Single Image Dehazing Using End-to-End Deep-Dehaze Network

Masud An-Nur Islam Fahim  
Chosun University, South Korea

Ho Yub Jung
Chosun University, South Korea

ABSTRACT
Atmospheric haze limits the performance of the camera sensor, which results in capturing degraded hazy images. Removal of this haze from the observed images is a complicated task because of its ill-posed nature. This study offers the Deep-Dehaze network to retrieve the haze-free image. For the given input, the proposed architecture uses four feature extraction module to perform nonlinear feature extraction. We improve the traditional Unet architecture and the residual network to design our architecture. We also introduced the L1 spatial-edge loss function, which enables our system to achieve better performance over the typical L1 and L2 loss function. Unlike other learning-based approaches, our network does not use any fusion connection for image dehazing. The experimental results show that our proposed Deep-Dehaze architecture surpasses previous state-of-the-art single image dehazing methods quantitatively and qualitatively. Our network achieves outstanding average PSNR score 24.5 on the RESIDE dataset.

CCS CONCEPTS
• • Computing methodologies; • Computational photography; Visual inspection.;

KEYWORDS
Additional Key Words and Phrases: End-to-End, Image dehazing, ResUnet, Seq-Residual block

ACM Reference Format:
Masud An-Nur Islam Fahim and Ho Yub Jung. 2020. Single Image Dehazing Using End-to-End Deep-Dehaze Network. In The 9th International Conference on Smart Media and Applications (SMA 2020), September 17–19, 2020, Jeju, Republic of Korea. ACM, New York, NY, USA, 6 pages.  
https://doi.org/10.1145/3426020.3426058

1 INTRODUCTION
Haze is a common atmospheric phenomenon during the presence of different sized floating particles in the air like water droplets, dust, smoke, and water particles. Images captured by the camera in this cluttered medium often contain complicated, non-linear, and data-dependent noise. So, these images are losing their contrast and visibility, which distorts image information. In any weather condition, air light and scattering light may degrade the image. This degradation results in the corruption of the image radiance information, which hinders many computer vision applications such as radar tracking, weather forecasting, traffic monitoring [14].

For image dehazing, most of the state of the art methods widely use the atmospheric scattering model [3–12, 15–17, 20–23] as noted below:

\[ I(x) = J(x)t(x) + A(1 - t(x)) \]  

Here, \( I(x) \) is the degraded hazy image, \( t(x) \) denotes the transmission map, \( J(x) \) is the haze free image, and \( A \) stands for the global atmospheric light. Incoming light to the camera sensor is not independent of the transmission map, which explicitly controls the visual quality of the output image from the sensor. A hazy image’s functional model fundamentally contains two parts: transmission information and the scattered atmospheric light. Both of them together comprise the hazy image, in theory. Assuming the atmospheric light \( A \) is homogeneous, the expression for the transmission map is:

\[ t(x) = e^{-\beta d(x)} \]

where, \( \beta \) represents attenuation coefficient of the atmosphere and \( d(x) \) is the depth information of the scene. Based upon the equation 1) and previously mentioned assumption for homogenous distribution, many studies have been proposed. We can generalize these studies into two classes: model-based methods [1, 2, 13, 18, 24], learning-based method [6, 12, 15, 16, 19, 22, 23].

In this study, we took a different approach to perform single image dehazing. Our network at the first stage takes the hazy input image without any preprocessing. Then this image plugged into four different feature extraction blocks to extract the hazy information from the image. Four feature extraction blocks are as follows: residual-Unet with two entirely different filter properties, conventional residual block, sequential residual block. In our residual Unet, we utilize only one residual connection for the Unet architecture. Moreover, this residual connection is a global residual connection which adds the input image with the extracted feature at the end of
the network. This connection is in contrast to the widely practiced local residual Unet architecture.

For the conventional residual block, we perform 15 continuous convolution operations upon the input image, followed by a global residual connection. In our residual block, we only use the ReLU activation function, and we do not use any batch normalization in the residual block. For our sequential-residual block, we concatenate the outputs from the residual blocks. After that, a simple convolution operation upon the concatenated tensor gives us the desired features. Now we subtract all the features from the input image, and we concatenate the resultants together into a single tensor. A final convolution operation upon this tensor gives us the desired haze-free image. We can summarize the overall contribution of our study as follows:

- Our Deep-Dehaze network enjoys end-to-end property for dehazing the hazy input images. The proposed Deep-Dhaze system achieves state-of-the-art performances by collaborating four different feature extraction modules.
- In this work, we propose a sequential-residual module for feature extraction, which outperforms the traditional residual block, dense block, and residual-attention block.
- Our exploited Residual-Unet uses only one global residual connection. This improvisation on the top of the Unet outperforms the performance original Unet-based module.
- Finally, we improvise the usual L1 loss function by utilizing it for both the spatial and edge domains. This loss function helps our network to retrieve the fine details without using any fusion or gated fusion operation. Also, our reconstructed outputs are free from the burning effect, unlike the prior-based approaches.

