IET Image Processing

Research Article

IET Journals
The Institution of Engineering and Technology

Deep pixel-to-pixel network for underwater image enhancement and restoration

ISSN 1751-9659
Received on 31st March 2018
Revised 12th September 2018
Accepted on 20th November 2018
doi: 10.1049/iet-ipr.2018.5237
www.ietdl.org

Xin Sun[1], Lipeng Liu[1], Qiong Li[1], Junyu Dong[1] ✉, Estanislau Lima[1], Ruiying Yin[1]

[1]College of Information Science and Engineering, Ocean University of China, No. 238 Songling Road, Qingdao, People's Republic of China

✉ *E-mail: sunxin1984@ieee.org*

Abstract: Turbid underwater environment poses great difficulties for the applications of vision technologies. One of the biggest challenges is the complicated noise distribution of the underwater images due to the serious scattering and absorption. To alleviate this problem, this work proposes a deep pixel-to-pixel networks model for underwater image enhancement by designing an encoding–decoding framework. It employs the convolution layers as encoding to filter the noise, while uses deconvolution layers as decoding to recover the missing details and refine the image pixel by pixel. Moreover, skip connection is introduced in the networks model in order to avoid low-level features losing while accelerating the training process. The model achieves the image enhancement in a self-adaptive data-driven way rather than considering the physical environment. Several comparison experiments are carried out with different datasets. Results show that it outperforms the state-of-the-art image restoration methods in underwater image defogging, denoising and colour enhancement.

1 Introduction

For a long time, underwater object detection is mainly based on sonar technologies. Many successful applications have been developed, such as seafloor mapping, underwater archaeology, salvage and oil industry. However, sonar technologies fail to meet the requirements of high precision underwater tasks, such as underwater object detection [1] and 3D surface reconstruction of underwater objects [2]. Recently, machine vision has drawn more and more attention in underwater applications such as underwater robot and biodiversity investigation system. Its advantage is that it captures the detail characters of the underwater target, which are valuable to object detection [3], recognition [4, 5] and surface reconstruction. The advanced imaging devices make it possible to gain high-resolution underwater images. However, the complexity of the underwater environment brings serious noises and poses great challenges to the acquisition of high-quality underwater visual images.

From the viewpoint of visual perception, the serious noises are caused by scattering and absorption due to the suspended particles in the turbid water environment. Furthermore, the absorption rates are different for various visible spectrums in such water environment. Colour degradation commonly occurs in the captured underwater images. Therefore, low quality and colour degradation are crucial obstacles to underwater vision technologies. Image enhancement and restoration have been studied for a long time in the area of image processing. It could also be a solution to improve the underwater images. One of the classic methods of normal natural image optimisation is histogram equalisation (HE), which can increase the global contrast of the image. It spreads out the most frequent intensity values in order to improve the vision effect. Some photograph enhancement methods try to solve the visual ambiguity problem by estimating the depth of the scene [6, 7]. Better to do so, these methods usually generate the spectral structure information first which is quite difficult in the underwater environment. A most recent work on image denoising is the dark channel priori (DCP) method proposed by He *et al.* [8]. The motivation of DCP method is based on the observation that few colour channels have very less intensity in the few pixels. So the atmospheric light can be calculated in most natural images, such as non-sky images. The DCP method has drawn high attention in the normal image enhancement. However, the underwater environment does not satisfy the priori conditions. Another research work is carried out from the point of view of optical engineering. These researchers consider that the main cause of low quality of underwater image is optical and physical changes. The intuitive solution is creating a physical model by the principle of physical attenuation in the underwater imaging process [9–11]. In such works, the experiment images are collected and utilised to fit the model parameters. For example, Lu *et al.* [10] established a physical model from two different angles to solve the scattering and colour distortion problem. First a background light estimator and a local adaptive filtering algorithm are employed to solve low contrast problem caused by scattering. Then a new underwater imaging model is introduced to supplement the attenuation of light in order to solve the blue tone problem of underwater images. Based on the optical physical characteristic, Liu *et al.* [12] proposed a deep sparse non-negative matrix factorisation method and showed effective performance for underwater image enhancement. Fattal [13] also introduced a physically sound method relying only on the assumption that the transmission and surface shading are locally uncorrelated. However, a well-designed and fitted model can only be suitable for one kind of noise images within the given condition. It might fail when the underwater environments are changed. In addition, recursive filtering and contrast stretching techniques are also employed for image enhancement [14–16].

