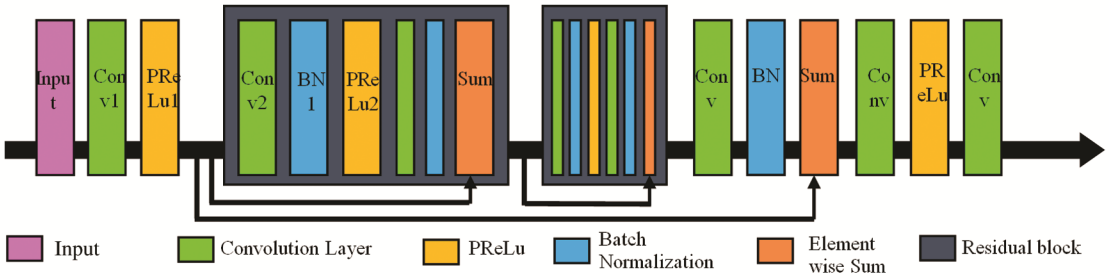

Fig. 4 — Generator model

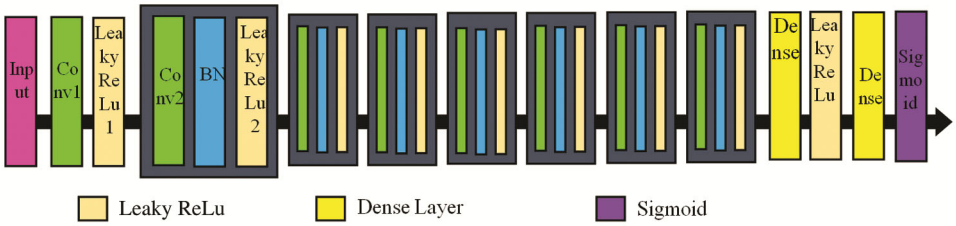

Fig. 5 — Discriminator model

- ***Activation Layer:*** The Parametric Rectified Linear Unit (ReLu) activation function introduces nonlinearity in the network. The result of the convolution operation is passed through the ReLu activation function. Therefore, the value in the final feature map is not a simple linear relationship. In the proposed model, PReLu is used. Unlike ReLu function which makes the negative values zero, a slope is added to the negative values while passing to the output.
- ***Batch Normalization:*** The output from the convolution layer can have values in a wide dynamic range. The batch normalization layer is added after the convolution layer to normalize the filter responses and bring the data to a uniform range.
- ***Element Wise Sum:*** This layer collects data from two different layers and adds the output.
- ***Residual Block:*** A residual block adds an alternate path to the layers in the deep learning model. This alternate path can bypass the function of a set of layers and directly connect two layers separated by the set of layers. This helps in eliminating the vanishing gradient problem.

Input frame is first fed to the input layer. This is followed by a convolution layer to extract the features. The extracted features are fed to a PReLu layer which acts as an activation function. This is followed by two residual blocks. The first residual block consists of size layers, Convolution, batch normalization and PReLu followed by Convolution, batch nor and element wise sum. This residual block is repeated one again. This is followed by a convolution, batch normalization and element wise sum. This element wise sum block adds the output of PReLu1 and residual block 2. The convolution and PReLu layers are added and finally, the model ends with a convolution layer. The proposed architecture of the generator is kept simple to reduce the runtime complexity. The image generated by the generator is sent the discriminator.

Discriminator Model

The discriminator model (Fig. 5) uses a complex structure to train the generator better. The discriminator consists of Input layer, convolution layer, leaky ReLu batch normalization, dense layer and sigmoid.

Leaky ReLu: Like PReLu, the leaky ReLu also creates a slope for negative values in the input, but the slope is smaller than that of a PReLu.

Dense Layer: A dense layer is a fully connected neuron structure where each neuron from the previous layers is connected to the present layer. This layer is generally used in classification problems.

Sigmoid: A sigmoid layer applies a function on the output in such a way that the output is bound in the interval of (0, 1). In this use case, the sigmoid function calculates the loss of the image created by the generator.

