Content-Based Hyperspectral Image Compression Using a Multi-Depth Weighted Map With Dynamic Receptive Field Convolution

Shaoming Pan, XiaoLin Gu, Yanwen Chong*, Yuanyuan Guo

The State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430072 (China)

Received 13 December 2021 | Accepted 21 June 2022 | Published 3 August 2022

unir
LA UNIVERSIDAD EN INTERNET

Abstract

In content-based image compression, the importance map guides the bit allocation based on its ability to represent the importance of image contents. In this paper, we improve the representational power of importance map using Squeeze-and-Excitation (SE) block, and propose multi-depth structure to reconstruct non-important channel information at low bit rates. Furthermore, Dynamic Receptive Field convolution (DRFc) is introduced to improve the ability of normal convolution to extract edge information, so as to increase the weight of edge content in the importance map and improve the reconstruction quality of edge regions. Results indicate that our proposed method can extract an importance map with clear edges and fewer artifacts so as to provide obvious advantages for bit rate allocation in content-based image compression. Compared with typical compression methods, our proposed method can greatly improve the performance of Peak Signal-to-Noise Ratio (PSNR), structural similarity (SSIM) and spectral angle (SAM) on three public datasets, and can produce a much better visual result with sharp edges and fewer artifacts. As a result, our proposed method reduces the SAM by 42.8% compared to the recently SOTA method to achieve the same low bpp (0.25) on the KAIST dataset.

Keywords

Compression, Dynamic Receptive Field Convolution, Hyperspectral Image, Importance Map, Multi-Depth.

DOI: 10.9781/ijimai.2022.08.004

I. Introduction

HYPERSPECTRAL images (HSIs) mainly own two kinds of redundancy, namely spectral similarity and spatial correlation [1]. As a typical 3D image, HSI compression has increasingly received attention in recent years to eliminate these two kinds of redundancy and achieve efficient image storage, transmission and processing [2]-[4].

Traditional lossy compression techniques, such as JPEG [5] and JPEG2000 [6] provide excellent rate-distortion performance for 2D imagery. In order to match the requirements of 3D image compression, a number of 3D compression algorithms including 3D-SPECK [7] and PCA+JPEG2000 [8] arise up for 3D HSI. However, these methods without the consideration of special characteristics of HSI by a direct extension from 2D to 3D may not fully satisfy the requirements of HSI compression [9]-[11], and the spectral fidelity of HSI cannot be guaranteed under the condition of effectively removing the spectral correlation of HSI.

In recent years, several DNNs-based lossy image compression methods [12]-[14] have achieved comparable performance to traditional methods [15],[16]. This is because deep convolutional network (DNNs) not only has good feature extraction ability, but also is good at flexible nonlinear analysis and comprehensive transformation of extracted spatial and spectral characteristics. The core research goal of DNNs-based lossy compression [17]-[19] is to balance compression ratio and the distortion to ensure the image quality [20],[21]. Bit-allocation based on the importance of image content has been effectively adopted in DNNs-based lossy image compression to achieve this goal [22],[23].

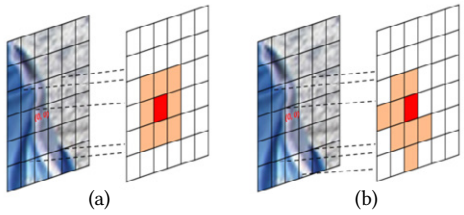


Fig. 1. The convolution is the process of the weighted summation. The red locations denote element to convolve, and the orange positions denote local receptive field to be weighted. (a) 3 × 3 instances of the normal convolution. (b) Our proposed Dynamic Receptive Field convolution (DRFc) with a kernel size of 3 × 3.

However, there are still several challenges in generating an accurate importance map based on the content of the image. An importance map is generally the representations produced by convolutional network that capture the salient contents of the image for bit allocation and compression rate control. In image compression, we usually want the bpp (bits per pixel) to be as small as possible, so a central theme of the importance map research is to search for more powerful representations that capture only the most salient properties of an image.

