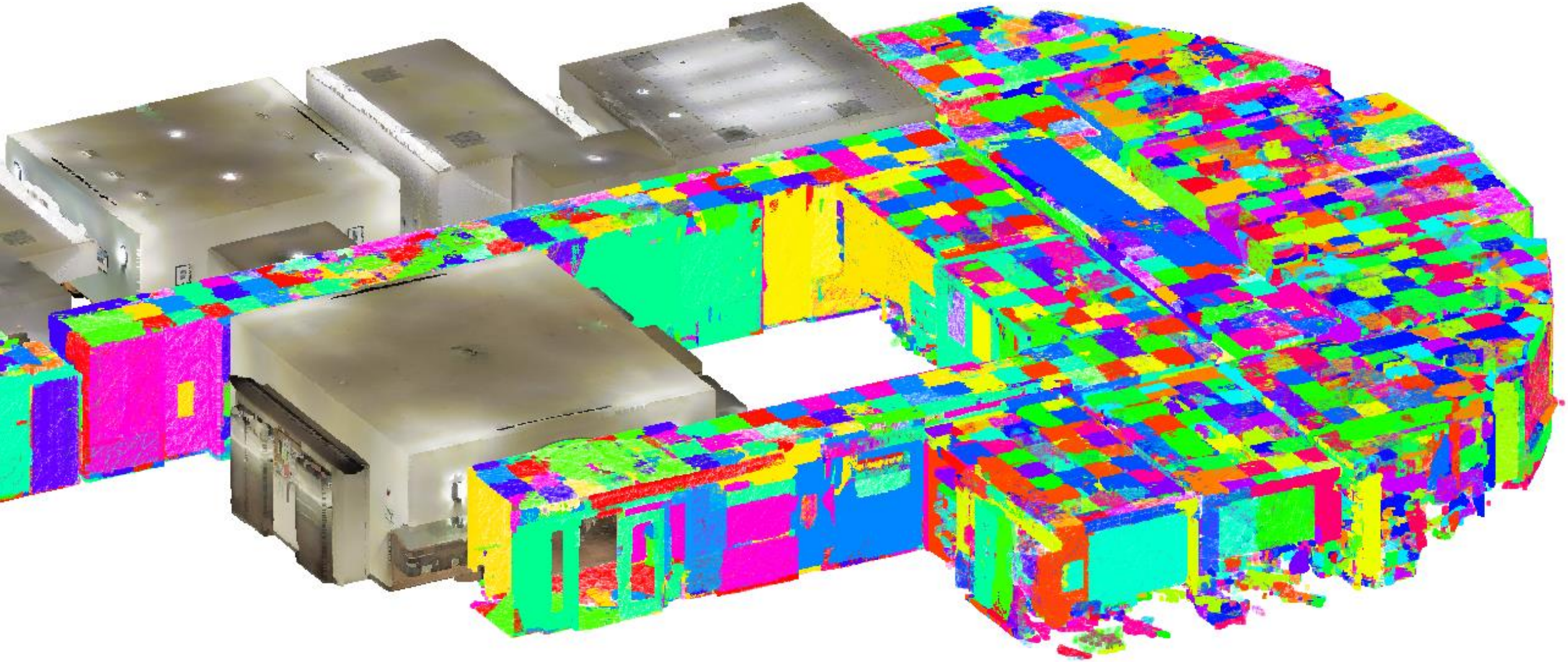


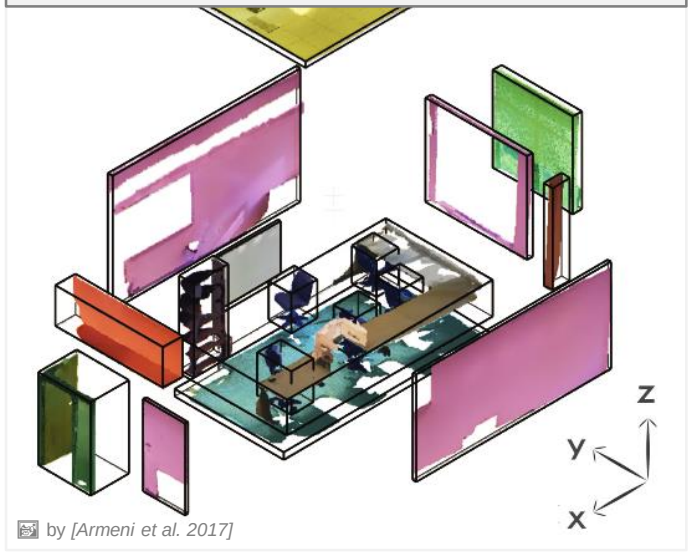
Nutzung von vortrainierten Vision Foundation Models bei der 3D-Punktwolkensegmentierung

Keno Moenck

IFPT

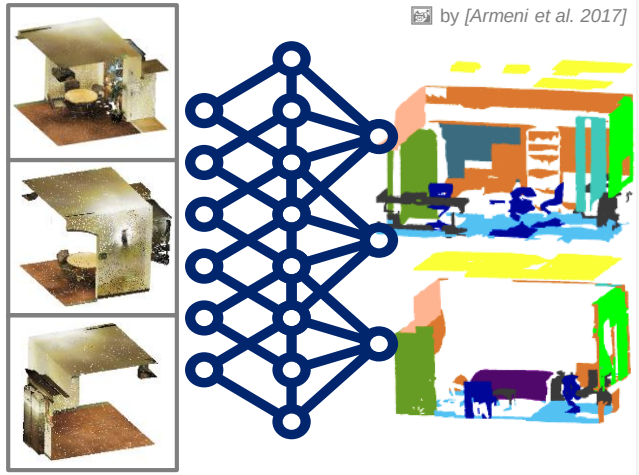
TUHH
Hamburg
University of
Technology

3D Scene Perception



by *[Armeni et al. 2017]*

State-of-the-art Approaches



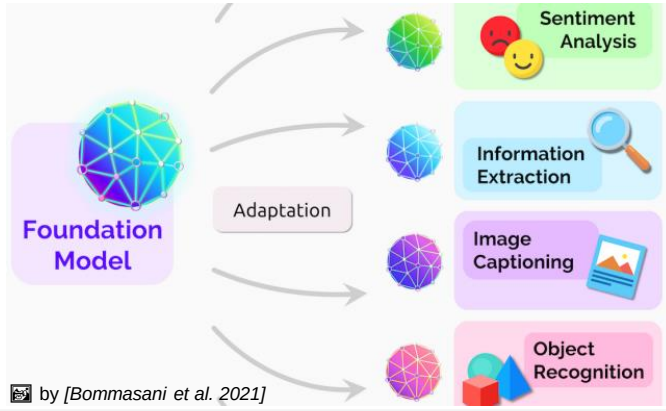
by *[Armeni et al. 2017]*

Rand-Domänen

Unspezifische Trainingsdaten

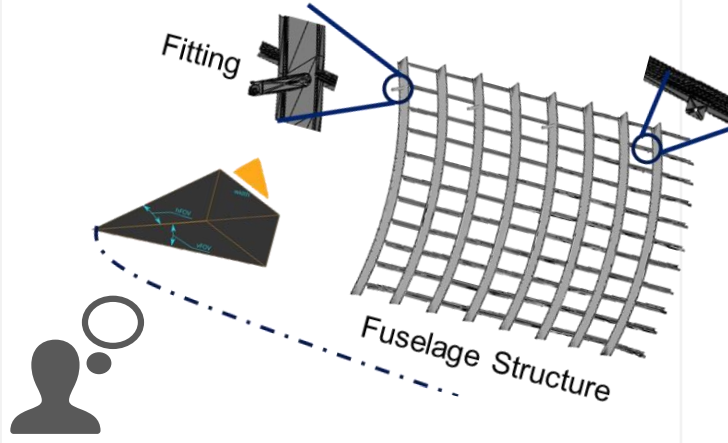
by *[Kirillov et al. 2023]*

Foundation Models



by *[Bommasani et al. 2021]*

Spezifische Trainingsdaten



[Awais et al. 2023]

