LEARNING SOURCE DOMAIN REPRESENTATIONS FOR ELECTRO-OPTICAL TO SAR TRANSFER

Boya Zeng, Marcel Hussing & Eric Eaton
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA
{boyazeng,mhussing,eeaton}@seas.upenn.edu

ABSTRACT

Embedding distribution alignment is an approach to transfer knowledge from label-abundant electro-optical (EO) images to the label-scarce synthetic aperture radar (SAR) modality. However, this approach assumes that it is possible to learn a useful and discriminative EO representation via a neural network. In this work, we study the properties of such a representation. We analyze a recent result showing that supervised contrastive learning can improve transfer performance and find that its reduction of the effective dimension of the embedding is crucial to successful transfer. We then show that directly optimizing for this property can yield even better down-stream accuracy. Finally, we show that the powerful representation of an EO foundation model is insufficient for alignment due to its generality, but that additional representation learning can recover alignment performance.

1 INTRODUCTION

Satellite imagery has emerged as a crucial mechanism to monitor the global environment and support reasoning about the state of our planet on a large scale (Kramer et al., 2002). Such awareness enables governments to enact preventative measures in cases of imminent natural disasters (Voigt et al., 2007), and to study and impede the effects of climate change (Yang et al., 2013). This imagery is collected across a wide range of data modalities, from multi-spectral electro-optical (EO) images to radar and other technologies. In particular, synthetic aperture radar (SAR) data has proven useful due to its ability to penetrate atmospheric obstacles (Koo et al., 2012; Cooke & Scott, 2019), such as extreme weather and lighting conditions, that affect EO images (Kim et al., 2021). Still, challenges such as speckle noise, geometric distortion, and absence of color can render SAR images difficult to interpret for humans (Zhang et al., 2023). Consequently, it would be desirable to employ machine learning to analyze SAR data, but this process is complicated by the relative lack of labeled SAR data. In addition, the restricted nature of many SAR technologies, coupled with interpretation difficulties, makes solutions such as crowd-sourcing SAR labels currently infeasible.

To address this problem, leveraging the abundance of labeled EO data, cross-modal EO-to-SAR transfer learning has achieved competitive performance on SAR classification (Rostami et al., 2019; Jeong et al., 2021; Tai et al., 2022). These methods extract useful information from plentifully labeled EO data and transfer this information to the SAR domain using few labeled SAR points. Such transfer can occur via fine-tuning (Zhang et al., 2018), embedding distribution alignment (Long et al., 2015), or adversarial training (Ganin et al., 2016). While few-shot methods have been extensively studied (Sun et al., 2021; Jeong et al., 2021; Tai et al., 2022), understanding the EO and SAR representation spaces is crucial to the development of new methods. Our work studies the impact of EO representations on distribution alignment methods, such as maximum mean discrepancy (MMD) (Gretton et al., 2006) and sliced Wasserstein distance (SWD) minimization (Rostami et al., 2019), which align the internal embedding of a SAR network with that of a high-performance EO model.

Hussing et al. (2022) showed that pretraining EO models with supervised contrastive learning (SupCon) (Khosla et al., 2020) can improve SAR transfer using SWD alignment. However, the work lacks analysis on the exact characteristics induced onto the EO embedding space. Our contribution is an extensive analysis of the embedding properties that allow for effective distribution alignment.

We identify contrastive learning's ability to align data as the key to improving transfer. The improved accuracy strongly correlates with the effective dimension of the learned EO embeddings, and we show that decreasing this dimension directly yields even higher performance. At this point, one might hypothesize that the high-performance neural network EO embeddings might not even be needed for performant transfer. We provide contrary evidence by showing that using Gaussian mixture models as a substitute is insufficient to achieve maximum performance. Lastly, we examine whether foundation models might substitute for the EO neural network, employing a recent foundation model for earth data (Jakubik et al., 2023). We hypothesize that the powerful backbone might yield even better embeddings for transfer but find that its generality is, in fact, harmful to task-specific alignment. The useful embedding properties we identified can be recovered in this model, but the resulting performance is equivalent to that of small neural network.

