Appendix

A.1 Dataset for Evaluation

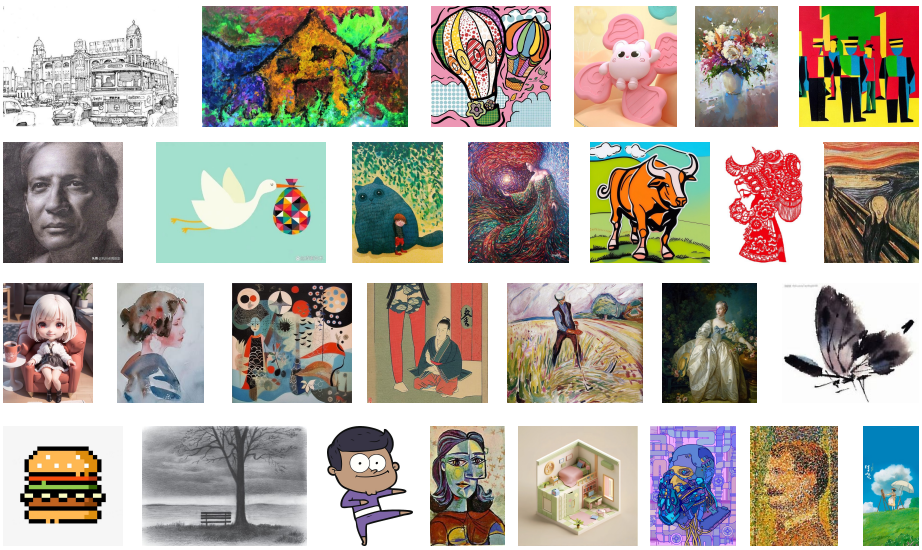

Figure 1: **The 28 style reference images with diverse styles for the generation models.**

In the experiment, we use a combination of 52 prompts and 28 style images, resulting in 1456 unique combinations, to guide the generation of images. These images are then used for quantitative evaluation. The list of prompts is shown in Tab. 1, using the same setting as in StyleAdapter [1]. The 28 style reference images are selected from the validation references to cover a wide range of styles, displayed in Fig. 1.

Table 1: **The list of prompts for experiments.**

Prompt
A robot.
A girl wearing a red dress, she is dancing.
A boy wearing glasses, he is reading a thick book.
A little cute boy.
A woman wearing a green sportswear, she is running.

A woman wearing a purple hat and a yellow scarf.
A man wearing a black leather jacket and a red tie.
A little boy with glasses and a watch.
A smiling little girl.
A little boy playing football.
A curly-haired boy.
A little girl holding flowers.
A lovely kitten walking in a garden.
A puppy sitting on a sofa.
A fluffy white rabbit with pink ears and nose.
A brown puppy with black spots and a red collar.
A black and white panda.
A dog in a bucket.
A cat wearing a hat.
A cute little fish in aquarium.
A bird in a word.
A kitten sleeping on a pillow.
A parrot singing a song.
A monkey playing with a banana.
A turtle wearing sunglasses.
A hamster eating a carrot.
A white rose.
A sunflower smiling at the sun.
A cactus wearing a hat.
A daisy with a ladybug on it.
A pine tree with a snowman hugging it.
A mushroom in winter.
A beautiful lotus.
A lotus with a frog meditating on it.
A cherry blossom.
A palm tree.
A river with rapids and rocks.
A creek with clear water and colorful pebbles.
A lake with calm water and reflections.
A waterfall with mist and rainbows.
A stone with a face carved on it, standing on a pedestal in a museum.
A stone with a hole in it.
A stone with a pattern of stripes on it.
A stone with a crack in it, holding a plant growing out of it.

A snowy mountain peak.
A mountain goat on a cliff.
A red baseball cap.
A football on the grass.
A motorcycle.
A modern house with a pool.
A house made of cardboard boxes.
A house covered with ice and snow.

A.2 Ablation study of hyper-parameter.

