ECCV'20 ONLINE
23-28 AUGUST 2020

DH3D: Deep Hierarchical 3D Descriptors for Robust Large-Scale 6DoF Relocalization



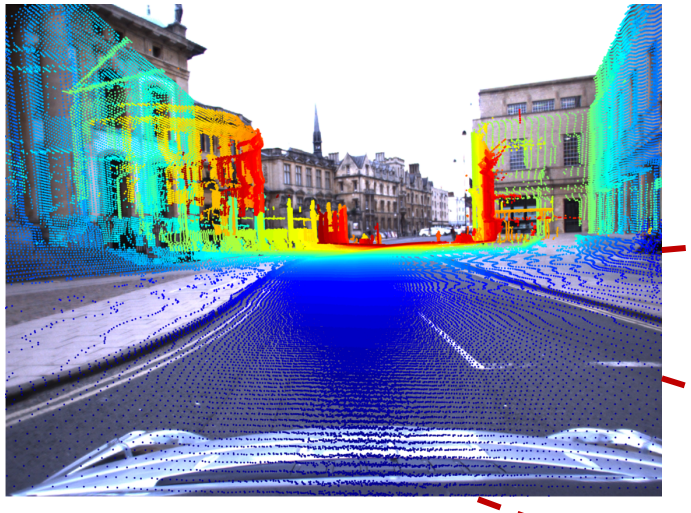
Juan Du,[1*] Rui Wang,[1,2*] Daniel Cremers[1,2]

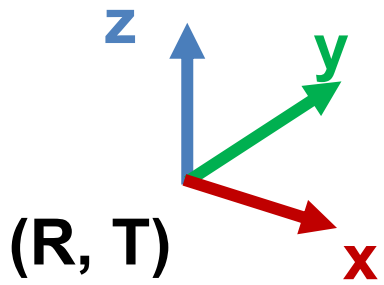
[1] Computer Vision Group
Technical University of Munich