2 Setting and Methodology

We consider a setting with access to extensive labeled EO data (X^S, Y^S), limited labeled SAR data (X^T, Y^T), and extensive unlabeled SAR data $\tilde{X}^T$.

2.1 Distribution Alignment

We use the EO-to-SAR neural network from Rostami et al. (2019) with two identical CNN encoders ϕ^S, ϕ^T for the EO and SAR domains and a shared classifier ψ (illustration in Appendix A). Let $Z^S = \{z_i^S | i = 1, 2, ..., |X^S|\} = \phi^S(X^S)$ denote the internal representation produced by the EO encoder and, similarly, $Z^T = \phi^T(X^T \cup \tilde{X}^T)$ is the representation produced by the SAR encoder. Further, let Z_c^S and Z_c^T denote the same embeddings conditioned on the data being from class c. We 1) pretrain the EO encoder and classifier with labeled EO data $(X^S,\ Y^S)$ and freeze the encoder, and 2) align the EO and SAR model's overall embedding distributions Z^S and Z^T and the class-conditioned embedding distributions Z_c^S and Z_c^T using unlabeled and labeled SAR data respectively. Distribution matching is done via minimization of either of the following metrics, SWD or MMD.

Sliced Wasserstein Distance is a fast approximation to the Wasserstein distance, a measure for optimal transport of probability distributions. It is calculated using the average Wasserstein distance between many one-dimensional projections of the two distributions (Rostami et al., 2019) as

$$\text{SWD}^2(Z^S, Z^T) = \frac{1}{H} \sum_{h=1}^{H} \sum_{i=1}^{m} \left| \gamma_h \cdot z_{s_h[i]}^S - \gamma_h \cdot z_{t_h[i]}^T \right|^2 \quad , \tag{1}$$

where H is the number of random projections γ, m is the EO and SAR alignment batch size, and s_h and t_h are the lists of indices that sort $\{\gamma_l \cdot z_i^S\}_{i=1}^m$ and $\{\gamma_l \cdot z_i^T\}_{i=1}^m$, respectively.

Maximum Mean Discrepancy compares two distributions in a reproducing kernel Hilbert space as

$$\text{MMD}^2(Z^S, Z^T) = \left\| \frac{1}{m} \sum_{i=1}^{m} f(z_i^S) - \frac{1}{n} \sum_{i=1}^{n} f(z_i^T) \right\|_{\mathcal{H}}^2 \quad , \tag{2}$$

where $\|\cdot\|_{\mathcal{H}}$ is the Hilbert-Schmidt norm and $f(\cdot)$ is the feature space map associated with the kernel map $k(Z^S, Z^T) = \left\langle f(Z^S), f(Z^T) \right\rangle$, a convex combination of linear kernels (Yan et al., 2017).

2.2 Learning source representations

We seek to understand what properties of source embeddings lead to efficient embedding distribution alignment. In general, the following loss function is minimized during EO training:

$$L(Z^S, Y^S, \theta) = L_{\text{CE}}\left(\psi_\theta(Z^S), Y^S\right) + \lambda L_{\text{rep}}\left(Z^S, Y^S, \theta\right) \quad , \tag{3}$$

where L_{CE} is the cross entropy loss and L_{rep} denotes one of the following representation losses.

The SupCon loss (Khosla et al., 2020) pulls data points from the same class (positives) together in the embedding space and pushes data points from different classes (negatives) away from each other. It uses random data augmentations of same-class images as inputs and is defined as

$$L_c(Z^S, Y^S, \theta) = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(w^T z_i \cdot w^T z_p / \tau)}{\sum_{a \in A(i)} \exp(w^T z_i \cdot w^T z_a / \tau)} \quad , \tag{4}$$

where $I = \{1, ..., 2|Z^S|\}$ is the set of indices for the $2|X^S|$ augmentations, $P(i) \subseteq I$ is the the set of indices of all positives distinct from i, w is a learnable linear projection layer applied to the embeddings, and τ is the temperature parameter controlling the emphasis on hard negative pairs.

