FreeFix: Boosting 3D Gaussian Splatting via Fine-Tuning-Free Diffusion Models

Supplementary Material

6. 3DGS Fisher Information Derivation

The uncertainty attribute of 3DGS in this paper is defined as:

$$\bar{\mathcal{C}}_{\mathcal{V};\mathcal{G}} = -\log p_f(\pi|\mathcal{V};\mathcal{G}) \tag{13}$$

Under the following regularity conditions, $-\log p_f(\pi|\mathcal{V};\mathcal{G})$ can be viewed as a loss term for Fisher information. It can also be expressed as an expectation term to represent Fisher information: $-\mathbb{E}_{\log p_f}[\frac{\partial^2 \log p_f(\pi|\mathcal{V};\mathcal{G})}{\partial \mathcal{G}^2}]$:

- The partial derivative of $p_f(\pi|\mathcal{V};\mathcal{G})$ with respect to $\mathcal{G}$ exists almost everywhere.
- The integral of $p_f(\pi|\mathcal{V};\mathcal{G})$ can be differentiated under the integral sign with respect to $\mathcal{G}$.
- The *support* of $p_f(\pi|\mathcal{V};\mathcal{G})$ does not depend on $\mathcal{G}$. In mathematics, the *support* of a real-valued function p_f is the subset of the function domain of elements that are not mapped to zero.

The volume rendering of 3D Gaussians meets these regularity conditions. With the consideration of $-\log p_f(\pi|\mathcal{V};\mathcal{G})$ can be regarded as the loss term of $\mathcal{L}$, the uncertain attribute of 3DGS can be represented as:

$$\begin{aligned}
\bar{\mathcal{C}}_{\mathcal{V};\mathcal{G}} &= -\mathbb{E}_{\log p_f}[\frac{\partial^2 \log p_f(\pi|\mathcal{V};\mathcal{G})}{\partial \mathcal{G}^2}] \\
&= \mathbb{E}_{\log p_f}[\frac{-\partial^2 \log p_f(\pi|\mathcal{V};\mathcal{G})}{\partial \mathcal{G} \partial \mathcal{G}^T}] \\
&= \mathbb{E}_{\log p_f}[\frac{\partial^2 \mathcal{L}(\mathcal{G})}{\partial \mathcal{G} \partial \mathcal{G}^T}] \\
&= \mathbf{H}^{''}[\pi|\mathcal{V};\mathcal{G}] \\
&= \nabla_{\mathcal{G}} \pi(\mathcal{V};\mathcal{G})^T \nabla_{\mathcal{G}} \pi(\mathcal{V};\mathcal{G})
\end{aligned} \tag{14}$$

7. Extrapolated Views Design

We design extrapolated testing views for the LLFF [18], Mip-NeRF 360 [1], and Waymo [29] datasets. The process for generating testing views in the Waymo dataset is straightforward; we translate the camera by 2 to 3 meters or rotate it by 10 to 15 degrees horizontally. However, the design for LLFF and Mip-NeRF 360 is not as straightforward, as we aim to construct extrapolated views that have ground truth images. For this reason, we cannot generate trajectories freely; instead, we need to create partitions for the testing and training sets. We present visualizations of the training and testing cameras in Fig. 8 from these scenes to illustrate the design of the extrapolated views. For some scenes where obvious extrapolated trajectories cannot be directly extracted, we aim to make the training views sparse in order to produce relative extrapolated trajectories.

8. Sampling Strategy

The supervisions during 3D refinement for $\mathcal{G}_i$ are sampled from current refined view $(\mathcal{V}_i^e, \hat{\mathcal{I}}_i^{e,f})$, previous refined views $\mathcal{F}_{i-1}$ and training views S_{train}. Each stage of 3D refinement aims to fit the newly refined 2D image while preserving rendering ability in the original training and previously refined views. The sampling strategy for training is structured as follows. During the first third of the 3D refinement steps, every three steps are designated as current-refine steps, using the current refine image $\hat{\mathcal{I}}_i^{e,f}$ to refine 3DGS. In the subsequent third of the 3D refinement steps, every five steps are defined as current-refine steps, and in the final third of the 3D refinement steps, every eight steps are designated as current-refine steps. For the remaining non-current-refine steps, we randomly select views from the training set $\mathcal{S}_{train}$ and the previous refined set $\mathcal{F}_{i-1}$, but with different selection weights. The probability of selecting views from $\mathcal{F}_{i-1}$ is lower compared to that of selecting views from $\mathcal{S}_{train}$.

