LW-UVENet: An Efficient and Lightweight Underwater Video Enhancement Network

1st Xinnan Fan
College of Information Science and Engineering
Hohai University
Changzhou, China
fanxn@hhu.edu.cn

2nd Qi Sun
College of Information Science and Engineering
Hohai University
Changzhou, China
231623010054@hhu.edu.cn

3rd Yuanxue Xin
College of Information Science and Engineering
Hohai University
Changzhou, China
xinyx@hhu.edu.cn

4th Pengfei Shi*
College of Artificial Intelligence and Automation
Hohai University
Changzhou, China
shipf@hhu.edu.cn

***Abstract*—Underwater Video Enhancement (UVE) aims to alleviate color distortion, low contrast, haze-like degradation, and motion blur caused by light absorption and scattering in water. Existing end-to-end UVE models, such as UVENet, already demonstrate the benefit of temporal modeling, but their computational cost and limited deblurring capability still restrict deployment on resource-constrained underwater robotic terminals. To address this practical bottleneck, we present LW-UVENet, a deployment-oriented lightweight redesign of UVENet for underwater video enhancement. The proposed network combines a Lightweight ConvNeXt-based Encoder, an Underwater Deblurring Module (UDM), an Enhanced Feature Alignment and Aggregation Module (Enhanced FAAM), a Lightweight Decoder, and a Lightweight Global Recovery Module (LW-GRM). Rather than introducing a completely new alignment paradigm, our goal is to integrate low-cost yet task-effective components to improve the quality–efficiency trade-off in underwater video enhancement. Experiments on the synthetic SUVE dataset and the real-world MVK dataset show that LW-UVENet maintains competitive and in some cases superior enhancement quality than the original UVENet, while reducing single-frame inference time from 0.241s to 0.111s and peak memory usage from 19.25G to 10.26G. These results indicate that LW-UVENet makes underwater video enhancement more practical as a front-end module for resource-limited underwater robots and other embodied visual systems.**

***Index Terms*—Underwater Video Enhancement, Lightweight Network, Deblurring, Temporal Feature Alignment**

I. Introduction

Underwater visual perception is fundamental to marine resource exploration, underwater archaeology, environmental monitoring, and robot navigation [1], [2]. However, underwater videos usually suffer from severe quality degradation. Different wavelengths attenuate at different rates, with red light vanishing fastest, which causes obvious blue-green color casts [3], [4]. At the same time, suspended particles scatter light, reducing contrast and introducing haze-like blur [5]. Camera motion and object motion further aggravate temporal instability.

Most prior work focuses on Underwater Image Enhancement (UIE), including physical-model inversion and deep learning methods based on CNNs [6], GANs [7], and Transformers [8]. These methods enhance single images effectively, but direct frame-by-frame use in videos often causes flicker and cannot exploit temporal cues. UVENet [9] is the first end-to-end model designed specifically for UVE and demonstrates the importance of temporal alignment and aggregation. Yet its standard ConvNeXt encoder, heavy decoder, and global recovery path still result in high computational cost, and it lacks a dedicated module for underwater deblurring.

To address these limitations, we propose LW-UVENet, a deployment-oriented and lightweight redesign of UVENet. The key idea is not to introduce a completely new alignment paradigm, but to preserve the effective temporal framework of UVENet while systematically replacing expensive components with task-aware lightweight modules. Our main contributions are summarized as follows:

- We redesign the encoder with lightweight ConvNeXt blocks, Efficient Channel Attention (ECA), and an underwater deblurring module to improve robustness to blur under tight computational budgets.
- We develop an enhanced FAAM based on grouped spatial shift, efficient feature fusion, and serial channel-spatial attention so that temporal information can be utilized without optical flow.
- We reconstruct the decoder and global recovery path with low-cost operators, enabling efficient image reconstruction and residual global color correction for deployment-oriented underwater video enhancement.

* Corresponding author.

This work is supported by the National Natural Science Foundation of China (Grant No. 62476080) and the National Key R&D Program of China (Grant/Award No. 2022YFB4703400).

