The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-4/W14-2025
9th International Workshop on Dynamic and Multi-dimensional GIS (DMGIS 2025), 22–24 August 2025, Beijing, China

Remote Sensing Image Super-Resolution Using Feature Grouped Multi-Scale Network

Wei-Tao Zhang[1,*], Nuo Xu[1], Yi-Bo Dang[1]

[1]School of Information Mechanics and Sensing Engineering, Xidian University, 710071, Xi'an, China
zhwt-work@foxmail.com, timxu219@gmail.com, dangyibo1111@163.com

Keywords: Image Super-Resolution, Multi-Scale Feature Extraction, Arbitrary-Scale Upsampling.

Abstract

With the rapid development of deep learning technology, remote sensing image super-resolution has made remarkable progress. Remote sensing images usually contain multiple objects with different scales, making it crucial to adopt multi-scale feature extraction methods. However, the existing multi-scale modules introduce a large number of parameters due to performing convolution operations with different kernel sizes on the input feature maps separately. To address this issue, this paper proposes a Grouped Multi-scale Feature Extraction (GMFE) module. By applying grouped convolutions along the depth dimension of the input feature maps, the number of parameters is effectively reduced. On this basis, we design a Feature-grouped Multi-scale Super-Resolution (FMSR) network. We propose an Edge Enhancement (EE) module integrated into the network to sharpen edges and enhance the visual quality of the image. Additionally, we introduce an arbitrary scale upsampling module, enabling a single trained model to perform image super-resolution reconstruction at arbitrary scales. Extensive experiments on the UC Merced and RSSCN7 datasets demonstrate that the proposed FMSR network achieves superior performance in both quantitative metrics and visual quality.

1. Introduction

With the rapid development of modern aerospace technology, remote sensing has become a crucial means of acquiring information about the Earth's surface and is widely applied in various fields such as land cover classification, agricultural monitoring, geological surveying, and military reconnaissance (Li et al., 2022, Aburaed et al., 2023, Hong et al., 2021, Li et al., 2021). However, due to factors such as cost, sensor limitations, and acquisition angles, traditional remote sensing images typically exhibit low spatial resolution, resulting in the loss of image details and posing challenges for image interpretation and analysis. Therefore, enhancing the resolution of remote sensing images and restoring more detailed information has become a critical research topic. To address these challenges, remote sensing image super-resolution (RSSR) technology has emerged. Among the RSSR methods, single image super-resolution (SISR) has become the preferred approach for many image enhancement and reconstruction tasks due to its relatively low computational complexity and simple input requirements.

Existing SISR techniques can be broadly categorized into three approaches: interpolation-based methods (Jo and Kim, 2021, Loghmani et al., 2018, Cherifi et al., 2020), reconstruction-based methods (Huang et al., 2018, Yang et al., 2018), and convolutional neural network (CNN)-based methods. Interpolation-based methods, such as nearest-neighbor interpolation and bicubic interpolation, are computationally simple and efficient. However, they fail to effectively recover image details and high-frequency information, limiting their ability to enhance image quality. Reconstruction-based methods, on the other hand, utilize techniques such as sparse representation and dictionary learning to formulate optimization problems for restoring image structures. Although these methods can better preserve image details, they are computationally intensive. With the rapid development of deep learning technology, CNN-based methods have become the dominant approach in the SISR field in recent years. By training on large datasets of low-resolution and corresponding high-resolution image pairs, these methods can automatically learn the mapping relationship between the two domains. In the field of computer vision, Dong et al.(Dong et al., 2015) first proposed Super-Resolution Convolutional Neural Network (SRCNN), which achieved end-to-end learning for mapping low-resolution images to high-resolution outputs. Based on SRCNN, Kim et al.(Kim et al., 2016) introduced Very Deep Super-Resolution Network (VDSR), which improved image reconstruction quality by increasing network depth and incorporating residual learning. Subsequently, Lim et al.(Lim et al., 2017) proposed Enhanced Deep Residual Networks (EDSR), which removed batch normalization layers and adopted a deeper network architecture to further enhance performance.