The output of the generator is fed to the input layer of the discriminator. The input layer is followed by a convolution layer. The output of the convolution layer

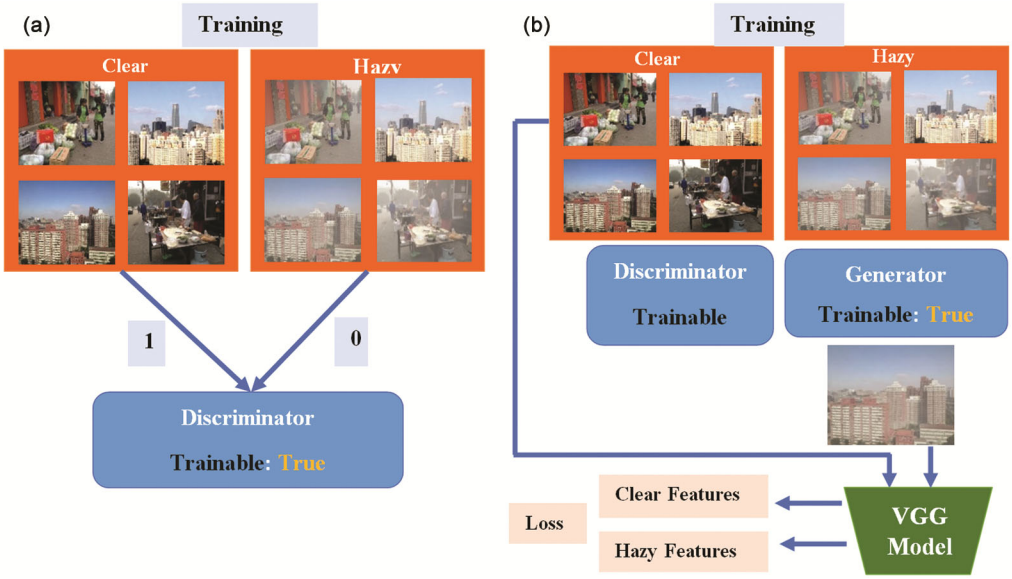

Fig. 6 — Complete working of the proposed GAN model: (a) Training the discriminator (b) Training the generator

is fed to the activation function, leaky ReLu. Based on the loss value of the discriminator, the generator model is retrained. This process is repeated until the loss is minimum.

Working of the Proposed GAN

The proposed GAN model consists of the aforementioned generator and discriminator architectures as shown in Figs 5 & 6. The discriminator model is first trained on the haze free images to make the discriminator aware of how a haze free image looks like. The discriminator is then made untrainable and generator is used to create a haze free image from a hazy input image. The generated fake haze free image is then used to compute the loss. Instead of directly calculating the loss, the trained VGG model is used to extract the features from the original haze free image and the generated haze free image. These features are used to calculate the loss and based on the loss value; the generator is trained to generate new images. This process is repeated until actual haze free images are obtained by the generator.

Step 1: Fetch a batch of hazy and clear images for training.

Step 2: The discriminator is made trainable using fake and real HR images as shown in Fig. 6 (a).

Step 3: Now, train the generator by fixing discriminator as non-trainable and create Fake images.

Step 4: Extract VGG features of clear images, to be used towards calculating loss.

Step 5: Extract VGG features of hazy free image in the generator.

Step 6: Calculate the loss in the VGG features as shown in the Fig. 6 (b).

Step 7: Iterate the steps to reduce the overall loss.

The loss functions used in the experimental analysis are:

- **Binary Cross entropy**: The binary cross entropy is a loss function commonly used to calculate binary values.

$$Binary\ Cross\ Entropy = -\frac{1}{n}\sum_{i=1}^{n} y_i \times log\hat{y}_i + (1 - y_i) \times \log(1 - \hat{y}_i) \quad \ldots (3)$$

where, $\hat{y}_i$ is the obtained output value, y_i is the actual output and n is the length of the input.

- **Mean Square Error (MSE):** The MSE value represents how close the actual output and expected outputs are.

$$MSE = \frac{1}{n}\sum_{i=1}^{size}(\hat{y}_i - y_i)^2 \quad \ldots (4)$$

where, n is the size of the input, $\hat{y}_i$ is the obtained output value, y_i is the actual output.

Based on these two loss values, the termination of the training is decided. The generator is trained until the loss is minimum.

