Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2021) 29: 1445 – 1463

doi:10.3906/elk-2005-78

An efficient deep learning based fog removal model for multimedia applications

Gaurav SAXENA[1,*], **Sarita SINGH BHADAURIA**[2]

[1]Department of Electronics and Communication Engineering, Jaypee University of Engineering & Technology, Guna, India

[2]Department of Electronics Engineering, Madhav Institute of Technology & Science, Gwalior, India

Received: 12.05.2020 • **Accepted/Published Online:** 19.11.2020 • **Final Version:** 31.05.2021

Abstract: In the present era of technology, several applications such as surveillances systems, security and object recognitions mainly depend on the contents of an image. In this context, the hazy/foggy environment and/or other adverse climatic conditions degrade the image contents that severely influences the result of related applications. The effective haze removal from a single image decides the reliability of these systems. The convolutional neural network (CNN) based techniques are widely used among the available image dehazing methods. However, in CNN based image dehazing techniques, the robustness and accuracy of the learning models are based on the improvement of transmission estimation without giving much concern to the atmospheric light. Therefore, in this paper, the accurate and efficient deep CNN based image dehazing model, which take care the minute information elements during the learning of feature map, is proposed. Besides, the proposed model handles the hallo, blocking artifacts, retainment of fine edges, white region handling, and color fidelity problems, which are primarily responsible for image sharpening and structural stability. For the evaluation of proposed method, the extensive experiments on synthetic and real world images are performed using existing and proposed techniques. The qualitative and quantitative analysis of experimental result shows that the proposed model is more efficient over the existing prior-based and learning-based methods.

Key words: Single image dehazing, atmospheric light, convolution neural network, transmissivity, image restoration

1. Introduction

Many real world problems can be resolved using image processing and it can only be achieved when the camera provides good quality images to get accurate and reliable results. However, in many cases during the acquisition of the image from the camera, it is observed that the different foreign particles (e.g. dust, water droplets, smog etc.) available in the atmosphere create a different density scattering medium for the transmission of light [1], which may not only affect the clarity of the image, but also reduce the performance of many multimedia applications. Therefore, dehazing/fog removal techniques play a significant role in these applications [1, 2]. There are several techniques suggested in the literature to minimize the effect of adverse climatic conditions [3, 4] during the restoration of the clear image from the input hazy/foggy image.

To understand the approaches suggested in the literature for the extraction of good quality image from its hazy counterpart, the hazy image formation model (atmospheric scattering model) [5] is briefly discussed. According to this model, the depth increases in the scene, and the contribution of light radiance from object collected by the camera becomes very less, whereas contribution of the light scattered by the foreign particles

*Correspondence: gauravsagar311@gmail.com

is more. This phenomenon creates haze in the captured image. Mathematically, the formation of haze by the scattering can be expressed as:

$$I_f(x) = R_d(x) \times t_r(x) + A(1 - t_r(x)); \quad t_r(x) = e^{-\beta.d(x)} \tag{1}$$

where $I_f(x)$, $R_d(x)$, $t_r(x)$, A and β represent the intensity of foggy image, scene radiance of recovered image (clear image), refined medium transmission function, depth of the scene, global atmospheric light, and atmospheric scattering coefficient, respectively. For the restoration of the scene radiance the Eq.(1) can be rewritten as:

$$R_d(x) = \frac{I_f(x) - A}{t_r(x)} + A \tag{2}$$

The intensity value of a foggy image $I_f(x)$ is only known and other parameters ($t_r(x)$ and A) are unknown in Eq.(2). The high-performance single image dehazing system requires an accurate estimation of these parameters. Hence, any failure in the estimation of these parameters leads to unsatisfied dehazing results. Therefore, the perfect estimation of these parameters is a challenging issue. There are various dehazing/defogging techniques suggested in the literature, which can be broadly classified into two categories. First category is based on some statistical assumptions (priors) to estimate the unknown parameters where the second one is based on the machine learning framework. In the prior based method, there are several assumptions, where some are well known like the change of brightness, variation in maximum contrast, Hue disparity, dark channel prior (DCP), the color attenuation prior (CAP) etc. Tan [6] has proposed a method to maximize the local contrast with the assumption that the local contrast of hazy image is comparatively less than the clear image. Retinex algorithms [7] and image fusion methods [8], have come in the existence to improve image contrast. He et al. [3] have suggested DCP to estimate the thickness of haze with the assumption that at least one channel contains low-intensity patches of the nonsky region whose intensity is very low. This approach is found to be cost-effective in terms of intensive computations, and sometimes it gives satisfying results.

In couple of years back, some techniques come into existence, which introduces learning based strategies to compute the coefficients of the assumption prior models. Zhu et al. [1] have proposed a fast algorithm based on CAP to restore the dense foggy images. In this method, a linear model has been developed to calculate the scene depth by using the statistics of brightness and saturation. Subsequently, Raikwar and Tapaswi [4] have proposed an improved linear depth model in which they have included the hue factor along with brightness and saturation, which is considered to improve the linearity of scene depth and computes linear coefficients of model through learning strategies. Further, Dat Ngo et al. [9] have proposed an improved color attenuation prior for dehazing and resolving the color distortion and background noise problem by using an efficient quad-decomposition algorithm. Although, these haze-relevant prior based methods are limited to some set of assumptions, as a result, they perform well on some set of images but not equally good for all kind of images. Therefore, the heuristic planned prior might be inadequate to completely catch the inherent qualities of the foggy image.

In recent years, machine learning framework has gained the remarkable success in the form of convolution neural network (CNN) and deep CNN for the more accurate estimation of haze removal parameters over the other methods in which hand crafted feature assumptions are considered. The method of CNN produced outstanding performance in image classification, segmentation, image enhancement, and restoration [10, 16].

Tang et al. [11] developed a random forest regression based model in a learning framework to investigate the four types of haze-relevant features for image dehazing. Ren et al. [12] proposed a single image dehazing technique based on the learning of mapping relation between hazy image and its transmission maps under the multiscale framework. This technique employed a two-fold convolutional neural network: the coarse network for the prediction of holistic transmission map and fine-scale network for the refinement of the transmission map. It requires haze density based parameter adjustments for obtaining good result for hazy image with different densities. However, it fails to remove the haze for dark images. Cai et al. [13] utilized a regression in deep convolution neural network to estimate the transmission of each pixel from its neighboring patch and developed end-to-end system for image dehazing. In the architecture of end-to-end model, the early layer is used for the feature extraction and rest of the layers support the regression. However, the model is unable to retain and it forwards the previously generated feature information. To make the learning procedure easier, Song et al.[14] added a novel ranking layer into the classical CNN framework, which keeps the structural and statistical attributes together. The addition of ranking layer increases the computational complexity, and the computations involved in feature extraction from the individual pixel are redundant, since the neighboring pixels have almost the same value. Li et al. [15] introduces all-in-one dehazing network (AOD-Net), and it is designed using reformulated atmospheric scattering model and generates the clean image through a light-weight CNN. Rashid et al. [16] have suggested a trainable CNN model which utilized the series of hidden layers for the filtration of low intensity pixels present in the input hazy image, further these filtered pixels are used for the estimation of transmission map to improve the visibility of dehazed image. However, this method provides the better performance only for outdoor images with low density fog. Recently, Ren et al. [17] have proposed an improved multiscale CNN architecture for the learning of parameters to establish the mapping between the haze image and its transmission maps. In this model, the scene transmissivity is initially computed coarsely and it is refined by a fine-scale network; the edge information is preserved by using the holistic edge guided refinement. However, it fails in case of nonuniform atmospheric light.

