MobileIE-Next: Improving Image Enhancement via Enhanced HDPA and Optimized Inference Pipeline

PANG, Hao-Chung
B11705006
Information Management Department
National Taiwan University

CHEN, Peng-Yu
B11705017
Information Management Department
National Taiwan University

LIU, QIAN-YI
B11705029
Information Management Department
National Taiwan University

YANG, Tung Wei
B11705057
Information Management Department
National Taiwan University

***Abstract*—We propose MobileIE-Next, an enhanced lightweight framework for real-time low-light image enhancement on mobile devices. While the original MobileIE [1] achieves strong efficiency with only 4K parameters, its Hierarchical Dual-Path Attention (HDPA) suffers from weak global representation and unstable local attention, leading to color distortions and contrast loss. We introduce Frequency-Guided HDPA (FG-HDPA), which incorporates frequency-domain descriptors and structure-aware local attention to improve texture preservation and color fidelity. We further propose EC-FG-HDPA, a simplified variant that reduces training time by 64.7% without sacrificing accuracy. In addition, we design an inference-time enhancement pipeline that combines the model output with a lightweight classical adjustment module and selects the better result via no-reference quality cues, yielding more stable and perceptually pleasing outputs under diverse lighting conditions. Experiments demonstrate consistent improvements on both LOLv1 [3] and UIEB [4] datasets: FG-HDPA achieves the best PSNR and LPIPS in both tasks, while EC-FG-HDPA obtains competitive second-best perceptual scores. Visual comparisons further confirm that our methods generate more faithful colors, clearer structures, and more stable illumination across diverse scenes.**

I. Introduction

Low-light image enhancement is a fundamental problem in computational photography and visual perception, directly affecting downstream tasks such as mobile imaging, nighttime surveillance, and real-time vision systems. Despite recent advances in lightweight enhancement networks, many existing methods still suffer from degraded structural fidelity, amplified noise, or unstable enhancement behavior when operating under extreme illumination conditions.

Mobile-oriented enhancement models prioritize efficiency, but this constraint often limits the expressiveness of global–local feature interactions. In particular, attention mechanisms designed for lightweight architectures tend to rely on overly simplified global descriptors and brightness-driven local responses, resulting in suboptimal structure preservation and perceptual quality. As demonstrated in recent benchmarks, such limitations manifest as inferior SSIM and LPIPS performance, even when PSNR improvements appear marginal.

In this work, we present MobileIE-Next, a refined low-light enhancement framework that targets the representational bottlenecks of global–local attention, Hierarchical Dual-Path Attention(HDPA) block in the previous model, MobileIE [1]. The key insight is that effective enhancement under low illumination requires structure-aware global reasoning and noise-robust local modulation, rather than solely brightness-based attention.

To this end, we introduce Frequency-Guided Hierarchical Dual-Path Attention (FG-HDPA), which augments spatial global descriptors with frequency-domain cues derived from FFT magnitude responses. This hybrid representation preserves texture, edge, and illumination patterns that are otherwise lost in conventional global average pooling. Simultaneously, we redesign the local attention pathway to replace max-pooling–based activation with channel-consensus structural aggregation, significantly improving stability and noise robustness.

Furthermore, we propose an Efficiency-Compliant FG-HDPA (EC-FG-HDPA) variant that reduces convolutional complexity within the Mobile Bottleneck Residual Convolution (MBRConv) blocks, substantially accelerating training while maintaining enhancement quality.

Extensive experiments demonstrate that MobileIE-Next consistently outperforms existing lightweight enhancement methods across PSNR, SSIM, and LPIPS metrics, achieving notable gains in structural similarity and perceptual quality. These results validate the effectiveness of frequency-guided global–local reasoning for low-light image enhancement under efficiency-constrained settings.

