Supplemental Material: DreamOmni: Unified Image Generation and Editing

Bin Xia [1], Yuechen Zhang [1], Jingyao Li[1], Chengyao Wang [1],
Yitong Wang[2], Xinglong Wu[2], Bei Yu[1], and Jiaya Jia [3]
[1] CUHK, [2] ByteDance Inc, [3] HKUST
https://zj-binxia.github.io/DreamOmni-ProjectPage/

The overview of the supplementary materials:

(1) We discuss current limitations of DreamOmni and provide future improvements (Sec. 1).

(2) We provide a detailed introduction to our synthetic collage data pipeline (Sec. 2).

(3) We provide more T2I visual results of DreamOmni (Sec. 3).

(4) We provide more Inpainting & Outpainting visual results of DreamOmni (Sec. 4).

(5) We provide more reference Image generation visual results of DreamOmni, including image-condition generation and subject-driven generation (Sec. 5).

(6) We provide more instruction-based editing visual results of DreamOmni (Sec. 6).

(7) We provide more drag editing visual results of DreamOmni (Sec. 7).

(8) We provide more segmentation & detection visual results of DreamOmni (Sec. 8).

1. Limitations and Future Work

As shown in Fig. 1, we provide multiple generation results of DreamOmni. Although our DreamOmni performs excellently in terms of overall composition, it sometimes struggles with generating certain fine details, such as the hands, positioning of legs, and other small elements, leading to occasional generation errors.

For future work, we can make improvements from several aspects: **(1)** Collect feedback data and introduce RLHF training to further enhance the model's ability to generate fine-grained details. **(2)** Incorporate more tasks, such as low-level tasks and in-context learning tasks, to further explore DreamOmni's capabilities as a foundational vision model. **(3)** Introduce knowledge distillation to reduce the number of model iterations, resulting in more timely generation and a better user experience.

2. Synthetic Collage Data Pipeline

In this section, we provide additional details about the our synthetic collage data pipeline introduced in the main paper. As shown in Fig. 2 (a), our approach uses an arbitrary number of matting images or stickers, arranging them randomly over various background images to efficiently generate diverse and accurate synthetic source and target images. Furthermore, as illustrated in Fig. 2 (b), this pipeline is leveraged to synthesize diverse editing data as well as T2I generation-augmented data. The efficiency of our pipeline enables the creation of large-scale, high-quality, accurate, and diverse datasets, supporting both scaling up pretraining and precise fine-tuning tasks.

3. More Results on T2I Generation

We provide additional T2I generation results in Figs. 3 and 4. Our DreamOmni can generate images in different styles and various aspect ratios.

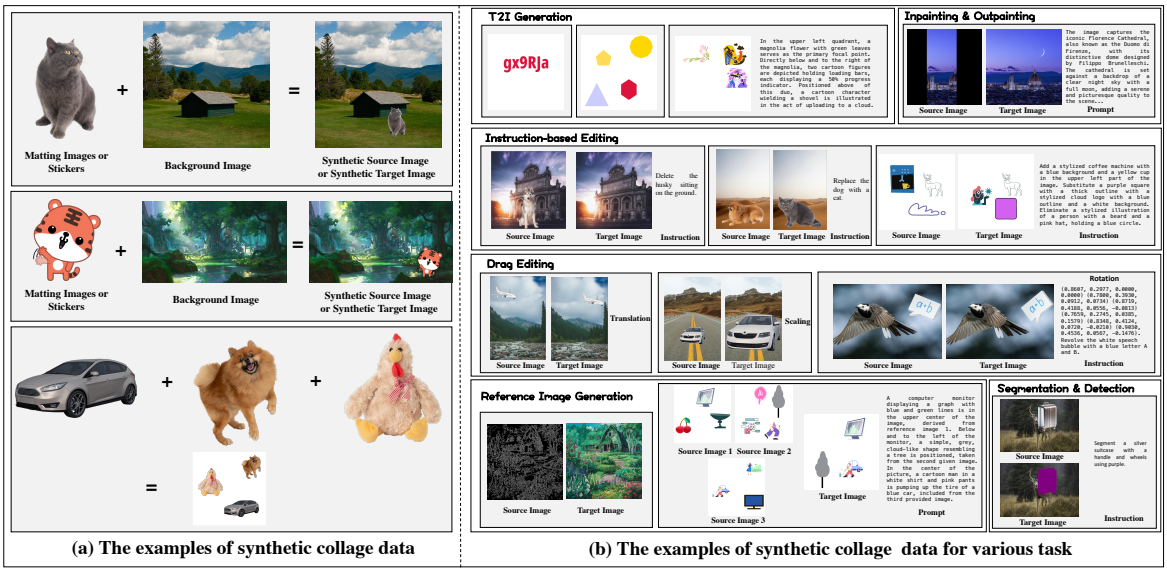
4. More Results on Inpainting & Outpainting

