Non Homogeneous Realistic Single Image Dehazing

Vinay P
PES University
Bangalore, India
vinaypurushothamnaidu@gmail.com

Abhisheka K S
PES University
Bangalore, India
ksabhisheka@gmail.com

Lithesh Shetty
PES University
Bangalore, India
shettylithesh24@gmail.com

Kushal T M
PES University
Bangalore, India
kush100kushal@gmail.com

Shylaja S S
PES University
Bangalore, India
shylaja.sharath@pes.edu

Abstract

A hazy image is one where atmospheric effects degrade the contrast and visibility of the image. It is often caused by the dispersion of light into the moisture particles present, smoke etc. This results in lower performance in high level vision tasks such as object detection, free space detection, scene understanding, etc. Hence the images have to be dehazed before applying other high level algorithms. Dehazing is the process of reconstructing the original colour and contrast of the image if taken in normal conditions. Image dehazing is a non-trivial task as it is hard to collect haze free ground truth images. Further, achieving dehazed images when variable haze is present is a significantly harder challenge. In this research, we propose the Non Homogeneous RESIDE dataset (NH-RESIDE) that contains images created synthetically using the principles of randomness and representativeness. Experimental results show that the model trained on our dataset produces visually more pleasing images with a much better dehazing effect on real world images. The model implemented in this paper also outperforms the state-of-the-art models by a huge margin on the NH-Haze dataset proposed by the NTIRE Non Homogeneous Dehazing Challenge at CVPR, achieving an average PSNR of 25.69 and an average SSIM of 0.80. It also achieves much better processing times when compared to other models, thereby facilitating real-time performance.

1. Introduction

Hazy images are caused due to multiple factors such as smoke, dust, fog, water droplets etc when light disperses into such media. A hazy image can be subjectively defined as an image with reduced visibility due to loss of colour and contrast. It can also be explained with the help of the atmospheric scattering model [9] [10] given by equation (1), where I(x) is the hazy image from the camera, A and t(x) are the global atmospheric light and the transmission map, respectively. J(x) is the scene radiance to be recovered.

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{1}$$

The presence of haze leads to performance degradation in other high level vision tasks like object detection, semantic segmentation, low light image enhancement, free space detection etc. Dehazing is the process of reconstructing the original colour and contrast of the image if taken in normal conditions. This is however a non-trivial and a highly challenging task because collection of real data itself is hard, to begin with. Consider a foggy scene, once the input hazy image is collected, it would be almost impossible to retake the ground truth image when the conditions are normal. This is because there could be moving objects and other external factors. It is even more tedious to collect hazy images of all kinds of distributions such as hazy images caused due to smoke, haze images caused due to fog and so on.

The type of haze is said to be non homogeneous when the degradation of contrast is not uniform throughout the scene *i.e.* when there are different concentrations of fog, mist or smoke in different areas of the image. In this paper we try to solve the problem for non homogeneous situations using a carefully created synthetic dataset.

The main contributions of this paper include :

1. NH-RESIDE, a dataset with non homogeneous hazy images of varying concentrations and patterns of haze.
2. A model architecture that outperforms the state-of-

the-art papers on the NH-Haze dataset proposed by the NTIRE Non Homogeneous Dehazing Challenge at CVPR 2020, also achieving PSNR and SSIM values more than the winners of the challenge by a considerable margin.

3. A model architecture that outperforms the state-of-the-art papers in terms of processing time, facilitating real-time performance and lower utilisation of compute power.

The remainder of this paper is structured as follows. Section 2 describes relevant previous work. Section 3 describes the dataset creation procedure, followed by the model architecture in section 4. Section 5 presents samples of the dehazed output for real-world non homogeneous hazy images. This section also compares the results of the model on the NH-Haze dataset with state-of-the-art papers. Section 6 contains the empirical analysis and finally, section 7 concludes our work.

2. Related Work

Early methods in image dehazing were prior based methods to estimate the transmission map. DCP [11] is one such prior based method that uses a dark channel prior. However, these techniques can lead to colour distortion in a few occasions. More recently however, many end to end learning based methods have been proposed that directly recover the haze free image.

