Deep Fully Convolutional Regression Networks for Single Image Haze Removal

Xi Zhao [#1], Keyan Wang [#2], Yunsong Li [#3], Jiaojiao Li [#4]

[#] *State Key Laboratory of Integrated Service Network, Xidian University, Xi'an 710071, China*
[1] xizhao24@gmail.com, [2,3] {kywang,ysli}@mail.xidian.edu.cn, [4] jjli@xidian.edu.cn

***Abstract*—Haze removal for a single image is known to be a challenging ill-posed problem in computer vision. The performance of existing prior-based image dehazing methods is limited by the effectiveness of hand-designed features. The emerging convolutional neural network (CNN) based approaches can remove haze with the automatically learned intrinsic mapping between the input hazy images and their corresponding transmission maps, but the recovered haze-free images sometimes are still unsatisfactory. In order to improve the dehazed images, we aim to develop an effective deep fully convolutional regression network for more accurate transmission estimation. Our network is an end-to-end regression system which take input of arbitrary size hazy image and predict correspondingly-sized transmission map. To train and evaluate deep network for image dehazing efficiently, we develop new outdoor synthetic training set respectively. In addition, we fully compare the existing CNN-based haze removal approaches with our algorithm on real-world images and our synthesized benchmark dataset. The experimental results demonstrate that our trained regression model achieves superior dehazing performance than the current state-of-the-art methods.**

***Index Terms*—Haze removal, convolutional neural network, transmission regressed model, deconvolution**

I. Introduction

Outdoor images captured in bad weather (*e.g.*, haze or fog) usually suffer from low contrast and poor visibility due to tiny particles in the atmosphere scattering the light through the transmission medium. The image degradation not only negatively impacts human perception, but also poses obstacles to many automatic computer vision tasks, such as automatic driving system, image classification and video analysis and recognition. Therefore, haze removal is highly desired to improve both the quality of the images and the performance of the computer vision systems.

Haze removal is a challenging inverse problem, its solution heavily hinges on the use of additional information or proper priors/assumptions of the underlying physical scenes. More recently, significant progresses have been made in single image dehazing based on the atmospheric scattering model [1] owing to the use of better assumptions and priors. Since the key issue of haze removal is estimating an accurate transmission map from an observed hazy image, almost all the existing dehazing approaches focus on the estimation of the transmission map, and many effective methods have been developed, among which the representative works include [2], [3]. He *et al.* [2] discover a useful prior named dark channel prior (DCP) for image dehazing, but their method is computationally expensive and fails to effectively handle the scene with bright objects in whose colors are close to the atmospheric light. To overcome these limitations and further improve dehazing quality, Meng *et al.* [3] adopt a contextual regularization method combined with boundary constraint to remove haze from a single image. Despite the significant progress in haze removal, the performance of existing state-of-the-art methods is still restricted by the effectiveness of these hand-designed haze-relevant priors or assumptions, such as dark channel, maximum contrast and so on. Consequently, the dehazed results are less satisfactory for some situations.

Inspired by the fact that a human brain can quickly identify the thickness of haze and recognize the vague objects from a hazy scene without any additional information, some scholars attempt to utilize the deep neural networks, a biologically inspired model for image dehazing. For example, as the first usage of deep learning for image dehazing, a trainable system based on CNN called DehazeNet is proposed by Cai *et al.* [4], which can learn the intrinsic mapping relationships between hazy image patches and their transmissions. Similarly, Ren *et al.* [5] develop a multi-scale convolutional neural network for effective transmission estimation. These dehazing approaches based on CNN achieve better results than aforementioned traditional methods because the CNN-based deep networks can learn haze-relevant features automatically via a data-driven approach, avoiding the limitation of the hand-crafted features. However, Cai *et al.* [4] and Ren *et al.*'s [5] trained models are limited by their patch-based training set and the indoor synthetic training data respectively.

