Pre-Training and Fine-Tuning for Satellite Image-to-Image Translation

Daiki Kimura[1] , **Tatsuya Ishikawa**[1] , **Masanori Mitsugi**[2] , **Yasunori Kitakoshi**[3] ,
Takahiro Tanaka[2] , **Naomi Simumba**[1] , **Kentaro Tanaka**[4] , **Hiroaki Wakabayashi**[4] ,
Masato Sampei[4] and **Michiaki Tatsubori**[1]
[1] **IBM Research,** [2] **IBM Japan Ltd.,** [3] **IBM Japan Digital Services Company,** [4] **Space Shift Inc.**
daiki@jp.ibm.com

Abstract

Geospatial technologies are increasingly important for various applications worldwide, including vegetation monitoring. Collecting ground truth data for specific geospatial tasks are challenging and time-consuming. Recently, the foundation model research have been explored, then pre-training on large-scale data and fine-tuning for specific tasks are important components of this technique. Although this approach can enhance the performance in downstream tasks such as satellite image translation, directly fine-tuning models pre-trained on natural images like ImageNet is suboptimal for geospatial data due to the inherent domain differences. In this paper, we propose a novel image translation approach with pre-training on geospatial-specific data and data augmentation. We present a case study where our method achieved outstanding results in a competition for inferring normalized difference vegetation index images from synthetic aperture radar data of cabbage farms. Our approach outperformed other methods with a 31% higher score than the second-ranked team and a 44% higher score than the average of the top five teams.

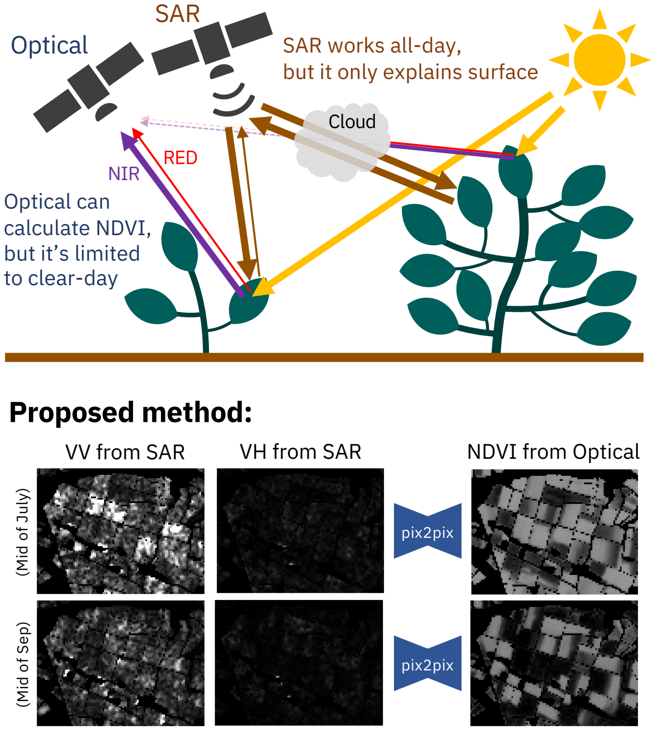

Figure 1: Overview image: The upper diagram represents the challenge, and the lower diagram illustrates the processing involved in the proposed method. While it is possible to obtain NDVI from optical satellites, a limitation is that they are only operational during clear daylight hours. On the other hand, SAR satellites offer all-day, but they can only provide information about surface shapes. Therefore, the proposed method aims to extract NDVI from SAR images. Note that, the NDVI values in July are higher (represented as whiter) due to the increased growth of vegetation.

1 Introduction

Climate change impacts vegetation distribution through rising temperatures, altered precipitation patterns, and shifting growing seasons. Monitoring vegetation changes is crucial for assessing ecosystem resilience and vulnerability. The Normalized Difference Vegetation Index (NDVI) is a valuable tool derived from optical satellite imagery, measuring reflected sunlight in the red and near-infrared bands. However, optical sensors have limitations like cloud cover and inability to capture data at night. Synthetic aperture radar (SAR) sensors offer advantages such as all-weather and day-night imaging. Estimating vegetation indices directly from SAR data can overcome optical imagery limitations, but mapping SAR data to vegetation indices presents challenges due to measurement differences. Advanced methodologies are needed for effective SAR data utilization. Image-to-image translation aims to transform input images while preserving visual characteristics. In vegetation monitoring, image-to-image translation can bridge the gap between SAR and vegetation index images. However, these techniques often require abundant training data, which is limited in the geospatial domain due to data collection challenges. Pre-training models on large-scale datasets like ImageNet [Deng *et al.*, 2009] does not translate well to geospatial data due to significant differences.

