RoRD : Rotation Robust Descriptors and Orthographic Views for Local Feature Matching

Udit Singh Parihar

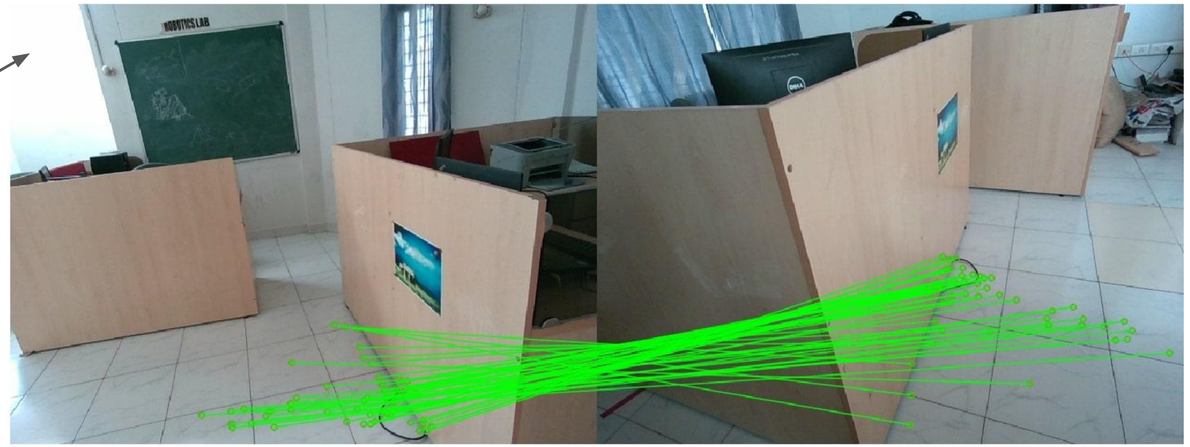
Accepted IROS 2021

Problem Statement

- Developing local descriptors invariant to high viewpoints
- Use in relative pose estimation
- Front-end pipelines in SLAM system
- Image Retrievals

Introduction

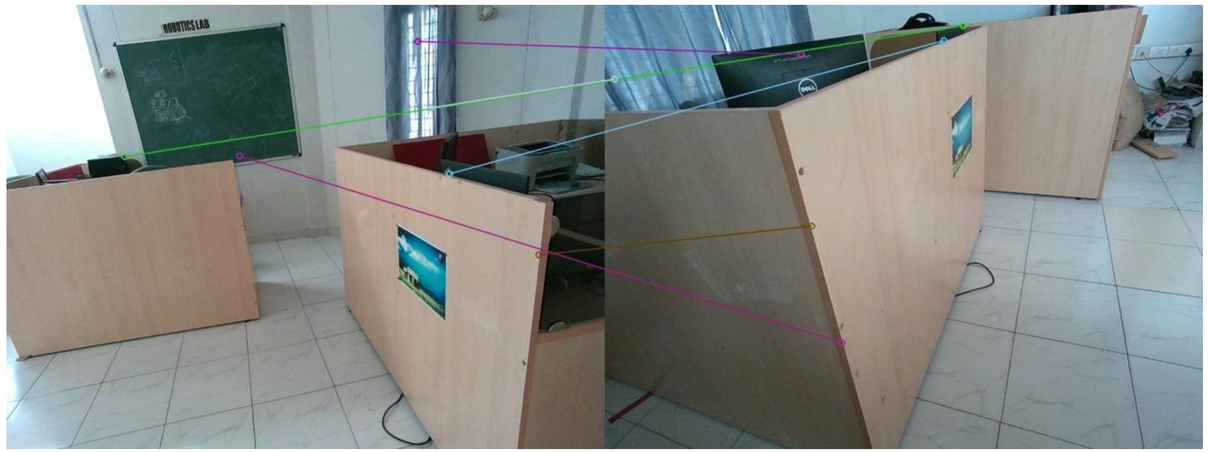
RORD (ours)



✔

Feature Correspondences

x



SIFT

Contributions

- Proposed deep learnt descriptors which works under extreme viewpoints
- Used self-supervised training to generate huge training data with simple homographic transformations
- Achieves good generalisation across datasets and perform state of art in variety of tasks
- Proposed dataset comprising of images from high change in camera viewpoint
- Dual headed architecture, where one head is rotation invariant and other is illumination invariant

Correspondences from RoRD

- Results from extreme viewpoints
- Works in both Indoor and outdoor

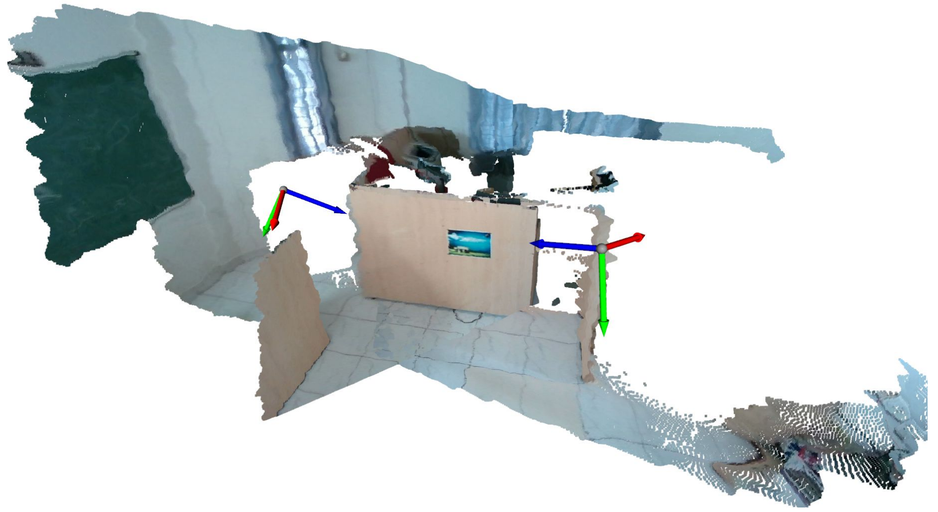
Improvements

- MMA in HPatches Dataset
- Pose estimation on DiverseView Dataset
- Use of rotation robust descriptors in Visual Place Recognition Task

Supervised Training

SfM Data

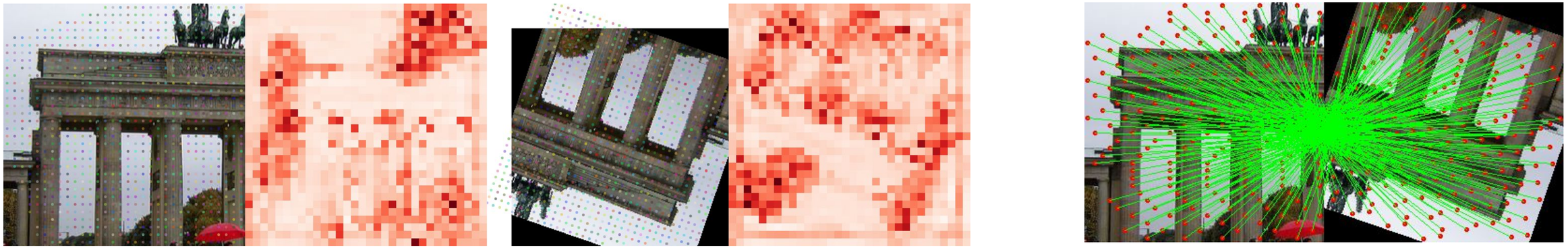
- Generates correspondences from monocular images by applying SfM
- Data labelling doesn’t require human labelling
- Inherit biases from SfM data pipeline which uses classical methods



Self Supervised Training

Homographic Transformations

- Training on primitive synthetic shapes and deploying on real world scenarios

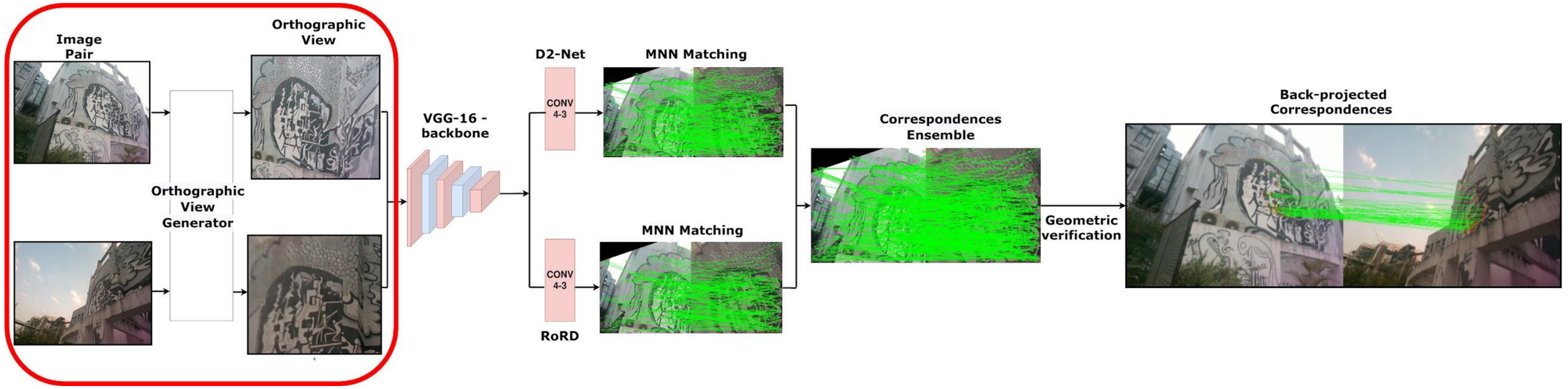


Orthographic Views

- Intermediate representation of scene as orthographic views
- Orthographic views improves image retrieval, feature matching and planning
- Ability to calculates orthographic views during 6 DoF camera motion