Although SISR techniques have made significant progress in natural image super-resolution, remote sensing imagery presents new challenges for this field. In recent years, various CNN models have been proposed to address the super-resolution of remote sensing images. Ma et al. (Ma et al., 2019) introduced the WTCRR, which combines wavelet transform and recursive Res-Net to reconstruct HR images of different frequency bands. Dong et al. (Dong et al., 2020) proposed the enhanced back-projection network (EBPN), which captures feature differences between channels by incorporating an attention mechanism. Wang et al. (Wang et al., 2022) developed the lightweight feature enhancement network (FeNet), which improves the network's representational capacity by designing lightweight lattice blocks as nonlinear feature extraction functions. Liu et al. (Liu et al., 2024) introduced SRDNet, a dual-domain network with hybrid convolution and progressive upsampling to exploit the spatial-spectral and frequency inform-

* Corresponding author: Wei-Tao Zhang

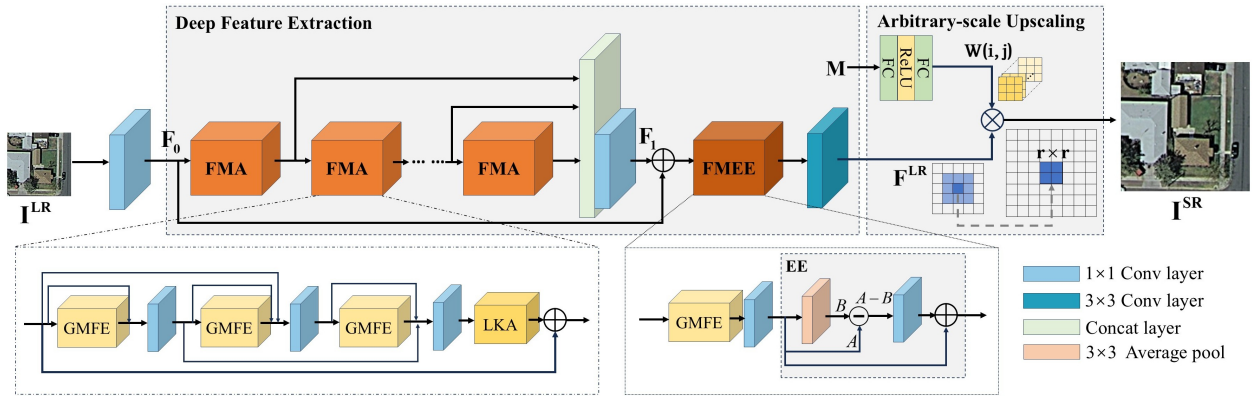

Figure 1. The architecture of proposed FMSR model.

ation of hyperspectral data. Despite the promising performance of existing SISR networks, many rely on multi-scale feature extraction to capture objects of different sizes, often at the cost of increased network parameters. This highlights the need for more efficient multi-scale extraction strategies. Moreover, current RSSR methods are generally restricted to fixed-scale super-resolution. Some employ multiple output branches for specific scaling factors, but this design introduces redundant upsampling modules, leading to higher computational costs and reduced efficiency. Therefore, it is essential to develop a unified and flexible arbitrary-scale upsampling module to enhance both the adaptability and efficiency of super-resolution networks.

To address the aforementioned issues, this paper adopts the concept of group convolution and proposes two Grouped Multi-Scale Feature Extraction modules (GMFE and GMFE+). Based on these modules, we further design the Feature Grouped Multi-Scale Attention module (FMA) and the Feature Grouped Multi-Scale Edge Enhancement module (FMEE). Finally, an arbitrary-scale upsampling module is introduced at the end of the network to reconstruct the feature maps. The main contributions of this paper can be summarized as follows:

1. To reduce the number of parameters while maintaining the feature extraction capability, this paper proposes the GMFE and GMFE+ modules based on the concept of group convolution. These modules significantly reduce the number of parameters compared to conventional feature extraction modules.

2. The clarity of an image is influenced by the sharpness of its edges. This paper further designs the FMEE module based on the GMFE module. By incorporating an Edge Enhancement (EE) mechanism, the FMEE module improves the clarity of image edges, particularly in scenarios with edge blurring and heavy noise.

