UNIVERSITÄT BONN

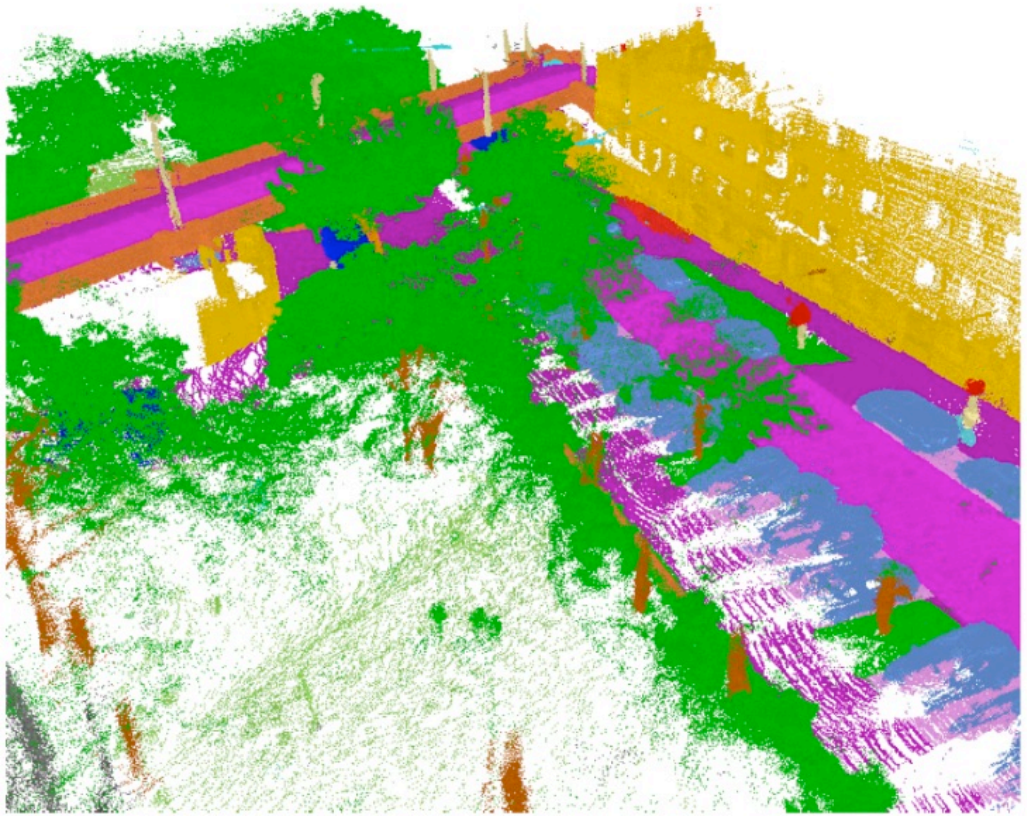
Juergen Gall

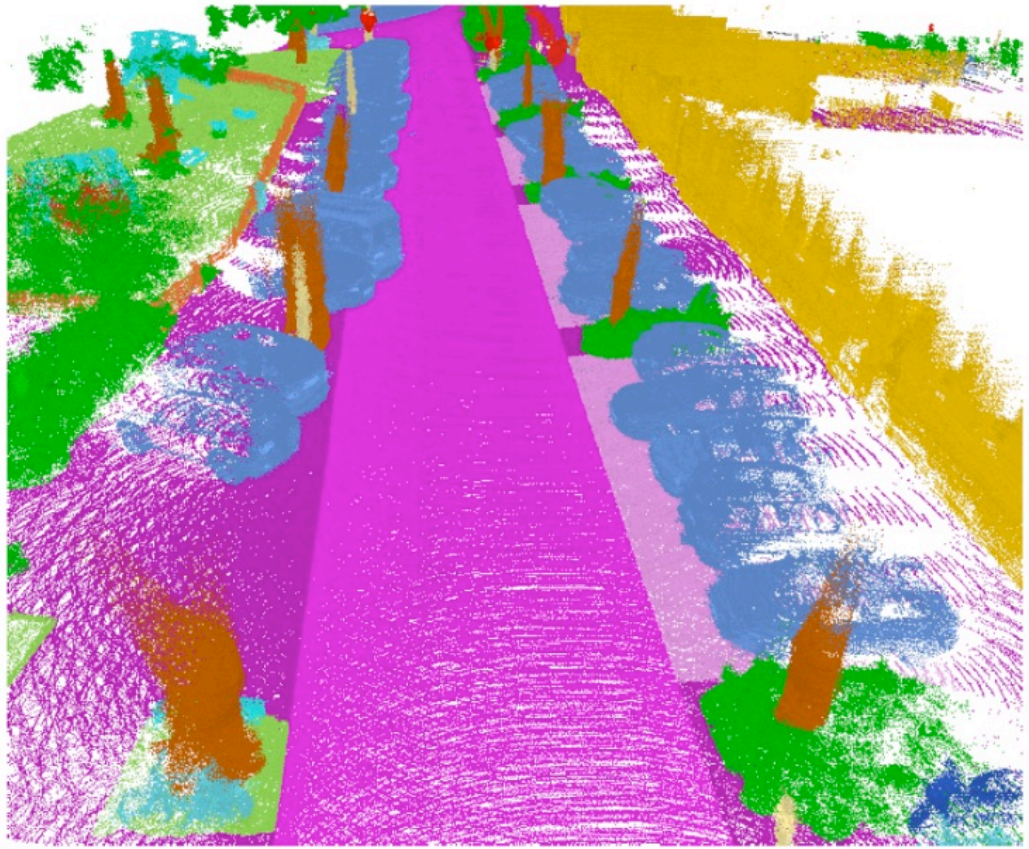
3D LiDAR-based Semantic Scene Understanding

Understanding LiDAR data

SemanticKITTI

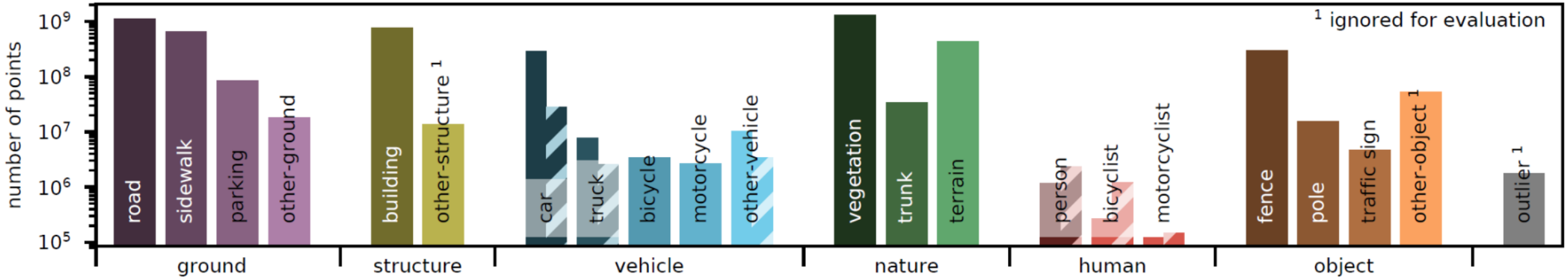
- Accurate dense annotations for each scan of all sequences from the KITTI Odometry Benchmark





Data and Annotation

- 43’000 scans from KITTI Vision Benchmark (Velodyne HDL-64E)
- Classes related to Mapillary Vistas and Cityscapes dataset:

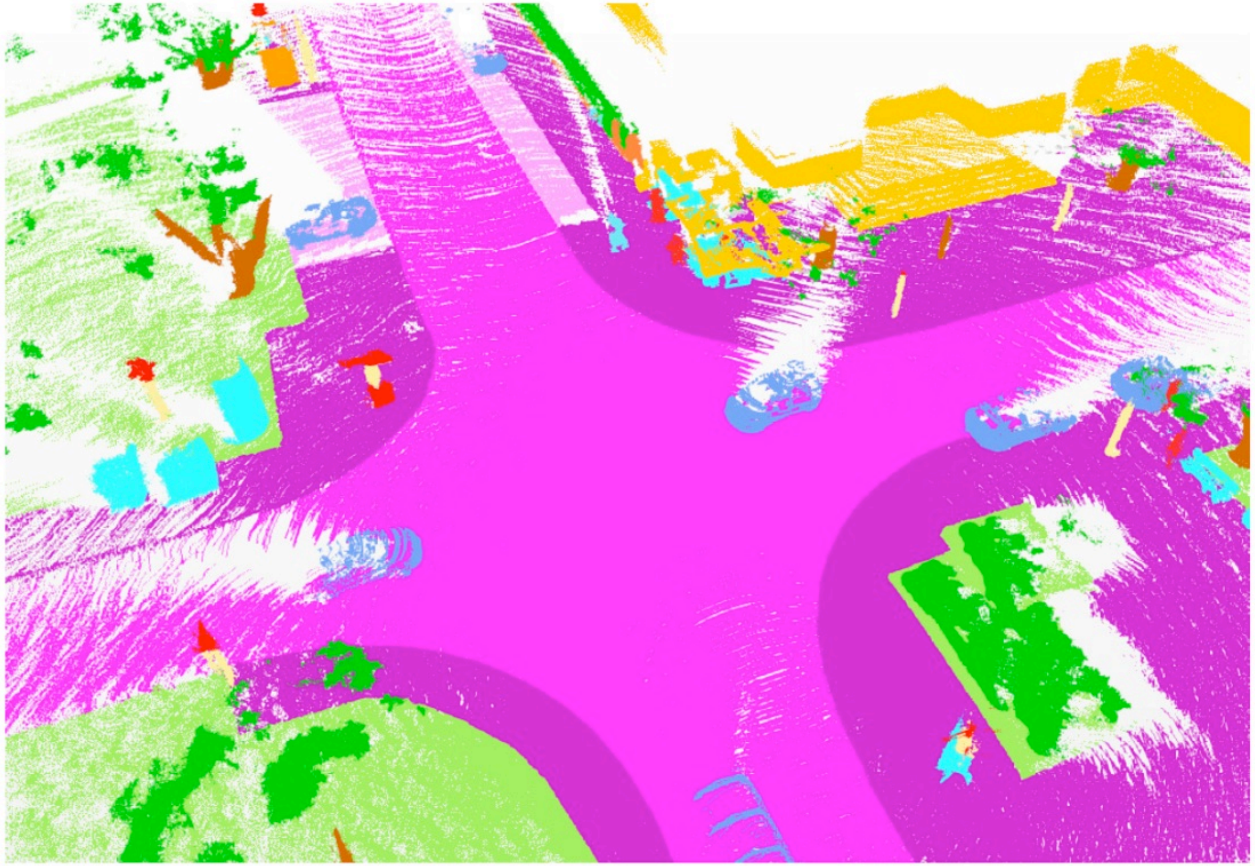


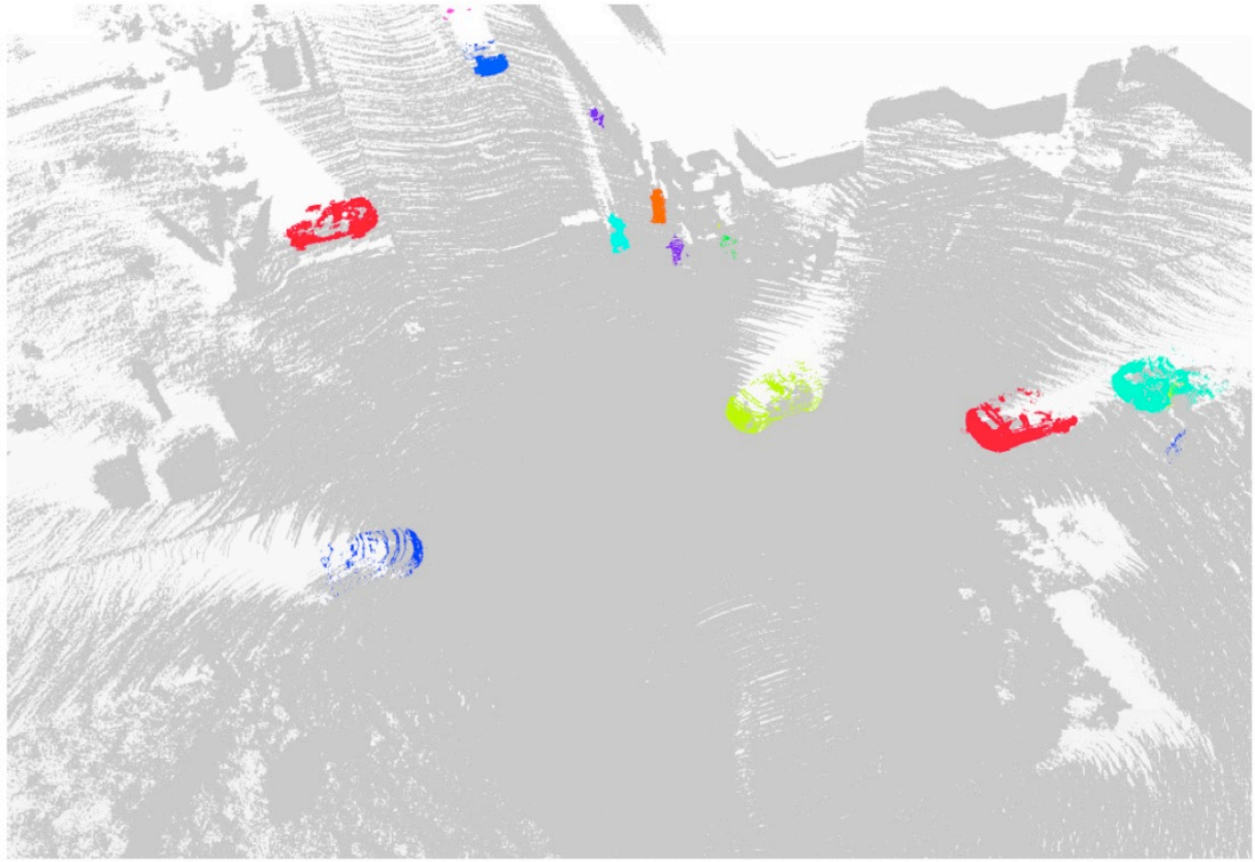
Moving objects (stripes)

[J. Behley et al. **SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences.** ICCV 2019; International Journal of Robotics Research 2021]

