A. Datasheet

A.1. Motivation

Why was the dataset created?

The dataset was created to facilitate future academic Computer Vision research about human aesthetic preference.

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

The dataset was created by researchers at MMLab, The Chinese University of Hong Kong.

A.2. 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, ratings; people, interactions between them; nodes, edges)

The instances are prompts and generated images, along with human preference choices among the images generated by the same prompt.

Are relationships between instances made explicit in the data (*e.g.* social network links, user/movie ratings, etc.)?

Yes, instances generated by the same user are identified by the same user id, which is anonymized for privacy.

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

There are 25,205 instances in the dataset.

What data does each instance consist of? "Raw" data (*e.g.* unprocessed text or images) or Features/attributes? Is there a label/target associated with instances? If the instances related to people, are sub-populations identified (*e.g.* by age, gender, etc.) and what is their distribution?

Each instance consists of $n \in 2, 3, 4$ image, one prompt and one human choice.

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.

Yes, we omit the specific parameters for generating the images, such as diffusion steps and guidance scale. They are omitted because we are more interested in the users' preference about the generated images, rather than how they

are created. Also, since the same batch of images (among which users make comparisons) are always generated with the same set of parameters except the random seed, they are irrelevant variables when studying human preferences.

Is everything included or does the data rely on external resources?

The dataset is self-contained.

Are there recommended data splits and evaluation measures? (*e.g.* training, development, testing; accuracy or AUC)

In our experiments, we use a training set of 20,205 instances and validation set of 5,000 images, which will be made public. We recommend using accuracy (%) with one decimal place.

Are there any errors, sources of noise, or redundancies in the dataset?

Yes. The users are not prompted to selected images fitting their preference, so there should be noise in the collected data.

Is the dataset self-contained, or does it link to or otherwise rely on external resources (*e.g.* websites, tweets, other datasets)?

The dataset is self-contained.

Does the dataset contain data that might be considered confidential (*e.g.* data that is protected by legal privilege or by doctorpatient confidentiality, data that includes the content of individuals non-public communications)?

No, the dataset is collected from the Stable Foundation Discord server, which is publicly available for any user with an account.

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

We collect images and their prompts from the Stable Foundation discord server. Even though the discord server has rules against users sharing any NSFW (not suitable for work, such as sexual and violent content) and illegal images, our dataset still contains some NSFW images and prompts that were not removed by the server moderators.

Does the dataset relate to people?

Yes, the prompts are written by users and the choices are made by users.

Does the dataset identify any subpopulations (*e.g.* by age, gender)?

No.

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?

No.

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)?

The dataset may contain sensitive data, because the prompts written by users may contain sensitive information, such as public figures and religious beliefs.

What experiments were initially run on this dataset? Have a summary of those results.

It has been used to validate the correlation between human preference and several popular image quality evaluation metrics, and serve as the training data for a human preference classifier. The results show that the tested metrics do not correlate well with human preference, and the correlation of the ViT-L/14 version of CLIP can be improved via fine-tuning on the dataset.

A.3. Data Collection Process

How was the data associated with each instance acquired?

The prompts, images and human choices are directly observable from the Stable Foundation Discord server.

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

Automatic scraping procedures were used to collect the data.

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

The dataset is not a sample of a larger set.

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

The authors of this paper were solely involved in the data collection process.

Over what time-frame was the data collected?

The dataset covers the chat history of dreambot channels between Dec. 2nd 2022 and Jan. 18th 2023.

Were any ethical review processes conducted (*e.g.* by an institutional review board)?

No official processes were conducted, due to the public nature of the data on Discord channel.

Does the dataset relate to people?

No.

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

The data was obtained from public messages in the Discord server.

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?

No analysis has been conducted.

A.4. Data Preprocessing

What preprocessing/cleaning was done? (*e.g.* discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instancess, processing of missing values)?

No preprocessing is done on the images and prompts.

A.5. Uses

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

As described in the paper, this dataset has been used for analysis about several image quality evaluation metrics and training the proposed human preference classifier.

Is there a repository that links to any or all papers or systems that use the dataset?

No.

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

It can be used for tasks related to human preference on generated images.

Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?

Yes. As discussed in Sec. 8, the dataset is biased towards the preference of the certain group of people that are active in the Stable Foundation Discord server.

Are there tasks for which the dataset should not be used?

No.

A.6. Data 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

Yes. Researchers at academic institutions will be able to request access to the dataset.

How will the dataset be distributed? (*e.g.* tarball on website, API, GitHub; does the data have a DOI and is it archived redundantly?)

We will provide download links for researchers on a GitHub repository.

When will the dataset be distributed?

Before April 15, 2023.

Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?

We will provide a terms of use agreement with the dataset. The dataset as a whole will be distributed under a non-commercial 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.

Unknown.

A.7. Dataset Maintenance

Who is supporting/hosting/maintaining the dataset?

The authors of this paper are maintainers of this dataset.

How can the owner/curator/manager of the dataset be contacted (*e.g.* email address)?

By email: wuxiaoshi@link.cuhk.edu.hk .

Is there an erratum?

