

A Multilevel Domain Similarity Enhancement Guided Network for Remote Sensing Image Compression

Cuiping Shi¹, Member, IEEE, Kaijie Shi¹, Fei Zhu¹, Zexin Zeng, and Ligu Wang², Member, IEEE

Abstract—Remote sensing image compression networks aim to enhance the similarity between the input image and the reconstructed image. The current network rarely considers the potential relationship between the compression features of different levels and the reconstruction features of the corresponding levels, which limits the improvement of remote sensing image compression performance. In this article, a concept of multilevel domain similarity is first proposed, which fully develops the multilevel domain similarity between the encoding and decoding processes to improve the quality of reconstructed images. On this basis, a multilevel domain similarity enhancement guided network (MDSNet) is proposed for remote sensing image compression. First, an efficient compression baseline network (BaselineA) was proposed, which realizes efficient image compression with low computational complexity. Second, a multilevel domain similarity enhancement module (MDEM) was designed, which improved the quality of the reconstructed image by enhancing the multilevel domain similarity. Third, a global information-enhanced attention module (GIE-AM) was constructed to enhance channel features and global features. Finally, under the guidance of the total loss ($\text{Loss}_{\text{Total}}$), which is constructed by the proposed MDEM loss (MDEM-Loss), an effective compression was implemented by the whole network for remote sensing image compression. Experimental results show that compared with some advanced compression models, the proposed MDSNet can significantly improve compression performance with lower computational complexity. In addition, the reconstructed images obtained by the proposed method can provide better classification performance, which further proves that the proposed MDSNet helps to preserve more important features of remote sensing images during the compression process.

Index Terms—Learned compression model, multihead self-attention (MHSA), multilevel domains, remote sensing image compression.

Received 19 June 2024; revised 11 September 2024; accepted 8 October 2024. Date of publication 18 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 Fundamental Research Funds in Heilongjiang Provincial Universities under Grant 145109145, and in part by Qiqihar University Graduate Innovative Research Project under Grant QUZLTS_CX2023025. (Corresponding author: Cuiping Shi.)

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

Kaijie Shi, Fei Zhu, and Zexin Zeng 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).

Ligu Wang is with the College of Information and Communication Engineering, Dalian Nationalities University, Dalian 116000, China (e-mail: wangliguo@hrbeu.edu.cn).

Digital Object Identifier 10.1109/TGRS.2024.3483312

1558-0644 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See <https://www.ieee.org/publications/rights/index.html> for more information.

I. INTRODUCTION

REMOTE sensing images are digital representations of the earth's surface information [1], [2], [3], [4], which have many features that are difficult to reflect in natural images, including land cover type, topography, landform, and temperature. Therefore, they are widely used in many fields such as environmental monitoring, meteorology, and geological sciences [5], [6], [7], [8]. With the rapid development of imaging technology, the spatial and spectral resolution of remote sensing images continue to increase, and the amount of data also increases exponentially. In addition, due to the influence of factors such as imaging angle, shooting distance, and atmospheric refraction, remote sensing images have the characteristics of high information entropy, rich texture information, mixed high-frequency and low-frequency features, and weak correlation, which make it difficult for traditional image compression methods to effectively compress remote sensing images [9], [10], [11].

At present, some research results have been achieved by common traditional image compression methods. For example, Báscones et al. [13] proposed a method that combines principal component analysis with JPEG2000 [12] to compress hyperspectral image data. Therefore, efficient remote sensing image compression networks need to be designed and developed urgently. In addition, BPG [14] and WebP [15], which have better performance, have also been born in the field of image compression. Among them, JPEG [16] and JPEG2000 consist of three parts, including image transformation, quantization, and entropy encoding. In general, first, the image is compressed and dequantized. Second, it retains valuable information through quantification. Finally, the entropy coding is used to compress the decorrelated coefficients. Traditional image compression methods can be divided into vector quantization [17], predictive coding [18], and transform coding [19]. Shannon's rate-distortion theory is based on vector quantization coding. Qian [20] proposed a fast vector quantization algorithm for multispectral image compression, which converts the input vectors into code word indices that match the input vectors in the codebook for data transmission and storage. Three-dimensional multiband linear predictor (MBLP) adopts prediction-based technology. It first eliminates the spatial redundancy of images and then predicts the frequency band. Finally, it encodes the prediction residuals through an entropy coder [21].

Three-dimensional set partitioning in hierarchical trees (SPIHT) is a transform method used for 3-D image compression, which applies 3-D wavelet transform in both spatial and spectral domains [22]. However, the application of traditional image compression methods to remote sensing image compression has the following problems, such as high computational complexity and lack of targeted processing for the characteristics of remote sensing images, including high information entropy, rich texture, mixed high-frequency and low-frequency features, and weak correlation. These limitations restrict the improvement of the compression performance of remote sensing images.

In order to seek breakthroughs, relevant researchers focus on neural network technology, which has become hot in recent years. Classical deep learning-based image compression frameworks mainly include autoencoder (AE) [23], [24] and variational AE (VAE) [25], [26]. The main reason is that the image compression task mainly consists of two symmetrical data processing processes, namely compression and reconstruction. There is a similar, reversible relationship between these two tasks, which means that networks with symmetrical codec structures will perform better. Among them, the VAE-based framework has more powerful image reconstruction capabilities. This is due to the fact that VAE has a continuous mapping space that AE does not have and can reconstruct smooth images. In recent years, many excellent VAE-based baseline networks have been proposed [27], [28], [29], [30], [31], [32], [33], which have achieved rate–distortion performance better than traditional image compression methods. The VAE-based image compression network consists of three parts, including encoder, entropy encoding, and decoder. These compression frameworks typically use neural networks to preliminarily compress blocks of data, then map pixel data into quantized representations, and finally use traditional encoding to compress the data into the smallest form of existence (bitstream). In addition, some compression models introduce entropy models into the framework, such as Laplace models, single Gaussian models, Gaussian mixture models, and factorized entropy models, to introduce prior information and achieve more accurate modeling [28], [30], [34], [35]. Therefore, many scholars have adopted deep learning-based compression models to achieve efficient remote sensing image compression [36], [37], [38], [39], [40]. It is worth mentioning that the common deep learning techniques used for remote sensing image compression mainly include three types: CNN-based image compression methods, Transformer-based image compression methods, and GAN-based image compression methods. Tang et al. [41] proposed an end-to-end image compression method by fusing graph attention and asymmetric convolutional neural networks (ACNNs). This method effectively overcomes the problem of common CNN paying too much attention to the local features of the image, promotes information interaction, and fully considers the interdependence between channels and position information. Li et al. [42] proposed a deep learning-based image compression network based on a Vision Transformer (ViT). This method blocks the input image and adopts different types of Transformer blocks in the encoder and decoder to achieve efficient image compression. At low bit rates, it achieves relatively good

rate–distortion performance. Han et al. [40] proposed a new end-to-end framework, namely the edge-guided adversarial network, which aims to preserve sharp edge information and texture information simultaneously. It leverages edge fidelity to constrain and guide the network in optimizing the reconstruction of image structures, aiming to address the issue of local smoothing in existing methods. The above methods achieve good rate–distortion performance, but they all enhance the quality of reconstructed images from a single factor (graph structure, long-distance context information, edge information) without comprehensively improving the quality of reconstructed images.

The essence of the remote sensing image compression framework is to improve the similarity between the original image and the reconstructed image. The similarity between the original image and the reconstructed image depends on a variety of factors, such as the reconstruction quality of high-frequency information, the recovery of long-distance contextual information, the influence of noise, and the strength of the correlation between different features. How to comprehensively consider various factors to improve the quality of remote sensing image reconstruction has become a serious challenge.

In order to alleviate the above problems, this article proposes a multilevel domain similarity enhancement guided network (MDSNet) for remote sensing image compression. It mainly enhances the overall similarity between the original remote sensing image and the reconstructed image (the similarity between the highest-level domains) from the perspective of enhancing the multilevel domain similarity of the front, so as to comprehensively improve the quality of the reconstructed remote sensing image. The MDSNet innovates and optimizes the network from three aspects. In the first aspect, the essence of the remote sensing image compression network is to enhance the similarity between the input image and the reconstructed image. In this article, the similarity of the highest-level domain is improved from the perspective of enhancing the multilevel domain similarity of the front. In this article, we consider the compression process of the encoder and the reconstruction process of the decoder as multilevel compression and multilevel reconstruction, respectively, and treat the multilevel image in the compression process and the multilevel image in the reconstruction process as two different domains. This article refers to this similarity between compressed and reconstructed domains at multiple levels as multilevel domain similarity. First, a multilevel feature interaction module (MFIM) and a multichannel information extraction module (MIEM) are designed. These two modules are used to form a multilevel domain similarity enhancement module (MDEM), so as to realize multichannel feature extraction and multilevel feature interaction. Second, this article proposes a MDEM loss (MDEM-Loss) to reconstruct the objective function of the network, so as to guide the compression model to compress and reconstruct remote sensing images with high quality. In the second aspect, remote sensing images contain abundant global information and channel characteristics. If this information is ignored, the model may lose its understanding of important contexts such as geomorphological features, and the current mainstream

methods do not achieve efficient extraction and combination of these two types of information. In this article, first, a channel extrusion block (CEB) is designed to enhance the channel information efficiently. In addition, the corresponding multi-head self-attention (MHSA) is used to efficiently extract global information. Finally, on the basis of these two modules, a global information-enhanced attention module (GIE-AM) was designed to enhance the extraction and combination of channel features and global features. In the third aspect, CNN has powerful feature extraction capabilities [43], [44], [45]. Researchers can adjust the size of the convolution kernel to control the range of receptive field, so as to adjust the focus of the network on extracting high- and low-frequency information. However, high-quality global information extraction means that larger convolutional kernels are selected, which leads to a sharp increase in the number of parameters. In addition, common networks have not effectively analyzed and processed the data distribution of images, which means that they lack research on the latent relationship between the number of channels and the image size in the process of compression and reconstruction. Therefore, this article designs an efficient remote sensing image compression baseline network (BaselineA). In view of the characteristics of remote sensing images, such as high information entropy, rich texture information, mixed high-frequency and low-frequency features, and weak correlation, high-quality image compression and reconstruction under low complexity are realized by selecting the reasonable convolution kernel sizes and redistributing the number of channels. In summary, this article constructs a low-complexity and high-performance MDSNet based on the proposed MFIM, MIEM, MDEM, MDEM-Loss, CEB, MHSA, GIE-AM, and BaselineA.

In this study, a large number of experiments were carried out on remote sensing image datasets such as San Francisco [46], NWPU-RESISC45 [47], and UC-Merced [48]. Experimental results show that compared with other methods, the MDSNet proposed in this article has superior performance in the evaluation indicators such as peak signal-to-noise ratio (PSNR) and multiscale structural similarity index metric (MS-SSIM). In addition, the effectiveness of MFIM, MIEM, MDEM, MDEM-Loss, CEB, MHSA, GIE-AM, and BaselineA was verified by a variety of ablation experiments.

The main contributions of this article are summarized as follows.

- 1) In this study, an MDEM was proposed, which includes two core components: MFIM and MIEM. The MDEM can realize multichannel feature extraction and multi-level feature interaction. In addition, an MDEM-Loss was proposed for reconstructing the objective function of the network. Therefore, the similarity between domains at the highest level is improved through this enhancement of the multilevel domain similarity of the front.
- 2) A GIE-AM was designed, whose core components mainly include CEB and MHSA, to efficiently enhance the extraction and fusion of channel features and global features.
- 3) An efficient compression baseline network (BaselineA) for remote sensing images was constructed. The network

considers the characteristics of high information entropy, rich texture, mixed high-frequency and low-frequency features, and weak correlation of remote sensing images and achieves excellent compression performance at low complexity by reasonably selecting the size of the convolution kernel and redistributing the number of channels.

- 4) In this study, MFIM, MIEM, MDEM, MDEM-Loss, CEB, MHSA, GIE-AM, and factorized entropy model are effectively embedded into BaselineA to construct a low-complexity and high-performance MDSNet. Through a large number of experiments on three datasets, the proposed MDSNet has been proven to significantly improve compression performance for remote sensing images with lower computational complexity.

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

II. METHODOLOGY

In this section, the proposed MDSNet is described in detail, as well as the efficient basic framework BaselineA and modules such as MFIM, MIEM, MDEM, MDEM-Loss, CEB, MHSA, and GIE-AM.

A. Overall Framework of the Proposed MDSNet

The proposed MDSNet deals with the task of remote sensing image compression from a new perspective of enhancing the multilevel domain similarity of the front. It mainly uses the efficient frameworks BaselineA, MDEM, and GIE-AM to achieve high-quality remote sensing image compression. The MDEM consists of three submodules, namely MFIM, MIEM, and MDEM-Loss. The GIE-AM consists of two submodules, namely CEB and MHSA. In addition, the loss function $\text{Loss}_{\text{total}}$ is established to coordinate the compression and reconstruction between the basic network and the proposed module, so as to achieve higher compression performance.

The proposed MDSNet uses VAE as the basic framework. MDSNet mainly consists of the following parts: encoder, decoder, MDEM-Loss, probability model, and an improved rate-distortion optimization, as shown in Fig. 1. In the image compression part, the encoder maps the remote sensing image data blocks into preliminarily compressed latent representation features by combining MFIM, MIEM, GIE-AM, and 2-D convolutions with different scales and numbers of channels. Then, the statistical redundancy is further removed by quantization, arithmetic coding (AE), and probability model, and the final bit stream is obtained. In the image decompression part, the model uses the mean μ_i and scale σ_i^2 parameters in the probability model, as well as arithmetic decoding (AD), GIE-AM, MFIM, MIEM, and 2-D convolutions with different scales and numbers of channels to process the bit stream, so as to reconstruct the image with high quality. All of the above

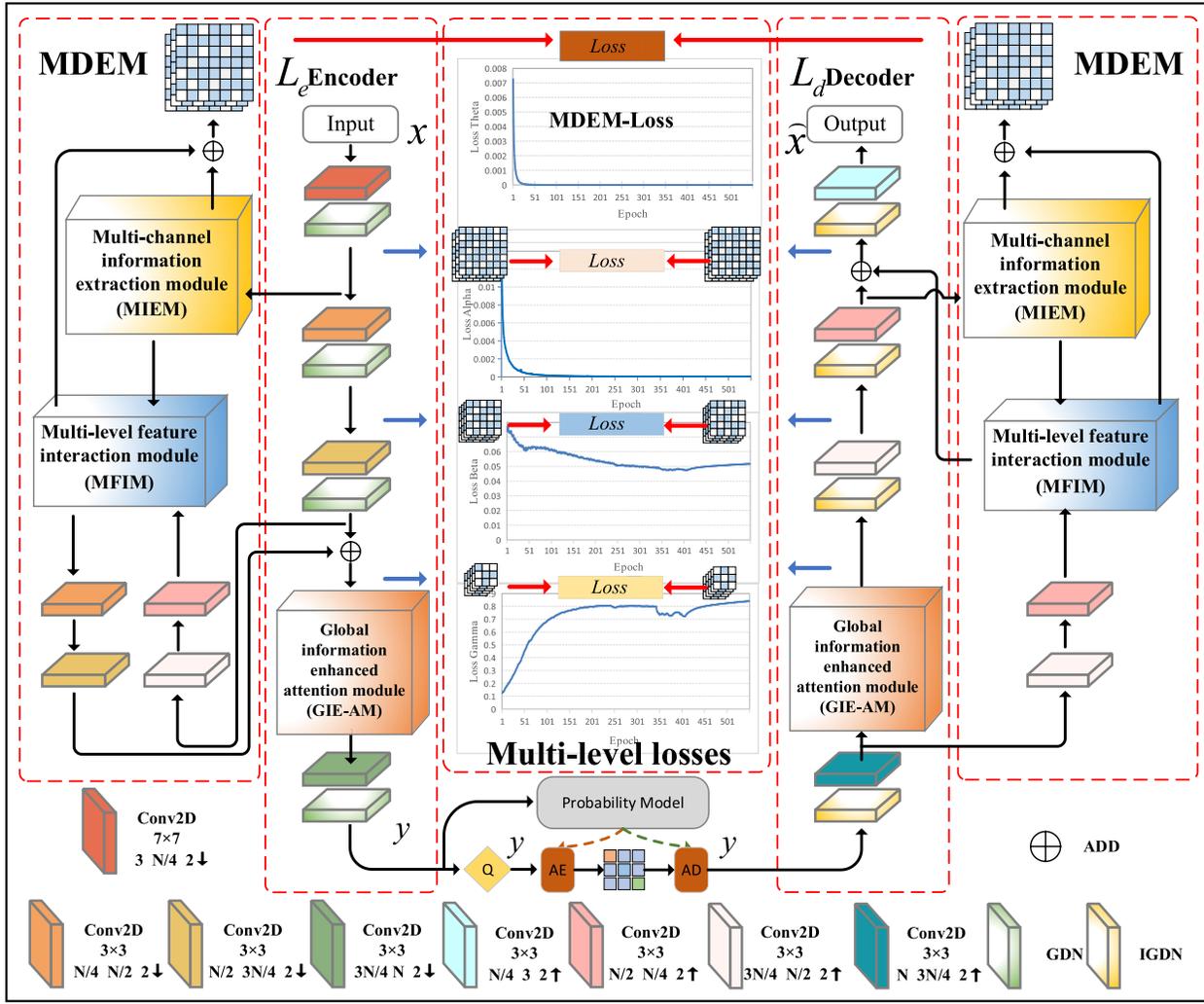


Fig. 1. Overall structure diagram of the proposed MDSNet.

work is carried out in an orderly and efficient manner under the guidance of the loss function $Loss_{total}$.

In Fig. 1, L_e and L_d are the main encoder and main decoder, respectively, which are utilized to learn the latent representation features of the remote sensing image. The probability model mainly includes a hyperprior network (hyper encoder, hyper decoder), Q (quantizer), AE, and AD, and the data between AE and AD are the smallest form of data (bitstream) existing in this model. In this article, the hyperprior network is adopted to learn the entropy model on which entropy coding depends. It is also used to generate the parameters of the entropy model (i.e., mean parameter μ_i and scale parameter σ_i^2), which is modeled as conditional Gaussian. The modules on the left and right sides represent MDEM, mainly including MFIM, MIEM, MDEM-Loss, and various convolutions. The MDEM-Loss is obtained by calculating the loss between the multilevel feature map obtained by MDEM in the compressed part and the multilevel feature map obtained by MDEM in the reconstructed part. The loss used here is a mean squared error (MSE), and $Loss_{\delta}$ is adopted to represent the MDEM-Loss. The multilevel losses in the middle represent the difference between the feature map of the compressed

domain and the feature map of the reconstructed domain when the domain is at the same level. $Loss_{\alpha}$ represents the difference between the first-order domains, which is called the first-level difference loss in this article. Similarly, $Loss_{\beta}$ represents the second-level difference loss. $Loss_{\gamma}$ represents the third-level difference loss. These three levels of loss are collectively referred to as the multilevel domain similarity of the front. The difference between the original image and the reconstructed image is called the zero-level difference loss or superlative-level difference loss. MSE can be expressed as

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

where m denotes the number of pixels, \hat{X} denotes the reconstructed image, and X denotes the original image. GIE-AM denotes the global information-enhanced attention module, which mainly includes CEB and MHSA. N denotes the number of channels, \downarrow denotes the downsampling, \uparrow denotes the upsampling. GDN denotes the generalized splitting normalization function, and IGDN denotes the inverse of GDN. They are nonlinear activation functions that are more suitable

for the normalization of image data than other normalization functions.

Specifically, based on the characteristics of remote sensing images, such as complex background, rich information, and weak correlation between features, this article studies the low-complexity and efficient remote sensing image baseline network (BaselineA), MDEM based on multiscale enhancement of similarity, and GIE-AM based on enhanced channel features and global features. Compared with the common remote sensing image compression model, BaselineA adopts a smaller convolutional kernel size to greatly reduce the number of parameters. At the same time, in the main encoder, some large convolutional kernels are used to compensate for the important long-distance context information loss caused by the use of small-size convolutional kernels. In addition, the number of channels of the convolutional layer in the overall framework was redistributed in order to extract the appropriate features at the appropriate compression level. MDEM is a multilevel domain feature map extraction module used to enhance the multilevel domain similarity of the front with MDEM-Loss, which mainly includes two components: MIEM for multichannel feature extraction and MFIM for multilevel feature interaction. MIEM is in the first-level domain, with a large feature map and more spatial information. Therefore, MIEM uses multiple point convolutions to extract and superimpose spatial information, so as to enhance spatial information. In addition, AvgPool2d and MaxPool2d are adopted to pool the third-level domain feature map and the first-level domain feature map obtained by MIEM, respectively. Then, the pooled feature maps are fused and strengthened by attention mechanism and residual operation. Finally, a multilevel domain feature map can be used for feedback to the network and calculation of MDEM-Loss. In addition, GIE-AM is mainly used to efficiently enhance channel features and global features, and its core components include CEB for enhancing channel information and MHSA for enhancing global context information. Then, CEB combined the attention mechanism and residual operation to reduce and ascend the dimension of the high-dimensional feature map. This can enhance the channel information in the remote sensing image and obtain a feature map for input to the MHSA. MHSA adopts a self-attention mechanism that can efficiently model long-distance information and the multiple heads to speed up the training speed of the network and the fusion of information on different subspaces. In general, MDSNet efficiently explores the latent relationship between remote sensing image compression and reconstruction through BaselineA, MDEM, and GIE-AM, so as to achieve high-quality compression and reconstruction work.

B. Multilevel Domain Similarity Enhancement Module

At present, the common remote sensing image compression methods are improved from one or several perspectives, such as strengthening the extraction of high-frequency information, improving the quality of global visual features, removing the complex background noise of remote sensing images, enhancing the correlation between features, and reducing spatial and spectral redundant information. However, these

methods are not very efficient in compressing and reconstructing various features of remote sensing images at the same time. Therefore, from the perspective of improving the overall similarity between the original remote sensing image and the reconstructed image, this study enhances the similarity between the highest-level domains by enhancing the multilevel domain similarity of the front. Therefore, this article constructs MDEM for enhancing multilevel domain similarity, which mainly includes three core components, namely MIEM for extracting multichannel information, MFIM for multilevel feature interaction, and MDEM-Loss for guiding the network to enhance multilevel domain similarity. The process of MIEM can be represented as

$$I_{MIEM} = W_{CONV1 \times 1} * Input + W_{CONV1 \times 1} * Input + W_{CONV1 \times 1} * Input. \quad (2)$$

Here, Input represents the input image data block, * represents the convolution, W represents the weight parameter of the point convolution, and I_{MIEM} represents the output of MIEM.

The process of MFIM in the main encoder can be represented as

$$\begin{aligned} I_{MFIM(Encoder)} &= I_{MIEM(Encoder)} \\ &\oplus (W_{MaxPool2d}(I_{MIEM(Encoder)}) + W_{AvgPool2d}(InputD) \\ &\quad \odot (W_{Softmax}(W_{MaxPool2d}(I_{MIEM(Encoder)})) \\ &\quad + W_{AvgPool2d}(InputD))). \end{aligned} \quad (3)$$

Here, $I_{MIEM(Encoder)}$ represents the output of the MIEM of the main encoder, W represents the weight parameters of the convolutional layer, the Softmax layer, and the pooling layer, \oplus represents the pointwise addition, \odot represents the Hadamard product, and $I_{MFIM(Encoder)}$ represents the multilevel domain feature map of the main encoder.

