

Received 21 December 2023, accepted 3 January 2024, date of publication 9 January 2024,
date of current version 16 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3351730

RESEARCH ARTICLE

Underwater Image Super-Resolution Using Frequency-Domain Enhanced Attention Network

XIN LIU¹, ZHENGXIANG GU¹, HAIMING DING², MIN ZHANG³, AND LI WANG⁴

¹School of Artificial Intelligence and Information Engineering, Jinken College of Technology, Nanjing 211156, China

²College of Computer and Information Engineering, Hohai University, Nanjing 211100, China

³Department of Information Engineering, Gannan University of Science and Technology, Ganzhou 341000, China

⁴School of Computer and Software, Nanjing Vocational University of Industry Technology, Nanjing 210023, China

Corresponding author: Li Wang (li1019wang@gmail.com)

This work was supported by the Science and Technology Research Project in the Department of Education of Jiangxi Province under Grant GJJ2203610.

ABSTRACT Underwater images super-resolution (SR) is a challenging task due to underwater images usually contain severely blurred details, color distortion, and low contrast. Although numerous deep learning-based methods have been developed to solve these problems, these methods suffer from huge model parameters and computations. To address this gap, we propose a frequency-domain enhanced attention network (FEAN), supported by a series of frequency-enhanced attention modules (FEAM), for accurate underwater SR. Specifically, we start by utilizing a Gaussian filter to decompose the features into high and low frequencies and pass them to the FEAM. Then, in the high-frequency path, we propose a multi-scale attention enhancement block (MAEB) to extract rich image texture information. While in the low-frequency path, we perform a simple convolutional operation to realize the brightness and contrast adjustment of the image. Further, we devise a channel attention fusion block (CAFB) to integrate the enhanced high and low-frequency features to further strengthen the powerful representational capability of the network. Finally, we employ two convolutions to further modulate the features on the high-frequency path for effective color bias correction and detail enhancement. Experimental results show that our FEAN performs better than other underwater SR methods on the USR-248 dataset, with PSNR values of 29.97 dB, 26.23 dB, and 23.99 dB, corresponding to $\times 2$, $\times 4$, and $\times 8$ scale factors.

INDEX TERMS Underwater image, super-resolution, frequency domain, attention mechanism.

I. INTRODUCTION

Underwater image super-resolution (SR) plays a crucial role in various fields such as scientific research and resource development. However, the harsh underwater imaging environment can lead to a series of problems such as reduced contrast, color distortion, scattering effects, and insufficient brightness in underwater images. Therefore, we need to seek ways to improve the quality of these images to meet our development needs. In our research, we will focus on image SR to address these issues.

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Moinul Hossain¹.

Image SR technology utilizes the internal information and statistical characteristics of images to restore lost high-resolution (HR) details from low-resolution (LR) images. In the fields of medical [1], [2], remote sensing [3], [4], surveillance [5], [6], [7], image SR technology has widespread applications. Classical image SR methods are usually based on techniques such as signal processing, interpolation, and filtering. However, due to the excessive manual design rules, they lack the ability to learn autonomously and exhibit limited effectiveness in dealing with diverse image processing needs. Recently, in the domain of SR, convolutional neural network (CNN) [8], [9], [10], [11], [12], [13], [14], [15], residual networks (ResNet) [16], [17], [18], [19], [20], [21], and generative adversarial

network (GAN) [22], [23], [24], [25], [26] have become major research trends. Dong et al. [9] used a three-layer convolutional network to fit nonlinear maps, obtaining good performance compared to traditional SR methods. Subsequently, Lim et al. [17] adopted a residual network and removed the batch normalization layer [27] to construct a deeper network and achieve significant SR accuracy improvement. Xu et al. [20] combined the ideas of residual connections and dense connections, enabling the network to capture image features from different levels and perspectives, which helps to improve the diversity and richness of features. Jiang et al. [21] designed a hierarchical dense residual block (HDB) to reduce the complexity and number of parameters introduced by dense connections. Zhang et al. [10] constructed an attention mechanism to automatically adjust the interdependence between feature channels. Although the above methods perform well in natural image SR tasks, they do not fully consider the unique characteristics of the underwater environment, such as color casts, color artifacts, and blurring. Accordingly, a wide variety of underwater SR networks have emerged to solve these adverse effects. Islam et al. [28] presented deep simultaneous enhancement and super-resolution (Deep SESR) for underwater SR reconstruction and enhancement. Wang et al. [29] proposed a progressive frequency-interleaved network (PFIN), that used a frequency-domain progressive mechanism to learn different information. Ren et al. [30] presented a novel U-Net-based reinforced Swin-Convs Transformer (URSCT) for simultaneous enhancement and SR. Chen et al. [31] constructed a progressive attentional learning (PAL) method for the underwater SR task to learn a nonlinear mapping from LR images to HR images. Sharma et al. [32] used the traditional pixel-wise and feature-based cost function to propose a novel framework, called Deep WaveNet, improving the spatial resolution of degraded underwater images. Similarly, Chen et al. [33] established a range-dependency learning network (RDLN) to model short and long-range dependency of multiscale features.

Although some progress has been made in underwater image SR, however, existing methods still face some challenges. On the one hand, most SR models have large parameter scales and high memory overheads, making it difficult to apply these models directly to underwater robots. On the other hand, few methods have used frequency domain-based strategies to adjust the high and low-frequency information of an image, thereby discouraging accurate reconstruction of the edges and texture of an image.

To address the aforementioned limitations, we propose a lightweight and effective frequency-domain enhanced attention network (FEAN) for underwater image SR reconstruction. Our FEAN architecture consists of multiple frequency-enhanced attention modules (FEAM), which can explore high and low-frequency components of the image to generate more delicate features. To be specific, we first utilize Gaussian filtering to decompose the features into high and low-frequency branches and process them separately using

different operations. Then, we devise a multi-scale attention enhancement block (MAEB) to catch realistic textures in the high-frequency branch, while leveraging simple convolution operation to enhance the low-frequency branch. Finally, we propose a channel attention fusion block (CAFB) to fuse and modulate frequency domain information according to the importance of features. Experimental results demonstrate that our FEAN exceeds mainstream SR approaches with fewer parameters and computations.

To summarize, our contributions can be highlighted in three main aspects:

- We propose a lightweight FEAN for underwater SR tasks. Thanks to the network backbone FEAM, supported by the MAEB and CAFB, our FEAN achieves excellent performance with low resource consumption.
- MAEB captures information at different scales and recalibrates features to focus on valuable characteristics, improving the representation of high-frequency branches and reconstructing the edge and texture of underwater images.
- CAFB aggregates high and low-frequency information effectively, and assigns different weights to different levels of information, jointly regulating the impact of information in the frequency domain.

II. RELATED WORK

A. UNDERWATER IMAGE SR

With the development of deep learning, image SR technology has become a hot research topic in computer vision tasks. Unlike natural images, underwater images are subject to complex lighting conditions and environments, which tend to have undesirable effects such as color deviation and detail blurring. Therefore, to further improve image reconstruction quality, plenty of works have been developed for underwater image SR tasks. In 2020, the first publicly available dataset USR-248 [24] was introduced for underwater SR tasks. Meanwhile, researchers proposed two fully-convolutional deep residual networks, named SRDRM and SRDRM-GAN, obtaining good performance and visual effects.

Deep SESR adopted the multi-scale strategy to mitigate the local distortion problem in underwater scenes. URSCT [30] integrates Swin Transformer into the U-Net model, thus enhancing its global feature capture capability and ensuring the improvement in the overall reconstruction results. As more and more underwater datasets become available, A. Aghelan [34] employed a pre-training and migration learning strategy to fine-tune Real-ESRGAN [35] using the USR-248 and UFO-120 datasets, generating higher-resolution and better-quality underwater images. RDLN [33] reduces the number of parameters of the network by introducing a channel-splitting module in multi-scale convolution. At the same time, it makes full use of the different properties of CNN and Transformer to learn the short-range and long-range dependencies to generate high-quality high-resolution images. Although many effective design methods have been developed to solve underwater SR

tasks, existing methods neglect the interaction of frequency information, resulting in poor image reconstruction.

