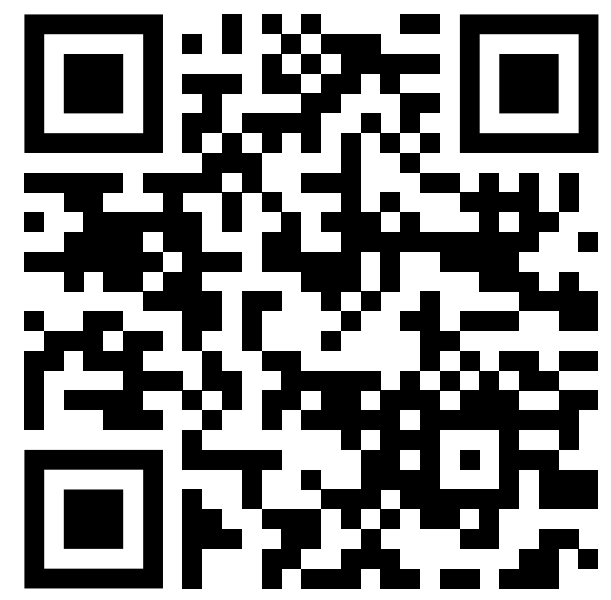




**MAX PLANCK INSTITUTE**  
FOR INFORMATICS

**SIC** Saarland Informatics  
Campus

**ETH** zürich



# Rewis3d

## Reconstruction Improves Weakly-Supervised Semantic Segmentation



Jonas Ernst\*



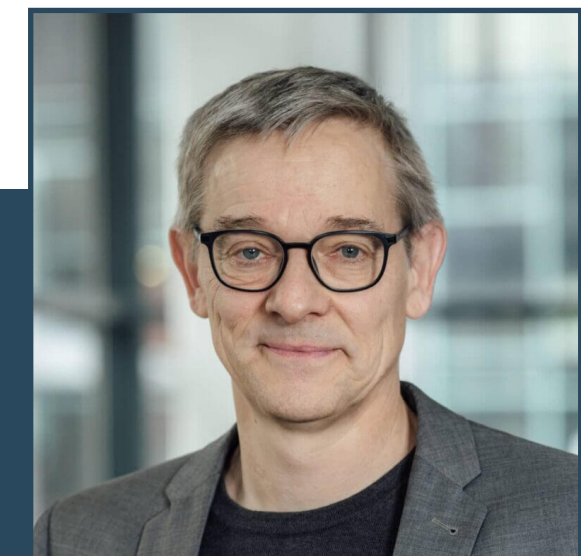
Wolfgang Boettcher\*



Lukas Hoyer



Jan Eric Lenssen

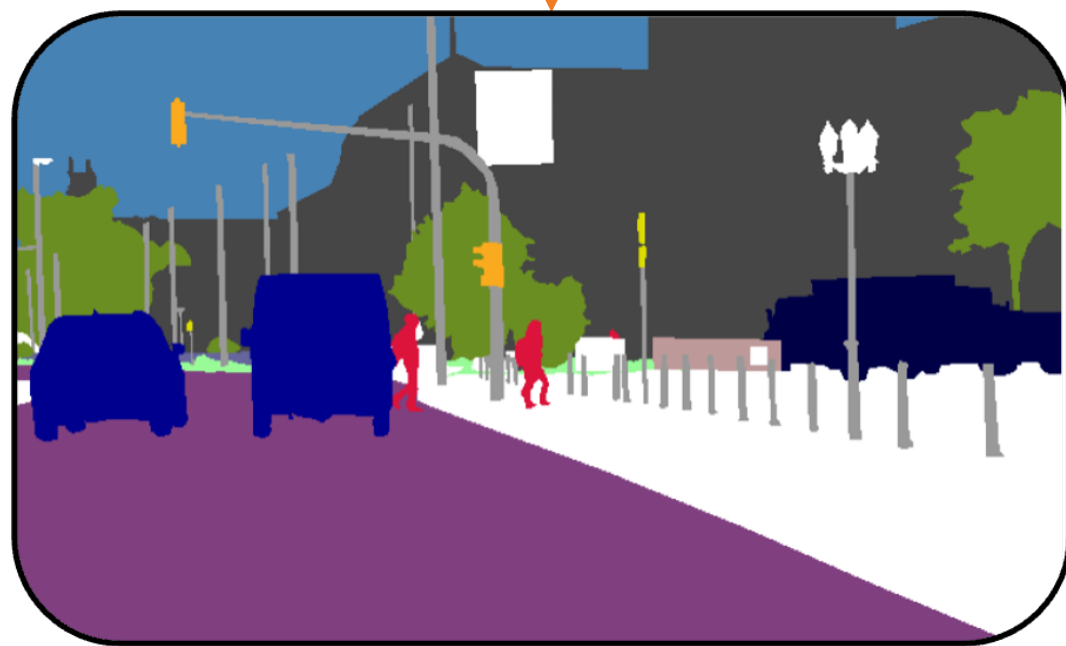


Bernt Schiele

# Dense Annotations are Expensive



- Reduce Ann. Effort → Sparse Labels
- Perform weakly supervised semantic segmentation (WSSSS)



Up to 2 hrs.



