

LCA-Net: Light Convolutional Autoencoder for Image Dehazing

Pavan A
CSE PESU
Bangalore, India
pavanappanna7@gmail.com

Adithya Bennur
CSE PESU
Bangalore, India
aditya.bennur@gmail.com

Mohit Gaggar
CSE PESU
Bangalore, India
gaggarmohit@gmail.com

Shylaja S S
CSE PESU
Bangalore, India
shylaja.sharath@pes.edu

Abstract—Image dehazing is a crucial image pre-processing task aimed at removing the incoherent noise generated by haze to improve the visual appeal of the image. The existing models use sophisticated networks and custom loss functions which are computationally inefficient and requires heavy hardware to run. Time is of the essence in image pre-processing since real time outputs can be obtained instantly. To overcome these problems, our proposed generic model uses a very light convolutional encoder-decoder network which does not depend on any atmospheric models. The network complexity-image quality trade off is handled well in this neural network and the performance of this network is not limited by low-spec systems. This network achieves optimum dehazing performance at a much faster rate, on several standard datasets, comparable to the state-of-the-art methods in terms of image quality.

Index Terms—Image dehazing, image pre-processing, light convolutional encoder-decoder

I. INTRODUCTION

Haze is an atmospheric phenomenon wherein the clarity of the sky is obscured due to the presence of smoke, dust and other dry particulate matter. This causes light to scatter and thus, deteriorates the image quality which in turn may affect the performance of several computer vision applications such as detection, classification and tracking. The atmospheric scattering model combines the haze-free input with the global atmospheric illumination, blended together with a transmission coefficient that models the haze effects. It is closely represented by the equation [1]:

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

where \mathbf{x} is the 2D pixel spatial location, $\mathbf{I}(\mathbf{x})$ is the observed haze image, $\mathbf{J}(\mathbf{x})$ is the haze-free scene radiance to be recovered, while $\mathbf{A}(\mathbf{x})$ and $\mathbf{t}(\mathbf{x})$ are two critical parameters. The first $\mathbf{A}(\mathbf{x})$, denotes the global atmospheric illumination, and $\mathbf{t}(\mathbf{x})$ is the transmission matrix.

Using the atmospheric model to build a neural network is also referred to as 'plug-in approach'. However, this approach may lack the features to detect intricate differences between a clean image and a hazy image. The viability of converting the problem of image reconstruction to a problem of estimation of \mathbf{A} and $\mathbf{t}(\mathbf{x})$ is very low. Besides, we cannot relate hazy and clear images with ease, reason being the relation may be too compound to be explained just by an equation. Thus, the plug-in approach may not be able to generalize the equation over

natural images. Our model simply builds a clear image from the hazy image by applying pooling and convolutions and does not refer to the above atmospheric model or the transmission map of the image for the process of dehazing of the image.

Image Dehazing is used in several image analyzing applications and other applications such as tracking, control systems, and intelligent surveillance where good quality images or videos are crucial for obtaining accurate results and optimum performance.

II. THE PROPOSED NETWORK

We shall propose a convolutional autoencoder in this section which is light both in terms of the architecture as well as its speed in dehazing an image. As the name suggests, this neural network has no special consideration towards image dehazing in its design. The following subsections outlines overall architecture, important features, main building blocks and the loss function used of the LCA-net.

A. Network Architecture

Figure 1 illustrates the overall architecture of LCA-Net. Our neural network consists of an encoder-decoder type of architecture. The input layer takes an image (512 x 512 x 3) which is padded and pushed to the first ReLU [2] activated convolutional layer consisting of 50 filters of size 3. This layer returns an output of original dimensions which acts as input for the average pooling layer which has a downsampling factor of 2, followed by the next convolutional layer where the cycle is repeated. This results in an encoded input of size 128 x 128 x 50. This stage is referred to as 'bottleneck'. The encoded input is then passed through 2 completely connected ReLU [2] activated dense layers consisting of 10 neurons each. Dense layers add an interesting non-linearity property, and can model any mathematical function. Thus, they help in extracting essential features efficiently. This output is passed to the decoder which consists of 2 pairs of deconvolutional and upsampling layers. The ReLU [2] activated deconvolutional layer consists of 50 filters of size 3 and the upsampling layer utilises a scale factor of 2. The resulting output of the decoder has a dimension of 512 x 512 x 50. The final output layer is a deconvolutional layer consisting of 3 filters of size 3 which

