Expanding to 4D Panoptic Segmentation using Scene Flow

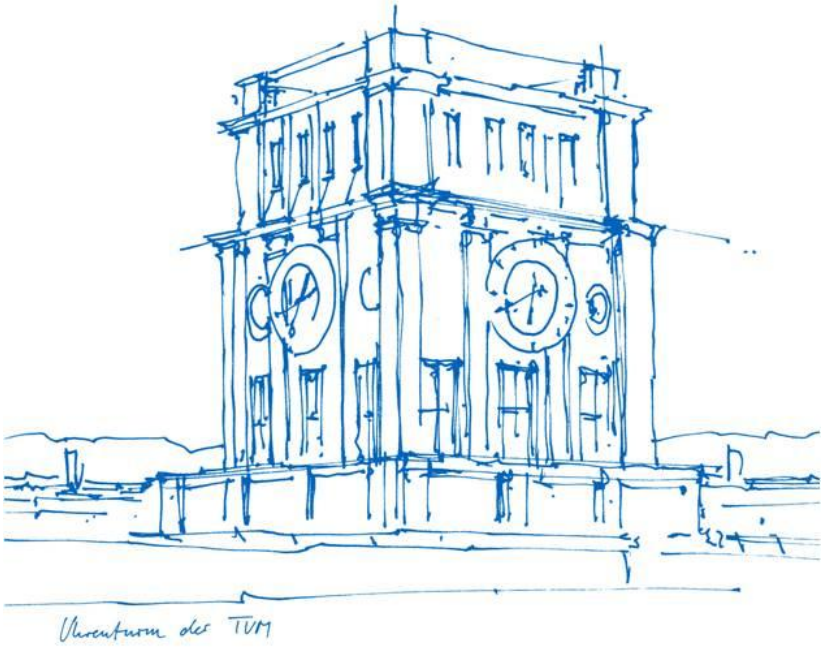
Niklas Bubeck, Doğaç Güzelkokar

Supervisor: Mariia Gladkova

Technische Universität München

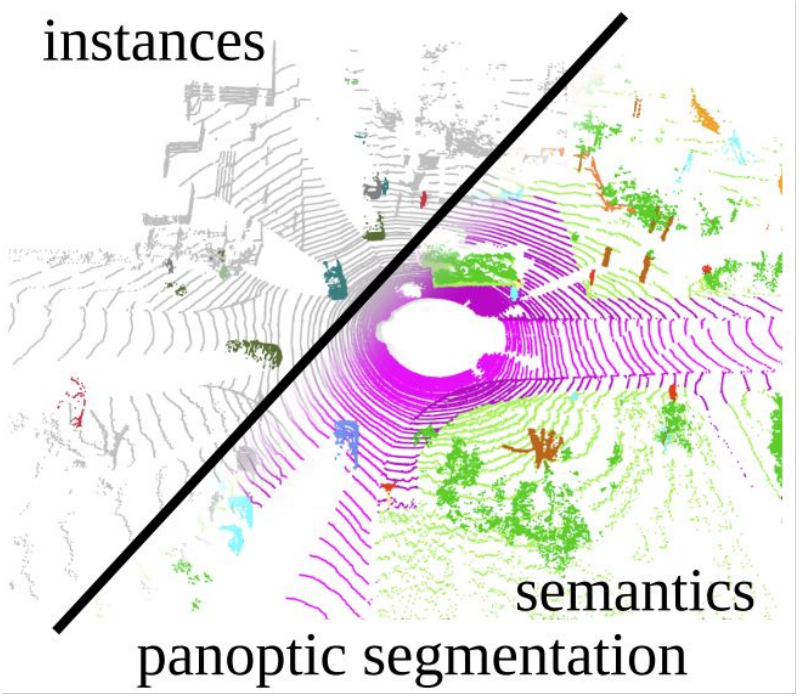
Chair of Computer Vision and Artificial Intelligence

August 30, 2022



What is 4D Panoptic Segmentation?

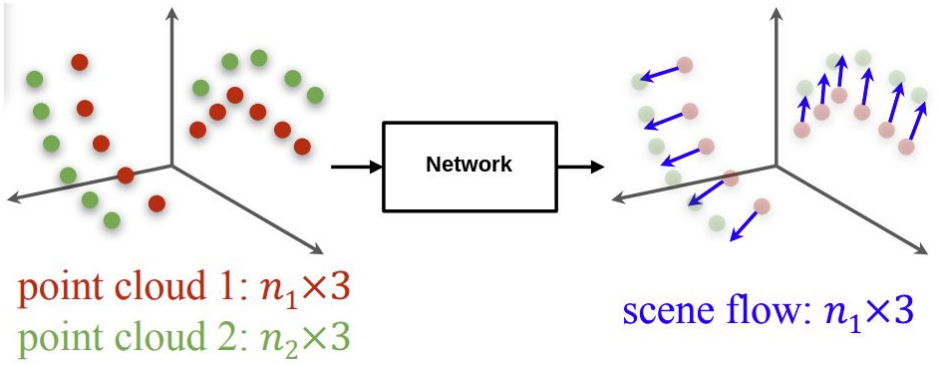
- Detect instances as well as semantics
- 4D refers to tracking instances over time



Mehmet Aygün, Aljoša Ošep et al., 4D Panoptic LiDAR Segmentation, 2021

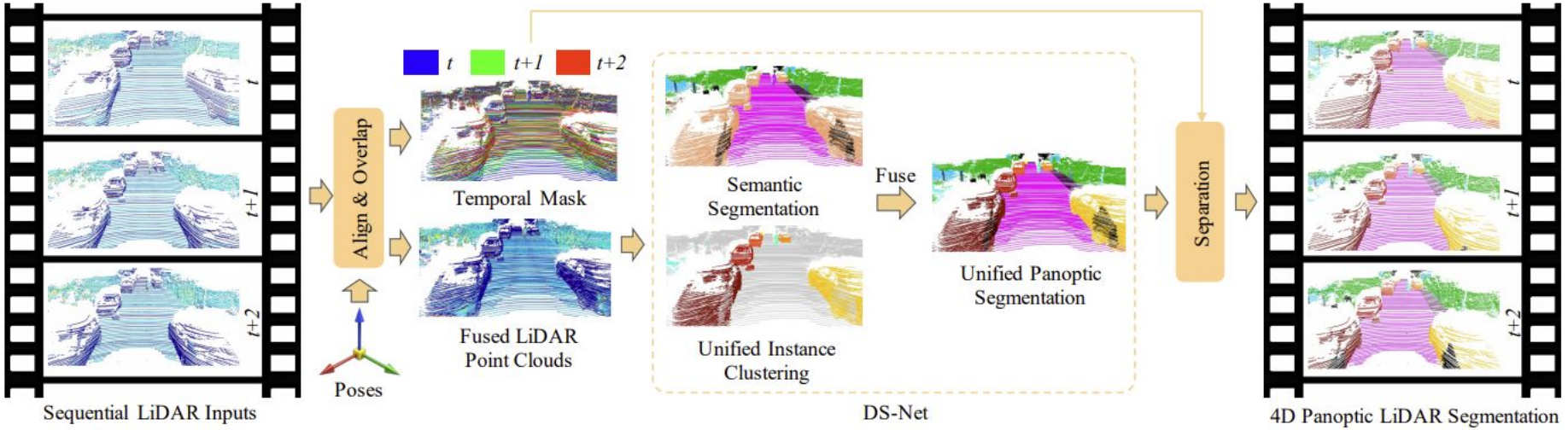
What is Scene Flow?

- Find the motion for each point
- Unordered points
- Different number of points
- 3D case of optical flow



Liu, Xingyu, et al. "Flownet3d: Learning scene flow in 3d point clouds." 2019.

Related Work: Dynamic-Shifting Net

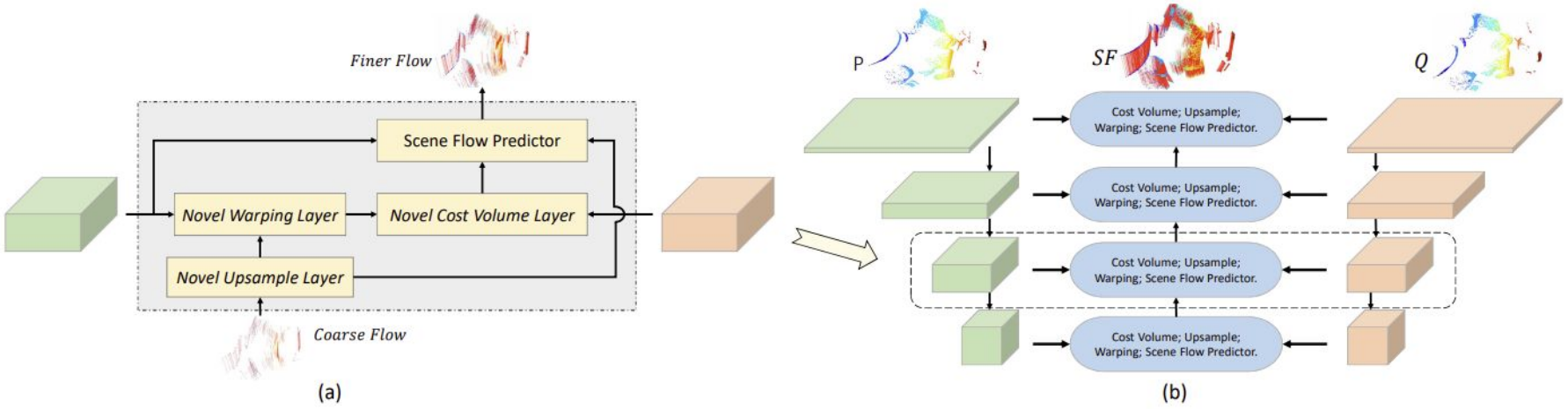


- Apply rigid transformation
- Solve 4D by aligning, predicting on aligned scans then separation

Why not compensate for object motion?

Fangzhou et al. , LiDAR-based 4D Panoptic Segmentation via Dynamic Shifting Network , 2022

Related Work: PointPWC



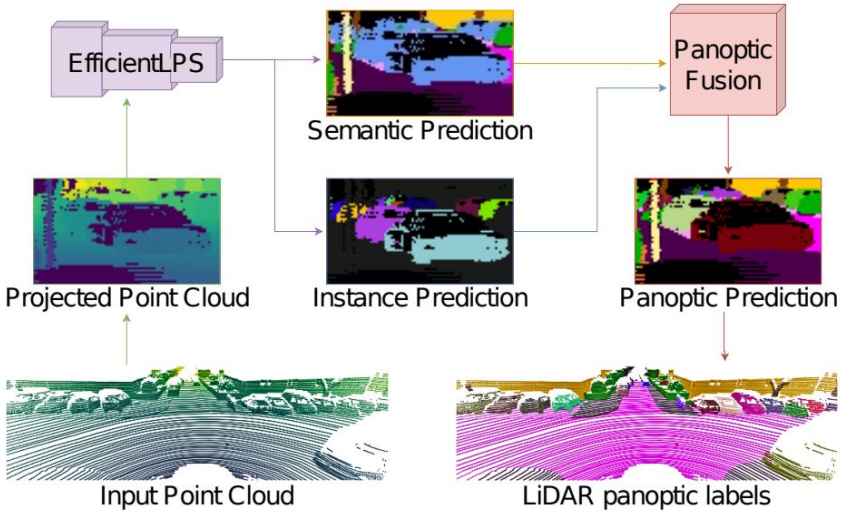
- Flow on different resolutions
- Captures details
- Introduces valuable loss functions

Why not train on semanticKitti dataset?

Wenxuan Wu et al., PointPWC-Net: Cost Volume on Point Clouds for (Self-)Supervised Scene Flow Estimation, 2020

Related Work: EfficientLPS

- Panoptic segmentation on image representation



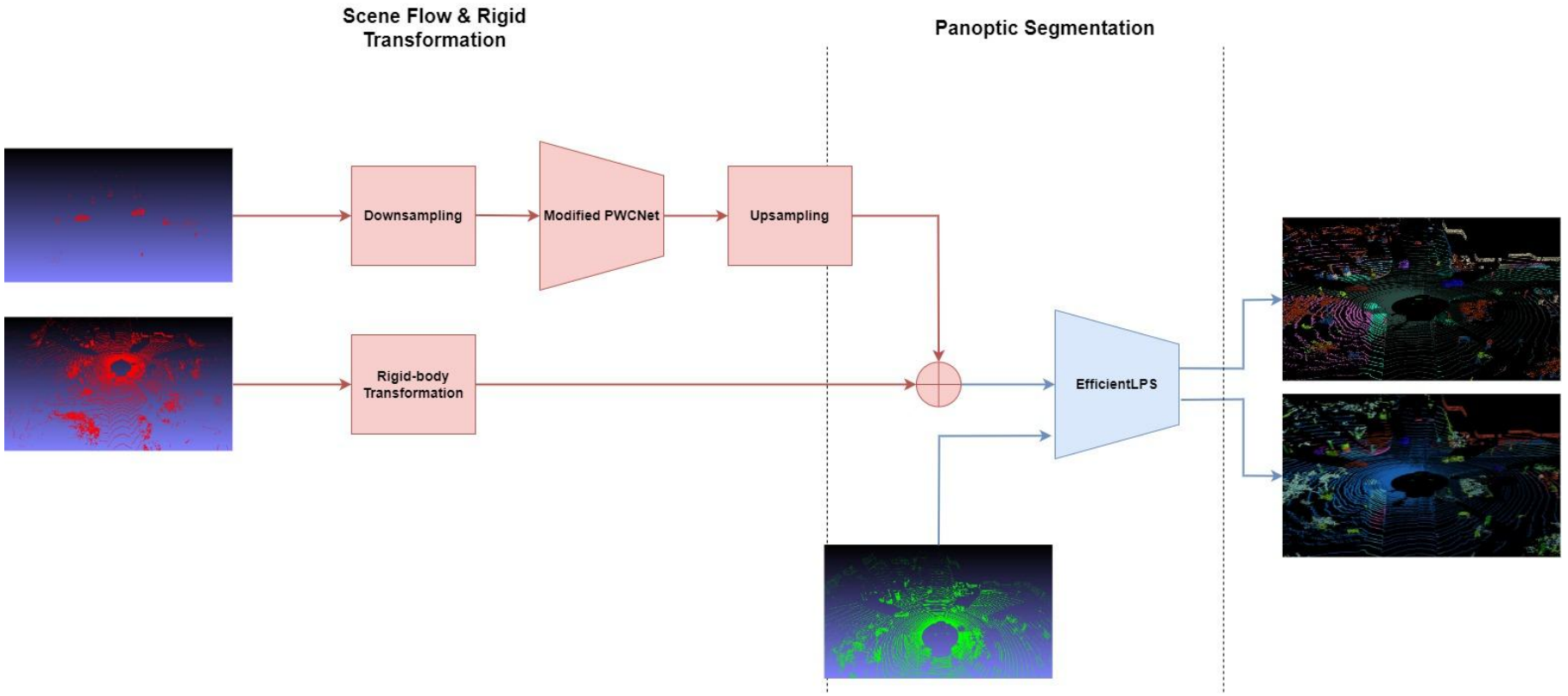
Why not use on aligned scans?

Kshitij Sirohi et al., EfficientLPS: Efficient LiDAR Panoptic Segmentation, 2021

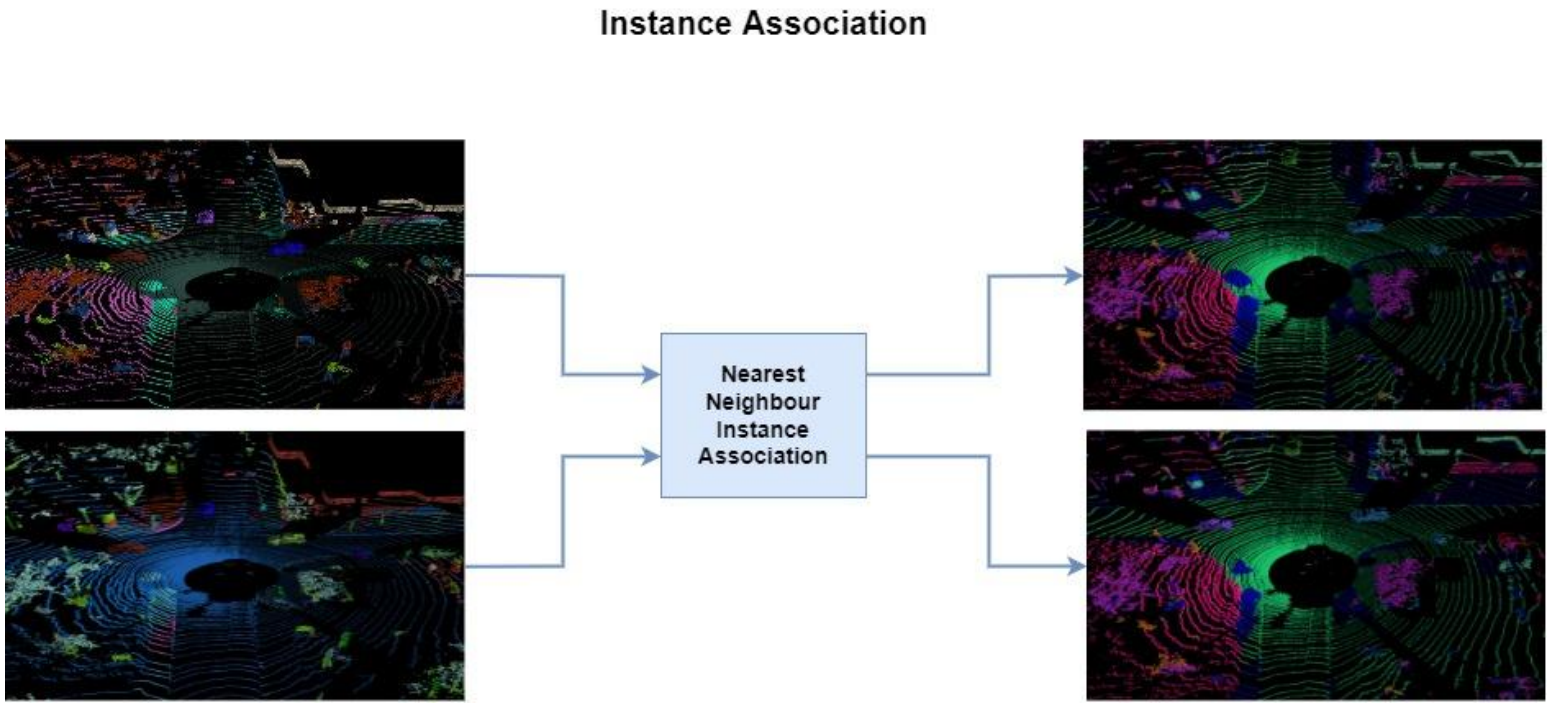
Our Contributions

- Overall concept that works with every scene flow and panoptic network
- Adapted scene flow working on lower hardware requirements and semanticKitti
- Intersection over union (IoU) instance matching to solve time dependency

Overall Workflow



Overall Workflow



Sampling

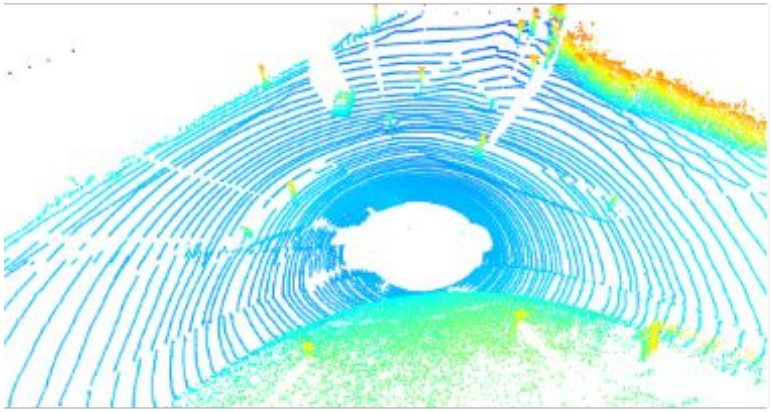
Why:

- Too much data
- A lot of noise
- High bias

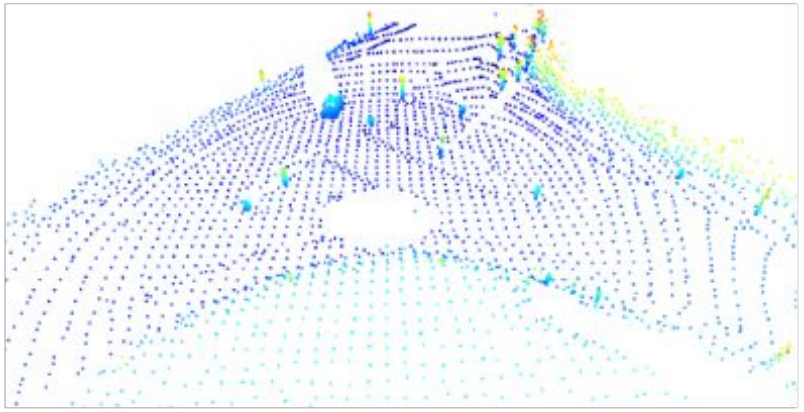
How:

- Use voxels as space representations
- Emphasize different classes

Downsampling to 50k



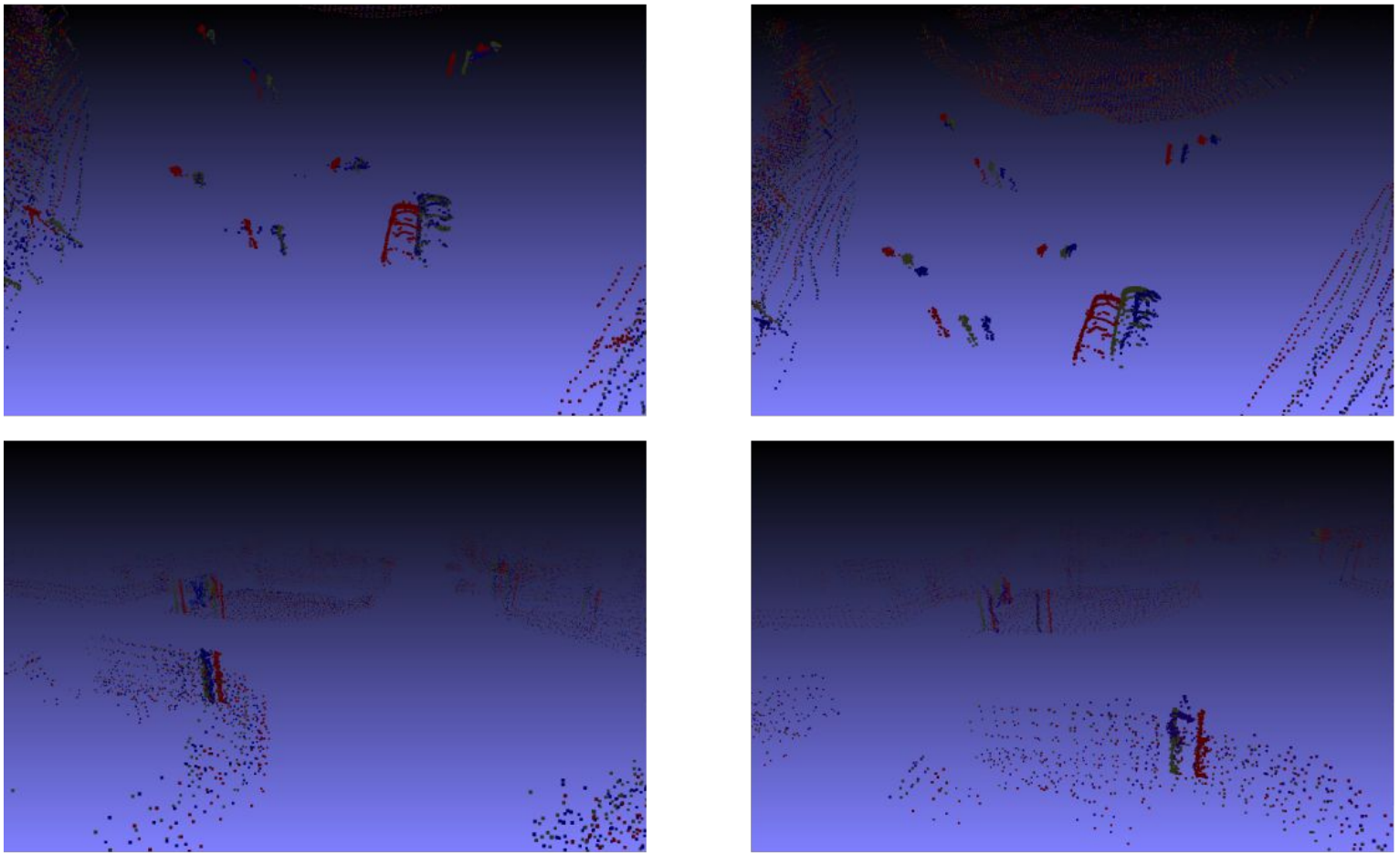
Downsampling to 6k emphasizing things classes



Scene Flow: Available methods

FLOT

PointPWC



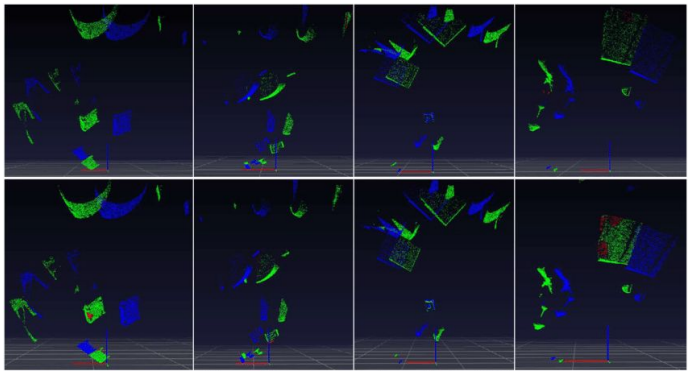
Pointcloud 1

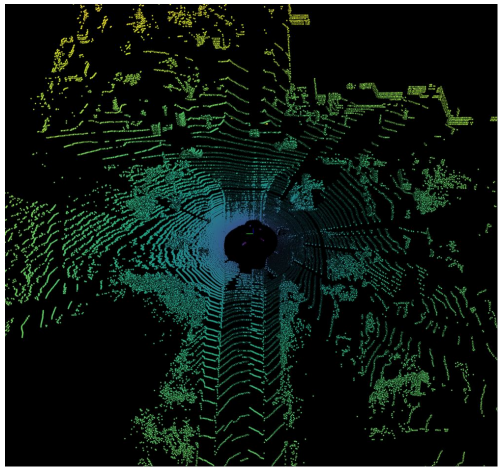
Pointcloud 2

Applied Scene Flow

Scene Flow: Problems

- RGBD vs. Pointcloud data
- 360deg fov → lower point density for individual objects
- No color values as features
- Higher number of points → exploding memory
- Unordered points with varying size → No bijective mapping possible
- Bidirectional reflectance disturbances cause many outliers





Mayer, Nikolaus, et al. "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation." 2016.

Behley, Jens, et al. "Semantickitti: A dataset for semantic scene understanding of lidar sequences." 2019.

Scene Flow: Losses

Chamfer Loss:

- Minimizes distance

$$\ell_C^l(P_w^l, Q^l) = \sum_{p_w^t \in P_w^l} \min_{q^l \in Q^l} \left\| p_w^l - q^l \right\|_2^2 + \sum_{q^l \in Q^l} \min_{p_w^l \in P_w^l} \left\| p_w^l - q^l \right\|_2^2$$

Curvature Loss:

- Local shape characteristics

$$\delta^l(p_i^l) = \frac{1}{|N(p_i^l)|} \sum_{p_j^l \in N(p_i^l)} (p_j^l - p_i^l)$$

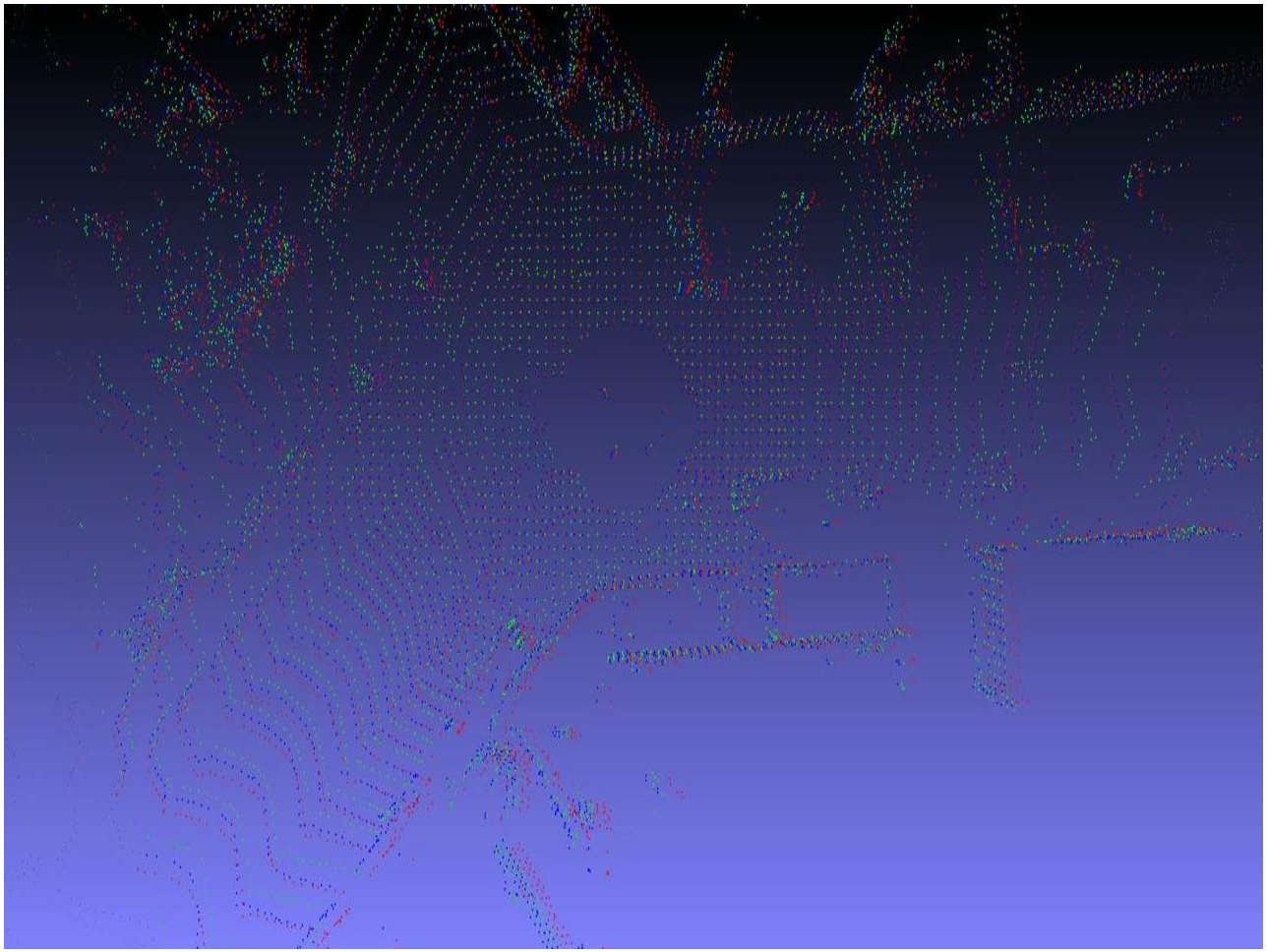
Smoothness Loss:

- Flow in local regions should be similar

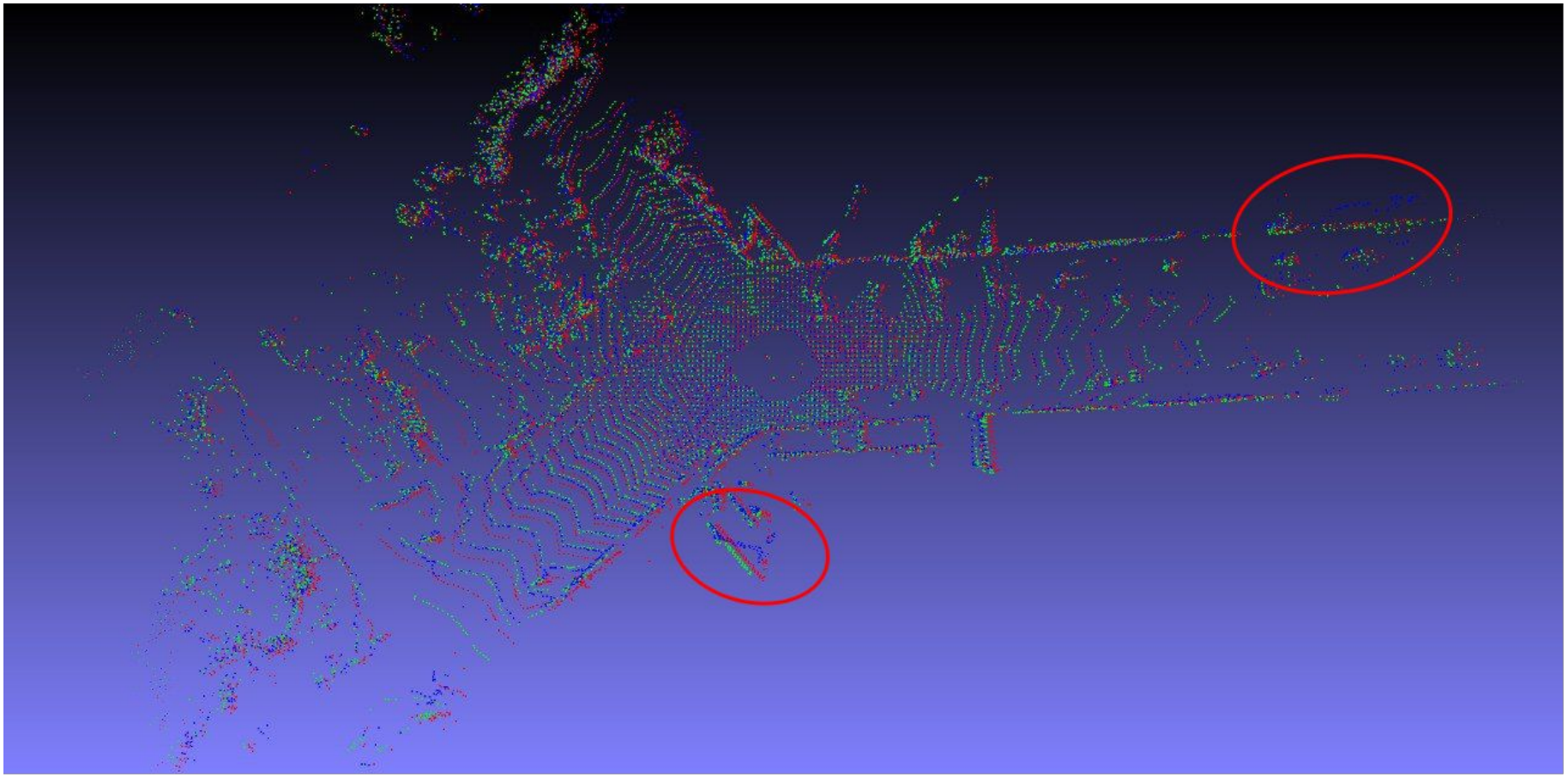
$$\ell_S^l(SF^l) = \sum_{p_i^l \in P^l} \frac{1}{|N(p_i^l)|} \sum_{p_j^l \in N(p_i^l)} \left\| SF^l(p_j^l) - SF^l(p_i^l) \right\|_2^2$$

Curvature and Smoothness act as constraint

Scene Flow: Using all losses



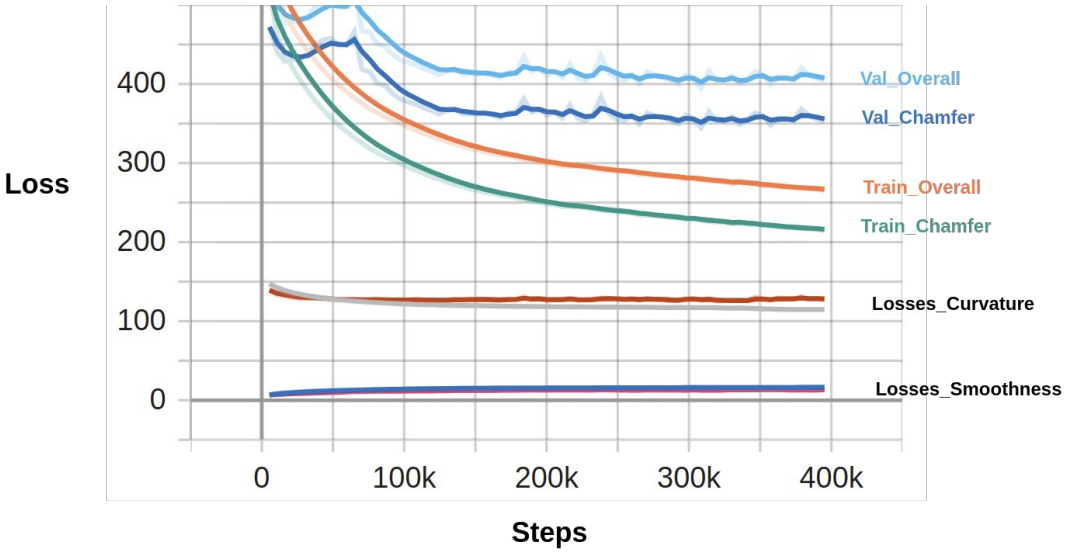
Scene Flow: Without constraints



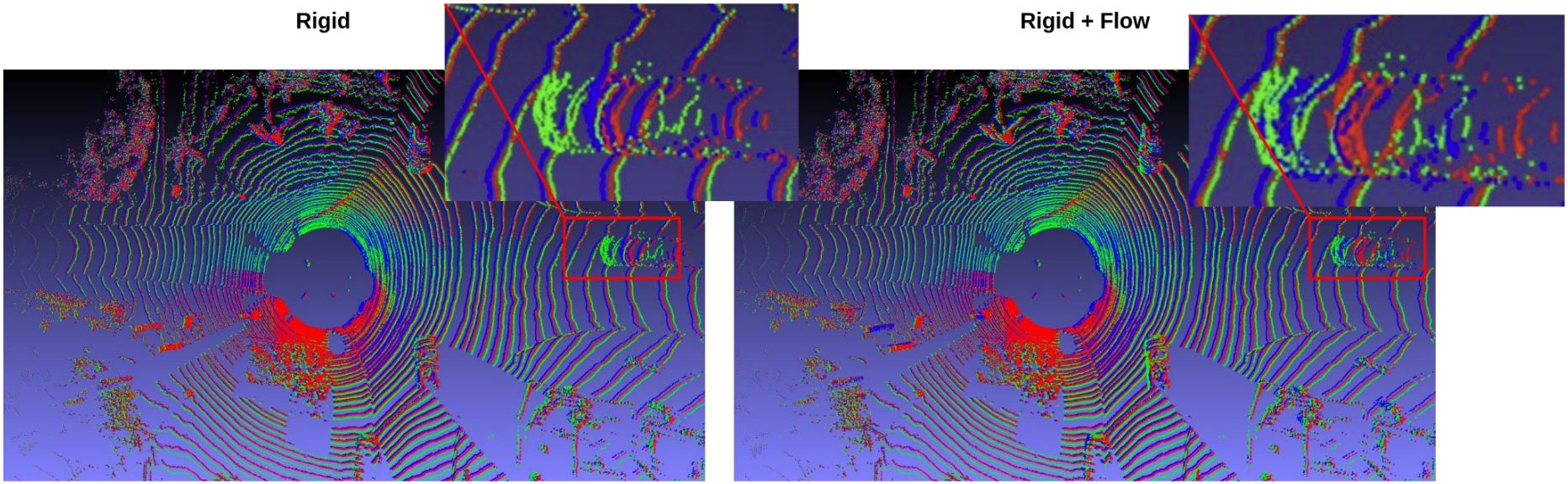
Scene Flow: Training

- Emphasize things classes
- Use low resolution stuff classes (better space representation)
- Padding for scans with less than 8192 points

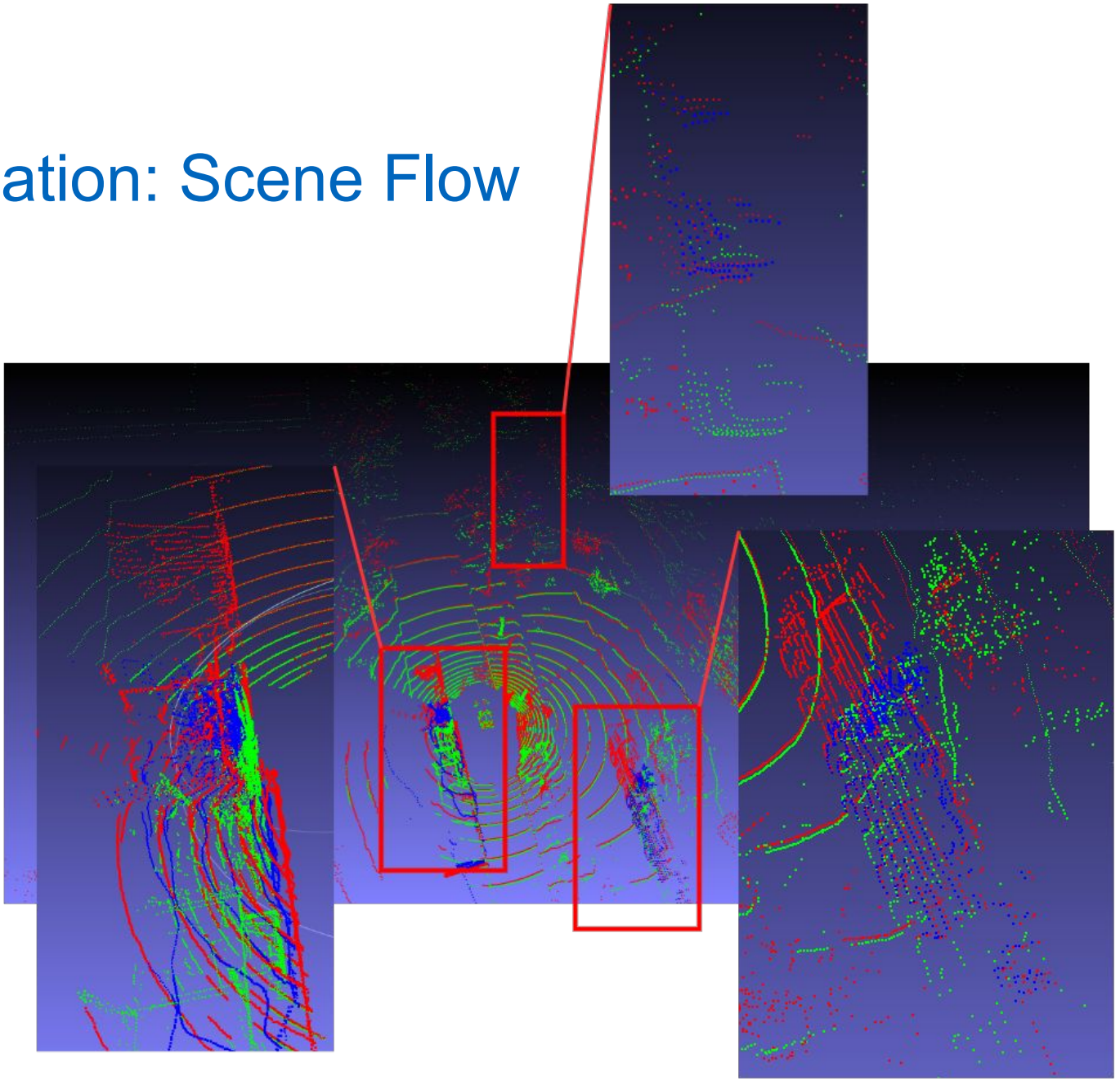
- Refine with only things classes and backprojection (t → t-1)



Evaluation: Scene Flow

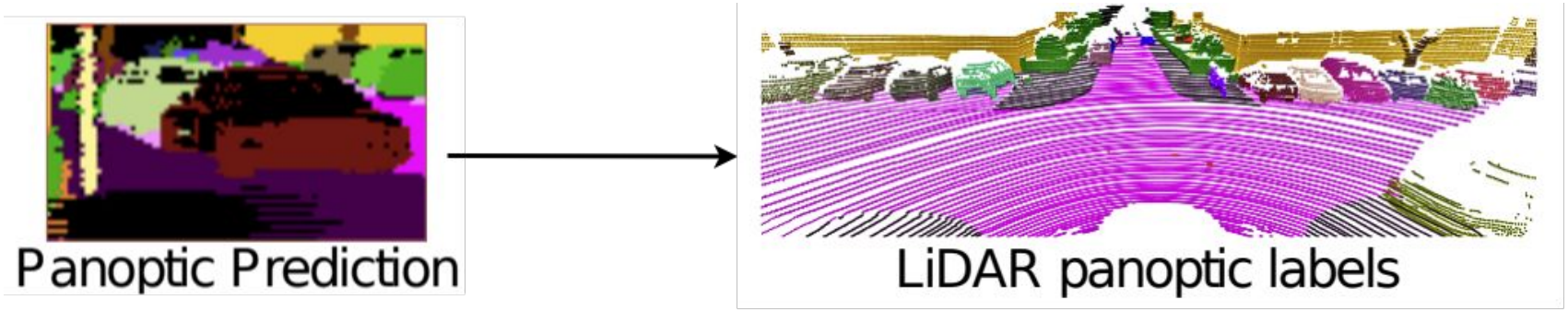


Evaluation: Scene Flow



Efficient LPS: Inference

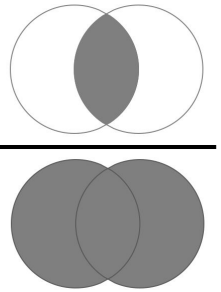
- Inference on aligned point cloud did NOT work
- Corrupted labels for the second point cloud of the stack
- Instead: Make inference on all single scans

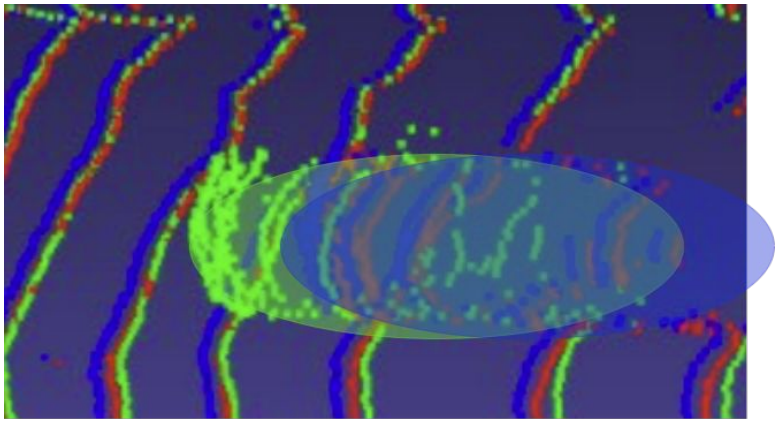


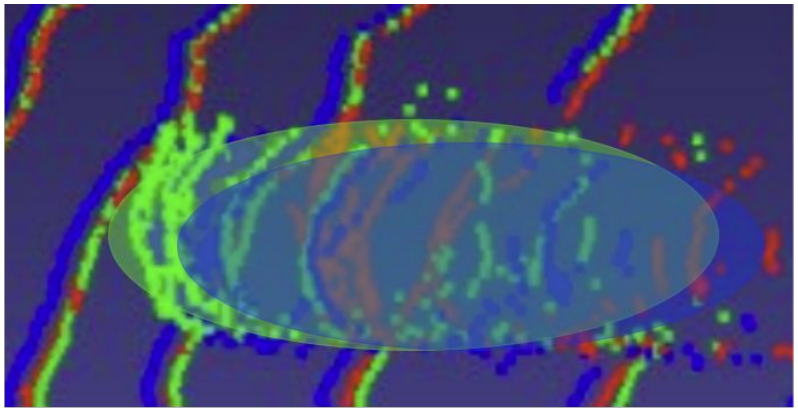
IoU-Matching

- Use masks for Instance ID's
- Find nearest neighbor (NN) for each point
- Assign major NN instance ID

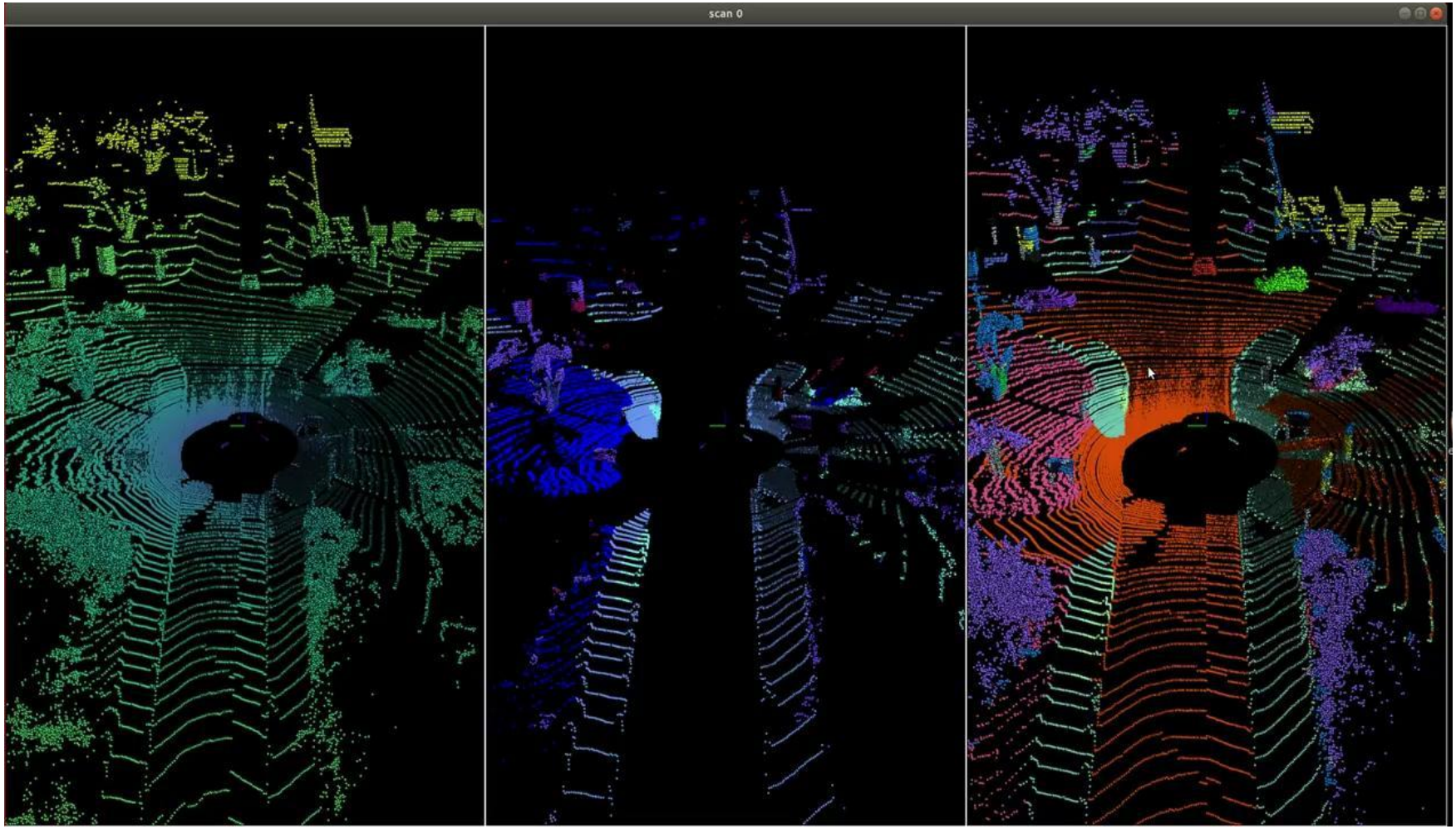
$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$







Evaluation: Panoptic Segmentation



Bottlenecks

- Hardware requirements and dependencies
- Sampling strategy introduces high variance
- No Fine-tuning/ H-param Optimization of the Scene Flow
- Robustness of instance mapping

Future Work

- Run a semantic/panoptic segmentation network to obtain labels for test case, do adaptive sampling based on the outputs and apply our workflow to the test sequences.

- Use better & more GPUs

- Train our own or retrain existing best panoptic segmentation models for running on temporally unified scans