The OLÉ loss (Lezama et al., 2018) pushes the inter-class subspace to be orthogonal through encouraging low-rank class-specific representations and high-rank overall representations:

$$L_o(Z^S, Y^S, \theta) = \sum\nolimits_{c=1}^{C} \max\left(\delta, \|Z_c^S\|_*\right) - \|Z^S\|_* \ , \tag{5}$$

where $\|\cdot\|_*$ is the matrix nuclear norm, C is the number of classes, and δ is a constant we set to 1.

The rank reduction loss (Park et al., 2022) encourages low-rank representations with highly correlated features through maximizing the off-diagonal terms of the normalized auto-correlation matrix for the representations zero-centered over the batch dimension:

$$L_r(Z^S, \theta) = -\sum_{i \neq j} \left(\frac{\sum_{k=1}^{m} \bar{Z}_{k,i} \bar{Z}_{k,j}}{\sqrt{\sum_{k=1}^{m} \bar{Z}_{k,i}^2} \sqrt{\sum_{k=1}^{m} \bar{Z}_{k,j}^2}} \right)^2 \quad \text{where} \quad \bar{Z}_{a,b} = Z_{a,b}^S - \frac{1}{m} \sum_{k=1}^{m} Z_{k,b}^S \ . \tag{6}$$

3 Experiments & Results

We use the four low-rise classes in the So2Sat dataset (Zhu et al., 2020) as a classification objective. For training, we use all labeled EO data and sample 128, 512, or $2,048$ labeled SAR data points per class. The remaining SAR data is used as unlabeled data. We report accuracy on a holdout set of labeled SAR data. Results are averaged over 5 trials. To assess properties of the embeddings, we consider 1) the *effective dimension* of the EO embeddings, defined as the fraction of principal components required to explain 90% of the variance (Wold et al., 1987) over the training data and 2) a measure of how close datapoints from each class are to each other called *inertia* (Chavent, 1998).

3.1 Representation learning for distribution alignment

Our first goal is to understand the benefits of SupCon pretraining originally reported in Hussing et al. (2022). For this, we look at a recent result (Wang & Isola, 2020) that provides an asymptotic decomposition of the contrastive loss into two terms: minimizing the mean distance between positive pairs (*alignment*) while promoting *uniformity* of points on the unit sphere. We train the EO model with different alignment and uniformity loss strengths and perform SWD transfer. Fig. 1 shows that decreasing uniformity and increasing alignment increases downstream performance. Further, increased accuracy is correlated with lower effective dimension of the embedding.

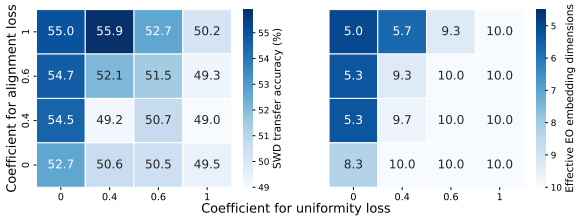

Figure 1: SWD transfer SAR accuracy (Left) and effective EO embedding dimension (Right) for different alignment & uniformity loss coefficients. Lower effective dimension correlates with transfer accuracy.

This observation inspired the introduction of two losses that consider the rank of the embedding matrix directly. The OLÉ loss maximizes the overall but minimizes the per class rank of the embedding, while the rank reduction loss reduces both. The results in Fig. 2 demonstrate that, as hypothesized, down-stream accuracy is correlated with low effective dimension and inertia. Rank reduction consistently leads to the best performance on SWD transfer.

