Mitigating coarse spatial reconstruction to generate missing bands for the HLS dataset

Vasileios Tsironis[1], Athina Psalta[1], Konstantinos Karantzalos[1]

[1] NTUA, Remote Sensing Lab, 15772 Zografou Athens, Greece - (tsironisbi, psaltaath, karank)@central.ntua.gr

Keywords: Spectral band generation, SSL-features, Data Harmonization, HLS dataset, Earth Observation

Abstract

The Harmonized LandSat-Sentinel (HLS) dataset has significantly advanced Earth Observation by integrating data from Landsat and Sentinel satellites. However, challenges persist in achieving spectral band parity between LandSat and Sentinel-derived HLS products. This paper presents an extended investigation aimed at enhancing spatial reconstruction accuracy to enable spectral band parity within HLS products. Building upon our previous work, which utilized generative neural networks to address partial feature mismatches between S30 and L30 products, we introduce a refined approach that fully integrates a Self-Supervised Learning (SSL)-pretrained encoder into a U-Net architecture. This method aims to access multi-scale features and improve spatial reconstruction accuracy, addressing the limitations in spatial resolution observed in our earlier study. Our methodology incorporates a comprehensive ablation study to assess various SSL-pretrained backbone architectures. Preliminary results demonstrate significant improvements in spatial reconstruction accuracy compared to our previous work. The adapted U-Net architecture, leveraging SSL-pretrained encoders, shows enhanced capability in capturing intricate spatial features within the HLS dataset. Our experiments demonstrate a substantial improvement in spatial resolution and feature reconstruction for L30 products, particularly in bands not natively present in Landsat data, paving the way for more accurate multi-sensor analyses.

1. Introduction

Earth Observation (EO) has been revolutionized by the integration of data from multiple satellite platforms, particularly the Landsat and Sentinel missions. The Harmonized Landsat - Sentinel (HLS) dataset (Claverie et al., 2018), developed by NASA, stands as a landmark achievement in this field, offering a unified resource for mid-resolution multispectral optical imagery. However, despite the significant strides made in harmonizing these datasets, there are still some persisting challenges, particularly in achieving spectral band parity between Landsat and Sentinel-derived HLS products.

The Harmonized Landsat and Sentinel-2 (HLS) initiative provides two key products: L30 (Landsat 8/9-based) and S30 (Sentinel-2-based). While these products are spatially and spectrally harmonized, they exhibit notable differences. S30 includes additional bands absent in L30, such as ’Red Edge’ wavelengths, which are crucial for applications like vegetation monitoring (Xie et al., 2018), crop disease detection and health assessment (Stamford et al., 2023), and precision agriculture (Segarra et al., 2020). These disparities highlight the need for platform-agnostic models that can utilize all available information across datasets (Kganyago et al., 2022).

Predicting spectral band images with arbitrary characteristics from diverse spectral inputs remains an under-explored EO task. Recent advancements in optical image simulation, such as conditional GANs for generating Sentinel-2 images from SAR data (He and Yokoya, 2018), show promising results. For such tasks, Encoder-decoder architectures like pix2pix (Isola et al., 2017) are common in image-to-image approaches. In addition the rise of Self-Supervised Learning (SSL) techniques, including MOCOv2 (Chen et al., 2020b) and SimCLR (Chen et al., 2020a), have demonstrated high-quality feature extraction without explicit supervision. In EO, large-scale datasets suitable for SSL pre-training have emerged, such as the Functional Map of the World (FMoW) (Christie et al., 2018), SSL4EO-L (Stewart et al., 2023) and SeeFar (Lowman et al., 2024), enabling further advancements in this field.

