Underwater Image Enhancement with Encoding-Decoding Deep CNN Networks

Xin Sun, Lipeng Liu, Junyu Dong*
College of Information Science and Engineering, Ocean University of China, 266100 Qingdao, PR.China

Abstract*—Image enhancement is an important topic in the field of machine vision and image processing. Complex underwater environment poses great challenge for machine vision due to the turbid water medium. Serious scattering and absorption make the underwater images have complicated noise distribution. This paper proposes an underwater image enhancement model based on Encoding-Decoding deep CNN networks. We employ the convolution layers as encoding while deconvolution layers as decoding. The model achieves the image enhancement in an end-to-end adaptive way rather than considering the physical environment. We provide several comparison experiments with different datasets. Experiments show that it outperforms state-of-the art underwater image restoration methods in underwater image defogging, denoising and color enhancement. Finally, the model is employed to handle the underwater images with the different levels of noise and shows good performance.

***Keywords*—Encoding-decoding; Underwater image; Color restoration; Image enhancement**

I. INTRODUCTION

Underwater object detection is mainly based on sonar technology for the cases of low-resolution applications. However, it fails to meet the high accuracy requirement of underwater discovery tasks [1]. Recent years, underwater vision has attracted more and more attentions in underwater applications such as underwater robot and coastal biodiversity investigation system. It can capture the detail information of the underwater object. High-quality underwater images have important applications in research related to underwater vision, such as coral image classification [2, 3], underwater 3D reconstruction [4]. However, the complexity of noise and the variability of underwater environment are great challenges for the acquisition of high quality underwater visual images.

Serious scattering and absorption caused by suspended particles in the turbid water environment make the underwater images have complicated noise distribution. Moreover, absorption rates are different in various visible spectrums. Therefore, color degradation also occurred in the underwater images. The quality and color restoration are important problems in the research of underwater image processing. In addition, the image enhancement is one of the effective ways to solve such problems. The current researches on underwater image enhancement mainly base on the physical properties analysis including visible spectrum attenuation and scattering during transmission, and the imaging properties of degraded images. All these methods can be summarized into the following categories.

Category one: priori approaches

In the case of the values of pixels gathering in small areas, the visual effect is bad for users. At the beginning, the classic histogram equalization (HE) algorithm was introduced to solve such problem, which transfers gray scale histogram of original image from concentration distribution to approximate average distribution. Therefore, the image vision effect can be improved. However, white drift phenomenon appears because of the different color temperature. Usually, the white balance algorithm is used to solve this kind of distortion, which consists of two steps: white point detection and then white point adjustment. He et al. [5] analyzed the large number of natural images and introduced the dark channel in the natural images. Then they proposed a dark channel priori image defog algorithm. Meanwhile, background light and transmission rate are also estimated for image defogging and enhancement [6]. Although the priori methods show good performance in natural image denoising and enhancement, they did not achieve good performance for underwater image enhancement. The reason is that low light conditions of underwater environment do not satisfy the prior conditions.

Category two: scene depth estimation

Some of the image enhancement algorithms solve the visual ambiguity problem in the imaging process by calculating the depth of the scene [7, 8] and Fattal et al [9]. They firstly generate the spectral structure information of the image. Then the corresponding scene depth should be estimated to realize the image enhancement. However, the spectral structure information is difficult to generate in the underwater environment.

Category three: physical modeling

The main cause of low quality of the underwater image is physical factors, such as optical. The intuitive solution is creating a physical model by the principle of physical attenuation in the underwater imaging process. In addition, the model parameters can be estimated by the test data. Then the model can be applied to underwater image denoising enhancement [10-12]. For example, a physical model is established from two different angles to solve the scattering and color distortion problem [11]. At first the background light estimator and local adaptive filtering algorithm are employed to solve low contrast problem caused by scattering. Secondly, a new underwater imaging model can be introduced to supplement the attenuation of light, in order to solve the blue tone problem of underwater images. Such kinds of physical model-based methods have achieved good enhancement effect in the experiment. However, a specific physical model can only be used for the noise environment with the given condition. It will fail when the environment of the suspended particles changing.