Geospatial-specific data and pre-training methods are essential.

This paper proposes a novel method using image-to-image translation, specifically the pix2pix framework [Isola *et al.*, 2017], to estimate NDVI from SAR images. Pre-training is conducted on a large dataset of Sentinel-2 optical satellite data. The model is then fine-tuned using paired SAR and optical image datasets from a targeted satellite. Additionally, we employ several data augmentation techniques during training and testing. The proposed method is compared to baselines on a specific dataset and evaluated in a public competition focusing on these challenges. The contributions of this paper are as follows:

- We propose a novel method for estimating NDVI from SAR satellite images using the pix2pix image-to-image translation framework which aims to providing a valuable tool for vegetation monitoring.
- We introduce pre-training and data augmentation techniques to enhance the accuracy and generalization capability for small size of dataset.
- We report that the proposed method outperforms other baselines and actually won the public competition [Space Shift Inc., 2023] among the 57 participants.

2 Related Work

2.1 Vegetation index

The Normalized Difference Vegetation Index (NDVI) is a numerical indicator used to assess and quantify the density and health of vegetation. NDVI measures the difference between the reflectance of near-infrared (NIR) and red light wavelengths, $\mathrm{NDVI} = \frac{\mathtt{NIR}-\mathtt{RED}}{\mathtt{NIR}+\mathtt{RED}}$. The values range from -1 to $+1$, where negative values indicate non-vegetated or non-photosynthetic surfaces, and positive values represent healthy and dense vegetation.

2.2 Alternative vegetation indices based on SAR

Several alternative vegetation indices based on SAR have been derived as fixed formulas such as Polarimetric Radar Vegetation Index (PRVI) and Radar Forest Degradation Index (RFDI) [Flores-Anderson *et al.*, 2019]. However, as the backscattering signal is sensitive to the physical structure of the object, the characteristics of SAR data for agricultural fields can depend strongly on the shapes, numbers, and canopy roughness of the targeted crops.

2.3 Machine learning-based approaches

Transfer learning to different domain is important research topic [Kimura *et al.*, 2013]. There is a study to estimate NDVI from SAR information [Roßberg and Schmitt, 2023]. However, as the available land use information is not about specific crop types but more abstract ones, e.g., grassland and forest, considering that the characteristics of the crops can significantly vary depending on the types, the practical usefulness of the approach may be limited. Contrary to the development of a globally applicable model with auxiliary information, in order to eliminate the differences of the crop types and growth phases, a set of hyperlocal and dynamic random forest models is proposed; each of which is trained for a specific crop field using relatively short period datasets and updated on the fly as new data becomes available [Pelta *et al.*, 2022]. However, the limitation on the dataset used in the training, particularly in shorter time period, may hinder to make use of potentially rich information contained in the long term observations to further improve the performance.

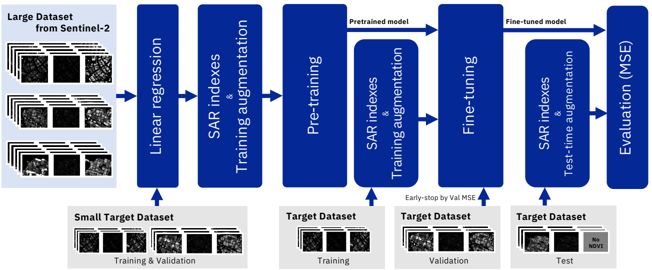

Figure 2: Proposed method: We have pre-training, fine-tuning, augmentation, linear regression for Sentinel-2 images, and part to convert indexes from SAR.

2.4 Image-to-image translation

In recent years, generative adversarial network (GAN) techniques in computer vision are very common and applied to many applications [Sampath *et al.*, 2021; Kimura *et al.*, 2020]. Then, the advent of image-to-image translation techniques has revolutionized the field of remote sensing image analysis [Jozdani *et al.*, 2022].

Cross-sensor transfer applications leveraging image-to-image translation in remote sensing have been discussed mainly focused on generating colored optical images from SAR data for genera scenes (e.g., [Zhang *et al.*, 2021] and [Ji *et al.*, 2021]). When we consider to apply these techniques to the image translation of SAR to NDVI and achieve practical estimation accuracy for NDVI of specific crops, we should have a way to make effective use of limited available dataset.