3.2 Condensing clusters with even simpler source representation

Given the benefits of low-rank embeddings, we investigate if even simpler representations might be sufficient and whether neural network pretraining is required. We first model the EO clusters for each class with a simple Gaussian mixture that was fit on the EO embedding (Rostami, 2021). Next, we go further by replacing the source embedding distribution for each class with Gaussian distributions centered at the corresponding one-hot vector. The results are presented in Fig. 3.

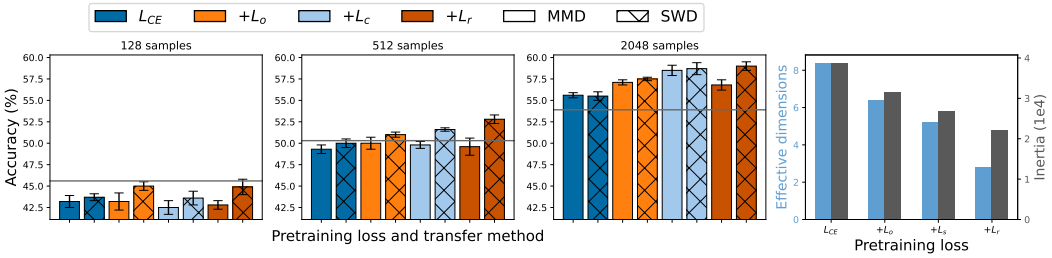

Figure 2: (Left) SAR accuracy when using the EO pretraining losses for MMD and SWD alignment. Colors indicate pretraining losses, patterns correspond to the the alignment methods. The gray line is the accuracy of direct supervised SAR training. The performance of both methods increases when using pretrained representations, especially when more data is available. (Right) Effective dimensions and inertia of the EO embedding. Lower values yield higher down-stream transfer accuracies.

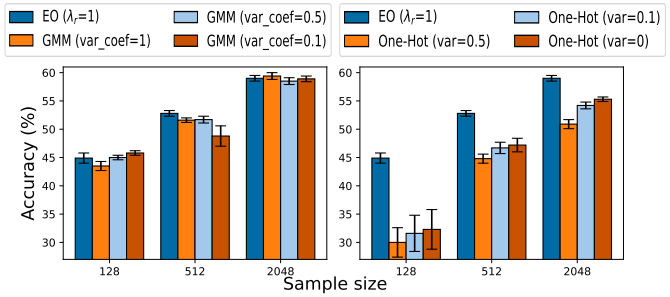

Figure 3: SWD transfer from GMMs and one-hot vectors with rank-reduced EO embeddings. GMMs are based on the EO embedding distribution. With sufficient samples, transfer from GMMs achieves reasonable performance but one-hot vectors provide poor targets.

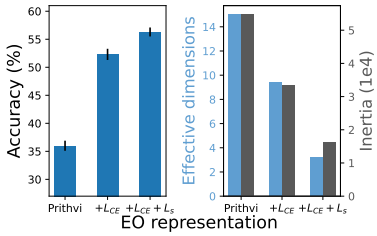

Figure 4: SAR Accuracy, effective dimensions, and inertia for Prithvi training. Original CNN transfer performance can be recovered via representation learning.

For the Gaussian mixture model as the source embedding, we obtain performance comparable to the standard distribution alignment accuracy on the original EO embeddings. However, scaling down the variances to decrease inertia does not lead to an improvement in performance. Further, we observe that SWD transfer performance from one-hot vectors is significantly worse than the standard SWD accuracy, especially with fewer labeled SAR images. This is partially caused by the lack of EO pretraining to fit the classifier. In addition, these results suggest that it is critical for the source embedding to contain fine-grained information about class distribution and inter-class relations.

3.3 Distribution alignment to a foundation model

Lastly, we investigate the transferability of embeddings provided by a recent foundation model for earth observation data called Prithvi (Jakubik et al., 2023). The results are provided in Fig. 4. Prithvi is a powerful feature extractor which might be useful for multi-modal transfer. Linear probing of Prithvi achieves 60.6% accuracy with only 256 labeled SAR samples on average for each class showing its promising ability to generalize to SAR data. Yet, we find that Privthi's embeddings of our data have high effective dimension and inertia; consequently, they are poor alignment targets.

