Land Use Prediction using Electro-Optical to SAR Few-Shot Transfer Learning

Marcel Hussing   **Karen Li**   **Eric Eaton**
University of Pennsylvania
{mhussing, karentli, eeaton}@seas.upenn.edu

Abstract

Satellite image analysis has important implications for land use, urbanization, and ecosystem monitoring. Deep learning methods can facilitate the analysis of different satellite modalities, such as electro-optical (EO) and synthetic aperture radar (SAR) imagery, by supporting knowledge transfer between the modalities to compensate for individual shortcomings. Recent progress has shown how distributional alignment of neural network embeddings can produce powerful transfer learning models by employing a sliced Wasserstein distance (SWD) loss. We analyze how this method can be applied to Sentinel-1 and -2 satellite imagery and develop several extensions toward making it effective in practice. In an application to few-shot Local Climate Zone (LCZ) prediction, we show that these networks outperform multiple common baselines on datasets with a large number of classes. Further, we provide evidence that instance normalization can significantly stabilize the training process and that explicitly shaping the embedding space using supervised contrastive learning can lead to improved performance.

1   Introduction

The United Nations has estimated that a large proportion of the Sustainable Development Goal indicators can be measured through use of geospatial data (Arnold et al., 2019). As an example, the goal of land use and land cover mapping is to measure the health of populations, urban areas, and ecosystems over time. Since urban areas are responsible for approximately 70% of the world's energy-related CO_2 emissions, tracking the development of cities plays a crucial role in climate change mitigation and adaptation (Lucon and Vorsatz, 2014). Interest in the topic has given rise to the collection of different forms of satellite data as a means to map out geospatial regions of the earth. One common form are hyperspectral electro-optical (EO) images, for which there exists an abundance of large labeled datasets. However, EO images cannot capture many factors of our earth's development; EO sensors are blocked by clouds, limited by the day-and-night cycle, and subject to distortion under various weather conditions. In situations that require imaging over extended time periods, researchers have started working with Synthetic Aperture Radar (SAR) data. SAR data is collected by a satellite sending out radar energy and recording the amount reflected back from the earth, making it insusceptible to many of the downsides of EO imaging. Unlike EO datasets, which can be easily interpreted and labeled through crowdsourcing, SAR data requires trained expertise for interpretation. As a result, labeled SAR datasets are both limited in quantity and costly to generate.

One approach to extract information from geospatial data with few available labels is transfer machine learning. The idea is to pre-train a neural network on a source domain with plentiful data (i.e., EO) and then fine-tune on a target domain for which little data is available (SAR). In this paper, we focus on a recent method by Rostami et al. (2019a,b), which uses sliced Wasserstein distance to align the distributions of the two modalities in a shared embedding space, and promote knowledge transfer between the classification of EO and SAR images. In doing so, one is not only able to use

the few labeled SAR data, but also make use of any available unlabeled data. The original work is limited in that it only does analysis on binary classification using a relatively simple dataset—ships on uncluttered backgrounds. More realistic settings typically involve multiple classes and cluttered scenes, presenting a substantially harder challenge. Further, the original papers do not explore what properties of the embedding space might lead to better alignment, required for these real-world tasks.

To gain insight for how to apply EO-to-SAR transfer in practice, this paper studies these issues of multi-class classification and properties of the embedding space. We performed the analysis using Sentinel-1 (EO) and Sentinel-2 (SAR) satellite data, and provide evidence that:
- SWD embedding alignment for EO-to-SAR transfer can be scaled to the multi-class setting,
- Applying instance normalization leads to more stable training and better performance, and
- Applying constrastive learning improves transfer performance of the SWD approach.

2   Data & Methods

So2Sat Dataset   Focusing on local climate zone classification, our analysis uses the So2Sat LCZ42 dataset (Zhu et al., 2020), which consists of LCZ labels for approximately half a million Sentinel-1 (EO) and Sentinel-2 (SAR) satellite image patches over 42 cities across the world. Each patch has a resolution of 32×32 pixels, and is labeled with one of 17 total classes (ten urban zone types, seven natural zone types). Each zone type is identified by a unique combination of surface structure, cover, and human activity at the local scale (10^2 to 10^4 m). Originally designed for urban heat island research, LCZ classification also provides significant information, such as the pervious surface fraction and surface albedo values of each zone, in applications to green infrastructure, population assessment, and ecosystem processes (Stewart and Oke, 2020).

