RESSCAL3D++: Joint Acquisition and Semantic Segmentation of 3D Point Clouds

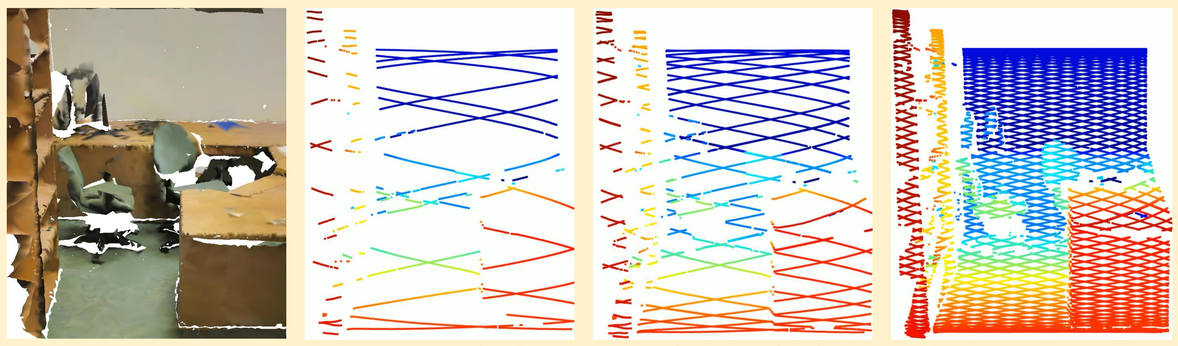
Remco Royen[1], Kostas Pataridis[2], Ward van der Tempel[2] and Adrian Munteanu[1]

[1] Vrije Universiteit Brussel, Department of Electronics and Informatics (ETRO)
[2] VoxelSensors

Full paper? Scan me

Introduction

VoxelSensors' ultra-low-power 3D sensor allows resolution scalable point cloud acquisition



3D scene | Low resolution t_1 time | Med. resolution t_2 time | High resolution t_3 time

This new technology allows resolution scalable processing

- Use acquisition latency to start processing
- Retrieve early predictions before dense point cloud is acquired
- Reduce overall processing time

Why lose (t_3 - t_1) potential processing time and not start processing during acquisition?

Dataset: VX-S3DIS

The first semantic segmentation dataset which mimics the working of a resolution-scalable 3D sensor

- Progressively captured using a Lissajous scanning pattern
- Enables intra-scan semantic processing
- Points are ordered in acquisition order.

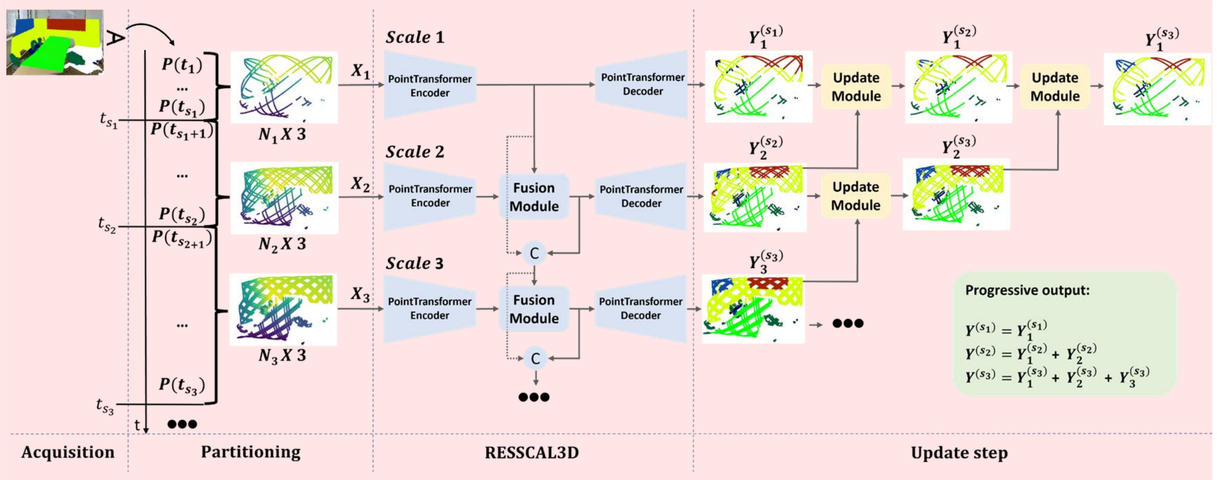
Sample until t = t_1 | Sample until t = t_2 | Sample until t = t_3