The process of MFIM in the main decoder is similar to that of the MFIM in the main encoder, except that the two input positions of the MFIM are reversed. This process can be represented as

$$\begin{aligned} I_{MFIM(Decoder)} &= I_{MIEM(Decoder)} \\ &\oplus (W_{AvgPool2d}(I_{MIEM(Decoder)}) + W_{MaxPool2d}(InputA) \\ &\quad \odot (W_{Softmax}(W_{AvgPool2d}(I_{MIEM(Decoder)})) \\ &\quad + W_{MaxPool2d}(InputA))). \end{aligned} \quad (4)$$

Here, $I_{MFIM(Decoder)}$ represents the multilevel domain feature map of the main decoder.

The process of MDEM-Loss can be expressed as

$$MDEM-LOSS = L_{MSE}(I_{MFIM(Encoder)}, I_{MFIM(Decoder)}). \quad (5)$$

Here, L_{MSE} denotes the loss measured by MSE. In this article, $Loss\delta$ is used to represent MDEM-Loss for convenience.

The overall structure of MDEM is shown in Fig. 2. It mainly includes MIEM, MFIM, and MDEM-Loss. MIEM uses three-point convolutions for the extraction and enhancement of multichannel information. The reason why MIEM uses multiple channels instead of a single channel for spatial information extraction is that the input to MIEM is a large-scale

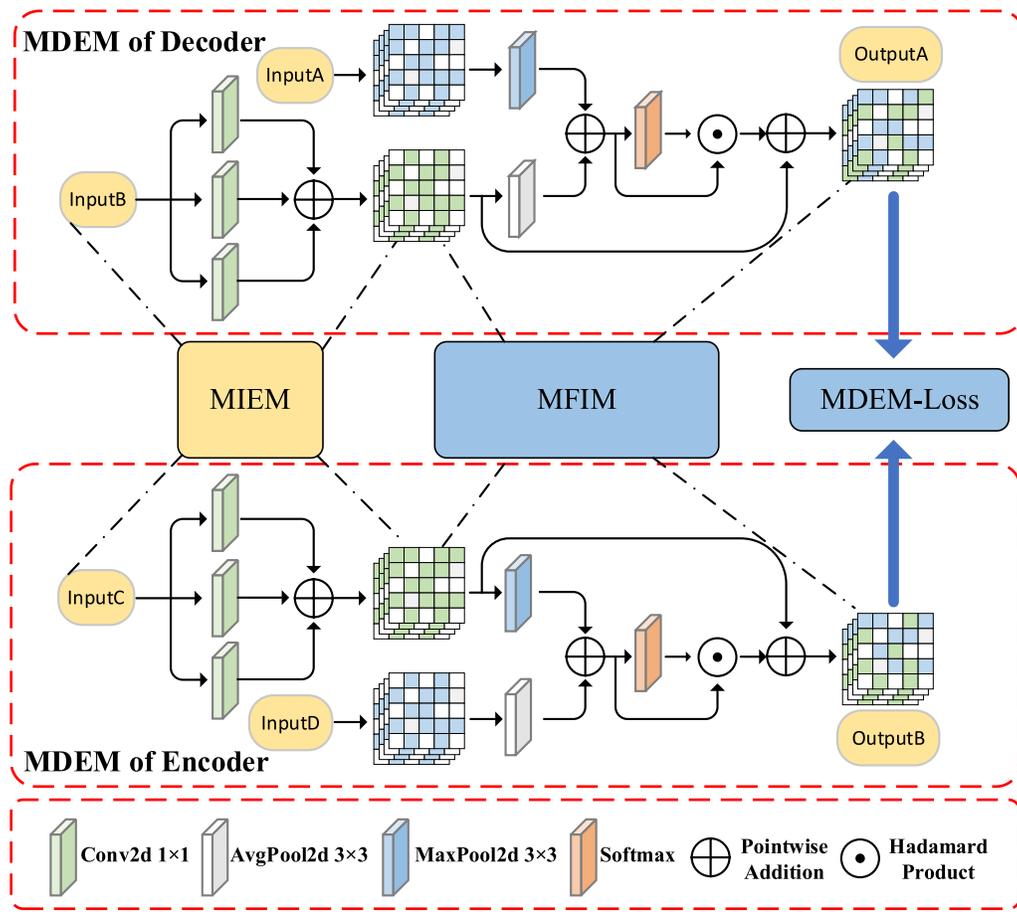


Fig. 2. Schematic of the designed MDEM. InputA represents the third-order reconstruction domain feature map, InputB represents the first-order reconstruction domain feature map, InputC represents the first-order compression domain feature map, InputD represents the third-order compression domain feature map, OutputA represents the multilevel domain feature map of the reconstructed domain, OutputB represents the multilevel domain feature map of the compressed domain, “MDEM of Encoder” represents the MDEM utilized in the main encoder part, while “MDEM of Decoder” represents the MDEM utilized in the main decoder part, and MDEM-Loss represents the loss between the multilevel domain feature map of the compressed part and the multilevel domain feature map of the reconstructed part. This loss is measured using the MSE.

feature map on the first-level domain. This kind of large-scale feature map leads to more spatial information that needs to be extracted. Therefore, the low information capacity of a single channel will lead to the loss of some spatial information. The reason for processing multichannel information by pointwise addition is that the similarity between the first-level domain feature maps has the greatest impact on the similarity between the original image and the reconstructed image. Then, when interacting the MIEM-processed first-level domain feature information with the third-level domain feature information, it is necessary to ensure that the first-level domain feature map accounts for a high proportion of the MFIM output (multilevel domain feature map). Therefore, pointwise addition is used to further enhance the spatial information on the first-level domain feature map. The core function of MFIM is to efficiently model the latent relationship between the first-level domain feature information and the third-level domain feature information that have been processed by MIEM. It realizes the effective fusion of these two types of information, so as to obtain a multilevel domain feature map with the first-level domain features as the main feature, the second-level domain features, and the third-level domain features as the supplement. First, the first-level domain feature information

and the third-level domain feature information enhanced by MIEM are extracted by AvgPooling2d and MaxPooling2d, respectively. Second, the information on different levels is fused by pointwise addition. Third, the fusion information on different levels is multiplied by the attention coefficient obtained by Softmax, so as to increase the discrimination between different features. Fourthly, in order to ensure the proportion of first-level domain features in multilevel domain features, this article introduces a branch of MIEM-enhanced first-level domain feature information at the end of MFIM. Finally, MDEM calculates the loss between the multilevel domain feature map of the main encoder and the multilevel domain feature map of the main decoder by using MSE, i.e., MDEM-Loss ($Loss_{\delta}$). In addition, in order to control the influence of MDEM-Loss on the final network performance, a weight coefficient (similarity weight ψ) is applied to $Loss_{\delta}$ which will be introduced into $Loss_{Total}$. This is explained in more detail in the Rate-Distortion Optimization section.

C. Global Information-Enhanced Attention Module

Remote sensing images are rich in global information and channel information. However, the existing remote sensing image methods either design a separate channel attention to

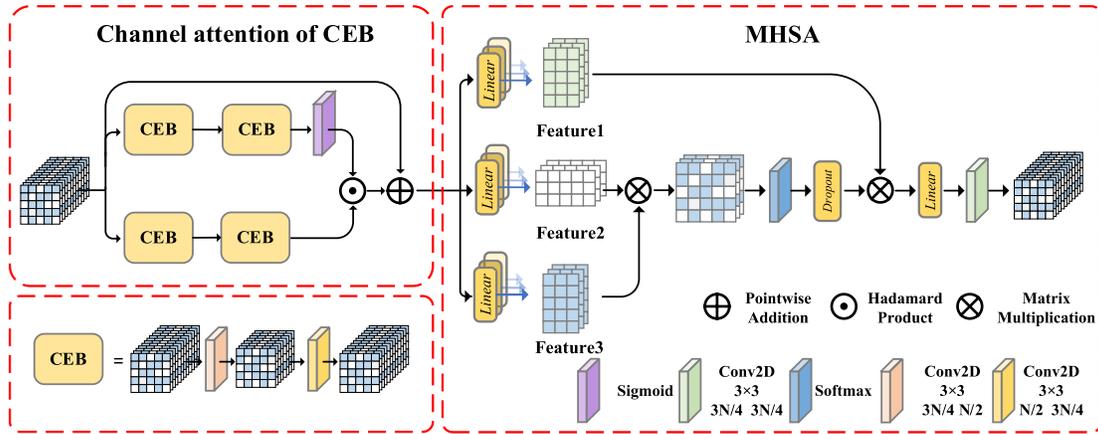


Fig. 3. Schematic of GIE-AM. The left half is the channel attention of CEB, and the right half is MHSA. Conv2D, 3×3 , $3N/4$, $N/2$ represent the parameter settings for convolution, respectively. 3×3 represents the size of the convolution kernel, $3N/4$ represents the number of input channels, $N/2$ represents the number of output channels, and the parameter settings of other convolutions are the same as above.

enhance the channel information or design the corresponding global attention module to enhance the global visual features, but they do not fully utilize the latent relationship between the channel information and the global context information. In this article, GIE-AM is designed to efficiently enhance and fuse channel information and global context information, which core components include channel attention of CEB and MHSA. Remote sensing images contain a large amount of channel information. Therefore, in neural networks, remote sensing image data are usually mapped from low-dimensional to high-dimensional to explore the latent relationship between channels. However, the number of bands in remote sensing images themselves is not as diverse as in hyperspectral images. If it is mapped to too high dimensions, the channel information feature will be attenuated and faded. This leads to the contradiction that remote sensing image data needs high-dimensional mapping, but cannot be mapped in high dimensions. Therefore, CEB is designed to first reduce the feature dimension and then restore the feature dimension. This makes the channel features more compact. The channel attention of CEB is

$$\text{Output} = \text{Input} + W_{\text{CEB}}(W_{\text{CEB}}(\text{Input})) \odot W_{\text{Sigmoid}}(W_{\text{CEB}}(W_{\text{CEB}}(\text{Input}))). \quad (6)$$

Here, Input represents the input data, W_{CEB} represents the weight parameter of CEB, W_{Sigmoid} represents the parameter of the Sigmoid layer, and \odot represents the Hadamard product.

Another core component is MHSA. ViT has excellent long-distance contextual information capture capability, which enables the ViT-based model to exhibit excellent performance across all levels of tasks. This powerful contextual information capture capability is mainly derived from its core module, i.e., self-attention. On the basis of self-attention, MHSA achieves faster training speed and information fusion in different subspaces by introducing the multihead mechanism. The structure diagram of MHSA is shown in Fig. 3. Its core is feature₁, feature₂, and feature₃, which map remote sensing image data into different projection spaces. Then, tensor multiplication is used to multiply the data on different projection spaces to enhance the global visual features. The process of the

MHSA is

$$\begin{aligned} \text{Output}_{\text{MHSA}} &= W_{\text{Conv}3 \times 3} * (W_{\text{Linear}}(\text{feature}_1 \otimes (W_{\text{Dropout}}(W_{\text{Softmax}}(\text{feature}_3 \\ &\quad \otimes \text{feature}_2))))). \quad (7) \end{aligned}$$

Here, W_{Linear} represents the weight parameter of the linear layer, W_{Dropout} represents the weight parameter of the dropout layer, W_{Softmax} represents the weight parameter of the Softmax layer, and \otimes represents matrix multiplication.

The process of GIE-AM feature extraction is shown in Algorithm 1.

It is worth mentioning that GIE-AM does not simply extract the rich channel features and global visual features in remote sensing images but effectively integrates the two. First, the channel attention of CEB combined with the attention mechanism was utilized to upgrade and reduce the dimensionality of the channel features multiple times, thus compressing the channel features and obtaining sufficient channel capacity. Second, MHSA maps high-quality features into different subspaces through the mechanism of multiple heads to further explore the global context relationship. Through the self-attention mechanism, the long-distance information and low-frequency features are effectively strengthened. Finally, through the combination of channel attention of CEB and MHSA, the feature map of remote sensing image with efficient fusion of channel features and global visual features is obtained.

