This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

MobileIE: An Extremely Lightweight and Effective ConvNet for Real-Time Image Enhancement on Mobile Devices

Hailong Yan[1], Ao Li[1], Xiangtao Zhang[1], Zhe Liu[1], Zenglin Shi[2], Ce Zhu[1], Le Zhang[1]*
[1]UESTC [2]Hefei University of Technology
yanhailong@std.uestc.edu.cn, lezhang@uestc.edu.cn

Abstract

Recent advancements in deep neural networks have driven significant progress in image enhancement (IE). However, deploying deep learning models on resource-constrained platforms, such as mobile devices, remains challenging due to high computation and memory demands. To address these challenges and facilitate real-time IE on mobile, we introduce an extremely lightweight Convolutional Neural Network (CNN) framework with around 4K parameters. Our approach integrates reparameterization with an Incremental Weight Optimization strategy to ensure efficiency. Additionally, we enhance performance with a Feature Self-Transform module and a Hierarchical Dual-Path Attention mechanism, optimized with a Local Variance-Weighted loss. With this efficient framework, we are the first to achieve real-time IE inference at up to 1,100 frames per second (FPS) while delivering competitive image quality, achieving the best trade-off between speed and performance across multiple IE tasks. The code will be available at https://github.com/AVC2-UESTC/MobileIE.git.

1. Introduction

Image enhancement (IE) aims to restore images degraded by factors such as camera limitations [86], poor lighting [48], or challenging environments (e.g., underwater [55]). With the growing prevalence of smart devices and embedded systems, real-time image enhancement has become essential in applications like mobile devices. These scenarios require not only high-quality output but also real-time processing, demanding efficient image enhancement within the constraints of limited computational resources [9, 16–18].

Recent advancements in Transformer-based [14, 42, 69] and Diffusion-based [79, 84, 96] methods have demonstrated significant progress in image enhancement (IE) tasks. However, their reliance on computationally expensive self-attention mechanisms and iterative diffusion processes makes them unsuitable for mobile devices. While lightweight models have been developed to reduce computational complexity using parameter compression and low-FLOPs architecture designs [57, 59, 63, 80, 83], these approaches often compromise enhancement quality, making it challenging to achieve high performance across a wide range of degradation scenarios [1]. Furthermore, the increasing demand for high-resolution images [69] exacerbates the computational burdens on mobile platforms.

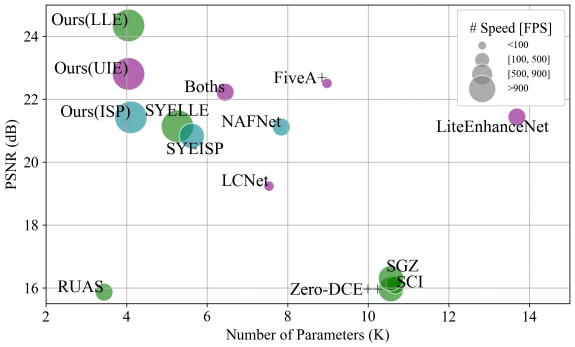


Figure 1. Efficiency comparison. MobileIE achieves superior performance in balancing speed, compactness, and accuracy.

Besides facing computational bottlenecks, current IE methods are typically optimized for specific degradation types, restricting their adaptability. Although advanced operators [10, 11] have shown effectiveness in certain tasks, their complexity makes them unsuitable for mobile deployment. Consequently, designing a universal and efficient network capable of real-time IE on resource-constrained devices remains a significant research challenge.

We believe that mobile-based IE should strike a balance between speed and performance, leveraging general architectures and hardware-friendly operators [66, 88]. To address these challenges, we introduce MobileIE, an efficient method designed for real-time enhancement in resource-constrained environments, optimizing both performance and resource utilization. Although degradation types vary widely, IE tasks share common requirements: global information for structural integrity and local details for accurate

*Corresponding author.

recovery. This commonality enables the use of universal modules across different tasks.

Building on this concept, MobileIE employs a streamlined topology and deployable operators to achieve efficient performance on resource-constrained devices. We decouple the training and inference phases, simplifying the feedforward structure for more efficient inference. At the core of our design is MBRConv, a Multi-Branch Reparameterized Convolution, which captures multi-scale features using various convolution sizes. These features are then processed through concatenation, compression, and mapping to achieve the desired output dimensions. Additionally, we introduce a Feature Self-Transform (FST) module, which enhances the feature representation capability of linear convolutions by capturing nonlinear relationships through secondary feature interactions.

To enhance focus on critical regions, we simplify the attention mechanism and develop a hardware-friendly Hierarchical Dual-Path Attention (HDPA) mechanism, which efficiently fuses global and local features. To overcome training bottlenecks, we introduce an Incremental Weight Optimization (IWO) Strategy and a Local Variance Weighted (LVW) Loss function. The IWO strategy freezes prior knowledge, integrates it into trainable convolutional kernels, and re-parameterizes these into new kernels to boost model performance. Meanwhile, the LVW Loss function enhances accuracy without increasing the number of parameters. Key contributions of this paper include:

- We introduce MobileIE, a framework designed for real-time image enhancement on mobile devices, featuring compact and efficient modules tailored for resource-constrained environments.
- We propose an Incremental Weight Optimization (IWO) Strategy and a Local Variance Weighted (LVW) Loss to address the challenges of training compact models, enhancing performance without adding complexity.
- We demonstrate that MobileIE achieves state-of-the-art speed and performance across three image enhancement tasks, sustaining over 100 FPS and enabling seamless deployment on mobile devices.

2. Related Works

2.1. Low-Level Vision Tasks

This section covers three key tasks: Low-Light Enhancement (LLE), Underwater Image Enhancement (UIE), and end-to-end Image Signal Processing (ISP).

LLE. LLE focuses on mitigating color distortion and noise in degraded images. Retinex-based [5, 46, 78, 81, 85] enhance images by separating illumination and reflectance. Zero-DCE [28] adjusts underexposed images using a luminance enhancement curve, and PairLIE [26] learns enhancement from varying lighting conditions. Transformer-based [5, 71, 76] and Mamba-based [93] methods have also achieved notable success, while Diffusion-based [39, 85] methods excel at generating accurate images in extremely low-light conditions, even with noise and limited data.

UIE. Compared to traditional UIE methods [24, 101], CNN-based [38, 52, 53, 61] directly learn feature representations from underwater images. Five A+ [40] proposed a two-stage framework with a pixel attention module for real-time enhancement. Transformer-based [41, 55] and Mamba-based [2, 12] models more effectively handle complex underwater scenes with their global receptive fields. Additionally, Diffusion-based [65, 96] approaches reduce noise and enhance contrast efficiently.

ISP. Learning-based ISP pipelines replace traditional multi-module systems by employing end-to-end models that directly process RAW input into RGB images. PyNet [35] focuses on mobile ISPs, while MW-ISP [34] and AWNet [20] improve performance using wavelet transforms. LiteISP [95] resolves misalignment between RAW and sRGB images, achieving better results through joint learning. Rawformer [56] eliminates the need for paired datasets with an unsupervised Transformer, and DiffRAW [84] introduces diffusion models into ISP. Several ISP challenge works [34, 36, 37] have also produced notable results.

2.2. Efficient Architectures Design

Efficient neural networks aim to balance computational complexity and performance. Models like MobileNet [32] and ShuffleNet [91] achieve high performance with fewer FLOPs, while GhostNet [29] and FasterNet [8] generate feature maps through cost-effective operations. StarNet [51] excels in extracting rich representations with star-shaped operations, and VanillaNet [7] demonstrates minimalism's power using only a few convolutions.

Furthermore, re-parameterization strategies have also proven successful in efficient network design. Models like ACNet [21], DBB [22], and RepVGG [23] use multi-branch topologies to accelerate inference, while RepGhost [23], RepViT [68], and MobileOne [66] adopt similar approaches to enhance performance. Several NTIRE challenge submissions [15, 45, 60] have utilized re-parameterized structures to accelerate inference in efficient IE tasks.

3. Proposed Method

3.1. Overall Pipeline

We aim to design an efficient IE model that balances parameters, speed, and performance. Following the simplicity principle [88], we adopted a streamlined topology based on basic operations, illustrated in Figure 2.

