Neural Compression for Multispectral Satellite Images

Woojin Cho*   **Steve Andreas Immanuel***   **Junhyuk Heo**   **Darongsae Kwon**
TelePIX
`{woojin, steve, hjh1037, darong.kwon}@telepix.net`

Abstract

Multispectral satellite images are essential for applications in agriculture, fisheries, and environmental monitoring. However, the high dimensionality, large data volumes, and diverse spatial resolutions across multiple channels present significant challenges for data compression and analysis. In this paper, we introduce ImpliSat, a unified framework specifically designed to address these challenges through efficient compression and reconstruction of multispectral satellite data. ImpliSat employs Implicit Neural Representations (INR) to model satellite images as continuous functions over coordinate space, capturing fine spatial details across varying spatial resolutions. Additionally, we propose a Fourier modulation algorithm that dynamically adjusts to the spectral and spatial characteristics of each channel, ensuring optimal compression while preserving critical image details.

1 Introduction

Satellite data is essential for research and applications such as climate change [1, 2] and marine ecosystem monitoring [3]. In particular, multispectral satellite imagery (MSI) enables detailed analysis of soil conditions [4], vegetation distribution [5, 6], and natural resource management [7], as it contains information collected across various wavelength bands.

In order to collect MSI, we typically deploy a satellite to orbit the earth at a certain altitude and control it from a ground station. During its orbit, a satellite captures information from the earth using broad ranges of sensor and store it in its onboard computer storage which are then sent to the ground station for further processing. However, a satellite can only establish communication with the ground station during its orbital pass, which is a period when the satellite is within a line of sight of the ground station. This can last from a few minutes to several hours depending on the orbit. Therefore, the satellite can only collect limited data to make sure that they can be fully transmitted during this period, otherwise they need to be discarded to make room for new data or wait for the next orbital pass in its cycle to transmit the remaining data. This motivates us to develop a compression algorithm specific for remote sensing image data.

Recent advancements in Implicit Neural Representation (INR) techniques offer a promising solution. INR represents data as coordinate-based functions, enabling the efficient compression of high-resolution imagery. One of the key advantages of INR is its ability to render continuous spatial data at arbitrary resolutions, making it particularly effective for reconstructing data at varying scales. When applied to satellite imagery, INR can handle the resolution differences across spectral bands and the diverse pixel value ranges more efficiently. However, existing INR methods have primarily focused on RGB images [8, 9] and 3D modeling [10, 11], with little research dedicated to multispectral satellite data, which involves varying wavelength bands and differing resolutions across channels.

As illustrated in Appendix E, the frequency spectrum of MSI data varies significantly depending on the spatial resolution, with each resolution displaying distinct spectral characteristics. 10m, 20m and 60m ground sample distance (GSD) images show different frequency characteristics, which suggests

---

*Equal Contribution

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

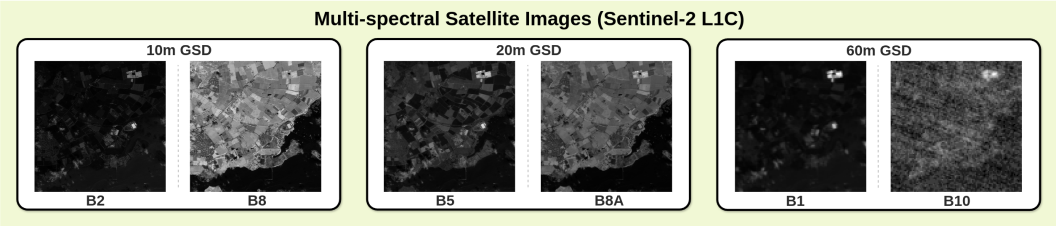


Figure 1: **MSI from the Sentinel-2 L1C dataset (London).** These images show 10m GSD (B2, B8), 20m GSD (B5, B8A), and 60m GSD (B1, B10) bands, illustrating the varying levels of detail captured at different resolutions. More detailed images can be found in the Appendix E and G.

that the Fourier features used to represent the data may benefit from adaptation based on the resolution. Therefore, unlike existing INR approaches that apply fixed Fourier features uniformly across all channels, there is motivation for adaptive modulation that adjusts the Fourier basis according to each band's specific resolution and spectral properties. This approach enables a more accurate and efficient representation of the data.

In this paper, we introduce **Impli**cit Neural Representations for Multispectral **Sat**ellite Images (**ImpliSat**), a novel INR-based compression framework specifically designed for MSI. ImpliSat models MSI data as coordinate-based functions, compressing and reconstructing the data while accounting for the differences in spatial resolution and channel information. Additionally, we propose a new meta-learning approach based on hypernetworks and Fourier modulation to optimize the compression of satellite imagery. This approach addresses the common challenges associated with satellite data, offering a more effective solution for compressing and processing MSI. Our ImpliSat framework addresses this gap by introducing a hypernetwork-based Fourier modulation technique that dynamically generates appropriate Fourier bases tailored to each spectral band's resolution. By conditioning the modulation on the unique resolution and channel information of each band, our approach ensures that each spectral channel is represented optimally, enhancing the efficiency of compression and accuracy of reconstruction.

2 Preliminary: Multispectral Satellite Images and Compression

Unlike standard RGB images, which consist of only three bands capturing visible light in the $400-700nm$ range, MSI encompasses multiple spectral bands, often exceeding ten, which includes non-visible wavelengths such as Near-Infrared (NIR) and Short-Wave Infrared (SWIR). Each spectral band has different pixel value range and covers a specific range of wavelengths which has different GSD depending on the sensor (cf. Figure 1). GSD denotes the actual ground distance between adjacent pixels (10m GSD means 1 pixel in the image represent 1m in actual distance). The bigger the GSD, the lower the spatial resolution, which means there will be fewer details visible in the image. This varying GSD between bands poses a unique challenge for data compression and processing.

Traditional image compression algorithms, such as JPEG [12] and PNG [13], are optimized for uniform resolution images and do not accommodate the complex, multi-resolution structure of MSI. More importantly, they are designed specifically for images with 3 (RGB) and 4 (RGBA) channels (for JPEG and PNG, respectively), thus cannot be applied for MSI. Existing deep learning-based image compression techniques [14] are also not designed to handle the diverse spatial resolutions and pixel value ranges inherent in MSI, leading to limitations in their performance when applied to such complex data as we show later in Section 4.

3 Proposed Methods

To compress and reconstruct MSI effectively, we propose the ImpliSat, which is based on modulated INR frameworks. Our key contribution is Fourier modulation, a novel modulation technique that leverages spectral information and low-rank adaptation [15] to efficiently handle the MSI data. ImpliSat consists of two main components: i) Hypernetwork that generates Fourier modulation vectors based on resolution and channel information, and ii) SIREN-based backbone INR model [8] that uses these modulation vectors to represent the each band of the multispectral image.

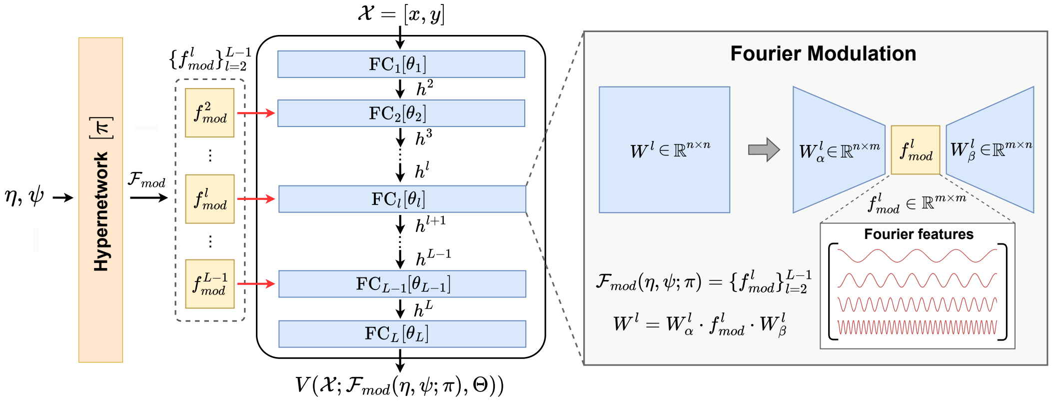


Figure 2: **Overall architecture of ImpliSat.** The left part shows the INR backbone, which takes spatial coordinates as input and is conditioned on resolution (ψ) and channel information (η). The right part demonstrates the Fourier modulation process, which effectively represent the MSI data.