We have followed the usual sequence for documenting of our study. Following section is focused on related works that reconstruct dehazed images. After this, we explain the network architecture and present relevant figures. Then, we compare our performance with other dehazing studies upon various datasets, which is followed by the conclusion.

2 RELATED WORK

Classical Prior-based methods operate upon two things: transmission map and the atmospheric light. At the very core, these prior dependent schemes try to improve the manipulation of the statistical property of the brightest intensity region. It’s common for these practices to focus on sky region segmentation, adaptive thresholding, air-light compression [19].

The primary dehaze-based method was Histogram equalization that introduces near-uniform probability density function to improve contrast enhancement by remapping the histogram of the images using [11]. Following that, homomorphic filtering combines illumination and reflectance multiplicatively and results in bleaching of the image near-uniform probability density function. Global contrast optimization schemes enhance the overexposed regions of the SDR images by applying the compression curve. Multi-scale Retinex [20] helps to increase the contrast in the luminance channel; however, the performance decreases more if the input image contains light noise. In [11], the authors combine single-scale Retinex theory with wavelet transformation to remove the input image’s haze. This study [7] use Spatio-Temporal Retinex-inspired Envelope with Stochastic Sampling scheme to approximate the haze-free image.

The dark channel prior is one of the oldest dehazing schemes [10]. This method is so influential that a bulk of the image dehazing study is standing upon this formula. This statistical observation relies mostly on most outdoor image regions having pixels that present low intensity in at least one of the color channels [10]. Even though prior dependent schemes are useful for image dehazing, they are not robust enough due to their inherent limitations in prior selection. Their assumption mostly covers the atmospheric haze at some scale but fails when it comes to the indoor fog or hazy urban images. The learning-based approaches take versatile datasets to learn the haze properties in different domains.

Many recent dehazing studies use deep neural networks to learn haze information for the given dataset and apply it to remove the haze from the image. These methods [6, 12, 16, 22] uses a deep convolutional network to estimate the transmission map, ambient light information form the given image, and later restore the clean image based on the learned information. This study [6] proposes an end-to-end dehazing network to estimate the transmission map from hazy images. Patel et al. [22] uses pyramid dense network to estimate the clear image from the given input corrupted image. AoD [12] net uses R-CNN to estimate the dehazed image without using the atmospheric light information. Ren et al. [16] use a coarse-to-fine scheme to accurately predict the transmission map for haze removal. This study [15] uses pix-2-pix network for image dehazing. They have converted the image dehazing problem into image translation problem for removing haze from the given image.

GAN-based methods also been using for approximating the haze-free image from the given corrupted image. FFDnet [23] is the first GAN-dehazing scheme that removes the haze from the picture by estimating the non-linear noise level map. Cycle dehaze GAN [9] performs dehazing on the top of the cycle GAN and uses perceptual and cycle consistency loss. By exploiting the Tiramisu model, conditional GAN [5] dehaze the hazy input image. CCGAN [21] uses a dehaze GAN and a classification network to predict the transmission map for accurate haze removal. CWGAN [8] learns to construct a clean image form the given corrupted image by using fundamental features of the Wasserstein GAN.

3 NETWORK DETAILS

As shown in Figure 2, our network learns the nonlinear relationship between the hazy input image and the output image. Our end-to-end deep-dehaze system consists of different sub-networks. Included sub-networks allows our end-to-end network to learn image features, spatial properties, haze patterns for the given input. We present the necessary description of these sub-modules below.

The proposed ResUnet block differs from the original Unet architecture with a single global residual connection. Established Unet with residual connections includes skipping links for every layer. These links allow the network to learn better than the original Unet architecture. The number of the channel increases for Unet after each pooling operation exponentially, which increases the cost for the skip connection at each layer. To reduce that, we use a single global skip connection for our Unet base architecture.
We have observed better performance for the Unet with global skip connection. In our end-to-end Deep-Dehaze network, we have used two ResUnet blocks for spatial feature learning. For the first ResUnet, channel increment starts from 32, and for the second, channel increases exponentially from 16.

