Journal of Southeast University (English Edition) Vol. 39, No. 4, pp. 410 - 415 Dec. 2023 ISSN 1003—7985

Two-stage attention for rapid underwater image enhancement

Yu Jing Zhang Le Wu Meng Jiang Shanghang Li Daojiang

(School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: A fast underwater image enhancement algorithm is proposed based on a two-stage attention mechanism to improve the quality of underwater degraded images. First, the proposed algorithm adopts a self-attention mechanism within features for enhancing the attention of the network to important information. Subsequently, a physical prior-based underwater transmission map is integrated into the network through a cross-attention mechanism for further enhancing the feature representation toward quality-degraded areas. Finally, a multiple joint loss function is designed using subjective and objective criteria for guiding the network to better visual enhancement effects. The experimental results on three benchmark datasets show that compared with five other underwater image enhancement methods, the proposed method obtains higher peak signal-to-noise ratio and structural similarity scores, exhibiting better performance. Therefore, the proposed method can effectively restore image color and texture details along with possessing real-time processing speed.

Key words: underwater image enhancement; self-attention; cross-attention; transmission map

DOI: 10.3969/j.issn.1003 - 7985.2023.04.010

As an indispensable component of an autonomous underwater vehicle (AUV)[1], the machine vision system has been widely developed for ocean observation, exploration, and extreme underwater environment operations. However, underwater images usually suffer from severe quality degradation, such as color deviation, low contrast, and blurred details, due to wavelength-dependent attenuation[2-3]. These degraded underwater images harm the development of high-level visual tasks, such as recognition and tracking. Therefore, underwater image enhancement (UIE) restores clear images from degraded images and is crucial for vision-guided AUVs.

Numerous UIE approaches have been proposed over the years and can be coarsely categorized as traditional and learning-based methods. Traditional methods tend to adopt general image enhancement methods targeted at onshore scenarios, i.e., histogram equalization[4], Retinex[5-6], and the image fusion method[7], completely ignoring the large domain shift. In addition, physical model-based methods[8-9] inverts the degradation process. Although their underlying theory is sound, these methods are usually unstable and sensitive in challenging underwater scenarios.

Recently, with the advancements in convolutional neural networks (CNNs), deep learning-based methods[10] have dominated the UIE field. U-color[11] designed a complex network comprising a multicolor space encoder network and a medium transmission-guided decoder network, but it was computationally inefficient. UGAN[12] and FUnIE-GAN[13] used the generative adversarial network (GAN) or Cycle-GAN to generate clear images. However, GAN-based methods are highly unstable and tend to produce undesirable artifacts. Recently, for the first time, U-shape[14] introduced a transformer to UIE, but its advantages in self-attention (SA) come at the cost of requiring many parameters. By contrast, the network architecture of Shallow-UWnet[15] is concise and thus exhibits favorable processing speed; however, its enhancement is inferior in quality. Moreover, most learning-based methods ignore the physical priors assumed in the degradation process of the traditional methods. Considering the abovementioned pros and cons, herein, we attempt to integrate the concise architecture, an attention mechanism, and a physical prior into a single model for finding the balance between effectiveness and efficiency.

In this paper, we present a two-stage attention (TSA)-based network for UIE for exploiting the physical model-based and data-driven methods. Specifically, SA in a transformer is first employed for spatial interaction, and the physical prior of a transmission map that can somewhat reflect the degree of regional degradation is further involved via cross-attention. Lastly, various losses are combined from restoration fidelity and visual similarity. With these improvements, the proposed method balances computational efficiency and modeling capacity.

1 Proposed Method

1.1 Network architecture

Fig. 1 illustrates an overview of the proposed network

Received 2023-07-07, **Revised** 2023-10-13.

Biographies: Yu Jing (1982—), female, doctor, associate research fellow, yujing@nwpu.edu.cn.

Foundation item: The Natural Science Basic Research Plan in Shaanxi Province of China (No. 2020JQ-208), Key Research and Development Program of Shaanxi (No. 2022GY-285, No. 2020SF-391), Foundation of Key Laboratory of Road Construction Technology and Equipment of Chang'an University(No. 300102259507).

