DS-Fusion: Artistic Typography via Discriminated and Stylized Diffusion Supplementary Material

1. Categories of Style Words

For the quantitative analysis in our paper, we tested 21 diverse categories of style words, covering animals (real or mythical), inanimate objects, professions, etc. The categories comprise the following: Astronaut, Mermaid, Butterfly, Music, Cactus, Octopus, Candle, Pilot, Cat, Plant, Chess, Robot, Cow, Shark, Dolphin, Socks, Dragon, Unicorn, Lion, Violin, and Zombie. In Fig. 1, we present some synthesized results from these 21.

2. Generating Style Images

To generate style images, we employ a Latent Diffusion Model which takes a style word and optionally a style attribute as the prompt. In practice, we also add two phrases into the prompt to obtain better styles images: (1) "cartoon illustration". The purpose is to generate images with a more graphical look for our stylization. (2) "white background". The purpose is to reduce noisy content in the backgrounds that could affect our stylization.

Some examples of style images generated for the 21 categories are shown in Fig. 2. The style images both guide the stylization of glyphs and ensure that the diffusion model will not forget the style concept. An ablation study was carried out (Fig. 3) to investigate the effects of not utilizing a "white background" during the generation of style images. While it is possible to achieve favorable results, there may be an increased occurrence of artifacts in the background, requiring additional effort to eliminate.

3. User Studies

Here, we provide details about our two user studies. The study's participants vary in art knowledge, gender, occupation, and education (Fig. 4). We start the studies by providing a description of Artistic Typography, as shown in Fig. 5. User interfaces for the first and second user studies are shown in Fig. 6 and 7. Fig. 8 shows the options of the "Rooster" example from the first study where we compare to DALL-E 2 [3], Stable Diffusion [4], and CLIPDraw [2]. Fig. 9 shows the options for the "Parrot" example used in the second study, where we compare our method to artist-designed typography. The options are randomly shuffled and the "They are equally good" option is appended to each question. In the second study, the artist-designed results are taken from a tutorial [1].

4. Custom Style Images

DS-Fusion can utilize custom styles images in addition to automatically generated ones. We manually drew six sketches of a novel cartoon character as the input and used a random conditioning prompt comprising seven characters. Fig. 10 exhibits the styles in our customized drawings naturally blended into the glyph "R".

5. Extensions to Common Glyphs/Shapes

DS-Fusion can also employ common shapes as the input glyphs, not limited to Latin alphabets. Fig. 11 demonstrates some examples of using objects such as chair, teddy bear, and car as the input glyphs, where "RABBIT", "ZOMBIE" and "UNICORN" are both the style words and conditioning prompts for chair, teddy bear and car, respectively.

6. Additional Results

In Fig. 12 and 13, we showcase more results of single-letter generation, where we compose stylized single letters into word images. In Fig. 14, we showcase more examples of multi-letter generation, where the input glyph is comprised of a set of letters, or a word, instead of a single letter. In the last row, we apply some style words to the glyphs whose content is different, to prove the generality of our method. We also show a side-by-side comparison with CLIPDraw [2] in Fig. 15.

DS-Fusion can produce diverse results for each combination of style word and glyph because it learns a style latent space for them. In Fig. 16, we demonstrate results sampled from different latent codes in both one-font and multiple-font input modes.

Figure 1. 21 categories of style words used in our quantitative analysis.

Figure 2. Examples of generated style images for the 21 categories of style words.

Figure 3. Effect of using a white background in style prompt when generating style images.

