IEICE TRANSACTIONS

on Fundamentals of Electronics, Communications and Computer Sciences

DOI:10.1587/transfun.2024EAP1038

Publicized:2024/10/15

A PUBLICATION OF THE ENGINEERING SCIENCES SOCIETY

The Institute of Electronics, Information and Communication Engineers
Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3 chome, Minato-ku, TOKYO, 105-0011 JAPAN

PAPER

FSRN: Feature separable reconstruction network for underwater optical image super-resolution

Lan XIE [†], Qiang WANG [†], Yongqiang JI[† a)], Yu GU [†], Gaozheng XU [†], Zheng ZHU [†], Yuxing WANG [†], Yuwei LI[†], ***Nonmembers***

SUMMARY Underwater image super-resolution reconstruction technologies have played a very important role in ocean resource exploration since it can significantly improve the clarity of underwater optical images. Although recent deep learning based methods have achieved promising performance in terrestrial image super-resolution, these methods lack sufficient capabilities to handle those dark, turbid and blurred underwater images. In this work, we propose a novel network, namely feature separable reconstruction network (FSRN), to separate the extraction features and the reconstruction features for better using of features in each layer, solving the problem of long-distance transmission of shallow features in the neural network. We design a depthwise separable convolutional residual block with large convolutional kernels(DWRB) to augment receptive fields, which improves the effectiveness of high-frequency feature extraction in the blur images. We further propose a channel attention mechanism based on the SE module and explore an optimal attention module insertion mode which pays more attention to the weight between reconstruction information, reducing information loss. Moreover, we also modify the convolutional kernel padding mode and propose a perceptual loss function with boundary clipping to avoid the inconsistent in feature extraction from boundary and non-boundary regions. Extensive experiments on underwater datasets demonstrate our proposed underwater super-resolution framework outperform over the state-of-the-art methods in terms of reconstruction accuracy and real-time performance.

key words: *Underwater image super-resolution reconstruction, deep learning, depthwise separable convolution, attention module insertion mode.*

1. Introduction

With the continuous exploration of marine resources, underwater robots are playing an increasingly important role in underwater target capture, underwater search and underwater investigation. Compared to sonar, laser and other devices, the optical camera is low cost and it can catch more types of visual information which could be conducive to obtain the information of marine environment by underwater robots. However, underwater optical images suffer from a degradation in imaging quality that is characterized by color deviation, low contrast and inconspicuous high-frequency details because of the complex underwater environment and underwater optical properties, which is expected to be resolved by the automatic underwater image super-resolution technology.

Current image super-resolution technologies [1-3] can improve the signal-to-noise ratio and clarity of images and restore image details, as well as lower the transmission bandwidth and storage space of cameras. However, we find that when these traditional image super-resolution methods which show good performance in the terrestrial image super-resolution field are directly applied to the underwater tasks, they can hardly maintain high accuracy and robustness. These methods mainly rely on some prior knowledge of the image to interpolate or establish degradation models as the basis for reconstructing high-resolution images. Nevertheless, different from terrestrial imagery, the prior knowledge of underwater images is difficult to be established accurately. Furthermore, the prior knowledge is only applicable to a specific scene, and the generalization ability becomes poor when encountering complex and diverse environments. Underwater image super-resolution in marine is still a challenging and crucial task.

This motivates the need to automatically learn the essential features of images through an automated end-to-end process which can achieve better super-resolution reconstructing accuracy and generalization performance without cumbersome prior knowledge. In recent years, the rapid development of deep learning has provided a new approach to the field of image super-resolution. Convolutional Neural Networks (CNNs) have been widely used as powerful characterization tools for image super-resolution reconstruction. Dong [4] used a three-layer end-to-end convolutional neural network based on deep learning to solve the problem of image super-resolution for the first time, called SRCNN, which achieved significant improvement compared to traditional image super-resolution methods. Shi [5] proposed ESPCN that eliminated pre-upsampling and performed upsampling through subpixel convolution after feature extraction, greatly saving a lot of unnecessary calculations. Lim [6] used a deeper neural network to extract features and added residual connections to solve the problem of difficulty in training a deep neural network. Zhang [7] introduced an attention mechanism into the field of image super-resolution and achieved good performance by stacking channel attention modules in feature extraction networks. Hui [8] proposed a lightweight network based on multi-information distillation, which greatly reduced the number of parameters and achieved good performance with a relatively shallow

[†] The authors are with Systems Engineering Research Institute of CSSC, Beijing 100036, China.
a) E-mail: ioajiyq@163.com (Corresponding author)

model. However, these superior methods based on deep learning still exist three challenges for underwater image super-resolution as follows:

· The underwater images are blurred and low in brightness. How to extract high-frequency features from blurred underwater optical images more efficiently.

· During our research, we find that low-level features that are rich in high-frequency features occupy an extremely important position in underwater image super-resolution reconstruction. How to solve the problem of long-distance transmission of shallow features in a deep neural network.

· How to use attention mechanisms more efficiently in underwater image super-resolution networks.