D. BaselineA

At present, the remote sensing image compression method based on deep learning mainly has the following problems: 1) in the process of compression and reconstruction at different levels, the number of channels is not allocated, resulting in the inability to fully utilize the neural network and 2) the size of some convolutional kernels is unreasonable, which leads to the slow training of the network under the same sensory field. In order to solve the above problems, BaselineA was designed to achieve the same receptive field and better rate-distortion

Algorithm 1 The Feature Extraction Process of Remote Sensing Images by GIE-AM**Input:** Remote sensing data $X \in \mathbb{R}^{b \times c \times h \times w}$ 1: for $i=1$ to T do2: Perform *CEB* twice, the result denoted as $X_1 \in \mathbb{R}^{b \times c \times h \times w}$.3: Perform *CEB* twice and *Sigmoid* once, the result denoted as $X_2 \in \mathbb{R}^{b \times c \times h \times w}$.4: Perform *Hadamard Product* of $X_1 \in \mathbb{R}^{b \times c \times h \times w}$ and $X_2 \in \mathbb{R}^{b \times c \times h \times w}$, the result denoted as $X_3 \in \mathbb{R}^{b \times c \times h \times w}$.5: Perform *Pointwise Addition* of $X_3 \in \mathbb{R}^{b \times c \times h \times w}$ and $X \in \mathbb{R}^{b \times c \times h \times w}$, the result denoted as $X_{CEB} \in \mathbb{R}^{b \times c \times h \times w}$.6: Perform *Flatten*, *Reshape* and *Linear*, the result denoted as $attn \in \mathbb{R}^{b \times n \times 3c}$.7: Perform *Split* and *Reshape*, the result denoted as $feature_3 \in \mathbb{R}^{b \times n \times c}$, $feature_2 \in \mathbb{R}^{b \times n \times c}$ and $feature_1 \in \mathbb{R}^{b \times n \times c}$.8: Perform *Reshape* and *Transpose*, the result denoted as $feature_3 \in \mathbb{R}^{b \times head \times n \times head}$, $feature_2 \in \mathbb{R}^{b \times head \times head \times n}$ and $feature_1 \in \mathbb{R}^{b \times head \times n \times head}$.9: Perform *Matrix multiplication* of $feature_3 \in \mathbb{R}^{b \times head \times n \times head}$ and $feature_2 \in \mathbb{R}^{b \times head \times head \times n}$, and then perform *Softmax*, *Dropout*, the result denoted as $attn_1 \in \mathbb{R}^{b \times head \times n \times n}$.10: Perform *Matrix multiplication* of $attn_1 \in \mathbb{R}^{b \times head \times n \times n}$ and $feature_1 \in \mathbb{R}^{b \times head \times n \times head}$, the result denoted as $attn_2 \in \mathbb{R}^{b \times head \times n \times head}$.11: Perform *Transpose*, *Flatten*, *Linear*, the result denoted as $attn_3 \in \mathbb{R}^{b \times n \times c}$.12: Perform *Transpose*, *Reshape*, the result denoted as $attn_4 \in \mathbb{R}^{b \times c \times h \times w}$.13: Perform *Conv*_{3×3}, the result denoted as $attn_{GIE-AM} \in \mathbb{R}^{b \times c \times h \times w}$.

end for

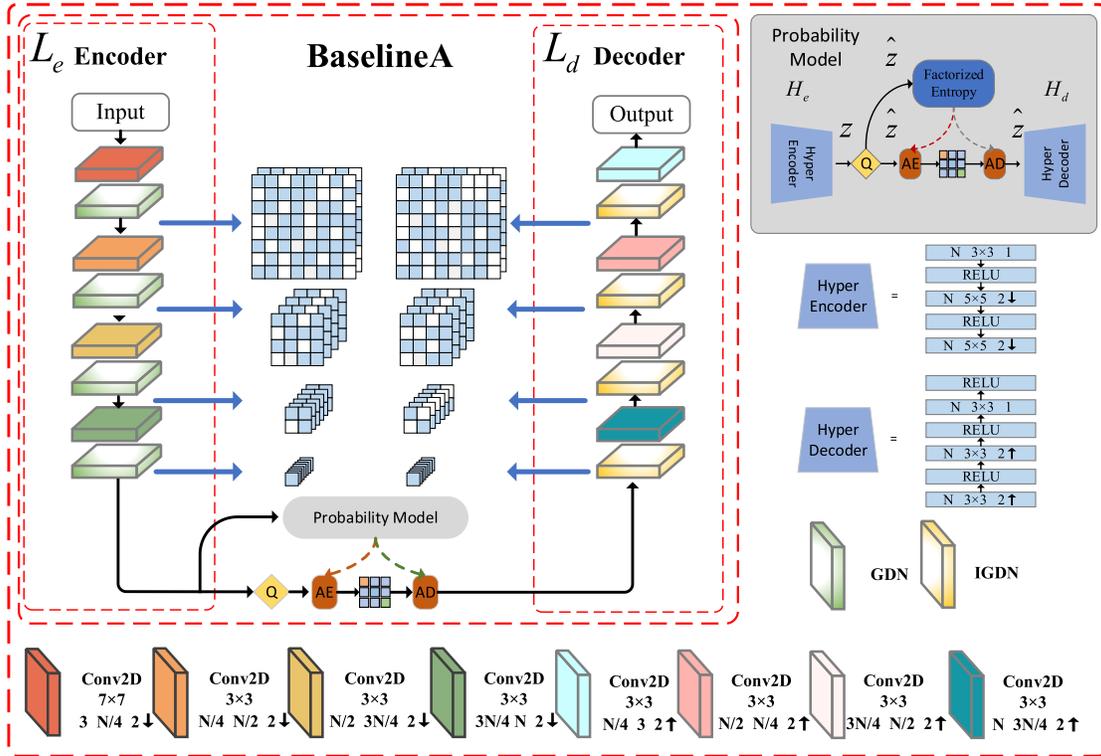
Output: Feature map $attn_{GIE-AM} \in \mathbb{R}^{b \times c \times h \times w}$ after feature extraction.

Fig. 4. Schematic of BaselineA.

performance with fewer parameters. BaselineA is shown in Fig. 4.

In order to make the reconstructed image smoother, BaselineA uses the common VAE as the basic framework. For problem 1), the number of channels in most of the convolutions in the common remote sensing image compression framework is set to the same number. However, as downsampling progresses, the size of the feature map continues to decrease, which means that there is less spatial information

available for learning. In this case, the network should pay more attention to the channel information of the image. Therefore, in this article, the number of downsampled channels in the main encoder is reassigned, that is, as downsampling proceeds, the number of channels is gradually increased, so as to focus more on spatial information extraction when the feature map space size is large and more on channel information extraction when the feature map space size is small. The main decoder goes through a similar process. For problem 2), the

main frame abandons the commonly used convolutional kernel size 5×5 and chooses a smaller convolutional kernel 3×3 to reduce the number of parameters. In addition, in order to compensate for the loss of receptive field caused by the small convolutional kernel, the convolution with size 7×7 is selected for compression during the first downsampling of the main encoder. The calculation process of the receptive field is

$$F(i) = (F(i+1) - 1) \times S + K. \quad (8)$$

Here, $F(i)$ represents the size of the receptive field in layer i , S represents the step size, and K represents the convolutional kernel size.

The process of calculating the parameter quantity of the convolution kernel is

$$P = h \times w \times c \times c. \quad (9)$$

Here, P represents the number of parameters, h represents the width of the convolutional kernel, w represents the height of the convolutional kernel, and c represents the number of channels.

These two formulas can be used to calculate the total number of parameters and the receptive field size consumed by convolution in the main encoder and main decoder. By calculation, the receptive field size of the main encoder is 61 and the number of parameters is $112.5C^2$ for the common VAE framework (convolutional kernel size 5×5 and channel set to $3C/4$). However, BaselineA (the main encoder) has the receptive field size of 63 and the number of parameters of $36.25C^2$. In comparison, it is found that the number of parameters of BaselineA is only one-third of that of common VAE frameworks under the premise of having almost the same receptive field. This undoubtedly has a huge parametric advantage.

E. Rate–Distortion Optimization

The goal of the compression frame is to achieve a balance between compression and distortion. Therefore, a common rate–distortion optimization strategy is introduced into the compression framework to guide the efficient training of the model, which can be expressed as

$$\arg \min \text{Loss}_{\text{Total}} = R + \lambda D. \quad (10)$$

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

$$R = R_{\hat{y}} + R_{\hat{z}}. \quad (11)$$

Here, the bit rate consists of the latent representation information \hat{y} together with the side information \hat{z}

$$R_{\hat{y}} = - \sum_i \log_2(p_{\hat{y}}(\hat{y}_i)) \quad (12)$$

$$R_{\hat{z}} = - \sum_i \log_2(p_{\hat{z}}(\hat{z}_i)). \quad (13)$$

Here, $p_{\hat{y}}$ is an entropy model that can be learned, and $p_{\hat{z}}$ represents a hyper encoder.

In order to further improve the quality of image compression, a novel rate–distortion optimization strategy is proposed, i.e., (14). MDEM-Loss (Loss_{δ}) is introduced in $\text{Loss}_{\text{Total}}$ to improve the multilevel domain similarity of the front, which is essentially a loss between the multilevel domain feature map of the main encoder and the multilevel domain feature map of the main decoder. This new $\text{Loss}_{\text{Total}}$ increases the similarity between the highest-level domains, thus efficiently optimizing the training of the network. This rate–distortion optimization strategy can be expressed as

$$\arg \min \text{Loss}_{\text{Total}} = R + \lambda(D + \psi \text{Loss}_{\delta}). \quad (14)$$

Here, ψ represents the weight coefficient of multilevel domain loss, and Loss_{δ} represents the MDEM-Loss.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A lot of experiments have been carried out on remote sensing image datasets such as San Francisco [46], NWPU-RESISC45 [47], and UC-Merced [48]. The selected datasets contain a wealth of ground object information, which can effectively evaluate the performance of the proposed MDSNet method. In this section, the proposed MDSNet method is compared with some excellent compression methods, including traditional codecs and deep learning-based compression models. Traditional image compression methods include JPEG2000 [12], BPG [14], and WebP [15], and compression models based on deep learning include Minnen et al. [28], Minnen et al. (mean) [28], Ballé et al. (hyperprior) [29], Ballé et al. (factorized-relu) [29], and Tong2023 [49]. Experimental results show that the proposed MDSNet method has the best performance in both PSNR and MS-SSIM evaluation indicators. In addition, the reconstructed images of different methods are evaluated by classification task, which further verifies the superiority of the MDSNet method.

A. Introduction to Remote Sensing Image Dataset

1) *San Francisco Dataset*: San Francisco is a dataset of remotely sensed images from [46]. San Francisco is a remote sensing image with resolution 17408×17408 . It contains information on various categories of features, such as buildings, coasts, roads, ports, and lakes. In this article, San Francisco was cropped at a resolution of 256×256 pixels, and 3000 valid images were selected from the cropped images to form a dataset. This dataset is divided into a training set, a validation set, and a test set at a ratio of 8:1:1. Fig. 5 shows a partial sample of this dataset.

2) *NWPU-RESISC45 Dataset*: 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 ports. The 140 images in each category were selected to form a dataset

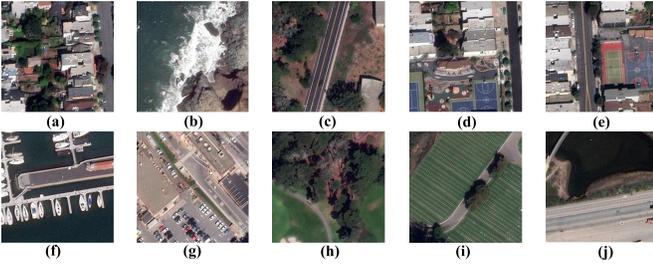


Fig. 5. Some images from the San Francisco dataset. (a) Building. (b) Coastline. (c) Highway. (d) Basketball court. (e) Tennis court. (f) Port. (g) Car park. (h) Forest. (i) Farmland. (j) Lake.

