DATI, AI E ROBOTICA @POLITO

RICERCA, TRASFERIMENTO TECNOLOGICO E SUPPORTO ALLE AZIENDE SUI TEMI FONDAMENTALI DEI BIG DATA, INTELLIGENZA ARTIFICIALE, ROBOTICA E RIVOLUZIONE DIGITALE

VIEWPOINT INVARIANT DENSE MATCHING FOR VISUAL GEOLOCALIZATION

Authors: Gabriele Berton, Carlo Masone, Valerio Paolicelli, Barbara Caputo

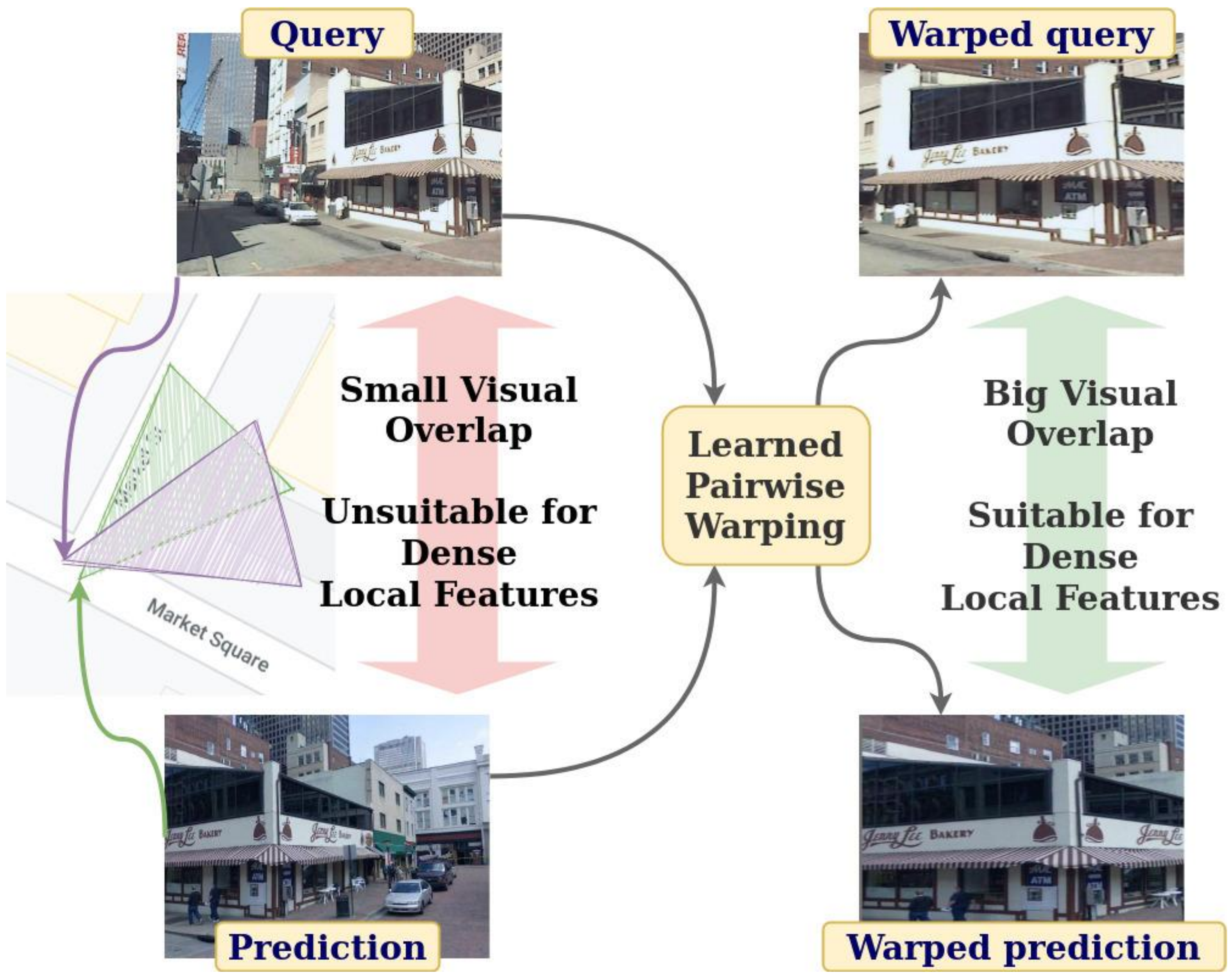
VISUAL GEOLOCALIZATION

- **Task:** recognize where a given photo was taken;
- **Common approach:** cast the problem as image retrieval

