

Supplementary Material: Inspecting the Geographical Representativeness of Images from Text-to-Image Models

Abhipsa Basu
abhypsabasu@iisc.ac.in
Vision and AI Lab
IISc Bangalore

R. Venkatesh Babu
venky@iisc.ac.in
Vision and AI Lab
IISc Bangalore

Danish Pruthi
danishp@iisc.ac.in
FLAIR Lab
IISc Bangalore

In this supplementary material, we provide country-wise statistics on participant ratings for the photo-realism of the generated images (discussed in Section 3 in the main paper). Additionally, we present further qualitative examples of images generated by the individual models.

A. Photo-realism of Generated Images

We discuss the effect of the photo-realism of the generated images on the human perceptions of geographical representativeness of the studied artifacts in subsection 3.4 in the main paper. Based on participant ratings on the realism of the images, we find that geographical representativeness and realism scores are correlated for both the models. Consistently, we observe that the participants provide an average self-evaluation score (rated by the participants to assess the impact of the photo-realism of the images on the scores they assigned for their geographical representativeness) of 3.5 on a scale of 1 to 5. Finally we find that that the average realism score assigned by participants is higher (averaged over all countries) when the prompt includes the country name, though, for some countries, e.g., the United States and Brazil, we see that the realism scores decrease upon including the country names in the prompt. We present the country-wise details on the realism values for the images generated by DALL·E 2 and Stable Diffusion in Table 1.

B. Qualitative Analysis

In Fig 3 in the main paper, we show examples of images depicting four nouns for both the unspecified and country-specific cases, for countries like Brazil, Mexico, *etc.* Here, we show images generated separately by both DALL·E 2 [1] and Stable Diffusion [2] for all the 10 nouns, whereas we choose one country from each continent: US, Chile, UK, Japan, South Africa, and Australia. The generated images are presented in Fig. 1 and 2 for DALL·E 2, and Fig. 3 and 4 for Stable Diffusion. For all the cases, we first show the images generated when the prompt does not specify the country name (denoted by unspecified), followed by those

generated with specific country names in the queries.

References

- [1] Aditya Ramesh, Prfulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 2022. 1
- [2] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, 2022. 1

Table 1. **Country wise photo-realism scores.** We present how the realism scores of images generated from DALL-E 2 (D2) and Stable Diffusion (SD) improve when the country name is specified in the text prompt. Additionally, in the last column we include the scores that users assign when asked about how the realism of images influenced their ratings about geographical representativeness.

