Preprints.org

DUVR-Net: A Dual-Path Underwater Video Restoration Network Based on Middle-Frame Residual and Transformation Convolution Kernel

Xun Cai , Yun Wang , Tianyou Wang , Yeqian Yang , Jianguo Lv , Kin-Tak Lau , Yanbo Gao *

Article

DUVR-Net: A Dual-path Underwater Video Restoration Network Based on Middle-Frame Residual and Transformation Convolution Kernel

Xun Cai [1], **Yun Wang** [2], **Tianyou Wang** [2], **Yeqian Yang** [2], **Jianguo Lv** [3], **Kin-Tak LAU** [4] **and Yanbo Gao** [1,*]

1   School of Software, Shandong University-Weihai Research Institute of Industrial Technology, Jinan & Weihai, China
2   School of Software, Shandong University, Jinan, China
3   Carbon Fiber Engineering Technology Research Center, School of Materials Science & Engineering, Shandong University, Jinan, China
4   Technological and Higher Education Institute of Hong Kong, Vocational Training Council, Hong Kong, China
*   Correspondence: ybgao@sdu.edu.cn

Abstract: Underwater video enhancement aims to improve the visibility and frame quality of underwater videos, which is crucial for marine and lake research and exploration. However, existing methods primarily rely on image enhancement algorithms that independently enhance individual frames, neglecting the frame-to-frame continuity and style similarity relationships inherent in underwater videos. Moreover, current methods require excessive parameters and slow training processes to accommodate 4K video enhancement. To address these challenges, we propose the Dual-path Underwater Video Restoration Network(DUVR-Net), which first enhances the middle frame and then transfers the enhancement effect to adjacent frames through a dual-path approach using residual compensation and transformed convolution kernels for inter-frame information transmission.Without the need to enhance each frame independently,our model requires only approximately 90k parameters. Experiments on real underwater videos demonstrate that our network can effectively accomplish 4K underwater video enhancement with impressive results.

Keywords: underwater video enhancement; residual patches; transformation convolution kernels; dual-channel processing

---

1. Introduction

Water is a crucial medium for human exploration and research, covering approximately 71% of Earth's surface. Underwater video recording plays a vital role in marine biology research, underwater archaeology, and ecological surveys [1–4]. However, due to the physical properties of water, such as light absorption, scattering, suspended particles, and other factors,underwater videos typically suffer from poor visibility, color distortion, and low contrast [5], which significantly hinder their practical applications. Traditional underwater video enhancement methods primarily rely on image processing algorithms to enhance each frame independently. While these methods have achieved some success in improving video quality, they face several key challenges. First, processing each frame independently ignores the inherent temporal continuity and style consistency between frames in a video, which may lead to temporal artifacts and inconsistent enhancement effects. The emergence of UVEB [6] provides a large-scale (4K) high-resolution underwater video enhancement dataset and introduces UVE-Net, which converts current frame information into convolution kernels and passes them to adjacent frames to achieve effective inter-frame information exchange. However, the UVE-Net model is massive, requiring substantial computational resources and parameters, resulting in slow training processes and limited practical applications.

To address these limitations, we propose a novel underwater video enhancement framework that efficiently processes video sequences while maintaining temporal coherence. Our method introduces a

dual-channel mechanism that first divides long video sequences into multiple shorter image sequences for input. For each image sequence, we extract the middle frame and enhance it using multi-scale and multi-channel attention mechanisms. The multi-channel processing better extracts image color features for enhancement, while multi-scale processing captures image details at different scales to enhance image detail enhancement. After enhancing the middle frame of the image sequence, the enhancement effect is propagated to adjacent frames through two complementary pathways. The effectiveness of our dual-channel mechanism is demonstrated in Figure 1, where we compare our method against existing approaches (UVE-Net [1] and PRWNet [7]) on a challenging underwater scene. As shown, our method better preserves the natural color and detail of underwater structures while effectively removing blur and color distortion present in the original frames.
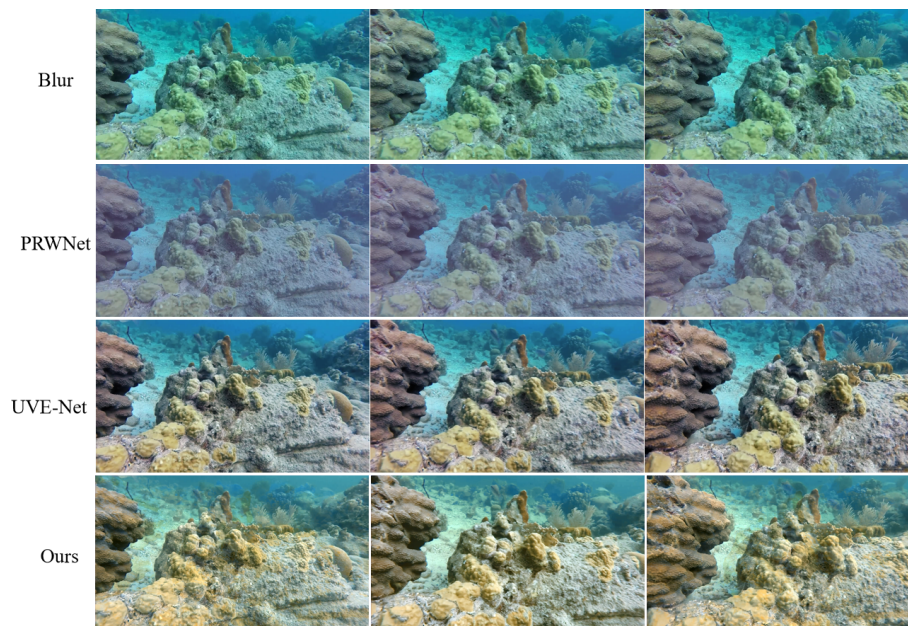