Main contributions are as follows:

- **Frequency-Guided Global Attention:** We enhance global attention by combining spatial pooling with FFT-based frequency cues, enabling better preservation of textures and structural details under low-light conditions.
- **Semantic-Aware Local Modulation:** We refine local attention by conditioning it on globally weighted features,

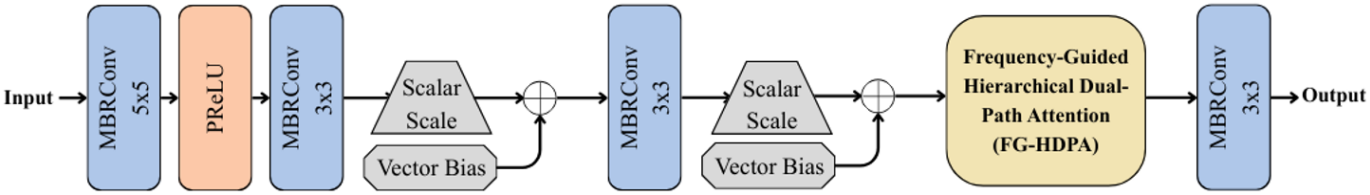


Fig. 1. The Full Structure and the Pipeline of Proposed Model

allowing the network to suppress noise and emphasize meaningful regions more robustly.

- **Efficiency-Compliant Variant:** We introduce EC-FG-HDPA, a simplified attention module that reduces training cost and latency while maintaining high visual quality and reparameterization support.
- **Strong Performance–Efficiency Trade-off:** Our model achieves state-of-the-art results in PSNR, SSIM, and LPIPS on standard benchmarks, all under strict constraints of model size, speed, and mobile deployment.

II. Related Work

A. Frequency-Aware Feature Modeling

Frequency-domain representations have been explored as an effective means of capturing global context in convolutional neural networks. Fast Fourier Convolution (FFC) [2] introduces a dual-branch architecture that explicitly processes feature maps in the Fourier domain, enabling efficient modeling of long-range dependencies and global structure. By performing convolution directly on complex-valued frequency representations, FFC [2] demonstrates strong performance in image generation and restoration tasks.

While effective, Fourier-domain convolution requires architectural modifications, complex-valued operations, and additional computational overhead, which may limit its applicability in lightweight or mobile-oriented enhancement networks. Moreover, directly learning frequency-domain filters can be unnecessary when frequency information is primarily needed to guide global reasoning rather than replace spatial feature learning.

Our work follows the philosophy of frequency-aware global modeling introduced by FFC [2], but adopts a fundamentally different and more lightweight formulation. Rather than performing convolution in the Fourier domain, we extract FFT magnitude responses to construct a frequency-guided global descriptor, which is integrated with spatial statistics to drive attention modulation.

III. Methodology

A. Observed Weaknesses in MobileIE [1]

1) Weak Global Representation: The original HDPA computes its global descriptor using global average pooling (GAP). Although computationally lightweight, GAP collapses all spatial structure into a mean vector, removing essential texture, edge, and frequency information. As a result, the global attention becomes shallow and biased toward global brightness rather than structural cues.

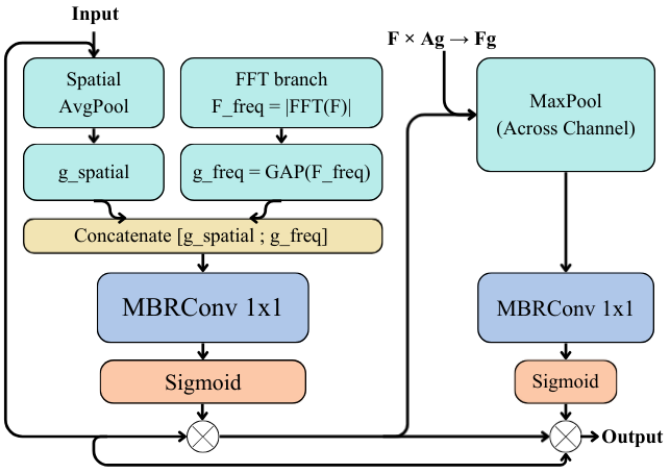


Fig. 2. Architecture of Frequency-Guided Hierarchical Dual-Path Attention (FG-HDPA) block.

To mitigate the loss of structural information, we introduce a Frequency-Guided descriptor. Specifically, we compute the FFT magnitude of intermediate features and apply average pooling to obtain a frequency-based descriptor. By concatenating spatial and frequency descriptors, the resulting hybrid global representation captures texture strength, noise patterns, and illumination variations more effectively.

2) Unstable Local Attention: The original local path relies on max pooling, effectively acting as a brightest-pixel detector. This leads to amplification of bright noise and suppression of low-contrast edges. Furthermore, the local attention lacks channel-level semantic reasoning, resulting in unstable enhancement behavior.