We provide additional inpainting & outpainting results in Fig. 5. Our DreamOmni can intelligently regenerate the areas of the image that are masked.

5. More Results on Reference Image Generation

We show additional image-conditioned generation results in Fig. 6. Moreover, we present additional subject-driven generation results in Fig. 7. For image-conditioned generation, our DreamOmni effectively adheres to both the image condition and the prompt, while simultaneously generating visually pleasing results. For subject-driven generation, our DreamOmni excels at accurately following user prompts while preserving the specified subject.

Figure 1. The limitations of DreamOmni. We provide multiple generation results of DreamOmni to identify some of its limitations. We can observe that, although DreamOmni performs well in terms of overall composition, it sometimes struggles with generating certain fine details, such as the hands, positioning of legs, and other small elements, leading to occasional generation errors.



(a) The examples of synthetic collage data **(b) The examples of synthetic collage data for various task**

Figure 2. The synthetic collage data pipeline. **(a)** We show how to use stickers and matting images to synthesize collage source images or target images for various editing and T2I generation tasks. **(b)** To overcome the difficulty and inefficiency in data creation and filtering for image editing, we propose a collage-based synthetic data pipeline. This pipeline enables the efficient creation of data for various editing tasks, such as adding, deleting, and replacement operations in instruction-based editing, as well as translation, scaling, and rotation in drag editing. Additionally, it supports reference image generation and segmentation & detection. Furthermore, our synthetic data pipeline enhances the accuracy of T2I generation, particularly for attributes related to text, geometry, color, position, and quantity. Due to space limitations, we have optionally shown the corresponding prompts or instructions for these cases.

6. More Results on Instruction-based Editing

We show additional instruction-based editing results in Fig. 8. Our DreamOmni can perform precise editing (including addition, removal, and replacement).

7. More Results on Drag Editing

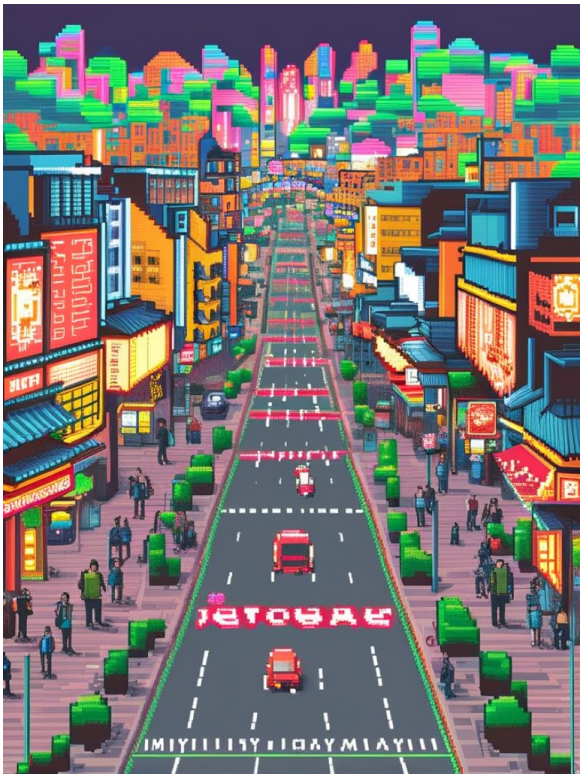
We show additional drag editing results in Fig. 9. DreamOmni accurately performs translation, rotation, and scaling edits.