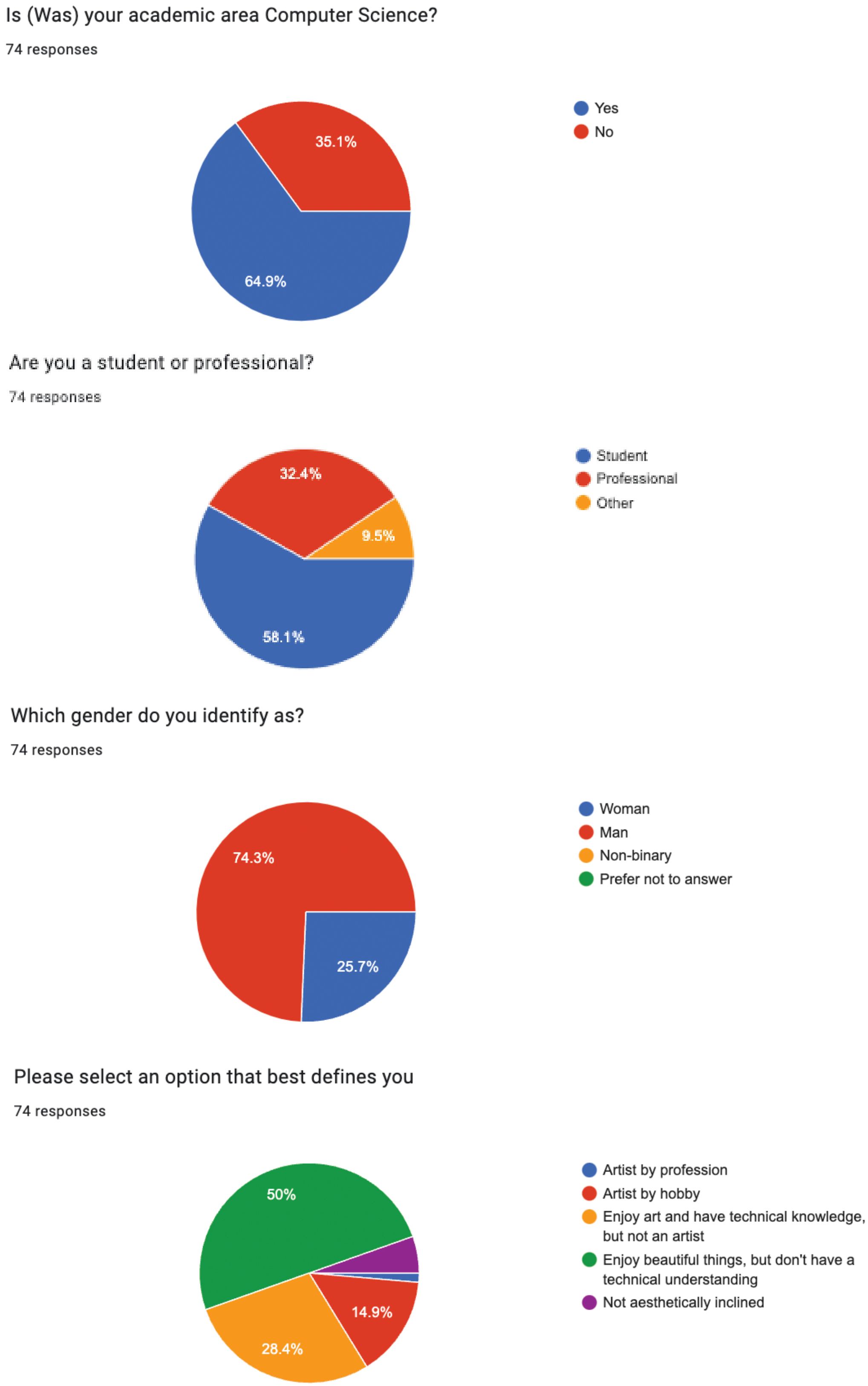
