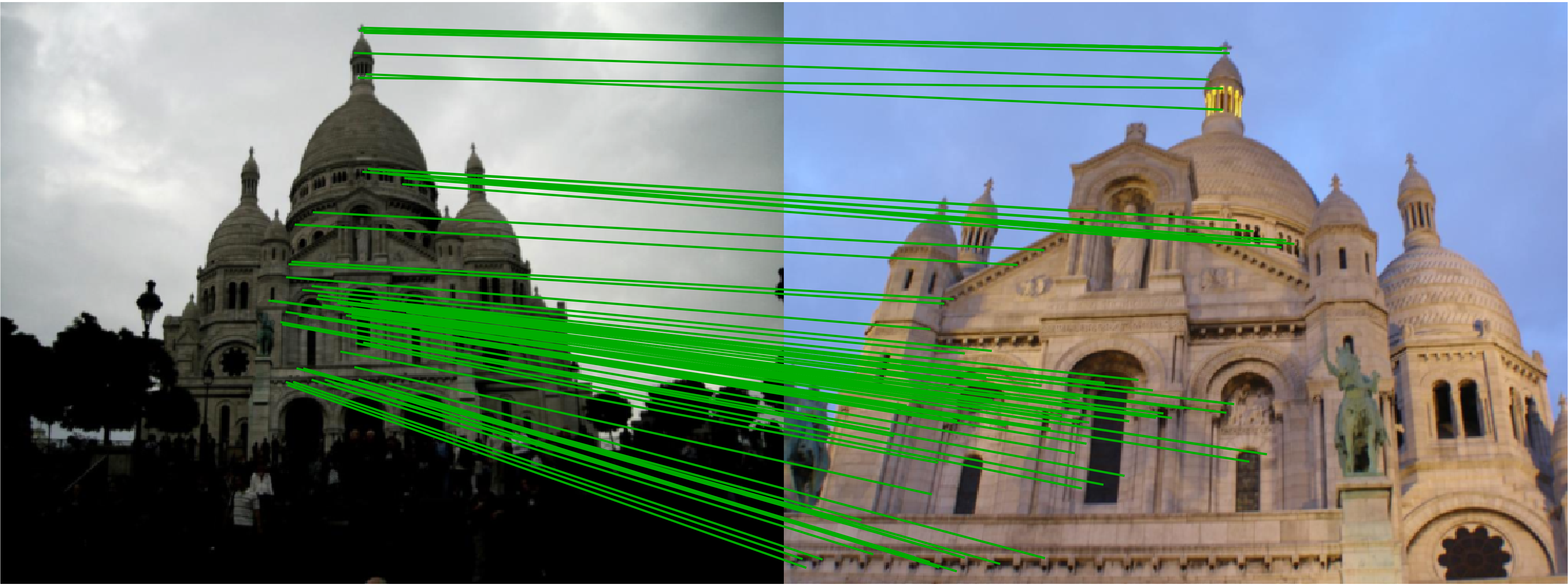


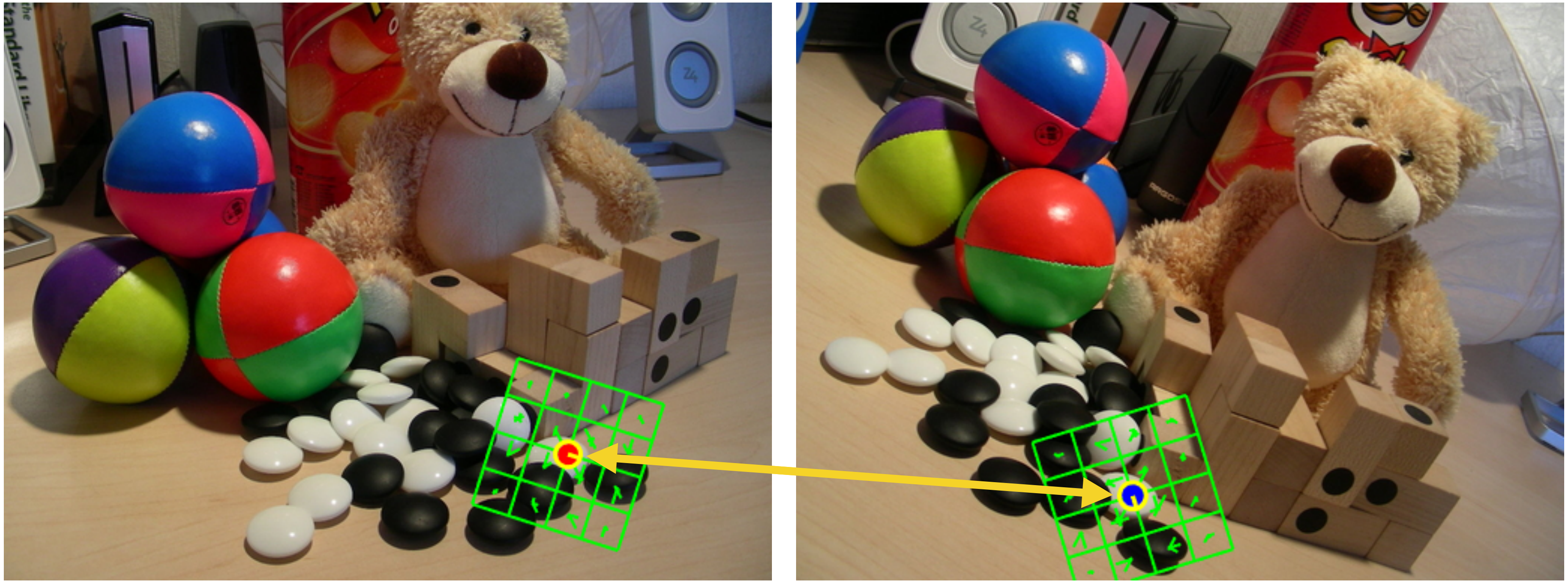
Learning to find good correspondences

K.M. Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, P. Fua

CVPR 2018 (Salt Lake City, UT, USA)

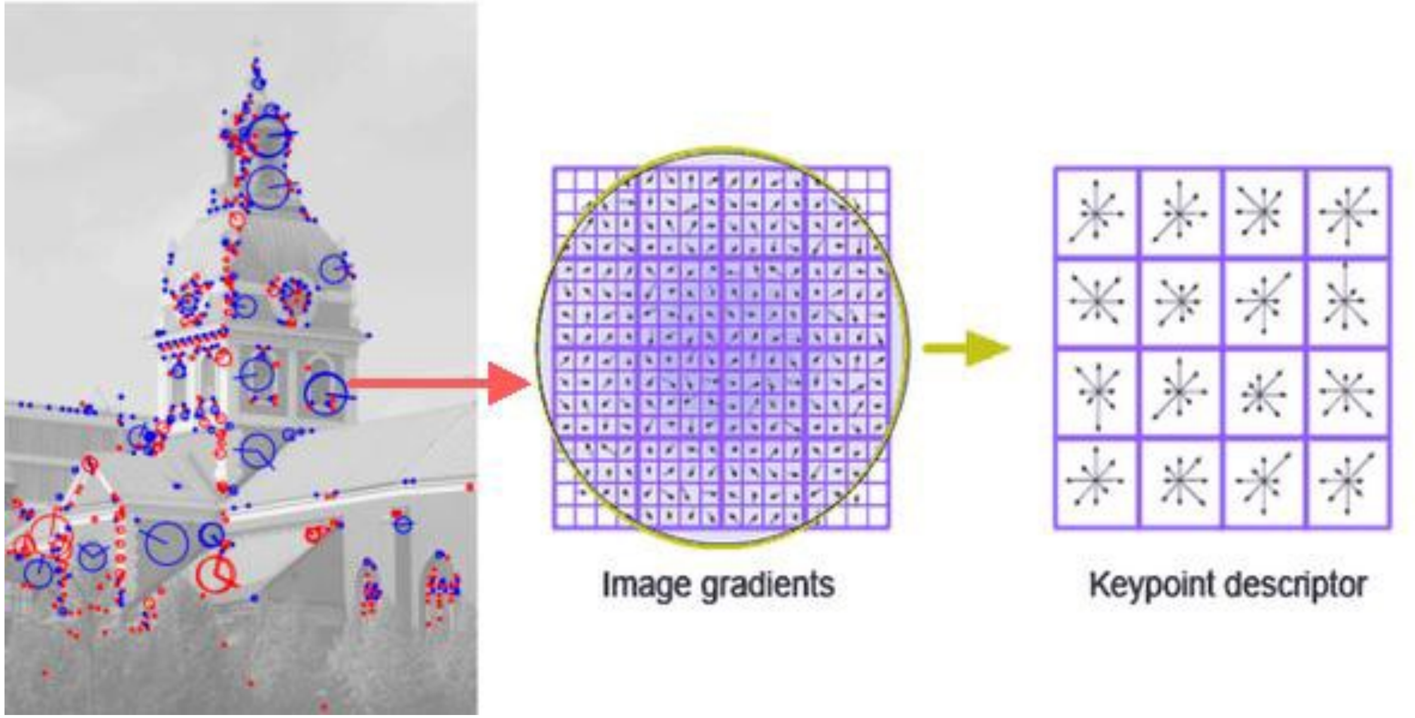
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
UVIC
TU Graz
SONY

