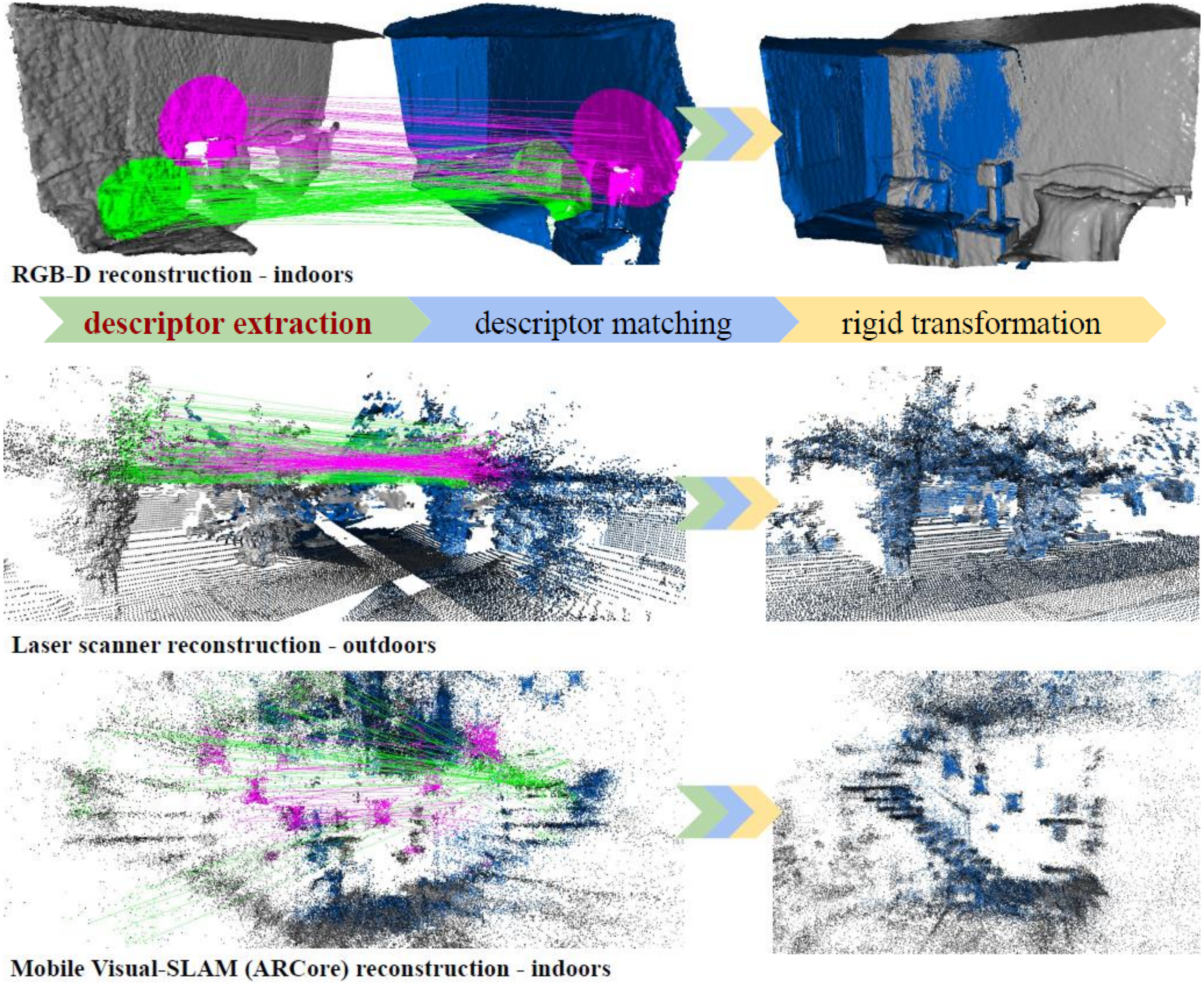
Distinctive 3D local deep descriptors

Fabio Poiesi and Davide Boscaini

TeV TECHNOLOGIES OF VISION

FONDAZIONE BRUNO KESSLER

Acknowledgement: FONDAZIONE CARITRO CASSA DI RISPARMIO DI TRENTO E ROVERETO


RGB-D reconstruction - indoors
descriptor extraction
descriptor matching
rigid transformation
Laser scanner reconstruction - outdoors
Mobile Visual-SLAM (ARCore) reconstruction - indoors