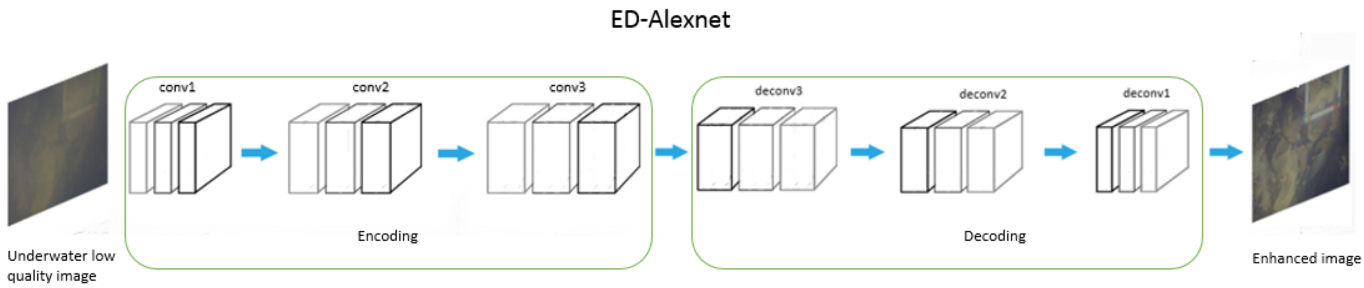


Fig. 1. The ED-Alex network structure

The fourth category: neural network + physical model

Such methods divide the image denoising enhancement into two stages [13] [14]. Firstly, the transmission of low quality image is estimated by neural network. Then the obtained transmission rate is substituted into a physical model as a parameter. Finally, the physical denoising model can enhance the images. Although this method is a kind of end-to-end neural network model, the output of the network is transmission rate rather than the final high-quality images. So these methods also fail when the environment is changing.

Over the past few years, deep learning achieves great performance in different areas, such as visual recognition [15] , natural language processing [16, 17]. And the convolution neural network (CNN) is one of the most popular methods and has many successful applications, such as image classification [18, 19] and image segmentation [20]. Inspired by the excellent performance of deep learning and the shortcomings of current underwater image enhancement algorithms, this paper proposes a encoding-decoding depth network based on convolution neural networks for underwater image enhancement.

The main contributions are summarized as follows: (1) We propose a convolution-deconvolution deep network architecture as the encoding-decoding procedure. (2) To overcome the limitation of insufficient training data, transfer learning is introduced to network training. The fine-tuning process makes the model gradually fit the requirement of the underwater image enhancement. (3) A pixel to pixel (end-to-end) network learning and image enhancement system is achieved without prior knowledge and physical models. (4) Our method shows good performance in underwater image enhancement.

The following sections of this paper are organized as follows: The second part describes the network architecture. The third part presents the details of the proposed network, and discusses the role of the proposed network structure in the process of image enhancement. The fourth part describes underwater image enhancement experiments. The fifth part concludes the paper.

II. ARCHITECTURE OF THE ENCODING-DECODING DEEP CNN NETWORKS

Alexnet has achieved very amazing results in the computer vision classification challenge [18]. Its excellent performance motives us to transfer its pre-trained model parameters to our underwater image enhancement task. The continuous convolution of the Alexnet network cannot restore the details of the low quality image. So we introduce the deconvolution layers to refine the texture after denoising. The architecture of our deep network can be regarded as an encoding-decoding symmetry Network. We name this convolution-deconvolution network as ED-Alexnet (Encoder-Deconder Alexnet). As shown in Fig. 1.

The network structure consists of two parts: convolution layers and deconvolution layers. From Fig. 1, it can be seen that the convolution and deconvolution layers are symmetrical. The symmetric deconvolution layers effectively refine the details of the feature map of the convolution layers. Our network structure is novel while imports part of the well trained layers from Alexnet [18]. We discard the full connection layer in Alexnet and only keep the first three layers. The full connection layer has achieved great results in image classification tasks; however, our goal is an end-to-end image enhancement task, which is different from the label-based classification. The full connection makes the feature mapping from two dimensions to one. It undoubtedly loses the two-dimensional information and fails for the underwater image enhancement. We also abandon the pooling layers. Pooling and unpooling layers make the edge of the object clearer in the object recognition and semantic segmentation. However pooling is unnecessary for the image enhancement and denoising tasks. The main reason is that the pooling layer does make the feature graphs denser in the multi-to-one mapping operation, while the corresponding unpooling layer will bring a lot of noise information. In the unpooling mapping, at most one value comes from the original feature map, and the remainings are artificially generated (in general, filled with the value of 0).