At this time, we are not aware of errors in our dataset. However, we will create an erratum as errors are identified.

Will the dataset be updated? If so, how often and by whom? How will updates be communicated? (*e.g.* mailing list, GitHub)

The dataset will be updated by the authors on an at-will basis (but no more than once a month).

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.

No such limits are established.

Will older versions of the dataset continue to be supported/hosted/maintained?

N/A

If others want to extend/augment/build on this dataset, is there a mechanism for them to do so? If so, is there a process for tracking/assessing the quality of those contributions. What is the process for communicating/distributing these contributions to users?

There will not be a mechanism to build on top of the dataset.

B. More Dataset Examples
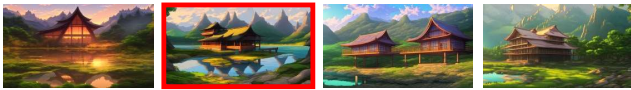
See Fig. 9 for more examples.

C. More Visualization

See Fig. 10 and Fig. 11 for more visualizations. We show that the adapted model generates images with less artifacts and are better aware of users' intentions.

maritime mistressin light ice armor on horse, highly detailed, digital painting, intense, sharp focus, photorealistic, detailed water, photorealistic oil painting full body, highly detailed, digital painting, intense, sharp focus

a beautiful modern wooden house near a lake in the middle of the mountains at sunrise, anime style, studio ghibli, genshin impact, digital art ...

a sculpture of a skeleton made of clocks and gears, cyberpunk art, by andrei kolkoutine, reddit, kinetic art, laser cut, with a skeleton army, lights, brown exoskeleton

batman mixed with spiderman, realistic, haze, ultra - detailed, film photography, light leaks, larry bud melman, trending on artstation, sharp focus ...

3 d octane render of a realistic hand with rough texture, intricate detail, high quality, studio lighting, anatomically correct, proper proportions, no deformity, perfect

car design sketch art station, octane render 8 k

tom clancy's the division, new york city, ruined city, night time, clouds, snow, intense cinematic shot :: a bald man from behind, with a red hat, wearing black streetwear ...

dog eating burrito

portrait of a dog queen wearing a gold crown and a queen dress
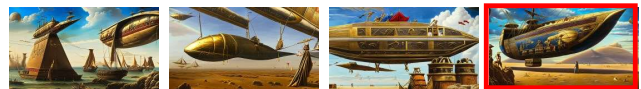
art by patrick woodroffe egyptian air ship sailing through the sky ultra hd 4 k ultra high detail

Figure 9. More examples of the collected data. The images are generated by Stable Diffusion with the prompts shown below each group of images. The preferred images are highlighted with red borders.

Original Model Regularization Only Adapted Model

a corgi sitting on the porch of a cabin enjoying the snow with the cabin in the
forest near a lake surrounded by mountains, hypermaximalist, elegant ...

draw of a person using a watering can to watering plants
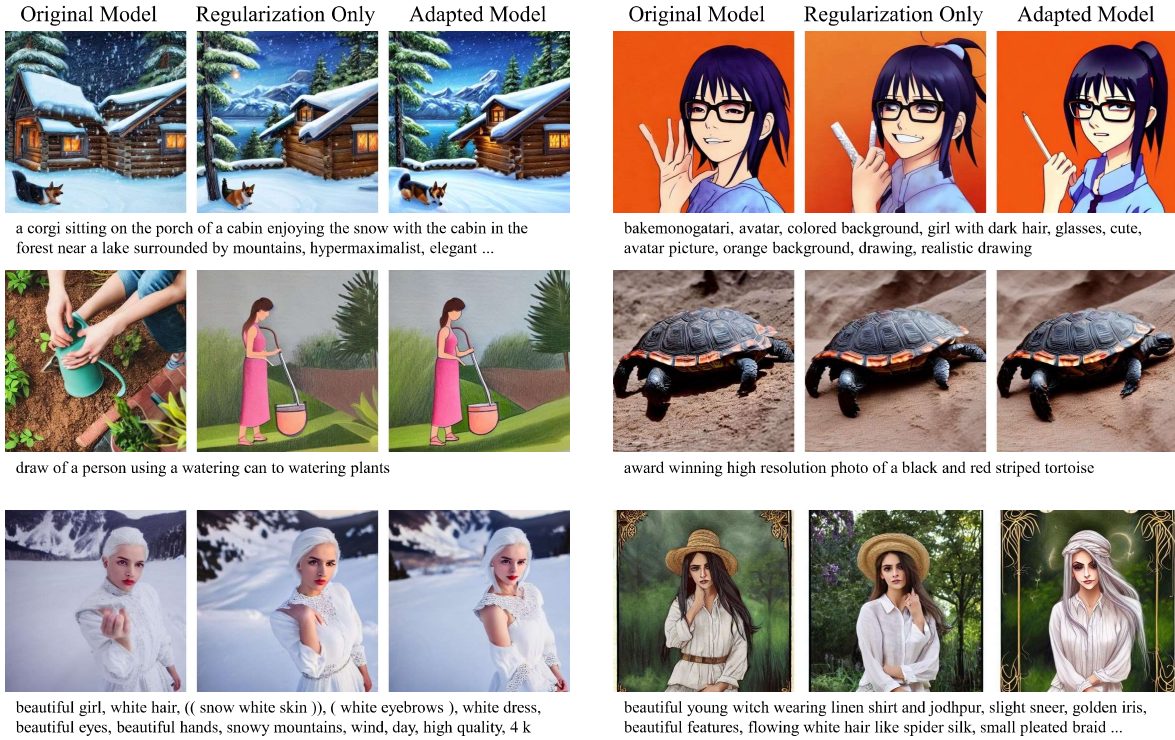
beautiful girl, white hair, ((snow white skin)), (white eyebrows), white dress,
beautiful eyes, beautiful hands, snowy mountains, wind, day, high quality, 4 k

