Image Dehazing Using U-net Architecture with Wavelet Feature Extraction

Aiman Jabaren
University of California, San Diego
ajabaren@ucsd.edu

Sina Shahsavari
University of California, San Diego
sshahsav@ucsd.edu

Abstract

Image dehazing is an ill-posed problem challenging many computer vision applications. It is an undesirable physical phenomena that occurs while capturing real pictures. Many methods and algorithms were developed for that end. Nonetheless, no generic high-quality solution has been proposed yet. Hence, we are proposing an end-to-end architecture that tackles the problem and outperforms state-of-the-art approaches. Deep network has shown promising results when it comes to image dehazing. We investigated and harnessed wavelet filterbank properties for feature extraction in U-net neural network. By doing so, we have combined U-net learning properties with image specific efficient features extraction. While enhancing the model performance, we explored the effect of different types of wavelet filterbanks as well as different number of wavelet levels on the quality of image effecting.

Index Terms— *Wavelet filterbank, feature extraction, multi-level wavelet, U-net, image dehazing,*

1. Introduction

In the wild, especially terrestrial areas, atmospheric suspended particles such as aerosols, water droplets, and other dry particles absorb and scatter light which lead to hazing the captured image on the camera, blurring it and fainting it. This results in the visual effect of contrast loss of far objects. Such a problem limits some high-level applications, such as photography stylization [2], semantic segmentation [29], scene understanding [25], and video surveillance [7]. Therefore, haze removal is desired for both consumer photography and computer vision applications.

Image dehazing problem endeavors extracting clear and haze-free images by removing haze effects from hazy images. However, it is a challenging problem because haze effect depends on many undetermined parameters such as objects' unknown depth.

Image dehazing is an ill-posed problem; There are several unknown optical parameters. The haze-free image and the atmospheric light transmission distribution are bilinearly coupled, making the problem non-convex and hard to solve.

There have been multiple approaches to solve the problem and modify the visual perception, including, contrast enhancement, histogram equalization [31], linear mapping, retinex-mapping [11] and gamma correction [13]. All these methods ignore haze thickness and variable objects' depths, therefore, failing to compensate adaptive haze degradation. Some of the previous attempts to amend this have used priors of multiple images. For instance, [22] utilized some user-specified information interactively in the physical model for haze removal. Such model is often underconstrained when only one image is used as an input because the light scattering is an unknown function of depth. Therefore, [26] has used multiple images (polarization) of the same scene in order to dehaze while using depth information.

Our proposed approach derives the haze without considering any prior knowledge and could remove haze artifact with a single input using U-net, which improves the state-of-the-art methods by applying wavelet transformation for feature extraction in a U-shape neural network.

In this paper, we discuss and introduce multiple topics. Section 2 provides physical model of hazy images and explains the problem. Section 3 provides a brief literature review of conventional single image dehazing as well as neural network methods and wavelet feature extraction. In Section 4, we describe 2D wavelet transformation, general U-net architecture and finally introduce our proposed architecture with wavelets. Section 5 presents our approaches and conducted experiments for enhancing the performance of the model by focusing on feature extraction blocks architecture. It includes experimental setup and dataset description. In section 6 we present results of our experiments, compare them with other methods and discuss them. Section 7 is the conclusion and finally in section 8 we discuss future work and potential future approaches.

2. Problem Statement

Multiple hazing models have been developed in order to predict the relation between the clear image and the hazy one. A famous dehazing model which utilizes physical parameters is:

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

where $I(x)$ is the hazy image and $J(x)$ is the clear image to be estimated, A is the global atmospheric light and $t(x)$ is the transmission matrix expressing unscattered light that reaches the camera.

In this model which assumes a homogeneous haze, the transmission matrix is expressed as $t(x) = e^{\beta d(x)}$, where β is the medium extinction coefficient and $d(x)$ is the scene depth. The many unknown parameters make this problem much harder to solve and many parameters need to be estimated. In previous conventional approaches, these parameters are usually assumed.

