Filtered Residual Compression for Satellite Images

Pascal Bacchus, Renaud Fraisse, Christine Guillemot, Aline Roumy

FILTERED RESIDUAL COMPRESSION FOR SATELLITE IMAGES

Pascal Bacchus[1]*, Renaud Fraisse*[2]*, Christine Guillemot*[1]*, Aline Roumy*[1]

[1]INRIA, Rennes, France, [2]Airbus Defence and Space, Toulouse, France
[1] *name.surname*@inria.fr, [2] renaud.fraisse@airbus.com

ABSTRACT

Learned image compression neural networks have difficulties adapting to certain satellite image characteristics, especially high frequencies that disappear at a high bit-rate in the blur generated in the reconstruction. To answer this problem we describe a joint end-to-end trainable neural network. It is separated into a general compression network and a smaller specialised network. We train a specialized network to compress the residual part of the image to best preserve the high-frequency details present in the satellite images. The proposed model achieves higher rate-distortion performance than current lossy image compression standards and also manages to retrieve details previously poorly reconstructed.

Index Terms— Deep Image Compression, Satellite Application, Residual Compression

1. INTRODUCTION

With the increasing amount of Earth observation data for all types of applications, there is a growing need to develop efficient compression solutions to transmit satellite images. Satellite images have the characteristic of having pixel-sized detail and therefore high frequencies. The major challenge in compressing these images is to be able to distinguish in the high-frequency information what is due to noise and what is the signal for accurate on-ground interpretation.

Learning-based compression, which has made great progress thanks to VAEs [1, 2, 3] (variational auto-encoder), is well suited to satellite images. Because these images have particular statistics, different from natural images, and that learning allows to adapt to these statistics [4, 5, 6, 7]. However, a limitation of these approaches is that they saturate at high rates and cannot reconstruct pixel details. Indeed, blur is an artefact added during compression that results in a lack of detail in fine-grained images such as satellite images with a tendency to have high distortion for high-frequency details. To mitigate this behaviour and still gain from improvements brought by auto-encoder networks, we propose to jointly compress the high frequencies contained in the residual image with a specialized network alongside a general compression scheme.

2. COMPRESSION SCHEME

The proposed general and specialized compression neural networks both share the same architecture, an auto-encoder based network with a hyper-prior as shown in Figure 1. This architecture is composed of two variational auto-encoders (VAE) following one another. The first one encodes the input into a latent representation that is then quantized and entropy coded into a bit-stream. The decoder of this auto-encoder transforms this bit-stream back into a reconstructed image. The second auto-encoder, the hyper-prior, extracts entropy parameters from the latent representation to improve the entropy model.

Compression networks are usually trained over a rate-distortion trade-off. However, this leads to a little blur added during compression due to the $l2$ norm being the reference pixel-based metric [4]. This is troublesome for satellite images as they have high entropy and thus high-frequency details that we want to recover for analysis purposes have disappeared because of a low-band filter effect of the compression. To mitigate this source of error we add a perceptual metric in our rate-distortion trade-off to preserve the high-frequency structures in our images, like stripped patterns. We know that there is a trade-off between distortion and perceptual distance [8]. However, there is a metric for which this trade-off is less severe, VGG [9]. Therefore, we define a loss function based on this network to extract the structures within our features and guide the learning towards a better reconstruction at high frequency [5].

$$P(x,\hat{x}) = \frac{1}{nm}(VGG_{0:2}(x,\hat{x})^2 + VGG_{0:4}(x,\hat{x})^2) \quad (1)$$

The first layers of VGG are selected because they are responsible for low-level shape extraction [10].

Learning parameters are now optimised with a rate-perception-distortion trade-off [5]:

$$\mathcal{L} = \lambda_1 D(x,\hat{x}) + \lambda_2 P(x,\hat{x}) + \alpha R(\hat{y}) \quad (2)$$

With D being the distortion metric MSE, P a perceptual loss based on VGG and R the rate.