8. More Results on Segmentation & Detection

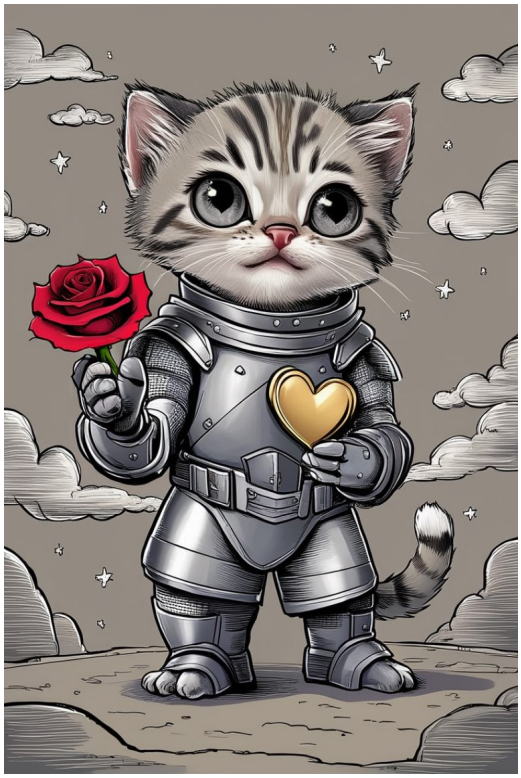
We show additional segmentation & detection results in Fig. 10. DreamOmni's segmentation & detection can actually be considered a subtask of instruction-based editing. We can see that DreamOmni accurately identifies the required objects and marks them with the specified colors.

An extreme close-up of an gray-haired man with a beard in his 60s, he is deep in thought pondering the history of the universe as he sits at a cafe in Paris, his eyes focus on people offscreen as they walk as he sits mostly motionless, he is dressed in a wool coat suit coat with a button-down shirt , he wears a brown beret and glasses and has a very professorial appearance.

A portrait of a delicate Elegant stylized female fatale, red lips, figure. She has striking blue eyes, full lips, and her face is painted with a mix of colors, predominantly reds and oranges. Her hair is depicted with bold strokes, and there are abstract elements, like splashes of paint and brush strokes, surrounding her. The overall style of the artwork is reminiscent of contemporary abstract portraiture, blending realistic features with abstract elements., painting, anime, conceptual art, ukiyo-e, dark fantasy



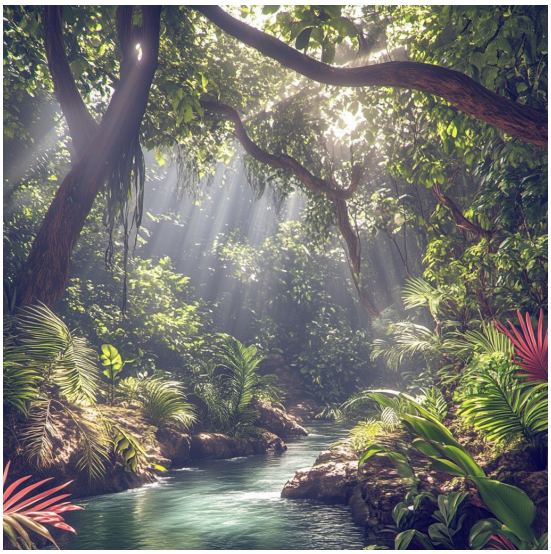
A pixel art landscape of Motoyama, Kochi at night, showcasing the bustling city streets and illuminated buildings. The vivid colors and detailed pixel art create an immersive atmosphere of the city's after-hours energy. The streets are filled with pedestrians, vehicles, and neon signs, giving a lively and dynamic impression of the district.



An adorable kitten . The character, with large round eyes and a small head, exudes cuteness and innocence. . Clad in an armored suit with protective parts, the character holds a red rose in its right hand, and a golden heart in the other hand, offering it to the viewer. The minimalistic background of floating clouds and stars immerses the viewer in an outer space setting, emphasizing the character's role as a cosmic explorer or protector.

Figure 3. More T2I generation visual results. Ours DreamOmni generate images in different styles.

