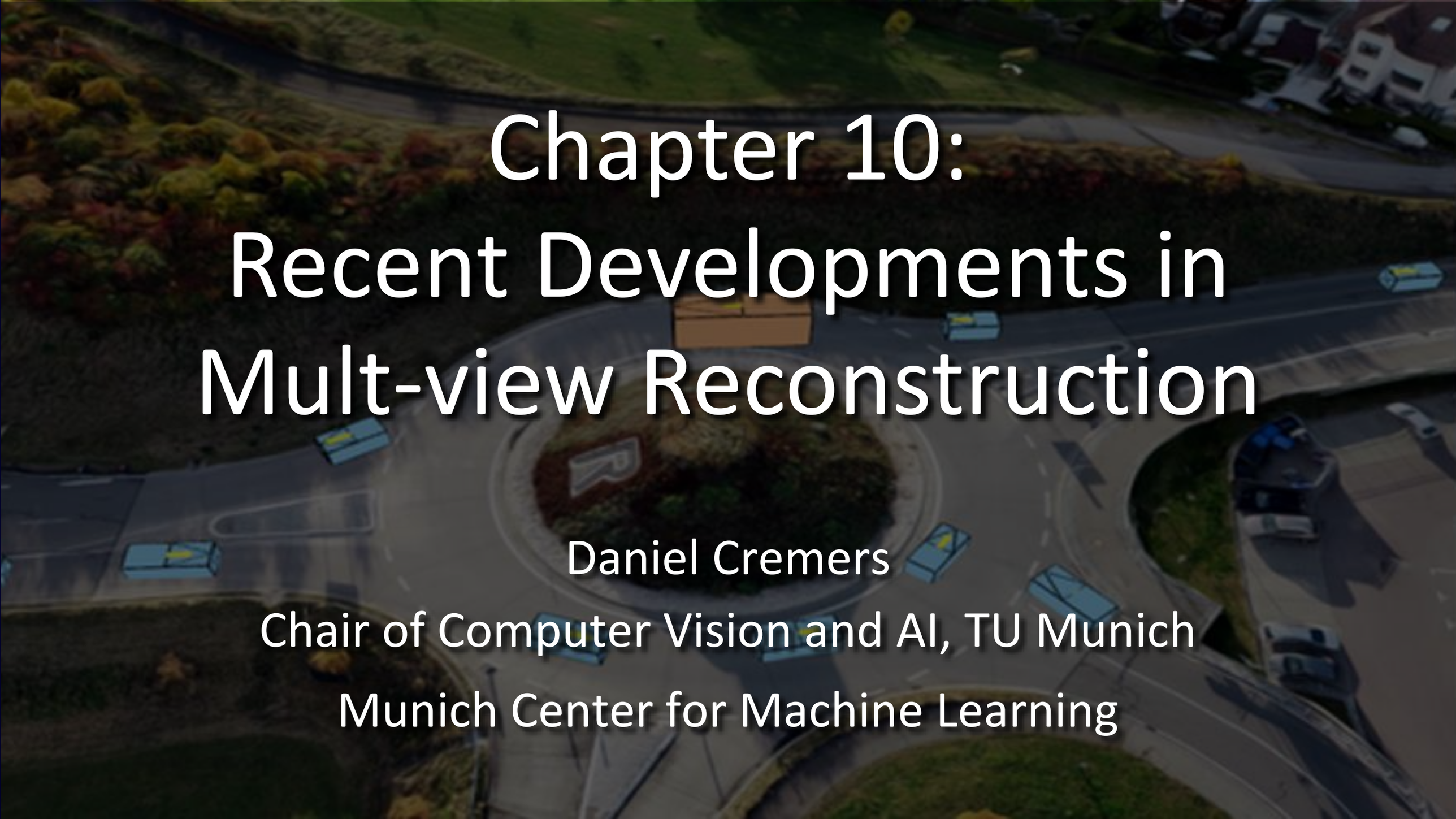


Chapter 10: Recent Developments in Mult-view Reconstruction

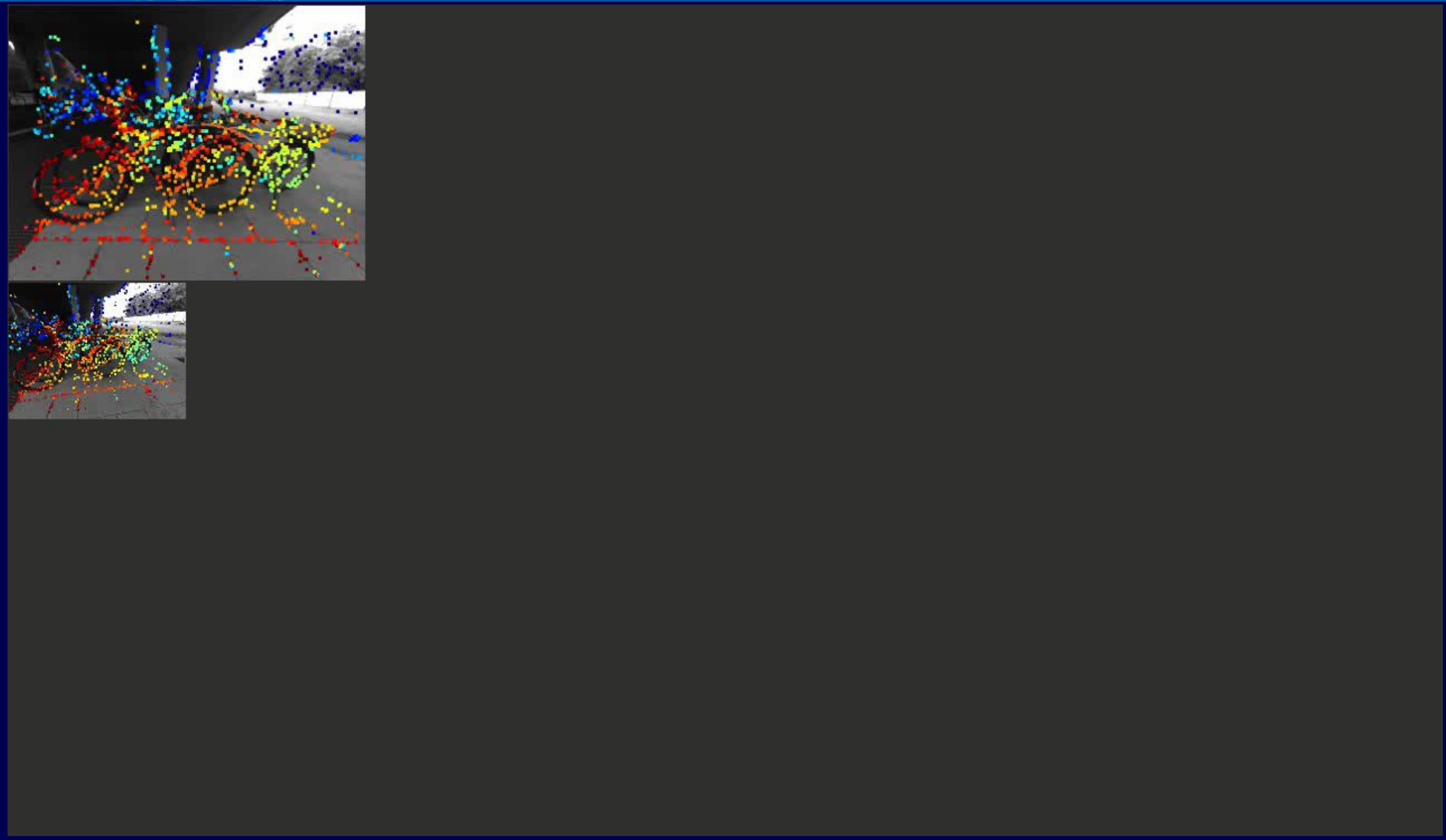
Daniel Cremers
Chair of Computer Vision and AI, TU Munich
Munich Center for Machine Learning

LSD SLAM: Large-Scale Direct SLAM

Engel, Schöps, Cremers, ECCV 2014:
LSD SLAM

Engel, Koltun, Cremers, PAMI 2018:
Direct Sparse Odometry

Direct Sparse Odometry



Engel, Koltun, Cremers, PAMI 2018

Sensor Integration: Stereo, IMU, RTK-GNSS,…

Mobile Mapping System:

- GPS-independent realtime localization
- Leveraging off-the-shelf cameras

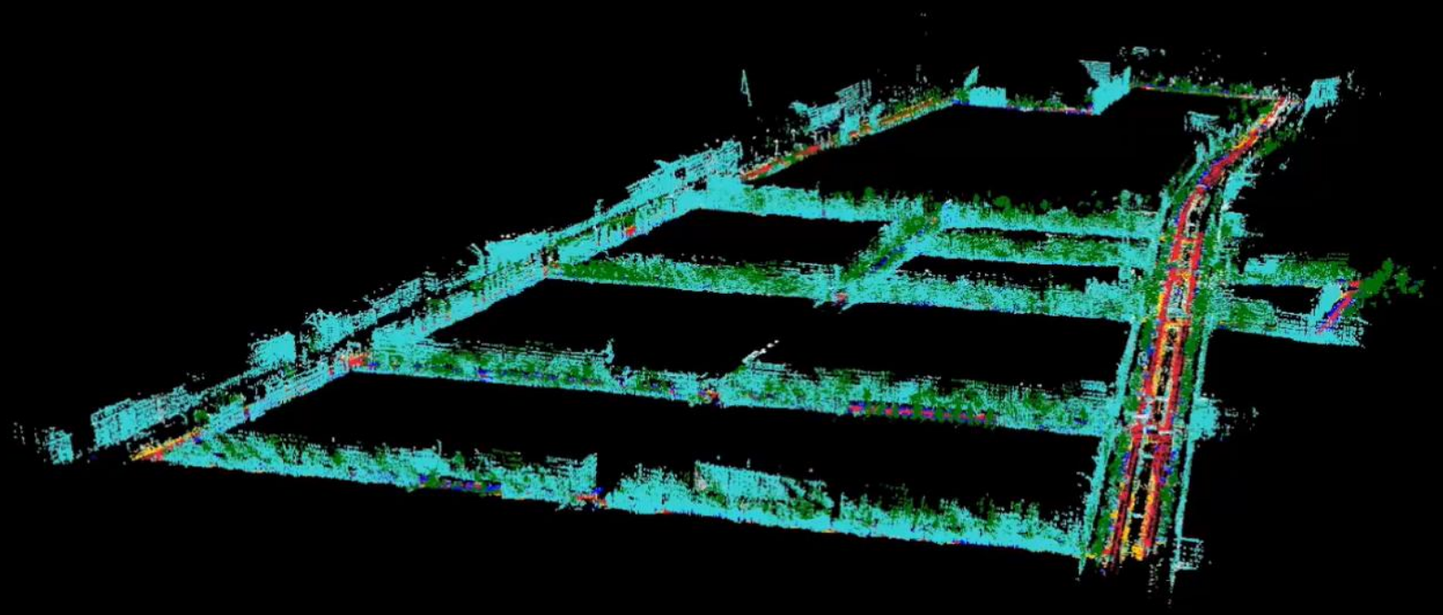
Spatial AI Cloud Platform:

- Access to navigation and map data
- Semantic scene understanding

Visual-inertial Odometry

Von Stumberg, Cremers, “Delayed Marginalization VI Odometry”, ICRA 2022

Realtime Large-Scale Semantic Mapping



drivable

sidewalk

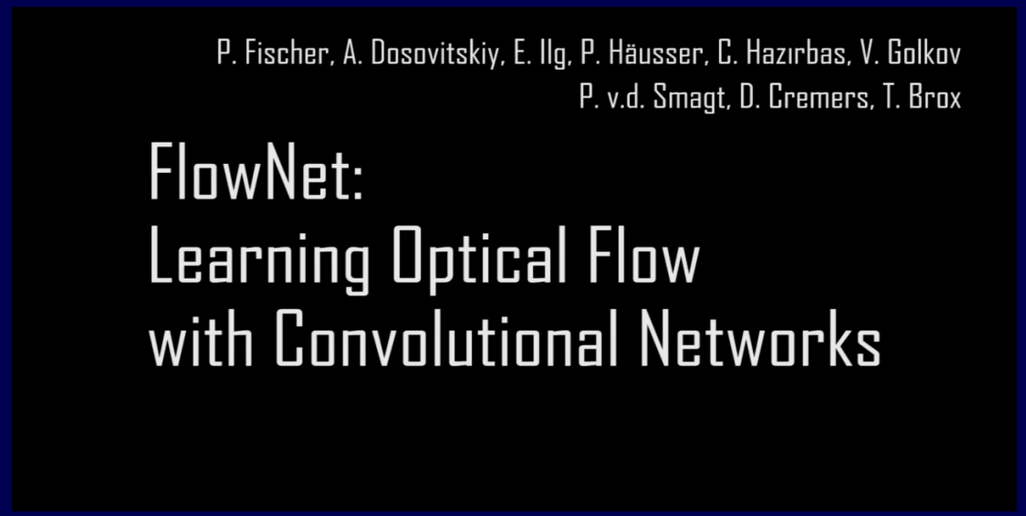
cars

buildings

vegetation

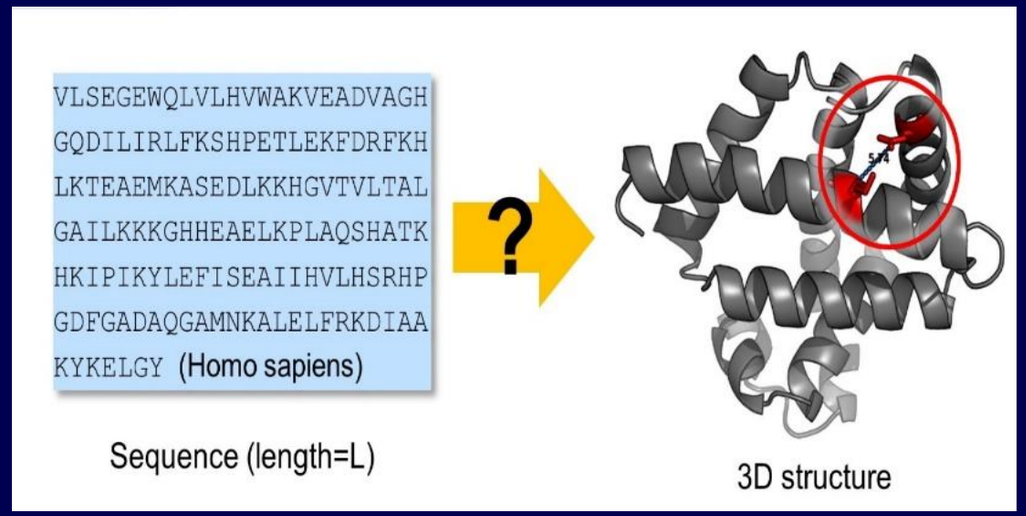
...

Deep Nets beyond Object Recognition



Dosovitsky et al., "FlowNet", ICCV 2015

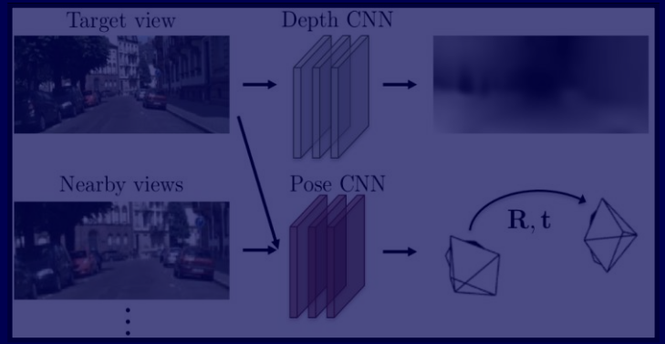
Badrinarayanan et al., "SegNet", arxiv'15



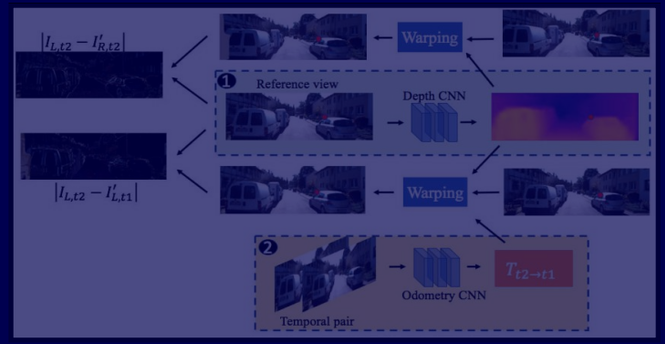
Golkov et al., NeurIPS '16

Caelles et al., CVPR 2018

Deep Nets for Visual SLAM



Zhou et al. CVPR 17

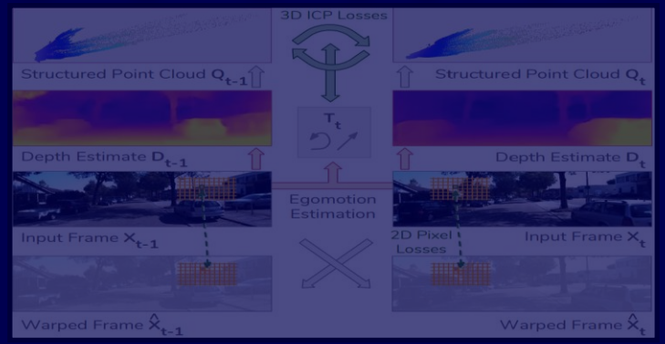


Zhan et al. CVPR 18

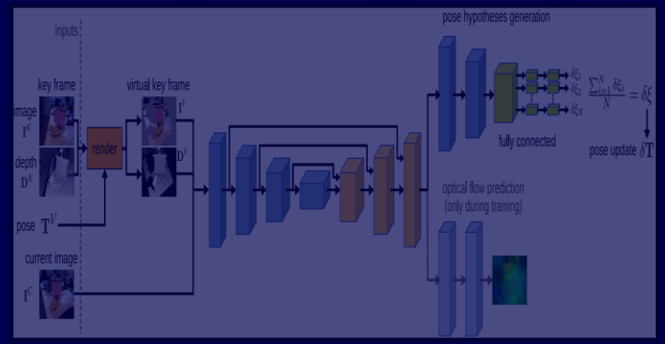
Initial methods were not state-of-the-art!

