International Journal of Science and Engineering Applications
Volume 14-Issue 12, 04 – 11, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1412.1002

Dynamic Large-Kernel Detail Enhancement Network for Underwater Image Super-Resolution

Tingting Jia*1st
School of Artificial Intelligence
Hubei University
Wuhan, China

Yuzhang Chen
School of Artificial Intelligence
Hubei University
Wuhan, China

Han Wang
School of Artificial Intelligence
Hubei University
Wuhan, China

Yuqi Ge
School of Artificial Intelligence
Hubei University
Wuhan, China

Abstract: Underwater imaging is degraded by light absorption, scattering, and suspended particles, which reduce clarity, suppress brightness, and diminish fine details.To address these challenges, we propose a lightweight Dynamic Large-Kernel Detail Enhancement Network (DLKDENet). The proposed framework integrates a Dynamic Global Feature Adaptation Module (DGFAN) and a Detail Perception Enhancement Module (DPEN), enabling coordinated optimization between global structural modeling and local detail restoration.DGFAN leverages large-receptive-field convolutions and a two-stage channel-attention strategy to attenuate noise and promote global structural consistency. In complement, DPEN enhances fine-grained texture reconstruction through overlapping patch decomposition and a gated convolution mechanism.The experimental results show that DLKDENet achieves consistent gains in PSNR and SSIM over several baseline models on the USR-248 and UFO-120 datasets. Furthermore, it delivers high-quality reconstruction with fewer parameters, demonstrating its efficiency and strong adaptability to underwater imaging conditions.

Keywords: Underwater Imaging, Super-Resolution, Large-Kernel Convolution, Transformer

1. Introduction

Accurate perception of the marine environment has become a fundamental requirement for national resource development, security, and oceanographic research. As ocean exploration extends to deeper, more distant, and more complex regions, underwater robots equipped with optical, acoustic, and inertial multimodal sensors are increasingly replacing manual operations, undertaking critical tasks such as environmental monitoring, seabed mapping, target localization, and resource assessment[1].However, the quality of underwater optical imaging is degraded by multiple factors, including medium absorption, backscattering, suspended particle distribution, and lighting fluctuations, which often result in color distortions, low contrast, and blurred details. Consequently, recovering high-fidelity structures and textures from low-quality underwater observations remains a critical scientific challenge for enabling intelligent underwater perception systems.Super-resolution (SR) techniques offer a promising approach to enhancing underwater image quality. As a fundamental task in computer vision, single image super-resolution (SISR) seeks to reconstruct high-resolution (HR) images from their low-resolution (LR) counterparts.Pioneering work such as SRCNN[2] first introduced convolutional neural networks into single-image super-resolution, enabling an end-to-end mapping between low- and high-resolution images. Subsequent approaches, including VDSR[3], EDSR[4], and SMSR[5], improved reconstruction performance by incorporating deeper architectures and residual learning mechanisms. Meanwhile, models such as RCAN[6] and RDN[7] further advanced feature representation through channel attention and dense connections, marking significant progress in the development of super-resolution techniques.Recent years have witnessed the rapid rise of Transformer-based architectures in image restoration, driven by their strong capability to model long-range dependencies. Methods such as SwinIR [8] and Restormer[9] leverage hierarchical windowed attention and gated feature modulation, respectively, and have demonstrated compelling performance across denoising, deblurring, and super-resolution tasks.

Despite rapid progress in general SISR techniques, their performance remains limited when addressing the complex degradations inherent to underwater images. To overcome these limitations, we propose the Dynamic Large-Kernel Detail Enhancement Network (DLKDENet), which jointly models global structural dependencies and fine-grained local details.At the core of the framework is the Dynamic Global-Feature Adaptation Network (DGFAN), which employs large-kernel convolutions to substantially expand the receptive field and incorporates a dynamic convolution aggregation mechanism that adaptively highlights structurally salient regions, thereby establishing a more reliable structural prior.In addition, we develop the Detail-Perception Enhancement Network (DPEN), which operates on overlapped local patches and employs lightweight self-attention to emphasize texture-rich regions, while a gated bottleneck design effectively discriminates true details from degradation-induced noise.By synergistically integrating DGFAN and DPEN, DLKDENet establishes a progressive optimization pipeline in which structural restoration and detail enhancement reinforce each other, ultimately improving both reconstruction fidelity and generalization in underwater super-resolution.