II. Related Work

A. Underwater Image Enhancement

UIE methods can be broadly divided into physical-model-based and non-physical-model-based approaches [10]. Physical methods restore images by estimating background light and transmission maps [11], [12], but underwater optical parameters are often spatially non-uniform and difficult to estimate accurately [13]. Non-physical methods directly optimize pixels or features. Traditional techniques such as MLLE improve color and local contrast but are prone to distortion in complex scenes [14]. Deep learning methods, including UGAN [15], Fusion-based enhancement [16], and NU^{2}Net [17], provide stronger restoration quality and perceptual consistency. However, most of them are designed for single images and do not explicitly model temporal consistency.

B. Underwater Video Enhancement and Lightweight Video Restoration

Earlier UVE solutions typically enhanced each frame independently and then applied temporal smoothing [18], [19]. Such two-stage strategies are not end-to-end and cannot fully exploit inter-frame information. Methods for video dehazing and spatio-temporal fusion partly inspired UVE research [20], [21], while UVENet [9] established an end-to-end benchmark for underwater video enhancement. In the broader video restoration field, methods such as BasicVSR++ and VRT achieve strong performance but often rely on optical flow, deformable alignment, or heavy propagation [22], [23]. These designs are less suitable for underwater robots with limited onboard resources. Shift-based alignment offers a low-cost alternative [24], and MobileNet together with ConvNeXt provide useful principles for lightweight backbone design [25], [26].

III. Method

A. LW-UVENet

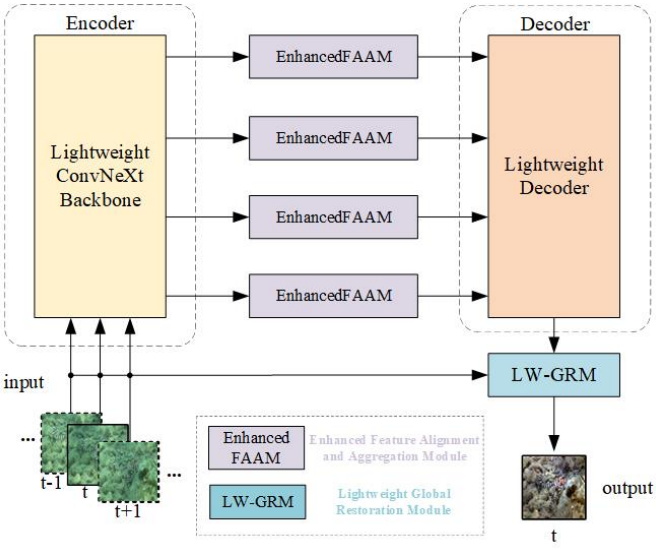

Fig. 1. Overall Architecture of LW-UVENet.

LW-UVENet keeps the multi-frame processing mode of UVENet. Given an input sequence of T underwater frames, the network predicts an enhanced center frame. As shown in Fig. 1, it consists of a lightweight encoder, an enhanced FAAM, a lightweight decoder, and an LW-GRM.

B. Lightweight ConvNeXt Encoder and UDM

The encoder is built on ConvNeXt but replaces the standard large-kernel block with a lightweight attention block. Specifically, 5×5 depthwise separable convolution is used instead of standard convolution to reduce redundancy, and ECA [27] is inserted after normalization to emphasize informative channels. In plain terms, each block first extracts spatial features efficiently, then reweights channels according to global context, and finally refines them through pointwise projection with a residual connection. Compared with squeeze-and-excitation style dimensionality reduction [28], ECA preserves local cross-channel interaction with negligible extra parameters.

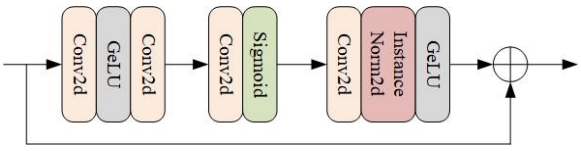

Fig. 2. Architecture of the Underwater Deblurring Module.

To strengthen blur handling, an Underwater Deblurring Module is added after each encoder stage, as shown in Fig. 2. Instead of performing an explicit Fourier transform, the module uses stacked convolutions to approximate frequency decomposition and highlight potentially high-frequency details. A spatial attention mask then suppresses noisy responses and emphasizes texture-rich regions. The reweighted features are normalized, activated, and added back to the input in a residual manner. This module improves recovery of blurred edges and fine structures while keeping the computation lightweight.