We propose a structure-aware local attention mechanism that first applies global-conditioned feature modulation before extracting pixelwise semantic responses across channels. By replacing brightness-driven max pooling with channel-consensus activation, the enhanced local descriptor becomes significantly more robust to noise and better aligned with the global semantic context.

B. FG-HDPA and EC-FG-HDPA

1) Frequency-Guided HDPA: To enhance the semantic richness of the original HDPA module, we propose **Frequency-Guided HDPA (FG-HDPA, shown in Fig 2)**, which incorporates both spatial context and frequency cues into the channel attention mechanism, while also improving the local attention path with semantically guided filtering.

Let $F \in \mathbb{R}^{B\times C\times H\times W}$ be the input feature map. We first compute a spatial descriptor via global average pooling (GAP), denoted as:

$$g_{spatial} = \text{GAP}(F). \tag{1}$$

To capture high-frequency texture and structural information, we apply 2D Fast Fourier Transform (FFT) on F and take the magnitude:

$$F_{freq} = |\text{FFT}(F)|, \tag{2}$$

which is interpolated back to spatial resolution and globally pooled:

$$g_{freq} = \text{GAP}(\text{Interpolate}(F_{freq}, H, W)). \tag{3}$$

The concatenated descriptor $[g_{spatial}; g_{freq}]$ is passed through two 1×1 convolutions with ReLU and sigmoid to generate global attention weights A_g:

$$A_g = \sigma(\text{Conv}^{(2)}_{1\times1}(\text{ReLU}(\text{Conv}^{(1)}_{1\times1}([g_{spatial}; g_{freq}])))). \tag{4}$$

This weight is applied to F for feature refinement:

$$F_g = A_g \cdot F. \tag{5}$$

For local attention, we reuse F_g to first compute a spatial importance map via channel-wise max pooling:

$$M = \max(F_g, \text{dim} = 1, \text{keepdim} = True), \tag{6}$$

which is then passed through a 1×1 convolution and sigmoid to produce A_l:

$$A_l = \sigma(\text{Conv}_{1\times1}(M)). \tag{7}$$

The final attention mask is obtained by element-wise multiplication of global and local attention:

$$A = A_g \cdot A_l, \tag{8}$$

and the final output is:

$$F_{out} = A \cdot F. \tag{9}$$

This dual-path attention enriches spatial understanding and texture awareness while maintaining compactness and efficiency, aligning well with mobile-friendly design.

The full model architecture is shown in Fig 1, where we replace HDPA in MobileIE with our FG-HDPA and we use scalar scale instead of vector scale.

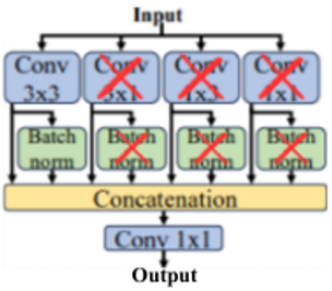


Fig. 3. The Architecture of revised MBRConv 3×3 for EC-FG-HDPA. The Architecture of revised MBRConv $N\times N$ only preserve the $N\times N$ Convolution layer and the corresponding BatchNorm.

2) Training Efficiency Improvements: To further accelerate training, we propose **Efficiency-Compliant FG-HDPA (EC-FG-HDPA)**, which reduces convolutional complexity inside the MBRConv block. Take MBRConv 3×3, shown in Fig. 3, as example. We eliminate all the convolution layers and batch normalizations, only reserving the 3×3 convolution layer and its corresponding batch normalization for MBRConv 3×3. This optimization decreases training time for LOLv1 [3] dataset from 17 seconds to 6 seconds per epoch on an RTX 5090 GPU, achieving a **reduction of approximately 64.7%** without degrading much performance.

C. Inference Enhancement

To improve prediction robustness under diverse lighting conditions, we design an inference-time enhancement pipeline that fuses learned and traditional priors. For each input image, we produce two candidate outputs: one directly from MobileIE-Next, and another refined by a lightweight **Local Adjustment Module (Loc-v3)**.

Loc-v3 applies classical image processing techniques including HSV-based shadow lifting, contrast sharpening, tone compression, and white balance adjustment, implemented using OpenCV and NumPy. These operations are efficient and require no training.

