Out with the Old?

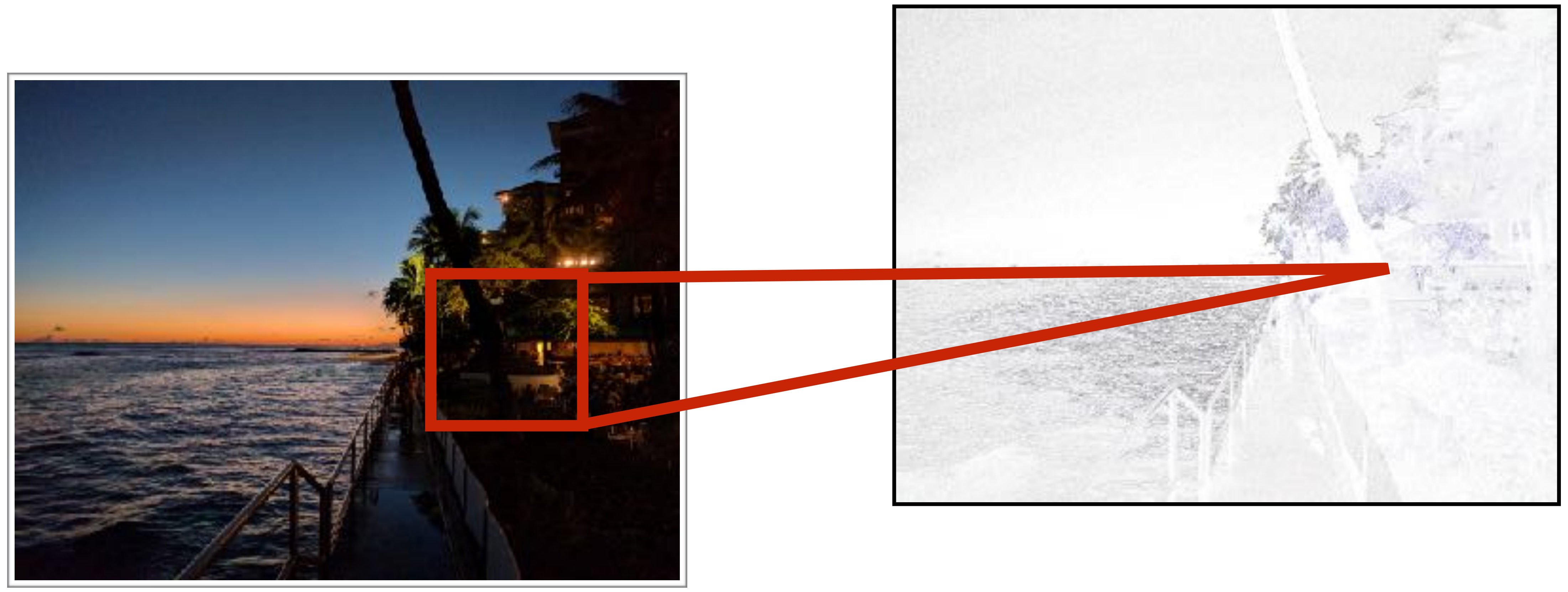
Convolutional Neural Networks for Feature Matching and Visual Localization