International Journal of Science and Engineering Applications
Volume 14-Issue 12, 04 – 11, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1412.1002

2. Method

2.1 Architecture

The overall architecture of DLKDENet comprises three main components: shallow feature extraction, deep feature extraction, and image reconstruction. The detailed network structure is illustrated in Figure 1.

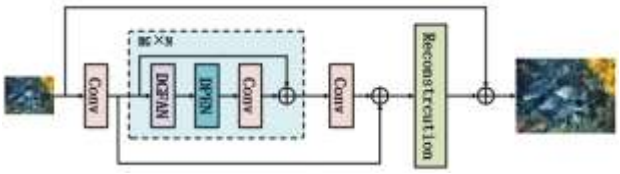

Figure. 1 Architecture of the proposed DLKDENet

In the shallow feature extraction stage, the low-resolution(LR) image I_{LR} is processed using a 3×3 convolution to obtain the shallow feature. These features F_0 are then fed into the deep feature extraction network to produce deep features F_n. The deep network is composed of multiple stacked residual groups, each consisting of three submodules: DGFAN, DPEN, and a convolutional layer. Residual connections are introduced to mitigate gradient vanishing during deep network training and to stabilize the optimization process.The resulting deep features F_n and shallow features F_0 are element-wise summed to integrate low-frequency content with high-frequency details, thereby enhancing the overall feature representation. The fused features are then passed through the image reconstruction network, where upsampling operations restore the spatial resolution. A residual connection with the bilinearly upsampled low-resolution image I_{LR} is applied to further refine fine details, ultimately producing the high-resolution output I_{HR} .

2.2 Dynamic Global-Feature Adaptive Network

To address the challenges of contrast degradation and structural blurring in underwater images caused by absorption and scattering, we propose the Dynamic Global-Feature Adaptive Network(DGFAN), which sequentially integrates the DynamicKernelUnit—built upon large-kernel convolutions to enhance edge and detail representation and mitigate blurring and texture loss—and a two-stage channel attention module (LGCA) that adaptively selects high-quality features while suppressing irrelevant or noisy information introduced by suspended particles and scattering. Residual connections are incorporated throughout the architecture to facilitate training and enable multi-scale feature extraction and enhancement tailored to underwater degradation characteristics. By achieving synergistic optimization of global and local feature extraction with adaptive feature modeling, the network provides efficient and highly expressive feature representations for underwater super-resolution, as illustrated in Figure 2.

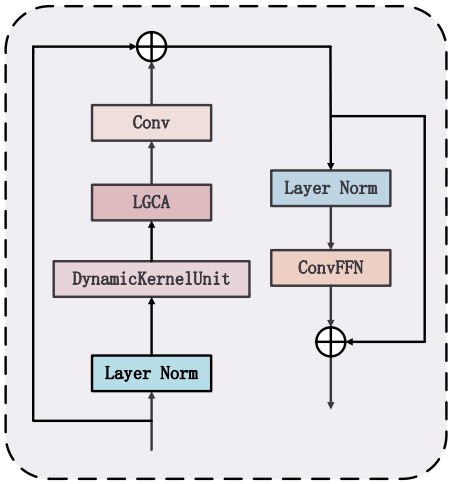

Figure. 2 Illustration of the proposed Dynamic Global-Feature Adaptive Network

2.2.1DynamicKernelUnit

Building on the large-kernel convolution concept [10], we propose the DynamicKernelUnit, a unified module in which the Large-Kernel Perception (LKP) first captures long-range spatial dependencies and overall scene layout to enhance underwater illumination, object contours, and global color consistency, followed by a Small-Kernel Aggregation (SKA) that, guided by LKP-generated weights, performs fine-grained modeling of local textures to improve edge and detail representation.