B. ATTENTION-BASED NETWORK

Attention mechanism is an effective optimization tool that focuses on important features to improve computational efficiency and is widely used in SR tasks. Zhang et al. [10] designed a residual channel attention block to learn the channel statistics among channels, boosting the discriminative ability of the network. Additionally, residual attention module (RAM) [36] and channel-wise and spatial feature modulation (CSFM) [37] combined channel attention and spatial attention for utilizing the interdependencies between channels and spatial features. Thanks to this advantage, the attention mechanism has also been introduced in underwater SR tasks to enhance the network reconstruction efficiency. For example, an attention-guided multipath cross-convolution neural network (AMPCNet) [38] introduced a channel attention mechanism in the field of underwater SR to assign more significant weights to high-frequency information, enhancing the edge and texture information in the image. In Deep WaveNet [32], the authors used the convolutional block attention module to regulate the flow of channel-specific information in the network adaptively. PFIN [29] devised a global spatial attention block to catch the meaningful context information by learning a group of weights for removing distorted information. Fu et al. [39] devised nonlocal attention and channel attention mechanisms respectively to mine and enhance more detailed features. These attention mechanism designs can capture valuable feature statistics, but they do not well consider feature different scale information and frequency information, which is not conducive to high-quality underwater image reconstruction. Inspired by the advantages of the attention mechanism, we design MAEB to strengthen the high-frequency features and recover more fine features further. Meanwhile, we propose a lightweight CAFB to explore the interaction of information in different frequency domains, improving the representational capability of the network.

III. PROPOSED METHOD

The overall architecture of the proposed FEAN is illustrated in Fig.1. The network can be divided into three stages: (1) Stage 1 focuses on extracting shallow features from the input image denoted as $X \in \mathbb{R}^{H \times W \times 3}$. (2) Stage 2 is specifically designed for frequency separation and enhancement. (3) Stage 3 aims to rebuild HR underwater images, and the final output is denoted as $S \in \mathbb{R}^{rH \times rW \times 3}$. Here, H and W represent the height and width of the input image, respectively. The r is the scaling factor used during the process.

Stage 1: As investigated in [40], we utilize a single convolutional layer to extract the shallow feature F_0 from the LR input X .

$$F_0 = H_{SFE}(X) \quad (1)$$

Here, $H_{SFE}(\cdot)$ denotes the shallow feature extraction convolution with a 3×3 kernel size. F_0 means the extracted shallow features, are then passed on to the next stage for further processing.

Stage 2: Non-uniform degradation in underwater scenes can seriously affect the quality of image restoration, and a model with rich feature representation capabilities is needed. Therefore, we explore different frequency domain information to enhance the texture details and color correction of the image. We use Gaussian filtering to separate the shallow features into high-frequency component F_{HH}^0 and low-frequency component F_{LL}^0 . Gaussian filtering is effective in suppressing high-frequency noise when performing signal or image smoothing, effectively improving the signal-to-noise ratio. And with the linear filtering method, Gaussian filtering is usually computationally faster. We then use K FEAMs to extract more finer features. The detailed description of FEAM will be provided in Section 3.3. This process can be represented as follows:

$$\begin{aligned} F_{SMB}^k, F_{HH}^k, F_{LL}^k \\ &= H_{FEAM}^k(F_{SMB}^{k-1}, F_{HH}^{k-1}, F_{LL}^{k-1}) \\ &= H_{FEAM}^k(H_{FEAM}^{k-1}(\cdots H_{FEAM}^1(F_0, F_{HH}^0, F_{LL}^0) \cdots)), \quad (2) \end{aligned}$$

where $F_{SMB}^k \in \mathbb{R}^{H \times W \times C}$, $F_{HH}^k \in \mathbb{R}^{H \times W \times C}$, $F_{LL}^k \in \mathbb{R}^{H \times W \times C}$ denote the output features of the k -th FEAM. After the step-by-step extraction by the FEAMs, the extracted high and low-frequency features are aggregated and smoothed using a 1×1 convolutional layer, as illustrated below:

$$F_{assemble} = F_0 + k_1 \times F_{SMB}^K + k_2 \times f_{1 \times 1}([F_{HH}^K, F_{LL}^K]), \quad (3)$$

where $[\cdot, \cdot]$ represents the concatenation operation along the channel dimension. $f_{1 \times 1}(\cdot)$ denotes a 1×1 convolutional layer. k_1 and k_2 represent the learnable scaling factors. $F_{assemble}$ refers to the aggregated features.

Stage 3: This stage is used for underwater image reconstruction, where the refined features are upsampled to the target HR image S . We refer to the technique from the paper [41] and use the method of using both H_{UP0} and H_{UP1} in the upsampling process. This approach is able to merge low-resolution image features and high and low frequency features, thus mitigating the loss of information. The process can be calculated as follows:

$$S = H_{UP0}(F_{SMB}^K) + H_{UP1}(F_{assemble}), \quad (4)$$

where both $H_{UP0}(\cdot)$ and $H_{UP1}(\cdot)$ contain a 3×3 convolution and a sub-pixel convolution.

A. MODEL LEARNING

To ensure a fair comparison with state-of-the-art competing methods, we employ the \mathcal{L}_1 loss function to preserve intricate textures and local structures. The training dataset consists of LR images and their corresponding HR images, denoted as: $\{X^{(i)}, S^{(i)}\}_{i=1}^N$, the goal of training FEAN is to optimize the

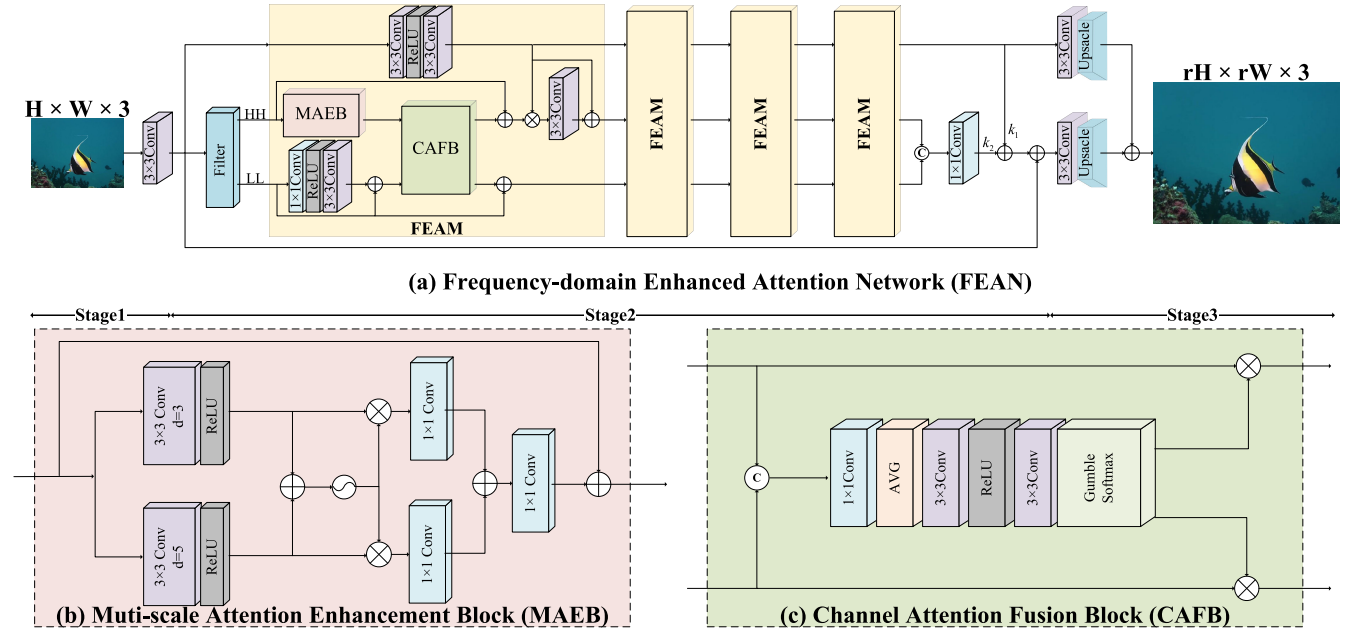


FIGURE 1. The overall pipeline of our proposed frequency-domain enhanced attention network (FEAN) is implemented by sequential stacking of FEAMs. (a) The architecture of the multi-scale attention enhancement block (MAEB). (b) The architecture of the channel attention fusion block (CAFB).

\mathcal{L}_1 loss function:

$$\mathcal{L}_1 = \frac{1}{N} \sum_{i=1}^n \|S^{(i)} - H_{FEAN}(X^{(i)})\|_1 \quad (5)$$

B. FREQUENCY ENHANCED ATTENTION MODULE (FEAM)

To achieve a higher quality of underwater images, the network needs to explore different frequency domain information of the image. In such a way, the network works well for color bias correction, contrast enhancement, and detail restoration. To this end, we propose FEAM that mainly consists of MAEB and CAFB to explore different frequency properties of underwater images. As depicted in Fig.1(a), the Gaussian filter decomposes features into high-frequency and low-frequency paths and then processes the frequency domain information in a heterogeneous manner. In the high-frequency path, we devise MAEB to further enhance the high-frequency features. While in the low-frequency path, we utilize a 1×1 convolution and a 3×3 convolution for cost-effective feature extraction. The low-frequency information of an image mainly reflects the parts of the image that change more slowly. By using a smaller convolutional kernel design, the filtering of local regions can be realized, thus reducing the high-frequency noise in the image and achieving the effect of adjusting the brightness and contrast.