In a dynamic tribute to Van Gogh's iconic style, imagine an oil painting bursting with love and colour. A canvas breathing with bold strokes and textured applications, inviting you into a world where the essence of twilight dances magic. In the foreground is a scene overlooking a tranquil lake. It contains a small Canadian country cottage with a wooden door. The cottage is surrounded by a lush field of wildflowers in different colors, and the background shows a calm blue lake. The sky should be bright, clear blue.



a mystical rainforest with exotic plants and trees, with sunlight filtering through the canopy, creating dappled light on a serene river flowing gently in its center. the scene is captured from an eye level perspective, highlighting rich textures of leaves and vines, creating a magical atmosphere. digital art, rendered using unreal engine for realistic lighting effects and color palette

A stunning underwater photograph capturing a sea turtle gracefully swimming near the ocean floor. The turtle's shell is a beautiful blend of greens and browns, perfectly camouflaging it among the coral. Sunlight filters through the water, casting dappled shadows and creating a serene ambiance. The background reveals a rich array of marine life, including colorful fish and swaying seagrass. The overall atmosphere of the scene is calm and tranquil, highlighting the beauty and fragility of the underwater world.

A cinematic illustration of a fantastical bird made of multi-colored glass shards. The bird is perched on a delicate piece of paper, creating a vivid splash of red, orange, yellow, green, blue, and purple hues. The background features a magical forest with shimmering leaves and vines, while a soft, ethereal glow emanates from the scene. The overall atmosphere is dreamy and otherworldly, capturing the essence of a fantastical glass-colored bird painting.

Figure 4. More T2I generation visual results. Ours DreamOmni generate images in different styles.

