Proceedings of Machine Learning Research 95:296-311, 2018 ACML 2018

# Underwater Image Restoration Based on Convolutional Neural Network

**Yan Hu** yanh_hy@163.com
**Keyan Wang** kywang@mail.xidian.edu.cn
**Xi Zhao** xizhao24@yeah.net
**Hui Wang** huiwang_ii@yeah.net
**Yunsong Li** ysli@mail.xidian.edu.cn
*State Key Laboratory of Integrated Services Networks, Xidian Univ., Xián, China*

## Abstract

Restoring degraded underwater images is a challenging ill-posed problem. Existing priors-based approaches have limited performance in many situations due to their hand-designed features. In this paper, we propose an effective convolutional neural network (CNN) based approach for underwater image restoration, which consists of a transmission estimation network (T-network) and a global ambient light estimation network (A-network). By learning the relationship between the underwater scenes and their corresponding blue channel transmission map and global ambient light respectively, we can recover and enhance the underwater images with an underwater optical imaging model. In T-network, we use cross-layer connection and multi-scale estimation to prevent halo artifacts and to preserve edge features. Moreover, we develop a new underwater image synthetic method for training, which can simulate underwater images captured in various underwater environments. Experimental results of synthetic and real images demonstrate that our restored underwater images exhibits more natural color correction and better visibility improvement against these state-of-the-art methods.

**Keywords:** underwater image restoration, convolutional neural network, transmission, color correction, contrast enhancement

## 1. Introduction

Underwater imaging is widely used in scientific research and technology such as marine biology and archaeology. Generally, captured underwater images are degraded by scattering and absorption. Scattering means a change of direction of light after collision with suspended particles, which causes the blurring and low contrast of images. Absorption means light absorbed by suspended particles which depends on the wavelength of each light beam. The light with shorter wavelength (i.e., green and blue light) travels longer in water. As a result, underwater images generally have predominantly green-blue hue. Contrast loss and color deviation are main consequences of underwater degradation processes, which bring difficulties to further processing. Hence, there is great significance to restore degraded underwater images.

Many methods are proposed to enhance this special degraded images, which can be classified into two categories. The first kind is image enhancement based method, such as white balance based methods Henke et al. (2013), color correction methods( Iqbal et al.

(2010), Zhang et al. (2017b)), retinex-based methods ( Fu et al. (2015), Zhang et al. (2017a))and fusion-based methods( Ancuti et al. (2011) Ancuti et al. (2018)). This kind of methods are not based on the physical degradation principle. Since they ignore the correlation between image degradation and depth, these kind of methods usually enhance the visual effect of the images, but cannot correctly reflect the true color features of scene.

The other kind of methods is image restoration based on a physical model. The atmospheric degradation model Chiang and Chen (2012) is widely used in underwater image restoration. According to this model, two important parameters, i.e., *transmission map* and *global ambient light*, need to be estimated for restoring the underwater images. Numerous works on parameters estimation have emerged in recent years. These methods can be divided into two categories, i.e., *prior driven models* and *data driven networks*.

In past years, some prior information has been proposed to estimate the transmission. Chiang and Chen (2012) restored an underwater image by the dark channel prior (DCP) He et al. (2011). However, red channel rapidly loses intensity in underwater environment, which leads to the failure of the DCP. Therefore, various modifications of the DCP were proposed for underwater circumstance. For example, Li et al. (2016b) ignored the red channel when calculating the dark channel; Galdran et al. (2015) proposed Red Channel Prior which can be regarded as a variant of the DCP. Although the accuracy of their estimated transmission map is improved, it is still not enough due to the decrease of the reliability of the prior. Besides, considering that the red color channel attenuates much faster than the green and blue ones, Carlevaris-Bianco et al. (2010) proposed a new prior which estimates the depth of the scene by the aid of attenuation difference. Li and Guo (2015) also used this prior. However, relying on color information, the prior underestimates the transmission of objects with green or blue color.

Theoretically, as the depth in the scene increases, the transmission decreases to zero gradually and the ambient light makes more significant contribution. So, image pixels with maximum depth are always used to estimate ambient light as reference pixels. For selecting these reference pixels accurately, two main rules are proposed. Chiang and Chen (2012) considered that ambient light can be assumed to be the pixel with the highest brightness value in an image. Galdran et al. (2015) and Li and Guo (2015) considered that red channel intensity of reference pixels is much lower than the other two channels. However, both two rules select reference pixels according to color information. Some objects with same color feature may interfere with the selection process, such as objects are blue-green colors. Moreover, sometimes there is no ideal reference pixels in an image. For example, there is no pixel with deep depth if an image photographed with a downward angle.

The above-mentioned conventional prior driven methods heavily rely on their priors. They may make large estimation error when their assumptions of priors and rules are not valid on specific data. To overcome this limitation, deep learning technology is used to estimate the unkown parameters. Recently, several Convolutional Neural Networks (CNNs) Cai et al. (2016), Ren et al. (2016), and Zhao et al. (2017) have been applied to estimate transmission. These deep learning models are trained with synthetic training set to regress transmission and obtain more refined restorations than conventional methods. However, due to assumption that three channels have same transmission, these dehazing methods only solve the influence of scattering. They improve image contrast, but cannot correct color cast of underwater images. Shin et al. (2016) proposed a method for estimating ambient light and local transmission of underwater images using same network architecture. Shin′s method is also used to dehaze the underwater images, moreover, they proposed a new method to estimate the global ambient light of the underwater images based on CNN.