Transfer Learning with Sliced Wasserstein Distance (SWD-transfer)   Our work builds on a recent method that trains a Y-shaped neural network (Rostami et al., 2019b) to transfer knowledge from the EO to the SAR domain. The network consists of two convolutional neural networks with identical architectures as EO and SAR encoders. The outputs of the two encoders map to a shared embedding space that is then fed into a classification network. First, the network is trained to classify a large amount of EO data to obtain a discriminative embedding produced by the EO encoder. Then, using a small amount of labeled and a large amount of unlabeled SAR data, the network is trained to align the distribution produced by the EO encoder with that created by the SAR encoder. The minimization of discrepancy between the encoded distributions is achieved via minimization of the sliced Wasserstein distance. SWD computes an approximation of the distance between the two distributions by projecting them along a set of random vectors and computing the mean over the 1-dimensional Wasserstein distances of those random projections. More formally, SWD computes

$$\mathcal{D}^2(p_{\mathcal{S}}, p_{\mathcal{T}}) \approx \frac{1}{L} \sum_{l=1}^{L} \sum_{m=1}^{M} |\langle \gamma_l, \phi(\mathbf{x}_{s_l[i]}^{\mathcal{S}}) \rangle - \langle \gamma_l, \psi(\mathbf{x}_{t_l[i]}^{\mathcal{T}}) \rangle|^2 \ , \tag{1}$$

where $\mathcal{S}$ and $\mathcal{T}$ are the source and target domain respectively, L is the number of random projections γ, $\phi(\cdot)$ and $\psi(\cdot)$ are the encoding functions, and $s_l[i]$ and $t_l[i]$ are sorting indices. The full approach (see Rostami et al. (2019b) for details) optimizes the objective

$$\min_{\mathbf{u},\mathbf{v},\mathbf{w}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(h_{\mathbf{w}}(\phi_{\mathbf{v}}(\mathbf{x}_i^s)), \mathbf{y}_i^s) + \sum_{i=1}^{O} \mathcal{L}(h_{\mathbf{w}}(\psi_{\mathbf{u}}(\mathbf{x}_i^t)), \mathbf{y}_i^t)$$
$$+ \alpha \mathcal{D}^2(\phi_{\mathbf{v}}(p_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}})), \psi_{\mathbf{u}}(p_{\mathcal{T}}(\mathbf{X}_{\mathcal{T}}))) + \beta \sum_{j=1}^{k} \gamma \mathcal{D}^2(\phi_{\mathbf{v}}(p_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}|C_j)), \psi_{\mathbf{u}}(p_{\mathcal{T}}(\mathbf{X}_{\mathcal{T}})|C_j)) \ , \tag{2}$$

where the first two terms are the classification losses for each domain with $\mathcal{L}$ being the cross-entropy loss, the third term is the distribution alignment term over the unlabeled data, and the final term is a class conditioned alignment of distribution to ensure correct matching of class modes.

Normalizing Embedding Spaces   In applying SWD-transfer to the Sentinel-1 and Sentinel-2 data, we found that it is sensitive to the scale of the input data. We conjecture that not only is the scale of the input of significance, but so is the scale across *all* layers of the neural network. Normalization is a common technique in neural networks to ensure that layer outputs with diverse ranges will have a proportionate impact on the prediction. The original work suggests to use batch normalization (Ioffe

and Szegedy, 2015), which computes mean and variance for every training batch and uses them for normalization. However, the wide range of the spectrum of the SAR data might distort estimates of those distribution measures. Instead, we consider that the specific nature of SAR data might benefit from using instance normalization (Ulyanov et al., 2016), which computes a per-data-point mean and variance to prevent instance-specific mean and covariance shifts.

Supervised Contrastive Learning We hypothesize that explicitly modeling the embedding space that is used for alignment can improve the performance of SWD-based transfer. To enforce a specific structure in the shared embedding space, we explore the impact of a recent supervised constrastive learning approach called SupCon (Khosla et al., 2020). Intuitively, contrastive learning enforces that data points from the same class are mapped closer together in embedding space and points from classes that are different ought to be pushed away from each other. By incorporating the contrastive loss into SWD-transfer, we hope to obtain a more discriminative embedding space and cleaner class-conditioned distribution boundaries in problems with large numbers of classes. Concretely, we add the following term to the optimization in Equation 2 when training the EO classifier:

$$\mathcal{L}_C = \sum_{i \in I} \frac{1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(\mathbf{z}_i \times \mathbf{z}_p / \tau)}{\sum_{a \in A(i)} \exp(\mathbf{z}_i \times \mathbf{z}_a / \tau)} \ , \tag{3}$$

where $\mathbf{z}_k = \phi_{\mathbf{v}}(\mathbf{x}_k)$, $i \in I$ is the index of an augmented data sample (e.g. rotation or translation of the original image) from a multiviewed batch, $A(i) \equiv I \setminus i$, and $P(i) \equiv \{p \in A(i) : \tilde{\mathbf{y}}_p = \tilde{\mathbf{y}}_i\}$ is the set of all positives in the batch distinct from i. A multiviewed batch is created by applying two random augmentations to every sample in a batch yielding a new batch of twice the original size.

3   Experiments & Results

We consider three configurations of the So2Sat dataset to show results on different levels of difficulties: four classes of different regions that contain *low-rise* buildings, all 10 *urban* classes, and the full dataset with *all* 17 classes. For the EO images, we compute an RGB representation of the image, and for the SAR data, we use the real and imaginary parts of the unfiltered VV channel. Further, we compare three different methods: (a) training a single neural network on the EO data and then *fine-tuning* it on the SAR domain, (b) training a classification network on the available labeled *SAR data only*, (c) and *SWD-transfer* (Rostami et al., 2019b), as described in Section 2. To evaluate effectiveness in the few-shot regime, we limit the number of available labeled SAR datapoints to *32*, *128*, *1024*, or use the *full* dataset. According to standard practice, we use train, validation, and test splits of the data, and report mean and standard error of the test accuracy over five random seeds.

Scaling Sliced Wasserstein Transfer to Multi-Class Problems The first question that we want to answer is whether or not the SWD-transfer method scales to multi-class problems. For this, we consider the three dataset settings (low-rise, urban, all) to evaluate whether adding more classes to the problem decreases the relative performance compared to the baselines. The results are depicted in Figure 1. First, we see that across all dataset settings, SWD-transfer outperforms the baselines in the low-data regime. With more classes and large amounts of data available, SAR-only training and fine-tuning outperform the SWD-transfer method. The results provide evidence that the SWD-transfer is indeed able to generalize to challenges with a large number of classes. However, the
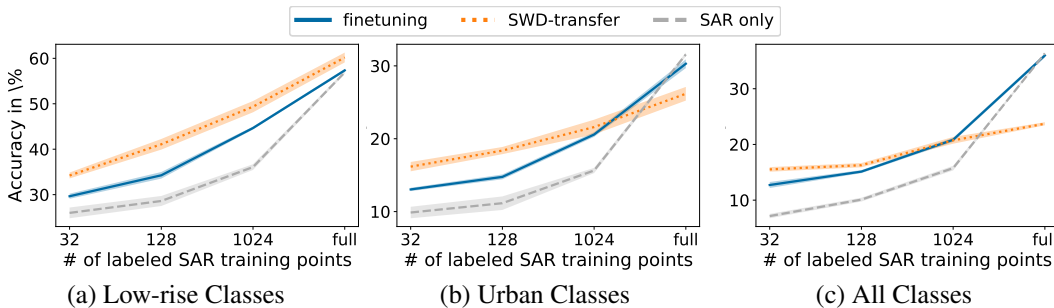

Figure 1: Mean test accuracy and standard error for varying numbers of classes. SWD-transfer outperforms the baselines in the low data regime but is overtaken when the full dataset is available.
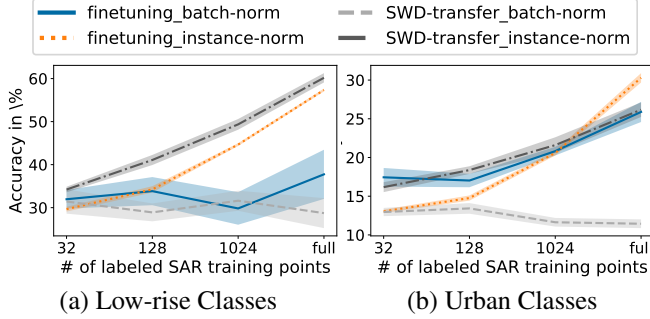
Figure 2: Mean test accuracy and standard error of batch and instance normalization for fine-tuning and SWD transfer. Instance-normalization consistently stabilizes SWD-transfer training and leads to improved performance on both methods.
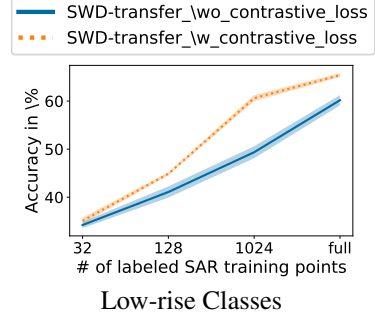
Figure 3: Mean test accuracy and standard error when pretraining without or with contrastive loss. The latter increases performance.

performance gap decreases as the number of classes increases, which is likely due to the learned EO embedding not being sufficiently discriminative.