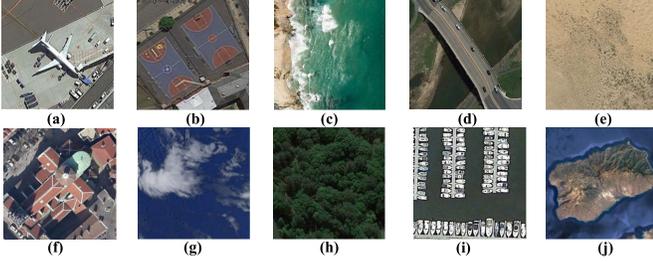


Fig. 6. Some images from the NWPU-RESISC45 dataset. (a) Airports. (b) Basketball court. (c) Sand beach. (d) Bridge. (e) Desert. (f) Church. (g) Clouds. (h) Forests. (i) Ports. (j) Island.

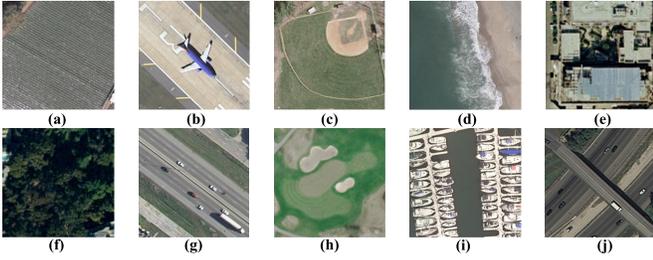


Fig. 7. Some images from the UC-Merced dataset. (a) Farmland. (b) Plane. (c) Baseball stadium. (d) Sand beach. (e) Building. (f) Forest. (g) Highway. (h) Golf course. (i) Port. (j) Overpass.

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 shows a partial sample of this dataset.

3) *UC-Merced Dataset*: 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. These images include farmland, airports, forests, ports, and islands. The UC-Merced dataset is divided into a training set, a validation set, and a test set at a ratio of 8:1:1. Fig. 7 shows a partial sample of this dataset.

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) *PSNR*: 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 represented as

$$\text{PSNR}(X, \hat{X}) = \frac{1}{C} \sum_{i=1}^C 10 \log_{10} \left(\frac{\max^2(X^i)}{\text{MSE}_i} \right). \quad (15)$$

Here, $\text{MSE}(X, \hat{X}) = (1/H \times W \times C) \|X - \hat{X}\|_F^2$, $\max^2(X^{(i)})$ represent the square of the largest pixel in the i th band, and C represents the number of bands.

2) *MS-SSIM*: 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 MS-SSIM can be expressed as

$$D_{\text{MS-SSIM}} = 1 - \prod_{m=1}^M \left(\frac{2\mu_X \mu_{\hat{X}} + N_1}{\mu_X^2 + \mu_{\hat{X}}^2 + N_1} \right)^{\alpha_m} \left(\frac{2\sigma_{X\hat{X}} + N_2}{\sigma_X^2 + \sigma_{\hat{X}}^2 + N_2} \right)^{\zeta_m}. \quad (16)$$

Here, M represents the different resolutions, μ_X and $\mu_{\hat{X}}$ represent the mean of the original image and the reconstructed image, σ_X and $\sigma_{\hat{X}}$ represent the standard deviation between the original image and the reconstructed image, $\sigma_{X\hat{X}}$ represent the covariance between the original image and the reconstructed image, α_m and ζ_m represent the relative importance between the two terms, and N_1 and N_2 are the 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 can be expressed as

$$\text{MS-SSIM} = -10 \log_{10}(1 - D_{\text{MS-SSIM}}). \quad (17)$$

3) *Classification Indicators of Remote Sensing Scenes*: In this article, two widely used remote sensing scene classification evaluation indicators are adopted 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 MDSNet method 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 and the main decoder, and the other is the auxiliary optimizer between the hyper encoder and the hyper decoder. For the main optimizer, the initial learning rate is set at 10^{-4} , and the optimal model of MDSNet 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.

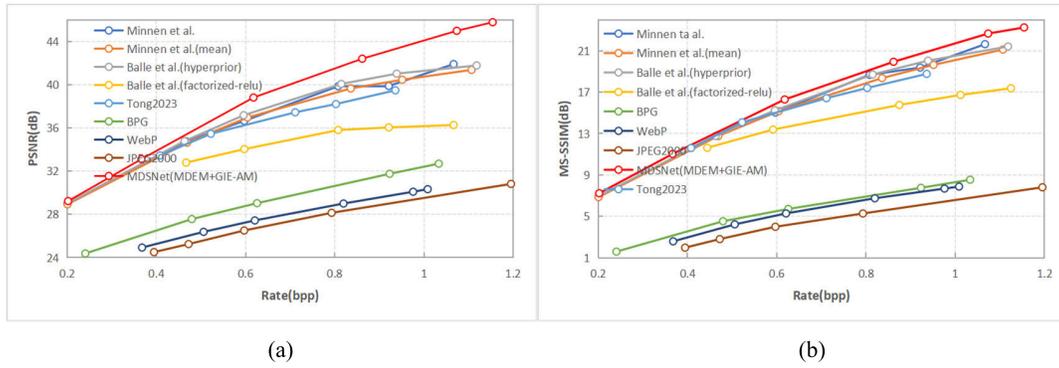


Fig. 8. Rate–distortion curves on San Francisco. (a) PSNR. (b) MS-SSIM.

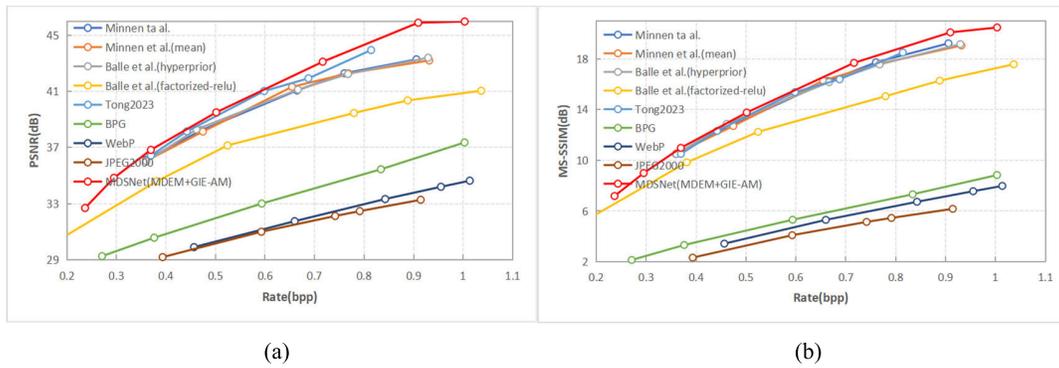


Fig. 9. Rate–distortion curves on NWPU-RESISC45. (a) PSNR. (b) MS-SSIM.

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@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 weight coefficient ψ for multilevel domain loss in MDEM-Loss is set to 0.1. In GIE-AM, the number of heads in the MHSA is set to 4. In the overall frame, the number of channels N is set to 256. In the classification of remote sensing scenes, the benchmark model used for testing was the efficient multiscale transformer and cross-level attention learning (EMTCAL) [50]. 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 compression methods were evaluated for rate–distortion performance using 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 datasets

San Francisco, NWPU-RESISC45, and UC-Merced, respectively. It can be seen that the rate–distortion performance of the method based on deep learning is better than that of the traditional image compression method. For traditional codec-based image compression methods, it can be found that the rate–distortion performance of BPG is better than that of WebP and JPEG2000 in most cases. The main reason is that BPG can encode different color channels separately through multichannel encoding techniques. This enhances the precise control of a wide range of detailed features, helping to reconstruct high-fidelity images. For the image compression method based on deep learning, the rate–distortion performance obtained by Ballé et al. (factorized-relu) [29] is relatively poor. The main reason is that it only uses a simple convolutional layer, which has a weak ability to extract features. It can stack the number of convolutional layers to improve the feature extraction ability to a certain extent, but this will greatly increase the number of parameters and inference time. Tong2023 achieved the highest PSNR and MS-SSIM rate–distortion performance among all methods except the MDSNet method on the datasets NWPU-RESISC45 and UC-Merced. The main reason is the use of an excellent attention mechanism and a more reasonable residual convolution module. However, Tong2023 achieves poor PSNR and MS-SSIM rate–distortion performance on the San Francisco dataset, which proves that the compression model is not robust. Minnen et al. [28], Minnen et al. (mean) [28], and Ballé et al. (hyperprior) [29], three deep learning-based image compression models, have achieved average rate–distortion

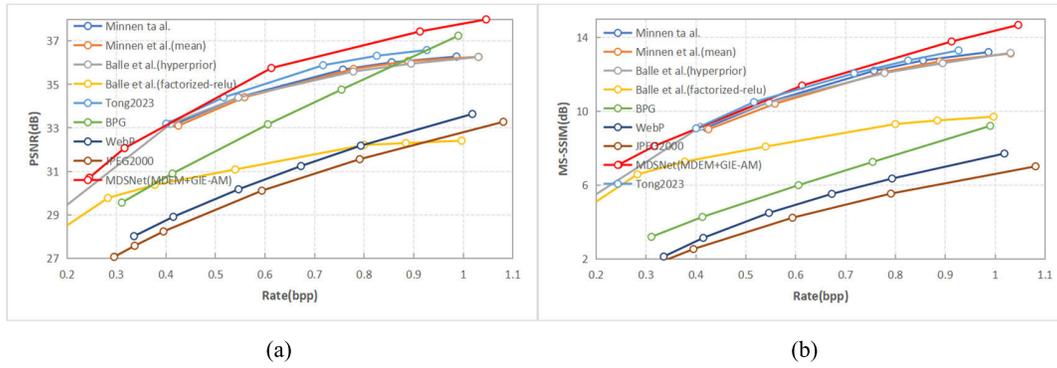


Fig. 10. Rate–distortion curves on UC-Merced. (a) PSNR. (b) MS-SSIM.

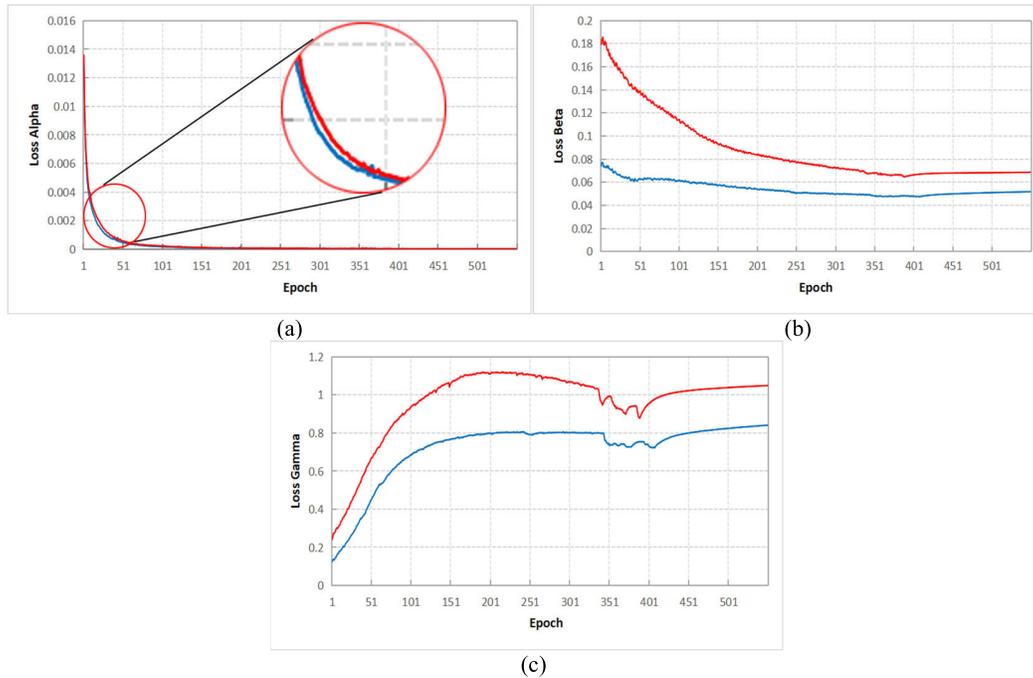


Fig. 11. Difference loss on the dataset San Francisco: (a) first-level difference loss, (b) second-level difference loss, and (c) third-level difference loss.

performance. The main reason is that although they have good convolution modules, they lack a strong attention mechanism and are more efficient $\text{Loss}_{\text{Total}}$. However, the MDSNet method proposed in this article achieves the highest PSNR and MS-SSIM rate–distortion performance on three datasets at the same time, and this performance advantage will be more prominent at high bits per pixel. This superior rate–distortion performance not only strongly proves the robustness of the MDSNet method but also proves the effectiveness of the MDSNet method, as well as MDEM, GIE-AM, and BaselineA.