The proposed MobileIE, built on Re-parameterization, integrates four key components: shallow feature extraction, deep feature extraction, feature transform, and an atten-

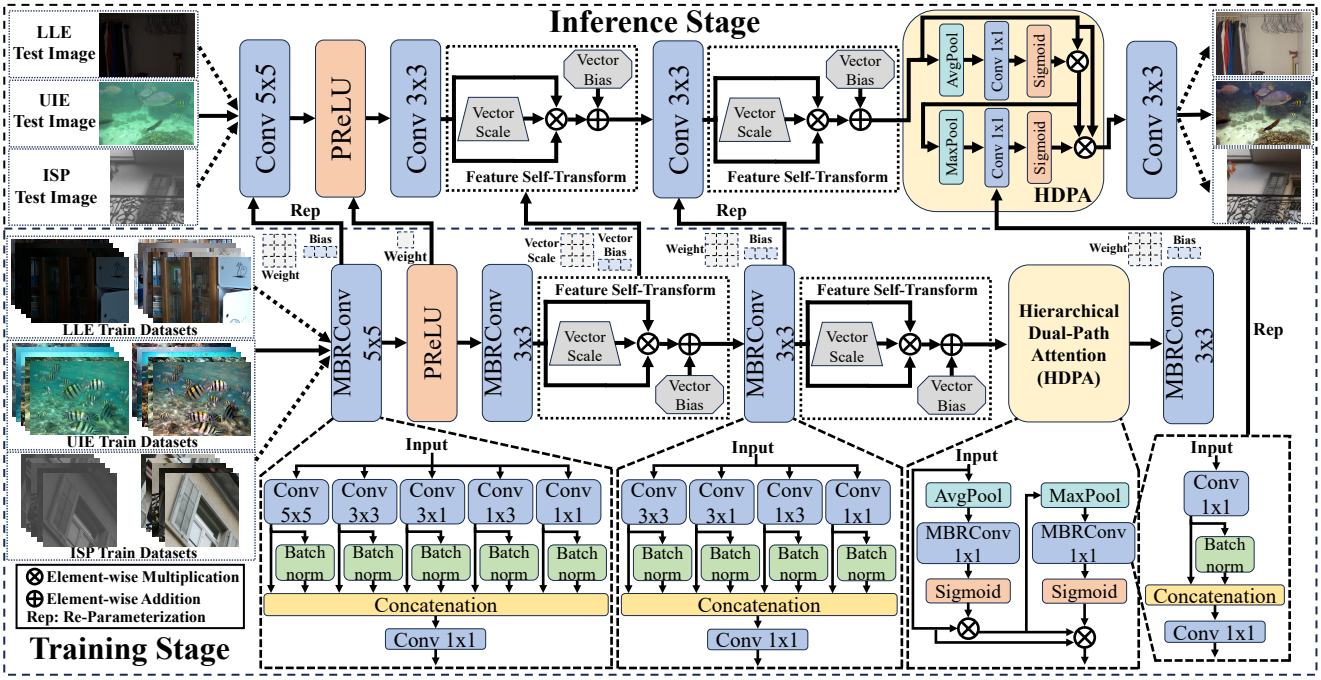


Figure 2. Architecture of the proposed MobileIE includes Multi-Branch Re-param Convolution (MBRConv), Feature Self-Transform (FST), and Hierarchical Dual-Path Attention (HDPA). During inference, MBRConv re-parameterizes into a standard convolution, reducing the model size to 4K parameters while maintaining training performance. The detailed architecture settings are provided in the Appendix.

tion mechanism. During training, the degraded image first passes through MBRConv5 $\times$ 5 and the PReLU activation function to extract shallow features. Two MBRConv3 $\times$ 3 and FST modules then process these to learn deeper features. The HDPA mechanism directs the model's focus to important regions, and MBRConv3 $\times$ 3 refines the output. To address training performance bottlenecks, we introduce the IWO strategy and LVW loss function.

In inference, all MBRConv layers are re-parameterized into standard convolutions, reducing parameters while maintaining performance. MobileIE's streamlined structure ensures fast inference and easy deployment on mobile devices. Despite its simplicity, MobileIE surpasses SOTA lightweight IE models, as shown in Figure 1.

3.2. Multi-Branch Re-param Convolution

Re-parameterization has achieved success in high-level vision tasks [21–23], but yields unsatisfactory results when directly applied to image enhancement [27]. We introduce MBRConv, specifically designed for image enhancement.

Figure 3(a) shows MBRConv with multiple convolutional branches of varying kernel sizes, capturing multi-scale features that are concatenated and integrated via a Conv 1 $\times$ 1 ($conv_{out}$). The branches are re-parameterized during inference into a single convolution, reducing computational cost while preserving training performance. Unlike previous Rep methods, MBRConv includes parallel Batch Norm (BN) layers in each branch. Although BN is less effective for IE tasks [74, 92], it enhances nonlinearity and cannot be replaced by other activations due to its unique ability to merge into a single convolution. The parallel BN layers retain both smoothed and original features, improving robustness across diverse data distributions. Additional details on MBRConv can be found in the Appendix.

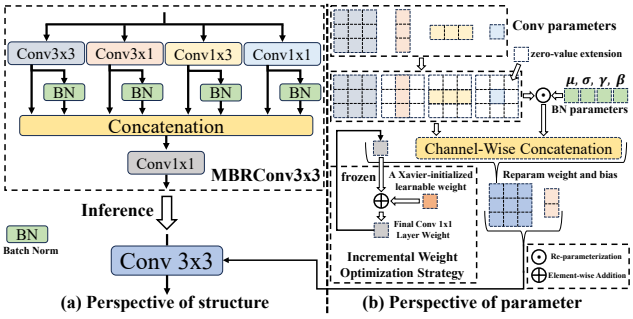


Figure 3. Structural of MBRConv3 $\times$ 3.

Although multi-branch topologies assist in feature extraction, compact networks still struggle to learn complex features, leading to stagnation in later training stages. Adjusting training hyperparameters often results in marginal improvements. To address this, we propose the Incremental Weight Optimization (IWO) strategy, which combines learnable weights with prior knowledge to enhance the in-

tegration performance of the $conv_{out}$ by better capturing relationships between features across layers in MBRConv.

As shown in Figure 3(b), IWO combines two weight components: frozen weights, W_{pre}, from prior training, and learnable weights, W_{learn}. These are fused as:

$$W_{final} = \text{Frozen}(W_{pre}) + W_{learn}, \quad (1)$$

where W_{pre} is the optimal weight from earlier training and remains frozen, and W_{learn} is dynamically updated during subsequent training. The final weight, W_{final}, is applied to the concatenated feature map:

$$y_{out} = F_{conv1\times1}(x_{concat}; W_{final}) + b_{conv_{out}}, \quad (2)$$

where x_{concat} denotes the concatenated multi-branch features, F_{conv} represents the convolution. y_{out} and $b_{conv_{out}}$ represent the output feature map and the bias, respectively. W_{pre} offers a stable initial feature representation, minimizing redundancy. W_{learn} refines these features, improving task-specific detail capture. IWO combines prior knowledge with learnable parameters, balancing knowledge transfer and feature refinement to improve feature integration.

3.3. Feature Self-Transform

While MBRConv captures multi-scale features, its linear operations restrict higher-order feature interactions. To address this, we introduce FST, which employs a quadratic interaction mechanism [51] to boost the model's nonlinear expressiveness. In FST, the input features are element-wise multiplied by themselves, scaled by a learnable parameter, $Scale$, and adjusted by a learnable $bias$ term.

$$\text{FST}(x) = Scale \cdot (x * x) + bias. \quad (3)$$

This quadratic interaction captures more complex relationships between features, unlike traditional linear combinations. A learnable $bias$ term is also applied channel-wise to fine-tune the output further. FST significantly improves feature representation by capturing higher-order interactions, while the learnable bias enhances adaptability across dimensions. This design balances improved expressiveness with computational efficiency, making FST suitable for lightweight models and real-time inference.