Citation: Yu Jing, Zhang Le, Wu Meng, et al. Two-stage attention for rapid underwater image enhancement[J]. Journal of Southeast University (English Edition), 2023, 39(4): 410 - 415. DOI: 10.3969/j.issn.1003 - 7985.2023.04.010.

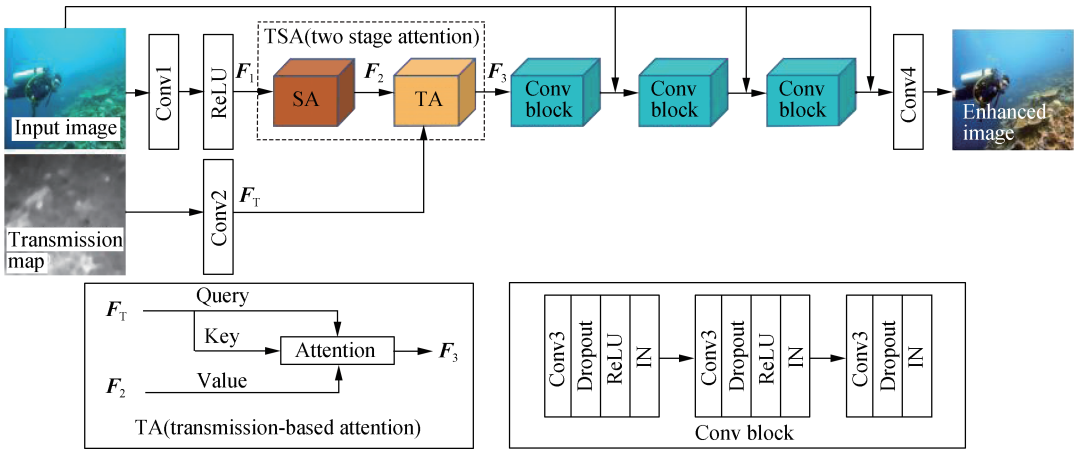

Fig. 1 Overview of the proposed network architecture

architecture. Obviously, the framework is extremely concise, although it is slightly more complex than Shallow-UWnet, i. e., it includes the proposed TSA before three fully convolutional blocks in series. In addition, skip-connection is used to avoid overfitting.

In particular, for an underwater image $\boldsymbol{I} \in \mathbf{R}^{H\times W\times 3}$ as input, the network applies a convolution layer with three kernels and a ReLU operation for extracting its shallow features $\boldsymbol{F}_1 \in \mathbf{R}^{H\times W\times 64}$. Subsequently, these features $\boldsymbol{F}_1$ pass through the TSA module, chained with three convolution blocks and output deep feature maps. This module includes two main components: SA and transmission-based attention (TA). An instance normalization layer is added to the convolution blocks to ensure the feature independence between each image. Lastly, a final convolution layer with three kernels generates the enhanced underwater image.

1.2 Two-stage attention

Unlike convolution, SA excels in capturing long-range dependencies and aggregating discriminative features. We introduce SA[16] to our framework for reinforcing the attention of the network on more serious color channels. The basic idea of SA is to adaptively select a small amount of useful information from input features and focus on that important information.

Given the features $\boldsymbol{F}_1$ from the first Conv-ReLU, SA calculates the correlation of the same set and outputs the features $\boldsymbol{F}_2 \in \mathbf{R}^{H\times W\times 64}$, as indicated by

$$\boldsymbol{F}_2 = \boldsymbol{F}_1 + \mathrm{softmax}(\boldsymbol{Q}^{\mathrm{T}}(\boldsymbol{F}_1)\boldsymbol{K}(\boldsymbol{F}_1))\boldsymbol{V}(\boldsymbol{F}_1) \tag{1}$$

where $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{V}$ represent 1×1 convolution for facilitating computation.

SA treats images indiscriminately regardless of the image quality in different regions, neglecting the uneven distribution of quality degradation in underwater images. Hence, to mitigate this issue, the physical prior is added to the degradation process.