Pipeline

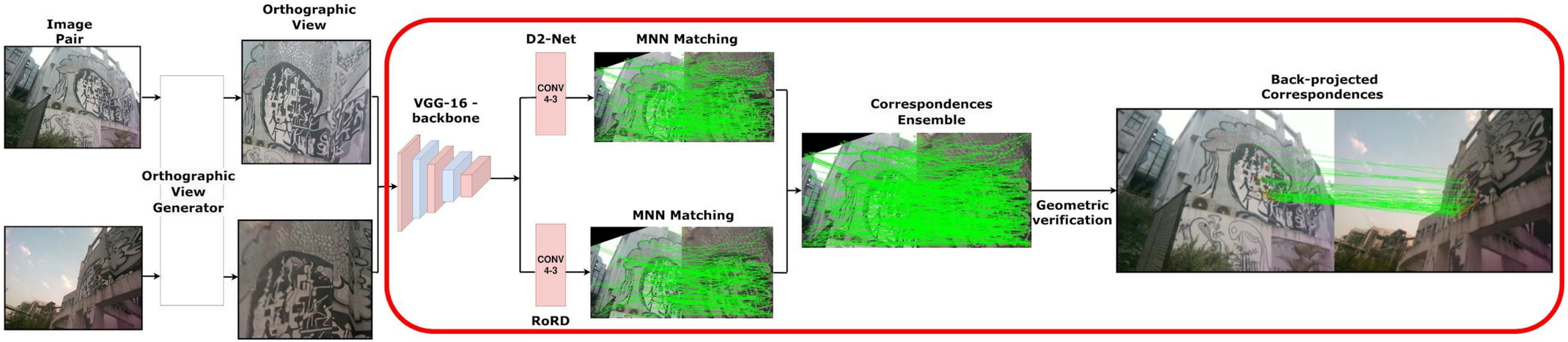
Orthographic view



- Orthographic view rectifies the perspectivity
- Planar patches extraction and surface normal calculation

Pipeline

Ensemble Architecture

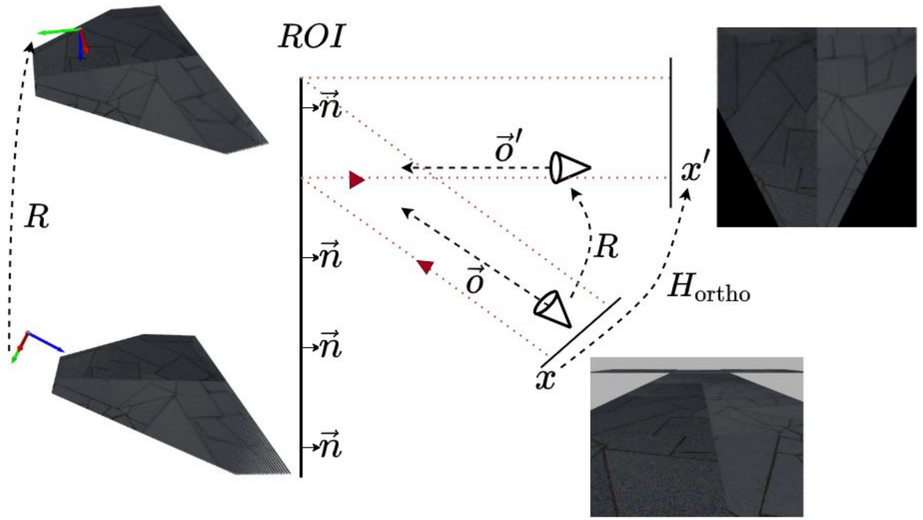


- Dual headed architecture for illumination and rotation invariance
- D2Net model is supervised using 3D depth and pose information
- RoRD head is trained using self-supervised rotational homographies

Orthographic View Generation

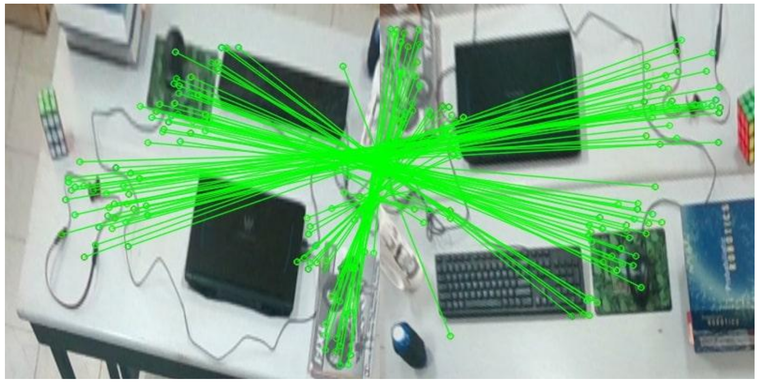
- Depth information along with desired ROI for 3D scene representation

- Virtual Camera anti parallel to surface normal



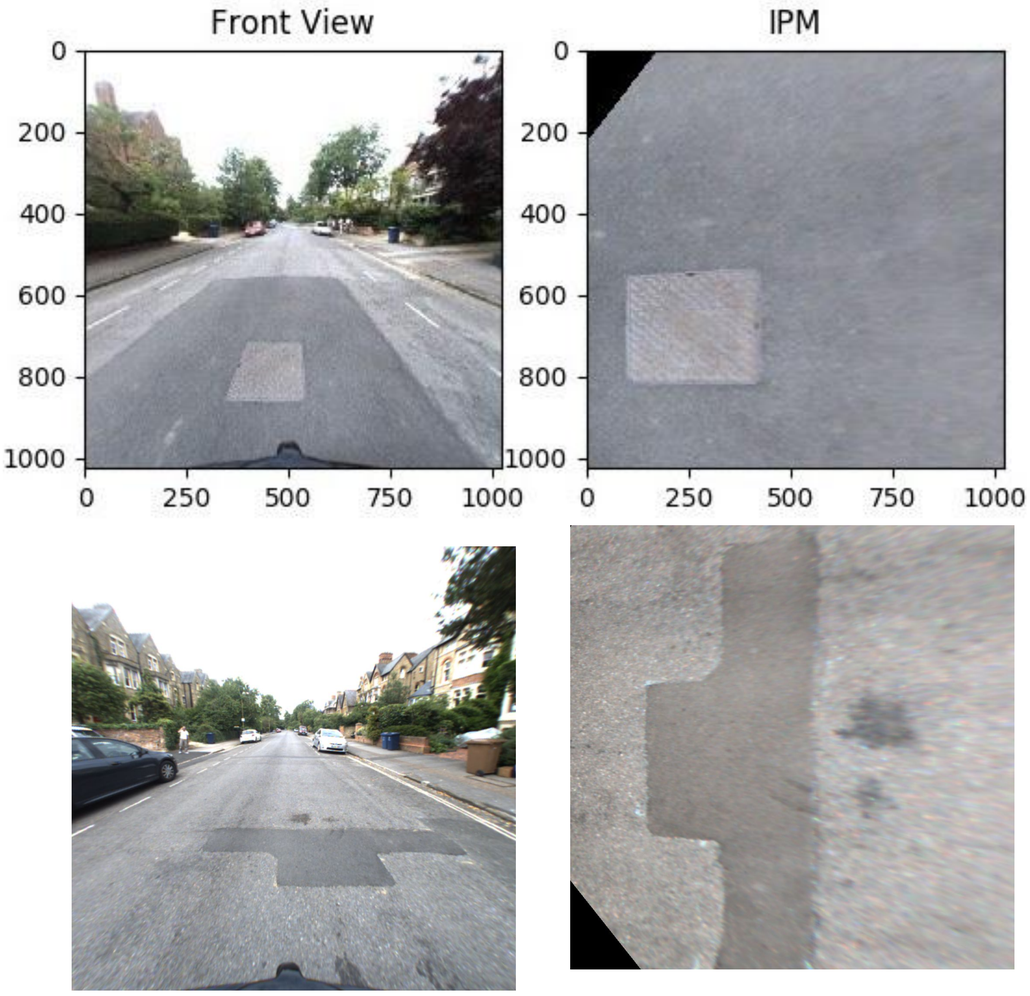
Ortographic to Perspective Matching

- Inverse homography to project correspondences from orthographic view to perspective view



