
ZAHA

UNIVERSITY OF CAMBRIDGE

CV4DT

CSIC Cambridge Centre for Smart Infrastructure & Construction

LAING O'ROURKE CENTRE for CONSTRUCTION ENGINEERING and TECHNOLOGY

Computer Vision for Digital Twins (CV4DT), where perception meets geometry

Dr Olaf Wysocki
Assistant Research Professor

CV4DT
Department of Engineering
University of Cambridge

About me

- **Assistant Research Professor (08/2025 – now)**

  @ University of Cambridge, leading the CV4DT group

- **PostDoc + PhD (09/2020 – 08/2025)**

  @ Technical University of Munich

  @ Photogrammetry & Remote Sensing, Prof Uwe Stilla

  3D Semantic Understanding / 3D Object Reconstruction

- **Previous stints in industry (e.g., Audi)**

CSIC Cambridge Centre for Smart Infrastructure & Construction

LAING O'ROURKE CENTRE for CONSTRUCTION ENGINEERING and TECHNOLOGY

CV4DT

UNIVERSITY OF CAMBRIDGE

UNIVERSITY OF CAMBRIDGE | CV4DT | CSIC Cambridge Centre for Smart Infrastructure & Construction | LAING O'ROURKE CENTRE for CONSTRUCTION ENGINEERING and TECHNOLOGY

Researchers

Dr Yixiong Jing

Postdoctoral Research Associate

Dr Brian Sheil

Laing O'Rourke Associate Professor in Construction Engineering

Dr Guangming Wang

Postdoctoral Research Associate

Haibing Wu

PhD Candidate

Dr Olaf Wysocki

Group Lead, Assistant Research Professor

Wanru Yang

PhD Candidate

Visitors

Daniel Lehmberg

Intern

Ziyang Xu

PhD Candidate

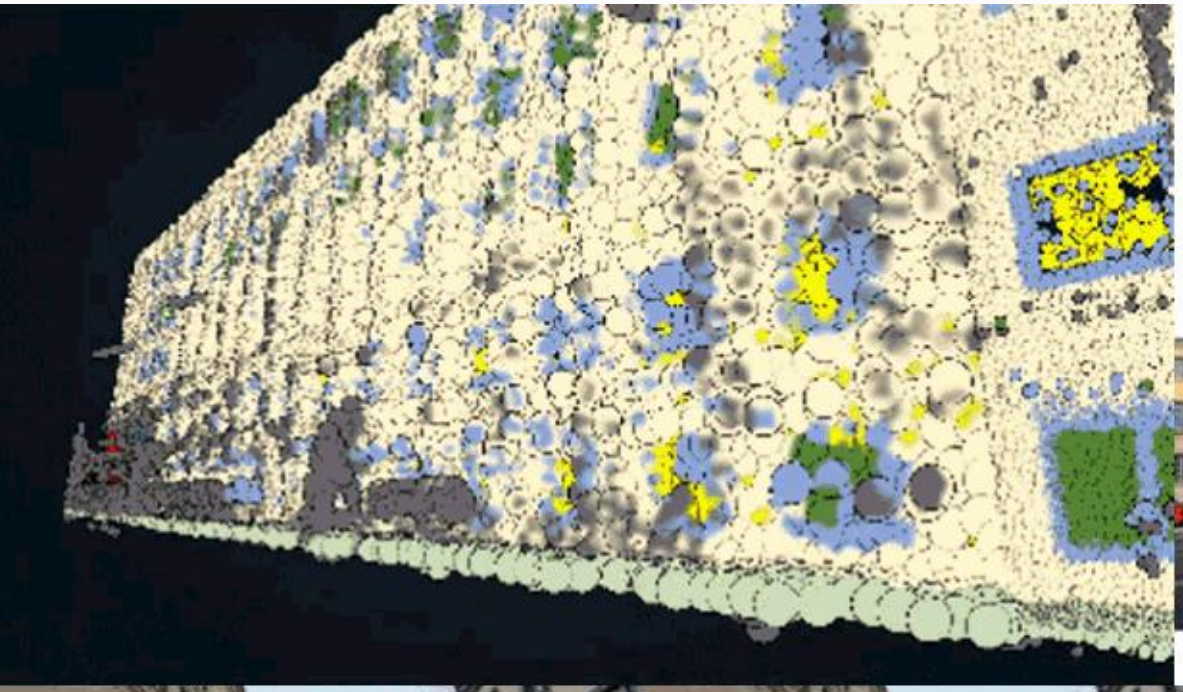
Qizhen Ying

PhD Candidate

Qilin Zhang

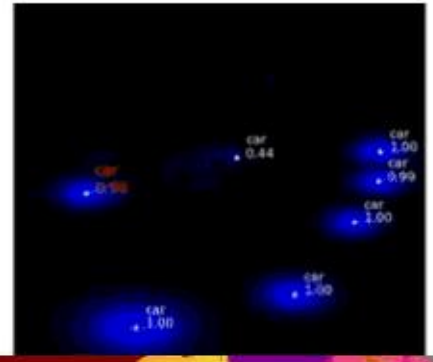
PhD Candidate

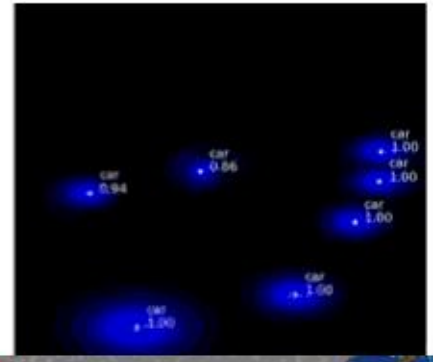
Perception is all you need

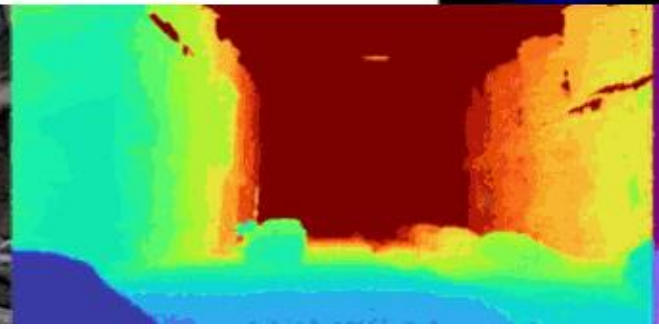


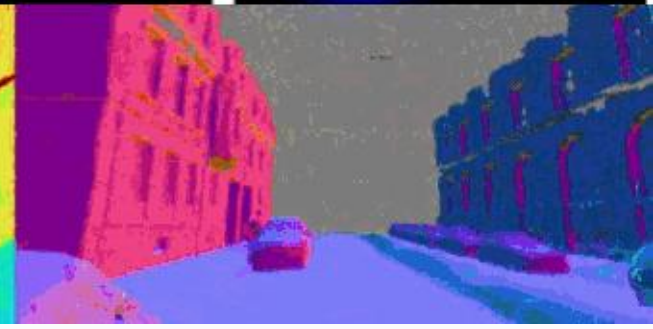
Before (LoD2)

After (LoD3)



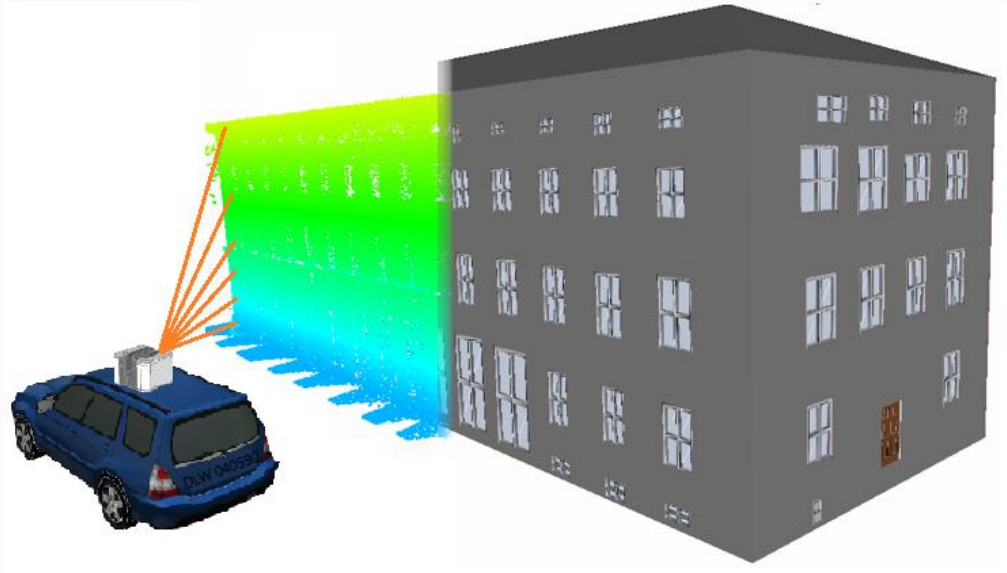


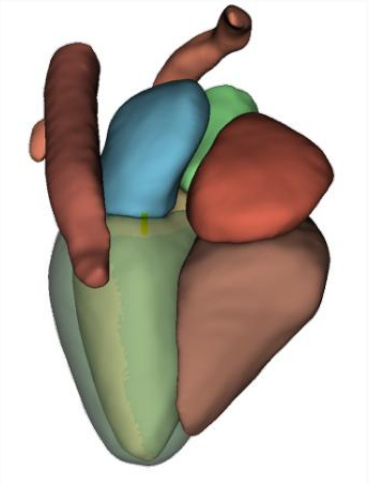


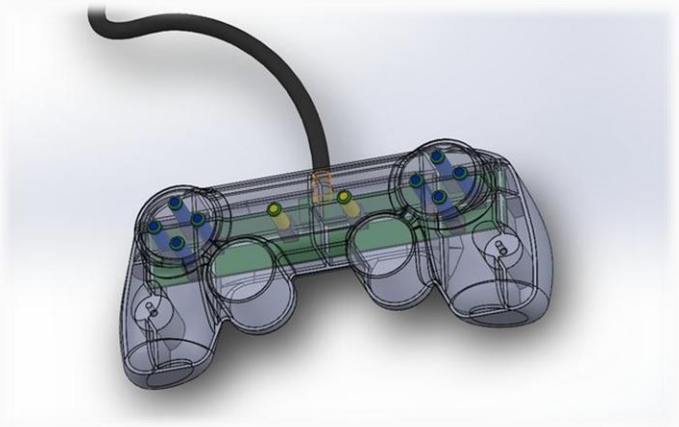


Motivation

- Digital Twins (DTs) necessary, e.g.,:
  - Gaming
  - Medical
  - Garment
  - Up to country-wide city models
- But DTs require:
  - 3D minimum-viable representation (think CAD)
  - Semantics
  - Accuracy & Completness
- 3D Computer Vision so-far:
  - Small-scale and…
  - Toy examples