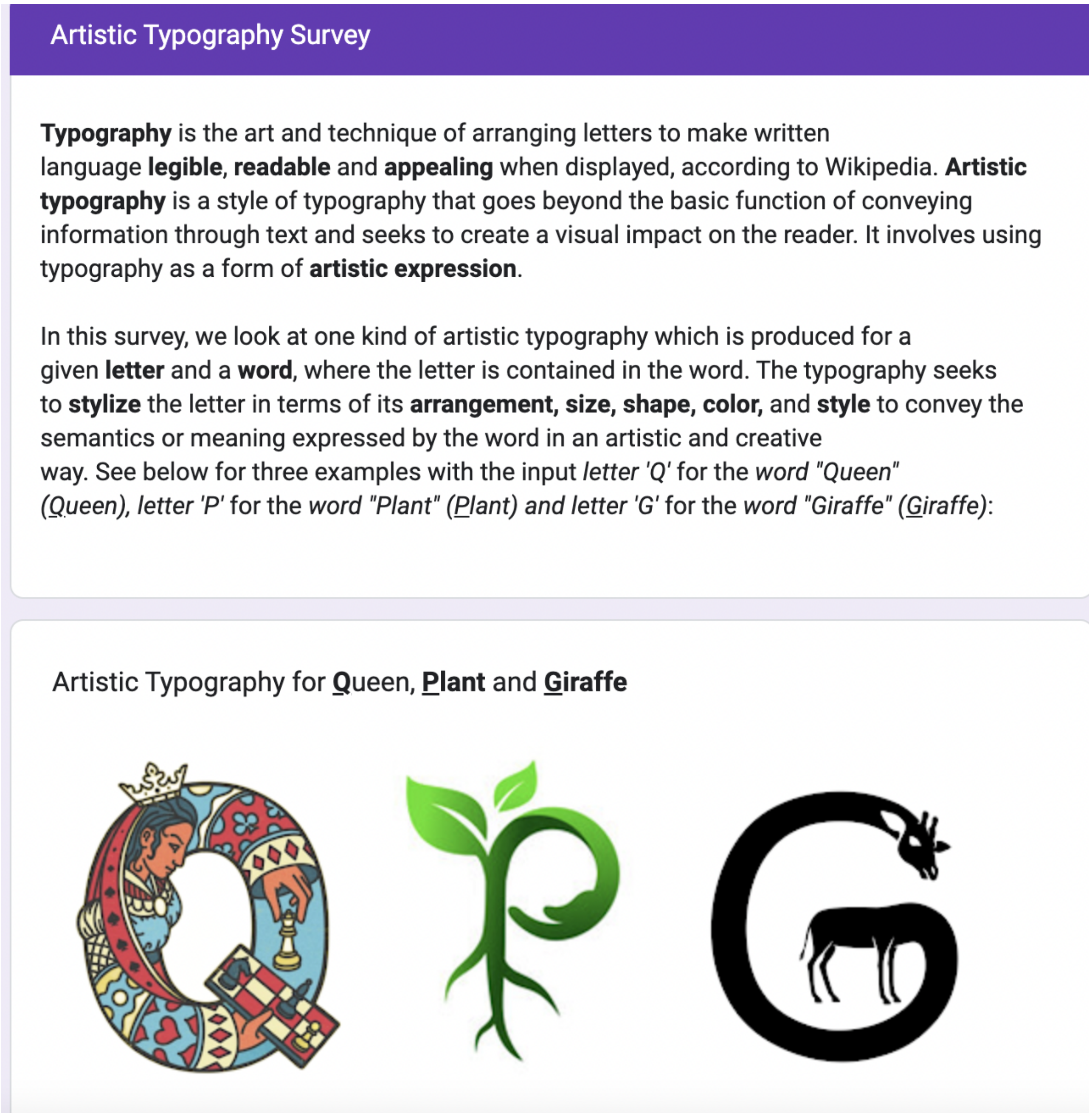