* Corresponding author.
E-mail address: ywchong@whu.edu.cn

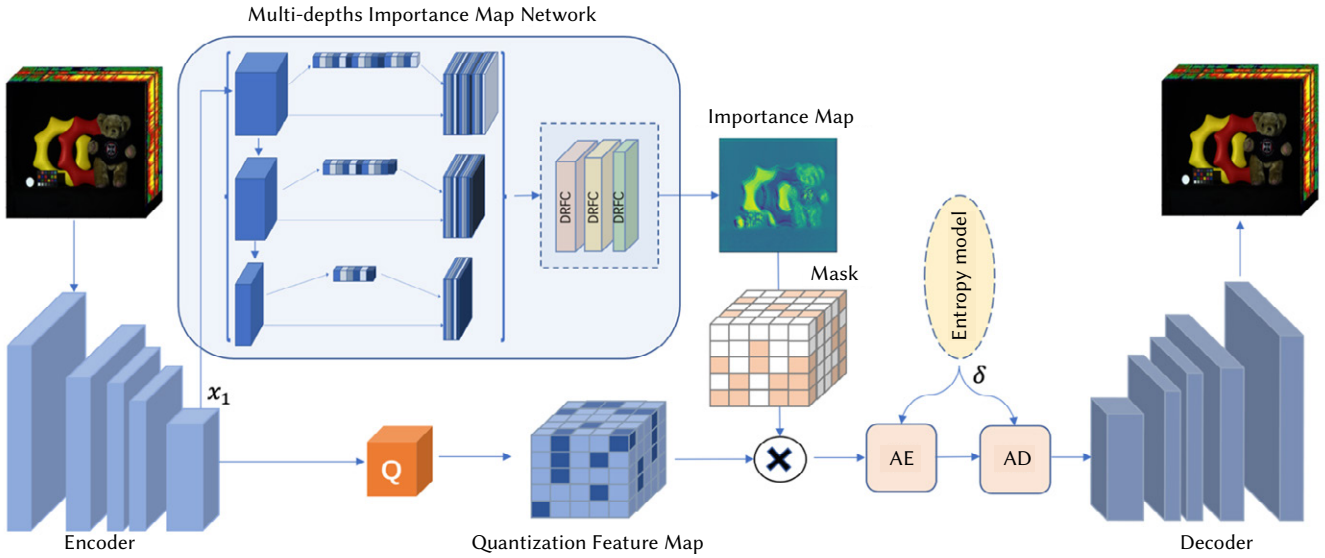


Fig. 2. Illustration of the proposed architecture for content-weighted image compression.

In addition, due to the fixed geometric structure of the convolution operator, as shown in Fig. 1(a), normal convolution has insufficient perception of edges in an image[24], resulting in a smaller value at the edge of the generated important map by CNNs. The larger the value of the importance map, the more bits for the image content are allocated. In this way, fewer bits will be allocated to the edges according to the importance map, and usually are inevitable in producing some visual artifacts, e.g., blurring and blocking in image reconstruction.

In this paper, we proposed a multi-depth importance map (MDIM) with Dynamic Receptive Field Convolution (DRFC) network (MDIMDRF), which is embedded into an encoder-decoder framework to produce an importance map and achieve content-based hyperspectral image compression. First, in our MDIM, we introduced the Squeeze-and-Excitation block(SE-block) to explicitly model the interdependencies between the channels of convolutional features and strengthen feature extraction of CNNs[25], thus improve the representational power of importance map. Since channel-wise information in single-depth importance map (SDIM) leads to excessive loss of non-important channels, and then compression performance often dramatically drops at low bpp, we designed the MDIM based on pyramid decomposition scheme to reconstruct non-important channel information at low bit rate. And then we introduced DRFC to greatly enhance CNNs' capability of extracting edge information. Finally, we replaced normal convolution with DRFc for the last three layers in MDIM and expected to improve the representation ability of important map synthetically.

To sum up, the main issues addressed in this paper are listed as follows:

1. Unlike other methods using simple convolution layers [22],[26]or residual blocks[23],[27] to obtain importance map, we designed MDIM to explicitly model the interdependence between feature channels and improve the representational power of importance map.
2. We reconsidered the guiding role of importance map to rate allocation in coding process, and retained the weak edges and mid-scale textures in the original image by increasing the weight of the regions with sharp edge of importance map.
3. The proposed compression framework can be end-to-end trained, and obtain significantly better results than state-of-the-art (SOTA) methods.

II. Methodology

As shown in Fig. 2, we proposed an end-to-end image compression model, which consists of encoder, MDIMDRF, entropy model, and decoder. Following[12], the encoder network consists of four convolutional layers and three generalized divisive normalization (GDN)[28]layers. The architecture of decoder is symmetric to that of the encoder. The MDIMDRF here can be understood as producing an importance map via MDIM and DRFc to obscure the non-important regions in the image so as to allocate more bits to the important regions.

A. Dynamic Receptive Field Convolution

As shown in Fig. 1(a), when convolving an edge pixel, the normal convolution unit samples the input feature map in a fixed receptive field, causing features to be influenced by irrelevant image content. For our DRFC, as shown in Fig. 1(b), after three steps (details in Fig. 3), we effectively find the $k \times k$ (k is the size of the convolution kernel) pixels with the strongest correlation with the convolution element as its dynamic receptive field.

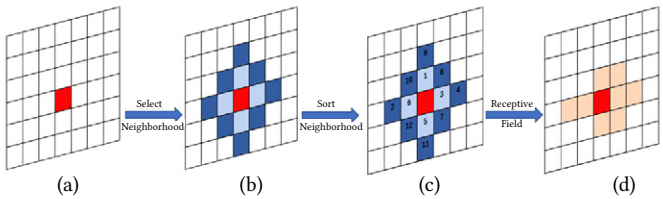


Fig. 3. Illustration of 3 × 3 Dynamic Receptive Field Convolution. The red grids denote pixels for convolution. Grids in light blue are first-order neighbors of the red, Grids in dark blue are second-order neighbors, and Grids in pink are the dynamic receptive field of the red.

For a $k \times k$ normal convolution, a receptive field $\mathcal{R}_{normal}$ (generally a $k \times k$ square grid) is constructed and moved over the input feature map x, with a scheduled step size s, The grid $\mathcal{R}_{normal}$ defines the receptive field size. For example, as shown in Fig. 1(a),

$$\mathcal{R}_{normal} = \{(-1,1), (0,1), \dots, (0,-1), (1,-1)\} \quad (1)$$

indicates the receptive field for a 3 × 3 normal convolution.

For each location p_0 on the output feature map y, we summate sampled values weighted by w and have

$$y(p_0) = \sum_{p_n \in \mathcal{R}_{normal}} w(p_n) \cdot x(p_0 + p_n) \quad (2)$$

where p_n enumerates the locations in $\mathcal{R}_{normal}$.