3 Proposed Method

We present our proposed method for estimating NDVI from SAR satellite images using the image-to-image translation framework. Additionally, we employ data augmentation techniques in training and test-time to enhance the diversity of the data and enable the model to handle different climatic conditions and SAR-specific characteristics.

3.1 Model

We use pix2pix model [Isola *et al.*, 2017] as an image-to-image translation method in this proposed method. The pix2pix is one of conditional GAN, and it consists of a generator network G and a discriminator network D, trained in an adversarial manner. The objective of this conditional GAN is,

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G), \quad (1)$$

$$\begin{aligned} \mathcal{L}_{cGAN}(G, D) =& \mathbb{E}_{x,y}[\log D(x, y)]+ \\ & \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))], \quad (2) \\ \mathcal{L}_{L1}(G) =& \mathbb{E}_{x,y,z}\left[\|y - G(x, z)\|_1\right], \quad (3) \end{aligned}$$

where G tries to minimize the objective against an adversarial D that tries to maximize it. We use U-Net [Ronneberger *et al.*, 2015] as a backbone of the generator network.

3.2 Pre-training images from Sentinel-2

We pre-train a generator network on optical data from Sentinel-2 satellite, which has relevant spectral bands for NDVI calculation, to improve our method for estimating NDVI from SAR data. We also utilize a linear regression model to correct the value ranges between Sentinel-2 and the targeted dataset, ensuring consistency and compatibility for training and evaluation. This improves the accuracy and reliability of the estimation and shows the value of data preprocessing and multi-sensor integration for vegetation monitoring and analysis.

3.3 Fine-tuning images from targeted satellite

Sentinel-2 and the targeted satellite data have variations and challenges due to different acquisition parameters and sensor characteristics. We fine-tune the model on the targeted satellite data to bridge the gap and align the datasets for NDVI estimation. Fine-tuning transfers the knowledge from the pretraining phase on Sentinel-2 data to the targeted satellite data, improving the accuracy and reliability of the estimation.

3.4 Data augmentation

We use data augmentation techniques to increase the diversity and robustness of the SAR-NDVI pairs dataset. We propose three types of data augmentation: SAR indexes transformation, training augmentation, and test-time augmentation.

1. **SAR indexes transformation:** We apply multiple transformations to generate indexes from SAR data, such as clipping, scaling, and simple equations. These transformations increase the variability of SAR data, improving the model's performance on diverse data conditions.

2. **Training augmentation:** We generate additional samples by applying random geometric transformations, such as, rotation, scaling, and flipping. This improves the model's generalization to unseen data.

3. **Test-time augmentation:** We apply data augmentation techniques to the test dataset, such as flipping, and value multiplying. After augmented samples are generated, it obtain predictions from each sample and aggregate the predictions to generate a final prediction. This improves the model's accuracy and reliability on SAR data variations.

4 Experiments

In this paper, we utilized SAR images obtained from the Sentinel-1 satellite, NDVI images obtained from the Sentinel-2 satellite as pre-training data, and NDVI images acquired from the Planet's satellite as target data.

Table 1: Search Conditions for downloading Sentinel-1 images

Parameter	Filter Value
File Type	L1 Detected High-Res Dual-Pol
Beam Mode	IW
Direction	Descending
Polarization	VV, VH

4.1 Datasets - SAR images

The SAR images we obtained are from Sentinel-1 operated by European Space Agency (ESA). We downloaded the SAR images from Alaska Satellite Facility Data Search under the conditions shown in Table 1. All available data during the experimental period were used, then 72 SAR images were taken.

We performed ortho correction to calculate σ_0 values for VV and VH polarizations via Sentinel Application Platform, provided by ESA. Then we computed resampling, region cropping and coordinate system tranformation by Geospatial Data Abstraction Library [GDAL/OGR contributors, 2023]. After all pre-processing is completed, we generated GeoTiff files containing the σ_0 values which is with 10 meters resolution and EPSG:32654 coordinate system.

4.2 Datasets - NDVI (Sentinel-2)

Sentinel-2 provides earth observation covering most of the lands by multi-spectral optical imagery consisting of 13 spectral bands with a resolution of 10 to 60 meters. The data product archive is made freely accessible and available and such a data distribution policy of the program opened an active research field of applying computer vision and image analysis techniques to satellite datasets.

The top-of-atmosphere reflectance data products are obtained via the Google Earth Engine as Harmonized Sentine-2 MSI: Multispectral Instrument, Level-1C. The NDVI values are calculated by equation (**??**) with the `NIR` and `RED` bands and have the same spatial resolution, coordinates, and coverage with the SAR dataset.

