Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 34(9): 6121–6152.
DOI: 10.3934/era.2026270
Received: 20 March 2026
Revised: 17 June 2026
Accepted: 29 June 2026
Published: 15 July 2026

Research article

Multi-stage dynamic wavelet gated aggregation network for underwater image denoising

Wei Li[1], Haiyan Xie[1,*], Jiaxi Li[1] and Huiyu Zhou[2]

[1] School of Science, Dalian Maritime University, Dalian 116026, China

[2] School of Computing and Mathematical Sciences, University of Leicester, Leicester, LE1 7RH, United Kingdom

* **Correspondence:** Email: xiehy@dlmu.edu.cn.

Abstract: Underwater image quality is inevitably degraded by the selective absorption and scattering of light, which introduce severe blurring and abrupt high-frequency noise. These interferences obscure latent structural information and make robust restoration a significant challenge. To address this, we present MWGANet, a multi-stage dynamic wavelet-gated aggregation network designed to decouple complex underwater noise distributions caused by environmental disturbances. The core of our architecture was a parallel strategy with three pathways tailored for feature extraction across domains. Specifically, the Gated Wavelet Transform Branch (GWTB) isolated noise spikes in the frequency domain to effectively safeguard critical structural textures from corruption. Our adaptive reuse branch (ARB) incorporated a dynamic aggregation block at multiple scales to facilitate feature flow across layers while adaptively filtering redundant information through learnable gating mechanisms. Furthermore, the dilated context branch (DCB) expanded the model's receptive field via graduated dilation rates to enhance the perception of diverse object priors. Extensive evaluations confirmed that MWGANet achieves a superior balance between denoising fidelity and computational efficiency. In particular, the model maintained a high-fidelity output on the EUVP dataset, achieving a PSNR of 31.06 dB even under severe synthetic additive white Gaussian noise (AWGN) ($\sigma = 50$). Such robustness and efficiency indicated that MWGANet has the potential to serve as a practical real-time solution for resource-constrained autonomous underwater vehicle (AUV) platforms.

Keywords: underwater image denoising; wavelet transform; attention mechanism; multi-scale aggregation; feature reuse

1. Introduction

In recent years, the rapid advancement of marine engineering has propelled underwater vision to the forefront of oceanic research, providing indispensable information for non-invasive marine exploration [1]. High-quality visual data are a prerequisite for autonomous underwater vehicles (AUVs) to perform complex tasks, including archaeology, search-and-rescue operations, and biological monitoring [2]. In practical AUV applications, restoration models are also expected to be computationally efficient since onboard perception systems often operate under limited memory and computing resources. However, underwater images are inherently affected by wavelength-selective absorption and light scattering, which induce chromatic shifts, contrast attenuation, and visibility degradation [3, 4]. In addition, suspended particles and sensor-level disturbances may further introduce local artifacts and stochastic perturbations into the captured signal. Although considerable research has been devoted to underwater color correction, enhancement, and dehazing [5, 6], underwater image denoising as a standalone restoration problem has received relatively limited attention.

Underwater image denoising is particularly challenging because noise components are often intertwined with weak structures and fine textures in visually degraded scenes [7]. In low-contrast regions, noise-induced high-frequency fluctuations may overlap with weak edges, object boundaries, and texture details, making it difficult to distinguish stochastic perturbations from meaningful image information. As a result, denoising methods tend to face a trade-off between residual noise and the over-smoothing of important structures [8]. This texture–noise ambiguity is more severe in underwater environments, where illumination is highly non-uniform and structural details are further degraded due to light absorption and scattering effects [6]. Therefore, an effective underwater denoising model should not only suppress noise but also preserve local structures and fine details.

In practical imaging systems, noise formation is influenced by photon statistics and electronic disturbances, which can be modeled by Poisson-Gaussian distributions under low illumination [9]. Real captured images often exhibit more complex noise statistics than synthetic additive white Gaussian noise (AWGN), which makes blind and real-world denoising more difficult [10]. In this work, AWGN is adopted as a simplified and controlled corruption model for quantitative evaluation. Although it does not fully reproduce real underwater noise statistics, AWGN enables standardized comparison across representative denoising methods under scalable noise levels (σ) and provides a controlled benchmark setting for future extensions to more realistic underwater noise and degradation models.

Under this controlled evaluation framework, denoising research has predominantly followed two trajectories: model-based restoration and learning-based approaches [11]. Classical model-based methods, such as non-local means (NLM) and block-matching 3D filtering (BM3D), suppress noise by exploiting statistical priors or transform-domain sparsity [12–15]; however, their underlying assumptions are frequently weakened when underwater image content contains low-contrast structures and spatially varying visual patterns. Consequently, such methods often suffer from residual noise or over-smoothed fine details. On the other hand, learning-based approaches, spanning CNNs, Transformers, and frequency-domain networks, have substantially raised performance benchmarks [7, 16, 17]. Nevertheless, directly applying general-purpose architectures to underwater images may be suboptimal, since they may not adequately address texture-noise ambiguity in underwater scenes, and high-capacity models can introduce considerable parameter overhead and inference latency.

Motivated by these limitations, we propose MWGANet, a lightweight Multi-path Wavelet Gated Aggregation Network engineered for underwater image denoising. Our core design philosophy treats denoising as a specialized signal purification task to jointly prioritize structural fidelity and computational efficiency. By processing features across parallel streams, MWGANet effectively addresses the texture-noise ambiguity without relying on heavy parameter redundancy. With only 0.79 M parameters, our network satisfies the strict power and latency constraints required for real-time perception on resource-constrained AUV platforms.

The main contributions of this paper are summarized as follows:

(1) We present a Gated Wavelet Transform Branch (GWTB) to resolve the inherent spatial ambiguity between stochastic noise and fine-grained textures. By exploiting the sparsity and invertibility of wavelet decomposition, this module achieves precise spectral disentanglement, isolating high-frequency noise components from structural information even in turbid underwater scenes.

(2) We propose a Multi-Scale Dynamic Aggregation Block (MSDAB) integrated with dense blocks within the Adaptive Reuse Branch (ARB) to facilitate efficient cross-layer feature propagation. Equipped with a learnable gating mechanism, the MSDAB adaptively suppresses scattering-induced feature redundancy while preserving critical discriminative cues, thereby maintaining robust representations even in severely degraded regions.

(3) We propose a Tri-pathway Parallel Paradigm to effectively resolve the coupled spatial and frequency-domain perturbations of underwater noise. By integrating a Wavelet Branch for spectral noise suppression, a Dilated Context Branch for multi-scale spatial modeling, and an ARB for progressive feature refinement, MWGANet achieves a favorable trade-off between denoising quality and computational cost.

Experimental results demonstrate that the proposed model not only effectively suppresses noise in underwater images but also proficiently preserves fine-grained details and color fidelity. The rest of this paper is structured as follows: In Section 2, we present the technical foundations and design motivations of MWGANet; in Section 3, we elaborate on the architecture of the proposed MWGANet and its core components; in Section 4, we describe the experimental setup, present a comprehensive analysis of the results and conduct ablation studies; and in Section 5, we conclude this paper and provide an overview of future work.

2. Related works

In this section, we outline the technical foundations and architectural paradigms that underlie the proposed MWGANet. The framework is structured around three core design motivations: wavelet-domain restoration, dynamic feature aggregation with adaptive channel allocation, and hierarchical feature fusion.

2.1. Wavelet-domain restoration

Owing to its spatial-frequency localization and multi-resolution analysis properties [18], the wavelet transform is inherently well-suited for image denoising tasks. The theoretical groundwork was established in [19] through the wavelet thresholding framework, and subsequent advances introduced

adaptive Bayesian thresholding [20] and probabilistic frameworks such as the Gaussian Scale Mixture model [21] to better capture natural image statistics. Nevertheless, these methods predominantly depend on handcrafted priors or computationally demanding parametric assumptions, which may limit their flexibility when dealing with complex image structures and spatially non-uniform visual patterns.

In the context of underwater image denoising, wavelet-based methods have been explored to attenuate noise across frequency scales by exploiting the observation that noise components are often concentrated in high-frequency sub-bands after decomposition. Adaptive wavelet sub-band thresholding has been shown to improve the balance between noise suppression and structural preservation by applying sub-band-specific thresholding strategies [22]. Subsequent studies have reported consistent findings through systematic evaluations of wavelet-based denoising methods for underwater image restoration [23].

However, these methods remain largely dependent on handcrafted thresholding criteria and sequential processing pipelines. In practical imaging systems, stochastic perturbations may arise from photon limitations and electronic sensor disturbances [9]. For underwater images, such perturbations are further entangled with weak edges, low-contrast textures, and spatially varying visual content, making fixed thresholding strategies less effective for detail-preserving denoising.

Moreover, integrating wavelet transforms into deep neural networks has emerged as a compelling paradigm for image restoration. Specifically, the Multi-Level Wavelet Convolutional Neural Network (MWCNN) replaces conventional pooling operators with discrete wavelet transforms to preserve spectral information across deep network layers [24]. Concurrently, predicting wavelet coefficients directly within the network architecture has been demonstrated to facilitate frequency-aware feature learning for super-resolution and image denoising pipelines [25]. Furthermore, the dual adaptive wavelet network (DAWN) incorporates directional attention within wavelet sub-bands to selectively restore degraded frequency components, effectively demonstrating the advantages of content-adaptive wavelet-domain processing [26].

Despite their architectural effectiveness, these pioneer methods predominantly employ static or rigid task-specific sub-band fusion strategies, which lack the spatial flexibility required to dynamically modulate frequency component weighting. Consequently, when transferred to highly non-uniform aquatic degradation fields, these frameworks fail to explicitly adapt their wavelet-domain representations to the localized structural characteristics and heterogeneous scattering profiles inherent to underwater imagery. Pioneering methods predominantly employ static or rigid task-specific sub-band fusion strategies that lack the spatial flexibility needed to dynamically modulate frequency-component weighting.

To address these limitations, we propose a Wavelet Gating Mechanism that dynamically weights frequency sub-bands. Unlike prior fixed-thresholding or static-fusion schemes, the proposed mechanism adaptively adjusts sub-band contributions according to content-dependent feature responses. This allows the network to suppress noise-related high-frequency disturbances while preserving weak edges and fine textures, leading to improved detail-preserving denoising performance under the controlled noise setting considered in this work.

2.2. Multi-order aggregation and dynamic channel allocation

Studies have shown that gated aggregation with multi-order receptive fields can effectively model complex structural patterns by capturing multi-scale spatial dependencies. The Hierarchical Order Recurrent Network (HorNet) [27] employs recursive gated convolutions to model high-order spatial

interactions, while the Visual Attention Network (VAN) [28] combines large-kernel convolutions with gating for long-range dependency modeling. The Multi-order Gated Aggregation Network (MogaNet) [29] further introduces multi-order gated aggregation, which decomposes channels into groups with increasing receptive fields to enable multi-scale feature interaction. However, its channel-splitting ratio is fixed, implicitly assuming that the relative importance of different receptive fields remains unchanged across inputs.

Several researchers have introduced dynamic selection mechanisms at the branch or group level. The Selective Kernel Network (SKNet) [30] uses attention to select among branches with different kernel sizes, and the Split-Attention Network (ResNeSt) [31] applies split-attention to fuse features within evenly divided channel groups. Moreover, dynamic convolution methods such as Conditionally Parameterized Convolution (CondConv) [32] and Omni-Dimensional Dynamic Convolution (ODConv) [33] adjust convolutional kernels conditioned on the input. However, in these methods, the number of channels assigned to each branch or group is usually predetermined. They dynamically reweight or parameterize features, but do not explicitly adjust the channel distribution ratio among branches with different receptive fields in an instance-dependent manner. This limitation is particularly relevant to underwater image denoising, where weak edges, fine textures, and noise-induced fluctuations are typically unevenly distributed across low-contrast regions. A rigid scale assignment may therefore over-smooth weak structures or leave residual noise in texture-rich areas.

Our design mitigates this rigidity through the MSDAB, an input-conditioned routing mechanism that adaptively partitions the channel budget across multi-order branches. By modulating the representational capacity assigned to different receptive fields on the fly, this architecture ensures that the structural preservation properties scale dynamically with local scene complexities.

2.3. Hierarchical feature fusion

Hierarchical feature fusion integrates representations from multiple network levels to jointly leverage low-level spatial details and deeper contextual information. Early methods, such as Fully Convolutional Networks (FCN) [34], adopted simple element-wise addition or concatenation to combine features from different layers, treating them uniformly. Later, the Dense Convolutional Network (DenseNet) [35] introduced dense connectivity to explicitly reuse features across layers, while attention-based mechanisms such as Squeeze-and-Excitation (SE) [36] and Convolutional Block Attention Module (CBAM) [37] improved fusion selectivity by re-weighting features according to channel or spatial relevance.