From the literature survey, following three observations are made. First, the saturation and naturality of the dehazed image are strongly dependent on the perfect level of transmissivity estimation, and it may be achieved by the effective utilization of intermediate processed information in the subsequent processing blocks, which is not effectively performed in the existing methods. Second, the feature extraction operation involves some redundant computations due to the consideration of individual pixels, whereas the features of neighboring pixels in a small section have nearly the same value, which can be exploited to reduce the some computations. Moreover, all CNN based techniques described in the literature use an input color image with three individual color channels (RGB). In CNN, these three color channels are filtered separately with same kernel value and produces the single channel by computing the pixel-wise average of the filtered colored channels. Here, it is possible to reduce the computation complexity by using hazy image information in other mode such HSI in place of RGB. Third, it is observed that most of the researchers pay more attention to accurate estimation of transmission map, while the global atmospheric light parameter also plays an important role in the quality of dehazed image. These observations motivate us to develop the image dehazing technique which may characterizes the optical atmospheric scattering model more accurately over the existing techniques.

Based on the observations above, the deep CNN-based image dehazing model is proposed. The aforementioned model uses the cascaded filtering approach to preserve the degraded edges, Intensity channel of HSI version of the image for transmission map estimation with reduced computational complexity. Besides, the proposed model takes care the hallo, blocking artifacts, and fine edges, which are primarily responsible for

image sharpening and structural stability. The qualitative and quantitative analysis of the simulation results obtained for various foggy images belong to real world, and artificially developed (synthetic) foggy images shows the effectiveness of the proposed image dehazing model. The key contributions in this paper are summarized as:

- The proposed deep CNN model is utilizing only the intensity component of the foggy input image for the estimation of the transmission matrix, which reduces the computational complexity up to 66%.
- Logarithmic filtering is employed for the correction of a dynamic range of medium transmission, which provides the effective restoration of foggy images.
- The new method is developed by cascading the minimum and median filters for the better estimation of atmospheric light (A) to minimize the chromatic distortions in the restored images.

The rest of the paper is organized as follows: The medium transmission estimation model using deep CNN is proposed in Section 2. Section 3 discusses the training of proposed deep CNN model and its application in image recovery model. The quantitative and qualitative comparative analysis of the experimental results is discussed in Section 4. Finally, conclusion of the work is presented in Section 5.

2. Proposed medium transmission estimation model using deep CNN

The proposed deep CNN based medium transmission estimation model is shown in Figure 1. It utilizes feed-forward network and HSI color model for the more accurate estimation of medium transmission with less computational complexity. This model has four major operations: (i) input hazy image conversion in HSI mode, (ii) patch extraction and nonlinear mapping, (iii) finely scaled feed-forward filters network, and (iv) local extremum and reconstruction of transmission map. All of these are shown in Figure 1 with dashed line boxes of purple, brown, violet, and blue colors, respectively.

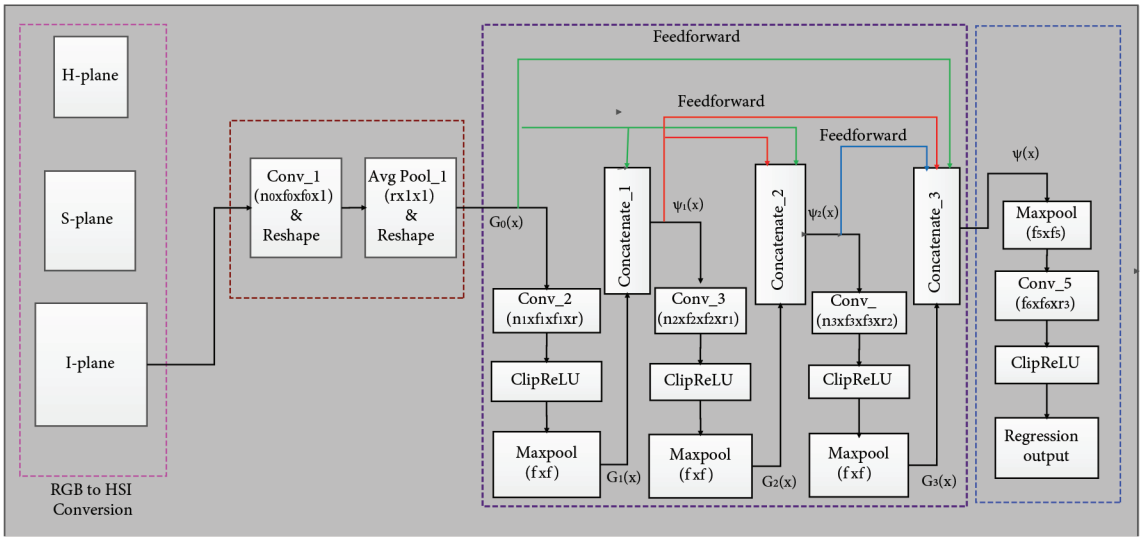


Figure 1. Proposed architecture of deep convolutional neural network.

(i) *Conversion of input hazy image in HSI mode*: Usually, the existing learning-based method uses the RGB fogy image as an input with size $(n \times m \times c)$, where n, m and c represent the number of rows, columns, and color channels, respectively. The network processes this foggy image (RGB channels) and generates estimated medium transmission map with only one single channel (i.e. $n \times m \times 1$), which contains all the required features. Thus, the processing of all three channels (i.e. RGB) does not provide any additional information. Besides, it involves more computations due to processing of these three channels. Therefore, it is possible to reduce the computational complexity by applying new strategy for the estimation of medium transmission. As we know from the coordinate geometry, whenever there is a requirement of extracting additional information, the transformation of the coordinate system is performed (i.e. Cartesian to cylindrical etc.). A similar phenomenon of transformation is available in the representation of the Cartesian RGB color model to the cylindrical HSI model [18]. The mathematical relation between RGB model and HSI model can be expressed as:

$$H = \begin{cases} 0, & \Delta = 0 \\ 60(\frac{G^{'} - B^{'}}{\Delta} \ mod\ 6), & C_{max} = R^{'} \\ 60(\frac{B^{'} - R^{'}}{\Delta} + 2), & C_{max} = G^{'} \\ 60(\frac{R^{'} - G^{'}}{\Delta} + 4, & C_{max} = B^{'} \end{cases} ; \quad S = \begin{cases} 0, & I = 0 \\ 1 - \frac{C_{min}}{I}, & I \neq 0 \end{cases} ; \quad I = \frac{R^{'} + G^{'} + B^{'}}{3} \tag{3}$$

where $R^{'} = \frac{R}{255}, G^{'} = \frac{G}{255}$, and $B^{'} = \frac{B}{255}$

$C_{max} = max(R^{'}, G^{'}, B^{'}); C_{min} = min(R^{'}, G^{'}, B^{'}); \Delta = C_{max} - C_{min}$

From the above equations, we have made the following observations:

- The hue (H) component includes majorly two out of three RGB image components and defines only the principal color information.
- The saturation (S) component value is lies between 0 to 1. It refers the amount of white light mixed with hue, which is not suitable for the brightness assessment.
- The intensity (I) component covers all the brightness information of all RGB components, and it is primarily required for medium transmission estimation.

Aforementioned observations reveals that, the input RGB image is converted to HSI mode and utilized only the I-channel as an input of the deep CNN. The utilization of a single channel saves the redundant $(f_1 \times f_1 \times n \times m \times 2)$ number multiplications and $(f_1 \times f_1 \times 2)$ number of addition in one filter operation of filter size $(f_1 \times f_1)$.

