



# SCRREAM : SCan, Register, REndeR And Map: A Framework for Annotating Accurate and Dense3D Indoor Scenes with a Benchmark

Hyunjun Jung<sup>1</sup>, Weihang Li<sup>1</sup>, Shun-Cheng Wu<sup>1</sup>, William Bittner<sup>1</sup>, Nikolas Brasch<sup>1</sup>, Jifei Song<sup>2</sup>, Eduardo Pérez-Pellitero<sup>2</sup>, Zhensong Zhang<sup>2</sup>, Arthur Moreau<sup>2</sup>, Nassir Navab<sup>1</sup>, Benjamin Busam<sup>1,3</sup>

<sup>1</sup> Technical University of Munich, <sup>2</sup> Huawei Noah's Ark, <sup>3</sup> 3dwe.ai

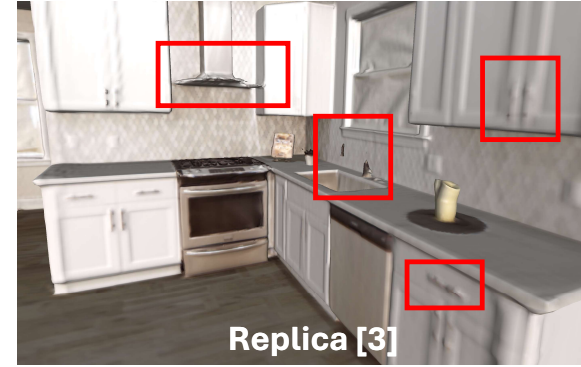
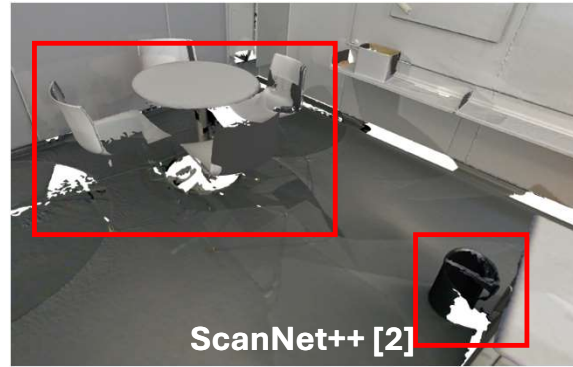
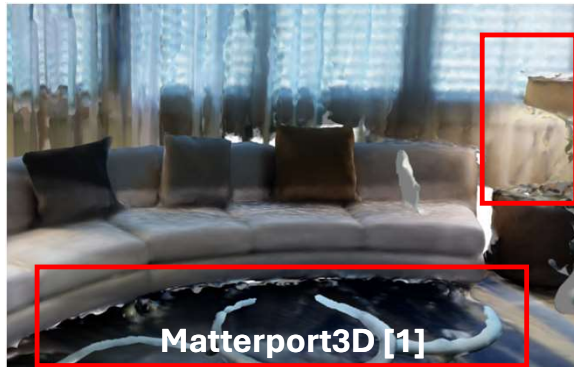
[hyunjun.jung@tum.de](mailto:hyunjun.jung@tum.de), [b.busam@tum.de](mailto:b.busam@tum.de)



Dec.12, 1630 Poster Session 4

# SCRREAM : Motivation

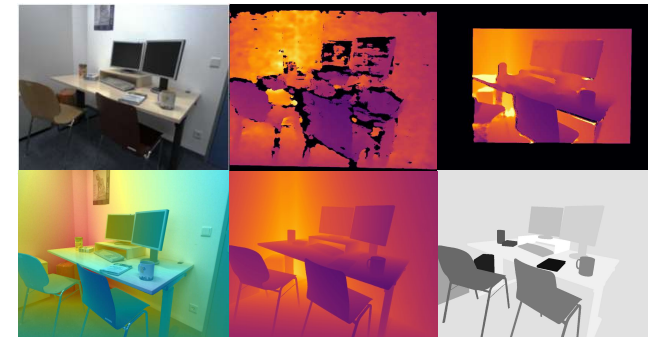
Commonly Used Indoor Dataset : Scan Entire Scene !



Indoor Dataset Annotated by SCRREAM Pipeline : Scan Individually and Register!



