DIVE-Net: Depthwise Multi-Dilation Vision Network for Efficient Underwater Image Enhancement

Sourav Phogat[1], Koyyada Dinesh Kumar[1], and Sujit Kumar Sahoo[1]

School of Electrical Sciences,
Indian Institute of Technology Goa, India
{sourav2414205,koyyada21242104,sujit}@iitgoa.ac.in

Abstract. Underwater image enhancement is challenging due to wavelength dependent attenuation, scattering, color distortion, and loss of local structural details. Recent CNN- and Transformer-based methods have improved restoration quality, but many of them rely on large restoration backbones with high parameter counts and computational cost. This limits their practicality for underwater surveillance, marine robotics, and embedded vision platforms. In this paper, we propose **DIVE-Net**, a Depthwise Multi-Dilation Vision Network for efficient underwater image enhancement. DIVE-Net follows a compact NAFNet-inspired restoration design, but replaces single-scale depthwise spatial mixing with a lightweight multi-scale depthwise convolution module. The proposed module uses three parallel depthwise branches with dilation rates of 1, 2, and 4, and combines them using learnable softmax weights to aggregate local, intermediate, and wider contextual information. The network is organized as a compact multi-scale encoder–decoder with residual output prediction at multiple resolutions. A learnable output residual scaling factor is further used to control the correction added to each scale output. Experiments on the paired UIEB test set show that DIVE-Net achieves 22.32 dB PSNR and 0.893 SSIM with only 0.74M parameters and 5.49G FLOPs. Compared with substantially larger underwater enhancement models, the proposed method provides a strong accuracy–efficiency trade-off. Results on unpaired U45 and C60 datasets further indicate competitive generalization without dataset-specific fine-tuning.

Keywords: Underwater image enhancement · Image restoration · Depthwise convolution · Multi-dilation aggregation · Lightweight network

1 Introduction

Underwater image enhancement aims to recover a visually clear and color-consistent image from degraded underwater observations. This task is important for marine robotics, underwater inspection, ecological monitoring, and visual perception in aquatic environments [1]. However, underwater images commonly suffer from color cast, low contrast, reduced visibility, haze-like degradation, and loss of fine structural details. These degradations mainly arise from

wavelength-dependent attenuation, scattering, and background veiling light [2–4]. Since longer wavelengths are absorbed more rapidly in water, the red channel is usually attenuated more strongly than the green and blue channels, causing blue-green dominance and poor chromatic fidelity [3, 4].

A commonly used underwater image formation model can be written as

$$I(\mathbf{x}) = J(\mathbf{x}) \odot T(\mathbf{x}) + B_\infty \odot (1 - T(\mathbf{x})), \tag{1}$$

where $I(\mathbf{x})$ is the observed underwater image, $J(\mathbf{x})$ is the latent clean image, $T(\mathbf{x})$ is the transmission map, B_∞ denotes the global background light, and $\odot$ represents element-wise multiplication. Although this model provides a useful physical description, recovering $J(\mathbf{x})$ from a single underwater image is highly ill-posed because both $T(\mathbf{x})$ and B_∞ are unknown and the degradation varies with scene depth, illumination, and water type [2, 4].

Traditional underwater enhancement methods use physical priors, depth or transmission estimation, wavelength compensation, color correction, and fusion strategies [5, 6]. These methods are efficient and interpretable, but their performance can degrade when the assumed priors fail under non-uniform illumination, strong scattering, or severe color distortion. Learning-based methods reduce this dependence on hand-crafted assumptions by learning degraded-to-clean mappings from data. CNN- and GAN-based methods such as Water-Net [7] and FUnIE-GAN [1] have shown the effectiveness of data-driven underwater enhancement, while recent Transformer-based methods improve representation capacity through long-range and channel-spatial modeling [8–10]. However, many high-performing models rely on large backbones, increasing parameters, FLOPs, memory usage, and inference cost. This can limit practical deployment in underwater surveillance, marine robotics, and embedded vision systems.