VX-S3DIS details

- Each point: <xyz, label, timestamp>
- Contains over 7000 samples from 168 rooms
- 11 semantic classes

Dataset Download

Method: RESSCAL3D++

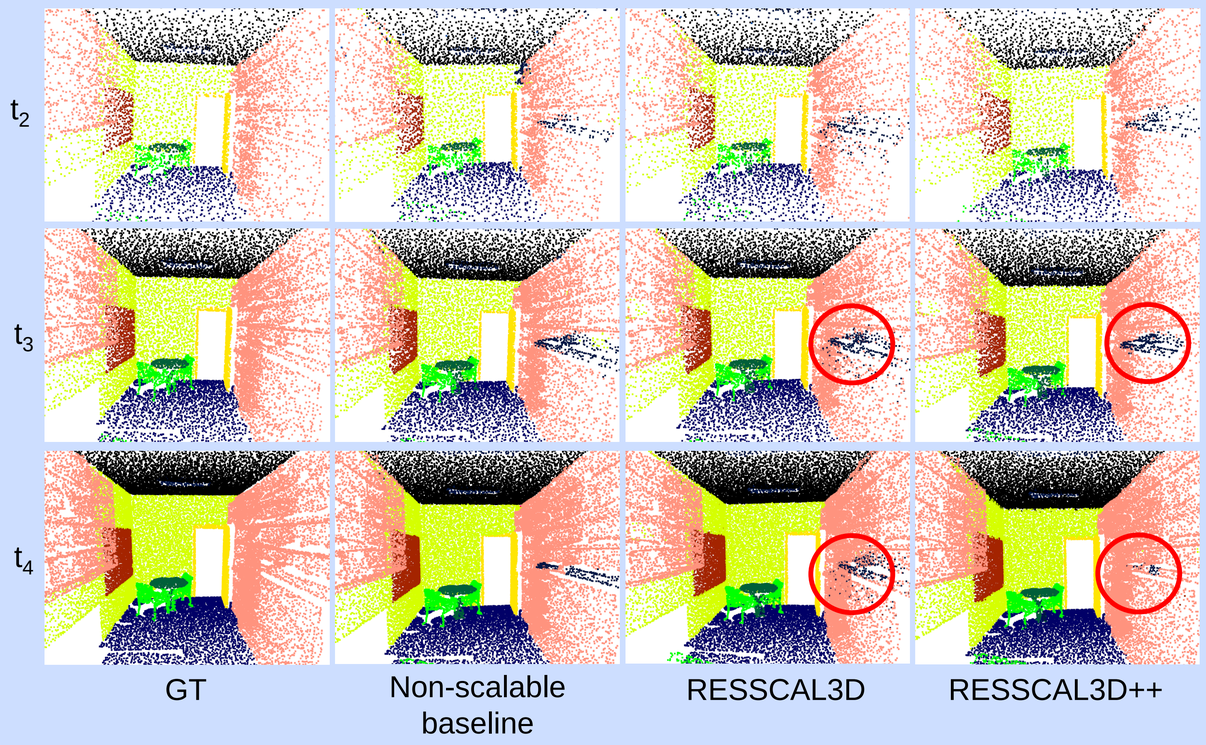


With $N_1 < N_2 < N_3 < \cdots$

- Real-time partitioning of point stream in non-overlapping partitions
- Enables asynchronous inference of scales.
- Features from previous scales are used as prior information
- The whole network is trained scale-per-scale
- An update module and strategy to refine early predictions

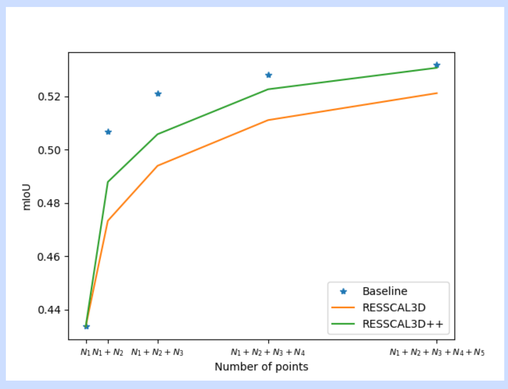
$$Y_i^{(s_{i+2})} = UM\left(Y_i^{(s_{i+1})}, Y_{i+1}^{(s_{i+2})}\right) = UM\left(UM\left(Y_i^{(s_i)}, Y_{i+1}^{(s_{i+1})}\right), UM\left(Y_{i+1}^{(s_{i+1})}, Y_{i+2}^{(s_{i+2})}\right)\right)$$