4.3 Datasets - NDVI (Planet)

Planet Labs, Inc. (Planet) operates several optical satellites. The satellites capture observations at fixed locations once per day, providing imagery in four bands (blue, green, red, and near-infrared) with a resolution of 3 meters. Planet offers images observed from three types of satellites: PS2, PS2.SD and PSB.SD. However, it is important to note that all product archives are not available free of charge.

In this study, we used atmospherically corrected Planet's satellite data. We obtained as much cloud-free observation data as possible during the experimental period, then we obtained 42 images from Planet. The NDVI image was resampled from resolution of 3 to 10 meters using nearest neighbor algorithm. We also filled zero value at non-farm regions of targeted vegetation, the farm regions were downloaded from open data provided by the Japanese Ministry of Agriculture, Forestry and Fisheries.

Table 2: Coordinates for area of interested in this study

Area	North	West	South	East
Training area	36.530	138.499	36.524	138.506
Validation area	36.515	138.472	36.508	138.483
Test area	36.544	138.483	36.538	138.493

Table 3: Ablation study: Comparison by mean-square error on validation dataset for the proposed method and some ablation methods. The check-mark indicates the use of the component.

	Pre-training	Fine-tuning	SAR indexes	Training aug	Test-time aug	MSE
Ours	✓	✓	✓	✓	✓	**.008661**
Ablation 1	-	✓	✓	✓	✓	.020866
Ablation 2	✓	-	✓	✓	✓	.009835
Ablation 3	✓	✓	-	✓	✓	.010978
Ablation 4	✓	✓	✓	-	✓	.018021
Ablation 5	✓	✓	✓	✓	-	.009048

4.4 Experimental details

The area of interest for this study was focused on Tsumagoi village in Gunma prefecture, Japan. We have training area, validation area, and test area; the coordinates for each area are shown in Tab. 2. We selected a cabbage farm as the specific target vegetation for our study, and we applied masking non-farm areas from all NDVI images. The experimental period for our study ranged from May 2017 to November 2017.

As described in Section 3, our approach involved pre-training on the Sentinel-2 dataset and fine-tuning on the images from Planet (dataset of training area in Tab. 2 is used). For the SAR indexes, we utilized clipped VV, 99.5%-range clipped VV, original VH, clipped VH, PRVI, RFDI, RVI4S1, RVI, $\mathtt{VH} + \mathtt{VV}$, and $\mathtt{VH} - \mathtt{VV}$, where the vegetation indexes are calculated by following equations [Pelta *et al.*, 2022].

$$\mathrm{PRVI} = \left(1 - \frac{\mathtt{VV}}{\mathtt{VH} + \mathtt{VV}}\right)\mathtt{VH} \tag{4}$$

$$\mathrm{RFDI} = \frac{\mathtt{VV} - \mathtt{VH}}{\mathtt{VH} + \mathtt{VV}} \tag{5}$$

$$\mathrm{RVI4S1} = \sqrt{\frac{\mathtt{VV}}{\mathtt{VH} + \mathtt{VV}}}\frac{4\mathtt{VH}}{\mathtt{VH} + \mathtt{VV}} \tag{6}$$

$$\mathrm{RVI} = \frac{4\mathtt{VH}}{\mathtt{VH} + \mathtt{VV}} \tag{7}$$

We trained 5,000 epochs in the pre-training phase, and we did early stopping based on validation mean-squared error in the fine-tuning phase.

4.5 Result - Ablation study

Table 3 presents the results of the ablation study conducted on the proposed method. Our findings demonstrate that the proposed method outperforms all the ablation models, each of which excludes a specific component. Notably, the accuracy of the Ablation 1 method, which excludes pre-training, was significantly lower; it means pre-training is the critical important in our approach. Additionally, the results emphasize the significance of training augmentation, as evidenced by the improved performance compared to Ablation 3. Furthermore, the inclusion of fine-tuning proved to be beneficial, considering the differences between images acquired from Sentinel-2 and those obtained from the targeted satellite. This observation further substantiates the positive impact of fine-tuning on enhancing accuracy in our proposed method. Our ablation study reinforces the importance of pre-training, training augmentation, and fine-tuning in achieving superior accuracy for NDVI estimation from SAR satellite imagery. These findings validate the effectiveness of our proposed approach in addressing the challenges posed by diverse data sources and contribute to advancing the field of SAR-based vegetation monitoring.