Torsten Sattler

Computer Vision and Geometry (CVG) Lab
ETH Zürich

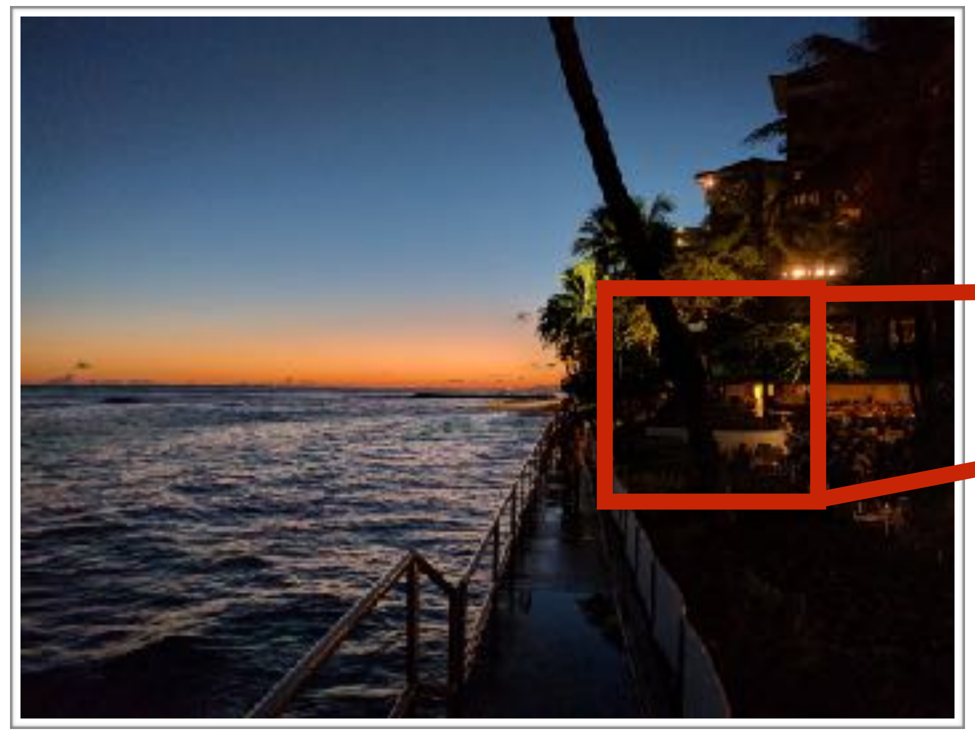
Convolutional Neural Networks

Convolutional Neural Networks



convolutions +
non-linearity

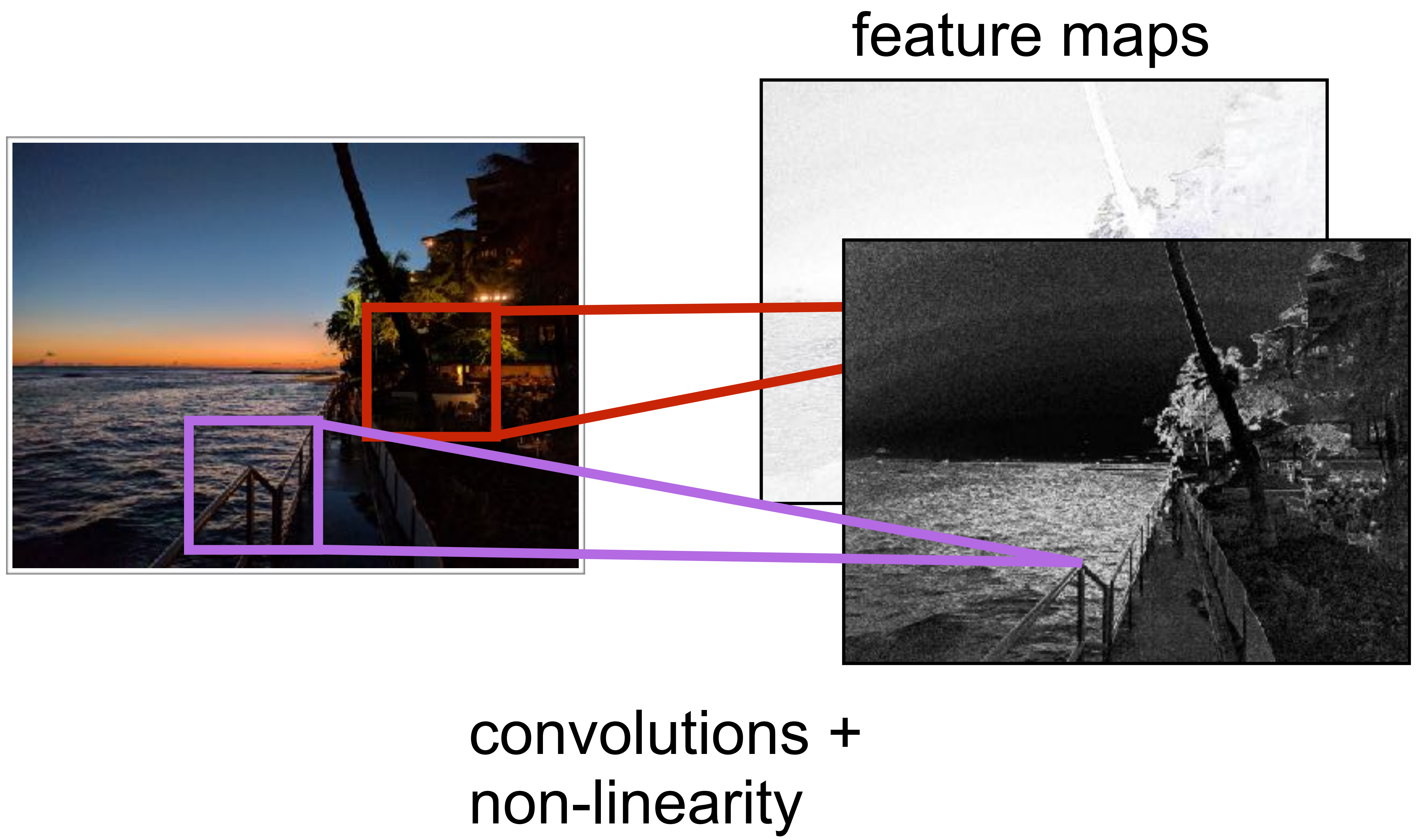
Convolutional Neural Networks



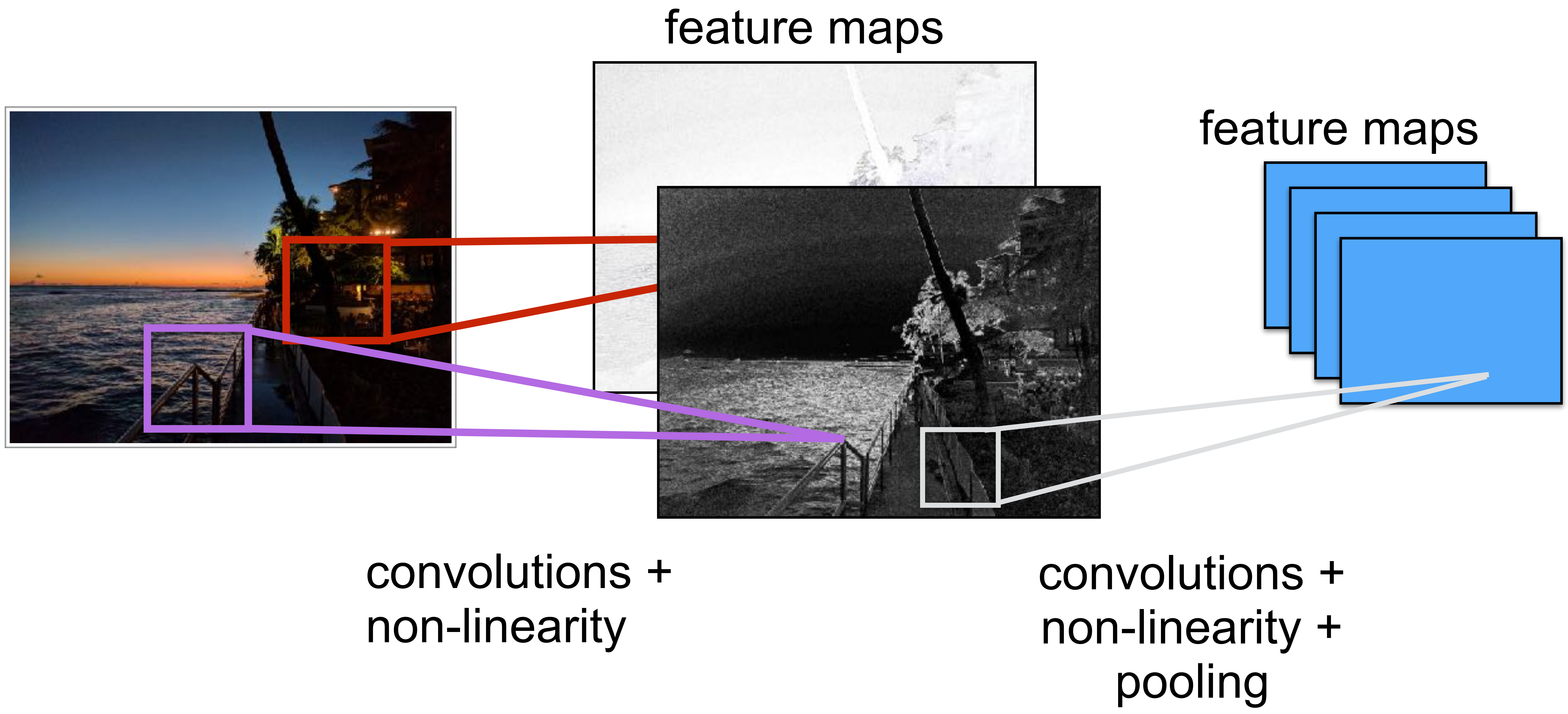
feature maps

convolutions +
non-linearity

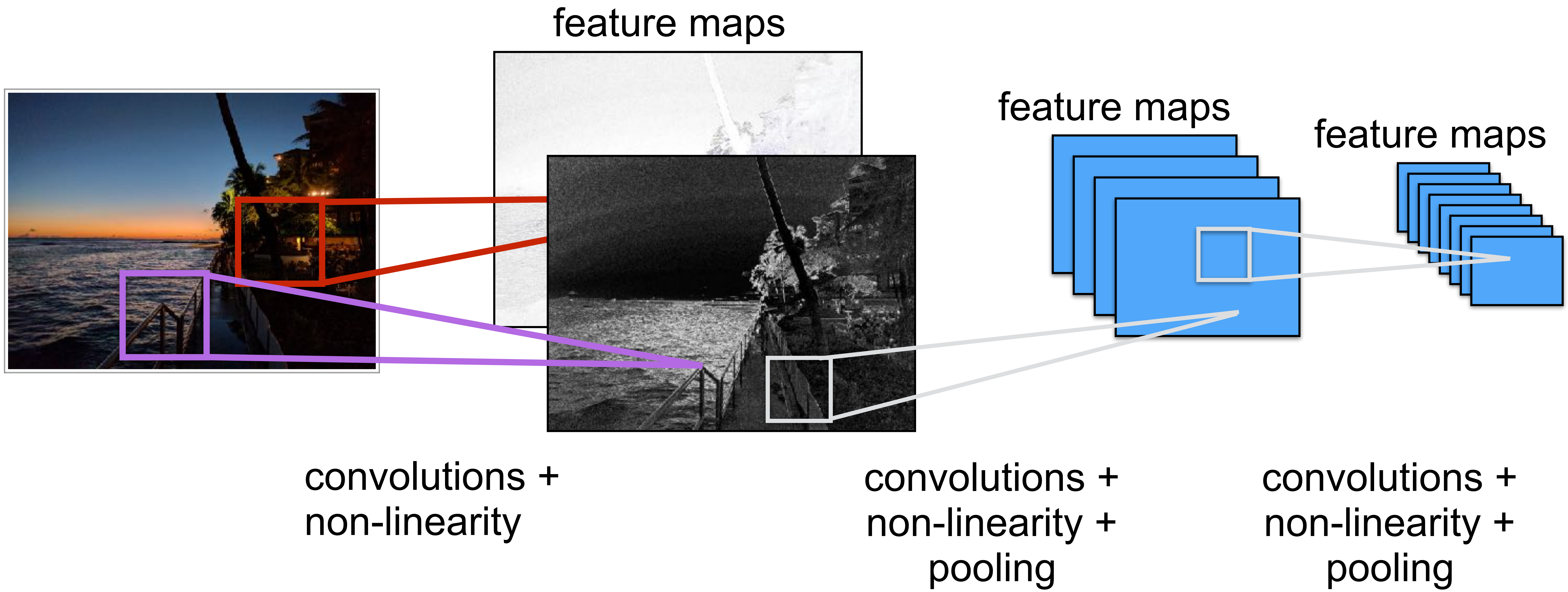
Convolutional Neural Networks



Convolutional Neural Networks



Convolutional Neural Networks

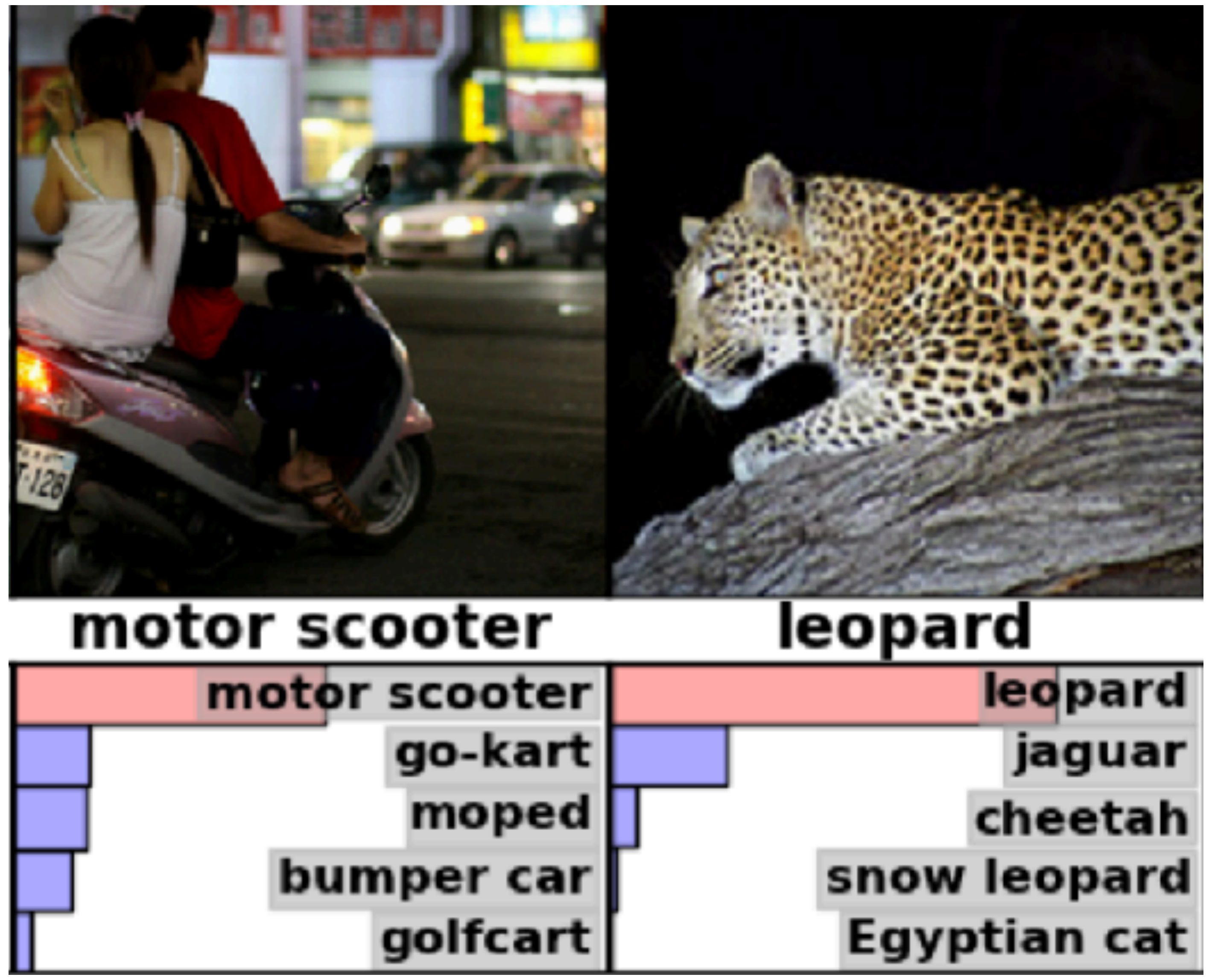


Convolutional Neural Networks

Convolutional Neural Networks

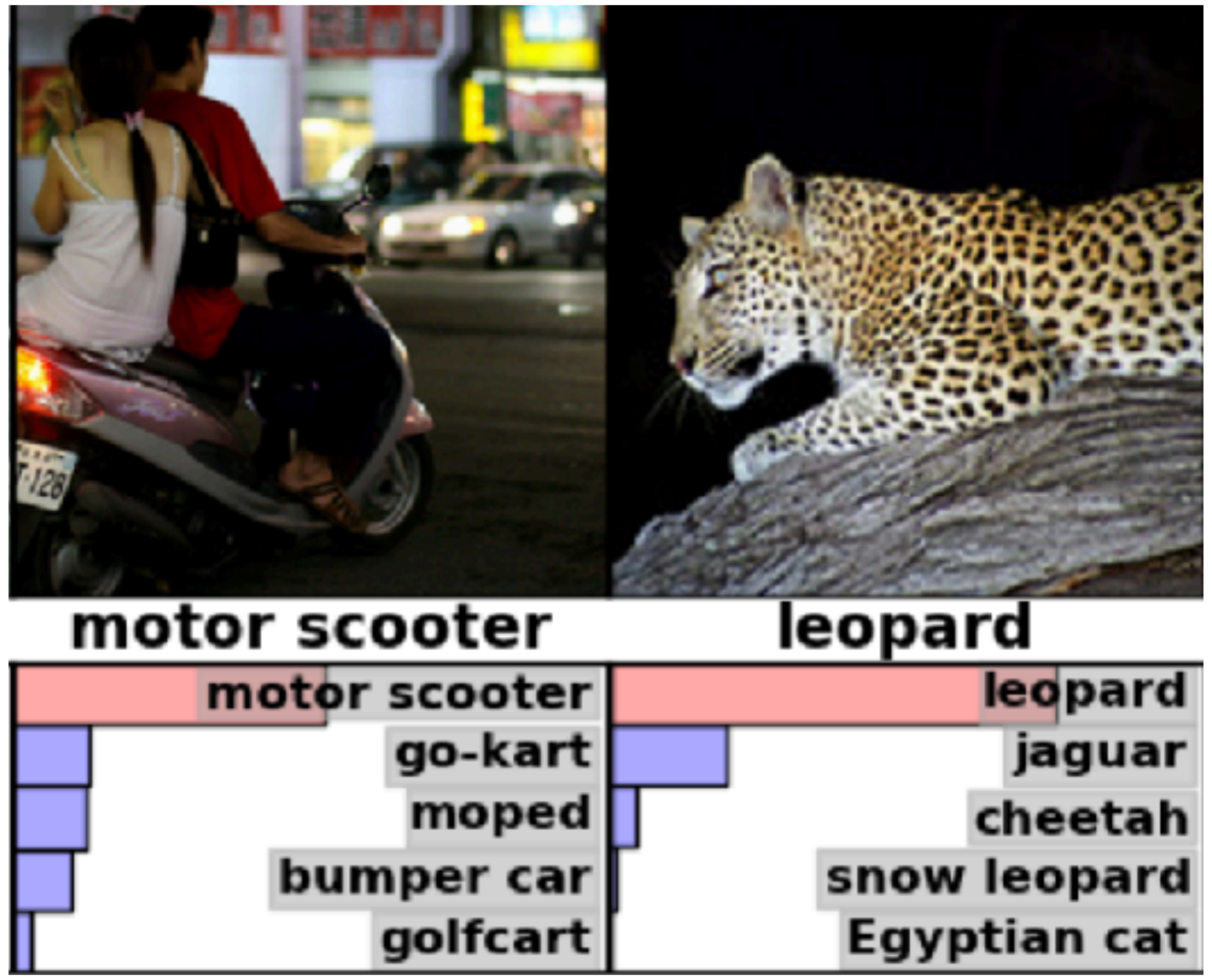
Deep Learning Revolution

Deep Learning Revolution

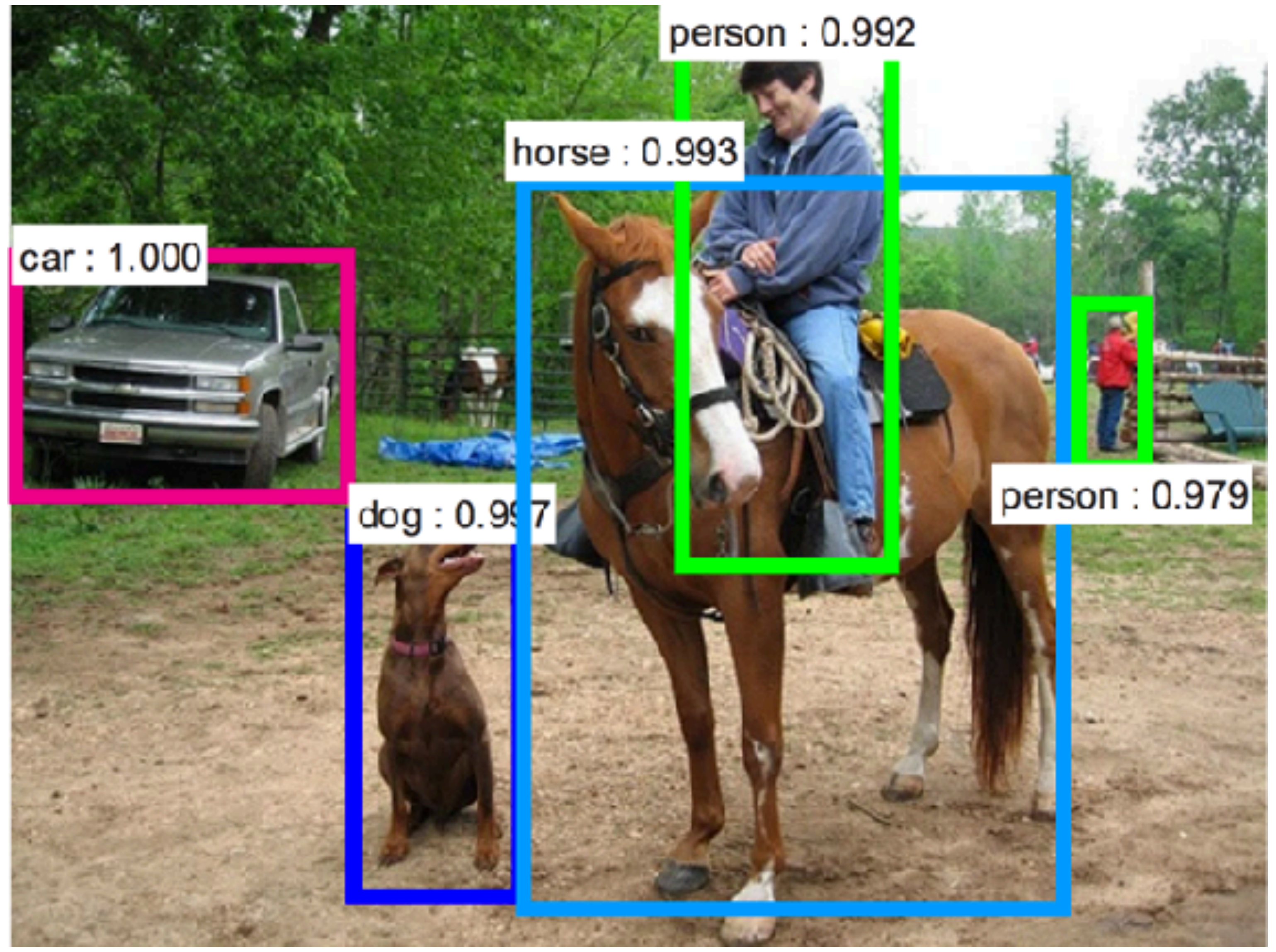


[Krizhevsky et al., ImageNet Classification with Deep Convolutional Neural Networks , NIPS 2012]

Deep Learning Revolution

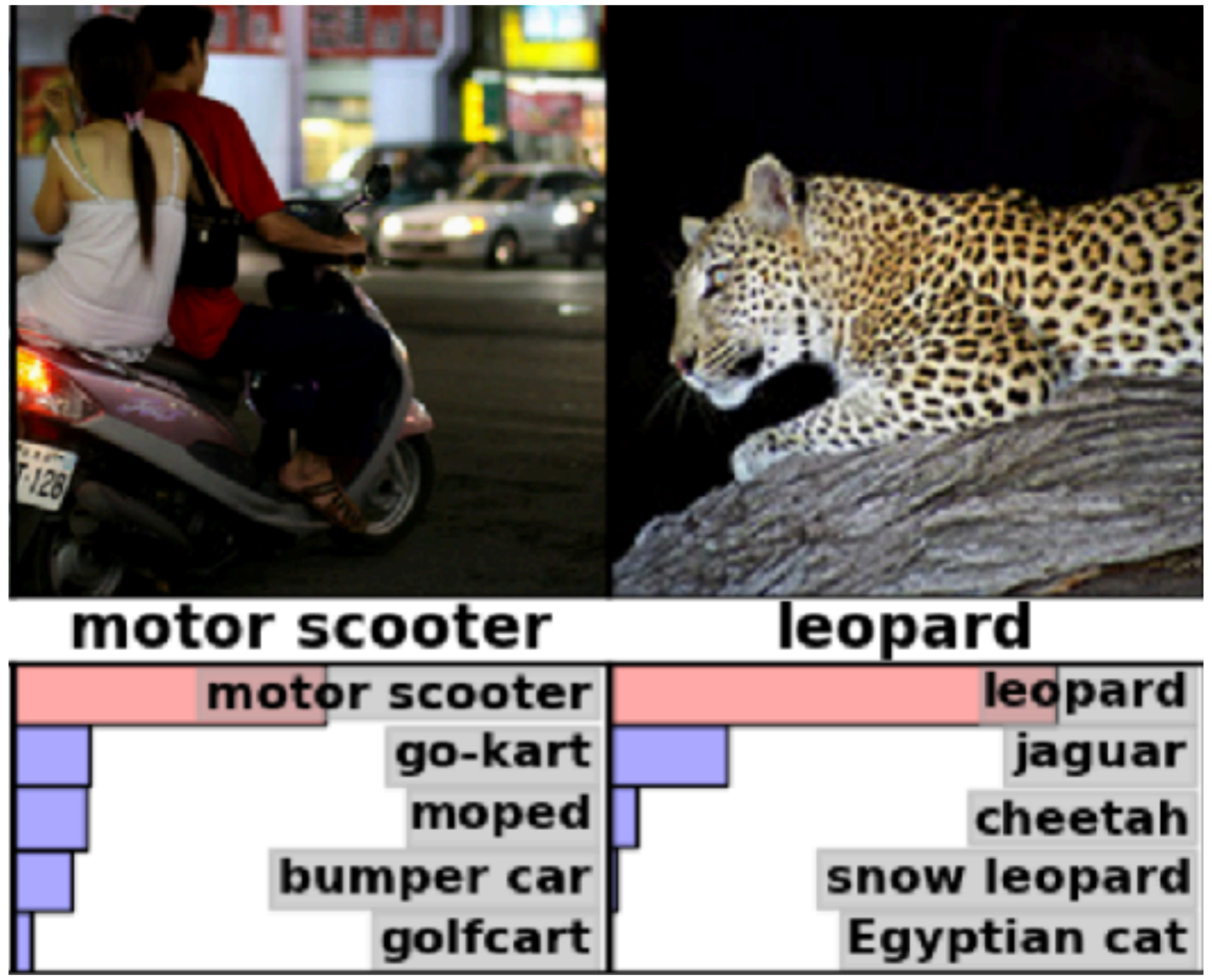


[Krizhevsky et al., ImageNet Classification with Deep Convolutional Neural Networks , NIPS 2012]

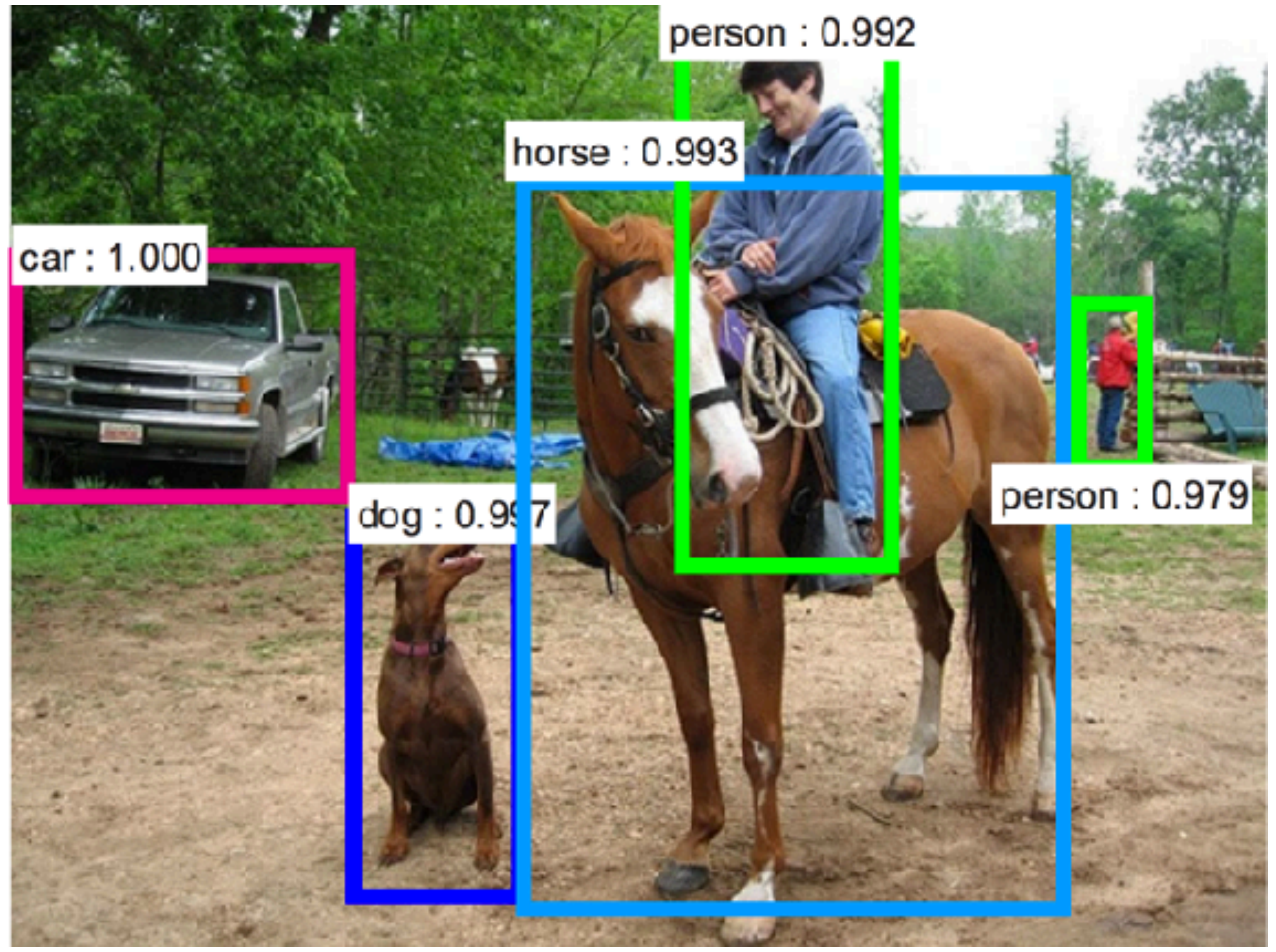


[Ren et al., Faster R-CNN: Towards real-time object detection with region proposal networks , NIPS 2015]

Deep Learning Revolution



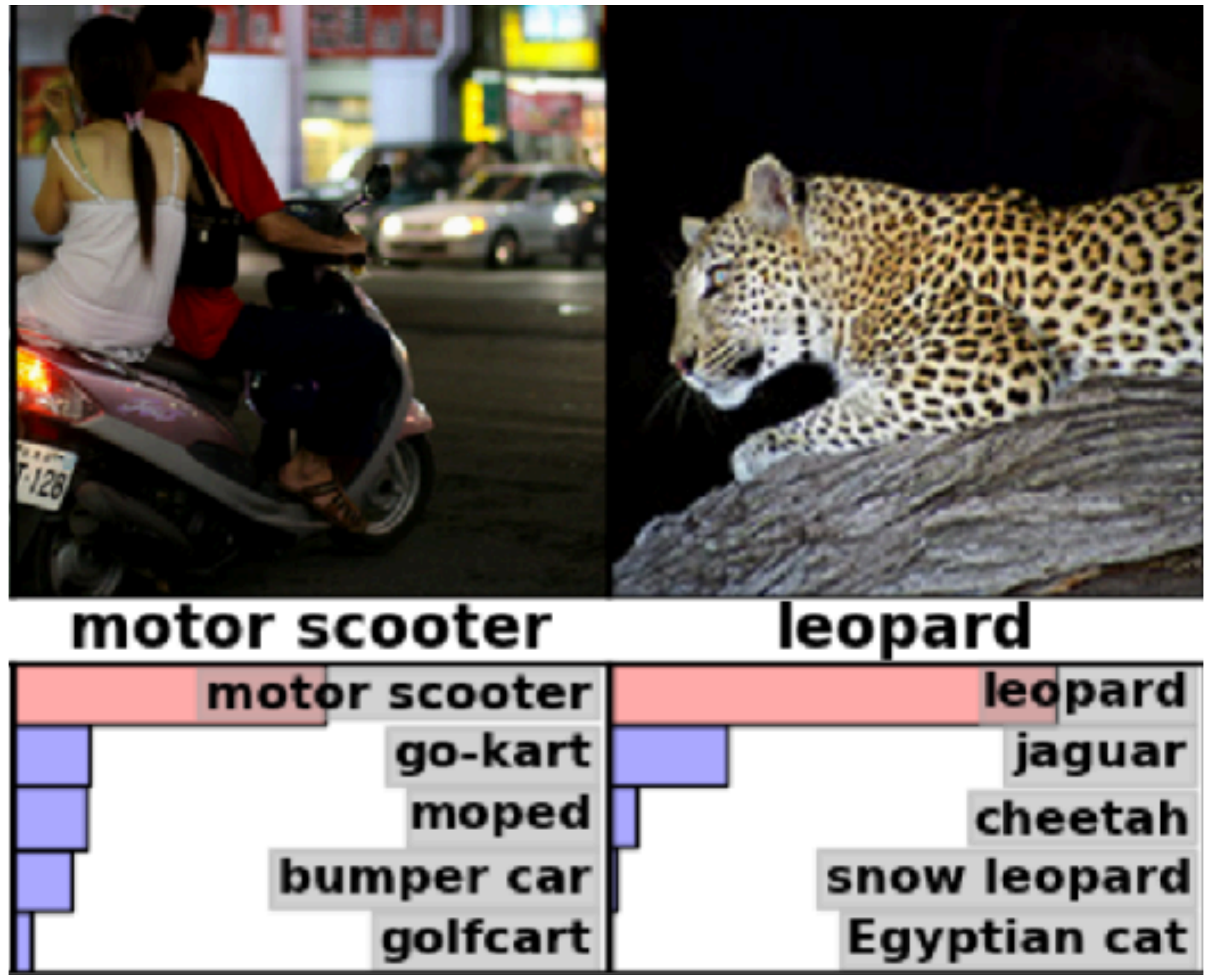
[Krizhevsky et al., ImageNet Classification with Deep Convolutional Neural Networks , NIPS 2012]



