Fixed-Quality Compression of Remote Sensing Images With Neural Networks

Sebastià Mijares i Verdú, Marie Chabert, 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 article 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 article, 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 artifacts even when coding neighboring areas at different qualities.**

***Index Terms*—Data compression, neural network applications, neural networks, optical data processing, remote sensing.**

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 minimizing a rate-distortion tradeoff [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 generalized 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 article, 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 six 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

Manuscript received 2 March 2024; revised 19 April 2024, 14 May 2024, and 25 June 2024; accepted 26 June 2024. Date of publication 3 July 2024; date of current version 12 July 2024. This work was supported in part by the Institute for Artificial and Natural Intelligence Toulouse (ANITI) under Grant ANR-19-PI3A-0004, in part by the Spanish Ministry of Science and Innovation (MICINN), in part by the European Regional Development Fund (FEDER), funded by under Grant MCIN/AEI/10.13039/501100011033/FEDER, UE, Grant PID2021-125258OB-I00 and Grant PRE2019-088824, in part by the Catalan Government under Grant SGR2021-00643, and in part by the "Data Compression and Machine Learning for Earth Observation Satellites" project under the Institute for Space Studies of Catalonia (IEEC) NewSpace 2024 initiative. *(Corresponding author: Sebastià Mijares i Verdú.)*

Sebastià Mijares i Verdú and Joan Serra-Sagristà are with the Department of Information and Communications Engineering, Universitat Autónoma de Barcelona, 08193 Cerdanyola del Vallès, Spain (e-mail: sebastia.mijares@uab.cat).

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

Thomas Oberlin is with the Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, 31055 Toulouse, France.

Digital Object Identifier 10.1109/JSTARS.2024.3422215

natural images [12], [13], [14], [16], [17] and remote sensing data [18], [19], [20], [21], [22], [23], [24], [25], and are the state of the art in those fields. The most successful methods, introduced in 2017, rely on autoencoders [11], [26]. These methods jointly optimize 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 article, 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 this article is organized 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 this article with a discussion of our findings.

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. Section 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. Section 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 [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" section, 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].

B. Image Compression Using Neural Networks

Contributions by Ballé et al. [11] introducing autoencoders, and later variational autoencoders (VAEs) 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 quantized 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 minimize a rate-distortion tradeoff 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 quantization (adding uniform noise, soft quantization, 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 tradeoff.

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 tradeoff (1), the λ parameter is a constant value set before training. The model is thus optimized for a specific tradeoff 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 [16], [24], [38]. 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 [24], [38]. Our proposed method uses the reduced-complexity multirate codec from Mijares i Verdú et al.'s [24] work, depicted in Fig. 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 quantization. In the decoder, the dequantized latent representation is pixelwise divided by this mask. The scaling or modulation allows to adapt the quantization step to reach a given rate-distortion tradeoff using a neural network trained for another tradeoff without training it again.

In Mijares i Verdú et al.'s [24] work, 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 Mijares i Verdú et al.'s [24] work. We note

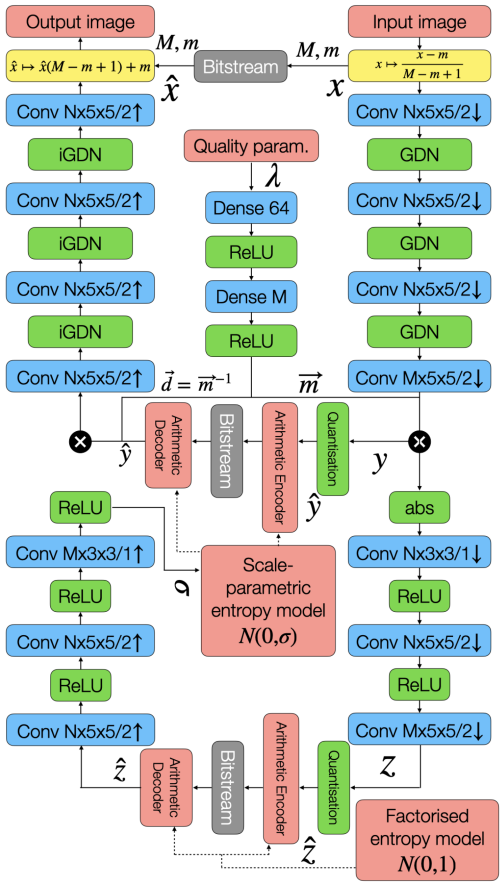


Fig. 1. Architecture used for the proposed method, based on the Mijares i Verdú et al.'s [24] work.

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. Fig. 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 tradeoff 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 nonlinearities used to derive the main transforms and the modulation vector. In this article, 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

