Optik - International Journal for Light and Electron Optics 248 (2021) 168197

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.com/locate/ijleo

Single Image Haze Removal Based on transmission map estimation using Encoder-Decoder based deep learning architecture

Sivaji Satrasupalli, Ebenezer Daniel *, Sitaramanjaneya Reddy Guntur

Department of Electronics and Communication Engineering, Vignan's Foundation for Science, Technology and Research, Guntur, Andhra Pradesh, 522213, India

ARTICLE INFO

Keywords:
Transmission map
Airlight
Dropout layer
Deep learning
Encoder-Decoder

ABSTRACT

Haze removal is an essential requirement in autonomous vehicle applications for identifying different objects on the road. Most of the available techniques are based on different constraints/ priors. The important parameters required for recovering the ground truth from hazy image are transmission map and air light. In this paper, we proposed a learning-based Encoder-Decoder deep learning architecture for transmission map estimation. Based on the assumption that at least twenty percent of the outdoor image includes with sky region and hence airlight is calculated as average of the twenty percent brightest pixels of the image. These two parameters namely transmission map and airlight were applied in atmospheric scattering model for ground truth image recovery. In Encoder-Decoder architecture, Max pooling layer, dropout layer was used for feature learning and efficient generalization respectively. The proposed architecture was trained on different datasets like NYU Depth data set, FRIDA and RESIDE Dataset for better generalization on unseen data. Experimental results shows that the proposed method has shown better performance compared to the existing state of the art methods.

1. Introduction

Haze forms due to refraction of light from the suspended particles like smoke, dust and hence it will limit the visibility. Haze generally results in reduced contrast and as haze concentration increases red, green and blue channel values will be close to air light [1]. Autonomous vehicles could not differentiate object in such scenarios and photography is also difficult. Hence for these reasons single image dehazing is required for video processing and photography. Removing haze is an ill posed problem because of unknown depth at every pixel [2]. Several techniques were proposed to reduce the haze, which are either based on many images/Single images. For example, many images were considered under different degrees of polarization [3]. In [4], multiple images were taken under different weather conditions and applied different constrains and [5] used depth map for recovering the ground truth. It is not always practical to have multiple images, so single image dehazing was popular recently.

Image enhancement methods are also applied to restore the ground truth from hazy images [6], based on histogram equalization, [7] based on contrast enhancement, [8] used weighted least squares for contrast enhancement [9], applied modified dark channel saturation prior and [10] based on saturation. In [11] an assumption that a clear image will have high contrast compared to hazy image, and contrast enhancement method applied based on Markov Random Field (MRF) [12], applied Independent Component Analysis (ICA) but very slow and not suitable for densely hazy images. K. He et al., in [13] Presented Dark Channel Prior based on the

* Corresponding author.
E-mail address: ebydaniel89@gmail.com (E. Daniel).

https://doi.org/10.1016/j.ijleo.2021.168197
Received 27 September 2021; Accepted 16 October 2021
Available online 24 October 2021

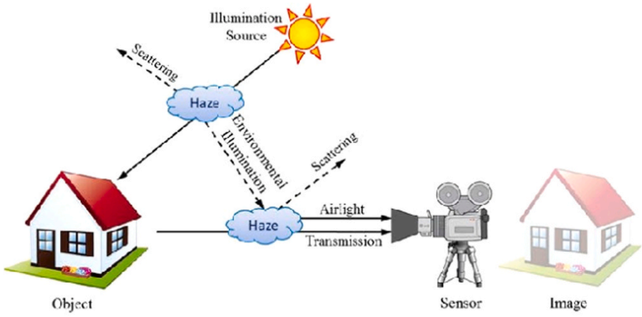

Fig. 1. Atmospheric Scattering [23].

statistical observation that the most of the clear images has at least one channel with very low pixel value and Depth of every pixel is calculated based in lowest pixel in the hazy image and recovered ground truth using atmospheric scattering model, but it is not shown better results for sky region. Later, many researchers have contributed by modifying the DCP (Dark Channel Prior). Complex and time consuming soft matting in [14] was replaced by median filtering [15], median of mean filtering [16], guided filter [17]. In [18] D. Berman et al. proposed haze lines based on the assumption that a clear image can be represented with few hundred colors, forms tight clusters turns into lines as haze concentration increases [19]. proposed color attenuation prior based on the assumption that value and saturation difference fairly represents the depth map of the image. Recently, learning based approaches were introduced by many authors. Convolutional Neural Networks have shown better performance in many high level tasks like object detection [20], classification [21] and segmentation [22], also being used for low level tasks. While priors based methods used statistical features, Learning based approaches using structural features. In [23] Cai, proposed a dehazenet for transmission map estimation [24], used All in One Deahze Net [25], used wavelet hybrid model for local and global featues.

