SliderSpace: Decomposing the Visual Capabilities of Diffusion Models

Supplementary Material

A. Principal Component Analysis

Our analysis reveals that concepts frequently encountered in training data (e.g., ”person”) exhibit greater variation compared to concepts that are either less diverse or less common (e.g., ”Van Gogh art” or ”waterfalls”). We demonstrate this by analyzing the principal components for each concept through PCA visualization in Figure A.1. Notably, the 50th principal component for the ”person” concept shows comparable variational magnitude to the 20th component of ”waterfalls,” highlighting the inherent variational differences across concepts. By discovering and uniformly sampling these variations, we effectively address the mode collapse problem in models, as shown in Figures E.2 and E.3.

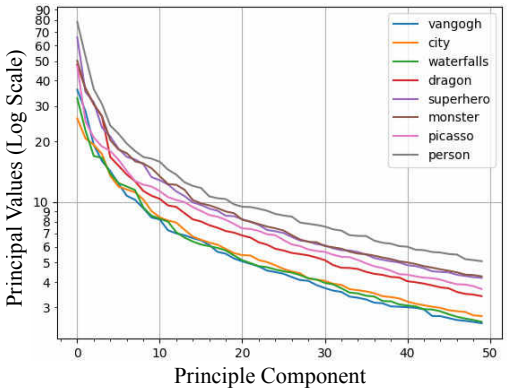

Figure A.1. Common concepts like “person” show higher variation in CLIP space compared to rarer concepts like “waterfalls”. 50th PCA component of “person” matches the 20th component of “waterfalls,” indicating the latter’s more limited variation.

B. Effect of TimeStep during Inference

The temporal application of sliders during inference significantly impacts both the precision and magnitude of image edit(Fig. E.4) for SDXL-DMD SliderSpace. When sliders are applied at all timesteps during inference, we observe strong semantic and structural changes in the generated image. But applying the slider after a few steps helps preserve the image structure while still enabling controlled edits. This latter approach facilitates more precise editing, albeit with subtler semantic alterations that can be amplified by increasing the slider strength parameter.

B.1. Choice of Semantic Embeddings

While our primary implementation uses CLIP embeddings for semantic decomposition, SliderSpace is compatible with various semantic encoders. Our experiments with alternative embeddings like DINO-v2 and FaceNet demonstrate the framework’s flexibility. As shown in Figure B.1. DINO-v2 shows comparable overall performance to CLIP, with each encoder exhibiting different strengths across various concepts. For person-specific concepts, using FaceNet embeddings enables the discovery of fine-grained facial semantic directions as seen in Figure 8

The choice of encoder can be tailored to the target domain - CLIP for general concepts, DINO-v2 for certain visual attributes, and specialized encoders like FaceNet for domain-specific applications. This flexibility allows SliderSpace to adapt to different use cases while maintaining its core benefits of unsupervised discovery and semantic consistency.

B.2. Hyperparameter Analysis

We analyze the impact of two key hyperparameters in SliderSpace: the number of PCA directions and the LoRA rank. Our experiments reveal that increasing PCA directions improves both knowledge coverage and output diversity up to about 40 dimensions, after which returns diminish. With just 10 directions, SliderSpace matches the FID scores of 64 manually created Concept Sliders when evaluated against artistic style distributions. Regarding model architecture, we find that lower-rank adaptors (particularly rank-one) efficiently capture variations with a fixed training budget, outperforming higher-rank versions while maintaining better FID scores than Concept Sliders across different ranks.

This analysis guides our choice of using rank-one adapters with 40 PCA directions as the default configuration, offering an optimal balance between performance and computational efficiency.

C. User Study

We conducted user studies to evaluate SliderSpace’s effectiveness through Amazon MTurk. For artistic evaluation (Sec 5.2), participants compared two 9-image grids - one generated by SliderSpace using 3 random sliders per image, and another by our baselines. Both sets used identical base prompts: “a building in a stunning landscape” and “a character in a scenic environment”. As shown in Fig E.12, participants rated which grid exhibited greater artistic diversity and utility for art applications. For conceptual evaluation (Sec 5.1), participants compared image grids based on diversity, generative utility, and creativity (Fig E.13). Grid presentation order was randomized across all experiments.

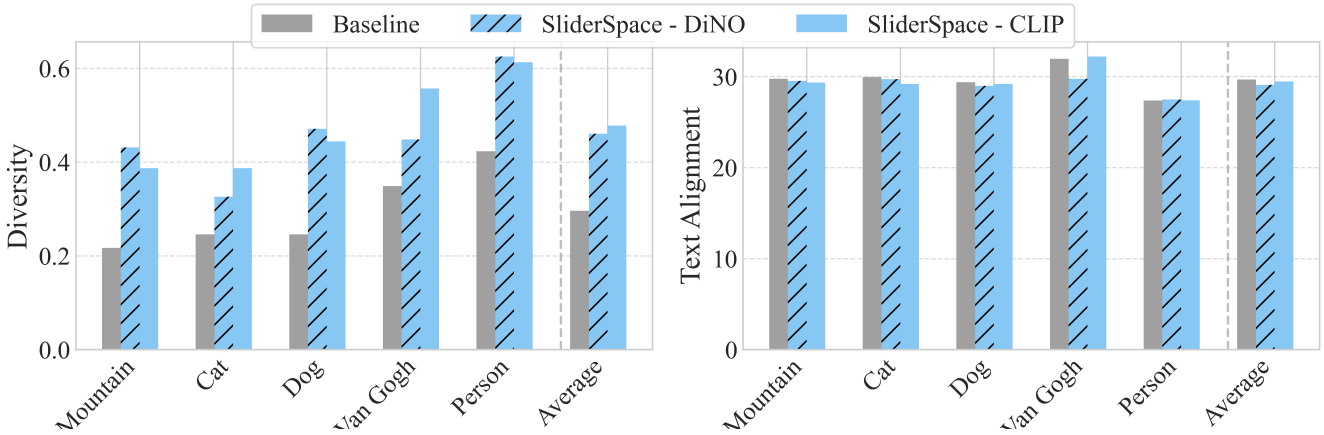

Figure B.1. SliderSpace shows similar diversity and text alignment when using either Dino-V2 or CLIP embeddings for PCA analysis.

