Graph-Based Image Fusion: Enhancing Multi-Modal Remote Sensing Data with GCNs

Johanna Garcia-Cardona*, Kevin Arias†, Antonio Ortega*, Henry Arguello†
*Dept. Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, USA
†Dept. of Electrical Engineering, Universidad Industrial de Santander, Bucaramanga, Colombia

***Abstract*—Multi-modal satellite data fusion is a key challenge in remote sensing, requiring the integration of data with differing spatial and temporal resolutions. Fusion methods must enhance image resolution while preserving the physical integrity of the underlying phenomena. Traditional fusion techniques, such as interpolation and statistical downscaling, often disrupt spatial consistency and introduce artifacts that distort physical representations. Graph-based methods offer a robust alternative by modeling diverse data sources into node-based signals and edge connections based on ancillary data. This work proposes a novel Graph Convolutional Network (GCN) framework for satellite image fusion, leveraging ancillary variables such as terrain elevation and vegetation indices to enhance spatially consistent data integration. A key advantage of this GCN approach is its ability to learn transferable edge relationships, enabling adaptation to new regions using ancillary data without retraining. Experimental results demonstrate that our GCN framework effectively captures spatial dependencies within the graph embedding while merging heterogeneous data sources, leading to improved image resolution and enhanced feature representation.**

***Index Terms*—Data Fusion, Graph Signal Processing (GSP), Graph Convolutional Networks (GCNs), Remote Sensing**

I. Introduction

The advancement of Earth observation technologies—including radiometers, synthetic aperture radars (SAR), and optical sensors—has provided unprecedented amounts of data for remote sensing. However, differences in spatial and temporal resolutions, sensor limitations, and data gaps present significant challenges to generating high-quality remote sensing products [1]. Addressing these issues requires data fusion techniques to improve the accuracy and resolution of satellite-derived variables, such as soil moisture (SM). These fusion techniques fall into two main categories: *image-only*, e.g., super-resolution and pan-sharpening, and *multimodal* approaches that integrate ancillary data, such as elevation or vegetation indices. This work focuses on the latter category, where additional sources of information are leveraged to improve estimation quality.

Image-only fusion methods often struggle with spatial consistency because they combine pixel values from different imaging sources, making them prone to missing critical terrain changes that could be obtained from other sources of information. This limitation can introduce artifacts that distort the representation of underlying physical processes [6]. For instance, pan-sharpening, extensively explored in remote sensing, enhances spatial resolution by merging low-resolution multispectral images with high-resolution auxiliary images, typically a panchromatic (PAN) band [7]. The most commonly used pan-sharpening algorithms, principal component analysis (PCA) and the Brovey transformation [8], have been applied to a wide range of image datasets. However, these methods can compromise spectral fidelity or result in the loss of valuable information from low-resolution data, as they do not account for the diverse land cover compositions represented in the images [9].

Machine learning (ML) methods can incorporate ancillary data, efficiently handle large datasets, and model complex nonlinear relationships [10]. ML-based downscaling techniques, for example, enhance spatial resolution by integrating ancillary variables such as vegetation indices and meteorological conditions. However, these approaches often rely on region-specific training data, limiting their ability to generalize across diverse landscapes [3], [11]. Deep learning-based super-resolution improves resolution by learning spatial patterns from low-resolution inputs but may produce blurred outputs, especially in areas with complex textures or sharp transitions [12]. This limitation often arises from the network's tendency to prioritize global structure over fine-grained variations, leading to oversmoothing. Furthermore, deep learning-based fusion techniques require large training datasets, which may not be readily available in remote sensing applications [4], [13].

Graph-based models [14], which can also incorporate ancillary data, are particularly well-suited for processing remote sensed signals that can exhibit irregularities [15], including (i) *spatial* irregularity, occurring when measurements, such as those from in-situ sensors or drones, are not acquired on a uniform grid, and (ii) *contextual* irregularity, which arises from prior knowledge or ancillary data that influence the relationships between samples. In this paper, we focus on this latter type of irregularity, arising from prior information, such as physical constraints or domain-specific correlations, that can introduce varying levels of dependence between observations. Path-wise graphs [16], have been used to enhance images by analyzing structural differences instead of pixel values. Similarly, superpixel-wise graphs [17] employ a sparsely constrained adaptive structure for a regression that improves change detection by leveraging structural consistencies. While these methods focus on capturing spatial relationships within images using graphs, they primarily rely on self-similarity and structure consistency without explicitly incorporating auxiliary information. To incorporate terrain data, the approach introduced in [18], [19] constructed a graph where node signals

TABLE I

COMPARISON OF DATA FUSION METHODS THAT CAN INCORPORATE AUXILIARY DATA. ▮ INDICATES HIGH-COMPLEXITY APPROACHES WITH EXTENSIVE DATA REQUIREMENTS AND CHALLENGING IMPLEMENTATION. ▮ REPRESENTS METHODS WITH MODERATE COMPLEXITY, BALANCING PERFORMANCE AND FEASIBILITY. ▮ HIGHLIGHTS TECHNIQUES KNOWN FOR THEIR ADAPTABILITY ACROSS DIVERSE DATASETS AND CONDITIONS.