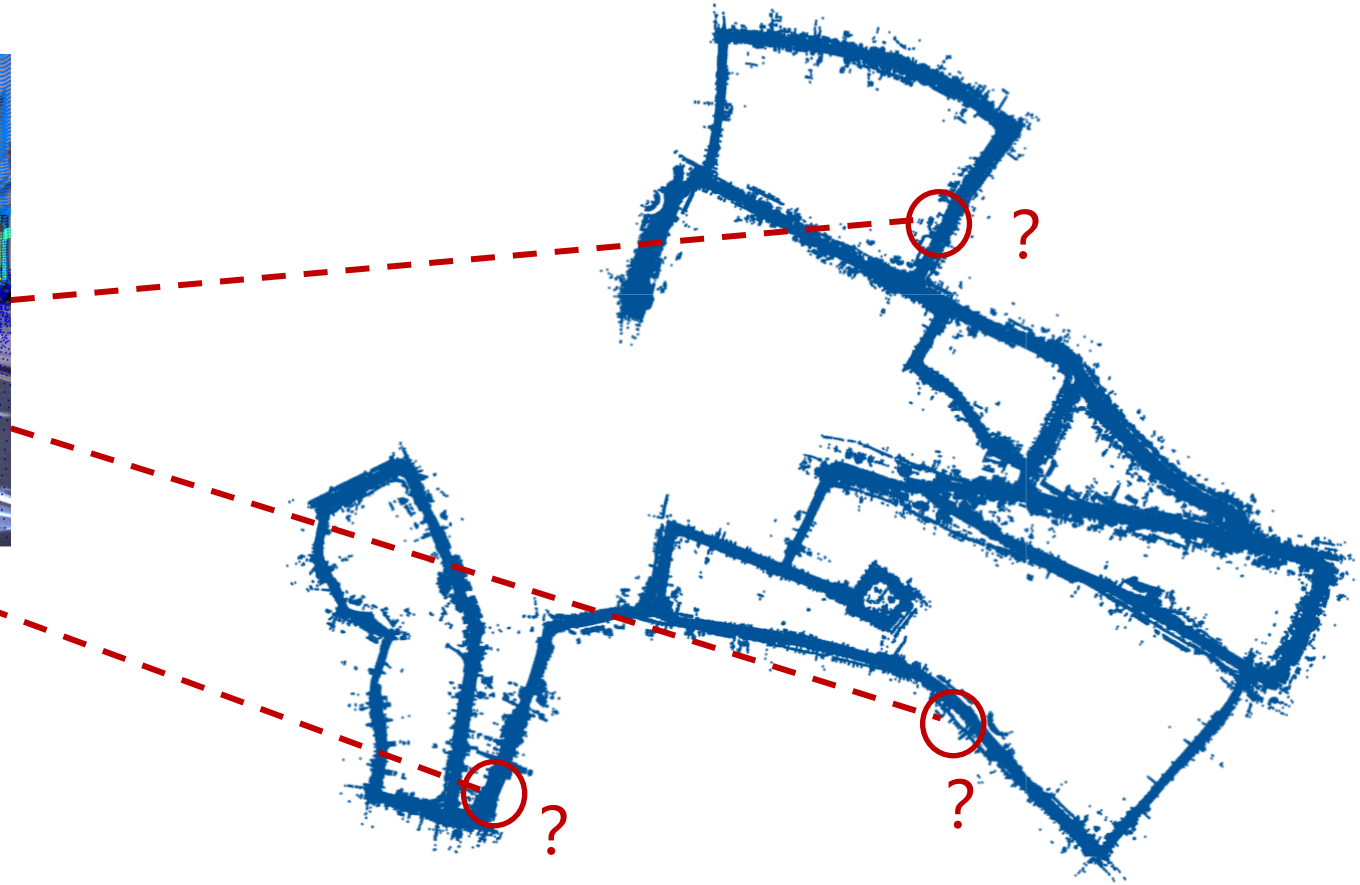
[2] Artisense

[*] Equal contribution

Target: 6DoF relocalization in a city-scale 3D map





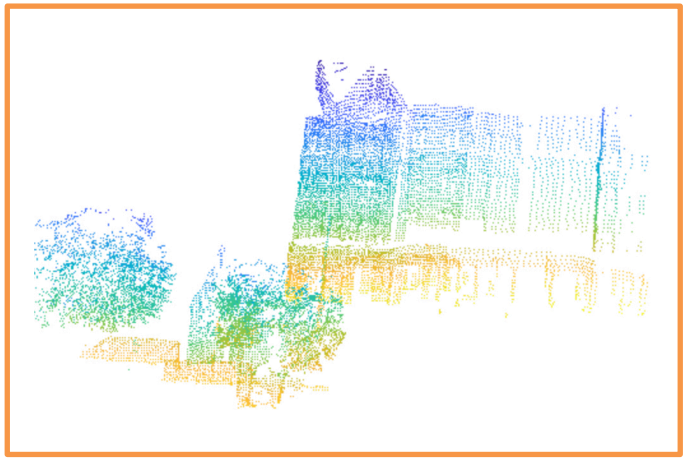


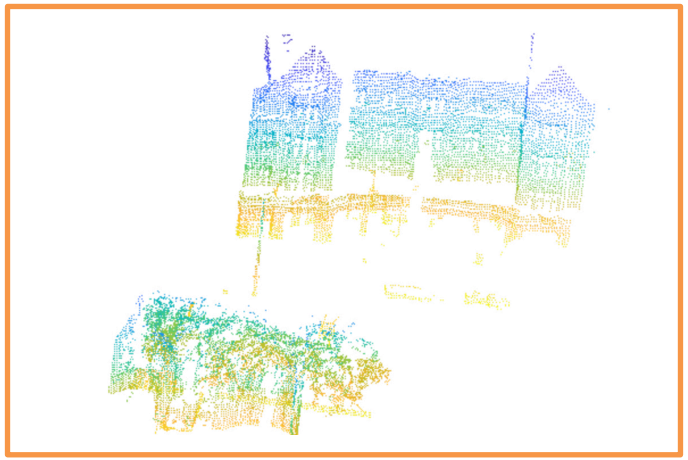
The Oxford RobotCar Dataset

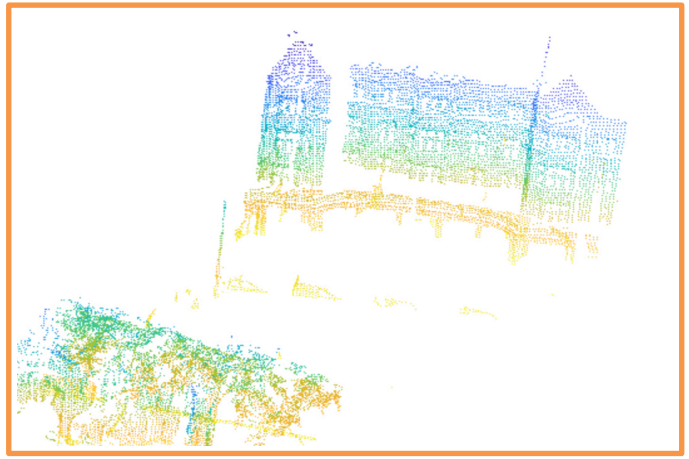
Limitations of image based approaches

illumination

viewpoint

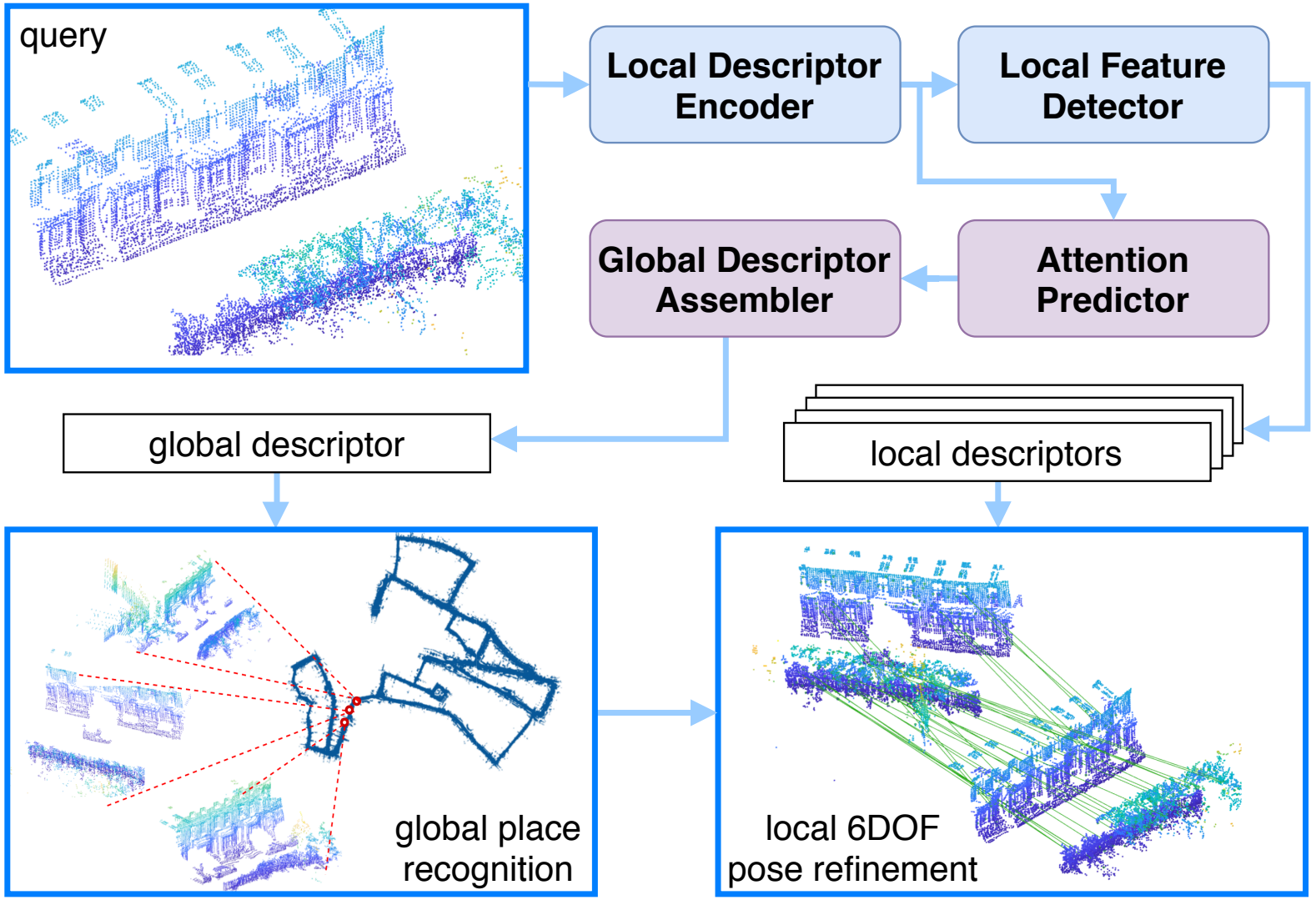






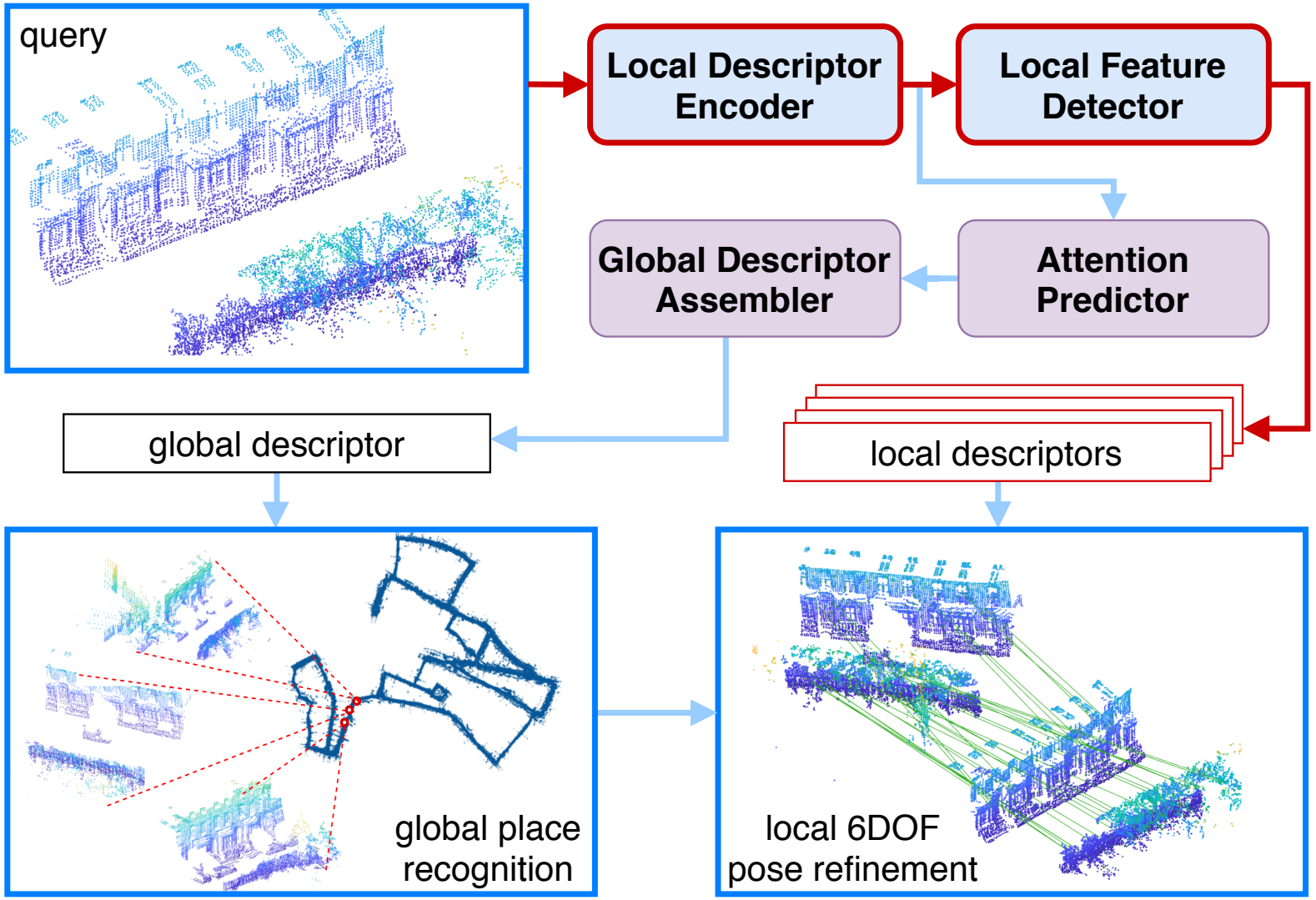
Goal: a data-driven 3D descriptor that captures geometric information