In our previous work (Tsironis et al., 2024), we addressed the partial feature mismatch between S30 and L30 products within the HLS dataset via introducing a novel approach utilizing generative neural networks, specifically an Encoder-Decoder architecture that leveraged advanced SSL-pretrained backbones alongside a versatile Fully Convolutional Network (FCN) architecture. This method demonstrated promising results in inferring missing bands in the L30 product, namely S30-exclusive bands *RedEdge-1*, *RedEdge-2*, *RedEdge-3* and *NIR-Broad*, partially bridging the gap between S30 and L30 products.

However, our initial approach revealed limitations, particularly in spatial resolution. While radiometrically satisfying, the derived bands exhibited a notable loss in spatial resolution. This was attributed to the FCN architecture decoding SSL-pretrained features that were spatially compressed by a factor of 8. These findings motivated us to further refine our methodology, with a specific focus on enhancing spatial reconstruction accuracy.

In this paper, we present an extended investigation that builds upon our previous work, introducing significant improvements to address the spatial resolution limitations. Our refined approach integrates a fully SSL-pretrained encoder into the U-Net architecture (Ronneberger et al., 2015), enabling access to multi-scale features and improving spatial reconstruction accuracy. This advancement targets to maintain the spectral fidelity achieved in our earlier work while significantly enhancing the spatial resolution of the derived bands. By addressing the spatial resolution limitations of our previous work, this study aims to further bridge the gap between S30 and L30 products within the HLS dataset. The resulting improvements have the potential to enhance a wide range of EO applications, such as land cover mapping (Karakizi et al., 2023) (Karakizi et al., 2024) and vegetation monitoring (Ouzoun et al., 2023), by providing more spatially accurate and spectrally complete data products.

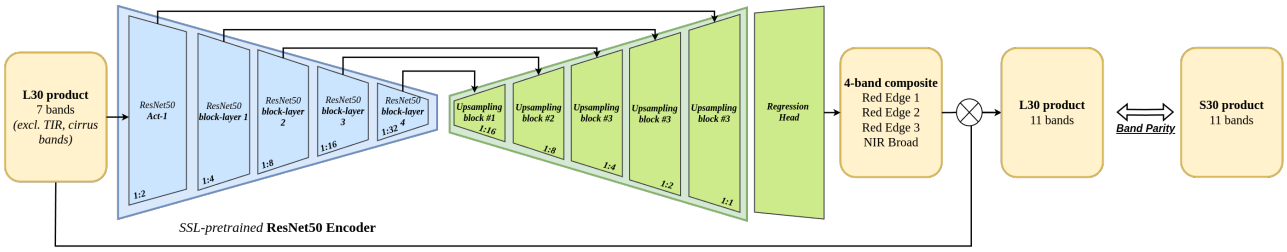


Figure 1. Proposed architecture consisting of a SSL-pretrained ResNet50 and a custom FCN U-Net like architecture.

In detail, the main contributions of this work are summarized as follows:

- Introduction of a novel architecture that fully integrates an SSL-pretrained encoder into a UNet framework, enabling multi-scale feature extraction for improved spatial reconstruction.
- Implementation of a comprehensive ablation study to systematically evaluate the performance of different SSL - pretrained backbone architectures in the context of HLS band prediction.
- Demonstration of substantial improvements in spatial resolution and feature reconstruction for L30 products, particularly in bands not natively present in Landsat data, enhancing the overall utility of the HLS dataset for Earth Observation applications.

2. Methodology

Our methodology builds upon our previous work (Tsironis et al., 2024), addressing the spatial resolution limitations by integrating a fully SSL-pretrained encoder into a U-Net architecture. This approach enables access to multi-scale features, significantly improving spatial reconstruction accuracy while maintaining spectral fidelity. The following subsections detail our enhanced model architecture, training pipeline and inference configuration.

2.1 SSL-pretrained Encoder

Similar to our previous work, we employ a SSL-pretrained ResNet50 model as our encoder, leveraging its superior feature representation capabilities. The model is pretrained using MOCOv2 (Chen et al., 2020b) on the SSL4EO-L dataset (Stewart et al., 2023), which is particularly suitable for Landsat data inference. Slight radiometric discrepancies between typical Landsat SR products and L30 products due to NBAR adjustment are negligible for the quality of the extracted SSL representations, as shown in our previous work.