3.1 Hypernetworks and Fourier Modulation

The hypernetwork parameterized by π in ImpliSat is designed to generate Fourier modulation vectors, following the approach introduced in [16], to adapt the INR model to varying resolutions η and channels ψ. Specifically, the hypernetwork takes the following inputs:

- **Resolution information** (η): Represents the spatial resolution of the input MSI, which can take values $\eta \in \{10, 20, 60\}$ (as provided by Sentinel-2 data).
- **Channel information** (ψ): Encoded as a one-hot vector, representing one of the 13 different spectral channels in the MSI data. (See Appendix C).

The Fourier modulation generated by the hypernetwork for each layer is represented as follows:

$$\mathcal{F}_{mod}(\eta, \psi; \pi) = \{f^l_{mod}\}^{L-1}_{l=2}, \qquad f^l_{mod} \in \mathbb{R}^{m\times m}. \tag{1}$$

In Equation 1, f^l_{mod} represents the Fourier modulation vector for l-th layer, with a size of $m \times m$, where $m < n$ is used to apply low-rank adaptation, reducing the dimension of the weights. The modulation vector f^l_{mod} is generated using a *cosine* function, as described in Equation 2.

$$\{f^l_{mod}\}^{L-1}_{l=2} = \{\cos(\Omega^l \odot \mathcal{Z} + \varphi^l)\}^{L-1}_{l=2}, \quad \Omega^l \in \mathbb{R}^{m\times m}, \ \varphi \in \mathbb{R}^m \tag{2}$$

where Ω^l and φ^l are the Fourier frequency matrix and phase matrix for l-th layer, respectively. $\mathcal{Z} \in \mathbb{R}^m$ is a vector sampled from a uniform distribution, and $\odot$ denotes element-wise multiplication. This Fourier modulation technique allows each layer of the backbone INR model to represent the data using Fourier bases adapted to the specific resolution and channel information of the target MSI.

3.2 SIREN-based INR Networks with Fourier-modulated Weights

The main INR network of ImpliSat is a SIREN-based INR model with L layers parameterized by $\Theta = \{\theta_l\}^L_{l=1}$, which takes spatial coordinates $\mathcal{X} \in \mathbb{R}^2$ as input and predicts the corresponding pixel value $V_{gt}(\mathcal{X}, \eta, \psi)$. Each layer of INR consists of fully connected (FC) layers, where each l-th layer is parameterized by $\theta_l = \{W^l, b^l\}$, with W^l as the weight matrix and b^l as the bias term. For example, the hidden state at $(l + 1)$-th layer is computed using sinusoidal activations as $h^{l+1} = \sin(W^l \cdot h^l + b^l)$, where h^l is the hidden state at l-th layer. This sinusoidal activation allow the model to learn high-frequency information, which is crucial for representing complex features in MSI data.

In our ImpliSat framework, to reduce computational cost, we apply low-rank adaptation. Instead of directly using the full-rank weight matrix, we form W^l by multiplying two learnable low-rank matrices, $W^l_\alpha \in \mathbb{R}^{n\times m}$ and $W^l_\beta \in \mathbb{R}^{m\times n}$, with the Fourier modulation vector f^l_{mod}.

$$W^l = W^l_\alpha \cdot f^l_{mod} \cdot W^l_\beta, \qquad W^l_\alpha \in \mathbb{R}^{n\times m}, \ W^l_\beta \in \mathbb{R}^{m\times n}, \ f^l_{mod} \in \mathbb{R}^{m\times m} \tag{3}$$

Table 1: PSNR and MSE comparison for the Fourier modulation method (ours) against baseline methods (shift and scale modulation)

Methods	London		Seoul		Merapi		Hawaii		Cairo	
	PSNR	MSE	PSNR	MSE	PSNR	MSE	PSNR	MSE	PSNR	MSE
Shift	30.252	9.437e-4	28.115	1.543e-3	28.567	1.391e-3	29.418	1.143e-3	29.966	1.008e-3
Scale	29.784	1.051e-3	27.876	1.631e-3	28.177	1.522e-3	29.043	1.247e-3	29.264	1.185e-3
Fourier (Ours)	**36.091**	**2.460e-4**	**33.773**	**4.195e-4**	**32.811**	**5.235e-4**	**35.589**	**2.761e-4**	**36.392**	**2.295e-4**

where, $f^l_{mod} \in \mathbb{R}^{m \times m}$ serves as the modulation vector for the l-th layer, ensuring that the weight matrix W^l is constructed efficiently through low-rank approximations. Further details of the modulation process are provided in Appendix F, which describes how the Fourier modulation vector adjusts the weights at each layer.

3.3 Objective Function

Our ImpliSat is trained by minimizing the reconstruction loss between the ground truth pixel values V_{gt} and the predicted values V_{pred}. The detailed formulation of the loss function ($\mathcal{L}$) is as follows:

$$\mathcal{L} = \frac{1}{\mathcal{N}\mathcal{M}} \sum_{i=1}^{\mathcal{N}} \sum_{j=1}^{\mathcal{M}} ||V_{pred}(\mathcal{X}_{i,j}; \mathcal{F}_{mod}(\eta_j, \psi_j; \pi); \Theta) - V_{gt}(\mathcal{X}_{i,j}, \eta_j, \psi_j)||^2 \tag{4}$$

Here, $\mathcal{N}$ and $\mathcal{M}$ are the number of training samples and number of spectral(or channels) information, respectively. $|| \cdot ||^2$ is the squared error between the ground truth and predicted values. This loss function guides the model to adjust its parameters to reconstruct the MSI, conditioned on the resolution η and channel information ψ.

4 Experiments

In this section, we demonstrate the performance of the ImpliSat framework using diverse MSI data for compression and reconstruction tasks. Additionally, we explore the results of Fourier modulation across various resolutions. For more details on experimental settings, please refer to Appendix A.

4.1 Experimental Settings

Experimental Setups We implement our ImpliSat framework and baselines in PYTHON 3.11.9, PYTORCH 2.3.1 [17], CUDA 12.0, and single NVIDIA RTX 3090 GPU. All models are trained for 10,000 iterations with approximately $200K$ trainable parameters (1MB per model checkpoint, around 10× smaller than the original image). We use the Adam optimizer [18] and early stopping algorithm.

Datasets We compile a dataset of five multispectral Sentinel-2 images, each capturing diverse environments: London (UK), Seoul (South Korea), Merapi (Indonesia), Hawaii (USA), Cairo (Egypt). The images vary in complexity and challenge, each with a size of 9.4MB. Additional details are provided in Appendix A.

4.2 Comparison with Existing Modulated INRs

To demonstrate the performance of ImpliSat, we compare our Fourier modulation approach with existing modulation techniques, specifically shift modulation used in [14, 19, 20] and scale modulation proposed in [14]. Shift modulation operates by learning bias terms for each layer, whereas scale modulation extends this concept by modulating the weights of each layer (cf. Appendix B). Both approaches are evaluated using PSNR and MSE to measure the accuracy of reconstructed images against the ground truth (cf. Appendix D). As shown in Table 1, Fourier

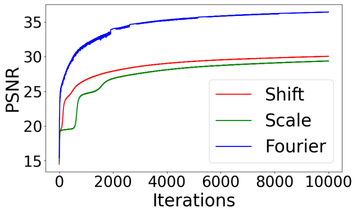
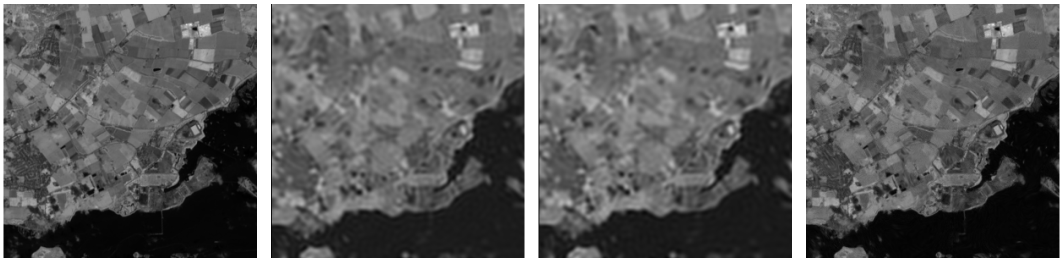
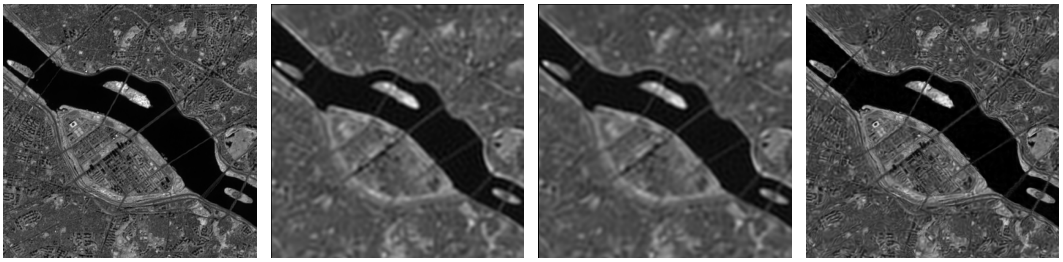


Figure 3: Comparison of PSNR across iterations for Shift, Scale, and Fourier (ours) methods of Cairo image.