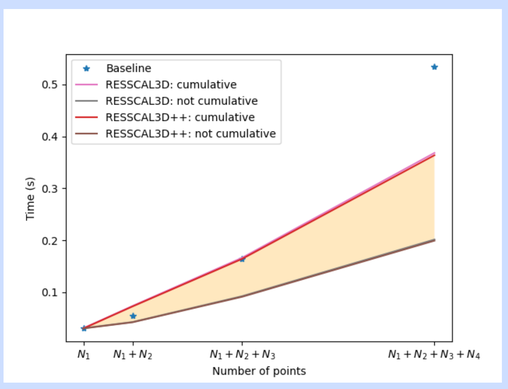
Experimental Results: S3DIS



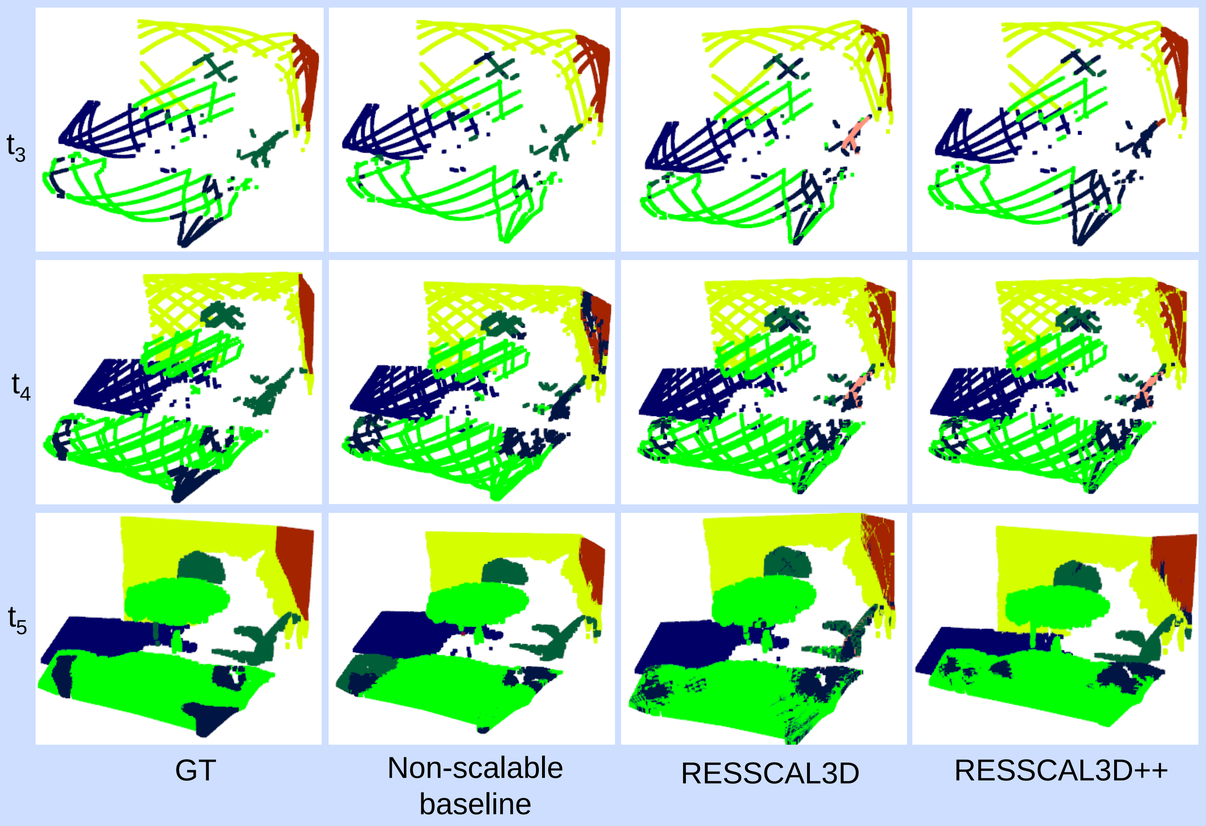
- RESSCAL3D++ refines earlier misclassifications

Experimental Results: VX-S3DIS





- RESSCAL3D++ reduces cost of scalability from 2.0% to 0.2%
- Earliest prediction at only 7% of total inference time
- Inference time at highest scale reduced by 15.6-63.9%
  - Actual inference latency depends on acquisition time



- RESSCAL3D++ reduces scale inconsistencies.

Conclusions

- VX-S3DIS: the first semantic segmentation dataset which mimics the working of a resolution-scalable 3D sensor
- **Joint acquisition and processing of point clouds**
  - Allows early decision-making
  - 15-64% faster inference than traditional methods
- RESSCAL3D++ reduces cost of scalability to 0.2%