3.4. Hierarchical Dual-Path Attention

We propose a simple HDPA mechanism to improve feature extraction in IE. HDPA selectively captures both global and local features, integrating spatial and channel interactions to improve feature precision and comprehensiveness. Its dual-path structure optimizes feature selection at each layer through two distinct pathways. HDPA operates in two steps:

(1) Global Feature Extraction: Global statistics are extracted using adaptive average pooling. Global features are processed by an MBRConv1x1 layer, and Sigmoid activation to generate channel-wise attention weights, modulating the importance of each channel. This step is expressed as:

$$A_g = \text{Sigmoid}(\text{MBRConv1} \times 1(\text{AvgPool}(F))), \quad (4)$$

where A_g denotes the global attention weights, and F represents the input features.

(2) Local Feature Enhancement: In the second pathway, local features are extracted using max pooling on globally enhanced features. The input is multiplied by the global weights $W_g = F * A_g$, followed by max pooling to capture local information. The local responses are processed by an MBRConv1 $\times$ 1, with Sigmoid activation generating the local attention weights.

$$A_l = \text{Sigmoid}(\text{MBRConv1} \times 1(\text{MaxPool}(W_g))), \quad (5)$$

where A_l represents the local attention weights.

Finally, the global and local attention weights are combined through element-wise multiplication to produce the final attention feature, which is applied to the input features:

$$\hat{F} = \text{HDPA}(F) = (A_g * A_l) * F, \quad (6)$$

where $\hat{F}$ donates the output feature after HDPA processing.

This dual-path hierarchical design balances global context and local feature enhancement, improving both depth and accuracy in feature extraction. Its simplicity reduces the computational cost, making it well-suited for resource-constrained environments.

3.5. Local Variance Weighted Loss

Compact CNNs, with limited parameters, struggle to capture rich features and are sensitive to extreme pixels that hinder optimization [1, 27]. While $L1$ loss is robust, its equal weighting limits outlier handling; $L2$ emphasizes outliers but may cause imbalance. To address this, we introduce Local Variance Weighted Loss (LVW), which adapts to local variability and mitigates outliers effectively.

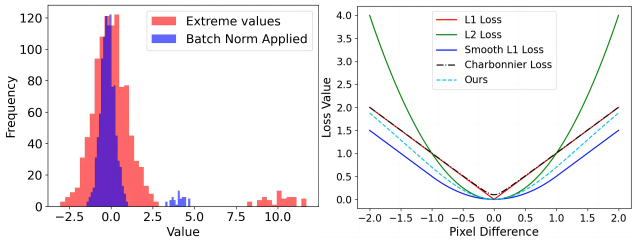


Figure 4. (a) Effect of Batch Norm on Data Distribution. (b) Comparison of Loss Functions for Pixel Difference Handling.

Let $\{\mathbf{O}, \mathbf{L}\} \in \mathbb{R}^{B\times C\times H\times W}$ denote the predicted output and ground-truth, where B, C, H, and W denote batch

Method	Venue	#Params↓	Model Size↓	Latency↓	Latency↓	FPS↑	LOLv1 [75]				LOLv2-Real [82]			
		(K)	(MB)	(GPU,ms)	(SoC,ms)	(600×400)	PSNR↑	SSIM↑	LPIPS↓	SCORE↑	PSNR↑	SSIM↑	LPIPS↓	SCORE↑
Kind++ [94]	IJCV'21	8,280	/	>500	>500	<10	17.75	0.766	0.198	<0.001	17.66	0.770	0.217	<0.001
DDNet [58]	IEEE TITS'24	5,389	20.56	32.910	>500	30.792	21.82	0.802	0.186	0.208	23.02	0.838	0.173	1.099
PairLIE [26]	CVPR'23	341.767	1.30	11.580	>500	86.354	19.51	0.736	0.248	0.024	19.88	0.778	0.234	0.040
IAT [19]	BMVC'22	86.856	0.33	6.204	202.33	161.186	23.38	0.808	0.216	9.604	25.46	0.843	0.182	171.690
Zero-DCE [28]	CVPR'20	79.416	0.30	2.539	82.94	393.910	14.86	0.559	0.335	<0.001	18.06	0.574	0.313	0.015
3DLUT [87]	IEEE TPAMI'20	593.500	2.26	1.176	/	850.340	17.59	0.721	0.232	0.218	19.68	0.637	0.224	6.000
Zero-DCE++ [44]	IEEE TPAMI'21	10.561	0.04	1.974	57.91	506.558	14.68	0.472	0.340	<0.001	17.23	0.412	0.319	0.006
SCI [50]	CVPR'22	10.671	0.04	4.773	116.8	209.525	14.90	0.531	0.341	<0.001	17.30	0.534	0.308	0.003
SGZ [98]	WACV'22	10.561	0.04	1.970	59.27	507.662	15.28	0.473	0.339	<0.001	17.34	0.409	0.322	0.007
RUAS [46]	CVPR'21	3.438	0.01	4.421	109.15	226.175	16.40	0.500	0.270	<0.001	15.33	0.488	0.310	<0.001
SYELLE [27]	ICCV'23	5.259	0.02	0.944	7.73	1059.732	21.03	0.794	0.219	2.428	21.26	0.801	0.308	3.340
Adv-LIE [72]	MMM'24	238.560	0.91	14.269	>500	70.083	23.02	0.808	0.203	2.535	21.95	0.844	0.192	0.575
Ours	/	4.047	0.02	0.895	6.72	1120.584	23.62	0.812	0.198	92.855	25.08	0.845	0.180	702.767

Table 1. Performance comparison of different low-light image enhancement models on LOL datasets. SCORE [37] represents a comprehensive measure of model performance and efficiency. The top results are marked: best in red and second in blue.

size, number of channels, height, and width of the image, respectively. The absolute difference between the predicted output and the ground truth at pixel (m, n) is calculated as:

$$\Delta_{m,n} = \|O_{m,n} - L_{m,n}\|_1. \tag{7}$$

For each predicted output, we calculate the local mean $\mu_{m,n}$ and variance $\sigma^2_{m,n}$ of $\Delta_{m,n}$ across the spatial dimensions H and W. This can be expressed as follows:

$$\begin{cases} \mu_{m,n} = \dfrac{1}{H \cdot W}(\sum\limits_{m=1}^{H}\sum\limits_{n=1}^{W}\Delta_{m,n}) \\ \sigma^2_{m,n} = \dfrac{1}{H \cdot W}(\sum\limits_{m=1}^{H}\sum\limits_{n=1}^{W}\Delta_{m,n} - \mu_{m,n})^2 \end{cases} \tag{8}$$

To demonstrate the impact of mean and variance on outliers, we applied BN to simulated data with both normal values and outliers (Figure 4(a)). BN balances sample contributions to mean and variance, reducing outliers' influence on the distribution. Based on these local statistics, we compute a weighted factor for each pixel. The weight W_Δ, is normalized by the absolute deviation of the error relative to the local mean and scaled according to the local variance:

$$W_\Delta = \mathrm{Tanh}(\frac{|\Delta_{m,n} - \mu_{m,n}|}{\sigma_{m,n} + \varepsilon}), \tag{9}$$

where ε is a small constant added to prevent division by zero. The Tanh function is applied to ensure the weights remain within a bounded range, facilitating a smooth transition between high-variance and low-variance regions.

The final loss is calculated by multiplying $\Delta_{m,n}$ with the corresponding local weighted factor:

$$\mathcal{L}_{LVW} = \frac{1}{H \cdot W}\sum_{m=1}^{H}\sum_{n=1}^{W}(W_\Delta \cdot \Delta_{m,n}). \tag{10}$$

$\mathcal{L}_{LVW}$ dynamically adjusts each pixel's contribution to the overall loss based on locally computed weights, thereby placing greater emphasis on optimizing anomalous pixels.

Figure 4(b) illustrates how different loss functions handle pixel differences, emphasizing their handling of prediction errors. $\mathcal{L}_{LVW}$ avoids the steep increase of $L2$ loss for large errors, reducing sensitivity to extreme pixels while preserving enough gradient response for small errors, balancing robustness and detail preservation.

4. Experiments

4.1. Experimental Settings

Implementation Details. We implemented MobileIE 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 [30] is applied with a fixed learning rate of 1e-6. The model is trained for 2,000 epochs, incorporating an Incremental Weight Optimization Strategy for improved convergence. For the ISP task, input data is preprocessed into a 256×256 Bayer pattern.

Dataset and Metrics. For LLE task, the LOLv1 [75] and LOLv2 [82] datasets are used for both training and testing. For UIE and ISP tasks, the UIEB [43] and ZRR [35] datasets are utilized, respectively.