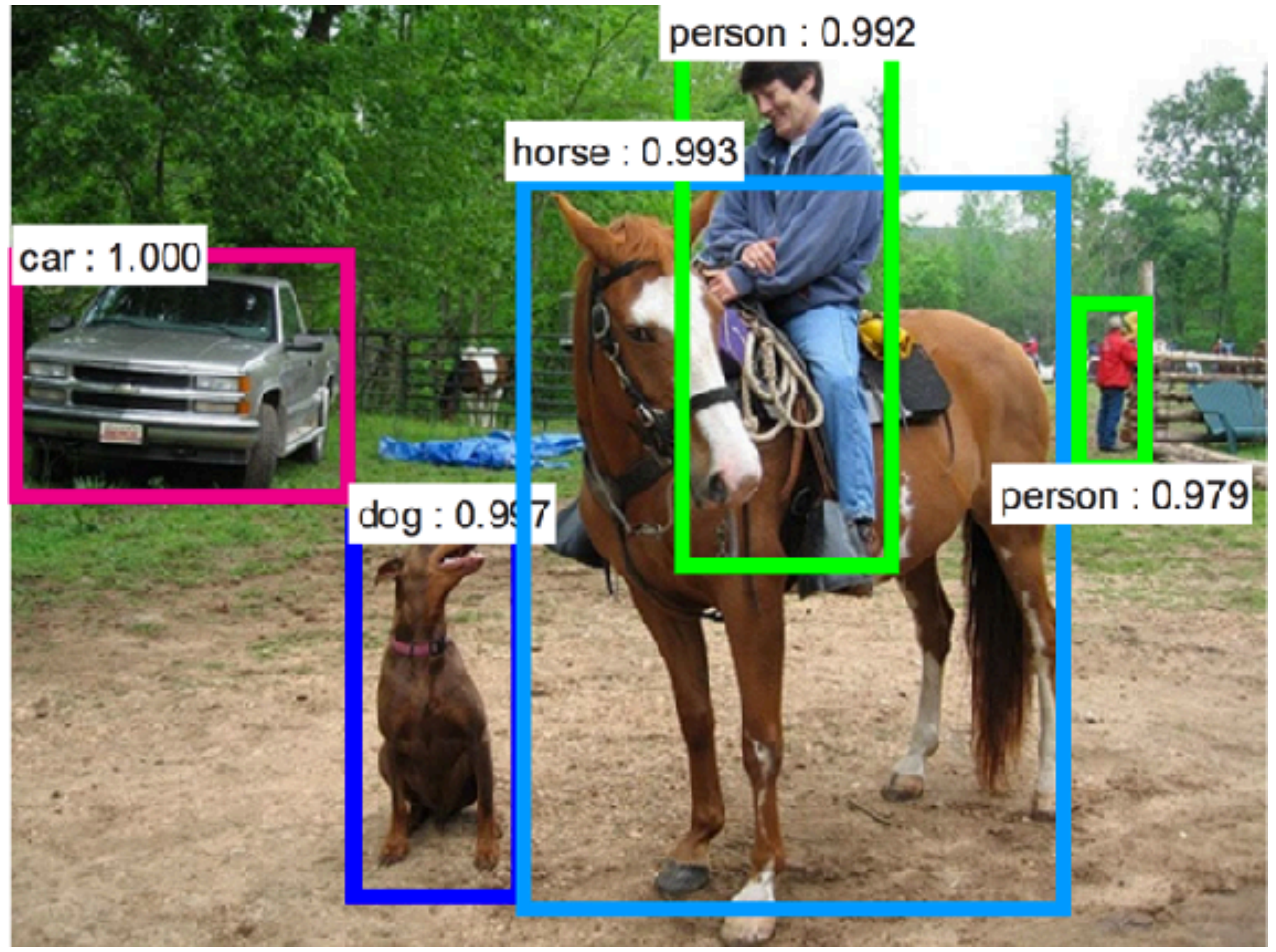
[Ren et al., Faster R-CNN: Towards real-time object detection with region proposal networks , NIPS 2015]

[Pohlen et al., Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes, CVPR 2017]

Deep Learning Revolution

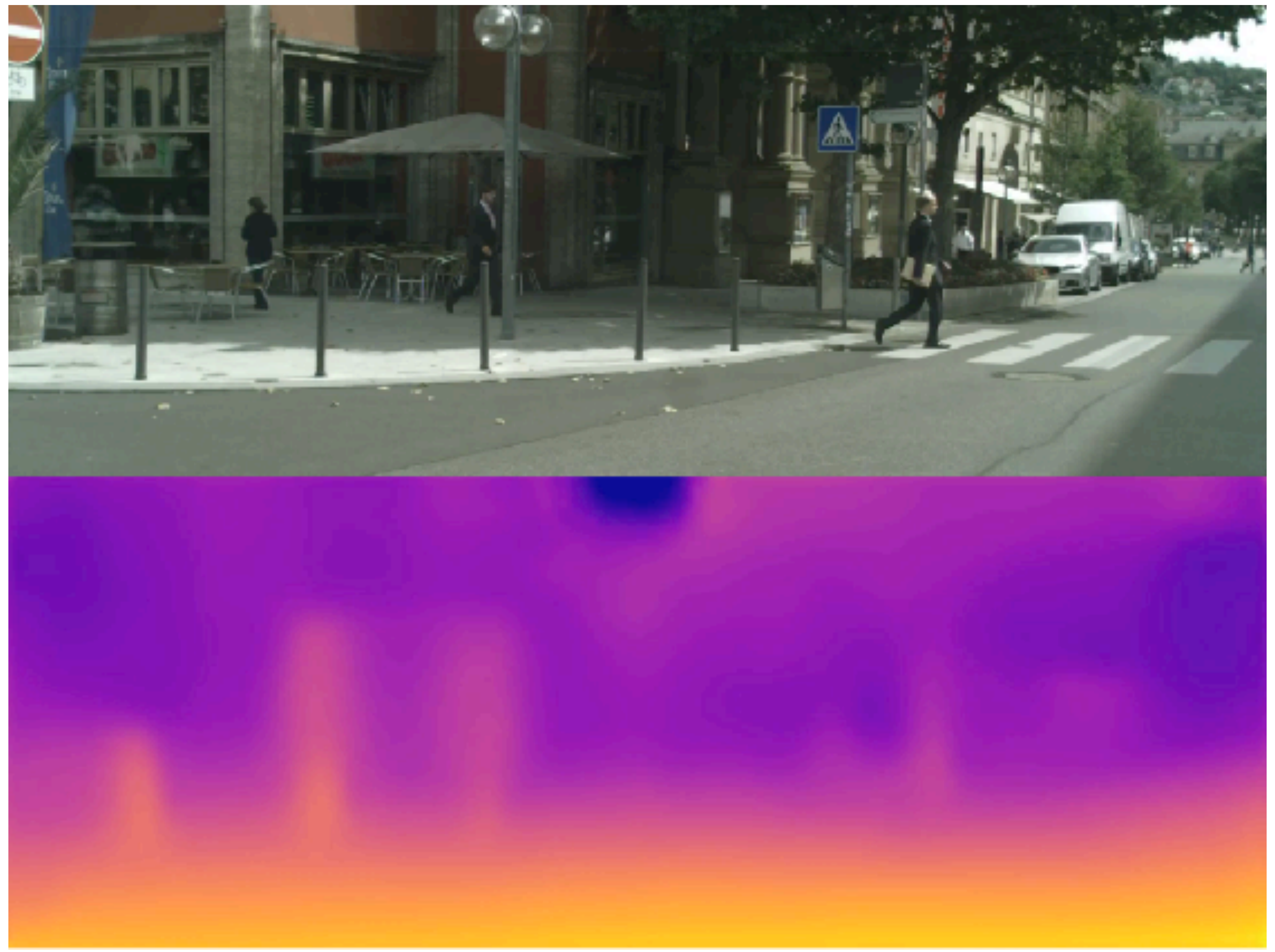


[Krizhevsky et al., ImageNet Classification with Deep Convolutional Neural Networks , NIPS 2012]



[Ren et al., Faster R-CNN: Towards real-time object detection with region proposal networks , NIPS 2015]

[Pohlen et al., Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes, CVPR 2017]



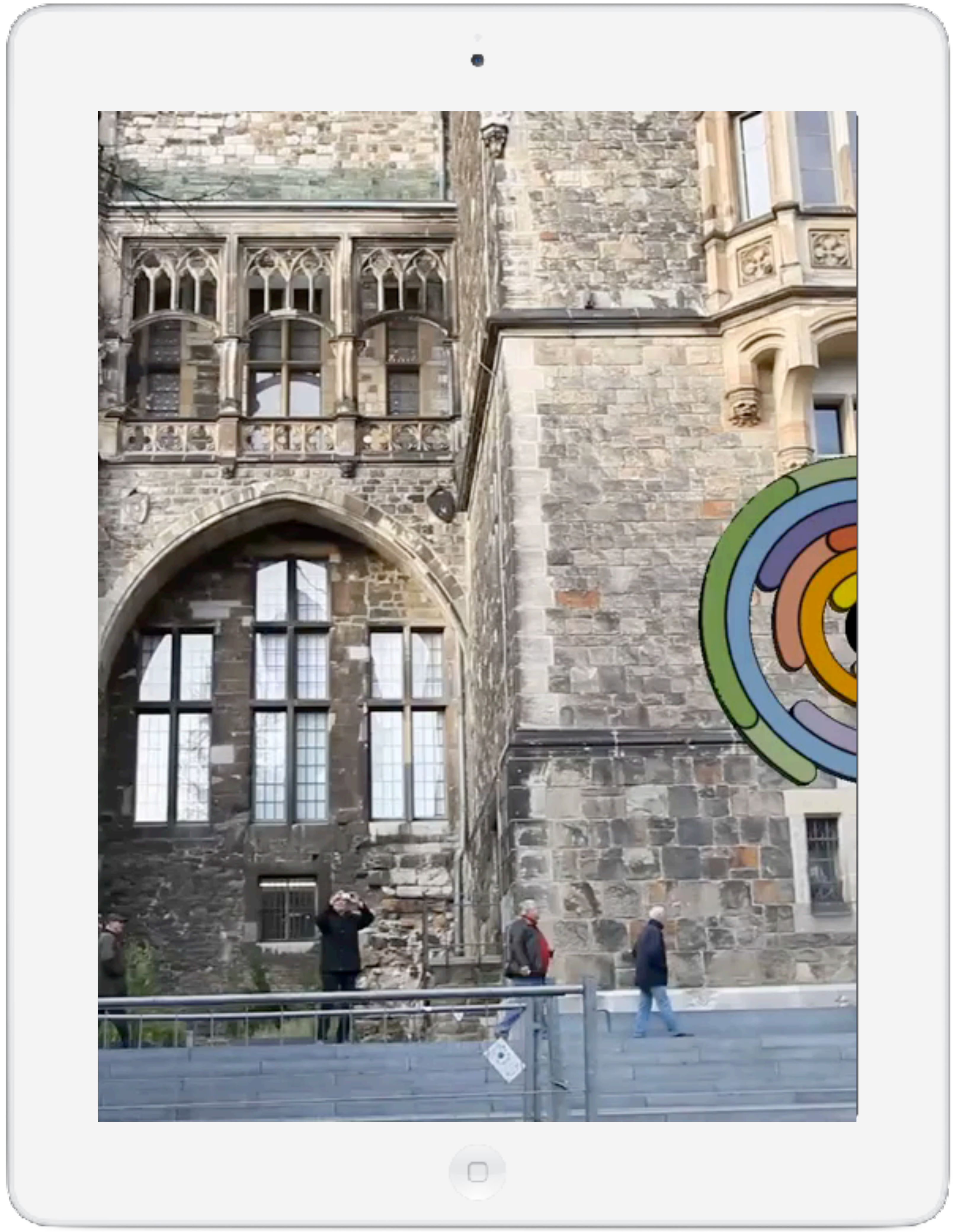
[Zhou et al., Unsupervised Learning of Depth and Ego-Motion from Video, CVPR 2017]

Overview

I. **CNNs for Visual Localization**

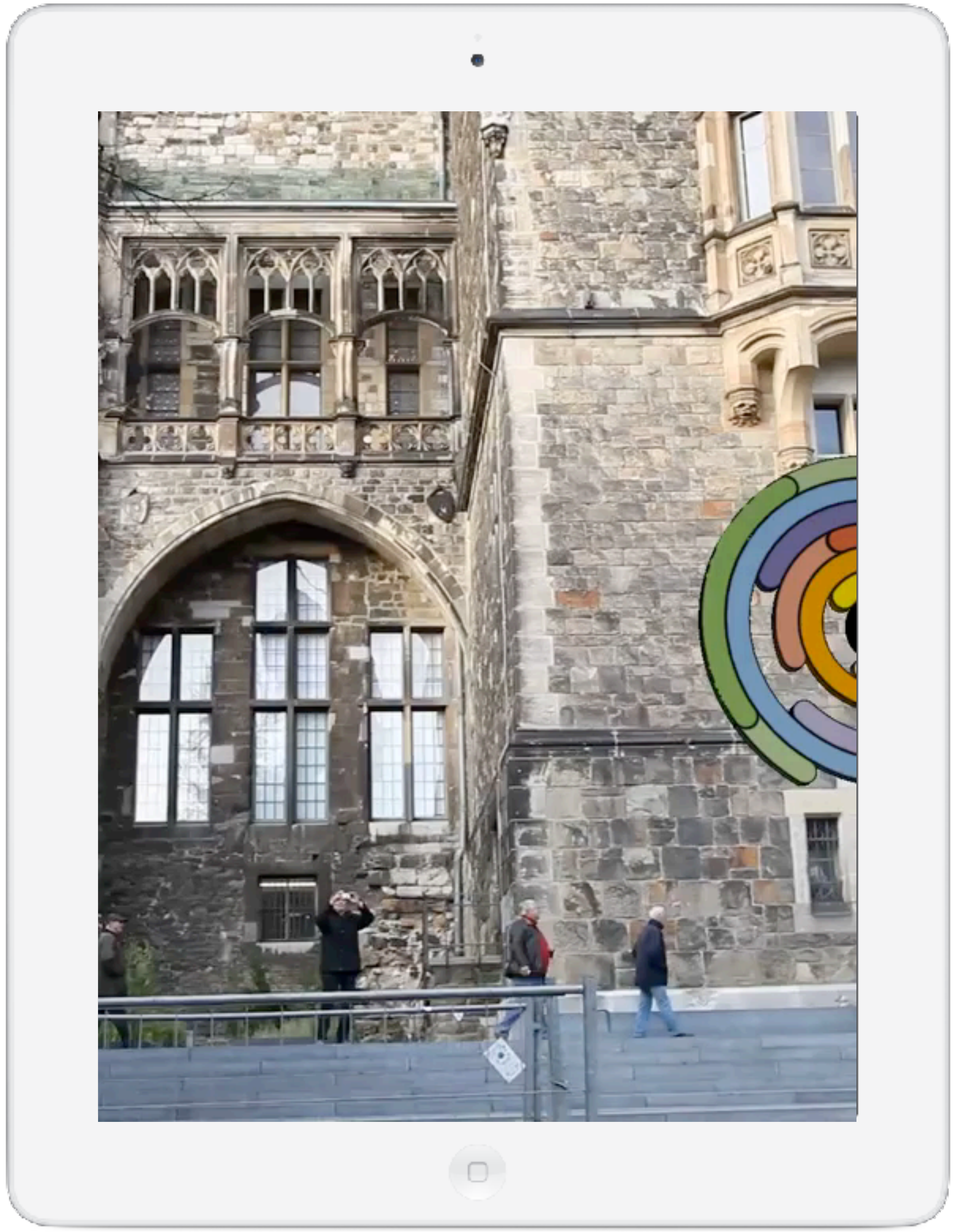
II. CNNs for Feature Detection & Description

