frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 05 January 2024
DOI 10.3389/fmars.2023.1328436

OPEN ACCESS

EDITED BY
Xiangrong Zhang,
Xidian University, China

REVIEWED BY
Deqiang Cheng,
China University of Mining and Technology, China
Wei Li,
Beijing Institute of Technology, China

*CORRESPONDENCE
Shichao Ma
mashch7@mail.sysu.edu.cn

RECEIVED 26 October 2023
ACCEPTED 11 December 2023
PUBLISHED 05 January 2024

CITATION
Liu B, Ning X, Ma S and Yang Y (2024) Multi-scale dense spatially-adaptive residual distillation network for lightweight underwater image super-resolution. *Front. Mar. Sci.* 10:1328436. doi: 10.3389/fmars.2023.1328436

Multi-scale dense spatially-adaptive residual distillation network for lightweight underwater image super-resolution

Bingzan Liu[1], Xin Ning[1], Shichao Ma[2*] and Yizhen Yang[1]

[1]School of Astronautics, Northwestern Polytechnical University, Xi'an, China, [2]School of Aeronautics and Astronautics, Sun Yat-Sen University, Shenzhen, China

Underwater images are typically of poor quality, lacking texture and edge information, and are blurry and full of artifacts, which restricts the performance of subsequent tasks such as underwater object detection, and path planning for underwater unmanned submersible vehicles (UUVs). Additionally, the limitation of underwater equipment, most existing image enhancement and super-resolution methods cannot be implemented directly. Hence, developing a weightless technique for improving the resolution of submerged images while balancing performance and parameters is vital. In this paper, a multi-scale dense spatially-adaptive residual distillation network (MDSRDN) is proposed aiming at obtaining high-quality (HR) underwater images with odd parameters and fast running time. In particular, a multi-scale dense spatially-adaptive residual distillation module (MDSRD) is developed to facilitate the multi-scale global-to-local feature extraction like a multi-head transformer and enriching spatial attention maps. By introducing a spatial feature transformer layer (SFT layer) and residual spatial-adaptive feature attention (RSFA), an enhancing attention map for spatially-adaptive feature modulation is generated. Furthermore, to maintain the network lightweight enough, blue separable convolution (BS-Conv) and distillation module are applied. Extensive experimental results illustrate the superiority of MDSDRN in underwater image super-resolution reconstruction, which can achieve a great balance between parameters (only 0.32M), multi-adds (only 13G), and performance (26.38 dB on PSNR in USR-248) with the scale of $\times 4$.

KEYWORDS

lightweight super-resolution reconstruction, multi-scale dense spatially adaptive residual distillation network, underwater image, deep learning, residual spatial adaptive feature attention

1 Introduction

It is widely recognized that 70% of the Earth is covered by water, indicating that it is a crucial development space for global ecology, resources, society, economy, and security Wang et al., 2017. In order to better utilize and develop marine resources, equipment such as unmanned underwater vehicles (UUVs), underwater probes, etc., have been widely used. The UUVs can perform diverse types of underwater detection tasks such as underwater resource detection Wang et al., 2023, coral reef inspection Mooney and Johnson, 2014, debris inspection Islam et al., 2019, and marine fisheries Barbedo, 2022, etc. During such processes, image synthesis and scene understanding based on high-resolution (HR) images are necessary. However, poor underwater visibility and the absorption and scattering effects of water contribute to the quality of underwater images are low, lack of details and edge information.

Specifically, various factors lead to such problems. The low contrast and darkness are caused by the decay of light rays with the increase of underwater depth Cheng et al., 2018. Additionally, the attenuation of red wavelengths of light when traveling through water causes underwater images to display bluish-greenish hues Ei et al., 2023. Furthermore, suspended particles can lead detailed textures to appear blurred Alenezi et al., 2022. On the other hand, limitations of hardware equipment and cost make it more difficult to acquire HR images directly. Therefore, many scholars have focused on researching underwater image processing with the aim of achieving HR images.

Nowadays, techniques for improving underwater images have become prominent. These techniques utilize prior knowledge via dark pass methodology Hu et al., 2018 as well as the Retinex algorithm Golts et al., 2020. Differential attenuation compensation (DAC) proposed by Lai et al., 2022 and Hybrid enhanced generative adversarial Network (HEGAN) designed by Li Y et al., 2022 are the typical examples. Nowadays, complex ocean exploration missions have high requirements for color distortion, image detail, contrast, and brightness Czub et al., 2018 but it is difficult to obtain enough prior knowledge for image preprocessing, especially in unfamiliar oceans. Consequently, end-to-end super-resolution (SR) based on convolutional neural networks (CNN) has gained significant interest from scholars in recent years. Chen et al. (Chen et al., 2019) applied modified dense blocks to CNN for SR of underwater images. Paper Helwig et al., 2023 integrated the residual dense block with the adaptive mechanism and proposed a residual-based underwater image SR method. Enlightened by IMDN, Yuan et al., 2023 incorporated the information distillation mechanism and spatial attention module into an ordinary residual network. Subsequently, in paper Li Z et al., 2022, blueprint separable convolution (BSC) was introduced to SR for underwater images. Aiming at improving the representational ability of high-frequency features, for underwater images, Restoration and Super-Resolution GAN (SRSRGAN) was introduced (Wang H. et al., 2023). However, few of these methods can strike a balance between performance and number of parameters and there is little research on deployable lightweight end-to-end method for underwater SR tasks. Therefore, it is essential to suggest a lightweight and high-quality SR method to achieve a balance between performance and computational cost.

To tackle these issues, a multi-scale dense spatially-adaptive residual distillation network (MDSRDN) is proposed. In the specification, a spatial feature Transformer layer and a multi-scale dense spatially-adaptive residual distillation module construct the main structure of MDSRDN, which can build long-range dependence like SwinIR. In addition, a residual spatial-adaptive feature attention (RSFA) is proposed aiming at realizing global feature extraction and obtaining enhanced multi-scale attention maps. Extensive experimental results exhibit that MDSRDN outperforms most of the state-of-the-art (SOTA) methods in low computational consumption for underwater image enhancement and super-resolution reconstruction. The main contributions of this paper are as follows:

- A multi-scale dense spatially-adaptive residual distillation network (MDSRDN) is proposed which can achieve a great balance between parameters, multi-adds, and performances. Furthermore, this network is an end-to-end mapping and do not need prior knowledge, which means the ability to deal with complex underwater scenarios.
- We design a multi-scale dense spatially-adaptive residual distillation module (MDSRD) to achieve multi-scale global-to-local feature extraction like a multi-head transformer, realize the generation of an attention map for spatially-adaptive feature modulation, and realize feature reuse in maximum.
- With the purpose of adaptive enhancing spatial features and acquiring long-range dependence, a lightweight and effective residual spatial-adaptive feature attention (RSFA) is constructed and a spatial feature transformer layer is introduced.
- Blue separable convolution (BS-Conv) which has been proven to be superior to deep separable convolution (DSC) and distillation module are applied to this network, which can reduce the number of parameters and multi-adds.
- Compared to current mainstream SR algorithms and enhancement techniques designed for underwater images, the MDSRDN presents clear advantages in terms of parameters, computational complexity, processing speed, and accuracy. The very fast running time and very low number of parameters determine that MDSRDN can be employed on edge underwater devices. Additionally, MDSRDN is generalized and achieves commendable results in benchmark datasets.

2 Related works

The scope of this paper concerns the super-resolution reconstruction of lightweight underwater images, as well as the enrichment of spatial attention maps and the extraction of spatial

features. Therefore, it is imperative to conduct a thorough review of pertinent literature within these categories.

2.1 Single image super-resolution

SR based on CNN is a method of restoring low-resolution (LR) images to HR images on the foundation of deep learning. The first model was introduced by Dong Dong et al., 2016, who used a three-layer CNN to learn the non-linear mapping relationship between LR to HR. Subsequently, an efficient sub-pixel convolutional network (ESPCN) was devised by Shi et al. Talab et al., 2019. They integrated sub-pixel convolution into the upsampling process which can expand the receptive field. These methods belong to the shallow feature extraction, which means high-frequency information such as detail texture could not be utilized. To address such issues, Kim et al. Kim et al., 2016a designed a very deep convolutional network (VDSR). Subsequently, the enhanced deep residual network (EDSR) Lim et al., 2017 added the number of layers to 160. Inspired by the EDSR, A vast number of in-depth networks like DRRN Tai et al., 2017 and DRCN Kim et al., 2016b emerged. Although these models achieved superior reconstruction results, the significant number of parameters and flops resulted in a substantial computational overhead, making it challenging to deploy these algorithms to UUVs.

In order to maintain the network lightweight enough, a very deep residual channel attention network (RCAN) Zhang et al., 2018 was proposed. This approach has the capacity to decrease parameter numbers by 40%, allowing for lighter network performance. Then channel attention (CA) was introduced to residual in the residual network (RIR), which meant that this network had the ability to bypass low-frequency information and focused on more important features with a smaller number of parameters. Afterward, an information distillation network (IDN) Hui et al., 2018 constructed the prototype of the distillation network. Inspired by IDN, Hui et al., 2019 modified the distillation block (DB) to an information multi-distillation block (IMDB), which can aggregate feature information by importance. Finally, a more concise feature extraction block realized by distillation connection was proposed in the residual feature distillation network (RFDN).

