



MTNet: A multi-task cascaded network for underwater image enhancement

Ju Moran¹ · Hu Qing¹

Received: 15 August 2022 / Revised: 21 August 2023 / Accepted: 8 September 2023 /
Published online: 25 September 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Underwater image enhancement has attracted much attention due to the rise of underwater vision research in recent years. In real-world underwater scene, the images are always with color distortion and low brightness and contrast because of light scattering and absorption, which hinders the practical applications of underwater images. To improve the quality of visual underwater scenes, in this paper, we introduced a Multi-Task Cascaded Network (MTNet) for underwater image enhancement, which contains three cascaded sub-tasks, namely color reconstruction task, contrast reconstruction task and content reconstruction task. For each task, the color loss, Hue Saturation Value (HSV) loss, Structure Similarity Index Measure (SSIM) loss and image gradient loss are employed to train MTNet in an end-to-end way. Furthermore, we design an Adaptive Fusion Module (AFM) to fuse the feature maps from different reconstruction task adaptively. To verify the performance of MTNet, we conducted the comparative experiments on both synthetic underwater images and real world underwater images. Experimental results show that our proposed method achieves better performance in both quantitative and qualitative evaluations.

Keywords Multi-task · Deep learning · Underwater image enhancement · Adaptive fusion

1 Introduction

Recently, underwater image enhancement has become a research hotspot in underwater vision [1], which has a wide variety of applications in marine archaeology [2], marine biology and marine ecological [3]. Recently, autonomous underwater vehicles have been widely employed to explore and develop the marine resources. However, the visual quality of underwater images hardly meets the expectations because the quality of the underwater images can be degraded by a lot of adverse effects, such as light scattering and wavelength dependent light absorption [4–6], which limits the performance of autonomous underwater vehicles to understand the underwater scene, as shown in Fig. 1. Therefore, it is necessary

✉ Ju Moran
jumoran@dlnu.edu.cn

¹ College of Information Science and Technology, Dalian Maritime University, Dalian 116026, Liaoning, China



Fig. 1 Samples of raw underwater images and its corresponding ground truth. Top row: raw underwater images; Bottom row: the corresponding ground truth underwater images

to develop effective methods to obtain higher quality underwater images for pleasant visual perception.

To address the above-mentioned problem, a lot of underwater image enhancement methods have been proposed and made notable progress. Traditional underwater image enhancement methods can be mainly classified into two groups: non-physical model based methods and physical model based methods. The former improves the quality of the underwater images by modifying the pixel value in the image. The latter builds a degradation model for the underwater images and obtains high quality images by estimating the parameters of the model. Recently, a variety of learning based underwater enhancement methods have been proposed and can be organized into two main categories: CNN-based methods and GAN-based methods. These learning-based approaches own powerful non-linear expression ability and generalization ability, which have achieved leading results in underwater image enhancement tasks.

Although learning based underwater image enhancement methods have achieved rapid development, there is still much room for improvement. Firstly, most of the existing CNN-based methods fuse the features directly by concatenating or residual operations such as [7, 8] which can't reflect the interdependencies of the features at different scales. Furthermore, the CNN-based underwater image enhancement methods usually apply SSIM loss, L1 loss, or perceptual loss to train the network, aiming to impose the texture, structure, content and semantics similarity on the predicted images. However,

the degraded underwater images are always with color distortion and low contrast because of light absorption and scattering, these methods haven't introduced specific color loss and contrast loss to correct the color casts and improve the contrast, which limits the enhancement quality of the degraded underwater image. Furthermore, the existing CNN-based underwater image enhancement methods have not simultaneously paid attention to the multiple factors that affect the visual perception of underwater images.

Given the above mentioned problems, in this paper, we proposed a Multi-Task Cascaded Network (MTNet) for underwater image enhancement, which contains three cascaded sub-tasks, namely color reconstruction task, contrast reconstruction task and content reconstruction task, as shown in Fig. 2. To correct the color casts, we introduced specific color loss to the color reconstruction task, which pays attention to the difference in colors between the images while eliminating texture and content comparison. To improve the contrast, we transformed the RGB color space to the HSV color space because the RGB color space can't directly reflect the contrast and brightness of the underwater image. HSV loss is used to learning the mapping function of saturation and brightness. To learn the texture and structure similarity from the ground truth image and sharpen the predicted underwater image, SSIM loss and image gradient loss are used for content reconstruction. Furthermore, we introduce an Adaptive Fusion Module (AFM) to fuse the feature maps from different reconstruction task. The comparative experiments are conducted on both synthetic underwater images and real world underwater images. Experimental results show that our proposed method achieves better performance in both qualitative and quantitative evaluations.

In summary, the main contributions of this paper can be listed as follows:

- We proposed a MTNet for underwater image enhancement, which contains three cascaded sub-tasks, namely color reconstruction task, contrast reconstruction task and content reconstruction task, aiming to reconstruct the color, the contrast and the content of the underwater image.