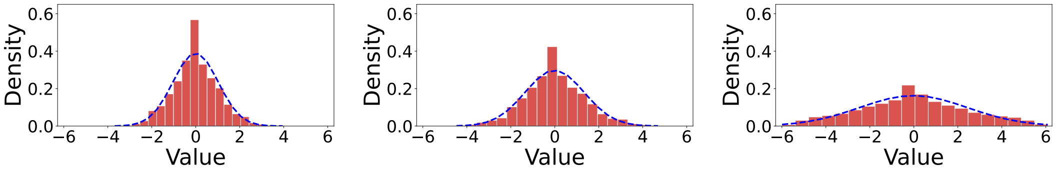


(a) Ground Truth [B7] (b) Shift (28.712) (c) Scale (28.532) (d) **Ours (35.751)**



(e) Ground Truth [B8A] (f) Shift (24.400) (g) Scale (24.195) (h) **Ours (29.994)**

Figure 4: Comparison of ground truth, shift modulation, scale modulation, and our proposed method with PSNR on the London (B7) and Seoul (B8A) image.



(a) 10m GSD (b) 20m GSD (c) 60m GSD

Figure 5: Density distribution of the frequency component of Fourier modulation vectors generated by the hypernetwork for different spatial resolutions: Figure 5(a) 10m GSD (B2), (b) 20m GSD (B7) 60m GSD (B10).

modulation consistently outperforms both shift and scale modulation. Notably, in complex areas such as Seoul, our approach demonstrates a 20% performance improvement over the baseline models. The results in Figure 4 further illustrate that our Fourier modulation captures sharper and more detailed structures in complex urban areas, such as the bridges and buildings visible in the Seoul B8A band. Unlike shift and scale modulation, which tend to blur fine details, our approach preserves clear edges and sharp features, leading to significantly better reconstructions.

Additionally, in Figure 3, our Fourier modulation achieves higher PSNR values compared to both shift and scale modulation throughout the training process on the Cairo dataset. Fourier modulation starts with a rapid increase in PSNR and continues to maintain superior performance as training progresses. While both shift and scale modulation exhibit slower convergence and lower peak PSNR, our approach effectively captures the fine details.

4.3 Analysis on Fourier Modulation Vectors

This section analyzes the behavior of Fourier modulation vectors generated by the hypernetwork across different resolutions. Figure 5 presents histograms depicting the density distribution of the frequency components ($\Omega \odot \mathcal{Z}$) for 10m, 20m, and 60m GSD, and the distribution of Fourier modulation frequencies varies notably with resolution. For 10m GSD, the values are more concentrated near zero, indicating stable modulation with minimal high-frequency emphasis, suitable for preserving fine details in high-resolution data. In contrast, the 60m GSD distribution is broader, suggesting a more variable modulation strategy that may account for the inherently smoother nature of low-resolution imagery. These results suggest that the hypernetwork successfully adjusts the frequency content of the modulation vectors according to the resolution-specific needs of the data. This adaptive capability underscores the flexibility of the Fourier modulation approach, enabling ImpliSat to handle the diverse challenges presented by multi-spectral satellite imagery.

5 Related Work

Machine Learning for Satellite Images Due to the distinctive characteristics of satellite imagery, such as multispectral data, and the high resolution of images, deep learning models have increasingly been employed for tasks like object detection [21], change detection [22], semantic segmentation [23] and data generation [24]. These models have been proven effective in handling the high-dimensional complexity of satellite data. However, there has been limited prior research on satellite image compression, primarily due to the complexity and variability of the data. Compressing satellite images while preserving essential details is challenging because of their high spatial and spectral resolution, which complicates the balance between compression ratio and image quality.

Recent works in neural rendering, such as Sat-NeRF [25] and EO-NeRF [26], apply Neural Radiance Fields (NeRF) techniques to satellite imagery. These models demonstrate the potential of implicit neural representations for capturing fine details in satellite data across different viewpoints Although primarily focused on rendering, these NeRF-based methods highlight the potential for implicit neural representations to revolutionize satellite image processing, including compression.

Implicit Neural Representations for Compression Implicit Neural Representation (INR) fundamentally works by taking coordinates as input and learning a mapping function to the corresponding function values, thereby enabling continuous data rendering [27, 28]. This characteristic allows resolution-invariant inference and makes pre-trained INR models highly effective for data compression [14]. While traditional INR frameworks were limited to representing a single data instance, recent studies have explored meta-learning-based approaches [29, 30] to develop INR frameworks capable of representing multiple datasets simultaneously. Notably, Functa frameworks [19, 20] introduce a highly effective neural compression algorithm by parameterizing datasets through an auto-decoding method [31] combined with a meta-learning-based INR framework. The data compressed into parameters in this manner can serve as input for other neural networks, enabling various downstream tasks without the need for decompression [32, 33]. Building on the strengths of such modulated INR frameworks, we develop a hypernetwork-based modulated INR framework designed to compress multispectral satellite images from various channels.

6 Conclusion

In this paper, we introduce ImpliSat, a novel framework designed for the efficient compression and reconstruction of multispectral satellite imagery using INRs with Fourier modulation. ImpliSat addresses the inherent challenges of MSI data, such as varying spatial resolutions and spectral characteristics, by adaptively modulating the Fourier basis according to each band's specific resolution and channel information. Our experimental results demonstrate that ImpliSat consistently outperforms existing modulation techniques, including shift and scale modulation, particularly in high-resolution and complex environments.

Limitation and Future Works Our current ImpliSat framework is specifically designed on multi-spectral images from the same geographical location. As a result, the pre-trained model struggles to generalize to unseen MSI data from different regions. This limitation restricts its broader applicability for global-scale satellite image analysis. In future work, we plan to extend ImpliSat by integrating auto-decoding techniques with meta-learning algorithms, enabling the model to handle MSI data from multiple locations simultaneously. This adaptation would allow the framework to generalize better across varying geographical regions and datasets. Additionally, we aim to develop more lightweight versions of the model and explore hardware acceleration techniques to facilitate real-time processing, which would significantly enhance the practical utility of ImpliSat in resource constrained environments such as onboard satellite systems.

References

[1] Jun Yang, Peng Gong, Rong Fu, Minghua Zhang, Jingming Chen, Shunlin Liang, Bing Xu, Jiancheng Shi, and Robert Dickinson. The role of satellite remote sensing in climate change studies. *Nature climate change*, 3(10):875–883, 2013.

[2] Joel R Norris, Robert J Allen, Amato T Evan, Mark D Zelinka, Christopher W O'Dell, and Stephen A Klein. Evidence for climate change in the satellite cloud record. *Nature*, 536(7614):72–75, 2016.

[3] Jeffrey G Masek. Stability of boreal forest stands during recent climate change: evidence from landsat satellite imagery. *Journal of biogeography*, 28(8):967–976, 2001.

[4] Leila Hassan-Esfahani, Alfonso Torres-Rua, Austin Jensen, and Mac McKee. Assessment of surface soil moisture using high-resolution multi-spectral imagery and artificial neural networks. *Remote Sensing*, 7(3):2627–2646, 2015.

[5] Edward Dwyer, S Pinnock, J-M Grégoire, and JMC Pereira. Global spatial and temporal distribution of vegetation fire as determined from satellite observations. *International Journal of Remote Sensing*, 21(6-7):1289–1302, 2000.

