An Enhanced Global Feature-Guided Network Based on Multiple Filtering Noise Reduction for Remote Sensing Image Compression

Cuiping Shi, *Member, IEEE*, Kaijie Shi, Fei Zhu, Zexin Zeng, Mengxiang Ding, and Zhan Jin

***Abstract*— Remote sensing images obtained at high altitudes often contain complete object or scene information, which makes their global visual features richer compared to natural images. In order to enhance the scope and multilevel characteristics of global visual features of remote sensing images, this article proposes an enhanced global feature-guided network based on multiple filtering noise reduction (GFRNet) for remote sensing image compression. First, a pyramid vision transformer (PVT) is introduced into remote sensing image compression for the first time. Based on this, a PVT compression branch (PVTCB) is designed, which can capture multilevel global visual features through a three-stage pyramid transformer module for image compression (TPTC) and utilizes filters to accurately control the output of TPTC. Second, a quadruple-filtered multicore noise reduction attention module (QFMR-AM) is constructed in the four-stage compression branch (FSCB) for denoising and enhancing multilevel features. Finally, a global visual feature guidance module (GVGM) is designed between FSCB and the four-stage reconstruction decoder (FSRD). By calculating the global visual feature loss Loss_{GVF} through GVGM, a novel rate-distortion $\text{Loss}_{\text{Total}}$ is constructed, making the network more focused on extracting global information. Experimental results show that compared with some advanced methods, the proposed GFRNet achieves better compression performance on multiple evaluation indicators. In addition, the reconstructed images obtained by the proposed GFRNet can provide better classification performance, which further proves that the proposed method helps to preserve more important features of remote sensing images during the compression process.**

***Index Terms*— Multihead self-attention, noise reduction, pyramid transformer, rate-distortion optimization, remote sensing image compression.**

I. Introduction

REMOTE sensing images have many unique land features, such as land cover, topography, landform, and temperature, which are usually not exhibited in natural images [1], [2]. Therefore, remote sensing images have been widely used in many fields, such as environmental monitoring, meteorology, and geological science [3], [4], [5], [6]. However, remote sensing images are usually captured by satellites at high altitude through the atmosphere, which inevitably has more background noise [7]. In addition, remote sensing images often contain complete object or scene information due to high-altitude shooting. Compared with natural images, the global visual features are richer [8]. This information has an important impact on the compression of texture features. Second, with the rapid development of remote sensing technology, the spatial and spectral resolution of remote sensing images continue to improve, and the amount of data also increase dramatically [9]. For these reasons, a specialized compression method suitable for the characteristics of remote sensing images is urgently needed.

At present, traditional image compression methods have achieved some results [10], [11]. For example, Báscones et al. [12] proposed a method to compress hyperspectral image data by combining principal component analysis and JPEG2000. The classical joint photographic experts group (JPEG) [13] and JPEG2000 [14] are mainly composed of three parts: image transformation, quantization, and entropy encoding. First, the image is transformed and dequantized. Then, important information is retained through quantification; Finally, the entropy coding is used to compress the correlation coefficient. In addition, better portable graphics (BPG) [15], [16] and WebP [17] with superior performance have been born in the field of image compression. Li et al. [16] used the mean deviation similarity index (MDSI) as an evaluation metric to improve BPG. A two-step compression strategy is used to provide more accurate remote sensing image quality control, which achieves consistency in compression efficiency and image quality [16]. Traditional image compression methods can be classified into vector quantization-based [18], predictive coding-based [19], and transform-based coding algorithms [20]. Qian [21] proposed a fast vector quantization compression algorithm for multispectral images. The core of this method is to replace the input vector with a codeword index that matches the codebook, so as to optimize the efficiency of data transmission and storage [21]. Three-dimensional (3-D)-multiband linear predictor (MBLP) uses predictive technology. First, spatially redundant information in the image is eliminated. Second, the current frequency band is predicted. Finally, the predicted residuals are encoded with the help of an entropy

Received 21 July 2024; revised 7 October 2024; accepted 12 October 2024. Date of current version 7 November 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42271409, in part by the Fundamental Research Funds in Heilongjiang Provincial Universities under Grant 145109145, and in part by the Qiqihar University Graduate Innovative Research Project under Grant QUZLTS_CX2023025. *(Corresponding author: Cuiping Shi.)*

Cuiping Shi is with the College of Information Engineering, Huzhou University, Huzhou 313000, China, and also with the Department of Communication Engineering, Qiqihar University, Qiqihar 161000, China (e-mail: shicuiping@qqhru.edu.cn).

Kaijie Shi, Fei Zhu, Zexin Zeng, Mengxiang Ding, and Zhan Jin are with the Department of Communication Engineering, Qiqihar University, Qiqihar 161000, China (e-mail: 2022910313@qqhru.edu.cn; 2022935750@qqhru.edu.cn; 2022910311@qqhru.edu.cn; 2021910321@qqhru.edu.cn; jinzhan@qqhru.edu.cn).

Digital Object Identifier 10.1109/TGRS.2024.3483871

decoder [22]. In addition, 3-D-set partitioning in hierarchical trees (SPIHT), as a transformation compression method for 3-D images, achieves efficient image compression by applying 3-D wavelet transform in the spatial and spectral domains [23]. However, traditional image compression methods have certain limitations. For example, JPEG, JPEG2000, and BPG are not designed to adjust the characteristics of remote sensing images. Therefore, these methods are limited in the compression performance of remote sensing images. Second, at high compression ratios, most traditional image compression methods exhibit relatively poor rate distortion performance. Finally, for the characteristics of remote sensing images, such as high information entropy, complex background noise, and abundant global information, the common traditional methods cannot make efficient adaptive adjustment. Therefore, these limitations restrict the improvement of the compression performance of remote sensing images.

In search of breakthroughs, researchers focus on deep learning technologies that have become hot in recent years [24], [25]. Classic deep learning-based image compression frameworks mainly include autoencoder (AE) [26], [27] and variational AE (VAE) [28], [29]. Riccardo et al. [26] proposed a deep convolutional AE-based compression network to solve the problem of processing large-scale data volumes generated by complex satellite instruments in the field of space science and satellite imagery. The network has advantages in both compression ratio and spectral signal reconstruction. In addition, it shows good robustness for data types larger than 8 bits [26]. Alves et al. [28] designed a low-complexity VAE to meet the computational resource constraints in satellite image compression. By reducing dimensions and simplifying the entropy model, the encoder reduces complexity while maintaining compression performance [28]. However, compared with AE-based frameworks, VAE-based frameworks have more powerful image reconstruction capabilities. The reason for this is that VAE has a continuous mapping space that AE does not have, and it can reconstruct images with smooth transitions between pixels. In recent years, some VAE-based baseline networks have been developed [30], [31], [32], [33], [34], [35], [36], which have demonstrated superior rate distortion performance compared with traditional image compression methods. These VAE-based image compression networks usually consist of three components: encoder, entropy encoder, and decoder. First, the neural network was used to preliminarily compress the image data block. Then, the compressed pixel data were mapped into a quantized representation. Finally, these data are further compressed into a bitstream form by traditional encoding techniques. In addition, in order to model more accurately, some compression models introduce entropy models, such as Laplace models, single-core gaussian models, hybrid gaussian models, and factorized entropy models into the framework to make full use of prior information [37], [38], [39], [40], [41], [42]. Based on the above theory, some scholars have developed many remote sensing image compression networks based on deep learning and achieved good rate distortion performance [43], [44], [45], [46], [47], [48], [49], [50].

Compared with natural images, remote sensing images contain rich global contextual features due to imaging at high altitudes. However, most of the current methods extract global information by introducing global feature modules into the backbone network. There are two problems with this approach: first, the global information obtained is relatively simple and lacks diversity; second, the scope of global contextual information is limited to the feature extraction process of local networks. Therefore, how to comprehensively and effectively extract global contextual information from remote sensing images, increase the diversity of global contextual features, and expand the scope of global contextual features has become a serious challenge that urgently needs to be solved in the field of remote sensing image compression.

In order to alleviate the above problems, this article proposes an enhanced global feature-guided network based on multiple filtering noise reduction for remote sensing image compression (GFRNet). By improving the multilevel nature of global information and expanding the influence range of global information, the global features in remote sensing images are enhanced, thereby improving the quality of reconstructed images. GFRNet mainly consists of the following parts: a dual-branch compression structure, including a pyramid vision transformer compression branch (PVTCB) and a four-stage compression branch (FSCB); entropy coding; and four-stage reconstruction decoder (FSRD). GFRNet mainly analyzes and optimizes from three aspects. First, remote sensing images are typically captured by satellites in space through complex clouds and atmospheres, inevitably carrying a significant amount of background noise. To address this issue, this article proposes a quadruple-filtered multicore noise reduction attention module (QFMR-AM) for denoising and enhancing multilevel features. Through multiple filters and convolutions at different scales, it can effectively reduce complex background noise. In addition, the efficient extraction, enhancement, and fusion of features at different scales are achieved. Second, the scope of the common global feature extraction module is limited. To solve this problem, this article constructs a global visual feature guidance module (GVGM) between the compressed part and the reconstruction part. The proposed global visual feature loss Loss_{GVF} is calculated by GVGM, and a new rate-distortion $\text{Loss}_{\text{Total}}$ is constructed. This design efficiently applies global context features to the entire network in the form of loss, thus effectively enhancing the quality of the global features. Third, the global features extracted by common methods lack hierarchy. Therefore, this article builds a PVTCB, which captures multilevel global information through the three-stage pyramid transformer module for image compression (TPTC), and uses filters to accurately control the output of TPTC, thereby optimizing the compression effect of remote sensing images. In summary, this article builds a high-performance GFRNet based on the proposed PVTCB, FSCB, FSRD, QFMR-AM, GVGM, and $\text{Loss}_{\text{Total}}$.

This study conducted sufficient experiments. Experimental results show that compared with some advanced compression methods, the proposed GFRNet can provide excellent

compression performance in peak signal-to-noise ratio (PSNR) and multiscale structural similarity index metric (MS-SSIM).

The main contributions of this article are summarized as follows.

1) A QFMR-AM is proposed to achieve noise reduction and multilevel feature enhancement. Through multiple filters and convolution at different scales, the module effectively reduces the complex background noise and efficiently fuses features of different scales.
2) A GVGM is designed, which constructs a novel rate-distortion $\text{Loss}_{\text{Total}}$ of the proposed network by calculating the global visual feature loss Loss_{GVF} between the compression part and the reconstruction part. Through this new loss, global features are efficiently applied to the entire network.
3) The pyramid vision transformer (PVT) was introduced into remote sensing image compression for the first time. Based on this, a PVTCB is constructed, which captures multilevel global information through a TPTC, and utilizes filters to accurately control the output of PVTCB to obtain multilevel global features.
4) This article effectively embeds PVTCB, FSCB, FSRD, QFMR-AM, GVGM, $\text{Loss}_{\text{Total}}$, and factorized entropy model to construct a high-performance enhanced GFRNet for remote sensing image compression. Through a large number of experiments on San Francisco, NWPU-RESISC45 and UC-Merced, the superior compression performance of GFRNet for remote sensing images is fully proved.

The remainder of the study is organized as follows: in Section II, relevant work will be discussed. In Section III, the proposed GFRNet framework and the details of each module are elaborated. In Section IV, this article comprehensively analyzes and compares the proposed GFRNet and other compression methods through a large number of experiments. In Section V, conclusions and future work are discussed.

II. Related Work

There are four main types of deep learning techniques used for remote sensing image compression: image compression methods based on convolutional neural networks (CNNs) [39], [42], [43], [47], [48], image compression methods based on transformer [40], [51], image compression methods based on generative adversarial networks (GANs) [52], [53], and image compression methods based on graph attention (GAT) [54].

A. CNN-Based Methods

In the CNN-based methods, Wang et al. [47] proposed a remote sensing image compression framework using historical images as a reference. Through double-ended reference downsampling encoding technology and correlation embedding, it effectively eliminates time redundancy and spatial redundancy and reduces spurious textures [47]. In addition, Shao et al. [42] proposed a discrete wavelet transform gaussian mixture model (DWTGMM) entropy model based on discrete wavelet transform (DWT) and Gaussian mixture model (GMM) for remote sensing image compression. The model obtains sparse representations through DWT and then uses GMM to model separately to estimate the probability distribution, which effectively improves the compression performance [42].

B. Transformer-Based Methods

In the transformer-based image compression methods, Li et al. [40] proposed a remote sensing image compression method. By distinguishing between objects and background areas and optimizing global and local information encoding, it achieves high-object fidelity compression at low bit rates [40]. In addition, Chuan et al. [51] proposed a remote sensing image compression network based on transformer and CNN, which can effectively reduce local and nonlocal redundancy and achieve efficient compression. The three-stage training strategy improves the generalization ability of the network, and the performance is better than that of traditional algorithms [51].

C. GAN-Based Methods

In the GAN-based image compression methods, Han et al. [52] proposed an edge-guided adversarial network designed to preserve sharp edge and texture information at the same time. It uses edge fidelity constraints to guide the network to optimize image content and structure, thereby solving the problem of local smoothness in the existing methods [52]. In addition, Kan et al. [53] proposed a remote sensing satellite image compression method based on conditional GANs, which improved the quality and detail of the reconstructed images by introducing the Laplacian of Gaussian loss and perception metrics [53].

D. GAT-Based Methods

In the GAT-based image compression methods, Pan et al. [54] proposed a hybrid attention compression network (HACN) for remote sensing image compression. By introducing the residual attention module (RAM) and the lightweight graph attention module (GAM), the network is able to capture spatial and cross-channel long-distance dependencies in the process of feature transformation [54].

Although the above methods achieve good compression performance, the quality of the global features extracted by these methods is relatively weak. Specifically, the influence range of global features is limited, and global features lack multilevel nature. These shortcomings lead to the suboptimal rate distortion performance of these models.

III. Methodology

In this section, the proposed GFRNet, as well as modules, such as QFMR-AM, GVGM, and PVTCB, will be introduced in detail.