4.2. Quantitative and Visual Comparisons

In this section, we compare the proposed MobileIE with current lightweight SOTA methods across three IE tasks, focusing on visual comparisons and performance metrics. Specifically, PSNR, SSIM, LPIPS, and SCORE [37] are used as evaluation metrics, while computational complexity is tested on a single NVIDIA 4090 (GPU) and a smartphone with a Snapdragon 8 Gen 3 System on Chip (SoC).

LLE: The quantitative results on the LOLv1 [75] and LOLv2 [82] datasets are presented in Table 1. MobileIE achieves superior or comparable PSNR and SSIM scores across all metrics, with an inference speed of **0.895 ms**. Compared to the lightweight SOTA method IAT [19], MobileIE delivers comparable PSNR gains while using only **4.7%** of the parameters, with a **6.9x** faster inference speed.

Method	Venue	#Params↓ (K)	Model Size↓ (MB)	Latency↓ (GPU,ms)	Latency↓ (SoC,ms)	FPS↑ (640×480)	UIEB [43] PSNR↑	SSIM↑	LPIPS↓	SCORE↑
FUnIE-GAN [38]	IEEE RA-L'20	7,020	26.78	3.698	76.61	270.428	19.72	0.845	0.239	0.101
Shallow-UWNet [53]	AAAI'21	219.456	0.84	11.158	400.01	89.620	16.69	0.747	0.365	<0.001
PUIE [25]	ECCV'22	1,401	5.34	24.000	>500	41.658	21.25	0.885	0.161	0.130
UIE-WD [52]	ICASSP'22	13,704	52.28	10.235	386.72	97.920	20.92	0.847	0.212	0.192
U-Shape [55]	IEEE TIP'23	22,817	87.04	38.484	>500	25.984	21.25	0.845	0.198	0.081
FiveA+ [40]	BMVC'23	8.974	0.03	11.700	423.43	90.224	22.51	0.902	0.165	1.525
Boths [47]	IEEE GRSL'23	6.447	0.02	2.567	58.04	389.431	22.23	0.904	0.156	2.173
SFGNet [97]	ICASSP'24	1,298	4.95	>100	>500	<10	21.66	0.871	0.191	<0.001
LiteEnhanceNet [90]	ESWA'24	13.688	0.05	6.654	190.02	150.289	21.44	0.903	0.168	0.608
LSNet [99]	Arxiv'24	7.534	0.03	59.519	>500	16.810	19.24	0.829	0.242	0.003
Ours	/	4.047	0.02	0.910	8.94	1099.370	22.81	0.906	0.155	29.711

Table 2. Performance comparison of different underwater image enhancement models on UIEB datasets. SCORE [37] represents a comprehensive measure of model performance and efficiency. The top results are marked: best in red and second in blue.

Figure 5 shows the visual results, where MobileIE produces higher-quality images, closely resembling the ground truth.

UIE: The quantitative results in Table 2 show that MobileIE achieves a PSNR of 22.81 dB. Remarkably, MobileIE surpasses FiveA+ [40] by 0.3 dB in PSNR while using only 45% of its parameters. Figure 6 provides visual comparisons, highlighting MobileIE's superior color restoration compared to other lightweight methods.

ISP: On the ZRR [43] dataset, the quantitative results in Table 3 show that MobileIE maintains competitive performance while delivering **3x** faster inference than existing lightweight SOTA models. Figure 7 provides the corresponding visual comparisons.

Mobile Deployment: As shown in Tables 1, 2, and 3, MobileIE's streamlined design (4K parameters and 0.924 GFLOPs) enables seamless deployment on commercial mobile devices, achieving real-time image enhancement at

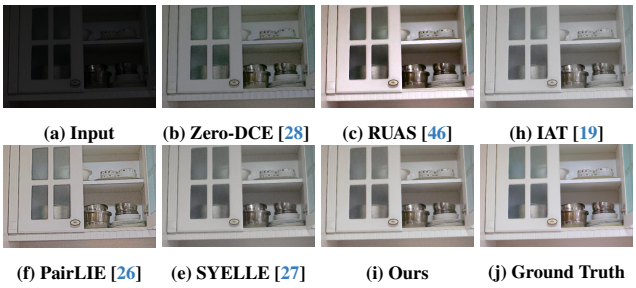


Figure 5. Visualization comparison of LLE on LOLv1 [75].

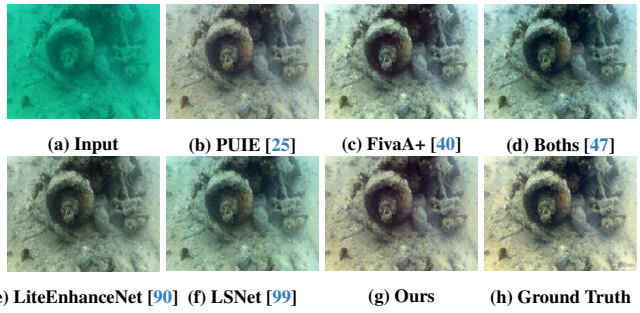


Figure 6. Visualization comparison of UIE on UIEB [43].

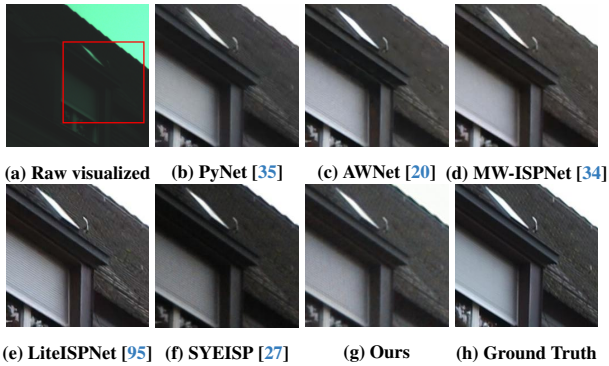


Figure 7. Visualization comparison of ISP on ZRR [35].

Method	Venue	#Params↓ (K)	Latency↓ (GPU,ms)	Latency↓ (SoC,ms)	FPS↑ (448×448)	ZRR [35] PSNR↑	SSIM↑	SCORE↑
PyNet [35]	CVPRW'20	47,548	118.339	>500	<10	21.19	0.747	0.024
AWNet(raw) [20]	ECCVW'20	45,408	102.932	>500	<10	21.42	0.748	0.038
MW-ISP [34]	ECCVW'20	29,200	/	/	/	21.16	0.732	/
LiteISP [95]	ICCV'21	11,900	46.326	>500	21.334	21.28	0.739	0.070
NAFNet [9]	ECCV'22	7.844	3.108	78.62	321.760	21.12	0.736	0.836
SYEISP [27]	ICCV'23	5.616	1.156	16.47	865.247	20.84	0.728	1.524
FourierISP [31]	AAAI'24	7,590	22.869	>500	43.731	21.65	0.755	0.237
Ours	/	4.104	1.020	14.40	980.516	21.43	0.731	3.913

Table 3. Performance comparison of different image signal processing models on ZRR datasets. SCORE [37] represents a comprehensive measure of model performance and efficiency. The top results are marked: best in red and second in blue.

over 100 FPS. This efficiency paves the way for real-time UHD images (2K-8K) enhancement on mobile platforms.

4.3. Ablation Studies and Analyses

We conducted a series of ablation studies to verify the effectiveness of MobileIE's modules. First, MBRConv was compared with other re-parameterization methods [23, 89, 92], and LVW loss was replaced by conventional loss functions. As shown in Table 4, both MBRConv and LVW loss significantly improved the model's performance. Importantly, MBRConv merges into standard convolutions during inference, adding no computational or memory overhead.

(1) Why does MBRConv + IWO work?