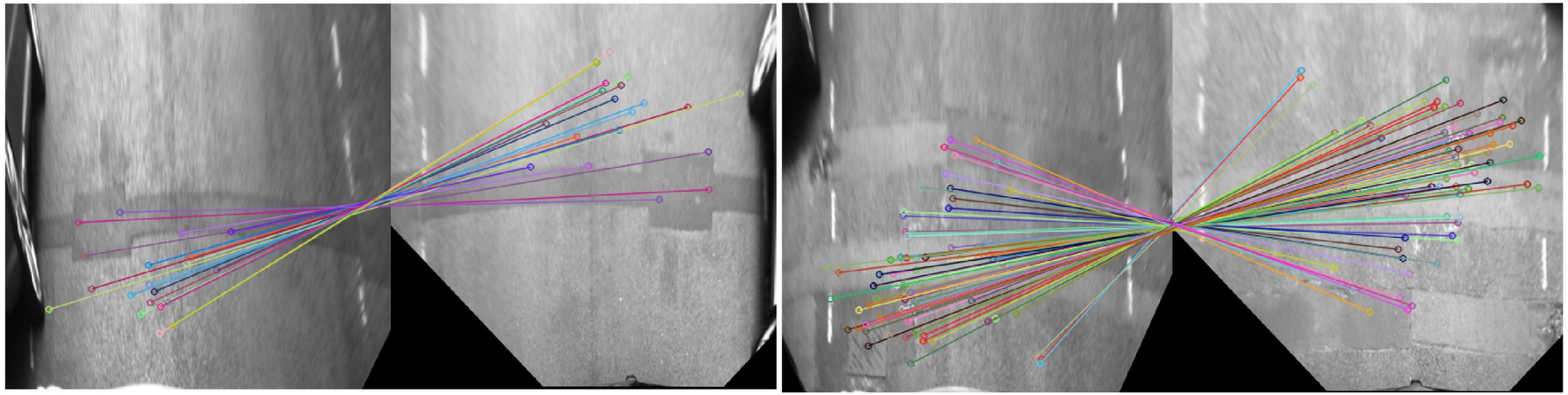
IPM for Autonomous Driving

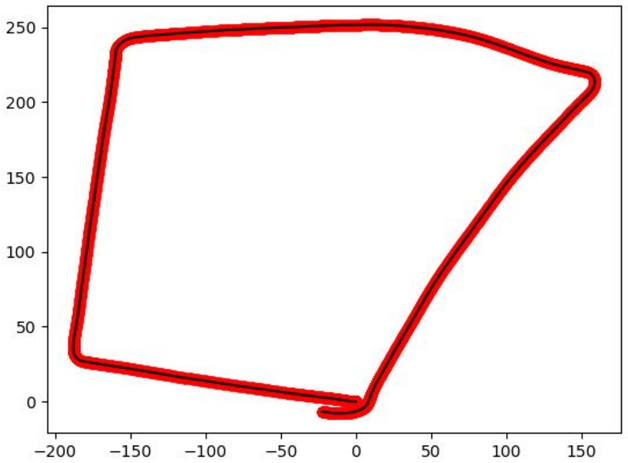
- Challenge of matching images from front and rear camera
- Discriminative road patches



VPR for Autonomous Driving

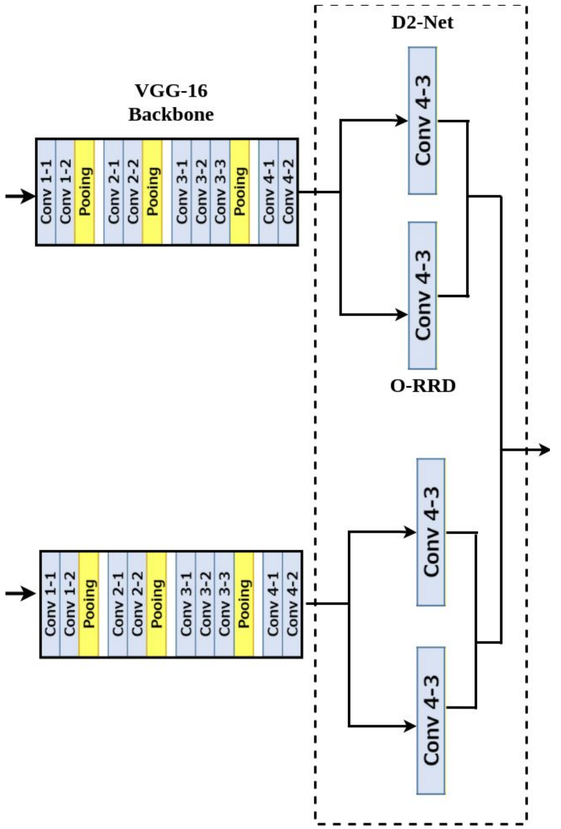
- Image pairs with maximum inliers are considered a match





Network Architecture

- VGG-16 common backbone architecture for better efficency
- Last layer is fine-tune for D2-Net and RoRd
- D2-Net head is train using SfM data while RoRD is trained using Homography data
- Feature correspondences are calculated independently for each head
- RANSAC based geometric verification for geometric verification



Loss Function

- Related via rotation homography
- Triplet margin loss, with margin
- p(c) is euclidean distance between descriptors
- Negative distance is calculated via hardest negative

$$p(c) = \|\hat{\mathbf{d}}_A^{(1)} - \hat{\mathbf{d}}_B^{(2)}\|_2$$

$$n(c) = \min\left(\|\hat{\mathbf{d}}_A^{(1)} - \hat{\mathbf{d}}_{N_2}^{(2)}\|_2, \|\hat{\mathbf{d}}_{N_1}^{(1)} - \hat{\mathbf{d}}_B^{(2)}\|_2\right)$$

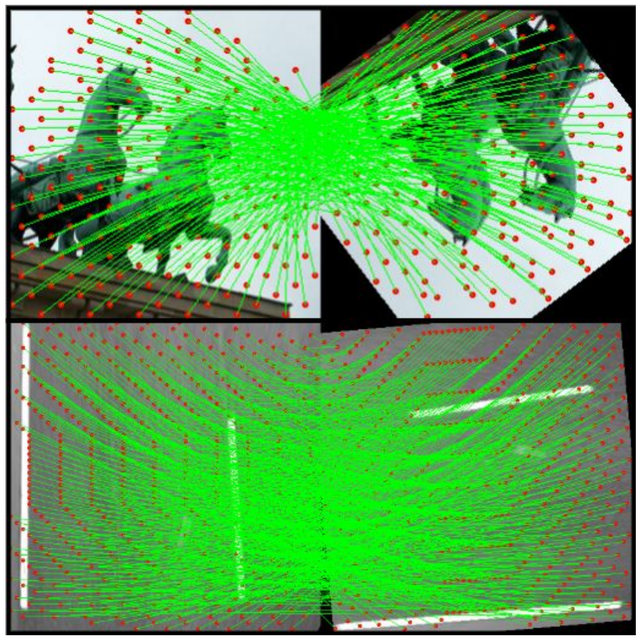
$$N_1 = \underset{P \in I_1}{\arg\min} \|\hat{\mathbf{d}}_P^{(1)} - \hat{\mathbf{d}}_B^{(2)}\|_2 \text{ s.t. } \|P - A\|_\infty > K$$

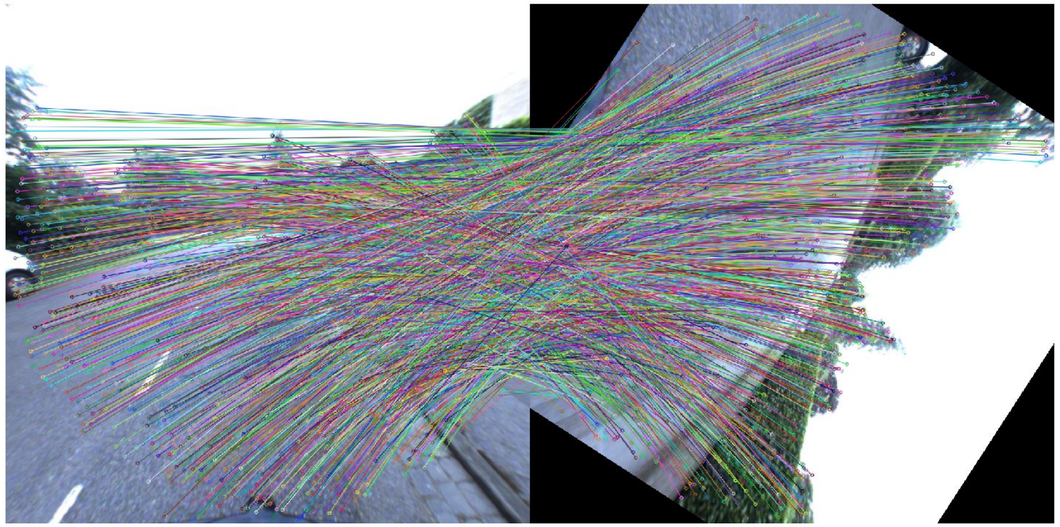
$$m(c) = \max\left(0, M + p(c)^2 - n(c)^2\right)$$