Study - Method	Model Complexity	Training Data Dependency	Transferability Across Regions	Limitations
Yu el al. (2021) - Random Forest [2]	▮ No need for data normalization. Rule-based approach.	▮ Requires training per categorized soil texture group.	▮ Region-specific models needed.	▮ Long training time, sensitive to imbalanced data.
Dorigo et al. (2021) - ANN [3]	▮ Detects non-linear relationships. Complex structure.	▮ Requires labeled datasets for supervised learning.	▮ Retraining needed for different landscapes.	▮ High computational demand, prone to overfitting.
Xu et al. (2021) - CNN [4]	▮ Specialized architecture with multiple layers for spatial data.	▮ Large training sample requirement.	▮ Requires domain-specific fine-tuning.	▮ High data demand, computationally expensive.
Llamas et al. (2022) - KNN [5]	▮ Simple, distance-based classification.	▮ Training requires dense sampling points for accuracy.	▮ Region-specific models are required.	▮ Fails with missing data, poor generalization in sparse datasets.
Graph-Based Processing	▮ Requires graph structure knowledge and relationships among nodes and edges.	▮ Does not need region-specific training data.	▮ Adaptable to diverse regions without retraining.	▮ Requires graph construction and edge definition.

represent the image, and edge connections encode terrain characteristics, enhancing resolution from multimodal data.

While both graph-based and ML methods can incorporate multimodal data, however, as illustrated in Table I, ML approaches [2]–[5] often require larger training datasets. Graph methods explicitly define dependencies through structural connections, allowing for more explicit modeling of spatial and contextual relationships. However, they usually rely on simple predefined formulas for edge weights [18], [19], limiting their ability to capture complex interactions. In contrast, our proposed method uses limited amounts of training data to learn functions that map differences ancillary data to edge weights, enabling a more adaptive and data-driven representation.

In this paper, we propose a graph convolutional network (GCN) approach that learns feature relationships to determine edge strengths rather than directly integrating geophysical data into the graph construction, as in [18]. This method enhances signal representation at the nodes while dynamically learning how auxiliary variables—such as terrain features, land cover, and other physical attributes—affect edge strength by capturing non-linear dependencies within both signal and edge weight domains. As a key advantage, this framework can train the embedded relationships of edges in one region and apply them to another for which ancillary datasets are available. This results in improved adaptability across geographically diverse areas without requiring region-specific model retraining.

The paper is structured as follows. Section II introduces the graph-based data fusion approach, followed by the methodology for incorporating GCNs. A case study on soil moisture enhancement is presented in Section III, with conclusions provided in Section IV.

II. PROPOSED METHODOLOGY

Remote sensing images can be modeled as graph signals on a graph, where each node corresponds to a sensed measurement. In previous work, horizontal and vertical edges connect neighboring nodes with non-negative weights w_{ij}, determined by the similarity of terrain characteristics at their respective locations [20]. In our proposed approach, we enhance this framework by employing GCNs to dynamically learn and refine edge weights. GCNs leverage relationships discovered within the auxiliary variables to adjust edge connections adaptively, without requiring predefined terrain-based similarities.

A. Graph Construction

Let $\mathbf{Y} \in \mathbb{R}^{n\times n}$ denote the low-resolution SM and $\mathbf{A} \in \mathbb{R}^{N\times N\times d}$ the ancillary variables, where d is the number of available variables. $\mathbf{Y}$ and $\mathbf{A}$ are used for estimating the target high-resolution SM $\mathbf{X} \in \mathbb{R}^{N\times N}$. We extract sets of K local co-registered patches $\mathcal{Y} = \left\{\mathbf{Y}^k \in \mathbb{R}^{p\times p}\right\}_{k=1}^{K}$, $\mathcal{A} = \left\{\mathbf{A}^k \in \mathbb{R}^{P\times P\times d}\right\}_{k=1}^{K}$ and $\mathcal{X} = \left\{\mathbf{X}^k \in \mathbb{R}^{P\times P}\right\}_{k=1}^{K}$.