2.2 The progress of the spatial attention module

A spatial attention module is an adaptive mechanism for selecting spatial regions, enabling attention to be focused appropriately. Spatial attention modules can be categorized into four types Guo et al., 2022: RNN-based methods, prediction of the relevant region explicitly, prediction of the relevant region implicitly, and self-attention-based methods. In 2014, Mnih et al., 2014 developed the recurrent attention model (RAM), which first gave the network the ability to determine where to focus its attention. Since then, many RNN-based methods have been designed such as Gregor et al., 2015 and Xu et al., 2015. Subsequently, spatial transformer networks (STN) were designed by Jaderberg et al., 2015 to make the network pay more attention to the most relevant regions. According to STN, subsequent works such as Dai et al., 2017 have achieved greater success. Furthermore, to capture long-range dependence, a recalibration function in the spatial domain was proposed Hu et al., 2018. Afterward, inspired by Hu et al., 2018, a point-wise spatial attention network (PSANet) was implemented to effectively capture long-range dependence.

With a substantial boost in hardware capabilities, transformers were introduced into the field of computer vision by Dosovitskiy et al., 2021. This new architecture was named vision transformer (ViT) and obtained favorable results on numerous benchmark datasets, particularly with large datasets. The point cloud transformer (Pct) proposed by Guo et al., 2021 is an excellent illustration. The initial transformer for super-resolution reconstruction (SR) was SwinIR Liang et al., 2021, which yielded superior outcomes compared to the majority of CNN-based methods. Later on, numerous advanced SR techniques were developed. Nevertheless, transformer-based approaches disregard the image's structural information by replacing two-dimensional matrices with a one-dimensional vector. Furthermore, the large number of references and computational costs make these methods infeasible to deploy on edge devices such as UUVs. Therefore, an alternative method MDSRDN which can acquire and enhance spatial features adaptively is proposed. The employment of multi-scale residual feature representation in this network facilitates the extraction of global features and fosters the realization of long-range dependence.

3 The proposed method

In this section, the proposed multi-scale dense spatially-adaptive residual distillation network (MDSRDN) is introduced in detail, followed by an elaborate presentation of the multi-scale dense spatially-adaptive residual distillation module and spatial feature transformer layer (SFT layer). Then, we introduce residual spatial-adaptive feature attention (RSFA) which is considered the most significant element of the MDSRDN.

3.1 The overall design of MDSRDN

Shallow feature extraction module, deep feature extraction module, and Upsampling module compose the whole structure of MDSRDN, which is exhibited in Figure 1.

The objective of MDSRDN, as depicted in Figure 1, is to gain multi-scale global and local features by utilizing MDSRD, which can operate similarly to multi-head transformers and SwinIR. Then, aiming at enriching spatial attention maps and enhancing the expression and representation of spatial features, SFTlayer and RSFA are introduced. The blue separable convolution which can decrease the number of parameters and multi-adds with the kernel $i \times i$ is denoted by $BSConv - i$ in Figure 1. Supposing that, I_{sfe}, I_{dfe}, and I_{SR}refer to the output of shallow feature extraction module, deep feature extraction module, and the final result of MDSRDN respectively, the entire process of MDSRDN can be considered in

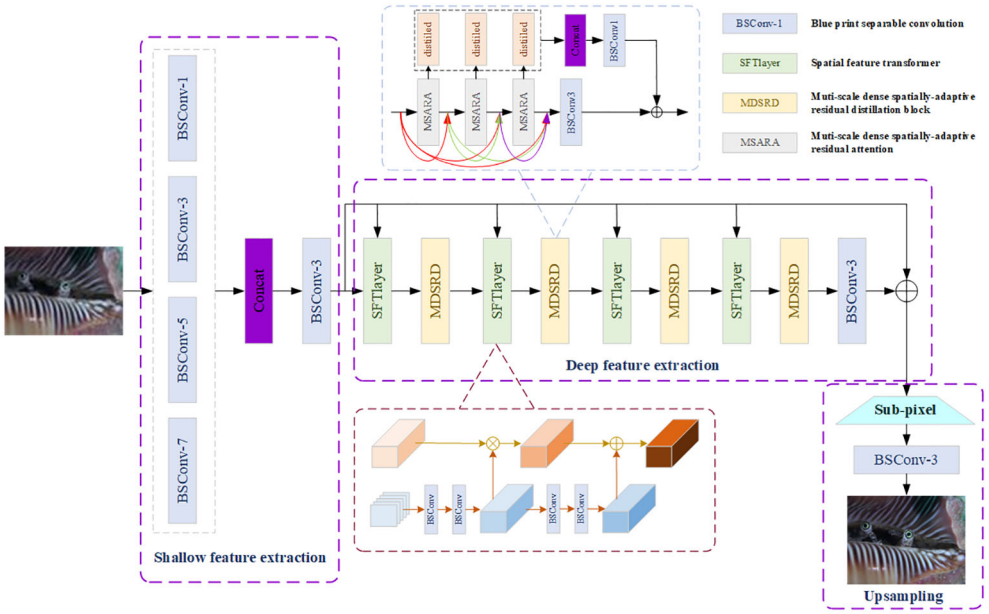

FIGURE 1
The main structure of MDSRDN.

Equation 1:

$$I_{sfe} = F_{sfe}(I_{LR})$$
$$I_{SR} = F_{up}(I_{sfe} + I_{dfe}), \quad (1)$$

where F_{sfe} and F_{up} denote the processing of shallow feature extraction and upsampling respectively. In detail, the upsampling module consists of a 3×3 BSC and a non-parametric sub-pixel operation. In addition, I_{LR} represents the input images belonging to $R^{H \times W \times C_{in}}$(H, W, and C_{in} are on behalf of the height, width, and input channel number of the input images, respectively).

3.2 Shallow feature extraction module

With the idea of Szegedy et al., 2015 and aiming at achieving more spatial features, a multi-scale shallow feature extraction module is proposed. Four different parallel BSCs with the kernel are applied to extract shallow features. 1×1, 3×3 BSC can obtain more local features for low-resolution (LR) images, and the BSC with kernel size equaling 5 and 7 determinate feature maps with a large receptive field. Figure 1 and Equation 2 illustrate the width-expanding part of shallow feature extraction.

$$L_i = F^i_{BSC}(I_{LR}) \quad (i = [1, 3, 5, 7]), \quad (2)$$

where i represents the kernel size and L_i denotes the output after different BSC operations. Subsequently, a concatenation operation along with a BSC decreases the number of channels to n, which stands for the output channel of the whole method. In this paper, we choose $n = 64$. The following equation (Equation 3) stands for the description of the output of the shallow feature extraction module:

$$I_{sfe} = F^3_{BSC}(Concat(L_1, L_3, L_5, L_7)) \quad (3)$$

3.3 Deep feature extraction module

The pivotal component of MDSRDN is the deep feature extraction module, encompassing multiple SFT layers and MDSRDs. More texture and edge features can be captured by this module.

3.3.1 Spatial feature transformer layer (SFT layer)

To improve spatial feature capture and acquisition of deeper spatial features, a SFT layer is incorporated (Wang et al., 2018) at the front of the MDSRD. By using this layer, the spatial attention maps constructed by MDSRD become more ample and more high-frequency features can be obtained. As is shown in Figure 1, the input of the SFT layer is the output of shallow feature extraction I_{sfe} and the result of the last MDSRD denoted as I^{j-1}_{MDSRDN}, where $j$stands for the $j-th$ SFT layer. Several 1×1BSC composes the mapping function denoted as Ψ and a modulation parameter pair can be acquired, which can be represented by (γ, β). Subsequently, the transformation of each intermediate feature map is carried out by scaling and shifting feature maps. This entire process is described in Equation 4.

$$(\gamma, \beta) = \Psi(I_{sfe})$$
$$I^j_{SFT} = \gamma \times I^{j-1}_{MDSRDN} + \beta, \quad (4)$$

where I^j_{SFT} refers to the final result of $j-th$ SFT layer.

