Supplementary Material: Appendix

Anjishnu Mukherjee, Ziwei Zhu, Antonios Anastasopoulos
Department of Computer Science
George Mason University, Fairfax, VA, USA
{amukher6, zzhu20, antonis}@gmu.edu

References

[1] Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. Quantifying the carbon emissions of machine learning, 2019. 5

[2] Maxim Tkachenko, Mikhail Malyuk, Andrey Holmanyuk, and Nikolai Liubimov. Label Studio: Data labeling software, 2020–2022. Open source software available from https://github.com/heartexlabs/label-studio. 6

1. Appendix

Table of Contents

1.1. Dataset details

Dollar Street concept classes (10) car, family snapshots, front door, home, kitchen, plate of food, cups/mugs/glasses, social drink, wall decoration, and wardrobe.

Dollar Street countries (63) South Africa, Serbia, Indonesia, Brazil, Kenya, India, Nigeria, France, Kazakhstan, United States, Philippines, Mexico, Sri Lanka, Netherlands, Thailand, Colombia, Pakistan, China, Russia, Egypt, Iran, United Kingdom, Romania, Spain, Turkey, Ukraine, Italy, Czech Republic, Denmark, Ethiopia, Jordan, Burundi, Burkina Faso, Malawi, Somalia, Zimbabwe, Haiti, Cote d'Ivoire, Myanmar, Papua New Guinea, Liberia, Cambodia, Bangladesh, Rwanda, Nepal, Palestine, Tunisia, Cameroon, Bolivia, Ghana, Vietnam, Guatemala, Mongolia, South Korea, Kyrgyzstan, Lebanon, Tanzania, Switzerland, Sweden, Canada, Peru, Austria and Togo.

Concept classes (10) for Dalle Street car, family snapshots, front door, home, kitchen, plate of food, cups/mugs/-glasses, social drink, wall decoration, and wardrobe.

Countries in our Dalle Street (67) Austria, Bangladesh, Bolivia, Brazil, Bulgaria, Burkina Faso, Burundi, Cambodia, Cameroon, Canada, China, Colombia, Cote d'Ivoire, Czech Republic, Denmark, Egypt, Ethiopia, France, Ghana, Greece, Guatemala, Haiti, India, Indonesia, Iran, Italy, Jordan, Kazakhstan, Kenya, Kyrgyzstan, Latvia, Lebanon, Liberia, Lithuania, Malawi, Mexico, Mongolia, Myanmar, Nepal, Netherlands, Nigeria, Pakistan, Palestine, Papua New Guinea, Peru, Philippines, Romania, Russia, Rwanda, Serbia, Somalia, South Africa, South Korea, Spain, Sri Lanka, Sweden, Switzerland, Tanzania, Thailand, Togo, Tunisia, Turkey, Ukraine, United Kingdom, United States, Vietnam, Zimbabwe

MaRVL Country to Language Mappings In the context of the MaRVL dataset, various languages are mapped to specific sub-regions based on the countries where these languages are predominantly spoken. The mapping is as follows:

- ``id'': The language code for Indonesian, which is primarily spoken in **Indonesia**, corresponds to the **South-eastern Asia** sub-region.
- ``sw'': The language code for Swahili, used in countries such as **Tanzania**, **Kenya**, and **Rwanda**, is mapped to the **Eastern Africa** sub-region.
- ``ta'': The language code for Tamil, spoken in **India** and **Sri Lanka**, is associated with the **Southern Asia** sub-region.
- ``tr'': The language code for Turkish, which is the official language of **Turkey**, falls under the **Western Asia** sub-region.
- ``zh'': The language code for Chinese, predominantly spoken in **China**, is linked to the **Eastern Asia** sub-region.

Table 1. Dataset Statistics

Sub-region	Eastern Africa	Eastern Asia	South-eastern Asia	Southern Asia	Western Asia	Caribbean	Central America	Central Asia	Eastern Europe	Melanesia
MaRVL	875	1107	1091	924	917	-	-	-	-	-
Dollar Street	310	313	578	839	128	56	12	20	136	14
DALL-E 3 Images	1052	438	840	742	600	160	176	280	741	147
Total	2237	1858	2509	2505	1645	216	188	300	877	161

Sub-region	Middle Africa	Northern Africa	Northern America	Northern Europe	South America	Southern Africa	Southern Europe	Western Africa	Western Europe	Total
MaRVL	-	-	-	-	-	-	-	-	-	4914
Dollar Street	107	81	317	51	447	60	223	262	183	4137
DALL-E 3 Images	303	289	465	736	605	139	740	888	594	9935
Total	410	370	782	787	1052	199	963	1150	777	18986

1.2. Prompt details

1.2.1 DALLE STREET generation

Prompt for data generation We use a simple template to prompt DALL-E 3 to generate images for a particular combination of country and category (Figure 1).

> **DALLE STREET Generation Prompt**
>
> A typical scene of **{category}** in **{country}**, culturally accurate and detailed.

Figure 1. We use a simple prompt that includes information about the concept class and the target country using a template, to generate our large scale dataset of DALL-E 3 images.

1.2.2 Cultural Awareness classifier

Prompt for classification of images We use a simple prompt to generate names of subregions from models when provided an image.

> **Classification Prompt**
>
> Strictly follow the United Nations geoscheme for subregions. Which geographical subregion of the United Nations geoscheme is this image from? Make an educated guess. Answer in one to three words.

Figure 2. We use a simple prompt to classify images in the data, by generating subregion labels.

1.2.3 Object Detection

Prompt used for Object Detection with GPT-4V We use a detailed prompt for GPT-4V to extract objects, colors and counts from images generated with DALL-E 3.

> **GPT-4V Object Detection Prompt**
>
> Give me a json output of the items you see in this image in both the foreground and background. Output the objects as a JSON with two fields: 'relevant_objects' for objects pertinent to the image category *concept* and 'other_objects' for all additional detected objects. Be as specific as possible. Within each field, for each detected object, include sub-fields describing object attributes like color, count, and anything else that is appropriate. For example, for buildings describe the architectural style in a sentence, for people describe clothing and headgear (if multiple colors and headgears are present, include the top three), for food items describe the exact type of food and include a brief recipe description, for pictures of rooms include objects in the background like mountains outside a window or paintings on the wall portraying something specific like a landmark or a particular type of scenery. For the counts of items, if the number of items is less than 10, give me exact numbers otherwise say more than 10.

Figure 3. We use a detailed prompt for GPT-4V to extract objects, colors and counts from images generated with DALL-E 3.

Prompt used for processing generated objects We use a detailed prompt for GPT-4 to process the dictionaries generated with the previous prompt into a simplified list along with some parsing rules to ensure correctness of the data structure.

GPT-4 Processing Object Detection Outputs

You will be provided a dictionary of items for the country **{country}** and concept **{concept}**. Summarize the dictionary into a list of comma-separated list of items with their respective colors. For example, *[red apple, blue car, green tree, house with a red roof and tinted glasses]*. Strictly follow the output format requested. The dictionary is as follows:

Figure 4. We use a detailed prompt for GPT-4 to process the dictionaries generated with the previous prompt into a simplified list along with some parsing rules to ensure correctness of the data structure.

1.3. LLM hyperparameters

We discuss the generation settings we used for our experiments, and also the associated costs and hardware.

1.3.1 Generation settings

- DALL-E 3 images are generated for `vivid` and `natural` settings for `standard` quality and size 1024×1024
- GPT-4 and GPT-4V generations are obtained for temperature $= 0.7$, top_p $= 0.95$, no frequency or presence penalty, no stopping condition other than the maximum number of tokens to generate, max_tokens $= 300$.
- LLaVA generations are obtained for temperature $= 1.0$ and top_p $= 1.0$, no penalties, and max_tokens $= 128$. The reason for using a slightly higher temperature and top_p is to have more consistent outputs. In our initial experiments, LLaVA did not perform as well in terms of following instructions at the same temperatue setting as GPT-4V.
- For Grounding DINO, we use `ShilongLiu/GroundingDINO` from Hugging Face and set both box and text thresholds to 0.25 for the grounded box generations.
- For Stable Diffusion, we use `stabilityai/stable-diffusion-2-inpainting` from Hugging Face, and replace the autoencoder with `stabilityai/sd-vae-ft-mse`. We also use a `DPMSolverMultistepScheduler` for speeding up the generation process. We add ``` ``intricate details. 4k. high resolution. high quality.'' ``` to the end of our prompt to get high quality images.