[6] Ronghua Ma, Hongtao Duan, Xiaohong Gu, and Shouxuan Zhang. Detecting aquatic vegetation changes in taihu lake, china using multi-temporal satellite imagery. *Sensors*, 8(6):3988–4005, 2008.

[7] Nathalie Pettorelli. *Satellite remote sensing and the management of natural resources*. Oxford University Press, 2019.

[8] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.

[9] Yannick Strümpler, Janis Postels, Ren Yang, Luc Van Gool, and Federico Tombari. Implicit neural representations for image compression. In *ECCV*, 2022.

[10] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.

[11] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10318–10327, 2021.

[12] Gregory K Wallace. The jpeg still picture compression standard. *Communications of the ACM*, 34(4):30–44, 1992.

[13] Thomas Boutell. *PNG: The Definitive Guide*. O'Reilly Media, Inc., 1997.

[14] Emilien Dupont, Hrushikesh Loya, Milad Alizadeh, Adam Goli'nski, Yee Whye Teh, and A. Doucet. Coin++: Neural compression across modalities. *TMLR*, 2022.

[15] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

[16] Kexuan Shi, Xingyu Zhou, and Shuhang Gu. Improved implicit neural representation with fourier reparameterized training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25985–25994, 2024.

[17] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[18] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, 2014.

[19] Emilien Dupont, Hyunjik Kim, SM Eslami, Danilo Rezende, and Dan Rosenbaum. From data to functa: Your data point is a function and you can treat it like one. In *ICML*, 2022.

[20] Matthias Bauer, Emilien Dupont, Andy Brock, Dan Rosenbaum, Jonathan Richard Schwarz, and Hyunjik Kim. Spatial functa: Scaling functa to imagenet classification and generation. *arXiv preprint arXiv:2302.03130*, 2023.

[21] Di Wang, Qiming Zhang, Yufei Xu, Jing Zhang, Bo Du, Dacheng Tao, and Liangpei Zhang. Advancing plain vision transformer toward remote sensing foundation model. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2022.

[22] Lazhar Khelifi and Max Mignotte. Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis. *Ieee Access*, 8:126385–126400, 2020.

[23] Nassim Ait Ali Braham, Conrad M Albrecht, Julien Mairal, Jocelyn Chanussot, Yi Wang, and Xiao Xiang Zhu. Spectralearth: Training hyperspectral foundation models at scale. *arXiv preprint arXiv:2408.08447*, 2024.

[24] Samar Khanna, Patrick Liu, Linqi Zhou, Chenlin Meng, Robin Rombach, Marshall Burke, David B Lobell, and Stefano Ermon. Diffusionsat: A generative foundation model for satellite imagery. In *The Twelfth International Conference on Learning Representations*, 2023.

[25] Roger Marí, Gabriele Facciolo, and Thibaud Ehret. Sat-nerf: Learning multi-view satellite photogrammetry with transient objects and shadow modeling using rpc cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1311–1321, 2022.

[26] Roger Marí, Gabriele Facciolo, and Thibaud Ehret. Multi-date earth observation nerf: The detail is in the shadows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2035–2045, 2023.

[27] Kyeong-Joong Jeong and Yong-Min Shin. Time-series anomaly detection with implicit neural representation. *arXiv preprint arXiv:2201.11950*, 2022.

[28] Zeyuan Chen, Yinbo Chen, Jingwen Liu, Xingqian Xu, Vidit Goel, Zhangyang Wang, Humphrey Shi, and Xiaolong Wang. Videoinr: Learning video implicit neural representation for continuous space-time super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2047–2057, 2022.

[29] Woojin Cho, Kookjin Lee, Donsub Rim, and Noseong Park. Hypernetwork-based meta-learning for low-rank physics-informed neural networks. *Advances in Neural Information Processing Systems*, 36, 2024.

[30] Woojin Cho, Minju Jo, Haksoo Lim, Kookjin Lee, Dongeun Lee, Sanghyun Hong, and Noseong Park. Parameterized physics-informed neural networks for parameterized pdes. *arXiv preprint arXiv:2408.09446*, 2024.

[31] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019.

[32] Luca De Luigi, Adriano Cardace, Riccardo Spezialetti, Pierluigi Zama Ramirez, Samuele Salti, and Luigi Di Stefano. Deep learning on implicit neural representations of shapes. *arXiv preprint arXiv:2302.05438*, 2023.

[33] Ziya Erkoç, Fangchang Ma, Qi Shan, Matthias Nießner, and Angela Dai. Hyperdiffusion: Generating implicit neural fields with weight-space diffusion. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 14300–14310, 2023.

A Detailed Information of Benchmark Dataset

We collected five different MSI of Sentinel-2 satellite with L1C processing level provided by European Space Agency (ESA) through Copernicus under Open Access compliant Creative Commons CC BY-SA 3.0 IGO licence. Each MSI has a resolution of $13 \times 600 \times 600$, denoting 13 different multispectral bands.

The Table 2 provides the geographical coordinates (Longitude and Latitude) of the benchmark data used in our experiments. Each location was chosen to evaluate the ability of ImpliSat framework to compress a variety MSI data.

Table 2: Geographic Coordinates of Benchmark Dataset

	London	**Seoul**	**Merapi**	**Hawaii**	**Cairo**
Latitude	51°26’11"N	37°31’30"N	07°32’29"N	19°31’12"N	30°01’29"N
Longitude	00°34’55"E	126°55’36"E	110°26’46"E	154°47’08"W	31°55’18"E
Environments	Urban	Urban	Volcano	Marine	Desert

Location Descriptions:

- **London** (United Kingdom): A densely populated urban area known for its diverse land cover, including built-up regions, parks, and water bodies. It provides a complex environment to test urban monitoring capabilities of the framework.
- **Seoul** (South Korea): A highly complex city characterized by a mix of high-rise buildings, infrastructure. The big river runs through the city, adding a unique urban water feature that further complicates the environments.
- **Merapi** (Indonesia): A volcanic region with active geological features. Merapi provides a challenging environment with diverse vegetation cover, useful for testing the model performance in monitoring volcanic landscape.
- **Hawaii** (USA): A tropical archipelago characterized by coastal zones, and diverse ecosystems. This location is ideal for evaluating multi-spectral imagery performance in coastal and marine.
- **Cairo** (Egypt): Located in a desert region with minimal vegetation, Cairo represents arid conditions. This site is significant for testing the framework’s capability in soil and water content analysis under extreme conditions.

B Description of modulated INRs

In this section, we describe the modulation techniques used in INR models, specifically shift and scale modulation.

Shift Modulation Shift modulation, introduced in [14], involves adding a learnable bias term to the output of each SIREN layer. The shift modulation is defined as:

$$h_{l+1} = \sin(W_l \cdot h_l + b_l + \mu_l). \tag{5}$$

where μ represents the modulation term specifically added to adjust the bias at each layer.

Scale Modulation Scale modulation extends the shift modulation by scaling the output of each SIREN layer. The scale modulation can be expressed as:

$$h_{l+1} = \sin(\kappa \odot (W_l \cdot h_l + b_l)). \tag{6}$$

where κ is the modulation vector that scales the output of the layer.

C Multispectral Satellite Imagery Bands

Table 3: Multispectral Bands in Sentinel-2 Satellite Imagery

Band Number	Band Description	Wavelength Range (nm)	GSD (m)	Usage
B1	Coastal aerosol	433–453	60	Atmospheric correction
B2	Blue	458–523	10	Water body analysis
B3	Green	543–578	10	Vegetation monitoring
B4	Red	650–680	10	Vegetation and soil analysis
B5	Red-edge 1	698–713	20	Vegetation monitoring
B6	Red-edge 2	733–748	20	Vegetation structure analysis
B7	Red-edge	773–793	20	Vegetation chlorophyll assessment
B8	NIR	785–900	10	Biomass and vegetation vigor monitoring
B8A	Narrow-NIR	855–875	20	Vegetation monitoring
B9	Water vapour	935–955	60	Cloud detection and water vapor analysis
B10	SWIR-Cirrus	1360–1390	60	Atmospheric correction and cirrus cloud detection
B11	SWIR	1565–1655	20	Cloud and snow monitoring
B12	SWIR	2100–2280	20	Soil and water content analysis

The Table 3 provides a summary of the multi-spectral bands provided by Sentinel-2, including their band number, wavelength ranges, GSD, and usages. The MSI data spans a wide range of wavelengths, covering both visible and non-visible spectrums. Each band provides unique information, making it suitable for various applications such as atmospheric correction, vegetation monitoring, and soil analysis.