In image denoising, hierarchical fusion has been widely adopted to preserve spatial details while effectively capturing multi-scale contextual information. The residual dense network (RDN) [38] combines residual learning with dense connections to fully exploit hierarchical features for image restoration. For underwater image content, preserving structural integrity during feature propagation is challenging because low contrast, non-uniform illumination, and weakened boundaries can make fine structures difficult to distinguish from noise-induced fluctuations. Jiang et al. [39] proposed a deep network for underwater image noise removal that leverages multi-level feature aggregation. However, it reuses preceding features in a relatively uniform manner, without explicitly modulating their contributions according to current feature relevance. Such uniform reuse may limit the adaptability of hierarchical fusion when weak structures and noise-related responses are unevenly distributed across feature levels [40].

The proposed Adaptive Feature Reuse Mechanism (AFRM) circumvents the uniform propagation constraint by conditioning multi-level cross-layer reuse on feature importance and inter-level correlations. This explicit modulation favors the retention of salient structural layouts while progressively filtering out noise-dominated and redundant representations across network depths, with implementation nuances detailed in Section 3.

3. Proposed model

In this paper, we propose MWGANet, a tri-pathway parallel framework designed for underwater image denoising. As illustrated in Figure 1, the network processes noisy feature maps through three concurrent streams: The DCB for capturing multi-scale spatial dependencies, the GWTB for frequency-aware feature decomposition, and the ARB for preserving fine-grained structural textures. This multi-branch parallel design is motivated by the characteristics of underwater image content, where weak structures, fine textures, and noise-induced fluctuations can be intertwined in spatial and frequency domains.

By jointly modeling these complementary dimensions, MWGANet captures more comprehensive spatial-frequency representations than conventional single-stream pipelines. Importantly, the proposed framework remains strictly denoising-oriented and is optimized end-to-end via a pixel-level reconstruction loss to estimate the clean image from its noisy observation. Instead of introducing additional enhancement or dehazing objectives, the network focuses on suppressing noise perturbations while preserving structural details and local textures. Finally, the multi-scale representations generated by the parallel branches are dynamically aggregated to reconstruct the denoised image.

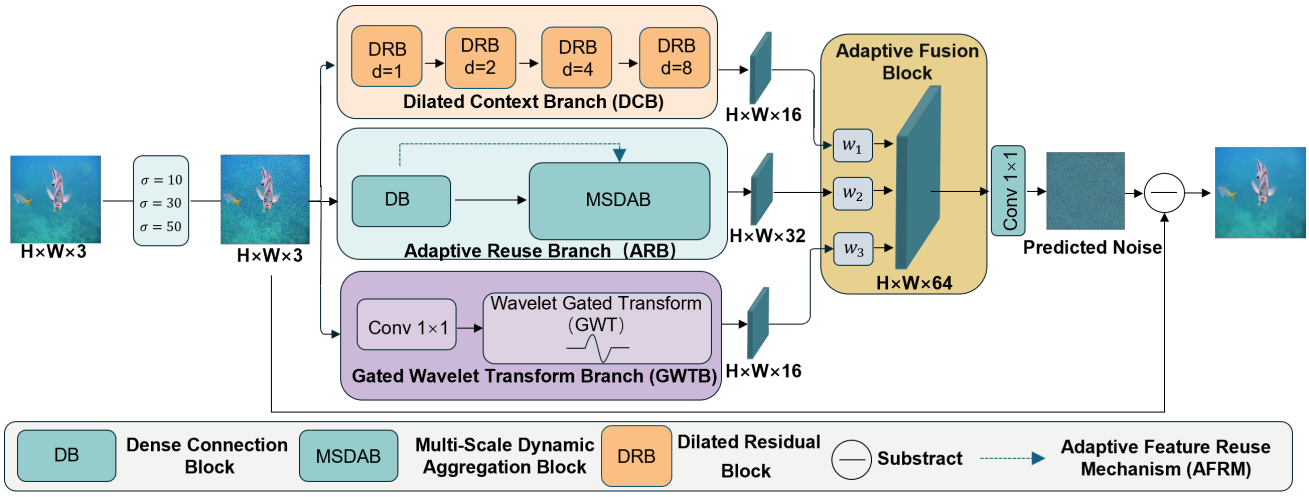

Figure 1. Multi-stage dynamic wavelet gated aggregation network.

Formally, let $I_{in} \in \mathbb{R}^{H\times W\times 3}$ denote the noisy underwater image input. The network first applies a shallow feature extraction layer to obtain the initial feature map F_0. Subsequently, F_0 is fed into three parallel branches to extract complementary representations:

$$F_{DCB} = \mathcal{H}_{DCB}(F_0), \quad F_{ARB} = \mathcal{H}_{ARB}(F_0), \quad F_{GWTB} = \mathcal{H}_{GWTB}(F_0), \tag{3.1}$$

where $\mathcal{H}_{DCB}(\cdot)$, $\mathcal{H}_{ARB}(\cdot)$, and $\mathcal{H}_{GWTB}(\cdot)$ represent the DCB, ARB, and GWTB, respectively. To effectively integrate these features, an Adaptive Fusion Block (AFB) is employed to assign dynamic weights to each branch. The fused feature F_{fuse} is formulated as:

$$F_{fuse} = \mathcal{H}_{AFB}(F_{DCB}, F_{ARB}, F_{GWTB}), \tag{3.2}$$

where $\mathcal{H}_{AFB}(\cdot)$ represents the AFB. The estimated noise map $\mathcal{R}$ is reconstructed via a 1×1 convolution layer. Subsequently, the restored image I_{clean} is obtained by:

$$I_{clean} = I_{in} - \mathcal{R}. \tag{3.3}$$

3.1. Dilated Context Branch (DCB)

To efficiently capture large-scale noise patterns and multi-scale spatial contexts, the first branch is designed as a DCB. As illustrated in Figure 2, the core components of this branch are cascaded Dilated Residual Blocks (DRBs). For an input feature map F_{in}, the dilated convolution with a dilation rate d is formulated as:

$$F_{out}(\mathbf{p}) = \sum_{\mathbf{k}\in\mathcal{K}} W(\mathbf{k}) F_{in}(\mathbf{p} + d \cdot \mathbf{k}), \tag{3.4}$$

where $\mathbf{p}$ denotes the spatial position, $\mathcal{K}$ is the kernel sampling grid, and W represents the learnable convolutional kernels.

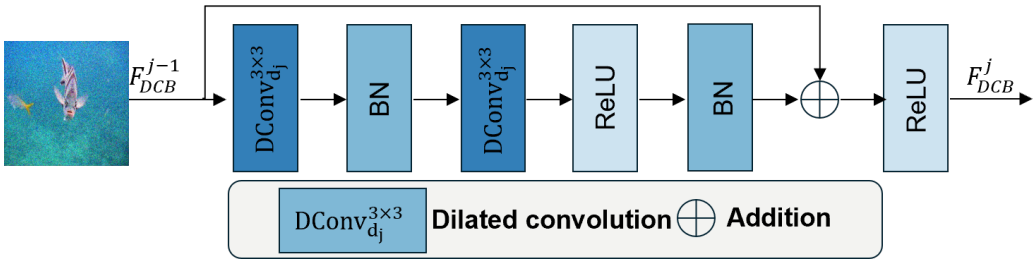

Figure 2. Dilated residual block.

The proposed DCB stacks four DRBs with progressively increasing dilation rates $d_j \in \{1, 2, 4, 8\}$. Specifically, given the branch input F_{DCB}^{0}, the transformation within the j-th DRB is defined as:

$$F_{DCB}^{j} = \mathcal{D}_{d_j}(F_{DCB}^{j-1}) = \mathrm{ReLU}\left(F_{DCB}^{j-1} + \mathrm{BN}\left(\mathrm{DConv}_{d_j}^{3\times3}\left(\mathrm{ReLU}\left(\mathrm{BN}\left(\mathrm{DConv}_{d_j}^{3\times3}(F_{DCB}^{j-1})\right)\right)\right)\right)\right), \tag{3.5}$$

where $\mathcal{D}_{d_j}(\cdot)$ denotes the DRB operator, $\mathrm{DConv}_{d_j}^{3\times3}$ is the 3×3 dilated convolution, and BN and ReLU represent Batch Normalization and Rectified Linear Unit, respectively. The element-wise residual addition within each block facilitates stable gradient propagation and encourages the network to effectively isolate noise-related components. The final output of this branch is denoted as $F_{DCB} = F_{DCB}^{4}$.

Although dilated convolutions effectively enlarge the receptive field without increasing computational overhead, high dilation rates can trigger the gridding effect, leading to sparse spatial sampling and a loss of local textural continuity. To mitigate this, we adopt a progressive dilation strategy with rates $\{1, 2, 4, 8\}$. This configuration ensures that the receptive fields of successive layers complement each other across continuous spatial scales, thereby suppressing gridding artifacts while maintaining a dense and robust contextual representation. The effectiveness of this hierarchical dilation scheme is further substantiated by the ablation analysis presented in Section 4.2.1.

3.2. Dense Connection Block (DB)

The second stream, designated the ARB, is engineered to exploit hierarchical features and fine-grained noise patterns through a cascaded architecture. As illustrated in Figure 3, the core of this branch comprises multiple Dense Connection Blocks (DBs) and MSDABs.

The DB facilitates maximal feature persistence by establishing direct connections between each layer and all its preceding layers. This dense connectivity ensures efficient information flow and encourages feature reusability, which is crucial for preserving subtle textures in complex underwater scenes. Specifically, let $x_0 = F_0$ denote the initial input to the ARB. For the l-th layer within the DB, the newly generated feature map y_l is formulated as:

$$y_l = H_l([x_0, x_1, \ldots, x_{l-1}]), \tag{3.6}$$

where $[\cdot]$ represents the concatenation operation along the channel dimension, and $H_l(\cdot)$ denotes the non-linear transformation function. In our implementation, $H_l(\cdot)$ is defined as a composite sequence of Batch Normalization (BN), Rectified Linear Unit (ReLU), and a 3×3 convolution:

$$H_l(\cdot) = \mathrm{Conv}_{3\times3}(\mathrm{ReLU}(\mathrm{BN}(\cdot))). \tag{3.7}$$

Each layer contributes k new feature maps to the cumulative feature stack. We set the growth rate $k = 16$ to strike a balance between representation capability and computational efficiency. The cumulative feature stack is recursively updated as:

$$x_l = [x_0, x_1, \ldots, x_{l-1}, y_l] \in \mathbb{R}^{(C_{l-1}+k)\times H\times W}, \tag{3.8}$$

where C_{l-1} denotes the channel dimension of the preceding feature stack. After passing through L layers, a transition operation $\tau(\cdot)$ is applied to compress redundant channels and restore the feature dimension:

$$F_{DB} = \tau([x_0, x_1, \ldots, x_L]), \tag{3.9}$$

where the transition function $\tau(\cdot)$ is implemented as a 1×1 convolution followed by normalization and activation:

$$\tau(\cdot) = \mathrm{Conv}_{1\times1}(\mathrm{ReLU}(\mathrm{BN}(\cdot))). \tag{3.10}$$

The resulting feature map F_{DB} is subsequently fed into the MSDAB for multi-scale dynamic aggregation.

3.3. Multi-Scale Dynamic Aggregation Block (MSDAB)

Inspired by the design of MOGA [29], we propose an improved MSDAB to replace the static fusion mechanism, as illustrated in Figure 3. While the original MOGA utilizes a fixed channel splitting strategy, its rigid allocation may fail to adapt to the spatially variant degradation in underwater environments. To address this limitation, we introduce an input-adaptive scaling and capacity reallocation framework within the Value Branch to continuously optimize the representation capability across receptive fields during inference.

The MSDAB takes the input feature map $F_{in} \in \mathbb{R}^{C\times H\times W}$ and first applies a 1×1 projection:

$$F_p = \mathrm{Conv}_{1\times1}(F_{in}). \tag{3.11}$$

To incorporate global context for effective noise disentanglement, the Feature DecomPosition Module (FDM) models the residual between local features and the global mean:

$$F_{dec} = \mathrm{SiLU}\left(F_p + \boldsymbol{\mu} \odot \mathrm{Sig}(F_p - \mathrm{GAP}(F_p))\right), \tag{3.12}$$

where GAP(·) denotes global average pooling, $F_{dec} \in \mathbb{R}^{C\times H\times W}$ denotes the output feature map of the FDM, SiLU(·) and Sig(·) are the activation functions, $\odot$ denotes element-wise multiplication, and $\boldsymbol{\mu} \in \mathbb{R}^C$ is a learnable scaling parameter initialized to zero. This ensures that the module initially acts as an identity mapping to promote training stability.