Local features **matter**



Keypoints provide us with a robust way to **match points across images.**

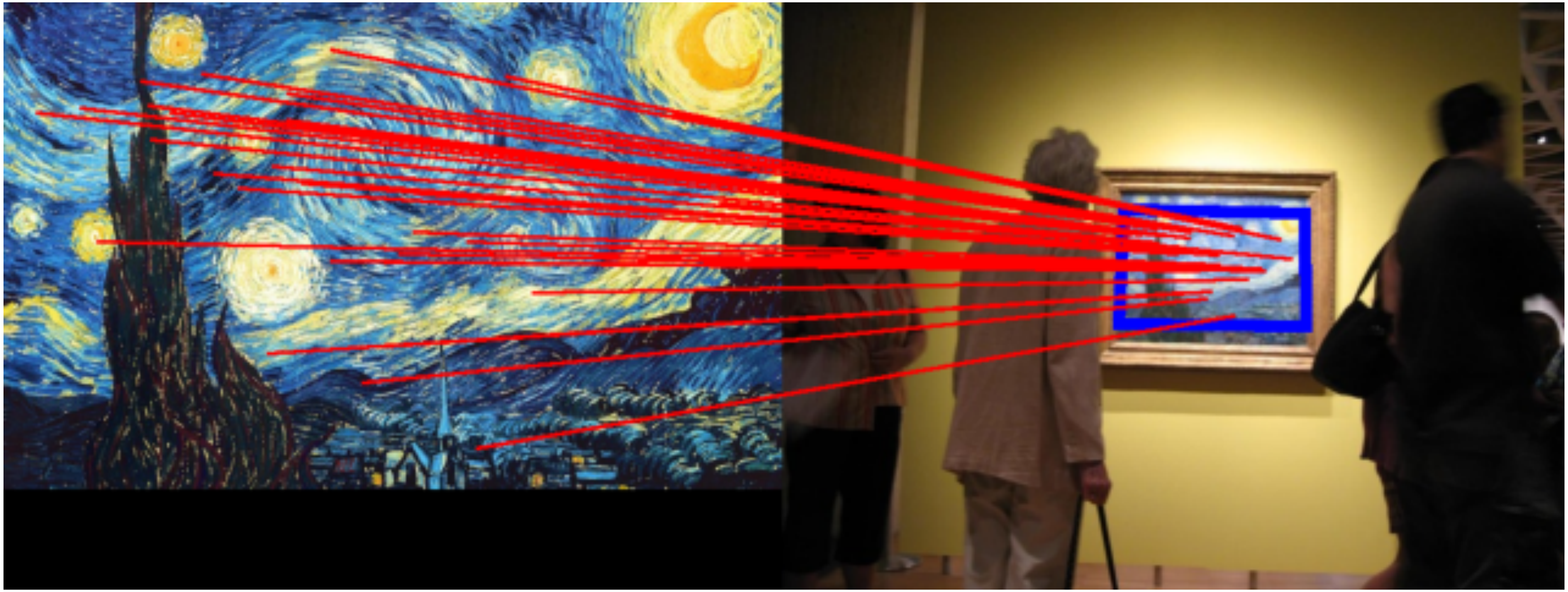
Local features **matter**



Keypoints: location (x, y), orientation, scale.
Descriptors: histograms of gradient orientations.

Source: http://medium.com/machine-learning-world/feature-extraction-and-similar-image-search-with-opencv-for-newbies-3c59796bf774

Local features **matter**



Matching large numbers of local features allows us to recover structure!

Source: OpenIMAJ (http://openimaj.org)

Why should I care?

David Lowe

Computer Science Dept., University of British Columbia
Verified email at cs.ubc.ca - Homepage

Computer Vision Object Recognition

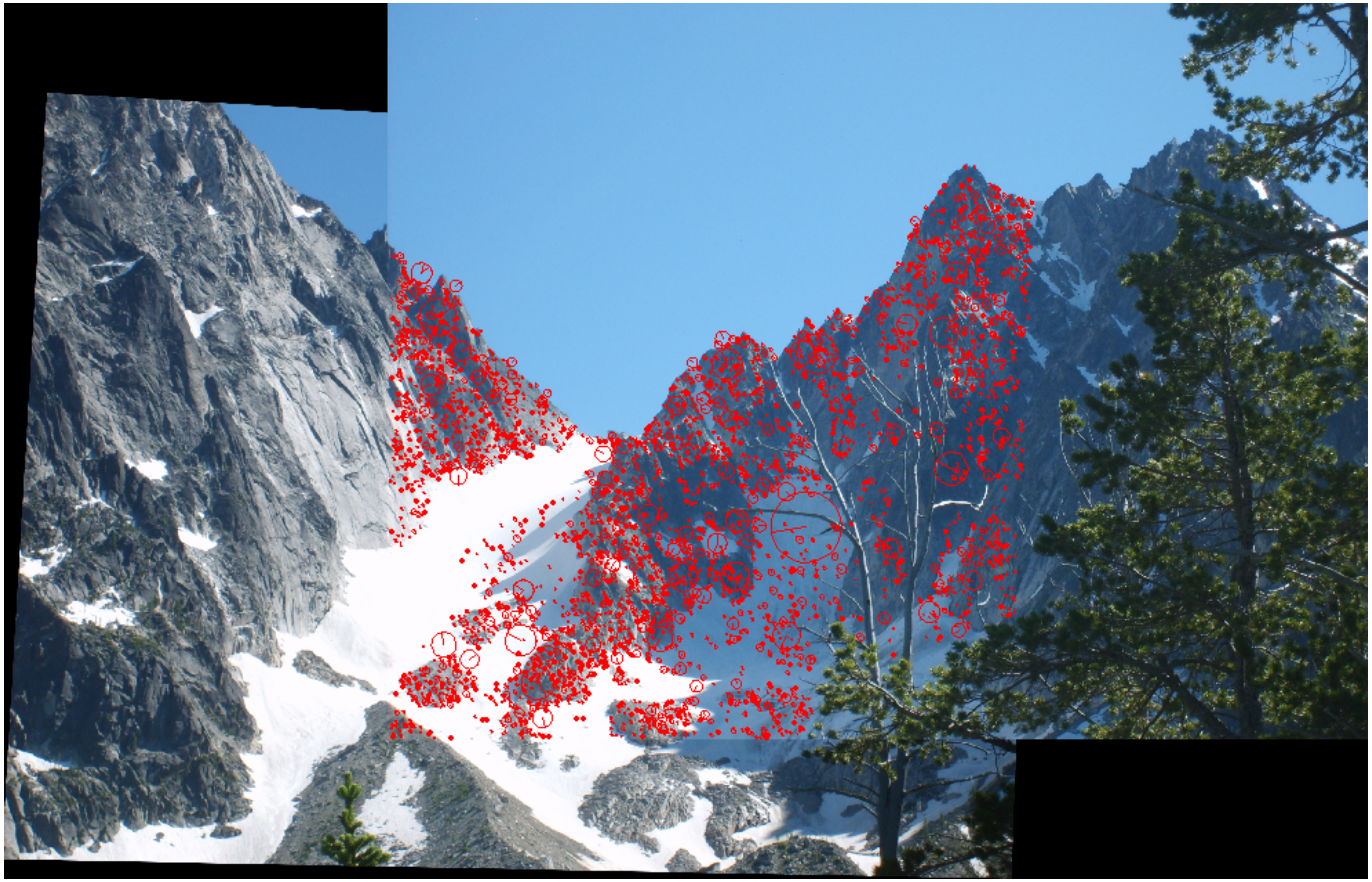
FOLLOW

SIFT

TITLE	CITED BY	YEAR
(journal paper) Distinctive image features from scale-invariant keypoints DG Lowe International journal of computer vision 60 (2), 91-110	48135	2004
(conference paper) Object recognition from local scale-invariant features DG Lowe International Conference on Computer Vision, 1999, 1150-1157	15815	1999
Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. M Muja, DG Lowe VISAPP (1) 2, 331-340	2463	2009
Automatic panoramic image stitching using invariant features M Brown, DG Lowe International Journal of Computer Vision 74 (1), 59-73	2015	2007
Perceptual Organization and Visual Recognition DG Lowe Kluwer Academic Publishers, Boston	1811 *	1985

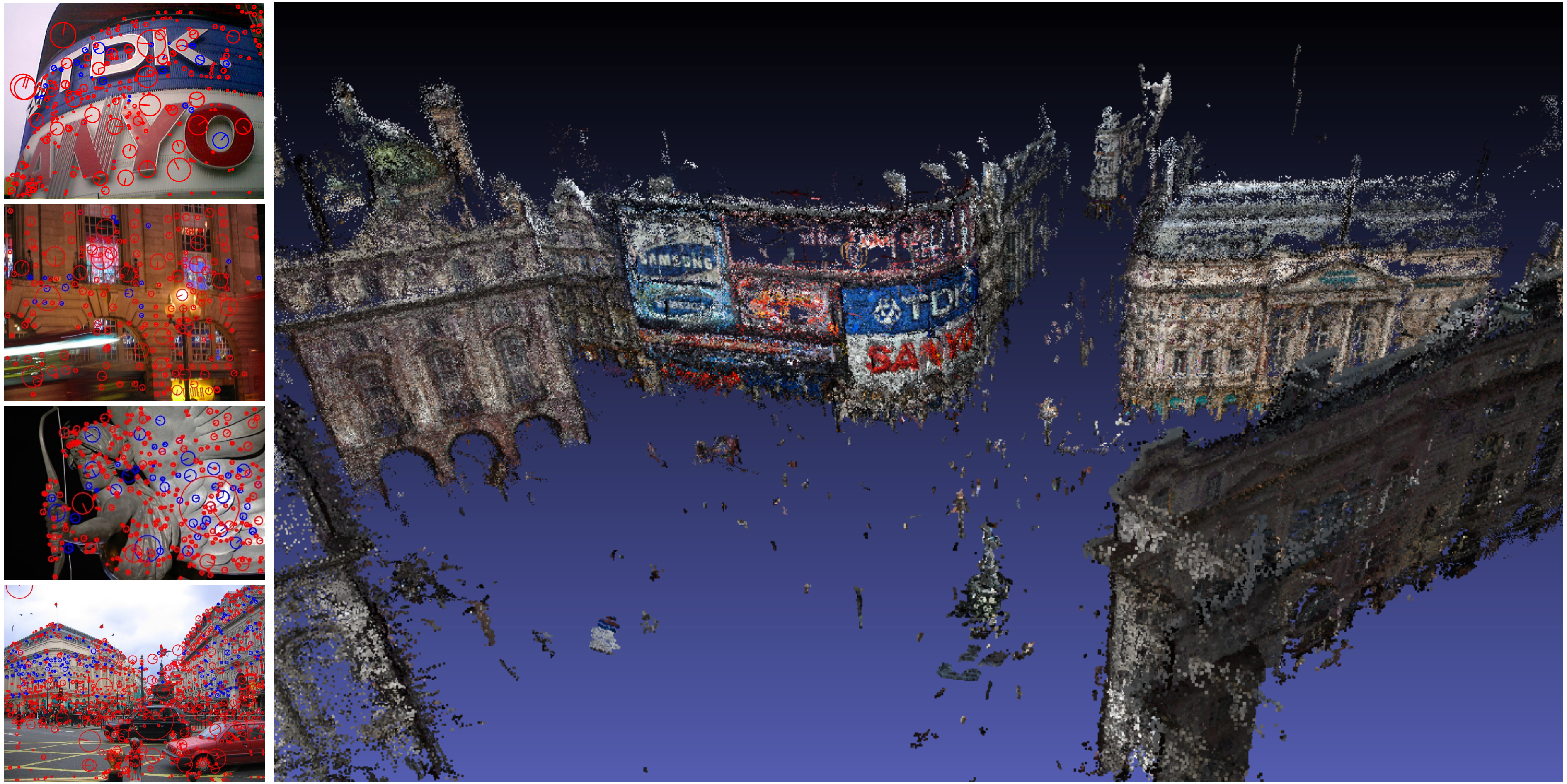
64k citations!