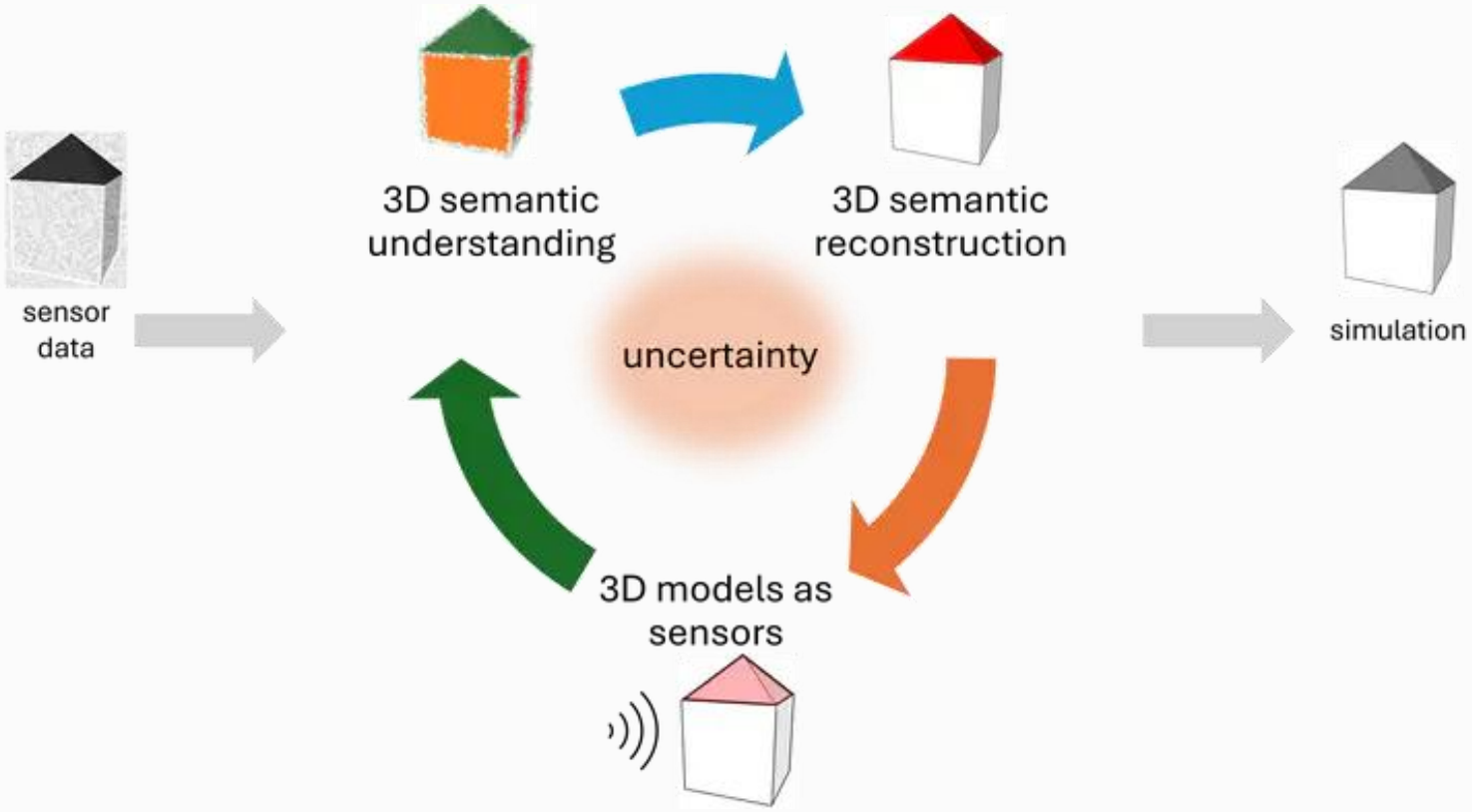


Wysocki O., Xia Y., Wysocki M., Grilli E., Hoegner L., Cremers D., Stilla U. *Scan2LoD3: Reconstructing Semantic 3D Building Models at LoD3 Using Ray Casting and Bayesian Networks,* ***CVPR '23***
Yang J., Sedykh E., Adhinarta JK., Le H., Fua P. *Generating Anatomically Accurate Heart Structures via Neural Implicit Fields* ***MICCAI 2024***
https://grabcad.com/library/gamepad-11 , GrabCAD 2025

CV4DT Vision

Focal points:

- 3D semantic understanding,
- 3D semantic reconstruction,
- 3D models as sensors,
- Uncertainty quantification – overarching all three aspects.



3D Computer Vision for Built Environment

Can we go beyond the toy examples?

a plush dragon toy

3D model (mesh)

Construction site

3D model (BIM)

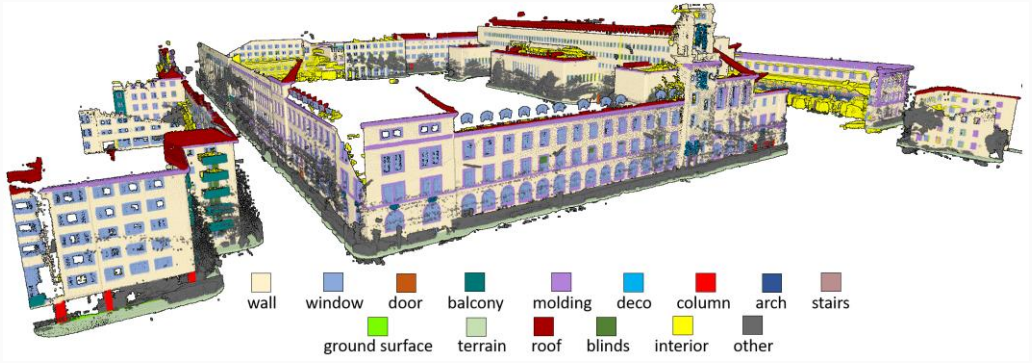
One-2-3-45++: Fast Single Image to 3D Objects with Consistent Multi-View Generation and 3D Diffusion, CVPR 2024

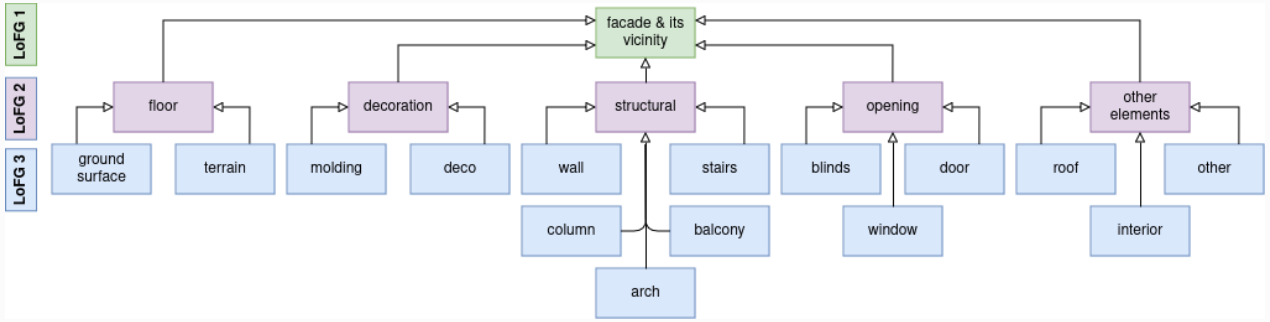
Benchmark Datasets: ZAHA

ZAHA: The largest facade segmentation dataset

Level of Facade Generalization (LoFG)

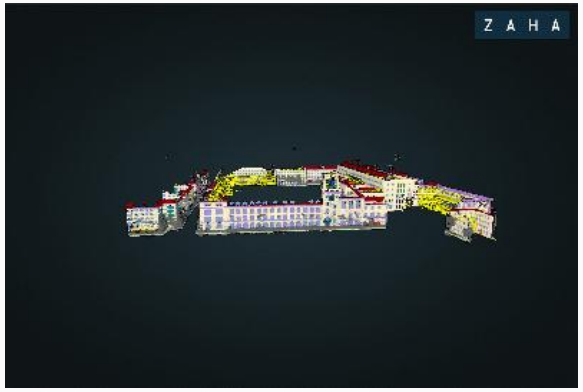
- 15 façade classes
- 3 levels of generalization
- Over 600 mln points

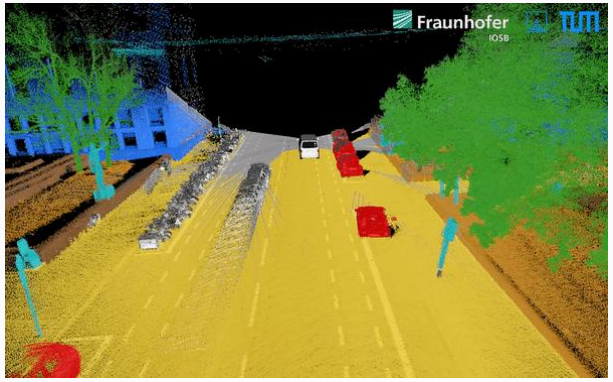


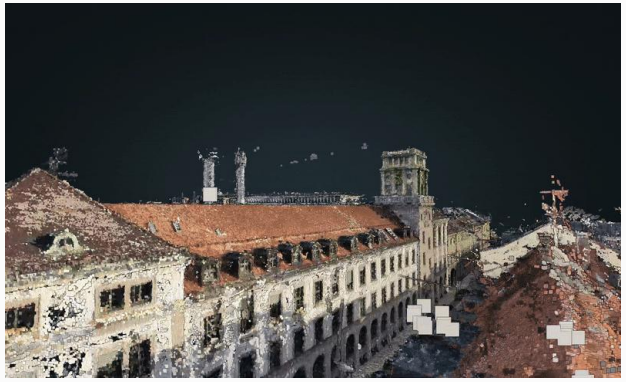


Wysocki O., Tan Y., Froech T., Xia Y., Wysocki M., Hoegner L., Cremers D., Holst Ch. *ZAHA: Introducing the Level of Facade Generalization and the Large-Scale Point Cloud Facade Semantic Segmentation Benchmark Dataset,* ***WACV '25***