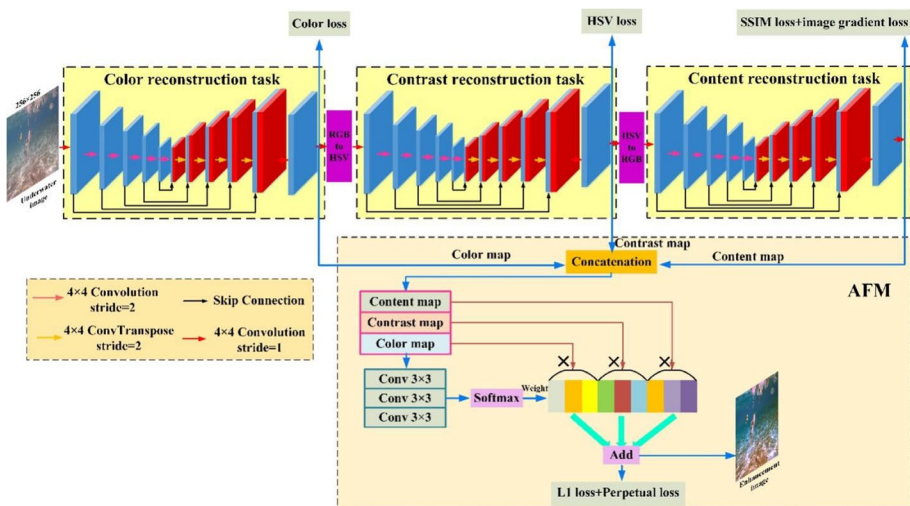


Fig. 2 The architecture of MTNet

- In MTNet, AFM is designed to fuse the feature maps from different reconstruction task.
- To correct the color casts, we introduced specific color loss to the color reconstruction task, which pays attention to the difference in colors between the images while eliminating texture and content comparison. For contrast reconstruction, HSV loss is used to learn the mapping function of saturation and brightness. SSIM loss and image gradient loss are used to learn the content similarity.
- The comparative experiments are conducted on both real world underwater images and synthetic underwater images with both the full-reference and non-reference metrics. Both the qualitative and quantitative experimental results show that our proposed method achieves better performance for underwater image enhancement.

The rest of the paper is organized as follows: Sect. 2 discusses the related works. Section 3 introduces the design of MTNet and loss function in detail. Section 4 conducts comparative experiments on both synthetic underwater images and real world underwater images and analyzes the experimental results with quantitative and qualitative evaluation. Section 5 is the conclusion of this paper.

2 Related works

Due to the importance of underwater image quality, a lot of underwater image enhancement methods have been proposed in recent years. Existing approaches can be classified into the following categories.

2.1 Non-physical model based methods

Non-physical model based methods aim to produce high quality underwater images without constructing any physical model by modifying image pixel values. The classical methods include White Balance (WB) color correction algorithm [9], gray world algorithm [10], Histogram Equalization (HE) algorithm [11] and fusion-based underwater image enhancement algorithm [12, 13] improved the contrast and saturation of underwater images both in HSV color space and RGB color space. [14, 15] reduce the number of over enhanced and under enhanced regions by a Rayleigh-stretched process based on [13, 16] proposed a two-step underwater image enhancement method which is used for image contrast enhancement and color correction. [17] applied Retinex algorithm to underwater image enhancement tasks, which consists of color correction, reflectance and illumination decomposition, and the enhancement of the reflectance and the illumination. [18] introduced an underwater image enhancement method via extended multi-scale Retinex. When non-physical model based methods are directly applied to the real underwater scene, there may be some problems such as color deviation and contrast deviation.

Image dehazing is close research area to the underwater enhancement. The image dehazing methods [19–23] are often used to enhance the quality of the underwater image. However, compared with the foggy image, the underwater image will suffer more serious distortion problems, such as reduced contrast and excessive blue-green. Therefore, the image dehazing methods fog removal method needs further improvement to achieve good results in underwater image enhancement tasks.

2.2 Physical model based methods

Physical model-based methods consider the image enhancement as an inverse problem, which constructs the underwater image degradation model and achieve image enhancement by estimating the parameters of the model. In 2006, [24] designed an adaptive filter to improve the underwater image quality based on the simplified Jaffe-McGlamery underwater model. [25] proposed to use Dark Channel Prior (DCP) and the wavelength-dependent compensation method to improve the visual perception of underwater images. [26] proposed an Underwater Dark Channel Prior (UDCP) which can estimate the medium transmission. Recently, [27] incorporated adaptive color correction to the model and proposed a Generalized Dark Channel Prior (GDGP). A Red Channel method is introduced in [28], which restores the colors associated with short wavelengths to recover the lost contrast of the underwater images. According to the relationships between inherent optical properties of water and the background color of the underwater images, [29] achieved better results for underwater image enhancement. According to the minimum information loss principle and optical properties of underwater images, [30] effectively improved the brightness and contrast of underwater images. Recently, [31] designed a physically accurate underwater image formation model improved by [6] to correct the color of underwater images. These physical model-based methods follow the simplified image formation models which achieve good performance for simple scenes, but for complex actual underwater scenes, there is still visually displeasing and unstable results.