A. Overall Framework of the Proposed GFRNet

The proposed GFRNet enhances the global contextual features of remote sensing images from the perspective of increasing the multilevel nature of global features and expanding the scope of global features, thereby improving the quality

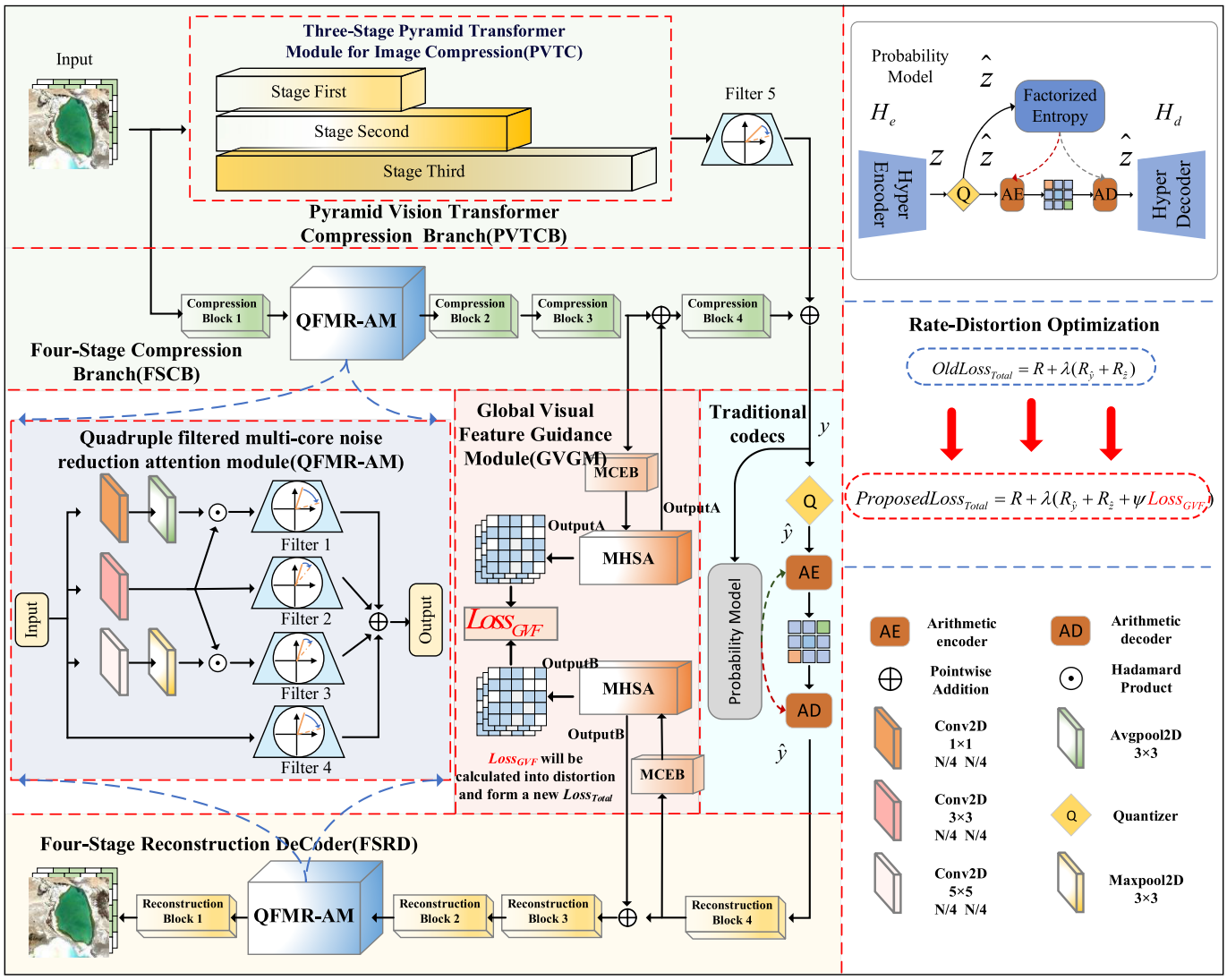


Fig. 1. Overall structure diagram of the proposed GFRNet.

of reconstructed images. It mainly includes QFMR-AM for noise reduction and multiscale feature enhancement, GVGM for efficient application of global features to the entire network in the form of loss, and PVTCB for enhancing the multilevel nature of global features in remote sensing images. QFMR-AM consists of convolutions of different scales and four filters (Filters1–4) for controlling the output. GVGM consists of three submodules, including the multichannel enhancement block (MCEB), the multihead self-attention module (MHSA), and the global visual feature loss Loss_{GVF}. PVTCB consists of two submodules, including TPTC and Filter5. In addition, in this article, Loss_{GVF} is designed to construct a new rate-distortion function $\text{Loss}_{\text{Total}}$, so as to coordinate the work of the backbone network and the proposed module.

Fig. 1 shows the overall structure of GFRNet. In this article, four compression blocks in FSCB and four reconstruction blocks in FSRD are designed by reasonably selecting the size of the convolution kernel and reassigning the number of channels, so as to achieve excellent compression performance at low complexity. Table I lists the parameters of compression blocks 1–4 and reconstruction blocks 1–4. Probability models mainly include hyperprior networks (hyperencoder and hyperdecoder), quantizer (Q), arithmetic encoding (AE), and arithmetic decoding (AD). Table II lists the parameters of hyperencoder and hyperdecoder. In this article, a superprior network is used to learn the probability model on which entropy coding depends. In addition, it is used to generate

TABLE I
SPECIFIC PARAMETERS OF COMPRESSION BLOCK AND RECONSTRUCTION BLOCK

Block	First layer	Second layer
Compression block 1	Conv2D 7×7 3 N/4 2↓	GDN
Compression block 2	Conv2D 3×3 N/4 N/2 2↓	GDN
Compression block 3	Conv2D 3×3 N/2 3N/4 2↓	GDN
Compression block 4	Conv2D 3×3 3N/4 N 2↓	GDN
Reconstruction block 1	Conv2D 3×3 N/4 3 2↑	IGDN
Reconstruction block 2	Conv2D 3×3 N/2 N/4 2↑	IGDN
Reconstruction block 3	Conv2D 3×3 3N/4 N/2 2↑	IGDN
Reconstruction block 4	Conv2D 3×3 N 3N/4 2↑	IGDN

TABLE II
SPECIFIC PARAMETERS OF HYPE RENCODER AND HYPERDECODER

	Hyper encoder	Hyper decoder
Layer1	Conv 3×3 N N 1	Conv 3×3 N N 2↑
Layer2	RELU	RELU
Layer3	Conv 3×3 N N 2↓	Conv 3×3 N N 2↑
Layer4	RELU	RELU
Layer5	Conv 3×3 N N 2↓	Conv 3×3 N N 1
Layer6	-	RELU

the parameters of the entropy model (mean parameter μ_i and scale parameter σ_i^2), which is modeled as conditional gaussian. GVGM is used to calculate the global visual feature loss Loss_{GVF} between the compressed and reconstructed parts, where Loss_{GVF} represents the difference between the global features of the compressed part and the global features of the reconstructed part. The smaller the loss value, the smaller the difference between the compressed part and the reconstructed part, and the higher the quality of the extracted global features. The loss here is mean squared error (MSE), which can be expressed as (1). In rate-distortion optimization, R represents the entropy rate, λ represents the penalty coefficient used to control different bit rates, $\hat{y}$ represents the latent representation information, and $\hat{z}$ represents the side information

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}(\hat{X} - X)^2 \tag{1}$$

where m denotes the number of pixels, $\hat{X}$ denotes the reconstructed image, and X denotes the original image.

In Tables I and II, N represents the number of channels, ↓ represents the downsampling, ↑ represents the upsampling, and RELU represents the linear rectification function. GDN stands for generalized split normalization function and IGDN stands for its inverse operation, which are nonlinear activation functions and are more suitable for normalizing image data than other normalization functions.

Here is how GFRNet works as a whole.

1) *Compression*: The remote sensing data block on a branch is PVTCB to obtain multilevel global features. On the other branch, the remote sensing data block obtains the shallow features after removing noise through compression block 1, QFMR-AM and compression blocks 2–3, and then inputs them into GVGM to obtain the global features (here, the global features wait for the subsequent feature alignment. On the one hand, it is returned to the network and is used to strengthen the quality of the global features). Then, the remote sensing data are input into compression block 4 and further compressed to obtain deep features. The features of the two branches are then fused and fed into the traditional codecs for further compression and probabilistic modeling.
2) *Reconstruction*: The bitstream that the model will get is reconstructed step by step. The remote sensing data are input into reconstruction block 4 for the first step of reconstruction, and the output feature maps of compression block 3 and reconstruction block 4 are input into GVGM for global feature alignment. Here, GVGM plays two roles: on the one hand, the difference between the global features of the compressed part and the reconstructed part is transmitted to the overall loss through Loss_{GVF}, so as to expand the scope of the global feature. On the other hand, GVGM will return the extracted global features to the network, thereby enhancing the quality of global features. After that, the obtained remote sensing data were successively reconstructed by reconstruction block 3, reconstruction block 2, QFMR-AM, and reconstruction block 1.

B. Quadruple Filtered Multicore Noise Reduction Attention Module

Remote sensing images are usually captured by satellites through complex atmospheric layers and inevitably have a lot of background noise. In addition, hyperspectral images usually have hundreds of bands, which can ensure the validity of the extracted features by extracting features in the band with less noise. However, remote sensing images usually have only a few bands, and there are fewer bands to choose from. Therefore, the noise in remote sensing images will bring serious negative effects to the tasks, such as object detection, scene classification, image compression, and so on. In this article, QFMR-AM is designed for the complex background noise in remote sensing images. The module uses multiple filters and multiscale convolution techniques to reduce background noise while efficiently extracting, enhancing, and fusing features at different scales.

Common noises in remote sensing images include speckle noise and salt-and-pepper noise. Speckle noise refers to the small areas of an image that have random abrupt changes in brightness or color. Salt-and-pepper noise, also known as impulse noise, is a random occurrence of white and black dots in an image. These noises manifest themselves in different ways, but they are essentially abrupt changes in pixel values. The essence of convolution is the linear addition of pixels in the local area, which alleviates the problem of pixel numerical mutation caused by noise, and the larger the convolutional kernel, the more noise can be reduced. Therefore, a four-branched QFMR-AM is constructed to effectively eliminate the influence of noise. The structure of the QFMR-AM is shown in Fig. 2. Data A in the Filter2 branch can be expressed as

$$X = \text{Conv}_{3\times3}(\text{Input}) \tag{2}$$

where Input represents the input image data block, and $\text{Conv}_{3\times3}$ represents a convolutional layer with a convolutional kernel size of 3.

The work process of the branch of Filter_1 can be expressed as

$$I_{\text{Filter}_1} = \text{Filter}_1(\text{Avgpool2D}_{3\times3}(\text{Conv}_{1\times1}(\text{Input})) \odot X) \tag{3}$$

where $\text{Conv}_{1\times1}$ represents the point convolution, $\text{Avgpool2D}_{3\times3}$ represents the average pooling, $\odot$ represents the Hadamard product, Filter_1 represents filter 1, and I_{Filter1} represents the output of the Filter1 branch.

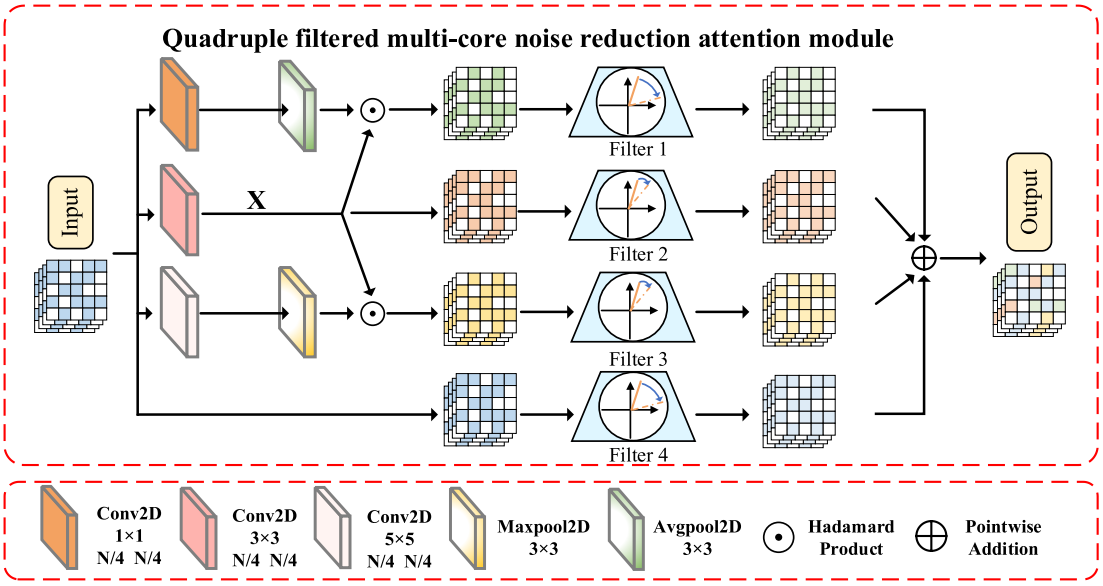


Fig. 2. Schematic of QFMR-AM. (Input represents the input feature map, output represents the output feature map, Filters1–4 represents filters with different mapping coefficients, that is, the pixel values of the feature map are mapped to a reasonable range. The mapping coefficient is 0.3 for Filter1, 0.7 for Filter2, 0.7 for Filter3, and 0.3 for Filter4. Conv2D 1 × 1 $N/4$ $N/4$ represents the parameter settings of the convolution, where 1 × 1 represents the size of the convolution kernel and $N/4$ represents the number of channels.)

The work process of the branch of Filter_2 can be expressed as

$$I_{\text{Filter2}} = \text{Filter}_2(\text{Conv}_{3\times3}(\text{Input})) \tag{4}$$

where Input represents the input image data block, $\text{Conv}_{3\times3}$ represents a convolutional layer with a convolutional kernel size of 3, Filter_2 represents filter 2, and I_{Filter2} represents the output of the Filter2 branch.

The work process of the branch of Filter_3 can be expressed as

$$I_{\text{Filter}_3} = \text{Filter}_3(\text{Maxpool2D}_{3\times3}(\text{Conv}_{5\times5}(\text{Input})) \odot X) \tag{5}$$

where $\text{Conv}_{5\times5}$ represents a convolutional layer with a convolutional kernel size of 5, $\text{Maxpool2D}_{3\times3}$ represents the maximum pooling, Filter_3 represents filter 3, and I_{Filter3} represents the output of the Filter3 branch.