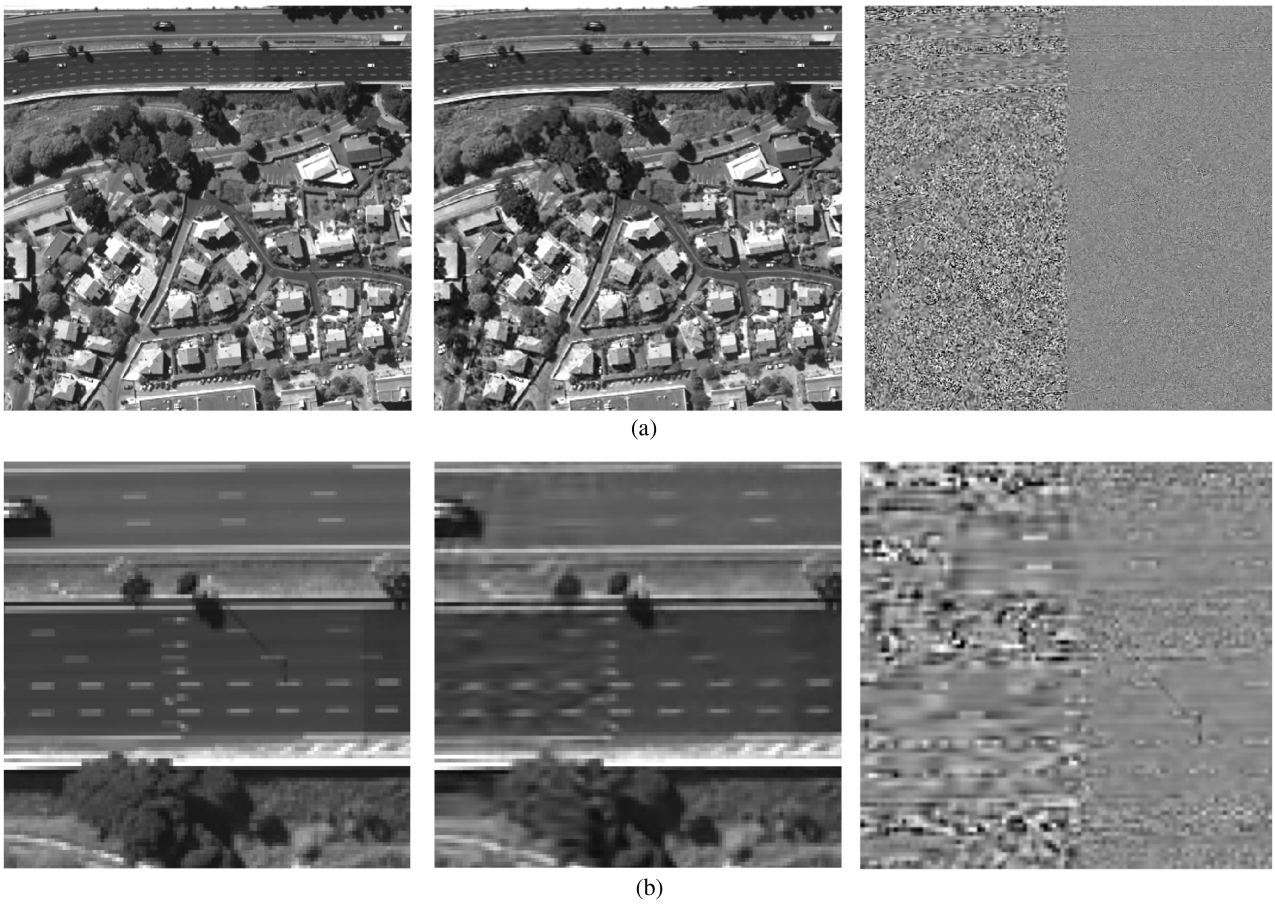


Fig. 2. Simulated Pléiades image compressed using minimum quality on the left half and maximum quality on the right half, resulting in 1.2 bps compression and an overall quality of 38.1 dB PSNR (© CNES 2024). (a) Original (left), reconstruction (center), and difference (right). (b) Zoom-in of the original (left), reconstruction (center), and difference (right).

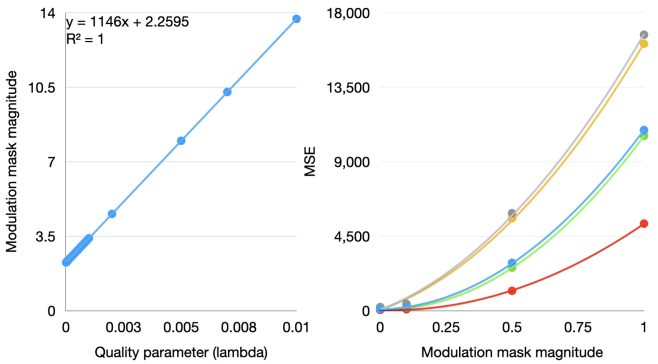


Fig. 3. Relationship between the parameter λ and the modulation mask magnitude $\bar{\mathcal{M}}$ in the Pléiades model (left) and between the modulation mask magnitude $\bar{\mathcal{M}}$ and the reconstruction MSE by our models on several Pléiades images individually, where each curve corresponds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

models, as shown in Fig. 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 quantization step size), as also shown in Fig. 3.

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 Mijares i Verdú et al.'s [24]

work trained for the intended data source, a λ can be estimated to compress the image at the target MSE denoted $\widehat{\text{MSE}}$. 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 quantization 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. Quantization in (5) can be reversed using

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

Using this quantization 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: 1) 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 2) a global target quality and at 3) a local quality with various-sized blocks; and finally 4) whether the proposed method introduces any artifacts upon visual inspection of the images. To evaluate these different aspects, four datasets 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.[1]

A. Datasets and Models

The general validity of the proposed method is assessed using four different remote sensing datasets as follows.

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 seven nonpanchromatic bands [39]. A total of 3584 images are used in training and 1280 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 five 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 dataset is only used to compare the rate-distortion performance of our model to state-of-the-art ML compression methods. The set contains 1403 images for training and 935 images for testing, all of varying sizes.

These datasets are the same as those used in Mijares i Verdú et al.'s [24] work 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 datasets 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 dataset. 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

Dataset	α	**a**	**b**
Pléiades	23.39	2620.50	0.78
Landsat	638.84	16 956.00	11.50
AVIRIS calibrated	35.91	1146.00	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 dataset, 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 (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 dataset the target quality was set to 40 dB PSNR, or 1677.7216 MSE. For the Landsat 8 dataset the target quality was set to 62 dB PSNR, or 2709.5 MSE. For the AVIRIS datasets, 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 $1.3 \times \text{MSE}_0$ of that band for calibrated data, and to $1.5 \times \text{MSE}_0$ of that band for uncalibrated data. Note that the target MSE was therefore uniform for all blocks in any given band.