To verify the stability of the trade-off between instruction following and style control of this method, we evaluate our method and IP-Adapter under different control weights, respectively. As shown in Fig.2, by applying different style control weights, the semantics of the generated images remain unchanged. And a higher weight has a greater impact on the style of the image.

A.3 Visualization of Style Similarity Score

In the experiment, we employ a style encoder to measure the style similarity between two images. In Fig. 3, we present different generated images and annotate their style similarity scores compared with their respective style reference images. It can be observed that the style encoder accurately assesses the similarity of styles between two images. Moreover, images with a style similarity score greater than 0.4 exhibit a favorable level of style similarity.

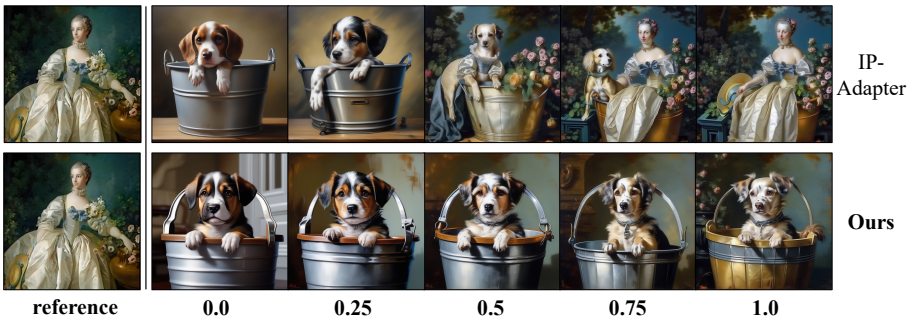

Figure 2: **Ablation study of hyper-parameter.**

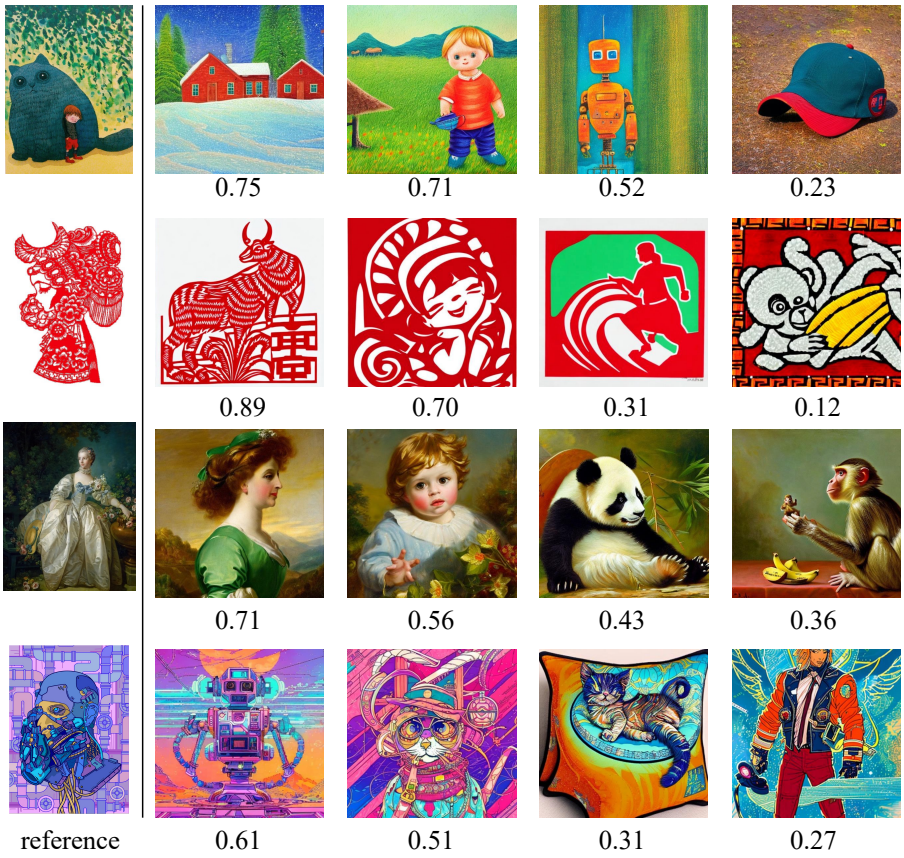

Figure 3: **Comparison of style similar scores between different pairs of style images and reference images.**

A.4 More Generated Images With More Styles

In Fig. 4, we present additional generated results. The first image in each row represents the style reference image, followed by the corresponding generated results. The prompts for each image are listed below the images. It can be observed that our method can generate images in various styles.

