LiDAR Localization and Perception for Autonomous Systems

Jean-Emmanuel Deschaud

Associate Professor at Centre for Robotics at Mines Paris – PSL

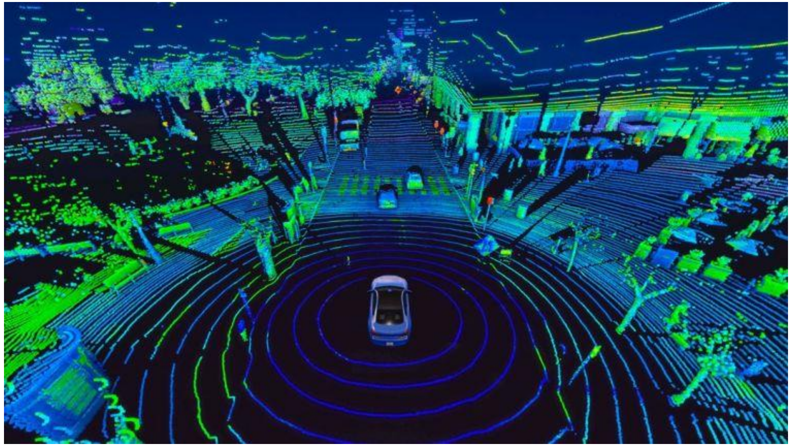
jean-emmanuel.deschaud@minesparis.psl.eu

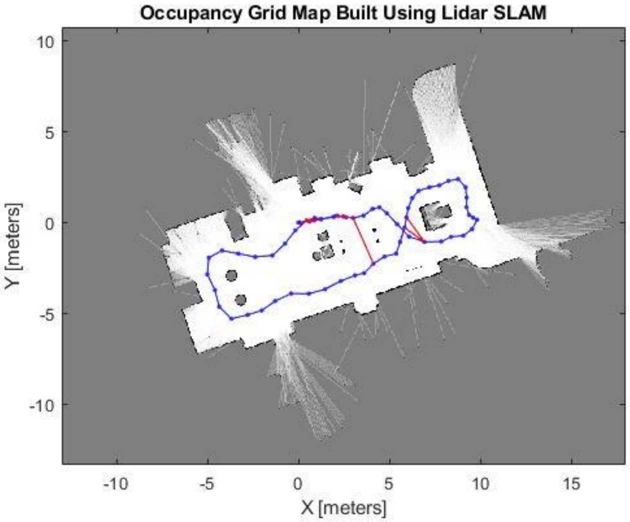
Part I : LiDAR Localisation

Terminology

LiDAR : **Li**ght **D**etection **A**nd **R**anging

SLAM : **S**imultaneous **L**ocalization **A**nd **M**apping





LiDAR sensors and their evolution

- Improved performance and lower prices

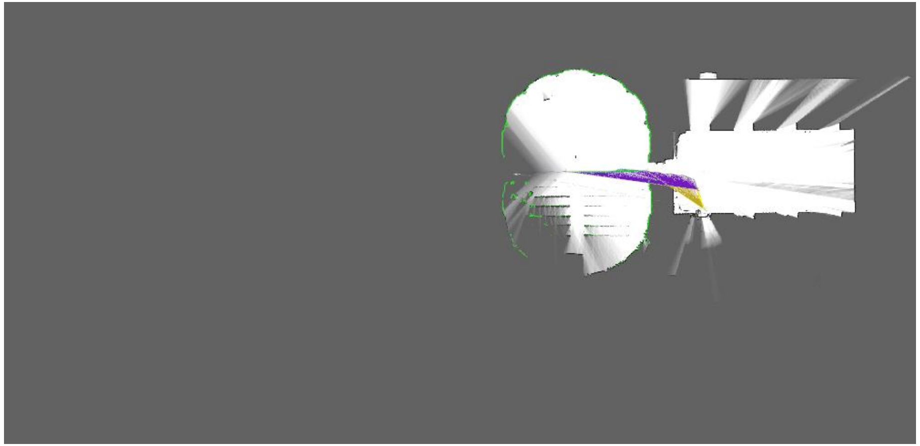
SICK LMS221
Weight 9kg
14k pts/s
~ 4-5 cm accuracy
3k euros in 2005

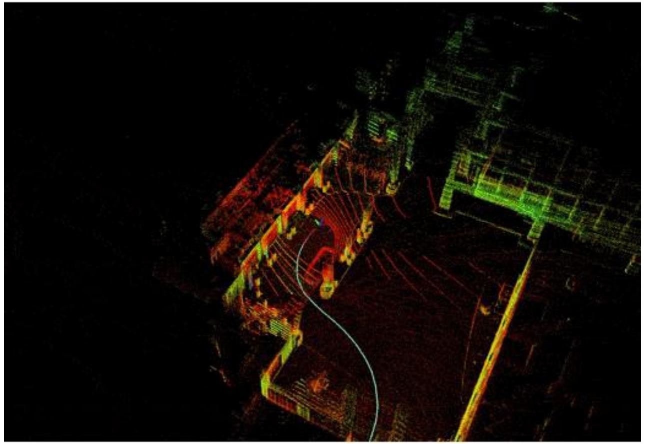
Velodyne HDL64
Weight 12kg
1,3M pts/s
~ 5-10 cm accuracy
120k euros in 2010

Hesai XT32
Weight 1kg
640k pts/s
~ 2 cm accuracy
4k euros in 2022

Objectives of LiDAR SLAM

- 2D sensors with 2D world moving towards a 3D world with 3D sensors



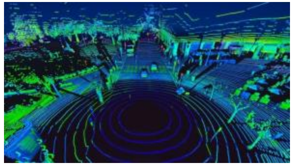


SLAM LiDAR 2D
Flat world
3 degrees of freedom

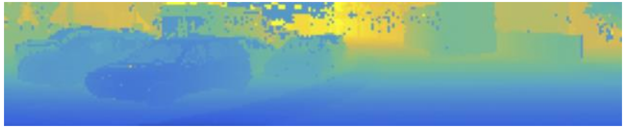
SLAM LiDAR 3D
Non flat world
6 degrees freedom

Architecture of a LiDAR SLAM

New Scan



Point cloud (3D)



Depth Image(2D)