1.3.2 Computation budget

- We spent about \$800 in total for DALL-E 3 generations. This was funded by a grant from Microsoft Azure.
- We spent about \$700 in total for GPT-4V `vision-preview` and GPT-4 `turbo` inference and across all experiments.
- For experiments with LLaVA, Stable Diffusion and Grounding DINO, we used a single instance of a Multi-Instance A100 GPU with 40GB of GPU memory, 3/7 fraction of Streaming Multiprocessors, 2 NVIDIA Decoder hardware units, 4/8 L2 cache size, and 1 node.
- Total emissions for API based models are estimated to be 25 kgCO_2 eq, of which 100 percents were directly offset by the cloud provider. Total emissions for our on-premise GPU usage is estimated to be less than 10 kgCO_2 eq. Estimations are conducted using the MachineLearning Impact calculator [1].

1.4. Human Study - the Annotators

Annotator Demographics All annotators have different demographic backgrounds (but are physically located in the USA currently) and are between 25-40 years old. Together, they are native to or have resided in more than countries and all 4 major regions from our dataset, and thus represent a strong sample of opinions. Roughly 40% identify as female, and the rest as male. In terms of an education background, 40% of the annotators are graduate students, whereas the rest includes working professionals from different backgrounds and also computer science faculty. In total, we have 14 annotators, recruited from different computer science labs at an university and also from a diverse set of social connections for this study. All the annotators have agreed to consent for using this data for research purposes. Our study qualifies for exemption from IRB as no PII is involved.

How many studies do we have

1. Human study to verify the quality of generated images
2. Human study to understand performance on our benchmark
3. Human validation for checking if artifacts are hallucinated
4. Human study to check if artifacts are culturally relevant
5. Human study for verifying people count associations
6. Human study for layout preservation after editing using CultureAdapt
7. Human study for cultural relevance after editing using CultureAdapt

Total number of annotations To validate our dataset and human baseline on cultural awareness, we have 14 people, each annotating 300 samples, resulting in 4200 annotations for one study and 8400 in total. For each of the other three studies (excluding the studies for edits), we have three people annotating 100 samples each for a study, resulting in 900 annotations in total. Finally, for the two editing-related studies, we have three people each looking at three images per sample (source image, edited image using our method, edited image using another method) for 100 samples, resulting in 900 annotations per study or 1800 in total. Combining all this, we have $8200 + 900 + 1800 = 10900$ annotation items to back the comprehensive quantitative evaluations we perform for each study in our paper.

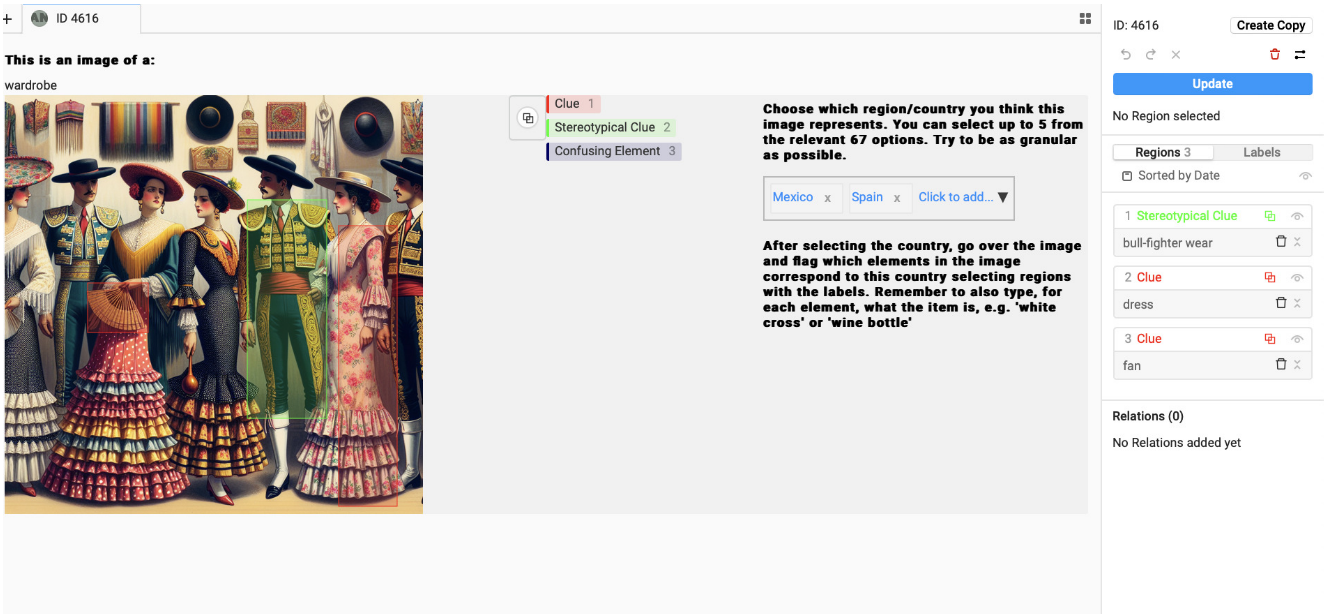
1.5. Human Study - the Interfaces

General Instructions for Study 1, 2 We use https://labelstud.io [2] to perform our human study. For each annotator, we create 2 tabs corresponding to the 2 studies, and ask them to solve them in numerical order to avoid getting influenced from seeing true labels first from the second study. Time taken to complete the first study is usually 2 hours, and the time taken to complete the second study is typically 30 minutes. All annotators will be compensated for their time with a $20 gift card upon completion of the task.

1.5.1 Study 1: verify the quality of generated images

Task 1 Instructions For every image, you have to make atleast 1 guess for the geographical region label, along with atleast 1 corresponding clue. If you are not sure what the clue is, add a question mark symbol at the end of it - example, headgear? bread?). Do not reverse image search or look anything up. Answer only using your knowledge or instinct. You can try guessing sub-region/region if you are not sure about country. You can use the knowledge of the fact that the image is generated by an AI conditioned on the provided prompt above the image. Note that you do not have to be correct!

Once the label is done, you need to add at least one bounding box somewhere in the image (it can be very specific and small or very broad or even the entire image) and then label that bounding box as either a clue or a stereotypical clue or a confusing element and then add a text description for the bounding box from the interface on the right (for example, a bounding box for a basket of baguettes can be a clue for France and the text description may be either something specific like "baguette" or something generic like "bread?"). The difference between stereotypical clue and regular clue is that stereotypical would be something like baguettes for France or "naan" for India or specific clothing styles for some country whereas a regular clue is



This is an image of a:

wardrobe

Choose which region/country you think this image represents. You can select up to 5 from the relevant 67 options. Try to be as granular as possible.

Mexico x Spain x

After selecting the country, go over the image and flag which elements in the image correspond to this country selecting regions with the labels. Remember to also type, for each element, what the item is, e.g. 'white cross' or 'wine bottle'

No Region selected

1 Stereotypical Clue — bull-fighter wear

2 Clue — dress

3 Clue — fan

No Relations added yet

Figure 5. Annotation Interface for Study 1

something that you are using to make your guess but you don't know enough about your guess to know what stereotypical clues might be associated with it, for example, sand for island countries.

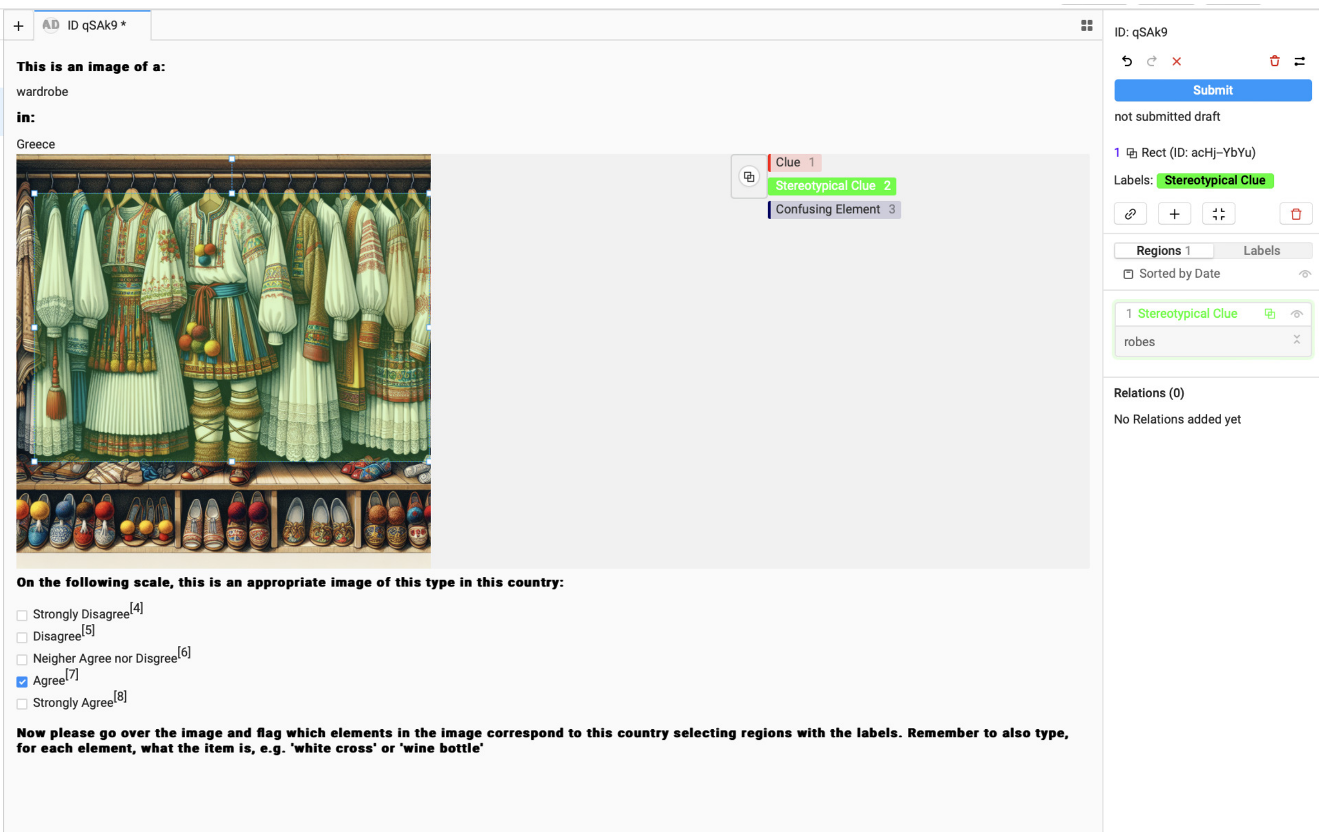
1.5.2 Study 2: human performance on our cultural awareness benchmark