2.3 Learning based methods

Recently, deep learning has been widely employed in the field of computer vision. A variety of learning based underwater enhancement methods have been proposed because these learning-based approaches own powerful non-linear expression ability and generalization ability. Learning based underwater enhancement methods can be organized into two main categories: GAN-based methods and CNN-based methods. [32] introduced an underwater image enhancement model, called WaterGAN. WaterGAN first generates synthetic training data from the in-air images and depth pairings. Then it uses a two-stage network to estimate the depth map and conducts color restoration. UWGAN [33] improved the WaterGAN and used Unet [34] to enhance the degraded underwater images. [35] proposed UWCNN to reconstruct the clear underwater image with MSE and SSIM loss which is trained by ten types of synthetic underwater images. [36] introduced a new real world underwater dataset and designed a novel network called WaterNet, which takes the images generated by WB, HE and Gamma Correction as input. In [37], both RGB color space and HSV color space are applied to design the underwater image enhancement network UIEC²-Net. More recently, an underwater enhancement network called Water CycleGAN [38] was proposed to improve the visual perception of the underwater images in a weakly supervised way. [39] introduced UGAN with simple structure using Generative Adversarial Network, aiming to enhance the visual perception for autonomous underwater robots. In [40], a large scale underwater dataset was presented. What is more, the author proposed a conditional generative adversarial network which is suitable for real-time visually-guided underwater robots. The above-mentioned learning based underwater image enhancement methods have not taken into account the reconstruction of the color, the content and the contrast simultaneously.

3 Our approach

In this section, we will first introduce the structure of MTNet. Then, the details of each reconstruction task and the design of AFM will be described. Finally, the design of loss function for each task will be described in detail.

3.1 Network Architecture

As shown in Fig. 2, we divided the underwater enhancement into three cascaded sub-tasks, namely color reconstruction task, contrast reconstruction task and content reconstruction task. For each sub-task, an encoder-decoder network like Unet [14] is designed to achieve feature extraction and feature map reconstruction. Residual block is taken as the basic unit of the encoder-decoder network because it is helpful for the reuse of the features from different layers. For the encoder, 4×4 convolution with stride 2 is used to down-sample the input. For the decoder, we utilize transpose convolution to up-sample the feature maps, aiming to generate the output with the same size as the input SAR image. Each convolution is followed by a Leaky-Relu activation and Batch Normalization. To achieve feature fusion, skip connections are used to concatenate the feature maps in the encoder to the ones in the decoder.

For color reconstruction task, the color sub-network takes the raw underwater image as input. To correct the color casts, we introduced specific color loss to the color reconstruction task, which pays attention to the difference in colors between the images while eliminating texture and content comparison. The output of the color reconstruction sub-network is color map.

For contrast reconstruction task, the contrast sub-network is cascaded to the color sub-network and takes the color map as input. To improve the contrast, we transformed the RGB color space to the HSV color space because the RGB color space can't directly present the brightness and contrast of the underwater image. HSV loss is used to learning the mapping function of saturation and brightness. We referred to [37] to transform the RGB color space to the HSV color space. The output of the contrast reconstruction sub-network is contrast map.

For content reconstruction task, the content sub-network is cascaded to the contrast sub-network and takes contrast map as input. To impose the texture and structure similarity on the predicted underwater image, SSIM loss is used for content reconstruction. What is more, to prevent producing blurry underwater images, image gradient loss is also introduced to the content.

As shown in Fig. 3, we design AFM to fuse the feature maps (color map, contrast map, content map) from different reconstruction task adaptively. To learn the importance of the feature maps from different sub-tasks, we first concatenate the color map, contrast map and content map in channel wise. Suppose x_{ij}^n and $weight_{ij}^n$ are the feature and the weight in the position (i, j) at channel n . Three 3×3 convolutions are used to learn the mapping from x_{ij}^n to $weight_{ij}^n$. The channel of each convolution is 64, 64 and 9. Then we utilize softmax function to compute the learnable weight for each reconstruction task. The learnable weight of each task will meet the formulas (1) and (2).

$$\sum_{n=1}^N weight_{ij}^n = 1 \quad (1)$$

where N represents the number of the feature maps in the network.

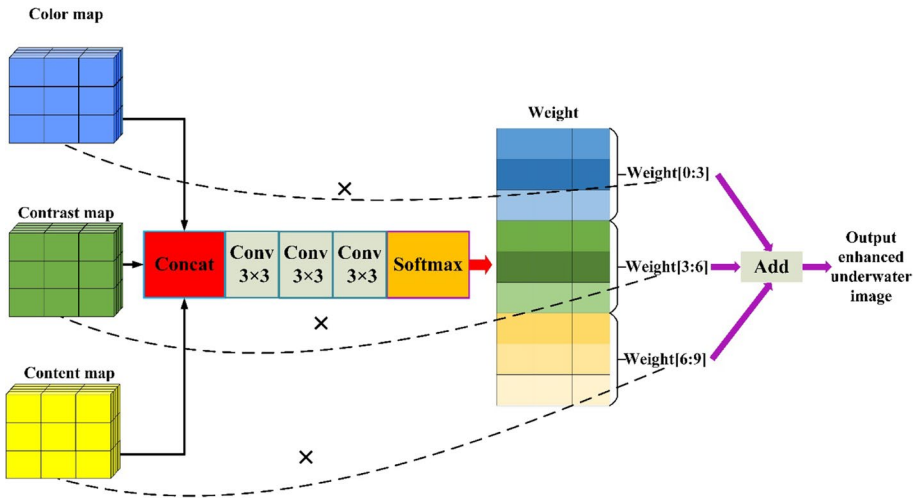


Fig. 3 The structure of AFM. reconstruction sub-task. The output of the content reconstruction sub-network is content map

$$weight_{i,j}^n \in [0, 1] \tag{2}$$

$weight_{i,j}^n$ reflects the importance of the features for each reconstruction task. Therefore, the output enhanced underwater image can be represented by (3).

$$\begin{aligned}
 Output &= weight[0 : 3] \times colormap \\
 &+ weight[3 : 6] \times contrastmap \\
 &+ weight[6 : 9] \times contentmap
 \end{aligned} \tag{3}$$

3.2 Design of multi-task loss function

The loss function of MTNet mainly consists of three sub-tasks, namely the color reconstruction task, the contrast reconstruction task and the content reconstruction task.