So far, some research works are trying to solve the image improvement problem from the data-driven technology. Over the past decade years, deep learning achieves great performance in different areas, such as visual detection [17] and scenes recognition [18, 19]. Recently, some works [20, 21] employed some deep learning methods on image denoising and enhancement. These works are evolved from the physical models. They first build a suitable physical model for certain underwater environment. Then the parameters, such as transmission rate, can be learned by a deep learning method. The learning procedure is a kind of end-to-end neural network training. The advantage of these methods is that the learned parameters are robust. However, it still works for special environment because the physical model is designed in conditions.

As known to all, the convolutional neural network (CNN) is the most popular method for image classification [22, 23] and segmentation [24]. The main reason is that convolution layers can filter most of the noises and capture the crucial details. In another word, the convolution operation should be suitable for denoising. The bad news is that the convolution operations smooth the image and drop some local structure information. Luckily, the deconvolution can recover the local information which is dropped

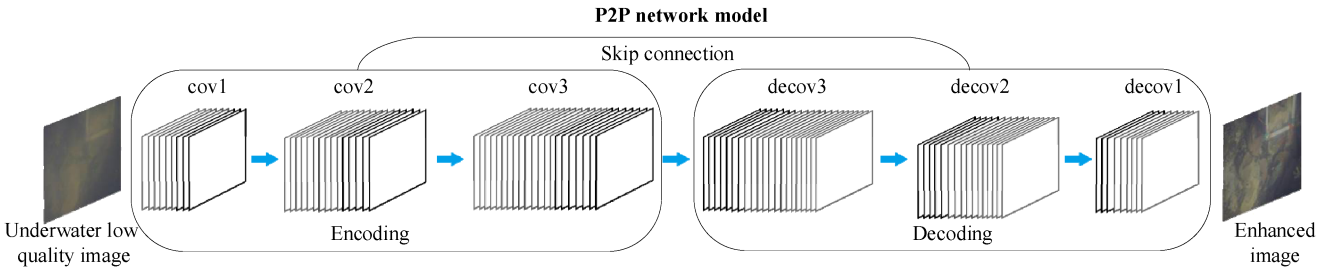

Fig. 1 *P2P network structure*

Table 1 Configuration of the deep model

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

during the convolution process. Motivated by the abilities of convolution and deconvolution, this paper proposes an encoding–decoding deep framework for pixel-to-pixel underwater image enhancement.

The following sections of this paper are organised as follows. The second section describes the framework of our proposed pixel-to-pixel deep model for image enhancement and restoration. The associated inner structure of the deep model and training methodology are illustrated in Section 3. Afterwards, Section 4 presents dataset-specific experiment details and experimental results. The conclusion of this work is presented in Section 5.

2 Pixel-to-pixel deep model

The CNNs achieved amazing results in the computer vision classification challenge [22]. The features extracted from each layer of the CNNs are able to capture all the important characteristics of an image. It motives us to employ such convolutional structures to filter the noises and retain the crucial details of the underwater image. However, the continuous convolutional operation of the CNNs cannot restore the details of the low-quality image. So the deconvolutional layers are introduced to refine the texture after denoising. The architecture of our deep network can be regarded as a symmetry encoding–decoding deep network, as shown in Fig. 1. The network architecture consists of two parts: convolutional layers and deconvolutional layers. From Fig. 1, it can be seen that convolutional and deconvolutional layers are symmetrical. The convolutional part is designed to filter the noises and keep the crucial detail characteristics. The symmetric deconvolution part is used to effectively refine the details of each feature map for the corresponding convolution layer. Moreover, skip connection is employed between the convolutional layer and deconvolutional layer. It first accelerates the training process. Detail texture information can be protected during the decoding process, because low-level features of the convolutional layers capture more fine details of the input images [25, 26].

As our goal is to achieve a data-driven based image enhancement model, the super-parameters of our network are essential. Therefore, the inner structure of the convolutional part imitates the popular Alexnet [22] network. In the later section, we will further discuss how to utilise the well-trained super-parameters from ImageNet. The convolutional part first keeps the first three layers and discards the full connection layers. The reason is that the full connection makes the feature mapping from two dimensions to one in order to input the vector to classifiers. Our purpose is to design a pixel-to-pixel network for image enhancement task which is different from the classification problem. Full connection undoubtedly loses the two-dimensional information and fails at the underwater image enhancement. Furthermore, we abandon the pooling layers. Pooling and unpooling layers can make the edge of the object clearer in the task of object recognition and semantic segmentation. However, it is unnecessary and harmful to 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 and spatial information within a receptive field is lost during pooling [27]. Meanwhile, the corresponding unpooling layer brings a lot of noise information. In the unpooling mapping, at most one value comes from the original feature map, and the remaining are artificially generated (in general, filled with the value of 0).