Task 2 Instructions For every image, select a single rating for "appropriateness" = "this image is one of the possible (stereo)typical representation of the mentioned category for the mentioned country". Then select at least one clue corresponding to your rating, similar to the first study.

Performance of Human Study participants on the Cultural Awareness Task We include anonymized performance statistics of each of our annotators to show differences in performance at the individual data point level for countries, subregions and continents.

User	Country Level	Subregion Level	Continent Level
User 1	46.53	70.14	90.97
User 2	16.67	31.94	78.47
User 3	11.19	31.47	70.63
User 4	32.81	50.78	72.66
User 5	21.13	51.41	75.35
User 6	10.87	32.61	71.01
User 7	28.67	64.34	84.62
User 8	22.14	43.57	69.29
User 9	7.09	34.04	70.92
User 10	26.43	62.86	83.57
User 11	20.71	50.71	87.14
User 12	4.86	18.05	75.69
User 13	3.57	17.85	72.14
User 14	6.47	37.41	75.53

Table 2. User accuracies across country, subregion and continent levels, rounded to two decimal places. At the country level, accuracy varies from 46% to about 4%, so the subregion level accuracies are a more reliable indicator of performance even for humans. Continent is the most generic label, and has very high accuracies from all participants.



This is an image of a:

wardrobe

in:

Greece

On the following scale, this is an appropriate image of this type in this country:

□ Strongly Disagree[4]
□ Disagree[5]
□ Neigher Agree nor Disgree[6]
☑ Agree[7]
□ Strongly Agree[8]

Now please go over the image and flag which elements in the image correspond to this country selecting regions with the labels. Remember to also type, for each element, what the item is, e.g. 'white cross' or 'wine bottle'

Figure 6. Annotation Interface for Study 2

1.5.3 Study 3, 4: artifact hallucination and cultural relevance

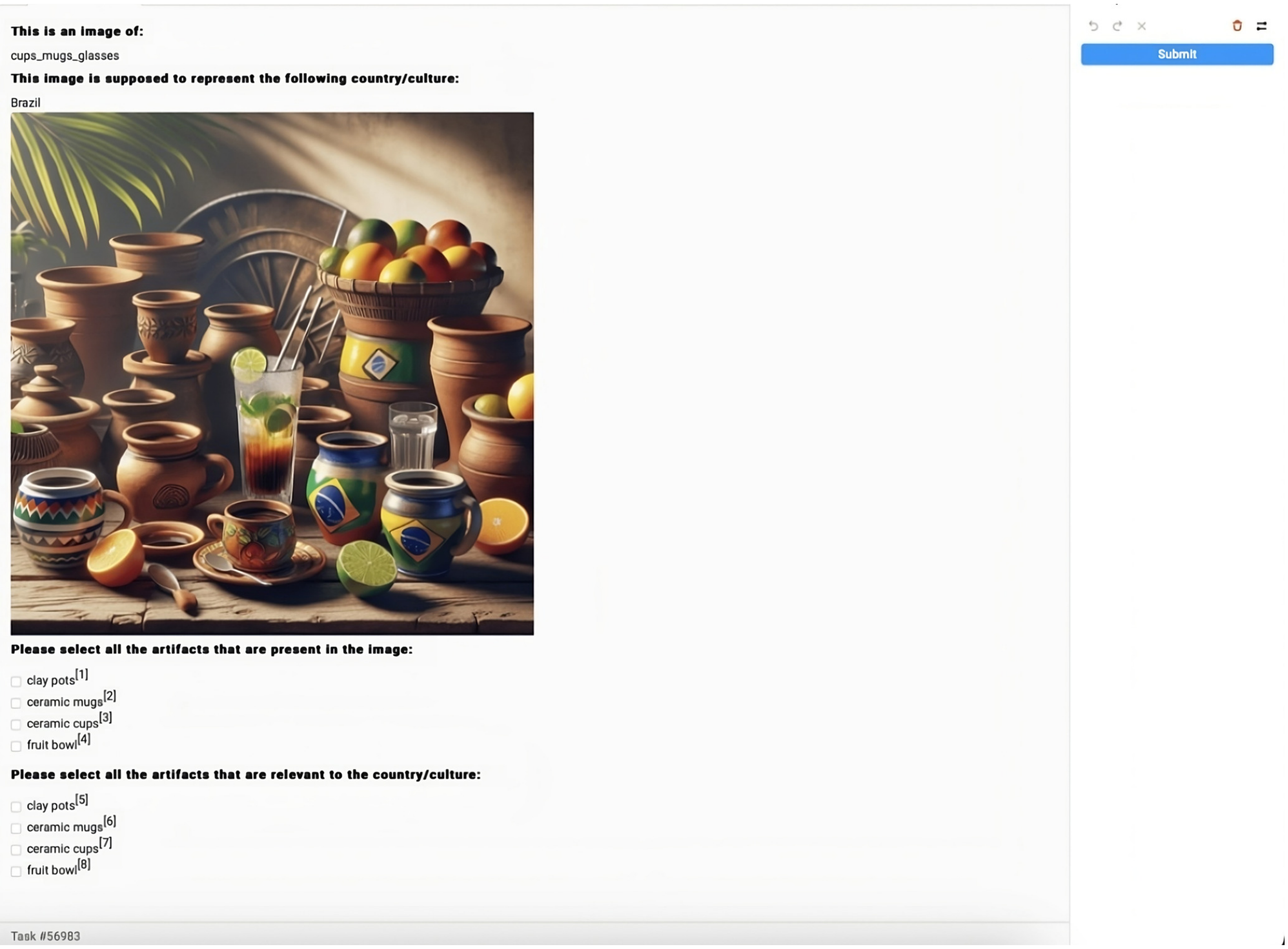
We provide an interface (Figure 7) that includes a multi-choice correct question answering setting, where we first ask a question about the artifacts that are present in the image and then ask if those artifacts are culturally relevant. For selecting these options, we sample randomly from all artifacts extracted for the accompanying image. For this study, we also provide the name of the concept class that the image is supposed to represent and the country, so that annotators have an idea about the object categories that they might be looking for. To account for cases where annotators might not be sure about any object and still choose one, in our analysis we only consider those cases where atleast two artifacts are marked as present. Responses to the second question is somewhat subjective as it depends on the cultural backgrounds of the annotators, but as we see in our analysis, the responses mostly agree and find more than half of the artifacts to be relevant.

1.5.4 Study 5: people count associations

We used an identical interface to the one for Study 3,4 to carry out this study, but instead of providing multiple correct options, we provide a direct True/False setting for annotators to answer if the count of people shown in the given image is approximately correct, i.e. falls in the correct count bucket (1-5, 5-10, more than 10). Our analysis reveals that people mostly agree with the count buckets generated, but on corroborating patterns with actual population statistics, the ordering of countries look slightly different, so we do not make any correlations about that in our analysis.