Figure 1. Visual results of different enhancement methods on a real underwater video.

In DUVE-Net, the first pathway is the residual-based repair pathway. Within the same short image sequence, frames exhibit temporal continuity and style consistency, so the enhancement effects needed for other frames share similarities with the enhancement effects obtained from the middle frame. Therefore, the enhancement effects needed for other frames can be derived from the middle frame's enhancement effects through inter-frame relationships. By calculating the difference between the enhanced and original middle frame, we obtain the enhancement effect residual of the middle frame. Considering the differences between frames, we calculate the inter-frame residual by comparing the middle frame with remaining adjacent frames. By applying convolution to the inter-frame residual as weight parameters for the middle frame's enhancement effect residual, we can obtain independent enhancement effects for each frame, which can be added to the corresponding original frames to achieve enhanced results. The residual-based repair pathway effectively transmits the enhancement effects of the middle frame.

The second pathway is the convolution-based repair pathway, which generates adaptive convolution kernels from the enhanced middle frame to process the original adjacent frames. Each original frame independently uses the enhanced middle frame to generate convolution kernels for convolution, allowing the middle frame's detail information to be transmitted to each frame. This design achieves effective inter-frame information transfer. The convolution-based repair pathway focuses more on the details in the enhancement effect transmission, reducing the impact of inter-frame differences.

The main contributions of our work can be summarized as follows:

1. We propose a novel underwater video enhancement framework that explicitly considers inter-frame relationships, ensuring temporal consistency in enhancement results.

2. We introduce a dual-channel mechanism combining residual-based repair pathway and convolution-based repair pathways, which effectively propagates enhancement information from middle frames to adjacent frames.

3. Our method achieves remarkable parameter efficiency, requiring only $\sim 90k$ parameters while maintaining high-quality enhancement effects for 4K underwater videos. As shown in Figure 2, our DUVR-Net achieves superior PSNR performance with significantly fewer parameters compared to existing methods.

4. Extensive experiments on real underwater videos demonstrate the effectiveness of our method in terms of both visual quality and computational efficiency.
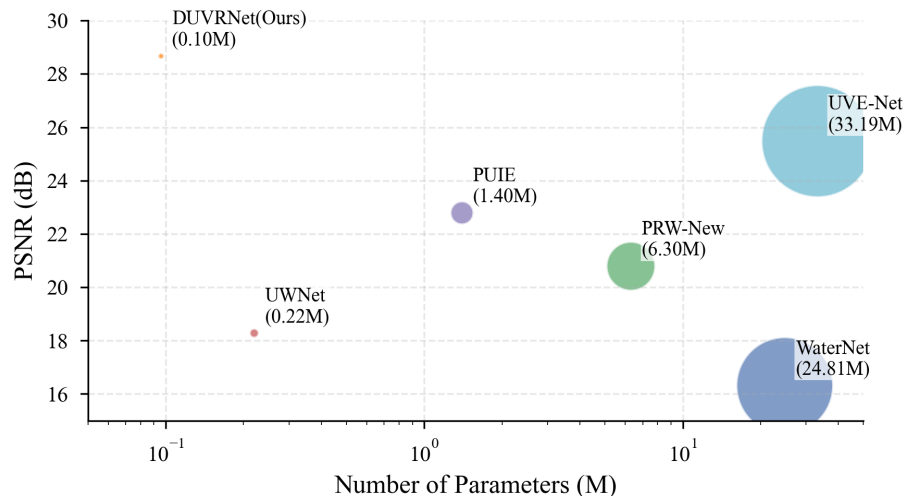

Figure 2. Comparison of recent state-of-the-art methods and our method.