Applications: panorama stitching

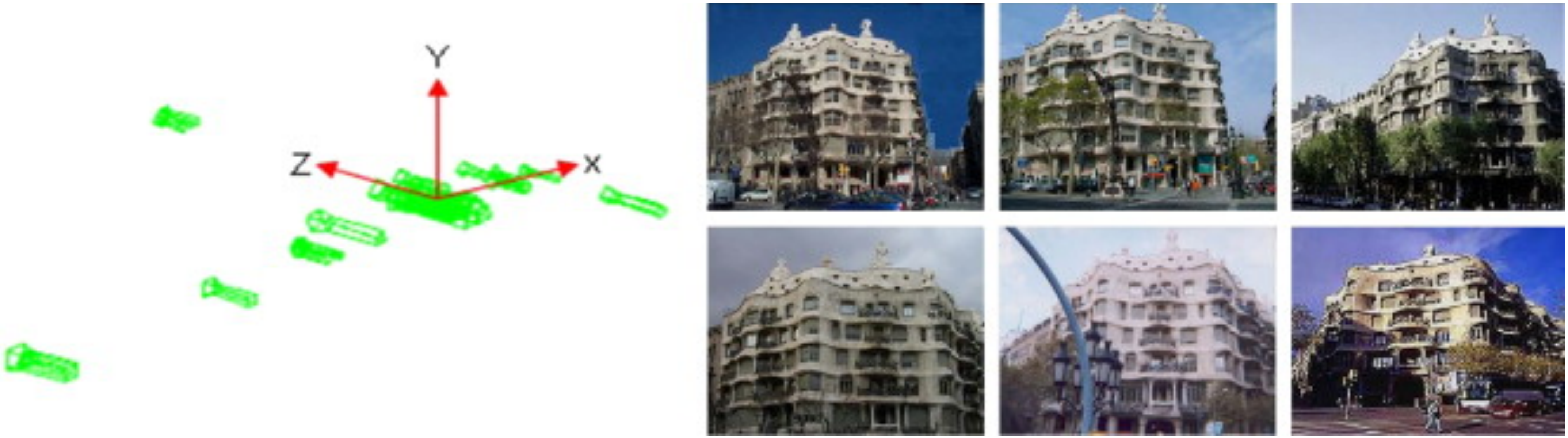


Source: http://karantza.org/wordpress/?p=10

Applications: 3D reconstruction



Applications: camera pose retrieval



Source: S.M. Yoon et al, Hierarchical image representation using 3D camera geometry for content-based image retrieval, EAAI 2014.

What about Deep Learning?

Recent works by the Computer Vision lab at EPFL:

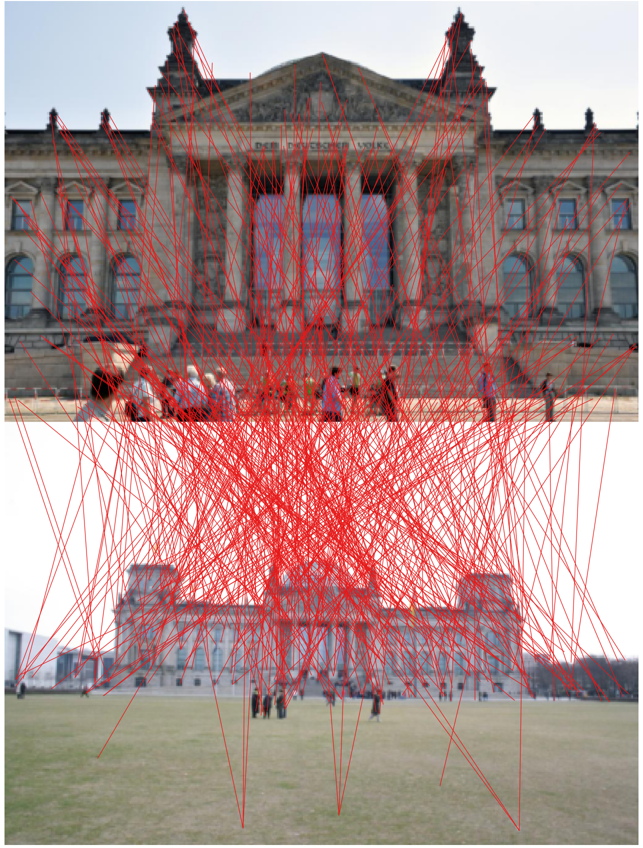
- TILDE: A Temporally Invariant Learned DEtector (CVPR'15).
- Discriminative learning of descriptors (ICCV'15).
- Learning to assign orientations to feature points (CVPR'16).
- LIFT: Learned Invariant Feature Transform (ECCV'16).
- Learning to find good correspondences (CVPR'18).
- LF-Net: Learning local features from images (arxiv'18).

What about Deep Learning?

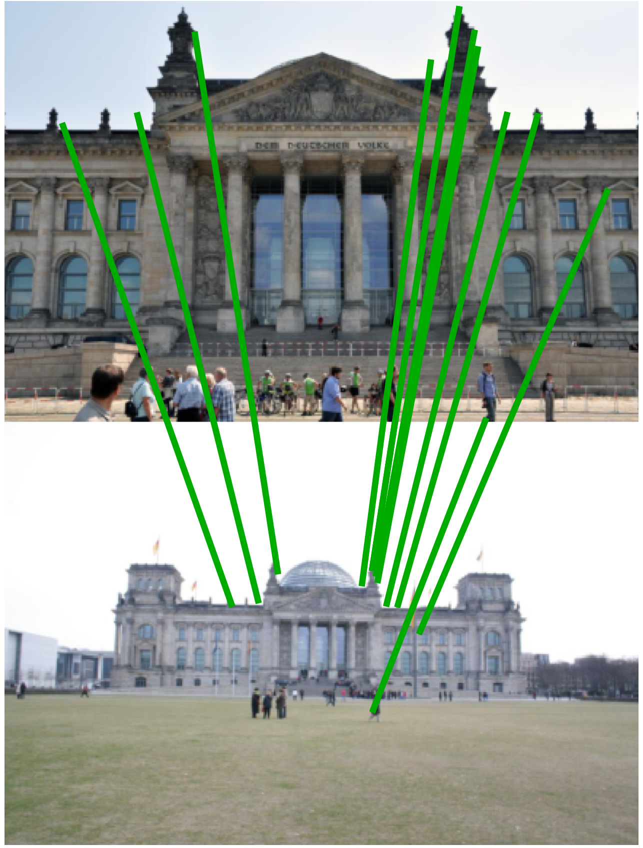
Recent works by the Computer Vision lab at EPFL:

- TILDE: A Temporally Invariant Learned DEtector (CVPR'15).
- Discriminative learning of descriptors (ICCV'15).
- Learning to assign orientations to feature points (CVPR'16).
- LIFT: Learned Invariant Feature Transform (ECCV'16).
- **Learning to find good correspondences (CVPR'18).**
- LF-Net: Learning local features from images (arxiv'18).