To optimise the various terms of the cost function, we use an multi-loss balancing strategy that automatically tunes both

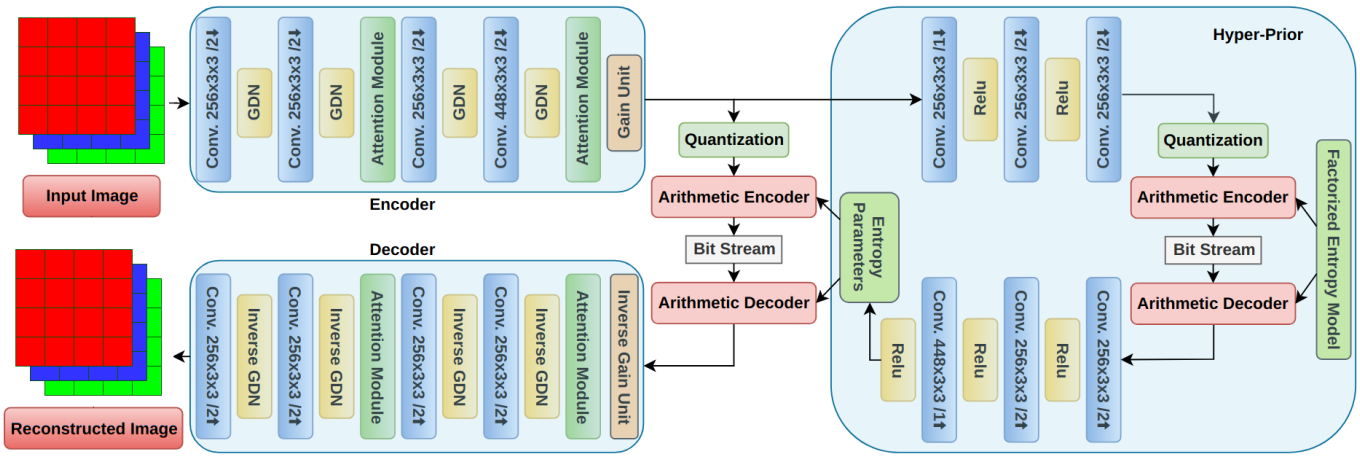


Fig. 1. General compression architecture [4]. The specialised architecture has the same structure with a reduced filter size.

λ_1 and λ_2 so that the loss terms are of the same order of magnitude during training [11].

$$\lambda_k(t) = 2.\frac{exp(\frac{w_k(t)}{T})}{\sum_i exp(\frac{w_i(t)}{T})}, w_k(t) = \frac{L_k(t-1)}{L_k(t-2)} \quad (3)$$

We can go even further to recover the high frequencies by filtering the high frequencies and retaining the typical patterns that clever filters struggle to find, such as striped patterns.

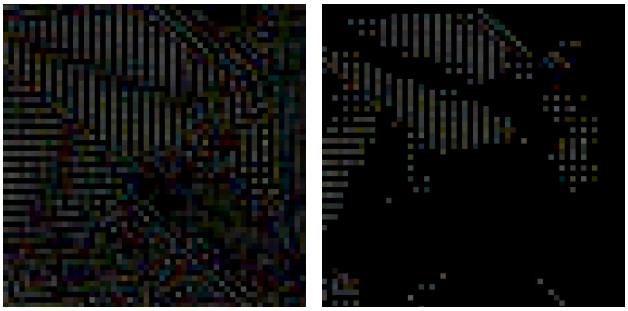
3. FILTERED RESIDUAL COMPRESSION

Instead of relying solely on a general compression network that inherently has troubles with high frequency details, we prefer specialized networks, each responsible for a determined part of the frequency spectrum. It leads to a generic heavy network for a general compression and a secondary lighter network to capture high frequency errors in the residual image. Those errors resulting from a poor reconstruction of high frequencies in general (noise accounted) and striped patterns in particular. The network used to compress the residual is a lighter version of the general compression scheme. Not all the residual image is compressed equally as we use a mask to remove as most unstructured noise as possible. This mask is obtained through filtering and thresholding of the luminance of the RGB residual image. High frequency details are poorly reconstructed compared to textures so we try to filter out random noise induced by compression and keep blocks of meaningful details. The following hand-crafted filter is used to keep only stripped patterns.