To address these problems, we propose a underwater super-resolution network named Feature Separable Reconstruction Network(FSRN),which takes advantage of the features from each layer to improve reconstruction accuracy. For the first challenge, we introduce a depthwise separable convolutional residual block (DWRB) to enhance the high-frequency feature representation. DWRB uses large convolutional kernels of 7×7 to make the network have a larger receptive field while simultaneously introduced the depthwise separable convolution structure to spare the large number of parameters. To resolve the second issue, we devise information distillation group which separates reconstruction information from feature extraction information to circumvent the issue of shallow feature information loss. For the third challenge, Improved SE module is proposed to enhance feature extraction capabilities. At the same time, we explore an optimal attention module insertion mode suitable for our network. Extensive experiments on underwater datasets demonstrate our proposed underwater super-resolution framework achieves better performance in reconstruction accuracy and real-time performance against the state-of-the-art methods. In a word, our contributions are summarized as follows:

(1) We propose a lightweight underwater super-resolution network which fully combines shallow and deep features, achieving accurate super-resolution reconstruction with fewer parameters and faster speed.

(2) We conduct experiments and explore an optimal attention module insertion mode suitable for underwater image super-resolution networks.

(3) We conduct extensive experiments on UFO-120, USR-248 and GBK-100 to verify that the proposed lightweight model has significant advantages in terms of reconstruction accuracy and real-time performance against other super-resolution methods.

2. Related works

2.1 Deep learning based image super-resolution methods

Convolutional Neural Networks (CNNs) have rapidly become popular in image and video processing, which also exhibits excellent fitting ability in the field of image super-resolution. Dong [4] proposed SRCNN, which used a three-layer end-to-end convolutional neural network to solve the problem of image super-resolution for the first time. Compared to traditional image super-resolution methods, it had achieved significant improvement. ESPCN [5] eliminated the pre-upsampling operation and utilized subpixel convolution for the first time to map low-resolution images to high-resolution images, reducing a large number of network parameters. In order to achieve better performance, the VDSR [9] trained a deeper neural network using residual learning, gradient clipping, and a high learning rate. Lim [6] also used a deeper neural network to extract image features, adding residual connections to solve the problem of difficulty in training a deep neural network. This method, called EDSR, has become a benchmark for many subsequent works. In order to fully utilize the features of each layer, the RDN [10] used dense residual blocks to enhance information flow and network expression capabilities.

In recent years, attention mechanisms and lightweight ideas have also been widely applied in the design of image super-resolution networks. RCAN [7] introduced a channel attention mechanism to learn the weight of each channel. Niu [11] proposed the HAN model, which not only used a channel attention mechanism but also performed attention modules to the spatial and layer domains. In lightweight designs, LapSRN [12] combined the Laplace pyramid with deep learning to propose a lightweight model. Hui et al. [13-14] struck a fair compromise between performance and computational complexity by lightweighting the network through information distillation. These general super-resolution structures provide a lot of inspiration for underwater image super-resolution technologies.

2.2 Super-resolution methods in underwater optical images

Among underwater research, studies on image super-resolution are not particularly prolific. Some traditional approaches primarily focus on enhancing underwater image reconstruction quality by deblurring, denoising, or de-scattering [15-17]. In addition, a series of deep learning based image super-resolution algorithms [18-20] have been applied to underwater optical image reconstruction thanks to some datasets, such as USR-248 [18], UFO-120 [19], and USR-2K [20], have been established and released, which alleviate the problem of dataset scarcity. Islam et al. [18] proposed a deep residual multiplier model, called SRDRM, to efficiently reconstruct more texture details of underwater images. Wang et al. [20] proposed a lightweight multistage information distillation network to balance model performance and computational speed in underwater image super-resolution tasks. Cherian et al. [21] proposed a practical underwater image super-resolution network, called AlphaSRGAN, which combined traditional image super-resolution approaches with deep learning methods. This method merged pre-processing images before feeding them into the generator network, which improved reconstruction performance and stability. Ren et al. [22] applied transformer structure to a U-shaped structure for underwater super-resolution reconstruction and achieved excellent performance. However, most of these underwater super-resolution methods use deeper and more complex networks which primarily focus on deep feature extraction to improve the quality of reconstruction, while ignoring the importance of shallow features for underwater super-resolution reconstruction. In this paper, we aim to design a lightweight underwater super-resolution model which can fully utilizes the features from each layer to improve reconstruction accuracy with fewer parameters and faster speed.

3. Methodology

3.1 The overall view of Feature Separable Reconstruction Network

Our proposed Feature Separable Reconstruction Network (FSRN) can separate reconstruction features from extraction features, making the final reconstruction feature map contain both strong semantic information and rich high-frequency features, solving the problem of shallow features requiring long-distance transmission. The overall structure is shown in Fig.1, which consists of the following components: (1) Shallow feature extraction layer H_{sf} to extract preliminary features F_{sf} while simultaneously to expand channel size. (2) Deep feature extraction backbone H_b :The backbone is composed of eight information distillation groups. Each information distillation group not only outputs feature extraction information F_B as input for the next group but also outputs reconstruction information F_S. (3) Reconstruction features fusion module H_c: In this part, we integrate the outputs of each information distillation group and fuse F_{sf} through residual structure [23] to make the reconstruction information contain features from each layer. (4) Reconstruction module H_{rec}: We use subpixel convolution [24] for upsampling. Then the upsampling result is convolved by 3×3 filters to obtain the final output.

