INVERSE ATMOSHPERIC SCATTERING MODELING WITH CONVOLUTIONAL NEURAL NETWORKS FOR SINGLE IMAGE DEHAZING

Zehan Chen, Yi Wang, Yuexian Zou*

ADSPLAB, School of ECE, Peking University, Shenzhen 518055, China
*E-mail: zouyx@pkusz.edu.cn

ABSTRACT

Single image dehazing is an ill-posed problem. Most existing works use the atmospheric scattering model (ASM) [1] and some natural priors to dehazing. Recently, DehazeNet [2] was developed using deep learning approach achieves the state-of-the-art results on many test hazy images, which motivates us to adopt the deep learning approach in this study. After carefully research on the essential principle of dehazing using ASM, we develop a novel end-to-end convolutional neural network to efficiently implement the inverse ASM for single image dehazing, which is termed as IASM-Net. Specifically, the ASM is equivalently decomposed into three sub-model and the IASM-Net is designed to model these sub-model efficiently. Our study shows that IASM-Net is jointly optimized by minimizing the model error between the output and the clear ground truth, which gives better performance compared existing CNN-based transmission model approach [1,2]. To facilitate this study, a sufficient aerial image set containing more than 6K aerial photos is built to train and test our IASM-Net. Experimental results validate the effectiveness and efficiency of our well designed IASM-Net.

Index Terms— Image dehazing, convolutional neural network, atmospheric scattering model

1. INTRODUCTION

Generally, the core of image dehazing is to estimate the scene transmission of the given hazy image. Then, based on the atmospheric scattering model (ASM), it is easy to infer the clear image free from haze. Existing mainstream methods exploit various kinds of natural priors to compute the transmission in the scenes, and achieve satisfying results. Tan proposes a novel haze removal method [3] under the assumption that the local contrast of the haze-free image is much higher than that in the haze image. Moreover, they get impressive results by maximizing the local contrast of the image based on Markov Random Field. Inspired by the statistical properties drawn from a large image set, He et al. discover a simple but effective prior named as dark channel prior (DCP), and apply it to restore haze free image form single input image [4]. However, such priors only work effectively when their assumptions are well met, and their assumptions do not hold for all cases. For example, when the values of scene objects are close to the atmospheric light, the assumption of DCP may become invalid. Because, in this case, the value in dark channel is close to zero.

Recently, with the pervasive usage and the corresponding success of deep learning in computer vision, deep learning approaches also have been employed to solve dehaze problem [2]. As an example, DehazeNet proposed by Bolun Cai, applies a convolutional neural network (CNN) to learn the transmission function using hazy image in supervised manner [2], which gives state-of-arts. It is noted that with the estimated transimission function from DehazeNet, the dehazed image is computed by using the atmospheric scattering model (ASM). Motivated by the success of DehazeNet and our aerial image dehazing task in hand, we carried out experiments to dehaze the aerial images and we found that DehazeNet cannot give good results in cetain conditions, such as there are green region. Some examples can be found in Figure 6. These observations tell us the dehazing approach proposed in [2] can be improved.

From application aspect, unmanned aerial vehicles (UAVs), considering their flexibility and utility, have been employed to carry out various tasks, such as city surveillance, aerial imagery, and so on. In aerial imagery, the visibility of the taken images is of great significance, while in the most industrialized cities haze does degrade the visual quality of images. Therefore, image dehazing is a highly desired technique, especially in the urban aerial imagery. In this study, we focus on the aerial image dehazing task for unmanned aerial vehicles and make an effort to develop more efficient dehazing method using deep learning approach.

As existing studies show that the hazy images can be modelled by ASM which is a hazy image generation model and essentially show how the hazy images is formed. Obviously, the inverse process of ASM is exactly the dehazing model (IASM). Inspire by this, we firstly propose an end-to-end fully convolutional neural network to model IASM and results in the IASM-Net. To satisfy the assumptions developed for ASM and IASM, we work on image patch domain. Specifically, for the IASM-Net, its traning pairs are hazy image patches and their ground truth clean image patches, which is an end-to-end dehazing method. Different from CNN-based approaches and DehazeNet, we do not estimate the intermeddle transmission model. Moreover, to further improvement the model convergence and computational efficiency, the IASM is decomposed into three sub-models and three sub-systems corresponding to the three sub-models are carefully designed, which is introduced in Section 2 in details.