The performances of these data driven learning-based methods are tightly depended on the quality of the training data. Cai et al. (2016), Ren et al. (2016), Zhao et al. (2017) synthetic images without color cast. Shin et al. (2016) generates a large number of patches with different color cast for training the network, which contains all kinds of color cast. In fact, due to complex underwater environment, the color cast of underwater images only contains various tones of blue or green. So, now, an effective underwater images synthetic method is lacked.

Our purpose in this paper is to explore underwater image restoration techniques with CNN. The contributions of this work are summarized as follows: 1) We propose a new underwater restoration algorithm using CNN which improves the image contrast and color cast. A new network for estimating transmission which can preserve fine spatial structures and edges features is proposed. And we also introduce a robust global ambient light estimation method based on CNN. 2) To improve the performance of the networks, we design a new underwater images synthetic method which can simulate underwater images captured in various underwater environments.

## 2. Optical Model

Following the previous research Chiang and Chen (2012), the simplified underwater optical imaging model can be described as:

$$I_c(x) = J_c(x)t_c(x) + A_c(1 - t_c(x)) \tag{1}$$

where $x$ denotes a pixel in the underwater image, $I_c(x)$ is the image captured by the camera, $J_c(x)$ is the scene radiance, $A_c$ is global ambient light, $t_c(x)$ is the transmission map which represents the residual energy ratio of the scene radiance reaching the camera. According to Schechner and Karpel (2004), $t_c(x)$ can be expressed as:

$$t_c(x) = e^{-\eta_c d(x)}, \quad c\epsilon\{r, g, b\} \tag{2}$$

where $d(x)$ is the object-camera distance, $\eta$ is attenuation coefficient, it is a sum of the absorption coefficient $\alpha$ and the scattering coefficient $\beta$, so $\eta = \alpha + \beta$.

In addition, after deriving formulas, Li et al. (2016a) found that the ratios of the total attenuation coefficients between different color channels in water can be expressed as:

$$\begin{gathered} \frac{\eta_r}{\eta_b} = \frac{(-0.00113\lambda_r + 1.62517)A_b}{(-0.00113\lambda_b + 1.62517)A_r} \\ \frac{\eta_g}{\eta_b} = \frac{(-0.00113\lambda_g + 1.62517)A_b}{(-0.00113\lambda_b + 1.62517)A_g} \end{gathered} \tag{3}$$

where $\frac{\eta_r}{\eta_b}$ and $\frac{\eta_g}{\eta_b}$ are the red-blue and green-blue total attenuation coefficient ratios, respectively. And $\lambda_c$ is the wavelength of different color channels, $\lambda_r$, $\lambda_g$, $\lambda_b$ are respectively 620nm, 540nm, and 450nm in general. Thus, the transmission maps of the green and red channels can be estimated as:

$$t_g(x) = (t_b(x))^{\frac{\eta_g}{\eta_b}}$$
$$t_r(x) = (t_b(x))^{\frac{\eta_r}{\eta_b}} \tag{4}$$

The purpose of restoring underwater image is to recover $J_c(x)$ from $I_c(x)$, thus $t_c(x)$ and $A_c$ need to be estimated first.

## 3. Proposed Algorithm

Aiming to improve the image contrast and color cast, an underwater image restoration approach based on CNN and underwater optical model is proposed in this paper. Like most of methods, the unkown parameters can be estimated respectively. The architecture of our approach is presented in Fig.1. It consists of three modules: A-network, T-network, and J-estimator. The A-network is used to estimate global ambient light. And the T-network is used to estimate blue channel transmission map of the underwater image. After that, the image is finally restored in J-estimator module.

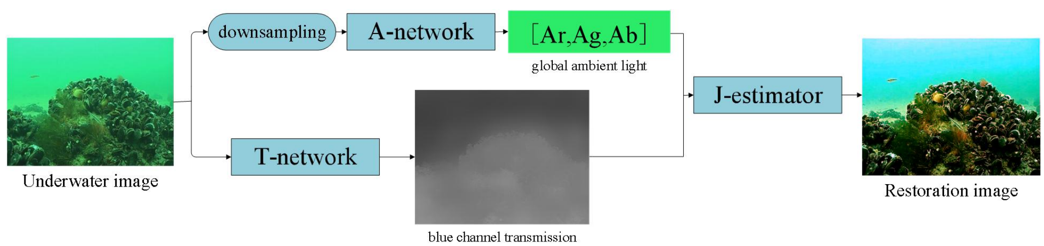



Figure 1: The architecture of our approach.

### 3.1. Model Architecture

**A-network:** As mentioned before, most of the conventional ambient light estimation methods select pixels with infinite depth to estimate ambient light. But the selection is often limited by camera angle and interfered by some special pixels. To address these issues and improve the robustness of the estimation, we proposed a new global ambient light estimation method based on CNN by learning the mapping between underwater images and their corresponding ambient light.

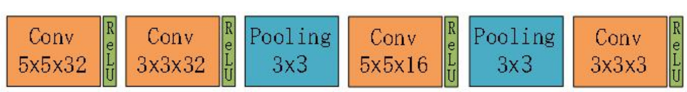



Figure 2: The architecture of A-network.

The illustration of the A-network architecture is given in Fig.2. It consists of mainly of two operations: convolution and max-pooling. The input of the A-network is an underwater image after downsampling. And the output is global ambient light, which size is the same as a pixel value. We use three convolution layers to extract features, and two max-pooling layers to overcome local sensitivity and to reduce the resolution of feature maps. The last layer is also a convolution layer for non-linear regression. In addition, we add the widely used ReLU layer after every convolution layer to avoid problems of slow convergence and local minima during the training phase.