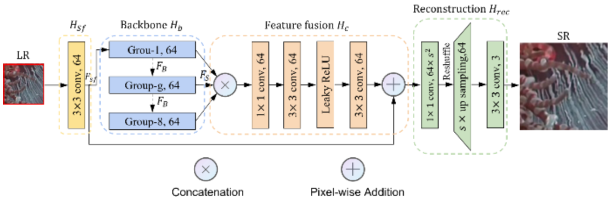

Fig. 1 The overall of FSRN.

3.2 Information distillation group

The proposed information distillation group is shown in Fig.2 The whole structure is divided into two parts: feature extraction backbone outputting feature extraction information and feature distillation fusion module outputting reconstruction information. The feature extraction backbone is composed of multiple depthwise separable convolutional residual blocks(DWRB) whose output is separated into two branches. Branch 1 serves as the input for the next residual block to continue extracting feature information. Branch 2 will be passed to the feature distillation fusion module. In this part, we first use a 1×1 convolution H_d to distill the output as reconstruction features, and then splice the distilled features of each residual block along the channel dimension. Afterward, a 1×1 convolution H_t is used to adjust the channel size. Note that we insert an optional attention module after the last DWRB and before the H_t, which we will introduce in Section 3.3. In addition, the feature extraction backbone of the information distillation group will have a long residual connection to ensure that the network can be trained more easily. This overall process can be expressed as:

$$(F_{s1}, F_{b1}) = R_1(F_{Bn-1}), \cdots, (F_{sM}, F_{bM}) = R_M(F_{bM-1}) \quad (1)$$

$$F_{Sn} = H_t[(Att_S(H_{d1}(F_{s1}) \otimes H_{d2}(F_{s2}) \otimes \cdots \otimes H_{dM}(F_{sM}))] \quad (2)$$

$$F_{Bn} = Att_b(F_{bM}) \oplus F_{Bn-1} \quad (3)$$

where $\otimes$ is the process of feature concatenation while $\oplus$ is the process of pixel-wise addition. Attention-b denotes the attention module inserted in feature extraction backbone while Attention-s denotes the attention module inserted in feature distillation fusion module. R_M denotes the Mth DWRB and (F_{sM}, F_{bM}) is used to show the output results of each DWRB.

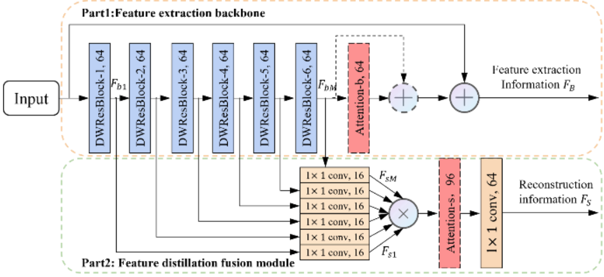

Fig. 2 Information distillation group.

The underwater images frequently exhibit severe blurring due to scattering, making a larger receptive field necessary for extracting high-frequency features. In order to augment receptive field and spare parameters, we combine large convolutional kernels and depthwise separable convolution [25] to propose the depthwise separable convolutional residual block(DWRB). Fig.3 shows the structure of the DWRB, which can be roughly divided into two stages: depthwise convolution and pointwise convolution. Depthwise convolution is a convolution where each input channel corresponds to only one convolutional kernel. Here, we use 7×7 large convolutional kernels. Pointwise convolution is a traditional 1×1 convolution that is used to exchange information between channels and adjust output channel size. In addition, inspired by Sandler et al. [26], we construct the backend of the DWRB into an anti-bottleneck structure. Finally, the output and input are fused through a shortcut.

The advantages of this structure are as follows: (1)

Using 7×7 large convolutional kernels increases the receptive field during feature extraction, making it easier to extract low-frequency and high-frequency features from blurred underwater images. (2) Adopting depthwise separable convolution effectively solves the problem of parameter increase caused by large convolutional kernels. This is also the core of the lightweight idea in this paper. (3)Performing a skip layer connection to integrate the output and input of the module, which makes the model produce a combined receptive field that can balance high-frequency and low-frequency features.

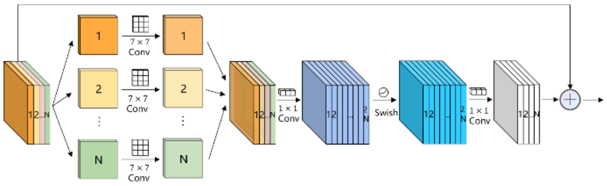

Fig. 3 Depthwise separable convolution residual block(DWRB).

3.3 Improved SENet

In this paper, we plan to apply the channel attention mechanism to the information distillation group, aiming to filter irrelevant information and highlight key information. Additionally, we design three basic attention module insertion modes. Attention-b is inserted after the last DWRB in the information distillation group. Attention-b+res adds a residual connection on the basis of Attention-b to reduce information loss. However, it also weakens the role of the attention module in filtering irrelevant information. Attention-s is inserted in front of the 1×1 convolution layer H_t to make the network pay more attention to the weight changes of the reconstructed features.