In this paper, aiming to improve the dehazing performance, we propose an effective deep fully convolutional regression network (DFCRN) for more accurate transmission estimation. Our contributions are mainly in three-fold. Firstly, we develop an end-to-end deep regression network endowed with novel UpConv units, which can automatically learn diverse haze-relevant features and then accurately predict the transmission map for an input hazy image of arbitrary size. Secondly, we apply a outdoor synthetic dataset for network training and build a new benchmark to evaluate the performance of the various haze removal algorithms. The evaluation dataset which consists of diverse hazy images and their corresponding transmission maps, is built by synthesizing clear images and their ground truth depth maps from the Make3D dataset [6]. Finally, we fully evaluate the proposed method against the existing CNN-based haze removal methods as well as some

978-1-5386-0462-5/17/$31.00 VCIP 2017, Dec. 10 – 13, 2017, St Petersburg, U.S.A.

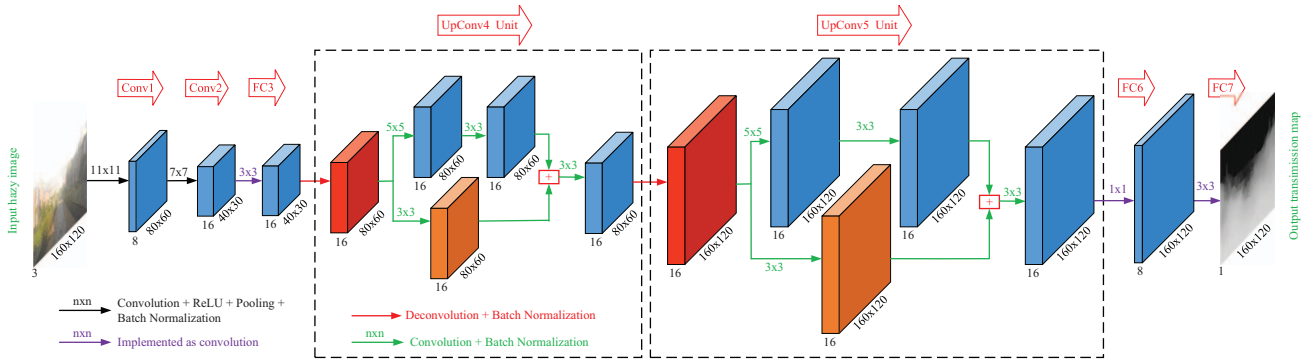


Fig. 1. Network architecture: Given a hazy image, the network maps the input image to feature maps and finally estimate the corresponding scene transmission. Conv x – Convolution layer + ReLU + Pooling + Batch Normalization; FC x – Fully connected layer (implemented as convolution). The black dotted rectangles show the details of UpConv4 and UpConv5 units. Each UpConv unit consists of a deconvolution layer, several convolution layers and batch normalization layers. In each deconvolution layer, the stride is 2 and kernel size is 2×2. For clarity, the deconvolution layer marks as red and the skip layer marks as orange.

state-of-the-art traditional dehazing methods on a number of datasets including synthetic and real-world hazy images. The experimental results demonstrate the superior performance of our network.

II. Image Dehazing with DFCRN

A. Background

The formulation of a hazy image can be modeled by the following atmospheric scattering model [1]:

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

where $I(x)$ and $J(x)$ are respectively the observed hazy image and the latent clear image, A is the globally constant atmospheric light, and $t(x) \in [0, 1]$ is the medium transmission map. When the atmosphere is assumed to be spatially homogenous, the transmission $t(x)$ can be defined as:

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

where β is a constant standing for the scattering coefficient of the atmosphere and $d(x)$ is the scene depth of image.

The key issue in haze removal is to estimate an accurate transmission map from an input hazy image. Once given the estimated transmission, the latent haze-free image can be easily restored via a simple pixel-wise operation.

