Fixed-Quality Compression of Remote Sensing Images With Neural Networks

Sebastià Mijares I Verdú, Marie Chabert, Thomas Oberlin, Joan Serra-Sagristà

Fixed-quality compression of remote sensing images with neural networks

Sebastià Mijares i Verdú, Marie Chabert, *Member, IEEE*, Thomas Oberlin, *Member, IEEE*, and Joan Serra-Sagristà, *Senior Member, IEEE*

***Abstract*—Fixed-quality image compression is a coding paradigm where the tolerated introduced distortion is set by the user. This paper proposes a novel fixed-quality compression method for remote sensing images. It is based on a neural architecture we have recently proposed for multirate satellite image compression. In this paper, we show how to efficiently estimate the reconstruction quality using an appropriate statistical model. The performance of our approach is assessed and compared against recent fixed-quality coding techniques and standards in terms of accuracy and rate-distortion, as well as with recent machine learning compression methods in rate-distortion, showing competitive results. In particular, the proposed method does not introduce artefacts even when coding neighbouring areas at different qualities.**

***Index Terms*—Remote sensing, Data compression, Neural networks, Neural network applications, Optical data processing.**

I. Introduction

DOWNLINK capacity is one of the key bottlenecks for remote sensing missions to this day. Limited downlink capacity restricts how much data can be captured by these missions, and thus how often can images be sensed or how many spectral bands can be recorded [1]. In turn, given the limited lifespan of satellites, especially that of smaller NewSpace missions, there is great interest in obtaining as much data as possible in the duration of the mission [2]. Data compression allows for more data to be transmitted through highly demanded channels, and it will become more and more crucial in remote sensing.

Lossless compression is the first choice for any mission, but, to meet the high demands placed on compression systems, *lossy compression* is often considered in remote sensing [3], [4], [5], [6]. Lossy compression can greatly reduce data volumes, at the cost of having reconstructions not identical to the original data. Controlling the loss of information recorded by the mission is critical for the final users, and thus is a key barrier to the adoption of lossy codecs which operate by minimising a rate-distortion trade-off [7]. An intermediate solution to this problem is *near-lossless* compression, another paradigm in between lossless and lossy compression where some loss is allowed within a preset error bound [8]. Near-lossless compression is generally concerned with pixel-by-pixel error, and it may be generalised to other distortion metrics [8]. Beyond having prior control of the quality at which the image is compressed, another concern is the uniformity of quality across regions in the image. *Fixed-quality compression* aims to compress at a user-defined quality requirement measured locally across the image. To that end, it allocates more bits to more complex regions of the image [9]. Quality may be measured by any distortion metric. Since some metrics are calculated over a certain region -such as Mean Square Error (MSE)-, when talking about local quality, we may consider the measurement of that quality within some reasonably-sized region. Near-lossless compression can be also seen as fixed-quality compression under the metric of Peak Absolute Error (PAE). Given a maximum error, a near-lossless codec compresses the image with a bounded PAE on a pixel by pixel basis.

In this paper we are specifically interested in using MSE or Peak Signal-to-Noise Ratio (PSNR) as the metrics, measured across the entire image, or in blocks of pixels. In particular, to compress an image so that each $p \times q$-pixel region is reconstructed at a given MSE/PSNR. The region size $p \times q$ is user-given, and may include setting a single region for the entire image.

Fixed-quality compression in Earth Observation is an industry need that has been investigated in recent years, with proposals based on CCSDS standards. Camarero *et al.* [9] and Blanes *et al.* [10] proposed a rate-allocation method for the wavelet-based CCSDS 122.0-B-1 to estimate the bitrate at which to encode block segments of the image so that the resulting segments are recovered at a desired quality, measured by the MSE.

Meanwhile, the most recent breakthrough in image compression has been the introduction of neural networks, particularly in the last 6 years [11], [12], [13], [14], [15]. Machine learning (ML)-based codecs have widely surpassed established conventional methods such as JPEG 2000 in lossy compression of natural images [12], [13], [16], [14], [17] and remote sensing data [18], [19], [20], [21], [22], [23], [24], [25], and

S. Mijares and J. Serra-Sagristà are with the Department of Information and Communications Engineering, Universitat Autònoma de Barcelona, Cerdanyola del Vallès 08193 Spain.

T. Oberlin is with ISAE-SUPAERO, Université de Toulouse, 31055 Toulouse, France.

Marie Chabert is with IRIT/INP-ENSEEIHT, Université de Toulouse, 31071 Toulouse, France.

Corresponding author is Sebastià Mijares i Verdú (sebastia.mijares@uab.cat).

This work was supported in part by the Institute for Artificial and Natural Intelligence Toulouse (ANITI) under grant agreement ANR-19-PI3A-0004; also by the Spanish Ministry of Science and Innovation (MICINN) and by the European Regional Development Fund (FEDER), funded by MCIN/AEI/10.13039/501100011033/FEDER, UE, under grants PID2021-125258OB-I00 and PRE2019-088824; also by the Catalan Government under grant SGR2021-00643; and also by the "Data Compression and Machine Learning for Earth Observation Satellites" project under the Institute for Space Studies of Catalonia (IEEC) NewSpace 2024 initiative.

Manuscript received March 2nd, 2024; revised June 19th, 2024.

are the state of the art in those fields. The most successful methods, introduced in 2017, rely on autoencoders [11], [26]. These methods jointly optimise neural network transforms, made of an encoder and a decoder, with respect to rate and distortion [15]. This basic setup can be further expanded by introducing a side network called the *hyperprior* that extracts context information from the latent representation to improve coding performance [12], [13]. Following this paradigm, state-of-the-art architectures have grown increasingly complex, using -among other developments- residual blocks instead of plain convolutional layers, and more sophisticated prior distributions and contexts for the arithmetic coder [14], [17].

Some key challenges to the practical adoption of ML-based codecs in remote sensing are computational complexity and variable-rate compression. Both have been addressed by recent proposals that surpass current standards such as JPEG 2000 and CCSDS-122 in lossy compression [20], [24], [25], and some methods have already been implemented for use in space such as in the Φ-sat-2 mission [27]. To the best of our knowledge, however, no proposals have been made on using neural networks for fixed-quality compression.

In this paper, a novel method for fixed-quality compression using neural codecs is proposed. Our method is based on reduced-complexity architectures proposed for onboard fixed-rate lossy compression and is, to the best of our knowledge, the first fixed-quality compression method using neural networks. We show that the proposed method can accurately compress at the desired quality, at rates highly competitive with standards currently in use in particular with methods with which fixed-quality compression has been carried out. Moreover, the desired quality can be achieved for multiple-sized regions of the image.

The rest of the paper is organised as follows: Section II discusses compression standards and methods to implement fixed-quality compression, and proposed ML architectures for onboard data compression. In Section III the proposed method is described, and its experimental merits are presented in Section IV. Finally, Section V concludes the paper with a discussion of our findings.

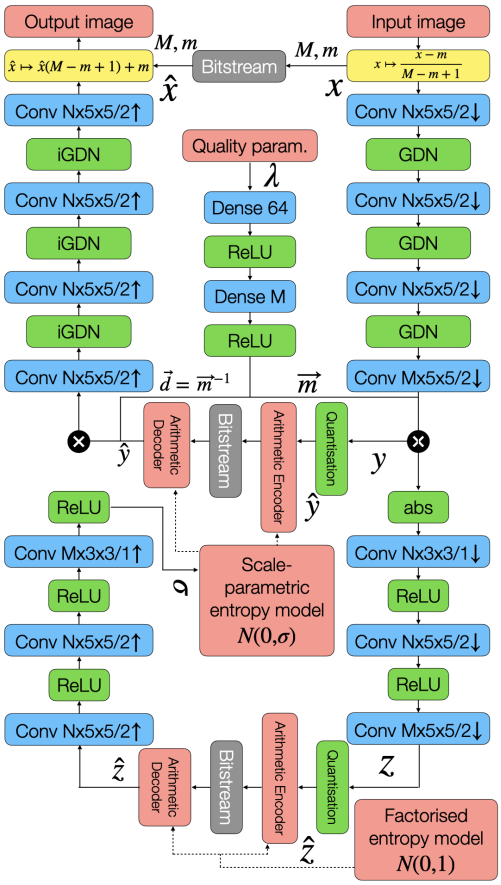


Fig. (1) Architecture used for the proposed method, based on the proposal from [24].