A key improvement in our current work is the access to multiple feature maps at various downsampling factors. Specifically, we extract features from five different levels of the ResNet50 architecture, corresponding to spatial downsampling factors of 2, 4, 8, 16 and 32. This multi-scale approach allows our model to capture both fine-grained details and broader contextual information, addressing the spatial resolution limitations observed in our previous work. The feature extraction process can be formalized as follows:

$$F_i = E_i(x), \quad i \in \{2, 4, 8, 16, 32\} \tag{1}$$

where E_i represents the encoder function up to the i-th downsampling level, x is the input image, and F_i is the resulting feature map at the corresponding downsampling factor.

As in our previous work, the encoder’s weights remain frozen during training to preserve the rich features learned through SSL pretraining and avoid overfitting to the image generation task.

2.2 U-Net Decoder

To fully utilize the multi-scale features extracted by our encoder, we implement a U-Net-based decoder to directly predic missing bands *RedEdge-1*, *RedEdge-2*, *RedEdge-3* and *NIR-Broad* for the L30 product. This architecture allows for effective integration of features at different spatial resolutions, enabling more accurate spatial reconstruction.

Our U-Net decoder consists of multiple upsampling blocks, each corresponding to a feature map from the encoder. Each block includes the following components:

1. Upsampling operation (bilinear) to match the spatial dimensions of the next level
2. Skip connection from the corresponding encoder level
3. Concatenation of upsampled features and skip connection features
4. Two 3x3 convolutional layers followed by batch normalization and ReLU activation

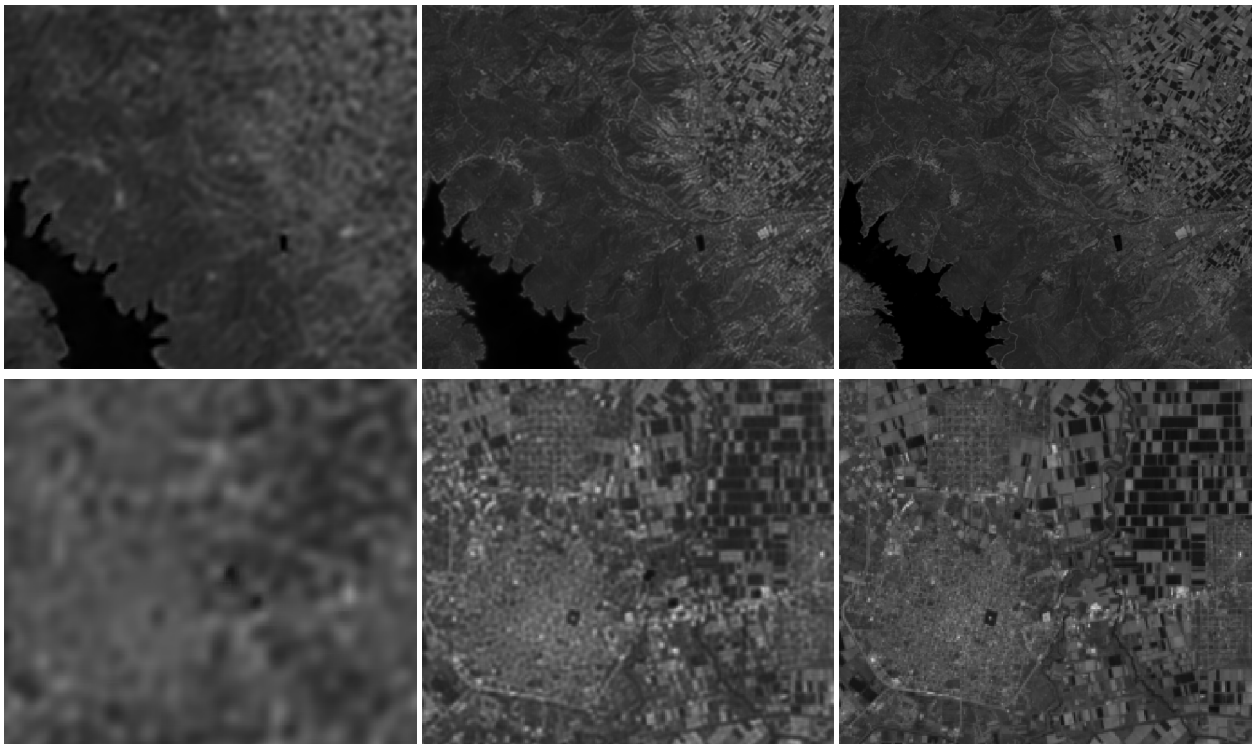
At the end of the Decoder, a convolutional regression head is attached using two 3x3 xonvolutional layers with ReLU intermediate activations and a final sigmoid activation layer to output reflectances ([0, 1] range).

