VISUAL IMAGE ANALYSIS OF HIGH-RATE LOCAL FIXED-QUALITY COMPRESSION WITH NEURAL NETWORKS

Sebastià Mijares[1,5], Mickaël Bruno[2], Christophe Latry[2], Marie Chabert[3], Thomas Oberlin[4], and Joan Serra-Sagristà[1,5]

[1]***Institut d'Estudis Espacials de Catalunya***
Carrer Esteve Terradas, 1, Oficina 212
Parc Mediterrani de la Tecnologia (PMT), Campus del Baix Llobregat – UPC
08860 Castelldefels, Spain
Email: mijares@ieec.cat

[2]***Centre National d'Études Spatiales (CNES)***
18 Av. Edouard Belin, 31401 Toulouse, France
Email: mickael.bruno@cnes.fr, christophe.latry@cnes.fr
[3]***INP-ENSEEIHT***
2 Rue Charles Camichel, 31000 Toulouse, France
Email: marie.chabert@toulouse-inp.fr
[4]***ISAE-SUPAERO***
10 Av. Edouard Belin, 31400 Toulouse, France
Email: thomas.oberlin@isae-supaero.fr
[5]***Universitat Autònoma de Barcelona***
Campus UAB, 08193 Cerdanyola del Vallès, Spain
Email: joan.serra@uab.cat

Abstract

Fixed-quality compression is a lossy compression paradigm in which data is encoded at a user-defined tolerated distortion. When that tolerated distortion is allocated in certain regions of the image, it is referred to as local fixed-quality compression. It consists in compressing more complex regions at a higher bit rate and less complex areas at a lower bit rate, in order to reach the same level of quality with a reduced overall bit rate. Indeed, the global trend in high resolution satellite imagery is to drive the compression algorithm with a distortion target rather than with a bit rate target, which allows to allocate the bit rate resource where it is needed [1].

The introduction of neural networks to data compression, particularly lossy image compression, has produced a breakthrough in the field. These novel methods outperform established algorithms that are the result of decades of innovation and fine-tuning. These advances have been applied to remote sensing satellite images through reduced-complexity algorithms that are compatible with on-board devices [2, 3]. Indeed, missions such as ESA's Phi-Sat 2 will soon be pioneering the usage of neural compression on board remote sensing satellites [4].

Recent contributions have proposed methods for local fixed-quality compression using neural networks for remote sensing data, presenting rate-distortion and accuracy results for rates up to 1.5 bps or 2.5 bps [5]. However in live applications target bit rates (and therefore image quality) are typically higher, so that distortion is at the level of instrumental noise. In this contribution, models for local fixed-quality compression at high qualities/bit rates are evaluated and visual image analysis is conducted on the artefacts. Furthermore, the computational capacity of the method is evaluated, showing it is compatible with new-generation on-board devices.

Key words: Fixed quality; Data compression; Visual analysis; Pléiades; Lossy compression.

1. INTRODUCTION

From the tallest of watchtowers that is Earth's orbit, satellites gaze upon the Earth as humanity's ultimate look-outs, constantly gathering information from all across the electromagnetic spectrum. In the HR (High Resolution) Earth Observation (EO) domain, instrument performance is constantly improving with higher spatial resolution in significant swath, expanding pixel dynamic range, and growing numbers of spectral bands, resulting in higher bit rate to manage at the output of the instrument. Threatening to negate these advances on the side of data collection, downlink capacity improvements –both in the number of ground stations and their associated transmission time window– have not removed transmission as a crucial bottleneck. In short, these satellites often capture data at a far higher rate than what can be directly sent down to their operators.

Consider, for example, missions such as CNES's Pléiades HR constellation (2011), which can generate up to 3.5 Gbit/s of high-resolution panchromatic and multispectral data, but downlink capacity is only 465 Mbit/s [6]. Similarly, ESA's Sentinel 2 mission (2015) multispectral sensors acquire data at a rate of 1.3 Gbit/s, while downlink capacity is only 560 Mbit/s [7]. Less recent missions such as NASA's Earth Observing 1 (EO-1, 2000) also serve as an example, in this case capturing hyperspectral data at a rate of 527 Mbit/s while downlink capacity was restricted to 105 Mbit/s [8]. The desired bit rates often cannot be achieved using lossless compression, hence lossy compression is used instead to meet those requirements. The aforementioned Pléiades mission uses CWICOM, an ASIC implementing the wavelet transform-based CCSDS 122.0-B-2 standard to compress the collected panchromatic data at 2.86 bits per sample (bps) and the multispectral data at 3.33 bps [6]. The Sentinel 2 mission also uses a wavelet transform-based lossy codec, coding at a rate of 4.15 bps [7]. The EO-1 mission used the Wideband Advanced Recorder Processor (WARP), which implemented lossless Rice coding capable of up to 1.8:1 compression ratio, insufficient to compress all the data the sensors could gather through the available channels [9].