4.6 Result - Winning in public competition

At same time, the proposed method demonstrated outstanding performance in a public competition [Space Shift Inc., 2023], which is our approach secured the **1st-place position** among the 57 participants. In the evaluation, our method surpassed other state-of-the-art approaches by achieving significantly lower mean square error in the test images obtained from the Planet. More specifically, our method marked $\mathtt{MSE}_{\mathtt{test}} = 0.005353$, it was 31% better than the second team, and 44% better than the average of others in top-5. This remarkable achievement further attests to the robustness and accuracy of our proposed method.

By achieving the top rank in the public competition, our method not only showcases its technical advancements but also establishes its practical significance in real-world applications. The recognition received from this competition highlights the potential impact of our method in the field of SAR-based vegetation monitoring and its ability to contribute to broader environmental management efforts. It solidifies our confidence in the effectiveness and competitiveness of our approach and strengthens its position as a cutting-edge solution in remote sensing and environmental studies.

5 Concluding Remarks

In this study, we proposed a method to estimate NDVI from SAR satellite imagery using pre-training with Sentinel-2 data. The method was validated against NDVI data from Planet, focusing on the Tsumagoi-area in Japan. Our results demonstrate that the proposed method outperforms other methods and marked as top-1 in a competition with 57 participants.

Future research can explore the application of the proposed method in other geographical regions and investigate additional data sources, such as Harmonized Landsat and Sentinel-2 [Claverie *et al.*, 2018], to further enhance its performance. This work aims to integrate SAR satellite imagery into vegetation monitoring, contributing to a better understanding and management of ecosystems in environmental and agricultural contexts.

References

[Claverie *et al.*, 2018] Martin Claverie, Junchang Ju, Jeffrey G Masek, Jennifer L Dungan, Eric F Vermote, Jean-Claude Roger, Sergii V Skakun, and Christopher Justice. The harmonized landsat and sentinel-2 surface reflectance data set. *Remote sensing of environment*, 219:145–161, 2018.

[Deng *et al.*, 2009] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[Flores-Anderson *et al.*, 2019] Africa Ixmucane Flores-Anderson, Kelsey E. Herndon, Rajesh Bahadur Thapa, and Emil Cherrington. *The SAR Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation*. NASA, Washington, DC, USA, 2019.

[GDAL/OGR contributors, 2023] GDAL/OGR contributors. *GDAL/OGR Geospatial Data Abstraction software Library*. Open Source Geospatial Foundation, 2023.

[Isola *et al.*, 2017] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017.

[Ji *et al.*, 2021] Guang Ji, Zhaohui Wang, Lifan Zhou, Yu Xia, Shan Zhong, and Shengrong Gong. Sar image colorization using multidomain cycle-consistency generative adversarial network. *IEEE Geoscience and Remote Sensing Letters*, 18(2):296–300, 2021.

[Jozdani *et al.*, 2022] Shahab Jozdani, Dongmei Chen, Darren Pouliot, and Brian Alan Johnson. A review and meta-analysis of generative adversarial networks and their applications in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 108:102734, 2022.

[Kimura *et al.*, 2013] Daiki Kimura, Ryutaro Nishimura, Akihiro Oguro, and Osamu Hasegawa. Ultra-fast multimodal and online transfer learning on humanoid robots. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 165–166, 2013.

[Kimura *et al.*, 2020] Daiki Kimura, Subhajit Chaudhury, Minori Narita, Asim Munawar, and Ryuki Tachibana. Adversarial discriminative attention for robust anomaly detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, March 2020.

[Pelta *et al.*, 2022] Ran Pelta, Ofer Beeri, Rom Tarshish, and Tal Shilo. Sentinel-1 to ndvi for agricultural fields using hyperlocal dynamic machine learning approach. *Remote Sensing*, 14(11):2600, 2022.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[Roßberg and Schmitt, 2023] Thomas Roßberg and Michael Schmitt. A globally applicable method for ndvi estimation from sentinel-1 sar backscatter using a deep neural network and the sen12tp dataset. *PFG*, 2023.

[Sampath *et al.*, 2021] Vignesh Sampath, Iñaki Maurtua, Juan Jose Aguilar Martin, and Aitor Gutierrez. A survey on generative adversarial networks for imbalance problems in computer vision tasks. *Journal of big Data*, 8:1–59, 2021.

[Space Shift Inc., 2023] Space Shift Inc. Competition page for satellite images analysis. https://connpass.com/event/264630/, 2023.

[Zhang *et al.*, 2021] Hongyan Zhang, Yiyao Song, Chang Han, and Liangpei Zhang. Remote sensing image spatiotemporal fusion using a generative adversarial network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(5):4273–4286, 2021.