2.3 Training Pipeline

Our training pipeline remains similar to our previous work, utilizing the S30 product of the HLS dataset. The 7 spectral bands equivalent to L8/9 OLI bands serve as input, while the remaining 4 bands are used as ground-truth labels. We train our model for 16 epochs using the AdamW optimizer with a learning rate of 1e-4. We do not employ any feature precomputing steps as no notable performance improvements were noticed. We employ a typical Mean Squared Error (MSE) loss to train our band prediction network:

$$\mathcal{L}_{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

where n is the total number of pixels across all bands in the image, y_i is the true value of the i-th pixel, and $\hat{y}_i$ is the predicted value of the i-th pixel.



(a) (b) (c)

Figure 2. Comparison of methods. (a) Previous work (FCN decoder, ResNet50 MoCo v2), (b) Ours (UNet decoder, ResNet50 MoCo v2), (c) Ground Truth (coincident S30) for two different scenes (top and bottom row respectively).

2.4 Inference Configuration

Inference is performed directly on the L30 product of HLS, producing 4 synthetic images corresponding to the extra bands found in the S30 product. The final result is an augmented L30 product that is spectrally equivalent to the S30 product while maintaining high spatial resolution. Using our method, HLS products can be used intercheangeably in a wide variety of downstream tasks without any further processing.

3. Experimental Evaluation

This section presents a comprehensive evaluation of our proposed method, comparing it with our previous work. We describe the dataset, evaluation criteria, and present both quantitative and qualitative results. Finally, we conduct an ablation study to analyze the impact of different SSL-pretrained backbones on our model’s performance.

3.1 Dataset

We utilize the Harmonized Landsat and Sentinel-2 (HLS) dataset (Claverie et al., 2018), maintaining consistency with our previous work. Our dataset comprises S30 and L30 products collected from several tiles spanning the central and northern parts of Greece for one month (July 2023). This configuration ensures diverse landscapes, including mountainous and flat terrains, various flora, and water surfaces, while minimizing cloud cover.

For training, we use the S30 data, divided into training and validation subsets with an 80%/20% ratio. The L30 data are used exclusively for evaluation. In all cases, we preprocess the data to mask out clouds, cloud shadows, and ice/snow.

3.2 Evaluation Criteria

Unlike our previous work, which used both coincident correlation and time-series interpolation, we now focus solely on coincident L30-S30 pairs for a more direct and accurate assessment. We employ the following standard image generation metrics:

- **Peak Signal-to-Noise Ratio (PSNR)**: Measures the ratio between the maximum possible signal power and the power of distorting noise. For a 16-bit unsigned integer (uint16) image:

$$\text{PSNR} = 20 \cdot \log_{10}\left(\frac{65535}{\sqrt{\text{MSE}}}\right) \tag{3}$$

  where MSE is the Mean Squared Error between the original and the generated image.