For the deconvolution layers of ED-Alexnet, we define the structure and the corresponding parameters be consistent with the convolution layers. The detailed configuration of the network is shown in TABLE I. The input of the ED-Alexnet model is a three-channel RGB image with size of 227*227. The size of image keeps unchanged by adding "pading" to feature maps during the process of convolution. The influence of network depth and parameters on the underwater image enhancement will be presented and discussed by experiments.

TABLE I. The ED-Alexnet configuration

ED-Alexnet		
Layer name	Kernel size	Output num
Conv1	11	96
Conv2	5	256
Conv3	3	384
Deconv1	3	384
Deconv2	5	256
Deconv3	11	3

III. ED-ALEXNET FOR UNDERWATER IMAGE ENHANCEMENT

ED-Alexnet is an end-to-end deep model for learning the mapping from low quality underwater images to high quality ones. In this section, we will describe the details of each layer in the ED-Alexnet, and discuss the role of each layer for the underwater image enhancement. Then we further suggest a transfer learning way to optimize the network parameters.

A. Strcture of each layer

1) Convolution and activation operations

Convolution layer of ED-Alexnet consists of a series of convolution filters that carry out the convolution operation on the input feature maps. The output of each convolution layer can be formulated as:

$$f_n^{l+1} = ReLU\left(\sum_m \left(f_m^l * k_{m,n}^{l+1}\right) + b_n^{l+1}\right) \quad (1)$$

where f_n^l and f_m^{l+1} are the corresponding maps of the current layer l and the following layer $l+1$, k represents the size of the convolution kernel, the index (m, n) shows the mapping relation from m^{th} feature map of the current layer to n^{th} feature map of the next, b_n^{l+1} is the bias, and the $ReLU(*)$ function represents the Rectified Linear Unit. From the left image of Fig. 2, we can see that the convolution is a multi-to-one mapping operation of the feature map. The noise of images can be filtered by the mapping procedure. By adding "padding", we can keep feature size unchanged before and after convolution. Through the cascade convolution filtering and activation operation, the original low quality image is enhanced by filtering the noise.

2) Deconvolution operations

Deconvolution was first proposed for visualizing the neural networks [21]. Hong et al. [22] employed the deconvolution to the image segmentation to get dense feature map from the unpooling. For the low-quality underwater image enhancement task, we introduce deconvolution operations to recover the missing details by the convolution operation. As shown in the right of Fig. 2, the deconvolution operation is one-to-multi mapping relation opposing to the convolution operation. The one-to-multi mapping operation makes the feature map larger than before. Therefore, we deduct the edge of the feature map to keep the size unchanged.

Both the convolution and deconvolution filters can be trained from data. We employ the back propagation algorithm to update the weights of the deep model.

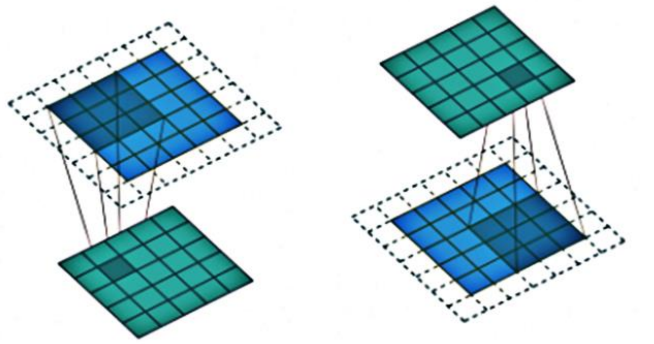


Fig. 2. Convolution and deconvolution. (Left) The upper feature map represents the original map in which the padding is shown in dotted line. The bottom is the feature map obtained by multi-to-one convolution operation. (Right) It shows the corresponding deconvolution process. The top is the original feature map, while the bottom is the feature map obtained after the one-to-multi deconvolution mapping.

B. Model optimization based on Transfer Learning

A large amount of training data is the key to train the parameters in the deep network. This makes it difficult to solve the problem in the field of underwater image processing with a small amount of dataset. Recent years, transfer learning is suggested to solve the problem of data starvation [23].Transfer learning can be summarized into two parts: instance based transfer learning and feature based transfer learning. The goal of instance-based transfer learning is to find out the suitable test data from the training data and transfer these examples to the training data. It makes the target task learn knowledge transfer quickly [23]. Feature based transfer learning is only based on the feature representation of source data. In the underwater image enhancement, the source and target data are different. So we use the feature transfer learning based on the heterogeneous space [24]. In this way, we can make the trained model more scalable.

