

An Underwater Image Enhancement Model Combining Physical Priors and Residual Network

Fan Xinnan, Zhou Xuan, Chen Hongzhu, Xin Yuanxue and Shi Pengfei*

Considering light absorbing and scattering problems in connection with wavelength can decrease the visibility, contrast and color distortion of images, we propose a new type of convolutional neural network with two training phases. Firstly, the coordinate attention module is integrated into the residual block of the residual group in the backbone network, which is used to strengthen the feature extraction capability of the network. Secondly, since the unrealistic image colors may degrade the image details, an unsupervised method that combines the physical prior knowledge and the real underwater images is proposed to finetune the backbone network. Furthermore, a model protection mechanism is designed to guarantee the successful execution of the training. The experimental results indicate the proposed model can effectively optimize the contrast, color and image quality of the underwater image. Compared with relevant algorithms, our UCIE and NIQE are respectively 0.525 and 4.149, which further verifies the superiority of the proposed model.

Introduction: Optical images are critical for underwater engineering exploration, archaeology, and biological research. However, getting high-quality images is a very difficult task. Light is highly attenuated and scatters as it travels through water, leading to color distortion and a bluish or greenish tint in underwater images. Additionally, suspended particles in the water can further reduce image clarity and contrast. Notable progress has been made to improve the visual quality of underwater images in recent years. Drews et al. [1] adapt the DCP [2] and create the Underwater Dark Channel Prior (UDCP). Kashif et al. [3] propose a sliding stretch-based method for underwater image perception (ICM). Song et al. [4] propose an underwater image scene based on the Underwater Light Attenuation Prior (ULAP). Zhuang et al. [5] process the underwater images with hyper-laplacian reflectance priors. Most of these algorithms are based on underwater optical imaging model, which focus on calculating the parameters by mathematical deduction. Despite having strong physical interpretability, these methods result in poor image quality.

The advent of deep neural networks has led to their application in underwater image enhancement. Wang et al. [6] propose a CNN-based underwater image enhancement network called UIE-Net. Li et al. [7] propose WaterGAN, which uses RGB-D images taken in air and underwater images as input to adversarially train a generative network. Fabbri et al. [8] propose UGAN, a generative adversarial network applicable to underwater scenes, which uses CycleGAN [9] to generate paired images and is able to provide a higher-quality training set for the model. Li et al. [10] improve the loss function of the traditional CycleGAN. Islam et al. [11] propose FunIEGAN, a real-time underwater image enhancement model based on conditional generative adversarial networks [12]. Liu et al. [13] proposed a dual-adversarial contrastive learning method for underwater image enhancement that alleviates the need for paired data in an unsupervised manner. Fu et al. [14] resolve underwater image enhancement into distribution estimation and consensus process. Although deep learning-based methods have shown superior performance compared to traditional methods, the limited available datasets often result in unnatural color representations. In addition, deep learning methods only pursue high similarity between the output image and the target image, while neglect whether it meets the human visual evaluation standards.

Therefore, we explore a method that combines the efficient performance and robustness of deep learning with the interpretability of physical models, and makes the output image color natural. Inspired by Chen et al. [15], we propose an end-to-end network with two training phases. In the first phase, we design and train a backbone network composed of residual blocks and jump connections with strong image enhancement capabilities. In the second phase, we add a physical prior fine-tuning module to further improve the backbone network and enhance the quality of image colors. The major contributions of the proposed method are as follows:

- We propose a novel residual attention block, utilizing the Coordinate Attention mechanism, as a fundamental component of the backbone network. This basic block enhances the network's feature extraction capability and makes the network easier to train.

- We combine several popular, well-grounded physical priors into an unsupervised correction module, which utilizes real underwater images (unpaired) to fine-tune the backbone.
- Our method outperforms existing algorithms in terms of objective metrics and visual inspection, demonstrating significant improvements in color saturation, contrast, and overall image clarity.