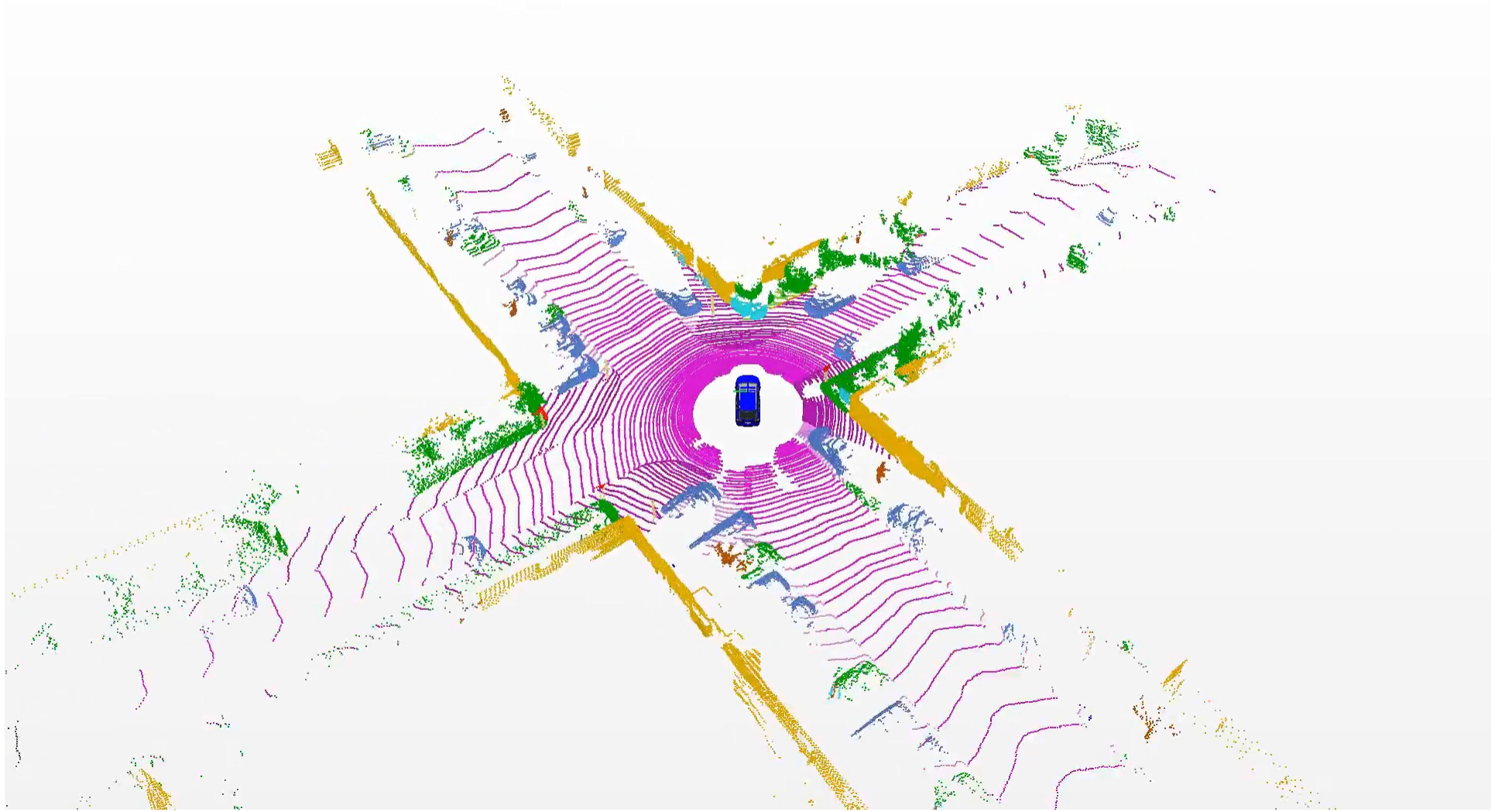
Data and Annotation

- Instance annotations



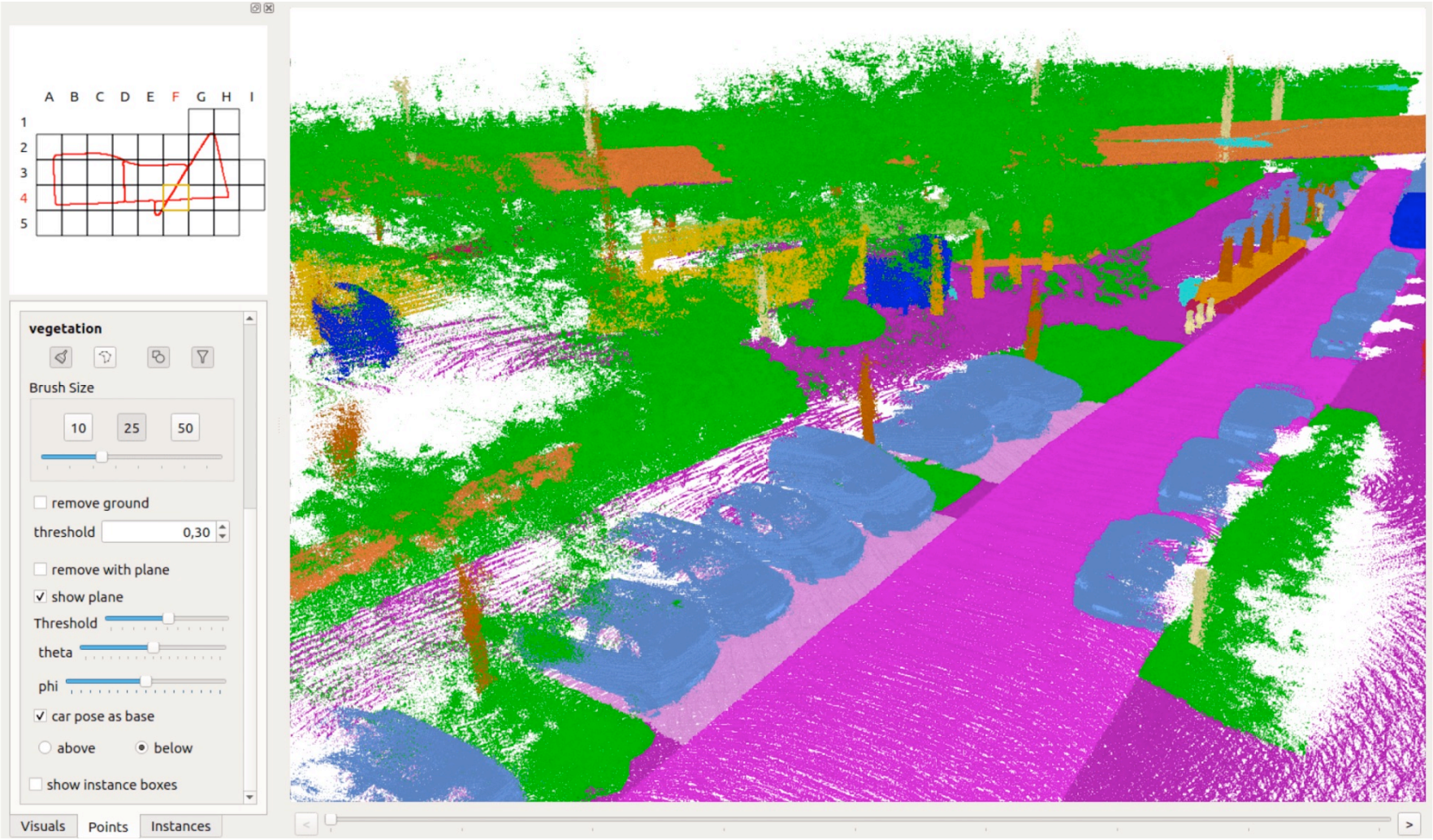


SemanticKITTI

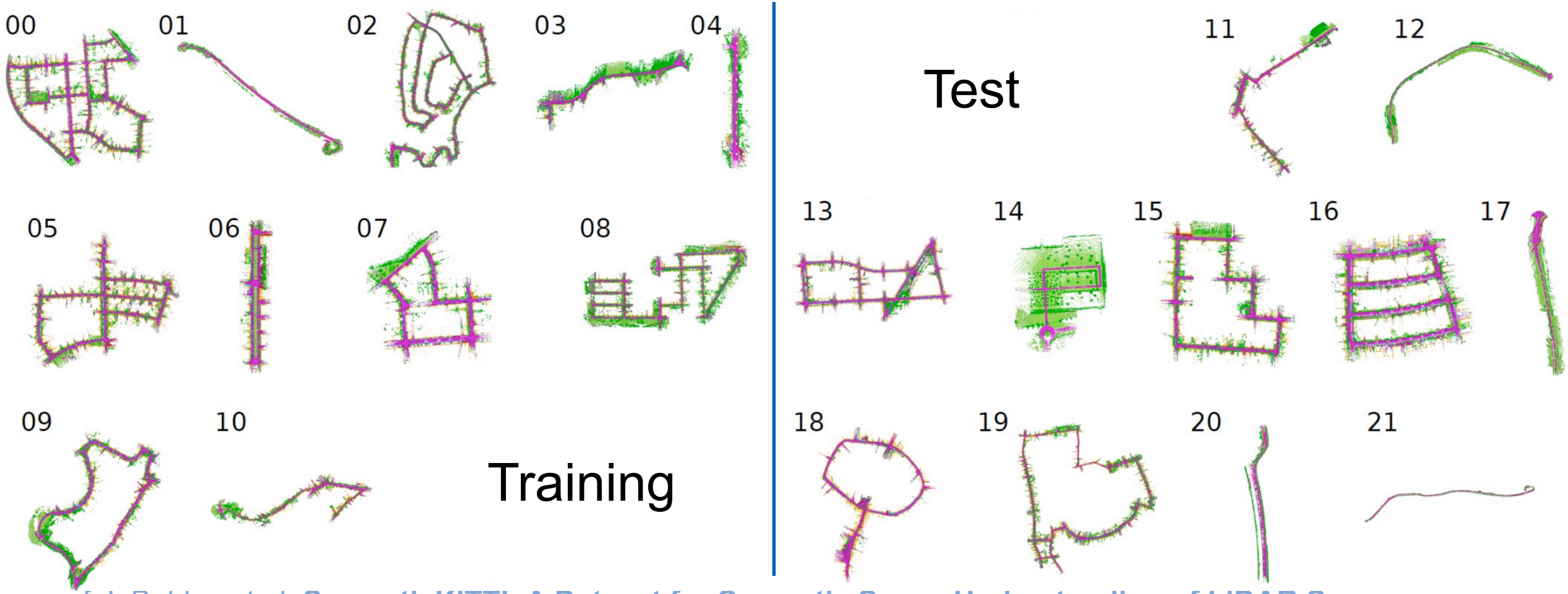


Annotation Process

- Instead of annotating single scans annotate entire regions
- SLAM with loop closure (manual correction in case of failure)

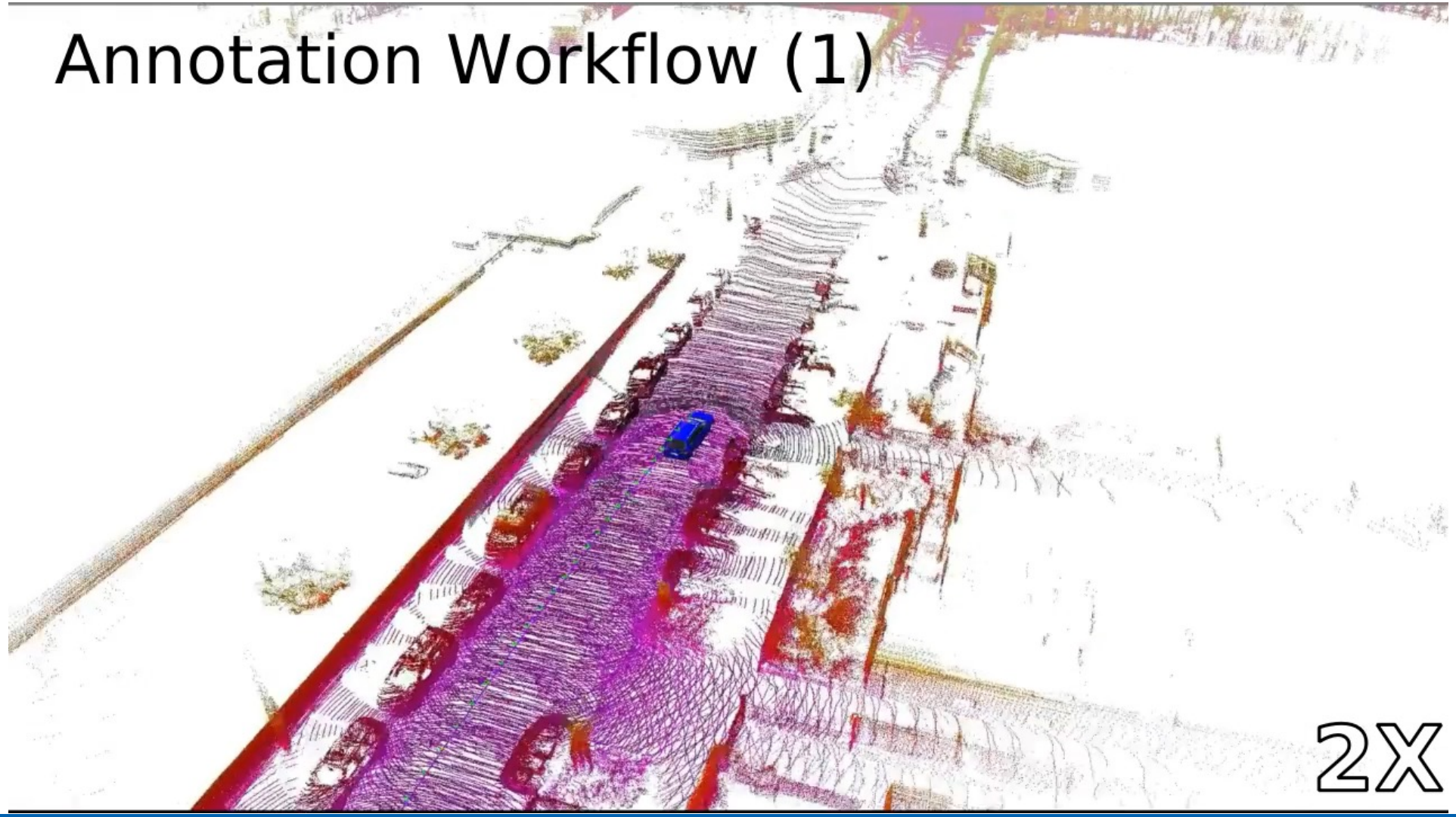


Data and Annotation

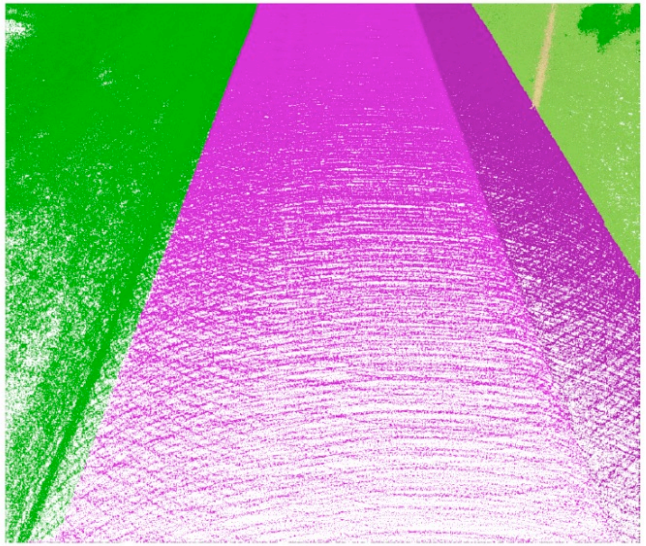


[J. Behley et al. **SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences.** ICCV 2019; International Journal of Robotics Research 2021]

Annotation Process

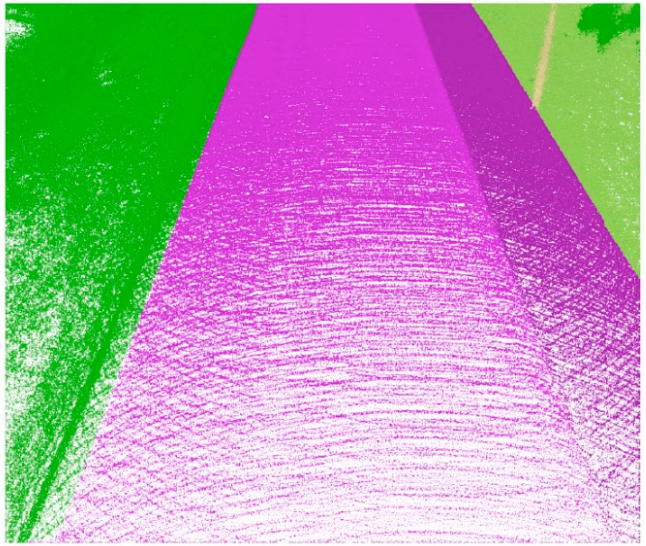


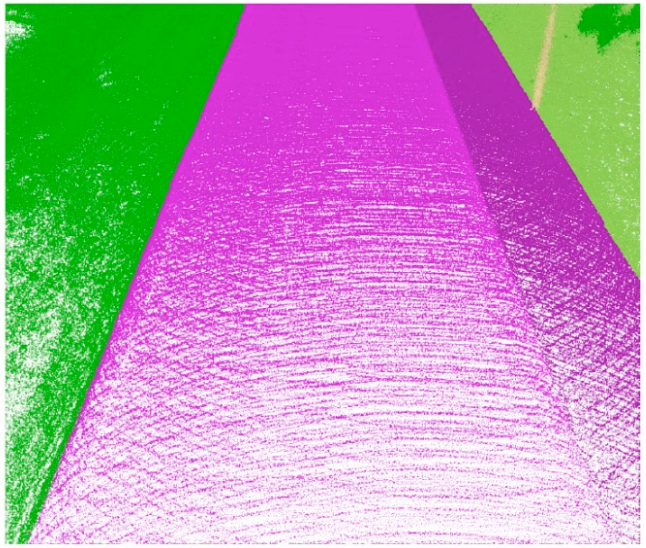
Verification Process



[J. Behley et al.
SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences.
ICCV 2019;
International Journal of Robotics Research 2021]