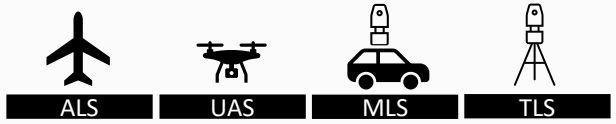
Benchmark Datasets: TUM2TWIN https://tum2t.win



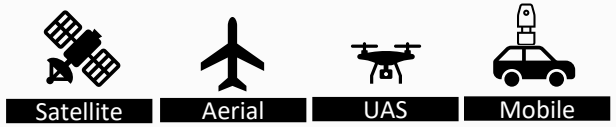




Point clouds



Images



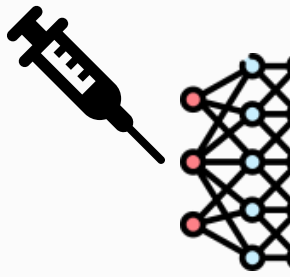
3D models

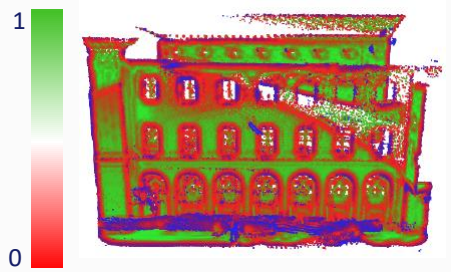
Wysocki O., Schwab B., et al. *TUM2TWIN: Introducing the Large-Scale Multimodal Urban Digital Twin Benchmark Dataset, Submitted to* ***ISPRS Journal of Photogrammetry and Remote Sensing***

3D Semantic Segmentation

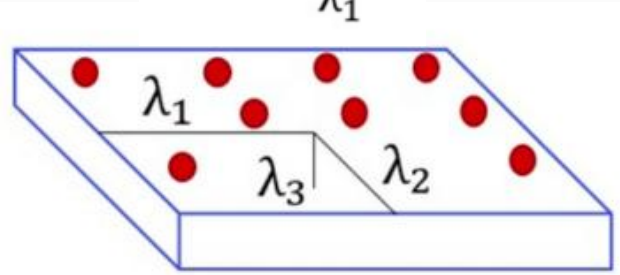
EARLy: How effectively can we use geometry for segmentation?

- Fusion of geometric features + learned features
  - EARLy + PointNet: ~81% (+9%)
  - EARLy + PointNet++: ~81% (+6%)
  - EARLy + Point Transformer (PT): ~88% (+10%)
- Faster convergence (2x Faster)
- Less compute needed (2x Less Parameters)

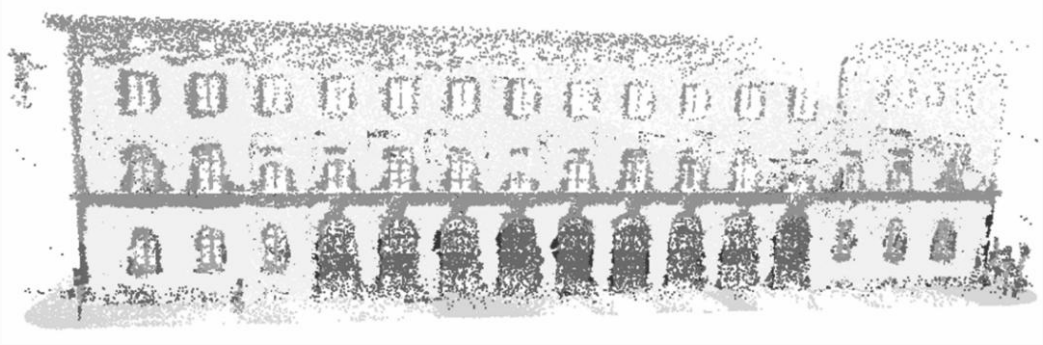




$$P_\lambda = \frac{\lambda_2 - \lambda_3}{\lambda_1}$$



Encoding planarity



Input - point cloud

Output - segmented point cloud

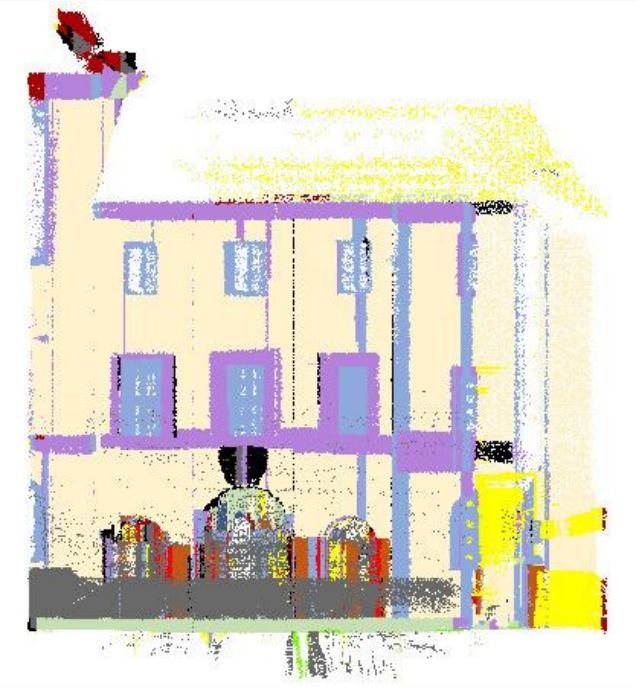
Tan, Y., Wysocki, O., Hoegner, L., and Stilla, U. Classifying point clouds at the facade-level using geometric features and deep learning networks, ***3D GeoInfo '23***

3D Semantic Segmentation

EARLy: How effectively can we use geometric prior?

Example of façade segmentation results

Before our EARLy



After our EARLy

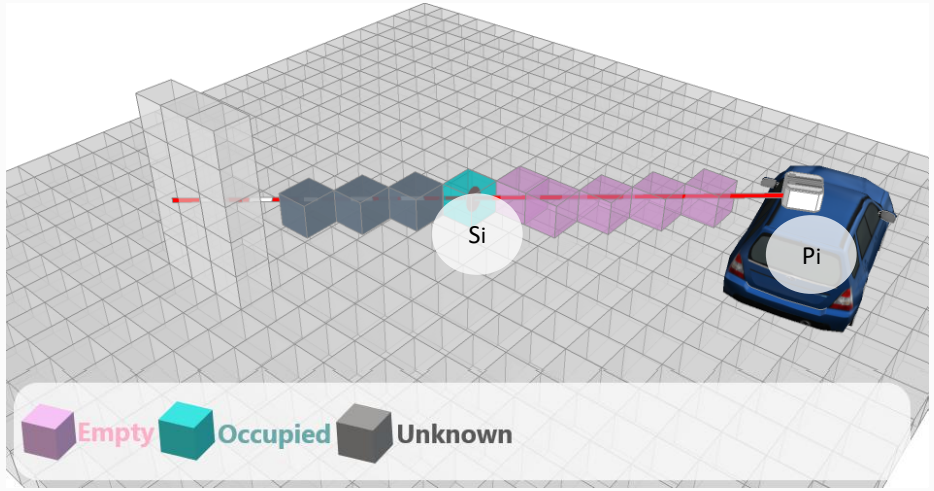
3D Semantic Segmentation & Reconstruction

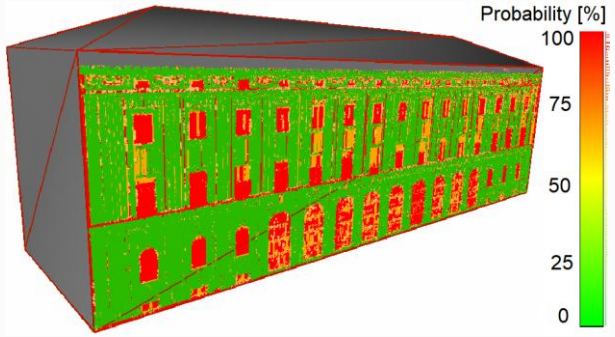
Laser scanner rays penetrate transparent objects ->

no point cloud!

Can we use it to our **advantage**?