Visual Localization



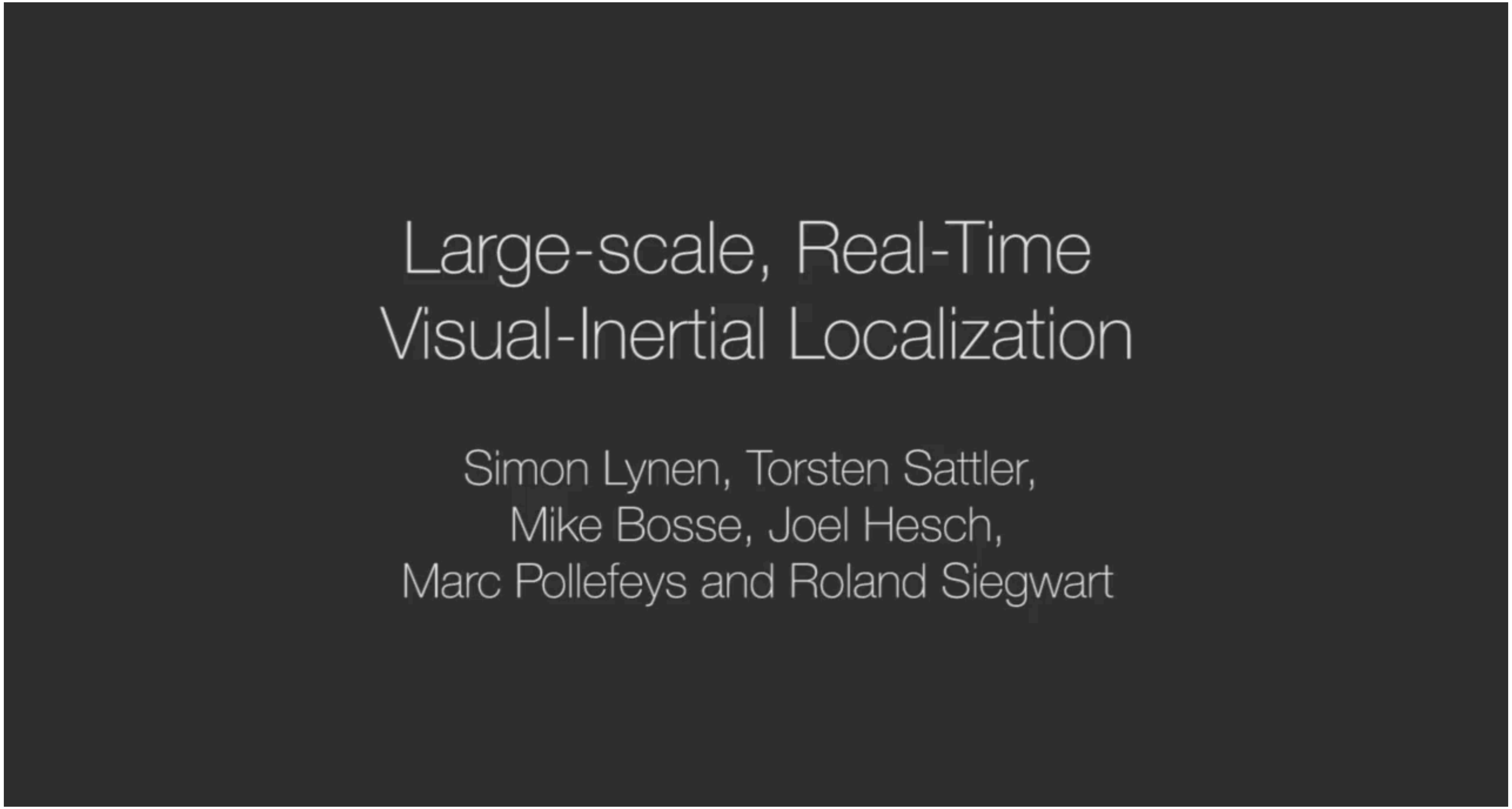
[Middelberg, **Sattler**, Untzelmann, Kobbelt, Scalable 6-DOF Localization on Mobile Devices. ECCV 2014]

Visual Localization



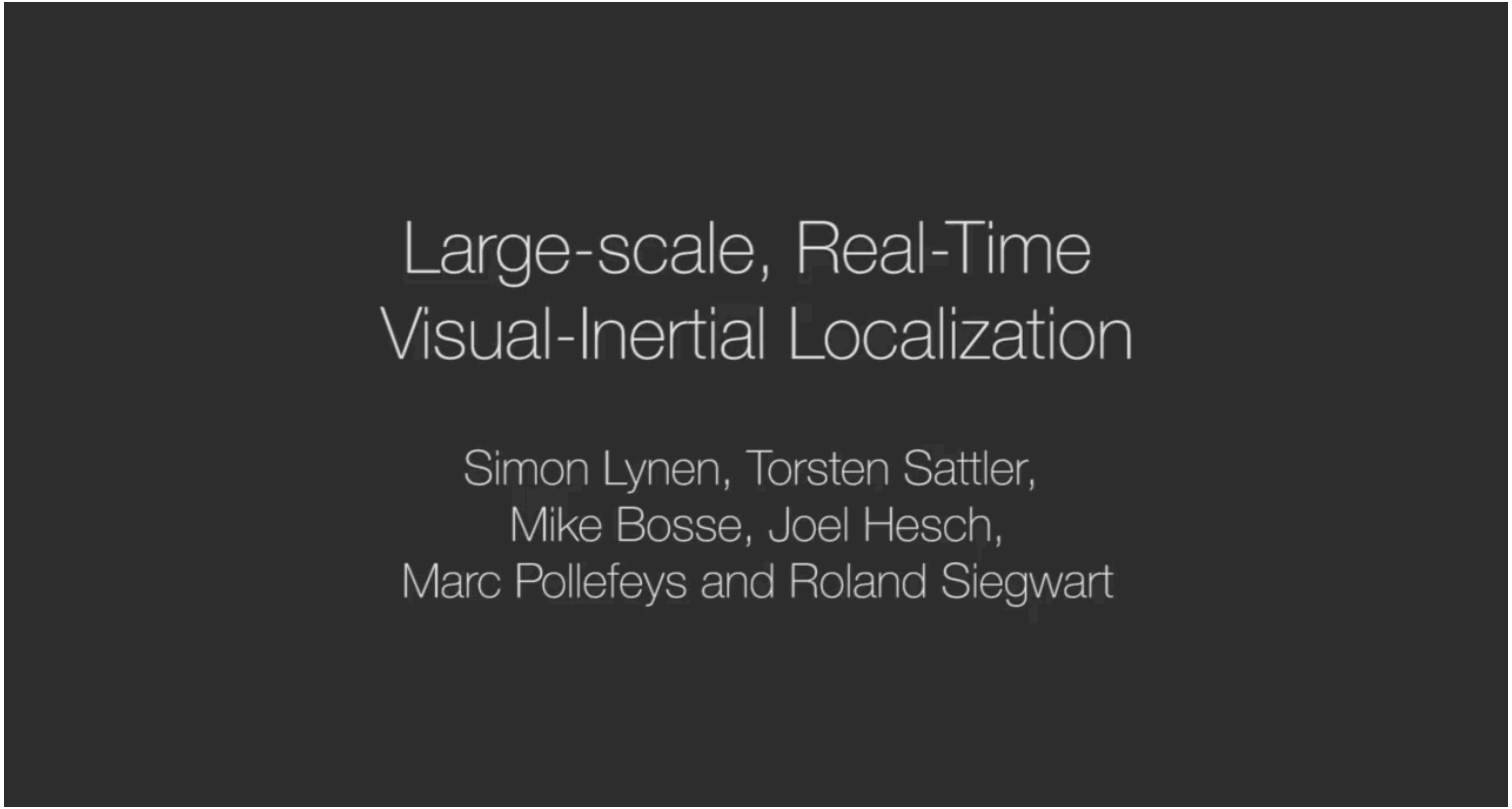
[Middelberg, **Sattler**, Untzelmann, Kobbelt, Scalable 6-DOF Localization on Mobile Devices. ECCV 2014]

Visual Localization



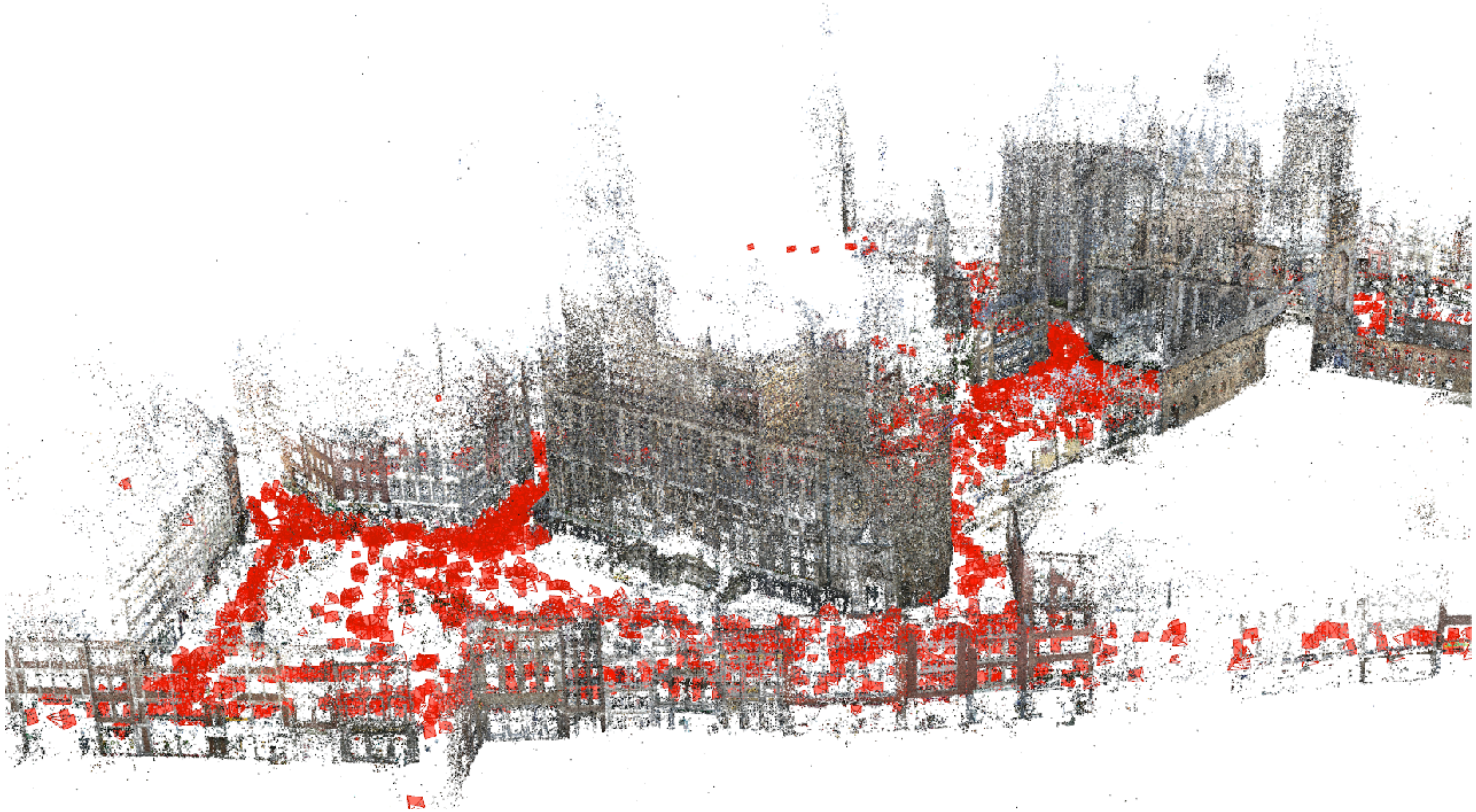
[Lynen, **Sattler**, Bosse, Hesch, Pollefeys, Siegwart, Large-scale Real-Time Visual-Inertial Localization. RSS 2015]

Visual Localization



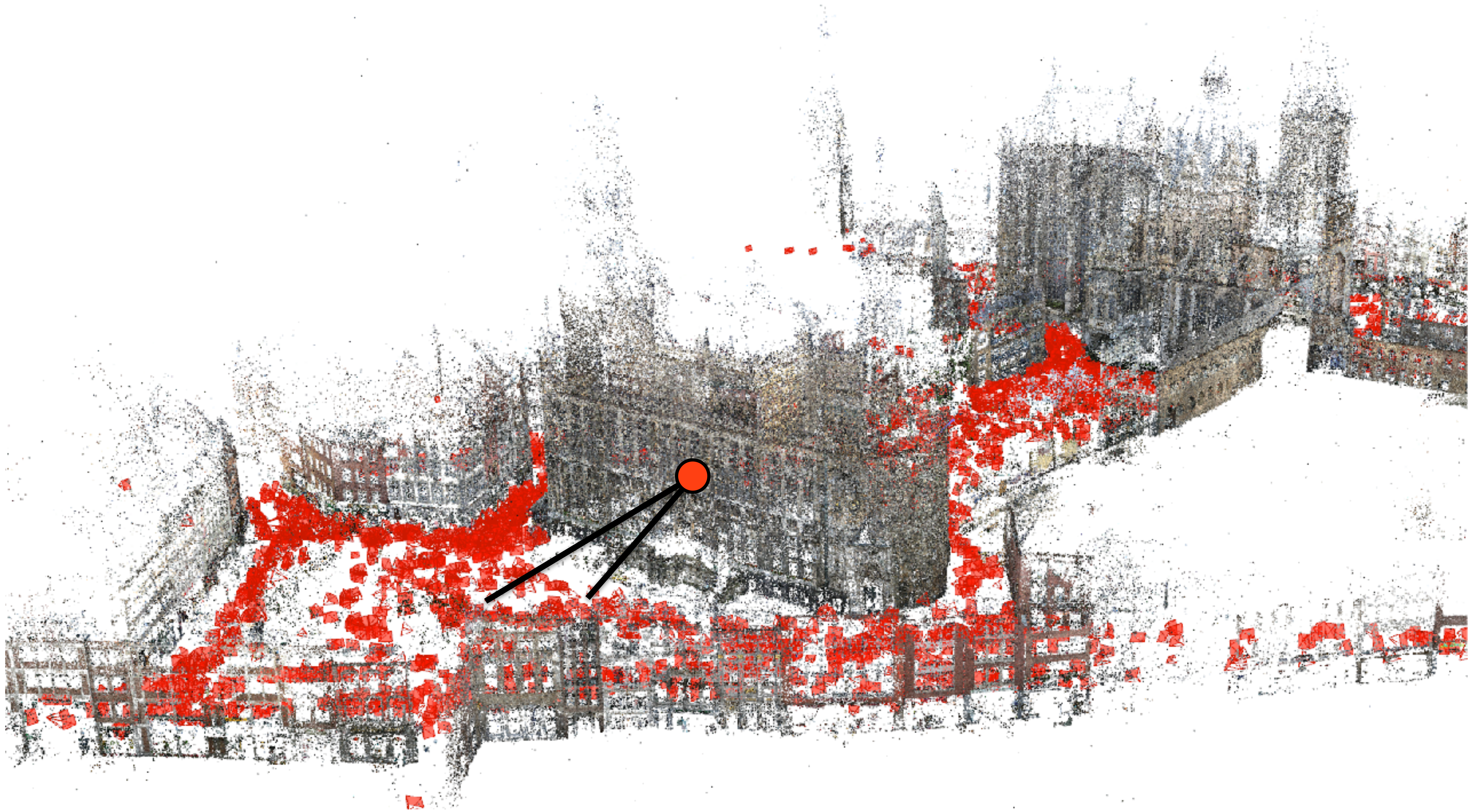
[Lynen, **Sattler**, Bosse, Hesch, Pollefeys, Siegwart, Large-scale Real-Time Visual-Inertial Localization. RSS 2015]

Classic Localization Pipeline



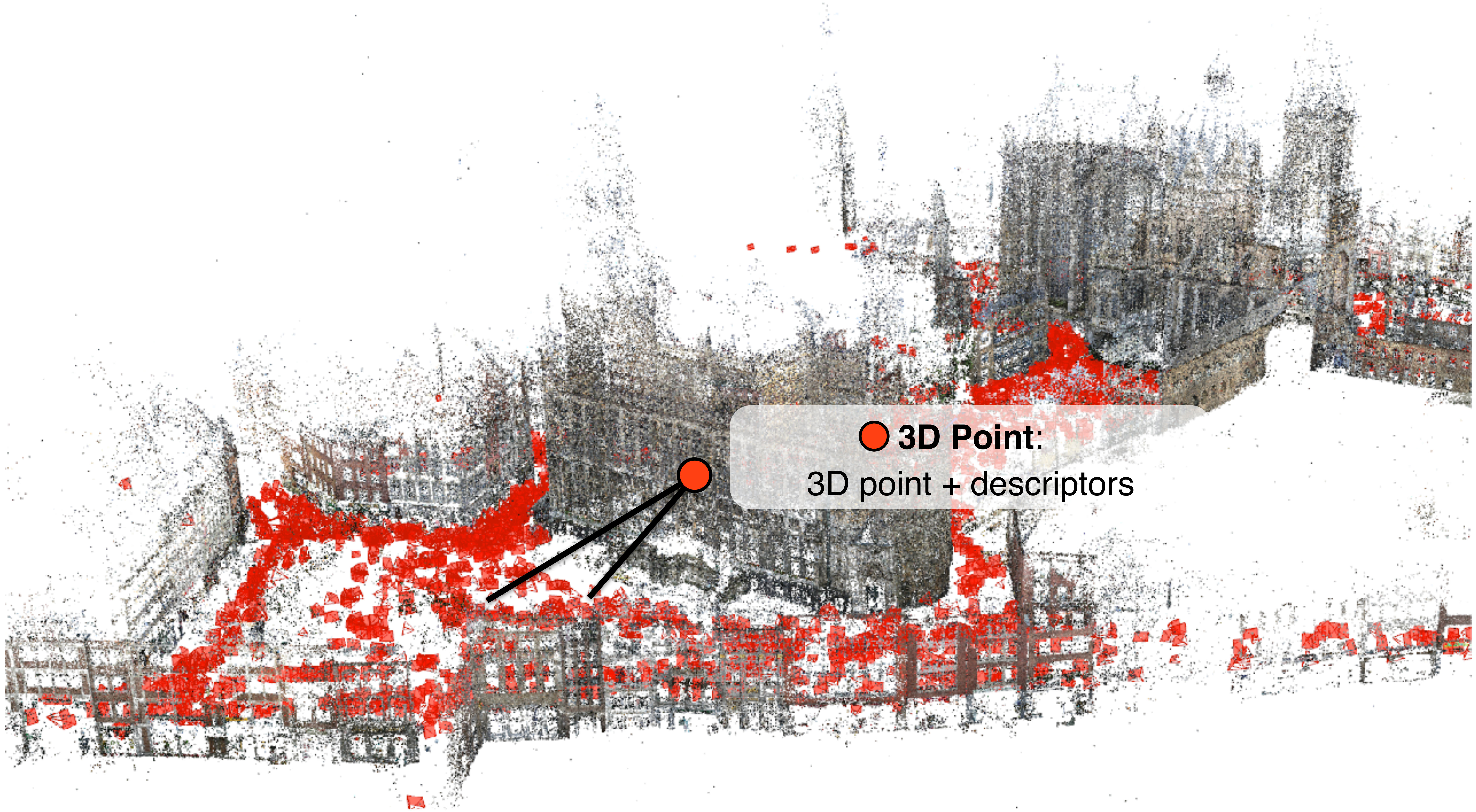
- Offline: Reconstruct scene using Structure-from-Motion