6DoF relocalization based on **Deep Hierarchical 3D Descriptors**



System Overview

6DoF relocalization based on **Deep Hierarchical 3D Descriptors**

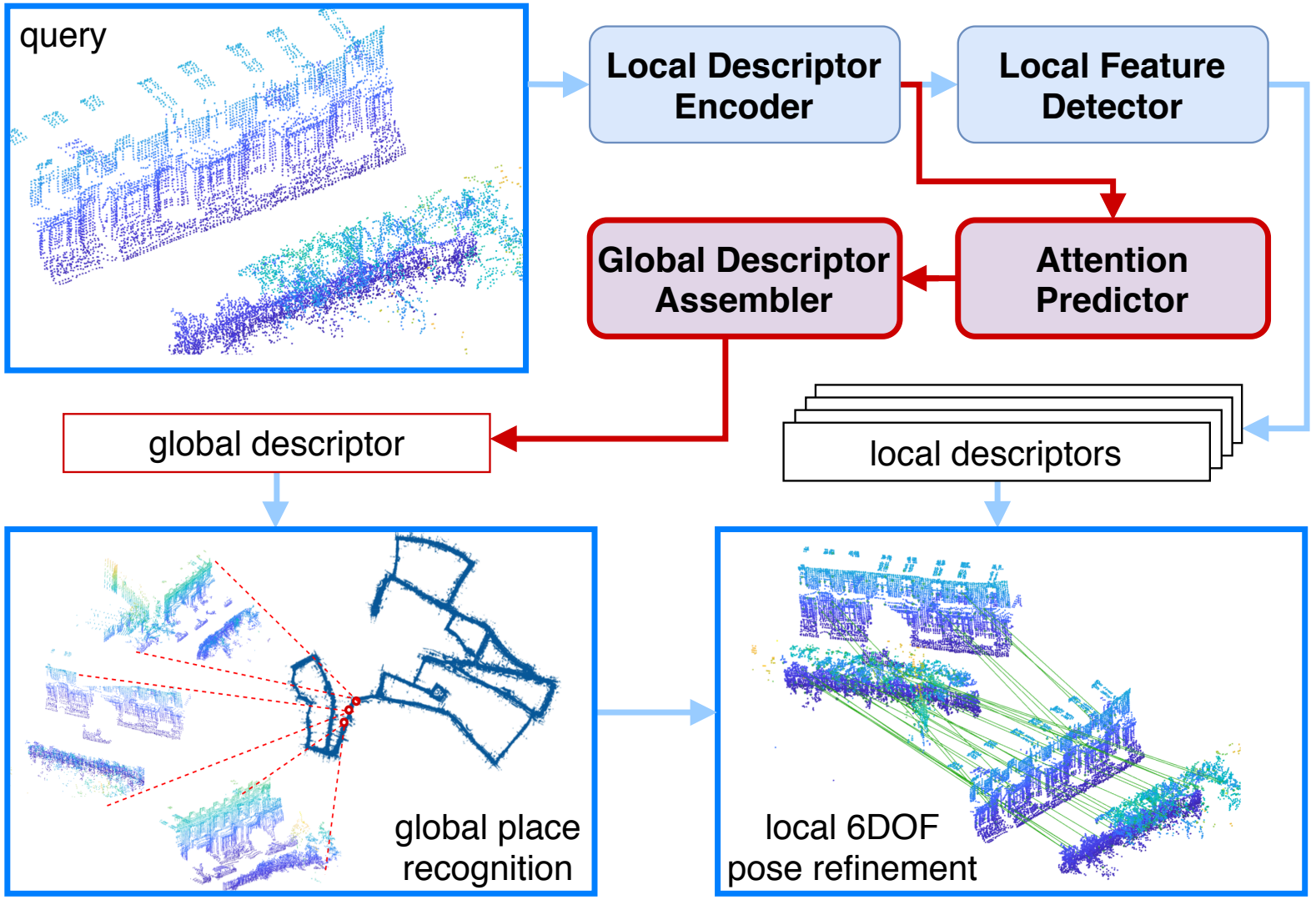


Networks extract

3D keypoints and

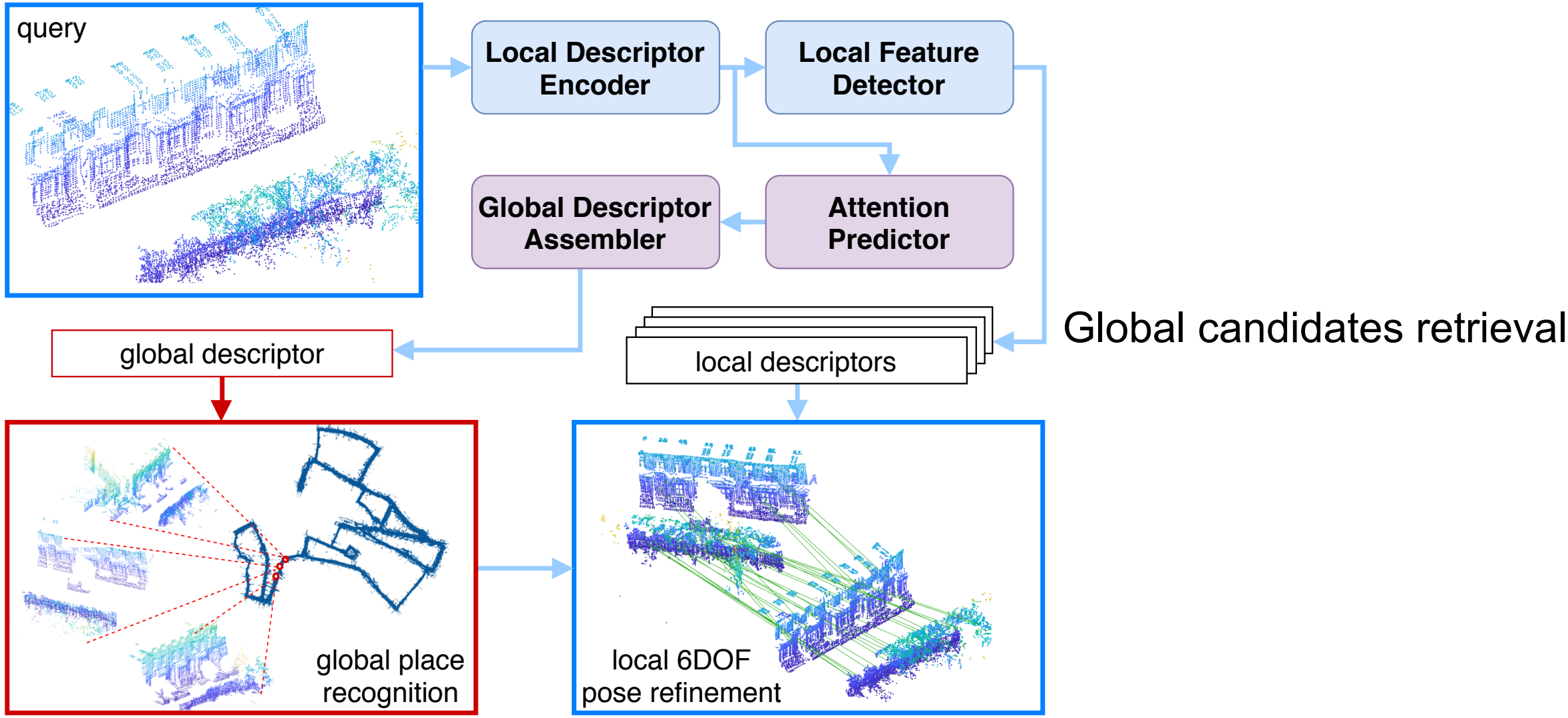
local descriptors

6DoF relocalization based on **Deep Hierarchical 3D Descriptors**

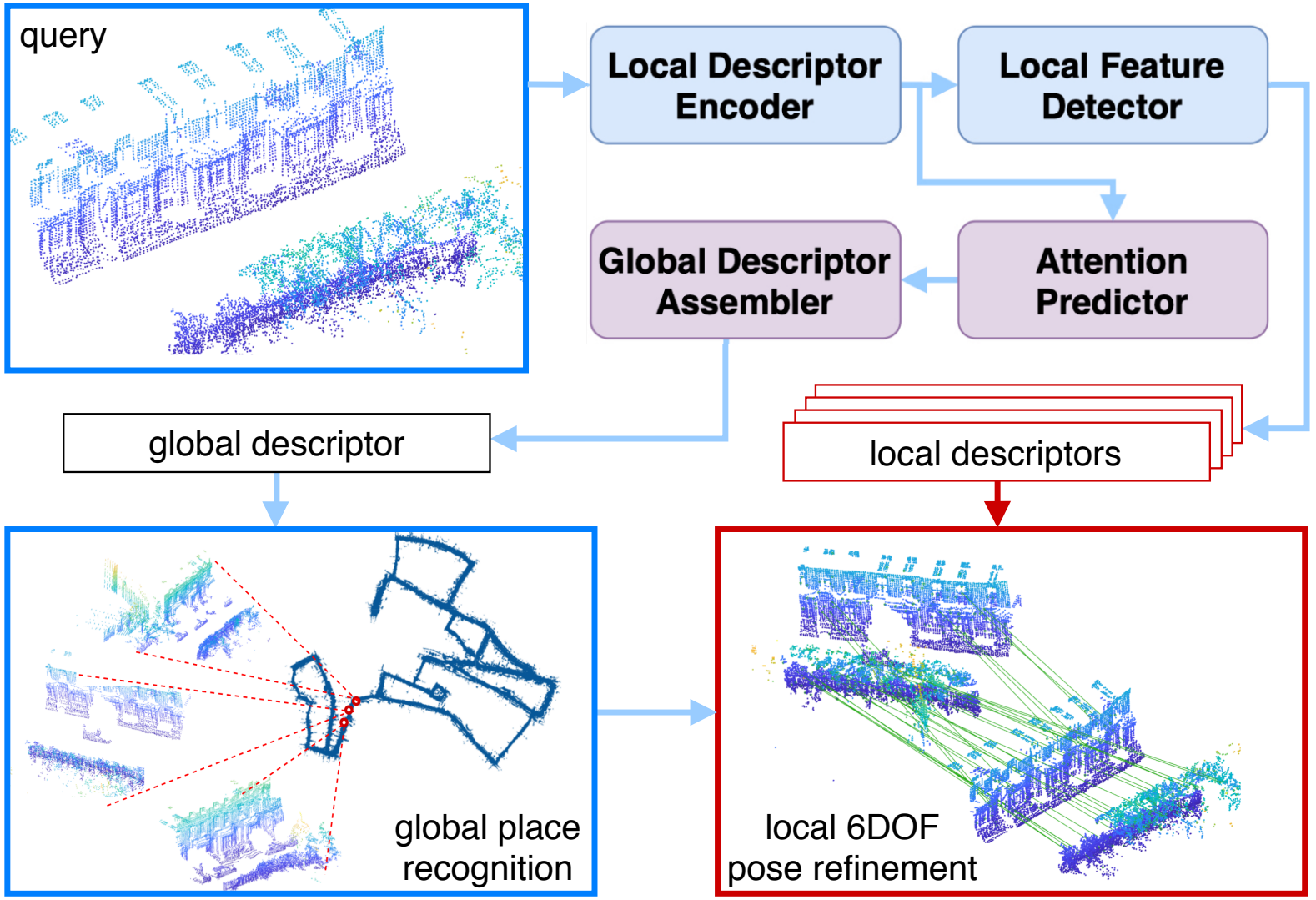


Networks aggregate local descriptors into a **3D global descriptor**

6DoF relocalization based on **Deep Hierarchical 3D Descriptors**



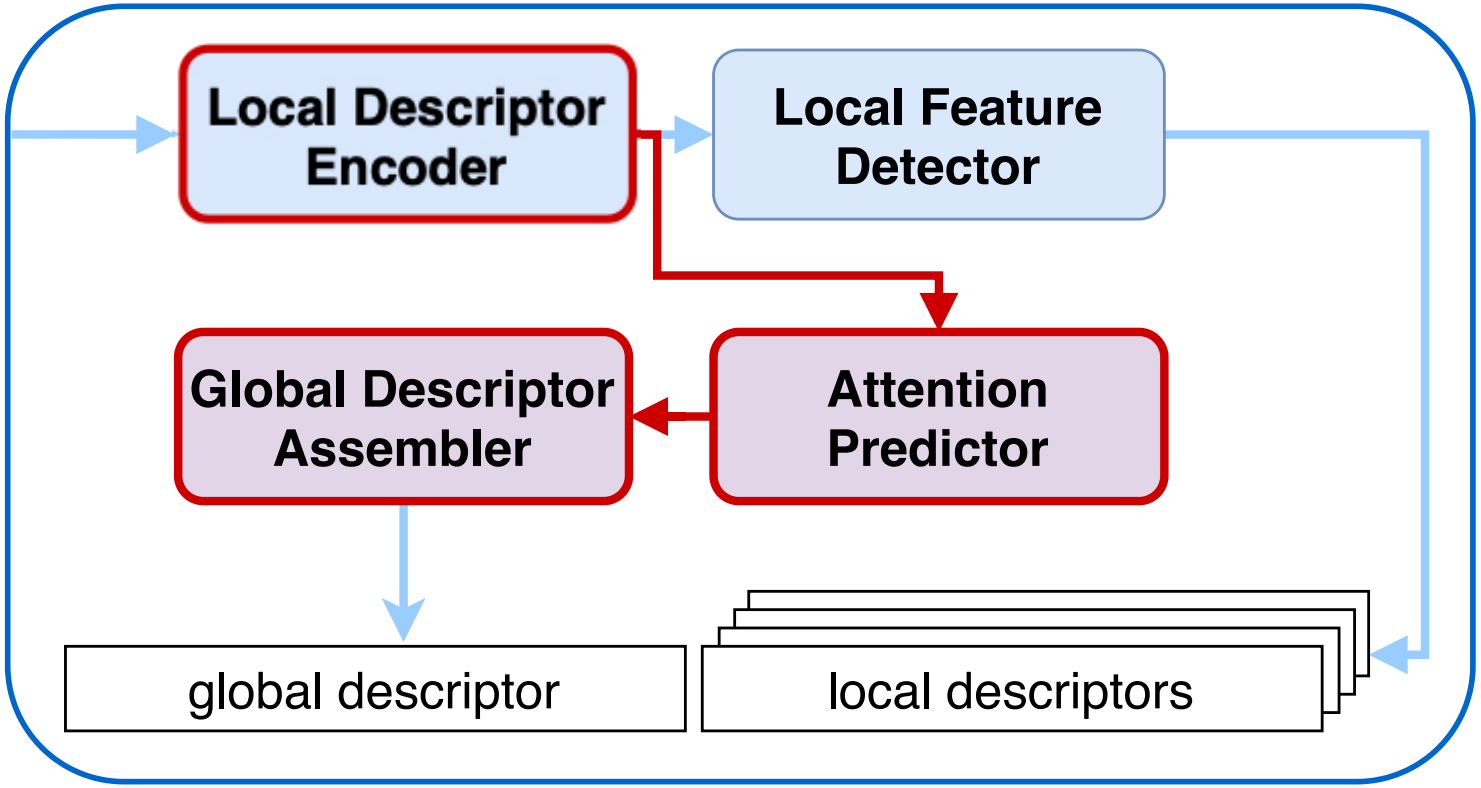
6DoF relocalization based on **Deep Hierarchical 3D Descriptors**