- Probabilistic ray tracing of points -> *Empty*, *Occupied*, *Unknown* voxels
- Comparison to 3D building model -> *Confirmed*, *Conflicted* voxels
- *Conflicted* voxels -> refinement required





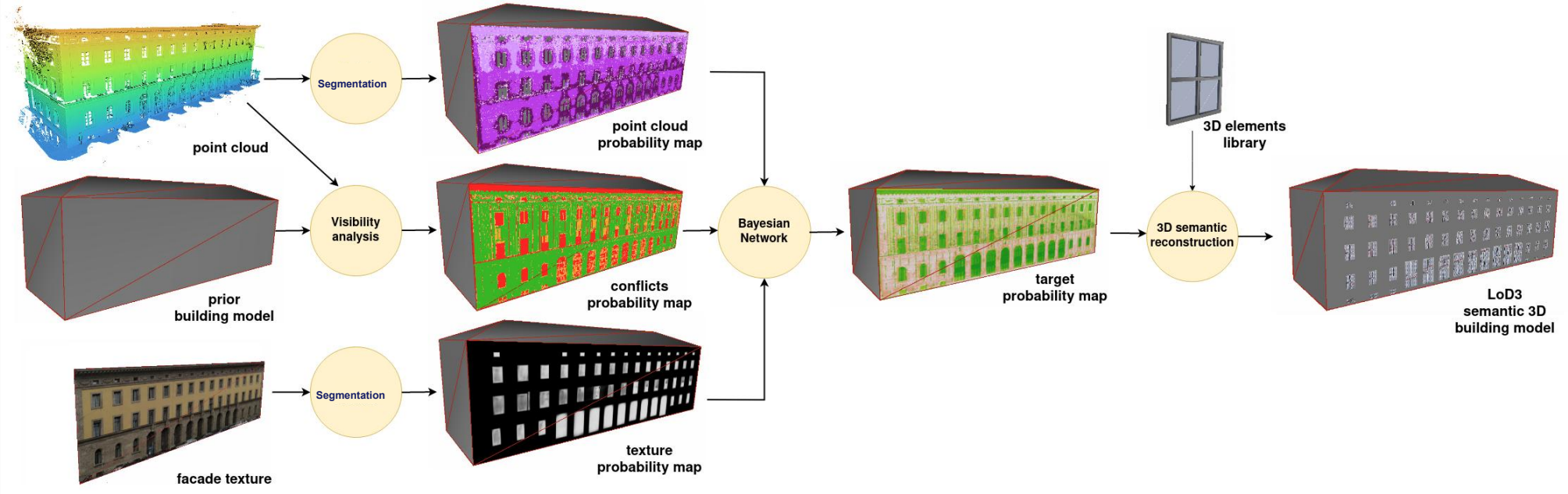
Wysocki O., Grilli E., Hoegner L., Stilla U. *Combining Visibility Analysis and Deep Learning for Refinement of Semantic 3D Building Models by Conflict Classification*, ***ISPRS Annals '22***

3D Semantic Segmentation & Reconstruction

How effectively can we use our strong prior for segmentation and reconstruction?

Scan2LoD3: Uncertainty-aware late fusion of semantic maps for more accurate 3D reconstruction

- Bayesian Network calculates target based on uncertain input
- Output -> Uncertainty-aware target map and reconstruction instances



Wysocki O., Xia Y., Wysocki M., Grilli E., Hoegner L., Cremers D., Stilla U. *Scan2LoD3: Reconstructing Semantic 3D Building Models at LoD3 Using Ray Casting and Bayesian Networks,* ***CVPR '23***

3D Semantic Segmentation & Reconstruction

Scan2LoD3:

Conflict Map + Point Cloud Map:

- Openings IoU: ~61% (+54% / +3%)

Conflict Map + Point Cloud Map + Texture Map:

- Openings IoU: ~76% (+59% / +18%)

Point-cloud-only:

- Openings IoU: 7%

Image-only:

- Openings IoU: 58%

windows with blinds

ground-truth *w/o images* *w/ images*

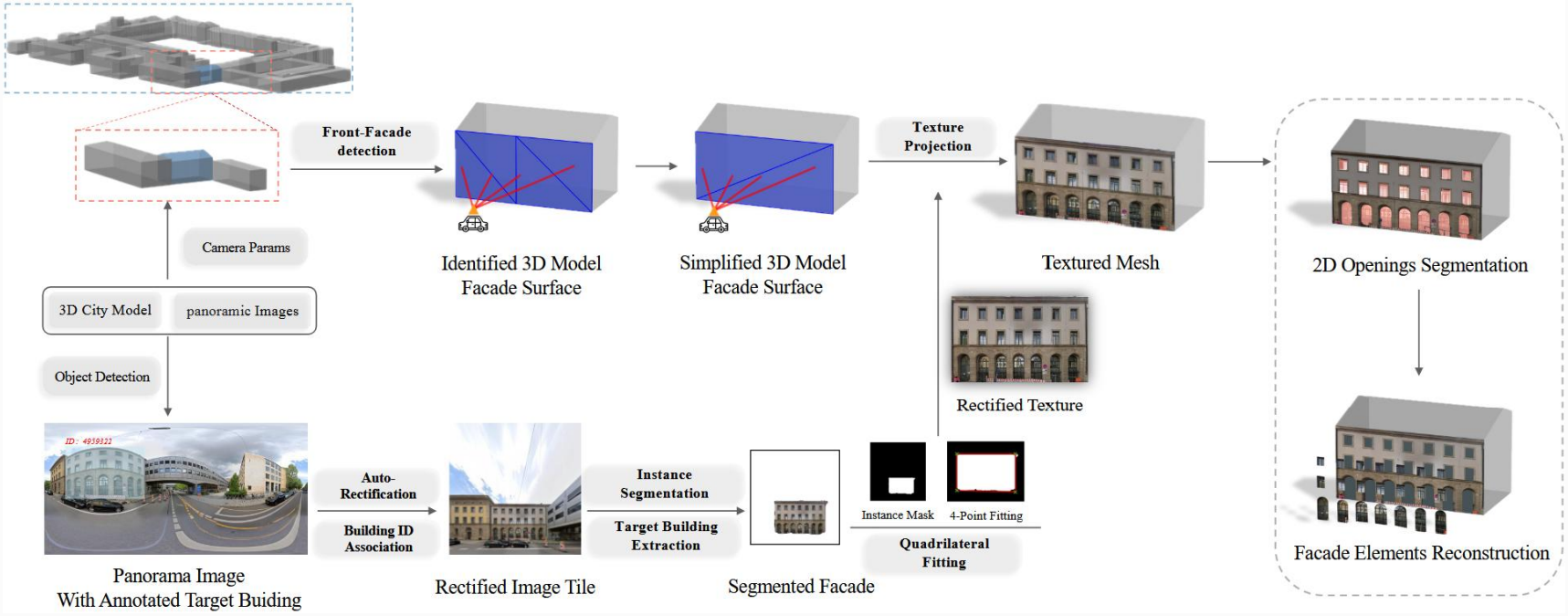
Wysocki O., Xia Y., Wysocki M., Grilli E., Hoegner L., Cremers D., Stilla U. *Scan2LoD3: Reconstructing Semantic 3D Building Models at LoD3 Using Ray Casting and Bayesian Networks, **CVPR '23***

3D Semantic Segmentation & Reconstruction

What happens if we do not have any scanners? Let use panoramic images (think Google Street View)!

Texture2LoD3: More accurate panoramic image projection -> more accurate segmentation & reconstruction

- Building models composed of planes – use them as rectification targets for panoramic images
  - Image2building and transform to quasi-orthorectified image

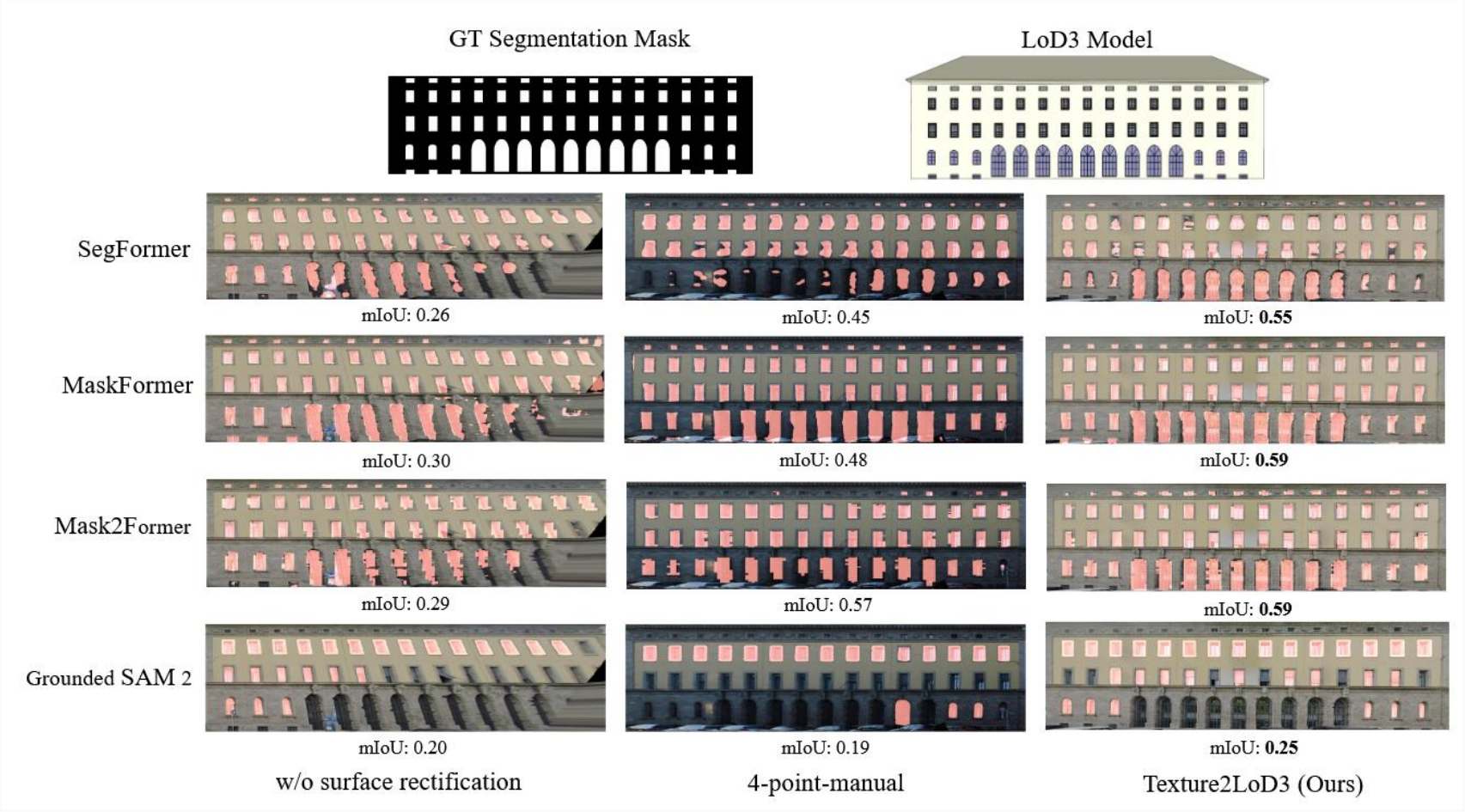


Tang W., Li W., Liang X., Wysocki O., Biljecki F., Holst Ch., Jutzi B. *Texture2LoD3: Enabling LoD3 Building Reconstruction With Panoramic Images*, CVPR '25