Global atmospheric light and transmission map are two important parameters for removing the haze. In this paper we proposed an eight-layer Encoder-Decoder based Architecture for transmission map estimation and airlight was estimated based on top 20% brightest pixel average values, scene radiance was restored based on atmospheric scattering model. Experiments have shown that the proposed method is better than the existing state of the art methods. Rest of the paper organized as follows. in section II, atmospheric scattering model and related works were discussed, in section III, block diagram for reconstruction is presented, experimental results were discussed in section IV and conclusion given in section V.

2. Related works

2.1. Atmospheric scattering model

Many image dehazing methods using single image have been proposed in the literature. Some of the best solutions reviewed in this section. Although different methods proposed for single image dehazing, all are based on same function called image dehazing model.

Haze formation mathematical model was first proposed by McCartney [26] and it was further improved by Narasimhan and Nayar [27]. Haze formation mathematically represented as

$$I(x,y) = J(x,y) * t(x,y) + \alpha(1 - t(x,y)) \tag{1}$$

Where $I(x,y)$ is the input hazy image, $J(x,y)$ is the recovered image after haze removal, $t(x,y)$ is medium transmission factor, (x,y) is special coordinates of the image and Fig. 1 gives the illustration of haze formation. From Eq. (1), Scene radiance $J(x,y)$ can be re recovered after estimating the medium transmission factor $t(x,y)$ and magnitude of global atmospheric light α.

The transmission factor of the medium $t(x,y)$ can be represented as,

$$t(x,y) = e^{-\beta.d(x,y)} \tag{2}$$

Where $d(x,y)$ is the depth of the scene radiance pixel at special location (x,y) and β is the scattering coefficient of the atmosphere. Based on Eq. (2), we can notice that as depth of the object tends to infinity, results in medium transmission tends to zero and Eq. (1) becomes,

$$I(x,y) = \alpha \tag{3}$$

Practically objects cannot be located at infinite distance, so α can be estimated based on top 20% pixels average based on the assumption that sky region generally takes around 20–25% of the image and which is farthest in the image.

2.2. Dark Channel Prior

Dark Channel Prior (DCP) was presented by K He. et al. in [13] based on the statistical observation that most of the outdoor haze free images will have at least one-color channel very close to zero called as dark/minimum channel. Minimum channel, considered to

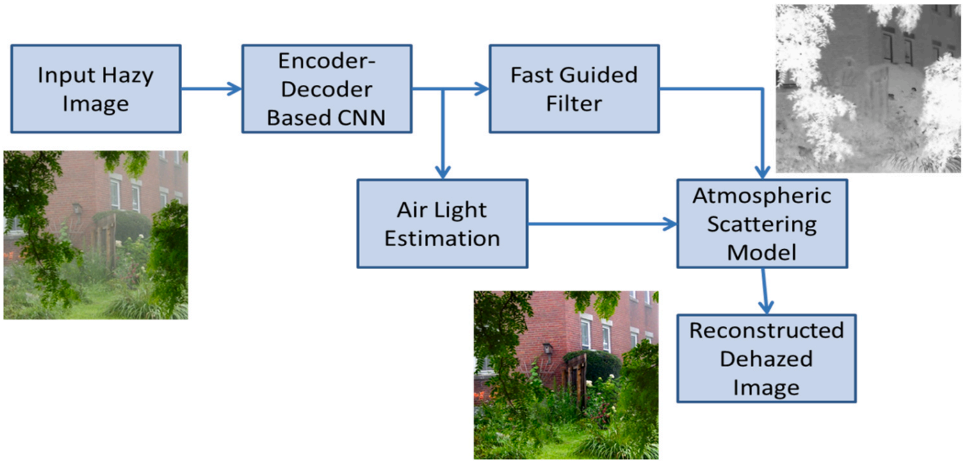

Fig. 2. Block diagram of proposed method for single image dehazing.

