How Biased Is Your Dataset?: Using VLM's to Audit Geographic Representation in Indoor Scene Datasets

Morayo Adeyemi, Daniela Galvez-Cepeda, Stefen Dodson, Aditya Ganeshan, Daniel Ritchie

BROWN Computer Science
Google Research

Introduction

- Increasingly, researchers are using **interior scene image datasets to train vision models** for various tasks, including object detection, visual question answering, autonomous navigation, etc.

- Such vision models have been shown to **reflect the biases** in their training data, including racial, gender, and cultural prejudice.

- For interior scene data, one potential form of bias is **geographic bias**, e.g. robots that only perform well in Eurocentric interiors.
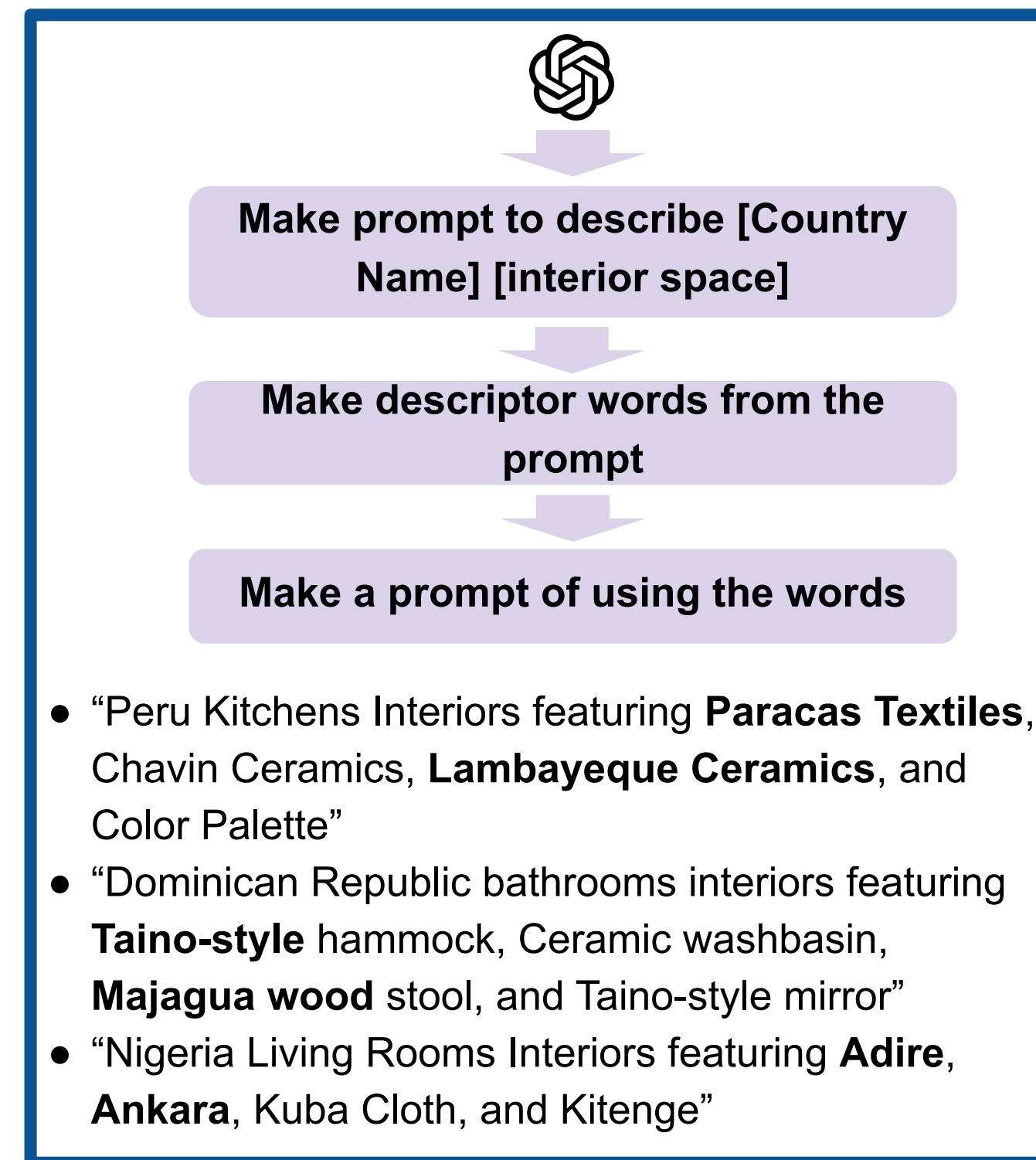
Goal