(ii) *Patch extraction and nonlinear mapping*: In this part of the CNN block, an image patch is extracted and give its required feature map using nonlinear mapping. It involves two layers, convolution followed by reshaping and average pooling followed by reshaping layers. Further, the patch of size $(f_0 \times f_0)$ is extracted using the convolution filter of size $(f_0 \times f_0 \times 1)$. To extract the haze relevant features using CNN, the input image is densely convolved by the n_0 number of appropriate filters in the first layer [21]. The n_0 maps generated from convolution layer are feed to the reshaping filter, which arranges the values

of each map in column-wise and produces n_0 number of columns. The second layer of this block consists of two sublayers, namely 'average pooling and reshaping layers', and these are used to develop the nonlinear activation function. In this layer, the batches of r number of columns from the received n_0 number of columns are selected sequentially and average pool each batch. Further, the reshaping filter is used to arrange each output column values in the two dimensions vector. In other words, this layer transforms the n_0 number of maps into r number of maps (where, $r < n_0$) by applying non linear-mapping. Based on the above discussion, the mathematical representation of the patch extraction and nonlinear mapping operations performed in the CNN model is expressed as:

$$g_0^{i,j} = W_0 * I + B_0; \quad G_0^i = \underset{j \in [1,r]}{Avg} [g_0^{i,j}(x)] \tag{4}$$

where, $*$ denotes the convolution operator, W_0 represents the filter weight matrices of size $(n_0 \times f_0 \times f_0 \times c_0)$, and B_0 shows the vector of biases or offset value of size $(1 \times n_0)$; n_0, f_0 and c_0 represent the number of filters, spatial size of single filter and number of color channels in input image I (in proposed model $c_0 = 1$), respectively. Intuitively, W_0 has n_0 number of filters each of kernel size $(f_0 \times f_0 \times c_0)$ to convolve the input image and produces n_0 number of feature maps. Further, average pooling is applied to map n_1-dimensional vectors into r-dimensional vectors.

(iii) *Finely scaled feed-forward filters network*: The violet dashed line box in Figure 1 shows the finely scaled feed-forward filters network. It has three sub-networks connected through the concatenation layers, where each subnetwork consists of convolution, clipped ReLU, and maxpool layers. These subnetworks collectively perform the sequential refinement of the individual features. The features generated from each subnetwork are retained by feed-forwarding these feature information to the next subnetwork as shown in Figure 1 by red, green, and blue colored connecting lines. Hence, each convolutional layer receives an additional input from all preceding subnetwork through feed-forward path and passes the processed output (feature map information) to all subsequent subnetworks.

The inner layer ($conv\_2$) of subnetworks is designed in such a manner that the kernel size of the convolutional network is going to be reduced (e.g. $f_1 > f_2 > f_3$) so that the outcome of each subnetwork provides more fine details of the features. The convolution layer is developed by taking stride-1 and suitable padding according to kernel size, which maintains the constant size of the feature maps. The next inner layer is a clipped rectified linear activation unit (ClipReLU), which is used to perform the threshold operation on input values. The operation of ClipReLU is mathematically expressed by the function $f(x)$ given below:

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x < t_h \\ t_h, & x \geq t_h \end{cases} \tag{5}$$

where x and t_h represent the value at the input of Clip ReLU and ceiling/threshold value, respectively. The Clipped ReLU upscales the negative input values to zero and higher input values ($x \geq t_h$) are downscale to ceiling value t_h. The estimated medium transmission map of the proposed model is expected to vary between 0 and 1. Therefore, the clipped ReLU function is employed as the activation function of each neuron by selecting the ceiling value $t_h = 1$.

The two supplementary key benefits provided by clipped ReLU are sparsity and minimization of the likelihood of vanishing-gradient effect. Moreover, the last layer of the subnetwork is the max-pooling layer. The objectives of using this layer are to just provide the nonlinearity and to extract sharp and smooth low-level features like edges, points, etc. The purpose of max-pooling is described in most of the literature as a down sampler for the size reduction of input features by 2 with the help of stride-2 [17, 19]. However, in this subnetwork, our concern is to collect the features as much as possible. Thus, we employed a max-pooling layer with the configuration of size (5×5), stride-1, and padding 'same' to maintain the size of the feature maps so that it can be easily adopted by concatenation layers in further stages. The mathematical representation of the subnetwork is as follows:

$$g_k^{i,j} = W_k^{i,j} * I^{i,j} + B_k \tag{6}$$

$$h_k^{i,j} = min(max(0, g_k^{i,j}(x)), t_h); \quad G_k^i(x) = \underset{y=\Omega(x)}{max}(h_k^{i,j}(y)) \tag{7}$$

where $\Omega(x)$ symbolizes the local window size $f_4 \times f_4$ centered at x and $*$ denotes the convolution operator; W_k and B_0 represent the filter weight matrices of size $(n_k \times f_k \times f_k \times c_k)$ and vector of offset values (biases) of size $(1 \times n_k)$, respectively, where n_k, f_k, and c_k, $j \in (1, 2, 3)$ represent the number of filters, spatial size of a single filter, and the number of input feature maps, respectively. There are three concatenation layers (Concatenation_1, Concatenation_2 and Concatenation_3) employed in the proposed model. Each concatenation layer combines the output of subnetwork(s) and feature maps coming from the feed-forward paths to strengthen feature propagation and to minimize the vanishing-gradient problem. The Concatenation_1, Concatenation_2 and Concatenation_3 layers generate $r + n_1$, $r + n_1 + n_2$ and $r + n_1 + n_2 + n_3$ number of features, respectively. The output $\Psi(x)$ of the finely-scaled feed-forward filters network block consists of $r + n_1 + n_2 + n_3$ (say p) number of features which can be expressed as:

$$\Psi(x) = concat(g_0^r, \Psi_1^{n_1}, \Psi_2^{n_2}, g_3^{n_3}) \tag{8}$$

where $g_0^r, \Psi_1^{n_1}, \Psi_2^{n_2}$ and $g_3^{n_3}$ are the outputs generated from patch extraction and nonlinear mapping block, Concatenation_1 layer, Concatenation_2 layer and maxpool layer of third subnetwork of finely scaled feed-forward filters network, respectively. This output is fed to the next processing block, which is discussed in the following subsection.

(iv) *Local extremum and reconstruction of feature map*: The problem of local sensitivity can be resolved by considering the maximum pixel value of the surrounding area. The spatial integration properties of the complex cells are employed with the help of a sequence of pooling operations for object recognition, which avoids the problem of local sensitivity [20]. Besides, the max-pooling helps to extract sharp and smooth features. Therefore, the last block of the proposed model employed the following local extremum operation on every channel of $\Psi(x)$.

$$G_4^l = \underset{y=\Omega(x)}{max}(\Psi^l(y)), \quad for \;\; l = 1, 2, 3, ..., p \tag{9}$$

where, $\Omega(x)$ is the local window of size $f_5 \times f_5$ centered at x, and G_4^l represents the processed l_{th} feature map. In the image processing, these features converge to the single feature using nonlinear regression operation. Hence, to integrate the local spatial information, the single convolution filter layer followed by ClipReLU activation function is employed in the proposed model. The non linear regression operation performed by this layer is expressed as

$$G_5 = min(max(0, W_4 * G_4 + B_4), 1) \tag{10}$$

where, W_4 is the weight matrix of size $1 \times f_6 \times f_6 \times p, B_4$ is a p-dimensional offset vector, and p is the number of channels (feature vectors).

3. Training of proposed deep CNN model and its application in image recovery model

This section discusses the training-dataset collection and generation of transmission with ground truth followed by network training methodology.

