A. Summary Video with Visual Results

We summarize our findings in a video, which outlines the three-stage DreamBooth3D method and includes a comparison to the baselines. We also show how our approach compares to other approaches via a user study. Finally, several example applications are shown, including material editing, accessorization, color changes, and pose changes.

B. NeRF Details

We use Mip-NeRF [**?**] as our choice of volumetric representation. Particularly, to render the color of a ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ cast into the scene, Mip-NeRF divides the ray into intervals and for each interval calculates the mean and variance (μ, Σ) of a conical frustum corresponding to the interval. These values are then used to encode the ray using integrated positional encoding

$$\gamma(\mu, \Sigma) = \left\{ \begin{bmatrix} \sin(2^l \mu) \exp(-2^{(2l-1)} \mathrm{diag}(\sigma)) \\ \sin(2^l \mu) \exp(-2^{(2l-1)} \mathrm{diag}(\sigma)) \end{bmatrix} \right\}_{l=0}^{L} \tag{1}$$

The learnt volume $\mathcal{N}_\phi$ is then used to generate albedo $\mathbf{c}$ and opacity σ.

$$\mathbf{c}, \sigma = \mathcal{N}_\phi(\gamma(\mu, \Sigma))$$

The final color is then calculated using numerical quadrature as in [**?**]. As in [**?**], we define $\Sigma = \lambda_t^2 I$, where, λ_t is annealed from a high to low value, to gradually introduce higher frequency components during the optimization. The NeRF volume is regularized using the orientation loss introduced in Ref-NeRF [**?**], to encourage better geometry. Particularly,

$$\mathcal{L}_{ori} = \sum_i \text{stop\_grad}(w_i) \max(0, \mathbf{n}_i . \mathbf{v}) \tag{2}$$

Where $\mathbf{n}_i$ is the normal direction at a point, w_i are rendering weights as defined in [**?**] and $\mathbf{v}$ is the lighting direction. An additional opacity loss is used to encourage foreground/ background separation

$$\mathcal{L}_{op} = \sqrt{(\sum_i w_i)^2 + 0.01} \tag{3}$$

The final NeRF regularization loss is then given by:

$$\mathcal{L}_{nerf} = \mathcal{L}_{op} + \mathcal{L}_{ori} \tag{4}$$

C. Additional results

Fig. 1 provides additional results with associated depths, normals and alpha maps to demonstrate the 3D consistency of our results on a variety of subjects. Fig. 2 shows multiple views of the assets rendered for the same subject with different text prompts.