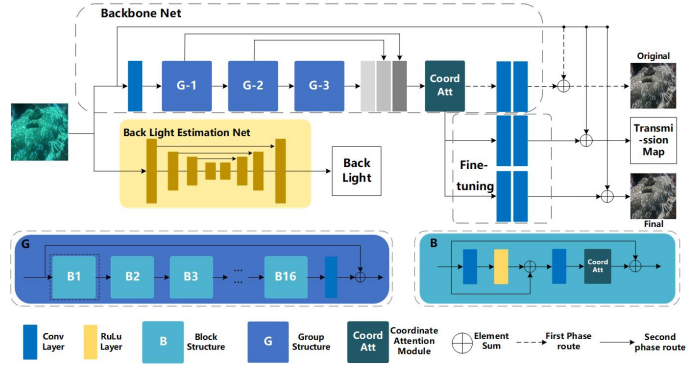


Fig. 1. Overview of the two-phase model.

Architecture : The complete model is shown in Fig.1. The backbone network consists of a shallow feature extraction layer, followed by three successive residual groups with skip connections that cascade the output features. The feature maps are then input to a Coordinate Attention module, inspired by Hou et al. [16]. As shown in Fig.2, the module aggregates features along two spatial directions and encodes them into direction-aware and position-sensitive attention maps that are applied complementarily to the feature maps. The output feature map is then processed through two convolutional layers and a skip connection that spans the entire network to produce the final clear image.

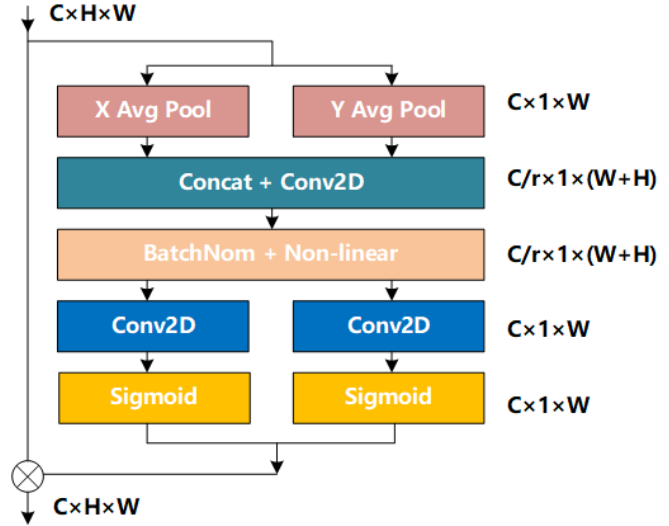


Fig. 2. Coordinate Attention module.

The residual group is comprised of multiple tandem residual blocks, a convolutional layer, and a skip connection, as depicted in the bottom left part of Fig. 1. Successive residual groups are added to increase the depth of the backbone network. Jump connection facilitates network training.

We propose a residual attention block consisting of a local residual structure and a Coordinate Attention module, as depicted in the bottom right part of Fig. 1. The residual attention block preserves shallow information and efficiently passes it to deeper layers, allowing the backbone network to focus on relevant feature information, such as complex textures and color biases in the image.

Training in the first phase: Underwater optical imaging model describes the process of underwater image quality deterioration, which can be

expressed as this equation:

$$I = Jt + B(1 - t) \quad (1)$$

where I denotes the underwater image captured by the camera, J is the clean underwater image, B is the global back light and t is the transmission map.

In the first phase, the backbone network output J and t by learning the relationship between the input low quality underwater image and the clean image. To estimate the back light B , we link the back light estimation network from DCPDN[17] with our backbone network. To ensure the accuracy of network outputs J, B and t , we set L_{joint} which consists of two essential components, L_J and L_I . The first component is L_J , which is formulated as follows:

$$L_J = \|J - J_o\|_1 \quad (2)$$

where J is the output clear underwater image and J_o is the input training ground truth.

We input J, B, t into the underwater optical imaging model showed in equation (1) to calculate the input image I . We calculate the loss between it and the input low quality underwater image, which is formulated as:

$$L_I = \|I - I_o\|_1 \quad (3)$$

where I_o is the input low quality underwater image.

Finally, we obtain the loss function L_{joint} for the training of backbone network:

$$L_{joint} = \alpha L_J + \beta L_I \quad (4)$$

where α and β act as hyper-parameters.