Summarily, our main contributions are given as follows.
1) A simple but high-performance end-to-end method, IASM-NET, for single image dehazing is proposed..
2) An aerial image dataset for dehazing task is built, which contains more than 6000 hazy aerial images.
3) Intensive experiments have been conducted to evlaute the dehazing performance. Comparisons are also given with the sate-of-arts.

The rest of this paper is organized as follows. In section II, we analyse and decompose ASM for haze removal, and then the design of our proposed IASM-Net is elaborated. In section III, our self-built aerial image set is described, and extensive experiments on natural hazy images are given to demonstrate the performance of IASM-Net. The conclusion is drawn in section IV.

2. METHODOLOGY

This section firstly describes the principle of ASM and IASM for dehazing task. Then the design of IASM-Net is given.

2.1. Dehazing with inverse atmospheric scattering model

In the atmospheric scattering model, the generation of a hazy image is given by:

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

where I is the observed hazy image, J is the real scene image (or clean image), T is the transmission matrix where $T(x,y) \in (0,1)$. A is the global atmospheric light factor, and (x,y) is the pixel index. Usually A can be considered as a constant due to its homogeneousness.

With simple manipulation from (1), $J(x,y)$ can be determined as follows:

$$
\begin{aligned}
J(x,y) &= \frac{I(x,y) - \big(1 - T(x,y)\big)A}{T(x,y)} \\
&= \frac{1}{T(x,y)}(I(x,y) - A) + A
\end{aligned} \tag{2}
$$

From the second line of (2), it is observed that the real scene $J(x,y)$ can be represented as three sub-models: 1) light normalization sub-model formulated by $I(x,y) - A$; 2) reciprocal transmission sub-model formulated as $1/T(x,y)$, and 3) fusion sub-model formulated by $\frac{1}{T(x,y)}(I(x,y) - A) + A$.

In details, for the light normalization model, it only contains a simple linear operation. For the reciprocal transmission sub-model, it is a reciprocal process of the transmission map according to some prior describe in [1] and [4]. Following [1,4], we derive how to determine $T(x,y)$. In most of the non-sky patches, at least one color channel has very low intensity at some pixels and its value tends to be zero. So we have:

$$\min_{c \in \{r,g,b\}} \left(\min_{(x',y') \in \Omega(x,y)} J^c(x',y') \right) = 0 \tag{3}$$

where J^c is the color channels of J and $\Omega(x,y)$ is a local patch centered at (x,y), We assume the atmospheric light A is give and A is always positive, and from (1) and (3) we obtain:

$$\frac{I(x,y)}{A} = \frac{T(x,y)J(x,y)}{A} + (1 - T(x,y)) \tag{4}$$

$$\min_{c \in \{r,g,b\}} \left(\min_{(x',y') \in \Omega(x,y)} \frac{J^c(x',y')}{A} \right) = 0 \tag{5}$$

we take the min operation in the local patch on the haze imaging Equation (4) and then take the min operation on the RGB channel, and obtain:

$$\left(\min_{c \in \{r,g,b\}} \left(\min_{(x',y') \in \Omega(x,y)} \frac{I^c(x',y')}{A^c} \right) \right) =$$

$$T(x,y) \left(\min_{c \in \{r,g,b\}} \left(\min_{(x',y') \in \Omega(x,y)} \frac{J^c(x',y')}{A^c} \right) \right) + (1 - T(x,y)) \tag{6}$$

where I^c and A^c are the color channels of I and A. Observe (5) and (6), we have:

$$\left(\min_{c \in \{r,g,b\}} \left(\min_{(x',y') \in \Omega(x,y)} \frac{I^c(x',y')}{A^c} \right) \right) = 0 + (1 - T(x,y)) \tag{7}$$

So the transmission matrix T can be computed using I and A in the following way:

$$T(x,y) = 1 - \omega \min_{c \in \{r,g,b\}} \left(\min_{(x',y') \in \Omega(x,y)} \frac{I^c(x',y')}{A^c} \right) \tag{8}$$

ω is a constant parameter to keep a little haze for the distant objects. It is obvious that T is the result from the nonlinear mapping of the local patches from I when considering A is constant. Moreover, T varies smoothly or stay stable locally according to the patch-based operation in (8).