C. Enhanced Feature Alignment and Aggregation Module

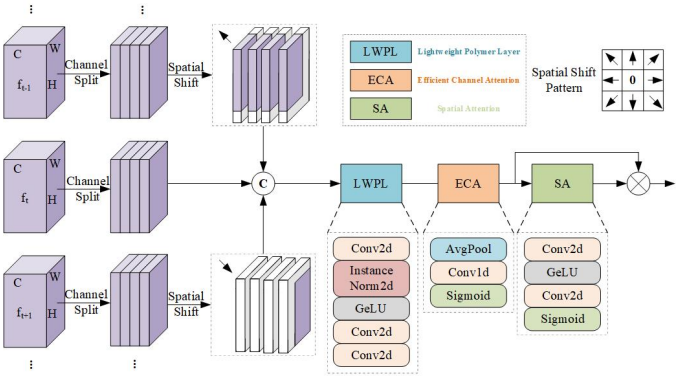

Fig. 3. Architecture of the Enhanced Feature Alignment and Aggregation Module.

The main challenge in underwater video enhancement is aligning non-rigid temporal information without expensive

motion estimation [29]. The enhanced FAAM in Fig. 3 adopts grouped spatial shift as a lightweight approximation to alignment. Relative to the center frame, neighboring frame features are shifted progressively according to temporal distance. Positive offsets move features toward the bottom-right and negative offsets move them toward the top-left. This simple strategy expands the temporal receptive field at nearly zero cost and is especially suitable for underwater videos dominated by camera translation, mild jitter, viewpoint drift, and small object motion.

After alignment, features from all frames are concatenated and fused by depthwise separable convolution. The fused representation is then refined by serial channel attention and spatial attention. In other words, the module first selects important channels, then suppresses background noise and misaligned regions spatially, and finally adds the refined result to the center-frame feature through a residual path [30]. This design compensates for the limited expressiveness of fixed shift alignment and preserves useful temporal cues without optical flow.

D. Lightweight Decoder and LW-GRM

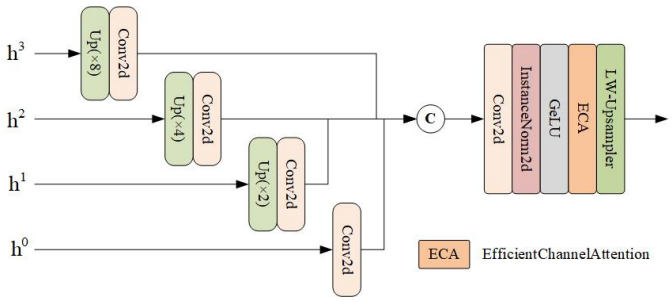

Fig. 4. Architecture of the Lightweight Decoder Module.

As shown in Fig. 4, the lightweight decoder reconstructs a high-resolution image from multi-scale aligned features [31]. Features from the four scales are first resized to the highest spatial resolution and compressed in channel dimension. They are then concatenated and fused using a compact 1×1 convolution, instance normalization, GELU, and ECA. Instead of standard progressive transposed convolutions, the decoder uses two-stage sub-pixel upsampling (PixelShuffle) [32]. This strategy restores image resolution efficiently and avoids the smoothing artifacts often observed with heavy deconvolution. A final lightweight convolutional refinement stage produces a preliminary reconstruction.