1.5.5 Study 6, 7: layout preservation and cultural relevance after editing

We use an interface (Figure 8) that shows the name of the source and target countries, the source image, and two edited images - one from our method CULTUREADAPT and one from `cap-edit`. We ask a somewhat subjective question regarding which image the annotator prefers, but responses vary widely for this question and do not correlate well across annotators, so we do not use these results for further analysis. For each of the edited images, we ask three questions - (1) if the edit maintains structural layout, (2) if the edit makes the image more culturally relevant to the target country and (3) if the edited image



This is an image of:

cups_mugs_glasses

This image is supposed to represent the following country/culture:

Brazil

Please select all the artifacts that are present in the image:

- □ clay pots[1]
- □ ceramic mugs[2]
- □ ceramic cups[3]
- □ fruit bowl[4]

Please select all the artifacts that are relevant to the country/culture:

- □ clay pots[5]
- □ ceramic mugs[6]
- □ ceramic cups[7]
- □ fruit bowl[8]

Task #56983

Figure 7. Annotation Interface for Study 3,4

might be toxic for people from target country. All of these questions require a response on a Likert scale, which we design using a system of stars, where users can select anywhere between 1 and 5 stars. Less than 1% of images are marked toxic across all annotations and annotators agree strongly on this metric. We report the analysis for the other two metrics in the main paper.

1.6. Additional Results

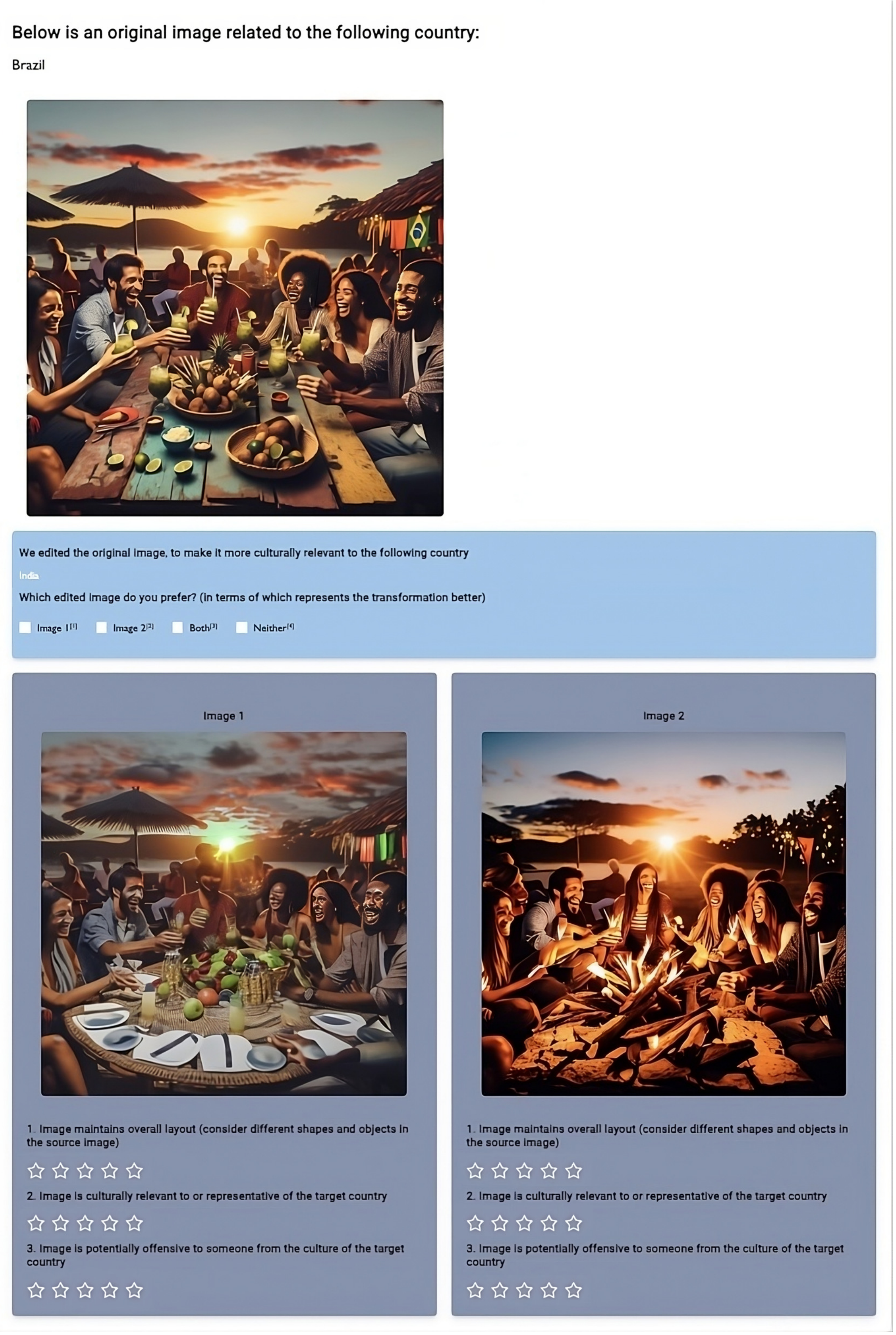
1.6.1 Task 1 - Cultural awareness

Income distribution performance for LLaVA We show the performance of LLaVA on the income quartile distribution basis for DOLLAR STREET.

Confusion matrix set 1 We show confusion matrices for both LLaVA and GPT-4V on DALLE STREET images.

Confusion matrix set 2 We show confusion matrices for both LLaVA and GPT-4V on DOLLAR STREET images.

Confusion matrix set 3 We show confusion matrices for both LLaVA and GPT-4V on MARVL images.



Below is an original image related to the following country:

Brazil

We edited the original image, to make it more culturally relevant to the following country

India

Which edited image do you prefer? (in terms of which represents the transformation better)

□ Image 1[1] □ Image 2[2] □ Both[3] □ Neither[4]

Image 1

1. Image maintains overall layout (consider different shapes and objects in the source image)

☆ ☆ ☆ ☆ ☆

2. Image is culturally relevant to or representative of the target country

☆ ☆ ☆ ☆ ☆

3. Image is potentially offensive to someone from the culture of the target country

☆ ☆ ☆ ☆ ☆

Image 2

1. Image maintains overall layout (consider different shapes and objects in the source image)

☆ ☆ ☆ ☆ ☆

2. Image is culturally relevant to or representative of the target country

☆ ☆ ☆ ☆ ☆

3. Image is potentially offensive to someone from the culture of the target country

☆ ☆ ☆ ☆ ☆

Figure 8. Annotation Interface for Study 6,7

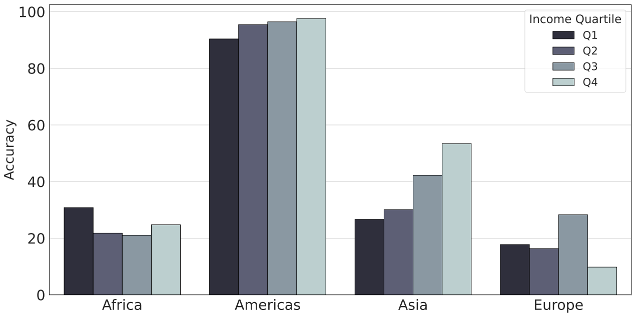


Figure 9. We normalize income data from DOLLAR STREET into region specific quartiles and plot corresponding accuracies for LLaVA.

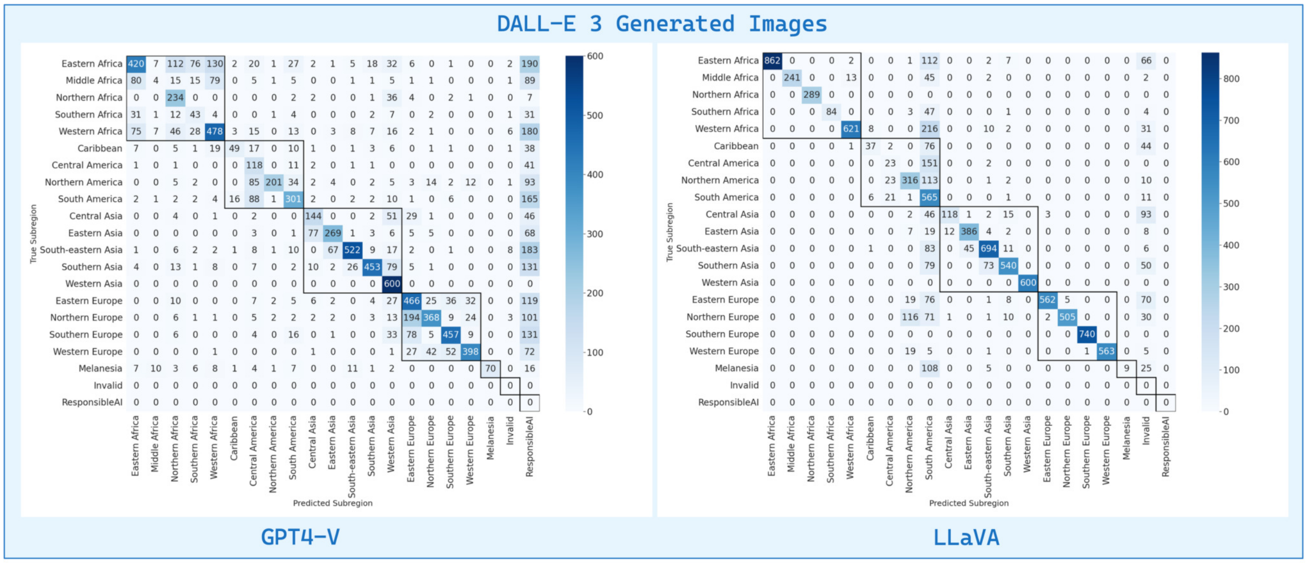


Figure 10. Confusion matrices for GPT-4V and LLaVA on the cultural awareness task for DALLE STREET images.

