POLITECNICO DI TORINO
Repository ISTITUZIONALE

COMPRESSION WITH SIDE INFORMATION, REVISITED: TAKING ADVANTAGE OF GROUND SEGMENT IMAGE ARCHIVES

Original
COMPRESSION WITH SIDE INFORMATION, REVISITED: TAKING ADVANTAGE OF GROUND SEGMENT IMAGE ARCHIVES / Dordoni, L., Valsesia, D., Magli, E.. - 1:(2026), pp. 1-8. (OBPDC 2026 - 10th International Workshop on On-Board Payload Data Compression Barcellona (Spa) 12-14 October 2026) [10.5281/zenodo.23231572].

Availability:
This version is available at: 11583/3016867 since: 2026-10-22T07:56:07Z

Publisher:
ESA

Published
DOI:10.5281/zenodo.23231572

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

18 November 2026

10th International Workshop on
ON-BOARD PAYLOAD DATA COMPRESSION
12-14 October 2026
Barcelona - SPAIN

COMPRESSION WITH SIDE INFORMATION, REVISITED: TAKING ADVANTAGE OF GROUND SEGMENT IMAGE ARCHIVES

Luca Dordoni[(1)], Diego Valsesia[(1)], Enrico Magli[(1)]

[(1)]Politecnico di Torino
Corso Duca degli Abruzzi 24, Torino, Italy
luca.dordoni@polito.it, diego.valsesia@polito.it, enrico.magli@polito.it

ABSTRACT

Nowadays vast archives of imagery of the Earth surface are available from past observations and keep being updated with new data. At the same time, compressions systems onboard satellites struggle to meet stringent bandwidth constraints. Previously acquired images could be fruitfully used as side information to reduce the amount of information to be transmitted, but they are unavailable to the satellites as they are stored in datacenters at the ground segment. In this work, we revisit the concept of distributed source coding where an encoder that has no explicit access to the side information can still compress to its conditional entropy, provided a suitable decoder exploiting it. In the past, this scheme was limited by the ability to model the correlation between the newly acquired image on the satellite and the archival data. However, we show that decoders based on deep learning can effortlessly address this shortcoming of older approaches. We present designs of compression systems where the onboard encoder operates independently on the images and ground-segment AI-based decoders decompress the encoded data together with the side information of previously acquired images to lead to a higher-quality reconstruction at a given rate. We experimentally show that significant rate-distortion gains can be achieved even when the side information is already compressed, distant in time or affected by clouds.

INTRODUCTION

Earth-observation (EO) satellites continuously acquire large volumes of imagery that must be transmitted to the ground under stringent bandwidth and onboard resource constraints. Image compression is therefore a fundamental component of satellite systems, and both conventional standards and, more recently, learned image compression methods have been developed to improve the trade-off between transmitted rate and reconstruction quality. At the same time, modern Earth-observation programs have accumulated extensive archives of repeated observations of the same geographical areas. These historical acquisitions contain information that is often strongly correlated with newly acquired images and could, in principle, be exploited to improve compression efficiency.

Recently, the focus on compression has been placed on AI algorithms as such methods have achieved increasingly competitive rate-distortion performance by jointly optimizing nonlinear analysis and synthesis transforms together with learned entropy models [1,2,3,4,18]. In parallel, multitemporal EO datasets have highlighted the large amount of spatial and semantic information shared by repeated acquisitions of the same scene [5,6,7,8]. Effectively exploiting this temporal redundancy therefore requires methods capable of identifying which information from a previous observation remains relevant to the current acquisition. However, this is not always practical. Previously acquired imagery is typically stored at the ground segment rather than on the satellite, while retaining and processing large archives onboard would introduce additional storage and computational requirements. Moreover, transmitting reference imagery would place an undesirable burden on the ground-to-space communication channel. For this reason, standard image compression has generally operated independently on each newly acquired image, neglecting ground archives.
Multitemporal EO naturally motivates a distributed source coding (DSC) formulation, to exploit correlated observations that may already be readily available. DSC was introduced with the pioneering Slepian-Wolf and Wyner-Ziv theorems [9,10], which established that correlation between distributed sources can be exploited without requiring the encoder of a compression model to access the correlated information, which can be available at the decoder only. Recent deep-learning methods focusing on highly correlated image pairs such as stereo views [11,12] have further shown that neural networks can effectively exploit decoder-only visual information for reconstruction.
The value of temporal redundancy has also recently been investigated specifically for remote-sensing image compression. Zhang et al. proposed a spatial-temporal context model that exploits information from previous acquisitions of the same geographical location to improve entropy estimation [19]. Similarly, Earth+ leverages historical Earth observations