The work process of the branch of QFMR-AM can be expressed as

$$I_{\text{QFMR-AM}} = I_{\text{Filter1}} \oplus I_{\text{Filter2}} \oplus I_{\text{Filter3}} \oplus I_{\text{Filter4}} \tag{6}$$

where I_{Filter4} represents the output of the Filter4 branch, $\oplus$ represents the pointwise addition, and $I_{\text{QFMR-AM}}$ represents the output of QFMR-AM.

Specifically, the Filter1 branch uses point convolution to extract a high-frequency feature with small-kernel convolution, which will contain more noise information. So, Filter1 limits it to a smaller pixel value by mapping, where the mapping coefficient is 0.3. The convolution kernels of the Filter2 and Filter3 branches are 3 × 3 and 5 × 5, respectively, which are used to obtain multiscale features in remote sensing images, and Filter2 and Filter3 map the features to larger pixel values as the backbone of this module. The mapping coefficient for Filter2 and Filter3 is 0.7. The Filter4 branch is similar to a residual structure that is used to speed up the training of the network and prevent overfitting. Unlike the residuals, this article adds a filter to this branch and maps it to smaller pixel values. This prevents the noise from the original block from being introduced into the final feature map again with the same high impact. In summary, the core function of the four filters here is a linear mapping of pixel values. The mapping coefficients of Filter1 and Filter4 are set at small values, and the main reason is that the point convolution in the Filter1 branch does not alleviate the problem of abrupt change of pixel values and still contains a certain amount of noise. The Filter4 branch, on the other hand, is similar to the residual branch and still contains a certain amount of noise. Therefore, the filter mapping coefficients of these two branches are set at a small value to reduce the effect of noise. In addition, convolution of different scales is used in the branches of Filter2 and Filter3, which fully alleviates the problem of numerical mutation caused by noise. Therefore, the mapping coefficient of these two branches is set at a large value, so that these two branches become the backbone of QFMR-AM. In addition, this module introduces Avgpool2D between Filter1 and Filter2 branches and Maxpool2D between Filter3 and Filter4 branches. In this way, the interaction between features of different scales is enhanced. Finally, the feature maps of the Filter1, Filter2, Filter3, and Filter4 branches are fused together by pointwise addition. Due to the filter's reasonable restriction and allocation of the output of each branch, the pixel values of the final features are still within a reasonable range.

C. Global Visual Feature Guidance Module

Remote sensing images have abundant global information, but the common way to obtain global features mainly exists in a certain part of the network in the form of a module. The global features extracted by these methods are either applied to a certain part or a certain branch, and their influence range

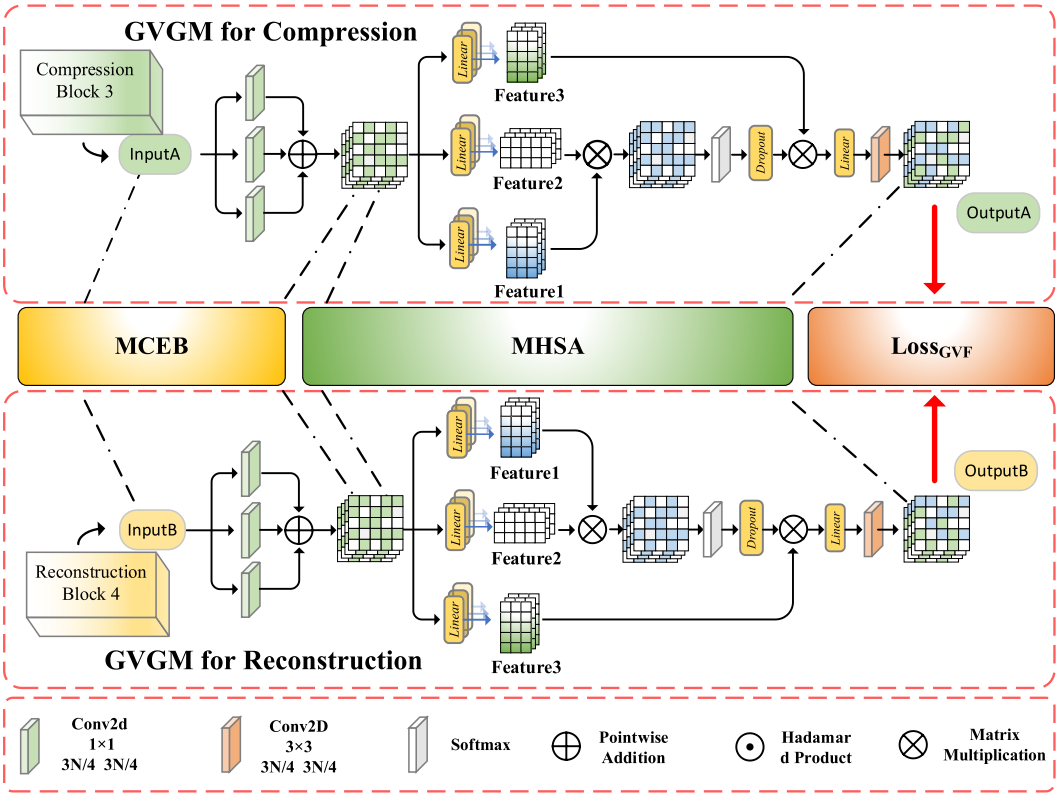


Fig. 3. Schematic of GVGM. (InputA represents the output feature map of compression block 3, and InputB represents the output feature map of reconstruction block 4, OutputA represents the processed global features of the compression part, and OutputB represents the processed global features of the reconstruction part. "GVGM for compression" represents the GVGM used to calculate the global features of the compressed part. "GVGM for reconstruction" represents the GVGM used to calculate the global features of the reconstructed part. Loss_{GVF} represents the difference between the global features of the compressed part and the global features of the reconstructed part, and this loss is MSE.)

is limited. Therefore, this article designs GVGM, which can apply global features to the entire network in the form of loss. The structure of GVGM is shown in Fig. 3. It mainly includes three core components, namely, MCEB for extracting multichannel features, MHSA for extracting global features, and Loss_{GVF} for calculating the difference in global features between the compression part and the reconstruction part.

MCEB uses three convolutions with kernel sizes of 1×1 to extract and enhance multichannel features. The reason for using multiple channels for spatial information extraction is that the size of the MCEB input feature map is relatively large, which leads to more spatial information to be extracted. Therefore, the low information capacity of a single channel will lead to the loss of part of the spatial information. The reason for adding the three features is that this makes the pixel values larger, which increases the difference between the different features and makes the output features more suitable for MHSA.

The process of MCEB can be represented as

$$I_{\text{MCEB}} = \text{Conv}_{1\times1}(\text{Input}) + \text{Conv}_{1\times1}(\text{Input}) + \text{Conv}_{1\times1}(\text{Input}) \quad (7)$$

where Input represents the input image data block, $\text{Conv}_{1\times1}$ represents point convolution, and I_{MCEB} represents the output of MCEB.

Another core component is MHSA: in recent years, vision transformer (ViT) has been widely used in the field of computer vision (CV). Its powerful long-range feature capture capability makes the ViT model excellent for a wide range of tasks. This capability is mainly due to its core module, the self-attention mechanism. MHSA introduces a multihead mechanism on the basis of self-attention, which not only improves the training speed but also realizes the fusion of different subspace features. MHSA maps remote sensing image data to different projection spaces through Feature1, Feature2, and Feature3. It uses tensor multiplication to fuse the features of different spaces, thereby enhancing the global features.

The process of the MHSA can be represented as

$$I_{\text{MHSA}} = \text{Conv}_{3\times3}(\text{Linear}(\text{Feature}_3 \otimes (\text{Dropout}(\text{Soft max}(\text{Feature}_1 \otimes \text{Feature}_2))) \quad (8)$$

where $\text{Conv}_{3\times3}$ represents convolution, Linear represents the linear layer, Dropout represents the randomly deactivated layer, Soft max represents the softmax layer, and $\otimes$ represents the matrix multiplication.

The last core component is Loss_{GVF}, which calculates the difference between the global features of the compressed part and the global features of the reconstructed part, and inputs the difference of this global feature into $\text{Loss}_{\text{Total}}$. The reasons are as follows.

1) The network can reduce the difference between the global features of the compressed part and the global features of the reconstructed part by decreasing the Loss_{GVF}. In this way, the similarity between the global

Algorithm 1 The Feature Extraction Process of Remote Sensing Images by GVGM

Input: Remote sensing data $X_{Compression} \in \mathbb{R}^{b\times c\times h\times w}$, $X_{Reconstruction} \in \mathbb{R}^{b\times c\times h\times w}$

1: **Stage1:**

2: Perform *MCEB*, $X_{Compression} \in \mathbb{R}^{b\times c\times h\times w}$ denoted as $X_1 \in \mathbb{R}^{b\times c\times h\times w}$

3: Perform *Flatten*, *Reshape* and *Linear*, the result denoted as $attn \in \mathbb{R}^{b\times n\times 3c}$

4: Perform *Split* and *Reshape*, the result denoted as $Feature_1 \in \mathbb{R}^{b\times n\times c}$, $Feature_2 \in \mathbb{R}^{b\times n\times c}$ and $Feature_3 \in \mathbb{R}^{b\times n\times c}$

5: Perform *Reshape* and *Transpose*, the result denoted as $Feature_1 \in \mathbb{R}^{b\times head\times n\times headd}$, $Feature_2 \in \mathbb{R}^{b\times head\times headd\times n}$, $Feature_3 \in \mathbb{R}^{b\times head\times n\times headd}$

6: Perform *MatrixMultiplication* of $Feature_1 \in \mathbb{R}^{b\times head\times n\times headd}$ and $Feature_2 \in \mathbb{R}^{b\times head\times headd\times n}$, and then perform *Softmax*, *Dropout*, the result denoted as $attn_1 \in \mathbb{R}^{b\times head\times n\times n}$

7: Perform *MatrixMultiplication* of $attn_1 \in \mathbb{R}^{b\times head\times n\times n}$ and $Feature_3 \in \mathbb{R}^{b\times head\times n\times headd}$, the result denoted as $attn_2 \in \mathbb{R}^{b\times head\times n\times headd}$

8: Perform *Transpose*, *Flatten*, *Linear*, the result denoted as $attn_3 \in \mathbb{R}^{b\times n\times c}$

9: Perform *Transpose*, *Reshape*, the result denoted as $attn_4 \in \mathbb{R}^{b\times c\times h\times w}$

10: Perform $Conv_{3\times 3}$, the result denoted as $attn_{GVGM(Compression)} \in \mathbb{R}^{b\times c\times h\times w}$

11: **Stage2:**

12: Replace the input with $X_{Reconstruction} \in \mathbb{R}^{b\times c\times h\times w}$, and perform Stage 1, the result denoted as $attn_{GVGM(Reconstruction)} \in \mathbb{R}^{b\times c\times h\times w}$

13: **Stage3:**

14: Calculate the *MSE* between $attn_{GVGM(Compression)} \in \mathbb{R}^{b\times c\times h\times w}$ and $attn_{GVGM(Reconstruction)} \in \mathbb{R}^{b\times c\times h\times w}$, the result denoted as $Loss_{GVF}$

end for

Output: $attn_{GVGM(Compression)} \in \mathbb{R}^{b\times c\times h\times w}$, $attn_{GVGM(Reconstruction)} \in \mathbb{R}^{b\times c\times h\times w}$ and $Loss_{GVF}$.

features of the compressed part and the global features of the reconstructed part is improved, and the quality of the global features is improved.

2) Loss_{GVF} is introduced into $\text{Loss}_{\text{Total}}$, so that the influence of the global feature is propagated to the entire network. In this way, the limitation of the scope of the global feature extraction module is solved.

The process of Loss_{GVF} can be represented as

$$\text{Loss}_{\text{GVF}} = L_{\text{MSE}}(I_{\text{MHSA(Compression)}}, I_{\text{MHSA(Reconstruction)}}) \quad (9)$$

where L_{MSE} represents the loss measured using MSE, $I_{\text{MHSA(Compression)}}$ represents the output of the MHSA of the compressed part, $I_{\text{MHSA(Reconstruction)}}$ represents the output of the MHSA of the reconstructed part, and Loss_{GVF} represents the loss of global visual features.

The feature extraction process of GVGM is described in Algorithm 1.

D. Pyramid Vision Transformer Compression Branch

Transformer initially made its mark in the field of natural language processing (NLP), with its self-attention mechanism making it adept at processing sequence data. As research deepened, scientists discovered that its potential was not limited to language and began to try to introduce it into the field of CV. By applying transformer to image data, researchers aim to capture global dependencies in images to improve tasks, such as image recognition and object detection, and have achieved good research results. However, no one has introduced PVT into the field of remote sensing image compression [55]. Therefore, for the first time, we have introduced it into the field of remote sensing image compression and make targeted optimization to make it more suitable for image compression tasks. Common transformer has some drawbacks: 1) the self-attention mechanism of the standard transformer causes significant computational and memory overhead when processing large images, because it performs pairwise attention calculations on all elements in the input sequence and 2) transformer typically outputs single-scale feature maps, which limits its application in tasks that require high-resolution output, such as pixel-level tasks. However, the PVT solves these problems well. First, PVT uses a pyramid structure to generate multiscale feature maps, which makes the model suitable for downstream tasks with different resolutions, especially for dense prediction tasks, such as object detection and semantic segmentation. Second, it uses fine-grained image blocks (4×4 pixels) as an input to learn high-resolution representations. Finally, spatial-reduction attention (SRA) is introduced, which effectively reduces the computation and memory consumption by reducing the dimensionality of the input space dimension before attention calculation. This enables PVT to handle high-resolution tasks more efficiently.

In our study, PVT is introduced into remote sensing image compression for the first time. It is improved to be more suitable for remote sensing image compression tasks. In this article, PVTCB is designed to increase the multilevel nature of the global feature. The structure of PVTCB is shown in Fig. 4. The process of PVTCB can be represented as

$$I_{\text{PVTCB}} = \text{Filter}_5\Big(\text{Stage3}_{B\times C_4\times \frac{H}{16}\times \frac{W}{16}}\Big(\text{Stage2}_{B\times C_3\times \frac{H}{8}\times \frac{W}{8}} \times \Big(\text{Stage1}_{B\times C_2\times \frac{H}{4}\times \frac{W}{4}}(\text{Input})\Big)\Big)\Big) \quad (10)$$

where B represents the batches, C represents the channels, H represents the height of the image, W represents the width of the image, Filter_5 represents filter 5, and I_{PVTCB} represents the output of PVTCB.