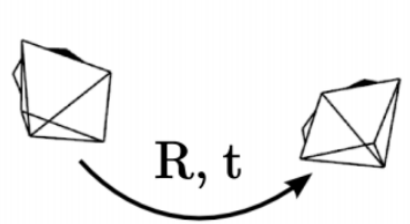
Matching keypoints



(a) Find putative matches



(b) Find inliers (e.g. RANSAC)

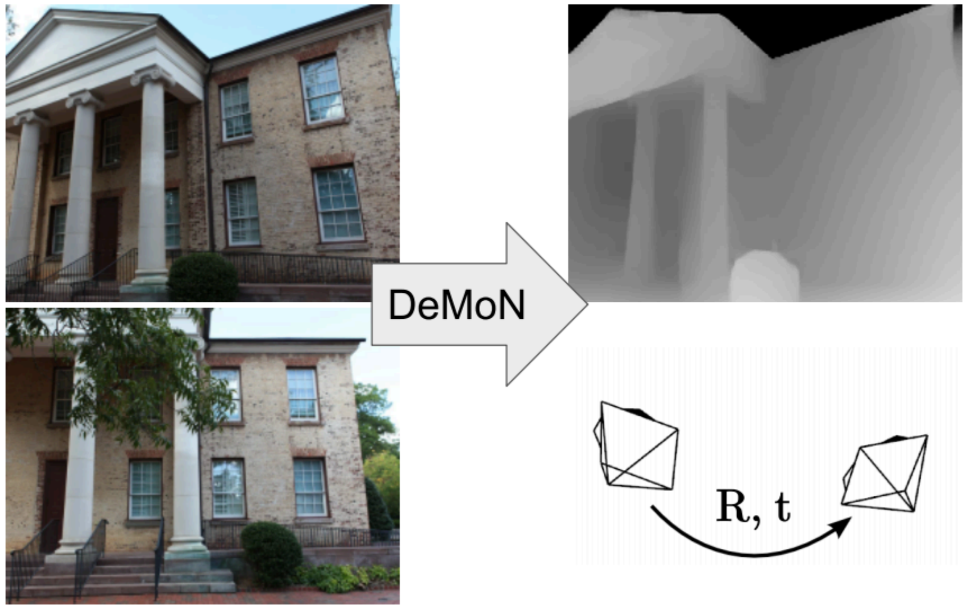


(c) Retrieve pose

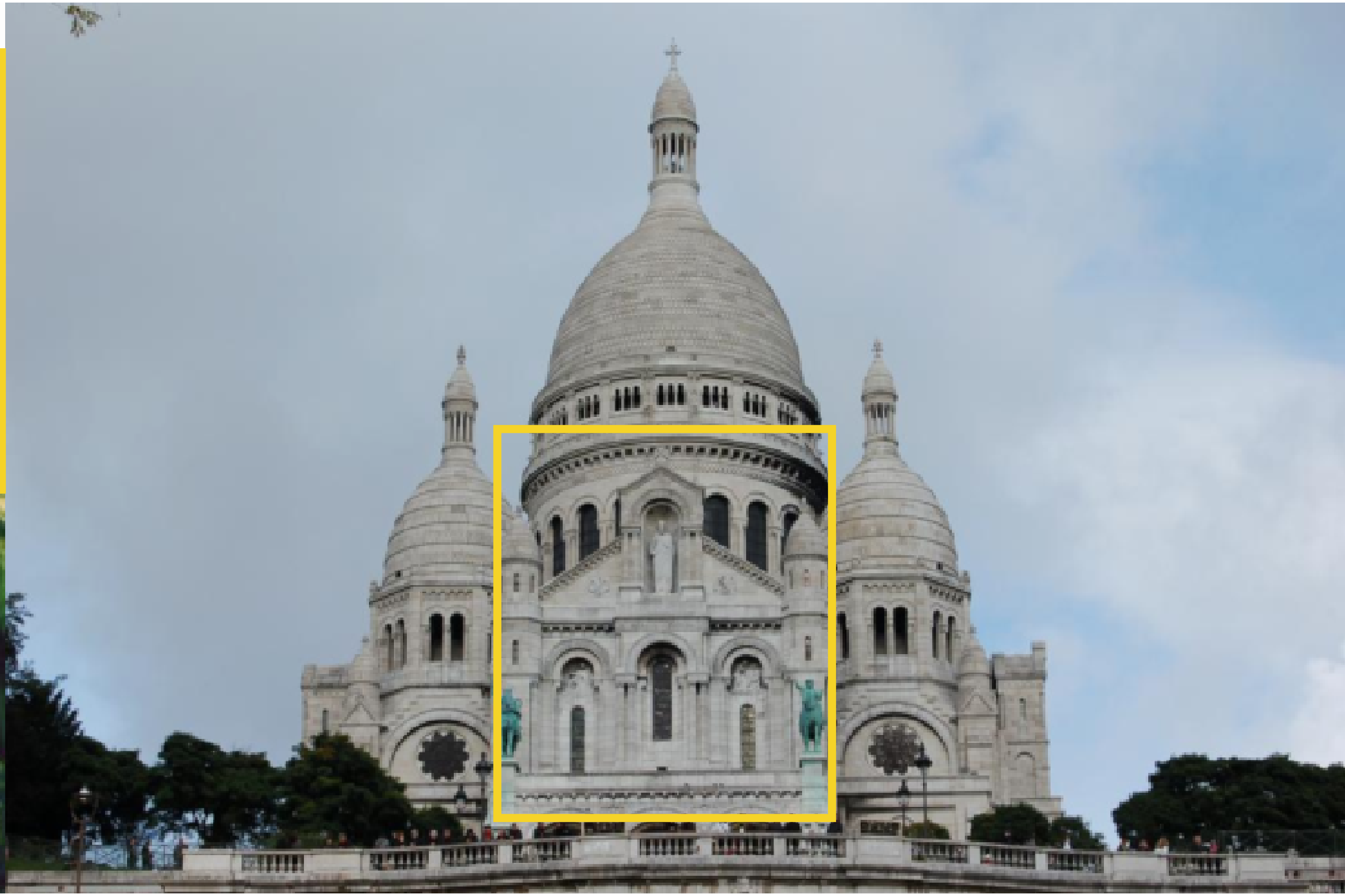
Fischler & Bolles, “Random Sample Consensus”. Comm. ACM, 1981

Dense matching with CNNs

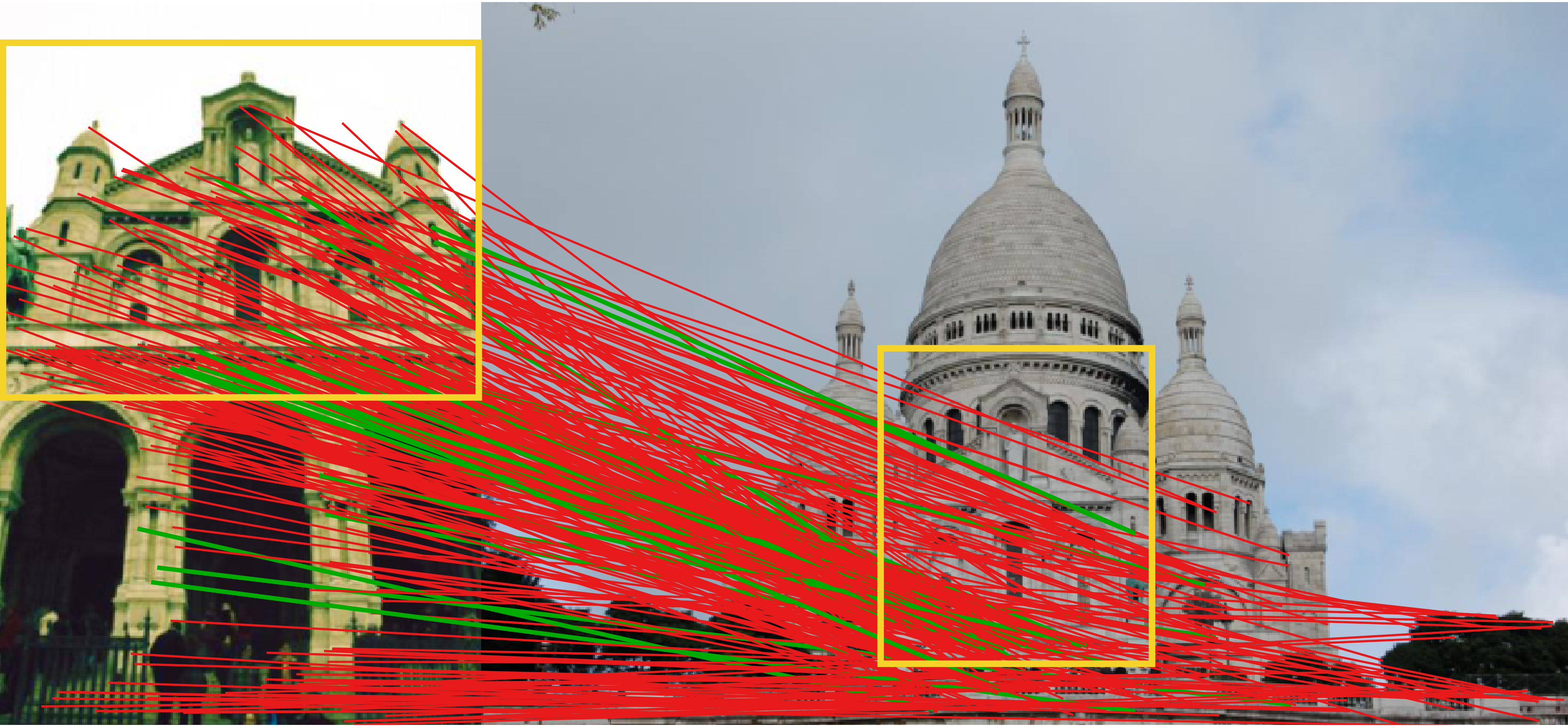
- Current focus of research:
  - ✤ Zamir et al, ECCV’16.
  - ✤ SfM-Net, arxiv’17.
  - ✤ DeMoN, CVPR’17.
  - ✤ Lowe et al, CVPR’17.
- Focus: video, small displacements.
- General case (wide baselines) remains unsolved.