available onboard as references to identify and transmit primarily the regions that have changed with respect to previous acquisitions [20]. These approaches demonstrate the compression potential of multitemporal imagery, but differ from our setting in that the historical information is made available during onboard encoding, posing a number of challenges, whereas we exploit it exclusively at the ground-segment decoder.

In this work, we investigate learned compression in a DSC setting by using previously acquired satellite images as decoder-side information (SI) for the compression of new observations. Each acquisition is encoded onboard without access to the SI, while the ground segment decoder combines the received representation with a registered historical acquisition used as SI to obtain an improved reconstruction. This arrangement preserves the operational simplicity of an independent onboard encoder and moves the additional processing to the ground, where both the image archive and potentially larger computational resources are available.
We investigate this setting using learned image compression models, such as autoencoder-based codecs [3,4]. We refer to this approach as *conditional reconstruction*: the current acquisition is encoded independently of any historical reference, while the decoder additionally receives a previous observation of the same area and uses it to assist reconstruction. In our experiments, the temporal reference is selected from an earlier revisit of the same satellite. Its usefulness can vary depending on the acquisition conditions and on how the reference itself was stored. We therefore study the effect of temporal separation and, in particular, compare uncompressed SI with references that have already undergone lossy compression.

The main contributions of this work are therefore:
1. a decoder-SI framework that exploits historical Earth-observation imagery while keeping the onboard encoder independent of the SI;
2. a practical realization of the framework, based on conditional reconstruction with end-to-end trained autoencoder neural networks;
3. an experimental evaluation showing rate-distortion improvements on multitemporal Sentinel-2 data, including analysis on the quality of the previously reconstructed acquisitions as side information.

BACKGROUND

Distributed Source Coding

Distributed source coding (DSC) addresses the compression of statistically dependent sources when they cannot be jointly encoded. The seminal Slepian-Wolf theorem established that two correlated discrete sources can be encoded independently and decoded jointly without asymptotic loss in compression efficiency: under lossless reconstruction, the same minimum sum rate as joint encoding, given by their joint entropy, can be approached provided that the individual rates satisfy the corresponding conditional-entropy bounds [9]. Wyner and Ziv subsequently considered the lossy counterpart of this problem, in which one source is encoded without access to a correlated signal that is available as side information only at the decoder, and characterized the corresponding rate-distortion function [10]. This asymmetric setting is particularly relevant to the scenario considered in this work, since historical satellite observations may be available at the ground segment while remaining inaccessible to the onboard encoder.

AI-based Compression

Modern learned image compression methods generally follow a nonlinear transform-coding framework. An analysis transform maps the input image to a latent representation, which is quantized and entropy coded, while a synthesis transform reconstructs the image from the decoded latents. Rather than designing these components independently, the analysis and synthesis transforms and the entropy model are jointly optimized according to a rate-distortion objective, $L = R + \lambda D(x, \hat{x})$, allowing the latent representation to adapt directly to the compression task [1,2,3,4,15,16].
A central component of these approaches is the probabilistic model used to estimate the distribution of the quantized latents, since more accurate probability estimates directly translate into more efficient entropy coding. Ballé et al. introduced a hierarchical entropy model based on a *hyperprior*, in which additional side latents are extracted from the main representation, transmitted as part of the bitstream, and used at the decoder to predict the statistics of the primary latents [1]. Some works combined this hierarchical representation with an autoregressive context model, exploiting both information provided by the hyperprior and dependencies among previously decoded latent symbols [2].
More recent learned codecs therefore seek to improve both the expressiveness of the transforms and the accuracy of the entropy model while retaining practical coding complexity. ELIC addresses this trade-off through an efficient transform