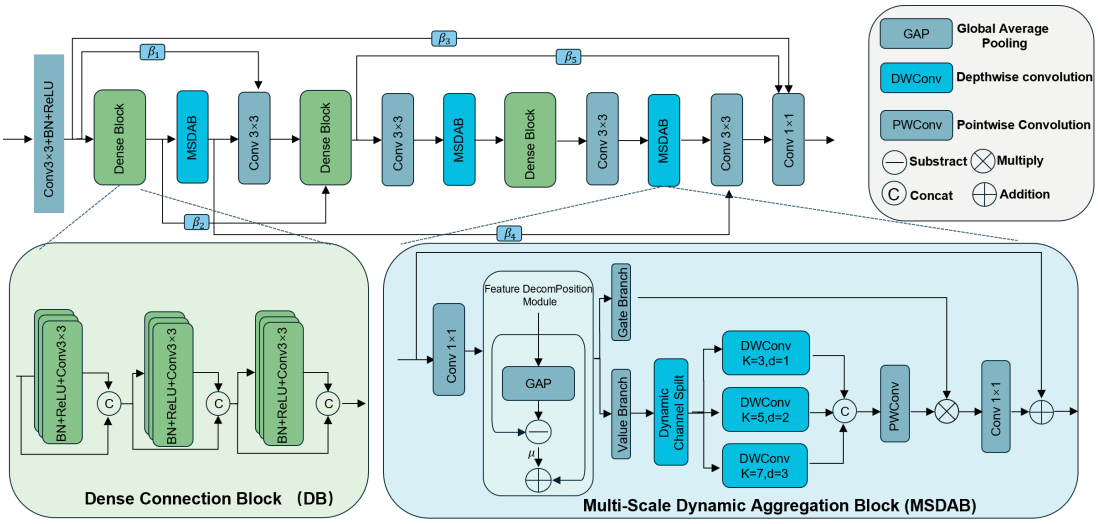

Figure 3. Schematic architectural design of the proposed Multi-Scale Dynamic Aggregation Block (MSDAB).

The decomposed feature $F_{dec} \in \mathbb{R}^{C\times H\times W}$ is subsequently processed by two parallel branches. The Gating Branch generates a modulation map $G = \mathrm{Conv}_{1\times1}(F_{dec})$. Furthermore, the Value Branch performs dynamic channel partitioning to allocate different discrete channel capacities to multi-scale convolutional branches according to the input content.

We first predict a dynamic scale-selection vector $\boldsymbol{\pi} = [\pi_1, \pi_2, \pi_3] \in \mathbb{R}^3$ through a lightweight selection network:

$$\boldsymbol{\pi} = \mathrm{Softmax}\left(\boldsymbol{\Theta}_{scale}(\mathrm{GAP}(F_{dec}))\right), \tag{3.13}$$

where $\boldsymbol{\Theta}_{scale}$ represents the learnable parameters of a lightweight MLP. The scale-selection vector is converted into continuous channel allocations:

$$\hat{C}_k = \pi_k(C - 3C_{min}) + C_{min}, \quad k = 1, 2, 3, \tag{3.14}$$

where $C_{min} = \lfloor 0.1C \rfloor$ is a minimum capacity floor to avoid degenerate allocation, which inherently preserves the total continuous channel budget, i.e., $\sum_{k=1}^{3} \hat{C}_k = C$. Then, the temporary discrete channel number C_k^{hard} for each subgroup is obtained using a residual allocation strategy:

$$C_k^{hard} = \left\lfloor \hat{C}_k \right\rfloor + \mathbb{I}\left(k \in \mathrm{Top}_r\left(\hat{C}_k - \lfloor \hat{C}_k \rfloor\right)\right), \tag{3.15}$$

where $r = C - \sum_k \lfloor \hat{C}_k \rfloor$ represents the total fractional residual, $\mathbb{I}(\cdot)$ is the indicator function, and $\mathrm{Top}_r(\cdot)$ selects the r subgroups with the largest fractional residuals, thereby guaranteeing that $\sum_{k=1}^{3} C_k^{hard} = C$.

During batch-wise computation, to maintain explicit scale-specific feature partitioning, we perform sample-wise channel partitioning along the batch dimension:

$$[F_{dec}^{(1)}, F_{dec}^{(2)}, F_{dec}^{(3)}] = \mathrm{Split}(F_{dec}, [C_1^{hard}, C_2^{hard}, C_3^{hard}]), \tag{3.16}$$

where $F_{dec}^{(k)} \in \mathbb{R}^{C_k^{hard} \times H \times W}$ denotes the split feature tensor for the k-th scale subgroup. Each dynamic subgroup is processed by a corresponding depth-wise convolution $DW_k(\cdot)$, where the kernel size is assigned as $2k + 1$ (yielding kernel sizes of 3, 5, and 7 for $k = 1, 2, 3$, respectively). To accommodate varying channel capacities, the full-sized weight tensors of the depth-wise convolutions are adaptively sub-sampled along the channel dimension up to the index C_k^{hard} for each sample during the functional forward invocation.

Since channel allocation involves discrete rounding, we adopt the Straight-Through Estimator (STE) to enable gradient propagation:

$$C_k = \mathrm{sg}[C_k^{hard} - \hat{C}_k] + \hat{C}_k, \tag{3.17}$$

where $\mathrm{sg}[\cdot]$ denotes the stop-gradient operator, which affects only gradient propagation and does not modify the forward routing behavior. Furthermore, to provide an auxiliary optimization path for the channel allocation network across the discrete channel partitioning process without altering the exact forward inference graph, we introduce a multiplicative scaling mechanism directly after each dynamic convolution:

$$\widetilde{DW}_k(F_{dec}^{(k)}) = DW_k(F_{dec}^{(k)}) \cdot \frac{C_k}{\mathrm{sg}[C_k^{hard}] + \epsilon}, \tag{3.18}$$

where $\widetilde{DW}_k(F_{dec}^{(k)})$ denotes the gradient-compensated convolution feature for the k-th scale branch, and ϵ is a small constant for numerical stability. In the forward pass, this scaling factor evaluates identically to 1.0, strictly preserving the physical convolution outputs. Although this surrogate gradient does not correspond to the true derivative of the hard routing process, it provides a stable optimization signal for the channel allocation network during training.

The outputs of all branches are concatenated and fused using a point-wise convolution to obtain the final value feature V:

$$V = \mathrm{Conv}_{pw}\left(\mathrm{Concat}_{k=1}^{3}(\widetilde{DW}_k(F_{dec}^{(k)}))\right), \tag{3.19}$$

where Conv_{pw} denotes the point-wise convolution used for cross-scale feature fusion.

Finally, the outputs are integrated via a Hadamard product and a residual connection to produce the final output feature map F_{MSDAB}:

$$F_{MSDAB} = F_{in} + \mathrm{Proj}_{out}(\mathrm{SiLU}(G) \odot \mathrm{SiLU}(V)), \tag{3.20}$$

where $G = \mathrm{Conv}_{1\times1}(F_{dec})$ and Proj_{out} denote the output projection layer.

3.4. Adaptive Feature Reuse Mechanism (AFRM)

To maximize the exploitation of hierarchical representations and mitigate computational redundancy, we implement the AFRM within the ARB. Unlike conventional static skip connections, the AFRM

dynamically recalibrates the contribution of reused features according to their semantic relevance to underwater restoration.

The core of this mechanism is the adaptive weighting unit, represented by the blue blocks labeled with $\boldsymbol{\beta}_i$ in Figure 3. For each skip connection path $i \in \{1, \ldots, 5\}$, a source feature map $F_{src,i}$ is extracted from a preceding stage. The corresponding weighting vector $\boldsymbol{\beta}_i$ and the recalibrated reused characteristic $F_{reused,i}$ are formulated as:

$$\boldsymbol{\beta}_i = \mathrm{Sig}\left(\mathrm{FC}\left(\mathrm{GAP}(F_{src,i})\right)\right), \tag{3.21}$$

$$F_{reused,i} = F_{src,i} \odot \boldsymbol{\beta}_i, \tag{3.22}$$

where GAP$(\cdot)$ denotes Global Average Pooling and FC$(\cdot)$ represents a fully connected layer.

Based on the spatial and semantic span of the skip connections shown in Figure 3, the AFRM categorizes feature reuse into short-term reuse (STR), medium-term reuse (MTR), and long-term reuse (LTR).

The STR path utilizes $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ to preserve local high-frequency details. Specifically, the output feature of the first MSDAB is enhanced by shallow reused features:

$$F^{(1)}_{stage} = F^{(1)}_{MSDAB} \oplus (\boldsymbol{\beta}_1 \odot F_{base}), \tag{3.23}$$

where $F^{(1)}_{stage}$ represents the integrated feature of the first stage, $\oplus$ denotes element-wise addition, and F_{base} serves as the source feature $F_{src,1}$.

The MTR path, modulated by $\boldsymbol{\beta}_4$, bridges intermediate stages to alleviate representation fading and enhance multi-scale structural consistency. The intermediate feature

$$F^{(1)}_{conv} = \mathrm{Conv}_{3\times3}(F^{(1)}_{stage}) \tag{3.24}$$

is directly propagated to the third aggregation stage:

$$F^{(3)}_{stage} = F^{(3)}_{MSDAB} + (\boldsymbol{\beta}_4 \odot F^{(1)}_{conv}). \tag{3.25}$$

Controlled by $\boldsymbol{\beta}_4$, the MTR path bridges intermediate stages to address medium-scale contrast degradation. It transmits the feature $F^{(1)}_{conv}$ (the output of the first stage's 3×3 convolution, serving as $F_{src,4}$) directly to the aggregation point of the third stage:

$$F^{(3)}_{stage} = F^{(3)}_{MSDAB} \oplus (\boldsymbol{\beta}_4 \odot F^{(1)}_{conv}). \tag{3.26}$$

The LTR paths, modulated by $\boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_5$, aggregate global information at the end of the branch via channel-wise concatenation:

$$F_{ARB} = \mathrm{Conv}_{1\times1}([\boldsymbol{\beta}_3 \odot F_{base}, \boldsymbol{\beta}_5 \odot F^{(2)}_{DB}, F^{(3)}_{stage}]), \tag{3.27}$$

where $F^{(2)}_{DB}$ (serving as $F_{src,5}$) is the feature map generated by the second Dense Block.

By explicitly defining the source features $F_{src,i}$ and their respective modulation vectors $\boldsymbol{\beta}_i$, the AFRM ensures a continuous and adaptive information flow. This enables the network to selectively emphasize relevant feature components, effectively suppressing redundant noise while preserving structural integrity in underwater scenes.

3.5. Gated Wavelet Transform Branch (GWTB)

To preserve high-frequency textures while suppressing underwater noise, the GWTB leverages the multi-resolution analysis capability of the Discrete Wavelet Transform (DWT). As illustrated in Figure 4, the GWTB decomposes the input feature map F_{in} into complementary spatial-frequency representations through a dual-path architecture.

The processing pipeline is formulated as follows. First, the input feature map is decomposed into four frequency sub-bands via DWT:

$$\{S_{LL}, S_{LH}, S_{HL}, S_{HH}\} = \text{DWT}(F_{in}), \tag{3.28}$$

where S_{LL} represents the low-frequency approximation component, while $\{S_{LH}, S_{HL}, S_{HH}\}$ correspond to horizontal, vertical, and diagonal high-frequency details, respectively.

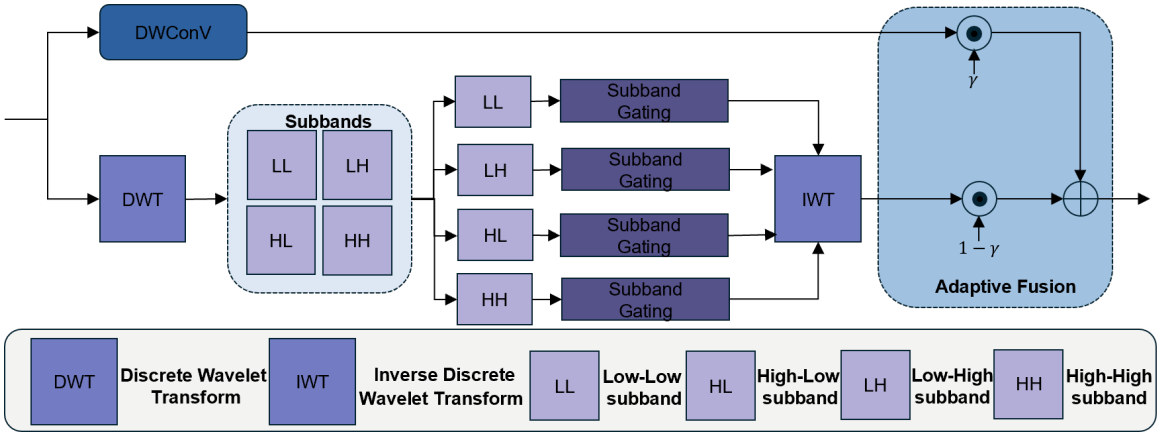

Figure 4. Gated Wavelet Transform Branch.

Each sub-band S_b ($b \in \{LL, LH, HL, HH\}$) is processed by a Subband Gating module, which acts as a dynamic spectral filter to recalibrate the importance of different frequency components. The gated sub-band is formulated as:

$$S'_b = \text{Sig}\left(\mathbf{\Theta}_2 \cdot \text{ReLU}\left(\mathbf{\Theta}_1 \cdot \text{GAP}(S_b)\right)\right) \odot \text{Conv}_{3\times3}(S_b), \tag{3.29}$$

where GAP$(\cdot)$ denotes Global Average Pooling, $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$ are the learnable weights of the MLP, and $\odot$ represents element-wise multiplication. The gating mechanism learns adaptive spectral importance to suppress noise-dominated frequency responses while preserving informative structural details.