Individually scanned high resolution meshes registered to the real scene



Real RGBD Sequence + Rendered Ground Truth

[1] A.Chang, A.Dai, T.Funkhouser, M.Halber, M.Niessner, M.Savva, S.Song, A.Zeng, Y.Zhang "Matterport3d: Learning from rgb-d data in indoor environments" 3DV 2017

[2] C.Yeshwanth, Y.C.Liu, M.Nießner, A.Dai "Scannet++: A high-fidelity dataset of 3d indoor scenes" ICCV 2023

[3] J.Straub, T.Whelan, L.Ma, Y.Chen, E.Wijmans, S.Green, J.J.Engel, R.Mur-Artal, C.Ren, S.Verma, A.Clarkson, M.Yan, B.Budge, Y.Yan, X.Pan, J.Yon, Y.Zou, K.Leon, N.Carter, J.Briales, T.Gillingham, E.Mueggler, L.Pesqueira, M.Savva, D.Batra, H.M.Strasdat, R.D.Nardi, M.Goesele, S.Lovegrove, R.Newcombe "The Replica dataset: A digital replica of indoor spaces" arXiv 2019

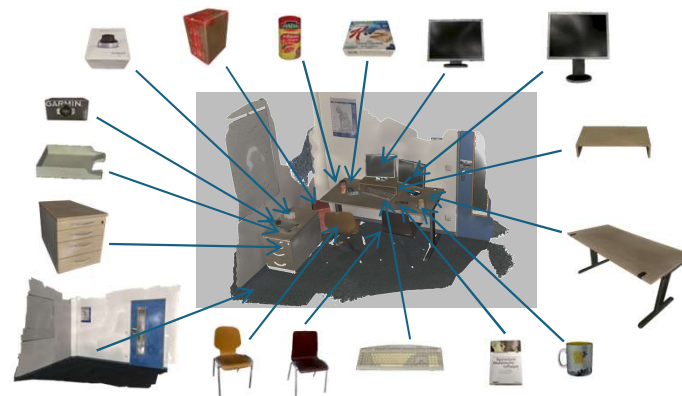
# SCRREAM : Intro

## 1. SCan



Scan Individual Objects, Furniture and Room

## 2. Register



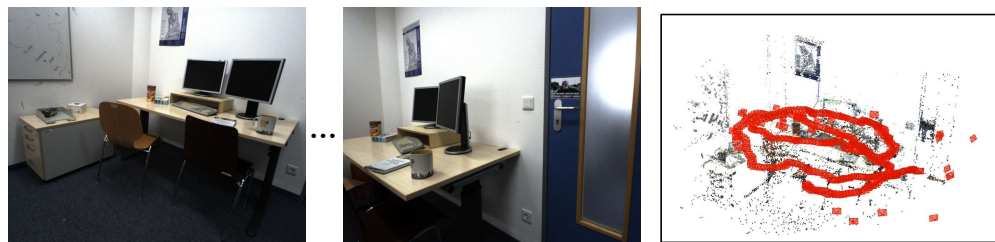
Register the Scanned Meshes on the Scene