Jaffe-McGlamery imaging model[17] is widely used in traditional underwater image restoration, which can be simply expressed as

$$\boldsymbol{I}^c(x) = \boldsymbol{J}^c(x)\boldsymbol{T}(x) + A^c(1-\boldsymbol{T}(x)) \qquad c\in\{\mathrm{r},\mathrm{g},\mathrm{b}\} \tag{2}$$

where $\boldsymbol{I}^c(x)$ is the degraded underwater image captured at point x; c represents the color channel; $\boldsymbol{J}^c(x)$ is the radiance of clear image; A^c is the homogeneous background light; $\boldsymbol{T}(x)$ is the medium transmission map indicating the degree of quality degradation in different regions. In this paper, the medium transmission map is estimated based on the general dark channel prior[18-19], which can be computed by

$$\hat{\boldsymbol{T}}(x) = \max_{y\in\boldsymbol{\Omega}(x)}\left(\frac{|A^c - \boldsymbol{I}^c(y)|}{\max(A^c, 1-A^c)}\right) \qquad c\in\{\mathrm{r},\mathrm{g},\mathrm{b}\} \tag{3}$$

where $\hat{\boldsymbol{T}}(x)$ is the estimated medium transmission map and $\boldsymbol{\Omega}(x)$ represents a local patch of size 15×15 centered at point x.

With the estimated transmission map, the problem is the way to utilize this physical information. Here, a cross-attention is introduced for incorporating the physical prior into the network because it enables the effective capture of the dependence between image and transmission map features by the network. To summarize, our TA can be formulated as

$$\boldsymbol{F}_3 = \boldsymbol{F}_2 + \mathrm{softmax}(1 - \boldsymbol{Q}^{\mathrm{T}}(\boldsymbol{F}_{\mathrm{T}})\boldsymbol{K}(\boldsymbol{F}_{\mathrm{T}}))\boldsymbol{V}(\boldsymbol{F}_2) \tag{4}$$

where $\boldsymbol{F}_{\mathrm{T}} \in \mathbf{R}^{H\times W\times 64}$ is the feature map obtained from the estimated transmission map and $\boldsymbol{F}_3 \in \mathbf{R}^{H\times W\times 64}$ is the feature refined through cross-attention. The product $\boldsymbol{Q}^{\mathrm{T}}(\boldsymbol{F}_{\mathrm{T}})\cdot\boldsymbol{K}(\boldsymbol{F}_{\mathrm{T}})$ computes the weight of the quality-degraded regions. The function $\mathrm{softmax}(1-\boldsymbol{Q}^{\mathrm{T}}(\boldsymbol{F}_{\mathrm{T}})\boldsymbol{K}(\boldsymbol{F}_{\mathrm{T}}))$ indicates that the larger attention weights should be assigned to regions with higher quality degradation. The use of inconsistent attenuation in different space areas further improves image enhancement. Therefore, we suppose that

TSA can improve the network adaptability at different turbidity regions of underwater images, thereby making our model more adaptable to diverse underwater scenes.

1.3 Loss function

To achieve a good balance between subjective and objective qualities, we design an integrated loss function including the objective criteria of MSE loss L_2 and Charbonnier loss L_{cha} and subjective criteria of SSIM loss L_{ssim}, perceptual loss L_{per}, and color loss L_{lab} for training. The overall loss is expressed as

$$L = \lambda_1 L_2 + \lambda_2 L_{\text{cha}} + \lambda_3 L_{\text{ssim}} + \lambda_4 L_{\text{per}} + \lambda_5 L_{\text{lab}} \quad (5)$$

The balance between the overall performance and local texture details is determined by the hyper-parameters. In our implementation, λ_1, λ_2, λ_3, and λ_4 are empirically set to 1, while λ_5 is set to 1×10^{-6} considering the loss range.