Initialization of movement

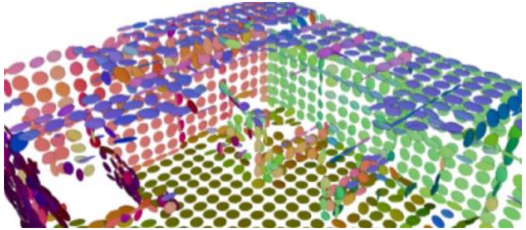
Association of points and neighborhoods

Iterations

Pose estimation

Update of the map

Map

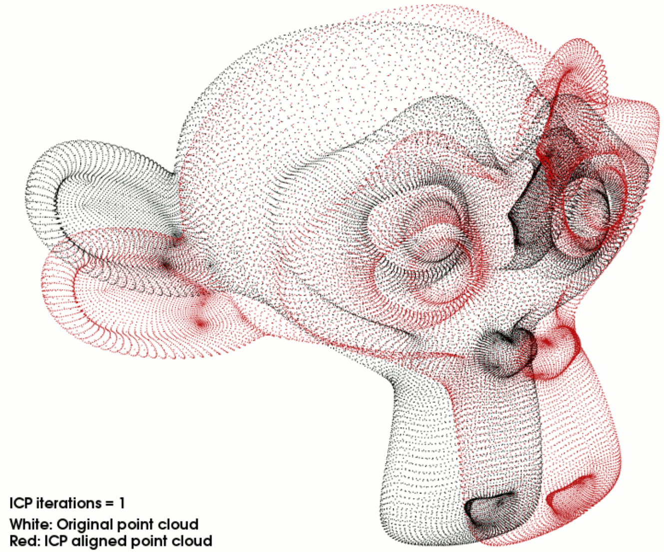


Iterative Closest Point (ICP)

- Besl and McKay article from 1992
- Based purely on geometry

- Method :
  - Two point clouds: one fixed and the other moving
  - Two steps at each iteration:
    - Nearest point neighborhood search
    - Estimation of the transformation that minimizes the distance

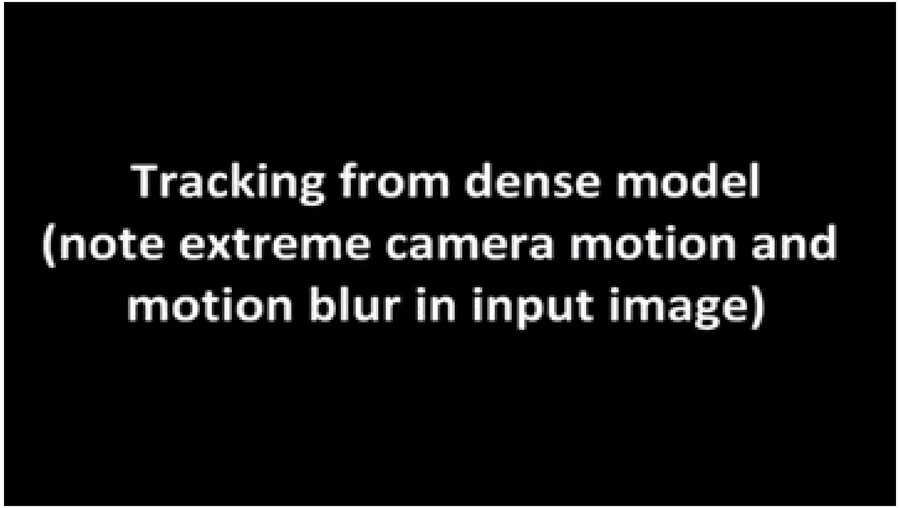
- Disadvantage: need good initialization
  -> Not too problematic for SLAM LiDAR



Besl, P. J., McKay N.D., "A Method for Registration of 3-D Shapes", T-PAMI, 1992

Improvements since 1992

- Initialization
  - Constant Velocity Model
  - Use another sensor: IMU
- Fast neighborhood search
  - Projective ICP
  - Subsampling
  - Map structure
- Map structure
  - Voxels
  - Surfels
  - Mesh
  - Implicit representation
  - Neural representation



Two examples of recent LiDAR SLAM:

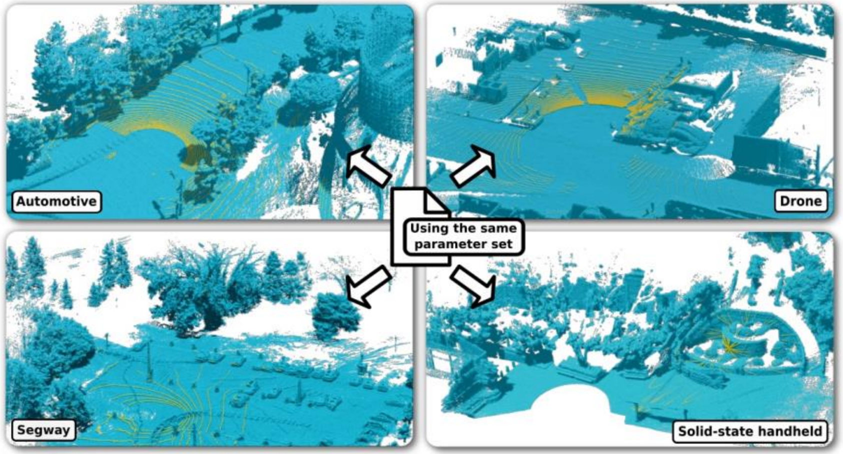
- KISS-ICP
- CT-ICP

Izadi et al, "KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera", ACM symposium on User interface software and technology, 2011.

KISS-ICP (2023)

- **KISS-ICP: K**eep **I**t **S**mall and **S**imple **ICP**
- University of Bonn
- Code available on Github: https://github.com/PRBonn/kiss-icp

- Map: a point cloud
  inside a sparse voxel grid
- Point-to-point ICP
- 1 pose per scan (6 degrees of freedom)

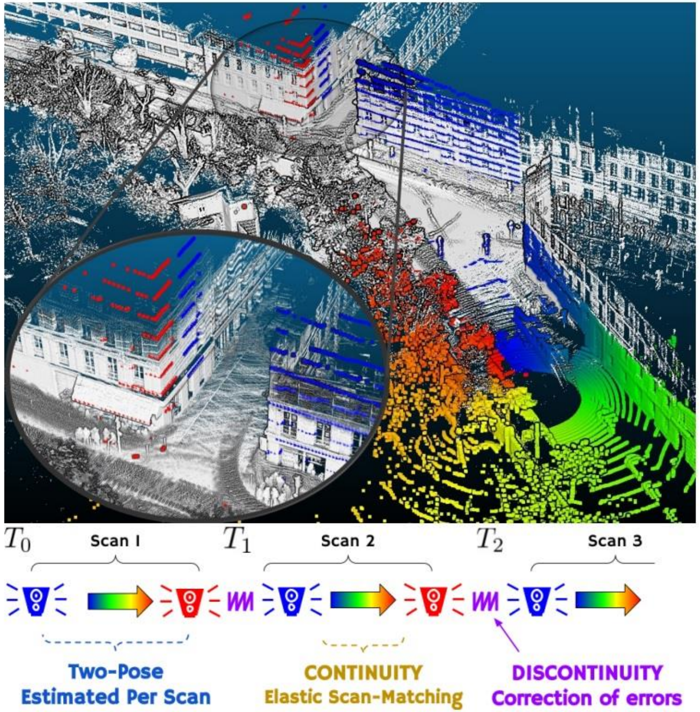


Vizzo et al, "KISS-ICP: In Defense of Point-to-Point ICP - Simple, Accurate, and Robust Registration If Done the Right Way", RA-L, 2023.