## 3. REnd



Synthetic Image Rendering and Feature De-projection

## 4. Map (real images)



Map Real Image Sequence to Synthetic Feature

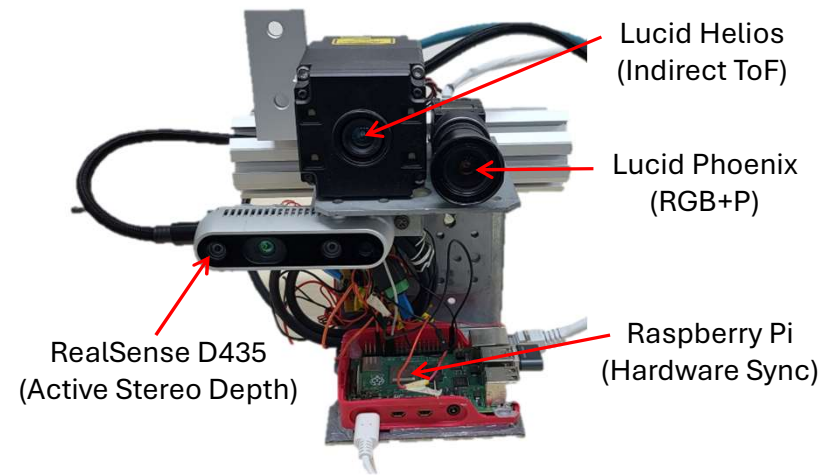
# SCRREAM : Hardware



(a) Shining 3D EinScan

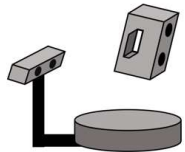


(b) Artec Leo Hand-Held Scanner

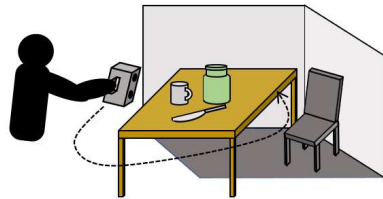


(c) Camera Rig

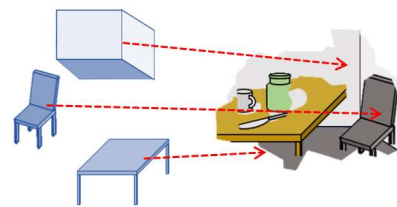
# SCRREAM : Pipeline



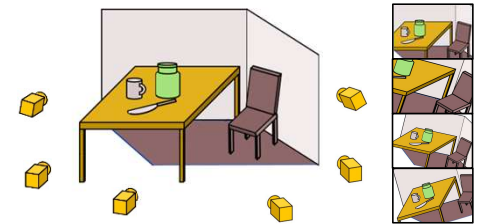
(a) Pre-scan objects, furniture and rooms



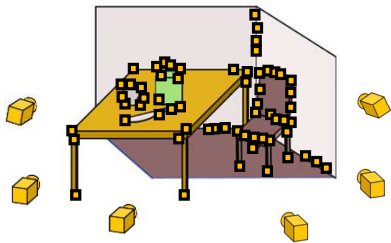
(b) Partial scanning of the scene



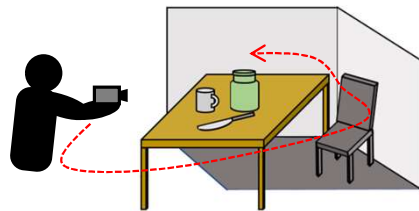
(c) Registering pre-scanned meshes to partially scanned mesh



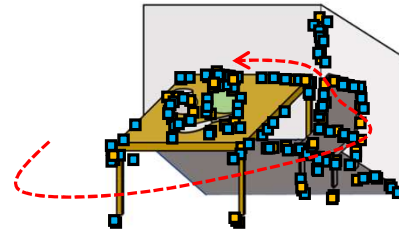
(d) Render the registered scene with realistic lighting



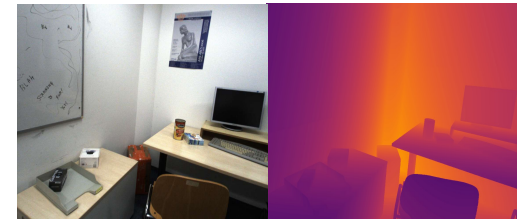
(e) Map feature with rendered image with GT pose



(f) Record a video sequence with the real camera rig

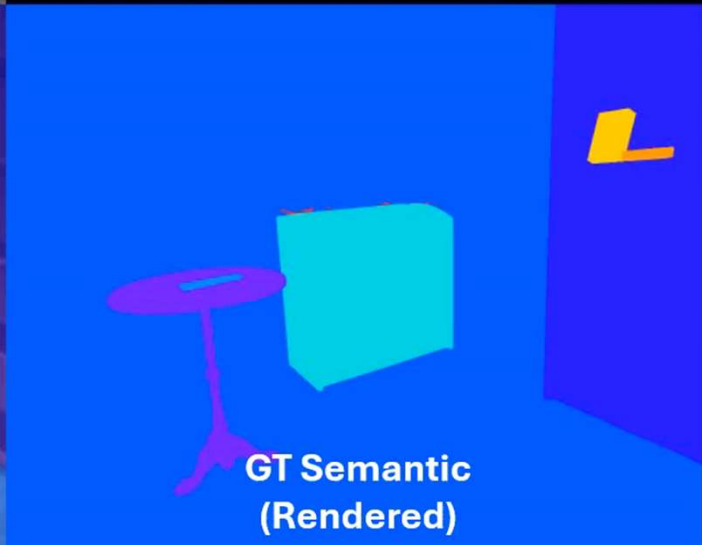
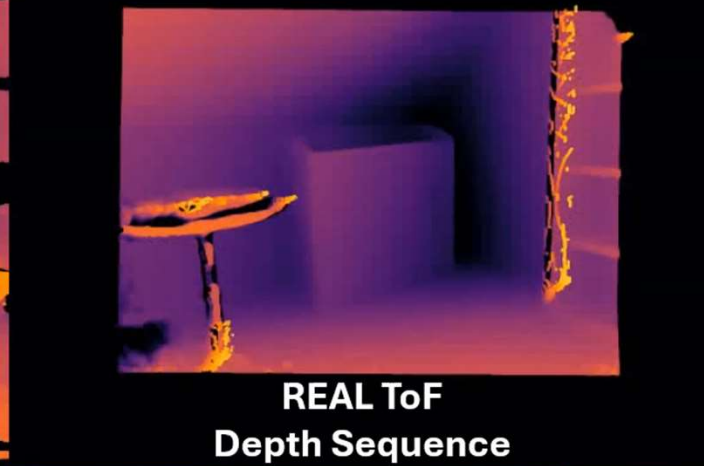
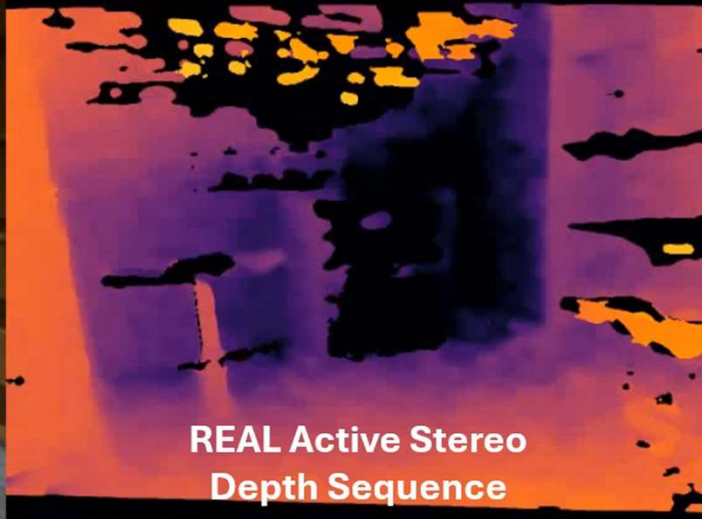