Ummenhofer et al. CVPR 17

Yin et al. CVPR 18



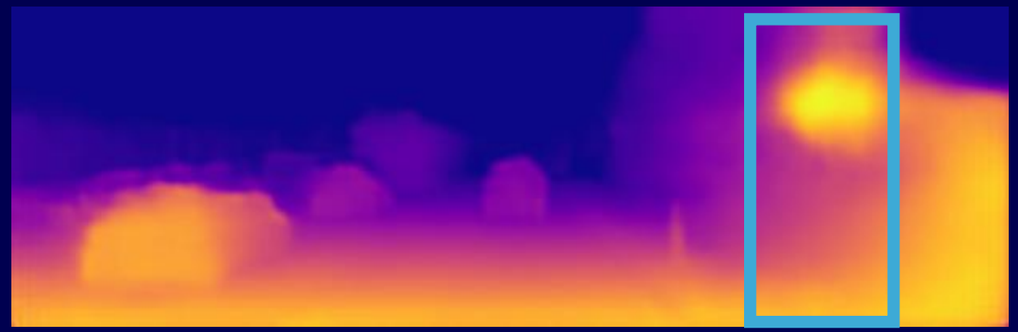
Mahjourian et al. CVPR 18



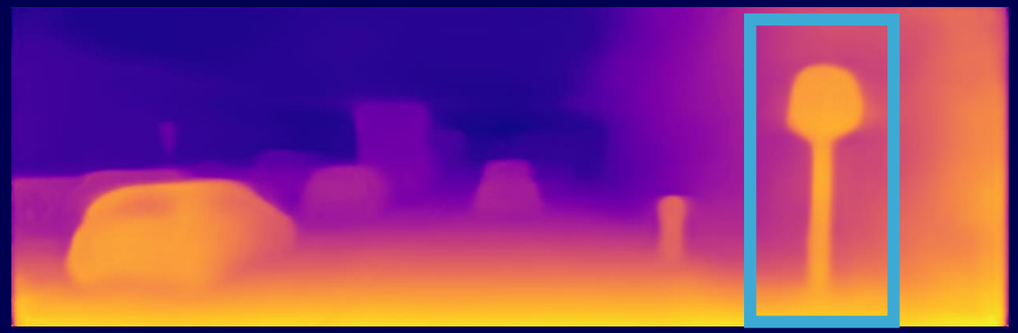
Zhou et al. ECCV 18

Deep Virtual Stereo Odometry

Deep Neural Network

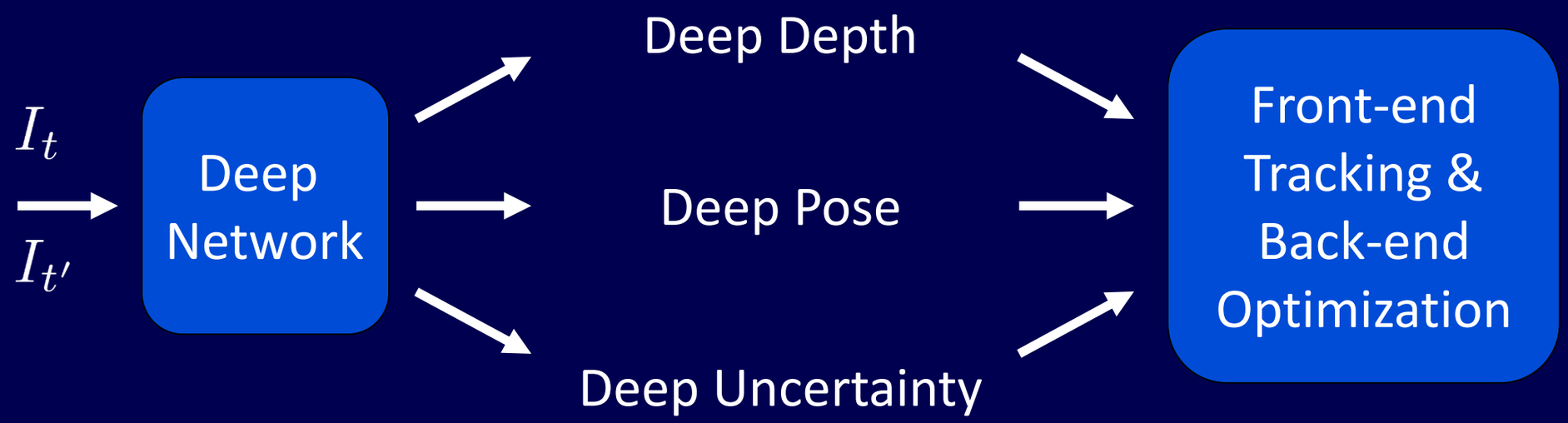


Kuznietsov et al. CVPR 2017



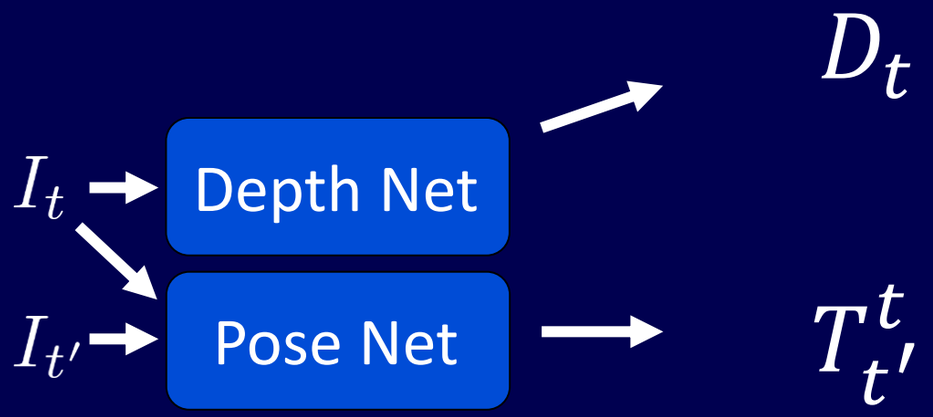
Yang, Wang, Stueckler, Cremers, ECCV 2018

D3VO: Depth, Pose & Uncertainty



Yang et al., "D3VO", CVPR '20

Deep Depth and Pose



I_t

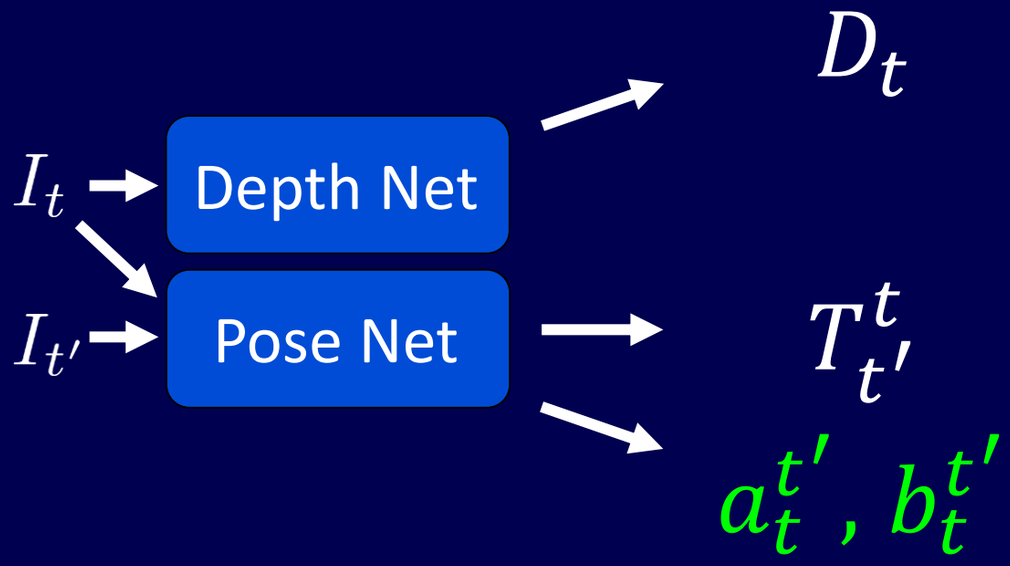
$I_{t'\to t}$

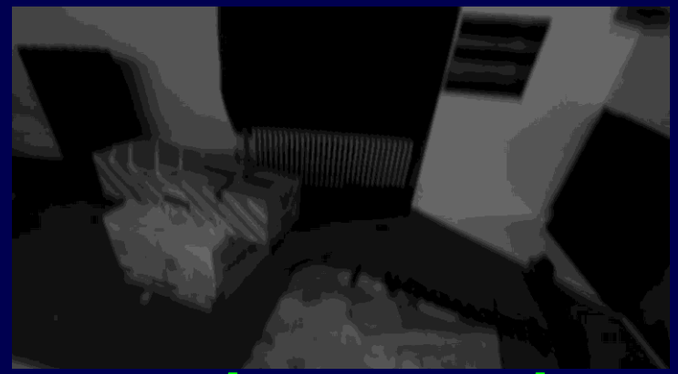
Self-supervised learning:

$$L_{self} = r(I_t, I_{t'\to t})$$

Yang et al., "D3VO", CVPR '20

Deep Affine Brightness Correction





$a_t^{t'} I_t + b_t^{t'}$

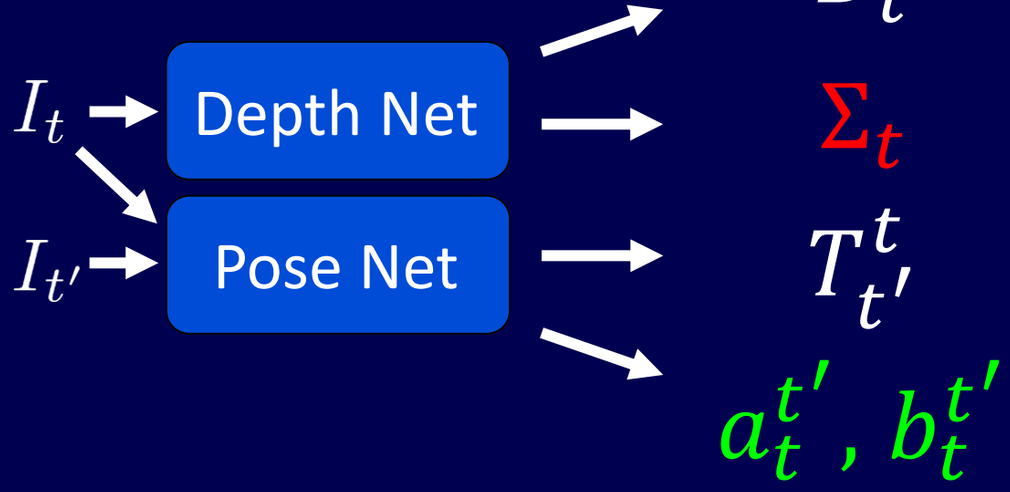
$I_{t' \to t}$

Self-supervised learning:

$$L_{self} = r(a_t^{t'} I_t + b_t^{t'}, I_{t' \to t})$$

Yang et al., "D3VO", CVPR '20

Deep Uncertainty



I_t

Σ_t

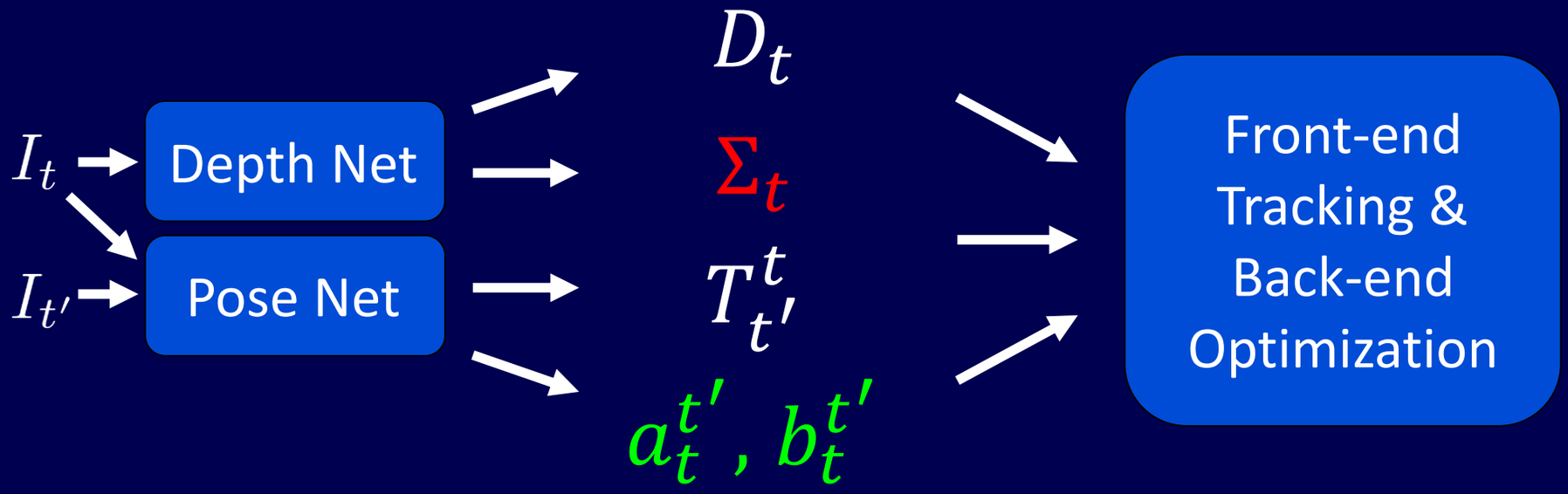
Self-supervised learning:

$$L_{self} = \frac{r(a_t^{t'} I_t + b_t^{t'}, I_{t' \to t})}{\Sigma_t} + \log \Sigma_t$$

Kendall et al., NeurIPS '17, Klodt et al. ECCV '18

Yang et al., "D3VO", CVPR '20

Predicting Depth, Pose & Uncertainty

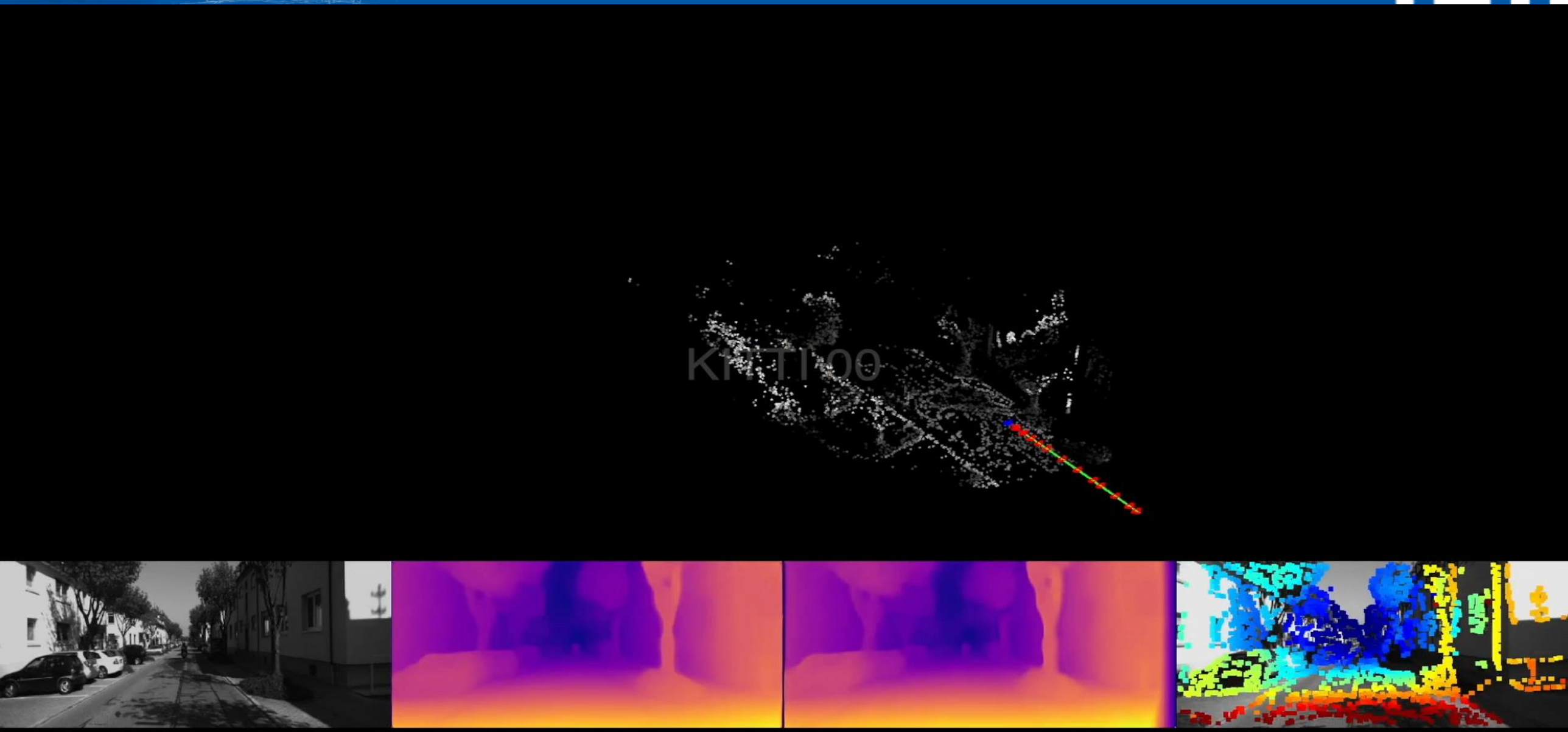


Self-supervised learning:

$$L_{self} = \frac{r(a_t^{t'} I_t + b_t^{t'}, I_{t' \to t})}{\Sigma_t} + \log \Sigma_t$$