- **Structural Similarity Index Measure (SSIM)**: Assesses the perceived quality of digital images. It is defined as:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{4}$$

  where μ_x and μ_y are the average of x and y, σ_x^2 and σ_y^2 are the variance of x and y, σ_{xy} is the covariance of x and y, and c_1 and c_2 are variables to stabilize the division with weak denominator.

- **Mean Absolute Error (MAE)**: Quantifies the average magnitude of the errors in a set of predictions, expressed as a percentage of reflectance:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (5)$$

where y_i is the true value, $\hat{y}_i$ is the predicted value, and n is the number of samples.

These metrics provide a comprehensive evaluation of both the overall radiometric accuracy and the structural fidelity of our generated bands.

3.3 Quantitative Evaluation

Our new approach demonstrates significant improvements over our previous work across all evaluation metrics. In Table 1 a comparison of the performance between our current and previous methods is presented.

Method	Band	PSNR(dB)	SSIM(%)	MAE(%)
Previous Work	RE-1	33.06	87.77	6.2
	RE-2	33.93	89.32	6.8
	RE-3	33.96	89.02	7.3
	NIR-B	33.91	88.65	7.3
Current Work	RE-1	33.13	88.65	5.8
	RE-2	34.13	90.33	6.0
	RE-3	34.24	90.48	6.2
	NIR-B	34.16	90.14	6.3

Table 1. Comparison with previous results.

As evident from the results, our new approach achieves higher PSNR and SSIM values, indicating better overall quality and structural similarity to the target Sentinel-2 bands. The lower MAE further confirms the improved accuracy of our synthesized bands.

3.4 Qualitative Evaluation

Visual inspection of our results reveals a massive improvement over our previous work, particularly regarding spatial resolution preservation. Figure 2 showcases a side-by-side comparison of the target Sentinel-2 band, our previous result, and our current result for a representative scene.

Our new approach preserves fine spatial details that were lost in our previous work, resulting in synthesized bands that closely resemble the target Sentinel-2 data. While there is still some minor scale loss, the improvement is substantial and significantly enhances the utility of our augmented L30 product for various Earth Observation applications.

3.5 Ablation Study

To understand the impact of different SSL-pretrained backbones on our model's performance, we conducted an ablation study comparing two backbone architectures (ResNet50 and ResNet18) and two SSL algorithms (SimCLR and MOCOv2). Table 2 presents the quantitative results of this study.

Quantitatively, the performance across all configurations is relatively close. However, qualitative analysis reveals significant differences:

- MOCOv2-based models consistently produce higher quality results compared to SimCLR-based models.

Backbone	SSL	PSNR(dB)	SSIM	MAE (%)
RedEdge-1 Band				
ResNet50	MOCOv2	**33.13**	**88.65**	**5.8**
	SimCLR	32.96	88.02	6.0
ResNet18	MOCOv2	33.01	88.58	5.8
	SimCLR	32.91	87.95	6.2
RedEdge-2 Band				
ResNet50	MOCOv2	**34.13**	**90.33**	**6.0**
	SimCLR	33.78	89.71	6.8
ResNet18	MOCOv2	33.98	90.17	6.4
	SimCLR	34.09	90.00	6.5
RedEdge-3 Band				
ResNet50	MOCOv2	**34.24**	**90.48**	**6.2**
	SimCLR	33.89	89.84	7.1
ResNet18	MOCOv2	34.10	90.33	6.7
	SimCLR	34.23	89.99	6.7
NIR-Broad Band				
ResNet50	MOCOv2	**34.16**	**90.14**	**6.3**
	SimCLR	33.96	89.71	6.8
ResNet18	MOCOv2	34.09	90.03	6.5
	SimCLR	34.11	89.52	6.9

Table 2. Ablation study results.