L_2 loss measures the difference between the enhanced image $\hat{\boldsymbol{J}}$ and reference image $\boldsymbol{J}$ as

$$L_2 = \sum_{i=1}^{H} \sum_{j=1}^{W} [\hat{\boldsymbol{J}}(i,j) - \boldsymbol{J}(i,j)]^2 \quad (6)$$

Charbonnier loss[20] adds a regularization term to L_1 loss and is expressed as

$$L_{\text{cha}} = \sum_{i=1}^{H} \sum_{j=1}^{W} \sqrt{[\hat{\boldsymbol{J}}(i,j) - \boldsymbol{J}(i,j)]^2 + \varepsilon} \quad (7)$$

In our implementation, ε is set to be 1×10^{-6}. This loss function is sufficiently robust to handle outliers and is thus stable during training. These objective criteria also tend to be global, which is easily observed.

We consider local structures and details that complement the abovementioned objective losses. To measure these factors, the similarity index (SSIM) loss is incorporated to impose structural similarity between the recovered and reference images. S represents SSIM. The loss function for the SSIM can be written as

$$L_{\text{ssim}} = 1 - S(\hat{\boldsymbol{J}}, \boldsymbol{J}) \quad (8)$$

In addition, the discrimination on high-level semantic feature representation is also considered important. We compute the perceptual loss on the VGG-19[21] network pretrained on the ImageNet dataset[22], as indicated by

$$L_{\text{per}} = \sum_{i=1}^{H} \sum_{j=1}^{W} | \boldsymbol{\Phi}_l(\hat{\boldsymbol{J}}(i,j)) - \boldsymbol{\Phi}_l(\boldsymbol{J}(i,j)) | \quad (9)$$

where $\boldsymbol{\Phi}_l$ is the l-th convolution layer of VGG-19.

To remove the color deviations of underwater images, we also introduce color loss in Lab space[14] because Lab space makes the color better distributed. It is defined as

$$L_{\text{lab}} = (\boldsymbol{l}_{\hat{J}}(i,j) - \boldsymbol{l}_J(i,j))^2 - \sum q(\boldsymbol{a}_{\hat{J}}) \times \log(q(\boldsymbol{a}_J)) - \sum q(\boldsymbol{b}_{\hat{J}}) \times \log(q(\boldsymbol{b}_J)) \quad (10)$$

where q stands for the quantization operator. Lab image $\boldsymbol{J}$ is divided into three channels, including lightness $\boldsymbol{l}_J$, component $\boldsymbol{a}_J$, and component $\boldsymbol{b}_J$. It is quantized to calculate the cross-entropy loss between the enhanced and reference images on different channels in Lab color spaces.

2 Experiments

2.1 Experimental settings

To evaluate our approach, three publicly available benchmarks are used: UIEB[11], LSUI[14], and UFO-12[23]. UIEB contains 890 real-world underwater images with manually selected reference images. The LSUI dataset includes 5 004 natural underwater image pairs and involves highly diverse underwater scenes, object categories, water types, and lighting conditions. The UFO-120 dataset comprises over 1 500 images collected from oceanic explorations from multiple locations.

Following U-shape[14], we randomly divide LSUI into Train-L4500 (4 500 images) and Test-L504 (504 images) for training and testing, respectively. Similarly, we divide UIEB into Train-U800 (800 images) and Test-U90 (90 images). Meanwhile, UFO-120 is only used for cross-dataset testing (1 500 images).

We evaluate using PSNR and SSIM metrics, reflecting the proximity to the reference. A higher PSNR score indicates closer image content, while a higher SSIM score reflects a more similar structure and texture.

Our network was implemented using the Pytorch framework with an NVIDIA GeForce RTX 3080 Ti GPU. We trained the network using an Adam optimizer with a learning rate of 2×10^{-4} and a layer dropout of 0.2. The model is trained for 500 iterations with a batch size of 1. The detailed architecture of our network is shown in Table 1.

Table 1 Detailed architecture of the proposed network

Layer	In channel	Out channel	Kernel size	Stride	Pad
Conv1	3	64	3	1	1
Conv2	1	3	3	1	1
Conv3	64	64	3	1	1
Conv4	64	3	3	1	1

2.2 Experimental results

2.2.1 Quantitative comparison

Quantitative results of different UIE algorithms on Test-U90 and Test-L504 are presented in Table 2. On these two datasets, the proposed method clearly outperforms all other competing methods in terms of PSNR and SSIM metrics. Our model trained on a larger dataset, namely LSUI, can achieve much higher PSNR and SSIM scores.