Graph Definition. For each patch in $\mathcal{Y}$ and $\mathcal{A}$, we define an undirected weighted graph, with k-th patch graphs denoted as $\mathcal{G}_y\left(\mathcal{V}_y, \mathbf{y}^k, \mathbf{W}_y^k\right)$ and $\mathcal{G}_a\left(\mathcal{V}_a, \mathbf{A}_{\text{vec}}^k, \mathbf{W}_a^k\right)$, where the nodes $\mathcal{V}_y = \left\{1, 2, \ldots, p^2\right\}$ and $\mathcal{V}_a = \left\{1, 2, \ldots, P^2\right\}$ correspond to the pixels in each patch. The information from $\mathbf{Y}^k$ and $\mathbf{A}^k$ is vectorized as $\mathbf{y}^k = \text{vec}\left(\mathbf{Y}^k\right) \in \mathbb{R}^{p^2}$ and $\mathbf{A}_{\text{vec}}^k = \text{vec}\left(\mathbf{A}^k\right) \in \mathbb{R}^{P^2\times d}$, representing the graph signals $\mathbf{y}^k : \mathcal{V}_y \to \mathbb{R}$ and $\mathbf{A}_{\text{vec}}^k : \mathcal{V}_a \to \mathbb{R}^d$ such that the signal value at each node equals the corresponding pixel intensity or the pixel intensities along the d ancillary variables. The symmetric matrices $\mathbf{W}_y^k \in \mathbb{R}^{p^2\times p^2}$ and $\mathbf{W}_a^k \in \mathbb{R}^{P^2\times P^2}$ represent the edge weights of the graphs $\mathcal{G}_y$ and $\mathcal{G}_a$, respectively. The values $\mathbf{W}_y^k(i,j)$ and $\mathbf{W}_a^k(i,j)$ quantify the similarity between pixels i and j, encoding the local structural relationships. Finally, we construct the fused graph $\mathcal{G}_f\left(\mathcal{V}_f, \mathbf{F}_{\text{vec}}^k, \mathbf{W}_f^k\right)$, where the signal $\mathbf{F}_{\text{vec}}^k \in \mathbb{R}^{P^2\times Z}$ across the nodes $\mathcal{V}_f = \left\{1, 2, \ldots, P^2\right\}$ is obtained by concatenating the features of both input sources.

$$\mathbf{F}_{\text{vec}}^k = \mathbf{F}_y^k \oplus \mathbf{F}_a^k : \mathcal{V} \to \mathbb{R}^Z, \tag{1}$$

where $\mathbf{F}_y^k \in \mathbb{R}^{P^2\times Z/2}$ in the upper arm and $\mathbf{F}_a^k \in \mathbb{R}^{P^2\times Z/2}$ in the lower arm of Fig. 1 are graph-based deep features extracted by GCNs from the graphs $\mathcal{G}_y$ and $\mathcal{G}_a$, respectively.

Definition of weights. The weight matrices are initialized based on the distance between pixels within the graph signals, such that $\bar{\mathbf{W}}_y^k(i,j) = 1 - |\mathbf{y}^k(i) - \mathbf{y}^k(j)|$, $\bar{\mathbf{W}}_a^k(i,j) = 1 - \frac{1}{d}\|\mathbf{A}_{\text{vec}}^k(i) - \mathbf{A}_{\text{vec}}^k(j)\|_1$ and $\bar{\mathbf{W}}_f^k(i,j) = 1 - \frac{1}{Z}\|\mathbf{F}_{\text{vec}}^k(i) - \mathbf{F}_{\text{vec}}^k(j)\|_1$. Then, we introduce a learnable function $\mathcal{R}_{\boldsymbol{\Theta}} : \mathbb{R}^{M\times M} \to \{0,1\}^{\frac{M(M-1)}{2}}$ parameterized by $\boldsymbol{\Theta}$ to control