The Alexnet model structure contains up to 60 million parameters [18]. A large data set is necessary to train such a large network. However, the number of underwater images is relatively limited for deep learning. The proposed ED-Alexnet is an extension of Alexnet, but the number of layers and parameters are larger than Alexnet. In order to solve the problem of the lack of underwater image data, the convolution layers of ED-Alexnet are initialized from the Alex model. We transfer the well-trained parameters as prior knowledge to our deep model. While the deconvolution layers are initialized with the Gauss distribution.

C. Training and testing

Our training images come from *3359*2307* underwater photos that are collected in real environment. The underwater noise scenes are simulated by adding *30*ml, *50*ml, *70*ml pure milk to 1 cubic meters of water respectively. In order to enlarge the data set, we crop patches (Stride is *66*66*) form all the pictures. We finally get *10000* training images and *2000* validation images for deep network learning. The training process of the network is an end-to-end training that maps low-quality images to high-quality images. The mean square error is used as the loss function in the training process:

$$L(\theta) = \frac{1}{n}\sum_{i=1}^{n}\|F(Y_i;\theta) - X_i\|^2 \qquad (2)$$

where n is the number of training samples, θ is the weight of the network, Y_i and X_i stand for the noise image and the clear image respectively [25]. Stochastic gradient descent method is used to minimize the loss function and standard back propagation is used to update network parameters. The Caffe toolkit [26] is used to implement the proposed network.

To show the generalization performance of the proposed network, evaluation will also be performed on publicly available underwater TURBID datasets [27].

IV. EXPERIMENT

In this section, we first analyze and discuss the influence of network depth and various configuration on the underwater image enhancement. Then, we compare the proposed method with the current popular dark channel prior method [5], histogram equalization (HE) and Fattal et al[9].

A. Evaluation method

Our underwater enhancement work has two purposes: removing noise and avoiding image distortion. To show the performance of noise removing, the peak signal-to-noise ratio (PSNR) is used as a quantitative assessment of the noise standard. We use PSNR index to evaluate image quality:

$$PSNR = 10\log_{10}\left(\frac{(2^n-1)^2}{MSE}\right) \qquad (3)$$

$$MSE = \frac{1}{H*W}\sum_{i=1}^{H}\sum_{j=1}^{W}(X(i,j) - Y(i,j))^2 \qquad (4)$$

where n is the pixel bit number. Formula (4) computes the mean square error between ground truth X and noisy image Y. H is the image's height and W is the image's width. The higher PSNR means the better denoising. We also hope to avoid the phenomenon of image distortion in the process of image enhancement. So the structural similarity index (SSIM) is suggested to evaluate the similarity between the enhanced image and the ground truth. SSIM measure the similarity between noisy image and groundtruth from three aspects: brightness, contrast, structure:

$$SSIM(X,Y) = \frac{(2\mu_x\mu_y + c_1)(2\delta_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\delta_x^2 + \delta_y^2 + c_2)} \qquad (5)$$

where μ_x and μ_y present the mean value of X and Y; δ_x and δ_y present the variance of X and Y; and δ_{xy} presents the covariance of X and Y. Therefore, the bigger the SSIM value, the smaller the distortion.

.