Local 6DoF

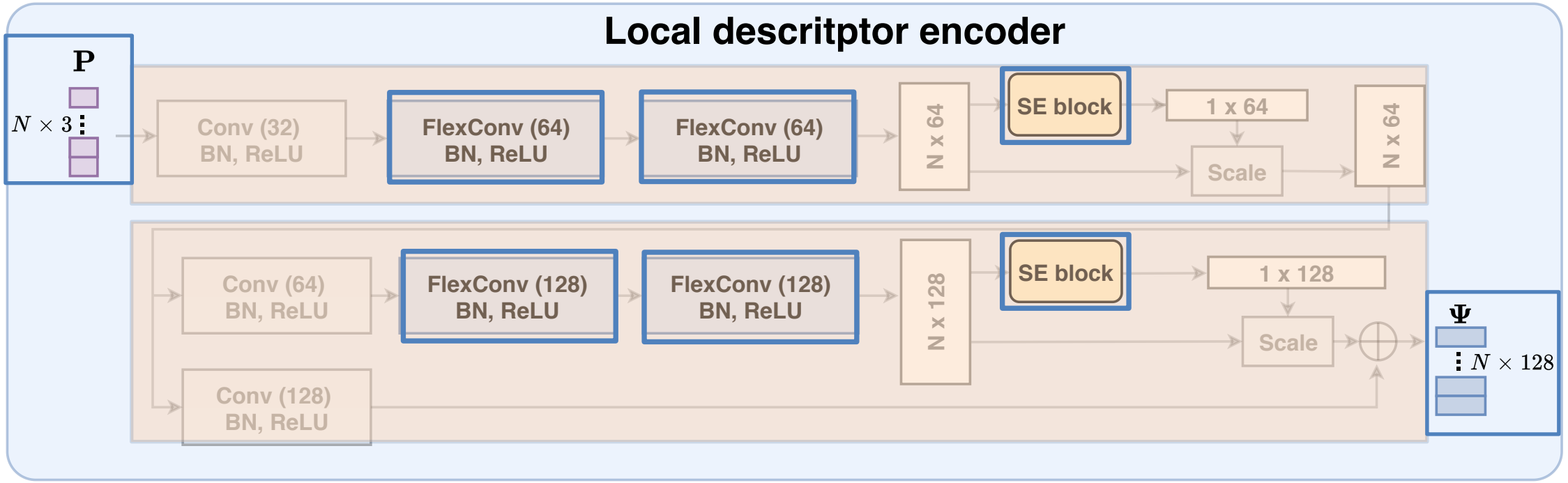
pose refinement

Network architecture

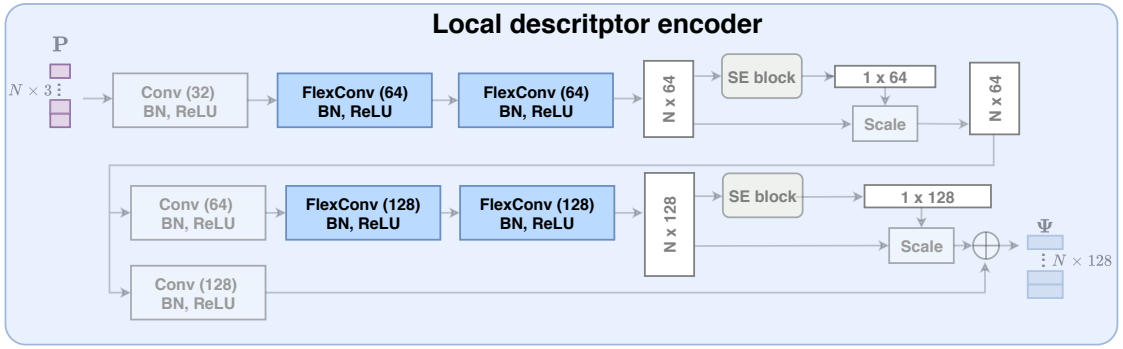


Reuse local descriptors

- Share low level geometric information
- One pass computation



- Integrate two levels of spatial information
- Two operations
  - Flex convolution
  - Squeeze-and-excitation block



- Interaction among unordered points - Euclidean space

Flex Convolution [Groh etal,2018]

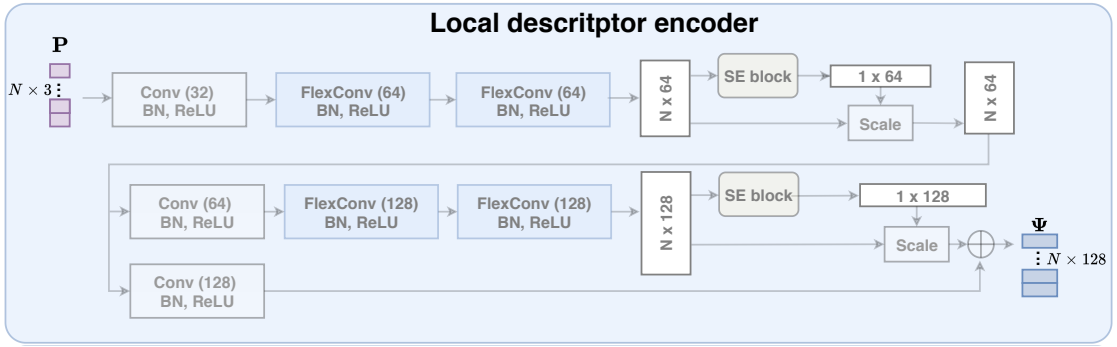
$\boldsymbol{p}_l$: a 3D point

$h(\boldsymbol{p}_l)$: feature function of $\boldsymbol{p}_l$

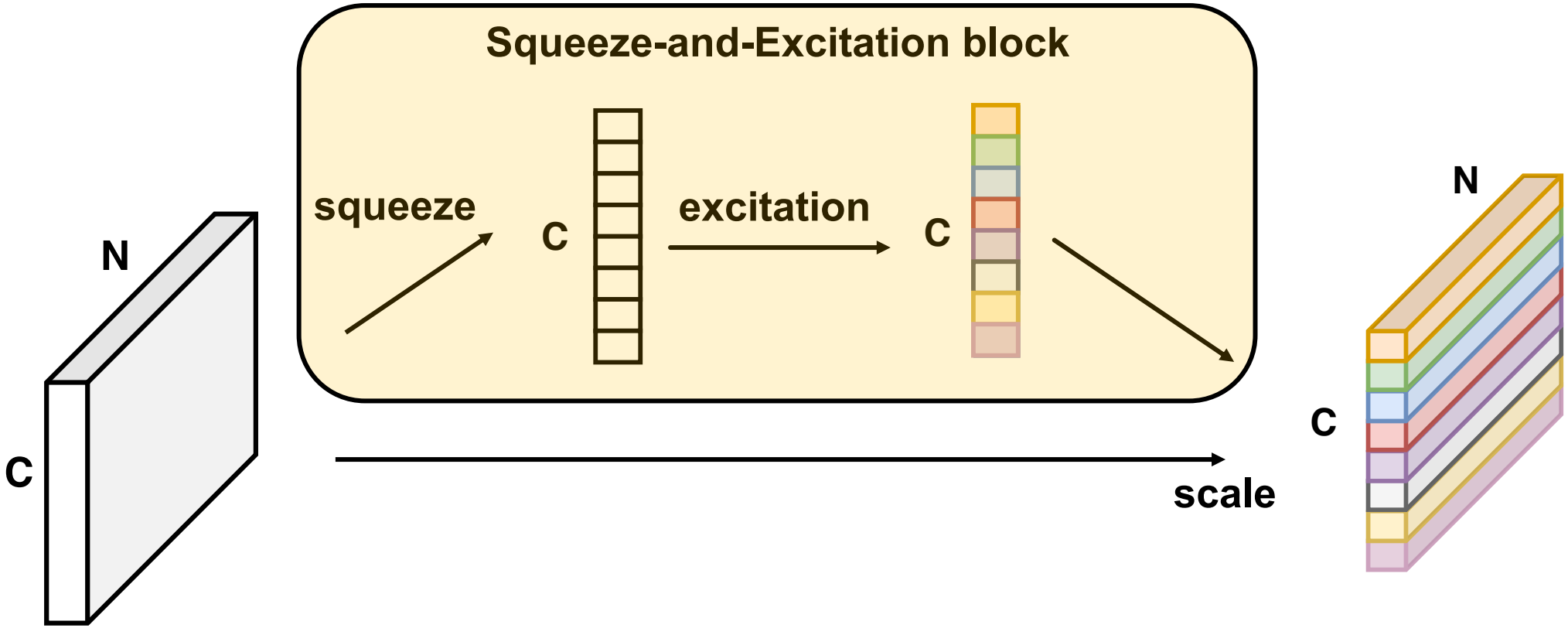
$N_k(\boldsymbol{p}_l)$: k neighbours of $\boldsymbol{p}_l$

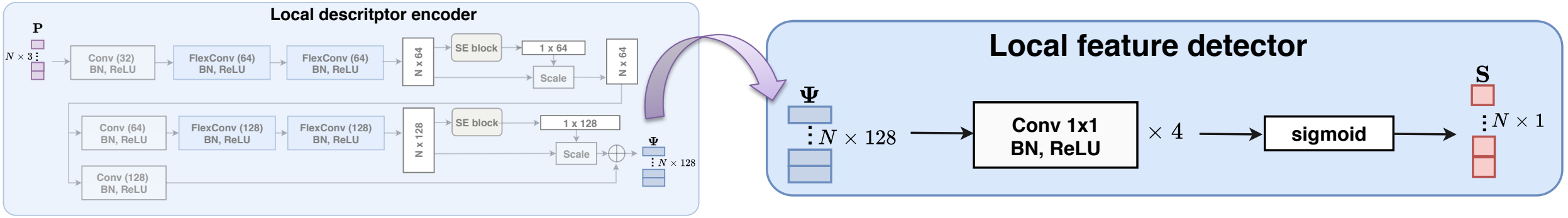
FlexConv: $f_{FlexConv}(\boldsymbol{p}_l) = \sum_{\boldsymbol{p}_{l_i} \in N_k(\boldsymbol{p}_l)} \omega(\boldsymbol{p}_{l_i}, \boldsymbol{p}_l) \cdot h(\boldsymbol{p}_{l_i})$

$$\omega(\boldsymbol{p}_{l_i}, \boldsymbol{p}_l \mid \theta, \theta_b) = \langle \theta, \boldsymbol{p}_{l_i} - \boldsymbol{p}_l \rangle + \theta_b$$