[1]Data is available online at: https://github.com/smijares/mcos2024/

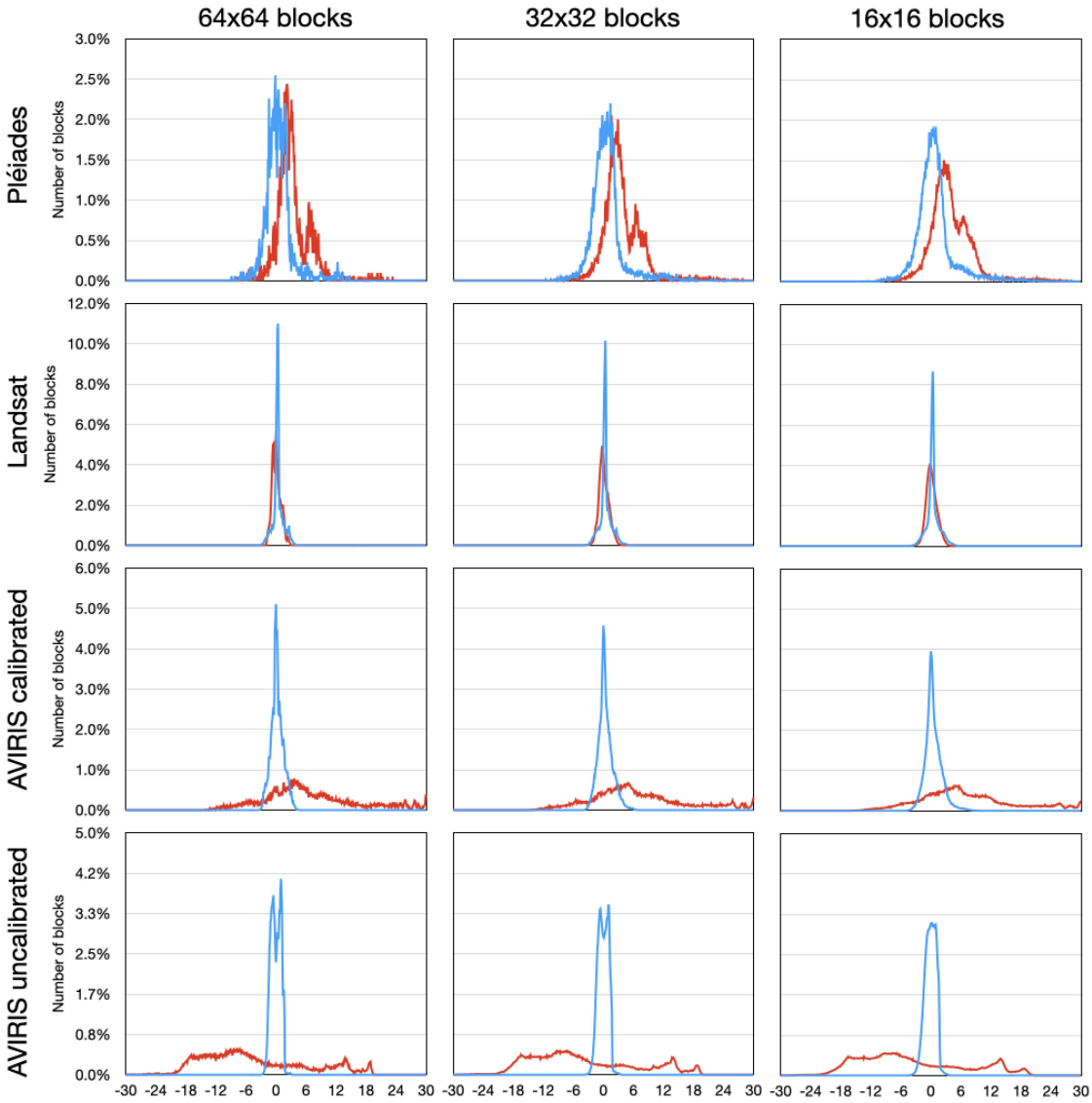


Fig. 4. Differences between the target PSNR and the obtained PSNR for each block in test images when using local fixed-quality compression (blue) or global fixed-quality compression (red).

Fig. 4 shows the histogram of differences between the target PSNR and the one obtained in reconstruction by the models at varying prediction block sizes for each of the datasets. Bin sizes for those histograms are 0.1 dB, so we have bins $(-0.05\ \text{dB}, 0.05\ \text{dB}]$, $(0.05\ \text{dB}, 0.15\ \text{dB}]$, 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.

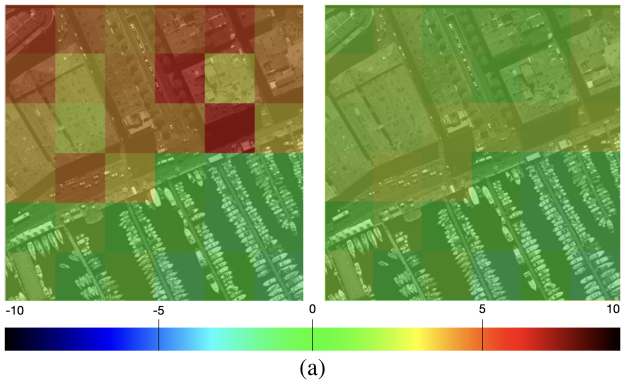
Table II shows the mean blockwise reconstruction target error by block size and dataset, and the standard deviation thereof. The proposed method is less accurate as block size decreases, and the differences between target and actual reconstruction quality are also more spread out as the block size decreases. 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, since $\text{PSNR}(\text{MSE}+\varepsilon) - \text{PSNR}(\text{MSE}) > \text{PSNR}(\text{MSE}) - \text{PSNR}(\text{MSE}-\varepsilon)$ for a fixed deviation $\varepsilon > 0$.