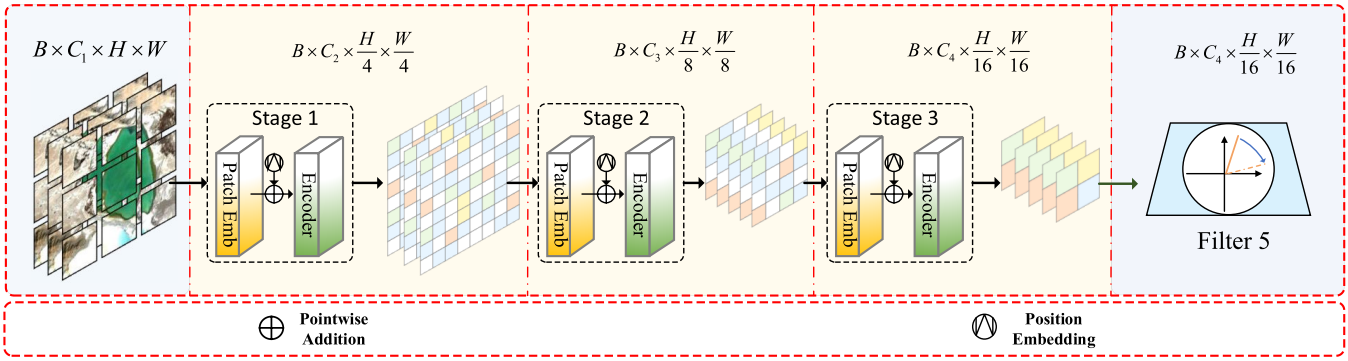


Fig. 4. Schematic of PVTCB, which is generally divided into two parts. The first part is TPTC (light yellow part in the figure), which consists of three stages, each of which consists of a patch embedding layer and a transformer encoder. According to the structure of the pyramid, the size of the data block changes from large to small, and the dimension changes from low to high dimension. The second part is Filter5, which has a mapping coefficient of 0.3.

The main function of PVTCB is to reduce the spatial size of remote sensing data blocks. Stage1 reduces the spatial size of the data to one-fourth of the original size, Stage2 reduces the spatial size of the data to half of the original size, and Stage3 reduces the spatial size of the data to half of the original size. In this way, the shallow remote sensing image features are compressed into deep features. PVTCB consists of two core components, including TPTC and Filter5. The original PVT consists of four stages. These four stages can be seen as four more complex downsampling. As the stage deepens, the size of the data block gradually decreases, and the number of channels increases. However, PVT reduces the size of the data block excessively and has a large number of parameters. Therefore, this article reconstructs the stage number and the key parts of the model to construct a TPTC that is more suitable for remote sensing image compression. The stage number has been refactored to three, which greatly reduces the complexity of the model. In this way, the output of PVTCB and the output of FSCB can be matched. In addition, the number of embedded dimensions for each stage is set to [64, 160, 256], which improves the model's ability to obtain channel features. The number of heads in each stage is set to [2], [4], and [8] to enhance the fusion of information in different subspaces. The encoder stacking depth of each stage is set to [2, 2, 2]. The space reduction scale of the sublayer is set to [4, 2, 1] to preserve more channel features. Finally, Filter5 is added to the end of TPTC to limit the pixel values of the multilevel global feature map to a reasonable range, so as to facilitate effective fusion with the output features of FSCB.

E. Rate-Distortion Optimization

The training goal of the compression framework is to achieve a balance between compression and distortion. To achieve this, a rate distortion optimization strategy is often added to the compression framework to guide the model for efficient training. In short, the strategy is designed to ensure that the data are compressed with as little information loss as possible. The rate distortion optimization strategy can be represented as

$$\arg\min \text{Loss}_{\text{Total}} = R + \lambda D \tag{11}$$

where R represents entropy rate, which is the cross-entropy between the latent edge distribution and the learning entropy model. D represents distortion between the original image and the reconstructed image. Different bitrates can be controlled by adjusting the penalty coefficient λ

$$R = R_{\hat{y}} + R_{\hat{z}} \tag{12}$$

where the bitrate consists of the latent representation information $\hat{y}$ together with the side information $\hat{z}$

$$R_{\hat{y}} = -\sum_i \log_2(p_{\hat{y}}(\hat{y})) \tag{13}$$

$$R_{\hat{z}} = -\sum_i \log_2(p_{\hat{z}}(\hat{z})) \tag{14}$$

where $p_{\hat{y}}$ is an entropy model that can be learned, and $p_{\hat{z}}$ represents the hyperencoder.

In order to further improve the quality of image compression, a novel rate distortion optimization strategy is proposed in this article. Loss_{GVF} is introduced into $\text{Loss}_{\text{Total}}$, so that the influence of the global feature is propagated to the entire network. This improves the quality of global features throughout the network. This novel rate distortion optimization strategy can be expressed as

$$\arg\min \text{ProposedLoss}_{\text{Total}} = R + \lambda(D + \psi \text{Loss}_{\text{GVF}}) \tag{15}$$

where ψ represents the coefficient of Loss_{GVF}.

IV. Experimental Results and Analysis

Sufficient experiments have been carried out on some remote sensing image datasets, including San Francisco [56], NWPU-RESISC45 [57], and UC-Merced [58]. These datasets contain a wealth of ground object information, which can effectively evaluate the performance of GFRNet. In this article, GFRNet is compared with some excellent compression methods, including traditional codecs and deep learning-based compression models, to verify the superiority of the proposed method. Traditional image compression methods include JPEG2000 [14], BPG [59], and WebP [17]. Compression models based on deep learning include Minnen et al. [31], Minnen et al. (mean) [31], Ballé et al. (hyperprior) [32], Ballé et al. (factorized-relu) [32], and Tong2023 [60]. Experimental

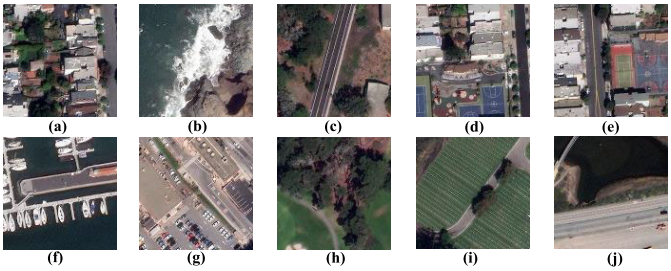


Fig. 5. Some images from San Francisco dataset. (a) Buildings, (b) coastline, (c) highway, (d) basketball court, (e) tennis court, (f) harbor, (g) parking lot, (h) forest, (i) farmland, and (j) lake.

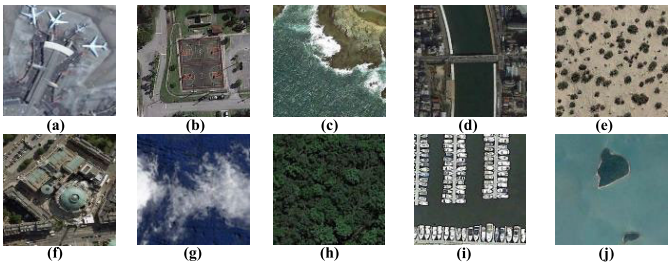


Fig. 6. Some images from NWPU-RESISC45 dataset. (a) Airport, (b) basketball court, (c) beach, (d) bridge, (e) desert, (f) church (g) clouds (h) forest, (i) port, and (j) island.

results show that the proposed GFRNet has the best compression performance in both PSNR and MS-SSIM. In addition, the quality of the reconstructed images obtained by different compression methods is evaluated through the classification task, which further verifies the superiority of GFRNet.

A. Introduction to Remote Sensing Image Dataset

1) Dataset San Francisco: San Francisco is a dataset of remotely sensed images from [56]. It is a remote sensing image with a resolution of 17 408 × 17 408, covering a variety of feature information, such as buildings, coasts, highways, ports, lakes, and so on. In this article, it is cropped to images with size 256 × 256, and 3000 valid images are selected to form the dataset. These images are divided into a training set, a validation set, and a test set at a ratio of 8:1:1. Fig. 5 shows some of the samples.

2) Dataset NWPU-RESISC45: NWPU-RESISC45 is provided by Northwestern Polytechnical University (NWPU). The dataset contains a total of 45 different remote sensing image scene categories. Each category contains 700 images, each with a resolution of 256 × 256 pixels. The dataset contains a variety of geographical environments and scenarios, including airports, deserts, churches, forests, and so on. The 140 images in each category were selected to form a dataset of 6300 remote sensing images, which was then divided into a training set, a validation set, and a test set at a ratio of 8:1:1. Fig. 6 gives some of the samples.

3) Dataset UC-Merced: UC-Merced is a remote sensing image dataset provided by the University of California, Merced. The UC-Merced dataset consists of 21 different categories, each consisting of 100 images. A total of 2100 images are included, each with a resolution of 256 × 256 pixels. The images include farmland, airports, forests, and other landform

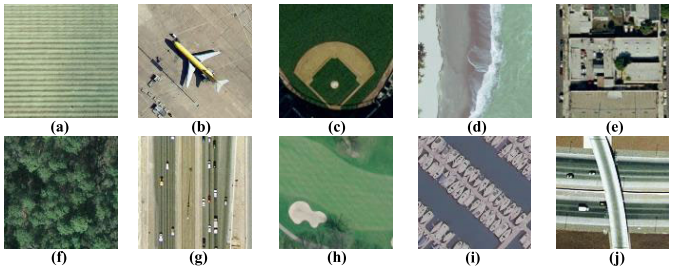


Fig. 7. Some images from UC-Merced dataset. (a) Farmland, (b) airplanes, (c) baseball stadiums, (d) beaches, (e) buildings, (f) forests, (g) roads, (h) golf courses, (i) ports, and (j) overpasses.

scenes. The dataset UC-Merced is divided into a training set, a validation set, and a test set at a ratio of 8:1:1. Fig. 7 shows some of the samples.

B. Evaluation Indicators

To evaluate the quality of reconstructed images, two commonly used evaluation metrics are adopted, i.e., PSNR and MS-SSIM. In the part of remote sensing scene image classification, the overall accuracy (OA) and confusion matrix (CM) are also used to measure the classification performance.

1) Peak Signal-to-Noise Ratio: PSNR compares the reconstructed image to the original image from the point of view of the mean square error. The higher the PSNR value, the higher the fidelity of the reconstructed image. The PSNR can be expressed as

$$\mathrm{PSNR}(X, \widehat{X}) = \frac{1}{C}\sum_{i=1}^{C} 10\log_{10}\left(\frac{\max^2(X^i)}{\mathrm{MSE}_i}\right) \tag{16}$$

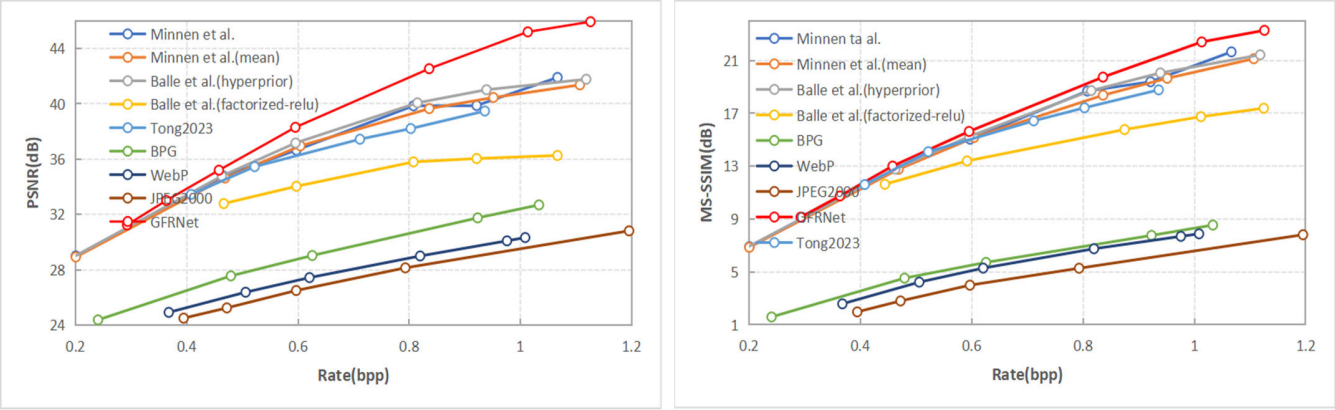
where MSE represents the mean square error between the original image and the reconstructed image. $\max^2(X^{(i)})$ represents the square of the largest pixel in band i. C represents the number of bands.

2) Multiscale Structural Similarity Index Metric: MS-SSIM is a multiscale structural similarity index. It measures the difference between the original image and the reconstructed image by merging image details at different resolutions. The value ranges from 0 to 1, with higher values indicating higher similarity and higher quality of the reconstructed image. The formula for MS-SSIM can be expressed as

$$D_{\text{MS-SSIM}} = 1 - \prod_{m=1}^{M}\left(\frac{2\mu_X\mu_{\widehat{X}} + C_1}{\mu_X^2 + \mu_{\widehat{X}}^2 + C_1}\right)^{\alpha_m}\left(\frac{2\sigma_{X\widehat{X}} + C_2}{\sigma_X^2 + \sigma_{\widehat{X}}^2 + C_2}\right)^{\zeta_m} \tag{17}$$

where M represents different resolutions, μ_X and $\mu_{\widehat{X}}$ represent the mean of the original image and the reconstructed image, respectively, σ_X and $\sigma_{\widehat{X}}$ represent the standard deviation between the original image and the reconstructed image, respectively, $\sigma_{X\widehat{X}}$ represents the covariance between the original image and the reconstructed image, α_m and ζ_m represent the relative importance between the two terms, and C_1 and C_2 are constant terms to prevent the divisor from being 0.