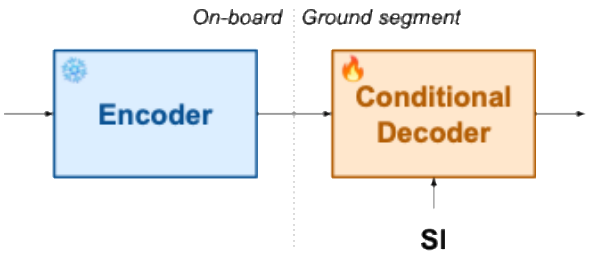


Fig. 1. Overview of the proposed *conditional reconstruction*. It employs a learned encoder and a conditional decoder that exploits the side information during image reconstruction.

architecture and a space-channel context model [3]. In particular, it exploits the nonuniform distribution of information across latent channels through uneven channel grouping and combines channel-wise dependencies with spatial context modeling. This design provides strong rate-distortion performance while avoiding the fully sequential decoding process associated with traditional autoregressive models and enabling more efficient parallel processing.

Visual Side Information and Multitemporal Compression

Deep-learning approaches have recently made decoder-SI significantly easier to exploit in practical modern compression systems. Ayzik et al. proposed a learned framework in which a correlated image available only at the decoder is used to improve reconstruction by identifying correspondences between the decoded image and the reference [13]. Subsequent works have explored learned common representations and attention-based mechanisms for combining coded information with correlated decoder-side images [11,12].
Multitemporal EO imagery provides a natural source of SI, but poses a more challenging correspondence problem. Repeated acquisitions of the same geographical area retain persistent structures such as roads, buildings, vegetation patterns, and other land features, but may also differ substantially because of seasonal evolution, illumination, atmospheric conditions, clouds, land-cover changes, and residual registration errors. Consequently, the decoder must determine which information from the historical observation remains relevant rather than directly transferring its content.
Temporal redundancy has also been investigated specifically for remote-sensing compression. Earlier approaches exploit a previous hyperspectral acquisition through temporal prediction or spectral-temporal decorrelation [14]. More recent learned methods such as TerraCodec-TT condition the coding process of a current acquisition on previously coded observations [17]. These approaches demonstrate the value of historical observations for compression, but differ from the setting considered here because the temporal reference is available during encoding. Our objective is instead to preserve an encoder that operates independently of archival imagery and introduce temporal information exclusively at the decoder.

METHOD

The proposed system uses historical observations exclusively during ground-segment reconstruction. Each newly acquired EO image is compressed onboard without requiring access to earlier observations of the same scene. After transmission, the ground segment retrieves a previous acquisition of the corresponding geographical area, registers it to the current image, and provides it to the reconstruction network as SI. Denoting the current image by x and the registered temporal reference by s, only x contributes to the transmitted representation. The historical observation enters the processing chain only at the receiver, so no SI-dependent operation is required onboard. For learned compression, we implement this principle by conditioning the synthesis stage on both the decoded latent representation and the temporal reference.
An overview of the configuration is shown in Fig. 1.

Conditional Reconstruction

The learned encoder generates and entropy-codes the representation of the current observation using only the current image. At the ground decoder, reconstruction is instead conditioned on both the decoded latents and the registered SI. We assume that a state-of-the-art autoencoder neural network is used for the encoder and decoder. The decoder architecture is then modified via an adapter to modulate the features of the image under decoding with the features of the SI.