Figure 4. Stats of participants of the two studies. The stats are combined from participants of the two studies.



Artistic Typography Survey

Typography is the art and technique of arranging letters to make written language **legible**, **readable** and **appealing** when displayed, according to Wikipedia. **Artistic typography** is a style of typography that goes beyond the basic function of conveying information through text and seeks to create a visual impact on the reader. It involves using typography as a form of **artistic expression**.

In this survey, we look at one kind of artistic typography which is produced for a given **letter** and a **word**, where the letter is contained in the word. The typography seeks to **stylize** the letter in terms of its **arrangement, size, shape, color,** and **style** to convey the semantics or meaning expressed by the word in an artistic and creative way. See below for three examples with the input *letter 'Q'* for the *word "Queen" (Queen), letter 'P'* for the *word "Plant" (Plant) and letter 'G'* for the *word "Giraffe" (Giraffe)*:

Artistic Typography for **Q**ueen, **Plant** and **Giraffe**

Figure 5. The definition of Artistic Typography.

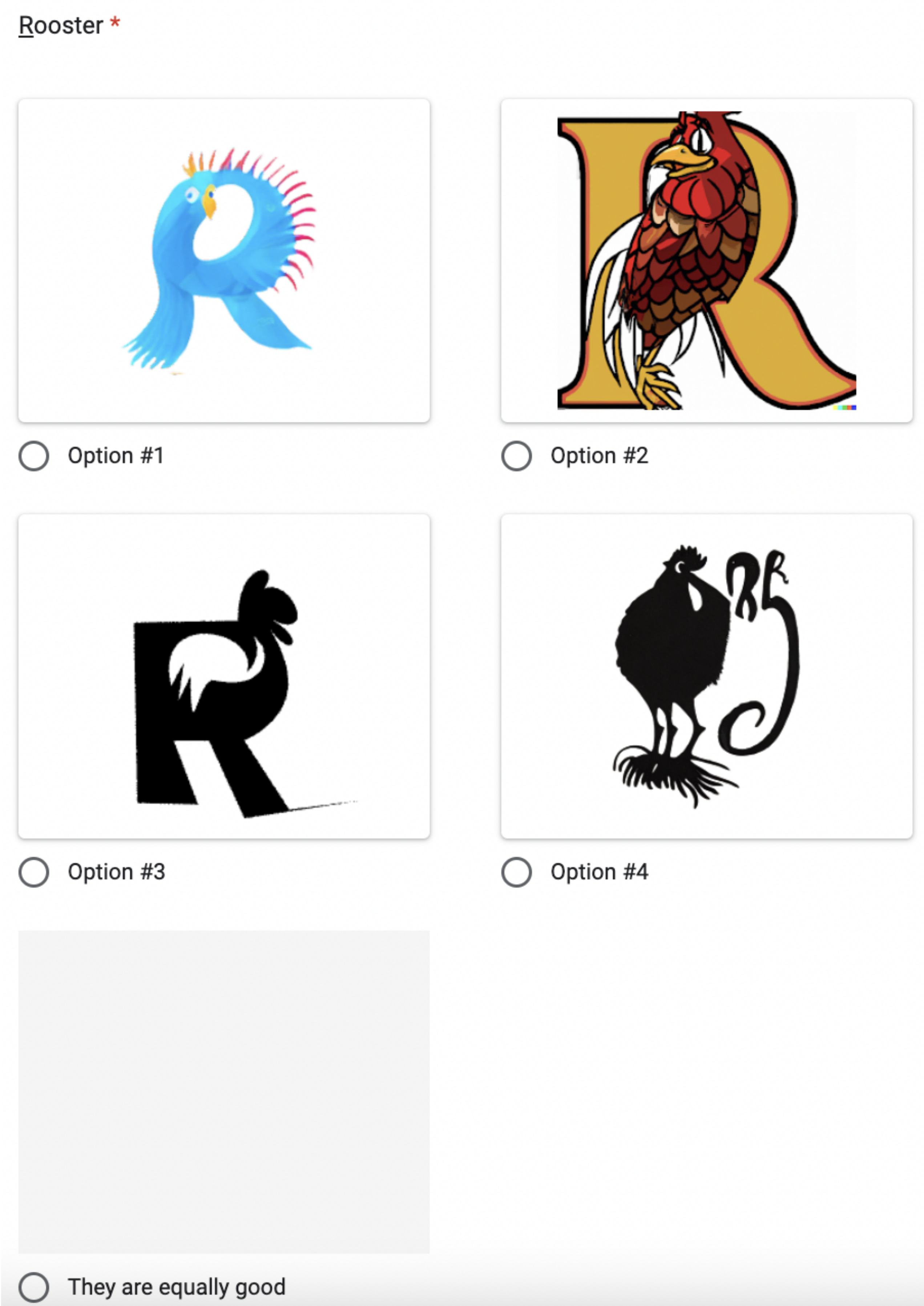


Figure 6. The user interface of the first user study.