Furthermore, we conduct a cross-dataset experiment. These methods are trained on Train-L4500 and tested on UFO-120. Table 2 indicates that among all these methods, our approach obtains the highest PSNR and SSIM

Table 2 Evaluation of various methods on UIEB, LSUI, and UFO-120 datasets

Method	Test-U90		Test-L504		UFO-120	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
UGAN	20.68	0.86	22.13	0.81	22.76	0.76
FUnIE-GAN	19.45	0.86	23.72	0.78	23.45	0.74
U-color	20.78	0.87	23.47	0.80	21.87	0.74
Shallow-UWnet	21.27	0.81	22.81	0.85	24.70	0.73
U-shape	22.54	0.82	26.45	0.85	24.20	0.76
Ours	24.96	0.88	28.15	0.88	25.55	0.79

scores, revealing the model's superior generalization capability over other methods.

2.2.2 Qualitative comparison

Fig. 2 shows a visual comparison of various methods on Test-L504. The numbers presented in the top-right corner of each image refer to its PSNR. Among these methods, Shallow-UWnet exhibits the largest number of color artifacts and haze. UGAN and FUnIE-GAN fail to recover color and structural texture details. U-color provides a better color appearance but fails to comprehensively enhance the details. Although U-shape has a relatively good visual quality for human observers, color artifacts exist in some regions. By contrast, our results are the most consistent with the reference image in terms of human visual perception and texture details. Moreover, Fig. 3 illustrates the cross-dataset enhancement results of Shallow-UWnet and our method on UFO-120 while training on Train-L4500. Our method still achieves higher quality than Shallow-UWnet. The PSNR of each image is presented in the top-right corner of each image.

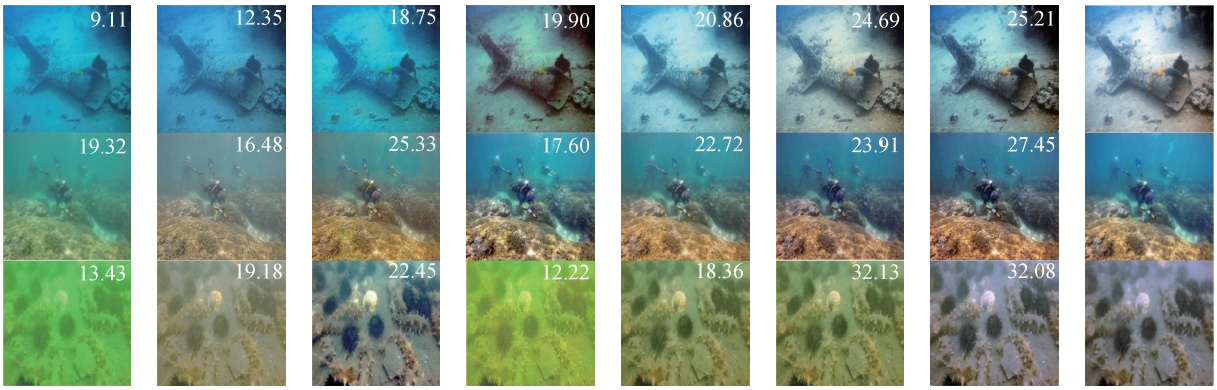

Fig. 2 Visual comparison of underwater images from TEST-L504. (a) Raw images; (b) Shallow-UWnet; (c) UGAN; (d) FUnIE-GAN; (e) U-color; (f) U-shape; (g) Our method; (h) Ground truth