For the deconvolutional part, the network structure and corresponding parameters are defined to be consistent with the layers of the convolutional part. The detailed configuration of the network is shown in Table 1. The input is a three-channel RGB image with size of 227*227. The influence of network depth and parameters on the underwater image enhancement will be discussed by experiments.

3 Inner structure of the deep model

The proposed model is a deep pixel-to-pixel neural network 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 networks and discuss the role of each layer for the underwater image enhancement. Then we further suggest a transfer learning way to optimise the network parameters.

3.1 Convolution and activation operations

The convolution layer consists of a series of convolution filters that execute the convolution operation on the input feature maps. The output of each convolution layer can be formulated as

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

where f_n^l and f_m^{l+1} are the corresponding feature 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 mth feature map of the current layer to nth feature map of the next, b_n^{l+1} is the bias and the $\mathrm{ReLU}(*)$ 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. To keep the size of image unchanged, we add 'padding' to feature maps during the process of convolution operation as the dotted lines illustrated in Fig. 2. Through the cascade convolution filtering and activation operation, the original low-quality image is enhanced by filtering the noise.

In the visual recognition task, activation function is usually used to compensate the linear model in the form of adding non-linear factor. In this work, we also use the activation function (ReLU) to retain and map features in our deep model, which can remove redundant features. It might be a nice way to remove high-frequency noise. In order to verify our assumption, we carry out experiment on image enhancement with ReLU and without ReLU, respectively. From Fig. 3, we can see that the ReLU activation functions improve the performance.

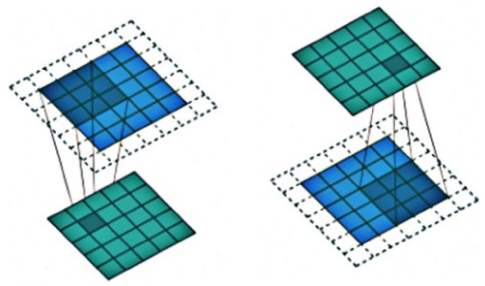

Fig. 2 *Convolution and deconvolution procedures*

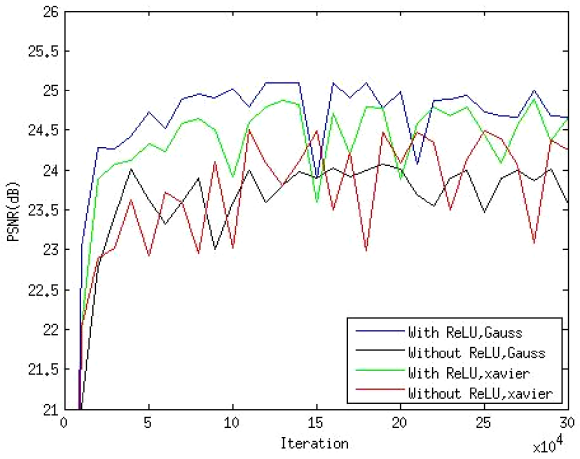

Fig. 3 *Comparison of with ReLU and without ReLU*

3.2 Deconvolution operations

Deconvolution was first proposed for visualising the neural networks [28]. It has been used to get dense feature maps from the unpooling in the work of image segmentation [29]. Here, we introduce deconvolution operations to recover the missing details during the convolution operations. As shown in the right of Fig. 2, the deconvolution operation is a one-to-multi-mapping relation opposing to the convolution operation. However, 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.

3.3 Skip connection

Skip-connection was first proposed to avoid the vanishing/exploding gradient problem [30]. It both improves the speed and accuracy of deep networks. In our encoding–decoding model, detail texture of the images might lose during the decoding process. A recent study showed that skip-connection acts like ensembles of relatively shallow networks [31]. In our work, skip connection is introduced to avoid low-level features losing. Detail texture information can be protected during the decoding process, because low-level features of the convolutional layers capture more fine details of the input images [25, 26] and contribute to the last few deconvolutional layers. Meanwhile, it can also accelerate the training process. As shown in Fig. 1, skip connection is employed between the convolutional layer and deconvolutional layer. Here skip connections perform identity mapping and the feature maps are added simply pixel by pixel.