Motivation

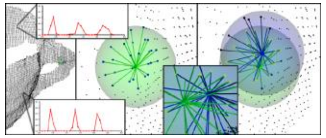
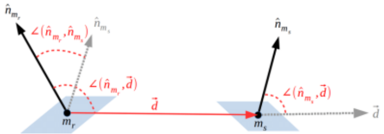
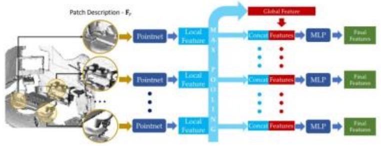
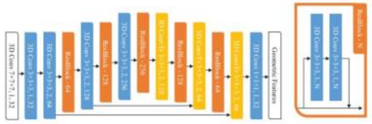
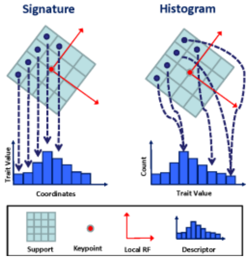
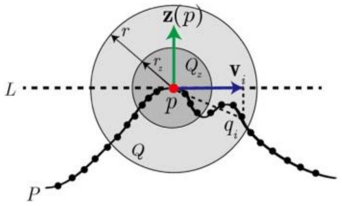
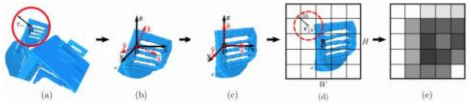
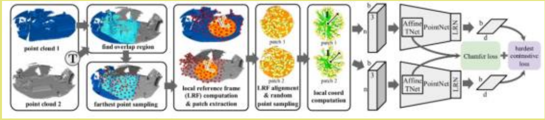
Compact descriptor

Efficient to compute

Generalise across sensor modalities

Learnable end-to-end

3D descriptors for PCD (overview)

	Hand crafted		Data driven	
One-stage (without LRF)	FPFH [1] 	PPF [2] 	PPFNet [5] 	FCGF [6] 
Two-stage (with LRF)	SHOT [3] 	TOLDI [4] 	3DSmoothNet [7]  **DIP** 	

LRF: Local Reference Frame

References

[1] R.B. Rusu et al., "Fast Point Feature Histograms (FPFH) for 3D registration," ICRA 2009

[2] B. Drost, et al., "Model globally, match locally: Efficient and robust 3d object recognition," CVPR 2010

[3] F. Tombari et al., “Unique Signatures of Histograms for Local Surface Description,” ECCV 2010

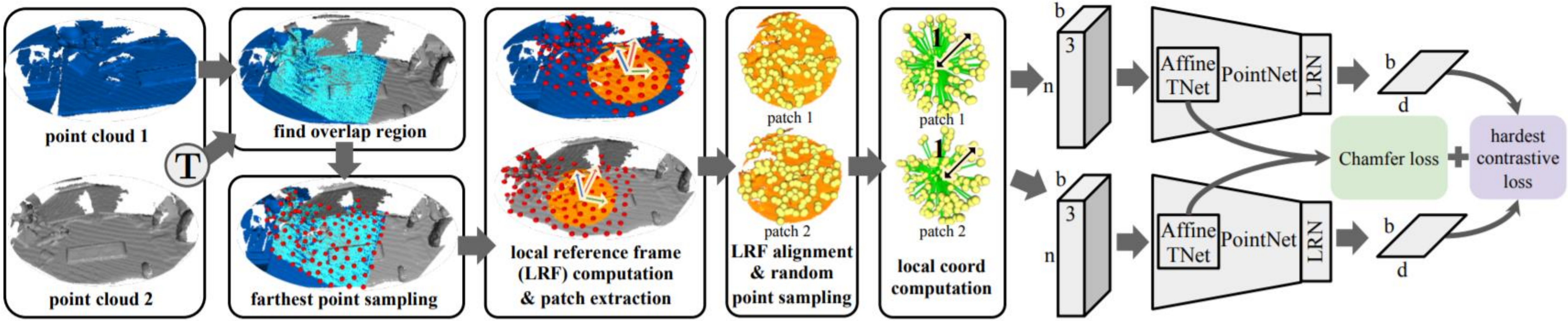
[4] Yang et al., “TOLDI: An effective and robust approach for 3D local shape description,” Patt. Rec. 2017

[5] Deng et al., “PPFNet: Global Context Aware Local Features for Robust 3D Point Matching,” CVPR 2018

