ROBUSTNESS TO GEOGRAPHIC DISTRIBUTION SHIFT USING LOCATION ENCODERS

Ruth Crasto
Microsoft
ruthcrasto@microsoft.com

ABSTRACT

Geographic distribution shift arises when the distribution of locations on Earth in a training dataset is different from what is seen at test time. The most common approaches to tackling geographic distribution shift treat regions delimited by administrative boundaries such as countries or continents as separate domains and apply standard domain adaptation methods, ignoring geographic coordinates that are often available as metadata. This paper proposes the use of location encoders for training models that are more robust to geographic distribution shift. We show how both simple sine-cosine encoders and pre-trained location encoders can be used to improve standard domain adaptation methods for the special case of geographic distribution shift. Our proposed methods achieve state-of-the-art results on geo-tagged imagery datasets from the WILDS benchmark.

1 INTRODUCTION

Empirical risk minimization (ERM) assumes a set of independent and identically distributed (*i.i.d*) training examples. The learned predictor is then expected to generalize well to unseen data at inference time with the assumption that this data is drawn independently from the same distribution. In many real-world applications, these assumptions do not hold (Dundar et al., 2007).

In this paper, we focus on a particular way of violating these assumptions known as *subpopulation shift* (Koh et al., 2021). In this setting, **1)** the training data is not *i.i.d* but rather drawn from many different distributions (domains), and **2)** the inference data is drawn from a different mixture of the same domains. Subpopulation shift arises naturally in many remote sensing applications (Ekim et al., 2024), where training and inference data may be drawn from different mixtures of geographic regions. We refer to this setting as *geographic distribution shift*.

A common approach to learning a predictor that is robust to geographic distribution shift is to treat different regions delimited by administrative boundaries such as countries or continents as different domains and apply standard domain adaptation methods (Koh et al., 2021). However, this approach may ignore intra-regional diversity and inter-regional similarities. Moreover, geographic coordinates that are often available as metadata are discarded.

The central idea of this paper is to use location encoders to model an underlying, continuous domain-assigning function in place of discrete region labels. Location encoders are parametric functions that map geographic coordinates into a learning-friendly higher-dimensional latent space (Mai et al., 2023b). With the rise of geo-tagged image datasets (Tang et al., 2015), location encoders have become a widely applicable tool and can be used to encode a rich domain-specific signal. In particular, geographic coordinates have proven to be a useful input for a variety of prediction tasks, including geo-localization (Vivanco et al., 2023), image classification (Aodha et al., 2019), regression (Klemmer et al., 2023), and super-resolution (Panangian & Bittner, 2025).

In this paper, we show how location encoders can be naturally integrated into various domain adaptation methods for improved robustness to geographic distribution shift. We conduct experiments using two remote sensing datasets from the WILDS benchmark: FMoW (Christie et al., 2018) and PovertyMap (Yeh et al., 2020). Empirically, we demonstrate that our proposed methods lead to significant improvements on worst-group performance metrics under geographic distribution shift. Notably, we achieve a new state-of-the-art result on the WILDS leaderboard for FMoW.

2 BACKGROUND

2.1 SUBPOPULATION SHIFT

What can go wrong when learning a predictor $f : X \rightarrow Y$ using ERM in the midst of a subpopulation shift? Intuitively, different domains may have different sets of features that are predictive of the target y, and the learned predictor must rely on the correct set of features for each domain without becoming dependent on spurious correlations. Standard ERM can fail in this setting by either underfitting to predictive features from under-represented domains in the training data or by overfitting to spurious features from over-represented domains. In both cases, the result is a predictor that performs significantly better on some domains than others at inference time. This bias is problematic when the mixture of domains at inference time is significantly different from what is seen during training or when the bias can have unfair or harmful consequences when deployed (Sagawa et al., 2020).

Many approaches to learning a predictor that is robust to subpopulation shift fall into one of three categories. **1)** The first seeks to strengthen the reliance of the predictor on domain-specific predictive features by using domain labels to condition the hypothesis space. Examples include feature-wise modulation (Lee et al., 2021) and mixture of experts models (Yao et al., 2024). **2)** The second seeks to reduce the reliance of the predictor on spurious correlations by learning features that satisfy some invariance condition across domains (Sun & Saenko, 2016) (Arjovsky et al., 2020) (Ganin et al., 2016). Note that some recent approaches are more targeted and can be considered part of both categories (Yao et al., 2022) (Gao et al., 2023). **3)** The third category consists of distributionally robust optimization approaches that aim to minimize risk according to an estimated worst-case data distribution (Sagawa et al., 2020). In this work, we focus on a subset of approaches from the first two categories, which we describe below.

2.2 CONDITIONAL MODELING METHODS

Domain conditional predictors (Lee et al., 2021)   The key idea behind domain conditional predictors (DCP) is to learn a latent representation of each domain and to condition the prediction of y on these domain representations using FiLM (Perez et al., 2018). The latent domain representations are obtained from an intermediate layer in a neural network that is trained to predict the domain label directly from the input x. This network is trained jointly with the predictor f. Please refer to Appendix A for the exact structure of the DCP loss function.

D^3G (Yao et al., 2024)   Given domain labels at training time, D^3G trains separate prediction heads for each domain in the training data, and jointly trains another model to predict pairwise similarities between domains (called domain relations) from metadata. At inference time, the predictions output from various heads are weighted by their relation to the input domain. The domain relation predictor is of the form:

$$\frac{1}{R} \sum_{r=1}^{R} S_C \left(w_r \cdot g(m_i), \ w_r \cdot g(m_{\text{head} j}) \right)$$

where S_C denotes cosine similarity, m_i, $m_{\text{head} j}$ is metadata representing domain i and the domain of the head j respectively, w_r is a set of learnable vectors, and g is a neural network. When applied to geographic distribution shift, D^3G uses continent-level labels as the domain metadata input to g.

2.3 INVARIANCE-BASED METHODS

Invariance-based domain adaptation methods seek to learn features that satisfy some invariance condition across domains during training. In this work, we focus on two methods that enforce invariance via a regularization term in the training loss:

IRM (Arjovsky et al., 2020)   The IRM regularizer encourages features such that a simple classifier common across domains on top of those features may be optimal for all domains in the training data.

CORAL (Sun & Saenko, 2016)   The CORAL regularizer encourages domain-pairwise differences between feature means and covariances to be small.
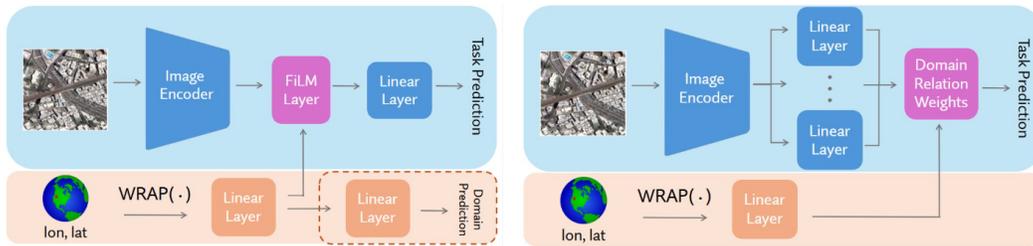
Figure 1: Illustration of our proposed approaches, DCP + WRAP (left) and D³G + WRAP (right). The domain predictor in dashed red outline is discarded at inference time.

3 METHODS

In this section, we explain how we propose to use location encoders to improve the standard domain adaptation methods described in the previous section for the case of geographic distribution shift.

3.1 LOCATION-AWARE CONDITIONAL MODELING

DCP + WRAP We hypothesize that in the case of geographic distribution shift, geographic coordinates are better predictors of domain than inputs x. We therefore propose to substitute the domain predictor in the original DCP method with 2 linear layers trained on top of a simple sine-cosine encoder which we refer to as WRAP, following Aodha et al. (2019) and Mai et al. (2023b):

$$\text{WRAP}(\phi, \theta) = [\ \sin(\phi)\ \cos(\phi)\ \sin(\theta)\ \cos(\theta)\ ]$$

where (ϕ, θ) is the longitude and latitude. We illustrate this approach in Figure 1. It can be interpreted as a way to condition the predictor f on a location embedding that is encouraged to be predictive of the corresponding domain label. Note that any standard location encoder can be used in place of WRAP, but to demonstrate the effectiveness of our approach, we chose a simple encoder for the scope of this work.