3. Related Works

3.1. Conventional Image Dehazing Methods

Image dehazing is an ill-posed and challenging problem, thus most of the work addressing it consider some form of prior knowledge while trying to estimate the transmission matrix and the global atmospheric light, respectively. In this section, we briefly review some of the significant previous work.

In [27], atmospheric artifacts are removed from training images by assuming the scene depth. In [4], non-local priors were used because authors assumed that in RGB space, each color cluster becomes a haze-line in the hazy image.

Based on another observation that in natural images, the local sample mean of pixel intensity is proportional to the standard deviation, Oakley *et al*. in [23] assumed air-light is constant over the entire image then estimated it given a single image. In another work by Tan [28] the contrast of the direct transmission is maximized while assuming a smooth layer of air-light, hence the image contrast is restored from a single input image.

The dark-channel prior model is proposed by He *et al*. in [9]. The dark-channel prior is based on the phenomenon has been observed in outdoor haze-free images and states that considering small patches of a haze-free image, the intensity values of pixels of at least one color channel are very low and even close to zero. The dark channel is defined as the following:

$$D(x) = \min_{y \in \Omega_r(x)} \left(\min_{c \in \{R,G,B\}} I^c(y) \right)$$

where $\Omega_r(x)$ is a $r \times r$ local patch centered at x, I^c is an RGB color channel of I. Since the dark channel feature has a high correlation to the amount of haze in the image, it can be used to estimate the medium transmission $t(x) \propto 1 - D(x)$ directly.

Besides prior assumptions, some proposed methods were fusion-based. One multi-scale fusion approach which proposed by Ancuti *et al*. [1], uses several inputs derived from the original image to obtain a visibility enhanced image. Moreover, fusion of high-frequency and low-frequency images derived from original images by discrete wavelet transform can be seen as fusion method.

3.2. Image Dehazing using Neural Networks

More recent approaches employ neural networks to predict and learn the parameters of the model implicitly without any prior assumption. Therefore, many deep learning-based methods for image dehazing are proposed. In [21], the transmission matrix and the global atmospheric light are simultaneously estimated by training a fully connected network.

In [24], a multi-scale CNNs (MSCNN) has been used to generate a coarse-scale transmission matrix. Li *et al*. in [16] have proposed Aod-net which is an end-to-end pipeline from hazy images to clear images. Furthermore, generative adversarial networks (GANs) [6] which is powerful in generating realistic images, are also used for image dehazing by Li *et al*. in [18].

3.3. Wavelets in Neural Network

In recent years, wavelets have recently emerged as an effective tool to analyze image information [20], because they provide a natural partition of the image spectrum into multi-scale and oriented sub-bands.

In [3], Bae *et al*. showed that, in single-image super-resolution problem, using wavelet feature extraction in the convolutional neural networks (CNNs) increases their performance and then proposed the wavelet residual network. Moreover, in [12], CNNs and multilayer convolution framelets for image reconstruction were built to preserve detail edges of images. Authors in [19] presented an efficient multi-scale correlated wavelet approach to solve image dehazing, which motivates us to train the neural network in the wavelet space. Therefore, we embed wavelets to our proposed network to extract hazy image features.

In a state of the art research [30], a U-net neural network is enhanced by replacing pooling layers with wavelet filter banks in which one level of Haar wavelets is used.

To sum up, we designed an end-to-end neural network with U-net architecture as our backbone model and added modified wavelet feature extraction blocks to increase accuracy.

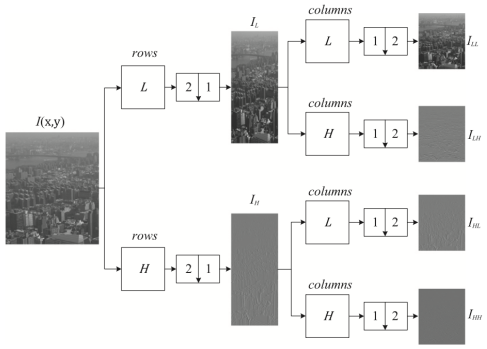

Figure 1. One level of 2D Discrete wavelet transform

4. Model

4.1. 2D Wavelets transform

One dimensional discrete wavelet transform consisted of two 1D functions, scaling function $(\phi(x))$ and wavelet function $(\psi(x))$, however when it comes to 2D signals such as images, 2D wavelet filters must be applied to separate input image into four frequency bands: low-low (LL), low-high (LH), high-low (HL) and high-high (HH). Therefore 2D DWT is formed by functions: $I_{LL}(x,y), I_{LH}(x,y), I_{HL}(x,y), I_{HH}(x,y)$.