3.4 Model optimisation based on transfer learning

Both the convolution and deconvolution filters can be trained by the data. We employ the backpropagation algorithm to update the weights of the deep model. Here a large amount of training data is the key to train the parameters in the deep network. However, underwater images are expensive to capture because of the complicated and expensive underwater scientific apparatus. Therefore, we cannot obtain sufficient training data. This makes it difficult to solve the problem in the field of underwater image processing.

Recent years, transfer learning is suggested to solve the problem of data starvation [32]. Transfer learning can be summarised in two parts: instance-based 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 [32]. Feature-based transfer learning is only based on the feature representation of source data. In the underwater image enhancement, we want to transfer the prior knowledge from a huge amount of natural images to insufficient underwater images. The source and target data are different. So we use the feature transfer learning based on the heterogeneous space [33]. In this way, we can make the trained model more scalable.

As mentioned above, the inner structure of the convolutional part imitates the Alexnet. We can transfer the well-learned knowledge, i.e. super-parameters, from the well-trained Alexnet. The Alexnet model structure contains up to 60 million parameters [22]. A large dataset ImageNet has been used to train such a large network. As the number of underwater images is relatively limited for deep learning, our proposed model cannot be trained well by the underwater images. In order to solve the problem of the lack of underwater image data, the convolution layers of our model are initialised from the Alexnet model. We transfer the well-trained parameters as prior knowledge to the first convolutional part of our deep model. The deconvolutional layers are initialised with Gauss distribution.

3.5 Training and loss function

The underwater training images come from 3359*2307 underwater photos which are collected in real environment. To simulate the different levels of underwater noise scenes, 30, 50 and 70 ml pure milk are poured into to 1 m^3 of water which, respectively, represent the low, medium and high degradation level of image. In order to enlarge the dataset, we crop patches (Stride is 66*66) from all these pictures. We finally get 10,000 training images and 2000 validation images for the deep network model training. The training procedure of the network is a pixel-to-pixel work that maps low-quality images to high-quality images. The mean square error is used as the loss function in the training process

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

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. We use Gaussian distribution to initialise deconvolution network, as shown in Fig. 3 that Gaussian initialisation strategies achieve better performance. Stochastic gradient descent method [34] is used to minimise the loss function. The most popular standard backpropagation method is used to update network parameters. The learning rate is set to be $1*10^{-7}$. For the network training, we finally get the best results when the number of iterations is 100,000 [35].

4 Experiment

In this section, we first analyse and discuss the influence of network depth and various configurations on the underwater image enhancement. Then, we compare the proposed method with the current popular dark channel prior method [8], HE and Fattal's work [13]. The Caffe toolkit [36] is used to implement the proposed network. To show the generalisation performance of the proposed network, the evaluation will also be performed on publicly available underwater TURBID datasets [37] and the underwater images captured in our VisionLab.

4.1 Evaluation method

Our underwater enhancement work has two goals: 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

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

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

where n is the pixel 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 the PSNR is, the better the result of denoising we could get. Meanwhile, we have 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 in three aspects: brightness, contrast, structure

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

where μ_x, μ_y, δ_x, δ_y, respectively, present the mean value and variance of X and Y, and δ_{xy} is the covariance of X and Y. Here, the bigger the value of SSIM is, the smaller the distortion could be achieved.

4.2 Results

To investigate the influence of the depth of the network on the denoising performance, we carry out different experiments without skip connection on tenth level degradation image from TURBID with various network configurations. As shown in Table 2, we can find that the deeper network does not mean the better performance.

Q6 **Table 2** Performance with different number of layers

	PSNR	SSIM
original image	22.8020	0.7223
2C-and-2D	24.3907	0.7202
3C-and-3D	**25.1086**	**0.7498**
5C-and-5D	23.5627	0.6755

Table 3 Performance of skip connection

	PSNR	SSIM
original image	22.8020	0.7223
without skip connection	23.5663	0.7302
with skip connection	**25.1086**	**0.7498**

For example, the performance of 3C-and-3D network (3 convolutional layers and 3 deconvolutional layers) performs better than the 5C-to-5D architecture. We believe the reason is that the padding operation in the convolution procedure brings noise. In the case of fewer network layers, the influence of padding noise is smaller. It can be much obvious when the network becomes deeper and deeper. Thus, for our encoding–decoding model, the deeper network does not mean the better underwater image enhancement performance. Moreover, we also carry out experiments to observe the performance improvement of skip connection. Here we fix the network architecture as 3C-and-3D network. The results are reported in Table 3. We can see that skip connection both improves the PSNR and SSIM.