II. Related work

This contribution follows on two distinct fields of research: fixed-quality compression for remote sensing data, and the usage of neural networks in image compression, with a particular focus on lightweight architectures for onboard application in both cases. Subsection II-A describes the current standards used in remote sensing lossy and near-lossless compression, and how these have been used for fixed-quality compression. Subsection II-B presents image compression using neural networks, application to remote sensing data and onboard considerations.

A. Onboard compression standards and fixed-quality compression

There are three main standards in lossy and near-lossless remote sensing image compression: JPEG 2000 [28], CCSDS 122.1-B-2 [29], and CCSDS 123.0-B-2 [30]. The first is a lossy –and lossless– codec standard based on the Discrete Wavelet Transform (DWT). The CCSDS 122.1-B-2 is another DWT-based lossy (and lossless) codec, similar to JPEG 2000, that introduces some modifications in order to reduce operational complexity. This standard is also in use in several ESA missions in implementations such as ESA's own CCSDS Wavelet Image COMpression (CWICOM) [31]. Attending to the complexity limitations, reduced-complexity DWT-based codecs are used in missions such as ESA's Copernicus Sentinel 2 and CNES's Pléiades. Finally, the CCSDS 123.0-B-2 standard describes a predictor-based codec that allows for lossless and near-lossless compression, distributed for remote-sensing missions under ESA's software SHyLoC [32].

As described in the introduction, there are two fixed-quality compression methods [9], [10] based on the CCSDS 122 standard. Both contributions estimate the rate or bit plane stopping point for the encoding process such that the reconstruction achieves a given quality. We thus compare the rate-distortion performance of our method against CCSDS 122 and JPEG 2000 as the reference standards. Observe that the rate-distortion performance of CCSDS 122 in fixed-quality mode should be similar or inferior to that of CCSDS 122 used in conventional mode. Similar to these contributions and to our own but in the realm of lossless compression, a variety of models have been developed to predict the rate of compression of remote sensing data achieved by a given method [33].

(a) Original (left), reconstruction (centre), and difference (right).

(b) Zoom-in of the original (left), reconstruction (centre), and difference (right).

Fig. (2) Simulated Pléiades image compressed using minimum quality on the left half and maximum quality on the right half, resulting in 1.2 bps compression and an overall quality of 38.1 dB PSNR (© CNES 2024).

B. Image compression using neural networks

Contributions by Ballé *et al.* introducing autoencoders [11], and later Variational Autoencoders (VAE) with a hyperprior network [12] have been a breakthrough in lossy image compression, and similar techniques have since become the state of the art in the field. Autoencoders used in compression are neural networks with two main parts: an *encoder* and a *decoder*. The encoder (more specifically the *analysis transform*) maps the input image into a *latent representation*, which is quantised and entropy-coded as the compressed image. To perform this entropy coding, most proposals use a hyperprior, which is a secondary network (also an autoencoder) which takes the latent representation as input to produce side information which can be decoded to obtain context information parameters for the aritmetic coding of the latent representation. The decoder performs the opposite operation to that of the encoder, applying a *synthesis transform* to the latent representation to produce the reconstructed image. These neural networks are all jointly trained to minimise a rate-distortion trade-off loss function:

$$L = R(\tilde{y}) + \lambda D(x, \hat{x}), \tag{1}$$

where $R(\tilde{y})$ is the compression rate of the latent representation y with some differentiable substitute for quantisation (adding uniform noise, soft quantisation, or one of other variants) [15], $D(x, \hat{x})$ is the distortion between the original image, x, and the reconstruction $\hat{x}$, and λ is a constant regulating the rate-distortion trade-off.

Neural network codecs have generally improved on the state of the art performance for natural image compression at the cost of an increasing computational complexity [34]. This has been achieved by introducing elements to the overall VAE framework detailed above such as residual blocks in the main transforms [14], more sophisticated entropy models to more efficiently encode the latent representations [13], [14], [35], and a variety of attention modules and mechanisms [36], [37]. While this may be computationally viable for regular compression applications on land, especially if GPUs can be used, this is a significant limitation if these compression techniques were to be deployed in low-power remote sensing platforms such as satellites. To that end, recent works have proposed reduced-complexity variants of these methods for remote sensing image compression, with complexities viable to be used on board [20], [24].

C. Modulated networks for variable bitrate

Classically, in the rate-distortion trade-off (1), the λ parameter is a constant value set before training. The model is thus optimised for a specific trade-off that leads to different rates for different input images. *Modulation* has been successfully applied to these architectures to make them capable of compression at multiple rates [38], [16], [24]. In general terms, it consists in a secondary network (the modulation network), which maps a parameter (in this case, λ) into an array of weights that are then used to scale the outputs of one or more layers in order to produce a different output. In the realm of multirate compression specifically, it has been found that modulating only in the latent layer can achieve practically the same rate-distortion results as the equivalent fixed-rate models, even in reduced-complexity architectures [38], [24]. Our proposed method uses the reduced-complexity multirate codec from [24], depicted in Figure 1. In the original architecture, variable rate is achieved by modulation: a lightweight side network maps the λ parameter into a scaling or modulation mask $\mathcal{M}$ with same size as the latent representation. Each pixel of $\mathcal{M}$ then multiplies the respective pixel of the latent representation before quantisation. In the decoder, the dequantised latent representation is pixel-wise divided by this mask. The scaling or modulation allows to adapt the quantisation step to reach a given rate-distortion trade-off using a neural network trained for another trade-off without training it again.

In [24], modulation has been used to achieve fixed-rate compression: each image is compressed at a user-defined bit rate, using a specific λ parameter obtained by binary search. However, this binary search strategy is impractical to achieve fixed-quality compression. Indeed, at each step of the binary search the reconstructed image quality would have to be derived which would require to apply the decoder each time. This is far too computationally costly to be done in practice.

III. Proposed method

As introduced at the end of the previous section, the proposed fixed-quality compression method is based on the modulated architecture from [24]. We note that, by using different scaling factors for different areas in the image, these areas can be compressed at different qualities to obtain a predefined rate or, alternatively, at different rates to target a predefined quality. This is the main principle underlying the proposed method. Figure 2 illustrates the impact of the scaling factor on the image quality: the usage of different scaling factors in different areas of the image leads to different reconstruction qualities, both visually and quantitatively in terms of MSE.

The parameter λ allows to reach a rate-distortion trade-off on the whole training set. This parameter does not guarantee a given rate and a given distortion for each input image individually. For a given image, the obtained rate and distortion will depend on its characteristics. Note also that the synthesis transform is not the exact inverse of the analysis transform. The reconstruction error measured by MSE has a minimum value, denoted as $\mathrm{MSE}_0(x)$, for each image x which varies significantly from one image to another. Nonetheless, we have identified linear and quadratic relations between the parameter λ and the reconstruction error. These relations can be associated to the proposed architecture, considering the number of layers and non-linearities used to derive the main transforms and the modulation vector. In this paper, these relations are obtained empirically.

On the one hand, the relation between the modulation mask magnitude, $\bar{\mathcal{M}} = \frac{\|\mathcal{M}\|_1}{n}$, and λ is essentially linear for various models, as shown in Figure 3. Note this does not mean that every individual pixel of the mask, $\mathcal{M}_i$, varies linearly with λ. On the other hand, for any given image, the reconstruction MSE increases quadratically with the inverse of $\bar{\mathcal{M}}$ (and thus with the quantisation step size), as also shown in Figure 3.