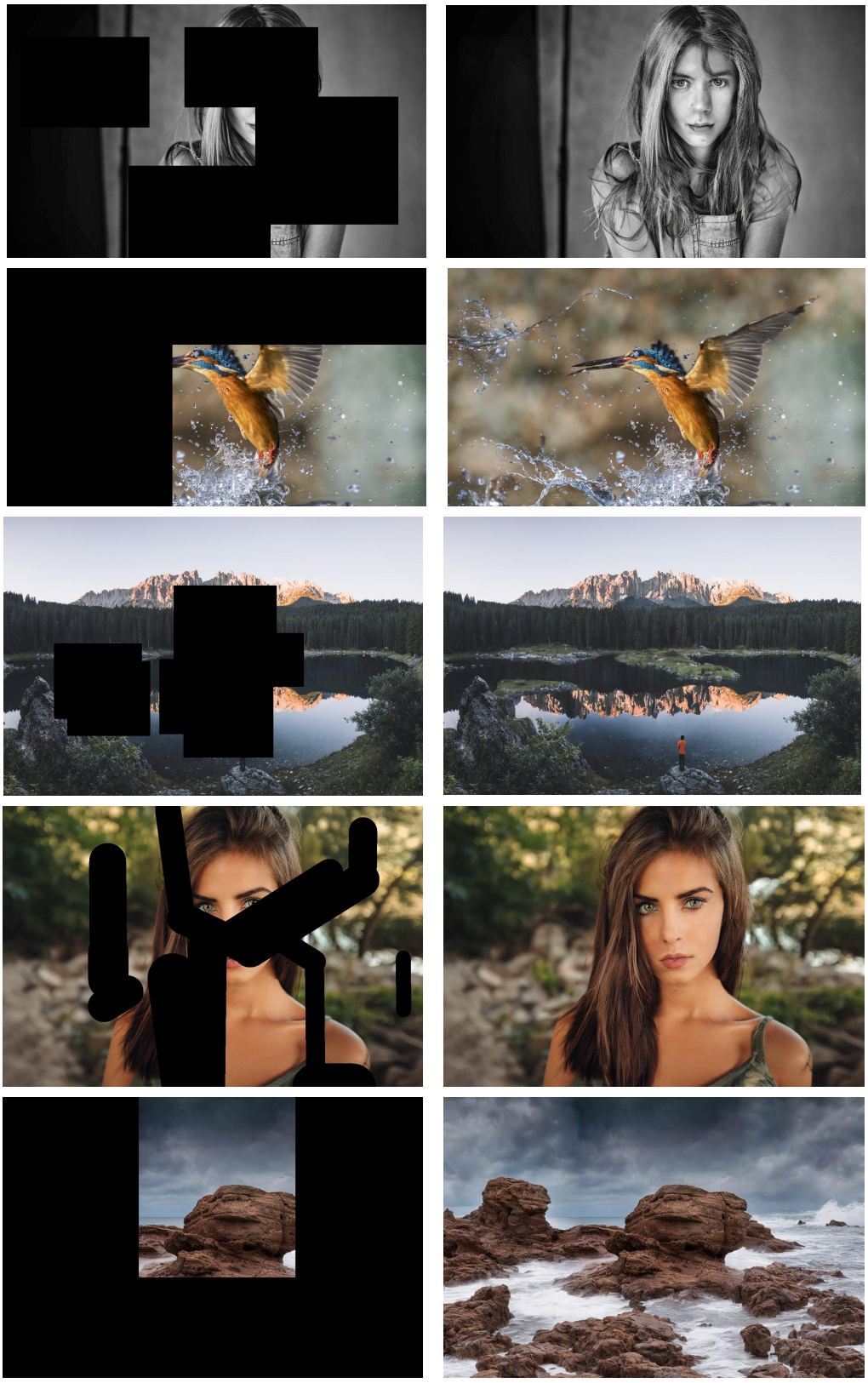


Figure 5. More inpainting & outpainting visual results.

Figure 6. More image-conditioned generation visual results.