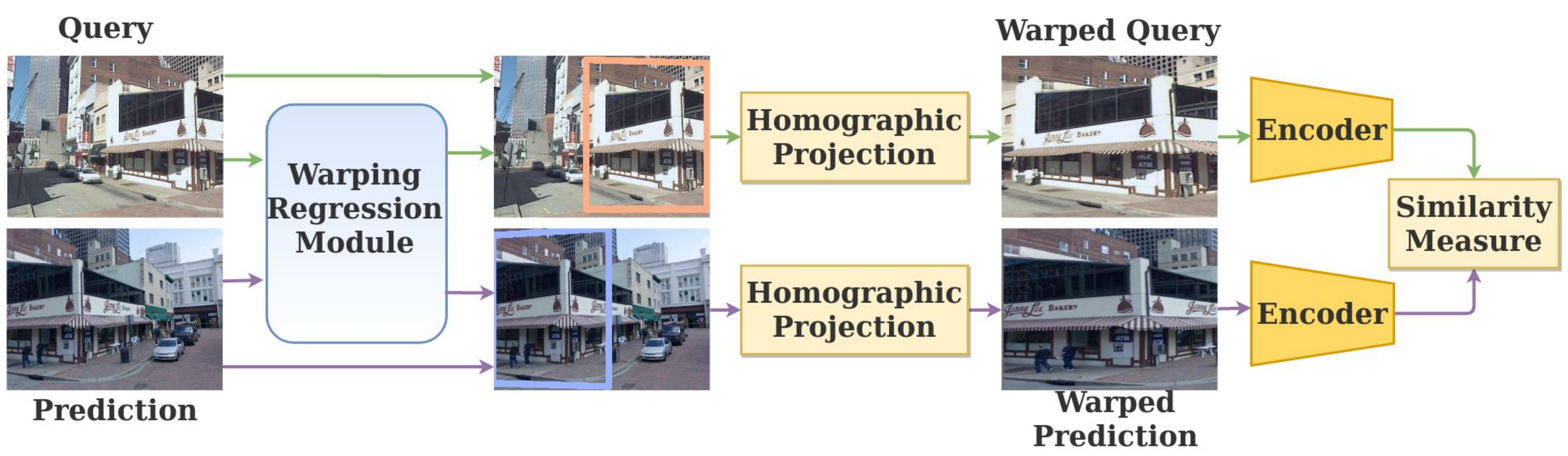
MOTIVATION AND NETWORK ARCHITECTURE

Task approached using global or local features:
- global features are compact, but lack robustness to clutter;
- sparse local features may fail with illumination or seasonal changes;
- dense local features are not robust to viewpoint shifts.

We use dense local features after warping the query-prediction pair, to achieve viewpoint invariancy.