3. An arbitrary-scale upsampling module is introduced to process remote sensing images, enabling a single trained model to perform image super-resolution reconstruction at arbitrary scales. This approach eliminates the need to train separate networks for different upsampling factors, thereby improving the generalizability and flexibility of the model.

2. Proposed Method

2.1 Network Architecture

As shown in Figure 1, the proposed FMSR network can be divided into two main components: deep feature extraction and arbitrary-scale upscaling. In the deep feature extraction section, we employ the FMA and FMEE modules to replace traditional feature extraction module. Let $\mathrm{Conv}(n_c, k, n_f)$ represent the convolution operation, where n_c, k and n_f denote the input channels, kernel size, and output channels, respectively. The operation $\mathrm{Conv}(c, 3, n)$ is used to extract the shallow features $\mathbf{F_0}$ from the input image $\mathbf{I^{LR}} \in \mathbb{R}^{c\times h\times w}$, c represents the number of channels in the low-resolution image, while h and w denote the height and width of the low-resolution image, respectively. Shallow feature $\mathbf{F_0}$ can be expressed as:

$$\mathbf{F_0} = \mathrm{W}^{(c,3,n)} \star \mathbf{I^{LR}} \tag{1}$$

Here, $\mathrm{W}^{(c,3,n)}$ represents the weight matrix consisting of n convolution kernels of size 3×3, and $\star$ denotes the convolution operation.

Subsequently, the shallow features $\mathbf{F_0}$ are fed into the deep feature extraction section. Let $f_{\mathrm{FMA}}(\cdot)$ and $f_{\mathrm{FMEE}}(\cdot)$ denote the computational processes of the FMA and FMEE modules, respectively, which will be described in detail in Section 2.2 and 2.3. The computation process of this section can be formulated as follows:

$$\mathbf{F_1} = \mathrm{W}^{(n,1,n)} \star \mathrm{cat}_{\mathrm{depth}}\left(f_{\mathrm{FMA}_1}(F_0), ..., f_{\mathrm{FMA}_{\mathrm{K..1}}}(F_0)\right) \tag{2}$$

$$\mathbf{F^{LR}} = \mathrm{W}^{(n,3,n)} \star f_{\mathrm{FMEE}}(\mathbf{F_1} + \mathbf{F_0}) \tag{3}$$

Here, K represents the number of FMA modules, $\mathrm{cat}_{\mathrm{depth}}(..., ...)$ represents the concatenation operation along the depth dimension. $\mathbf{F_1}$ represents the feature map extracted after passing through K FMA modules, while $\mathbf{F^{LR}}$ represents the deep features enhanced by the FMEE module.

Finally, they are fed into the arbitrary-scale upsampling module, denoted by $f_{\mathrm{upscale}}(\cdot)$, to generate super-resolution images at any desired scaling factor. The computation process of this section can be formulated as follows:

$$\mathbf{I}^{\mathbf{SR}} = f_{\text{upscale}}(\mathbf{M}, \mathbf{F}^{\mathbf{LR}}) \tag{4}$$

where $\mathbf{M}$ represents the location projection matrix, and $\mathbf{I}^{\mathbf{SR}}$ denotes the output high-resolution image. The computational process of the above module will be described in detail in Section 2.4.

2.2 Grouped Multi-Scale Feature Extraction

Figure 2 illustrates the structures of the traditional multi-scale feature extraction module, the GMFE module, and the GMFE+ module. The traditional multi-scale feature extraction module applies several sets of convolution kernels with different sizes to the input feature map to capture multi-scale features, which results in a significant increase in the number of parameters. Inspired by the concept of group convolution, this paper proposes the GMFE and GMFE+ modules. Specifically, in the GMFE module, the input feature map $\mathbf{F_i}^{\mathbf{GMFE}}$ is first evenly divided into four groups along the depth dimension, denoted as $\{\mathbf{F_{ik}}^{\mathbf{GMFE}}\}, k = 1, 2, 3, 4$. Each group of feature maps is then processed by convolution operations with different kernel sizes $\{\text{Conv}(n, i, n)\}, i = 1, 3, 5, 7$. Finally, the four output feature maps are concatenated along the depth dimension to produce a feature map with the same shape as the input. This feature extraction operation can be expressed as:

$$\mathbf{F_{out}}^{\mathbf{GMFE}} = \text{cat}_{\text{depth}}(\text{W}^{(n,2k-1,n)} \star \mathbf{F_{ik}}^{\mathbf{GMFE}}), k = 1, 2, 3, 4 \tag{5}$$

where

$$\{\mathbf{F_{i1}}^{\mathbf{GMFE}}, \mathbf{F_{i2}}^{\mathbf{GMFE}}, \mathbf{F_{i3}}^{\mathbf{GMFE}}, \mathbf{F_{i4}}^{\mathbf{GMFE}}\} = \text{split}_{\text{depth}}(\mathbf{F_i}^{\mathbf{GMFE}}) \tag{6}$$

Here, $\text{split}_{\text{depth}}(..., ...)$ represents the operation of evenly splitting along the depth dimension. This operation utilizes group convolution to divide the input channels, which not only preserves the multi-scale feature extraction capability but also significantly reduces the number of parameters.

GMFE+ is an enhanced version of the GMFE module with significantly improved feature extraction capabilities. Compared to GMFE, the output feature maps of GMFE+ retain the feature outputs obtained through sequential processing with convolution kernels of different scales. The detailed process is as follows: First, the input feature map $\mathbf{F_i}^{\mathbf{GMFE+}}$ undergoes a 3×3 convolution operation. The resulting feature map $\mathbf{F_{i1}}^{\mathbf{GMFE+}}$ is then evenly divided into two groups along the depth dimension, referred to as $\mathbf{F_{i11}}^{\mathbf{GMFE+}}$ and $\mathbf{F_{i12}}^{\mathbf{GMFE+}}$. $\mathbf{F_{i12}}^{\mathbf{GMFE+}}$ is processed by a 5×5 convolution operation, while $\mathbf{F_{i11}}^{\mathbf{GMFE+}}$ remains unchanged. Next, the output feature map $\mathbf{F_{i2}}^{\mathbf{GMFE+}}$ from the 5×5 convolution is further divided into two parts along the depth dimension, referred to as $\mathbf{F_{i21}}^{\mathbf{GMFE+}}$ and $\mathbf{F_{i22}}^{\mathbf{GMFE+}}$. $\mathbf{F_{i22}}^{\mathbf{GMFE+}}$ undergoes a 7×7 convolution operation, while $\mathbf{F_{i21}}^{\mathbf{GMFE+}}$ remains unchanged. Finally, the output from the 5×5 convolution, the output from the 7×7 convolution, and the two unmodified feature groups are concatenated along the depth dimension to produce a feature map with the same shape as the input. The output feature map is denoted as $\mathbf{F_{out}}^{\mathbf{GMFE+}}$.

2.3 Attention and Edge Enhancement Modules

The FMA and FMEE modules are both designed based on the GMFE architecture. Inspired by (Guo et al., 2023), the FMA module utilizes a large-kernel attention mechanism to capture long-range dependencies, thereby enhancing the network's ability to perceive and process image details. Traditional large-kernel convolutions significantly increase the network's parameter count, which not only leads to higher computational costs but may also result in performance bottlenecks. To mitigate the increase in computation and parameters, we adopt a convolutional decomposition strategy to construct the LKA module in FMA. Specifically, traditional large-kernel convolutions can be decomposed into three parts: first, spatial local convolutions (depthwise convolutions) handle the local detail information in the image; second, spatial long-range convolutions (dilated convolutions) capture relationships between distant pixels by increasing the span of the convolution kernels; and finally, channel convolutions (1×1 convolutions) are used to integrate information across the channel dimension. Let $\mathbf{F_{in}}^{\mathbf{LKA}}$ and $\mathbf{F_{out}}^{\mathbf{LKA}}$ denote the input and output of the LKA module, respectively. The computational process of the module can be expressed as:

$$\mathbf{F_{out}}^{\mathbf{LKA}} = \mathbf{X_{attention}} \otimes \mathbf{F_{in}}^{\mathbf{LKA}} \tag{7}$$

where

$$\mathbf{X_{attention}} = \text{W}^{(n,1,n)} \star (\text{Conv}_{\text{DW-D}}(\text{Conv}_{\text{DW}}(\mathbf{F_{in}}^{\mathbf{LKA}}))) \tag{8}$$

Here, $\text{Conv}_{\text{DW-D}}(\cdot)$ represents the depthwise dilated convolution, $\text{Conv}_{\text{DW}}(\cdot)$ refers to the depthwise convolution, and $\mathbf{X_{attention}}$ indicates the attention map, where the values in the attention map represent the importance of each feature. The symbol $\otimes$ represents element-wise product.