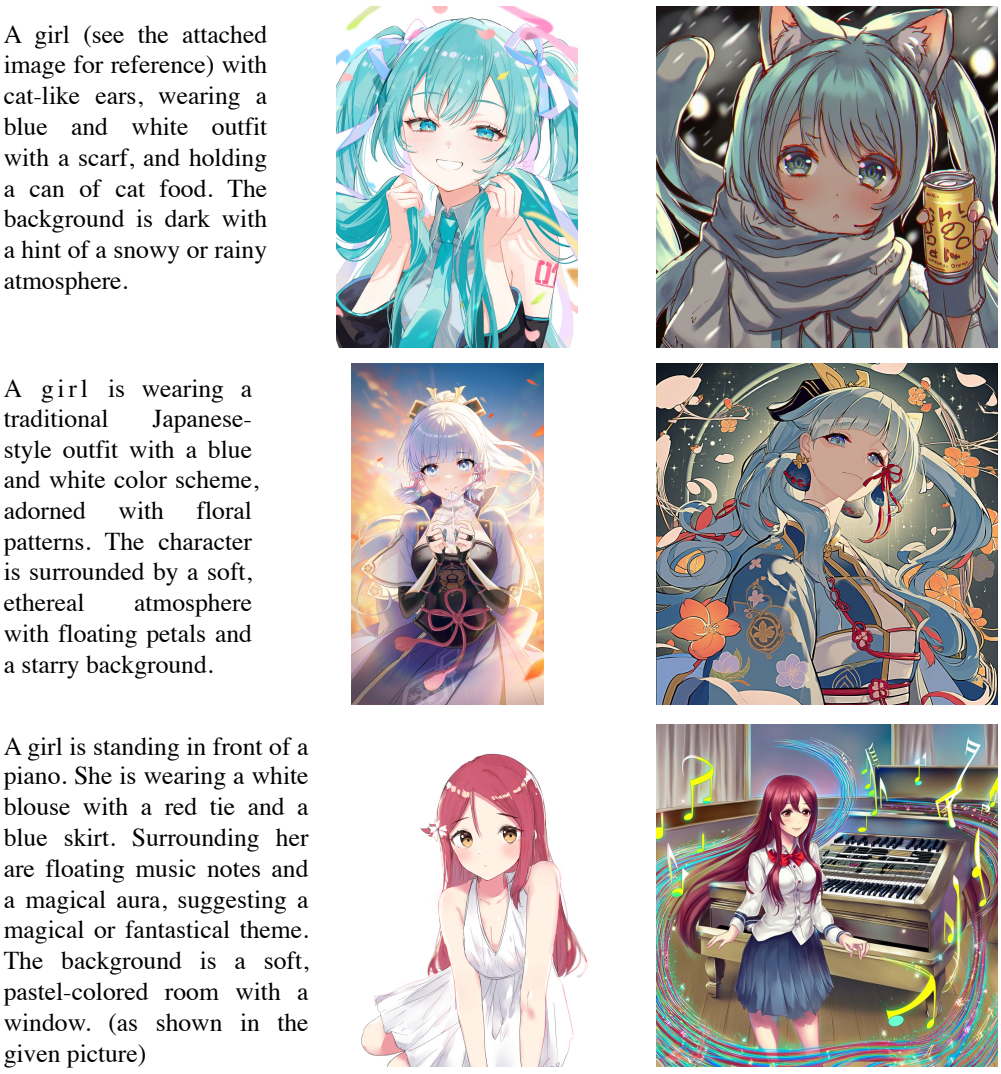


Figure 7. More subject-driven generation visual results.

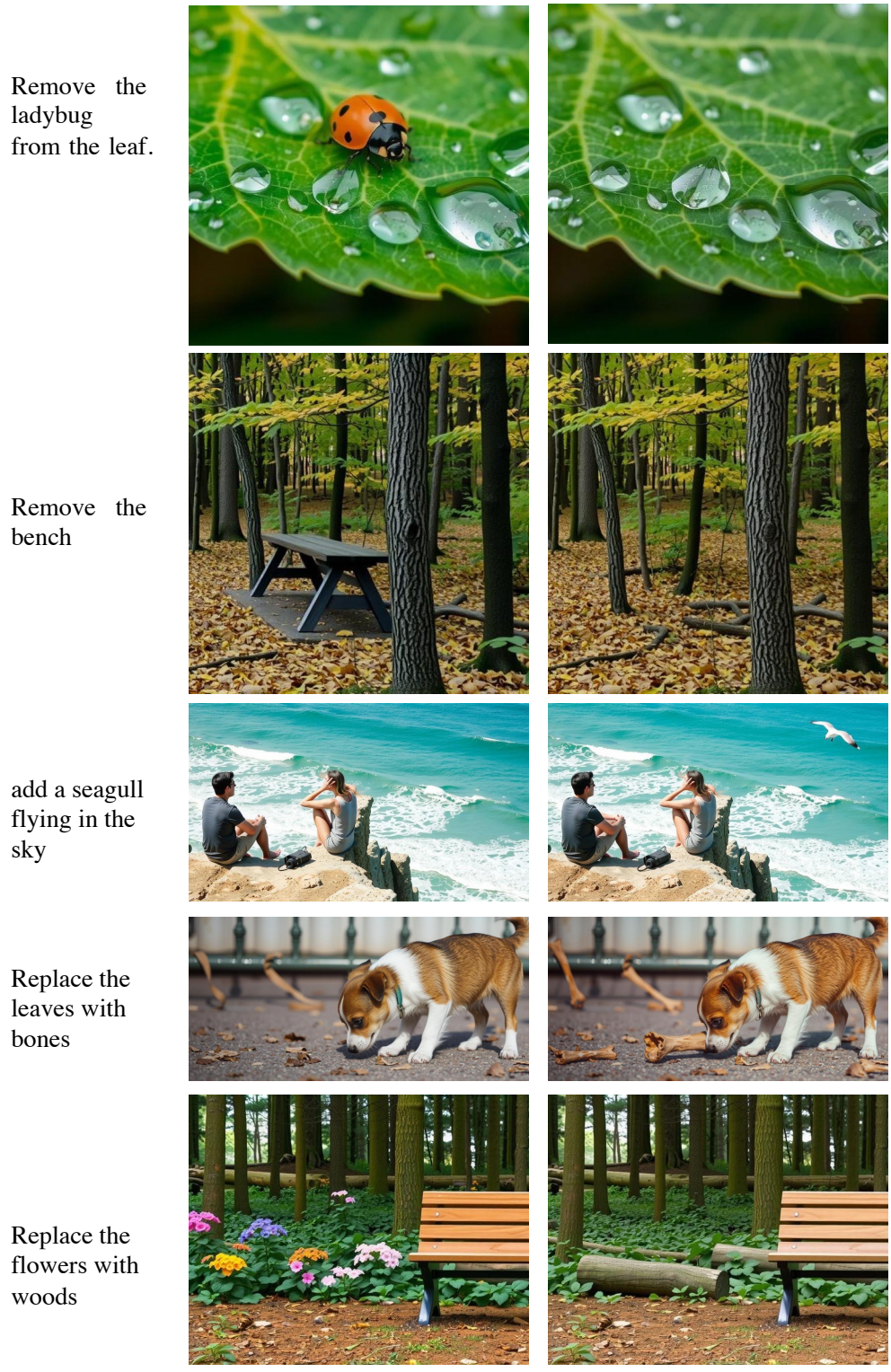


Figure 8. More instruction-based editing visual results.