The AVIRIS uncalibrated scenes present a clearly different distribution from the other datasets, attributed to the usage of

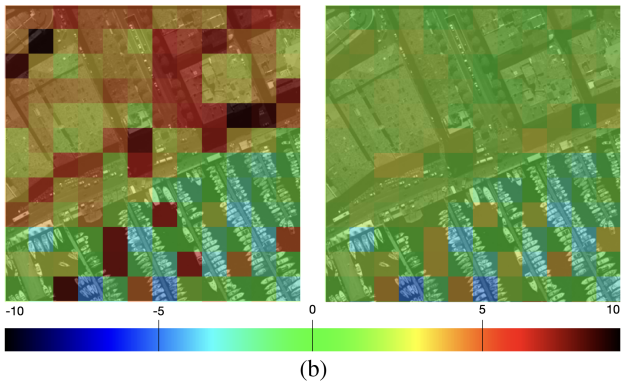
TABLE II
MEAN ERROR IN PSNR QUALITY PREDICTION BY BLOCK SIZE

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

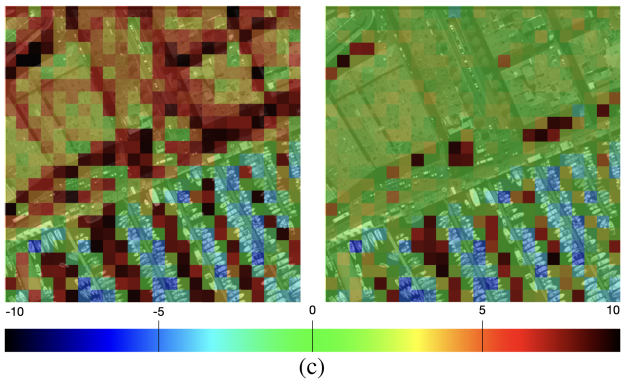
Average and standard deviation.



(a)

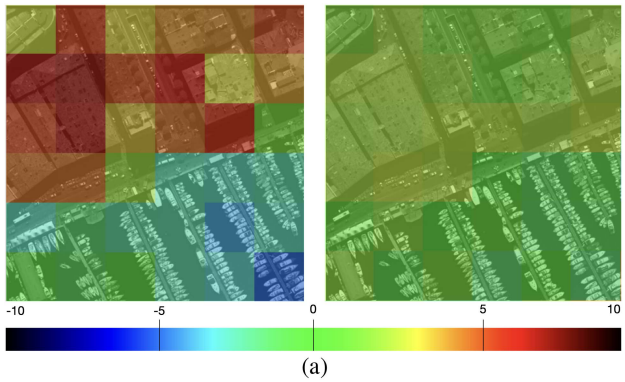


(b)

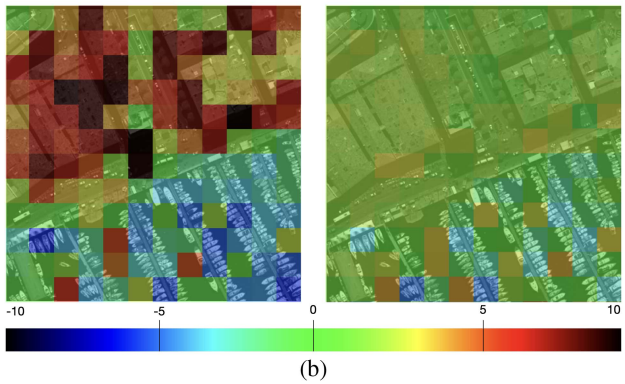


(c)

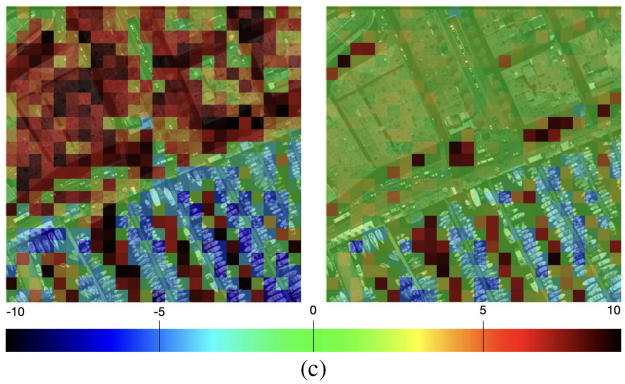
Fig. 5. Differences between the obtained PSNR and the target PSNR (35.71 dB) using fixed global quality (left) and fixed local quality (right) for each block in an example Pléiades image. Note that a positive error means the reconstruction quality was above the set target (© CNES 2024). (a) 64×64 block PSNR errors comparison. (b) 32×32 block PSNR errors comparison. (c) 16×16 block PSNR errors comparison.



(a)



(b)



(c)

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). (a) 64×64 block PSNR errors comparison. (b) 32×32 block PSNR errors comparison. (c) 16×16 block PSNR errors comparison.

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 dataset than for the others in both mean of the prediction and standard deviation of the errors distribution.

Fig. 5 shows the blockwise 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 localized quality as opposed to using a flat global quality, as was indeed shown in general in the previous histograms. Fig. 6 shows a similar blockwise prediction error to Fig. 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.