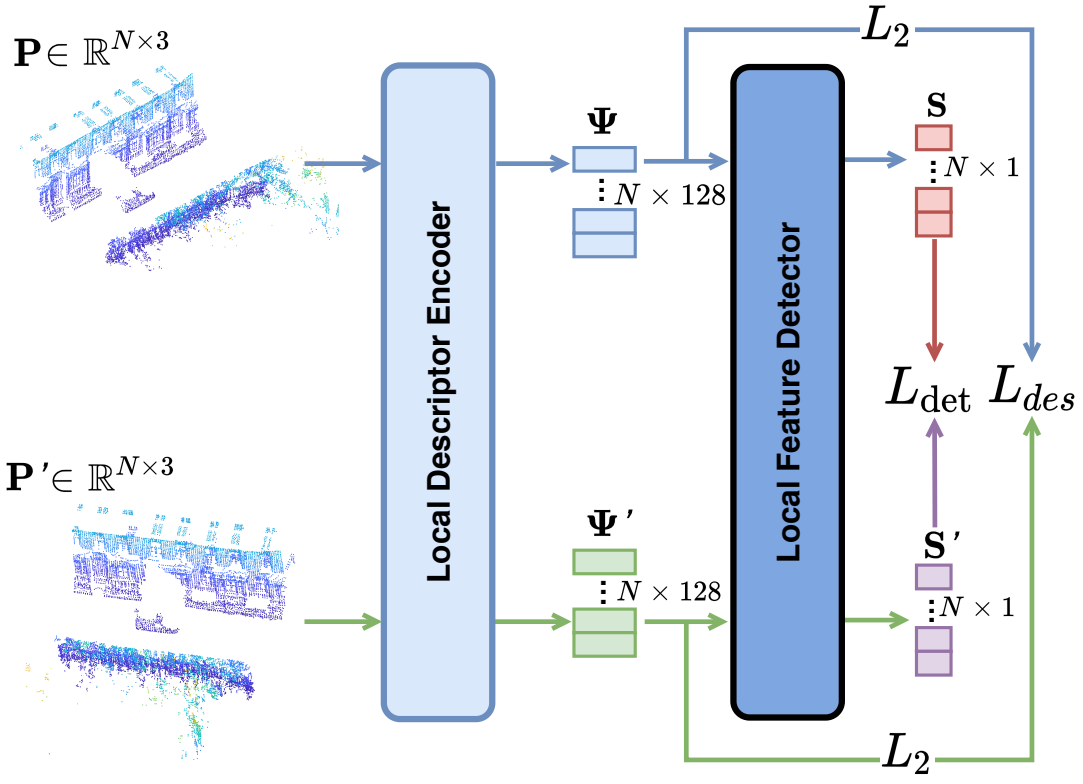


- Interaction among unordered points - feature space



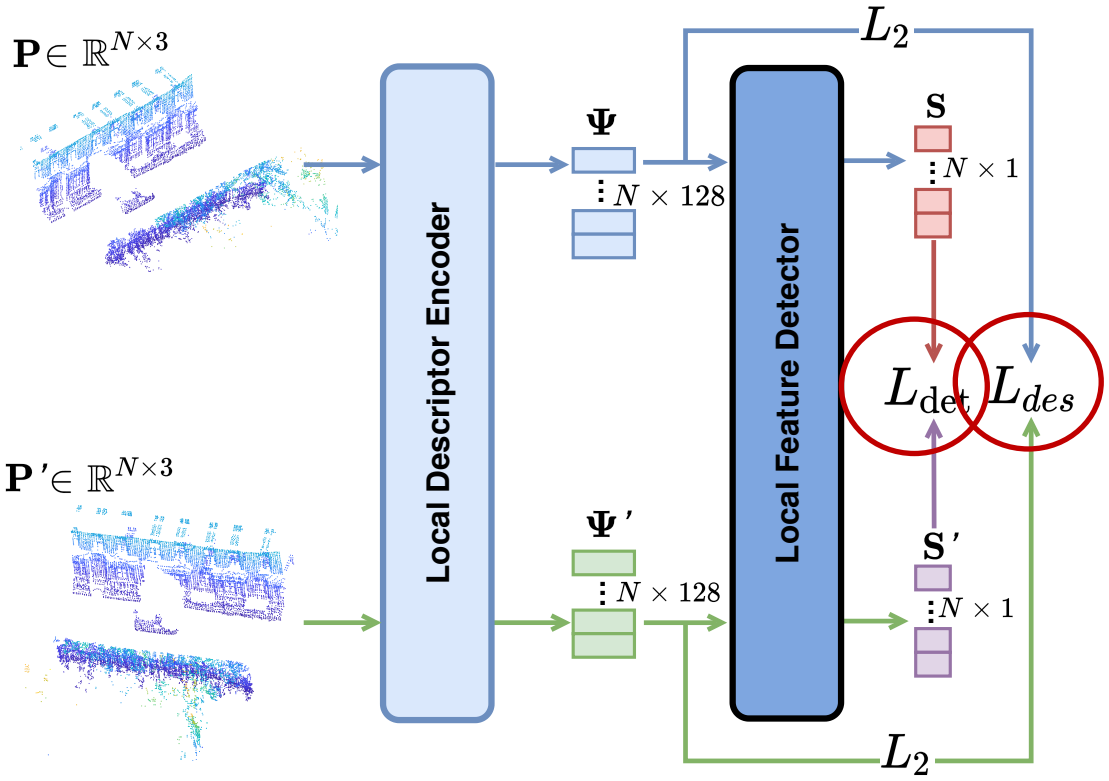


- ***Describe-and-detect*** approach to employ high-level information



- Applying arbitrary rotations around the upright axis
- A Siamese network jointly learns ***detection*** and ***description***

Network training: local part



- descriptor loss
  - N-tuple loss [Groh etal,2018]

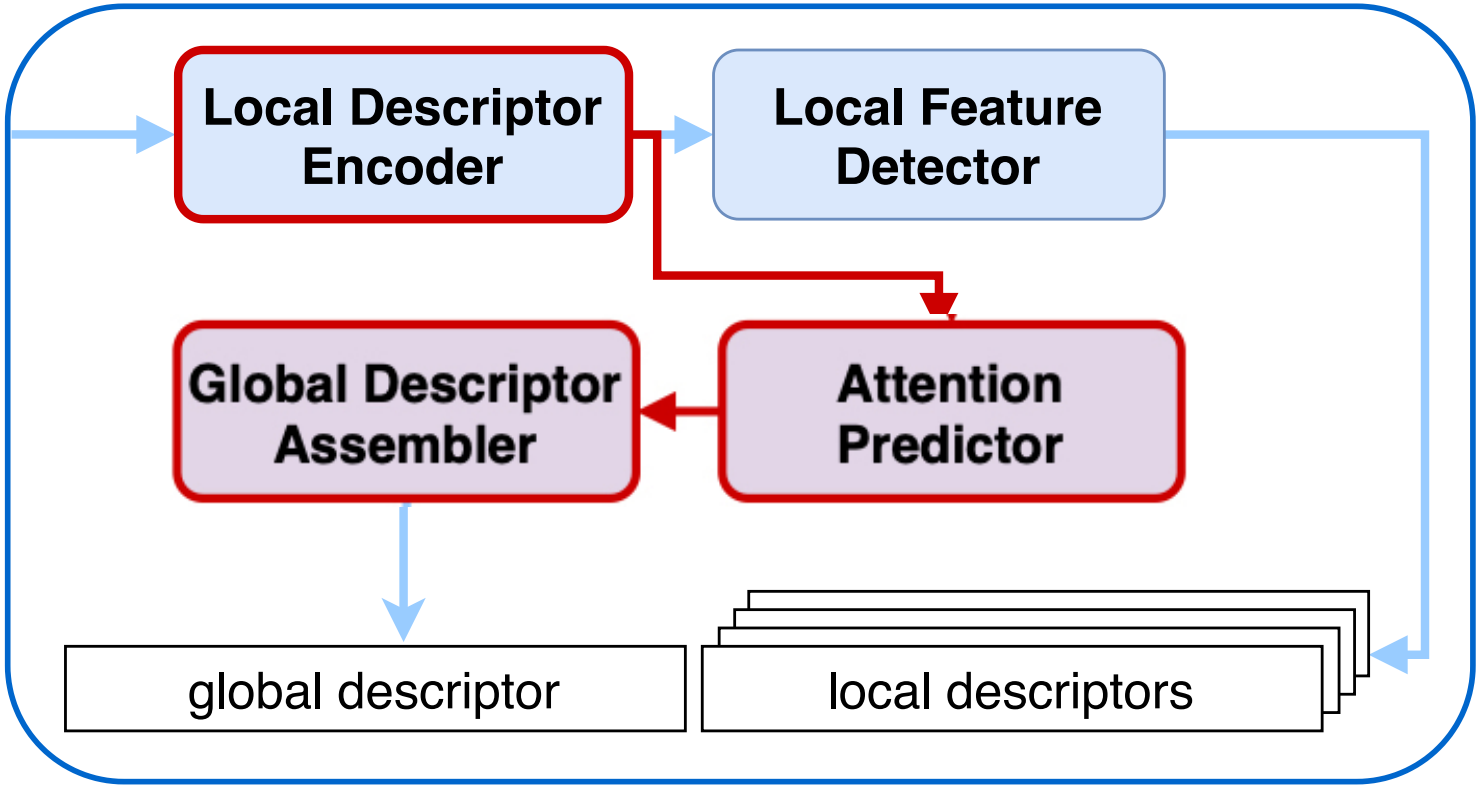
$$L_{des} = \sum \left(\frac{M \circ \mathrm{D}}{\|M\|_F^2} + \eta \frac{\max(\mu - (1-M) \circ D, 0)}{N^2 - \|M\|_F^2} \right)$$

- detector loss
  - Unsupervised training
  - AR (Average successful Rate) :

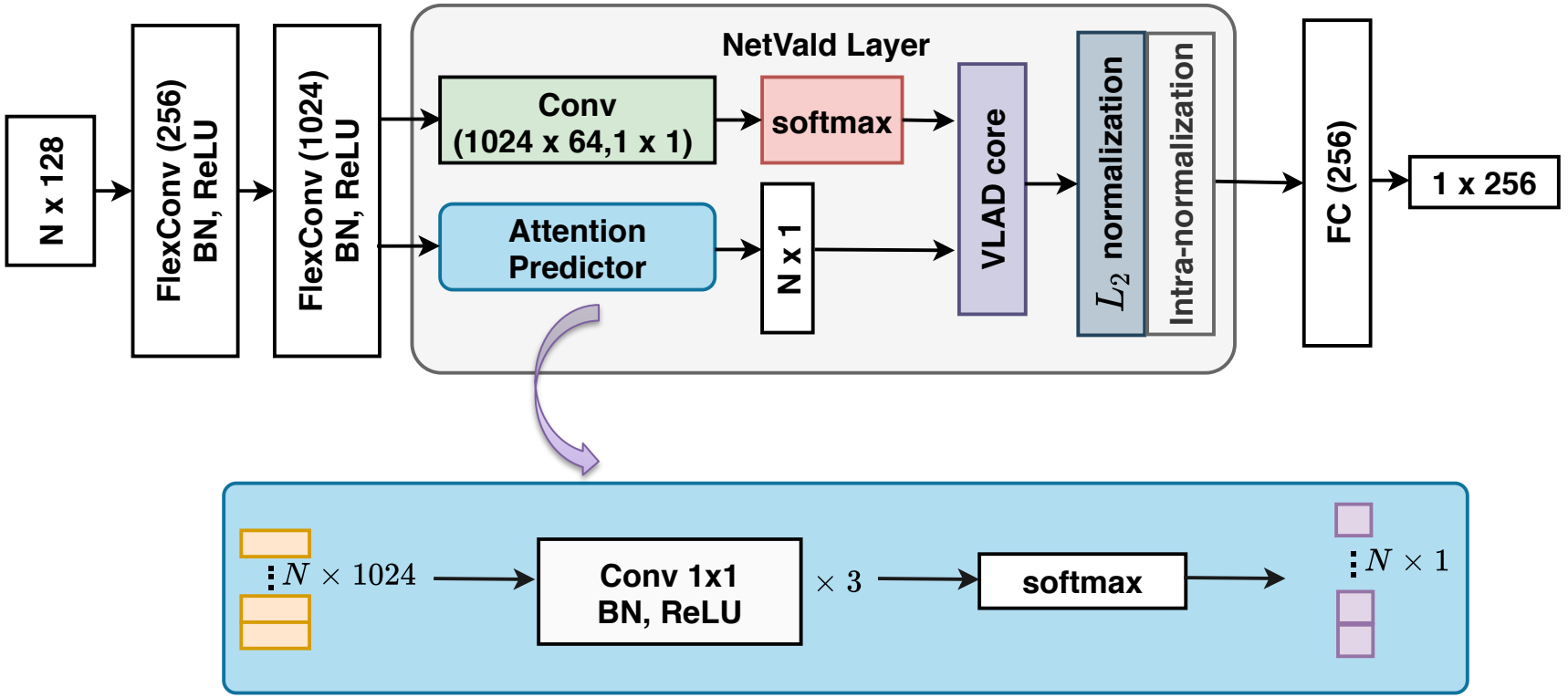
$AR_i = \frac{1}{k}\sum_{j=1}^{k} c_{ij}$, where c_{ij} = 1 if at least one correct correspondence can be found in the first j candidates otherwise is 0.