Based on the analysis and decomposition of IASM above, we make an effort to incorporate the structure of the sub-models in our end-to-end IASM-net design. The details are given next.

2.2. The design of IASM-Net

First of all, the IASM-Net is an end-to-end dehazing network working based on image pataches. IASM-Net takes a hazy image as input and output the dehazed one. Since dehazing process is patch-based and the operation in every patch has no difference, we choose convolutional neural network (CNN) and just make use of the CNN's local sensing and weight sharing properties, without fully connection layer [5].

According to the IASM model given in the second line of (2), essentially, the designed IASM-Net will implement the nonlinear mapping in the form of $F(I)$, where I is the hazed image pactch. Considering the sub-model explained in Section 2,1, we design F to consist the following three modules:

1. The transmission module (TM): this module estimates local transmission on the extracted overlapping patches from the hazy image I.
2. The light normalization module (LNM): this module calculates the global atmospheric light based on extracted overlapping patches from the hazy image I.
3. The fusion module (FM): this module fuses the transmission information from the transmission module and the light normalization module into one image, and tunes it for better reconstruction and visual appealing.

The architecture of our proposed IASM-Net is illustrated in Figure 1. Every convolutional layer has a kernel size of 3×3 except conv4 which has a kernel size of 5×5. And every operation of convolutional layer has a padding size of 1 except conv4 which has a padding size of 2. The conv1, conv2 and conv3 output 32 feature maps while the conv4 outputs 3 feature maps. So we input a three channels RGB image to the IASM-Net and obtain an output of three channels RGB image. This so called end-to-end operation makes our method simple and efficient. As shown in Figure 1, the transmission module consists of four layers, including two convolutional layers and two Rectified Linear Unit (ReLU) layers [6]. The nonlinear mapping brought by the convolutional operation and nonlinear active function can well estimate the transmission map using sufficient training data. Besides, the convolutional operation introduced by the convolutional layers ensure the local properties of T. Thus, transmission module does inherit the core properties of the light normalization operation in IASM. Let $F_{TM}(I)$ denote the output of transmission module, and it can be denoted as:

$$F_{TM}(I) = \text{ReLU}(W_2 \text{ReLU}(W_1 I + B_1) + B_2) \tag{9}$$

where W_i and B_i are the weights and biases of the convolutional layer Conv$i$, and $\text{ReLU}(x) = \max(x, 0)$.

Referring to Figure 1, for the light normalization module, one convolutional layer with a fixed bias (which is set to $-A$) is used. Here the fixed bias is just used to simulate the light normalization operation. And the convolutional operations are employed to slightly adjust the normalized hazy images for better local perception. Let $F_{LNM}(I)$ denote the output of light normalization module, and it is given as:

$$F_{LNM}(I) = W_3 I - A \tag{10}$$

From Figure 1, we can see that, for the fusion module, it firstly applies an elementwise layer with the product operation to
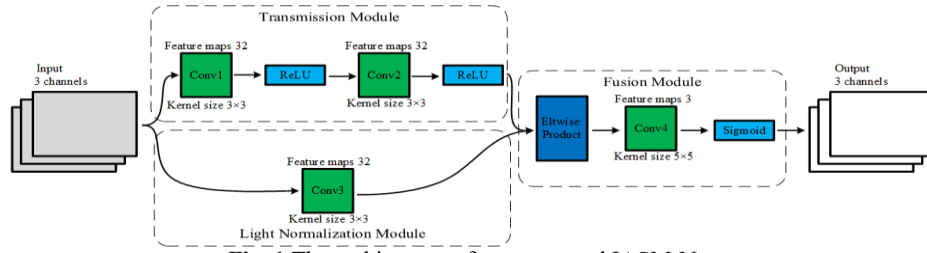
Fig .1 The architecture of our proposed IASM-Net