We conjecture that condensing the embedding dimension via contrastive learning might enable us to use alignment methods. We train two additional layers on top of Prithvi to classify our EO data and align to the corresponding embedding instead. However, the results show that we do not gain the desired performance increase; the resulting accuracy is comparable to that of the CNN encoder from previous sections. We believe alignment is not able to transfer the fine-grained details required to disentangle complex EO and SAR scenes that Prithvi can capture (see Appendix C for details).

4 Conclusion and Future Work

We highlight the importance of representation learning for EO to SAR embedding alignment and trace its efficacy to the rank of the EO encoding. We hope these insights will be useful for developing methods in the future that do not require low-dimensional shaping and can align complex distributions. Investigation of such methods will give us the ability to do few-shot alignment to powerful pretrained foundational models and improve our ability to leverage multi-modal satellite data.

Acknowledgments

This research was partially supported by the DARPA SAIL-ON program under contract HR001120C0040, the DARPA ShELL program under agreement HR00112190133, the Army Research Office under MURI grant W911NF20-1-0080, and the DARPA Triage Challenge program under agreement HR001123S0011. The opinions expressed in this article are the authors' own and do not necessarily reflect the views of DARPA, the US Army, or the US government. Approved for public release; distribution is unlimited.

References

Marie Chavent. A monothetic clustering method. *Pattern Recognition Letters*, 19(11):989–996, 1998.

Colin LV Cooke and K Andrea Scott. Estimating sea ice concentration from sar: Training convolutional neural networks with passive microwave data. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7):4735–4747, 2019.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59):1–35, 2016.

Arthur Gretton, Karsten Borgwardt, Malte Rasch, Bernhard Schölkopf, and Alex Smola. A kernel method for the two-sample-problem. *Advances in neural information processing systems*, 19, 2006.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16000–16009, 2022.

Marcel Hussing, Karen Li, and Eric Eaton. Land use prediction using electro-optical to sar few-shot transfer learning. *arXiv preprint arXiv:2212.03084*, 2022.

Johannes Jakubik, Sujit Roy, CE Phillips, Paolo Fraccaro, Denys Godwin, Bianca Zadrozny, Daniela Szwarcman, Carlos Gomes, Gabby Nyirjesy, Blair Edwards, et al. Foundation models for generalist geospatial artificial intelligence. *arXiv preprint arXiv:2310.18660*, 2023.

Somi Jeong, Youngjung Kim, Sungho Kim, and Kwanghoon Sohn. Enriching sar ship detection via multistage domain alignment. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2021.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020.

Junhee Kim, Sujin Shin, Sungho Kim, and Youngjung Kim. Eo-augmented building segmentation for airborne sar imagery. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2021.

Voon Koo, Yee Kit Chan, Gobi Vetharatnam, Ming Yam Chua, Chot Hun Lim, Chee Siong Lim, CC Thum, Tien Sze Lim, Zahid bin Ahmad, Khairul Annuar Mahmood, et al. A new unmanned aerial vehicle synthetic aperture radar for environmental monitoring. *Progress In Electromagnetics Research*, 122:245–268, 2012.

Herbert J Kramer et al. *Observation of the Earth and its Environment: Survey of Missions and Sensors*, volume 10. Springer, 2002.

José Lezama, Qiang Qiu, Pablo Musé, and Guillermo Sapiro. Ole: Orthogonal low-rank embedding-a plug and play geometric loss for deep learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8109–8118, 2018.

Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pp. 97–105. PMLR, 2015.

Geon Yeong Park, Chanyong Jung, Sangmin Lee, Jong Chul Ye, and Sang Wan Lee. Self-supervised debiasing using low rank regularization. *arXiv preprint arXiv:2210.05248*, 2022.