D³G + WRAP We hypothesize that in the case of geographic distribution shift, geographic coordinates are better predictors of domain relation than discrete region labels. We therefore propose to modify the original D³G method to use geographic coordinates for the metadata m_i input to the domain relation predictor, and to use a linear layer trained on top of WRAP features to encode the input domain in place of g (while $m_{\text{head }j}$ and its encoding are unchanged). This can be interpreted as a way of predicting domain relations from location embeddings that have been encouraged to align more closely to the corresponding domain label.

3.2 LOCATION-AWARE INVARIANCE-BASED METHODS

IRM/CORAL + SatCLIP We hypothesize that pre-trained location embeddings that capture the visual characteristics of given locations across the globe can model domain assignments that are more suitable for imposing invariance conditions than administrative boundaries. We propose to use SatCLIP location embeddings (Klemmer et al., 2023) for clustering locations, and to treat each cluster as a separate domain. Specifically, we perform k-means clustering on normalized SatCLIP embeddings for all locations in the training data, where k is a hyperparameter. For validation and inference, we use nearest-neighbor cluster assignment, i.e. we do not include validation and test sets when clustering. Note that other pre-trained location encoders such as GeoCLIP (Vivanco et al., 2023) or CSP (Mai et al., 2023a) may be used in place of SatCLIP. A comparison of how different location encoders perform in conjunction with our proposed methods is left for future work.

4 EXPERIMENTS

Datasets We conduct experiments on two datasets from the WILDS benchmark: **1)** FMoW (Christie et al., 2018) is a land-use classification dataset from satellite imagery covering over 200

Table 1: Our experiment results on the official WILDS OOD test splits. Following the WILDS guidelines, these results are averaged across 3 random seeds for FMoW and across the 5 official data folds for PovertyMap. Standard error is reported in parentheses.

	FMoW		PovertyMap	
	Avg. Acc. ($\uparrow$)	Worst Acc. ($\uparrow$)	Avg. r ($\uparrow$)	Worst r ($\uparrow$)
ERM	67.1 (0.07)	47.6 (0.86)	0.78 (0.02)	0.45 (0.02)
D3G	65.8 (0.24)	50.4 (0.92)	0.76 (0.03)	0.37 (0.04)
IRM	64.7 (0.12)	48.4 (1.34)	0.77 (0.02)	0.41 (0.04)
CORAL	65.3 (0.25)	49.4 (0.73)	0.76 (0.03)	0.39 (0.05)
ERM + WRAP	66.7 (0.34)	47.5 (0.55)	0.77 (0.02)	0.44 (0.03)
DCP + WRAP	66.9 (0.21)	51.8 (0.86)	0.78 (0.02)	**0.51 (0.04)**
D3G + WRAP	66.1 (0.26)	**53.8 (0.22)**	0.76 (0.02)	0.43 (0.03)
IRM + SatCLIP	65.2 (0.58)	50.6 (0.76)	0.76 (0.03)	0.41 (0.04)
CORAL + SatCLIP	66.2 (0.33)	50.1 (0.98)	0.78 (0.02)	0.44 (0.05)

countries. Each continent is treated as a different domain. **2)** PovertyMap (Yeh et al., 2020) is a dataset of multi-spectral Sentinel-2 imagery for asset wealth prediction (a regression task). The data is grouped into two domains corresponding to whether the image was taken in an urban or rural area.

Architecture   For both datasets, the base model we train is an image encoder with a linear prediction head. For FMoW we finetune CLIP ViT-L/14 (Radford et al., 2021) as the image encoder and for PovertyMap we use a randomly-initialized multi-spectral ResNet-18 as in Koh et al. (2021). A detailed report of hyperparameters can be found in Appendix A.

Baselines   In addition to the standard domain adaptation baselines (D^3G, IRM, and CORAL), we experiment with two baseline methods: standard ERM, and concatenating WRAP-encoded location with the output from the image encoder before the linear prediction head (ERM + WRAP).