combine the output from the transmission module and light normalization module respectively. Then, the fusion module employs a convolutional layer with a fixed bias (which is set to A) and a sigmoid function to tune the aforementioned product for better dehazing effects. Let $F_{\text{FM}}(I)$ denote the output of fusion module, and it is computed as:

$$F_{\text{FM}}(I) = S(W_4(F_{\text{TM}}(I) \odot F_{\text{LNM}}(I)) + B_4) \qquad (11)$$

where $S(x) = 1/(1 + e^{-x})$ and $\odot$ is the Hadamard product.

Let I_i and J_i denote the i_{th} input hazy image and real scene image, respectively, and $F(I_i)$ is the output from our IASM-Net. Θ is the parameters of the IASM-Net. Given a set of the hazy images $\{I_i\}$ and its corresponding real scene images $\{J_i\}$ where $i = 1, 2, \ldots, N$, we adopt the following commonly used L_2 loss function:

$$l(\Theta) = \frac{1}{N} \sum_{i=1}^{N} ||F(I_i) - J_i||^2 \qquad (12)$$

The loss is minimized by stochastic gradient descent with the standard back propagation.

It is noted that most of the mainstream DNN based dehazing methods are designed to learn the the transmission model, and then apply ASM to estimate the real scene image. We are not exactly sure the theoretical mechnism difference between our IASM-Net and the conventional DNN-based dehazing methods including DehazeNet [2]. Our experimental results show that our IASM-Net ourperforms the conventional approaches. One of the possible argument is that, for the conventional DNN-based dehazing methods, the combination of optimal models doesn't guarantee a global optimum.e conventional DNN-based dehazing methods, the combination of optimal models doesn't guarantee a global optimum.

3. EXPERIMENTS AND RESULTS

In this section, we firstly introduce how we build the training and testing datasets. Then the experimental settings and relevant compared results are given.

3.1. Dataset

Here we build an aerial image set for dehazing task, which contains more than six thousands of high-resolution (4000×3000 with RGB three channels) aerial images captured by unmanned aerial vehicle. Specifically, only a small portion of them (50) is used to train the IASM-Net. In order to improve the generalization performance, we also collect another 20 high-resolution images from the internet that contain landscape and portrait. These 70 images are split into two parts. The first part owns 63 images and is used for training. The rest 7 images are used for testing.

One challenge of training a CNN model to dehaze is that the hazy image of natural scene and its medium transmission maps are not massively available. To handle such challenge, hazy image synthesize method is employed. Specifically, based on the prior assumption we can regard the medium transmission map $T(x, y)$ as a constant t in a small patch. Hence hazy images can be generated by a given value of t based on the following equation.

$$I_i = J_i t + A(1 - t) \qquad (13)$$

where A is set to 1 and t is a random values between 0.6 and 1. The synthesized hazy images are generated based on patches. For training set, patches whose size is 32×32 are extracted from these high-resolution images with a step size 16. Totally, the 63 high-resolution images and their corresponding hazy images will generate near 180K patch pairs (hazy ones and clear ones), meaning N in (12) is around 180K in our experiments. We follow the same way to construct the testing set and the testing set has near 20K image patch pairs.