So in the first training phase, the overall loss is as follows:

$$L_{first} = L_{joint} \quad (5)$$

Thanks to the well-designed backbone network, our model produces satisfactory clean results and accurate physical parameters in the first phase.

Training in the second phase: To mitigate computational errors arising from scene changes and the absence of clean underwater images, we employ the physical prior as an unsupervised training method during the second phase. By inputting real-world underwater images, we obtain a physical output interface that can accurately capture image characteristics. Then we connect it to the output part of the backbone. This interface contains two branches, each consisting of two convolution layers that output a transmission map and a clear image, respectively. The following physical priors are applied in our model:

BCP:Applying the bright channel prior to the underwater dark environment enhancement results in a significant improvement in global illumination while recovering more detail. We incorporate this prior into the training using the following loss function:

$$L_{bcp} = \|t - \tilde{t}\|_1 \quad (6)$$

where t is the transmission map estimated by the BCP and \tilde{t} is the transmission map estimated by the backbone network proposed in the first phase.

DCP:Using only BCP can lead to an over-brightened and unrealistic image. In this paper, we simultaneously utilize both DCP and the BCP to achieve an optimal illumination balance. We combine the strengths of previous DCP techniques[18] with the following loss function:

$$L_{dcp} = E(t, \tilde{t}) = t^T L t + \lambda(t - \tilde{t})^T (t - \tilde{t}) \quad (7)$$

where t is the transmission map estimated by the underwater DCP, \tilde{t} is the transmission map estimated by the backbone network, λ is the hyperparameter, and L is the Laplacian matrix.

CLAHE:Contrast limited adaptive histogram equalization has shown promise in mitigating the blurring that commonly affects underwater images. We performed loss calculations on both the output of our model and the processing results obtained with CLAHE:

$$L_{clahe} = \|J - J_{clahe}\|_1 \quad (8)$$

where J is the output of backbone network and J_{clahe} is the result of CLAHE enhancement.

At this point we obtain the physical priors loss unit:

$$L_{phy} = aL_{bcp} + bL_{dcp} + cL_{clahe} \quad (9)$$

Due to the shift in training focus during fine-tuning, the backbone network's enhancing ability is weakened. To solve this problem, we put forward a protective mechanism to help our model memorize the previous image-enhancing task in the first phase. In practical terms, we replicate the backbone network in the first phase as N_0 and compare the output of the new network N_1 in the second phase with that of N_0 . We minimize the loss function $L_{protect}$ to achieve this:

$$L_{protect} = \|F_p - F_{0p}\|_1 + \|F_{up} - F_{0up}\|_1 \quad (10)$$

Where F_p and F_{0p} are the output feature maps of N_0 and N_1 for paired images, respectively. F_{up} and F_{0up} are the output feature maps of N_0 and N_1 for unpaired images.

In summary, the overall loss function in the second phase is as follows:

$$L_{second} = L_{joint} + L_{phy} + L_{protect} \quad (11)$$

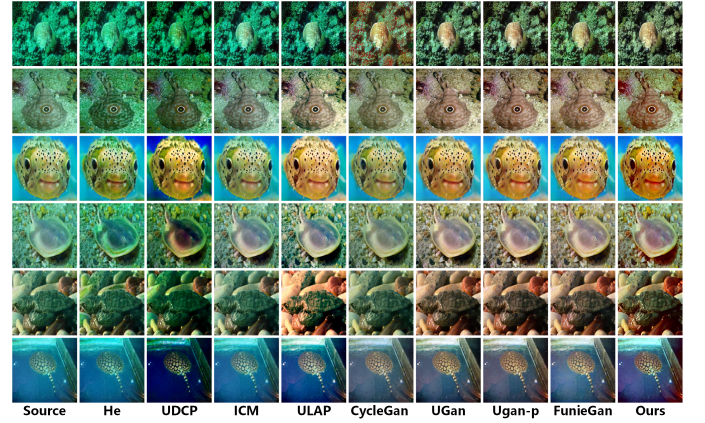


Fig. 3. Comparison of different methods.