Yang et al., "D3VO", CVPR '20

Deep Nets for Visual SLAM



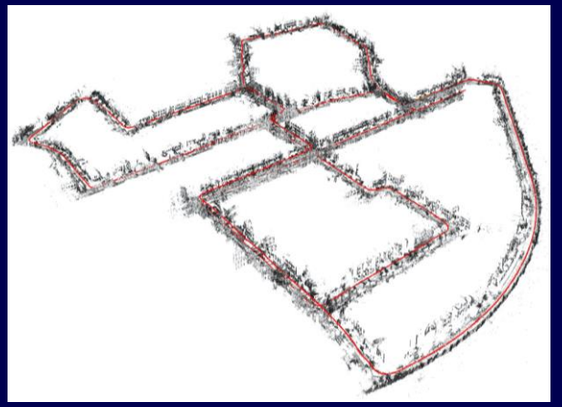
Experiments: Visual Odometry

KITTI: classical methods

		mean
Mono	ORB-SLAM	37.0
Stereo	Stereo DSO	0.89
Mono	D3VO	**0.82**

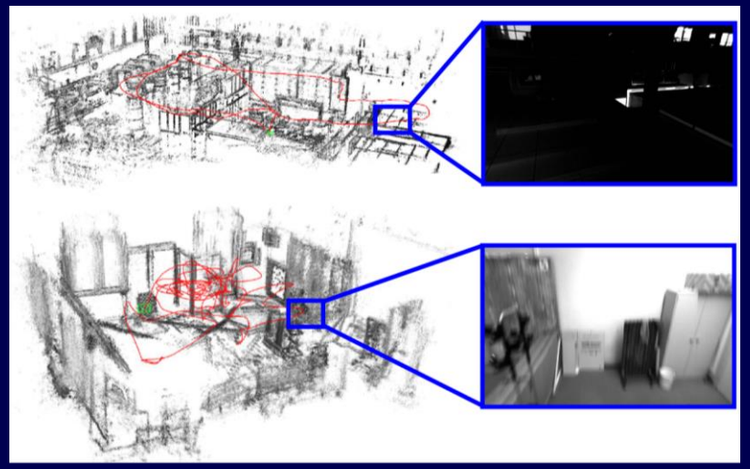
KITTI: deep methods

		Seq. 09	Seq. 10
End-to-end	Gordon et al.	2.7	6.8
Hybrid	Zhan et al.	2.61	2.29
Hybrid	D3VO	**0.78**	**0.62**



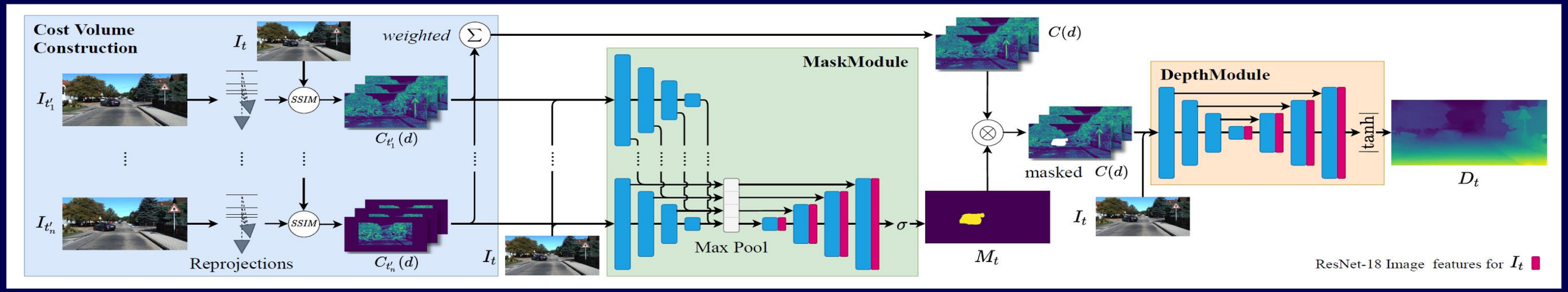
EuRoC: classical methods

		mean
Mono	DSO	0.48
Mono-VIO	VI-DSO	0.11
Stereo-VIO	Basalt	**0.08**
Mono	D3VO	**0.08**



Yang et al., "D3VO", CVPR '20

Dense Reconstructions from a Single Camera



cost volume

mask module

depth module

Aggregate brightness from multiple warped frames.

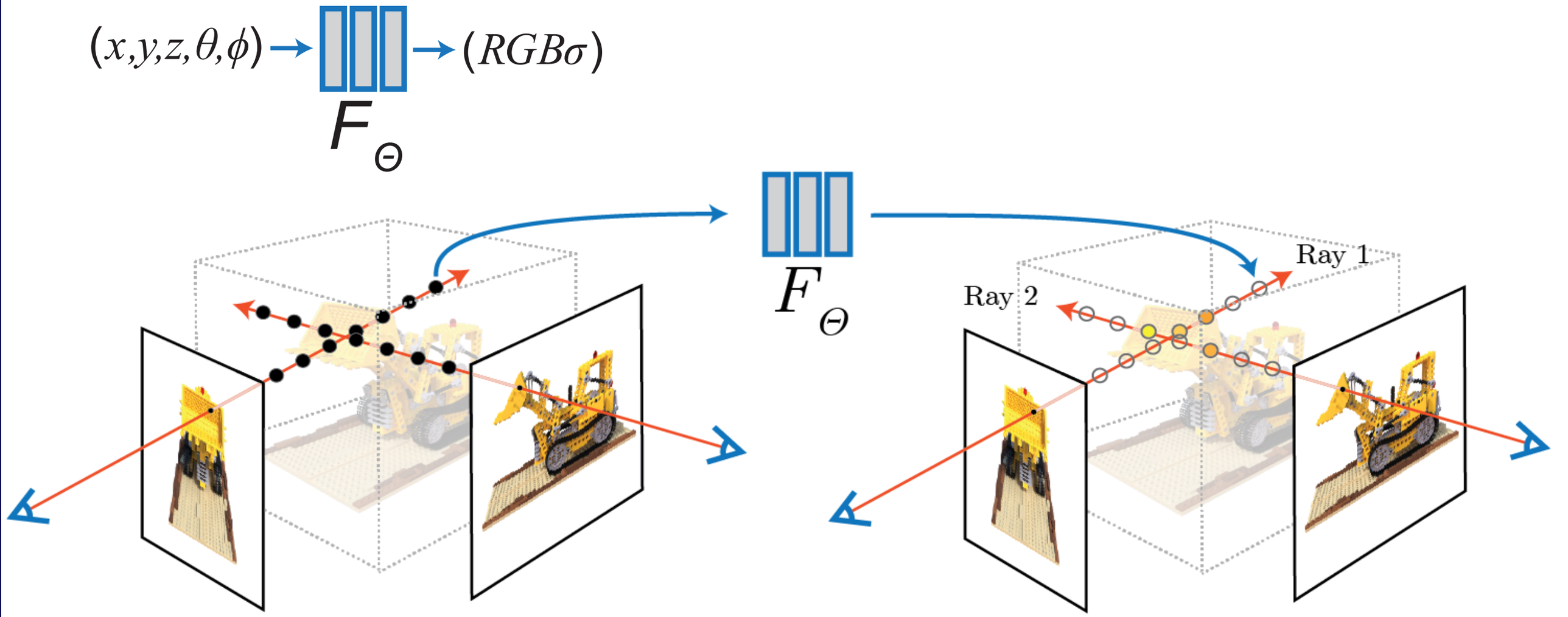
Filter out moving objects based on lack of brightness consistency.