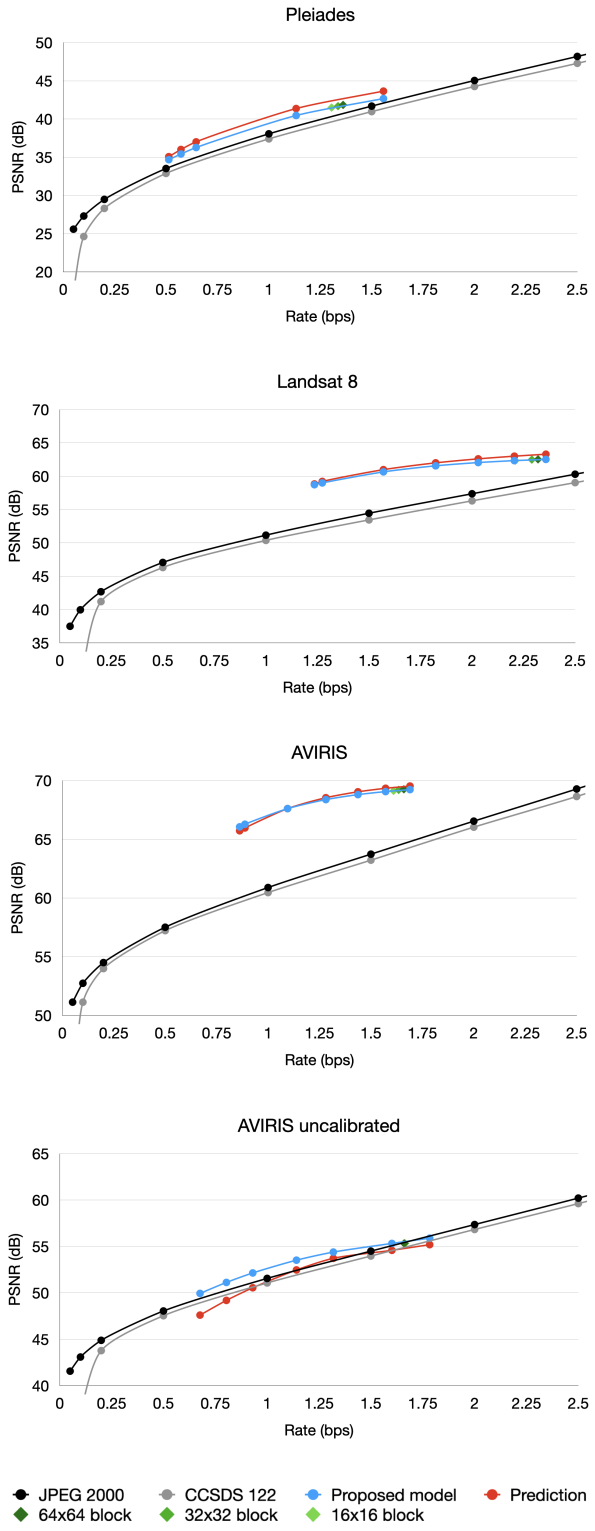


Fig. 7. Rate-distortion performance of our models and reference standards on Pléiades, Landsat 8 OLI, and AVIRIS (calibrated and uncalibrated) datasets.

C. Rate-Distortion Performance

Fig. 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 Mijares i Verdú et al.'s [24] work, 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 datasets except for the AVIRIS uncalibrated dataset at some lower bitrates. This difference can be attributed due to the small dataset available, which is then more susceptible to specific differences between images taking great effect.

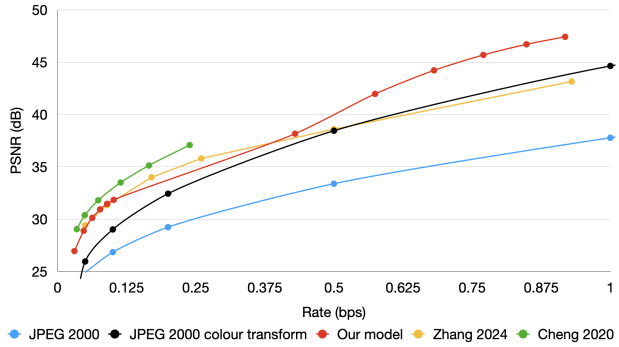


Fig. 8. Rate-distortion performance of our model, the reference standards, and the Zhang et al. 2024 [25] and the Cheng et al. 2020 [14] models.

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 Fig. 7. Having no downside in the rate results by varying the size of prediction blocks, the tradeoff in block-size choice is in MSE accuracy: as discussed in Section 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 blockwise 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 available state-of-the-art model [14], [43]. Results, depicted in Fig. 8 show our 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 substantially 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. Fig. 9 shows an example of how local prediction can better recover some of the features