Training Data

Homographic Transformations

- Training data from phototourim and Oxford Robot Car Dataset

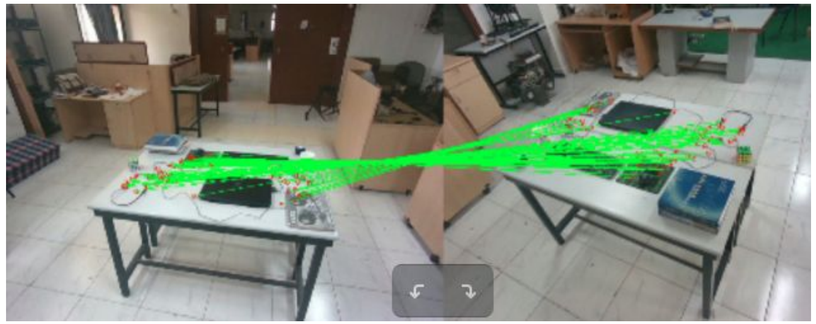




Results

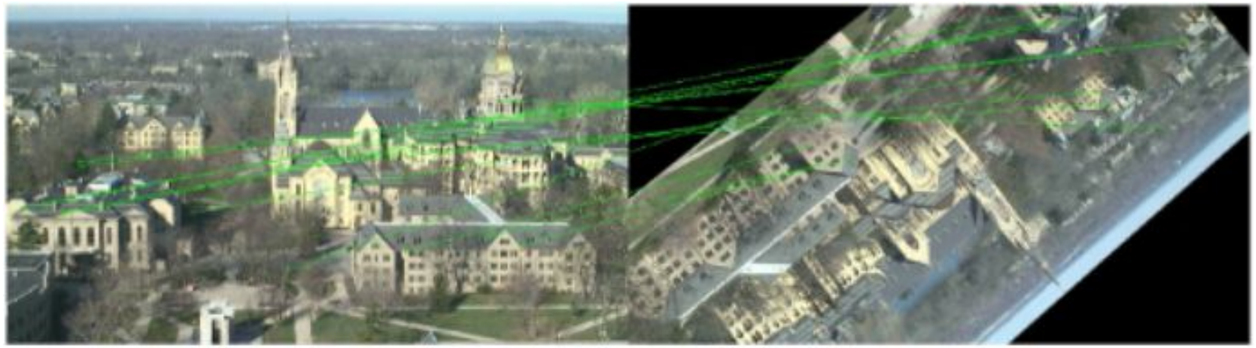
Datasets and Tasks

- MMA on Hpatches Dataset
- VPR on Oxford RobotCar Dataset
- Pose Estimation Error on DiverseView Dataset

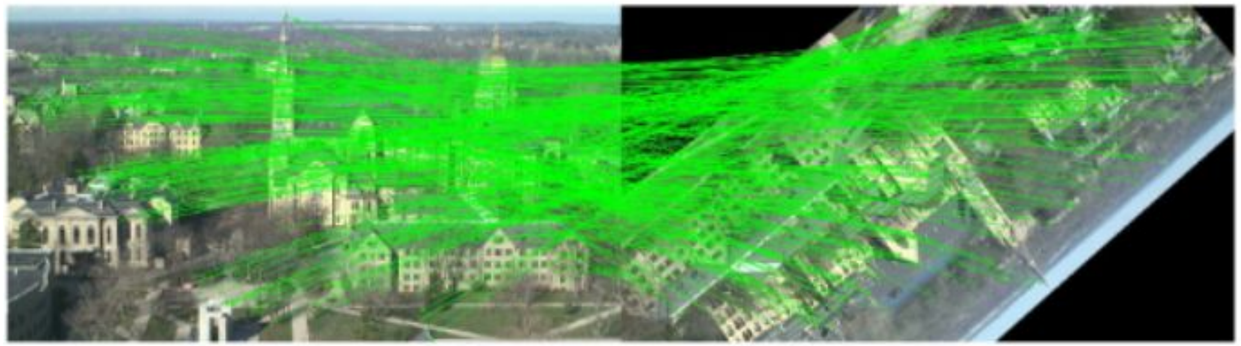


HPatches Dataset

- D2 Net



- RoRD



- Extended Hpatches Dataset to scenes having high rotation
- Evaluated for Mean matching accuracy, by using ground truth homography

Qualitative Results (MMA)

- Comparison on Standard and extended HPatches Dataset

Model	Standard	Rotated	Average
SIFT [7]	0.53/0.54/0.54	0.00/0.00/0.01	0.26/0.27/0.28
SuperPoint [3]	0.69/0.71/0.73	0.21/0.22/0.22	0.45/0.46/0.48
D2-Net [1]	**0.73/0.81/0.84**	0.17/0.20/0.22	0.45/0.50/0.53
(Ours) RoRD	0.68/0.75/0.78	**0.46/0.57/0.62**	**0.57/0.66/0.70**
(Ours) RoRD Comb.	0.71/0.78/0.81	0.44/0.54/0.59	**0.57/0.66/0.70**
(Ours) RoRD + CE	**0.79/0.84/0.86**	**0.48/0.59/0.64**	**0.64/0.72/0.75**