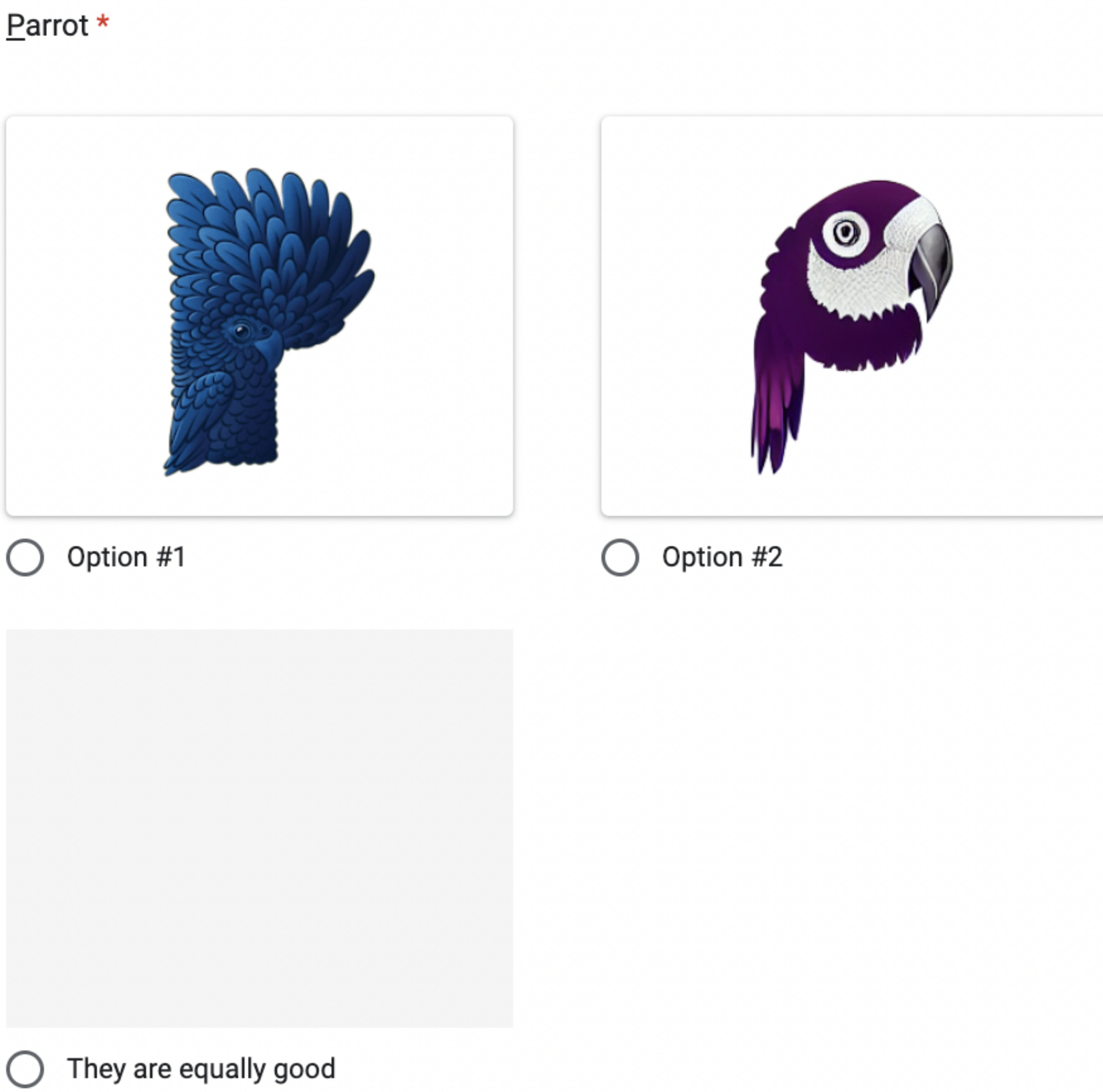


Figure 7. The user interface of the second user study.