For the color reconstruction task, to impose the color similarity on the predicted underwater image, we applied Gaussian blur operator on the predicted and ground truth underwater image to eliminate texture and content comparison and compute the L1 loss. Color loss can be computed by:

$$\mathcal{L}_{color} = \|X(\hat{I}_{colormap}) - X(I_{colormap})\|_1 \tag{4}$$

where $X(\cdot)$ represents the blurred images computed by a Gaussian blur operator, which can be written as:

$$X(I) = \sum_{k,l} I(i+k, j+l) \cdot G(k, l) \tag{5}$$

where the Gaussian blur operator $G(k, l)$ is written as:

$$G(k, l) = A \times \exp\left(-\frac{(k - \mu_x)^2}{2\sigma_x} - \frac{(l - \mu_y)^2}{2\sigma_y}\right) \quad (6)$$

where $A = 0.053$, $\mu_{x,y} = 0$, $\sigma_{x,y} = 0$.

For the contrast reconstruction task, to further improve the contrast and saturation of the predicted underwater images, we transform RGB color space to HSV color space and compute the HSV loss as follows:

$$\mathcal{L}_{HSV} = \|\hat{S}\hat{V}\cos(\hat{H}) - SV\cos(H)\|_1 \quad (7)$$

where H , S and V are the hue, saturation and value in the HSV color space, $H \in [0, 2\pi)$, $S \in [0, 1]$, $V \in [0, 1]$. With HSV loss, the luminance, saturation and color of the underwater images can be refined through value-channel, saturation-channel and hue-channel, respectively.

For the content reconstruction task, we first apply SSIM loss to impose the texture and structure similarity on the predicted underwater image. The SSIM value is computed within a 11×11 patch for each pixel in the image as the following formula.

$$SSIM(x) = \frac{2\mu_I(x)\mu_{\hat{I}}(x) + c_1}{\mu_I^2(x) + \mu_{\hat{I}}^2(x) + c_1} \cdot \frac{2\sigma_{II}(x) + c_2}{\sigma_I^2(x) + \sigma_{\hat{I}}^2(x) + c_2} \quad (8)$$

where $\mu_I(x)$ and $\mu_{\hat{I}}(x)$ are the mean of predicted content map and the ground truth underwater image; $\sigma_I(x)$ and $\sigma_{\hat{I}}(x)$ are the standard deviation of predicted content map and the ground truth underwater image; $\sigma_{II}(x)$ represents the cross-covariance; c_1 and c_2 are set to 0.02 and 0.03, respectively.

Then, the SSIM loss can be computed by

$$\mathcal{L}_{SSIM} = 1 - \frac{1}{N} \sum_{i=1}^N SSIM(x_i) \quad (9)$$

where N indicates the number of the underwater images of each batch.

To prevent producing blurry underwater images, we also introduce image gradient loss to the content reconstruction sub-task.

$$\mathcal{L}_{GL} = \sum_{i,j} \left| |I_G(i,j) - I_G(i-1,j)| - |I_P(i,j) - I_P(i-1,j)| \right| + \left| |I_G(i,j-1) - I_G(i,j)| - |I_P(i,j-1) - I_P(i,j)| \right| \quad (10)$$

where I_P and I_G are the output content map and the ground truth underwater image.

According to the formula (3), we can get the predicted enhanced underwater image. To ensure our predicted enhanced underwater image are enough close to the real underwater images, we use L1 loss to preserve overall similarity, which can be represented as:

$$\mathcal{L}_{l1} = \|\hat{I} - I\|_1 \quad (11)$$

where \hat{I} and I are the predicted enhanced underwater image and the ground truth underwater image.

To preserve the semantic information, we also introduce the perceptual loss. The perceptual loss is defined based on VGG network, which can be computed by

$$\mathcal{L}_{per} = \frac{1}{C_j H_j W_j} \sum_{i=1}^N \|\phi_j(\hat{I}_i) - \phi_j(I_i)\| \quad (12)$$

where N represents the number of each batch; $C_j H_j W_j$ are the channel, height and width of the feature map in j th layer; ϕ_j represents the specific j th layer of VGG-19.

Therefore, the total loss can be calculated by summing the loss at each scale.

$$\mathcal{L}_{total} = \mathcal{L}_{color} + \mathcal{L}_{HSV} + \mathcal{L}_{SSIM} + \mathcal{L}_{GL} + \mathcal{L}_{l1} + \mathcal{L}_{per} \quad (13)$$

where \mathcal{L}_{color} is for the color reconstruction task, \mathcal{L}_{HSV} is for contrast reconstruction task, \mathcal{L}_{SSIM} and \mathcal{L}_{GL} are for content reconstruction task, \mathcal{L}_{l1} and \mathcal{L}_{per} are for the predicted enhanced underwater image.

4 Experiments

4.1 Experimental setup

To demonstrate the performance of MTNet, we do the quantitative and qualitative experiments with traditional underwater image enhancement methods and learning based underwater image enhancement methods on both synthetic and real world underwater images. These comparative methods include Contrast Limited Adaptive Histogram Equalization (CLAHE), White Balance, Gamma Correction, Dark channel prior, UGAN, FUnIE-GAN, UWCNN, WaterNet, UIEC²-Net. For fair comparison, we ran the source codes to generate the best results. In this section, we will introduce the comparative experiments and analyze the experimental results in detail.