The FMEE module incorporates an Edge Extraction (EE) module to enhance the edge and texture information in the feature map. Specifically, the input feature map $\mathbf{F_{in}}^{\mathbf{EE}}$ is first subtracted by a smoothed version of the feature map to extract the edge information. Then, a 1×1 convolution is applied to fuse the depthwise channel information. Finally, a residual connection is made with the input feature map to obtain the output feature map $\mathbf{F_{out}}^{\mathbf{EE}}$. The EE module can be expressed as follows:

$$\mathbf{F_{out}}^{\mathbf{EE}} = \text{W}^{(n,1,n)} \star (\mathbf{F_{in}}^{\mathbf{EE}} - \text{AveragePool}_{3\times3}(\mathbf{F_{in}}^{\mathbf{EE}})) + \mathbf{F_{in}}^{\mathbf{EE}} \tag{9}$$

2.4 Arbitrary Scale Upsampling Module

The upsampling module needs to find a mapping from the low-resolution feature to the high-resolution image. Specifically, for each low-resolution feature map pixel $\mathbf{F}^{\mathbf{LR}}(i', j')$, $r*r$ specific filter weights $\text{W}(i, j)$ needs to be found to map it to the super-resolved image $\mathbf{I}^{\mathbf{SR}}(i, j)$. r represents the scaling factor, and $r * r$ indicates that each pixel in the low-resolution feature map will be mapped to $r * r$ pixels in the high-resolution image. The computational process of the module can be expressed as follows:

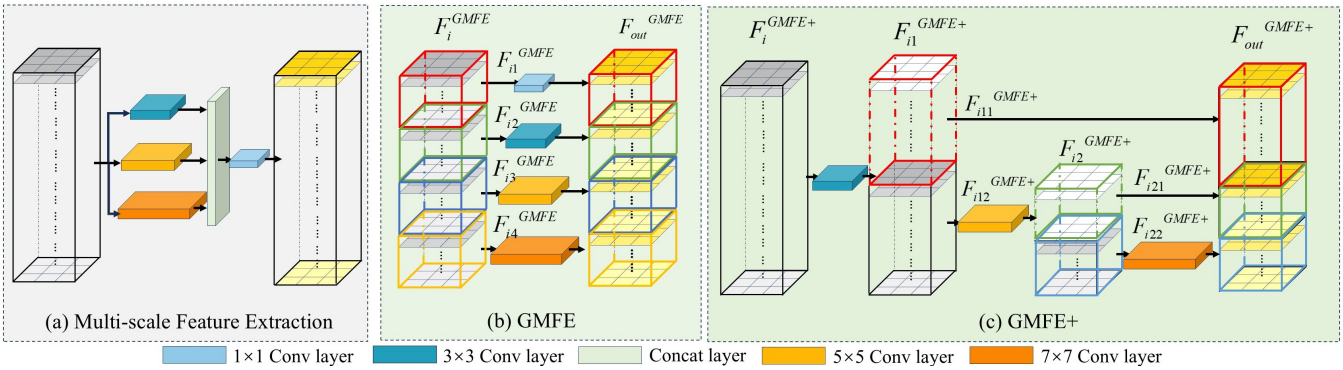

Figure 2. Multi-scale feature extraction (a) Traditional multi-scale feature extraction module. (b) GMFE. (c) GMFE+.

$$\mathbf{I}^{\mathbf{SR}}(i,j) = \Phi(\mathbf{F}^{\mathbf{LR}}(i',j'), \mathrm{W}(i,j)) \tag{10}$$

Here, $\mathrm{W}(i,j)$ denotes the filter weights corresponding to the pixel (i,j) in $\mathbf{I}^{\mathbf{SR}}$. $\Phi(\cdot)$ represents the operation of computing the product and summation of the corresponding elements.

