SAR Image Compression Based on Low-frequency Suppression and Target Perception

Jiawen Deng[1,2,3], Lijia Huang[1,2] and Yifan Wu[1,2,3]

[1] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
[2] Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China
[3] School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100190, China

Abstract. Synthetic Aperture Radar (SAR) images are widely used in the field of remote sensing. To store and transmit the growing amount of SAR image data, more efficient compression algorithms are required. In this paper, a new framework for compressing SAR images is proposed based on deep learning. Firstly, we propose a new two-stage transformation based on low-frequency suppression of input data to achieve high information entropy and low quantization loss. To explore the redundancy between regions of interest and regions of non-interest, an image compression model for target perception is proposed to convert the input SAR image into a compact latent representation together with the target perception map. To evaluate the performance of the algorithm, we conducted experiments on SAR image dataset. The results show that the proposed algorithm is significantly better than the traditional processing algorithm in terms of data input, and can effectively preserve and improve the performance of the image compression model.

Keywords: Lossy compression, Image learning compression, Low-frequency suppression, Target perception, Synthetic Aperture Radar (SAR) image compression.

1 Introduction

In recent years, various types of SAR systems have been developed around the world, and the research and launch of various SAR satellites are increasing. The application of SAR images in various fields such as aerospace is increasing, and the data and quality of SAR images are constantly improving [1]. SAR images have the inherent characteristics of complex data structure and concentrated pixel distribution, with high resolution, large imaging range and high data volume [2,3]. At the same time, there are few targets of concern such as ships, and wide non-areas of concern such as sea surface, resulting in the high cost of transporting, storing, and managing SAR images. At the same time, it brings a huge obstacle to downstream tasks such as SAR image detection, recognition, segmentation, etc.

EAI MLICOM 2023, December 17, Beijing, People's Republic of China

DOI 10.4108/eai.17-12-2023.2342474

In 2016, CNN networks were first applied in the field of image compression. Balle proposed an end-to-end image compression framework based on variational autoencoder convolutional neural network [4]. In 2018, Balle et al. improved the end-to-end convolutional neural network compression framework. This framework uses a variational autoencoder to process the data, and proposes a hyper-prior network to capture the latent data structures [5]. Minnen et al. improved the entropy coding link and designed an enhanced entropy coding context model based on Balle's work [6], which is the first deep learning compression coding method that is better than BPG in the objective evaluation of images.

Li and Liu introduced CNN into remote sensing image compression, using a simple two-layer CNN as the baseline. The results obtained by DCT transformation and entropy encoding and decoding were better than those of BPG [7]. Xu et al. proposed a variational autoencoder SAR image compression model with a priori model, which combines residual blocks with transformations, and the results are better than JPEG, JPEG000 and Li [8]. Zhang et al. changed the single Gaussian model to a mixed Gaussian model for fitting and estimating the model parameters, which outperformed the traditional compression methods and learning-based algorithms on both ICEYE and Sandia datasets [9]. To explore spatial redundancy, Fu et al. introduced local and global context information, and introduced multiple residual modules instead of convolution to improve the performance of the model [10]. In the same year, Fu et al. proposed a model consisting of multiple prior networks to deeply explore redundancy in spatial results [11].

However, most models often do not pay attention to the errors at the data input level, and apply traditional methods such as linear stretching for quantification, resulting in the loss of input features [12]. In fact, most models only focus on global and local spatial redundancy, and do not pay attention to the difference between the target of interest and the non-target of interest [7-11], which will lead to the sub-optimal compression performance of the model.

To reduce the loss of image features and explore the redundancy in the spatial structure, a SAR image compression model based on low-frequency suppression and target perception was proposed. Our main work contributions are summarized below:

(i) In this paper, a SAR image compression model based on two-stage low-frequency suppression and target perception guidance are proposed, and the better image compression index is achieved through experiments.

(ii) This paper mainly constructs a two-stage transformation operator at the input data level to suppress low-frequency input data, achieving peak signal-to-noise ratio and low quantization loss data input.

