

FRUC

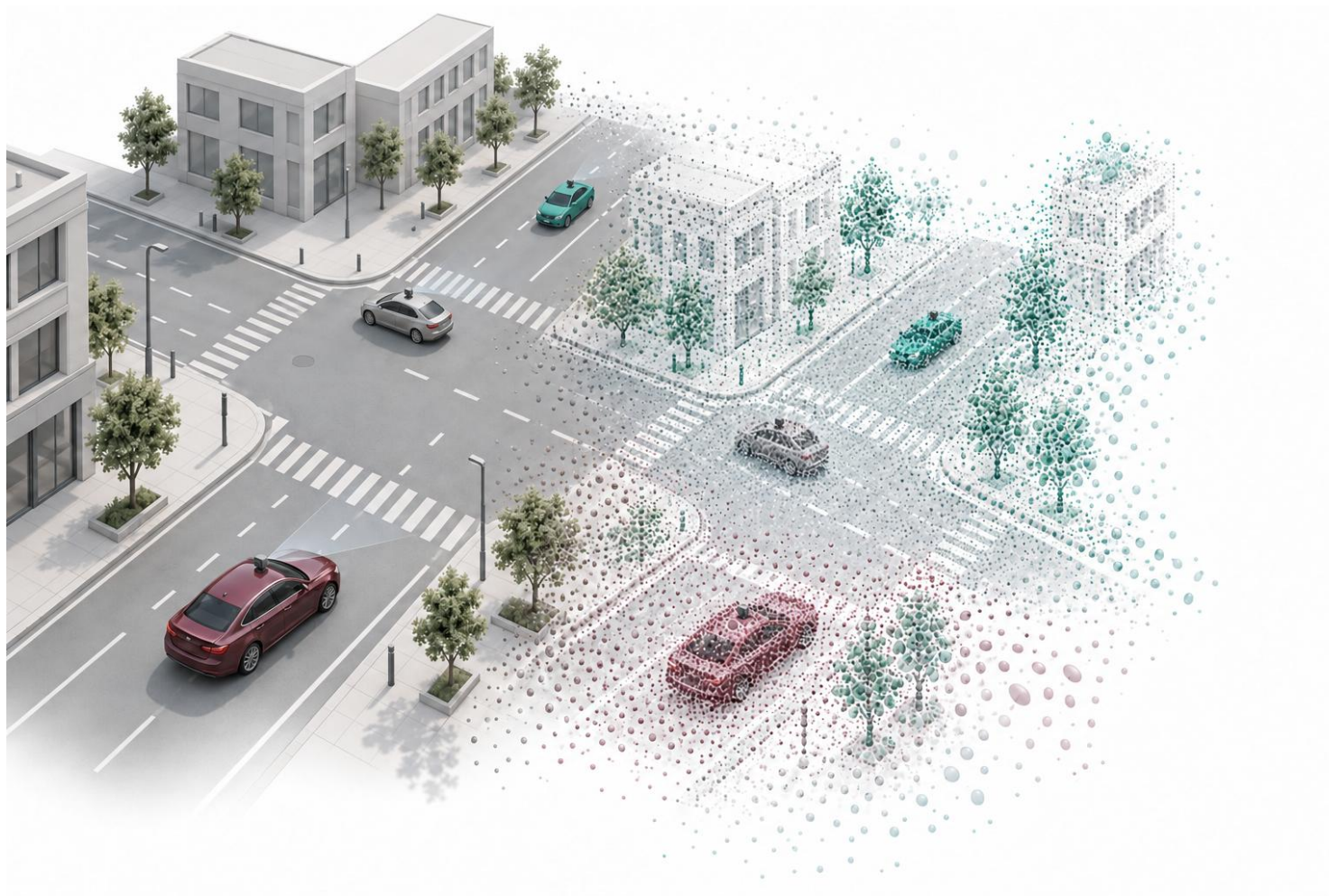
Feedforward Dynamic Scene Reconstruction from Uncalibrated Collaborative Driving Views

Yihang Tao, Yu Guo, Zhengru Fang, Haonan An, Yuguang Fang

Hong Kong JC STEM Lab of Smart City, City University of Hong Kong

NeurIPS 2026

Dynamic reconstruction for driving



On-board reasoning

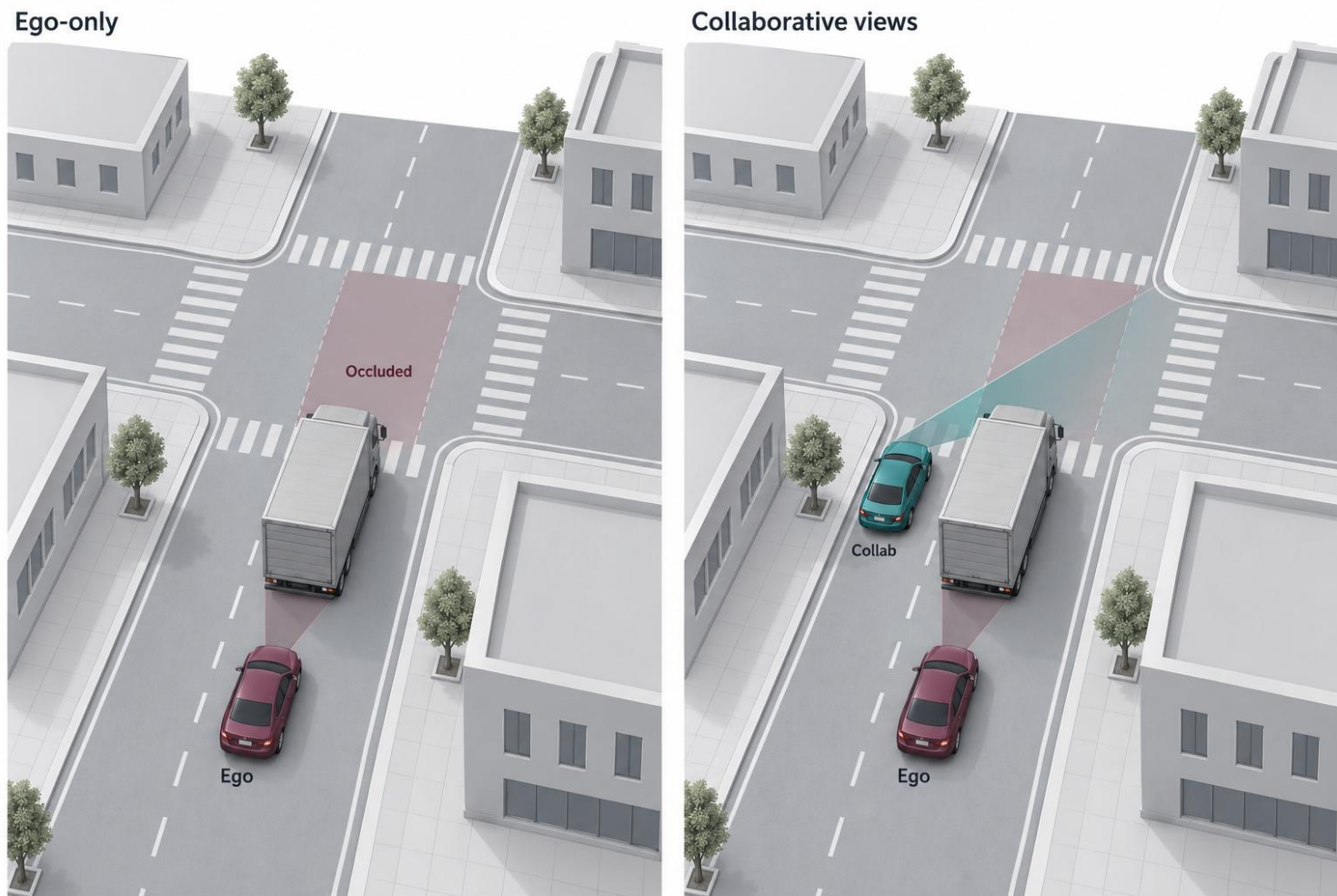
Novel-view synthesis and what-if analysis from sparse driving observations.

Simulation at scale

Driving logs become dynamic digital twins for closed-loop simulation and evaluation.

Feedforward 3DGS enables single-pass reconstruction. Unseen regions still need complementary evidence.

The ego view leaves blind spots



Removing an occluder exposes missing geometry unless the background is observed.

Physical occlusion

Foreground objects hide parts of the static background.

Complementary views

A second agent may observe the geometry that the ego camera cannot see.

Why collaboration is difficult

Unstructured geometric misalignment



Collaborative view



Ego view

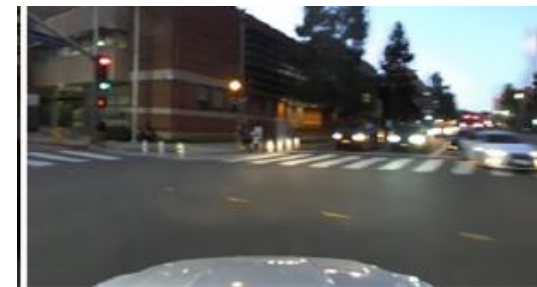
Independent cameras observe different viewpoints. Uncalibrated motion and changing overlap complicate geometric alignment.

Complete blind spots while preserving reliable ego-visible geometry.

Destructive semantic interference



DGGT

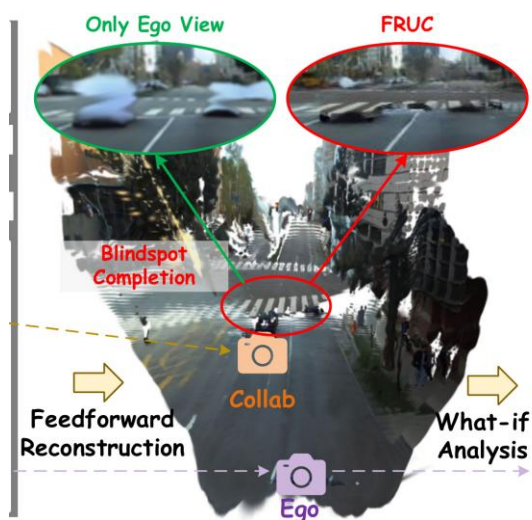


FRUC

Cross-agent context can disturb reliable ego geometry. FRUC targets these artifacts through ego-conditioned correction.

Contributions

01 Uncalibrated reconstruction

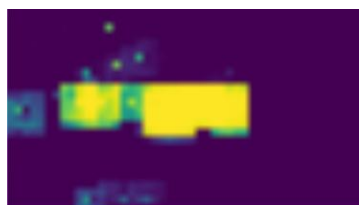


One feedforward pass reconstructs a dynamic scene from uncalibrated collaborative views.

02 Ego-guided supplementation



Dynamic footprint



Occlusion prior

Causal occlusion reasoning and residual denoising recover hidden geometry while protecting reliable observations.

03 Evidence on real-world datasets



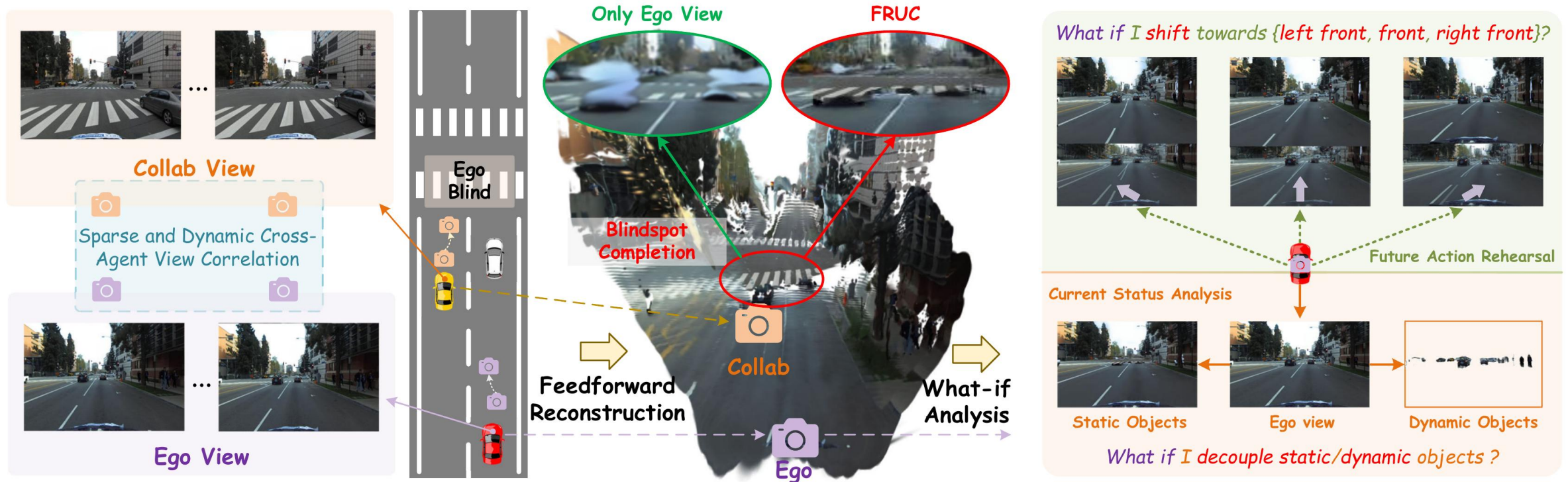
V2X-Real



UrbanIng-V2X

V2X-Real and UrbanIng-V2X evaluations cover rendering quality, blind-spot completion and generalization.

Problem setup



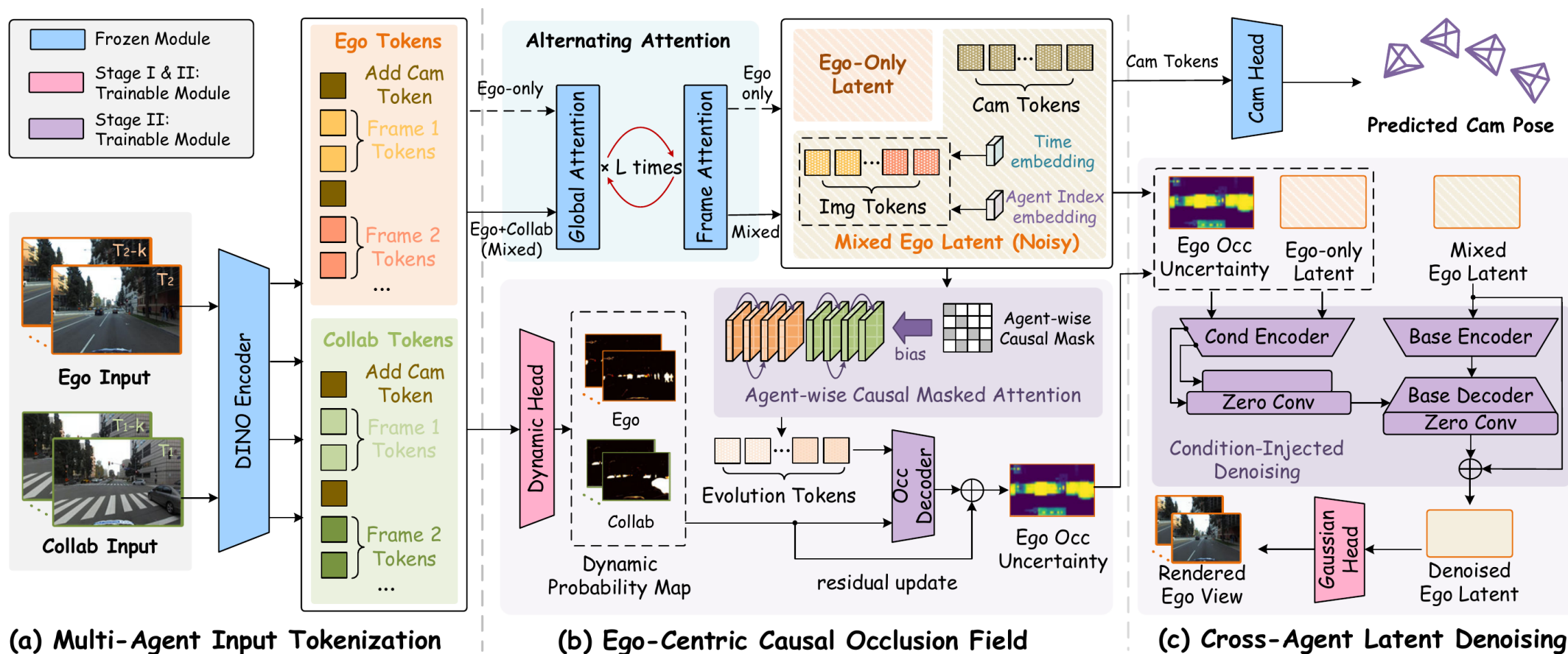
Input: ordered collaborative RGB views

Ego/collab identity and frame order are known.
Precise inter-agent calibration is not required.

Output: an ego-centric dynamic scene

Static and dynamic 3D Gaussians support novel views, scene editing and what-if analysis.

Methodology: FRUC framework



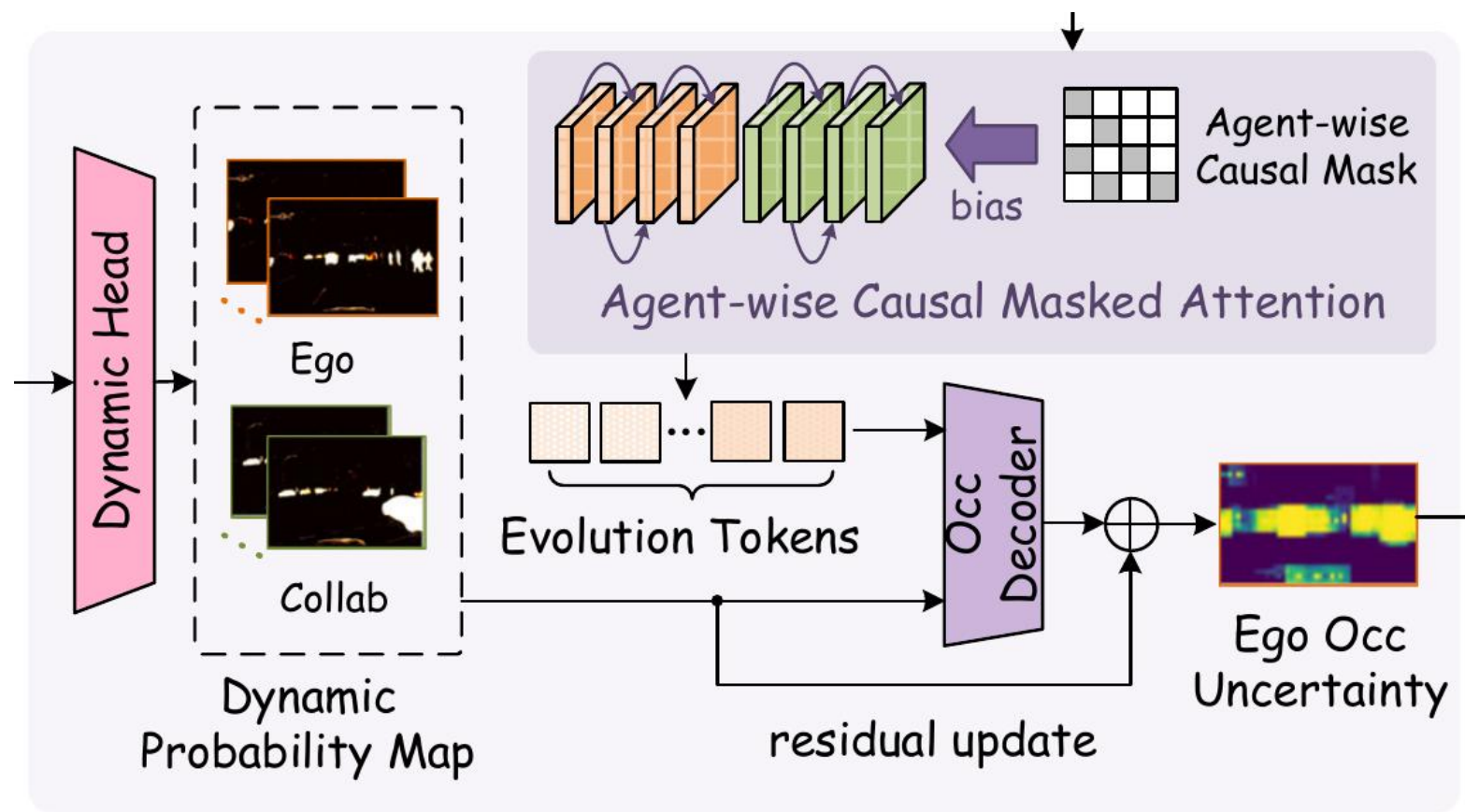
Locate uncertain geometry

The causal occlusion field models where moving objects hide the scene.

Refine the collaborative latent

Ego-conditioned residual denoising supplements blind spots while preserving visible geometry.

Ego-Centric Causal Occlusion Field



(b) Ego-Centric Causal Occlusion Field

Dynamic footprint

Shallow DINO features identify movable objects.

Causal motion prior

Same-agent attention uses past and current frames to expand the footprint along motion.

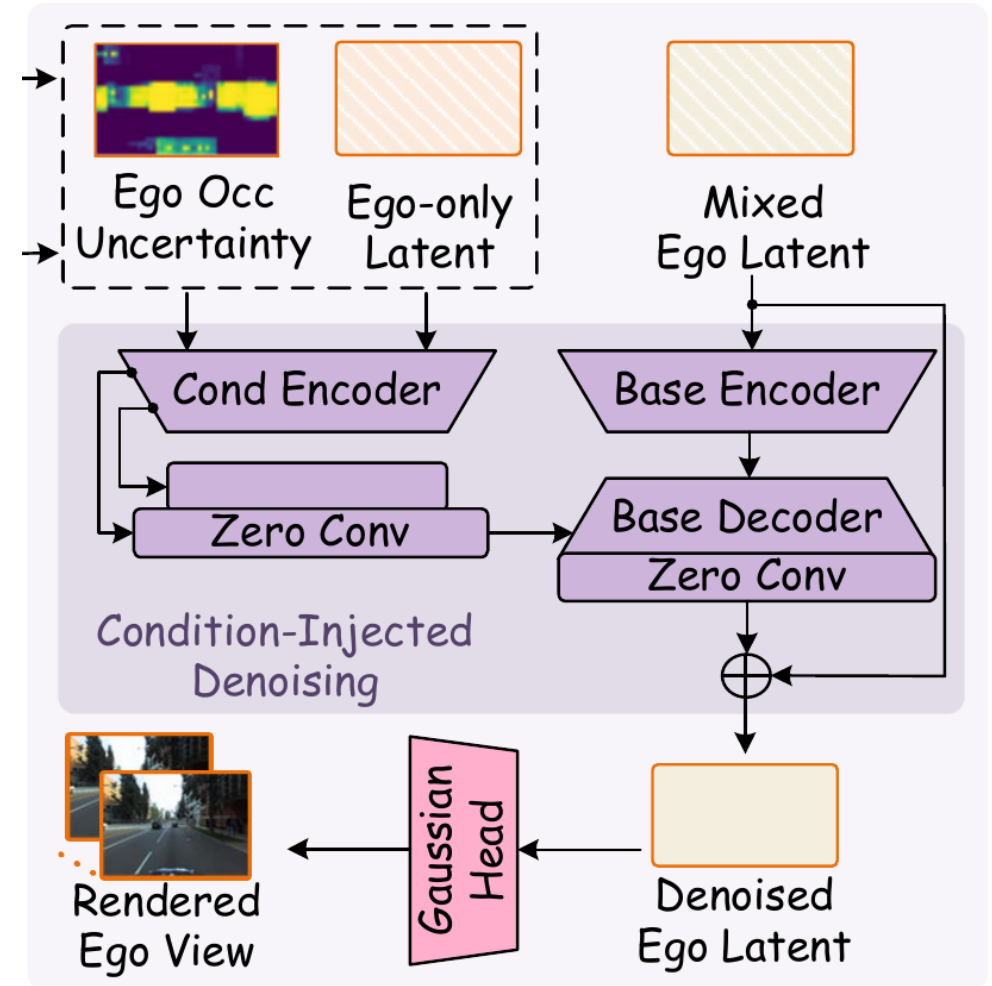
Cross-Agent Latent Residual Denoising

Ego-only reference

The conditional branch encodes reliable ego geometry together with the occlusion prior.

Targeted residual correction

Zero-initialized convolutions stabilize adaptation and protect reliable visible geometry.

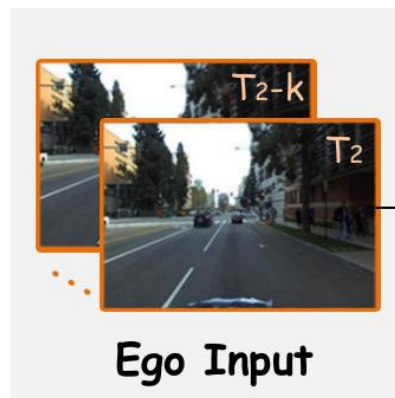


(c) Cross-Agent Latent Denoising

Two-stage training

I

Single-Agent Pre-training



Frozen

VGGT backbone

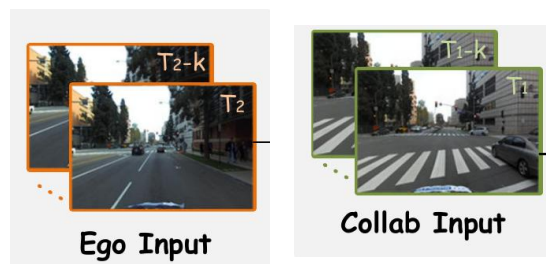
Train

Gaussian, sky and dynamic heads

Single-agent sequences. Photometric L1 + LPIPS and dynamic-mask BCE supervision.

II

Cross-Agent Adaptation



Frozen

VGGT, dynamic and sky heads

Train

FRUC modules and remaining decoders

Initialize from Stage I. Collaborative sequences with photometric and auxiliary objectives.

Experiments: datasets and protocols

Dataset	Split	Original data					Added / parsed metadata				Samples (K)
		RGB	Frame ID	Camera params	Ego pose	Scene context	LiDAR	Sky mask	Dynamic mask	View association	
V2X-Real	Train	Yes	Yes	No	No	No	No	Yes	Yes	No	103.5
V2X-Real	Val	Yes	Yes	No	No	No	No	Yes	Yes	Yes	29.1
UrbanIng-V2X	Train	Yes	Yes	No	No	No	No	Yes	Yes	No	114.5
UrbanIng-V2X	Val	Yes	Yes	No	No	No	No	Yes	Yes	Yes	44.6

Multi-frame interpolation

Ego + collab at $t - 1$ and $t + 1$ reconstruct the ego view at t . This is the default setting.

Single-frame extrapolation

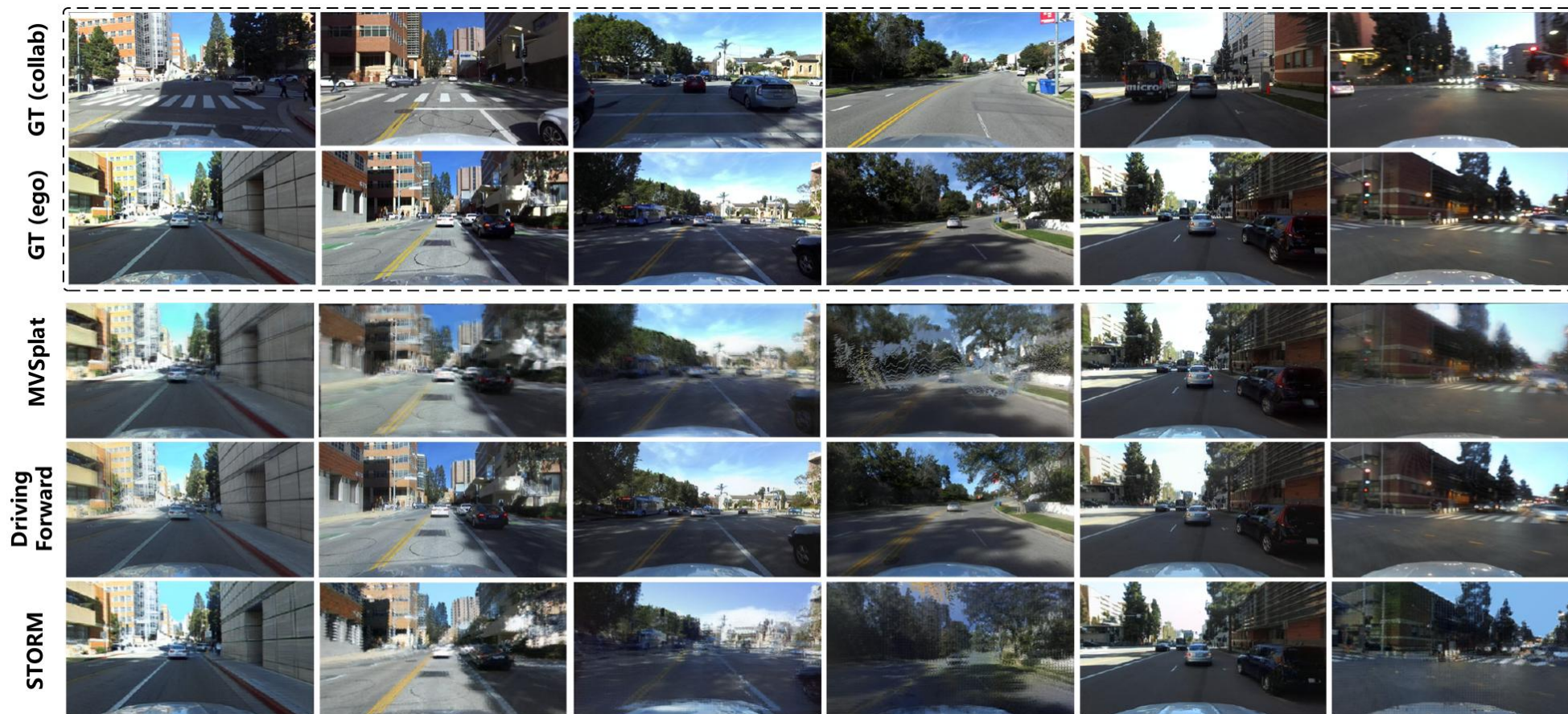
Ego + collab at t predict the ego view at $t + 1$. Counts above are thousands of view pairs.

V2X-Real: quantitative comparison

Method	Full image			Dynamic-only			Blind-spot	Inference
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	NIQE $\downarrow$	Time $\downarrow$
EmerNeRF	21.65	0.636	0.583	17.79	0.979	0.043	—	10 min
3DGS	20.19	0.616	0.592	16.85	0.976	0.035	—	14 min
V2X-Gaussians	<u>22.20</u>	0.784	0.409	23.87	0.994	<u>0.011</u>	—	21 min
MVSplat	16.73	0.478	0.281	15.18	0.952	0.049	—	0.31 s
DrivingForward	17.37	0.596	0.239	14.46	0.955	0.047	—	0.15 s
STORM	16.27	0.579	0.379	14.36	0.953	0.069	8.682	0.35 s
AnySplat	19.18	0.605	0.219	16.69	0.959	0.046	—	<u>0.29 s</u>
DGGT	21.85	<u>0.804</u>	<u>0.129</u>	<u>24.76</u>	<u>0.992</u>	0.129	<u>4.301</u>	0.54 s
FRUC	25.89	0.883	0.076	25.93	0.994	0.007	3.949	0.77 s

FRUC: 25.89 dB full-image PSNR, 3.949 blind-spot NIQE and 0.77 s inference.

V2X-Real: reconstruction quality



Intermediate ego timestamp

Uncalibrated ego and collaborative observations reconstruct the held-out ego view.

Challenging driving geometry

Occlusion, changing viewpoints and dynamic traffic make cross-agent context difficult to align.

V2X-Real: collaborative detail recovery



More faithful scene structure

The examples retain road layout and object boundaries across the displayed scenes.

Complementary geometry

FRUC uses collaborative context to supplement ego observations and improve local detail.

UrbanIng-V2X: zero-shot transfer

Method	Full image			Dynamic-only			Blind-spot
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	NIQE $\downarrow$
MVSplat	18.58	0.463	0.380	15.15	0.966	0.056	-
DrivingForward	18.09	0.676	0.335	15.66	0.970	0.037	-
STORM	17.30	0.604	0.458	14.16	0.967	0.068	11.781
AnySplat	21.31	0.552	<u>0.222</u>	17.88	0.974	<u>0.028</u>	-
DGGT	<u>22.23</u>	<u>0.724</u>	0.286	<u>23.71</u>	0.993	0.208	<u>7.509</u>
FRUC	25.07	0.795	0.144	24.61	0.993	0.008	5.270

Transfer without fine-tuning

Direct inference uses weights trained on V2X-Real.

Strong cross-dataset quality

FRUC reaches 25.07 dB PSNR and 5.270 NIQE on UrbanIng-V2X.

UrbanIng-V2X: visual generalization



A new driving domain

Day and night scenes test transfer beyond the V2X-Real training distribution.

Two evaluation settings

The paper compares direct zero-shot transfer with fine-tuning on UrbanIng-V2X.

UrbanIng-V2X: fine-tuned reconstruction

Method	Full image			Dynamic-only			Blind-spot
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	NIQE $\downarrow$
MVSplat	19.24	0.467	0.345	15.86	0.969	0.039	-
DrivingForward	21.80	<u>0.754</u>	0.258	16.71	0.974	<u>0.028</u>	-
STORM	17.36	0.614	0.336	14.22	0.969	0.079	6.268
AnySplat	22.45	0.579	<u>0.176</u>	19.20	<u>0.976</u>	0.025	-
DGGT	<u>25.37</u>	0.745	0.275	<u>24.63</u>	0.993	0.285	<u>5.728</u>
FRUC	27.26	0.996	0.005	27.64	0.819	0.079	3.854

Improved full-image quality

FRUC reaches 27.26 dB full-image PSNR and 3.854 blind-spot NIQE after fine-tuning.

Metric-specific differences

Other baselines lead dynamic-only SSIM and LPIPS.

UrbanIng-V2X: structure and boundaries



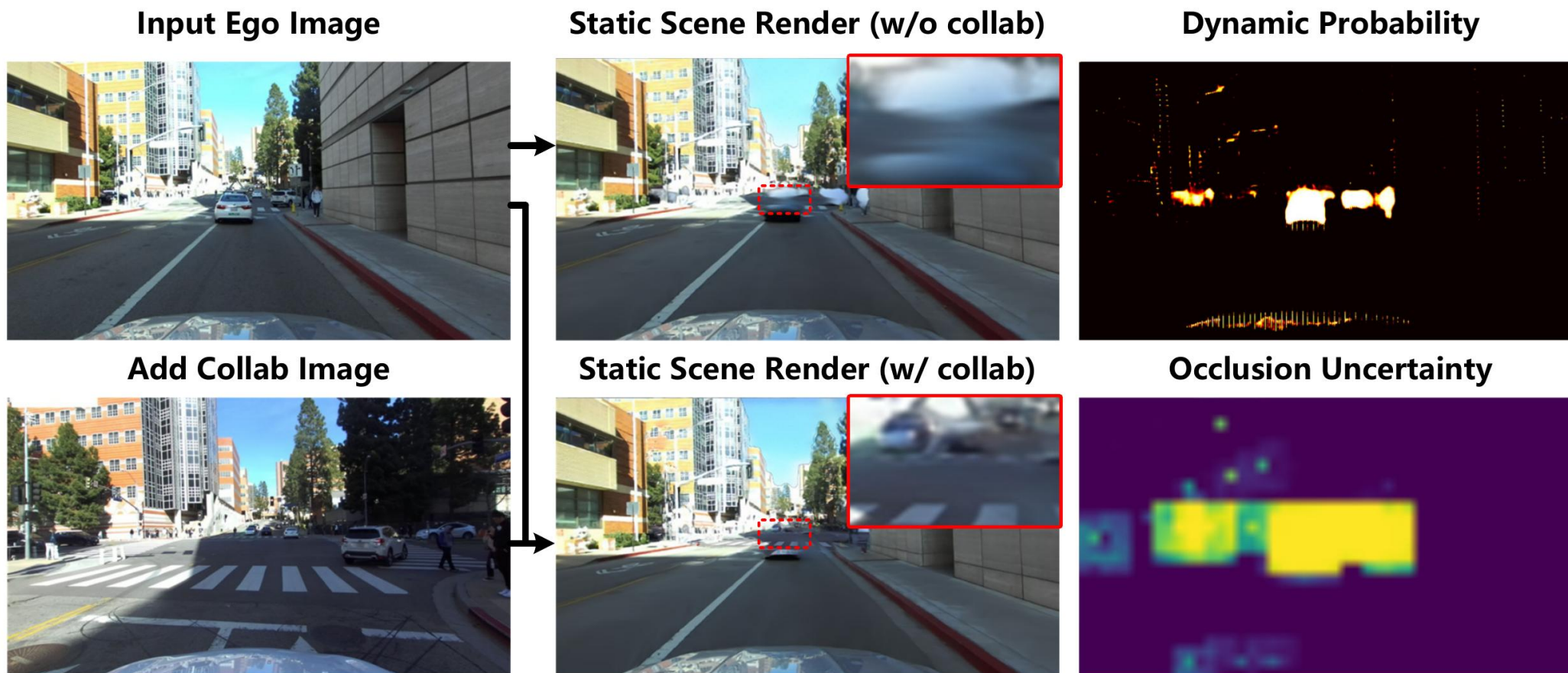
Zero-shot scene structure

FRUC preserves clearer scene layout and object boundaries in the shown examples.

Benefits of fine-tuning

Local sharpness and geometric consistency improve after adapting to the new dataset.

Blind-spot completion



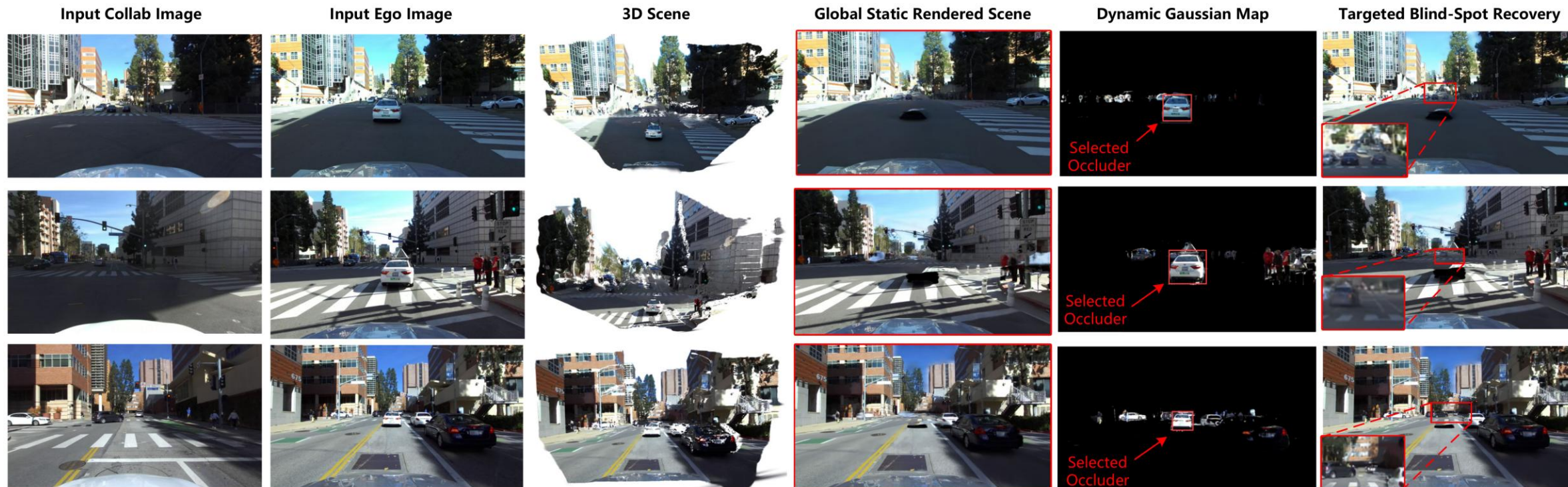
Reveal the static background

Removing dynamic Gaussians exposes geometry hidden by foreground objects.

Use collaborative evidence

Additional views reduce holes and blurring. Blind-spot NIQE: 3.949 (FRUC) vs. 4.301 (DGGT).

Scene editing and what-if analysis



Remove selected occluders

Static/dynamic decoupling supports targeted removal of foreground objects.

Recover the hidden background

Collaborative geometry supports the edited view and targeted blind-spot recovery.

Ablation study

CALRD	COF	Aux.	Full image			Dynamic-only			Blind-spot
			PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	NIQE $\downarrow$
No	No	Yes	21.94	0.799	0.138	24.71	0.985	0.098	4.181
Yes	No	Yes	<u>23.97</u>	<u>0.836</u>	<u>0.119</u>	<u>24.96</u>	<u>0.989</u>	<u>0.077</u>	<u>4.017</u>
Yes	Yes	No	22.76	0.801	0.121	24.92	0.987	0.084	4.309
Yes	Yes	Yes	25.89	0.883	0.076	25.93	0.994	0.007	3.949

Each component matters

CALRD, COF and auxiliary objectives jointly improve the reconstruction.

Full configuration

The complete model gives the best reported quality and lowest NIQE among these variants.

Single-frame future-view prediction

Method	Full image			Dynamic-only			Blind-spot
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	NIQE $\downarrow$
DrivingForward	15.02	<u>0.682</u>	0.289	<u>20.02</u>	0.986	0.017	-
AnySplat	<u>19.04</u>	0.603	<u>0.224</u>	16.60	<u>0.979</u>	0.031	-
DGGT	18.02	0.585	0.226	16.57	0.978	0.209	<u>6.021</u>
FRUC	21.47	0.773	0.156	20.50	<u>0.979</u>	<u>0.022</u>	3.987

Extrapolation from one timestamp

Ego and collaborative views at t predict the ego view at $t + 1$.

Quality across regions

FRUC leads full-image metrics, dynamic PSNR and NIQE. DrivingForward leads dynamic SSIM / LPIPS.

Number of collaborative input views

Method	Inputs		Full image			Dynamic-only			Blind-spot
	Ego	Collab	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	NIQE $\downarrow$
DGGT	2	0	24.80	0.838	0.179	27.48	0.995	0.105	4.938
DGGT	2	2	21.85	0.804	0.129	24.76	0.992	0.129	4.301
DGGT	2	4	20.77	0.773	0.229	23.12	0.991	0.161	4.185
FRUC	2	0	27.02	0.893	0.006	27.93	0.995	0.005	4.985
FRUC	2	2	25.89	0.883	0.076	25.93	0.994	0.007	3.949
FRUC	2	4	23.33	0.839	0.117	24.54	0.993	0.008	3.936

Improved blind-spot completion

For FRUC, NIQE improves from 4.985 to 3.936 as the collaborative input count increases.

A reconstruction tradeoff

Full-image PSNR decreases. More external context does not improve every metric.

Takeaways and project resources

Targeted collaboration

FRUC completes dynamic ego blind spots using uncalibrated collaborative views in one forward pass.

Current limitations

Severe occlusion with weak cross-agent semantic correlation, and inaccurate dynamic maps, can still limit reconstruction and editing.



Project Page

yihangtao.github.io/FRUC

[Paper](#)

[Code](#)

[Weights](#)