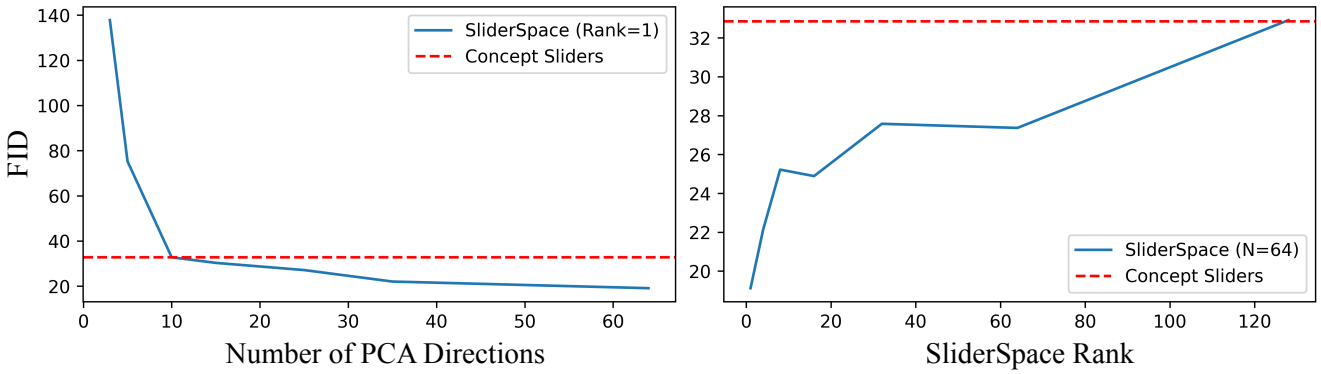

Figure B.2. Concept Sliders Comparison & Hyperparameter analysis: (Left) Impact of PCA directions: SliderSpace with 10 directions matches the FID of 64 Concept Sliders. More directions, upto 40, leads to improved FID. (Right) Effect of LoRA rank: Given a fixed training budget rank-one sliders are efficient than higher rank versions and outperforms Concept Sliders

D. Qualitative Results

D.1. Art Exploration

We identify the top-36 distinct art directions discovered by SDXL-DMD2 SliderSpace for the concept "artwork in the style of famous artist" in Figures E.5 and E.6. Additionally, we showcase various combinations of SliderSpace samples in Figures E.8 and E.9, where we randomly sample three sliders to generate images for both characters and buildings (used in our art experiments and user studies in Section 5.2). The top 18 art styles discovered by SDXL-SliderSpace are presented in Figure E.7.

D.2. Diversity Enhancement

We provide additional qualitative examples demonstrating how our generic diversity sliders mitigate mode collapse in distilled models. Our observations indicate that distilled models such as DMD2 [51] tend to generate visually similar images for identical prompts, despite different random seed initializations. Through our trained diversity sliders, which are model-agnostic, we successfully counter mode collapse (Section 5.3). As quantitatively validated in Table 3, the diversity SliderSpace significantly improves image variation, achieving FID scores comparable to the base model.

D.3. Concept Decomposition

We present qualitative examples of concept decomposition using the SDXL-DMD2 [51] SliderSpace in Figures E.14–E.19. Furthermore, we demonstrate SliderSpace's versatility across various models, including SDXL-Turbo [43] (Figures E.20, SDXL-Base [34] (Figures E.7), and the state-of-the-art transformer-based FLUX Schnell models (Figures 1 and E.21). We note that Claude3.5 [3] generated captions are not always accurate. For instance, in Figure E.14, Claude annotates one of the sliders as "Black Lab Technician", but it is not visually distinct whether the slider is 'lab technician' or a 'scientist'.

E. Ablations

We analyze the key components of our method and validate their necessity: (1) the semantic orthogonality objective, (2) expanding diversity of training samples, and (3) CLIP

embedding analysis. Figure E.1 shows qualitative examples and FID measures on art exploration experiments. In both the qualitative and quantitative experiments, we find that uniqueness criteria in Eqn 5 is very important to get diverse discovery of SliderSpace. When we extract a naive-SliderSpace by training multiple sliders on a single concept using regular customization [29, 41] loss and no contrastive objective, many redundant and junk directions appear, as shown in Fig. E.1(a). This baseline is equivalent to Liu et al. [31]. Similarly, by applying our objective (Eq. 5) on diffusion output space $\tilde{x}_{0,t}$ (Eq. 2) rather than CLIP space, SliderSpace discovers directions that are more relevant in color and shape but not semantic variations, as shown in Fig. E.1(b). This baseline is slider equivalent version of NoiseCLR [9]. Finally, diversity expansion of training data (Fig. E.1 d,e), helps with expanding a diverse set of sliders. This can be used to improve the variation across sliders. We use SDXL for generating images in concept and art experiments. For diversity experiments, we use LLM prompt expansion as we compare against SDXL as baseline.

.

Figure E.1. We conduct our ablations on the art-exploration application and show FID scores as a measure of diversity. SliderSpace contrastive objective (Eq. 5 is essential for discovering diverse directions. Ablating CLIP space analysis and performing spectral analysis in diffusion output space (Eq. 2) results in sliders that control color, texture and shapes. We also find that expanding the training data diversity using LLM enhanced prompts and base SDXL models can help with improved distilled model's SliderSpace diversity

Figure E.2. We show a few possible variations possible with SliderSpace directions. For a given seed and prompt, users can sample different combinations of sliders from SliderSpace and generate unique and diverse outputs (all variations from a single prompt and seed). We show this for the concept "Van Gogh" SliderSpace on SDXL-DMD2 [51].

