GeoDE: A Geographically Diverse Evaluation Dataset

Vikram V. Ramaswamy, Sing Yu Lin, Dora Zhao,
Aaron B. Adcock, Laurens van der Maaten,
Deepti Ghadiyaram, Olga Russakovsky

Rethinking image dataset collection

Large-scale datasets start by scraping images from the web:

- Stereotypes, underrepresentation
  - Who is pictured in images? How are they pictured?
- Geographic bias
  - What countries do images come from?
- Consent
  - Who created the image? Can we use their work to train an AI model?
- Privacy
  - Do people know that their likeness is being used to train an AI model?

Vikram V. Ramaswamy, et al. GeoDE: a Geographically Diverse Evaluation Dataset for Object Recognition. *NeurIPS D&B, 2023.*

Advantages of crowd-sourcing

- Target specific distributions of data
- Have consent; can ensure no recognizable people or PII in images
- Images are not created to generate excitement or novel content

Disadvantages of crowd-sourcing

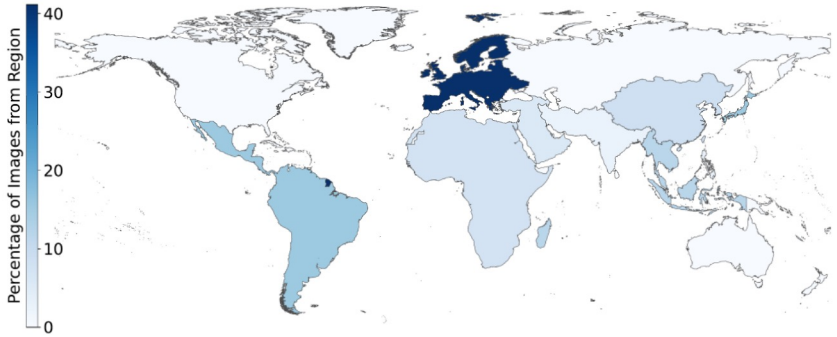
- Cost!

GeoDE: Geographically Diverse Evaluation dataset

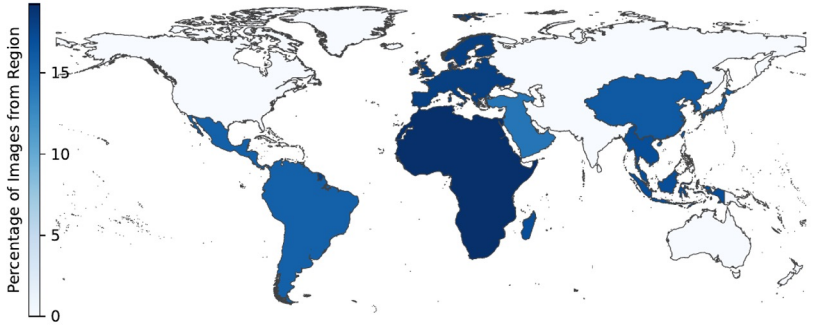
- Solicited photographs from people around the world.
- 61,940 images from 6 different regions and 40 different objects across the world
  - Roughly balanced across region and object.

Vikram V. Ramaswamy, et al. GeoDE: a Geographically Diverse Evaluation Dataset for Object Recognition. *NeurIPS D&B, 2023.*

Takeaway 1: Can ensure specific distributions



GeoYFCC [DRPM'21]



GeoDE

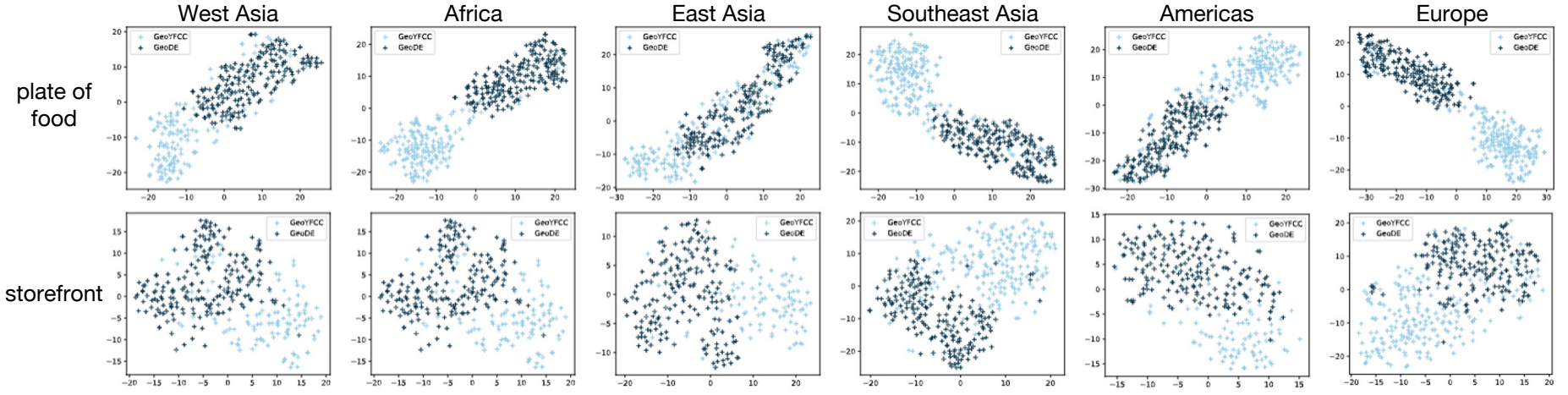
[DRPM'21] Abhimanyu Dubey et. al. Adaptive Methods for Real-World Domain Generalization. *CVPR* 2021.

Takeaway 2: Crowd-sourced images look different

Vikram V. Ramaswamy, et al. GeoDE: a Geographically Diverse Evaluation Dataset for Object Recognition. *NeurIPS D&B, 2023.*

Takeaway 2: Crowd-sourced images look different

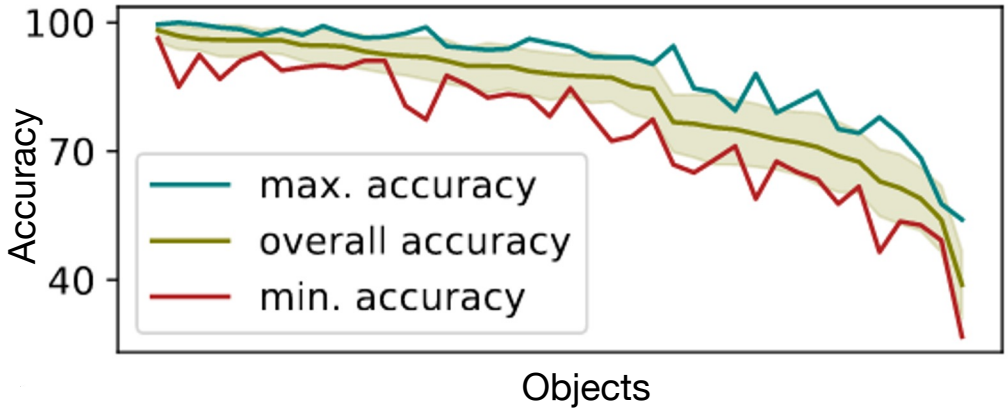
Feature representations between crowd-sourced and web-scraped images are very different



Vikram V. Ramaswamy, et al. GeoDE: a Geographically Diverse Evaluation Dataset for Object Recognition. *NeurIPS D&B, 2023.*

Takeaway 3: GeoDE can find gaps in large models

- Example: CLIP [RKH+'21]
- Discrepancy among regions: 31/40 objects have one region whose accuracy falls outside the 95% confidence interval



[RKH+'21] Alec Radford et al. "Learning transferable visual models from natural language supervision." *ICML*, 2021.

Takeaway 4: Training with GeoDE improves model performance

Adding GeoDE to ImageNet improves performance on DollarStreet [RDK'22]

	Africa	America	Asia	Europe	Overall
ImageNet	45	64	58	75	60
+ GeoDE	**55**	**74**	**68**	**80**	**69**

[RDK+'22] William A. Gaviria Rojas et al. "The Dollar Street Dataset." *NeurIPS D&B.* 2022

Conclusions

- Constructed GeoDE: a crowd-sourced dataset
  - GeoDE helps measuring performance discrepancies across different regions
  - Training with GeoDE can improve model performance
- More generally, given the current issues with webscraping images to construct datasets, crowdsourcing images might be a viable path forward.