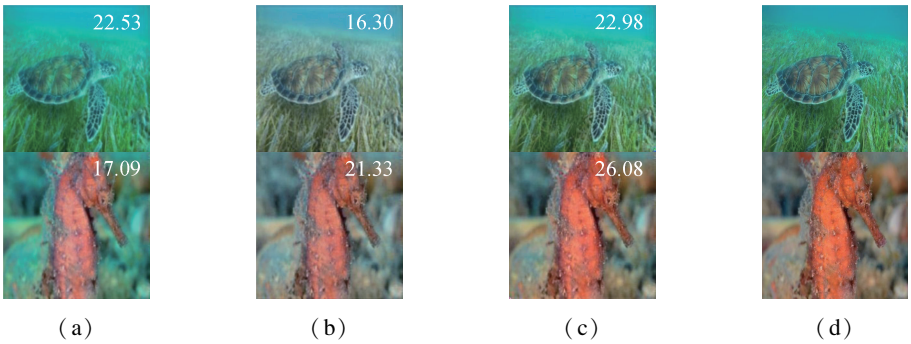

Fig. 3 Visual comparison of underwater images sampled from UFO-120. (a) Raw images; (b) Shallow-UWnet; (c) Our method; (d) Ground truth

2.3 Model capacity and efficiency

Table 3 presents the model size and inference time. All methods are tested on a PC with a single NVIDIA GeForce RTX 3080 Ti GPU. Shallow-UWnet and our method clearly outperform others, while U-color is the most time-consuming among all because of its complex network. Our approach can process one image in 0.03 s, indicating its superior computational efficiency.

Table 3 Model capacity and efficiency comparison

Method	$P/10^6$	R/s
UGAN	57.17	0.81
FUnIE-GAN	4.21	0.20
U-color	148.77	7.80
Shallow-UWnet	0.22	0.02
U-shape	11.09	0.09
Ours	0.31	0.03

Note: P is the number of trainable parameters; R is the inference time of an image.

2.4 Ablation study

To demonstrate the effect of each component of our model, we conduct the following ablation studies on UIEB, including the model without a TSA module (w/o TSA), the model without $L_{cha} + L_{ssim} + L_{lab}$ (w/o Loss), and the full model. As presented in Table 4, our full model achieves the best quantitative performance, verifying that our TSA module and the delicately designed loss functions are integral to the proposed method.

Table 4 Ablation study

Method	w/o TSA	w/o Loss	Full model
PSNR	23.68	22.15	24.96
SSIM	0.87	0.82	0.88

2.5 Application test

To further verify the effectiveness of our method for high-level vision tasks, it is applied as a preprocessing for an underwater depth estimation algorithm[24] and a semantic segmentation method[25]. Fig. 4 shows a visual comparison of various enhancing methods for these two tasks. Our method obviously generates more accurate and consistent estimates of depth maps and semantic segmentation than other methods, indicating its superiority for high-level vision tasks.

3 Conclusions

1) We combine a concise CNN with an attention mechanism in a transformer via comprehensive investigations of the status quo of UIE, integrating the physical prior of the transmission map into attention.

2) The fidelity and similarity losses are adopted to guide the network to obtain better objective indicators and visual quality, thereby achieving a good balance between subjective and objective qualities.

3) Extensive experiments prove that the proposed method achieves state-of-the-art performance on several recent benchmarks in terms of visual quality and quantitative metrics.