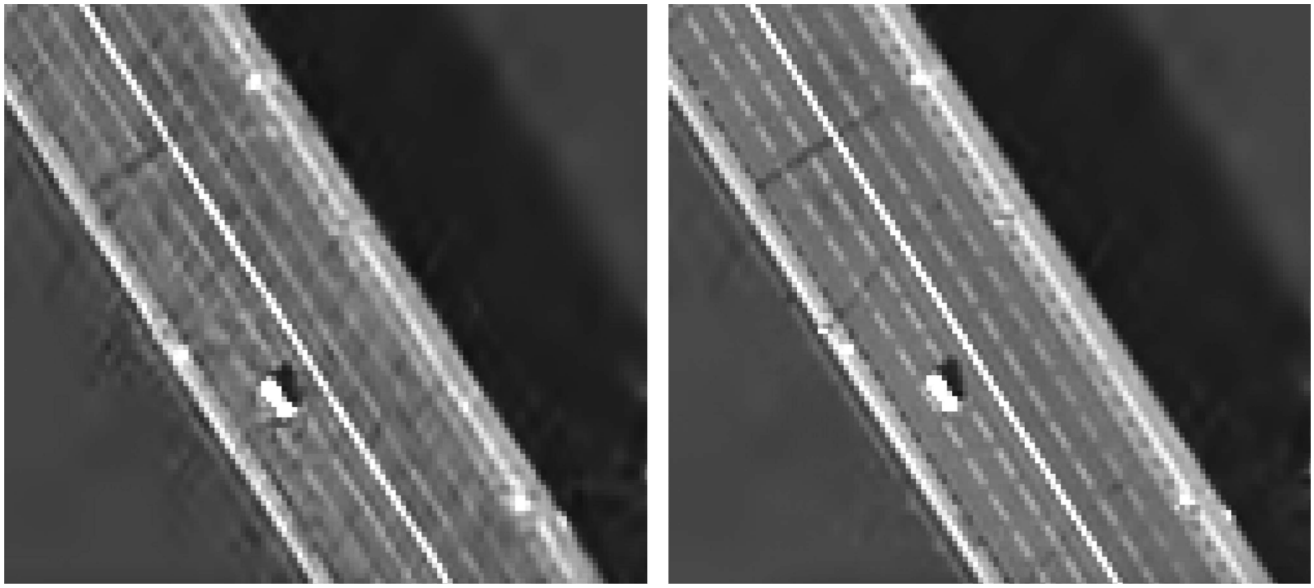


Fig. 9. Zoom-in of a Pléiades image compressed at a global fixed-quality (left) and at a local 16 × 16-block fixed-quality (right) with a compression target PSNR of 37.47 dB (© CNES 2024).

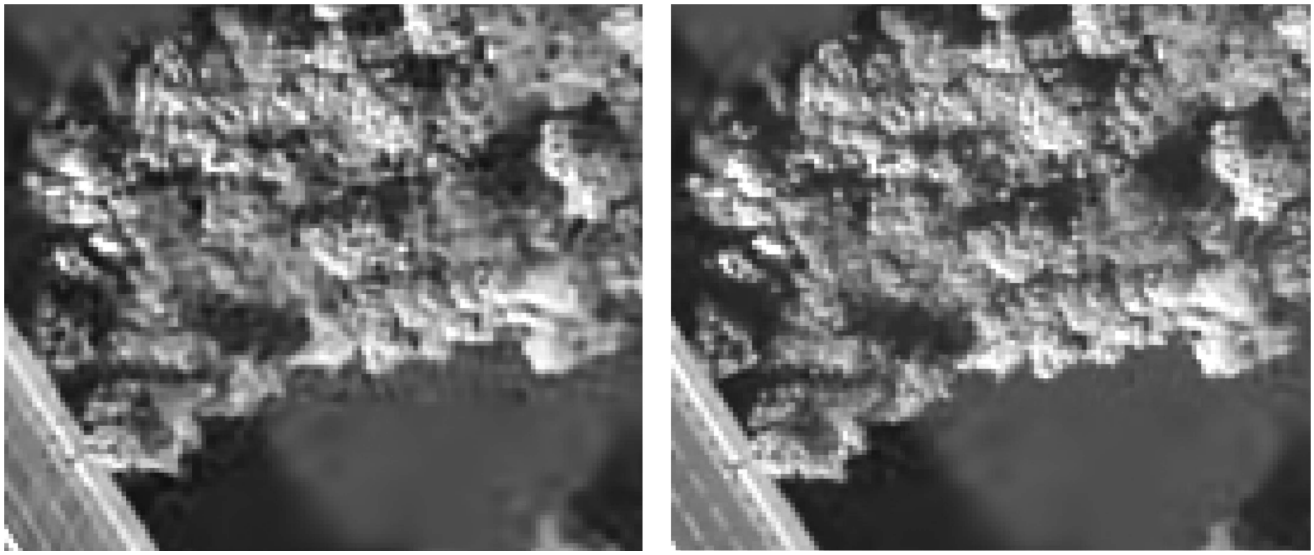


Fig. 10. Zoom-in of a Pléiades image compressed at a global fixed-quality (left) and at a local 16 × 16-block fixed-quality (right) with a compression target PSNR of 37.47 dB. Original is (© CNES 2024).

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. Fig. 10 shows another section of this scene involving vegetation at the water's edge. The global fixed-quality reconstruction produces visible noise artifacts on the water areas at the edge of the vegetation, which local fixed-quality compression does not produce.

Fig. 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 elementwise 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.