2. Related Work

2.1. Underwater Image Enhancement Methods

Underwater Image Enhancement methods can be broadly categorized into physical model-based methods, physical model-free methods, and deep learning-based methods. Physical model-based methods construct models by considering image degradation factors in specific environments, thereby reversing or compensating for degradation effects and improving image quality [8–10]. Physical model-free methods directly adjust image pixel values through histogram equalization [11] and color balance, bypassing the complex optical propagation models to enhance underwater images through direct manipulation of pixel intensity distributions and color characteristics. With the successful application of deep learning in computer vision tasks, various innovative architectures have emerged. Notable networks include WaterNet [12], which employs adaptive filters [13] and CNNs [14] for quality improvement; Ucolor [15], which utilizes medium transmission-guided multi-color space embedding; and FA+Net [16], which incorporates color enhancement and pixel attention modules. Transformer-based [17] solutions such as U-shape [18] have also demonstrated excellence in capturing global features and handling long-range dependencies. Recent lightweight networks developed based on generative adversarial networks, such as FspiralGAN10 [19], are more suitable for resource-constrained environments. Although these networks have made significant progress, challenges still persist in color distortion correction, texture preservation, and computational efficiency. In the field of UIE, balancing quality and efficiency remains a more critical issue.

2.2. Underwater Video Enhancement Methods

Early underwater video enhancement methods primarily extended single-frame image enhancement algorithms to video processing through frame-by-frame processing. Although some methods improved efficiency through fast filtering algorithms or lightweight deep networks, this independent frame processing often led to temporal artifacts and inter-frame flickering. To address these issues,

researchers began exploring temporal relationships between frames. For instance, Ancuti et al. [20] introduced temporal bilateral filtering to achieve inter-frame smoothing, while Li et al. [21] extended video dehazing algorithms to underwater scenarios. Qing et al. [22] proposed a spatiotemporal information fusion method to estimate consistent transmission maps and background light.Meanwhile, recent advances in video restoration have provided new insights for underwater video enhancement. Researchers have proposed various models to capture inter-frame relationships and aggregate temporal information, including optical flow estimation, deformable convolutions, and cross-frame attention layers. Notable examples include Shift-Net14 [23], which introduces grouped spatiotemporal displacement to implicitly capture inter-frame correspondences, and UVE-Net [6], which innovatively transforms frame information into convolution kernels for efficient inter-frame information exchange. Inspired by these approaches, we propose DUVR-Net, a dual-path method that simultaneously utilizes residual compensation and transformative convolution kernels for underwater video enhancement.

3. Methods

3.1. Overview of DUVR-Net

The input to DUVR-Net consists of consecutive frames $\{x_i \in \mathbb{X}^{3 \times H \times W}\}_{i=t-k}^{t+k}$ from an underwater video, where t denotes the index of the middle frame, and $T = 2k + 1$ represents the total number of input frames. The output of DUVR-Net comprises enhanced consecutive underwater video frames $\{y_i \in \mathbb{Y}^{3 \times H \times W}\}_{i=t-k}^{t+k}$. The operation of DUVR-Net can be formulated as shown in Equation (1):

$$\{y_i = DUVR\_Net(x_i)\}_{i=t-k}^{t+k} \tag{1}$$

$\mathbb{X}$ and $\mathbb{Y}$ denote the original low-quality underwater videos and their corresponding enhanced versions, respectively.The overall architecture of DUVR-Net is illustrated in Figure 3.

3.2. Middle Frame Enhancement Module

The Middle Frame Enhancement Module primarily focuses on enhancing individual middle frames, with its network architecture specifically designed for image-mode enhancement. To address the challenges of contrast deficiency and color distortion during the enhancement process, we adopt the enhancement strategy from FA+Net [16], implementing a dual-branch enhancement mode comprising multi-channel and multi-scale pathways during middle frame enhancement. The multi-channel branch transforms the input into features with varying channel dimensions to better capture color feature distributions across RGB channels. In the multi-scale branch, the input is converted into features with different heights and widths, utilizing 1×1 convolutions to perform consecutive processing of individual image pixels, which not only reduces the parameter burden but also better captures color information for each pixel. The information fusion of the dual branches is accomplished through Inverse Fourier Transform [24], with a hyperparameter α controlling the fusion ratio between spatial and frequency domain information.Subsequently, the fused features undergo fine-grained enhancement through the serial application of multi-scale modules and spatial attention mechanisms, ultimately producing the enhanced middle frame.

3.3. Video Enhancement Module

The video enhancement module takes as input a sequence of consecutive frames from a short underwater video, along with their corresponding enhanced high-quality intermediate frame. The output is the enhanced consecutive underwater video frames corresponding to this input sequence.The enhancement is achieved by propagating information from the enhanced middle frame to the remaining frames through residual and Transformation convolution kernels.
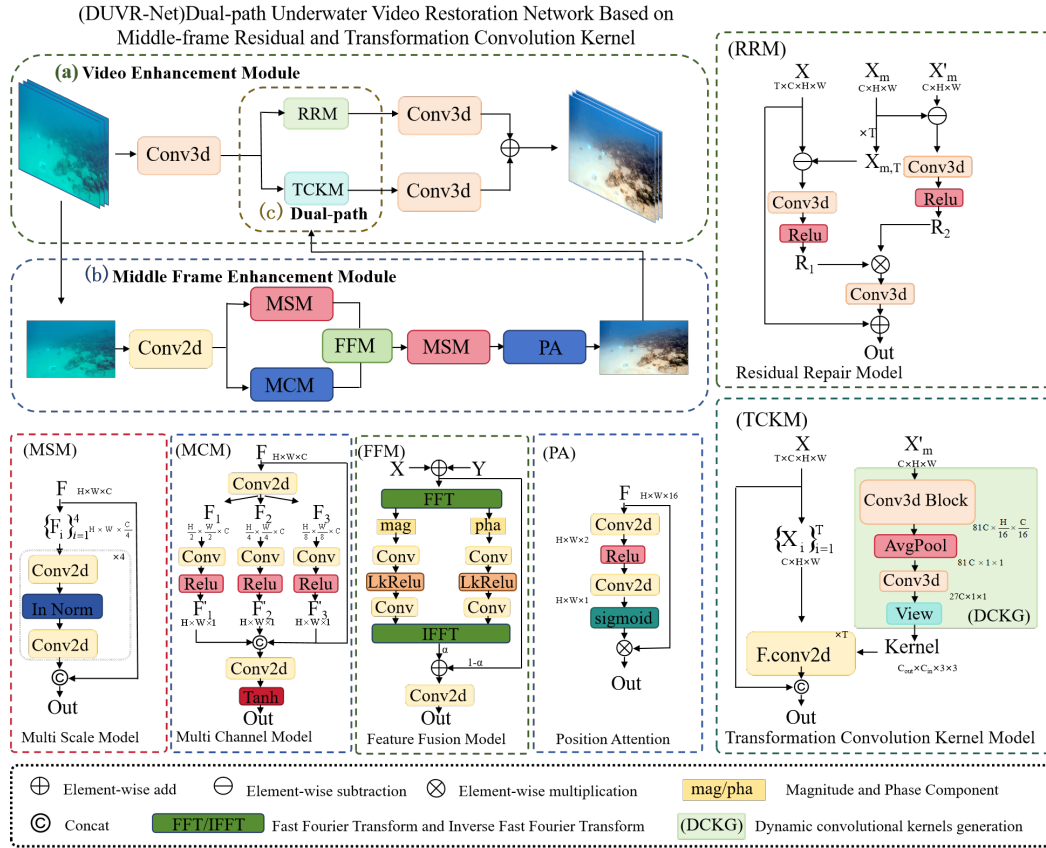
Figure 3. The overall architecture of DUVR-Net is composed of a middle frame enhancement stage and a global enhancement stage. The core components of the network comprise: **(a)** Video Enhancement Module performs the overall frame quality improvement through a dual-path mechanism; **(b)** Middle Frame Enhancement Module focuses on processing and enhancing the center frame; and **(c)** Dual-channel Information Transmission implements inter-frame information exchange through two parallel paths: residual value transmission and transformation convolution kernels, enabling effective feature propagation across frames.

3.3.1. Residual Repair Model

The Residual Modification Model (RMM) aims to extract the enhancement effects obtained from the middle frame and modify these effects according to the differences between each frame and the middle frame, subsequently applying the modified results to the input frame sequence. The RMM takes as input both a consecutive frame sequence requiring enhancement and its enhanced middle frame. By extracting, customizing, and propagating the enhancement effects instead of enhancing each frame individually, this approach significantly reduces the model's computational complexity and parameter count.

The key operations of RMM are as follows, given the feature $\{x_i \in \mathbb{X}^{3 \times H \times W}\}_{i=t-k}^{t+k}$ and the enhanced middle frame $\hat{x}_t$ derived from x_t.