Considering separability property of 2D DWT:

$$I_{LL}(x,y) = \phi(x)\phi(y)$$

$$I_{LH}(x,y) = \phi(x)\psi(y)$$

$$I_{HL}(x,y) = \psi(x)\phi(y)$$

$$I_{HH}(x,y) = \psi(x)\psi(y)$$

Thus to calculate the four sub-images it suffices to apply 1D $\phi(x)$ and $\psi(x)$ to each row in the image, followed by vertical filtering and down-sampling using two filters to each column.

In contrast, inverse DWT merges four sub-images to the original image by up-sampling using the inverse form of the same filters. The 2D wavelet decomposition using 1D DWT on each dimension is illustrated in Fig.1.

4.2. U-net

Among different architectures of deep neural networks, U-net network has shown better results in image dehazing. U-net architecture, which acts as an auto-regression model, consists of a contraction and an expansion path. In contraction path pooling or down-sampling layers increase the receptive fields of features and help learning more complex features in low-resolution space. Afterwards, un-pooling or up-sampling layers in expansion path propagate context information to higher resolution space by considering spatial information of the image and lead to an image with the same resolution as the input.

4.3. Wavelet U-net

In Fig.2 we see the architecture of a Wavelet-U-Net which has been employed as a base network for image dehazing in this project.

The input of the network is a single hazy image which has 3 RGB channel. In each layer of the U-net network, each input channel is passed through four 2D wavelet filterbanks by applying functions in Eq.(1) which produces four output channels with half the dimensions size. After filtering channels, a 3×3 convolution kernel is applied. Then, Convolution and DWT are applied iteratively to extract the features and decrease the resolution of feature maps.

After extracting the features vector, at the bottom of the U-net, a residual block (illustrated in Fig.2, brown lines) is applied, which is inspired by ResNet [10], to learn the haze directly from low-resolution features of the hazy image, and subtract it from the hazy image features, thus extracting the haze free features. At the bottom of the model, a haze-free image feature map vector has been extracted. In the expansion path, inverse wavelet transfomation is applied on the channels of the feature map followed by a 3x3 deconvolution kernel at each layer of the U-net until a clear image is extracted. The DWT outputs in the contraction path are concatenated to the outputs of the IDWT in the expansion path similarly to the original U-net. The network architecture is displayed in Fig.2.

5. Experiments

5.1. Dataset

For the purpose of training and evaluating the performance of our developed models, we use RESIDE dataset [17]. The RESIDE dataset contains clear images as ground truth. It also contains hazy images which were created synthetically according to the model of the hazy image in Eq.(1). We use 500 outdoor and in the wild clear images form this dataset for the purpose of training the model. As for hazy images, eleven hazy images are used per clear image by varying 7 different values of β which resemble the various amount of haze in an image. The atmospheric light A varies in the hazy images between 0.8 and 1. However, $d(x)$ which is the depth of different objects in an image, is constant in hazy images. Before feeding the images to the network both clear and hazy images are resized to 480×640 pixels.

5.2. Experiment 1

In wavelet based image dehazing the selection of wavelets is an essential task which determines the dehazed image quality. There are several choices of wavelets with