Subsequently, different frequency features will be passed to the CAFB for further fusion and enhancement. More importantly, we adopt two convolutions to preserve the original spatial information, and then modulate the features in the high-frequency branch, referred to as the spatial modulation block (SMB). The process can be expressed as

follows:

$$\begin{aligned} F_{SMB}^k &= H_{SMB}^k(F_{SMB}^{k-1}) \\ F_{HH}^{k'} &= H_{MAEB}^k(F_{HH}^{k-1}) \\ F_{LL}^{k'} &= f_{3 \times 3}(f_{1 \times 1}(F_{LL}^{k-1})) \\ F_{LL}^k, F_{HH}^{k''} &= H_{CAFB}^k(F_{HH}^{k'}, F_{LL}^{k'}) \\ F_{HH}^k &= f_{3 \times 3}(F_{HH}^{k''} \otimes F_{SMB}^k), \end{aligned} \quad (6)$$

where $H_{SMB}^k(\cdot)$ represents two 3×3 convolutional operations in the k -th SMB. $H_{MAEB}^k(\cdot)$ and $H_{CAFB}^k(\cdot)$ denote the k -th MAEB module and the k -th CAFB module respectively. The $f_{3 \times 3}(\cdot)$ denotes the convolution with a 3×3 kernel size. The $f_{1 \times 1}(\cdot)$ denotes the convolution with a 1×1 kernel size. The $F_{HH}^{k'} \in \mathbb{R}^{H \times W \times C}$ represents the high-frequency features enhanced by MAEB, while the $F_{HH}^{k''} \in \mathbb{R}^{H \times W \times C}$ denotes the high-frequency features adjusted in CAFB. The $F_{LL}^{k'} \in \mathbb{R}^{H \times W \times C}$ denotes the augmented low-frequency feature. For brevity, the activation function and residual representation are omitted.

Multi-scale Attention Enhancement Block (MAEB) can distill richer texture features to help high qualitative image reconstruction in the high-frequency branch, as depicted in Fig.1(b). MAEB achieves a more comprehensive focus on high-frequency features by considering feature information at different scales to better preserve image details. Firstly, we adopt two convolutions with different dilation rates to catch two different features $F_{d=5}^{k-1} \in \mathbb{R}^{H \times W \times C}$ and $F_{d=3}^{k-1} \in \mathbb{R}^{H \times W \times C}$, expanding the learning of contextual cues. Here, we employ two 3×3 convolutional operations, corresponding to dilation factors of d equal to 3 and 5, respectively. This configuration effectively expands the receptive field without introducing additional parameters or computational

complexity. Then we fuse multi-scale features and employ the Sigmoid operation to generate distinct attention weights α . Next, the attention weights recalibrate features at different scales to further highlight information-rich regions. Finally, we integrate these features using different convolution operations, and a residual connection is introduced to deliver the flow of information. The above process can be expressed as follows:

$$\begin{aligned} \alpha &= \sigma[f_{3 \times 3}^{d=3}(F_{HH}^{k-1}) + f_{3 \times 3}^{d=5}(F_{HH}^{k-1})] = \sigma(F_{d=3}^{k-1} + F_{d=5}^{k-1}) \\ H_{MAEB}^k &= f_{1 \times 1}[f_{1 \times 1}(\alpha \cdot F_{d=3}^{k-1}) + f_{1 \times 1}(\alpha \cdot F_{d=5}^{k-1})] + F_{HH}^{k-1} \end{aligned} \quad (7)$$

Here, σ denotes the Sigmoid function. The $f_{3 \times 3}^{d=3}$ and $f_{3 \times 3}^{d=5}$ represent the expansion convolution with expansion coefficients of 3 and 5, respectively. For brevity, the residual connections are omitted.

Channel Attention Fusion Block (CAFB) aims to merge the high and low-frequency information to enhance the representation of important features and reduce the impact of noise. CAFB integrates the effects of different frequency domain information on image restoration, and focuses more effectively on the high and low frequency features of the key parts by integrating the frequency domain information in order to improve the network's ability to reconstruct the details of underwater degraded images. Initially, we concatenate frequency domain information and employ 1×1 convolution to obtain fused features. Then we perform feature statistics on each channel of feature maps by employing average pooling. Subsequently, we enhance feature statistics through two 3×3 convolutions. Next, Gumbel Softmax is applied to yield channel attention weights β . Gumbel Softmax introduces a more flexible and differentiable strategy, enabling the model to handle discrete random variables more effectively. In CAFB, this method enables the model to selectively focus on high-frequency and low-frequency features. Meanwhile, by introducing continuity in discrete selection, the model maintains differentiability, allowing for the use of backpropagation algorithms for training. Finally, the channel attention weights recalibrate the features on the high and low frequency paths respectively, resulting in the ultimate feature representation. Through this process, we effectively integrate various information, thereby elevating the feature expression capacity. The process can be expressed as follows:

$$\begin{aligned} \beta &= \bar{w}(f_{3 \times 3}(f_{3 \times 3}(\text{avg}(f_{1 \times 1}([F_{HH}^{k'}, F_{LL}^{k'}]))) \\ H_{CAFB}^k &= \beta \cdot F_{HH}^{k'} + \beta \cdot F_{LL}^{k'}, \end{aligned} \quad (8)$$

where $\bar{w}(\cdot)$ represents the function of Gumbel Softmax. The $\text{avg}(\cdot)$ represents the operation of average pooling.

IV. EXPERIMENTS

In this section, we begin by describing the datasets used in the experiments and providing implementation details for network training. Subsequently, we assess the performance

of our proposed network on two publicly available datasets. We then examine the contributions of individual components within our proposed network.

A. DATASET AND EXPERIMENTAL SETUP

In this research, we employ publicly accessible underwater image datasets for training our network, specifically the USR-248 dataset [24] and the UFO-120 dataset [28]. The USR-248 dataset is a comprehensive collection of high-resolution underwater images and their corresponding low-resolution pairs. It includes LR images of sizes 320×240 , 160×120 , and 80×60 , corresponding to downsample factors of $\times 2$, $\times 4$, and $\times 8$, while the HR images are of size 640×480 . A total of 1060 RGB image pairs are used for training and validation, along with an additional 248 test images. The UFO-120 dataset comprises 1500 paired images for training and an additional 120 images for testing. The HR images in this dataset have dimensions of 640×480 , while the LR images come in sizes of 320×240 , 213×160 , and 160×120 , corresponding to downsample factors of $\times 2$, $\times 3$, and $\times 4$, respectively. The performance of image SR was thoroughly assessed using three criteria: namely, the peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and underwater image quality measure (UIQM).

We employ the Adam optimizer to minimize the objective function in our network, with optimizer parameters $\beta_1=0.9$, $\beta_2=0.999$, and $\varepsilon = 10^{-8}$. The initial learning rate is set to $1e-3$, and it undergoes halving every 200 epochs. To accommodate memory limitations [42], each batch is composed of 32 LR patches of size 60×60 for the SR task, and the number of FEAM is set to $K=4$ and the number of hidden layers is set to $C=64$. The random seed was set to 111 throughout our experiments, facilitating result reproducibility. Our model is implemented using the PyTorch framework and executed on an NVIDIA RTX 3060 GPU.

B. EXPERIMENTAL EVALUATION ON THE UFO-248 DATASET

We compare our proposed FEAN with several SR networks, including SRCNN [9], VDSR [16], DSRCNN [43], EDSRGAN [17], SRGAN [22], ESRGAN [23], LatticeNet [44], SRDRM [24], SRDRM-GAN [24], AMPCNet [38], PAL [31], and RDLN [33]. The experimental results can be seen in Table 1. In comparison to popular SR works, our FEAN shows competitive results on all scale factors. Compared to SR methods used for natural images such as LatticeNet, our method boosts PSNR, SSIM, and UIQM by up to 2.19%, 18.5%, and 4.88%, respectively. This is because our FEAN takes into account the interaction between different frequency information, making it better used for underwater image recovery. In addition, our FEAN obtains optimal and sub-optimal performance compar to underwater SR methods. For example, our network yields gains of 0.16dB and 0.02 PSNR and SSIM at a large scale factor $\times 8$ than AMPCNet, even using fewer computations. Further, RDLN

TABLE 1. Quantitative results on the USR-248 dataset with scale factors of $\times 2$, $\times 4$, and $\times 8$ for underwater-image SR. Bold indicates the best performance. When calculating FLOPs, the input tensor sizes are 320×240 for $\times 2$, 160×120 for $\times 4$, and 80×60 for $\times 8$.