(iii) we construct a compression model guided by target perception graph, guide the allocation of compression bit rate, explore the redundancy between focused and non-focused targets, and achieve a higher level of information fidelity for image compression models that focus on target perception.

2 System Model

This section describes the models that are common to the entire system. Figure 1 introduces the recommended model structure, which introduces a hyper-prior network based on the variational self-encoding and decoding model [4-6], which is used to capture the structural redundancy between the latent representation feature maps, resulting in higher network accuracy, better compression performance, and less loss of complex data. The optimization problem of the algorithm can be modeled as a variational autoencoder model [13], the entropy model corresponds to the variational autoencoding model with a prior of the hidden layer representation, and the edge information σ and u generated by the hyper-prior network are a priori of the entropy model.

In Figure 1, our baseline model is the hyper-prior network. The first part is the pre-processing and post-processing modules, which can suppress the low-frequency points of the image by constructing a two-stage transformation operator, and realize the data post-processing restoration with low ingestion parameters. After preprocessing, the data was pre-trained and the model ViT was used to obtain the quality feature map [14]. The quality feature map and input data are together used as inputs to obtain the potential representation by the encoder g_a. Then the processed data is quantized, entropy encoded and decoded, and the image is reconstructed by the decoder g_s. The quantized data probability model is generally regarded as a joint known distribution, and then the entropy of the image is encoded and decoded by calculating its probability model [15], which is used for image storage and transportation.

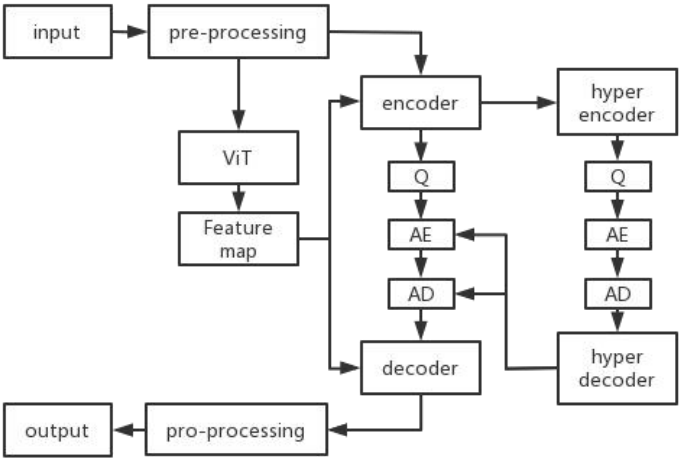

Fig. 1. Image compression model for hyper-prior networks.

However, there is a gap between the known distribution of the joint and the actual distribution y, as the actual distribution is unknown, so it is necessary to make the probability model distribution as close as possible to the actual distribution. In the process of image entropy encoding, if the parameters of the probabilistic model $\tilde{y}$ are known, then good image compression encoding can be achieved in arithmetic encoding and decoding. To reduce the gap between the probabilistic model and the actual model, the boundary variable z is introduced through the hyper-prior coding network to achieve accurate estimation of the probabilistic model：

$$z = h_a(y; \emptyset_h) \tag{1}$$

where $\emptyset_h$ refers to the parameters in the hyper-prior encoding network. After quantization, entropy encoding and decoding are redefined as $\tilde{z}$ and the distribution of y can be expressed by using the parameter generation model h_s. This can be expressed as:

$$p_{\hat{y}|\hat{z}}(\hat{y}|\hat{z}) \leftarrow h_s(\hat{z}; \theta_h) \tag{2}$$

where θ_h refers to the parameter of the hyper-prior decoding network h_s, $p_{\hat{y}|\hat{z}}(\hat{y}|\hat{z})$ means the estimate of the distribution model. Here we estimate the distribution of our model by gaussian mixture model [16,17], it can be expressed as:

$$p_{\hat{y}|\hat{z}}(\hat{y}|\hat{z}) \sim \sum_k N_k\left(\mu_k, \sigma_k^2\right) \tag{3}$$

μ and σ is the estimation of the mean and standard deviation of the Gaussian model. In this paper, we select k=3 in practice, and the parameter fitting estimation is carried out by three Gaussian mixed models. Then the processed parameters $\hat{y}$ is input into the image decoder g_s to realize the reconstruction of the image.