Dataset To evaluate the enhanced capacity of MTNet, we conduct comparative experiments on both the synthetic underwater images and real world underwater images. We first evaluate the performance of MTNet on the synthetic dataset generated by RGB-D NYU-v2 indoor dataset. We also conduct the comparative experiments on the real world underwater images from UIBE dataset [36] which owns a diversity of scenes and underwater content.

Implementation details The experiments are implemented by an Intel i7-5930 k processor, 32 GB RAM and 1 NVIDIA GeForce GTX 3090. For training, both the synthetic underwater images based on NYU-v2 and real world underwater images from UIBE are used as input. There are 2000 images in training set. The input images are resized to 320×320 . The models are conducted on Pytorch deep learning framework and trained by stochastic gradient descent (SGD) for optimization without any augmentation. For testing, there are 90 real world underwater images and 90 synthetic underwater images in the testing set. The initial learning rate of our model is set to 0.0001, which will decrease to 0.000001 during training. We set the batch size to 24 and the total epoch to 300.

Evaluation metrics For full-reference indicators, we use the Peak Signal-to-Noise Ration (PSNR), Mean Square Error (MSE) and Structural Similarity (SSIM) to objectively evaluate the enhanced capacity of MTNet. In terms of PSNR and MSE, the higher PSNR or the lower MSE represents the recovery underwater images is more close to the ground truth underwater image. For SSIM, the higher value denotes the texture and the structure is more

Table 1 Full reference underwater image quality evaluation on synthetic underwater images

Enhancement method	MSE	SSIM	PSNR(dB)
CLAHE	4732.5086	0.6029	13.6337
White Balance	6534.3900	0.6859	13.1364
Gamma Correction	5271.5483	0.5687	12.4780
Dark channel prior	3054.2229	0.6973	16.2842
UGAN	1837.2495	0.6406	18.7508
FUnIE-GAN	2998.3801	0.6522	17.7959
UWCNN	2694.7167	0.6997	18.1389
WaterNet	1741.7989	0.7651	18.6515
UIEC ² -Net	1272.5980	0.7989	20.9121
MTNet	1255.6526	0.8125	21.8064

Table 2 Non-reference underwater image quality evaluation on synthetic underwater images

Enhancement method	UIQM	UCIQE
CLAHE	2.8401	0.3524
White Balance	2.2567	0.4909
Gamma Correction	2.2435	0.3070
Dark channel prior	2.8305	0.4306
UGAN	3.2613	0.5011
FUnIE-GAN	2.8645	0.4758
UWCNN	2.9274	0.4279
WaterNet	2.9407	0.4859
UIEC ² -Net	2.5990	0.5051
MTNet	3.0251	0.5107

close to ground truth. Meanwhile, we also employ Underwater Image Quality Measure (UIQM) and Underwater Color Image Quality Measure (UCIQE) for non-reference underwater image quality evaluation. In case of UIQM and UCIQE, higher value means better underwater enhancement performance.

4.2 Performance comparison on synthetic underwater images

To evaluate the performance achieved by the proposed MTNet, We compare MTNet with several state-of-the-art underwater image enhancement methods on the synthetic underwater testing set. The comparative experiments are conducted on the synthetic underwater testing set, which includes 90 underwater images.

Table 1 shows the quantitative comparison of different underwater enhancement methods in terms of MSE, PSNR and SSIM on the synthetic underwater testing set. The best enhancement results are in bold. It is obvious that our proposed MTNet obtains the best performance compared with both the traditional underwater enhancement methods and the deep learning based methods across all the full-reference metrics. In terms of SSIM, our proposed MTNet is 0.8943 higher than the second best enhancement method.

To further evaluate the enhancement performance of MTNet, we also employ UIQM and UCIQE for non-reference underwater image quality evaluation. Table 2 describes the average values on 90 testing underwater images. It is easy to see that our proposed

method obtains higher UIQM than other underwater enhancement methods. Furthermore, MTNet achieves the second best UCIQE value, which is larger than most of the methods. Both the full-reference and non-reference metrics prove our proposed network has better capacity for underwater enhancement.

To qualitatively evaluate the detection performance of MTNet, Fig. 4 shows the visualization of the comparative results on synthetic underwater testing set. It is obvious that the underwater images are always with color shift and low brightness and contrast because of light scattering and absorption. Most of the traditional underwater enhancement methods are not sensitive to brightness and saturation and may introduce color casts.