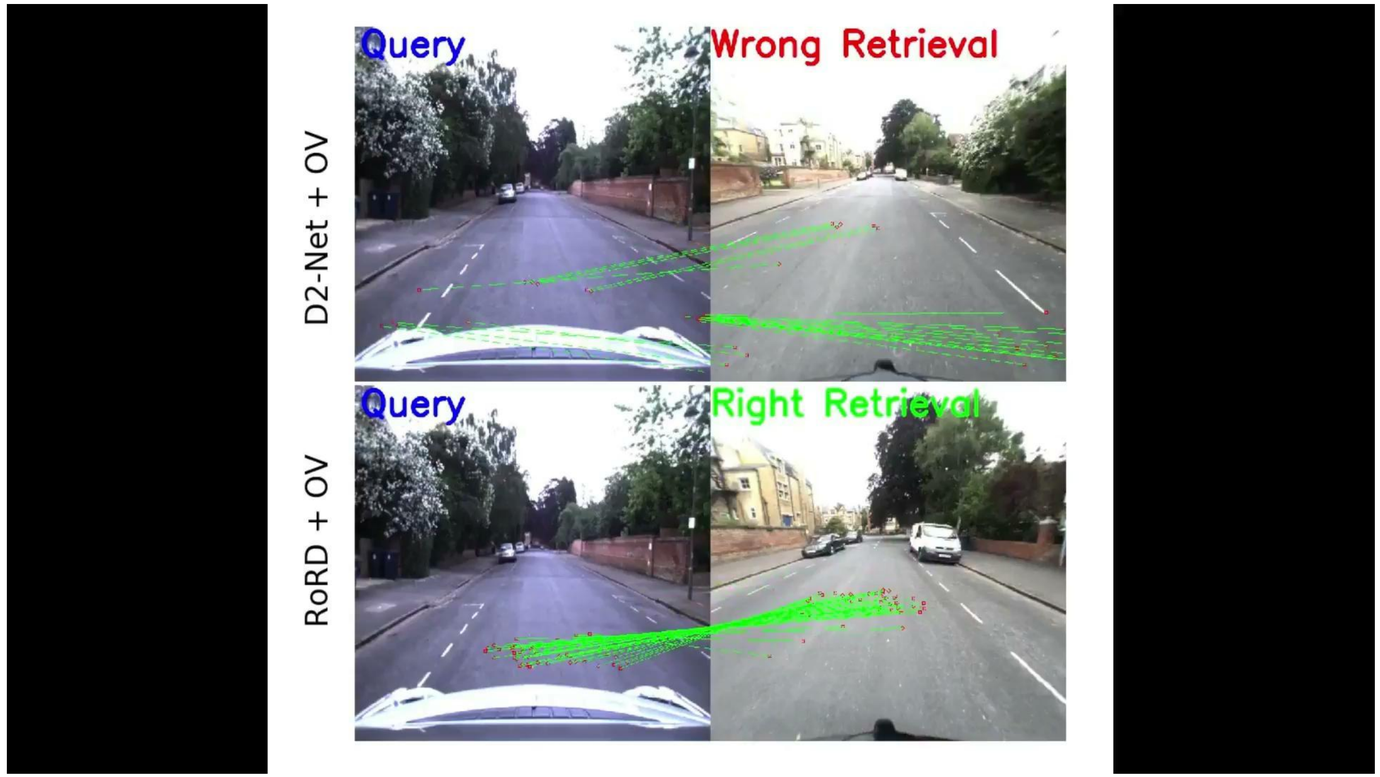
Oxford RobotCar Dataset

- D2-Net with Orthographic View

- RoRD with Orthographic View

VPR Results with Front and Rear Camera

- Testing VPR results on the sequence not seen during training



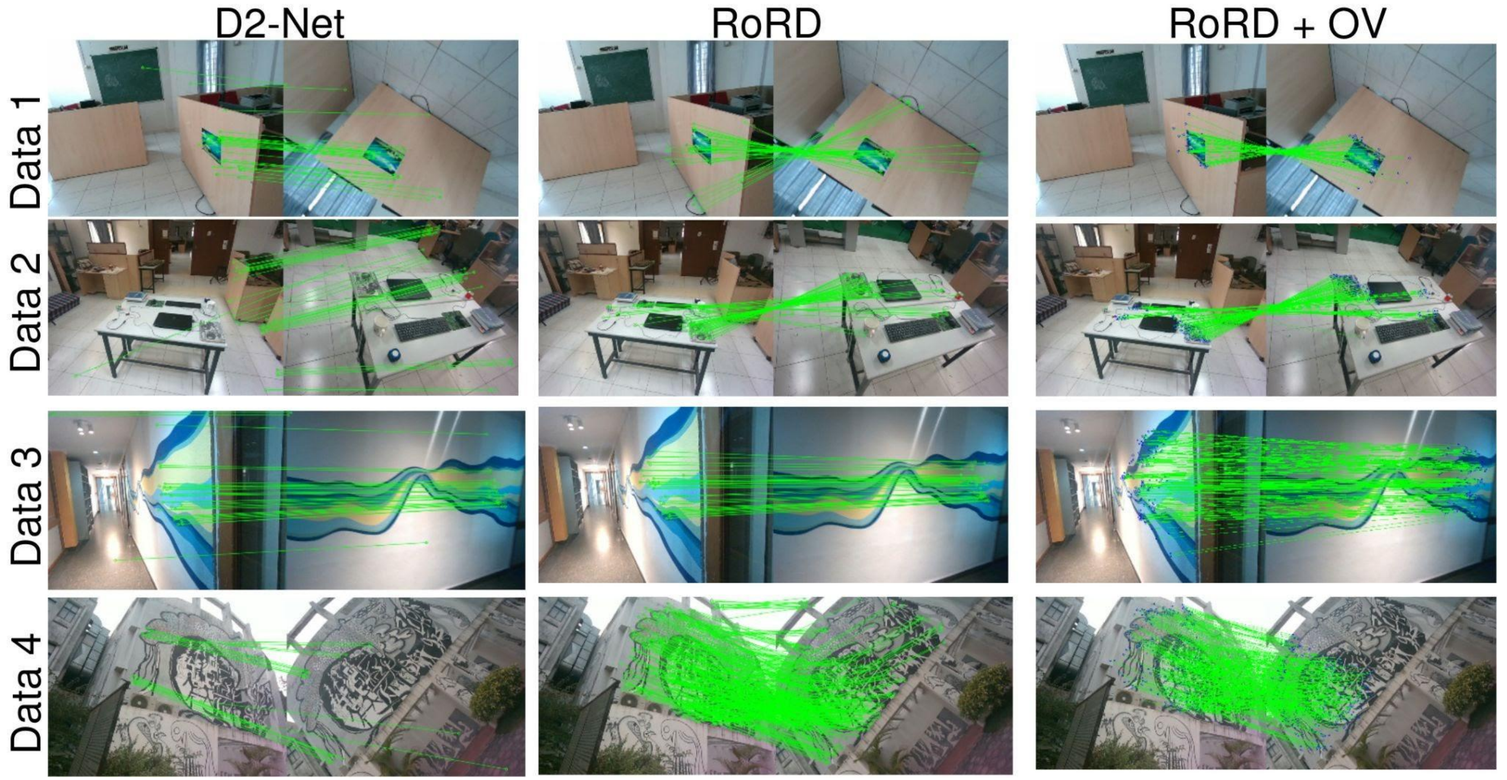
Video

Recall for VPR

- Obtain twice the recall compared to second best on challenging Oxford RobotCar Dataset

Model	Recall
LoST [12]	10.30
D2-Net [1]	22.16
SuperPoint+SuperGlue [2]	25.47
SIFT (Rectified Features) [6]	**34.97**
RoRD (ours)	**70.31**

