E2S EMBED 2SCALE

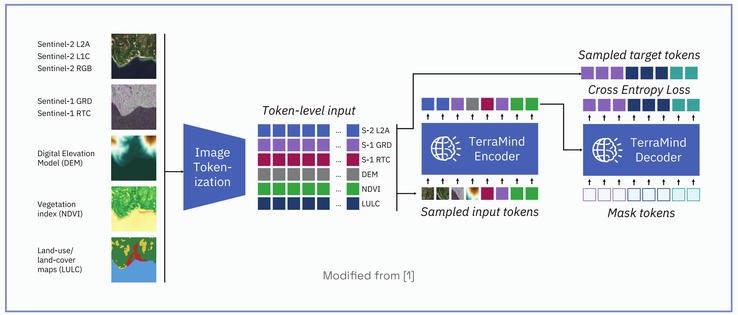
REVOLUTIONISING GEOSPATIAL DATA WITH AI COMPRESSION

Scalable Efficient Compression in Large-Scale Earth Observation

ERIK SCHEURER, JIANGTAO WANG, ROCCO SEDONA, STEFANO MAUROGIOVANNI, BENEDIKT BLUMENSTIEL, JOHANNES JAKUBIK, PAOLO FRACCARO, THOMAS BRUNSCHWILER, STEFAN KESSELHEIM, GABRIELE CAVALLARO

Motivation

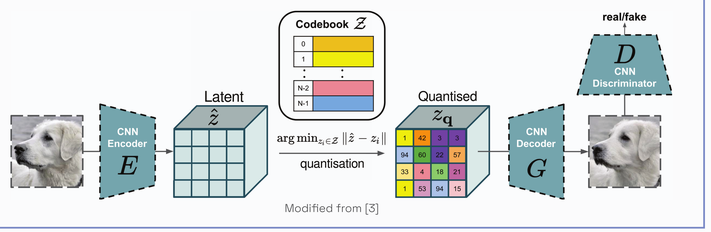
- Continuous EO acquisitions generate petabyte-scale archives
- Larger deep learning models demand more training data and I/O
- Vector-quantisation (VQ) compression stores only codebook indices, reducing size by > 450x
- Transformer backbones like TerraMind [1] already rely on a fixed VQ codebook
- Experiments use MajorTOM-Core dataset [2]



Modified from [1]

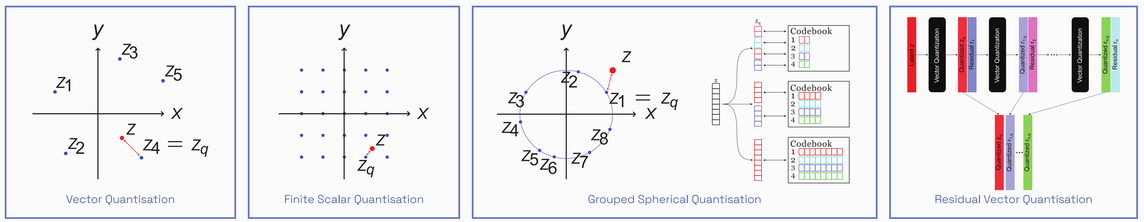
Vector Quantization

- Encoder (VQ-GAN [3]) produces latent representation
- Quantiser assigns each latent vector to its nearest codebook entry
- Decoder reconstructs the image from the quantised embeddings
- An optional discriminator influences training dynamics



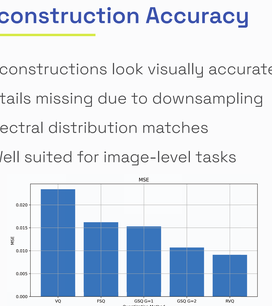
Modified from [3]