As for the parameter optimization problem of the whole network, the overall rate-distortion function is used [18,19], and the loss function of the model can be expressed as:

$$L = R + \lambda D \tag{4}$$

where D represents the distortion between the original image and the reconstructed image, which is usually replaced by MSE, PSNR, and SSIM [20,21]. R represents the compressed bitrate of the overall framework:

$$R = R_{\hat{y}} + R_{\hat{z}} \tag{5}$$

The hyper-prior network and the underlying convolutional network are jointly trained. By using different parameters λ for weighting, the optimal image compression reconstruction performance at different bitrates is achieved in rate-distortion optimization.

3 Proposed Algorithm

In this section, we mainly focus on the design and introduction of low-frequency suppression algorithms and target perception models.

3.1 Two-stage Low-frequency Suppression Algorithm

When the original data is linear to 0-255, most of the gray scale is within 0-30. Direct quantification of raw data results in significant. Therefore, it is necessary to design a operator to suppress the low-frequency part of the data and expand the high-frequency part of the data to achieve low information loss at the input level. Traditional low-frequency suppression processing methods include linear stretching, histogram equalization, and power conversion method.

The method of linear stretching is to stretch the data from 0-255, with pixels greater than a specific threshold discarded as input data. This method can save most of the information data, but there is a greater loss of information at the point with high pixel value, resulting in a large loss of information at the input data level.

The central idea of the histogram equalization method is to expand the low-frequency suppression and high-frequency, but the degree of suppression and expansion varies according to the number of pixels in the range. Within a certain range, it is like the linear stretching method and requires additional recording of larger parameters, which is inconsistent with the original intention of image compression.

The method of power conversion is a trade-off between the linear stretching method and the histogram equalization method, and the power-based method only needs to record one parameter and can better achieve the suppression of low frequency, but with the different parameters in the model, the high-frequency part is expanded too much.

Therefore, based on the existing methods, a two-stage low-frequency suppression operator is designed, which realizes high-frequency expansion by linear stretching of the main high-frequency part, and realizes low-frequency suppression by compressing the sub-high-frequency and low-frequency parts, to achieve the purpose of low information loss and few parameters at the data input level.

SAR Model Analysis. SAR images can be described by Rayleigh distribution, and the original Rayleigh distribution can be fitted with only one parameter σ to represent the gray distribution of the original data, and the probability distribution function of Rayleigh distribution can be expressed as:

$$y = \frac{x}{\sigma^2} * e^{-\frac{x^2}{2*\sigma^2}} \tag{6}$$

The power conversion method and histogram equalization method have achieved better quantitative reconstruction than linear, and at the same time meet the properties of tail compression and front end stretching. However, the stretch of the front end may be too large depending on hyper parameter value, and the histogram equalization method introduces more additional parameters. Therefore, it is necessary to find a

transformation function that satisfies the following two points: The transformation function and its inverse transformation function should be continuous and have as few parameters as possible; The front-end stretching should not be extreme at 0, and the ideal stretching method should be histogram-like balanced, weighted by the size of the grayscale histogram, and the degree of stretching should be gentle.

Two-stage Low-frequency Suppression Model. To meet the above two points, the linear and power two-stage low-frequency suppression function g(x) is constructed:

$$g(x) = \begin{cases} g(t) * x/t & \text{while } x < t \\ 255 * \log(\frac{x}{k} + 1)/\log\left(\frac{255}{k} + 1\right) & \text{while } x \geq t \end{cases} \quad (7)$$

In function 7, g(x) has two hyperparameters t and k which we need to optimize, where g(t) is:

$$g(t) = 255 * \log(\frac{t}{k} + 1)/\log\left(\frac{255}{k} + 1\right) \quad (8)$$