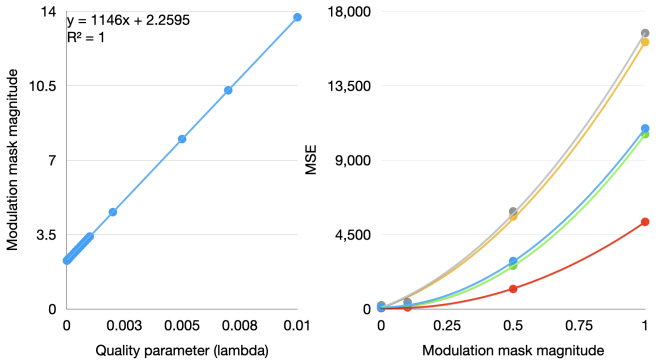


Fig. (3) Relationship between the parameter λ and the modulation mask magnitude $\bar{\mathcal{M}}$ in the Pléiades model (left) and between the modulation mask magnitude $\bar{\mathcal{M}}$ and the reconstruction MSE by our models on several Pléiades images individually, where each curve corresponds to a different image (right).

This relation is, furthermore, proportional to the baseline reconstruction loss, $\mathrm{MSE}_0(x)$, for a given image. As a result, we can propose the following model for the MSE of a given image x as a function of parameter λ:

$$\hat{\mathrm{MSE}} = \mathrm{MSE}_0 + \alpha\,\mathrm{MSE}_0\,\frac{1}{\left(2\bar{\mathcal{M}}(\lambda)\right)^2}, \tag{2}$$

where α is derived by regression from a subset of the training set. Similarly, due to the aforementioned linear relation between the parameter λ and the magnitude of the modulation mask $\bar{\mathcal{M}}$, we can find the regression coefficients a, b such that:

$$\bar{\mathcal{M}}(\lambda) = a\lambda + b. \tag{3}$$

Combining (2) and (3), we can produce the following function to obtain what parameter $\hat{\lambda}$ to use to encode an image at a target MSE, denoted as $\hat{\mathrm{MSE}}$:

$$\hat{\lambda} = \frac{1}{a}\left(\sqrt{\frac{\alpha\,\mathrm{MSE}_0(x)}{4\left(\hat{\mathrm{MSE}} - \mathrm{MSE}_0(x)\right)}} - b\right). \tag{4}$$

Derived from (4), the outline of the proposed method is as follows: using a model as those from [24] trained for the intended data source, a λ can be estimated to compress the image at the target MSE denoted $\hat{\mathrm{MSE}}$. Furthermore, this λ

can be adjusted, not only globally, but regionally, in blocks of down to 16×16 pixels to target different MSEs on these blocks.

The array of λ produced is an array of 32-bit floating point values, 256 times smaller than the original image (spatially), and must be transmitted as side information to the decoder for reconstruction. This side information corresponds to an additional 0.125 bits per sample (bps) if raw transmitted. To reduce this side information, the proposed method includes quantisation of the λ's into 8-bit integer values using

$$Q = \left\lfloor 255 \frac{\lambda - \lambda_{\min}}{\lambda_{max} - \lambda_{\min}} \right\rceil, \tag{5}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the minimum and maximum values of λ used in the training of the model. Quantisation in Equation (5) can be reversed using

$$\hat{\lambda} = \frac{(\lambda_{\max} - \lambda_{\min})\, Q}{255} + \lambda_{\min}. \tag{6}$$

Using this quantisation of the λ array, the volume of the side information required is now only 0.03125 bps.

IV. Experimental results

The proposed method has different aspects to be assessed: (i) whether it is competitive with fixed-quality compression and lossy compression standards and other ML methods in terms of rate-distortion performance; whether the proposed fixed-quality method is accurate in compressing at (ii) a global target quality and at (iii) a local quality with various-sized blocks; and finally (iv) whether the proposed method introduces any artefacts upon visual inspection of the images. To evaluate these different aspects, four data sets of various characteristics are selected; they are described in section IV-A. We have used several test models –trained for the previous datasets on quality accuracy–, listed in section IV-B. Rate-distortion performance is evaluated in section IV-C, where we also evaluate our method on an additional dataset for comparison with recent ML remote sensing data compression techniques. Some of the reconstructed images are visually inspected to compare the local with the global fixed-quality compression in section IV-D. Our implementation is available in an open GitHub repository https://github.com/smijares/mcos2024/.

A. Data sets and models

The general validity of the proposed method is assessed using four different remote sensing data sets:

1) 12-bit simulated panchromatic Pléiades images of 50 cm resolution. A total of 96 images are used in training and 32 images in testing, all with size 820×820. For this dataset, we use $\lambda \in [10^{-5}, 10^{-3}]$ in the model's training.
2) 16-bit L1-processed Landsat 8 OLI images with 30 m spatial resolution, taking frame by frame 7 non-panchromatic bands [39]. A total of 3,584 images are used in training and 1,280 in testing, all with size 512×512. For this dataset, we use $\lambda \in [10^{-6}, 5 \cdot 10^{-4}]$ in the model's training.
3) 16-bit AVIRIS calibrated scenes, taking frame by frame all the 224 spectral bands [40]. The images are of 30 m resolution. A total of 180 scenes are used for training and 20 scenes for testing, with size 512×512. For this dataset, we use $\lambda \in [10^{-5}, 5 \cdot 10^{-3}]$ in the model's training.
4) 16-bit AVIRIS uncalibrated scenes, taking frame by frame all the 224 spectral bands [41]. The images are of 30 m resolution. A total of 5 scenes are used for training and testing with size 512×512. For this dataset, we use $\lambda \in [10^{-6}, 10^{-4}]$ in the model's training.
5) 8-bit RGB images of the DOTA v1.0 set [42]. This data set is only used to compare the rate-distortion performance of our model to state-of-the-art ML compression methods. The set contains 1,403 images for training and 935 images for testing, all of varying sizes.

These data sets are the same as those used in [24] as well as a small set of AVIRIS uncalibrated data to show the proposed method still holds for data before calibration and the DOTA v1.0 dataset for reference with recent proposed methods in rate-distortion. A model is trained for each of the data sets using $N = 64$ and $M = 192$ as the numbers of filters, as well as an additional model using $N = 128$ and $M = 384$ for higher bitrates for the DOTA v1.0 data set. The resulting regression coefficients for fixed-quality compression are as in Table I, calculated as described in Section III.

TABLE (I) Regression coefficients for fixed-quality compression.

Data set	α	**a**	**b**
Pléiades	23.39	2,620.50	0.78
Landsat	638.84	16,956.00	11.50
AVIRIS calibrated	35.91	1,146.00	2.26
AVIRIS uncalibrated	206.61	62,764.60	2.37

B. Quality accuracy

To assess the accuracy of the proposed models in terms of fixed-quality compression, blocks of 64×64, 32×32, and 16×16 pixels are used, as well as the global prediction results. For each data set, a target MSE is set for all images. Every image is transformed, the MSE_0 of each block is measured, finding the necessary $\hat{\lambda}$ using Equation (4) for each block, and the image is compressed using the resulting array of $\hat{\lambda}$. The image is then decompressed, and the full MSE, as well as that for each block, is measured.

For the Pléiades data set the target quality was set to 40 dB PSNR, or 1,677.7216 MSE. For the Landsat 8 data set the target quality was set to 62 dB PSNR, or 2,709.5 MSE. For the AVIRIS data sets, the target quality was different for each of the bands, as, due to the large variability among them, there was no viable common MSE. Instead, for each band, the target MSE was set to $1.3 \times \mathrm{MSE}_0$ of that band for calibrated data, and to $1.5 \times \mathrm{MSE}_0$ of that band for uncalibrated data. Note that the target MSE was therefore uniform for all blocks in any given band.