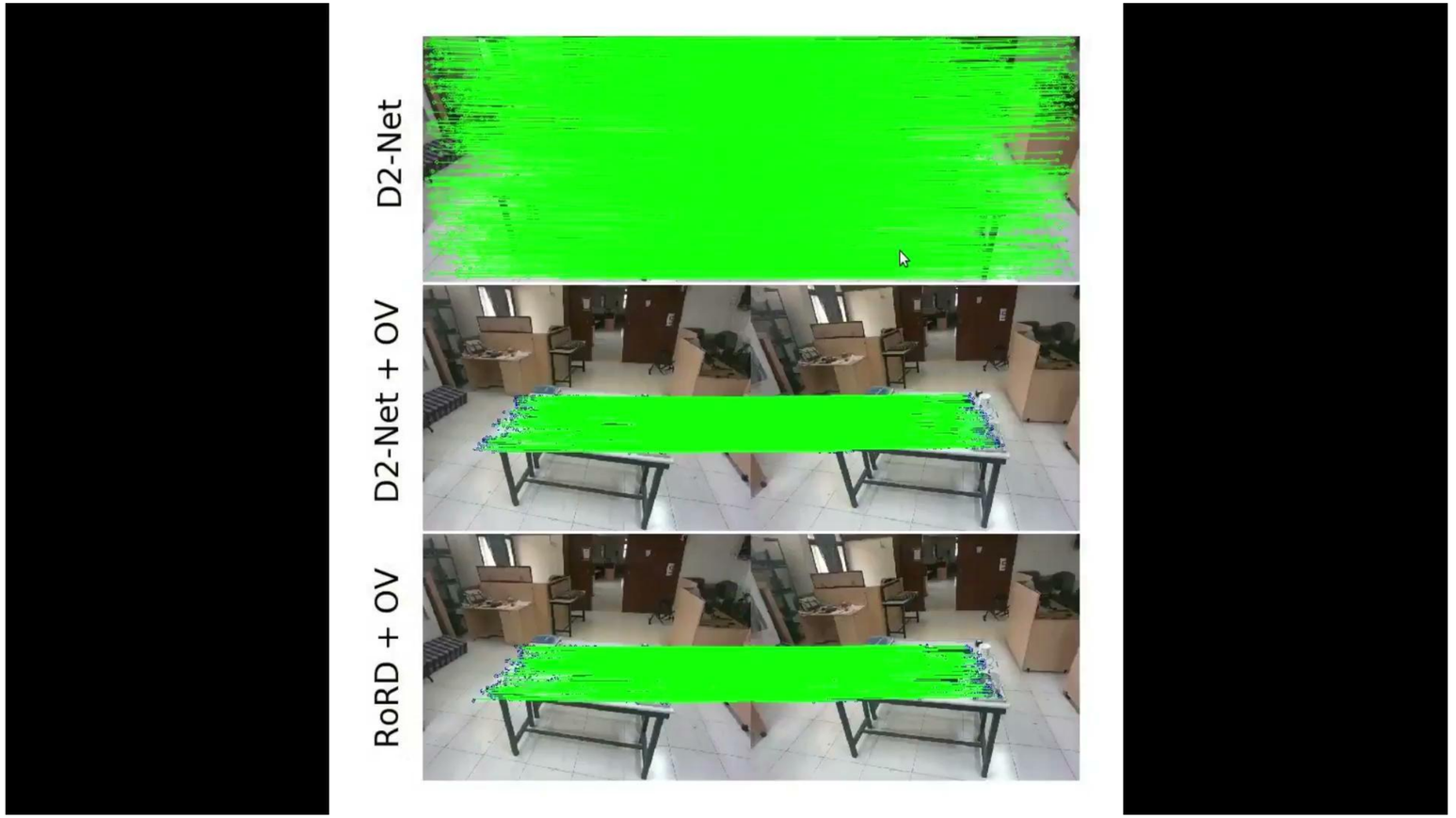
DiverseView Dataset



- Orthographic views in tandem with RoRD gives best correspondences

Pose Estimation Results

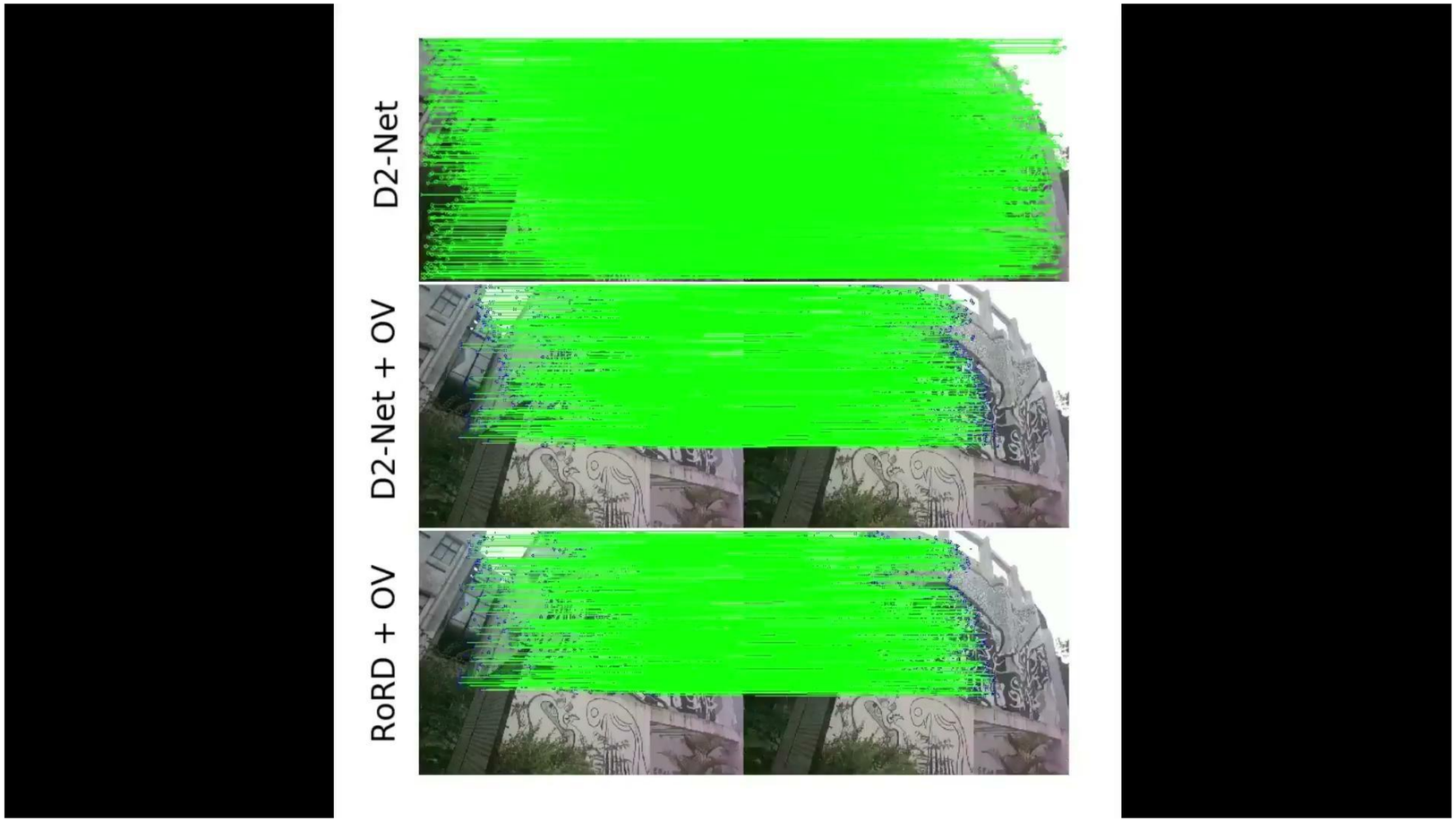
Indoor



Video

Pose Estimation Results

Outdoor



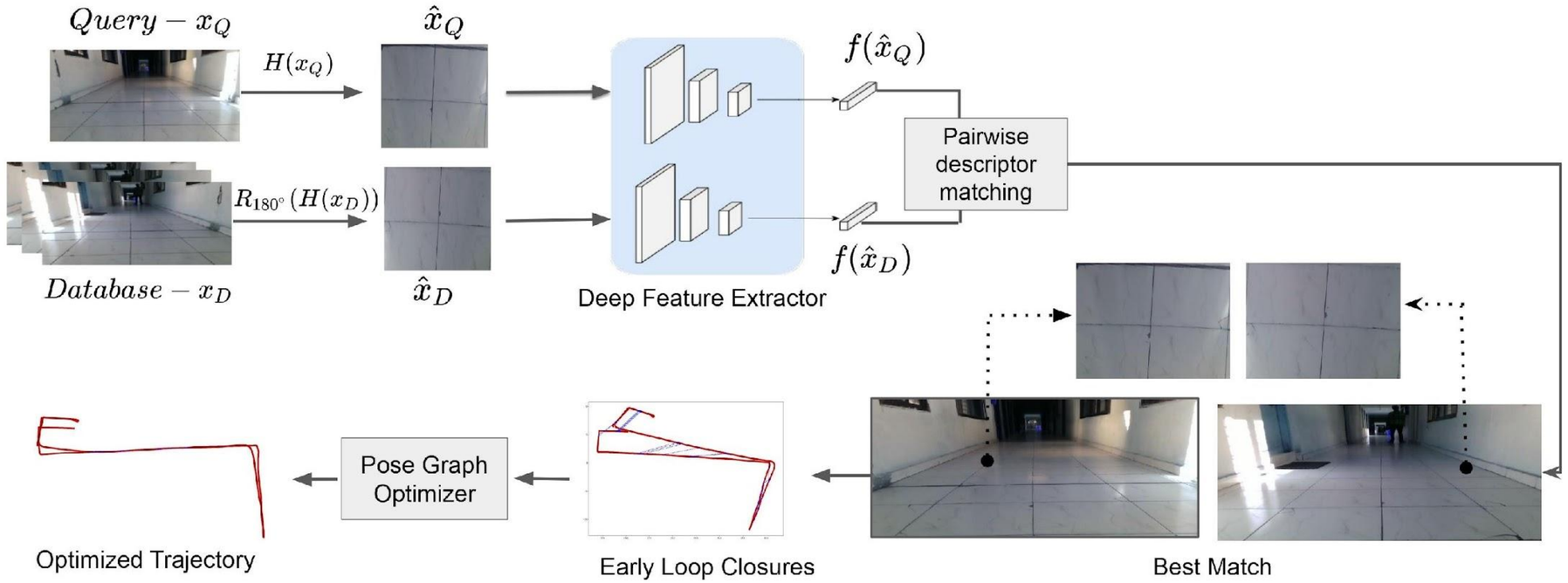
Video

Ablation Study for Pose Estimation

Model	Sequence 1	Sequence 2	Sequence 3
Perspective View			
SuperPoint+SuperGlue [2]	58.71/1.18	79.03/1.47	49.27/4.45
D2-Net [1]	68.76/1.02	80.28/1.44	48.33/4.55
SIFT [7]	48.43/1.12	**59.81**/1.30	12.74/1.47
(Ours) RoRD	**18.12/0.42**	**61.90/1.15**	**10.17/1.38**
(Ours) RoRD + CE	**20.69/0.44**	65.55/**1.21**	**10.37/1.41**
Proposed Orthographic View (OV)			
SuperPoint+SuperGlue [2]	54.73/0.80	77.85/1.13	44.61/4.37
D2-Net [1]	63.69/0.87	80.71/1.29	51.04/4.80
SIFT [7]	13.04/0.26	17.84/0.34	**5.11/0.76**
(Ours) RoRD	**7.71/0.18**	**8.58/0.20**	**7.79/1.03**
(Ours) RoRD + CE	**8.66/0.18**	**16.32/0.32**	8.33/1.07