Because underwater videos often retain global color cast and brightness attenuation even after local restoration [33], we propose a Lightweight Global Recovery Module inspired by Gray World correction [34]. As shown in Fig. 5, the original frame sequence and the preliminary reconstruction are concatenated and passed through a multi-scale feature extractor using 3×3, 5×5, and 7×7 branches [35]. These features are fused and globally pooled to predict a residual RGB scaling vector. The final output is obtained by applying

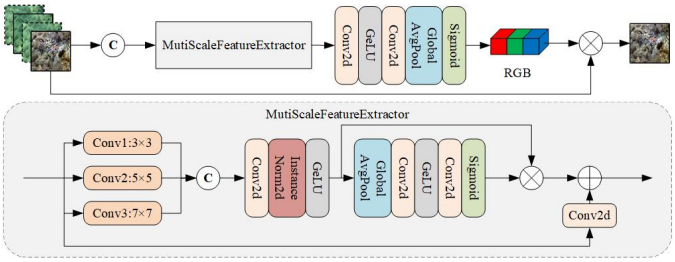

Fig. 5. Architecture of the Lightweight Global Restoration Module.

this global color calibration to the reconstructed image. Since local degradation has already been addressed upstream, such a lightweight residual correction is sufficient in many underwater scenes while adding little overhead.

IV. Experiments

A. Experimental Setup

Training uses the paired synthetic SUVE dataset [9], which contains 660 training video pairs and 180 test video pairs across diverse underwater conditions. Real-scene inference is evaluated on the MVK dataset [36]. Following the original setting, 45 challenging real videos from 36 locations are selected and sampled at 5 frames per second to form a real test set of more than 7,000 frames.

Frame-level evaluation includes PSNR, SSIM, UIQM [37], and UCIQE [38]. Video-level evaluation uses brightness flicker error based on MSE of the Mean Absolute Brightness Difference (MABD) vectors [39] and the Color Distribution Consistency (CDC) index [40]. Efficiency is measured by average single-frame inference time and peak memory during inference. The model is trained with Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay 10^{-4}), an initial learning rate of 4×10^{-4} with cosine annealing, an L_1 loss, and batch size 16. Experiments are conducted in PyTorch 2.7.1 with CUDA 12.9 on an NVIDIA GeForce RTX 5090 GPU.

B. Comparative Experiments

To fully evaluate the performance of the proposed LW-UVENet, we compared it with various mainstream methods, including physical model methods, single-frame enhancement deep learning methods, and the original UVENet. All comparison methods were retrained or fine-tuned on the same SUVE dataset to ensure fairness.

Table I reports the results on SUVE. LW-UVENet preserves competitive or slightly better enhancement quality than UVENet while exhibiting a markedly stronger deployment profile. The gain in PSNR over UVENet is relatively modest (0.08 dB), indicating that the main contribution of the proposed model is not a dramatic fidelity increase, but the preservation of restoration quality under much lower computational cost. Specifically, LW-UVENet improves SSIM from 0.908 to 0.911 and reduces MSE(MABD) from 0.86 to 0.81, which suggests that the lightweight redesign does not compromise structural fidelity or brightness stability.

TABLE I
COMPARISON RESULTS ON THE SUVE DATASET. IN THE "MEMORY" COLUMN, * INDICATES CPU USAGE, WHILE ENTRIES WITHOUT * CORRESPOND TO GPU USAGE. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Methods	PSNR↑	SSIM↑	UCIQE↑	UIQM↑	MSE(MABD)↓	CDC↓	Inference Times(s)	Memory(G)
Fusion [16]	17.08	0.781	**0.598**	2.796	25.22	0.0114	5.932	0.91*
UGAN [15]	20.03	0.812	0.578	**3.108**	16.45	0.0125	0.341	2.03*
MLLE [14]	16.98	0.678	0.589	1.998	177.64	**0.0086**	7.261	1.22*
NU^{2}Net [17]	26.55	0.904	0.575	2.743	1.69	0.0145	0.752	1.58*
UVENet [9]	27.53	0.908	0.587	2.549	0.86	0.0127	0.241	19.25
Ours	**27.61**	**0.911**	0.589	2.758	**0.81**	0.0124	**0.111**	**10.26**

TABLE II
COMPARISON RESULTS ON THE MVK DATASET. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Methods	UCIQE↑	UIQM↑	CDC↓
Fusion [16]	0.613	2.779	0.0147
UGAN [15]	0.556	**3.028**	0.0378
MLLE [14]	0.587	2.101	0.0125
NU^{2}Net [17]	0.562	2.818	0.0205
UVENet [9]	0.601	2.806	**0.0118**
Ours	**0.615**	2.912	0.0141

More importantly, latency is reduced from 0.241s/frame to 0.111s/frame, corresponding to about a 53.9% reduction, while memory usage drops from 19.25G to 10.26G, i.e., about 46.7% lower than UVENet. These results indicate a substantially improved quality–efficiency trade-off for deployment on resource-constrained underwater platforms. Compared with frame-by-frame methods such as Fusion and MLLE, LW-UVENet also achieves far lower flicker-related error, with MSE(MABD) reduced to 0.81, confirming the benefit of temporal modeling.