| Scale | Method | FLOPs(G) | Params(M) | PSNR(dB) | SSIM | UIQM |
|-----------------|-----------------|-----------|-----------|--------------|-------------|-------------|
| $\times 2$ | SRCNN [9] | 21.30 | 0.06 | 26.81 | 0.76 | 2.74 |
| | VDSR [16] | 205.28 | 0.67 | 28.98 | 0.79 | 2.57 |
| | DSRCNN [43] | 54.22 | 1.11 | 27.14 | 0.77 | 2.71 |
| | EDSRGAN [17] | 273.34 | 1.38 | 27.12 | 0.77 | 2.67 |
| | SRGAN [22] | 377.76 | 5.95 | 28.08 | 0.78 | 2.74 |
| | ESRGAN [23] | 4274.68 | 16.70 | 26.66 | 0.75 | 2.70 |
| | LatticeNet [44] | 56.84 | 0.76 | 29.47 | 0.80 | 2.65 |
| | SRDRM [24] | 203.91 | 0.83 | 28.36 | 0.80 | 2.78 |
| | SRDRM-GAN [24] | 289.38 | 11.31 | 28.55 | 0.81 | 2.77 |
| | AMPCNet [38] | - | 1.15 | 29.54 | 0.80 | 2.77 |
| | PAL [31] | 203.82 | 0.83 | 28.41 | 0.80 | - |
| | RDLN [33] | 74.86 | 0.84 | 29.96 | 0.83 | 2.68 |
| | FEAN(ours) | 71.06 | 0.94 | 29.97 | 0.82 | 2.76 |
| | $\times 4$ | SRCNN [9] | 21.30 | 0.06 | 23.38 | 0.67 |
| VDSR [16] | | 205.28 | 0.67 | 25.70 | 0.68 | 2.44 |
| DSRCNN [43] | | 15.77 | 1.11 | 23.61 | 0.67 | 2.36 |
| EDSRGAN [17] | | 206.42 | 1.97 | 21.65 | 0.65 | 2.40 |
| SRGAN [22] | | 529.86 | 5.95 | 24.76 | 0.69 | 2.42 |
| ESRGAN [23] | | 1504.09 | 16.70 | 23.79 | 0.66 | 2.38 |
| LatticeNet [44] | | 14.61 | 0.78 | 26.06 | 0.65 | 2.43 |
| SRDRM [24] | | 291.73 | 1.90 | 24.64 | 0.68 | 2.46 |
| SRDRM-GAN [24] | | 377.20 | 12.38 | 24.62 | 0.69 | 2.48 |
| AMPCNet [38] | | - | 1.17 | 25.90 | 0.66 | 2.58 |
| PAL [31] | | 303.42 | 1.92 | 24.89 | 0.69 | - |
| RDLN [33] | | 29.56 | 0.84 | 26.16 | 0.66 | 2.38 |
| FEAN(ours) | | 18.56 | 0.98 | 26.23 | 0.70 | 2.58 |
| $\times 8$ | | SRCNN [9] | 21.30 | 0.06 | 19.97 | 0.57 |
| | VDSR [16] | 205.28 | 0.67 | 23.58 | 0.63 | 2.17 |
| | DSRCNN [43] | 6.15 | 1.11 | 20.14 | 0.56 | 2.04 |
| | EDSRGAN [17] | 189.69 | 2.56 | 19.87 | 0.58 | 2.12 |
| | SRGAN [22] | 567.88 | 5.95 | 20.14 | 0.60 | 2.10 |
| | ESRGAN [23] | 811.44 | 16.70 | 19.75 | 0.58 | 2.05 |
| | LatticeNet [44] | 4.05 | 0.86 | 23.88 | 0.54 | 2.21 |
| | SRDRM [24] | 313.68 | 2.97 | 21.20 | 0.60 | 2.18 |
| | SRDRM-GAN [24] | 399.15 | 13.45 | 20.25 | 0.61 | 2.17 |
| | AMPCNet [38] | - | 1.25 | 23.83 | 0.62 | 2.25 |
| | PAL [31] | 325.51 | 2.99 | 22.51 | 0.63 | - |
| | RDLN [33] | 18.23 | 0.84 | 23.91 | 0.54 | 2.18 |
| | FEAN(ours) | 5.44 | 1.15 | 23.99 | 0.64 | 2.21 |

adopts the Transformer structure to improve reconstruction accuracy, but its evaluation metrics are much lower than our proposed FEAN. This indicates that our proposed network does not require complex mechanisms to improve the results, but exploits the proposed FEAM to explore and utilize different frequency characteristics to achieve better performance. Moreover, we present visual comparisons on the USR-248 dataset in Fig. 2. We can find that our FEAN yields better visual effects, whose recovered image produces sharper texture and detail information. AMPCNet and PAL show severe blurring artifacts. In contrast, our proposed FEAN can effectively solve these issues by exploring frequency domain cues.

C. EXPERIMENTAL EVALUATION ON THE UFO-120 DATASET

We compare our proposed FEAN with SRCNN [9], SRGAN [22], LatticeNet [44], ESRGAN [45], HNCT [46], SRDRM [24], SRDRM-GAN [24], SwinIR [47], Deep WaveNet [32], AMPCNet [38], and RDLN [33]. As demonstrated in Table 2, we can see that our FEAN obtains more competitive performance than existing SR networks. For example, in the

case of scale factor $\times 2$, our FEAN achieves the highest PSNR, SSIM, and UIQM scores. Although the UIQM of FEAN has lower scores on scale factors $\times 4$ and $\times 8$ compared to Deep WaveNet and RDLN, the difference is only 0.01. Additionally, we present visual comparisons in Fig. 3. It can be seen that the results reconstructed by AMPCNet and Deep WaveNet generate severe blurring artifacts. RDLN suffers from over-saturation of colors, while our FEAN obtains the most visually pleasant restored images.

D. ABLATION STUDY

To explicitly exhibit how our proposed components improve restoration results, we respectively remove MAEB, CAFB, and SMB to retrain the network, and the experimental results are shown Table 3. When FEAN is without MAEB, the performance drops dramatically, and the PSNR, SSIM, and UIQM are respectively reduced by 0.46dB, 0.022dB, and 0.04dB.

1) ABLATION OF PROPOSED CAFB IN FEAN

We compare the performance of the Squeeze-and-Excitation block (SEB) [48] and CAFB modules on the dataset

