



# DreamBeast: Distilling 3D Fantastic Animals with Part-Aware Knowledge Transfer

## Supplementary Material

$\alpha_{self}$	$\alpha_{cross}$	CLIP Score $\uparrow$		
		B/32	B/16	L/14
0.8	0.6	$0.286 \pm 2.3e^{-4}$	$0.288 \pm 3.7e^{-4}$	$0.246 \pm 3.9e^{-4}$
	0.9	$0.285 \pm 2.6e^{-4}$	$0.289 \pm 3.5e^{-4}$	$0.245 \pm 4.6e^{-4}$
	1.2	$0.284 \pm 3.2e^{-4}$	$0.288 \pm 4.5e^{-4}$	$0.242 \pm 5.9e^{-4}$
1.2	0.6	$0.284 \pm 4.2e^{-4}$	$0.289 \pm 3.7e^{-4}$	$0.244 \pm 5.3e^{-4}$
	0.9	$0.283 \pm 2.7e^{-4}$	$0.288 \pm 4.7e^{-4}$	$0.243 \pm 4.6e^{-4}$
	1.2	$0.284 \pm 2.6e^{-4}$	$0.288 \pm 3.8e^{-4}$	$0.244 \pm 4.4e^{-4}$

Table 1. Performance comparison of different attention modulation hyper-parameters.

## Appendix

This supplementary material contains additional ablations (Appendix A), more detailed quantitative results (Appendix B), non-animal part-aware asset generated by DreamBeast compared to MVDream [2] (Appendix C), running cost breakdown (Appendix D), more qualitative results of fantastic animals (Appendix F), failure case analysis (G), and more implementation details (Appendix H).

### A. Additional Ablation Study

We also conducted an ablation study on the cross-attention modulation factor ( $\alpha_{cross}$ ) and self-attention modulation factor ( $\alpha_{self}$ ), as detailed in Table 1. The performance of DreamBeast remains stable across a wide range of values for the modulation factors.

### B. Detailed Human Study Breakdown

We present more detailed statistics of the results in Figure 5, showing the trend that human evaluators consistently prefer the results generated by DreamBeast.

### C. Non-animal Part-aware Asset Generation

While our main manuscript focuses on the generation of 3D fantastical beasts, we also observed that our model performs exceptionally well with non-animal, part-specific 3D assets. As demonstrated in Figures 1, 2, and 3, DreamBeast continues to excel in generating part-aware 3D non-animal assets, whereas MVDream [2] struggles with this task. We hope that our framework can be extended to more general applications, which we leave for future exploration.

### D. Detailed Running Speed Breakdown

The part-affinity map extraction process takes 41.84 seconds for a single view. The Part-affinity NeRF requires just 0.06 seconds per optimization step, and the attention-modulated SDS process takes 0.27 seconds per step. Although the part-affinity map extraction is time-consuming, it is still more efficient and significantly faster than directly applying SDS + SD3.

### E. Learned Part-affinity NeRF

We include 10 videos of the rendered Part-affinity NeRF in the folder.

### F. More Qualitative Results

We demonstrate more qualitative results in Figure 6. All the results show that DreamBeast is able to generate part-aware 3D animal asset. We also include 12 rendered videos of RGB, normal map, and opacity map in the folder.

### G. Failure Case Analysis

There are also instances where DreamBeast fails to produce the expected results. The first type of failure occurs when the part-affinity map is misplaced, causing the body parts to be incorrectly positioned (as shown in the upper left of Figure 7). The second type of failure happens when two body parts are too similar, making it difficult for DreamBeast to distinguish between them. For instance, in the upper-right example of Figure 7, the terms “body” and “trunk” are similar, leading to the generated result having a mix of animal features. Additionally, the model sometimes misinterprets parts semantically, as seen in Figure 4, where “white gun” in the prompt is generated as a “black gun.”