Wimbauer et al., "MonoRec: Monocular Dense Reconstruction", CVPR '21

Dense Reconstructions from a Single Camera

Wimbauer et al., “MonoRec: Monocular Dense Reconstruction”, CVPR ‘21

Neural Radiance Fields for Novel View Synthesis

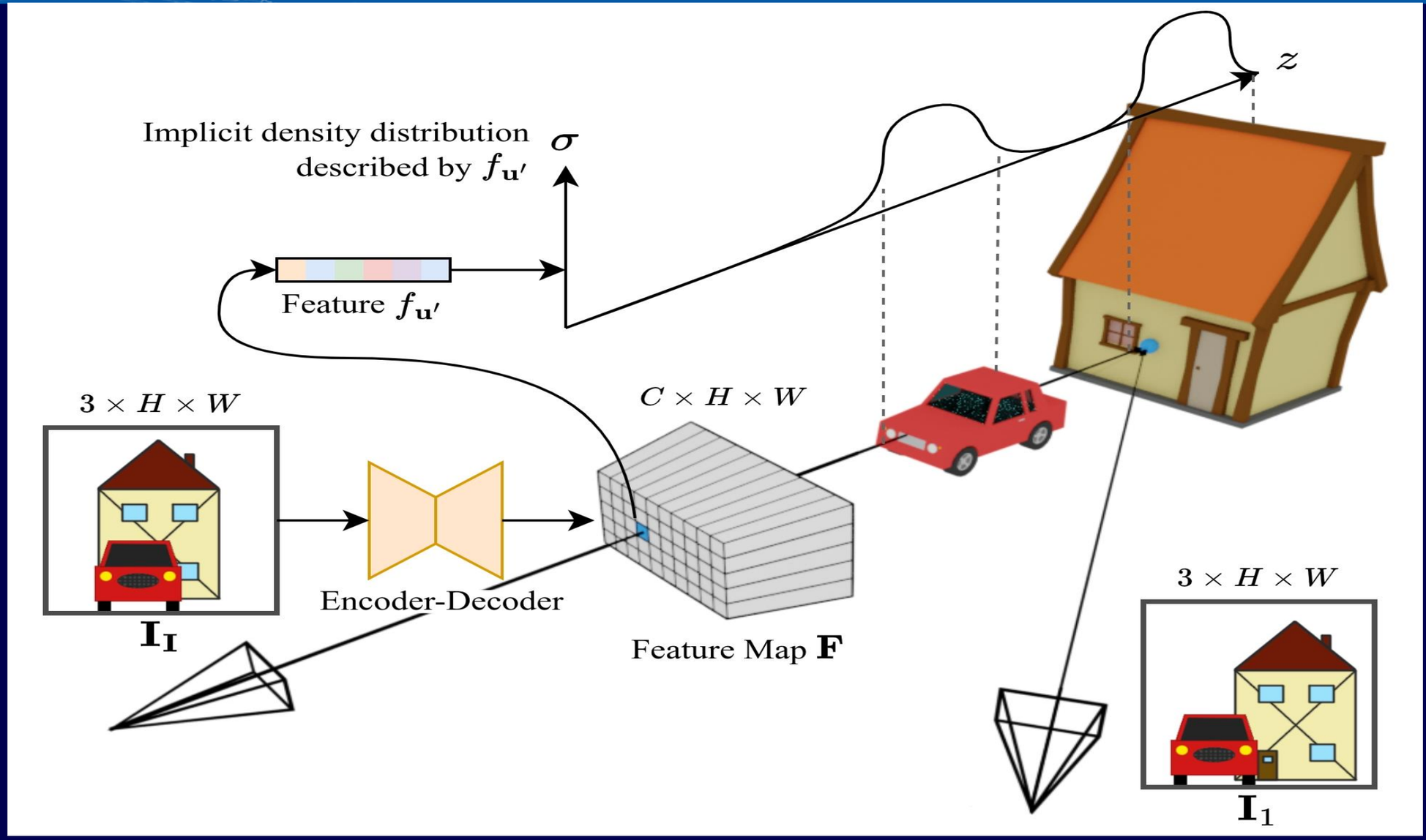


Mildenhall et al., "Neural Radiance Fields", ECCV 2020

Neural Radiance Fields for Novel View Synthesis

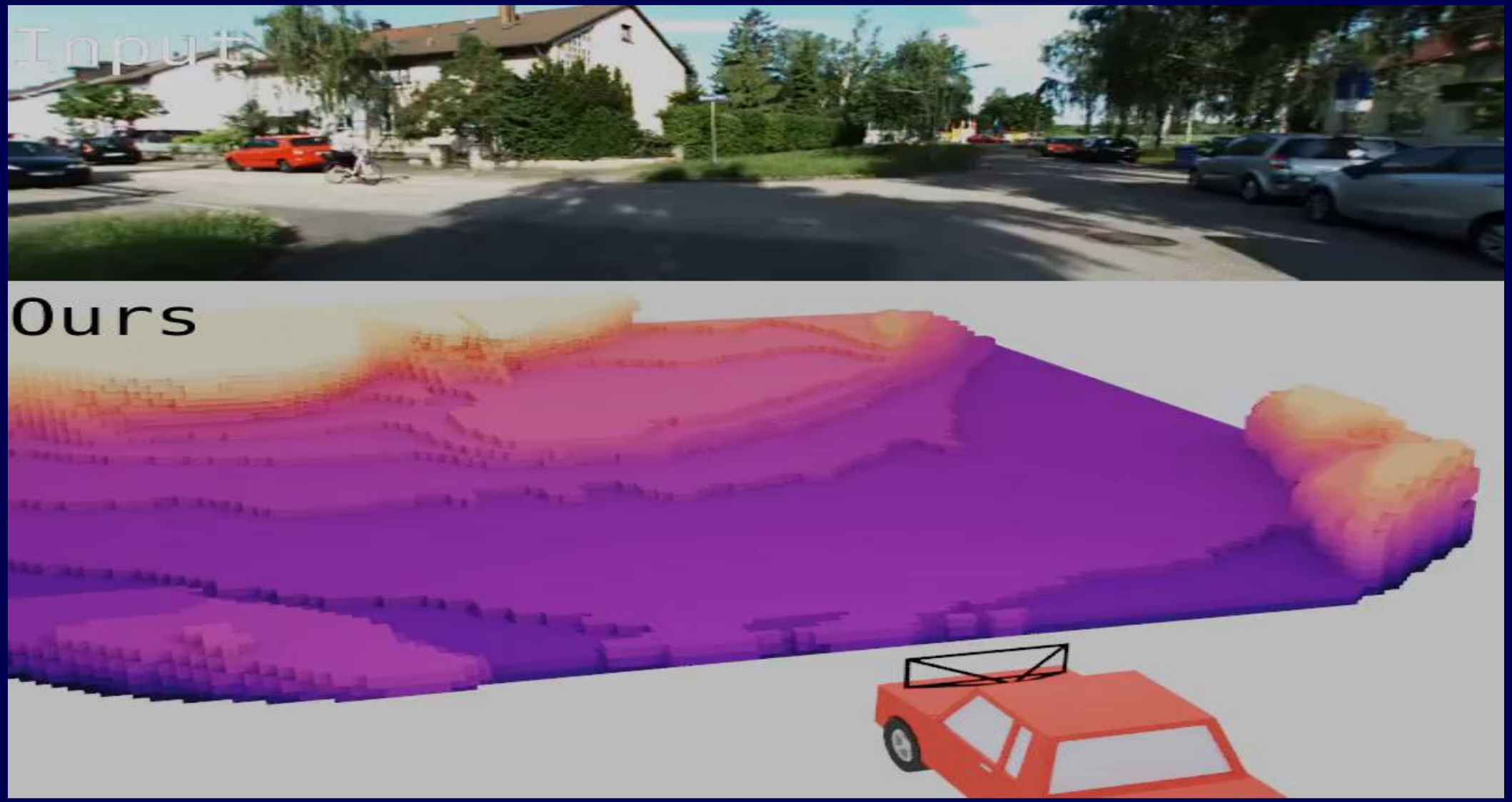
Mildenhall et al., “Neural Radiance Fields”, ECCV 2020

Density-Fields for Single View Reconstruction



Wimbauer et al., "Behind the Scenes: Density Fields for Single View Reconstruction", CVPR '23

Density-Fields for Single View Reconstruction



Wimbauer et al., “Behind the Scenes: Density Fields for Single View Reconstruction”, CVPR ‘23

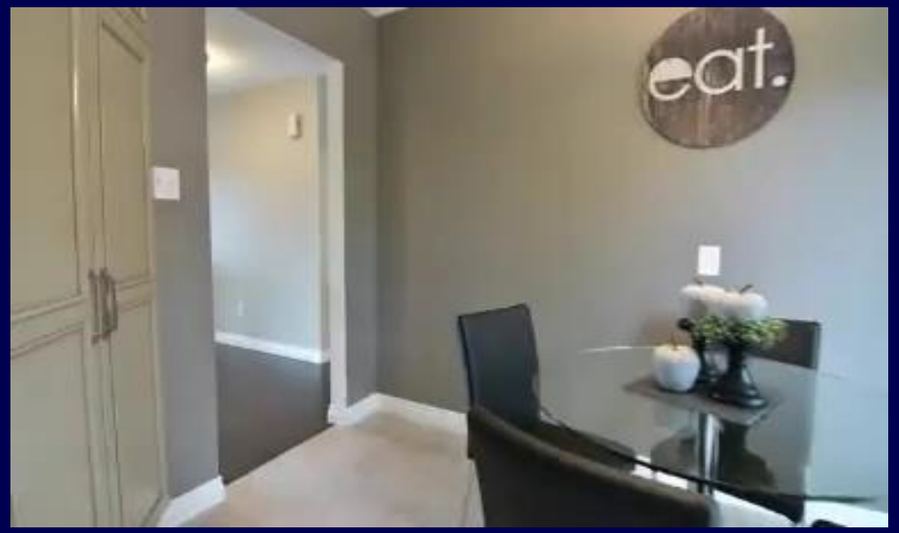
Novel View Synthesis Results





Wimbauer et al., "Behind the Scenes: Density Fields for Single View Reconstruction", CVPR '23