In order to clearly compare the differences in MS-SSIM values, they are converted into decibel values. This process



(a) (b)

Fig. 8. Rate distortion curves on San Francisco. (a) PSNR and (b) MS-SSIM.

can be expressed as

$$\text{MS-SSIM} = -10\log_{10}(1 - D_{\text{MS-SSIM}}). \tag{18}$$

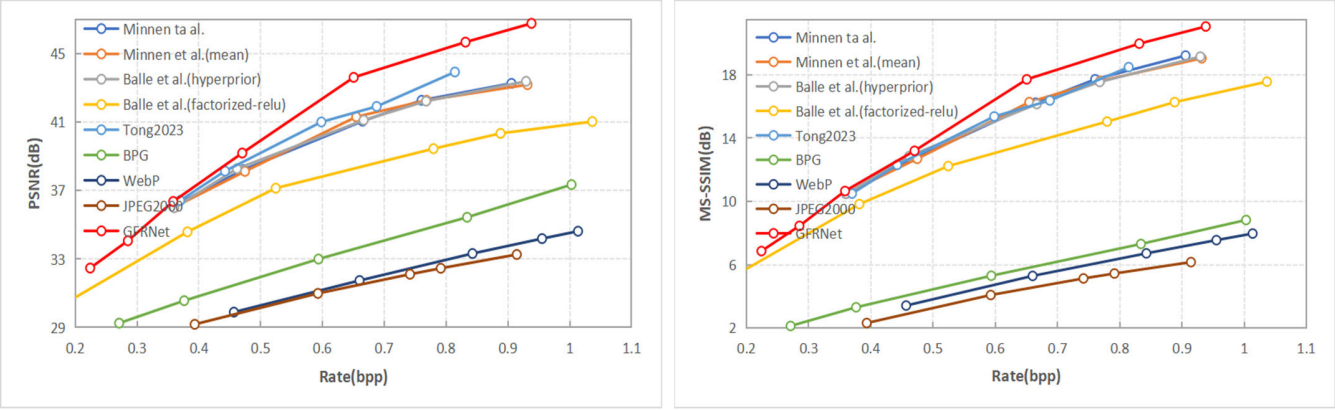
3) Classification Indicators of Remote Sensing Scenes: In this article, two widely used remote sensing scene classification evaluation indicators are selected to measure the quality of the reconstructed image, including OA and CM. The OA value is obtained by dividing the number of correctly classified images by the total number of test images, and it reflects the overall performance of a classification model. CM reflects the degree of confusion and detailed classification errors between different scene categories. Each row in the CM represents the true category, and each column represents the predicted category.

C. Experimental Environment and Parameter Settings

In this study, the proposed GFRNet is implemented by PyTorch. The Adam optimizer was chosen. There are two optimizers in this network, one is the main optimizer between the main encoder (PVTCB and FSCB) and the main decoder (FSRD), and the other is the auxiliary optimizer between the hyperencoder and the hyperdecoder. For the main optimizer, the initial learning rate is set at 10^{-4}, and the optimal model of GFRNet will be stored when the learning rate decays to 10^{-6} during network training. For the auxiliary optimizer, its initial learning rate is set at 10^{-3}. During training, the batch size is set to 8. In this experiment, the neural network models are trained on an NVIDIA GeForce RTX 3090, and the traditional codecs are performed on a CPU (i9-9900K CPU at 3.60 GHz). For the sake of fairness, all experiments in this article were conducted in the above environment. The penalty coefficient λ used in this article is [0.660, 0.508, 0.211, 0.072, 0.033, 0.013, 0.007]. The mapping coefficients of filters 1–5 are [0.3, 0.7, 0.7, 0.3, 0.3, 0.3]. In GVGM, the number of heads in the MHSA is set to 4. In the proposed rate distortion optimization strategy, the coefficient ψ of Loss_{GVF} is set to 0.065. N is set to 256 in compression block, reconstruction block, hyperencoder, and hyperdecoder. In the classification of remote sensing scenes, the benchmark model used for testing was efficient multiscale transformer and cross-level attention learning (EMTCAL) [61]. The dataset used for training is NWPU-RESISC45, and the training-to-test ratio is 10%–90%. The images used for compression and the images used for remote sensing scene classification training are not crossed. The reconstructed images are only used for testing the classification performance, not for the training of the classification network.

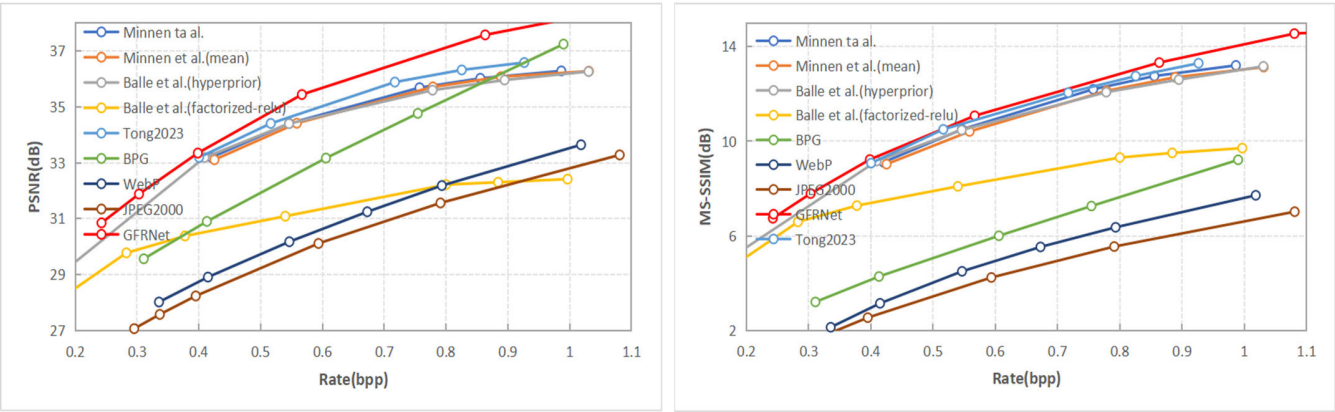
D. Rate Distortion Performance

In this experiment, all models were evaluated for rate distortion performance by PSNR and MS-SSIM. In this article, eight comparison methods are selected, including three traditional image compression methods and five image compression methods based on deep learning. Figs. 8–10 show the rate distortion performance curves obtained by different compression methods on the dataset San Francisco, NWPU-RESISC45, and UC-Merced, respectively. In traditional image compression methods, BPG shows better rate distortion performance than WebP and JPEG2000 in most cases. This is mainly due to BPG's multichannel coding technology. This technique allows for independent encoding of different color channels, which in turn enables fine control over detailed features, helping to reconstruct high-quality images. For the image compression methods based on deep learning, the rate-distortion performance demonstrated by Ballé et al. (factorized-relu) [32] is relatively poor. This is mainly due to the fact that it employs only simple convolutional layers, which have limited capabilities in feature extraction. Although the feature extraction ability can be improved to a certain extent by increasing the number of convolutional layers, this will significantly increase the number of parameters of the model and prolong the reconstruction time. On the dataset NWPU-RESISC45 and UC-Merced, the Tong2023 method achieves the highest PSNR and MS-SSIM rate distortion performance except GFRNet. This is mainly due to its excellent attention mechanism and a more reasonable residual convolution module. However, on the dataset San Francisco, the rate distortion performance of the Tong2023 method is poor, which indicates that the compression model is less robust. The other comparison methods based on deep learning have average performance, mainly



(a) (b)

Fig. 9. Rate distortion curves on NWPU-RESISC45. (a) PSNR and (b) MS-SSIM.



(a) (b)

Fig. 10. Rate distortion curves on UC-Merced. (a) PSNR and (b) MS-SSIM.

because they lack a strong attention mechanism and excellent rate distortion optimization strategies. GFRNet proposed in this article achieves the highest PSNR and MS-SSIM rate distortion performance on three datasets at the same time. On the dataset San Francisco, specifically, at 1.1 bpp, GFRNet achieves PSNR improvements of 9.3%, 11.0%, 9.9%, 24.1%, and 12.2% compared to that of Minnen et al. [31], Minnen et al. (mean) [31], Ballé et al. (hyperprior) [32], Ballé et al. (factorized-relu) [32], and Tong2023, respectively. In addition, GFRNet achieves MS-SSIM improvements of 6.5%, 10.6%, 8.6%, 31.1%, and 11.8% compared to that of Minnen et al. [31], Minnen et al. (mean) [31], Ballé et al. (hyperprior) [32], Ballé et al. (factorized-relu) [32], and Tong2023, respectively. This superior rate distortion performance not only strongly proves the robustness of GFRNet but also strongly proves the effectiveness of PVTCB, QFMR-AM, GVGM, and the proposed rate distortion optimization strategy in GFRNet.

E. Visualization Comparison of Reconstructed Images

In order to further verify the effectiveness of GFRNet, this article visually compares the reconstructed images of different methods. Figs. 11 and 12 are the reconstructed images on the dataset San Francisco and the dataset UC-Merced, respectively, and their local enlarged images. The images in the visualization experiment were all reconstructed images with a bit rate of 0.25 bpp. Taking Fig. 11 as an example, from left to right, the reconstructed image of the original image, the reconstructed image of the eight comparison methods, and the reconstructed image of GFRNet. In the traditional image compression method, the rate distortion performance of BPG is significantly better than that of JPEG2000 and WebP. In the enlarged image of the reconstructed image, compared with JPEG2000 and WebP, the roof of the BPG method retains more texture information. However, the JPEG2000 and WebP reconstruction areas have lost most of the detail features and are blurry. The main reason for this phenomenon is that the BPG method has a multichannel encoding technique, which has a stronger ability to reconstruct detailed features. The comparison method based on deep learning generally achieves better visualization than the traditional image compression method, but it is still inferior to GFRNet. In Fig. 11, some artifacts and noise are prevalent in the reconstructed images of Minnen et al. [31], Minnen et al. (mean) [31], Ballé et al. [32], and Ballé et al. (factorized-relu) [32]. This results in a blurry image. The transitions between the pixels of the reconstructed image of these four comparison methods are too coarse, which leads to color flattening and distortion. Finally, comparing Tong2023 with GFRNet, the tree of the enlarged image in GFRNet retains more texture features and sharper edges of objects. As a result, GFRNet achieved the best visualization

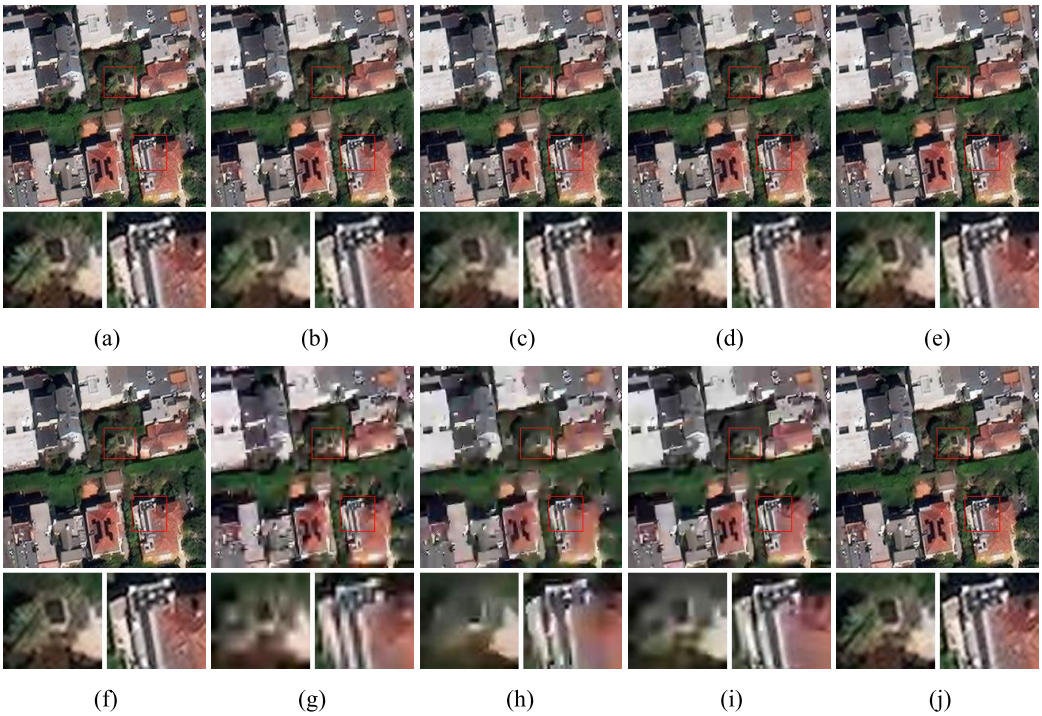


Fig. 11. Visual comparison of reconstructed images obtained by different methods on the dataset San Francisco. (a) Original, (b) Minnen et al. [31] (bpp: 0.252; PSNR: 29.92; MS-SSIM: 7.91), (c) Minnen et al. (mean) [31] (bpp: 0.249; PSNR: 29.38; MS-SSIM: 7.36), (d) Ballé et al. (hyperprior) [32] (bpp: 0.251; PSNR: 29.53; MS-SSIM: 7.47), (e) Ballé et al. (factorized-relu) [32] (bpp: 0.252; PSNR: 29.25; MS-SSIM: 7.49), (f) Tong2023 (bpp: 0.252; PSNR: 30.28; MS_SSIM: 8.10), (g) JPEG2000 (bpp: 0.262; PSNR: 22.27; MS-SSIM: 1.15), (h) Webp (bpp: 0.255; PSNR: 22.71; MS-SSIM: 1.19), (i) BPG (bpp: 0.270; PSNR: 23.96; MS-SSIM: 1.93), and (j) GFRNet (bpp: 0.252; PSNR: 30.10; MS-SSIM: 8.26).