Middle frame enhancement residual(R_1) extraction:

$$R_1 = \hat{x}_t - x_t \tag{2}$$

Inter-frame residual(R_2) calculation:

$$R_2 = \{x_i - x_t\}_{i=t-k}^{t+k} \tag{3}$$

Residual modulation(R_3),$\odot$ represents Element-wise multiplication:

$$R_3 = \text{expand}(R_1) \odot R_2 \tag{4}$$

Enhancement effect integration:

$$\text{Out}_{RMM} = \mathbb{X} + R_3 \tag{5}$$

In summary, the complete processing of RMM can be expressed as:

$$\text{Out}_{RMM} = \mathbb{X} + \{\text{expand}(\hat{x}_t - x_t) \odot (x_i - x_t)\}_{i=t-k}^{t+k} \tag{6}$$

3.3.2. Transformation Convolution Kernel Model

The transformation convolution kernel converts the enhanced intermediate frame into a 3×3 convolution kernel through Dynamic Convolution Kernels Generation (DCLG) and applies convolution operation to the current frame, thereby obtaining information from the enhanced intermediate frame and achieving inter-frame information transfer.The process of TCKM can be expressed as:

$$\text{Out}_{TCKM} = \text{DCLG}(\hat{x}_t) * \{x_i\}_{i=t-k}^{t+k} \tag{7}$$

where $\{x_i\}_{i=t-k}^{t+k}$ represents the features of the input consecutive frames. $\hat{x}_t$ represents the enhanced middle frame derived from x_t. $(*)$ represents convolution.

3.4. Loss Function

We calculate the loss of the middle frames and video as the total loss function $\mathcal{L}$:

$$\mathcal{L} = \mathcal{L}_{\text{Char}}(y_t, \hat{x}_t) + \mathcal{L}_{\text{Percep}}(y_t, \hat{x}_t) + \lambda \left[\mathcal{L}_{\text{Char}}\{(y_i, \hat{x}_i)\}_{i=t-k}^{t+k} + \mathcal{L}_1\{(y_i, \hat{x}_i)\}_{i=t-k}^{t+k} \right] \tag{8}$$

where $\mathcal{L}_{\text{Char}}$ refers to Charbonnier Loss [25], $\mathcal{L}_{\text{Percep}}$ refers to Perceptual Loss [26], $\mathcal{L}_1$ refers to L_1 Loss [27]. $\{y_i\}_{i=t-k}^{t+k}$ is the GT of sample $\{x_i\}_{i=t-k}^{t+k}$, and y_t is the GT of x_t. λ is a weighting parameter that controls the contribution of video-related losses, and its value increases as the number of epochs grows.

4. Results

4.1. Settings

4.1.1. Dataset

We used the large-scale high-resolution Underwater Video Enhancement Baseline (UVEB) [6] released on July 7, 2024, as our data source. The UVEB dataset comprises videos from multiple countries, encompassing various scenes and video degradation types to accommodate diverse and complex underwater environments. Due to the extensive size of the UVEB dataset, we utilized the first 150 pairs of video sequences in our experiment, consisting of 150 source videos and 150 target videos. By extracting one frame every six frames, we obtained a total of 16,836 images, including 8,418 source image sequences and 8,418 target image sequences. The ratio of training set to test set was set at 9:1, resulting in 7,563 images for training and 855 images for testing.

4.1.2. Implementation Details

Our experiments are implemented using the PyTorch [28] framework with two NVIDIA Tesla V100-SXM2-16GB GPUs. We use an ADAM optimizer for network optimization. The initial learning rate is set to 4×10^{-4}. The total number of epochs is 100. The batch size is 4, and the patch size of input video frames is 512×512, with each sequence containing either 3 or 5 frames.

4.2. Comparisons with State-of-the-Art Methods

4.2.1. Quantitative Comparison

In quantitative measurement, we use peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and mean square error (MSE) [29] as metrics for image quality. Floating point operations

(FLOPs) [30] and parameter count (Params) [31] serve as indicators of model efficiency and performance. From these quantitative results, our method outperforms others in terms of image quality. As shown in Table 1, for 4K video, our model uses the minimal computational and parameter resources, in the same order of magnitude as FA+Net, which has the smallest model size in image processing tasks. DUVRNet has only 1/350 of the parameter count of UVE-Net while improving the PSNR metric by 3.19dB, demonstrating the superiority and feasibility of this algorithm

Table 1. Experimental results on UVEB dataset.The top two results in each column are marked in red and blue, respectively.