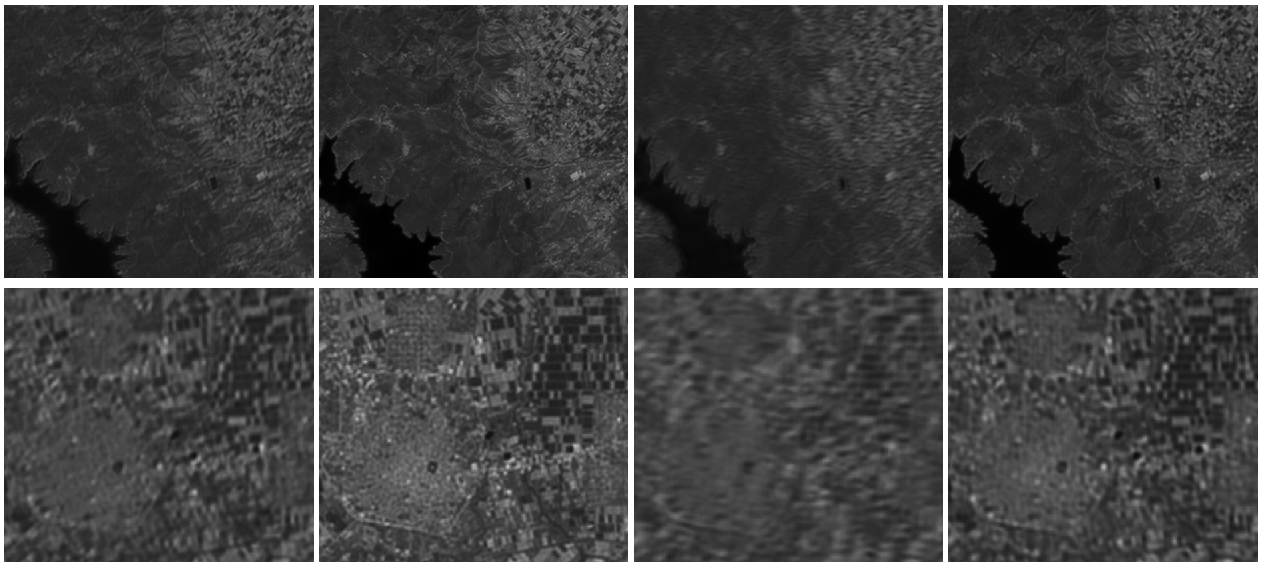
- SimCLR-based models occasionally generate minor artifacts such as double edges and fuzzy boundaries, not present in MoCov2 - based results.
- ResNet50 versions generate slightly higher resolution results, closer to the target resolution, compared to their ResNet18 counterparts.

Figure 3 illustrates these qualitative differences for a sample scene. Based on these results, we conclude that the MoCov2 pretrained ResNet50 backbone provides the best balance of quantitative performance and qualitative output quality for our task.

4. Conclusions

This work presents a significant advancement in the harmonization of Landsat and Sentinel-2 data, building upon our previous efforts to create a more unified and versatile EO dataset. Our improved approach, which integrates a fully SSL-pretrained encoder into a U-Net architecture, demonstrates substantial enhancements in both spectral fidelity and spatial resolution preservation. Our new method significantly reduces the scale loss observed in our previous work, preserving fine spatial details crucial for various EO applications. The quantitative results show improvements across all evaluation metrics, indicating better alignment with the target Sentinel-2 bands. Our ablation study reveals the effectiveness of MOCOv2-pretrained models, particularly when combined with the ResNet50 architecture.

While our results show significant progress, there remain opportunities for further improvement. Future work could explore the integration of additional contextual information, such as seasonal variations or geographical metadata, to further refine the band synthesis process. Additionally, investigating the applicability of this method to other satellite sensors could further expand its utility in the EO community.



(a) (b) (c) (d)

Figure 3. Ablation study. (a) ResNet50 SimCLR, (b) ResNet50 MoCo v2, (c) ResNet18 SimCLR, (d) ResNet18 MoCo v2. for two different scenes (top and bottom row respectively).