This motivates the question: *can a compact network achieve competitive underwater enhancement by using more effective spatial feature aggregation instead of a heavy backbone?* Underwater degradation is inherently multi-scale. Local receptive fields are useful for restoring edges, textures, and object boundaries, while wider receptive fields help correct color cast, scattering, and background veiling effects. Standard depthwise convolution is computationally efficient [18], but it performs spatial mixing at only one receptive-field scale. Dilated convolution can enlarge the receptive field without downsampling [19], but a single dilation rate is still limited for diverse underwater degradation.

In this work, we propose **DIVE-Net**, a Depthwise Multi-Dilation Vision Network for efficient underwater image enhancement. DIVE-Net follows a compact NAFNet-inspired residual and gated restoration design [17], but replaces single-scale depthwise spatial mixing with a Depthwise Multi-Dilation Aggregation (DMDA) block. The DMDA block uses three parallel depthwise branches with dilation rates of 1, 2, and 4 to capture local, intermediate, and wider contextual information. The branch outputs are combined using learnable softmax weights and projected using a shared pointwise convolution. Since only the depthwise spatial filtering operation is replicated, the additional parameter overhead remains small.

The complete DIVE-Net adopts a lightweight multi-scale encoder–decoder with skip fusion and residual output prediction at multiple resolutions. Lower-resolution outputs encourage coarse color and contrast correction, while the final high-resolution output focuses on fine-detail restoration. A learnable output residual scaling factor further controls the correction added to each scale output. The main contributions are summarized as follows:

- We propose **DIVE-Net**, a compact underwater image enhancement network designed to improve the accuracy–efficiency trade-off without relying on a heavy restoration backbone.
- We introduce a **Depthwise Multi-Dilation Aggregation** block that replaces single-scale depthwise spatial mixing with learnable multi-dilation aggregation using dilation rates of 1, 2, and 4.
- We integrate the proposed block into a NAFNet-inspired residual and gated restoration framework with multi-scale residual output prediction and learnable output residual scaling.
- We evaluate the proposed method on paired and unpaired underwater benchmarks and compare it with traditional and recent learning-based methods.

2 Related Work

2.1 Underwater Image Enhancement

Underwater image enhancement methods can be broadly grouped into traditional model-based methods and learning-based methods. Model-based approaches describe underwater degradation using attenuation and background veiling light, and then estimate physical quantities such as transmission, scene depth, or background light for restoration [2–4]. Wavelength compensation and depth-estimation-based methods improve visibility by compensating channel-dependent attenuation [3, 5], while fusion-based methods combine multiple enhanced images using perceptual weights to improve contrast, saturation, and color balance [6]. Although these methods are interpretable, their performance depends on the reliability of hand-crafted priors and may degrade in complex underwater scenes.

Learning-based methods formulate underwater enhancement as a data-driven restoration problem. FUnIE-GAN introduced a fast enhancement framework for underwater visual perception [1], while Water-Net and the UIEB dataset provided an important benchmark for supervised underwater image enhancement [7]. More recent methods use stronger architectures, including U-Shape Transformer [8], UDAformer [9], and AquaClarity [10], to model long-range dependencies and channel-spatial interactions. Trinity-Net is also used as a strong restoration comparison because of its gradient-guided Swin Transformer design [13]. These methods improve restoration quality, but they often increase model size and computational complexity. In contrast, DIVE-Net focuses on improving spatial feature aggregation within a compact network.

2.2 Lightweight Multi-Scale Feature Aggregation

Lightweight restoration is important for resource-constrained underwater platforms. Depthwise separable convolution reduces computation by separating spatial filtering and channel mixing [18], and NAFNet showed that a simple residual and gated design can be highly effective for image restoration [17]. However, a standard depthwise convolution uses a single receptive-field scale. Multi-scale context is important for underwater enhancement because fine details require local information, whereas color cast and scattering correction require broader spatial context. Dilated convolution increases the receptive field without reducing resolution [19], but using only one dilation rate limits contextual diversity. The proposed DMDA block addresses this by using parallel depthwise branches with dilation rates of 1, 2, and 4 and learnable softmax weighting, enabling adaptive multi-scale spatial aggregation with low computational overhead.