In our Dynamic Receptive Field Convolution, as shown in Fig. 3, we generate an irregular receptive field applying the following steps to the elements for convolution: (1) assemble a fixed-size neighborhood $\mathcal{R}^o_{DRFc}$ for each element;(2) sort the neighborhood and create receptive field $\mathcal{R}_{DRFc}$; (3) learn the receptive field representations with CNN. As shown in Fig. 3(b),

$$\mathcal{R}^o_{DRFc} = \{(-2,0),(0,-2),(-1,1),\dots,(1,-1),(2,0),(0,2)\} \quad (3)$$

Equation (2) becomes

$$y(p_0) = \sum_{p'_n \in \mathcal{R}_{DRFc}} w(p'_n) \cdot x(p_0 + p'_n) \quad (4)$$

where p'_n enumerates the locations in $\mathcal{R}_{DRFc}$.

TABLE I. RECEPTIVEFIELD: CREATE RECEPTIVE FIELD

1. input: Neighborhood $\mathcal{R}^o_{DRFc}$ of location p_0, Convolution kernel size k, Moran's Index m
2. output: Receptive Field $\mathcal{R}_{DRFc}$ of p_0
3. Compute an order r of the elements of $\mathcal{R}^o_{DRFc}$, subject to
$$\forall p, q \in \mathcal{R}^o_{DRFc}: m(p,p_0) < m(q,p_0) \leftrightarrow r(p) < r(q)$$
4. $\mathcal{R}_{DRFc}$ = top k^2 elements in $\mathcal{R}^o_{DRFc}$ according to r
5. return $\mathcal{R}_{DRFc}$

Illustrated in Fig. 3, Table I gives the procedures of creating receptive field by imposing an order on the elements $\mathcal{R}^o_{DRFc}$ via a correlation measure Moran's Index as

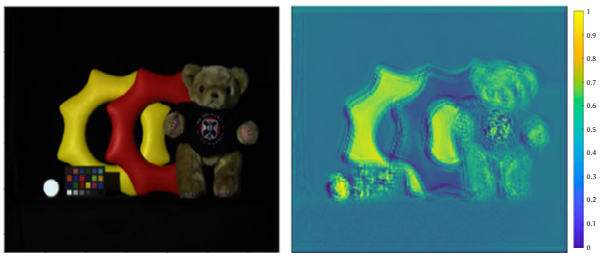
$$\mathrm{m}(p,q) = \frac{\sum_{i=1}^{c} w_{x_p,x_q}(x_{pi}-\overline{x_p})(x_{qi}-\overline{x_q})}{s_{x_p} s_{x_q} \sum_{i=1}^{c} x_{pi} x_{qi}} \quad (5)$$

where x_p, x_q be the vector at location p, q; c is the length of the tensor x_p, x_q; x_{pi}, x_{qi} is the i-th valve of x_p, x_q; w_{x_p,x_q} is the weight of spatial autocorrelation, which is generally the reciprocal of the distance between x_p and x_q; s_{x_p}, s_{x_q} is the variance of x_p, x_q.

The basic idea is to select the points in the adjacency domain in turn that have a high correlation with the center point in turn and apply them to each input channel if and only if they have similar structural roles in two feature maps.

B. Multi-depth Importance Map Network

When we encode an input image, we tend to allocate the bits efficiently according to spatial variant local image content, that is, fewer bits should be allocated to the smooth regions while more bits should be allocated to the regions with more information content, which makes it possible to improve the reconstructed image quality



(a) Original image (b) Importance map

Fig. 4. Illustration of Importance map.

while improving the compression ratio. For example, given the image in Fig. 4(a), it is natural to be interested in the teddy bear and two-colored circles, which are called the important regions. It is reasonable to allocate more bits to the teddy bear and two-colored circles and fewer bits to black background.

Thus we first designed a single-depth importance map network(SDIM) of four convolution layers[22] to retain the most important features of the image and generate an importance map to guide the allocation of bits. To improve the representational power of importance map, we strengthen feature extraction of CNNs using SE block[25] via modelling the correlation between feature channels and adjusting the feature map according to the correlation degree. Secondly, in order to compensate the excessive loss of non-important channels caused by channel-wise operation at low bit rate, we adopt a multi-depth importance map network (MDIM) based on pyramid decomposition scheme to reconstruct non-important channel information. As shown in Fig.5, we obtain the sub-importance maps generated by feature maps of different depths respectively, the results of each depth are weighted and summed to produce an importance map.

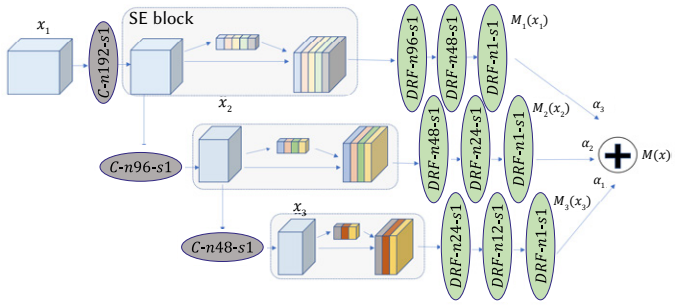


Fig. 5. Illustration of the MDIM's pyramidal decomposition structure with 3 depths. It's noted that "C-n192-s1" represents a CNN layer with 192 filters and a stride of 1 and "DRFC-n96-s1" represents a DRFC layer with 96 filters and a stride of 1 where the normal convolution unit is replaced by Dynamic Receptive Field convolution unit.