  - $L_{det} = \frac{1}{N}\sum_{1}^{N} 1 - [\kappa(1 - s_i) + s_i \cdot AR_i]$

Network architecture



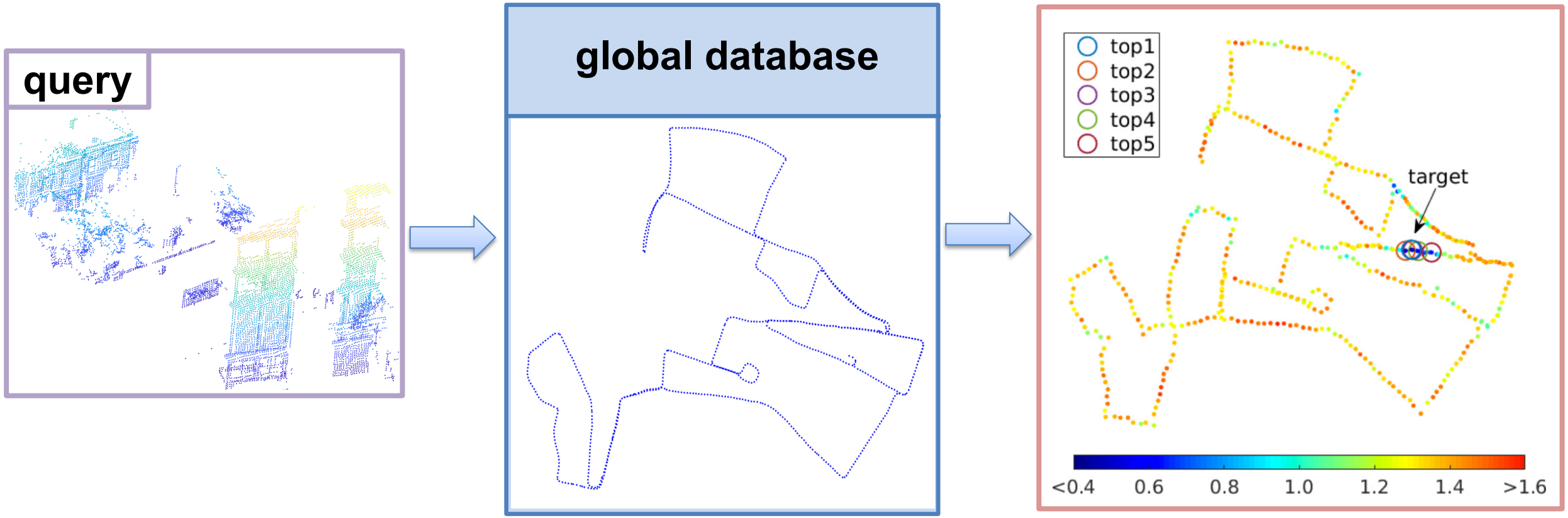
Network training: global part



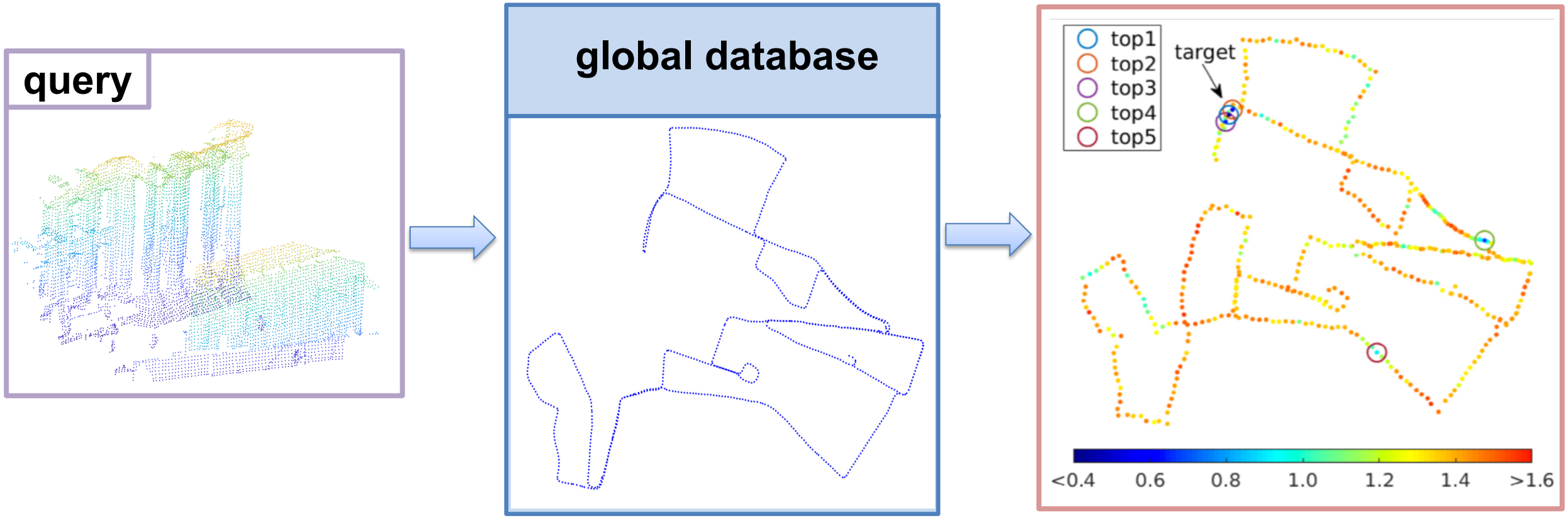
Global descriptor loss

- Lazy quadruplet loss (following PointNetVLAD [Uy etal,2018])

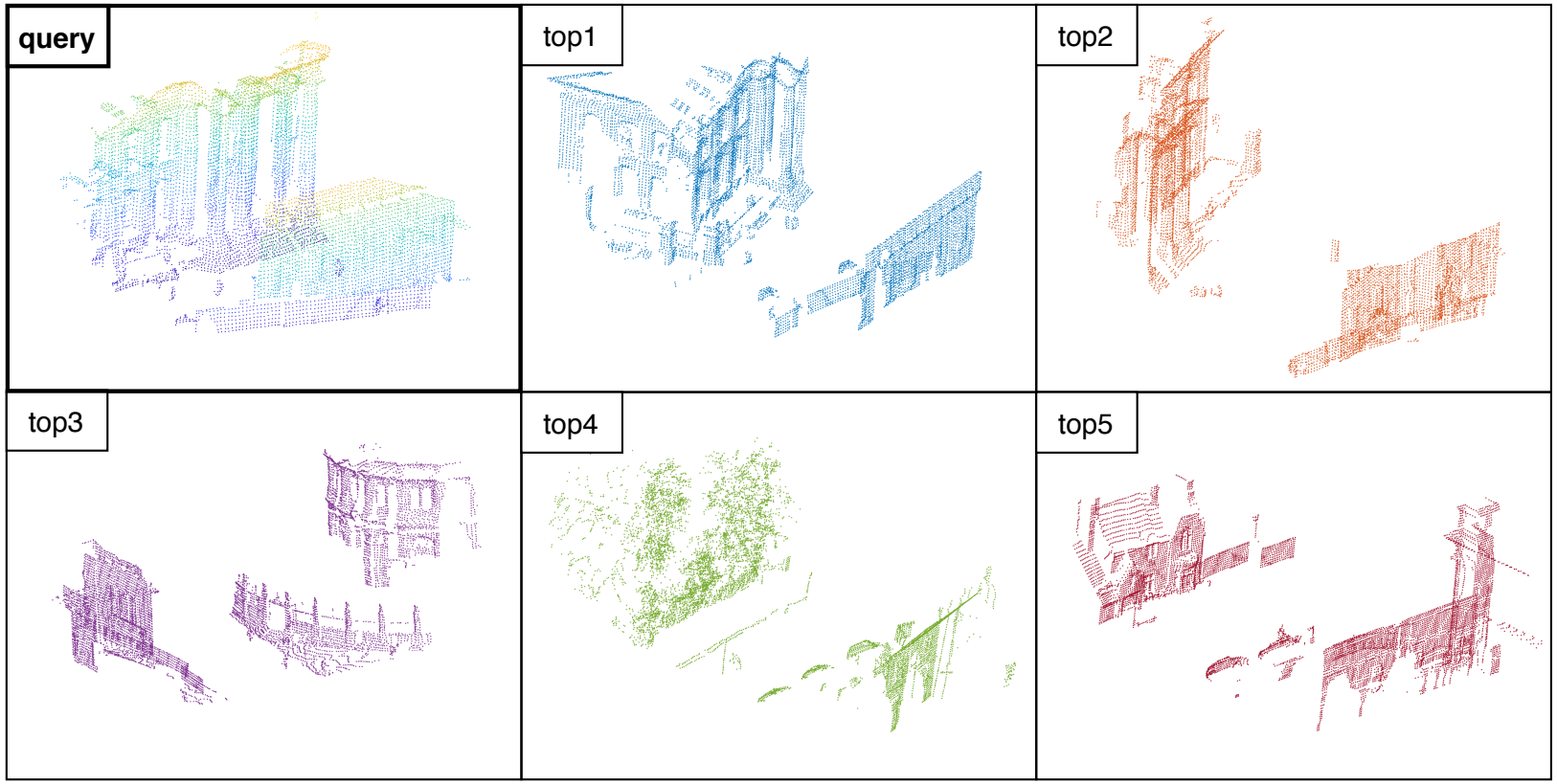
Experiments: point cloud retrieval (the Oxford RobotCar)

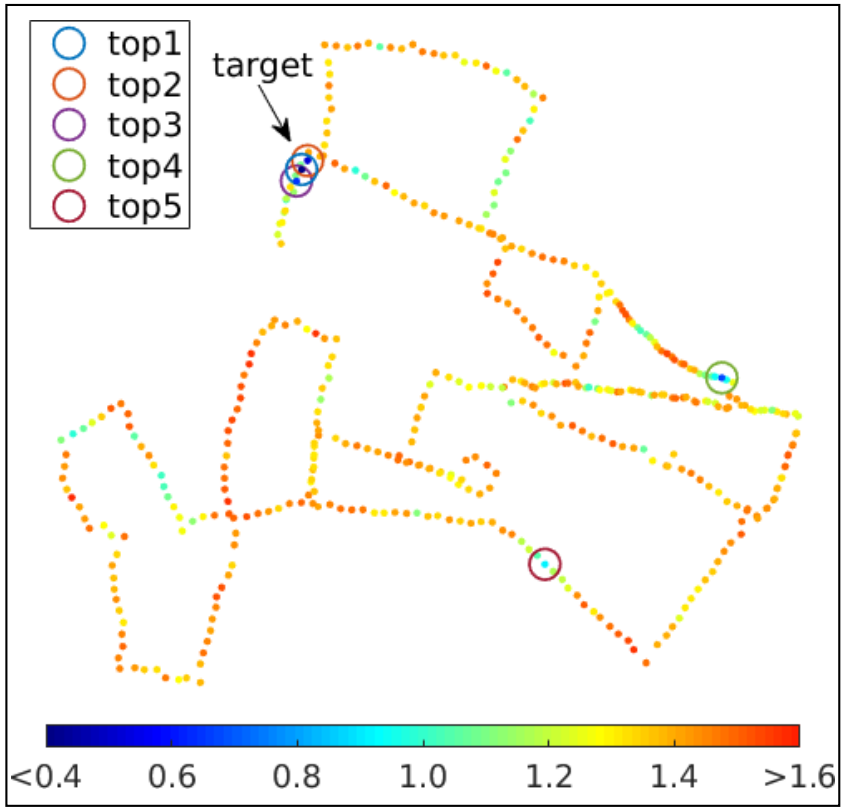


Experiments: point cloud retrieval (the Oxford RobotCar)