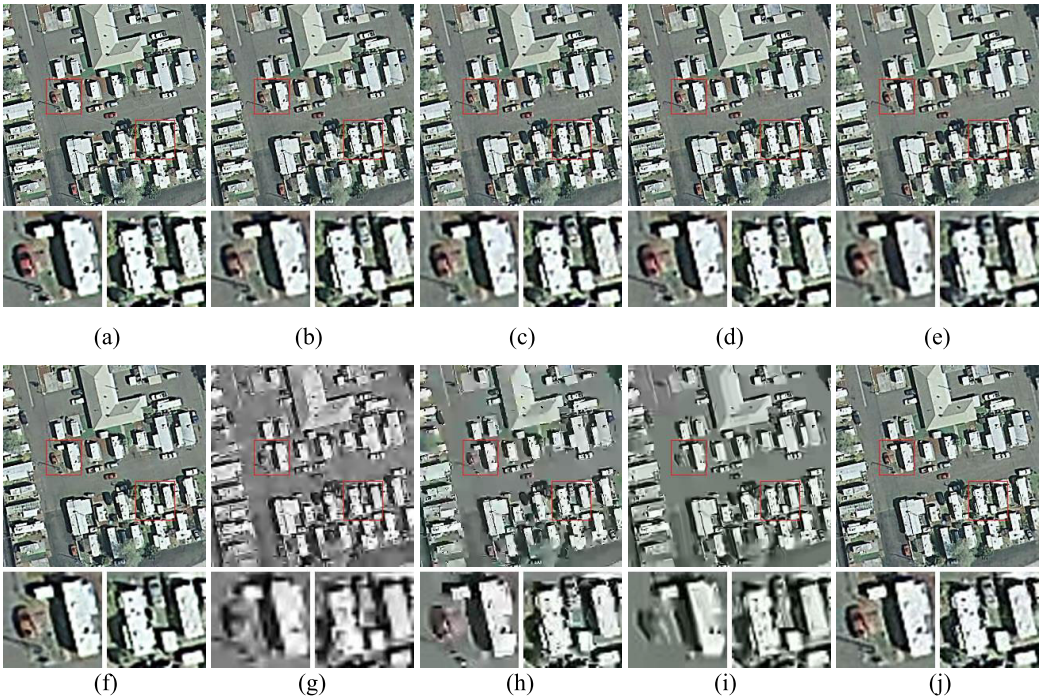
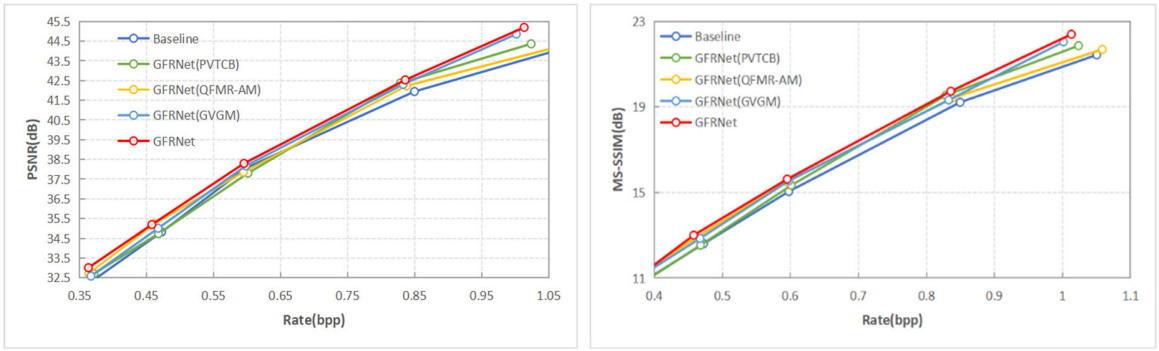


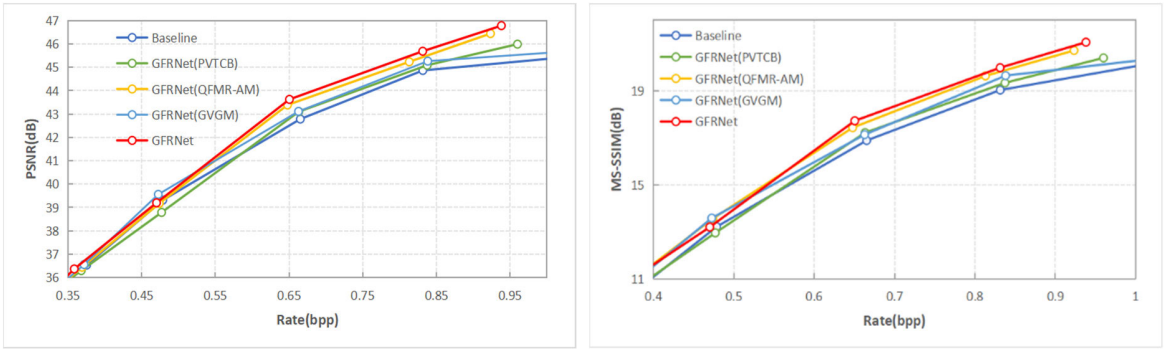
Fig. 12. Visual comparison of reconstructed images obtained by different methods on the dataset UC-Merced. (a) Original, (b) Minnen et al. [31] (bpp: 0.249; PSNR: 30.37; MS-SSIM: 6.10), (c) Minnen et al. (mean) [31] (bpp: 0.248; PSNR: 30.05; MS-SSIM: 5.69), (d) Ballé et al. (hyperprior) [32] (bpp: 0.249; PSNR: 30.39; MS-SSIM: 6.18), (e) Ballé et al. (factorized-relu) [32] (bpp: 0.250; PSNR: 28.93; MS-SSIM: 5.30), (f) Tong2023 (bpp: 0.251; PSNR: 29.59; MS-SSIM: 5.11), (g) JPEG2000 (bpp: 0.261; PSNR: 16.82; MS-SSIM: 0.75), (h) Webp (bpp: 0.486; PSNR: 19.93; MS-SSIM: 2.46), (i) BPG (bpp: 0.276; PSNR: 18.49; MS-SSIM: 0.97), and (j) GFRNet (bpp: 0.249; PSNR: 30.76; MS-SSIM: 6.93).

on the dataset San Francisco. In addition, in Fig. 12, GFRNet achieves the best visualization on the dataset UC-Merced. This also verifies the robustness of GFRNet. The above experiments fully prove the rationality of the three working principles of GFRNet (removing complex background noise, enhancing the multilevel characteristics of global features, and expanding the



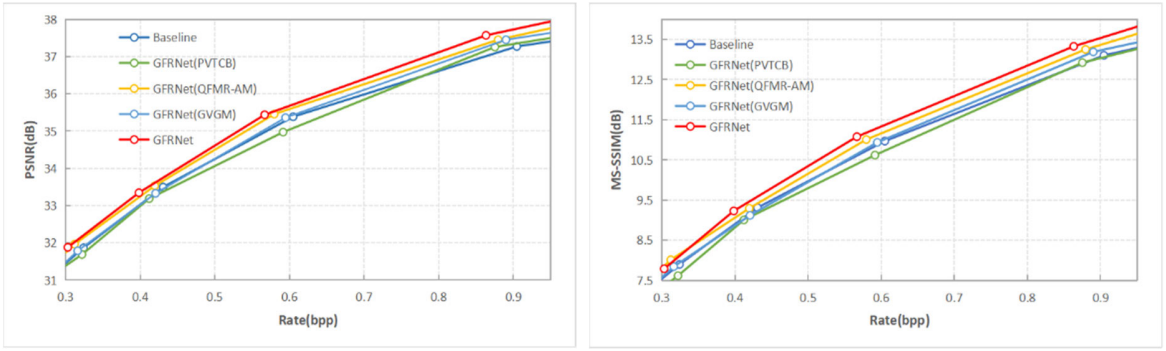
(a) (b)

Fig. 13. Ablation results of different methods on the San Francisco dataset. (a) PSNR and (b) MS-SSIM.



(a) (b)

Fig. 14. Ablation results of different methods on the NWPU-RESISC45 dataset. (a) PSNR and (b) MS-SSIM.



(a) (b)

Fig. 15. Ablation results of different methods on the UC-Merced dataset. (a) PSNR and (b) MS-SSIM.

scope of global features) and the efficiency of the proposed rate distortion optimization strategy.

F. Ablation Experiments

In this article, sufficient ablation experiments were carried out to verify the effectiveness of the proposed components, such as PVTCB, QFMR-AM, and GVGM. Figs. 13–15 are the results of ablation experiments on the dataset San Francisco, NWPU-RESISC45, and UC-Merced, respectively: 1) baseline represents the baseline network; 2) GFRNet (PVTCB) stands for the integration of PVTCB on the basis of baseline; 3) GFRNet (QFMR-AM) represents the integration of QFMR-AM on the basis of baseline; 4) GFRNet (GVGM) stands for GVGM integrated on the basis of baseline; and 5) GFRNet stands for baseline, which integrates PVTCB, QFMR-AM, and GVGM. It should be noted that the GVGM here includes the proposed rate distortion optimization strategy. As can be seen from Figs. 13–15, the rate distortion performance of baseline is the lowest in most cases across all three datasets. GFRNet (PVTCB) is better than baseline at the same bit rate, which verifies the effectiveness of the method to improve the quality of the reconstructed image by improving the multilevel characteristics of global features. The rate distortion performance of GFRNet (QFMR-AM) is also better than that of baseline at the same bit rate, which proves that removing complex background noise from remote sensing images is of great significance. The rate distortion performance of GFRNet (GVGM) is better than that of baseline at the same bit rate, which fully demonstrates the positive impact of expanding the scope of global features on improving the quality of

TABLE III
COMPARISON OF DIFFERENT LOSS FUNCTIONS

	Baseline	RMSE	MAE	SmoothL1Loss	MSE
PSNR	32.38	32.56	32.41	32.36	32.58
MS-SSIM	10.58	10.77	10.61	10.62	10.62

TABLE IV
COMPARISON OF PARAMETERS AND PERFORMANCE OF DIFFERENT MAIN ENCODERS

	Baseline	Baseline+Balle	Baseline+Cheng	Baseline+PVTCB
GPU Memory	1.6GB	2.1GB	4.0GB	2.9GB
Flops	4.01G	9.37G	45.78G	6.03G
Parameters	5.02M	10.21M	14.35M	10.34M
PSNR	43.91	43.47	44.11	44.36
MS-SSIM	21.43	21.49	21.88	21.86

the reconstructed image. In addition, GFRNet achieves the best rate distortion performance at the same bit rate. This phenomenon shows that PVTCB, QFMR-AM, and GVGM achieve efficient feature extraction by removing background noise, enhancing the multilevel characteristics of global features, and expanding the scope of global features under the guidance of the proposed rate distortion optimization strategy.

In addition, the function used to calculate the loss in GVGM was ablated. We add GVGM on baseline and then replace the MSE in GVGM with different losses. There are four types of loss, including MSE, root mean square error (RMSE), mean absolute error (MAE), and SmoothL1Loss. The MSE is used to calculate the average of the squares of the difference between the predicted value and the true value. RMSE is the square root of MSE. MAE is used to calculate the average of the absolute value of the difference between the predicted value and the true value. SmoothL1Loss combines the advantages of the L1 norm and the L2 norm. When the difference between the predicted value and the true value is small, the L2 norm is used; when the difference is large, the L1 norm is used, which balances the effects of small and large errors. The results of the experiment are shown in Table III. The dataset used here is San Francisco. The results here are all around 0.36 bpp. As can be seen from Table III, MSE achieves the best rate distortion performance. The main reason for this is that MSE is calculated by squaring and averaging the prediction error (the difference between the true value and the predicted value) for each sample. The squared operation amplifies the effect of large errors, making the model more focused on reducing predictions that are far from the true value. This sensitivity helps the model to adjust the parameters more precisely during training to reduce differences between global features.

In order to verify the effectiveness of the proposed PVTCB, PVTCB is compared with two publicly available convolution-based main encoders. Here, we have chosen the two main encoders of the public network, Ballé et al. [32] and Cheng et al. [33]. Then, the bpp is 1, and the test parameters include GPU memory, floating point operations (FLOPs), parameters, PSNR, and MS-SSIM. The result is shown in Table IV. The dataset used here is San Francisco. Through comparison, baseline + PVTCB achieved the best PSNR, and the MS-SSIM value was almost the same as that of the best baseline + Cheng. In terms of computing resources, baseline + PVTCB is slightly more than baseline + Balle, but far less than baseline + Cheng. In terms of GPU memory, FLOPs, and parameters, baseline + PVTCB is only 72.5%, 13.2%, and 72.1% of baseline + Cheng, respectively. This shows that PVTCB achieves high rate distortion performance with low complexity. The main reason for this phenomenon is that although PVTCB uses the transformer framework, the number of stages is reduced, and all parameters (including the number of dimensions, the number of heads in self-attention, and the stacking depth) are optimized.

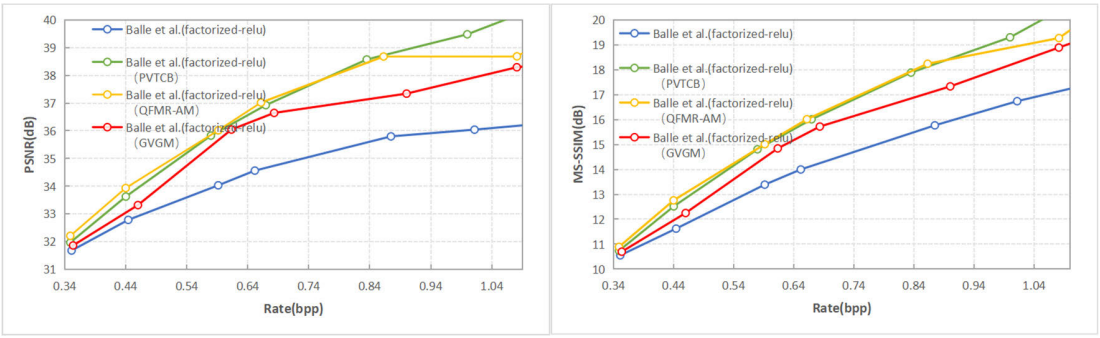
In addition, the visualization of ablation experiments for each module was performed. Among them, GFRNet (PVTCB), GFRNet (QFMR-AM), and GFRNet (GVGM) are 0.988%, 1.019%, and 0.618% higher than that of baseline on PSNR, respectively. GFRNet (PVTCB), GFRNet (QFMR-AM), and GFRNet (GVGM) were increased by 0.189%, 2.741%, and 2.363%, respectively, compared with baseline on MS-SSIM. The visualization results are shown in Fig. 16, and three modules show good visualization results with little to no noise, and the high-quality global characteristics are retained. This is mainly due to the noise suppression by QFMR-AM and the enhancement of global features by GVGM and PVTCB. This fully verifies the effectiveness of each module.