Methods	PSNR↑	SSIM↑	MSE↓	FLOPs↓	Params↓
UDCP [32]	12.08	0.749	1.454	-	-
WaterNet [14]	16.31	0.797	1.161	193.70G	24.81M
PRW-Net [7]	20.79	0.823	0.981	223.40G	6.30M
UWNet [33]	18.28	0.782	1.314	304.75G	0.22M
PUIE [34]	22.79	0.810	0.937	423.05G	1.40M
FA+Net [16]	23.06	0.855	0.895	9.36G	0.009M
UVE-Net [6]	25.48	0.910	0.862	732.02G	33.19M
DUVRNet(Ours)	28.67	0.951	0.825	12.15G	0.096M

4.2.2. Qualitative Comparison

As shown in Figure 4, we visually compared our network enhancement results with other methods. Due to its simple network structure, PRW-Net cannot adequately restore underwater images, resulting in enhanced images with low saturation. The enhancement results presented by PRW-Net appear unnatural in color and generally too dark. In contrast, our method demonstrates superior ability in removing color distortion while restoring appropriate brightness levels.
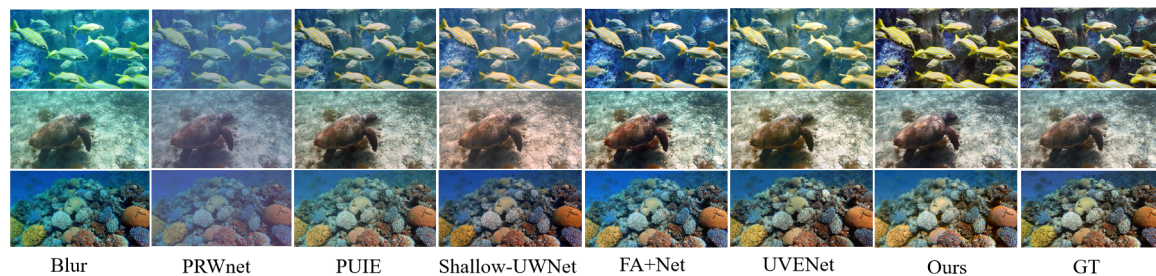

| Blur | PRWnet | PUIE | Shallow-UWNet | FA+Net | UVENet | Ours | GT |

Figure 4. Visual comparisons with state-of-the-art methods on real underwater scenes.

4.3. Ablation Studies

4.3.1. Effectiveness of Middle Frame Enhancement Module

To explore the impact of the middle frame enhancement module on the entire network, we replaced it with a residual neural network and examined the effects of this substitution. As shown in Table 2, MFEM represents the Middle Frame Enhancement Module in the model. Comparing columns (a) and (d) in the Table 2, MFEM enables the network to achieve better quality improvement with fewer parameters. Additionally, as demonstrated in Figure 5, our implemented MFEM exhibits superior capability in recovering color and brightness information in the images.

4.3.2. Effectiveness of RRM and TCKM

The results in columns (b), (c) and (d) of Table 2 verify the effectiveness of RRM and TCMK. From columns (b) and (d), we observe that the introduction of the RRM module not only reduces the MSE from 1.935 to 0.882 but also accelerates the inference time. The visual results in Figure 5 demonstrate that the absence of RRM and TCKM modules may lead to imprecise enhancement effects, resulting in artifacts such as ghosting and color distortion. The incorporation of both RRM and TCKM modules contributes to faster and better enhancement performance.
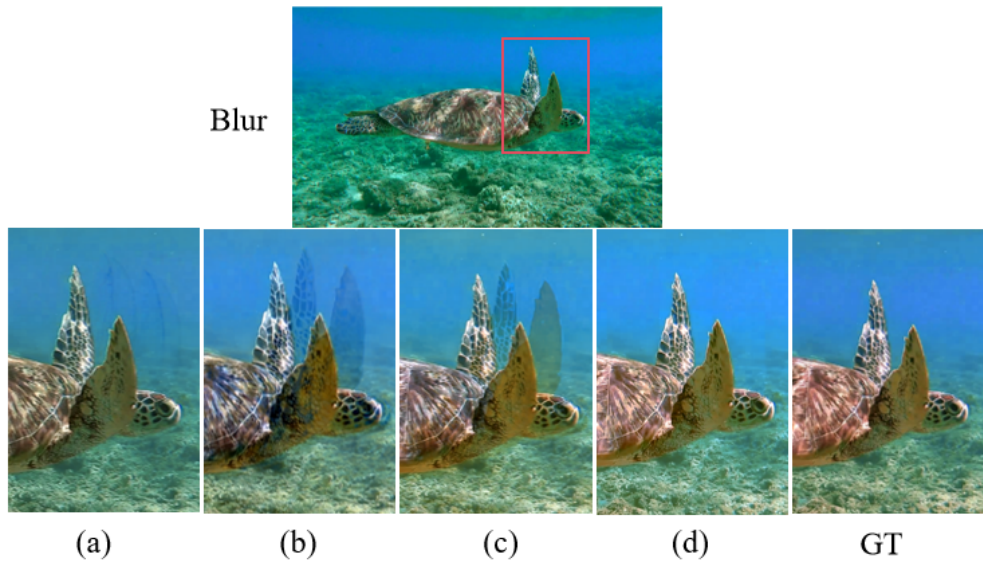
Figure 5. Visual comparison of ablation results.