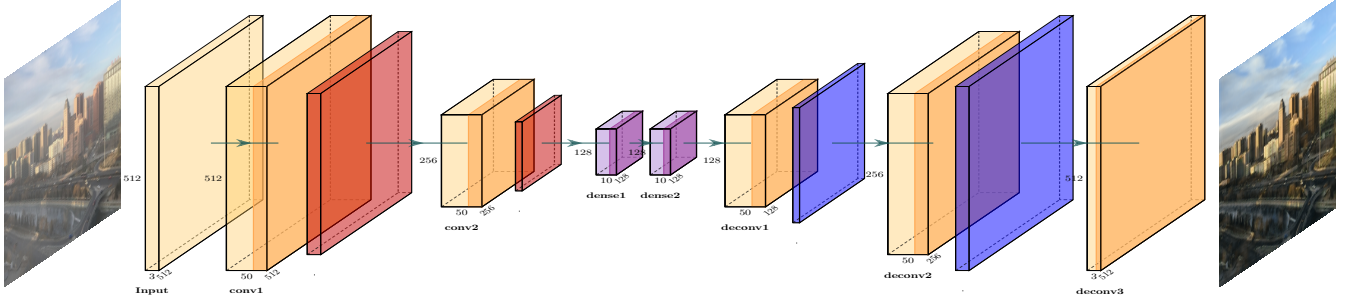


Fig. 1: The architecture of LCA-Net

are linearly activated producing a dehazed image whose size is identical to the input image.

B. Loss function and Training setting

The loss function used in this model is mean squared error(MSE). There are high chances of encountering unexpected changes in intensity of pixel values. MSE helps average the variation of extreme values, supporting a smooth gradient descent making it an optimal loss function.

$$L = \frac{1}{N} \sum_{x=1}^N \sum_{i=1}^3 \left\| \hat{J}_i(x) - J_i(x) \right\|^2 \quad (2)$$

where $J_i(x)$ is the intensity of the i^{th} color channel of pixel x in the output image (the ground truth), and N is the total number of pixels. Figure 2 shows the loss function decreasing with increase in the number of epochs.

The optimizer used is ADAM [3] in its default settings which are: learning rate = 0.001 , $\beta_1 = 0.9$, $\beta_2 = 0.999$.

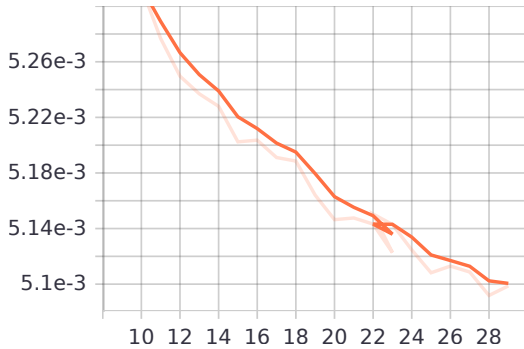


Fig. 2: Epoch loss

C. Datasets

Like most of the previous CNN based dehazing techniques our autoencoder is also trained on a custom dataset. Our dataset consists a total of 2061 clear images, we then select 35 varying haze levels on each image to constitute a total of 72135 haze images borrowed from the RESIDE dataset [4]. There are 3 test sets borrowed from RESIDE-standard [4] dataset. The test sets, known as the Synthetic Objective

Testing Set (SOTS-outdoor and SOTS-indoor) contain 500 synthetic outdoor and 500 synthetic indoor hazy images respectively. They are generated using images in NYU2 and are different from the training set images. The third test set is known as the Hybrid Subjective Testing Set (HSTS), which contains 10 synthetic outdoor and 10 real-world outdoor hazy images to evaluate qualitative visual performance.

III. BENCHMARKING AND RESULTS

The performance of the neural network is determined by evaluation metrics such as PSNR(Peak Signal to Noise Ratio) and SSIM(Structural Similarity Index). PSNR is essentially the ratio between maximum intensity and MSE. High PSNR value implicates low noise, the vice-versa is also true. SSIM as the name says measures the similarity between 2 images. It also improves on methods like PSNR and MSE.

Table I, Table II and Table III shows the PSNR and SSIM scores of our LCA-Net compared to nine other state-of-the-art methods [5]–[13] on the HSTS, SOTS-indoor and SOTS-outdoor test sets respectively. The results of previous dehazing methods for Table I, Table II and Table III are from, [4] and [13]. Nonetheless, since our model benchmarks on the same dataset, the comparison is fair.

Table I shows the PSNR, SSIM scores and average time taken to dehaze an image. Our model when compared to other state-of-the-art method [5]–[13] comes in first and the SSIM only comes second to DehazeNet [10] in outdoor HSTS. Due to the light nature of our model it outperforms all other models in the run-time per image, making it one of the fastest methods for image dehazing.