We introduce the Improved SENet, adding channel maximum information on the basis of SENet [27], as shown in Fig.4. We first perform global average pooling and global maximum pooling on the input to extract the average and maximum values of each channel as two feature statistics.

$$A_{Max} = GMP(F_{IN}),\ A_{avg} = GAP(F_{IN}) \tag{4}$$

Then, similar to SEnet, the two statistics will be fed into an shared MLP which consists of three fully connected layers, with the second fully connected layer reducing the number of neurons to $\frac{1}{8}$ of the original channel size.

After going through the Sigmoid function, the output result is multiplied by 2 to get the final weight of each channel. The purpose of multiplying by 2 here is to map the weight to the interval between 0 and 2 which is conducive to train network.

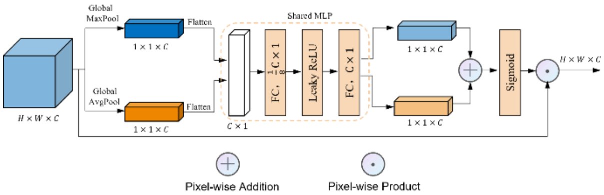

Fig. 4 Improved SENet with maxpooling information.

3.4 Loss function

In previous research on image super-resolution, few have explored the effect of image boundaries on model training. During our research, we find that the extraction of image boundary features is inconsistent with that of non-boundary features for convolutional kernels due to translation invariance. To address the above issue, the c pixels of the image boundary will not be used for calculating losses.

We proposed a loss function with boundary clipping named perceptual loss function which consists of three parts: pixel loss, content loss and generative adversarial loss. As shown in Eq. 5.

$$L_P = L_1 + L_C + \lambda L_{GAN} \tag{5}$$

Pixel loss can often make the I_{SR} have a high PSNR. In this paper, L_1 loss function is used to train the model. As shown in the Eq. 6.

$$L_1 = \frac{1}{(w-2c)(h-2c)} \sum_{x=c+1}^{w-c} \sum_{y=c+1}^{h-c} |I_{HR}(x,y) - I_{SR}(x,y)| \tag{6}$$

where w and h refer to the width and height of the image, c denotes the number of boundary clipping pixels, I_{SR} is the reconstructed super-resolution image and I_{HR} is the original high-resolution image.

The content loss measures the degree of similarity by comparing the high-level feature differences between I_{SR} and I_{HR}. We use VGG-19 to extract features, select feature maps of several layers in VGG, and calculate the loss of feature maps between I_{SR} and I_{HR} using L_1 loss function. As shown in the Eq. 7.

$$L_C = \frac{1}{(w_{i,j}-2c)(h_{i,j}-2c)} \sum_{x=c+1}^{w_{i,j}-c} \sum_{y=c+1}^{h_{i,j}-c} a_{i,j} |\phi_{i,j}[I_{HR}(x,y)] - \phi_{i,j}[I_{SR}(x,y)]| \tag{7}$$

where $\phi_{i,j}$ represents the feature map output by the jth convolution layer before the ith maximum pooling layer in the VGG network, a denotes the loss weight of this feature map. In this paper, we select these layers of $\{\phi_{1,1}, \phi_{2,1}, \phi_{3,1}, \phi_{4,1}, \phi_{5,1}\}$ and the values of a of these layers are $\{0.1, 0.1, 1, 1, 1\}$ respectively.

Using generative adversarial loss functions can effectively restore texture details and generate visually more realistic images [28-30]. In this paper, we use U-Net [31] as the discriminator (as shown in Fig.5). The generative adversarial loss function is the binary cross entropy loss which is shown in Eq.8 and the loss function used for discriminator training is shown in Eq.9.

$$L_{GAN} = \frac{1}{(w-2c)(h-2c)} \sum_{x=c+1}^{w-c} \sum_{y=c+1}^{h-c} -LOG\ [D(I_{SR}(x,y))] \tag{8}$$

$$L_D = -\mathbb{E}_{x\sim P_{HR}}[LOG(D(x))] - \mathbb{E}_{z\sim P_{SR}}[LOG(1 - D(z))] \tag{9}$$

where D denotes the discriminator, x represents the real image data and z represents the image data generated by the generator G.

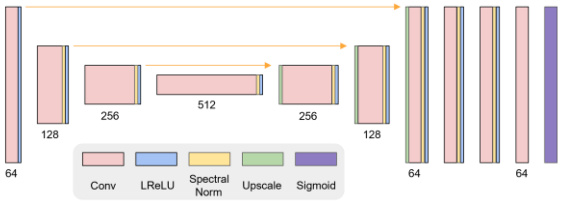

Fig. 5 Generative adversarial network discriminator based on U-Net.

Using the above loss functions with boundary clipping, the model will ignore the difference between image boundary and non-boundary feature extraction, but this also leads to a black edge in the reconstruction results. In order to make the image boundary pixels of I_{SR} more natural visually, the padding mode of replication is used in all convolutional layers in this paper.