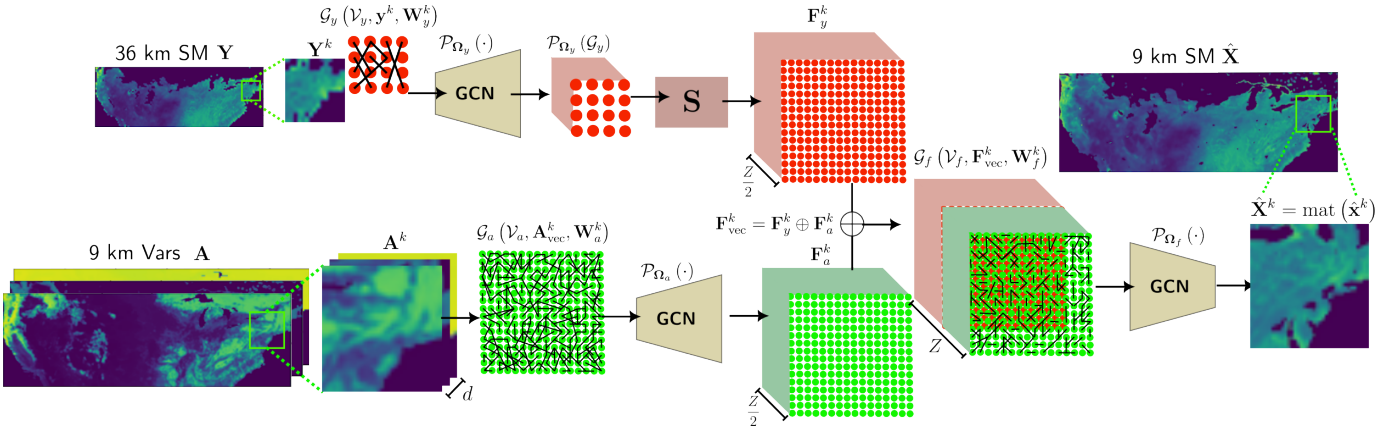

Fig. 1. High-resolution SM is reconstructed from a fused graph representation that integrates multi-source data. Independent graph representations are computed for low-resolution SM (36 km) and high-resolution variables (9 km) such as VWC, VOD, and LST, then expanded into latent representations via GCNs that leverage graph connectivity. These latent representations—comprising the high-resolution variable features and the upsampled low-resolution SM features—are concatenated, and a subsequent GCN learns their connectivity to reconstruct the high-resolution SM.

weight sparsity. This function takes a weight matrix of M nodes (where $M = p^2$ or $M = P^2$, depending on the signal resolution) and outputs a binary mask that determines whether each weight should be retained (1) or removed (0). To obtain the binary mask, we first compute a feature matrix $\mathbf{U}$, where each element $\mathbf{U}(i,j)$ represents a learned feature for the edge between nodes i and j. The feature matrix is computed as $\mathbf{U} = \mathcal{C}_{\boldsymbol{\Theta}}\left(\text{vec}_{\triangle}\left(\bar{\mathbf{W}}\right), \{(i,j)\}^{\frac{M(M-1)}{2}}\right)$ where $\mathcal{C}_{\boldsymbol{\Theta}}$ is a 1D convolutional layer with positional encoding. This function learns position-aware features for the $\frac{M(M-1)}{2}$ node pairs corresponding to the upper triangular part $\text{vec}_{\triangle}\left(\bar{\mathbf{W}}\right)$. The final weight matrix $\bar{\mathbf{W}} \in \mathbb{R}^{M \times M}$ is obtained as

$$\mathbf{W}(i,j) = \bar{\mathbf{W}}(i,j) \cdot \mathcal{R}_{\boldsymbol{\Theta}}\left(\bar{\mathbf{W}}\right)(i,j), \tag{2}$$

where $\mathcal{R}_{\boldsymbol{\Theta}}\left(\bar{\mathbf{W}}\right)(i,j) = \text{sign}\left(\mathbf{U}(i,j)\right)$ corresponds to the values of the binary mask. (2) ensures that only selected weights remain in the final graph representation while others are pruned based on learned features. The weights are estimated for $i < j$ as the undirected topology is fully determined by its upper triangular part. Following the general formulation in (2), we learn the weights as $\mathbf{W}_y^k(i,j) = \bar{\mathbf{W}}_y^k(i,j) \cdot \mathcal{R}_{\boldsymbol{\Theta}_y}\left(\bar{\mathbf{W}}_y^k\right)(i,j)$, $\mathbf{W}_a^k(i,j) = \bar{\mathbf{W}}_a^k(i,j) \cdot \mathcal{R}_{\boldsymbol{\Theta}_a}\left(\bar{\mathbf{W}}_a^k\right)(i,j)$ and $\mathbf{W}_f^k(i,j) = \bar{\mathbf{W}}_f^k(i,j) \cdot \mathcal{R}_{\boldsymbol{\Theta}_f}\left(\bar{\mathbf{W}}_f^k\right)(i,j)$ with trainable parameters $\boldsymbol{\Theta}_y$, $\boldsymbol{\Theta}_a$ and $\boldsymbol{\Theta}_f$ for the graphs of low-resolution SM, ancillary variables and fused, respectively.