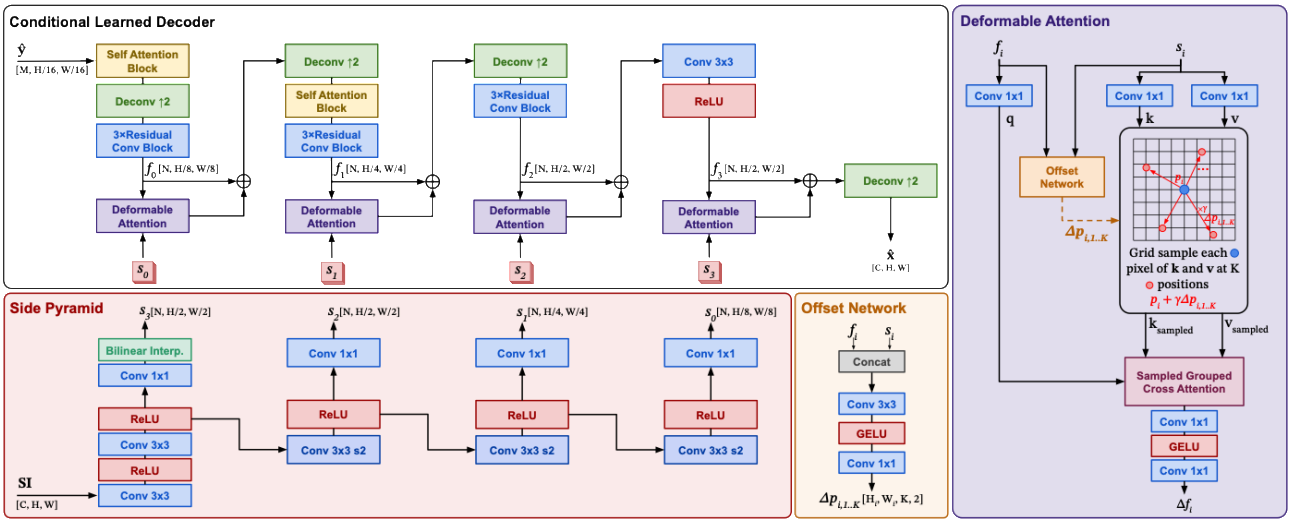


Fig. 2. Multiscale SI adapter used to inject temporal-reference features into the learned decoder.

SI is introduced in the decoder neural network through a *deformable attention adapter* attached to multiple stages of the decoder. An overview of the overall architecture of the adapter as integrated for the ELIC decoder is shown in Fig. 2. A separate multiscale branch extracts features from the SI at multiple resolutions, while the current decoder features guide the selection of a sparse set of relevant positions from which temporal information is sampled. Cross-attention is then used to select and aggregate the useful SI features before injecting them residually into the reconstruction stream. This spatially adaptive mechanism is intended to handle imperfect geometric correspondence between multitemporal acquisitions. Moreover, the attention operations is input-adaptive and allows to reject the influence of SI regions affected by clouds, illumination changes, or actual modifications of the observed surface.

Training is performed end-to-end with the standard rate-distortion criterion, allowing the learned representation to adapt to the fact that additional temporal information will only become available at the decoder.

EXPERIMENTS

We evaluate the *conditional reconstruction* system assuming that SI is already stored at the ground segment as the image acquired during a previous revisit. Its acquisition rate is therefore excluded from the cost of the current transmission. The experiments first compare different SI-quality conditions and subsequently investigate how the benefit changes when increasingly old acquisitions are selected as reference.

Data and Experimental Setup

The evaluation is performed on multitemporal Sentinel-2 L1C data from WorldStrat. The dataset comprises 3109 training locations, 392 validation locations, and 389 test locations, each associated with 16 observations acquired between 2017 and 2022. Neighboring acquisitions are separated by roughly five days at the median, while a full temporal sequence spans about 150 days. Cloudy observations are retained rather than filtered out, making the data representative of realistic acquisition conditions.