Table II shows the results on real underwater videos. LW-UVENet achieves the best UCIQE and a competitive UIQM, indicating stronger color correction and a favorable balance between contrast and clarity in real scenes. In particular, the UCIQE score reaches 0.615, which is higher than UVENet (0.601) by about 2.3%, suggesting better correction of underwater color cast in real-world videos. The UIQM score of 2.912 also surpasses UVENet (2.806) and NU^{2}Net (2.818), showing that the proposed method maintains good perceptual quality beyond the synthetic benchmark. Its CDC is slightly higher than that of UVENet but remains much lower than that of UGAN, indicating that temporal color transitions remain stable overall. Although MLLE obtains a slightly lower CDC value, its substantially inferior UIQM (2.101) and the visual results suggest that this apparent temporal smoothness is achieved at the cost of detail preservation and enhancement quality. These observations support the robustness of the proposed lightweight redesign under real underwater conditions.

The qualitative comparison in Fig. 6 leads to the same conclusion. Raw frames exhibit severe bluish-green cast and blurred details. Fusion introduces visible distortion, UGAN suffers from obvious temporal brightness inconsistency, and MLLE loses details. In contrast, LW-UVENet better restores coral textures and fish contours while maintaining natural transitions from Frame 3 to Frame 43.

C. Ablation Analysis and Discussion

Table III shows that different modules contribute differently. Replacing the original ConvNeXt block with the lightweight block improves both accuracy and efficiency. Adding UDM further enhances deblurring and frame fidelity, though it introduces a small memory increase compared with using the lightweight block alone. Enhanced FAAM mainly improves temporal aggregation and alignment quality. The decisive deployment gain appears after the decoder and LW-GRM are also redesigned, showing that the practical efficiency improvement comes from full-link co-design rather than from any single module alone. The final model greatly reduces runtime and memory while preserving strong restoration performance.

The experiments suggest several practical implications. First, the proposed method should be understood as a quality-preserving lightweight redesign rather than a purely accuracy-driven replacement. On SUVE, the gains in PSNR and SSIM over UVENet are small, but the reduction in latency and memory is substantial. This is important in underwater operation because video enhancement is rarely the final task. Instead, it is usually a front-end stage before detection, tracking, mapping, measurement, or robot control. In such pipelines, a moderate reduction in enhancement cost can translate into a larger system-level benefit because more resources remain available for downstream perception and decision-making.

Second, the ablation results show that lightweighting underwater video enhancement cannot rely on only one local substitution. The lightweight block improves efficiency early, but the final deployment gain emerges only when the encoder, temporal aggregation, decoder, and global correction path are redesigned together. This observation explains why some intermediate variants achieve higher PSNR than the final model while still failing to deliver the best practical trade-off. In other words, the objective of LW-UVENet is not to maximize one benchmark metric in isolation, but to reach a more balanced operating point for real deployment.

Third, the proposed temporal alignment strategy reflects an explicit design choice. Grouped spatial shift is less expressive than optical-flow-based or deformable alignment, especially under severe non-rigid motion. However, underwater robotic video often contains dominant camera translation, mild viewpoint drift, and local object motion mixed with blur and scattering. Under this condition, a fixed low-cost alignment