[6] Choy et al. “Fully Convolutional Geometric Features,” ICCV 2019

[7] Gojcic et al., “The Perfect Match: 3D Point Cloud Matching with Smoothed Densities,” CVPR 2019

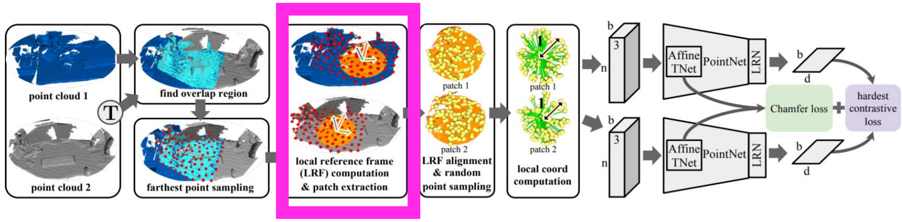
How do we learn DIPs?

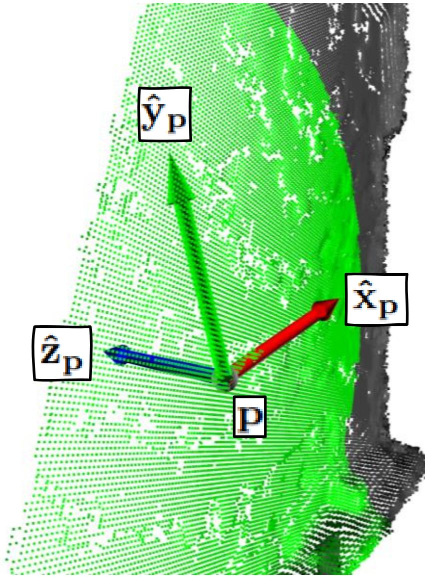


[Qi2017a] [Yang2017] [Gojcic2019] [Qi2017b]

[Qi2017a] Qi et al., “PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space,” NeurIPS 2017
[Yang2017] Yang et al., “TOLDI: An effective and robust approach for 3D local shape description,” Pattern Recognition 2017
[Gojcic2019] Gojcic et al., “The Perfect Match: 3D Point Cloud Matching with Smoothed Densities,” CVPR 2019
[Qi2017b] Qi et al., “PointNet: Deep learning on point sets for 3D classification and segmentation,” CVPR 2017

TOLDI LRF [Yang2017,Gojcic2019]





$$\tilde{\Sigma}_\mathcal{S} = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{p}_i \in \mathcal{S}} (\mathbf{p}_i - \mathbf{p})(\mathbf{p}_i - \mathbf{p})^\mathrm{T}$$

$$\mathcal{S} = \{\mathbf{p}_i : ||\mathbf{p}_i - \mathbf{p}||_2 \leq r_{LRF}\}$$

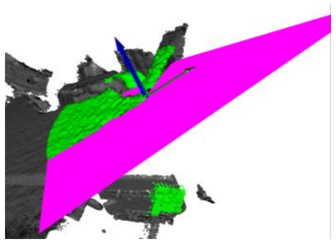
$$\hat{\mathbf{z}}_\mathbf{p} = \begin{cases} \hat{\mathbf{n}}_\mathbf{p}, & \text{if } \sum_{\mathbf{p}_i \in \mathcal{S}} \langle \hat{\mathbf{n}}_\mathbf{p}, \overrightarrow{\mathbf{p}_i\mathbf{p}} \rangle \geq 0 \\ -\hat{\mathbf{n}}_\mathbf{p}, & \text{otherwise} \end{cases}$$

$\hat{\mathbf{n}}_\mathbf{p}$ eigenvector corresponding to the smallest eigenvalue of $\tilde{\Sigma}_\mathcal{S}$ (normalised)

$$\hat{\mathbf{y}}_\mathbf{p} = \hat{\mathbf{x}}_\mathbf{p} \times \hat{\mathbf{z}}_\mathbf{p}$$

$$\hat{\mathbf{x}}_\mathbf{p} = \frac{1}{|| \sum_{\mathbf{p}_i \in \mathcal{S}} \alpha_i \beta_i \mathbf{v}_i ||_2} \sum_{\mathbf{p}_i \in \mathcal{S}} \alpha_i \beta_i \mathbf{v}_i$$

$$\mathbf{v}_i = \overrightarrow{\mathbf{p}\mathbf{p}_i} - \langle \overrightarrow{\mathbf{p}\mathbf{p}_i}, \hat{\mathbf{z}}_\mathbf{p} \rangle \hat{\mathbf{z}}_\mathbf{p}$$