Dong *et al*. have proposed a method called MSBDN-DFF [1] which uses an encoder-decoder based architecture with error feedback and boosted modules. These are the two principal ideas in their work. The model implemented contains dense feature fusion primarily derived from the U-net architecture. They have also incorporated a boosting strategy called Strengthen-Operate-Subtract in the architecture of the decoder. Since the U-net model aggressively reduces the spatial information they have designed a dense feature fusion block to remedy the problem.

Chen *et al*. have presented a two stage framework called PSD [2] (Principled Synthetic to Real Dehazing) where they address the problem of generalizing the models to the real world. Their method involved supervised pre-training and then using unlabelled real hazy images to fine tune the network in an unsupervised fashion.

Liu *et al*. [6] have addressed the issue of domain shift between synthetic and real world by Disentangled Consistency mean-teacher network (DMT-net). This involves disentangling feature representation into 3 component maps using Disentangled Image Dehazing network (DID-net) and boosting single image dehazing by DMT-net with unlabelled real data.

Wang *et al*. [4] have proposed the FFA-Net architecture, which uses feature fusion to combine pixel attention mechanism along with channel attention mechanism. The have proposed the FA block based on attention which has given great results and has also been adopted by other methods. FFA-Net produces visually very pleasing images, however, it is computationally intensive.

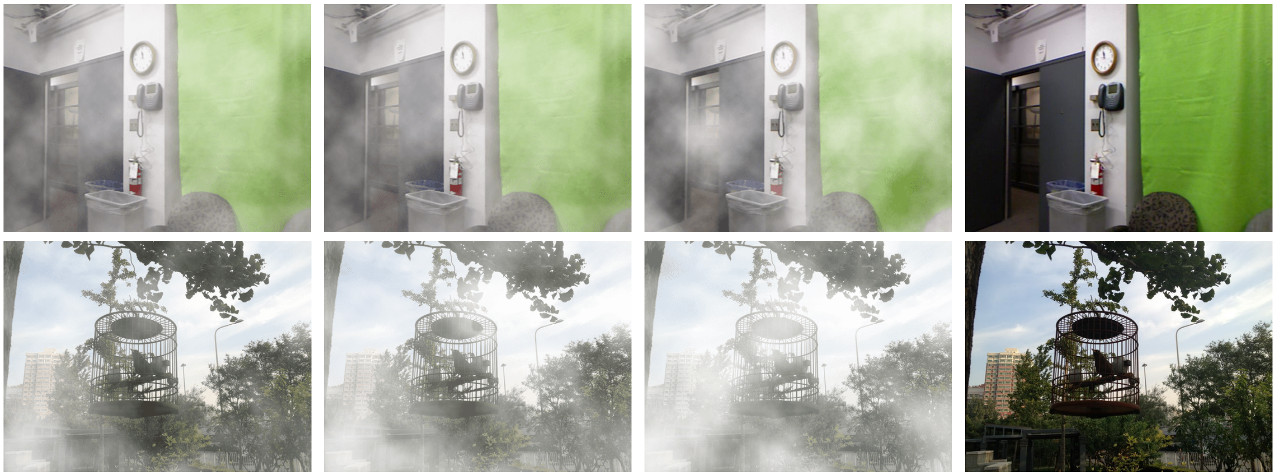
Shyam *et al*. [8] have proposed a greedy data augmentation technique for training the model. The augmented training has increased efficiency on non synthetic datasets and they also have focused on the dual challenge of domain and haze distributions that significantly reduces the performance of dehazing models.

Most of the deep learning models learn from the positive samples (clear image) and never exploit the negative samples (hazy image), Wu *et al*. [3] addresses this by pushing the prediction away from the hazy image and closer to a clear image. Their method uses less number of parameters giving higher PSNR values, however, the drawback is that when then negative samples were increased, the model would take longer training time than usual.

