OT ON THE MAP: QUANTIFYING DOMAIN SHIFTS IN GEOGRAPHIC SPACE

Anonymous authors
Paper under double-blind review

ABSTRACT

In computer vision and machine learning for geographic data, out-of-domain generalization is a pervasive challenge, arising from uneven global data coverage and distribution shifts across geographic regions. Although task-specific models are frequently trained in one region and deployed in another, there is no principled method for determining when this cross-region adaptation will be successful. *Location embeddings*, globally available representations of geographic characteristics such as terrain and environmental conditions, can provide a basis for comparing geographic domains. Using location embeddings, we propose a novel strategy for predicting geospatial domain transfer performance with Optimal Transport methods (GEOSPOT). In our experiments, GEOSPOT distances emerge as effective predictors of cross-domain transfer difficulty, with distances based solely on location embeddings outperforming those computed from image or text features in two of four experimental settings. We further demonstrate that location embeddings provide information comparable to image/text embeddings, despite relying solely on longitude-latitude pairs as input. This allows users to predict how well a geospatial model trained in a given geographic domain will perform out-of-distribution, even when the exact downstream task is unknown, or no task-specific data are available. Building on these findings, we show that GEOSPOT can also guide the selection of source data for transfer to a target geographic region, without requiring any target-domain input features or labels.

1 INTRODUCTION

Machine learning (ML) with geospatial data has been used for a range of impactful applications, such as crop yield prediction (Ansarifar et al., 2021), disaster forecasting (Linardos et al., 2022), and pollution monitoring (Hu et al., 2017). However, the quantity, quality and accessibility of geospatial data differ across geographic regions due to factors such as population, environmental and socioeconomic differences (Sudmanns et al., 2020; Rolf et al., 2024). Due to the limited availability and quality of labeled datasets, domain scientists commonly train models on data-rich regions (often in Western countries) and then transfer the model's predictive capabilities to data-poor target regions (often in the Global South) via domain adaptation. However, transferring ML models between geographic domains is not always straightforward. Distinct climates, cultures, and environmental conditions cause distribution shifts between regions (Sudmanns et al., 2020; Federici et al., 2021) that can lead to differences in data appearance, labels, and distribution. Consequently, ML models may struggle to generalize effectively to unseen geographic domains (Rolf et al., 2024). Predicting when and where a model will transfer is therefore a prerequisite for the reliable deployment of geospatial ML models, yet accurately and precisely comparing geographic datasets remains an open challenge.

A common heuristic for geospatial model transferability is geographic distance, following Tobler's first law of geography: "all things are related, but near things are more related" (Tobler, 1970). This heuristic is supported by empirical evidence; for example, species distribution modeling experiments have found that transferability decreased with geographic distance (Rousseau & Betts, 2022; Nguyen & Leung, 2022). Additionally, geographic distance between domains requires no knowledge of the underlying data distribution—a strength in that it's easy to compute, but also a critical weakness, as it ignores attributes of the data collected in each region. A second, complementary notion of distance instead compares domains in feature space, for example by using optimal transport (OT) on feature embeddings of the data (Alvarez-Melis & Fusi, 2020). Such feature-based distances

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

directly reflect similarity in the observable data (e.g., images, text, or other measurements), but they require that the data already exist for both domains—a costly and often impractical assumption in real-world geospatial settings. Moreover, these distances are not inherently aware of geographic structure, leaving important information unused.

General-purpose geospatial embeddings, called *location embeddings*, present a way to combine both of these notions of distance, as they encode rich information about terrain and environmental conditions from geospatial data (Klemmer et al., 2025) without requiring the actual data or labels. By viewing these datasets as empirical distributions over location embeddings, we can use these location embeddings to compare entire geographic domains, rather than solely individual pairs of locations. Specifically, by using Optimal Transport, we can lift pairwise distance in location embedding space to a dataset or domain-level distance, which in turn is based on both spatial proximity and encoded ground conditions. In this work, we propose geospatial optimal transport (GEOSPOT), an optimal transport dataset distance between geographic domains based on location embeddings. In four experimental settings across three datasets, we demonstrate that GEOSPOT:

- **reliably predicts transfer performance between geographic domains using pretrained location embeddings** from GeoCLIP and SatCLIP, and using Optimal Transport to quantify domain similarity;
- **outperforms other distributional and non-distributional domain-distance baselines, and is comparable to feature-based dataset distances**, with GEOSPOT outperforming image/text-based distances in two of four experimental settings;
- **can be extended with input modality information via GEOSPOT-FUSE**, which combines location embeddings with task-specific embeddings; and
- **supports principled dataset selection for geographic transfer**, identifying useful training domains for a target region without requiring input features or labels from that domain.

2 RELATED WORK

Geographic distribution shifts. Machine Learning models trained on geospatial data often face substantial performance degradation when deployed in new regions due to *geographic distribution shifts*. Such shifts arise from regional variations in terrain, seasonality, and acquisition conditions (Ekim et al., 2025; Rolf et al., 2024), as well as differences in sensing instruments and collection protocols (Lynch & Wookey, 2021). For instance, while satellite images of houses in the United States and Egypt depict the same conceptual object "house", differences in architectural style, urban planning, and environmental factors produce distinct visual signatures. Previous work provides empirical evidence for these distribution shifts, e.g., highlighting performance gaps when training data lack geographic diversity (Nachmany & Alemohammad, 2019; Shankar et al., 2017), and investigates what geographic regions transfer best to others (Kaur et al., 2026).

Efforts to address such shifts in geospatial ML have focused primarily on *post hoc* adaptation and robustness. This includes work on out-of-distribution detection (Gawlikowski et al., 2021; Dimitrić et al., 2023; Bellier & Audebert, 2024; Ekim et al., 2025) and improved domain adaptation methods for remote sensing imagery (Lin et al., 2019; Fang et al., 2022; Makkar et al., 2022; Ismael et al., 2023). Recent studies have incorporated relevant metadata—particularly geographic coordinates—into adaptation frameworks to enhance cross-region transfer (Yao et al., 2024; Crasto & Rolf, 2026). In contrast, our work focuses on *predicting* domain transfer success *a priori*, even for the challenging case where no data beyond coordinates is available for the target domain. Furthermore, unlike work on quantifying feature-space distances for individual samples (Meyer & Pebesma, 2021), we propose to compare entire geospatial data distributions, introducing a principled way to anticipate geospatial adaptation challenges before model training.

Dataset distances for domain adaptation. Rigorous, well-defined methods for quantitatively comparing datasets can guide domain adaptation decisions, and several distinct notions of dataset distances have been explored to predict cross-domain transfer success. Discrepancy distances measure the similarity over a set of functions from a hypothesis class (Kifer et al., 2004; Ben-David et al., 2006) yet are often computationally intractable or infeasible to compute. Other measures of distance employ information-theoretic measures of similarity (Achille et al., 2019) or learned domain relations (Yao et al., 2024), but depend on the model used to determine distances. More recent work leverages

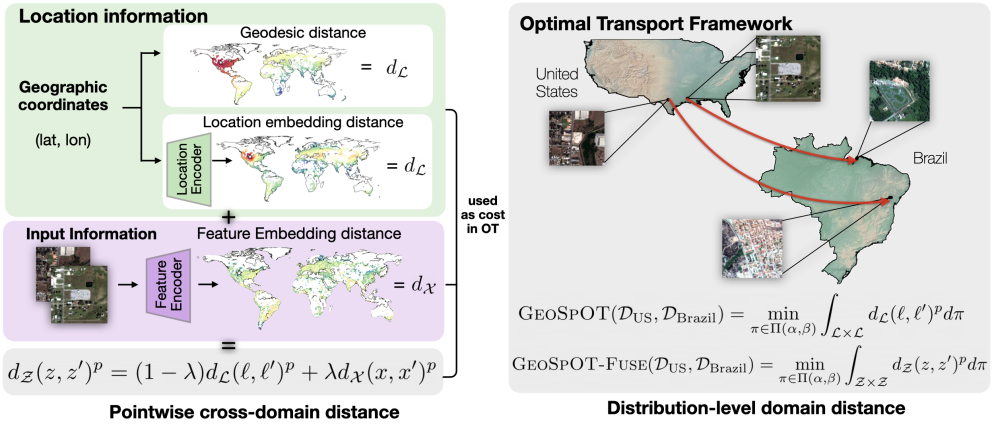


Figure 1: **Overview of GeoSpOT domain distances:** (Left) Pointwise cross-domain computation according to feature (image) embedding distance, raw geographic distance, or location embedding distance, highlighting the difference between the approaches. (Right) Distribution-level domain distance between the United States and Brazil. Red arrows represent transported mass under the coupling resulting from solving the optimal transport problem. Coupled images share similar visual features despite being far apart geographically.

Optimal Transport (OT) as a flexible way to define dataset distances, encoding aspects of the data that are most relevant for learning, with extensions incorporating label information (Courty et al., 2017; Alvarez-Melis & Fusi, 2020; Tan et al., 2021), group structure (Vayer et al., 2020), or hierarchical relationships (Yurochkin et al., 2019). Among these, Tan et al. (2021) combine the OT distance between domains with a label-dependent task distance, but in our label-agnostic setting, this reduces to the standard OT (Sinkhorn) distance between domains. Similarly, Alvarez-Melis & Fusi (2020) incorporate label information by modifying the ground cost to combine feature- and label-based distances. OT approaches have been applied to geospatial ML, e.g. for target shift (Redko et al., 2019) and training on noisy data (Damodaran et al., 2020). To the best of our knowledge, our approach is the first to use OT on location embeddings, and we specifically focus on predicting cross-region transfer in the setting of geographically separate domains, explicitly encoding feature and spatial information in the OT ground cost. This geographically-aware OT formulation tailors dataset distances to geospatial applications, improving predictive power for transfer across regions.

3 Methods

Our goal is to define a notion of similarity or distance between *geospatial domains* (i.e. geographic areas) that captures differences in the underlying terrain and environmental conditions across domains. Given a source domain $\mathcal{D}_s$ and a target domain $\mathcal{D}_t$, we seek a principled distance between them that reflects location-level differences, without necessarily requiring access to features (e.g., images, text) collected at these locations.

Optimal Transport. Let $\mathcal{L}$ denote the location (embedding) space. We consider two finite datasets, $\mathcal{D}_s = \{\ell_s^{(i)}\}_{i=1}^{n_s}$ and $\mathcal{D}_t = \{\ell_t^{(j)}\}_{j=1}^{n_t}$, corresponding to the source and target domains, which we treat as *empirical probability measures*: $\alpha = \frac{1}{n_s}\sum_{i=1}^{n_s} \delta_{\ell_s^{(i)}}, \quad \beta = \frac{1}{n_t}\sum_{j=1}^{n_t} \delta_{\ell_t^{(j)}}$, where δ_ℓ denotes the Dirac delta measure centered at ℓ. Given a pointwise "ground" cost function $c_{\mathcal{L}}(\cdot,\cdot) : \mathcal{L}\times\mathcal{L} \to \mathbb{R}^+$ that quantifies the cost of matching locations across domains (and is often taken to be a distance metric $d_{\mathcal{L}}$), optimal transport lifts this into a notion of distribution-level discrepancy between the entire domains through the following optimization problem:

$$\text{OT}(\alpha, \beta) = \min_{\pi \in \Pi(\alpha,\beta)} \int_{\mathcal{L}\times\mathcal{L}} c_{\mathcal{L}}(\ell, \ell') \text{d}\pi(\ell, \ell'),$$

where $\Pi(\alpha, \beta)$ is the set of joint probability distributions over $\mathcal{L}\times\mathcal{L}$ with marginals α and β. Intuitively, OT finds a coupling π that reallocates "mass" from points in the source domain to points

in the target domain as efficiently as possible under the ground cost $c_{\mathcal{L}}$. The value of the minimum cost coupling between domains can be understood as a geometric-statistical quantification of their dissimilarity.

Geospatial Optimal Transport (GEOSPOT). Geospatial domains can refer to distinct spatial units such as countries, continents, ecological regions, or grid cells on a longitude/latitude mesh. Their distinction may refer either to natural, physical attributes of the planet (e.g., ecological regions, climate zones) or to man-made definitions (e.g., voting districts, countries). We define the GEOSPOT distance between two geospatial domains as the OT distance over their location embeddings:

$$\text{GEOSPOT}(\mathcal{D}_s, \mathcal{D}_t) = \min_{\pi \in \Pi(\alpha,\beta)} \int_{\mathcal{L}\times\mathcal{L}} d_{\mathcal{L}}(\ell, \ell')^p \, \mathrm{d}\pi(\ell, \ell'), \tag{1}$$

where $d_{\mathcal{L}}(\ell, \ell') = 1 - \phi(\ell)^\top \phi(\ell')/\|\phi(\ell)\|\|\phi(\ell')\|$ is the cosine distance between pretrained location embeddings $\phi(\ell)$, $\phi(\ell')$ obtained from a location encoder. Because GEOSPOT relies only on location, it can be computed for any pair of domains without requiring that the target dataset's features already exist — a key advantage in data-scarce settings. For an informative baseline, we can replace $d_{\mathcal{L}}$ with the raw geographic (arc) distance d_{arc} between latitude–longitude coordinates. This uses a purely spatial notion of domain distance and serves as an OT-based analog of Tobler's heuristic. In our experiments, we compare both variants of $d_{\mathcal{L}}$.

GEOSPOT-FUSE: Incorporating Feature Information. While GEOSPOT captures similarity in terrain and environmental conditions through location alone, it does not have direct access to the features (e.g., images, text) that predictive models operate on. When such features are available for both domains, we propose GEOSPOT-FUSE, a variant that also incorporates feature-level similarity. Let $\mathcal{Z} \triangleq \mathcal{X} \times \mathcal{L}$ represent the joint feature-location space, and let $\mathcal{D}_s = \{z_s^{(i)}\}_{i=1}^{n_s}$, $\mathcal{D}_t = \{z_t^{(j)}\}_{j=1}^{n_t}$ with $z = (x, \ell) \in \mathcal{Z}$. We define a pointwise distance on $\mathcal{Z}$ as a convex combination of feature and location distances:

$$d_{\mathcal{Z}}(z, z')^p = (1-\lambda)\, d_{\mathcal{L}}(\ell, \ell')^p + \lambda\, d_{\mathcal{X}}(x, x')^p, \tag{2}$$

where the hyperparameter λ controls the relative contribution of feature vs. location information, $d_{\mathcal{X}}$ is a cosine-based dissimilarity between feature embeddings extracted from a pretrained image or text encoder, and $d_{\mathcal{L}}$ is the location embedding distance as defined above. When both $d_{\mathcal{X}}$ and $d_{\mathcal{L}}$ are valid metrics and $p \geq 1$, this ℓ_p combination defines a metric on the joint space $\mathcal{Z}$. Here, however, we take $d_{\mathcal{X}}$ to be a cosine-based dissimilarity, which does not strictly satisfy the triangle inequality. In such cases, $d_{\mathcal{Z}}$ should be interpreted more generally as a ground cost rather than a true metric—a distinction that does not affect the validity of the OT formulation, which only requires nonnegative costs. With this, we define

$$\text{GEOSPOT-FUSE}(\mathcal{D}_s, \mathcal{D}_t) = \min_{\pi \in \Pi(\alpha,\beta)} \int_{\mathcal{Z}\times\mathcal{Z}} d_{\mathcal{Z}}(z, z')^p \, \mathrm{d}\pi(z, z'). \tag{3}$$

In practice, computing GEOSPOT-FUSE requires choosing λ in Equation (2), which weights feature-level and location-level information. We explore this design choice in our experiments. GEOSPOT can be seen as the special case $\lambda = 0$, obtained without requiring access to feature data at all.

4 EXPERIMENTAL SETUP

We conduct experiments on three datasets spanning different modalities, scales, class distributions, and geographic shifts. Preprocessing details are provided in the Appendix and code release.

GeoYFCC. GeoYFCC (Dubey et al., 2021) is a large-scale domain generalization benchmark with over 1.1 million geotagged samples from 1,261 classes and 62 country domains, subsampled from the YFCC100M collection of Flickr images and text descriptions (Thomee et al., 2016). Samples may have multiple labels; following the original work, we convert the task to single-label classification by duplicating entries. To evaluate GEOSPOT across modalities, we separately use the georeferenced images (GeoYFCC-Image) and text captions (GeoYFCC-Text).

fMoW-WILDS. fMoW-WILDS (Koh et al., 2021), adapted from the Functional Map of the World dataset (Christie et al., 2018), contains 224×224 RGB satellite images labeled with one of 62

land-use or building classes. We focus on spatial shifts and ignore the temporal component. To ensure reliable training and evaluation, we retain countries with abundant training data spanning most classes, yielding 152,196 samples from 25 countries.

GeoDE. GeoDE (Ramaswamy et al., 2023) contains 61,940 crowd-sourced images, approximately balanced across 40 common object classes and 6 global regions. To ensure sufficient data per domain, we retain countries with at least 1,000 samples, yielding 61,378 images from 19 countries.

Geographic Domain Definition. We use countries as geographic domains for all datasets. This follows GeoYFCC, which defines its 62 domains as countries (Dubey et al., 2021), and is the finest partition consistently available across all three datasets. We use this partition since each dataset contains sufficient data at each country level for model training, and using countries allows for a sufficient number of source–target domain pairs. Future work could evaluate alternative geographic domain definitions on suitably constructed datasets.

Embeddings and Model Training. We generate *location embeddings* from longitude and latitude using pretrained SatCLIP (Klemmer et al., 2025) and GeoCLIP (Vivanco Cepeda et al., 2023) encoders. Both require only coordinates at inference time, provide global coverage, and have publicly available weights. They also differ along two key axes: pretraining signal (satellite imagery for SatCLIP versus ground-level photographs for GeoCLIP) and positional encoding (spherical harmonics versus hierarchical random Fourier features). We extract *feature embeddings* using commonly used pretrained models: ResNet50 pretrained on ImageNet for images (He et al., 2016) and BERT for text (Devlin et al., 2019).

We compute GeoSpOT dataset distances with the Python Optimal Transport (POT) library (Flamary et al., 2021; 2024). For efficiency, we use the entropically regularized Sinkhorn divergence rather than exact OT, with regularization parameter $\epsilon = 0.01$. For downstream training, we fine-tune pretrained ResNet50 models for image tasks and BERT models for text tasks, replacing the original classifier head with a linear layer matching the number of target classes. We report additional results for a ViT on GeoYFCC-Image and fMoW-WILDS datasets in Appendix Figure 6. We use separate learning rates for the pretrained backbone and new head to preserve pretrained representations while allowing faster adaptation of the task-specific layer.

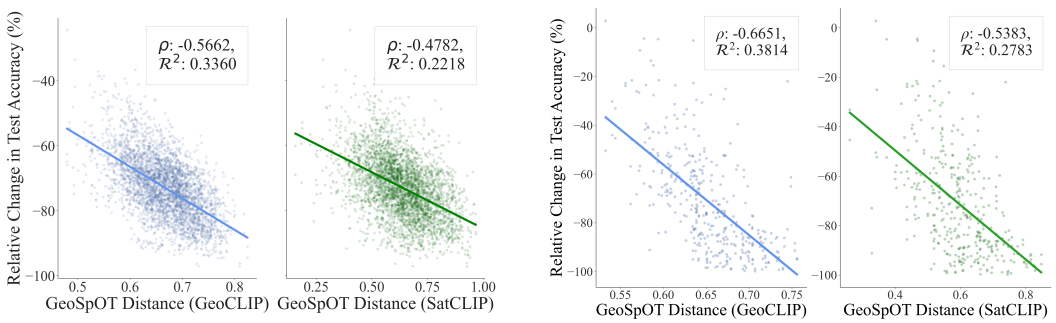
5 Predicting Domain Transfer Performance

We assess the quality of GeoSpOT distances by examining its correlation with out-of-distribution model performance of a trained classifier across different domain pairs. For each source-target domain pair $(\mathcal{D}_s, \mathcal{D}_t)$, we train two models $(\mathcal{M}_s, \mathcal{M}_t)$ on the respective training sets of $\mathcal{D}_s$ and $\mathcal{D}_t$, and we evaluate both models on the test set of the target domain. We measure the relative change (Δ) in test performance with respect to in-distribution training as

$$\Delta_{\mathcal{D}_s,\mathcal{D}_t} = \frac{\text{Acc}_{\mathcal{D}_t}(\mathcal{M}_s) - \text{Acc}_{\mathcal{D}_t}(\mathcal{M}_t)}{\text{Acc}_{\mathcal{D}_t}(\mathcal{M}_t)} \times 100\%, \tag{4}$$

where $\text{Acc}_{\mathcal{D}_t}(\mathcal{M}_s)$ represents performance of a model transferred from the source domain to the target domain, and $\text{Acc}_{\mathcal{D}_t}(\mathcal{M}_t)$ represents in-distribution performance of a model trained and tested on the target domain. Values closer to zero indicate better transferability, while larger negative values reflect poorer cross-domain generalization. To evaluate how well GeoSpOT distances correlate with domain transfer performance, we report Spearman's rank correlation (ρ) and the coefficient of determination ($\mathcal{R}^2$) of GeoSpOT distances with $\Delta_{\mathcal{D}_s,\mathcal{D}_t}$ across pairs of source and target domains, which is standard practice in the dataset-distance and transferability-estimation literature (Alvarez-Melis & Fusi, 2020).

Across all datasets (see Appendix Figure 5), transfer performance plots exhibit the expected negative correlation with $\Delta_{\mathcal{D}_s,\mathcal{D}_t}$, i.e., larger GeoSpOT distances are associated with larger drops in test accuracy of target domains, consistently across datasets, which demonstrates that GeoSpOT can predict transfer performance across diverse datasets such as image, text, and remote-sensing tasks. In Figure 2, we show results on GeoYFCC-Text and fMoW-WILDS to exemplify GeoSpOT's predictive performance on datasets distinct from the pretraining datasets of the location encoders. While GeoSpOT-GeoCLIP may be a strong predictor of transfer performance on GeoYFCC-Image (Appendix Figure 5), since GeoCLIP is pretrained on a similar modality, we also see GeoSpOT-GeoCLIP is a strong predictor of transfer performance on fMoW-WILDS dataset (Figure 2b; with



(a) **GeoYFCC-Text** (b) **fMoW-WILDS**

Figure 2: **GeoSpOT distances correlate with transfer performance (i.e., performance difference) between train and test domains**, measured in zero-shot transfer settings for (a) GeoYFCC-Text and (b) fMoW-WILDS. Points represent source-target domain pairs. For each plot, we show the best-fit line and report Spearman's rank correlation coefficient (ρ) and the coefficient of determination ($\mathcal{R}^2$). Analogous plots for the remaining datasets can be found in the Appendix.

$\rho \approx -0.65$), even though the task is predicting land use category from satellite images, and GeoYFCC-Text (Figure 2a; with $|\rho| \approx 0.57$). This suggests that GeoSpOT distances capture spatial variation associated with transfer difficulty across different data modalities and tasks. Additionally, while GeoSpOT-SatCLIP also correlates with transfer performance (e.g., $|\rho| \approx 0.54$ on fMoW-WILDS), notably, it is outperformed by GeoSpOT-GeoCLIP on this dataset ($|\rho| \approx 0.67$), even though fMoW-WILDS is a dataset of satellite images. Although the scatter plots in Figure 2 and Appendix Figure 5 show not all variation in transfer performance is captured by GeoSpOT, these results demonstrate that GeoSpOT provides a useful, label-free proxy for assessing the difficulty of transferring between geographic domains.

6 Comparison with Other Dataset Distances

Aggregate Transfer Prediction. To further evaluate GeoSpOT as a formulation for geographic domain distance, we compare it against two sets of dataset distance baselines. First, we compare GeoSpOT with alternative *location-based* distances to assess whether an optimal transport formulation provides a more informative measure of geographic dataset similarity. We consider both non-distributional distances, including average geodesic distance and average cosine similarity, and distributional distances, including Maximum Mean Discrepancy (MMD) and Fréchet Distance (FD). Detailed formulations and implementation details for these baseline distances are provided in Appendix Section A.1. Second, we compare dataset distances on location embeddings with *input-modality* (feature-based) dataset distances computed on image/text embeddings. Because the latter methods have access to the full input modality, our objective is not necessarily to outperform them, but rather to assess how effectively geographic information alone can predict transferability despite using substantially less information about the datasets.

Among distances computed solely from GeoCLIP embeddings, GeoSpOT achieves the strongest results across all four datasets, outperforming average cosine distance, MMD, and FD. With SatCLIP embeddings, GeoSpOT similarly outperforms both MMD and FD on GeoYFCC-Image and GeoYFCC-Text and achieves the strongest $\mathcal{R}^2$ on fMoW-WILDS. Moreover, across all datasets, GeoSpOT with location embeddings substantially outperforms distances computed directly from raw coordinates using geodesic distance. Notably, average geodesic distance explains almost none of the variation in transfer performance ($\mathcal{R}^2 \leq 0.03$ on all datasets), suggesting that geographic proximity alone is a poor guide to transferability between these domains. Together, these comparisons show that both components of GeoSpOT are important: modeling domains as distributions with optimal transport and representing locations using semantically meaningful embeddings.

Next, we compare location-based distances with distances computed directly from input-modality (IM) embeddings. Despite having no access to the downstream input data, location-only GeoSpOT is competitive with, and often better than, input-modality distances. GeoSpOT-GeoCLIP outper-

Table 1: **Computed on the same location embeddings, GEOSPOT predicts geographic transfer performance better than cosine, MMD and FD distances, and approaches or exceeds distances computed on input-modality (IM) embeddings.** Entries report $|\rho|/\mathcal{R}^2(\uparrow)$ between dataset distance and $\Delta_{\mathcal{D}_s,\mathcal{D}_t}$ across all domain pairs. GEOSPOT uses location embeddings, while OT baselines use ResNet50/BERT embeddings of the input modality. Rows are grouped by the embedding space the distance is computed in; best results within each embedding-space group are underlined, and best results for a dataset overall are in bold.

Embedding space	Dataset distance	GeoYFCC Image	GeoYFCC Text	fMoW-WILDS	GeoDE
Coordinates only					
Geodesic	Avg. Geodesic	0.14 / 0.01	0.22 / 0.03	0.02 / 0.00	0.01 / 0.00
	MMD	0.45 / 0.21	0.46 / 0.21	0.22 / 0.08	0.03 / 0.01
	OT	0.14 / 0.02	0.43 / 0.16	0.02 / 0.01	0.01 / 0.00
IM embeddings					
ResNet50/BERT	Avg. Cosine	0.17 / 0.06	0.01 / 0.00	0.27 / 0.05	0.28 / 0.07
	MMD	0.70 / 0.44	0.54 / 0.17	0.60 / 0.15	0.60 / 0.35
	FD	**0.74** / 0.54	0.53 / 0.27	0.59 / 0.20	0.35 / 0.21
	OT	0.72 / 0.52	0.41 / 0.16	0.49 / 0.16	**0.63 / 0.39**
Location embeddings					
GeoCLIP	Avg. Cosine	0.32 / 0.09	0.29 / 0.08	0.39 / 0.19	0.38 / 0.14
	MMD	0.66 / 0.43	0.52 / 0.29	0.65 / 0.35	0.51 / 0.25
	FD	0.46 / 0.20	0.50 / 0.27	0.61 / 0.37	0.34 / 0.14
	GEOSPOT	0.70 / 0.48	0.57 / 0.34	**0.67 / 0.38**	0.54 / 0.31
SatCLIP	Avg. Cosine	0.45 / 0.20	0.41 / 0.15	0.56 / 0.27	0.28 / 0.08
	MMD	0.43 / 0.18	0.37 / 0.14	0.41 / 0.14	0.04 / 0.01
	FD	0.37 / 0.15	0.44 / 0.19	0.55 / 0.27	0.29 / 0.10
	GEOSPOT	0.55 / 0.30	0.48 / 0.22	0.54 / 0.28	0.22 / 0.06
Location + IM embeddings					
IM + Geodesic	OT	0.60 / 0.30	0.46 / 0.18	0.36 / 0.14	0.21 / 0.09
IM + GeoCLIP	GEOSPOT-FUSE	**0.74 / 0.56**	**0.60 / 0.37**	0.63 / 0.27	0.59 / 0.37
IM + SatCLIP	GEOSPOT-FUSE	0.67 / 0.45	0.50 / 0.24	0.60 / 0.29	0.29 / 0.10

forms all ResNet50-based distances on fMoW-WILDS and exceeds all BERT-based distances on GeoYFCC-Text. On GeoYFCC-Image, GEOSPOT performs comparably to OT on ResNet50 features. Additionally, incorporating both sources of information further improves transfer prediction: GEOSPOT-FUSE with GeoCLIP achieves the strongest results on GeoYFCC-Image and GeoYFCC-Text and improves over the corresponding location-only GEOSPOT variants. These results show that location embeddings capture transfer-relevant information that is not fully contained in input-modality embeddings and can therefore provide a complementary signal for predicting geographic domain transfer. Interestingly, we see that on GeoDE, GEOSPOT-GeoCLIP outperforms GEOSPOT-SatCLIP ($|\rho| \approx 0.54$ vs. $|\rho| \approx 0.22$). This is likely due to the differences in the location encoders' pretraining data and the downstream task: GeoCLIP, pretrained on ground-level social media images, can capture the variation in GeoDE, which is composed of crowd-sourced photographs of household objects. In contrast, SatCLIP, pretrained on satellite imagery, does not.

Transfer Prediction from a Fixed Source Domain. We also consider the practically relevant setting in which the source domain is fixed; this setting can determine if the dataset distance can identify target regions where the specific model will transfer successfully. In Table 2, we see that GEOSPOT-GeoCLIP exhibits the strongest association with transfer performance for both the US and all source domains on average (full results in Appendix Table 4). However, we do note that the domain distance measures have notable variation between the source domains as seen in the standard deviations, so GEOSPOT may not be uniformly the strongest domain distance measure for every source domain.

Sensitivity to Hyperparameter λ in GEOSPOT-FUSE. In Table 1, we use GEOSPOT-FUSE with a default $\lambda = 0.5$, which controls the contribution of location relative to input-modality

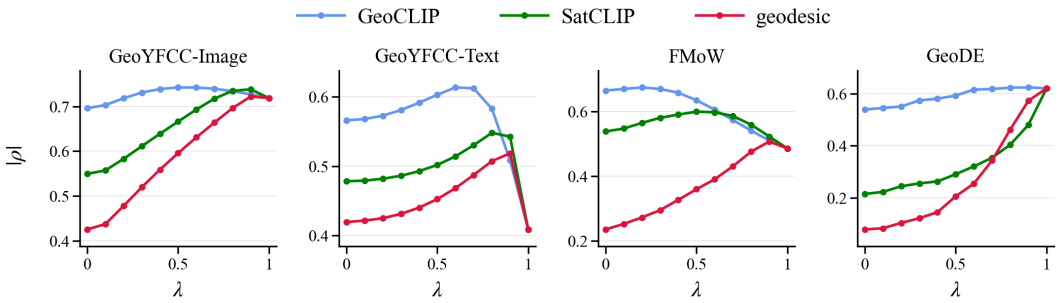


Figure 3: **Predictive capabilities of GEOSPOT-FUSE can be improved by tuning λ.** We analyze the rank correlation coefficient (ρ) for different values of λ in GEOSPOT-FUSE, where $\lambda = 0$ is location-only GEOSPOT, $\lambda = 1$ is input-modality (IM)-only OT, $\lambda = 0.5$ represents our default value for GEOSPOT-FUSE, and $0 < \lambda < 1$ is analyzed for the sensitivity analysis.

information. Although we do not expect this value to be optimal, we use it to avoid assuming prior knowledge of the domains or optimal weighting. We examine sensitivity to λ in Figure 3. Across all four datasets, GEOSPOT-FUSE with GeoCLIP generally yields the flattest curves, indicating relatively small gains from adding input-modality information. This suggests that GeoCLIP already captures much of the signal relevant to OT-based transfer prediction. In contrast, SatCLIP- and Geodesic-based distances generally benefit more from input-modality information, with steeper gains as λ increases, except for SatCLIP on fMoW-WILDS.

Table 2: Association between dataset distance and transfer performance for the US source domain and averaged across all source domains on GeoYFCC-Text. The US column reports the single $|\rho|$ and $\mathcal{R}^2$ value for this source domain, while "All sources" reports the mean $\pm$ standard deviation across source domains.

	US		**All sources**	
	$\lvert\rho\rvert$	$\mathcal{R}^2$	$\lvert\rho\rvert$	$\mathcal{R}^2$
Cosine	0.29	0.14	0.28±0.18	0.12±0.10
FD	0.38	0.13	0.41±0.17	0.22±0.12
MMD	0.34	0.14	0.45±0.13	0.23±0.12
GEOSPOT	**0.40**	**0.22**	**0.46±0.14**	**0.26±0.11**

For GeoYFCC (Image and Text) and fMoW-WILDS, the strongest association for each GEOSPOT-FUSE variant occurs at an intermediate λ, suggesting complementary information is present in location and input-modality embeddings. While $\lambda = 0.5$ improves the GeoCLIP variant on both GeoYFCC datasets, it is consistently competitive across all variants only on GeoYFCC-Text. Nonetheless, GEOSPOT remains comparable to input-modality-only OT in all but one setting (SatCLIP on GeoDE). Given this sensitivity across datasets and location embeddings, we leave adaptive tuning of λ to future work.

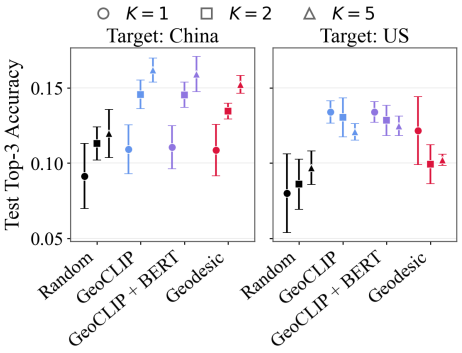
7 TARGET-AWARE CONSTRAINED DATASET SELECTION

Lastly, we investigate whether GEOSPOT distances can guide data subset selection to improve model performance under a constrained training budget, a setting that mirrors common challenges in real-world geospatial ML deployment. This is particularly useful with GEOSPOT, as it is based on location-only inputs; as such, informed dataset selection is possible before any training data is actually collected. For these experiments, we design an algorithm (Figure 4 (a)), which greedily selects K source domains by iteratively minimizing the GEOSPOT distances of the pooled distribution and the target domain; we then sample equally from selected domains up to the training budget. We compare against two baselines under the same training budget: (1) *random subset selection*, where K source domains are selected uniformly at random and sampled in equal proportions, and (2) *geodesic-OT subset selection*, which applies the same selection procedure using OT with raw geodesic distance, to isolate the utility of using location embeddings over geographic proximity.

Figure 4 shows the results on the GeoYFCC-Text dataset for two target countries (China and the United States) using $K \in \{1, 2, 5\}$ source domains under a total budget of $N = 2,000$ data points. Results for additional datasets and budgets are provided in Appendix Section F. Notably, all methods outperform Random for each K and for both target domains. For both China and the US, GEOSPOT-

Input: Source domains $\{\mathcal{D}_i^s\}_{i=1}^N$; target domain $\mathcal{D}^t$;
number of source domains $K \leq N$.
Output: Selected source domains $\mathcal{S}_K$.
$\mathcal{I}_0 \leftarrow \{1, \ldots, N\}$
$\mathcal{S}_0 \leftarrow \emptyset$
for $t = 1$ **to** K **do**
  $i_t \leftarrow \arg\min_{i \in \mathcal{I}_{t-1}} \text{GEOSPOT}\left(\mathcal{S}_{t-1} \cup \{\mathcal{D}_i^s\}, \mathcal{D}^t\right)$
  $\mathcal{I}_t \leftarrow \mathcal{I}_{t-1} \setminus \{i_t\}$
  $\mathcal{S}_t \leftarrow \mathcal{S}_{t-1} \cup \{\mathcal{D}_{i_t}^s\}$

(a) Domain selection algorithm



(b) Domain-selection performance

Figure 4: **GEOSPOT enables target-aware source-domain selection.** (a) Our greedy domain selection algorithm iteratively selects the source domain that minimizes the GEOSPOT distance between the aggregated selected domains and the target domain. (b) Zero-shot target-domain performance on GeoYFCC-Text of models trained on $N = 2{,}000$ samples obtained using different source-domain selection methods. Models are trained for 5 random seeds. Since GeoYFCC has 1,261 classes, we report the top-3 accuracy.

GeoCLIP and GEOSPOT-FUSE with GeoCLIP+BERT perform similarly, while Geodesic-OT tends to underperform for some values of K. Increasing K tends to improve performance for China but interestingly, can hurt performance for the United States. Additionally, the decrease in performance observed when increasing K in the US is smaller for GEOSPOT and GEOSPOT-FUSE than it is for Geodesic-OT. We also note that the greedy algorithm selects complementary, not necessarily nearest, domains. For China, GeoCLIP and GeoCLIP+BERT select Taiwan then Hong Kong, while Geodesic-OT selects Taiwan then South Korea. For the US, GeoCLIP selects Canada, Australia, and France; GeoCLIP+BERT selects Canada, Australia, and the UK, likely reflecting language similarity; and Geodesic-OT instead selects Canada, the Bahamas, and Mexico. In these experiments, we test with target domains of the US and China, which are generally regarded as large countries with more heterogeneous data; future work could investigate dataset selection when the target is a smaller geographic domain, with more homogeneous data.

8 DISCUSSION

In this work, we introduce GEOSPOT, a geographically aware, OT-based distance measure computed from pretrained location embeddings. Despite requiring only geographic coordinates, GEOSPOT provides a useful proxy for cross-domain transfer difficulty. We additionally introduce GEOSPOT-FUSE, which combines location embeddings and input-modality features when available, providing complementary information to allow improved prediction of transfer performance. Across three diverse geospatial datasets, GEOSPOT and GEOSPOT-FUSE distances consistently predict cross-domain transfer difficulty. The performance of GEOSPOT distances based on location embeddings alone is particularly noteworthy, demonstrating that meaningful, task-agnostic estimates of geospatial domain transferability can be obtained *even before any observations, features, or labels* are collected. Beyond prediction, we show that GEOSPOT can guide training data selection for zero-shot domain adaptation and visualize where models are likely to transfer successfully—or benefit most from additional data—before any data acquisition occurs.

Our findings suggest several directions for future work. First, GEOSPOT-FUSE could benefit from methods that tune λ without access to target-domain performance. Second, extending GEOSPOT beyond image and text classification, using different geographic domains (e.g., continents, states, regions) and evaluating additional location encoders, including emerging satellite embedding models, would further test its generality. Finally, our dataset selection experiments primarily establish feasibility; future work could improve the selection strategy's effectiveness and scalability.

AI USE STATEMENT

In this work, we have used generative AI tools to aid in implementing methods, namely model training code, subset selection experiments, and visualization code. We also used generative AI to give suggestions on editing the writing to improve readability. We have not used generative AI tools for dataset construction, processing, analysis, or interpreting results, and tasks involving mathematical claims are not relevant to this work.

We have reviewed all AI-assisted work. For example, model training code and subset selection code was manually reviewed and tested by two authors. For writing, no generative AI was given any edit access to this document, and suggestions were manually implemented. No generative AI was used to obtain or format references and citations. We take responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

In the appendix, we describe all datasets in details, including any preprocessing steps. We additionally provide model details and hyperparameters used. We will make our repository public upon acceptance.

REFERENCES

Alessandro Achille, Michael Lam, Rahul Tewari, Avinash Ravichandran, Subhransu Maji, Charless C. Fowlkes, Stefano Soatto, and Pietro Perona. Task2vec: Task embedding for meta-learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.

David Alvarez-Melis and Nicolo Fusi. Geometric dataset distances via optimal transport. In *Advances in Neural Information Processing Systems*, volume 33, pp. 21428–21439, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/f52a7b2610fb4d3f74b4106fb80b233d-Paper.pdf.

Javad Ansarifar, Lizhi Wang, and Sotirios V Archontoulis. An interaction regression model for crop yield prediction. *Scientific reports*, 11(1):17754, 2021.

Georges Le Bellier and Nicolas Audebert. Detecting out-of-distribution Earth observation images with diffusion models. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 481–491, 2024. doi: 10.1109/CVPRW63382.2024.00053.

Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. Analysis of representations for domain adaptation. volume 19, 2006. URL https://proceedings.neurips.cc/paper_files/paper/2006/file/b1b0432ceafb0ce714426e9114852ac7-Paper.pdf.

Gordon Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. Functional map of the world. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6172–6180, 2018. doi: 10.1109/CVPR.2018.00646.

Nicolas Courty, Rémi Flamary, Amaury Habrard, and Alain Rakotomamonjy. Joint distribution optimal transportation for domain adaptation. volume 30, 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/0070d23b06b1486a538c0eaa45dd167a-Paper.pdf.

Ruth Crasto and Esther Rolf. Latent domain modeling improves robustness to geographic shifts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Findings*, pp. 2419–2428, 2026.

Bharath Bhushan Damodaran, Rémi Flamary, Vivien Seguy, and Nicolas Courty. An entropic optimal transport loss for learning deep neural networks under label noise in remote sensing images. *Computer Vision and Image Understanding*, 191:102863, 2020. doi: https://doi.org/10.1016/j.cviu.2019.102863. URL https://www.sciencedirect.com/science/article/pii/S1077314219301559.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, 2019. doi: 10.18653/v1/N19-1423. URL https://aclanthology.org/N19-1423/.

Dajana Dimitrić, Vladimir Risojević, and Mitar Simić. Nearest neighbor based out-of-distribution detection in remote sensing scene classification. In *2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH)*, pp. 1–4, 2023. doi: 10.1109/INFOTEH57020.2023.10094135.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

Abhimanyu Dubey, Vignesh Ramanathan, Alex Pentland, and Dhruv Mahajan. Adaptive methods for real-world domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14340–14349, 2021.

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. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2256–2265, 2025. doi: 10.1109/CVPRW67362.2025.00213.

Zhuoqun Fang, Yuexin Yang, Zhaokui Li, Wei Li, Yushi Chen, Li Ma, and Qian Du. Confident learning-based domain adaptation for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16, 2022. doi: 10.1109/TGRS.2022.3166817.

Marco Federici, Ryota Tomioka, and Patrick Forré. An information-theoretic approach to distribution shifts. In *Advances in Neural Information Processing Systems*, volume 34, pp. 17628–17641, 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/93661c10ed346f9692f4d512319799b3-Paper.pdf.

Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. POT: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021. URL http://jmlr.org/papers/v22/20-451.html.

Rémi Flamary, Cédric Vincent-Cuaz, Nicolas Courty, Alexandre Gramfort, Oleksii Kachaiev, Huy Quang Tran, Laurène David, Clément Bonet, Nathan Cassereau, Théo Gnassounou, Eloi Tanguy, Julie Delon, Antoine Collas, Sonia Mazelet, Laetitia Chapel, Tanguy Kerdoncuff, Xizheng Yu, Matthew Feickert, Paul Krzakala, Tianlin Liu, and Eduardo Fernandes Montesuma. POT python optimal transport (version 0.9.5), 2024. URL https://github.com/PythonOT/POT.

Jakob Gawlikowski, Sudipan Saha, Anna Kruspe, and Xiao Xiang Zhu. Out-of-distribution detection in satellite image classification. In *RobustML Workshop at ICLR 2021*, pp. 1–5, 2021. URL https://elib.dlr.de/142285/.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

Ke Hu, Ashfaqur Rahman, Hari Bhrugubanda, and Vijay Sivaraman. Hazeest: Machine learning based metropolitan air pollution estimation from fixed and mobile sensors. *IEEE Sensors Journal*, 17(11):3517–3525, 2017.

Sarmad F. Ismael, Koray Kayabol, and Erchan Aptoula. Unsupervised domain adaptation for the semantic segmentation of remote sensing images via one-shot image-to-image translation. *IEEE Geoscience and Remote Sensing Letters*, 20:1–5, 2023. doi: 10.1109/LGRS.2023.3281458.

Amandeep Kaur, Mirali Purohit, Gedeon Muhawenayo, Esther Rolf, and Hannah Kerner. Pretrain Where? Investigating how pretraining data diversity impacts geospatial foundation model performance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 7949–7959, 2026.

Daniel Kifer, Shai Ben-David, and Johannes Gehrke. Detecting change in data streams. In *Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30*, VLDB '04, pp. 180–191, 2004.

Konstantin Klemmer, Esther Rolf, Caleb Robinson, Lester Mackey, and Marc Rußwurm. SatCLIP: Global, general-purpose location embeddings with satellite imagery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 4347–4355, 2025.

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, Etienne David, Ian Stavness, Wei Guo, Berton Earnshaw, Imran Haque, Sara M Beery, Jure Leskovec, Anshul Kundaje, Emma Pierson, Sergey Levine, Chelsea Finn, and Percy Liang. 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, 2021. URL https://proceedings.mlr.press/v139/koh21a.html.

Yong-Xiang Lin, Daniel Stanley Tan, Wen-Huang Cheng, Yung-Yao Chen, and Kai-Lung Hua. Spatially-aware domain adaptation for semantic segmentation of urban scenes. In *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 1870–1874, 2019. doi: 10.1109/ICIP.2019.8803192.

Vasileios Linardos, Maria Drakaki, Panagiotis Tzionas, and Yannis L. Karnavas. Machine learning in disaster management: Recent developments in methods and applications. *Machine Learning and Knowledge Extraction*, 4(2):446–473, 2022. ISSN 2504-4990. doi: 10.3390/make4020020. URL https://www.mdpi.com/2504-4990/4/2/20.

Jack Lynch and Sam Wookey. Leveraging domain adaptation for low-resource geospatial machine learning. 2021. URL https://arxiv.org/abs/2107.04983.

Nikhil Makkar, Lexie Yang, and Saurabh Prasad. Adversarial learning based discriminative domain adaptation for geospatial image analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:150–162, 2022. doi: 10.1109/JSTARS.2021.3132259.

Hanna Meyer and Edzer Pebesma. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, 12(9):1620–1633, 2021. doi: https://doi.org/10.1111/2041-210X.13650. URL https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13650.

Yoni Nachmany and Hamed Alemohammad. Detecting roads from satellite imagery in the developing world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019.

Dat Nguyen and Brian Leung. How well do species distribution models predict occurrences in exotic ranges? *Global Ecology and Biogeography*, 31(6):1051–1065, 2022. doi: https://doi.org/10.1111/geb.13482. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.13482.

Vikram V. Ramaswamy, Sing Yu Lin, Dora Zhao, Aaron Adcock, Laurens van der Maaten, Deepti Ghadiyaram, and Olga Russakovsky. GeoDE: a geographically diverse evaluation dataset for object recognition. In *Advances in Neural Information Processing Systems*, volume 36, pp. 66127–66137, 2023. doi: 10.52202/075280-2888. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/d08b6801f24dda81199079a3371d77f9-Paper-Datasets_and_Benchmarks.pdf.

Ievgen Redko, Nicolas Courty, Rémi Flamary, and Devis Tuia. Optimal transport for multi-source domain adaptation under target shift. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 849–858. PMLR, 2019. URL https://proceedings.mlr.press/v89/redko19a.html.

Esther Rolf, Konstantin Klemmer, Caleb Robinson, and Hannah Kerner. Position: Mission critical – satellite data is a distinct modality in machine learning. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=PQ0ERKKYJu.

Josée S. Rousseau and Matthew G. Betts. Factors influencing transferability in species distribution models. *Ecography*, 2022(7):e06060, 2022. doi: https://doi.org/10.1111/ecog.06060. URL https://nsojournals.onlinelibrary.wiley.com/doi/abs/10.1111/ecog.06060.

Shreya Shankar, Yoni Halpern, Eric Breck, James Atwood, Jimbo Wilson, and D. Sculley. No classification without representation: Assessing geodiversity issues in open data sets for the developing world. 2017. URL https://arxiv.org/abs/1711.08536.

Martin Sudmanns, Dirk Tiede, Hannah Augustin, and Stefan Lang. Assessing global sentinel-2 coverage dynamics and data availability for operational Earth observation (eo) applications using the eo-compass. *International Journal of Digital Earth*, 13(7):768–784, 2020. doi: 10.1080/17538947.2019.1572799. URL https://doi.org/10.1080/17538947.2019.1572799.

Yang Tan, Yang Li, and Shao-Lun Huang. OTCE: A transferability metric for cross-domain cross-task representations. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15774–15783, 2021. doi: 10.1109/CVPR46437.2021.01552.

Bart Thomee, David A. Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. YFCC100M: the new data in multimedia research. *Commun. ACM*, 59(2):64–73, 2016. doi: 10.1145/2812802. URL https://doi.org/10.1145/2812802.

Waldo R. Tobler. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46:234–240, 1970. doi: 10.2307/143141. URL https://www.tandfonline.com/doi/abs/10.2307/143141.

Titouan Vayer, Ievgen Redko, Rémi Flamary, and Nicolas Courty. Co-optimal transport. In *Advances in Neural Information Processing Systems*, volume 33, pp. 17559–17570, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/cc384c68ad503482fb24e6d1e3b512ae-Paper.pdf.

Vicente Vivanco Cepeda, 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*, volume 36, pp. 8690–8701. Curran Associates, Inc., 2023. doi: 10.52202/075280-0379. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/1b57aaddf85ab01a2445a79c9edc1f4b-Paper-Conference.pdf.

Huaxiu Yao, Xinyu Yang, Xinyi Pan, Shengchao Liu, Pang Wei Koh, and Chelsea Finn. Improving domain generalization with domain relations. In *International Conference on Learning Representations*, volume 2024, pp. 42624–42646, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/ba9fed4258bf3eac96b02cc99f17a82b-Paper-Conference.pdf.

Mikhail Yurochkin, Sebastian Claici, Edward Chien, Farzaneh Mirzazadeh, and Justin Solomon. Hierarchical optimal transport for document representation. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/8b5040a8a5baf3e0e67386c2e3a9b903-Paper.pdf.

Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.