Results   In Table 1 we report performance on the WILDS OOD test set averaged across all samples and across all samples from the worst-performing domain. Validation set results can be found in Appendix A. With the exception of IRM on PovertyMap, our proposed location-aware variants for all domain adaptation methods outperform their vanilla counterparts on worst-group performance. In particular, D^3G + WRAP tops the official WILDS leaderboard for FMoW. Moreover, DCP + WRAP achieves state-of-the-art performance on both datasets and is the only domain adaptation method to outperform standard ERM on PovertyMap. Interestingly, location-aware conditional modeling methods tend to significantly outperform the ERM + WRAP baseline, suggesting that location encoders yield the greatest improvement in worst-group performance when used in conjunction with standard domain adaptation methods.

5   CONCLUSION

It is written in Dundar et al. (2007) that though sample correlations are ubiquitous in real-world datasets, "we do not appreciate the benefits of modeling these correlations, and the ease with which this can be accomplished algorithmically." In this paper, we show how location encoders can be leveraged as a simple and effective tool for modeling correlations between the visual characteristics of different regions on Earth. We demonstrate that using them in conjunction with standard domain adaptation methods can greatly improve robustness to geographic distribution shift on benchmark datasets. Future work may involve investigating our proposed methods for other tasks beyond regression and classification, extending them to other settings such as unsupervised domain adaptation (Sagawa et al., 2022), or comparing different location features and pre-training strategies.

REFERENCES

Oisin Mac Aodha, Elijah Cole, and Pietro Perona. Presence-only geographical priors for fine-grained image classification. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9595–9605, 2019.

Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. Invariant risk minimization. 2020. URL https://arxiv.org/abs/1907.02893.

Gordon Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. Functional map of the world. In *CVPR*, 2018.

Murat Dundar, Balaji Krishnapuram, Jinbo Bi, and R. Bharat Rao. Learning classifiers when the training data is not iid. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, IJCAI'07, pp. 756–761, 2007.

Burak Ekim, Girmaw Abebe Tadesse, Caleb Robinson, Gilles Hacheme, Michael Schmitt, Rahul Dodhia, and Juan M. Lavista Ferres. Distribution shifts at scale: Out-of-distribution detection in earth observation. *arXiv preprint arXiv:2412.13394*, 2024. URL https://arxiv.org/abs/2412.13394.

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. URL http://jmlr.org/papers/v17/15-239.html.

Irena Gao, Shiori Sagawa, Pang Wei Koh, Tatsunori Hashimoto, and Percy Liang. Out-of-domain robustness via targeted augmentations. In *Proceeding of the Fortieth International Conference on Machine Learning*, 2023.

Konstantin Klemmer, Esther Rolf, Caleb Robinson, Lester Mackey, and Marc Rußwurm. Satclip: Global, general-purpose location embeddings with satellite imagery. *arXiv preprint arXiv:2311.17179*, 2023. URL https://arxiv.org/abs/2311.17179.

Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, Tony Lee, et al. Wilds: A benchmark of in-the-wild distribution shifts. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5637–5664. PMLR, 18–24 Jul 2021.

Dar-Shyang Lee, Jianqiao Feng, Joao Monteiro, Vincent Dumoulin, and Xavier Gibert. Domain conditional predictors for domain adaptation. *Proceedings of Machine Learning Research (PMLR)*, 148:193–220, 2021. URL http://proceedings.mlr.press/v148/monteiro21a.html.

Gengchen Mai, Ni Lao, Yutong He, Jiaming Song, and Stefano Ermon. Csp: Self-supervised contrastive spatial pre-training for geospatial-visual representations. In *International Conference on Machine Learning*. PMLR, 2023a.

Gengchen Mai, Yao Xuan, Wenyun Zuo, Yutong He, Jiaming Song, Stefano Ermon, Krzysztof Janowicz, and Ni Lao. Sphere2vec: A general-purpose location representation learning over a spherical surface for large-scale geospatial predictions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 202:439–462, 2023b. URL https://www.sciencedirect.com/science/article/pii/S0924271623001818.

Daniel Panangian and Ksenia Bittner. Can location embeddings enhance super-resolution of satellite imagery? *arXiv preprint arXiv:2501.15847*, 2025. URL https://arxiv.org/abs/2501.15847.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron C. Courville. Film: Visual reasoning with a general conditioning layer. In *AAAI*, 2018.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763. PMLR, 18–24 Jul 2021.