3D Semantic Segmentation & Reconstruction

Texture2LoD3:

- +10% IoU vs baselines
- On par with manual



Tang W., Li W., Liang X., Wysocki O., Biljecki F., Holst Ch., Jutzi B. *Texture2LoD3: Enabling LoD3 Building Reconstruction With Panoramic Images*, CVPR '25

3D Semantic Segmentation & Reconstruction

[GS]

Gaussian Splatting reconstruction yielding appealing visualization

[GS]

Gaussian Splatting reconstruction in the practical, built environment scenario

3D Models as Sensors & Reconstruction

Gaussian Splatting (GS) – State-of-the-art

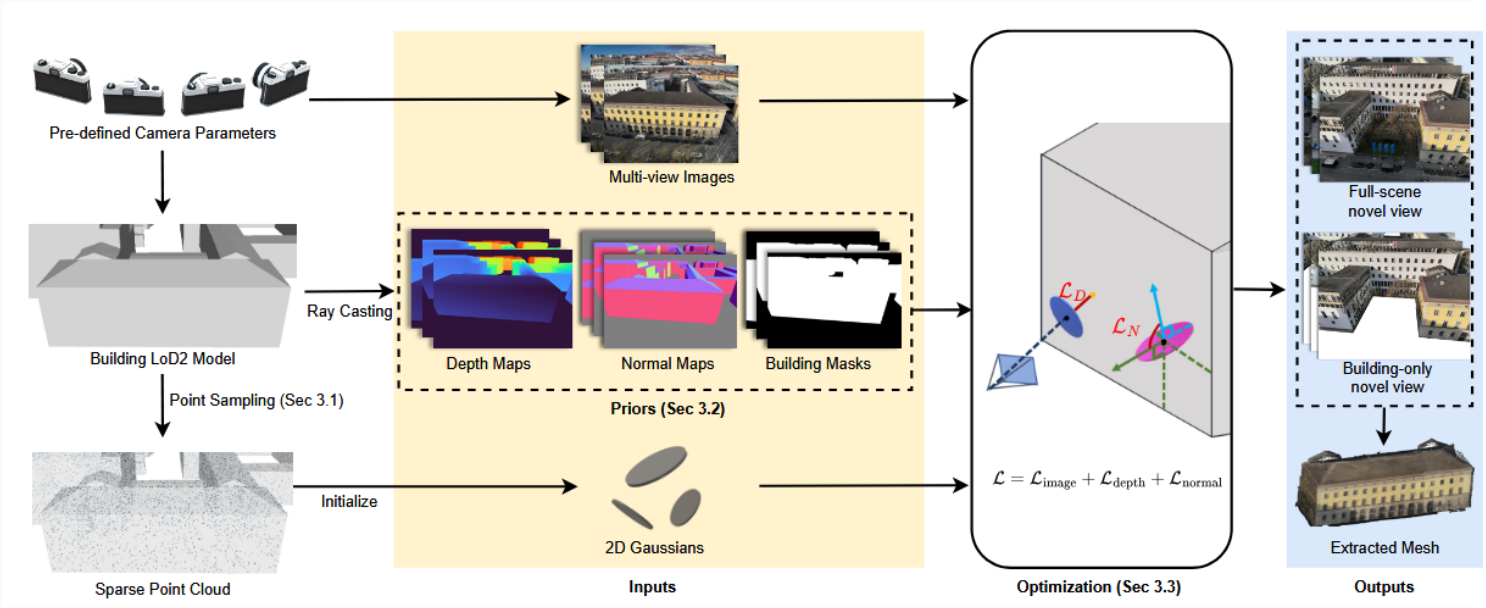
[3]

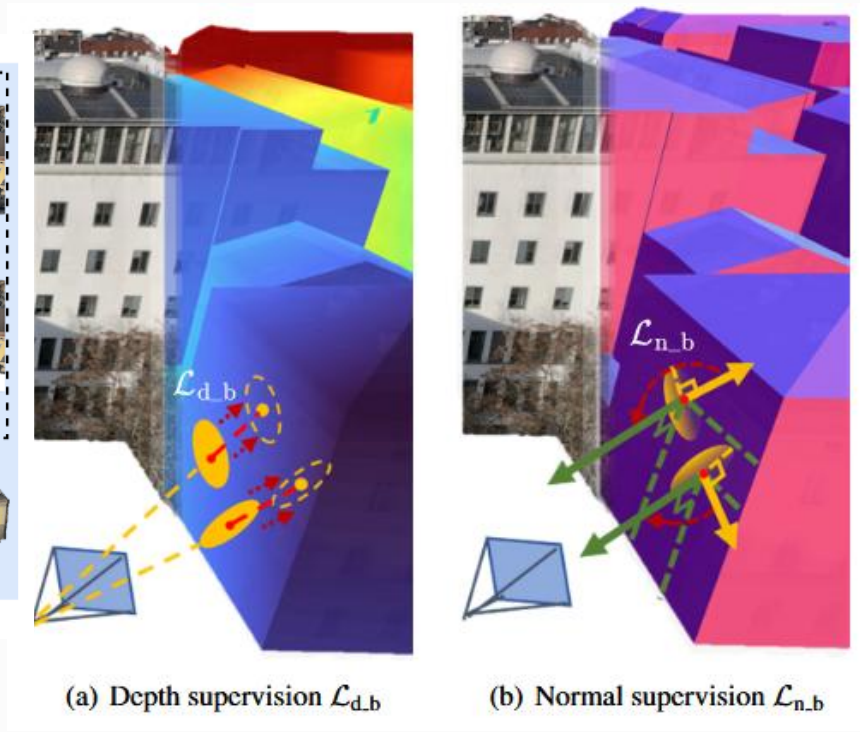
3D Semantic Reconstruction & 3D Models as Sensors

Prior-driven Gaussian Splatting

GS4B: Can we recycle worldwide-available 3D models to initialize + optimize GS reconstruction?

- Leverage prior model approximations to reduce noise and outliers
- Supervise by model-extracted depth and normals





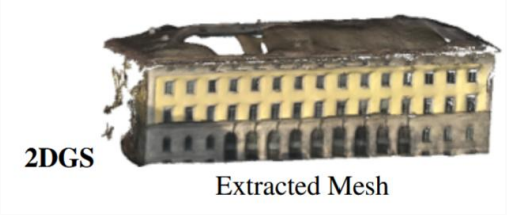
Zhang, Q., Wysocki, O., Jutzi, B. *GS4Buildings: Prior-Guided Gaussian Splatting for 3D Building Reconstruction* ***3DGeoInfo '25***

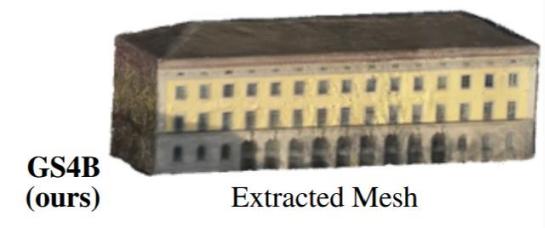
3D Models as Sensors & Reconstruction

GS4Buildings:

+33% higher accuracy,

+64% higher completness





Our GS4B

Our GS4B (Building-only mode)

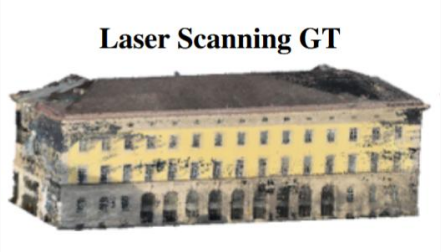
*Zhang, Q., Wysocki, O., Jutzi, B. GS4Buildings: Prior-Guided Gaussian Splatting for 3D Building Reconstruction **3DGeoInfo '25***

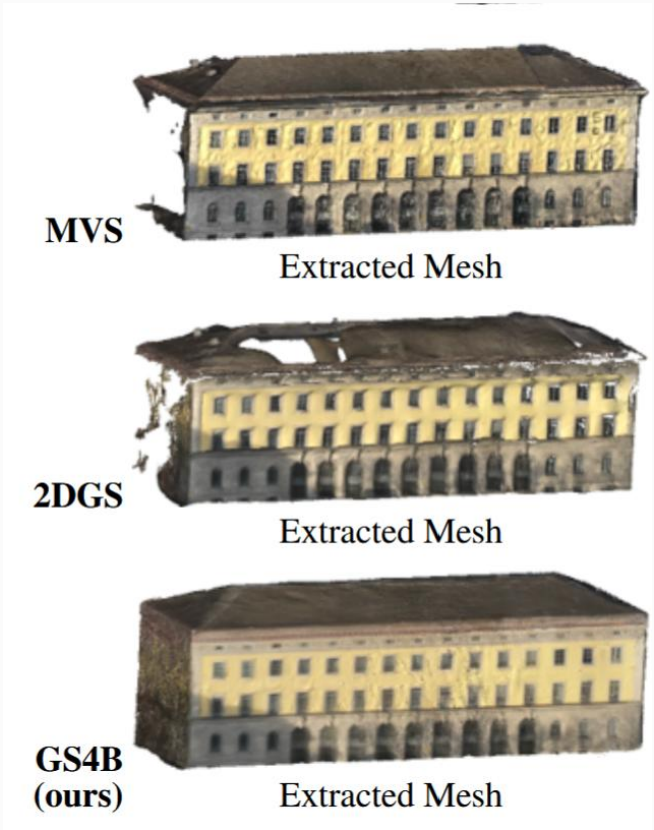
3D Semantic Reconstruction & 3D Models as Sensors