B. DFCRN Architecture

Fig. 1 illustrates the detailed architecture of our proposed network, called deep fully convolutional regression network (DFCRN). The entire deep network is an end-to-end system based on CNN, which takes a RGB hazy image as input and finally regresses its corresponding transmission map. It is mainly composed of two parts — *Conv units* and *UpConv units*. Every Conv unit consists of a convolutional layer, followed by a Rectified Linear Unit (ReLU) and a max-pooling, whereas each UpConv unit is composed of deconvolution, ReLU and convolution layers. The activation function used in each layer is ReLU, and each fully connected layer is implemented as convolution. In addition, we apply an extra batch normalization layer [7] in every Conv and UpConv unit so as to prevent overfitting and speed up convergence. We argue that the benefits of our developed UpConv units lie in expanding the feature map to ensure pixels-to-pixels regressed transmission map and fusing different level features to learn fine spatial structures. More details of the UpConv unit are described as follows.

In the DFCRN architecture, our goal is to obtain an output predicted transmission map with the same size as the input hazy image. But the pooling layer, designed for filtering noisy activations in a lower layer, always results in a shrinkage of the feature maps whose size is reduced to a quarter after each max-pooling. To resolve such issue, Ren *et al.* [5] directly apply an Un-sampling layer after every pooling layer. Inspired by the up-projection blocks in Laina *et al.*'s network [8] which aims to increase the spatial resolution of feature maps, we adopt the similar up-projection blocks in this paper and furthermore improve it by using a *deconvolution layer* to replace the un-pooling layer, as shown in Fig. 1 (see the red label in rectangle). That because the output activation map of un-pooling layer is enlarged yet sparse, whereas, the deconvolution layer in the UpConv unit which performs the reverse operation of convolution and pooling can generate enlarged and dense activation maps with finer spatial structures.

Moreover, there is a *skip convolution* layer in each UpConv unit (see the blue label in rectangle) that by-pass two convolutions and then is summed to their output. The skip convolution layer is used to solve the degradation or vanishing gradients problems in deep networks and fuse different scale features. Therefore, our estimated transmission map is able to preserve fine spatial structures with supervision from the different level fused features.

C. Training

1) Training data: Different from the synthetic indoor training data used in [5], we develop a new synthetic outdoor training set for network training. First, we collect 710 haze-free color images from the Internet and then use Liu *et al.*'s [9] depth map estimation model to generate their corresponding depth maps. Next, we randomly select 690 clean images

and their estimated depth maps to construct our training set. Given a clear image $J(x)$, the corresponding depth map $d(x)$, random atmospheric light A and eighteen randomly sampled $\beta \in [1.0, 2.5]$, we synthesize eighteen hazy images via Eq. (1) and (2). To diminish the uncertainty of variables in learning, the atmospheric light A is set to [1, 1, 1]. We finally have more than 12,000 hazy images and their transmission maps in the training set, which are all resized to 160×120 pixels. Similarly, we use the rest 20 clear images to generate 100 hazy images for cross validation (five random $\beta \in [1.0, 2.5]$).

2) Training method: We implement the network training based on Caffe [10] framework and use Euclidean loss function for training the network in Fig. 1. Additionally, the classical back-propagation algorithm and stochastic gradient descent approach are used to optimize our network, where mini-batch, weight decay and momentum are set to 8, 0.0001, 0.9, respectively. Initial learning rate is 0.0009, and decreased by 0.1 every 100K iterations. The optimization is stopped at 300K iterations (about 190 epochs). Training takes around 13 hours in a single Nvidia Tesla K80 GPU.

D. Haze removal with the trained network

The first step of dehazing is to estimate the atmospheric light A. In [5] the brightest pixel value among the 0.1% darkest pixels of the estimated transmission map is selected as A. This method is likely to be affected by estimation errors and results in color distortion. In this study, we choose the median value of all the 0.1% pixels having the largest dark channel values as A.