### H. More Implementation Details

We chose GPT-4o-mini as the Large Language Model (LLM) to extract part-specific prompts from the original global prompt. Additionally, we implemented a part-specific prompt checker to verify that the tokens of the extracted part-specific prompts are also tokens of the original global prompt to prevent hallucinated body part prompts. However, users also have the option to input part-specific prompts manually, making GPT-4o-mini an optional component in our pipeline. For the part affinity map extractor,

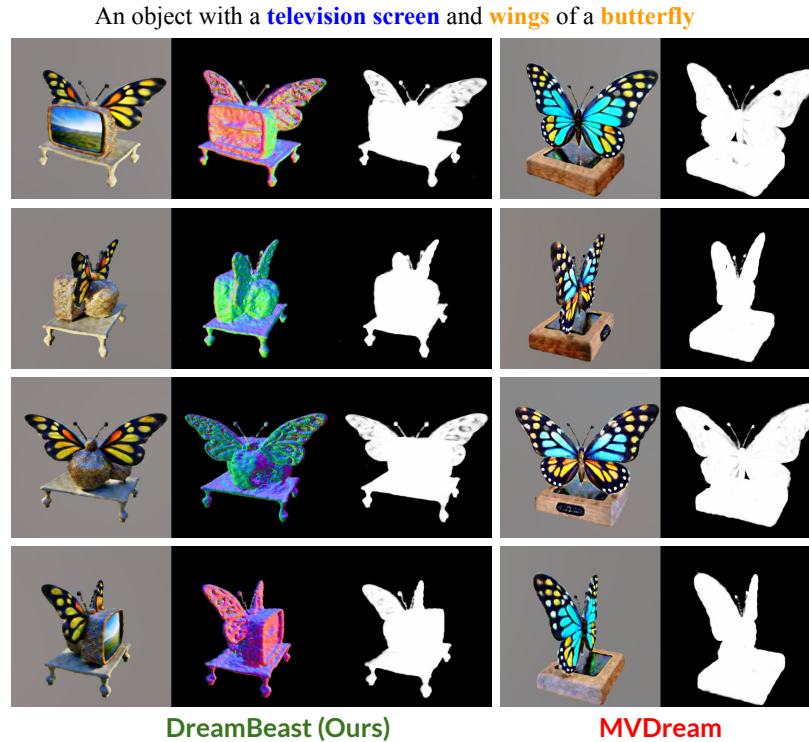


Figure 1. Non-animal result generated by DreamBeast

we employed Stable Diffusion 3 medium [1]. This cross-attention map operates in the latent space at a resolution of  $H = 128, W = 128$ .

The Part-affinity NeRF is represented by an MLP with a single hidden layer of 64 neurons, and we process 128 samples per ray during rendering. We employ the AdamW optimizer with a learning rate of 0.001, carrying out the optimization over 5000 steps. We use Google Form to make our human study.

## References

- [1] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, Kyle Lacey, Alex Goodwin, Yannik Marek, and Robin Rombach. Scaling rectified flow transformers for high-resolution image synthesis, 2024. 2
- [2] Yichun Shi, Peng Wang, Jianglong Ye, Mai Long, Kejie Li, and Xiao Yang. Mvdream: Multi-view diffusion for 3d generation, 2024. 1