3.3.2 Multi-scale dense spatially-adaptive residual distillation module

The MDSRD module is a pivotal component of the MDSRDN, which comprises three dense cross-multi-scale spatially-adaptive residual attention (MSARA). Due to the reuse of features and distillation mechanism, spatial features can be obtained and utilized in maximum and the whole module can remain lightweight. Enriching spatial attention maps, building long-range dependence, and achieving global-to-local feature extraction are the primary functions of MSARA, which will be discussed in detail in the upcoming section. The blue box in Figure 1 showcases the processing of MDSRD. We employ I^1_{MSARA}, I^2_{MSARA}, I^3_{MSARA} to express the consequences of the first, second, and the last MSARA. So, the result of MDSRD can be interpreted in Equation 5:

$$\begin{aligned} I^i_{dis-m}, I^i_{rem-m} &= split(F^i_{BSC}(I^j_{SFT}))\ \ (i = 3, 5) \\ I^m_{MSARA} &= F_{MSARA}(I^i_{rem-m}) + \sum_1 I^{m-1}_{MSARA} + I^j_{SFT}\ \ (m \in [1, 3]) \\ I_{dis} &= F^1_{BSC}(I^i_{dis-1}, I^i_{dis-2}, I^i_{dis-3}) \\ I_{MDSRD} &= I_{dis} + F^3_{BSC}(I^m_{MSARA}), \end{aligned} \tag{5}$$

where *split* represents the distillation mechanism, and I^i_{dis-m}, I^i_{dis-m} denotes the distillation part and the remained part of features, which m refers to the serial number of MSARA and i stands for the kernelsize of BSC. In this paper, the distillation rate is set as 0.25. Subsequently, we introduce F^i_{BSC} as the function of BSC, and I_{MDSRD} is used as the output of MDSRD.

3.3.3 Multi-scale spatially-adaptive residual attention

As is shown in Figure 2, MSARA enhances the capture of spatial features and enriches attention maps through multi-scale and residual mechanisms. Then, a residual spatially adaptive feature attention module is used to establish long-range dependencies and extract more local features. Subsequently, local features are extracted by using an enhanced spatial attention module (ESA) Liu J. et al., 2020 and several BSCs. Therefore, MSARA determines that we can obtain multi-scale global-to-local features like multi-head transformers and promote the performance of our model.

MSARA is divided into two parts, the global feature extraction part, and the local feature extraction part, with the output of the global feature extraction part being assumed as $I_{gl-MSARA}$. So, the entire process of the global feature extraction part can be expressed through Equation 6.

$$I_{gl-MSARA} = F_{RSFA}(I^5_{rem}) + F_{RSFA}(I^3_{rem}) + I^5_{rem} + I^3_{rem}, \tag{6}$$

where F_{RSFA} stands for the operation of RSFA.

For the local feature extraction part, we introduce an enhanced spatial attention module (ESA) to realize spatial-dimension response and BSC is used to obtain local features. Equation 7 expresses the consequence of MSARA denoted as I_{MSARA}.

$$\begin{aligned} I_{ESA} &= F_{ESA}(F^3_{BSC}(I_{gl-MSARA})) \\ I_{MSARA} &= F^3_{BSC}(I_{ESA} \times (F_{RSFA}(I^3_{rem}) + I^3_{rem})) + F^5_{BSC} \\ &\quad (I_{ESA} \times (F_{RSFA}(I^5_{rem}) + I^5_{rem})) \end{aligned} \tag{7}$$

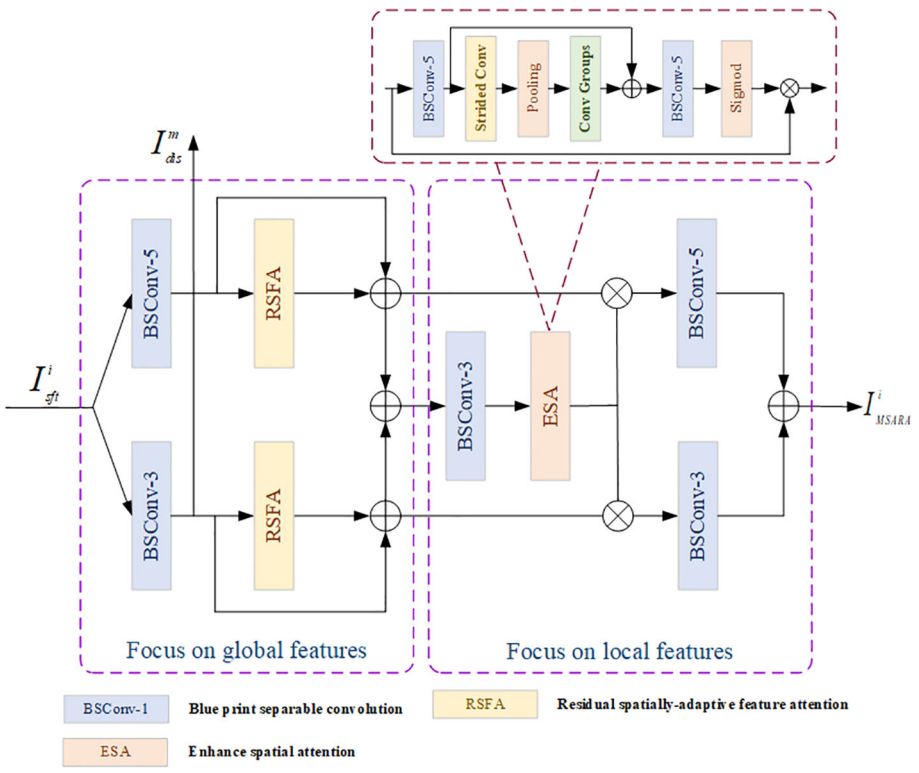

FIGURE 2
The overall design of MSARA.

where F_{ESA} refers to the mapping function of ESA, which is similar to the ESA in (Liu J. et al., 2020). In this paper, BSC is used to replace convolution in ESA. Noteworthily, $\times$ represents the matrix multiplication. Table 1 exhibited a pytorch-like process of MSARA, which can express the method intuitively. When x is the input of MSARA, *self*.*RSFA* denotes the proceeding of RSFA and *self*. *GELU* is introduced as the activation function.

3.3.4 Residual spatially-adaptive feature attention

Taking inspiration from Sun et al.'s (Sun et al., 2023) proposal of spatially adaptive feature modulation (SAFM), we propose RSFA, which enables the construction of long-range dependence from multi-scale feature representations and the enhancement of spatial feature capturing via a residual block (RSB). Therefore, more texture features and edge profiles can be obtained. The key structure of RSFA is comprised of an RSB and a SAFM, which is visually displayed in Figure 3. Furthermore, as is shown in Haase and Amthor, 2020, BSC is superior to SDC in most vision tasks. Therefore, we selected BSC to replace SDC which is chosen by Haase and Amthor, 2020.

For RSB, two 5×5 BSCs and an active function configure the main structure of RSB. So, the result of MDSRD can be interpreted in Equation 8:

$$I_{RSB} = F^5_{BSC}(\sigma(F^5_{BSC}(I_{rem}))), \tag{8}$$

where σ represents the Grelu function and I_{RSB} denotes to the output of RSB. The following equation (Equation 9) can express the procedure of revised SAFM.

$$\begin{gathered}[X_0, X_1, X_2, X_3] = split(I_{RSB}) \\ \widehat{X_0} = F^3_{BSC}(X_0) \\ \widehat{X_k} = \uparrow_p (F^3_{BSC}(\downarrow_{p/2^k}(X_k))), \quad (i \in [1,3])\end{gathered} \tag{9}$$

TABLE 1 The whole proceeding of MSARA by using a pytorch-like pseudocode.

A pytorch-like multi-scale spatially-adaptive residual attention
def forward (*self*, *x*) :
$x\_conv1 = self.GELU(self\ .\ BSConv3(x))$
$x\_conv2 = self.GELU(self\ .\ BSConv5(x))$
$xc1\_rem,\ xc1\_dis = split(x\_conv1)$
$xc2\_rem,\ xc2\_dis = split(x\_conv2)$
$x1\_gl = self.RSFA(xc1\_rem) + xc1\_rem$
$x2\_gl = self.RSFA(xc2\_rem) + xc2\_rem$
$g\_gl = x1\_gl + x2\_gl$
$g\_lo = self\ .\ ESA(self.GELU(self.BSConv3(g\_gl)))$
$x\_out1 = self.GELU(self.BSConv3(g\_lo \times x1\_gl))$
$x\_out2 = self.GELU(self.BSConv5(g\_lo \times x2\_gl))$
return $x\_out1 + x\_out2$

where *split* denotes the channel-wise distillation mechanism, X_k is the result after distillation and $\hat{X}_k$ represents the consequence of every branch. Furthermore, $\uparrow_p$ and $\downarrow_{p/2^k}$ refer to the upsampling features at a specific level to the original resolution p and downsampling features to the size of $p/2^k$, respectively. Subsequently, a concatenation operation, a BSC along with a matrix multiplication aggregate these features together, which is shown in Equation 10.

$$\begin{gathered}\widehat{X} = F^1_{BSC}(Concat([\widehat{X_0}, \widehat{X_1}, \widehat{X_2}, \widehat{X_3}])) \\ I_{RSFA} = \sigma(\widehat{X}) \times I_{rem},\end{gathered} \tag{10}$$

where I_{RSFA} stands for the output of RSFA.

3.4 Loss function

On the basis of the state-of-the-art methods, l_1 the loss function is utilized as the loss function of DSRDN. The expression of l_1 loss function is defined in Equation 11:

$$L(\Theta) = \frac{1}{N}\sum_{I=1}^{N} \| I_i^{SR} - I_i^{HR} \|_1 \tag{11}$$

where Θ denotes the learnable parameters and $\|.\|1$ refers to l_1 norm and N stands for the number of training samples. What's more, I_i^{SR} and I_i^{HR} represent the reconstructed images applying DPLKA and the corresponding ground-truth images, respectively.