Results and Discussion

This section presents the experimental results carried out to evaluate the proposed model. As

discussed in the proposed model, the dataset is constructed with different levels of haze. The clear images are read and haze is added using Eqs (1 & 2). The input clear frame is shown in Fig. 7 whereas Fig. 8 shows the hazy frames. The haze is added by varying the scattering coefficient. The scattering coefficient is varied from 0.25 to 1.0.

Fig. 7 — Input clear frame

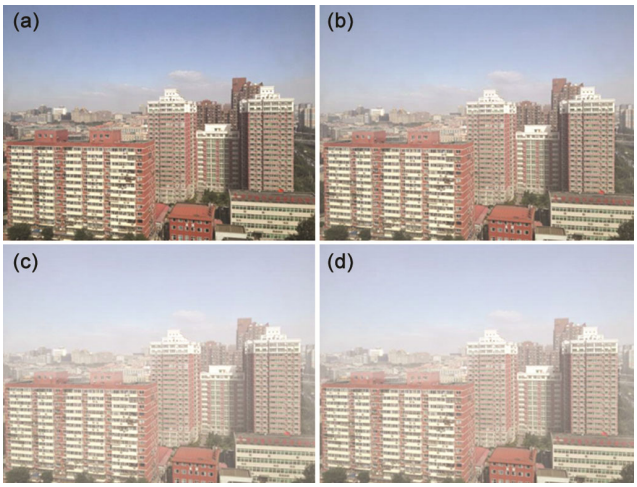

Fig. 8 — Hazy frames generated with β (scattering coefficient) varying from 0.25 to 1.0 as (a) β = 0.25, (b) β = 0.5, (c) β = 0.75, and (d) β = 1

These images are trained with separate models for each scattering coefficient. Once the models are ready, test images are read and the haze is estimated. The haze estimation results are shown in Fig. 9. The result of estimation with input haze image β = 0.25 is shown in Fig. 9(a). The estimated haze value obtained is 0.23. Based on the estimated haze value the appropriate trained model can be selected. Similarly, Fig. 9 (b) shows the results of haze estimation model where the actual β is 0.5 and the estimated β is 0.52. The results of β estimation model are shown Fig. 9 (c) where the actual β is 0.75 and the estimated β is 0.77. The results of haze estimation model (Fig. 9 (d)) shows that the actual β is 1 and the estimated β is 1.08.

The results of the proposed dehazing model are shown in Fig. 10. The input is a hazy image with β = 0.25 shown in Fig.10 (a). The image is dehazed with the proposed model after haze estimation as shown in 10 (b). The Dehazed Image obtained by the proposed model for β = 0.25 has PSNR of 30.05, the time of Execution is 56 milli seconds and the model size is 5.13 MB. The Figs (a), (c), (d) and (e) show the hazy images with β varying from 0.25 to 1.0. The images (b), (d), (f) and (h) in Fig. 10 show the dehazed images. The PSNR for these images are 30.05, 29.4, 28.8 and 27.8 respectively. The hazy images with β varying from 0.25 to 1.0 are shown in Figs 11(a), (c), (d) and (e). The images shown in Figs 11(b), (d), (f) and (h) show the dehazed images. The PSNR for these four images are 31.2, 30.14, 29.07 and 28.45 respectively.

The comparative analysis of the proposed model is given in Table 1. The existing models based on Pix2Pix, CycleGAN and DCGAN have been compared to the proposed model. The PSNR of the proposed model is 30.5 where the existing models produced a