Table 4 shows the performance of our model in comparison with HE, DSNMF, white balance (WB) and DCP for underwater TURBID images dataset with different level of degradation. I10, I12, I14, I16 and I18 stand for the degradation level of 10, 12, 14, 16 and 18 of the images in TURBID dataset. From the experimental results, we can see that our method achieves promising improvement than others.

Figs. 4–6 show the enhanced performance of different methods at the I10, I14 and I16 turbidity level of TURBID dataset. These images represent three different levels of degradation with different amount of milk added. For each figure, clear image without milk is shown in the first column. The second column represents a degraded image corresponding to turbidity levels. The rest columns, respectively, show the results of our P2PNets method, HE, DSNMF, WB and DCP.

From these results of figures, we can see that the proposed P2PNet method restore the turbid images better than others. For example, HE method makes serious exposure problem. The DSNMF and WB methods also produce noise while enhancing. It can be noticed that DCP method improves the luminance. However, it also causes serious colour distortion. Our method achieves promising results, which improves the luminance while does not make any colour distortion.

Fig. 7 shows the effectiveness of the P2PNet and the other comparison methods on the underwater image collected in our Lab. We can see that serious exposure and colour distortion appeared with HE method. DCP method performs well on the corners of image in good light condition, which can be seen from the lower right corner of the colour board. However, it failed at the top left of the colour board where the light condition is bad. The method of DSNMF failed in both scenes. The effect of WB method is not obvious. We can see that our proposed method performs much better than others do. In addition, our method achieves similar computational performance with the traditional methods, e.g. HE (0.73 s), WB (0.8 s) and P2PNet (0.81 s).

5 Conclusion

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

Table 4 Comparison results with various images enhancement methods

	Noise	P2PNets	HE	DSNMF	WB	DCP
			PSNR			
I10	22.8020	**26.3050**	11.7740	11.6988	17.3238	21.1060
I12	20.6982	**25.5723**	11.3553	12.1146	16.6753	21.0353
I14	20.5300	**24.6234**	11.2682	11.7496	16.5328	20.9521
I16	19.4013	**22.9310**	10.9658	11.8287	16.2316	20.2567
I18	19.4793	**21.6675**	10.6441	11.3534	15.9372	19.8501
			SSIM			
I10	0.7223	**0.7762**	0.6193	0.6909	0.7196	0.7582
I12	0.6284	**0.6981**	0.5345	0.5844	0.6140	0.6539
I14	0.6191	**0.6756**	0.5396	0.5700	0.6027	0.6436
I16	0.5931	**0.6312**	0.4846	0.5383	0.5730	0.6083
I18	0.5783	**0.5986**	0.4398	0.5139	0.5543	0.5881