We retain the B04, B03, B02, and B08 channels, corresponding respectively to red, green, blue, and near-infrared information. Sentinel-2 digital numbers are scaled to obtain reflectance values, which are subsequently restricted to the [0,1] interval. Images from the same temporal sequence are aligned prior to training by estimating a subpixel translation relative to the latest acquisition through phase cross-correlation. Shift estimation operates on standardized high-pass versions of the images, reducing sensitivity to broad radiometric variations. Cloud masks restrict the computation to jointly clear pixels, and the estimated registration is rejected whenever the valid area falls below 10% or the displacement magnitude exceeds 20 pixels. Each training epoch uses 20000 randomly drawn 128 × 128 patches, whereas testing is

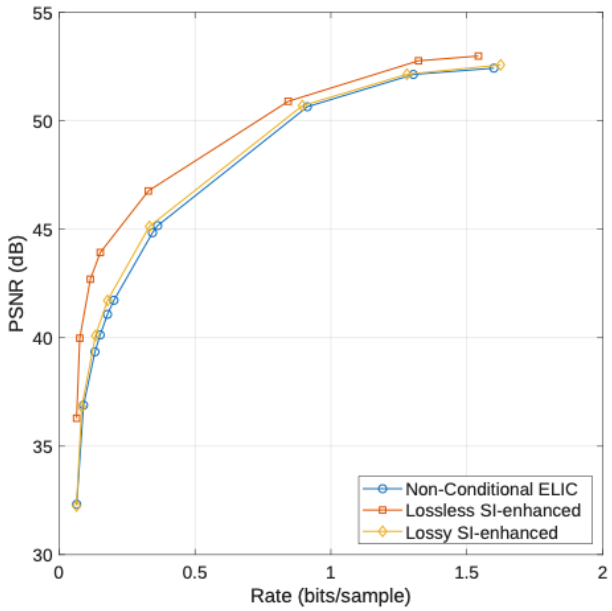


Fig. 3. Rate-distortion performance of conditional ELIC on WorldStrat using the second-to-last acquisition as SI at different quality levels.

carried out on fixed central crops. Temporal ordering is preserved throughout evaluation. Bitrate is expressed in bits per pixel per channel by normalizing the actual entropy-coded bitstream size by the number of image samples HWC.
Reconstruction targets the final acquisition of each sequence, $t = T-1$, using the immediately preceding observation, $t = T-2$, as the temporal reference. Reported PSNR values are obtained by first evaluating each of the 389 test locations independently and then averaging over locations. Since the reference is assumed to be already present in the ground archive, only the bitstream associated with the current acquisition contributes to the reported rate.

Conditional Reconstruction Experiments

ELIC [3] is adopted as the learned compression backbone because it provides a useful compromise between coding performance and model complexity. The proposed temporal modules are inserted only in its decoder, while a standard ELIC implementation provides the SI-free baseline. Independent networks are optimized for the different rate-distortion operating points, using 600 training epochs and an identical sampling and augmentation pipeline across all configurations.

The configuration uses 192 internal feature channels and a latent dimensionality of 320. At each adapter stage, nine SI positions are sampled for every spatial location, with groups of 16 channels and an offset scale of 0.1. Samples outside the feature map have zero padding. Queries, keys, and values are generated through 1×1 projections, and the attended SI contribution is further transformed by two pointwise convolutions separated by GELU before residual fusion with the decoder stream. The original reconstruction head is left unchanged.
We evaluate two quality levels for the SI. For the lossless-reference experiment, the decoder receives the registered $t = T-2$ observation without any compression degradation. In the lossy case, the temporal reference is instead generated by passing the earlier acquisition through an independently trained ELIC model at the matching operating point; the reconstructed output of that model is then used as SI. A dedicated conditional network is optimized for each combination of reference quality and rate-distortion setting.

The SI conditioning advantage is made clear in Fig. 3. The uncompressed reference produces the largest shift in the rate-distortion curve, while using a previously compressed observation still improves over independent ELIC, albeit by a smaller margin. Thus, even an imperfect archived reconstruction retains information that is complementary to the current bitstream.