E. Analysis and Verification of MDEM

In this section, some experiments are carried out to verify the effectiveness of the proposed MDEM. The baseline network is BaselineA. The red curve is the BaselineA without MDEM and is denoted as the original difference loss. The blue curve is BaselineA with MDEM added, which is denoted as the optimization difference loss. Figs. 11 and 12 show the difference loss curves on San Francisco and NWPU-RESISC45, respectively. In addition, parts of Figs. 11(a) and 12(a) are

enlarged to show the relationship between the curves more clearly. MDEM is designed to enhance the similarity between the highest-level domains by enhancing the multilevel domain similarity of the front. Through comparison, it is found that on the San Francisco dataset, whether it is the first-level difference loss, the second-level difference loss, or the third-level difference loss, their losses are significantly reduced after the introduction of MDEM in the network. This means a significant reduction in the difference between the multilevel image features in the compression domain and the multilevel image features in the reconstruction domain. In other words, the similarity between the compressed domain features and the reconstructed domain features is greatly enhanced. The difference loss experiments on NWPU-RESISC45 also confirm the effectiveness of the MDEM and also illustrate the strong robustness of MDEM.

F. Visualization Comparison of Reconstructed Images

In order to further verify the effectiveness of the proposed MDSNet, the reconstructed images obtained by different

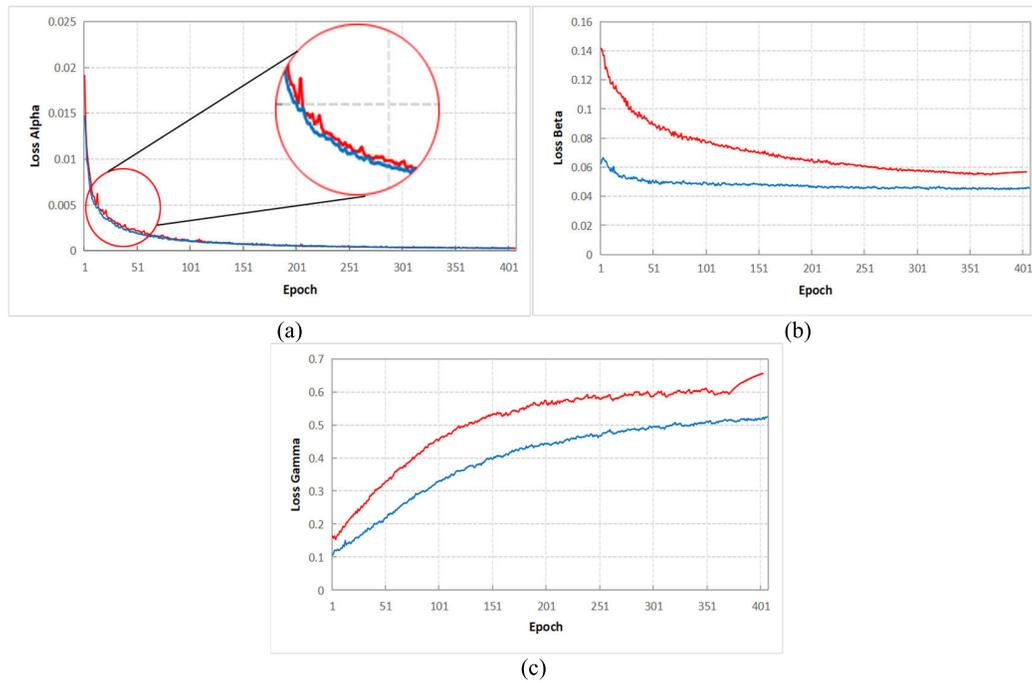


Fig. 12. Difference loss on the dataset NWPU-RESISC45: (a) first-level difference loss, (b) second-level difference loss, and (c) third-level difference loss.

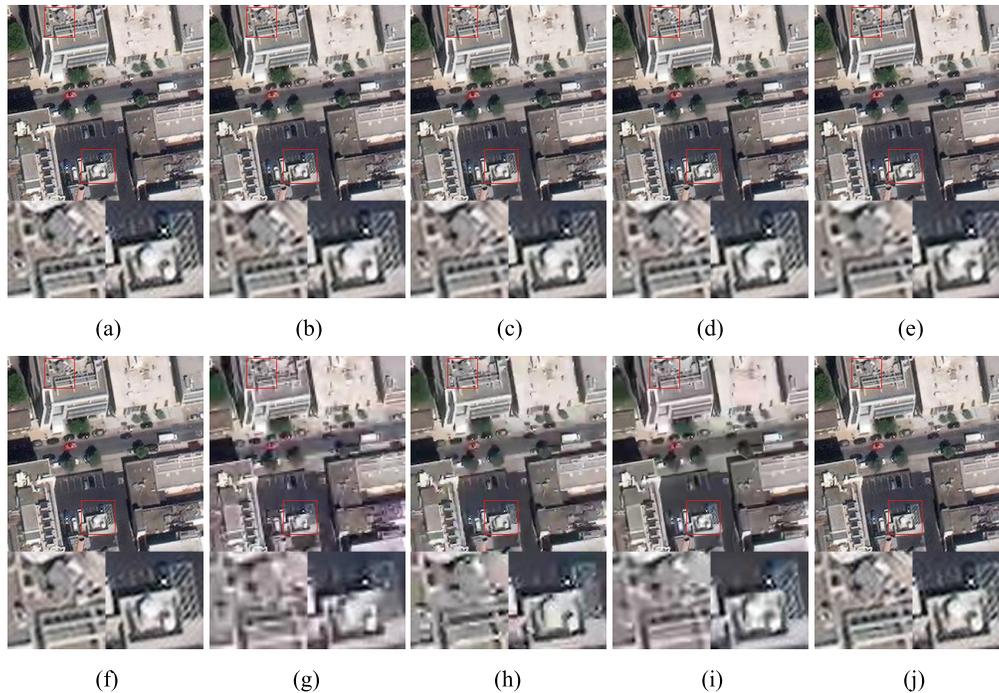


Fig. 13. Visual comparison of reconstructed images obtained by different methods on the San Francisco dataset. (a) Original. (b) Minnen et al. (bpp: 0.297; PSNR: 31.12; MS-SSIM: 9.48) [28]. (c) Minnen et al. (mean) (bpp: 0.298; PSNR: 30.93; MS-SSIM: 9.44) [28]. (d) Ballé et al. (hyperprior) (bpp: 0.294; PSNR: 30.93; MS-SSIM: 9.52) [29]. (e) Ballé et al. (factorized-relu) (bpp: 0.295; PSNR: 29.41; MS-SSIM: 8.86) [29]. (f) Tong2023 (bpp: 0.303; PSNR: 30.92; MS-SSIM: 9.40). (g) JPEG2000 (bpp: 0.309; PSNR: 22.73; MS-SSIM: 2.46). (h) Webp (bpp: 0.413; PSNR: 24.48; MS-SSIM: 4.39). (i) BPG (bpp: 0.307; PSNR: 24.65; MS-SSIM: 3.97). (j) MDSNet (bpp: 0.301; PSNR: 31.36; MS-SSIM: 9.86.)

compression methods are compared. Figs. 13 and 14 show the comparison results of different methods on the San Francisco dataset and the NWPU-RESISC45 dataset, respectively. Fig. 13 shows the visual experimental results with a bit rate of 0.3 bpp, and Fig. 14 shows the visual experimental results with a bit rate of about 0.25 bpp. Fig. 13 includes the original image and nine reconstructed images obtained by different

methods. Fig. 13 shows a diagram of a city building. The roof area in the upper left corner and the square building area in the middle area were enlarged. For the traditional image compression method, the rate-distortion performance of BPG is significantly higher than that of JPEG2000 and WebP. In the image reconstructed by BPG, the edge texture of the object in the roof area in the upper left corner is

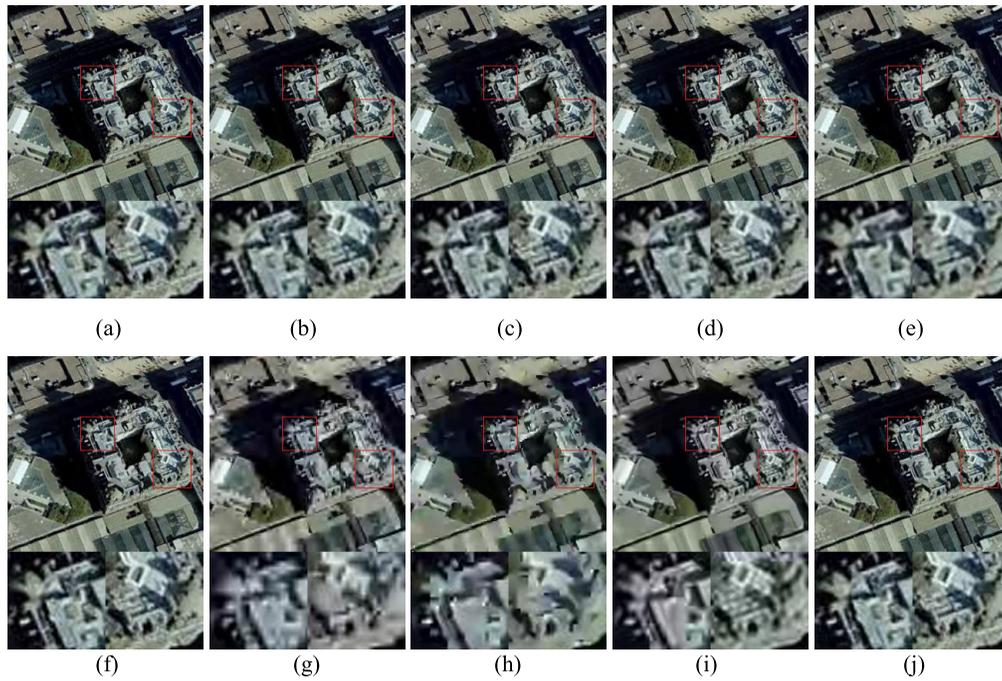


Fig. 14. Visual comparison of reconstructed images obtained by different methods on the NWPU-RESISC45 dataset. (a) Original. (b) Minnen et al. (bpp: 0.260; PSNR: 33.26; MS-SSIM: 7.73) [28]. (c) Minnen et al. (mean) (bpp: 0.257; PSNR: 33.20; MS-SSIM: 7.67) [28]. (d) Ballé et al. (hyperprior) (bpp: 0.263; PSNR: 33.30; MS-SSIM: 7.64) [29]. (e) Ballé et al. (factorized-relu) (bpp: 0.261; PSNR: 31.87; MS-SSIM: 7.49) [29]. (f) Tong2023 (bpp: 0.263; PSNR: 33.57; MS-SSIM: 7.76). (g) JPEG2000 (bpp: 0.257; PSNR: 22.40; MS-SSIM: 1.92). (h) Webp (bpp: 0.249; PSNR: 22.69; MS-SSIM: 2.13). (i) BPG (bpp: 0.259; PSNR: 24.26; MS-SSIM: 3.72). (j) MDSNet (bpp: 0.262; PSNR: 33.92; MS-SSIM: 8.02).

clearer, and the square building area in the middle area retains more stripe information. However, the JPEG2000 and WebP reconstruction areas have lost some of the details and are blurred. The main reason is that BPG uses multichannel encoding technology, which improves the ability to reconstruct detailed features. For the image compression method based on deep learning, the reconstructed images show a relatively good visualization, but there is still a certain gap with the proposed MDSNet method. In Fig. 13, there are some artifacts and noise in the local area of the reconstructed image by Minnen et al. (mean) [28], Ballé et al. [29], Ballé et al. (factorized-relu) [29], and Tong2023. In the reconstructed images of Minnen et al. (mean) [28], Ballé et al., Ballé et al. (factorized-relu) [29], and Tong2023, the discrimination between pixels at the edge of the image is too low, and the transition between pixels is too coarse. This leads to color flattening and distortion in the reconstructed image. Finally, compared with Minnen et al. [28], which has a relatively good visual effect in the comparison method, the reconstructed image quality of MDSNet has more detailed features. However, Minnen et al.'s [28] reconstruction image is too smooth between pixels, which leads to the loss of some detailed information. In Fig. 14, MDSNet also achieves the best visualization result. From the perspective of visualization, the above experiments further illustrate the effectiveness of the proposed MDSNet.