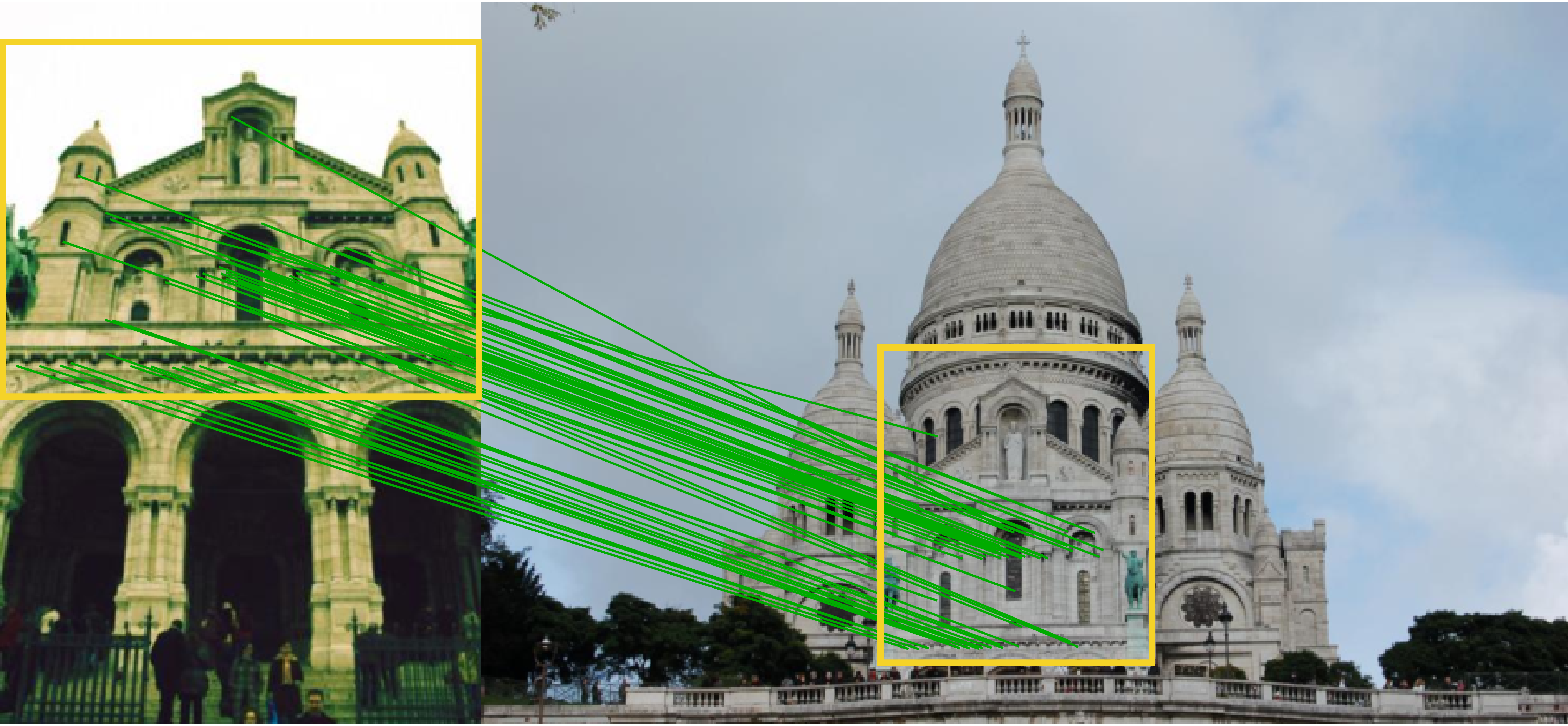
Where’s the challenge?

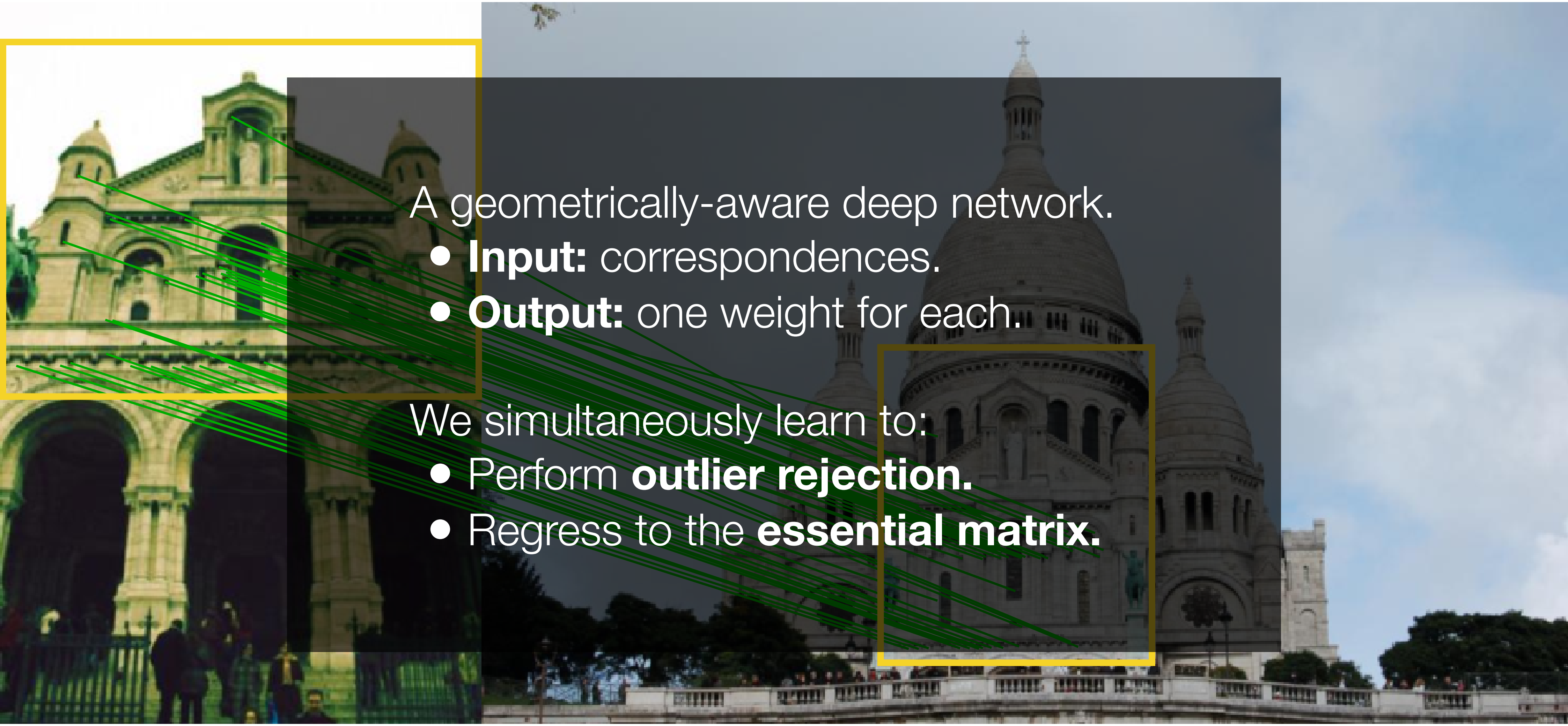
RANSAC: not always enough



Geometry to the rescue



Geometry to the rescue



A geometrically-aware deep network.

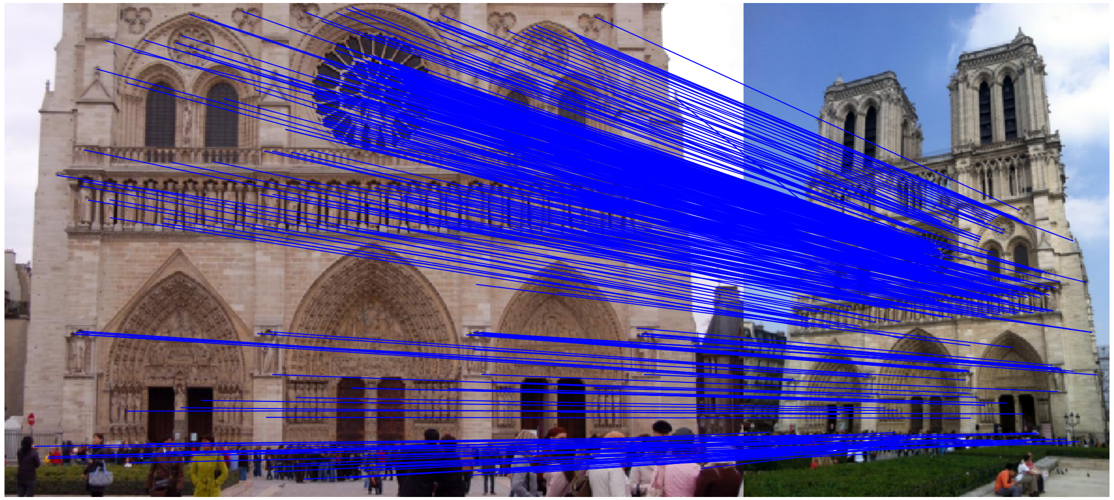
- **Input:** correspondences.
- **Output:** one weight for each.