Furthermore, the refined sub-bands are reconstructed back into the spatial domain through the Inverse Wavelet Transform (IWT):

$$F_{wavelet} = \text{IWT}(S'_{LL}, S'_{LH}, S'_{HL}, S'_{HH}). \tag{3.30}$$

Furthermore, a parallel spatial branch employs a depth-wise convolution to preserve spatial consistency:

$$F_{dw} = \text{DWConv}(F_{in}). \tag{3.31}$$

Finally, an adaptive fusion strategy integrates the spatial and spectral representations:

$$F_{GWTB} = \boldsymbol{\gamma} \odot F_{dw} + (1 - \boldsymbol{\gamma}) \odot F_{wavelet}, \tag{3.32}$$

where γ is a learnable balancing parameter that dynamically controls the contributions of spatial and spectral features.

3.6. Adaptive Fusion Block (AFB)

Conventional multi-branch image restoration architectures commonly rely on heuristic integration strategies, such as direct averaging or static weighted summation. However, such rigid fusion mechanisms fail to capture the statistical priorities of different degradation dimensions on average across the domain. To address this limitation, we propose the AFB, which adaptively optimizes the structural contribution of each branch through a task-level joint learning process.

In contrast to AFRM, which selectively refines degraded shallow features to preserve fine-grained local details, the AFB implements a dataset-adaptive fusion baseline to integrate heterogeneous representations from multiple parallel branches. Moreover, the AFB integrates complementary representations emphasizing adaptive cross-layer feature reuse, multi-scale contextual aggregation, and frequency-aware feature enhancement. By combining these complementary restoration representations, the proposed framework alleviates the limitations of single-path feature extraction and enhances the reconstruction of degraded underwater structures and textures, thereby achieving high-fidelity underwater image denoising.

The final fused representation is formulated as:

$$F_{fuse} = w_1 F_{DCB} + w_2 F_{ARB} + w_3 F_{GWTB}, \tag{3.33}$$

where w_i denotes the learned task-adaptive fusion weight assigned to the corresponding branch. While the preceding MSDAB module captures highly dynamic, instance-specific input adaptivity, the AFB is intentionally designed to model global prior behaviors to stabilize the multi-scale aggregation. To ensure numerical stability and a normalized weight distribution during multi-branch integration, a Softmax-based normalization strategy is adopted:

$$w_i = \frac{\exp(\alpha_i)}{\sum_{j=1}^{3} \exp(\alpha_j)}, \tag{3.34}$$

where α_i are dataset-adaptive parameters that are automatically optimized via backpropagation during the training phase to model the optimal baseline contribution of each pathway, satisfying:

$$\sum_{i=1}^{3} w_i = 1. \tag{3.35}$$

During inference, these weights remain static to serve as a robust structural foundation. By optimally balancing contextual representations from the DCB, structural features from the ARB, and spectral priors from the GWTB, the proposed AFB achieves robust and task-adaptive underwater image restoration across degradation conditions while preventing overfitting to individual anomalous frames.

3.7. Loss function

The network training employs a composite loss function, combining the advantages of L_{MAE} (Mean Absolute Error) and L_{MSE} (Mean Squared Error) losses [41]. The total loss function is defined as:

$$L_{\text{total}} = \lambda_1 L_{\text{MAE}} + \lambda_2 L_{\text{MSE}}, \tag{3.36}$$

where the L_{MAE} loss term is utilized to preserve the overall structural information of the image, and the L_{MSE} loss term optimizes pixel-level reconstruction accuracy. Following the established optimization protocol in the recent underwater image denoising literature [42], the hyperparameters are set to $\lambda_1 = 100$ and $\lambda_2 = 0.1$. Adopting this domain-specific configuration ensures that the gradient updates are appropriately guided by structural constraints via the MAE term, while the pixel-level MSE term acts as a regularizer to stabilize convergence without inducing over-smoothing artifacts in turbid underwater scenes.

4. Experiments and results

In this section, we provide a systematic empirical evaluation of the proposed MWGANet architecture designed for underwater image denoising tasks under a rigorously controlled synthetic noise framework. Initially, the experimental configurations are formally established, delineating the essential baseline dataset properties, the AWGN-based degradation protocol, specific network implementation details, and the quantitative evaluation metrics. Subsequently, comprehensive ablation studies are conducted to isolate and verify the individual architectural contributions and effectiveness of our core structural components. Following this verification, the performance of MWGANet is benchmarked against representative state-of-the-art denoising methods through both quantitative indicators and qualitative visual comparisons. Finally, an in-depth computational complexity investigation is presented, focusing on parameter overhead and real-time inference efficiency to precisely assess the computational cost of the proposed framework.

4.1. Experimental settings

In this subsection, we outline the foundational configuration utilized throughout our empirical evaluation. The structural characteristics of the utilized underwater image benchmarks are first elaborated, along with the concrete AWGN synthesis protocol deployed to generate degraded-pristine image pairs for network training and validation. Subsequently, the precise implementation details, encompassing network hyperparameter training settings and optimization strategies, are explicitly provided. Finally, quantitative evaluation metrics are defined to establish consistent and objective mathematical validation criteria for rigorous performance comparison against existing image denoising methods.

4.1.1. Datasets and augmentation

To rigorously evaluate the localized denoising capability and structural preservation performance of the proposed method under diverse and visually complex aquatic conditions, publicly available underwater image benchmarks are introduced, namely UFO-120 [43], EUVP [44], and UIEB [5]. Characterized by non-uniform illumination topologies, low-contrast gradients, and intricate color distributions, these datasets encompass rich underwater scenes with highly challenging visual attributes, thereby presenting significantly more demanding content distributions compared to standard terrestrial natural-image benchmarks.

It should be emphasized that our primary objective of this study is confined to image denoising rather than standard underwater image restoration or color correction. Consequently, for paired benchmarks such as UFO-120, EUVP, and UIEB, only the high-quality reference maps are utilized as the clean ground truth targets, whereas the original degraded-reference image pairs are deliberately excluded from the supervised restoration learning pipeline.

For the underwater datasets, degraded input samples are constructed by injecting synthetic AWGN with varying standard deviations ($\sigma = 10, 30, 50$) into the corresponding reference images. Under this formulation, the synthetically added Gaussian noise serves as the unique degradation vector introduced during the network training and empirical evaluation phases. Such a protocol enables a highly controlled assessment of the network's specialized noise-removal capabilities while implicitly preserving the challenging visual characteristics inherent to underwater imagery. Detailed structural statistics of the adopted datasets are systematically summarized in Table 1.

In addition to the underwater benchmarks mentioned above, several standard terrestrial databases, including DIV2K [45], Kodak24 [46], and CBSD68 [47], are also utilized to comprehensively assess the general image denoising proficiency of the proposed framework. Consisting of high-quality natural scenes, these datasets are widely acknowledged and adopted in standard image restoration literature to verify cross-domain robustness.

Table 1. Detailed statistics of the employed datasets.

Dataset	Type	Paired		Unpaired		Key features
		Train	Test	Train	Test	
UFO-120	Syn.	1500	120	–	–	CycleGAN + Blur
EUVP	Mixed	11,435	515	6335	613	Diverse scenes
UIEB	Real	800	90	–	60	Real benchmark

4.1.2. Implementation details

We use the UFO-120 dataset for underwater experiments. A total of 14 high-quality reference images were selected exclusively from the 1500-image training split of UFO-120, ensuring strict data separation from the 120 held-out test images used for evaluation. To avoid subjective content bias while maintaining diversity, these images are randomly sampled under the constraint that they collectively cover four critical visual dimensions: diverse color casts, varying turbidity levels, varying lighting conditions, and distinct structural contents. A detailed ablation study demonstrating the robustness of this constrained sampling strategy is provided in Section 4.3.4. These images are cropped into 41×41 patches to align with the network's receptive field and ensure computational efficiency. Synthetic noise was generated by adding AWGN with noise levels $\sigma \in [0, 50]$. While real underwater degradation involves complex coupled mechanisms (scattering, absorption, sensor noise), the use of AWGN provides a controlled and reproducible benchmark consistent with established image denoising protocols [7]. The model's robustness to this distribution shift is further validated by its competitive performance on the real-world UIEB test sets, which contain authentic underwater degradation. To enhance model generalization, we apply geometric augmentations—including random horizontal/vertical flips and rotations ($90°, 180°, 270°$)—yielding a total of 14,784 synthetic underwater samples. For comparative terrestrial denoising, we follow an identical protocol using 12 images from the DIV2K dataset, generating 126,720 training patches.

All experiments are implemented in PyTorch on an Ubuntu 20.04 server equipped with an NVIDIA RTX 2080 Ti GPU (11 GB). The proposed MWGANet is trained for 80 epochs using the Adam optimizer with an initial learning rate of 1×10^{-3} and a cosine annealing scheduler. To optimize memory usage and stability, we employ Automatic Mixed Precision (AMP) with a gradient accumulation step of 4 (effective batch size of 32), taking approximately 18 hours for the underwater dataset and 24 hours for the terrestrial dataset. To ensure a rigorous and fair evaluation, all compared baseline methods are fully retrained from scratch using their official source codes under the identical data environment, sharing the exact same data split, 41×41 patch size, and noise generation protocol. Each model is optimized using its respective paper-recommended optimal hyperparameters (e.g., specific optimizers and learning rate schedules) to avoid sub-optimal performance. This unified setup guarantees that the resulting performance disparities purely reflect architectural capacities rather than optimization discrepancies.

4.1.3. Evaluation metrics

To comprehensively and objectively evaluate the image restoration performance of different denoising methods from multiple dimensions, four mainstream quantitative metrics are introduced in this study, namely Peak Signal-to-Noise Ratio (PSNR) [48], Structural Similarity (SSIM) [49], Normalized Mean Square Error (NMSE), and Learned Perceptual Image Patch Similarity (LPIPS) [50]. PSNR and NMSE are utilized to quantify the pixel-level fidelity and pixel-wise error accumulation, whereas SSIM and LPIPS are leveraged to assess the high-level structural conformity and deep feature-based human perceptual alignment, respectively. The mathematical boundaries and respective validation criteria of these evaluation metrics are systematically detailed in Table 2.

Table 2. Delineation of mathematical ranges and effectiveness criteria of the adopted evaluation metrics.

Metric	Range	Effectiveness criteria
PSNR	[20 dB, 40 dB]	A higher metric value denotes suppressed pixel-wise distortion and superior localized image denoising capabilities.
SSIM	$[0, 1]$	Values approaching unity reflect enhanced high-frequency structural preservation and superior objective visual quality.
NMSE	$[0, \infty)$	A smaller value approaching zero indicates prioritized error minimization and enhanced numerical fidelity reconstruction.
LPIPS	$[0, 1]$	Diminished metric scores imply high-level semantic feature consistency and superior human perceptual similarity.

4.2. Ablation studies

To systematically isolate the individual architectural contributions and empirical efficacy of each core component integrated within the proposed framework, comprehensive ablation studies are executed for the UFO-120 dataset at a baseline noise intensity of $\sigma = 30$. Evaluating structural variants exclusively for the UFO-120 benchmark is methodologically justified by the intrinsic formulations of our task paradigm. Since the target corruption vector across all four adopted benchmarks is identically distributed with synthetic AWGN, the underlying statistical distribution and mathematical nature of the noise remain strictly invariant. Consequently, the primary variable manifested across these databases resides solely in

the textural complexity and distribution profiles of the underlying clean imagery. Among the available evaluation benchmarks, UFO-120 exhibits the most severe degradation characteristics, typified by non-uniform illumination topologies and heavily attenuated contrast gradients. Treating this dataset as a rigorous, high-fidelity "stress test" ensures that a module capable of distinguishing stochastic noise from such severely obscured structural textures will inherently possess robust generalization capabilities to milder visual domains.