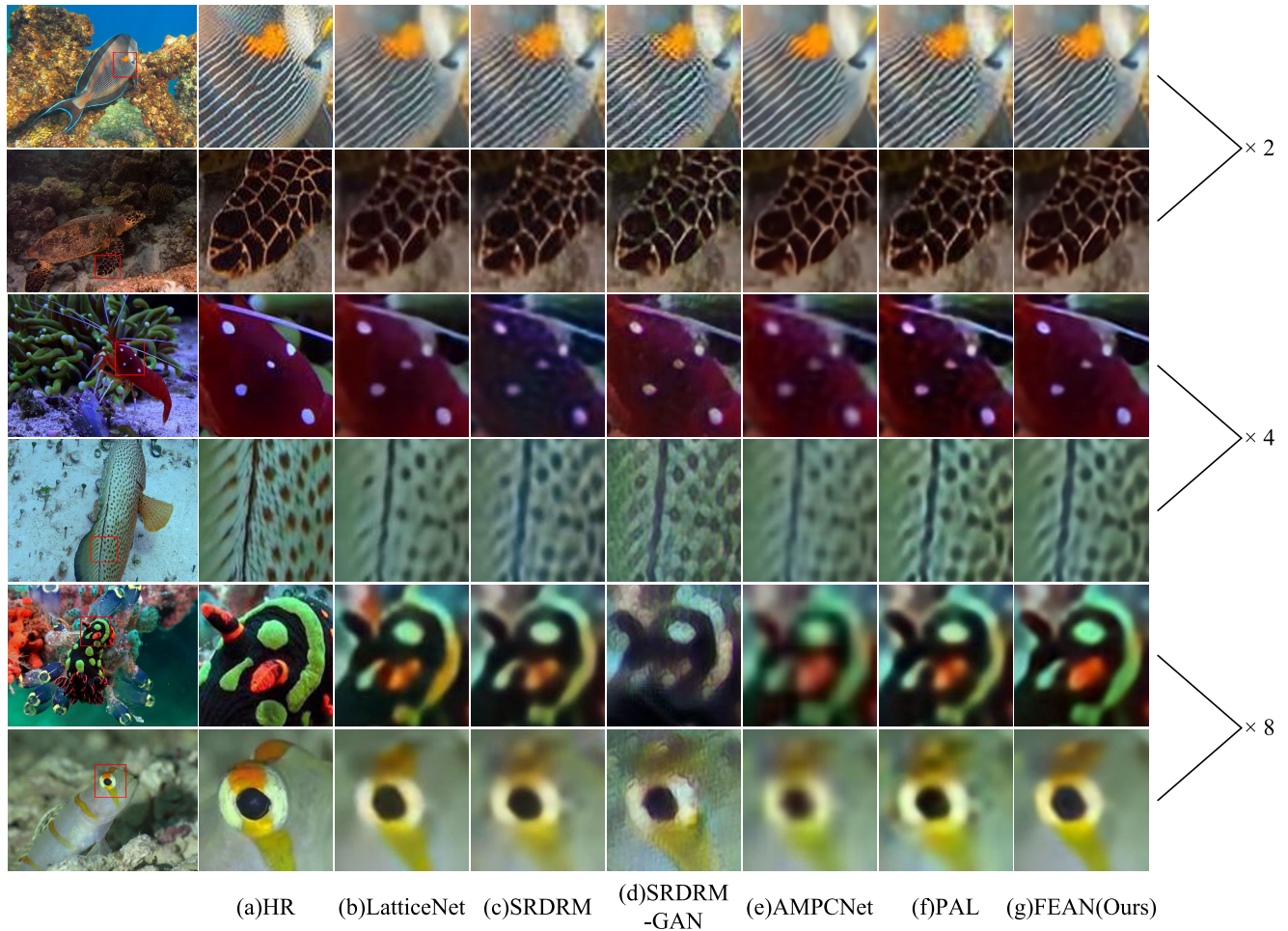


FIGURE 2. Visual comparison of our model against existing methods using the USR-248 dataset with scale factors $\times 2$, $\times 4$, and $\times 8$. (L to R) The reference underwater image (HR), LatticeNet, SRDRM, SRDRM-GAN, AMPCNet, PAL, and our FEAN.

TABLE 2. Quantitative results on the UFO-120 dataset with scale factors of $\times 2$, $\times 3$, and $\times 4$ for underwater image SR. Bold indicates the best performance.

| Method | Params(M) | | | PSNR(dB) | | | SSIM | | | UIQM | | |
|-------------------|------------|------------|------------|--------------|--------------|--------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | $\times 2$ | $\times 3$ | $\times 4$ | $\times 2$ | $\times 3$ | $\times 4$ | $\times 2$ | $\times 3$ | $\times 4$ | $\times 2$ | $\times 3$ | $\times 4$ |
| SRCNN [9] | 0.06 | 0.06 | 0.06 | 24.75 | 22.22 | 19.05 | 0.72 | 0.65 | 0.56 | 2.39 | 2.24 | 2.02 |
| SRGAN [22] | 5.95 | 5.95 | 5.95 | 26.11 | 23.87 | 21.08 | 0.75 | 0.70 | 0.58 | 2.44 | 2.39 | 2.56 |
| LatticeNet [44] | 0.76 | 0.77 | 0.78 | 25.86 | 26.13 | 25.10 | 0.71 | 0.71 | 0.70 | 2.97 | 2.94 | 2.94 |
| ESRGCNN [45] | 1.53 | 1.53 | 1.53 | 25.82 | 26.19 | 25.20 | 0.72 | 0.71 | 0.70 | 2.98 | 2.96 | 2.85 |
| HNCT [46] | 0.36 | 0.36 | 0.36 | 25.73 | 25.86 | 24.91 | 0.71 | 0.71 | 0.69 | 2.96 | 2.88 | 2.84 |
| SRDRM [24] | 0.83 | - | 1.90 | 24.62 | - | 23.15 | 0.72 | - | 0.67 | 2.59 | - | 2.57 |
| SRDRM-GAN [24] | 11.31 | - | 12.38 | 24.61 | - | 23.26 | 0.72 | - | 0.67 | 2.59 | - | 2.55 |
| SwinIR [47] | 0.88 | 0.89 | 0.90 | 25.70 | 26.41 | 25.35 | 0.75 | 0.75 | 0.71 | 3.02 | 2.97 | 2.94 |
| Deep WaveNet [32] | 0.28 | 0.28 | 0.29 | 25.71 | 25.23 | 25.08 | 0.77 | 0.76 | 0.74 | 2.99 | 2.96 | 2.97 |
| AMPCNet [38] | 1.15 | 1.16 | 1.17 | 25.24 | 25.73 | 24.70 | 0.71 | 0.70 | 0.70 | 2.93 | 2.85 | 2.88 |
| RDLN [33] | 0.84 | 0.84 | 0.84 | 25.96 | 26.55 | 25.37 | 0.76 | 0.74 | 0.73 | 2.98 | 2.98 | 2.94 |
| FEAN(ours) | 0.94 | 0.98 | 1.15 | 25.95 | 26.59 | 25.39 | 0.75 | 0.76 | 0.71 | 3.00 | 2.97 | 2.96 |

USR-248. The experimental results are presented in Table 4. In contrast to the FEAN model using SEB, FEAN model presents approximate results at scale factor $\times 2$, however, its PSNR improves by 0.012dB and 0.02dB at scale factor $\times 4$ and $\times 8$, respectively.

2) ABLATION OF PROPOSED GAUSSIAN FREQUENCY SEPARATION BLOCK IN FEAN

We compare the FEAN models with the Gaussian frequency separation module removed on the USR-248 dataset. As seen

from the experimental results in Table 4, the FEAN model with Gaussian frequency separation performs better in both PSNR and SSIM values at different deflation scales.

3) ABLATION OF PROPOSED MAEB WITH VARYING DILATION RATES IN FEAN

We conduct comparative experiments on the dilation convolution of MAEB with dilation rates set to 2 and 4, compared to the previous settings of 3 and 5, and the detailed results are shown in Table 4. MAEB with dilation rates of

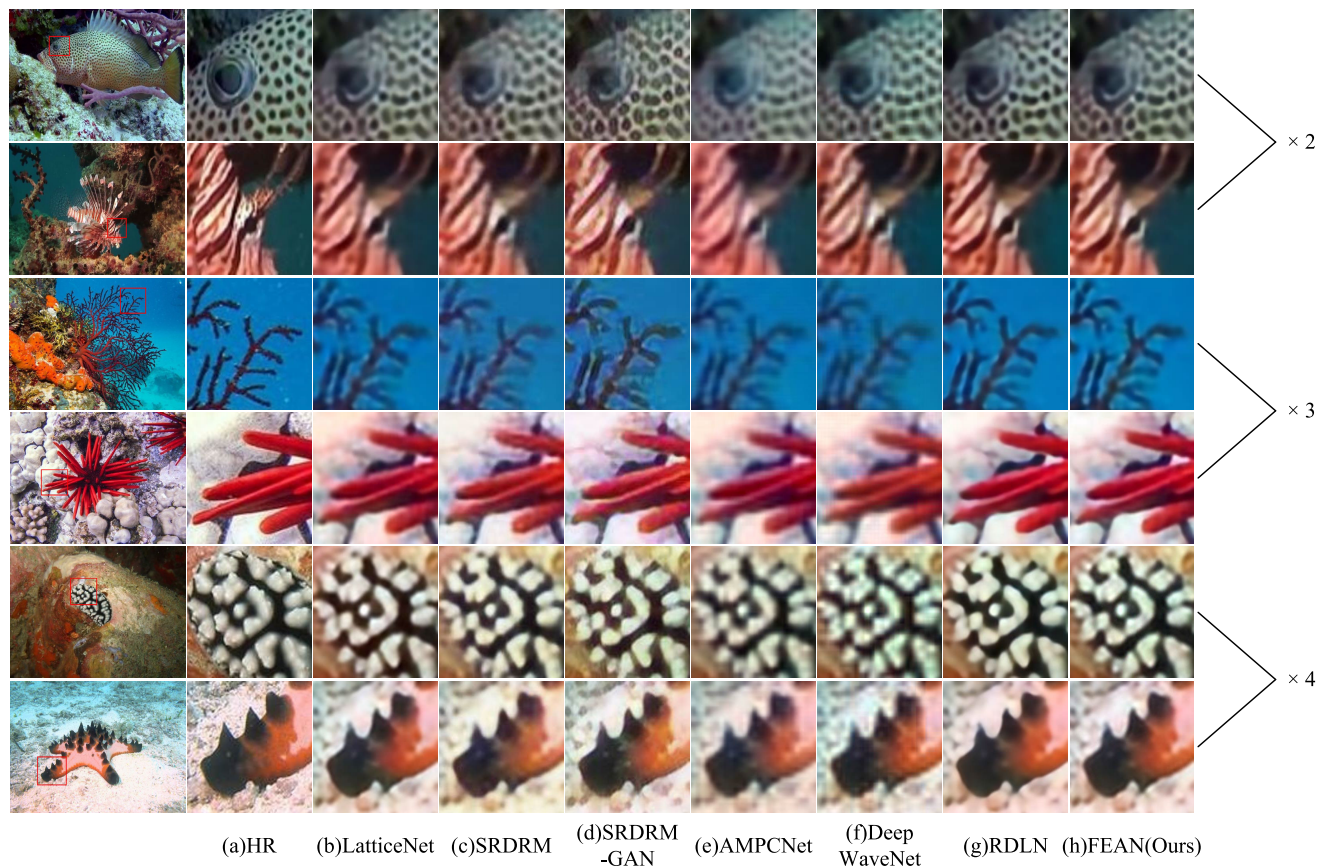


FIGURE 3. Visual comparison of our model against existing methods using the UFO-120 dataset with scale factors $\times 2$, $\times 3$, and $\times 4$. (L to R) The reference underwater image (HR), LatticeNet, SRDRM, SRDRM-GAN, AMPCNet, Deep WaveNet, RDLN, and our FEAN.

3 and 5 exhibits superior performance while maintaining the same number of parameters. We believe that adopting a larger convolutional dilation rate helps to capture a wider range of high-frequency features, thus enhancing the details of underwater image SR.

In addition, as demonstrated in Fig.4, we present the average gray-scale feature maps with and without MAEB enhancement on the high-frequency branch at different stages. We can see that the features enhanced by MAEB have sharper and finer texture details.

Although the networks with and without CAFB have smaller gains in SSIM and UIQM, they differ by 0.14 dB in PSNR. This is because CAFB can effectively gather high and low-frequency features to restore abundant texture characteristics. Fig.5 illustrates the heatmap of the high and low-frequency branches with and without CAFB processing for the first and fourth phases. The first row indicates the high-frequency branches and the second row indicates the low-frequency branches. As Fig.5 (c) and (d) demonstrate, the model processed with CAFB better enhances the frequency domain information, exhibiting more positive responses at the target object.

Moreover, FEAN augmented with SMB acquires a PSNR of 0.13 dB and a UIQM of 0.06 improvements, with increasing parameters by only 80K. The convergence results

TABLE 3. Ablation studies of proposed components on the UFO-120 with scale factor $\times 4$.

| Method | Params | PSNR(dB) | SSIM | UIQM |
|---------------|--------|----------|--------|------|
| FEAN w/o MAEB | 646K | 24.97 | 0.6893 | 2.95 |
| FEAN w/o CAFB | 934K | 25.29 | 0.7094 | 2.97 |
| FEAN w/o SMB | 905K | 25.30 | 0.7105 | 2.93 |
| FEAN | 979K | 25.43 | 0.7113 | 2.99 |

for different components are shown in Fig.6, it can be seen that FEAN without SMB converges more slowly. These results reveal the validity and rationality of the proposed components, which jointly contribute to the performance and computational efficiency.

E. EFFECTIVENESS OF MULTI-MODAL OBJECTIVE FUNCTION

We investigate the influence of loss function for the underwater SR task, as shown in Table 5. We conduct three loss function experiments, namely, L1, L2, and L1+L2. The experimental results show that the L1 loss function performs well in terms of PSNR and SSIM values, which are ahead of L2 by 0.172 and 0.025, respectively. Additionally, L1 outperforms L1+L2 in both PSNR (with a lead of 0.179dB) and SSIM (with a lead of 0.024). It is to be noted that the L1 loss function is superior in PSNR and SSIM, but it has some deficiencies in UIQM index compared to L1+L2. This

TABLE 4. Ablation studies of proposed components on the USR-248. "FEAN w/o Gaussian" denotes the model that does not use Gaussian filtering for frequency separation, "FEAN with MAEB(2,4)" denotes the model with MAEB dilation rates set to 2 and 4, and "FEAN with SEB" denotes the model that uses SEB instead of CAFB.

| Method | Scale | FLOPs(G) | Params(M) | PSNR(dB) | SSIM | UIQM |
|---------------------|-------|----------|-----------|----------|-------|-------|
| FEAN | ×2 | 71.06 | 0.94 | 29.865 | 0.816 | 2.754 |
| | ×4 | 18.56 | 0.98 | 26.188 | 0.694 | 2.601 |
| | ×8 | 5.44 | 1.15 | 23.946 | 0.632 | 2.301 |
| FEAN w/o Gaussian | ×2 | 71.04 | 0.94 | 29.862 | 0.805 | 2.762 |
| | ×4 | 18.56 | 0.98 | 26.171 | 0.687 | 2.609 |
| | ×8 | 5.44 | 1.15 | 23.894 | 0.629 | 2.315 |
| FEAN with MAEB(2,4) | ×2 | 71.06 | 0.94 | 29.866 | 0.817 | 2.786 |
| | ×4 | 18.56 | 0.98 | 26.179 | 0.693 | 2.609 |
| | ×8 | 5.44 | 1.15 | 23.924 | 0.632 | 2.288 |
| FEAN with SEB | ×2 | 68.54 | 0.90 | 29.867 | 0.817 | 2.773 |
| | ×4 | 17.93 | 0.94 | 26.176 | 0.694 | 2.601 |
| | ×8 | 5.28 | 1.11 | 23.926 | 0.634 | 2.288 |

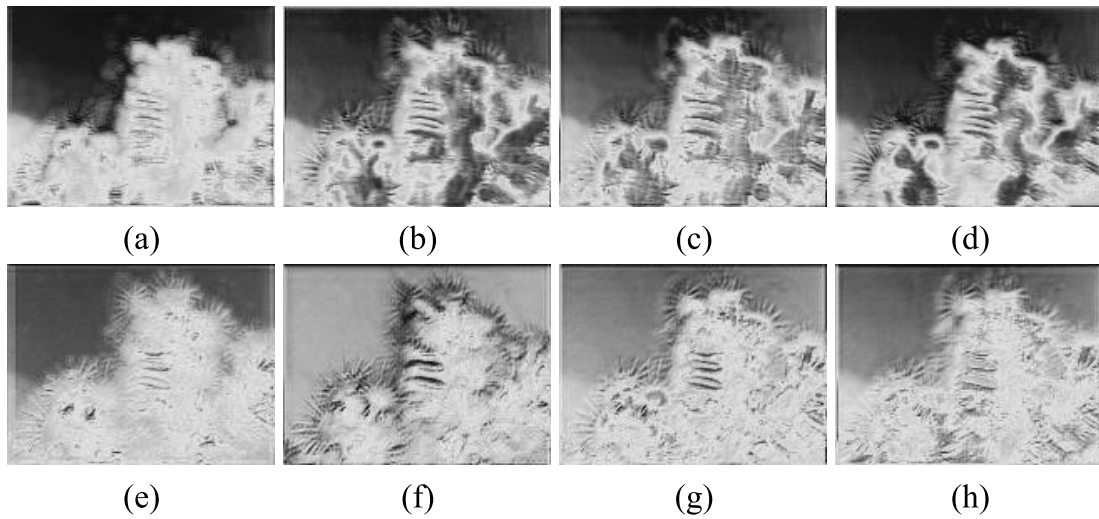


FIGURE 4. Visualized feature maps of FEAN with and without MAEB using grayscale colormap. (a)-(d) show feature maps of high-frequency branch without MAEB; (e)-(h) show feature maps of high-frequency branch with MAEB.

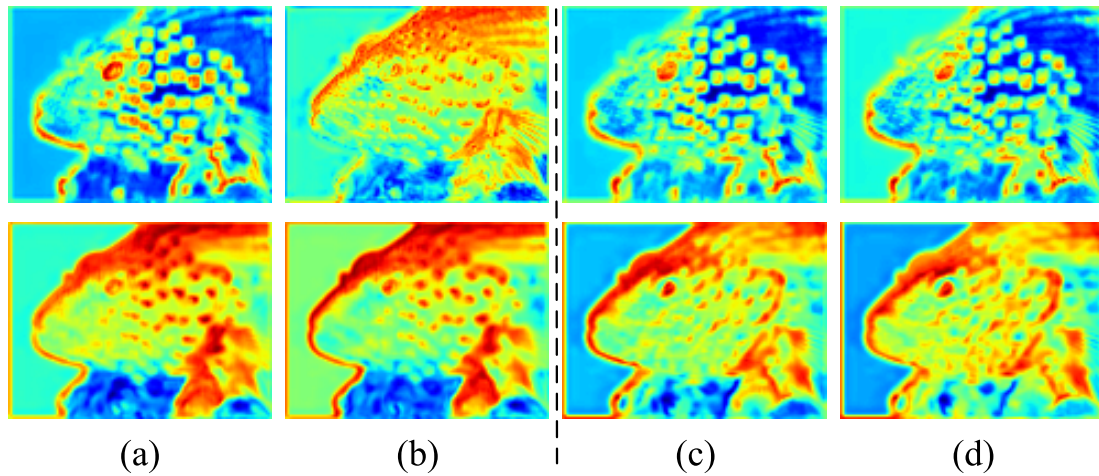


FIGURE 5. Visualized feature maps of FEAN without and with CAFB. (a) and (b) show the heatmaps on the high and low-frequency branches without CAFB; (c) and (d) show the heatmaps on the high and low-frequency branches with CAFB.

is probably because the L1 loss function pays more attention to minimizing the absolute error in the optimization process and ignores some image features related to human visual perception.