The inverse function of g(x) can be expressed as $g^{-1}(t)$ which is used for post-processing:

$$g^{-1}(x) = \begin{cases} t * x/g(t) & \text{while } x < g(t) \\ k * \left(10^{x*\log\left(\frac{255}{k}+1\right)/255} + 1\right) & \text{while } x \geq g(t) \end{cases} \quad (9)$$

Quantitative Loss Function Analysis. The MSE quantization loss function directly quantized by the traditional method of raw data can be defined as:

$$loss_{raw} = \int_0^{255} y(x) * (x - round(x))^2 = \sum_{m=0}^{255} \int_{m-0.5}^{m+0.5} y(x) * (x - m)^2 dx \quad (10)$$

Where x in (m-0.5, m+0.5), curve y is continuously smoothed and simplified to a straight-line segment $y_{new} = y'(m) * (x - m) + y(m)$. And the curve is symmetrical with respect to m, y(m) center in (m-0.5, m+0.5). So, the y(x) term in the Function 10 can be reduced to the constant y(m). The MSE loss function is a constant with respect to the σ:

$$loss_{raw} = \sum_{m=0}^{255} y(m) \int_{m-0.5}^{m+0.5} (x - m)^2 dx = \sum_{m=0}^{255} y(m)/12 \cong \frac{1 - e^{-\frac{255^2}{2\sigma^2}}}{12} \quad (11)$$

After a two-stage low-frequency rejection transformation, the MSE loss function $loss_{pro}$ is quantified:

$$loss_{pro} = \int_0^{255} y(x) * (x - round(x))^2 = \sum_0^{255} \int_{g^{-1}(m-0.5)}^{g^{-1}(m+0.5)} y(x) * (x - g^{-1}(m)) dx \quad (12)$$

Parametric Analysis of Loss Functions. The loss problem can be further simplified as a loss problem of 0-t internal tension and T-255 internal compression. After simplification, x is regarded as a straight-line segment in $g^{-1}(m)\ to\ g^{-1}(m+1)$, and the simplified result is in Function 12,13:

$$\mathrm{loss_{pro}} = \sum_{g(0)}^{g(255)} y(g^{-1}(m)) \int_{g^{-1}(m-0.5)-g^{-1}(m)}^{g^{-1}(m+0.5)-g^{-1}(m)} x^2 dx \tag{13}$$

$$\mathrm{loss_{pro}} = \sum_{g(0)}^{g(t)} \frac{y\left(g^{-1}(m)\right)}{12\left(\frac{g(t)}{t}\right)^3} + \sum_{g(t)}^{g(255)} y(g^{-1}(m))((h(m+0.5))^3 + (h(m-0.5))^3)/3 \tag{13}$$

Where $h(m) = g^{-1}(m+0.5) - g^{-1}(m)$. Let $l = g(t,k)/t$, and $l' = (255-kt)/(255-t)$ means the average value of the derivatives of the change function from t to 255 l'=(255-kt)/(255-t). So, the first and second terms of loss are as follows:

$$\mathrm{loss_{pro1}} = \sum_{g(0)}^{g(t)} y(g^{-1}(m))/(12l^3) = \frac{1-e^{-\frac{t^2}{2\sigma^2}}}{12l^3} \tag{14}$$

$$\mathrm{loss_{pro2}} = \frac{e^{-\frac{t^2}{2\sigma^2}} - e^{-\frac{255^2}{2\sigma^2}}}{12(l')^2} \cong \frac{(255-t)^2 e^{-\frac{t^2}{2\sigma^2}}}{12(255-lt)^2} \tag{15}$$

Therefore, the loss function $\mathrm{loss_{pro}}$ after simplified processing is expressed as:

$$\mathrm{loss_{pro}} = \frac{1-e^{-\frac{t^2}{2\sigma^2}}}{12l^2} + \frac{(255-t)^2 e^{-\frac{t^2}{2\sigma^2}}}{12(255-lt)^2} \tag{16}$$

$loss_{pro}$ (l, t, σ) is a resolvable loss function in the range of 1<l<255/t,0<t. Therefore, the final solution to the problem boils down to:

$$\min_{l,t\epsilon D}\ \mathrm{loss_{pro}}(l,t,\sigma) \text{ subject to } 1 < l < \frac{255}{t}, 0 < t, l = \frac{g(t,k)}{t} \tag{17}$$

From Function 10-17, we can get a minimum quantization loss parameter. In fact, if the actual σ is determined, there is a definite k and t to minimize the loss, so that a bag-of-words model corresponding to the σ can be obtained, and the optimal parameters can be selected to realize the processing of the input data.