To generate the filter weights $\mathrm{W}(i,j)$, it is necessary to construct a location projection matrix $\mathbf{M}$ as the input to the module. The weight prediction matrix should include the positional projection information between (i',j') and (i,j), as well as the image's scaling factor r. Let $\vec{\mathbf{v}}_{(i,j)}$ represent the vector in the weight prediction matrix. $\mathbf{M}$ can be expressed as follows:

$$\mathbf{M} = \left[\vec{\mathbf{v}}_{(0,0)}, \vec{\mathbf{v}}_{(0,1)}, ..., \vec{\mathbf{v}}_{(i,j)}, ..., \vec{\mathbf{v}}_{(H,W)}\right] \tag{11}$$

where

$$\vec{\mathbf{v}}_{(i,j)} = \left[\frac{i}{r} - \left\lfloor\frac{i}{r}\right\rfloor, \frac{j}{r} - \left\lfloor\frac{j}{r}\right\rfloor, \frac{1}{r}\right]^T \tag{12}$$

Here, $(\frac{i}{r}, \frac{j}{r})$ represents the projection of the pixel index (i,j) in the high-resolution image $\mathbf{F}^{\mathbf{LR}}$ onto the low-resolution feature map $\mathbf{I}^{\mathbf{SR}}$. Since r can be any positive number greater than 1, $\frac{i}{r}$ and $\frac{j}{r}$ may take non-integer values. Therefore, it is necessary to introduce $\frac{i}{r} - \lfloor\frac{i}{r}\rfloor$ and $\frac{j}{r} - \lfloor\frac{j}{r}\rfloor$ to represent the projection offset of the pixel, where $\lfloor\rfloor$ represents the floor operation. Then, the location projection matrix is passed through two fully connected layers to obtain the filter weights $\mathrm{W}(i,j)$, which, when applied to $\mathbf{F}^{\mathbf{LR}}(i',j')$, result in the output super-resolved image $\mathbf{I}^{\mathbf{SR}}$.

3. Experimental Results

3.1 Datasets

To evaluate the effectiveness of the proposed method, we selected the UC Merced and RSSCN7 remote sensing datasets for training. The UC Merced dataset is a high-resolution remote sensing image dataset consisting of 21 land cover classes, including urban, forest, water, agricultural areas, airports, etc. Each class contains 100 images, for a total of 2100 images. The image size is 256×256 pixels. In our experiments, 90 images from each class were selected as the training set and 10 images as the test set. Therefore, the training set consists of 1890 images, and the test set consists of 210 images. The high-resolution images in both the training and test sets have a pixel size of 256×256, with the corresponding low-resolution images obtained by bicubic interpolation downsampling from the high-resolution images.

The RSSCN7 dataset contains remote sensing images from 7 categories, including urban, forest, grassland, farmland, and water bodies. Each category includes 400 images, for a total of 2800 images. The image size is 400×400 pixels. The RSSCN7 dataset provides rich land cover information, making it particularly suitable for multi-classification tasks on high-resolution remote sensing images. In our experiments, 360 images from each class were selected as the training set and 40 images as the test set. Therefore, the training set consists of 2520 images, and the test set consists of 280 images. The high-resolution images in both the training and test sets have a pixel size of 400×400, with the corresponding low-resolution images obtained by bicubic interpolation downsampling from the high-resolution images.

3.2 Experimental Settings

This study primarily focuses on image upscaling tasks with scale factors of ×2 and ×4. The FMSR network is capable of performing image upscaling for different scale factors using the same module, while in other networks, the reconstruction part adjusts the upsampling operation according to the specific scale factor. The batch size for training data is set to 32. To enhance the diversity of training samples, data augmentation techniques are employed, including random horizontal flipping and random rotations (90°, 180°, 270°). During training, the L1 loss function is used as the optimization objective, and the Adam optimizer is utilized for training, with hyperparameters set to $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate is set to 1×10^{-4} , with a learning rate decay strategy to ensure stability and convergence during the training process. To evaluate the performance of the network, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) are used as evaluation metrics. PSNR is used to measure the quality of image reconstruction, with higher values indicating better reconstruction quality; SSIM reflects the structural similarity between images, with values closer to 1 indicating better visual quality. All experiments are implemented and trained using the PyTorch framework on a TITAN RTX GPU.