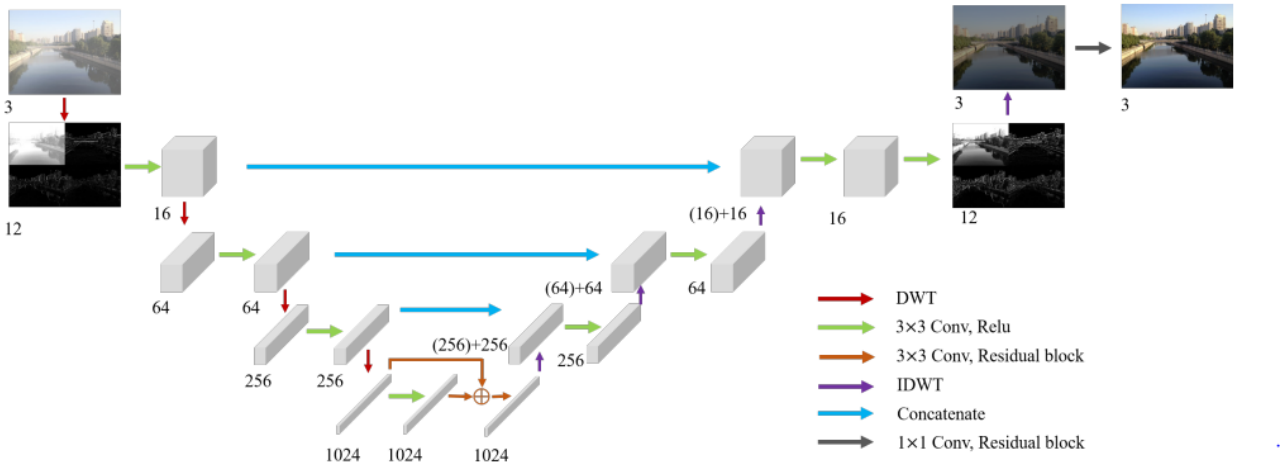

Figure 2. Overview of the proposed Model (wavelet-U-net) for single image dehazing. The digits under the blocks are the numbers of channels and digits in parentheses mean concatenated layers

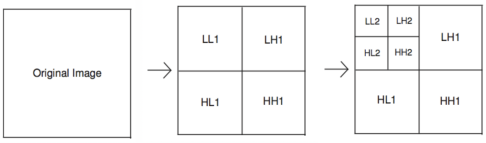

Figure 3. two level of 2D DWT

different properties. Therefore, in the first experiment, we aim to investigate the different wavelet types performance in image dehazing and features extraction when employed in the model described in section 4. We use different bases of wavelet filters. We have used Haar, Daubechies, Symlets [5] and Biorthogonal wavelets [20]. we are interested in finding whether certain wavelets outperform others and investigate their properties and advantages in image dehazing U-net. Some of the properties that we think might affect the performance are: **Orthogonality, symmetry, number of vanishing moments, power perseverance and filter size.**

5.3. Experiment 2

In this experiment, we aim to enhance image dehazing by boosting feature extraction using multiple levels wavelets filterbanks. Therefore, as it can be seen in Fig.3 we have replaced the first layer's Haar wavelet filterbank by a two level Haar wavelet filterbank. By applying a two level wavelet feature bank, we would split the input image to narrower frequency sub-bands in the low frequencies where most of the crucial image and haze features are located.

5.4. Experimental Setup

We have used Pywt [15] package for implementing wavelet transforms and Pytorch for building our model. Adam [14] is used as training optimizer. The initial learning rate is set to 10^{-4}. As for the loss function, L2 loss which computes MSE loss is used. The batch size is 40 for training and 10 for validation. We train for 60 epochs. Other parameters are initialized randomly.

6. Results and Discussion

6.1. Experiment 1

In the first experiment, we have trained the model using different types of wavelet filterbanks. We would like to see how different wavelet filterbank types perform as feature extraction blocks in our model.

In Fig.5 and Fig.4 we can observe that test PSNR increase as well as test and training loss decrease during the training process. The results of the different wavelet filterbanks are summarized in Table 1. The different wavelet filterbanks have resulted in different PSNR and LOSS when testing the model.