CT-ICP (2022)

- **CT-ICP: C**ontinuous-**T**ime **ICP**
- Mines Paris – PSL with Kitware
- Github: https://github.com/jedeschaud/ct_icp

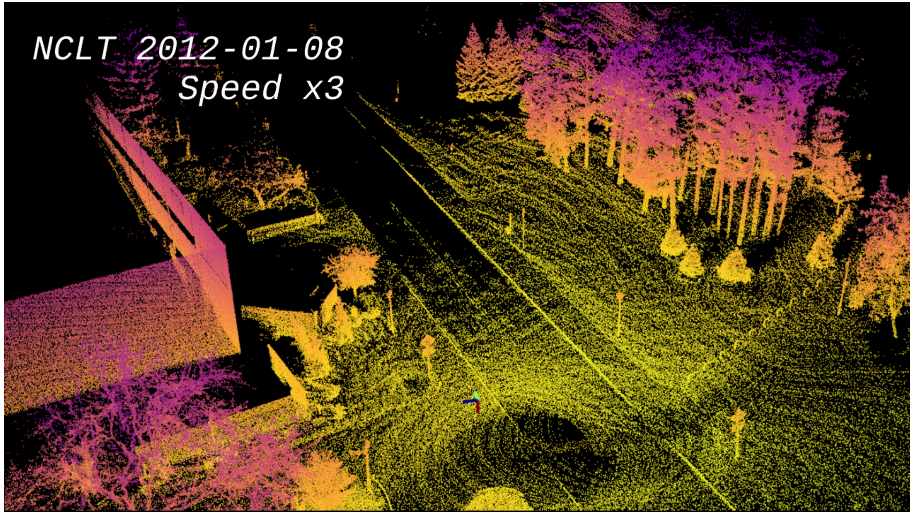
- Map: a point cloud inside a sparse voxel grid
- Point-to-plane ICP
- 2 poses per scan (12 degrees of freedom)



Dellenbach et al, "CT-ICP: Real-time Elastic LiDAR Odometry with Loop Closure", ICRA, 2022.

CT-ICP (2022)

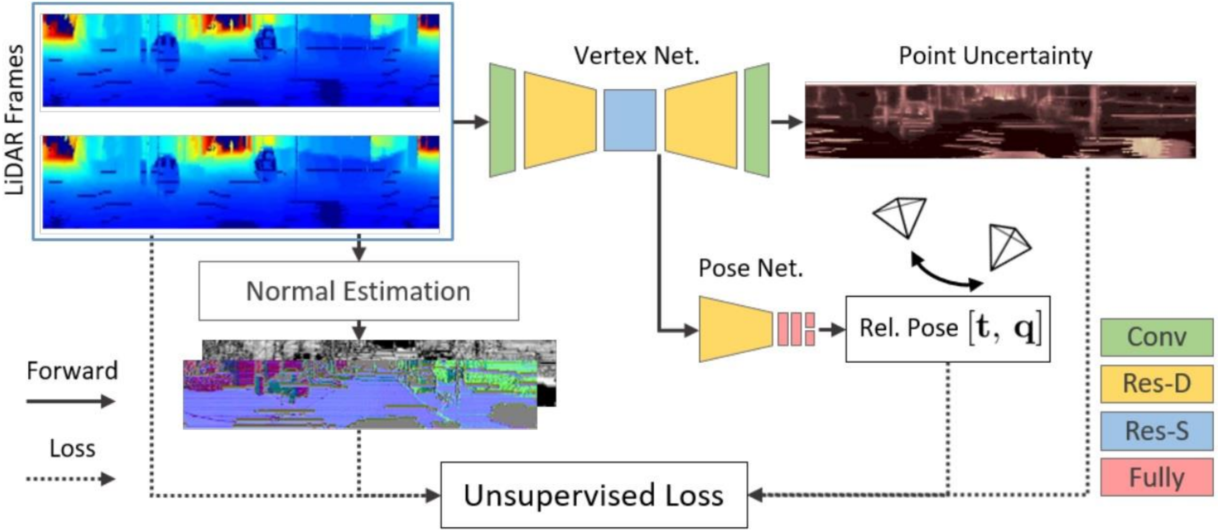
- Qualitative results on NCLT Dataset:
  NCLT Dataset : The University of Michigan North Campus Long-Term Vision and LiDAR Dataset



Dellenbach et al, "CT-ICP: Real-time Elastic LiDAR Odometry with Loop Closure", ICRA, 2022.

SLAM with neural networks

- Method DeepLO:

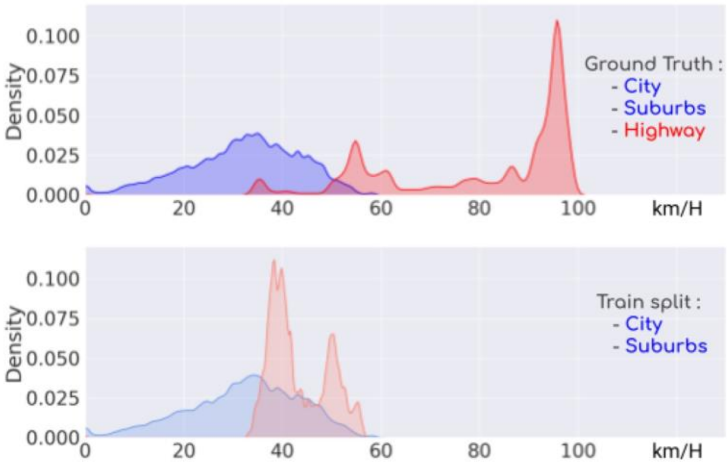


Cho et al, "Unsupervised Geometry-Aware Deep LiDAR Odometry", ICRA, 2020

Geometric SLAM vs Deep SLAM

- Study and comparison:

	00–08*	09–10
PoseNet (ICP loss)	2.24	7.65
PoseNet (Weighted ICP loss)	1.49	7.19
NI + P-F2F	40.1	30.4
CV + P-F2F	1.46	1.7
EI + P-F2F	1.47	1.9
NI + Kd-F2F	24.18	14.04
CV + Kd-F2F	1.41	1.84
EI + Kd-F2F	1.41	1.87