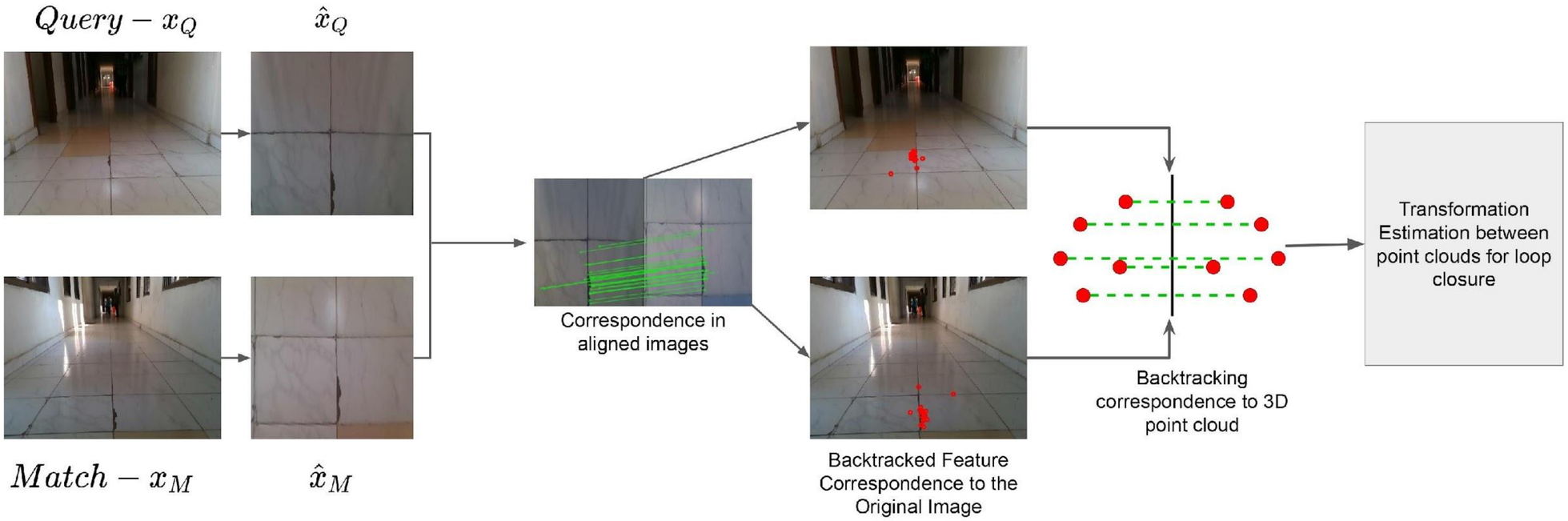
- Rotation and Translation error

Opposite View Loop Closures in SLAM



- Global feature matching for VPR and local feature matching for transformations

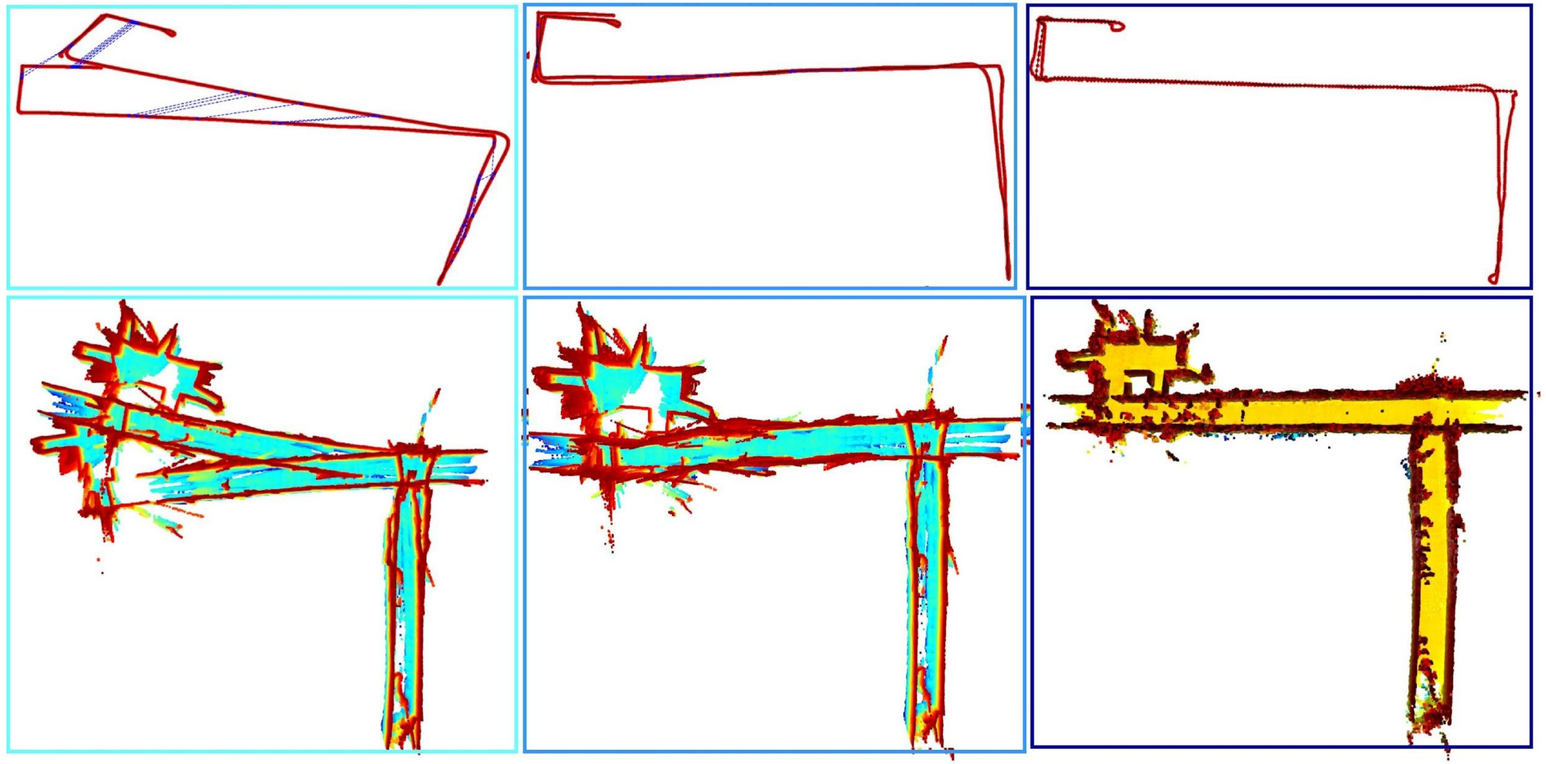
Transformation Estimation using Rotation Invariant Descriptors



Video Results of Loop Closure in Lab Dataset

Early Bird: Loop Closures from Opposing Viewpoints for Perceptually-Aliased Indoor Environments

Pose Graph Optimization on Lab Dataset



- Extended RTABMAP to work with scenes where robot revisits the scene

Code and Dataset

- Code and Dataset are publicly available
- https://github.com/UditSinghParihar/RoRD

Topological Mapping for Manhattan-like Repetitive Environments

Sai Shubodh Puligilla [*], Satyajit Tourani [*], Tushar Vaidya [*],
Udit Singh Parihar [*], Ravi Kiran Sarvadevabhatla and K. Madhava Krishna

*Denotes authors with equal contribution

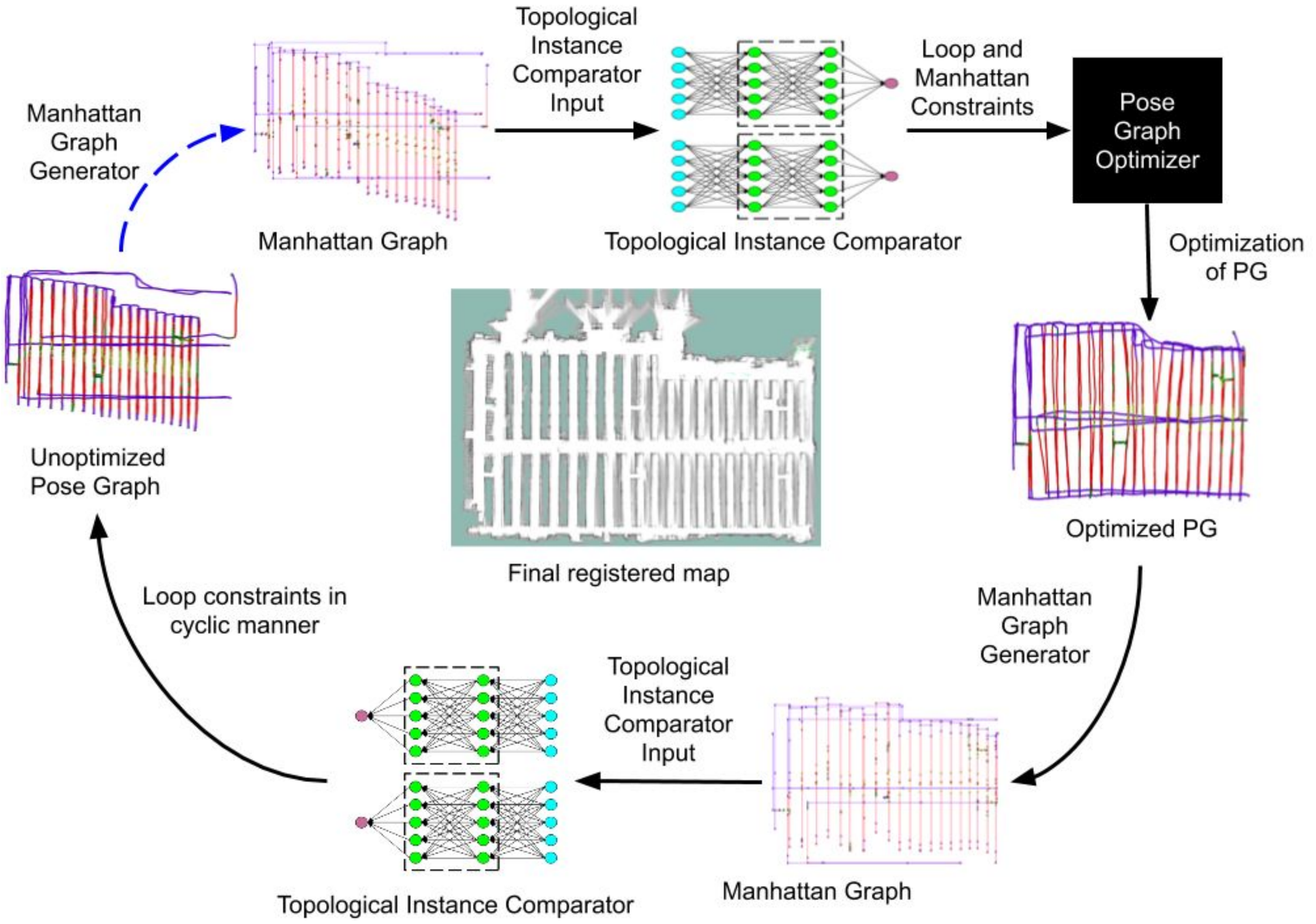
Accepted to International Conference on Robotics and Automation 2020

Problem Formulation

1. Loop closure constraints are proposed by MLP and computed using ICP.
2. Manhattan constraints are obtained from orthogonal relations between nodes in Manhattan graph.

