SimpleRecon
3D Reconstruction Without 3D Convolution

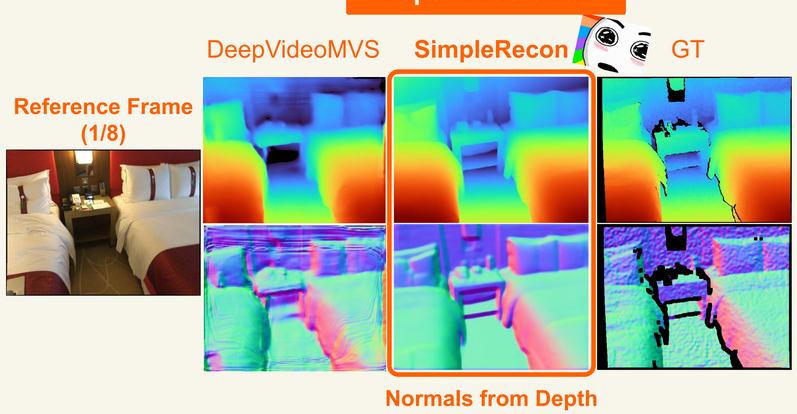
Mohamed Sayed[2*] **John Gibson**[1] **Jamie Watson**[1] **Victor Adrian Prisacariu**[1,3]
Michael Firman[1] **Clément Godard**[4*]

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

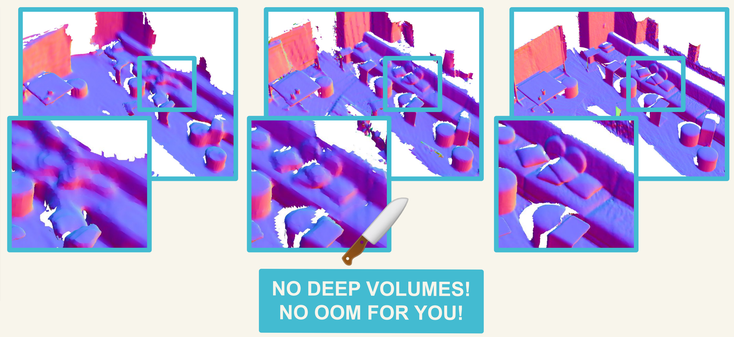
SimpleRecon

- Posed RGB Images
- Sharp metric monocular depth
- **SOTA** monocular depth from video
- **SOTA** 3D **recon**struction
- **Simple** architecture, no 3D convs
- Fast 3D recon via off-the-shelf fusion
- Novel *metadata*