3 Proposed Method

3.1 Overall Architecture

Fig. 1 shows the overall architecture of the proposed DIVE-Net. Given an underwater input image $I \in [0,1]^{3\times H\times W}$, the network predicts enhanced outputs at three resolutions. The input image is first padded using reflection padding so that the spatial size is compatible with the encoder–decoder downsampling factor. The network also generates downsampled inputs $I_{1/2}$ and $I_{1/4}$ using bilinear interpolation. These downsampled images are used as base images for residual prediction at lower scales.

The model contains a compact stem, three encoder stages, a bottleneck stage, two decoder stages, skip fusion modules, and three output heads. Let F_1, F_2, and F_3 denote the encoder features at three resolutions. The shallow feature is first extracted as

$$F_1 = \mathcal{E}_1(\mathcal{S}(I)), \tag{2}$$

where $\mathcal{S}(\cdot)$ denotes the stem and $\mathcal{E}_1(\cdot)$ denotes the first encoder block stack. The deeper encoder features are obtained as

$$F_2 = \mathcal{E}_2(\mathcal{D}_1(F_1)), \quad F_3 = \mathcal{E}_3(\mathcal{D}_2(F_2)), \tag{3}$$

where $\mathcal{D}_1$ and $\mathcal{D}_2$ denote downsampling layers. The bottleneck output is processed and decoded progressively as

$$D_3 = \mathcal{B}_3(F_3), \tag{4}$$

$$D_2 = \mathcal{B}_2\left(\mathcal{F}_2(\mathcal{U}_3(D_3), F_2)\right), \quad D_1 = \mathcal{B}_1\left(\mathcal{F}_1(\mathcal{U}_2(D_2), F_1)\right), \tag{5}$$

where $\mathcal{U}_i$ denotes upsampling, $\mathcal{F}_i$ denotes skip fusion, and $\mathcal{B}_i$ denotes decoder block stacks.

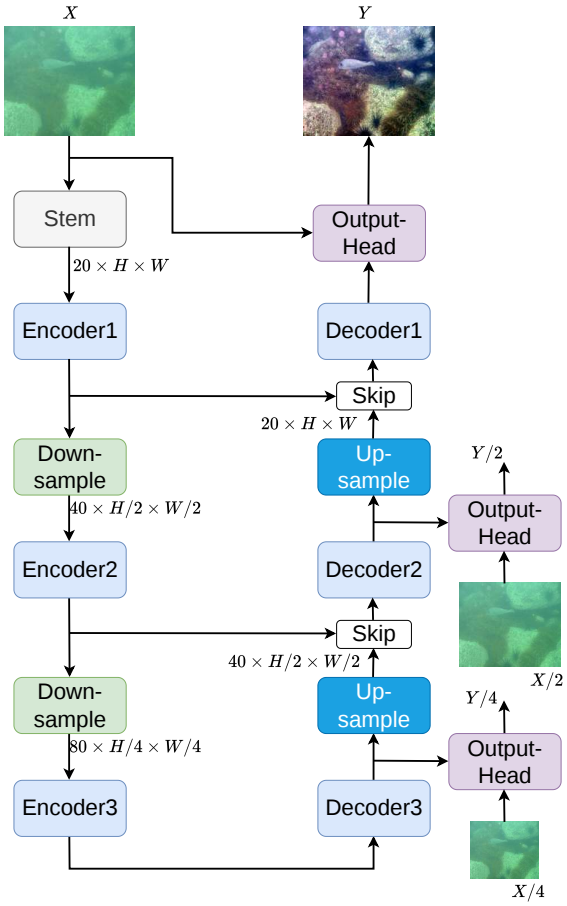

Fig. 1. Overall architecture of the proposed DIVE-Net. The input image is processed by a compact encoder–decoder network with three encoder stages, two decoder stages, skip fusion, and multi-scale output heads. The network predicts residual-enhanced outputs at 1/4, 1/2, and full resolutions, enabling coarse color/contrast correction at lower scales and fine-detail restoration at the final scale.