A person in a **red lolita dress**, wearing a **yellow cowboy hat**



Figure 2. Non-animal result generated by DreamBeast

A **car** with **airplane wings**

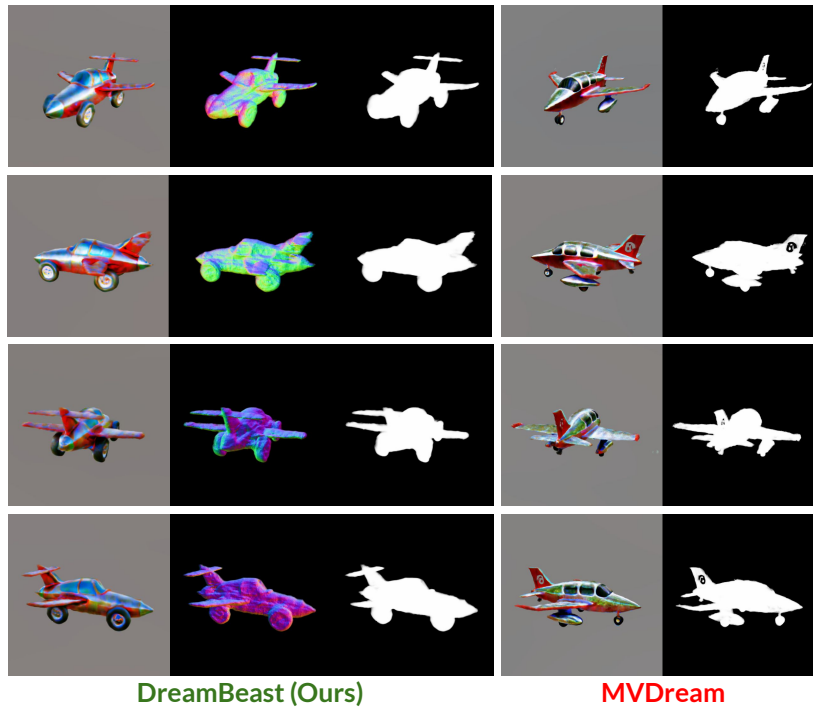


Figure 3. Non-animal result generated by DreamBeast

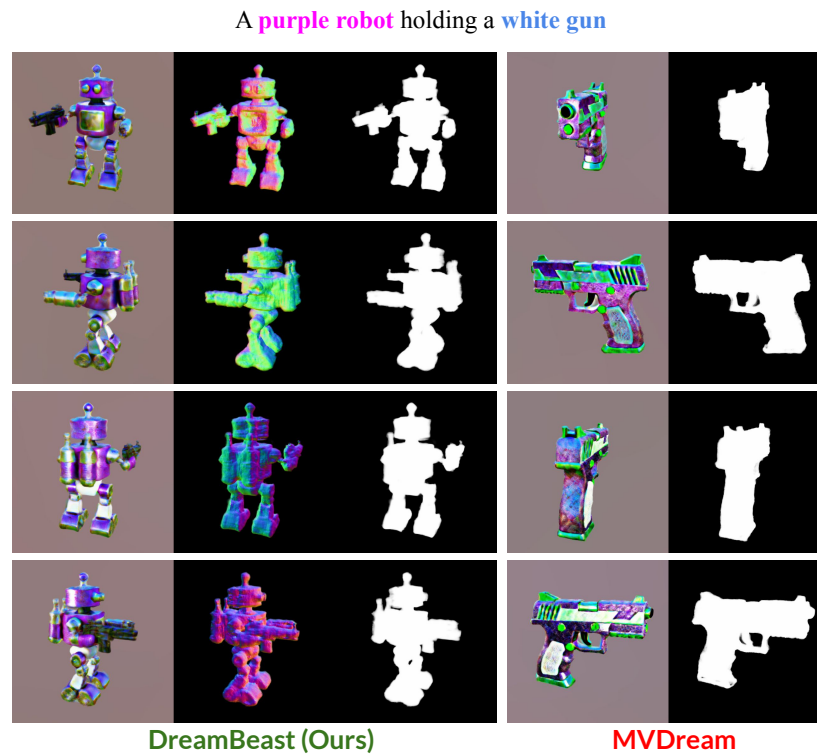


Figure 4. Non-animal result generated by DreamBeast

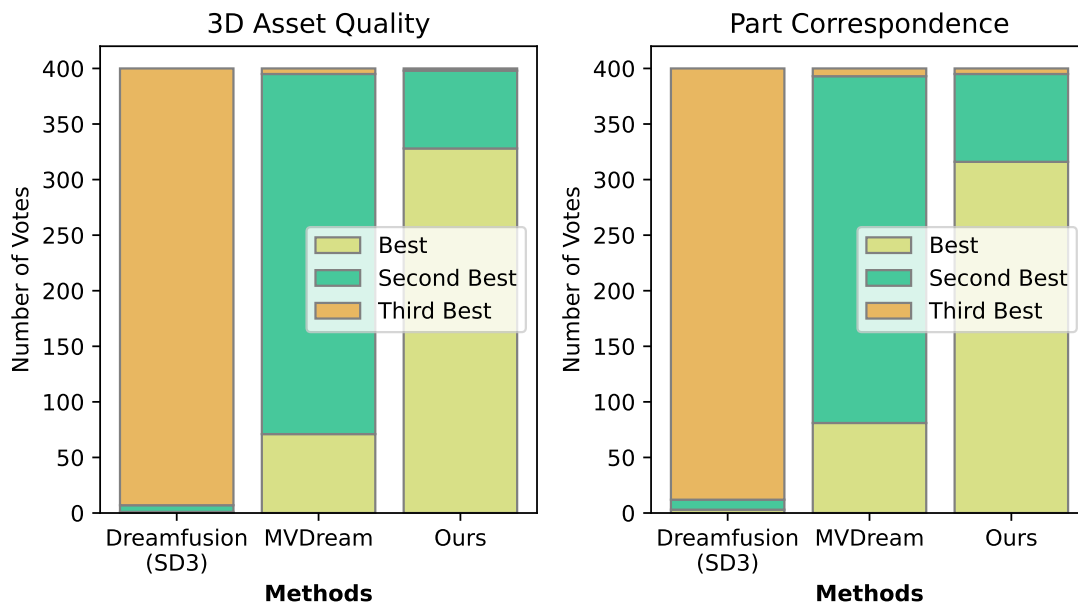
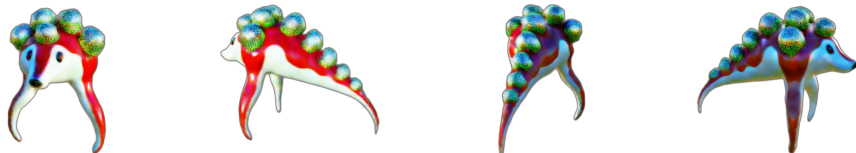


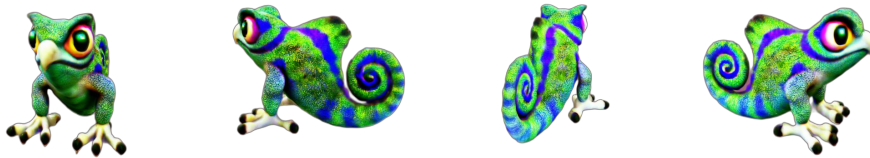
Figure 5. Human study results in more detail.



An creature with a **body** of a **guinea pig** and a **tail** of a **scorpion**



An creature with a **head** of a **fox** and a **tentacles** of a **jellyfish**



An creature with a **body** of a **chameleon** and a **eyes** of an **owl**



An creature with a **body** of a **newt** and a **beak** of a **toucan**



An creature with a **body** of a **monkey** and a **wings** of a **bat** and **snout** of a **pig**