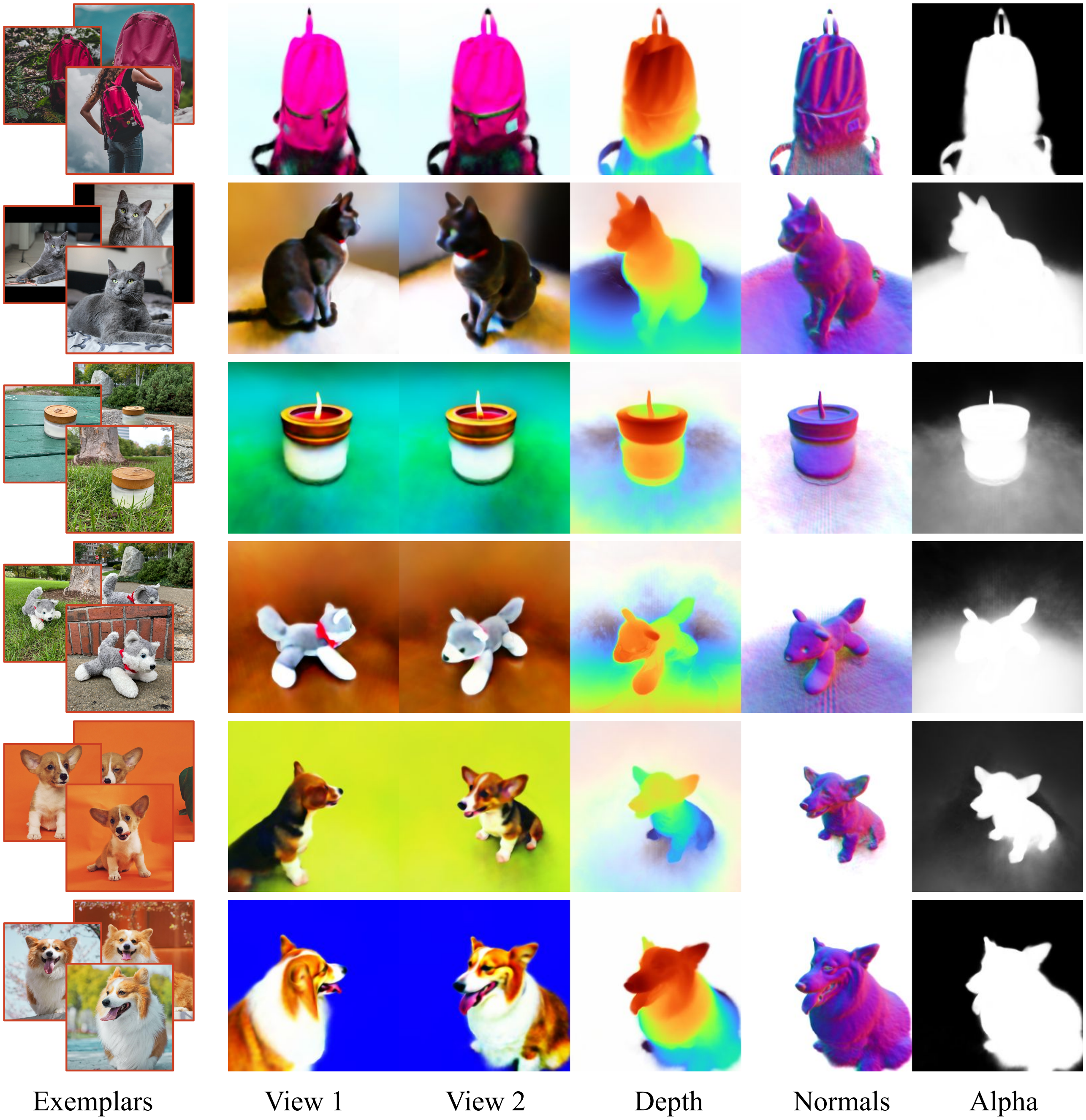

Figure 1: **Additional Results**. Dreambooth3D can produce 3D consistent volumes from text prompts. The figure shows generated assets for the base prompt ”A photo of <v>” where <v> is the subject presented in the first column. Column 2 and 3 shows two different views of the rendered volume. Column 3,4 and 5 shows the depth, normals and opacity of the second view respectively.

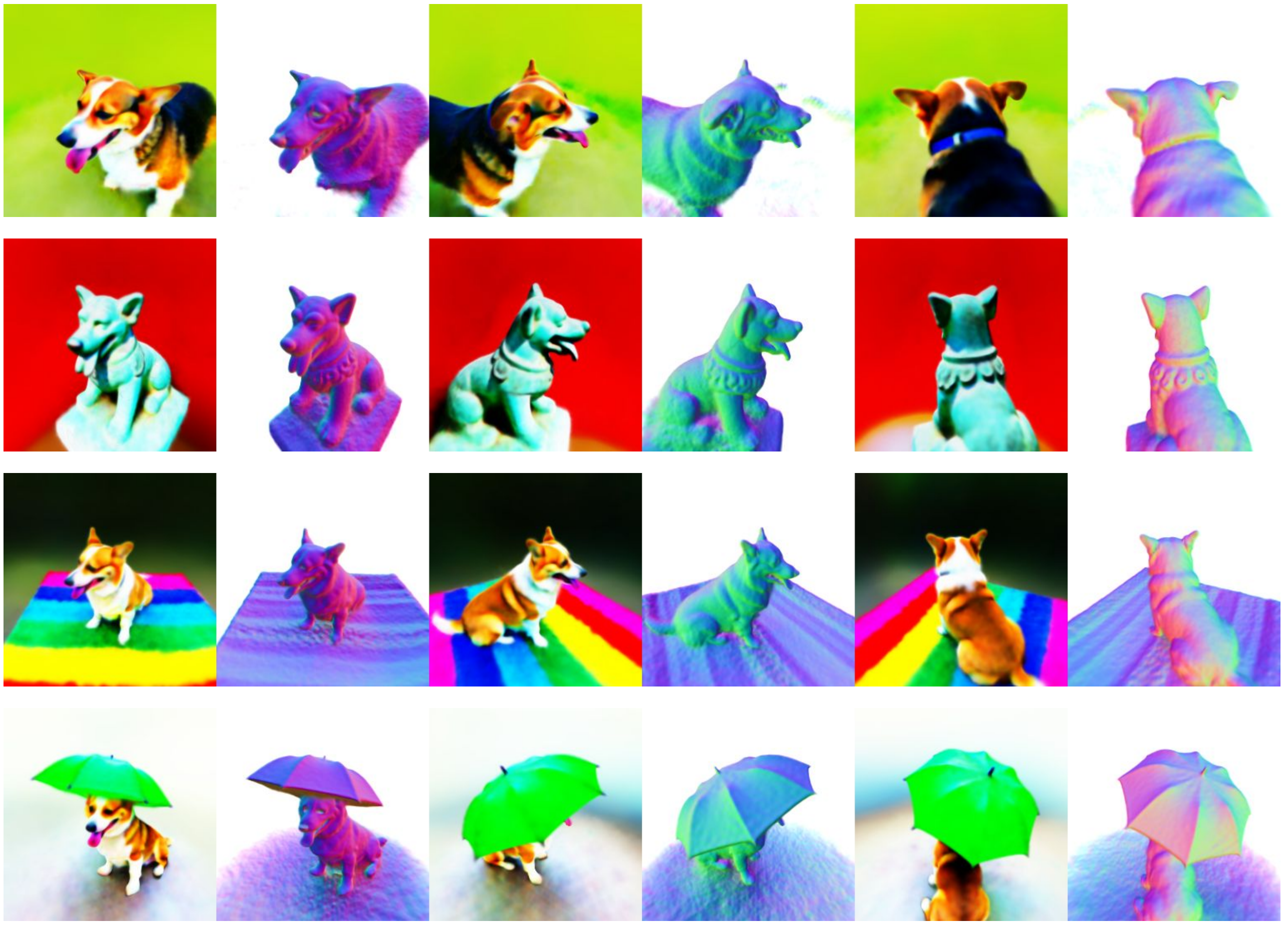

Figure 2: **Additional Results**. Dreambooth3D is capable of a number of accessorization, composition, material editing tasks through text prompting. An example of this form of prompting is "A photo of <v> wearing a green umbrella". Row 1 shows the rendered geometry and normals of the base subject, and the subsequent rows show material-edited, composited, or accessorized variants. Row 2 demonstrates a material edit to change the dog into a stone statue. Row 3 composites a rainbow carpet into the scene. Row 4 adds a green umbrella.