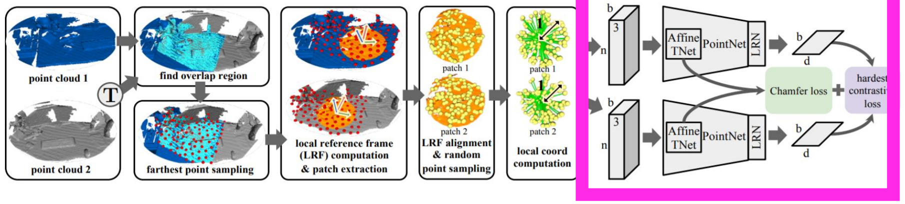


$$\alpha_i = (r_{LRF} - ||\mathbf{p} - \mathbf{p}_i||_2)^2$$
$$\beta_i = \langle \overrightarrow{\mathbf{p}\mathbf{p}_i}, \hat{\mathbf{z}}_\mathbf{p} \rangle^2$$

[Yang2017] Yang et al., “TOLDI: An effective and robust approach for 3D local shape description,” Pattern Recognition 2017

[Gojcic2019] Gojcic et al., “The Perfect Match: 3D Point Cloud Matching with Smoothed Densities,” CVPR 2019

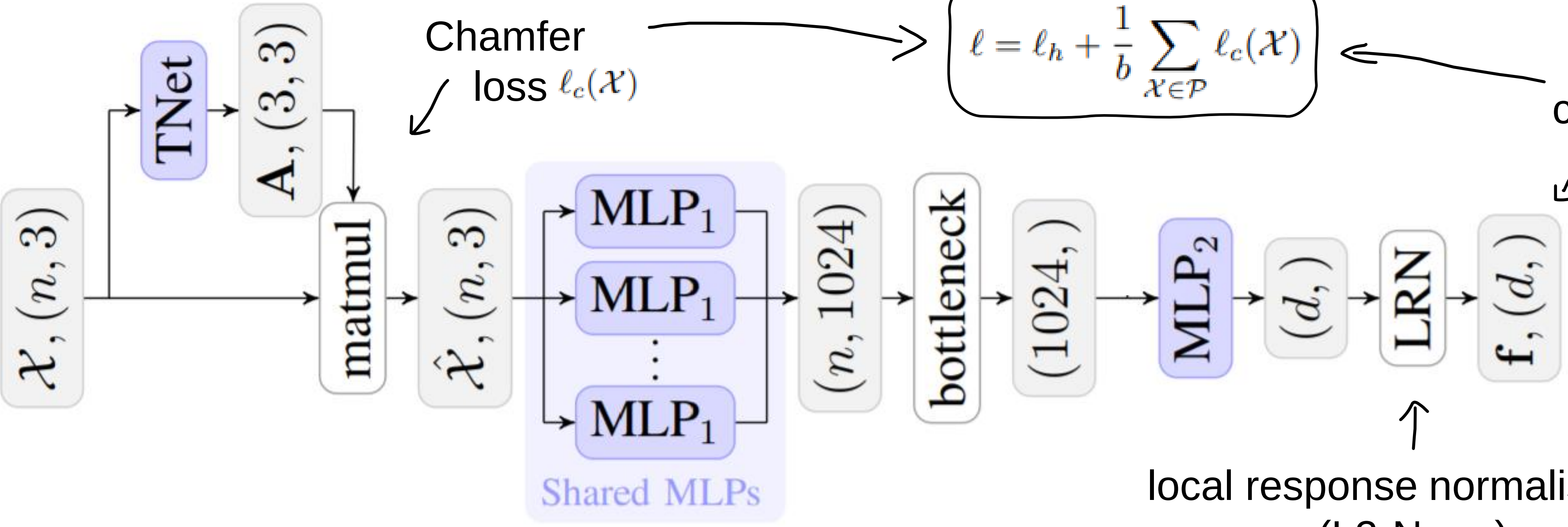
PointNet [Qi2017]



Chamfer loss $\ell_c(\mathcal{X})$

$$\ell = \ell_h + \frac{1}{b}\sum_{\mathcal{X}\in\mathcal{P}} \ell_c(\mathcal{X})$$

