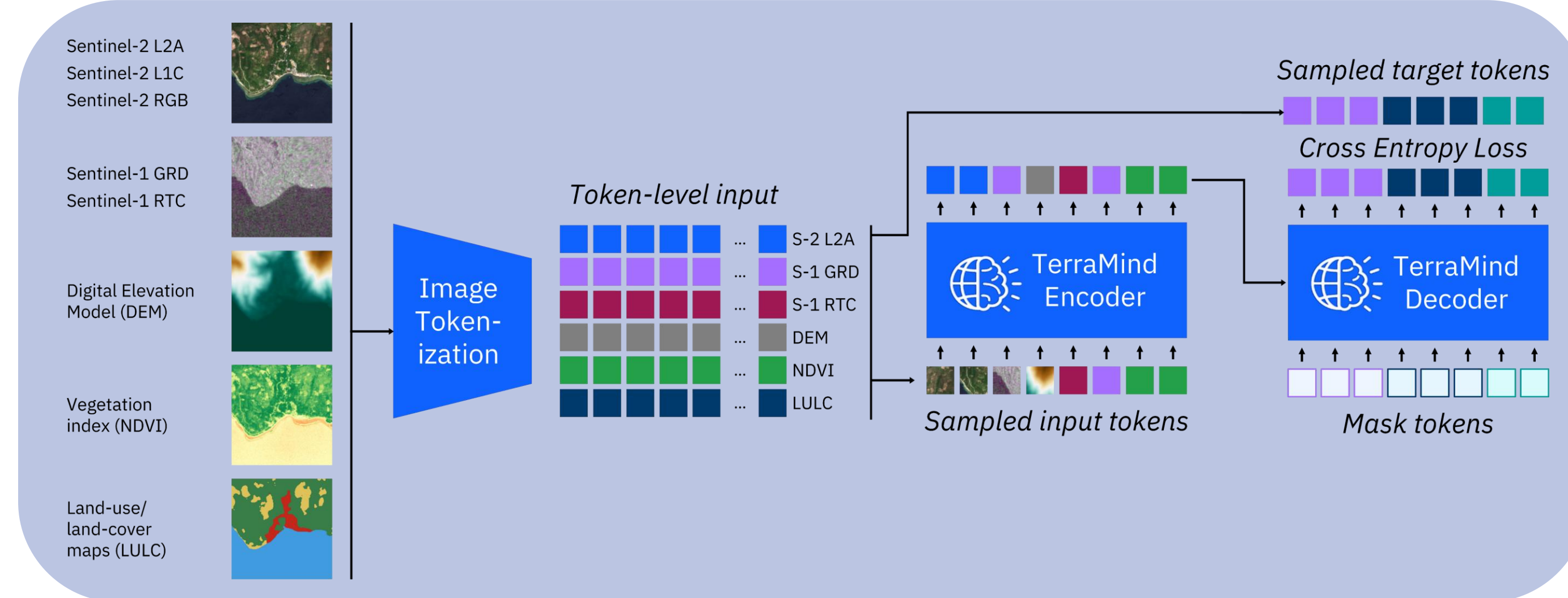


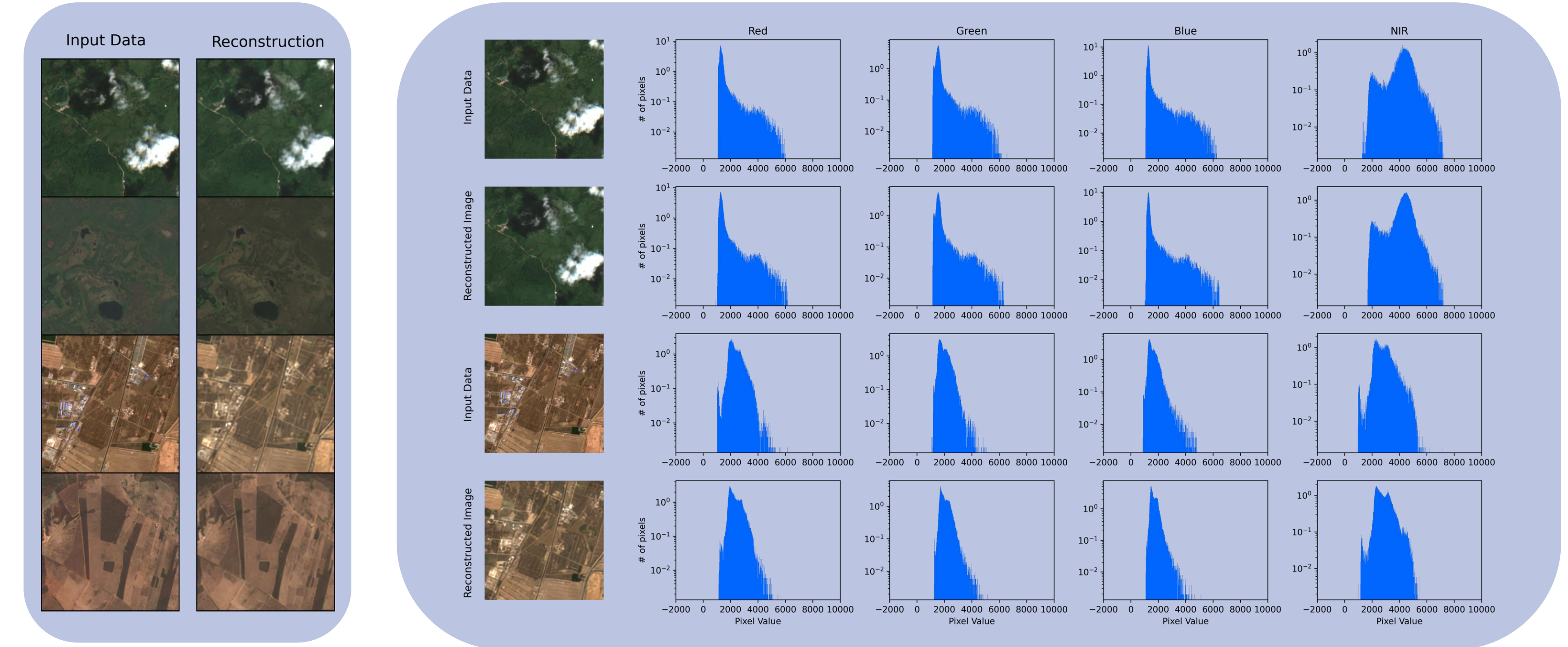
Motivation

- Remote sensing missions generate massive datasets through continuous acquisition.
- Larger models need more training data
- Compression eases transmission and storage
 - Quantization to store only indices not vectors
- Transformer-architectures like TerraMind [1] use fixed tokens from quantization
- MajorTOM-Core dataset from TerraMesh [2]