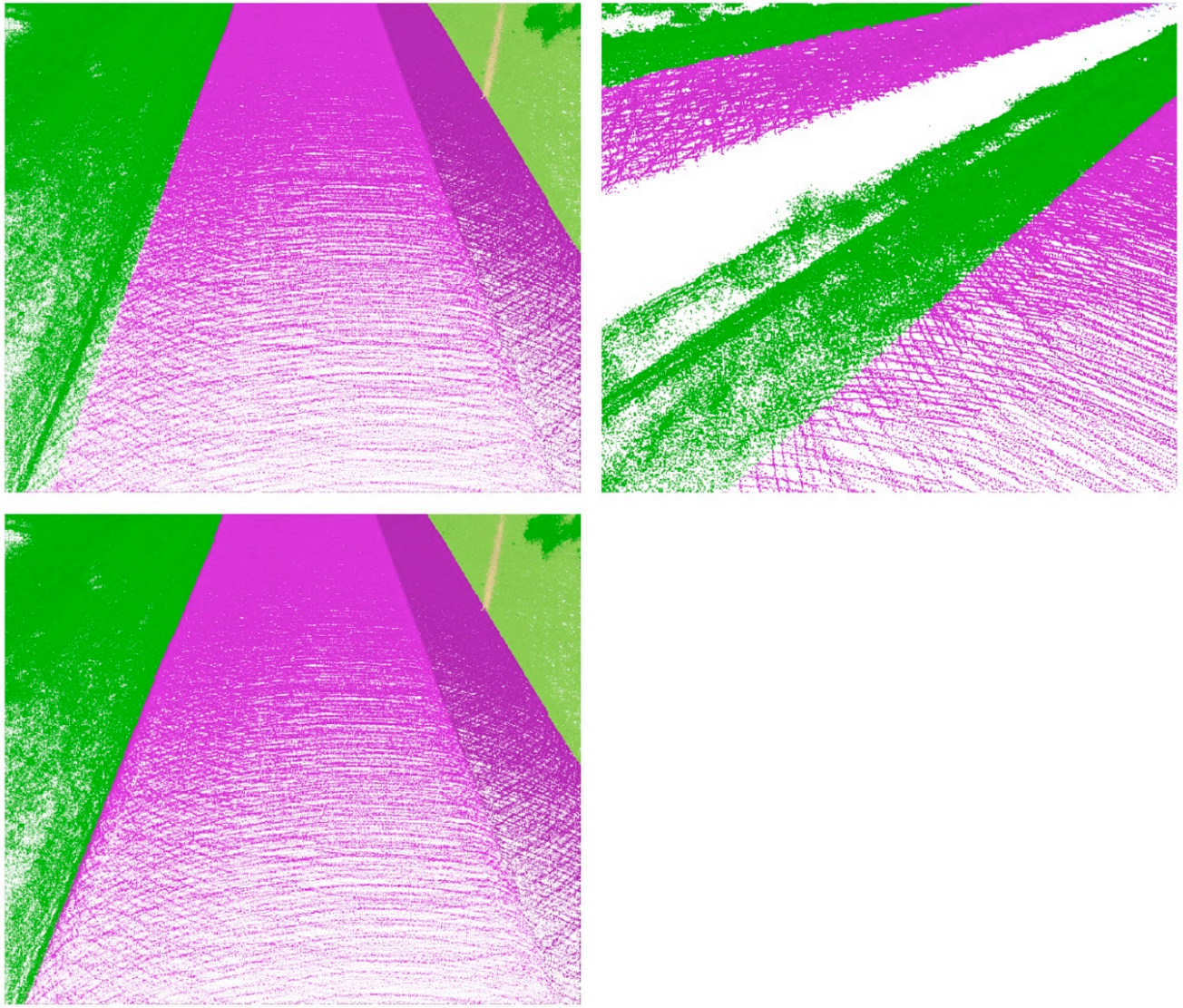
Verification Process





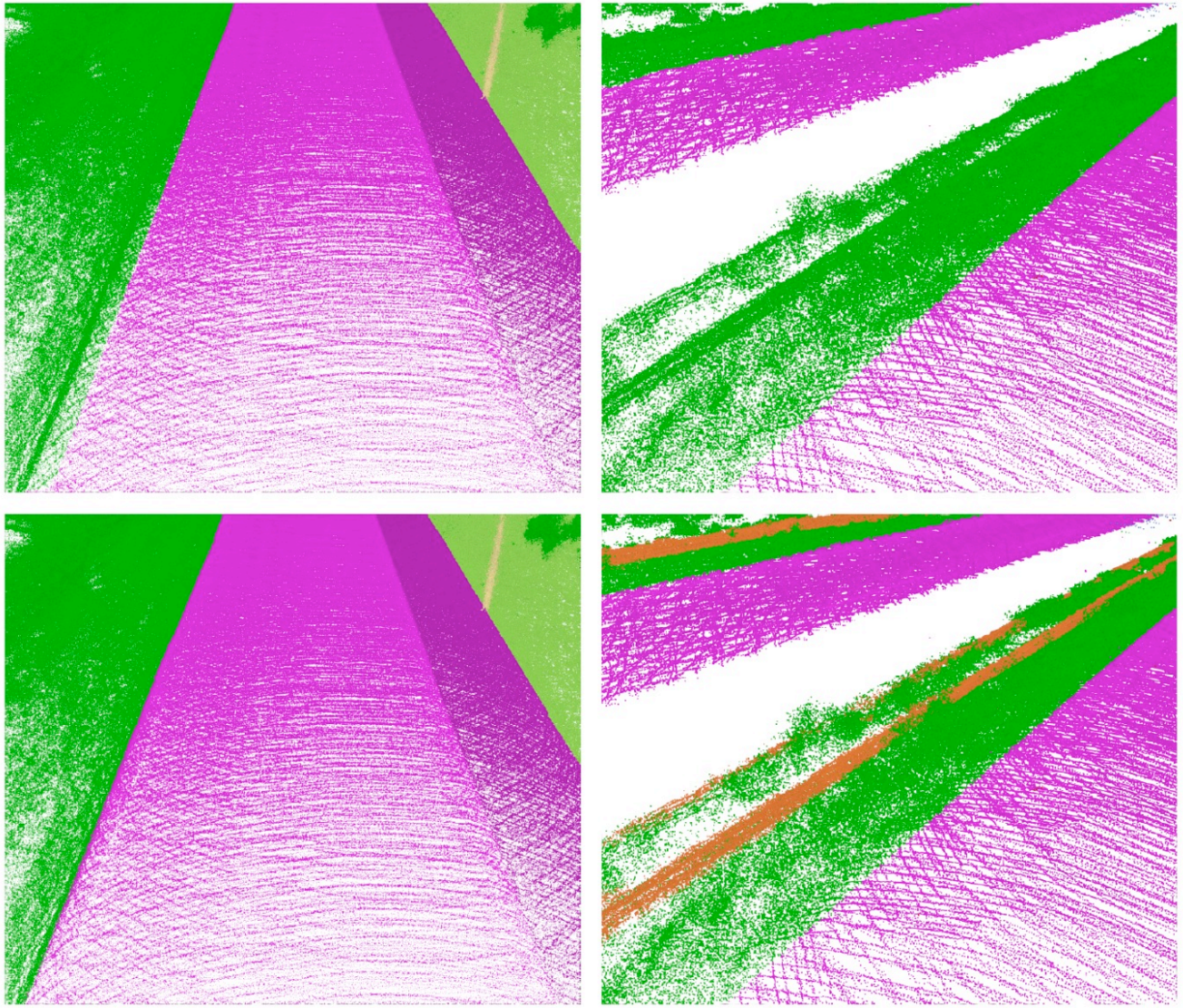
[J. Behley et al. **SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences.** ICCV 2019; International Journal of Robotics Research 2021]

Verification Process



[J. Behley et al. **SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences.** ICCV 2019; International Journal of Robotics Research 2021]

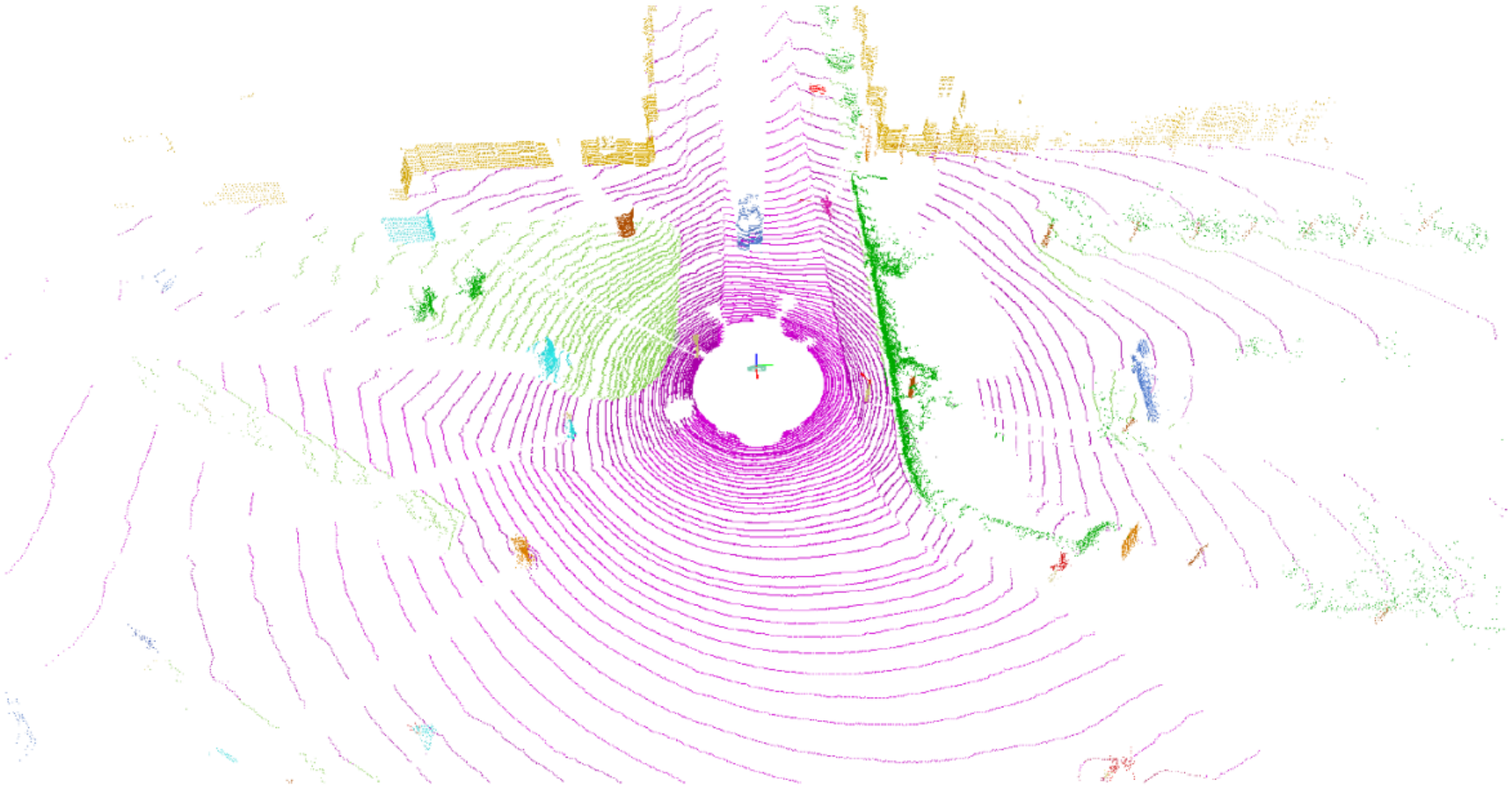
Verification Process



[J. Behley et al. **SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences.** ICCV 2019; International Journal of Robotics Research 2021]

Tasks: Semantic Segmentation

- Single scan
- Sequence of 5 scans



[J. Behley et al. **SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences.** ICCV 2019; International Journal of Robotics Research 2021]

Semantic Segmentation

- Accuracy is important (SemanticKitti benchmark)

Single Scan | Multiple Scans

Approach	Paper	Code	mIoU	Classes (IoU)	Details
2DPASS			72.9		
Point-Voxel-KD			71.2		
RPVNet			70.3		
AF2S3Net			69.7		
Cylinder3D			67.8		
SPVNAS			66.4		
JS3C-Net			66.0		
AMVNet			65.3		
Lite-HDSeg			63.8		

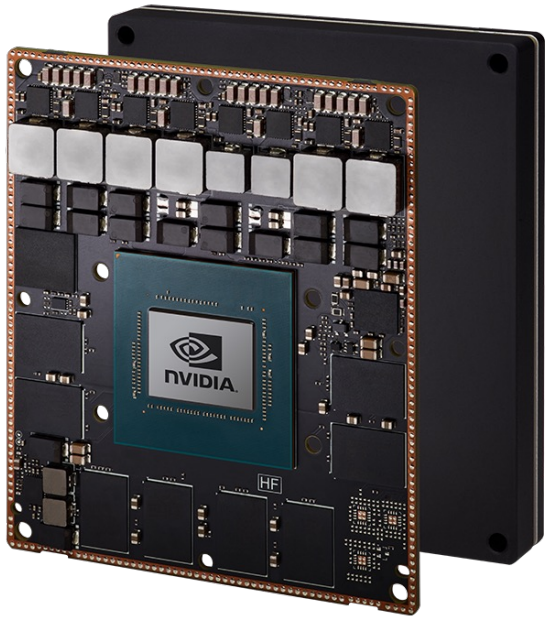
http://semantic-kitti.org

Semantic Segmentation

- Accuracy is important
- Runtime is important for autonomous driving

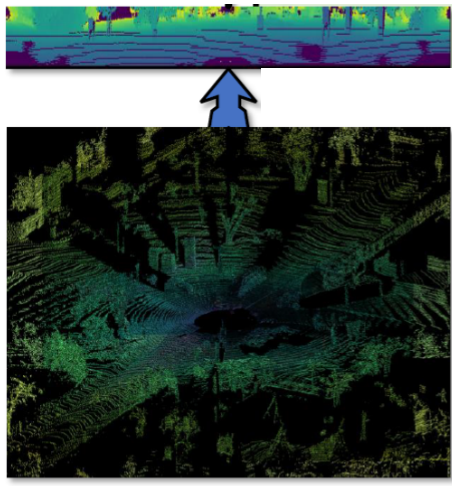
Semantic Segmentation

- Accuracy is important
- Runtime on an embedded system is important for autonomous driving



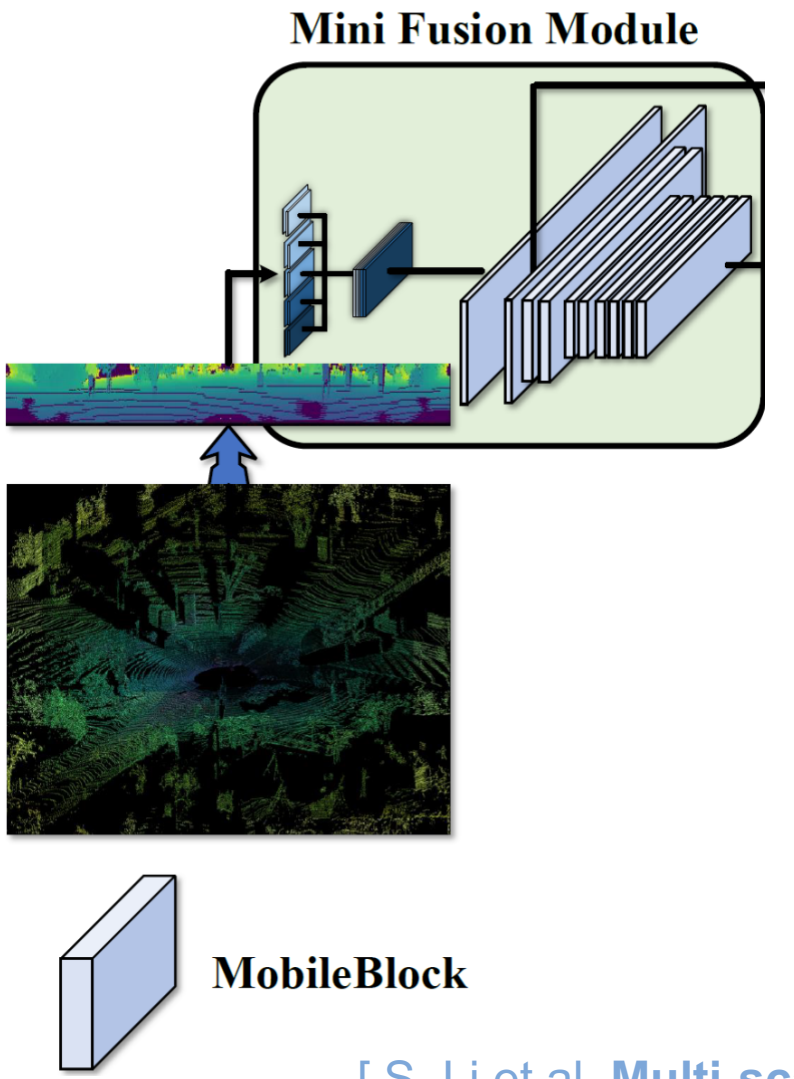
[S. Li et al. **Multi-scale Interaction for Real-time LiDAR Data Segmentation on an Embedded Platform.** IEEE Robotics and Automation Letters 2021]