Hardest contrastive loss ℓ_h

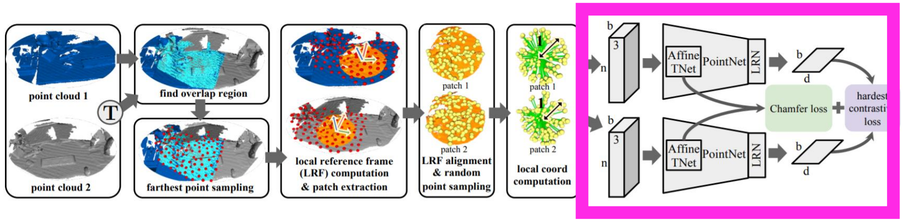


local response normalisation (L2-Norm)

- **input**: LRF-rotated patch points (n = number of points)
- **output**: descriptor (d = 32)
- trained via Siamese approach

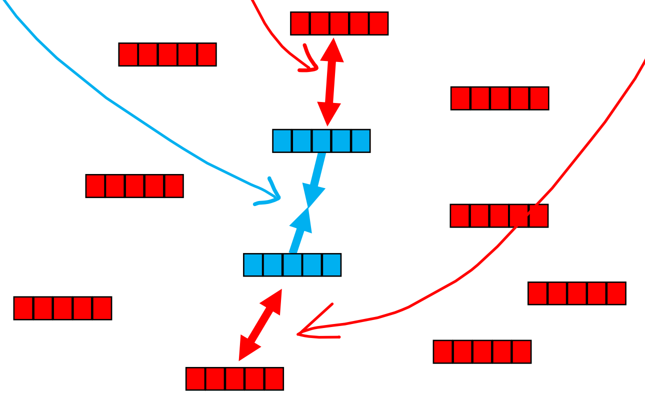
[Qi2017b] Qi et al., “PointNet: Deep learning on point sets for 3D classification and segmentation,” CVPR 2017

Hardest contrastive loss



Intuition: within a batch, make the descriptors of positive pairs close in the embedding space while making the other descriptors distant

$$\ell_h = \frac{1}{b}\sum_{(\mathbf{f},\mathbf{f}')\in\mathcal{C}_+}\left(\boxed{\frac{1}{|\mathcal{C}_+|}\left[d(\mathbf{f},\mathbf{f}') - m_+\right]_+^2} + \boxed{\frac{1}{2|\mathcal{C}_-|}\Big[m_- - \underbrace{\min_{\tilde{\mathbf{f}}\in\mathcal{C}_-} d(\mathbf{f},\tilde{\mathbf{f}})}_{d(\mathbf{f},\mathbf{f}_-)}\Big]_+^2} + \boxed{\frac{1}{2|\mathcal{C}_-|}\Big[m_- - \underbrace{\min_{\tilde{\mathbf{f}}\in\mathcal{C}_-} d(\mathbf{f}',\tilde{\mathbf{f}})}_{d(\mathbf{f}',\mathbf{f}'_-)}\Big]_+^2}\right)$$



$\mathbf{f}$ descriptor

$(\mathbf{f}, \mathbf{f}')$ pair of anchors

$(\mathbf{f}_-, \mathbf{f}'_-)$ hardest negatives

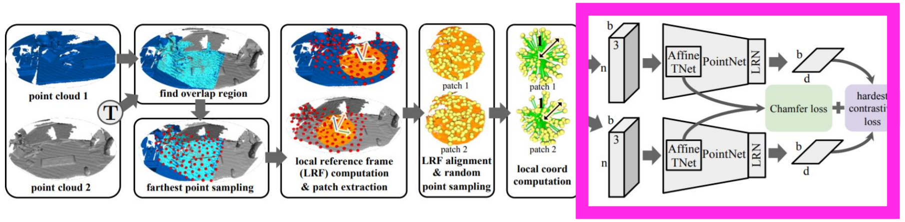
m_- m_+ margins

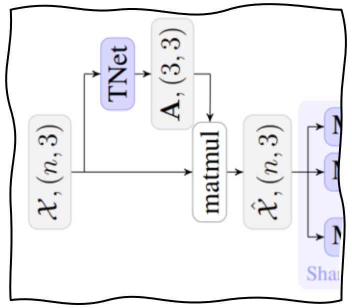
Descriptors (illustration)