3.3 Comparison With Other Methods

To demonstrate the superiority of the proposed method, we selected six image super-resolution algorithms for comparison,

Figure 3. Image super-resolution performance of different models
(a) UC Merced dataset. (b) RSSCN7 dataset.

Table 1. PSNR and SSIM of different methods under scaling factors of ×2 and ×4

Methods	UC Merced		RSSCN7	
	×2	×4	×2	×4
Bicubic	31.12 / 0.8983	25.52 / 0.6973	28.71 / 0.7940	25.66 / 0.6145
SRCNN	32.88 / 0.9240	26.28 / 0.7361	29.16 / 0.8042	25.96 / 0.6279
VDSR	33.84 / 0.9360	27.39 / 0.7720	29.34 / 0.8100	26.08 / 0.6368
RDN	34.38 / 0.9404	27.84 / 0.7924	30.05 / 0.8235	26.14 / 0.6433
MHAN	34.42 / 0.9324	27.62 / 0.7736	30.14 / 0.8212	26.20 / 0.6429
Omni-SR	34.32 / 0.9335	27.80 / 0.7652	30.11 / 0.8197	26.16 / 0.6421
FMSR	**34.63 / 0.9412**	**27.91 / 0.7932**	**30.19 / 0.8245**	**26.26 / 0.6437**

including Bicubic, SRCNN(Dong et al., 2015), VDSR(Kim et al., 2016), RDN(Zhang et al., 2018), MHAN(Zhang et al., 2020), and Omni-SR(Wang et al., 2023). Tables 1 present the performance of each algorithm on the UC Merced and RSSCN7 datasets, where each model performs image super-resolution using scaling factors of ×2 and ×4, respectively. Based on the experimental results, FMSR demonstrates superior performance in terms of PSNR and SSIM metrics compared to other super-resolution models. Figures 3 present the visual results of image super-resolution obtained using different models on the UC Merced and RSSCN7 datasets. As shown in the comparison images, FMSR outperforms the other models in restoring image edges and demonstrates significantly less distortion in the recovery of object shapes.

3.4 Ablation Studies

This paper proposes the GMFE and GMFE+ modules, which utilize grouped convolutions to extract features at multiple scales for better handling of complex land cover information and details in remote sensing images. We conducted a series of ablation experiments to validate the effectiveness of each module, and the experimental results are presented in Table 2, Here, FE represents the conventional multi-scale feature extraction module, while GMFE and GMFE+ denote the proposed Grouped Multi-Scale Feature Extraction modules. The results demonstrate that the super-resolution network achieves the best performance when the GMFE+, LKA, and EE modules are used simultaneously. When the input channel size is 64, the parameter counts of model using GMFE and GMFE+ are approximately reduced by 311k and 248k, respectively, compared to the conventional multi-scale feature extraction module, representing only 6.74% and 25.48% of the total parameter size.

Table 2. Comparison of the Super-Resolution Performance of the Network Using Different Blocks

FE	GMFE	GMFE+	LKA	EE	PSNR/SSIM	Param
✓	×	×	✓	✓	27.81 / 0.7873	333.6k
×	✓	×	✓	✓	27.78 / 0.7881	22.5k
×	×	✓	✓	✓	**27.91 / 0.7932**	85k
×	×	✓	✓	×	27.82 / 0.7883	83k
×	×	✓	×	✓	27.75 / 0.7876	77k

4. Conclusion

This paper presents a Feature-grouped Multi-scale Super-Resolution (FMSR) network built upon the proposed Grouped Multi-scale Feature Extraction (GMFE) module, which leverages grouped convolution to reduce parameters while preserving feature extraction capacity. The network integrates a Large Kernel Attention (LKA) mechanism and an Edge Enhancement (EE) module to further boost performance. An arbitrary-scale upsampling module is also introduced, allowing flexible super-resolution at any scale with a single model. Extensive experiments on the UC Merced and RSSCN7 datasets demonstrate that FMSR achieves superior quantitative and visual results, confirming its effectiveness and generalizability for remote sensing image super-resolution.