To determine if datasets of interior spaces (referred to as the training sets) have adequate representation of non-European/American spaces.
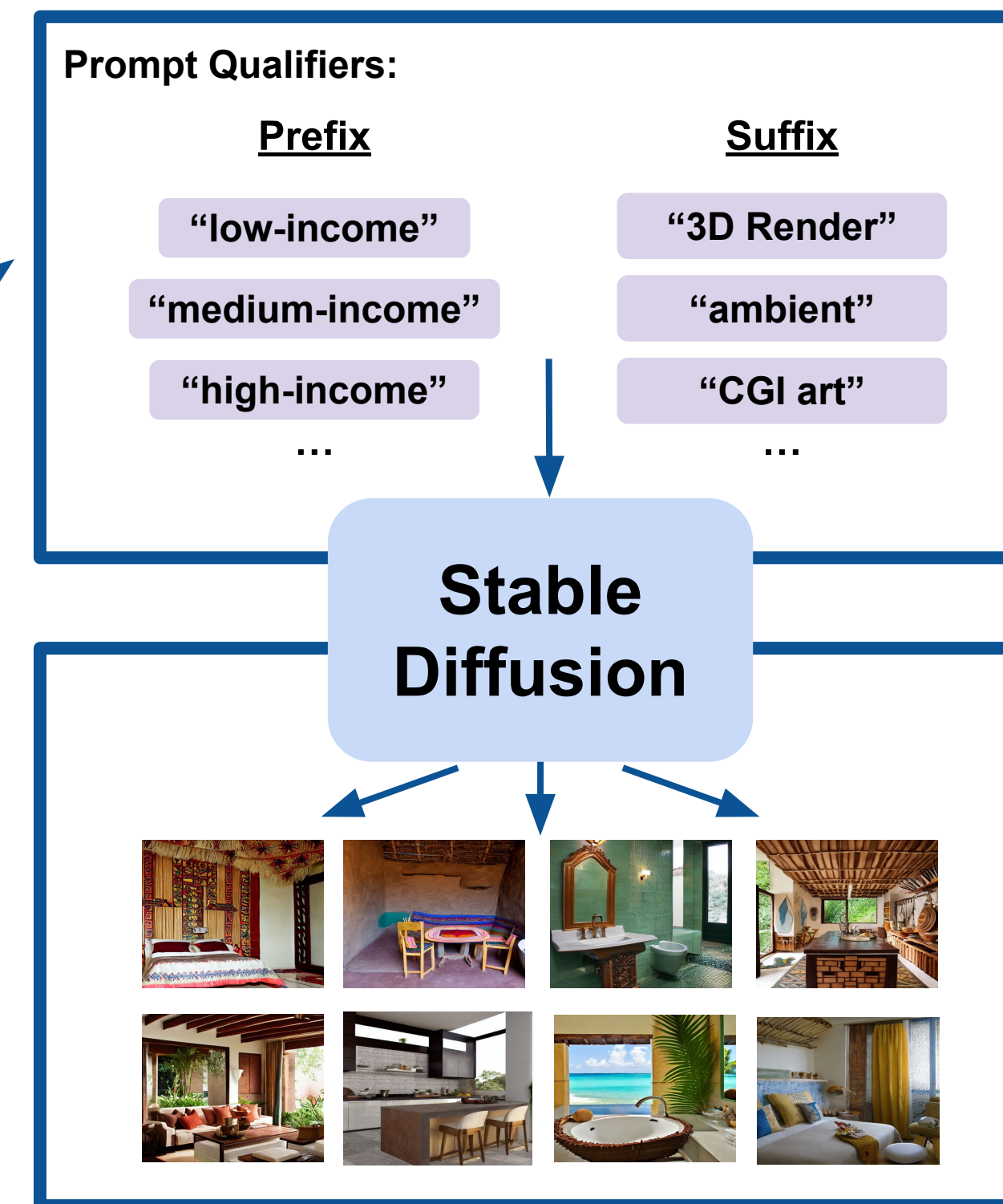
Proposed Approach

