KDBTS:

Boosting Self-Supervision for Single View Scene completion via **K**nowledge **D**istillation

keonhee-han.github.io/publications/kdbts/


Keonhee Han[*,1]


Dominik Muhle[*,1,2]


Felix Wimbauer[,1,2]


Daniel Cremers[1,2]

[1]Technical University of Munich

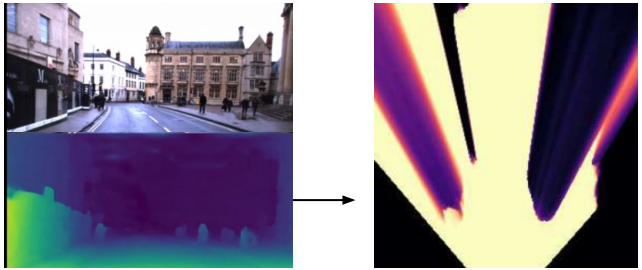
[2]MCML

TECHNISCHE UNIVERSITÄT MÜNCHEN

Munich Center for Machine Learning

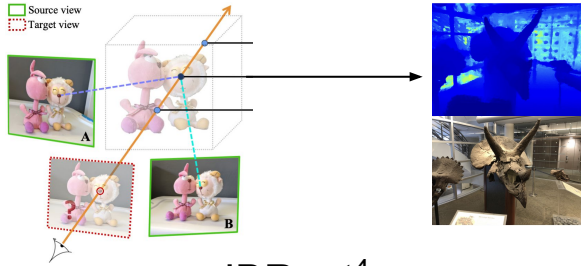
KDBTS

- Density Field: **Image conditioned** prediction via feature map. e.g. PixelNeRF[2]



Monodepth2[1]

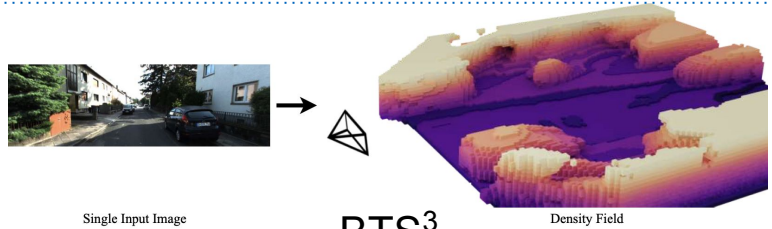
❌ Limited to **visible** areas



IBRnet[4]

✅ Visibility **Aggregation** - multi views

❌ Requires additional **posed inputs**



BTS[3]