Especially for complex underwater environment The deep learning based underwater enhancement methods achieves relatively good enhancement performance. Our proposed MTNet can effectively hinder the color casts and improve the brightness and saturation of the underwater images even with complex underwater environment, which

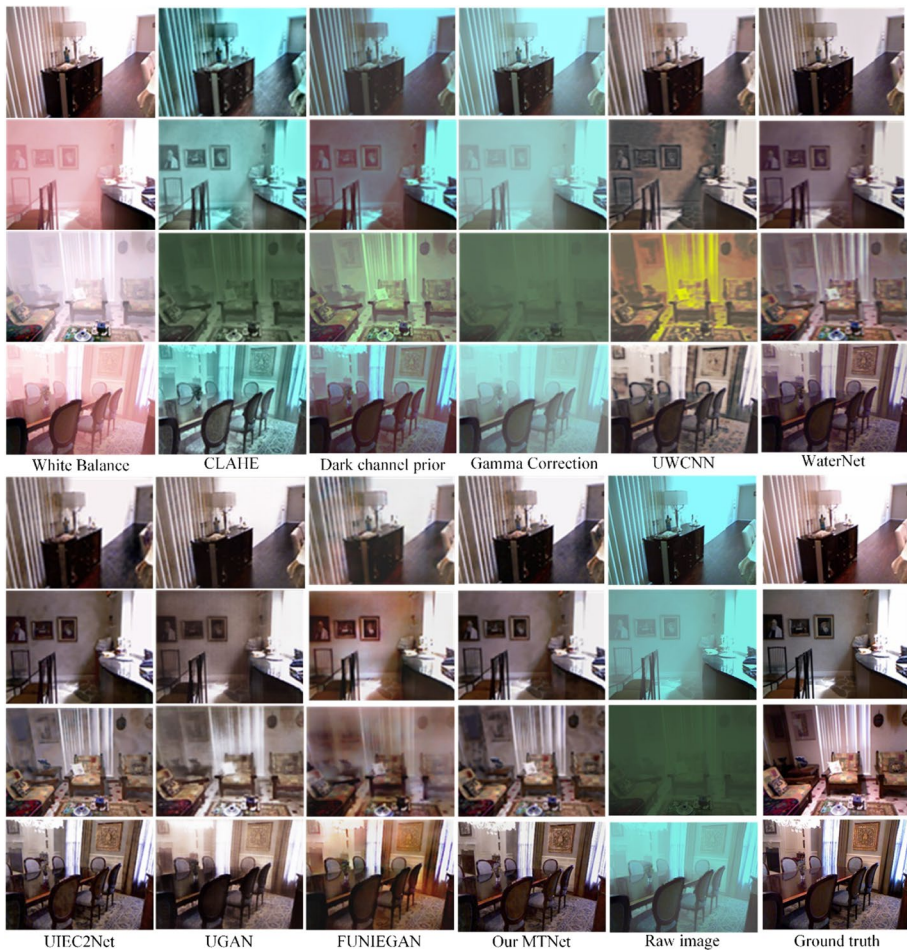


Fig. 4 Visualization of the comparative results on synthetic underwater testing set

Table 3 Full reference underwater image quality evaluation on real world underwater images

Enhancement method	MSE	SSIM	PSNR(dB)
CLAHE	1561.0980	0.8159	19.5163
White Balance	1329.1862	0.8456	23.5742
Gamma Correction	2164.0459	0.7462	17.9730
Dark channel prior	2850.7680	0.7143	16.2015
UGAN	557.1345	0.7800	23.1740
FUnIE-GAN	1171.7098	0.8029	20.4427
UWCNN	1404.7385	0.8322	19.8689
WaterNet	1235.1448	0.8360	20.0336
UIEC ² -Net	486.7439	0.9048	24.5574
MTNet	448.7683	0.9121	25.4068

Table 4 Non-reference underwater image quality evaluation on real world underwater images

Enhancement method	UIQM	UCIQE
CLAHE	2.6975	0.4223
White Balance	2.5679	0.4341
Gamma Correction	2.4647	0.3547
Dark channel prior	2.6282	0.3847
UGAN	3.2885	0.4241
FUnIE-GAN	3.2696	0.4325
UWCNN	3.0355	0.3799
WaterNet	2.9394	0.4190
UIEC ² -Net	2.9871	0.4469
MTNet	3.0402	0.4470

can produce a good and pleasant perception. The visual results in Fig. 4 agree with the non-reference metrics in Table 2.

4.3 Performance comparison on real world underwater images

To further validate the performance of MTNet, we also conduct comparative experiments on real world underwater images, which includes 90 testing images. The test results between MTNet and the state-of-the-art enhancement methods are described in Table 3. Similarity to Sect. 4.2, MSE, PSNR and SSIM are employed to evaluate the enhanced underwater images. Our proposed MTNet also achieves the best enhancement performance across all the full-reference metrics. Compared with the second best enhancement method, the PSNR and SSIM have been improved 0.0073 and 0.8494 by MTNet.

Meanwhile, we also use UCIQE and UIQM non-reference metrics to verify the performance of MTNet. As shown in Table 4, our proposed method also performs the best in terms of UIQM on real world underwater dataset. Although the UCIQE value of MTNet is not the highest, it still achieves the second best.

Similarly, to qualitatively evaluate the performance of MTNet, Fig. 5 shows the visualization of comparative results among different underwater enhancement methods on real world underwater dataset. The deep learning based enhancement methods outperform most of the traditional underwater enhancement methods. The enhanced images produced by