In conclusion, this work represents a meaningful step towards creating a more unified and versatile Earth Observation dataset. By bridging the gap between Landsat and Sentinel-2 data products, we contribute to the development of more robust, consistent, and long-term Earth Observation analyses, ultimately supporting better-informed decision-making in critical areas like climate change monitoring, land use management, and environmental conservation.

5. Acknowledgments

Part of this research was supported by the research project Bi-CUBES “Analysis-Ready Geospatial Big Data Cubes and Cloud-based Analytics for Monitoring Efficiently our Land & Water” funded by HFRI (grant: 03943).

References

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020a. A simple framework for contrastive learning of visual representations. *International conference on machine learning*, PMLR, 1597–1607.

Chen, X., Fan, H., Girshick, R., He, K., 2020b. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*.

Christie, G., Fendley, N., Wilson, J., Mukherjee, R., 2018. Functional map of the world. *CVPR*.

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., Justice, C., 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote sensing of environment*, 219, 145–161.

He, W., Yokoya, N., 2018. Multi-temporal sentinel-1 and-2 data fusion for optical image simulation. *ISPRS International Journal of Geo-Information*, 7(10), 389.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A. A., 2017. Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Karakizi, C., Gounari, O., Sofikiti, E., Begkos, G., Karantzalos, K., Symeonakis, E., 2023. Assessing the contribution of optical and sar data for fractional savannah woody vegetation mapping. *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 3118–3121.

Karakizi, C., Okujeni, A., Sofikiti, E., Tsironis, V., Psalta, A., Karantzalos, K., Hostert, P., Symeonakis, E., 2024. Mapping savannah woody vegetation at the species level with multispecral drone and hyperspectral EnMAP data. *arXiv preprint arXiv:2407.11404*.

Kganyago, M., Adjorlolo, C., Sibanda, M., Mhangara, P., Laneve, G., Alexandridis, T., 2022. Testing Sentinel-2 spectral configurations for estimating relevant crop biophysical and biochemical parameters for precision agriculture using tree-based and kernel-based algorithms. *Geocarto International*, 1–25.

Lowman, J., Zheng, K. L., Fraser, R., The, J. V. G., Valipour, M., 2024. SeeFar: Satellite Agnostic Multi-Resolution Dataset for Geospatial Foundation Models. *arXiv preprint arXiv:2406.06776*.

Ouzoun, M., Falagas, A., Gounari, O., Kandylakis, Z., Karakizi, C., Karantzalos, K., 2023. Per-parcel high-resolution mapping of critical crop-growth parameters with remote sensing. *Precision agriculture’23*, Wageningen Academic, 971–977.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, Springer, 234–241.

Segarra, J., Buchaillot, M. L., Araus, J. L., Kefauver, S. C., 2020. Remote sensing for precision agriculture: Sentinel-2 improved features and applications. *Agronomy*, 10(5), 641.

Stamford, J. D., Vialet-Chabrand, S., Cameron, I., Lawson, T., 2023. Development of an accurate low cost NDVI imaging system for assessing plant health. *Plant Methods*, 19(1), 9.

Stewart, A. J., Lehmann, N., Corley, I. A., Wang, Y., Chang, Y.-C., Braham, N. A. A., Sehgal, S., Robinson, C., Banerjee, A., 2023. Ssl4eo-l: Datasets and foundation models for landsat imagery.

Tsironis, V., Psalta, Athena El Saer, A., Karantzalos, K., 2024. Generating sentinel-2 additional bands from landsat 8/9 for hls dataset with deep convolutional networks. *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*.

Xie, Q., Dash, J., Huang, W., Peng, D., Qin, Q., Mortimer, H., Casa, R., Pignatti, S., Laneve, G., Pascucci, S. et al., 2018. Vegetation indices combining the red and red-edge spectral information for leaf area index retrieval. *IEEE Journal of selected topics in applied earth observations and remote sensing*, 11(5), 1482–1493.