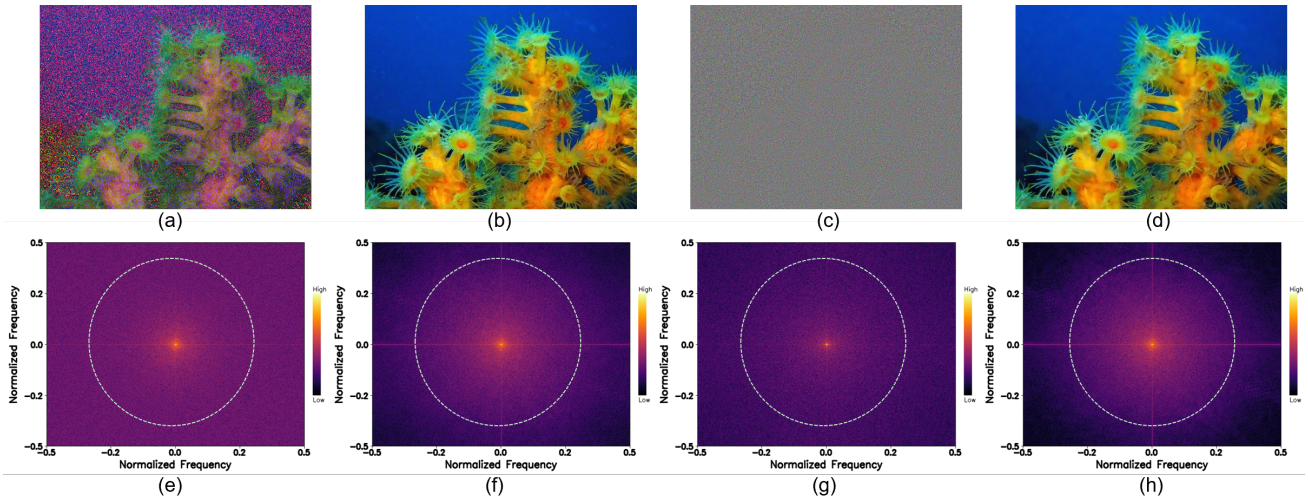

Figure 5. Joint spatial–frequency domain analysis of the proposed framework. Top row: Spatial-domain representations of the (a) degraded input, (b) ground truth (GT), (c) learned residual map, and (d) restored image. Bottom row: Corresponding normalized 2D FFT magnitude spectra of (a)–(d). The dashed circle indicates the low-frequency region centered at the DC component. Compared with the degraded input, the restored result suppresses excessive high-frequency noise while preserving frequency characteristics consistent with the GT image.

Table 3. Average test results of different network configurations on the underwater dataset UFO-120 (bold indicates the best result).

Network architecture	ARB	DCB	GWTB	PSNR (dB)	SSIM	NMSE
(a)	✓	×	×	31.95	0.9849	0.0032
(b)	✓	✓	×	32.10	0.9857	0.0031
(c)	✓	×	✓	32.14	0.9858	0.0031
(d)	✓	✓	✓	**32.23**	**0.9861**	**0.0030**
(e)	×	×	×	20.50	0.8126	0.0417

As evaluated in Table 3, the primitive baseline configuration (e) yields a PSNR of 20.50 dB and an NMSE of 0.0417, serving as the empirical baseline for comparison. After introducing the ARB backbone in configuration (a), the PSNR increases substantially to 31.95 dB, validating the effectiveness of the ARB backbone in improving restoration performance. Based on this backbone, the further integration of DCB (configuration b) and GWTB (configuration c) improves the PSNR to 32.10 dB and 32.14 dB, respectively, suggesting that multi-scale spatial context modeling and frequency-domain

priors provide additional restoration cues. Finally, the complete MWGANet configuration (d) achieves the best overall performance, with 32.23 dB PSNR, 0.9861 SSIM, and 0.0030 NMSE, indicating the effectiveness of the joint architectural design.

Although the quantitative results in Table 3 show the progressive contribution of different architectural components, global metrics such as PSNR, SSIM, and NMSE mainly reflect spatially averaged restoration accuracy and may not fully characterize the frequency-domain behavior of the entire network. Therefore, a 2D Fast Fourier Transform (FFT) analysis is further introduced to examine whether the restored image exhibits a spectral distribution consistent with the ground-truth image, and to provide additional evidence for the noise-residual separation behavior of MWGANet.

To avoid ambiguity, it should be clarified that the decoupling discussed in this work refers to the separation of additive Gaussian-noise-related residual components from the underlying underwater image content, rather than a complete physical decomposition of all real-world underwater degradation factors (e.g., scattering, color attenuation, blur, and non-uniform illumination). This content-noise decoupling behavior is explicitly substantiated by the frequency-domain observations. As illustrated by the normalized FFT magnitude spectrum in Figure 5(e), the noisy input exhibits a broader spectral distribution with noticeable energy diffusion toward peripheral high-frequency regions, which is consistent with the spectral perturbations introduced by additive Gaussian white noise. In contrast, the GT spectrum in Figure 5(f) presents a more compact low-frequency-dominated distribution, where the spectral energy gradually decays from the center to higher-frequency regions.

As shown in Figure 5(g), the learned residual is mainly distributed over dispersed high-frequency regions, with relatively weak energy around the spectral center. This distinct high-frequency concentration directly validates that the residual branch successfully decouples and captures the Gaussian-noise-related perturbations rather than corrupting the dominant low-frequency structural content of the clean target. After restoration, the spectrum in Figure 5(h) becomes more consistent with that of the GT. The excessive high-frequency diffusion observed in the noisy input is suppressed, while the central low-frequency components and major spectral patterns are better preserved. These observations support that MWGANet performs structure-aware Gaussian noise suppression and residual separation, rather than indiscriminate spectral smoothing.

4.2.1. Ablation study on the DCB

To optimize the topology of the DCB, we evaluate the effect of using different numbers of cascaded Dilated Residual Blocks (DRBs). As shown in Figure 6, the single-DRB configuration, denoted by the green curve, shows unstable convergence and limited denoising performance, mainly due to its restricted receptive field.

Increasing the number of DRBs consistently improves performance. The four-DRB configuration, represented by the red curve, achieves the best evaluation results with a smooth convergence trajectory. This suggests that cascading DRBs with heterogeneous dilation rates effectively enlarges the receptive field and captures multi-scale contextual information, which is beneficial for suppressing underwater noise while preserving structural details.

Although deeper configurations may further increase model capacity, the trends in Figure 6 indicate diminishing returns. The performance gain from three to four DRBs becomes much smaller than that from one to three DRBs, and the NMSE reduction in Figure 6(a) tends to plateau. Therefore, adding more DRBs would likely bring limited improvement while increasing computational cost and the risk of

overfitting. Based on this trade-off, four cascaded DRBs are adopted in the proposed DCB.

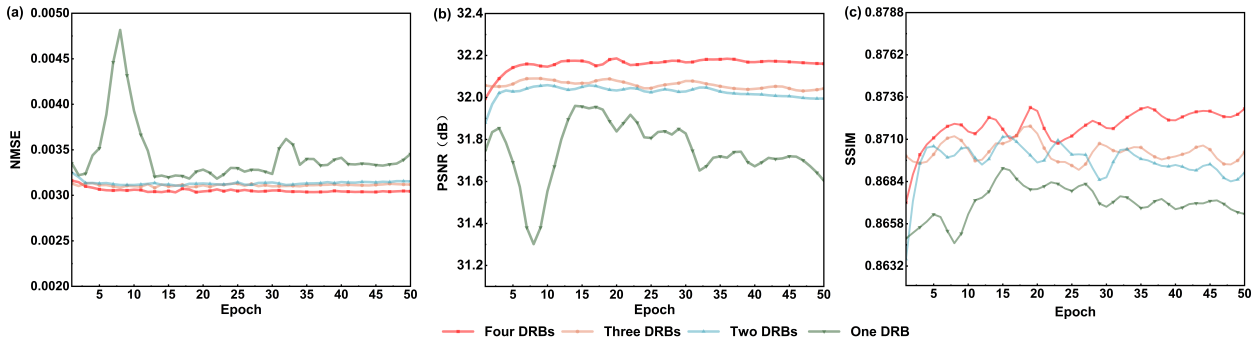

Figure 6. Impact of the number of cascaded Dilated Residual Blocks (DRBs) within the DCB on network convergence and evaluation metrics: (a) NMSE, (b) PSNR (dB), and (c) SSIM.

4.2.2. Ablation study on the GWTB

To quantitatively evaluate image restoration performance and isolate the specialized impact of the gating pipeline, an independent ablation study focusing on the GWTB is performed. As summarized systematically in Table 4, the architectural variant formulated without the gating strategy (GWTB without Gate) yields a constrained performance baseline. In contrast, the fully integrated gating mechanism employed in the complete GWTB configuration yields a slight but consistent improvement, achieving a PSNR of 32.23 dB, an SSIM of 0.9861, and a reduced NMSE of 0.0030.

Table 4. Quantitative ablation study of the gating mechanism in GWTB.

Network architecture	PSNR (dB)	SSIM	NMSE
GWTB w/o Gate	32.11	0.9857	0.0031
GWTB	**32.23**	**0.9861**	**0.0030**

Although the numerical improvement reported in Table 4 is relatively modest, the empirical error distributions provide a more intuitive visualization of the functional decoupling capacity introduced by the GWTB. As illustrated in the localized error maps in Figure 7(b), the restoration result obtained with the complete GWTB configuration demonstrates exceptional decoupling efficacy. The background stochastic noise is universally suppressed, and structural residual errors are highly minimized, leaving only a faint, clean blue-green profile across the scene. This visual evidence firmly confirms that the gating pipeline successfully identifies and preserves critical structural boundaries while cleanly stripping away ambient degradations.

In sharp contrast, as demonstrated in Figure 7(c), completely deactivating the gating workflow triggers a severe performance collapse. The residual error map exhibits conspicuous, dense, and highly intense structural residual responses (strongly approaching the 1.0 threshold in the colorbar) tightly clustered around the intricate coral silhouettes and edge boundaries. Concurrently, the flat background regions are contaminated with diffuse, unstructured errors. This reveals that without the frequency-selective gating constraint, the network fails to decouple coupled stochastic noise-related spikes from structurally meaningful high-frequency components, causing the structural anatomy to bleed heavily

into the error space. This comparative analysis firmly verifies that the GWTB acts as a spatial-frequency decoupling pivot, preventing the blurring of sharp textures and ensuring that compound underwater noise is isolated.

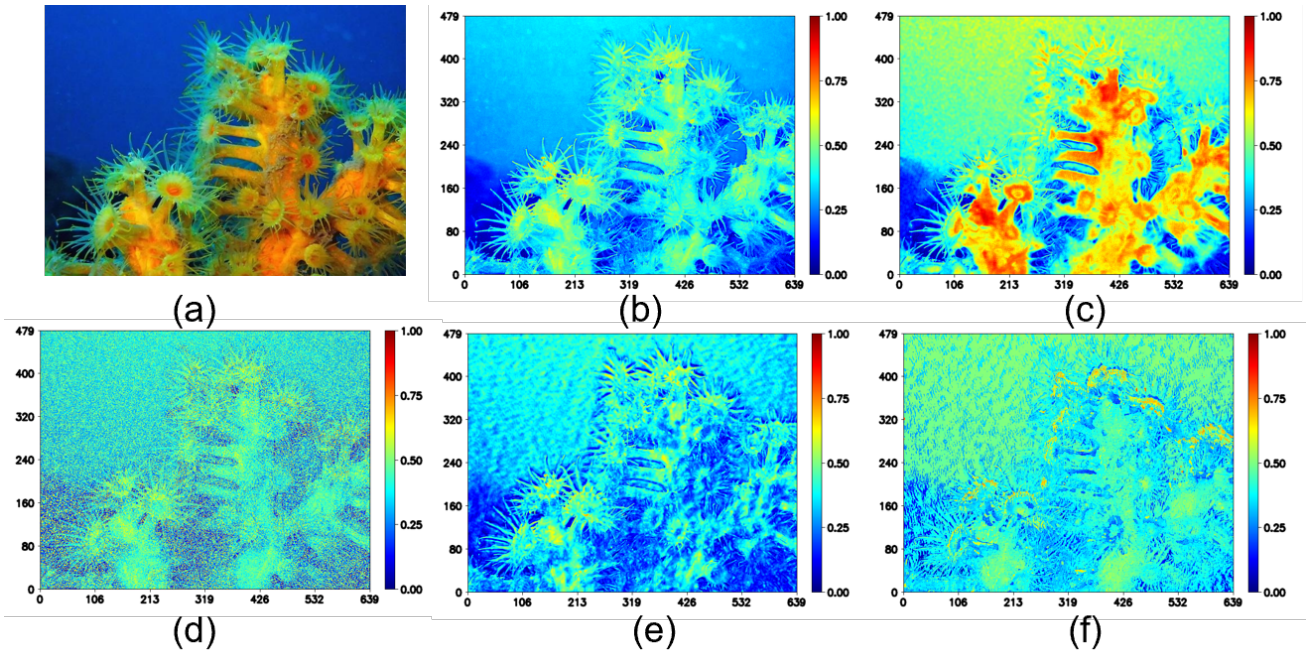

Figure 7. Visual comparison of the ground-truth image and residual error heatmaps from branch-level ablation experiments on the underwater image dataset. (a) Ground-truth image; (b)–(f) normalized residual error maps under different ablation settings: (b) GWTB, (c) GWTB w/o Gate, (d)ARB w/o AFRM, (e)ARB w/ MOGA, and (f) AFRM+MSDAB. Blue and red indicate lower and higher residual errors relative to the ground truth, respectively.