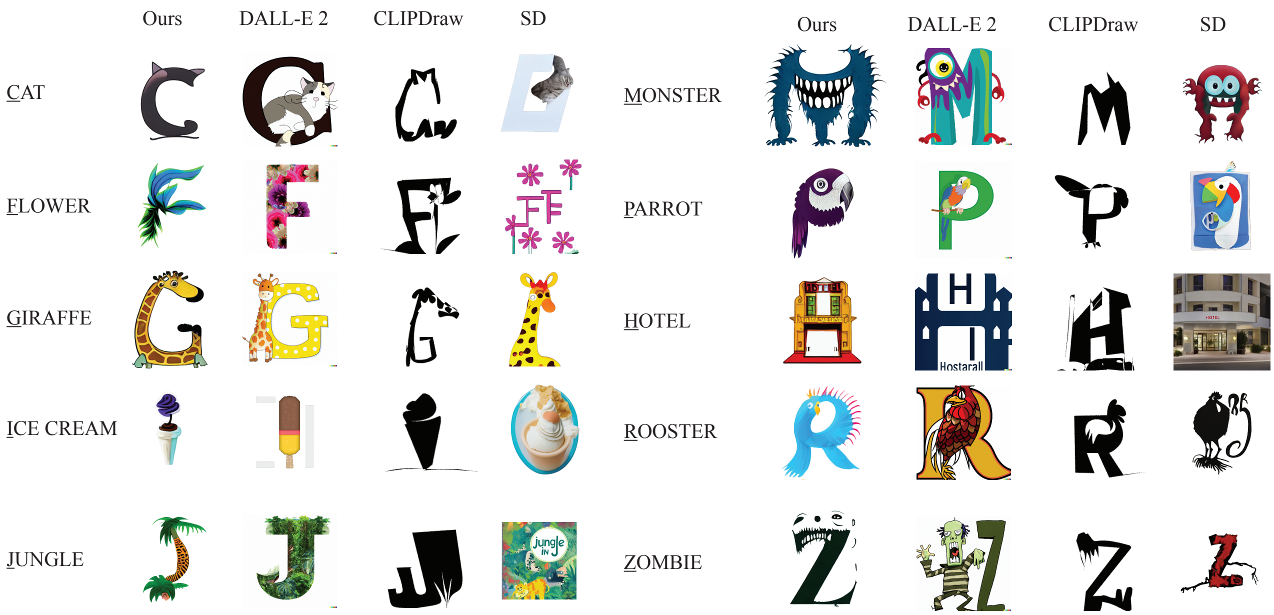


Figure 8. The options of the first user study, except “They are equally good”.

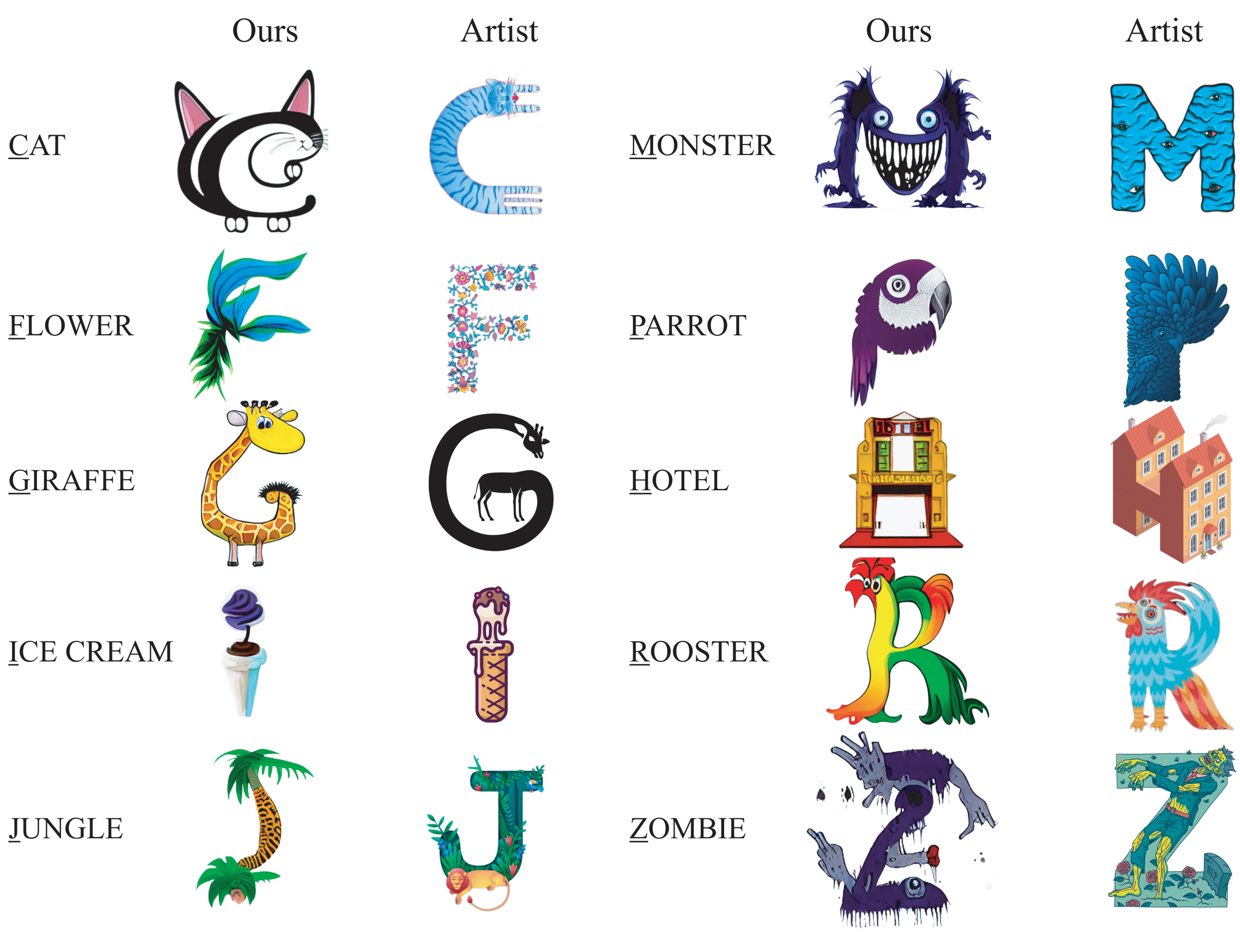


Figure 9. The options of the second user study, except “They are equally good”.