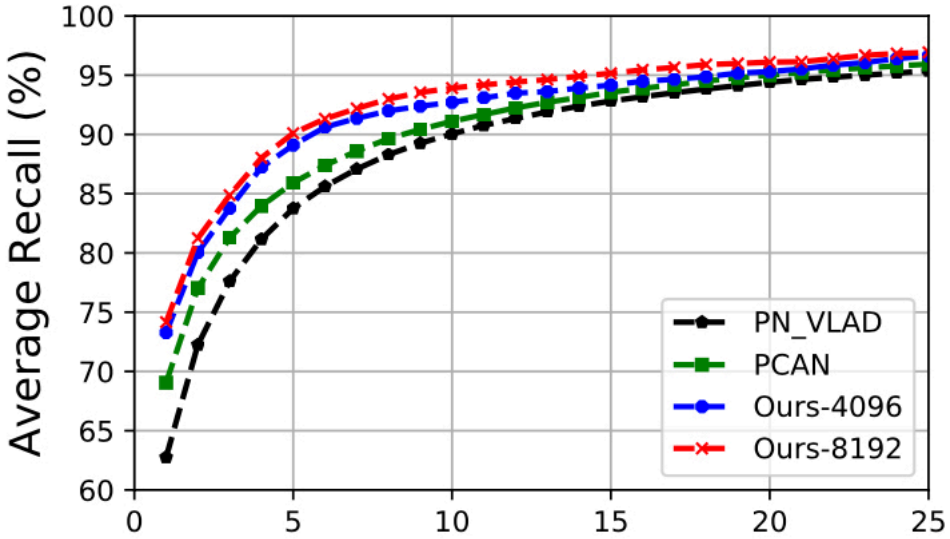


Experiments: point cloud retrieval (the Oxford RobotCar)





Experiments: point cloud retrieval (the Oxford RobotCar)

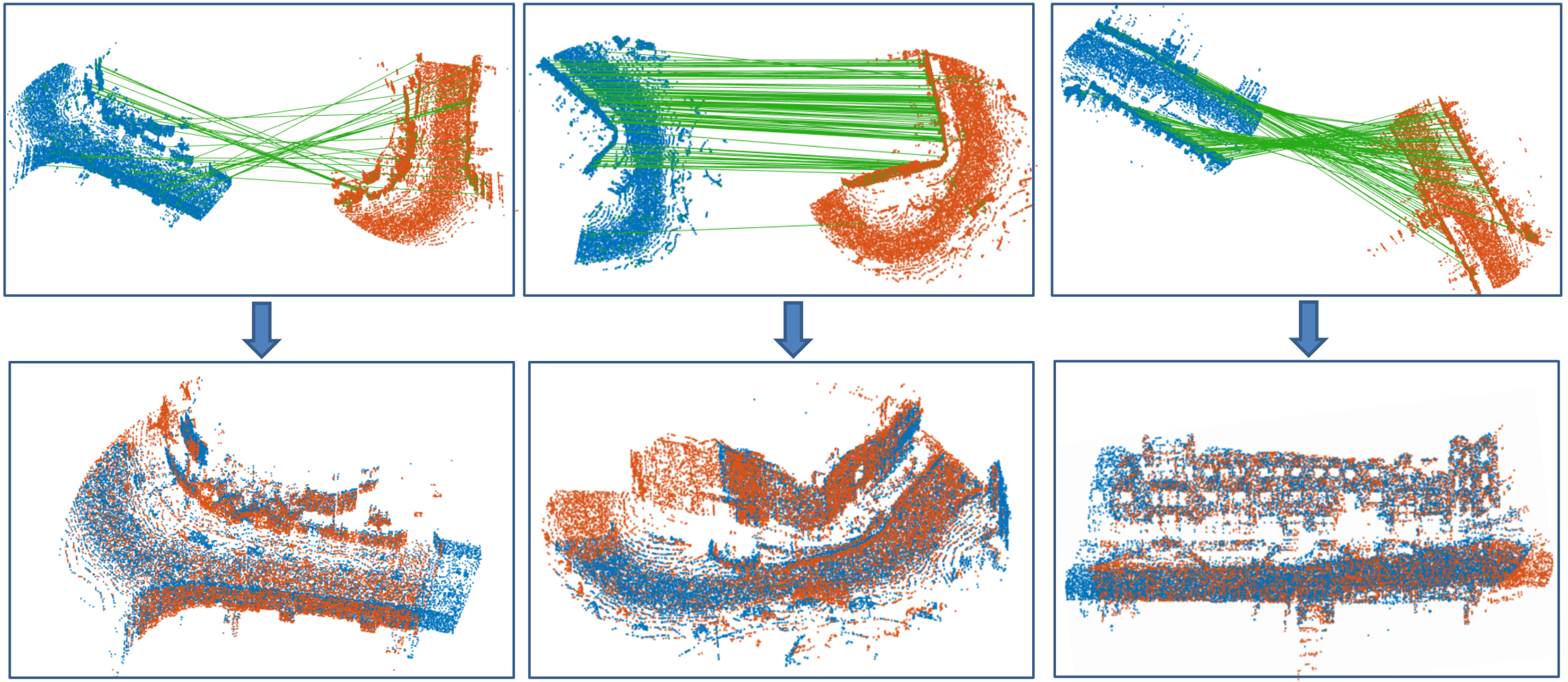


Average recall of the top 25 retrievals

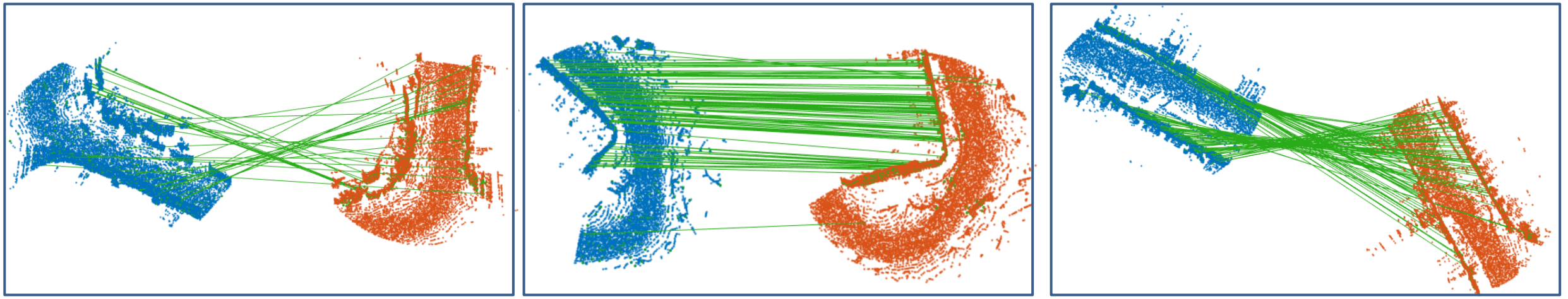
Method	Ours(8192)	Ours(4096)	PN_VLAD	PCAN	PN_MAX
Ave recall @ 1%	85.30	84.26	81.01	83.81	73.44
Ave recall @ 1	74.16	73.28	62.18	69.76	58.46

Average recall (%) at top 1% and at top 1

Experiments: point cloud registration (the Oxford RobotCar)



Experiments: point cloud registration (the Oxford RobotCar)

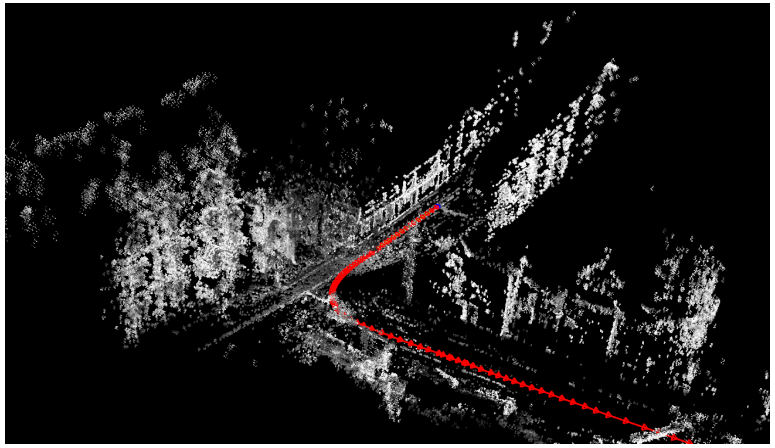
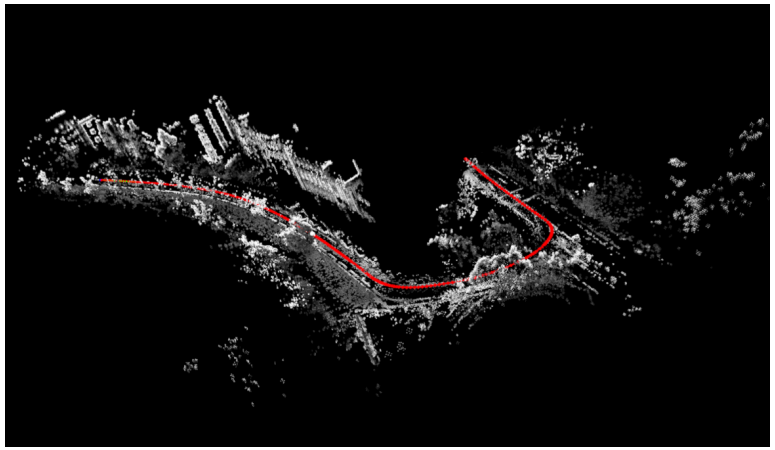
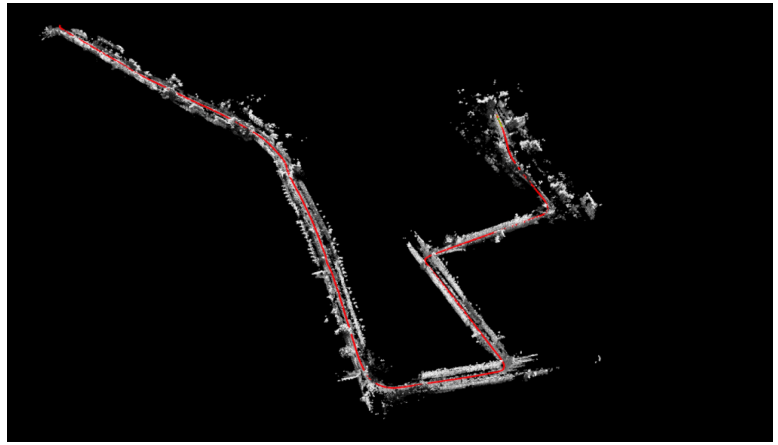


