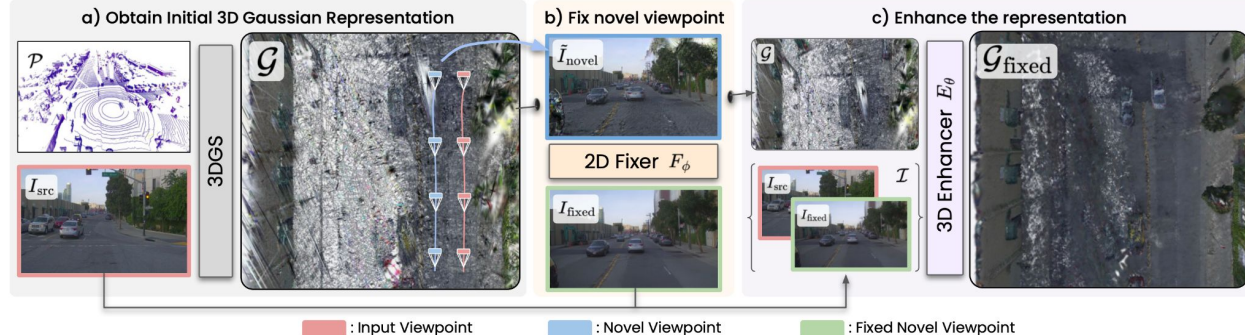


## Motivation

- Recent reconstruction methods such as 3D Gaussian Splatting (3DGS) enable various applications in robotics and autonomous driving through real-time, photorealistic rendering
- However, 3DGS overfits to input views and exhibits noticeable artifacts at large novel viewpoints, limiting downstream application

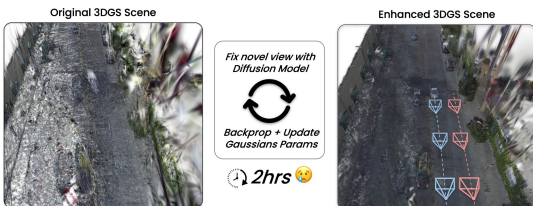


## GenRe: Diffusion-guided Generalizable 3D Enhancer



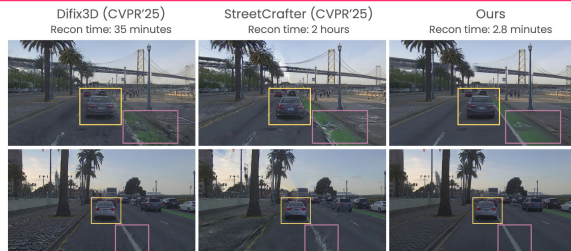
## Existing Work: 2D Diffusion-based Fixer

- 2D diffusion-based fixer to generate pseudo ground-truth images at novel views for reconstruction
- Limitations:** 1) costly per-scene optimization; 2) the distilled representations remain fragile and fail to generalize beyond limited synthesized views



**Goal:** A generalizable 3D Enhancer that achieves consistent novel view rendering within 2 minutes

## Comparison to SoTA methods



COMPARISON ON HARD EXTRAPOLATED VIEWS.

Methods	Extrapolation (Hard)		Recon time Minute↓
	FID@4m↓	FID@5m↓	
<b>Standalone reconstruction</b>			
3DGS	205.09	231.70	41.90
StreetGS	153.76	176.43	47.62
SplatAD	188.00	210.24	113.62
G3R	174.64	191.33	<b>0.90</b>
Ours (GNet)	<b>124.65</b>	<b>138.30</b>	0.96
<b>Reconstruction with neural fixers</b>			
StreetCrafter	118.94	141.14	127.33
Difix3D	137.80	167.28	35.32
Ours (GenRe+)	<b>102.73</b>	<b>114.19</b>	<b>2.77</b>

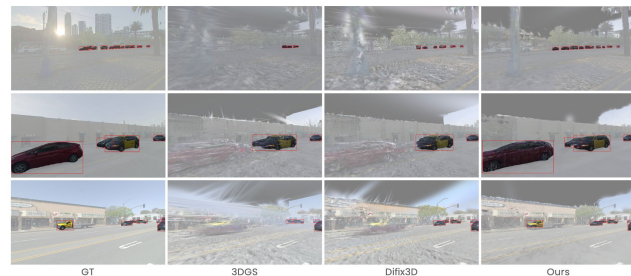
3D ENHANCER PLUGS INTO DIFFERENT 3DGS-BASED METHODS.

Methods	FID@3m↓
3DGS	154.21
+ GenRe	<b>99.69</b>
G3R	132.46
+ GenRe (zero-shot)	100.67
+ GenRe (fine-tune)	<b>91.12</b>

## Downstream Experiments

### Downstream evaluation

Front-camera recon → domain-gap eval on front-left camera for 2D detection/segmentation



### Downstream training

Relative 3D Detection Improvement over Real-only

Methods	mAP	AP@1m	AP@2m	AP@3m
Real-only	0.00%	0.00%	0.00%	0.00%
Real + Sim (3DGS)	0.78%	14.1%	-0.40%	-1.60%
Real + Sim (Ours)	<b>8.20%</b>	<b>23.5%</b>	<b>10.1%</b>	<b>3.66%</b>

### Limitations

- visible artifacts under very large viewpoint changes;
- does not fully complete object and background in occluded or unobserved regions