D Metrics: PSNR Calculation

Because each band in MSI has different pixel value range, before calculating the PSNR, we apply min-max normalization for each band independently, given by:

$$\mathbf{I'}_i = \frac{\mathbf{I}_i - min(\mathbf{I}_i)}{max(\mathbf{I}_i) - min(\mathbf{I}_i)}, \tag{7}$$

where $\mathbf{I}_i$ denotes the i-th band of the MSI. Given a ground truth MSI V_{gt} and prediction V_{pred}, we calculate the MSE for i-th band as follows:

$$\mathrm{MSE}_i = ||V_{gt,i} - V_{pred,i}||^2, \tag{8}$$

After that we average the MSE across all $\#\mathcal{N}_c$ bands ($\mathcal{N}_c$: Number of bands),

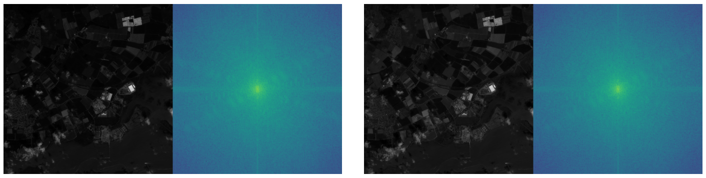
$$\overline{\mathrm{MSE}} = \frac{1}{\mathcal{N}_c} \sum_{i=1}^{\mathcal{N}_c} \mathrm{MSE}_i, \tag{9}$$

Finally, we calculate the PSNR for the whole bands using the following formula:

$$\mathrm{PSNR} = -10 \log \overline{\mathrm{MSE}} \tag{10}$$

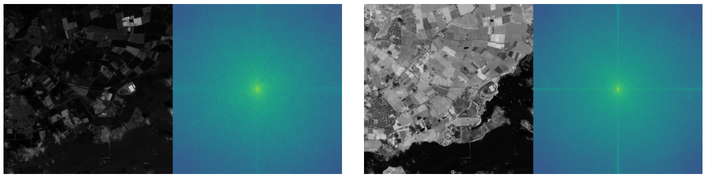
Alternatively, we can calculate the independent PSNR for each band by applying Equation 10 to MSE_i directly. Note that the results that we presented in Table 1 in the main paper are the PSNR for all the bands as reflected in Equation 10, while the results in Appendix G are the PSNR for each band independently.

E Fourier Spectrum of Multispectral Images



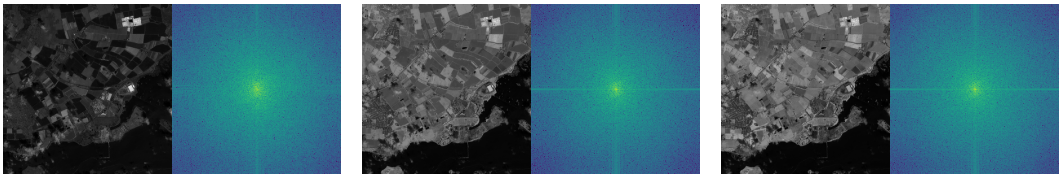
(a) B2 (10m GSD)

(b) B3 (10m GSD)



(c) B4 (10m GSD)

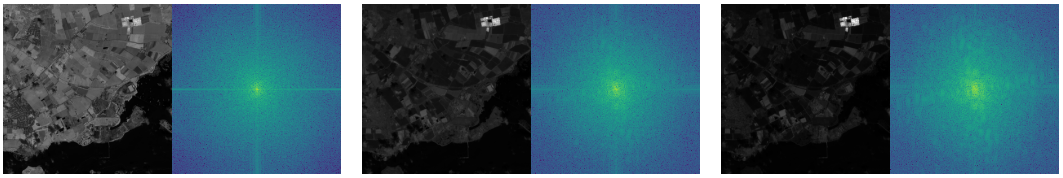
(d) B8 (10m GSD)



(e) B5 (20m GSD)

(f) B6 (20m GSD)

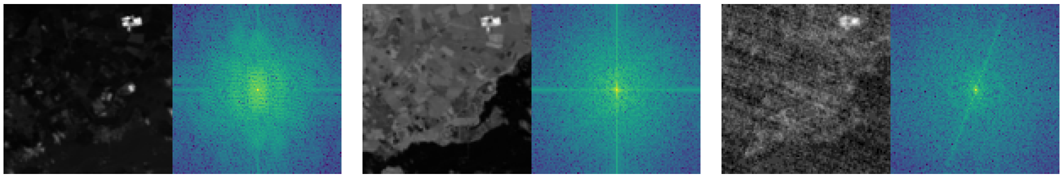
(g) B7 (20m GSD)



(h) B8A (20m GSD)

(i) B11 (20m GSD)

(j) B12 (20m GSD)



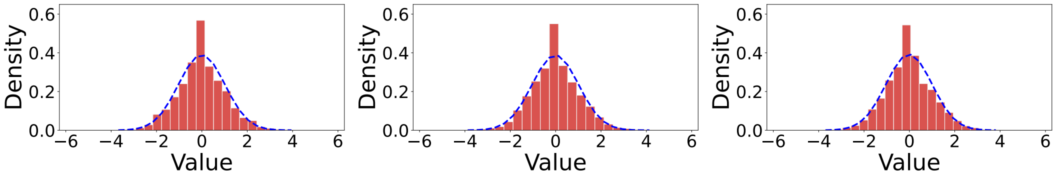
(k) B1 (60m GSD)

(l) B9 (60m GSD)

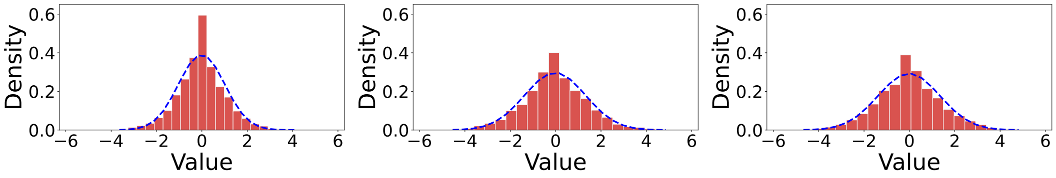
(m) B10 (60m GSD)

Figure 6: **The Fourier spectrum results for multispectral image(London).** First two rows from top represent 10m GSD bands (B2, B3, B4, B8), followed by six images represent 20m GSD bands (B5, B6, B7, B8A, B11, B12). The final row depicts 60m GSD bands (B1, B9, B10). In each image pair, the left panel represents the ground truth, while the right panel displays the result after applying Fast Fourier Transform (FFT).

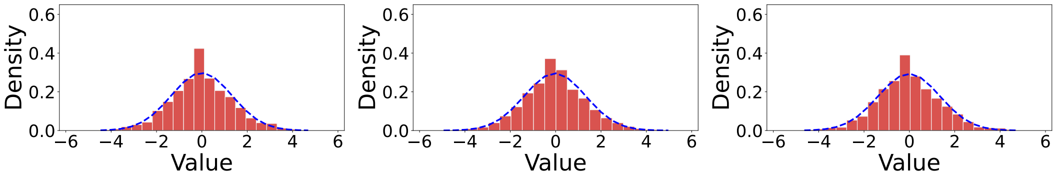
F Distribution of Fourier Modulation Vectors



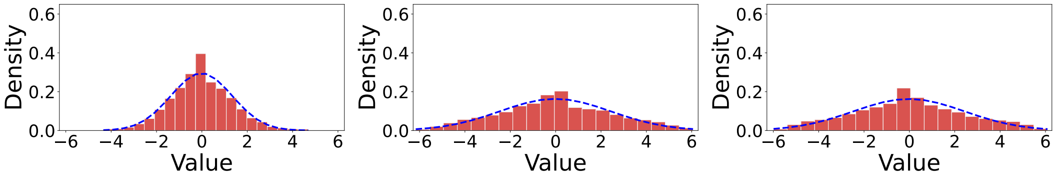
(a) Band 2 (10m GSD) (b) Band 3 (10m GSD) (c) Band 4 (10m GSD)