Figure 4 shows the histogram of differences between the target PSNR and the one obtained in reconstruction by the

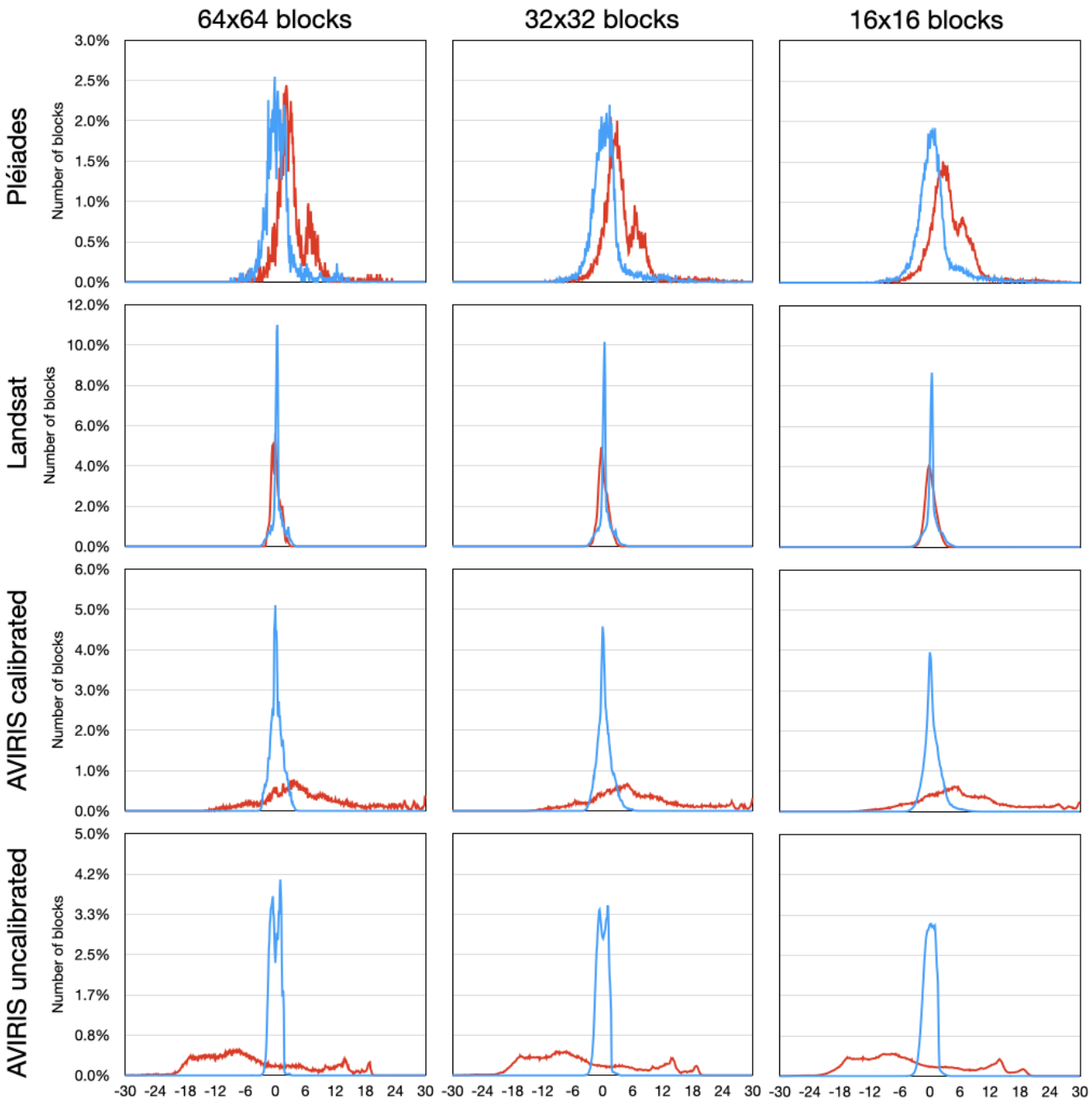
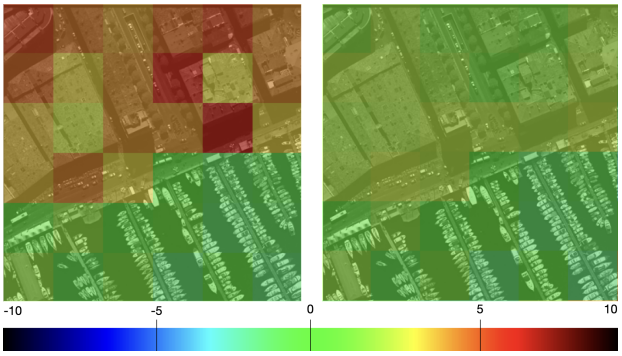


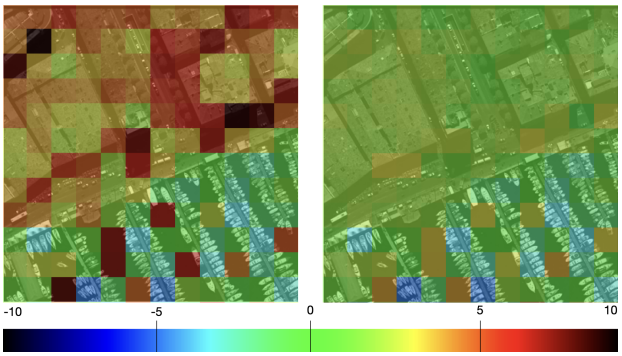
Fig. (4) Differences between the target PSNR and the obtained PSNR for each block in test images when using local fixed-quality compression (blue) or global fixed-quality compression (red).

models at varying prediction block sizes for each of the data sets. Bin sizes for those histograms are 0.1 dB, so we have bins $(-0.05\ \mathrm{dB}, 0.05\ \mathrm{dB}]$, $(0.05\ \mathrm{dB}, 0.15\ \mathrm{dB}]$, etc.. For reference, those also show the difference in PSNR across the same block sizes when the images are compressed using a global PSNR target rather than a local block-by-block target. It is clear from these histograms that, in comparison with using a global fixed-quality, the proposed local fixed-quality method recovers the local blocks accurately at the target PSNR, with differences with the target being greatly concentrated around 0 dB at all block sizes.

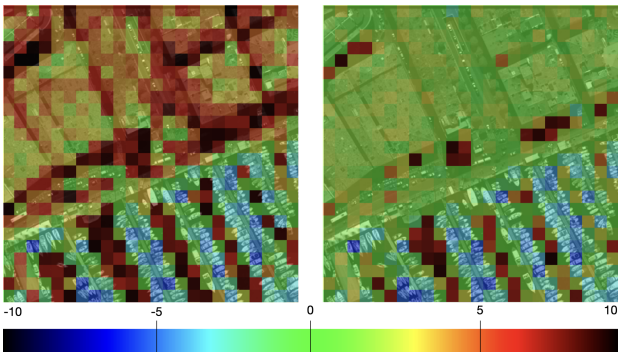
Table II shows the mean block-wise reconstruction target error by block size and data set, and the standard deviation thereof. The proposed method is less accurate as block size decreases, and the differences between target and actual reconstruction quality are also more spread out as the block size decreases. This is as could be expected: the statistical model makes less accurate predictions when given a smaller context, as outlier features can be more dominant (i.e., take up more pixels relative to the size of that context). The offset of our predictions being slightly conservative in PSNR can be attributed to the relation between MSE and PSNR,



(a) 64×64 block PSNR errors comparison.



(b) 32×32 block PSNR errors comparison.

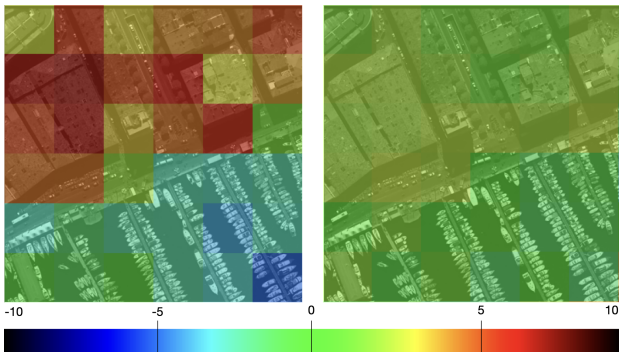


(c) 16×16 block PSNR errors comparison.