4 Experiments

4.1 Preparation of experiments and dataset

In this study, MDSRDN is trained by the publicly available underwater image datasets USR-248 and UFO-120. With the purpose of demonstrating the generalization of MDSRDN, Set5, Set14, and urban100 are also used for testing. The USR-248 dataset comprises 1060 pairs of samples for training and 248 pairs of samples for testing. Additionally, bicubic interpolation with 20% Gaussian noise is utilized as the down-sampling method $\times$ 2, $\times$ 4, and $\times$ 8scale factors can be obtained in USR-248. Whereas the UFO-120 dataset contains 1500 paired images for training and 120 images for testing. Aiming at maintaining consistency with other approaches, the peak signal-to-noise ratio (PSNR), structure similarity index (SSIM), and underwater image quality measure (UIQM) are designated as evaluation criteria. Particularly, the value of UIQM can be acquired by Equation 12.

$$I_{UIQM} = c_1 \times I_{UICM} + c_2 \times I_{UISM} + c_3 \times I_{UIConM}, \tag{12}$$

where I_{UICM} denotes coloration, I_{UISM} represents sharpness and I_{UIConM}indicates contrast. Furthermore, c_1, c_2, and $c_3$are fixed constants which are configured as 0.0282, 0.2953, and 3.5753.

As is shown in Table 2, NVIDIA GeForce RTX-4090 and Intel i9-13900k are used as the hardware platform of our experiments. Subsequently, pytorch-3.8.0 is the underlying framework for the whole network. Adam optimizer is applied in MDSRDN to

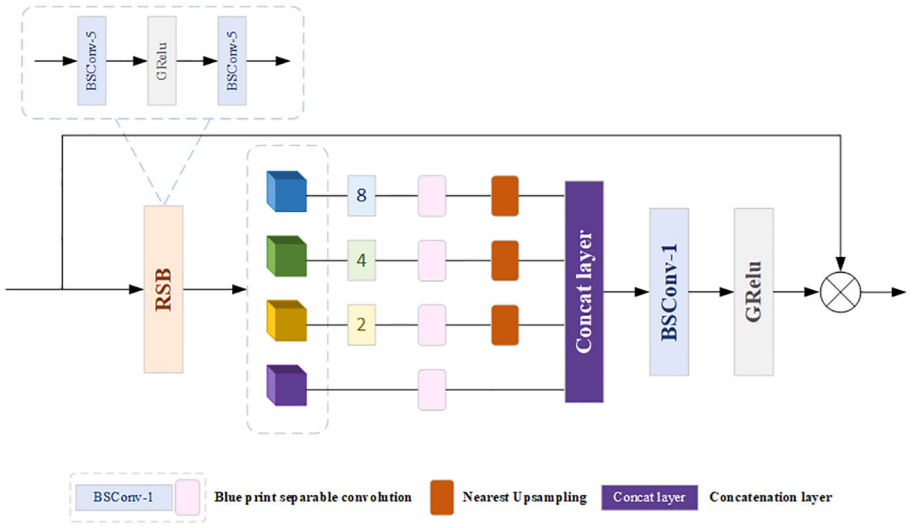

FIGURE 3
The main structure of RSFA.

minimize the object function. $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 1 \times 10^{-8}$ are the value of the active parameter in the optimizer. The training epoch is served as the 800, the batch size of the training dataset is 16 and the input patch size is 192 × 192. What's more, the initial rate is installed $2 \times e^{-3}$ and attenuates to half of its original every 200 epochs.

4.2 SR results of underwater images

4.2.1 Comparison of the USR-248 dataset on quality

Various methods, including SRCNN, VDSR, DSRCANN Dong et al., 2018, SRGAN Ledig et al., 2017, ESRGAN, SRDRM-GAN Islam et al., 2020, HNCT Fang et al., 2022, RDLN Chen et al., 2023, HAN Niu et al., 2020, SAN Liu R. et al., 2020, PAL Chen et al., 2020, AMPCNet Zhang Y. et al., 2022, PFIN Wang et al., 2022, IPT Chen et al., 2021, ELAN Zhang X. et al., 2022 are utilized for SR task comparison with MDSRDN. Table 3 demonstrates the outcomes of these methods. The results with the best performance are highlighted in bold, while those with second-best performance are italicized and displayed in blue.

As demonstrated in Table 3, MDSRDN exhibits significant advantages across all scales. When compared to other methods with the same scale, MDSRDN surpasses them all in terms of PSNR, SSIM, and UIQM. MDSRDN improves the PSNR, SSIM, and UIQM by about 8.9%, 7.7%, and 2.5% respectively compared with SRGAN. Subsequently, though MDSRDN achieves the second-best result × 4, 0.7% lower than AMPCNet on UIQM, a 5.1% enhancement on PSNR and a 7.5% elevation on SSIM can justify the excellence of MDSRDN. What's more, compared with some large deep networks (LDN) such as PAL and HAN, MDSRDN does not achieve the best results on SSIM and UIQM with the scale of × 8. Nevertheless, MDSRDN still obtains the best performance on PSNR. What's more, the parameters and flops of PAL and HAN determined that they could not be deployed on underwater edge devices.

TABLE 2 The initial parameters and hardware-software designment.

Mame	Numerical value
Hardware-platform-GPU	NVIDIA GeForce RTX-4090
Hardware-platform-CPU	Intel i9-13900k
Software-platform	Pytorch-3.8.0
Batch-size	16
Train-epoch	800
Patch-size	192 × 192
Initial learning-rate	$2 \times e^{-3}$
Optimizer	Adam

4.2.2 Comparison of computational cost on the USR-248 dataset

As shown in Table 3, the flops and parameters of our module exhibit competitiveness across all methods. We calculate the output resolution as 720p (i.e., 1080 × 720). As is widely known, all methods are competitive state-of-the-art lightweight methods. Figure 4 visualizes the relationship between PSNR, flops, and parameters. From Figure 4 and Table 3, our MDSRDN can achieve a great balance between computational cost and performance. MDSRDN decreases parameters by 65%, 24%, and 54% with the scale ×2 compared with RDLN, HNCT, and PAL. In addition, it demonstrates exceptional performance on manifestations, resulting in improvements of 0.61, 1.25, and 2.16 dB on PSNR ×2. With the scale ×4, our MRSRDN also achieves good grades. The PSNR is improved by 0.26 dB and 0.24 dB, and the number of parameters was reduced by 61% and 62% at the same time compared with EALN and RDLN, which were proposed in

TABLE 3 Some norms on the USR-248 dataset with the scale of ×2, ×4, and ×8.

Scale	Method	FLOPS(G)	Params(M)	PSNR (dB)	SSIM	UIQM
×2	SRCNN (2014)	21.30	0.06	26.81	0.76	2.74
	VDSR (2016)	205.28	0.67	28.98	0.79	2.57
	SRGAN (2017)	377.76	5.95	28.05	0.78	2.74
	ESRGAN (2017)	4274.68	16.70	26.66	0.75	2.70
	DSRCNN (2018)	54.22	1.11	27.14	0.77	2.71
	SRDRM-GAN (2020)	289.38	11.31	28.55	0.81	2.77
	SAN (2020)	1204.48	15.71	29.48	*0.80*	2.65
	PAL (2020)	203.82	0.83	28.41	*0.80*	–
	HAN (2020)	1216.57	15.92	28.67	0.79	2.53
	IPT (2021)	–	11.3	29.33	0.80	–
	HNCT (2022)	–	0.38	29.32	0.82	2.66
	PFIN (2022)	76.83	1.32	29.94	**0.83**	*2.80*
	AMPCNet (2022)	–	1.15	29.54	*0.80*	2.77
	ELAN (2022)	153	0.818	*30.14*	**0.83**	–
	RDLN (2023)	–	0.84	29.96	**0.83**	2.68
	MDSRDN (ours)	103.4	0.29	**30.57**	**0.83**	**2.81**
×4	SRCNN (2014)	21.30	0.06	23.38	0.67	2.38
	VDSR (2016)	205.28	0.67	25.70	0.68	2.44
	SRGAN (2017)	529.86	5.95	24.76	0.69	2.42
	ESRGAN (2017)	1504.09	16.70	23.79	0.66	2.38
	DSRCNN (2018)	15.77	1.11	23.61	0.67	2.36
	SRDRM-GAN (2020)	377.20	12.38	24.62	0.69	2.48
	SAN (2020)	312.86	15.86	26.00	0.65	2.40
	PAL (2020)	303.42	1.92	24.89	0.69	–
	HAN (2020)	315.88	16.07	25.26	0.59	*2.56*
	IPT (2021)	–	11.4	25.82	0.69	-
	HNCT (2022)	–	0.78	26.06	0.66	2.41
	PFIN (2022)	19.65	1.34	*26.25*	*0.70*	2.53
	AMPCNet (2022)	–	1.17	25.90	0.66	**2.58**
	ELAN (2022)	150	0.82	26.12	0.70	-
	RDLN (2023)	–	0.84	26.16	0.69	2.48
	MDSRDN (ours)	25.95	0.32	**26.38**	**0.71**	*2.56*
×8	SRCNN (2014)	21.30	0.06	19.97	0.57	2.01
	VDSR (2016)	205.28	0.67	23.58	0.63	2.17
	SRGAN (2017)	567.88	5.95	20.14	0.60	2.10
	ESRGAN (2017)	811.44	16.70	19.75	0.58	2.05
	DSRCNN (2018)	6.15	1.11	20.14	0.56	2.04
	SRDRM-GAN (2020)	399.15	13.45	20.25	0.61	2.17
	SAN (2020)	89.96	16.01	23.78	0.53	2.19