9. Additional Experiments

9.1. More Comparisons with Baselines

We provide more qualitative comparisons in Fig. 9. The quantitative comparisons on each scene are shown in Tab. 4, Tab. 5, and Tab. 6. Additionally, Fig. 11 shows the quantitative comparisons between FreeFix and NVS-Solver [45].

9.2. Uncertainty as Guidance

In this paper, we apply certainty as guidance during denoising. In this subsection, we provide a comparison between using the uncertainty mask from [7] as guidance and our certainty mask as guidance. Specifically, for rendered uncertain masks $\mathcal{M}^{\bar{c}}$, we use $1 - \mathcal{M}^{\bar{c}}$ as guidance to experiment on Garden in Mip-NeRF 360. As shown in Fig. 10 and Tab. 7, the images generated using uncertainty masks as guidance exhibit significant inconsistency, resulting in less satisfying performance.

9.3. Ablation on Affine Transform

We apply an affine transform during 3D refinement to prevent 3DGS from learning slightly different color styles generated by diffusion models. In this subsection, we present an ablation study for this component on Garden in Mip-NeRF 360. As shown in Tab. 8, although removing the affine transform slightly improves PSNR, it results in a decrease in SSIM and LPIPS. Furthermore, as illustrated in Fig. 12, removing the affine transform results in large floaters in testing views, which can significantly lower human sensory preference.

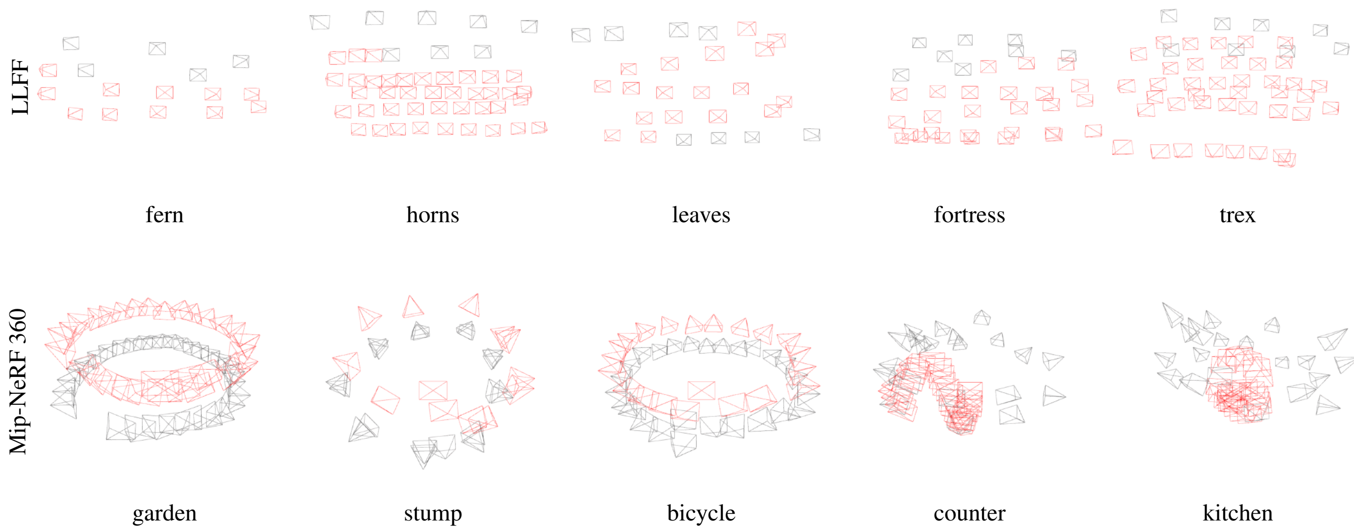


Figure 8. **Design of Training and Testing Views Design.** We design partitions to conduct experiments on extrapolated testing views. Training views and Testing views are highlighted with their respective colors.

