Supplementary Materials

1. Image Augmentation

We have used eight image augmentation methods of the Torchvision v0.13 library, which are widely applied in many solutions of computer vision problems. Below is the list[1] of the functions we applied in all of our experiments.

List 1. Image Augmentation Functions

1. RandomInvert
2. RandomPosterize
3. RandomSolarize
4. RandomAdjustSharpness
5. RandomAutocontrast
6. RandomEqualize
7. RandomHorizontalFlip
8. RandomVerticalFlip

As can be noted from List 1, it does not contain common geometric augmentation functions (such as *RandomRotate*, *RandomShear*, etc...), because they easily cause augmentation leakages. Those functions are selected by experiments to avoid having the same problem.

2. Prompts Complexity

In Section 4.3 of the main paper, we highlighted the specificity of the prompts we use for the few-shot diffusion process for style extraction. The prompts can be very diverse, containing descriptions about the environment, colors, styles, or even parameters of the camera and details about image resolution. The complexity of the prompts is important, so we defined three types of them in our experiments: *Multiplex*, *Contrary*, and *Ordinary*.

Multiplex - The prompts of this type usually are long, containing more than 30 words. They aim to provide details about the environment, objects, structures, and visual appearance with many words (see Table 1). They even contain specifications about the camera device or image resolutions (even though the model synthesizes fixed-size images). The main issue of these prompts is the distance from the common captions in the embedding space. We believe that the multiple words of too complex prompts push the representation farther from a small cluster in the embedding space we have during training, which makes the model forget about fine-tuned information of few-shot examples. Figure 1 contains examples where our model ignores the style of the few-shot dataset (*flat design*) it was trained on.

Contrary - As we focus on the style in our few-shot diffusion process, it is tricky when the prompts contain details about the opposite appearances. Some characteristics or phrases are contrary to the style expressed in few-shot examples, which worsens the performance of the model. E.g. few-shot dataset contains only *line-art* images, but the prompt includes phrases such as *red flower, yellowish, warm lights, etc*. Other examples can be found in Table 1 and Figure 1. In common, contrary words are style specifications, environmental descriptions, color, or many other features.

Ordinary - The prompts which are not *Multiplex* or *Contrary*, considered to be *Ordinary* prompts. We used only ordinary prompts in all our experiments and their visualizations, to express the style of few-shot examples.

3. Sparsity

The number of examples and the complexity of their captions affect the generalization of few-shot image synthesis. The technique of sparse updating helps to improve the generalization of the model. We have an ablation study for different sparsity levels in Figure 2. As can be seen from the comparison, 1% updating misses many visual features of few-shot examples (see last two columns of Figure 2). Other sparsity levels, such as 20% or 100% often have close results.

4. More complex text prompts & failure cases.

A few additional results with more complex text prompts are shown in Figure 3.

[1] Augmentation functions: https://pytorch.org/vision/stable/auto_examples/plot_transforms.html

Multiplex - Too complex prompts
• "A tree to the left of the park bench on the lawn, and golden god rays shine through the gaps in the branches and leaves, insane details, dramatic lighting, unreal engine 5, concept art, greg rutkowski, james gurney, johannes voss, hasui kawase" • "The exterior of a house in devonshire that was built in the 1970s and is rumoured to be haunted, painterly, offset printing technique, photographed on kodachrome by brom, robert henri, walter popp, various refining methods, micro macro autofocus, ultra definition, award winning photo" • "Intricately detailed porcelain carved chrysalis, explosion of butterflies, fantasy pop surrealism by peter mohrbacher, james jean, alena aenami"
Contrary - Opposite words in the prompts
• "An old 1800's chalet with dusty floor, old library with sunset light through dusty windows, old fireplace, old chairs, hyperdetailed, artstation, cgsociety, 8k" • "Stargate made of stone that form a circle, cinematic view, epic sky, detailed, concept art, low angle, high detail, warm lighting, volumetric, godrays, vivid, beautiful, trending on artstation, by jordan grimmer, huge scene, grass, art greg rutkowski" • "Inside a realistic, 4k, octane render, raindrop, a dystopian futuristic city with heavy smog and tall buildings with neon signs and video billboards, dimly lit by the sun. diffused lighting, highly detailed digital art, trending on artstation"
Ordinary - Plain type of prompts we used
• "A man is walking in the park with his dog" • "A boy is playing with butterflies" • "Full moon, sky with full of stars, and moaning wolf in the forest"

Table 1. Complexities of the prompts

References

[1] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S Sara Mahdavi, Rapha Gontijo Lopes, et al. Photorealistic text-to-image diffusion models with deep language understanding. *arXiv preprint arXiv:2205.11487*, 2022. 4

Figure 1. Samples generated by the corresponding prompts of Table 1 with our model trained on *flat design* dataset.

Figure 2. Samples are generated with **1% sparsity, 10% sparsity, 20% sparsity,** and **100% sparsity (full)** accordingly by using "*small house in the forest, dark night, leaves in the air, mushrooms, animals, gibli, james gilleard, atey ghailan, lois van baarle, jesper ejsing, pop art patterns, exquisite lighting, clear focus, very coherent, plain background, very detailed*" prompt.

(a) (b) (c) (d)

Figure 3. Generation with complex prompts: *woman facing to the front / a castle on a hilltop beside a lake with swans / paying for a quarter-sized pizza with a pizza-sized quarter / a storefront with 'Hello World' written on it*. The latter two prompts from DrawBench [1] often fail other models too.