Given the estimated medium transmission $t(x)$ and atmospheric light A, the final scene radiance $J(x)$ is recovered easily by Eq.(1), which is rewritten as follows according to Meng *et al.*'s method [3]:

$$J(x) = \frac{I(x) - A}{[\max\{t(x), \gamma\}]^{\delta}} + A\,, \tag{3}$$

where γ is a constant with small value (typically 0.05) to avoid dividing by zero, and the exponent δ is a optional parameter used to adjust the estimated medium transmission.

III. Experiments

To testify the effectiveness of the proposed dehazing algorithm, we evaluate the performance of our method comprehensively on two different synthetic datasets and a number of real-world hazy images, with comparisons to several state-of-the-art approaches.

A. Quantitative comparison on synthetic images

We randomly select 50 clean images and their corresponding depth maps from the NYU Depth dataset [11] to synthesize 50 transmission maps and hazy images for performance evaluation. Similarly, we synthesize 40 hazy images and transmission maps from the Make3D dataset[1] [6] as well. All of these synthetic data are different from those for training.

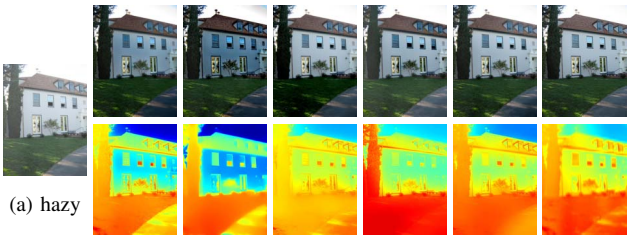
[1]http://make3d.cs.cornell.edu/data.html

TABLE I
AVERAGE PSNR AND SSIM OF DEHAZED RESULTS ON DIFFERENT SYNTHETIC DATASET.

Dataset	Metrics	He *et al.* [2]	Meng *et al.* [3]	DehazeNet [4]	Ren *et al.* [5]	Ours
Make3D	PSNR	16.04	21.42	20.01	22.64	**23.43**
	SSIM	0.85	0.94	0.87	0.94	**0.96**
NYU	PSNR	17.50	19.78	19.03	19.05	**21.91**
	SSIM	0.85	0.88	0.86	0.86	**0.91**

TABLE II
AVERAGE PSNR AND SSIM FOR TRANSMISSION MAP ESTIMATION ON DIFFERENT SYNTHETIC DATASET

Dataset	Metrics	He *et al.* [2]	Meng *et al.* [3]	DehazeNet [4]	Ren *et al.* [5]	Ours
Make3D	PSNR	13.74	17.38	16.82	**19.55**	19.53
	SSIM	0.78	0.90	0.89	0.88	**0.92**
NYU	PSNR	17.59	18.80	19.95	19.02	**21.76**
	SSIM	0.83	0.90	**0.92**	0.90	**0.92**



(a) hazy (b) [2] (c) [3] (d) [4] (e) [5] (f) ours (g) GT

Fig. 2. Qualitative comparison of different methods on developed Make3D synthetic haze dataset. GT means ground truth. The pseudo-color images in (b) – (g) are their corresponding transmission maps.

We perform a quantitative comparison against four state-of-the-art dehazing methods [2]–[5] using the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) evaluation metrics. As shown in Table I and Table II, our proposed algorithm performs better than the other state-of-the-art dehazing methods [2]–[5].

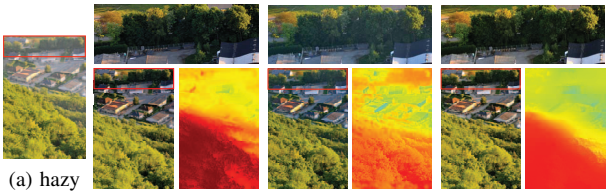
Fig. 2 shows visual comparisons of different approaches on the developed synthetic dataset. We observe that the dehazed images by [2] and [3] tend to exhibit darker visibility and color distortions (see the color of house) resulting from their overestimate of the haze thickness. Similarly, the DehazeNet [4] also suffers from these problems as shown in the Fig. 2(d) (see the grass in bottom right corner). Although Ren *et al.*'s method [5] achieves better dehazed results than [2]–[4], there is still an amount of haze remaining in Fig. 2(e) (see the tree and grass area). In contrast, the dehazed results and the estimated transmission maps by the proposed algorithm are closer to the ground truth, as depicted in Fig. 2(f).