Classic Localization Pipeline



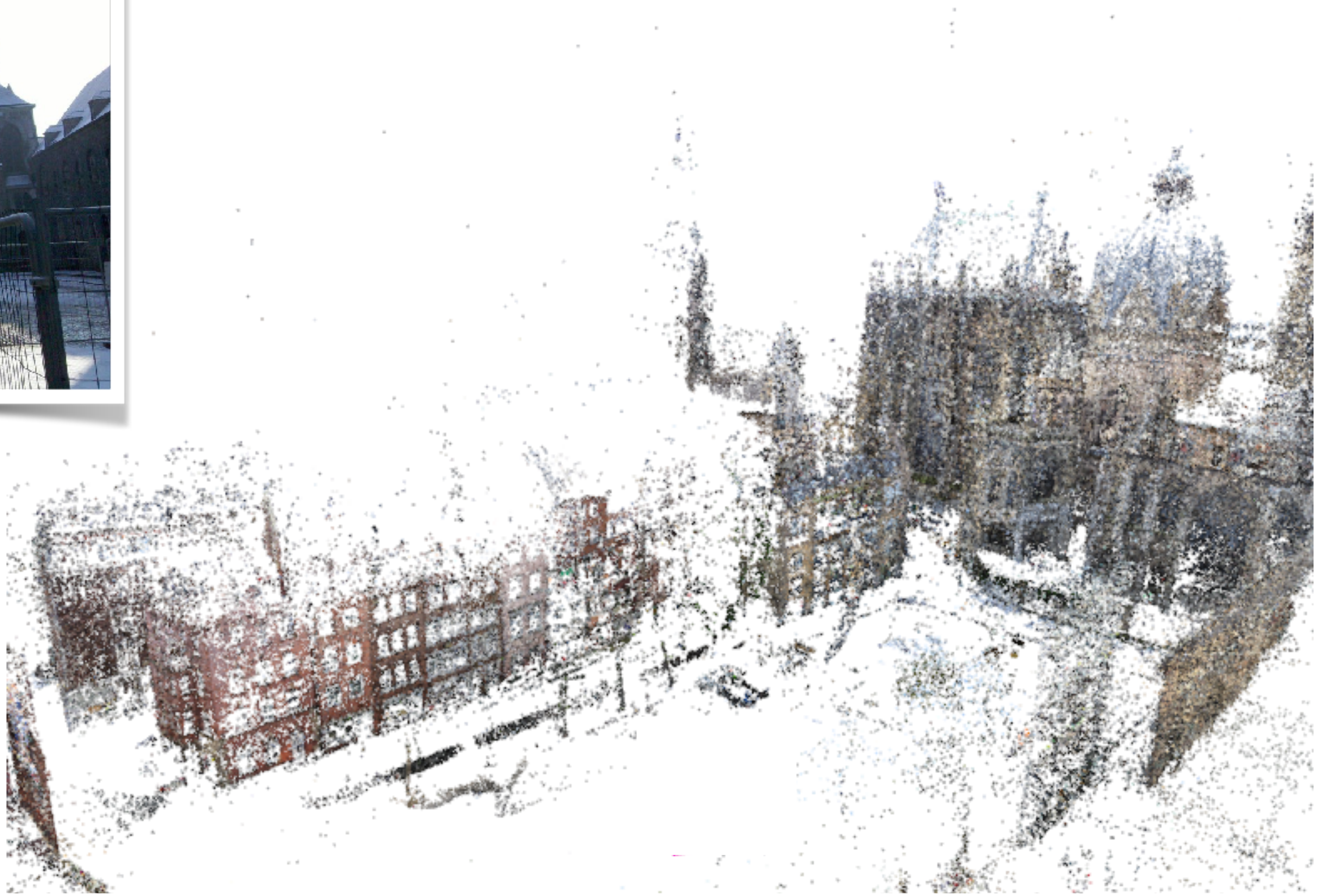
- Offline: Reconstruct scene using Structure-from-Motion

Classic Localization Pipeline

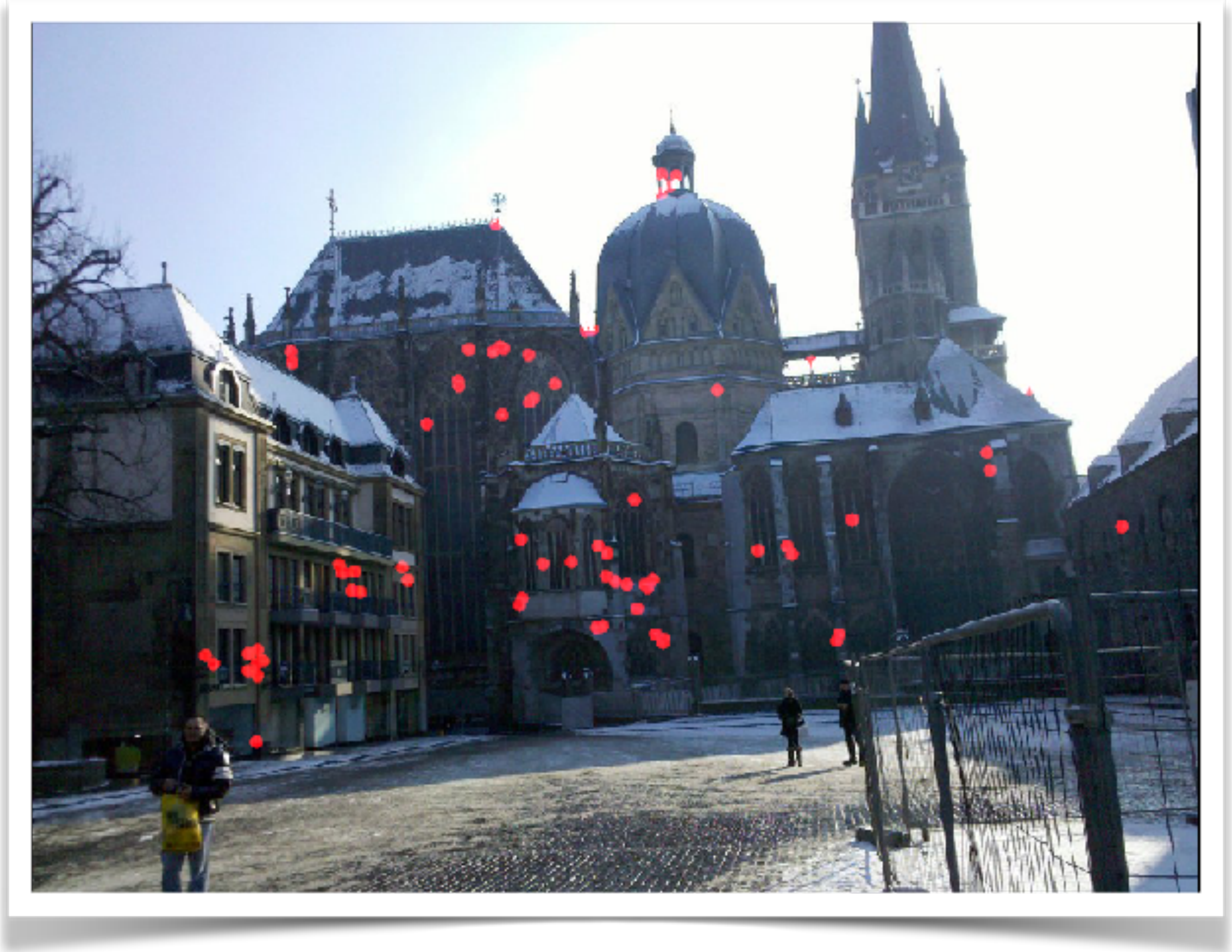


- Offline: Reconstruct scene using Structure-from-Motion
- Associate each 3D point with local image descriptors (SIFT)