Our residual block follows the typical design flow. We have included the skip connection to learn the spatial haze patterns form the given input. It is a well-known fact that residual connection can attain the underlying nonlinear relation between the input and the output image. For this reason, in ill-posed computer vision problems, the residual module is a popular choice. In our residual block, we have used 15 consecutive convolution operation before the skip connection. After learning the haze patterns at 15 layers, we have added the input image to the end of the module. For this module, we use 32 filters from the second convolution layer to the 14th layer. For the first layer and the last layer, we have used three channels. Our kernel width is 7 for each layer in the residual module.

The proposed Seq-Residual block is the simple extension of the typical residual block, as we describe above. In this module, we have used three consecutive residual blocks, one individual convolution operation after each residual block. Then we have concatenated...
the residual blocks into a single tensor for the global feature learning. For each res blocks, we have used different kernel widths. Our kernel sizes are 3, 5, and 7. For the individual convolution operation, the kernel width is 3. At the final stage, we perform the last convolution operation to extract the concatenated tensor’s global feature information.

After extracting desired features using the modules mentioned above, we perform subtract operation between the hazy input image and the output of each module. Following that, we concatenate all of them into a single tensor. This tensor is the most crucial feature bulk that we extract from the corrupted input image in each epoch. In this tensor, we concatenate the multi-scale spatial information, inter-pixel relationship, transmission light features. A final convolution operation upon this tensor completes our network’s single image dehazing task. For ill-posed computer vision problems, it is reasonable to choose the loss function from the regression loss family. Typically, people use L1 or L2 loss function. However, L1 is preferable over the L2 loss function due to the outlier scenario. In practice, L1 is more robust for outliers over the L2 loss function. For our network, we use a combination of L1 loss function for the spatial and edge domains. Usually, the system approximates the clean image from the given corrupted image based on the nonlinear relationship between them. In this manner, the performance of the network varies from image to image. Also, the spatial complexity of the given image makes this approximation more challenging. Due to this reason, it is common to see the resultant image contains less amount of inter-pixel variance than the ground truth image. In other words, the reconstructed image is more likely to be he smoothed out version of the original image. To address this, we use the L1 loss function for both the edge and the spatial domains. We augmented the loss value from both of them in a weighted manner to constitute our proposed loss function. Our choice of weights is 0.85 and 0.15 for the spatial domain and the edge domain, respectively. In this way, we have achieved a sharper image compared to the traditional L1 loss function. The equation of the proposed loss function is below:

\[
L(x, y) = 0.85 * L^\text{spatial}_1 + 0.15 * L^\text{edge}_1 \tag{3}
\]

Here, \(L^\text{spatial}_1\) is the L1 loss function for the spatial domain and \(L^\text{edge}_1\) is for edge domain. The equation of the L1 loss function as follows:

\[
L(x, y) = \frac{1}{n} \sum_{i=0}^{n} |y_i - y'_i| \tag{4}
\]

From the equation 3 comes from the empirical evidence. The next section contains detailed result analysis for our proposed network.

4 RESULT ANALYSIS

In this study, we primarily use the RESIDE dataset for single image dehazing. We have selected three subsets of the RESIDE dataset. They are indoor training set, outdoor training set, synthetic testing set. For validation, we have used indoor and outdoor validation set from the RESIDE dataset. For synthetic haze validation, we have also used the validation set from the DIV2k dataset. For the extreme dehazing performance, we have used the dataset curated by the Ancuti et al.\[3\]. This dataset contains extremely hazy images with their respective ground truths.

We have used following training setup for our network. We sample our training images into 128 by 128 patches to reduce the training complexity. For our scheme, we use the Keras framework to design the dehazing network. For the ADAM optimizer, we keep the usual training parameter. We train the Deep-Dehaze network with the batch size 16.

For the comparative analysis, we use three different methods against our study. These state-of-the-art studies are Dark channel prior, Non-local prior, All-in-one dehazing net. We use PSNR, SSIM, and MSE metric to demonstrate quantitative performance. The synthetic dataset from the RESIDE validation, DIV2k validation, and the extreme dehaze validation dataset from the Ancuti et al.\[3\] comes with the ground truth images. Figure 6 is showing the visual comparison between the proposed study and the studies mentioned above. In Figure 7, we present the comparative analysis for the extreme dehazing. The left column of the Figure 6 contains the input hazy image. From this figure, we can see the side by side comparison of the single image dehazing performance. In the rightmost column, we see the dehazing performance for the dark channel prior \[10\]. As seen from the picture, a dark channel prior helps the user remove atmospheric hazes from the given image effectively. However, it faces challenges with the burning-effect, which is common for typical dehazing methods. Due to the burning-effect, many dehazing schemes reduce the overall light information that decreases vital details. This method also reconstructs the given input image with an unnatural contrast, which degrades the total fidelity of the input image.