We simultaneously learn to:

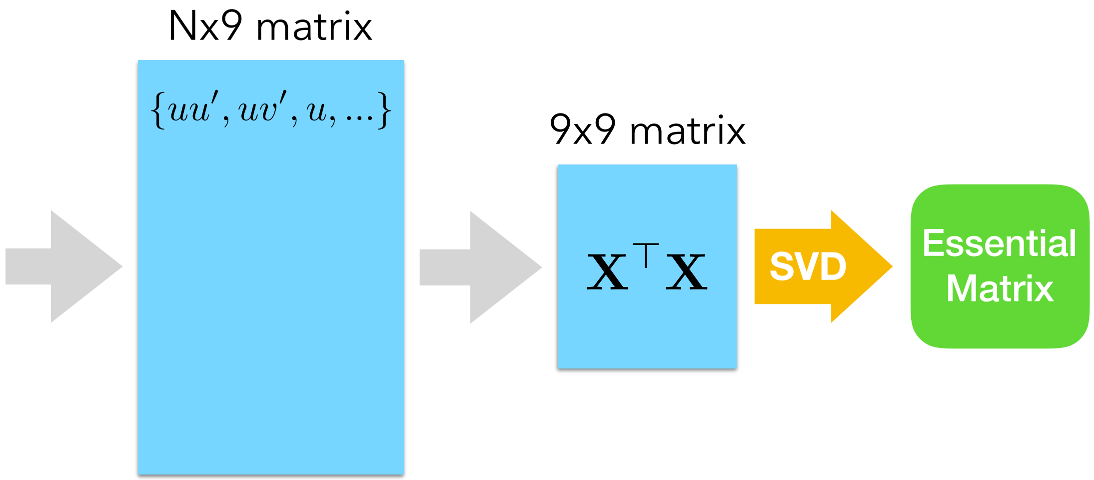
- Perform **outlier rejection.**
- Regress to the **essential matrix.**

Computing the Essential matrix

Closed form solution: **8-point algorithm**



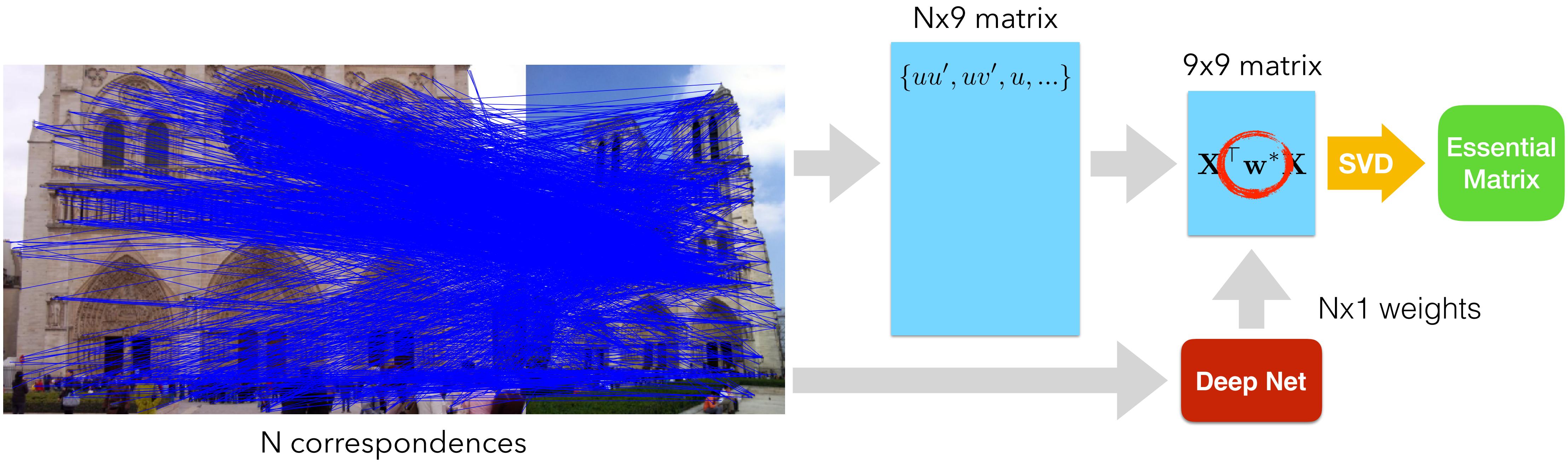
N correspondences



Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections”. Nature, 1981.