(Continued)

TABLE 3 Continued

Scale	Method	FLOPS(G)	Params(M)	PSNR (dB)	SSIM	UIQM
	PAL (2020)	325.51	2.99	22.51	**0.63**	–
	HAN (2020)	90.71	16.22	23.17	0.48	**2.47**
	IPT (2021)	–	–	22.87	0.58	-
	HNCT (2022)	–	0.86	23.88	0.54	2.21
	PFIN (2022)	5.36	1.44	23.96	0.55	2.27
	AMPCNet (2022)	–	1.25	23.83	*0.62*	2.25
	ELAN (2022)	–	–	23.76	*0.62*	–
	RDLN (2023)	–	0.84	*23.91*	0.54	2.18
	MDSRDN (ours)	7.95	0.34	**24.16**	*0.62*	*2.31*

The results with the best performance are highlighted in bold, while those with second-best performance are italicized and displayed in blue.

2022 and 2023. On ×8, an excellent performance is achieved by MDSRDN. However, on UIQM and SSIM, the MDSRDN achieved the second-best result. However, the parameter and flops quantities of SAN and PAL make them unsuitable for lightweight purposes. When compared to the lightweight networks AMPCNet and RDLN, which were introduced in 2022 and 2023, MDSRDN shows an increase of 0.33 dB and 0.25 dB in PSNR and a reduction of 73% and 59.5% in the number of parameters. Consequently, this is a noteworthy accomplishment. To better validate the deployment ability of our method, running-time tests are performed on the test dataset of USR-248. Experimental results exhibit that the running time on the experimental platform is only 120ms, 89ms, and 45ms. The parameters, flops, and running time indicate that our method is lightweight enough to deploy on underwater edge devices for underwater observation tasks.

4.2.3 Comparison of the USR-248 dataset on quantity

The visualized SR image is presented in Figure 5. Figure 5A represents SR images with the scale of x2. SR images with the scale of x4are expressed in Figures 5B, C is used to exhibit the SR images with the scale of x8. Figure 5 demonstrates some of the SR results of

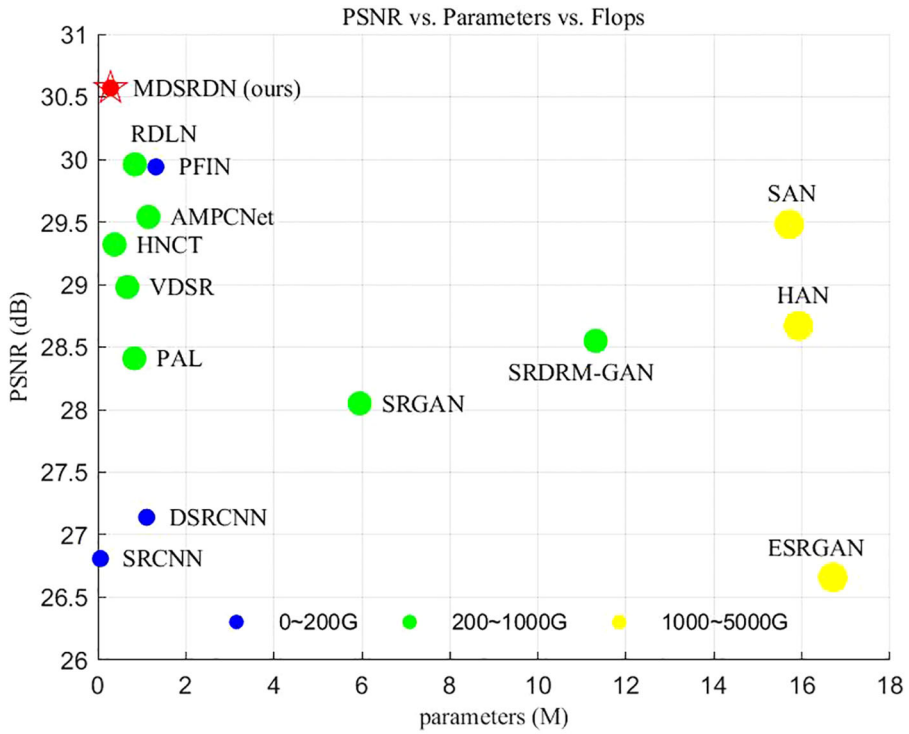

FIGURE 4
Comparison with state-of-the-art models on PSNR, parameters, and flops.

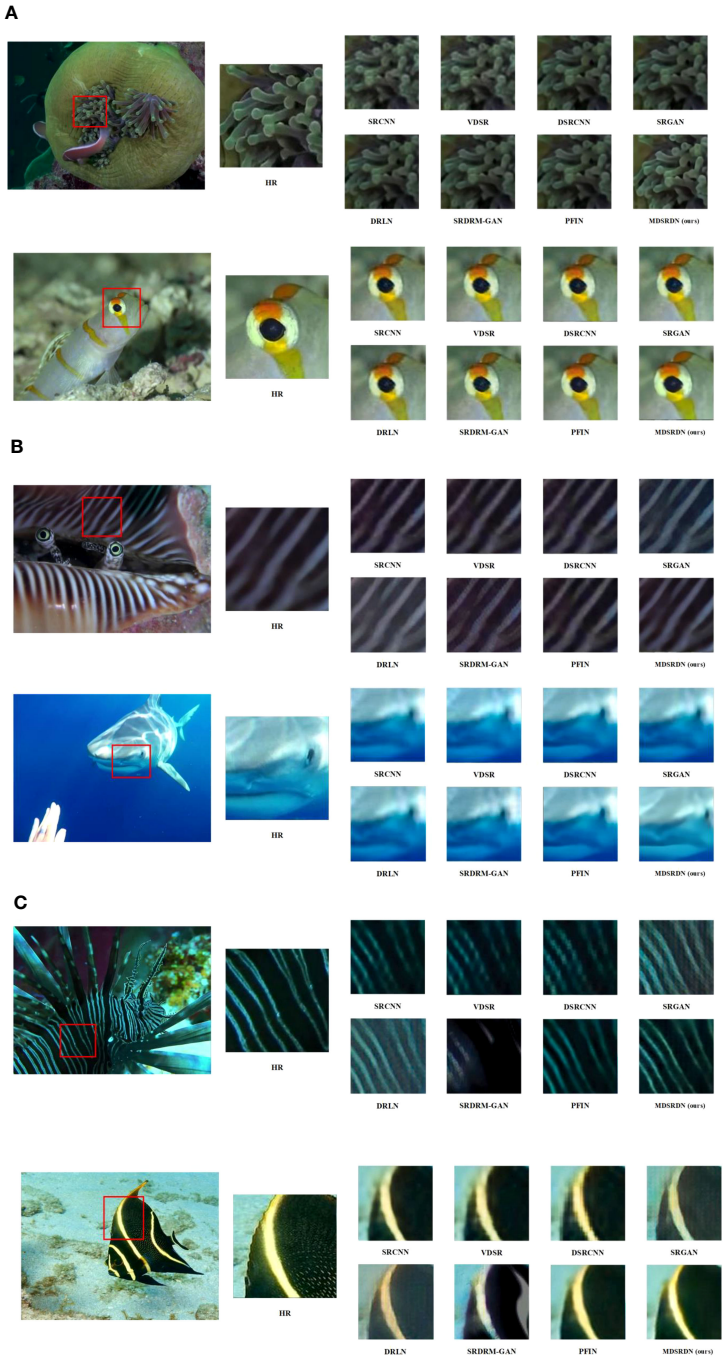

FIGURE 5
(A) Visual comparison between different methods on USR-248, with a scale factor of ×2. **(B)** Visual comparison between different methods on USR-248, with a scale factor of ×4. **(C)** Visual comparison between different methods on USR-248, with a scale factor of ×8.

underwater images, with the scale factor of x2, x4, x8. It is obvious that our MDSRDN achieves more detailed features and high-frequency information. For example, in the second image of Figure 5B, the color of the head of the shark is more similar to the HR image and the image recovery results are also clearer. In the first image of Figure 5B, more clearly detailed textures are represented, and the performance of contrast and other aspects of colors are superior. Then, it is obvious that the first image in Figure 5A and the first image in Figure 5C, the section we have boxed contains a multitude of detailed texture information. Compared with other methods, our MDSRDN outperforms in expressing detailed texture features. It is of note that the SR images reconstructed by VDSR, SRGAN, and DRLN et al. have significant edge artifacts, which lead to the SR images having blurred edges and poor reconstruction. Furthermore, with the increase of the scale, the artifacts become more pronounced such as the second image of Figure 5C, in which, the SR images of the yellow region have severe edge and border artifacts restored by

VDSR, DRLN, and SRDRAM-GAN. While clearer boundary information and fewer artifacts are demonstrated on MDSRDN. What's more, MDSRDN's recovery of sharpness and contrast is clearly superior compared to other methods. The second image of Figure 5C indicates that for the yellow region with uneven sharpness and relatively low contrast, our method achieves the best effect. Therefore, MDSRDN has advantages on edge texture features, detailed textures and colors, artifacts, and SR quality.