4. Experiment and result

4.1 Experimental data and preprocessing

Currently, there are relatively few publicly available underwater image super-resolution datasets. Therefore, we integrated two publicly available datasets, USR-248 and UFO-120, to evaluate our approach. USR-248 includes 1060 images in the training set and 248 images in the testing set. We selected 60 images randomly from the training set as the validation set. The UFO-120 has 1500 training images and 120 test images. This dataset performs Gaussian blur in the downsampling process, which increases the difficulty of image super-resolution reconstruction. In addition, in order to further verify the generalization performance of the model, we randomly selected 100 seabed images from the publicly available data of IMOS and established a dataset called GBR-100 which was only used as a testing set. In order to further expand the dataset, we applied random horizontal flipping and random rotating to increase the number of training images to 10 times, which improved the rotation invariance and scale invariance of the algorithm.

4.2 Implementation details

After balancing the network depth and the real-time performance, the standard FSRN used 8 information distillation groups, each of which contained 6 DWRBs. During training, the training images were randomly cropped into several 64×64 images as input, and the batch size was set to 16. We trained the model using the Adam optimizer and perceptual loss function with 4-pixel boundary clipping. All convolutional layers adopted the padding mode of replication. The initial learning rate was 3×10^{-4}, and then halved the learning rate at $2 \times 10^{5}th$, $3.5 \times 10^{5}th$, $4.5 \times 10^{5}th$ and $4.75 \times 10^{5}th$ iterations. The training set was trained for a total of 500000 iterations. PSNR and SSIM were used as evaluation indicators in the experiment. And the Average Inference Time(AIT) of a single image across the three datasets is used to measure the real-time performance of each model.

4.3 Ablation experiment

In this subsection, we first conducted ablation experiments to explore the optimal attention module insertion mode suitable for FSRN. We selected three classic attention modules, including SE [27], ECA [32], and CBAM [33], and tested the effects of five different insertion modes on FSRN, including Att-b, Att-s, Att-b+res, Att-b+Att-s, Att-b+res+Att-s. In this part, we adjusted the FSRN to include 6 information distillation groups, each containing 4 common convolutional residual blocks with kernel size of 3×3 instead of the DWRB. The results are shown in Table 1. Then we performed ablation experiments to verify the effectiveness of the improvements proposed in this paper, including DWRB, Improved SE module with global maximum pooling information and perceptual loss function with boundary clipping. In this part, we used the standard FSRN with the optimal attention module insertion mode as the baseline. The results are shown in Table 2.

Optimal attention module insertion mode. It can be concluded from Table 1 that after adding attention modules to the FSRN, the model's performance is greatly improved. For most attention modules, the Att-s insertion mode can produce more superior performance. It is worth noting that the use of Att-b or Att-b+res on the basis of Att-s will lead to the degradation of network performance, suggesting that inserting attention modules into the feature extraction backbone negatively impacts feature transmission. The ideal mode to insert an attention module, in our opinion, is Att-s.

Effectiveness of DWRB. We compare the results of the DWRB and the traditional convolutional residual block with kernel size of 3×3 .We find that using DWRB only brings a slight improvement in PSNR and SSIM. But the network parameters based on DW are only 1.32 M , which is significantly smaller than the network based on common convolutional residual blocks, which are 2.14M.

Table 1 Comparison between results of different attention module insertion modes.

Attention	Att-b	Att-b+res	Att-s	PSNR/SSIM
None				30.9615/0.8547
SE	√			31.29/0.8559
		√		31.33/0.8568
			√	31.35/0.8576
	√		√	31.35/0.8572
		√	√	31.35/0.8574
ECA	√			31.31/0.8564
		√		31.29/0.8560
			√	31.36/0.8577
	√		√	31.36/0.8579
		√	√	30.92/0.8500
CBAM	√			31.31/0.8564
		√		31.35/0.8573
			√	31.36/0.8578
	√		√	31.29/0.8561
		√	√	31.35/0.8576

Note: None represents not using attention modules in our network. The optimal result is marked in red, and the suboptimal result is marked in blue.

Effectiveness of Improved SENet. It is concluded from experiments that compared to only using SENet, combining global average pooling information and global maximum pooling information bring a significant improvement in PSNR and SSIM in all the three datasets without increasing the number of parameters.

Effectiveness of perpetual loss with boundary clipping. In addition, we find that compared to L_1 loss function without boundary clipping, using the perceptual loss function with boundary clipping can further improve the final performance of the model, especially in the UFO-120 and GBR-100 datasets. Therefore, we believe that the problem of image boundary feature extraction will seriously affect the generalization ability of the model.

Table 2 Comparison between results of different functional modules (PSNR/SSIM).

Module or Method			USR-248	UFO-120	GBR-100
DW	MP	ECL			
√	√	√	**31.42/0.8594**	**27.39/0.7808**	**38.92/0.9394**
	√	√	31.39/0.8586	27.35/0.7819	38.90/0.9392
√		√	31.35/0.8577	27.23/0.7790	38.80/0.9387
√	√		31.36/0.8577	26.91/0.7780	38.71/0.9386

Note: DW, MP, and ECL respectively represent depthwise separable convolutional residual block, SE module with global maximum pooling information and perceptual loss function with boundary clipping. The baseline has been bolded. The optimal result is marked in red, and the suboptimal result is marked in blue.

