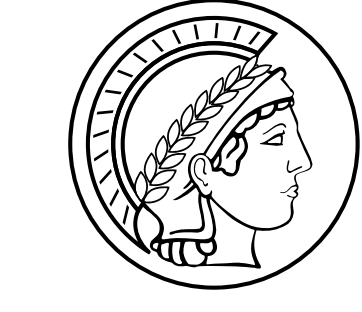


OctNetFusion: Learning Depth Fusion from Data

Gernot Riegler¹, Ali Osman Ulusoy^{2,3}, Horst Bischof¹, Andreas Geiger^{2,4}

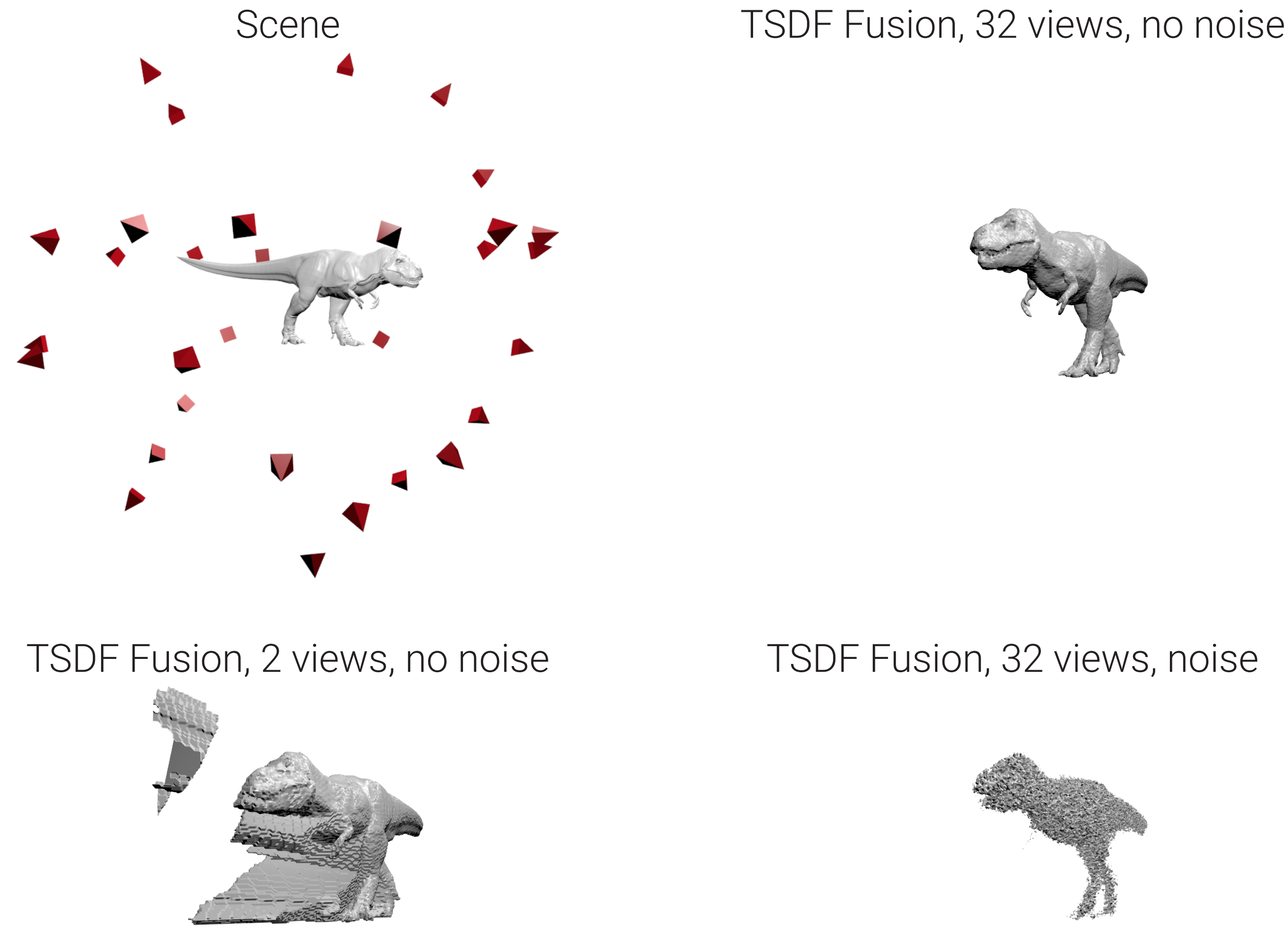
¹Graz University of Technology ²MPI for Intelligent Systems Tübingen ³Microsoft ⁴ETH Zürich



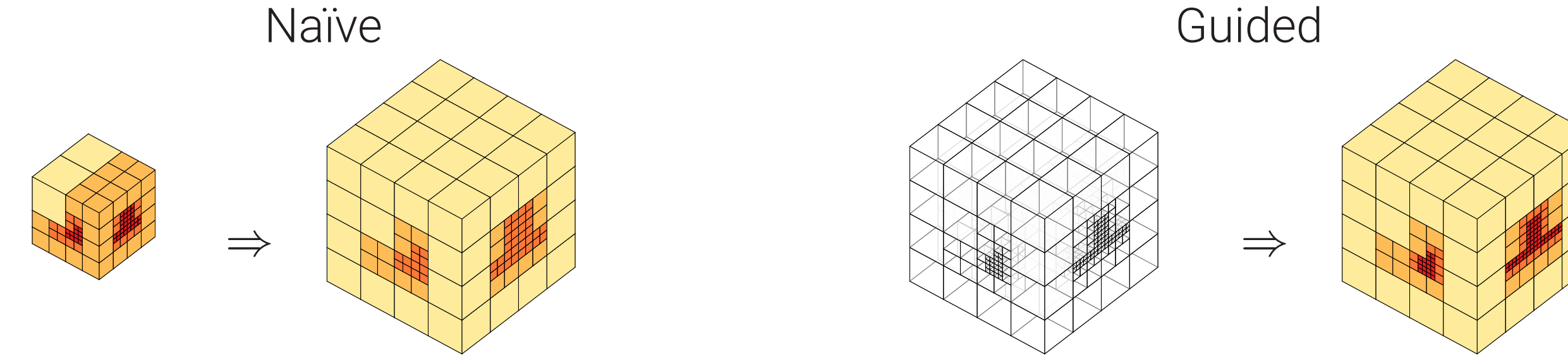
Max Planck Institute
for Intelligent Systems
Autonomous Vision Group



Motivation

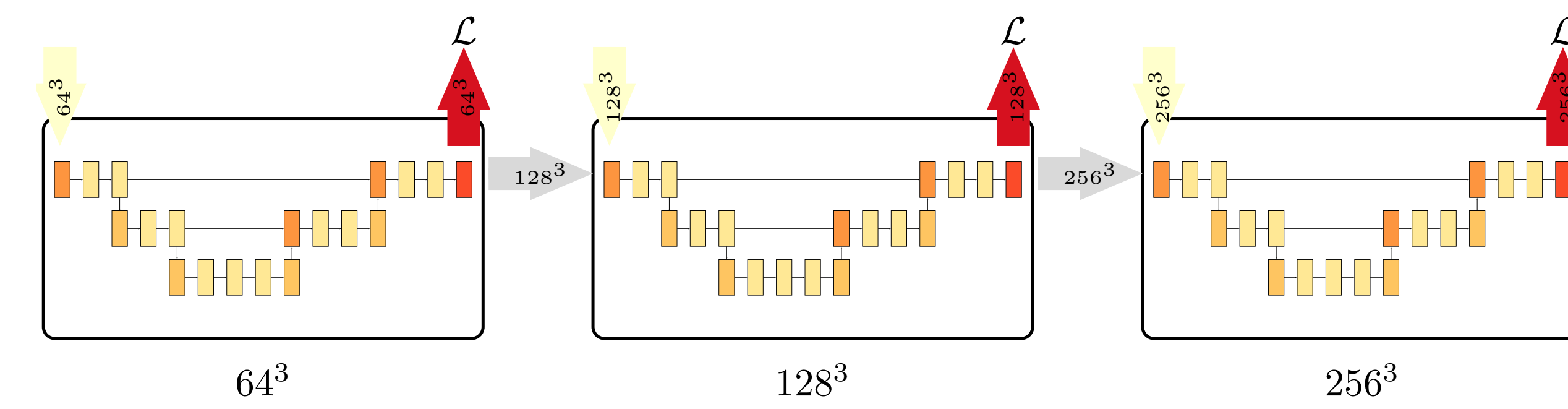


OctNet Unpooling

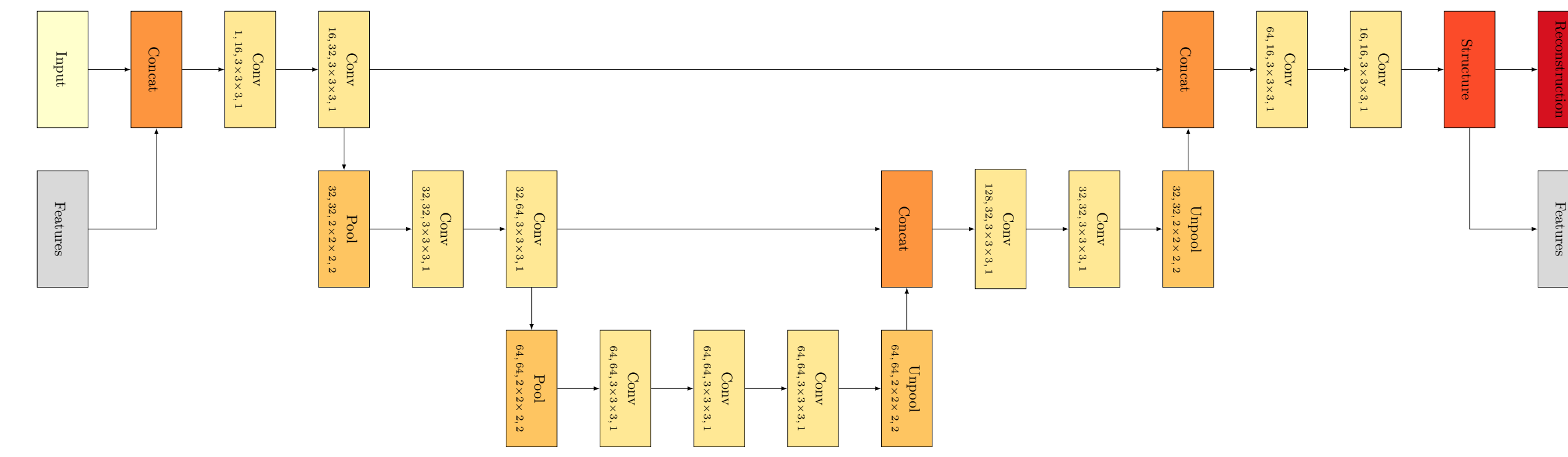


OctNetFusion

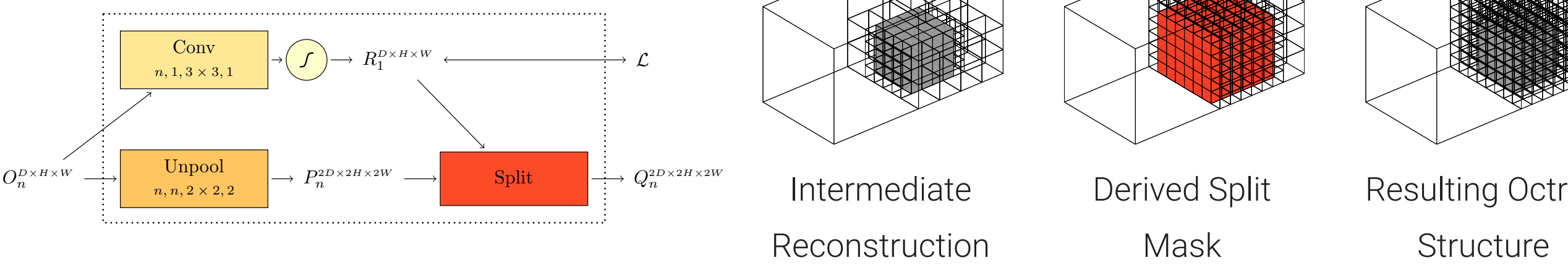
Architecture



Encoder-Decoder Module



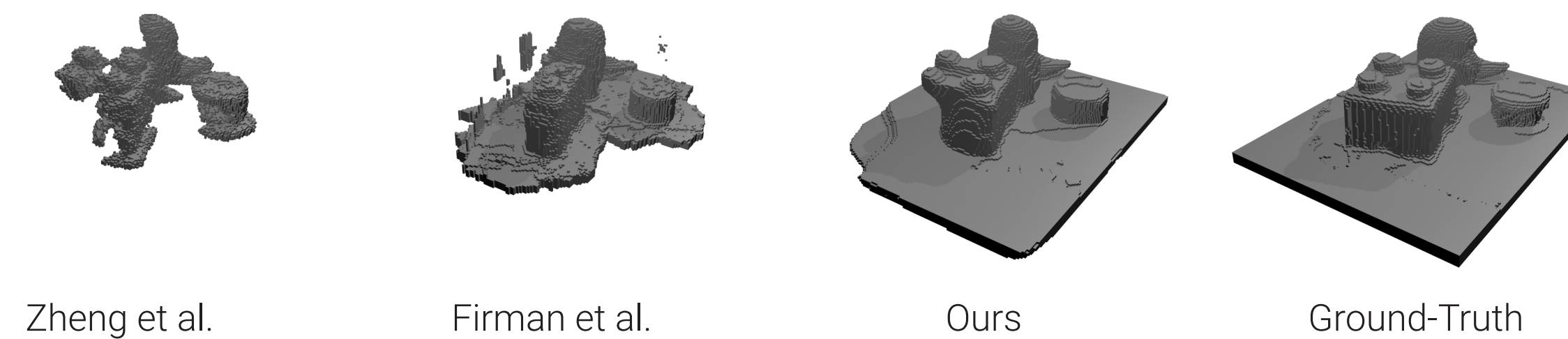
Structure Module



Volumetric Shape Completion

Voxlets Dataset [3]

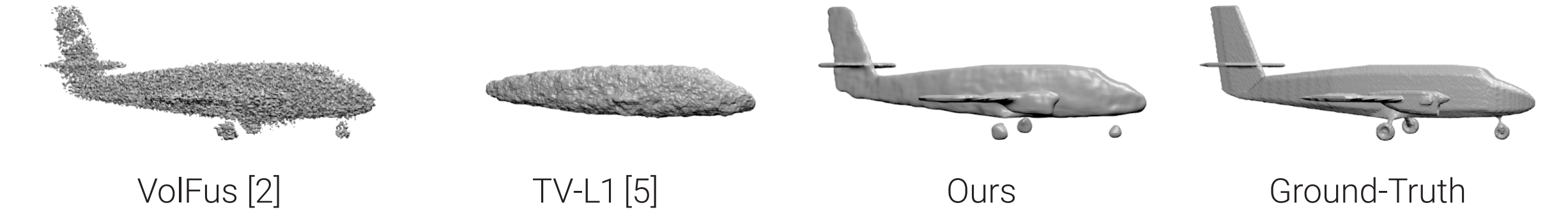
Method	IoU	Precision	Recall
Zheng et al.*	0.528	0.773	0.630
Firman et al.*	0.585	0.793	0.658
Firman et al.	0.550	0.734	0.705
Ours	0.650	0.834	0.756



Volumetric Depth Fusion

ModelNet: Input Encoding

	MAD (mm)					
	VolFus	TV-L1	Occ	TDF + Occ	TSDF	TSDF Hist
64 ³	4.136	3.899	2.095	1.987	1.710	1.715
128 ³	2.058	1.690	0.955	0.961	0.838	0.736
256 ³	1.020	0.778	0.410	0.408	0.383	0.337

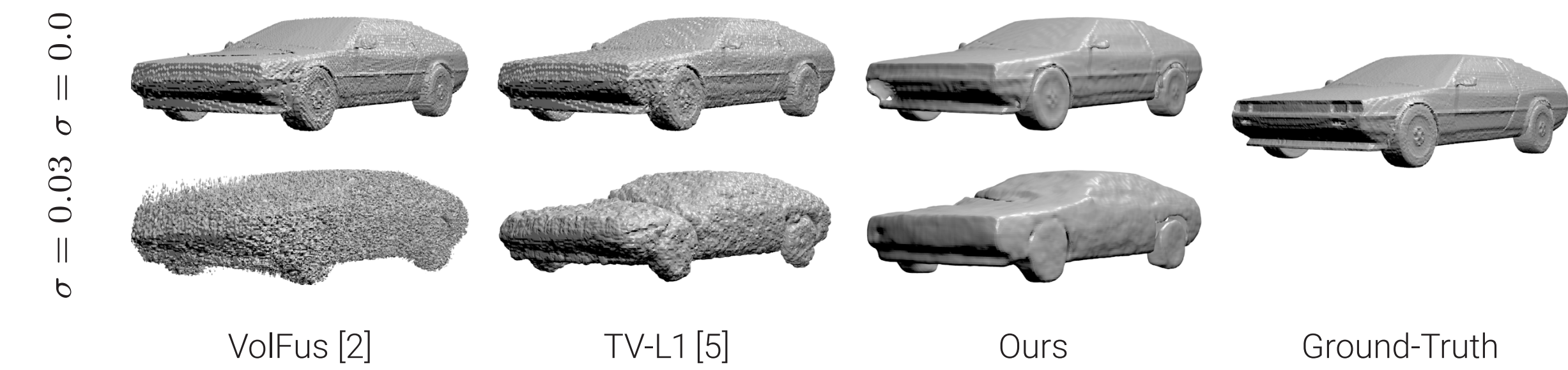


ModelNet: Number of Input Views

	MAD (mm)											
	views=1			views=2			views=4			views=6		
	VolFus	TV-L1	Ours	VolFus	TV-L1	Ours	VolFus	TV-L1	Ours	VolFus	TV-L1	Ours
64 ³	59.295	48.345	7.855	15.626	13.267	2.755	4.136	3.899	1.715	3.171	2.905	1.484
128 ³	29.795	26.525	3.853	7.850	6.999	1.333	2.058	1.690	0.736	1.648	1.445	0.661
256 ³	14.919	14.529	1.927	3.929	3.537	0.616	1.020	0.778	0.337	0.842	0.644	0.360

ModelNet: Varying Input Noise

	MAD (mm)											
	$\sigma = 0.00$			$\sigma = 0.01$			$\sigma = 0.02$			$\sigma = 0.03$		
	VolFus	TV-L1	Ours	VolFus	TV-L1	Ours	VolFus	TV-L1	Ours	VolFus	TV-L1	Ours
64 ³	3.020	3.272	1.647	3.439	3.454	1.487	4.136	3.899	1.715	4.852	4.413	1.938
128 ³	1.330	1.396	0.744	1.647	1.543	0.676	2.058	1.690	0.736	2.420	1.850	0.804
256 ³	0.621	0.637	0.319	0.819	0.697	0.321	1.020	0.778	0.429	1.188	0.858	0.402



Kinect Object Scans [1]

	MAD (mm)					
	views=10			views=20		
	VolFus	TV-L1	Ours	VolFus	TV-L1	Ours
64 ³	103.855	25.976	22.540	72.631	22.081	18.422
128 ³	58.802	12.839	11.827	41.631	11.924	9.637
256 ³	31.707	5.372	4.806	22.555	5.195	4.110

References

- [1] S. Choi et al. "A Large Dataset of Object Scans". In: *arXiv.org* 1602.02481 (2016).
- [2] B. Curless and M. Levoy. "A Volumetric Method for Building Complex Models from Range Images". In: *SIG-GRAPH*. 1996.
- [3] M. Firman et al. "Structured Prediction of Unobserved Voxels From a Single Depth Image". In: *CVPR*. 2016.
- [4] G. Riegler et al. "OctNet: Learning Deep 3D Representations at High Resolutions". In: *CVPR*. 2017.
- [5] C. Zach et al. "A Globally Optimal Algorithm for Robust TV-L1 Range Image Integration." In: *ICCV*. 2007.
- [6] B. Zheng et al. "Beyond Point Clouds: Scene Understanding by Reasoning Geometry and Physics". In: *CVPR*. 2013.