Specifically, the input feature $F_i \in \mathbb{R}^{H\times W\times C}$ is first processed by the LKP module, where a 1×1 convolution reduces the channel dimension by half to lower computational cost, followed by a ReLU activation to enhance nonlinear feature representation. A subsequent 9×9 large-kernel depthwise separable convolution captures long-range spatial context, expanding the receptive field to encode the global scene layout. Further pointwise convolutions are applied for channel remapping and activation, after which a 1×1 convolution generates dynamic convolution kernel weights. Group normalization is employed to stabilize parameter optimization, and the output is ultimately reshaped into G groups of 5×5 convolution kernels, providing adaptive weights $W \in \mathbb{R}^{H\times W\times D}$ for subsequent feature aggregation. The entire LKP process can be formally expressed as follows:

$$F_{LKP} = W_i = PC\left(PC\left(DW_{9\times9}\left(PC\left(F_i\right)\right)\right)\right) \quad (1)$$

where $W_i \in \mathbb{R}^{D}$ denotes the generated weight for F_i , $PC(\cdot)$ represents a pointwise convolution, and $DW_{9\times9}(\cdot)$ refers to the 9×9 large-kernel depthwise separable convolution used for subsequent feature aggregation, and F_{LKP} corresponds to the output of the LKP module.

In the SKA stage, the dynamic convolutional kernel weights generated by LKP are applied to the input feature through a group convolution operation. Specifically, the channels of the feature map F_i are partitioned into G groups, and all channels within each group share the same aggregation weights. This design reduces memory usage and computational cost, thereby making the module suitable for lightweight underwater SR

models.For each input feature map F_i , the corresponding weight W_i generated by the large-kernel perception step is reshaped to produce $W_i \in \mathbb{R}^{G\times K_S\times K_S}$, where $K_S \times K_S$ denotes a 5×5 convolution kernel. These weights are then used to aggregate highly correlated contextual information within a $K_S \times K_S$ spatial neighborhood centered on each position of F_i . This adaptive weighting enables fine-grained feature representation and enhances the model's sensitivity to dynamic and complex variations across different underwater backgrounds. The overall SKA process can be formulated as:

$$F_{SKA} = W_{ig} * N_{K_S}(F_{ic}) \quad (2)$$

where N_{K_S} denotes the $K_S \times K_S$ neighborhood centered at the input feature map F_i , and F_{ic} represents the c-th channel belonging to the g-th channel group. By performing a convolutional operation between the neighborhood $N_{K_S}(F_{ic})$ and the corresponding dynamic weights $W_{ig} \in \mathbb{R}^{K_S\times K_S}$, we obtain the aggregated feature representation F_{SKA} .

The enhanced feature produced by the SKA stage is first normalized using BatchNorm2d to stabilize its statistical distribution and is then added to the original input through a residual connection. This design preserves the low-level semantic information of the input while effectively integrating the global context captured by LKP and the fine-grained local details emphasized by SKA. This design thereby avoids the computational overhead associated with using large-kernel convolutions alone and mitigates the limited receptive field inherent to small-kernel operations. The overall formulation of the DynamicKernelUnit is expressed as follows:

$$F_{DKU} = BN\left(F_{SKA}\left(F_i, F_{LKP}(F_i)\right)\right) + F_i \quad (3)$$

where F_{DKU} denotes the output of the DynamicKernelUnit, and $BN(\cdot)$ represents batch normalization. The DynamicKernelUnit effectively balances global context modeling with local detail enhancement, offering a more precise adaptation to complex visual scenes compared to conventional large-kernel convolutions, and significantly improving feature integration in lightweight models.