B. Qualitative comparison on real-world images

Fig. 3 compares the dehazed results of our algorithm against the state-of-the-art methods on several real-world images. More results are provided in our supplementary material. As shown in Fig. 3(b) and 3(c), the dehazing methods of [2] and [3] often produce obvious color distortion, especially in the sky region. The dehazed results by [4] and [5] have some remaining haze as shown in Fig.3(d) and 3(e). In addition, [5] also generates over-enhanced results in some regions (for

(a) hazy image (b) He *et al.* [2] (c) Meng *et al.* [3] (d) DehazeNet [4] (e) Ren *et al.* [5] (f) ours

Fig. 3. Qualitative comparison of different methods on real-world hazy images.



(a) hazy (b) [4] (c) [5] (d) ours

Fig. 4. Comparison of transmission maps estimated by different CNN-based methods. The pseudo-color images in (b) – (d) are the corresponding transmission maps. The three top images are the zoomed area (see the red rectangle) of the corresponding dehazed image.

instance, the sky region of the second-row image in Fig. 3(e)). In contrast, our dehazed results are visually more comfortable in dense haze regions and free from oversaturation or artifacts.

Fig. 4 shows visual comparisons of the dehazed results by different CNN-based methods. The results obtained by existing CNN-based methods [4], [5] either overestimate the transmission map or still contain some haze (see the zoomed region in Fig. 4(b) and 4(c)). On the contrary, our algorithm estimates the transmission map more accurately and recovers better haze-free images, and contains natural color information and high contrast.

IV. CONCLUSION

In this paper, we propose a deep fully convolutional regression network (DFCRN) combined with a novel UpConv unit for haze removal. Different from the traditional prior-based approaches, the proposed framework automatically learns a regression model for transmission map estimation and subsequently performs image dehazing using atmospheric scattering model. In addition, we develop a new outdoor synthetic dataset based on the learned depth map from the single clear image to train and optimize the proposed network. We finally evaluate the performance of our algorithm aganinst the state-of-the-art dehazing methods including two CNN-based methods. Both the results from synthetic and real-world images demonstrate that our approach is capable of estimating the transmission map effectively and achieving better visibilty of dehazed images. In the future, we will extend our regression network to different kinds of hazy images, like the nighttime scene images *etc.*, and parallelize our trained model in GPU for real-time processing.

ACKNOWLEDGMENT

This work was jointly supported by the National Natural Science Foundation of China (No. 61301291), the 111 Project (B08038), and Shaanxi province science and technology innovation team project (2013KCT-02).

REFERENCES

[1] E. J. McCartney, "Optics of the atmosphere: scattering by molecules and particles," *New York, John Wiley and Sons, Inc., 1976. 421 p.*, 1976.

[2] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2011.

[3] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *IEEE international conference on computer vision (ICCV)*, 2013, pp. 617–624.

[4] 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*, vol. 25, no. 11, pp. 5187–5198, 2016.

[5] 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 (ECCV)*. Springer, 2016, pp. 154–169.

[6] A. Saxena, M. Sun, and A. Y. Ng, "Make3d: Learning 3d scene structure from a single still image," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 5, pp. 824–840, 2009.

[7] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[8] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab, "Deeper depth prediction with fully convolutional residual networks," in *3D Vision (3DV), 2016 Fourth International Conference on*. IEEE, 2016, pp. 239–248.

[9] F. Liu, C. Shen, and G. Lin, "Deep convolutional neural fields for depth estimation from a single image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5162–5170.

[10] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[11] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *European Conference on Computer Vision (ECCV)*. Springer, 2012, pp. 746–760.