		3DGS	FreeFix + Flux	FreeFix + SDXL	ViewExtrapolator [16]	NVS-Solver [45]	Difix3D+ [37]
Fern	PSNR ↑	17.78	19.3	19.39	18.63	12.65	18.5
	SSIM ↑	0.603	0.656	0.658	0.619	0.375	0.631
	LPIPS ↓	0.289	0.243	0.245	0.3	0.551	0.265
Flower	PSNR ↑	18.64	18.95	18.54	17.59	11.04	19.07
	SSIM ↑	0.575	0.612	0.605	0.527	0.253	0.594
	LPIPS ↓	0.265	0.254	0.265	0.367	0.654	0.244
Fortress	PSNR ↑	16.97	21.33	20.32	21.97	12.8	17.87
	SSIM ↑	0.689	0.751	0.729	0.702	0.387	0.712
	LPIPS ↓	0.205	0.194	0.255	0.25	0.473	0.166
Horns	PSNR↑	16.76	19.06	18.95	18.17	11.81	17.78
	SSIM↑	0.588	0.69	0.685	0.615	0.336	0.63
	LPIPS↓	0.322	0.28	0.3	0.36	0.588	0.294
Leaves	PSNR↑	14.6	16.51	16.63	14.49	9.94	14.82
	SSIM↑	0.432	0.525	0.53	0.382	0.115	0.438
	LPIPS↓	0.303	0.222	0.22	0.333	0.636	0.303
Room	PSNR↑	23.68	25.02	25.22	18.47	13.53	24.67
	SSIM↑	0.868	0.9	0.9	0.782	0.609	0.883
	LPIPS↓	0.196	0.143	0.146	0.457	0.465	0.173
Trex	PSNR↑	18.27	20.7	20.45	18.53	12.15	19.33
	SSIM↑	0.676	0.763	0.758	0.674	0.382	0.721
	LPIPS↓	0.275	0.212	0.228	0.3	0.553	0.229