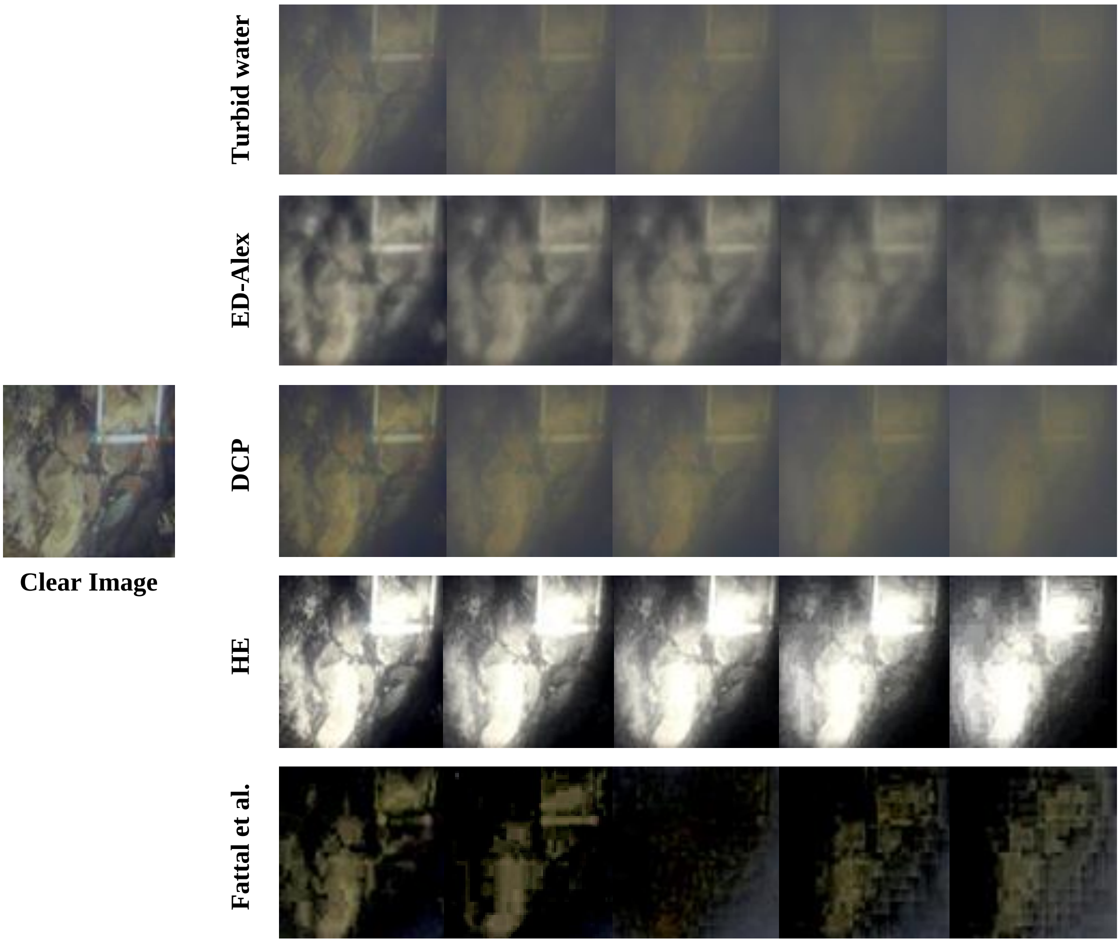


Figure. 3. Comparison output with various image enhancement methods on the same image.

B. Denoising performance

To investigate the influence of the depth of network on the denoising performance, we carry out different experiments on the tenth level degradation image from TURBID with various network configurations. As shown in TABLE II, we found that the deeper network does not mean the better performance. For example, the performance of conv3-deconv3 (3 layers convolution and 3 layers deconvolution) is better than conv5-deconv5. We believe that the padding operation in the convolution procedure brings noise. In the case of fewer network layers, the influence of padding is small. And it becomes much obvious when the network is deeper. So, for the proposed ED-Alexnet, the deeper network doesn't mean the better underwater image enhancement performance.

TABLE II. The performance of various network configuration.

	PSNR	SSIM
5con-5decon	23.5627	0.6755
3con-3decon	**25.1086**	**0.7498**
2con-2decon	24.3907	0.7202

TABLE III shows the performance of ED-Alexnet in comparison with dark channel priori (DCP), HE and Fattal et al [9] methods on underwater TURBID images dataset with different level of degradation. I10, I12, I14, I16 and I18 stand for images in TURBID dataset which degradation degree is 10, 12, 14, 16 and 18. From the experimental results, we can see that our method achieves promising results than others. Fig. 3 shows the output of various image enhancement methods applied on underwater images from TURBID dataset. These images are captured on the same area with different degradation level. These images represent four different levels of degradation with different amount of milk added. Clear image without milk is shown on the left. Pictures in the first row present noise image at different degradation levels respectively. And the second, third, fourth, and fifth rows shows the results from ED-Alex, DCP, HE and Fattal et al respectively.