3.2 Target Perception Image Compression Model

Model Building. Most of the existing deep learning SAR image compression is compressed and reconstructed through the encoding and decoding network, which often only focuses on the compression performance and indicators at the global level. So it is often difficult to achieve a higher degree of information fidelity for the local target of interest, which leads to the redundancy of non-interest region information. Therefore, by constructing the object perception map, the compression model of high

information retention for the target of interest and specific regions can be realized, which can further reduce the redundancy of non-interest region information in the image.

The model is a hyper-prior network model, mainly composed of a main encoding and decoding network and a hyper-prior encoding and decoding network. The quality mapping information is introduced to allocate the bit rate, achieving the goal of reducing the redundancy of the non-interest area information. As shown in Figure 2, the importance guide map is introduced into the compression model. The importance guide feature map is extracted by the ViT pre-trained network, and the pre-trained feature map is used to guide the bit rate allocation, the bit rate weight of the feature region is increased, and the importance guide map and the input data are fused with spatial features SFF [22]. By up sampling, down sampling and spatial fusion of feature maps and guide maps from different angles to achieve the purpose of deepening the importance of the guide map in the network, the image compression model of the overall quality map guidance control is realized by setting k equal divisions of the guide map weight parameters 0-1, randomly selected for training, and inferred according to the required fidelity of the target of concern during inference.

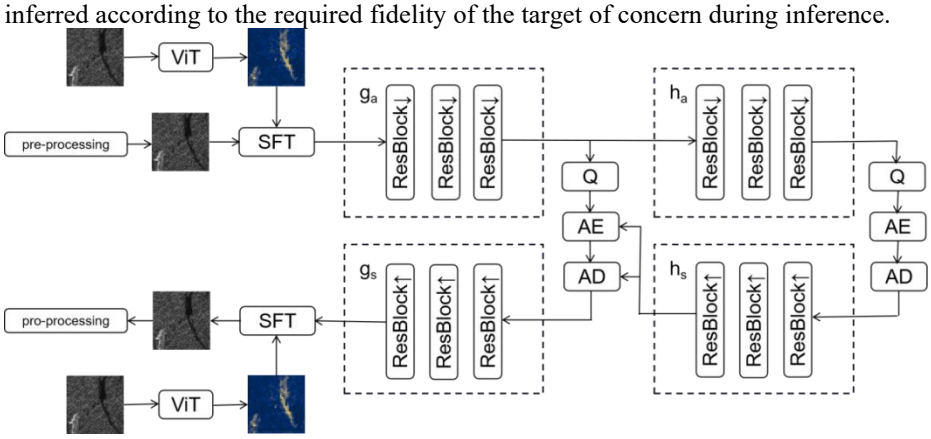

Fig.2. Target perception image compression network model

Construction of Multiplicative Loss Function for Target Perception. Under the inference of the variational model, the traditional loss function focuses on the trade-off of the code rate and distortion, and the loss function L is shown in Function 4. In our target perception model, the loss function focuses on the loss on the target with a higher priority, so the distortion D is modified and the weighted sum with the target-aware map is defined as:

$$D_{map} = f(D, map) = f(\frac{1}{n}\sum_{i=1}^{n}(x_i - x_i^{'})^2, map) = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_i^{'})^2 * (1 + map_i) \quad (18)$$

Therefore, the loss function L is improved as:

$$L = R + \lambda D_{map} \quad (19)$$

4 Experiments

This section mainly presents the experimental results of the proposed model.

