

Breaking Semantic Boundaries: Distribution-Guided Semantic Exploration for Creative Generation

Supplementary Material

A. Additional Results

A.1. Additional Results on Conditional Exploration Guided by Class Distributions

We provide additional qualitative results to evaluate DisTok’s capacity for fine-grained composition under distribution-conditional generation. As shown in Fig. 10, DisTok successfully synthesizes coherent and visually unified concepts across diverse class distributions, including cases with more than three classes and imbalanced ratios.

Notably, DisTok preserves the target class proportions with high fidelity: dominant classes manifest through salient visual features, while minor classes contribute subtle but recognizable details. In contrast to baseline diffusion models, which often produce fragmented or spatially disjoint outputs (Fig. 3), DisTok achieves seamless fusion across concepts, even under complex distributional conditions. These results underscore the expressiveness of DisTok’s latent space and its robustness in modeling intricate, distribution-guided visual compositions.

A.2. Additional Results on Conditional Exploration Guided by Class Pairs

Fig. 11 presents additional qualitative results generated by DisTok, where each example is conditioned on a concept pair, represented either as two discrete classes or as an equivalent class distribution (e.g., (50% Ant, 50% Dog)). These visualizations illustrate DisTok’s ability to synthesize coherent hybrid concepts that reflect integrated semantics rather than simple spatial composition or surface-level blending.

DisTok consistently produces unified and semantically coherent outputs under both balanced and imbalanced class distributions, preserving fine-grained attributes from each source concept. In comparison, search-based approaches like BASS [17] often suffer from significant computational overhead due to iterative sampling or optimization and struggle to maintain semantic balance—frequently resulting in dominant features or incomplete compositions (Fig. 4). These shortcomings limit their scalability and applicability in practical settings. In contrast, DisTok enables efficient one-shot inference with strong compositional generalization, highlighting its effectiveness for controllable and scalable creative synthesis from class-conditioned inputs.

A.3. Additional Results on Unconditional Exploration

To further evaluate DisTok’s ability to support open-ended, unconditional generation, we present additional qualitative results obtained via direct latent sampling without any class-based conditioning. Latent vectors are sampled from multiple zero-mean, unit-variance distributions—including Gaussian, Cauchy, Laplace, and Uniform—and decoded into creative concept tokens. As shown in Fig. 12, DisTok produces diverse and semantically meaningful outputs across all distributions, highlighting the structure and generality of its latent space.

The generated concepts display substantial morphological and semantic diversity, demonstrating DisTok’s ability to synthesize novel visual forms beyond predefined prompts or control conditions. Unlike conditional generation paradigms grounded in text pairs or class distributions, which inherently constrain exploration to user-specified inputs, DisTok enables fully open-ended synthesis. These results highlight its potential as a scalable and autonomous framework for unconstrained concept discovery.

A.4. Style-Aware Adaptability of Creative Concepts

We assess the adaptability of DisTok-generated concepts to diverse visual styles by integrating creative tokens into natural language prompts that specify stylistic variations, such as watercolor painting, line drawing, and others (Fig. 13).

Across all styles, the generated images retain the semantic structure and visual identity of the original concepts, despite substantial changes in appearance induced by stylistic prompts. This consistency indicates that DisTok produces concept tokens that are semantically disentangled from style and can generalize across varied visual domains. Such invariance to stylistic prompts highlights DisTok’s versatility in real-world creative workflows, including animation, game asset generation, and visual storytelling. It enables scalable and consistent concept deployment across heterogeneous visual settings, supporting controllable style adaptation while preserving conceptual fidelity.

B. More Details on Evaluation

B.1. Prompts Used for GPT-4o Evaluation

We conduct an objective evaluation of the creativity of images generated by DisTok and other methods using GPT-4o, assessing four key dimensions: Conceptual Integration, Alignment with Prompt, Originality, and Aesthetic Quality.