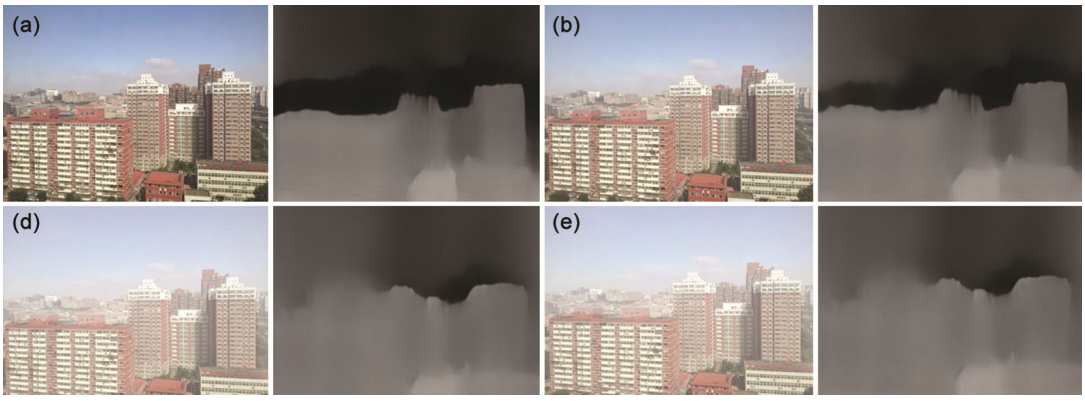

Fig. 9 — Haze estimation results with estimated haze (a) 0.23, (b) 0.52, (d) 0.77, and (e) 1.08

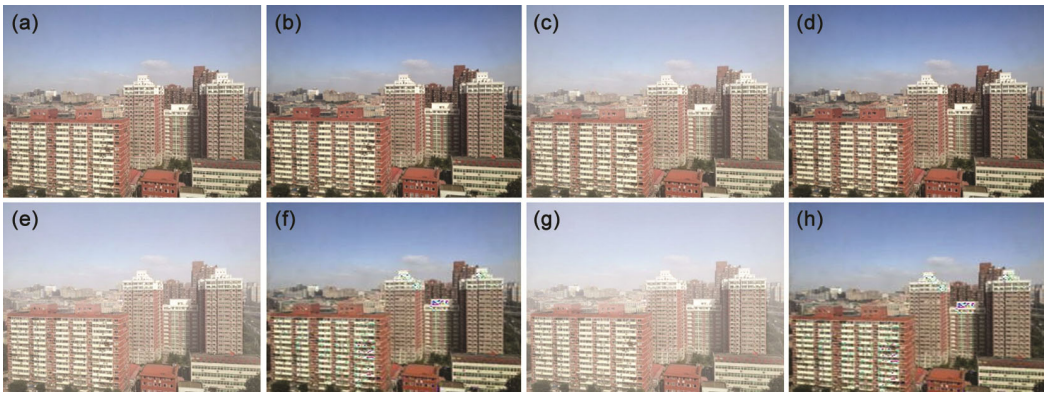

Fig. 10 — Dehazing results of images with β (scattering coefficient) varying from 0.25 to 1.0 as (a) β = 0.25, (b) dehazed image, PSNR = 30.05, (c) β = 0.5, (d) dehazed image, PSNR = 29.4, (e) β = 0.75, (f) dehazed image, PSNR = 28.8, (g) β = 1, (h) dehazed image PSNR = 27.8

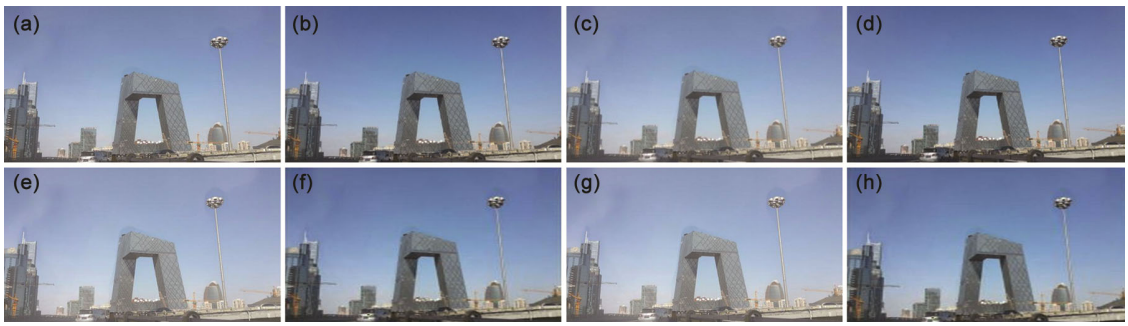

Fig. 11 — Dehazing results of images with β (scattering coefficient) varying from 0.25 to 1.0 as (a) β = 0.25, (b) dehazed image, PSNR = 31.2, (c) β = 0.5, (d) dehazed image, PSNR = 30.14, (e) β = 0.75, (f) dehazed image, PSNR = 29.07, (g) β = 1, and (h) dehazed image, PSNR = 28.45

Fig. 12 — Video dehazing results for (a) input video, (b) hazy video with β = 0.6, (c) result of existing method, average PSNR: 22.5, (d) result of proposed method, Average PSNR: 26.8