The three output heads predict residual corrections over the corresponding base images. For compact notation, let $\mathcal{S} = \{1, 1/2, 1/4\}$ denote the output scales. The residual prediction at scale s is written as

$$\hat{J}_s = I_s + \alpha_s \odot \mathcal{H}_s(D_s), \quad s \in \mathcal{S}, \tag{6}$$

where I_s is the input image resized to scale s, D_s is the corresponding decoder feature, $\mathcal{H}_s(\cdot)$ is the output head, and α_s is a learnable channel-wise residual scaling factor. The final enhanced image is $\hat{J}_1$, while $\hat{J}_{1/2}$ and $\hat{J}_{1/4}$ are used only for multi-scale supervision during training.

3.2 Depthwise Multi-Dilation Aggregation Block

The main component of the proposed network is the Depthwise Multi-Dilation Aggregation (DMDA) block. A standard depthwise convolution applies spatial filtering using a single receptive-field scale. In underwater enhancement, this can be restrictive because different degradations require different spatial contexts. Fine structures such as edges and textures benefit from local filtering, whereas color cast, scattering, and background veiling effects require broader contextual information. To address this limitation, the proposed DMDA block uses three parallel depthwise branches with dilation rates of 1, 2, and 4.

Given an input feature $X \in \mathbb{R}^{C\times H\times W}$, the three branch responses are computed as

$$Y_1 = \mathrm{DW}_{3\times3}^{d=1}(X), \quad Y_2 = \mathrm{DW}_{3\times3}^{d=2}(X), \quad Y_4 = \mathrm{DW}_{3\times3}^{d=4}(X), \tag{7}$$

where $\mathrm{DW}_{3\times3}^{d}$ denotes a depthwise 3×3 convolution with dilation rate d.

$$\mathbf{w} = \mathrm{softmax}(\mathbf{a}), \tag{8}$$

where $w_1 + w_2 + w_4 = 1$. The aggregated feature is then computed as

$$Y = w_1Y_1 + w_2Y_2 + w_4Y_4. \tag{9}$$

Finally, a shared pointwise projection mixes channel information:

$$\mathrm{DMDA}(X) = \phi\left(\mathrm{PW}_{1\times1}(Y)\right), \tag{10}$$

where $\mathrm{PW}_{1\times1}$ denotes a pointwise convolution and $\phi(\cdot)$ is the GELU activation. Since only the depthwise spatial filtering operation is replicated and the pointwise projection is shared, the block captures multi-scale context with limited parameter overhead.

3.3 Restoration Block

The restoration block follows a lightweight residual design inspired by NAFNet [17]. Given a feature X, the first branch performs normalized feature extraction using pointwise expansion, simple gating, DMDA-based spatial mixing, channel attention, and pointwise projection. The second branch acts as a compact feed-forward refinement module. Unlike a standard NAFNet-style block that uses a single depthwise convolution for spatial mixing, the proposed block uses the DMDA module to capture local, intermediate, and wider underwater degradation context. Fig. 2 shows the internal structure of the restoration block.

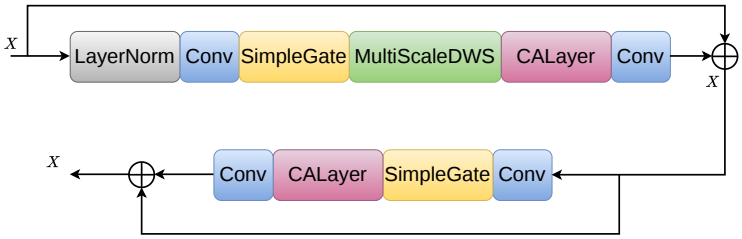

Fig. 2. Structure of the proposed restoration block. The block follows a NAFNet-inspired residual and gated design, where single-scale depthwise spatial mixing is replaced by the proposed DMDA module.

The block is written as

$$X' = X + \beta \odot \mathcal{M}(\mathrm{LN}(X)), \tag{11}$$

$$Z = X' + \gamma \odot \mathcal{G}(\mathrm{LN}(X')), \tag{12}$$

where $\mathcal{M}(\cdot)$ is the DMDA-based spatial mixing branch, $\mathcal{G}(\cdot)$ is the gated feed-forward branch, $\mathrm{LN}(\cdot)$ is channel-wise layer normalization, and β and γ are learnable residual scaling parameters. These residual scaling parameters are initialized to zero, allowing the network to gradually learn stable residual corrections during training.