The detailed evaluation prompts provided to GPT-4o are as follows, inspired by those proposed in CreTok [9]:

The subject of this evaluation is an image that represents a mixture of a Alpaca and a Bee (or a Zebra (34%), a Moth (23%), a Dog (20%) and a Chicken (23%)). The objective is to assess the creativity of an entity that synthesizes two distinct concepts as delineated in the provided prompt. Accordingly, please evaluate the creativity of images generated by various methodologies for the identical prompt, utilizing the following criteria on a scale from 1 to 10:

1. *Conceptual Integration (1-10): This criterion gauges the degree to which the image manifests a coherent and integrated concept, as opposed to merely placing two independent elements side by side. A high score signifies that the elements are intricately merged, creating a new, unified entity.*

2. *Alignment with Prompt (1-10): This evaluates the extent to which the image conforms to and encapsulates the specific combination of concepts described in the prompt. The image should refrain from including irrelevant elements that detract from the primary concepts. A high score is allocated when the image closely adheres to the specifications of the prompt.*

3. *Originality (1-10): This assesses the innovativeness of the concept portrayed in the image. The depicted concept should not mimic existing animals, plants, or widely recognized mythical creatures unless specifically mentioned in the prompt. Images that present a distinctive and inventive amalgamation receive a high score.*

4. *Aesthetic Quality (1-10): This criterion scrutinizes the visual appeal of the image, focusing on color harmony, the balance and arrangement of elements, and the overall visual impact. A high score is awarded when the image is not only conceptually robust but also visually engaging.*

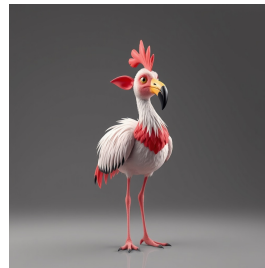
In conclusion, based on the aforementioned criteria, provide a comprehensive creative assessment (1-10) for each image and articulate specific justifications for your rating.

B.2. More Details on User Study

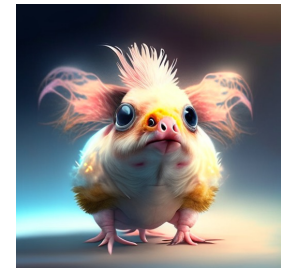
To ensure a comprehensive and balanced assessment of visual creativity, our user study includes 100 participants drawn from a diverse range of professional and non-professional backgrounds in art, design, and vision-related fields. This diversity allows us to evaluate whether the generated concepts satisfy both expert-level criteria and general audience preferences. The study interface is shown in Fig. 9.

For each question, you will be shown a pair of images. Please select the one you consider more creative, based on your personal and intuitive understanding of creativity. There is no single correct answer—choose the image that you feel best reflects novelty, originality, and conceptual integration.

*25. (0.2 Flamingo, 0.25 Chicken, 0.55 Pig)



69612



ddb44

Figure 9. Interface of the User Study.