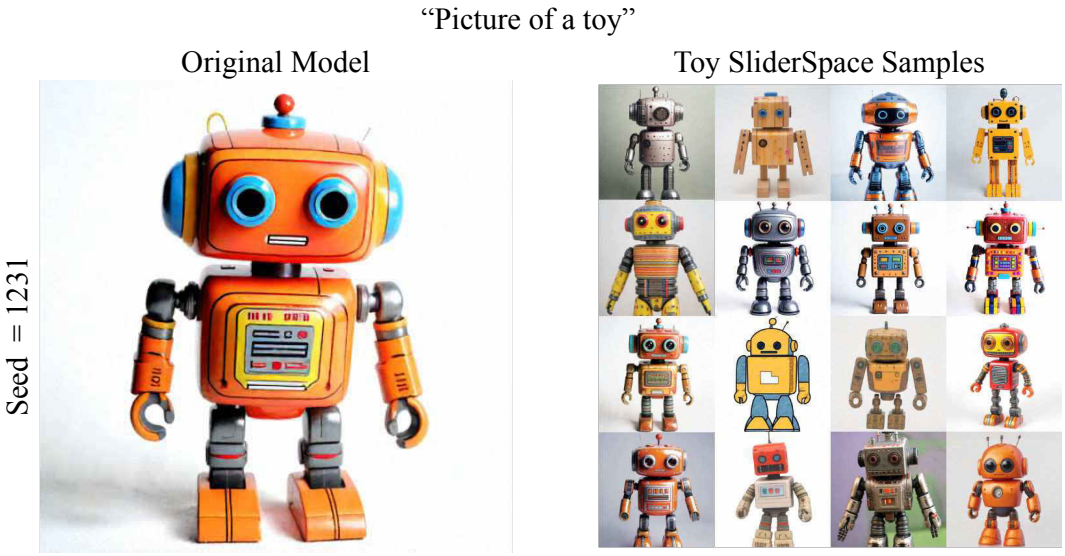

Figure E.3. We show a few possible variations possible with SliderSpace directions. For a given seed and prompt, users can sample different combinations of sliders from SliderSpace and generate unique and diverse outputs (all variations from a single prompt and seed). We show this for the concept “Toy” SliderSpace on SDXL-DMD2 [51].

Figure E.4. The choice of timestep at which sliders are applied can have an effect on the preciseness of the sliders. We show that when the sliders are applied to all the timesteps in inference, the images look different from the original models images for the same prompt and seed. But skipping the first timestep can lead to precise edits (similar observations as [16])

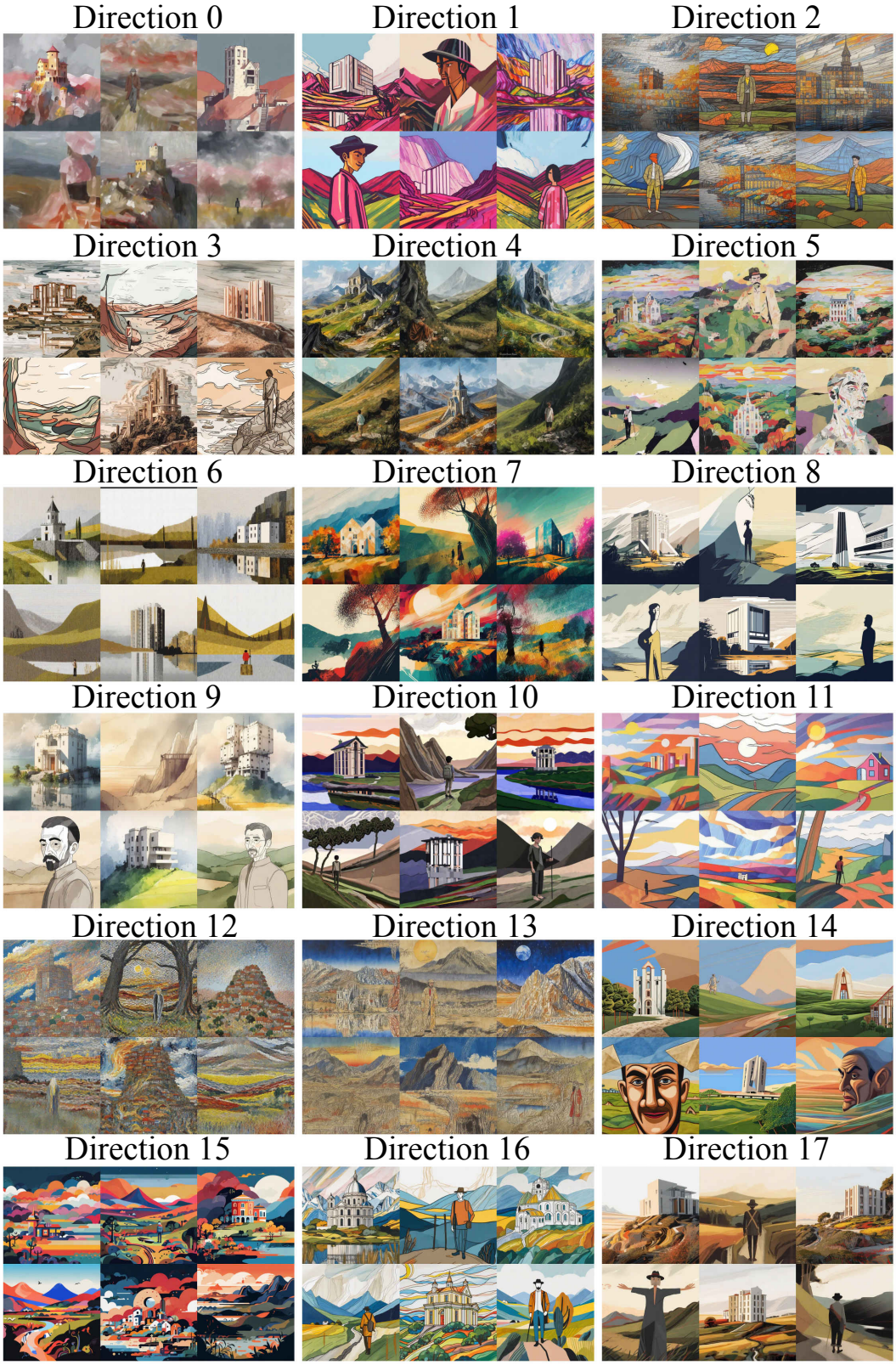

Figure E.5. We show the top 18 art directions that are discovered in the SDXL-DMD2 [51] SliderSpace for the concept “art”.

Figure E.6. We show the top 18-36 art directions that are discovered in the SDXL-DMD2 [51] SliderSpace for the concept "art".

Figure E.7. We show the top 18 art directions that are discovered in the SDXL [34] SliderSpace for the concept “art”.

Figure E.8. We show samples from our art experiments 5.2. We sample random 3 sliders from the SDXL-DMD2 [51] SliderSpace for the concept “art” and generate images for the prompt “a building in a stunning landscape the style of a famous artist”.