To automatically select the better result, we compute four no-reference quality cues: edge preservation, color shift (relative to input), contrast gain, and a saturation-based failure indicator. A simple scoring function ranks both candidates, and the one with higher perceptual quality is returned.

This strategy enhances visual stability and perceptual quality without retraining, and maintains compatibility with mobile or real-time deployment scenarios.

IV. Experiments

A. Experimental Settings

1) Implementation Details: We implemented MobileIE-Next based on MobileIE [1] in PyTorch. The model uses the Adam optimizer with a cosine annealing learning rate schedule, starting at 0.001. The learning rate is reset every 50 epochs with gradual decay. A 10-epoch warm-up phase is applied with a fixed learning rate of 1e-6. The model is trained for 2,000 epochs.

TABLE I

PERFORMANCE COMPARISON OF DIFFERENT LOW-LIGHT IMAGE ENHANCEMENT MODELS ON LOLV1 [3] DATASET. THE TOP RESULTS ARE MARKED: BEST IN BOLD AND SECOND WITH UNDERLINE.

Method	PSNR↑	SSIM↑	LPIPS↓
Kind++ [8]	17.75	0.766	0.198
DDNet [9]	21.82	0.802	0.186
PairLIE [10]	19.51	0.736	0.248
IAT [7]	23.38	0.808	0.216
Zero-DCE [5]	14.86	0.559	0.335
3DLUT [11]	17.59	0.721	0.232
Zero-DCE++ [12]	14.68	0.472	0.340
SCI [13]	14.90	0.531	0.341
SGZ [14]	15.28	0.473	0.339
RUAS [6]	16.40	0.500	0.270
SYELLE [15]	21.03	0.794	0.219
Adv-LIE [16]	23.02	0.808	0.203
MobileIE [1]	23.62	0.812	0.198
Ours(EC-FG-HDPA)	<u>23.87</u>	<u>0.867</u>	<u>0.145</u>
Ours(FG-HDPA)	**23.88**	**0.875**	**0.138**

(a) Input (b) Zero-DCE [5] (c) RUAS [6] (d) IAT [7]

(e) MobileIE [1] (f) EC-FG-HDPA (g) FG-HDPA (h) Ground Truth

Fig. 4. Visualization comparison of LLE on LOLv1 [3].

2) Dataset and Metrics: For Low-Light Enhancement(LLE) task, the LOLv1 [3] dataset is used for training and testing. For Underwater Image Enhancement(UIE), the UIEB [4] dataset is utilized.

B. Quantitative and Visual Comparisons

1) LLE: We train and evaluate our method on the LOLv1 [3] low-light enhancement dataset, consisting of 485 training pairs and 15 testing pairs. Each pair includes a low-light image and a corresponding normal-light ground truth.

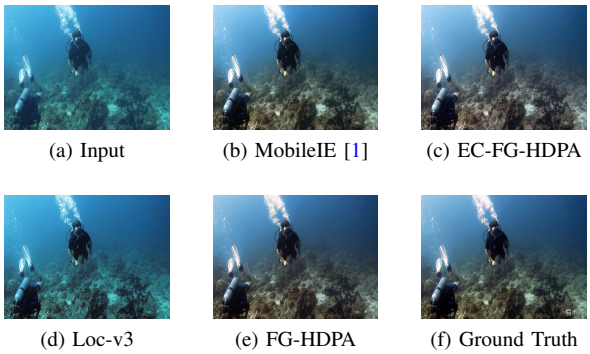
Evaluation is conducted using standard metrics including peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and perceptual similarity (LPIPS). These metrics assess pixel accuracy, structural consistency, and perceptual quality respectively.

As results shown in Table I, FG-HDPA and EC-FG-HDPA improve PSNR, SSIM, and LPIPS over the baseline MobileIE [1] model. Visual comparisons are shown in Fig. 4, which further demonstrate that the results of FG-HDPA and EC-FG-HDPA are more similar to the ground truth. MobileIE [1] also brightens the image effectively; however, the color of the glass remains slightly gray. Moreover, the glass color in the FG-HDPA result contains more blue, which is much closer to the ground truth. Although EC-FG-HDPA does not restore the glass color as well as FG-HDPA, the blue color is still visible in the glass with reduced gray tones.