Input *Model* *Output*

Traditionell

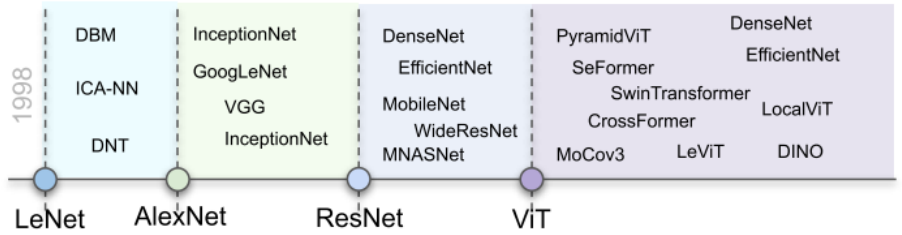
„the image shows …“

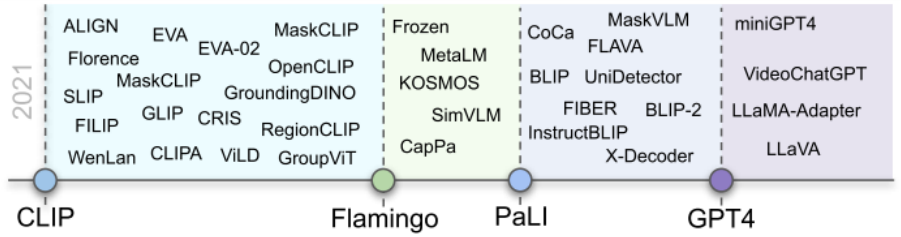
Text-Prompt

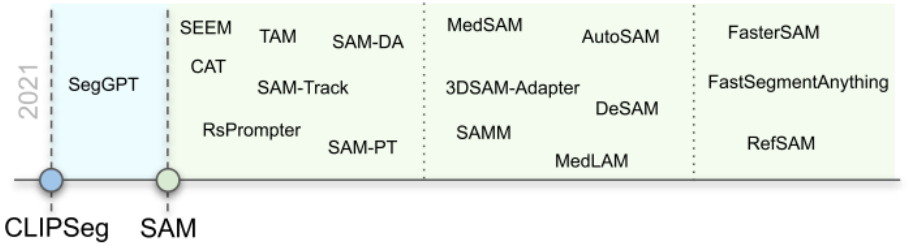
Visuell-Prompt

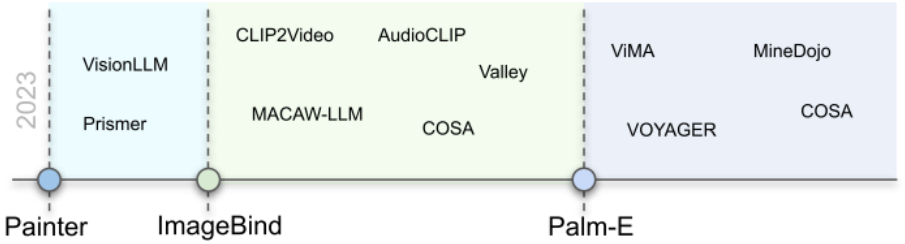
„the image shows …“

Multi-Modal-Prompt

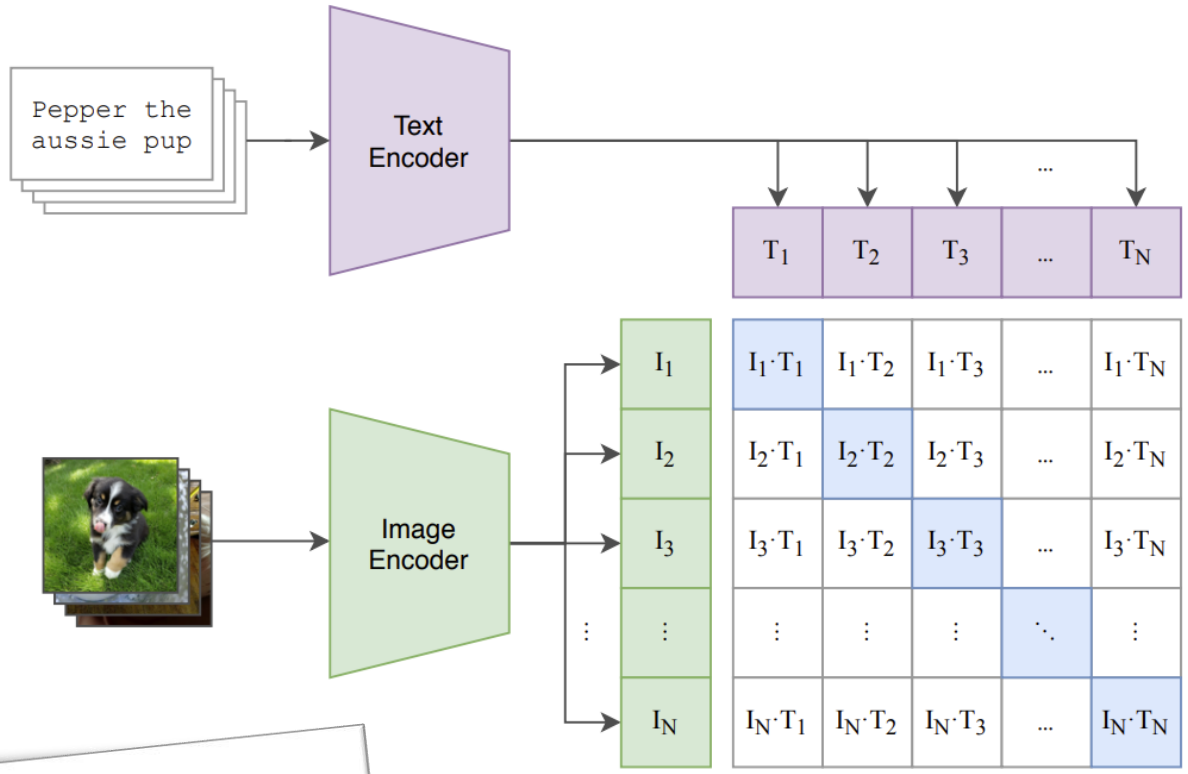


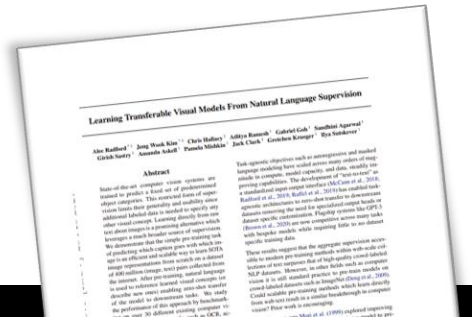




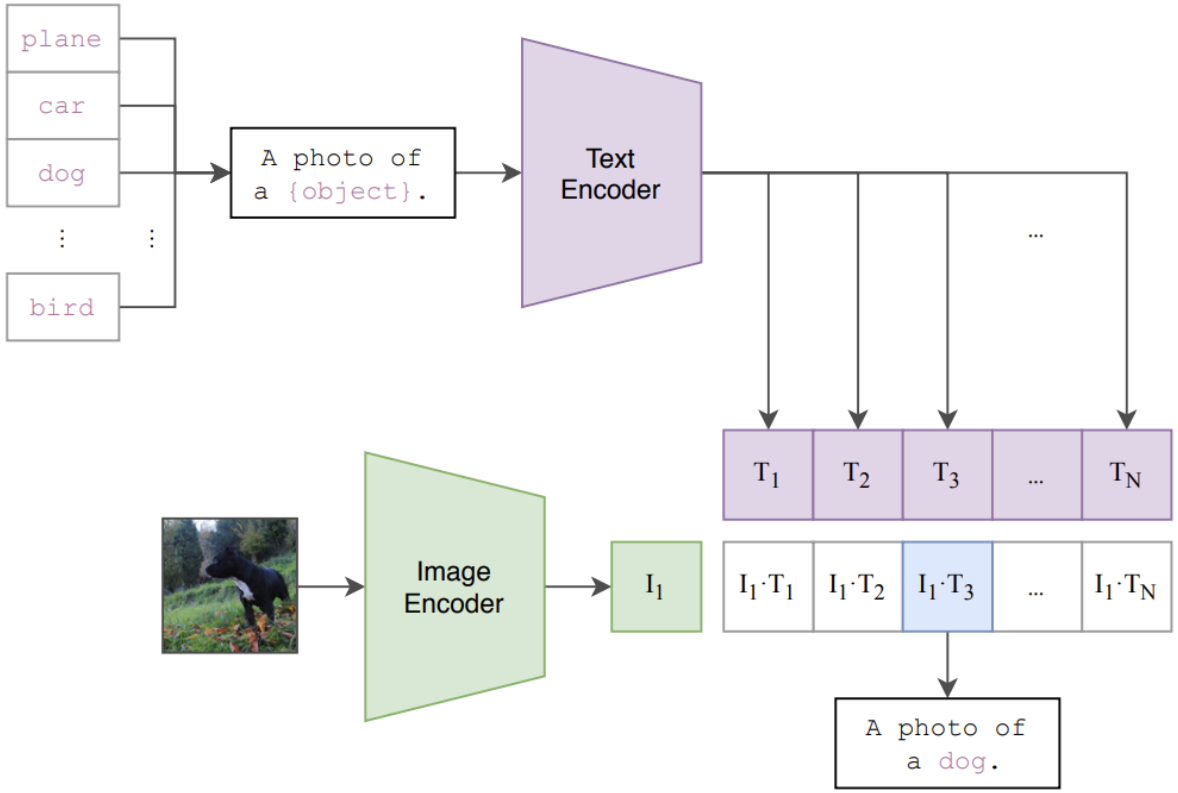


Contrastive Pre-Training

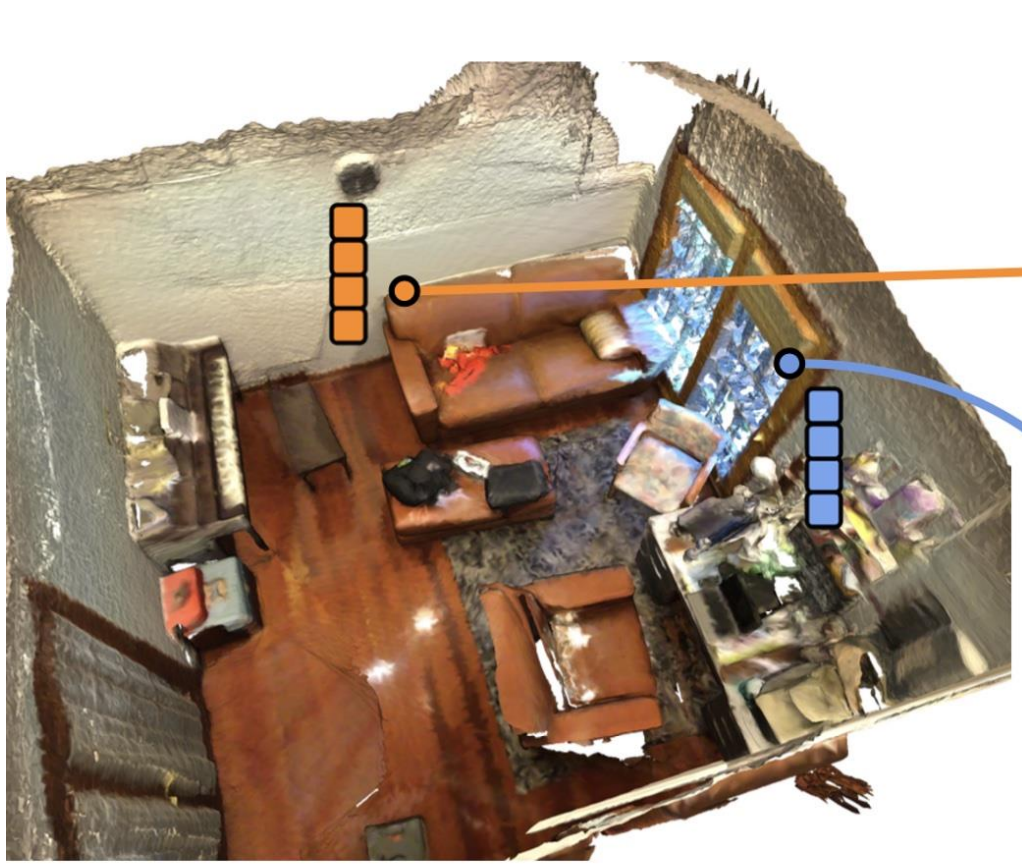



Learning Transferable Visual Models From Natural Language Supervision

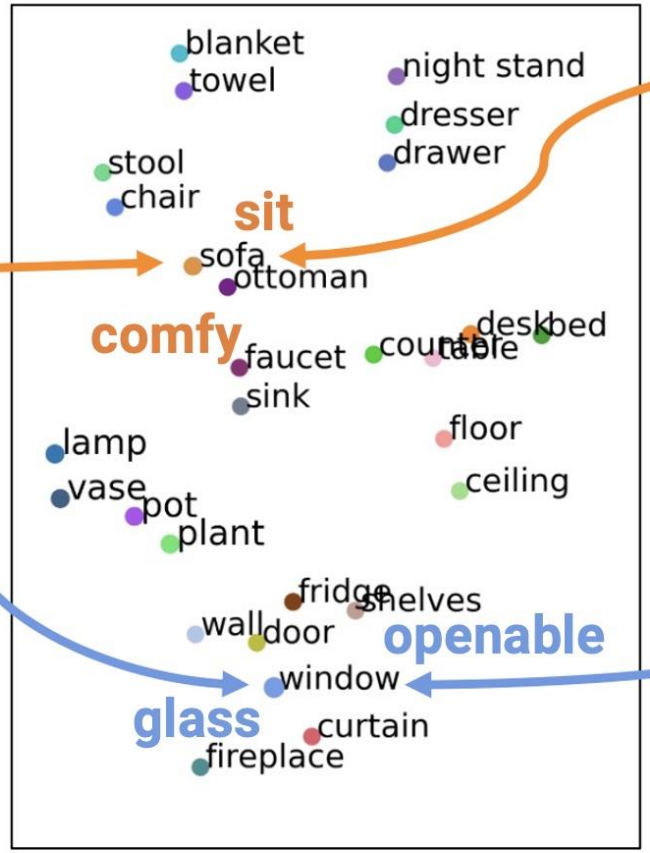
Zero-shot Prediction



[Radford et al. 2021]



3D Geometry



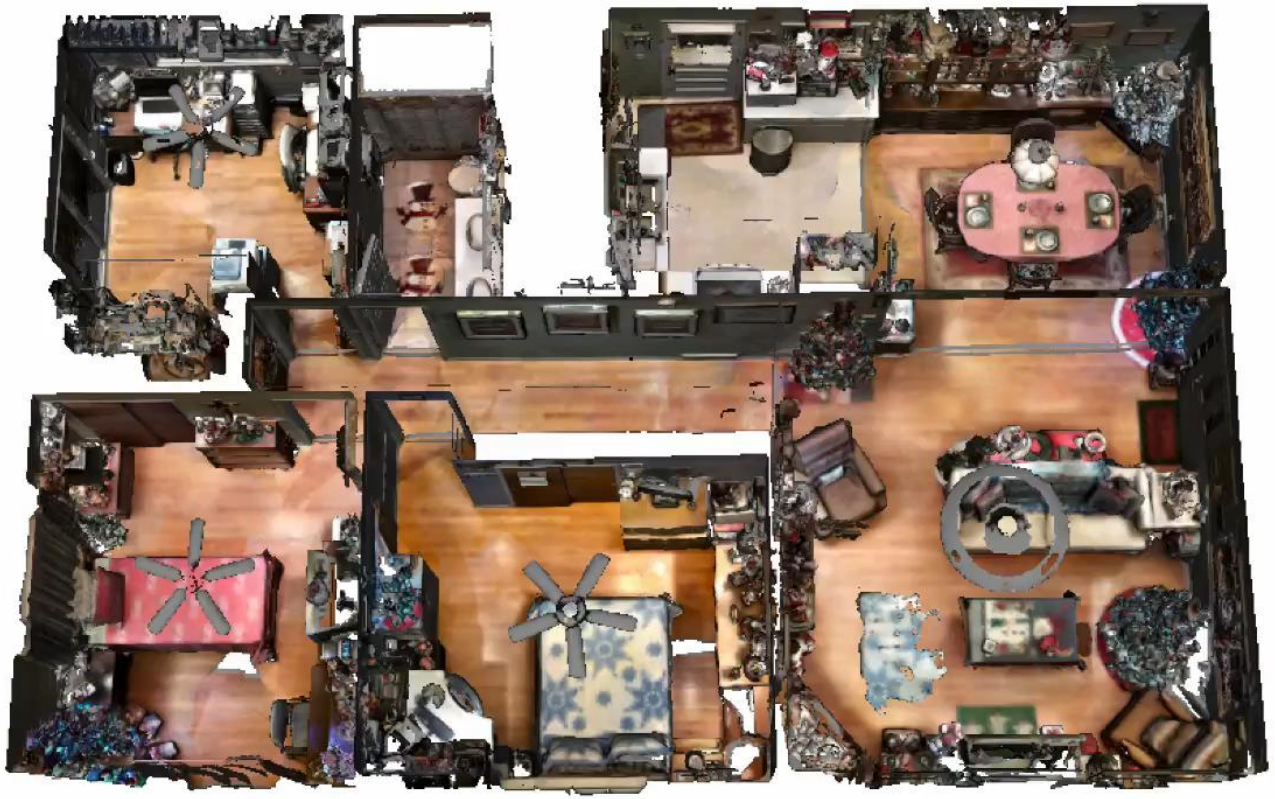
CLIP Text Features

(visualize with T-SNE)