G. Ablation Experiments

In this article, some ablation experiments are performed to verify the effectiveness of each component. Figs. 15 and 16 show the results of ablation experiments on the San Francisco

dataset and the UC-Merced dataset, respectively. In these figures, 1) MDSNet represents BaselineA, which is the baseline network of this ablation experiment; 2) MDSNet (MDEM) represents the integration of MDEM based on BaselineA; 3) MDSNet (GIE-AM) represents the integration of GIE-AM based on BaselineA; and 4) MDSNet (MDEM + GIE-AM) stands for MDEM and GIE-AM based on BaselineA. As can be seen from Figs. 15 and 16, the rate–distortion performance of BaselineA is the lowest in most cases. MDSNet (MDEM) outperforms BaselineA at the same bit rate. This proves the effectiveness of the method to improve the quality of the reconstructed image by enhancing multilevel domain similarity of the front. MDSNet (GIE-AM) also outperforms BaselineA at the same bit rate. This indicates that the channel information and global visual features in remote sensing images are of great significance for the high-quality reconstruction of images. The rate–distortion performance of MDSNet (MDEM + GIE-AM) is optimal at the same bit rate, and this performance advantage will be more prominent as the bit rate increases. This phenomenon shows that BaselineA, MDEM, and GIE-AM achieve efficient fusion and can retain sufficient high-frequency and low-frequency information to support high-quality image reconstruction.

H. Discussion

In this section, the effectiveness of the MDSNet is further verified from a new perspective by using the reconstructed image for remote sensing scene image classification. The dataset selected is NWPU-RESISC45. The image compression methods used for comparison include Minnen et al. [28], Minnen et al. (mean) [28], Ballé et al. (hyperprior) [29],

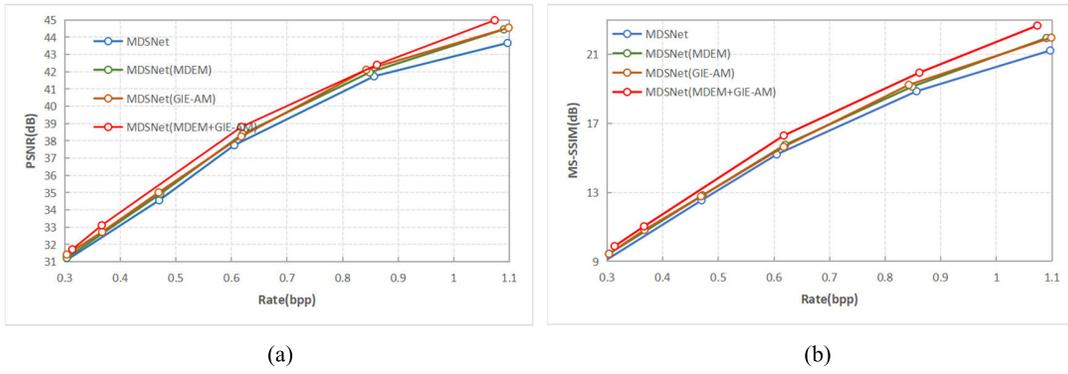


Fig. 15. Ablation results of different methods on the San Francisco dataset. (a) PSNR. (b) MS-SSIM.

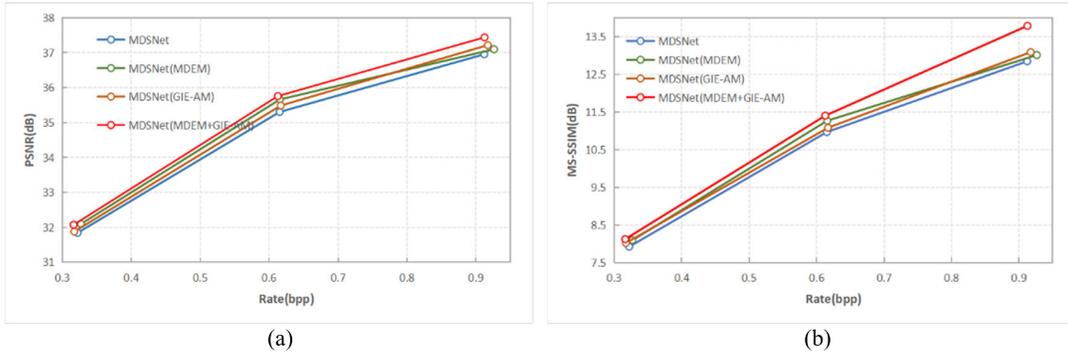


Fig. 16. Ablation results of different methods on the UC-Merced dataset. (a) PSNR. (b) MS-SSIM.

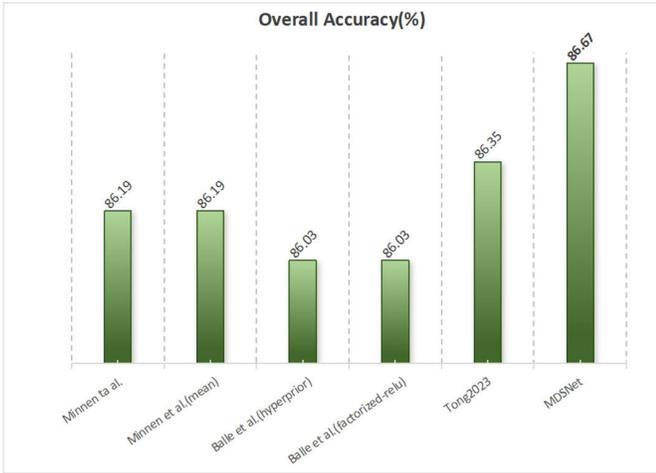


Fig. 17. OA of the reconstructed images obtained by different compression methods for remote sensing scene image classification.

Ballé et al. (factorized-relu) [29], and Tong2023. The benchmark model for remote sensing scene classification is EMTCAL [50], and the training–testing ratio is 10%–90%. The reconstructed images are not used for the training of remote sensing scene classification, but only participate in the testing of classification performance. In order to ensure the fairness of the experiment, the reconstructed images of different compression methods were obtained at a bit rate of 0.81 bpp. Fig. 17 shows the OA obtained by different reconstructed images for remote sensing scene image classification. In terms of OA, the proposed MDSNet obtains the highest

OA value and achieves the best classification performance, which is higher 0.55% than that of Minnen et al. [28], higher 0.55% than that of Minnen et al. (mean) [28], higher 0.73% than that of Ballé et al. (hyperprior) [29], higher 0.73% than that of Ballé et al. (factorized-relu) [29], and higher than 0.37% than that of Tong2023. Fig. 18 demonstrates the CMs of reconstructed images obtained by different compression methods when they are used for remote sensing scene classification. In Fig. 18, the classification effect of Lake, stadium, and river in MDSNet’s CM is better than those of other comparison methods. This is mainly due to the fact that there are a large number of global visual features in these types of scenes, and the GIE-AM in MDSNet just effectively enhances the channel features and global features. For remote sensing scene image classification, the protection of complex ensemble structure features and global visual features is the key factor to improve the classification performance. In the proposed MDSNet, the similarity of the overall geometric structure between the original image and the reconstructed image is enhanced by multilevel domain similarity through MDEM. Then, the channel features and global visual features are efficiently enhanced by GIE-AM. Finally, high-quality discriminant features are obtained. This is the reason why the proposed MDSNet network reconstruction image achieves the best performance in remote sensing scene classification.

I. Complexity Analysis

In order to compare the computational complexity and time consumption of different compression methods, some

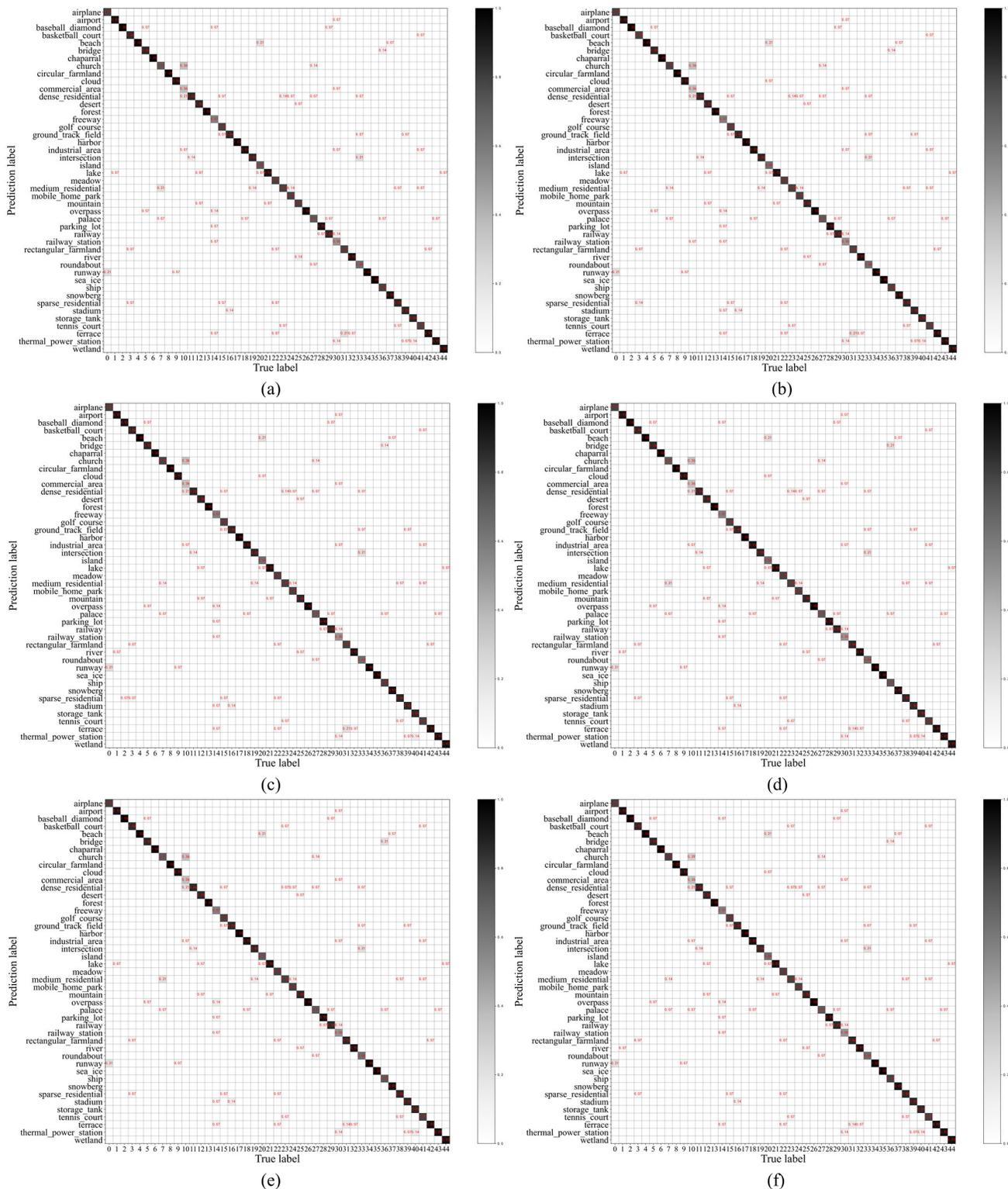


Fig. 18. CM of the reconstructed images by different compression methods. (a) Minnen et al. [28]. (b) Minnen et al. (mean) [28]. (c) Ballé et al. (hyperprior) [29]. (d) Ballé et al. (factorized-relu) [29]. (e) Tong2023. (f) MDSNet.