3.1. Training-dataset collection and generation of ground truth

In order to train the weights and biases of the deep learning network, it is not economical to collect a huge amount of the labeled data [21]. On the other side, it is more difficult to collect the pair of hazy and clear images of the same scene without any geometrical deflection for the training dataset. In [13], Cai et al. suggested an interesting method for training data generation using haze formation model expressed mathematically by Eq.(1). This method considered the two hypotheses: (i) the particular content of image may be seen at any depth of scene, in other words, it is medium transmission independent, and (ii) image pixel in the small patch is available at similar depth. Based on these hypotheses, we have collected the clear-scene images $R_d(x)$ of different domains (e.g. natural, buildings, city landscape etc.) from online Google image database, and these are arbitrarily sampled in 16×16 image patches $R_d^p(x)$. Using the dataset generation process shown in Figure 2, the synthesized hazy image patches $I_f^p(x)$ called 'training dataset' are generated from the collected haze-free image patches $R_d^p(x)$, random transmission map $t_r(x)$, and atmospheric light A. Here, we have considered that the value of medium transmission $t_r(x)$ is uniformly distributed and lies between 0 and 1 i.e. $t_r(x) \in (0, 1)$, and the value of atmospheric light is assumed to be fixed for the dataset generation as $A = [1, 1, 1]$.

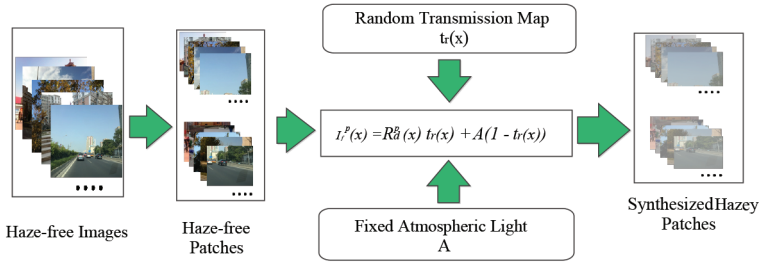


Figure 2. Block diagram of the training dataset generator.

3.2. Training methodology

The supervised learning method is used to train the proposed network by computing the weights (W_i) and biases (B_i),$\forall i \in (0,4)$ parameters. The optimum value of weights and biases are estimated by proper mapping between I-channel (of input hazy image patch) and corresponding transmission maps. The proper mapping between I-channel and transmission map is achieved by minimizing the mean square error (MSE) between reconstructed transmission $t_j^r(x)$ and the ground truth medium transmission $t_j^g(x)$. The MSE is considered as loss function and it is defined as:

$$L(t_j^r(x), t_j^g(x)) = \frac{1}{N}\sum_{j=1}^{N} \parallel t_j^r(x) - t_j^g \parallel^2 \tag{11}$$

where N is the number of synthesized hazy image patches in the training dataset. This loss function is used to train the proposed model. Besides, the adaptive moment estimation (Adam) method along with the standard back-propagation algorithm is utilized to train the network. The Adam optimization is a computationally efficient method developed for gradient-based optimization of stochastic objective functions. It utilizes the running averages of the gradients along with the second moments of the gradients [22]. This trained deep network will be utilized for the automatic extraction of the haze-relevant features as compared to conventional handcrafted methods. A detailed configuration and parameter setting of proposed deep network as shown in Figure 1 is summarized in Table 1. Our network includes 5 convolutional layers, 2 reshaping layers, 5 pooling layers, and clipReLU activations. This architecture consists of overall 21 layers and 26 connections.

3.3. Haze-free image recovery model using proposed trained deep CNN model

The image recovery model using proposed deep CNN is shown in Figure 3, which is used to recover the haze-free image from its counterpart hazy image. Our major contributions in this model are in the blue colored blocks. This model involves six operations (i) medium transmission estimation, (ii) atmospheric light estimation, (iii) logarithmic filtering, (iv) refined transmission map, (v) radiance recovery, (vi) post-processing. The descriptions of these operations are as follows:

(i) *Medium transmission estimation*: The medium transmission parameter is estimated using proposed deep CNN model (shown in Figure 1). This parameter is utilized for estimation of atmospheric light.

(ii) *Estimation of atmospheric light*: The atmospheric light parameter is also an important role in the perfect recovery of haze-free image. In the literature [1, 3, 23], with the consideration that the airlight is available at far from the observer means at the infinite depth or it is available on the topmost region of the image (i.e. at sky-region) and takes the brightest pixel value as an atmospheric light A. In some cases the atmospheric light is selected as maximum value of pixel among the different color planes RGB (i.e. max [RGB]) [4]. Unfortunately, in many cases the input image doesn't contain sky region or it may also contain the objects of brightest value (white). In all these cases, the assumption fails and the atmospheric light is wrongly selected which may create adverse effect in dehazing.

Table 1. The architectural details of the proposed deep CNN model.

Section	Type	Input size	Number of Filters	Filter Size	Output Size	Learnable Weight (W) Bias(B)	Padding
Patch Extraction & reshaping	Convolution & reshape	16 × 16 × 1 16 × 16 × 16	16 04	5 × 5 × 1 -	16 × 16 × 1 256 × 16 × 1	W: 5 × 5 × 1 × 16 B: 1 × 1 × 16	'same' -
	Avg.Pool &reshap	256 × 16 × 1 256 × 4 × 1	04 04	1 × 4 -	256 × 4 × 1 16 × 16 × 4	- -	- -
Finely scaled feed-forward filters network	Convolution	16 × 16 × 4	16	7 × 7 × 4	16x16x1	W: 5 × 5 × 1 × 16	3
	ClipReLU	16 × 16 × 16	01	1 × 1	256 × 16 × 1	B: 1x1x16	-
	Max Pooling	16 × 16 × 16	01	5 × 5	256 × 4 × 1	-	'same'
	Concatenation	2-inputs	-	-	16 × 16 × 4	-	-
	Convolution	16 × 16 × 20	16	5 × 5 × 20	16 × 16 × 16	W: 5 × 5 × 20 × 16	2
	ClipReLU	16 × 16 × 16	01	1 × 1	16 × 16 × 16	B: 1 × 1 × 16	-
	Max Pooling	16 × 16 × 16	01	5 × 5	16 × 16 × 16	-	'same'
	Concatenation	3-inputs	-	-	16 × 16 × 40	-	-
	Convolution	16 × 16 × 40	16	3 × 3 × 40	16 × 16 × 16	W: 3 × 3 × 40 × 16	1
	ClipReLU	16 × 16 × 16	01	1 × 1	16 × 16 × 16	B: 1 × 1 × 16	-
	Max Pooling	16 × 16 × 16	01	5 × 5	16 × 16 × 16	-	'same'
	Concatenation	4-inputs	-	-	16 × 16 × 80	-	-
Local extremum & reconstruction of feature map	Convolution	16 × 16 × 80	-	7 × 7	6 × 6 × 80	W: 6 × 6 × 80 × 1	-
	ClipReLU	6 × 6 × 80	01	6 × 6 × 80	1 × 1 × 1	B: 1 × 1 × 1	0
	Max Pooling	1 × 1 × 1	-	1 × 1	1 × 1 × 1	-	-

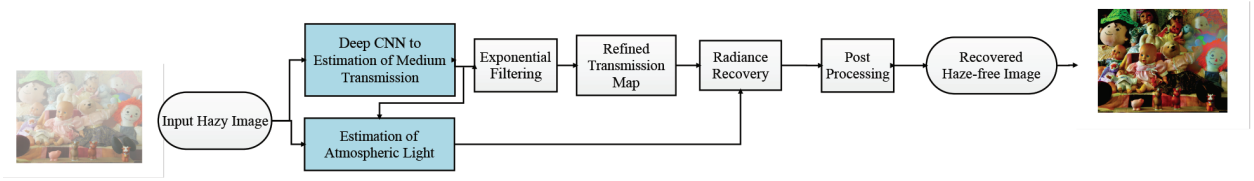


Figure 3. Complete haze-free image recovery model.