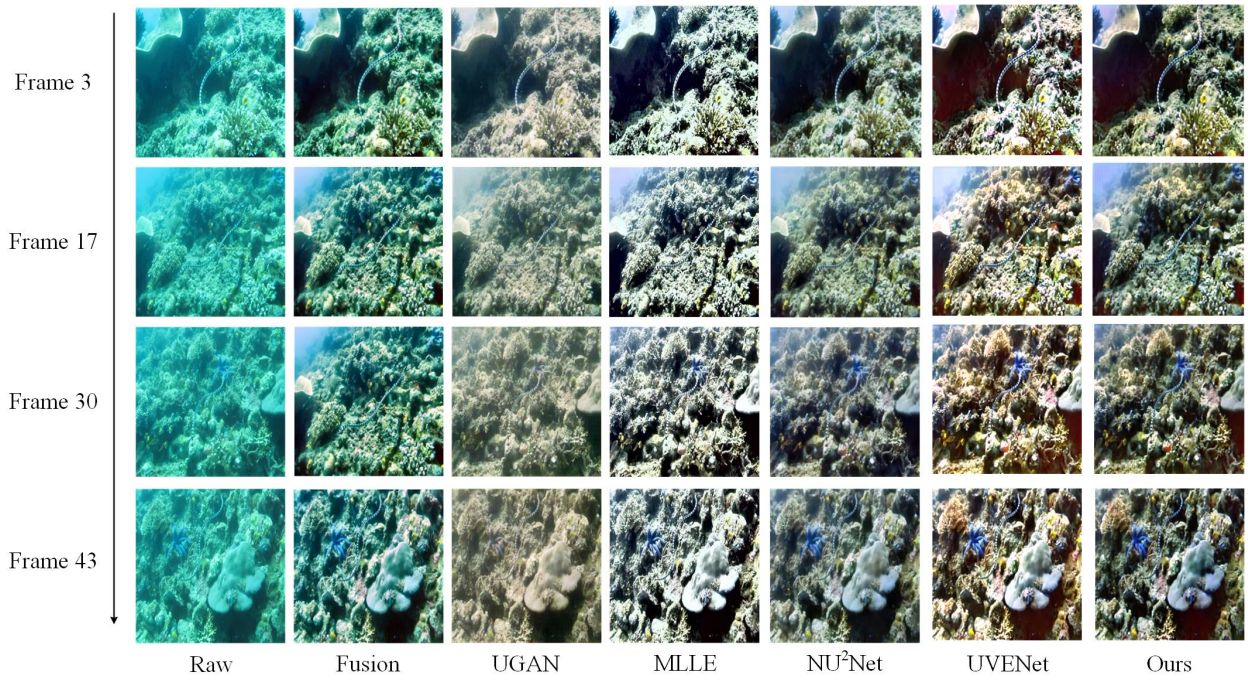

Fig. 6. Qualitative comparison results on the MVK dataset.

TABLE III
ABLATION STUDY RESULTS. IN THE MODEL DESCRIPTION, A DENOTES LW-BLOCK, B DENOTES UDM, C DENOTES ENHANCED FAAM, D DENOTES LW-DECODER, AND E DENOTES LW-GRM. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Experiment ID	Model Description	PSNR↑	SSIM↑	UCIQE↑	UIQM↑	MSE(MABD)↓	CDC↓	Inference Times(s)	Memory(G)
0	UVENet(Baseline)	27.54	0.908	0.587	2.549	0.86	0.0127	0.221	19.25
1	w/ A	27.98	0.922	0.588	2.579	0.83	0.0124	0.172	17.01
2	w/ A+B	28.35	0.925	0.588	2.537	**0.81**	0.0139	0.221	17.16
3	w/ A+B+C	**28.39**	**0.926**	0.588	2.534	0.86	0.0126	0.215	17.31
4	w/ A+B+C+D	28.17	0.925	0.587	2.539	0.88	0.0139	0.219	17.31
5	w/ A+B+C+D+E	27.61	0.911	**0.589**	**2.758**	1.05	**0.0124**	**0.111**	**10.26**

prior, followed by efficient fusion and serial attention, is often sufficient to recover useful temporal information. The results on MVK indicate that this approximation remains effective in real scenes, even though it does not attempt to model every motion pattern explicitly.

The behavior of LW-GRM is also worth noting. The final model does not always achieve the absolute best full-reference metric among the ablation variants, but it attains the highest UCIQE and UIQM in the final configuration. This suggests that lightweight global residual color calibration is particularly valuable for real underwater enhancement, where overall color cast and brightness attenuation strongly affect perceptual quality. Thus, the model balances local restoration and global correction instead of over-optimizing for only pixel-level similarity on synthetic data.