The Importance of Normalization in Sliced Wasserstein Transfer This experiment explores how layer normalization can affect the performance of SWD-based transfer. Figure 2 compares the performance of the SWD-transfer model and the fine-tuning approach when trained with batch or instance normalization. In most cases, SWD-transfer with batch normalization is not better than random guessing and instance normalization consistently stabilizes training and outperforms its counterpart. Additionally, instance normalization can improve the performance of the fine-tuning method. While in many cases the fine-tuning method works well with batch normalization, SWD-based transfer seems to *require* instance normalization. This suggests that instance normalization is necessary in order for the approach to learn the embedding distribution alignment for SAR data.

Embedding Modeling during Pretraining using Contrastive Learning Finally, we analyze the effect of modeling the shared embedding space in order to explicitly obtain more discriminative embeddings. We applied the SupCon method (Section 2) to the EO training cycle, while the SAR training remained unchanged from the original SAR-transfer formulation. Results on the low-rise dataset setting (Figure 3) shows that the SupCon-pretrained SWD-based transfer network outperforms the original SWD-transfer method across all data regimes. This presents evidence that shaping the internal representations of the neural network via contrastive learning facilitates the alignment of the distributions by creating more discriminative embeddings.

4   Conclusion and Future Work

We identified several shortcomings of the original SAR-transfer method when applied to more realistic SAR data, and developed enhancements to compensate, moving towards EO-to-SAR transfer in more practical settings. The additions of instance normalization and supervised contrastive learning provide significant improvements on the base method, bringing us closer to deploying the approach in the wild. However, we highlight that none of the results achieve the very high accuracies required for automated land cover mapping, leaving much room for improvement. In the future, applying semi-supervised contrastive learning during the SAR training phase is an interesting idea to explore.

Acknowledgements

References

Stephan Arnold, Jun Chen, and Olav Eggers. Global and complementary (non-authoritative) geospatial data for sdgs: Role and utilisation, 2019. URL `https://ggim.un.org/documents/Report_Global_and_Complementary_Geospatial_Data_for_SDGs.pdf`.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France, 07–09 Jul 2015. PMLR.

Prannay Khosla, Piotr Terwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, 2020. doi: https://doi.org/10.48550/arXiv.2004.11362.

Oswaldo Lucon and Diana Vorsatz. *Climate Change 2014: Mitigation of Climate Change*, chapter Chapter 9 - Buildings. Cambridge: Cambridge University Press, 2014.

Mohammad Rostami, Soheil Kolouri, Eric Eaton, and Kyungnam Kim. Deep transfer learning for few-shot sar image classification. *Remote Sensing*, 11(11), 2019a.

Mohammad Rostami, Soheil Kolouri, Kyungnam Kim, and Eric Eaton. Sar image classification using few-shot cross-domain transfer learning. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 907–915, 2019b. doi: http://dx.doi.org/10.1109/CVPRW.2019.00120.

Iain Stewart and Tim Oke. Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society*, 93:1879–1900, 2020. doi: https://doi.org/10.1175/BAMS-D-11-00019.1.

Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Instance normalization: The missing ingredient for fast stylization. *CoRR*, abs/1607.08022, 2016. URL `http://arxiv.org/abs/1607.08022`.

Xiao Xiang Zhu, Jingliang Hu, Chunping Qiu, Yilei Shi, Jian Kang, Lichao Mou, Hossein Bagheri, Matthias Haberle, Yuansheng Hua, Rong Huang, Lloyd Hughes, Hao Li, Yao Sun, Guichen Zhang, Shiyao Han, Michael Schmitt, and Yuanyuan Wang. So2sat lcz42: A benchmark data set for the classification of global local climate zones. *IEEE Geoscience and Remote Sensing Magazine*, 8(3):76–89, 2020. doi: http://dx.doi.org/10.1109/MGRS.2020.2964708.