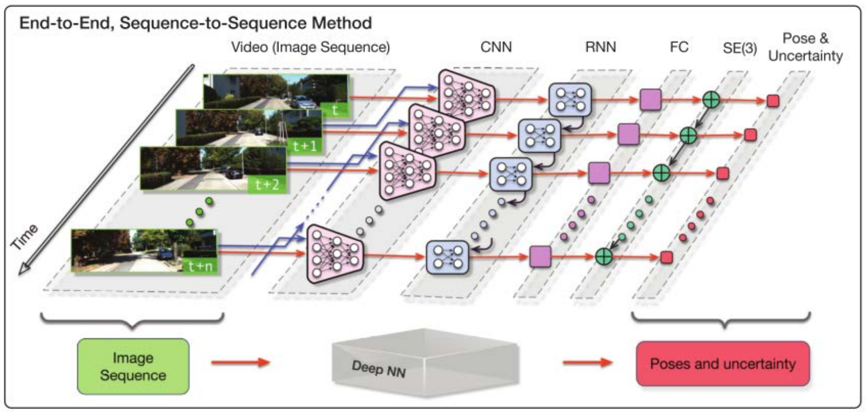
IROS 2019 and Trends of Deep Learning on Robotics

2019.11.27.
임형태
shapelim@kaist.ac.kr

Introduction: Limitation of End-to-end Approaches of Pose Regression



On KITTI Odometry Benchmark Suite

	Method	Setting	Code	Translation	Rotation	Runtime	Environment	Compare
1	F-SLAM			0.00 %	0.0000 [deg/m]	0.01 s	1 core @ 2.5 Ghz (C/C++)	
2	V-LOAM			0.55 %	0.0013 [deg/m]	0.1 s	2 cores @ 2.5 Ghz (C/C++)	
	J. Zhang and S. Singh: Visual-lidar Odometry and Mapping: Low drift, Robust, and Fast. IEEE International Conference on Robotics and Automation(ICRA) 2015.							
3	LOAM			0.57 %	0.0013 [deg/m]	0.1 s	2 cores @ 2.5 Ghz (C/C++)	
	J. Zhang and S. Singh: LOAM: Lidar Odometry and Mapping in Real- time. Robotics: Science and Systems Conference (RSS) 2014.							
4	IMLS-SLAM++			0.61 %	0.0014 [deg/m]	1.3 s	1 core @ >3.5 Ghz (C/C++)	
5	SOFT2			0.65 %	0.0014 [deg/m]	0.1 s	2 cores @ 2.5 Ghz (C/C++)	
	I. Cvišić, J. Ćesić, I. Marković and I. Petrović: SOFT-SLAM: Computationally Efficient Stereo Visual SLAM for Autonomous UAVs. Journal of Field Robotics 2017.							
6	IMLS-SLAM			0.69 %	0.0018 [deg/m]	1.25 s	1 core @ >3.5 Ghz (C/C++)	
	J. Deschaud: IMLS-SLAM: Scan-to-Model Matching Based on 3D Data. 2018 IEEE International Conference on Robotics and Automation (ICRA) 2018.							
7	MC2SLAM			0.69 %	0.0016 [deg/m]	0.1 s	4 cores @ 2.5 Ghz (C/C++)	
	F. Neuhaus, T. Koss, R. Kohnen and D. Paulus: MC2SLAM: Real-Time Inertial Lidar Odometry using Two-Scan Motion Compensation. German Conference on Pattern Recognition 2018.							
⋮								
124	DeepVO			24.55 %	0.0489 [deg/m]	1 s	1 core @ 2.5 Ghz (Python)	

Wang *et al.*, "DeepVO: Towards End to End Visual Odometry with Deep Recurrent Convolutional Neural Networks.," *Proc of the IEEE ICRA, 2017*

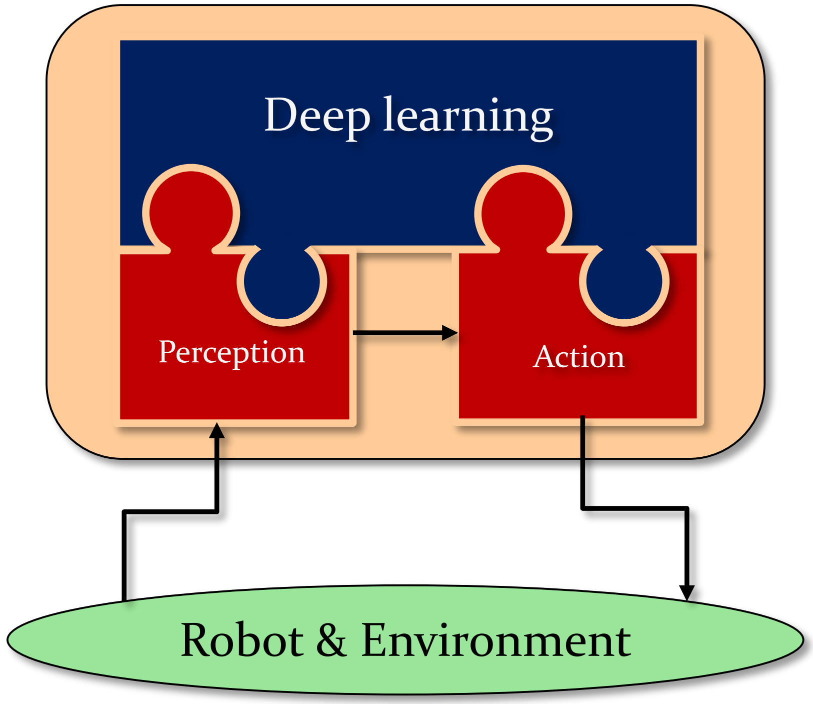
Wang *et al.*, "End-to-end, sequence-to-sequence probabilistic visual odometry through deep neural networks.," *The International Journal of Robotics Research*, 2017

Introduction: Limitation of End-to-end Approaches of Pose Regression

Brown: Trajectories of train data
Green : Trajectories of test data

Sattler *et al.*, "Understanding the Limitations of CNN-based Absolute Camera Pose Regression.," CVPR 2019

Trends of Deep Learning



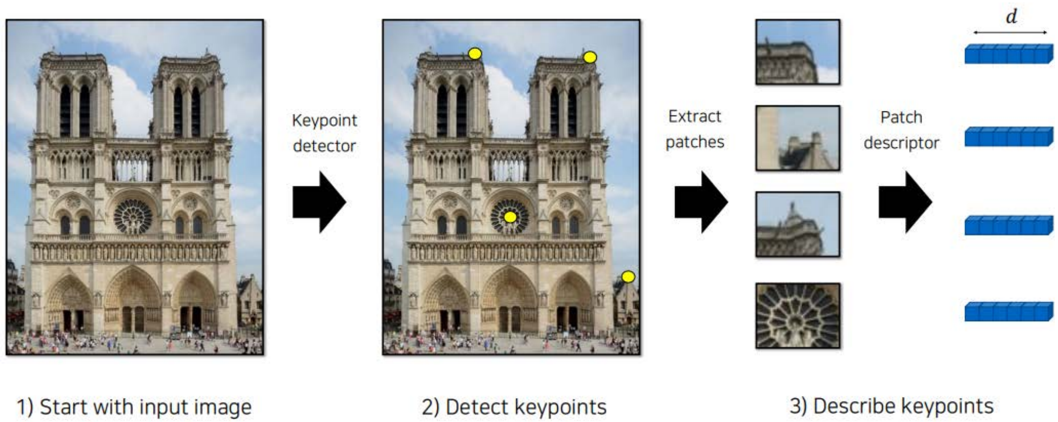
[Picture from A.Saffiotti.]