2.2.2Dual-Stage Local-Global Channel Attention

The Dual-Stage Local-Global Channel Attention (LGCA) module evaluates channel importance from both local dependencies and global correlations, enabling fine-grained channel-wise feature enhancement. Through its two-stage weighting mechanism, LGCA progressively refines channel responses while preserving the original spatial structure, effectively emphasizing salient features and suppressing channel-level interference caused by light scattering and color attenuation in underwater images, thereby improving the discriminative capacity of the feature representations.

In the first stage, the module focuses on local channel dependencies and neighborhood patterns, emphasizing critical features such as edges and textures. This operation enables continuous modeling of inter-channel relationships without introducing significant additional parameters, allowing the model to more accurately capture fine-grained complementary interactions between adjacent features, thereby enhancing the response of texture and edge regions. The process can be formally expressed as follows:

$$F_{first} = Sigmoid\left(Conv1d_{3\times3}\left(AP(F_{DKU})\right)\right) \otimes F_{DKU} \quad (4)$$

where F_{first} denotes the output feature from the first-stage channel attention, $AP(\cdot)$ represents adaptive average pooling, $Conv1d_{3\times3}$ is a one-dimensional convolution with a kernel size of 3×3 , and $Sigmoid(\cdot)$ denotes the activation function.

In the second stage, channel-wise features are globally recalibrated to capture semantic-level importance, leveraging cross-channel non-linear interactions to generate global channel attention weights for a refined feature selection. Specifically, the first-stage output features are again processed by adaptive average pooling to produce a channel-level global descriptor. This descriptor is then passed through a two-layer multi-layer perceptron (MLP) with ReLU activation for dimensionality reduction, non-linear transformation, and expansion. The final global channel attention weights are obtained via a Sigmoid function and multiplied element-wise with the first-stage weights to yield the refined output. The formulation for this stage is expressed as follows:

$$F_{second} = Sigmoid\left(AP\left(F_{first}\right)\right) \otimes F_{first} \quad (5)$$

where F_{second} denotes the output features of the second-stage channel attention. This stage aims to integrate global contextual information and recalibrate the overall feature distribution, thereby balancing the local enhancement achieved in the preceding stage.

2.3 Detail Perception Enhancement Network

To address the limitations in early-stage detail restoration, we propose the Detail Perception Enhancement Network (DPEN), which achieves a breakthrough in local feature refinement through the synergistic design of block-wise self-attention and gated feature enhancement. In DPEN, the input feature map is first partitioned into overlapping local patches according to a predefined stride and patch size. Using a local feature partitioning function[11], the entire feature map is divided into multiple local blocks, enabling more precise modeling of texture structures and edge details within the image.Within each local feature block, a multi-head self-attention mechanism based on the Transformer architecture[8] is applied to effectively capture spatial dependencies and enhance the modeling of edge structures and texture details. The processed blocks are then reassembled using a local feature reconstruction function, restoring them to their original positions. This ensures that the reconstructed features are continuous, the boundaries appear natural, and the overall structural consistency of the image is preserved.Subsequently, a multi-stage bottleneck convolutional structure is employed to enhance salient features while suppressing low-value and redundant information, enabling adaptive refinement of complex underwater scattering and attenuation characteristics. This design excels in preserving fine details and maintaining structural integrity, ultimately improving the quality of reconstructed images through multi-stage collaborative operations. The architecture of the DPEN network is illustrated in Figure 3.

International Journal of Science and Engineering Applications
Volume 14-Issue 12, 04 – 11, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1412.1002