$$Kernel = \frac{1}{28}\begin{pmatrix} 2 & 0 & 2 & 0 & 2 \\ 0 & 2 & 0 & 2 & 0 \\ 2 & 0 & 4 & 0 & 2 \\ 0 & 2 & 0 & 2 & 0 \\ 2 & 0 & 2 & 0 & 2 \end{pmatrix}$$

Images we are working on, have a 50 cm spatial resolution and it results in having details at a pixel level. So we focus on patterns that have a 2-pixel period. This is displayed in the filter with a spacing of each component. This filter highlights areas of a certain pattern at the expense of the other. We threshold the result to retain homogeneous patches of errors. Finally, this mask is used on the original residual image to remove unstructured noise expensive to compress.

In Figure 2, we distinguish easily between random and structured errors. The whole process of filter and threshold is to emphasize those meaningful areas that have not been properly compressed so that the specialized compression network only focuses on those few patches of errors.



(a) True residual (b) Post-processing residual

Fig. 2. Effect of our filtering/thresholding mask on residual error

4. EXPERIMENTS

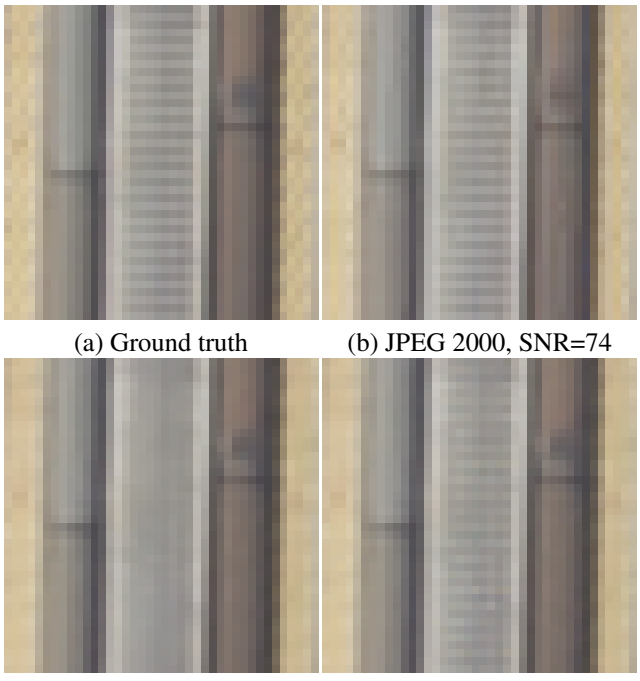
4.1. Training Details

The data set used includes 300 12-bits RGB satellite images (2000x2000) with 50cm geometric resolution [4], 5% are used for testing, and the rest for training. All images are

located around Lyon, France with images of the city, the outskirts and the surrounding countryside. To achieve rotational invariance, each batch of photos is cropped into patches and randomly enhanced with rotation. The neural network compression models were created using the CompressAI Python module [12], an overlay for neural network compression models. Experiments were conducted on NVIDIA A40 GPUs for 200 epochs. Encoder and decoder inference times are around 1 and 1.5 seconds, respectively.

4.2. Qualitative Results

Figure 3 shows the results obtained with different methods in comparison with the ground truth, for a satellite image with a geometric resolution of 50 cm of a train station shelter roof. The reference image is compressed to 2 bpp using JPEG 2000 reference processing, the compression network and the compression network with residual compression added. All the images are well reconstructed since we are aiming for a high bit rate. However, high-frequency details, such as the striped patterns on the roof, have disappeared for model (c) despite a better SNR than JPEG 2000. Model (d) with the addition of residual compression comes close to recovering all these details. Our model with both a general and specialized residual compression recovers locally much more information as some large striped patterns found in cities (pedestrian crossings, rooftops) that were unsuccessfully reconstructed are now compressed with low distortion.



(a) Ground truth (b) JPEG 2000, SNR=74

(c) Network, SNR=112 (d) Network+Residual, SNR=113

Fig. 3. Visual comparison of compressed images at the same bit rate (2 bpp) with the ground truth.