There is also a problem performance degradation when domain shifts are confronted, this happens when there is a density gap between the datasets. Chang *et al*. propose DAMIX [7] which generates a synthetic hazy image according to the haze density of the target domain. The network has 2 branches to generate preliminary dehazed images later both are combined by a weight generator, both focusing on different contaminated areas. DAMIX is a robust algorithm that generates natural samples while maintaining haze density diversity.

Zheng *et al*. [5] have employed a 3 CNN architecture to solve the problem of dehazing. They have also proposed an efficient architecture that can be used to dehaze high resolution images. The first CNN is used to obtain haze related features at a smaller resolution, the second CNN is used to learn guidance maps at full resolution and a third CNN is used to fuse the features into a haze-free image. The model architecture is efficient and scales well for high resolution images. Hence this model primarily solves the problem of high computational complexity of the other methods, and also achieves great PSNR values. The model implemented in our research is inspired by the UHD model.

Dehamer [16] is the latest and best performing model and is proposed by Chongyi *et al*. Integrating CNNs and transformers seems to be the main idea of their work. This is challenging, as some aspects of transformers may be undesired for the task of reconstruction. It is hard to directly get the benefits of transformers. They propose techniques such as modulation matrices to overcome some of the problems mentioned. Their model outperforms all other models on several image dehazing benchmarks. Our implementation outperforms theirs on the NH-Haze dataset and achieves competitive performance on other benchmarks while being radically efficient.



(a) Low (b) Medium (c) High (d) GT

Figure 1: NH-RESIDE Dataset

Figure 2: Image degradation when trained on NH-Haze

3. The Dataset

The first non homogeneous dehazing dataset created was the NH-Haze dataset. This dataset contains hazy images that were prepared using professional haze generating machines. It was created as part of the NTIRE 2020 Non Homogeneous Dehazing Challenge [12]. This dataset contains images that mostly look like smoke, hence when a different colour and distribution of haze is found, there are chances of image degradation as pointed out by the DAMIX paper [7]. The dataset also contains only 45 training samples which may not help in generalising well for other real world images. This is shown in figure (2) where the dehazed image suffers from artifacts. Training on the proposed dataset actually solves this problem and can be seen in section 5.

Hence there is a need for a realistic looking non homogeneous dataset. We propose the Non Homogeneous RESIDE dataset (NH-RESIDE) that was created using the clear images from the official RESIDE [13] dataset. RESIDE stands for REalistic Single Image DEhazing. Our dataset is manually prepared using a brush from Adobe Photoshop which generates cloud-like haze.

3.1. Creating realistic images

To generate a hazy image that looks real, we first apply one layer of haze with a random pattern and with a certain opacity level. The subsequent layer of haze contains a new random pattern with a different opacity level. This procedure creates one style of haze. We create 10 styles for each ground truth image, to get the exact same footprint as the official RESIDE dataset. We also apply each style to 30 different images as creating manually for each image would be very time consuming and laborious. Hence by following this process we generate over 14000 indoor hazy images and 21000 outdoor hazy images, with over 700 styles of haze, and every ground truth image contains 10 hazy images. We adopt two main principles for the dataset creation process.

1. Randomness - The patterns of haze created should be as random as possible to account for real world performance. Hence as explained before, we use multiple