4.1 Data and Evaluation Indicators

The main data used in this article is the SAR data of Sentinel-1, which covers marine terrain and land terrain, including details such as ships, houses, etc., and textures including roads and ocean ripples. Sentinel-1 images cover a wide range and high complexity, and are also one of the most classic SAR images. It can provide unpolarized complex data and quadrupolar zed complex data, and at the same time, it has a marine ship scenario, which can provide similar targets of interest, so it is of great value to study Sentinel-1 data.

The evaluation of the information retention ability of the compression reconstruction model is mainly carried out through the image evaluation index, such as peak signal-to-noise ratio (PSNR), multi-scale structural similarity (MS-SSIM) and bit rate (BPP). The main metric we used in this article is the peak signal-to-noise ratio (PSNR), which measures the distortion of the original and reconstructed images as:

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right) \tag{20}$$

where MAX is the maximum number of pixels in an image, and MSE is the MSE mean square error of the original image and the reconstructed image. The unit of PSNR is decibels (dB), and the larger the PSNR, the smaller the difference between the original image and the reconstructed image.

4.2 Low-frequency Suppression Algorithm Experiments

In the process of low-frequency suppression algorithm, the first step is the estimation of the data distribution parameter σ, so that each data is represented by a single σ parameter, and the fitting result is shown in Figure 3, and the σ=6.42 in the figure.

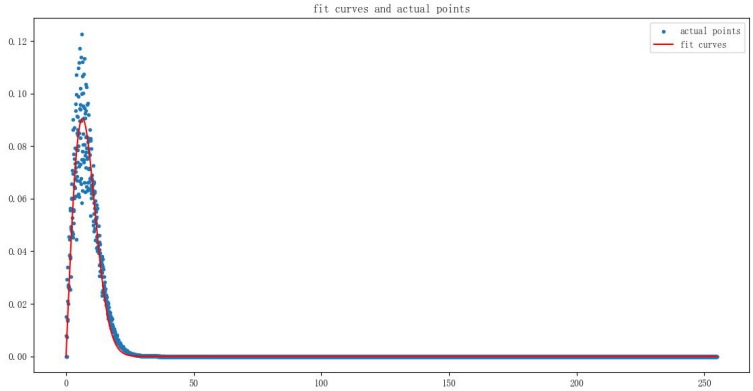

Fig.3. Fit curves and actual points where σ=6.42

In fact, when the parameters k and t are fitted, the 3.1 principle can be used to fit and solve $loss_{pro}(k, t)$ under a specific σ, and the results are shown in Figure 4, and the optimal parameters k and t under a given σ can be obtained. The corresponding t and k in the figure are 27 and 1e-8 respectively. In fact, under the condition that a specific t is reasonable, the PSNR loss for k is not much different in a relatively large range, so it can directly correspond to the given parameter σ according to the results of the sampled bag-of-word model and interpolation.

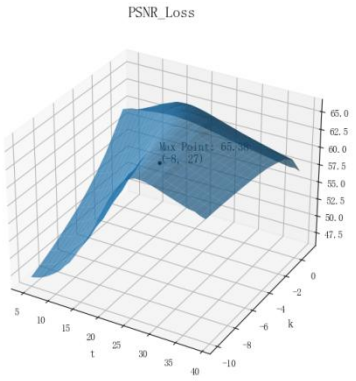

Fig.4. Evaluation index chart PSNR vs k, t where σ=6.42

The traditional processing methods include linear processing and power processing, and the experiment results of the PSNR indicators of the three treatments are compared in Table 1, and the proposed algorithm PSNR of 65.38 is the better algorithm.

Table 1. PSNR of the traditional linear method, traditional power method and proposed.

Algorithm	PSNR
Traditional linear method	42.93
Traditional power method	58.92
Proposed	65.38

4.3 Object Perception Compression Experiments

We selected a sub-image of the Sentinel-1 for experiments. Firstly, we processed the image by the low-frequency suppression algorithm. Then we extracted the target perception quality map and experimented the compression model. In the actual perception image extraction, we used the ViT preprocessing model to extract the multi-head attention feature map. The final extracted feature maps containing 12

heads are shown in Figure 5. Figure. 6 is the result of the traditional method and the low-frequency suppression algorithm feature map. The experimental results show that the proposed method can effectively retain and fuse the image features, to improve the image quality.