experiments are conducted. All the compression methods are tested on the same device and in the same environment. The evaluation indicators include the amount of parameters, floating point operations (FLOPs), GPU memory, compression time, and reconstruction time. Table I shows the computational complexity of these comparative methods

and the proposed MDSNet method when the input image size is $3 \times 256 \times 256$. By comparison, it is found that the number of parameters of MDSNet is second only to Ballé et al. (factorized-relu) [29]. It is worth mentioning that although the number of parameters of Ballé et al. (factorized-relu) [29] is less than that of the proposed MDSNet,

TABLE I
COMPLEXITY COMPARISONS OF DIFFERENT COMPRESSION METHODS

	Minnen et al.	Minnen et al.(mean)	Balle et al. (hyperprior)	Balle et al.(factorized-relu)	Tong2023	MDSNet
Parameter	12.05M	11.04M	9.91M	5.56M	27.55M	8.64M
FLOPs	27.04G	26.78G	26.49G	25.95G	67.21G	13.90G
GPU Memory	3.0GB	2.8GB	2.8GB	1.6GB	4.2GB	3.1GB
Compression time	0.5931s	0.0712s	0.0731s	0.0321s	0.7014s	0.1059s
Reconstruction time	1.0781s	0.0745s	0.0775s	0.0390s	1.3214s	0.0980s

the PSNR and MS-SSIM rate–distortion performance of Ballé et al. (factorized-relu) [29] is much lower than that of our method at the same bit rate. By comparing FLOPs, it can be found that MDSNet achieves the minimal FLOPs, which is only 51.41%, 51.90%, 52.47%, 53.56%, and 20.68% of the FLOPs of Minnen et al. [28], Minnen et al. (mean) [28], Ballé et al. (hyperprior) [29], Ballé et al. (factorized-relu) [29], and Tong2023, respectively. This fully demonstrates the superiority of the proposed MDSNet. For GPU Memory, MDSNet is in the middle of all methods. The reason is that some networks in MDEM require parallel computing, which consumes a certain amount of GPU resources. For compression time and reconstruction time, it can be found that the codec speed of MDSNet is much faster than that of Minnen et al. [28] and Tong2023 methods, and the codec time is only one-tenth of them. However, it is slightly inferior to the methods such as Minnen et al. (mean) [28], Ballé et al. (hyperprior) [29], and Ballé et al. (factorized-relu) [29], but the PSNR and MS-SSIM of MDSNet are significantly higher than these methods at the same bit rate. These experiments strongly demonstrate that MDSNet can achieve excellent rate–distortion performance with low complexity.

IV. CONCLUSION

This article proposes a novel MDSNet for the compression of remote sensing images. First, BaselineA was constructed by rationally allocating the number of channels in the network and redesigning the convolutional kernel size. It can achieve a large receptive field and good rate–distortion performance under the premise of low parameter quantity. Second, MDEM was constructed by designing MIEM for extracting multichannel information, MFIM for multilevel feature interaction, and MDEM-Loss for guiding the network to improve multilevel domain similarity. It can improve the quality of the final reconstructed image by enhancing the multilevel domain similarity of the front. Third, GIE-AM was constructed by designing a CEB to enhance channel features and an MHSA to improve the ability to capture global visual features, which could achieve efficient enhancement and fusion of channel features and global visual features. Finally, MDSNet was efficiently trained under the guidance of the proposed Loss_{Total} strategy, which can achieve the optimum rate–distortion performance compared with other comparison methods and obtains the reconstructed image with the best classification performance. Specifically, at 1.1 bpp, MDSNet achieves PSNR improvements of 7.4%, 8.8%, 7.7%, 23.9%, and 11.1% compared

to that of Minnen et al. [28], Minnen et al. (mean) [28], Ballé et al. (hyperprior) [29], Ballé et al. (factorized-relu) [29], and Tong2023, respectively. However, MDEM only fuses and improves the similarity of features on the domain at three levels. In the future, we will further explore the possibility of more level domain feature interaction and similarity improvement. 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] H. Fang, S. Guo, X. Wang, S. Liu, C. Lin, and P. Du, “Automatic urban scene-level binary change detection based on a novel sample selection approach and advanced triplet neural network,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5601518.
- [2] Y. Zhao, P. Chen, Z. Chen, Y. Bai, Z. Zhao, and X. Yang, “A triple-stream network with cross-stage feature fusion for high-resolution image change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5600417.
- [3] S. Mei, C. Song, M. Ma, and F. Xu, “Hyperspectral image classification using group-aware hierarchical transformer,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5539014.
- [4] M. Ma, S. Mei, F. Li, Y. Ge, and Q. Du, “Spectral correlation-based diverse band selection for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5508013.
- [5] 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.
- [6] 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. 1784–1797, 2022.
- [7] 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.
- [8] 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.
- [9] Z. Xing et al., “Flood vulnerability assessment of urban buildings based on integrating high-resolution remote sensing and street view images,” *Sustain. Cities Soc.*, vol. 92, May 2023, Art. no. 104467.
- [10] G.-S. Xia et al., “AID: A benchmark data set for performance evaluation of aerial scene classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.

- [11] X. Lu, B. Wang, X. Zheng, and X. Li, "Exploring models and data for remote sensing image caption generation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2183–2195, Apr. 2018.
- [12] *JPEG2000 Official Software OpenJPEG*. Accessed: 2015. [Online]. Available: <https://jpeg.org/jpeg2000/software.html>
- [13] 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.
- [14] F. Bellard. *BPG Image Format*. Accessed: 2018. [Online]. Available: <http://bellard.org/bpg/>
- [15] M. Maldonado and J. WebP, *A New Web Oriented Image Format*. Barcelona, Spain: Universitat Oberta de Catalunya, 2010.
- [16] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34, no. 4, pp. 30–44, Apr. 1991.
- [17] 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.
- [18] 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.
- [19] F. Aulí-Llinàs, M. W. Marcellin, J. Serra-Sagrasta, and J. Bartrina-Rapesta, "Lossy-to-lossless 3D image coding through prior coefficient lookup tables," *Inf. Sci.*, vol. 239, pp. 266–282, Aug. 2013.
- [20] 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.
- [21] R. Pizzolante and B. Carpentieri, "Multiband and lossless compression of hyperspectral images," *Algorithms*, vol. 9, no. 1, p. 16, Feb. 2016.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] J. Balle, V. Laparra, and E. P. Simoncelli, "End-to-end optimization of nonlinear transform codes for perceptual quality," in *Proc. Picture Coding Symp. (PCS)*, Sep. 2016, pp. 1–5.
- [28] D. Minnen, J. Ballé, and G. D. Toderici, "Joint autoregressive and hierarchical priors for learned image compression," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 10794–10803.
- [29] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational image compression with a scale hyperprior," 2018, *arXiv:1802.01436*.
- [30] 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. 7936–7945.
- [31] Z. Guo, Z. Zhang, and R. Feng, "Causal contextual prediction for learned image compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 2329–2341, Jun. 2021.
- [32] 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.
- [33] M. Cao et al., "Entropy modeling via gaussian process regression for learned image compression," in *Proc. Data Compres. Conf. (DCC)*, 2022, pp. 163–172.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] S. Xiang and Q. Liang, "Remote sensing image compression with long-range convolution and improved non-local attention model," *Signal Process.*, vol. 209, Aug. 2023, Art. no. 109005.
- [38] S. Zhao, S. Yang, J. Gu, Z. Liu, and Z. Feng, "Symmetrical lattice generative adversarial network for remote sensing images compression," *ISPRS J. Photogramm. Remote Sens.*, vol. 176, pp. 169–181, Jun. 2021.
- [39] L. Zhang, X. Hu, T. Pan, and L. Zhang, "Global priors with anchored-stripe attention and multiscale convolution for remote sensing image compression," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 138–149, 2024.
- [40] P. Han, B. Zhao, and X. Li, "Edge-guided remote-sensing image compression," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5524515.
- [41] Z. Tang, H. Wang, X. Yi, Y. Zhang, S. Kwong, and C.-C. J. Kuo, "Joint graph attention and asymmetric convolutional neural network for deep image compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 1, pp. 421–433, Jan. 2023.
- [42] B. Li, J. Liang, and J. Han, "Variable-rate deep image compression with vision transformers," *IEEE Access*, vol. 10, pp. 50323–50334, 2022.
- [43] H. Ma, D. Liu, R. Xiong, and F. Wu, "iWave: CNN-based wavelet-like transform for image compression," *IEEE Trans. Multimedia*, vol. 22, no. 7, pp. 1667–1679, Jul. 2020.
- [44] 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.
- [45] S. Mei, R. Jiang, M. Ma, and C. Song, "Rotation-invariant feature learning via convolutional neural network with cyclic polar coordinates convolutional layer," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5600713.
- [46] Accessed: May 2023. [Online]. Available: <https://resources.maxar.com/product-samples/analysis-ready-data-san-francisco-california>
- [47] 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.
- [48] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, Nov. 2010, pp. 270–279.
- [49] 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.
- [50] 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. She is currently a Professor with the Department of Communication Engineering, Qiqihar University, Qiqihar, China. From 2024, she also works with the College of Information Engineering, Huzhou University, Huzhou, China. She has published two academic books about remote sensing image processing and more than 90 papers 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 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.



Zexin Zeng received the bachelor's degree from Guangzhou Maritime University, Guangzhou, 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.



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.



Ligu Wang (Member, IEEE) received the M.S. and Ph.D. degrees in signal and information processing from Harbin Institute of Technology, Harbin, China, in 2002 and 2005, respectively.

From 2006 to 2008, he held a Post-Doctoral Research position with Harbin Engineering University, Harbin. Since 2020, he has been working with the College of Information and Communication Engineering, Dalian Nationalities University, Dalian, China. He has published two books about hyperspectral image processing and more than 130 papers in journals and conference proceedings. His main research interests include remote sensing image processing.

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: A Multilevel Domain Similarity Enhancement Guided Network for Remote Sensing Image Compression

作者: Shi, CP (Shi, Cuiping); Shi, KJ (Shi, Kaijie); Zhu, F (Zhu, Fei); Zeng, ZX (Zeng, Zexin); Wang, LG (Wang, Ligu)

来源出版物: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING 卷: 62 文献号: 5645819 DOI: 10.1109/TGRS.2024.3483312 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:001350106800009

语言: English

文献类型: Article

地址: [Shi, Cuiping; Shi, Kaijie; Zhu, Fei; Zeng, Zexin] Qiqihar Univ, Dept Commun Engn, Qiqihar 161000, Peoples R China.

[Shi, Cuiping] Huzhou Univ, Coll Informat Engn, Huzhou 313000, Peoples R China.

[Wang, Ligu] Dalian Nationalities Univ, Coll Informat & Commun Engn, Dalian 116000, Peoples R China.

通讯作者地址: Shi, CP (通讯作者), Qiqihar Univ, Dept Commun Engn, Qiqihar 161000, 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; wangliguo@hrbeu.edu.cn

Affiliations: Qiqihar University; Huzhou University; Dalian Minzu University

IDS 号: L3Y3M

ISSN: 0196-2892

eISSN: 1558-0644

来源出版物页码计数: 19

特此证明