3.2. Experimental settings

We implement our IASM-Net using Caffe package [7]. The learning rate is fixed to 0.0001. The maximum training iteration is set to 30000. The tendency of testing loss is similar to the training loss. They are shown in Figure 2. The value of training loss and testing loss declines soon and reaches stabilization within 1000 iterations. So our IASM-Net is easy to train.
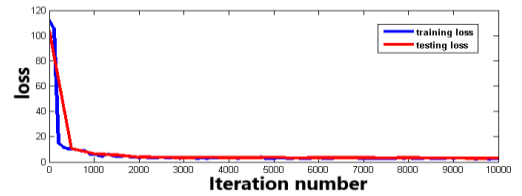

Fig. 2 the tendency of IASM-Net's training loss and testing loss with iteration number

3.3. Experimental Results

In order to prove the validity of our neural network architecture design, we only change the 4 convolutional layers to be cascaded and remain all the other setting to perform the controlled experiments (termed as plainNet). We show the neural network architecture of plainNet in Figure 3. The value of training loss and testing loss reach stabilization within 2000 iterations which is 2 time of the IASM-Net's. So the design of our IASM-Net can also bring the advantage of neural network training. The experimental results are show in Figure 4. We can see that our well designed neural network IASM-Net can provide excellent dehazing results while the plain neural network plainNet can't. So we can prove the validity of our IASM-Net.

To verify whether our IASM-Net can generate the transmission map of a hazy image, we extract the output of the transmission module and show it in Figure 5. In Figure 5 we can see that our design of transmission module work well.

The compared algorithms include He's DCP [4] based method and Cai's DehazeNet [2]. From the Figure 6, we show two dehazing results and its details of our method and the compared algorithms. It is noted that the result of our IASM-Net outperforms that of He's method and DehazeNet. The first row of the Figure 6 shows the local patches of the sky region, where it is clear that He's method is not able to deal with this case. The sky region in the result yielded by DehazeNet is slightly purple while IASM-

Fig .3 Architecture of the plain neural network (plainNet)

(a)   (b)   (c)
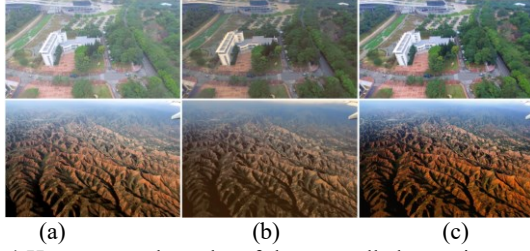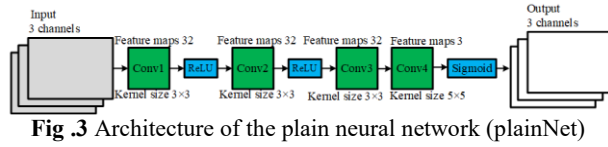
Fig .4 Haze removal results of the controlled experiments. (a) Input hazy images. (b) Results of plainNet. (c) Results of our IASM-Net.

(a)   (b)   (c)   (d)
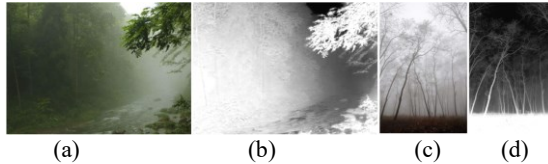
Fig .5 The transmission map generated by the transmission module of our IASM-Net. (a)(c) hazy input image. (b)(d) corresponding transmission map.

Net's result is better on visual perception. In the third row we show another local patches, where we can see that the yellow flowers and the river in the result of IASM-Net's is clearer than others. In the fourth and fifth rows, it is obvious that He's method failed again to deal with the sky region. And it has some color bias. The result of DehazeNet in this case is too gloomy. From the third picture of the fifth row we can see that the sky is a little dim and the trees almost loss its texture. Meanwhile, the result produced by IASM-Net recovers more details of the trees' texture than the result of DehazeNet. And it's sky region looks nature.