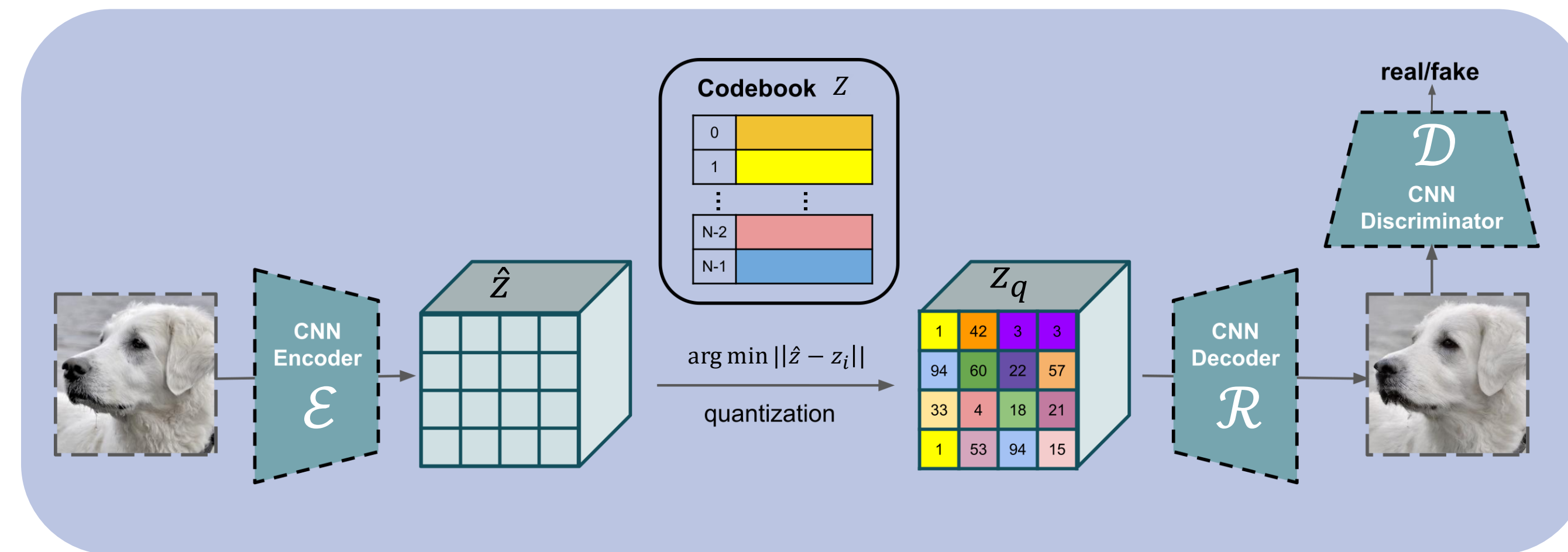
Reconstruction Accuracy

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



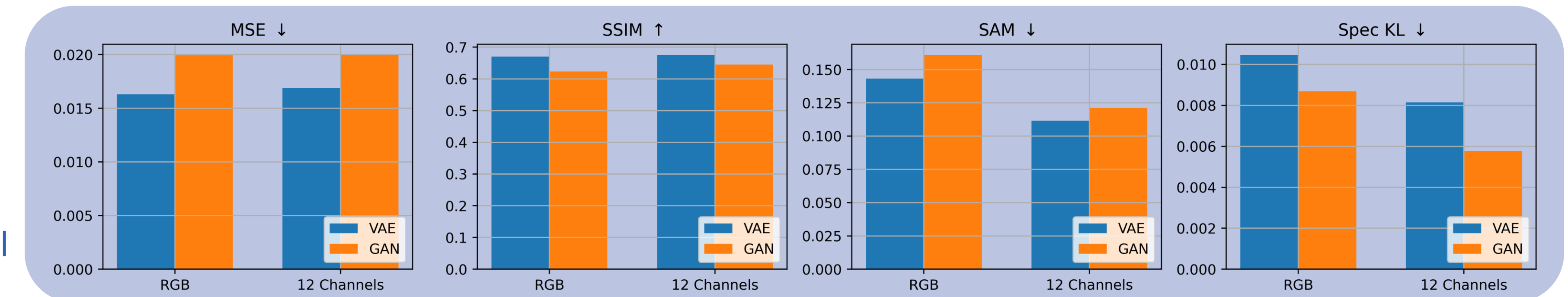
Vector Quantization

- VQ-GAN [3] encodes image into latent vectors
- Quantization maps latent vector to closest codebook entry
- Decoder reconstructs image from quantized embedding
- Optional discriminator to influence training dynamics

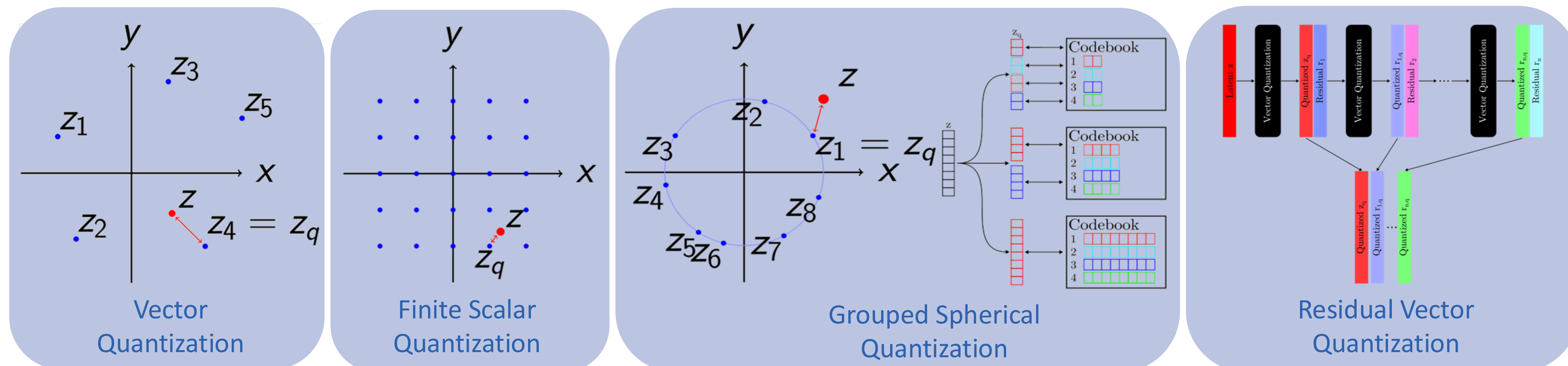


Using a Discriminator

- For non-EO data: Discriminator found to improve visual fidelity of images [4]
- For EO data: Not important how images look: spectrum must match, and each pixel should be as close to original as possible