4.2.4 Comparison with the UFO120 dataset

More experiments are conducted on the UFO120 dataset. Table 4 expresses the consequences of different methods. SRCNN, SRGAN, SRDRM-GAN, AMPCNet, Deep WaveNet Sharma et al., 2023, SRERM Islam et al., 2020, SRResNet Lin et al., 2017, RDLN, URSCT (Ren et al., 2022), and IPT are applied to compare with MDSRDN. Similar to Table 3, the best results are bolded in the box, and the second-best results are expressed in blue italics.

As is shown in Table 4, the proposed MDSRDN achieves the best or the second-best results in various indicators. For the scale factor $\times 2$, there is a slight underperformance against URSCT on SSIM. However, the PSNR improves by 0.08 dB and the UIQM is salable at the same time. In particular, the performance of DRLN is the most comparable method to MDSRDN especially on PSNR. While the MDSRDN and RDLN share similarities in metrics such as PSNR and SSIM, the former has only 40% of the parameters of the latter. Furthermore, MDSRDN performs better in terms of PSNR and SSIM metrics with the scale of ×2 and ×4. Additionally, Deep WaveNet is also a favorable competitor. However, it has a better performance on SSIM and a proximate degree on UIQM, 1.2%, 4.6%, and 1.3% decrease on PSNR with the scale of ×2, ×3 and ×4 signifies the loss of competition.

Figures 6 and 7 demonstrate the visualized SR image with the scale of ×2 and ×4. Figure 6 further exhibits that our method has advantages in the recovery of texture features, especially for the detailed textures of fish heads. In comparison to other techniques, our method produces a much clearer reconstruction of the white texture of the head, without any edge artifacts. In Figure 7, it is obvious that the MDSRDN performs well on color deviation and clarity of SR images. In addition, the unpleasant artifacts are removed in MDSRDN, while it is nasty on Deep WaveNet. In contrast, the stronger color correction ability and the superior capacity of reconstructing texture features allow the MDSRDN more competent for the SR task of underwater images.

Running-time is an important norm to express the lightweight of our method. By using the test dataset of UFO120, the running time of our method is 79ms, 50ms, and 16ms with the scale of ×2, ×4, and ×8.

4.3 Ablation study

In this section, several ablation experiments have been conducted. To demonstrate the impact of different blocks, several ablation experiments such as the effect without the SFT layer and the effect of RSFA.

4.3.1 The effect of the SFT layer

The SFT layer is applied to enhance the ability to capture spatial features and obtain deep spatial features. To showcase its importance, we removed the SFT layer from the deep feature extraction module. Therefore, the final construction of the ablation model is depicted in Figure 8 and the results of ablation experiments are showcased in Table 5. It is worth mentioning that, to maintain the parameters and flops, we added an MDSRD module. Even though, from Table 5, with the scale of ×2 and ×4, the SFT layer still achieves the best performance. It is obvious that in PSNR, a 0.33 dB and a 0.3 dB improvement can be obtained on the USR-248 dataset and a promotion of 0.18 dB and 0.12 dB can be achieved, respectively. With the application of the SFT layer, more spatial features can be achieved which is beneficial for the

TABLE 4 Some norms on the UFO120 dataset with the scale of ×2, ×3, and ×4.

Method	PSNR (dB)			SSIM			UIQM		
	× 2	× 3	× 4	× 2	× 3	× 4	× 2	× 3	× 4
SRCNN (2014)	24.75	22.22	19.05	0.72	0.65	0.56	2.39	2.24	2.02
SRGAN (2017)	26.11	23.87	21.08	0.75	0.70	0.58	2.44	2.39	2.56
SRResNet (2017)	25.23	23.85	19.13	0.74	0.68	0.56	2.42	2.18	2.09
SRDRM (2020)	24.62	–	23.15	0.72	–	0.67	2.59	–	2.57
SRDRM-GAN (2020)	24.61	–	23.26	0.72	–	0.67	2.59	–	2.55
IPT (2021)	25.68	25.16	23.00	0.74	0.71	0.71	2.68	2.64	2.67
AMPCNet (2022)	25.24	25.73	24.70	0.71	0.70	0.70	2.98	*2.96*	2.85
URSCT (2022)	*25.96*	–	23.59	**0.80**	–	0.66	–	–	–
Deep WaveNet (2023)	25.71	25.23	25.08	0.77	**0.76**	**0.74**	**2.99**	*2.96*	**2.97**
RDLN (2023)	*25.96*	**26.55**	*25.37*	0.76	0.74	*0.73*	2.98	**2.98**	2.94
MDSRDN (ours)	**26.04**	*26.41*	**25.53**	0.78	*0.75*	*0.73*	**2.99**	**2.98**	*2.96*

The results with the best performance are highlighted in bold, while those with second-best performance are italicized and displayed in blue.

FIGURE 6
Visual comparison between different methods on UFO120, with a scale factor of ×2.

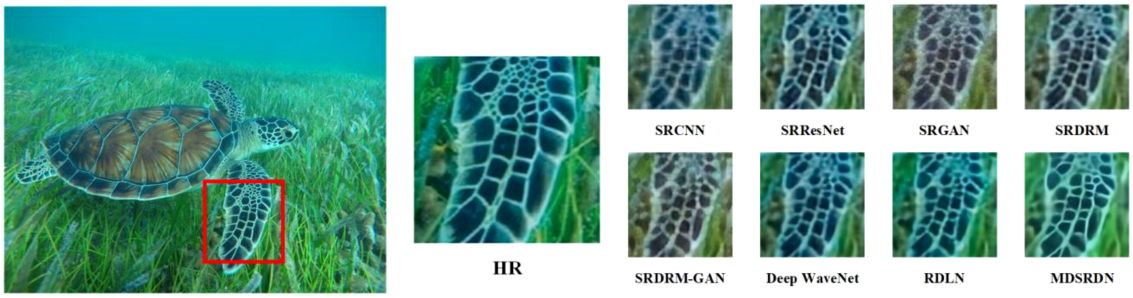

FIGURE 7
Visual comparison between different methods on UFO120, with a scale factor of x4.

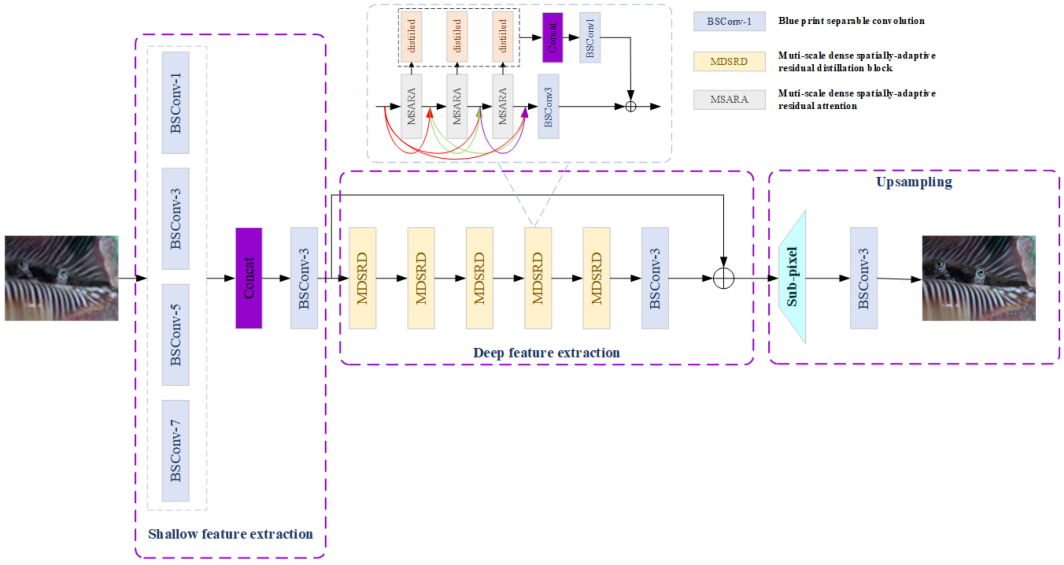

FIGURE 8
The structure of the ablation model without the SFT layer.

TABLE 5 The effect of the SFT layer.