Let x_m denotes the input of the m-th layer of MDIM, and also x_1 denotes the original output of encoder. $M_m(x_m)$ represents the output of the m-th layer. In our paper, we sequentially set m to 1, 2, and 3 to individually produce a feature map containing different channel information with only one channel and the same size as the encoder output. The results of each scale are weighted and summed to produce the final importance map $M(x) = \alpha_1 M_1(x_1) + \alpha_2 M_2(x_2) + \alpha_3 M_3(x_3)$. What's more, DRFc instead of normal convolution is used in the last three layers of MDIM to enhance feature extraction of edge pixels, thus increasing the weight of regions with sharp edges or rich textures.

III. EXPERIMENTS

To evaluate the performance of the proposed compression model, we compared our model with traditional compression methods, i.e., 3D-SPECK[7], PCA+JPEG2000[8], and DNNs-based compression models, i.e., factorized prior[12], hyperprior[29] on different datasets. All DNNs-based experiments are conducted on a server equipped with the NVIDIA GeForce RTX 3090Ti graphics card.

We used three standard HSI datasets to train and test our proposed compression framework: KAIST[30], CAVE[31] and ICVL[32]. KAIST is a high-resolution dataset containing 30 images of size 2704×3376×31, CAVE consists of 28 images of 31×512×512 and ICLR consists of 201 images of 1300×1392×31. A total of 20000 patches with a size of 31×256×256 were sampled from both the original images and their enhancement (such as flipping and rotating at different angles). The data are divided into a training data set, a testing data set and a validation data set. Specifically, 60% of the images were used for training, 20% for testing and 20% for validation. Please note that all the test images are not included in the training dataset. Several original images from KAIST, CAVE and ICLR dataset are shown in Fig. 6, Fig. 7 and Fig. 8, respectively.

Fig. 6. Original image from KAIST.

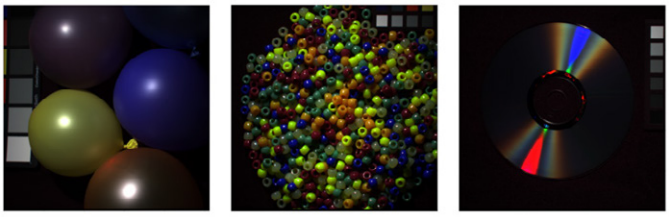


Fig. 7. Original image from CAVE.

Fig. 8. Original image from ICVL.

A. Performance Metrics

To quantitative evaluate the performance of proposed model, we used the following indexes as Peak Signal-to-Noise Ratio (PSNR)[18, 33], Structural Similarity Index Measure (SSIM) [33],[34] and Spectral Angle Mapper (SAM)[35].

1. Peak Signal-to-Noise Ratio

The ratio between the input image and the reconstructed image is known as PSNR. Also, the PSNR is measured based on the Mean Square Error (MSE)[36]. Please note that the PSNR for HSIs in this paper is calculated as in (6),

$$PSNR = \frac{1}{C}\sum_{i=1}^{C} 10log_{10}\left(\frac{p_{max}^2}{MSE}\right) \tag{6}$$

Where p_{max} denotes the maximum value in the i-th band of HSIs, and the unit of PSNR is dB.

2. Structural Similarity Index Measure

It is used to evaluate the distortion between the input image x and the reconstructed image x^*, it can be defined as in (7),

$$SSIM(x, x^*) = \frac{1}{C}\sum_{i=1}^{C} \frac{(2\mu_{x_i}\mu_{x_i^*} + a_1)(2\sigma_{x_i x_i^*} + a_2)}{(\mu_{x_i}^2 + \mu_{x_i^*}^2 + a_1)(\sigma_{x_i}^2 + \sigma_{x_i^*}^2 + a_2)} \tag{7}$$

Where C is the number of bands of input image x, x_i is i-th band of x, and μ_{x_i}, σ_{x_i} are the corresponding mean and variance. a_1, a_2 is constant.

3. Spectral Angle Mapper

The spectrum of each pixel in HSIs is regarded as a high-dimensional vector, and the similarity between the two spectrums is measured by calculating the Angle between the two vectors. Note that a small SAM value indicates less spectral distortion.

B. Training Details and Parameter Settings

Our objective is to minimize the weighted sum of the rate loss and distortion loss, R + λD, where λ governs the trade-off between the two terms. Thus, we trained the model on the batch of size B, and defined the loss function $\mathcal{L}$ of our model on the entire batch:

$$\mathcal{L} = \frac{1}{B}\sum_{i=1}^{B}\{\mathcal{L}_R(c, x^i) + \lambda\mathbb{E}[d(x^i, {}^{\circ}\widehat{x^i})]\} \tag{8}$$

where c is the code of the input image x^i. $\mathcal{L}_D(c, x^i)$ denotes the rate loss and $d(x^i, {}^{\circ}\widehat{x^i})$ is the expected difference between the reconstruction $\widehat{x^i}$ and the original image x^i, as measured by Mean Square Error (MSE) in order to be consistent with PCA+JPEG2000[8].

Firstly, we set the weights α_1, α_2, and α_3 in the MDIM to 1/2, 1/4, and 1/4, respectively. During the training process, we set the batch parameter B to 8 and the model is iteratively trained 300 times on the dataset. In addition, the initial learning rate is set to 10^{-4}, and performs stochastic gradient descent[37] using the Adam algorithm[38].

With this setup, we trained a total of 24 separate models: half of the models with MDIM and half without; half of the models with DRFc, and half without; finally, each of these combinations with 6 different values of λ in order to cover a range of rate-distortion tradeoffs.