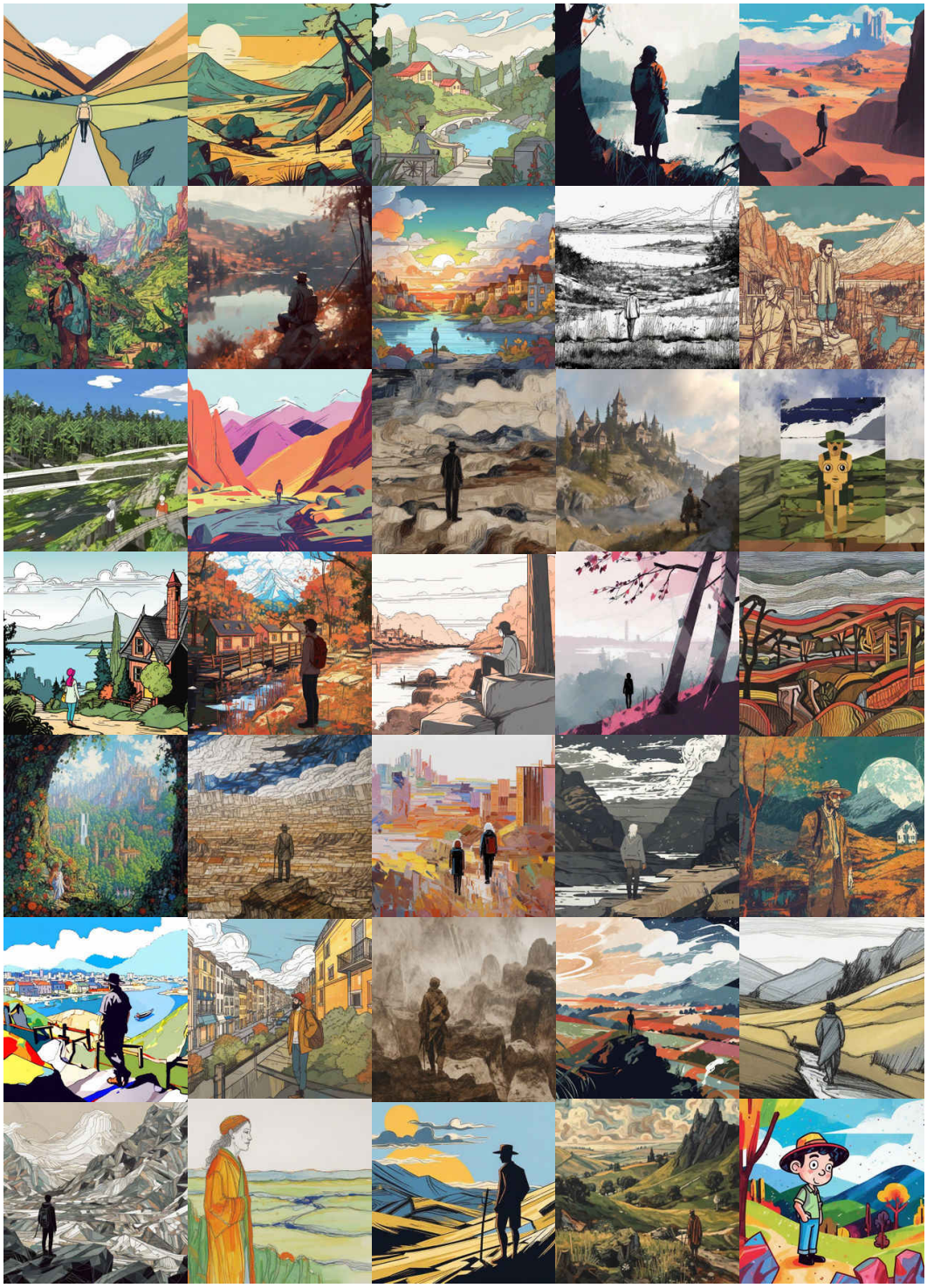

Figure E.9. We show samples from our art experiments 5.2. We sample random 3 sliders from the SDXL-DMD2 [51] SliderSpace for the concept “art” and generate images for the prompt “a character in a scenic environment the style of a famous artist”.

Figure E.10. We show samples from our diversity experiments 5.3. We sample random 3 sliders from the SDXL-DMD2 [51] diversity SliderSpace. We find that the common diversity sliderspace has a visual improvement in diversity and reverses the mode collapse in the distilled models

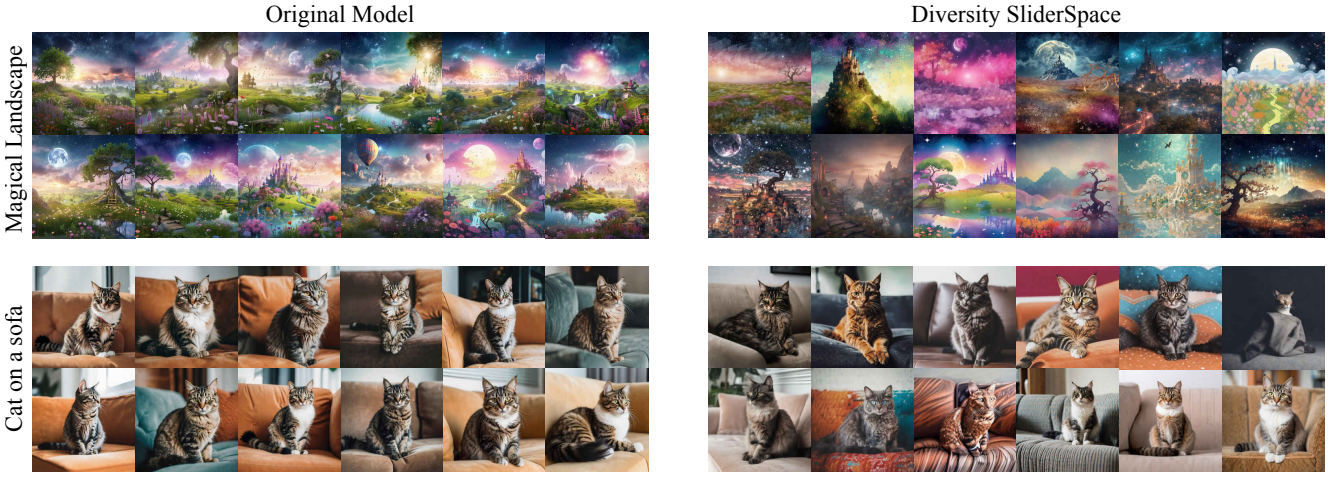

Figure E.11. We show samples from our diversity experiments 5.3. We sample random 3 sliders from the SDXL-DMD2 [51] diversity SliderSpace. We find that the common diversity sliderspace has a visual improvement in diversity and reverses the mode collapse in the distilled models