Controlling image quality at the receiver end is critical, especially for high-cost HR-EO missions such as Pléiades, expected to deliver precise high-quality imagery. Traditionally, lossy compression of satellite data used fixed-rate compression, owing to the goal of fitting the collected data through the tight bottleneck. In addition, fixed-rate compression yielded predictable data volumes that are easier to operate. However, in recent years the global trend in high-resolution satellite imagery has been quality-driven lossy compression: compression at a user-defined distortion target which allows to allocate the bit rate resource where it is needed, what is referred to as *fixed-quality compression* [1]. Fixed-quality compression in Earth Observation is an industry need that has been investigated in recent years, with proposals based on CCSDS standards [1, 10].

More recent contributions have pioneered the usage of end-to-end optimised neural network codecs for fixed-quality compression, achieving rate-distortion performance results surpassing those obtained by the CCSDS standards or JPEG 2000 [5]. This later contribution published results for rates much lower than those used in practice (up to 1.5 bps to 2.5 bps in different data sources), at which the differences between reconstructions and the original are clear to the naked eye. Following that thread, in this contribution we analyse the usage of neural networks for fixed-quality compression at the higher qualities required from practical applications. A study of the accuracy and rate-distortion performance is conducted, as well as expert visual analysis of the artefacts introduced by this fixed-quality compression method. Furthermore, as this is to be an on board-compatible compression method, complexity and runtime results are also provided.

2. RELATED WORK

As described in the introduction, there are two fixed-quality compression methods [11, 10] based on the CCSDS 122.0-B-2 standard. Both methods 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 122.0-B-2 and JPEG 2000 as the reference standards. It is observed that the rate-distortion performance of CCSDS 122.0-B-2 in fixed-quality mode should be similar or inferior to that of CCSDS 122.0-B-2 used in conventional mode. These advancements will be deployed in future missions such as the CO3D CNES constellation [12].

The introduction of autoencoders has 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 maps the input image into a *latent representation*, which is quantised and entropy-coded as the compressed image. To perform this entropy coding, most models use a *hyperprior*, 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 arithmetic coding of the latent representation. The decoder performs the opposite operation to that of the encoder, mapping the latent representation into the reconstructed image. These neural networks are all jointly trained to minimise a rate-distortion trade-off loss function, $L = R(\tilde{y}) + \lambda D(x, \hat{x})$, where $R(\tilde{y})$ is the bit rate of the latent representation y with some differentiable substitute for quantisation (adding uniform noise,

soft quantisation, or one of other variants), $D(x,\hat{x})$ is the distortion between the original image, x, and the reconstruction $\hat{x}$, and λ is a constant regulating the rate-distortion trade-off.

Improvements on the state of the art achieved by neural-network codecs for natural image compression has generally come at the expense of greater computational complexity [13]. This is a significant challenge for on-board compression, as the original Ballé *et al.* 2018 models were already too computationally costly for on-board hardware. To address this, several contributions have proposed reduced-complexity variants of these methods for remote sensing image compression, compatible with an on-board implementation [2, 3].

In the rate-distortion trade-off the λ parameter is a constant value set before training in the vast majority of autoencoder architectures for compression. The model is thus optimised for a specific trade-off, so that an image is compressed at a single bit rate. *Modulation* has been successfully applied to these architectures to make them capable of compression at multiple rates [14, 15, 3]. 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 [14, 3].

Based on a modulated autoencoder, recent contributions have pioneered the usage of neural networks for fixed-quality compression. Noting that modulation masks may be different across the image, a method has been proposed to predict the λ parameter to use in each latent pixel so that the reconstruction is at a target MSE [5]. This method is shown to be generally accurate and to hold the highly competitive performance of neural networks, but it was only evaluated at bit rates far lower than those used in real-world applications (up to between 1.5 bps and 2.5 bps).