Method	USR-248			UFO120		
	PSNR	SSIM	UIQM	PSNR	SSIM	UIQM
SFT layer ×2	30.57	0.83	2.81	26.04	0.78	2.99
w/o SFT layer ×2	30.24	0.82	2.80	25.86	0.75	2.97
SFT layer ×4	26.38	0.71	2.56	25.53	0.73	2.96
w/o SFT layer ×4	26.16	0.69	2.53	25.41	0.71	2.94

TABLE 6 The effect of RSFA.

Method	USR-248			UFO120		
	PSNR	SSIM	UIQM	PSNR	SSIM	UIQM
RSFA ×2	30.57	0.83	2.81	26.04	0.78	2.99
w/o ESA ×2	30.14	0.80	2.80	25.76	0.77	2.96
w/o CA ×2	29.97	0.80	2.79	25.69	0.75	2.95
w/o SAFM ×2	30.01	0.78	2.80	25.71	0.76	2.97

acquisition of spatial attention mechanism maps and adaptive spatial response.

4.3.2 The effect of RSFA

We design RSFA to generate an enhancement attention map for spatially adaptive feature extraction, obtain long-range dependence, and realize global feature extraction. To evaluate the effectiveness of RSFA, various attention modules including ESA, channel attention (CA), and SAFM are employed. As is exhibited in Table 6, RSFA performs excellent than all of them. It is observed that RSFA promotes the PSNR in the range of 0.43, 0.6, and 0.56 dB in USR-248. Furthermore, the parameters and flops in comparison to ESA and CA are relatively minor. Subsequently, on UFO120, with the promotion of 0.28, 0.35, and 0.33 dB, our RSFA gets excellent results. Consequently, with the ability to capture global features and achieve long-range dependence, spatial attention maps can optimize and acquire richer features adaptively.

5 Conclusion

This paper presents the MDSRDN, a multi-scale dense spatially-adaptive residual distillation network (MDSRDN) for SR tasks of underwater images. The idea of MDSRDN is realizing end-to-end feature mapping without prior knowledge which is suitable for different scenarios. By constructing MDSRD, multi-scale global-to-local feature extraction like multi-head transformer can be realized, an attention map for a spatially-adaptive feature can be generated, and the feature can be reused maximally. RSFA is a crucial aspect of MDSRD as it successfully enhances spatial features and acquires long-range dependence in an adaptive manner. Furthermore, with the application of multi-scale shallow feature extraction and the introduction of the SFT layer, the attention maps on spatial-dimension compose more texture and edge features, which can be selected adaptively by RSFA. The model also maintains controlled flops and parameters to enable efficient deployment on underwater edge devices. Comprehensive experimental results on USR-248 and UFO120 indicate that the proposed method can achieve realistic colors, abundant detail features, and clear texture features. Thus, MDSRDN attains a remarkable balance between performance and computational costs while having excellent generalizing prowess. Consequently, MDSRDN has application value for underwater image super-resolution reconstruction and ocean observation, which can be deployed on underwater edge devices and sensors.

In forthcoming endeavors, our aim is to introduce deformable convolution to RSFA with the objective of enhancing the capacity to present spatial features.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://drive.google.com/drive/folders/1dCe5rlw3UpzBs25UMXek1JL0wBBa697Q; https://www.v7labs.com/open-datasets/ufo-120. Further inquiries can be directed to the corresponding authors.

Author contributions

BL: Conceptualization, Methodology, Visualization, Writing – original draft. XN: Conceptualization, Resources, Writing – original

draft. SM: Methodology, Supervision, Writing – review & editing. YY: Software, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This project was sponsored by the Open Research Fund of CAS Key Laboratory of Space Precision Measurement Technology (Grant no. SPMT-2022–06).

Acknowledgments

The authors are very grateful to the reviewers for their valuable comments and suggestions.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alenezi, F., Armghan, A., and Santosh, K. C. (2022). Underwater image dehazing using global color features. *Eng. Appl. Artif. Intell.* 116 (October), 105489. doi: 10.1016/j.engappai.2022.105489

Barbedo, J. G. A. (2022). A review on the use of computer vision and artificial intelligence for fish recognition, monitoring, and management. *Fishes* 7 (6), 335. doi: 10.3390/fishes7060335

Chen, L. B., Chen, Y. Z., Wang, X. C., Zou, P., and Hu, X. M. (2019). Underwater image super-resolution reconstruction method based on deep learning. *J. Comput. Appl.* 39, 2738–2743. doi: 10.1109/access.2019.3004141

Chen, Z., Liu, C., Zhang, K., Chen, Y., Wang, R., and Shi, X. (2023). Underwater-image super-resolution via range-dependency learning of multiscale features. *Comput. Electrical Eng.* 110 (April), 108756. doi: 10.1016/j.compeleceng.2023.108756

Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., et al. (2021). "Pre-trained image processing transformer," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (Nashville, TN, USA: IEEE), 12294–12305. doi: 10.1109/CVPR46437.2021.01212

Chen, X., Wei, S., Yi, C., Quan, L., and Lu, C. (2020). "Progressive attentional learning for underwater image super-resolution," in *Proceedings of the International Conference on Intelligent Robotics and Applications* (Kuala Lumpur, Malaysia: Springer cham)), 12595, 233–243. doi: 10.1007/978-3-030-66645-3_20

Cheng, N., Zhao, T., Chen, Z., and Fu, X. (2018). "Enhancement of Underwater images by super-resolution generative adversarial networks," in *Proceedings of the 10th International Conference on Internet Multimedia Computing and Service*. (New York, NY, USA: ACM) 1–4. doi: 10.1145/3240876.3240881

Czub, M., Kotwicki, L., Lang, T., Sanderson, H., Klusek, Z., Grabowski, M., et al. (2018). Deep sea habitats in the chemical warfare dumping areas of the Baltic Sea. *Sci. Total Environ.* , 616–617, 1485–1497. doi: 10.1016/j.scitotenv.2017.10.165

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., et al. (2017). "Deformable convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*. (Venice, Italy: IEEE), 764–773. doi: 10.1109/ICCV.2017.89

Dong, L. F., Gan, Y. Z., Mao, X. L., Yang, Y.-B., and Shen, C. (2018). Learning deep representations using convolutional auto-encoders with symmetric skip connections. *ICASSP IEEE Int. Conf. Acoustics Speech Signal Process. - Proc.* 020214380026), 3006–3010. doi: 10.1109/ICASSP.2018.8462085

Dong, C., Loy, C. C., He, K., and Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (2), 295–307. doi: 10.1109/TPAMI.2015.2439281

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2021). "an image is worth 16X16 words: transformers for image recognition at scale," in *ICLR 2021 - 9th International Conference on Learning Representations*. (Vienna, Austria: OpenReview.net), 16. doi: 10.48550/arXiv.2010.11929

Ei, X. I. M., Iufen, X. Y. E., Ang, J. U. W., Ang, X. U. L. I. W., Anjie, H., Uang, H., et al. (2023). UIEOGP : an underwater image enhancement method based on optical geometric properties. *Optic Express* 31 (22), 36638–36655. doi: 10.1109/10.1364/oe.499684

Fang, J., Lin, H., Chen, X., and Zeng, K. (2022). "A hybrid network of CNN and transformer for lightweight image super-resolution," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. (New Orleans, LA, USA: IEEE), 1102–1111. doi: 10.1109/CVPRW56347.2022.00119

Golts, A., Freedman, D., and Elad, M. (2020). Unsupervised single image dehazing using dark channel prior loss. *IEEE Trans. Image Process.* 29, 2692–2701. doi: 10.1109/TIP.2019.2952032

Gregor, K., Danihelka, I., Graves, A., Rezende, D. J., and Wierstra, D. (2015). "DRAW: A recurrent neural network for image generation," in *32nd International Conference on Machine Learning, ICML 2015*. (Lille, France: ACM), Vol. 2. 1462–1471. doi: 10.48550/arXiv.1502.04623

Guo, M. H., Cai, J. X., Liu, Z. N., Mu, T. J., Martin, R. R., and Hu, S. M. (2021). PCT: Point cloud transformer. *Comput. Visual Media* 7 (2), 187–199. doi: 10.1007/s41095-021-0229-5

Guo, M. H., Xu, T. X., Liu, J. J., Liu, Z. N., Jiang, P. T., Mu, T. J., et al. (2022). Attention mechanisms in computer vision: A survey. *Comput. Visual Media* 8 (3), 331–368. doi: 10.1007/s41095-022-0271-y

Haase, D., and Amthor, M. (2020). "Rethinking depthwise separable convolutions: How intra-kernel correlations lead to improved mobilenets," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (Seattle, WA, USA: IEEE), 14588–14597. doi: 10.1109/CVPR42600.2020.01461

Helwig, N. E., Hong, S., and Hsiao-wecksler, E. T. (2023). Underwater image reconstruction method based on improved residual network. *Comput. Sci.* 6, 1671–1698. doi: 10.16526/j.cnki.11-4672/tp.2023.06.029

Hu, J., Shen, L., Albanie, S., Sun, G., and Vedaldi, A. (2018). Gather-excite: Exploiting feature context in convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 11, 9401–9411. doi: 10.5555/3327546.3327612