Shiori Sagawa, Pang Wei Koh, Tatsunori B. Hashimoto, and Percy Liang. Distributionally robust neural networks. In *8th International Conference on Learning Representations, ICLR*, 2020. URL https://arxiv.org/abs/1911.08731.

Shiori Sagawa, Pang Wei Koh, Tony Lee, Irena Gao, Sang Michael Xie, Kendrick Shen, Ananya Kumar, Weihua Hu, Michihiro Yasunaga, Henrik Marklund, Sara Beery, Etienne David, Ian Stavness, Wei Guo, Jure Leskovec, Kate Saenko, Tatsunori Hashimoto, Sergey Levine, Chelsea Finn, and Percy Liang. Extending the wilds benchmark for unsupervised adaptation. In *International Conference on Learning Representations (ICLR)*, 2022.

Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *Computer Vision – ECCV 2016 Workshops*, pp. 443–450. Springer International Publishing, 2016.

Kevin Tang, Manohar Paluri, Li Fei-Fei, Rob Fergus, and Lubomir Bourdev. Improving image classification with location context. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1008–1016, 2015.

Vicente Vivanco, Gaurav Kumar Nayak, and Mubarak Shah. Geoclip: Clip-inspired alignment between locations and images for effective worldwide geo-localization. In *Advances in Neural Information Processing Systems*, 2023.

Huaxiu Yao, Yu Wang, Sai Li, Linjun Zhang, Weixin Liang, James Zou, and Chelsea Finn. Improving out-of-distribution robustness via selective augmentation. In *Proceeding of the Thirty-ninth International Conference on Machine Learning*, 2022.

Huaxiu Yao, Xinyu Yang, Xinyi Pan, Shengchao Liu, Pang Wei Koh, and Chelsea Finn. Improving domain generalization with domain relations. In *The Twelfth International Conference on Learning Representations, ICLR*, 2024.

Christopher Yeh, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. Using publicly available satellite imagery and deep learning to understand economic well-being in africa. *Nature Communications*, 11, 2020.

A APPENDIX

A.1 HYPERPARAMETERS

In this section, we describe the training hyperparameters that are specific to each dataset (FMoW and PovertyMap) and specific to each domain adaptation method (DCP, D³G, IRM, and CORAL).

FMoW Each model is trained for 5 epochs on a single GPU with early stopping on validation loss. We train with batch size 16 and use the Adam optimizer with initial learning rate 10^{-4} that decays by a factor of 0.96 each epoch. The CLIP ViT-L/14 backbone is finetuned with the same optimizer and learning rate schedule but with an initial learning rate of 10^{-5}. We also use gradient accumulation to achieve an effective batch size of 64. Finally, we normalize all images using ImageNet mean and standard deviation, and apply random horizontal flip.

PovertyMap As in Koh et al. (2021), each model is trained for 200 epochs on a single GPU with early stopping on validation r. We train with batch size 64 and use the Adam optimizer with initial learning rate 10^{-3} that decays by a factor of 0.96 each epoch. As in Koh et al. (2021), we apply random horizontal and vertical flip and color jitter to each image.

DCP The linear layer used for domain prediction is also trained with Adam and step decay learning rate schedule (decayed by a factor of 0.96 per epoch) but with initial learning rate 0.1 times the initial learning rate of the prediction head (i.e. 0.1×10^{-4} for FMoW and 0.1×10^{-3} for PovertyMap). The DCP loss function includes two terms: the task prediction (TP) loss, and the domain prediction (DP) loss. It is of the form:

$$\mathcal{L}_{TP} + \lambda \mathcal{L}_{DP}$$

where $\mathcal{L}$ is classification loss (cross entropy) and λ is a hyperparameter. We experimented with values of λ in $\{0., 0.1, 0.2\}$ and found that $\lambda = 0.2$ performed best across all experiments. Note that DCP + WRAP with $\lambda = 0$ is equivalent to conditioning prediction on WRAP-encoded location features with FiLM.