By keeping all these facts in mind, we have developed a simple algorithm for the estimation of atmospheric light using the hazy image formation model given in Eq.(1). From Eq.(1), it can be seen that if $t_r(x) \to 0$ then $I_f(x)$ is approximately equal to A. Thus, based on this approximation, we utilizes the darkest pixels of learned medium transmission $t(x)$ and input foggy image $I_f(x)$ to develop an algorithm for the estimation of atmospheric light which avoids the wrong estimation of the value A as in case of existing approaches. The proposed atmospheric light estimation algorithm (given below) employed the median filtering which provides the capability to handle the white region during the estimation of A.

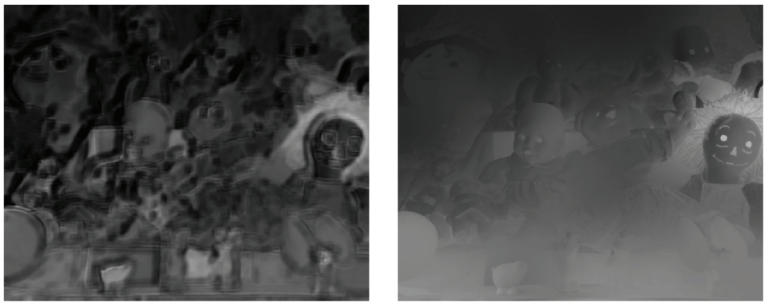
(iii) *Logarithmic filtering*: In general, the density of fog is increases from bottom to top of the hazy image due to the fact that the 2D image $(height \times width)$ captured from the camera is the projection of 3D scene $(height \times width \times depth)$. Therefore, the depth information in the estimation of medium transmission is missing in the captured image and it's extraction from single 2D image is a critical task. However, by simply viewing the image, human brain can perceived the depth on the basis of their past experiences. The most interesting aspect is to be noticed that whenever anyone can see that at the 2D projection created by the camera projects the near objects to camera in lower side and the farther ones on the top of the image. Thus, it can be considered that the depth varies from bottom to top with in the image, and, in the similar way, haze density also varies from close object to distant. Hence, the bottom part of image is not much affected than the upper or top side. Due to this type of haze distribution creates the uneven results of dehazing, all the existing methods treat the filtration process constantly throughout the complete image. The parameters responsible for the dehazing are medium transmission and atmospheric light. Thus, there is a requirement of some transformation of the intensity impact less on the dark pixel in comparison to brightest one. It is clear that the intensity transformation of the transmission map is required in the image recovery model. Therefore, we have employed the logarithmic filtering on a rough transmission map generated from the CNN model. The logarithmic filtering performs the operation of transformation by replacing the pixel value with its logarithmic value. This will create an expansion in lower range magnitudes and compression in higher range of magnitudes input image. Mathematically, the logarithmic filtering operation can be expressed as

$$\hat{t} = c \ln(1 + (e^{\sigma} - 1) \times t(x)) \tag{12}$$

$$c = \frac{255}{\log[1 + max(t(x))]}; \quad \sigma = median[t(x)] \tag{13}$$

where, c and σ represent the scaling factor of image quantization and scaling factor of the input to logarithmic function, respectively. After the transformation of the range values of transmission map the output of logarithmic filtering block is $\hat{t}(x)$ and it is send to subsequent block of refined transmission map.

(iv) *Refined transmission map*: The refined transmission map removes the problem of blocking artifacts, which is generated due to the utilization of extremum filter (last block of Figure 1) in deep CNN. Therefore, the guided image filtering is utilized to perform smoothing of an image and this filter also preserves the edges during smoothing process. In this context, the I-channel of an input image is selected as guidance image so that it can preserve the structure. The guided image filtering block receives the rough transmission map $\hat{t}(x)$ and produces the refined transmission map $t_r(x)$ by suppressing the effect of blocking artifacts. The problem of blocking artifacts and its resolved version by refined transmission map can be observed easily in the images shown in Figure 4.



(a) (b)

Figure 4. Complete haze-free image recovery model.

(v) *Radiance recovery*: After the estimation of refined transmission map $t_r(x)$ and atmospheric light A, the last step of processing is to compute the radiance with the utilization of image formation model [5] in the reverse manner. In the literature [1, 3, 4, 8], the transmission map is estimated by $t_r(x) = e^{-\beta.d(x)}$, where β is the scattering coefficient and its value is selected according to the nature of fog (i.e homogenous or heterogeneous). In many of the articles, it can be seen that researchers assumed that the atmospheric scattering is homogenous and consider the value of $\beta = 1$ but it is not always true[1, 3]. However, in proposed CNN model, the transmission map is directly computed from the input image, so scattering coefficient of β is distributed according to the nature of the atmospheric scattering. Further, the radiance recovery block in Figure 3 computes the radiance $R_d(x)$ of haze free image using refined transmission map $t_r(x)$ and atmospheric light A according to Eq.(2).

(vi) *Postprocessing*: In the post processing, the quality recovered radiance $R_d(x)$ is enhanced by adaptive histogram equalization (AHE). This operation is required for the removal of the intrinsic noise, improve the contrast, and to maintain the sharpness. During the process of AHE, the input image is subdivided into tiny section called tiles. The AHE performs operation on each tile to improve the individual contrast. Further, the bilinear interpolation technique is utilized to combine the neighbouring tiles, which eradicate the synthetically induced edges and improve the definition of edges in each area of a clear image. After the post processing, the haze free clear image is available for the different multimedia applications.

4. Simulation results and comparative analysis

The dehazing method enhances the details of visibility, texture, and edges along with preserving the image structure and color. Therefore, the quality assessment results in the fact that the different methods should

be compared to visualize the effect of visibility, color refurbishment, and structure restoration. The experimental analysis for proposed and existing state-of-the-art methods have been carried out on MATLAB-19(a) (MathWorks, Inc., Natick, MA, USA) with PC having 64 bit Intel I7 Processor, 8GB RAM and NVIDIA GeForce GTX 1650 GDDR5 4GB VRAM Graphics. To evaluate and compare the quality of image dehazing methods, the various test fogy images of different domains available in the data set of Waterloo image and vision computing (IVC) image database [24], NITRE [25] and RESIDE [26] are considered for experiments. For the performance evaluation of the proposed method, the experiment has been performed for proposed method, existing prior-based [3, 4] and deep-learning based defogging [12, 13, 15, 17] methods. The result obtained from the experiments are analysed qualitatively and quantitatively and discussed in following subsections.

4.1. Qualitative assessment

In single image dehazing method there is an unavailability of the reference image; the dehazing quality of different method can be assessed by the viewer's opinion regarding accomplishment of the objective of image dehazing. For the qualitative analysis, the nine hazy images from different categories are processed using proposed and existing image dehazing methods, and processed out images are shown in Figures 5-9. Figures 5, 6, 7, 8 and 9 show the set of foggy images belong to indoor-outdoor scene, deep-depth scene, researcher challenging and synthesized images, RESIDE dataset images along with reference images, respectively. It can be seen from Figures 5-9 that the dehazed image obtained from the proposed model has significantly better quality over the dehazed images generated by the state-of-the-art methods [3], [4], [12], [13], [15] and [17]. Besides, the image recovered from the proposed method also maintained the naturality of the image, which can be observed from Figure 5(h), where the colour of fruits and painting on table top in the first image as well as tiles of the garden in second image restored perfectly as natural.