be very low vale and it can be calculated in a local window with size $r\ X\ r$ is:

$$D(x,y) = min_{x,y\,\in\,\Omega_r^{(z)}}\left(min_{c\in(\mathrm{r,g,b})}^{I^c}{}^{(x,y)}\right) \tag{4}$$

Where I^c is input hazy image with three channels namely red, green, and blue. Ω_r^z is the local window of size $r\ X\ r$ centered at z.as haze concentration increases, $D(x,y)$ decreases proportionally and hence transmission value decreases. Transmission map approximately calculated using Eq. (5).

$$t(x,y) = 1 - D(x,y) \tag{5}$$

2.3. Color attenuation prior

Zhu et al. in [19] introduced a color attenuation prior based on observation that the saturation decreases and value increases as haze concentration rises. Hence the saturation and value difference fairly represents the depth map.

$$D(x,y)\propto(I^V(x,y) - I^s(x,y)) \tag{6}$$

Where $I^V(x,y)$ value of input hazy image and $I^s(x,y))$ is saturation of the image.

2.4. Dehazenet

B. Cai et al. in [23] presented a DehazeNet for estimating the transmission map. DehazeNet was designed to get the haze related features like Dark channel Prior using a filter with center value as -1, maximum contrast using round filter, and color attenuation prior by transforming RGB image to HSV image. Maxpool layer was added to remove the local estimation error and BReLU (Bilateral Rectified Linear Unit) was used as activation function to limit the transmission map value to be within specified range. The model was trained on over 1,00,000 patches of size 16×16. Mean Square Error was the basis for learning the network parameters.

3. Proposed method

The proposed model is aimed to recover the haze free image by estimating the two important parameters called airlight and transmission maps. Here, Encoder-Decoder based deep learning architecture is used for transmission map estimation. Also, we chosen fast-guided filter to smoothen the transmission map by preserving the edge information. The airlight was estimated based on the assumption that sky is the farthest object in any image, generally accounts 20% area of the image, specifically in autonomous vehicle applications. Hence, we used an average of brightest 20% pixels was considered as airlight. The block diagram of proposed model is given in Fig. 2.

3.1. Network model

Transmission map estimation is very important parameter in recovering the haze free image. In this paper, we introduced a Encoder-Decoder based deep learning architecture. The neural network architecture is given in Fig. 3, which includes two layer encoder, two layer decoder and fully connected layer. The size of the hazy input image applied to the network is resized to $256 \times 256{\times}3$. Input image is convolved with 60 kernals of size 3×3 and produces an output of size $256 \times 256 \times 60$. The output of

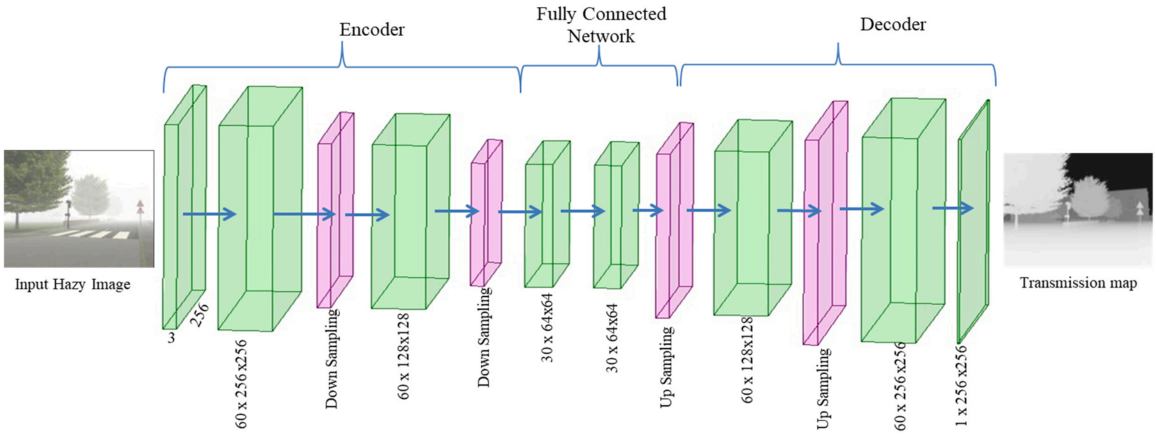

Fig. 3. CNN architecture for transmission map estimation.

the first layer is applied to down sampling layer by replacing a 2×2 matrix with its maximum values, results in output size of $128 \times 128 \times 60$ and the next layer produces $64 \times 64 \times 30$. Further upscaling was used in decoder block to obtain an output with dimensions $1 \times 256 \times 256$. ReLU activation function was used for all the layers except output layer. BReLU (Bilateral Rectified Linear Unit) was used as activation function to get values of transmission map within specified range in the output layer.