The major difference between our algorithm and the method proposed by Shin et al. (2016) lies in the following two aspects. Firstly, we estimate three channel values of the ambient light at the same time by our A-network, instead of one by one like Shin et al. (2016), which reduces the number of parameters. Secondly, considering that the depth information is helpful to ambient light estimation, we adopt the lower resolution underwater images which generated by depth map and underwater optical model as the training samples of our network, rather than small local patches which lack of global transmission information. In this way, we can get a more accurate estimate because of a better training set. Moreover, because the image details are not important when we estimate global ambient light. So, we reduce the size of A-network training images to improve the training speed.

**T-network:** Recently, several CNN architectures were proposed on similar topics for estimating one channel transmission map. They assume that transmission of the three channels is the same. However, the assumption fails in the problem of underwater image restoration. We need to estimate three channel transmission respectively. For reducing complexity of training, we estimate the blue channel transmission by CNN and estimate the other two channel with the help of Ep(4). So, we simplify the problem to one channel transmission estimation problem. Since $I_c(x)$ is dependent on $t_c(x)$ according to Ep(1), we build a CNN model, and train the model by minimizing the reconstruction errors between its output $t_b(x)$ and the ground truth blue channel transmission map.

The U-net structure is often adopted in the network to estimate transmission map, such as the deep fully convolutional regression network(DFCRN) which proposed by Zhao et al. (2017). The U-net structure is used to expand the receptive field, filter noisy and reduce the network parameters. But the pooling layer results in a shrinkage of the feature maps and a loss of detail features. For obtaining an output with the same size as the input image, Zhao adopted an up-projection blocks. The transmission estimated by the network can preserve fine spatial structures, but can not preserve some detail features well(e.g., edges).

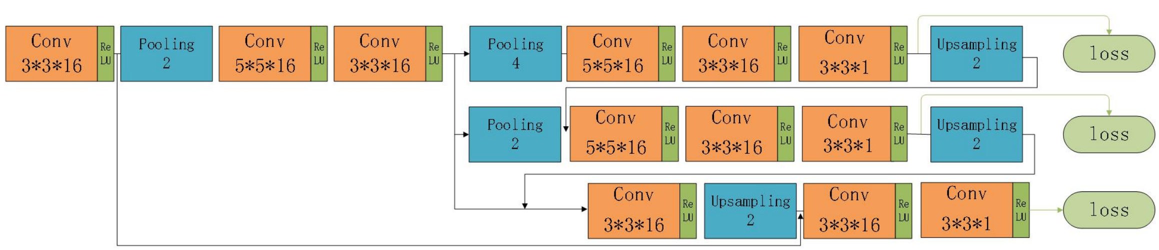



Figure 3: The architecture of T-network.

For restoring the underwater images, we design a T-network to estimate blue channel transmission and the T-network architecture is given in Fig.3. We also adopt the U-net structure. But for preserving detail features, we adopt cross-layer connection and multi-scale estimation. The connection between the first convolutional layer and the penultimate convolutional layer is used to compensate for the information loss, especially edge information. Moreover, we adopt a multi-level pyramid pooling as the second pooling layer, which helps that features from different scales are embedded in the final result Zhang and Patel (2018). Each scale size is half the size of its next scale. Inspired by Ren's fusion network Ren et al. (2018), after a multi-level pyramid pooling, we fuse multi-scale transmission map. There is an up-sampling layer after the output of every scale. And the output of small scale will be added to the next scale as a feature map. The multi-scale approach provides a convenient way to incorporate local image details over varying resolutions Ren et al. (2018). We perform estimation by varying the image resolution in a small-to-large manner to prevent halo artifacts and preserve edge features.

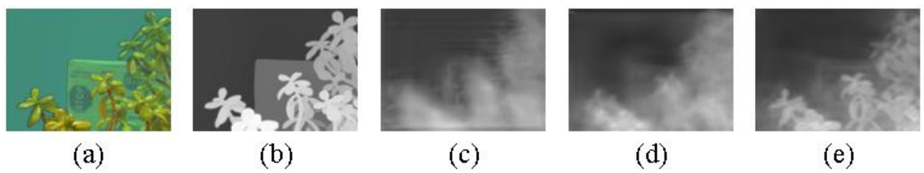



Figure 4: Transmission map estimation results using different modules.(a)underwater images (b)ground truth (c)U-net structure + up-projection blocks(DFCRN) (d) U-net structure + up-projection blocks + multi-scale estimation (e) U-net structure + multi-scale estimation + cross layer connection(our T-network)

To verify the effectiveness of our T-network, we train three different networks by the same training dataset generated by our method. Fig.4 shows transmission estimated by them. As shown in Fig.4(c), the U-net structure(DFCRN) can not preserve edge features well. After adopting multi-scale estimation method to the U-net structure, the result shows more details. And when adopt multi-scale estimation and cross-layer connection to the U-net (our T-network), the result preserves much sharper edges. The comparison experiment proves the T-network proposed in this paper can better refine the detail for objects.