(g) Obtain the camera pose by feature matching on the synthetic feature



(h) Render 3D information by using the camera pose and the mesh

Two Stage Mapping



# SCRREAM : Indoor Reconstruction and SLAM Dataset



Scene01



Scene02



Scene03



Scene11



Scene04



Scene05



Scene06



Scene10



Scene07



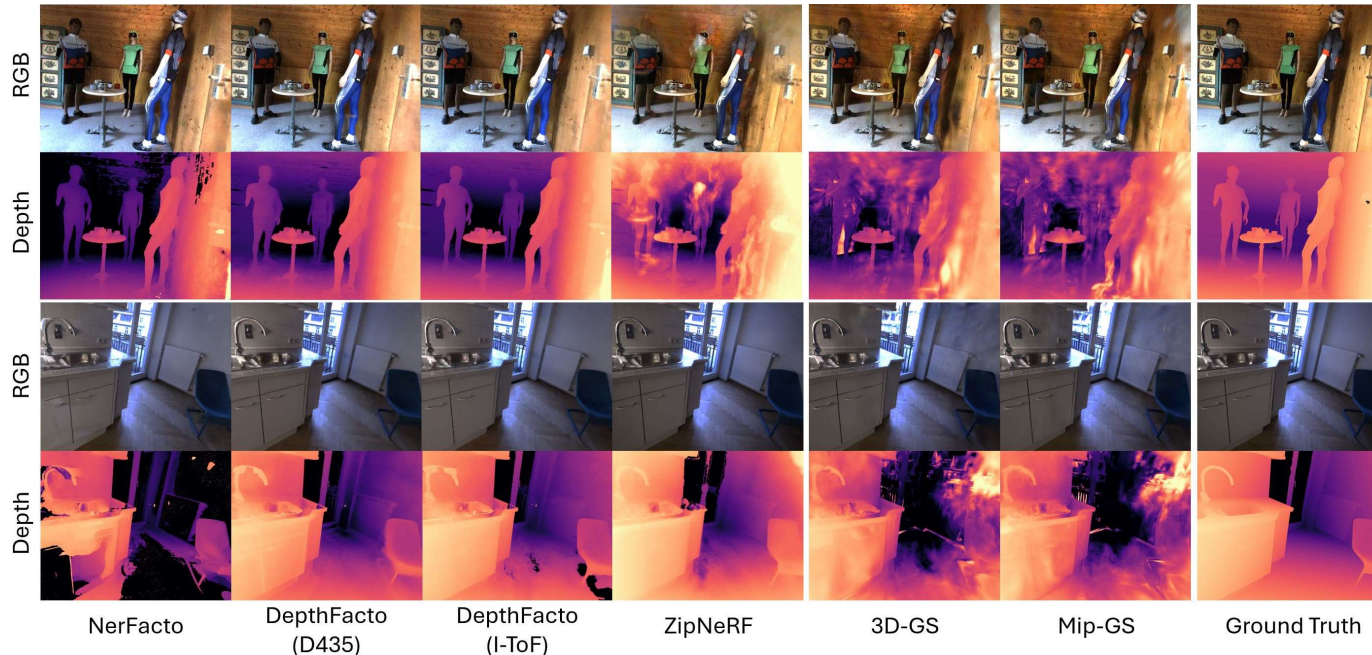
Scene08



Scene09

# SCRREAM : Indoor NVS Benchmark

Evaluation	RGB			Depth					
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	RSME $\downarrow$	Abs Rel $\downarrow$	Sqr Rel $\downarrow$	<1.25 <sup>1</sup> $\uparrow$	<1.25 <sup>2</sup> $\uparrow$	<1.25 <sup>3</sup> $\uparrow$
NeRFacto [30]	22.645	0.765	0.343	1.244	0.645	1.231	0.429	0.597	0.726
Depth-Facto [30] (Active Stereo)	24.502	0.786	0.324	<b>0.218</b>	<b>0.079</b>	<b>0.059</b>	<b>0.968</b>	<b>0.981</b>	<b>0.988</b>
Depth-Facto [30] (ToF)	24.540	0.788	0.323	0.336	0.093	0.149	0.926	0.943	0.957
Zip-NeRF [31]	<b>28.315</b>	0.783	<b>0.259</b>	0.493	0.245	0.189	0.546	0.825	0.887
Gaussian-Splatting [13]	25.943	0.801	0.328	0.526	0.218	0.174	0.589	0.806	0.883
Mip-Splatting [32]	25.925	<b>0.802</b>	0.327	0.532	0.223	0.178	0.594	0.794	0.875



[13] B.Kerbl, G.Kopanas, T.Leimkühler, G.Drettakis, "3d gaussian splatting for real-time radiance field rendering" ACM Transactions on Graphics 2023.

[30] M.Tancik, E.Weber, E.Ng, R.Li, B.Yi, T.Wang, A.Kristoffersen, J.Austin, K.Salahi, A.Ahuja "Nerfstudio: A modular framework for neural radiance field development" ACM SIGGRAPH 2023

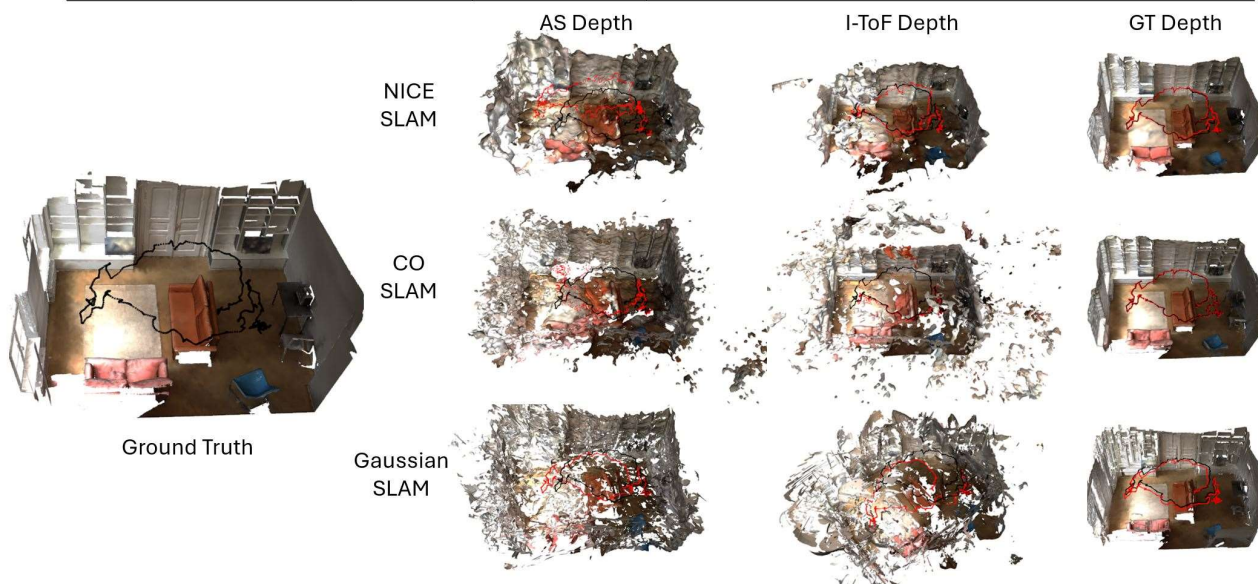
[31] J.T.Barron, B.Mildenhall, D.Verbin, P.P.Srinivasan, P.Hedman "Zip-nerf: Anti-aliased grid-based neural radiance fields" CVPR 2023

[32] Z.Yu, A.Chen, B.Huang, T.Sattler, A.Geiger "Mip-splatting: Alias-free 3d gaussian splatting" ArXiv 2023.



# SCRREAM : Indoor Reconstruction and SLAM Benchmark

Evaluation		Tracking	Mapping		
Methods	Depth	ATE↓ [cm]	Acc↓ [cm]	Comp↓ [cm]	Comp Ratio↑ [%]
NICE-SLAM [34]	AS	39.09	23.64	14.51	37.49
	ToF	24.02	21.60	14.63	46.18
	GT	5.47	2.91	5.91	77.11
CO-SLAM [35]	AS	39.19	32.38	12.95	40.07
	ToF	45.94	58.22	8.02	57.68
	GT	4.18	1.87	1.69	96.52
Gaussian-SLAM [36]	AS	38.20	37.94	11.74	41.36
	ToF	83.53	43.67	9.06	53.79
	GT	5.00	2.25	1.70	94.89



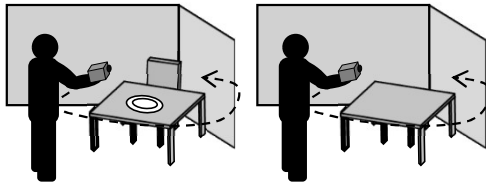
[34] Z. Zhu, S. Peng, V. Larsson, W. Xu, H. Bao, Z. Cui, M. R. Oswald, and M. Pollefeys, "Nice-slam: Neuralimplicit scalable encoding for slam," CVPR 2022.

[35] H. Wang, J. Wang, and L. Agapito, "Co-slam: Joint coordinate and sparse parametric encodings for neuralreal-time slam," CVPR, 2023

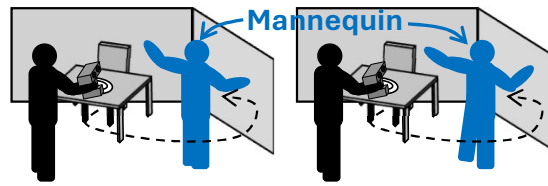
[36] V. Yugay, Y. Li, T. Gevers, and M. R. Oswald, "Gaussian-slam: Photo-realistic dense slam with gaussiansplatting," 2023.

# SCRREAM : Possible Variations

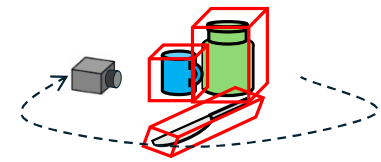
## Possible Dataset Variants



(a) Scene Editing Dataset



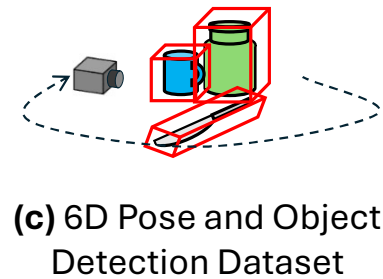
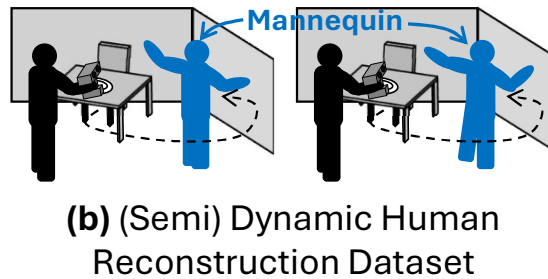
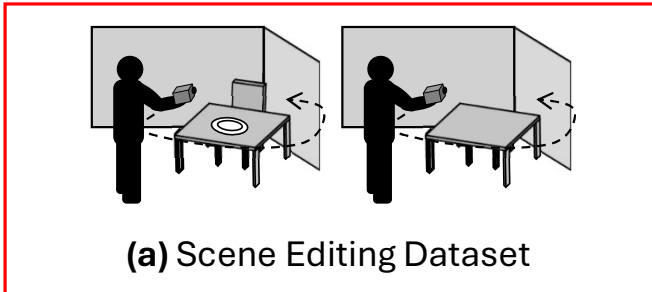
(b) (Semi) Dynamic Human Reconstruction Dataset



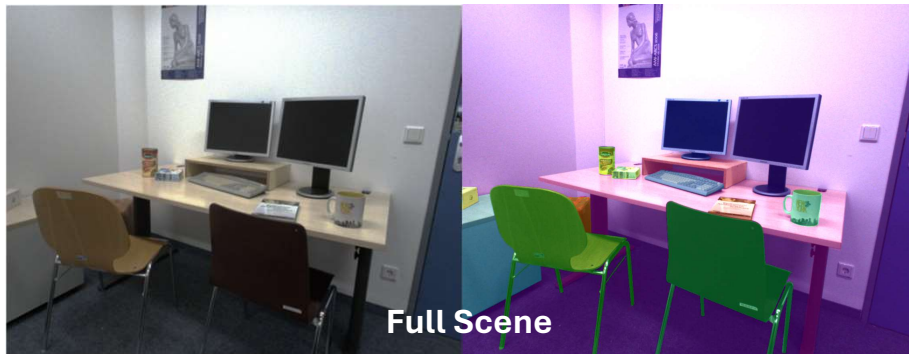
(c) 6D Pose and Object Detection Dataset

# SCRREAM : Possible Variations

## Possible Dataset Variants



## (a) Scene Editing Dataset

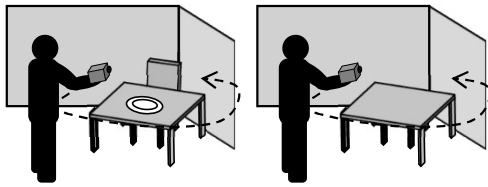


Object  
Removal  
→

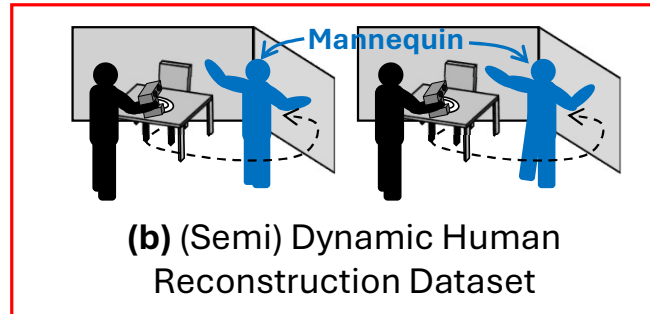


# SCRREAM : Possible Variations

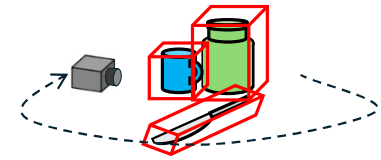
## Possible Dataset Variants



(a) Scene Editing Dataset



(b) (Semi) Dynamic Human Reconstruction Dataset



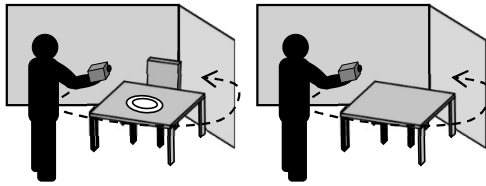
(c) 6D Pose and Object Detection Dataset

## (b) (Semi) Dynamic Human Reconstruction Dataset

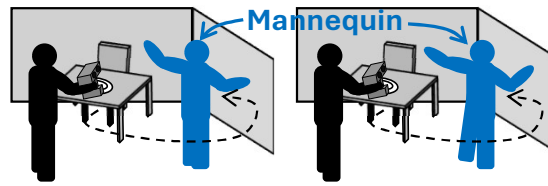


# SCRREAM : Possible Variations

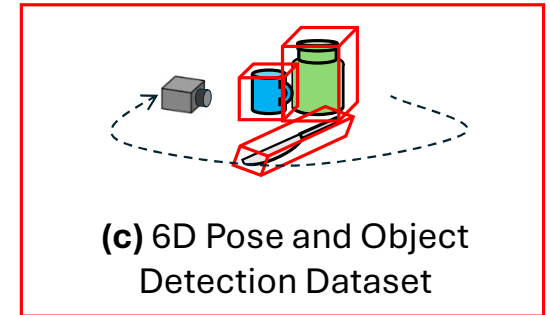
## Possible Dataset Variants



(a) Scene Editing Dataset

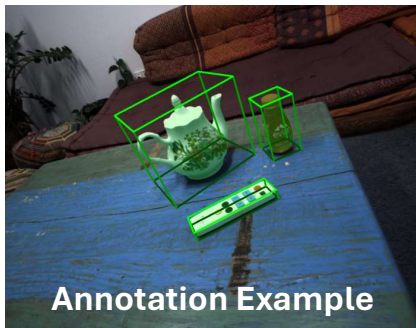


(b) (Semi) Dynamic Human Reconstruction Dataset

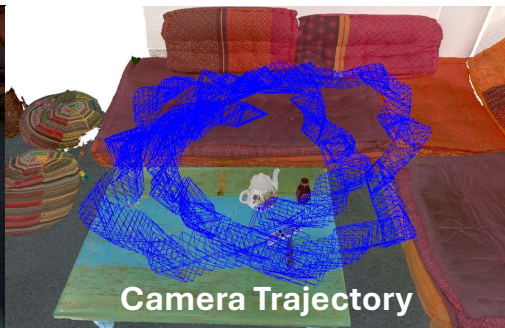


(c) 6D Pose and Object Detection Dataset

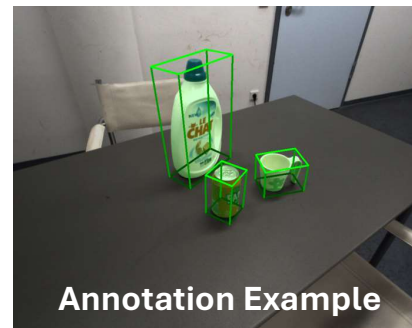
## (c) 6D Pose and Object Detection Dataset



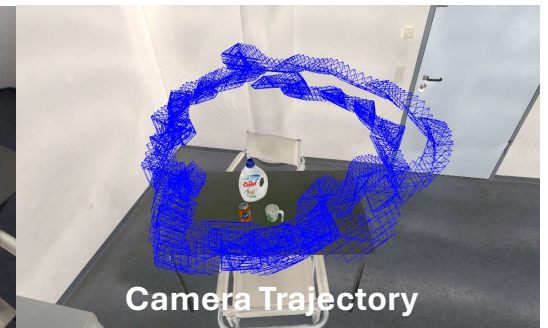
Annotation Example



Camera Trajectory



Annotation Example



Camera Trajectory



# SCRREAM : SCan, Register, REndeR And Map: A Framework for Annotating Accurate and Dense3D Indoor Scenes with a Benchmark

Hyunjun Jung<sup>1</sup>, Weihang Li<sup>1</sup>, Shun-Cheng Wu<sup>1</sup>, William Bittner<sup>1</sup>, Nikolas Brasch<sup>1</sup>, Jifei Song<sup>2</sup>, Eduardo Pérez-Pellitero<sup>2</sup>, Zhensong Zhang<sup>2</sup>, Arthur Moreau<sup>2</sup>, Nassir Navab<sup>1</sup>, Benjamin Busam<sup>1,3</sup>

<sup>1</sup> Technical University of Munich, <sup>2</sup> Huawei Noah's Ark, <sup>3</sup> 3dwe.ai

[hyunjun.jung@tum.de](mailto:hyunjun.jung@tum.de), [b.busam@tum.de](mailto:b.busam@tum.de)



Dec.12, 1630 Poster Session 4