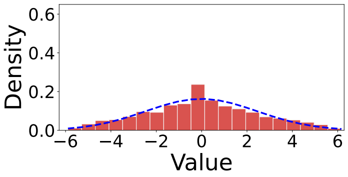
(d) Band 8 (10m GSD) (e) Band 5 (20m GSD) (f) Band 6 (20m GSD)



(g) Band 7 (20m GSD) (h) Band 8A (20m GSD) (i) Band 11 (20m GSD)



(j) Band 12 (20m GSD) (k) Band 1 (60m GSD) (l) Band 9 (60m GSD)



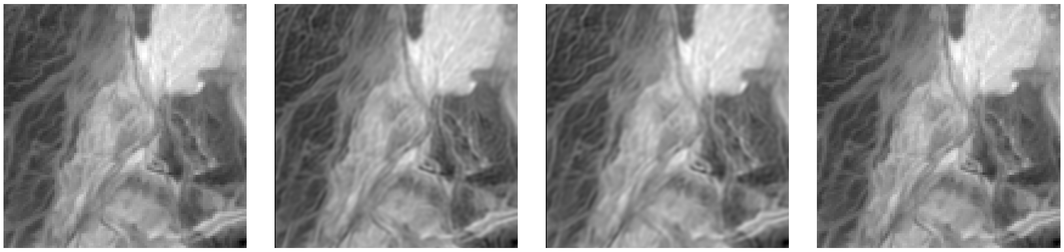
(m) Band 10 (60m GSD)

Figure 7: **The probability density function (PDF) of the Fourier modulation vector.** Red histogram represents PDF of Fourier modulation vector which are the output value of hypernetwork. To provide more precise analysis, we illustrate the normalized distribution using a blue dashed line.

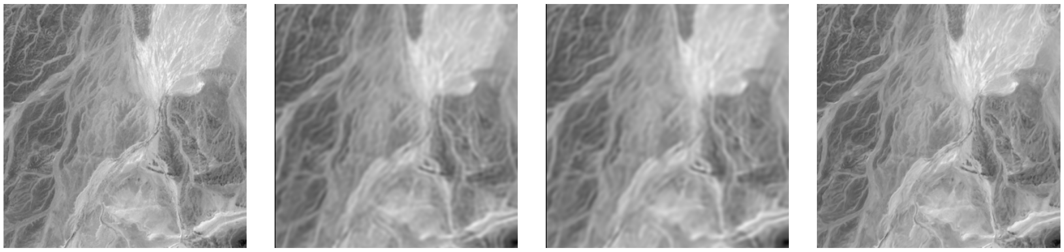
G Additional Experiment Results

In this section, we present a more detailed experiment results for all datasets. ‘Shift’ denotes the shift modulation and ‘Scale’ denotes the scale modulation. Explanations of each modulation can be found in Appendix B.

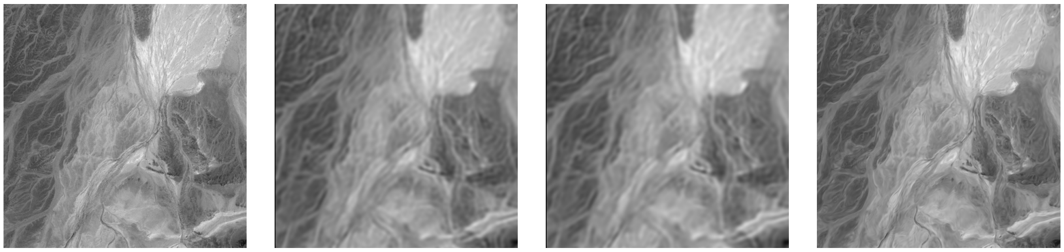
G.1 Cairo



(a) Ground Truth [B1] (b) Shift (33.372) (c) Scale (32.886) (d) **Ours (52.833)**



(e) Ground Truth [B5] (f) Shift (31.002) (g) Scale (30.330) (h) **Ours (36.871)**



(i) Ground Truth [B2] (j) Shift (29.257) (k) Scale (28.777) (l) **Ours (32.124)**

Figure 8: Comparison of ground truth, shift modulation, scale modulation, and our proposed method with PSNR across multibands (B1, B5, B2) of the Cairo image. Each row represents a different GSD: 60m, 20m, and 10m, respectively.

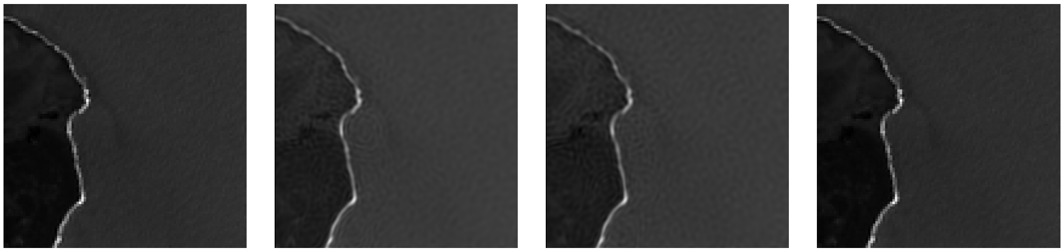
Table 4: Detailed PSNR for each band of Cairo image.

Methods	B1	B2	B3	B4	B5	B6	B7	B8	B8A	B9	B10	B11	B12
Shift	33.372	29.257	29.841	30.258	31.002	31.399	31.517	31.042	31.946	34.259	23.242	34.830	34.167
Scale	32.886	28.777	29.209	29.779	30.330	30.807	30.949	30.513	31.324	33.754	22.331	34.042	33.356
Fourier (Ours)	**52.833**	**32.124**	**32.877**	**33.289**	**36.871**	**37.170**	**37.213**	**34.051**	**37.315**	**53.009**	**56.384**	**40.183**	**39.541**

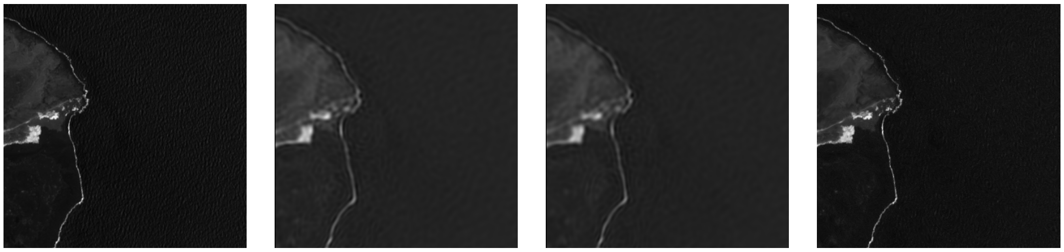
Table 5: Detailed MSE for each band of Cairo image.

Methods	B1	B2	B3	B4	B5	B6	B7	B8	B8A	B9	B10	B11	B12
Shift	4.601e-4	1.186e-3	1.037e-3	9.422e-4	7.940e-4	7.246e-4	7.051e-4	7.866e-4	6.388e-4	3.751e-4	4.740e-3	3.289e-4	3.831e-4
Scale	5.145e-4	1.325e-3	1.200e-3	1.052e-3	9.268e-4	8.304e-4	8.038e-4	8.885e-4	7.373e-4	4.213e-4	5.846e-3	3.942e-4	4.618e-4
Fourier (Ours)	**5.209e-6**	**6.132e-4**	**5.156e-4**	**4.689e-4**	**2.055e-4**	**1.919e-4**	**1.900e-4**	**3.935e-4**	**1.856e-4**	**5.001e-6**	**2.299e-6**	**9.587e-5**	**1.111e-4**

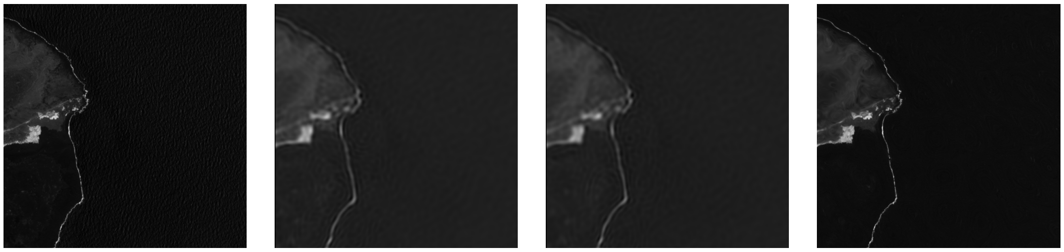
G.2 Hawaii



(a) Ground Truth [B1] (b) Shift (33.754) (c) Scale (33.438) (d) **Ours (48.434)**



(e) Ground Truth [B7] (f) Shift (30.933) (g) Scale (30.654) (h) **Ours (35.241)**



(i) Ground Truth [B8] (j) Shift (30.743) (k) Scale (30.531) (l) **Ours (33.338)**

Figure 9: Comparison of ground truth, shift modulation, scale modulation, and our proposed method with PSNR across multibands (B1, B7, B8) of the Hawaii image. Each row represents a different GSD: 60m, 20m, and 10m, respectively.