4.4 Comparison with other algorithms

In order to verify the effectiveness and high performance of the FSRN in underwater optical image super-resolution tasks, we compared our method with other state-of-the-art super-resolution frameworks, including eight general super-resolution methods and three underwater super-resolution methods, when the magnification is 2 times and 4 times. All the super-resolution frameworks were trained for 500,000 iterations in the training sets of USR-248 and UFO-120. The results are in Table 3.

Table 3 Comparison between different algorithms in underwater datasets (PSNR/SSIM).

Model	Params(M)	USR-248	UFO-120	GBR-100	AIT(ms)
Upscale: × 2					
EDSR[6]	40.73	31.35/0.8607	27.28/0.7810	38.59/0.9390	268.57
LapSRN[12]	0.44	28.75/0.8327	25.63/0.7390	33.72/0.9269	112.41
MemNet[34]	2.91	30.17/0.8335	26.34/0.7404	41.67/0.9690	89.85
RDN[10]	22.12	31.05/0.8557	27.13/0.7767	38.22/0.9370	116.67
RCAN[7]	15.44	31.42/0.8590	27.50/0.7823	38.90/0.9394	131.67
IMDN[14]	0.69	31.22/0.8552	27.09/0.7755	38.69/0.9379	123.26
HAN[11]	15.92	31.39/0.8580	26.97/0.7787	38.81/0.9389	447.14
SwinIR[35]	11.68	30.64/0.8480	26.86/0.7654	37.58/0.9335	441.89
SRDRM[18]	3.52	31.40/0.8588	27.33/0.7819	38.65/0.9374	145.77
SRDRM-GAN[18]	3.52	30.62/0.8473	26.85/0.7650	37.64/0.9341	144.36
Deep SESR[19]	10.05	29.93/0.8368	28.57/0.8014	37.62/0.9338	129.24
FSRN	**1.32**	**31.42/0.8594**	**27.39/0.7808**	**38.92/0.9394**	**94.85**
Upscale: × 4					
EDSR[6]	43.09	27.70/0.7206	-	34.23/0.8590	147.14
LapSRN[12]	0.87	26.81/0.6869	-	33.53/0.8440	72.23
MemNet[34]	2.91	27.06/0.6908	-	36.44/0.9005	54.72
RDN[10]	22.27	27.68/0.7203	-	34.18/0.8578	88.15
RCAN[7]	15.59	27.70/0.7214	-	34.24/0.8592	80.95
IMDN[14]	0.72	27.63/0.7172	-	34.18/0.8575	61.20
HAN[11]	16.07	27.73/0.7214	-	34.24/0.8593	306.26
SwinIR[35]	11.90	27.67/0.7202	-	34.18/0.8580	299.83
SRDRM[18]	8.10	27.68/0.7205	-	34.22/0.8587	140.63
SRDM-GAN[18]	8.10	27.59/0.7159	-	34.20/0.8582	139.57
Deep SESR[19]	10.05	27.39/0.7134	-	34.17/0.8577	99.82
FSRN	**1.46**	**27.71/0.7214**	**-**	**34.25/0.8593**	**40.73**

Note: Our models have been bolded. The optimal result is marked in red, and the suboptimal result is marked in blue.

Fig.6 shows the results of underwater image super-resolution reconstruction of various models. Compared with the results in Fig.6, EDSR performs well in all the three datasets, but its network parameters exceed 40M which cannot meet real-time requirements. MemNet achieves the highest PSNR and SSIM in GBR-100, but performes poorly in UFO-120 where it even results in severe color deviation. The underwater super-resolution method Deep SESR achieves the highest PSNR and SSIM in UFO-120, but also performed poorly in USR-248 and GBR-100. The generalization performance of MemNet and Deep SESR is poor in underwater environment. In addition, EDSR, LapSRN, RDN, and SwinR all exhibit severe blue-green bias in UFO-120. Our proposed method FSRN reveals an advanced performance in underwater datasets. Especially in USR-248, FSRN achieves the highest PSNR and SSIM with only 1.32M parameters. Compared to deep and large super-resolution networks, FSRN uses less parameter to achieve higher or similar PSNR and SSIM values. Compared to other lightweight networks, PSNR and SSIM are worth a significant premium. Furthermore, it can be seen from Fig.6 that FSRN shows a superior generalization performance in all the three datasets. FSRN can repair the texture and color and enhance the details of underwater images, making the image structure closer to the real condition.

In terms of real-time performance, as shown in Table 3, when the scale factor is ×2, FSRN only takes 94.85ms to infer a LR image, surpassing most other methods. When the scale factor is ×4, FSRN's inference time is only 40.73ms, which is the fastest. These all demonstrate that our model has a huge advantage in real-time performance.