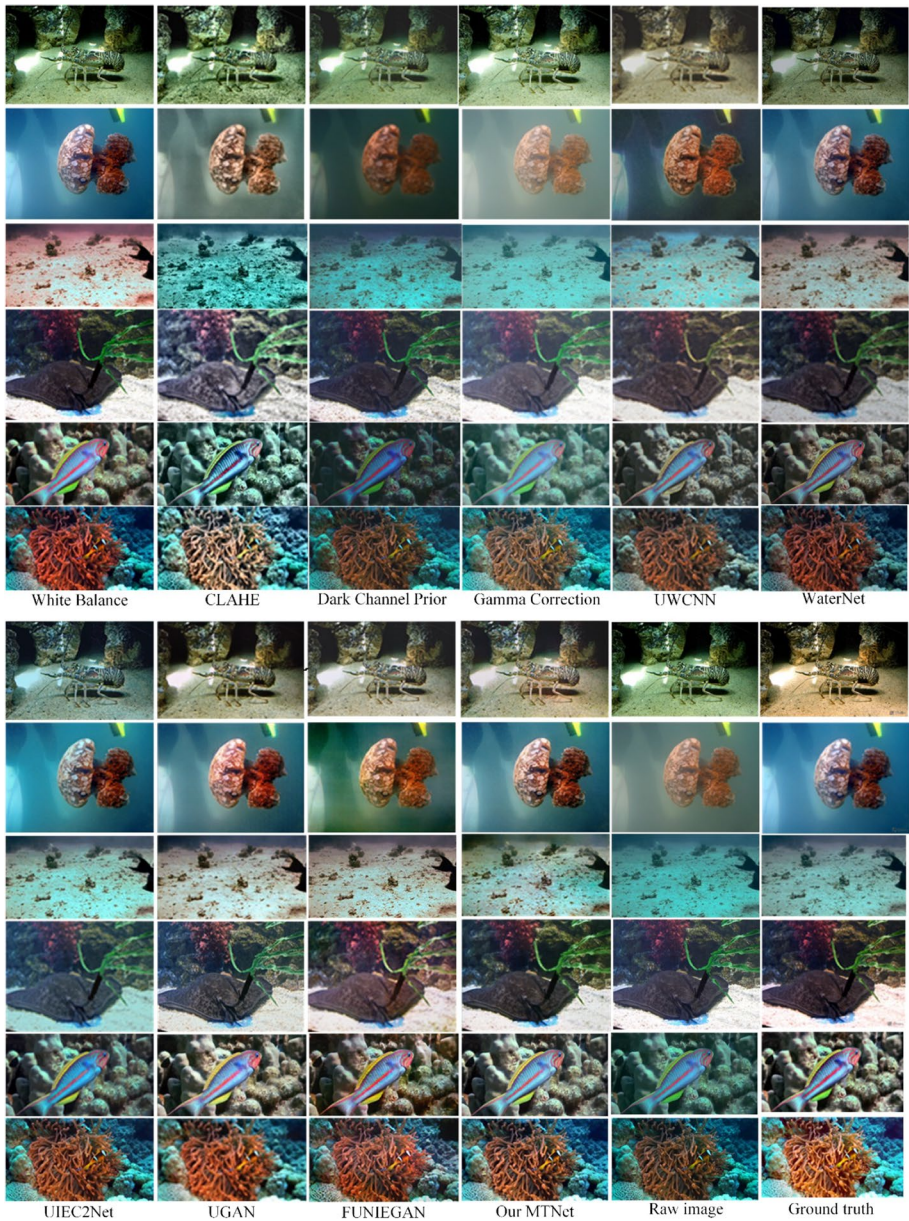


Fig. 5 Visualization of the comparative results on real world underwater testing set

MTNet are natural without introducing artificial colors and MTNet can effectively enhance the brightness and contrast, which is similar to the ground truth underwater images.

To sum up, both the comparative experiments on the synthetic underwater testing set and the real world underwater testing set demonstrate our MTNet outperforms other state-of-the-art underwater enhancement methods.

5 Conclusion

In this paper, a multi-task cascaded network is introduced to improve the visual perception for underwater image, which contains three cascaded sub-tasks, namely color reconstruction task, contrast reconstruction task and content reconstruction task. For each task, the color loss, HSV loss, SSIM loss and image gradient loss are employed to train MTNet in an end-to-end way. Furthermore, we introduce an AFM to fuse the feature maps from different reconstruction task adaptively. To verify the performance of MTNet, we conducted the comparative experiments on synthetic underwater images and real world underwater images with both the full-reference and non-reference metrics. Experimental results demonstrate that our proposed method can efficiently improve the underwater images quality and outperforms other underwater image enhancement methods in both qualitative and quantitative evaluations.

Acknowledgements The authors acknowledge National Natural Science Foundation of China (Grant no. 62201114) and the Fundamental Research Funds for the Central Universities (Grant no. 3132023233).

Data availability There are no data available for this paper.

Declarations

Conflict of Interest We declare that we have no conflict of interest.