Table 1 — Comparative analysis

Parameter	Pix2pix[28]	CycleGAN[29]	DCGAN[30]	Proposed method
PSNR	21.78	21.98	21.82	30.05
SSIM	0.8557	0.8573	0.8567	0.9812
Model Size	207 MB	204 MB	204 MB	5.13
Execution time per image (CPU)	8 seconds	8 seconds	8 seconds	3.5 seconds
Execution time per image (GPU)	—	—	—	55 ms

PSNR of 21.78, 21.98 and 21.82 respectively.[28–30] The SSIM of the proposed model is 0.9812, whereas the existing models obtained 0.8557, 0.8573 and 0.8567 respectively. The model size obtained by the proposed model is very less compared to the existing models, which is only 5.13 MB. The execution time of the proposed model is 3.5 seconds on CPU and 55 ms on GPU.

The proposed model is used to dehaze a video as shown in Fig. 12. The Fig. 12 (a) shows a frame of

clear video. Fig. 12 (b) shows the haze added frame with β = 0.6. Fig. 12 (c) shows the result of existing method shown in with a PSNR of 22.5.[31] Fig. 12 (d) shows the result of the proposed model with a PSNR of 26.8.

Conclusions

In this study, a fast video dehazing method based on the Generative Adversarial Network (GAN) model is presented. The depth of the scene is retrieved using a pre-trained monocular depth ResNet model to estimate the amount of haze in the input video. A suitable model is chosen and trained for the unique hazy circumstances based on the quantity of haze. The generator model is kept basic in the proposed study to obtain results more quickly and in real-time. To increase the generator's effectiveness, the discriminator is kept complex. VGG feature loss is used in place of the conventional loss function to improve dehazing. When compared to other models, the suggested model gave superior outcomes. The majority of the frames' PSNR is higher than 32. The suggested model is suitable for video dehazing since the execution duration is less than 60 milli sec.

References

1 Ullah H & Mehmood I, Real-time video dehazing for industrial image processing, *13th Int Conf Softwar Knowl Inform Manag Appl* (IEEE) 2019) 19–24.

2 Wang Z, Video Dehazing Based on Convolutional Neural Network Driven Using Our Collected Dataset, *J Phys Conf Ser*, **1544(1)** (2020) 012156, doi: 101088/1742-6596/1544/1/012156

3 Crebolder J M & Sloan R B, Determining the effects of eyewear fogging on visual task performance, *Appl Ergon*, **35(4)** (2004) 371–381, doi: 101016/japergo200402005

4 Soma P & Jatoth R K, An efficient and contrast-enhanced video de-hazing based on transmission estimation using HSL color model, *Vis Comput*, **38(7)** (2021) 2569–2580, doi: 101007/s00371-021-02132-3

5 Singh D, Garg D & Pannu H S, Efficient Lands at image fusion using fuzzy and stationary discrete wavelet transform, **65(2)** (2017) 108–114, doi: 101080/1368219920171289629

6 Sakaridis C, Dengxin D & Luc V G, Semantic foggy scene understanding with synthetic data, *Int J of Comput Vis*, **126(9)** (2018) 973–992.

7 Singh D, A Comprehensive Review of Computational Dehazing Techniques, *Arch Comput Methods Eng*, **26** (2019) 1395-1413**.**

8 Li B, Peng Xi, Zhangyang W, Jizheng Xu & Dan F, Aod-net: All-in-one dehazing network, in *Proc IEEE Int Conf Comput Vis* (IEEE) 2017, 4770–4778, doi: 101109/ICCV2017511

9 Kaiming He, Sun J & Xiaoou T, Single image haze removal using dark channel prior, *IEEE Trans Pattern Anal Mach* Intel,l **33(12)** (2010) 2341–2353.

10 NishinoKo, Louis K & Stephen L, Bayesian defogging, *Int J Comput Vis*, **98(3)** (2012) 263–278, doi: 101007/s11263-011-0508-1

11 Sharma N, Kumar V & Singla S K, Single Image Defogging using Deep Learning Techniques: Past, Present and Future, *Arch Comput Methods Eng*, **28(7)** (2021) 4449– 4469.