**J-estimator**: After $A_c$ and $t_b(x)$ are estimated by the A-network and the T-network, we first recover $t_g(x)$ and $t_r(x)$ using Eq.(4). Then, according to Eq.(1), scene radiance can be restored as follows:

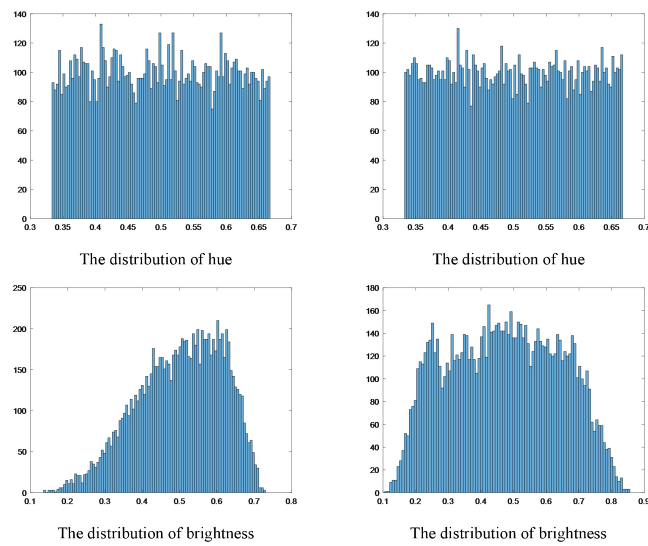
$$J_c(x) = \frac{I_c(x)-B_c}{t_c(x)} + B_c \tag{5}$$

### 3.2. Training Data and Method

**Training Data:** The quality of the training data plays an important role in the performance of network architectures. Training of CNN requires a pair of underwater images and corresponding parameters (e.g. transmission map or global ambient light). It is very difficult to obtain such a training dataset via experiments. Therefore, as many researchers have done, we synthesize training images using underwater optical model and publicly available depth datasets.

Since underwater images generally have predominantly green-blue hue, we propose a synthetic method which enables us to synthesize the underwater images with different green-blue hue color distortions. Our goal is to simulate underwater images captured in various underwater environments.

First, to simulate different underwater environments, we select indoor depth dataset and outdoor depth dataset simultaneously for constructing training dataset. We choose Middlebury Stereo dataset Scharstein and Szeliski (2003), Scharstein et al. (2014) as indoor depth dataset. Besides, we use clear outdoor images from the Internet and Liu's depth map estimation model Liu et al. (2015) to generate outdoor depth dataset. They have abundant colors in clean images and better edge-preserving ability in depth maps. Then we reduce images size of the datasets and cut them into smaller ones with a canonical size of 160×160 pixels.




(a) (b)

Figure 5: Hue and brightness distribution of global ambient light. (a)uniform distribution to generate $A_r$ (b)revised method to generate $A_r$ (note that the value of blue-green hue is 0.33-0.66)

Having 477 clean images $J(x)$ and corresponding depth maps $d(x)$, we generate random $\eta_b \epsilon [0.5, 2.5]$, $A_c \epsilon [A_r, A_g, A_b]$ and synthesize images using the physical model mentioned in section 2. Here, we use the relationship between the transmission of three channels in Li et al. (2016a) to reduce unknown parameters. Hence, we just generate blue channel attenuation coefficient. After blue channel transmission is calculated via Eq.(2), the other two channels can be calculated via Eq.(4) and the underwater image can be generated via Eq.(1). Because longer wavelengths travel shorter in water, red channel attenuates much faster and underwater images generally have predominantly green-blue hue. We assume that $A_r$ is smaller than $A_g$ and $A_b$, so we generate random $A_r \epsilon [0.1, 0.6]$, $A_g \epsilon [A_r, 1]$ and $A_b \epsilon [A_r, 1]$. The dataset generated in this manner enables the network to estimate more accurate transmission and ambient light of underwater images with different green-blue hue color cast.

Because ambient light contributes more to simulate various underwater environments. We make a statistic of the hue and brightness of ambient light generated by our method. At first, we follow uniform distribution to generate $A_c$. As shown in Fig.5(a), ambient light generated by this method contain all sorts of green-blue hue color, but have uneven distribution of brightness. For simulating various underwater environments, we modify the distribution of $A_c$ by increasing the proportion of dark background light. And the way is shown as Table.1. For comparing the effect of two $A_c$ distribution, two datasets with different $A_c$ distribution are used to train out network respectively. The restoration comparison is showed in Fig.6. After increasing the proportion of dark background light, we simulate more varied underwater environments and our networks perform better especially in dark underwater images.

**Table 1.** Method to Generate $A_c$

| $A_c$ distribution | | ratio |
|---|---|---|
| $A_r \in [0.1,0.6]$ | $A_{gb} \in [Ar,1]$ | 6 |
| $A_r \in [0.1,0.4]$ | $A_{gb} \in [Ar,0.7]$ | 2 |
| $Ar \in [0.1,0.25]$ | $A_{gb} \in [Ar,0.4]$ | 1 |
| $Ar \in [0.1,0.15]$ | $A_{gb} \in [Ar,0.3]$ | 1 |

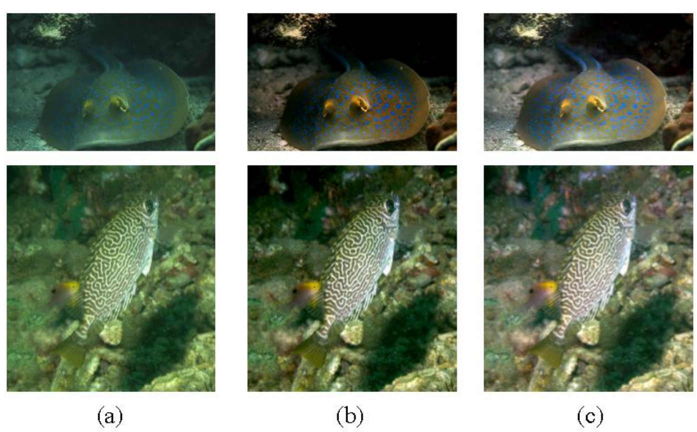


Figure 6: The results comparison of before and after modifying $A_c$ distribution. (a)raw underwater images (b)before modifying $A_c$ distribution (c)after modifying $A_c$ distribution

Finally, we generate 13780 underwater images and their corresponding transmission maps to train T-network, and 20670 underwater images and their corresponding ambient light to train A-network. Note that the training images for training A-network are resized to 49×49 pixels.