3. FIXED-QUALITY COMPRESSION METHOD

As introduced at the end of the previous section, we use the fixed-quality compression method from [5], which in turn is based on the reduced-complexity modulated architecture from [3], illustrated in Figure 1. The main underlying principle is 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.

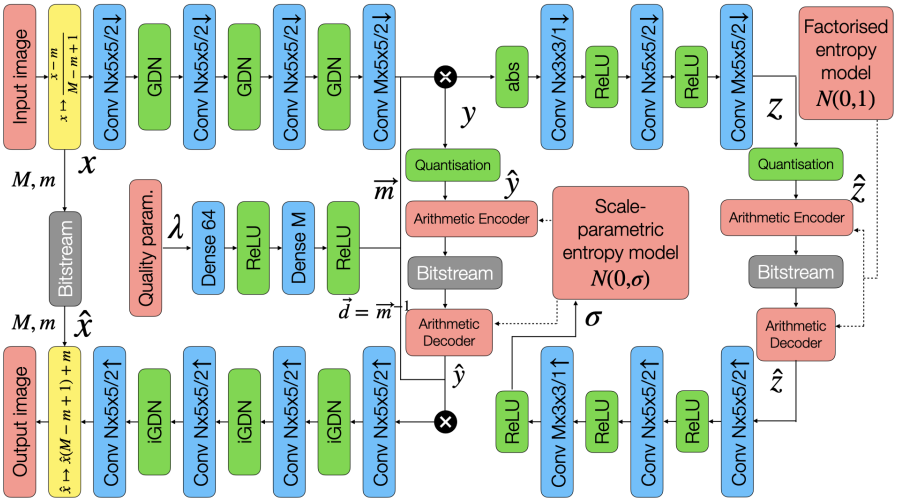

Figure 1: Architecture used for the fixed-quality method, based on the proposals from [3, 5].

The input parameter to the modulation network is the one regulating the rate-distortion trade-off, λ, which on its own does not provide any information on the rate or quality an image will be compressed at. To perform fixed-quality compression, a statistically-derived function is proposed to predict the value of λ to be used in modulation. The prediction function takes two inputs: the target MSE, $\hat{\text{MSE}}$, and the MSE of the reconstructed image without quantising the latent space, MSE_0. The resulting formula is

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

Parameters a and b correspond to the linear relation between λ and the magnitude of the modulation mask $\bar{\mathcal{M}} = \frac{\|\mathcal{M}\|_1}{n}$, while parameter α is derived by regression from a subset of the training set with the formula

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

Using (1), a λ value can be predicted for each latent pixel, which in turn corresponds to a 16×16-pixel block in the final reconstruction, so fixed-quality compression can be applied in regions defined in such 16×16-pixel regions.

3.1. Complexity analysis

The complexity of our models can be characterised by the number of floating-point operations (FLOP) it performs for every pixel in our data. Let N_{pixels} denote the number of pixels in the input image. Given the l-th convolutional layer with N_{in} input features and N_{out} output filters, using $k \times k$ kernels and a stride length of s_l, the number of parameters introduced by that layer will be $N_{par} = N_{in}N_{out}k^2 + N_{out}$, and the number of FLOPs performed by that layer will be $\mathrm{FLOP}_l = \frac{N_{par}N_{pixels}}{\prod_{l=1}^{L} s_l^2}$. For the general divisive normalisation (GDN) [16] layers, the number of parameters and FLOP is the same as that for a convolutional layer with kernel size 1 and stride 1. In this contribution, our models are parametrised for higher bit rates, thus with a larger number of filters, setting $N = 64$ and $M = 384$ as in Figure 1.

The number of operations per pixel of our model, calculated as in [5], is 13,767 FLOP/pixel. This computational complexity is only 18% more than the Alves 2021 reduced-complexity model [2] and 12% less complex than the Mijares 2023 reduced-complexity model [3], thus is compatible with currently available on-board hardware. The recent Φ-Sat-1 ESA mission, pioneering the usage of AI on board, carries the Movidius Myriad 2 [17], which can perform up to 1 TFLOP/s/W. With that available power, our model could process 72.6 Mpixels/s. This would be more than sufficient for a mission such as Pléiades, which captures a daily average of 27.6 Mpixels/s with a maximum acquisition rate of 375 Mpixels/s [18]. In comparison, the Ballé *et al.* 2018 model, with numbers of filters aimed at lower bit rates, would be able to process only 23.4 Mpixels/s, insufficient for a mission such as Pléiades.