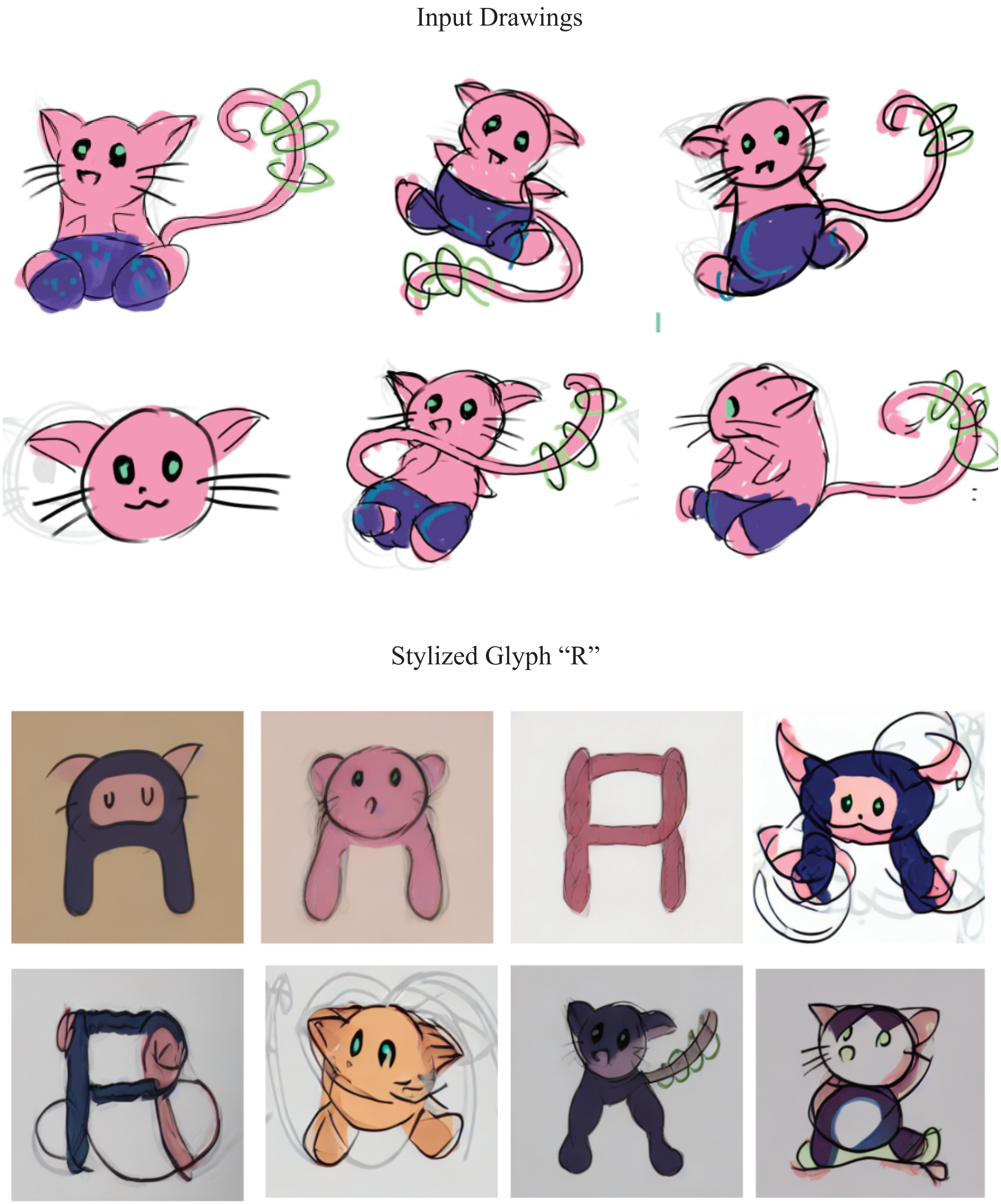


Figure 10. Utilizing customized style images as input.