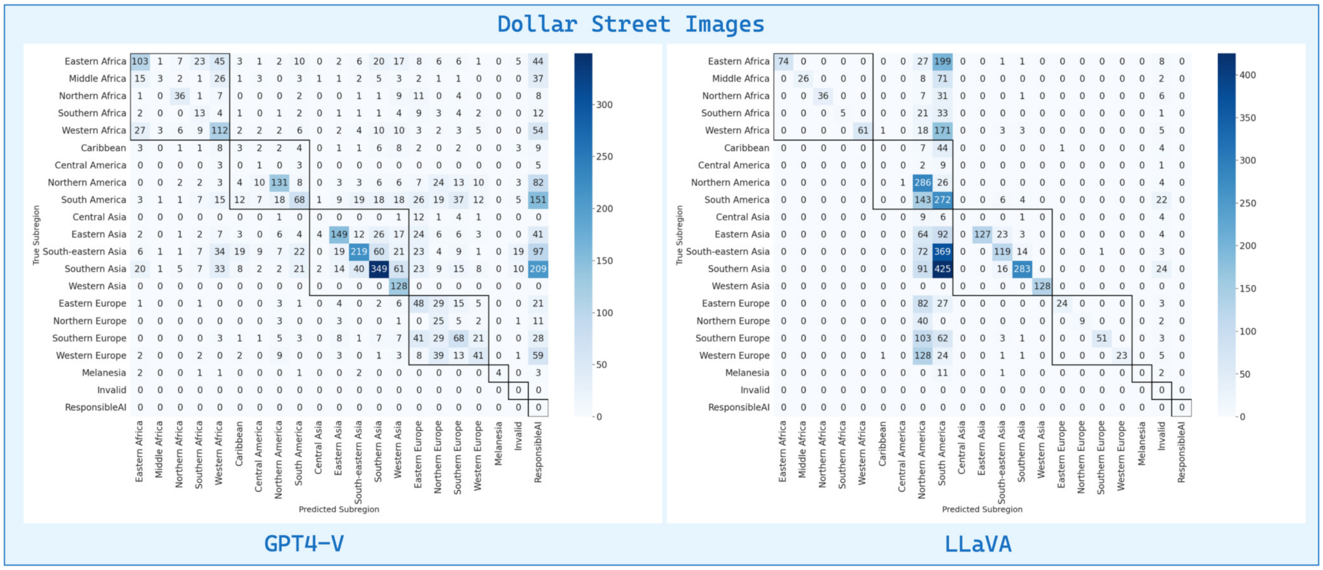


Figure 11. Confusion matrices for GPT-4V and LLaVA on the cultural awareness task for DOLLAR STREET images.

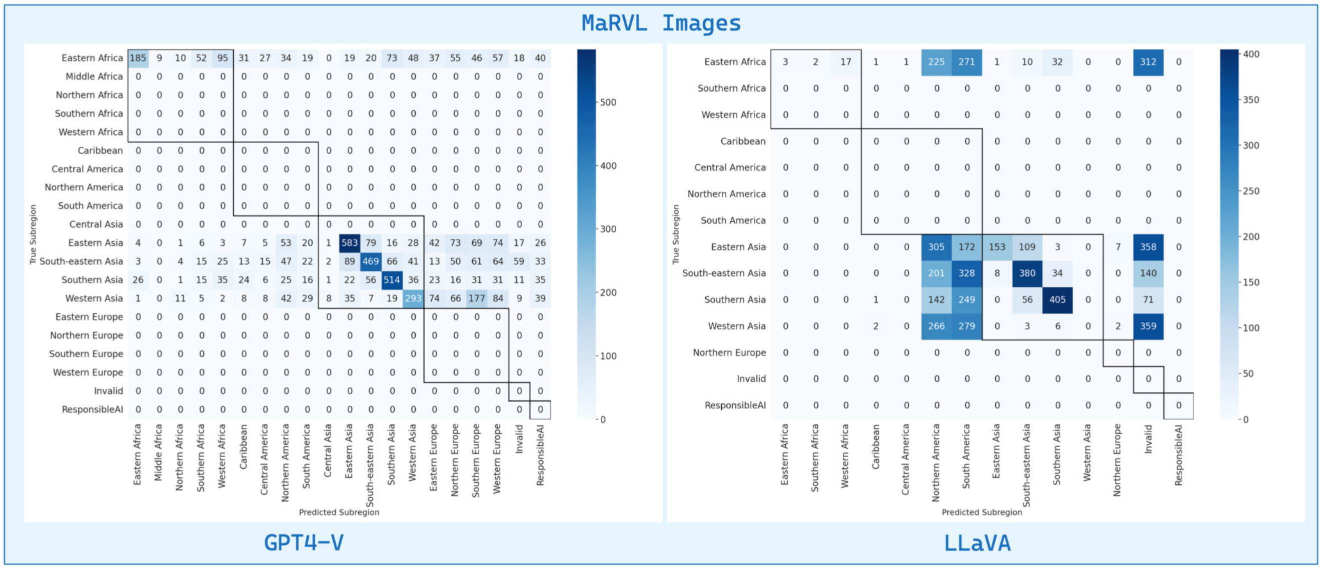


Figure 12. Confusion matrices for GPT-4V and LLaVA on the cultural awareness task for MaRVL images.

1.6.2 Task 2 - Artifacts

How many artifacts did we identify using GPT-4V Table 3 shows the counts across each of the 67 countries for the number of artifacts identified. adj implies unique artifacts are a combination of words and adjectives that appear before it to quantify count or color. no_adj implies only the raw words identified. Figure **??** includes a distribution of the TD-IDF scores for these (country, artifact) pairs and high scores that lie outside the range as given by the mean and the standard deviation of the distribution (i.e larger than 3.01 or smaller than 0.47) would imply strongly correlated artifacts for a given country, and there exists 4019 such items from this data. Further human filtering can be done to remove common words like table or mailbox or dresses to find unique and interesting associations like pretzels in Austria, zinnias in Bolivia and many more, some of which we report in Table 4.

	Austria	**Bangladesh**	**Bolivia**	**Brazil**	**Bulgaria**	**Burkina Faso**	**Burundi**	**Cambodia**	**Cameroon**	**Canada**
adj	248	283	279	264	264	288	292	276	251	292
no_adj	154	148	141	157	138	153	141	161	133	170
	China	**Colombia**	**Cote d'Ivoire**	**Czech Republic**	**Denmark**	**Egypt**	**Ethiopia**	**France**	**Ghana**	**Greece**
adj	275	268	270	276	278	276	252	278	265	270
no_adj	124	134	146	160	168	152	139	157	137	151
	Guatemala	**Haiti**	**India**	**Indonesia**	**Iran**	**Italy**	**Jordan**	**Kazakhstan**	**Kenya**	**Kyrgyzstan**
adj	272	301	264	312	246	262	280	257	270	269
no_adj	163	160	147	172	123	135	158	139	149	147
	Latvia	**Lebanon**	**Liberia**	**Lithuania**	**Malawi**	**Mexico**	**Mongolia**	**Myanmar**	**Nepal**	**Netherlands**
adj	269	244	276	267	281	266	282	292	290	260
no_adj	156	133	163	159	152	133	156	155	156	146
	Nigeria	**Pakistan**	**Palestine**	**Papua New Guinea**	**Peru**	**Philippines**	**Romania**	**Russia**	**Rwanda**	**Serbia**
adj	279	254	270	279	265	276	272	244	288	250
no_adj	152	144	137	141	137	152	149	144	161	143
	Somalia	**South Africa**	**South Korea**	**Spain**	**Sri Lanka**	**Sweden**	**Switzerland**	**Tanzania**	**Thailand**	**Togo**
adj	291	272	279	263	250	260	265	272	273	279
no_adj	165	165	144	149	150	157	155	145	154	140
	Tunisia	**Turkey**	**Ukraine**	**United Kingdom**	**United States**	**Vietnam**	**Zimbabwe**	**Total**		
adj	249	254	267	281	273	291	311	18212		
no_adj	133	132	148	181	162	163	166	10035		

