3DGS-Enhancer: Enhancing Unbounded 3D Gaussian Splatting with View-consistent 2D Diffusion Priors

Xi Liu*, Chaoyi Zhou*, Siyu Huang

Visual Computing Division, School of Computing, Clemson University

Clemson Vision Learning Lab

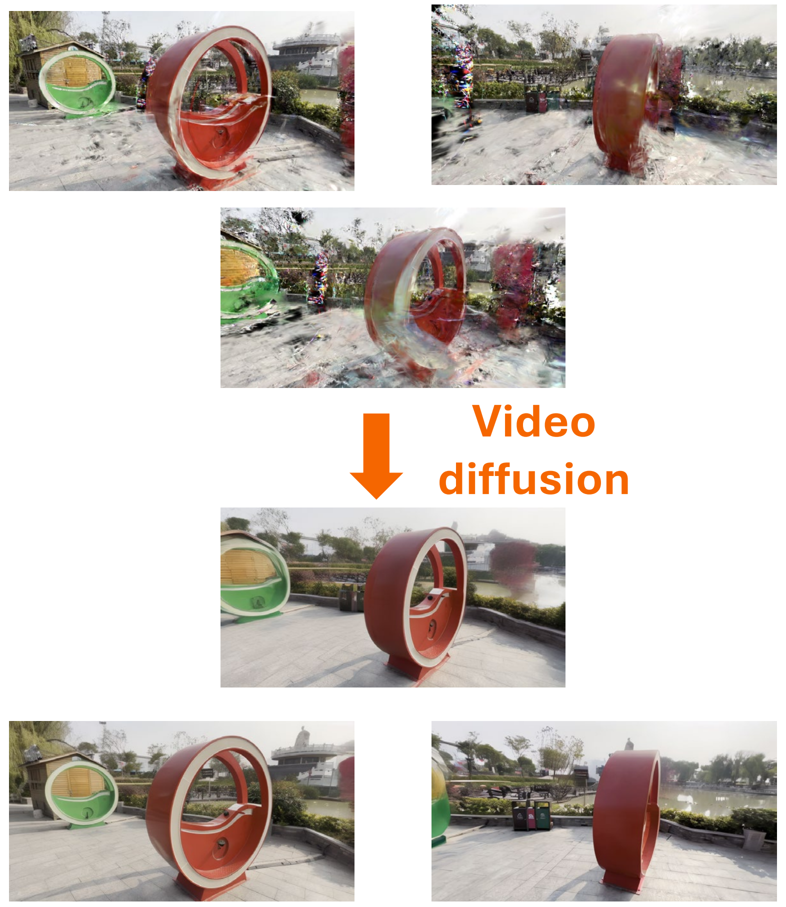
NEURAL INFORMATION PROCESSING SYSTEMS

Motivation and Key Hypothesis

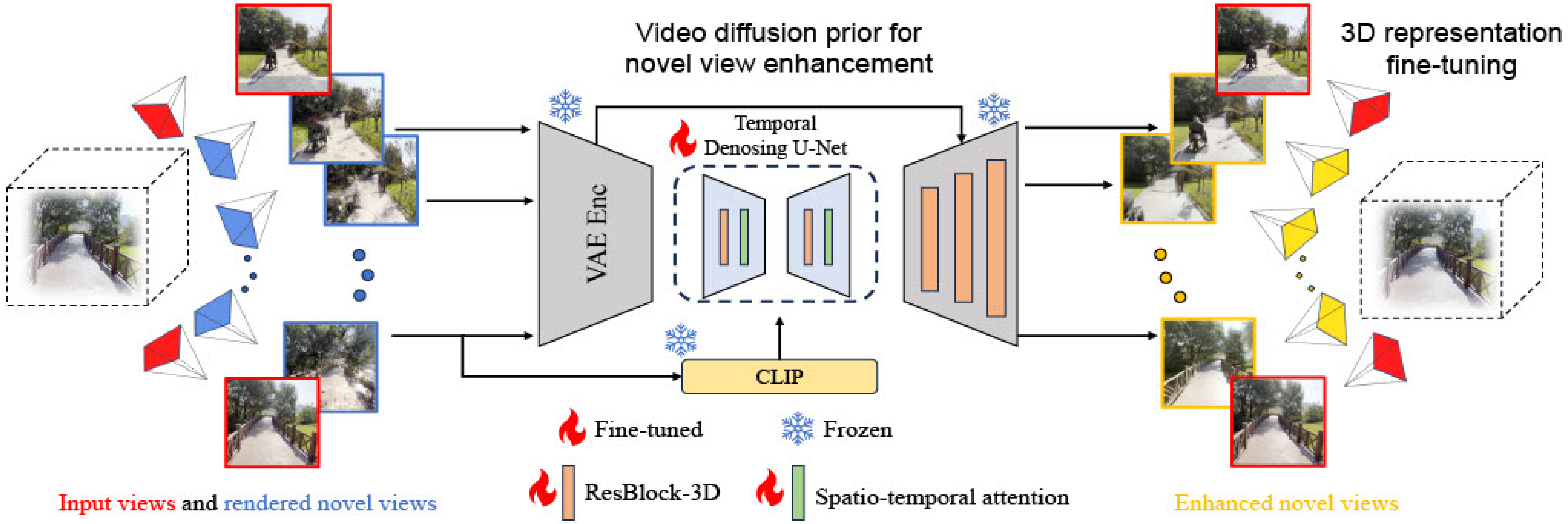
Problem: Sparse-view 3D reconstructions often result in significant **artifacts** and **inconsistencies**.