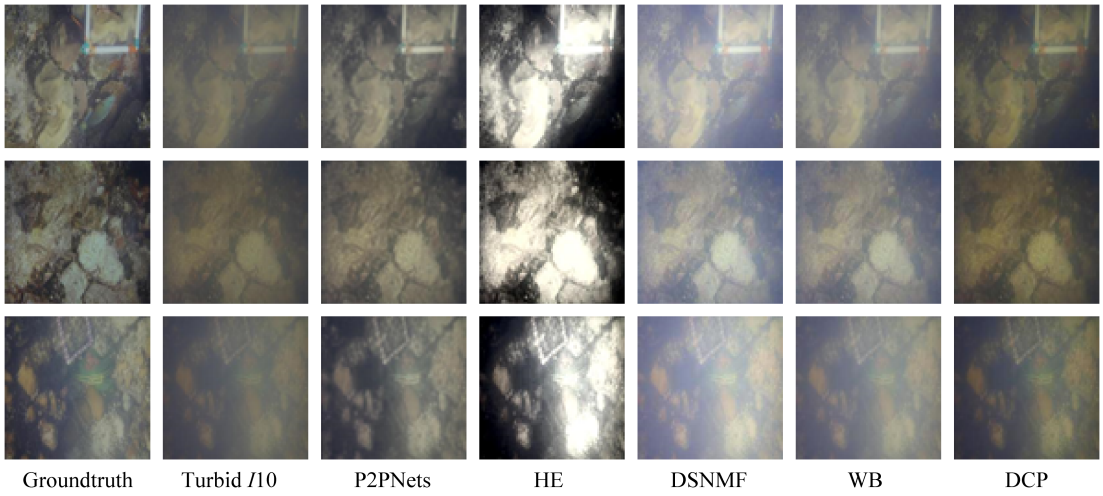

Fig. 4 *Comparison experiments on the turbidity turbid images of level-I10*

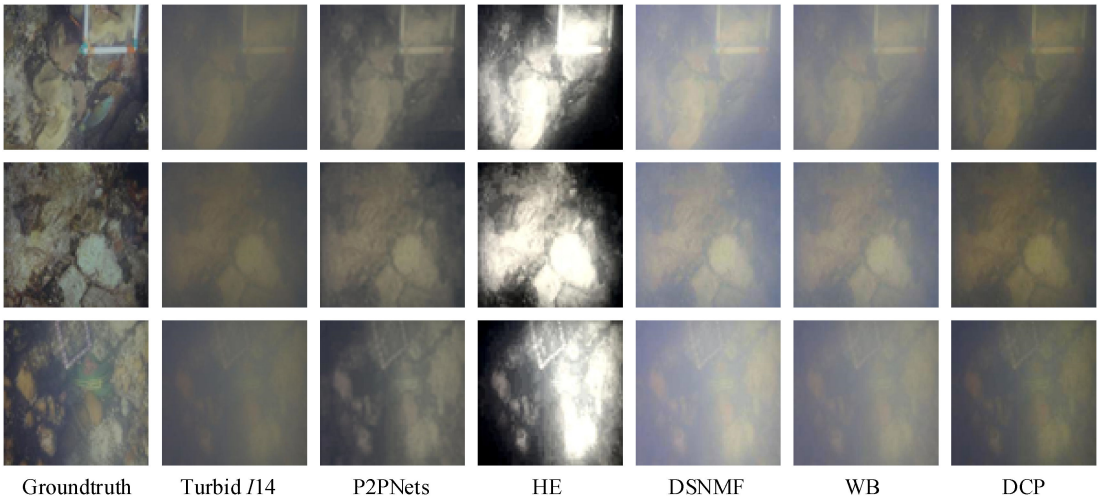

Fig. 5 *Comparison experiments on the turbidity turbid images of level-I14*

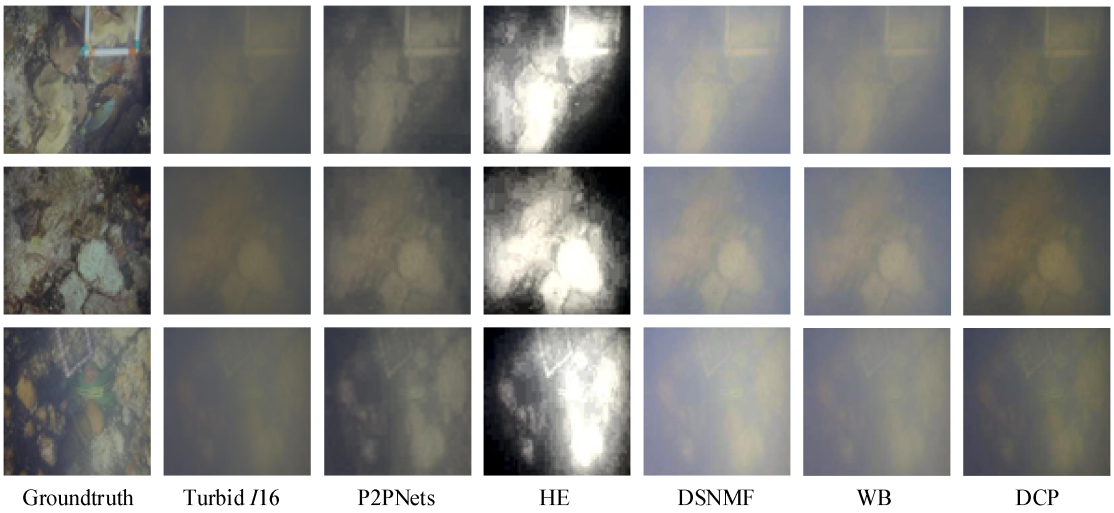

Fig. 6 *Comparison experiments on the turbidity turbid images of level-I16*

this paper, we proposed an underwater image enhancement model based on encoding–decoding deep CNN networks. We employ the convolutional layers as encoding while deconvolutional layers as decoding. The model achieves the image enhancement in a pixel-to-pixel 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.

6 Acknowledgments

This work was partially supported by the Key Research and Development Program of Shandong Province (no. GG201703140154), National Science Foundation of China (no. U1706218, 41741007, 41401174) and Applied Basic Research Programs of Qingdao (no. 18–2-2-38-jch).