C. Comparison of Rate–Distortion Performance

In this subsection, we evaluate the performance improvements of the proposed model quantitatively, and rate–distortion curves for different methods on KAIST, CAVE and ICLR datasets are provided in Fig. 9, respectively.

Firstly, we compare the PSNR and SSIM performance of our proposed method with PCA+JPEG2000 and 3D SPEAK as well as the methods proposed in[12, 29].As seen from Fig. 9, our method outperforms traditional methods[7, 8] and DNNs-based methods[12, 29] at a wide range of bpp on three datasets. Although the PSNR and SSIM performance of our method owns only a relatively small advantages on CAVE dataset comparing with factorized prior[12] and hyperprior[29], the corresponding performance improvement is particularly obvious on KAIST and ICLR dataset. This is because the spatial resolution of individual KAIST dataset is almost 35 times higher than that of CAVE dataset and there are 50 times more ICLR training patches than CAVE training patches, larger dataset makes the model more fully trained and the test performance better.

Next, we further compare the SAM performance of different methods based on the work of[35], As seen from Fig. 9, the average rate-distortion curves of SAM show that the proposed method can significantly outperform other methods on three datasets and the SAM performance of our method is still superior to other methods at low bpp on CAVE dataset. For example, compared with factorized prior[12] and hyperprior[29], the SAM of our proposed model is reduced by 0.03 and 0.02 when bpp is 0.25, respectively. A strong explanation is that the importance map network designed in our proposed model takes full account of spectral similarity, and retains spectral characteristic information to the maximum extent.

D. Comparison of Visual Quality

The visual quality comparisons of the reconstructed HSIs in low compression rates for three datasets are provided in Table II. As can be seen from Table II, traditional compression methods such as 3D SPEAK and PCA+JPEG 2000 inevitably produce obvious blurring, ringing in the second and third columns, which can seriously affect the human visual experience. The methods[12, 29] suppress the artifacts effectively, but there are still some blur effects along the edges visible in the fourth and fifth columns. In contrast, our method overcomes the above flaws, and some important edges and textures are well-retained and thus the reconstructed image owns better visual quality due to the bit-allocation guided by the importance map.

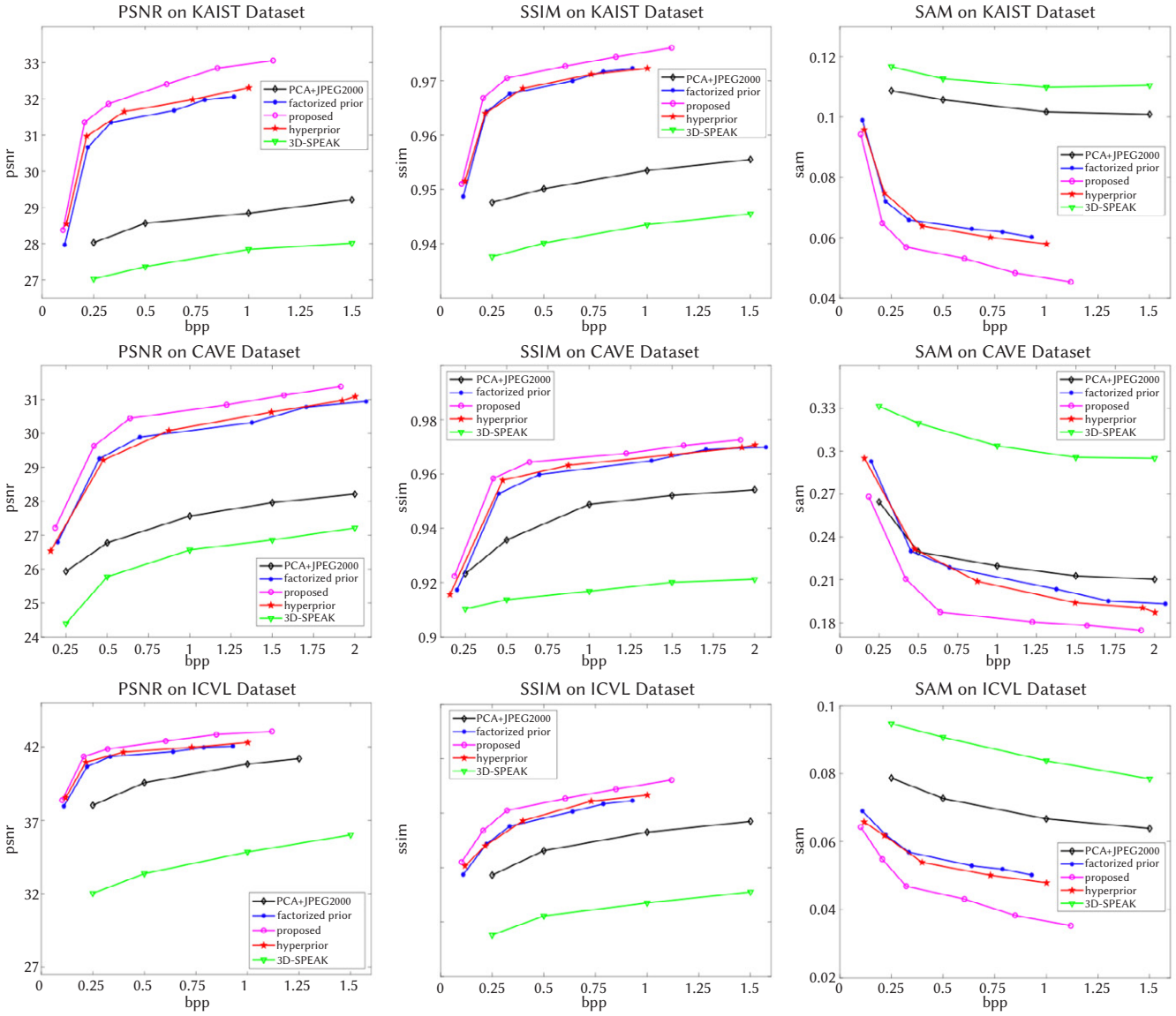


Fig. 9 Comparison of the ratio-distortion curves by different metrics: PSNR, SSIM, and SAM