✅ Reasons about the **whole scene**

❌ Limited accuracy due to **single view** input

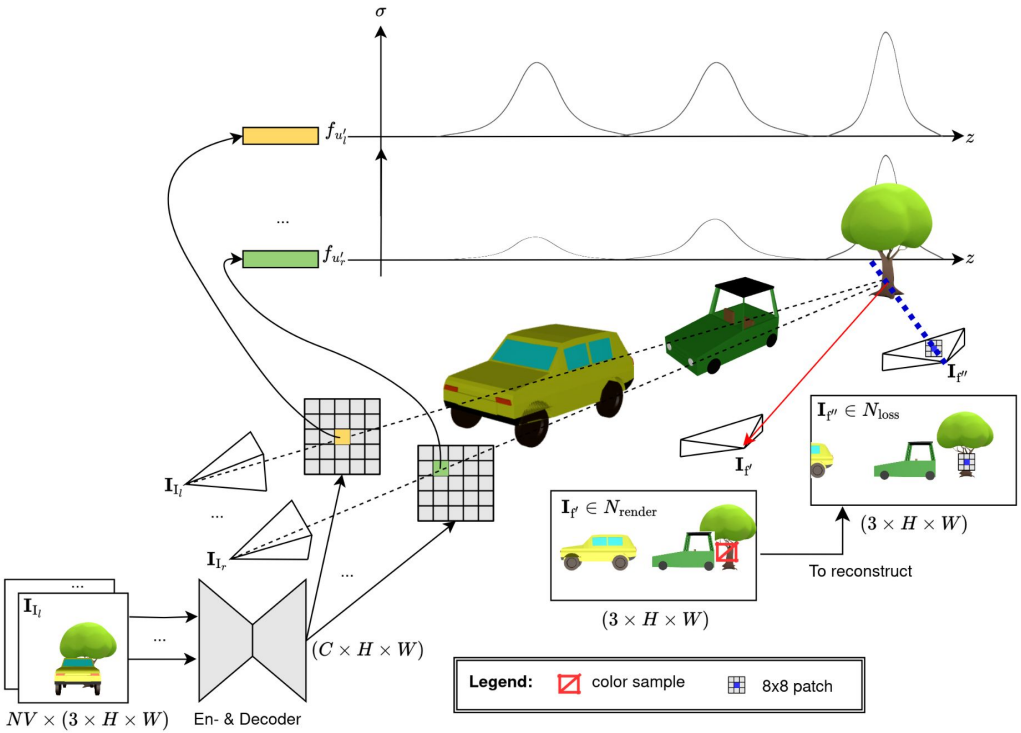
MVBTS **KDBTS**

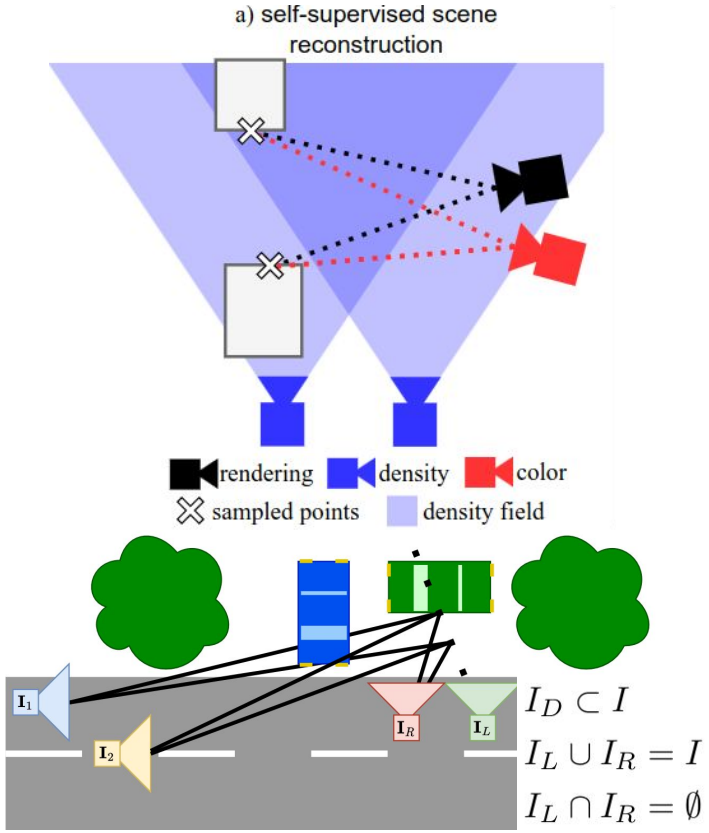
✅ 1. Exploiting visibility of **Multi-Views** via confidence-based view aggregation.

✅ 2. Improving the single view scene completion by **K**nowledge **D**istillation.

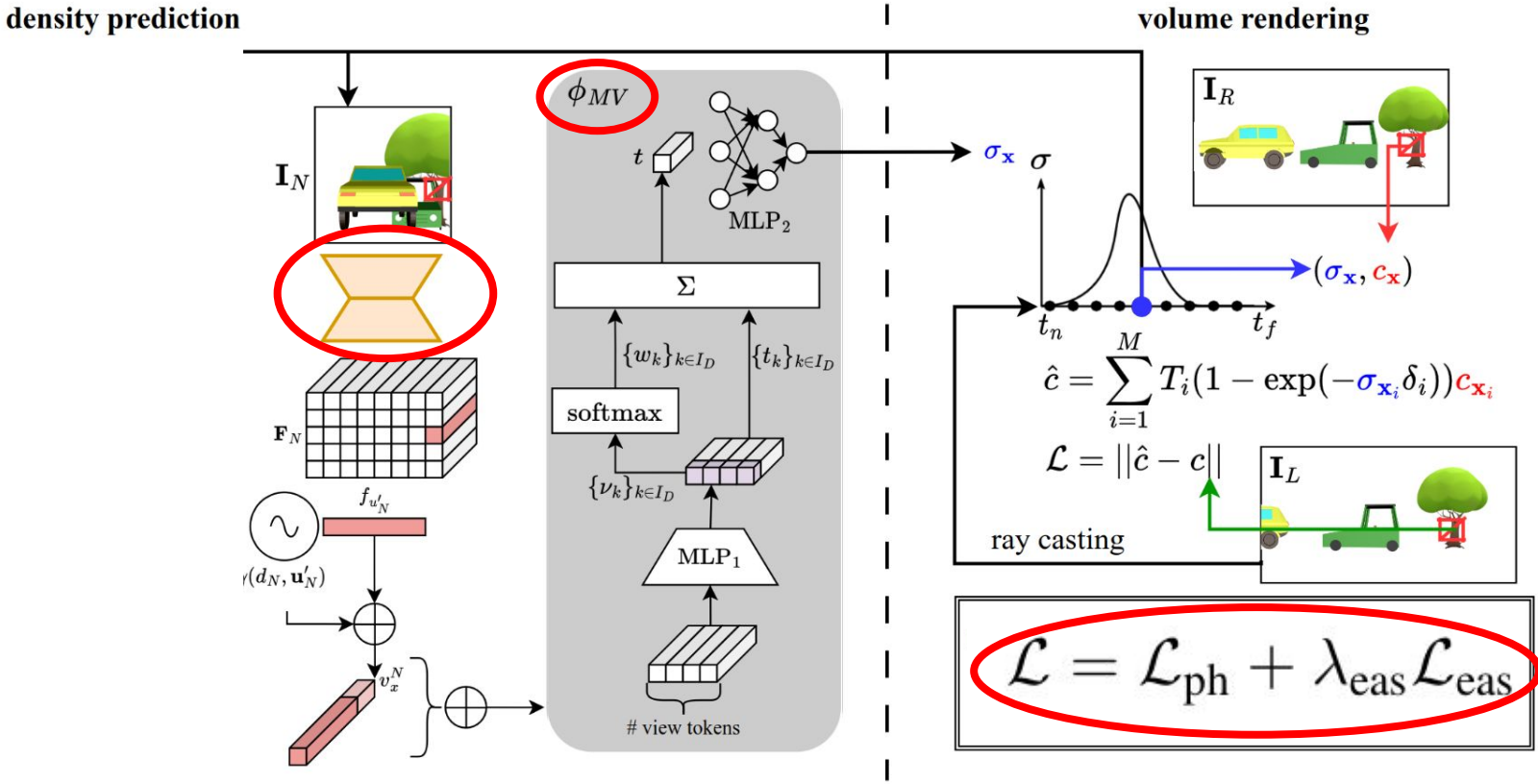
[1] Godard, Clément, et al. Digging into self-supervised monocular depth estimation., ICCV 2019
[2] Yu, Alex, et al. pixelnerf: Neural radiance fields from one or few images, CVPR 2021
[3] Wimbauer, Felix, et al. Behind the scenes: Density fields for single view reconstruction, CVPR 2023
[4] Wang, Qianqian, et al. Ibrnet: Learning multi-view image-based rendering, CVPR 2021

Self-Supervised Training in Multi-Views



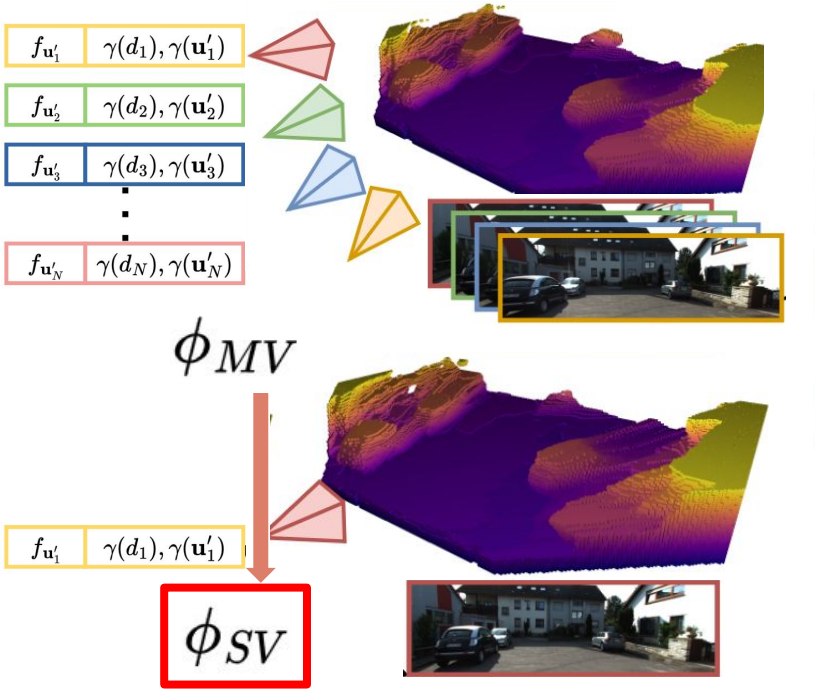


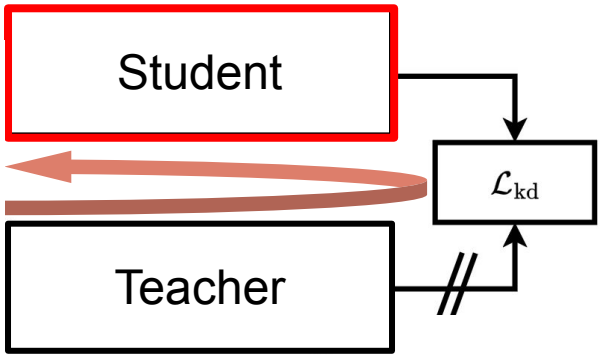
Model Architecture - MVBTS



Model Architecture - KDBTS

- Direct supervision by freezing all, except Single-View (Student) network
- Faster inference by smaller network, No posed inputs required





$$\mathcal{L}_{\text{kd}} = \|\sigma_{\mathbf{x},\text{MV}} - \sigma_{\mathbf{x},\text{SV}}\|$$

Datasets & Training setup

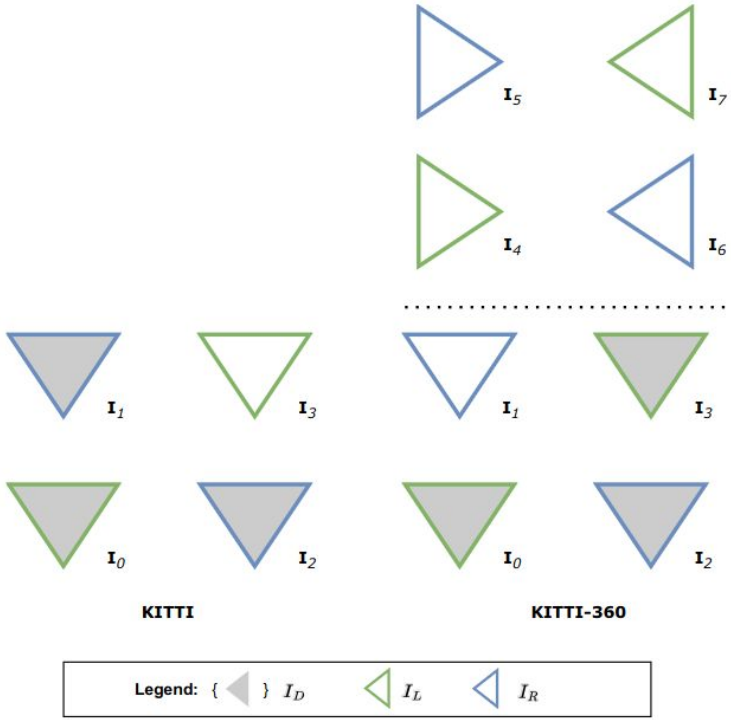
- Training data benchmark for MVBTS, and its frame setup.

KITTI-360

KITTI

Dataset	Split.	# Train	# Val	# Test
KITTI	Eigen[1] [9]	39810	4424	697
KITTI-360	BTS[2] [49]	98008	11451	446

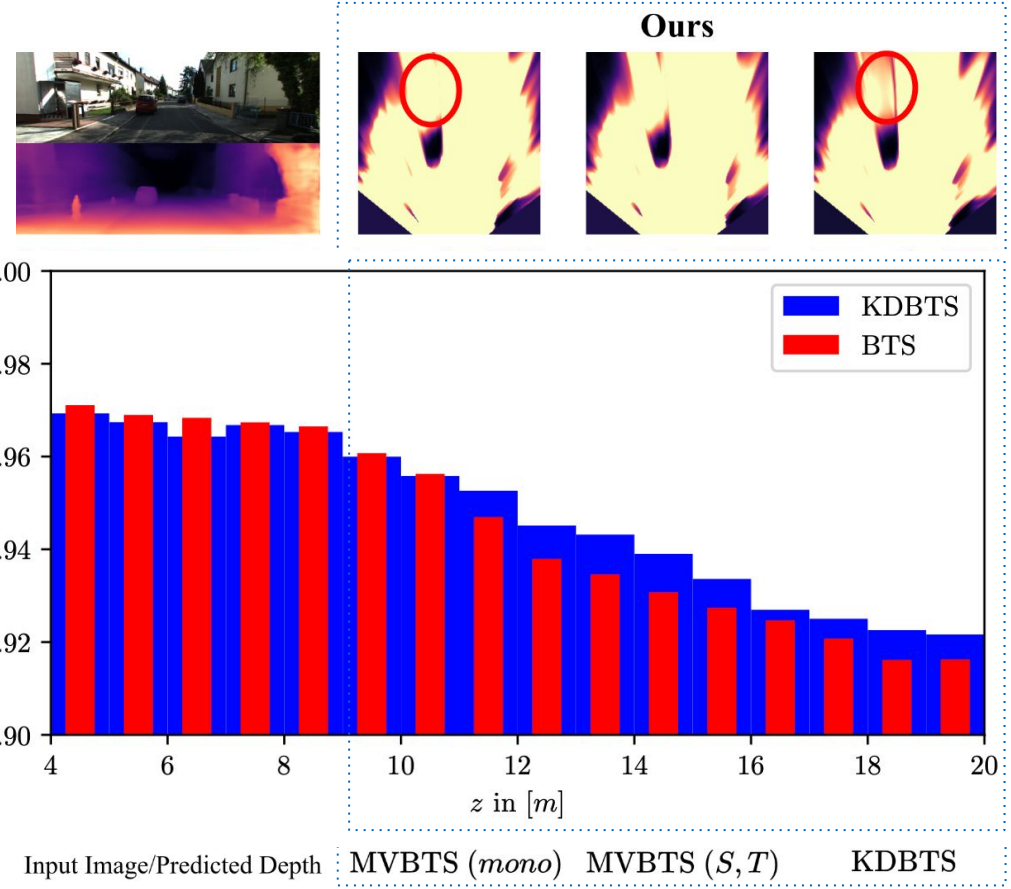
Frame Arrangement

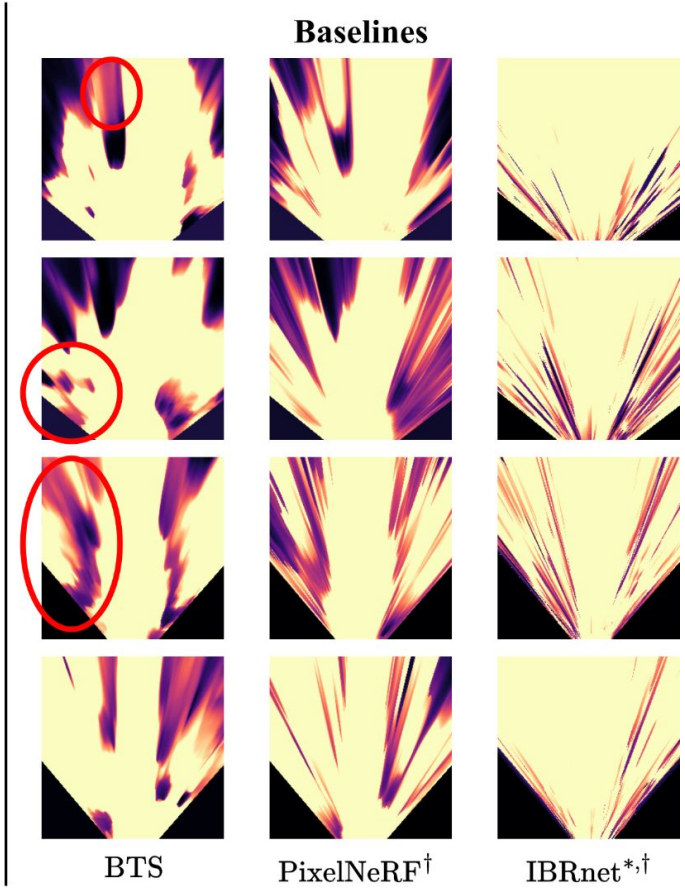


[1] Eigen et al. Depthmap prediction from a single image using a multi-scale deep network.
[2] Wimbauer, Felix, et al. Behind the scenes: Density fields for single view reconstruction, CVPR 2023

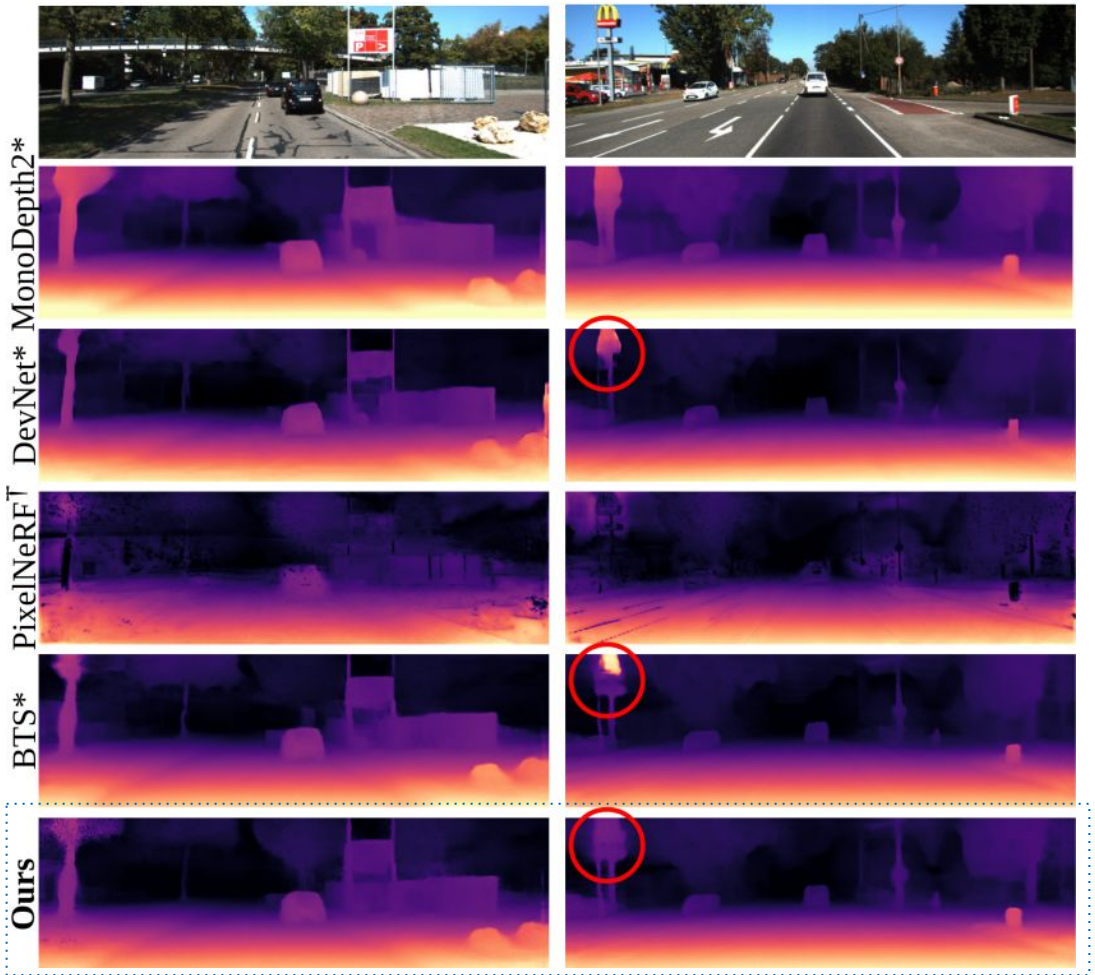
Occupancy Estimation - KITTI 360

- KDBTS handles occlusions as good as MVBTS, and outperform baselines.
- More **accurate** occupancy estimation in further distance.





Depth Estimation - KITTI



Quantitative Evaluation

- KDBTS handles occlusions as much as MVBTS, and outperform BTS baseline.

Method	Multi View	O_{acc} ↑	O_{prec} ↑	O_{rec} ↑	IE_{acc} ↑	IE_{prec}	IE_{rec} ↑
PixelNeRF[2] [53]	✗	93.82%	51.94%	69.43%	61.33%	37.86%	42.21%
BTS[1] [49]	✗	94.47%	58.73%	84.24%	77.04%	**54.21%**	43.99%
MVBTS (**Ours**)	✗	**94.76%**	**60.83%**	84.51%	78.00%	53.69%	**44.04%**
KDBTS (**Ours**)	✗	**94.76%**	60.68%	**84.78%**	**78.30%**	53.62%	44.00%
IBRnet[3] [46]	✓	96.03%	4.14%	4.78%	34.36%	32.97%	**96.02%**
IBRnet[3] (depth + 4m) [46]	✓	**98.12%**	43.35%	**86.01%**	59.67%	25.18%	9.85%
MVBTS (**Ours**)	✓	94.91%	**61.73%**	85.78%	**79.47%**	**55.08%**	45.23%

[1] Yu, Alex, et al. pixelnerf: Neural radiance fields from one or few images, CVPR 2021
[2] Wimbauer, Felix, et al. Behind the scenes: Density fields for single view reconstruction, CVPR 2023
[3] Wang, Qianqian, et al. Ibrnet: Learning multi-view image-based rendering, CVPR 2021


ours (KDBTS)

CVPR
JUNE 17-21, 2024
SEATTLE, WA

Munich Center for Machine Learning

TECHNISCHE UNIVERSITÄT MÜNCHEN

KDBTS:

Boosting Self-Supervision for Single View Scene completion via **K**nowledge **D**istillation

- ✅ **MVBTS** extends existing density field prediction to a confidence based **M**ulti-**V**iew setting
- ✅ **K**nowledge **D**istillation improves the single view scene completion, **KDBTS**.
- ✅ State-of-the-art occupancy prediction on KITTI-360 due to better occlusion reasoning

For **code** and **pretrained models**, please visit our **project page**:

keonhee-han.github.io/publications/kdbts/