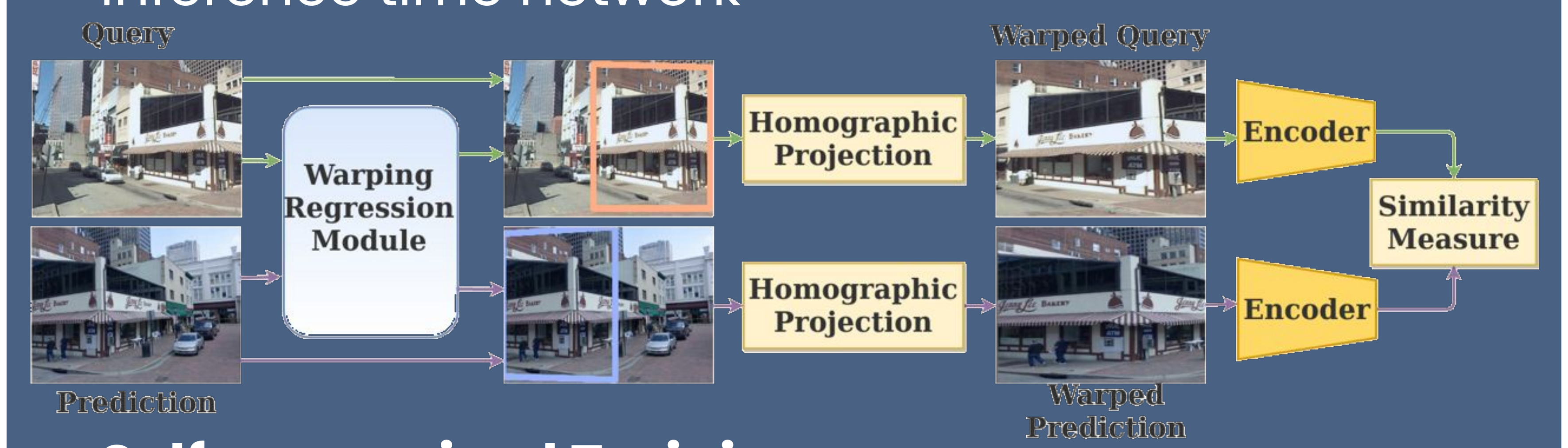


Architecture



TEST AND TRAIN TIME

Inference time network



Self-supervised Training

Generate training data from a single image:
- given an image, draw two intersecting trapezoids;
- apply homography and extract two images;
- the intersection is known!
- given the two images, train a CNN to estimate their intersection.

QUANTITATIVE AND QUALITATIVE RESULTS

Backbone	Method	Pitts30k 10m	Pitts30k 25m	Pitts30k 50m	R-Tokyo 10m	R-Tokyo 25m	R-Tokyo 50m
VGG16	GeM	55.4	70.6	76.3	33.7	40.4	45.8
VGG16	GeM + GeoWarp **(ours)**	**65.3**	**79.2**	**83.1**	**49.2**	**55.9**	**58.6**
VGG16	NetVLAD	67.2	82.5	86.5	50.2	56.9	59.9
VGG16	NetVLAD + GeoWarp **(ours)**	**70.2**	**83.3**	**86.7**	**54.5**	**61.6**	**63.6**
ResNet-50	GeM	68.3	81.4	84.4	36.0	41.8	45.5
ResNet-50	GeM + GeoWarp **(ours)**	**70.4**	**82.7**	**85.6**	**44.1**	**49.5**	**51.9**
ResNet-50	NetVLAD	70.2	84.3	87.3	65.0	72.4	74.4
ResNet-50	NetVLAD + GeoWarp **(ours)**	**72.1**	**84.8**	**87.8**	**69.7**	**74.4**	**75.4**

Backbone	Method	Pitts30k 10m	Pitts30k 25m	Pitts30k 50m	R-Tokyo 10m	R-Tokyo 25m	R-Tokyo 50m
-	DenseVLAD	63.6	77.3	81.6	35.4	39.4	40.1
-	SuperGlue	72.0	**84.9**	**88.1**	65.3	73.1	74.7
ResNet-50	NetVLAD	70.2	84.3	87.3	65.0	72.4	74.4
ResNet-50	NetVLAD + QE + DBA	68.6	83.7	87.1	66.7	72.1	74.1
ResNet-50	NetVLAD + diffusion	67.6	81.6	85.1	62.3	69.0	72.4
ResNet-50	DELG	65.4	83.0	88.0	60.6	73.0	75.1
ResNet-50	NetVLAD + GeoWarp **(ours)**	**72.1**	84.8	87.8	**69.7**	**74.4**	**75.4**

CONCLUSIONS

Warping query-predictions pairs and using dense local features presents a reliable way to be resistant to viewpoint shifts, as well as temporal variations.