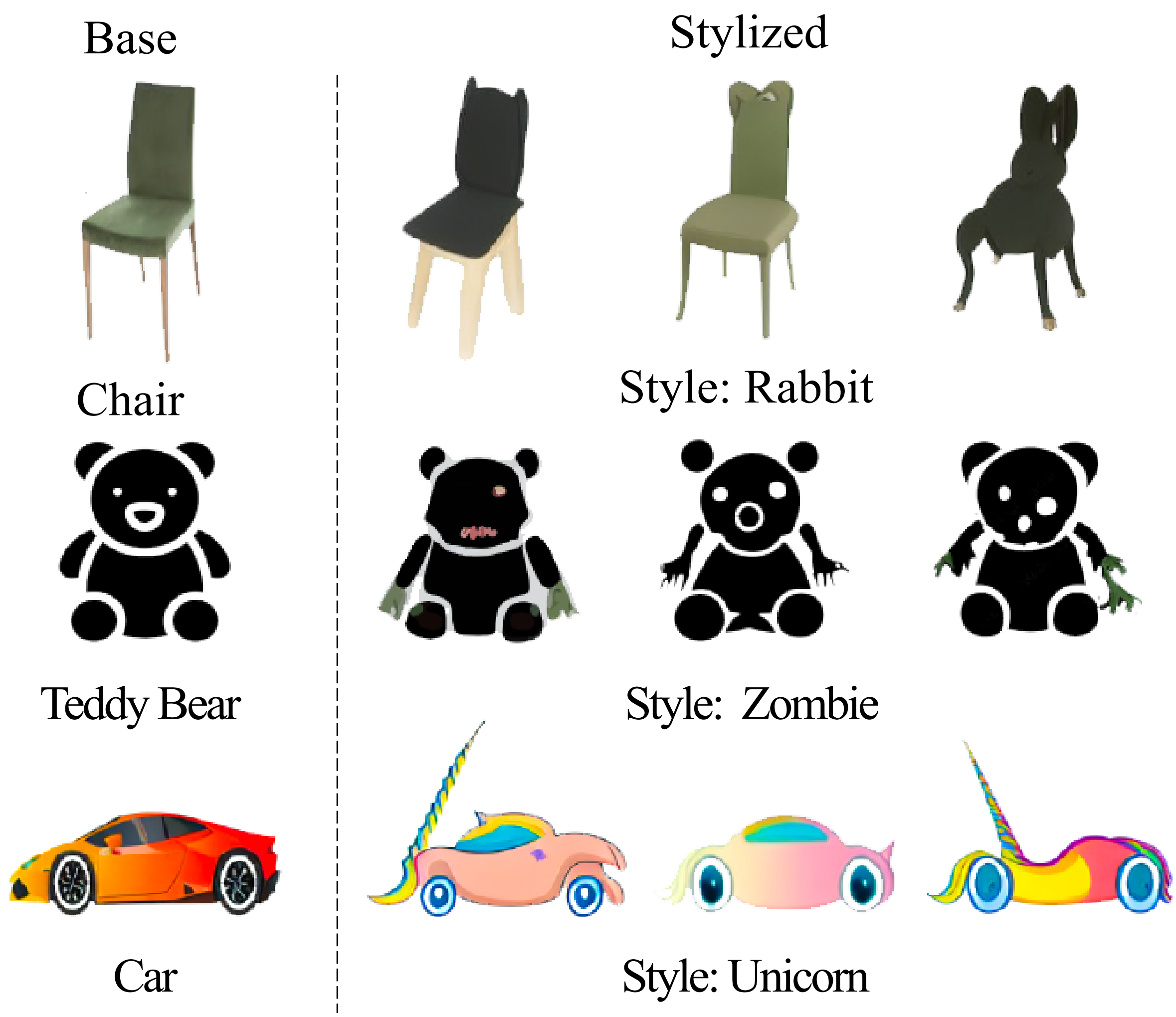


Figure 11. Employing common shapes as input glyphs.

Figure 12. More examples of single-letter generation (Part 1).

Figure 13. More examples of single-letter generation (Part 2).

Figure 14. More examples of multi-letter generation.

Figure 15. Multi-letter results compared with CLIPDraw [2].

Figure 16. Generating diverse results for given inputs in both one-font and multi-font input.

References

[1] Create Artistic Typography Designs with Amadine, howpublished = https://amadine.com/useful-articles/create-artistic-typography-designs-with-amadine. 1

[2] Kevin Frans, Lisa B Soros, and Olaf Witkowski. Clipdraw: Exploring text-to-drawing synthesis through language-image encoders. *arXiv preprint arXiv:2106.14843*, 2021. 1, 15

[3] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 2022. 1

[4] 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*, pages 10684–10695, 2022. 1