IWO enhances the kernel skeletons. To explore the impact of IWO on MBRConv kernels, we visualized the

Datasets	UIEB [43]				
Metrics	PSNR↑	SSIM↑	LPIPS↓	LOE↓	MAE↓
Only inference network	21.48	0.887	0.192	0.108	0.086
$L1$ loss	22.20	0.902	0.168	0.098	0.078
$L2$ loss	21.74	0.894	0.175	0.095	0.083
Smooth $L1$ loss	22.27	0.904	0.164	0.097	0.078
Charbonnier loss	22.31	0.905	0.162	0.097	0.077
Robust loss [4]	22.34	0.905	0.167	0.099	0.077
LVW loss	**22.57**	**0.906**	**0.160**	**0.097**	**0.075**
Datasets	LOLv1 [75]+LOLv2-Real [82]				
RepVGG [23]	22.69	0.821	0.202	0.257	0.089
RepNAS [89]	21.90	0.810	0.210	0.259	0.099
ECBSR [92]	23.96	0.816	0.204	0.240	0.080
MBRConv(No BN)	23.27	0.821	0.232	0.258	0.082
MBRConv(No IWO)	24.02	0.823	0.199	0.257	0.078
Ours	**24.35**	**0.829**	**0.189**	**0.254**	**0.074**

Table 4. Performance comparison of Loss and Re-param modules.

weight increments ΔW in the last convolutional layer of the MobileIE model trained on the UIEB dataset. Here, ΔW is defined as the difference between the weights trained with IWO (W_{IWO}) and those trained without it (W_{base}), i.e., $\Delta W = W_{IWO} - W_{base}$. As shown in Figure 8, most ΔW values reinforce the center-row and center-column characteristics of the convolutional kernels, which contributes to improved model performance [6, 21].

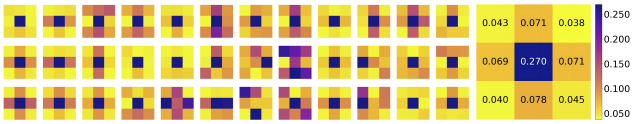


Figure 8. Visualization of kernel differences pre- and post-IWO.

IWO reduces channel redundancy, validated by comparing the last Conv1×1 layer of MBRConv5×5 before and after IWO optimization. Channel similarity was measured using Kullback-Leibler (KL) [6, 73, 100] divergence, with higher values indicating lower redundancy. We trained MobileIE on the LOLv1, saving checkpoints at different stages. Applying softmax across channels and calculating KL divergence between channel pairs, we generated KL matrices at different training stages (Figure 9). As training progressed, KL divergence increased, and after IWO, it rose significantly, underscoring MBRConv's effectiveness in reducing redundancy and improving feature representation.

Impact of IWO on Training Convergence. Similar to the Pre-training stage, W_{pre} is the well-performing weight obtained after the first 1000 epochs. As shown in the training curves (Figure 10), without IWO, the loss stagnates in the later training stages, whereas with IWO, it continues to decrease. This demonstrates that IWO facilitates further optimization of the weights, enhancing the model's convergence performance. The oscillation is due to cosine decay.

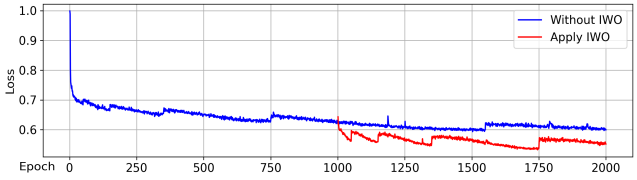
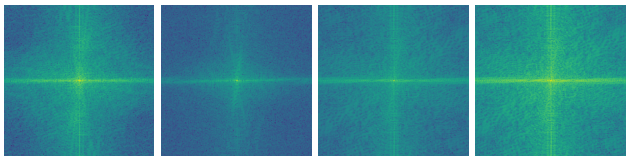


Figure 10. Training loss curves showing the effect of IWO.

(2) FST can enhance nonlinear feature interaction. The FST uses a nonlinear squaring transformation to enhance high-frequency sensitivity, highlighting image details and edges (Figure 11). In the Fourier spectrum, squaring shows a stronger high-frequency response than ReLU, effectively preserving detail. The learnable parameters, $Scale$, and $bias$, further adaptively adjust the feature dynamic range, enhancing restoration performance.



(a) Original Feature:F **(b) ReLU(F)** **(c) $F * F$** **(d) FST(F)**

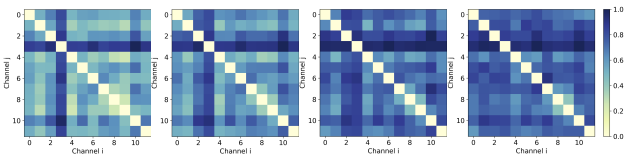
Figure 11. Feature Transformation Fourier Spectrum.

Figure 12 shows that the FST achieves a faster convergence rate during training compared to other feature transform methods. In the initial epochs, the FST rapidly reduces loss, demonstrating its ability to enhance feature representation early and accelerate convergence. Additionally, the FST's final stabilized loss is lower than other methods, highlighting its robustness and effectiveness in complex tasks. Table 5 further confirms that FST delivers the best feature transform performance on the UIEB [43] dataset.

Methods	PSNR↑	SSIM↑	LPIPS↓	LOE↓	MAE↓
Baseline	21.12	0.884	0.199	0.101	0.088
ADD	21.50	0.896	0.178	0.107	0.084
Scale∗ADD	21.34	0.893	0.173	0.105	0.087
ADD+Bias	21.58	0.890	0.177	0.093	0.086
CAT	21.44	0.893	0.181	0.095	0.085
MUL	21.80	0.901	0.173	0.095	0.084
MUL+Bias	21.84	0.901	0.168	0.092	0.081
Scale∗MUL	21.95	0.902	0.167	0.092	0.081
Ours	**22.60**	**0.906**	**0.157**	**0.092**	**0.075**

Table 5. Performance comparison of different feature transform. "Baseline" represents the model without any feature transform.

(3) Dual-path fusion enables precise feature capture. How can one achieve both efficiency and precise capture of local and global feature dependencies? Following



(a) Epoch = 100 **(b) Epoch = 500** **(c) Epoch = 1,000** **(d) Applying IWO**

Figure 9. KL similarity matrix for the last convolutional layer of MBRConv at different training stages.

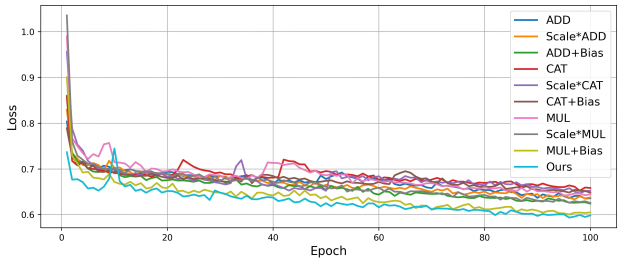
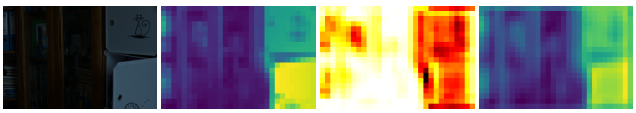


Figure 12. Training loss comparison of different feature transform.

Occam's Razor principle, HDPA adopts a streamlined design language, utilizing a dual-path structure and adaptive feature interaction to decompose and aggregate features hierarchically, thereby extracting critical information across different scales. This dual-path design also supports mutual optimization during backpropagation.



(a) RGB image **(b) Input feature** **(c) HDPA heat map** **(d) Output feature**

Figure 13. Visualization of feature maps at different stages.

Figure 13 illustrates HDPA's effectiveness. The initial convolution preserves key edges and structures, while the HDPA heatmap highlights important regions, focusing attention on key details. This produces a sharper output feature map, demonstrating how HDPA's attention improves image enhancement by refining key details.

Method	#Params↓	FLOPs↓	Latency↓	Latency↓	LOLv1 [35]			
	(K)	(G)	(GPU,ms)	(SoC,ms)	PSNR↑	SSIM↑	LPIPS↓	SCORE↑
SE-Net[33]	4.179	0.922	0.948	7.41	22.27	0.804	0.217	17.259
CBAM[77]	4.254	0.945	6.573	95.71	22.38	0.796	0.231	1.556
ECA-Net[70]	**3.870**	0.922	0.805	6.65	21.79	0.799	0.221	9.886
EA[62]	4.491	1.057	1.232	11.45	19.93	0.779	0.260	0.436
NAM[49]	3.891	0.930	0.905	7.02	20.10	0.761	0.252	0.900
SCA[9]	4.023	0.922	**0.769**	6.33	20.82	0.786	0.250	2.707
DFC[64]	3.979	**0.921**	0.890	**6.29**	21.50	0.802	0.208	6.992
EMA[54]	4.239	1.086	1.569	9.79	20.78	0.792	0.233	1.656
SPAN[67]	5.511	1.308	1.835	10.24	19.79	0.767	0.260	0.401
CGA[13]	6.330	1.486	1.437	9.62	22.33	0.803	0.209	14.447
LKA[3]	5.247	1,236	1.082	8.96	19.76	0.782	0.227	0.440
Ours	4.035	0.924	0.886	6.72	**23.26**	**0.805**	**0.213**	**75.079**

Table 6. Performance comparison of different Attention mechanisms on LOLv1 [75] dataset. SCORE [37] reflects the efficiency.

Finally, we replaced it with mainstream attention mechanisms and conducted ablation experiments, as shown in Table 6. With minimal added parameters (+0.168 K) and FLOPs (+5.76 M), HDPA achieved a PSNR gain of +3.08 dB. Tests on mobile devices confirmed HDPA's high performance with minimal latency, underscoring the efficiency of our edge-focused IE design in MobileIE.

(4) Effectiveness of LVW in Outlier Optimization. We observed that the pixel error histograms follow a Laplace-like distribution, suggesting the model optimizes accurate pixels while neglecting outliers [27]. Based on $L1$ Loss, LVW adjusts weights to increase the constraint on outliers. Normalization standardizes errors, bringing the data distribution closer to uniform and ensuring balanced pixel weights. Additionally, we visualized the pixel prediction error histograms for different loss functions, as shown in Figure 14. LVW exhibits lower mean and median pixel errors, improving effectiveness in optimizing challenging regions.

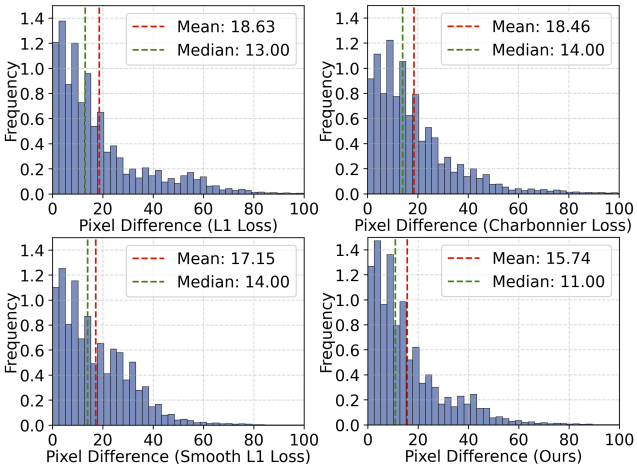


Figure 14. Visualization of pixel prediction error histograms for various loss functions.

(5) Module-Wise Lightweight Contributions. Table 7 demonstrates that re-parameterization significantly reduces model complexity during inference. IWO and LVW introduce no extra cost, making them suitable for efficient deployment. Re-parameterization combined with cost-free strategies enables efficient and lightweight inference.

Model	Training	w/o FST	w/o HDPA	Inference	w/o FST	w/o HDPA	IWO	LVW
Params (K)	47.211	47.211	43,971	4.047	4.047	3.867		
Model Size (MB)	0.18	0.18	0.17	0.02	0.02	0.01	Cost-Free	
FLOPs (G)	11.042	11.042	17.099	0.924	0.924	0.919	Optimization	
Latency (ms)	9.58	9.52	8.99	0.895	0.864	0.732		

Table 7. Ablation on different module efficiency.

5. Conclusion

In this paper, we introduce MobileIE, a lightweight CNN with only 4K parameters designed for real-time image enhancement on mobile devices. MobileIE combines simplicity with high performance across low-light image enhancement, underwater image enhancement, and image signal processing tasks. Key components such as the MBRConv, Feature Self-Transform module, Hierarchical Dual-Path Attention mechanism, and Local Variance-Weighted loss enable MobileIE to achieve superior results while maintaining an inference time of around 0.9 ms. Furthermore, MobileIE can be easily deployed on any commercial mobile device with minimal computational overhead and efficient resource utilization. Future work will explore optimizations to expand MobileIE's application to more different tasks.

Acknowledgements. This work was supported by the Key Program for International Cooperation of Ministry of Science and Technology, China (No.2024YFE0100700).

References

[1] Lusine Abrahamyan, Valentin Ziatchin, Yiming Chen, and Nikos Deligiannis. Bias loss for mobile neural networks. In *ICCV*, pages 6556–6566, 2021. 1, 4

[2] Guanhua An, Ao He, Yudong Wang, and Jichang Guo. Uw-mamba: Underwater image enhancement with state space model. *IEEE Signal Processing Letters*, 2024. 2

[3] Reza Azad, Leon Niggemeier, Michael Hüttemann, Amirhossein Kazerouni, Ehsan Khodapanah Aghdam, Yury Velichko, Ulas Bagci, and Dorit Merhof. Beyond self-attention: Deformable large kernel attention for medical image segmentation. In *WACV*, pages 1287–1297, 2024. 8

[4] Jonathan T Barron. A general and adaptive robust loss function. In *CVPR*, pages 4331–4339, 2019. 7

[5] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *ICCV*, pages 12504–12513, 2023. 2

[6] Zhicheng Cai, Xiaohan Ding, Qiu Shen, and Xun Cao. Refconv: Re-parameterized refocusing convolution for powerful convnets. *arXiv preprint arXiv:2310.10563*, 2023. 7

[7] Hanting Chen, Yunhe Wang, Jianyuan Guo, and Dacheng Tao. Vanillanet: the power of minimalism in deep learning. *NeurIPS*, 36, 2024. 2

[8] Jierun Chen, Shiu-hong Kao, Hao He, Weipeng Zhuo, Song Wen, Chul-Ho Lee, and S-H Gary Chan. Run, don't walk: chasing higher flops for faster neural networks. In *CVPR*, pages 12021–12031, 2023. 2

[9] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *ECCV*, pages 17–33. Springer, 2022. 1, 6, 8

[10] Xiangyu Chen, Ruiwen Zhen, Shuai Li, Xiaotian Li, and Guanghui Wang. Mofa: A model simplification roadmap for image restoration on mobile devices. In *ICCV*, pages 1322–1332, 2023. 1

[11] Yuantao Chen, Runlong Xia, Kai Yang, and Ke Zou. Gcam: lightweight image inpainting via group convolution and attention mechanism. *International Journal of Machine Learning and Cybernetics*, 15(5):1815–1825, 2024. 1

[12] Zhihao Chen and Yiyuan Ge. Mambauie&sr: Unraveling the ocean's secrets with only 2.8 flops. *arXiv preprint arXiv:2404.13884*, 2024. 2

[13] Zixuan Chen, Zewei He, and Zhe-Ming Lu. Dea-net: Single image dehazing based on detail-enhanced convolution and content-guided attention. *IEEE TIP*, 2024. 8

[14] Haram Choi, Cheolwoong Na, Jihyeon Oh, Seungjae Lee, Jinseop Kim, Subeen Choe, Jeongmin Lee, Taehoon Kim, and Jihoon Yang. Reciprocal attention mixing transformer for lightweight image restoration. In *CVPR*, pages 5992–6002, 2024. 1

[15] Marcos V Conde, Eduard Zamfir, Radu Timofte, Daniel Motilla, Cen Liu, Zexin Zhang, Yunbo Peng, Yue Lin, Jiaming Guo, Xueyi Zou, et al. Efficient deep models for real-time 4k image super-resolution. ntire 2023 benchmark and report. In *CVPR*, pages 1495–1521, 2023. 2

[16] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Focal network for image restoration. In *ICCV*, pages 13001–13011, 2023. 1

[17] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Image restoration via frequency selection. *IEEE TPAMI*, 2023.

[18] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Revitalizing convolutional network for image restoration. *IEEE TPAMI*, 2024. 1

[19] Ziteng Cui, Kunchang Li, Lin Gu, Shenghan Su, Peng Gao, Zhengkai Jiang, Yu Qiao, and Tatsuya Harada. You only need 90k parameters to adapt light: a light weight transformer for image enhancement and exposure correction. In *BMVC*, 2022. 5, 6

[20] Linhui Dai, Xiaohong Liu, Chengqi Li, and Jun Chen. Awnet: Attentive wavelet network for image isp. In *ECCVW*, pages 185–201. Springer, 2020. 2, 6

[21] Xiaohan Ding, Yuchen Guo, Guiguang Ding, and Jungong Han. Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks. In *ICCV*, pages 1911–1920, 2019. 2, 3, 7

[22] Xiaohan Ding, Xiangyu Zhang, Jungong Han, and Guiguang Ding. Diverse branch block: Building a convolution as an inception-like unit. In *CVPR*, pages 10886–10895, 2021. 2

[23] Xiaohan Ding, Xiangyu Zhang, Ningning Ma, Jungong Han, Guiguang Ding, and Jian Sun. Repvgg: Making vgg-style convnets great again. In *CVPR*, pages 13733–13742, 2021. 2, 3, 6, 7

[24] Paulo LJ Drews, Erickson R Nascimento, Silvia SC Botelho, and Mario Fernando Montenegro Campos. Underwater depth estimation and image restoration based on single images. *IEEE computer graphics and applications*, 36(2):24–35, 2016. 2

[25] Zhenqi Fu, Wu Wang, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Uncertainty inspired underwater image enhancement. In *ECCV*, pages 465–482. Springer, 2022. 6

[26] Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Learning a simple low-light image enhancer from paired low-light instances. In *CVPR*, pages 22252–22261, 2023. 2, 5, 6

[27] Weiran Gou, Ziyao Yi, Yan Xiang, Shaoqing Li, Zibin Liu, Dehui Kong, and Ke Xu. Syenet: A simple yet effective network for multiple low-level vision tasks with real-time performance on mobile device. In *ICCV*, pages 12182–12195, 2023. 3, 4, 5, 6, 8

[28] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *CVPR*, pages 1780–1789, 2020. 2, 5, 6

[29] Kai Han, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, and Chang Xu. Ghostnet: More features from cheap operations. In *CVPR*, pages 1580–1589, 2020. 2

[30] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *CVPR*, pages 16000–16009, 2022. 5

[31] Xuanhua He, Tao Hu, Guoli Wang, Zejin Wang, Run Wang, Qian Zhang, Keyu Yan, Ziyi Chen, Rui Li, Chengjun Xie, et al. Enhancing raw-to-srgb with decoupled style structure in fourier domain. In *AAAI*, pages 2130–2138, 2024. 6

[32] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *ICCV*, pages 1314–1324, 2019. 2

[33] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR*, pages 7132–7141, 2018. 8

[34] Andrey Ignatov, Radu Timofte, Zhilu Zhang, Ming Liu, Haolin Wang, Wangmeng Zuo, Jiawei Zhang, Ruimao Zhang, Zhanglin Peng, Sijie Ren, et al. Aim 2020 challenge on learned image signal processing pipeline. In *ECCVW*, pages 152–170. Springer, 2020. 2, 6

[35] Andrey Ignatov, Luc Van Gool, and Radu Timofte. Replacing mobile camera isp with a single deep learning model. In *CVPRW*, pages 536–537, 2020. 2, 5, 6, 8

[36] Andrey Ignatov, Cheng-Ming Chiang, Hsien-Kai Kuo, Anastasia Sycheva, and Radu Timofte. Learned smartphone isp on mobile npus with deep learning, mobile ai 2021 challenge: Report. In *CVPR*, pages 2503–2514, 2021. 2

[37] Andrey Ignatov, Radu Timofte, Shuai Liu, Chaoyu Feng, Furui Bai, Xiaotao Wang, Lei Lei, Ziyao Yi, Yan Xiang, Zibin Liu, et al. Learned smartphone isp on mobile gpus with deep learning, mobile ai & aim 2022 challenge: report. In *ECCVW*, pages 44–70. Springer, 2022. 2, 5, 6, 8

[38] Md Jahidul Islam, Youya Xia, and Junaed Sattar. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 5(2):3227–3234, 2020. 2, 6

[39] Hai Jiang, Ao Luo, Haoqiang Fan, Songchen Han, and Shuaicheng Liu. Low-light image enhancement with wavelet-based diffusion models. *TOG*, 42(6):1–14, 2023. 2

[40] Jingxia Jiang, Tian Ye, Jinbin Bai, Sixiang Chen, Wenhao Chai, Shi Jun, Yun Liu, and Erkang Chen. Five a+ network: You only need 9k parameters for underwater image enhancement. In *BMVC*, 2023. 2, 6

[41] Raqib Khan, Priyanka Mishra, Nancy Mehta, Shruti S Phutke, Santosh Kumar Vipparthi, Sukumar Nandi, and Subrahmanyam Murala. Spectroformer: Multi-domain query cascaded transformer network for underwater image enhancement. In *WACV*, pages 1454–1463, 2024. 2

[42] Ao Li, Le Zhang, Yun Liu, and Ce Zhu. Feature modulation transformer: Cross-refinement of global representation via high-frequency prior for image super-resolution. In *ICCV*, pages 12514–12524, 2023. 1

[43] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE TIP*, 29:4376–4389, 2019. 5, 6, 7

[44] Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE TPAMI*, 44(8):4225–4238, 2021. 5

[45] Yawei Li, Yulun Zhang, Radu Timofte, Luc Van Gool, Lei Yu, Youwei Li, Xinpeng Li, Ting Jiang, Qi Wu, Mingyan Han, et al. Ntire 2023 challenge on efficient super-resolution: Methods and results. In *CVPR*, pages 1922–1960, 2023. 2

[46] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *CVPR*, pages 10561–10570, 2021. 2, 5, 6

[47] Xu Liu, Sen Lin, Kaichen Chi, Zhiyong Tao, and Yang Zhao. Boths: Super lightweight network-enabled underwater image enhancement. *IEEE Geoscience and Remote Sensing Letters*, 20:1–5, 2023. 6

[48] Xiaoning Liu, Zongwei Wu, Ao Li, Florin-Alexandru Vasluianu, Yulun Zhang, Shuhang Gu, Le Zhang, Ce Zhu, Radu Timofte, Zhi Jin, et al. Ntire 2024 challenge on low light image enhancement: Methods and results. *arXiv preprint arXiv:2404.14248*, 2024. 1

[49] Yichao Liu, Zongru Shao, Yueyang Teng, and Nico Hoffmann. Nam: Normalization-based attention module. *arXiv preprint arXiv:2111.12419*, 2021. 8

[50] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *CVPR*, pages 5637–5646, 2022. 5

[51] Xu Ma, Xiyang Dai, Yue Bai, Yizhou Wang, and Yun Fu. Rewrite the stars. In *CVPR*, pages 5694–5703, 2024. 2, 4

[52] Ziyin Ma and Changjae Oh. A wavelet-based dual-stream network for underwater image enhancement. In *ICASSP*, pages 2769–2773, 2022. 2, 6

[53] Ankita Naik, Apurva Swarnakar, and Kartik Mittal. Shallow-uwnet: Compressed model for underwater image enhancement (student abstract). In *AAAI*, pages 15853–15854, 2021. 2, 6

[54] Daliang Ouyang, Su He, Guozhong Zhang, Mingzhu Luo, Huaiyong Guo, Jian Zhan, and Zhijie Huang. Efficient multi-scale attention module with cross-spatial learning. In *ICASSP*, pages 1–5, 2023. 8

[55] Lintao Peng, Chunli Zhu, and Liheng Bian. U-shape transformer for underwater image enhancement. *IEEE TIP*, 32: 3066–3079, 2023. 1, 2, 6

[56] Georgy Perevozchikov, Nancy Mehta, Mahmoud Afifi, and Radu Timofte. Rawformer: Unpaired raw-to-raw translation for learnable camera isps. *arXiv preprint arXiv:2404.10700*, 2024. 2

[57] Nianzu Qiao, Changyin Sun, and Lu Dong. Semi-supervised feature distillation and unsupervised domain adversarial distillation for underwater image enhancement. *IEEE TCSVT*, 2024. 1

[58] Jingxiang Qu, Ryan Wen Liu, Yuan Gao, Yu Guo, Fenghua Zhu, and Fei-Yue Wang. Double domain guided real-time low-light image enhancement for ultra-high-definition transportation surveillance. *IEEE Transactions on Intelligent Transportation Systems*, 2024. 5

[59] Nareddy Kartheek Kumar Reddy, Mani Madhoolika Bulusu, Praveen Kumar Pokala, and Chandra Sekhar Seelamantula. Quantized proximal averaging networks for compressed image recovery. In *CVPR*, pages 4633–4643, 2023. 1