4.2.3. Ablation study on the ARB

To further verify the contribution of different components in the ARB, a branch-level ablation study is conducted. As shown in Table 5, ARB w/o AFRM obtains a PSNR of 32.07 dB, an SSIM of 0.9856, and an NMSE of 0.0031. Replacing the branch with MOGA slightly degrades the performance to 31.96 dB in PSNR and 0.9854 in SSIM. In contrast, the complete ARB with MSDAB and AFRM achieves the best results, reaching 32.23 dB PSNR, 0.9861 SSIM, and 0.0030 NMSE. Although the numerical improvements are moderate, the consistent gains indicate that the combination of MSDAB and AFRM is more suitable for the proposed restoration framework.

Table 5. Quantitative ablation study of key modules in ARB.

Network architecture	PSNR (dB)	SSIM	NMSE
ARB w/o AFRM	32.07	0.9856	0.0031
ARB w/ MOGA	31.96	0.9854	0.0031
ARB w/ MSDAB+AFRM	**32.23**	**0.9861**	**0.0030**

To further explain these differences, the residual error heatmaps in Figure 7 are analyzed, where

warmer colors denote larger reconstruction errors and cooler colors indicate smaller residuals. As shown in Figure 7(d), removing AFRM leads to dense scattered errors in smooth background regions and object boundaries, explaining its lower quantitative performance. In Figure 7(e), MOGA reduces part of the background residuals, but continuous errors remain around coral edges, indicating insufficient fine-structure recovery.

By comparison, Figure 7(f) shows the lowest residual intensity and a more homogeneous error distribution. Most regions are dominated by cooler colors, with only weak errors near complex boundaries. This suggests that the collaboration between MSDAB and AFRM effectively reduces scattered residuals and yields more stable reconstruction results, which is consistent with the improved quantitative performance. Overall, the quantitative metrics and heatmap visualization confirm the effectiveness of the complete ARB design.

4.2.4. Ablation study on the AFB

Quantitative evaluations exploring alternative cross-branch feature fusion paradigms inside the AFB are detailed in Table 6. Rigid algebraic operators, such as element-wise summation (Sum) and static concatenation (Concat), result in constrained denoising performance. Such static operations fail to accommodate the highly non-linear, spatially non-uniform scattering and absorption processes inherent to underwater channels.

In contrast, the proposed Full Fusion strategy achieves a notable optimization breakthrough, yielding the highest fidelity metrics (32.23 dB PSNR and 0.9861 SSIM) alongside minimized residual variance. Specifically, our adaptive framework surpasses the rigid Sum and Concat baselines by 0.52 dB and 3.33 dB in PSNR, respectively. These empirical results verify that static fusion alternatives are insufficient for cross-layer feature alignment. Our dynamic design, conversely, successfully harnesses multi-branch synergies to suppress persistent background artifacts while reinforcing structural clarity.

Table 6. Quantitative comparison of distinct feature fusion paradigms.

Fusion strategy	PSNR (dB)	SSIM	NMSE
Equal	31.45	0.9847	0.0035
Concat	28.90	0.9812	0.0060
Sum	31.71	0.9849	0.0033
Full (Proposed)	**32.23**	**0.9861**	**0.0030**

4.2.5. Robustness of source image selection

In our underwater experiments, the training set is constructed by sampling 14 full-resolution reference images and cropping them into 14,784 local patches. To investigate whether our selection might introduce content bias, we conduct an ablation study to evaluate the model's stability across different data sampling iterations. We compare the performance of our original randomly sampled baseline (Subset 3) against two independent random selections of 14 images (Subsets 1 and 2). Crucially, all three trials are independently sampled under the identical constraint of covering the four key visual dimensions (turbidity, color cast, lighting, and content) from the UFO-120 training split.

As reported in Table 7, the performance fluctuations across the three distinct sets are negligible. The model achieves an average PSNR of 32.20 ± 0.03 dB, an SSIM of 0.9860 ± 0.0001, and a stable NMSE of 0.003 ± 0.000. Such exceptionally low variance demonstrates that our patch-generation

strategy effectively breaks image-level content dependencies. This proves that extracting patches from a small, dimension-constrained random subset of 14 images is sufficient to capture unbiased and highly generalizable local structural information. This confirms that the model's performance is stable and not reliant on any specific combination of images, provided the fundamental physical visual dimensions are represented.

Table 7. Robustness evaluation under different selections of 14 source images.

Training trial	PSNR (dB)	SSIM	NMSE
Subset 1	32.18	0.9859	0.003
Subset 2	32.20	0.9860	0.003
Subset 3 (Baseline)	32.23	0.9861	0.003
Mean ± Std	**32.20 ± 0.03**	**0.9860 ± 0.0001**	**0.003 ± 0.000**

4.3. Comparison with state-of-the-art methods

To comprehensively evaluate the performance of the proposed denoising model in underwater image processing, we conduct systematic denoising experiments on the UFO-120, UIEB,and EUVP datasets. Furthermore, comprehensive comparisons are performed against various state-of-the-art methods, including UDRN [42], ADNet [51], ECNDNet [52], DnCNN [7], DudeNet [53], DRDNet [54], and MWDCNN [55], under different noise levels. We include DRDNet as a higher-capacity baseline to test whether the proposed lightweight design remains competitive against deeper models. We also acknowledge the emergence of Transformer-based methods in 2024–2026. While these approaches represent the latest SOTA, they often entail heavy computational burdens unsuitable for real-time AUV tasks. Therefore, our comparative analysis focuses on established CNN architectures to benchmark the proposed method primarily against models with comparable efficiency profiles.

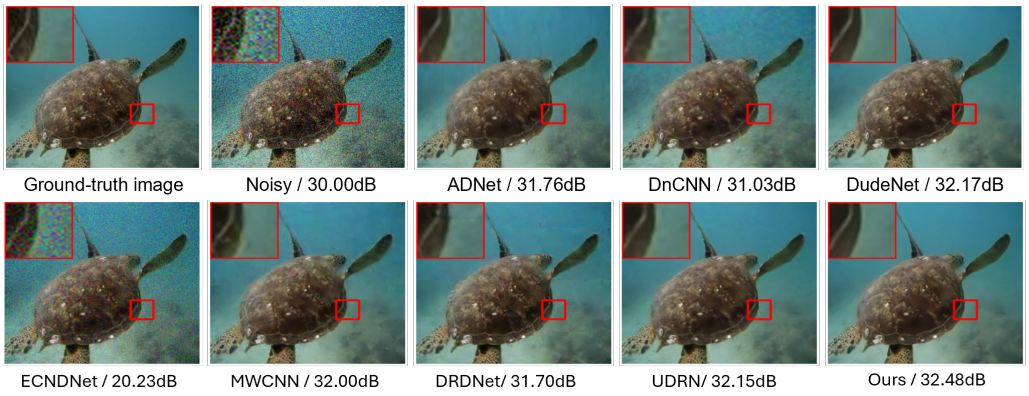

Figure 8. Visual comparison of the UFO-120 dataset obtained by different methods at a noise level of 30.

Table 8 summarizes the comparative results on the UIEB dataset under three noise intensities. Our proposed MWGANet consistently achieves the highest PSNR and SSIM across all evaluated scenarios. At the $\sigma = 10$ level, our method reaches a PSNR of 36.64 dB and a near-perfect SSIM of 0.9947. While ADNet and DRDNet show competitive NMSE and LPIPS performances under low-to-medium

noise, their advantage diminishes as degradation intensifies. A key observation is the robustness of our architecture in severe noise conditions. When the noise level escalates to $\sigma = 50$, most conventional denoisers suffer significant fidelity loss; for example, the PSNR of ECNDNet drops sharply to 16.36 dB. In contrast, MWGANet maintains a robust PSNR of 29.27 dB and an SSIM of 0.9721, representing a 0.47 dB and 0.52 dB improvement over the second-best performers (MWDCNN and ADNet, respectively). More importantly, compared with ECNDNet, our method improves the PSNR by 12.91 dB, which verifies its excellent resilience. Furthermore, at this extreme noise level, our approach yields the best LPIPS (0.2615), suggesting that the integration of frequency-domain priors and multi-scale context is particularly effective for recovering perceptual details from heavily corrupted underwater images.

To further evaluate the generalization of MWGANet across heterogeneous underwater conditions, we conduct extensive experiments for the EUVP and UFO-120 datasets. These benchmarks represent a wide spectrum of aquatic optical degradations, ranging from blue-green ocean tints to turbid harbor waters. The comparative performance across PSNR, SSIM, and LPIPS metrics is visualized in Figure 8.

As illustrated by the bar charts, our method achieves a decisive margin over state-of-the-art (SOTA) algorithms, particularly in low-visibility regimes ($\sigma = 50$). For instance, for the EUVP dataset at $\sigma = 50$, ECNDNet suffers a catastrophic failure with a PSNR of 16.46 dB and SSIM of 0.7177. In contrast, MWGANet sustains a high-fidelity output of 31.06 dB and 0.9754, representing a remarkable 14.6 dB improvement in PSNR and a 0.2577 gain in SSIM over ECNDNet. This significant gap underscores the resilience of our wavelet-based architecture in decoupling complex underwater noise from latent scene structures.

The visual results in Figures 9 and 10 further elucidate these quantitative gains. Competing methods often struggle with color-cast-dependent noise, frequently introducing residual speckle artifacts or over-smoothing the fine-grained textures of marine organisms. In contrast, MWGANet effectively suppresses high-frequency noise while preserving sharp edges and structural coherence, which is reflected in its superior LPIPS scores that consistently lead the second-best performer by a noticeable margin.

The consistent lead in SSIM across four datasets, spanning synthetic and real-world underwater scenarios, validates the structural-preserving capability of our model. By incorporating global contextual information and frequency-domain priors, MWGANet provides a promising framework for underwater vision tasks, such as AUV navigation and marine biological surveys, and demonstrates enhanced adaptability to complex underwater illumination variations.

Table 8. Performance comparison of various methods for the UIEB dataset (bold indicates the best result).

Dataset	Evaluation	Noise	UDRN	ADNet	ECNDNet	DnCNN	DudeNet	DRDNet	MWDCNN	Ours
UIEB	PSNR	10	36.28	36.36	28.11	34.04	36.41	35.74	36.24	**36.64**
		30	30.39	30.95	20.45	29.66	30.39	30.76	31.11	**31.44**
		50	27.48	28.75	16.36	28.20	27.46	28.38	28.80	**29.27**
	SSIM	10	0.9240	0.9942	0.9558	0.9498	0.9247	0.9283	0.9403	**0.9947**
		30	0.8095	0.9807	0.8205	0.8637	0.8075	0.8252	0.8467	**0.9830**
		50	0.7271	0.9675	0.6802	0.8078	0.7186	0.7424	0.7690	**0.9721**
	NMSE	10	0.0063	**0.0014**	0.0061	0.0025	0.0060	0.0016	0.0065	**0.0014**
		30	0.0233	0.0049	0.0482	0.0065	0.0233	0.0050	0.0207	**0.0045**
		50	0.0446	0.0080	0.1235	0.0088	0.0450	0.0085	0.0346	**0.0073**
	LPIPS	10	0.0639	0.0664	0.2627	0.0849	0.0622	**0.0552**	0.0604	0.0595
		30	0.1955	0.2114	0.3976	0.2281	0.2129	**0.1760**	0.1827	0.1803
		50	0.3008	0.2968	0.6168	0.2968	0.3060	0.2638	0.2651	**0.2615**

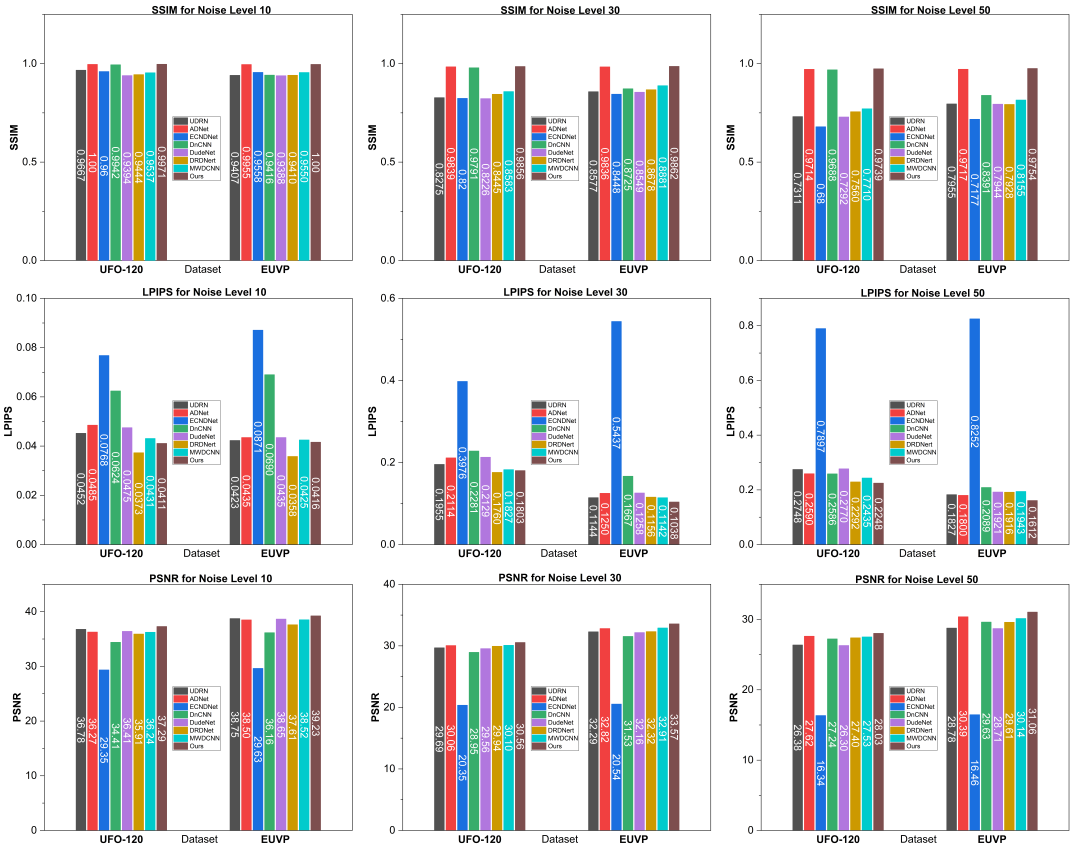

Figure 9. Visual comparison of the EUVP dataset obtained by different methods at a noise level of 30.