Table 3. Salient Artifact Statistics (**adj** indicates descriptors like color, etc are part of the artifact name, whereas **no_adj** indicates the artifact name does not have such descriptors)

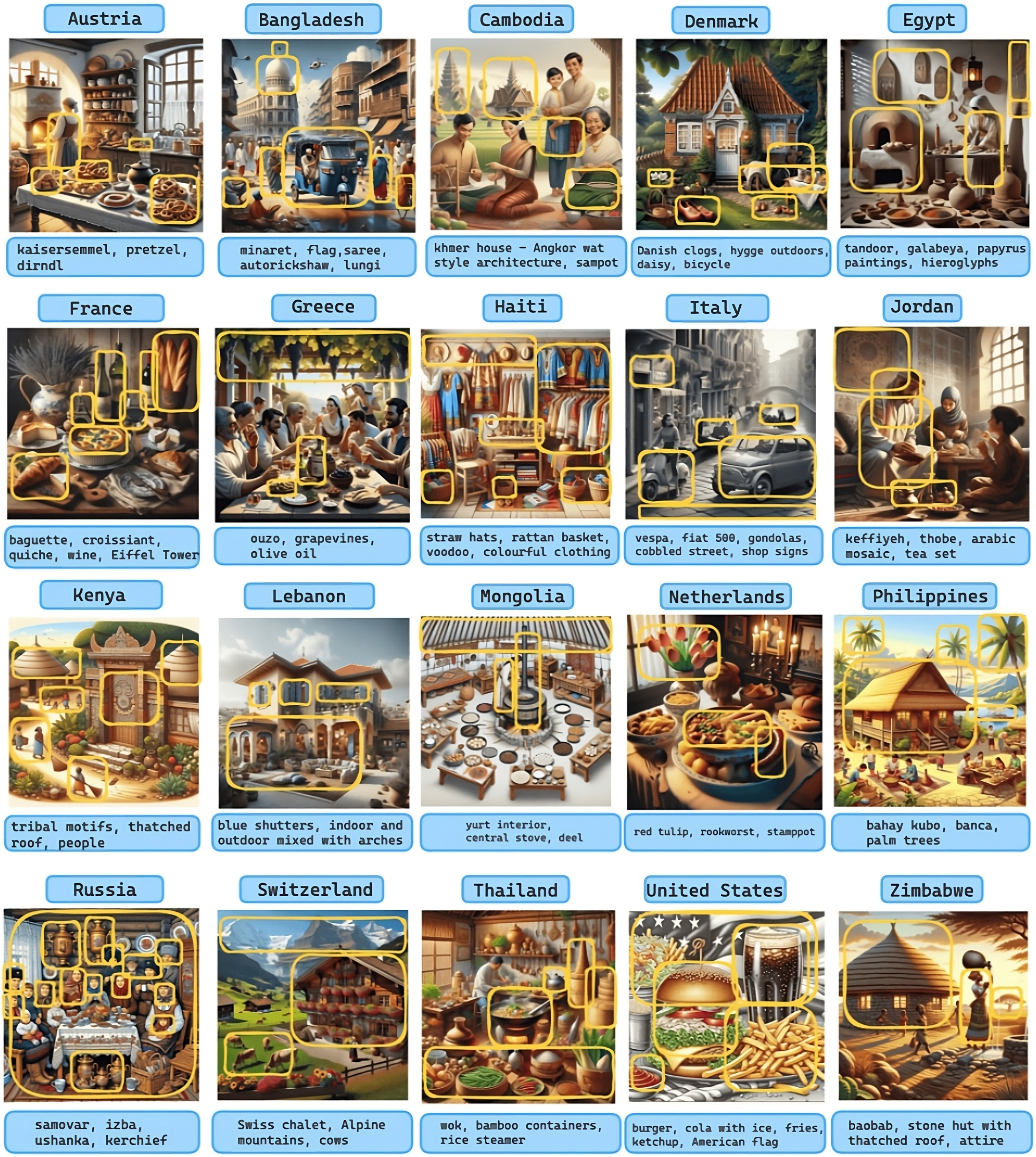

Austria
kaisersemmel, pretzel, dirndl
Bangladesh
minaret, flag,saree, autorickshaw, lungi
Cambodia
khmer house - Angkor wat style architecture, sampot
Denmark
Danish clogs, hygge outdoors, daisy, bicycle
Egypt
tandoor, galabeya, papyrus paintings, hieroglyphs
France
baguette, croissiant, quiche, wine, Eiffel Tower
Greece
ouzo, grapevines, olive oil
Haiti
straw hats, rattan basket, voodoo, colourful clothing
Italy
vespa, fiat 500, gondolas, cobbled street, shop signs
Jordan
keffiyeh, thobe, arabic mosaic, tea set
Kenya
tribal motifs, thatched roof, people
Lebanon
blue shutters, indoor and outdoor mixed with arches
Mongolia
yurt interior, central stove, deel
Netherlands
red tulip, rookworst, stamppot
Philippines
bahay kubo, banca, palm trees
Russia
samovar, izba, ushanka, kerchief
Switzerland
Swiss chalet, Alpine mountains, cows
Thailand
wok, bamboo containers, rice steamer
United States
burger, cola with ice, fries, ketchup, American flag
Zimbabwe
baobab, stone hut with thatched roof, attire

Figure 13. We identify more than 18,000 unique cultural artifacts across all countries as part of our second task, and then filter them to find salient ones. This figure shows strongest correlated artifacts for 20 randomly picked countries.
Note that these associations are extracted from LMM generations and may not always be accurate.

Color associations on DALL-E 3 images We find that countries are more likely to be associated with particular colors, with some showing prominently strong associations.

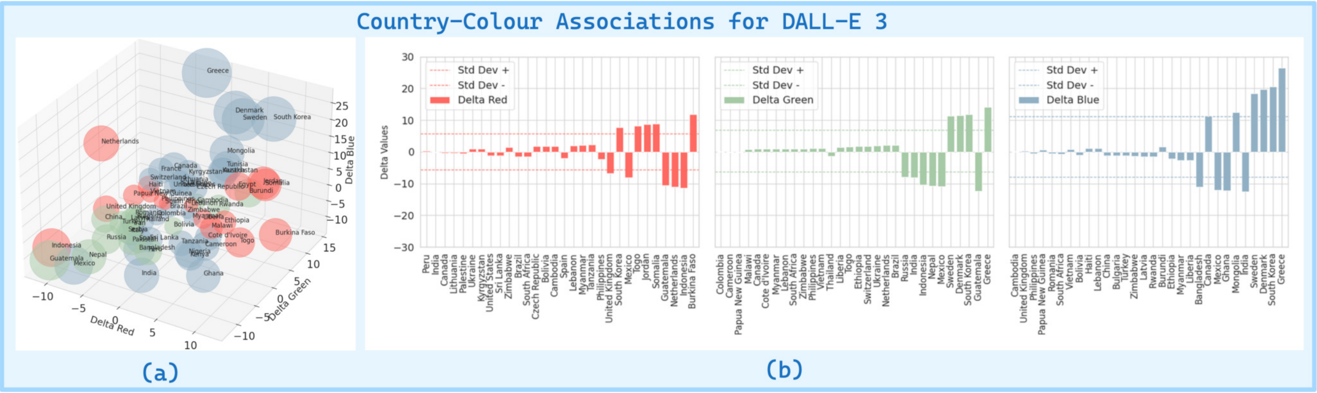


Figure 14. We explore how countries are distributed on a color spectrum by first calculating a global average RGB vector for DALLE STREET images and then defining deltas along each axes aggregated at the country level. **Takeaway**: We find interesting associations - Greece is strongly correlated with blue, Burkina Faso with red.

People-count associations on DALL-E 3 images Distributions validated by humans do not always correlate with actual population statistics.

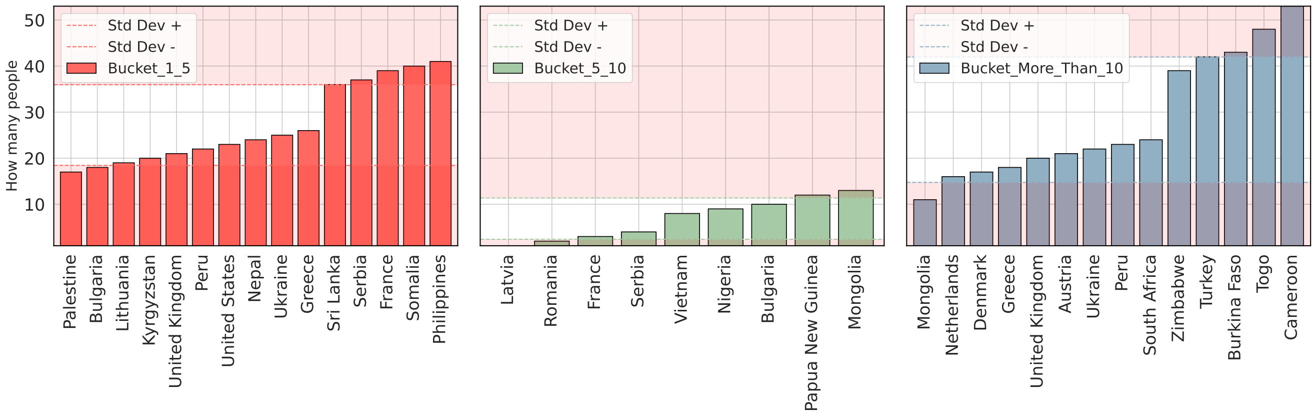


Figure 15. Here, we look at buckets of people counts in DALLE STREET images aggregated at the country level, each of the subplots representing one bucket. **Takeaway**: Counts of people in images may not always accurately reflect population densities of the corresponding countries to scale.