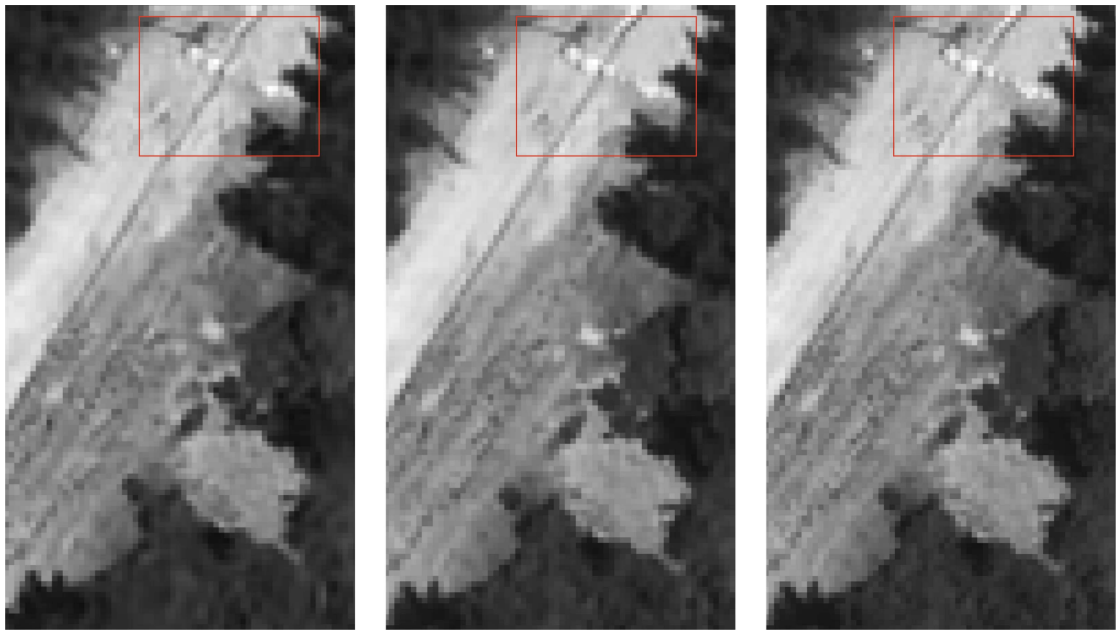


Fig. 11. Zoom-in of a Pléiades image compressed at a global fixed-quality (left) and at a local 16 × 16-block fixed-quality (middle) with a compression target PSNR of 37.47 dB, with the original image (right) for reference (© CNES 2024).

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

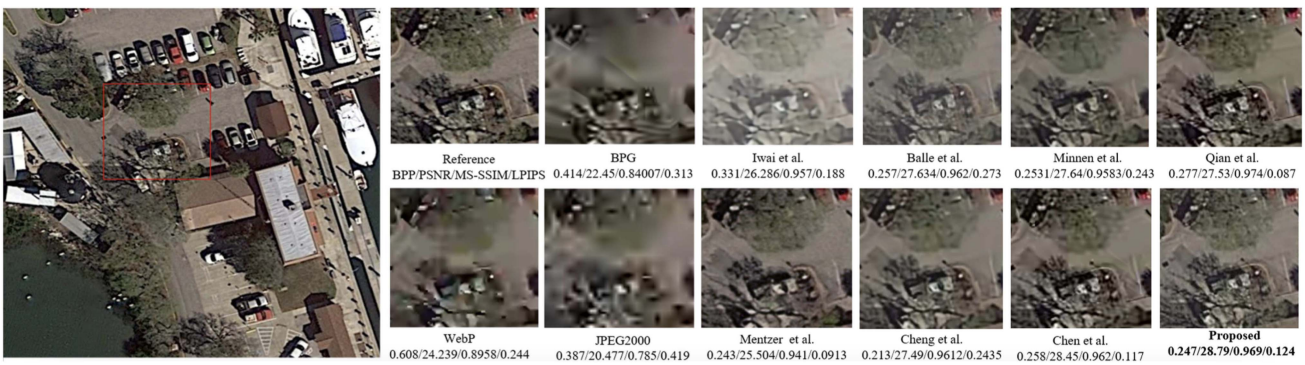


Fig. 13. Visual example of decompressed image and specific zoom area as presented in Zhang et al.'s [25] work. Values in each subfigure correspond to the referenced bpp, PSNR, MS-SSIM, LPIPS of compression with each method.

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 Zhang et al.'s [25] work. Fig. 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 Zhang et al.'s [25] work, shown in Fig. 13, and in line with the overall rate-distortion results shown in Fig. 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 tradeoff 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 tradeoff 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 artifacts appearing 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 article 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 Geosci. Remote Sens. Mag.*, vol. 2, no. 3, pp. 8–26, Sep. 2014.

[2] G. Denis et al., "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 Tech. Dig.*, 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 Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 236–250, Jan. 2011.

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

[6] J.-M. Delvit, C. Thiebaut, C. Latry, G. Blanchet, and R. Camarero, "A pipeline to improve compressed image quality," in *Proc. Int. Conf. Space Opt.-ICSO*, vol. 11180, 2019, pp. 2705–2713.

[7] F. García-Vílchez et al., "On the impact of lossy compression on hyperspectral image classification and unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 253–257, Mar. 2011.

[8] F. Aràndiga, P. Mulet, and V. Renau, "Lossless and near-lossless image compression based on multiresolution analysis," *J. Comput. Appl. Math.*, 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 bitrate on-board image compression for future CNES missions," in *Proc. Conf. SPIE - Int. Soc. Opt. Eng.*, 2012, vol. 8514, Art. no. 851402.

[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," *Int. J. Remote Sens.*, vol. 39, no. 7, pp. 1953–1970, 2018.

[11] J. Ballé, V. Laparra, and E. Simoncelli, "End-to-end optimised image compression," in *Proc. Int. Conf. Learn. Representations*, 2017.

[12] J. Ballé, D. C. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational image compression with a scale hyperprior," in *Proc. Int. Conf. Learn. Representations*, 2018.

[13] D. C. Minnen, J. Ballé, and G. Toderici, "Joint autoregressive and hierarchical priors for learned image compression," in *Proc. 32nd Conf. Neural Inf. Process. Syst.*, 2018.

[14] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Learned image compression with discretized Gaussian mixture likelihoods and attention modules," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 7936–7945.

[15] J. Ballé et al., "Nonlinear transform coding," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 2, pp. 339–353, Feb. 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 Process. Lett.*, 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 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 10532–10541.

[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 Sens.*, vol. 13, no. 1, 2021, Art. no. 9.

[19] J. Li and Z. Liu, "Efficient compression algorithm using learning networks for remote sensing images," *Appl. Soft Comput. J.*, vol. 100, 2021, Art. no. 106987.

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

[21] V. Alves de Oliveira et al., "Satellite image compression and denoising with neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4504105.