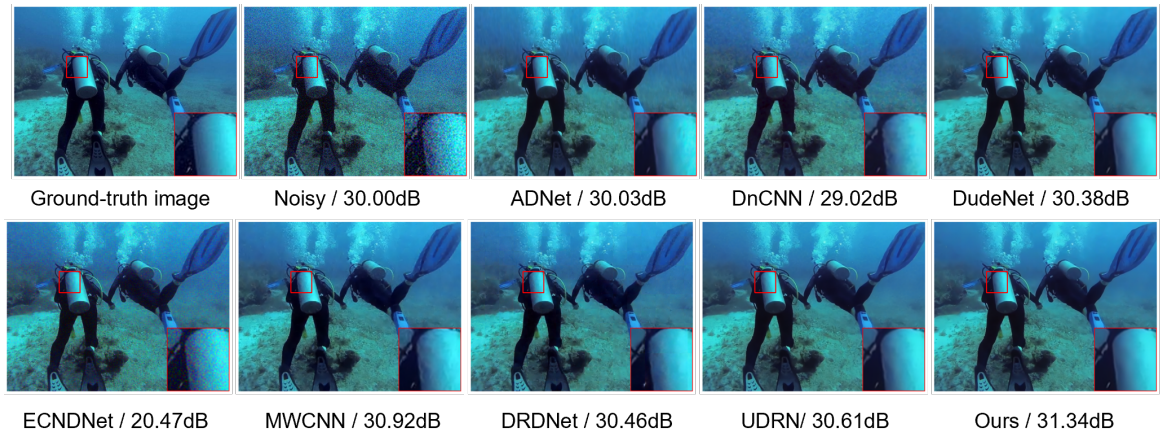

Figure 10. Quantitative comparison of different denoising methods on the UFO-120 and EUVP datasets under Gaussian noise levels of 10, 30, and 50. The first, second, and third rows correspond to SSIM, LPIPS, and PSNR, respectively.

4.4. Generalization performance comparison

To evaluate generalization, supplementary experiments are conducted for two standard natural image datasets: CBSD68 and Kodak24. These datasets provide a benchmark for cross-domain robustness, as their feature distributions differ significantly from underwater imagery. We test all models at three noise levels ($\sigma = 10, 30, 50$). To ensure a fair comparison, all CNN-based baselines (UDRN, ADNet, ECNDNet, DnCNN, DudeNet, DRDNet, and MWDCNN) are retrained from scratch using the identical experimental protocol, including training data, optimizer, and evaluation pipeline, as described in Section 4. Performance metrics are summarized in Tables 9 and 10.

As detailed in Table 9, MWGANet achieves top-tier performance across all noise intensities ($\sigma \in \{10, 30, 50\}$). At the lower noise floor ($\sigma = 10$), our method yields a PSNR of 35.82 dB and an SSIM of 0.9597, establishing a strong upper bound for high-frequency detail reconstruction. Notably, as the degradation intensifies to $\sigma = 50$, MWGANet sustains a robust PSNR of 27.60 dB and an SSIM of 0.7917, outperforming the second-best method (UDRN) by 0.03dB and 0.0015, respectively. A critical observation is the architectural stability of MWGANet when subjected to significant domain shifts. While ECNDNet suffers from a catastrophic performance collapse at higher noise levels, with PSNR dropping to 16.22 dB—, our method maintains a graceful and predictable degradation curve, achieving a 11.28 dB higher PSNR than ECNDNet at $\sigma = 50$. Furthermore, although MWDCNN and DRDNet employ complex feature-extraction schemes, they fail to match the structural fidelity of our approach. This superiority is visually corroborated in Figure 11; the zoomed-in patches ($\sigma = 30$) demonstrate that MWGANet preserves sharper boundaries and finer textures than SOTA baselines. Such results provide compelling evidence that while MWGANet is tailored for underwater optical complexities, its fundamental denoising mechanism effectively captures universal image manifolds, ensuring robust generalization to natural scene imagery.

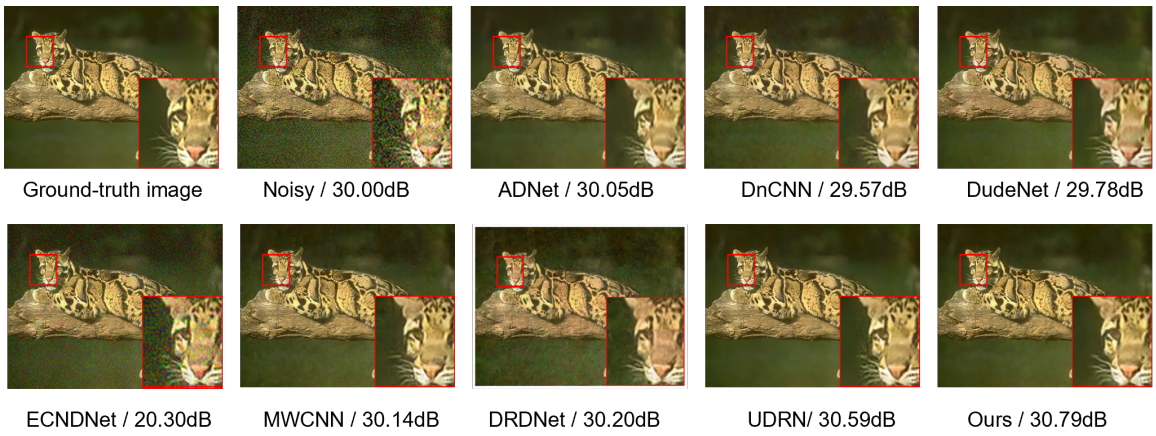

Figure 11. Visual comparison of the CBSD68 dataset obtained by different methods at a noise level of 30.

Table 10 presents the quantitative evaluation of the Kodak24 dataset, a canonical benchmark for high-fidelity image restoration. These results further corroborate the cross-domain adaptability of MWGANet across visual manifolds. At the low-noise regime ($\sigma = 10$), our method establishes a new performance ceiling with a PSNR of 36.50 dB and an SSIM of 0.9530, surpassing the second-best

performer (UDRN) by 0.13 dB in PSNR. While maintaining a lower NMSE (0.0012), our approach highlights its precision in recovering latent clean signals from minimal corruption.

As the noise floor rises, the resilience of our architecture becomes more pronounced. At $\sigma = 30$, MWGANet achieves a PSNR of 30.96 dB, yielding a 0.11 dB improvement over UDRN. More significantly, under severe noise ($\sigma = 50$), our method sustains a robust 28.64 dB, while ECNDNet suffers from a catastrophic performance collapse, dropping to 16.19 dB. In this extreme scenario, MWGANet achieves a 12.45 dB higher PSNR than ECNDNet, demonstrating exceptional stability. This consistency suggests that the integration of multi-scale wavelet priors and attentive residual learning effectively captures universal image regularities that are invariant to specific domain characteristics. The collective evidence from CBSD68 and Kodak24 confirms that although MWGANet is optimized for the intricate physics of underwater light propagation, it serves as a highly versatile denoising solution capable of maintaining structural integrity even when subjected to significant distributional shifts.

Table 9. Performance comparison with other methods for the CBSD68 dataset (bold indicates the best result).

Methods	CBSD68								
	$\sigma = 10$			$\sigma = 30$			$\sigma = 50$		
	PSNR↑	NMSE↓	SSIM↑	PSNR↑	NMSE↓	SSIM↑	PSNR↑	NMSE↓	SSIM↑
UDRN	35.81	0.0014	0.9566	29.95	0.0057	0.8607	27.57	0.0099	0.7905
ADNet	35.80	0.0014	0.9574	29.71	0.0059	0.8564	27.34	0.0105	0.7805
ECNDNet	29.16	0.0061	0.9573	20.23	0.0476	0.839	16.22	0.1194	0.7094
DnCNN	34.02	0.0022	0.9416	29.07	0.0068	0.8522	26.83	0.0117	0.7811
DudeNet	35.72	0.0014	0.9588	29.71	0.0060	0.8595	27.07	0.0112	0.7812
DRDNet	35.23	0.0016	0.9399	29.33	0.0062	0.8297	26.71	0.0113	0.7309
MWDCNN	35.49	0.0078	0.9484	29.61	0.0309	0.8458	26.98	0.0555	0.7594
Ours	**35.82**	**0.0014**	**0.9597**	**29.96**	**0.0057**	**0.8663**	**27.60**	**0.0099**	**0.7917**

Table 10. Performance comparison with other methods for the Kodak24 dataset (bold indicates the best result).

Methods	Kodak24								
	$\sigma = 10$			$\sigma = 30$			$\sigma = 50$		
	PSNR↑	NMSE↓	SSIM↑	PSNR↑	NMSE↓	SSIM↑	PSNR↑	NMSE↓	SSIM↑
UDRN	36.37	0.0012	0.9501	30.85	0.0046	0.8560	28.58	0.0078	0.7927
ADNet	36.13	0.0013	0.9478	30.42	0.0050	0.8448	28.12	0.0087	0.7762
ECNDNet	29.25	0.0060	0.9504	20.25	0.0472	0.8428	16.19	0.1197	0.7255
DnCNN	34.95	0.0017	0.9416	30.07	0.0053	0.8466	27.96	0.0088	0.7849
DudeNet	36.01	0.0014	0.9493	30.27	0.0052	0.8531	27.68	0.0094	0.7636
DRDNet	35.42	0.0015	0.9241	29.88	0.0055	0.8110	27.22	0.0101	0.7085
MWDCNN	36.06	0.0069	0.9387	30.55	0.0253	0.8402	28.06	0.0451	0.7618
Ours	**36.50**	**0.0012**	**0.9530**	**30.96**	**0.0045**	**0.8636**	**28.64**	**0.0078**	**0.7992**

4.5. Parameters and inference speed

As illustrated in Table 11, MWGANet comprises only 0.79M parameters, representing a 27% reduction compared to DudeNet (1.08 M) and being significantly more lightweight than DRDNet (199.17 M) and MWDCNN (74.75 M). Regarding inference efficiency in Table 12, MWGANet processes

a 512 × 512 image in 95.99 ms. While this latency exceeds that of ultra-lightweight models such as DnCNN (∼46 ms), it remains well within the 10 FPS requirement for real-time underwater applications. Remarkably, MWGANet achieves a 20× speedup over the heavyweight MWDCNN while maintaining superior denoising performance. Such computational efficiency suggests the potential suitability of MWGANet for deployment on power-constrained AUVs.

Table 11. Complexity comparison of different denoising networks.

Methods	Parameters
UDRN	819,239
ADNet	521,493
ECNDNet	520,330
DnCNN	558,336
DudeNet	1,079,591
DRDNet	199,168,875
MWDCNN	74,751,067
Ours	**790,264**

Table 12. Running time comparison of different denoising methods for noisy images of various sizes. All times are measured in milliseconds (ms).

Image size	Inference time (ms)							
	UDRN	ADNet	ECNDNet	DnCNN	DudeNet	DRDNet	MWDCNN	Ours
32 × 32	3.3788	**1.5963**	1.6570	1.6961	2.6201	21.9009	16.9253	5.6595
64 × 64	3.6219	**1.6059**	1.6652	2.1109	2.8506	22.2580	29.6199	5.6701
128 × 128	3.9283	2.6056	**2.5592**	2.5826	4.6319	22.1229	71.8877	6.3552
256 × 256	15.9818	9.2212	**9.0247**	9.1429	18.7659	46.3175	240.7773	15.7956
512 × 512	79.3022	45.7478	**44.9099**	45.9489	85.0960	195.4539	2002.5156	95.9951
1024 × 1024	315.3802	178.1981	**177.9611**	178.8045	335.8508	534.7421	7953.5781	385.3915

5. Conclusions

In this paper, we present a robust underwater image denoising framework named MWGANet, designed to address the deleterious effects of light absorption and scattering in aquatic environments. Recognizing that underwater noise primarily manifests as high-frequency abrupt signals that obscure latent scene geometry, our model employs a three-pathway parallel architecture to decouple interference across domains. By integrating densely connected blocks, dilated residual blocks, and wavelet gating structures, the network effectively captures multi-scale features while preserving structural fidelity. The introduction of the MSDAB further enables adaptive feature decomposition, enabling the model to selectively suppress stochastic noise. Moreover, our results across four datasets, including UIEB, EUVP, and terrestrial benchmarks, demonstrate that MWGANet consistently surpasses state-of-the-art methods in denoising effectiveness and computational efficiency. The model maintains exceptional resilience under extreme noise conditions ($\sigma = 50$), achieving a superior balance between restoration accuracy

and resource consumption. Such performance indicates the potential applicability of MWGANet to vision-based underwater tasks on resource-constrained AUV platforms, where it can provide enhanced inputs for subsequent detection and segmentation operations.