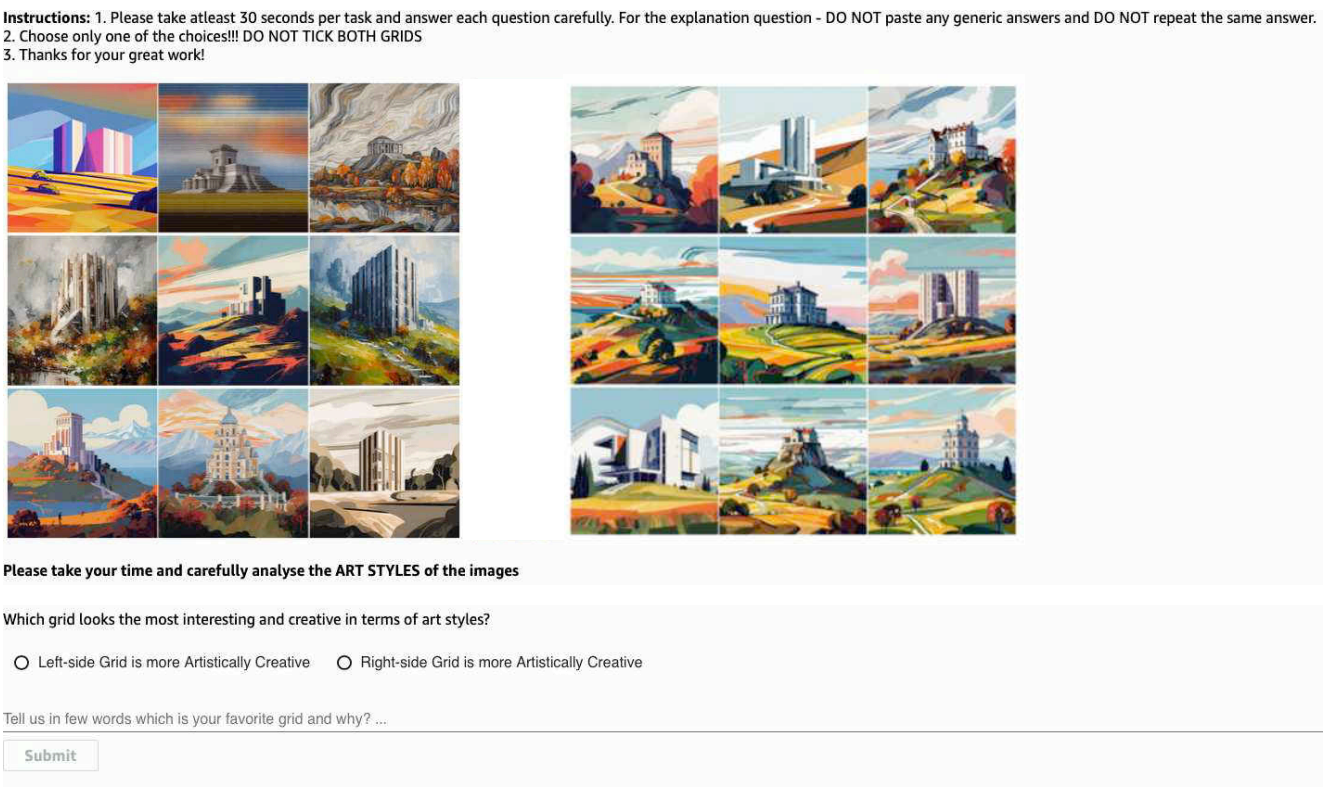

Figure E.12. User study interface on Amazon Mechanical Turk. Users are shown images randomly sampled from SliderSpace or our baselines (Sec: 5.2, and asked to identify the grid with most creative art renditions.

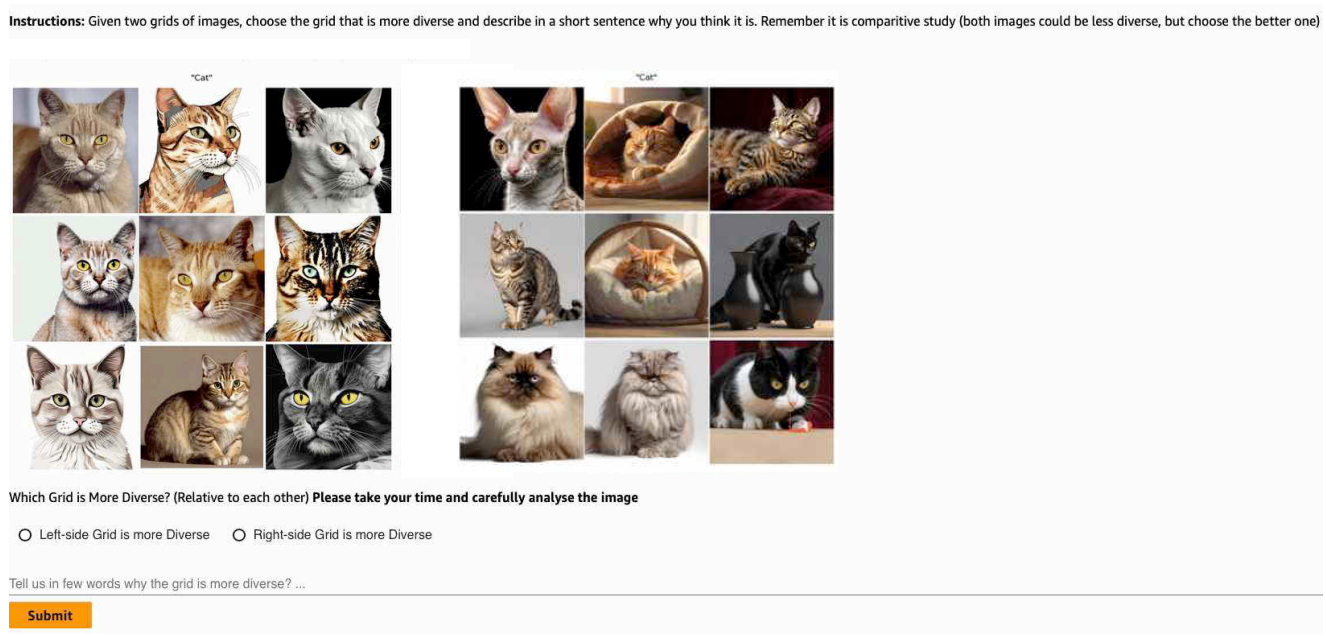

Figure E.13. User study interface on Amazon Mechanical Turk. Users are shown images randomly sampled from SliderSpace or our baselines (Sec: 5.1, and asked to identify the grid with most diverse outputs.