3.2. Loss function

Generally big datasets are required for training the model. We have used the data from three different datasets namely NYU2 [22], FRIDA [28] and Outdoor Training Set (OTS) of RESIDE [29] for training the network. Including Outdoor Training Set (OTS) from RESIDE dataset, around one lakh samples were used for training the network. As the training data taken from different resources, network is less prone to overfitting and resulted in better generalization. The proposed model is for estimating the transmission map, but datasets are having depth map references, so transmission map was approximated with the Eq. (7).

$$t(x,y) = 1 - d(x,y) \tag{7}$$

The network weights were initialized using Gaussian random variables and drop out layer [30] is used for regularization of the network for avoiding overfitting. The dataset is divided into training and testing sets with a ratio of 70:30. Mean Square Error between the actual output and true value was the loss function and gradient descent algorithm with Adam optimizer is used to update the weights with learning rate 0.001.

$$L(\Theta) = \frac{1}{N}\sum_{i=1}^{N} \|f(I_i;\Theta) - t_i\|^2 \tag{8}$$

The network parameters namely weight matrix and bias will be updated based on simple loss function called Mean Square Error (MSE). In Eq. (8), I_i is applied Hazy input, t_i is ground truth transmission map and Θ is network parameter.

3.3. Reconstruction

After estimating the two important parameters called transmission map and airlight, dehazed image can be calculated from the following equation by reorganizing Eq. (1).

$$J(x,y) = \frac{I(x,y) + \alpha(t(x,y) - 1)}{t(x,y)} \tag{9}$$

In Eq. (9), transmission map $t(x,y)$ is limited in the range (0.10,0.95) to avoid the color saturation problem. Fast guided filter [33] is applied to the output of Encoder-Decoder architecture to get the transmission map. Eqs. (10)–(12) are governing equations for applying fast guided filter for smoothening the transmission map by preserving the edge information.

$$t(x,y) = \overline{a_k}E_i + \overline{b_k} \tag{10}$$

$$a_k = \frac{\frac{1}{|\omega|}\sum_{i\in\omega_k}(E_i p_i - \mu_k \overline{p}_k)}{\sigma_k^2 + \epsilon} \tag{11}$$

$$b_k = \overline{p}_k - a_k\mu_k \tag{12}$$

Table 1
Average SSIM and PSNR values on validation images of NYU Depth dataset.

Parameter	CAP [19]	Cai [23]	DCP [13]	Proposed
PSNR	19.42	19.14	18.51	**19.76**
SSIM	0.8424	0.7615	0.8174	**0.8911**

Table 2
Average SSIM and PSNR values on middleburry dataset.

Parameter	CAP [19]	Cai [23]	DCP [13]	Proposed
PSNR	21.49	21.09	19.2	**21.7**
SSIM	0.8861	0.8749	0.8581	**0.9012**

Table 3
Average MSE between dehazed and ground truth of Middlebury stereo database.

Parameter	CAP [19]	Cai [23]	DCP [13]	Proposed
MSE	371.50	421.26	647.24	356.84

Table 4
Execution time (sec) taken for 256 × 256 × 3 image.

Parameter	CAP [19]	Cai [23]	DCP [13]	Proposed
Time (sec)	0.61	1.41	10.2	**1.1**


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

Fig. 4. Results on synthetic images: (a) Hazy input image (b) CAP [19] (c) Cai [23] (d) DCP [13] and (e) proposed.

Where $\overline{a_k}$, $\overline{b_k}$ are averages of a and b respectively. E_i is the output of the CNN architecture, μ_k and σ_k are the average and variance of the input E_i in the local window called k. In Eq. (9), α is called airlight & it can be estimated with the following Eq. (13).

$$\alpha = \frac{\sum_{i=0}^{n} E_i}{n} \tag{13}$$

Where n is considered as 20% of the total size of the image, which is 256 × 256 and E_i is the output of the Encoder-Decoder architecture called transmission map.