Classic Localization Pipeline

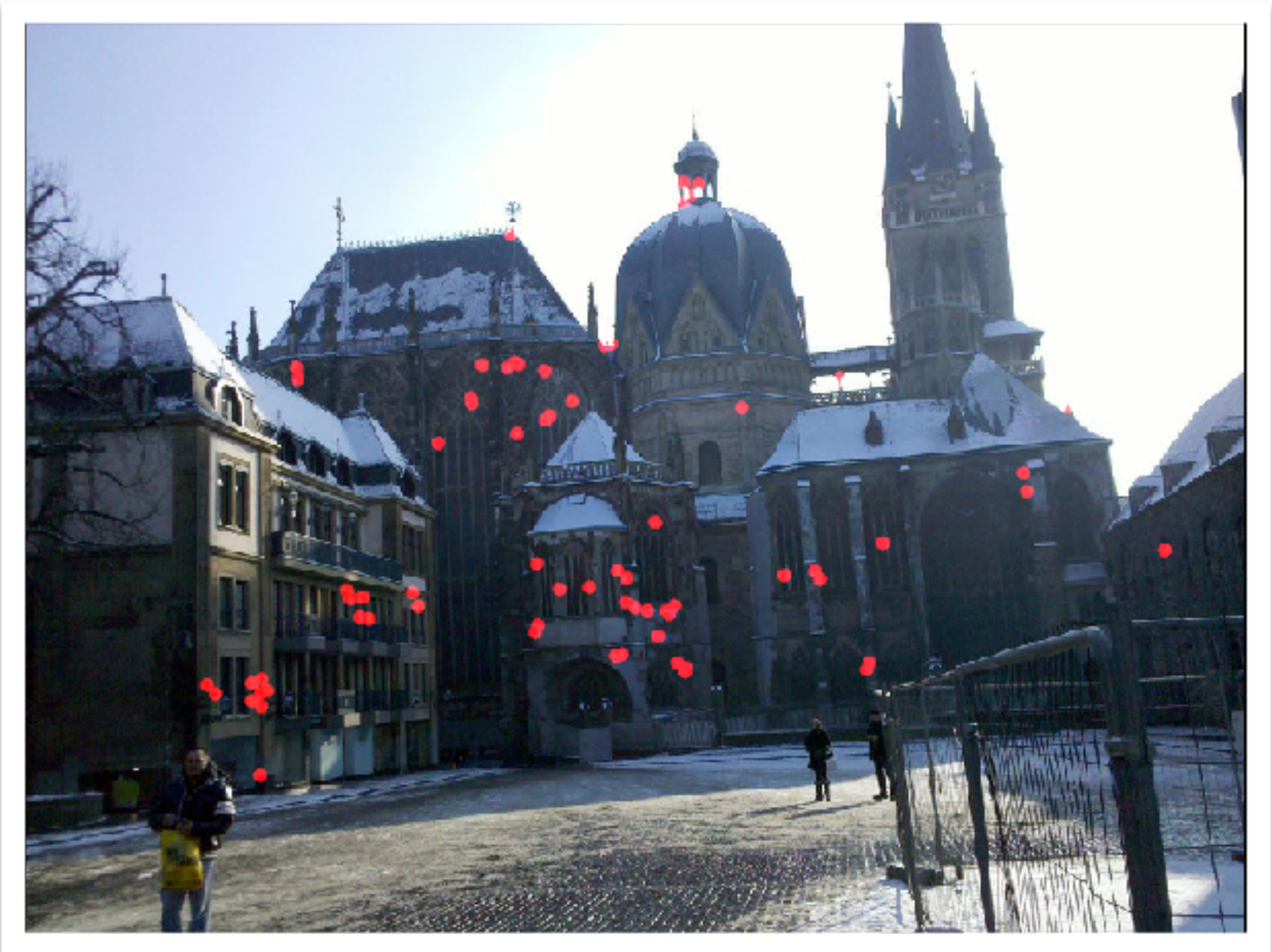


Classic Localization Pipeline



Extract Local Features

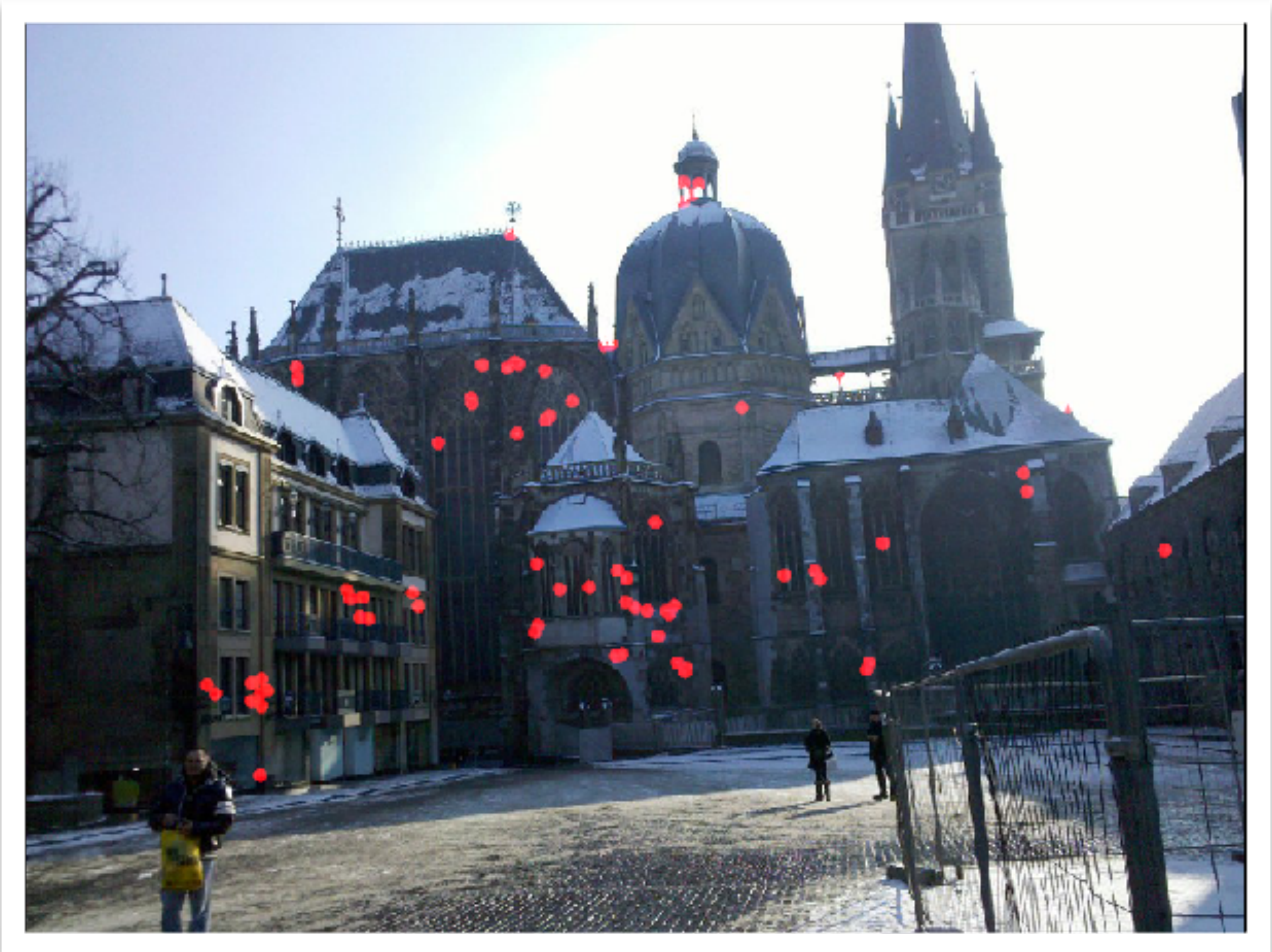
Classic Localization Pipeline

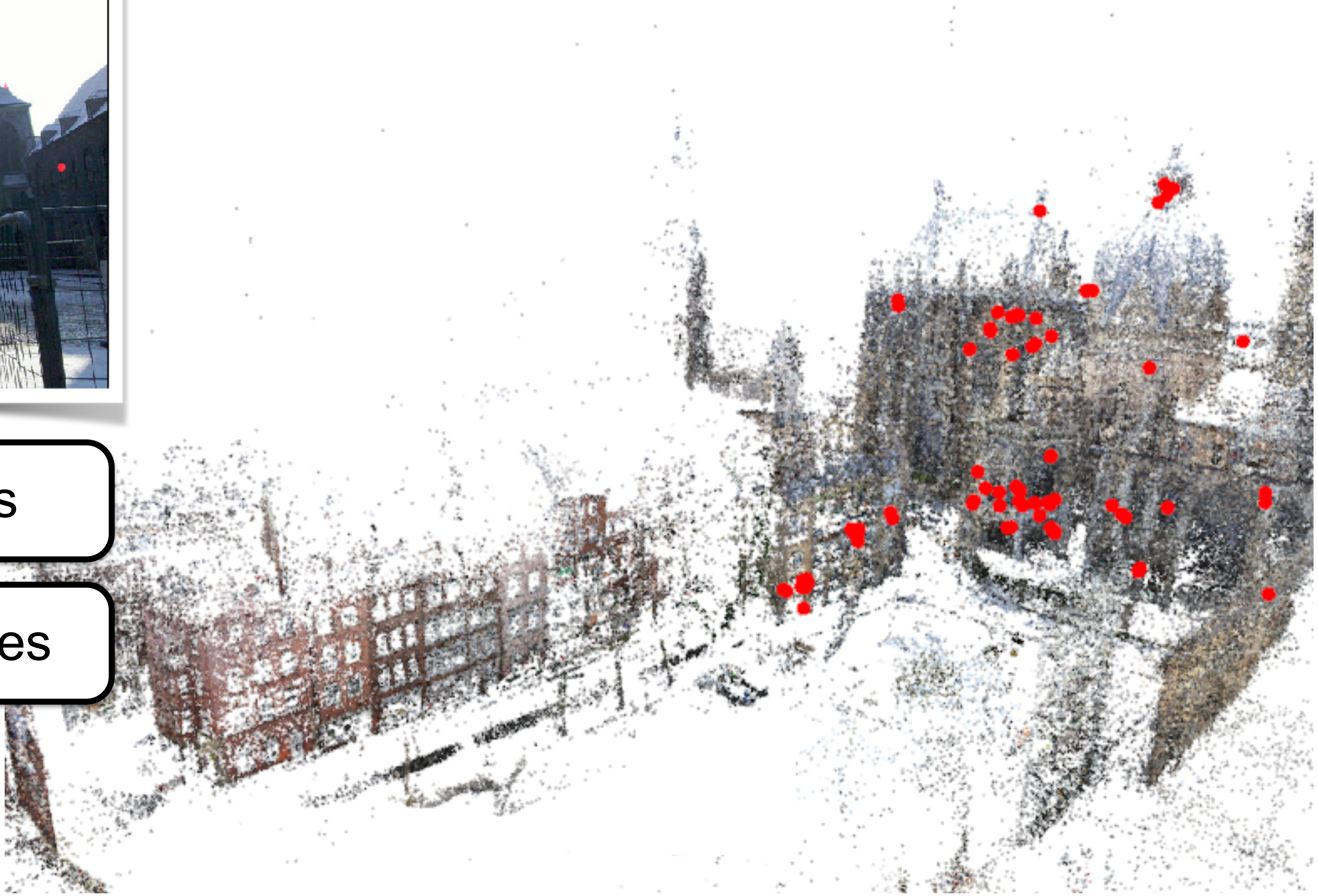


Extract Local Features

Establish 2D-3D Matches

Classic Localization Pipeline





Extract Local Features

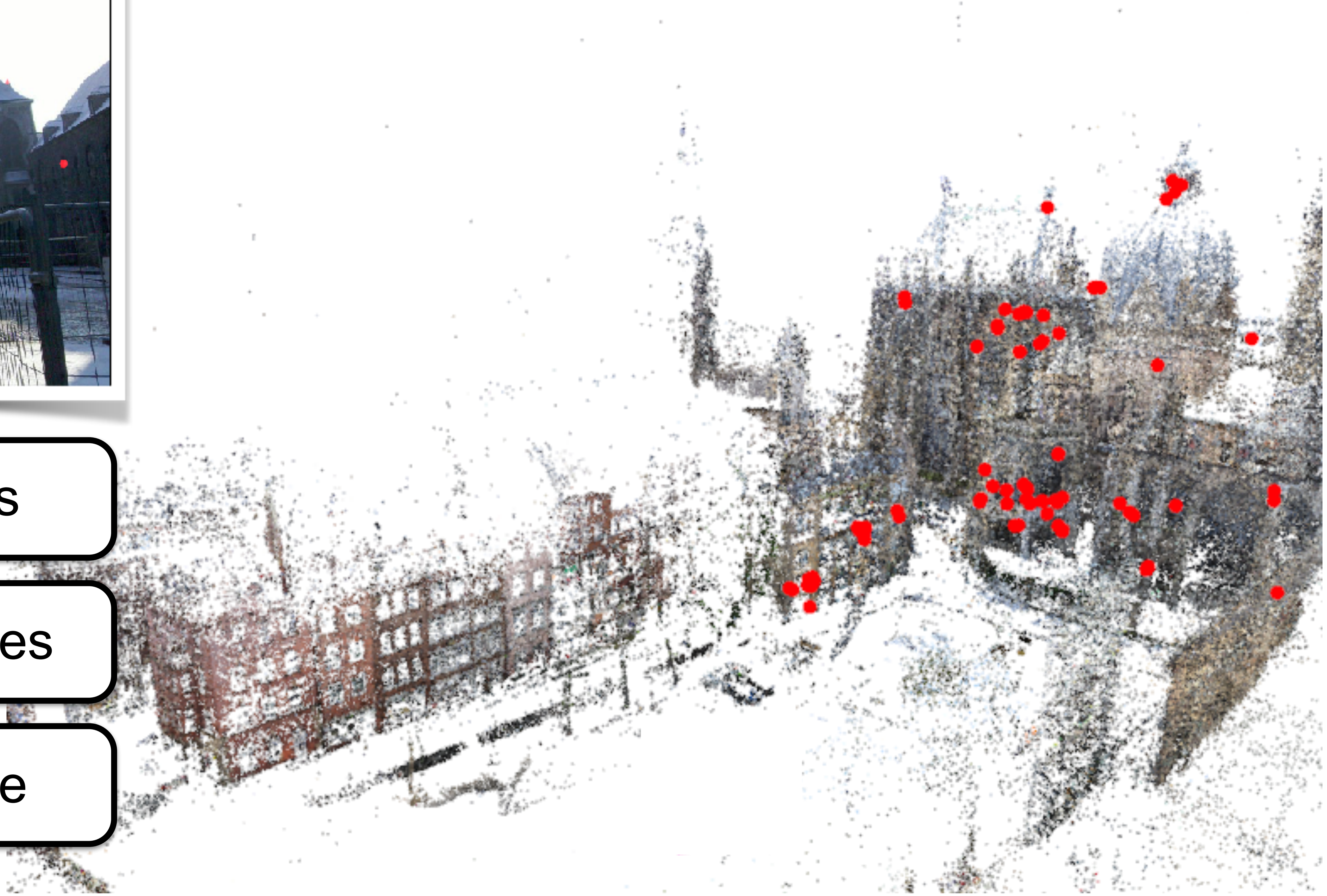
Establish 2D-3D Matches

Classic Localization Pipeline

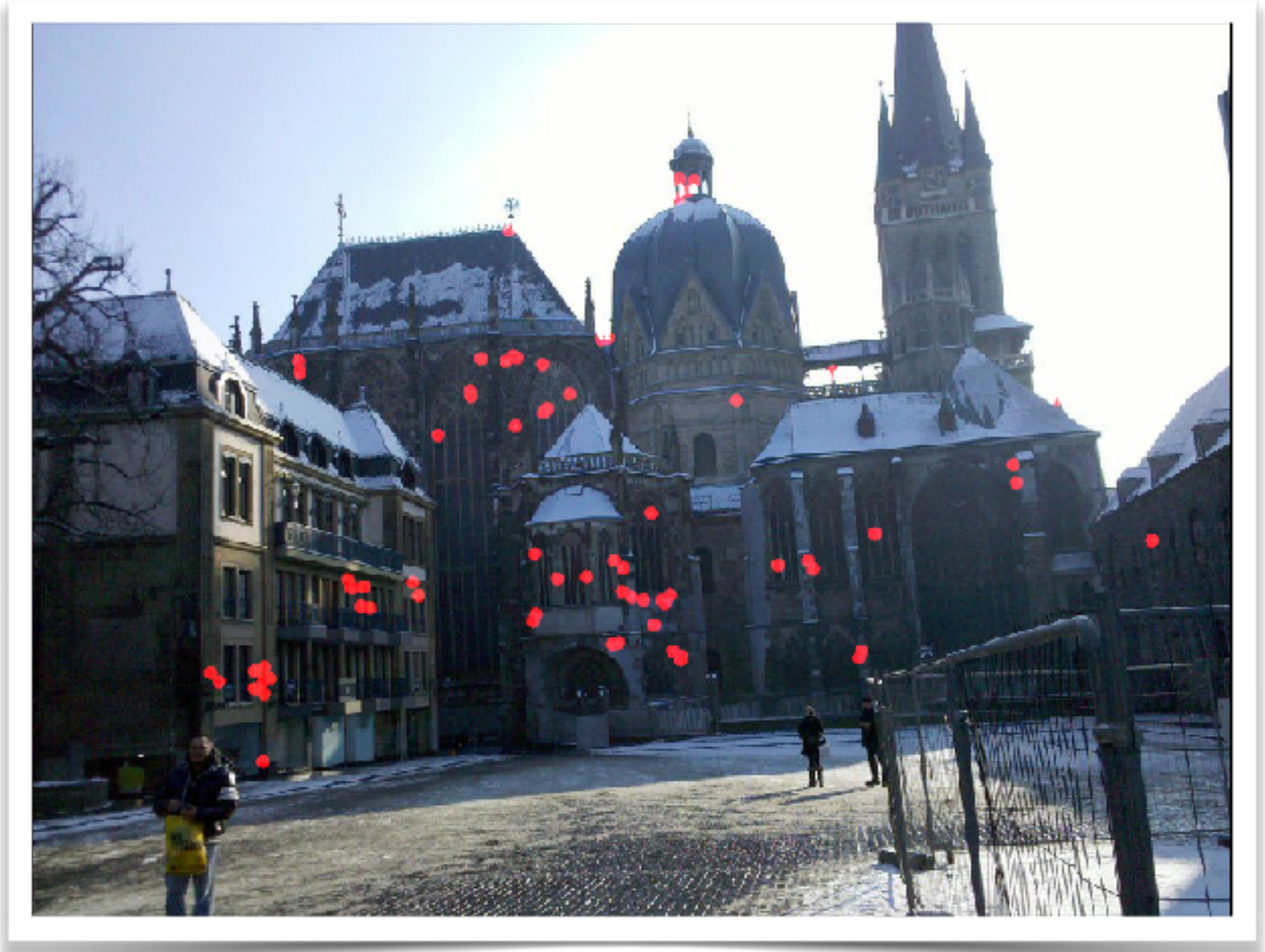
Extract Local Features

Establish 2D-3D Matches

Estimate Camera Pose



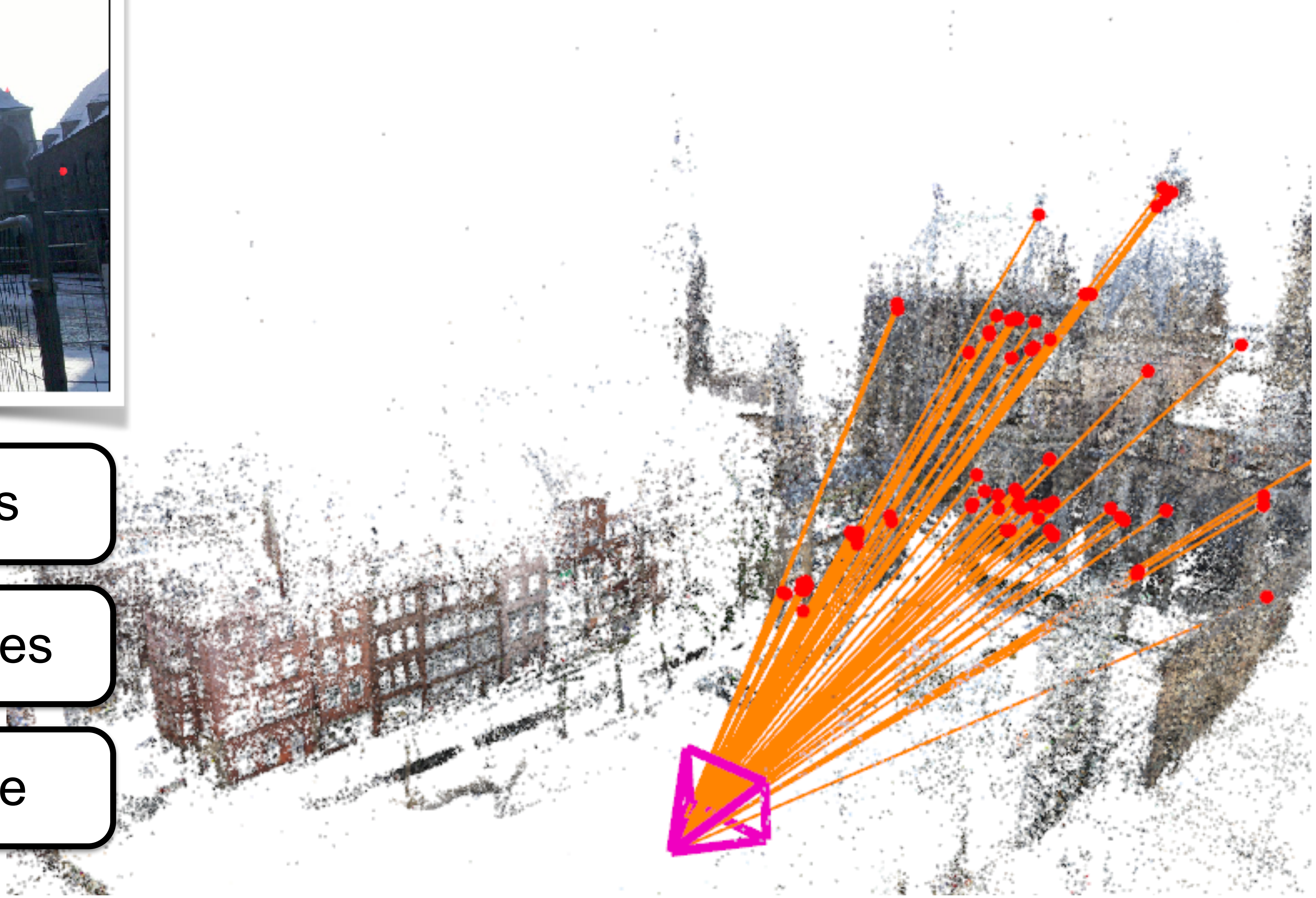
Classic Localization Pipeline



Extract Local Features

Establish 2D-3D Matches

Estimate Camera Pose

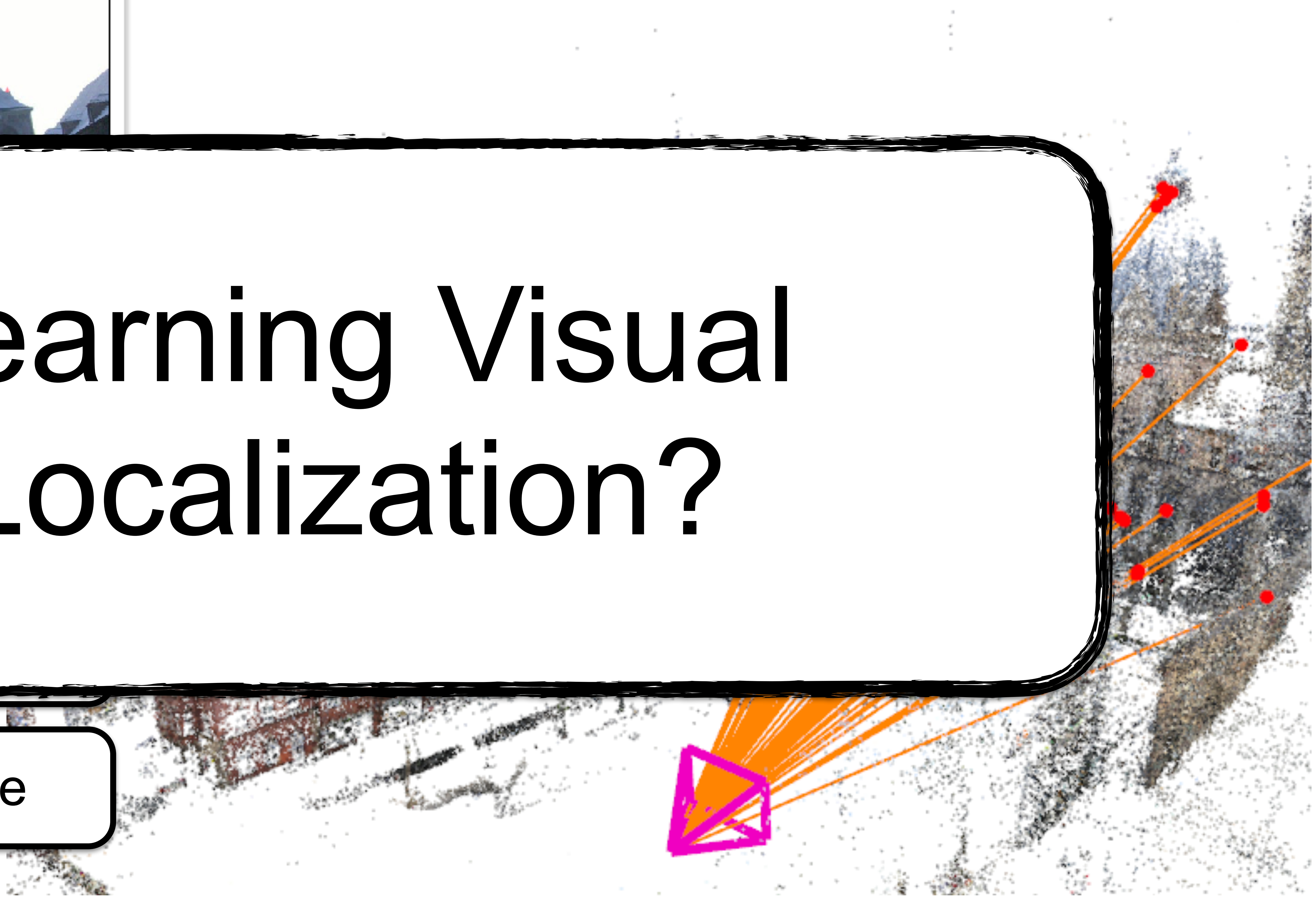


Classic Localization Pipeline

Ext

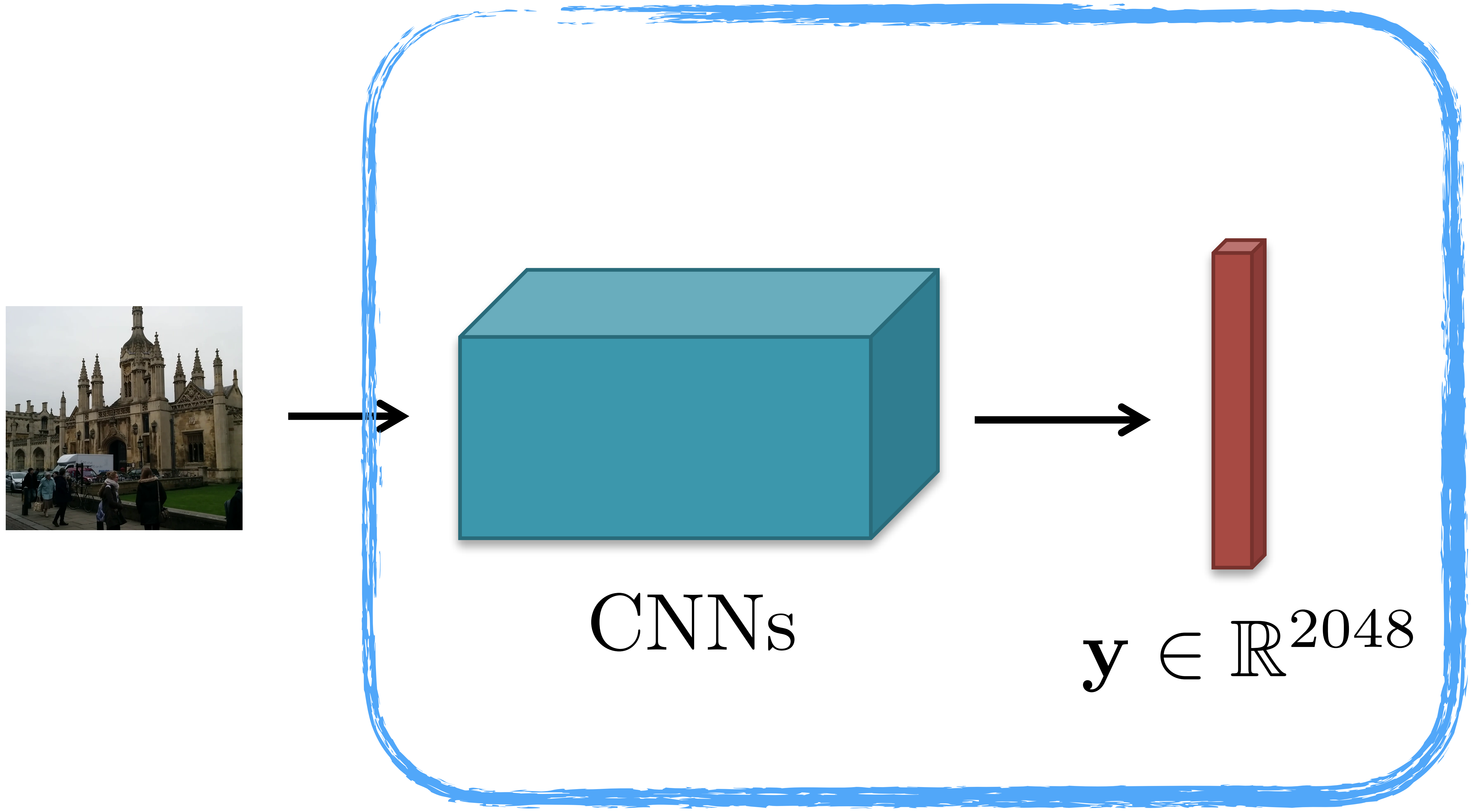
Estab

Estimate Camera Pose



Learning Visual Localization?