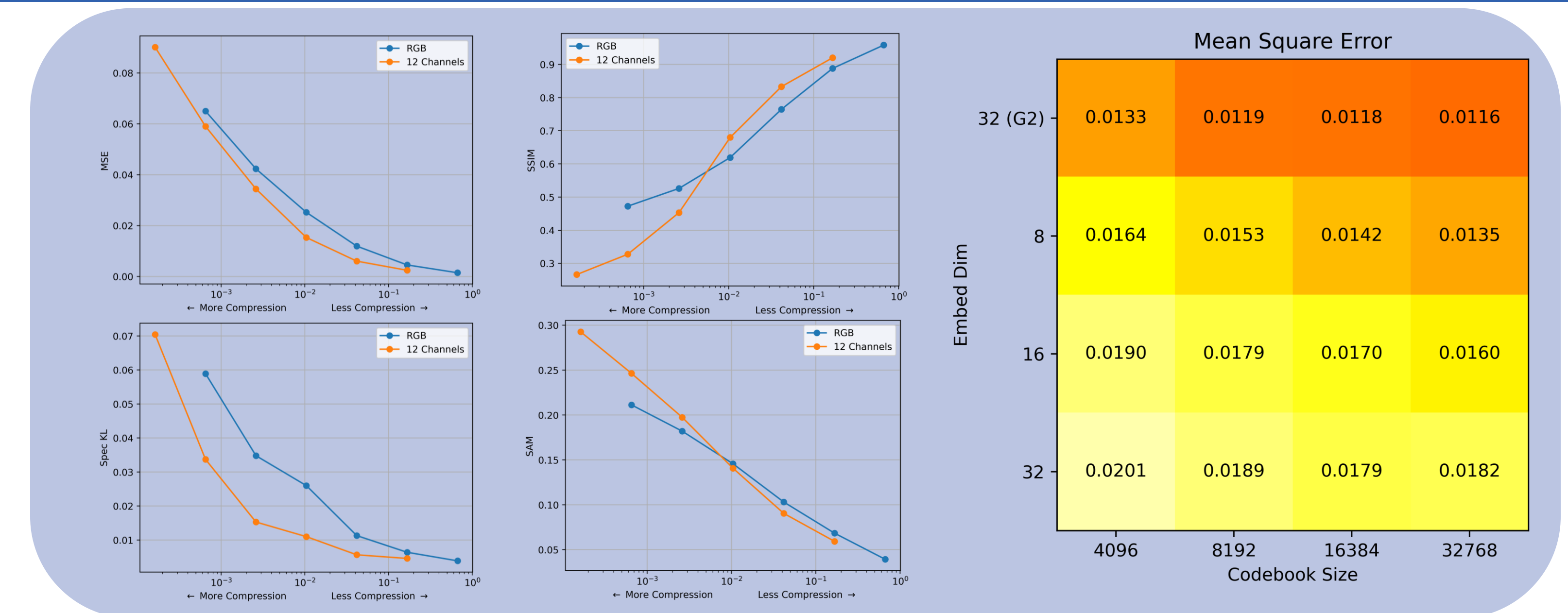


Quantization Methods



VQ-VAE for Compression

- Additional correlations in multispectral data aids compression
- Larger codebook size improve reconstruction fidelity
- Larger embedding dimension underutilized and harder to optimize codebook
 - Decomposed VQ helps utilize 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'epé, N. (2025). TerraMind: A Planetary Mosaic of Multimodal Earth Observation Data.
- [2] Blumenstiel, B., Fraccaro, P., Marsocci, V., Jakubik, J., Maurogiovanni, S., Czerkaski, M., Sedona, R., Cavallaro, G., Brunschwiler, T., Bernabé-Moreno, J., & Long'epé, 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.