[60] Bin Ren, Yawei Li, Nancy Mehta, Radu Timofte, Hongyuan Yu, Cheng Wan, Yuxin Hong, Bingnan Han, Zhuoyuan Wu, Yajun Zou, et al. The ninth ntire 2024 efficient super-resolution challenge report. In *CVPR*, pages 6595–6631, 2024. 2

[61] Prasen Sharma, Ira Bisht, and Arijit Sur. Wavelength-based attributed deep neural network for underwater image restoration. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(1):1–23, 2023. 2

[62] Zhuoran Shen, Mingyuan Zhang, Haiyu Zhao, Shuai Yi, and Hongsheng Li. Efficient attention: Attention with linear complexities. In *WACV*, pages 3531–3539, 2021. 8

[63] Xiangsheng Shi, Xuefei Ning, Lidong Guo, Tianchen Zhao, Enshu Liu, Yi Cai, Yuhan Dong, Huazhong Yang, and Yu Wang. Memory-oriented structural pruning for efficient image restoration. In *AAAI*, pages 2245–2253, 2023. 1

[64] Yehui Tang, Kai Han, Jianyuan Guo, Chang Xu, Chao Xu, and Yunhe Wang. Ghostnetv2: Enhance cheap operation with long-range attention. *NeurIPS*, 35:9969–9982, 2022. 8

[65] Yi Tang, Hiroshi Kawasaki, and Takafumi Iwaguchi. Underwater image enhancement by transformer-based diffusion model with non-uniform sampling for skip strategy. In *ACM MM*, pages 5419–5427, 2023. 2

[66] Pavan Kumar Anasosalu Vasu, James Gabriel, Jeff Zhu, Oncel Tuzel, and Anurag Ranjan. Mobileone: An improved one millisecond mobile backbone. In *CVPR*, pages 7907–7917, 2023. 1, 2

[67] Cheng Wan, Hongyuan Yu, Zhiqi Li, Yihang Chen, Yajun Zou, Yuqing Liu, Xuanwu Yin, and Kunlong Zuo. Swift parameter-free attention network for efficient super-resolution. In *CVPR*, pages 6246–6256, 2024. 8

[68] Ao Wang, Hui Chen, Zijia Lin, Jungong Han, and Guiguang Ding. Repvit: Revisiting mobile cnn from vit perspective. In *CVPR*, pages 15909–15920, 2024. 2

[69] Cong Wang, Jinshan Pan, Wei Wang, Gang Fu, Siyuan Liang, Mengzhu Wang, Xiao-Ming Wu, and Jun Liu. Correlation matching transformation transformers for uhd image restoration. In *AAAI*, pages 5336–5344, 2024. 1

[70] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *CVPR*, pages 11534–11542, 2020. 8

[71] Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Bjorn Stenger, and Tong Lu. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *AAAI*, pages 2654–2662, 2023. 2

[72] William Y Wang, Lisa Liu, and Pingping Cai. Adversarially regularized low-light image enhancement. In *International Conference on Multimedia Modeling*, pages 230–243. Springer, 2024. 5

[73] Xudong Wang and X Yu Stella. Tied block convolution: Leaner and better cnns with shared thinner filters. In *AAAI*, pages 10227–10235, 2021. 7

[74] Xintao Wang, Chao Dong, and Ying Shan. Repsr: Training efficient vgg-style super-resolution networks with structural re-parameterization and batch normalization. In *ACM MM*, pages 2556–2564, 2022. 3

[75] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *BMVC*, 2018. 5, 6, 7, 8

[76] Yanjie Wen, Ping Xu, Zhihong Li, and Wangtu Xu ATO. An illumination-guided dual attention vision transformer for low-light image enhancement. *PR*, page 111033, 2024. 2

[77] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *ECCV*, pages 3–19, 2018. 8

[78] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *CVPR*, pages 5901–5910, 2022. 2

[79] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. In *ICCV*, pages 13095–13105, 2023. 1

[80] Xiaogang Xu, Shu Kong, Tao Hu, Zhe Liu, and Hujun Bao. Boosting image restoration via priors from pre-trained models. In *CVPR*, pages 2900–2909, 2024. 1

[81] Hailong Yan, Junjian Huang, and Tingwen Huang. Igdnet: Zero-shot robust underexposed image enhancement via illumination-guided and denoising. *arXiv preprint arXiv:2507.02445*, 2025. 2

[82] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *CVPR*, pages 3063–3072, 2020. 5, 7

[83] Juncheol Ye, Hyunho Yeo, Jinwoo Park, and Dongsu Han. Accelir: Task-aware image compression for accelerating neural restoration. In *CVPR*, pages 18216–18226, 2023. 1

[84] Mingxin Yi, Kai Zhang, Pei Liu, Tanli Zuo, and Jingduo Tian. Diffraw: Leveraging diffusion model to generate dslr-comparable perceptual quality srgb from smartphone raw images. In *AAAI*, pages 6711–6719, 2024. 1, 2

[85] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *ICCV*, pages 12302–12311, 2023. 2

[86] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *ECCV*, pages 492–511. Springer, 2020. 1

[87] Hui Zeng, Jianrui Cai, Lida Li, Zisheng Cao, and Lei Zhang. Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. *IEEE TPAMI*, 44(4):2058–2073, 2020. 5

[88] Jiangning Zhang, Xiangtai Li, Jian Li, Liang Liu, Zhucun Xue, Boshen Zhang, Zhengkai Jiang, Tianxin Huang, Yabiao Wang, and Chengjie Wang. Rethinking mobile block for efficient attention-based models. In *ICCV*, pages 1389–1400, 2023. 1, 2

[89] Mingyang Zhang, Xinyi Yu, Jingtao Rong, and Linlin Ou. Repnas: Searching for efficient re-parameterizing blocks. In *ICME*, pages 270–275, 2023. 6, 7

[90] Song Zhang, Shili Zhao, Dong An, Daoliang Li, and Ran Zhao. Liteenhancenet: A lightweight network for real-time single underwater image enhancement. *Expert Systems with Applications*, 240:122546, 2024. 6

[91] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *CVPR*, pages 6848–6856, 2018. 2

[92] Xindong Zhang, Hui Zeng, and Lei Zhang. Edge-oriented convolution block for real-time super resolution on mobile devices. In *ACM MM*, pages 4034–4043, 2021. 3, 6, 7

[93] Xuanqi Zhang, Haijin Zeng, Jinwang Pan, Qiangqiang Shen, and Yongyong Chen. Llemamba: Low-light enhancement via relighting-guided mamba with deep unfolding network. *arXiv preprint arXiv:2406.01028*, 2024. 2

[94] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *IJCV*, 129:1013–1037, 2021. 5

[95] Zhilu Zhang, Haolin Wang, Ming Liu, Ruohao Wang, Jiawei Zhang, and Wangmeng Zuo. Learning raw-to-srgb mappings with inaccurately aligned supervision. In *ICCV*, pages 4348–4358, 2021. 2, 6

[96] Chen Zhao, Weiling Cai, Chenyu Dong, and Chengwei Hu. Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration. In *CVPR*, pages 8281–8291, 2024. 1, 2

[97] Chen Zhao, Weiling Cai, Chenyu Dong, and Ziqi Zeng. Toward sufficient spatial-frequency interaction for gradient-aware underwater image enhancement. In *ICASSP*, pages 3220–3224, 2024. 6

[98] Shen Zheng and Gaurav Gupta. Semantic-guided zero-shot learning for low-light image/video enhancement. In *WACV*, pages 581–590, 2022. 5

[99] Fuheng Zhou, Dikai Wei, Ye Fan, Yulong Huang, and Yonggang Zhang. A 7k parameter model for underwater image enhancement based on transmission map prior. *arXiv preprint arXiv:2405.16197*, 2024. 6

[100] Yuefu Zhou, Ya Zhang, Yanfeng Wang, and Qi Tian. Accelerate cnn via recursive bayesian pruning. In *ICCV*, pages 3306–3315, 2019. 7

[101] Peixian Zhuang, Jiamin Wu, Fatih Porikli, and Chongyi Li. Underwater image enhancement with hyper-laplacian reflectance priors. *IEEE TIP*, 31:5442–5455, 2022. 2