Table II shows the dehaze results on SOTS-indoor dataset. Our model performs inadequately when bench-marked on indoor SOTS and low-light images as they were trained on outdoor images. This phenomenon can also be explained due to the drastically varying feature vectors of indoor and outdoor images. Deficient performance on low-light images is also explained by the fact that low-light images have contrasted feature vectors. Figure 3 illustrates the comparison

	DCP [5]	FVR [6]	BCCR [7]	GRM [8]	NLD [9]	DehazeNet [10]	MSCNN [11]	AOD-Net [12]	CAE [13]	LCA-Net
PSNR	14.84	14.48	15.08	18.54	18.92	24.48	18.64	20.55	20.08	24.734
SSIM	0.7609	0.7624	0.7382	0.8184	0.7411	0.9153	0.8168	0.8973	0.8169	0.8951
Time(s)	1.62	6.79	3.85	83.96	9.89	2.51	2.60	0.65	1.13	0.3546

TABLE I: Dehaze results on synthetic hazy images from HSTS

	DCP [5]	FVR [6]	BCCR [7]	GRM [8]	NLD [9]	DehazeNet [10]	MSCNN [11]	AOD-Net [12]	CAE [13]	LCA-Net
PSNR	16.62	15.72	16.88	18.86	17.29	21.14	17.57	19.06	24.56	18.23
SSIM	0.8179	0.7483	0.7913	0.8553	0.7489	0.8472	0.8102	0.8504	0.9126	0.7808

TABLE II: Dehaze results on Indoor hazy images from SOTS

	DCP [5]	FVR [6]	BCCR [7]	GRM [8]	NLD [9]	DehazeNet [10]	MSCNN [11]	AOD-Net [12]	LCA-Net
PSNR	18.54	16.61	17.71	20.77	19.52	26.84	21.73	24.08	23.37
SSIM	0.71	0.7236	0.7409	0.7617	0.7328	0.8264	0.8313	0.8726	0.8763

TABLE III: Dehaze results on Outdoor hazy images from SOTS

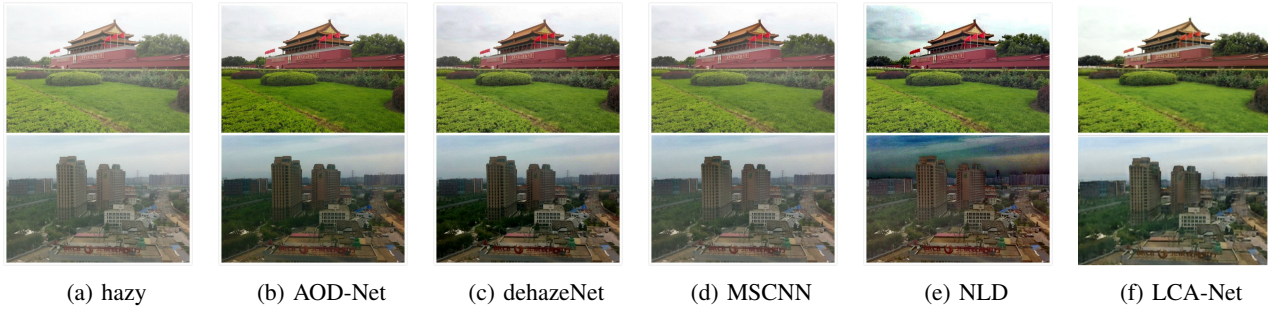


Fig. 3: Comparison of various model benchmarks

of different models' performances on standard hazy images from the RESIDE dataset [4].

Table III shows the PSNR and SSIM scores of models [5]–[12] on SOTS-outdoor images. Our model comes third behind DehazeNet [10] and AOD-Net [12] in PSNR and first in the SSIM metric. Therefore, our model is efficient in dehazing outdoor real world and synthetic images.

IV. CONCLUSION AND FUTURE WORK

This paper presented a light convolutional autoencoder which quickly helps remove haze in images. Further it performs better than other state-of-the-art image dehazing methods with regard to PSNR and the time taken to dehaze an image. The image quality is comparable to the state of the art methods. Since it is a light weighted network, it is also compatible with low-spec computers. The downside of this network is its inefficient performance on low-light images and dense-haze images and our future work would be centered

on eliminating this flaw to provide fine performance on all categories of images.

REFERENCES

- [1] Srinivasa G Narasimhan and Shree K Nayar. Vision and the atmosphere. *International journal of computer vision*, 48(3):233–254, 2002.
- [2] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3, 2013.
- [3] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [4] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019.
- [5] 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, Dec 2011.
- [6] J. Tarel and N. Hautire. Fast visibility restoration from a single color or gray level image. In *2009 IEEE 12th International Conference on Computer Vision*, pages 2201–2208, Sep. 2009.
- [7] Gaofeng Meng, Ying Wang, Jiangyong Duan, Shiming Xiang, and Chunhong Pan. Efficient image dehazing with boundary constraint and contextual regularization. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2013.

- [8] Chen Chen, Minh N. Do, and Jue Wang. Robust image and video dehazing with visual artifact suppression via gradient residual minimization. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 576–591, Cham, 2016. Springer International Publishing.
- [9] Dana Berman, Tali treibitz, and Shai Avidan. Non-local image dehazing. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [10] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, Nov 2016.
- [11] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 154–169, Cham, 2016. Springer International Publishing.
- [12] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [13] R. Chen and E. M. Lai. Convolutional autoencoder for single image dehazing. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 4464–4468, Sep. 2019.