~ 20 sec.

~ 2 min.

~ 7 min.

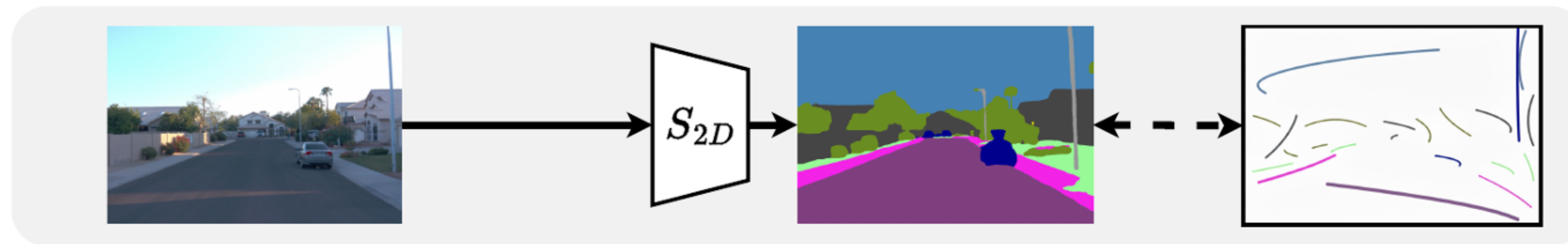
# What Makes this Challenging?



- Missing explicit class boundary information
- Multiple object scales not properly represented in labels
- Aggravated Co-Occurrence Bias