An creature with a **body** of a **salamander** and a **head** of a **kangaroo**

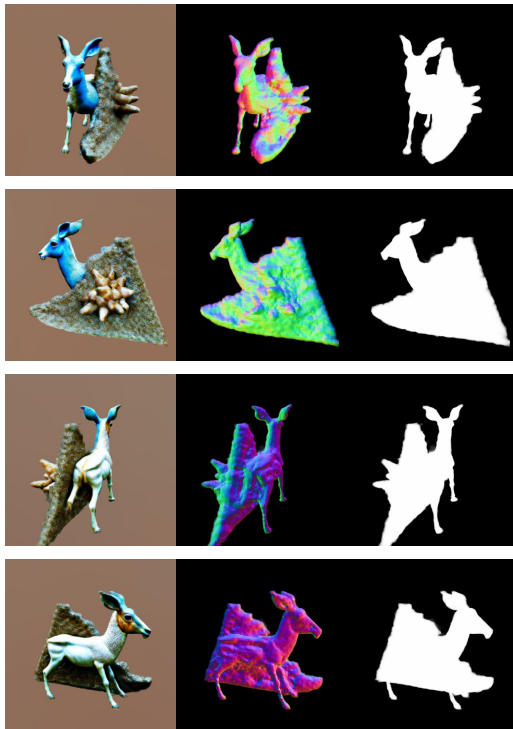


An creature with a **body** of a **seal** and a **mane** of a **lion** and **fins** of a **goldfish**



An creature with a **body** of a **cheetah** and a **head** of a **dodo**

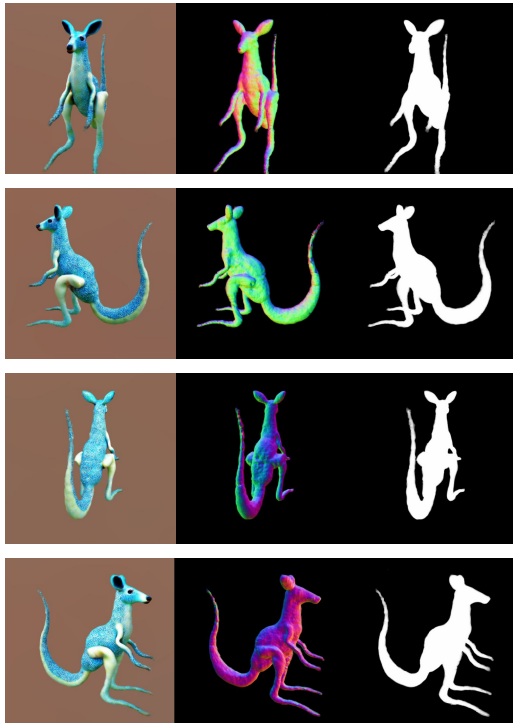
A creature with a **body** of a **gazelle** and the **shell** of a **barnacle**



A creature with a **body** of a **hawk** and the **trunk** of a **aardvark**



A creature with a **body** of a **kangaroo** and the **tentacles** of a **jellyfish**



A creature with a **body** of a **buffalo** and the **beak** of a **duck** and **claws** of a **lobster**

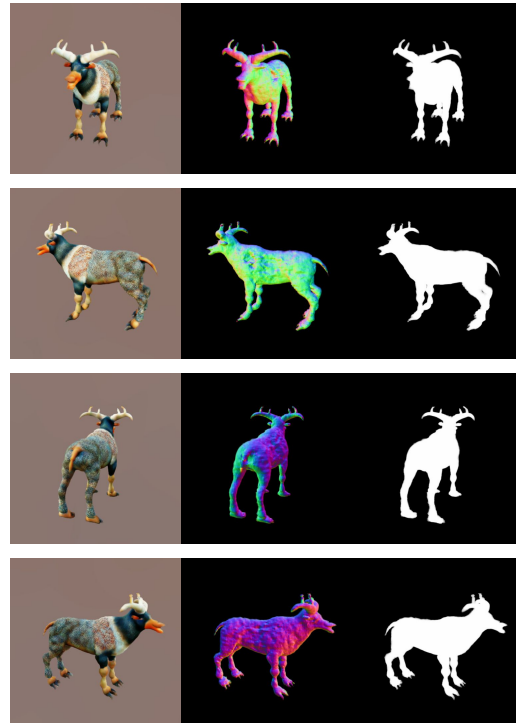


Figure 7. Failure case generated by DreamBeast