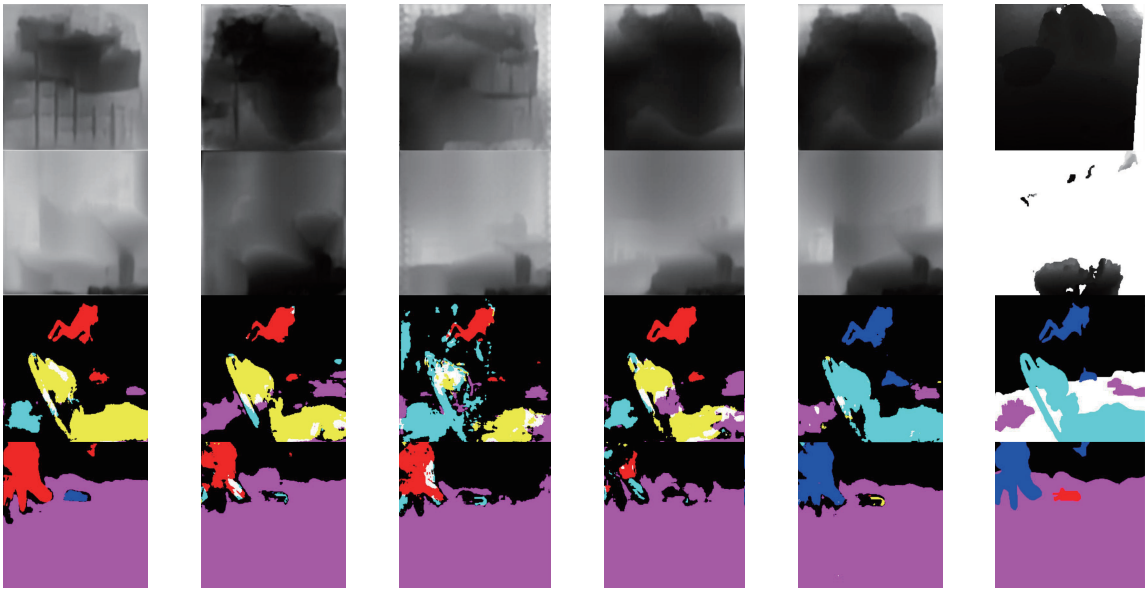

Fig. 4 Underwater depth estimates and semantic segmentations of different methods. (a) Raw image; (b) UGAN; (c) FUnIE-GAN; (d) U-color; (e) Our method; (f) Ground truth

References

[1] Zhao W, Qi H, Jiang Y, et al. A convolutional neural network accelerator for real-time underwater image recognition of autonomous underwater vehicle[J]. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 2021, **235**(10): 1839-1848. DOI: 10.1177/0959651820958208.

[2] Akkaynak D, Treibitz T, Shlesinger T, et al. What is the space of attenuation coefficients in underwater computer vision[C]//2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA, 2017: 568-577. DOI: 10.1109/CVPR.2017.68.

[3] Zhuang P, Wu J, Porikli F, et al. Underwater image enhancement with hyper-Laplacian reflectance prior[J]. *IEEE Transactions on Image Processing*, 2022, **31**: 5442-5455. DOI: 10.1109/TIP.2022.3196546.

[4] Hummel R. Image enhancement by histogram transformation[J]. *Computer Graphics and Image Processing*, 1977, **6**(2): 184-195. DOI: 10.1016/S0146-664X(77)80011-7.

[5] Fu X, Fan Z, Ling M, et al. Two-step approach for single underwater image enhancement[C]//2017 *International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*. Xiamen, China, 2017: 789-794. DOI: 10.1109/ISPACS.2017.8266583.

[6] Zhuang P, Ding X. Underwater image enhancement using an edge-preserving filtering Retinex algorithm[J]. *Multimedia Tools and Applications*, 2020, **79**(25/26): 17257-17277. DOI: https://doi.org/10.1007/s11042-019-08404-4.