Despite promising performance, several avenues remain for further advancement. The formation of underwater image noise is inherently complex and diverse, where different noise types exhibit significantly varied distributions. Mapping this complex noise domain to a high-quality image domain is essentially a many-to-many task, which poses challenges for standard supervised training. The reliance on Euclidean distance (L1/L2 loss) may cause the training process to converge to an average level, potentially limiting the restoration of fine-grained details in non-stationary environments. In addition, the evaluation mainly relies on AWGN, which cannot fully capture the complicated degradation characteristics encountered in real underwater scenarios, such as wavelength-dependent light absorption, scattering, turbidity, and mixed noise distributions. Therefore, in future work, we will focus on incorporating more realistic underwater degradation models and physics-aware priors to better bridge the gap between synthetic benchmarks and practical underwater imaging conditions.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgment

This work is supported in part by the Basic Scientific Research Project of the Department of Education of Liaoning Province under grant 984250013, and in part by the Joint Program of the Liaoning Provincial Science and Technology Plan (Key RD Program Project), Project No. 2025JH2/101800002, and the China Scholarship Council (Grant No. 202506570050).

Conflict of interest

The authors declare there are no conflicts of interest.

References

1. D. Akkaynak, T. Treibitz, A revised underwater image formation model, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018), 6723–6732. https://doi.org/10.1109/CVPR.2018.00703
2. C. Li, W. Liu, G. Gong, X. Ding, X. Zhong, SU-YOLO: Spiking neural network for efficient underwater object detection, *Neurocomputing*, **644** (2025), 130310. https://doi.org/10.1016/j.neucom.2025.130310
3. J. S. Jaffe, Computer modeling and the design of optimal underwater imaging systems, *IEEE J. Ocean. Eng.*, **15** (1990), 101–111. https://doi.org/10.1109/48.50695
4. R. Schettini, S. Corchs, Underwater image processing: State of the art of restoration and image enhancement methods, *EURASIP J. Adv. Signal Process.*, **2010** (2010), 746052. https://doi.org/10.1155/2010/746052

5. C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, et al., An underwater image enhancement benchmark dataset and beyond, *IEEE Trans. Image Process.*, **29** (2020), 4376–4389. https://doi.org/10.1109/TIP.2019.2955241
6. S. Anwar, C. Li, Diving deeper into underwater image enhancement: A survey, *Signal Process.: Image Commun.*, **89** (2020), 115978. https://doi.org/10.1016/j.image.2020.115978
7. K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising, *IEEE Trans. Image Process.*, **26** (2017), 3142–3155. https://doi.org/10.1109/TIP.2017.2662206
8. Y. Blau, T. Michaeli, The perception-distortion tradeoff, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018), 6228–6237. https://doi.org/10.1109/CVPR.2018.00652
9. A. Foi, M. Trimeche, V. Katkovnik, K. Egiazarian, Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data, *IEEE Trans. Image Process.*, **17** (2008), 1737–1754. https://doi.org/10.1109/TIP.2008.2001399
10. S. Guo, Z. Yan, K. Zhang, W. Zuo, L. Zhang, Toward convolutional blind denoising of real photographs, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2019), 1712–1722. https://doi.org/10.1109/CVPR.2019.00182
11. C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, C. W. Lin, Deep learning on image denoising: An overview, *Neural Networks*, **131** (2020), 251–275. https://doi.org/10.1016/j.neunet.2020.07.025
12. A. Buades, B. Coll, J. M. Morel, A non-local algorithm for image denoising, in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, (2005), 60–65. https://doi.org/10.1109/CVPR.2005.38
13. K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3-D transform-domain collaborative filtering, *IEEE Trans. Image Process.*, **16** (2007), 2080–2095. https://doi.org/10.1109/TIP.2007.901238
14. S. Gu, L. Zhang, W. Zuo, X. Feng, Weighted nuclear norm minimization with application to image denoising, in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, (2014), 2862–2869. https://doi.org/10.1109/CVPR.2014.366
15. J. Mairal, F. Bach, J. Ponce, G. Sapiro, A. Zisserman, Non-local sparse models for image restoration, in *2009 IEEE 12th International Conference on Computer Vision*, (2009), 2272–2279. https://doi.org/10.1109/ICCV.2009.5459452
16. S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M. H. Yang, Restormer: Efficient transformer for high-resolution image restoration, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2022), 5728–5739. https://doi.org/10.1109/CVPR52688.2022.00564
17. Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, H. Li, Uformer: A general U-shaped transformer for image restoration, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2022), 17683–17693. https://doi.org/10.1109/CVPR52688.2022.01716
18. S. G. Mallat, A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intell.*, **11** (1989), 674–693. https://doi.org/10.1109/34.192463

19. D. L. Donoho, De-noising by soft-thresholding, *IEEE Trans. Inf. Theory*, **41** (1995), 613–627. https://doi.org/10.1109/18.382009

20. S. G. Chang, B. Yu, M. Vetterli, Adaptive wavelet thresholding for image denoising and compression, *IEEE Trans. Image Process.*, **9** (2000), 1532–1546. https://doi.org/10.1109/83.862633

21. J. Portilla, V. Strela, M. J. Wainwright, E. P. Simoncelli, Image denoising using scale mixtures of Gaussians in the wavelet domain, *IEEE Trans. Image Process.*, **12** (2003), 1338–1351. https://doi.org/10.1109/TIP.2003.818640

22. C. J. Prabhakar, P. U. P. Kumar, An image based technique for enhancement of underwater images, *Int. J. Mach. Intell.*, **3** (2010), 217–224.

23. M. Jian, X. Liu, H. Luo, X. Lu, H. Yu, J. Dong, Underwater image processing and analysis: A review, *Signal Process. Image Commun.*, **91** (2021), 116088. https://doi.org/10.1016/j.image.2020.116088

24. P. Liu, H. Zhang, K. Zhang, L. Lin, W. Zuo, Multi-level wavelet-CNN for image restoration, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, (2018), 773–782. https://doi.org/10.1109/CVPRW.2018.00121

25. H. Huang, R. He, Z. Sun, T. Tan, Wavelet-SRNet: A wavelet-based CNN for multi-scale face super resolution, in *2017 IEEE International Conference on Computer Vision (ICCV)*, (2017), 1689–1697. https://doi.org/10.1109/ICCV.2017.187

26. K. Jiang, Z. Wang, P. Yi, C. C. Wang, Z. Wang, Z. Han, et al., Direction-aware attention wavelet network for image deraining, in *Proceedings of the 31st ACM International Conference on Multimedia*, (2023), 7065–7074. https://doi.org/10.1145/3581783.3611697

27. Y. Rao, W. Zhao, Y. Tang, J. Zhou, S. N. Lim, J. Lu, HorNet: Efficient high-order spatial interactions with recursive gated convolutions, *Adv. Neural Inf. Process. Syst.*, **35** (2022), 10353–10366.

28. M. H. Guo, C. Z. Lu, Z. N. Liu, M. M. Cheng, S. M. Hu, Visual attention network, *Comput. Vis. Media*, **9** (2023), 733–752. https://doi.org/10.1007/s41095-023-0364-2

29. S. K. Li, X. Wang, H. Zuo, C. Qian, S. Pu, M. Li, MogaNet: Multi-order gated aggregation network, in *International Conference on Learning Representations*, (2024), 37393–37427.

30. X. Li, W. Wang, X. Hu, J. Yang, Selective kernel networks, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2019), 510–519. https://doi.org/10.1109/CVPR.2019.00060

31. H. Zhang, C. Wu, Z. Zhang, Y. Zhu, H. Lin, Z. Zhang, et al., ResNeSt: Split-attention networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, **44** (2022), 5765–5776. https://doi.org/10.1109/CVPRW56347.2022.00309

32. B. Yang, G. Bender, Q. V. Le, J. Ngiam, CondConv: Conditionally parameterized convolutions for efficient inference, preprint, arXiv:1904.04971.

33. C. Li, A. Zhou, A. Yao, Omni-dimensional dynamic convolution, preprint, arXiv:2209.07947.

34. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2015), 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965

35. G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), 4700–4708. https://doi.org/10.1109/CVPR.2017.243

36. J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018), 7132–7141. https://doi.org/10.1109/CVPR.2018.00745

37. S. Woo, J. Park, J. Y. Lee, I. S. Kweon, CBAM: Convolutional block attention module, in *Computer Vision – ECCV 2018*, (2018), 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

38. Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image super-resolution, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018), 2472–2481. https://doi.org/10.1109/CVPR.2018.00262

39. Q. Jiang, Y. Chen, G. Wang, T. Ji, A novel deep neural network for noise removal from underwater image, *Signal Process. Image Commun.*, **87** (2020), 115921. https://doi.org/10.1016/j.image.2020.115921

40. H. Sheng, X. Wang, C. Zhang, J. Wang, P. Duan, Y. Wang, AIGC video detection based on the fusion of spatial-frequency-optical flow multimodal features, *J. Syst. Eng. Electron.*, **36** (2026), 1–15. https://doi.org/10.23919/JSEE.2026.000049

41. H. Zhao, O. Gallo, I. Frosio, J. Kautz, Loss functions for image restoration with neural networks, *IEEE Trans. Comput. Imaging*, **3** (2016), 47–57. https://doi.org/10.1109/TCI.2016.2644865

42. J. Yang, H. Xie, N. Xue, A. Zhang, Research on underwater image denoising based on dual-channel residual network, *Comput. Eng.*, **49** (2023), 188–198. https://doi.org/10.19678/j.issn.1000-3428.0064662

43. M. J. Islam, Y. Xia, J. Sattar, Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception, in *Robotics: Science and Systems (RSS)*, 2020. https://doi.org/10.15607/RSS.2020.XVI.018

44. M. J. Islam, Y. Xia, J. Sattar, Fast underwater image enhancement for improved visual perception, *IEEE Robot. Autom. Lett.*, **5** (2020), 3227–3234. https://doi.org/10.1109/LRA.2020.2974710

45. R. Timofte, E. Agustsson, L. Van Gool, M. Yang, L. Zhang, B. Lim, et al., NTIRE 2017 challenge on single image super-resolution: Dataset and study, in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, (2017) 1122–1131. https://doi.org/10.1109/CVPRW.2017.149

46. R. Franzen, Kodak lossless true color image suite, 1999.

47. D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, (2001), 416–423. https://doi.org/10.1109/ICCV.2001.937655

48. T. Q. Huynh-Thu, M. Ghanbari, Scope of validity of PSNR in image/video quality assessment, *Electron. Lett.*, **44** (2008), 800–801. https://doi.org/10.1049/el:20080522

49. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity, *IEEE Trans. Image Process.*, **13** (2004), 600–612. https://doi.org/10.1109/TIP.2003.819861

50. R. Zhang, P. Isola, A. A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018), 586–595. https://doi.org/10.1109/CVPR.2018.00068

51. C. Tian, Y. Xu, W. Zuo, L. Fei, C. W. Lin, Attention-guided CNN denoising network (ADNet), *Neural Networks*, **121** (2020), 461–473. https://doi.org/10.1016/j.neunet.2019.12.024

52. C. Tian, Y. Xu, L. Fei, J. Wang, J. Wen, N. Luo, Enhanced CNN for image denoising, *CAAI Trans. Intell. Technol.*, **4** (2019), 17–23. https://doi.org/10.1049/trit.2018.1054

53. C. Tian, Y. Xu, W. Zuo, B. Du, C. W. Lin, D. Zhang, Designing and training of a dual CNN for image denoising, *Knowl.-Based Syst.*, **226** (2021), 107145. https://doi.org/10.1016/j.knosys.2021.106949

54. I. Batool, M. Imran, A dual residual dense network for image denoising, *Eng. Appl. Artif. Intell.*, **147** (2025), 110275. https://doi.org/10.1016/j.engappai.2025.110275

55. C. Tian, M. Zheng, W. Zuo, B. Zhang, Y. Zhang, D. Zhang, Multi-stage image denoising with the wavelet transform, *Pattern Recognit.*, **134** (2023), 109050. https://doi.org/10.1016/j.patcog.2022.109050

AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)