R2D2: Repeatable and Reliable Detector and Descriptor

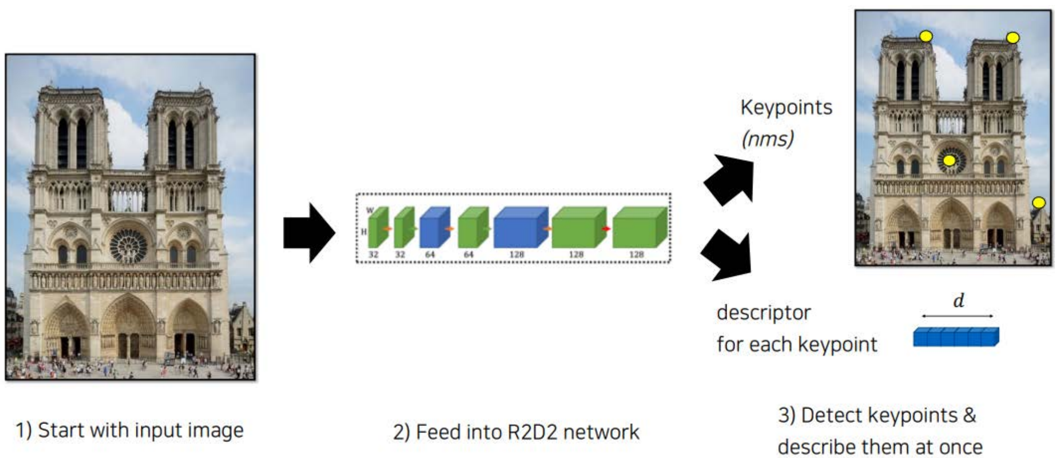
- Motivation
  - Structure-based methods perform well and the critical part is feature extraction and matching
  - A robust feature detector enables robust visual localization
  - ...and improves many other applications such as object detection, VSLAM and SfM

R2D2: Repeatable and Reliable Detector and Descriptor

- Classical methods: *Detect-**then**-describe*



- Their approach: *Detect-**and**-describe*



R2D2: Repeatable and Reliable Detector and Descriptor

- Repeatability
  - Image locations that are invariant to usual image transformations

  (기하학적, 시각적 변화에도 같은 위치에서 keypoints가 검출되는 정도)
- Reliability
  - Image locations that are good (discriminative and robust) for matching purpose *(그 feature descriptor와 주위 영역과 구분이 되는 정도)*

Reliable?

Repeatable?	No	Yes
No	①	②
Yes	③	④

➔ *All cases are possible:*
reliability and repeatability
are independent

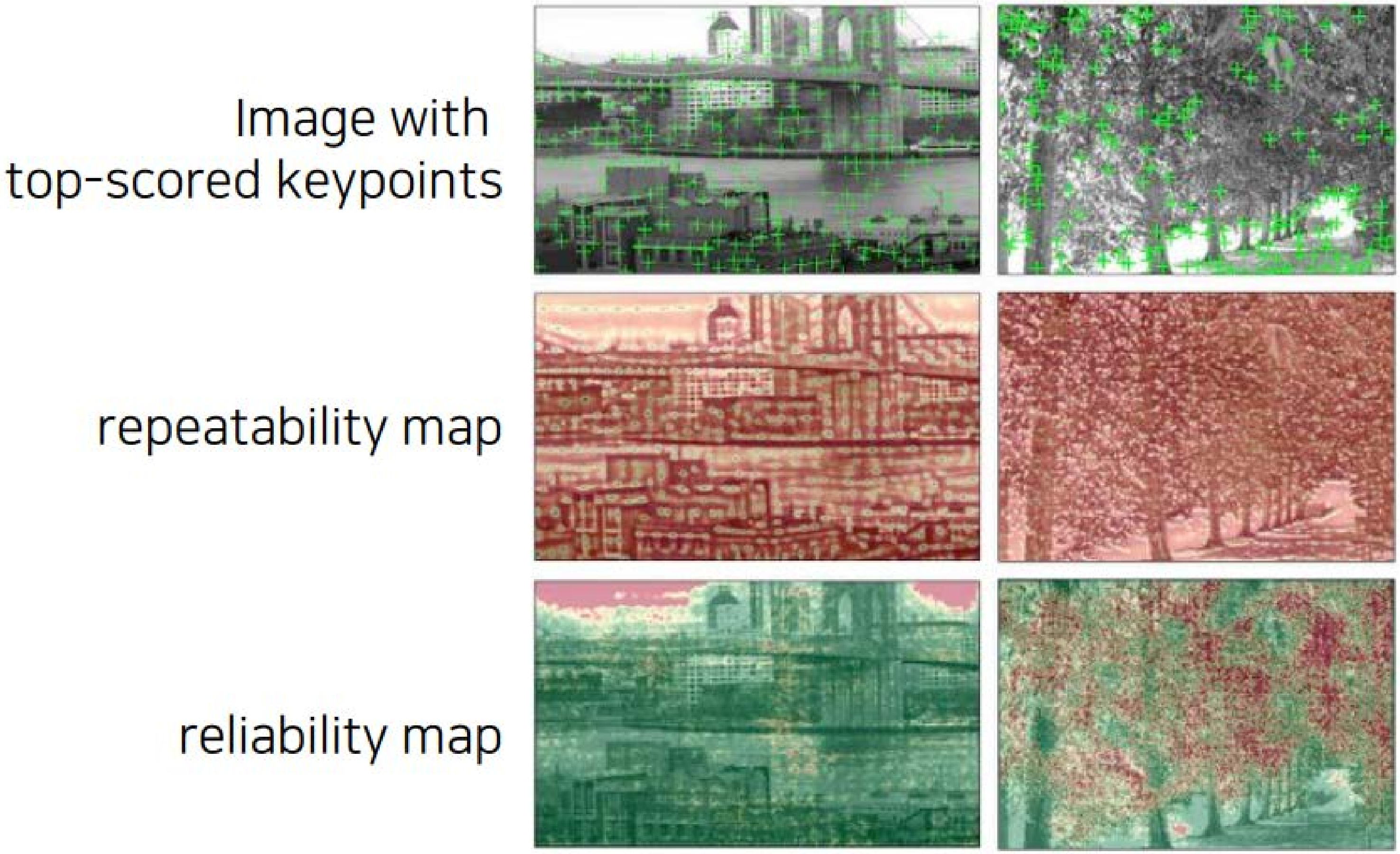
https://deview.kr/data/deview/2019/presentation/[133]Deview2019_Martin_Humenberger_small.pdf

