



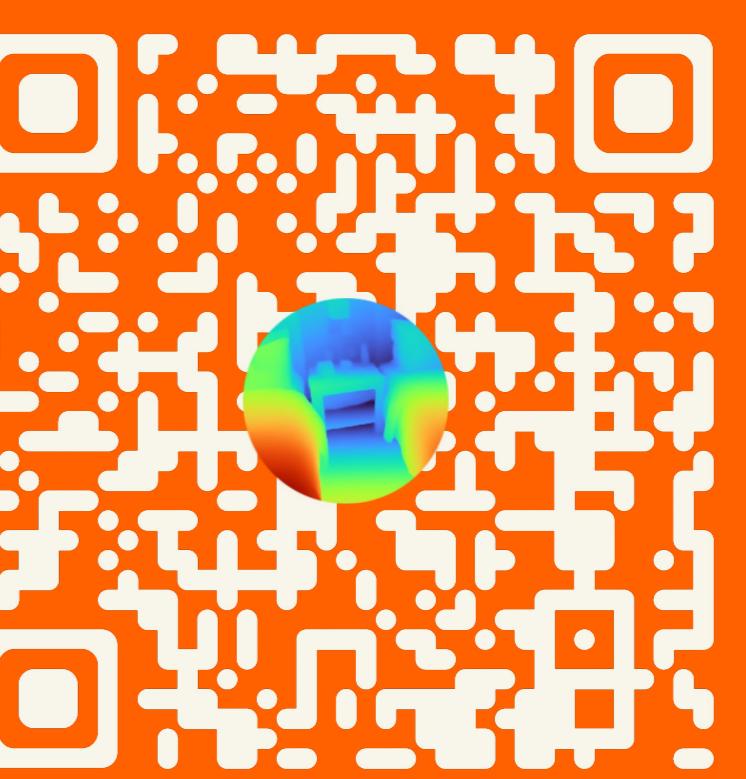
SimpleRecon

3D Reconstruction Without 3D Convolution

Mohamed Sayed^{2*} John Gibson¹ Jamie Watson¹ Victor Adrian Prisacariu^{1,3}

Michael Firman¹ Clément Godard^{4*}

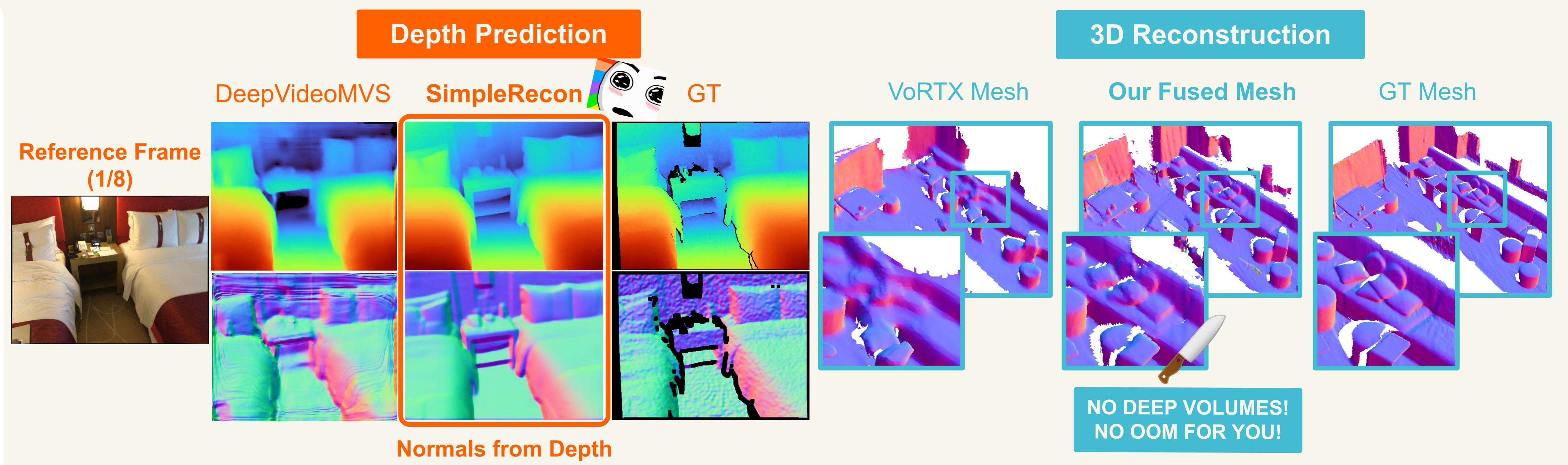
¹Niantic ²University College London ³University of Oxford ⁴Google
*Work done while at Niantic, during Mohamed's internship.