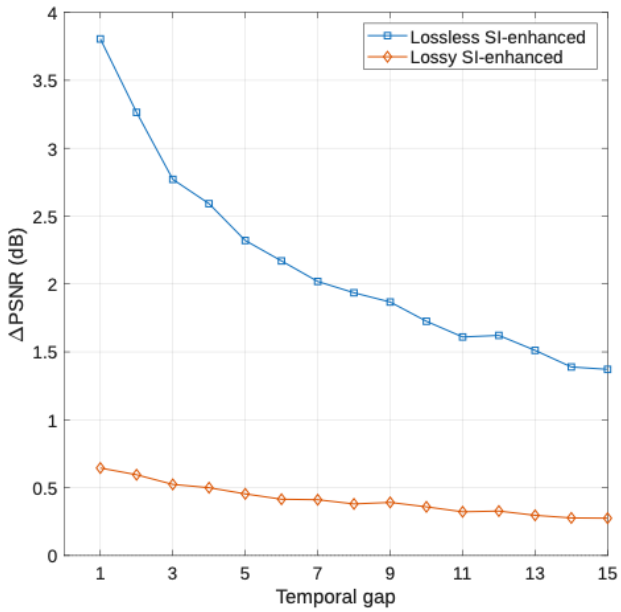


Fig. 4. Sensitivity to side-information recency, measured as the PSNR difference with respect to plain ELIC at the same operating point, as a function of the side-information acquisition time.

Influence of SI Age

The effectiveness of SI is expected to decrease as its temporal distance from the current acquisition increases, since older observations are more likely to exhibit differences due to seasonal evolution, illumination changes, atmospheric conditions, or modifications of the observed scene. We therefore evaluate the sensitivity of conditional reconstruction to SI recency using ELIC under the same single-acquisition setting, as shown in Fig. 4. The target image is always the most recent acquisition, while progressively older observations from the WorldStrat sequence are used as SI. A temporal gap of one corresponds to the immediately preceding acquisition, whereas larger gaps indicate increasingly older references. Performance is measured as the PSNR gain over non-conditional ELIC at the same operating point. As the age of the reference increases, its contribution becomes progressively weaker: lossless SI provides roughly a 3.8 dB advantage at a separation of about ten days, compared with approximately 1.4 dB after around five months. This trend is consistent with the decreasing correspondence between acquisitions over time. Still, conditional reconstruction remains better than the SI-free baseline at every tested gap, suggesting that stable geographical content remains informative even after substantial temporal separation.

CONCLUSION

This work investigated the use of historical Earth-observation imagery as decoder-SI for improving the compression of newly acquired satellite images. The proposed framework keeps the onboard encoder completely independent of the SI and introduces temporal information only at the ground segment, where historical acquisitions and larger computational resources are naturally available. In *conditional reconstruction*, a neural network compression pipeline adopts a conditional decoder which is augmented with a multiscale *deformable attention adapter* that progressively incorporates features extracted from the historical acquisition with the features of the image under decoding.
Experiments on multitemporal Sentinel-2 imagery from WorldStrat demonstrate that SI can provide substantial rate-distortion improvements. With the ELIC compression model used as a backbone, both lossless SI and lossy-compressed SI are beneficial for conditional reconstruction. The analysis of SI recency further shows that temporally closer observations provide the largest benefits, as expected, but that the gain remains positive even when considerably older acquisitions are used. This indicates that persistent spatial structures in Earth-observation imagery can remain useful for reconstruction despite temporal differences.

Overall, the results support decoder-SI conditioning as a practical way of exploiting the growing archives of repeated satellite observations without increasing onboard storage requirements or requiring access to historical imagery during encoding. Future work will investigate more advanced SI selection strategies, robustness to larger scene changes and cloud contamination. In addition, we plan to extend the proposed framework to a non-AI setting, where the backbone compression algorithm is non-learnable.

ACKNOWLEDGEMENT

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. This work is supported by ERC grant 101199648 - IntelliSwarm.