From the second right column, non-local image scheme shoes superior image dehazing performance over the dark channel prior. Compared to the dark channel prior, this method is contrast-adaptive and more immune to the pixel burning-effect. The middle column shows the image dehazing performance of the AoD net. Even though its overall haze removal performance is better than

Figure 5: Seq-Residual module for our network.
other methods, reconstructed images are less sharp than the other studies. The second left column is showing the dehazing performance of the proposed study. We can see that the proposed research can dehaze the green leaf region most effectively compared to the other studies. For the other studies, pixel burning-effect is strongly visible in the leaf region of the house image. We can also see that the inherent contrast appears to be more natural for the proposed study compared to other studies. From figure 7, we can see the demonstration of the extreme image dehazing performance. This picture contains the hazy images and the ground truth image from the Ancuti et al.[3]. From this photo, our approach can extract the most pixel information from the given hazy image. For the bicycle image, we can see a generator in the room from the ground-truth image. From the dehazed images, the generator is more pleasantly visible for the proposed study.

For the quantitative analysis, we use the PSNR, SSIM, and the MSE metric for the comparison. Table 1 is showing the quantitative comparison for the validation dataset from the RESIDE dataset. In Table 2, we present the comparison for the extreme dehazing dataset. Table 3 contains the quantitative comparison for the validation dataset from the DIV2K dataset. From these tables, we can see that the proposed study achieves the best results on average for every dataset.
In summary, our model performs better than other state-of-the-art studies. Proposed study achieves better visual performance as from the Figure 6 and Figure 7. Also, this study shows its efficacy for quantitative analysis as from the tables above. However, our method is not absolute in terms of performance. In some cases, other studies perform better than our scheme. For example, from Figure 6, we can see that our method poorly dehazed the road image. Even though that image contains better contrast, more haze is on the road for our model than the other studies. In our future work, we would like to address issues like this. We hope to work with more complex haze scenarios to achieve better performance in the extreme dehazing dataset. We would also like to improve the network’s performance by instance- learning. By doing so, we can improve the overall adaptability of the proposed network. Furthermore, our network is still limited to extreme dehazing tasks, which will be considered for future work.

5 CONCLUSION

In this study, we propose an end-to-end image dehazing study to recover the clean image from the given input. In our research, we have introduced several improvements over usual deep learning architectures. We also improve the performance of the typical L1 loss function for low-level computer vision tasks. Our study covers the image dehazing performance over several datasets. In every case, our scheme achieves better results compared to the state-of-the-art studies. In our future work, we are planning to implement this work for real-time video and detection tasks.

ACKNOWLEDGMENTS

This study was supported by research fund from Chosun University, 2020.

Table 1: Quantitative comparison for RESIDE dataset [12]

| PSNR | SSIM | MSE | Methods         |
|------|------|-----|-----------------|
| 23.08 | .7882 | 265.2 | AoD net [12]    |
| 17.35 | .7160 | 285.71 | He et al. [10]  |
| 18.89 | .7309 | 282.34 | Non-local [4]   |
| 24.5  | .8176 | 258.83 | Proposed study  |

Table 2: Quantitative comparison for Ancuti et al. [3]

| PSNR | SSIM | MSE | Methods         |
|------|------|-----|-----------------|
| 14.78 | .6136 | 431.03 | AoD net [12]    |
| 12.83 | .5800 | 443.72 | He et al. [10]  |
| 10.152 | .5561 | 458.48 | Non-local [4]   |
| 16.03 | .6945 | 421.25 | Proposed study  |

Table 3: Quantitative comparison for DIV2K dataset [12]

| PSNR | SSIM | MSE | Methods         |
|------|------|-----|-----------------|
| 27.9  | .8827 | 78.33 | AoD net [12]    |
| 26.03 | .8305 | 82.40 | He et al. [10]  |
| 24.62 | .8561 | 89.28 | Non-local [4]   |
| 29.92 | .9101 | 73.56 | Proposed study  |