Figure E.14. We show the SliderSpace discovered in SDXL-DMD2 4-step model [51] for the concept "Scientist"

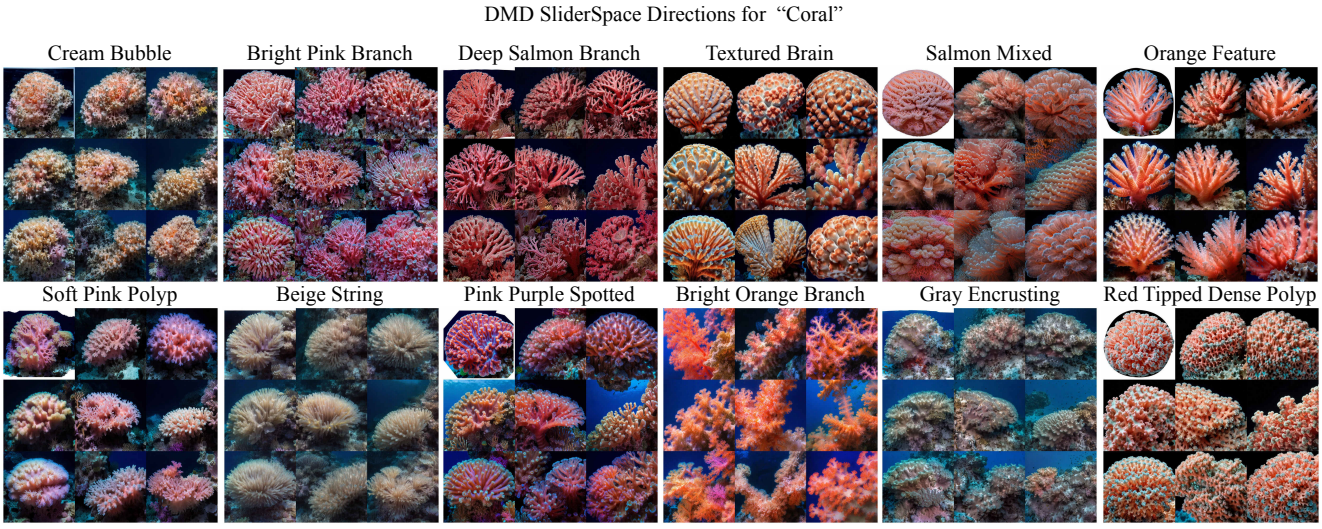

Figure E.15. We show the SliderSpace discovered in SDXL-DMD2 4-step model [51] for the concept "Coral"

Figure E.16. We show the SliderSpace discovered in SDXL-DMD2 4-step model [51] for the concept "Cowboy"

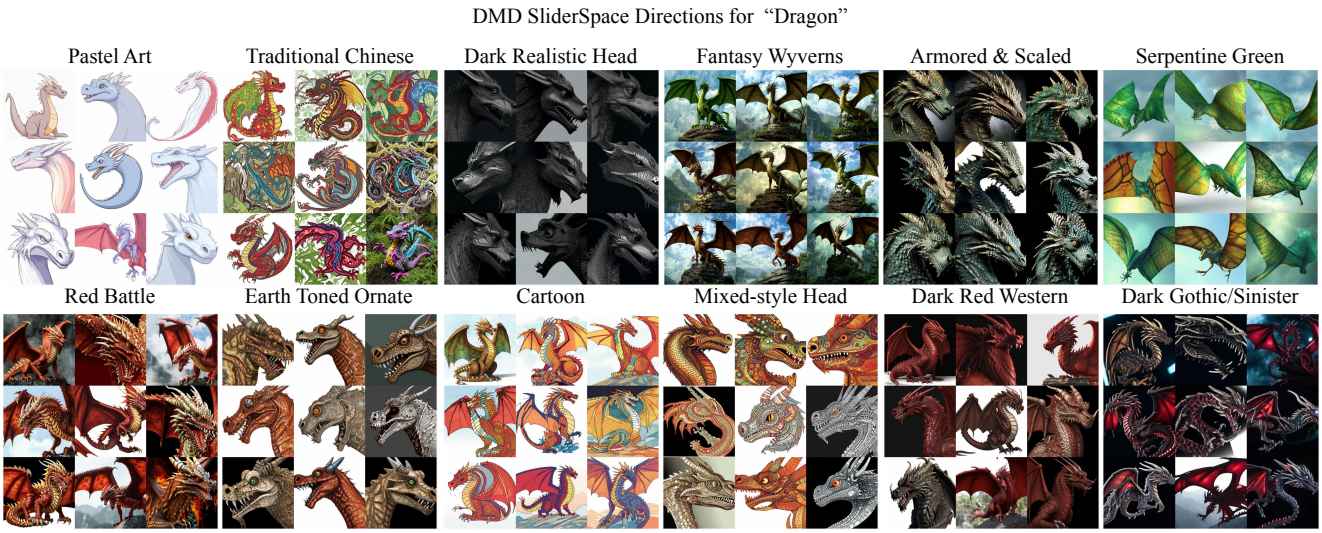

Figure E.17. We show the SliderSpace discovered in SDXL-DMD2 4-step model [51] for the concept "Dragon"