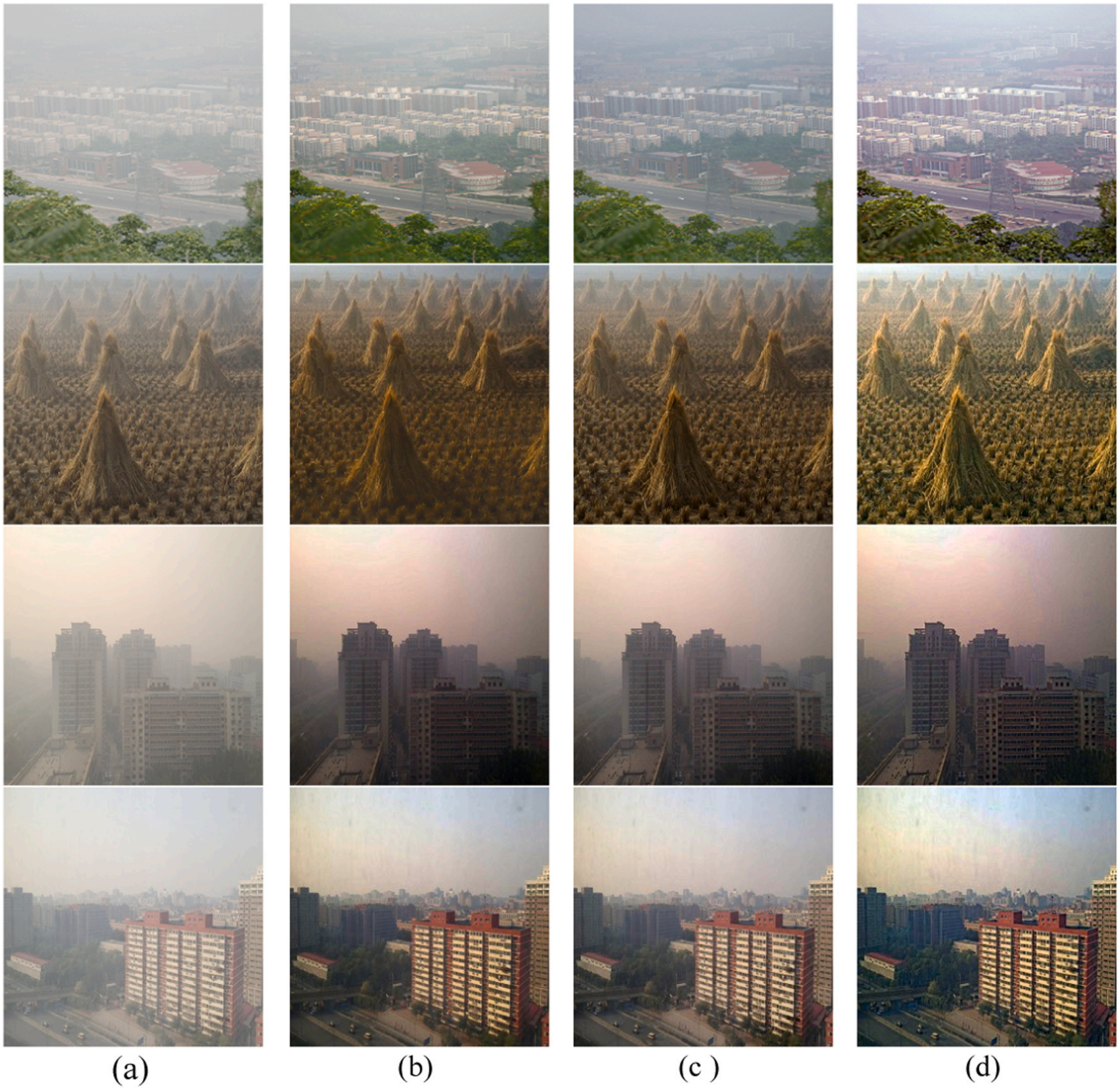

Fig. 5. Results on real world images: (a) Hazy input image (b) CAP [19] (c) Cai [23] (d) DCP [13] and (e) proposed.