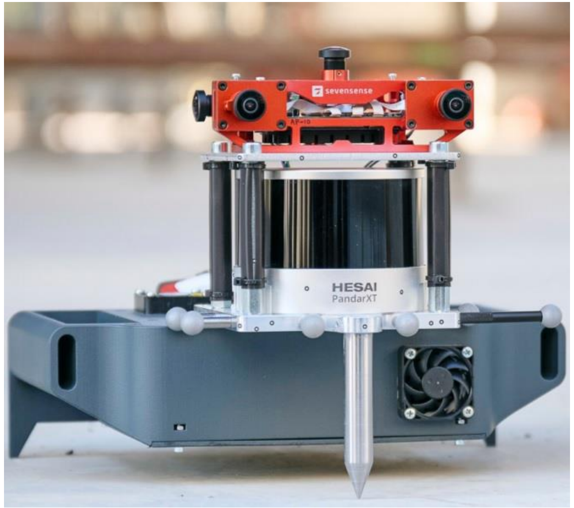
Dellenbach et al, "What's in My LiDAR Odometry Toolbox?", IROS, 2021

Evaluation: HILTI SLAM Challenge

- Datasets and challenges organised by HILTI (company in construction)
- SLAM in difficult environments with real conditions (mainly construction sites)

- 3 challenges :
  - IROS2021 (October 2021)
  - ICRA2022 (May 2022)
  - ICRA2023 (May 2023)

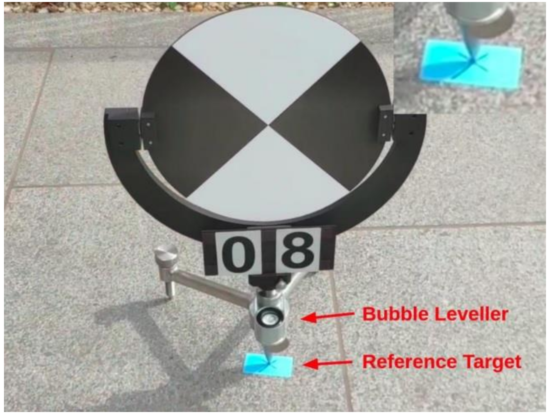
- Sensors in session 2022
  - Cameras Sevensense Alphasense
  - LiDAR Hesai PandarXT-32
  - IMU Bosch BMI085



Zhang et al, "Hilti-Oxford Dataset: A Millimeter-Accurate Benchmark for Simultaneous Localization and Mapping", RA-L, 2023.

HILTI SLAM Challenge 2022

- 16 sequences (between 2min et 18min)
- Quantification of the quality of SLAM LiDAR
  - Topographic quality laser surveys
  - 95% of positionning scans with uncertainty less than 3mm

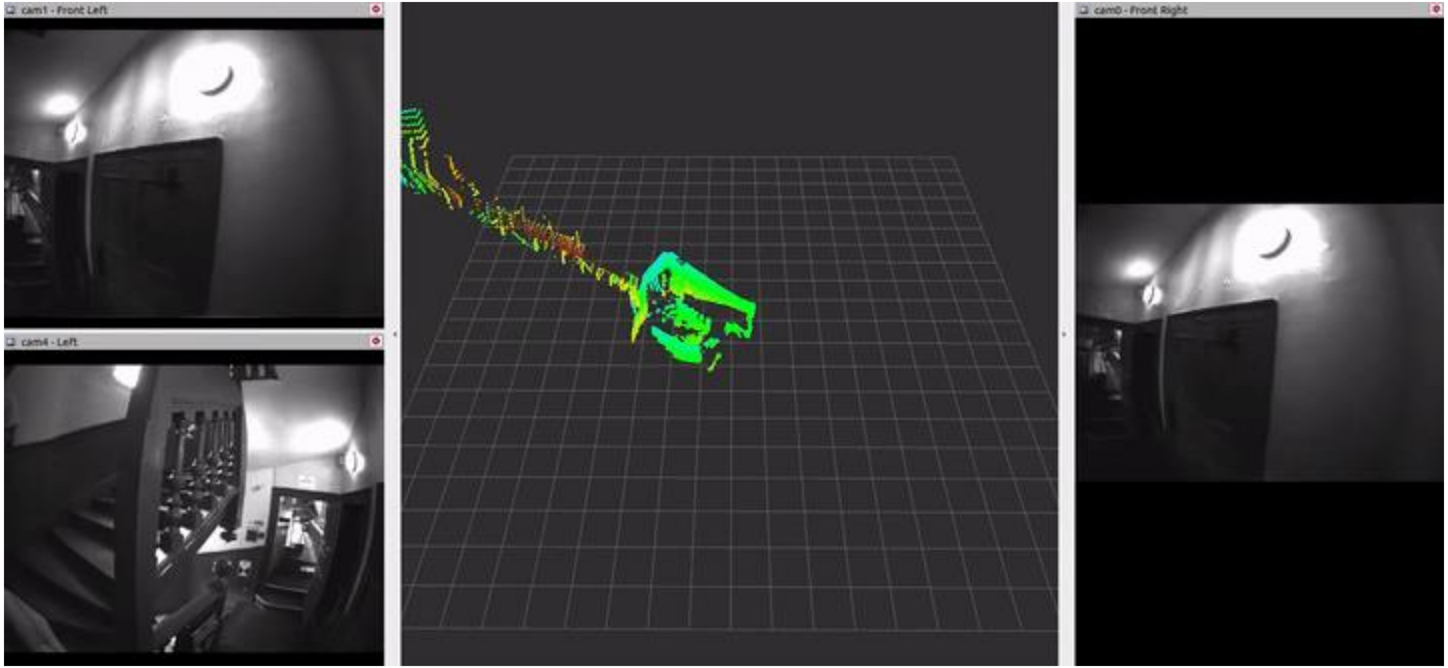


HILTI SLAM Challenge 2022

- Sequence « Sheldonian Theatre » inside Oxford University (~ 6min)
- Considered the most difficult of the 2022 Challenge

HILTI SLAM Challenge 2022

- Sequence « Sheldonian Theatre » inside Oxford University (20s sample)



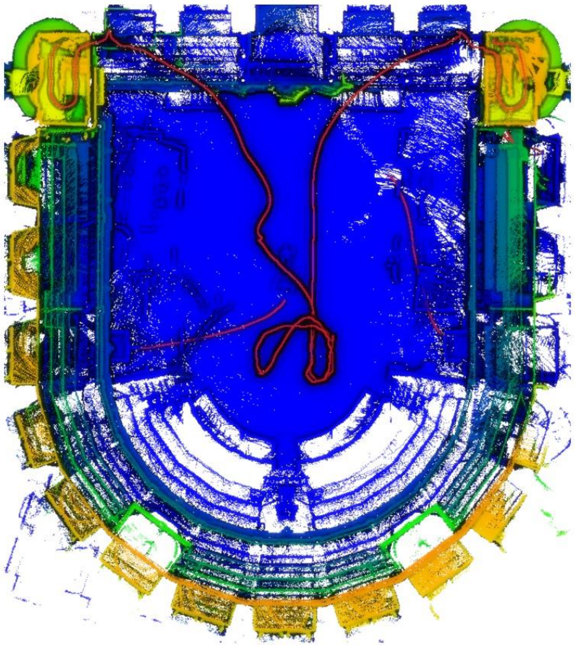
HILTI SLAM Challenge 2022

- Results:
    - 42 submissions from academic and industrials groups
    - SLAM LiDAR more accurate than SLAM based on cameras
    - IMU essential to manage rapid movements
    - Quality higher than organizations’ expectations -> modification of the points system