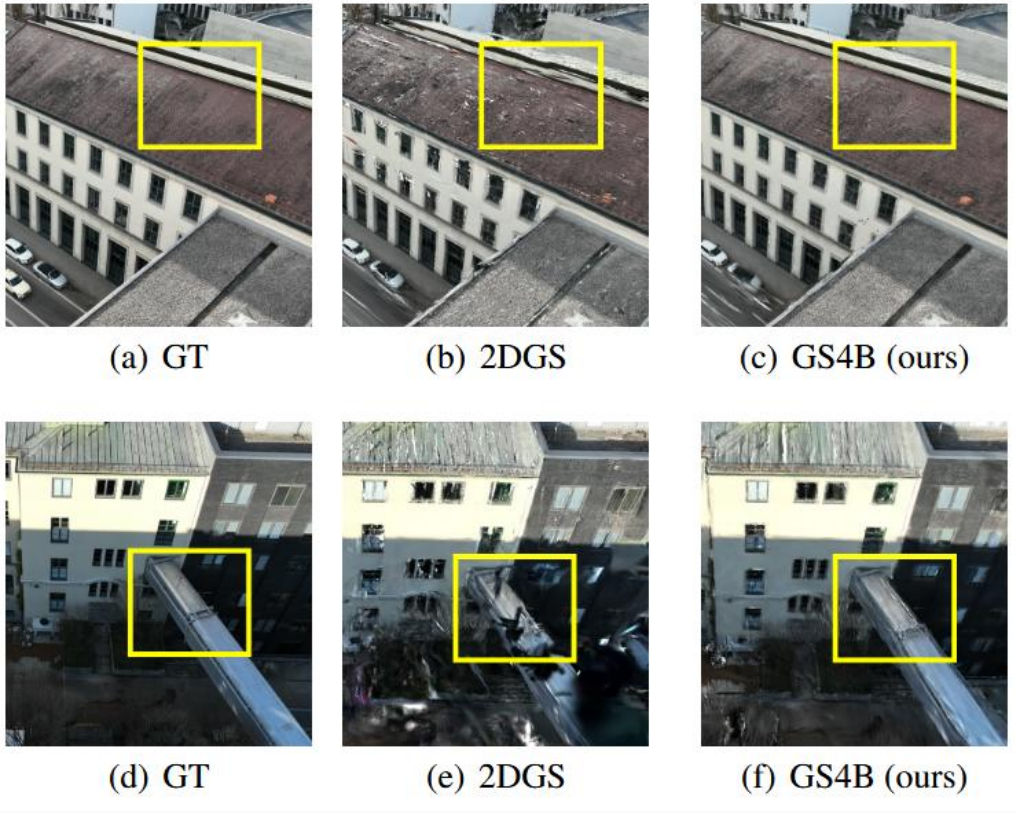
GS4B:

+33% higher accuracy,

+64% higher completness







Zhang, Q., Wysocki, O., Jutzi, B. *GS4Buildings: Prior-Guided Gaussian Splatting for 3D Building Reconstruction* ***3DGeoInfo '25***

3D Uncertainty-aware Reconstruction

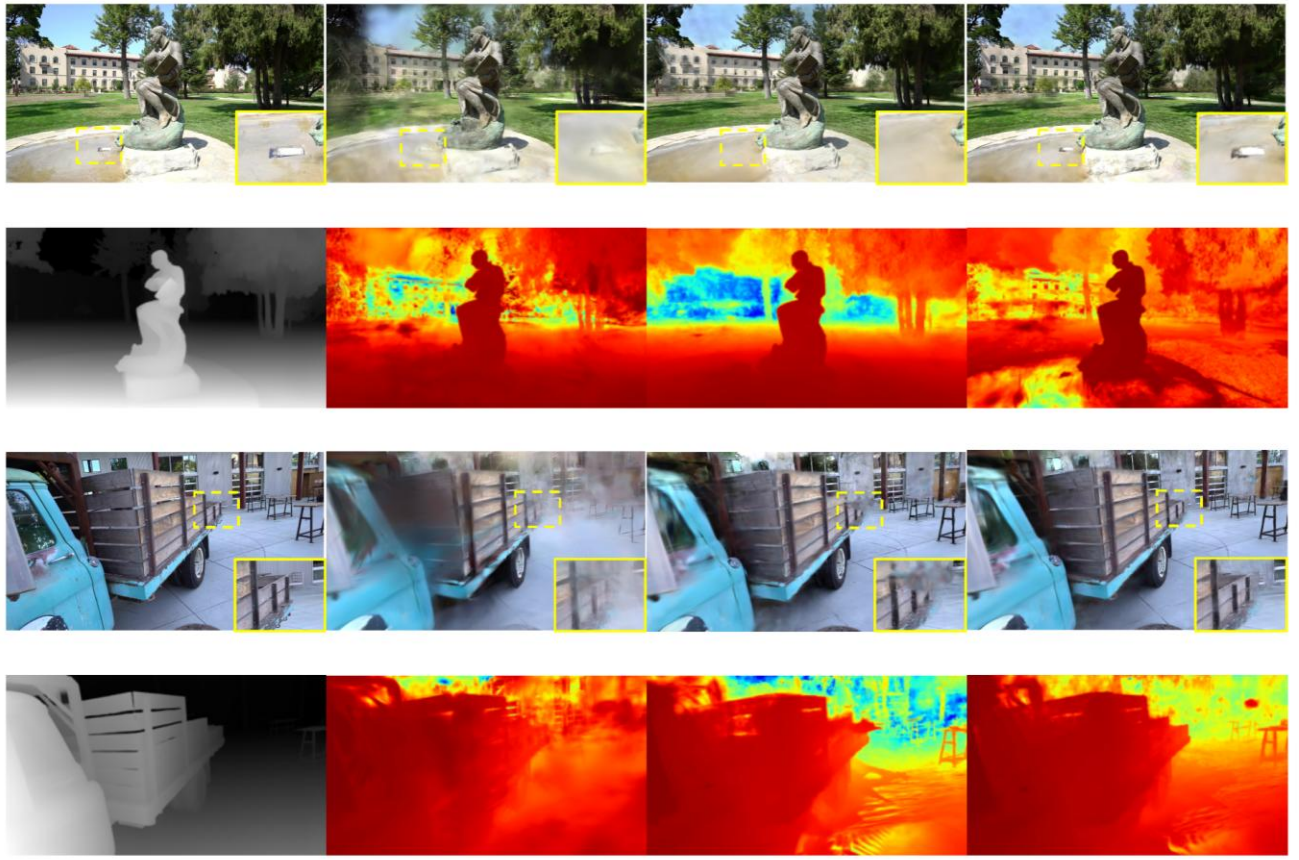
CD-GS: Uncertainty-aware Gaussian Splatting

Including confidence for RGB edge + depth maps

Confidence-Aware Depth Regularization:

- Generates confidence maps for each depth map
- Adaptively adjusts depth loss during optimization
- Enables reconstructing:
  - Small objects
  - Corner cases (literally as well!)

Ground truth point cloud | M3C2 distance visualization



Ground Truth | 3DGS | DRGS | CDGS (Ours)

Zhang, Q., Wysocki, O., Urban, S., Jutzi, B. *CDGS: Confidence-aware depth regularization for 3D Gaussian Splatting*, LowCost 3D '24

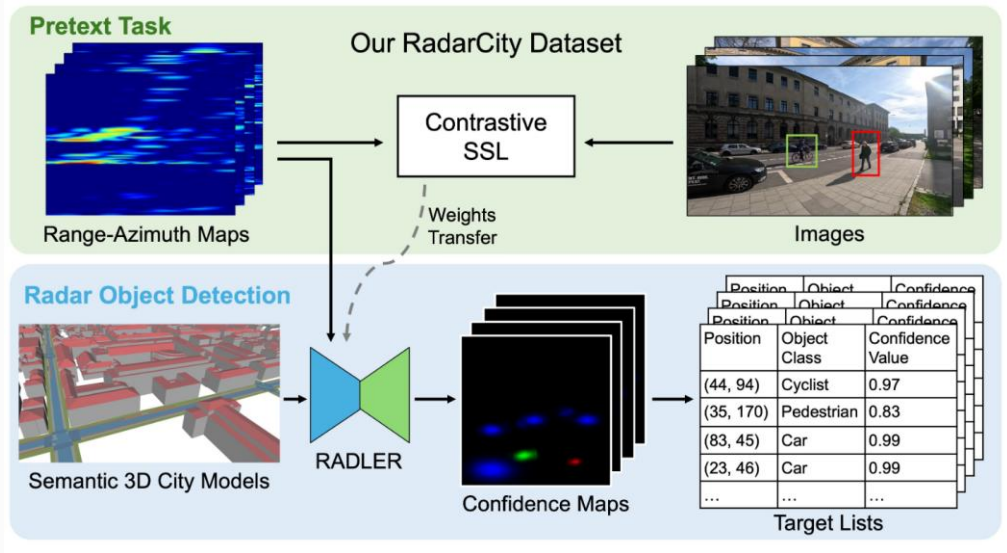
3D Models as Sensors

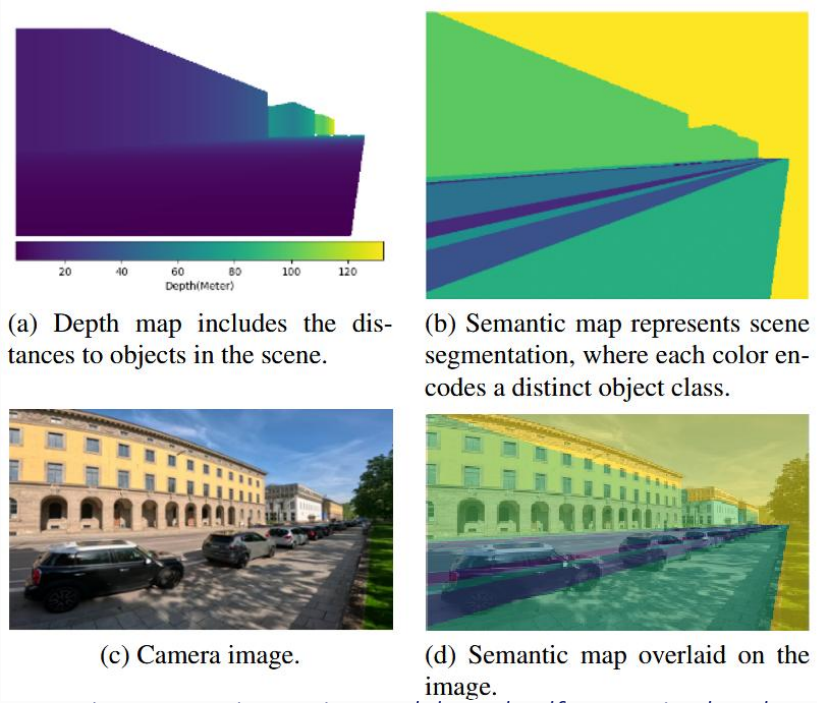
Can we use 3D semantic models for radar object detection?

Yes!

It turns out that:

- We get the depth estimation + semantics of the scene for free
- SSL used for getting the dynamic objects features
- We can improve detection results by ca. +7% vs SOTA





(a) Depth map includes the distances to objects in the scene.

(b) Semantic map represents scene segmentation, where each color encodes a distinct object class.

(c) Camera image.

(d) Semantic map overlaid on the image.

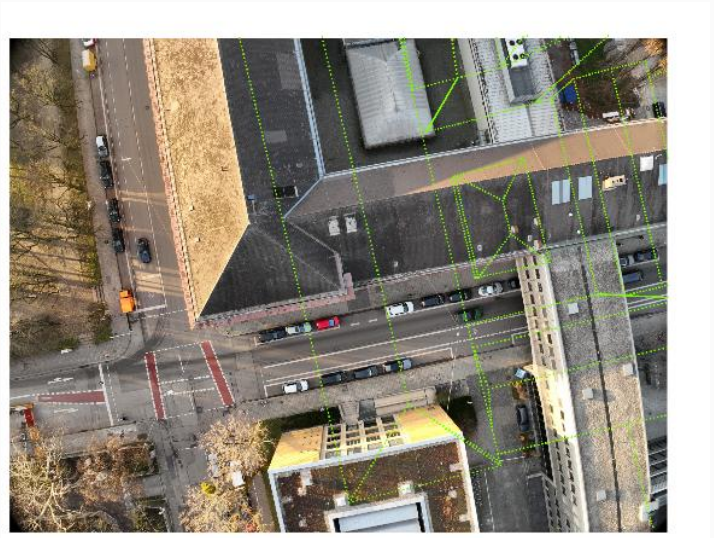
Luo Y., Hoffmann R., Xia Y., Wysocki O., Schwab B., Kolbe TH., Cremers D. *RADLER: Radar Object Detection Leveraging Semantic 3D City Models and Self-Supervised Radar-Image Learning, **CVPR '25***

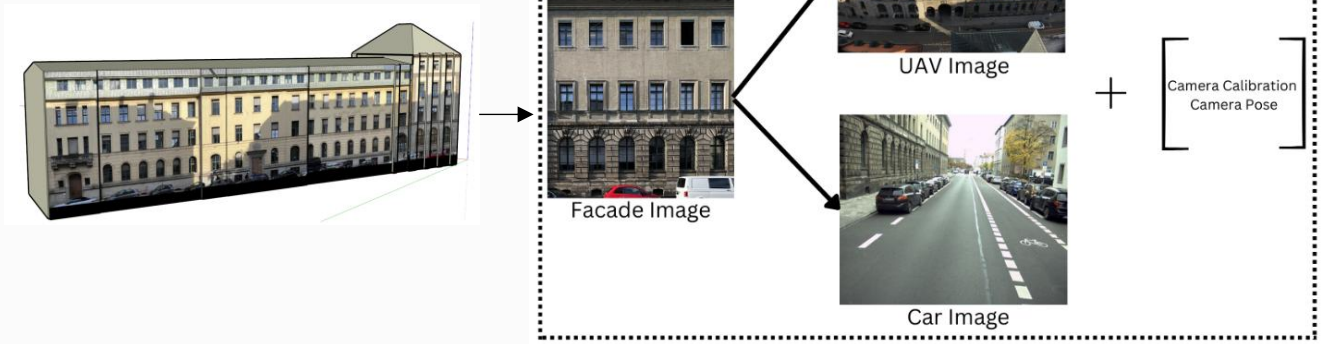
3D Models as Sensors

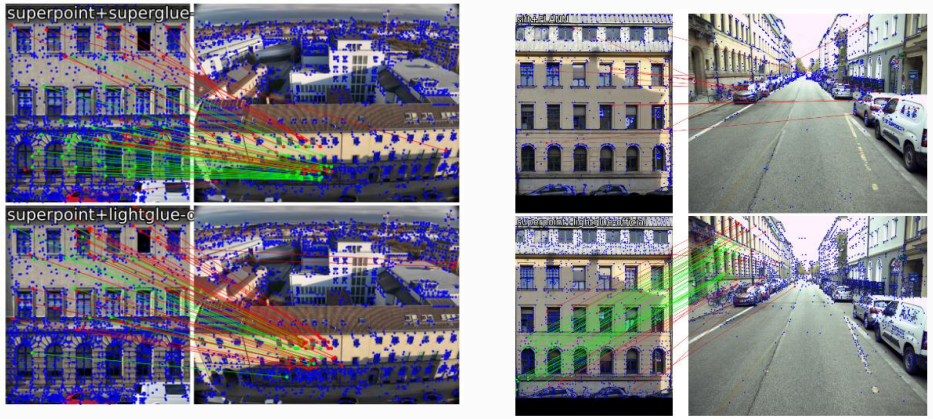
Can we use 3D semantic models for localization without GNSS?
Yes!

2D image to 3D geometry problem

- We get the 3D global (worldwide) coordinates for free
- Drastic changes in representation still challenging
- Still possible to retrieve inliers (+optional radiometry in limited cases)







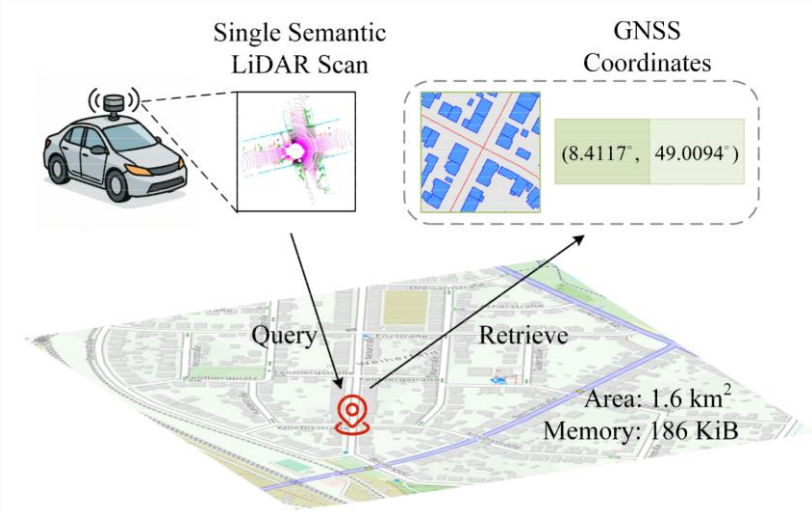
Gaisbauer S., Gyawali P., Zhang Q., Wysocki O., Jutzi B. *To Glue or Not to Glue? Classical vs Learned Image Matching for Mobile Mapping Cameras to Textured Semantic 3D Building Models,* ***ISPRS Annals '25*** */// Bieringer A., Wysocki O., Tuttas S., Hoegner L., Holst Ch. Analyzing the Impact of Semantic LoD3 Building Models on Image-based Vehicle Localization,* ***3D GeoInfo '24***

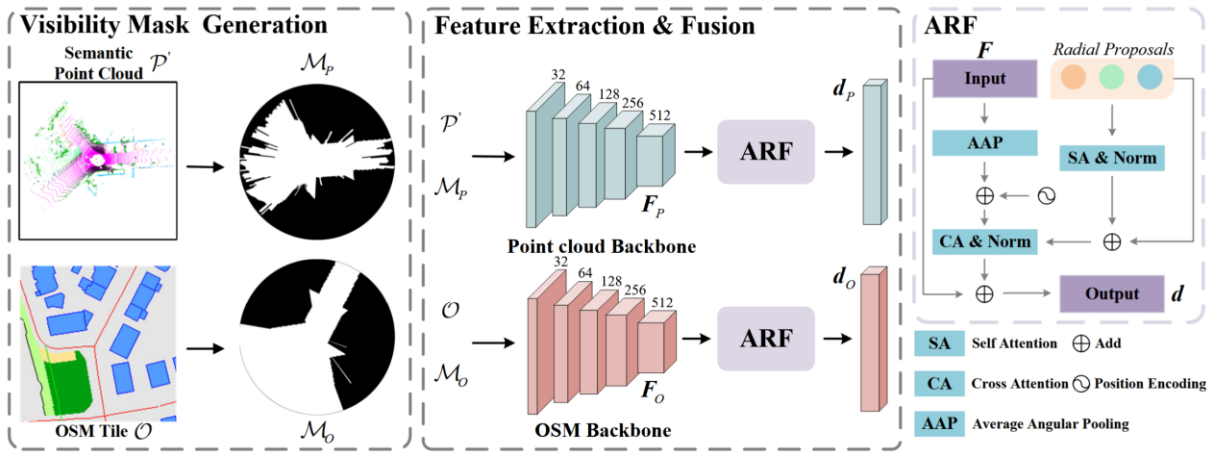
3D Models as Sensors

Can we use semantics of OpenStreetMap for LiDAR place recognition?

Turns out that we need only one scan to get meter-level accuracy

- Generate visibility mask to attend to visible regions
- Siamese CNN to get local feature maps in BEVs (point cloud + OSM)
- Adaptive Radial Fusion (ARF) for dynamic radial feature weighting
- Achieving +16% higher recall (KITTI, KITTI-360)





Kang S., Liao MY., Xia Y., Wysocki O., Jutzi B., Cremers D. *OPAL: Visibility-aware LiDAR-to-OpenStreetMap Place Recognition via Adaptive Radial Fusion, Accepted to* ***CoRL '25***

Outlook

Digital Twins – they need 3D vision

Sense → Understand & Reconstruct → Act → Update

Unsolved!

Mission of CV4DT!

~~Virtual~~ reality

3D computer vision for built environment

Can we go beyond the toy examples?

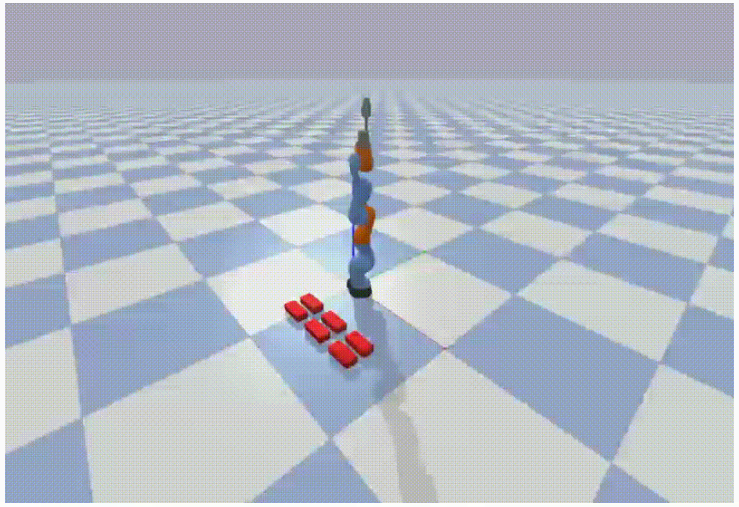
a plush dragon toy

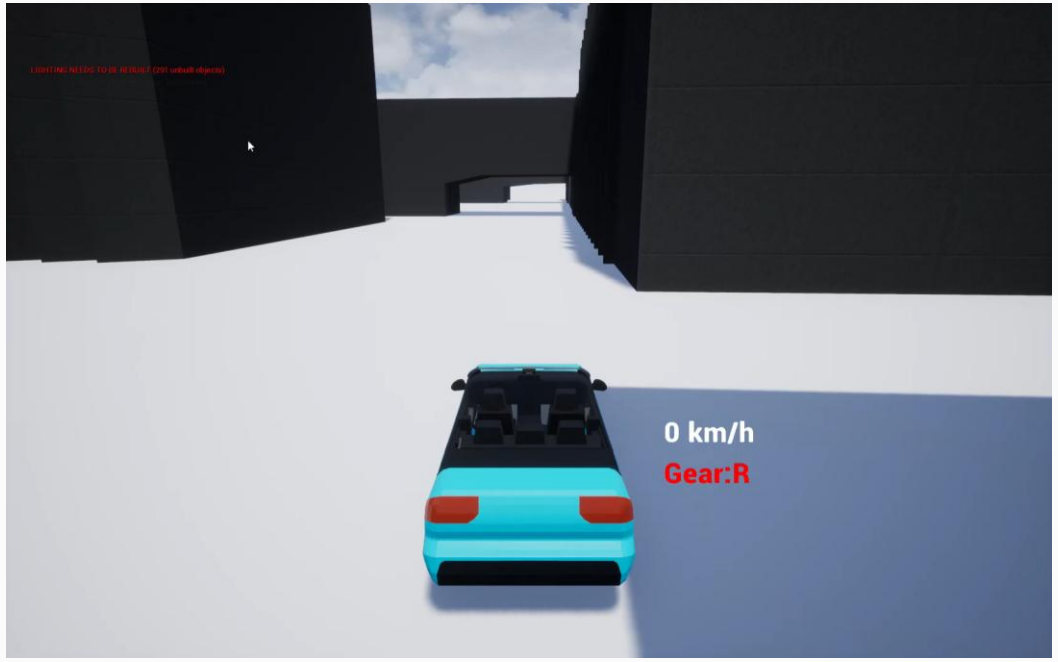
A construction site

One-2-3-45++: Fast Single Image to 3D Objects with Consistent Multi-View Generation and 3D Diffusion, CVPR 2024

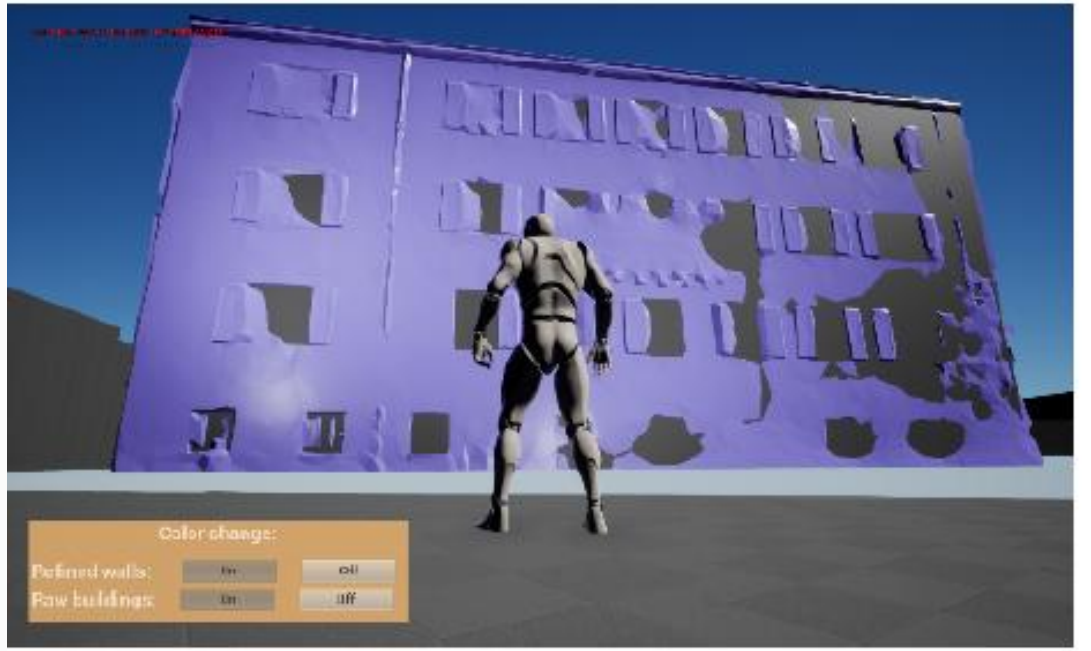
World Models – they need 3D vision

3D object-oriented semantic environment as physics prior to 2D/3D perception



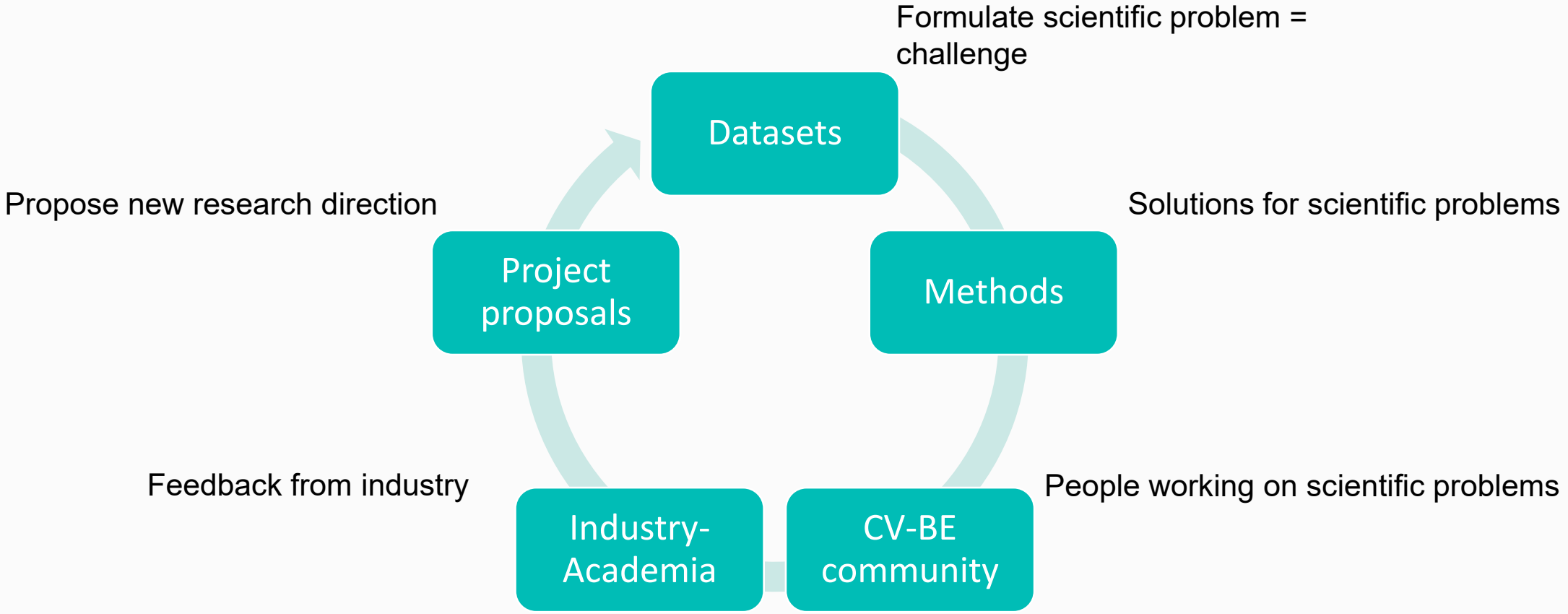


Demo game – Driving in 3D reconstructed urban environment of Munich, Germany



Demo game – Walking in 3D reconstructed urban environment of Ingolstadt, Germany (toggle on and off improved models)

CV4DT – On the mission to create the built environment ecosystem



CamCV – Cambridge Community for Computer Vision

Gathering CV-related researchers of Cambridge!

- Webpage
- Regular meet-ups
- Invited talks
- Pre-conference events (e.g., CVPR/ICCV… paper presentations in Cambridge)
- …and more?

UNIVERSITY OF CAMBRIDGE

CV4DT

CSIC Cambridge Centre for Smart Infrastructure & Construction

LAING O'ROURKE CENTRE for CONSTRUCTION ENGINEERING and TECHNOLOGY

CV4DT