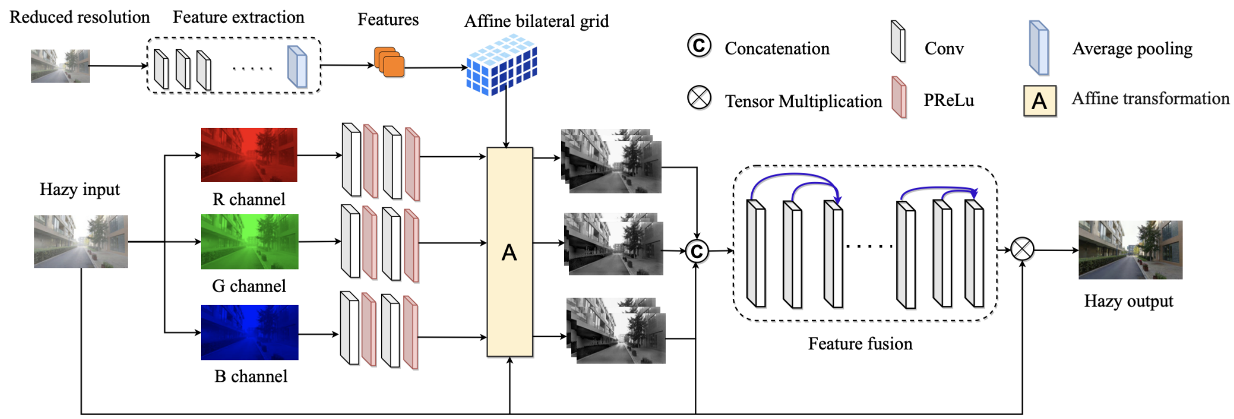


Figure 3: Model architecture of UHD [5]

layers of haze and since we apply one style of haze to 30 different images, for each of those images it would still be different context as there would be different objects in varying positions.

2. Representativeness - In the 10 styles that we create per image, we use up a range of opacity levels varying from very thin haze all the way to thick haze. This once again, ensures real-world performance.

Few images with varying thickness of non homogeneous haze are shown in figure (1) along with the ground truth images.

4. Model Architecture

We improve the model architecture of the UHD [5] method which results in the best PSNR and SSIM values on the NH-Haze dataset. The architecture consists of 3 ConvNets. The first CNN is used to obtain haze related features at a smaller resolution, the second CNN is used to learn guidance maps at full resolution and a third CNN is used to fuse the features into a haze-free image. This model architecture enables processing high definition images which take a lot of processing power when other methods are used. The model architecture from the original paper is shown in figure (3).

Firstly, in order to make the model more lightweight and improve processing time, we implement depthwise separable convolution [15] instead of normal convolution. The depthwise separable convolution consists of two stages, a depthwise operation and a pointwise operation. Instead of the filter convolving with the whole volume, we first convolve k filters if the volume has k channels, each one operating on a single channel, and then use a (1 x 1) convolution to add up the outputs from each channel as shown in figure 4. This would essentially allow us to operate with the same dimensions but with much fewer parameters and fewer floating point operations.

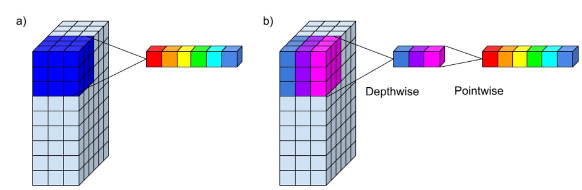


Figure 4: Depthwise separable convolution

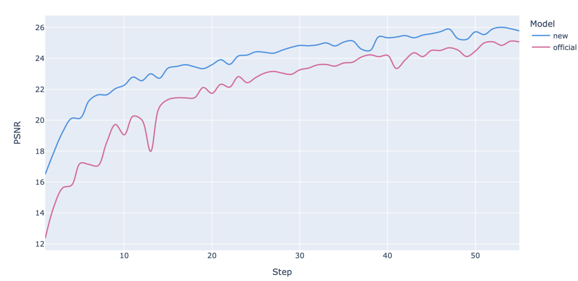


Figure 5: Faster convergence and better PSNR

Secondly, using depthwise separable convolutions also meant we could use a deeper U-Net [14] block capable of better feature extraction. The official implementation contains a U-net block which downsamples from 64 channels to 128, to 256, to 512 and then to 1024 channels before upsampling back to 64. Our implementation allows us to use a bigger module that downsamples from 64 channels to 128, to 256, to 512, to 1024, to 2048 and then to 4096 channels before upsampling back to 64 channels. And because of depthwise separable convolutions, it still means that we were able to reduce the model parameters from 34M to 25M

(1a) Input (1b) MSBDN [1] (1c) FFA-Net [4]

(1d) UHD [5] (1e) Dehamer [16] (1f) Ours

(2a) Input (2b) MSBDN [1] (2c) FFA-Net [4]

(2d) UHD [5] (2e) Dehamer [16] (2f) Ours

Figure 6: Qualitative comparison on real world images

while achieving faster convergence and better PSNR and SSIM values for the same hyperparameter configuration. This is shown in figure (5). This figure shows the training stats on the NH-Haze dataset. They were trained for about 7500 steps (1 step = 1 mini-batch update) with a learning rate of 0.0003 and batch size of 10. We used the Nvidia Tesla P100 GPU for this purpose.