3.4 Multi-Scale Output Supervision

The network predicts enhanced images at 1/4, 1/2, and full resolutions. This design encourages the deeper stages to correct coarse color and contrast degradation, while the final decoder stage focuses on fine spatial details. During training, each predicted image is compared with the corresponding downsampled ground truth. The total training loss is computed over all output scales as

$$\mathcal{L} = \sum_{s \in \mathcal{S}} \left(\mathcal{L}^s_{\mathrm{rec}} + \lambda_{\mathrm{freq}} \mathcal{L}^s_{\mathrm{freq}} \right), \tag{13}$$

where $\mathcal{L}^s_{\mathrm{rec}} = \|\hat{J}_s - J_s\|_1$ and $\mathcal{L}^s_{\mathrm{freq}} = \|\mathcal{F}(\hat{J}_s) - \mathcal{F}(J_s)\|_1$, with $\mathcal{F}(\cdot)$ the 2-D DFT. All scales are weighted equally, and $\lambda_{\mathrm{freq}} = 0.1$. Here, J_s is the ground-truth image downsampled to scale s.

4 Experiments and Results

4.1 Datasets and Evaluation Metrics

We evaluate the proposed method on paired and unpaired underwater image enhancement benchmarks. For supervised training and full-reference evaluation,

we use the paired subset of UIEB [7]. Following the standard split, 790 paired images are used for training and 100 paired images are used for testing. The paired UIEB test set is evaluated using PSNR and SSIM [14], where higher values indicate better restoration fidelity.

To evaluate generalization to unseen underwater images, we further test the trained model on U45 [12] and C60 [7]. U45 contains 45 real underwater images with common underwater degradations, while C60 denotes the 60 challenging unpaired images from UIEB. No additional training or fine-tuning is performed on U45 or C60. Since reference images are unavailable for these datasets, we report UCIQE [16] and NIQE [15]. Higher UCIQE indicates better underwater color image quality, while lower NIQE indicates better naturalness.

4.2 Implementation Details

The proposed DIVE-Net is trained using the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The batch size is set to 8, and the model is trained for 500 epochs. The initial learning rate is set to 8×10^{-4} and is gradually reduced to 1×10^{-6} using cosine annealing with a linear warm-up over the first 3 epochs. Gradients are clipped to a maximum ℓ_2 norm of 0.001 at each training step. The learnable output residual scaling factor α_s in each output head is initialised to 0.1, while the residual scaling parameters β and γ inside each restoration block are initialised to zero, allowing the network to gradually learn stable residual corrections during training.

For fair comparison, the learning-based competing methods, including UDAformer [9], Trinity-Net [13], and AquaClarity [10], are retrained on the same UIEB training split using their reported training settings and loss functions. The traditional methods, Single [5] and Color-Fusion [6], are directly evaluated on the same test images. All reported PSNR and SSIM values on UIEB are computed on the same 100-image test split. For U45 and C60, the UIEB-trained models are directly tested without any dataset-specific fine-tuning. Parameters and FLOPs are reported under the same evaluation setting for all learning-based methods, and FLOPs are computed for an input resolution of 256×256.

4.3 Quantitative Comparison on UIEB

Table 1 reports the full-reference comparison on the paired UIEB test set. Traditional methods such as Single [5] and Color-Fusion [6] obtain lower PSNR and SSIM, indicating the limitation of fixed enhancement priors under diverse underwater conditions. Recent learning-based methods improve the restoration quality but often require larger models and higher computational cost.

DIVE-Net achieves the highest PSNR among the compared methods while using only 0.74M parameters and 5.49G FLOPs. Compared with AquaClarity [10], DIVE-Net improves PSNR from 21.92 dB to 22.32 dB, while reducing the parameter count by about 54.7× and the computational cost by about 13.1×. Although AquaClarity obtains a slightly higher SSIM, DIVE-Net remains very close with 0.893 SSIM. In terms of efficiency, DIVE-Net uses only 7.7% of the parameters

Table 1. Full-reference comparison on the paired UIEB test set. Best and second-best results are shown in bold and underlined, respectively.