Input Image & Drag Points

DreamOmni Result

Figure 9. More drag editing visual results.

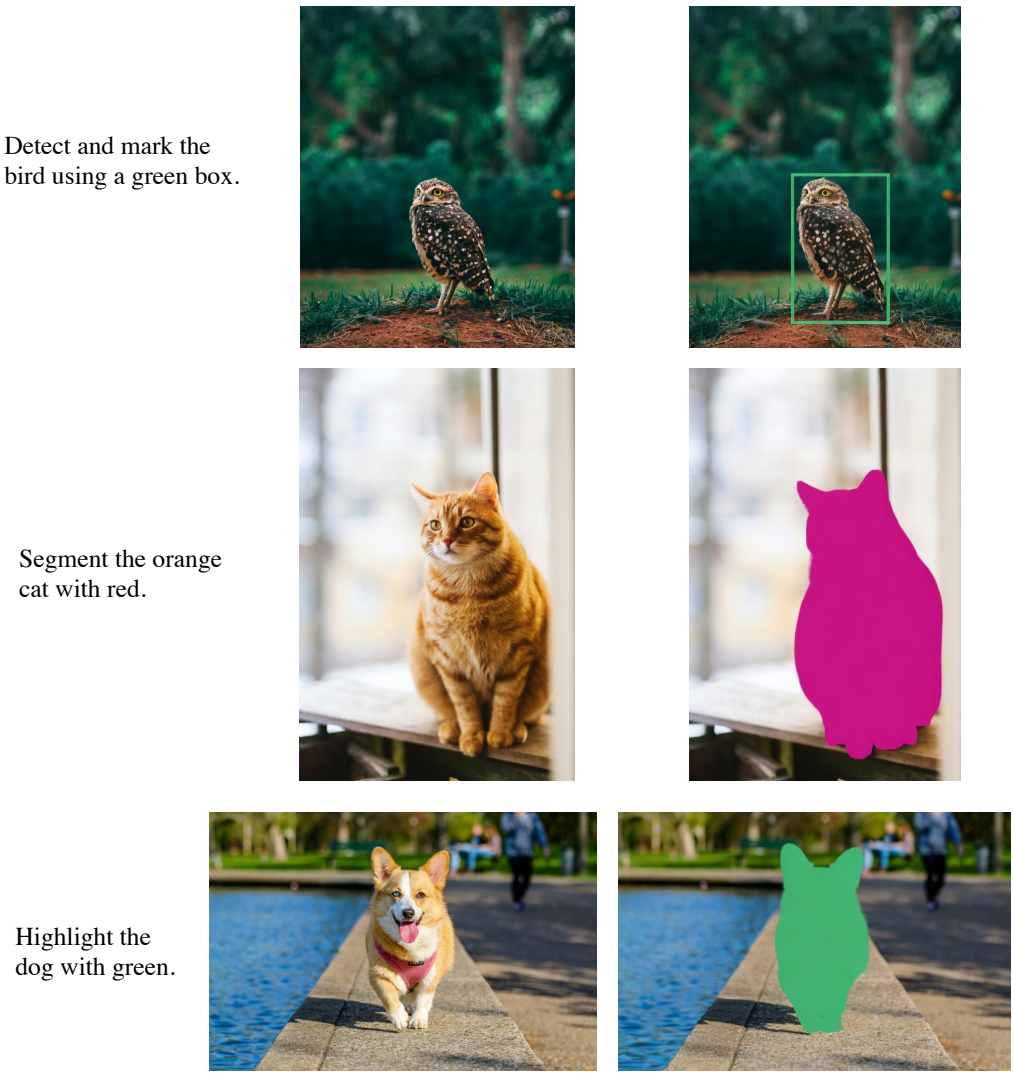


Figure 10. More segmentation & detection visual results.