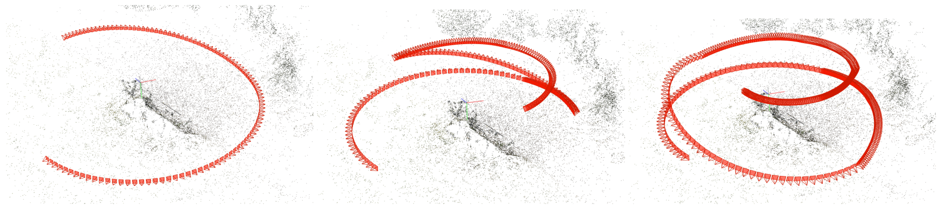
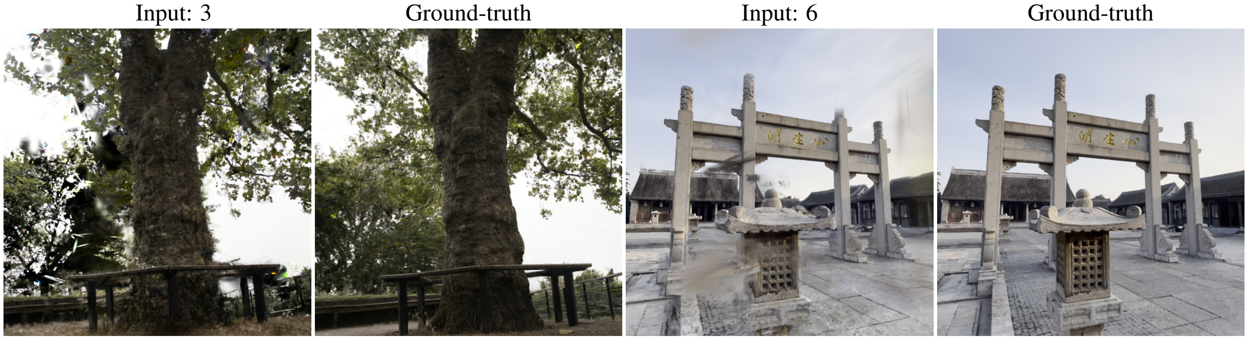


Input: 3 Input: 6 Input: 9

Figure 6: The fitting trajectories under different number of input views.

Input: 3 Ground-truth Input: 6 Ground-truth



Input: 9 Ground-truth Input: 5% Ground-truth

Figure 7: The low and high quality image pairs created in our 3DGS Enhancement dataset.

A Details of 3DGS Enhancement Dataset

For our 3DGS Enhancement Dataset, constructed based on DL3DV, we randomly select 120 scenes to create the training set for our video diffusion model and 30 scenes as the test set. By following previous works, we use the standard train/test split, selecting every 8th frame of the remaining frames for evaluation.

To create image pairs simulating the artifacts due to the lack of input views in novel view synthesis problem, we render the image pairs from pairs of low-high quality 3DGS models. Specifically, the input views for the high-quality model consist of all images in the original dataset, while the inputs for the low-quality model are a subset uniformly sampled from the original dataset. To add more complexity, we sample the subset according to a certain number (e.g., 3, 6, 9) or a certain ratio (e.g., 5%). With the aim to fully capture the distribution of artifacts created by the sparse input views and train the video diffusion model with smoother inputs, we propose a heuristic trajectory fitting algorithm, as shown in Figure 6, proving a sequence of cameras by interpolating the low or high-quality model's input views. Specifically, if the original camera trajectories are smooth and simple, such as those of DL3DV, we use the high-quality input views as the reference to fit the trajectories. For complex trajectories, such as those in Mip-NeRF 360, we use the low-quality input to avoid significantly poor rendering results, which would lead to unreasonable artifact distributions. As a result, we render a large number of image pairs with and without artifacts, as shown in Figure 7, at a resolution of 512×512, leading to powerful video diffusion priors with high view consistency and photo-realism.

B Details of Comparison Baselines

For the evaluation datasets, we compare against the standard 3D Gaussian Splatting [16] (which is also the reconstruction pipeline used in our work), and the state-of-the-art few-view NVS regularization methods, including Mip-NeRF[1], FreeNeRF [40], Zip-NeRF [3], and RegNeRF [24]. We also compare to some few-shot NVS methods using generative priors including ZeroNVS [30], and ReconFusion [37].

For the evaluation of MipNeRF, FreeNeRF, RegNeRF, and DNGaussian on DL3DV and Mip-NeRF 360 dataset, we follow the original configurations and code shared by the authors. Additionally, we use random point cloud as the initialization for 3DGS, following the implementations from DNGaussian. We also decrease the batch size for RegNeRF from 4096 to 512 according to the limited computation resource.