4. EXPERIMENTAL RESULTS

4.1. Data set

To evaluate this method, 12-bit simulated panchromatic images of 50 cm resolution are used with the same Point Spread Function (PSF) and signal-to-noise ratio (SNR) as Pléiades panchromatic images (which are 70cm-resolution images at the acquisition, zoomed during ground processing). This set of images has been computed from airborne 10cm-resolution images and is representative of landscape diversity (rural, urban, industrial, coastal areas etc...). Simulation includes convolution with Pléiades PSF and noise simulation, taking into account the Pléiades noise model. A total of 96 images are used in training and 32 images in testing, all with size 820×820. More information on the Pléiades mission is available on the ESA eoPortal [18].

4.2. Fixed-quality compression performance

The proposed fixed-quality compression method is evaluated in its accuracy in allocating quality to each block as well as in rate-distortion performance. The accuracy of the method compares the usage of *local* fixed-quality compression –in which a quality target is set for every $w \times h$-pixel block–, with *global* fixed-quality compression, where a single quality target is set for the full image. Figure 2 shows the distribution of prediction errors between local and global fixed-quality compression. The target MSE is set to the instrumental noise (local or global), and images are compressed using local (in blocks of 16×16 pixels) or global fixed quality. First, the ratio between the target and MSE and the real MSE is measured, which is > 1 if the reconstruction was higher quality than the target, and is < 1 if it was lower quality than the target. Second, the difference between the real peak signal-to-noise ratio (PSNR) and the target PSNR is calculated, which is positive if the reconstruction was higher quality than the target, and negative if it was lower than the target. Similar to the results presented in [5] local fixed-quality compression is more accurate than global fixed-quality compression at higher bit rates and qualities.

Figure 3 depicts the rate-distortion performance of our model compared to JPEG 2000 and CCSDS 122.0-B-2. As expected from [5], our model outperforms these conventional methods at the selected higher bit rates. Average instrumental noise level in Pléiades data, the average target quality for lossy compression of these images, is 48 MSE. At that quality, our