TABLE III, Comparison results with various image enhancement methods

PSNR					
	Noise	ED-Alexnet	DCP	HE	Fattal
I10	22.8020	**25.1086**	21.1060	11.7740	10.3240
I12	20.6982	**24.0255**	21.0353	11.3553	9.8126
I14	20.5300	**23.5742**	20.9521	11.2682	8.7241
I16	19.4013	**21.1036**	20.2567	10.9658	9.4470
I18	19.4793	**20.5589**	19.8501	10.6441	10.5408
SSIM					
	Noise	ED-Alexnet	DCP	HE	Fattal
I10	0.7223	0.7498	**0.7582**	0.6193	0.1521
I12	0.6284	**0.6701**	0.6539	0.5345	0.4267
I14	0.6191	**0.6623**	0.6436	0.5396	0.1314
I16	0.5931	**0.6101**	0.6083	0.4846	0.1230
I18	0.5783	**0.5908**	0.5881	0.4398	0.2152

We can see that DCP method can improve the luminance but cause the color distortion, which is also shown in Table 3. HE method makes serious exposure problem. The method of Fattal et al. fails the underwater task. Our method achieves promising results. It both improves the luminance and does not make color distortion.

Figure 4 shows the effectiveness of the proposed method and comparison methods on the underwater image collected in our Lab. We can see that serious exposure and color distortion appeared with HE method. DCP method performs well on the better light condition, which can be seen from the lower right corner of the color board. However, it failed at the top left of the color board where the light condition is bad. The method of Fattal et al. failed in both scenes. In addition, we can see that our proposed method performs much better than others do.

V. CONCLUTION

Underwater imaging plays an important role in marine research. Due to the special physical properties of underwater environments, underwater images are different from common ones such as complicated noise distribution, serious scattering and absorption. In this paper, we proposed an underwater image enhancement model based on Encoding-Decoding deep CNN networks. We employ the convolution layers as encoding while deconvolution layers as decoding. The model achieves the image enhancement in an end-to-end adaptive way rather than considering the physical environment. We provide several comparison experiments with different datasets. Our method shows good performance in underwater image enhancement.

REFERENCES

[1] D. M. Kocak, F. R. Dalgleish, F. M. Caimi, and Y. Y. Schechner, "A Focus on Recent Developments and Trends in Underwater Imaging," *Marine Technology Society Journal,* vol. 42, pp. 52-67, 2008.

[2] A. S. M. Shihavuddin, N. Gracias, R. Garcia, and J. Escartin, "Automated classification and thematic mapping of bacterial mats in the North Sea," in *Oceans*, 2013, pp. 1-8.

[3] M. D. Stokes and G. B. Deane, "Automated processing of coral reef benthic images," *Limnology & Oceanography Methods,* vol. 7, p. 157–168, 2009.

[4] V. Brandou, A. G. Allais, M. Perrier, and E. Malis, "3D Reconstruction of Natural Underwater Scenes Using the Stereovision System IRIS," in *Oceans*, 2007, pp. 1-6.

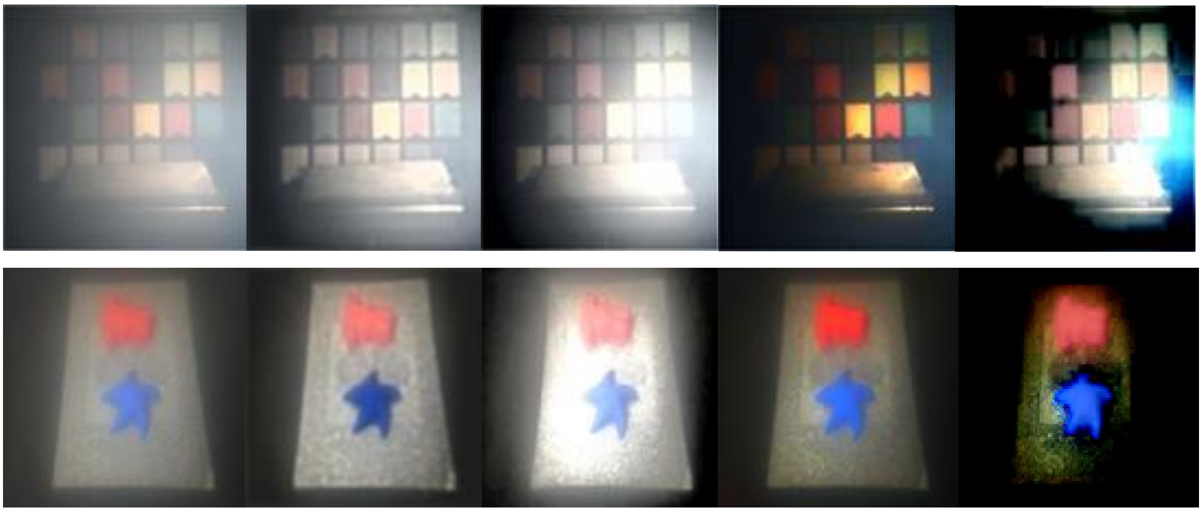
[5] K. He, J. Sun, and X. Tang, "Single Image Haze Removal Using Dark Channel Prior," *IEEE Transactions on Pattern Analysis & Machine Intelligence,* vol. 33, pp. 2341-53, 2011.