	Lead Organization	Algorithm	Sensors Used			Odometry		SLAM			Same	Results	
			Lidar	IMU	Cam. (#)	Type	Real-Time	Global BA	Causal	LC	Params	ATE	Score
1	CSIRO	Wildcat SLAM [15]	✓	✓		SW Opt.	✓	✓	✗	✓	✓	2.07	563.8
2	Vision & Robotics	MC2SLAM [17]	✓	✓		SW Opt.	✓	✓	✗	✓	✓	3.94	443.8
3	HKU	FastLIO2[18], BALM [19]	✓	✓		Filter	✓	✓	–	✓	–	5.94	400.4
4	KAIST	Based on [18], [20]	✓	✓		Filter	–	✓	✗	✓	–	19.02	317.5
5	Beihang Uni.	Based on [18], [21]	✓	✓	✓(2)	Filter	✓	✗	✓	✗	✗	22.59	311.6
6	Luxembourg Uni.	Based on [18], [22]	✓	✓	✓(1)	Filter	✓	✗	✓	✗	✓	20.49	303.8
7	MINES ParisTech	CT-ICP [23]	✓	✓		Opt.	✓	✗	✓	✓	–	7.72*	272.8
8	AIST	VITAMIN-E [24], [25]	✓	✓	✓(3)	SW Opt.	✓	✗	✓	✗	✓	16.16	260.5
9	HKUST & Georgia Tech	Based on [18], [26]	✓	✓	✓(5)	Filter	✓	✓	✗	✓	✓	47.50	257.6
10	KTH & NTU	VIRAL SLAM [27]	✓	✓		SW Opt.	✗	✗	✓	✗	✓	6.90	251.9
		Vision-only Results											
1	TUM	OKVIS2.0 [16]		✓	✓(5)	SW Opt.	✓	✓	✗	✓	✓	25.36	32.5
2	Stuttgart Uni. & TUM	Based on [28]		✓	✓(4)	SW Opt.	✗	✗	✓	✗	✗	42.04	22.2

HILTI SLAM Challenge 2022

- Results: Wildcat SLAM with 4cm error on average on sequence « Sheldonian Theatre »



Application: mapping of cities

- L3D2 prototype at Centre for Robotics at Mines Paris - PSL

Application: mapping of campus

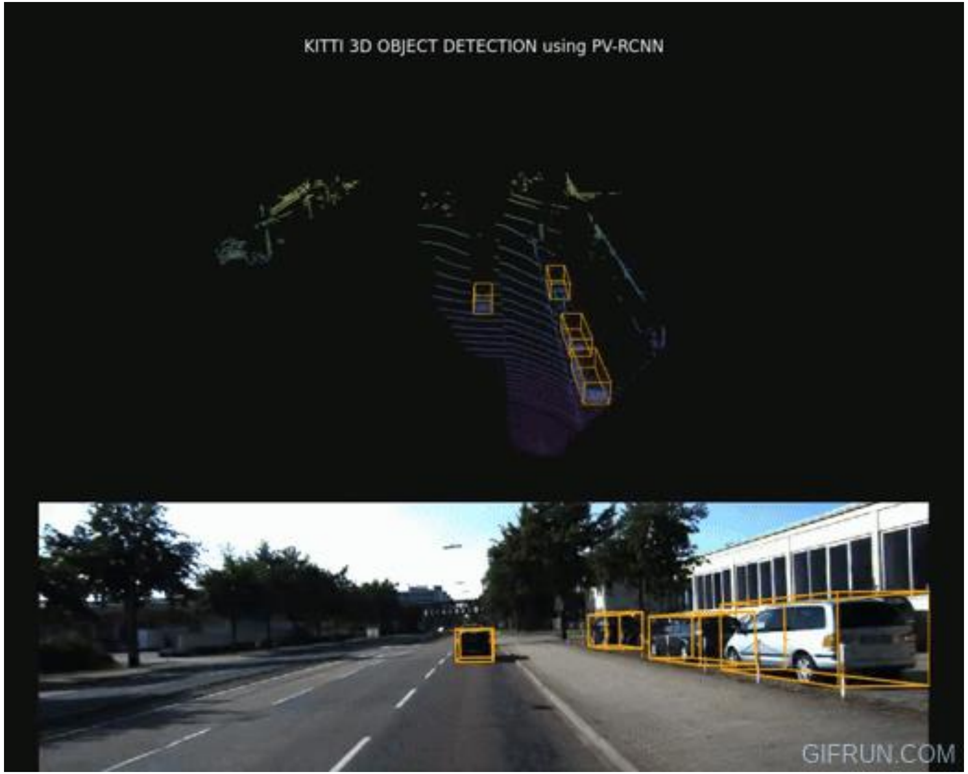
- Handheld system Rock Robotic R3 Pro

https://cloud.rockrobotic.com/share/463abbf8-5ee6-4f5c-81c5-6f6154cca039

Part II : LiDAR Perception

Tasks for LiDAR-based Perception

- **Semantic Segmentation**
- Object Detection
- Instance Segmentation
- Panoptic Segmentation
- Tracking



The goal of LiDAR Semantic Segmentation

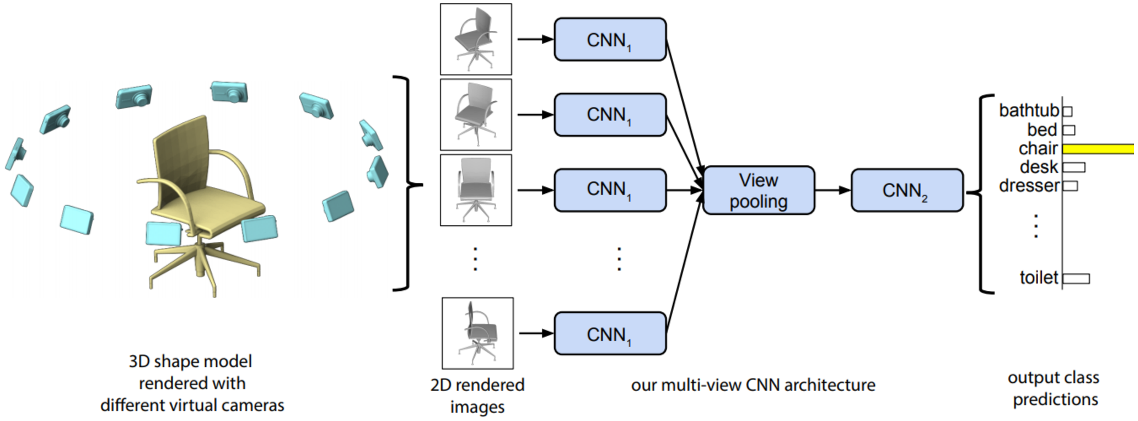
- Predict a class for each 3D point

How to do deep learning on point cloud?