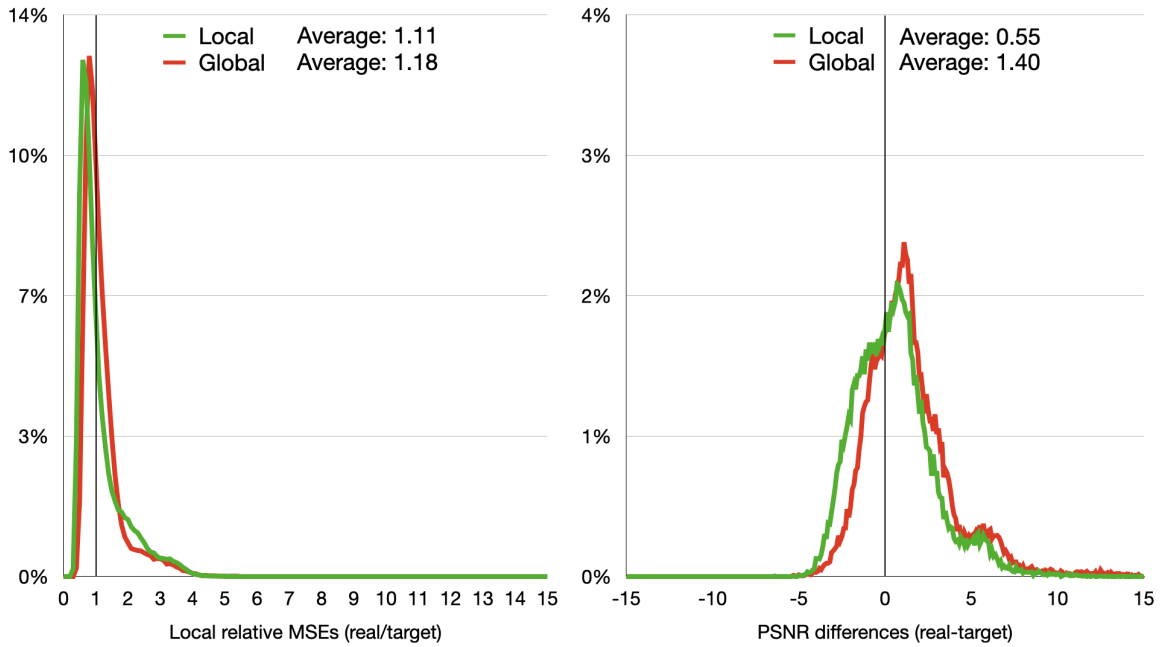

Figure 2: Histograms with the ratio between the real MSE and the target MSE when using local fixed-quality compression (left) and with the difference between the real PSNR and target PSNR (right). Global fixed-quality in this case refers to the usage of a single quality parameter for the full scene. Target quality is given by the local or global instrumental noise.

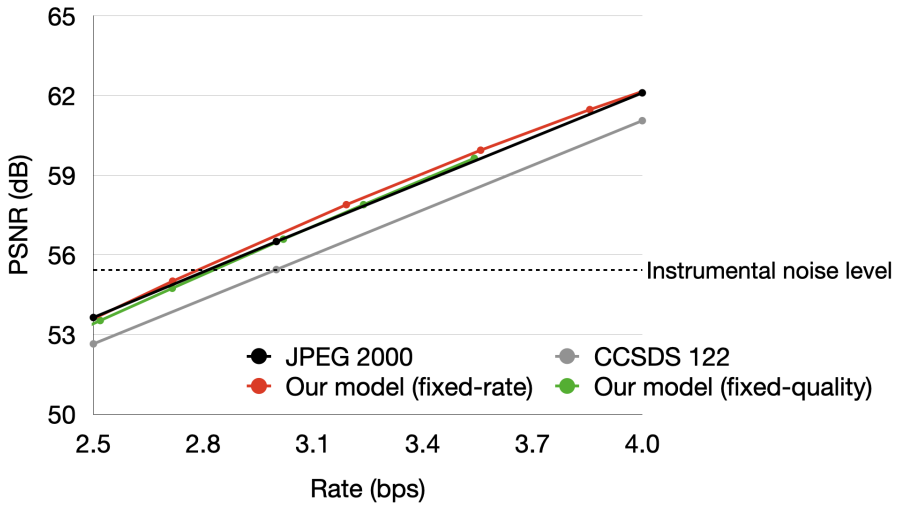

Figure 3: Rate-distortion performance of our model compared to JPEG 2000 and CCSDS 122.0-B-2.

model compresses the test data at an average 2.77 bps, compared to the 2.83 bps of JPEG 2000 and the 3.00 bps of the CCSDS 122.0-B-2. In other words, the proposed model can reduce data rate by 7.66% compared to the CCSDS 122.0-B-2 in fixed-rate mode.

The rate-distortion performance of our model is also evaluated in fixed-quality mode. Target MSE is set to be a constant value for all images, locally in 16×16 blocks. Results shown in Figure 3 show a slight decline in performance in comparison to fixed-rate compression, though this is not in contradiction with results reported in [5], where no significant performance trade-off was observed. For any given image, rate-distortion performance is indeed equivalent between fixed-rate and fixed-quality compression. However, not all images can be compressed at the same qualities using the same model, and so if target MSE is set too high or too low for a certain image, results for that image may be subobptimal, thus negatively affecting the average performance across the entire test set. This is in contrast with fixed-quality compression methods proposed for CCSDS 122.0-B-2, which do negatively affect rate-distortion performance in all cases [11].

4.3. Visual image analysis

In order to verify that the compression-decompression behaves properly, a simplified visual analysis is performed. The goal of this visual analysis is to confirm that the algorithm does not generate structured artefacts and/or any loss that can damage visual image interpretation. For this analysis two sets of images are used: reference (noise-free) images, and those same images with simulated instrumental noise. Noisy images are compressed using our model with local fixed-quality MSE targets set by the instrumental noise in 16×16-pixel blocks, and reconstructions are compared to the reference images. Figure 4 shows an example of the images used.

The visual analysis shows that the model does not generate structured artefacts, even in very noisy areas. These noisy areas are modified, but the compression noise (error) is statistically random. The amplitude of the compression noise is additive to the instrumental noise, as these are both random and independently distributed. This property is very interesting because it means that we have a precise noise model for the compressed image, which allows to use efficient denoising method during on ground processing. The amplitude of the compression noise depends on the local signal level. We can observe low compression noise in darker areas (low level of signal) and higher compression noise in brighter areas.