TABLE II

PERFORMANCE COMPARISON OF DIFFERENT UNDERWATER IMAGE ENHANCEMENT MODELS ON UIEB [4] DATASETS. THE RESULTS WERE RUN BY OURSELVES SINCE OUR TRAIN SET AND TEST SET ARE DIFFERENT COMPAIRING TO THE ORIGINAL PAPER OF MOBILEIE [1]. THE TOP RESULTS ARE MARKED: BEST IN BOLD AND SECOND WITH UNDERLINE.

Method	PSNR↑	SSIM↑	LPIPS↓
MobileIE [1]	<u>22.96</u>	**0.908**	**0.095**
Ours(EC-FG-HDPA)	22.03	0.889	0.102
Ours(FG-HDPA)	**23.12**	<u>0.899</u>	<u>0.096</u>



(a) Input (b) MobileIE [1] (c) EC-FG-HDPA

(d) Loc-v3 (e) FG-HDPA (f) Ground Truth

Fig. 5. Visualization comparison of UIE on UIEB [4].

2) UIE: Our training and evaluation for our model on UIE tasks are conducted on UIEB [4] dataset, which contains 796 pairs of training set and 94 pairs test set. Each pair is composed of a low-light image and a ground truth. The evaluation is also performed using PSNR, SSIM and LPIPS, just as LLE tasks.

The quantitative results in Table II shows that although our models are not the best method under SSIM and LPIPS metrics, FG-HDPA outperform other models on PSNR while only tiny differences on LPIPS compared with MobileIE [1].

The visual results shown in Fig. 5 further demonstrate that for the whole image, FG-HDPA derives a closer match to the ground truth; however, the result generated by MobileIE [1] has more contrast between light and dark areas. This is the aspect we can improve since we focus more on managing global information. EC-FG-HDPA also recovers the image well, but the rock or seabed should contain stronger green instead of being so dim and gray. In addition, the Loc-v3 refinement produces a noticeably more natural underwater appearance, restoring the bluish–green water tone and depth cues that MobileIE tends to over-sharpen, giving the final result a clearer sense of being "in the water."

V. CONCLUSION

We presented **MobileIE-Next**, a frequency-aware enhancement framework that advances the MobileIE [1] architecture by addressing key limitations in its Hierarchical Dual-Path Attention (HDPA) module.

To better model global structure and local semantic cues under low-light conditions, we proposed the **Frequency-**

Guided HDPA (FG-HDPA) block. By incorporating FFT-derived frequency features into global attention and refining local attention through global-aware modulation, the model significantly improves its ability to enhance detail-rich and noise-prone regions.

We further introduced **EC-FG-HDPA**, a computationally efficient variant that reduces training time and parameter overhead, enabling mobile-friendly deployment without compromising visual quality.

In addition, we designed an inference enhancement strategy that fuses learned predictions with traditional image priors via a lightweight scoring-based selection mechanism. This pipeline enhances robustness across varying lighting conditions without retraining.

Comprehensive experiments demonstrate that MobileIE-Next achieves superior performance on PSNR, SSIM, and LPIPS, while maintaining a compact size and fast runtime, validating the effectiveness of our frequency-guided dual-path design for efficient and perceptually robust image enhancement.

REFERENCES

[1] H. Yan, A. Li, X. Zhang, Z. Liu, Z. Shi, C. Zhu, and L. Zhang, "MobileIE: An Extremely Lightweight and Effective ConvNet for Real-Time Image Enhancement on Mobile Devices," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2025.

[2] L. Chi, B. Jiang, and Y. Mu, "Fast Fourier Convolution," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 4479–4488, 2020.

[3] W. Chen, W. Wang, W. Yang, and J. Liu, "Deep Retinex Decomposition for Low-Light Enhancement," in *Proc. British Machine Vision Conference (BMVC)*, 2018.

[4] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An Underwater Image Enhancement Benchmark Dataset and Beyond," arXiv:1901.05495, 2019.

[5] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1780–1789.

[6] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo, "Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10561–10570.

[7] Z. Cui, K. Li, L. Gu, S. Su, P. Gao, Z. Jiang, Y. Qiao, and T. Harada, "You only need 90k parameters to adapt light: A lightweight transformer for image enhancement and exposure correction," *arXiv preprint arXiv:2205.14871*, 2022.