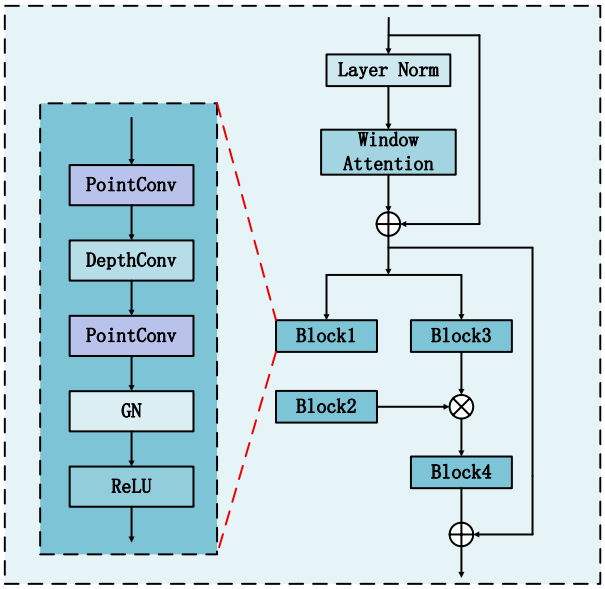

Figure. 3 Structure of the Detail Perception Enhancement Network.

In the DPEN, features from the preceding stage are first partitioned into overlapping local patches, over which a multi-head self-attention mechanism computes contextual feature weights. The resulting representations are then processed by the Gated Feature Enhancement module, which employs a cascaded bottleneck convolutional structure comprising Block1 through Block4. Specifically, Block1 and Block2 concentrate on refining and enhancing local details, whereas Block3 and Block4 establish long-range semantic dependencies. Meanwhile, the cross-block residual connection ensures that low-level edge and texture information is preserved throughout deeper layers, effectively mitigating detail degradation associated with increased network depth. This design thereby enables robust feature enhancement under complex underwater scattering and attenuation conditions. For a given feature map F_{DGFAN} , the computational process of the DPEN network can be formulated as follows:

$$\begin{aligned} F_{LN} &= LayerNorm\left(F_{DGFAN}\right) \\ F_O &= MSA\left(F_{LN}W_Q, F_{LN}W_K, F_{LN}W_V\right) + F_{DGFAN} \\ F_{DPEN} &= Block_4\left(Block_2\left(Block_1\left(F_O\right)\right) \square Block_3\left(F_O\right)\right) + F_O \end{aligned} \quad (6)$$

where F_{DPEN} denotes the output of the DPEN module, and W_Q , W_K , and W_V are the weight matrices shared across blocks. The DPEN module enables the model to establish stronger correlations between global and local information across different spatial locations, significantly enhancing the modeling of fine-grained textures and long-range dependencies. Consequently, it achieves superior image reconstruction quality, excelling in both detail preservation and structural integrity.Each bottleneck convolution $Block(\square)$ consists of a pointwise convolution $PC(\square)$ followed by a depthwise convolution $DW(\square)$, together with group normalization $GN(\square)$ and a ReLU activation. This design reduces computational complexity while maintaining adequate feature representation capacity.

The bottleneck convolution blocks, denoted as Block(.), employ differentiated Group Normalization (GN) configurations and depthwise convolution strategies to achieve multi-level feature enhancement. Specifically, Block1 and Block2 use GN with the number of groups set to $\frac{1}{16}$ of the channel dimension, combined with a 3×3 depthwise convolution with a padding of 1, focusing on the extraction and enhancement of local structural features. This design, leveraging dense receptive-field coverage provided by small convolution kernels, effectively captures edge contours and fine texture details while preserving the spatial resolution of the feature maps. In contrast, Block3 and Block4 adopt 1×1 depthwise convolutions with a padding of 0, combined with distinct GN group strategies. Block3 maintains a group number equal to $\frac{1}{16}$ of the channel dimension, whereas Block4 uses a fixed configuration of 8 groups. This differentiated design enables multi-scale feature interactions, transitioning from local detail refinement to global semantic modeling. The use of depthwise convolutions reduces computational cost while facilitating cross-channel feature reorganization and information integration, allowing the model to adaptively modulate feature responses across channels. The computational process can be formalized as follows:

$$F_{Block_i} = \mathrm{Re}LU\left(GN\left(PC\left(DW\left(PC\left(F_O\right)\right)\right)\right)\right), i = 1,2,3,4 \quad (7)$$

where $PC(\square)$ denotes the pointwise convolution operation, and $DW(\square)$ represents the depthwise convolution operation, combined with Group Normalization $GN(\square)$ and the $\mathrm{Re}LU(\square)$ activation function. This design enables the network to adapt to detail attenuation caused by scattering and absorption in underwater environments, while significantly enhancing texture preservation and local structural representation at a relatively low computational cost. Consequently, it provides the subsequent reconstruction network with high-quality, discriminative feature representations.