Table 2. Ablation studies of major components.

	(a)	(b)	(c)	(d)
MFEM		✓	✓	✓
RRM	✓		✓	✓
TCKM	✓	✓		✓
PSNR	25.00	25.11	27.12	28.67
MSE	1.979	1.935	1.239	0.825
FLOPs	10.34	13.05	12.79	12.156
Params(G)	5.48	10.18	11.04	0.096

5. Discussion

Compared to traditional frame-by-frame processing methods, DUVR-Net not only has a smaller model parameter count but also demonstrates the capability to process 4K videos. The model shows considerable potential in practical applications such as marine research and underwater archaeology. It can provide high-quality input for downstream tasks like underwater object detection.Although DUVR-Net is specifically tailored for underwater videos, its network architecture is still based on image enhancement networks. The dual-channel mechanism primarily functions to propagate image enhancement effects to adjacent frames. Therefore, in specific scenarios (such as scenes with intense motion), there might be certain limitations. Future work will focus on incorporating other advanced technologies (such as adaptive parameter adjustment) to make the model more adaptable to multi-environmental tasks. Additionally, attention will be paid to optimizing the enhancement strategy for intermediate frames.

6. Conclusions

In this paper, we present DUVR-Net, a lightweight model for 4K underwater video enhancement that aims to overcome the limitations of previous frame-by-frame enhancement methods. Our approach employs a dual-channel strategy that effectively reduces computational complexity by separately transmitting enhanced color content with attention mechanisms and restored details. DUVR-Net currently has the smallest parameter count among underwater video enhancement models, while experiments validate its superior performance in underwater video enhancement tasks. We believe that DUVR-Net provides new design concepts and directions for underwater video enhancement, with practical applications in marine research and underwater archaeology.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be available upon request.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

UVEB	Underwater Video Enhancement Benchmark
RRM	Residual Repair Model
TCKM	Transformation Convolution Kernel Model
MSM	Multi Scale Model
MCM	Multi Channel Model
FFM	Feature Fusion Model
PA	Position Attention
DCLG	Dynamic Convolution Kernels Generation
MFEM	Middle Frame Enhancement Module

References

1. Li, J.; Xu, W.; Deng, L.; Xiao, Y.; Han, Z.; Zheng, H. Deep Learning for Visual Recognition and Detection of Aquatic Animals: A Review. *Rev. Aquacult.* **2023**, *15*, 409–433.
2. Guyot, A.; Lennon, M.; Thomas, N.; Gueguen, S.; Petit, T.; Lorho, T.; Hubert-Moy, L. Airborne Hyperspectral Imaging for Submerged Archaeological Mapping in Shallow Water Environments. *Remote Sens.* **2019**, *11*, 2237.
3. Lee, D.; Kim, G.; Kim, D.; Myung, H.; Choi, H. T. Vision-Based Object Detection and Tracking for Autonomous Navigation of Underwater Robots. *Ocean Eng.* **2012**, *48*, 59–68.
4. Bell, K. L.; Chow, J. S.; Hope, A.; Quinzin, M. C.; Cantner, K. A.; Amon, D. J.; McClain, C. R. Low-Cost, Deep-Sea Imaging and Analysis Tools for Deep-Sea Exploration: A Collaborative Design Study. *Front. Mar. Sci.* **2022**, *9*, 873700.
5. Zhou, J.; Yang, T.; Zhang, W. Underwater Vision Enhancement Technologies: A Comprehensive Review, Challenges, and Recent Trends. *Appl. Intell.* **2023**, *53*, 3594–3621.
6. Xie, Y.; Kong, L.; Chen, K.; Zheng, Z.; Yu, X.; Yu, Z.; Zheng, B. Uveb: A Large-Scale Benchmark and Baseline Towards Real-World Underwater Video Enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; **2024**; pp 22358–22367.
7. Huo, F.; Li, B.; Zhu, X. Efficient Wavelet Boost Learning-Based Multi-Stage Progressive Refinement Network for Underwater Image Enhancement. In Proceedings of the IEEE/CVF International Conference on Computer Vision; **2021**; pp 1944–1952.
8. Li, C.; Hu, E.; Zhang, X.; Zhou, H.; Xiong, H.; Liu, Y. Visibility Restoration for Real-World Hazy Images via Improved Physical Model and Gaussian Total Variation. *Front. Comput. Sci.* **2024**, *18*, 181708.
9. Liu, Y.; Yan, Z.; Tan, J.; Li, Y. Multi-Purpose Oriented Single Nighttime Image Haze Removal Based on Unified Variational Retinex Model. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *33*, 1643–1657.
10. He, K.; Sun, J.; Tang, X. Single Image Haze Removal Using Dark Channel Prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 2341–2353.