Semantic Segmentation

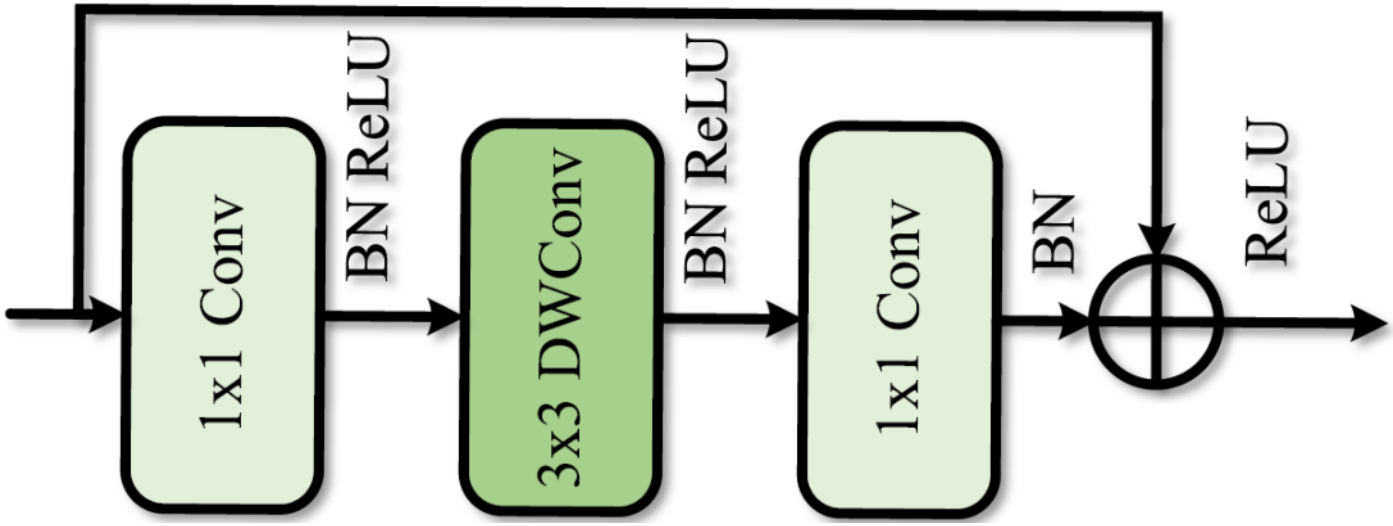


[S. Li et al. **Multi-scale Interaction for Real-time LiDAR Data Segmentation on an Embedded Platform.** IEEE Robotics and Automation Letters 2021]

Semantic Segmentation

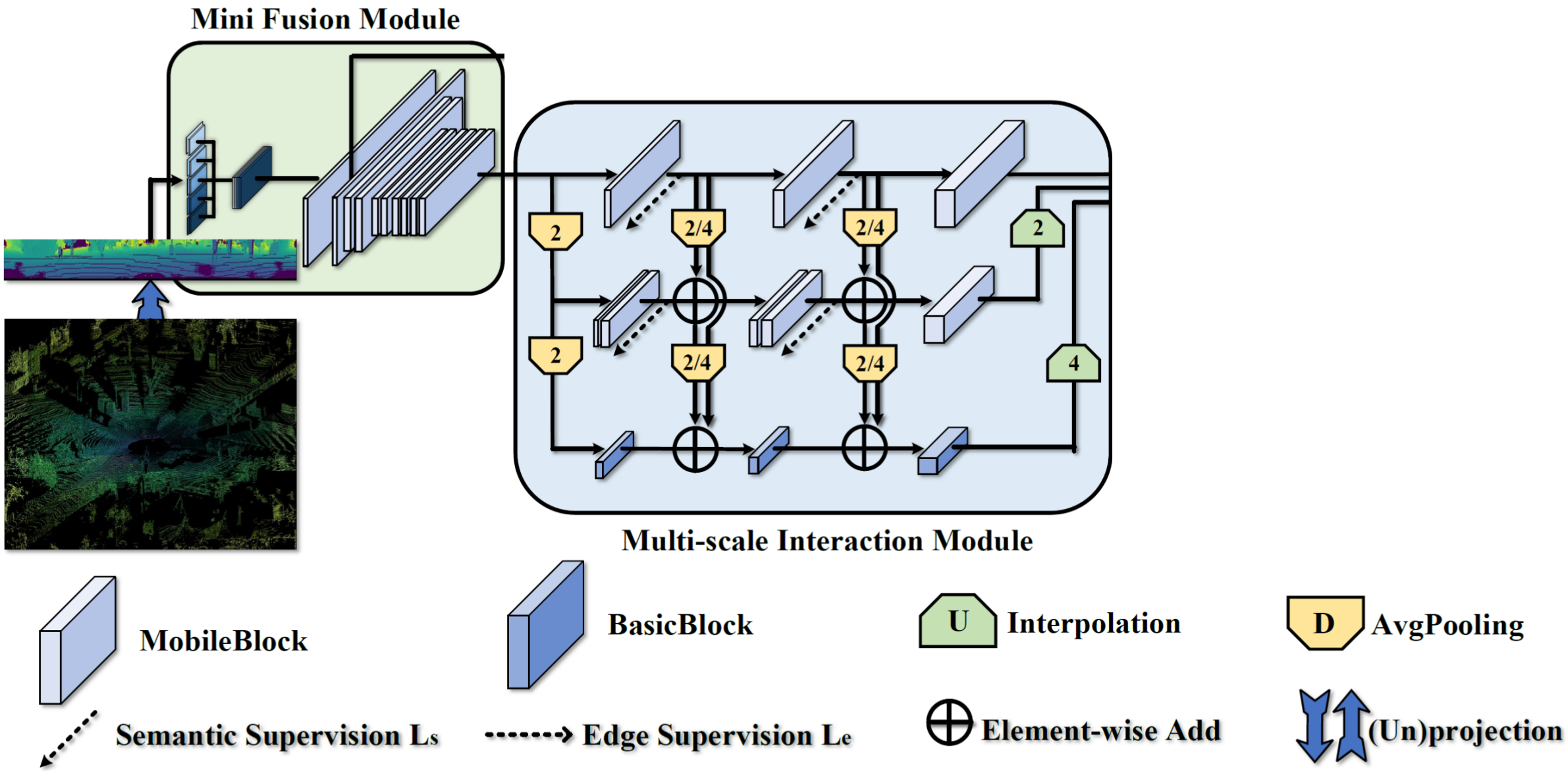


- 5 dimensional input (x, y, z, depth, remission)
- Each dimension is processed independently and then fused
- Uses 2D convolutions and MobileBlocks:

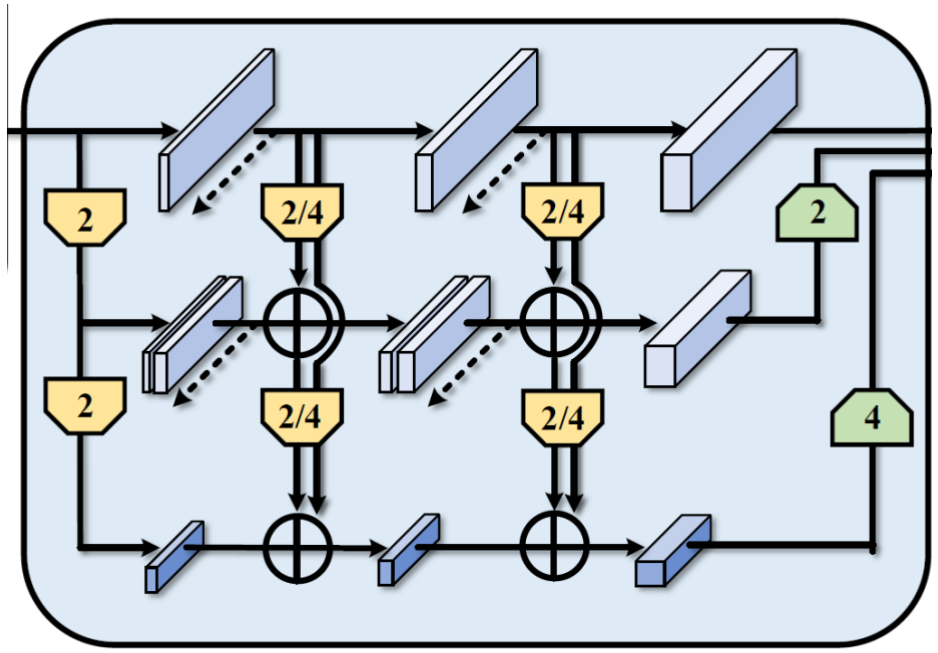


[S. Li et al. **Multi-scale Interaction for Real-time LiDAR Data Segmentation on an Embedded Platform.** IEEE Robotics and Automation Letters 2021]

Semantic Segmentation

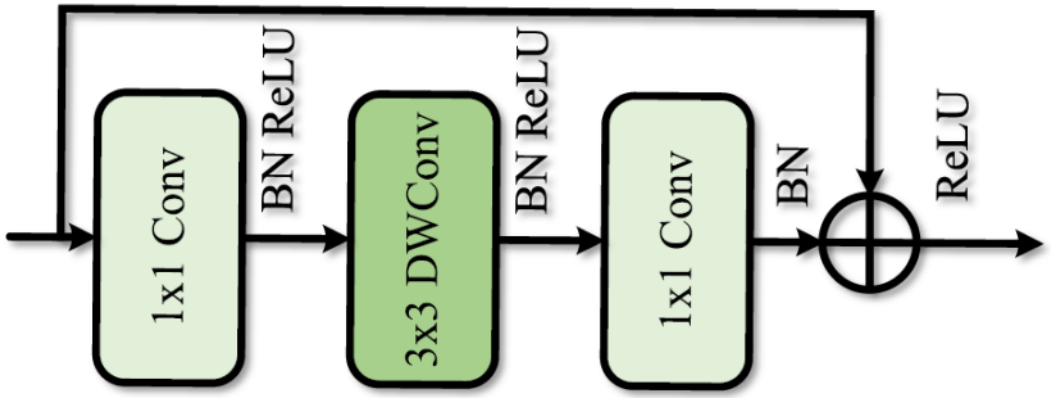


Semantic Segmentation

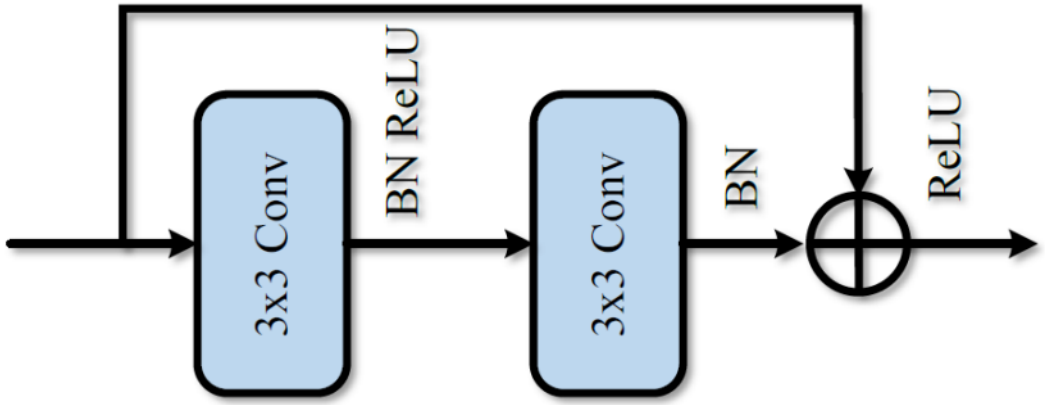


Multi-scale Interaction Module

MobileBlock (computational cost: low)



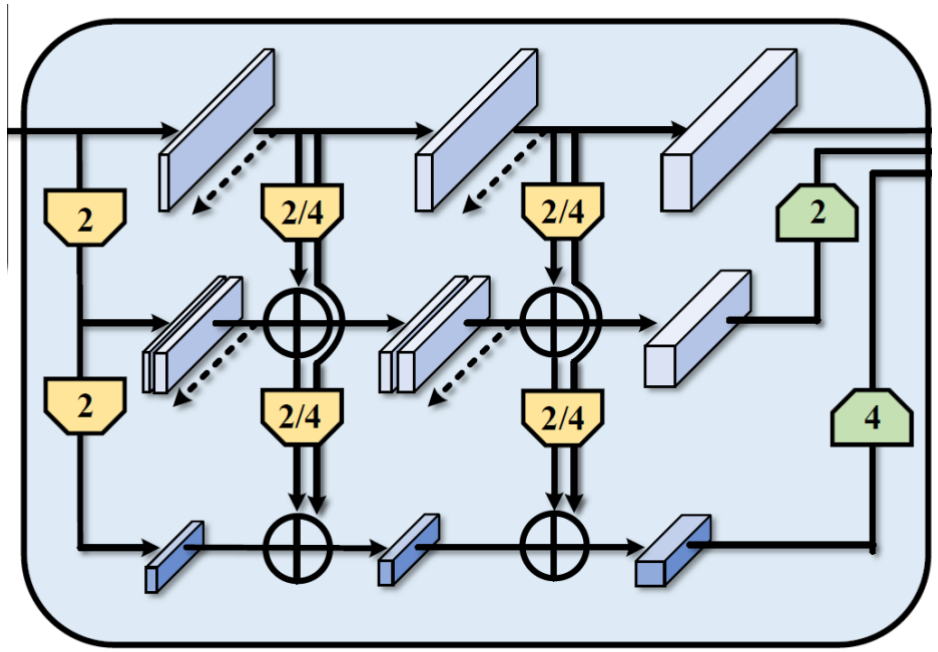
BasicBlock: (computational cost: high)



[S. Li et al. **Multi-scale Interaction for Real-time LiDAR Data Segmentation on an Embedded Platform.** IEEE Robotics and Automation Letters 2021]

Semantic Segmentation

- SPS (scans per second on GPU)



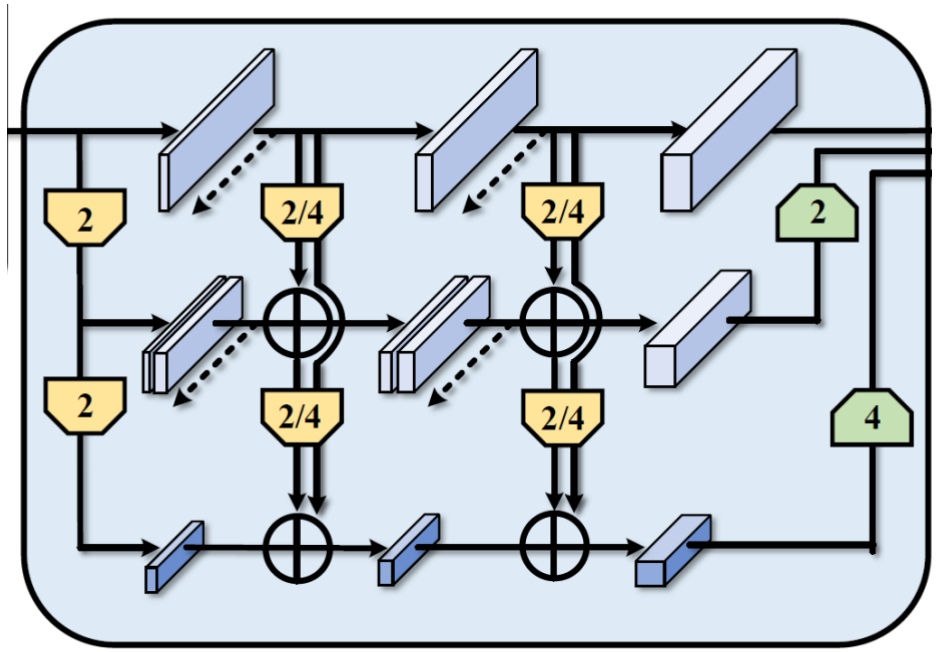
Multi-scale Interaction Module

Top	Middle	Bottom	Param (M)	GFLOP	SPS	mIoU
3×MB	3×MB	3×MB	0.6	6.00	61	49.7

[S. Li et al. **Multi-scale Interaction for Real-time LiDAR Data Segmentation on an Embedded Platform.** IEEE Robotics and Automation Letters 2021]

Semantic Segmentation

- SPS (scans per second on GPU)



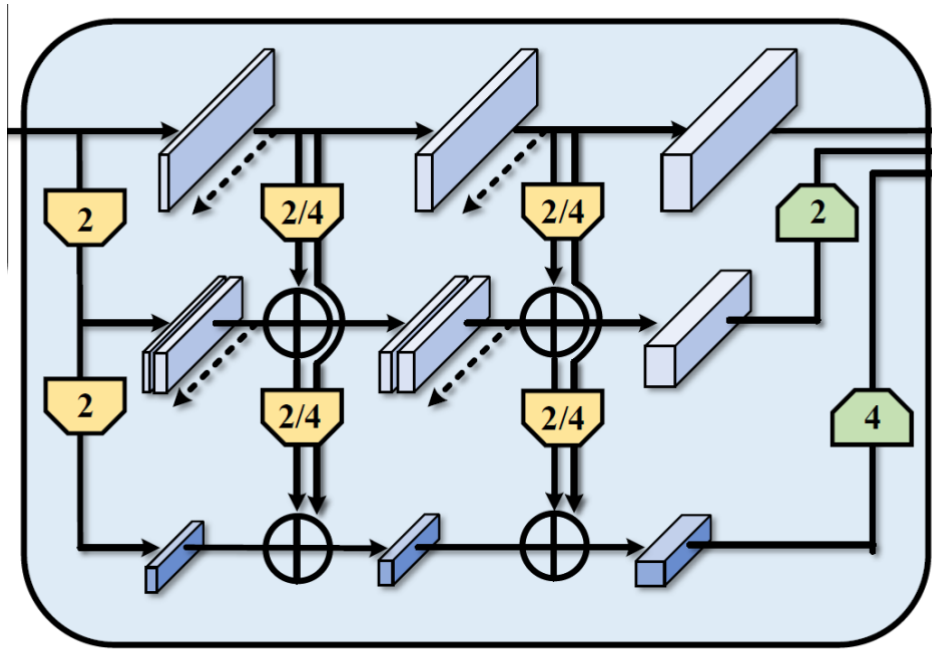
Multi-scale Interaction Module

Top	Middle	Bottom	Param (M)	GFLOP	SPS	mIoU
3×MB	3×MB	3×MB	0.6	6.00	61	49.7
5×MB	5×MB	5×MB	0.6	6.30	58	50.2

[S. Li et al. **Multi-scale Interaction for Real-time LiDAR Data Segmentation on an Embedded Platform.** IEEE Robotics and Automation Letters 2021]

Semantic Segmentation

- SPS (scans per second on GPU)



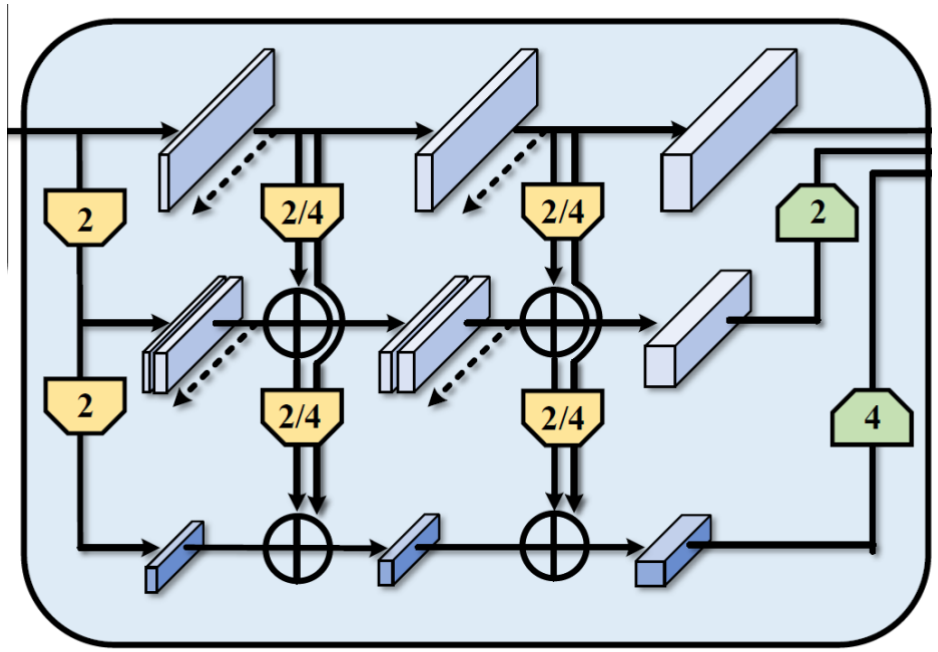
Multi-scale Interaction Module

Top	Middle	Bottom	Param (M)	GFLOP	SPS	mIoU
3×MB	3×MB	3×MB	0.6	6.00	61	49.7
5×MB	5×MB	5×MB	0.6	6.30	58	50.2
3×BB	3×BB	3×BB	2.0	11.04	33	51.2

[S. Li et al. **Multi-scale Interaction for Real-time LiDAR Data Segmentation on an Embedded Platform.** IEEE Robotics and Automation Letters 2021]

Semantic Segmentation

- SPS (scans per second on GPU)

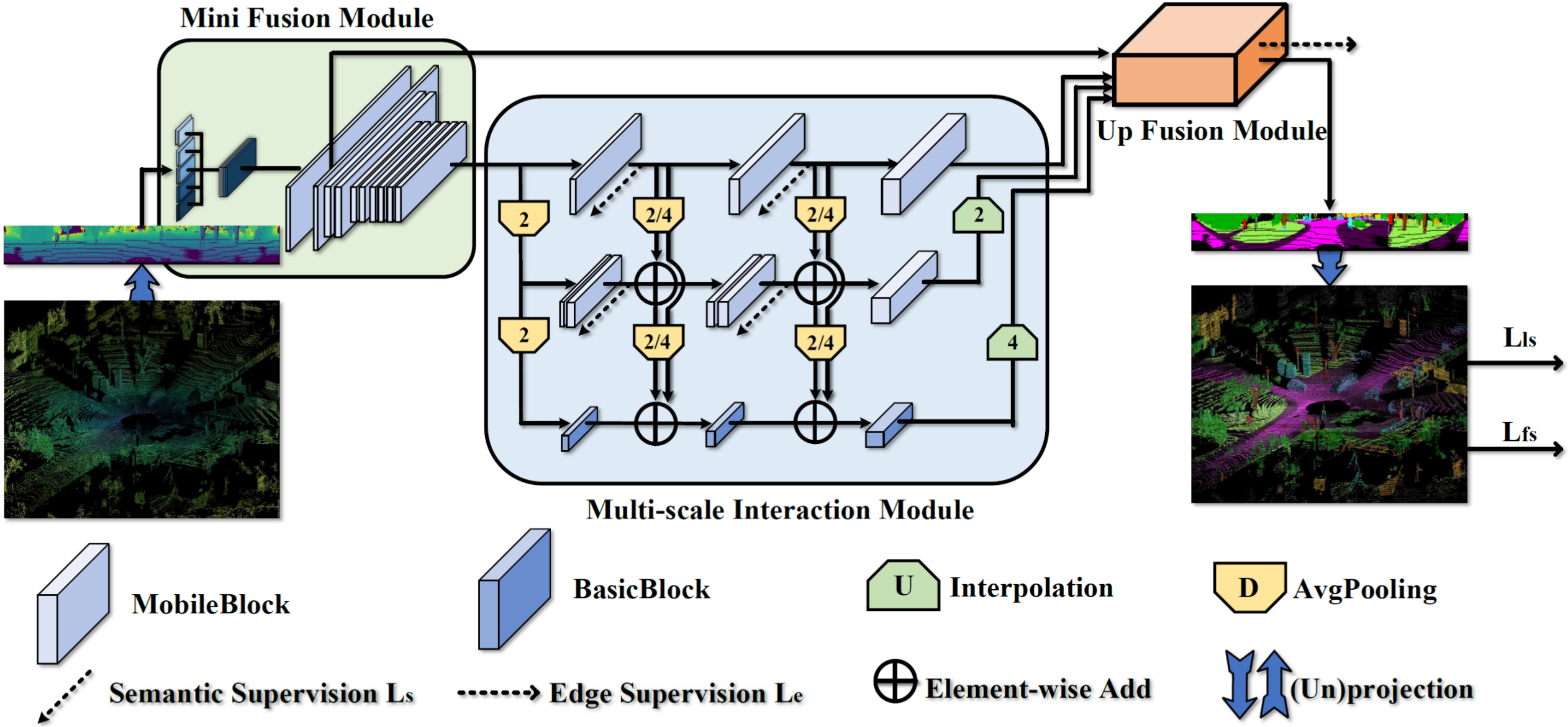


Multi-scale Interaction Module

Top	Middle	Bottom	Param (M)	GFLOP	SPS	mIoU
3×MB	5×MB	3×BB	1.0	6.20	59	51.8
3×MB	3×MB	3×MB	0.6	6.00	61	49.7
5×MB	5×MB	5×MB	0.6	6.30	58	50.2
3×BB	3×BB	3×BB	2.0	11.04	33	51.2

[S. Li et al. **Multi-scale Interaction for Real-time LiDAR Data Segmentation on an Embedded Platform.** IEEE Robotics and Automation Letters 2021]

Semantic Segmentation



Results on Jetson AGX

- Used TensorRT
- k-NN is a simple postprocessing
- Data captured with 10Hz

Methods	Resolutions	GFLOPs	SPS	mIoU
RangeNet53 [10]	64×2048	360.5	7	49.9
	64×1024	180.3	11	45.4
	64×512	90.1	22	39.3
RangeNet53 + k-NN [10]	64×2048	360.5	5	52.2
	64×1024	180.3	8	48.0
	64×512	90.1	13	41.9

Results on Jetson AGX

- Used TensorRT
- k-NN is a simple postprocessing
- Data captured with 10Hz

Methods	Resolutions	GFLOPs	SPS	mIoU
RangeNet53 [10]	64×2048	360.5	7	49.9
	64×1024	180.3	11	45.4
	64×512	90.1	22	39.3
RangeNet53 + k-NN [10]	64×2048	360.5	5	52.2
	64×1024	180.3	8	48.0
	64×512	90.1	13	41.9
MINet	64×2048	6.2	24	52.4
	64×1024	3.2	47	49.1
	64×512	1.7	80	45.0
MINet + k-NN	64×2048	6.2	13	55.2
	64×1024	3.2	18	52.4
	64×512	1.7	21	48.5

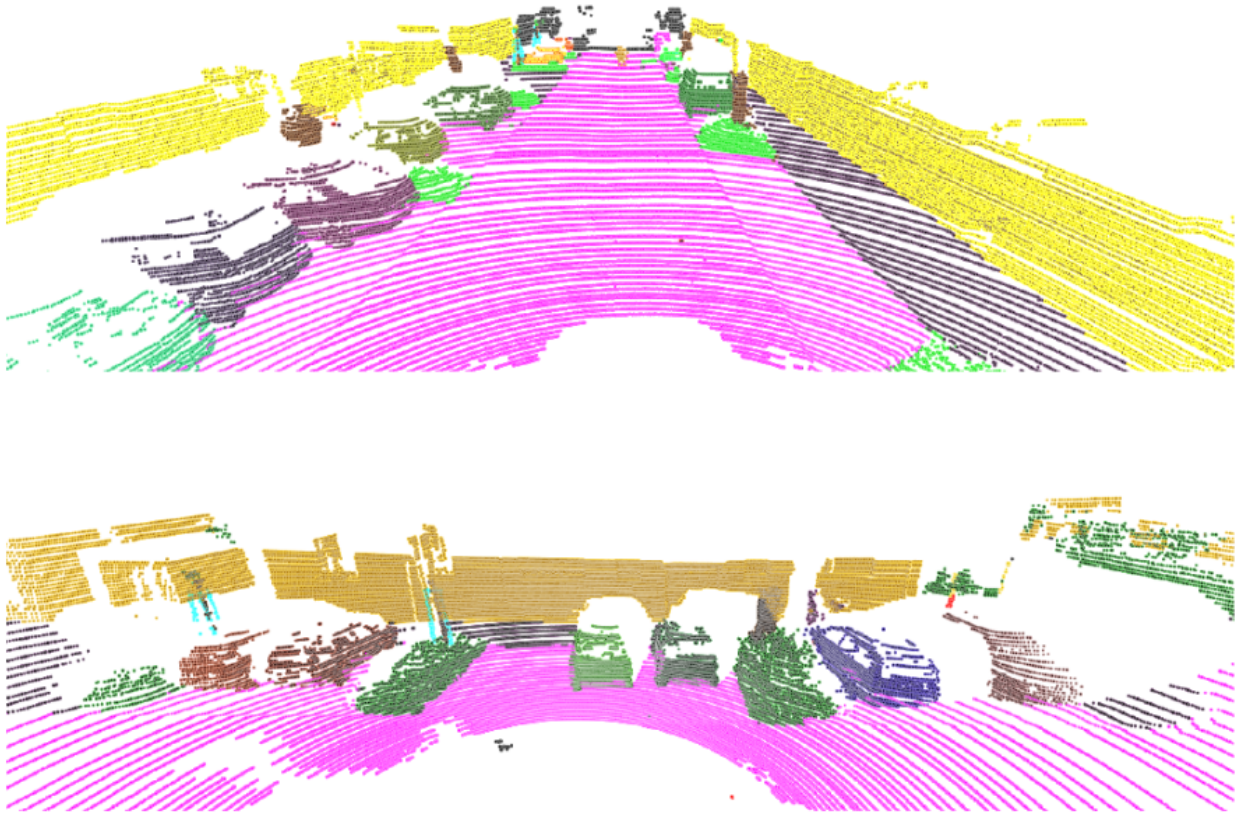
Results on Jetson AGX

- Used TensorRT
- k-NN is a simple postprocessing
- Data captured with 10Hz

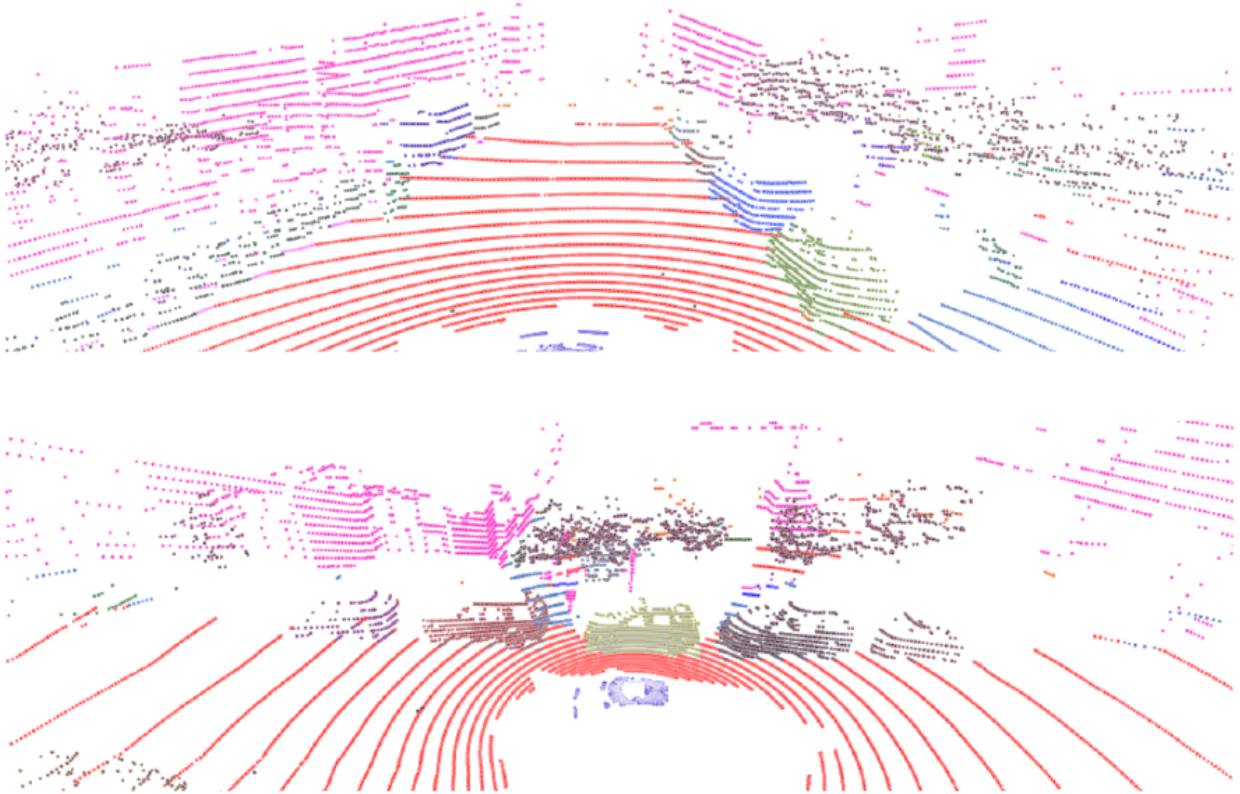
Methods	Resolutions	GFLOPs	SPS	mIoU
RangeNet53 [10]	64×2048	360.5	7	49.9
	64×1024	180.3	11	45.4
	64×512	90.1	22	39.3
RangeNet53 + k-NN [10]	64×2048	360.5	5	52.2
	64×1024	180.3	8	48.0
	64×512	90.1	13	41.9
MINet	64×2048	6.2	24	52.4
	64×1024	3.2	47	49.1
	64×512	1.7	80	45.0
MINet + k-NN	64×2048	6.2	13	55.2
	64×1024	3.2	18	52.4
	64×512	1.7	21	48.5

Tasks: Panoptic Segmentation

SemanticKITTI

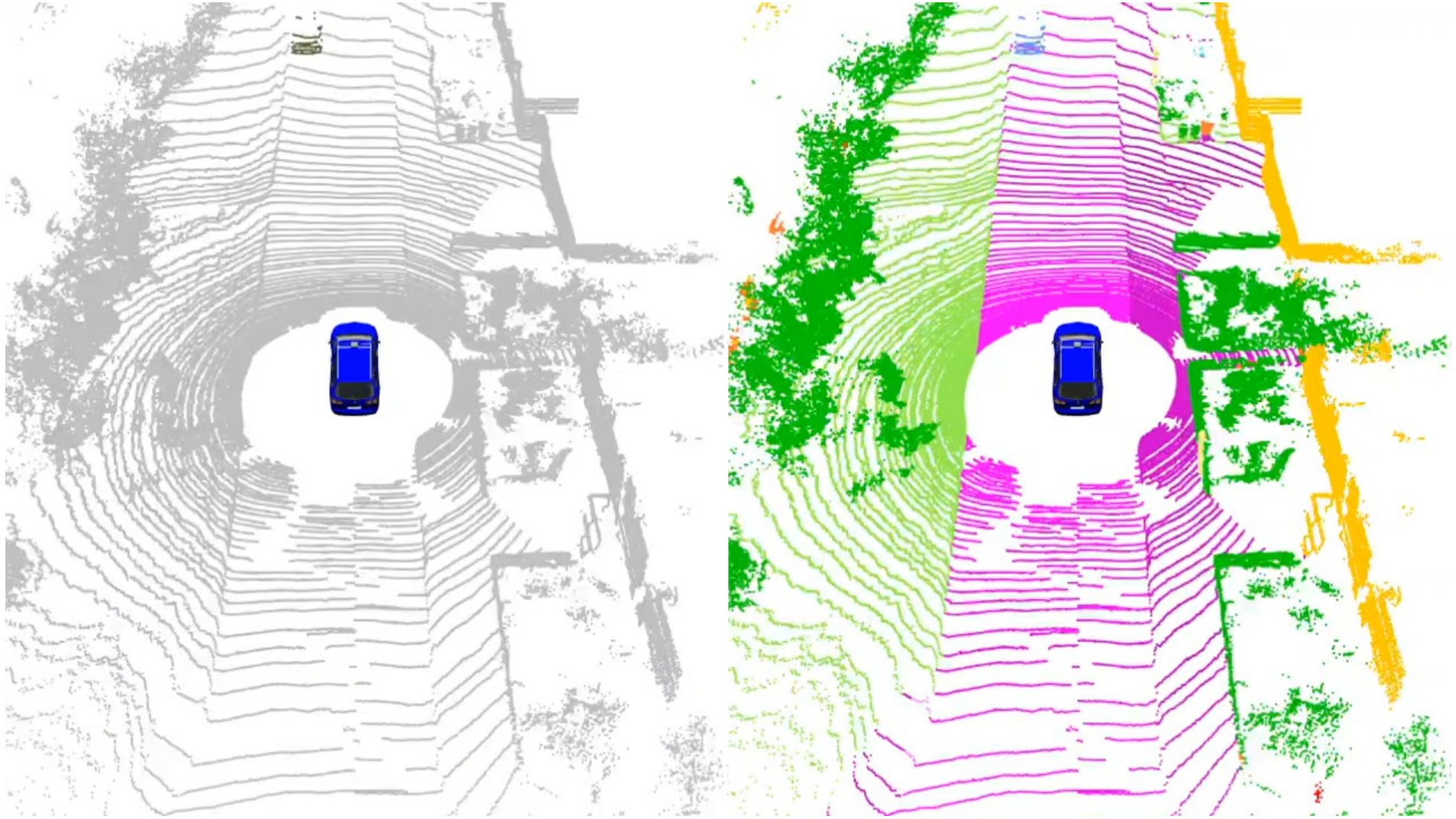


nuScenes



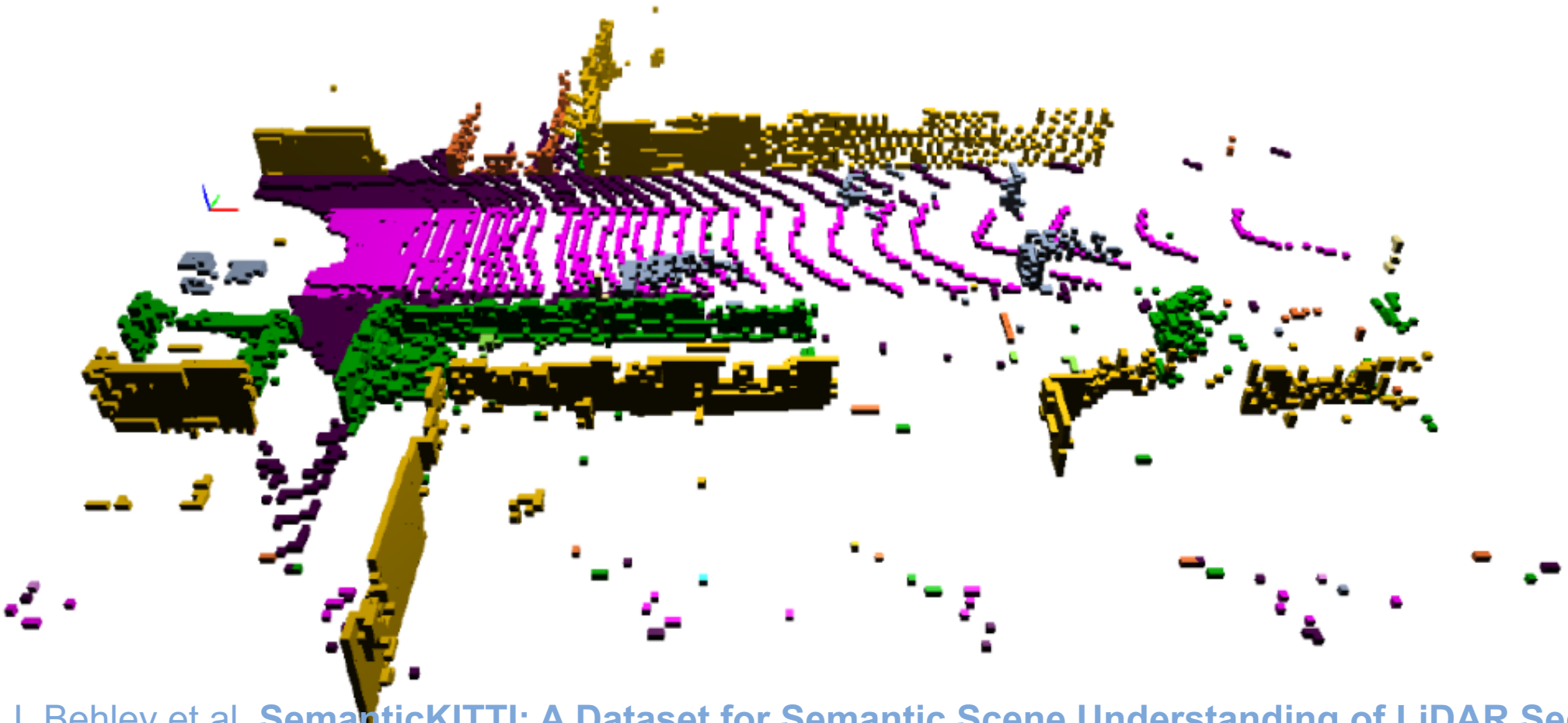
[K. Sirohi et al. **EfficientLPS: Efficient LiDAR Panoptic Segmentation.** IEEE Transactions on Robotics 2022]

Tasks: 4D Panoptic Segmentation



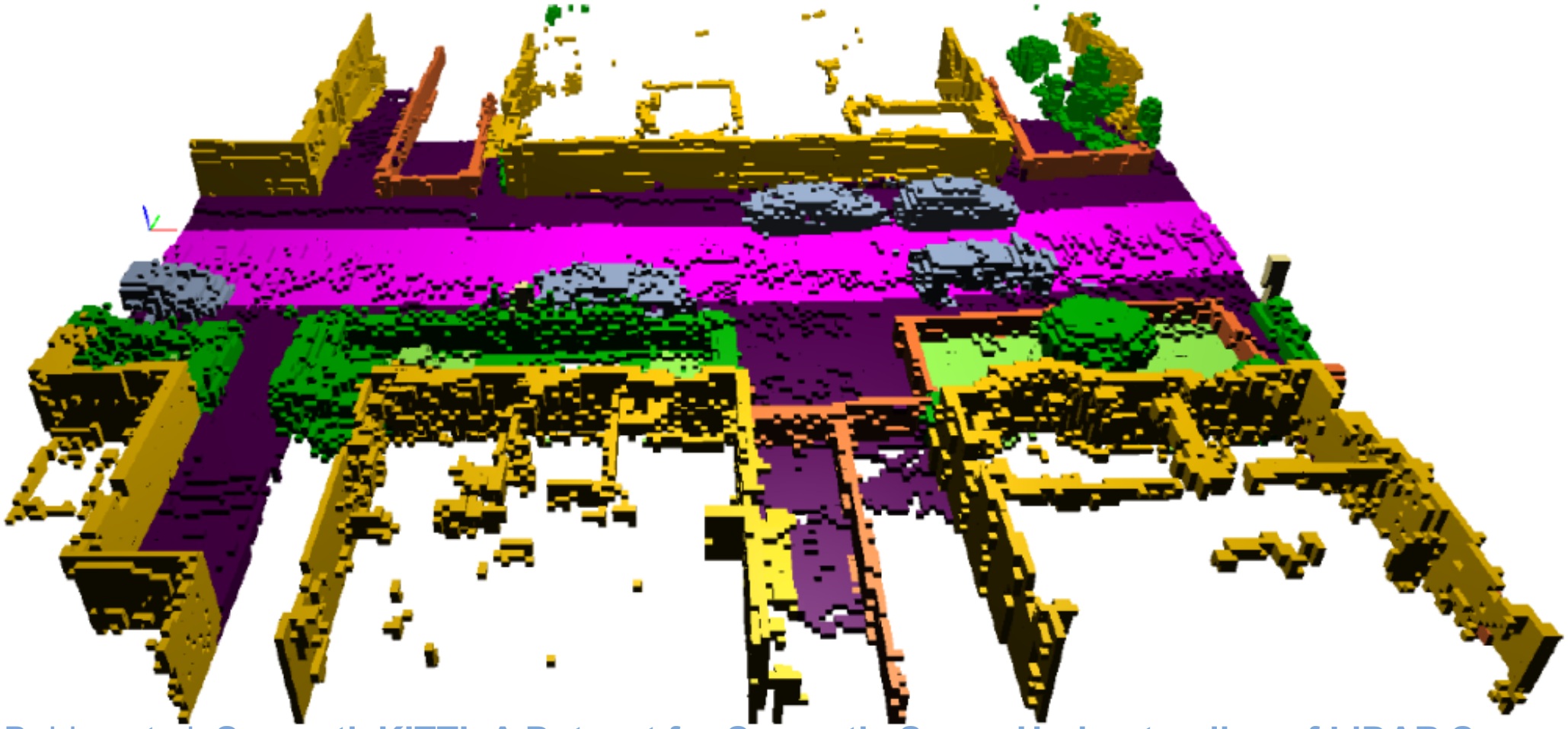
[M. Aygun et al. **4D Panoptic Lidar Segmentation.** CVPR 2021]

Tasks: Semantic Scene Completion



[J. Behley et al. **SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences.** ICCV 2019; International Journal of Robotics Research 2021]

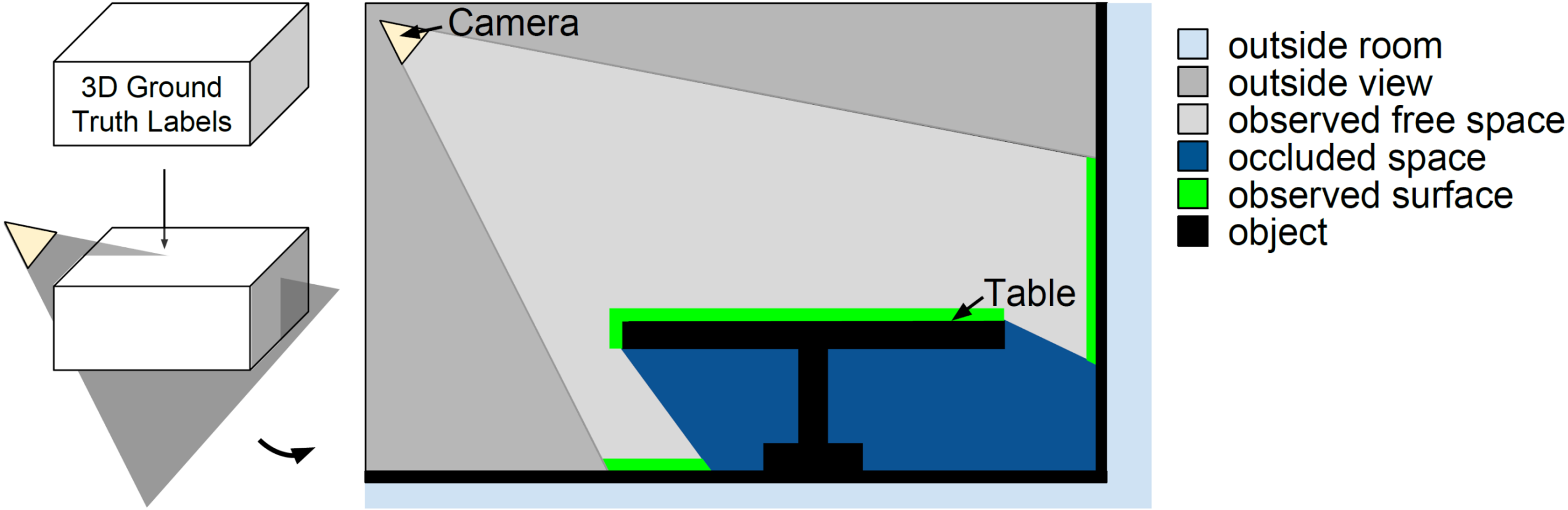
Tasks: Semantic Scene Completion



[J. Behley et al. **SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences.** ICCV 2019; International Journal of Robotics Research 2021]

Semantic Scene Completion

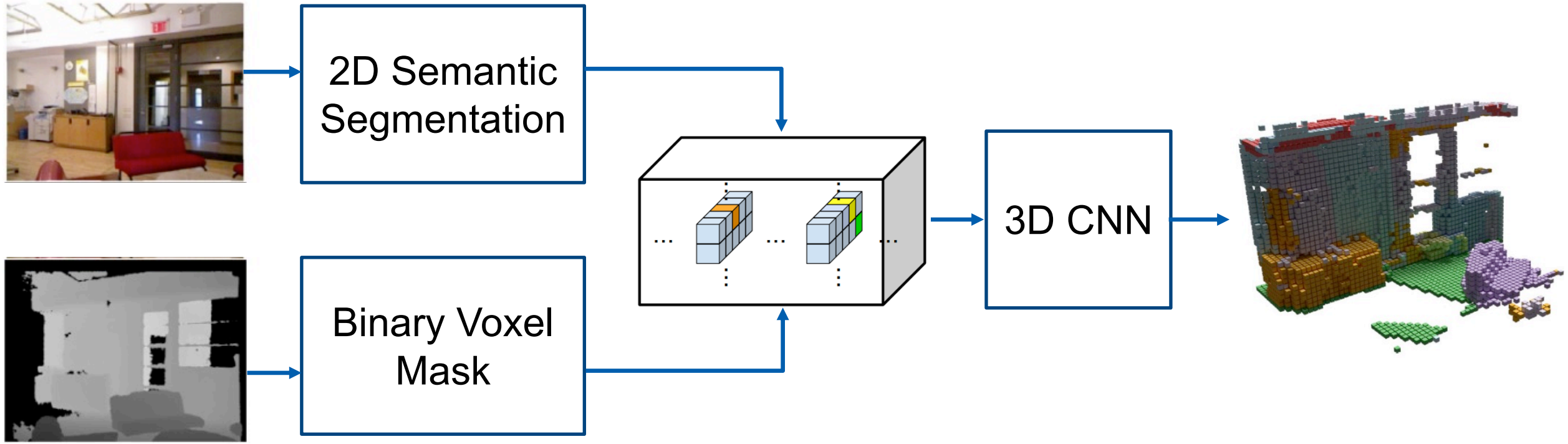
- Protocol defined by Song et al. (Indoor scenes)



[S. Song et al. **Semantic Scene Completion from a Single Depth Image.** CVPR 2017]

Semantic Scene Completion

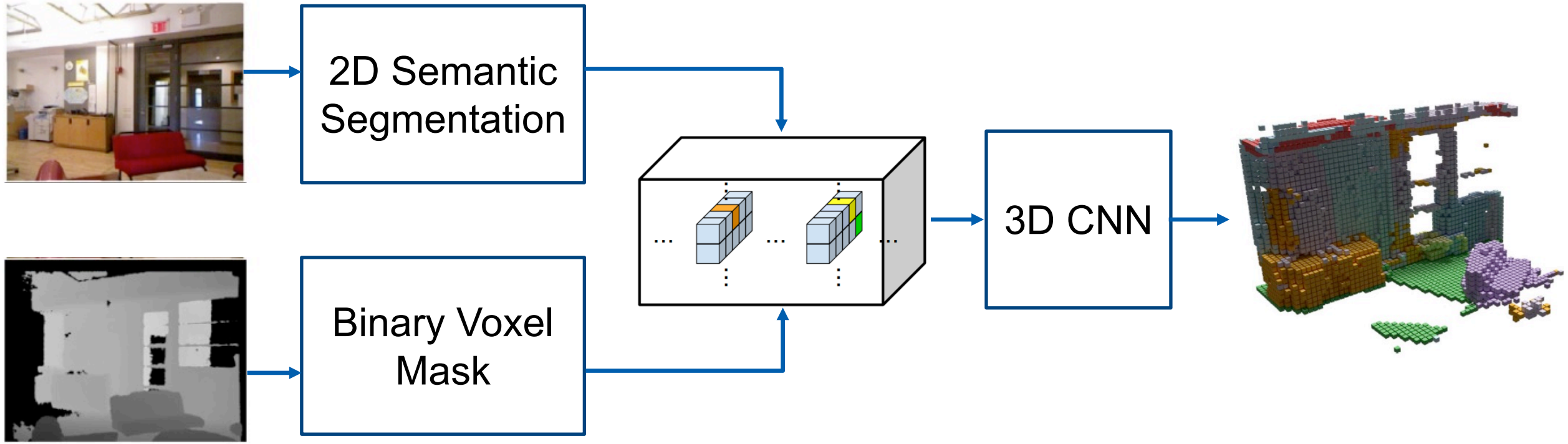
- Indoor: RGB + depth as input



[M. Garbade et al. **Two Stream 3D Semantic Scene Completion.** Multimodal Learning and Applications Workshop 2019]

Semantic Scene Completion

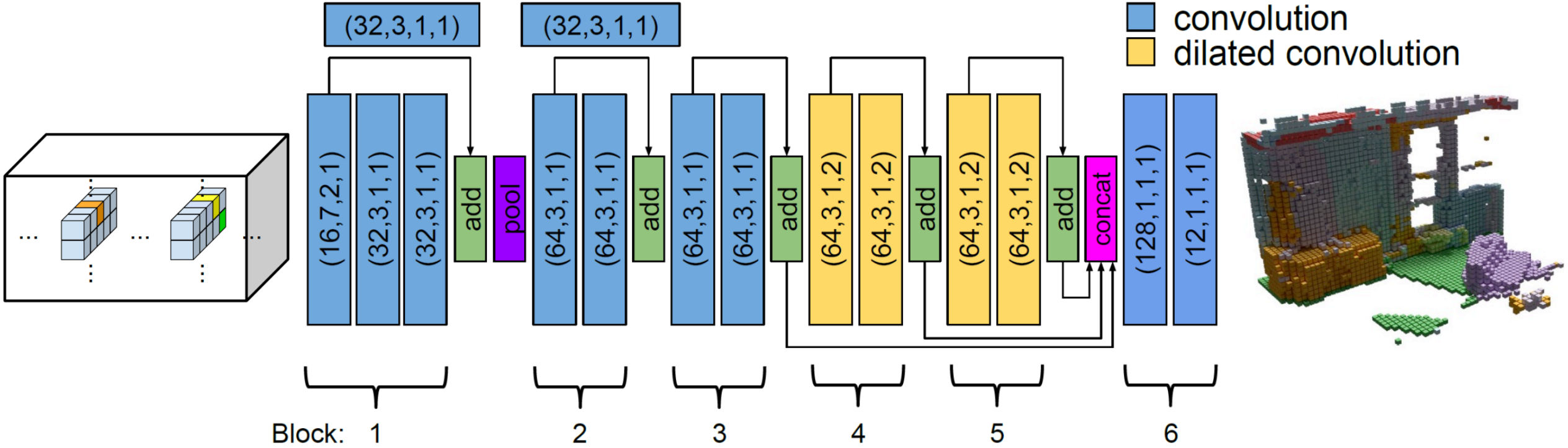
- Indoor: RGB + depth as input



[M. Garbade et al. **Two Stream 3D Semantic Scene Completion.** Multimodal Learning and Applications Workshop 2019]

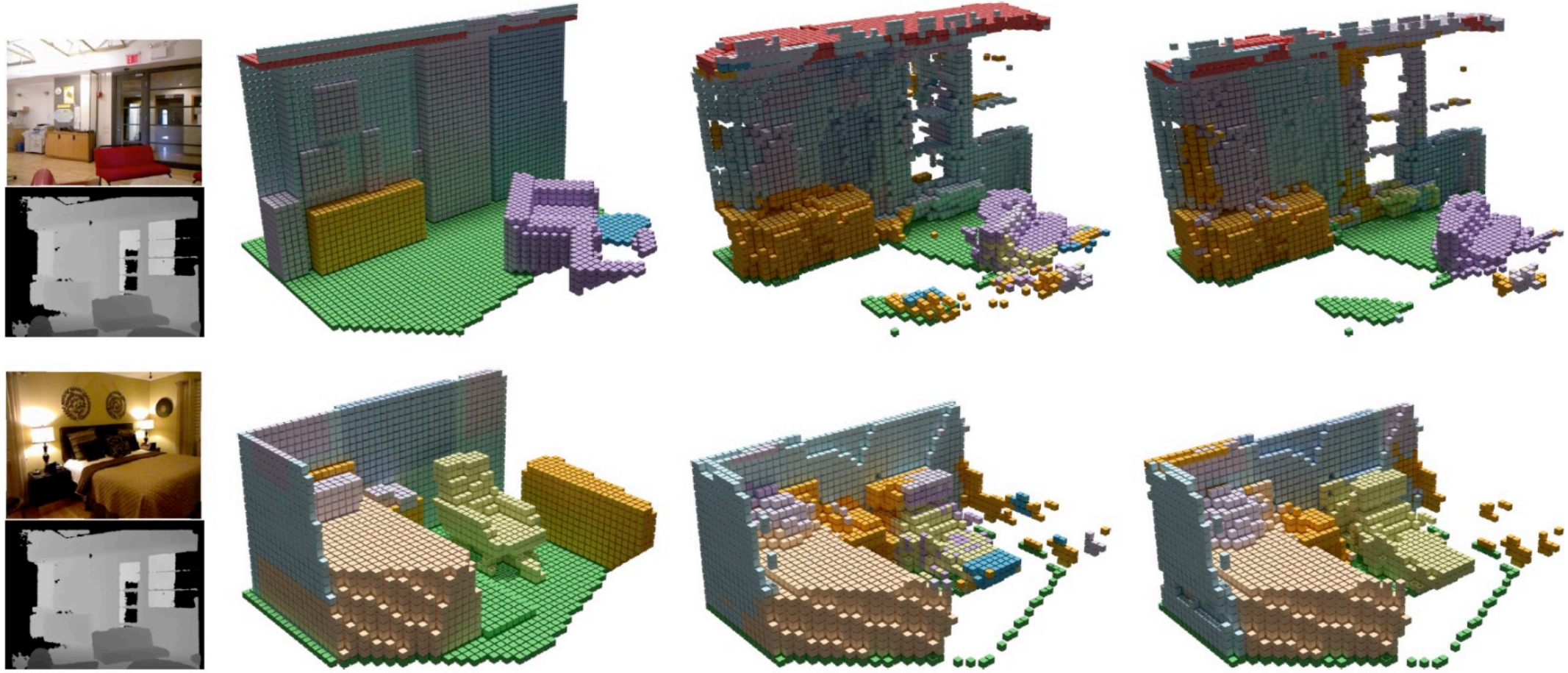
Semantic Scene Completion

- Indoor: RGB + depth as input



[M. Garbade et al. **Two Stream 3D Semantic Scene Completion.** Multimodal Learning and Applications Workshop 2019]

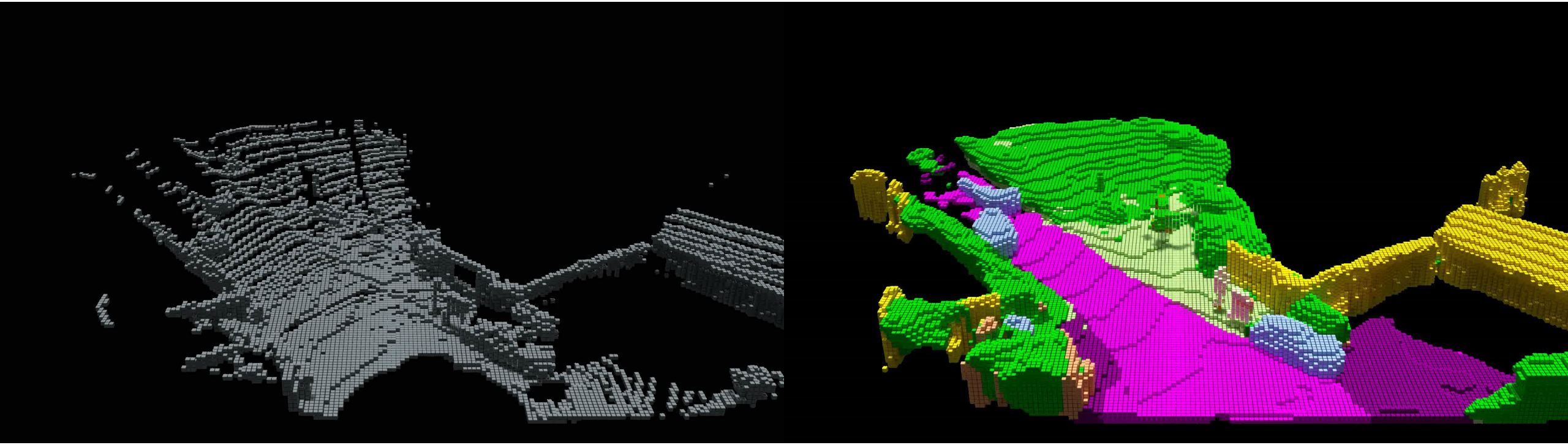
Semantic Scene Completion



Input | Ground Truth | Song et al. | Proposed

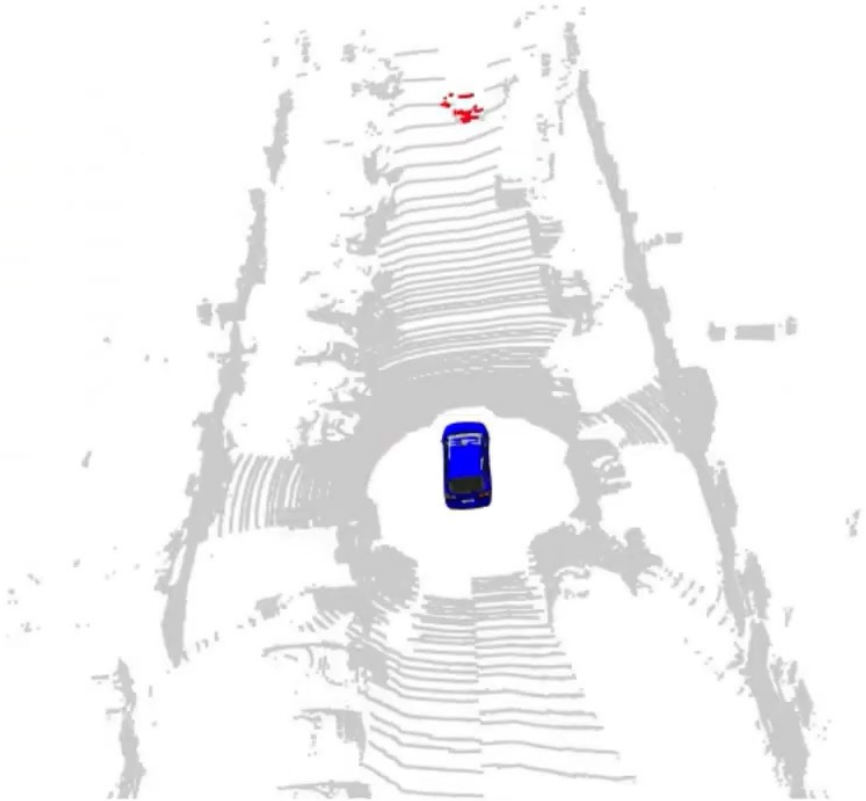
Semantic Scene Completion

- Anticipate semantic geometry in front of car



[J. Behley et al. **SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences.** ICCV 2019]
[M. Garbade et al. **Two Stream 3D Semantic Scene Completion.** Multimodal Learning and Appl. Workshop 2019]

Task: Moving Object Segmentation



[X. Chen et al. **Moving Object Segmentation in 3D LiDAR Data: A Learning-based Approach Exploiting Sequential Data.** IEEE Robotics and Automation Letters 2021]

How to deal with rare examples?

[Rob Siegel;
https://www.hagerty.com]

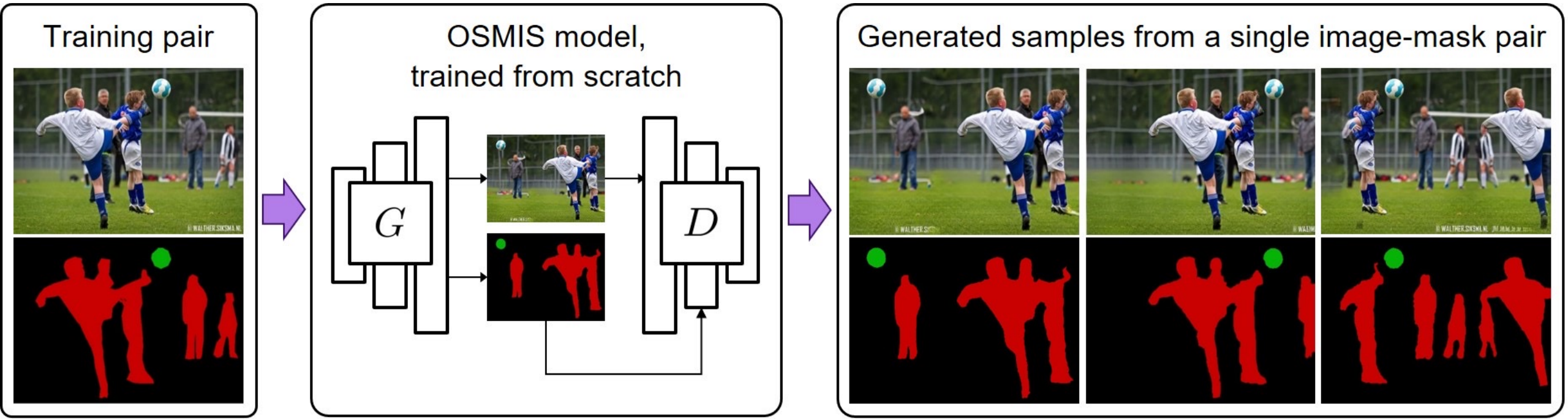
One-shot Image and Label Synthesis

- Given a single annotated training images

One-shot Image and Label Synthesis

- Given a single annotated training images
- How can you generate multiple annotated training images?

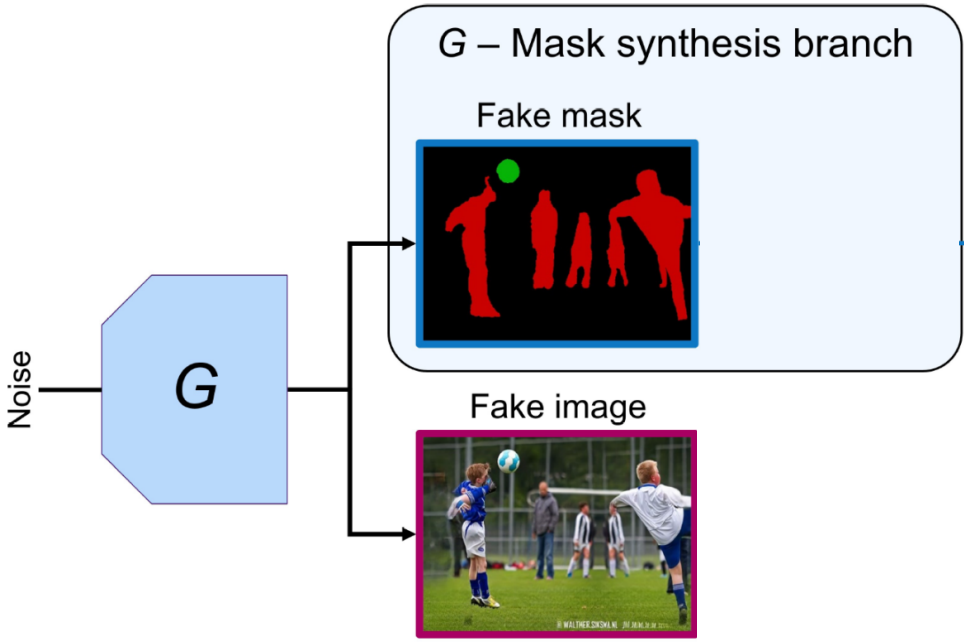
One-shot Image and Label Synthesis



OSMIS = One-**S**hot **M**ask and **I**mage **S**ynthesis

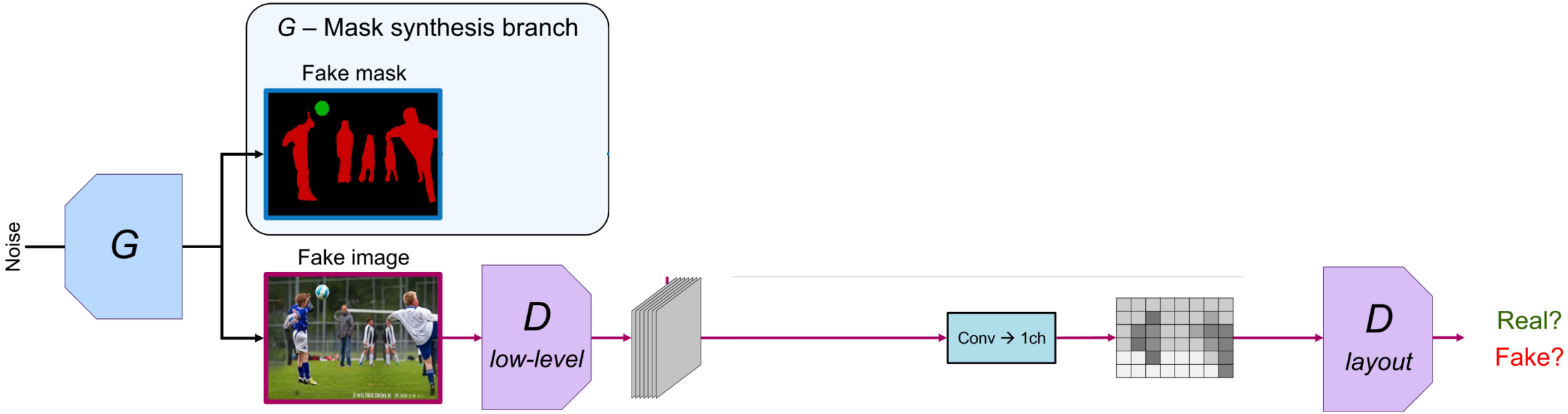
[V. Sushko et al. **One-Shot Synthesis of Images and Segmentation Masks.** WACV 2023]

One-shot Image and Label Synthesis



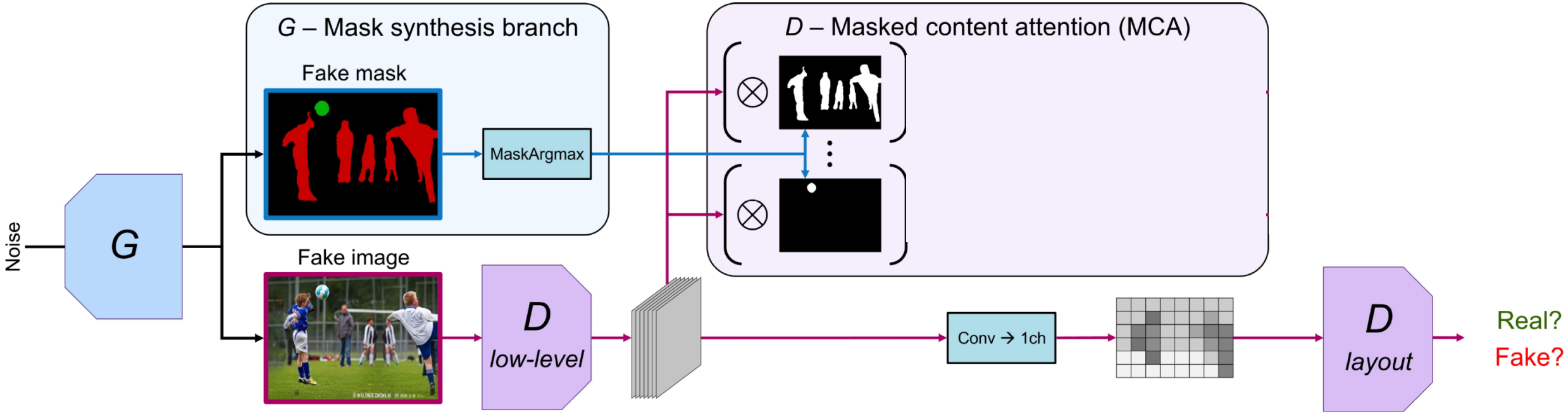
[V. Sushko et al. **One-Shot Synthesis of Images and Segmentation Masks.** WACV 2023]

One-shot Image and Label Synthesis



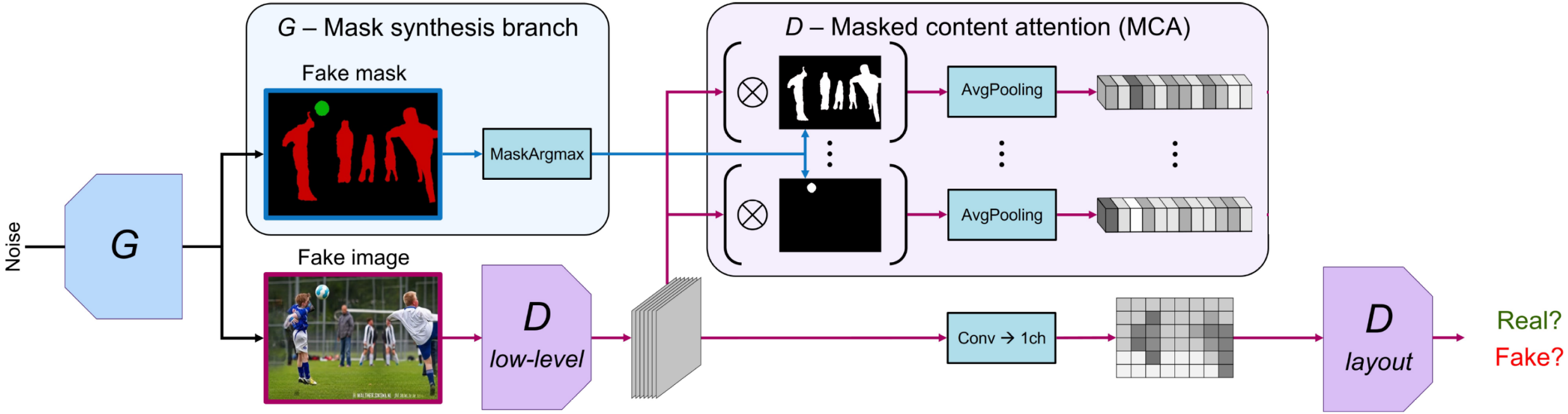
[V. Sushko et al. **One-Shot Synthesis of Images and Segmentation Masks.** WACV 2023]

One-shot Image and Label Synthesis



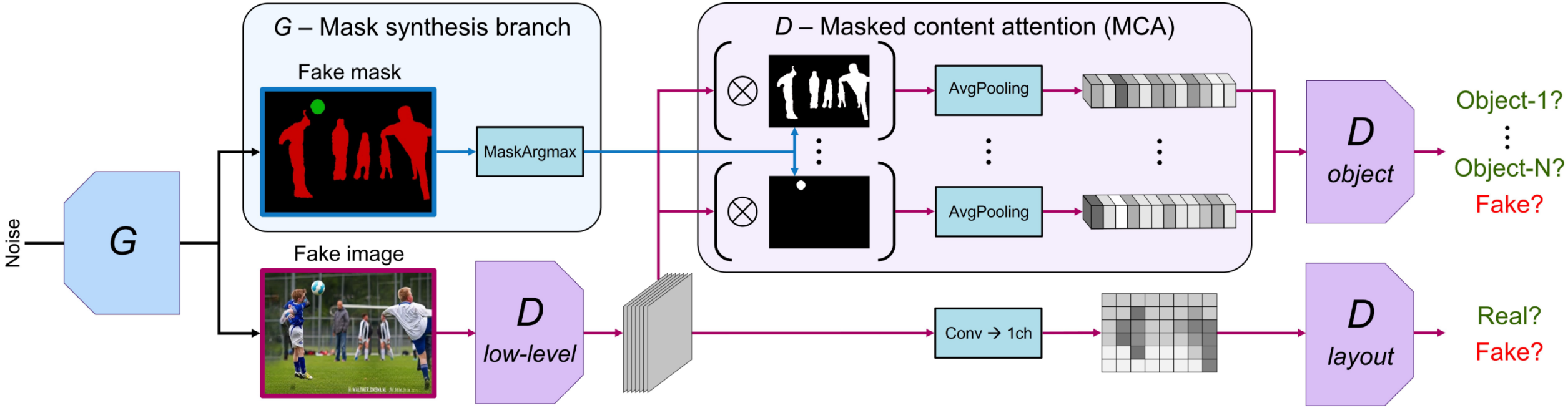
[V. Sushko et al. **One-Shot Synthesis of Images and Segmentation Masks.** WACV 2023]

One-shot Image and Label Synthesis



[V. Sushko et al. **One-Shot Synthesis of Images and Segmentation Masks.** WACV 2023]

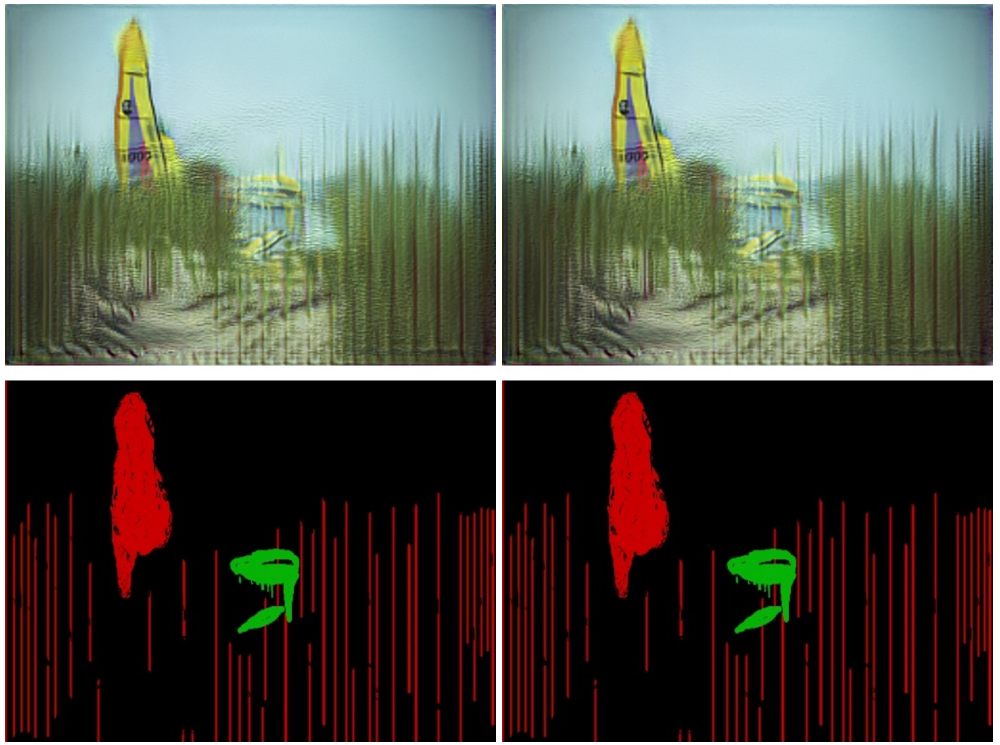
One-shot Image and Label Synthesis



[V. Sushko et al. **One-Shot Synthesis of Images and Segmentation Masks.** WACV 2023]

Previous Works

Training pair



SemanticGAN
[Li et al., CVPR21]

DatasetGAN*
[Zhang et al., CVPR21]

One-shot Image and Label Synthesis

Training pair

Proposed

[V. Sushko et al. **One-Shot Synthesis of Images and Segmentation Masks.** WACV 2023]

Thank you for your attention!

European Research Council

Deutsche Forschungsgemeinschaft

LAMARR
Institute for Machine Learning and Artificial Intelligence

PHENOROB

UNIVERSITÄT BONN