Method	Params	FLOPs	PSNR↑	SSIM↑
Single [5]	–	–	15.56	0.577
Color-Fusion [6]	–	–	17.05	0.610
UDAformer [9]	9.60M	41.59G	21.83	0.884
Trinity-Net [13]	9.02M	17.79G	20.07	0.890
AquaClarity [10]	40.47M	71.71G	<u>21.92</u>	**0.894**
DIVE-Net	**0.74M**	**5.49G**	**22.32**	<u>0.893</u>

Table 2. No-reference comparison on U45 and C60. Best and second-best results are shown in bold and underlined, respectively.

Method	U45		C60	
	UCIQE↑	NIQE↓	UCIQE↑	NIQE↓
Single [5]	0.60	4.16	**0.59**	5.31
UDAformer [9]	0.60	4.20	0.56	4.35
Trinity-Net [13]	0.58	4.03	0.56	5.70
AquaClarity [10]	**0.61**	<u>3.83</u>	<u>0.57</u>	**3.87**
DIVE-Net	**0.61**	**3.80**	<u>0.57</u>	<u>3.92</u>

of UDAformer and 1.8% of the parameters of AquaClarity, while still achieving the best PSNR on UIEB. These results show that depthwise multi-dilation aggregation provides a strong accuracy–efficiency trade-off without relying on a heavy restoration backbone.

4.4 No-Reference Evaluation on U45 and C60

Table 2 reports the no-reference comparison on U45 and C60. On U45, DIVE-Net obtains the best NIQE score and matches the best UCIQE score. On C60, DIVE-Net achieves the second-best NIQE score and remains competitive in UCIQE. These results suggest that the proposed model generalizes reasonably well to unpaired underwater images without dataset-specific fine-tuning.

No-reference metrics should be interpreted carefully because they emphasize different perceptual properties. UCIQE mainly reflects underwater color quality, contrast, and saturation, whereas NIQE measures natural image statistics. Therefore, the no-reference results are used as complementary evidence together with the full-reference UIEB results and qualitative comparisons.

4.5 Ablation Study

We design the ablation study to examine the effect of each progressive design choice in DIVE-Net. The baseline uses a plain residual block and applies the loss

Table 3. Ablation study of DIVE-Net on the paired UIEB test set. HR denotes loss applied only at the final high-resolution output, MS denotes multi-scale loss, and ORS denotes learnable output residual scaling.

Model	Block / Mixing	Loss	ORS	Params	PSNR↑	SSIM↑
B0	ResBlock	HR	×	0.631M	18.87	0.792
B1	RestBlock + DWS	HR	×	0.714M	21.76	0.875
B2	RestBlock + MS-DWS	HR	×	0.731M	22.05	0.887
B3	RestBlock + MS-DWS	MS	×	0.744M	22.19	0.890
DIVE-Net	RestBlock + MS-DWS	MS	✓	**0.744M**	**22.32**	**0.893**

DWS denotes depthwise separable convolution, and MS-DWS denotes the proposed multi-scale depthwise convolution.

only at the final high-resolution output. The next variant replaces the plain residual block with a NAFNet-inspired restoration block using standard depthwise separable convolution. We then introduce the proposed multi-scale depthwise convolution to test whether multiple receptive-field scales improve underwater restoration. Next, we add multi-scale loss to supervise the outputs at different resolutions. Finally, the full model introduces a learnable residual scaling factor in each output head before adding the predicted residual to the corresponding base image.

Table 3 shows a consistent improvement as each component is introduced. Replacing the plain residual block with the restoration block improves PSNR from 18.87 dB to 21.76 dB, showing the importance of the NAFNet-inspired residual and gated design. Replacing standard DWS with MS-DWS further improves PSNR from 21.76 dB to 22.05 dB and SSIM from 0.875 to 0.887, confirming the benefit of multi-dilation spatial aggregation. Adding multi-scale supervision improves the result to 22.19 dB and 0.890 SSIM, indicating that intermediate-scale outputs help the model learn coarse color and contrast correction. Finally, learnable output residual scaling gives the best result of 22.32 dB and 0.893 SSIM without increasing the parameter count.

4.6 Qualitative Comparison

Fig. 3 shows representative visual comparisons on paired and unpaired underwater images. The paired UIEB examples include reference images, so PSNR and SSIM values are reported below each enhanced output. The U45 example is unpaired; therefore, the comparison focuses on visual quality, color balance, and contrast restoration. Overall, DIVE-Net improves visibility and color consistency while preserving local scene structures.