D³G There are two training hyperparameters used in D³G, denoted λ and β in the original paper. Similar to DCP, D³G has two loss terms: the task prediction loss (denoted $\mathcal{L}_{pred}$ in the original paper) computed from the prediction head for the current domain, and the consistency loss ($\mathcal{L}_{rel}$) that is computed from all other prediction heads whose predictions are weighted by their relation to the current domain. The final loss is of the form:

$$\mathcal{L}_{pred} + \lambda \mathcal{L}_{rel}$$

The β hyperparameter is used to average fixed and learned domain relations (see Yao et al. (2024) for more details). We use a value of $\lambda = 0.5$ and $\beta = 0.8$ for all experiments.

IRM The IRM loss function includes two terms: the first is the task prediction loss and the second is the IRM penalty term, weighted by λ which is a hyperparameter. In all experiments with IRM, we use $\lambda = 0.1$ and compute IRM penalty on the final output from the image encoder (before the linear prediction head).

CORAL The CORAL loss function includes two terms: the first is the task prediction loss and the second is the CORAL penalty, computed as the sum of **1)** the squared L2 norms of all domain-pairwise differences between feature means in the current batch, and **2)** the squared Frobenius norms of all domain-pairwise differences between feature covariances in the current batch. The second term is weighted by a hyperparameter λ. In all experiments with CORAL, we use $\lambda = 0.1$ and compute the CORAL penalty on the final output from the image encoder (before the linear prediction head).

A.2 FULL EXPERIMENT RESULTS

In this section, we extend Table 1 to include results on the official WILDS OOD validation sets for FMoW (Table 2) and PovertyMap (Table 3).

Table 2: Our experiment results on the official WILDS OOD validation and test splits for FMoW. Following the WILDS guidelines, these results are averaged across 3 random seeds. Standard error is reported in parentheses.

	Validation Set		Test Set	
	Avg. Acc. (↑)	Worst Acc. (↑)	Avg. Acc. (↑)	Worst Acc. (↑)
ERM	73.31 (0.50)	62.22 (0.08)	67.1 (0.07)	47.6 (0.86)
D3G	71.52 (0.17)	59.28 (0.85)	65.8 (0.24)	50.4 (0.92)
IRM	71.07 (0.26)	60.73 (1.30)	64.7 (0.12)	48.4 (1.34)
CORAL	71.55 (0.14)	59.74 (0.54)	65.3 (0.25)	49.4 (0.73)
ERM + WRAP	73.06 (0.55)	63.18 (1.29)	66.7 (0.34)	47.5 (0.55)
DCP + WRAP	72.56 (0.22)	**63.55 (1.55)**	66.9 (0.21)	51.8 (0.86)
D3G + WRAP	71.7 (0.11)	63.51 (2.35)	66.1 (0.26)	**53.8 (0.22)**
IRM + SatCLIP	70.84 (0.29)	60.31 (1.21)	65.2 (0.58)	50.6 (0.76)
CORAL + SatCLIP	71.6 (0.31)	61.77 (0.87)	66.2 (0.33)	50.1 (0.98)

Table 3: Our experiment results on the official WILDS OOD validation and test splits for PovertyMap. Following the WILDS guidelines, these results are averaged across the 5 official data folds. Standard error is reported in parentheses.

	Validation Set		Test Set	
	Avg. r (↑)	Worst r (↑)	Avg. r (↑)	Worst r (↑)
ERM	0.80 (0.02)	0.48 (0.04)	0.78 (0.02)	0.45 (0.02)
D3G	0.80 (0.01)	0.46 (0.04)	0.76 (0.03)	0.37 (0.04)
IRM	0.78 (0.02)	0.43 (0.02)	0.77 (0.02)	0.41 (0.04)
CORAL	0.78 (0.03)	0.45 (0.05)	0.76 (0.03)	0.39 (0.05)
ERM + WRAP	0.79 (0.02)	0.48 (0.04)	0.77 (0.02)	0.44 (0.03)
DCP + WRAP	0.80 (0.02)	0.50 (0.02)	0.78 (0.02)	**0.51 (0.04)**
D3G + WRAP	0.80 (0.02)	0.48 (0.03)	0.76 (0.02)	0.43 (0.03)
IRM + SatCLIP	0.79 (0.02)	0.44 (0.03)	0.76 (0.03)	0.41 (0.04)
CORAL + SatCLIP	0.80 (0.02)	**0.50 (0.03)**	0.78 (0.02)	0.44 (0.05)