F. MODEL COMPLEXITY

To exhibit the computation efficiency of the proposed method, we evaluate it in terms of the number of parameters, FLOPs, average time, and computational efficiency,

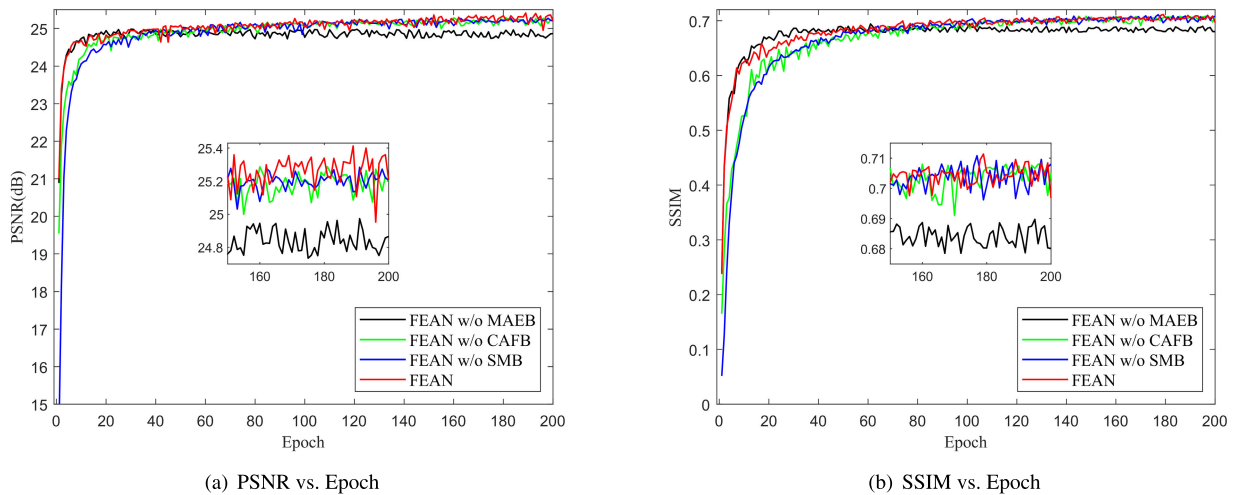


FIGURE 6. Convergence analysis of different components on UFO-120 with scale factor $\times 4$, where the x-axis is the number of Epochs, the y-axis in (a) is the PSNR value, and the y-axis in (b) is the SSIM value.

TABLE 5. Ablation studies show the effectiveness of different loss functions on the USR-248 with scale factors $\times 4$ dataset.

| Type | L1 | L2 | PSNR(dB) | SSIM | UIQM |
|------|----|----|---------------|--------------|--------------|
| 1 | ✓ | | 26.188 | 0.694 | 2.601 |
| 2 | | ✓ | 26.016 | 0.669 | 2.711 |
| 3 | ✓ | ✓ | 26.009 | 0.670 | 2.726 |

TABLE 6. Computation efficiency of representative methods with scale factors $\times 4$ on the USR-248 dataset.

| Method | FLOPs(G) | Params(M) | Average(s) | Efficiency(fps) |
|----------------|----------|-----------|------------|-----------------|
| SRCNN [9] | 21.30 | 0.06 | 0.04595 | 21.76 |
| VDSR [16] | 205.28 | 0.67 | 0.04795 | 20.85 |
| DSRCNN [43] | 15.77 | 1.11 | 0.03519 | 28.42 |
| EDSRGAN [17] | 206.42 | 1.97 | 0.04385 | 22.80 |
| SRResNet [22] | 85.49 | 1.59 | 0.03633 | 27.52 |
| SRGAN [22] | 529.86 | 5.95 | 0.05365 | 18.64 |
| ESRGAN [23] | 1504.09 | 16.70 | 0.12110 | 8.26 |
| SRDRM [24] | 291.73 | 1.90 | 0.04994 | 20.02 |
| SRDRM-GAN [24] | 377.20 | 12.38 | 0.05218 | 19.16 |
| FEAN(Ours) | 18.56 | 0.98 | 0.04671 | 21.41 |

respectively. We select some representative methods to compute the inference time on the USR-248 dataset, which are executed on the same device with the RTX3060. As Table 6 displays, the average time per image of our FEAN is achieved at 0.04671s (21.41fps), which is faster than most popular underwater approaches. As analyzed in Section IV-B, the quantitative results show that our proposed method obtains better restoration accuracy with appropriate computational complexity and superior computational efficiency. This is because our FEAN is composed of a sequence of FEAMs that keep the entire network light enough, facilitating fast inference in the network. Therefore, we can conclude that our proposed method strikes a good balance between network reconstruction performance and computational efficiency.

V. CONCLUSION

In this work, we propose a lightweight and effective frequency-domain enhanced attention network (FEAN) for

underwater SR. To be specific, frequency enhanced attention modules (FEAM), composed of attention enhancement block (MAEB) and channel attention fusion block (CAFB), acts as the backbone of the network inferring rich texture details in a coarse-to-fine manner. MAEB can capture more useful information in the high-frequency branch to facilitate the reconstruction of HR images. Further, CAFB is proposed to aggregate high and low-frequency information, which can achieve color deviation correction and detail enhancement. Extensive experiments show that our FEAN attains better performance with less model complexity than mainstream underwater SR methods. However, the diversity of underwater environments and limited training data limit the generalization ability of the model. In the future, we will focus on physical model learning to build a network that can adapt to a variety of imaging conditions in order and construct a large-scale dataset of real underwater images to improve the performance of super-resolution models for underwater images.

REFERENCES

- [1] C. Zhou, K. Qiu, C. Chen, D. Zhang, and Y. Guo, "Video super-resolution for wireless capsule endoscopy imaging sensor," *IEEE Sensors J.*, vol. 22, no. 17, pp. 17283–17290, Sep. 2022.
- [2] X. Duan, Y. Chen, M. Li, Y. Rong, R. Xie, and J. Han, "UArch: A super-resolution processor with heterogeneous triple-core architecture for workloads of U-Net networks," *IEEE Trans. Biomed. Circuits Syst.*, vol. 17, no. 3, pp. 633–647, Mar. 2023.
- [3] Y. Xiao, Q. Su, Q. Yuan, D. Liu, H. Shen, and L. Zhang, "Satellite video super-resolution via multiscale deformable convolution alignment and temporal grouping projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5610819.
- [4] Y. Xiao, Q. Yuan, K. Jiang, J. He, Y. Wang, and L. Zhang, "From degrade to upgrade: Learning a self-supervised degradation guided adaptive network for blind remote sensing image super-resolution," *Inf. Fusion*, vol. 96, pp. 297–311, Aug. 2023, doi: 10.1016/j.inffus.2023.03.021.
- [5] K. Jiang, Z. Wang, P. Yi, G. Wang, K. Gu, and J. Jiang, "ATMFN: Adaptive-threshold-based multi-model fusion network for compressed face hallucination," *IEEE Trans. Multimedia*, vol. 22, no. 10, pp. 2734–2747, Oct. 2020.