4. Experimental results

4.1. Objective analysis

We have compared the proposed approach with state of the art methods: color attenuation prior, dehazenet, dark channel prior. As we have both hazy and ground truth images, enabled us to validate the proposed method based on Structural similarity Index (SSIM) [31], Peak Signal to Noise Ratio (PSNR) [31] and Mean Square Error(MSE). Table 1, Table 2 and Table 3 shows the performance comparison of the proposed model with the existing methods on validation images of NYU Depth dataset and Middleburry dataset [32]. The proposed solution is based on encoder-decoder based neural network architecture. Our model is trained on around one lakh images of size 256 $\times$ 256$\times$3 with different types of data sets helped us to achieve better performance. The following Eq. (14) [31], (15) and (16) [31] were used to calculate the performance metrics.

$$PSNR(\hat{y}, y) = 10 \log_{10}\left(\frac{255^2}{MSE(\hat{y}, y)}\right) \tag{14}$$

$$MSE(\widehat{y}, y) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\widehat{y}_{i,j} - y_{i,j}\right)^2 \tag{15}$$

$$\mathrm{SSIM}(\hat{y}, y) = \frac{(2\mu_{\hat{y}}\mu_y + C_1)(2\sigma_{\hat{y}y} + C_2)}{(\mu_{\hat{y}}^2 + \mu_y^2 + C_1)(\sigma_{\hat{y}}^2 + \sigma_y^2 + C_2)} \tag{16}$$

4.2. Execution time

The proposed method was trained in google colabs and downloaded the model, used for the reconstruction of the image in jupyter notebook of anaconda environment. Table 4 shows the comparison of the execution time taken by different methods. Our method took approximately 1.1 s for input image of size $256 \times 256 \times 3$ and it is better than DCP & dehazenet but not as fast as Color attenuation prior. All the algorithms have been executed in MATLAB (2019b) with the same computer (LENOVO Laptop with i5-3210M processor @ 2.50 GHz and 8 GB RAM) but the proposed method excuted in jupyter notebook, however hardware is same.

4.3. Subjective analysis

We have extensively tested the proposed method on both synthetic and natural images. Figs. 4 and 5 shows the comparison of the proposed method with the state-of-the-art existing methods namely color attenuation prior, dehazenet and the dark channel prior. It is noted that the results of the proposed method have shown better contrast in heavy hazy regions but partly suffers from color shifting in the sky region. As the proposed method is intended for object detection in autonomous vehicle applications, detection of the objects in the foreground of the image is important compared to sky region.

5. Conclusion

In this paper, we had proposed an Encoder-Decoder based deep learning architecture for estimating the transmission map and refined using fast guided filter. The proposed architecture was trained on approximately a lakh image with 100 epochs, batch size 500 and mean square error as loss function. After 30 epochs mean square value was hovering about 1.56e-2 and shows no improvement further and BreLU activation function in the output layer. Moreover, we had proposed a new method for calculating the airlight based on assumption that at least 20% area of the image covers sky region in the outdoor images, hence considered the average of top 20% pixel values as airlight. These two parameters were used in atmospheric scattering model and reconstructed the dehazed image and shown better performance in terms of SSIM, PSNR and MSE. We had considered constant airlight value, but it can be learned from the network for further improvement in performance.

Declaration of Competing Interest

We declare that there is no conflict of interest.