positive

negative

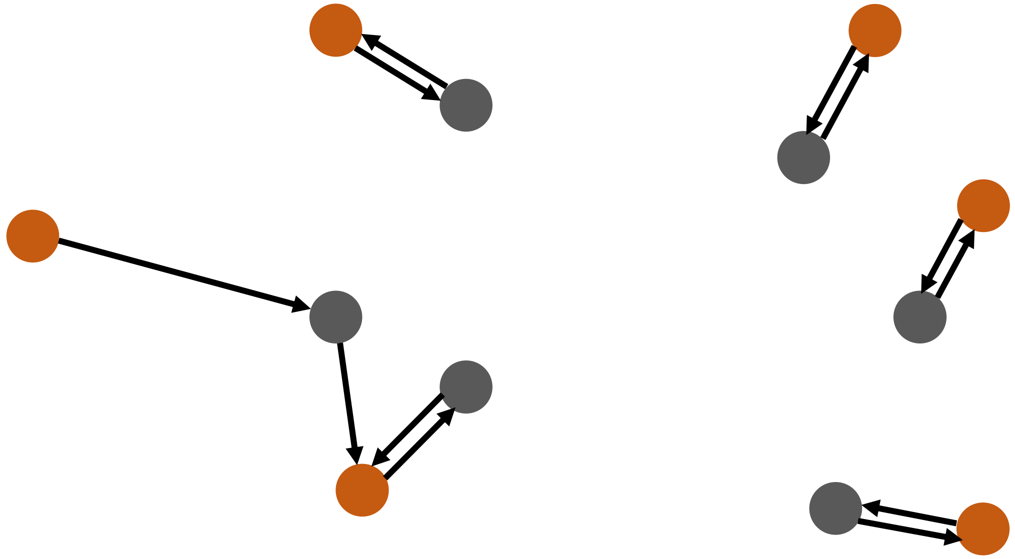
Chamfer loss





$$\ell_c(\mathcal{X}) = \frac{1}{2n}\left(\sum_{\mathbf{x}\in\mathcal{X}} \min_{\mathbf{x}'\in\mathcal{X}'} \|\mathbf{A}\mathbf{x}-\mathbf{A}'\mathbf{x}'\|_2 + \sum_{\mathbf{x}'\in\mathcal{X}'} \min_{\mathbf{x}\in\mathcal{X}} \|\mathbf{A}\mathbf{x}-\mathbf{A}'\mathbf{x}'\|_2\right)$$

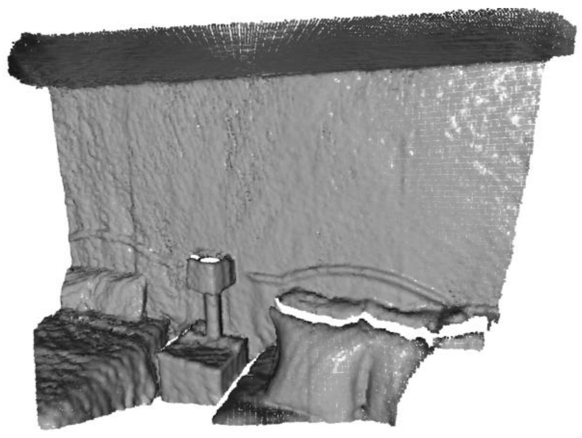
$\mathbf{A} \in \mathbb{R}^{3\times 3}$ unconstrained



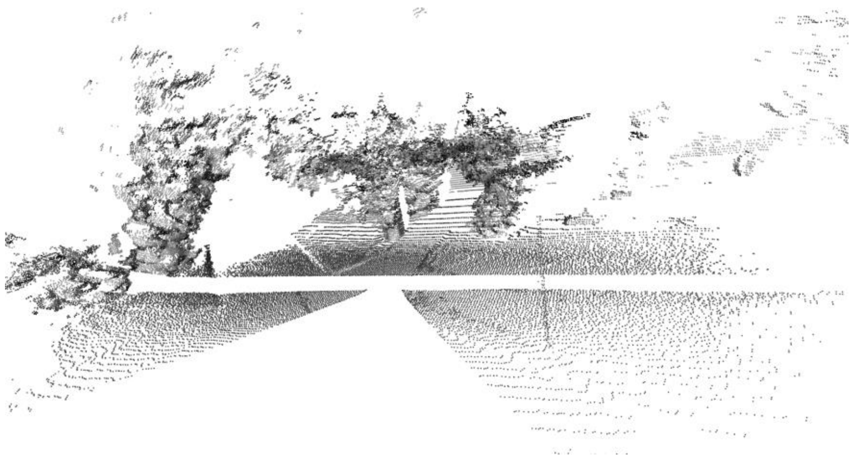
tried $\ell_{\text{reg}} = \|\mathbf{I} - \mathbf{A}\mathbf{A}^\top\|_F^2$ → $\mathbf{A} \to \mathbf{I}$ → no contribution