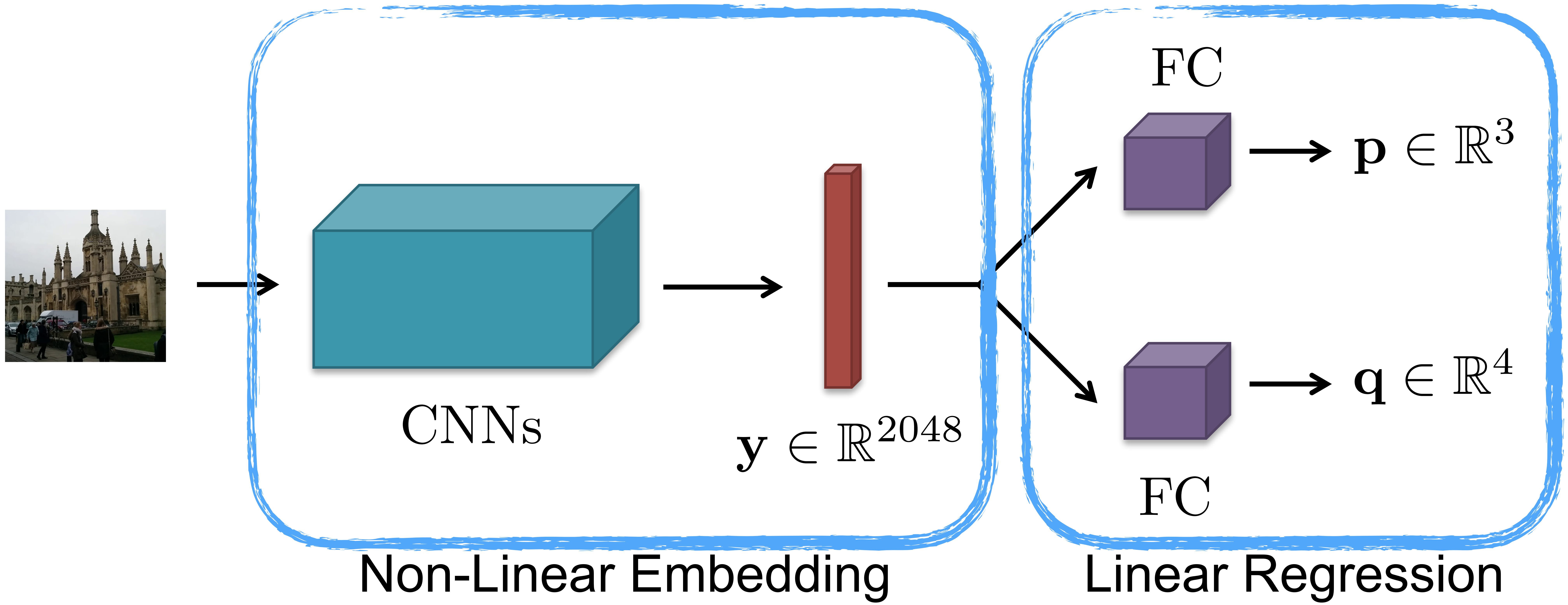
CNN-based Localization (PoseNet)



Non-Linear Embedding

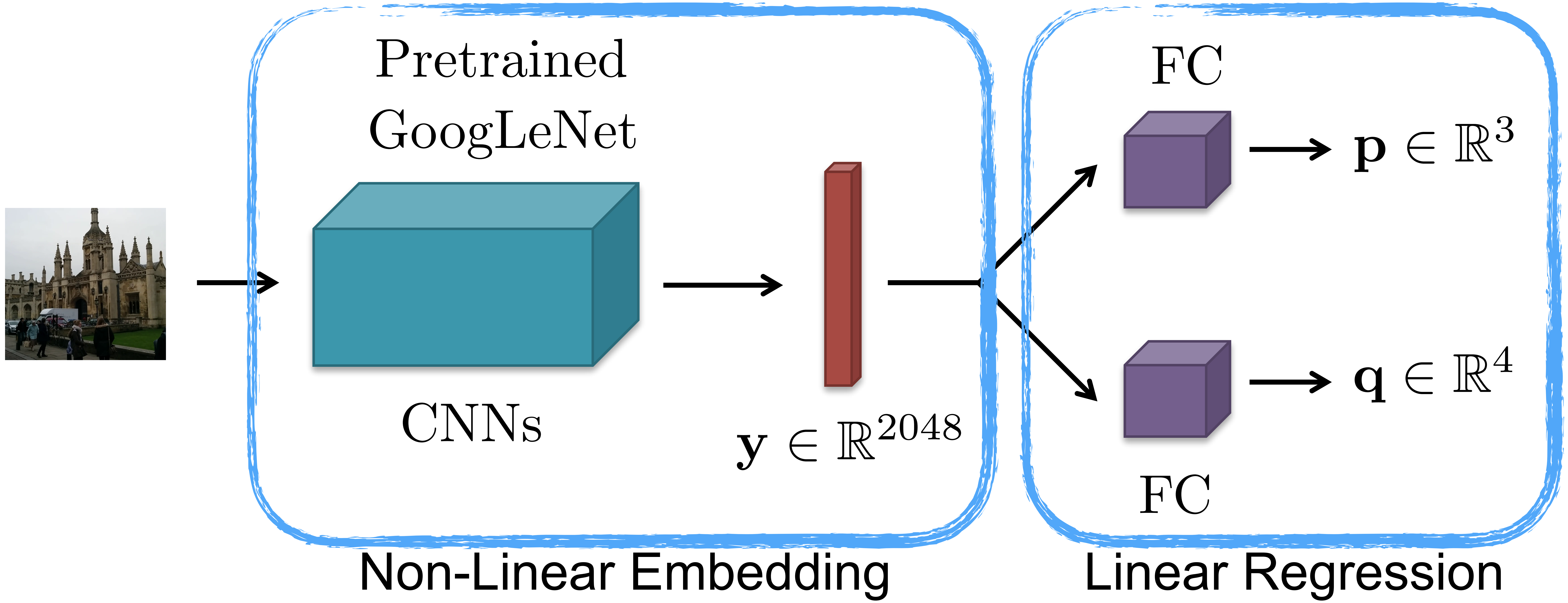
[Kendall, Grimes, Cipola, PoseNet: A convolutional network for real-time 6-dof camera relocalization. ICCV 2015]

CNN-based Localization (PoseNet)



[Kendall, Grimes, Cipola, PoseNet: A convolutional network for real-time 6-dof camera relocalization. ICCV 2015]

CNN-based Localization (PoseNet)



[Kendall, Grimes, Cipola, PoseNet: A convolutional network for real-time 6-dof camera relocalization. ICCV 2015]

CNN-based Localization

[Walch, Hazirbas, Leal-Taixé, **Sattler**, Hilsenbeck, Cremers, Image-based localization using LSTMs for structured feature correlation. ICCV, 2017]

CNN-based Localization

[Walch, Hazirbas, Leal-Taixé, **Sattler**, Hilsenbeck, Cremers, Image-based localization using LSTMs for structured feature correlation. ICCV, 2017]

Training PoseNet

- Input: Images I_i with known 6DOF camera pose $(\hat{\mathbf{c}}_i, \hat{\mathbf{q}}_i)$

[Kendall, Grimes, Cipola, PoseNet: A convolutional network for real-time 6-dof camera relocalization. ICCV 2015]
[Kendall, Cipola, Geometric loss functions for camera pose regression with deep learning. CVPR 2017]

Training PoseNet

- Input: Images I_i with known 6DOF camera pose $(\hat{\mathbf{c}}_i, \hat{\mathbf{q}}_i)$
- Non-geometric loss function:

$$L_i = \|\mathbf{c}_i - \hat{\mathbf{c}}_i\|_2 + \beta \cdot \left\| \mathbf{q}_i - \frac{\hat{\mathbf{q}}_i}{\|\hat{\mathbf{q}}_i\|} \right\|_2$$

[Kendall, Grimes, Cipola, PoseNet: A convolutional network for real-time 6-dof camera relocalization. ICCV 2015]
[Kendall, Cipola, Geometric loss functions for camera pose regression with deep learning. CVPR 2017]

Training PoseNet

- Input: Images I_i with known 6DOF camera pose $(\hat{\mathbf{c}}_i, \hat{\mathbf{q}}_i)$
- Non-geometric loss function:

$$L_i = \|\mathbf{c}_i - \hat{\mathbf{c}}_i\|_2 + \beta \cdot \left\| \mathbf{q}_i - \frac{\hat{\mathbf{q}}_i}{\|\hat{\mathbf{q}}_i\|} \right\|_2$$

- Geometric loss function: Minimize re-projection error of 3D points visible in image

[Kendall, Grimes, Cipola, PoseNet: A convolutional network for real-time 6-dof camera relocalization. ICCV 2015]
[Kendall, Cipola, Geometric loss functions for camera pose regression with deep learning. CVPR 2017]

Deep vs. Classical Localization

Measure: Median position [m] / orientation [deg] error

[Walch, Hazirbas, Leal-Taixé, **Sattler**, Hilsenbeck, Cremers, Image-based localization using LSTMs for structured feature correlation. ICCV, 2017]

Deep vs. Classical Localization

Measure: Median position [m] / orientation [deg] error

original PoseNet	1.92m 5.40°	2.31m 5.38°	1.46m 8.08°	2.65m 8.48°	0.32m 8.12°	0.47m 14.4°	0.29m 12.0°	0.48m 7.68°	0.47m 8.42°	0.59m 8.64°	0.47m 13.8°

Cambridge Landmarks (outdoor)

7 Scenes (indoor)

[Walch, Hazirbas, Leal-Taixé, **Sattler**, Hilsenbeck, Cremers, Image-based localization using LSTMs for structured feature correlation. ICCV, 2017]

Deep vs. Classical Localization

Measure: Median position [m] / orientation [deg] error

original PoseNet	1.92m 5.40°	2.31m 5.38°	1.46m 8.08°	2.65m 8.48°	0.32m 8.12°	0.47m 14.4°	0.29m 12.0°	0.48m 7.68°	0.47m 8.42°	0.59m 8.64°	0.47m 13.8°
PoseNet + LSTM	0.99m 3.65°	1.51m 4.29°	1.18m 7.44°	1.52m 6.68°	0.24m 5.77°	0.34m 11.9°	0.21m 13.7°	0.30m 8.08°	0.33m 7.00°	0.37m 8.83°	0.40m 13.7°

Cambridge Landmarks
(outdoor)

7 Scenes
(indoor)

[Walch, Hazirbas, Leal-Taixé, **Sattler**, Hilsenbeck, Cremers, Image-based localization using LSTMs for structured feature correlation. ICCV, 2017]

Deep vs. Classical Localization

Measure: Median position [m] / orientation [deg] error

original PoseNet	1.92m 5.40°	2.31m 5.38°	1.46m 8.08°	2.65m 8.48°	0.32m 8.12°	0.47m 14.4°	0.29m 12.0°	0.48m 7.68°	0.47m 8.42°	0.59m 8.64°	0.47m 13.8°
PoseNet + LSTM	0.99m 3.65°	1.51m 4.29°	1.18m 7.44°	1.52m 6.68°	0.24m 5.77°	0.34m 11.9°	0.21m 13.7°	0.30m 8.08°	0.33m 7.00°	0.37m 8.83°	0.40m 13.7°
PoseNet + geometric loss	0.88m 1.04°	3.20m 3.29°	0.88m 3.78°	1.57m 3.32°	0.13m 4.48°	0.27m 11.3°	0.17m 13.0°	0.19m 5.55°	0.26m 4.75°	0.23m 5.35°	0.35m 12.4°

Cambridge Landmarks
(outdoor)

7 Scenes
(indoor)

[Walch, Hazirbas, Leal-Taixé, **Sattler**, Hilsenbeck, Cremers, Image-based localization using LSTMs for structured feature correlation. ICCV, 2017]

Deep vs. Classical Localization

Measure: Median position [m] / orientation [deg] error

original PoseNet	1.92m 5.40°	2.31m 5.38°	1.46m 8.08°	2.65m 8.48°	0.32m 8.12°	0.47m 14.4°	0.29m 12.0°	0.48m 7.68°	0.47m 8.42°	0.59m 8.64°	0.47m 13.8°
PoseNet + LSTM	0.99m 3.65°	1.51m 4.29°	1.18m 7.44°	1.52m 6.68°	0.24m 5.77°	0.34m 11.9°	0.21m 13.7°	0.30m 8.08°	0.33m 7.00°	0.37m 8.83°	0.40m 13.7°
PoseNet + geometric loss	0.88m 1.04°	3.20m 3.29°	0.88m 3.78°	1.57m 3.32°	0.13m 4.48°	0.27m 11.3°	0.17m 13.0°	0.19m 5.55°	0.26m 4.75°	0.23m 5.35°	0.35m 12.4°
[Sattler et al., PAMI 2017]	**0.42**m **0.55**°	**0.44**m **1.01**°	**0.12**m **0.40**°	**0.19**m **0.54**°	**0.04**m **1.96**°	**0.03**m **1.53**°	**0.02**m **1.45**°	**0.09**m **3.61**°	**0.08**m **3.20**°	**0.07**m **3.37**°	**0.03**m **2.22**°

Cambridge Landmarks (outdoor) — first four result columns; 7 Scenes (indoor) — last seven result columns

[Walch, Hazirbas, Leal-Taixé, **Sattler**, Hilsenbeck, Cremers, Image-based localization using LSTMs for structured feature correlation. ICCV, 2017]

Deep vs. Classical Localization

Results on Dubrovnik dataset:

	Quantile Errors [m]		
	25%	50%	75%
PoseNet + geometric loss	-	7.9	-

[**Sattler**, Torii, Sivic, Pollefeys, Taira, Okutomi, Pajdla,Are Large-Scale 3D Models Really Necessary for Accurate Visual Localization? CVPR 2017]

Deep vs. Classical Localization

Results on Dubrovnik dataset:

	Quantile Errors [m]		
	25%	50%	75%
PoseNet + geometric loss	-	7.9	-
Image Retrieval (**No Pose Estimation**)	0.9	2.9	9.0

[**Sattler**, Torii, Sivic, Pollefeys, Taira, Okutomi, Pajdla,Are Large-Scale 3D Models Really Necessary for Accurate Visual Localization? CVPR 2017]

Deep vs. Classical Localization

Results on Dubrovnik dataset:

	Quantile Errors [m]		
	25%	50%	75%
PoseNet + geometric loss	-	7.9	-
Image Retrieval (**No Pose Estimation**)	0.9	2.9	9.0
[Sattler et al., PAMI 2017]	0.5	1.3	5.0

[**Sattler**, Torii, Sivic, Pollefeys, Taira, Okutomi, Pajdla,Are Large-Scale 3D Models Really Necessary for Accurate Visual Localization? CVPR 2017]

Deep vs. Classical Localization

Results on Dubrovnik dataset:

	Quantile Errors [m]		
	25%	50%	75%
PoseNet + geometric loss	-	7.9	-
Image Retrieval (**No Pose Estimation**)	0.9	2.9	9.0
[Sattler et al., PAMI 2017]	0.5	1.3	5.0
[Zeisl et al., ICCV 2015]	0.2	0.6	2.1

[**Sattler**, Torii, Sivic, Pollefeys, Taira, Okutomi, Pajdla,Are Large-Scale 3D Models Really Necessary for Accurate Visual Localization? CVPR 2017]

A Hard Example

[Walch, Hazirbas, Leal-Taixé, **Sattler**, Hilsenbeck, Cremers, Image-based localization using LSTMs for structured feature correlation. ICCV, 2017]

A Hard Example

original PoseNet	1.87m, 6.14○
PoseNet + LSTM	**1.31**m, **2.79**○
[Sattler et al., PAMI 2017]	SfM failed

[Walch, Hazirbas, Leal-Taixé, **Sattler**, Hilsenbeck, Cremers, Image-based localization using LSTMs for structured feature correlation. ICCV, 2017]

My Take

- PoseNet + variants learn mapping from visual appearance to 6D pose space

My Take

- PoseNet + variants learn mapping from visual appearance to 6D pose space
- In theory: Possible to learn camera pose regression (for known camera intrinsics)

My Take

- PoseNet + variants learn mapping from visual appearance to 6D pose space
- In theory: Possible to learn camera pose regression (for known camera intrinsics)
- In practice: Probably not enough training data to learn mapping that generalizes away from training data