References

Aburaed, N., Alkhatib, M. Q., Marshall, S., Zabalza, J., Al Ahmad, H., 2023. A review of spatial enhancement of hyperspectral remote sensing imaging techniques. *IEEE J Sel Top Appl Earth Obs Remote Sens.*, 16, 2275–2300.

Cherifi, T., Hamami-Metiche, L., Kerrouchi, S., 2020. Comparative study between super-resolution based on polynomial interpolations and whittaker filtering interpolations. *Proc. 2020 1st Int. Conf. Commun., Control Syst. Signal Process*, IEEE, 235–241.

Dong, C., Loy, C. C., He, K., Tang, X., 2015. Image super-resolution using deep convolutional networks. *IEEE Trans Pattern Anal Mach Intell.*, 38(2), 295–307.

Dong, X., Xi, Z., Sun, X., Yang, L., 2020. Remote sensing image super-resolution via enhanced back-projection networks. *Proc. IGARSS 2020 - IEEE Int. Geosci. Remote Sens. Symp.*, IEEE, 1480–1483.

Guo, M.-H., Lu, C.-Z., Liu, Z.-N., Cheng, M.-M., Hu, S.-M., 2023. Visual attention network. *Comput. Vis. Media*, 9(4), 733–752.

Hong, D., Han, Z., Yao, J., Gao, L., Zhang, B., Plaza, A., Chanussot, J., 2021. SpectralFormer: Rethinking hyperspectral

image classification with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–15.

Huang, Y., Li, J., Gao, X., He, L., Lu, W., 2018. Single image super-resolution via multiple mixture prior models. *IEEE Trans Image Process*, 27(12), 5904–5917.

Jo, Y., Kim, S. J., 2021. Practical single-image super-resolution using look-up table. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 691–700.

Kim, J., Lee, J. K., Lee, K. M., 2016. Accurate image super-resolution using very deep convolutional networks. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 1646–1654.

Li, J., Zhang, Z., Tian, Y., Xu, Y., Wen, Y., Wang, S., 2021. Target-guided feature super-resolution for vehicle detection in remote sensing images. *IEEE geoscience and remote sensing letters*, 19, 1–5.

Li, R., Zheng, S., Duan, C., Wang, L., Zhang, C., 2022. Land cover classification from remote sensing images based on multi-scale fully convolutional network. *Geo-spatial Inf. Sci.*, 25(2), 278–294.

Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K., 2017. Enhanced deep residual networks for single image super-resolution. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 136–144.

Liu, T., Liu, Y., Zhang, C., Yuan, L., Sui, X., Chen, Q., 2024. Hyperspectral image super-resolution via dual-domain network based on hybrid convolution. *IEEE Trans Geosci Remote Sens.*

Loghmani, G. B., Zaini, A. M. E., Latif, A. M., 2018. Image zooming using barycentric rational interpolation. *J. Math. Ext.*, 12(1), 67–86.

Ma, W., Pan, Z., Guo, J., Lei, B., 2019. Achieving super-resolution remote sensing images via the wavelet transform combined with the recursive res-net. *IEEE Trans Geosci Remote Sens.*, 57(6), 3512–3527.

Wang, H., Chen, X., Ni, B., Liu, Y., Liu, J., 2023. Omni aggregation networks for lightweight image super-resolution. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 22378–22387.

Wang, Z., Li, L., Xue, Y., Jiang, C., Wang, J., Sun, K., Ma, H., 2022. FeNet: Feature enhancement network for lightweight remote-sensing image super-resolution. *IEEE Trans Geosci Remote Sens.*, 60, 1–12.

Yang, Q., Zhang, Y., Zhao, T., Chen, Y., 2018. Single image super-resolution using self-optimizing mask via fractional-order gradient interpolation and reconstruction. *ISA Trans.*, 82, 163–171.

Zhang, D., Shao, J., Li, X., Shen, H. T., 2020. Remote sensing image super-resolution via mixed high-order attention network. *IEEE Trans Geosci Remote Sens.*, 59(6), 5183–5196.

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y., 2018. Residual dense network for image super-resolution. *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2472–2481.