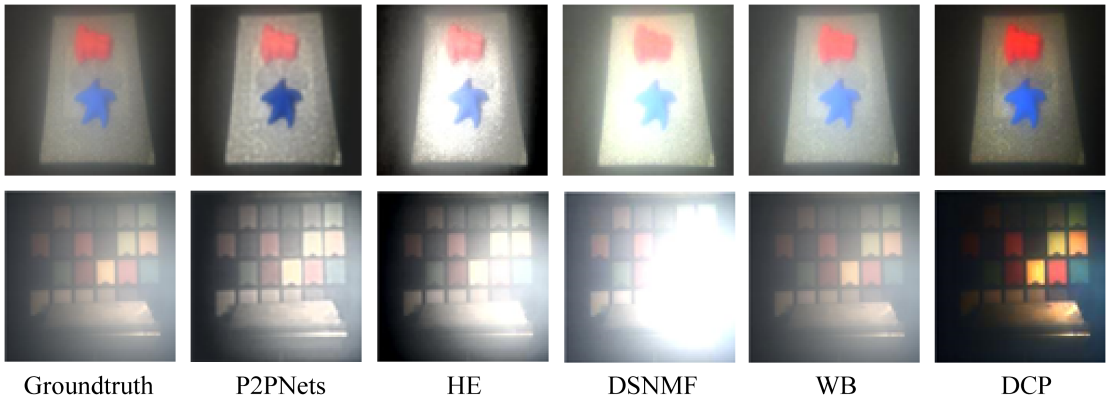

Fig. 7 *Comparison experiments on the underwater image collected in our lab*

7 References

[1] Kocak, D.M., Dalgleish, F.R., Caimi, F.M., *et al.*: 'A focus on recent developments and trends in underwater imaging', *Mar. Technol. Soc. J.*, 2008, **42**, pp. 52–67

[2] Brandou, V., Allais, A.G., Perrier, M., *et al.*: '3D reconstruction of natural underwater scenes using the stereovision system IRIS'. Oceans, 2007, pp. 1–6

[3] Sun, X., Shi, J., Liu, L., *et al.*: 'Transferring deep knowledge for object recognition in Low-quality underwater videos', *Neurocomputing*, 2018, **275**, pp. 897–908

[4] Shihavuddin, A.S.M., Gracias, N., Garcia, R., *et al.*: 'Automated classification and thematic mapping of bacterial mats in the north Sea'. Oceans, 2013, pp. 1–8

[5] Stokes, M.D., Deane, G.B.: 'Automated processing of coral reef benthic images', *Limnol. Oceanogr. Methods*, 2009, **7**, pp. 157–168

[6] Kopf, J., Neubert, B., Chen, B., *et al.*: 'Deep photo: model-based photograph enhancement and viewing', *ACM Trans. Graph.*, 2008, **27**, p. 116

[7] Hautiere, N., Tarel, J.P., Aubert, D.: 'Towards fog-free in-vehicle vision systems through contrast restoration'. IEEE Conf. on Computer Vision and Pattern Recognition, CVPR '07, 2007, pp. 1–8

[8] He, K., Sun, J., Tang, X.: 'Single image haze removal using dark channel prior', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011, **33**, pp. 2341–2353

[9] Guo, Y., Liu, H., Chen, Y., *et al.*: 'Color restoration method for underwater objects based on multispectral images'. Oceans, 2016, pp. 1–5

[10] Lu, H., Li, Y., Zhang, L., *et al.*: 'Contrast enhancement for images in turbid water', *J. Opt. Soc. Am. A Opt. Image Sci. Vis.*, 2015, **32**, pp. 886–893

[11] Boffety, M., Galland, F.: 'Phenomenological marine snow model for optical underwater image simulation: applications to color restoration'. Oceans, 2012, pp. 1–6

[12] Liu, X., Zhong, G., Liu, C., *et al.*: 'Underwater image colour constancy based on DSNMF', *IET Image Process.*, 2017, **11**, pp. 38–43

[13] Fattal, R.: 'Single image dehazing', *ACM*, 2008, **27**, pp. 1–9

[14] Albu, F., *et al.*: 'One scan shadow compensation and visual enhancement of color images'. IEEE Int. Conf. on Image Processing, 2010

[15] Singh, G., Khosla, A., Anwar, M.I.: 'Spatial domain color image enhancement based on local processing'. Int. Conf. on Signal Processing and Integrated Networks, 2016

[16] Jourlin, M, Pinoli, J.C.: 'Logarithmic image processing*: The mathematical and physical framework for the representation and processing of transmitted images', *Adv. Imaging Electron Phys.*, 2001, **115**, pp. 129–196