# Rewis3d - Leverage 3D Reconstruction

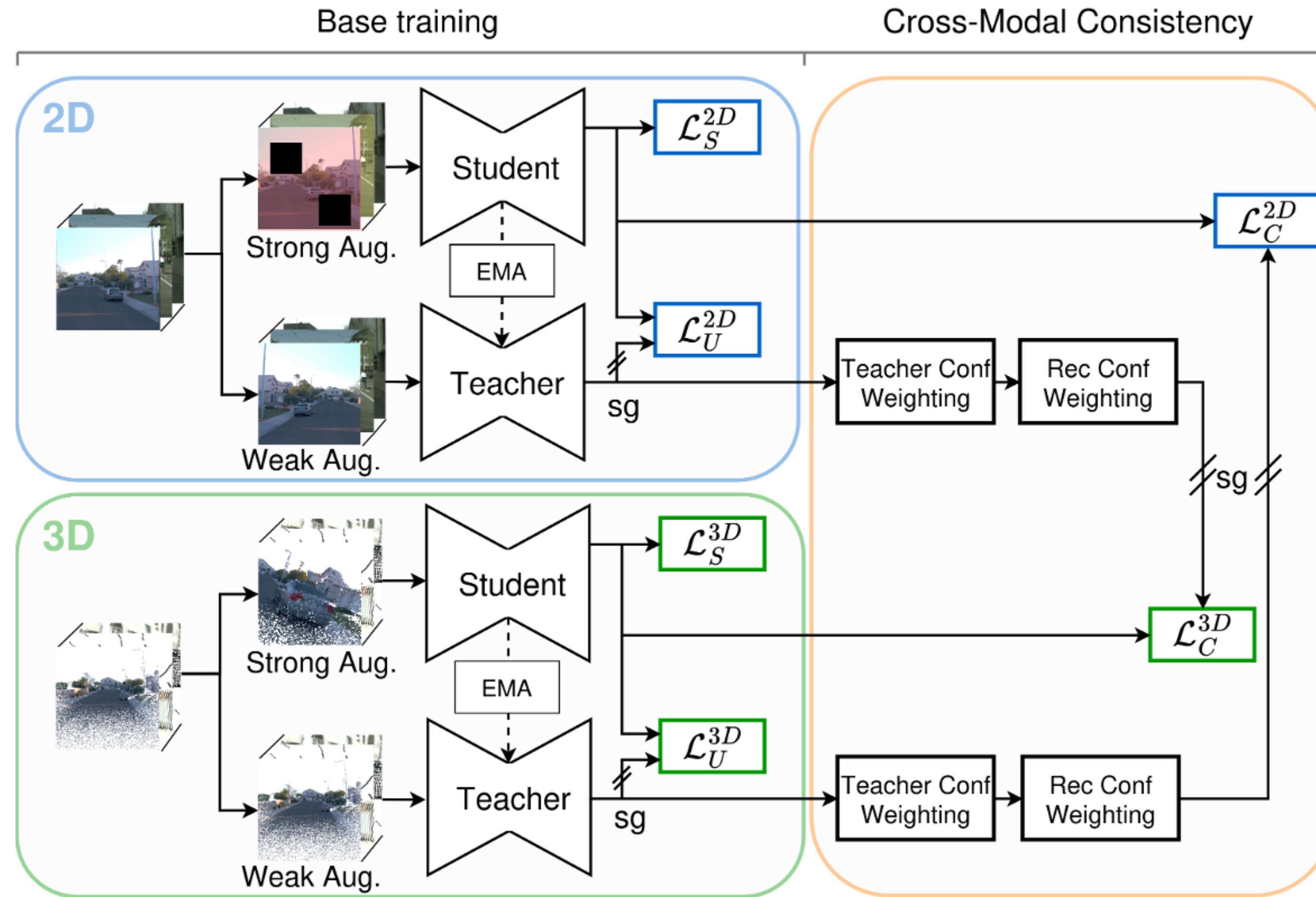
## Prior Work



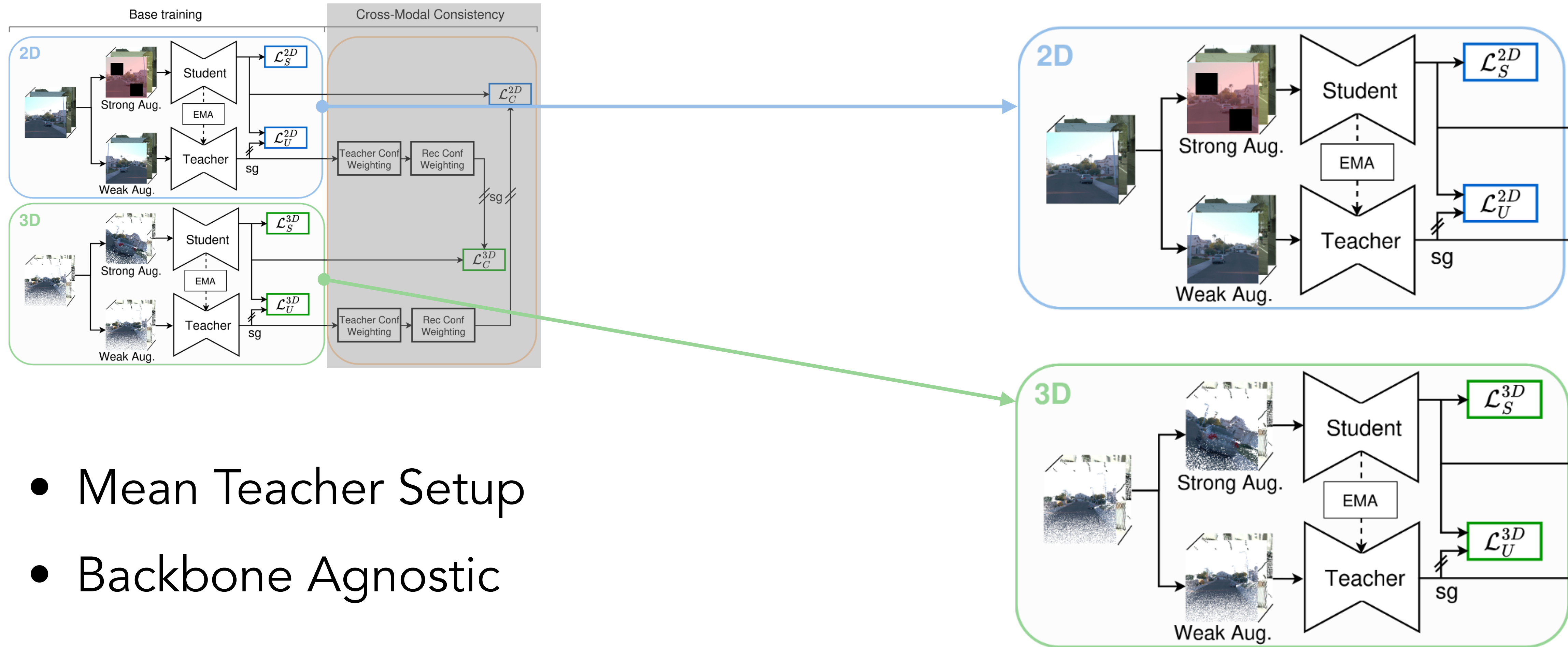
- Single Image - Label Pairs
- Operate in 2D modality alone
- Rely on regularizations to apprehend class boundaries