3. Experiments

To evaluate the effectiveness of the proposed DLKDENet, we conducted a comprehensive analysis of its reconstruction performance in comparison with mainstream networks, including SRCNN[2], SMSR[5], SwinIR[8], LKDN[12], AGDN[13], HPINet[11], and CATANet[14]. Both quantitative metrics and qualitative visual assessments were employed to assess the reconstruction results. Additionally, ablation studies were performed to validate the contributions of the individual modules within the DLKDENet architecture.

We employed two real-world underwater image datasets, USR-248 and UFO-120, for both training and evaluation. These datasets were meticulously constructed and publicly released by the Interactive Robotics and Vision Laboratory at the University of Minnesota. Network performance was rigorously quantified using three complementary metrics: peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and underwater image quality measure (UIQM). All images were transformed into the YCbCr color space, with evaluations conducted specifically on the luminance (Y) channel to ensure precision in assessing perceptual fidelity.

3.1 Comparisons with the state-of-the-arts

To evaluate the performance and advantages of DLKDENet across different super-resolution scales, we conducted a quantitative analysis comparing it with several benchmark networks. Table 1 presents the results of three image quality

metrics—PSNR, SSIM, and UIQM—on the USR-248 and UFO-120 datasets for reconstruction scales of ×2 and ×4. In the table, bolded values indicate the best performance, while values highlighted in blue denote the second-best results.

Table 1. Quantitative Comparison of Performance Across Different Networks

Method	Scale	Params (M)	USR-248			UFO-120		
			PSNR(dB)	SSIM	UIQM	PSNR(dB)	SSIM	UIQM
SRCNN	×2	0.067	29.82	0.8120	2.5220	25.75	0.7031	2.3837
SMSR	×2	0.985	31.10	0.8456	2.6513	27.09	0.7662	2.4331
SwinIR-light	×2	0.878	31.56	0.8589	2.6759	27.40	0.7731	2.4801
LKDN	×2	0.303	31.53	0.8527	2.6761	27.26	0.7728	2.5195
AGDN	×2	0.289	31.52	0.8615	2.7103	27.39	0.7815	2.5210
HPINet	×2	0.780	31.53	0.8621	2.6749	27.34	0.7766	2.4923
CATANet	×2	0.477	31.59	0.8629	**2.7198**	27.40	0.7772	2.5202
DLKDENet	×2	0.391	**31.60**	**0.8632**	2.7131	**27.52**	**0.7826**	**2.5232**
SRCNN	×4	0.067	26.07	0.6758	2.3471	24.49	0.6189	2.2754
SMSR	×4	0.985	27.19	0.7098	2.4717	26.11	0.6731	2.3773
SwinIR-light	×4	0.897	27.66	0.7137	2.4801	26.32	0.6890	2.3928
LKDN	×4	0.321	27.60	0.7186	2.4831	26.31	0.6911	2.3801
AGDN	×4	0.293	27.69	0.7168	2.4978	26.30	0.6911	2.3899
HPINet	×4	0.900	27.57	0.7100	2.4920	26.33	0.7008	2.4010
CATANet	×4	0.535	**27.72**	**0.7213**	2.4957	26.34	**0.7109**	2.3905
DLKDENet	×4	0.449	27.70	0.7210	**2.5144**	**26.37**	**0.7109**	**2.4011**

Experimental results demonstrate that, under the ×2 super-resolution task, DLKDENet achieves superior performance across all evaluation metrics compared to other benchmark networks, delivering enhanced image quality while maintaining a compact parameter footprint. On the USR-248 dataset, DLKDENet's UIQM score differs from CATANet by approximately 0.0067, and its SSIM performance is comparable, yet its parameter count is significantly lower. These findings indicate that DLKDENet preserves robust feature representation and reconstruction capabilities within a lightweight design, exemplifying an optimal balance between performance and computational efficiency.