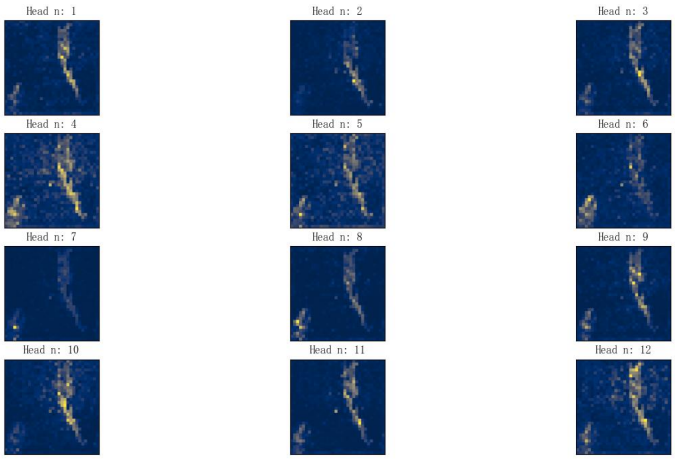

Fig.5. Multi-head attention feature map

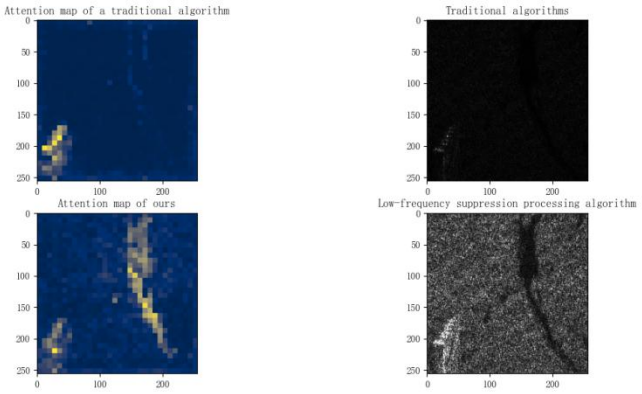

Fig.6. Feature map of traditional method and proposed

In our subsequent experiments, we utilized a hyper-prior model as backbone. And we performed spatial features fusion of the quality map and the input data. Throughout the training phase, we use the PyTorch framework and use the Adam optimizer to train the model. The batch size is set to 8 and the initial learning rate is 0.0001. The size of 256x256 pixels and 3000 epochs were applied on a NVIDIA GeForce RTX 3080 Ti GPUs. Table 2 shows the compression performance and experimental results

of the SAR model for low-frequency rejection. Experimental results show that when the BPP is about 0.35, the PSNR performance of the recommendation model is about 5db better than that of JPEG, and the proposed method can effectively retain and improve the performance of the image compression model.

Table 2. PSNR of Object Perception Compression Model.

Algorithm	BPP	PSNR
Traditional linear method	0.35	26.72
Proposed	0.34	31.35

5 Conclusion

Firstly, we construct a two-stage transformation operator for the input data to suppress the low-frequency of the input data, to achieve the data input with high peak signal-to-noise ratio and low quantization loss. Secondly, a compression model guided by the perceptual graph of the focus on the target is constructed to guide the allocation of the compression bitrate, and the redundancy between the target of interest and the non-target of interest is explored, to realize the image compression model with a higher degree of information fidelity for the local target of concern. Experimental results show that the proposed method can effectively reduce the loss of the input layer in the SAR image and improve the compression performance of the model. The significance of this study lies in the application of deep learning technology in SAR image processing. A two-stage low-frequency suppression algorithm and an image compression model for target perception are introduced, and their effectiveness is experimentally verified.