4.3. Quantitative Results

As an objective metric, we use the signal-to-noise ratio (SNR). Given a reference signal I and its damaged processed image $\hat{I}$ we compute the SNR [13]:

$$SNR = \sqrt{\frac{\sum_{pix} I[pix]^2}{\sum_{pix}(I[pix] - \hat{I}[pix])^2}} \quad (4)$$

This gives information on the quantity of noise that has been added to the signal during compression and facilitates comparison between degraded signals.

Figure 4 evaluates the efficiency of our general compression network. We compare it to JPEG 2000 as it is similar to the standard used for RGB images [14] using DCT transforms. For deep learned algorithms comparison, we re-train the compression network from [6] on our data set. Our general compression outperforms all other models, especially for high bit-rate. The effect of the residual compression is not added to the graph because its impact on the overall compression performance is low in comparison of the general compression network. On a quantitative side, the specialized network will increase the performances up to 2 SNR at a rate increase of 0.3 bpp. As seen in previous sections, the gain of this secondary compression network is more about locally recovering high-frequency details.

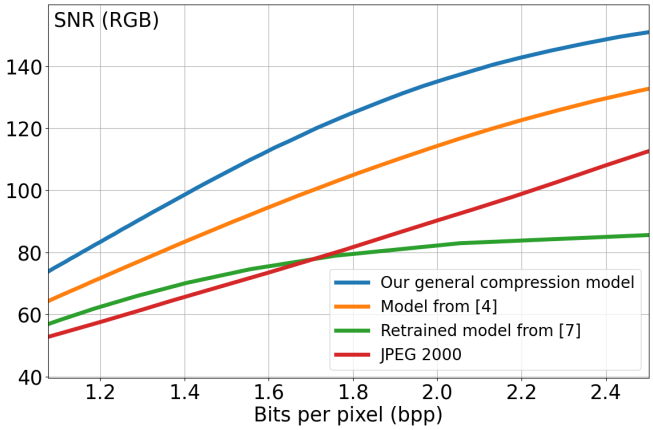


Fig. 4. Our general network compared to baselines

5. CONCLUSION

In this work, we have proposed a compression model designed for RGB satellite images with increased distortion performance compared to traditional sequential processing. The reconstruction is further improved by adding a second network dedicated to compressing the image residue. This compressed residual allows much more information to be recovered locally as some striped patterns found in cities that could not be fully reconstructed are now compressed with low distortion.

6. REFERENCES

[1] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimized image compression," in *ICLR*, 2017.

[2] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational image compression with a scale hyperprior," in *ICLR*, 2018.

[3] D. Minnen, J. Ballé, and G. D. Toderici, "Joint autoregressive and hierarchical priors for learned image compression," in *NeurIPS*, 2018.

[4] P. Bacchus, R. Fraisse, A. Roumy, and C. Guillemot, "Quasi lossless satellite image compression," in *IGARSS 2022*, 2022, pp. 1532–1535.

[5] P. Bacchus, R. Fraisse, A. Roumy, and C. Guillemot, "Joint Compression and Demosaicking for Satellite Images," in *ICASSP 2023*, June 2023.

[6] 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," *Remote Sensing*, vol. 13, no. 3, 2021.

[7] V. Alves de Oliveira, M. Chabert, T. Oberlin, C. Poulliat, M. Bruno, C. Latry, M. Carlavan, S. Henrot, F. Falzon, and R. Camarero, "Satellite image compression and denoising with neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, 2022.

[8] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. jun 2018, IEEE.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.

[10] M. Saeed Rad, B. Bozorgtabar, U. V. Marti, M. Basler, H. Kemal Ekenel, and J. P. Thiran, "SROBB: targeted perceptual loss for single image super-resolution," *CoRR*, vol. abs/1908.07222, 2019.

[11] S. Liu, E. Johns, and A. J. Davison, "End-to-end multi-task learning with attention," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1871–1880, 2019.

[12] 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.

[13] R. Corsino González, R. E. Woods, and B. R. Masters, "Digital image processing, third edition," *Journal of Biomedical Optics*, vol. 14, pp. 376, 2009.

[14] Consultative Committee for Space Data Systems (CCSDS), *Image data compression CCSDS 122.0-B-1*, CCSDS, 2005.