| Countries | DALL-E 2 | | Stable Diffusion | | Self-Assessed |
|----------------|------------|-------------|------------------|-------------|---------------|
| | w/ country | Unspecified | w/ country | Unspecified | |
| United States | 3.82 ±0.37 | 4.00 ±0.36 | 3.88 ±0.40 | 4.12 ±0.32 | 3.88 ±0.46 |
| Canada | 4.06 ±0.27 | 3.88 ±0.49 | 3.06 ±0.44 | 3.19 ±0.40 | 3.31 ±0.68 |
| Mexico | 4.23 ±0.72 | 3.31 ±0.58 | 3.23 ±0.63 | 2.54 ±0.66 | 3.62 ±0.77 |
| Brazil | 4.07 ±0.34 | 4.27 ±0.39 | 3.40 ±0.45 | 3.27 ±0.34 | 3.53 ±0.58 |
| Chile | 4.37 ±0.33 | 4.16 ±0.47 | 3.05 ±0.45 | 2.26 ±0.54 | 3.58 ±0.45 |
| United Kingdom | 4.46 ±0.34 | 3.92 ±0.58 | 3.38 ±0.55 | 3.54 ±0.51 | 3.77 ±0.65 |
| Italy | 3.73 ±0.50 | 3.20 ±0.62 | 3.47 ±0.52 | 2.60 ±0.58 | 3.40 ±0.69 |
| Spain | 4.11 ±0.37 | 4.00 ±0.75 | 3.67 ±0.53 | 3.11 ±0.65 | 2.89 ±0.72 |
| Greece | 4.32 ±0.29 | 3.74 ±0.48 | 4.00 ±0.39 | 3.47 ±0.55 | 3.42 ±0.51 |
| Poland | 4.73 ±0.22 | 4.13 ±0.41 | 3.93 ±0.34 | 2.8 ±0.55 | 3.27 ±0.68 |
| Portugal | 4.35 ±0.40 | 3.88 ±0.56 | 4.11 ±0.46 | 3.29 ±0.51 | 3.47 ±0.77 |
| Belgium | 3.75 ±0.39 | 3.75 ±0.44 | 3.40 ±0.32 | 2.80 ±0.45 | 3.05 ±0.47 |
| Czech Republic | 3.67 ±0.38 | 3.44 ±0.56 | 3.11 ±0.43 | 2.72 ±0.48 | 3.56 ±0.52 |
| Hungary | 4.26 ±0.35 | 3.84 ±0.47 | 4.00 ±0.36 | 3.53 ±0.42 | 3.21 ±0.65 |
| Slovenia | 3.78 ±0.29 | 3.61 ±0.51 | 3.27 ±0.37 | 2.61 ±0.47 | 3.33 ±0.56 |
| Germany | 4.17 ±0.28 | 3.83 ±0.38 | 3.50 ±0.38 | 3.33 ±0.44 | 4.11 ±0.40 |
| Latvia | 4.56 ±0.28 | 3.33 ±0.58 | 2.72 ±0.40 | 2.39 ±0.47 | 3.67 ±0.58 |
| Estonia | 4.21 ±0.31 | 3.47 ±0.42 | 3.00 ±0.36 | 3.11 ±0.44 | 3.74 ±0.50 |
| Finland | 4.16 ±0.39 | 3.95 ±0.42 | 3.26 ±0.46 | 3.00 ±0.51 | 3.74 ±0.60 |
| France | 4.00 ±0.39 | 3.63 ±0.47 | 3.31 ±0.39 | 3.37 ±0.44 | 3.00 ±0.55 |
| India | 4.31 ±0.26 | 3.92 ±0.47 | 3.62 ±0.47 | 3.85 ±0.37 | 4.16 ±0.44 |
| Japan | 3.89 ±0.37 | 3.67 ±0.46 | 2.94 ±0.52 | 2.78 ±0.55 | 3.22 ±0.50 |
| South Korea | 4.32 ±0.29 | 3.42 ±0.47 | 3.21 ±0.47 | 3.05 ±0.61 | 3.89 ±0.38 |
| Israel | 4.54 ±0.34 | 4.15 ±0.36 | 2.92 ±0.58 | 3.31 ±0.69 | 2.62 ±0.75 |
| Australia | 3.93 ±0.43 | 3.80 ±0.62 | 3.93 ±0.54 | 3.47 ±0.52 | 3.67 ±0.71 |
| New Zealand | 3.75 ±0.44 | 3.40 ±0.56 | 2.85 ±0.38 | 2.65 ±0.49 | 3.35 ±0.52 |
| South Africa | 4.06 ±0.61 | 3.94 ±0.50 | 3.06 ±0.66 | 3.44 ±0.55 | 3.50 ±0.65 |
| Average | 4.14 ±0.08 | 3.73 ±0.10 | 3.38 ±0.10 | 3.09 ±0.11 | 3.50 ±0.12 |



Figure 1. Qualitative examples of images of five common nouns (City, Beach, House, Festival and Road) generated by DALL-E 2, for the default (unspecified) and the country specific prompts. From the shown examples, we note that the default generations often reflect artifacts from the US, whereas the representativeness is lower for Japan and Chile. The trend remains similar even after specifying the country name in the prompts, though the **GR** scores increase for all countries. (SA: South Africa, Aus: Australia)