Experiments

- Training
  - 3DMatch dataset
  - about 16K point-cloud pairs
  - each pair is 256 descriptors
  - 40 epochs
- Testing
  - 3DMatch, ETH
- Evaluation
  - Feature Matching Recall **[Deng2018]**



3DMatch **[Zeng2017]**
RGBD



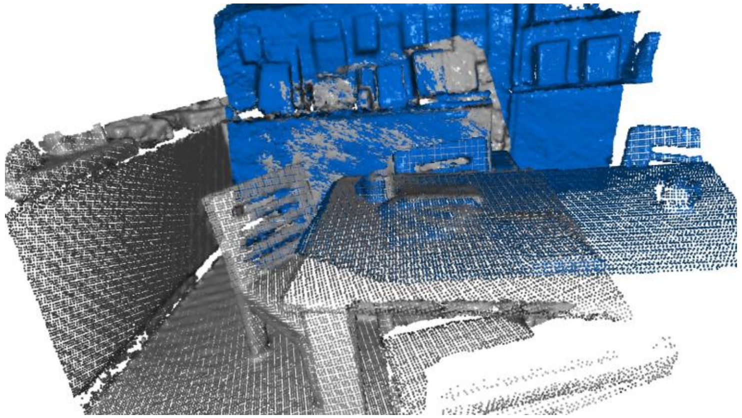
ETH **[Pomerleau2017]**
LIDAR

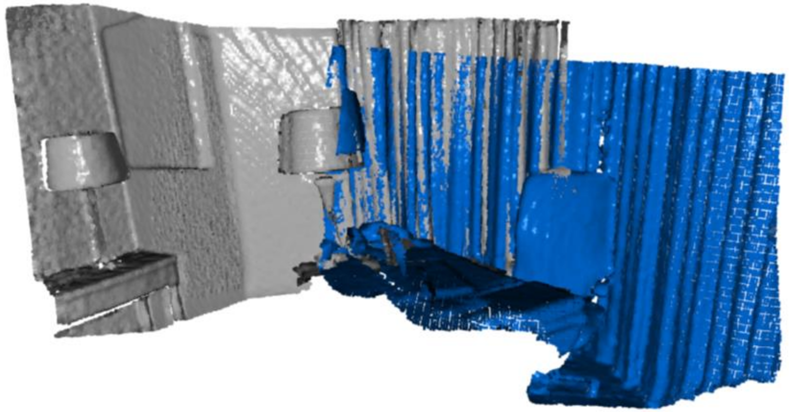
[Zeng2017] Zeng et al., "3DMatch: Learning the matching of local 3D geometry in range scans," CVPR 2017

[Pomerleau2017] Pomerleau et al., "Challenging data sets for point cloud registration algorithms," IJRR 2012

[Deng2018] Deng et al., “PPFNet: Global Context Aware Local Features for Robust 3D Point Matching,” CVPR 2018