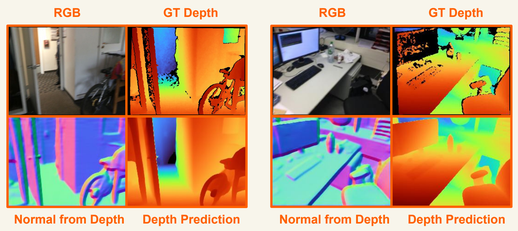
Depth Prediction



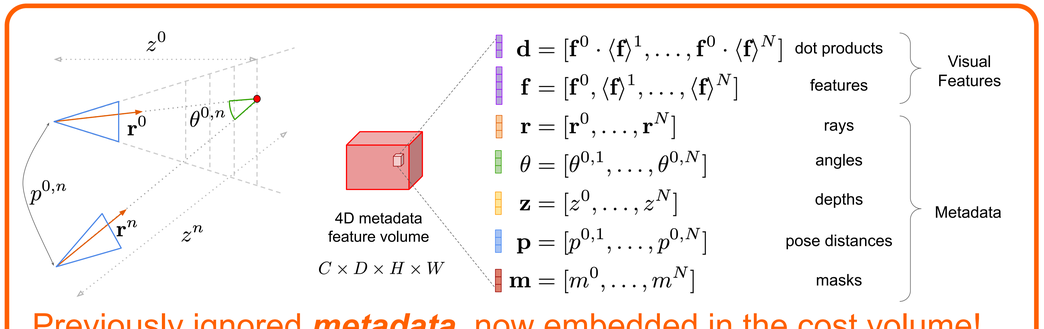
3D Reconstruction



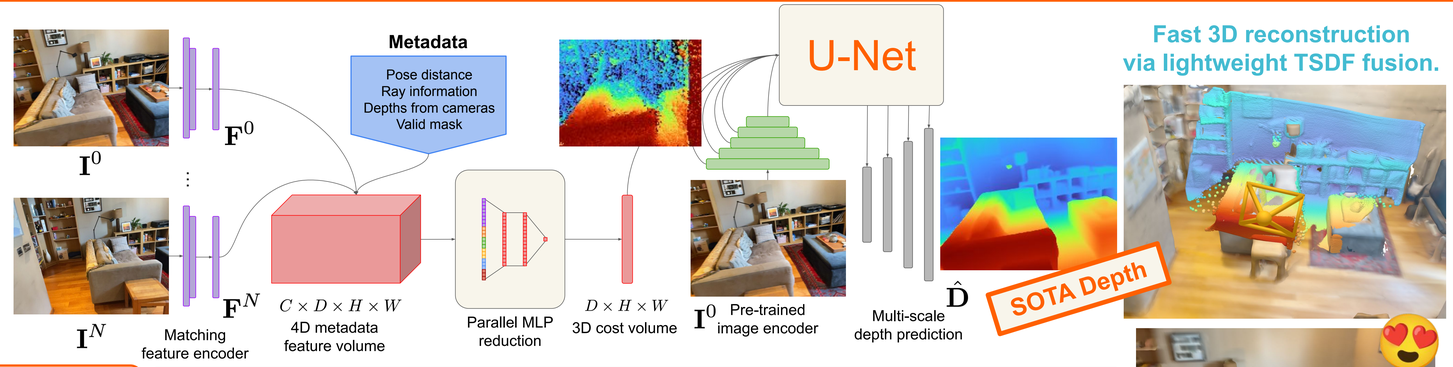
NO 3D CONVS! NO LSTMs!
Fast 3D Reconstruction!

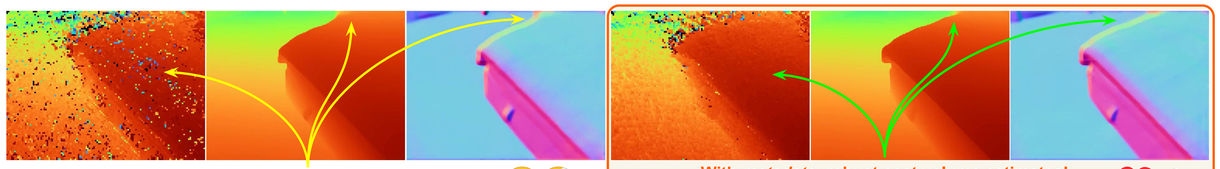


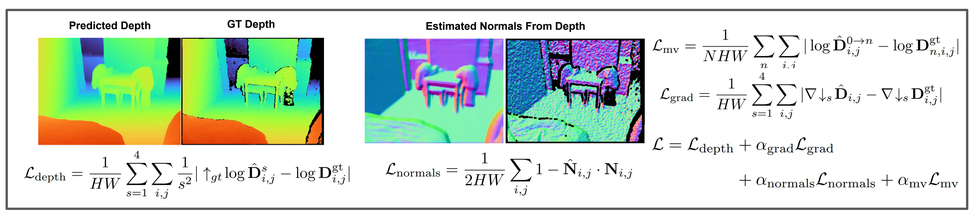
Method



Previously ignored ***metadata***, now embedded in the cost volume!







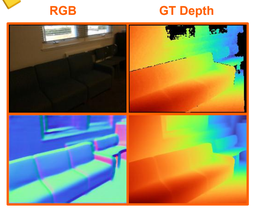
Depth Results

	ScanNetv2					7Scenes				
	Abs Diff↓	Abs Rel↓	Sq Rel↓	$\delta < 1.05$ ↑	$\delta < 1.25$ ↑	Abs Diff↓	Abs Rel↓	Sq Rel↓	$\delta < 1.05$ ↑	$\delta < 1.25$ ↑
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

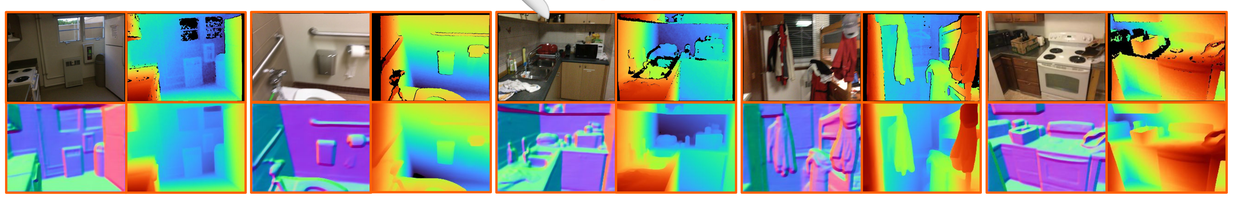
*trained with more data

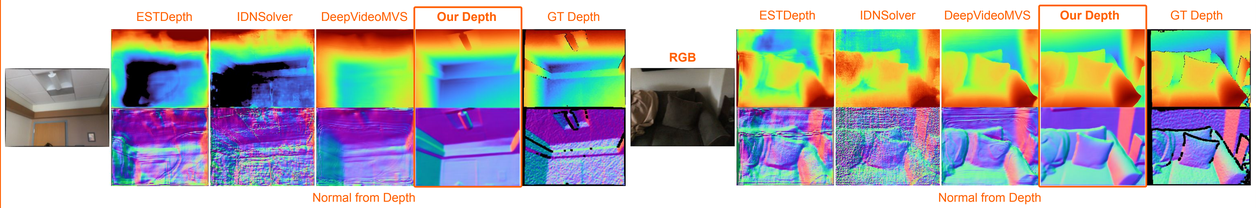
metadata leads to higher accuracy

accurate metric depth



sharp edges





Reconstruction Results

ScanNetv2 (**Ours** uses online DeepVideoMVS keyframes)

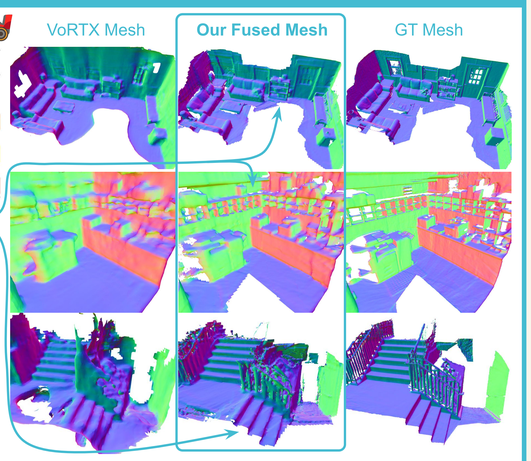
	Volumetric	Comp↓	Acc↓	Chamfer↓	Prec↑	Recall ↑	F-Score ↑	**Update**	**Latency** (ms)↓
COLMAP [58]	No	10.22	11.88	11.05	0.509	0.474	0.489		-
ATLAS [46]	Yes	7.16	7.61	7.38	0.675	0.605	0.636		382ms
NeuralRecon [65]	Yes	5.09	9.13	7.11	0.630	0.612	0.619		90ms
3DVNet [54]	Yes	7.72	6.73	7.22	0.655	0.596	0.621		23875ms
TransformerFusion [2]	Yes	5.52	8.27	6.89	0.728	0.600	0.655		326ms
VoRTX [64]	Yes	4.31	7.23	5.77	0.767	0.651	0.703		4550ms
Ours	No	5.53	6.09	5.81	0.686	0.658	0.671		72ms

fast updates

Fusion

	Volumetric	Comp↓	Acc↓	Chamfer↓	Prec↑	Recall ↑	F-Score ↑
3DVNet	Yes	5.84	7.53	6.68	0.704	0.620	0.655
Ours	No	4.99	9.14	7.06	0.646	0.681	0.659
Ours FPN, Offline source frames	No	4.71	8.84	6.78	0.669	0.692	0.676

sharp and accurate detail



Out of 3D domain reconstruction (multi story house)

Metadata helps in other MVS pipelines!

	Abs Diff↓	Abs Rel↓	Sq Rel↓	$\delta < 1.05$ ↑	$\delta < 1.10$ ↑	Acc↓	Comp↓	Overall↓
CasMVSNet	11.05	0.0143	2.153	92.80	95.95	**0.347**	0.341	0.344
CasMVSNet + **Metadata**	**10.58**	**0.0133**	**2.067**	**93.64**	**96.15**	0.368	**0.314**	**0.341**