Hui, Z., Gao, X., Yang, Y., and Wang, X. (2019). "Lightweight image super-resolution with information multi-distillation network," in *Proceedings of the 27th ACM International Conference on Multimedia*. (Nice, France: ACM), 2024–2032. doi: 10.1145/3343031.3351084

Hui, Z., Wang, X., and Gao, X. (2018). "Fast and accurate single image super-resolution via information distillation network," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (Salt Lake City, UT, USA: IEEE), 723–731. doi: 10.1109/CVPR.2018.00082

Islam, M. J., Ho, M., and Sattar, J. (2019). Understanding human motion and gestures for underwater human–robot collaboration. *J. Field Robotics* 36 (5), 851–873. doi: 10.1002/rob.21837

Islam, M. J., Sakib Enan, S., Luo, P., and Sattar, J. (2020). "Underwater image super-resolution using deep residual multipliers," in *Proceedings - IEEE International Conference on Robotics and Automation*. (Paris, France: IEEE), 900–906. doi: 10.1109/ICRA40945.2020.9197213

Jaderberg, M., Simonyan, K., Zisserman, A., and Kavukcuoglu, K. (2015). Spatial transformer networks. *Adv. Neural Inf. Process. Syst.* 2015-Janua, 2017–2025.

Kim, J., Lee, J. K., and Lee, K. M. (2016a). "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV, USA: IEEE) 2016 (12), 1646–1654. doi: 10.1109/CVPR.2016.182

Kim, J., Lee, J. K., and Lee, K. M. (2016b). "Deeply-recursive convolutional network for image super-resolution," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV, USA: IEEE) 26 (12), 1637–1645. doi: 10.1109/CVPR.2016.181

Lai, Y., Zhou, Z., Su, B., Xuanyuan, Z., Tang, J., Yan, J., et al. (2022). Single underwater image enhancement based on differential attenuation compensation. *Front. Mar. Sci.* 9 (November). doi: 10.3389/fmars.2022.1047053

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., et al. (2017). "Photo-Realistic single image super-Resolution using a generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Honolulu, HI, USA: IEEE) Vol. 2. 4681–4690. doi: 10.1109/cvpr.2017.19

Li, Z., Liu, Y., Chen, X., Cai, H., Gu, J., Qiao, Y., et al. (2022). "Blueprint separable residual network for efficient image super-resolution," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. (New Orleans, LA, USA: IEEE), Vol. 2022 (6), 832–842. doi: 10.1109/CVPRW56347.2022.00099

Li, Y., Yang, D., Liu, L., and Wang, Y. (2022). Underwater image enhancement based on generative adversarial networks. *J. Mar. Sci. Eng.* 56 (2), 134–142. doi: 10.16183/j.cnki.jsjtu.2021.075

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., and Timofte, R. (2021). "SwinIR: image restoration using swin transformer," in *Proceedings of the IEEE International Conference on Computer Vision*. (Montreal, BC, Canada: IEEE), 1833–1844. doi: 10.1109/ICCVW54120.2021.00210

Lim, B., Son, S., Kim, H., Nah, S., and Lee, K. M. (2017). "Enhanced deep residual networks for single image super-resolution," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. (Honolulu, HI, USA: IEEE), 1132–1140. doi: 10.1109/CVPRW.2017.151

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). "Feature pyramid networks for object detection," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. (Honolulu, HI, USA: IEEE), 936–944. doi: 10.1109/CVPR.2017.106

Liu, R., Su, Z., Lin, G., and Zhou, F. (2020). "Second-order attention network for magnification-arbitrary single image super-resolution," in *Proceedings of the 8th International Conference on Digital Home*. (Dalian, China: IEEE), 127–134. doi: 10.1109/ICDH51081.2020.00030

Liu, J., Zhang, W., Tang, Y., Tang, J., and Wu, G. (2020). "Residual feature aggregation network for image super-resolution," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (Seattle, WA, USA: IEEE), Vol. 1, 2356–2365. doi: 10.1109/CVPR42600.2020.00243

Mnih, V., Heess, N., Graves, A., and Kavukcuoglu, K. (2014). Recurrent models of visual attention. *Adv. Neural Inf. Process. Syst.* 3 (January), 2204–2212. doi: 10.1016/j.cviu.2019.05.001

Mooney, J. G., and Johnson, E. N. (2014). A comparison of automatic nap-of-the-earth guidance strategies for helicopters. *J. Field Robotics* 27 (6), 1–17. doi: 10.1002/rob

Niu, B., Wen, W., Ren, W., Zhang, X., Yang, L., Wang, S., et al. (2020). "Single image super-resolution via a holistic attention network," in *European Conference on Computer Vision*. (Glasgow, UK: Springer Cham), Vol. 12350. 191–207. doi: 10.1007/978-3-030-58610-2_47

Ren, T., Xu, H., Jiang, G. Y. M., Zhang, X., Wang, B., and Luo, T.. (2022). Reinforced Swin-Convs Transformer for Simultaneous Underwater Sensing Scene Image Enhancement and Super-resolution. *IEEE T. Geosci. Remote.* 60, 1–16. doi: 10.1007/978-3-030-58610-2_47

Sharma, P., Bisht, I., and Sur, A. (2023). Wavelength-based attributed deep neural network for underwater image restoration. *ACM Trans. Multimedia Comput. Commun. Appl.* 19 (1), 1–23. doi: 10.1145/3511021

Sun, L., Dong, J., Tang, J., and Pan, J. (2023) *Spatially-adaptive feature modulation for efficient image Super-Resolution*. Available at: http://arxiv.org/abs/2302.13800.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (Boston, MA, USA: IEEE), 1–9. doi: 10.1109/CVPR.2015.7298594

Tai, Y., Yang, J., and Liu, X. (2017). "Image super-resolution via deep recursive residual network," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition*. (Honolulu, HI, USA: IEEE), 2790–2798. doi: 10.1109/CVPR.2017.298

Talab, M. A., Awang, S., and Najim, S. A. D. M. (2019). "Super-Low Resolution Face recognition using integrated efficient sub-pixel convolutional neural network (ESPCN) and convolutional neural network (CNN)," in *Proceedings of the 2019 IEEE International Conference on Automatic Control and Intelligent Systems*. (Selangor, Malaysia: IEEE), 331–335. doi: 10.1109/I2CACIS.2019.8825083

Wang, J., Li, Q., Fang, Z., Zhou, X., Tang, Z., Han, Y., et al. (2023). YOLOv6-ESG: A lightweight seafood detection method. *J. Mar. Sci. Eng.* 11 (8), 1623. doi: 10.3390/jmse11081623

Wang, X., Nian, R., He, B., Zheng, B., and Lendasse, A. (2017). *Underwater image super-resolution reconstruction with local self-similarity analysis and wavelet decomposition*. (Aberdeen, Aberdeen, UK: IEEE), 1–6. doi: 10.1109/OCEANSE.2017.8084745

Wang, L., Xu, L., Tian, W., Zhang, Y., Feng, H., and Chen, Z. (2022). Underwater image super-resolution and enhancement via progressive frequency-interleaved network. *J. Visual Communication Image Representation* 86 (May), 103545. doi: 10.1016/j.jvcir.2022.103545

Wang, X., Yu, K., Dong, C., and Change Loy, C. (2018). "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 606–615. doi: 10.1109/CVPR.2018.00070

Wang, H., Zhong, G., Sun, J., Chen, Y., Zhao, Y., Li, S., et al. (2023). Simultaneous restoration and super-resolution GAN for underwater image enhancement. *Front. Mar. Sci.* 10 (June). doi: 10.3389/fmars.2023.1162295

Xu, K., Ba, J. L., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., et al. (2015). "Show, attend and tell: Neural image caption generation with visual attention," in *Proceedings of the 32nd International Conference on Machine Learning*, (Lille, France: ACM) 3, 2048–2057. doi: 10.48550/arXiv.1502.03044

Yuan, H. C., Kong, L. D., Zhang, S. S., Gao, K., and Yang, Y. Y. (2023). Underwater image super-resolution reconstruction algorithm based on information distillation mechanism. *Laser Optoelectronics Prog.* 60, 1210017. doi: 10.3788/LOP221324

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y. (2018). Image super-Resolution using very deep residual channel attention networks. *Lecture Notes Comput. Sci.* 11213, 294–310. doi: 10.1007/978-3-030-01234-2_18

Zhang, Y., Yang, S., Sun, Y., Liu, S., and Li, X. (2022). Attention-guided multi-path cross-CNN for underwater image super-resolution. *Signal Image Video Process.* 16 (1), 155–163. doi: 10.1007/s11760-021-01969-4

Zhang, X., Zeng, H., Guo, S., and Zhang, L. (2022). "Efficient long-range attention network for image super-resolution," in *Proceedings of the European Conference on Computer vision*. (Tel Aviv, Israel: Springer, Cham), 649–667. doi: 10.1007/978-3-031-19790-1_39