NO 3D CONVS! NO LSTMs!
Fast 3D Reconstruction!

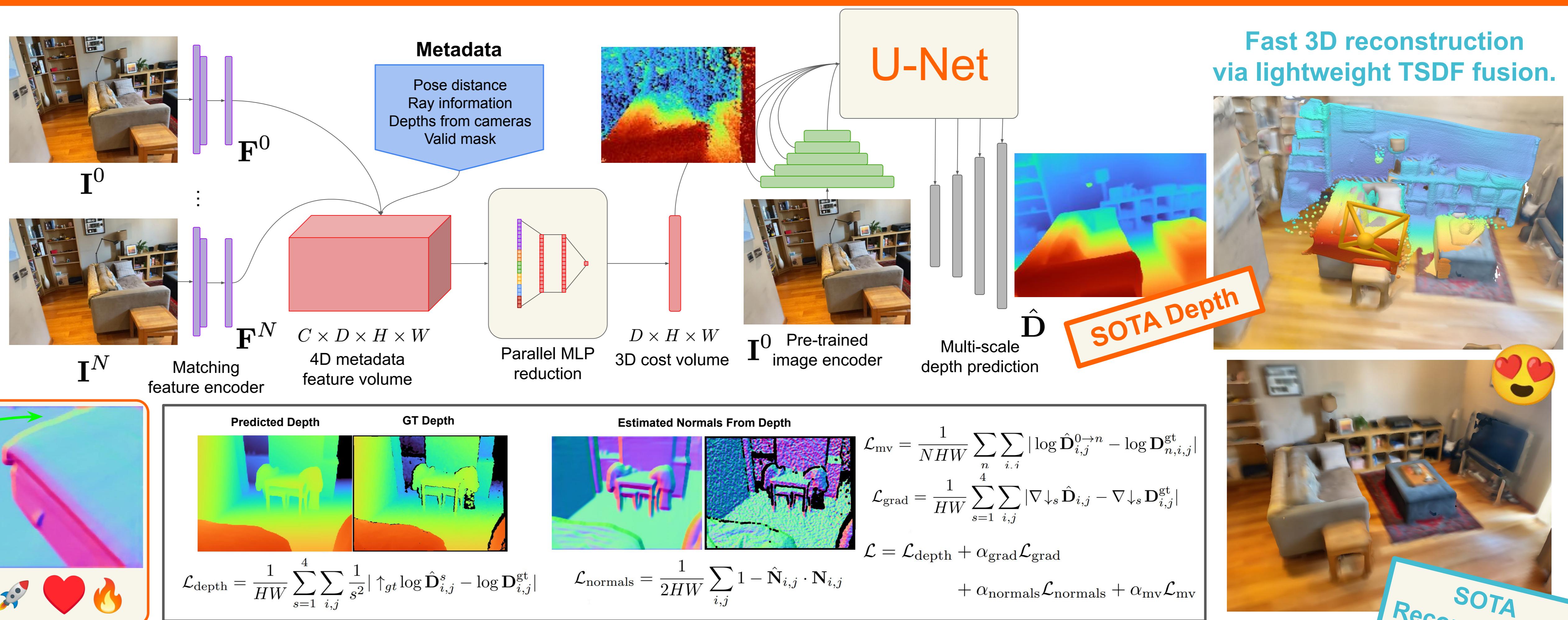
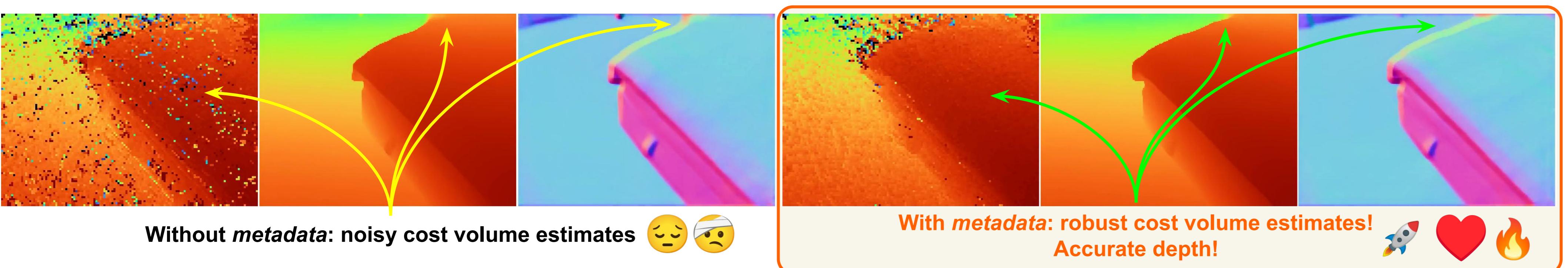
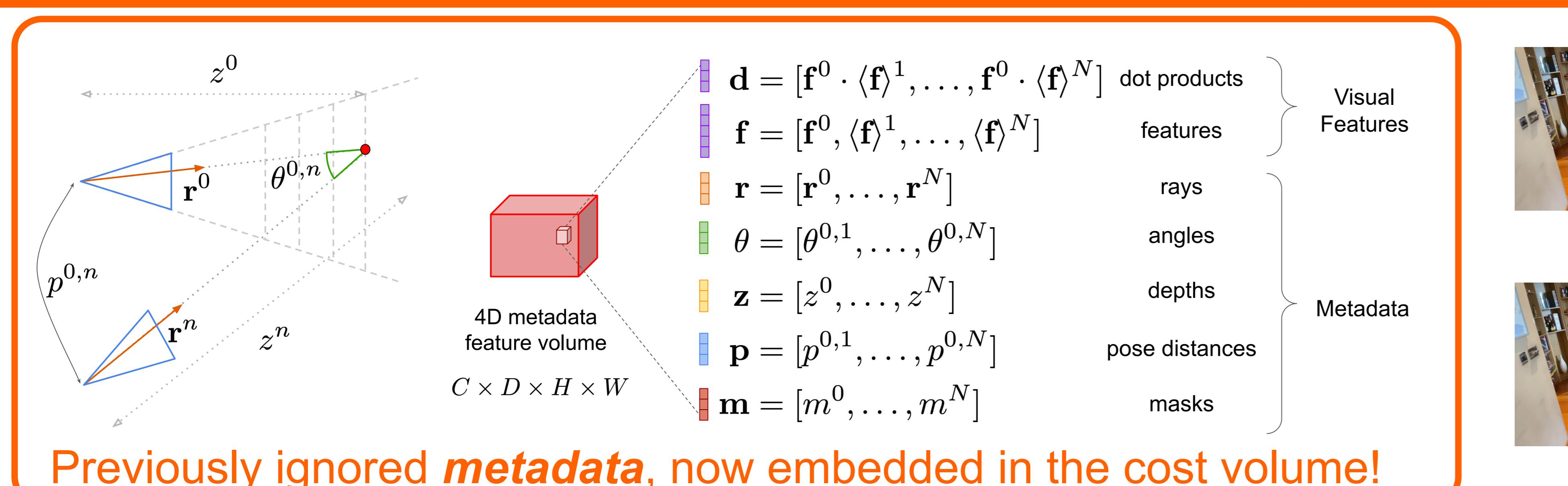
SimpleRecon