B. Graph Convolution Network Architecture

We propose a supervised network for high-resolution reconstruction, leveraging graph-based local relationships to fuse multiple sources $\mathbf{y}^k$ and $\mathbf{A}_{\text{vec}}^k$. The fusion occurs in the space of their features, with $\mathbf{F}_y^k = \mathbf{S}\ \mathcal{P}_{\boldsymbol{\Omega}_y}(\mathcal{G}_y)$ and $\mathbf{F}_a^k = \mathcal{P}_{\boldsymbol{\Omega}_a}(\mathcal{G}_a)$ corresponding to the upper and lower arms of the network in Fig. 1 where $\mathbf{S} \in \mathbb{R}^{P^2 \times p^2} = \text{softmax}\left(\bar{\mathbf{S}}\right)$ is a softmax-normalized upsampling matrix, initialized from a random matrix $\bar{\mathbf{S}} \sim \mathcal{N}(0,1)^{P^2 \times p^2}$, ensuring a smooth and probabilistic mapping of low-dimensional features to a high-dimensional space. The GCNs $\mathcal{P}_{\boldsymbol{\Omega}_y}$ and $\mathcal{P}_{\boldsymbol{\Omega}_a}$ follow a two-layer architecture [21], effectively leveraging structural advantages of graphs to capture local relationships and dependencies between connected nodes for an enhanced representation. The concatenated features $\mathbf{F}_y^k \oplus \mathbf{F}_a^k$ form the signal on the fused graph $\mathcal{G}_f$, which is then passed through the GCN $\mathcal{P}_{\boldsymbol{\Omega}_f}$ to estimate the target high-resolution SM as:

$$\hat{\mathbf{x}}^k = \mathcal{P}_{\boldsymbol{\Omega}_f}\left(\mathcal{G}_f\left(\mathcal{V}_f, \mathbf{F}_y^k \oplus \mathbf{F}_a^k, \mathbf{W}_f^k\right)\right), \tag{3}$$

where the network output $\hat{\mathbf{X}}^k = \text{mat}\left(\hat{\mathbf{x}}^k\right)$ is the matrix form of $\hat{\mathbf{x}}^k$ representing the high-resolution SM estimate for the k-th patch. The optimization minimizes (i) the distance between the target $\mathbf{x}^k$ and estimate $\hat{\mathbf{x}}^k$ (quality term), (ii) the distance between the target $\mathbf{x}^k$ and a rough upsampling $\mathbf{S}\,\mathbf{y}^k$, regularizing $\mathbf{S}$ acts as a consistent upsampling operator, as it is used to upsample low-resolution features and, (iii) the error between distance-based target weights $\bar{\mathbf{W}}_x^k(i,j) = 1 - |\mathbf{x}^k(i) - \mathbf{x}^k(j)|$ and $\bar{\mathbf{W}}_f^k(i,j)$, enforcing alignment of fuse graph weights with target weights, shown beneficial for GCN-based reconstruction. The optimization problem is:

$$\boldsymbol{\Theta}^*, \boldsymbol{\Omega}^*, \mathbf{S}^* = \underset{\boldsymbol{\Theta}, \boldsymbol{\Omega}, \mathbf{S}}{\arg\min} \frac{1}{2} \sum_k \|\mathbf{x}^k - \hat{\mathbf{x}}^k\|_2^2 + \|\mathbf{x}^k - \mathbf{S}\mathbf{y}^k\|_2^2 + \|\text{vec}_{\Delta}(\bar{\mathbf{W}}_x^k) - \text{vec}_{\Delta}(\bar{\mathbf{W}}_f^k)\|_2^2. \tag{4}$$

where $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_y, \boldsymbol{\Theta}_a, \boldsymbol{\Theta}_f\}$ is the set of trainable parameters for the graph weights, $\boldsymbol{\Omega} = \{\boldsymbol{\Omega}_y, \boldsymbol{\Omega}_a, \boldsymbol{\Omega}_f\}$ denote the GCN learnable kernel and $\mathbf{S}$ is the learnable upsampling matrix, all these represent the trainable parameters of our network.

III. Case Study

As a case study for our data fusion method, we focus on the Soil Moisture Active Passive (SMAP) mission. SMAP used an L-band radiometer (36 km resolution) and an L-band radar (3 km resolution) for soil moisture (SM) estimation [22], [23]. After the radar's failure, SMAP relied solely on coarse radiometer-derived brightness temperatures (Tb), limiting spatial resolution. This highlights the need for advanced fusion techniques to reconstruct high-resolution soil moisture (SM) maps from coarse-resolution radiometric observations, using auxiliary terrain information. We evaluated our GCN-based