We also trained our model on the new proposed NH-RESIDE dataset. The training was done along with the official RESIDE dataset. This was because the type of haze looks roughly same and the model would learn to dehaze both uniform and non homogeneous hazy images. Since the combined dataset contains more than 70,000 examples of hazy images we had to use a larger batch size of 52 and the much faster Nvidia A100 80GB GPU. The learning rate was set to 0.0003 and the model was trained for 15000 steps.

5. Qualitative Comparison

To show the effectiveness of the proposed dataset, we have selected very tricky real world non homogeneous images. The images are from Formula 1, in wet weather conditions. In fact, these are the hardest kind of hazy images because the hazy appearance is caused due to thin droplets of water. There is a limit to which information can be extracted, and it gets very challenging in situations like these.

As it can be seen from figure (6), 1a represents a non homogeneous hazy image with moderate concentration. All the state-of-the-art models perform a good level of dehazing, but our implementation produces the most dehazed effect after training with our NH-RESIDE dataset. 2a is another non homogeneous image but with much thicker haze. Once again, it can be seen that our implementation is more effective.

(a) Input (b) Our result (c) Ground truth

Figure 7: Results on NH-Haze

Method	Metric	
	Avg PSNR	Avg SSIM
DCP [11]	12.72	0.44
DehazeNet [19]	11.76	0.40
AODNet [17]	15.69	0.57
GridDehazeNet [18]	18.33	0.67
FFANet [4]	18.13	0.65
MSBDN [1]	17.97	0.66
Dehamer [16]	20.66	0.68
Ours	25.69	0.80

Table 1: PSNR, SSIM - NH-Haze dataset

Resolution	Grid [18]	Dehamer [16]	FFA [4]	Ours
960 x 540	7.68s	13.33s	60.09s	2.80s
1280 x 720	16.01s	25.61s	146.26s	4.63s
1920 x 1080	38.06s	64.05s	325.80s	6.34s

Table 2: Processing time comparison on an Intel core i5 processor

The improved UHD model was also trained on the NH-Haze dataset. This dataset was proposed by the NTIRE Non Homogeneous Dehazing challenge. It contains real hazy images that were created using professional haze generation machines. The dataset contains 55 image pairs, the first 45 images form the train split, the next 5 form the validation split and the last 5 images provided by the competition form the test split. The dehazed images on NH-Haze dataset can be found in figure (7). The enhanced dehazed images are remarkably similar to the ground truth. We have achieved the best PSNR and SSIM values on this dataset. We chose PSNR and SSIM in our work as are standard image metrics that are generally used to compare result with the ground truth, SSIM also takes into account the texture and other details. The best scores from the NTIRE Non homogeneous dehazing challenge were a PSNR of 21.91 and an SSIM of 0.71, both from different teams. According to the Dehamer [16] paper, their model achieved a PSNR of 20.66 and an SSIM of 0.6844. The model implemented in this paper outperforms these scores by a huge margin and achieves a PSNR of 25.69 and an SSIM of 0.80. These numbers are summarised in the table (1). Our outputs also do not contain any form of colour distortion.

6. Empirical Analysis

State-of-the-art methods like MSBDN, FFA-Net, Dehamer etc are computationally very intensive. It makes these models very hard to use in real-time applications like a video feed. It is also a challenge to process such networks on low compute environments like mobile phones. The modified and improved UHD [5] model implemented in this paper however, is light both in terms of parameters and processing time. This method can be easily implemented for low compute environments.