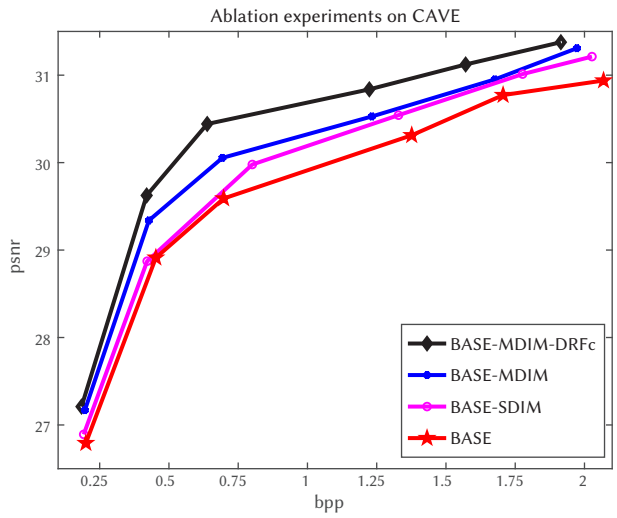


Fig. 10: Illustration of the results of the ablation experiment.

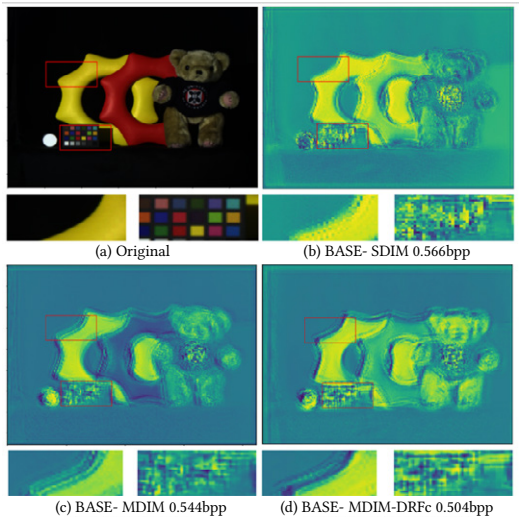
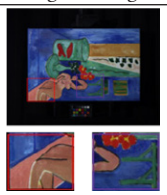
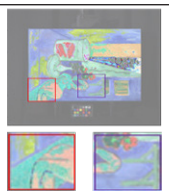
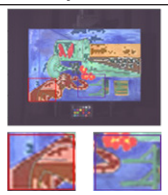
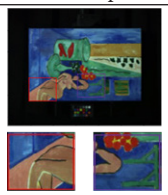
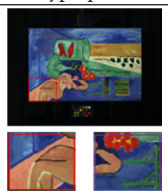
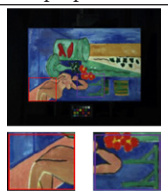
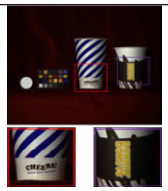
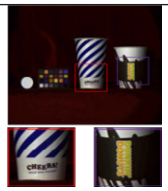
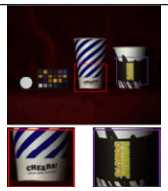
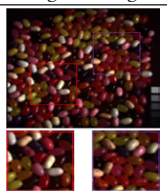
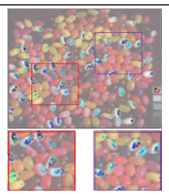
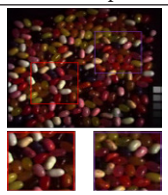
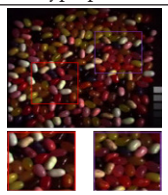
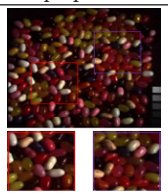
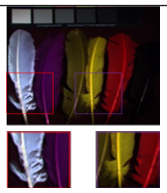
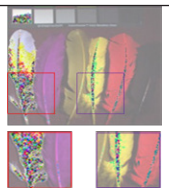
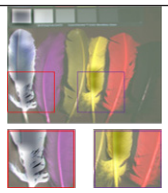
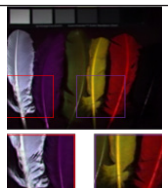
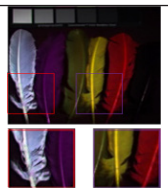


Fig. 11. The important maps obtained by different models. The right color bar shows the palette on the number of bits.