# Rewis3D - Pipeline

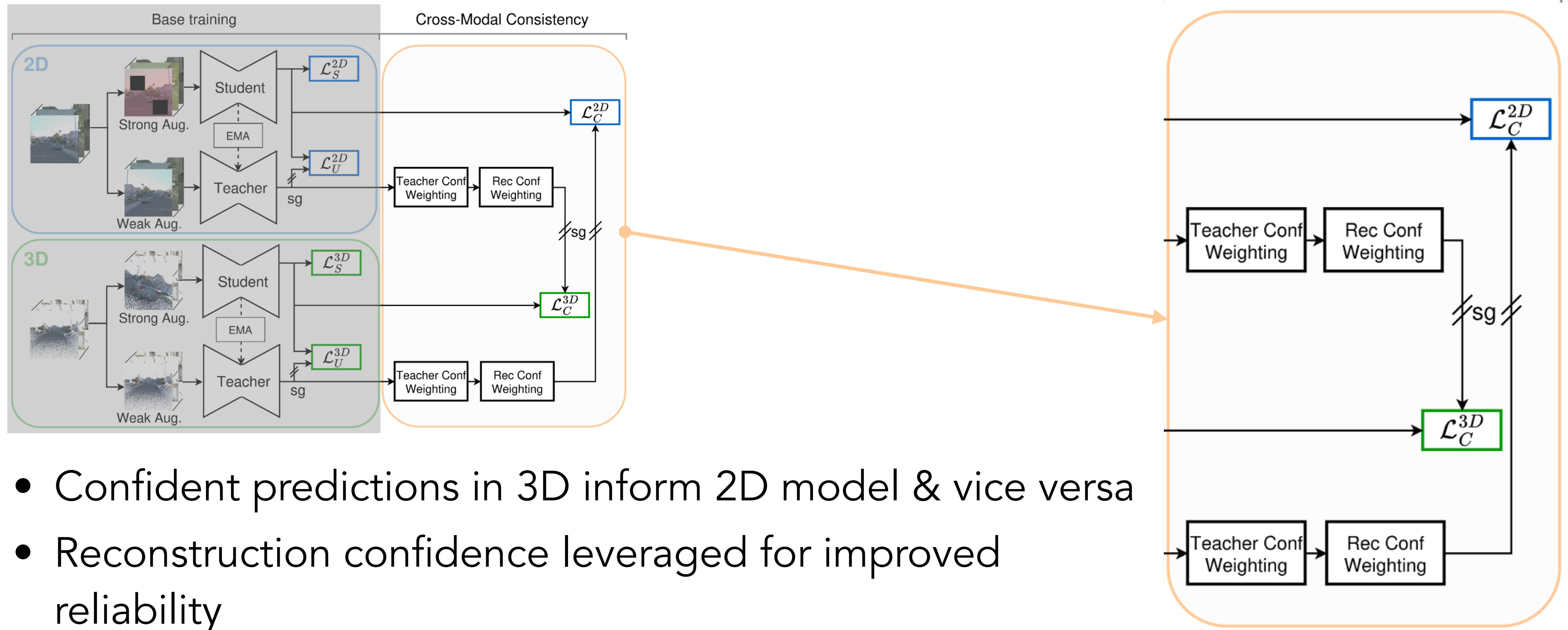


# Rewis3D - Pipeline



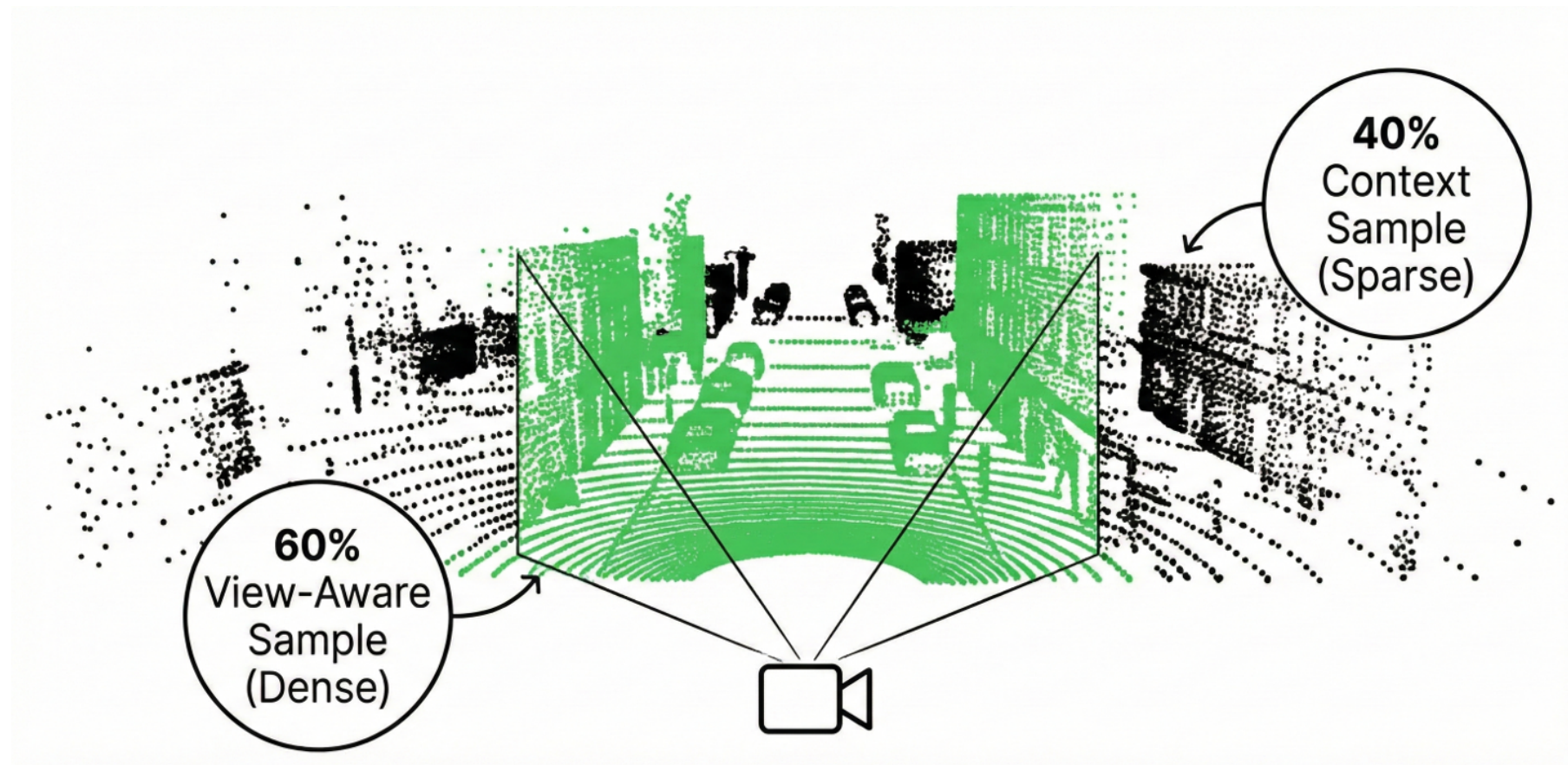
- Mean Teacher Setup
- Backbone Agnostic

# Rewis3D - Pipeline

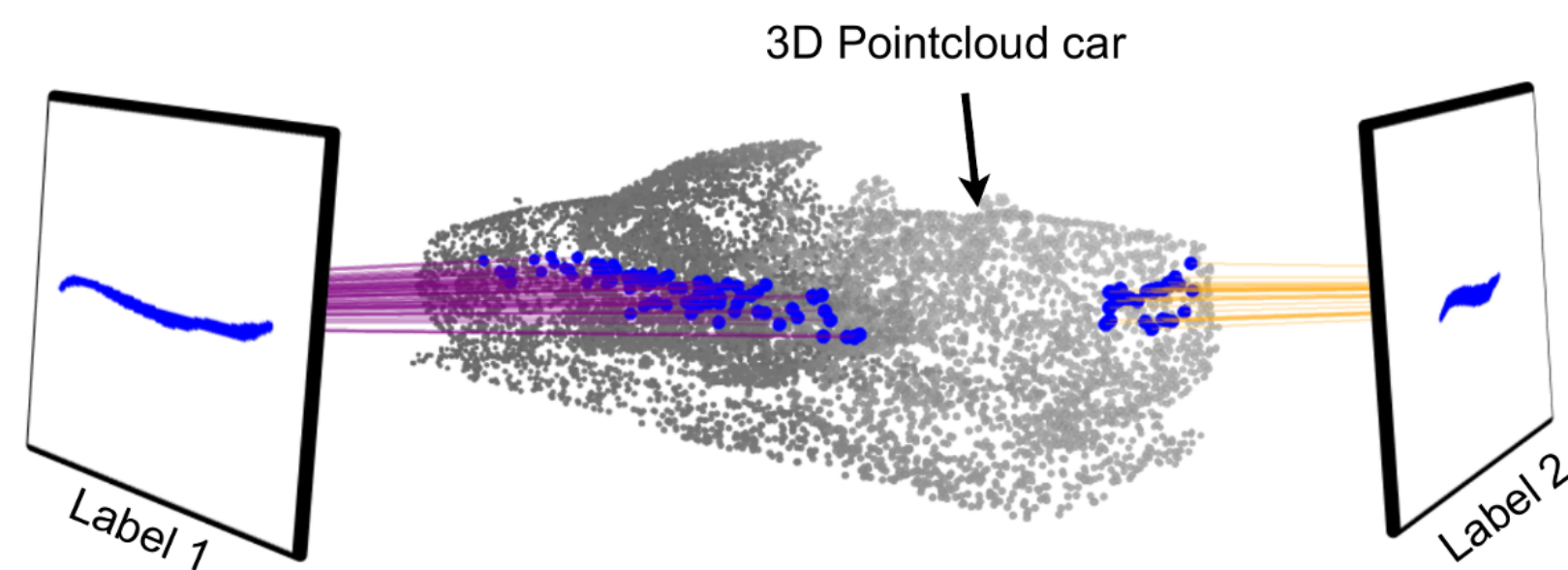


- Confident predictions in 3D inform 2D model & vice versa
- Reconstruction confidence leveraged for improved reliability

# How to Exploit Cross Modality?

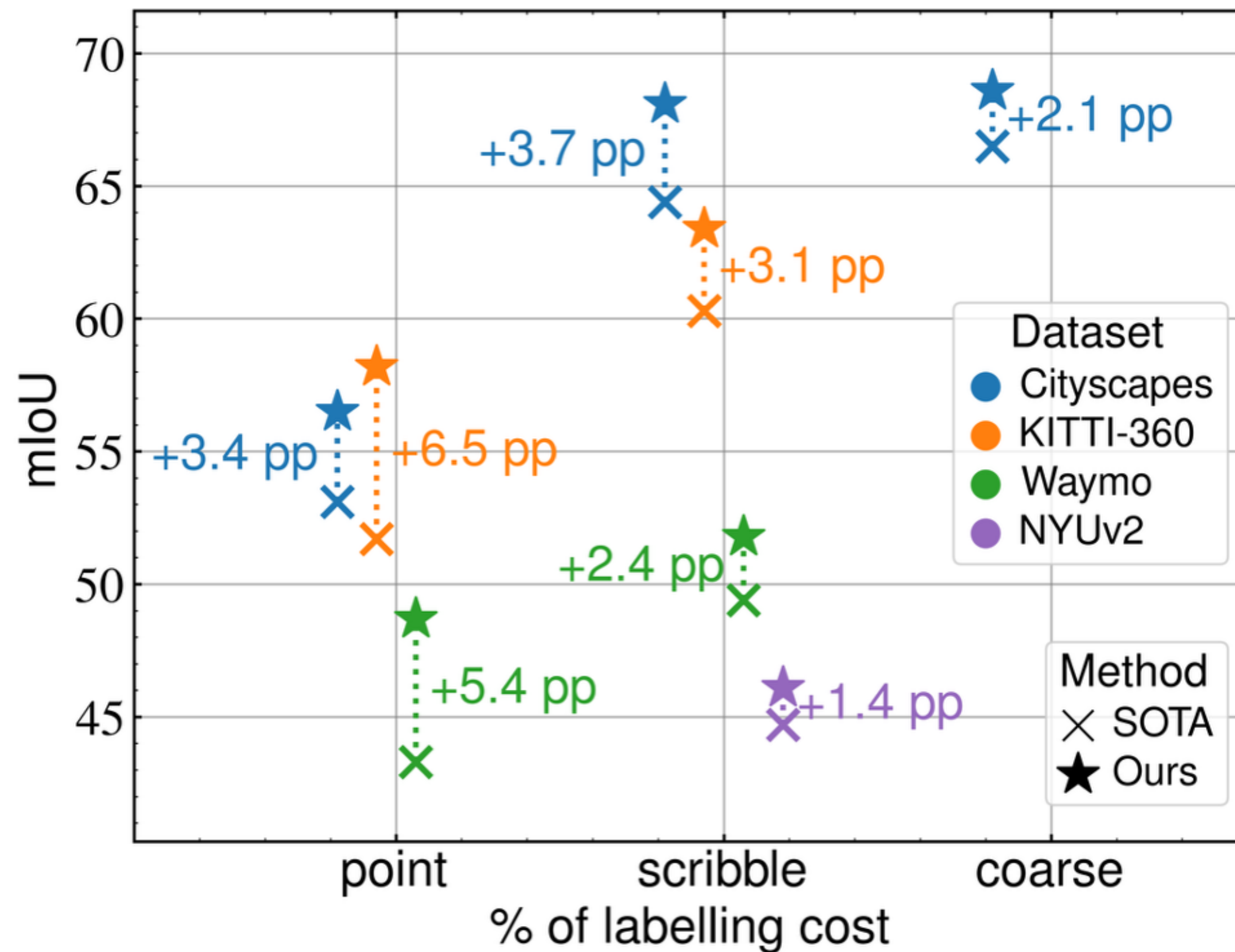


- Reconstruct 3D scene from multiple images



- Aggregating points from multiple frames affords label accumulation

# Rewis3d Outperforms for All WSSS Scenarios



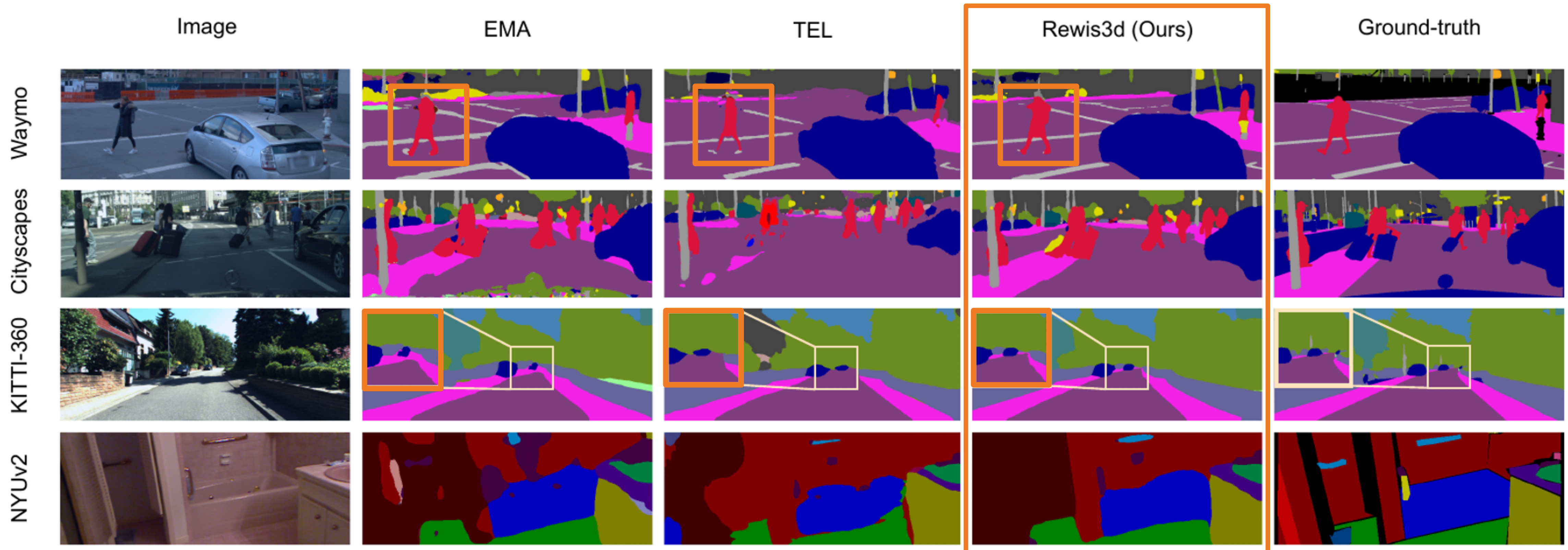
- ✓ Indoor
- ✓ Outdoor
- ✓ Point Labels
- ✓ Scribble Labels
- ✓ Coarse Labels
- ✓ Single-View
- ✓ Image-Sequence

# Rewis3d Outperforms for all WSSS Scenarios

Method	Ours	Backbone	3D Supervision	Waymo		KITTI-360		NYUv2	
				mIoU	SS/FS (%)	mIoU	SS/FS (%)	mIoU	SS/FS (%)
Fully Supervised		Segformer-B4	–	59.0	–	68.4	–	51.1	–
EMA		Segformer-B4	–	49.4	83.7	60.3	88.2	42.9	84.0
SASFormer		Segformer-B4	–	37.8	64.1	46.4	67.8	44.7	87.5
TEL		DeepLabV3+	–	42.4	71.9	59.2	86.6	38.3	75.0
<b>Ours (Real 3D)</b>	✓	Segformer-B4	LiDAR/Depth	<u>51.8</u>	<u>87.8</u>	<u>61.7</u>	<u>90.2</u>	<u>44.7</u>	<u>87.6</u>
<b>Ours (Recon)</b>	✓	Segformer-B4	Recon.	<b>53.3</b>	<b>90.3</b>	<b>63.4</b>	<b>93.4</b>	<b>46.1</b>	<b>90.2</b>

- Performance Gap to fully supervised closed to **>90 %**
- Improved performance irrespective of LiDAR or reconstructed 3D

# Rewis3d Improves Class Boundaries



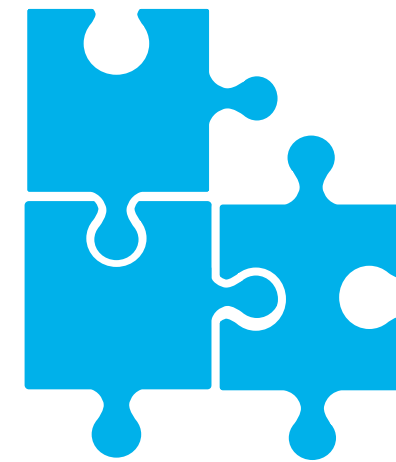
- Sharper, more faithful class boundaries
- Improved long-range details

# Takeaways



## SOTA Performance

Outperforms existing methods by 2-7% mIoU on major datasets.



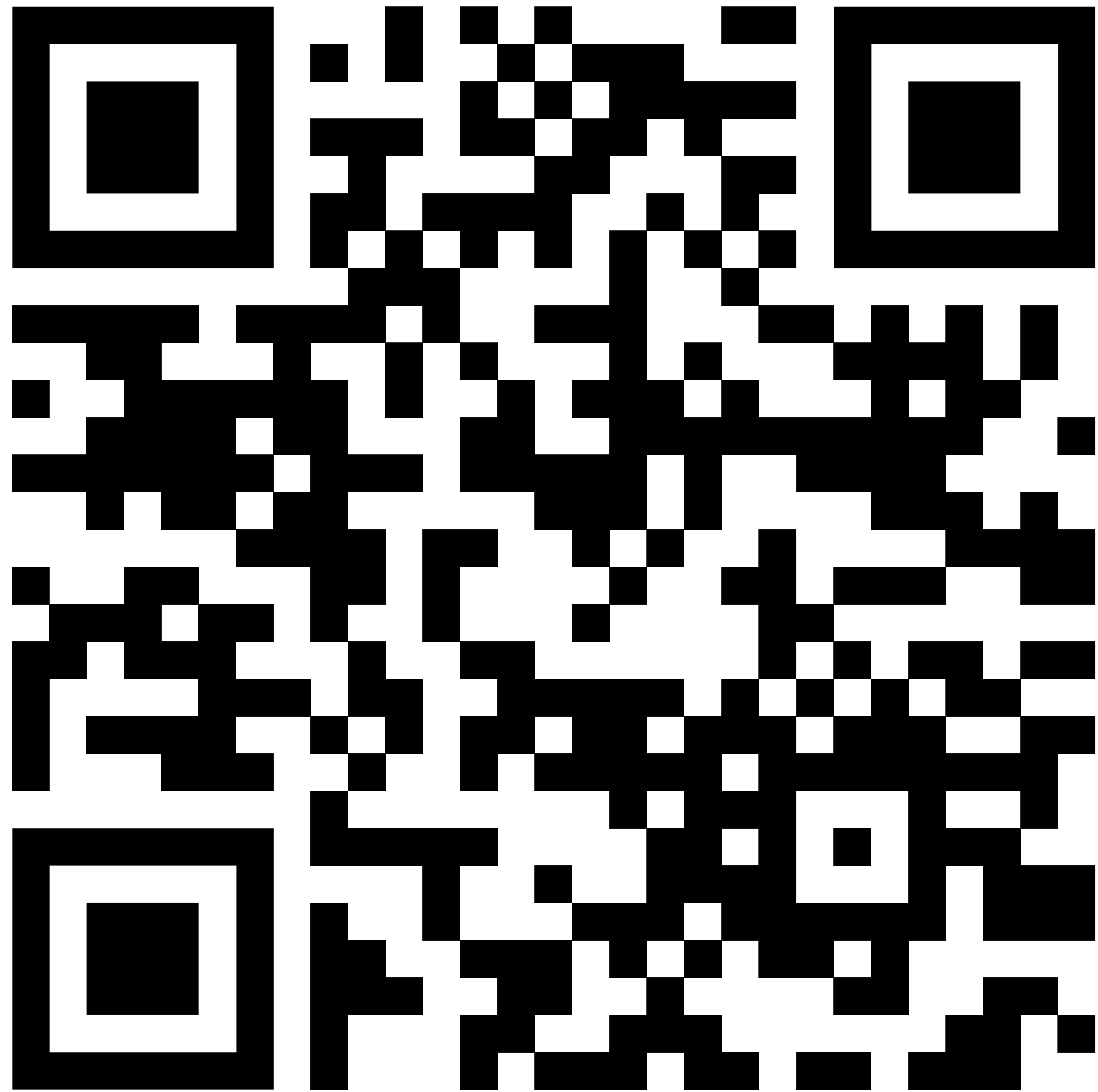
## Geometry is Key

Proves that 3D structure is a powerful supervisor for 2D tasks.



## Versatility through Reconstruction

Uses image reconstruction. No LiDAR required. No inference overhead.



- For further information, please visit the project page:  
[rewis3d-mpi.github.io/Rewis3d/](https://rewis3d-mpi.github.io/Rewis3d/)
- For the codebase, please go to our GitHub page:  
[github.com/Rewis3d-MPI/Rewis3d](https://github.com/Rewis3d-MPI/Rewis3d)