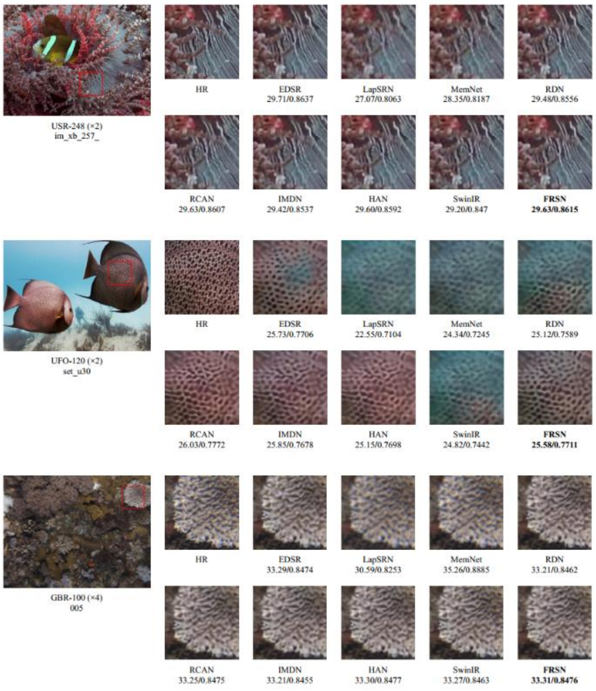

Fig. 6 Subjective comparison between the reconstruction result of advanced image super-resolution models.

5. Discussion

This paper analyzes the challenging problems in underwater image super-resolution reconstruction technology. Firstly, we find that underwater image super-resolution technologies heavily rely on shallow features during reconstruction. However, most current deep learning based super-resolution networks, such as EDSR, only use the last layer of features for reconstruction, ignoring the importance of shallow features, which are not suitable for complex underwater environments. In addition, using channel attention mechanism can indeed significantly improve the performance of super-resolution networks in underwater environments. RCAN using channel attention mechanism performs well in the three datasets, even achieving the highest PSNR and SSIM in UFO-120. However, we also find that inserting channel attention modules into the feature extraction backbone will cause information loss to a certain extent. FSRN separates reconstruction features from each layer, making the final reconstruction information contain rich shallow-level and high-level features, solving the problem of long-distance transmission of shallow features, greatly improving the reconstruction performance. Moreover, FSRN uses the Att-s attention module insertion mode, which does not affect the backbone extracting features, reducing information loss. In summary, FSRN effectively solves the two major challenges of underwater super-resolution reconstruction technology, achieving superior performance with only 1.32M parameters.

6. Conclusion

In this paper, we propose a deep learning based underwater super-resolution framework named FSRN to effectively improve the super-resolution reconstruction performance of underwater optical images. We design a depthwise separable convolutional residual block with large convolutional kernels to extract features, which augments the receptive field of the model, making the model have stronger high-frequency feature extraction ability while saving a lot of parameters. We change the convolutional kernel padding mode and propose a perceptual loss function with boundary clipping to avoid the inconsistent in feature extraction from boundary and non-boundary regions. We propose a multi-branch information distillation group that separates feature extraction information and reconstruction information, solving the problem of long-distance transmission of shallow features in deep neural networks. We improve the SE module combining maxpooling information and explore an optimal attention module insertion mode that reduces feature information loss caused by stacking attention modules. Experiments show that our underwater image super-resolution algorithm has significant improvements compared to other algorithms, achieving high performance with fewer parameters.

References

[1] KEYSR. Cubic convolution interpolation for digital image processing[J]. IEEE Trans-actions on Acoustics, Speech, and Signal Processing,1981, 29(6): 1153-1160.

[2] ZHANG L, WU X. An edge-guided image interpolation algorithm via directional filtering and data fusion[J]. IEEE Transactions on Image Processing, 2006,15(8):2226-2238.

[3] SUN J, XU Z, SHUM H Y. Image super-resolution using gradient profile prior[C]y/2008 IEEE Conference on Computer Vision and Pattern Recognition.2008: 1-8.

[4] DONG C, LOY C C, HE K, et al. Image Super-Resolution Using Deep Convolutional Networks[J]. IEEE T Pattern Anal, 2016,38(2): 295-307.

[5] SHI W, CABALLERO J, HUSZAR F, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network[C]V/2016IEEEConference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 1874-1883.

[6] LIM B, SON S, KIM H, et al. Enhanced Deep Residual Networks for Single Image Super-Resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017: 136-144.

[7] ZHANG Y, LI K, LI K, et al. Image Super-Resolution Using Very Deep Residual Channel Attention Networks[C]/Proceedings of the European Conference on Computer Vision (ECCV).2018: 286-301.

[8] HUIZ, GAO X, YANG Y, et al. Lightweight Image Super-Resolution with Information Multi-distillation Network[C]//Proceedings of the 27th ACM International Conference on Multimedia. Nice France: ACM, 2019: 2024-2032.

[9] KIM J,LEEJK, LEE K M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1646-1654.

[10] ZHANG Y，TIAN Y, KONG Y, et al. Residual Dense Network for Image Super-Resolution[C]/2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT: IEEE, 2018: 2472-2481.

[11] NIU B, WEN W,REN W, et al. Single Image Super-Resolution via a Holistic Attention Network[M]/VEDALDI A, BISCHOF H, BROX T, et al. Computer Vision - ECCV2020: volume 12357.Cham: Springer International Publishing,2020: 191-207.

[12] LAI W S, HUANG J B, AHUJA N, et al. Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution[C]V/Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:624-632.

[13] HUI Z, WANG X, GAO X. Fast and Accurate Single Image Super-Resolution via ln-formation Distillation Network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:723-731.

[14] HUI Z, GAO X, YANG Y, et al. Lightweight Image Super-Resolution with Information Multi-distillation Network[C]//Proceedings of the 27th ACM International Conference on Multimedia. Nice France: ACM, 2019: 2024-2032.