The current study still has clear limitations. Power consumption on embedded devices was not measured directly, and human subjective evaluation was not conducted. Statistical significance testing was also not included. In addition, although MVK evaluation indicates useful real-scene generalization, the synthetic-to-real gap remains an open issue, especially for water types, lighting conditions, and motion patterns not well represented in SUVE. These limitations do not change the main conclusion of this work, but they define the next steps needed before full deployment in long-duration underwater missions.

V. CONCLUSION AND FUTURE WORK

This paper presents LW-UVENet, a lightweight and deployment-oriented redesign of UVENet for underwater video enhancement. By combining lightweight backbone transformation, explicit underwater deblurring, flow-free temporal aggregation, and a lightweight reconstruction path, the proposed method achieves a more favorable balance between enhancement quality and computational efficiency. Experimental results on SUVE and MVK show that LW-UVENet preserves competitive restoration quality while substantially reducing inference latency and memory usage relative to UVENet, making it more suitable for resource-constrained underwater robotic systems. Nevertheless, the current study still has several limitations: it has not yet included embedded power measurements, human perceptual evaluation, or statis-

tical significance testing, and the synthetic-to-real domain gap has not been explicitly addressed by adaptation techniques. These issues will be investigated in future work together with quantization and more hardware-aware deployment studies.

REFERENCES

[1] X. Ji, X. Wang, L.-Y. Hao, and C.-T. Cai, "Cfenet: Cost-effective underwater image enhancement network via cascaded feature extraction," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108561, 2024.

[2] D. Akkaynak and T. Treibitz, "Sea-thru: A method for removing water from underwater images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 1682–1691.

[3] A. S. A. Ghani and N. A. M. Isa, "Enhancement of low quality underwater image through integrated global and local contrast correction," *Applied Soft Computing*, vol. 37, pp. 332–344, 2015.

[4] W. Zhang, L. Zhou, P. Zhuang, G. Li, X. Pan, W. Zhao, and C. Li, "Underwater image enhancement via weighted wavelet visual perception fusion," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 4, pp. 2469–2483, 2023.

[5] Z. Wang, M. Hu, and K. Zhang, "Underwater turbid media stokes-based polarimetric recovery," *Sensors*, vol. 24, no. 5, p. 1367, 2024.

[6] L. O. Chua and T. Roska, "The cnn paradigm," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 40, no. 3, pp. 147–156, 2002.

[7] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[9] D. Du, E. Li, L. Si, F. Xu, and J. Niu, "End-to-end underwater video enhancement: Dataset and model," *arXiv preprint arXiv:2403.11506*, 2024.

[10] D. Du, E. Li, L. Si, W. Zhai, F. Xu, J. Niu, and F. Sun, "Uiedp: Boosting underwater image enhancement with diffusion prior," *Expert Systems with Applications*, vol. 259, p. 125271, 2025.

[11] S. Cheng, Z. Jin, X. Wu, and J. Liang, "Transmission map and background light guided enhancement of unpaired underwater image," *Neurocomputing*, vol. 621, p. 129270, 2025.

[12] 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.

[13] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, "Automatic red-channel underwater image restoration," *Journal of Visual Communication and Image Representation*, vol. 26, pp. 132–145, 2015.

[14] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, and C. Li, "Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement," *IEEE Transactions on Image Processing*, vol. 31, pp. 3997–4010, 2022.

[15] 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.

[16] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 81–88.

[17] C. Guo, R. Wu, X. Jin, L. Han, W. Zhang, Z. Chai, and C. Li, "Underwater ranker: Learn which is better and how to be better," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 1, 2023, pp. 702–709.

[18] K. Hu, C. Weng, Y. Zhang, J. Jin, and Q. Xia, "An overview of underwater vision enhancement: From traditional methods to recent deep learning," *Journal of Marine Science and Engineering*, vol. 10, no. 2, p. 241, 2022.