- Projection on images
- Convolution on a sparse grid of voxels
- Convolution on a point cloud

Projection on images

- Project on cameras:
  - Random view, LiDAR sensor view, bird's-eye view
  - Perspective, parallel, cylindrical projection…
  - Keeping depth, accumulation, RGB data…
- Use classic 2D convolutional neural networks



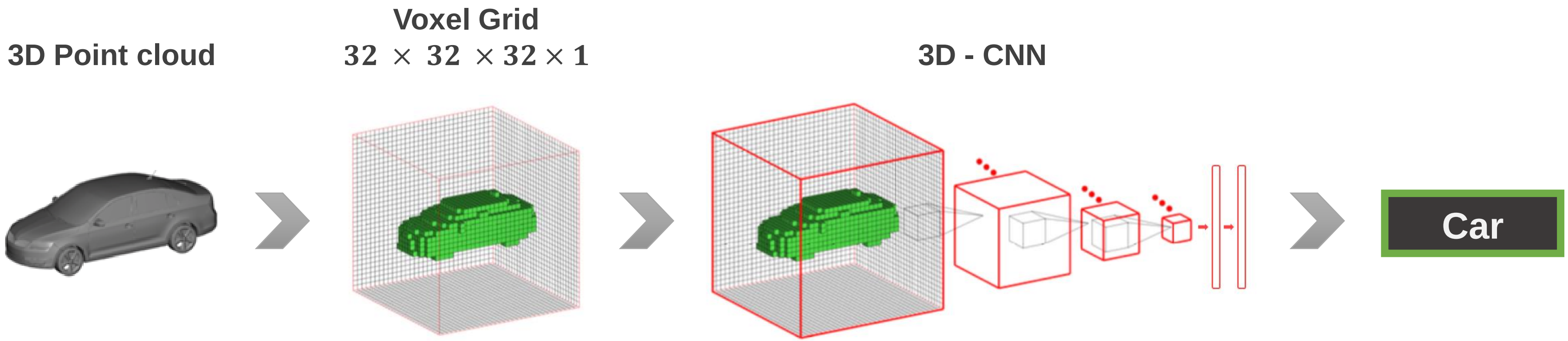
Projection on images

- RangeNet++: LiDAR Semantic Segmentation using range images
  - Loss of information in the projection

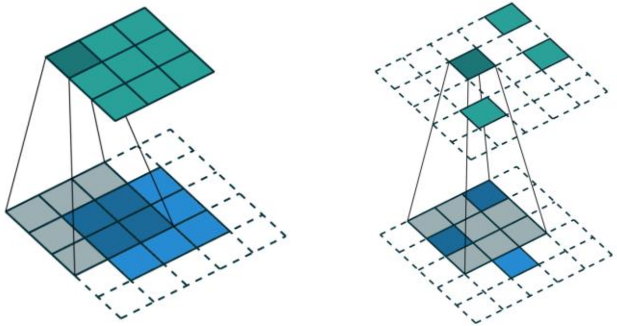
Milioto et al, "RangeNet++: Fast and Accurate LiDAR Semantic Segmentation", IROS, 2019.

Convolution on a sparse grid of voxels

- Point cloud transformed into an occupancy grid (voxels)

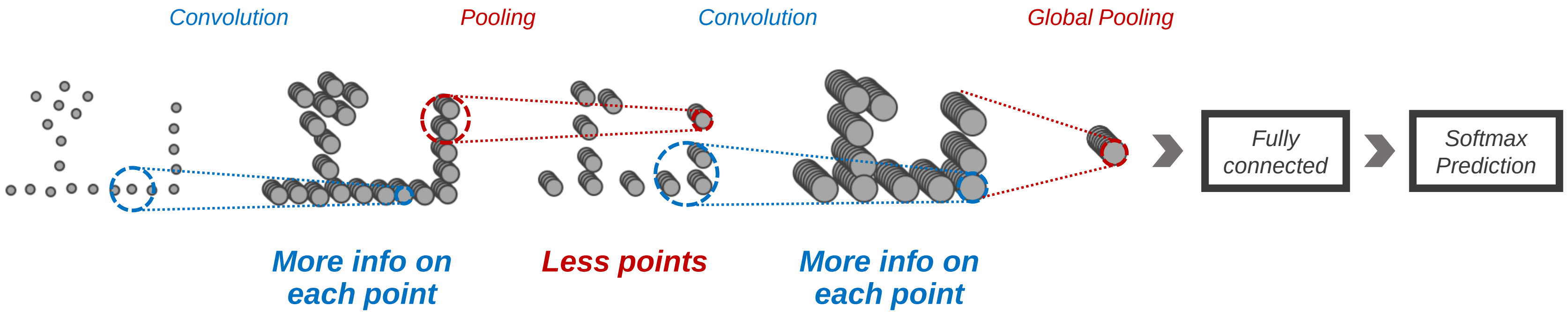


- Replace dense operations with sparse operations
  - Example SRU-Net as Sparse Residual U-Net