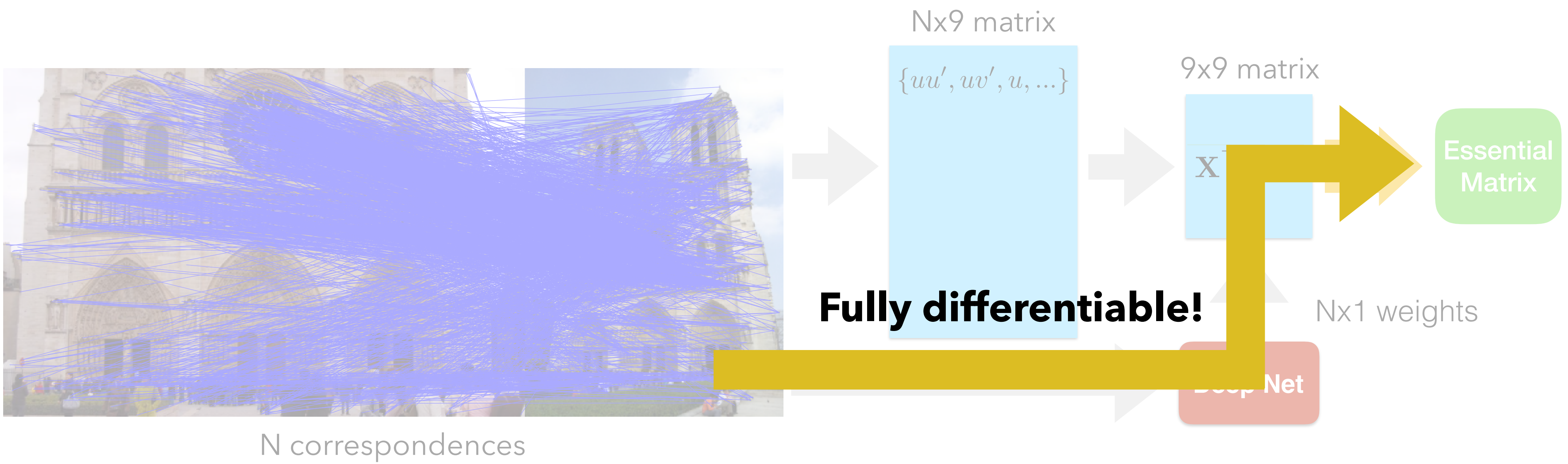
Learning to compute weights

We learn to compute **weights for the 8-point algorithm**



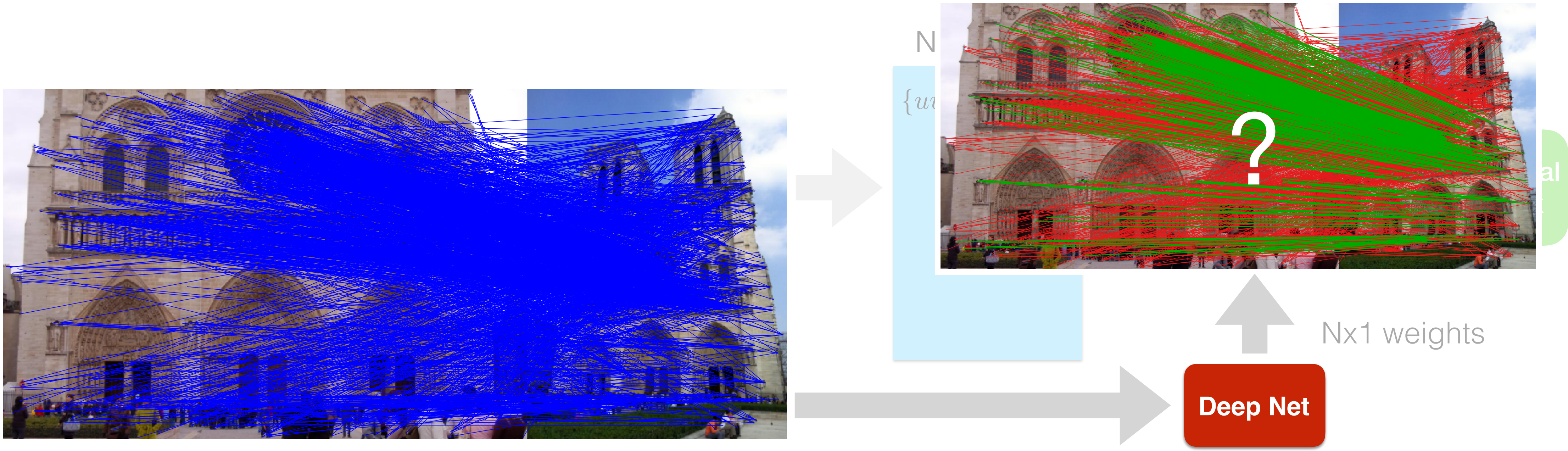
Learning to compute weights

We learn to compute **weights for the 8-point algorithm**



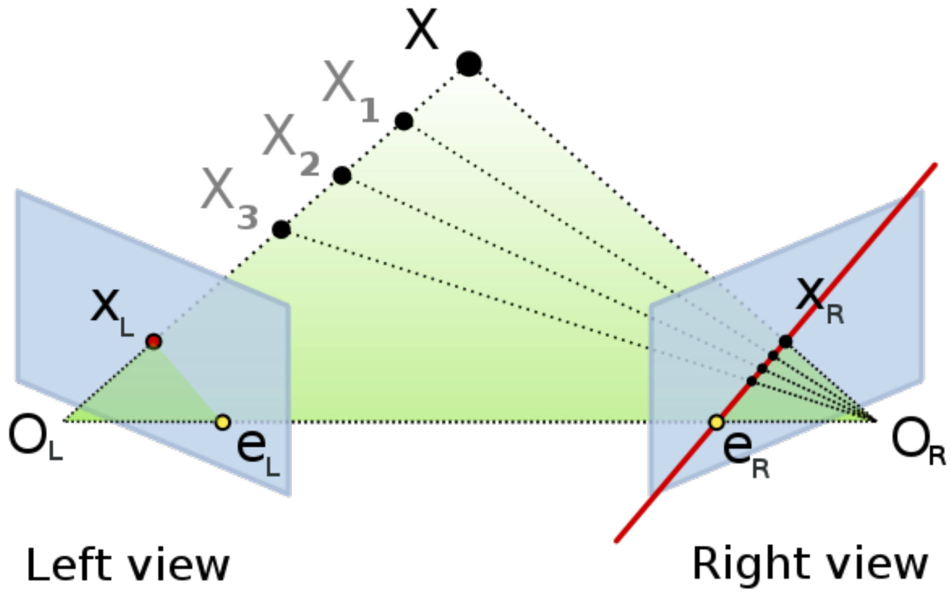
Learning to compute weights

We learn to compute **weights for the 8-point algorithm**



N correspondences

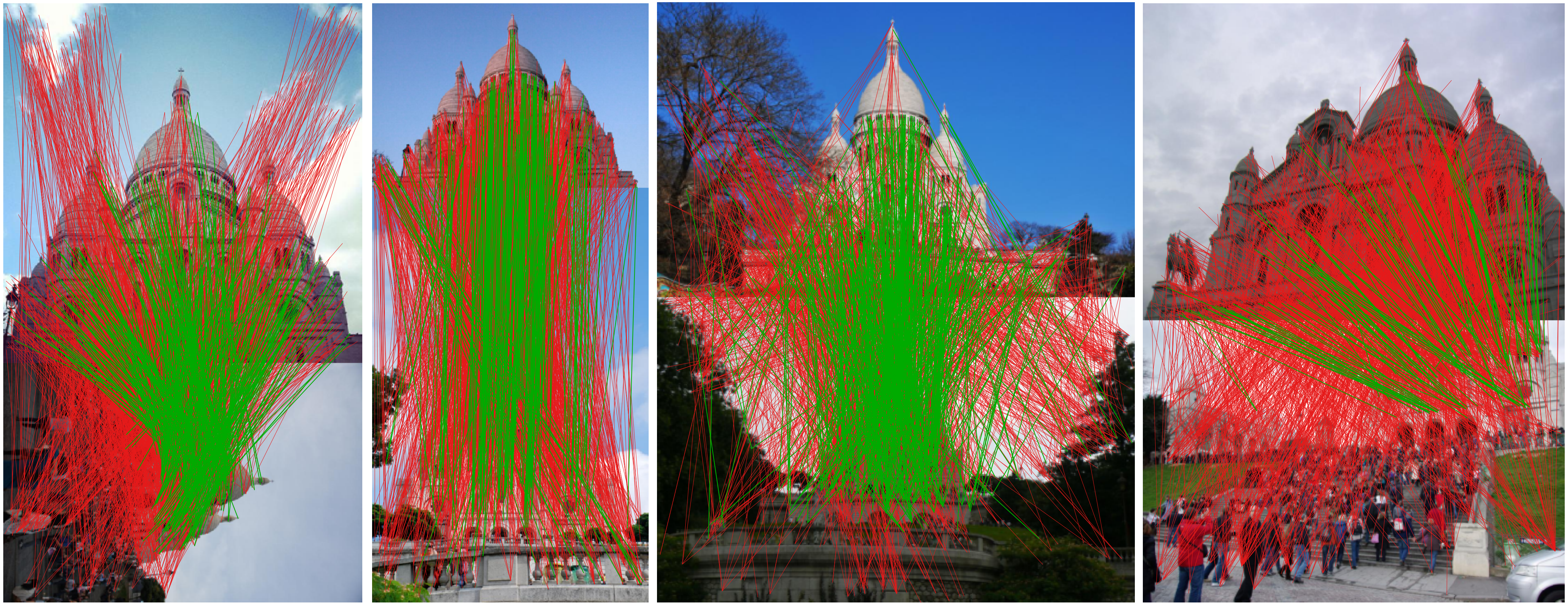
Exploiting epipolar geometry



We do not have dense depth data. But we have the **ground truth camera poses.**
With **epipolar geometry** we know that points in image 1 map to lines in image 2.

Source: wikipedia (https://en.wikipedia.org/wiki/Epipolar_geometry)

Adding a classification loss



Not perfect (point ↔ line)! But good enough for a **supervision signal.**

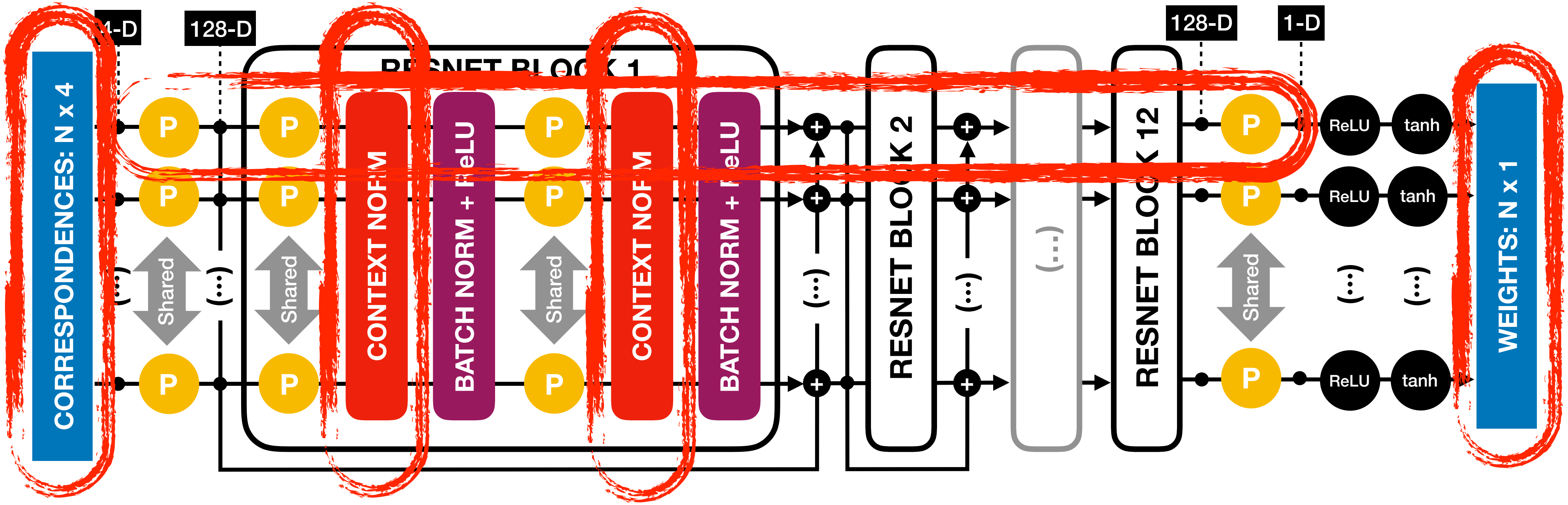
Hartley & Zisserman, “Multiple view geometry in computer vision”, 2000.

Complete formulation

We jointly train for outlier rejection and regression to the Essential matrix by minimizing the hybrid loss:

$$\mathcal{L}(\Phi) = \sum_{k=1}^{P} \Big(\underbrace{\alpha \mathcal{L}_x(\Phi, \mathbf{x}_k)}_{\text{Classification (Inliers vs outliers)}} + \underbrace{\beta \mathcal{L}_e(\Phi, \mathbf{x}_k)}_{\text{Regression (which inliers help us retrieve E?)}} \Big)$$

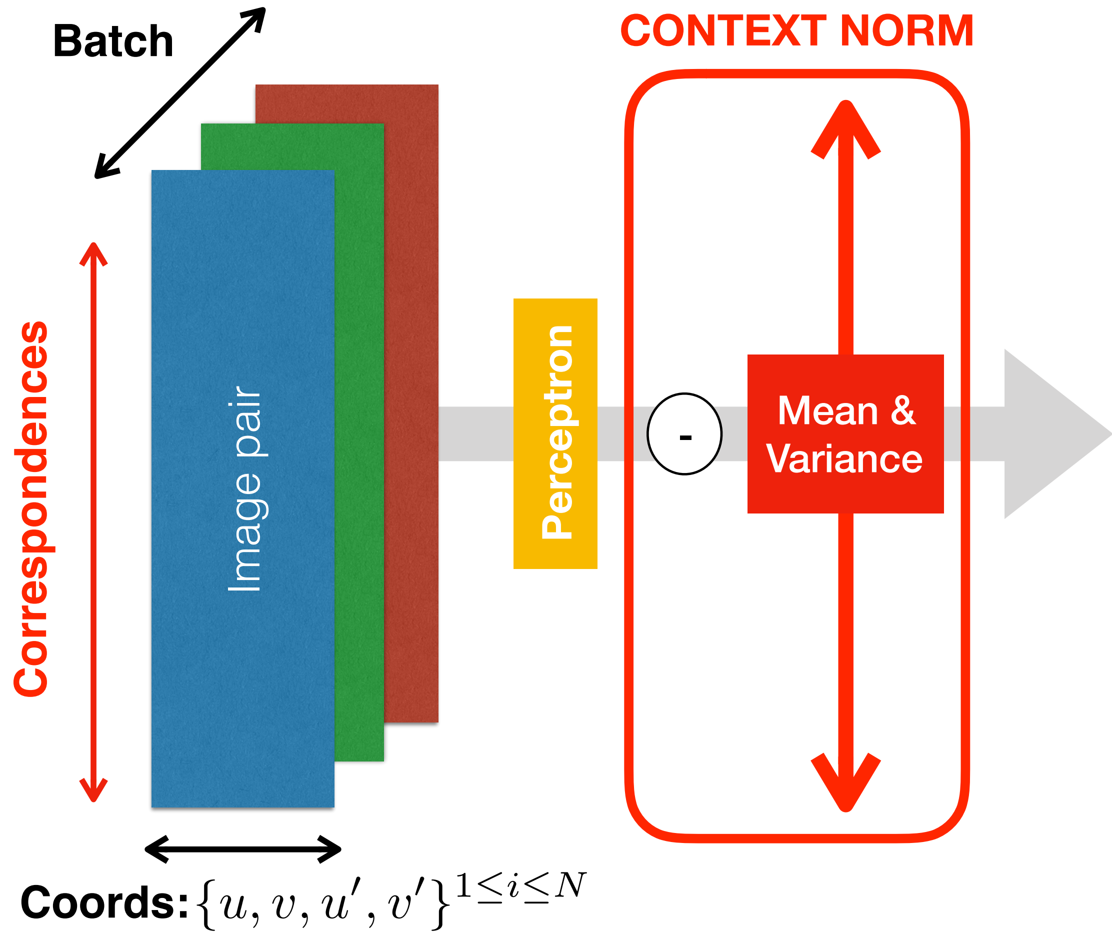
Our network



- **Input:** putative matches (SIFT+NN). Coordinates only: $\{u, v, u', v'\}^{1 \leq i \leq N}$
- **Output:** Weights, encoding inlier probability.
- **Network:** MLPs. Global context embedded via Context Normalization.