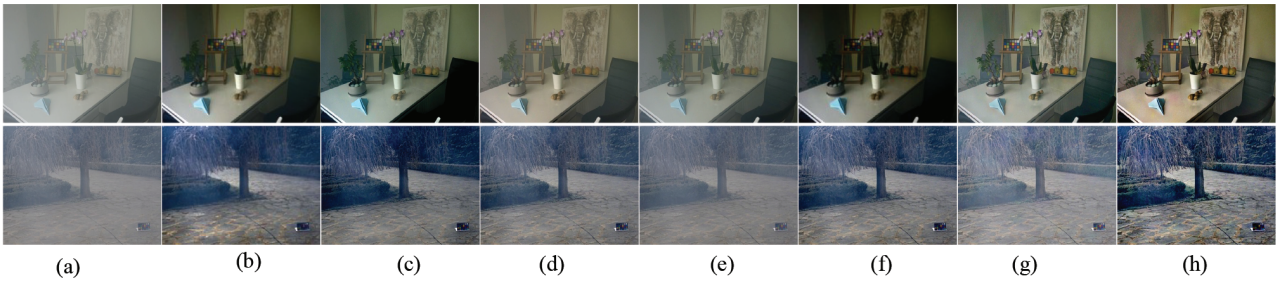


Figure 5. Comparison of the proposed with existing methods on indoor and outdoor images of NITRE dataset images. The first column shows foggy indoor-outdoor images and others represent the restored images by [3], [4], [12], [13], [15], [17] and proposed methods, respectively.

4.2. Quantitative assessment

The quantitative assessment approach refers to the performance evaluation of image dehazing methods on the basis of parameters describing the quality of image. It is broadly classified in two types namely reference and nonreference based assessments, where term reference denotes the availability of clear image along with its foggy version. In the nonreference-based assessment, input hazy and dehazed output images are utilized to evaluate the noticeable improvement in the dehazed image. The parameters frequently used for the nonreference based quantitative assessment of image dehazing methods are blind assessment discriptors (e), $(\overline{r})$, (σ) [27], image visibility measure (IVM) [28], contrast gain (CG) [29], visual contrast measure (VCM) [30], histogram

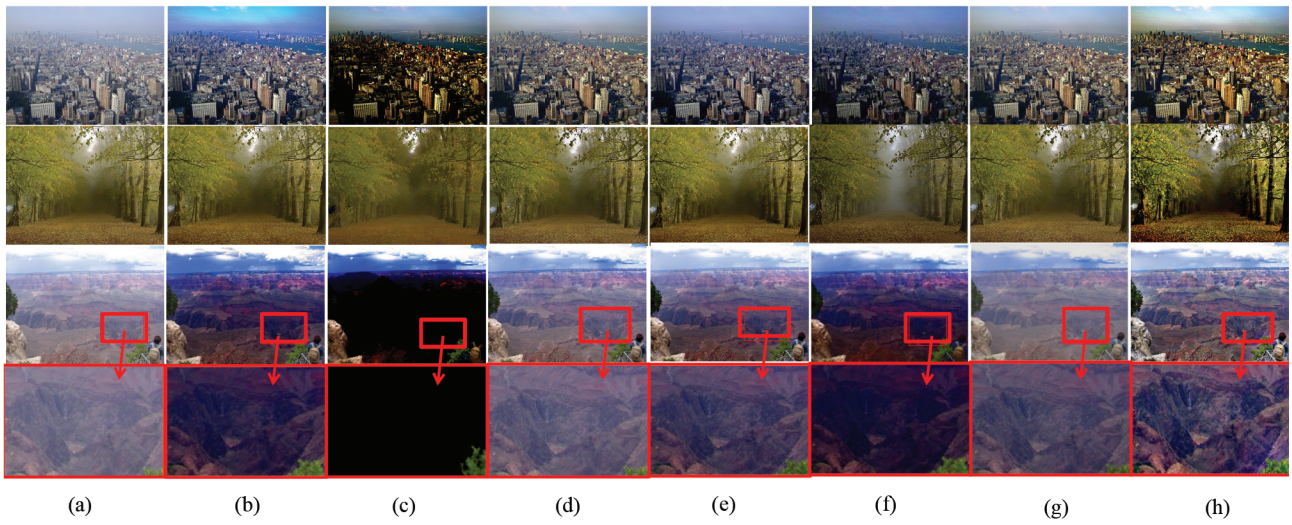


Figure 6. Visual comparison on foggy image of deep-depth scene (a) foggy image, (b–g) dehazed image using [3], [4], [12], [13], [15], [17], and (h) proposed method.

correlation coefficient (HCC) [31], and universal quality index (UQI) [32]. In general, the good quality dehazed image has larger values for e, $\bar{r}$, CG, IVM, VCM, HCC, UQI and smaller value for σ.

The simulation is carried out using the proposed and existing methods for dehazing of three set of images shown in Figures 6(a), 7(a) and 8(a). The processed images are shown in Figures (6-8) and the values of evaluation parameters are listed in the Table 2.

Figure 6(a) shows the three foggy images containing deep-depth scenes in which the fog affects deeply insight of the scene with variable density. In Figure 6(b-g), it can be seen that the existing methods produce the restored images with good quality; however, color restoration fails in some cases that leads to unnatural look. On the other hand, the restored images are generated using the proposed method, which enhances the visibility and maintain the color distribution till depth of scene, and it can also be observed from the zoomed version of selected patch (marked with red color) shown in the last row of Figure 6. The improvement in the values of quality parameters shown in Table 2 is validating the effectiveness of the proposed method.

Figure 7(a) shows three challenging foggy image of Community Park, dog pair images, and pair of girls. The fog in these images are spreaded in foreground and background with equal density. The defogged images (Figure 7(h)) obtained from the proposed method show the enhancement of foreground as well as the background of the foggy image, while the defogged images (Figure 7(b-g)) generated using the existing methods enhanced either foreground or background of the recovered. From Table 2, it is clear that the proposed method provides improvements in the quality parameters over the existing methods which verified the aforementioned observations.

The set of normally distributed synthetic foggy images shown in Figure 8 (a) are considered for the performance evaluation of the image dehazing methods. It can be observed from Figure 8 that all methods perform well but at the same time they are unable to handle the intensity of fog. The method of [3] produces good quality images over the other existing methods, which handles the intensity of fog perfectly but unable to maintain contrast up to the desired extent. Whereas, the images obtained using proposed method produce significant improvement in quality over the existing methods. The quantitative parameters of proposed method

Figure 7. Visual comparison on challenging foggy images (a) foggy image (b–g) restored image using[3], [4], [12], [13], [15], [17], and (h) proposed method.

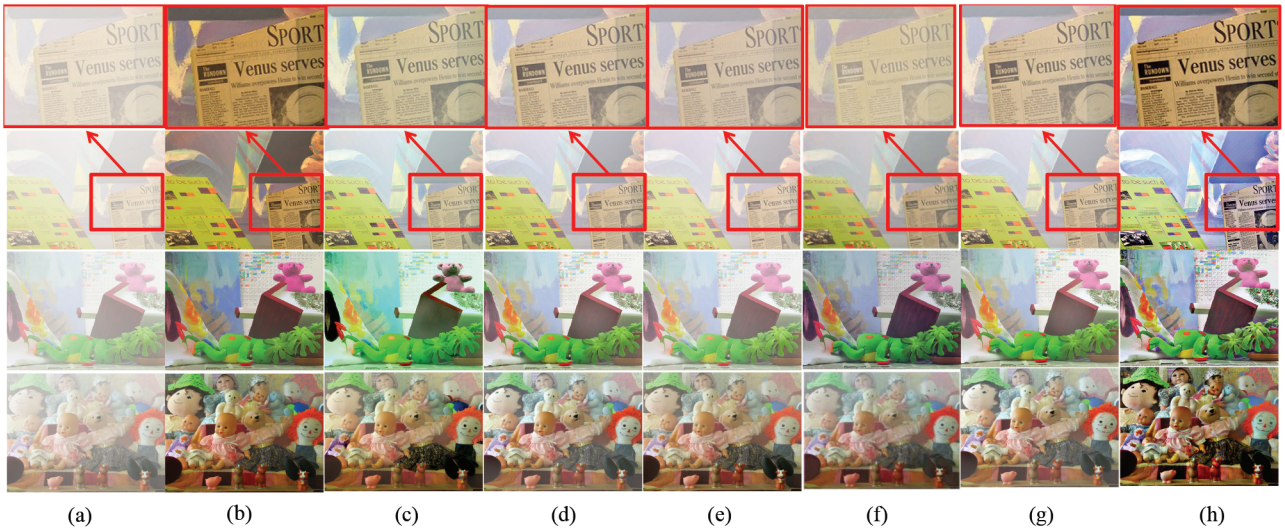


Figure 8. Visual comparison on synthetic foggy images of IVC dataset (a) foggy image (b–g) restored image using [3], [4], [12], [13], [15], [17], and (h) proposed method.

are given in Table 2, and improvement in the values validate the observations.