REFERENCES

[1] Cosmin Ancuti, Codruta Ancuti, O Codruta, Christophe De Vleeschouwer, and Alan C. Bovik. 2016. Night-time dehazing by fusion.
[2] Cosmin Ancuti, Codruta Ancuti, Chris Herrmann, and Philippe Belaerts. 2010. A fast semi-inverse approach to detect and remove the haze from a single image. In Asian Conference on Computer Vision (WOCV’07). Springer, Queenstown, New Zealand, Article 7, 501-514 pages.
[3] Codruta Ancuti, Cosmin Ancuti, and Radu Timofte. 2020. NH-HAZE: An Image Dehazing Benchmark with Non-Homogeneous Hazy and Haze-Free Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. IEEE.
[4] Dana Berman, Shai Avidan, et al. 2016. Non-local image dehazing. In Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE.
[5] Raj N Bharath and N Venkateswaran. 2020. Single image haze removal using a generative adversarial network. In 2020 International Conference on Wireless Communications, Signal Processing and Networking (WSPNET). IEEE.
[6] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei QIng, and Dacheng Tao. 2016. Dehazenet: An end-to-end system for single image haze removal. IEEE Transactions on Image Processing 25, 11 (Dec 2016), 5187–5198.
[7] Vincent Whanwoo de Bravo and Jon Yogve Harderberg. 2015. Stress for dehazing. In 2015 Colour and Visual Computing Symposium (CVCS). 1–6.
[8] Joshua Peter Ebenezer, Bijayalaxmi Das, and Sudipta Mukhopadhyay. 2019. Single Image Haze Removal Using Conditional Wasserstein Generative Adversarial Networks. In 2019 27th European Signal Processing Conference (EUSIPCO). IEEE.
[9] Deniz Engin, Anil Gen, and Ekeren Hazim Kemel. 2018. Cycle-dehaze: Enhanced cyclergan for single image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. IEEE.
[10] Kaiming He, Jian Sun, and Xiaoou Tang. 2011. Single image haze removal using dark channel prior. IEEE transactions on pattern analysis and machine intelligence 33, 12 (Dec 2011), 2341–2353. https://doi.org/10.1109/TPAMI.2010.168
[11] Haadi Ibrahim and Seng Chuan Hoo. 2014. Local contrast enhancement utilizing bidirectional switching equalization of separated and clipped subhistograms. Mathematical Problems in Engineering 2014 (Feb 2014). https://doi.org/10.1155/2014/848615
[12] Boyi Li, Xinlian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. 2017. AoD-net: All-in-one dehazing network. In Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE.
[13] Gaoqiang Meng, Duan yang, Xiang Jianyong, Shiming Xiang, and Chunhong Pan. 2013. Efficient image dehazing with boundary constraint and contextual regularization. In Proceedings of the IEEE international conference on computer vision. IEEE, Sydney, Australia, 617–624
[14] Dubok Park, Hyungjo Park, David K Han, and Hanoek Ko. 2014. Single image dehazing with image entropy and information fidelity. In 2014 IEEE International Conference on Image Processing (ICIP). IEEE, Paris, France.
[15] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. 2019. Enhanced pix2pix dehazing network: In Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE.
[16] Wening Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaoichun Cao, and Ming-Hsuan Yang. 2016. Single image dehazing via multi-scale convolutional neural networks. In European conference on computer vision. Springer.
[17] Zhu Rong and Wang Li Jun. 2014. Improved wavelet transform algorithm using color attenuation prior. IEEE transactions on image processing 27, 9 (Oct 2017), 4608–4622. https://doi.org/10.1109/TIP.2018.2839891
[18] Chih-Tsung Shen and Wen-Liang Hwang. 2009. Image enhancement using retinex with robust envelope. In Proceedings of the IEEE international conference on computer vision and pattern recognition. IEEE.
[19] Jiawei Shen, Zhiyuan Li, Lei Yu, Gui-Song Xia, and Wen Yang. 2020. Implicit Euler ODE Networks for Single-Image Dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. IEEE, Seattle,WA,USA, 218–219.
[20] Bin Xie, Fan Guo, and Zixing Cai. 2010. Improved single image dehazing using dark channel prior and multi-scale retinex. In 2010 International Conference on Intelligent System Design and Engineering Application (1). IEEE, Chinghua,China, 848–851.
[21] Pei Yanting, Yaping Huang, and Xingyan Zhang. 2019. Classification-driven Single Image Dehazing. ArXiv abs/1911.09389 (2019).
[22] He Zhang and Vishaal M Patel. 2018. Densely connected pyramid dehazing network. In Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE.
[23] Kai Zhang, wangmeng Zuo, and Lei Zhang. 2017. FFIDNet: Toward a fast and flexible solution for CNN-based image denoising. IEEE Transactions on Image Processing 27, 9 (Oct 2017), 4608–4622. https://doi.org/10.1109/TIP.2018.2839891
[24] Qingsong Zhu, Jianming Mai, and Ling Shao. 2015. A fast single image haze removal algorithm using color attenuation prior. IEEE transactions on image processing 24, 11 (2015), 3522–3533. https://doi.org/10.1109/TIP.2015.2466191