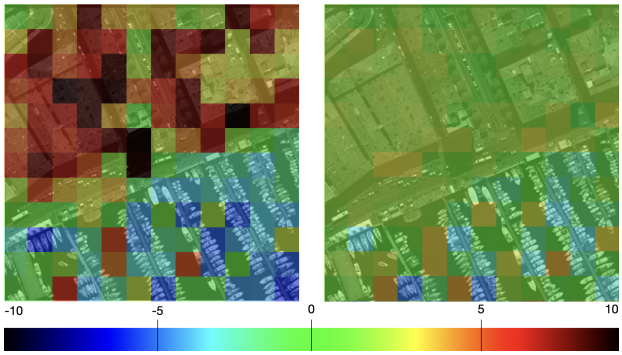
Fig. (5) Differences between the obtained PSNR and the target PSNR (35.71 dB) using fixed global quality (left) and fixed local quality (right) for each block in an example Pléiades image. Note that a positive error means the reconstruction quality was above the set target (© CNES 2024).

since $\mathrm{PSNR}(\mathrm{MSE}+\varepsilon) - \mathrm{PSNR}(\mathrm{MSE}) > \mathrm{PSNR}(\mathrm{MSE}) - \mathrm{PSNR}(\mathrm{MSE}-\varepsilon)$ for a fixed deviation $\varepsilon > 0$.

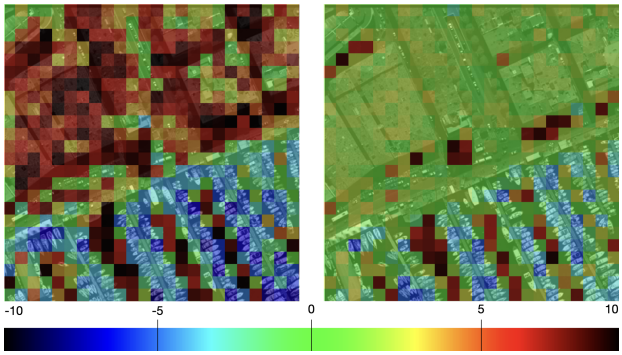
The AVIRIS uncalibrated scenes present a clearly different distribution from the other data sets, attributed to the usage of a much smaller number of scenes in both training and testing. The distribution is clearly bimodal, indicating there are two "main" classes of images (and blocks), for which our prediction coefficients take an average, thus underpredicting in one class, and overpredicting in the other. Nonetheless, we may observe how, in numerical terms, our predictions are more accurate for this data set than for the others in both mean of the prediction and standard deviation of the errors distribution.



(a) 64×64 block PSNR errors comparison.



(b) 32×32 block PSNR errors comparison.



(c) 16×16 block PSNR errors comparison.

Fig. (6) Differences between the obtained PSNR and the global/target PSNR using the Cheng 2020 architecture [14] (left) and our proposed fixed local quality (right) for each block in an example Pléiades image. Since the Cheng 2020 architecture does not allow to adjust the compression rate or quality continuously, the target PSNR is set to the overall PSNR obtained for the image. Note that a positive error means the reconstruction quality was above the set target (© CNES 2024).

TABLE (II) Mean error in PSNR quality prediction by block size. Average and standard deviation.

Block size	**Pléiades**	**Landsat**	**AVIRIS**	**AVIRIS uncal.**
Global	0.11; 1.44	-0.21; 0.69	-0.10; 0.56	0.03; 0.76
64×64	0.51; 2.77	0.42; 0.99	0.26; 1.23	-0.01; 0.96
32×32	0.63; 3.17	0.43; 1.10	0.38; 1.41	-0.03; 1.01
16×16	0.81; 3.72	0.43; 1.19	0.50; 1.67	-0.04; 1.09

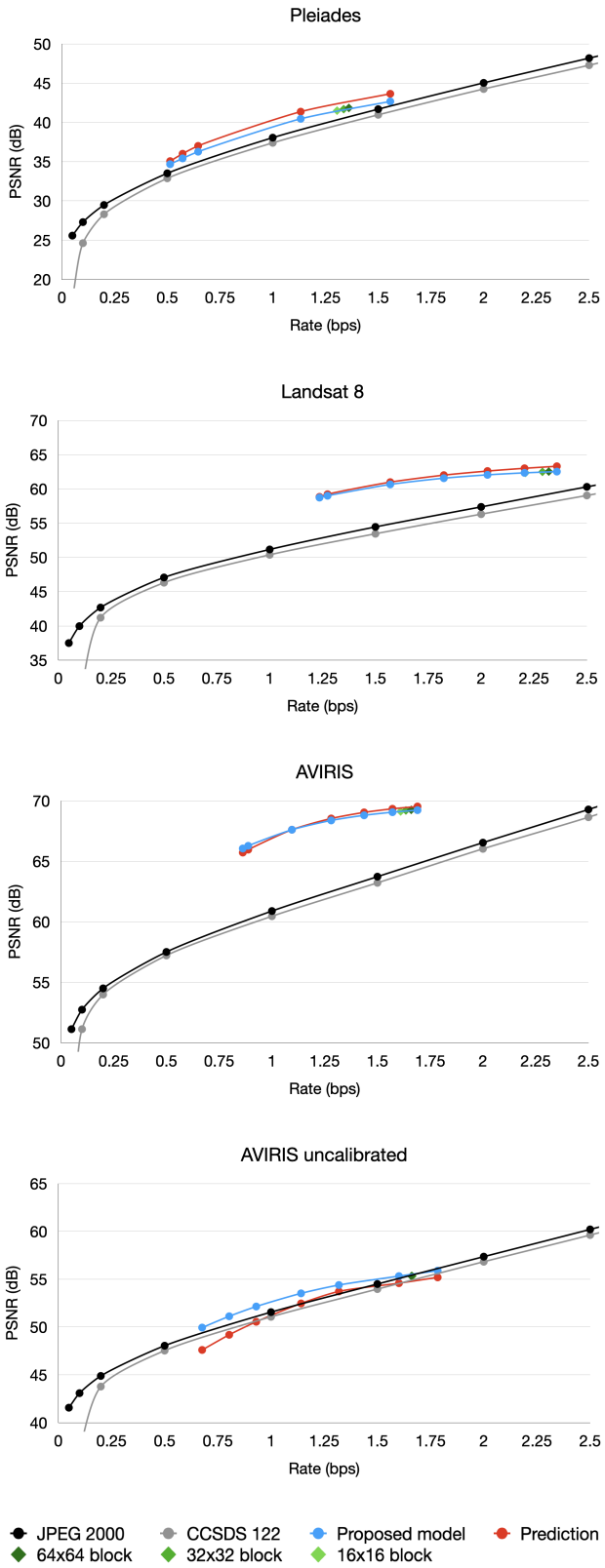


Fig. (7) Rate-distortion performance of our models and reference standards on Pléiades, Landsat 8 OLI, and AVIRIS (calibrated and uncalibrated) datasets.

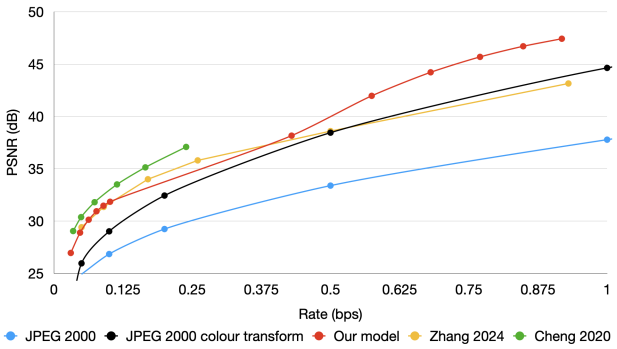


Fig. (8) Rate-distortion performance of our model, the reference standards, and the Zhang *et al.* 2024 [25] and the Cheng *et al.* 2020 [14] models.

Figure 5 shows the block-wise prediction error in a specific example image when compressed with a fixed global quality and when compressed using local quality at different block sizes. Clearly, the target qualities were far more accurate when using localised quality as opposed to using a flat global quality, as was indeed shown in general in the previous histograms. Figure 6 shows a similar block-wise prediction error to Figure 5, where our fixed-quality compression method is compared to the Cheng 2020 architecture [14]. Again, it is shown how compressing at a local fixed-quality achieves homogeneous MSE results than the alternative.