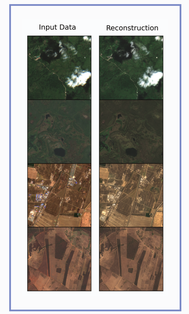
Quantization Methods

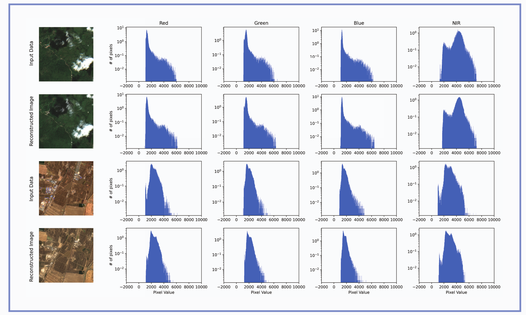


Reconstruction Accuracy

- Reconstructions look visually accurate
- Details missing due to downsampling
- Spectral distribution matches
  - Well suited for image-level tasks

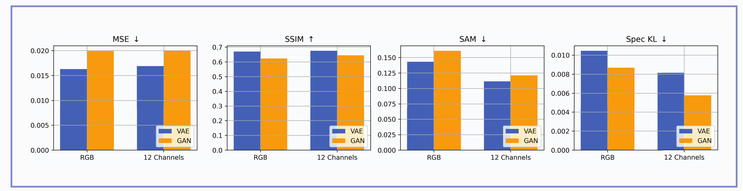






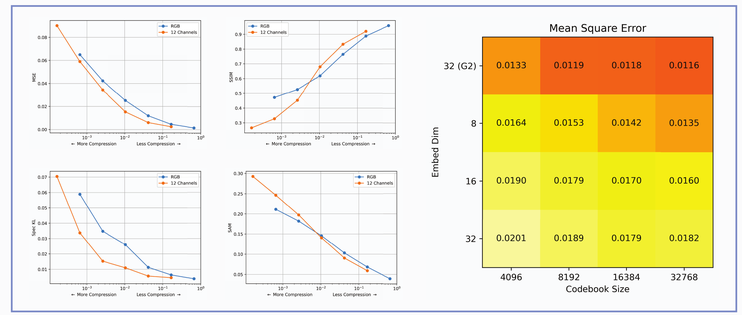
Using a Discriminator

- For non-EO data: Discriminator found to improve visual fidelity of images [4]
- For EO data: Beyond visual aspect, spectra must match, and each pixel should be as close to original as possible



VQ-VAE for Compression

- Additional correlations in multispectral data aids compression
- Larger codebook sizes improve reconstruction fidelity
- Larger embedding dimension underutilised and harder to optimise codebook
  - Decomposed VQ helps utilise large latent space [4]



Bibliography

[1] Jakubik, J., Yang, F., Blumenstiel, B., Scheurer, E., Sedona, R., Maurogiovanni, S., Bosmans, J., Dionelis, N., Marsocci, V., Kopp, N., Ramachandran, R., Fraccaro, P., Brunschwiler, T., Cavallaro, G., Bernabé-Moreno, J., & Long'ep'e, N. (2025). TerraMind: Large-Scale Generative Multimodality for Earth Observation.

[2] Blumenstiel, B., Fraccaro, P., Marsocci, V., Jakubik, J., Maurogiovanni, S., Czerkawski, M., Sedona, R., Cavallaro, G., Brunschwiler, T., Bernabé-Moreno, J., & Long'ep'e, N. (2025). TerraMesh: A Planetary Mosaic of Multimodal Earth Observation Data.

[3] Esser, P., Rombach, R., & Ommer, B. (2021). Taming transformers for high-resolution image synthesis. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 12873–12883.

[4] Wang, J., Qin, Z., Zhang, Y., Hu, V. T., Ommer, B., Briq, R., & Kesselheim, S. (2024). Scaling Image Tokenizers with Grouped Spherical Quantization. arXiv preprint arXiv:2412.02632.

Co-funded by the European Union

UK Research and Innovation

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

Download Poster:

Visit our website: