3D Reconstruction Meets Semantics

Reconstruction Challenge 2018

Radim Tyleček, Torsten Sattler,
Thomas Brox, Marc Pollefeys, Robert B. Fisher, Theo Gevers

EU project TrimBot2020

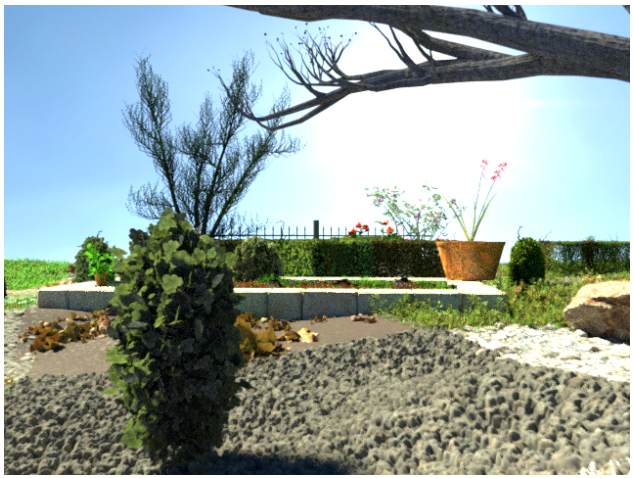
ECCV 2018 Workshop, Munich

Outline

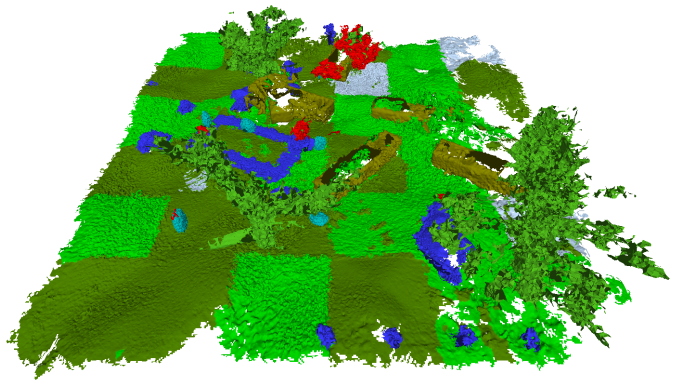
- Challenge Goals
- Garden Dataset
- Evaluation
- Results

3DRMS Challenge Goals

- **Input**: Set of images and known camera poses
- **Goal**: Create a semantically annotated 3D model of the scene
  - Compute depth maps for the images
  - Fuse them together into a single 3D model
  - Incorporate information from the semantics
- **Categories**: Semantics/Geometry, Synthetic/Real

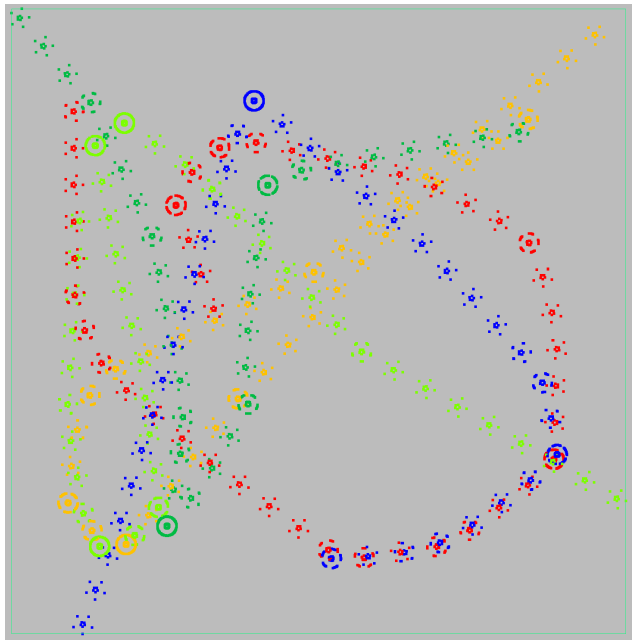


$\Rightarrow$



Virtual and Real Garden Dataset

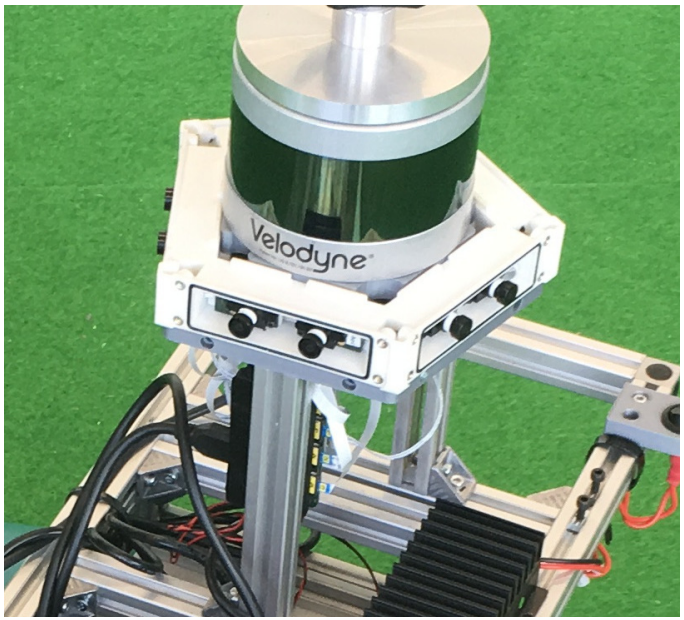
- Four **training** scenes with 20 sequences
  - Total 20k synthetic images
  - Ground truth semantic annotations and depth maps
  - A semantically annotated 3D point cloud
  - Different environment for each sequence (clear, cloudy, overcast, sunset, twilight)
- A **testing** scene with 5 sequences
  - 5k synthetic images with camera poses
  - Ground truth for evaluation only
- A **validation** sequence (2017 dataset)
  - 500 real images with camera poses and calibration

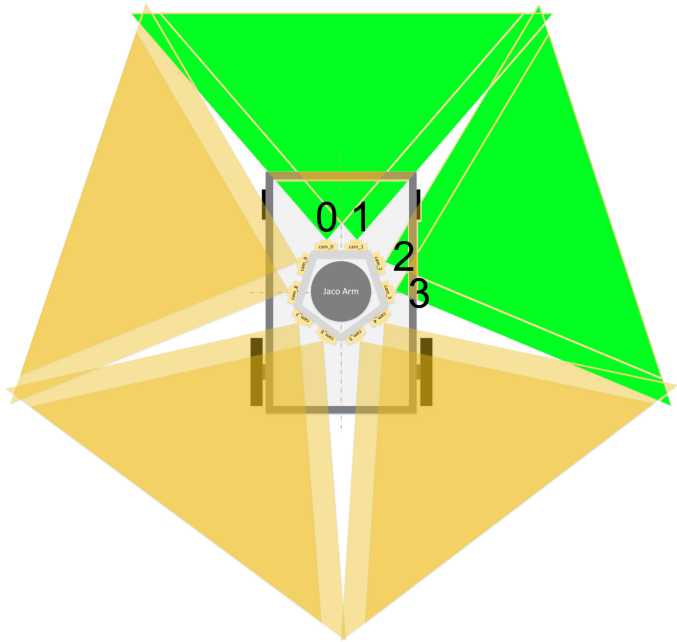


Randomly generated trajectories for the test scene (unique color for each sequence)

Garden Dataset: Image Data

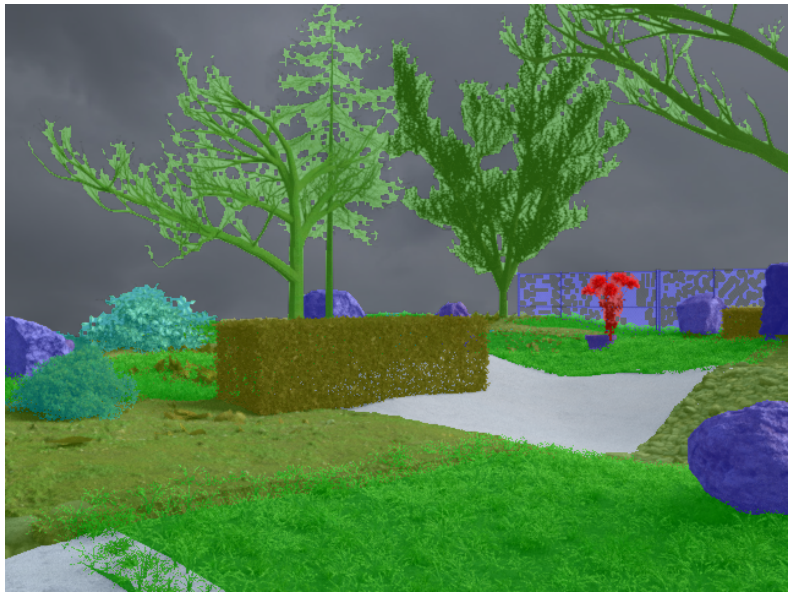
- Pentagonal camera rig with 10 cameras
- Synthetic data: 5 camera pairs
  - VGA resolution (640x480), color
- Real data: 2 camera pairs (0-1, 2-3)
  - WVGA resolution (752x480), color/greyscale

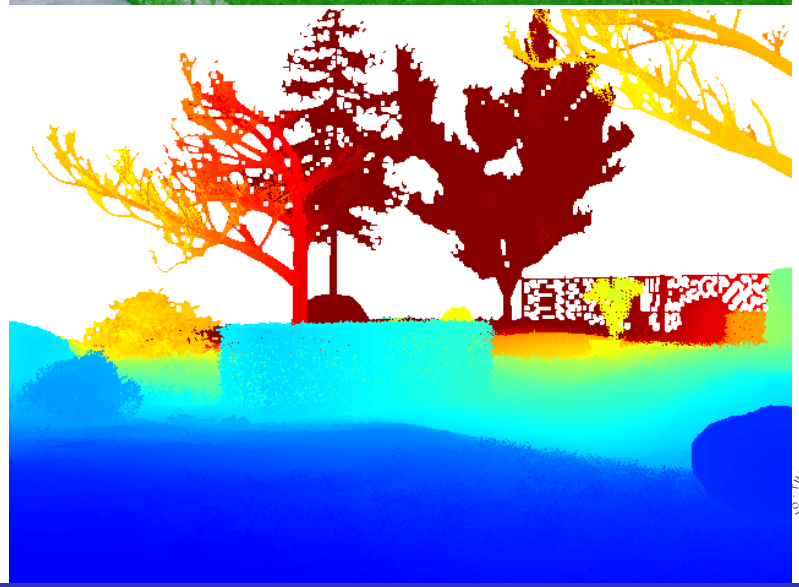




Garden Dataset: Semantics and Depth

- Set of 9 classes we distinguish
  - Grass (light green)
  - Ground (brown)
  - Pavement (grey)
  - Hedge (ochre)
  - Topiary (cyan)
  - Rose (red)
  - Obstacle (blue)
  - Tree (dark green)
  - Background (black)
- Depth map from projected points



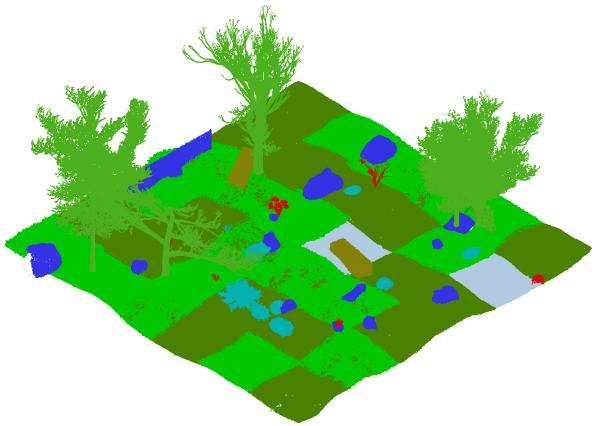
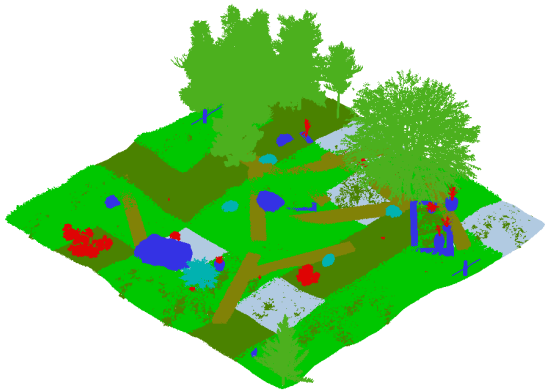
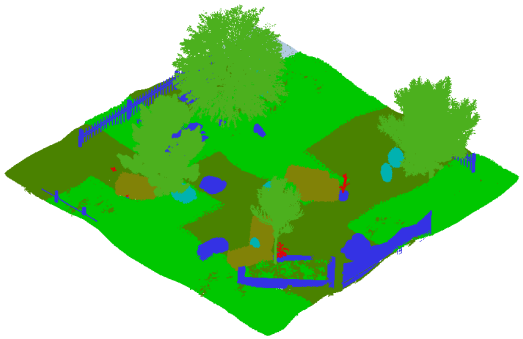
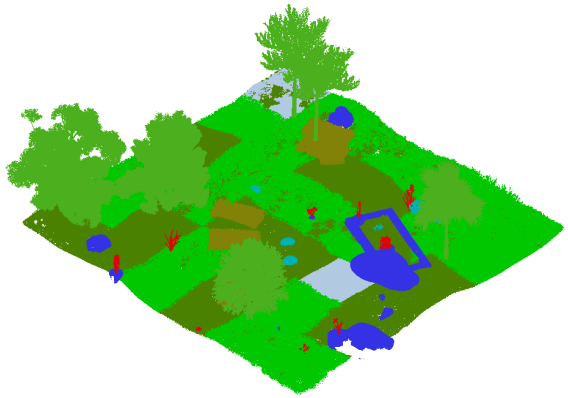


Garden Dataset: Training Point Clouds




Dataset rendered by Hoang-An Le (UvA)

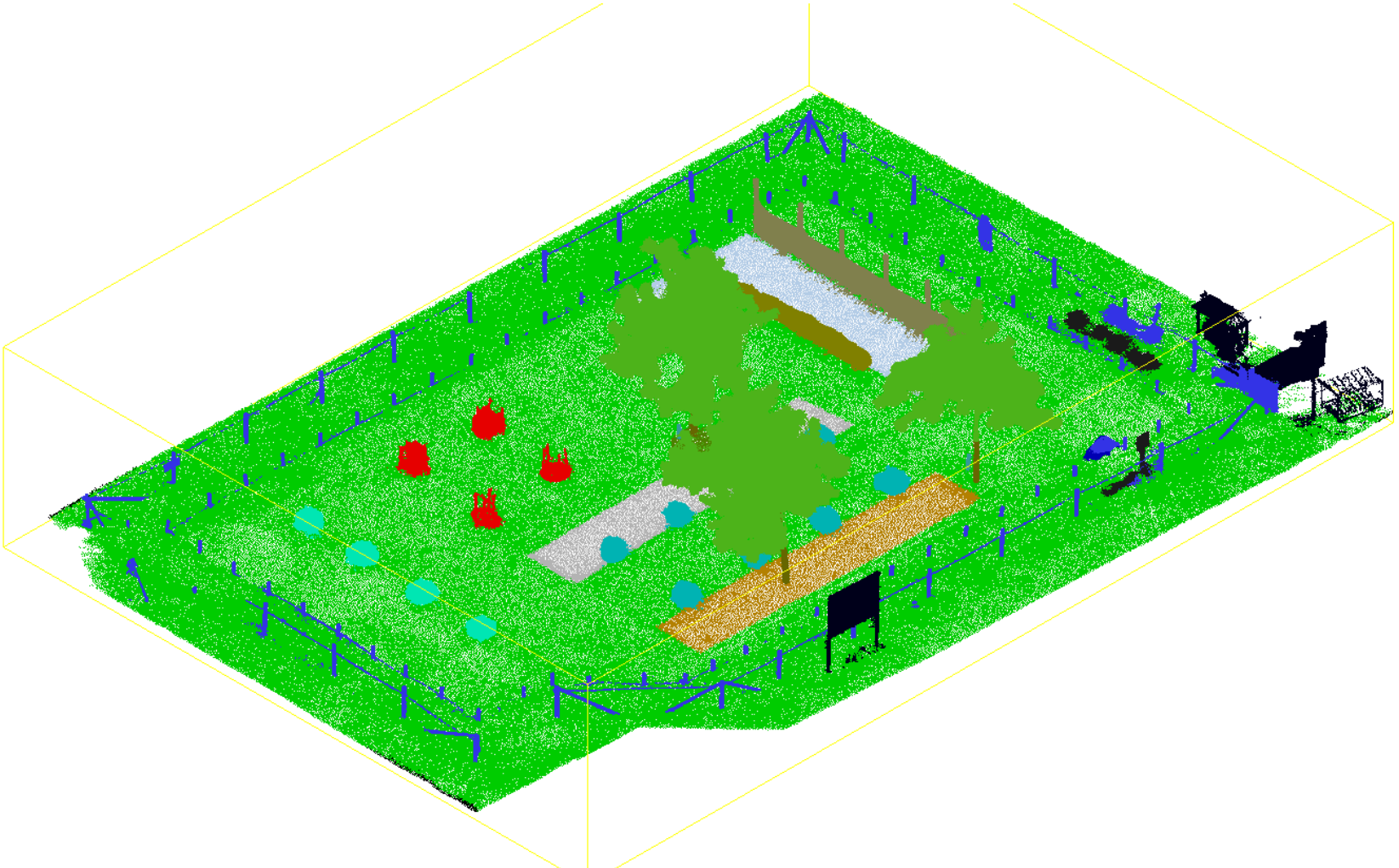
Garden Dataset: Training Point Clouds

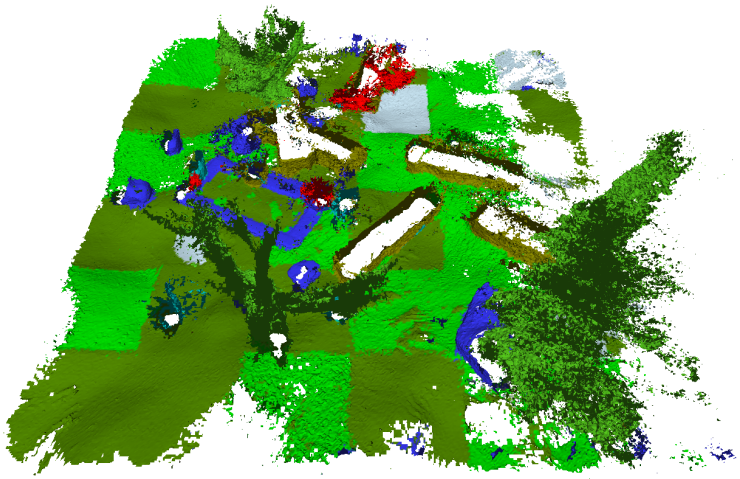
Dataset rendered by Hoang-An Le (UvA)