C. Rate-distortion performance

Figure 7 shows the rate-distortion performance of our proposed models in comparison with JPEG 2000 and CCSDS 122, as well as the expected PSNR curves of our models at the different λ used. As expected from [24], our models clearly surpass both standards in this regard. Furthermore, our PSNR predictions are less than 1 dB off from the actual quality obtained on average for all data sets except for the AVIRIS uncalibrated data set at some lower bitrates. This difference can be attributed due to the small data set available, which is then more susceptible to specific differences between images taking great effect.

Rate results in fixed-quality compression achieved by our models are identical to those obtained in the regular rate-distortion curve, as also shown on Figure 7. Having no downside in the rate results by varying the size of prediction blocks, the trade-off in block-size choice is in MSE accuracy: as discussed in Subsection IV-B, using larger blocks makes each block individually more accurate. However, using smaller blocks one can expect the reconstruction to allocate higher quality to smaller ”complex” regions, and the overall PSNR (as opposed to the average of block-wise PSNR) to be closer to the target PSNR.

As a comparison with recent published ML methods, we train and evaluate our model for the DOTA v1.0 dataset [42] and compare with a recent state-of-the-art codec for remote sensing images using results published for said data [25] and with a widely referenced and publically avaliable state-of-the-art model [14], [43]. Results, depicted in Figure 8 show our

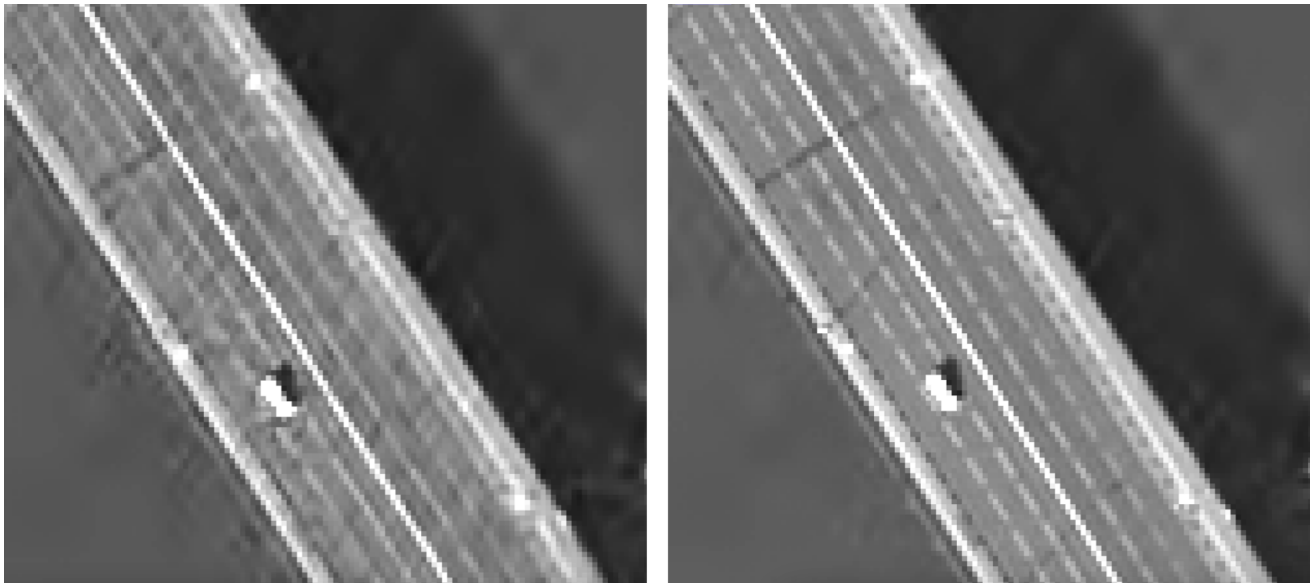


Fig. (9) Zoom-in of a Pléiades image compressed at a global fixed-quality (left) and at a local 16×16-block fixed-quality (right) with a compression target PSNR of 37.47 dB (© CNES 2024).

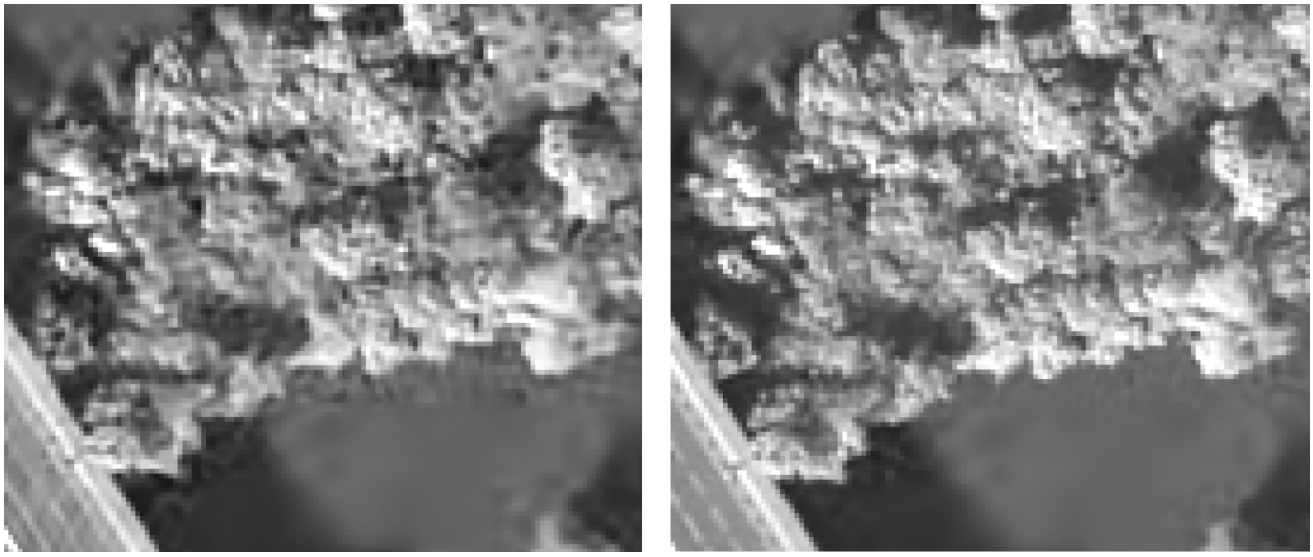


Fig. (10) Zoom-in of a Pléiades image compressed at a global fixed-quality (left) and at a local 16×16-block fixed-quality (right) with a compression target PSNR of 37.47 dB. Original is (© CNES 2024).

model matches the published performance of the Zhang 2024 architecture at lower rates and widely surpasses their proposal at rates approaching 1 bps, while the Cheng 2020 architecture surpasses both at the available rates. Furthermore, it must be noted that both comparison methods are subtantially more computationally complex than our own, and neither allow continuous adaptation of bitrates as our own proposal.

D. Visual inspection

A central aim of fixed-quality compression from the final users' perspective is to allocate more bits to more "complex" regions of the image, so that the reconstruction of said regions is more accurate. To assess whether some of those more complex regions are indeed being better recovered when using local MSE prediction as opposed to global MSE prediction, an image is compressed using (a) a global MSE target and using (b) the same MSE target in 16×16 blocks. Figure 9 shows an example of how local prediction can better recover some of the features in the images: the original image contains a road bridge across a large body of water that encompasses the full image. That contrast between the water and land features of the image is exploited by local fixed-quality compression to produce a far better reconstruction of the image: compared to global fixed-quality compression, the bridge edges are far less noisy and the elements in the road (the dashed line, most notably) are well preserved only when using local fixed-quality compression at this rate/quality. Figure 10 shows another section of this scene involving vegetation at the water's edge. The global fixed-quality reconstruction produces visible noise artefacts on the water areas at the edge of the vegetation, which