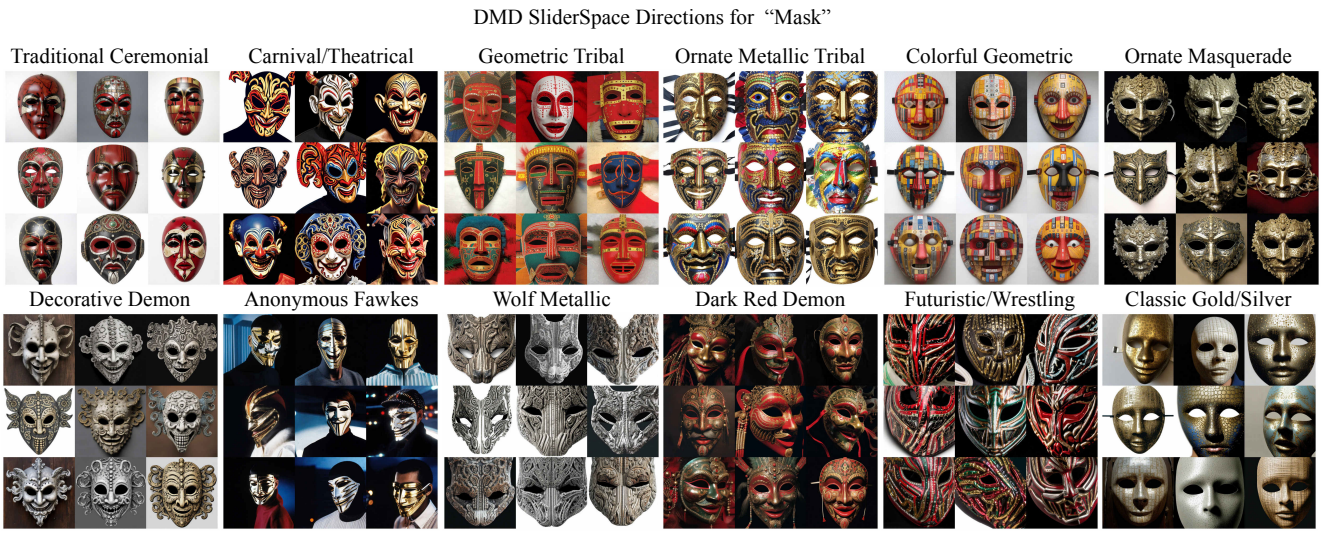

Figure E.18. We show the SliderSpace discovered in SDXL-DMD2 4-step model [51] for the concept "Mask"

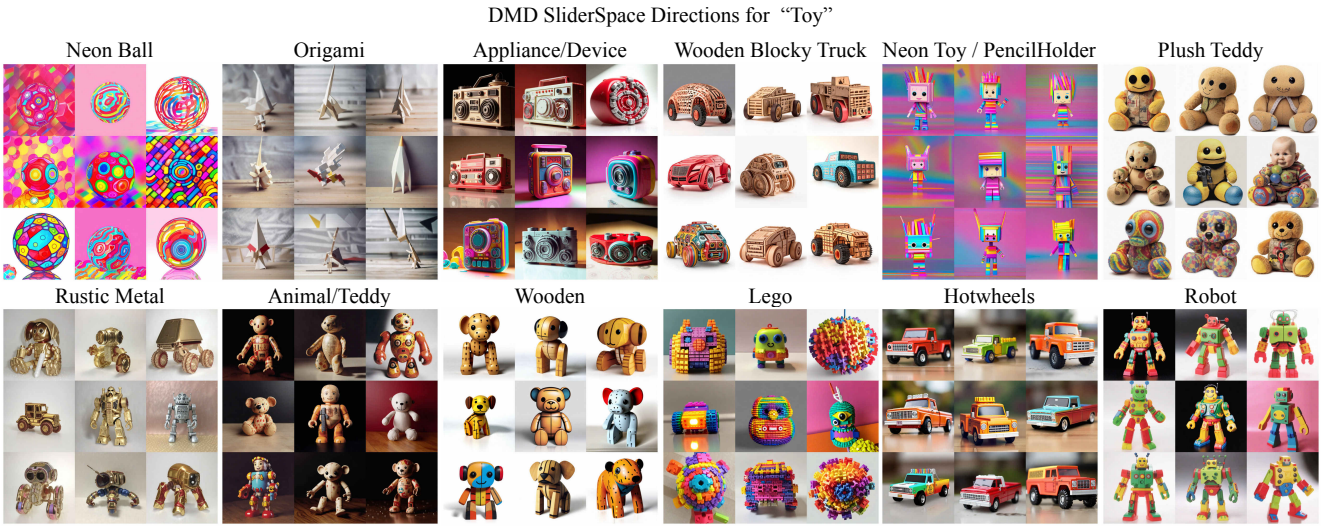

Figure E.19. We show the SliderSpace discovered in SDXL-DMD2 4-step model [51] for the concept "Toy"