KAIST

URBAN ROBOTICS LAB

R2D2: Repeatable and Reliable Detector and Descriptor

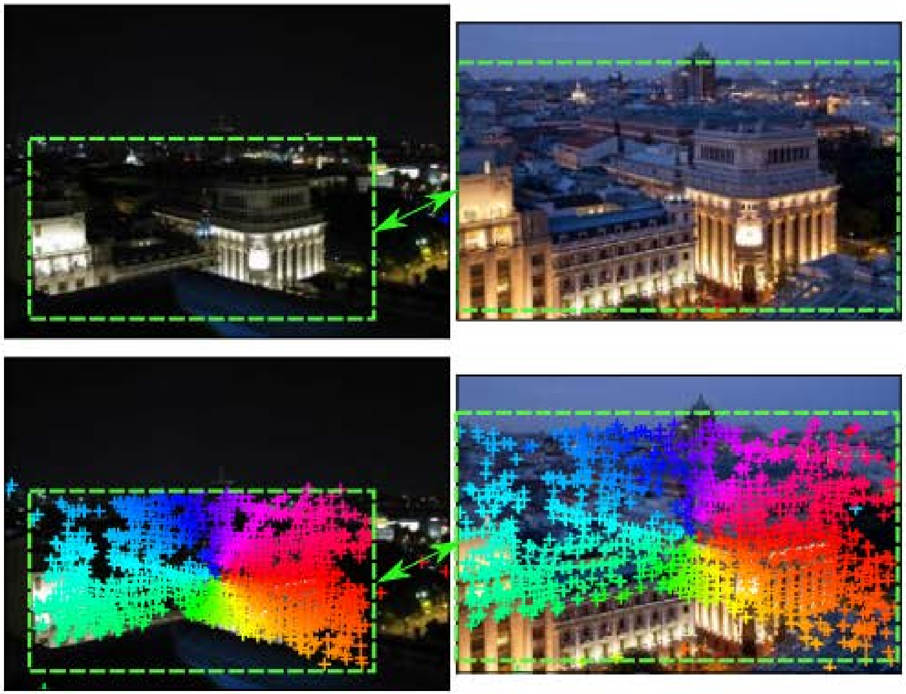
- Results



https://deview.kr/data/deview/2019/presentation/[133]Deview2019_Martin_Humenberger_small.pdf

R2D2: Repeatable and Reliable Detector and Descriptor

- Results

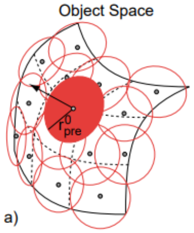


https://deview.kr/data/deview/2019/presentation/[133]Deview2019_Martin_Humenberger_small.pdf

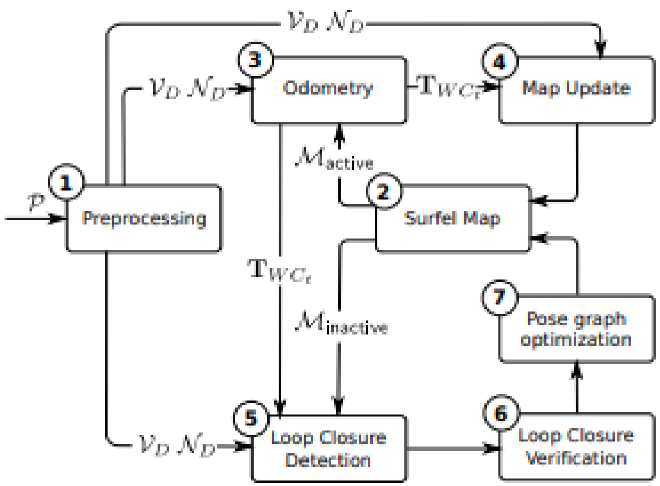
SuMa++: Efficient LiDAR-based Semantic SLAM

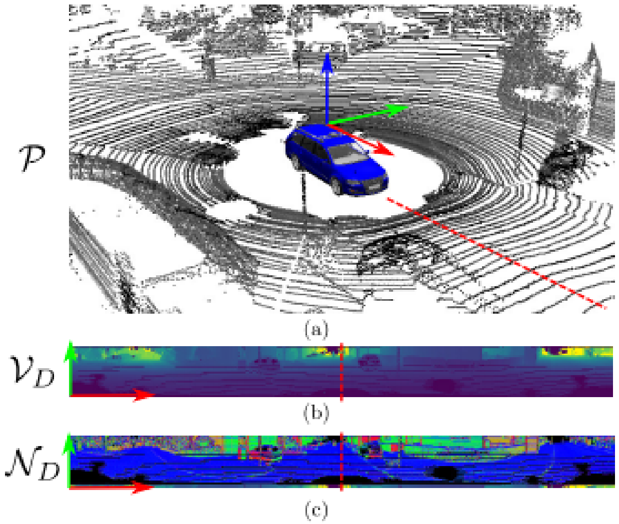
- Previous work
  - Efficient Surfel-Based SLAM using 3D Laser Range Data in Urban Environments.

※**Surfel**: Surface element



$v_s \in \mathbb{R}^3$
$n_s \in \mathbb{R}^3$
$r_s \in \mathbb{R}$





$\mathcal{P}$ Point cloud
$\mathcal{V}_D$ Vertex map
$\mathcal{N}_D$ Normal map

Active: the area of the model most recently observed

Inactive: Older parts of the map which have not been observed in a period of time δ_t

T. Whelan *et al.*, "ElasticFusion: Dense SLAM Without A Pose Graph". *In Proc. of Robotics: Science and Systems*, 2015.

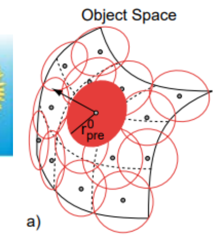
Pfister, Hanspeter, et al. "Surfels: Surface elements as rendering primitives." Proceedings of the 27th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co., 2000.

Behley, Jens, and Cyrill Stachniss. "Efficient Surfel-Based SLAM using 3D Laser Range Data in Urban Environments." Robotics: Science and Systems. 2018.