Garden Dataset: Test Point Cloud

Garden Dataset: Validation Point Cloud



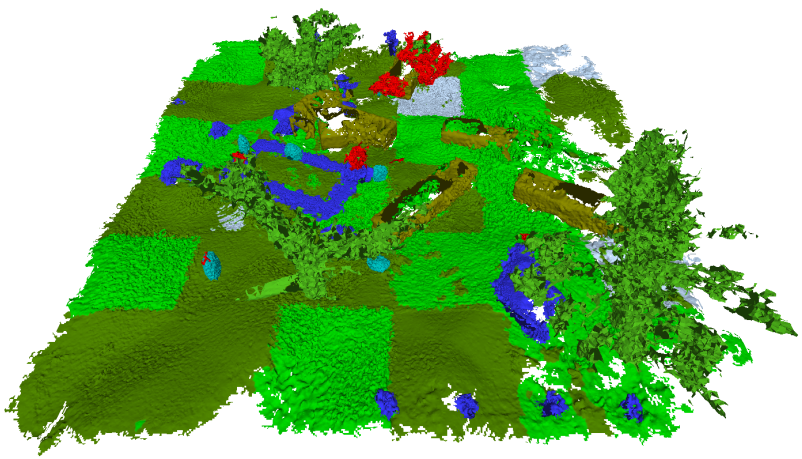
Submissions vs. Ground Truth



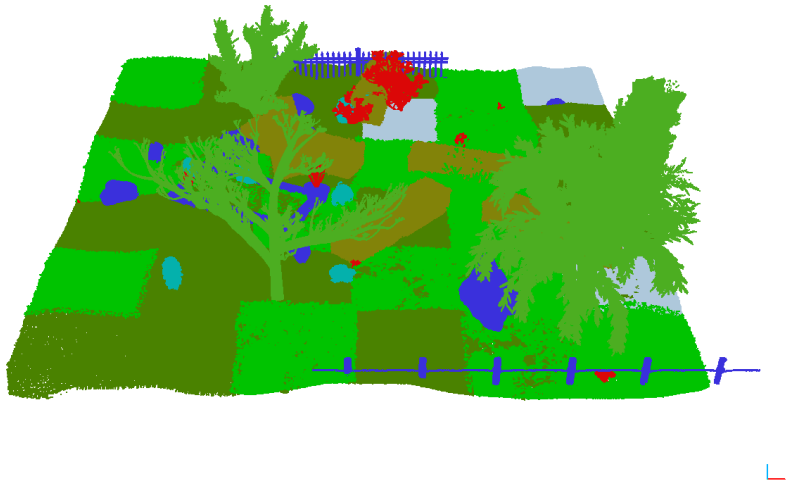
DTIS

GT

Submissions vs. Ground Truth

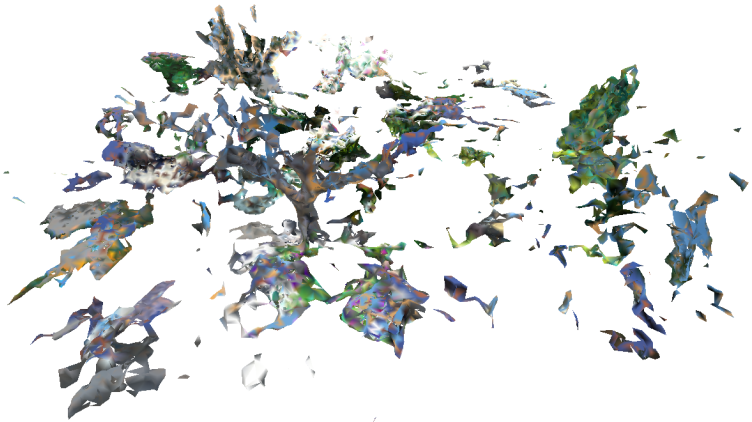


HAB



GT

Submissions vs. Ground Truth



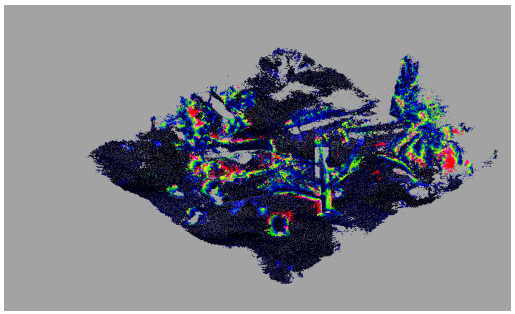
LAPSI

GT

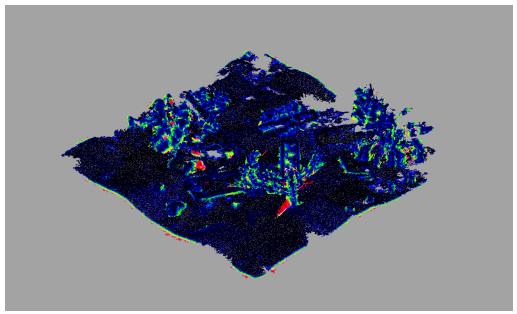
Evaluation Methodology

- Test GT
  - delimited by the bounding box of the test area
- 3D geometry
  - Submissions cropped to bounding box +1 m
  - Cumulative histograms of distances (mesh $\rightleftarrows$ GT points)
  - **Accuracy** is distance d (in m) such that 90% of the reconstruction is within d of the ground truth mesh
  - **Completeness** is the percent of points in the GT point cloud that are within 5 cm of the reconstruction
- Semantics
  - Labels assigned to vertices or faces of the 3D model
  - Projected to all test images
  - **Prediction accuracy** of the pixels corresponding to the 3D test part

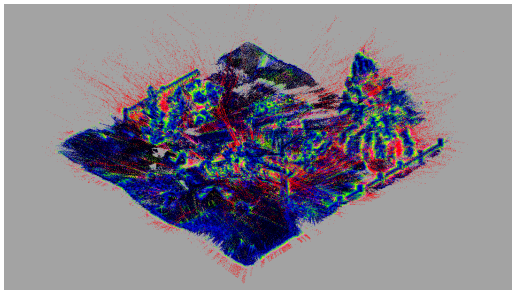
Evaluation: 3D Geometry



DTIS

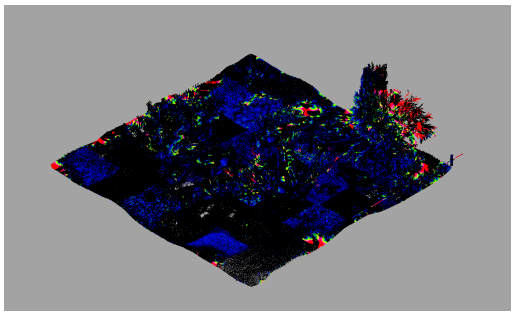


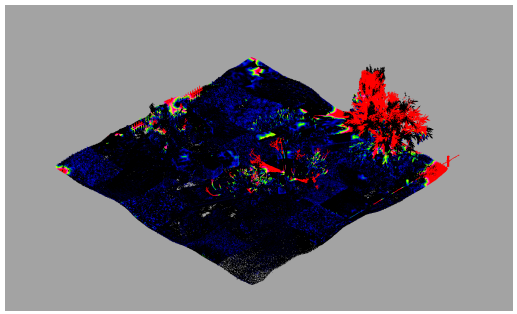
HAB

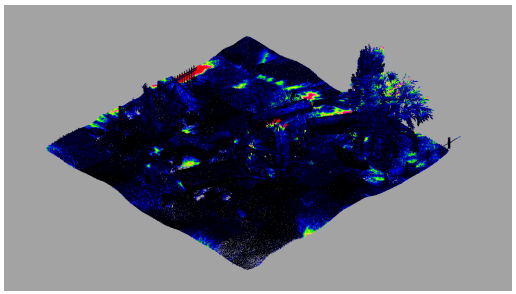


COLMAP[1] (baseline)

Accuracy





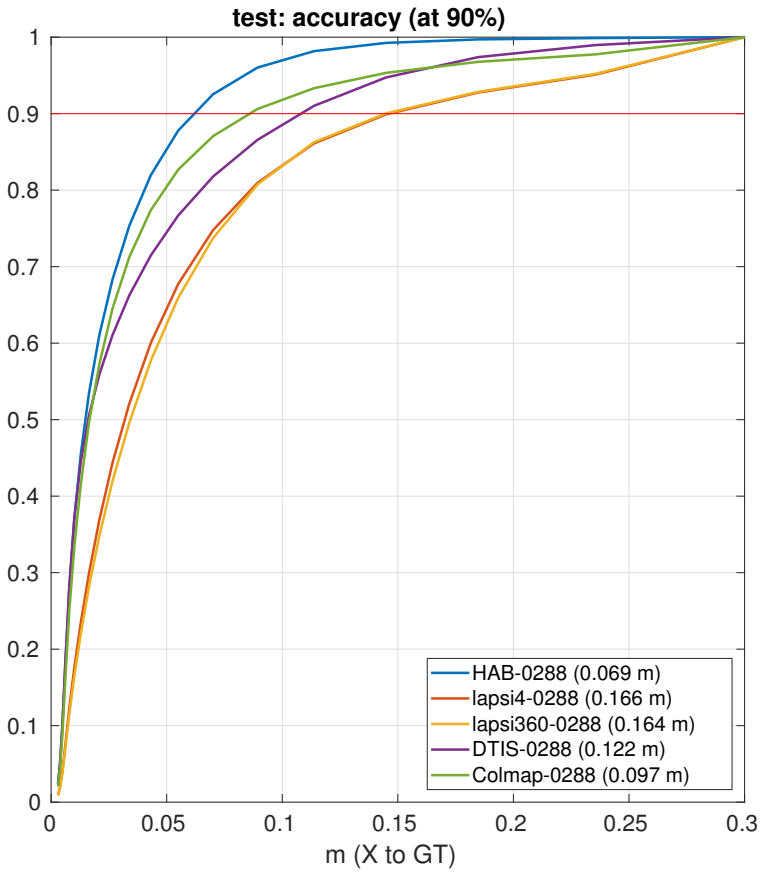


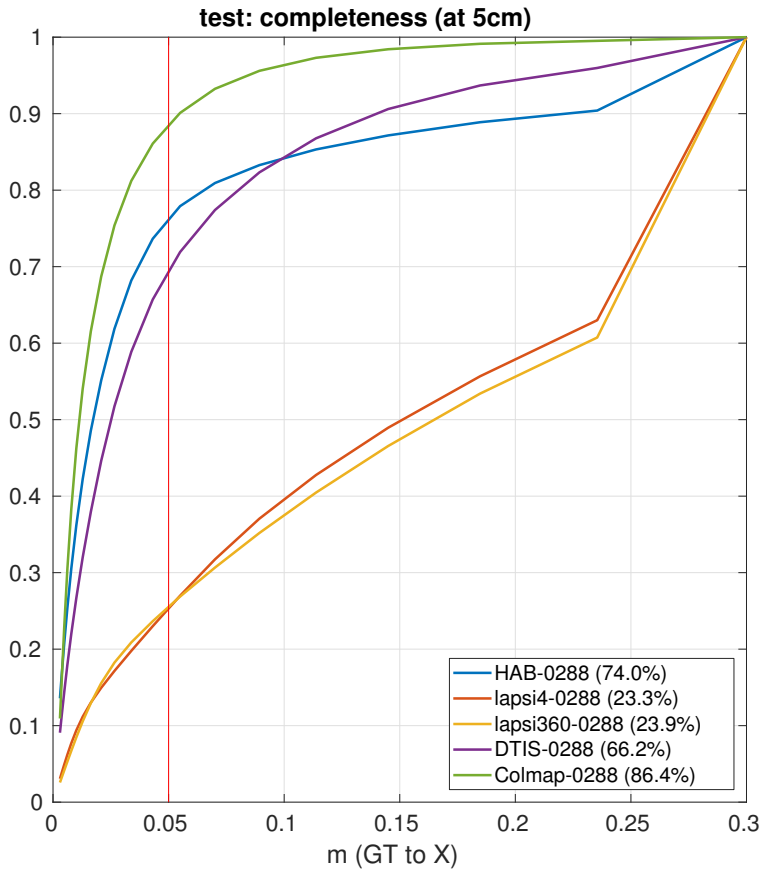
Completeness

- Distances [0-1m]: Cold colors indicate well reconstructed segments
- Hot colors indicate noisy surface (accuracy) or missing parts (completeness).

[1] Schönberger et al.: Structure-from-motion revisited. CVPR, 2016.

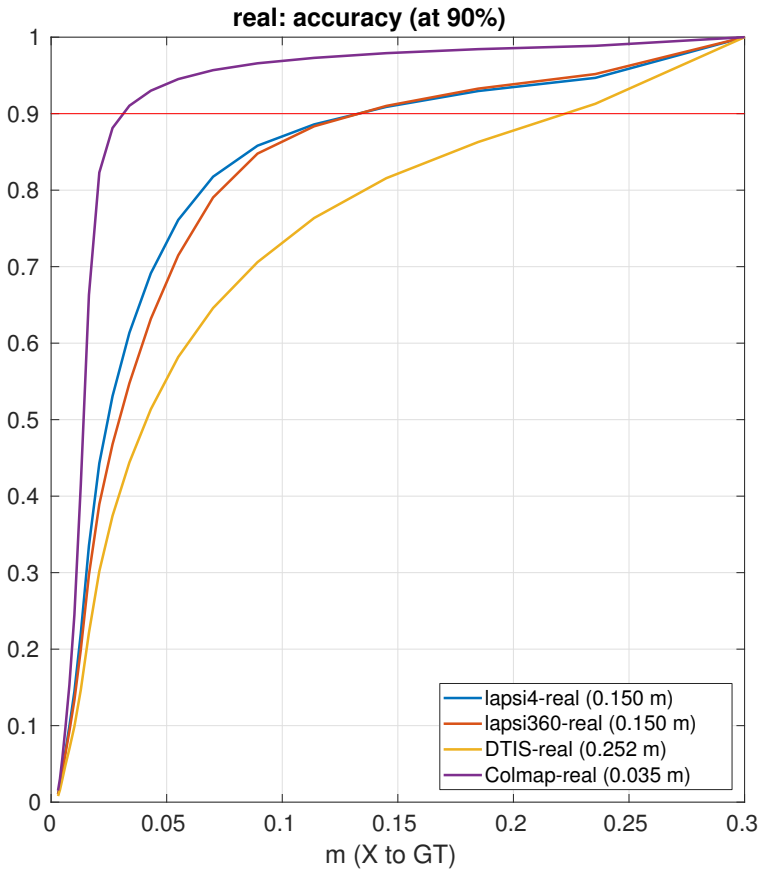
Evaluation: 3D Geometry (synthetic data)

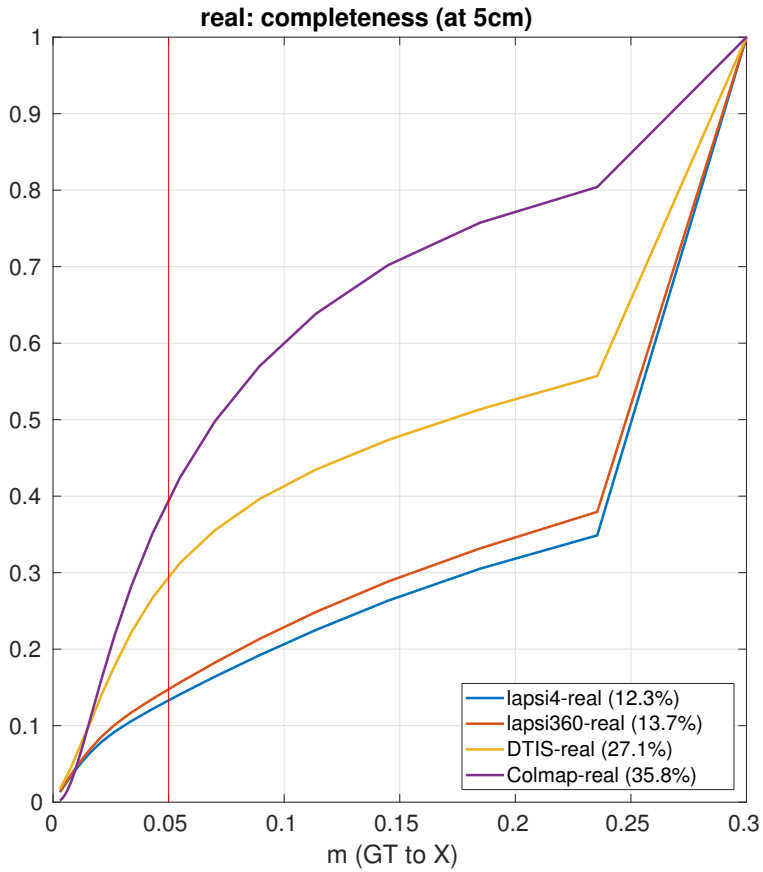




- Cumulative plots of distances (mesh $\rightleftarrows$ GT points)

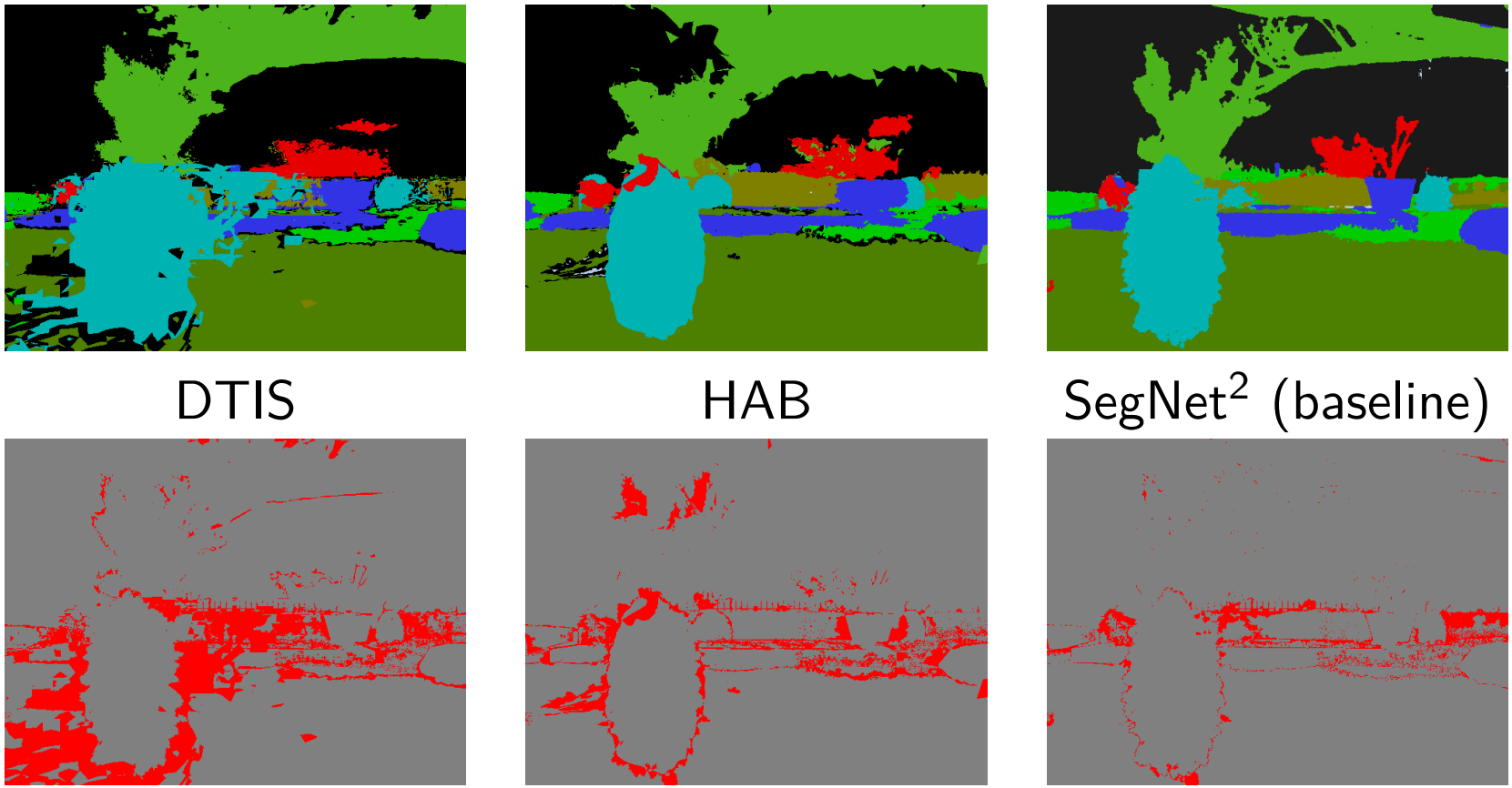
Evaluation: 3D Geometry (real data)





- Cumulative plots of distances (mesh $\rightleftarrows$ GT points)

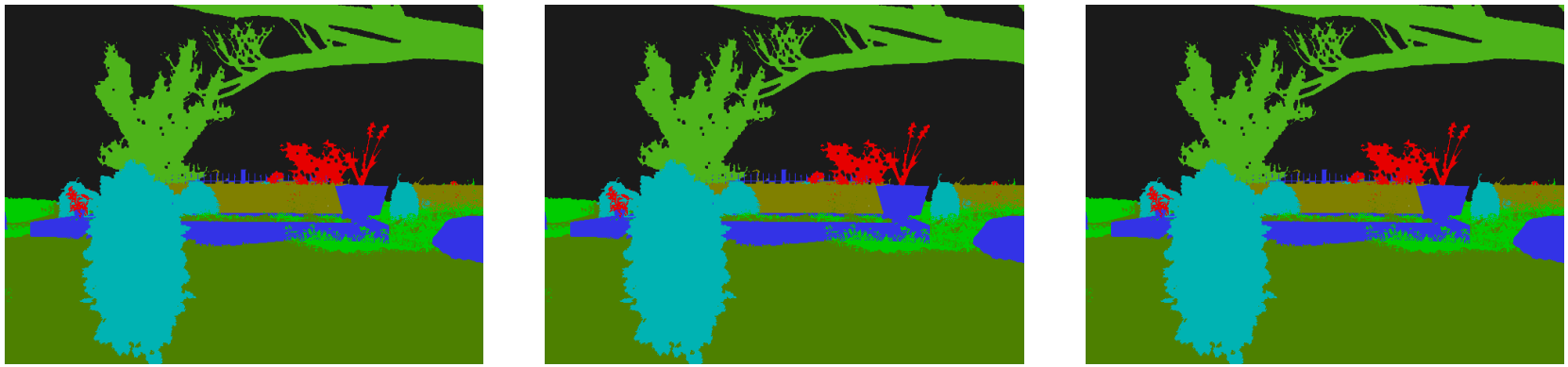
Evaluation: Semantics (synthetic data)



- Error: *red* incorrect pixels, *grey* correct, *black* not evaluated
- SegNet: pre-trained with ImageNet + challenge training set

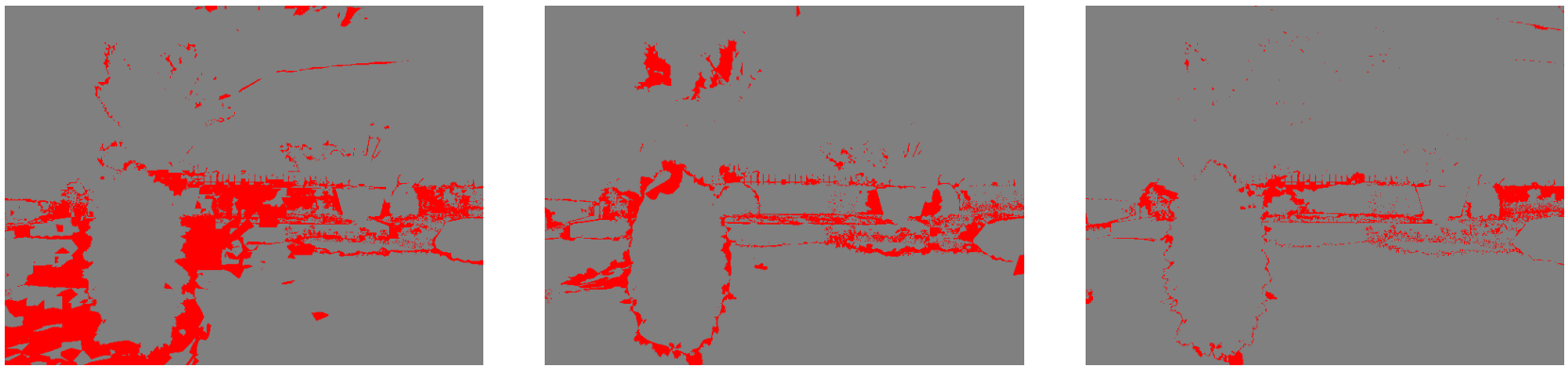
[2] Badrinarayanan et al.: SegNet: A deep convolutional encoder-decoder architecture for image segmentation. PAMI, 2017.

Evaluation: Semantics (synthetic data)



GT

DTIS | HAB | SegNet[3] (baseline)

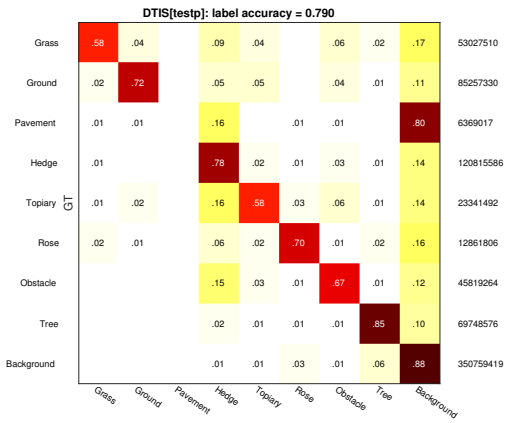


Error mask

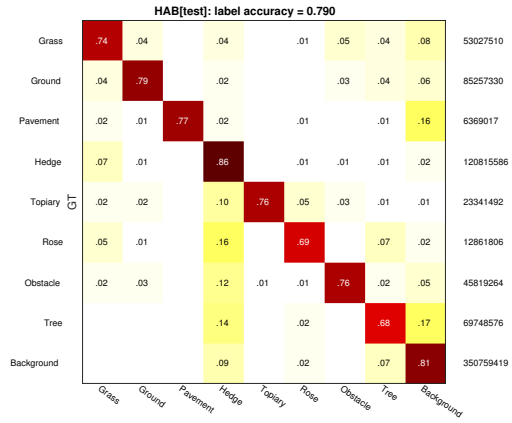
- Error: *red* incorrect pixels, *grey* correct, *black* not evaluated
- SegNet: pre-trained with ImageNet + challenge training set

[3] Badrinarayanan et al.: SegNet: A deep convolutional encoder-decoder architecture for image segmentation. PAMI, 2017.

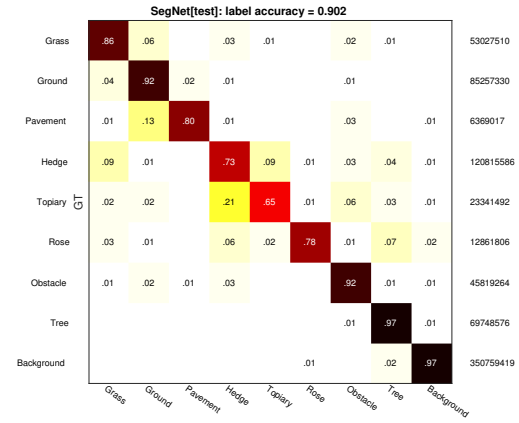
Evaluation: Semantics (synthetic data)



79.0%
DTIS[3D]



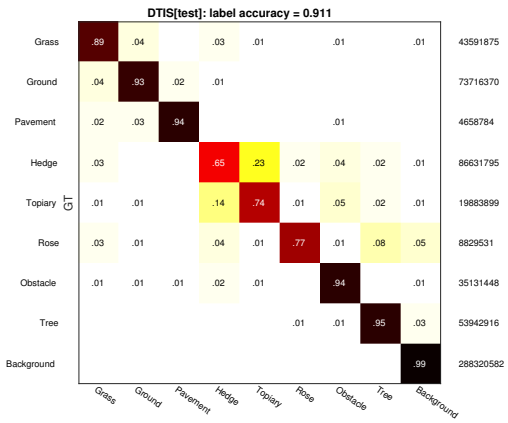
79.0%
HAB

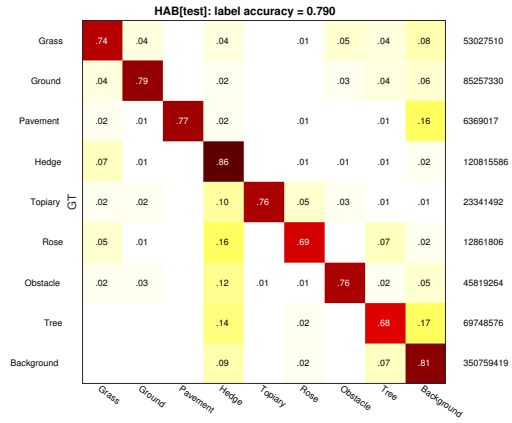


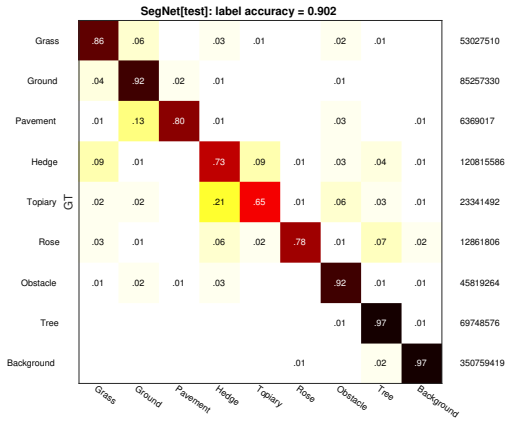
90.2%
SegNet (baseline)

- Confusion matrix: *dark* on diagonal indicates good match of the prediction with GT labels
- Semantic accuracy: pixelwise ratio of correct predictions over all test images

Evaluation: Semantics (synthetic data)







91.1%
DTIS[2D]

79.0%
HAB

90.2%
SegNet (baseline)

- Confusion matrix: *dark* on diagonal indicates good match of the prediction with GT labels
- Semantic accuracy: pixelwise ratio of correct predictions over all test images

Evaluation: Summary of Performance

Method	3D Reconstruction		*Semantic*
	Accuracy	*Completeness*	*Accuracy*
DTIS	0.122 m	66.2 %	**91.1 %**
HAB	**0.069 m**	**74.0 %**	79.0 %
LAPSI	0.164 m	23.9 %	
Baseline	*0.097 m*	*86.4 %*	*90.2 %*

Test set (Synthetic data)

Evaluation: Summary of Performance

Method	3D Reconstruction		*Semantic*
	Accuracy	*Completeness*	*Accuracy*
DTIS	0.25 m	**27.1 %**	**65.7 %**
HAB			
LAPSI	**0.15 m**	13.7 %	
Baseline	*0.035 m*	*35.8 %*	*85.4 %*

Validation set (Real data)

Challenge Results Summary

- Best performers for **synthetic** data
  - Category A/B - 3D Semantic mesh: HAB
  - Category C - 2D Semantic image annotations: DTIS
- Best performer for **real** data: DTIS

Congratulations and thanks to all participants!

We are open to additional submissions.