In Fig. 6, the loss and PSNR have been plotted as functions of the different wavelet types. We can observe that SYM2 has performed best, yielding the highest PSNR and lowest loss, while it was followed by Haar. The results strengthen the conventional assumption that Haar is a good image feature extractor. Haar is traditionally used as a feature extractor for image related applications. It preserves and extracts most of the important features in images such as vertical, horizontal and diagonal edges of the image. It is

Wavelet type	PSNR	MSE
Haar	22.78	0.0053
DB2	20.71	0.0085
DB3	22.46	0.0057
DB9	18.12	0.0154
Bior1.3	21.01	0.0079
Bior3.1	20.70	0.0085
Bior2.2	21.56	0.0069
Bior3.5	19.14	0.0122
sym2	**23.72**	**0.0042**
sym3	22.26	0.0059

Table 1. Results for first experiment

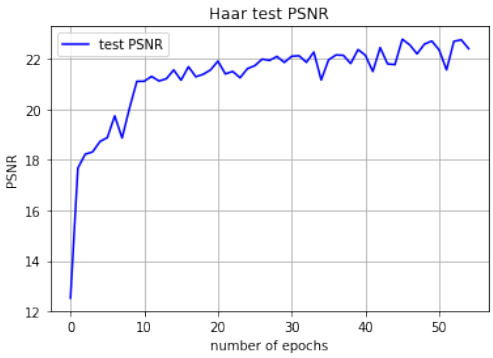

Figure 4. Test PSNR vs Epochs while using Haar Wavelet transform in U-net

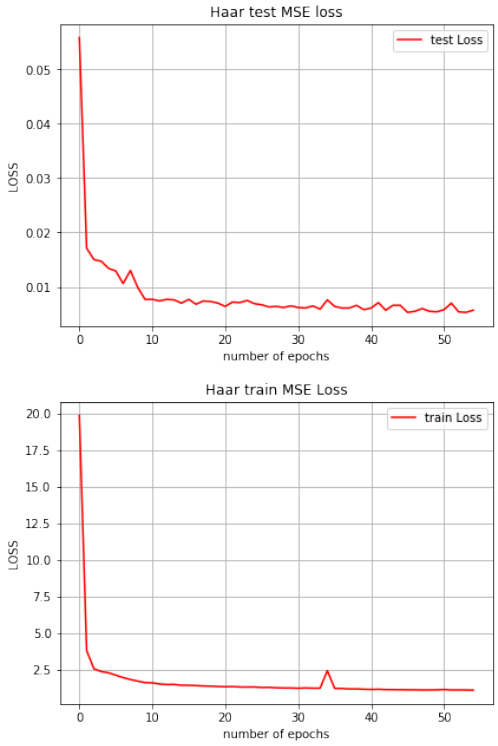

Figure 5. Test & Train Loss vs Epochs while using Haar Wavelet transmission in U-net

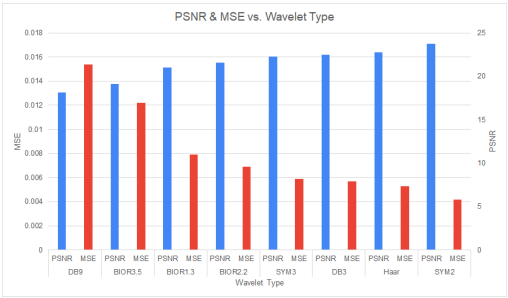

Figure 6. Exp1: PSNR and MSE of different wavelet types in Wavelet U-net for image dehazing

also symmetric which is common in images and important for human vision. It also preserves power and has a vanishing moment of one which means that a second wavelet decomposition of high frequency part of the image doesn't carry much information. Therefore it suffices to apply a second level of wavelet decomposition only on low frequency section since it would extract significantly more features.

As for the outperformance of SYM2, it does satisfy some of the desired properties of Haar and DBs such as orthogonality, however, it is nearly symmetric unlike DB and also has a greater number of vanishing moments than Haar.

As for biorthogonal, it has performed worse than other types which was unexpected since it is biorthogonal and symmetric. It could be due to the fact that it does not preserve power.

6.2. Experiment 2

In experiment 2, we aim to investigate image dehazing performance when a second wavelet level is added for feature extraction in the first layer of the model. In Fig. 8 and Fig. 7, the PSNR and MSE of the two models are plotted, respectively. In Tabel 2 we can see that adding a second level has enhanced and boosted image deahzing significantly. The PSNR has reached **25.056** which shows a remarkable improvement in image dehazing by about 4 units. Based on the results mentioned in [19] and [8] haze is mostly distributed on the low frequency spectrum, in addition, it is known that most of the significant features in natural images are in the low frequencies. Hence, adding a second level of wavelet decomposition in feature extraction blocks has boosted their performance by leading to more valuable extracted features of the image, thus enhancing the haze removal performance.