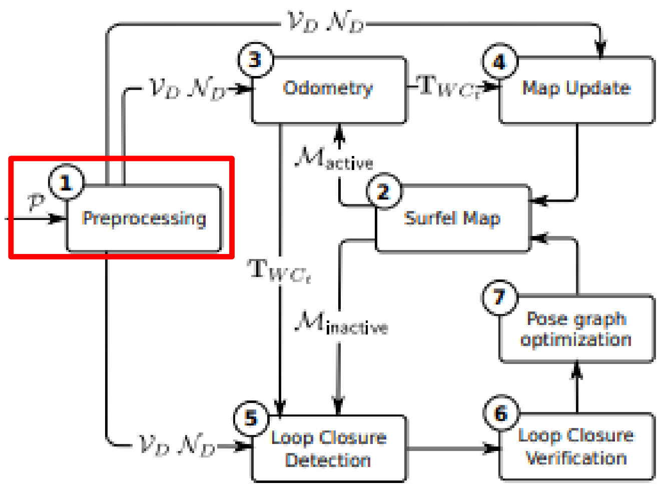
SuMa++: Efficient LiDAR-based Semantic SLAM

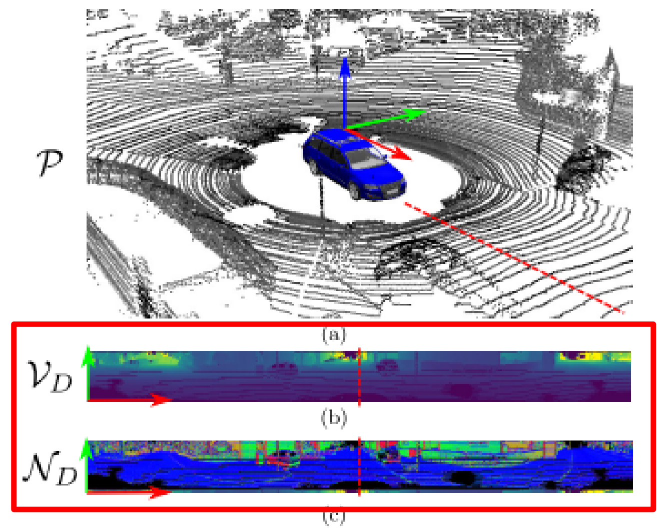
- Previous work
  - Efficient Surfel-Based SLAM using 3D Laser Range Data in Urban Environments.

※**Surfel**: Surface element



$v_s \in \mathbb{R}^3$
$n_s \in \mathbb{R}^3$
$r_s \in \mathbb{R}$





$\mathcal{P}$ Point cloud
$\mathcal{V}_D$ Vertex map
$\mathcal{N}_D$ Normal map

Active: the area of the model most recently observed

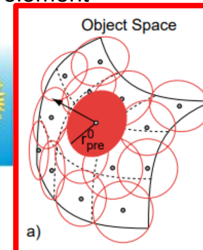
Inactive: Older parts of the map which have not been observed in a period of time δ_t

T. Whelan *et al.*, "ElasticFusion: Dense SLAM Without A Pose Graph". *In Proc. of Robotics: Science and Systems*, 2015.
Pfister, Hanspeter, et al. "Surfels: Surface elements as rendering primitives." Proceedings of the 27th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co., 2000.
Behley, Jens, and Cyrill Stachniss. "Efficient Surfel-Based SLAM using 3D Laser Range Data in Urban Environments." Robotics: Science and Systems. 2018.

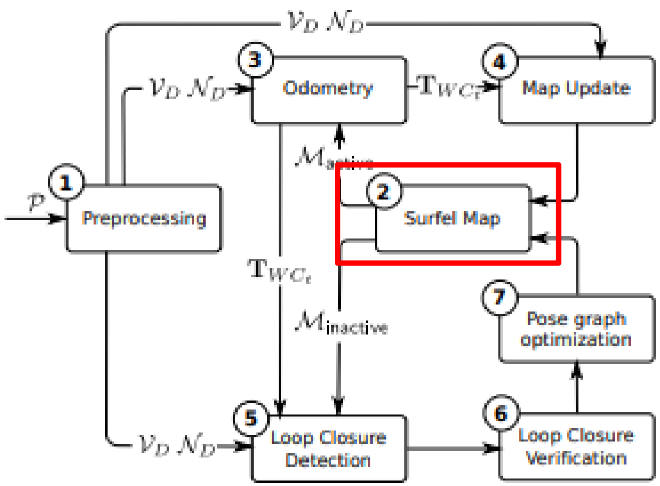
SuMa++: Efficient LiDAR-based Semantic SLAM

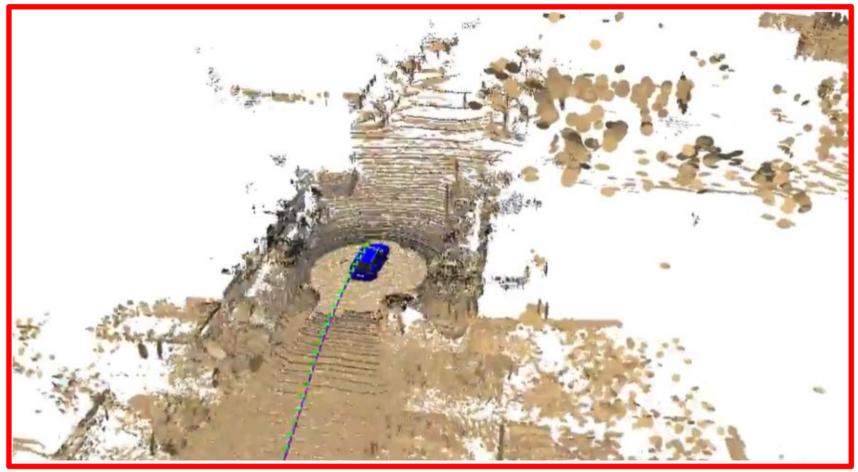
- Previous work
  - Efficient Surfel-Based SLAM using 3D Laser Range Data in Urban Environments.

※**Surfel**: Surface element



$v_s \in \mathbb{R}^3$
$n_s \in \mathbb{R}^3$
$r_s \in \mathbb{R}$





$\mathcal{P}$ Point cloud
$\mathcal{V}_{\mathcal{D}}$ Vertex map
$\mathcal{N}_{\mathcal{D}}$ Normal map

Active: the area of the model most recently observed

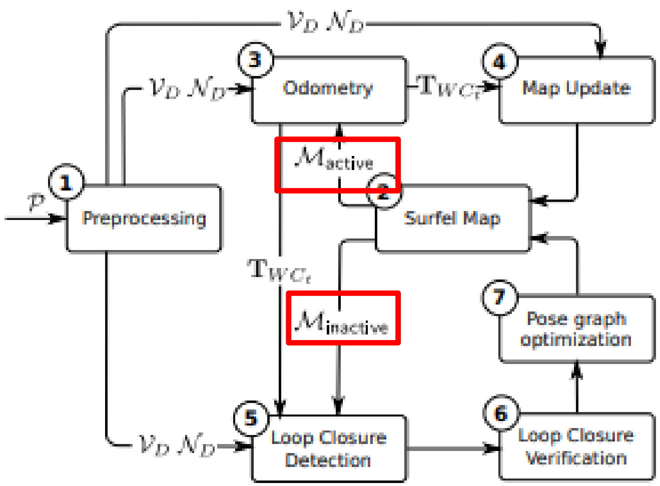
Inactive: Older parts of the map which have not been observed in a period of time δ_t