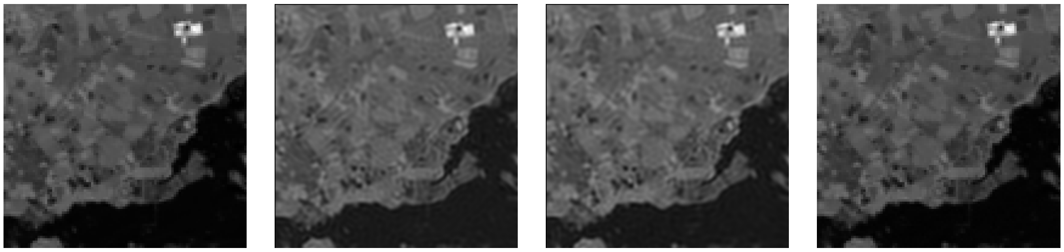
Table 6: Detailed PSNR for each band of Hawaii image.

Methods	**B1**	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	**B8**	**B8A**	**B9**	**B10**	**B11**	**B12**
Shift	33.754	31.535	31.681	31.530	32.341	31.978	30.933	30.743	31.792	36.075	21.958	30.649	28.697
Scale	33.438	31.282	31.485	31.330	31.947	31.730	30.654	30.531	31.521	35.937	21.405	30.447	28.516
Fourier (Ours)	**48.434**	**35.283**	**35.276**	**34.489**	**36.385**	**35.647**	**35.241**	**33.338**	**35.698**	**47.267**	**52.140**	**33.518**	**32.063**

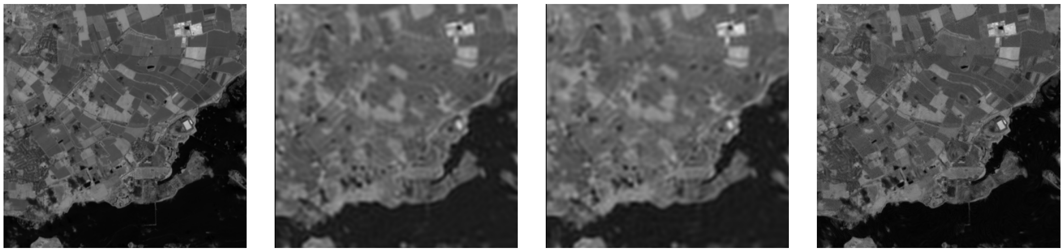
Table 7: Detailed MSE for each band of Hawaii image.

Methods	**B1**	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	**B8**	**B8A**	**B9**	**B10**	**B11**	**B12**
Shift	4.213e-4	7.023e-4	6.791e-4	7.031e-4	5.834e-4	6.342e-4	8.067e-4	8.427e-4	6.619e-4	2.469e-4	6.371e-3	8.611e-4	1.350e-3
Scale	4.531e-4	7.444e-4	7.104e-4	7.362e-4	6.387e-4	6.715e-4	8.602e-4	8.848e-4	7.046e-4	2.549e-4	7.237e-3	9.021e-4	1.407e-3
Fourier (Ours)	**1.434e-5**	**2.963e-4**	**2.968e-4**	**3.558e-4**	**2.299e-4**	**2.725e-4**	**2.991e-4**	**4.637e-4**	**2.693e-4**	**1.876e-5**	**6.110e-6**	**4.448e-4**	**6.219e-4**

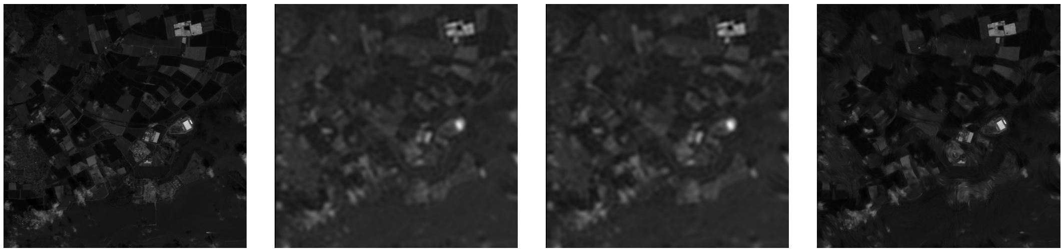
G.3 London



(a) Ground Truth [B9] (b) Shift (33.490) (c) Scale (33.015) (d) **Ours (50.814)**



(e) Ground Truth [B6] (f) Shift (29.249) (g) Scale (28.768) (h) **Ours (35.876)**



(i) Ground Truth [B3] (j) Shift (32.516) (k) Scale (32.074) (l) **Ours (35.144)**

Figure 10: Comparison of ground truth, shift modulation, scale modulation, and our proposed method with PSNR across multibands (B9, B6, B3) of the London image. Each row represents a different GSD: 60m, 20m, and 10m, respectively.

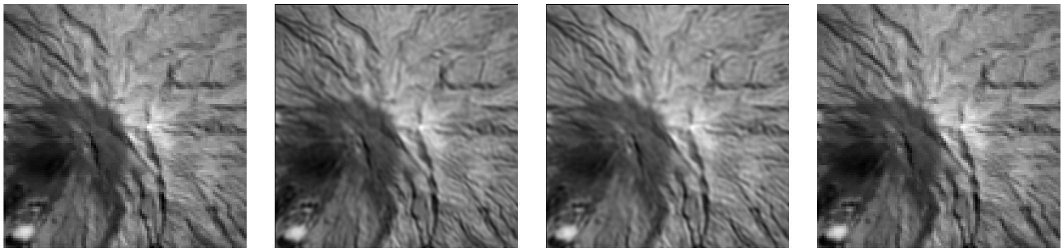
Table 8: Detailed PSNR for each band of London image.

Methods	**B1**	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	**B8**	**B8A**	**B9**	**B10**	**B11**	**B12**
Shift	35.323	32.849	32.516	30.831	31.257	29.249	28.718	25.133	30.326	33.490	27.604	35.088	35.216
Scale	32.774	32.260	32.074	30.280	30.531	28.768	28.532	24.965	30.018	33.015	27.006	34.586	34.640
Fourier (Ours)	**50.365**	**35.660**	**35.144**	**34.138**	**37.312**	**35.876**	**35.751**	**29.639**	**36.329**	**50.814**	**49.715**	**40.061**	**41.087**

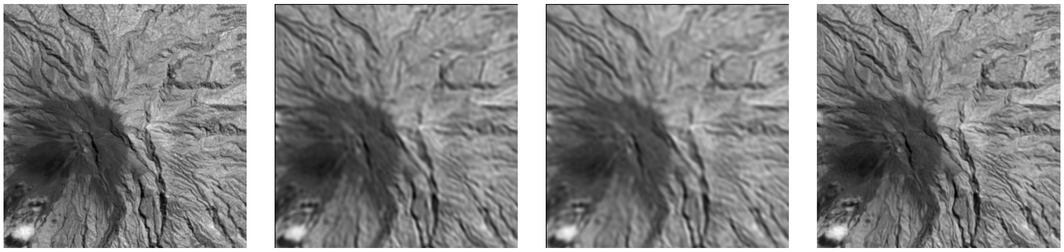
Table 9: Detailed MSE for each band of London image.