References

1. Wu, Z., Hou, B., Jiao, L.: Multiscale CNN with autoencoder regularization joint contextual attention network for SAR image classification. IEEE Transactions on Geoscience and Remote Sensing 59(2), 1200–1213 (2020).
2. DeGraaf, S.R.: Sar imaging via modern 2-D spectral estimation methods. IEEE Transactions on Image Processing 7(5), 729–761 (1998).
3. Pestel-Schiller, U., Ostermann, J.: Subjective evaluation of compressed SAR images using JPEG and HEVC intra coding: Sometimes, compression improves usability. In: 2018 15th European Radar Conference (EuRAD). pp. 154–157. IEEE (2018).
4. Ballé, J., Laparra, V., Simoncelli, E.P.: End-to-end optimized image compression. arXiv preprint arXiv:1611.01704 (2016).
5. Ballé, J., Minnen, D., Singh, S., Hwang, S.J., Johnston, N.: Variational image compression with a scale hyperprior. arXiv preprint arXiv:1802.01436 (2018).
6. Minnen, D., Ballé, J., Toderici, G.D.: Joint autoregressive and hierarchical priors for learned image compression. Advances in neural information processing systems 31 (2018).

7. Li, J., Liu, Z.: Multispectral transforms using convolution neural networks for remote sensing multispectral image compression. Remote Sensing 11(7), 759 (2019).
8. Xu, Q., Xiang, Y., Di, Z., Fan, Y., Feng, Q., Wu, Q., Shi, J.: Synthetic aperture radar image compression based on a variational autoencoder. IEEE Geoscience and Remote Sensing Letters 19, 1–5 (2021).
9. Zhang, L., Pan, T., Huang, Y., Qu, L., Liu, Y.: Sar image compression using discretized gaussian adaptive model and generalized subtractive normalization. IEEE Geoscience and Remote Sensing Letters 19, 1–5(2022).
10. Fu, C., Du, B., Zhang, L.: Sar Image Compression Based on Multi-Resblock and Global Context. IEEE Geoscience and Remote Sensing Letters 20, 1–5 (2023).
11. Fu, C., Du, B.: Remote Sensing Image Compression Based on the Multiple Prior Information. Remote Sensing 15(8), 2211 (2023).
12. Ross, T.D., Worrell, S.W., Velten, V.J., Mossing, J.C., Bryant, M.L.: Standard SAR ATR evaluation experiments using the MSTAR public release data set. In: Algorithms for Synthetic Aperture Radar Imagery V. vol. 3370, pp. 566–573. SPIE (1998).
13. Sun, Y., Li, L., Ding, Y., Bai, J., Xin, X.: Image Compression Algorithm Based on Variational Autoencoder. In: Journal of Physics: Conference Series. vol. 2066, p. 012008. IOP Publishing (2021).
14. Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., et al.: A survey on vision transformer. IEEE transactions on pattern analysis and machine intelligence 45(1), 87–110 (2022).
15. Wu, C.P., Kuo, C.C.J.: Efficient multimedia encryption via entropy codec design. In: Security and Watermarking of Multimedia Contents III. vol. 4314, pp. 128–138. SPIE (2001).
16. Murphy, K.P.: Machine learning: a probabilistic perspective. Cambridge, Massachusetts, USA: MIT press (2012).
17. McLachlan, G.J., Rathnayake, S.: On the number of components in a Gaussian mixture model. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 4(5), 341–355 (2014).
18. Ortega, A., Ramchandran, K.: Rate-distortion methods for image and video compression. IEEE Signal processing magazine 15(6), 23–50 (1998).
19. Berger, T.: Rate-distortion theory. Wiley Encyclopedia of Telecommunications (2003).
20. Sara, U., Akter, M., Uddin, M.S.: Image quality assessment through FSIM, SSIM, MSE and PSNR a comparative study. Journal of Computer and Communications 7(3), 8–18 (2019).
21. Hore, A., Ziou, D.: Image quality metrics: PSNR vs. SSIM. In: 2010 20th international conference on pattern recognition. pp. 2366–2369. IEEE (2010).
22. Liu, S., Huang, D., Wang, Y.: Learning spatial fusion for single-shot object detection. arXiv 2019. arXiv preprint arXiv:1911.09516 (1911).