Table 4. **Quantitative Comparison with Baselines for each scene in LLFF.**

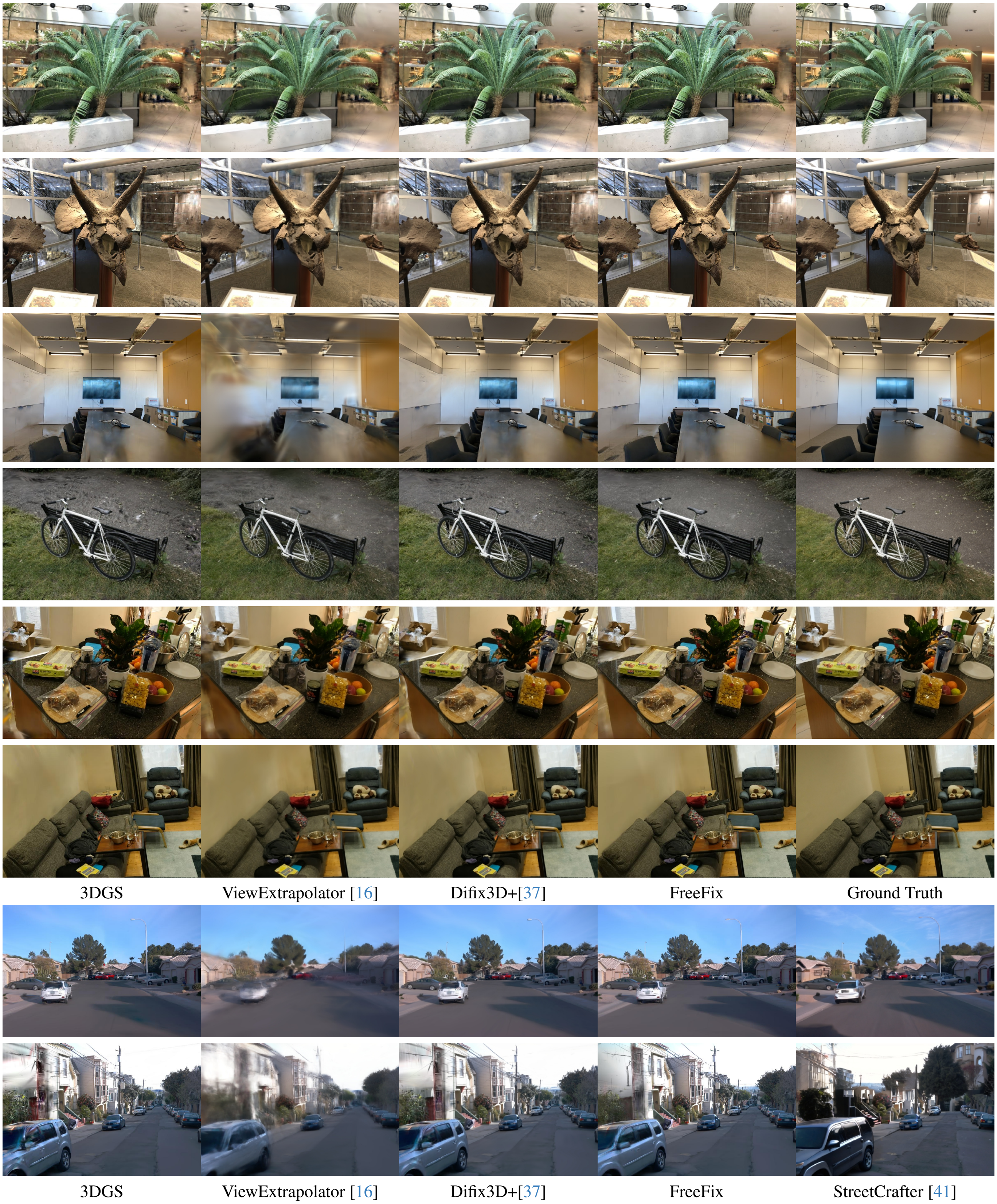


Figure 9. **Additional Qualitative Comparisons**


3DGS Ground Truth Uncertainty Certainty

Figure 10. **Generated Results Comparison between Uncertainty and Certainty as Guidance.**

		3DGS	FreeFix + Flux	FreeFix + SDXL	ViewExtrapolator [16]	NVS-Solver [45]	Difix3D+ [37]
Bicycle	PSNR↑	20.71	22.61	22.48	20.0	14.58	21.39
	SSIM↑	0.497	0.589	0.588	0.482	0.266	0.519
	LPIPS↓	0.327	0.267	0.269	0.419	0.626	0.293
Bonsai	PSNR↑	23.68	24.5	24.07	22.01	10.27	24.19
	SSIM↑	0.828	0.837	0.829	0.725	0.221	0.841
	LPIPS↓	0.147	0.132	0.14	0.205	0.632	0.128
Counter	PSNR↑	22.2	23.29	23.06	22.01	10.56	23.03
	SSIM↑	0.788	0.806	0.803	0.762	0.281	0.806
	LPIPS↓	0.157	0.149	0.152	0.199	0.65	0.137
Garden	PSNR↑	18.38	19.72	19.42	17.86	12.41	19.09
	SSIM↑	0.415	0.52	0.517	0.409	0.234	0.449
	LPIPS↓	0.357	0.288	0.294	0.505	0.626	0.305
Kitchen	PSNR↑	22.58	23.97	22.9	19.65	12.46	23.02
	SSIM↑	0.759	0.776	0.765	0.586	0.296	0.773
	LPIPS↓	0.199	0.168	0.18	0.396	0.618	0.172
Room	PSNR↑	26.3	26.9	26.79	25.06	10.42	26.7
	SSIM↑	0.87	0.884	0.88	0.813	0.345	0.877
	LPIPS↓	0.099	0.098	0.106	0.171	0.67	0.093
Stump	PSNR↑	18.97	20.14	20.06	19.31	16.45	19.6
	SSIM↑	0.343	0.415	0.414	0.356	0.222	0.359
	LPIPS↓	0.386	0.351	0.355	0.431	0.597	0.339