4.4. Runtime

The encoding time of our model and the CCSDS 122.0-B-2 standard are assessed on low-power hardware. For our model we use our own PyTorch [19] implementation for this test. For the CCSDS 122.0-B-2 standard we use an implementation developed by GICI. Both codecs are evaluated on a Raspberry Pi 3 model B+ with an ARM64 CPU. Encoding times are measured up to the writing to disk of the compressed data.

Our model required an average encoding time of 3.97 s per image, processing 66,031 pixels/s, while the CCSDS 122.0-B-2 standard required 1.65 s per image, or 158,875 pixels/s. As expected, the CCSDS 122.0-B-2 standard is faster than our models, which are 60% slower than the standard. These results, however, do not consider the usage of parallelisation devices such as GPU or TPU, which may reduce encoding time by our models beyond that of the CCSDS 122.0-B-2 standard.

5. CONCLUSIONS

A fixed-quality compression method using neural networks is evaluated for the higher bit rates that are more commonly used in practical applications. The rate-distortion performance and fixed-quality compression accuracy of the method are shown to be qualitatively similar to those found for lower bit rates. Furthermore, visual analysis on the reconstructed images is conducted, finding the model does not introduce structured artefacts, and that the compression noise is independently distributed from –and therefore additive with– the instrumental noise, allowing for efficient denoising to be performed on the ground. Finally, the execution time of the model is compared to that of the CCSDS 122.0-B-2 standard on a low-power device, finding the former to be 60% slower than the latter.

ACKNOWLEDGMENTS

This work was supported in part by the Centre National d'Études Spatiales (CNES) and 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) and the Catalan Government in the framework of the New Space Strategy of Catalonia 2024.