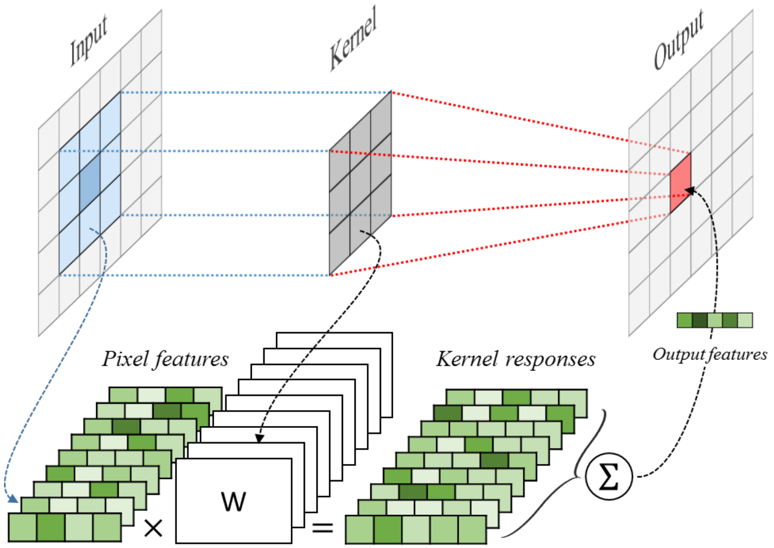
Convolution on a point cloud

- Point cloud as support for features inside the network

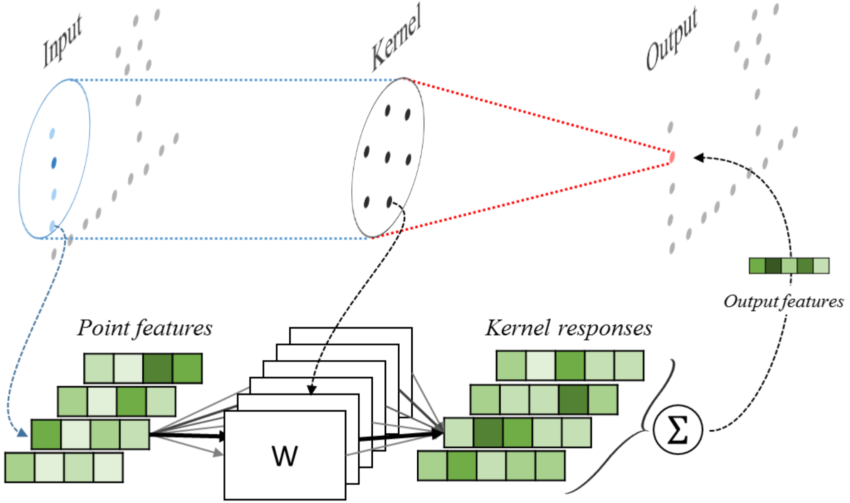


Convolution on a point cloud

- KPConv: point-based convolutions



Hard alignment of the weights based on input pixel positions



Soft alignment of the weights based on the correlation between input points and kernel points

Thomas et al, "KPConv: Flexible and Deformable Convolution for Point Clouds", ICCV, 2019.

Convolution on a point cloud

- KPConv results

Thomas et al, "KPConv: Flexible and Deformable Convolution for Point Clouds", ICCV, 2019.

Convolution on a point cloud

- KPConv results on Paris-Lille-3D:
  - 82.0% mIoU (jaccard index averaged by class)
  - 95.3% overall accuracy (on all points)

Thomas et al, "KPConv: Flexible and Deformable Convolution for Point Clouds", ICCV, 2019.

LiDAR-based Perception in autonomous driving

- 3D neural networks obtain good results on datasets in autonomous driving
  - 60 - 70 mIoU on SemanticKITTI
  - 70 - 80 mIoU on nuScenes

- Two main issues:
  - 3d neural networks like SRU-Net or KPConv are slow
  - Not robust: a model trained on one dataset does not perform well on another dataset

- Dealing with speed:
  - Decrease the size of the network
  - Knowledge distillation
  - Apply 3d neural networks on some parts of the scene (-> 3DLabelProp)

- Dealing with robustness
  - **Domain Generalization** for LiDAR Semantic Segmentation

What is Domain Generalization?

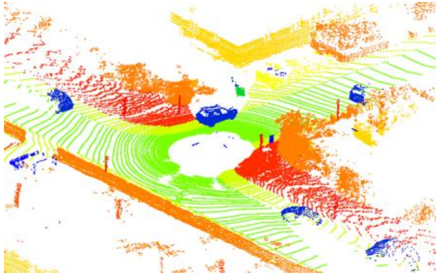
- Definition:
  - “*The idea of Domain Generalization is to learn from one or multiple training domains, to extract a domain-agnostic model which can be applied to an unseen domain*”

    (Definition taken from paperswithcode.com)

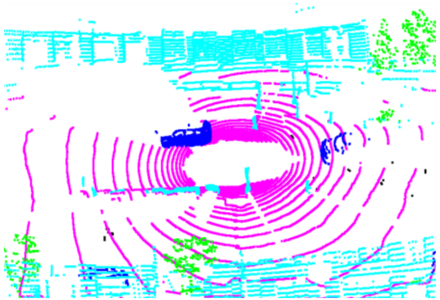
- Main domain shifts in LiDAR perception for autonomous driving
  - Scene shifts
  - Appearance shifts
  - Sensor shifts

What is Domain Generalization?

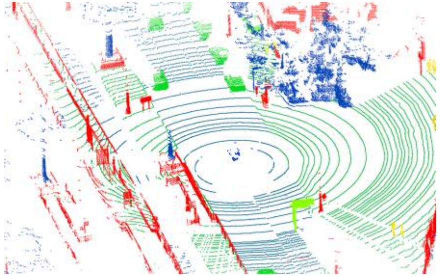
- Example of datasets in autonomous driving with some domain shifts



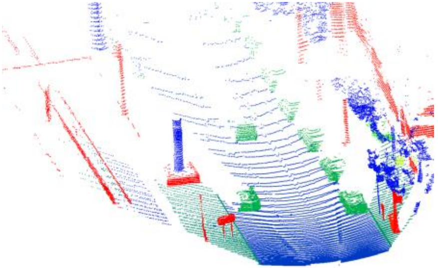
SemanticKITTI
(Velodyne HDL64)



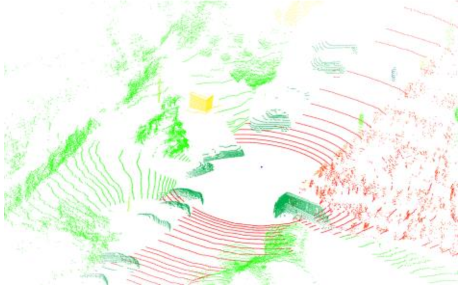
nuScenes
(Velodyne HDL32)



PandaSet
(Hesai Pandar64)



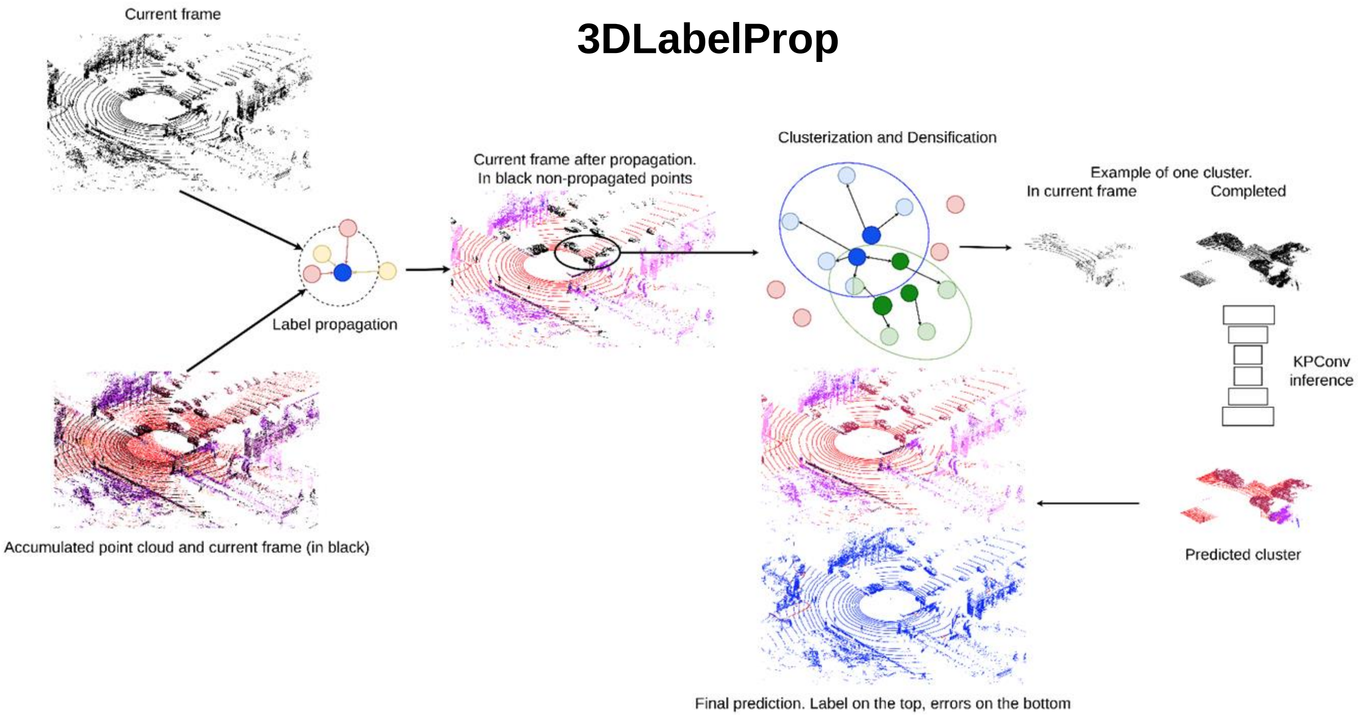
PandaSet
(Hesai Solid-State Pandar-150)