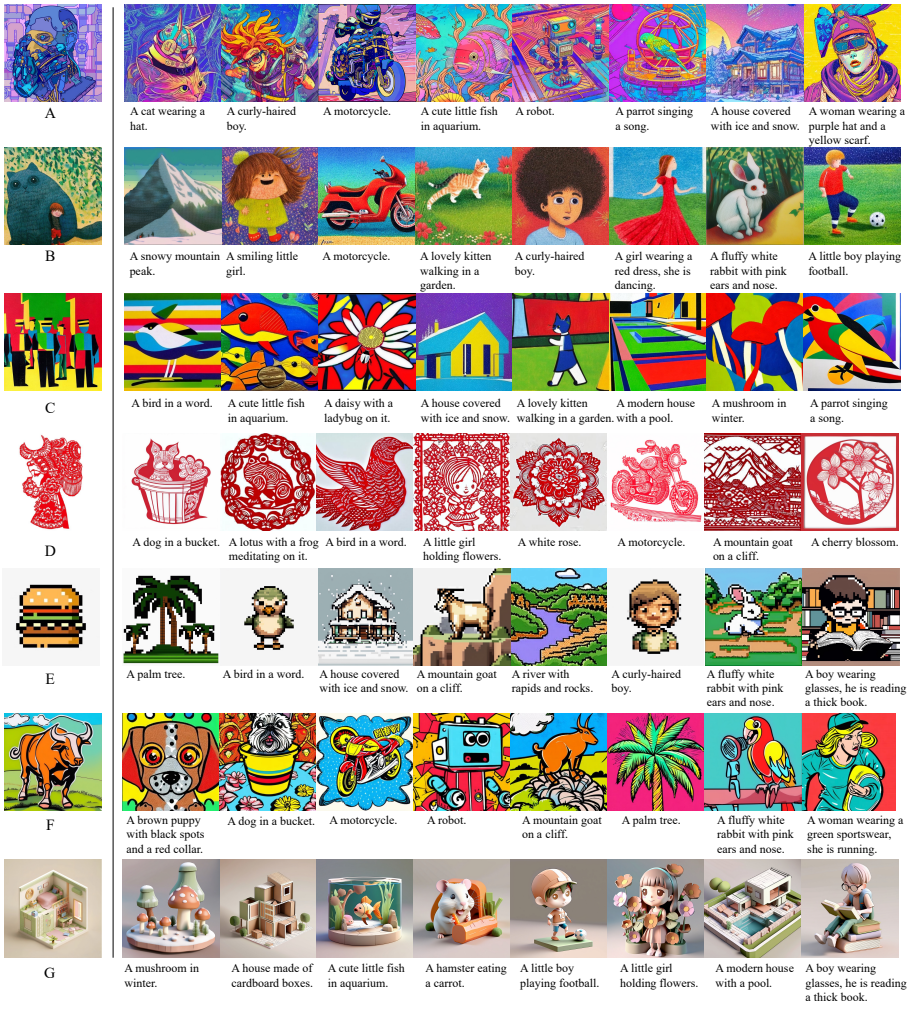

Figure 4: **Additional visual results generated by our algorithm.** The first image on the left of each row is the corresponding style reference image, followed by the prompt and its corresponding generated results.

References

[1] Wang, Z., Wang, X., Xie, L., Qi, Z., Shan, Y., Wang, W., Luo, P.: Styleadapter: A single-pass lora-free model for stylized image generation. arXiv preprint arXiv:2309.01770 (2023)