REFERENCES

[1] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston. "Variational image compression with a scale hyperprior." *In International Conference on Learning Representations (ICLR), 2018.*

[2] D. Minnen, J. Ballé, and G. D. Toderici. "Joint autoregressive and hierarchical priors for learned image compression." *Advances in neural information processing systems* 31 (2018).

[3] 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 *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5708-5717. IEEE, 2022.

[4] W. Jiang, J. Yang, Y. Zhai, F. Gao, and R. Wang. "MLIC++: Linear complexity multi-reference entropy modeling for learned image compression." *ACM Transactions on Multimedia Computing, Communications and Applications* 21, no. 5 (2025): 1-25.

[5] J. Cornebise, I. Oršolić, and F. Kalaitzis. "Open high-resolution satellite imagery: The worldstrat dataset–with application to super-resolution." *Advances in Neural Information Processing Systems* 35 (2022): 25979-25991.

[6] P. Ebel, Y. Xu, M. Schmitt, and X. X. Zhu. "SEN12MS-CR-TS: A remote-sensing data set for multimodal multitemporal cloud removal." IEEE Transactions on Geoscience and Remote Sensing 60 (2022): 1-14.

[7] V. S. F. Garnot and L. Landrieu. "Panoptic segmentation of satellite image time series with convolutional temporal attention networks." In *2021 IEEE/CVF International Conference on Computer Vision (ICCV),* pp. 4852-4861. IEEE, 2021.

[8] A. Van Etten, D. Hogan, J. Martinez Manso, J. Shermeyer, N. Weir, and R. Lewis. "The multi-temporal urban development spacenet dataset." In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6394-6403. IEEE, 2021.

[9] D. Slepian and J. Wolf. "Noiseless coding of correlated information sources." *IEEE Transactions on information Theory* 19, no. 4 (1973): 471-480.

[10] A. Wyner and J. Ziv. "The rate-distortion function for source coding with side information at the decoder." *IEEE Transactions on information Theory* 22, no. 1 (1976): 1-10.

[11] N. Mital, E. Özyılkan, A. Garjani, and D. Gündüz. "Neural distributed image compression using common information." In *2022 Data Compression Conference (DCC)*, pp. 182-191. IEEE, 2022.

[12] N. Mital, E. Özyilkan, A. Garjani, and D. Gündüz. "Neural distributed image compression with cross-attention feature alignment." In *Proceedings of the IEEE/CVF Winter conference on applications of computer vision*, pp. 2498-2507. 2023.

[13] S. Ayzik and S. Avidan. "Deep image compression using decoder side information." In *European Conference on Computer Vision*, pp. 699-714. Cham: Springer International Publishing, 2020.

[14] W. Zhu, Q. Du, and J. E. Fowler. "Multitemporal hyperspectral image compression." *IEEE Geoscience and Remote Sensing Letters* 8, no. 3 (2010): 416-420.

[15] L. Theis, W. Shi, A. Cunningham, and F. Huszár. "Lossy image compression with compressive autoencoders." *In International Conference on Learning Representations (ICLR), 2017.*

[16] O. Rippel and L. Bourdev. "Real-time adaptive image compression." In *International conference on machine learning*, pp. 2922-2930. PMLR, 2017.

[17] J. Costa Watanabe, I. Wittmann, B. Blumenstiel, and K. Schindler. "TerraCodec: Compressing Optical Earth Observation Data." In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 271-287. Cham: Springer Nature Switzerland, 2026.

[18] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto. "Learned image compression with discretized gaussian mixture likelihoods and attention modules." In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7936-7945. IEEE, 2020.

[19] J. Zhang, R. Dong, J. Zheng, M. Chen, L. Zhang, Y. Zhao, and H. Fu. "Spatial-temporal context model for remote sensing imagery compression." In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 6063-6072. 2024.

[20] K. Du, Y. Cheng, P. Olsen, S. Noghabi, and J. Jiang. "Earth+: On-board satellite imagery compression leveraging historical earth observations." In *Proceedings of the 30th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 1*, pp. 361-376. 2025.