Interesting associations We show examples of interesting associations identified by models and humans at the country level, for our artifact extraction task.

Table 4. Interesting associations and their explanations for various countries.
Note that these associations are extracted from LMM generations and may not always be accurate.

Country	Interesting Associations and Explanations
Austria	**Dirndl**: A traditional dress worn in Austria and parts of Germany. **Pretzel**: A type of baked bread product, often associated with German-speaking countries. **Lederhosen**: Traditional leather shorts worn by men in the Alpine regions.
Bangladesh	**Lungi**: A traditional garment worn by men, usually a wraparound skirt. **Kurti**: A traditional garment worn by women, often paired with leggings or a skirt.

Continued on next page

Country	Interesting Associations and Explanations
	Harmonium: A musical instrument commonly used in South Asian music.
Bolivia	**Zinnias**: A type of flower native to the region, known for its bright colors and significance in local celebrations. **Llama**: A domesticated South American camelid, significant in Bolivian culture. **Chullos**: Knitted hats, typically with ear flaps, that are traditional to the Andes.
Brazil	**Bikini**: Associated with the famous beaches of Brazil. **Lychee**: A tropical fruit found in Brazil. **Samba**: A Brazilian music genre and dance style.
Bulgaria	**Spanakopita**: A savory pastry filled with spinach and feta cheese. **Moussaka**: A layered dish with eggplant, potatoes, and minced meat. **Terracotta**: Refers to clay-based unglazed or glazed ceramic.
Cameroon	**Kaftans**: A type of long robe worn in many African countries. **Fufu**: A dough-like food made from cassava or yams. **Savanna**: A type of ecosystem common in Cameroon, characterized by grassland with scattered trees.
Canada	**Poutine**: A dish consisting of fries topped with cheese curds and gravy. **Moose**: A large mammal found in Canada. **Snowmobile**: A vehicle designed for travel on snow, common in Canadian winters.
China	**Changshan**: A traditional Chinese garment for men. **Baozi**: A type of Chinese steamed bun with fillings. **Lion Dance**: A traditional dance in Chinese culture performed during the Lunar New Year and other cultural events.
Ethiopia	**Injera**: A sourdough flatbread and a staple food in Ethiopia. **Wat**: A traditional Ethiopian stew. **Shawl**: Often worn by Ethiopian women as part of traditional attire.
France	**Camembert**: A famous French cheese. **Baguette**: A long, thin loaf of French bread. **Beret**: A soft, round, flat-crowned hat associated with French culture.
Germany	**Oktoberfest**: An annual beer festival and cultural event in Munich. **Bratwurst**: A type of German sausage. **Dirndl**: Traditional dress worn by women during Oktoberfest and other occasions.
Greece	**Toga**: A garment worn in ancient Greece. **Dolma**: A dish made of grape leaves stuffed with rice or meat. **Moussaka**: A layered dish with eggplant, meat, and béchamel sauce.
India	**Sari**: A traditional garment worn by women. **Lassi**: A yogurt-based drink. **Rangoli**: A form of art created on the floor using colored rice, sand, or flower petals.
Japan	**Kimono**: A traditional Japanese garment. **Sushi**: A popular Japanese dish. **Tatami**: A type of mat used as a flooring material in traditional Japanese rooms.
Mexico	**Sombrero**: A wide-brimmed hat traditionally worn in Mexico. **Tacos**: A traditional Mexican dish. **Guacamole**: A Mexican avocado-based dip or spread.
Morocco	**Tagine**: A North African dish named after the earthenware pot in which it is cooked.

Continued on next page

Country	Interesting Associations and Explanations
	Kaftan: A long robe worn in Morocco. **Mint Tea**: A popular beverage in Morocco, often served as a welcoming gesture.
Nepal	**Topi**: A traditional hat worn in Nepal. **Himalayas**: The mountain range running across Nepal. **Dal Bhat**: A traditional Nepalese dish consisting of lentils and rice.
Peru	**Chullo**: A traditional hat with earflaps. **Llama**: A significant animal in Peruvian culture. **Ponchos**: Traditional clothing made from wool.
Thailand	**Tuk-tuk**: A common form of transportation in Thailand. **Pad Thai**: A popular Thai noodle dish. **Elephant**: An animal deeply ingrained in Thai culture and symbolism.
Togo	**Kente Cloth**: A traditional fabric made of silk and cotton, known for its vibrant colors and patterns. **Yam Festival**: A major cultural festival celebrating the harvest of yams. **Agbadza Dance**: A traditional dance performed during festivals and ceremonies.
Tunisia	**Shisha**: A popular water pipe used for smoking flavored tobacco. **Harissa**: A spicy chili paste that is a staple in Tunisian cuisine. **Mosaic Art**: Intricate and colorful tile art that is significant in Tunisian culture.
Turkey	**Evil Eye**: A common talisman believed to protect against negative energy. **Baklava**: A sweet pastry made of layers of filo filled with nuts and honey. **Whirling Dervishes**: A religious dance performed by Sufi practitioners.
Ukraine	**Pysanky**: Traditional Ukrainian Easter eggs decorated with intricate designs. **Borscht**: A beet soup that is a key part of Ukrainian cuisine. **Vyshyvanka**: Traditional Ukrainian embroidered shirts.
United Kingdom	**Afternoon Tea**: A British tradition involving tea and a variety of snacks. **Red Telephone Box**: Iconic public telephone booths found throughout the UK. **Fish and Chips**: A classic British dish of battered fish and fried potatoes.
United States	**Route 66**: A historic highway symbolizing the American road trip. **Thanksgiving**: A national holiday celebrating the harvest and other blessings. **Statue of Liberty**: A symbol of freedom and democracy in the US.
Vietnam	**Ao Dai**: A traditional Vietnamese dress for women. **Pho**: A Vietnamese noodle soup that is a staple dish. **Conical Hat (Non La)**: A traditional hat made of bamboo and palm leaves.
Zimbabwe	**Mbira**: A traditional musical instrument also known as the thumb piano. **Great Zimbabwe**: The ruins of an ancient city, significant in Zimbabwean history. **Victoria Falls**: One of the largest and most famous waterfalls in the world, located on the border between Zimbabwe and Zambia.

Table 5. Cultural artifacts for various countries based on human annotations.
Note that these artifacts are based on subjective perceptions of our human annotators and may not be completely accurate always.

Country	Cultural Artifacts
Austria	beer, sausage, dirndl
Bangladesh	rice, saree, fish
Bolivia	colorful clothes, poncho, hats

Continued on next page

Country	Cultural Artifacts
Brazil	brazilian flag, tropical fruit, colorful pottery
Bulgaria	clothing, rugs, door
Burkina Faso	dry, black people, straw basket
Burundi	rice, beans, bananas
Cambodia	buddhism, buddhist art, clothing
Cameroon	african people, bananas, beans
Canada	maple leaf, canadian flag, poutine
China	characters, chinese food, lanterns
Colombia	coffee, rice, avocado
Cote d'Ivoire	black people, dry, african outfit
Czech Republic	beer, dress, czech
Denmark	danish flag, beer, windmill
Egypt	hieroglyphs, egyptian art, islamic clothing
Ethiopia	coffee, colors, clay pots
France	baguette, cheese, wine
Ghana	black people, african necklaces, clothing
Greece	blue and white, sea, olives
Guatemala	mayan art, tortilla, beans
Haiti	black people, rice, beans
India	naan, curry, sari
Indonesia	buddhism, rice, clothing
Iran	islamic art, kebab, persian rug
Italy	pizza, pasta, wine
Jordan	clothing, arabic, islamic art
Kazakhstan	clothing, houses, islamic art
Kenya	african people, african art, corn
Kyrgyzstan	clothing, islamic art, rugs
Latvia	clothing, beer, bread
Lebanon	arabic clothing, hummus, bread
Liberia	rice, black people, palm trees
Lithuania	clothing, food, beer
Malawi	hut, corn, black people
Mexico	sombrero, tequila, tortilla
Mongolia	yurt, dumplings, clothing
Myanmar	buddhist art, rice, pagoda
Nepal	buddhist elements, hindu elements, rice
Netherlands	windmill, cheese, dutch clothing
Nigeria	rice, yams, african clothing
Pakistan	clothing, curry, sombrero
Palestine	arabic art, hummus, bread
Papua New Guinea	black people, tropical fruit, coconut
Peru	inca clothing, machu picchu, andes mountains
Philippines	rice, tropical vegetation, cooking
Romania	clothing, sheep, ceramic pots
Russia	fur hat, warm clothes, vodka
Rwanda	african art, beans, dark-skinned people
Serbia	clothing, beer, sausages
Somalia	islamic art, banana, rice
South Africa	african art, corn, hat
South Korea	korean characters, kimchi, korean dress
Spain	flamenco, paella, bull fighting