Novel View Synthesis Results

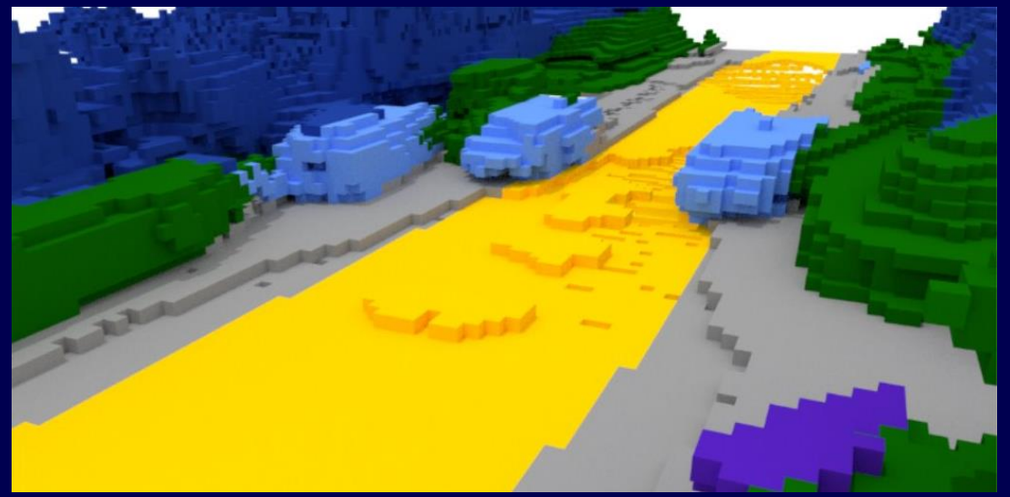


Wimbauer et al., “Behind the Scenes: Density Fields for Single View Reconstruction”, CVPR ‘23

Self-supervised Semantic Scene Completion

Color image

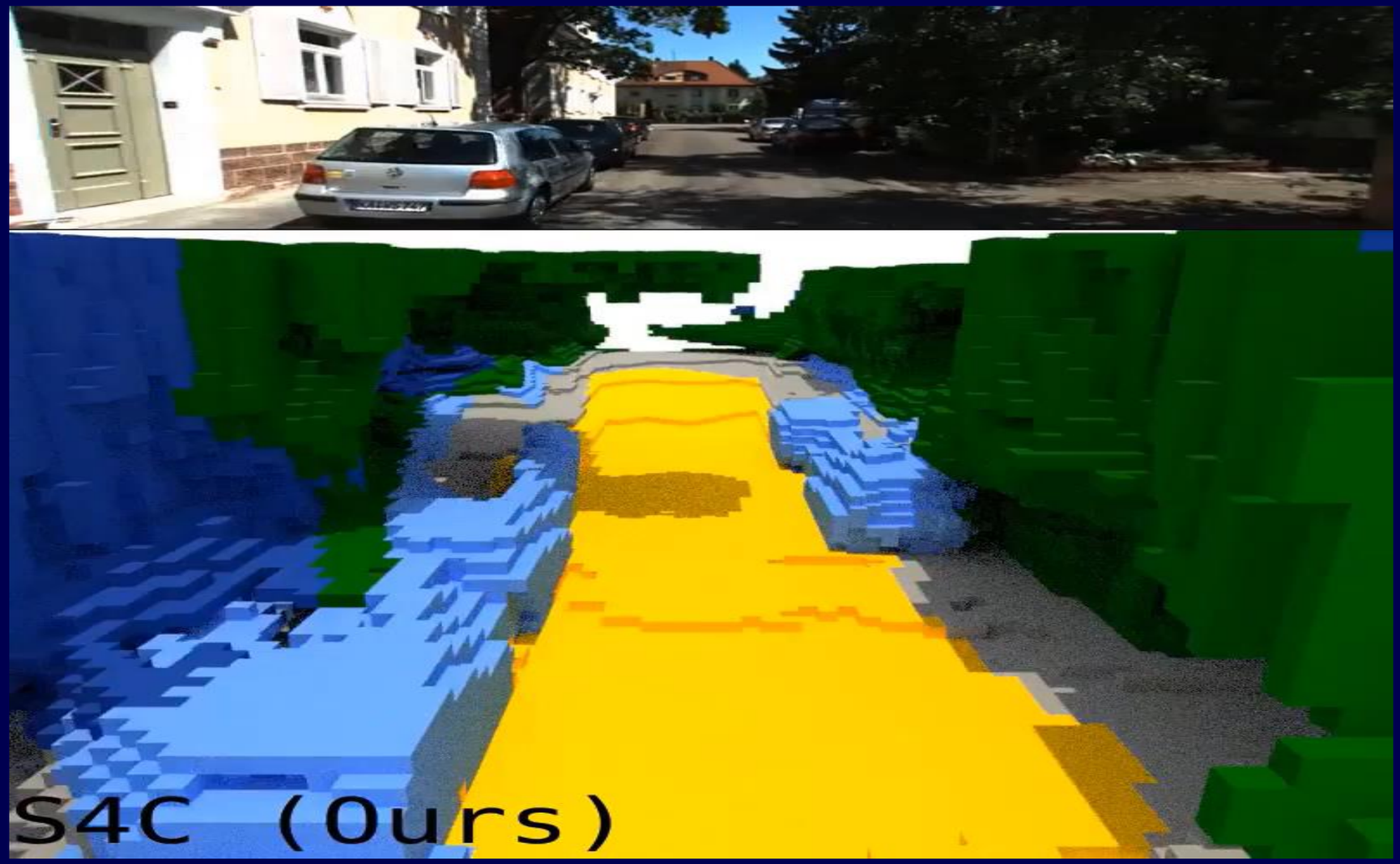
(Depth / Lidar)



Complete **Geometry** and **Semantics**

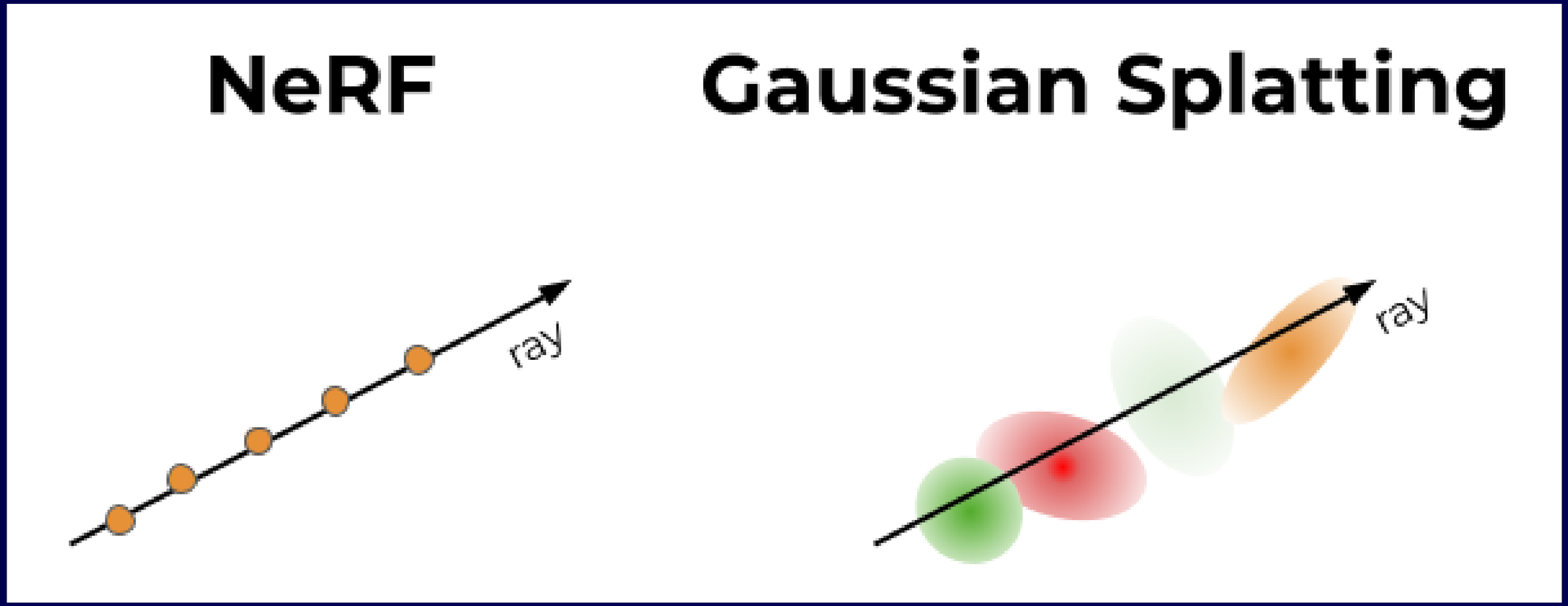
Hayler et al., “S4C: Self-Supervised Semantic Scene Completion”, 3DV ‘24