In evaluations on the USR-248 and UFO-120 datasets, DLKDENet and other benchmark networks generally exhibit lower evaluation metrics for ×4 super-resolution reconstruction compared to ×2, indicating the commonly observed performance degradation at higher upscaling factors. Specifically, on the USR-248 dataset, DLKDENet achieves a UIQM score of 2.5144, outperforming LKDN and CATANet and demonstrating superior image quality and perceptual consistency. On the UFO-120 dataset, DLKDENet attains a PSNR 0.03 dB higher than CATANet while maintaining a smaller parameter count, indicating that the network preserves strong detail reconstruction and visual quality while maintaining a lightweight architecture. In summary, DLKDENet demonstrates robust performance and superior structural restoration across varying reconstruction scales and diverse underwater scenarios.

In addition to quantitatively evaluating the performance of DLKDENet using objective metrics, we further conducted a comparative analysis from the perspective of subjective visual perception. To visually illustrate differences in detail

International Journal of Science and Engineering Applications
Volume 14-Issue 12, 04 – 11, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1412.1002

restoration and texture reconstruction across networks, representative samples with upscaling factors of ×2 and ×4 were selected from the USR-248 and UFO-120 test sets, respectively. The red-framed regions in the low-resolution images were locally enlarged to facilitate detailed comparison.

Super-Resolution Reconstruction Results (×2): For the ×2 super-resolution reconstruction, as illustrated in Figure 4, a scene featuring the Amphiprion species from the USR-248 dataset was selected. DLKDENet more accurately restores fine textures, with clear and continuous edge contours. The background rocks and algae exhibit rich structural layers, resulting in an overall visual appearance that closely approximates the high-resolution reference image.

Super-Resolution Reconstruction Results (×4): When the reconstruction scale is increased to ×4, as shown in Figure 5, a coral scene from the UFO-120 dataset is selected. DLKDENet effectively preserves the structural continuity and depth hierarchy of coral branches, substantially mitigating texture discontinuities and artifacts. The network also delivers natural color reproduction and rich details, resulting in perceptual quality that significantly surpasses that of other benchmark methods.

Figure 4. Qualitative comparison of ×2 super-resolution reconstruction for im_xb_1146 from the USR-248 dataset.

Figure 5. Qualitative comparison of ×4 super-resolution reconstruction for set_o49 from the UFO-120 dataset.

3.2 Ablation Study

To systematically evaluate the effectiveness and design rationale of DLKDENet's key modules, ablation experiments were conducted under ×2 super-resolution conditions using the USR-248 and UFO-120 datasets. These experiments primarily examined the impact of each module within the deep feature extraction stage on reconstruction performance.The deep feature extraction stage is designed to comprehensively capture image edge information while establishing efficient interactions between local and global features, thereby enhancing overall reconstruction quality. To further assess the contribution of individual components, the DKU, LGCA, and DPEN modules were separately removed, and the network was evaluated on the USR-248 and UFO-120 datasets. Table 2 presents the PSNR, SSIM, and UIQM results corresponding to different module configurations.

Table 2. Quantitative Comparison of the Impact of Individual Modules in the Deep Feature Extraction Network on DLKDENet Performance

datasets	Params(M)	DGFAN		DPEN	PSNR	SSIM	UIQM
		DKU	LGCA				
USR-248	0.344	×	√	√	31.53	0.8619	0.7002
	0.384	√	×	√	31.49	0.8610	2.6889
	0.326	√	√	×	31.55	0.8620	2.7079
	0.337	×	×	√	31.57	0.8623	0.7119
	0.272	×	×	×	31.43	0.8591	2.6830
	0.391	√	√	√	31.60	0.8632	2.7131