The blind test of subjective evaluation was performed to confirm the excellence of our method. We selected 10 hazy aerial images form the internet for dehazing with He's method, DehazeNet and our method. Each method produced 10 outputs. We shuffle the results of each hazy image and respectively let 6 testing persons to pick out the clearest one with respect to the original hazy image. The method of the results that was picked out the most number of times implies it is the outstand one. The Figure 7 shows the result of the blind test and we can know that our method has the hightest times of being picked out.
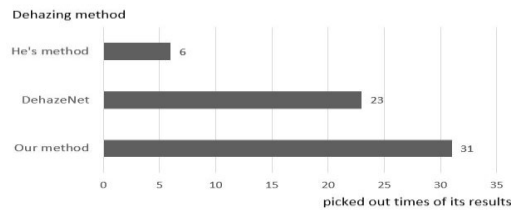
Fig. 7 the result of subjective evaluation blind test for three dehazing methods

We also compare our method with the compared algorithms on peak signal-to-noise ratio (PSNR). 20 high-resolution images are randomly chosen from our self-built aerial image set as the ground truth images. Based on the aforementioned synthesize method, we synthesize hazy images using the ground truth images and the value of t varies from 0.6 to 0.8 with a step size as 0.1. Then hazy

(a)   (b)   (c)   (d)
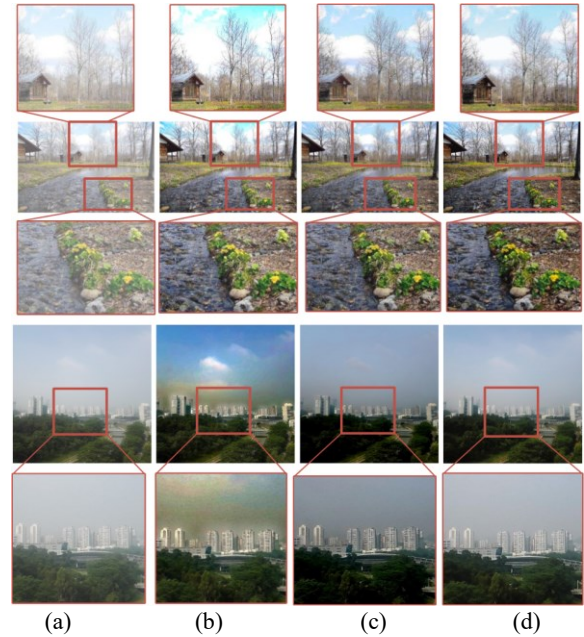
Fig .6 Haze removal results with aerial images. (a) Input hazy images. (b) Results of He's method. (c) Results of DehazeNet. (d) Results of our IASM-Net. The 2ed and 4th rows show the whole images while the 1st, 3rd and 5th show the local patch images for details comparison.

images are dehazed by using different methods, and the PSNR values of these methods are computed by their results and the ground truth images. The results are given in Table 1.

From Table 1, it is noteworthy that our method outperforms He's method and DehazeNet. As t is equal to 0.6, 5.8dB and 1.5dB gain is achieved over He's method and DehazeNet, respectively. With the increase of t, the performance of He's method improves, that of DehazeNet degrades, while our IASM-Net stay stable.

Table 1. the results of PSNR on 20 images

t \\ Methods	He's method	DehazeNet	Ours
0.6	21.6501	25.9501	**27.4507**
0.7	24.7070	26.8732	**27.3748**
0.8	26.7914	24.6418	**27.4507**

4. CONCLUSIONS

In this paper, we proposed a novel end-to-end deep convolutional neural network for single image dehazing, which is named as IASM-Net. The design of IASM-Net is inspired by the essential inverse atmospheric scattering model (IASM) to remove haze. By simulating IASM with specific three sub-model structures, IASM-Net retains the good interpretability and faster convergence and computational efficiency. To facilitate this research, a self-built aerial hazy image dataset has been established which has more than 6000 hazy aerial images. Experimental results verify that our IASM-Net performs well and achieves the state-of-the-art dehazing effect.

ACKNOWLEDGMENT

5. 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] 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, pp. 5187-5198, 2016.

[3] R. T. Tan, "Visibility in bad weather from a single image," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8.

[4] 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, pp. 2341-2353, 2011.

[5] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[6] Nair, Vinod, and Geoffrey E. Hinton. "Rectified linear units improve restricted boltzmann machines." Proceedings of the 27th international conference on machine learning (ICML-10). 2010.

[7] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding[C]//Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014: 675-678.