[17] Wang, Q., Wan, J., Yuan, Y.: 'Locality constraint distance metric learning for traffic congestion detection', *Pattern Recognit.*, 2017, **75**

Q1 [18] Wang, Q., Gao, J., Yuan, Y.: 'A joint convolutional neural networks and context transfer for street scenes labeling', *IEEE Trans. Intell. Transp. Syst.*, 2018, **19**, pp. 1457–1470

[19] Wang, Q., Wan, J., Yuan, Y.: 'Deep metric learning for crowdedness regression', *IEEE Trans. Circuits Syst. Video Technol.*, 2017, p. 1

Q2 [20] Cai, B., Xu, X., Jia, K., *et al.*: 'Dehazenet: an end-to-end system for single image haze removal', *IEEE Trans. Image Process.*, 2016, **25**, pp. 5187–5198

[21] Ren, W., Liu, S., Zhang, H., *et al.*: '*Single image dehazing via multi-scale convolutional neural networks*' (Springer International Publishing, 2016), pp. 154–169

Q4 [22] Krizhevsky, A., Sutskever, I., Hinton, G.E.: 'Imagenet classification with deep convolutional neural networks', *Adv. Neural. Inf. Process. Syst.*, 2012, **25**, pp. 1097–1105

[23] Simonyan, K., Zisserman, A.: 'Very deep convolutional networks for large-scale image recognition', *Comput. Sci.*, 2014

Q5 [24] Shelhamer, E., Long, J., Darrell, T.: 'Fully convolutional networks for semantic segmentation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2016, **79**, pp. 1337–1342

[25] Shelhamer, E., Long, J., Darrell, T.: 'Fully convolutional networks for semantic segmentation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, **39**, pp. 640–651

[26] Hariharan, B., Arbelaez, P., Girshick, R., *et al.*: 'Hypercolumns for object segmentation and fine-grained localization'. IEEE Conf. on Computer Vision and Pattern Recognition, 2015, pp. 447–456

[27] Chertov, R., Fahmy, S., Shroff, N.B.: 'High fidelity denial of service (DoS) experimentation'. Proc. of the DETER Community Workshop on Cyber Security Experimentation, June 2006

[28] Zeiler, M.D., Taylor, G.W., Fergus, R.: 'Adaptive deconvolutional networks for mid and high level feature learning'. IEEE Int. Conf. on Computer Vision, ICCV 2011, Barcelona, Spain, November, 2011, pp. 2018–2025

[29] Noh, H., Hong, S., Han, B.: 'Learning deconvolution network for semantic segmentation'. IEEE Int. Conf. on Computer Vision, 2015, pp. 1520–1528

[30] He, K., Zhang, X., Ren, S., *et al.*: 'Deep residual learning for image recognition', 2015, pp. 770–778

[31] Szegedy, C., Liu, W., Jia, Y., *et al.*: 'Going deeper with convolutions'. Proc. of the IEEE Conf. Computer Vision and Pattern Recognition, 2015, pp. 1–9

[32] Dai, W., Yang, Q., Xue, G.R., *et al.*: 'Boosting for transfer learning'. Int. Conf. on Machine Learning, 2007, pp. 193–200

[33] Dai, W., Chen, Y., Xue, G.R., *et al.*: 'Translated learning: transfer learning across different feature spaces'. Conf. on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December, 2008, pp. 353–360

[34] Lecun, Y., Bottou, L., Bengio, Y., *et al.*: 'Gradient-based learning applied to document recognition', *Proc. IEEE*, 1998, **86**, pp. 2278–2324

[35] Williams, R.J., Peng, J.: 'An efficient gradient-based algorithm for on-line training of recurrent network trajectories', *Neural Comput.*, 1990, **2**, pp. 490–501

[36] Jia, Y., Shelhamer, E., Donahue, J., *et al.*: 'Caffe: convolutional architecture for fast feature embedding', Eprint Arxiv, 2014, pp. 675–678

[37] Duarte, A., Codevilla, F., Gaya, J.D.O., *et al.*: 'A dataset to evaluate underwater image restoration methods'. Oceans, 2016, pp. 1–6

Author Queries

Q Please make sure the supplied images are correct for both online (colour) and print (black and white). If changes are required please supply corrected source files along with any other corrections needed for the paper.
Q1 Please provide page range in Ref. [17].
Q2 Please confirm the given page number in Ref. [19].
Q3 Please provide volume number in Refs. [19, 36].
Q4 Please provide location (country) of publisher in Ref. [21].
Q5 Please provide volume number, page range in Ref. [23].
Q6 Please provide the significance of bold values used in Tables 2-4