Here, we provide some more details about our experiments.

- In Sec. A, we provide the datasheet for GeoDE, as in [4].
- In Sec. B, we describe our heuristic to select object categories in more detail.
- In Sec. C, we compare the GeoDE feature space to that of ImageNet [2].
- In Sec. D, we provide results when finetuning pre-trained models rather than just training the final layer of a ResNet.
- In Sec. E, we give more details about GeoDE, including the counts of images of different regions and categories, as well as more sample images from this dataset.

A Datasheet for GeoDE

We include the datasheet for GeoDE below, based on Datasheets for Datasets [4].

MOTIVATION

For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

GeoDE was created for 2 purposes: (1) To construct a more geographically diverse dataset for training and evaluation and (2) to understand what it would take to crowd-source an image dataset from scratch.

Who created this dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

GeoDE was created via a collaboration between the Princeton Visual AI lab and Meta research.

What support was needed to make this dataset? (e.g., who funded the creation of the dataset? If there is an associated grant, provide the name of the grantor and the grant name and number, or if it was supported by a company or government agency, give those details.)

Creation of GeoDE was partially supported by the National Science Foundation under Grant No. 2145198. It was also supported by Meta AI and the Princeton SEAS Howard B. Wentz, Jr. Junior Faculty Award to OR.

Any other comments?

N/A
COMPOSITION

What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.

GeoDE consists of images of 40 different categories from 6 different regions. These categories and regions are fully listed in tables 2 and 3 of the main paper. Additionally, meta data for each image includes the type of phone used to take the picture, the GPS coordinates of the image, whether there are people present in the background of the image (note that there are no recognizable people in the dataset), and if a large fraction of the image consists of trees.

How many instances are there in total (of each type, if appropriate)?

There are 61,940 images.

Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).

GeoDE consists of a sample of images. The larger set would include all possible objects from all possible regions of the world. GeoDE is balanced across 6 regions and 40 objects, (regions chosen to maximise geodiversity).

What data does each instance consist of? “Raw” data (e.g., unprocessed text or images) or features? In either case, please provide a description.

Each instance consists of an image, location, object in the image, whether there are people in the background, whether a significant portion of the image contains trees.

Is there a label or target associated with each instance? If so, please provide a description.

Yes, each image is labelled with one of 40 objects.

Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.

No

Are relationships between individual instances made explicit (e.g., users’ movie ratings, social network links)? If so, please describe how these relationships are made explicit.

N/A

Are there recommended data splits (e.g., training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.

No, we created training, validation and test splits by randomly splitting the dataset.

Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.

Possible errors include issues with labels of objects and countries. To the best of our knowledge this is limited to lesser than 1% of the dataset.
Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they existed at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a future user? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.

GeoDE is self-contained.

Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals’ non-public communications)? If so, please provide a description.

No, GeoDE does not contain confidential data.

Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.

No.

Does the dataset relate to people? If not, you may skip the remaining questions in this section.

No.

Does the dataset identify any subpopulations (e.g., by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.

N/A.

Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset? If so, please describe how.

N/A.

Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals racial or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.

N/A.

Any other comments?

N/A.

---

**COLLECTION**

How was the data associated with each instance acquired? Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.

Participants from across the world took photos of different objects and submitted it. They were compensated for their efforts. This data was manually checked and verified to contain the object.
Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created. Finally, list when the dataset was first published.

Data was collected in 2022. It was first publicly released in January 2023.

What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)? How were these mechanisms or procedures validated?

We used manual human curation: participants took photos of different objects. All images were verified by Appen’s quality analysis team.

What was the resource cost of collecting the data? (e.g. what were the required computational resources, and the associated financial costs, and energy consumption - estimate the carbon footprint. See Strubell et al. [9] for approaches in this area.)

Total cost for all images was $54,000, not including researcher time. There were no models involved in collecting the dataset.

If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?

It is not a sample from a larger dataset.

Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?

Participants from across the world were tasked with taking photos of specific objects and paid for their time. We partnered with Appen (www.appen.com), who recruited and compensated the workers. Compensation varied depending on the region, but we got assurances that the pay was appropriate for the work.

Were any ethical review processes conducted (e.g., by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.

No.

Does the dataset relate to people? If not, you may skip the remainder of the questions in this section.

No.

Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (e.g., websites)?

N/A

Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.

N/A

Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the
individuals consented.

N/A

If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate)

N/A

Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.

N/A

Any other comments?

N/A

PREPROCESSING / CLEANING / LABELING

Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remainder of the questions in this section.

No.

Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)? If so, please provide a link or other access point to the “raw” data.

N/A

Is the software used to preprocess/clean/label the instances available? If so, please provide a link or other access point.

N/A

Any other comments?

N/A

USES

Has the dataset been used for any tasks already? If so, please provide a description.

GeoDE has been used to evaluate large scale models for geographical bias in this paper.

Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.

No.

What (other) tasks could the dataset be used for?

Additional uses of GeoDE could be trying to understand geodiversity of current datasets, to understand how webscraping and crowd-collected images differ (a small analysis done in section 5 of the main paper).
Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a future user might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other undesirable harms (e.g., financial harms, legal risks)? If so, please provide a description. Is there anything a future user could do to mitigate these undesirable harms?

All images in GeoDE were collected by participants with smartphones. Thus, the dataset does not exhibit economic diversity.

Are there tasks for which the dataset should not be used? If so, please provide a description.

N/A

Any other comments?

N/A

DISTRIBUTION

Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.

How will the dataset will be distributed (e.g., tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?
GeoDE is available as a .zip file to download.

When will the dataset be distributed?
GeoDE is currently available.

Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.
No, GeoDE is released under a CC-BY license.

Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.
No

Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.
No

Any other comments?

N/A
MAINTENANCE

Who is supporting/hosting/maintaining the dataset?
Currently, GeoDE is being hosted by the Princeton computer science department, specifically, Dr. Vikram Ramaswamy and Prof. Olga Russakovsky. For the long term, we are considering one of two options: partnering with Common Visual Data Foundation (CVDF; [http://www.cvdfoundation.org/](http://www.cvdfoundation.org/)) or utilizing [https://researchdata.princeton.edu/news/2023-05-25/coming-soon-princeton-data-commons](https://researchdata.princeton.edu/news/2023-05-25/coming-soon-princeton-data-commons) (we’ve seen internal versions which look great for our use cases but are waiting for it to become public).

How can the owner/curator/manager of the dataset be contacted (e.g., email address)?
Questions can be emailed to vr23@cs.princeton.edu

Is there an erratum? If so, please provide a link or other access point.
No.

Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)?
If so, please describe how often, by whom, and how updates will be communicated to users (e.g., mailing list, GitHub)?
Yes, the dataset will be updated as needed, by Vikram Ramaswamy. Updates will be posted on the GitHub repo along with the website, on how to access the corrected version.

If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.
N/A

Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to users.
No, older versions will not continue to be hosted, however, we will provide information on our github as well as the webpage, with a script to update the dataset (if applicable).

If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to other users? If so, please provide a description.
No, there is no current mechanism to do so. Users can provide feedback / corrections to the dataset on github, which we will use to update the dataset.

Any other comments?
N/A

B Selecting object categories for GeoDE

In this section, we provide more details about the object selection heuristic we employed. We mainly used the GeoYFCC [3] dataset that was constructed to be geodiverse.
Leave one out training

curler, fan, footstool, chili, coconut, toilet, canoe, motorboat, mountain-bike, stupa, villa, backpack,
baseball-glove, basin, basket, bat, bathtub, battery, beer-mug, belt, blade, bowl, bowl, broom, bucket, carryall,
case, cash-machine, cassette, cleaver, cologne, cooler, counter, dinner-dress, dinner-jacket, dish, gown, grille,
hammer, jacket, kettle, microphone, parka, porch, rack, remote-control, sandal, scale, shelf, shot-glass, stereo,
stocking, stool, sweater, tape, teddy, timer, tripod, trouser, turntable, wardrobe, weight, wok, woodcarving,
hot-pot, chewing-gum, cucumber, lime, fig, pineapple, jackfruit, kiwi, mango, basil, garlic, sage, lager, ale,
porter, stout, champagne, rum, tequila, vodka, whiskey, mocha

Linear SVM for region

mountain-bike, bicycle, raft, ferry, ship, kayak, streetcar, bus, impala, car, footstool, bench, chair, mushroom,
breakfast, vegetable, dessert, dinner, door, bowler-hat, house, building, chandelier, lamp, light, light, castle, acropolis,
fortress, tower, palace, dome, architecture, memorial, statue, sculpture, gravestone, arch, temple, stupa,
monastery, church, cathedral, chapel, mosque, signboard, grocery-store, shop, kitchen, lantern, doll, couti, cork,
prime, alp, shore, curler, cologne, seashore, gnu, hog, giraffe, arctic, ice-rink, ski, elephant, guinness, makeup,
circuit, geysers, skyscraper, hippopotamus, basketball, painball, sword, hijab, fortification, craft, clock, stage,
tractor, dagger, defile, bikini, swing, windmill, motor, brick, snowboard, course, volleyball, display, opera,
ailing, playground, veranda, wind-instrument, city-hall, ruin, portfolio, newspaper, airbus, bridge, airfield,
global-positioning-system, brake-drum, kid, mangrove, motor-scooter, crane, intersection, plain, column,
wardrobe, interface, guitar, costume, grand-piano, aircraft, factory, seaside, ball, sweet, gravy-boat, spotlight,
american-bison, sail, beer, pier, road, tulip, grass, miniskirt, willow, flood, street, roof, slide, cliff, track, train,
vehicle, boot, world, patio, window, rainbow, beacon, sidewalk, organ-pipe, tank, cable-car, grey, hall, map,
cattle, airport, school, mountain, promontory, monkey, motorcycle, bubble, black, mirror, golf-club, skateboard,
computer, university, denim, sky, rock, earphone, descent, garden, hill, library, tea, blush-wine, radio, hill,
sunglass, ballpark, apparel, web, field-glass, reef, fountain, downhill, pen, cable, step, graffiti, conveyance,
facl, hovel, umbrella, iron, cloud, strand, toilet, walker, valley, airplane, cup, base, wire, camel, pizza,
bathroom, lounge, dock, van, circuit-board, bell, sheep, book, fish, canyon, fire, array, rangefinder, coca-cola.

Clustering

acropolis, cork, coati, footstool, stupa, impala, chili, prime, cologne, gnu, guinness, alp, hog, shore, boater,
waller, plain, hippopotamus, raft, chandelier, curler, giraffe, arctic, bowler-hat, castle, geysers, boot, streetcar,
rum, hijab, ski, temple, windmill, dagger, fortification, snowboard, coffee, ice-rink, display, cathedral, bench,
bikini, lantern, slope, elephant, strand, sword, painball, gravestone, tulp, golf-club, downhill, swing, volleyball,
mushroom, monasterey, american-bison, stage, cup, church, wardrobe, wind-instrument, skyscraper, sweet,
course, tower, opera, sketch, circuit, chapel, col, motor, clock, raling, mangrove

Table 1: Prospective tags identified from GeoYFCC [3]. Tags in red seemed hard to picture. Tags in blue are of animals that might be hard to crowdsourc.

Implementation details. Features for GeoYFCC were extracted using a ResNet50 [5] pretrained on ImageNet [2]. We used Logistic regression, Linear SVM and KMeans clustering implementations from the sklearn library [6]. We used continents as regions. GeoYFCC [3] contains over 1200 tags, we ignored all tags with counts in the bottom 20th percentile, giving us a total of 745 tags. After identifying these tags, we first pruned them by removing tags that did not appear to correspond to an object. Examples of this include “arctic”, “descent”, etc. Second, we removed tags corresponding to wild animals, since these would not be found in all regions. Examples of tags removed in this
manner were “gnu”, “camel”, etc. This gave us a list of 265 tags. Third, these tags were sometimes
variants of objects, for example, we had tags like “stupa”, “temple”, “church”, “mosque” and “chapel”.
Thus we grouped tags based on meaning. Other examples included “breakfast”, “dinner” and “dessert”
as “plate of food”, “stool”, “footstool” and “bench” as “chair”, etc. This gave us a list of objects we
could collect, e.g. “religious buildings”, “plate of food”, “toy”.
We also provide the user demographics of the participants (Fig. 1). As shown, images in GeoDE were
provided by people of varying genders, ethnicities and ages.

C Comparison with ImageNet

In this section, we run more comparisons of GeoDE with ImageNet [2] and DollarStreet [8].

C.1 Comparison to ImageNet

We note that the comparison to GeoYFCC [3] in Sec. 5 in the main text required us to use tags which
are noisy. Thus, to compare GeoDE to another web-scraped dataset, we compare GeoDE to ImageNet
in this section, checking how much the feature spaces differ.
We find a subset of ImageNet21k as outlined in Sec. 6, and extract features using a PASS [1] trained
ResNet50 [5] model. Other implementation details remain the same as in Sec 5 in the main paper.
We first use a Logistic regression model to predict the dataset that the features are taken from and this
has an accuracy of 96.0%, showing that the feature space is very different. We also visualize TSNE
plots of different objects in figure 2.

C.2 Comparison with DollarStreet

Compared to GeoDE, DollarStreet [8] contains a lot more categories, thus resulting in much fewer
images per category (on average, the top 40 object categories in DollarStreet contain only 382
images, whereas GeoDE has an average of 1548 images per category). Thus, when filtering images to
comprise of only common categories between both datasets, we end up with much fewer images for
DollarStreet. We run small scale tests to understand how these images differ from those in GeoDE.

Relative value of an image. Similar to the canonical work on dataset bias [10], we measure the
relative value of an image from GeoDE and DollarStreet. That is, we measure the number of
training images needed for strong cross-dataset generalization. Concretely, we select 13 classes
from DollarStreet which (1) appear in GeoDE, and (2) have more than 100 images. We restrict
both datasets to these 13 classes, resulting in 4,788 images for DollarStreet and 17,245 images for
GeoDE. We now extract features using a PASS trained network, and train linear models to predict
the 13 classes. First, we train a baseline model on 250 randomly sampled DollarStreet images and
evaluate it on the remaining DollarStreet images; we then train models with increasing numbers of
images from GeoDE until we match the accuracy of the DollarStreet-trained baseline. This occurs
with 3,000 GeoDE training images. Next, we do a similar process for a baseline trained on 250
Figure 2: We visualize the TSNE plots for several object classes using ImageNet and GeoDE. While the features do overlap slightly, on the whole, they are very different for dataset distributions, even within each category.

Figure 3: We visualize the probabilities assigned by CLIP to different images from the same class as a box plot. We see different scenarios emerge: the variance of scores is large for all regions, as in the case of spices; the variance is large for certain regions, as in the case of stoves from Africa and East and Southeast Asia, or the scores are much lower for a specific region, as in the case of religious buildings in East Asia.

D More results using GeoDE as an evaluation dataset

Here, we provide more analysis performed on using GeoDE as an evaluation dataset. As shown in sec. 5 of the main text, CLIP models perform worse on certain objects (e.g., “house”, “spices”, “medicine”, etc.). Visualizing the probabilities assigned to different images from the same class as a box plot, we see different scenarios emerge: the variance of scores is large for all regions, as in the case of spices; the variance is large for certain regions, as in the case of stoves from Africa and East and Southeast Asia, or the scores are much lower for a specific region, as in the case of religious buildings in East Asia.

E More results when training with GeoDE

In this section, we provide results for the incremental training with GeoDE for different regions that were not presented in Sec 7, and provide results when fine-tuning a ResNet50 model, rather than freezing the layer weights.
We visualize the increase in accuracy as images are incrementally added in from a region. We find that while adding any GeoDE regional images increases the performance of the model on European images, it has a larger effect on the region the images were drawn from.

### E.1 Results from incrementally adding additional regions

We visualize the improvement in the accuracy as we incrementally add images from different regions (Sec. 6.2 in the main text). We can see that the performance both within the specific region and in Europe (compared to Americas when considering Europe) increase with the additional GeoDE data. We see that the increase within the region is larger than that of the control, showing that these images are from different domains. (Fig. 4)

### E.2 Results from finetuning a ResNet50 model

**Implementation details.** We use a ResNet50 model pretrained on Imagenet and fine-tune the weights using different fractions of the ImageNet and GeoDE datasets as mentioned in Sec. 7 in the main paper. We train the model with an SGD optimizer, learning rate = 0.1, and momentum=0.9. Other implementation details remain the same as before.

**Results.** While the overall trend of the results are the same, we see that these results are slightly noisier, potentially because the model overfits to the small training set (Fig. 5).

### E.3 Performance on individual countries

We measure the performance of models on individual countries within each region of GeoDE. We use two different models: first, we use the CLIP model to understand performance of current models on GeoDE. Next, we train a simple linear model of features extracted from PASS for the GeoDE dataset, and measure performance on individual countries. Implementation details for the CLIP model are the same as in Sec. 6 in the main paper, implementation details for the GeoDE model are as in section 7.1 of the main paper. Accuracies are computed for all countries with at least 25 images in the test set.

Tab. 2 summarizes our results for both. When using a pre-trained model like CLIP, we do notice differences between accuracies of countries within each region, however, we note that the overall trend in accuracy remains roughly the same. When training with images from GeoDE, discrepancies between countries within each region further reduces.

### F More details about the dataset

In this supplementary section, we provide counts of the objects per region in GeoDE as well as more examples of images from this dataset.
Figure 5: We visualize the increase in accuracy as images are incrementally added in from a region when fine-tuning a ResNet50 model. Similar to Fig. 10 in the main text, we see that adding in GeoDE images increases performance more in the region than a control.

<table>
<thead>
<tr>
<th>Region</th>
<th>Average</th>
<th>STD</th>
<th>Minimum</th>
<th>Country</th>
<th>Average</th>
<th>STD</th>
<th>Minimum</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>79.4</td>
<td>2.9</td>
<td>75.4</td>
<td>Nigeria</td>
<td>86.7</td>
<td>1.6</td>
<td>84.5</td>
<td>Egypt</td>
</tr>
<tr>
<td>Americas</td>
<td>84.4</td>
<td>0.2</td>
<td>84.1</td>
<td>Argentina</td>
<td>88.8</td>
<td>1.0</td>
<td>87.7</td>
<td>Argentina</td>
</tr>
<tr>
<td>East Asia</td>
<td>80.2</td>
<td>2.0</td>
<td>77.2</td>
<td>China</td>
<td>89.0</td>
<td>1.1</td>
<td>87.9</td>
<td>Japan</td>
</tr>
<tr>
<td>Europe</td>
<td>85.9</td>
<td>3.6</td>
<td>78.3</td>
<td>Portugal</td>
<td>91.9</td>
<td>1.8</td>
<td>89.2</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>South East Asia</td>
<td>82.7</td>
<td>1.1</td>
<td>81.4</td>
<td>Indonesia</td>
<td>88.3</td>
<td>1.0</td>
<td>87.0</td>
<td>Indonesia</td>
</tr>
<tr>
<td>West Asia</td>
<td>82.3</td>
<td>2.0</td>
<td>79.1</td>
<td>Jordan</td>
<td>88.9</td>
<td>3.6</td>
<td>84.2</td>
<td>Saudi Arabia</td>
</tr>
</tbody>
</table>

Table 2: We measure the performance of a CLIP based model as well as model trained on GeoDE on individual countries within each region. For each region, we report the average accuracy, the standard deviation, and the country with the minimum accuracy for all countries with over 25 images. We find that while a pretrained model does show significant discrepancies among countries within the same region, training on GeoDE data does reduce this.

As mentioned before, GeoDE is mostly balanced across both region and object: for most part, we were able to get at least 150 images per region per object, with a few exceptions (“wheelbarrow” in 2 regions; “monument”, “boat” and “flag” in 1 region). See Tab. 3 for full counts.

We also provide more examples of the images from GeoDE in the figures Figs. 6 to 19.

References

Figure 6: Randomly chosen images for “backyard” for 3 regions. We notice that some of these are backyards made of concrete (West Asia: r2c1, r3c4, etc., Africa: r3c1, r4c9, r5c3, etc.,) or do not contain lawns (West Asia: r3c1, r2c5, etc., Africa: r5c1-5, etc., East Asia: r2c4, r5c1, r5c4-6, etc.)
Figure 7: Randomly chosen images for “backyard” for the 3 other regions. Again, we see that regions tend to have backyards made of concrete or paved (Southeast Asia: r1c8, r2c5, r3c3 as examples, Americas: r1c9, r1c10, r2c10, etc., Europe: r3c2, r4c5, etc.), or do not contain lawns (Southeast Asia: r1c1, r1c9, etc., Americas: r3c2, r3c3, etc., Europe: r3c2, r5c9, etc.)
Figure 8: Randomly chosen images for “bicycle” for 3 regions. While most images are of standard bicycles, we notice a couple of interesting images: tricycle (West Asia: r4c9), rickshaws (Africa: r2c8, r2c10, r4c8, r4c10), and motorized cycles (West Asia: r1c8). There are also a lot of children’s bicycles.
Figure 9: Randomly chosen images for “bicycle” for the 3 other regions. We see more motorized cycles (Southeast Asia: r5c6) as well as several children’s bicycles.
Figure 10: Randomly chosen images for “boat” for 3 regions. We see a variety of boats including larger ships in West Asia (r1c1, r1c2, r1c5, r4c1, r5c1), smaller kayaks and canoes in Africa (r1c8-9, r2c1-4, r4c2, etc.), and a mix in East Asia.
Figure 11: Randomly chosen images for “boat” for the 3 other regions. We again see a variety of boats ranging from motor boats in Europe and the Americas to smaller boats in Southeast Asia.
Figure 12: Randomly chosen images for “cleaning equipment” for 3 regions. This appears to be a diverse category within all regions containing images of mops, buckets, products, brooms, etc.
Figure 13: Randomly chosen images for “cleaning equipment” for the 3 other regions. This appears to be a diverse category within all regions containing images of mops, buckets, products, brooms, etc.
Figure 14: Randomly chosen images for “spices” for 3 regions. We see a wide range of containers, ranging from packets (mostly in Africa), glass jars (in West Asia) to some bottles (all regions).
Figure 15: Randomly chosen images for "spices" for 3 regions. We see a wide range of containers, ranging from packets (some in Southeast Asia and Americas) to bottles (some in Southeast Asia).
Figure 16: Randomly chosen images for “stove” for 3 regions. We see that Africa and East Asia contain one-burner and two burner stoves (along with 4 burner stoves). We also see a variety of stoves in terms of induction, gas, ovens, etc.
Figure 17: Randomly chosen images for “stove” for 3 regions. We see that Southeast Asia contains one-burner and two burner stoves (along with 4 burner stoves). We also see a variety of stoves in terms of induction, coils, gas, ovens, etc.
Figure 18: Randomly chosen images for “waste container” for 3 regions. We see that different regions have containers of varying sizes (Africa seems to be smaller than West Asia or East Asia), and have different closing mechanisms (see West Asia r5c6 as an interesting example.) East Asia also tends to have segregated waste containers.
<table>
<thead>
<tr>
<th>Waste containers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Southeast Asia</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Americas</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Europe</strong></td>
</tr>
</tbody>
</table>

Figure 19: Randomly chosen images for “waste container” for 3 regions. We see that different regions have containers of varying sizes (Europe seems to have containers of very different sizes) and have different closing mechanisms (see Southeast Asia r2c6 as an interesting example.)
<table>
<thead>
<tr>
<th>Object</th>
<th>West Asia</th>
<th>Africa</th>
<th>East Asia</th>
<th>Southeast Asia</th>
<th>Americas</th>
<th>Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td>backyard</td>
<td>216</td>
<td>670</td>
<td>192</td>
<td>218</td>
<td>217</td>
<td>226</td>
</tr>
<tr>
<td>bag</td>
<td>267</td>
<td>397</td>
<td>370</td>
<td>593</td>
<td>298</td>
<td>437</td>
</tr>
<tr>
<td>bicycle</td>
<td>237</td>
<td>257</td>
<td>298</td>
<td>235</td>
<td>228</td>
<td>241</td>
</tr>
<tr>
<td>boat</td>
<td>162</td>
<td>227</td>
<td>174</td>
<td>237</td>
<td>84</td>
<td>222</td>
</tr>
<tr>
<td>bus</td>
<td>203</td>
<td>240</td>
<td>223</td>
<td>214</td>
<td>217</td>
<td>226</td>
</tr>
<tr>
<td>candle</td>
<td>232</td>
<td>244</td>
<td>220</td>
<td>239</td>
<td>188</td>
<td>270</td>
</tr>
<tr>
<td>car</td>
<td>242</td>
<td>331</td>
<td>276</td>
<td>235</td>
<td>273</td>
<td>363</td>
</tr>
<tr>
<td>chair</td>
<td>279</td>
<td>365</td>
<td>326</td>
<td>512</td>
<td>344</td>
<td>349</td>
</tr>
<tr>
<td>cleaning</td>
<td>259</td>
<td>284</td>
<td>307</td>
<td>305</td>
<td>270</td>
<td>361</td>
</tr>
<tr>
<td>equipment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cooking pot</td>
<td>216</td>
<td>270</td>
<td>228</td>
<td>202</td>
<td>213</td>
<td>304</td>
</tr>
<tr>
<td>dog</td>
<td>219</td>
<td>194</td>
<td>185</td>
<td>244</td>
<td>206</td>
<td>193</td>
</tr>
<tr>
<td>dustbin</td>
<td>220</td>
<td>423</td>
<td>266</td>
<td>203</td>
<td>271</td>
<td>294</td>
</tr>
<tr>
<td>fence</td>
<td>259</td>
<td>322</td>
<td>244</td>
<td>302</td>
<td>226</td>
<td>282</td>
</tr>
<tr>
<td>flag</td>
<td>206</td>
<td>265</td>
<td></td>
<td>139</td>
<td>206</td>
<td>272</td>
</tr>
<tr>
<td>front door</td>
<td>210</td>
<td>254</td>
<td>216</td>
<td>224</td>
<td>200</td>
<td>235</td>
</tr>
<tr>
<td>hairbrush/comb</td>
<td>269</td>
<td>255</td>
<td>307</td>
<td>300</td>
<td>290</td>
<td>431</td>
</tr>
<tr>
<td>hand soap</td>
<td>222</td>
<td>208</td>
<td>277</td>
<td>191</td>
<td>245</td>
<td>362</td>
</tr>
<tr>
<td>hat</td>
<td>209</td>
<td>297</td>
<td>337</td>
<td>316</td>
<td>294</td>
<td>336</td>
</tr>
<tr>
<td>house</td>
<td>199</td>
<td>437</td>
<td>208</td>
<td>195</td>
<td>277</td>
<td>194</td>
</tr>
<tr>
<td>jug</td>
<td>217</td>
<td>211</td>
<td>186</td>
<td>249</td>
<td>236</td>
<td>194</td>
</tr>
<tr>
<td>light fixture</td>
<td>234</td>
<td>344</td>
<td>248</td>
<td>209</td>
<td>191</td>
<td>300</td>
</tr>
<tr>
<td>light switch</td>
<td>215</td>
<td>240</td>
<td>246</td>
<td>273</td>
<td>273</td>
<td>234</td>
</tr>
<tr>
<td>lighter</td>
<td>221</td>
<td>312</td>
<td>225</td>
<td>237</td>
<td>217</td>
<td>268</td>
</tr>
<tr>
<td>medicine</td>
<td>242</td>
<td>286</td>
<td>310</td>
<td>330</td>
<td>328</td>
<td>300</td>
</tr>
<tr>
<td>monument</td>
<td>161</td>
<td>191</td>
<td>186</td>
<td>183</td>
<td>254</td>
<td>245</td>
</tr>
<tr>
<td>plate of food</td>
<td>211</td>
<td>480</td>
<td>294</td>
<td>364</td>
<td>241</td>
<td>304</td>
</tr>
<tr>
<td>religious building</td>
<td>222</td>
<td>230</td>
<td>204</td>
<td>226</td>
<td>197</td>
<td>229</td>
</tr>
<tr>
<td>road sign</td>
<td>226</td>
<td>416</td>
<td>258</td>
<td>270</td>
<td>235</td>
<td>284</td>
</tr>
<tr>
<td>spices</td>
<td>243</td>
<td>250</td>
<td>331</td>
<td>216</td>
<td>290</td>
<td>300</td>
</tr>
<tr>
<td>stall</td>
<td>143</td>
<td>215</td>
<td>203</td>
<td>227</td>
<td>197</td>
<td>221</td>
</tr>
<tr>
<td>storefront</td>
<td>209</td>
<td>306</td>
<td>191</td>
<td>240</td>
<td>243</td>
<td>204</td>
</tr>
<tr>
<td>stove</td>
<td>199</td>
<td>553</td>
<td>191</td>
<td>262</td>
<td>206</td>
<td>282</td>
</tr>
<tr>
<td>streetlight/lantern</td>
<td>202</td>
<td>346</td>
<td>211</td>
<td>196</td>
<td>208</td>
<td>227</td>
</tr>
<tr>
<td>toothbrush</td>
<td>264</td>
<td>258</td>
<td>330</td>
<td>361</td>
<td>337</td>
<td>270</td>
</tr>
<tr>
<td>toothpaste/toothpaste</td>
<td>209</td>
<td>288</td>
<td>269</td>
<td>230</td>
<td>245</td>
<td>315</td>
</tr>
<tr>
<td>toy</td>
<td>224</td>
<td>221</td>
<td>280</td>
<td>292</td>
<td>323</td>
<td>287</td>
</tr>
<tr>
<td>tree</td>
<td>226</td>
<td>308</td>
<td>245</td>
<td>357</td>
<td>300</td>
<td>328</td>
</tr>
<tr>
<td>truck</td>
<td>205</td>
<td>246</td>
<td>207</td>
<td>231</td>
<td>212</td>
<td>225</td>
</tr>
<tr>
<td>waste container</td>
<td>231</td>
<td>213</td>
<td>209</td>
<td>213</td>
<td>211</td>
<td>253</td>
</tr>
<tr>
<td>wheelbarrow</td>
<td>122</td>
<td>267</td>
<td>130</td>
<td>197</td>
<td>152</td>
<td>243</td>
</tr>
</tbody>
</table>

Table 3: We show the counts of objects per region in GeoDE. **Bolded** are the ones categories for which we were not able to get 175 images per region.