UFO-120	0.344	×	√	√	27.43	0.7702	2.5192
	0.384	√	×	√	27.43	0.7725	2.5159
	0.326	√	√	×	27.41	0.7730	2.5099
	0.337	×	×	√	27.45	0.7735	2.5202
	0.272	×	×	×	27.38	0.7684	2.4651
	0.391	√	√	√	27.52	0.7826	2.5232

Analysis of the experimental results reveals that the removal of any single module leads to a noticeable decline in PSNR and SSIM across both datasets. This finding underscores the critical role of each module in feature enhancement and information integration: the DKU module effectively captures high-frequency edge details, the LGCA module facilitates adaptive fusion of local and global features, and the DPEN module preserves feature integrity through overlapping patch partitioning. The synergistic design of these three components enables DLKDENet to robustly enhance edge structures of complex underwater targets, such as marine organisms and vehicles, thereby achieving superior image reconstruction quality.

4. Conclusion

Underwater images are inherently affected by light absorption, scattering, and interference from suspended particles. These factors commonly result in color distortion, reduced contrast, and blurred textures, which compromise the accurate perception and utilization of visual information. To address these challenges, we propose a lightweight Dynamic Large-Kernel Detail Enhancement Network (DLKDENet) for underwater image super-resolution.The network implements a multi-stage reconstruction framework, progressing from shallow texture extraction through deep semantic enhancement to fine-grained texture restoration. By incorporating the Dynamic Global Feature Adaptive Network (DGFAN) and the Detail Perception Enhancement Network (DPEN), DLKDENet effectively balances global structural consistency with local detail refinement.Experimental results on the USR-248 and UFO-120 real-world underwater datasets demonstrate that DLKDENet consistently outperforms multiple state-of-the-art networks across various reconstruction scales, achieving superior performance in objective metrics such as PSNR, SSIM, and UIQM, while maintaining strong visual consistency and structural integrity. Ablation studies further confirm the effectiveness of each module and their synergistic contributions. Overall, DLKDENet strikes a well-balanced trade-off among feature representation, detail preservation, and computational efficiency, offering an effective and robust approach for underwater image super-resolution.

5. References

[1] Raveendran S, Patil M D, Birajdar G K. Underwater image enhancement: a comprehensive review, recent trends, challenges and applications[J]. Artificial Intelligence Review, 2021, 54: 5413-5467.

[2] Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resol ution[C]//Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, Septe mber 6-12, 2014, Proceedings, Part IV 13. Springer International Publishing, 2014: 184-199.

[3] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1646-1654.

[4] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017: 136-144.

[5] Wang L, Dong X, Wang Y, et al. Exploring sparsity in image super-resolution for efficient inf erence[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recogniti on. 2021: 4917-4926.

[6] Zhang Y, Li K, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 286-301.

[7] Zhang Y, Tian Y, Kong Y, et al. Residual dense network for image super-resolution[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 2472-2481.

[8] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, LucVan Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF international conference on computer vision, pages 1833–1844,2021. 1, 2, 3, 4, 5, 6, 7.

[9] Zamir S W, Arora A, Khan S, et al. Restormer: Efficient transformer for high-resolution image restoration[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 5728-5739.

[10] Wang A, Chen H, Lin Z, et al. LSNet: See Large, Focus Small[C]//Proceedings of the Computer Vision and Pattern Recognition Conference. 2025: 9718-9729.

[11] Liu J, Chen C, Tang J, et al. From coarse to fine: Hierarchical pixel integration for lightweight image super-resolution[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2023, 37(2): 1666-1674.

[12] Xie C, Zhang X, Li L, et al. Large kernel distillation network for efficient single image super -

resolution[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 1283-1292.

[13] Wang H, Wei Z, Tang Q, et al. Attention guidance distillation network for efficient image super-resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 6287-6296.

[14] Liu X, Liu J, Tang J, et al. CATANet: Efficient Content-Aware Token Aggregation for Lightweight Image Super-Resolution[C]//Proceedings of the Computer Vision and Pattern Recognition Conference. 2025: 17902-17912.