SemanticPOSS
(Hesai Pandora-40)

Sensor-agnostic LiDAR Semantic Segmentation

3DLabelProp



Sanchez et al, "Domain generalization of 3D semantic segmentation in autonomous driving", ICCV, 2023.

Sensor-agnostic LiDAR Semantic Segmentation

- 3DLabelProp quantitative results

Method	Input type	$\text{mIoU}^{NS}_{\mathcal{L}_{NS}}$	$\text{mIoU}^{SK}_{\mathcal{L}_{NS \cap SK}}$	$\text{mIoU}^{SK32}_{\mathcal{L}_{NS \cap SK}}$	$\text{mIoU}^{P64}_{\mathcal{L}_{NS \cap PS}}$	$\text{mIoU}^{PFF}_{\mathcal{L}_{NS \cap PS}}$	$\text{mIoU}^{SP}_{\mathcal{L}_{NS \cap SP}}$
KPConv [30]	Point-based	63.1	44.9	50.6	25.0	16.9	60.7
SPVCNN [28]	Voxel & point-based	67.2	49.4	53.2	43.7	11.1	**64.8**
C3D [47]	Cylindrical voxel-based	70.2	31.7	46.1	15.8	4.7	42.8
3DLabelProp (Ours)	Point-based & 4D	**71.0**	**60.5**	**62.5**	**65.4**	**66.7**	64.3

Generalization results when trained on nuScenes (NS) and tests on nuScenes (NS), SemanticKITTI (SK), SemanticKITTI32 (SK32), Panda64 (P64), PandaFF (PFF), and SemanticPOSS (PS)

Sensor-agnostic LiDAR Semantic Segmentation

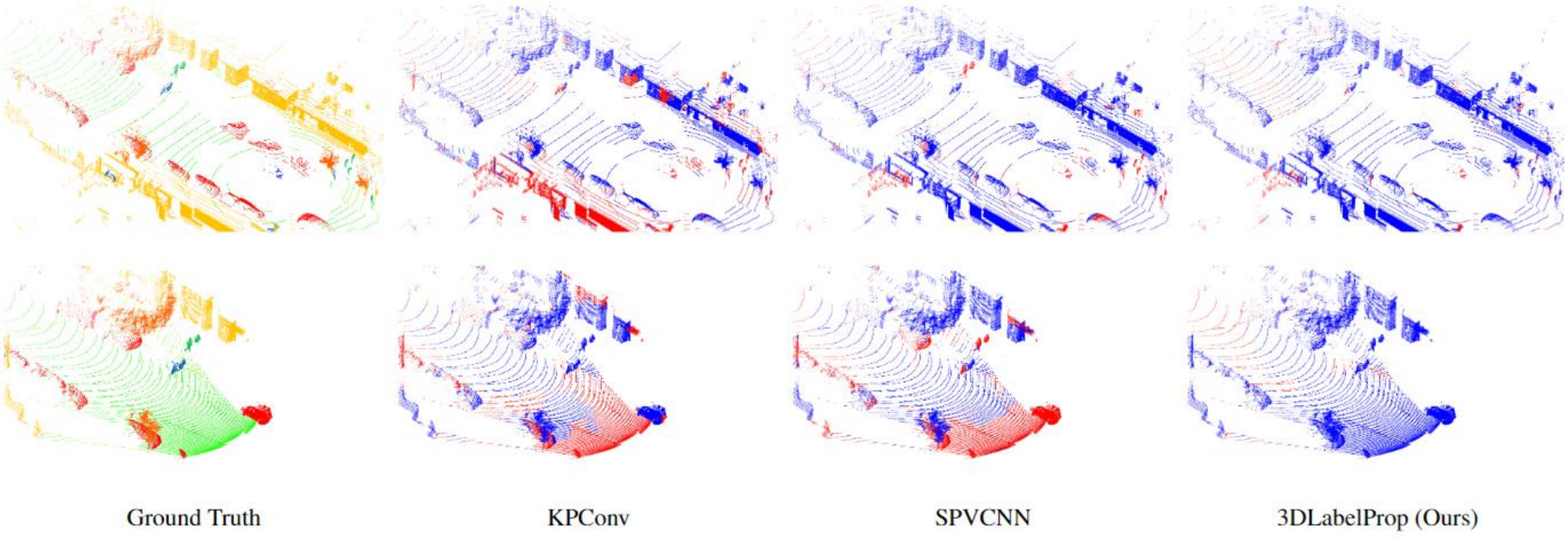
- 3DLabelProp speed

	$\text{mIoU}^{SK}_{\mathcal{L}_{SK}}$	$\text{mIoU}^{SP}_{\mathcal{L}_{SK \cap SP}}$	$\text{mIoU}^{NS}_{\mathcal{L}_{SK \cap NS}}$	FPS
KPConv w/ reflectivity	59.9	33.1	47.6	0.2
KPConv	58.3	39.1	46.7	0.2
KPConv multiframe	53.0	47.2	44.2	0.05
3DLabelProp (Ours)	60.8	50.4	44.4	1.0

Comparison of 3DLabelProp method with KPConv singe-frame and KPConv multi-frame

Sensor-agnostic LiDAR Semantic Segmentation

- 3DLabelProp qualitative results



Qualitative results when trained on SemanticKITTI and tested on Panda64 (top row) and PandaFF (bottom row)

Questions ?