**Training Method:** Our CNN-based regression tasks optimized over pixel-wise L2-norm (Euclidean loss) and L1-norm between the predicted and ground truth images. The A-network is learned by minimizing Euclidean loss. And the T-network is learned by minimizing a weighted combination of the pixel-wise Euclidean loss and L1-norm loss, and it is defined as follows: $Loss_t = 0.7L2 + 0.3L1$. Where $L2$ is the Euclidean loss, $L1$ is the L1-norm loss. Additionally, the back-propagation algorithm and the widely-used stochastic gradient descent (SGD) algorithm are used to train our models. In T-network, We use a batch size of 8 images, the initial learning rate is 0.001 and decreased by 0.1 after every 1k iterations. The optimization is stopped at 20k iterations. Weight decay and momentum are 0.0001 and 0.9. In A-network, mini-batch, weight decay and momentum are 128, 0.005 and 0.9, respectively. The initial learning rate is 0.001 and decreased by 0.1 after every 1k iterations. The optimization is stopped at 20k iterations.

## 4. Experimental Results

We quantitatively evaluate the proposed algorithm on synthetic datasets and real-world underwater images, with comparisons to several state-of-the-art methods.

### 4.1. Evaluation on Parameters Estimation

In Fig.7, we compare our transmission map and global ambient light estimated results on synthetic underwater images with other methods. Theoretically, objects with deeper depth will have smaller transmission. And the value of image pixels whose transmission is close to 0 are approximate to the value of ambient light.

Galdran et al. (2015) selects pixels with maximum depth(or minimum transmission) in an underwater image as global ambient light. And the method estimates transmission and ambient light based on the Red Channel Prior. As shown in Fig.7(b), the method overestimates the transmission of objects with small depth. Besides, the estimation of objects in blue or green color may inaccurate due to the limitation of the prior. For improving the robustness of the method, Shin et al. (2016) and us estimate transmission and ambient light based on CNN. Based on the hypothesis that transmission of a local region is the same, Shin synthesizes a large number of small local patches as training data, and the network is used to estimate transmission of local region. Because only local information is considered, transmission estimation is influnced by color information. It brings some estimation errors as shown in Fig7.(c). Besides, without helping of depth information, ambient light estimated by Shin et al. (2016) is affected by the overall tone of the image.

We propose a synthetic method which enables us to simulate underwater images captured in various underwater environments. Our method estimates transmission and global ambient light by learning the mapping between underwater images and their corresponding transmission and ambient light respectively. As shown in Fig7., the transmission and global ambient light estimated by our method are closer to the ideal value. Moreover, Galdran and Shin use guided filter to refine the transmission after estimating, which helps to form a

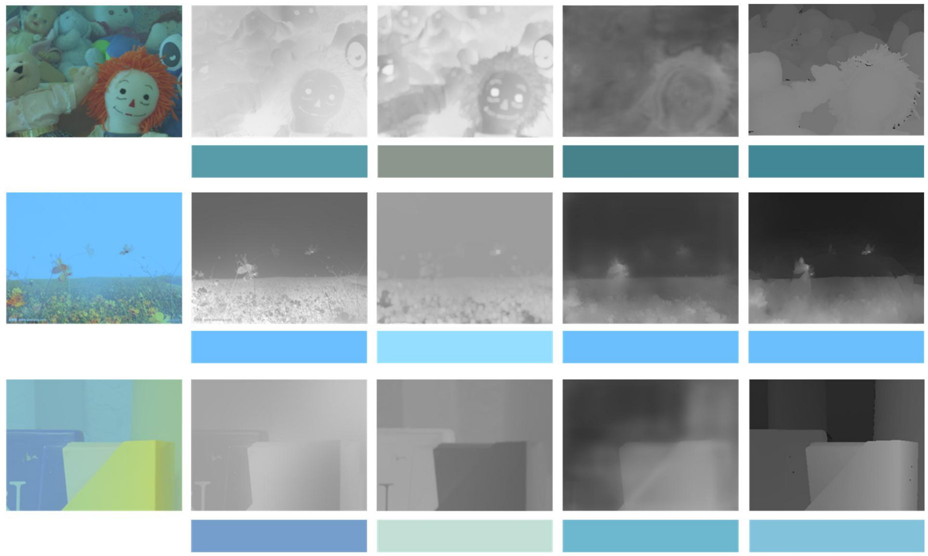



Figure 7: Visual comparison for transmission map and global ambient background light estimation results on synthetic underwater images. (a)Original images (b) Galdran et al. (2015) (c) Shin et al. (2016) (d)Proposed method (e)Ground truth images.

more accurate edges. Our T-network preserves edges information by cross layer connection and multi-scale estimation.

### 4.2. Qualitative Evaluation on Synthetic Images

We rely on several referenced metrics to evaluate similarity between ground truth images and restored images. We synthesize 30 indoor underwater images using Middlebury Stereo dataset and other 30 outdoor underwater images using Make3D dataset Saxena et al. (2009) and Liu′s model. We evaluate the proposed algorithm with Zhang, Galdran, Shin using Features Similarity (FSIM), the Peak Signal-to-Noise Ratio (PSNR) and color difference formula (CIEDE2000) metrics. FSIM is used to measure the features similarity of two images. FSIM characterizes the image quality based on human visual system by the phase congruency and the image gradient magnitude. A lager value of FSIM indicates that two images is more similar on local structure and contrast. PSNR is used to measure image distortion, which based on error between the corresponding pixels. The human visual characteristic is not taken into account. A lager value of PSNR indicates higher image quality. CIEDE2000 is used to measure the color difference of two images. A smaller value of CIEDE2000 indicate a more effective color correction. As shown in Table 1, our algorithm performs better on the three metrics. Our results are closer to ground truth images.