TABLE II. Images Produced By Different Compression Models At Different Compression Rates. All Images Are Visualized With the Same Ordinal Band

Original image	3D-SPEAK	PCA+JPEG2000	factorized prior	hyperprior	proposed
KAIST dataset (29,19,9)					
					
	0.5 bpp PSNR:18.46 SSIM:0.908 SAM:0.430	0.5 bpp PSNR:22.53 SSIM:0.935 SAM:0.263	0.515bpp PSNR:39.43 SSIM:0.982 SAM:0.124	0.543bpp PSNR:40.01 SSIM:0.984 SAM:0.112	0.499bpp PSNR:40.29 SSIM:0.985 SAM:0.102
					
	0.7 bpp PSNR:21.42 SSIM:0.31 SAM:0.72	0.7 bpp PSNR:29.16 SSIM:0.80 SAM:0.42	0.713bpp PSNR:41.94 SSIM:0.992 SAM:0.089	0.748bpp PSNR:43.13 SSIM:0.993 SAM:0.0998	0.676bpp PSNR:44.57 SSIM:0.9946 SAM:0.073
CAVE dataset (24,6,25)					
Original image	3D-SPEAK	PCA+JPEG2000	factorized prior	hyperprior	proposed
					
	0.7 bpp PSNR:23.68 SSIM:0.93 SAM:0.42	0.7 bpp PSNR:28.96 SSIM:0.94 SAM:0.26	0.743bpp PSNR:29.49 SSIM:0.961 SAM:0.210	0.723bpp PSNR:29.13 SSIM:0.958 SAM:0.234	0.711bpp PSNR:29.23 SSIM:0.956 SAM:0.216
					
	0.5bpp PSNR:24.69 SSIM:0.67 SAM:0.76	0.5bpp PSNR:33.16 SSIM:0.84 SAM:0.45	0.565bpp PSNR:31.66 SSIM:0.961 SAM:0.188	0.526bpp PSNR:32.15 SSIM:0.969 SAM:0.179	0.483bpp PSNR:33.14 SSIM:0.975 SAM:0.158
ICVL dataset (29,19,9)					
Original image	3D-SPEAK	PCA+JPEG2000	factorized prior	hyperprior	proposed
					
	0.7bpp PSNR:28.58 SSIM:0.65 SAM:0.07	0.7bpp PSNR:45.41 SSIM:0.993 SAM:0.036	0.737bpp PSNR:53.39 MS-SSIM:0.999 SAM:0.037	0.710bpp PSNR:54.27 MS-SSIM:0.999 SAM:0.035	0.703bpp PSNR:54.48 MS-SSIM:0.999 SAM:0.029
					
	0.5bpp PSNR:26.06 SSIM:0.486 SAM:0.113	0.5bpp PSNR:44.33 SSIM:0.993 SAM:0.056	0.517bpp PSNR:47.81 MS-SSIM:0.999 SAM:0.038	0.539bpp PSNR:50.19 MS-SSIM:0.999 SAM:0.030	0.525bpp PSNR:51.25 MS-SSIM:0.999 SAM:0.026

E. Ablation Experiments

To assess the role of MDIM and DRFc, we trained a baseline model by removing MDIMDRF from our framework. We designed the following four models according to whether the presence of SDIM, MDIM, and DRFc in the architecture: (1) BASE: the baseline model; (2) BASE-SDIM: BASE with SDIM; (3) BASE-MDIM: BASE with MDIM; (4) BASE-MDIM-DRFc: BASE with MDIM and DRFc.

As shown in Fig. 10, at the same bpp, BASE-MDIM-DRFc has the best performance while BASE has the worst performance. BASE-MDIM performs better than BASE-SDIM at low bpp, which proves MDIM's help in reconstructing the non-important channels of convolutional features at low bpp. In Fig. 11, we can observe the blurring artifacts and color distortion in (b) and (c). In contrast, the results in (d) exhibit much clearer and is much more consistent with human visual perception.

IV. Discussion

In our proposed end-to-end compression framework, we design the multi-depths importance map network based on pyramidal decomposition, and produce an importance map to guide bit rates allocation and further compress the code by entropy coding. At the same time, we introduce Dynamic Receptive Field convolution to increase the weight of the importance map in the edge area to solve the distortion caused by insufficient feature representation to edge of normal convolution.

Rate-distortion performance in Fig. 9 clearly shows that our proposed method outperforms conventional and DNNs-based methods at a wide range of bpp. In addition, as shown in Fig. 11, the existence of multi-depth importance map and Dynamic Receptive Field convolution have significant influence on the performance improvement. In addition, to achieve PSNR of 40 and MI-SSIM of 0.95, the average time to encode and decode the image is 49 *ms* and 11 *ms*, running on the GeForce RTX 3090Ti.

V. Conclusion

In this paper, we proposed a content-based compression system for hyperspectral images. In the proposed system, we designed MDIM and DRFc to improve representability of the importance map so as to allocate bits precisely for different contents. Our models can be end-to-end learned on a training set. Experimental results clearly show the superiority of our model in retaining HSI's spectral structure characteristics and extracting edge content, resulting in significant image reconstruction quality.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant Nos. 62072345, 41671382), State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing Special Research Funding. The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

References

[1] E. Christophe, C. Mailhes, and P. Duhamel, "Hyperspectral Image Compression: Adapting SPIHT and EZW to Anisotropic 3-D Wavelet Coding," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society,* vol. 17, no. 12, p. 2334, 2008.

[2] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, no. Dec., pp. 279-317, 2019.

[3] B. Penna, T. Tillo, E. Magli, and G. Olmo, "Transform Coding Techniques for Lossy Hyperspectral Data Compression," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 45, no. 5, pp. 1408-1421, 2007.

[4] Y. Qu, H. Qi, and C. Kwan, "Unsupervised Sparse Dirichlet-Net for Hyperspectral Image Super-Resolution," in *Conference on Computer Vision and Pattern Recognition*, 2018.

[5] G. K. Wallace, "The JPEG Still Picture Compression Standard."