Hypothesis: Multi-view Consistency in 3D space can be learned as **Temporal Consistency** in 2D space.

Solution: Fine-tune a **multi-view consistent video diffusion** model using 3D reconstruction data to enhance unbounded 3D Gaussian splatting (3DGS)



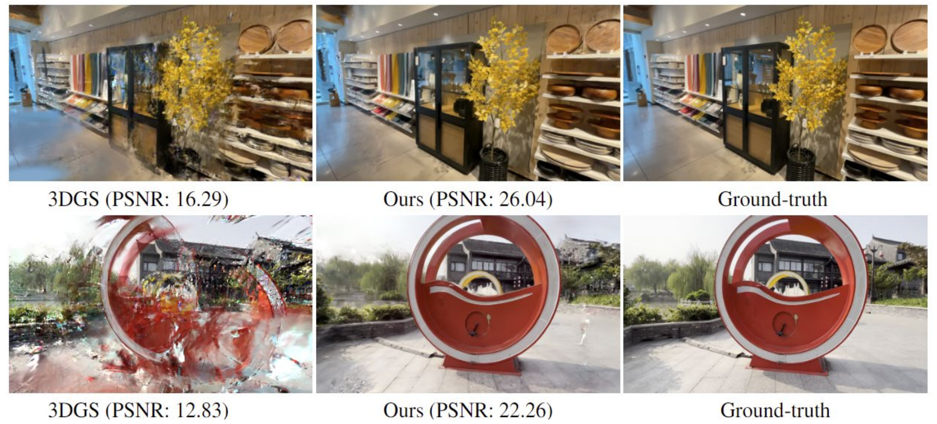
Method



1. **Video Diffusion Prior:** Finetune an **view-consistency** video diffusion model for enhancing novel view synthesis
2. **Spatial-Temporal Decoder:** Boost the and mitigate the **VAE degradation**
3. **Confidence-aware**: Reduce the impact of non-photorealistic and inconsistent regions generated by diffusion models.

Experiments

DL3DV-10K In-Domain



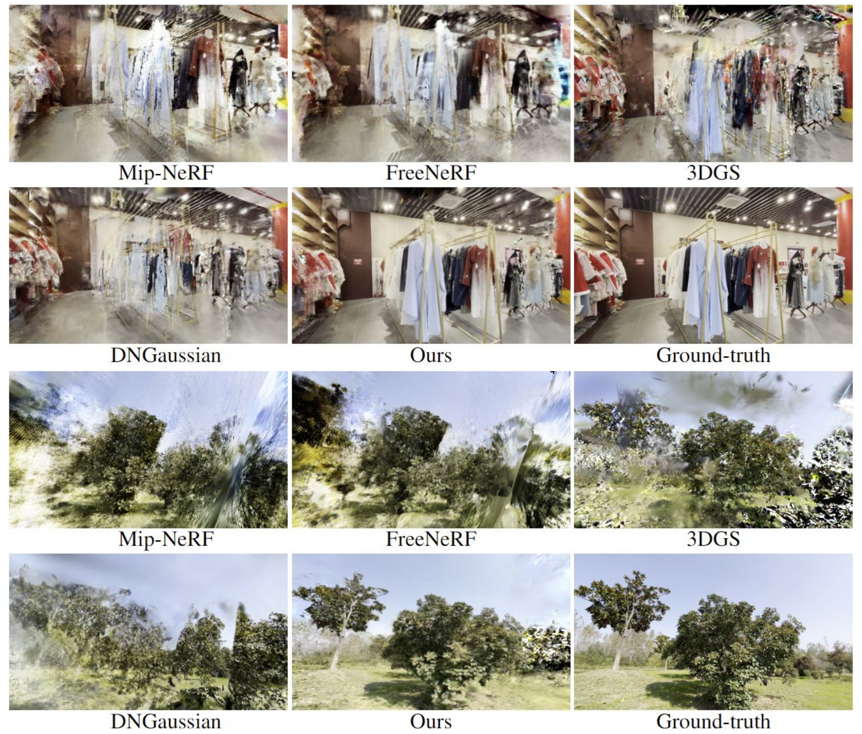
Method	3 views PSNR↑	3 views SSIM↑	3 views LPIPS↓	6 views PSNR↑	6 views SSIM↑	6 views LPIPS↓	9 views PSNR↑	9 views SSIM↑	9 views LPIPS↓
DL3DV (130 training scenes, 20 test scenes)									
Mip-NeRF [1]	10.92	0.191	0.618	11.56	0.199	0.608	12.42	0.218	0.600
RegNeRF [27]	11.46	0.214	0.600	12.69	0.236	0.579	12.33	0.219	0.598
FreeNeRF [43]	10.91	0.211	0.595	12.13	0.230	0.576	12.85	0.241	0.573
3DGS [18]	10.97	0.248	0.567	13.34	0.332	0.498	14.99	0.403	0.446
DNGaussian [19]	11.10	0.273	0.579	12.67	0.329	0.547	13.44	0.365	0.539
3DGS-Enhancer (ours)	**14.33**	**0.424**	**0.464**	**16.94**	**0.565**	**0.356**	**18.50**	**0.630**	**0.305**

Mip-NeRF360 Out-of-Domain

Method	6 views PSNR↑	6 views SSIM↑	6 views LPIPS↓	9 views PSNR↑	9 views SSIM↑	9 views LPIPS↓
Mip-NeRF360 (all test scenes)						
Mip-NeRF	13.08	0.159	0.637	13.73	0.189	0.628
RegNeRF	12.69	0.175	0.660	13.73	0.193	0.629
FreeNeRF	12.56	0.182	0.646	13.20	0.198	0.635
3DGS	11.53	0.144	0.651	12.65	0.187	0.607
DNGaussian	11.81	0.208	0.689	12.51	0.228	0.683
3DGS-Enhancer (ours)	**13.96**	**0.260**	**0.570**	**16.22**	**0.399**	**0.454**

We achieved **SOTA** performance on both **in-domain** and **out-of-domain** data.

Method Comparison 3 Views



Preserve the **structural details** of the image, such as **text**.
mitigate **color shift** of diffusion model.

STD Ablation study 9 views

Video diffusion	STD (temporal layers)	color correction	PSNR ↑	SSIM ↑	LPIPS ↓
✓	-	-	18.11	0.591	0.312
✓	✓	-	18.44	0.625	0.306
✓	✓	✓	**18.50**	**0.630**	**0.305**

PSNR: 18.11 -> 18.44

Confidence Ablation Study (all 3-6-9 views)

Video diffusion	Real image	Image confidence	Pixel confidence	PSNR↑	SSIM↑	LPIPS↓
✓	-	-	-	14.33	0.476	0.422
✓	✓	-	-	17.01	0.553	0.361
✓	✓	✓	-	17.29	0.570	0.354
✓	✓	-	✓	17.16	0.564	0.351
✓	✓	✓	✓	**17.34**	**0.574**	**0.351**

PSNR: 17.01 -> 17.34