3DMatch dataset

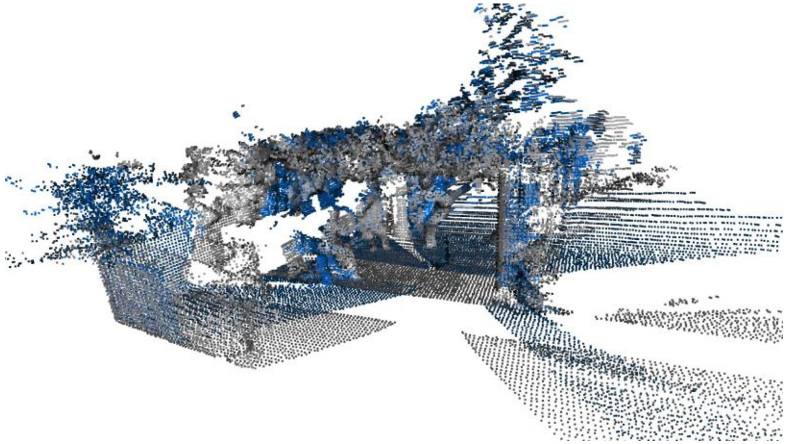


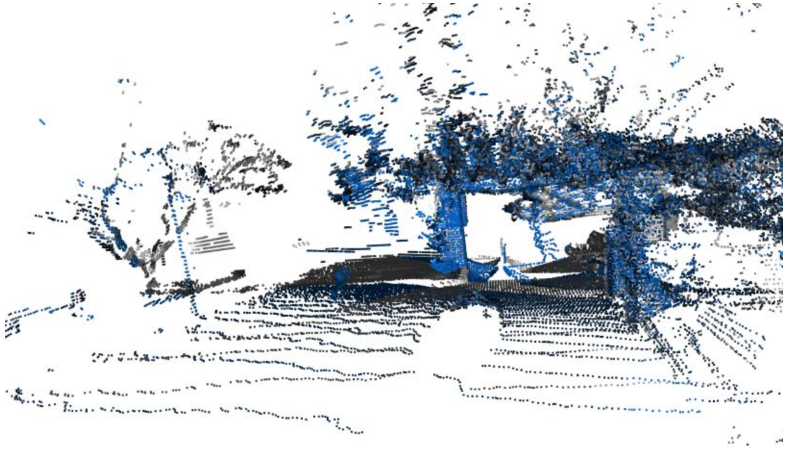


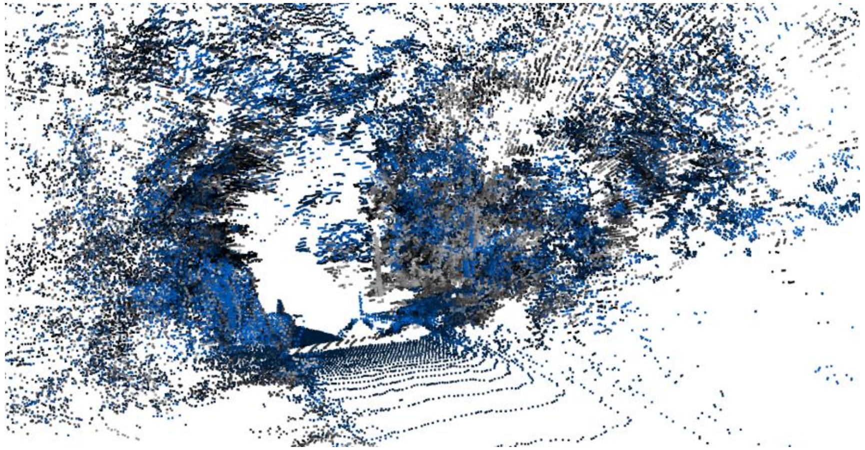
FEATURE-MATCHING RECALL ON THE 3DMATCH DATASET [4].

Method	3DMatch Ξ	3DMatch std	3DMatchRotated Ξ	3DMatchRotated std	Feat. dim.	Time [ms]
Spin [3]	.227	.114	.227	.121	153	.133
SHOT [17]	.238	.109	.234	.095	352	.279
FPFH [16]	.359	.134	.364	.136	33	.032
USC [36]	.400	.125	-	-	1980	3.712
CGF [37]	.582	.142	.585	.140	32	1.463
3DMatch [4]	.596	.088	.011	.012	512	3.210
Folding [5]	.613	.087	.023	.010	512	.352
PPFNet [6]	.623	.108	.003	.005	64	2.257
PPF-FoldNet [7]	.718	.105	.731	.104	512	.794
DirectReg [8]	.746	.094	-	-	512	.794
CapsuleNet [9]	.807	.062	.807	.062	512	1.208
PerfectMatch [10]	.947	.027	.949	.024	32	5.515
FCGF [11]	.952	.029	.953	.033	32	.009
D3Feat [12]	**.958**	.029	**.955**	.035	32	-
DIP	.948	.046	.946	.046	32	4.870

ETH Dataset







FEATURE-MATCHING RECALL ON THE ETH DATASET [22].

Method	Gazebo Summer	Gazebo Winter	Wood Autumn	Wood Summer	Average
FPFH [16]	.386	.142	.148	.208	.221
SHOT [17]	.739	.457	.609	.640	.611
3DMatch [4]	.228	.083	.139	.224	.169
CGF [37]	.375	.138	.104	.192	.202
PerfectMatch [10]	**.913**	.841	.678	.728	.790
FCGF [11]	.228	.100	.148	.168	.161
D3Feat [12]	.859	.630	.496	.480	.563
DIP	.908	**.886**	**.965**	**.952**	**.928**

Conclusions

- Compact descriptor
- Efficient to compute
- Generalise across sensor modalities
- Learnable end-to-end

- Some interesting new research can be explored
  - e.g. 6D pose estimation

CODE AVAILABLE
https://github.com
/fabiopoiesi/dip