Self-supervised Semantic Scene Completion



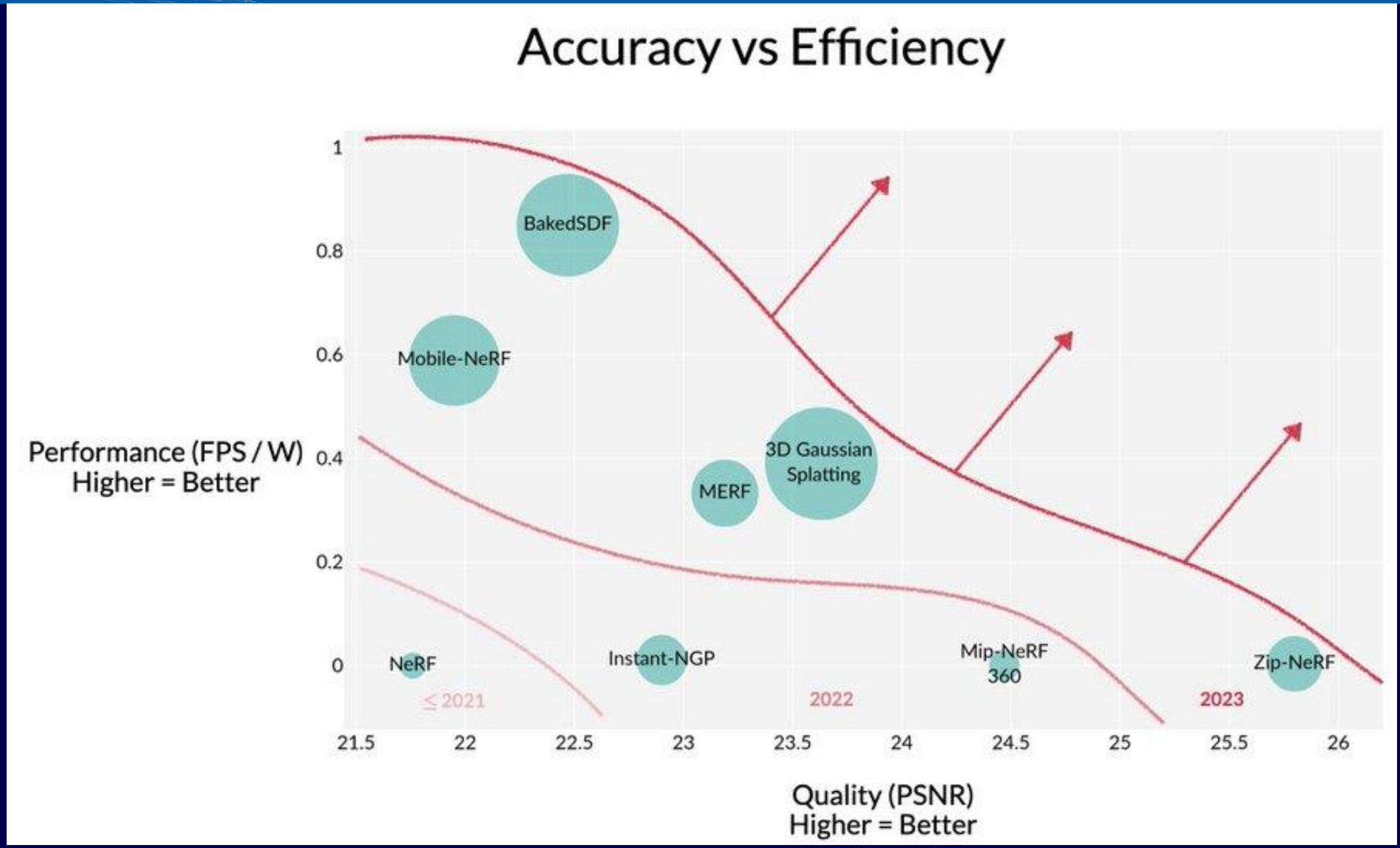
Hayler et al., “S4C: Self-Supervised Semantic Scene Completion”, 3DV ‘24

NeRFs vs. Gaussian Splatting



Lee Alan Westover 1991, Kerbl et al., "3D Gaussian Splatting", Siggraph 2023

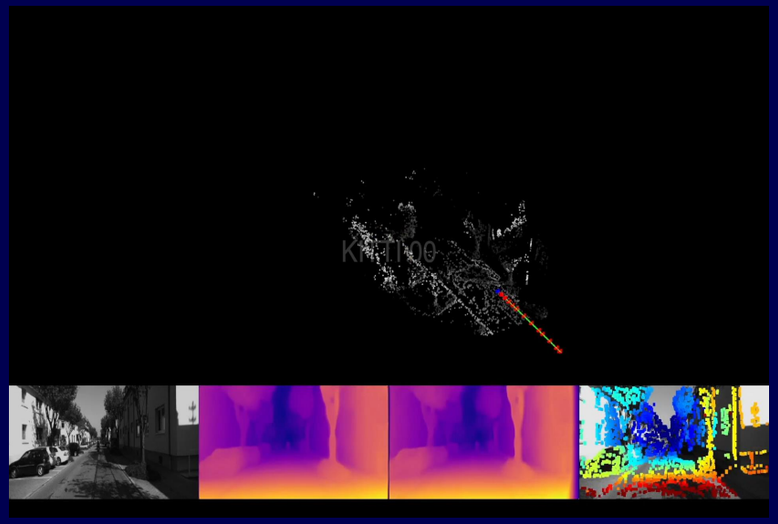
Gaussian Splatting vs. NeRFs



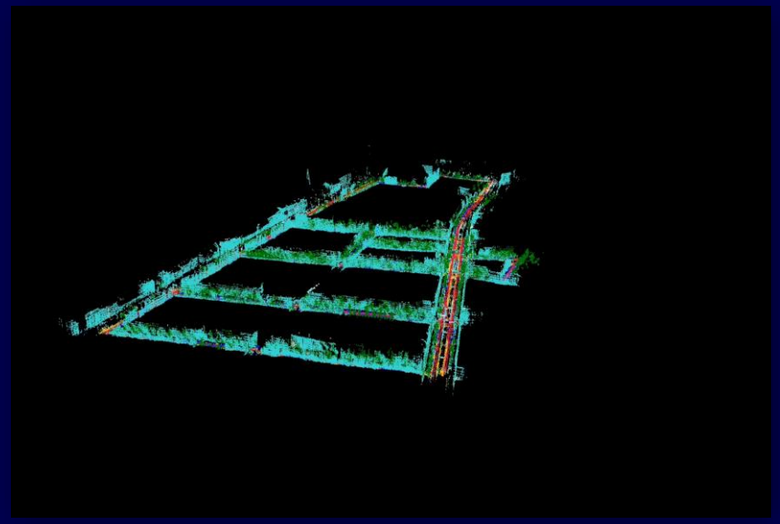
Source: Jon Barron

Summary

DM-VIO for visual-inertial odometry

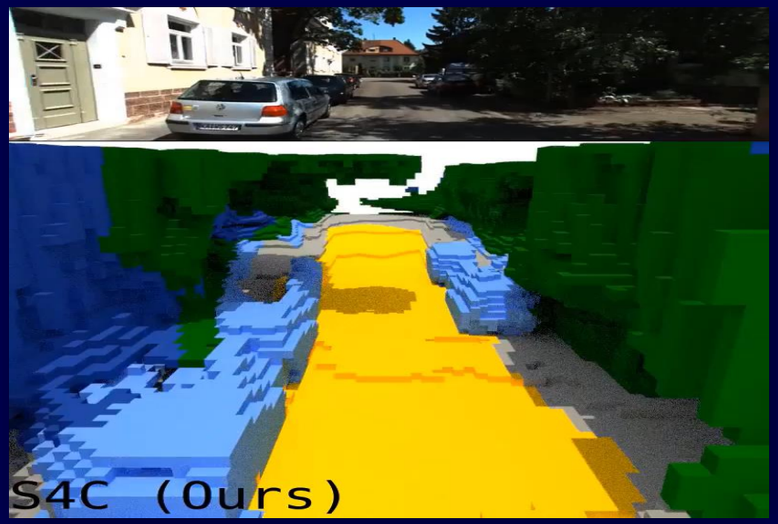


Deep & direct visual SLAM

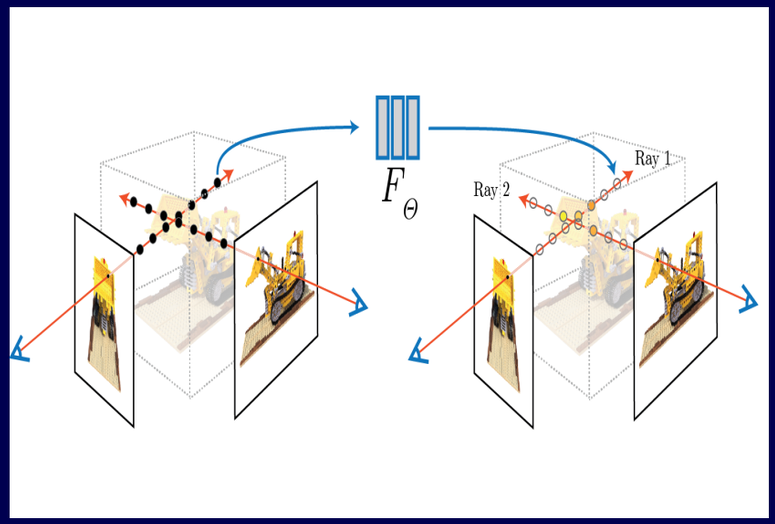


3D semantic mapping

Monoocular dense mapping



Semantic scene completion



Neural radiance fields & splatting