#Wavelet levels	PSNR	MSE
1	21.157	0.0077
2	**25.056**	**0.0031**

Table 2. Results for second experiment

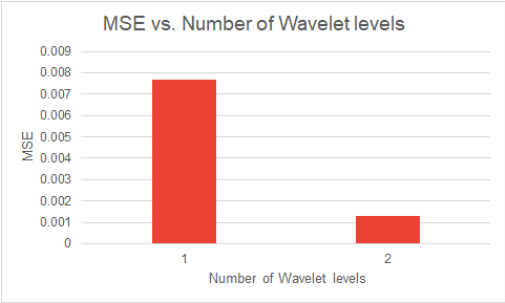

Figure 7. Exp2: MSE vs number of levels of wavelet decomposition in the first layer of Wavelet U-net

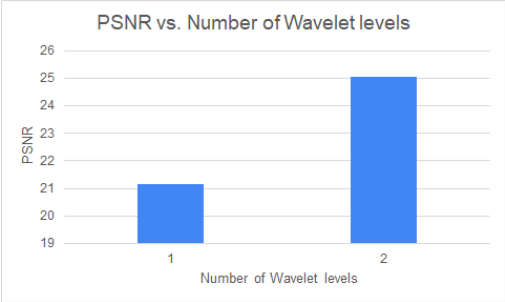

Figure 8. Exp2: PSNR vs number of levels of wavelet decomposition in the first layer of Wavelet U-net

6.3. Model Evaluation and Comparison

For further investigation, we have compared and plotted our developed models' yielded PSNR in both experiments versus state of the art models and algorithms that were developed and tested using the RESIDE dataset: [16],[24],[19],[9],[12]. In Fig.9 we can observe that adding a second level of wavelet decomposition in the model's first layer has generated better results (PSNR=25.056) than all previous image dehazing models and methods shown. It extracted more image and haze features which boosted the dehazing significantly. This method utilizes machine learning concepts while training, but it is also image specific because it extracts image features according to Haar wavelet filterbank. In other words, in addition to the U-net parameter learning, the new architecture combines learning the new parameters with utilizing wavelets filterbanks for feature extraction and image dehazing.

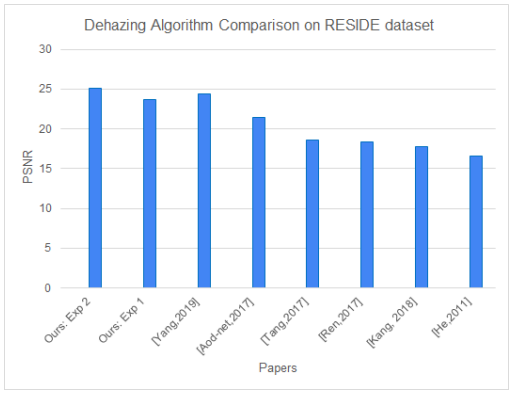

Figure 9. Comparing our best results with literature results on RESIDE dataset

7. Conclusion

In this paper, we have investigated and boosted image dehazing U-net based architecture using enhanced wavelet filterbanks as feature extraction blocks in the U-net. Previous papers have either used a simple U-net for dehazing or added a simple Haar wavelet filterbank for feature extraction. However, we have compared and tested multiple wavelets filterbanks on the RESIDE dataset and saw that SYM2 has performed best in image dehazing due to better performance in image and haze features extraction than other wavelet types. SYM2 has outperformed other wavelet types as a result of multiple properties that contributes to image and haze feature extraction as discussed earlier.
We have also boosted the performance and outperformed previous work by extracting more features when adding a second level in the wavelet filterbank for the first layer of the model.