[15] Y. Chen, B. Yang, M. Xia, W. Li, K. Yang, and X. Zhang. Model-based Super-resolution Reconstruction Techniques for Underwater Imaging. In Photonics and Optoelectronics Meetings (POEM): Optoelectronic Sensing and Imaging, volume 8332, page 83320G. International Society for Optics and Photonics, 2012.

[16] E. Quevedo, E. Delory, G. Callico, F. Tobajas, and R. Sarmiento. Underwater Video Enhancement using Multicamera Super-resolution. Optics Communications, 404:94102, 2017.

[17] H. Lu, Y. Li, S. Nakashima, H. Kim, and S. Serikawa. Underwater Image Super-resolution by Descattering and Fusion.IEEE Access, 5:670–679, 2017.

[18] ISLAM M J, SAKIB ENAN S, LUO P, et al. Underwater Image Super-Resolution using Deep Residual Multipliers[C]V/2020 IEEE International Conference on Robotics and Automation (ICRA).2020:

900-906.

[19] ISLAM MJ, LUO P, SATTAR J. Simultaneous Enhancement and Super-Resolution of Underwater Imagery for Improved Visual Perception[Z]. 2020.arXiv: 2002.01155.

[20] WANG H, WU H, HU Q, et al. Underwater image super-resolution using multi-stage information distillation networks[J]. J Vis Commun Image R, 2021,77: 103136.

[21] Cherian, A.K.; Poovammal, E. A Novel AlphaSRGAN for Underwater Image Super Resolution. Comput. Mater. Contin. 2021,69, 1537–1552.

[22] Ren, T.; Xu, H.; Jiang, G.; Yu, M.; Zhang, X.; Wang, B.; Luo, T. Reinforced Swin-Convs Transformer for Simultaneous Underwater Sensing Scene Image Enhancement and Super-Resolution. IEEE Trans. Geosci. Remote Sens. 2022, 60, 4209616.

[23] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]/2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016:770-778.

[24] SHI W, CABALLEROJ, HUSZARF, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network[C]/2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA:IEEE,2016: 1874-1883.

[25] CHOLLET F. Xception: Deep Learning With Depthwise Separable Convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2017: 1251-1258.

[26] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks[C]/2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT: IEEE,2018:4510-4520.

[27] HU J, SHEN L, SUN G. Squeeze-and-Excitation Networks[C]/Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2018: 7132-7141.

[28] LEDIG C, THEIS L, HUSZAR F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[C]V/2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, 2017: 105-114.

[29] WANG X, YUK, WU S, et al. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks[C]//Proceedings of the European Conference on Computer Vision (ECCV) Workshops. 2018: 0-0.

[30] WANG X, XIE L, DONG C, et al. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data: arXiv:2107.10833[M]. arXiv, 2021.

[31] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]/NAVAB N, HORNEGGER J, WELLS W M, et al. Lecture Notes in Computer Science: Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015.Cham: Springer International Publishing,2015:234-241.

[32] WANG Q, et al .ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C]//2020IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Seattle, WA, USA: IEEE,2020: 11531-11539.

[33] WOO S, PARK J,LEE J Y, et al. CBAM: Convolutional Block Attention Module[C]V/Proceedings of the European Conference on Computer Vision (ECCV).2018:3-19.

[34] Tai Y, Yang J, Liu X, et al. Memnet: A persistent memory network for image restoration[C]//Proceedings of the IEEE international conference on computer vision. 2017: 4539-4547.

[35] LIANG J, CAO J, SUN G, et al. SwinIR: Image Restoration Using Swin Transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 1833-1844.

Lan Xie received the B.S degree in Computer and Application from Beijing University of Technology in 1996. She is currently employed by Systems Engineering Research Institute of CSSC.

Qiang Wang received the B.S. and M.S. degrees in Mechanical Engineering from Xian Jiaotong University in 2001 and 2004. He is currently employed by Systems Engineering Research Institute of CSSC.

Yongqiang Ji received the Ph.D degree from Institute of Acoustics, Chinese Academy of Sciences in 2015. He is currently employed by Systems Engineering Research Institution of CSSC.

Yu Gu received the B.S. and M.S. degrees in Control Science And Engineering from Harbin Engineering University in 2010 and 2013. He is currently employed by Systems Engineering Research Institution of CSSC.

Gaozheng Xu received the B.S. degree in Beijing Institute of Technology in 2010. He is currently employed by Systems Engineering Research Institution of CSSC.

Zheng Zhu received the B.S. degree in Beihang University in 2014. He is employed by Systems Engineering Research Institution of CSSC from 2014 to now.

Yuxing Wang received the B.S. degree in Science from Shaanxi Normal University in 2006 and the M.S. degree in Computer System Architecture from Donghua University in 2010. After graduation, he has been engaged in technical research on information systems. He works now with Systems Engineering Research Institute of CSSC.

Yuwei Li recieved offshore and ship engineering master degree from Technique University Harburg Hamburg in 2016. During 2016-2019, he worked as ship structure engineer at Chinese Ship Design and Research Center. Now he works in CSSC Systems Engineering Research Institute.