1 Create appropriate prompts for a given interior using a Natural Language Processing (NLP) model

2 Use a text-to-image diffusion model to generate images from the NLP prompts

3 Use set distance metrics in CLIP embedding space to measure the similarities between the generated dataset and the training dataset
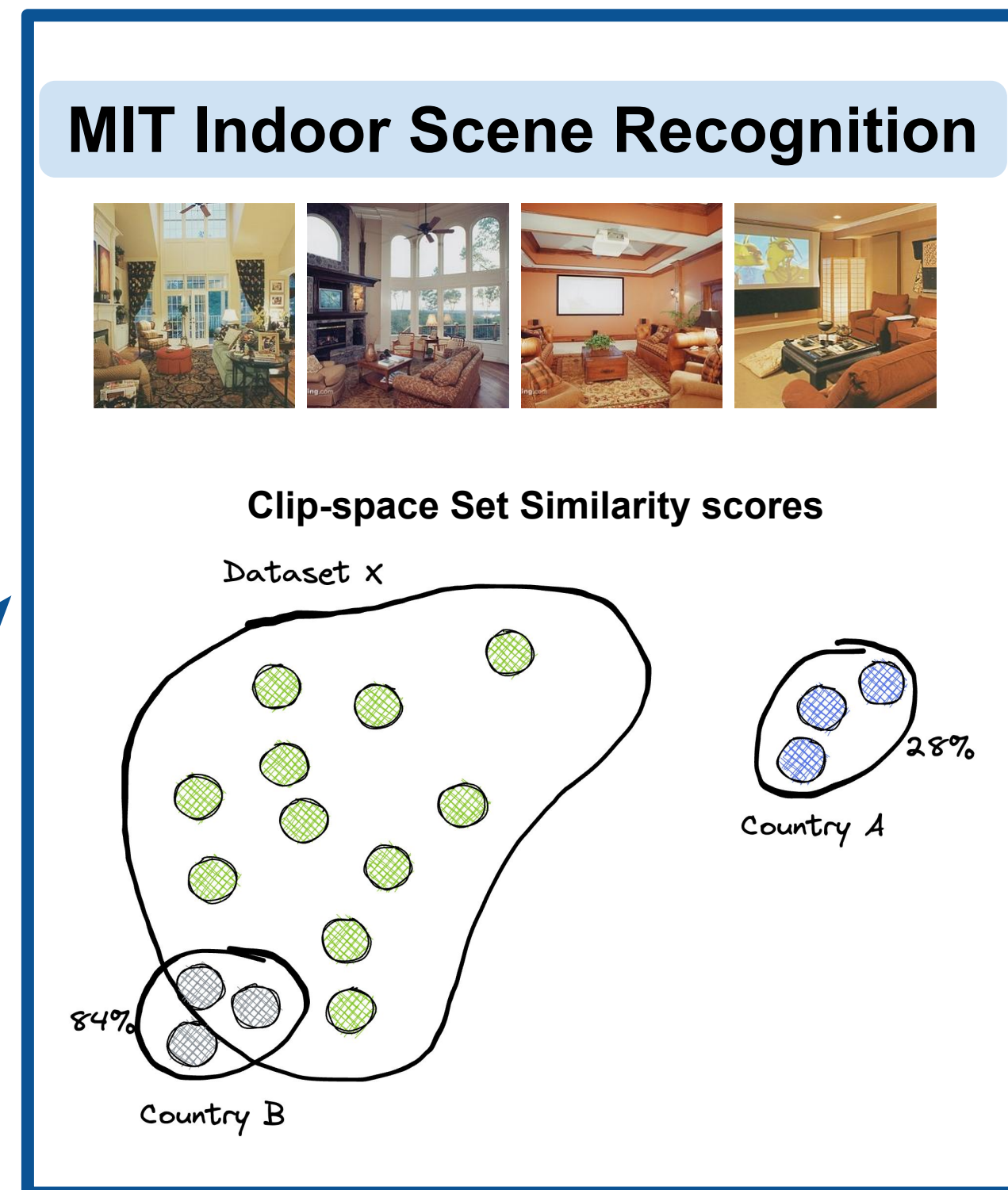
Module 1: Prompts

Make prompt to describe [Country Name] [interior space]

Make descriptor words from the prompt

Make a prompt of using the words

- "Peru Kitchens Interiors featuring **Paracas Textiles**, Chavin Ceramics, **Lambayeque Ceramics**, and Color Palette"
- "Dominican Republic bathrooms interiors featuring **Taino-style** hammock, Ceramic washbasin, **Majagua wood** stool, and Taino-style mirror"
- "Nigeria Living Rooms Interiors featuring **Adire**, **Ankara**, Kuba Cloth, and Kitenge"

Module 2: Images

Prompt Qualifiers:

Prefix	Suffix
"low-income"	"3D Render"
"medium-income"	"ambient"
"high-income"	"CGI art"
...	...

Stable Diffusion

Module 3: Comparisons

MIT Indoor Scene Recognition

Clip-space Set Similarity scores

Dataset X

28% Country A

84% Country B

Comparison with Dataset

	Country Image	Most Similar in HyperSim	Most Similar in MIT	Most Similar in IKEA
United States				
Peru				
Dominican Republic				
Nigeria				

Normalized Chamfer Similarity	Hyper Sim	MIT	Ikea
United States	41%	76%	59%
Peru	37%	65%	43%
Dominican Republic	44%	70%	48%
Nigeria	38%	59%	35%

Conclusion

There is a **lack of geographic representation** in interior scene image datasets. Due to this underrepresentation, systems trained on this data may exhibit **bias towards Eurocentric interiors**.

Future Work

How to improve generated images? **Clip Inversion.**

How to act on this analysis? Use generative VLMs to improve coverage of **different cultures and income levels**.