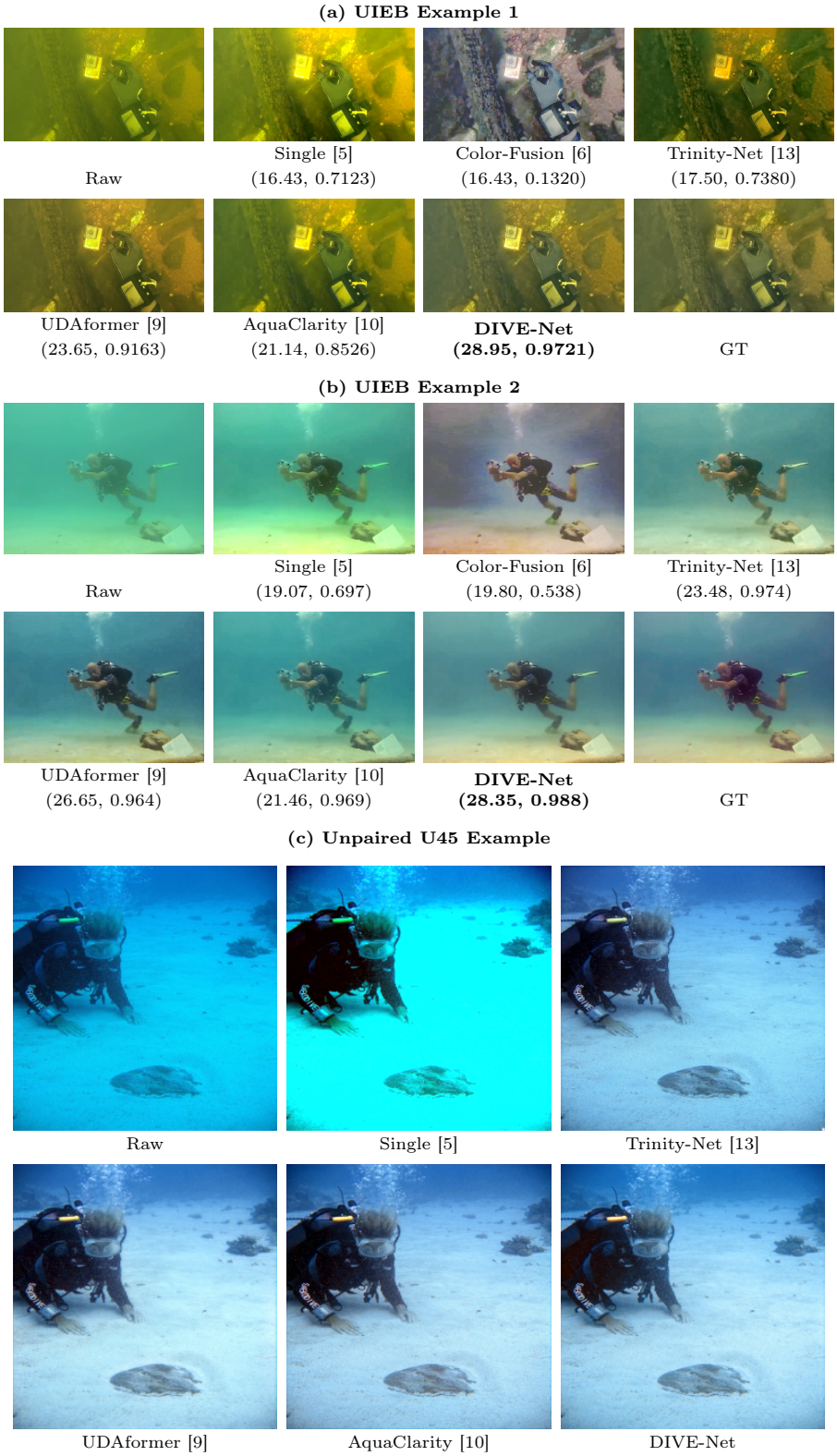

Fig. 3. Qualitative comparison on two paired UIEB examples and one unpaired U45 example. The values below each enhanced UIEB image denote PSNR and SSIM for the corresponding example. DIVE-Net provides stronger color correction, contrast enhancement, and structural preservation while maintaining a lightweight design.