Mohammad Rostami. Lifelong domain adaptation via consolidated internal distribution. *Advances in neural information processing systems*, 34:11172–11183, 2021.

Mohammad Rostami, Soheil Kolouri, Eric Eaton, and Kyungnam Kim. Deep transfer learning for few-shot sar image classification. *Remote Sensing*, 11(11):1374, 2019.

Xian Sun, Bing Wang, Zhirui Wang, Hao Li, Hengchao Li, and Kun Fu. Research progress on few-shot learning for remote sensing image interpretation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:2387–2402, 2021.

Yuan Tai, Yihua Tan, Shengzhou Xiong, Zhaojin Sun, and Jinwen Tian. Few-shot transfer learning for sar image classification without extra sar samples. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:2240–2253, 2022.

Stefan Voigt, Thomas Kemper, Torsten Riedlinger, Ralph Kiefl, Klaas Scholte, and Harald Mehl. Satellite image analysis for disaster and crisis-management support. *IEEE transactions on geoscience and remote sensing*, 45(6):1520–1528, 2007.

Tongzhou Wang and Phillip Isola. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *International Conference on Machine Learning*, pp. 9929–9939. PMLR, 2020.

Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

Hongliang Yan, Yukang Ding, Peihua Li, Qilong Wang, Yong Xu, and Wangmeng Zuo. Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2272–2281, 2017.

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

Chongqi Zhang, Ziwen Zhang, Yao Deng, Yueyi Zhang, Mingzhe Chong, Yunhua Tan, and Pukun Liu. Blind super-resolution for sar images with speckle noise based on deep learning probabilistic degradation model and sar priors. *Remote Sensing*, 15(2):330, 2023.

Di Zhang, Jia Liu, Wang Heng, Kaijun Ren, and Junqiang Song. Transfer learning with convolutional neural networks for sar ship recognition. In *IOP Conference Series: Materials Science and Engineering*, volume 322, pp. 072001. IOP Publishing, 2018.

Xiao Xiang Zhu, Jingliang Hu, Chunping Qiu, Yilei Shi, Jian Kang, Lichao Mou, Hossein Bagheri, Matthias Haberle, Yuansheng Hua, Rong Huang, et al. So2sat lcz42: A benchmark data set for the classification of global local climate zones [software and data sets]. *IEEE Geoscience and Remote Sensing Magazine*, 8(3):76–89, 2020.

A NETWORK ARCHITECTURE DETAILS

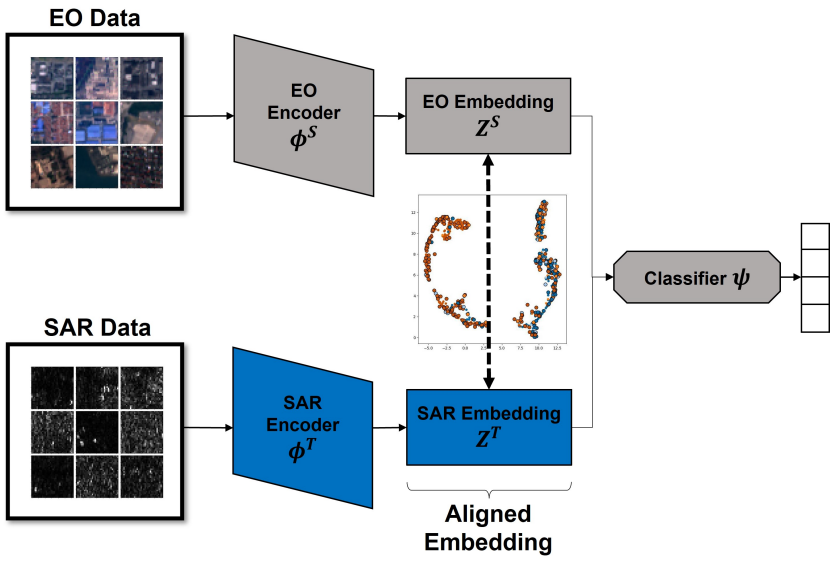

Figure 5: Abstract visualization of the neural network architecture we use. The gray color indicates the parts that are frozen after pretraining and the blue color indicates the parts that are used for alignment.