- [6] K. Jiang, Z. Wang, P. Yi, T. Lu, J. Jiang, and Z. Xiong, "Dual-path deep fusion network for face image hallucination," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 1, pp. 378–391, Jan. 2022.
- [7] M. Hu, K. Jiang, Z. Wang, X. Bai, and R. Hu, "CycMuNet+: Cycle-projected mutual learning for spatial-temporal video super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 13376–13392, Jul. 2023.
- [8] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.
- [9] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 184–199. [Online]. Available: <https://api.semanticscholar.org/CorpusID:18874645>
- [10] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Computer Vision—ECCV 2018 (Lecture Notes in Computer Science)*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, p. 294–310.
- [11] B. Niu, W. Wen, W. Ren, X. Zhang, L. Yang, S. Wang, K. Zhang, X. Cao, and H. Shen, *Single Image Super-Resolution Via a Holistic Attention Network (Lecture Notes in Computer Science)*, vol. 12357. Cham, Switzerland: Springer, 2020, pp. 191–207. [Online]. Available: https://link.springer.com/10.1007/978-3-030-58610-2_12
- [12] Z. Wang, G. Gao, J. Li, Y. Yu, and H. Lu, "Lightweight image super-resolution with multi-scale feature interaction network," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2021, pp. 1–6. [Online]. Available: <https://api.semanticscholar.org/CorpusID:232335914>
- [13] L. Wang, K. Li, J. Tang, and Y. Liang, "Image super-resolution via lightweight attention-directed feature aggregation network," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 19, no. 2, pp. 1–23, Feb. 2023, doi: [10.1145/3546076](https://doi.org/10.1145/3546076).
- [14] Y. Chen, K. Niu, Z. Zeng, and Y. Pan, "A wavelet based deep learning method for underwater image super resolution reconstruction," *IEEE Access*, vol. 8, pp. 117759–117769, 2020.
- [15] Y. Li, C. Ma, T. Zhang, J. Li, Z. Ge, Y. Li, and S. Serikawa, "Underwater image high definition display using the multilayer perceptron and color feature-based SRCNN," *IEEE Access*, vol. 7, pp. 83721–83728, 2019.
- [16] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.
- [17] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.
- [18] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in *Proc. 15th Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany. Berlin, Heidelberg: Springer-Verlag, Sep. 2018, pp. 527–542, doi: [10.1007/978-3-030-01237-3_32](https://doi.org/10.1007/978-3-030-01237-3_32).
- [19] A. Muqet, J. Hwang, S. Yang, J. H. Kang, Y. Kim, and S.-H. Bae, "Multi-attention based ultra lightweight image super-resolution," in *Proc. ECCV Workshops*, 2020, pp. 103–118. [Online]. Available: <https://api.semanticscholar.org/CorpusID:221834582>
- [20] J. Xu, Y. Chae, B. Stenger, and A. Datta, "Dense bynet: Residual dense network for image super resolution," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 71–75.
- [21] K. Jiang, Z. Wang, P. Yi, and J. Jiang, "Hierarchical dense recursive network for image super-resolution," *Pattern Recognit.*, vol. 107, Nov. 2020, Art. no. 107475. [Online]. Available: <https://api.semanticscholar.org/CorpusID:219734185>
- [22] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA: IEEE, Jul. 2017, pp. 105–114. [Online]. Available: <http://ieeexplore.ieee.org/document/8099502/>
- [23] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Computer Vision—ECCV 2018 Workshops*. Munich, Germany. Berlin, Germany: Springer-Verlag, 2019, pp. 63–79, doi: [10.1007/978-3-030-11021-5_5](https://doi.org/10.1007/978-3-030-11021-5_5).
- [24] M. J. Islam, S. Sakib Enan, P. Luo, and J. Sattar, "Underwater image super-resolution using deep residual multipliers," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 900–906.
- [25] K. Prajapati, V. Chudasama, H. Patel, K. Upla, R. Ramachandra, K. Raja, and C. Busch, "Unsupervised single image super-resolution network (USISResNet) for real-world data using generative adversarial network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Seattle, WA, USA: IEEE, Jun. 2020, pp. 1904–1913. [Online]. Available: <https://ieeexplore.ieee.org/document/9151093/>
- [26] K. Zhang, J. Liang, L. Van Gool, and R. Timofte, "Designing a practical degradation model for deep blind image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4771–4780.
- [27] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, vol. 37, 2015, p. 448–456.
- [28] M. Jahidul Islam, P. Luo, and J. Sattar, "Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception," 2020, *arXiv:2002.01155*. [Online]. Available: <https://api.semanticscholar.org/CorpusID:211020567>
- [29] L. Wang, L. Xu, W. Tian, Y. Zhang, H. Feng, and Z. Chen, "Underwater image super-resolution and enhancement via progressive frequency-interleaved network," *J. Vis. Commun. Image Represent.*, vol. 86, Jul. 2022, Art. no. 103545.
- [30] T. Ren, H. Xu, G. Jiang, M. Yu, X. Zhang, B. Wang, and T. Luo, "Reinforced swin-convs transformer for simultaneous underwater sensing scene image enhancement and super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4209616.
- [31] X. Chen, S. Wei, C. Yi, L. Quan, and C. Lu, "Progressive attentional learning for underwater image super-resolution," in *Proc. Int. Conf. Intell. Robot. Appl.*, 2020, pp. 233–243. [Online]. Available: <https://api.semanticscholar.org/CorpusID:231589470>
- [32] P. Sharma, I. Bisht, and A. Sur, "Wavelength-based attributed deep neural network for underwater image restoration," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 19, no. 1, pp. 1–23, Jan. 2023, doi: [10.1145/3511021](https://doi.org/10.1145/3511021).
- [33] Z. Chen, C. Liu, K. Zhang, Y. Chen, R. Wang, and X. Shi, "Underwater-image super-resolution via range-dependency learning of multiscale features," *Comput. Electr. Eng.*, vol. 110, Sep. 2023, Art. no. 108756.
- [34] A. Aghelan, "Underwater image super-resolution using generative adversarial network-based model," in *Proc. 13th Int. Conf. Comput. Knowl. Eng. (ICCKE)*, 2023, pp. 480–484, doi: [10.1109/ICCKE60553.2023.10326266](https://doi.org/10.1109/ICCKE60553.2023.10326266).
- [35] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1905–1914. [Online]. Available: <https://api.semanticscholar.org/CorpusID:236171006>
- [36] J.-H. Kim, J.-H. Choi, M. Cheon, and J.-S. Lee, "Ram: Residual attention module for single image super-resolution," 2018, *arXiv:1811.12043*. [Online]. Available: <https://api.semanticscholar.org/CorpusID:53980605>
- [37] Y. Hu, J. Li, Y. Huang, and X. Gao, "Channel-wise and spatial feature modulation network for single image super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3911–3927, Nov. 2020.
- [38] Y. Zhang, S. Yang, Y. Sun, S. Liu, and X. Li, "Attention-guided multi-path cross-CNN for underwater image super-resolution," *Signal, Image Video Process.*, vol. 16, no. 1, pp. 155–163, Feb. 2022.
- [39] B. Fu, L. Wang, R. Wang, S. Fu, F. Liu, and X. Liu, "Underwater image restoration and enhancement via residual two-fold attention networks," *Int. J. Comput. Intell. Syst.*, vol. 14, no. 1, p. 88, 2020.
- [40] S. Anwar and N. Barnes, "Densely residual Laplacian super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1192–1204, Mar. 2022.
- [41] J. Yu, Y. Fan, J. Yang, N. Xu, Z. Wang, X. Wang, and T. Huang, "Wide activation for efficient and accurate image super-resolution," 2018, *arXiv:1808.08718*. [Online]. Available: <https://api.semanticscholar.org/CorpusID:52098212>
- [42] H. Wang, H. Wu, Q. Hu, J. Chi, X. Yu, and C. Wu, "Underwater image super-resolution using multi-stage information distillation networks," *J. Vis. Commun. Image Represent.*, vol. 77, May 2021, Art. no. 103136. [Online]. Available: <https://api.semanticscholar.org/CorpusID:235565167>
- [43] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," 2016, *arXiv:1606.08921*.

- [44] X. Luo, Y. Xie, Y. Zhang, Y. Qu, C. Li, and Y. R. Fu, "LatticeNet: Towards lightweight image super-resolution with lattice block," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 272–289. [Online]. Available: <https://api.semanticscholar.org/CorpusID:221725942>
- [45] C. Tian, Y. Yuan, S. Zhang, C.-W. Lin, W. Zuo, and D. Zhang, "Image super-resolution with an enhanced group convolutional neural network," *Neural Netw.*, vol. 153, pp. 373–385, Sep. 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0893608022002143>
- [46] J. Fang, H. Lin, X. Chen, and K. Zeng, "A hybrid network of CNN and transformer for lightweight image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*. Piscataway, NJ, USA: IEEE, Jun. 2022, pp. 1102–1111. [Online]. Available: <https://ieeexplore.ieee.org/document/9857212/>
- [47] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1833–1844.
- [48] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.



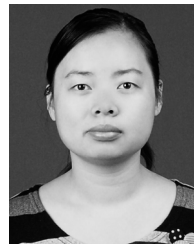
HAIMING DING received the B.S. degree in communication engineering from the Wanjiang University of Technology, Ma'anshan, China, in 2022.

His research interests include deep learning and image and video processing.



XIN LIU received the M.S. degree in computer science and technology from Hohai University, Nanjing, China, in 2023.

Since 2023, he has been with the Jinken College of Technology. He is the author of one article and three inventions. His research interests include deep learning and image and video processing.



MIN ZHANG received the master's degree in computer application technology from the Jiangxi University of Technology, China, in 2008.

She is currently with the Gannan University of Science and Technology. Her research interests include image processing and signal processing.



ZHENGXIANG GU received the M.S. degree in computer science and technology from Southeast University, Nanjing, China, in 2003.

Since 2003, he has been with the Jinken College of Technology, where he became the Vice-Senior, in 2020. He is the author of four books, more than 12 articles, and more than seven inventions. His research interest includes image processing.



LI WANG received the M.S. degree in electronic and communication engineering and the Ph.D. degree in information and communication engineering from Hohai University, Nanjing, China, in 2019 and 2023, respectively.

Her research interests include deep learning and image and video processing.

...