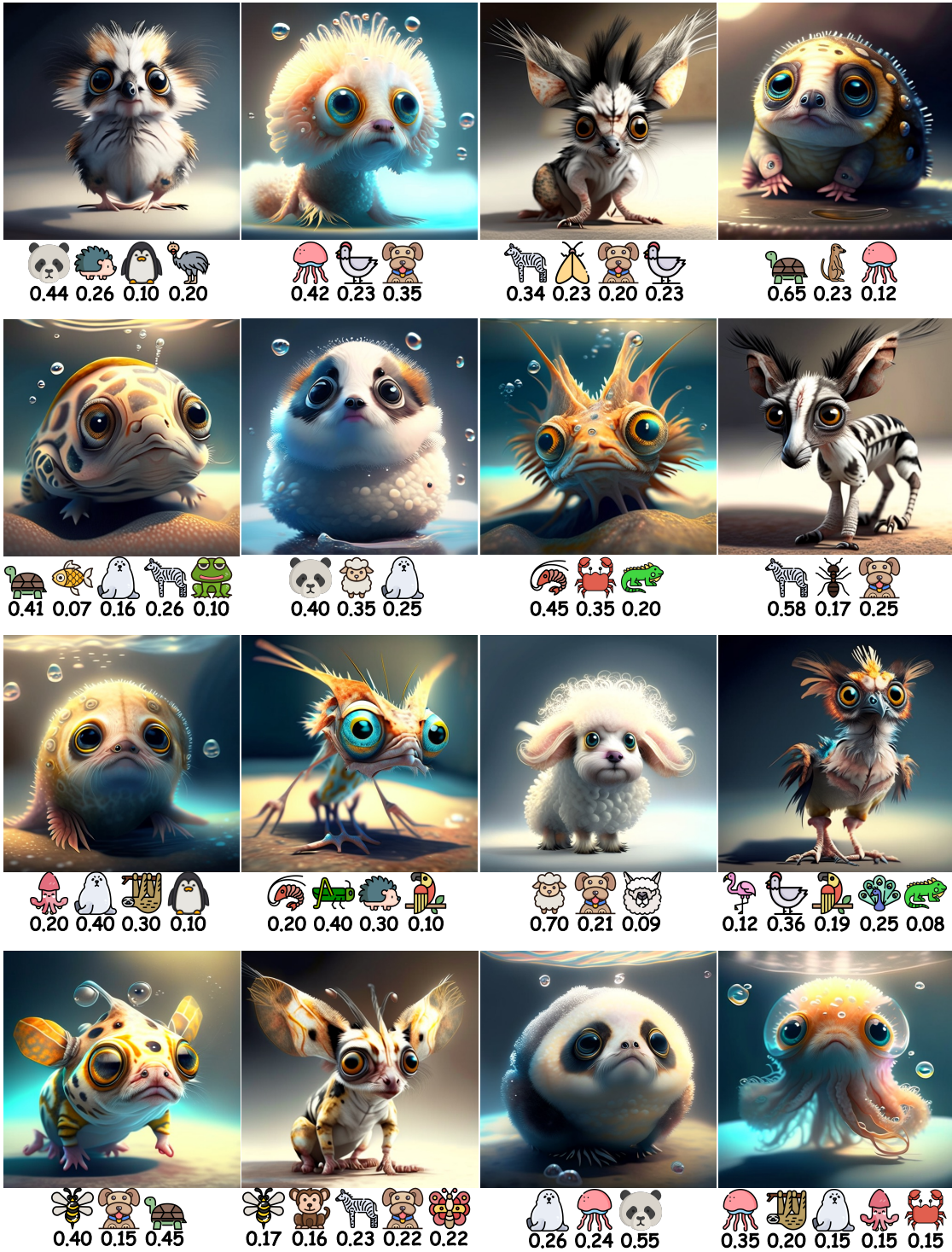


Figure 10. Additional creative concepts generated by DisTok on semantic exploration guided by class distributions.



Figure 11. Additional creative concepts generated by DisTok on semantic exploration guided by class pairs.