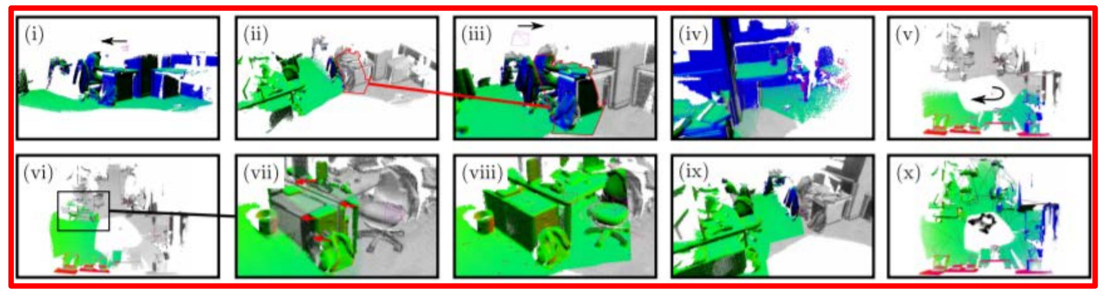
T. Whelan *et al.*, "ElasticFusion: Dense SLAM Without A Pose Graph". *In Proc. of Robotics: Science and Systems*, 2015.
Pfister, Hanspeter, et al. "Surfels: Surface elements as rendering primitives." Proceedings of the 27th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co., 2000.
Behley, Jens, and Cyrill Stachniss. "Efficient Surfel-Based SLAM using 3D Laser Range Data in Urban Environments." Robotics: Science and Systems. 2018.

SuMa++: Efficient LiDAR-based Semantic SLAM

- Previous work
  - Efficient Surfel-Based SLAM using 3D Laser Range Data in Urban Environments.

※**Surfel**: Surface element





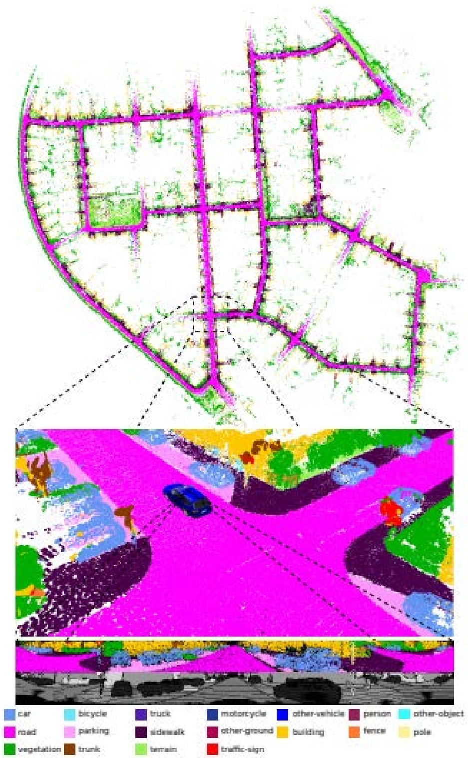
$\mathcal{P}$ Point cloud
$\mathcal{V}_D$ Vertex map
$\mathcal{N}_D$ Normal map

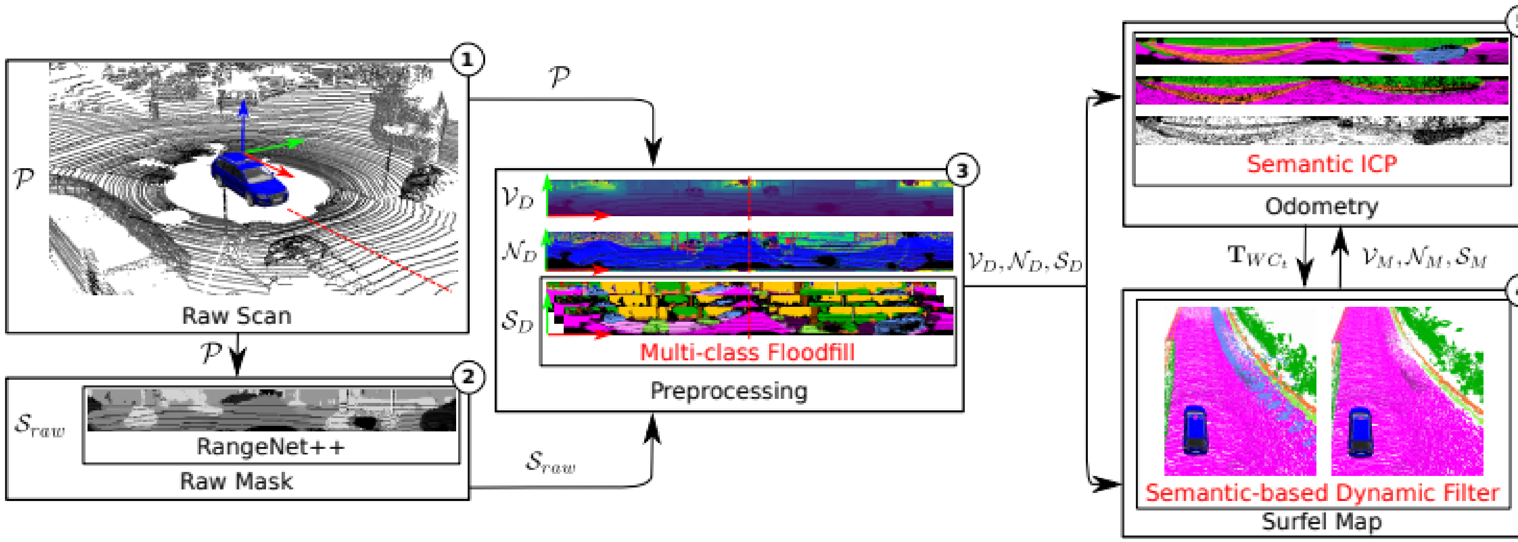
Active: the area of the model most recently observed

Inactive: Older parts of the map which have not been observed in a period of time δ_t

T. Whelan *et al.*, "ElasticFusion: Dense SLAM Without A Pose Graph". *In Proc. of Robotics: Science and Systems*, 2015.
Pfister, Hanspeter, et al. "Surfels: Surface elements as rendering primitives." Proceedings of the 27th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co., 2000.
Behley, Jens, and Cyrill Stachniss. "Efficient Surfel-Based SLAM using 3D Laser Range Data in Urban Environments." Robotics: Science and Systems. 2018.