[8] Y. Zhang, X. Guo, J. Ma, W. Liu, and J. Zhang, "Beyond brightening low-light images," *Int. J. Comput. Vision*, vol. 129, no. 4, pp. 1013–1037, Apr. 2021, doi: 10.1007/s11263-020-01407-x.

[9] J. Qu, R. W. Liu, Y. Gao, Y. Guo, F. Zhu, and F.-Y. Wang, "Double domain guided real-time low-light image enhancement for ultra-high-definition transportation surveillance," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 8, pp. 9550–9562, Aug. 2024.

[10] Z. Fu, Y. Yang, X. Tu, Y. Huang, X. Ding, and K.-K. Ma, "Learning a simple low-light image enhancer from paired low-light instances," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 22252–22261.

[11] H. Zeng, J. Cai, L. Li, Z. Cao, and L. Zhang, "Learning image-adaptive 3D lookup tables for high performance photo enhancement in real-time," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 4, pp. 2058–2073, Apr. 2022.

[12] C. Li, C. Guo, and C. C. Loy, "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4225–4238, Aug. 2022.

[13] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5637–5646.

[14] S. Zheng and G. Gupta, "Semantic-guided zero-shot learning for low-light image/video enhancement," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 581–590.

[15] W. Gou, Z. Yi, Y. Xiang, S. Li, Z. Liu, D. Kong, and K. Xu, "Syenet: A simple yet effective network for multiple low-level vision tasks with real-time performance on mobile device," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 12182–12195.

[16] W. Y. Wang, L. Liu, and P. Cai, "Adversarially regularized low-light image enhancement," in *Proc. Int. Conf. Multimedia Modeling (MMM)*, Amsterdam, The Netherlands, Jan.–Feb. 2024, pp. 230–243.

[17] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3227–3234, Apr. 2020, doi: 10.1109/LRA.2020.2974710.

[18] A. Naik, A. Swarnakar, and K. Mittal, "Shallow-UWnet: Compressed model for underwater image enhancement (student abstract)," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, vol. 35, no. 18, May 2021, pp. 15853–15854, doi: 10.1609/aaai.v35i18.17923.

[19] Z. Fu, W. Wang, Y. Huang, X. Ding, and K.-K. Ma, "Uncertainty inspired underwater image enhancement," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2022, pp. 465–482.

[20] Z. Ma and C. Oh, "A wavelet-based dual-stream network for underwater image enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2022, pp. 2769–2773.

[21] L. Peng, C. Zhu, and L. Bian, "U-shape transformer for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 32, pp. 3066–3079, 2023.

[22] J. Jiang, T. Ye, J. Bai, S. Chen, W. Chai, S. Jun, Y. Liu, and E. Chen, "Five A^+ network: You only need 9K parameters for underwater image enhancement," *arXiv preprint arXiv:2305.08824*, 2023.

[23] X. Liu, S. Lin, K. Chi, Z. Tao, and Y. Zhao, "Boths: Super lightweight network-enabled underwater image enhancement," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[24] C. Zhao, W. Cai, C. Dong, and Z. Zeng, "Toward sufficient spatial-frequency interaction for gradient-aware underwater image enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2024, pp. 3220–3224.

[25] S. Zhang, S. Zhao, D. An, D. Li, and R. Zhao, "LiteEnhanceNet: A lightweight network for real-time single underwater image enhancement," *Expert Syst. Appl.*, vol. 240, p. 122546, 2024.

[26] F. Zhou, D. Wei, Y. Fan, Y. Huang, and Y. Zhang, "A 7K-parameter model for underwater image enhancement based on transmission map prior," *arXiv preprint arXiv:2405.16197*, 2024.

TABLE III
TEAM CONTRUBUTION

Name	Student ID	Contributions
PANG, Hao-Chung	B11705006	· Attempted to use Restormer to improve performance. · Write report and slides. · Film part of the video for presentation.
CHEN, Peng-Yu	B11705017	· Proposed EC-FG-HDPA. · Train and test EC-FG-HDPA on LOLv1 and UIEB dataset. · Train and test FG-HDPA and MobileIE on UIEB dataset. · Write report and slides. · Film part of the video for presentation.
LIU, QIAN-YI	B11705029	· Proposed Loc-v3 · Write part of the slides.
YANG, Tung Wei	B11705057	· Design and train FG-HDPA model · Write report and slides. · Film part of the video for presentation.