Moreover, to get a good insight in quantitative analysis, the reference based assessment is carried out by using an input hazy image along with its ground truth and restored output images for the computation of the peak signal to noise ratio (PSNR) and structural similarity index (SSIM) [23]. The good quality of the restored image produces a high value of PSNR and SSIM. These parameters are computed by conducting the test on synthetic foggy images available in RESIDE [26] dataset with its ground truth image shown in Figure 9. Besides, patches are extracted from the ground truth, foggy as well as restored images, and their zoomed version is shown in the last row of Figure 9. It can be observed from the zoomed images that the image restored from the proposed method provides good quality compared to the other existing methods. Further, the average values of the reference-based parameters shown in Table 3 also justify the performance of proposed method, and it appears that the proposed algorithm improves the values of PSNR and SSIM.

Table 2. The quantitative comparison of results for existing and proposed methods.

Input image	Quality assessment	Para-meters	Existing methods						Proposed Method
			[3]	[4]	[12]	[13]	[15]	[17]	
Image in row-1 (Figure 6(a))	Visibility	e	1.083	0.313	1.22	0.263	0.58	1.245	**1.597**
		$\overline{r}$	1.355	1.184	1.27	1.156	1.34	1.29	**1.411**
		CG	0.242	1.421	0.31	0.351	**0.53**	0.216	1.295
		VCM	73.654	69.230	70.12	69.038	59.78	71.923	**75.961**
		IVM	7.504	8.760	7.87	7.649	7.19	7.696	**9.175**
	Color refurbishment	σ	0.0018	0.317	0.00019	0.0017	0.00082	0.0002	**0.00016**
		HCC	0.031	-0.062	0.43	0.253	-0.090	**0.513**	-0.092
	Structure restoration	UQI	0.883	0.426	0.87	0.826	0.87	0.904	**0.921**
Image in row-1 (Figure 7(a))	Visibility	e	19.705	14.657	4.4	20.890	16.74	3.835	**22.912**
		$\overline{r}$	1.411	1.705	1.8	1.585	1.91	1.172	**2.326**
		CG	0.449	0.480	0.112	0.797	0.36	0.066	**0.632**
		VCM	44.666	52.833	36.8	48.833	38.6	37.833	**57.166**
		IVM	8.167	7.834	4.23	9.393	8.93	4.373	**9.319**
	Color refurbishment	σ	0.0086	0.0972	0.02	0.0980	**0.0063**	0.0106	0.0085
		HCC	-0.213	-0.135	**0.28**	-0.175	-0.21	0.266	-0.191
	Structure restoration	UQI	0.312	0.477	0.30	0.287	0.36	0.324	**0.487**
Image in row-1 (Figure 8(a))	Visibility	e	13.879	10.791	10.12	8.469	13.56	10.627	**13.759**
		$\overline{r}$	3.118	2.31	2.48	1.918	5.12	2.348	**4.983**
		CG	0.18	0.096	0.087	0.064	0.18	0.099	**0.261**
		VCM	45.228	50.091	30.45	50.682	34.21	30.682	**53.637**
		IVM	3.908	3.232	3.11	2.698	3.73	3.208	**4.268**
	Color refurbishment	σ	0	0	0	0	0.00001	0	**0**
		HCC	-0.273	0.351	0.298	0.573	0.56	0.318	**0.583**
	Structure restoration	UQI	0.781	0.97	0.96	**0.989**	0.86	0.971	0.89

4.3. Time complexity analysis

The time complexity of any method can be evaluated based on the number of operation performed for the execution of the method. An input image of size $(n \times m \times 3)$ has three channels which are utilized in the existing methods; whereas, the proposed method is based on the single channel (I-channel) of HSI version of input image. Hence, the proposed method involves operations on the single channel only. In general, for the computation of convolution operation on $(n \times m \times 1)$ image with $(f_1 \times f_1)$ sized filter requires $(n \times m \times f_1 \times f_1)$ multiplication and $((n \times m) \times (f_1 \times f_1 - 1))$ addition operations. In CNN model, the convolution is the main operation and aforementioned numbers of computations are repeatedly required each time. In the proposed method, the patch extraction and reshaping stages required 32 times less computations in comparison to the existing methods. The overall processing time of the proposed and existing methods of [3], [4], [12], [13], [15] and [17] are evaluated and shown in Table 4. It can be observed from Table 4 that the processing time of the

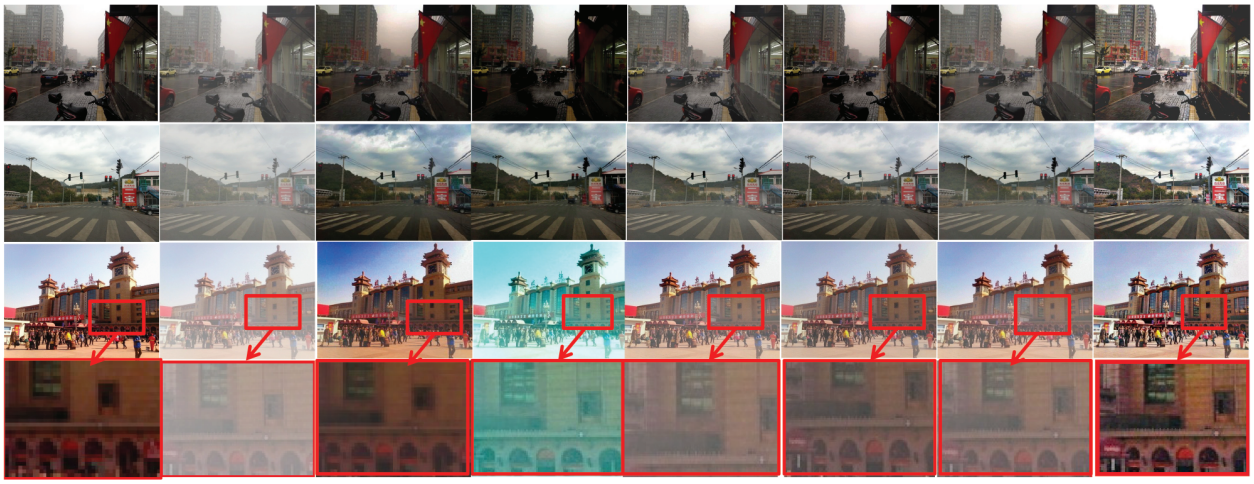


Figure 9. Visual comparison on synthetic foggy images of RESIDE dataset (a) ground-truth image (b) foggy image (c–g) restored image using [3], [4], [12], [13], [17], and (h) proposed method.

proposed method is less as compared to the other existing methods. Therefore, it can be concluded that the proposed method is faster as compared to the others existing methods.