11. Ghani, A. S. A.; Isa, N. A. M. Underwater Image Quality Enhancement Through Integrated Color Model with Rayleigh Distribution. *Appl. Soft Comput.* **2015**, *27*, 219–230.

12. Li, C.; Guo, C.; Ren, W.; et al. An Underwater Image Enhancement Benchmark Dataset and Beyond. *IEEE Trans. Image Process.* **2019**, *29*, 4376–4389.

13. Ferrara, E. Fast Implementations of LMS Adaptive Filters. *IEEE Trans. Acoust., Speech, Signal Process.* **1980**, *28*, 474–475.

14. Krizhevsky, A.; Sutskever, I.; Hinton, G. E. Imagenet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*.

15. Li, C.; Anwar, S.; Hou, J.; Cong, R.; Guo, C.; Ren, W. Underwater Image Enhancement via Medium Transmission-Guided Multi-Color Space Embedding. *IEEE Trans. Image Process.* **2021**, *30*, 4985–5000.

16. Jiang, J.; Ye, T.; Bai, J.; et al. Five A+ Network: You Only Need 9K Parameters for Underwater Image Enhancement. arXiv preprint arXiv:2305.08824. **2023**.

17. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Polosukhin, I. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.

18. Peng, L.; Zhu, C.; Bian, L. U-Shape Transformer for Underwater Image Enhancement. *IEEE Trans. Image Process.* **2023**, *32*, 3066–3079.

19. Han, R.; Guan, Y.; Yu, Z.; Liu, P.; Zheng, H. Underwater Image Enhancement Based on a Spiral Generative Adversarial Framework. *IEEE Access* **2020**, *8*, 218838–218852.

20. Ancuti, C. O.; Ancuti, C.; De Vleeschouwer, C.; Bekaert, P. Color Balance and Fusion for Underwater Image Enhancement. *IEEE Trans. Image Process.* **2017**, *27*, 379–393.

21. Li, C.; Anwar, S.; Porikli, F. Underwater Scene Prior Inspired Deep Underwater Image and Video Enhancement. *Pattern Recognit.* **2020**, *98*, 107038.

22. Qing, C.; Yu, F.; Xu, X.; Huang, W.; Jin, J. Underwater Video Dehazing Based on Spatial–Temporal Information Fusion. *Multidimens. Syst. Signal Process.* **2016**, *27*, 909–924.

23. Li, D.; Shi, X.; Zhang, Y.; Cheung, K. C.; See, S.; Wang, X.; Li, H. A Simple Baseline for Video Restoration with Grouped Spatial-Temporal Shift. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; **2023**; pp 9822–9832.

24. Frigo, M.; Johnson, S. G. FFTW: An Adaptive Software Architecture for the FFT. In Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98; **1998**; Vol. 3, pp 1381–1384.

25. Sun, D.; Roth, S.; Black, M. J. A Quantitative Analysis of Current Practices in Optical Flow Estimation and the Principles Behind Them. *Int. J. Comput. Vis.* **2014**, *106*, 115–137.

26. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In Computer Vision–ECCV 2016: 14th European Conference; **2016**; pp 694–711.

27. Zhao, H.; Gallo, O.; Frosio, I.; Kautz, J. Loss Functions for Image Restoration with Neural Networks. *IEEE Trans. Comput. Imaging* **2016**, *3*, 47–57.

28. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Chintala, S. Pytorch: An Imperative Style, High-Performance Deep Learning Library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.

29. Wang, Z.; Bovik, A. C. Mean Squared Error: Love It or Leave It? A New Look at Signal Fidelity Measures. *IEEE Signal Process. Mag.* **2009**, *26*, 98–117.

30. Howard, A. G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Adam, H. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint arXiv:1704.04861. **2017**.

31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; **2016**; pp 770–778.

32. Drews, P.; Nascimento, E.; Moraes, F.; Botelho, S.; Campos, M. Transmission Estimation in Underwater Single Images. In Proceedings of the IEEE International Conference on Computer Vision Workshops; **2013**; pp 825–830.

33. Naik, A.; Swarnakar, A.; Mittal, K. Shallow-UWNet: Compressed Model for Underwater Image Enhancement (Student Abstract). In Proceedings of the AAAI Conference on Artificial Intelligence; **2021**; Vol. 35, pp 15853–15854.

34. Fu, Z.; Wang, W.; Huang, Y.; Ding, X.; Ma, K. K. Uncertainty Inspired Underwater Image Enhancement. In European Conference on Computer Vision; **2022**; pp 465–482.