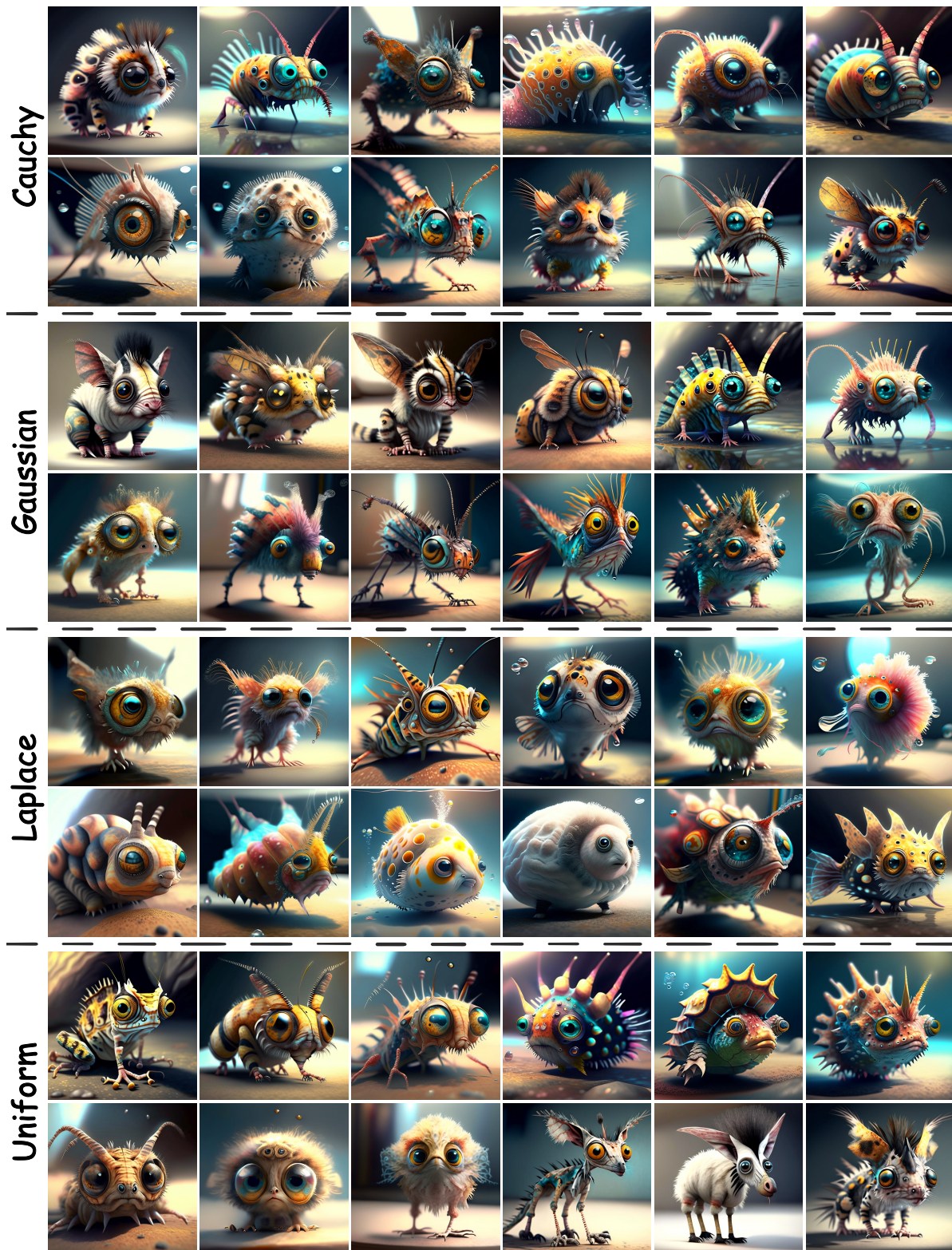


Figure 12. Creative concepts generated unconditionally by DisTok through latent vector sampling from Cauchy, Gaussian, Laplace, and Uniform distributions.

Professional high-quality art of a full-body <crit>, photorealistic rendering, cartoonish yet realistic style, centered composition, 4k, HQ



A watercolor painting of a <crit>.



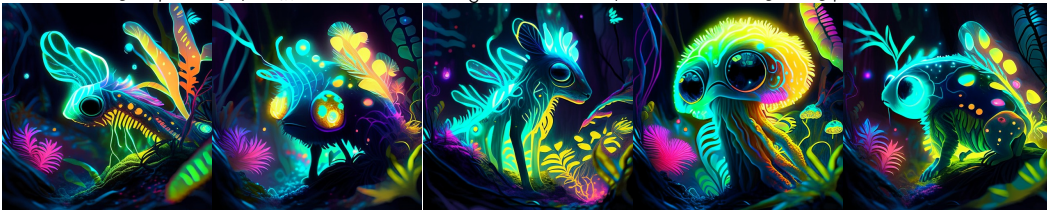
A line drawing of a <crit>.