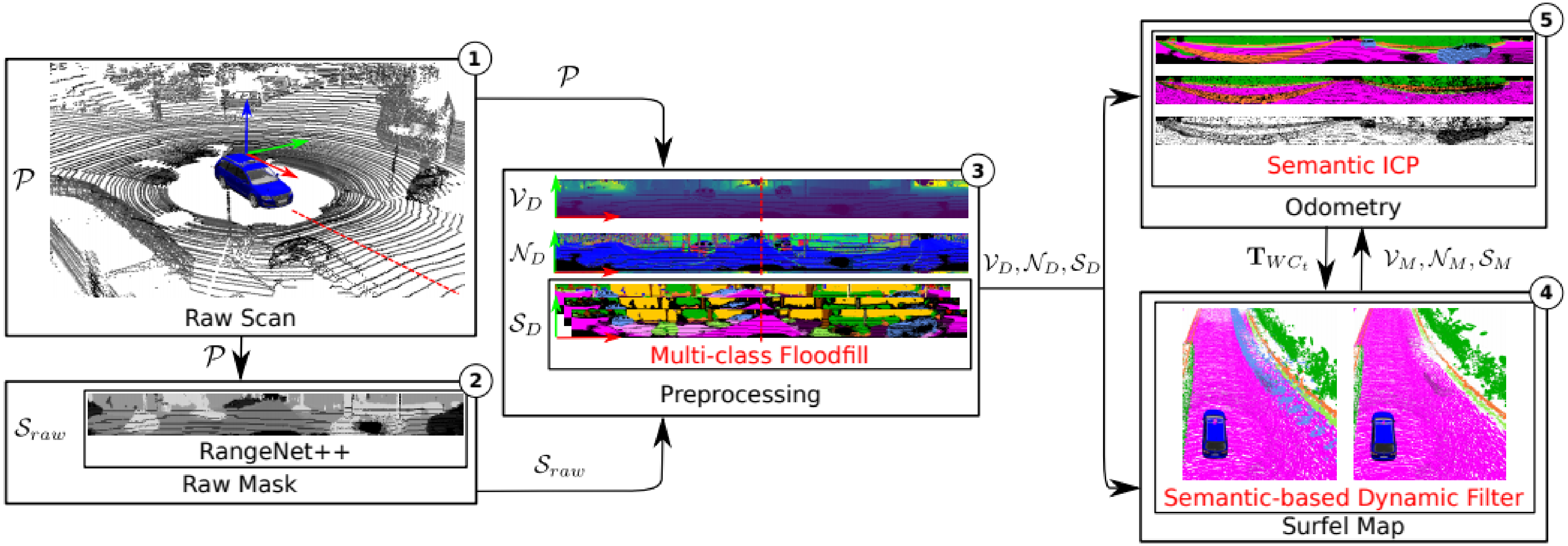
SuMa++: Efficient LiDAR-based Semantic SLAM





SuMa++: Efficient LiDAR-based Semantic SLAM

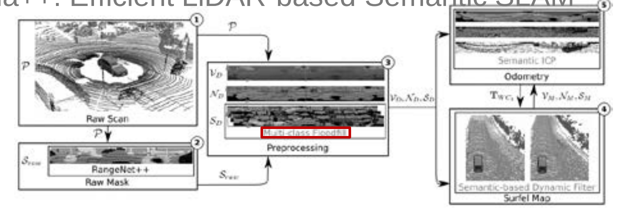
- Semantic segmentation using RangeNet++
  - Multi-class Floodfill
  - Semantic ICP
  - Dynamic detection and removal
    - Compare the new observation S_D and the world model S_M

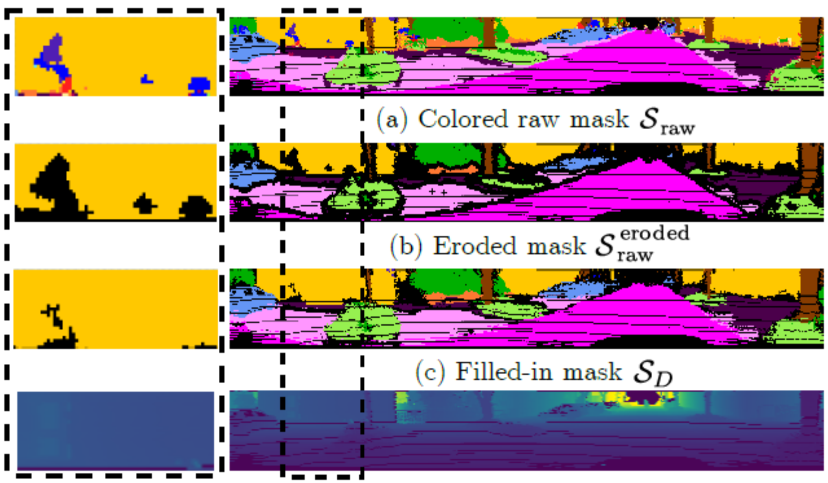


SuMa++: Efficient LiDAR-based Semantic SLAM

- 1. Multi-class Floodfill
  - ① Use an erosion to remove boundary labels and small areas of wrong labels resulting in S_{raw}^{eroded}
  - ② Fill-in eroded labels w/ neighboring labels to get a more consistent result S_D

SuMa++: Efficient LiDAR-based Semantic SLAM





(a) Colored raw mask $\mathcal{S}_{\text{raw}}$

(b) Eroded mask $\mathcal{S}_{\text{raw}}^{\text{eroded}}$

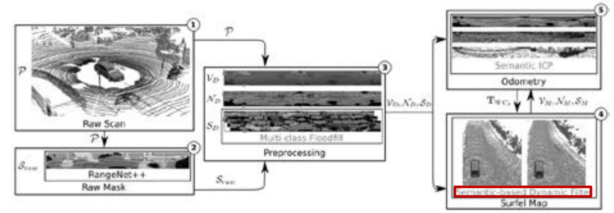
(c) Filled-in mask $\mathcal{S}_D$

(e) Zoomed-up

(d) Depth image generated from $\mathcal{V}_D$

SuMa++: Efficient LiDAR-based Semantic SLAM

SuMa++: Efficient LiDAR-based Semantic SLAM



2. Filtering Dynamics using Semantics

- Stable log odds term, $l_s^{(t)}$

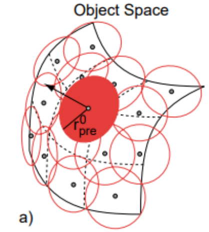
SuMa

$$l_s^{(t)} = l_s^{(t-1)} + \text{odds}\left(p_{\text{stable}} \cdot \exp\left(-\frac{\alpha^2}{\sigma_\alpha^2}\right) \exp\left(-\frac{d^2}{\sigma_d^2}\right)\right) - \text{odds}(p_{\text{prior}}), \tag{4}$$

SuMa++

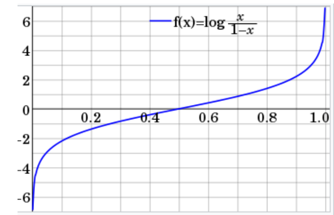
$$l_s^{(t)} = l_s^{(t-1)} + \text{odds}\left(p_{\text{stable}} \exp\left(-\frac{\alpha^2}{\sigma_\alpha^2}\right) \exp\left(-\frac{d^2}{\sigma_d^2}\right)\right) - \text{odds}(p_{\text{prior}}) - \text{odds}(p_{\text{penalty}}),$$

Object Space



$v_s \in \mathbb{R}^3$
$n_s \in \mathbb{R}^3$
$r_s \in \mathbb{R}$

Odds(log odds or logit)