- ↓ Posed RGB Images
- ↑ Sharp metric monocular depth
- + SOTA monocular depth from video
- + SOTA 3D reconstruction
- Rocket Simple architecture, no 3D convs
- Rocket Fast 3D recon via off-the-shelf fusion
- Star Novel metadata



Placeholder for Samsung A8 tablet (sized)

Method

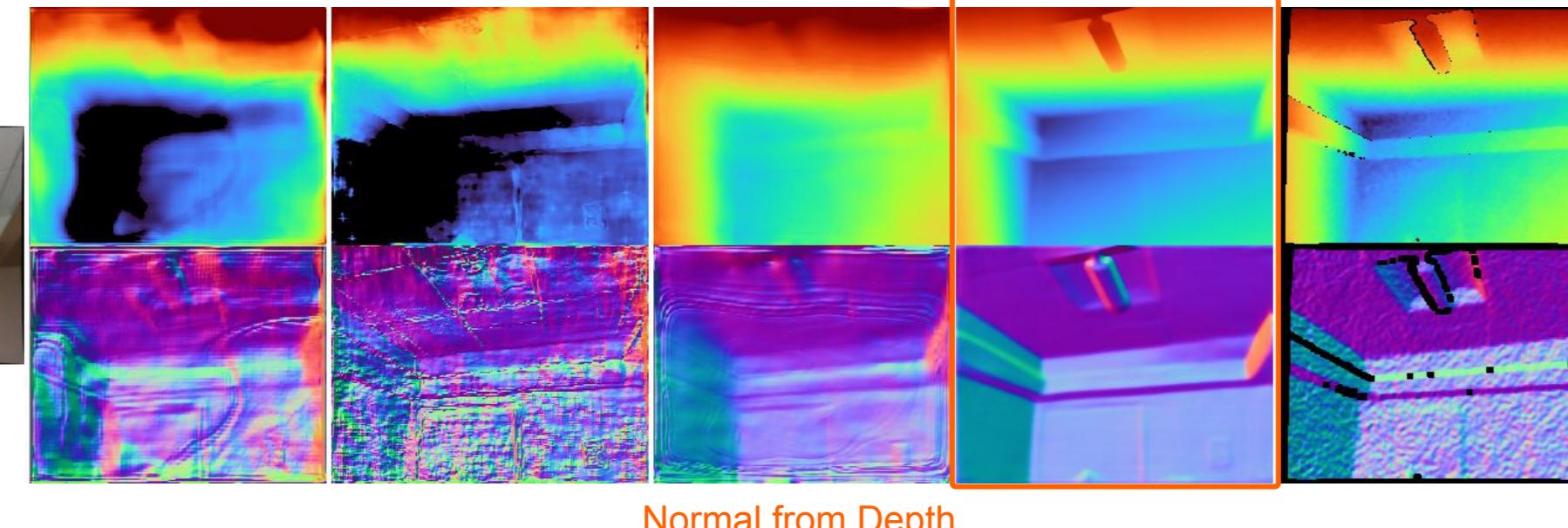
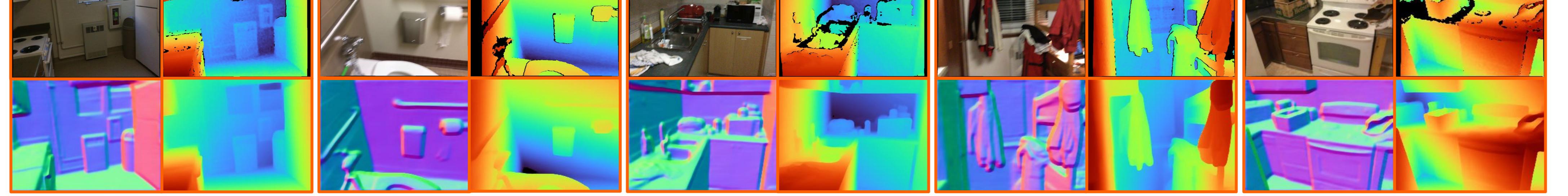


Depth Results

ScanNetv2						7Scenes					
Abs Diff↓	Abs Rel↓	Sq Rel↓	$\delta < 1.05 \uparrow$	$\delta < 1.25 \uparrow$		Abs Diff↓	Abs Rel↓	Sq Rel↓	$\delta < 1.05 \uparrow$	$\delta < 1.25 \uparrow$	
DPSNet [26]	0.1552	0.0795	0.0299	49.36	93.27	0.1966	0.1147	0.0550	38.81	87.07	
MVDepthNet [70]	0.1648	0.0848	0.0343	46.71	92.77	0.2009	0.1161	0.0623	38.81	87.70	
DELTAS [62]	0.1497	0.0786	0.0276	48.64	93.78	0.1915	0.1140	0.0490	36.36	88.13	
GPMVS [23]	0.1494	0.0757	0.0292	51.04	93.96	0.1739	0.1003	0.0462	42.71	90.32	
DeepVideoMVS, fusion [12]*	0.1186	0.0583	0.0190	60.20	96.76	0.1448	0.0828	0.0335	47.96	93.79	
Ours (no metadata)	0.0941	0.0467	0.0139	70.48	97.84	0.1105	0.0617	0.0175	57.30	97.02	
Ours	0.0885	0.0434	0.0125	73.16	98.09	0.1045	0.0575	0.0153	59.78	97.38	

*trained with more data

metadata leads to higher accuracy



Reconstruction Results