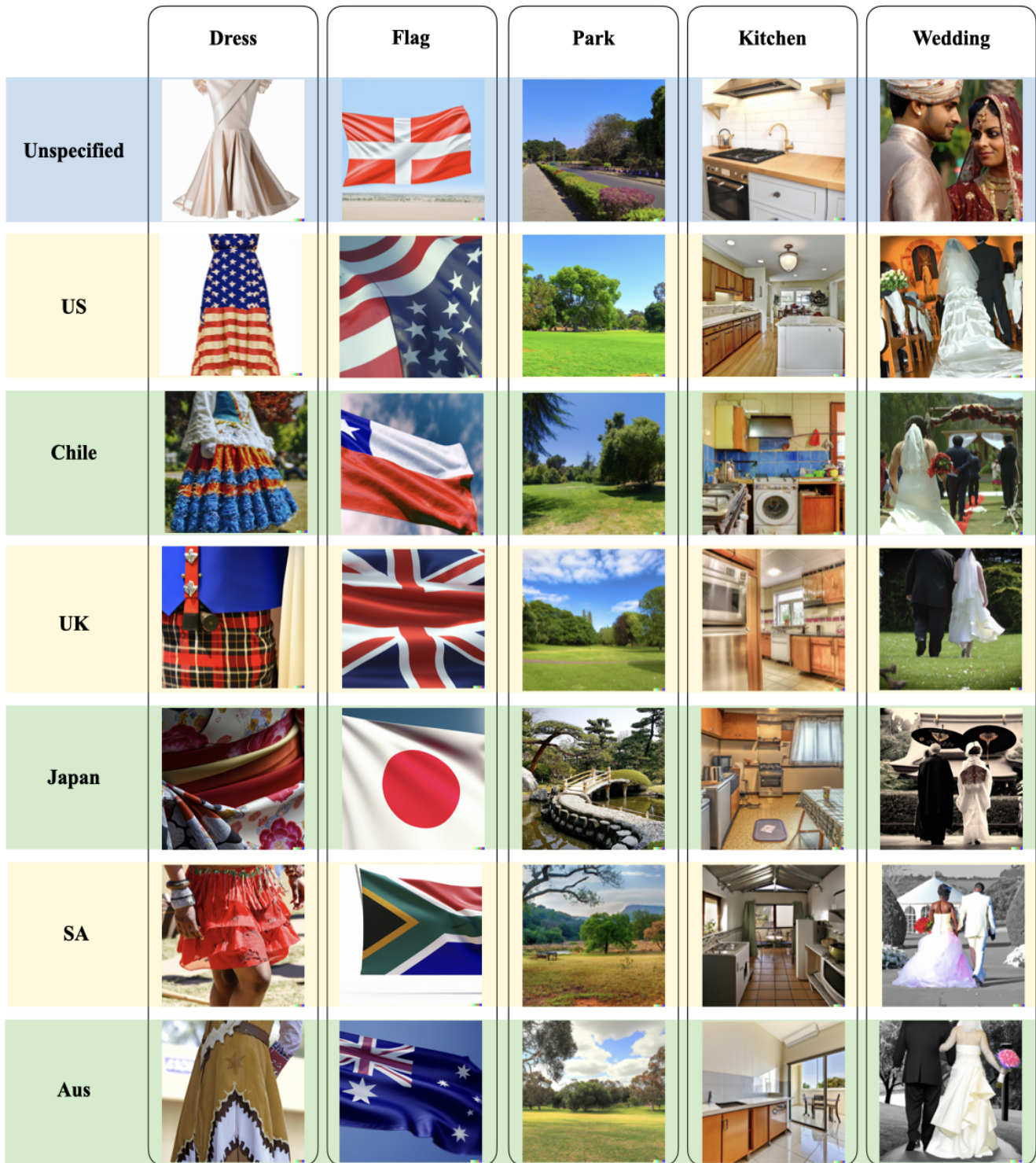


Figure 2. Qualitative examples of images of the five other common nouns (Dress, Flag, Park, Wedding and Kitchen) generated by DALL-E 2 for the default (unspecified) and the country specific prompts. From the shown examples, we note that the default generations often reflect artifacts from the US, whereas the representativeness is lower for Japan and Chile. The trend remains similar even after specifying the country name in the prompts, though the **GR** scores increase for all countries. (SA: South Africa, Aus: Australia)