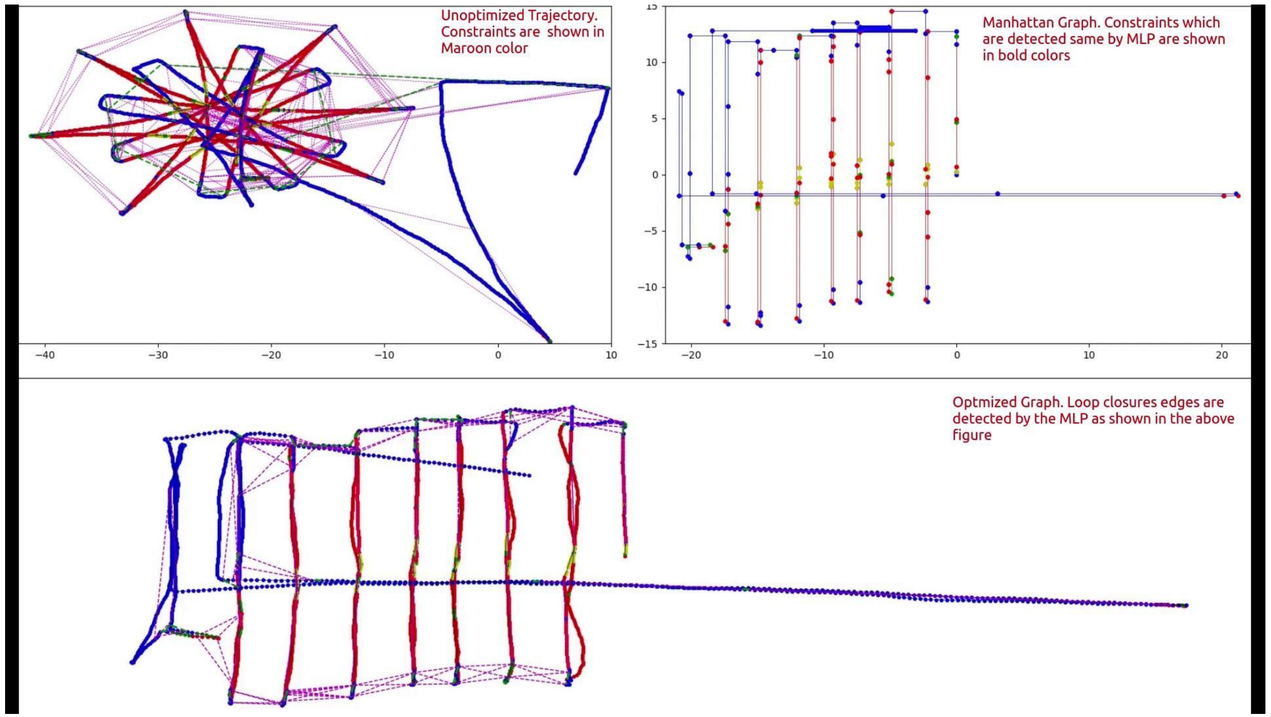
$$X^* = \underset{X}{argmax}\, P(X|U) = \underset{X}{argmax} \prod_i P(x_{i+1}|x_i, u_i) \times \underbrace{\prod_{i \in C(M_i), j \in C(M_j)} P(x_j|x_i, c_{ij})}_{Loop\ Closure\ Constraints} \times \underbrace{\prod_{i \in N(M_i), j \in N(M_j)} P(x_j|x_i, m_{ij})}_{Manhattan\ Constraints}$$

Pose Graph Optimization Pipeline



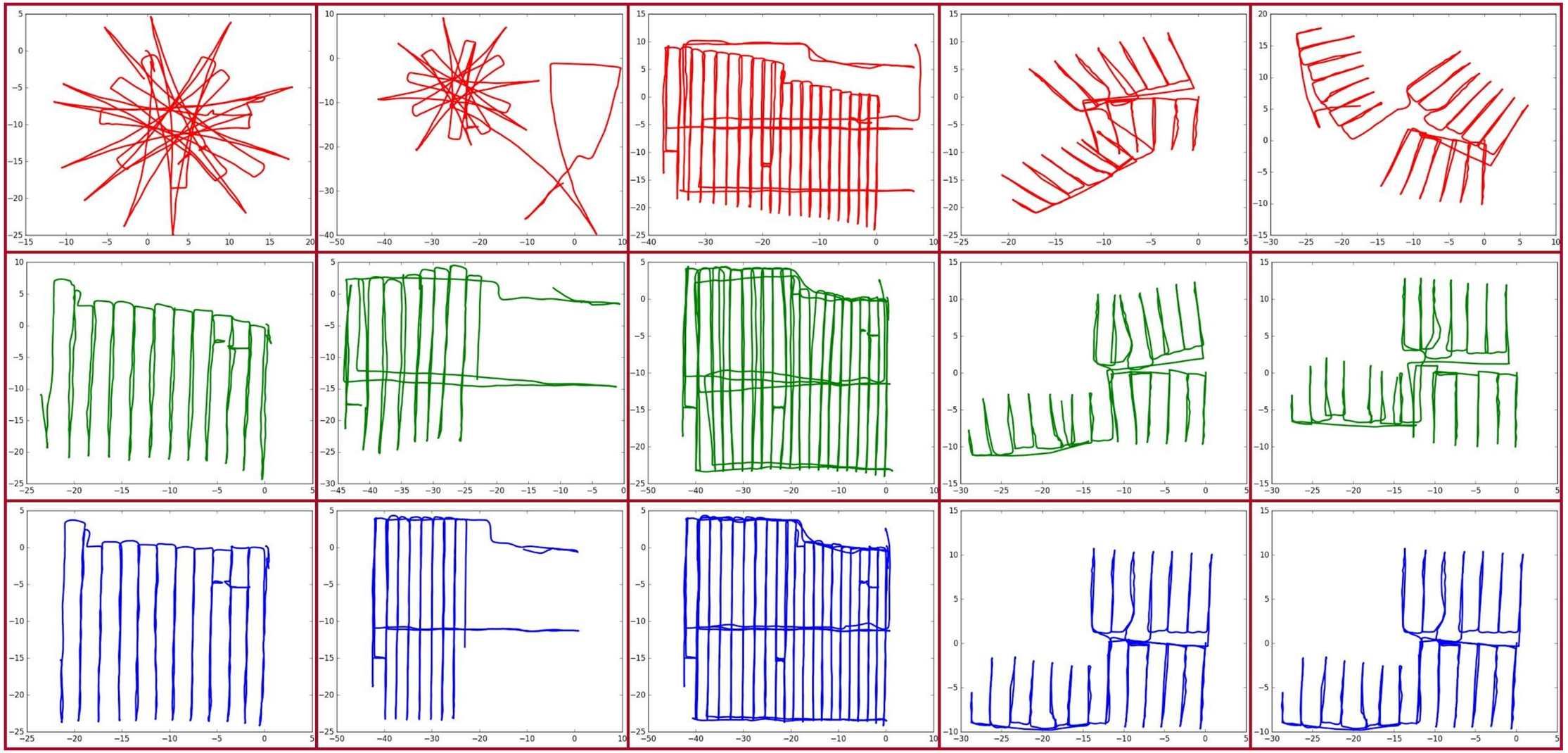
RESULTS

1. Nodes of the pose graph are labelled based on the topological labels by CNN to create a topological graph



1. A highly distorted unoptimized pose graph with labelled topologies is recovered using an intermediate Manhattan Graph.
2. A loop detection event between two Manhattan Nodes in the Manhattan Graph is shown in bold color. This results in both Manhattan and/or Loop constraints.

RECOVERED TRAJECTORIES



Top row shows unoptimized trajectories, middle row shows trajectories recovered using our pipeline and last row shows ground truth trajectories

Improving RTABMAP with Topological Mapping

Benchmarking RTABMAP

1. Evaluation of RTABMAP on warehouse dataset leads to detection of many False Positive loop closure constraints, due to repeating corridors.
2. Incorporation of Topological constraints leads to better ATE.
3. Our topological comparator utilizes geometric structure of the topological representation.

RTABMAP	RTABMAP + Topological Constraints
4.45	3.36

References

- RoRD - Rotation Robust Descriptors and Orthographic Views for Local Feature Matching, IEEE IROS 2021
- Early Bird : Loop Closures from Opposing Viewpoints for Perceptually-Aliased Indoor Environments, VISAPP 2020
- Topological Mapping for Manhattan-like Repetitive Environments, IEEE ICRA 2020

Thanks