RGB Images

[Peng et al. 2022]

Text queries:



[Peng et al. 2022]

Vision Foundation Models: Segment Anything Model (SAM)

[Kirillov et al. 2023]
[Moenck et al. 2023]

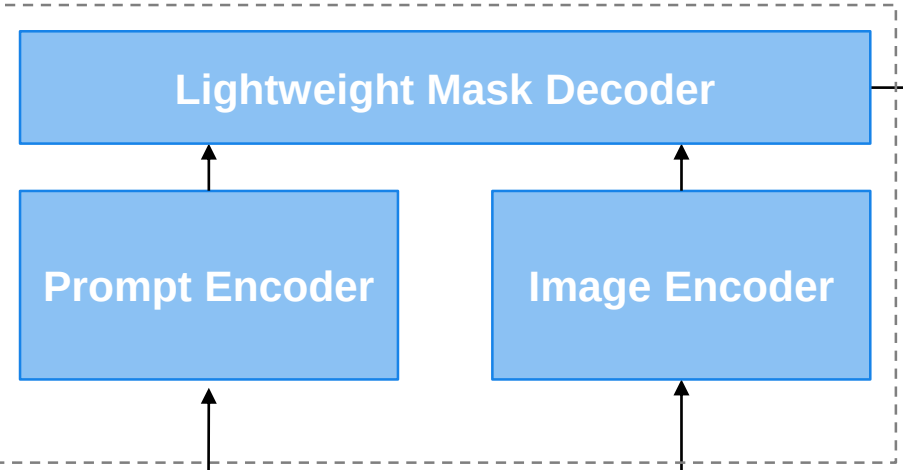

Segment Anything

Alexander Kirillov[1,2,4] Eric Mintun[2] Nikhila Ravi[1,2] Hanzi Mao[2] Chloe Rolland[3] Laura Gustafson[3]
Tete Xiao[3] Spencer Whitehead Alexander C. Berg Wan-Yen Lo Piotr Dollár[4] Ross Girshick[4]
[1]project lead [2]joint first author [3]equal contribution [4]directional lead
Meta AI Research, FAIR

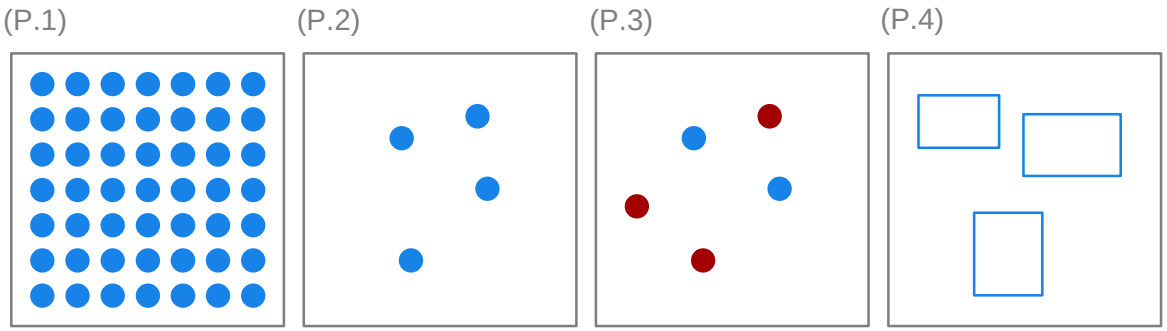
Figure 1: We aim to build a foundation model for segmentation by introducing three interconnected components: a promptable segmentation task, a segmentation model (SAM) that powers data annotation and enables zero-shot transfer to a range of tasks via prompt engineering, and a data engine for collecting SA-1B, our dataset of over 1 billion masks.

Abstract

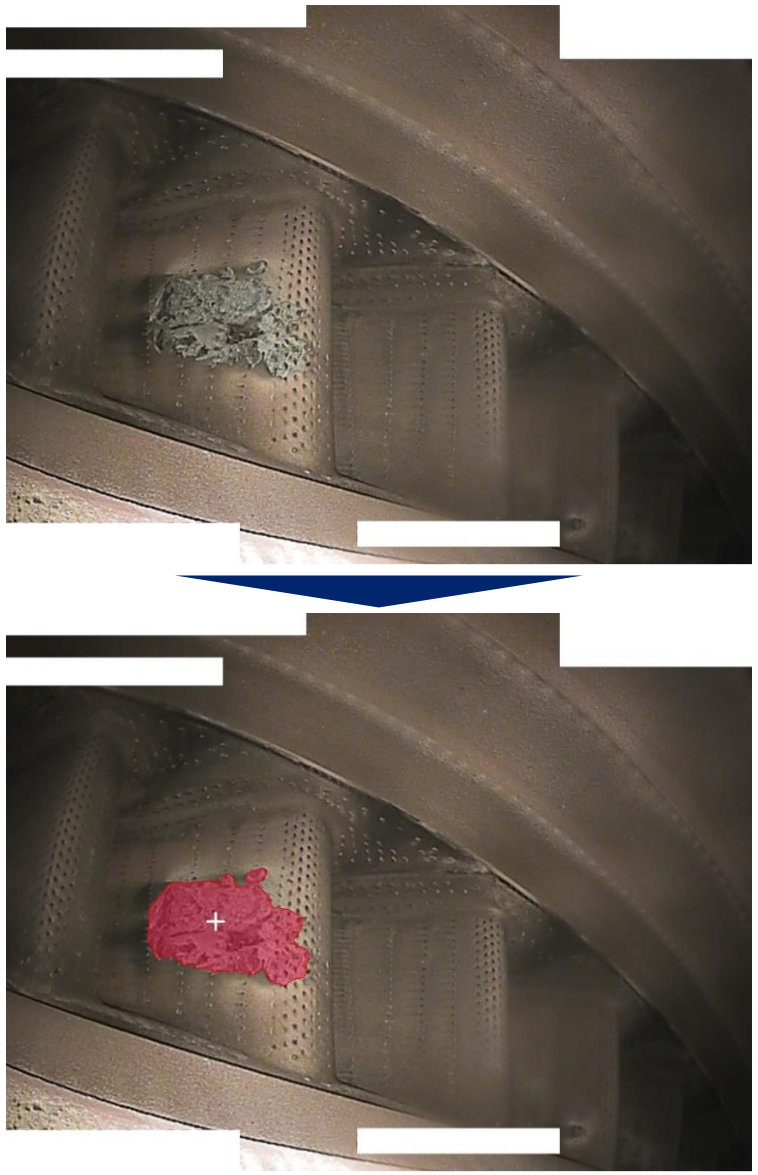
1. Introduction

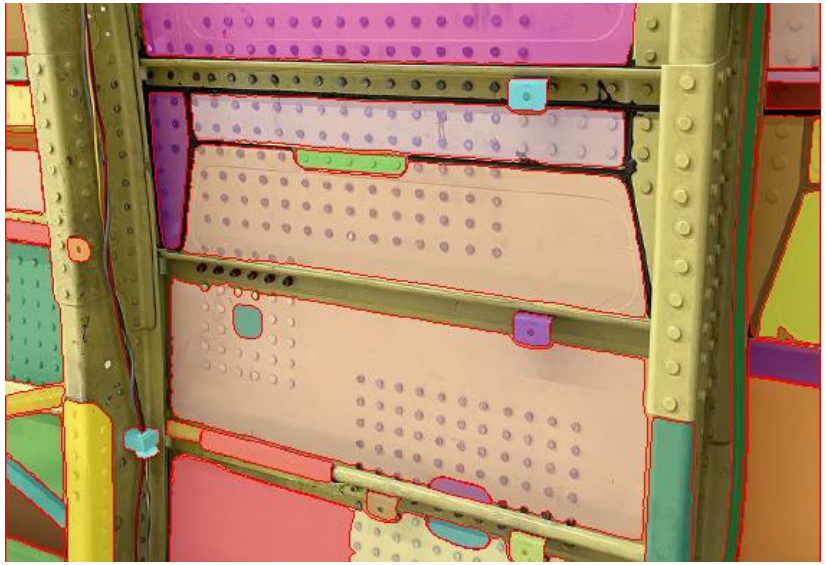


three levels of masks



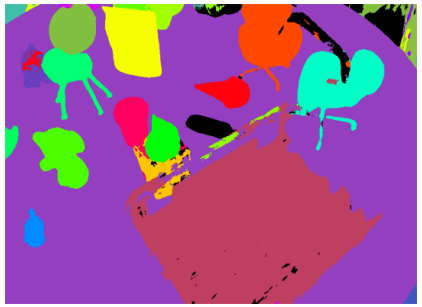
- SA-1b Datensatz: 11 Mio. Bilder / 1+ Mrd. Masken
- Datensatz durch iterative Erstellung: Modell-assistiert, semi-automatisch, voll-automatisiert
- Hierarchische Masken: Whole, Part, Sub-part
- Maskierung von semantisch-sinnvollen Regionen
- Klassen-agnostisch



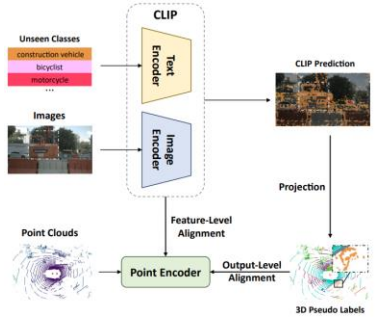


Super-/Semi-supervised

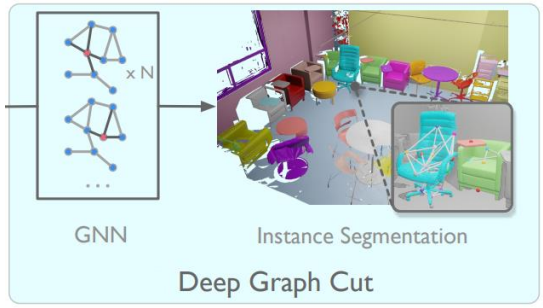
[Peng et al. 2022]



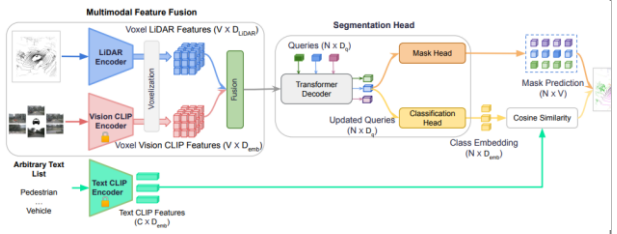
[Ye et al. 2023]



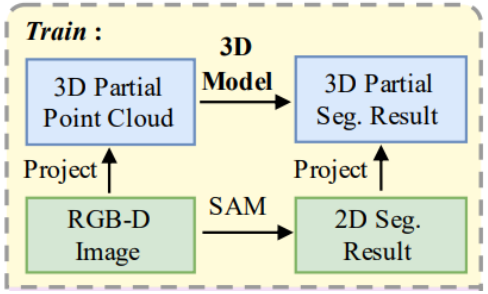
[Wang et al. 2023]



[Guo et al. 2023]



[Xiao et al. 2023]

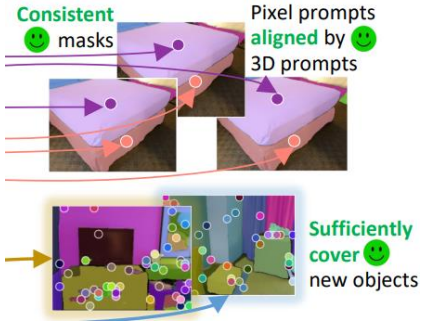


[Huang et al. 2023]

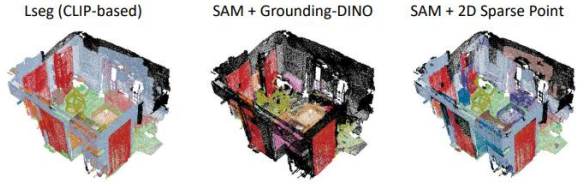
Training-free

SAM3D [Yang et al. 2023]

[Poux 2023]

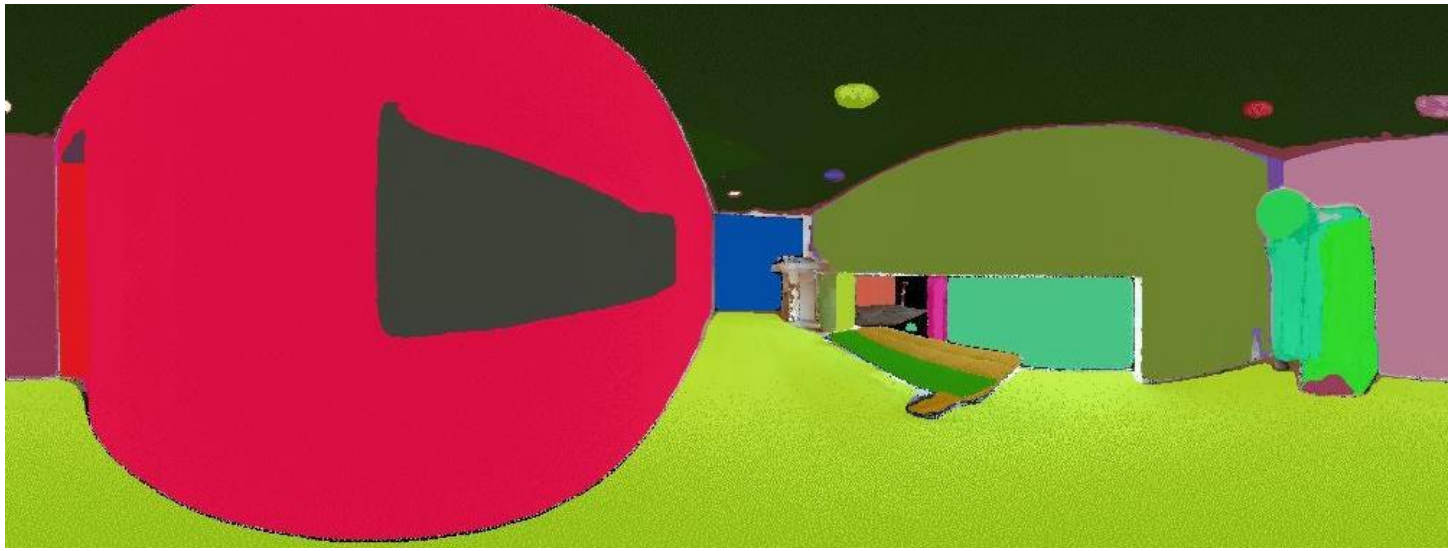


[Xu et al. 2023]



[Dong et al. 2023]

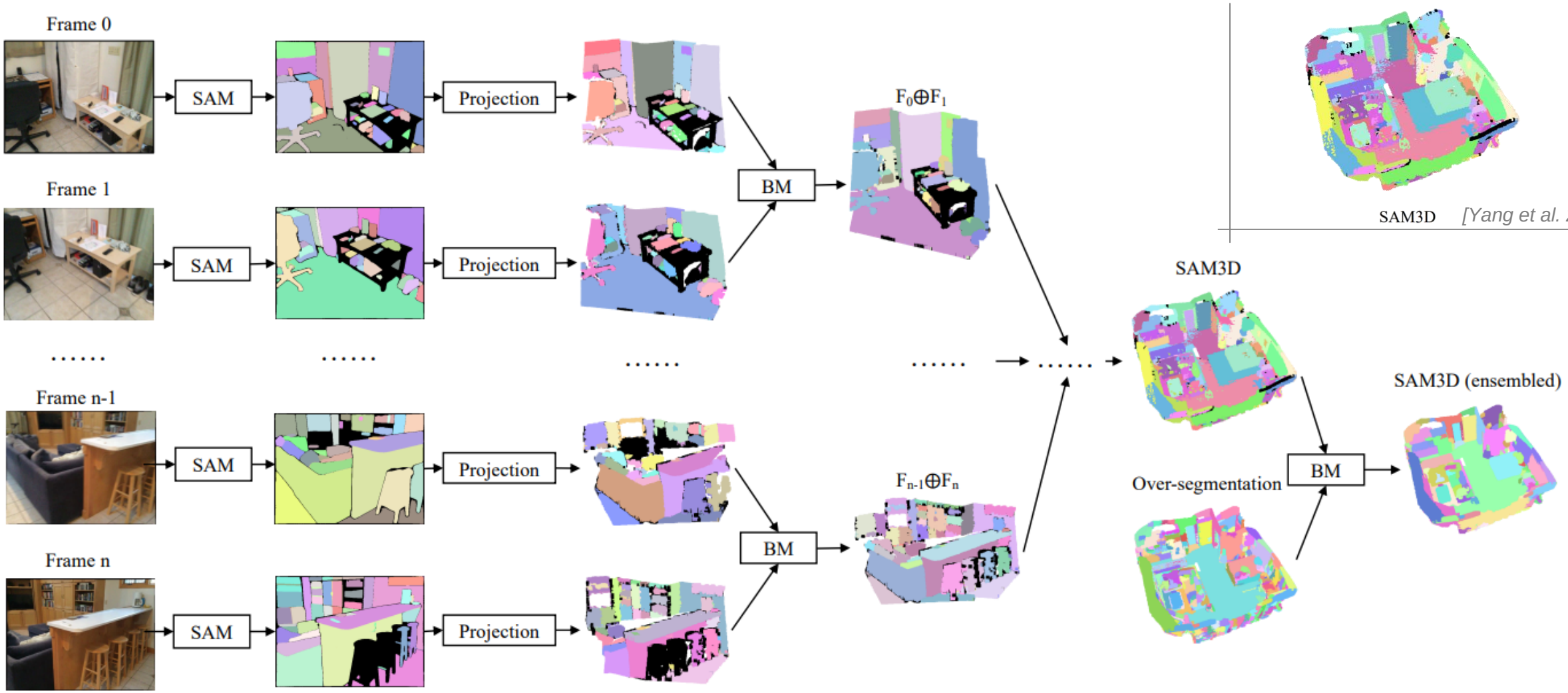
[Yin et al. 2023]



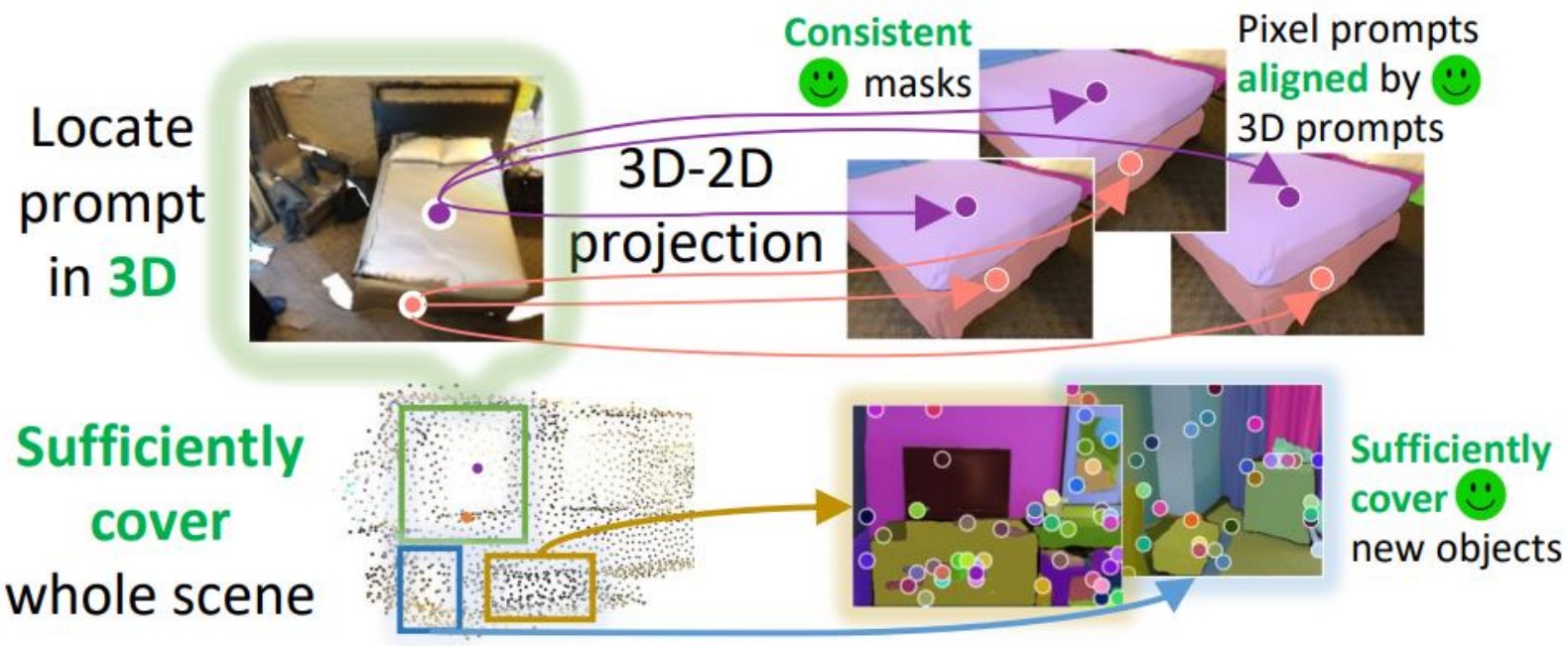
Training-free

[Poux 2023]

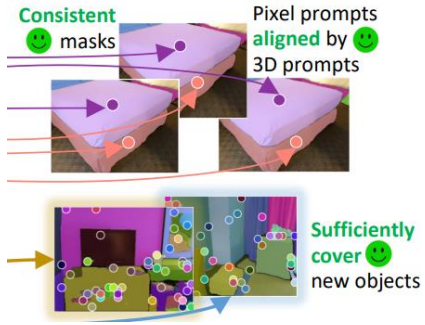
Training-free



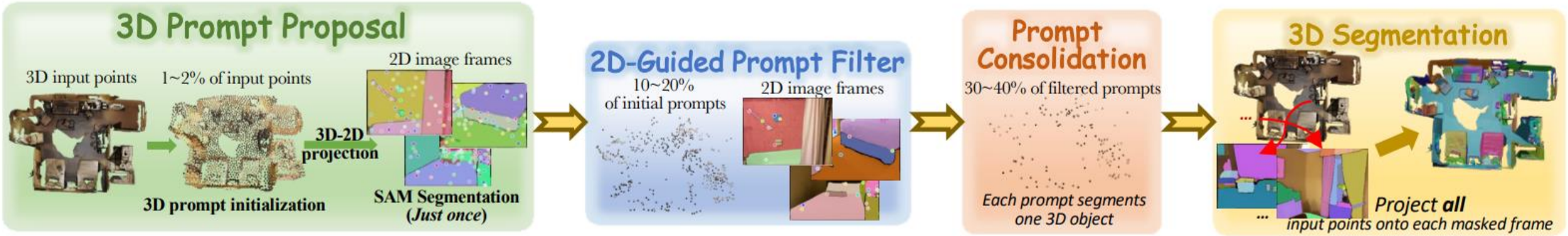
SAM3D *[Yang et al. 2023]*



Training-free

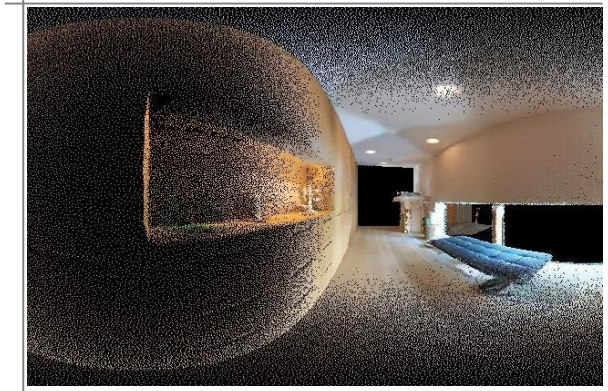


[Xu et al. 2023]



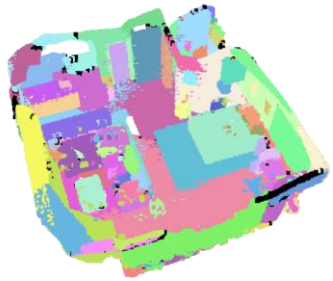
Point Cloud Projection

[Poux 2023]