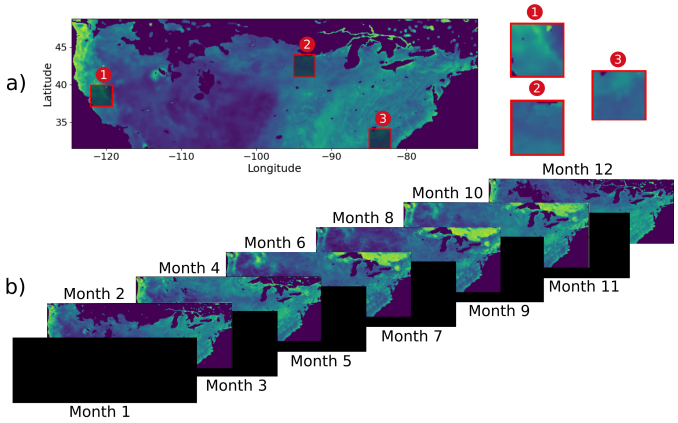

Fig. 2. Setups for spatial (a) and temporal (b) shot reconstructions: (a) Missing spatial regions (red) are reconstructed using a model trained on the remaining data. (b) Data from odd-numbered months (black) is reconstructed using a model trained on even-numbered months.

approach using SMAP mission data in the continental United States.

A. Experimental Setup

We used two spatial resolutions: 36 km (coarse) and 9 km (fine). Soil moisture is the target signal, while the ancillary variables—vegetation optical depth, land surface temperature, and vegetation water content—were available at 9 km. These ancillary variables were used to enhance the signal on the nodes. The interpolation performance was assessed by comparing estimates with the closest ground-based observations.

B. Ablation Studies

We conduct two ablation studies to assess the generalizability of our network. Inference is tested when training data lacks local spatial regions at all time instants, Fig. 2(a), and when the entire terrain is missing from training for specific months (temporal shot), Fig. 2(b).

Spatial Reconstruction: Spatially delimited regions known as 1) Tonzi Ranch, 2) South Fork and 3) LittleRiver as enumerated in Fig. 2(a) are removed from the training dataset throughout all months, ensuring that the network does not know any spatial details of these local regions.Fig. 3 confirms the ability of our graph representation and network, with consistent estimations across regions and months, demonstrating reliable reconstruction in unseen regions.

Temporal Shot Reconstruction: Temporal shots are removed from training as shown in Fig. 2(b), ensuring the network has no prior knowledge of any spatial details of specific months while retaining information from contiguous ones. This study evaluates the capability of the network to generalize to missing months given prior knowledge of the same region at different times. Visual and PSNR results in Fig. 4 confirm the effectiveness of our graph representation in reconstructing missing temporal shots.

C. Discussion

The results summarized in Table II demonstrate the effectiveness of our proposed method in estimating soil moisture

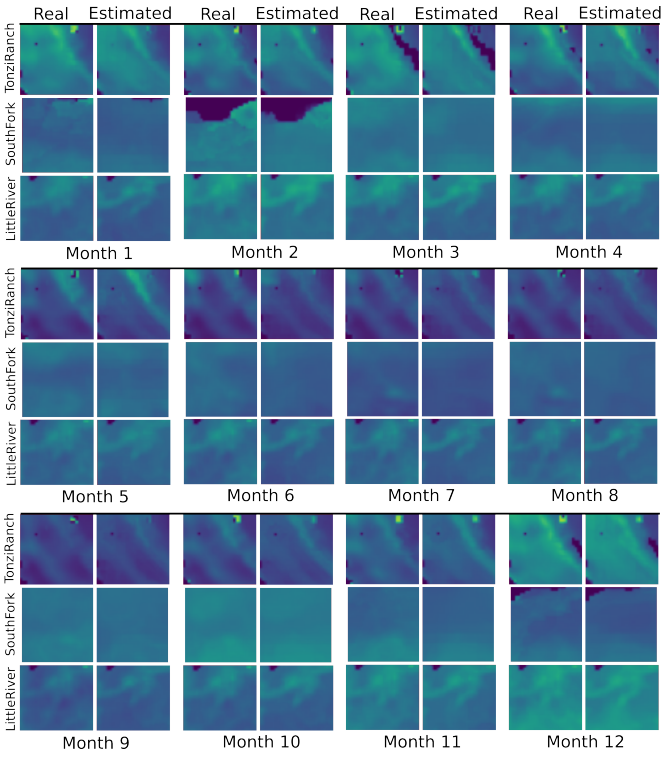

Fig. 3. Spatial reconstruction: 9 km SM inference for Tonzi Ranch, South Fork, and Little River, excluded during training. Each row shows estimations alongside corresponding real data, organized by month (bottom of each panel).

