

Supplementary Materials: Consistent123: One Image to Highly Consistent 3D Asset Using Case-Aware Diffusion Priors

Anonymous Authors

In this section, we will cover the implementation details, more qualitative results, and more 3D assets reconstructed by our method that are not shown in the paper due to space constraints. We hope that these materials will help the reader to better understand our paper.

1 3D RECONSTRUCTED ASSETS BY CONSISTENT123

We invite readers to explore the 3D assets reconstructed by Consistent123 in Figure 1 and Figure 2.

2 METRICS

On GSO, we measure **Chamfer Distance**, **volumetric IoU**, **multi-view PSNR** and **multi-view LPIPS**. The ground truth 3D mesh and the reconstructed 3D mesh are first normalized within the unit cube. The chamfer distance is defined as follows:

$$\begin{aligned} \text{CD}(S_1, S_2) = & \frac{1}{S_1} \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2^2 + \\ & \frac{1}{S_2} \sum_{y \in S_2} \min_{x \in S_1} \|y - x\|_2^2 \end{aligned} \quad (1)$$

where S_1 and S_2 denote the sets of vertices sampled from the real mesh and the reconstructed mesh, respectively, and in our experiments the number of sampled points is set to 2000. If this distance is smaller, it means that the reconstruction is better. Volumetric IoU is defined as the intersection of two meshes divided by the union of two meshes. Specifically, the meshes are first converted to the voxel representation, and we compute the volumetric IoU at resolution 64^3 . Since 3D ground truths are available, we can get rendered images from any viewpoint, which can be used as ground truth images to compute multi-view PSNR and multi-view LPIPS. Specifically, we select 18 viewpoints, each represented by azimuth and elevation. The 18 azimuth angles are valued as:

$$\{0^\circ, 20^\circ, 40^\circ, 60^\circ, 80^\circ, 100^\circ, 120^\circ, 140^\circ, 160^\circ, 180^\circ, 200^\circ, 220^\circ, 240^\circ, 260^\circ, 280^\circ, 300^\circ, 320^\circ, 340^\circ\} \quad (2)$$

The 18 elevation angles are valued as:

$$\{0^\circ, 0^\circ, -30^\circ, -15^\circ, -10^\circ, -5^\circ, -5^\circ, 0^\circ, 0^\circ, 0^\circ, 5^\circ, 5^\circ, 10^\circ, 15^\circ, 30^\circ, 0^\circ, 0^\circ\} \quad (3)$$

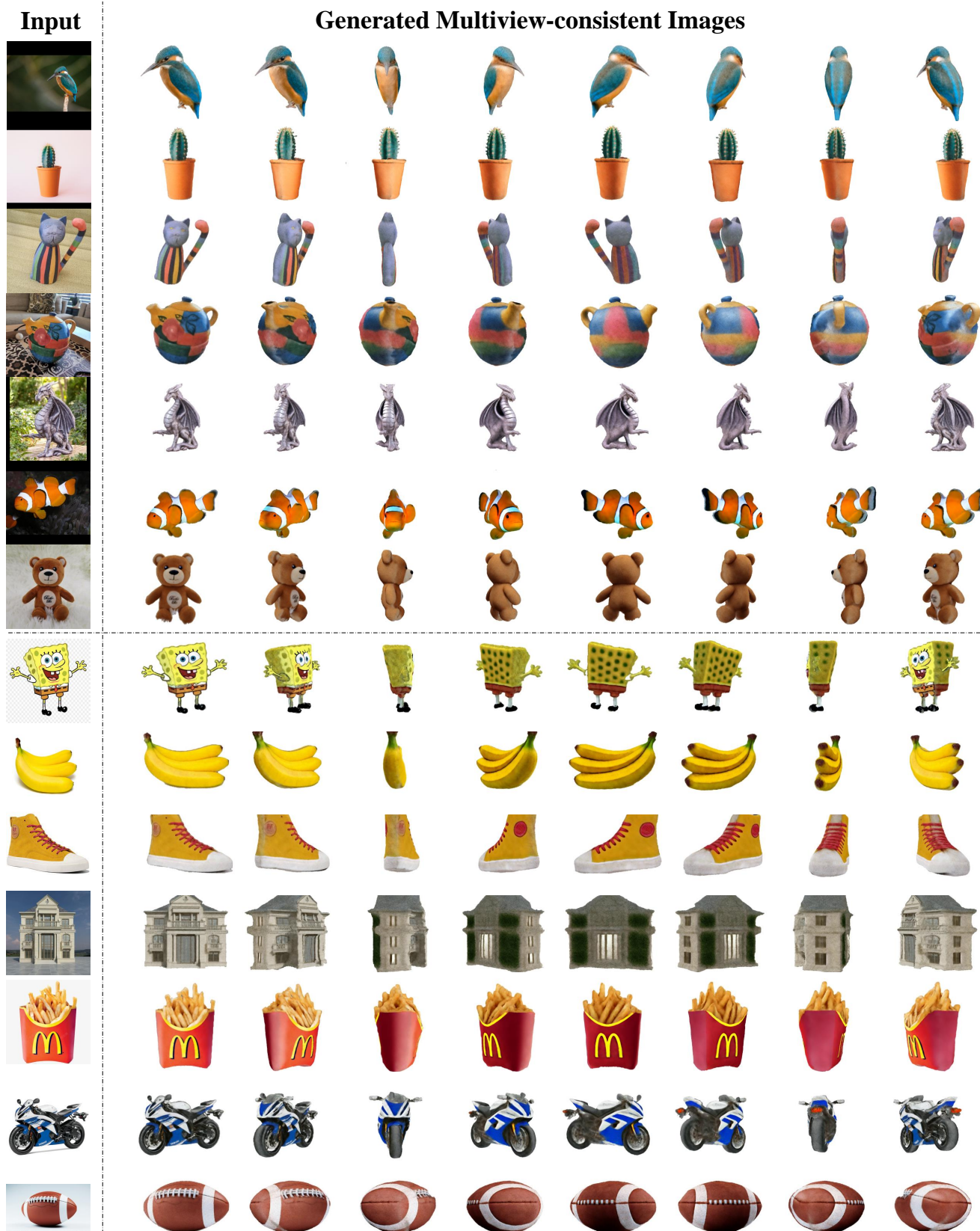


Figure 1: The reconstruction of Consistent123 on RealFusion15 and C10. We showcase the 3D model from 8 perspectives.



Figure 2: The 3D assets obtained by Consistent123.