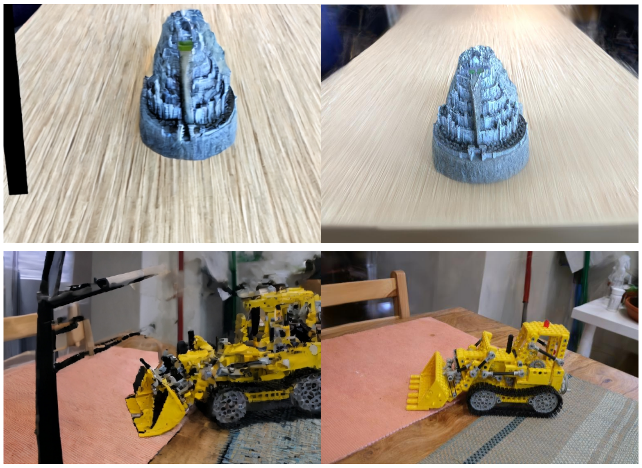
Table 5. **Quantitative Comparison with Baselines for each scene in Mip-NeRF 360.**

	3DGS	FreeFix + Flux	FreeFix + SDXL	ViewExtrapolator [16]	NVS-Solver [45]	Difix3D+ [37]	StreetCrafter [41]
Seq102751-Trans	0.181	0.169	0.176	0.242	0.282	0.173	0.225
Seq134763-Rot	0.133	0.125	0.133	0.155	0.314	0.114	0.112
Seq134763-Trans	0.156	0.144	0.134	0.184	0.213	0.142	0.178
Seq143481-Rot	0.113	0.112	0.103	0.124	0.323	0.124	0.122
Seq148697-Rot	0.1	0.089	0.094	0.175	0.281	0.089	0.124
Seq177619-Rot	0.214	0.204	0.21	0.182	0.31	0.2	0.262
Seq177619-Trans	0.187	0.182	0.197	0.192	0.296	0.163	0.192

Table 6. **Quantitative Comparison with Baselines for each scene in Waymo.** The metric in this table is KID ↓.

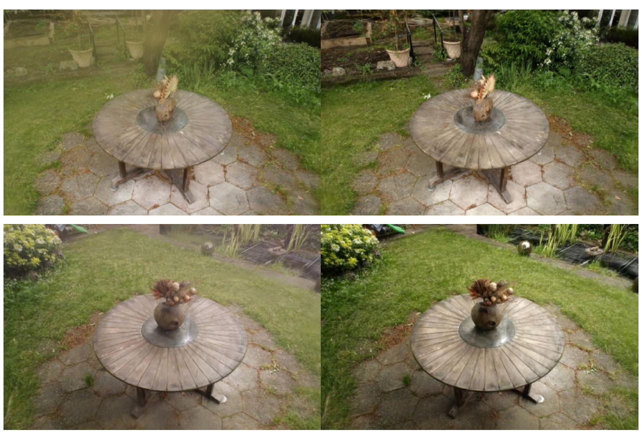
	PSNR↑	SSIM↑	LPIPS↓
Uncertainty Mask	19.30	0.515	0.310
Certainty Mask	**19.72**	**0.520**	**0.287**

Table 7. **Quantitative Comparison between Uncertainty and Certainty as Guidance**.



NVS-Solver [45] FreeFix

Figure 11. **Comparisons on FreeFix and NVS-Solver.** The less satisfying results may lead by inaccurate depth and warp results.



w/o Affine w/ Affine

Figure 12. **Comparison on Affine Transform Ablation Study.** The absence of the affine transform can lead to significant floaters in the testing views.

	PSNR↑	SSIM↑	LPIPS↓
FreeFix w/o Affine	**20.03**	0.517	0.317
FreeFix	19.72	**0.520**	**0.287**

Table 8. **Ablation Study on Affine Transform**. Although the affine transform results in a slight decrease in PSNR, this component helps to avoid significant floaters, thereby enhancing SSIM, LPIPS, and overall subjective quality.