Continued on next page

Country	Cultural Artifacts
Sri Lanka	buddhist art, buddhist symbols, spicy food
Sweden	northern european clothing, fish, snowy landscape
Switzerland	alps, swiss cheese, chocolate
Tanzania	african art, rice, meat
Thailand	buddhist art, thai food, clothing
Togo	african clothing, cloth patterns, wood carvings
Tunisia	arabic art, couscous, arched doorways
Turkey	turkish coffee, rugs, kebabs
Ukraine	clothing patterns, flower designs, ukrainian food
United Kingdom	pubs, fish and chips, tea
United States	american flag, burgers, jeans
Vietnam	conical hats, pho, pagodas
Zimbabwe	thatched huts, african clothing, animal carvings

1.6.3 Task 3 - Edits

Source-Target Pair	**Similarity `SSIM`**		**Metric M_1**		**Metric M_2**	
	cap-edit	**CULTUREADAPT**	**cap-edit**	**CULTUREADAPT**	**cap-edit**	**CULTUREADAPT**
Brazil-India	0.96	0.93	0.72	0.68	0.95	0.95
Brazil-Nigeria	0.96	0.93	0.62	0.47	0.89	0.91
Brazil-Turkey	0.96	0.93	0.69	0.52	0.95	0.94
Brazil-USA	0.96	0.93	0.28	0.35	0.91	0.83
Average	*0.96*	*0.93*	*0.58*	*0.51*	*0.93*	*0.91*
India-Brazil	0.97	0.94	0.63	0.60	0.96	0.90
India-Nigeria	0.96	0.94	0.69	0.68	0.94	0.90
India-Turkey	0.97	0.94	0.56	0.57	0.92	0.89
India-USA	0.96	0.94	0.67	0.63	0.93	0.88
Average	*0.97*	*0.94*	*0.63*	*0.62*	*0.94*	*0.89*
Nigeria-Brazil	0.96	0.92	0.29	0.41	0.76	0.85
Nigeria-India	0.96	0.92	0.62	0.67	0.87	0.93
Nigeria-Turkey	0.96	0.91	0.66	0.63	0.91	0.93
Nigeria-USA	0.96	0.92	0.50	0.60	0.90	0.91
Average	*0.96*	*0.92*	*0.51*	*0.58*	*0.86*	*0.90*
Turkey-Brazil	0.97	0.94	0.46	0.41	0.89	0.84
Turkey-India	0.97	0.95	0.62	0.59	0.88	0.88
Turkey-Nigeria	0.97	0.94	0.67	0.64	0.93	0.90
Turkey-USA	0.96	0.94	0.29	0.43	0.88	0.89
Average	*0.97*	*0.94*	*0.51*	*0.51*	*0.89*	*0.88*
USA-Brazil	0.97	0.94	0.53	0.46	0.88	0.89
USA-India	0.98	0.94	0.18	0.62	0.58	0.91
USA-Nigeria	0.98	0.94	0.20	0.46	0.61	0.84
USA-Turkey	0.98	0.94	0.21	0.46	0.64	0.87
Average	*0.97*	*0.94*	*0.28*	*0.50*	*0.68*	*0.88*
Overall Average	**0.97**	**0.94**	**0.50**	**0.54**	**0.85**	**0.89**

Table 6. Mean Similarity (`SSIM`) Scores, Metric M_1, and Metric M_2 for `cap-edit` and CULTUREADAPT grouped by Source and Target Country

CULTUREADAPT quantitative metrics

More examples of edits using CULTUREADAPT We include more examples of edits across different concept classes and source-target pairs using our CULTUREADAPT pipeline in Figure 16. As can be seen, the pipeline is only constrained by the two bottlenecks of object detection and diffusion based inpainting, which sometimes may detect objects incorrectly or not generate consistent images of human faces for example.

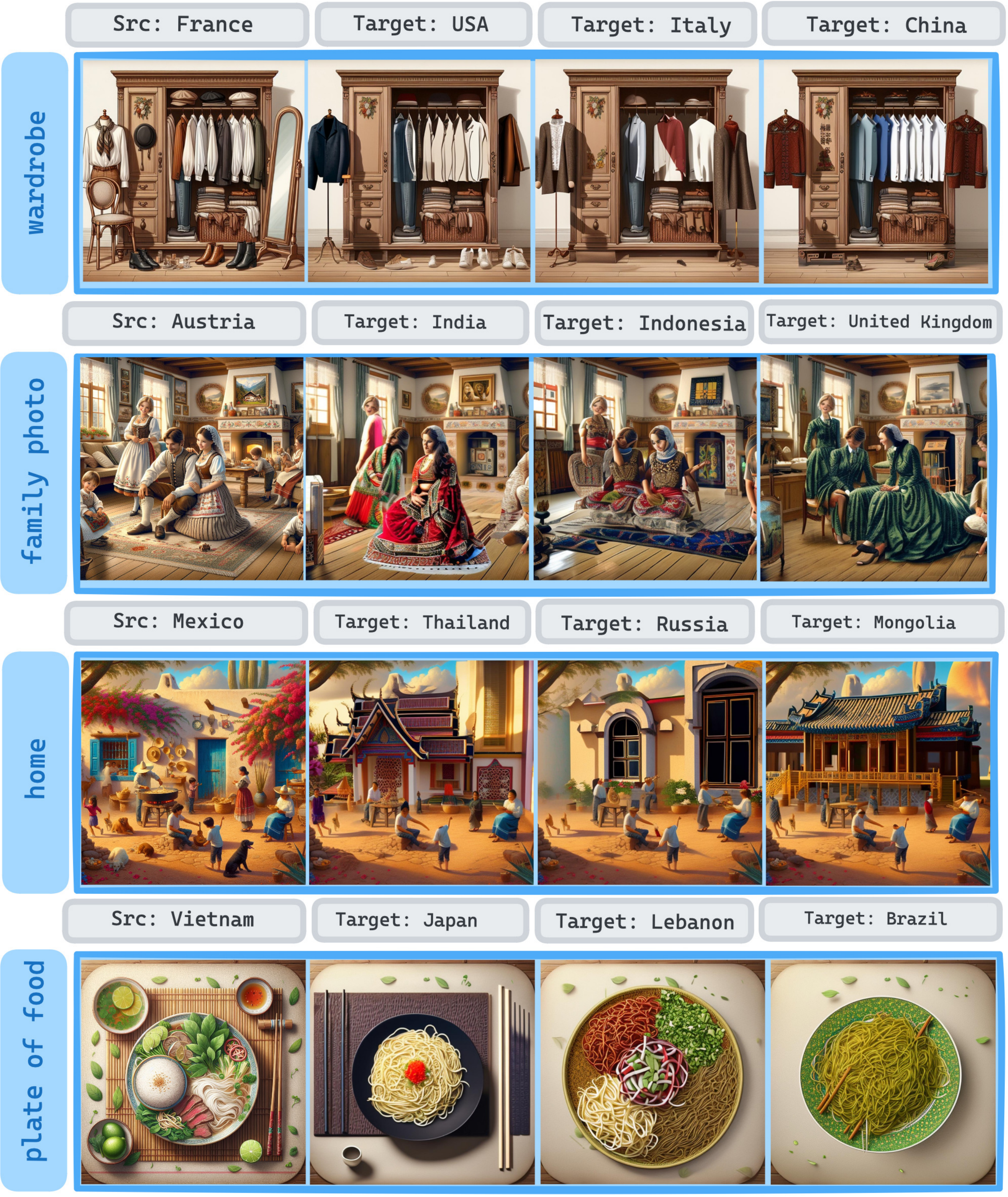


Figure 16. We show examples of edits made using our CULTUREADAPT pipeline across 4 different concept classes and 12 pairs of unique source, target combinations to illustrate both cases where our pipeline excels and also where it is limited by the parts it is composed of. For all of these edits, our metric success criteria of $\Delta_1 < 0$ and $\Delta_2 > 0$ is satisfied.