[7] Dhivya R, Prakash R, Mohanraj M R. Color balance and fusion for underwater image enhancement [J]. *IEEE Transactions on Image Processing*, 2018, **27**(1): 379 - 393. DOI: 10.1109/TIP.2017.2759252.
[8] Li C, Guo J, Cong R, et al. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior[J]. *IEEE Transactions on Image Processing*, 2016, **25**(12): 5664 - 5677. DOI: 10.1109/TIP.2016.2612882.
[9] Zhou Y, Wu Q, Yan K, et al. Underwater image restoration using color-line model[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, **29**(3): 907 -911. DOI: 10.1109/TCSVT.2018.2884615.
[10] Wang Y, Zhang J, Cao Y, et al. A deep CNN method for underwater image enhancement[C]//2017 *IEEE International Conference on Image Processing* (*ICIP*). Beijing, China, 2017: 1382 - 1386. DOI: 10.1109/ICIP.2017.8296508.
[11] Li C, Anwar S, Hou J, et al. Underwater image enhancement via medium transmission-guided multi-color space embedding[J]. *IEEE Transactions on Image Processing*, 2021, **30**: 4985 - 5000. DOI: 10.1109/TIP.2021.3076367.
[12] Fabbri C, Jahidul Islam M, Sattar J. Enhancing underwater imagery using generative adversarial networks[C]//2018 *IEEE International Conference on Robotics and Automation* (*ICRA*). Brisbane, Australia, 2018: 7159 - 7165. DOI: 10.1109/ICRA.2018.8460552.
[13] Islam M J, Xia Y, Sattar J. Fast underwater image enhancement for improved visual perception[J]. *IEEE Robotics and Automation Letters*, 2020, **5**(2): 3227 - 3234. DOI: 10.1109/LRA.2020.2974710.
[14] Peng L, Zhu C, Bian L. U-shape transformer for underwater image enhancement[J]. *IEEE Transactions on Image Processing*, 2023, **32**: 3066 -3079. DOI: 10.1109/TIP.2023.3276332.
[15] Naik A, Swarnakar A, Mittal K. Shallow-UWnet: Compressed model for underwater image enhancement(student abstract)[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, **35**(18): 15853 - 15854. DOI: 10.1609/aaai.v35i18.17923.
[16] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]//*Proceedings of the* 31*st International Conference on Neural Information Processing Systems* (*NIPS*'17). Red Hook, NY, USA, 2017: 6000 -6010. DOI: 10.48550/arXiv.1706.03762.
[17] Trucco E, Olmos-Antillon A T. Self-tuning underwater image restoration[J]. *IEEE Journal of Oceanic Engineering*, 2006, **31**(2): 511 - 519. DOI: 10.1109/JOE.2004.836395.
[18] Peng Y T, Cao K, Cosman P C. Generalization of the dark channel prior for single image restoration[J]. *IEEE Transactions on Image Processing*, 2018, **27**(6): 2856 - 2868. DOI: 10.1109/TIP.2018.2813092.
[19] Chen C F, Fan Q, Panda R. CrossViT: Cross-attention multi-scale vision transformer for image classification [C]//2021 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). Montreal, Canada, 2021: 347 -356. DOI: 10.1109/ICCV48922.2021.00041.
[20] Lai W-S, Huang J-B, Ahuja N, et al. Fast and accurate image super-resolution with deep laplacian pyramid networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, **41**(11): 2599 - 2613. DOI: 10.1109/TPAMI.2018.2865304.
[21] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2019-09-04) [2023-06-29]. https://arxiv.org/abs/1409.1556.
[22] Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database[C]//2009 *IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL, USA. 2009: 248 - 255. DOI: 10.1109/CVPR.2009.5206848.
[23] Islam M J, Luo P, Sattar J. Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception [EB/OL]. (2020-07-16) [2023-06-29]. https://arxiv.org/abs/2002.01155.
[24] Gupta H and Mitra K. Unsupervised single image underwater depth estimation [C]//2019 *IEEE International Conference on Image Processing* (*ICIP*). Taipei, China, 2019: 624 -628. DOI: 10.1109/ICIP.2019.8804200.
[25] Islam M J, Edge C, Xiao Y, et al. Semantic segmentation of underwater imagery: Dataset and benchmark[C]//2020 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*). Las Vegas, NV, USA, 2020: 1769 -1776. DOI: 10.1109/IROS45743.2020.9340821.

基于两阶段注意力机制的快速水下图像增强方法

蔚 婧 张 乐 吴 萌 江尚航 李道江

(西北工业大学航海学院,西安 710072)

摘要:为提升水下退化图像的质量,提出了一种基于两阶段注意力机制的快速水下图像增强算法.首先,在特征内采用自注意力机制来加强网络对重要信息的关注.然后,通过交叉注意力机制,将基于物理先验的水下传输图融入网络,以增强网络对质量退化区域的特征表达.最后,从主观和客观标准出发,设计了多项联合损失函数,引导网络获得更好的视觉增强效果.3 个基准数据集上的实验结果表明,与 5 种水下图像增强方法相比,所提方法的峰值信噪比和结构相似性分数更高,性能明显更优.该方法不仅能有效恢复图像颜色和纹理细节,而且具备实时高效的处理速度.

关键词:水下图像增强;自注意力;交叉注意力;传输图

中图分类号:TP391