References

1. Sheinin M, Schechner Y (2016) The next best underwater view. *Proc IEEE Int Conf Comput Vis Pattern Rec (CVPR)* 3764–3773.
2. Ludvigsen M, Sortland B, Johnsen G et al (2007) Applications of georeferenced underwater photo mosaics in marine biology and archaeology. *J Oceanogr* 20(4):140–149
3. Strachan N (1993) Recognition of fish species by colour and shape. *J Image Vis Comput* 11(1):2–10
4. Jaffe J (1990) Computer modeling and the design of optimal underwater imaging systems. *IEEE J Ocean Eng* 15(2):101–111
5. Hou W, Woods S, Jarosz E et al (2012) Optical turbulence on underwater image degradation in natural environments. *Appl Opt* 15(14):2678–2686
6. Akkaynak D, Treibitz T (2018) A revised underwater image formation model. *Proc IEEE Int Conf Comput Vis Pattern Rec (CVPR)* 6723–6732
7. Li J, Skinner KA, Eustice RM et al (2017) WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robot Autom Lett* 3(1):387–394
8. Wang N, Zhou Y, Han F, Zhu H, Zheng Y (2019) Uwgan: underwater GAN for real-world underwater color restoration and dehazing. *arXiv preprint arXiv:1912.10269*
9. Liu YC, Chan WH, Chen YQ (1995) Automatic white balance for digital still camera. *IEEE Trans Consum Electron* 41(3):460–466
10. Rizzi A, Gatta C, Marini D (2002) Color correction between gray world and white patch//Human vision and electronic imaging VII. *Int Soc Optics Photonics* 4662:367–375
11. Hummel R (1977) Image enhancement by histogram transformation. *Comput Graphics Image Process* 6(2):184–195
12. Ancuti C, Ancuti C O, Haber T et al (2012) Enhancing underwater images and videos by fusion//2012 IEEE Conference on Computer Vision and Pattern Recognition. *IEEE* 81–88
13. Iqbal K, Odetayo M, James A (2010) Enhancing the low quality images using unsupervised colour correction method. *Proc IEEE Int Conf Syst Man Cybern* 1703–1709
14. Ghani A, Isa N (2015) Underwater image quality enhancement through integrated color model with Rayleigh distribution. *Appl Soft Comput* 27:219–230
15. Ghani A, Isa N (2015) Enhancement of low quality underwater image through integrated global and local contrast correction. *Appl Soft Comput* 37:332–344
16. Fu X, Fan Z, Ling M (2017) Two-step approach for single underwater image enhancement. *Symp IEEE Intell Signal Process Commun Syst* 789–794

17. Fu X, Zhang P, Huang Y et al (2014) A retinex-based enhancing approach for single underwater image. *Proc IEEE Int Conf Image Process (ICIP)* 4572–4576
18. Zhang S, Wang T, Dong J et al (2017) Underwater image enhancement via extended multi-scale Retinex. *Neurocomput* 245(5):1–9
19. Kansal I, Kasana SS (2020) Improved color attenuation prior based image de-fogging technique. *Multimed Tools Appl* 79:12069–12091
20. Kansal I, Kasana SS (2018) Fusion based Image De-fogging using Dual Tree Complex Wavelet Transform. *Int J Wavelets Multiresolution Inf Proc* 16(06):1850054
21. Kansal I, Kasana SS (2018) Minimum preserving subsampling based fast image de-fogging. *J Mod Optics* 65(18):2103–2123
22. Kansal I, Kasana SS (2017) Weighted image de-fogging using luminance dark prior. *J Mod Optics* 64(19):2023–2034
23. Othman MK, Abdulla AA (2022) Enhanced Single Image Dehazing Technique based on HSV Color Space. *UHD J Sci Technol* 6(2):135–146
24. Trucco E, Olmos-Antillon AT (2006) Self-tuning underwater image restoration. *IEEE J Ocean Eng* 31(2):511–519
25. Chiang J, Chen Y (2012) Underwater image enhancement by wavelength compensation and dehazing. *IEEE Trans Image Process* 21(4):1756–1769
26. Drews-Jr P, Nascimento E, Botelho S et al (2016) Underwater depth estimation and image restoration based on single images. *IEEE Comput Graph Appl* 36(2):24–35
27. Peng Y, Cao T, Cosman P (2018) Generalization of the dark channel prior for single image restoration. *IEEE Trans Image Process* 27(6):2856–2868
28. Galdran A, Pardo D, Picn A (2015) Automatic red-channel underwater image restoration. *J Vis Commu Image Repr* 26:132–145
29. Zhao X, Jin T, Qu S (2015) Deriving inherent optical properties from background color and underwater image enhancement. *Ocean Eng* 94:163–172
30. Li C, Guo J, Chen S et al (2016) Underwater image restoration based on minimum information loss principle and optical properties of underwater imaging. *Proc IEEE IntConf Image Process (ICIP)* 1993–1997
31. Akkaynak D, Treibitz T (2019) Sea-thru: a method for removing water from underwater images. *Proc IEEE Int Conf Comput Vis Pattern Rec (CVPR)* 1682–1691
32. Li J, Sinner K, Eustice R et al (2018) WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robot Autom Lett* 3(1):387–394
33. Wang N, Zhou Y, Han F, Zhu H, Zheng Y (2019) Uwgan: underwater gan for real-world underwater color restoration and dehazing. *arXiv preprint arXiv:1912.10269*
34. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image .segmentation//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham 234–241.
35. Li C, Anwar S, Porikli F (2020) Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Rec* 98:107038–107049
36. Li C, Guo C, Ren W, Cong R, Hou J, Kwong S, Tao D (2019) An underwater image enhancement benchmark dataset and beyond. *IEEE Trans Image Process* 29:4376–4389
37. Wang Y, Guo J, Gao H, Yue H (2021) UIEC²-Net: CNN-based underwater image enhancement using two color space. *Signal Proc: Image Commun* 96
38. Li C, Guo J, Guo C (2018) Emerging from water: Underwater image color correction based on weakly supervised color transfer. *IEEE Signal Process Lett* 25(3):323–327
39. Islam MJ, Xia Y, Sattar J (2020) Fast underwater image enhancement for improved visual perception. *IEEE Robot Autom Lett* 5(2):3227–3234
40. Silberman N, Hoiem D, Kohli P, Fergus R (2016) Indoor segmentation and support inference from rgb-d images. *Proc Eur Conf Comput Vis (ECCV)* 746–760

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.