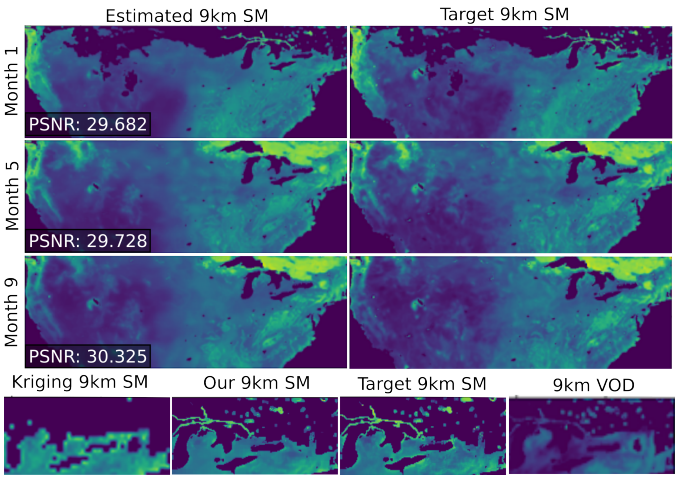

Fig. 4. Temporal shot reconstruction: 9 km SM estimations (left) for some missing odd-numbered months (rows) closely match the target SM (right). PSNR results validate reconstruction quality despite the model lacking 9 km SM data for those months. A zoomed-in view of the last row highlights the local consistency of our method in recovering fine details, which are missing from image-only approaches like Kriging but can be extracted from ancillary variables such as VOD.

(SM) from enhanced satellite images, validated against in situ measurements. Correlation and error performance metrics indicate that the proposed GCN model not only maintains spatial consistency in the reconstructed images, but also achieves high accuracy, producing estimates that align with ground-based observations. Furthermore, a comparison with nearest neigh-

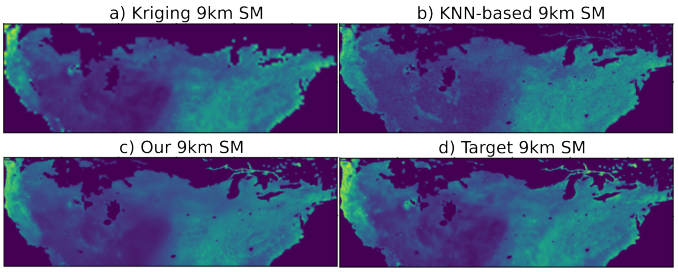

Fig. 5. 9 km SM estimations from (a) image-only Kriging, (b) a KNN-based approach using ancillary variables, and (c) our proposed method. Visual results indicate that our approach more closely resembles the target shown in (d).

bor, kriging (a statistical interpolation technique that estimates values using neighboring data points), and other graph-based methods highlights the superior performance of our approach, demonstrating improved estimation accuracy. Visualizations in Fig. 5 also demonstrate enhanced visual quality compared to methods that rely solely on image data, such as Kriging, and those that incorporate ancillary variables, such as the KNN-inspired method. The spatial reconstruction study, where regions such as Tonzi Ranch, South Fork, and Little River were entirely omitted from the training dataset, shows that the pre-trained model successfully generalizes to previously unseen terrain regions. This confirms that our approach does not rely on predefined feature-to-feature relationships but rather on a more robust representation of the feature space within the network's learned embedding.

TABLE II
COMPARISON RESULTS FOR DIFFERENT METHODS ESTIMATING IN-SITU MEASUREMENTS, THE ERRORS ARE ESTIMATED IN TERM OF THE STATE-OF-THE-ART METRICS.

	Nearest-Neighbor	Kriging	Graph-Based [19]	Ours
MAE ↓	0.090	0.085	0.0729	**0.062**
RMSE ↓	0.110	0.094	0.083	**0.074**
ubRMSE ↓	0.090	0.088	-	**0.061**
bias ↓	0.087	0.084	-	**0.033**
R ↑	0.749	0.703	-	**0.803**

IV. CONCLUSION AND FUTURE WORK

A key advantage of our approach is its ability to train on regions with specific terrain characteristics and generalize to new areas. This ensures that the learned representations remain applicable across different geographic regions, provided ancillary datasets are available. Such adaptability is particularly valuable in remote sensing, where terrain and environmental conditions vary widely. The proposed GCN framework provides a flexible and adaptive approach for modeling feature relationships, effectively capturing nonlinear dependencies while maintaining spatial consistency in reconstructed images. Future research will focus on extending this framework to enhance temporal resolution, enabling improved monitoring of dynamic environmental processes. Additionally, we aim to explore its scalability to datasets with sparse measurements, broadening its applicability across diverse remote sensing challenges.