[6] R. Fattal, "Single image dehazing," vol. 27, pp. 1-9, 2008.

[7] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen*, et al.*, "Deep photo: model-based photograph enhancement and viewing," *Acm Transactions on Graphics,* vol. 27, p. 116, 2008.

[8] N. Hautiere, J. P. Tarel, and D. Aubert, "Towards Fog-Free In-Vehicle Vision Systems through Contrast Restoration,"

in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, 2007, pp. 1-8.
[9] R. Fattal, "Single image dehazing," *Acm Transactions on Graphics,* vol. 27, pp. 1-9, 2008.
[10] Y. Guo, H. Liu, Y. Chen, and W. Riaz, "Color restoration method for underwater objects based on multispectral images," in *Oceans*, 2016, pp. 1-5.
[11] H. Lu, Y. Li, L. Zhang, and S. Serikawa, "Contrast enhancement for images in turbid water," *Journal of the Optical Society of America A Optics Image Science & Vision,* vol. 32, pp. 886-93, 2015.
[12] M. Boffety and F. Galland, "Phenomenological marine snow model for optical underwater image simulation: Applications to color restoration," in *Oceans*, 2012, pp. 1-6.
[13] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An End-to-End System for Single Image Haze Removal," vol. 25, 2016.
[14] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M. H. Yang, *Single Image Dehazing via Multi-scale Convolutional Neural Networks*, 2016.
[15] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," *IEEE Transactions on Pattern Analysis & Machine Intelligence,* vol. 38, pp. 142-158, 2016.
[16] J. Andreas, M. Rohrbach, T. Darrell, and K. Dan, "Learning to Compose Neural Networks for Question Answering," pp. 1545-1554, 2016.
[17] X. Li, T. Qin, J. Yang, and T. Y. Liu, "LightRNN: Memory and Computation-Efficient Recurrent Neural Networks," 2016.
[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in neural information processing systems,* vol. 25, pp. 1097-1105, 2012.
[19] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *Computer Science,* 2014.
[20] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Transactions on Pattern Analysis & Machine Intelligence,* vol. 79, pp. 1337-1342, 2016.
[21] M. D. Zeiler, G. W. Taylor, and R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning," in *IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November*, 2011, pp. 2018-2025.
[22] H. Noh, S. Hong, and B. Han, "Learning Deconvolution Network for Semantic Segmentation," in *IEEE International Conference on Computer Vision*, 2015, pp. 1520-1528.
[23] W. Dai, Q. Yang, G. R. Xue, and Y. Yu, "Boosting for transfer learning," in *International Conference on Machine Learning*, 2007, pp. 193-200.
[24] W. Dai, Y. Chen, G. R. Xue, Q. Yang, and Y. Yu, "Translated Learning: Transfer Learning across Different Feature Spaces," in *Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December*, 2008, pp. 353-360.
[25] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE,* vol. 86, pp. 2278-2324, 1998.
[26] Jia, Yangqing, Shelhamer, Evan, Donahue, Jeff, *et al.*, "Caffe: Convolutional Architecture for Fast Feature Embedding," *Eprint Arxiv,* pp. 675-678, 2014.
[27] A. Duarte, F. Codevilla, J. D. O. Gaya, and S. S. C. Botelho, "A dataset to evaluate underwater image restoration methods," in *Oceans*, 2016, pp. 1-6.



Turbid water **ED-Alexnet** **HE** **DCP** **Fattal et al**.

Figure 4. Effectiveness of the proposed method and comparison methods in underwater image collected in our lab.