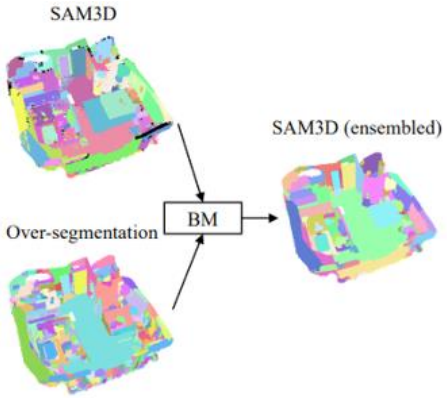


- Niedrige Informationsdichte
- Zusammenführen verschiedener Aufnahmen

SAM3D

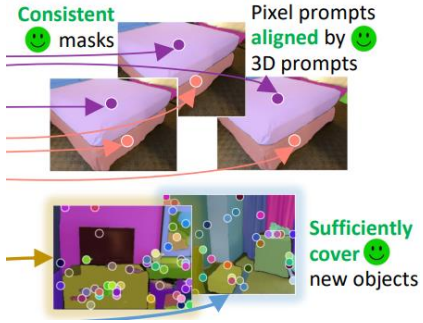


SAM3D *[Yang et al. 2023]*



- Bi-Directional Merging (BM) überleben nur große Masken
- Large-scale efficient

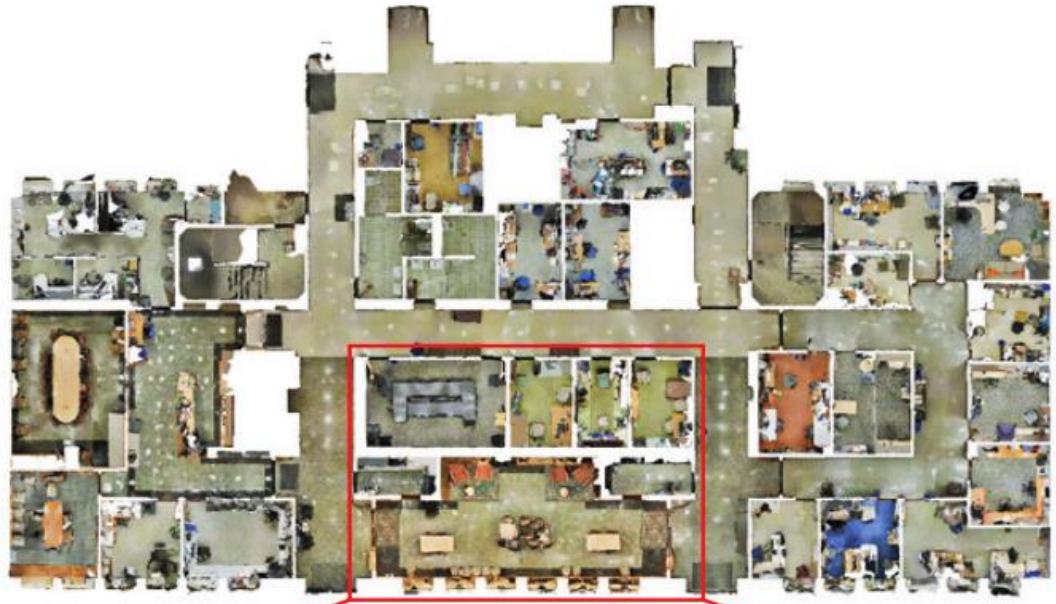
SAMPro3D

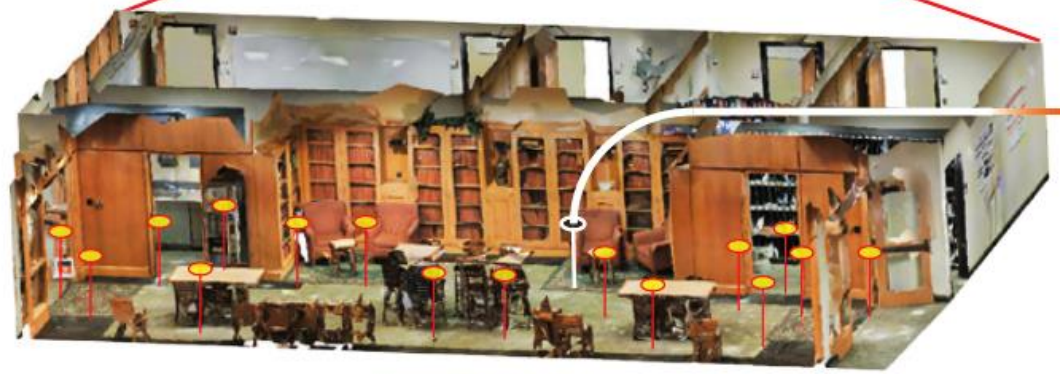


[Xu et al. 2023]

- Hoher Detaillierungsgrad
- Large-scale infeasible

[Armeni et al. 2023]





3D Mesh

RGB

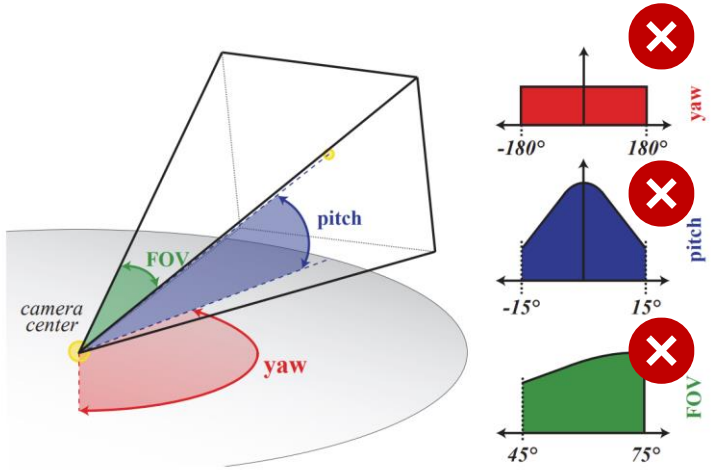
Depth

Surface Normals

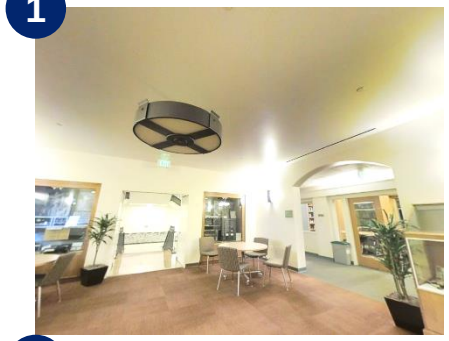
2D Semantics

3D Semantics

floor wall column door table chair board sofa bookcase clutter

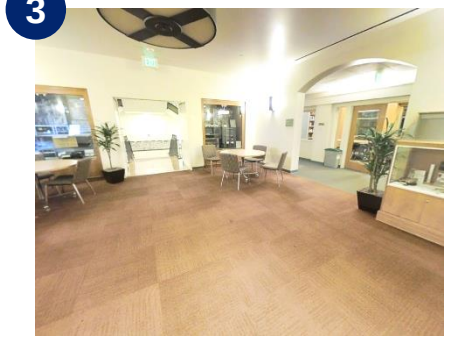


1



2

3

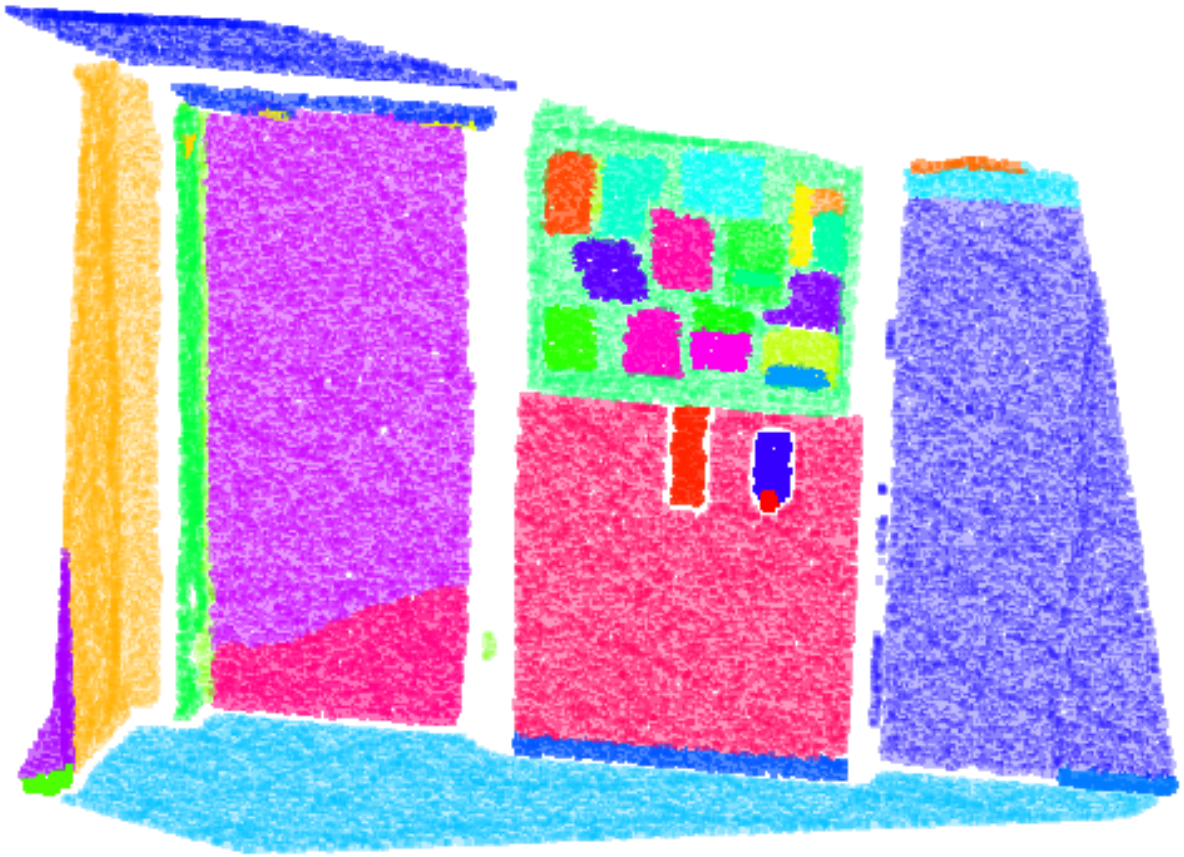


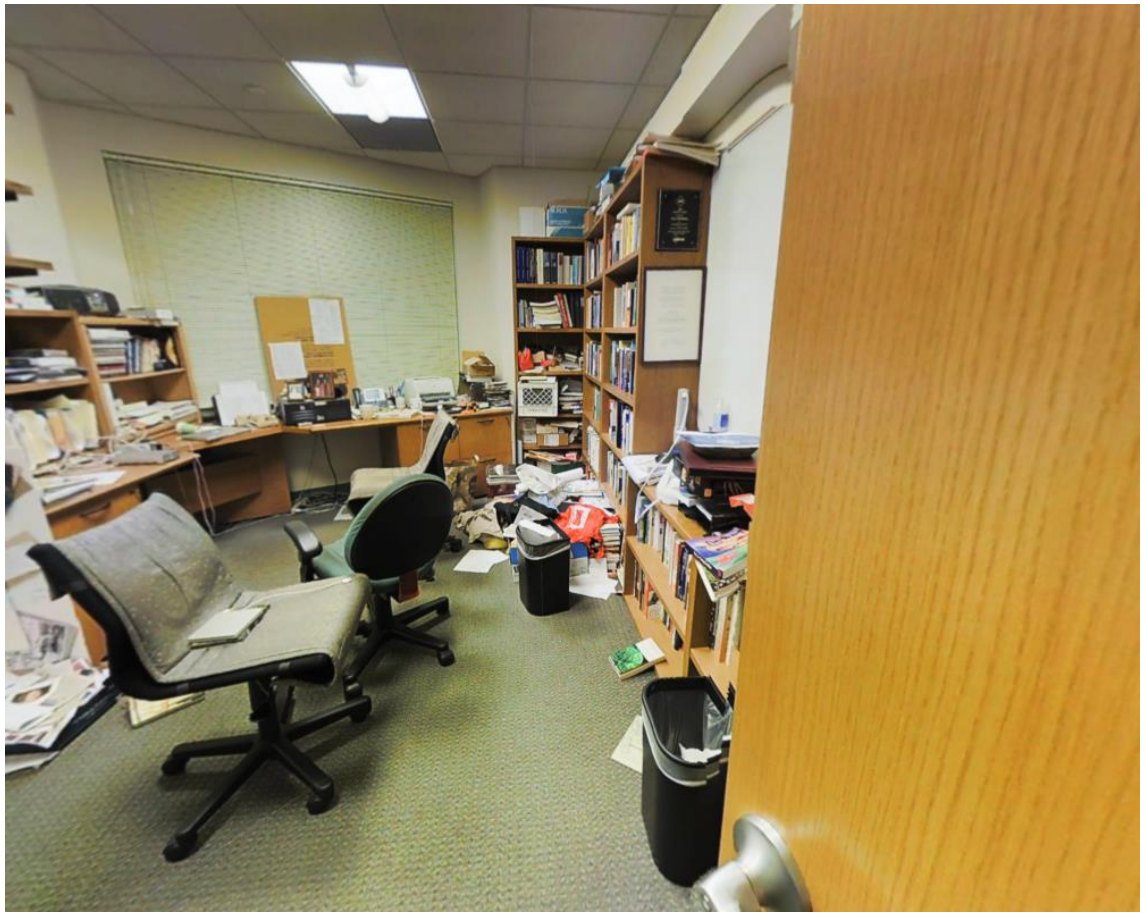
...

...

by *[Armeni et al. 2017]*

by *[Armeni et al. 2017]*





by *[Armeni et al. 2017]*

by *[Armeni et al. 2017]*

x

1

x + 2

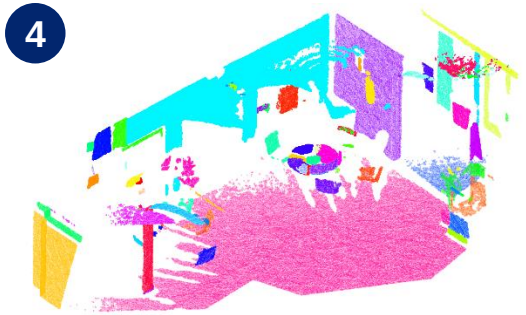
3

x + 1

2

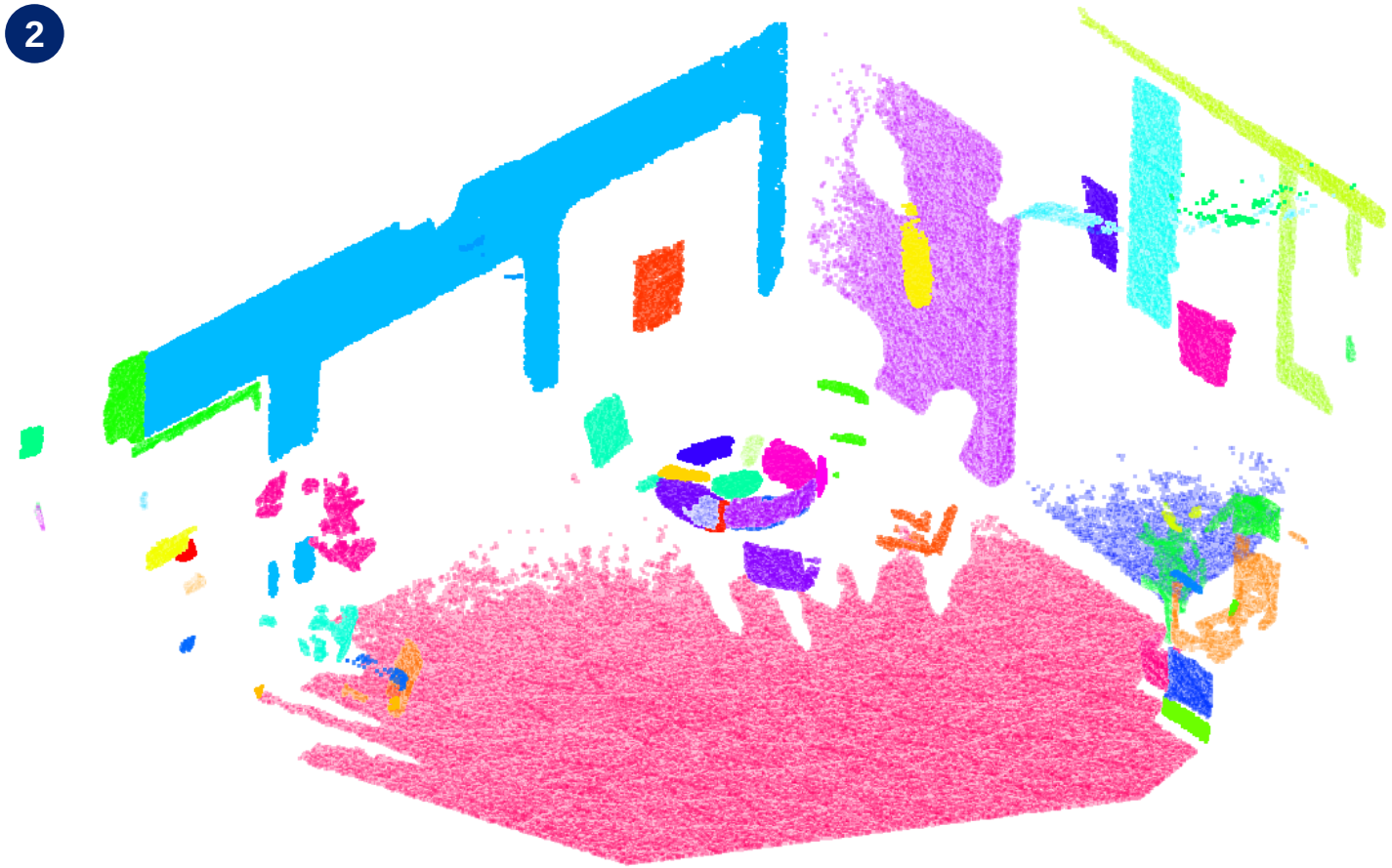
x + 3

4

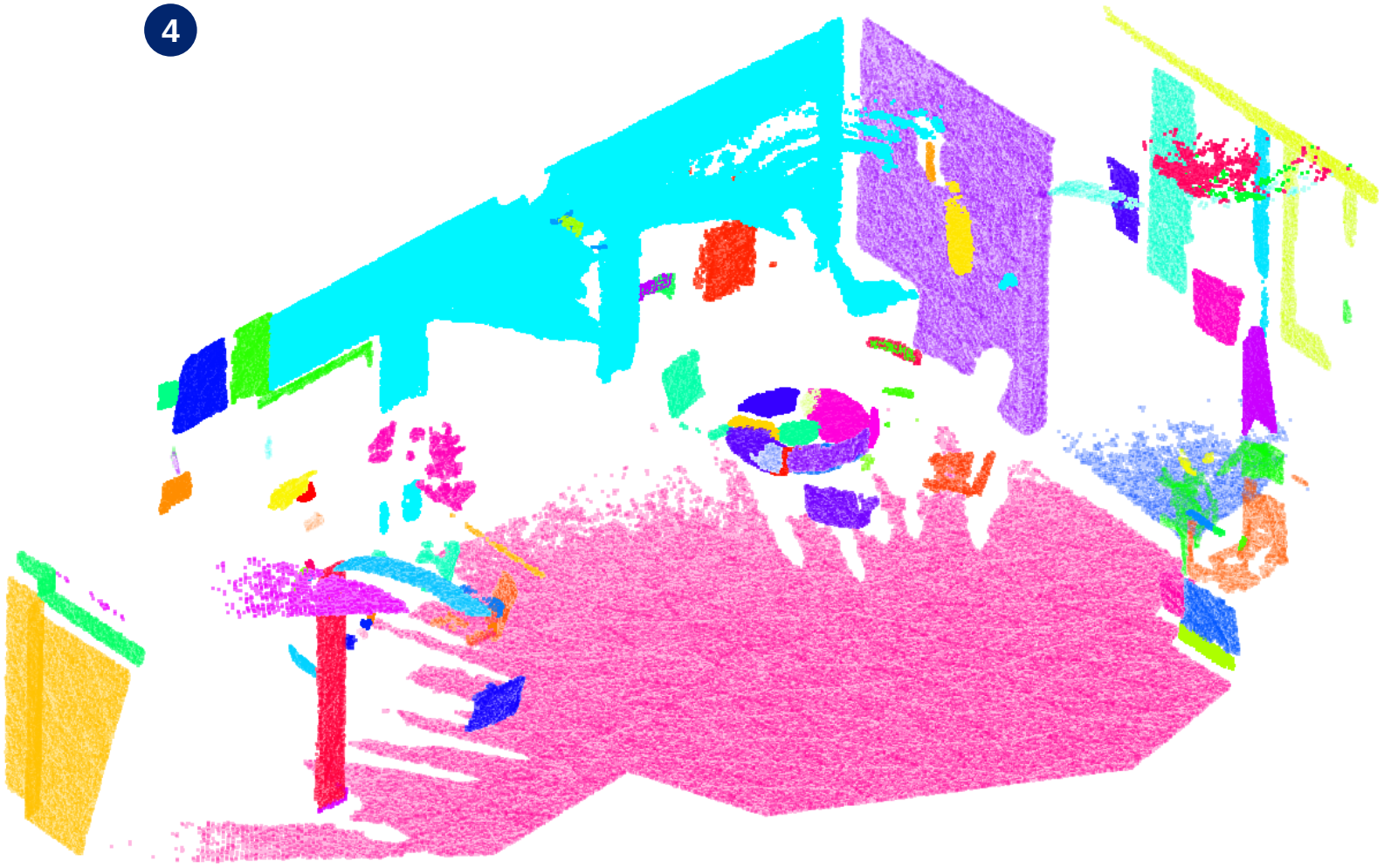


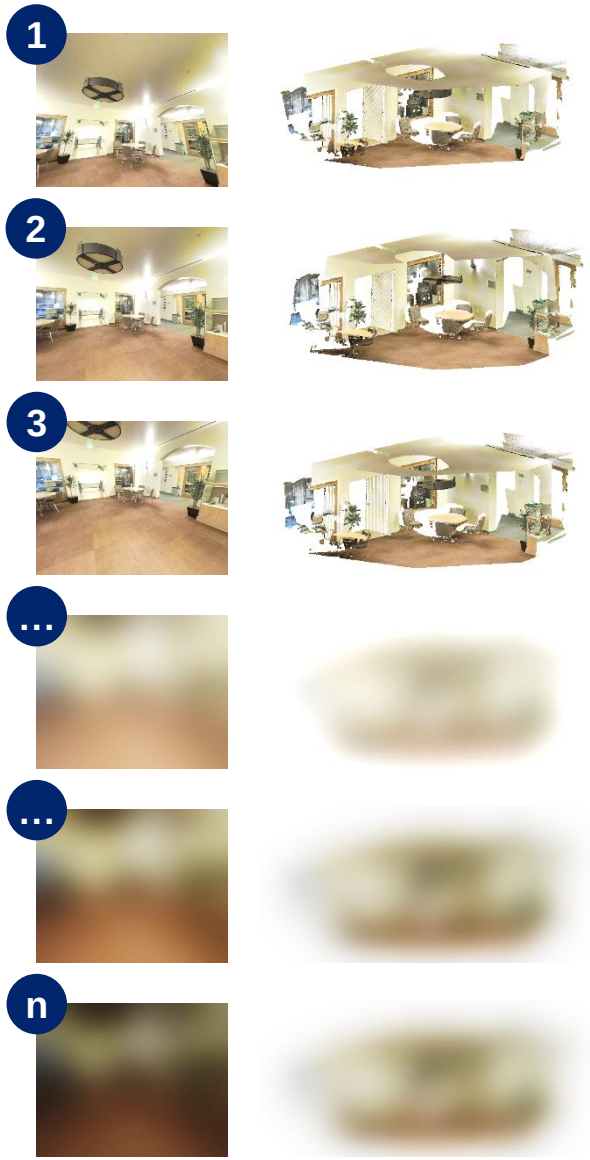
by *[Armeni et al. 2017]*

1


2

3

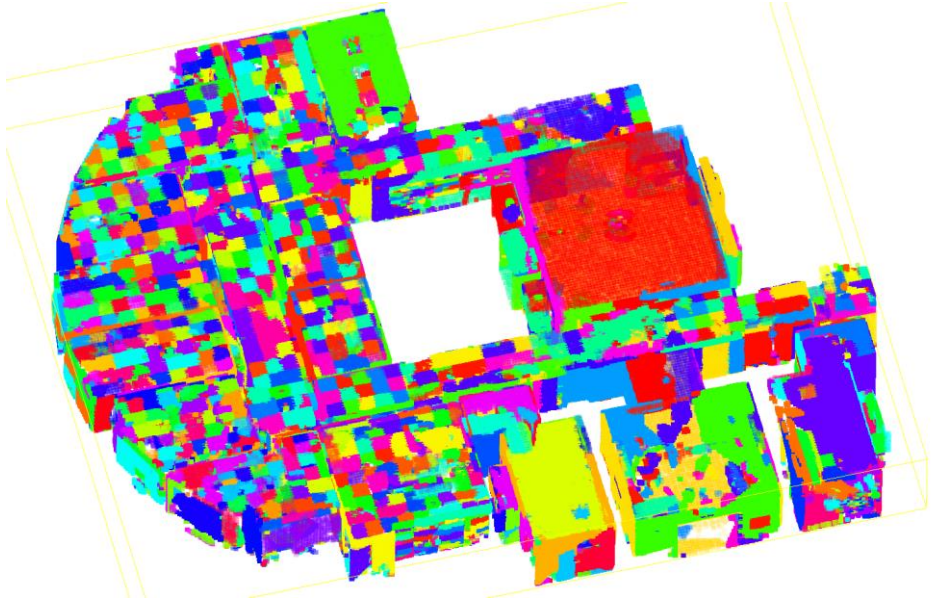

4

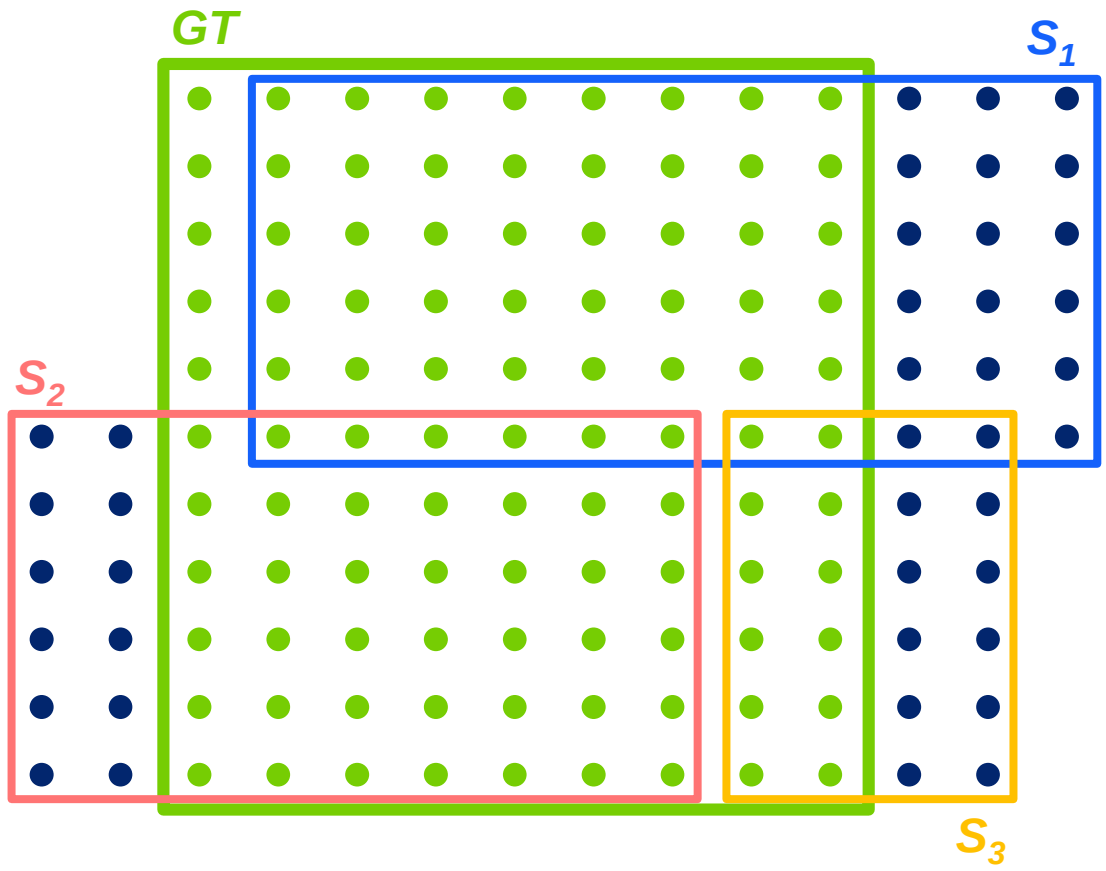


1 Propagating Known Masks

2 Automatic Segmentation

3 Post-Processing



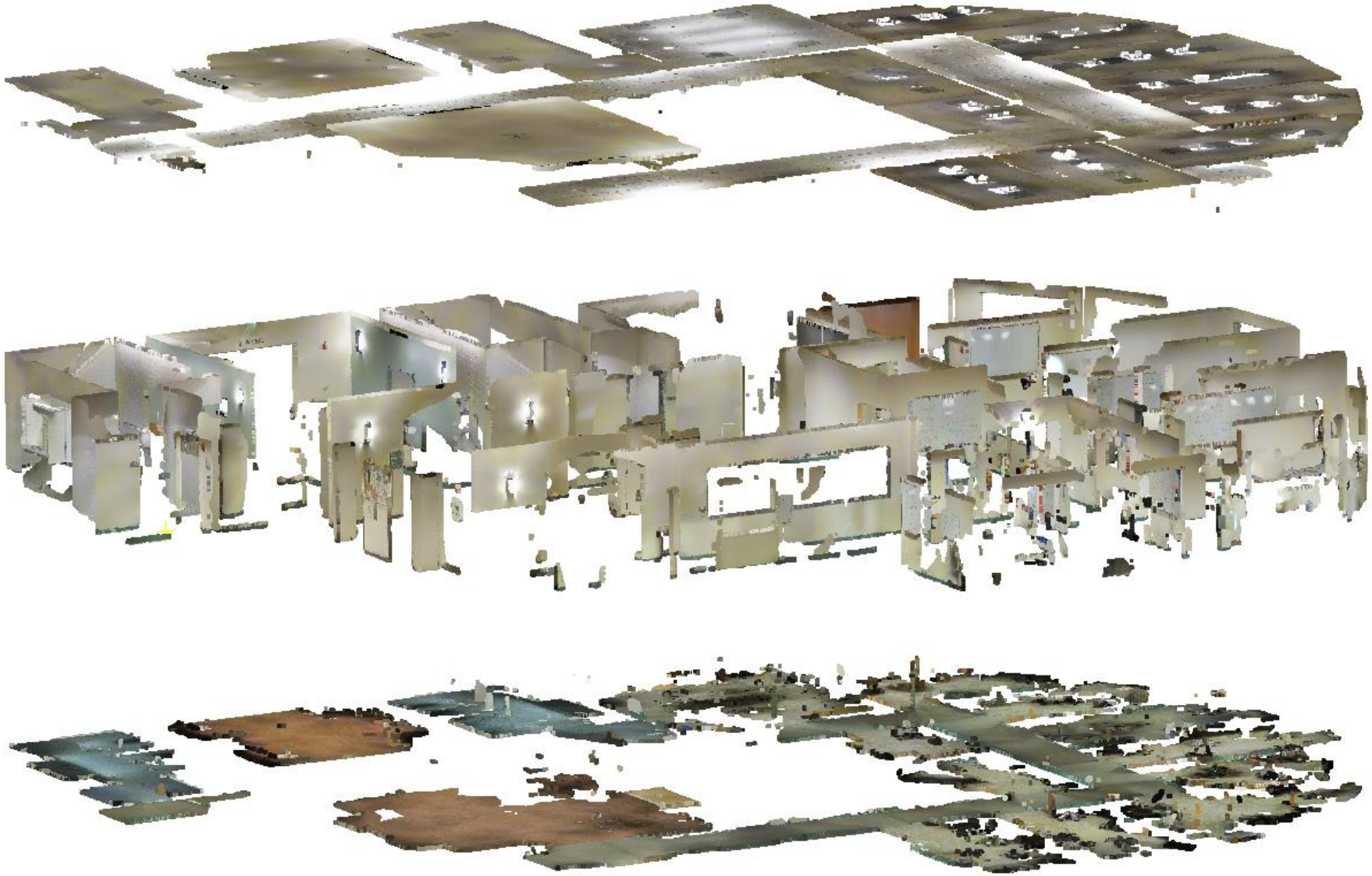


$$\mathbf{S_x} = \mathbf{M}_x, wenn \frac{|\boldsymbol{M_x} \cap \boldsymbol{GT}|}{|\boldsymbol{M_x}|} > \delta ,$$

$$\boldsymbol{S} = \boldsymbol{S_1} \cup \boldsymbol{S_2} \cup \boldsymbol{S_3} \cup \ \ldots ,$$

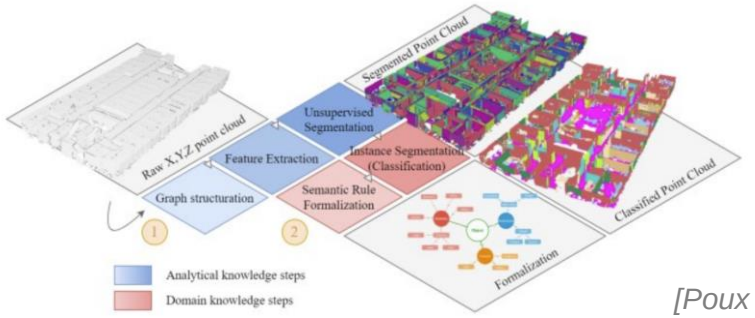
$$\boldsymbol{mIoU_\delta} = \frac{|\boldsymbol{S} \cap \boldsymbol{GT}|}{|\boldsymbol{S} \cup \boldsymbol{GT}|}$$

			Structural Elements							Furniture				
		Mean	*Ceiling*	*Floor*	*Wall*	*Beam*	*Column*	*Windows*	*Door*	*Table*	*Chair*	*Sofa*	*Bookcase*	*Board*
Semantic Segmentation	**PTv3** *Area5* [Wu et al. 2023]	**74,7**												
	PTv3 *6-fold* [Wu et al. 2023]	**80,8**												
	PointNet *Area 5* [Charles et al. 2017]	**41,1**												
	PointNet *6-fold* [Charles et al. 2017]	**47,6**												
	Ours *Area3* $[\delta = 0.7]$	**76,9**	89,3	90,9	74,4	60,0	47,7	88,3	74,7	72,0	68,4	88,5	86,1	85,4
Instance Segmentation	**Ours** *Area3* $[\delta = 0.7]$	**62,1**	76,5	76,3	47,4	57,5	46,2	87,9	56,0	68,2	63,3	87,6	78,8	79,7



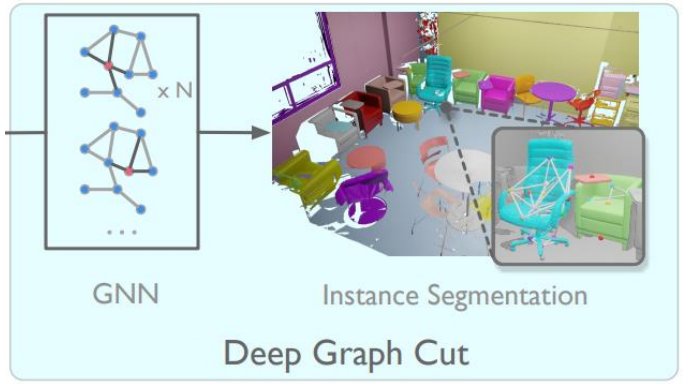
Prozesskettenschritt

- **Unsupervised:** Nutzung der klassen-agnostischen Features



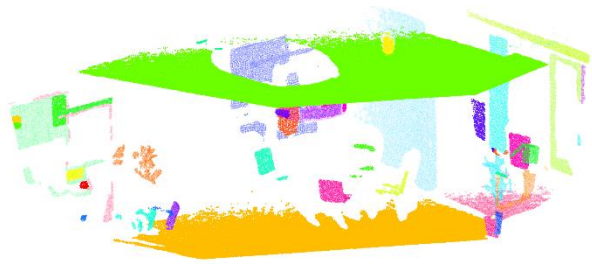
[Poux et al. 2020]

- **Supervised:** “Superpoints”



[Guo et al. 2023]

Anwendungen



Hamburg University of Technology
Institute of Aircraft Production Technology

Keno Moenck
keno.moenck@tuhh.de

TUHH
Hamburg
University of
Technology

Supported by:

Federal Ministry
for Economic Affairs
and Climate Action

on the basis of a decision
by the German Bundestag

TUHH

Literaturverzeichnis

[Armeni et al. 2016] Armeni, Iro; Sener, Ozan; Zamir, Amir R.; Jiang, Helen; Brilakis, Ioannis; Fischer, Martin; Savarese, Silvio (2016): **3D Semantic Parsing of Large-Scale Indoor Spaces.** In : 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA, 27.06.2016 - 30.06.2016: IEEE, pp. 1534–1543.

[Charles et al. 2017] Charles, R. Qi; Su, Hao; Kaichun, Mo; Guibas, Leonidas J. (2017): **PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation**. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[Armeni et al. 2017] Armeni, Iro; Sax, Sasha; Zamir, Amir R.; Savarese, Silvio (2017): **Joint 2D-3D-Semantic Data for Indoor Scene Understanding.**

[Laukotka et al. 2019] Laukotka, Fabian; Oltmann, Jan; Krause, Dieter: **A digitized approach to reduce assembly conflicts during aircraft cabin conversions.** In Dieter Krause, Kristin Paetzold, Sandro Wartzack (Eds.): DFX 2019: Proceedings 2019 (DfX).

[Deneke et al. 2019] Deneke, Constantin; Moenck, Keno; Schueppstuhl, Thorsten (2021): **Augmented Reality Based Data Improvement for the Planning of Aircraft Cabin Conversions.** In Association for Computing Machinery, New York, NY, United States (Ed.): 2021 The 8th International Conference on Industrial Engineering and Applications (Europe) (ICIEA). ICIEA 2021-Europe: 2021 The 8th International Conference on Industrial Engineering and Applications.

[Poux et al. 2020] Poux, F.; Ponciano, J. J. (2020): **Self-learning Ontology for Instance Segmentation of 3D Indoor Point Cloud.** In Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLIII-B2-2020, pp. 309–316. DOI: 10.5194/isprs-archives-XLIII-B2-2020-309-2020.

[Radford et al. 2021] Radford, Alec; Kim, Jong Wook; Hallacy, Chris; Ramesh, Aditya; Goh, Gabriel; Agarwal, Sandhini et al. (2021): **Learning Transferable Visual Models From Natural Language Supervision.**

[Bommasani et al. 2021] Bommasani, Rishi; Hudson, Drew A.; Adeli, Ehsan; Altman, Russ; Arora, Simran; Arx, Sydney von et al. (2021): **On the Opportunities and Risks of Foundation Models.**

[Moenck et al. 2022] Moenck, Keno H. W.; Laukotka, Fabian N.; Deneke, Constantin; Schüppstuhl, Thorsten; Krause, Dieter; Nagel, Thorsten J. (2022): **Towards an Intelligent Digital Cabin Twin to Support an Aircraft's Retrofit and Base Maintenance.** In SAE (Ed.): SAE Technical Paper Series. AeroTech, Mar. 15, 2022

[Peng et al. 2022] Peng, Songyou; Genova, Kyle; Jiang, Chiyu "Max"; Tagliasacchi, Andrea; Pollefeys, Marc; Funkhouser, Thomas (2022): **OpenScene: 3D Scene Understanding with Open Vocabularies.**

[Moenck et al. 2023] Moenck, Keno; Wendt, Arne; Prünte, Philipp; Koch, Julian; Sahrhage, Arne; Gierecker, Johann et al. (2023): **Industrial Segment Anything – a Case Study in Aircraft Manufacturing, Intralogistics, Maintenance, Repair, and Overhaul.**

[Kirillov et al. 2023] Kirillov, Alexander; Mintun, Eric; Ravi, Nikhila; Mao, Hanzi; Rolland, Chloe; Gustafson, Laura et al. (2023): **Segment Anything.**

[Yang et al. 2023] Yang, Yunhan; Wu, Xiaoyang; He, Tong; Zhao, Hengshuang; Liu, Xihui (2023): **SAM3D: Segment Anything in 3D Scenes.**

[Xu et al. 2023] Xu, Mutian; Yin, Xingyilang; Qiu, Lingteng; Liu, Yang; Tong, Xin; Han, Xiaoguang (2023): **SAMPro3D: Locating SAM Prompts in 3D for Zero-Shot Scene Segmentation.**

[Awais et al. 2023] Awais, Muhammad; Naseer, Muzammal; Khan, Salman; Anwer, Rao Muhammad; Cholakkal, Hisham; Shah, Mubarak et al. (2023): **Foundational Models Defining a New Era in Vision: A Survey and Outlook.**

[Poux 2023] Ph.d., Florent Poux (2023): **Segment Anything 3D for Point Clouds: Complete Guide** | Towards Data Science. In Towards Data Science, 12/13/2023. Available online at https://towardsdatascience.com/segment-anything-3d-for-point-clouds-complete-guide-sam-3d-80c06be99a18, checked on 1/26/2024.

[Dong et al. 2023] Dong, Shichao; Liu, Fayao; Lin, Guosheng (2023): **Leveraging Large-Scale Pretrained Vision Foundation Models for Label-Efficient 3D Point Cloud Segmentation.**

[Yin et al. 2023] Yin, Yingda; Liu, Yuzheng; Xiao, Yang; Cohen-Or, Daniel; Huang, Jingwei; Chen, Baoquan (2023): **SAI3D: Segment Any Instance in 3D Scenes.**

[Ye et al. 2023] Ye, Mingqiao; Danelljan, Martin; Yu, Fisher; Ke, Lei (2023): **Gaussian Grouping: Segment and Edit Anything in 3D Scenes.**

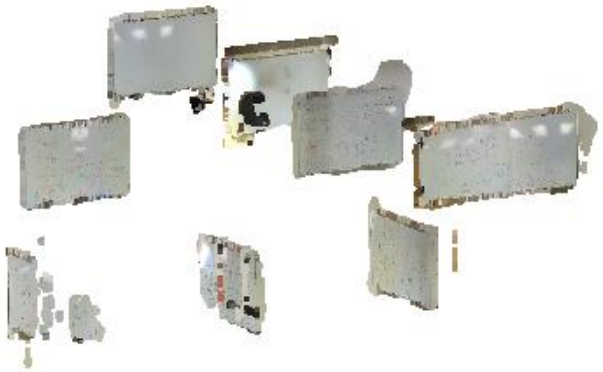
[Wang et al. 2023] Wang, Yuanbin; Huang, Shaofei; Gao, Yulu; Wang, Zhen; Wang, Rui; Sheng, Kehua et al. (2023): **Transferring CLIP's Knowledge into Zero-Shot Point Cloud Semantic Segmentation.**

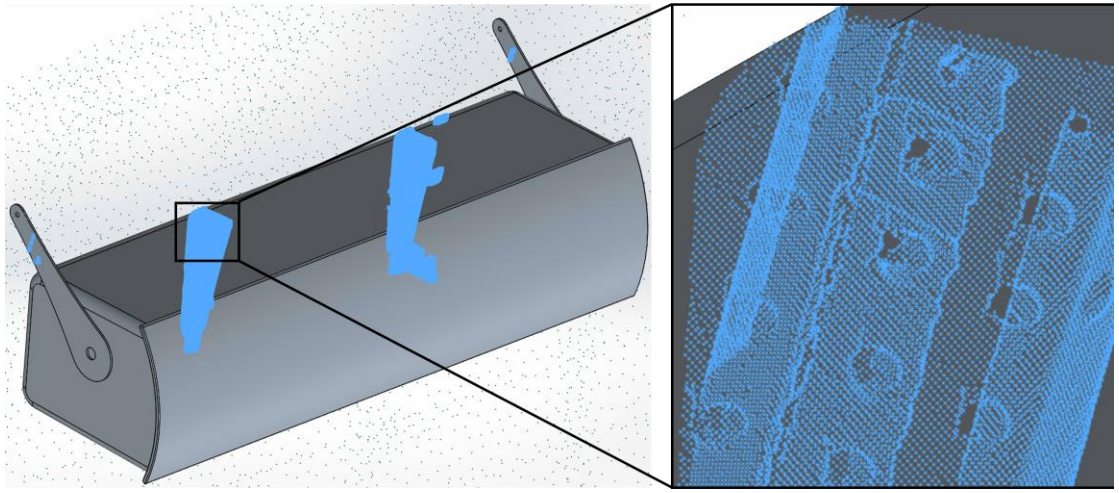
[Guo et al. 2023] Guo, Haoyu; Zhu, He; Peng, Sida; Wang, Yuang; Shen, Yujun; Hu, Ruizhen; Zhou, Xiaowei (2023): **SAM-guided Graph Cut for 3D Instance Segmentation.**

[Xiao et al. 2023] Xiao, Zihao; Jing, Longlong; Wu, Shangxuan; Zhu, Alex Zihao; Ji, Jingwei; Jiang, Chiyu Max et al. (2024): **3D Open-Vocabulary Panoptic Segmentation with 2D-3D Vision-Language Distillation.**

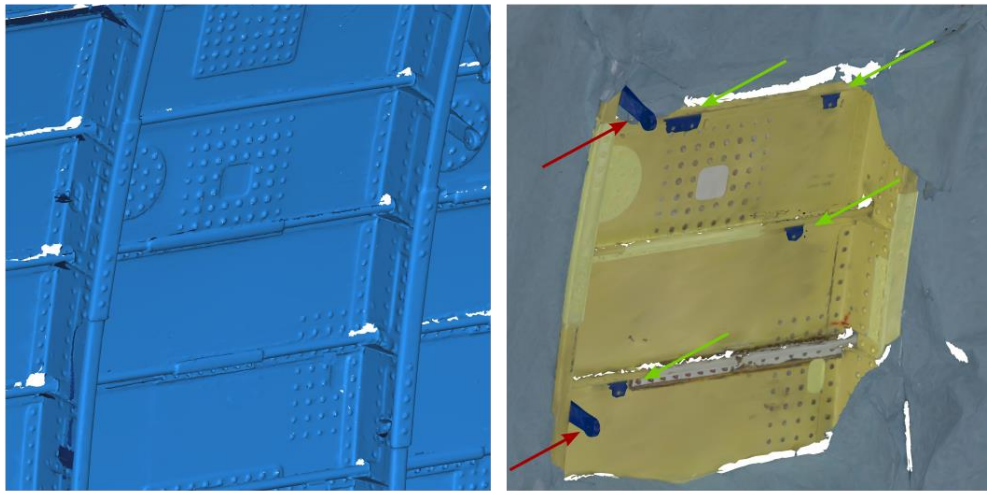
[Huang et al. 2023] Huang, Rui; Peng, Songyou; Takmaz, Ayca; Tombari, Federico; Pollefeys, Marc; Song, Shiji et al. (2023): **Segment3D: Learning Fine-Grained Class-Agnostic 3D Segmentation without Manual Labels.**

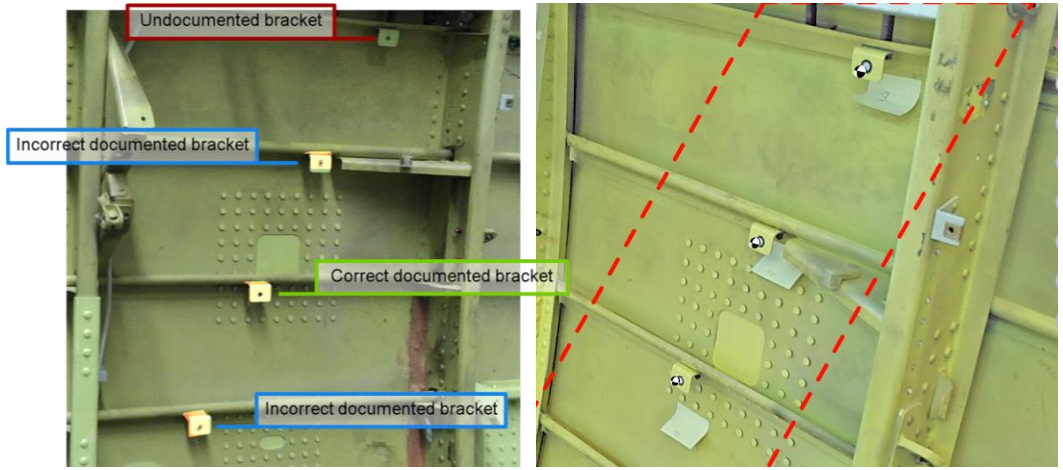
[Wu et al. 2023] Wu, Xiaoyang; Jiang, Li; Wang, Peng-Shuai; Liu, Zhijian; Liu, Xihui; Qiao, Yu et al. (2023): **Point Transformer V3: Simpler, Faster, Stronger.**





[Laukotka et al. 2019]





[Deneke et al. 2019]

[Moenck et al. 2022]