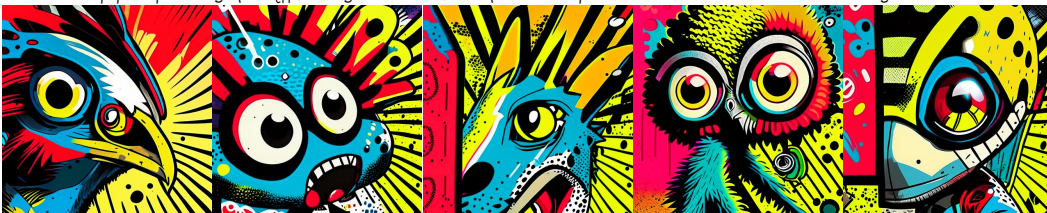
A vibrant, surrealist photo of a <crit>, dreamlike forest background with soft, pastel colors and a touch of fantasy art style.



A neon-glow painting of a <crit>, set in an otherworldly, bioluminescent forest with vibrant, glowing plants and creatures.



A pop art painting of a <crit>, set against a bold, colorful backdrop with comic book elements and striking contrasts.



A glitch art image of a <crit>, set in a distorted, digital landscape with fragmented visuals and vibrant, pixelated colors.

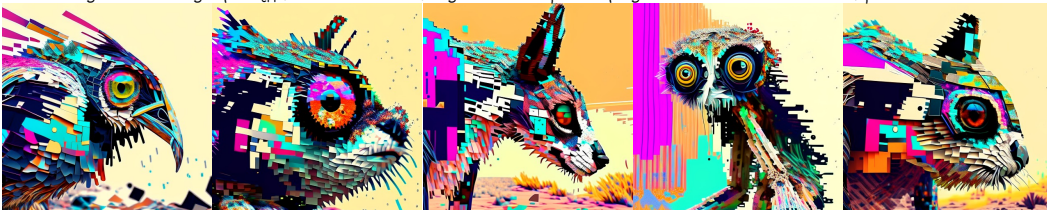


Figure 13. Creative concepts rendered in diverse visual styles by integrating tokens generated by DisTok with style-specific natural language prompts.

DisTok	Prompt Engineering	
		A fuzzy, insectoid-mammal hybrid with striped yellow-black fur, oversized teal eyes, large translucent ears, and tufted antennae; depicted in a soft, cartoon-realistic style with warm ambient lighting and fine fur texture.
		A fantastical deep-sea creature with multiple large golden eyes, glowing antennae, striped blue-orange exoskeleton, and soft spines; rendered in a surreal, semi-photorealistic cartoon style with warm lighting and sandy seabed setting.
		A surreal insect-like creature with fish scales, oversized yellow eyes, antennae, thin crab-like legs, and vibrant orange-blue textures; stylized in a semi-realistic, cartoonish render with sharp lighting and indoor background.
		A fluffy, bird-like creature with spindly legs, oversized blue-yellow eyes, messy white and brown feathers, and a tiny beak; rendered in a stylized photorealistic cartoon style with soft focus and warm indoor lighting.
		A tiny, wide-eyed hybrid creature with oversized orange ears, spiky blue fur, delicate limbs, and a mouse-like body; stylized in a hyper-detailed, cartoon-realistic blend with soft lighting and warm tones.
		A round, pufferfish-like creature with a soft, plush texture, large blue-ringed eyes, light peach spots, translucent fins, and a cartoonish yet photorealistic underwater style, lit from above.
		A round, fuzzy insect-mammal hybrid with striped legs, glossy black eyes, colorful peacock-like tail spots, and tufts of fur; depicted in a soft, stylized cartoon-realistic render with muted tones and focused lighting.
		A whimsical amphibian-insect hybrid with large bulging eyes, striped limbs, spiny back, and textured yellow-orange skin; rendered in a semi-realistic cartoon style with sharp detailing and soft, cave-like lighting.
		A lanky, moth-like creature with oversized amber eyes, long patterned wings, thin limbs, soft white fur, and blue-tinted chitin; illustrated in a whimsical cartoon-realistic style with warm indoor lighting and fine detail.

Figure 14. Detailed prompts generated by GPT-4o for each DisTok-generated concept, which are subsequently used for image synthesis with Kandinsky.