

Era3D: High-Resolution Multiview Diffusion Using Efficient Row-wise Attention

Table 2: Memory usage and running time of the pipeline with 512 resolution and xFormer.

	Memory usage (G)			Running time (ms)		
	MV Attn	Other parts	Total	MV Attn	Other parts	Total
Dense	1.42		2.40	22.96		29.13
Epipolar	1.71	~1.0	2.81	20.03	~6.5	26.75
Ours	1.08		2.09	1.86		8.31

Table 3: Ablation of EFReg on GSO datasets with various elevation (α) and focal lengths (f). We report the Chamfer Distance ($\downarrow$).

Pose	$f=35$	$f=50$	$\alpha=0$ $f=85$	$f=105$	$f=135$	$\alpha=-10$	$\alpha=10$	$f=\infty$ $\alpha=20$	$\alpha=30$	$\alpha=40$	$\alpha=0$ $f = \infty$
w/o EFRreg	0.0237	0.0233	0.020	0.022	0.0217	0.0221	0.0217	0.0225	0.0231	0.0228	0.0217
w EFRreg	0.0223	0.0219	0.0216	0.0214	0.0214	0.0217	0.0216	0.0216	0.0219	0.0217	0.0213

Table 4: Quantitative evaluation of SSIM and PSNR.

Method	RealFusion	Zero-1-to-3	SyncDreamer	Wonder3D	Ours
SSIM($\uparrow$)	0.722	0.779	0.798	0.811	0.837
PNSR($\uparrow$)	15.26	18.93	20.05	20.83	22.74

Figure 1: View number ablation.

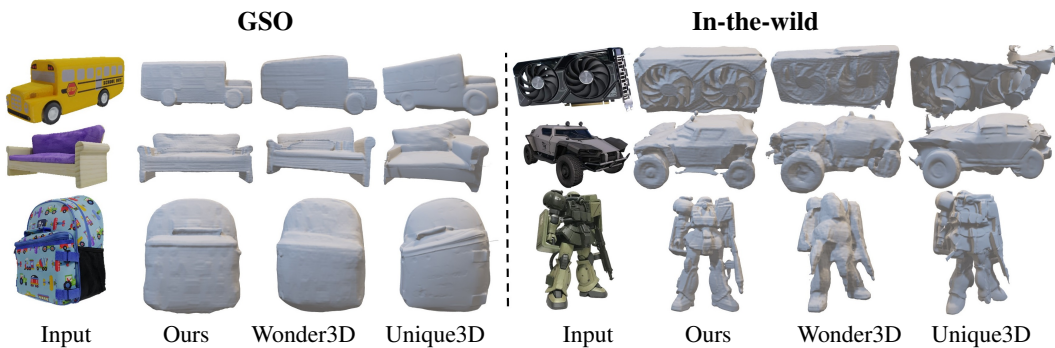
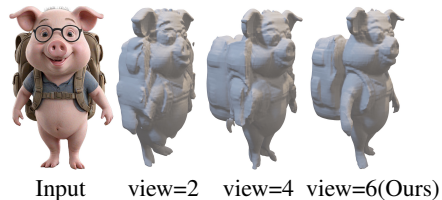


Figure 2: More comparisons w.r.t distortion problem.