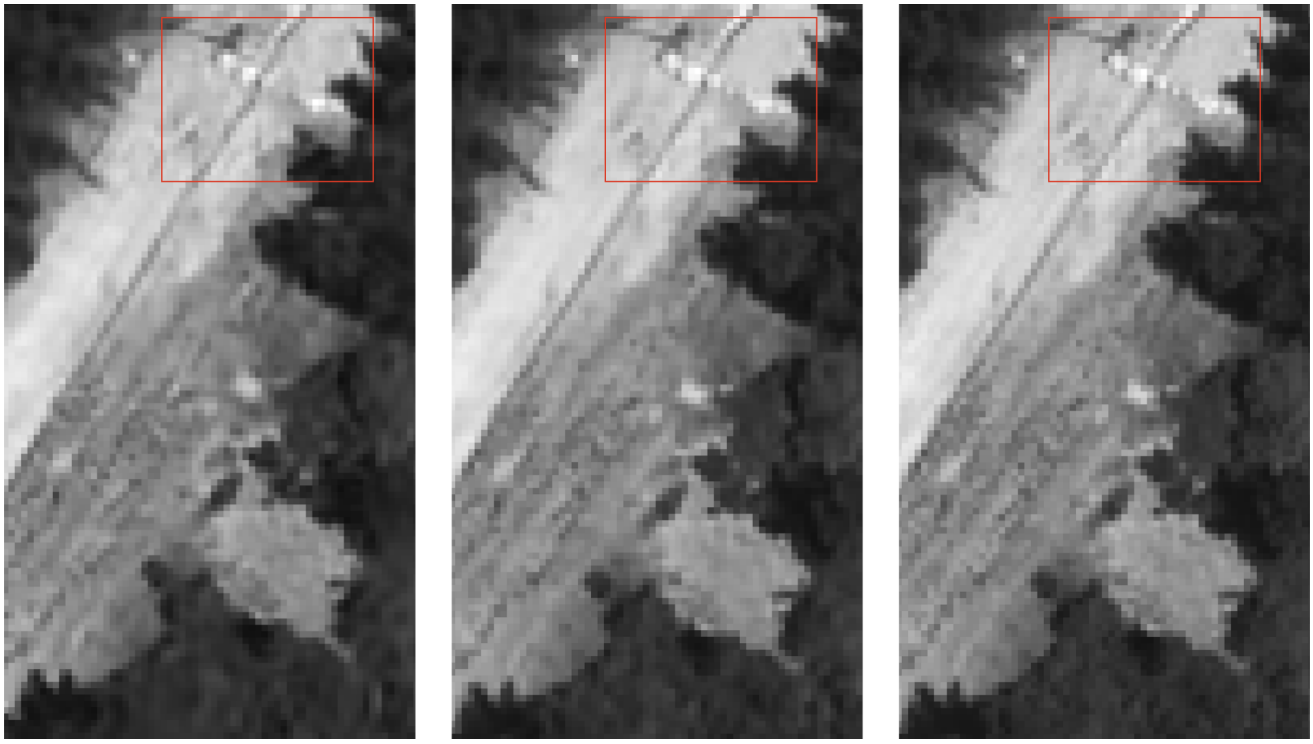


Fig. (11) Zoom-in of a Pléiades image compressed at a global fixed-quality (left) and at a local 16×16-block fixed-quality (middle) with a compression target PSNR of 37.47 dB, with the original image (right) for reference (© CNES 2024).

Fig. (12) Visual example of decompressed image and specific zoom area presented in [25]. This image was compressed at 0.096 bps (0.28 bpp) obtaining a reconstruction quality of 29.65 dB PSNR.

local fixed-quality compression does not produce.

Figure 11 provides yet another example, there of a forested area in France. The contrast between the vegetation regions of the image and the bright clear ones of roads, paths and agricultural fields give local fixed-quality compression an advantage over global fixed-quality compression. In particular, consider the elements by the path at the top of the image, highlighted in a red box. While the local fixed-quality reconstruction recovers them all, the global fixed-quality reconstruction obscures the identification of some of these elements. These small element-wise differences are of key importance for a high-precision mission such as Pléiades whose purpose is precisely to capture features at this scale for civilian and security purposes.

For completion, we also make a visual inspection of the results obtained by our models in the DOTA set using the same image and detail region as in [25]. Figure 12 shows an image decompressed using our model at 0.096 bps (or 0.28 bpp), for which the obtained quality was 29.65 dB PSNR. Results for the specific example image, both quantitatively and visually, are similar to those presented for the model in [25], shown in Figure 13, and in line with the overall rate-distortion results shown in Figure 8.

V. Conclusion

A fixed-quality compression method for remote sensing data using ML is proposed, in which the reconstruction quality is estimated from the contents of the image before compression. This proposed method is shown to be generally accurate in compression at a target MSE, set globally or locally on various-sized regions when evaluated on several sources of remote sensing data, including uncalibrated data as is captured by the sensor. A trade-off exists between the target region's size and fixed-quality accuracy: the smaller the region, the lower the accuracy in predicting reconstruction quality. The rate-distortion performance of the proposed fixed-quality method is identical to that of using a homogeneous λ for the whole image, thus no trade-off in compression ratio needs to be considered to use this method. The proposed fixed-quality compression method is particularly beneficial in images where high-frequency features or textures are on top of low-frequency backgrounds. For a given compression rate, by allocating different qualities to those contrasting elements, the proposed method allows a better reconstruction of the high-frequency ones. Furthermore, the method is shown to produce smooth reconstructions, without artefacts appearing

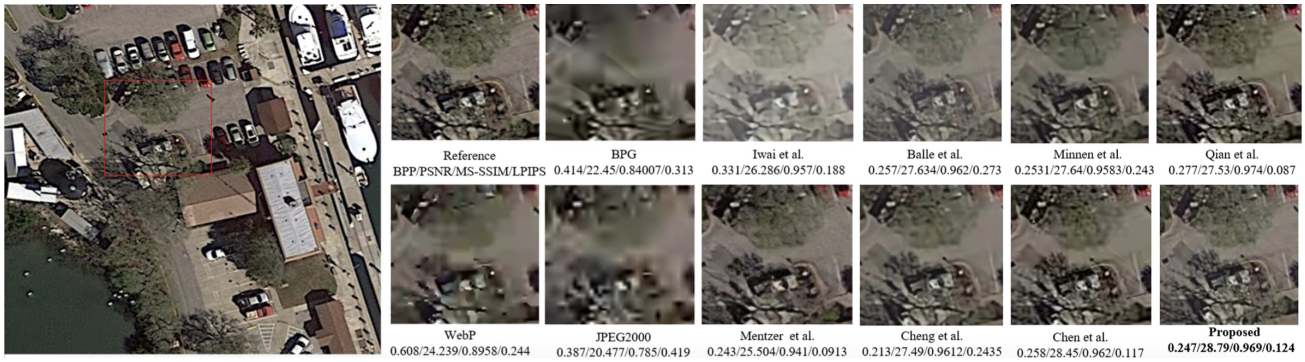


Fig. (13) Visual example of decompressed image and specific zoom area as presented in [25]. Values in each subfigure correspond to the referenced bpp, PSNR, MS-SSIM, LPIPS of compression with each method.

between regions compressed at different qualities, even when there is high contrast between those qualities. The method does increase the computational complexity of the encoding process, however, as the main inverse transform needs to be applied in order to calculate the baseline MSE for the quality adjustment.

ACKNOWLEDGMENT

Authors in this paper acknowledge collaboration from the *Centre National d'Études Spatiales* (CNES) in the obtention of Pléiades simulated data used for the experiments.

REFERENCES

[1] I. Blanes, E. Magli, and J. Serra-Sagrista, "A tutorial on image compression for optical space imaging systems," *IEEE Geoscience and Remote Sensing Magazine*, vol. 2, no. 3, pp. 8–26, 2014.

[2] G. Denis, A. Claverie, X. Pasco, J.-P. Darnis, B. de Maupeou, M. Lafaye, and E. Morel, "Towards disruptions in earth observation? new earth observation systems and markets evolution: Possible scenarios and impacts," *Acta Astronautica*, vol. 137, pp. 415–433, 2017.

[3] N. D. Beser, "Space Data Compression Standards," *Johns Hopkins APL Technical Digest*, vol. 15, no. 3, pp. 206–223, 1994.