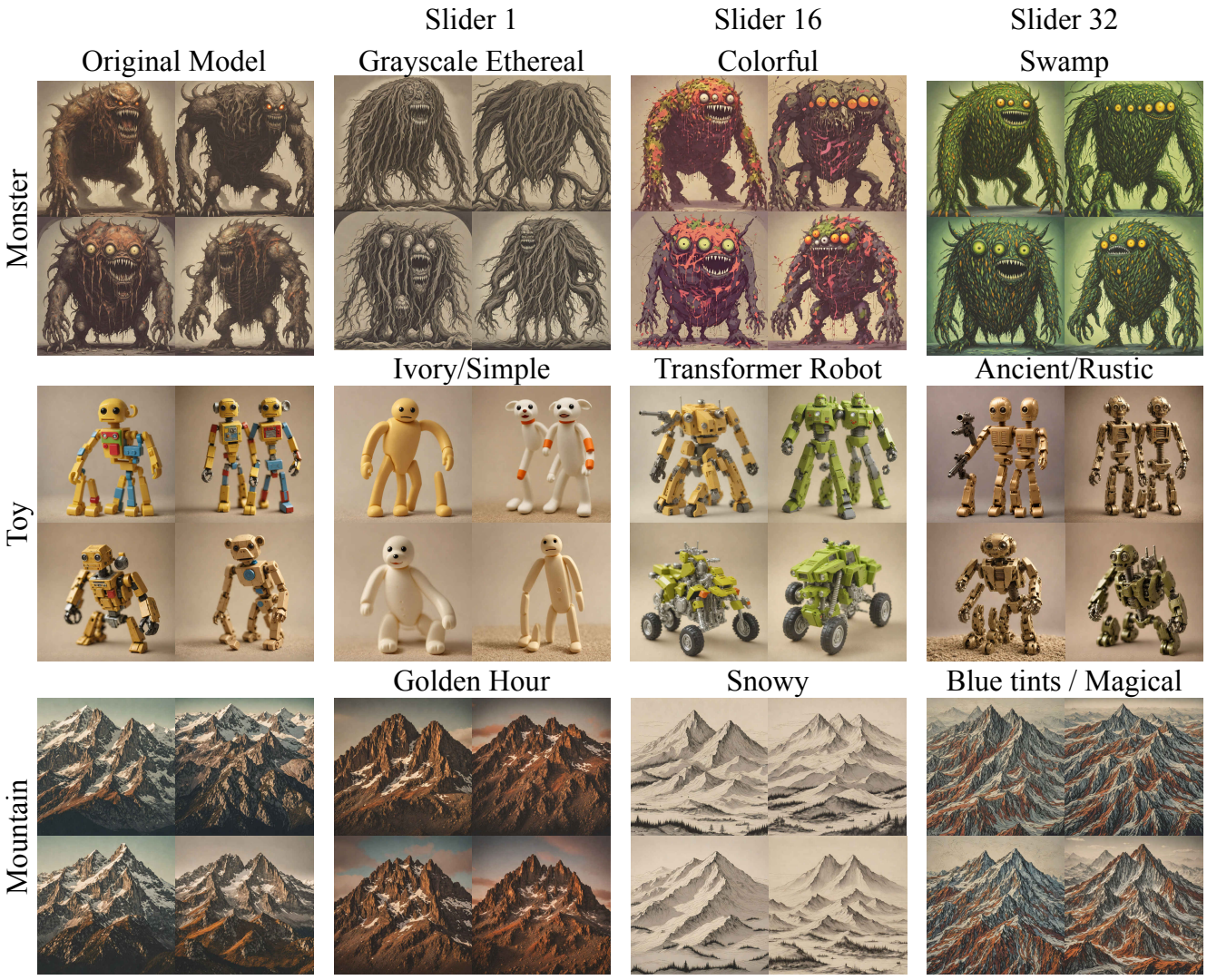

Figure E.20. We show the SliderSpace discovered in SDXL-Turbo 4-step model [43] and how they can be used for precise control of image generation

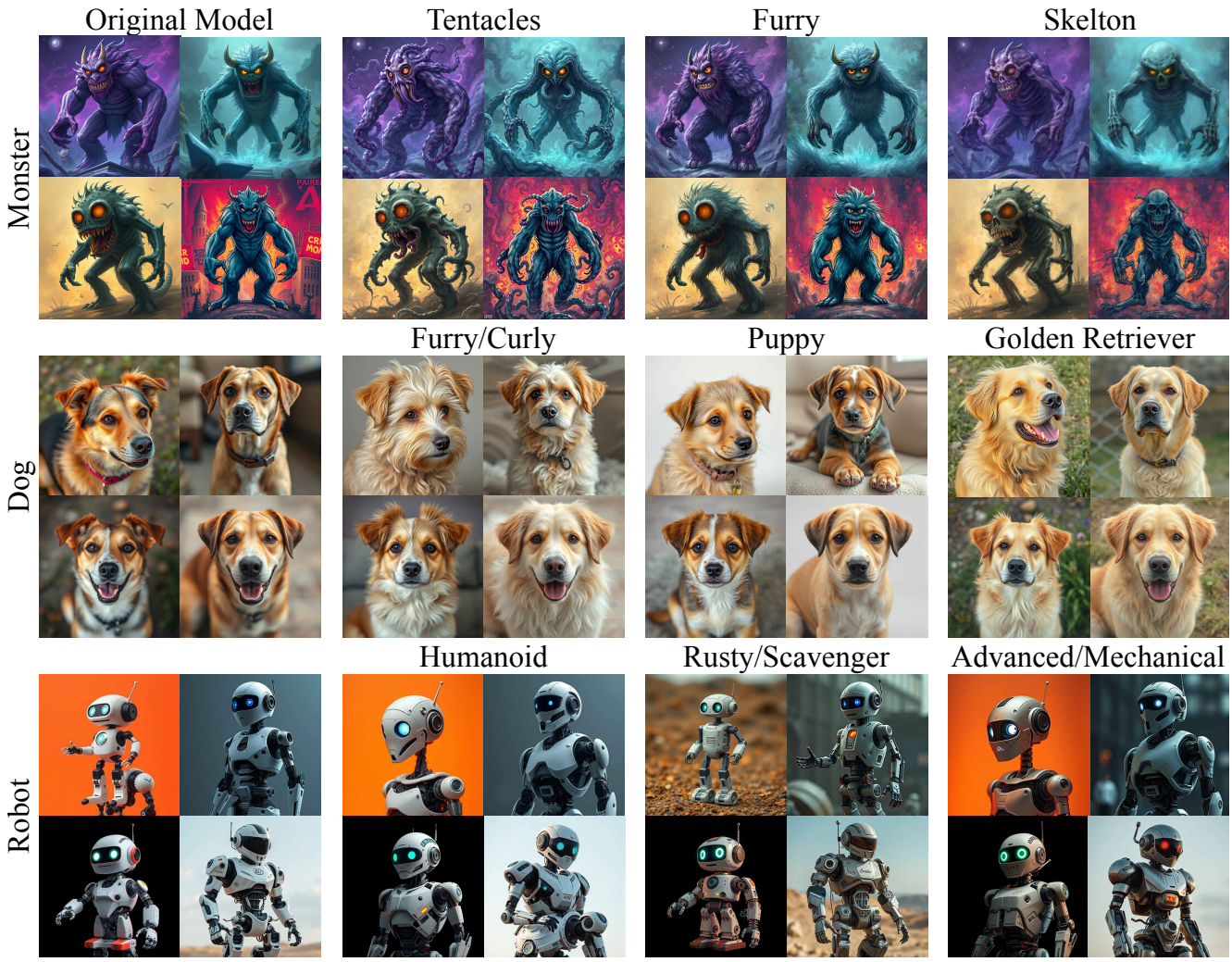

Figure E.21. We show the SliderSpace discovered in FLUX Schnell model [4] for concepts “monster” and “dog”