α: The angle between surfel's normal n_s and normal of the measurement
d: distance of the measurement in respect to the associated surfel

SuMa++: Efficient LiDAR-based Semantic SLAM

- 2. Filtering Dynamics using Semantics
  - Stable log odds term, $l_s^{(t)}$

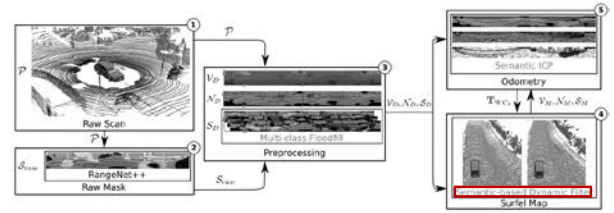
SuMa
$$l_s^{(t)} = l_s^{(t-1)} + \text{odds}\left(p_{\text{stable}} \cdot \exp\left(-\frac{\alpha^2}{\sigma_\alpha^2}\right) \exp\left(-\frac{d^2}{\sigma_d^2}\right)\right) - \text{odds}(p_{\text{prior}}), \quad (4)$$

⬇

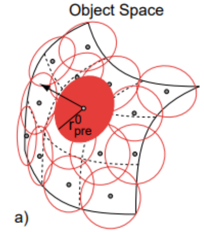
SuMa++
$$l_s^{(t)} = l_s^{(t-1)} + \text{odds}\left(p_{\text{stable}} \exp\left(-\frac{\alpha^2}{\sigma_\alpha^2}\right) \exp\left(-\frac{d^2}{\sigma_d^2}\right)\right) - \text{odds}(p_{\text{prior}}) - \text{odds}(p_{\text{penalty}}),$$

α: The angle between surfel's normal n_s and normal of the measurement
d: distance of the measurement in respect to the associated surfel

SuMa++: Efficient LiDAR-based Semantic SLAM

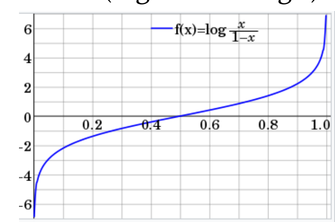


Object Space



$v_s \in \mathbb{R}^3$
$n_s \in \mathbb{R}^3$
$r_s \in \mathbb{R}$

Odds(log odds or logit)



(a) SuMa (b) SuMa++ (c) SuMa_nomovable

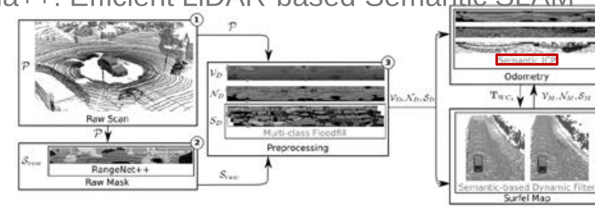
KAIST

URL URBAN ROBOTICS LAB

SuMa++: Efficient LiDAR-based Semantic SLAM

- 3. Semantic ICP

SuMa++: Efficient LiDAR-based Semantic SLAM



Point-to-plane error term

$$E(\mathcal{V}_D, \mathcal{V}_M, \mathcal{N}_M) = \sum_{\mathbf{u} \in \mathcal{V}_D} \underbrace{\mathbf{n}_u^\top \cdot \left(\mathbf{T}^{(k)}_{C_{t-1}C_t} \mathbf{u} - \mathbf{v}_u \right)}_{r_{\mathbf{u}}}{}^2$$

where each vertex $\mathbf{u} \in \mathcal{V}_D$ is projectively associated to a reference vertex $\mathbf{v}_u \in \mathcal{V}_M$ and its normal $\mathbf{n}_u \in \mathcal{N}_M$ via

$$\mathbf{v}_u = \mathcal{V}_M \left(\Pi \left(\mathbf{T}^{(k)}_{C_{t-1}C_t} \mathbf{u} \right) \right), \tag{6}$$

$$\mathbf{n}_u = \mathcal{N}_M \left(\Pi \left(\mathbf{T}^{(k)}_{C_{t-1}C_t} \mathbf{u} \right) \right). \tag{7}$$

$$E(\mathcal{V}_D, \mathcal{V}_M, \mathcal{N}_M) = \sum_{\mathbf{u} \in \mathcal{V}_D} \boxed{w_{\mathbf{u}}} \underbrace{\mathbf{n}_{\mathbf{u}}^\top \left(\mathbf{T}^{(k)}_{C_{t-1}C_t} \mathbf{u} - \mathbf{v}_{\mathbf{u}} \right)}{}^2,$$

Within ICP, we compute the **weight** $w_{\mathbf{u}}^{(k)}$ for the **residual** $r_{\mathbf{u}}^{(k)}$ in iteration k as follows:

$$w_{\mathbf{u}}^{(k)} = \boxed{\rho_{\text{Huber}}\left(r_{\mathbf{u}}^{(k)}\right)} \boxed{C_{\text{semantic}}(\mathcal{S}_D(\mathbf{u}), \mathcal{S}_M(\mathbf{u}))} \boxed{\mathbb{I}\left\{ l_s^{(k)} \geq l_{stable} \right\}} \tag{6}$$

where $\rho_{\text{Huber}}(r)$ corresponds to the Huber norm, given by:

$$\boxed{\rho_{\text{Huber}}(r) = \begin{cases} 1 & , \text{if } |r| < \delta \\ \delta |r|^{-1} & , \text{otherwise.} \end{cases}} \tag{7}$$

For **semantic compatibility** $C_{\text{semantic}}((y_{\mathbf{u}}, P_{\mathbf{u}}), (y_{v_{\mathbf{u}}}, P_{v_{\mathbf{u}}}))$, the term is defined as:

$$\boxed{C_{\text{semantic}}(\cdot,\cdot) = \begin{cases} P(y_{\mathbf{u}}|\mathbf{u}) & , \text{if } y_{\mathbf{u}} = y_{v_{\mathbf{u}}} \\ 1 - P(y_{\mathbf{u}}|\mathbf{u}) & , \text{otherwise.} \end{cases}}, \tag{8}$$

the residual. By $\mathbb{I}\{a\}$, we denote the indicator function that is 1 if the argument a is true, and 0 otherwise.