[19] C. Tang, U. F. von Lukas, M. Vahl, S. Wang, Y. Wang, and M. Tan, "Efficient underwater image and video enhancement based on retinex," *Signal, Image and Video Processing*, vol. 13, no. 5, pp. 1011–1018, 2019.

[20] Z. Li, P. Tan, R. T. Tan, D. Zou, S. Zhiying Zhou, and L.-F. Cheong, "Simultaneous video defogging and stereo reconstruction," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 4988–4997.

[21] C. Qing, F. Yu, X. Xu, W. Huang, and J. Jin, "Underwater video dehazing based on spatial–temporal information fusion," *Multidimensional Systems and Signal Processing*, vol. 27, no. 4, pp. 909–924, 2016.

[22] K. C. Chan, S. Zhou, X. Xu, and C. C. Loy, "Basicvsr++: Improving video super-resolution with enhanced propagation and alignment," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5972–5981.

[23] J. Liang, J. Cao, Y. Fan, K. Zhang, R. Ranjan, Y. Li, R. Timofte, and L. Van Gool, "Vrt: A video restoration transformer," *IEEE Transactions on Image Processing*, vol. 33, pp. 2171–2182, 2024.

[24] D. Li, X. Shi, Y. Zhang, K. C. Cheung, S. See, X. Wang, H. Qin, and H. Li, "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.

[25] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[26] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 11 976–11 986.

[27] S. Girepunje and P. Singh, "Ecam-net: A lightweight convolutional neural network enhanced with efficient channel attention for lung cancer classification," *Procedia Computer Science*, vol. 260, pp. 126–133, 2025.

[28] H. Bao and J. Ma, "A novel frame funnel-shaped cnn with se block for epilepsy detection in eeg signals," in *2025 10th International Conference on Intelligent Computing and Signal Processing (ICSP)*. IEEE, 2025, pp. 471–477.

[29] K. Hu, Y. Meng, Z. Liao, L. Tang, and X. Ye, "Enhancing underwater video from consecutive frames while preserving temporal consistency," *Journal of Marine Science and Engineering*, vol. 13, no. 1, 2025.

[30] H. Liang, R. Yang, H. Peng, and Y. Yang, "Multi-scale attention lightweight network for underwater schooling fish detection and behavioral analysis," in *Proceedings of the 2nd Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Digital Economy and Artificial Intelligence*, 2025, pp. 233–241.

[31] S. Fang and H. Gan, "Unfolding-associative encoder-decoder network with progressive alignment for pansharpening," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 13 651–13 661.

[32] D. Choi, Y. No, J. Lee, and D. Kim, "Fourier-guided attention upsampling for image super-resolution," *arXiv preprint arXiv:2508.10616*, 2025.

[33] G. Luo, H. Li, H. Wang, H. Sui, X. Zhang, R. Zhang, B. Chen, and Z. Jiang, "A lightweight underwater image and video enhancement method based on multi-scale feature fusion," *Frontiers in Marine Science*, vol. 12, p. 1725829, 2025.

[34] W. WU and M. YAO, "Optimization of underwater images based on gray world algorithm and jaffe-mcglamery models," *FRONTIERS*, vol. 11, no. 1, pp. 80–84, 2025.

[35] X. Lu, M. Suganuma, and T. Okatani, "Cascaded multi-scale attention for enhanced multi-scale feature extraction and interaction with low-resolution images," *arXiv preprint arXiv:2412.02197*, 2024.

[36] Q.-T. Truong, T.-A. Vu, T.-S. Ha, J. Lokoč, Y.-H. Wong, A. Joneja, and S.-K. Yeung, "Marine video kit: a new marine video dataset for content-based analysis and retrieval," in *International Conference on Multimedia Modeling*. Springer, 2023, pp. 539–550.

[37] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541–551, 2015.

[38] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 6062–6071, 2015.

[39] H. Jiang and Y. Zheng, "Learning to see moving objects in the dark," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 7324–7333.

[40] Y. Liu, H. Zhao, K. C. Chan, X. Wang, C. C. Loy, Y. Qiao, and C. Dong, "Temporally consistent video colorization with deep feature propagation and self-regularization learning," *Computational visual media*, vol. 10, no. 2, pp. 375–395, 2024.