12 Peng S J, Zhang H, Liu X, Fan W, Zhong B & Du J X, Real-time video dehazing via incremental transmission learning and spatial-temporally coherent regularization, *Neurocom puting*, **458** (2021) 602–614, doi: 101016/jneuco m202002134

13 Ren W, Zhang J, Xu X, Ma L, Cao X, Meng G & Liu W, Deep Video Dehazing with Semantic Segmentation, *IEEE Trans Image Process*, **28(4)** (2019) 1895–1908, doi: 101109/TIP20182876178

14 Runde Li, Progressive deep video dehazing without explicit alignment estimation, *arXiv*, (2021).

15 Zhong L, Hao Z, Yuanyuan S, Zhuhong S & Hui D, Fast video dehazing using per-pixel minimum adjustment, *Math Probl Eng* (2018) 1–8.

16 Boyi Li, Xiulian P, Zhangyang W, Jizheng Xu & Dan F, End-to-end united video dehazing and detection, in *Proc AAAI Conf Artif Intell*, **32(1)** (2018).

17 Zhang J, Liang Li, Yi Z, Guoqiang Y & Xiaochun C, Video dehazing with spatial and temporal coherence, *Vis Comput*, **27(6)** (2011) 749–757, doi: 101007/s00371-011-0569-8

18 CaiB, Xiangmin Xu, Kui Jia, Chunmei Q & Dacheng T, Dehazenet: An end-to-end system for single image haze removal, *IEEE Trans Image Process*, **25(11)** (2016) 5187–5198.

19 Xie Li, Hao W, Zhuowei W & Lianglun C, DHD-Net: A Novel Deep-Learning-based Dehazing Network, *In Int Joint Conf on Neural Networks* (IEEE) 2020, 1–7.

20 GoodfellowI, Jean P-A, Mehdi M, Bing Xu, David W-F, Sherjil O, Aaron C&Yoshua B, Generative adversarial networks, *Commun ACM*, **63(11)** (2020) 139–144.

21 Feng Y, A Survey on Video Dehazing Using Deep Learning, *J Phys Conf Series*, **1487(1)** (2020) 012018, doi: 101088/1742-6596/1487/1/012018

22 Gonog, Li & Yimin Z, A review: generative adversarial networks, *14th IEEE Conf Ind Electron Appl* (IEEE) 2019, 505–510.

23 Mathieu M, Camille C & Yann L, Deep multi-scale video prediction beyond mean square error, *arXiv preprint arXiv*, (2015).

24 Xiong W, Wenhan L, Lin Ma, Wei L & Jiebo L, Learning to generate time-lapse videos using multi-stage dynamic generative adversarial networks, *Proc IEEE Conf Comput Vis Pattern Recognit* (IEEE) 2018) 2364–2373.

25 GuiJ, Zhenan S, Yonggang W, Dacheng T & Jieping Ye, A review on generative adversarial networks: Algorithms theory and applications, *IEEE Trans Knowl Data Eng* (IEEE) 2021, 23 November, doi: 101109/TKDE20213130191

26 Nuha A, Arcot S, Nadine M & Andgelareh M, Video Generative Adversarial Networks: A Review, *ACM Comput Surv*, **55(2)** (2022) 1–25.

27 Godard C, Oisin M A, Michael F & Gabriel J B, Digging into self-supervised monocular depth estimation, *Proc*

IEEE/CVF Int Conf Comput Vis (IEEE) 2019, 3828–3838.

28 Isola P, Zhu J-Y, Zhou T & Alexei A E, Image-to-image translation with conditional adversarial network, *In Proc IEEE Conf Comput Vis Pattern Recognit* (IEEE) 2017, 1125–1134.

29 Chen J, Wu C, Hu C & Peng C, Unsupervised dark-channel attention-guided cyclegan for single-image dehazing, *Sensors*, **20(21)** (2020) 6000.

30 WuM, Xin J, Qian J, Shin-jye L, Wentao Li, Guo L & Shaowen Y, Remote sensing image colorization using symmetrical multi-scale DCGAN in YUV color space, *The Vis Comput*, **37(7)** (2021) 1707–1729.

31 Tianyang D, Guoqing Z, Jiamin W, Yang Y & Ying S, Efficient traffic video dehazing using adaptive dark channel prior and spatial–temporal correlations, *Sensors*, **19(7)** (2019) 1593