REFERENCES

[1] M. Burke, A. Driscoll, D. B. Lobell, and S. Ermon, "Using satellite imagery to understand and promote sustainable development," *Science*, vol. 371, no. 6535, Mar. 2021.
[2] Y. Yu, L. Renzullo, and S. Tian, "Continental scale downscaling of awra-l analysed soil moisture using random forest regression," in *Proc. MODSIM'21*, 12 2021.
[3] W. Dorigo, I. Himmelbauer, D. Aberer, and L. Schremmer, "The international soil moisture network: Serving earth system science for over a decade," *Hydrol. Earth Syst. Sci.*, vol. 25, pp. 5749–5804, 2021.
[4] W. Xu, Z. Zhang, Z. Long, and Q. Qin, "Downscaling smap soil moisture products with convolutional neural network," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. PP, pp. 1–1, 03 2021.
[5] R. Llamas, L. Valera, P. Olaya, M. Taufer, and R. Vargas, "Downscaling satellite soil moisture using a modular spatial inference framework," *Remote Sens.*, vol. 14, pp. 3137, 06 2022.
[6] U. Gangkofner, P. Pradhan, and D. Holcomb, "Optimizing the High-Pass Filter Addition Technique for Image Fusion," *Photogramm. Eng. Remote Sens.*, vol. 74, Sept. 2008.
[7] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 313–317, 2004.
[8] D. Rodríguez-Esparragón, J. Marcello, F. Eugenio, A. García-Pedrero, and C. Gonzalo-Martin, "Object-based quality evaluation procedure for fused remote sensing imagery," *Neurocomputing*, vol. 255, 04 2017.
[9] F. Dadrass Javan, F. Samadzadegan, and F. Fathollahi, "Spectral and spatial quality assessment of ihs and wavelet based pan-sharpening techniques for high resolution satellite imagery," *Adv. Image Video Process.*, vol. 6, 04 2018.
[10] Y. Liu, W. Jing, Q. Wang, and X. Xia, "Generating high-resolution soil moisture by using spatial downscaling techniques: a comparison of six machine learning algorithms," *Adv. Water Resour.*, vol. 141, pp. 103601, 05 2020.
[11] I. Senanayake, K. Lakshani, I.-Y. Yeo, S.-C. Han, and P. Dahlhaus, "Spatial downscaling of satellite-based soil moisture products using machine learning techniques: A review," *Remote Sens.*, vol. 16, pp. 2067, 06 2024.
[12] M. Jiang, H. Shen, J. Li, Q. Yuan, and L. Zhang, "A differential information residual convolutional neural network for pansharpening," *ISPRS J. Photogramm. Remote Sens.*, vol. 163, pp. 257–271, 05 2020.
[13] G. Vivone, "Robust band-dependent spatial-detail approaches for panchromatic sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. PP, pp. 1–13, 04 2019.
[14] A. Ortega, *Introduction to Graph Signal Processing*, Cambridge University Press, 2022.
[15] A. Sanfeliu, R. Alquézar, J. Andrade, J. Climent, F. Serratosa, and J. Vergés, "Graph-based representations and techniques for image processing and image analysis," *Pattern Recognit.*, vol. 35, pp. 639–650, 2002.
[16] Y. Sun, L. Lei, X. Li, X. Tan, and G. Kuang, "Structure consistency-based graph for unsupervised change detection with homogeneous and heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–21, 2022.
[17] Y. Sun, L. Lei, D. Guan, M. Li, and G. Kuang, "Sparse-constrained adaptive structure consistency-based unsupervised image regression for heterogeneous remote-sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.
[18] J. Garcia-Cardona, A. Ortega, and N. Rodriguez-Alvarez, "Graph-Based Interpolation for Remote Sensing Data," in *Proc. EUSIPCO'22*, Aug. 2022, pp. 1791–1795.
[19] J. Garcia-Cardona, W.-Y. Lu, and A. Ortega, "Multimodal graph-based fusion to enhance satellite image resolution," in *Proc. CAMSAP'23*, 12 2023, pp. 31–35.
[20] J. Garcia-Cardona, A. Ortega, and N. Rodriguez-Alvarez, "Downscaling smap soil moisture with ecostress products using a graph-based interpolation method," in *Proc. IGARSS'22*, 2022, pp. 6169–6172.
[21] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.
[22] P. E. O'Neill, S. Chan, E. G. Njoku, T. Jackson, R. Bindlish, and J. Chaubell, "Smap l3 radiometer global daily 36 km ease-grid soil moisture, version 6," 2022.
[23] Steve Ramsey, "Nasa soil moisture radar ends operations, mission science continues," July 2015, NASA Press Release.