Table 3. Reference-based quantitative comparison on image dataset [26]

Parameter	Existing methods						Proposed method
	[3]	[4]	[12]	[13]	[15]	[17]	
SSIM	0.78	0.82	0.81	0.82	0.83	0.85	**0.87**
PSNR	14.61	16.20	18.04	19.79	20.01	20.29	**21.80**

Table 4. Processing time (second) comparison of the proposed and existing methods

Size	Existing methods						Proposed method
	[3]	[4]	[12]	[13]	[15]	[17]	
$370 \times 290 \times 3$	4.8	3.5	1.31	1.1	0.83	1.3	**0.85**
$410 \times 380 \times 3$	5.01	3.6	1.36	1.25	0.91	1.4	**0.88**
$341 \times 512 \times 3$	5.2	3.8	1.4	1.8	0.98	1.6	**0.91**
$384 \times 512 \times 3$	5.5	4.0	1.5	1.82	1.09	1.63	**0.94**
$450 \times 600 \times 3$	5.9	4.2	1.68	1.85	1.14	1.72	**1.02**
$512 \times 512 \times 3$	6.1	4.5	1.7	1.87	1.24	1.79	**1.13**

5. Conclusion

In this paper, an efficient deep CNN based image dehazing model is proposed, which also takes care the minute information elements during the learning of feature map. In this model, the transmission map is estimated only using I-channel of HSI image, which helps to reduce the computational complexity. Further, an algorithm

for the accurate estimation of atmospheric light parameter is proposed to remove the white region handling problem with improved dehazing results. Besides, the proposed model handles the hallo, blocking artifacts, retainment of fine edges, and color fidelity problems, which are primarily responsible for image sharpening and structural stability. For the evaluation of proposed method, the extensive experiments on synthetic and real world images are performed using existing and proposed techniques. The qualitative and quantitative analysis of experimental results shows that the proposed model is more efficient over the existing prior-based and learning-based methods.

References

[1] Zhu Q, Mai J, Shao L. A fast single image haze removal algorithm using color attenuation prior. IEEE Transactions on Image Processing 2015; 24(11) : 3522-3533.

[2] Berman D, Treibitz T, Avidan. Single image dehazing using haze-lines. IEEE Transactions on Pattern Analysis and Machine Intelligence 2020; 42(3): 720-734.

[3] He K, Sun J, Tang X. Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence 2011; 33(12): 2341-2353.

[4] Raikwar SC, Tapaswi s. An improved linear depth model for single image fog removal. Multimedia Tools and Applications 2018; 77(15): 19 719–19 744.

[5] Narasimhan SG, Nayar SK. Vision and the atmosphere. International Journal of Computer Vision 2002; 48(3): 233-254.

[6] Tan RT. Visibility in bad weather from a single image. In: IEEE 2008 Conference on Computer Vision and Pattern Recognition; Anchorage, AK; 2008, pp. 1-8.

[7] Zhou J, Zhou F. Single image dehazing motivated by retinex theory. In: 2013 IMSNA 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation IEEE; Toronto, ON; 2013, pp. 243-247.

[8] Ancuti CO, Ancuti C. Single image dehazing by multi-scale fusion. IEEE Transactions on Image Processing 2013; 22(8): 3271-3282.

[9] Ngo D, Lee GD, Kang B. Improved color attenuation prior for single-image haze removal. Journal of Applied Sciences 2019; 9(19): 1-22.

[10] Vo AT, Tran HS, Le TH. Advertisement image classification using convolutional neural network. In: 2017 KSE 9th International Conference on Knowledge and Systems Engineering; Hue, 2017, pp. 197-202.

[11] Tang K, Yang J, Wang J. Investigating haze-relevant features in a learning framework for image dehazing. In: 2014 Proceedings of the IEEE conference on computer vision and pattern recognition; Columbus, OH, USA; 2014, pp. 2995-3002.

[12] Ren W, Liu S, Zhang H, Pan J, Cao X, et al. Single image dehazing via multi-scale convolutional neural networks. In: European Conference on Computer Vision; Amsterdam, The Netherlands; 2016, pp. 154–169.

[13] Cai B, Xu X, Jia K, Qing C, Tao D. Dehazenet: An end-to-end system for single image haze removal. IEEE Transactions on Image Processing 2016; 25(11): 5187-5198.

[14] Song Y, Li J, Wang X, and Chen X. Single image dehazing using ranking convolutional neural network. IEEE Transactions on Multimedia 2018; 20(6):1548-1560.

[15] Li B, Peng X, Wang Z, Xu J and Feng D. AOD-Net: All-in-one dehazing network. In: IEEE International Conference on Computer Vision (ICCV); Venice, 2017: pp. 4780-4788.

[16] Rashid H, Zafar N, Iqbal MJ,Dawood H, Dawood H. Single image dehazing using cnn. In: 2019 Procedia Computer Science; 147,2019, pp.124-130.

[17] Ren W, Pan J, Zhang H , Cao X, Yang MH. Single image dehazing via multi-scale convolutional neural networks with holistic edges. International Journal of Computer Vision 2020; 128(1): 240-259.

[18] Hanbury A. Constructing cylindrical coordinate colour spaces. Pattern Recognition Letters 2008; 29(4): 494–500.

[19] Huang J, Jiang W, Li L, Wen Y, Zhou G. Deeptransmap: a considerably deep transmission estimation network for single image dehazing. Multimedia Tools and Applications 2019;78(21): 30 627-30 649.

[20] Lampl I, Ferster D, Poggio T, Riesenhuber M. Intracellular measurements of spatial integration and the max operation in complex cells of the cat primary visual cortex. Journal of Neurophysiology 2014 ; 92(5): 2704-2713.

[21] Li J, Li G, Fan H. Image dehazing using residual-based deep cnn. IEEE Access 2018; 6: 26 831–26 842.

[22] Kingma D, Lei Ba J. Adam: a method for stochastic optimization 3rd international Conference learn. Representations (Preprint 1412.6980 v9), 2015.

[23] Raikwar SC, Tapaswi S. Adaptive dehazing control factor based fast single image dehazing. Multimedia Tools and Applications 2020; 79(1-2): 891-918.

[24] Ma K, Liu W, Wang Z. Perceptual evaluation of single image dehazing algorithms.In:2015 ICIP IEEE International Conference on Image Processing; 2015. pp. 3600–3604.

[25] Arad B, Ohad Ben-Shahar, Radu Timofte et al. NTIRE 2018 challenge on spectral reconstruction from RGB Images In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, 2018, pp. 1042-1049.

[26] Li B, Ren W, Fu D, Tao D, Feng D, et al. Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing 2019; 28(1): 492-505.

[27] Hautière N, Tarel JP, Aubert D, Dumont E. Blind contrast enhancement assessment by gradient ratioing at visible edges. Image Analysis & Stereology 2008; 27(2): 87-95.

[28] Yu X, Xiao C, Deng M, Peng L. A classification algorithm to distinguish image as haze or non-haze. In: 2011 Sixth International Conference on Image and Graphics; 2011. pp. 286-289.

[29] Tripathi AK, Mukhopadhyay S. Removal of fog from images: A review. IETE Technical Review 2012; 29(2):148-156.

[30] Jobson DJ, Rahman Z, Woodell GA, Hines GD. A comparison of visual statistics for the image enhancement of foresite aerial images with those of major image classes. In: Defence and security Symposium, Orlando, Florida, FL, USA; 2006.

[31] Yu J, Xu DB, Liao QM. Image defogging: a survey. Journal of Image and Graphics 2011; 16(9): 1561-1576.

[32] Wang Z, Bovik AC. A universal image quality index. IEEE Signal Processing Letters 2002; 9(3): 81-84.