Original Model Regularization Only Adapted Model

bakemonogatari, avatar, colored background, girl with dark hair, glasses, cute,
avatar picture, orange background, drawing, realistic drawing

award winning high resolution photo of a black and red striped tortoise

beautiful young witch wearing linen shirt and jodhpur, slight sneer, golden iris,
beautiful features, flowing white hair like spider silk, small pleated braid ...

Figure 10. The adapted model generates images with less artifacts. Images in the same group are generated with the same prompt and random seed.

Original Model Regularization Only Adapted Model

an otherworldly creature made of crystals, magnificent, majestic, highly intricate,
realistic photography, incredibly detailed, ultra high resolution, 8 k ...

scary glowing ghost pirateship with skeletons crewmembers in a tempest under a
skull shaped moon by dan mumford and aleksi briclot, impactful rimlighting ...

creative font, letter b, sharp cyberpunk details, hooded panda in shadows, sharp
focus, elegant, high detail, illustration, pixiv, digital art

Original Model Regularization Only Adapted Model

full body humanoid cyborg art by marco plouffe, by cedric seaut, studio lighting,
artstation, high resolution, overdetailed, 4 k, hd, optimal render ...

monkey gamer wearing red - pink hoodie and a cap, aesthetic, stunning, highly
detailed, smooth, soft focus highly detailed
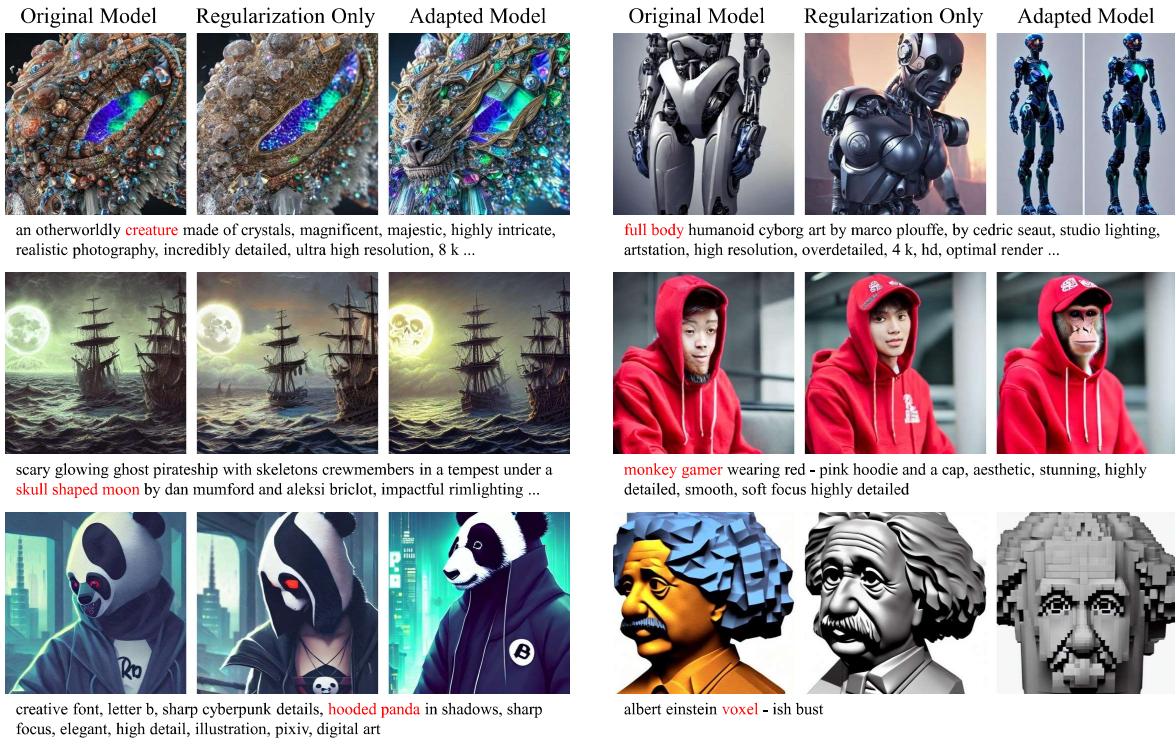
albert einstein voxel - ish bust

Figure 11. The adapted model generates images that better capture user intentions. Images in the same group are generated with the same prompt and random seed.

Figure 12. Screenshot of the user-study interface.