Method (det + desc)	RTE (m)	RRE (deg)	Succ. Rate	Iterations
3DFeatNet +3DSmoothNet	0.34	1.34	95.1%	7280
DH3D + 3DSmoothNet	0.32	1.22	96.0%	3904
3DFeatNet + 3DFeatNet	0.30	1.07	98.1%	2940
3DFeatNet + **DH3D**	0.32	1.24	95.4%	2489
DH3D + DH3D	0.23	1.04	98.5%	1972

Point cloud registration performance on the Oxford RobotCar LiDAR points

Experiments: point cloud registration on **vSLAM points**

- Point clouds generated by Stereo DSO
- Images from 8 sequences covering different conditions
- 318 pairwise poses for testing (no fine-tunning)

Experiments: point cloud registration on **vSLAM points**

- Point clouds generated by Stereo DSO
- Images from 8 sequences covering different conditions
- 318 pairwise poses for testing (no fine-tunning)

Method (det + desc)	RTE (m)	RRE (deg)	Succ. Rate
3DFeatNet +3DSmoothNet	0.38	2.22	66.6%
DH3D + 3DSmoothNet	0.35	2.01	77.9%
3DFeatNet + 3DFeatNet	0.92	1.97	84.1%
3DFeatNet + **DH3D**	0.74	2.38	80.9%
DH3D + **DH3D**	0.36	1.58	90.6%

Point cloud registration performance on vSLAM points

Experiments: point cloud registration on **vSLAM points**

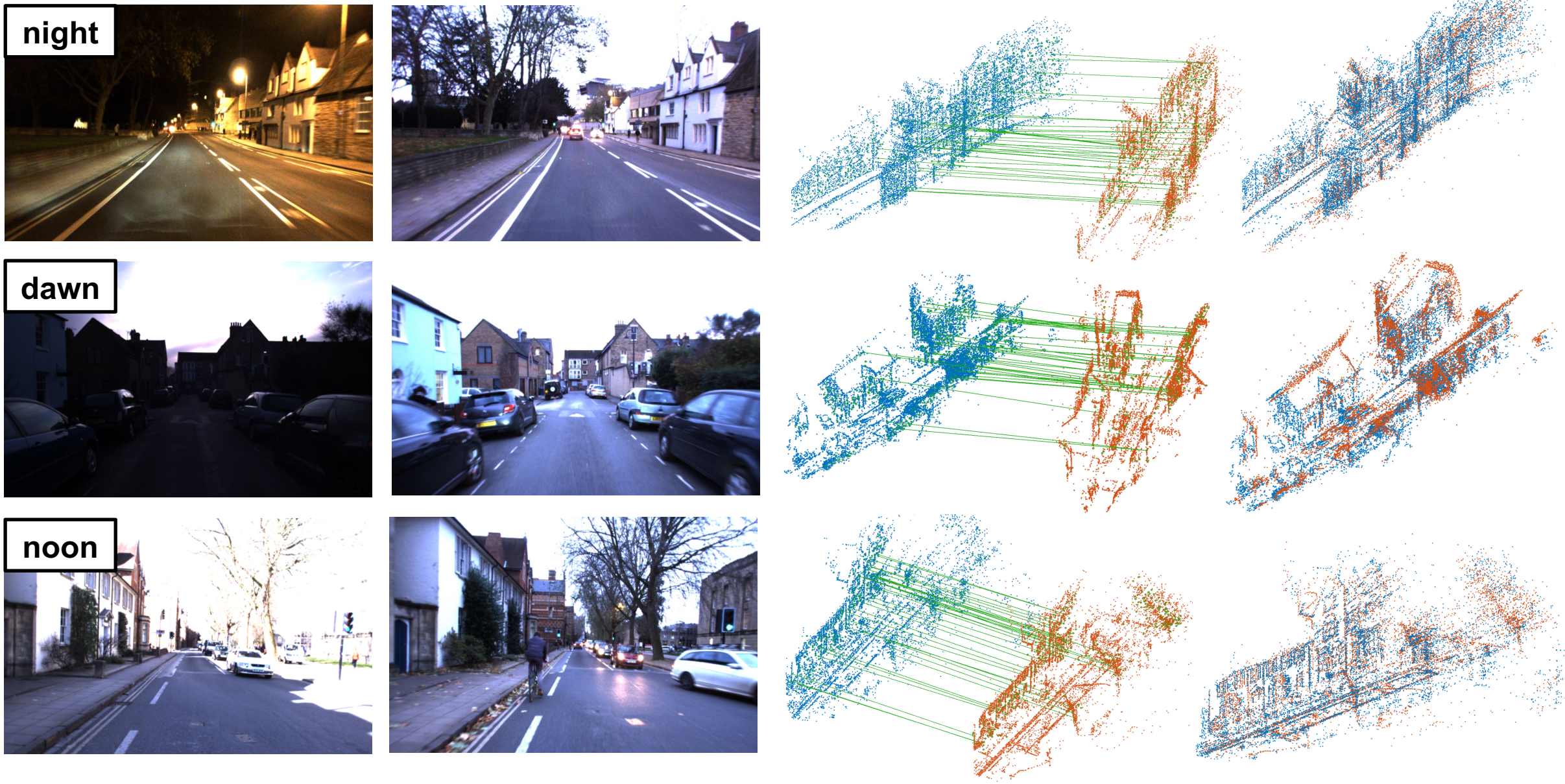
- Point clouds generated by Stereo DSO
- Images from 8 sequences covering different conditions
- 318 pairwise poses for testing (no fine-tunning)

Method (det + desc)	RTE (m)	RRE (deg)	Succ. Rate
3DFeatNet +3DSmoothNet	0.38 (0.34)	2.22 (1.34)	66.6% (95.1%)
DH3D + 3DSmoothNet	0.35 (0.32)	2.01 (1.22)	77.9% (96.0%)
3DFeatNet + 3DFeatNet	0.92 (0.30)	1.97 (1.07)	84.1% (98.1%)
3DFeatNet + **DH3D**	0.74 (0.32)	2.38 (1.24)	80.9% (95.4%)
DH3D + **DH3D**	0.36 (0.23)	1.58 (1.04)	90.6% (98.5%)

Point cloud registration performance on vSLAM points

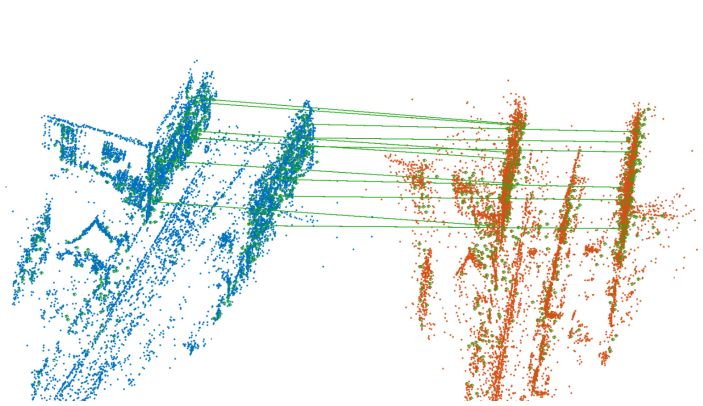
* Results on LiDAR points are shown in parentheses for comparison

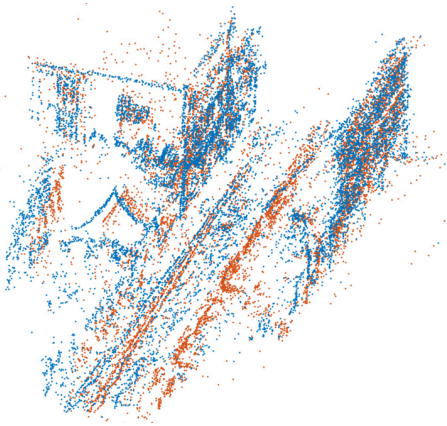
Experiments: point cloud registration on **vSLAM points**



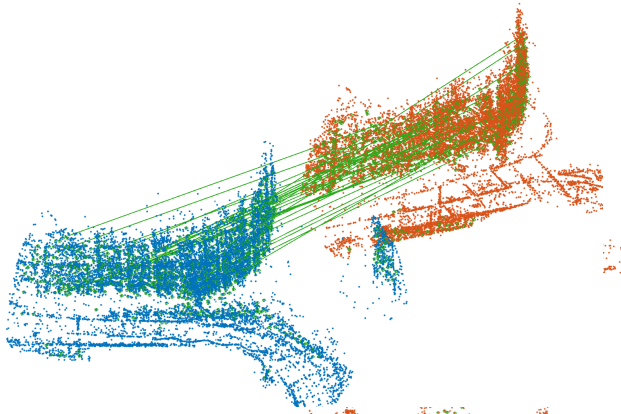
Experiments: point cloud registration on **vSLAM points**

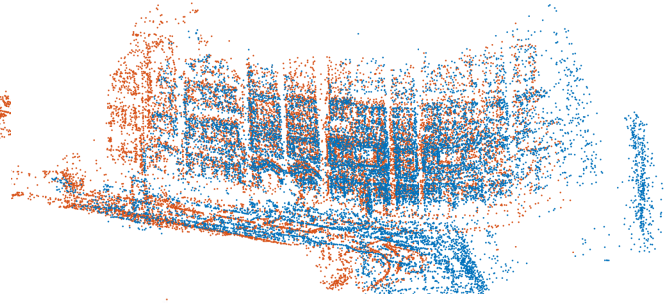
rain



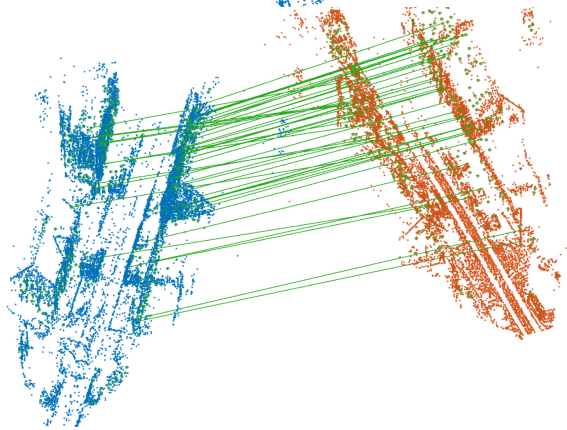


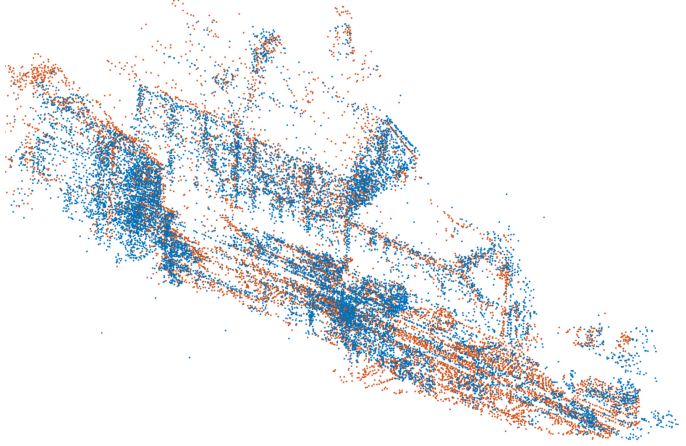
construction





snow





Please visit our project page for

code and other materials:

https://vision.in.tum.de/research/vslam/dh3d

Thank you!