Embedding context

- Non-parametric normalization of the mean/std of feature maps.
- Applied over each image pair in the batch separately.
- Also known as Instance Norm, used in image stylization.



Results

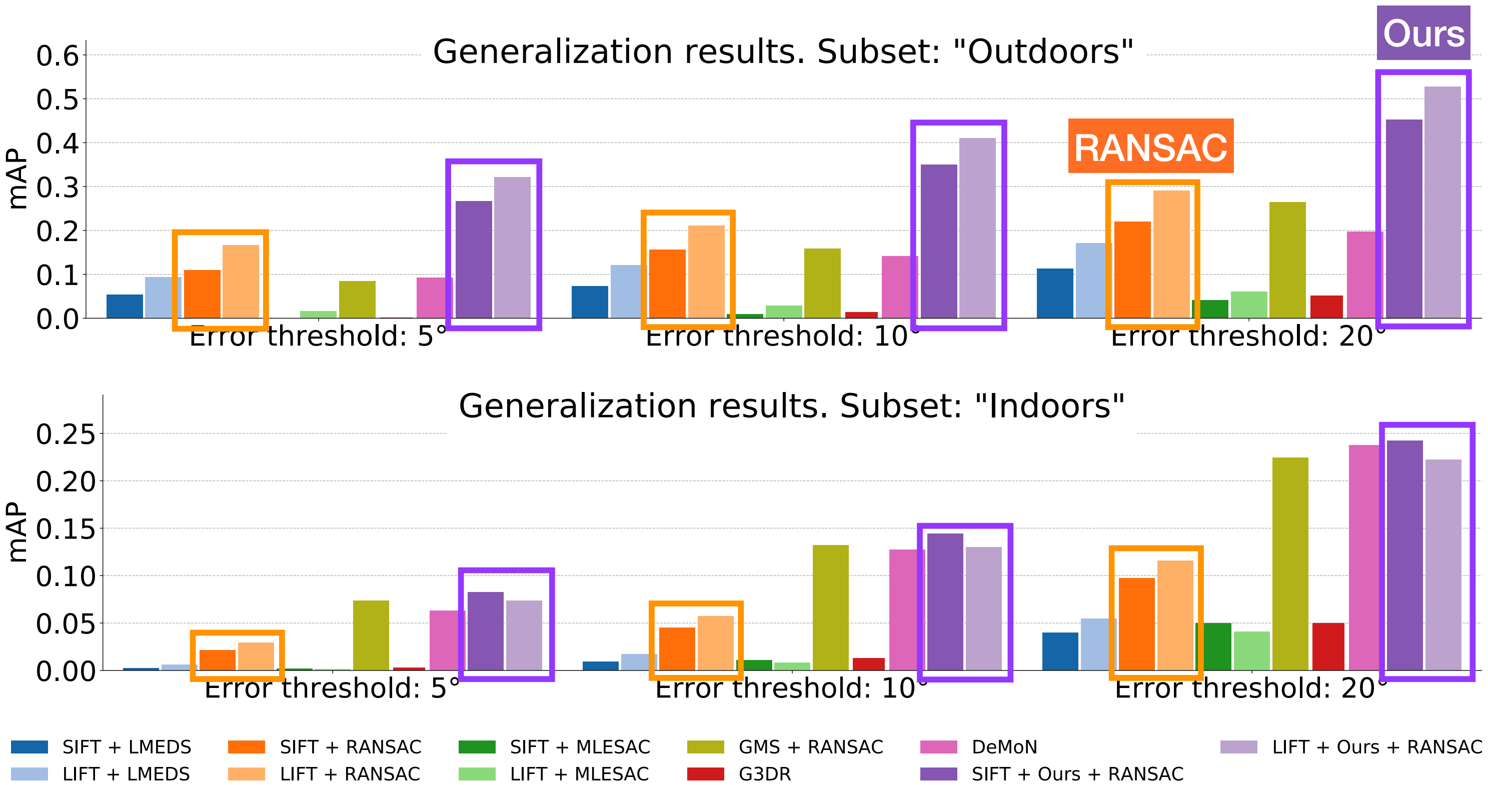
Train on only **two sequences:** one indoors & one outdoors (10k pairs from each):

(i) St. Peter's Square (2.5k images)

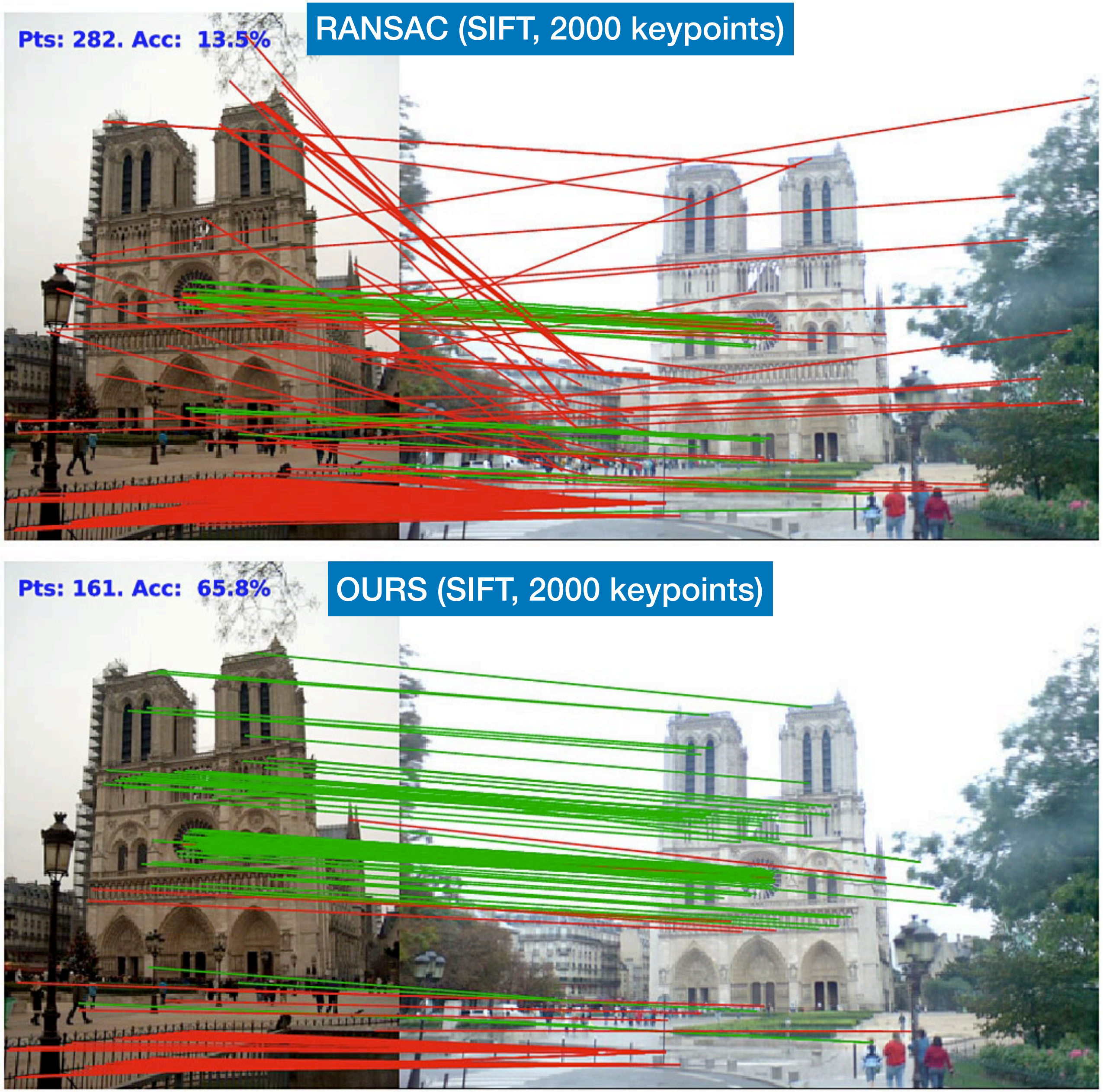
(ii) Brown (video, 8k images)

Test on **completely different** sequences (1k pairs from each):

Results



Outdoors: great performance. **Indoors:** slightly better than dense methods.


Pts: 282. Acc: 13.5%
RANSAC (SIFT, 2000 keypoints)
Pts: 161. Acc: 65.8%
OURS (SIFT, 2000 keypoints)

Collaborators

Kwang Yi
(U. Victoria)

Eduard Trulls
(EPFL)

Yuki Ono
(Sony)

Mathieu Salzmann
(EPFL)

Vincent Lepetit
(U. Bordeaux)

Pascal Fua
(EPFL)

Code and models: **github.com/vcg-uvic/learned-correspondence-release**

Thanks for your attention.
Questions?