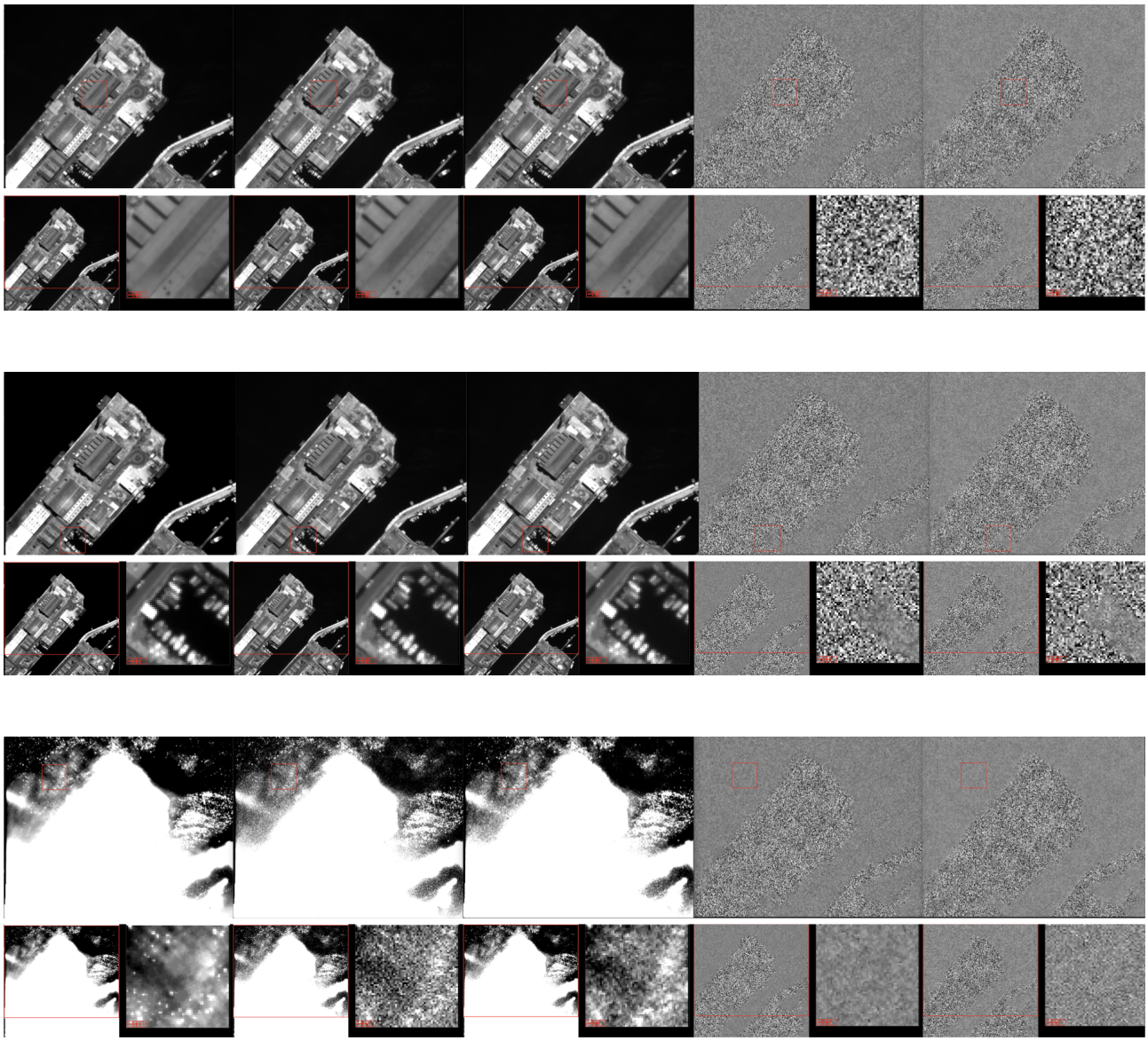

Figure 4: Visual comparisons of a (left) reference image, (middle-left) the simulated noise image, (middle) the noisy image compressed at fixed quality with local quality target equal to the local instrumental noise, (middle-right) the instrumental noise, and (right) the compression noise, as well as detail zoom-in of several areas of the image (© CNES 2024).

REFERENCES

[1] R. Camarero, C. Thiebaut, C. Latry, M. Albinet, J.-M. Delvit, and X. Delaunay, "Future CNES high-resolution remote sensing missions: novel image compression approaches for on-board processing units," in *Proceedings Volume 9501, Satellite Data Compression, Communications, and Processing XI*, 2015.

[2] 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.

[3] 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.

[4] 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.

[5] S. Mijares i Verdú, M. Chabert, T. Oberlin, and J. Serra-Sagristà, "Fixed-Quality Compression of Remote Sensing Images With Neural Networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 12 169–12 180, 2024.

[6] M. A. Gleyzes, L. Perret, and P. Kubik, "Pléiades system architecture and main performances," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXIX-B1, pp. 537–542, 2012.

[7] S. J. Baillarin, A. Meygret, C. Dechoz, B. Petrucci, S. Lacherade, T. Tremas, C. Isola, P. Martimort, and F. Spoto, "Sentinel-2 Level 1 Products and Image Processing Performances," *International Archives of the Photogrammery, Remote Sensing and Spatial Information Sciences*, vol. XXXIX-B1, 2012.

[8] EO Portal, "EO-1 (Earth Observing-1)," https://www.eoportal.org/satellite-missions/eo-1#spacecraft, 2023.

[9] Smith, Terry and Kessler, John, "Earth Orbiter 1: Wideband Advanced Recorder and Processor (WARP)," in *Government Microelectronics and Applications Conference (GOMAC)*, 1999.

[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] 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.

[12] E. Cucchetti, C. Latry, G. Blanchet, J.-M. Delvit, and M. Bruno, "Onboard/on-ground image processing chain for high-resolution earth observation satellites," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIII-B3-2021, pp. 755–762, 2021. [Online]. Available: https://isprs-archives.copernicus.org/articles/XLIII-B3-2021/755/2021/

[13] 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

[14] 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.

[15] 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.

[16] J. Ballé, V. Laparra, E. Simoncelli, M. W. Marcellin, and J. Malo, "Density Modelling of Images Using a Generalised Normalization Transformation," *International Conference on Learning Representations (ICLR)*, 2016.

[17] G. Giuffrida, L. Fanucci, G. Meoni, M. Batič, L. Buckley, A. Dunne, C. van Dijk, M. Esposito, J. Hefele, N. Vercruyssen, G. Furano, M. Pastena, and J. Aschbacher, "The phi-sat-1 mission: The first on-board deep neural network demonstrator for satellite earth observation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[18] EO Portal, "Pleiades-HR (High-Resolution Optical Imaging Constellation of CNES)," https://www.eoportal.org/satellite-missions/pleiades, 2023.

[19] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf