1 INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD
2 Mohamed bin Zayed University of Artificial Intelligence
3 UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386
4 vlm run
5

Leveraging 2D Priors and SDF Guidance for Urban Scene Rendering

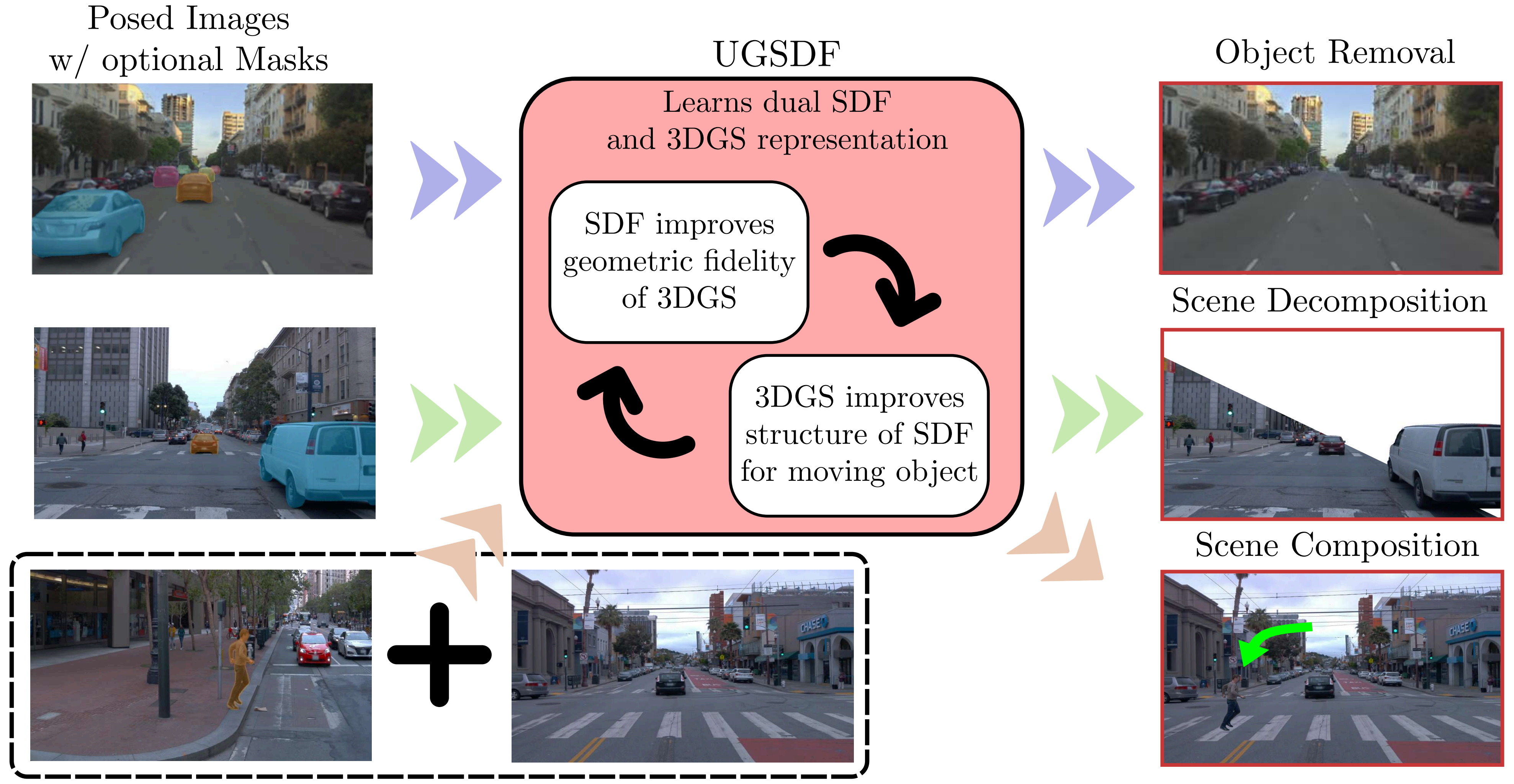
Siddharth Tourani[2,3], Jayaram Reddy[1], Akash Kumbar[1], Satyajit Tourani[1],
Nishant Goyal[5], Madhava Krishna[1], N Dinesh Reddy[4], Muhammad Haris Khan[2]

https://dynamic-ugsdf.github.io/

ICCV HONOLULU HAWAII
OCT 19-23, 2025

Overview: Rendering Dynamic Scenes from 2D Cues

Problem: Dynamic urban scene reconstruction need heavy 3D(LiDAR, tracklets, templates).
Our Solution: UGSDF fuses **SDF precision** with **3DGS realism**, supervised by 2D priors.
Key Mechanism: Bidirectional guidance: SDFs structure 3DGS; 3DGS refines detail.
Outcomes: Fewer 3D requirements, sharper geometry, photoreal renderings.


Posed Images w/ optional Masks

UGSDF

Learns dual SDF and 3DGS representation

SDF improves geometric fidelity of 3DGS

3DGS improves structure of SDF for moving object

Object Removal

Scene Decomposition

Scene Composition

UGSDF's SDF↔3DGS synergy enables high-quality object removal and scene (de)composition.

Quantitative Results

Method	Scene Reconstruction PSNR ↑	SSIM ↑	LPIPS ↓	Novel View Synthesis PSNR ↑	SSIM ↑	LPIPS ↓
StreetGS [90]	29.11	0.921	0.117	25.71	0.764	0.218
S3Gaussians [29]	31.35	0.911	0.106	26.82	0.788	0.226
4DGF [16]	29.08	0.929	0.110	28.31	0.859	0.206
OmniRe [13]	33.79	0.942	0.105	29.35	0.780	0.186
UGSDF (*Ours*)	34.19	0.944	0.104	30.63	0.871	0.129
UGSDF w/o LiDAR (*Ours*)	33.98	0.942	0.105	30.32	0.871	0.145

Table 1. Evaluation on the Waymo Open Dataset.

Method	Scene Reconstruction PSNR ↑	SSIM ↑	LPIPS ↓	Novel View Synthesis PSNR ↑	SSIM ↑	LPIPS ↓
StreetGS [90]	26.73	0.883	0.162	25.61	0.803	0.211
4DGF [16]	27.16	0.885	0.149	26.47	0.761	0.237
OmniRe [13]	27.95	0.895	0.147	26.59	0.816	0.191
UGSDF (*Ours*)	29.55	0.934	0.105	28.63	0.926	0.123

Table 2. Evaluation on the KITTI Dataset.

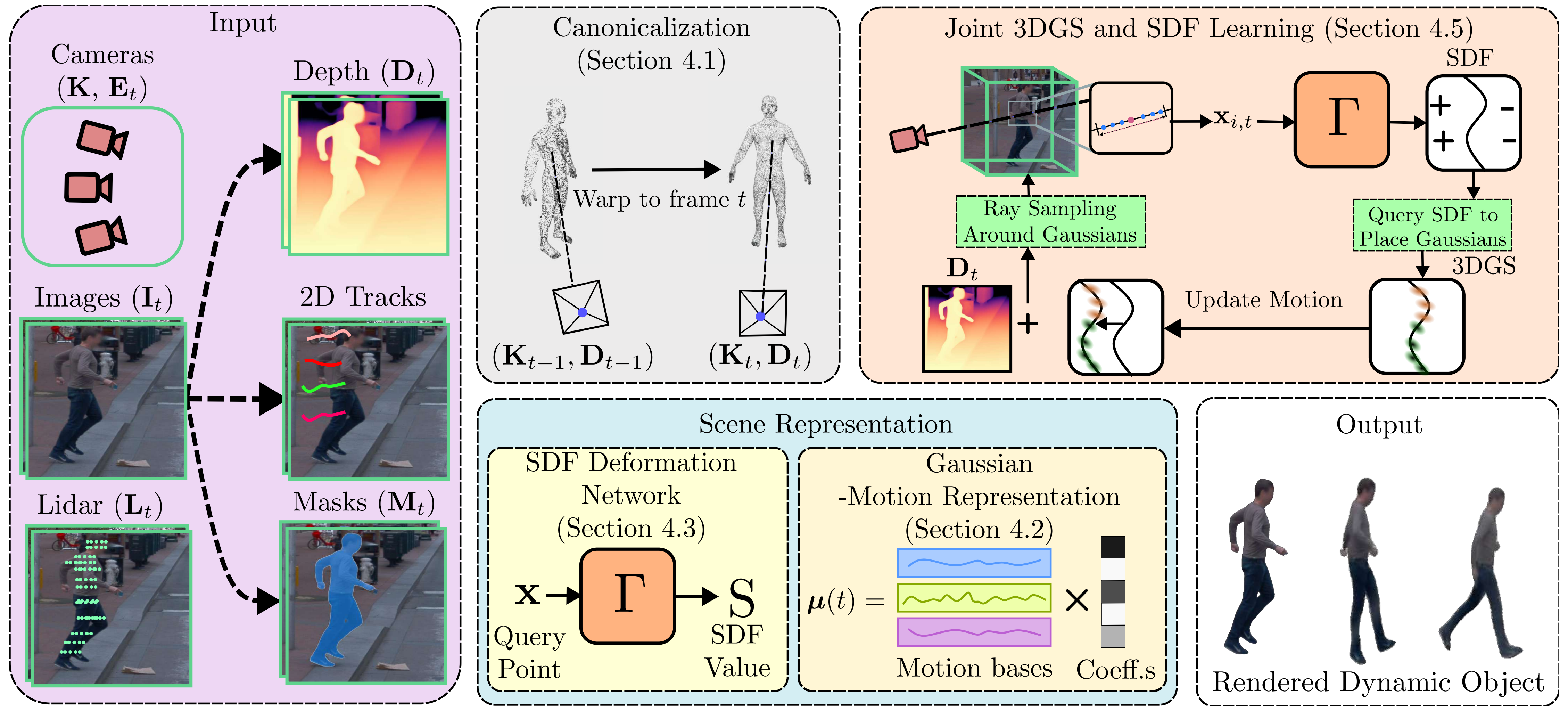
The *best* and *second* best results are highlighted.

Methods	Scene Reconstruction				Novel View Synthesis			
	Human		Vehicle		Human		Vehicle	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
EmerNeRF[92]	22.88	0.578	24.65	0.723	20.32	0.454	22.07	0.609
DeformGS[94]	17.80	0.460	19.53	0.570	17.30	0.426	18.91	0.530
PVG[12]	24.06	0.703	25.02	0.787	21.30	0.576	22.28	0.679
StreetGS[90]	16.83	0.420	27.73	0.880	16.55	0.393	26.71	0.846
OmniRe [13]	28.15	0.845	28.91	0.892	24.36	0.727	27.57	0.858
UGSDF (Ours)	27.89	0.839	30.34	0.906	25.48	0.758	28.68	0.872
UGSDF w/o LiDAR (Ours)	27.89	0.825	29.91	0.902	25.14	0.742	28.23	0.847

Table 3. Evaluation of methods on Human and Vehicle categories

UGSDF: Method at a Glance

Input: • Posed RGB • Object Masks • 2D Tracks • Metric Depth (± LiDAR)


Input

Cameras $(\mathbf{K}, \mathbf{E}_t)$

Depth $(\mathbf{D}_t)$

Images $(\mathbf{I}_t)$

2D Tracks

Lidar $(\mathbf{L}_t)$

Masks $(\mathbf{M}_t)$

Canonicalization (Section 4.1)

Warp to frame t

$(\mathbf{K}_{t-1}, \mathbf{D}_{t-1})$ $(\mathbf{K}_t, \mathbf{D}_t)$

Joint 3DGS and SDF Learning (Section 4.5)

SDF

$\mathbf{x}_{i,t}$ Γ

Ray Sampling Around Gaussians

Query SDF to Place Gaussians

$\mathbf{D}_t$

3DGS

Update Motion

Scene Representation

SDF Deformation Network (Section 4.3)

$\mathbf{x}$ Γ S

Query Point

SDF Value

Gaussian -Motion Representation (Section 4.2)

$\boldsymbol{\mu}(t) =$ $\times$

Motion bases

Coeff.s

Output

Rendered Dynamic Object

Overview Pipeline

4.1 Canonicalization

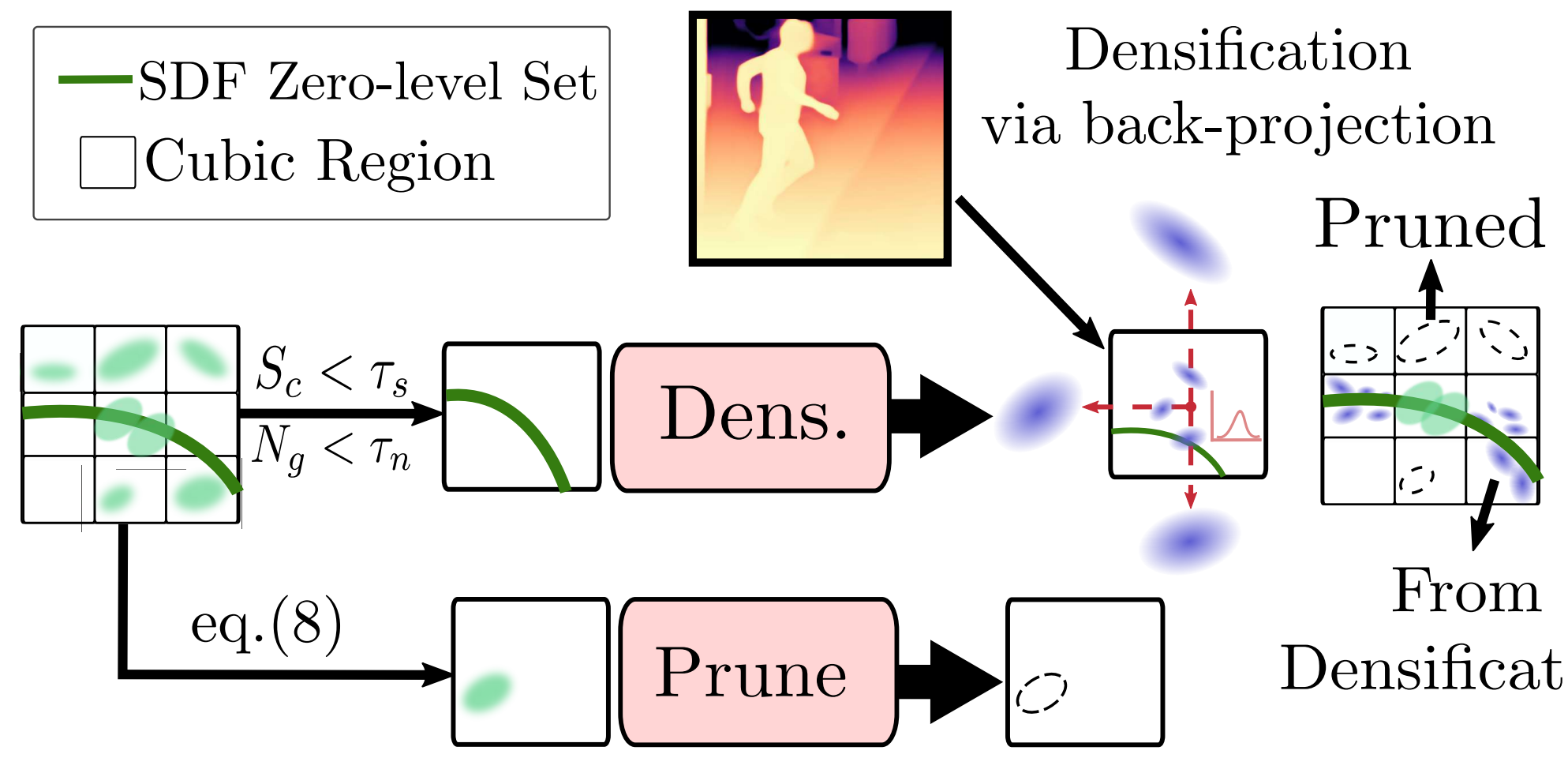
- Backproject depth → object point clouds
- Warp to canonical with 2D tracks; initialize 3D Gaussians

4.2 3DGS Motion

- Sparse basis for $\mu(t), q(t)$;
- Supervise with rendered image/depth/mask

4.3 SDF Deformation

- Multi-res feature grid + MLPs
- $\phi_{def}, \phi_{hyp}, \phi_{sdf}$ $(+\ \phi_{rgb})$ → SDF(+color)


SDF Zero-level Set

☐ Cubic Region

Densification via back-projection

$S_c < \tau_s$

$N_g < \tau_n$

Dens.

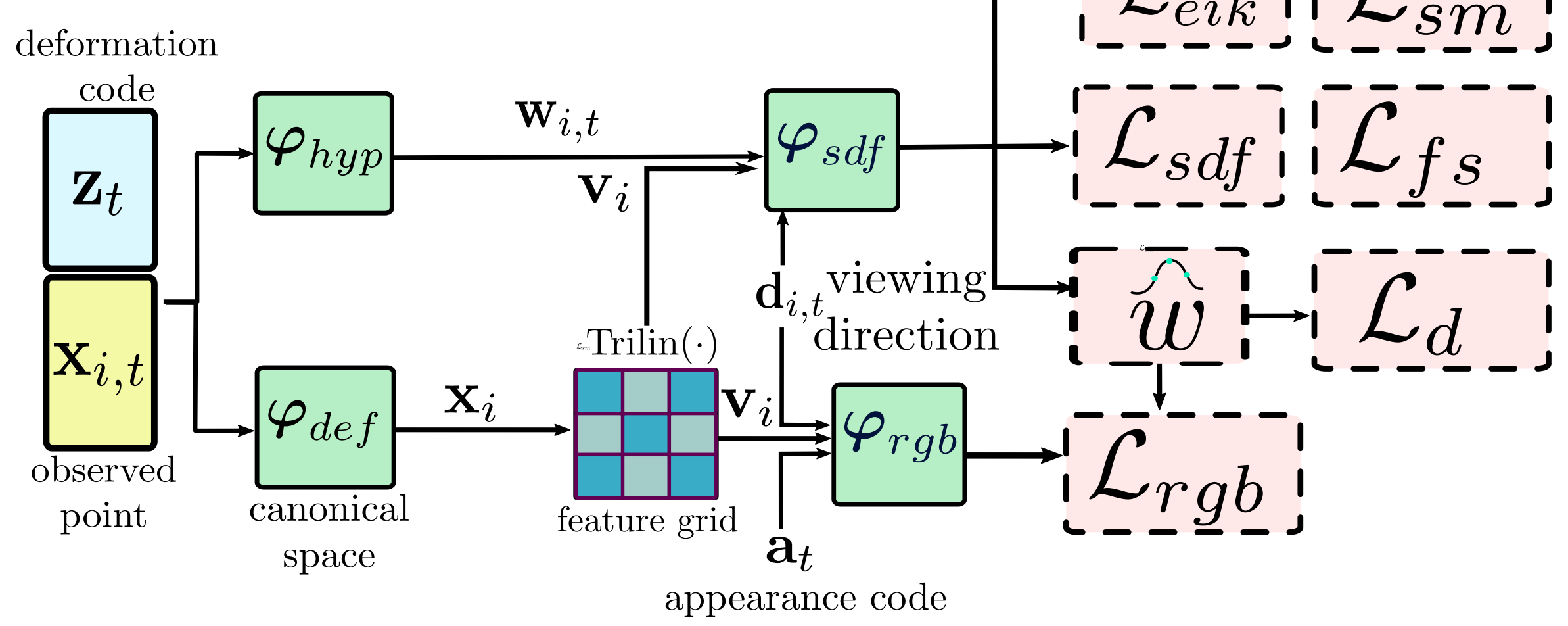
Pruned

eq.(8)

Prune

From Densification

SDF Guidance for Gaussian Primitive Distribution


deformation code

$\mathbf{z}_t$

$\mathbf{x}_{i,t}$

observed point

φ_{hyp}

φ_{def}

canonical space

$\mathbf{w}_{i,t}$

$\mathbf{v}_i$

$\mathbf{x}_i$

Trilin(·)

feature grid

φ_{sdf}

$\mathbf{d}_{i,t}$ viewing direction

φ_{rgb}

$\mathbf{a}_t$

appearance code

$\mathcal{L}_{eik}$ $\mathcal{L}_{sm}$

$\mathcal{L}_{sdf}$ $\mathcal{L}_{fs}$

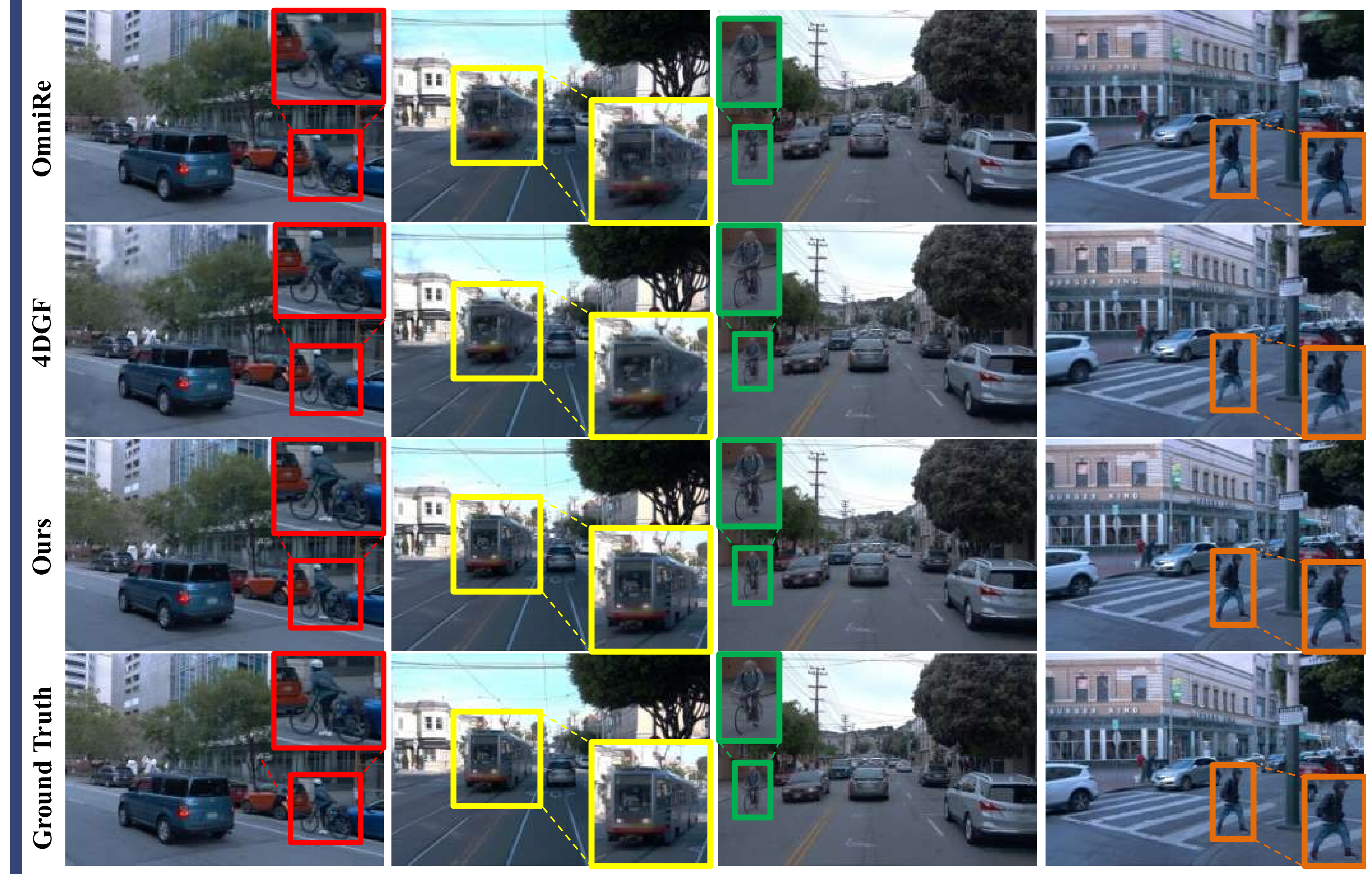
$\hat{w}$ $\mathcal{L}_d$

$\mathcal{L}_{rgb}$

SDF Deformation Network

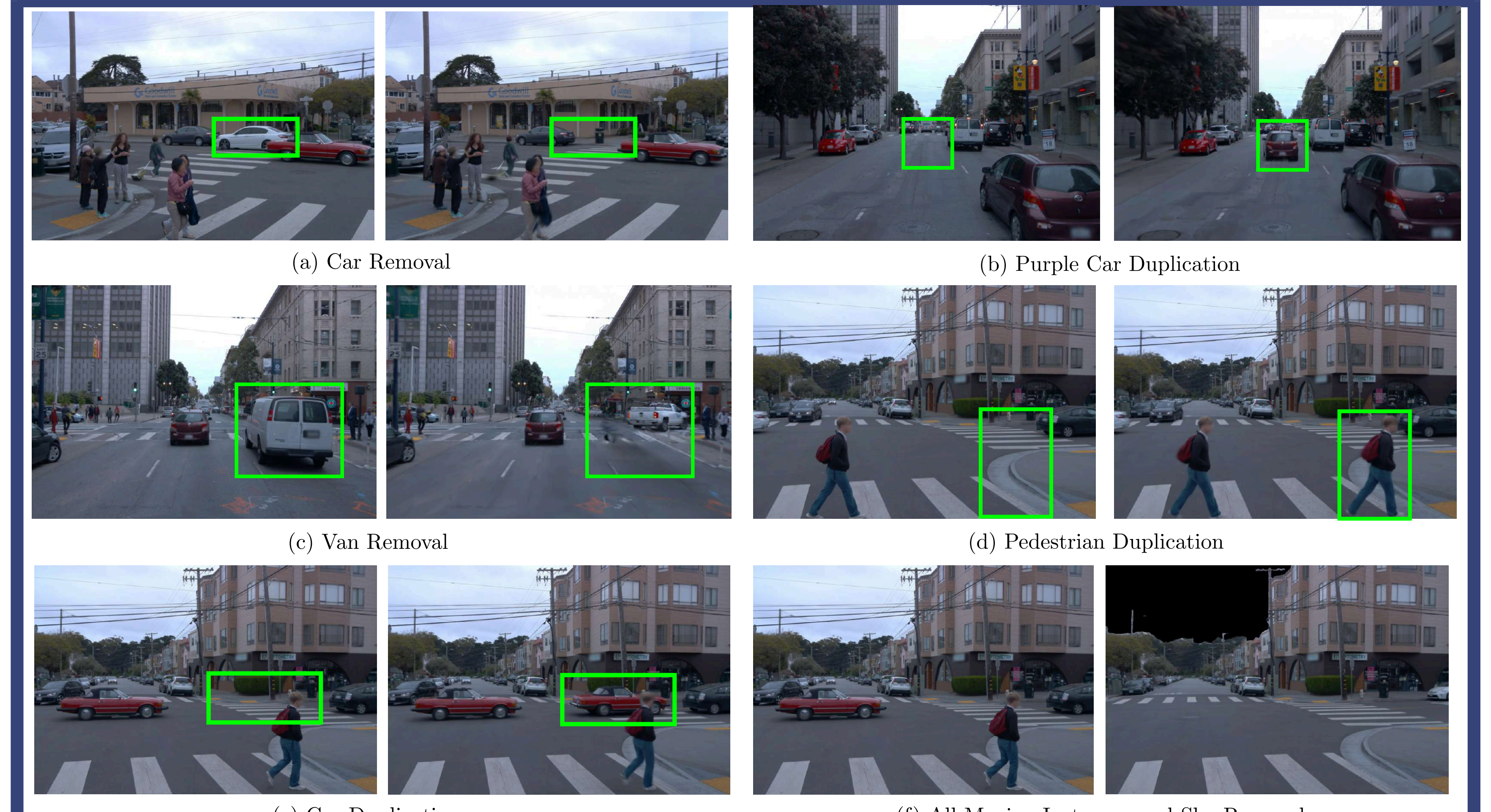
Takeaway
Dual representation + 2D Priors ⇒ robust geometry and high fidelity rendering — no 3D templates/tracklets

Qualitative Results: Cleaner Dynamics



Cleaner surfaces and steadier motion vs. baselines

Downstream Task: Edit & Simulate Dynamic Scenes


(a) Car Removal
(b) Purple Car Duplication
(c) Van Removal
(d) Pedestrian Duplication
(e) Car Duplication
(f) All Moving Instances and Sky Removal

Stable geometry enables robust dynamic editing