G. Generalization Experiments of Modules

In order to verify the generalization performance of the proposed PVTCB, QFMR-AM and GVGM, some generalization experiments are carried out. It should be noted that the GVGM here includes the proposed new rate distortion optimization strategy. In this article, a public deep learning-based image compression algorithm (Ballé et al. (factorized-relu) [32]) is selected as the baseline network, and then, each module is embedded into the baseline network to verify the generalization of the module. The dataset used here is San Francisco. Fig. 17 shows the rate distortion performance curve of the network after the introduction of each module. The distortion rate performance of the baseline network Ballé et al. (factorized-relu) [32] is at the lowest. After introducing the corresponding modules, Ballé et al. (factorized-relu) (PVTCB) [32], Ballé et al. (factorized-relu) (QFMR-AM) [32], and Ballé et al. (factorized-relu) (GVGM) [32] have all achieved effective improvements in terms of PSNR and MS-SSIM. This

(a) (b) (c) (d) (e)

Fig. 16. Visual comparison of reconstructed images obtained by baseline with different modules on the dataset San Francisco. (a) Original, (b) baseline (bpp: 0.371; PSNR: 32.38; MS-SSIM: 10.58), (c) GFRNet (PVTCB) (bpp: 0.371; PSNR: 32.70; MS-SSIM: 10.60), (d) GFRNet (QFMR-AM) (bpp: 0.363; PSNR: 32.71; MS-SSIM: 10.87), and (e) GFRNet (GVGM) (bpp: 0.367; PSNR: 32.58; MS-SSIM: 10.83).



(a) (b)

Fig. 17. Generalization experimental results of different modules on the San Francisco dataset. (a) PSNR and (b) MS-SSIM.

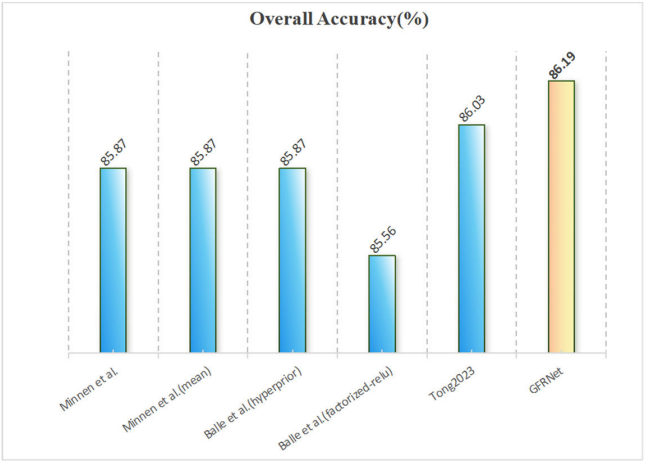


Fig. 18. OA of the reconstructed image obtained by different compression methods in remote sensing scene classification (the dataset used is NWPU-RESISC45).

fully proves that the removal of complex background noise, the increase of the multilevel nature of global features, and the expansion of the influence of global features have effectively promoted the compression of remote sensing images. It also strongly proves the generalization of each module. It is worth mentioning that the improvement of the PVTCB module on the baseline network Ballé et al. (factorized-relu) [32] is surprising. At 1.1 bpp, compared with Ballé et al. (factorized-relu) [32], Ballé et al. (factorized-relu) (GVGM) [32] and Ballé et al. (factorized-relu) (QFMR-AM) [32], respectively, improved 5.6% and 6.7%, while Ballé et al. (factorized-relu) (PVTCB) [32] reached an astonishing 11.7%. The reason for this phenomenon is that Ballé et al. (factorized-relu) [32] does not include a network that can efficiently extract multilevel global features, and PVTCB just makes up for this shortcoming.

H. Classification of Remote Sensing Scene Images

In this article, the reconstructed images obtained by different compression methods are used for remote sensing scene image classification, so as to verify the effectiveness of GFRNet from the perspective of application. The dataset selected is NWPU-RESISC45. The image compression methods used for comparison include Minnen et al. [31], Minnen et al. (mean) [31], Ballé et al. (hyperprior) [32], Ballé et al. (factorized-relu) [32], and Tong2023. The benchmark model for remote sensing scene classification is EMTCAL. In order to ensure the fairness of the experiment, the reconstructed images of different methods were obtained at a bit rate of 0.6 bpp. Fig. 18 shows OA obtained by different methods of reconstructed images for scene classification of remote sensing images. In terms of OA, the proposed GFRNet obtains the highest OA, which is higher 0.37% than that of Minnen et al. [31], higher 0.37% than that of Minnen et al. (mean) [31], higher 0.37% than that of Ballé et al. (hyperprior) [32], higher 0.73% than that of Ballé et al. (factorized-relu) [32], and higher than 0.19% than that of Tong2023.

Fig. 19 demonstrates the confusion matrices of reconstructed images of Minnen et al. [31], Minnen et al. (mean) [31], Ballé et al. (hyperprior) [32], Ballé et al. (factorized-relu) [32], Tong2023, and GFRNet when they are used for remote sensing scene classification. In Fig. 19, the classification effect of lake, beach, golf course, and intersection in GFRNet's CM is better than that of other comparison methods. This is mainly due to the fact that there are many

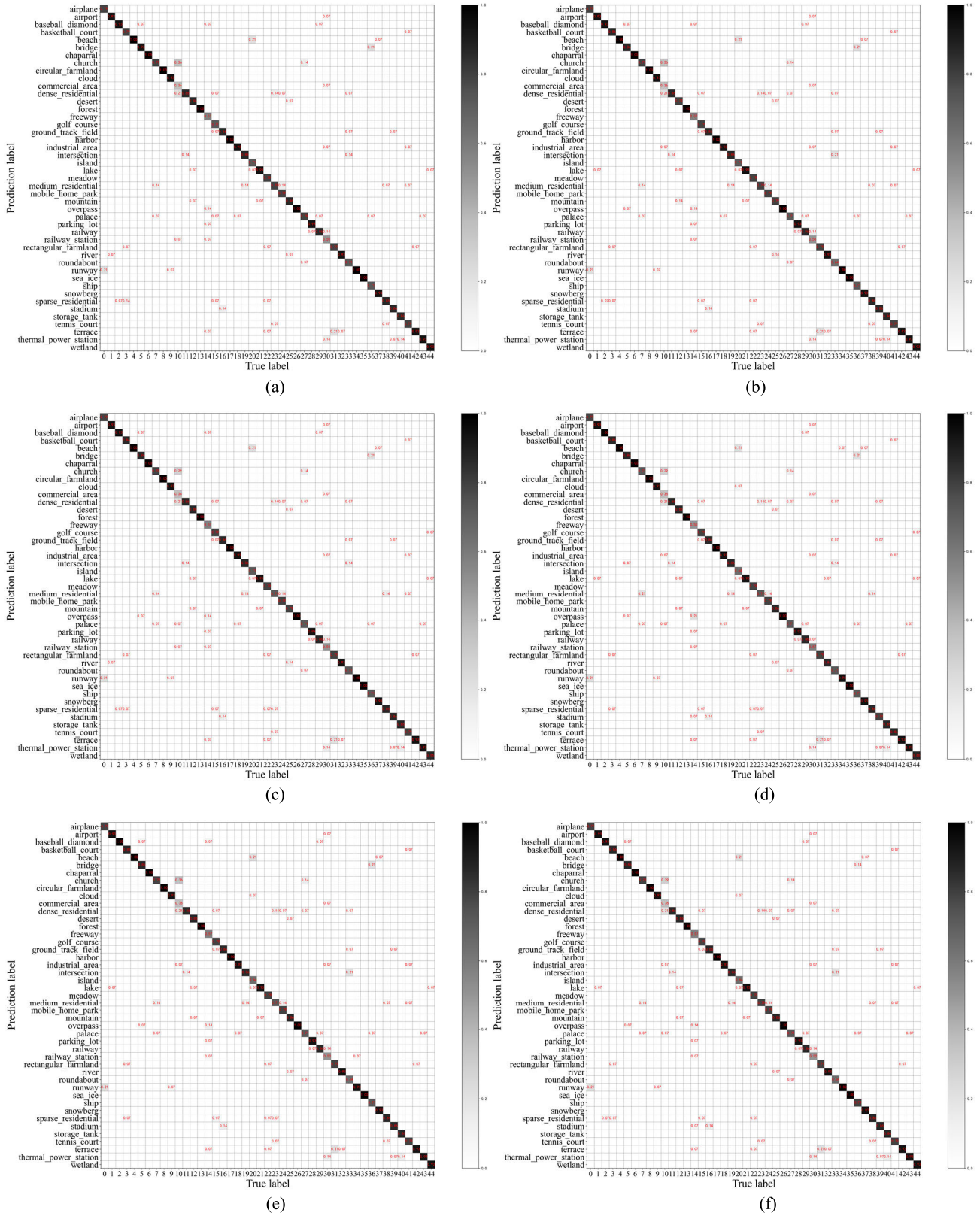


Fig. 19. CM of the reconstructed image by different methods. (a), (b), (c), (d), (e), and (f) correspond to Minnen et al. [31], Minnen et al. (mean) [31], Ballé et al. (hyperprior) [32], Ballé et al. (factorized-relu) [32], Tong2023, and GFRNet, respectively.

global features in these types of scenes, and PVTCB and GVGM in GFRNet just enhance the multilevel characteristics of global features and expand the scope of global features. Such high-quality global features greatly improve the quality of the final discriminant features. This is the reason why the reconstructed image obtained by the proposed GFRNet achieves the best performance in remote sensing scene classification.

TABLE V
COMPLEXITY COMPARISONS OF DIFFERENT COMPRESSION METHODS

	Minnen et al.	Minnen et al.(mean)	Balle et al. (hyperprior)	Balle et al. (factorized-relu)	Tong2023	GFRNet
Parameter	12.05M	11.04M	9.91M	5.56M	27.55M	11.001M
FLOPs	27.04G	26.78G	26.49G	25.95G	67.21G	11.97G
GPU Memory	3.0GB	2.8GB	2.8GB	1.6GB	4.2GB	4.2GB
Compression time	0.5931s	0.0712s	0.0731s	0.0321s	0.7014s	0.1794s
Reconstruction time	1.0781s	0.0745s	0.0775s	0.0390s	1.3214s	0.0895s

I. Complexity Analysis

In order to fairly compare the computational complexity and resource consumption of different compression methods, all compression methods are tested on the same device and in the same environment. The evaluation indicators include parameter, FLOPs, GPU memory, compression time, and reconstruction time. Here, the input image size is $3 \times 256 \times 256$. The experimental results are listed in Table V. Here, M stands for million, G stands for billion, GB stands for gigabyte, and S stands for seconds. It can be seen that the parameter of GFRNet is the third least among all methods. It is worth mentioning that although the parameter of Ballé et al. (factorized-relu) [32] is less than the proposed GFRNet, its PSNR and MS-SSIM are much lower than our method at the same bit rate. By comparing FLOPs, it can be found that GFRNet has achieved the fewest FLOPs, which are only 44.27%, 44.70%, 45.19%, 46.13%, and 17.81% of Minnen et al. [31], Minnen et al. (mean) [31], Ballé et al. (hyperprior) [32], Ballé et al. (factorized-relu) [32], and Tong2023, respectively. This fully illustrates the superiority of the GFRNet. Compared to GPU memory, GFRNet consumes the most of all methods. The reason for this is that the design of the two-branch structure requires a lot of parallel computing, which leads to a large GPU memory overhead. Comparing the compression time and the reconstruction time, it can be found that the time consumption of GFRNet is medium, but at the same bit rate, the PSNR and MS-SSIM of GFRNet are significantly better than other comparison methods. In addition, the compression time of GFRNet is twice as long as the reconstruction time. The main reason for this phenomenon is that the compressed network adopts a double-branch structure, which will have higher computational complexity than the single-branch reconstruction network. These experiments strongly demonstrate that GFRNet can achieve excellent rate distortion performance at a relatively low complexity.

V. CONCLUSION

In this article, a GFRNet is proposed for the compression of remote sensing images. First, a QFMR-AM is designed for noise reduction and multilevel information enhancement. Second, a PVTCB is constructed to capture multilevel global information. Third, a GVGM is proposed, which is utilized to calculate a novel Loss_{GVF}, and thus, a $\text{Loss}_{\text{Total}}$ of the network is constructed. Finally, all the modules and networks in this article are trained to focus more on global feature extraction under the guidance of a new rate-distortion function $\text{Loss}_{\text{Total}}$. Compared with other methods, the proposed GFRNet achieves the best rate distortion performance. Moreover, classification task is applied to evaluate the influence of the reconstructed images obtained by different compression methods on the application, and it is proved that the proposed GFRNet can provide the best classification performance. This shows that the proposed method can retain the important information in remote sensing images more effectively. In the future, we will explore how to integrate the global feature loss at more levels into $\text{Loss}_{\text{Total}}$ in a more reasonable way. In addition, we will further carry out more detailed hierarchical processing on the compression and reconstruction process of remote sensing images. By reducing the information gap between the latent representation feature and the specific task, the compression performance of remote sensing images can be further improved.

ACKNOWLEDGMENT

The authors would like to thank the handling editor and the anonymous reviewers for their careful reading and helpful remarks.

REFERENCES

[1] J. Feng et al., "Class-aligned and class-balancing generative domain adaptation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5509617.
[2] J. Feng, Q. Jiang, J. Zhang, Y. Liang, R. Shang, and L. Jiao, "CFDRM: Coarse-to-fine dynamic refinement model for weakly supervised moving vehicle detection in satellite videos," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5626413.
[3] W. Tang, F. He, A. K. Bashir, X. Shao, Y. Cheng, and K. Yu, "A remote sensing image rotation object detection approach for real-time environmental monitoring," *Sustain. Energy Technol. Assessments*, vol. 57, Jun. 2023, Art. no. 103270.
[4] P. H. T. Gama, H. N. Oliveira, J. Marcato, and J. Dos Santos, "Weakly supervised few-shot segmentation via meta-learning," *IEEE Trans. Multimedia*, vol. 25, pp. 7980–7991, 2022.
[5] W. Han et al., "A survey of machine learning and deep learning in remote sensing of geological environment: Challenges, advances, and opportunities," *ISPRS J. Photogramm. Remote Sens.*, vol. 202, pp. 87–113, Aug. 2023.
[6] X. Wang, C. Wang, X. Jin, and H. Wang, "Coordinated analysis of county geological environment carrying capacity and sustainable development under remote sensing interpretation combined with integrated model," *Ecotoxicol. Environ. Saf.*, vol. 257, Jun. 2023, Art. no. 114956.
[7] J. Kang, R. Fernandez-Beltran, X. Kang, J. Ni, and A. Plaza, "Noise-tolerant deep neighborhood embedding for remotely sensed images with label noise," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2551–2562, 2021.
[8] P. Zheng, J. Jiang, Y. Zhang, C. Zeng, C. Qin, and Z. Li, "CGC-net: A context-guided constrained network for remote-sensing image super resolution," *Remote Sens.*, vol. 15, no. 12, p. 3171, Jun. 2023.