8. Future Work

Our results are promising and show enhancement in image dehazing performance and accuracy when extracting more features. We would like to investigate more types and approaches for wavelet feature extraction using U-net. Therefore, we would like to design other experiments for that end. The first method is adding a second level wavelet to all network layers. A second method that we would like to explore is concatenating 2 different wavelets filterbanks. By stacking different types of wavelets, more distinguished features would be extracted, therefore, enhancing performance. A third method that might show some improvement is adding wavelet packets to the wavelet filterbanks blocks.

References

[1] C. O. Ancuti and C. Ancuti. Single image dehazing by multi-scale fusion. *IEEE Transactions on Image Processing*, 22(8):3271–3282, 2013.

[2] R. Azami and D. Mould. Detail and color enhancement in photo stylization. In *Proceedings of the symposium on Computational Aesthetics*, page 5. ACM, 2017.

[3] W. Bae, J. Yoo, and J. Chul Ye. Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 145–153, 2017.

[4] D. Berman, S. Avidan, et al. Non-local image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1674–1682, 2016.

[5] I. Daubechies. *Ten lectures on wavelets*, volume 61. Siam, 1992.

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[7] K. Goyal and J. Singhai. Review of background subtraction methods using gaussian mixture model for video surveillance systems. *Artificial Intelligence Review*, 50(2):241–259, 2018.

[8] J. He, F. Z. Xing, R. Yang, and C. Zhang. Fast single image dehazing via multilevel wavelet transform based optimization. *arXiv preprint arXiv:1904.08573*, 2019.

[9] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010.

[10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[11] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997.

[12] E. Kang, W. Chang, J. Yoo, and J. C. Ye. Deep convolutional framelet denosing for low-dose ct via wavelet residual network. *IEEE transactions on medical imaging*, 37(6):1358–1369, 2018.

[13] J. Katajamäki. Methods for gamma invariant colour image processing. *Image and Vision Computing*, 21(6):527–542, 2003.

[14] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[15] G. R. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary. Pywavelets: A python package for wavelet analysis. *J. Open Source Softw*, 4(36):1237, 2019.

[16] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4770–4778, 2017.

[17] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018.

[18] R. Li, J. Pan, Z. Li, and J. Tang. Single image dehazing via conditional generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8202–8211, 2018.

[19] X. Liu, H. Zhang, Y.-m. Cheung, X. You, and Y. Y. Tang. Efficient single image dehazing and denoising: An efficient multi-scale correlated wavelet approach. *Computer Vision and Image Understanding*, 162:23–33, 2017.

[20] S. Mallat. *A wavelet tour of signal processing*. Elsevier, 1999.

[21] R. Mondal, S. Santra, and B. Chanda. Image dehazing by joint estimation of transmittance and airlight using bi-directional consistency loss minimized fcn. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 920–928, 2018.

[22] S. G. Narasimhan and S. K. Nayar. Chromatic framework for vision in bad weather. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, volume 1, pages 598–605. IEEE, 2000.

[23] J. P. Oakley and H. Bu. Correction of simple contrast loss in color images. *IEEE Transactions on Image Processing*, 16(2):511–522, 2007.

[24] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang. Single image dehazing via multi-scale convolutional neural networks. In *European conference on computer vision*, pages 154–169. Springer, 2016.

[25] C. Sakaridis, D. Dai, S. Hecker, and L. Van Gool. Model adaptation with synthetic and real data for semantic dense foggy scene understanding. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 687–704, 2018.

[26] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar. Instant dehazing of images using polarization. In *CVPR (1)*, pages 325–332, 2001.

[27] K. Tan and J. P. Oakley. Enhancement of color images in poor visibility conditions. In *Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101)*, volume 2, pages 788–791. IEEE, 2000.

[28] R. T. Tan. Visibility in bad weather from a single image. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[29] F. Z. Xing, E. Cambria, W.-B. Huang, and Y. Xu. Weakly supervised semantic segmentation with superpixel embedding. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 1269–1273. IEEE, 2016.

[30] H.-H. Yang and Y. Fu. Wavelet u-net and the chromatic adaptation transform for single image dehazing. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 2736–2740. IEEE, 2019.

[31] H. Zhu, F. H. Chan, and F. K. Lam. Image contrast enhancement by constrained local histogram equalization. *Computer vision and image understanding*, 73(2):281–290, 1999.