The processing time for these SOTA models is shown in the table (2). We conduct these experiments for 3 standard resolutions - 540p, 720p, 1080p. For video applications, 720p (1280 x 720) is almost always the standard resolution to achieve the best trade-off between quality and computational resources. Our implementation, while staying competitive in metrics like PSNR and SSIM has performed better and has the lowest processing time. This implementation achieves a minimum of 20x speedup over FFA-Net.

7. Conclusion

In this research, NH-RESIDE, a dataset with non homogeneous hazy images of varying concentrations and patterns of haze was proposed. The dataset was build on the principles of randomness and representativeness and contains images with haze of a wide range of opacity levels. A model architecture that outperforms the state-of-the-art papers on the NH-Haze dataset and also outperforms in terms of processing time, facilitating real-time performance and lower utilisation of compute power was implemented. Experimental results show that our dataset is effective in solving the problem for real-world non homogeneous images better than other datasets.

References

[1] Hang, Dong and Jinshan, Pan and Zhe, Hu and Xiang, Lei and Xinyi, Zhang and Fei, Wang and Ming-Hsuan, Yang. “Multi-Scale Boosted Dehazing Network with Dense Feature Fusion”, CVPR 2020.

[2] Chen, Zeyuan and Wang, Yangchao and Yang, Yang and Liu, Dong. “PSD: Principled Synthetic-to-Real Dehazing Guided by Physical Priors”, CVPR 2021.

[3] Haiyan Wu, Yanyun Qu, Shaohui Li, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, Lizhuang Ma. “Contrastive Learning for Compact Single Image Dehazing”, CVPR 2021.

[4] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong, Huizhu Jia. “FFA-Net: Feature Fusion Attention Network for Single Image Dehazing”. AAAI 2020

[5] Zhuoran Zheng, Wenqi Ren, Xiaochun Cao, Xiaobin Hu, Tao Wang, Fenglong Song, Xiuyi Jia. “Ultra High Definition Image Dehazing via Multi-Guided Bilateral Learning”. CVPR 2021.

[6] Ye Liu, Lei Zhu, Shunda Pei, Huazhu Fu, Jing Qin, Quing Zhang, Liang Wan, Wei Feng. “From Synthetic to Real: Image Dehazing Collaborating with Unlabeled Real Data”. ACMMM 2021

[7] Chia-Ming Chang, Chang-Sung Sung, Tsung-Nan Lin. “Density-Aware Data Augmentation for Unsupervised Domain Adaptation on Single Image Dehazing (DAMix)”. CVPR 2021.

[8] Pranjay Shyam, Kuk-Jin Yoon, Kyung-soo Kim. “Towards Domain Invariant Single Image Dehazing”. AAAI 2021.

[9] Cartney, E. J. ”Optics of the atmosphere: scattering by molecules and particles”. New York, John Wiley and Sons, Inc., 1976. 421 p.

[10] Narasimhan, S. G., and Nayar, S. K.” Chromatic framework for vision in bad weather”. In Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000.

[11] Kaiming He, Jian Sun, Xiaoou Tang. ”Single image haze removal using dark channel prior”. IEEE transactions on pattern analysis and machine intelligence 33(12):2341–2353.

[12] NTIRE 2020 Challenge on NonHomogeneous Dehazing.

[13] Li, Boyi and Ren, Wenqi and Fu, Dengpan and Tao, Dacheng and Feng, Dan and Zeng, Wenjun and Wang, Zhangyang. ”Benchmarking Single-Image Dehazing and Beyond”. IEEE Transactions on Image Processing, 2019.

[14] Olaf Ronneberger, Philipp Fischer, Thomas Brox. ”U-Net: Convolutional Networks for Biomedical Image Segmentation”. MICCAI, 2015.

[15] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam. ”MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications”. CoRR, 2017.

[16] Chunle Guo, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, Chongyi Li. ”Image Dehazing Transformer with Transmission-Aware 3D Position Embedding”. CVPR, 2022.

[17] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. Aod-net: All-in-one dehazing network. ICCV, 2017.

[18] X. Liu, Y. Ma, Z. Shi, and J. Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. CVPR, 2019.

[19] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao. An end-to-end system for single image haze removal. TIP, 2016.