Experiments: We utilize data from the Enhancing Underwater Visual Perception (EUVP), established by Islam et al. [7]. On the whole, 5.1K paired data are input to the first phase while 2K real underwater images are input to the second phase, with all images cropped to 128*128 size as input. The model is trained for 100 epochs by the Adam optimizer in the first phase and 4 epochs in the second phase. We adopt the cosine annealing strategy (He et al.[19]) to adjust the learning rate. The training batch size is set to 4. As for the loss function, we set $\alpha = 1$, $\beta = 0.8$, $a = 0.05$, $b = 0.001$ and $c = 1$.

Table 1: Quantitative results using UCIQE and NIQE metrics

Method	UCIQE	NIQE
He	0.460	4.441
UDCP	0.516	4.480
ULAP	0.485	4.591
ICM	0.447	4.405
CycleGan	0.430	5.181
UGan	0.468	4.975
UGan-p	0.480	5.057
FunIEGAN	0.460	5.063
Ours	0.525	4.149

The evaluation metrics used in this article are UCIQE (Underwater Color Image Quality Evaluation metrics) and NIQE (Natural Image Quality Evaluator), both of which do not need to be compared with reference images when used. Compared with PSNR and SSIM, UCIQE and NIQE have more visual reference value and are specific for underwater images. We compare our method with several traditional algorithms: He[2], UDCP[1], ICM[3], and ULAP[4], as well as several deep learning methods: CycleGan[9], UGan[8], UGan-P[8], and FunIEGAN[11]. Table I present the numerical scores for the two metrics, with bold font indicating the best results. The results demonstrate that our method achieves optimal performance on UCIQE and NIQE evaluations. Our UCIQE and NIQE reach 0.525 and 4.149 respectively, with a 0.009-0.095 increase in UCIQE and a 0.256-1.032 decrease in NIQE compared to other methods.

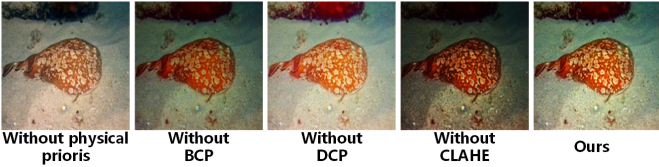


Fig. 4. Results of ablation experiment on physical priors.

The processing results of each algorithm are shown in Fig.3. The results indicate that He's method has negligible impact on underwater image quality. UDCP enhances the contrast of the images, but does not solve the problem of color distortion, resulting in images with unrealistic green and blue hues and low image brightness in some cases (e.g., image4, image5, and image6). ICM improves greenish hues, but at the cost of image contrast and overall quality. ULAP further improves greenish hues, but may lead to overexposure (e.g., the turtle in the image5). UGan and UGan-p show promising results in correcting blue color bias, but make the picture too bright (e.g., the rocky coral reef in the image1). FunIEGAN slightly improves overall image quality, but lacks detail and brightness in some areas (e.g., the fish patterns in the image2 and the image3). Compared to the above methods, our model significantly improves the underwater image coloration problem, and further enhances the image contrast with brighter and more vivid colors.

Table 2: Comparison of ablation experiment results

Number of residual groups	UCIQE	NIQE
1	0.492	4.071
2	0.501	4.103
4	0.498	4.125
3(ours)	0.518	4.139

Ablation experiments: We conduct an ablation experiment on the number of residual groups in the network structure. The number of residual groups is set to 1, 2, and 4. Then we randomly select 200 images for evaluation. The test results are summarized in TableIII.

As the number of residual groups increases, the feature extraction capability increases and the output image quality improves. However, the performance score decreases when the number of residual groups reaches four, potentially due to the increased difficulty of model training and overfitting resulting from the model's increased depth.

In addition, we conduct an ablation experiment on the physical priors in the second phase. As shown in Fig.4, the visual results reveal that the backbone network without physical priors fine-tuning displays strong enhancement potential, but the contrast and colors could be further improved. In contrast, the model lacking BCP produced dark images with incomplete texture information, while the model lacking DCP produced slightly blurred images with exposure issues. The image lacking CLAHE guidance exhibited uneven lighting and poor results. By integrating all three physical priors, the resulting images demonstrate the most complete texture effects, rich colors, strong contrast, and excellent visual outcomes.