KAIST

URL
URBAN ROBOTICS LAB

DeepPCO: End-to-End Point Cloud through Deep Parallel Neural Network

- Parallel Neural Network shows more better performance than other Deep Learning-based approaches

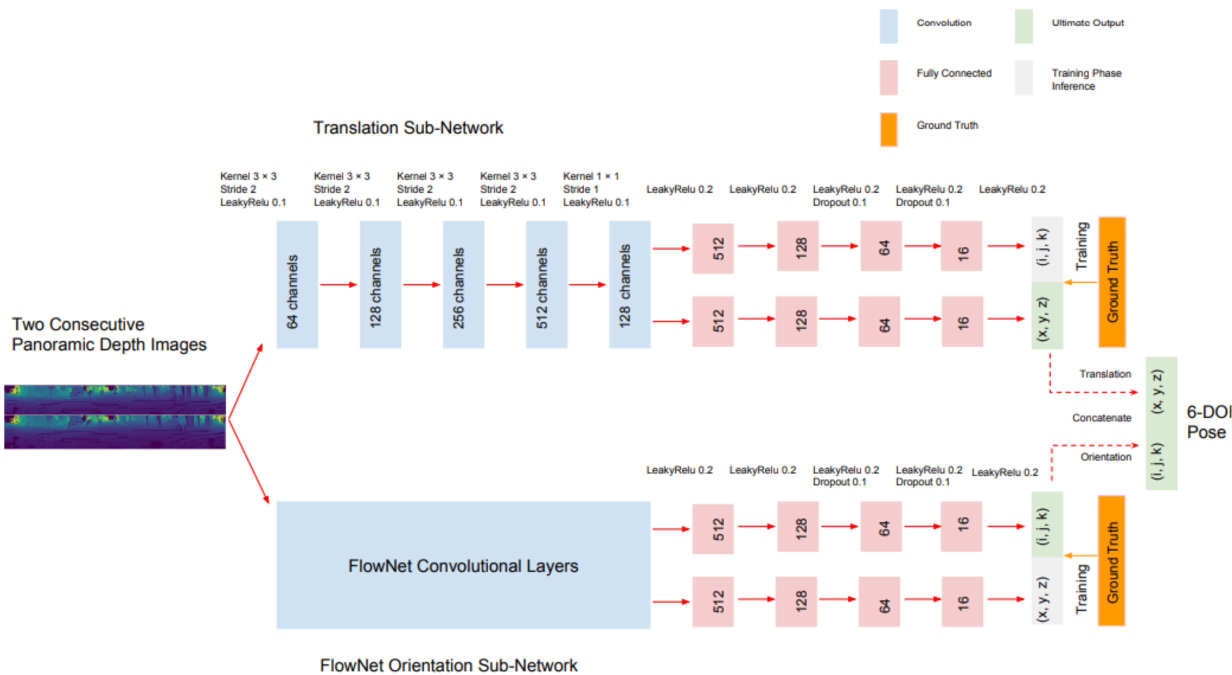
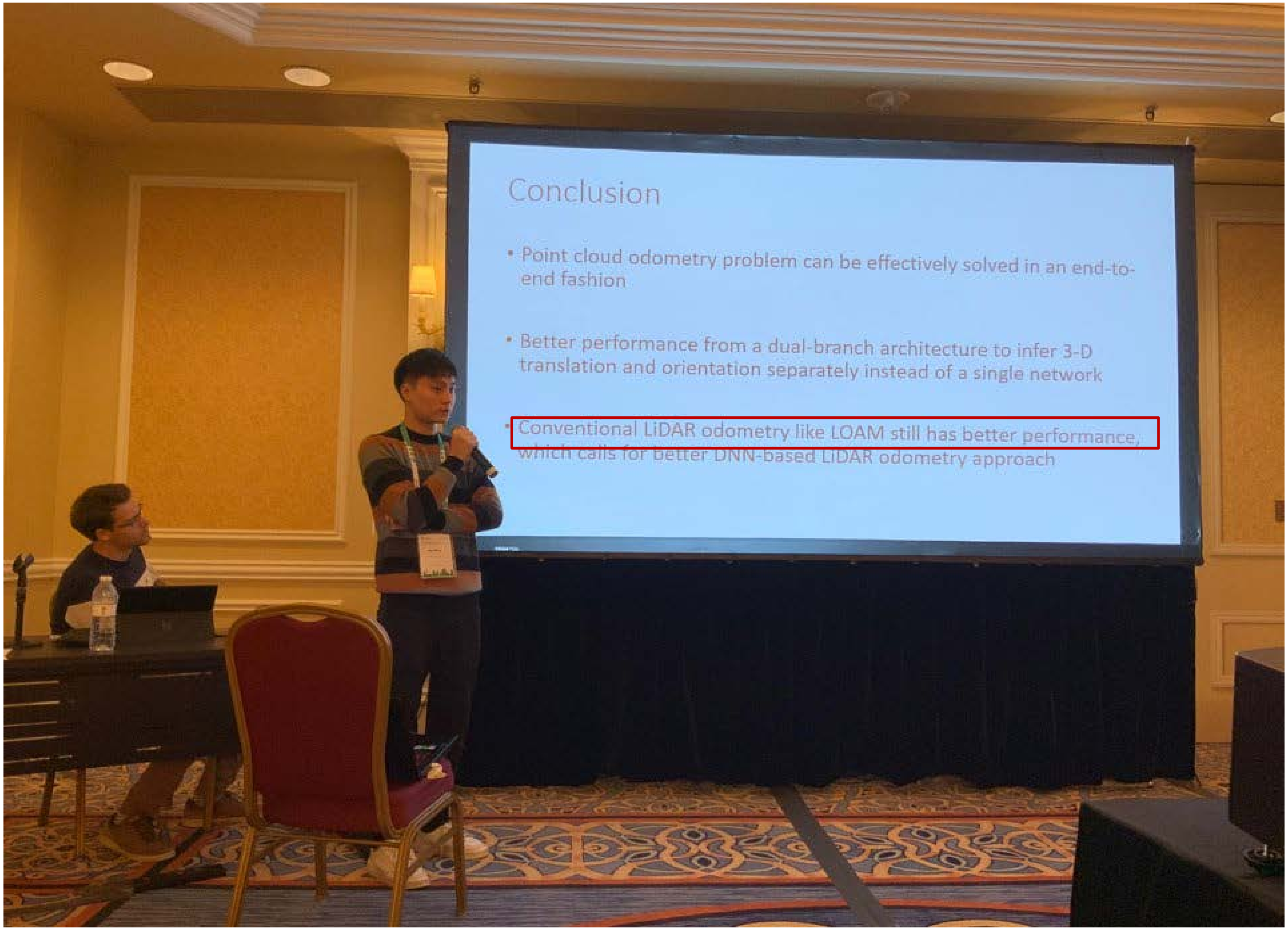


TABLE I: Test results on KITTI sequences: t_{rel} is averaged translation RMSE between prediction and ground truth, and r_{rel} is averaged orientation RMSE between prediction and ground truth. The measurement units are textbfm).

	Sequence 04		Sequence 10	
Model	t_rel	r_rel	t_rel	r_rel
Two-stream [16]	0.0554	0.0830	0.0870	0.1592
ResNet18 [10]	0.1094	0.0602	0.1443	0.1327
DeepVO [20]	0.2157	0.0709	0.2153	0.3311
PointNet [17]	0.0946	0.0442	0.1381	0.1360
PointGrid [12]	0.0550	0.0690	0.0842	0.1523
DeepPCO (Ours)	**0.0263**	**0.0305**	**0.0247**	**0.0659**

DeepPCO: End-to-End Point Cloud through Deep Parallel Neural Network



Q & A