B HYPERPARAMETERS

Table 1: Direct SAR Training

Hyperparameter	**Configuration**
Optimizer	Adam
Learning Rate	1e-3
Weight Decay	0
Duration	50 epochs
Batch Size	32

Table 2: SWD and MMD Transfer

Hyperparameter	**Configuration**
Pretrain Optimizer	Adadelta
Pretrain Learning Rate	1e-3
Pretrain Weight Decay	0
Pretrain Duration	100 epochs
Pretrain Batch Size ($L_{CE}/L_c/L_o/L_r$)	32/32/32/32
Pretrain Loss Coefficients ($\lambda_{CE}/\lambda_c/\lambda_o/\lambda_r$)	1/1/1/1e-3
Optimizer	Adam
Learning Rate	1e-3
Weight Decay	0
Duration	200 updates
Alignment Batch Size	1024

Table 3: SWD Transfer from Prithvi

Hyperparameter	Configuration
Pretrain Optimizer	Adadelta
Pretrain Learning Rate	1e-2
Pretrain Weight Decay	0
Pretrain Duration	100 epochs
Pretrain Batch Size (L_{CE}/L_c)	32/1024
Pretrain Loss Coefficients (λ_{CE}/λ_c)	1/1
Optimizer	Adam
Learning Rate	1e-3
Weight Decay	0
Duration	300 updates
Alignment Batch Size	1024

Table 4: SWD Transfer from MAE

Hyperparameter	Configuration
MAE Encoder Depth	6
MAE Decoder Depth	2
MAE Training Optimizer	Adam
MAE Training Learning Rate	1e-4
MAE Training Weight Decay	0
MAE Training Duration	200 epochs
MAE Training Batch Size	64
Optimizer	Adam
Learning Rate	1e-3
Weight Decay	0
Duration	1000 updates
Alignment Batch Size	1024

C ADDITIONAL RESULTS ON UNSUPERVISED REPRESENTATIONS

Table 5: Accuracy, effective dimensions, and inertia of SWD transfer from Masked Autoencoders trained on EO data with different embedding dimensions.

Model	MAE (32)	MAE (64)	MAE (128)	MAE (256)
SWD accuracy	$52.2_{\pm 1.1}$	$52.9_{\pm 1.1}$	$51.8_{\pm 0.7}$	$51.8_{\pm 0.5}$
Effective dimension	$11.6_{\pm 1.0}$	$18.0_{\pm 2.0}$	$26.8_{\pm 1.4}$	$32.8_{\pm 2.1}$
Inertia	$28108.4_{\pm 2546.0}$	$51677.6_{\pm 3886.2}$	$122952.8_{\pm 6874.4}$	$278275.6_{\pm 36346.7}$

In section 3.3, we found that the Prithvi model is not a good target to align to without additional supervised training. We want to examine whether this is a function of the unsupervised pretraining nature of the model, the large size of the embedding or the quantity and variety of data it was trained on. To do so, we train our own masked auto-encoder (MAE) transformer (He et al., 2022) on the EO data from the So2Sat dataset. We use this model and align to its embedding as a proxy for the Prithvi model. The results are reported in Table 5.

We find that the poor performance is not necessarily a function of the unsupervised training procedure as alignment to our custom MAEs leads to significantly higher performance than alignment to Prithvi. Additionally, we see that the embedding dimension is relatively unimportant across our MAE results and all accuracies are constant even though we obtain increasing numbers of effective dimensions and inertia with larger embeddings. These insights indicate that the EO data distribution a model is trained on is key in whether alignment is possible or not.