Methods	**B1**	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	**B8**	**B8A**	**B9**	**B10**	**B11**	**B12**
Shift	2.936e-4	5.189e-4	5.603e-4	8.258e-4	7.487e-4	1.189e-3	1.344e-3	3.067e-3	9.276e-4	4.477e-4	1.736e-3	3.099e-4	3.009e-4
Scale	5.280e-4	5.943e-4	6.202e-4	9.376e-4	8.849e-4	1.328e-3	1.402e-3	3.188e-3	9.959e-4	4.995e-4	1.993e-3	3.479e-4	3.435e-4
Fourier (Ours)	**9.193e-6**	**2.717e-4**	**3.059e-4**	**3.857e-4**	**1.857e-4**	**2.584e-4**	**2.660e-4**	**1.087e-3**	**2.328e-4**	**8.291e-6**	**1.068e-5**	**9.861e-5**	**7.786e-5**

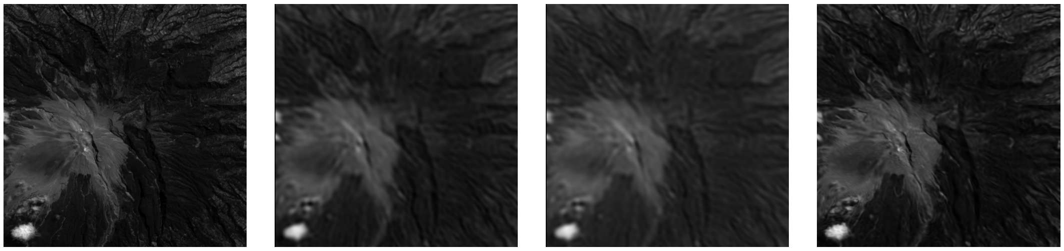
G.4 Merapi



(a) Ground Truth [B9] (b) Shift (29.758) (c) Scale (29.562) (d) **Ours (45.940)**



(e) Ground Truth [B8A] (f) Shift (25.749) (g) Scale (25.356) (h) **Ours (31.018)**



(i) Ground Truth [B4] (j) Shift (30.183) (k) Scale (29.634) (l) **Ours (32.650)**

Figure 11: Comparison of ground truth, shift modulation, scale modulation, and our proposed method with PSNR across multibands (B9, B8A, B4) of the Merapi image. Each row represents a different GSD: 60m, 20m, and 10m, respectively.

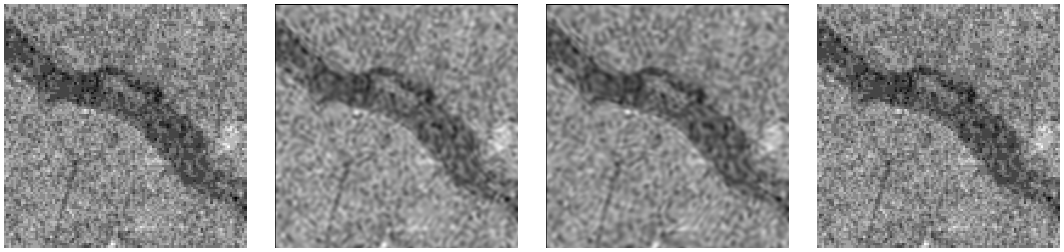
Table 10: Detailed PSNR for each band of Merapi image.

Methods	**B1**	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	**B8**	**B8A**	**B9**	**B10**	**B11**	**B12**
Shift	36.823	33.036	31.401	30.183	30.343	26.510	25.803	23.690	25.749	29.758	31.593	29.899	37.073
Scale	36.270	32.682	31.059	29.634	29.882	26.109	25.445	23.418	25.356	29.562	30.760	29.428	35.975
Fourier (Ours)	**49.430**	**35.216**	**33.398**	**32.650**	**33.722**	**31.511**	**30.852**	**26.448**	**31.018**	**45.940**	**45.071**	**34.321**	**42.444**

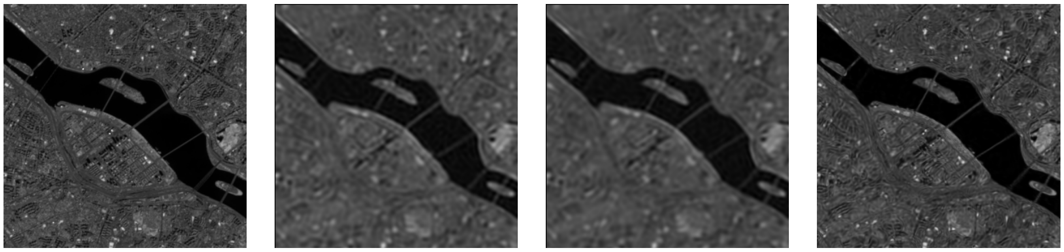
Table 11: Detailed MSE for each band of Merapi image.

Methods	**B1**	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	**B8**	**B8A**	**B9**	**B10**	**B11**	**B12**
Shift	2.078e-4	4.971e-4	7.243e-4	9.587e-4	9.241e-4	2.234e-3	2.629e-3	4.276e-3	2.662e-3	1.057e-3	6.929e-4	1.024e-3	1.962e-4
Scale	2.360e-4	5.392e-4	7.837e-4	1.088e-3	1.027e-3	2.450e-3	2.855e-3	4.552e-3	2.914e-3	1.106e-3	8.394e-4	1.141e-3	2.526e-4
Fourier (Ours)	**1.140e-5**	**3.009e-4**	**4.573e-4**	**5.433e-4**	**4.244e-4**	**7.061e-4**	**8.220e-4**	**2.266e-3**	**7.910e-4**	**2.547e-5**	**3.111e-5**	**3.697e-4**	**5.697e-5**

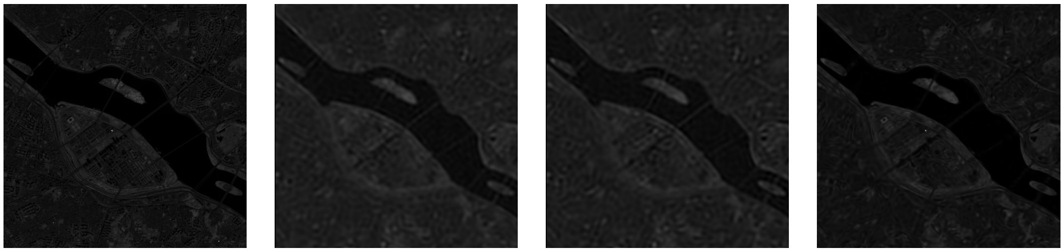
G.5 Seoul



(a) Ground Truth [B10] (b) Shift (24.631) (c) Scale (24.326) (d) **Ours (51.222)**



(e) Ground Truth [B11] (f) Shift (27.047) (g) Scale (26.748) (h) **Ours (32.270)**



(i) Ground Truth [B8] (j) Shift (32.826) (k) Scale (32.454) (l) **Ours (34.540)**

Figure 12: Comparison of ground truth, shift modulation, scale modulation, and our proposed method with PSNR across multibands (B10, B11, B8) of the Seoul image. Each row represents a different GSD: 60m, 20m, and 10m, respectively.

Table 12: Detailed PSNR for each band of Seoul image.

Methods	**B1**	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	**B8**	**B8A**	**B9**	**B10**	**B11**	**B12**
Shift	30.128	36.101	35.476	32.686	28.891	27.230	25.220	32.826	24.400	29.025	24.631	27.047	29.931
Scale	29.736	35.614	35.007	32.512	28.689	27.051	25.037	32.454	24.195	28.965	24.326	26.748	29.683
Fourier (Ours)	**48.224**	**37.726**	**36.884**	**34.035**	**32.524**	**31.161**	**29.999**	**34.540**	**29.944**	**48.538**	**51.222**	**32.270**	**34.373**

Table 13: Detailed MSE for each band of Seoul image.

Methods	**B1**	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	**B8**	**B8A**	**B9**	**B10**	**B11**	**B12**
Shift	9.710e-4	2.454e-4	2.834e-4	5.388e-4	1.291e-3	1.892e-3	3.006e-3	5.217e-4	3.631e-3	1.252e-3	3.442e-3	1.974e-3	1.016e-3
Scale	1.063e-3	2.746e-4	3.157e-4	5.607e-4	1.352e-3	1.972e-3	3.136e-3	5.684e-4	3.806e-3	1.269e-3	3.693e-3	2.115e-3	1.076e-3
Fourier (Ours)	**1.505e-5**	**1.688e-4**	**2.049e-4**	**3.949e-4**	**5.593e-4**	**7.653e-4**	**1.000e-3**	**3.516e-4**	**1.013e-3**	**1.400e-5**	**7.548e-6**	**5.929e-4**	**3.653e-4**