Fig.8 shows six input underwater images which are synthesized from the clean images with known depth maps. Zhang et al. (2017b) enhances underwater images by color correction and illumination adjustment. Without using physical model, the method improves

**Table 2.** Average Metrics on Synthetic Datasets

| Average Metrics | | Zhang | Galdran | Shin | ours |
|---|---|---|---|---|---|
| Indoor | FSIM | 0.9472 | 0.9791 | 0.9003 | **0.9822** |
| | PSNR | 16.6522 | 21.7243 | 19.6418 | **22.8509** |
| | CIEDE2000 | 15.1097 | 14.2608 | 17.6405 | **10.7806** |
| Outdoor | FSIM | 0.9177 | **0.9731** | 0.9080 | 0.9720 |
| | PSNR | 17.6601 | 21.8570 | 16.5396 | **24.3791** |
| | CIEDE2000 | 15.5413 | 10.0548 | 16.9021 | **7.0146** |

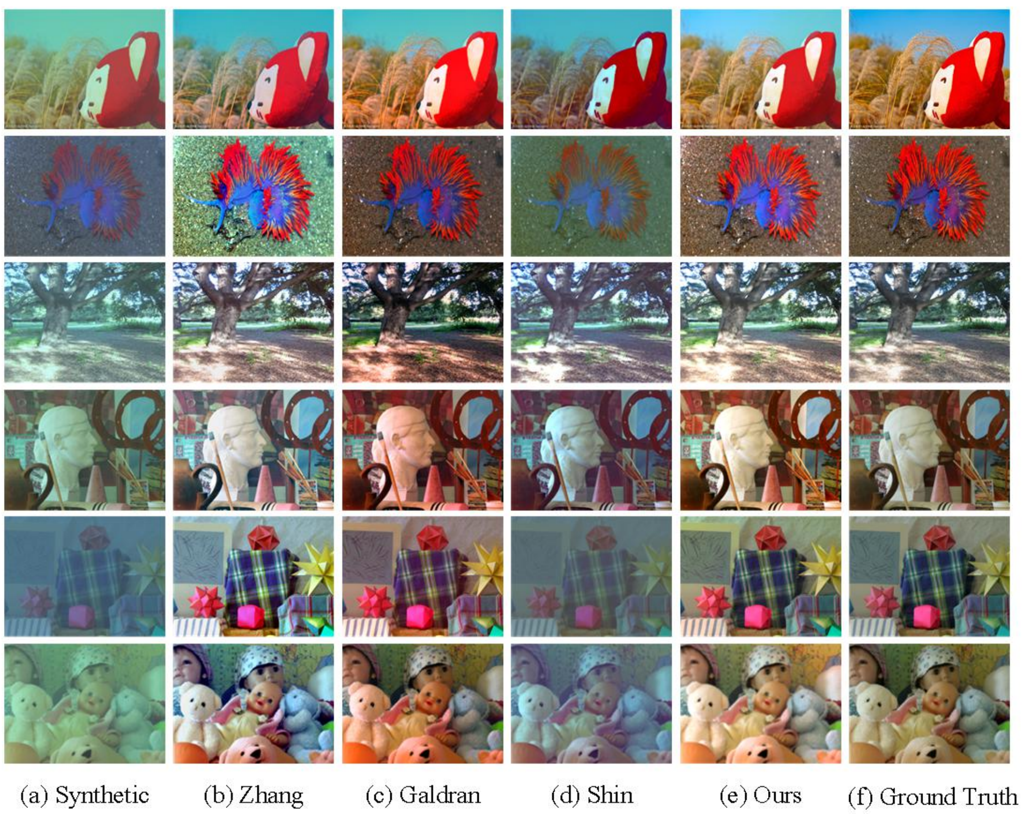



Figure 8: Visual comparison for restoration results on synthetic underwater images (a)Original images (b) Zhang et al. (2017b) (c) Galdran et al. (2015) (d) Shin et al. (2016) (e)Proposed method (f)Ground truth images.

the visual effect of the images, but the hue of images is a little bit different from the ground truth. Galdran et al. (2015) proposed the Red Channel Prior for restoring underwater images. The prior develops from the Dark Channel Prior. The results sometimes look relatively more reddish. Shin et al. (2016) estimates local transmission and global ambient light based on a CNN. It still has a low contrast due to overestimation of transmission. And the color cast of underwater images is not well corrected. In contrast, the restoration results by the proposed algorithm are close to the ground truth images. Our method improves the image contrast and corrects the image color cast.

### 4.3. Evaluation on real-world Images

Fig.9 shows a qualitative comparison with several state-of-the-art underwater restoration algorithms on real-world underwater images.

Fu et al. (2015) and Zhang et al. (2017b) restore underwater images based on image enhancement method, as shown in Fig.9(b)(c). Fu proposed a retinex-based method to enhance underwater image. The method enhance the details of the images, but the color of images is not natural enough. Zhang enhances underwater image via color correction and illumination adjustment. The method is able to increase the contrast and unveil color of the raw underwater images, but its results are not as natural as our results.

He et al. (2011), Galdran et al. (2015) and Li et al. (2016a) restore underwater images based image restoration method. They estimate unkown parameters(transmission and global ambient light) via their prior. Based on statistics on clear outdoor images, He proposed the Dark Channel Prior which thinks that the minimum value of three channels will be larger when depth increase. However, red channel rapidly loses intensity with the increase of depth in underwater environment, which leads to the failure of the DCP. Due to ignore the attenuation difference between three channels, as shown fig.9(d), the method enhances the contrast but can not correct the color. Galdran proposed a Red Channel method, which can be interpreted as a variant of the Dark Channel method. However, the restoration images sometimes seem a little bit reddish, as shown fig.9(e). Built on a minimum information loss principle, Li estimates transmission of red channel and restores image via the physical model same with us. In order to further enhance the color, Li enhances images based on histogram distribution at last. As shown fig.9(f), the method enhances the contrast and color of underwater images, but the results are not natural.

For estimating more accurate parameters, Shin et al. (2016) and us estimate transmission and global ambient light by deep learning. Shin generates local patch rather than images as training dataset. Lacking of global information, the estimation of this method is interfered by color information. To improve the performance of the networks, we design a underwater images synthetic method which can simulate underwater images captured in various underwater environments. The results of Shin on real-world images are low contrast. In contrast, our approach achieves a natural color correction and enhances visibility significantly.

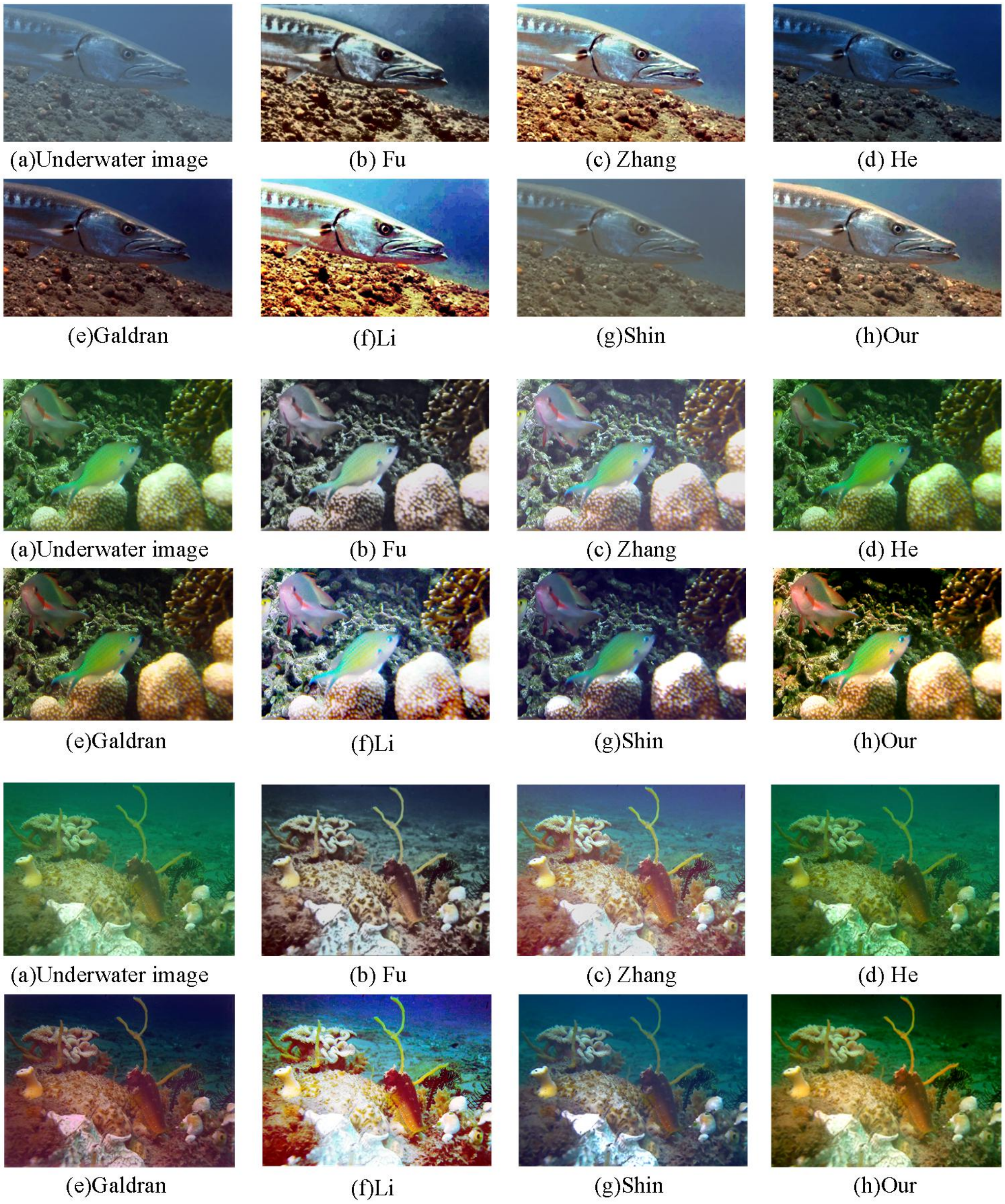



Figure 9: Visual comparison for restoration results on real underwater images (a)Underwater images (b) Fu et al. (2015) (c) Zhang et al. (2017b) (d) He et al. (2011) (e) Galdran et al. (2015) (f) Li et al. (2016a) (g) Shin et al. (2016) (h)Proposed method.

## 5. Conclusions

In this paper, we propose a novel approach to restore underwater images based on CNN. First, we propose a new underwater images synthetic method which can simulate underwater images captured in various underwater environments. It helps to improve the performance of our networks. And we design a new transmission estimation network, which is able to preserve details information well by cross-layer connection and multi-scale estimation. Besides, we also introduce a robust global ambient light estimation method based on CNN. The qualitative and quantitative evaluations show that the proposed method can effectively restore natural color and increase contrast. In the future, we will extend our approach to more challenging underwater scene, like underwater images with serious color cast and poor visibility.

## Acknowledgments

This work was jointly supported by the National Natural Science Foundation of China (No. 61301291) and the 111 Project (B08038).

## References

Codruta-Orniana Ancuti, Cosmin Ancuti, Tom Haber, and et al., editors. *Fusion-based restoration of the underwater imagespages 1557-1560*, IEEE International Conference on Image Processing, 2011.

Codruta-Orniana Ancuti, Cosmin Ancuti, Christophe-De Vleeschouwer, and et al., editors. *Color Balance and Fusion for Underwater Image Enhancementpages 379-393*, volume 27 of *IEEE Transactions on Image Processing*, 2018.

Bo-lun Cai, Xiang-min Xu, Kui Jia, and et al., editors. *DehazeNet: An End-to-End System for Single Image Haze Removal*, volume 25 of *IEEE Transactions on Image Processing, pages 5187-5198*, 2016.

Nicholas Carlevaris-Bianco, Anush Mohan, and Ryan M. Eustice, editors. *Initial results in underwater single image dehazing*, volume 27 of *IEEE Conference on OCEANS, pages 1-8*, 2010.

Jonh Y. Chiang and Ying-Ching Chen, editors. *Underwater image enhancement by wavelength compensation and dehazing*, volume 21 of *IEEE Transactions on Image Processing, pages 1756-1769*, 2012.

Xue-yang Fu, Pei-xian Zhuang, Yue Huang, and et al., editors. *A retinex-based enhancing approach for single underwater image, pages 4572-4576*, IEEE International Conference on Image Processing, 2015.

Adrian Galdran, David Pardo, Artzai Picon, and Aitor Alvarez-Gila, editors. *Automatic red-channel underwater image restoration*, volume 26 of *Journal of Visual Communication and Image Representation, pages 132-145*, 2015.

Kai-Ming He, Jian Sun, and Xiao-ou Tang, editors. *Single image haze removal using dark channel prior*, volume 33 of *IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 2341-2353*, 2011.

Birgit Henke, Matthias Vahl, Zhiliang Zhou, and et al., editors. *Removing color cast of underwater images through non-constant color constancy hypothesis*, Image Enhancement, pages 20-24, 2013.

Kashif Iqbal, Michael Odetayo, Anne James, and et al., editors. *Enhancing the low quality images using Unsupervised Colour Correction Method*, volume 25 of *IEEE International Conference on Systems Man and Cybernetics, pages 1703-1709*, 2010.

Chong-yi Li and Ji-chang Guo, editors. *Underwater image enhancement by dehazing and color correction*, volume 24 of *Journal of Electronic Imaging, no.3*, 2015.

Chong-Yi Li, Ji-Chang Guo, Run-Min Cong, and et al., editors. *Underwater Image Enhancement by Dehazing With Minimum Information Loss and Histogram Distribution Prior*, volume 25 of *IEEE Transactions on Image Processing, pages 5664-5677*, 2016a.

Chong-Yi Li, Ji-Chang Guo, Yan-Wei Pang, and et al., editors. *Single underwater image restoration by blue-green channels dehazing and red channel correction*, IEEE International Conference on Acoustics, Speech and Signal Processing, pages 1731-1735, 2016b.

Fa-yao Liu, Chun-hua Shen, Guo-sheng Lin, and et al., editors. *Deep convolutional neural fields for depth estimation from a single imagepages 5162-5170*, the IEEE Conference on Computer Vision and Pattern Recognition, 2015.

Wen-qi Ren, Si Liu, Hua Zhang, and et al., editors. *Single Image Dehazing via Multi-scale Convolutional Neural Networks*, European Conference on Computer Vision, pages 154-169, 2016.

Wen-qi Ren, Lin Ma, Jia-wei Zhang, and et al., editors. *Gated Fusion Network for Single Image Dehazing*, Computer Vision and Pattern Recognition, 2018.

Ashutosh Saxena, Min Sun, and Andrew Y. Ng, editors. *Make3D: learning 3D scene structure from a single still image*, volume 31 of *IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 824-840*, 2009.

Daniel Scharstein and Richard Szeliski, editors. *High-Accuracy Stereo Depth Maps Using Structured Light*, volume 1 of *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 195-202*, 2003.

Daniel Scharstein, Heiko Hirschmuller, York Kitajima1, and et al., editors. *High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth*, Pattern Recognition, pages 31-42, 2014.

Yoav Y. Schechner and Nir Karpel, editors. *Clear underwater vision*, volume 1 of *Computer Vision and Pattern Recognition, pages 536-543*, 2004.

Young-Sik Shin, Young-gun Cho, Gaurav Pandey, and et al., editors. *Estimation of ambient light and transmission map with common convolutional architecture*, IEEE Conference on OCEANS, pages 1-7, 2016.

He Zhang and Vishal M. Patel, editors. *Densely Connected Pyramid Dehazing Network*, Computer Vision and Pattern Recognition, 2018.

Shu Zhang, Ting Wang, Jun-yu Dong, and et al., editors. *Underwater image enhancement via extended multi-scale Retinexpages 1-9*, volume 245 of *Neurocomputing*, 2017a.

Wen-hao Zhang, Ge Li, Zhen-qiang Ying, and et al., editors. *A New Underwater Image Enhancing Method via Color Correction and Illumination Adjustment*, IEEE International Conference on Visual Communications and Image Processing, 2017b.

Xi Zhao, Ke-yan Wang, Yun-song Li, and et al., editors. *Deep Fully Convolutional Regression Networks for Single Image Haze Removal*, IEEE International Conference on Visual Communications and Image Processing, 2017.