[4] B. Li, R. Yang, and H. Jiang, "Remote-Sensing Image Compression Using Two-Dimensional Oriented Wavelet Transform," *IEEE Transactions on Geoscience and Remote sensing*, vol. 49, no. 1, pp. 236–250, 2010.

[5] *Pléiades Imagery User Guide*, Astrium, EADS, October 2012.

[6] J.-M. Delvit, C. Thiebaut, C. Latry, G. Blanchet, and R. Camarero, "A pipeline to improve compressed image quality," in *International Conference on Space Optics—ICSO 2018*, vol. 11180. SPIE, 2019, pp. 2705–2713.

[7] F. García-Vílchez, J. Muñoz-Marí, M. Zortea, I. Blanes, V. González-Ruiz, G. Camps-Valls, A. Plaza, and J. Serra-Sagristà, "On the impact of lossy compression on hyperspectral image classification and unmixing," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 2, pp. 253–257, 2011.

[8] F. Aràndiga, P. Mulet, and V. Renau, "Lossless and near-lossless image compression based on multiresolution analysis," *Journal of Computational and Applied Mathematics*, vol. 242, pp. 70–81, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0377042712004682

[9] R. Camarero, X. Delaunay, and C. Thiebaut, "Fixed-Quality/Variable Bit-Rate On-Board Image Compression for Future CNES Missions," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 8514, p. 851402, 10 2012.

[10] I. Blanes, M. Albinet, R. Camarero, and J. Serra-Sagristà, "Almost Fixed Quality Rate-Allocation Under Unequal Scaling Factors for On-Board Remote-Sensing Data Compression," *International Journal of Remote Sensing*, vol. 39, no. 7, pp. 1953–1970, 2018.

[11] J. Ballé, V. Laparra, and E. Simoncelli, "End-to-End Optimised Image Compression," *International Conference on Learning Representations (ICLR)*, 2017.

[12] J. Ballé, D. C. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational Image Compression with a Scale Hyperprior," *International Conference on Learning Representations (ICLR)*, 2018.

[13] D. C. Minnen, J. Ballé, and G. Toderici, "Joint Autoregressive and Hierarchical Priors for Learned Image Compression," in *NeurIPS*, 2018.

[14] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Learned Image Compression With Discretized Gaussian Mixture Likelihoods and Attention Modules," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 7936–7945.

[15] J. Ballé, P. A. Chou, D. Minnen, S. Singh, N. Johnston, E. Agustsson, S. J. Hwang, and G. Toderici, "Nonlinear Transform Coding," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 2, pp. 339–353, 2021.

[16] F. Yang, L. Herranz, J. v. d. Weijer, J. A. I. Guitián, A. M. López, and M. G. Mozerov, "Variable Rate Deep Image Compression With Modulated Autoencoder," *IEEE Signal Processing Letters*, vol. 27, pp. 331–335, 2020.

[17] Z. Cui, J. Wang, S. Gao, T. Guo, Y. Feng, and B. Bai, "Asymmetric Gained Deep Image Compression With Continuous Rate Adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 10 532–10 541.

[18] F. Kong, K. Hu, Y. Li, D. Li, and S. Zhao, "Spectral–Spatial Feature Partitioned Extraction Based on CNN for Multispectral Image Compression," *Remote Sensing*, vol. 13, no. 1, 2021.

[19] J. Li and Z. Liu, "Efficient Compression Algorithm Using Learning Networks for Remote Sensing Images," *Applied Soft Computing Journal*, vol. 100, p. 106987, 2021.

[20] V. Alves de Oliveira, M. Chabert, T. Oberlin, C. Poulliat, M. Bruno, C. Latry, M. Carlavan, S. Henrot, F. Falzon, and R. Camarero, "Reduced-Complexity End-to-End Variational Autoencoder for on Board Satellite Image Compression," *MDPI Remote Sensing*, vol. 13, no. 3, 2021.

[21] ——, "Satellite Image Compression and Denoising With Neural Networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[22] F. Kong, K. Hu, Y. Li, D. Li, X. Liu, and T. S. Durrani, "A spectral-spatial feature extraction method with polydirectional cnn for multispectral image compression," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 2745–2758, 2022.

[23] Y. Guo, Y. Tao, Y. Chong, S. Pan, and M. Liu, "Edge-guided hyperspectral image compression with interactive dual attention," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–17, 2023.

[24] S. Mijares i Verdú, M. Chabert, T. Oberlin, and J. Serra-Sagristà, "Reduced-complexity multirate remote sensing data compression with neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.

[25] L. Zhang, X. Hu, T. Pan, and L. Zhang, "Global priors with anchored-stripe attention and multiscale convolution for remote sensing image compression," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 138–149, 2024.

[26] L. Theis, W. Shi, A. Cunningham, and F. Huszár, "Lossy Image Compression with Compressive Autoencoders," *International Conference on Learned Representations (ICLR)*, 03 2017.

[27] G. Guerrisi, F. D. Frate, and G. Schiavon, "Artificial intelligence based on-board image compression for the phi-sat-2 mission," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 8063–8075, 2023.

[28] *Information technology - JPEG 2000 image coding system - Part 1: Core coding system*, ISO/IEC, December 2000.

[29] *Spectral Preprocessing Transform for Multispectral and Hyperspectral Image Compression - Recommended Standard CCSDS 122.1-B-1*, Consultative Committee for Space Data Systems (CCSDS), September 2017.

[30] *Low-complexity Lossless and Near-lossless Multispectral and Hyperspectral Image Compression - Recommended Standard CCSDS 123.0-B-2*, Consultative Committee for Space Data Systems (CCSDS), February 2019.

[31] J.-L. Poupat and R. Vitulli, "Cwicom: A highly integrated & innovative ccsds image compression asic," *DASIA 2013-DAta Systems In Aerospace*, vol. 720, p. 62, 2013.

[32] Y. Barrios, A. J. Sánchez, L. Santos, and R. Sarmiento, "Shyloc 2.0: A versatile hardware solution for on-board data and hyperspectral image compression on future space missions," *IEEE Access*, vol. 8, pp. 54 269–54 287, 2020.

[33] X. Cheng and Z. Li, "Predicting the lossless compression ratio of remote sensing images with configurational entropy," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 11 936–11 953, 2021.

[34] D. He, Z. Yang, W. Peng, R. Ma, H. Qin, and Y. Wang, "Elic: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5708–5717, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:247594672

[35] Y. Hu, W. Yang, and J. Liu, "Coarse-to-Fine Hyper-Prior Modeling for Learned Image Compression," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 11 013–11 020, Apr. 2020.

[36] M. Li, W. Zuo, S. Gu, D. Zhao, and D. Zhang, "Learning convolutional networks for content-weighted image compression," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3214–3223.

[37] J. Yang, C. Yang, Y. Ma, S. Liu, and R. Wang, "Learned low bit-rate image compression with adversarial mechanism," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020, pp. 575–579.

[38] T. Dumas, A. Roumy, and C. Guillemot, "Autoencoder based image compression: Can the learning be quantization independent?" in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 1188–1192.

[39] "Landsat 8 Google Cloud Data Portal," https://cloud.google.com/storage/docs/public-datasets/landsat, accessed: 2021-06-06.

[40] "AVIRIS Data Portal, Jet Propulsion Laboratory, NASA," https://aviris.jpl.nasa.gov/dataportal/, accessed: 2021-11-06.

[41] "Test Data for the CCSDS 123.0-B standard, Space Link Services (SLS) Collaborative Work Environment, CCSDS." https://cwe.ccsds.org/sls/default.aspx, accessed: 2015-05-20.

[42] J. Ding, N. Xue, G.-S. Xia, X. Bai, W. Yang, M. Yang, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, "Object detection in aerial images: A large-scale benchmark and challenges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.

[43] J. Bégaint, F. Racapé, S. Feltman, and A. Pushparaja, "Compressai: a pytorch library and evaluation platform for end-to-end compression research," *arXiv preprint arXiv:2011.03029*, 2020.