References

[1] Dai X., Yuan X., Zhang J., Zhang L., Improving the performance of vehicle detection system in bad weathers, IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, China, 2016.
[2] S.G. Narasimhan, S.K. Nayar, Vision and the atmosphere, Int. J. Comput. Vis. 48 (3) (2002) 233–254.
[3] Y.Y. Schechner, S.G. Narasimhan, S.K. Nayar, Instant dehazing of images using polarization, IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2001, pp. I–325.
[4] S.G. Narasimhan, S.K. Nayar, Contrast restoration of weather degraded images, IEEE Trans. Pattern Anal. Mach. Intell. 25 (6) (2003) 713–724.
[5] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, D. Lischinski, Deep photo: Model-based photograph enhancement and viewing, ACM Trans. Graph. 27 (5) (2008), 116.
[6] Z. Xu, X. Liu, N. Ji, Fog removal from color images using contrast limited adaptive histogram equalization, Image and Signal Processing, 2nd International Congress on. IEEE, 2009, pp. 1–5.
[7] S.G. Narasimhan, S.K. Nayar, Contrast restoration of weather degraded images, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) 713–724.
[8] Huang Z., Zhu Z., An Q., Wang Z., Fang H., Global-local image enhancement with contrast improvement based on weighted least squares, Optik. 2021.
[9] Ma R., Zhang S., An improved color image defogging algorithm using dark channel model and enhancing saturation, Optik, 2019.
[10] J. McDonald, The saturation adjustment in numerical modelling of fog, J. Atmos. Sci. 20 (5) (1963) 476–478.
[11] R.T. Tan, "Visibility in bad weather from a single image," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.
[12] R. Fattal, Single image dehazing, ACM Trans. Graph. (TOG) vol. 27 (3) (2008) 72–79.
[13] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. vol. 33 (12) (2011) 2341–2353.
[14] A. Levin, D. Lischinski, Y. Weiss, A closed-form solution to natural image matting, IEEE Trans. Pattern Anal. Mach. Intell. 30 (2) (2008) 228–242.
[15] K.B. Gibson, D.T. Vo, T.Q. Nguyen, An investigation of dehazing effects on image and video coding, IEEE Trans. Image Process. vol. 21 (2) (2012) 662–673.
[16] J.-P. Tarel and N. Hautiere, Fast visibility restoration from a single color or gray level image, in IEEE International Conference on Computer Vision, 2009, pp. 2201–2208.
[17] K. He, J. Sun, X. Tang, Guided image filtering, IEEE Trans. Pattern Anal. Mach. Intell. 35 (6) (2013) 1397–1409.
[18] D. Berman, T. Treibitz, S. Avidan, Air-light estimation using haze-lines. In Computational Photography (ICCP), 2017 IEEE International Conference on, pages 1–9, 2017.
[19] Q. Zhu, J. Mai, L. Shao, A fast single image haze removal algorithm using color attenuation prior, IEEE Trans. Image Process. 24 (11) (2015) 3522–3533.
[20] R. Fattal, Dehazing using color-lines, ACM Trans. Graph. (TOG) 34 (1) (2014) 13–14.
[21] S. Diamond, V. Sitzmann, S. Boyd, G. Wetzstein, and F. Heide, Dirty pixels: Optimizing image classification architectures for raw sensor data. arXiv preprint arXiv:1701.06487, 2017.
[22] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, Indoor segmentation and support inference from rgbd images. In European Conference on Computer Vision, pages 746–760. Springer, 2012.
[23] B. Cai, X. Xu, K. Jia, C. Qing, D. Tao, Dehazenet: an end-to-end system for single image haze removal, IEEE Trans. Image Process. 25 (11) (2016) 5187–5198.

[24] B. Li, X. Peng, Z. Wang, J. Xu, D. Feng, Aod-net: All-in-one dehazing network, Proceedings of the IEEE international conference, 2017.
[25] Dharejo FA, Zhou Y., Deeba F., Jatoi MA, Khan MA, Mallah GA, Ghaffar A., Chhattal M., Du Y., Wang X., A deep hybrid neural network for single image dehazing via wavelet transform. Optik, 2021.
[26] E.J. McCartney, Optics of the atmosphere: scattering by molecules and particles, vol. 1, John Wiley and Sons, Inc,, 1976.
[27] S.K. Nayar and S.G. Narasimhan, Vision in bad weather, IEEE International Conference on Computer Vision, vol. 2, 1999, pp. 820– 827.
[28] Tarel J.P., Hautiere, N.Cord, Improved visibility of road scene images under heterogeneous fog, IEEE Intelligent Vehicles Symposium, 2010, pp 478–485.
[29] Li B., Ren W., Fu D., Reside: A benchmark for single image dehazing, arXiv preprint,2017.
[30] N. Srivastava, G. Hinton, A. Krizhevsky, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.
[31] Hore A., Ziou D., Image quality metrics: PSNR vs. SSIM, international conference on pattern recognition, 2010, pp 2366–2369.
[32] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, Int. J. Comput. Vis. vol. 47 (1–3) (2002) 7–42.
[33] He K., Sun.J., Fast guided filter, arXiv preprint arXiv:1505.00996, 2015.