Conclusion: In this letter, we propose a two-phase training model for enhancing underwater images. In the first phase, we incorporate Coordinate Attention module into the residual block to create a deep residual network, which serves as an efficient model backbone. In the second phase, we fine-tune the backbone guided by three physical priors to get the final model. We integrate deep learning methods with physical priors to ensure that the model maintains the strong feature capture ability of convolutional neural networks while also having empirical control over image quality. Our experiments demonstrate that our method performs better than other methods in both UCIQE and NIQE. However, due to the high computational requirements of our model, the processing time for images is lengthy. Future work will focus on simplifying the model to reduce computational complexity.

Acknowledgment: This work was supported in part by the National Key R&D Program of China, Grant/Award Number:2022YFB4703402, and in part by the Fundamental Research Funds for the Central Universities (B220202020).

Fan Xinnan, Zhou Xuan, Chen Hongzhu and Xin Yuanxue (College of Information Science and Engineering, Hohai University, Changzhou, China) Shi Pengfei(College of Artificial Intelligence and Automation, Hohai University, Changzhou, China)

E-mail: shipf@hhu.edu.cn

References

- 1 P. Drews, E. Nascimento, F. Moraes, S. Botelho, and M. Campos, "Transmission estimation in underwater single images," in *Proceedings of the IEEE international conference on computer vision workshops*, 2013, pp. 825–830.
- 2 K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.
- 3 K. Iqbal, R. A. Salam, A. Osman, and A. Z. Talib, "Underwater image enhancement using an integrated colour model," *IAENG International Journal of computer science*, vol. 34, no. 2, 2007.
- 4 W. Song, Y. Wang, D. Huang, and D. Tjondronegoro, "A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration," in *Advances in Multimedia Information Processing-PCM 2018: 19th Pacific-Rim Conference on Multimedia, Hefei, China, September 21-22, 2018, Proceedings, Part I* 19. Springer, 2018, pp. 678–688.
- 5 P. Zhuang, J. Wu, F. Porikli, and C. Li, "Underwater image enhancement with hyper-laplacian reflectance priors," *IEEE Transactions on Image Processing*, vol. 31, pp. 5442–5455, 2022.
- 6 Y. Wang, J. Zhang, Y. Cao, and Z. Wang, "A deep cnn method for underwater image enhancement," in *2017 IEEE international conference on image processing (ICIP)*. IEEE, 2017, pp. 1382–1386.
- 7 J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, "Watgan: Unsupervised generative network to enable real-time color correction of monocular underwater images," *IEEE Robotics and Automation letters*, vol. 3, no. 1, pp. 387–394, 2017.
- 8 C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 7159–7165.
- 9 P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- 10 N. J. Li Qing-Zhong, Bai Wen-Xiu, "Underwater image color correction and enhancement based on improved cycle-consistent generative adversarial networks," *ACTA AUTOMATICA SINICA*, vol. 46, pp. 1–10, 2020.
- 11 M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3227–3234, 2020.
- 12 M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.
- 13 R. Liu, Z. Jiang, S. Yang, and X. Fan, "Twin adversarial contrastive learning for underwater image enhancement and beyond," *IEEE Transactions on Image Processing*, vol. 31, pp. 4922–4936, 2022.
- 14 Z. Fu, W. Wang, Y. Huang, X. Ding, and K.-K. Ma, "Uncertainty inspired underwater image enhancement," in *European Conference on Computer Vision*. Springer, 2022, pp. 465–482.
- 15 Z. Chen, Y. Wang, Y. Yang, and D. Liu, "Psd: Principled synthetic-to-real dehazing guided by physical priors," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 7180–7189.
- 16 Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 13 713–13 722.
- 17 H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3194–3203.
- 18 A. Golts, D. Freedman, and M. Elad, "Unsupervised single image dehazing using dark channel prior loss," *IEEE transactions on Image Processing*, vol. 29, pp. 2692–2701, 2019.
- 19 T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li, "Bag of tricks for image classification with convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 558–567.