[9] J. Nunez, O. Fors, X. Otazu, V. Pala, R. Arbiol, and M. T. Merino, "A wavelet-based method for the determination of the relative resolution between remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2539–2548, Sep. 2006.

[10] C. Sun, X. Fan, and D. Zhao, "Lossless recompression of JPEG images using transform domain intra prediction," *IEEE Trans. Image Process.*, vol. 32, pp. 88–99, 2023.

[11] C. De Cea-Dominguez, J. C. Moure-Lopez, J. Bartrina-Rapesta, and F. Auli-Llinas, "GPU-oriented architecture for an end-to-end image/video codec based on JPEG2000," *IEEE Access*, vol. 8, pp. 68474–68487, 2020.

[12] D. Báscones, C. González, and D. Mozos, "Hyperspectral image compression using vector quantization, PCA and JPEG2000," *Remote Sens.*, vol. 10, no. 6, p. 907, Jun. 2018.

[13] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34, no. 4, pp. 30–44, Apr. 1991.

[14] *JPEG2000 Official Software OpenJPEG*. Accessed: 2015. [Online]. Available: https://jpeg.org/jpeg2000/software.html

[15] B. Kovalenko, V. Lukin, S. Kryvenko, V. Naumenko, and B. Vozel, "BPG-based automatic lossy compression of noisy images with the prediction of an optimal operation existence and its parameters," *Appl. Sci.*, vol. 12, no. 15, p. 7555, Jul. 2022.

[16] F. Li, V. Lukin, O. Ieremeiev, and K. Okarma, "Quality control for the BPG lossy compression of three-channel remote sensing images," *Remote Sens.*, vol. 14, no. 8, p. 1824, Apr. 2022.

[17] M. Maldonado, "A new web oriented image format," Universitat Oberta de Catalunya, Barcelona, Spain, Tech. Rep. 81, 2010.

[18] Z. Wang, N. M. Nasrabadi, and T. S. Huang, "Spatial–spectral classification of hyperspectral images using discriminative dictionary designed by learning vector quantization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4808–4822, Aug. 2014.

[19] Y. Hu, W. Yang, Z. Ma, and J. Liu, "Learning end-to-end lossy image compression: A benchmark," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4194–4211, Mar. 2021.

[20] F. Aulí-Llinàs, M. W. Marcellin, J. Serra-Sagrista, and J. Bartrina-Rapesta, "Lossy-to-lossless 3D image coding through prior coefficient lookup tables," *Inf. Sci.*, vol. 239, pp. 266–282, Aug. 2013.

[21] S.-E. Qian, "Hyperspectral data compression using a fast vector quantization algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1791–1798, Aug. 2004.

[22] R. Pizzolante and B. Carpentieri, "Multiband and lossless compression of hyperspectral images," *Algorithms*, vol. 9, no. 1, p. 16, Feb. 2016.

[23] L. Thornton, J. Soraghan, R. Kutil, and M. Chakraborty, "Unequally protected SPIHT video codec for low bit rate transmission over highly error-prone mobile channels," *Signal Process., Image Commun.*, vol. 17, no. 4, pp. 327–335, Apr. 2002.

[24] J. Feng, G. Bai, D. Li, X. Zhang, R. Shang, and L. Jiao, "MR-selection: A meta-reinforcement learning approach for zero-shot hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5500320.

[25] J. Feng, Z. Gao, R. Shang, X. Zhang, and L. Jiao, "Multi-complementary generative adversarial networks with contrastive learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5520018.

[26] R. La Grassa, C. Re, G. Cremonese, and I. Gallo, "Hyperspectral data compression using fully convolutional autoencoder," *Remote Sens.*, vol. 14, no. 10, p. 2472, May 2022.

[27] J. Liu, F. Yuan, C. Xue, Z. Jia, and E. Cheng, "An efficient and robust underwater image compression scheme based on autoencoder," *IEEE J. Ocean. Eng.*, vol. 48, no. 3, pp. 925–945, Jul. 2023.

[28] V. A. de Oliveira et al., "Reduced-complexity end-to-end variational autoencoder for on board satellite image compression," *Remote Sens.*, vol. 13, no. 3, p. 447, Jan. 2021.

[29] Q. Xu, Y. Xiang, and Z. X. Di, "Synthetic aperture radar image compression based on a variational autoencoder," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.

[30] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimization of nonlinear transform codes for perceptual quality," in *Proc. Picture Coding Symp. (PCS)*, Dec. 2016, pp. 1–5.

[31] D. Minnen, J. Ballé, and G. D. Toderici, "Joint autoregressive and hierarchical priors for learned image compression," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–10.

[32] J. Ballé, D. Minnen, S. Singh, S. Jin Hwang, and N. Johnston, "Variational image compression with a scale hyperprior," 2018, *arXiv:1802.01436*.

[33] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Learned image compression with discretized Gaussian mixture likelihoods and attention modules," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7939–7948.

[34] Z. Guo, Z. Zhang, R. Feng, and Z. Chen, "Causal contextual prediction for learned image compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 2329–2341, Apr. 2022.

[35] T. Chen, H. Liu, Z. Ma, Q. Shen, X. Cao, and Y. Wang, "End-to-end learnt image compression via non-local attention optimization and improved context modeling," *IEEE Trans. Image Process.*, vol. 30, pp. 3179–3191, 2021.

[36] M. Cao et al., "Entropy modeling via Gaussian process regression for learned image compression," in *Proc. Data Compress. Conf. (DCC)*, Mar. 2022, pp. 163–172.

[37] D. Liu, X. Sun, F. Wu, and Y.-Q. Zhang, "Edge-oriented uniform intra prediction," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1827–1836, Oct. 2008.

[38] F. Kong, T. Cao, Y. Li, D. Li, and K. Hu, "Multi-scale spatial–spectral attention network for multispectral image compression based on variational autoencoder," *Signal Process.*, vol. 198, Sep. 2022, Art. no. 108589.

[39] C. Fu, B. Du, and L. Zhang, "SAR image compression based on multi-resblock and global context," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[40] J. Li and X. Hou, "Object-fidelity remote sensing image compression with content-weighted bitrate allocation and patch-based local attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 2004314.

[41] J. Gao, Q. Teng, X. He, Z. Chen, and C. Ren, "Mixed entropy model enhanced residual attention network for remote sensing image compression," *Neural Process. Lett.*, vol. 55, no. 7, pp. 10117–10129, Dec. 2023.

[42] S. Xiang, Q. Liang, and L. Fang, "Discrete wavelet transform-based Gaussian mixture model for remote sensing image compression," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 3000112.

[43] S. Xiang and Q. Liang, "Remote sensing image compression based on high-frequency and low-frequency components," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5604715.

[44] Y. Guo, Y. Chong, Y. Ding, S. Pan, and X. Gu, "Learned hyperspectral compression using a student's T hyperprior," *Remote Sens.*, vol. 13, no. 21, p. 4390, Oct. 2021.

[45] M. Zhao, R. Yang, M. Hu, and B. Liu, "Deep learning-based technique for remote sensing image enhancement using multiscale feature fusion," *Sensors*, vol. 24, no. 2, p. 673, Jan. 2024.

[46] G. Sumbul, J. Xiang, and B. Demir, "Towards simultaneous image compression and indexing for scalable content-based retrieval in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5630912.

[47] H. Wang, L. Liao, J. Xiao, W. Lin, and M. Wang, "Uplink-assist downlink remote sensing image compression via historical referecing," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5621415.

[48] M. Liu, L. Tang, L. Fan, S. Zhong, H. Luo, and J. Peng, "CARNet: Context-aware residual learning for JPEG-LS compressed remote sensing image restoration," *Remote Sens.*, vol. 14, no. 24, p. 6318, Dec. 2022.

[49] W. Ye, W. Lei, W. Zhang, T. Yu, and X. Feng, "GFSCompNet: Remote sensing image compression network based on global feature-assisted segmentation," *Multimedia Tools Appl.*, vol. 83, no. 25, pp. 67103–67127, Jan. 2024.

[50] S. Xiang, Q. Liang, and P. Tang, "Task-oriented compression framework for remote sensing satellite data transmission," *IEEE Trans. Ind. Informat.*, vol. 20, no. 3, pp. 3487–3496, Mar. 2024.

[51] C. Fu and B. Du, "Remote sensing image compression based on the multiple prior information," *Remote Sens.*, vol. 15, no. 8, p. 2211, Apr. 2023.

[52] P. Han, B. Zhao, and X. Li, "Edge-guided remote-sensing image compression," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5524515.

[53] K. Cheng, Y. Zou, Y. Zhao, H. Jin, and C. Li, "A remote sensing satellite image compression method based on conditional generative adversarial network," in *Image and Signal Processing for Remote Sensing XXIX*, vol. 12733. Bellingham, WA, USA: SPIE, 2023, pp. 322–331.

[54] T. Pan, L. Zhang, Y. Song, and Y. Liu, "Hybrid attention compression network with light graph attention module for remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[55] W. Wang et al., "Pyramid vision transformer: A versatile backbone for dense prediction without convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 568–578.
[56] [Online]. Available: https://resources.maxar.com/product-samples/analysis-ready-data-san-francisco-california
[57] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
[58] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, Nov. 2010, pp. 270–279.
[59] F. Bellard. *BPG Image Format*. Accessed: 2018. [Online]. Available: http://bellard.org/bpg/
[60] K. Tong, Y. Wu, Y. Li, K. Zhang, L. Zhang, and X. Jin, "QVRF: A quantization-error-aware variable rate framework for learned image compression," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2023, pp. 1310–1314.
[61] X. Tang, M. Li, J. Ma, X. Zhang, F. Liu, and L. Jiao, "EMTCAL: Efficient multiscale transformer and cross-level attention learning for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5626915.

Cuiping Shi (Member, IEEE) received the M.S. degree from Yangzhou University, Yangzhou, China, in 2007, and the Ph.D. degree from Harbin Institute of Technology (HIT), Harbin, China, in 2016.

From 2017 to 2020, she was a Post-Doctoral Researcher with the College of Information and Communications Engineering, Harbin Engineering University, Harbin. Since 2024, she has been working with the College of Information Engineering, Huzhou University, Huzhou, China. She is currently a Professor with the Department of Communication Engineering, Qiqihar University, Qiqihar, China. She has published two academic books about remote sensing image processing and more than 90 articles in journals and conference proceedings. Her main research interests include remote sensing image processing pattern recognition and machine learning.

Dr. Shi's Doctoral Dissertation won the Nomination Award of Excellent Doctoral Dissertation of Harbin University of Technology (HIT) in 2016.

Kaijie Shi received the bachelor's degree from Heilongjiang University of Science and Technology, Harbin, China, in 2021. He is currently pursuing the master's degree with Qiqihar University, Qiqihar, China.

His research interests include remote sensing image compression and machine learning.

Fei Zhu received the bachelor's degree from Luoyang Institute of Science and Technology, Luoyang, China, in 2021. He is currently pursuing the master's degree with Qiqihar University, Qiqihar, China.

His research interests include hyperspectral image processing and machine learning.

Zexin Zeng received the bachelor's degree from Guangzhou Maritime University, Guangdong, China, in 2021. He is currently pursuing the master's degree with Qiqihar University, Qiqihar, China.

His research interests include hyperspectral image processing and machine learning.

Mengxiang Ding received the bachelor's degree from WeiFang Medical University, Weifang, China, in 2021, and the master's degree from Qiqihar University, Qiqihar, China, in 2024.

His research interests include remote sensing image processing, machine learning, and deep learning.

Zhan Jin received the B.S. degree in electrical and information engineering from Heilongjiang University, Harbin, China, in 2005, and the M.S. and Ph.D. degrees in information and communication engineering from Harbin Engineering University, Harbin, in 2009 and 2020, respectively.

She has been working with the College of Communication and Electronic Engineering, Qiqihar University, Qiqihar, China, since 2009, and is currently an Associate Professor. She has published ten articles and one academic book about sparse adaptive filtering. Her research interests include signal processing and sparse adaptive filtering.

SCI 收录报告

经查 **Web of Science-Core Collection**，石翠萍提供的如下文章已经被 **SCI-Expanded**（科学引文索引）收录，其收录记录简要信息摘选如下：

标题: An Enhanced Global Feature-Guided Network Based on Multiple Filtering Noise Reduction for Remote Sensing Image Compression

作者: Shi, CP (Shi, Cuiping); Shi, KJ (Shi, Kaijie); Zhu, F (Zhu, Fei); Zeng, ZX (Zeng, Zexin); Ding, MX (Ding, Mengxiang); Jin, Z (Jin, Zhan)

来源出版物: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING **卷:** 62 **文献号:** 5646220 **DOI:** 10.1109/TGRS.2024.3483871 **Published Date:** 2024

Web of Science 核心合集中的 "被引频次": 0

被引频次合计: 0

入藏号: WOS:001350106800040

语言: English

文献类型: Article

地址: [Shi, Cuiping] Huzhou Univ, Coll Informat Engn, Huzhou 313000, Peoples R China. [Shi, Cuiping; Shi, Kaijie; Zhu, Fei; Zeng, Zexin; Ding, Mengxiang; Jin, Zhan] Qiqihar Univ, Dept Commun Engn, Qiqihar 161000, Peoples R China.

通讯作者地址: Shi, CP (通讯作者)，Huzhou Univ, Coll Informat Engn, Huzhou 313000, Peoples R China.

电子邮件地址: shicuiping@qqhru.edu.cn; 2022910313@qqhru.edu.cn; 2022935750@qqhru.edu.cn; 2022910311@qqhru.edu.cn; 2021910321@qqhru.edu.cn; jinzhan@qqhru.edu.cn

Affiliations: Huzhou University; Qiqihar University

IDS 号: L3Y3M

ISSN: 0196-2892

eISSN: 1558-0644

来源出版物页码计数: 20

特此证明

湖州师范学院图书馆

查证人：吕佳燕

2024 年 11 月 27 日

湖州师范学院图书馆 查索专用章