5 Conclusion

This paper presented DIVE-Net, a lightweight underwater image enhancement network based on depthwise multi-dilation aggregation. The proposed DMDA block uses three parallel depthwise branches with dilation rates of 1, 2, and 4 and combines them using learnable softmax weights. This allows the network to capture local, intermediate, and wider contextual information while keeping the parameter overhead small. The complete model follows a compact multi-scale encoder–decoder structure and predicts residual enhanced images at multiple resolutions. Experiments on paired and unpaired underwater benchmarks demonstrate that DIVE-Net provides a favorable trade-off between restoration quality and computational efficiency. Ablation studies further verify the contribution of multi-dilation depthwise aggregation and learnable branch weighting.

References

1. Islam, M.J., Xia, Y., Sattar, J.: Fast Underwater Image Enhancement for Improved Visual Perception. IEEE Robotics and Automation Letters **5**(2), 3227–3234 (2020)
2. Schechner, Y.Y., Karpel, N.: Recovery of Underwater Visibility and Structure by Polarization Analysis. IEEE Journal of Oceanic Engineering **30**(3), 570–587 (2005)
3. Chiang, J.Y., Chen, Y.C.: Underwater Image Enhancement by Wavelength Compensation and Dehazing. IEEE Transactions on Image Processing **21**(4), 1756–1769 (2012)
4. Akkaynak, D., Treibitz, T.: A Revised Underwater Image Formation Model. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6723–6732 (2018)
5. Peng, Y.T., Zhao, X., Cosman, P.C.: Single Underwater Image Enhancement Using Depth Estimation Based on Blurriness. In: IEEE International Conference on Image Processing (ICIP), pp. 4952–4956 (2015)
6. Ancuti, C.O., Ancuti, C., De Vleeschouwer, C., Bekaert, P.: Color Balance and Fusion for Underwater Image Enhancement. IEEE Transactions on Image Processing **27**(1), 379–393 (2018)
7. Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., Tao, D.: An Underwater Image Enhancement Benchmark Dataset and Beyond. IEEE Transactions on Image Processing **29**, 4376–4389 (2020)
8. Peng, L., Zhu, C., Bian, L.: U-Shape Transformer for Underwater Image Enhancement. IEEE Transactions on Image Processing **32**, 3066–3079 (2023)
9. Shen, Z., Xu, H., Luo, T., Song, Y., He, Z.: UDAformer: Underwater Image Enhancement Based on Dual Attention Transformer. Computers & Graphics **111**, 77–88 (2023)
10. Chen, X., Li, Z., Zhang, X., Lei, Y., Dong, Y., Zhang, X., Pun, C.M.: AquaClarity: Vision Transformer Enhances Underwater Image via Nonlinear Feature Fusion and Color Fidelity Improvement. IEEE Transactions on Consumer Electronics **72**(1), 1122–1129 (2026)
11. He, K., Sun, J., Tang, X.: Guided Image Filtering. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(6), 1397–1409 (2013)
12. Li, H., Li, J., Wang, W.: A Fusion Adversarial Underwater Image Enhancement Network with a Public Test Dataset. CoRR abs/1906.06819 (2019)

13. Chi, K., Yuan, Y., Wang, Q.: Trinity-Net: Gradient-Guided Swin Transformer-Based Remote Sensing Image Dehazing and Beyond. IEEE Transactions on Geoscience and Remote Sensing **61**, 1–14 (2023)
14. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
15. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a Completely Blind Image Quality Analyzer. IEEE Signal Processing Letters **20**(3), 209–212 (2013)
16. Yang, M., Sowmya, A.: An Underwater Color Image Quality Evaluation Metric. IEEE Transactions on Image Processing **24**(12), 6062–6071 (2015)
17. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple Baselines for Image Restoration. In: European Conference on Computer Vision Workshops (ECCVW), pp. 17–33 (2022)
18. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint arXiv:1704.04861 (2017)
19. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(4), 834–848 (2018)