[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 J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 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 Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5500817.

[24] S. Mijares i Verdú, M. Chabert, T. Oberlin, and J. Serra-Sagristà, "Reduced-complexity multirate remote sensing data compression with neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 6011705.

[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 J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 138–149, 2024.

[26] L. Theis, W. Shi, A. Cunningham, and F. Huszár, "Lossy image compression with compressive autoencoders," in *Proc. Int. Conf. Learned Representations*, 2017.

[27] G. Guerrisi, F. D. Frate, and G. Schiavon, "Artificial intelligence based on-board image compression for the ϕ-Sat-2 mission," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 8063–8075, 2023.

[28] *Information Technology - JPEG 2000 Image Coding System - Part 1: Core Coding System*, ISO/IEC, Dec. 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), Sep. 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), Feb. 2019.

[31] J.-L. Poupat and R. Vitulli, "CWICOM: A highly integrated & innovative CCSDS image compression ASIC," in *Proc. Conf. Data Syst. Aerosp.*, 2013, vol. 720, Art. no. 62.
[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. 54269–54287, 2020.
[33] X. Cheng and Z. Li, "Predicting the lossless compression ratio of remote sensing images with configurational entropy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 11936–11953, 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," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 5708–5717. [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," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, pp. 11013–11020.
[36] M. Li, W. Zuo, S. Gu, D. Zhao, and D. Zhang, "Learning convolutional networks for content-weighted image compression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 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 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 575–579.
[38] T. Dumas, A. Roumy, and C. Guillemot, "Autoencoder based image compression: Can the learning be quantization independent?," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2018, pp. 1188–1192.
[39] "Landsat 8 Google cloud data portal." Accessed: Jun. 6, 2021. [Online]. Available: https://cloud.google.com/storage/docs/public-datasets/landsat
[40] "AVIRIS data portal, Jet propulsion laboratory, NASA." Accessed: Nov. 6, 2021. [Online]. Available: https://aviris.jpl.nasa.gov/dataportal/
[41] "Test data for the CCSDS 123.0-B standard, space link services (SLS) collaborative work environment, CCSDS." Accessed: May 20, 2015. [Online]. Available: https://cwe.ccsds.org/sls/default.aspx
[42] J. Ding et al., "Object detection in aerial images: A large-scale benchmark and challenges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7778–7796, Nov. 2022.
[43] J. Bégaint, F. Racapé, S. Feltman, and A. Pushparaja, "CompressAI: A pytorch library and evaluation platform for end-to-end compression research," 2020, *arXiv:2011.03029*.

Sebastià Mijares i Verdú received the bachelor's degree in mathematics from the Universidad Complutense de Madrid, Madrid, Spain, in 2019, and the M.Sc. degree in advanced mathematics and mathematical engineering from Universitat Politècnica de Catalunya, Barcelona, Spain, in 2020. He has been working toward the Ph.D. degree in computer science with Universitat Autónoma de Barcelona, Bellaterra, Spain, since 2021.

He is currently a Research Engineer with the Institut d'Estudis Espacials de Catalunya, Barcelona, Spain. His research interests include data compression using neural networks and the usage of neural networks in remote sensing.

Marie Chabert received the engineer's degree in electronics and signal processing from ENSEEIHT, Toulouse, France, and the M.Sc. degree in signal processing from the National Polytechnic Institute of Toulouse, Toulouse, France, both in 1994, and the Ph.D. degree in signal processing and the Habilitation to direct research degree from the National Polytechnic Institute of Toulouse, in 1997 and 2007, respectively.

She is currently a Professor in signal and image processing with ENSEEIHT, National Polytechnic Institute of Toulouse, part of the University of Toulouse where she is giving lectures and labs in signal processing, probability, and statistics. She is the Head of the Signal and Communication research team of the Institut de Recherche en Informatique de Toulouse (IRIT), Toulouse, France. Her research interests include statistical modeling of heterogeneous data, image, and restoration as well as change detection in remote sensing.

Thomas Oberlin (Member, IEEE) received the M.S. degree in applied mathematics from Université Joseph Fourier, Grenoble, France, in 2010, and the engineer's degree in computer science from the Grenoble Institute of Technology, Saint-Martin-d'Héres, France, in 2010. He received the Ph.D. in applied mathematics from the University of Grenoble, Grenoble, France, in 2013.

In 2014, he was a Postdoctoral Fellow in signal processing and medical imaging with Inria Rennes, Rennes, France, before joining as an Assistant Professor INP Toulouse—ENSEEIHT and the IRIT Laboratory, Université de Toulouse, Toulouse, France. In 2019, he joined ISAE-SUPAERO, Université de Toulouse, where he has been a Professor in imaging and machine learning, since 2022. His research interests include signal, image, and data processing and particular time–frequency analysis, representation learning, and regularizations for inverse problems.

Dr. Oberlin has been an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING, since 2022.

Joan Serra-Sagristà (Senior Member, IEEE) received the Ph.D. degree in computer science from Universitat Autónoma Barcelona (UAB), Barcelona, Spain, in 1999.

He is currently a Full Professor with the Department of Information and Communications Engineering, UAB. From 1997 to 1998, he was with the University of Bonn, Bonn, Germany, funded by German Academic Exchange Service (DAAD) and by Catalan Savings Bank "La Caixa." He has coauthored more than 150 publications. His research interests include source coding/data compression, with special attention to image coding for remote sensing and telemedicine applications.

Dr. Serra-Sagristà was the recipient of the Spanish Intensification Young Investigator Award in 2006. He serves or has served as Senior Area Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING, the Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, and *SPIE Journal of Electronic Imaging*, and a Program Committee co-chair for IEEE Data Compression Conference.