[6] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 still image compression standard," *IEEE Signal Processing Magazine,* vol. 18, no. 5, pp. 36 - 58, 2001.

[7] X. Tang and W. A. Pearlman, "Lossy-To-Lossless Block-Based Compression of Hyperspectral Volumetric Data," in *IEEE*, 2006.

[8] Q. Du and J. E. Fowler, "Hyperspectral Image Compression Using JPEG2000 and Principal Component Analysis," *IEEE Geoscience & Remote Sensing Letters*, vol. 4, no. 2, pp. 201-205, 2007.

[9] H. Ying and G. Liu, "Hyperspectral image lossy-to-lossless compression using the 3D Embedded Zeroblock Coding alogrithm," in *International Workshop on Earth Observation & Remote Sensing Applications*, 2008, pp. 1-6.

[10] A. Karami, S. Beheshti, and M. Yazdi, "Hyperspectral image compression using 3D discrete cosine transform and support vector machine learning," in *International Conference on Information Science*, 2012.

[11] E. Christophe, D. Leger, and C. Mailhes, "Quality criteria benchmark for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 9, pp. 2103-2114, 2005.

[12] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end Optimized Image Compression," 2016.

[13] J. Lee, S. Cho, and S. K. Beack, "Context-adaptive Entropy Model for End-to-end Optimized Image Compression," *arXiv*, 2018.

[14] L. Galteri, L. Seidenari, M. Bertini, and A. Bimbo, "Deep Generative Adversarial Compression Artifact Removal," *IEEE Transactions on Multimedia,* 2017.

[15] H. A. Jawdhari, "Hyperspectral image compression using sparse representations and wavelet transform based spectral decorrelation," 2017.

[16] A. Karami, "Compression of hyperspectral images using discerete wavelet transform and tucker decomposition," *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing,* vol. 5, no. 2, pp. 444-450, 2012.

[17] C. Kwan and J. Larkin, "New Results in Perceptually Lossless Compression of Hyperspectral Images," *Journal of Signal and Information Processing,* vol. 10, no. 3, pp. 96-124, 2019.

[18] L. Theis, W. Shi, A. Cunningham, and F. Huszár, "Lossy Image Compression with Compressive Autoencoders," 2017.

[19] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. V. Gool, "Practical Full Resolution Learned Lossless Image Compression," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[20] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *IEEE Conference on Computer Vision & Pattern Recognition*, 2011.

[21] Y. Blau and T. Michaeli, "Rethinking Lossy Compression: The Rate-Distortion-Perception Tradeoff," 2019.

[22] L. Mu, W. Zuo, S. Gu, D. Zhao, and D. Zhang, "Learning Convolutional Networks for Content-weighted Image Compression," 2017.

[23] L. Wu, K. Huang, and H. Shen, "A GAN-based Tunable Image Compression System," 2020.

[24] J. Dai *et al.*, "Deformable Convolutional Networks," *IEEE*, 2017.

[25] H. Jie, S. Li, S. Gang, and S. Albanie, "Squeeze-and-Excitation Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. PP, no. 99, 2017.

[26] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Performance Comparison of Convolutional AutoEncoders, Generative Adversarial Networks and Super-Resolution for Image Compression," 2018.

[27] B. Zla, A. Sy, A. Yw, and B. Jza, "Discrimination of the fruits of Amomum tsao-ko according to geographical origin by 2DCOS image with RGB and Resnet image analysis techniques," *Microchemical Journal,* 2021.

[28] J. Ballé, V. Laparra, and E. P. Simoncelli, "Density Modeling of Images using a Generalized Normalization Transformation," *arXiv*, 2015.

[29] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational

image compression with a scale hyperprior," 2018.
[30] I. Choi, M. Kim, D. Gutierrez, D. Jeon, and G. Nam, "High-quality hyperspectral reconstruction using a spectral prior," 2017.
[31] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum," *IEEE transactions on image processing,* vol. 19, no. 9, pp. 2241-2253, 2010.
[32] K. Yi *et al.*, "CLEVRER: CoLlision Events for Video REpresentation and Reasoning," 2019.
[33] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. V. Gool, "Conditional Probability Models for Deep Image Compression," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.
[34] G. Toderici *et al.*, "Full Resolution Image Compression with Recurrent Neural Networks," *IEEE Computer Society,* 2016.
[35] F. A. Kruse *et al.*, "The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data," *Remote Sensing of Environment,* 1993.
[36] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. V. Gool, "Generative Adversarial Networks for Extreme Learned Image Compression," 2018.
[37] L. Zhou, C. Cai, Y. Gao, S. Su, and J. Wu, "Variational Autoencoder for Low Bit-rate Image Compression," 2018.
[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.

Shaoming Pan

Shaoming Pan received the B.S. and M.S. degrees from Huazhong University of Science and Technology, China, in 1998, Ph.D. degree from Wuhan University, China. He is an Associate Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China. He is mainly engaged in the research on image processing, multimedia communication and spatial information storage.

Xiaolin Gu

Xiaolin Gu is currently pursuing the master's degree with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China. Her main research interest is hyperspectral image compression based on neural network.

Yanwen Chong

Yanwen Chong received the B.S. degree from the Qufu Normal University, China, in 1995, M.S. and Ph.D. degrees from Wuhan University, China, in 1998 and 2001. He is a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China. His research interests include image processing, computer vision and pattern recognition.

Yuanyuan Guo

Yuanyuan Guo is currently pursuing the Ph.D. degree with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China. Her research interests are mainly within the fields of hyperspectral image compression, model-based deep learning.