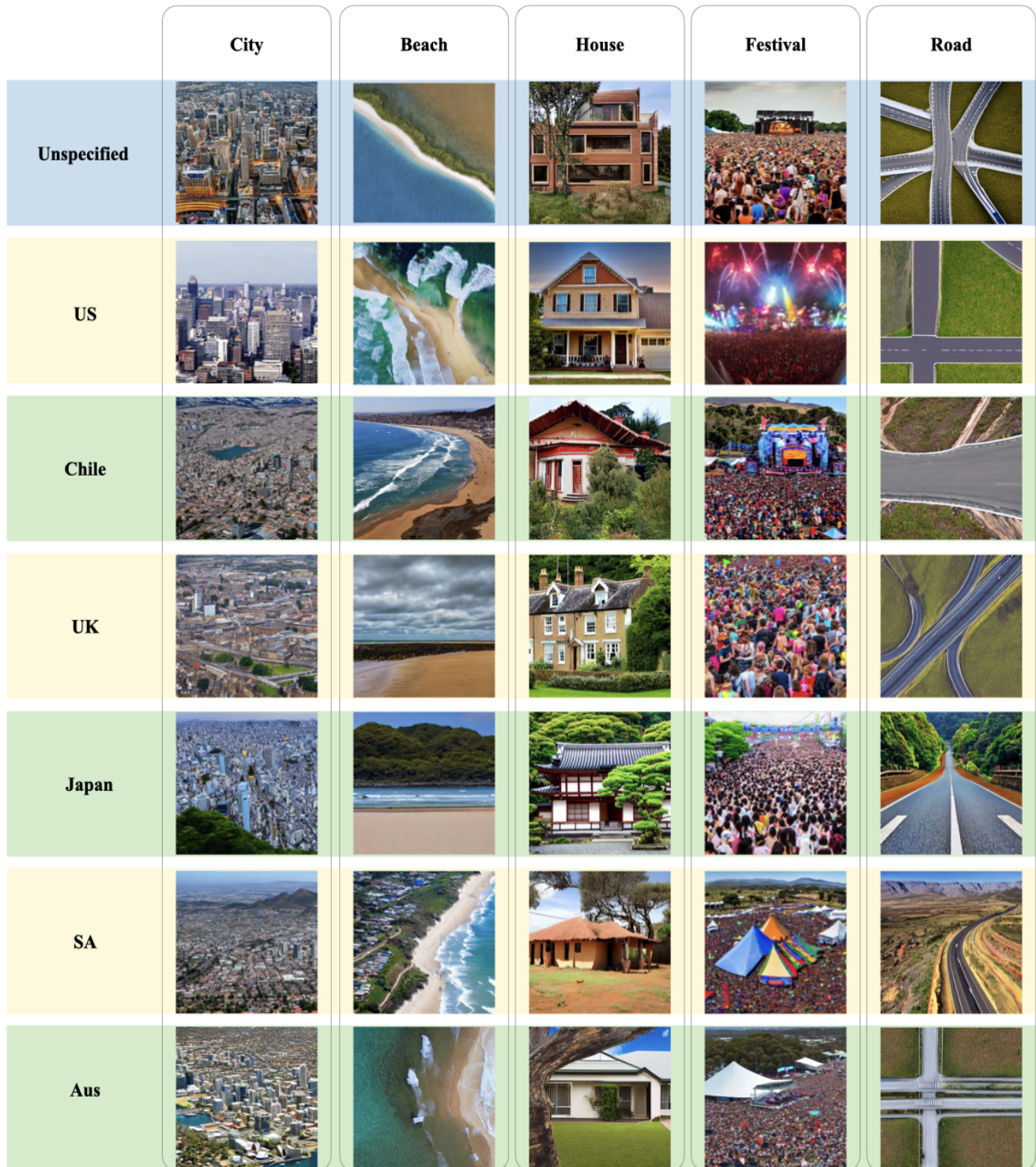


Figure 3. Qualitative examples of images of the five other common nouns (City, Beach, House, Festival, Road) generated by Stable Diffusion, for the default (unspecified) and the country specific prompts. Like the DALL·E 2 generated images, we note that the default generations often reflect artifacts from the US, whereas the representativeness is lower for Japan and Chile. The trend remains similar even after specifying the country name in the prompts, though the **GR** scores increase for all countries. (SA: South Africa, Aus: Australia)




































| | Dress | Flag | Park | Kitchen | Wedding |
|-------------|---|---|---|---|---|
| Unspecified |  |  |  |  |  |
| US |  |  |  |  |  |
| Chile |  |  |  |  |  |
| UK |  |  |  |  |  |
| Japan |  |  |  |  |  |
| SA |  |  |  |  |  |
| Aus |  |  |  |  |  |

Figure 4. Qualitative examples of images of the five other common nouns (Dress, Flag, Park, Kitchen, Wedding) generated by Stable Diffusion, for the default (unspecified) and the country specific prompts. Once again, we note that the default generations often reflect artifacts from the US, whereas the representativeness is lower for Japan and Chile. The trend remains similar even after specifying the country name in the prompts, though the **GR** scores increase for all countries. (SA: South Africa, Aus: Australia)