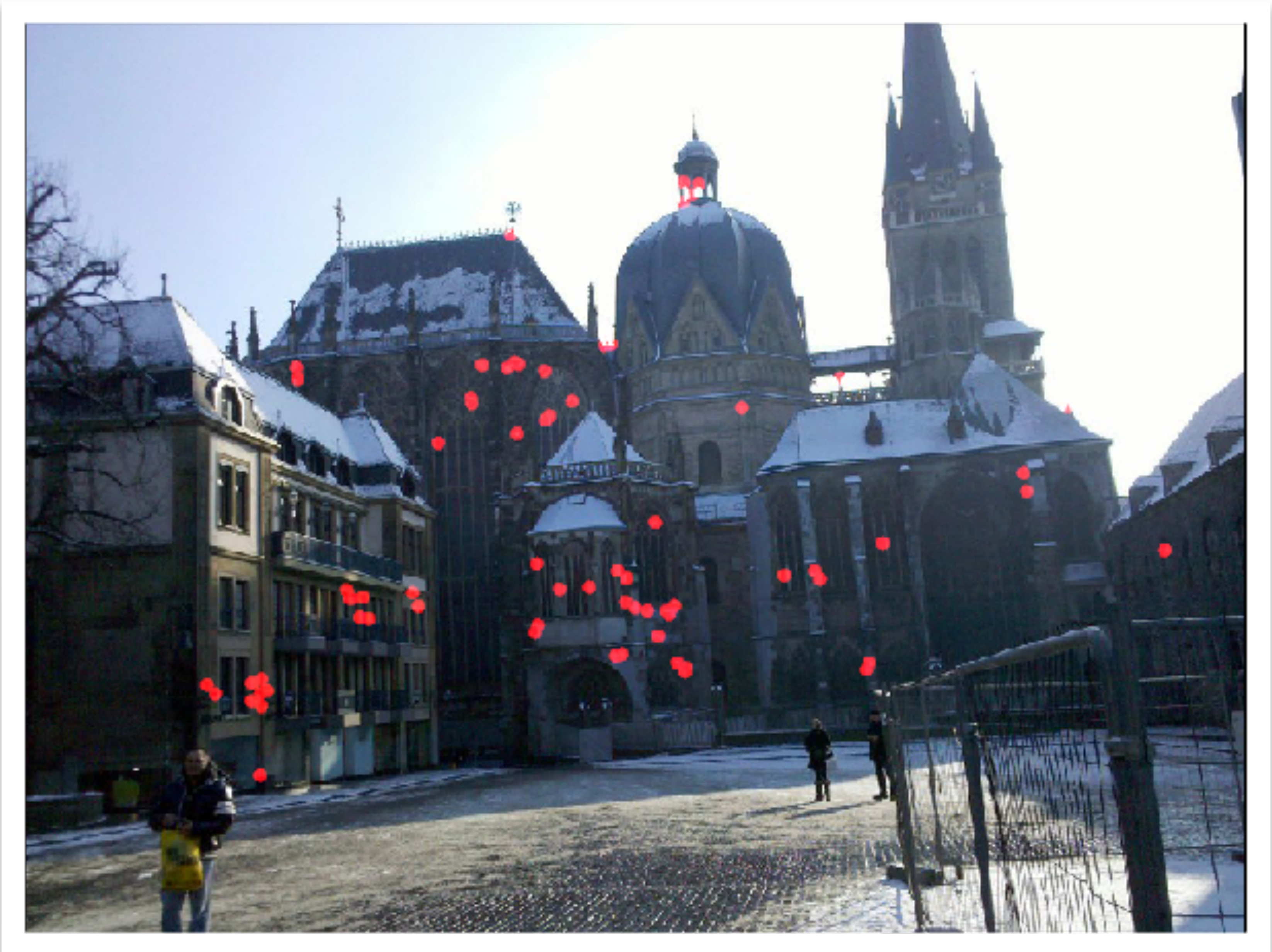
My Take

- PoseNet + variants learn mapping from visual appearance to 6D pose space
- In theory: Possible to learn camera pose regression (for known camera intrinsics)
- In practice: Probably not enough training data to learn mapping that generalizes away from training data
- Promising results for hard scenes in which feature-based approaches fail

My Take

- PoseNet + variants learn mapping from visual appearance to 6D pose space
- In theory: Possible to learn camera pose regression (for known camera intrinsics)
- In practice: Probably not enough training data to learn mapping that generalizes away from training data
- Promising results for hard scenes in which feature-based approaches fail
- Why learn full pose estimation pipeline?

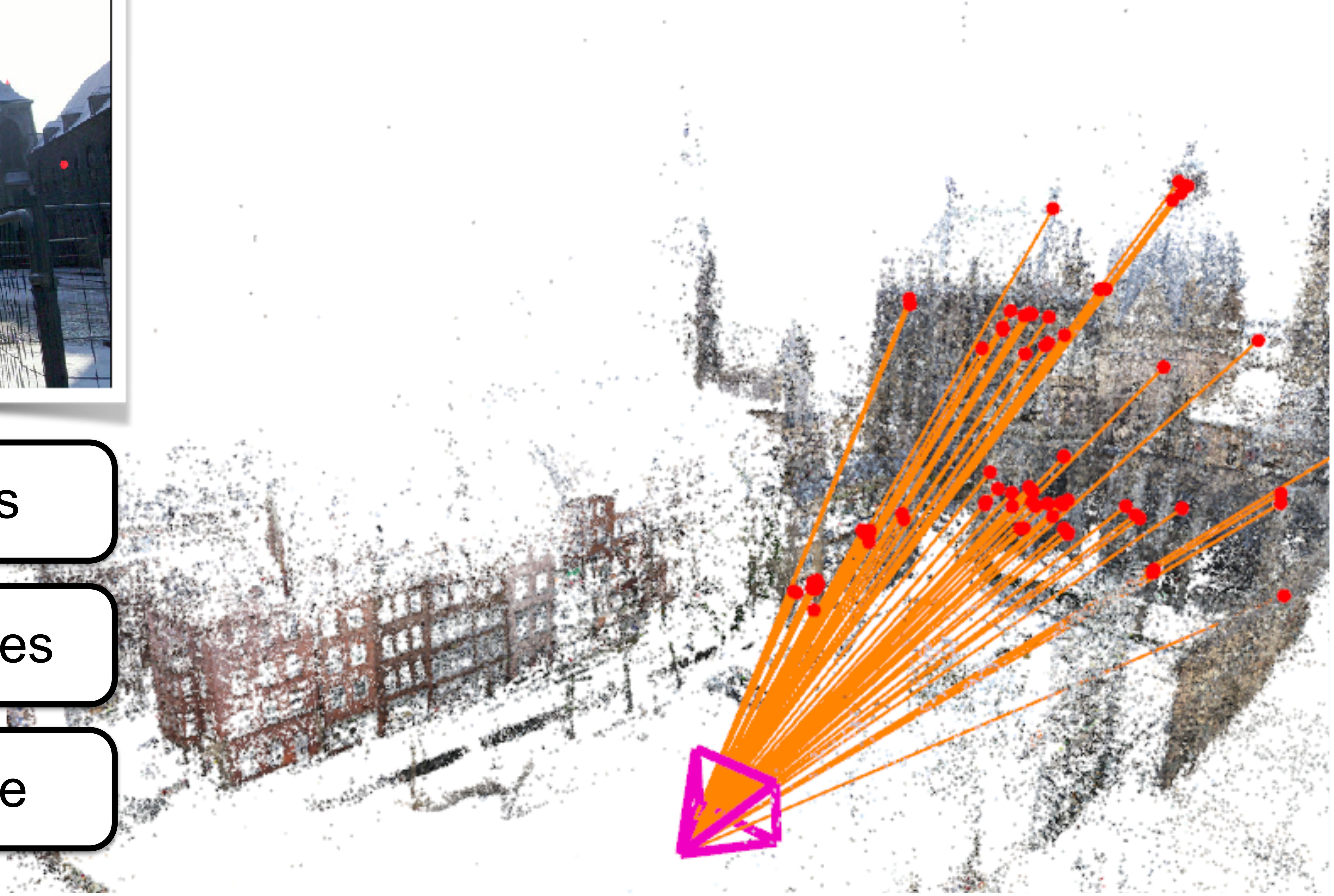
Classic Localization Pipeline



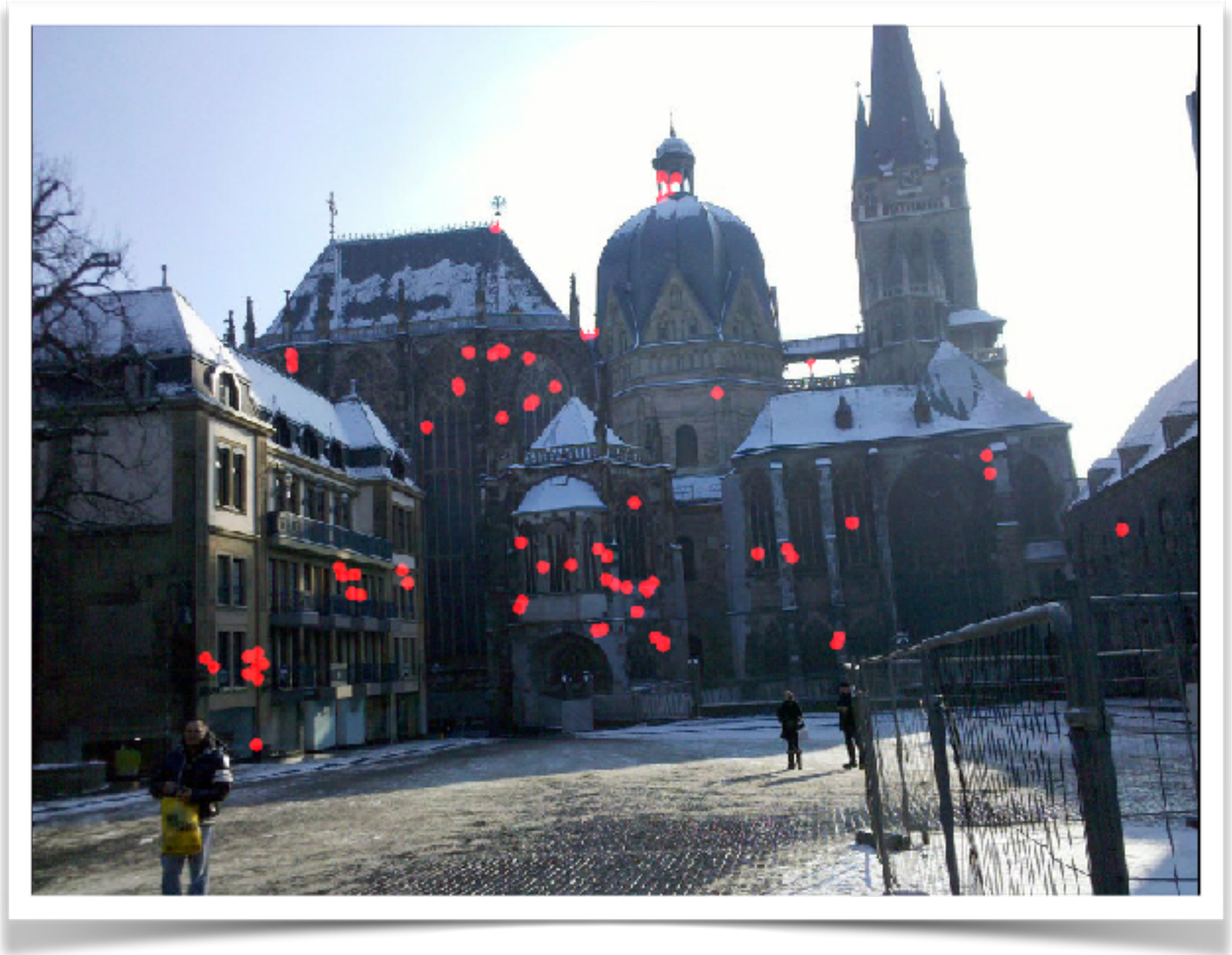
Extract Local Features

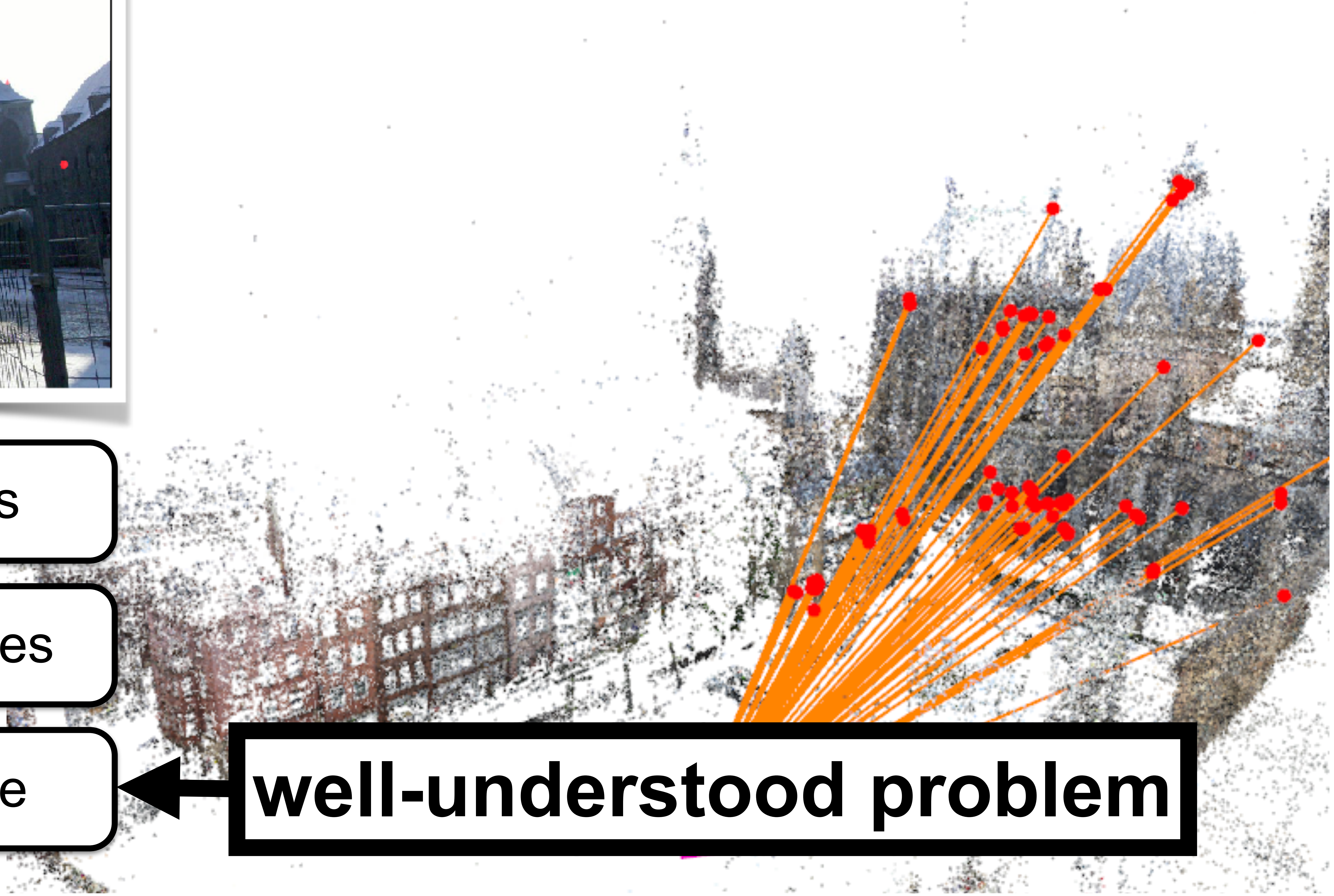
Establish 2D-3D Matches

Estimate Camera Pose



Classic Localization Pipeline





Extract Local Features

Establish 2D-3D Matches

Estimate Camera Pose

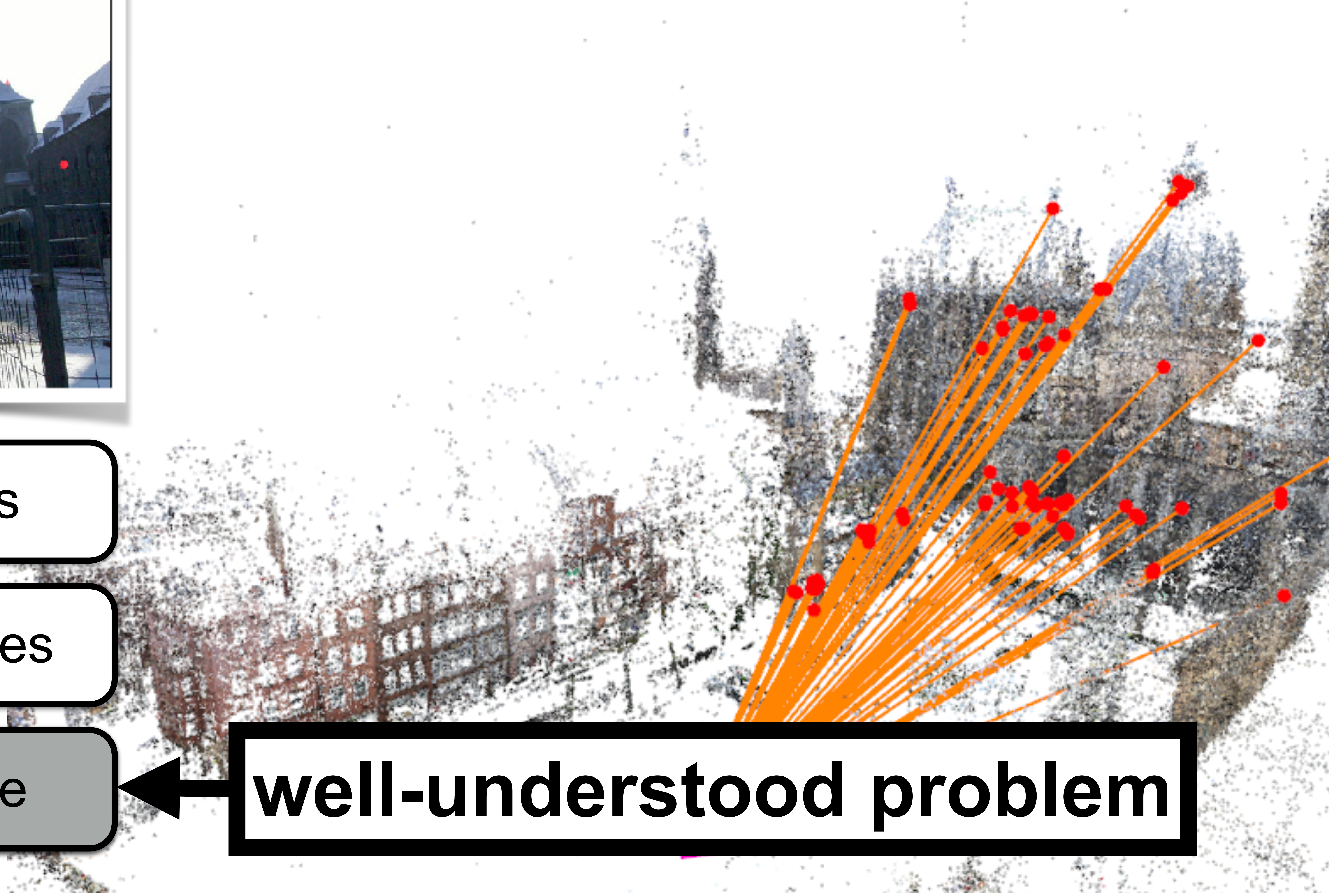
well-understood problem

Classic Localization Pipeline

Extract Local Features

Establish 2D-3D Matches

Estimate Camera Pose



well-understood problem

Classic Localization Pipeline

Extract Local Features

Establish 2D-3D Matches ← **nearest neighbor search**

Estimate Camera Pose ← **well-understood problem**

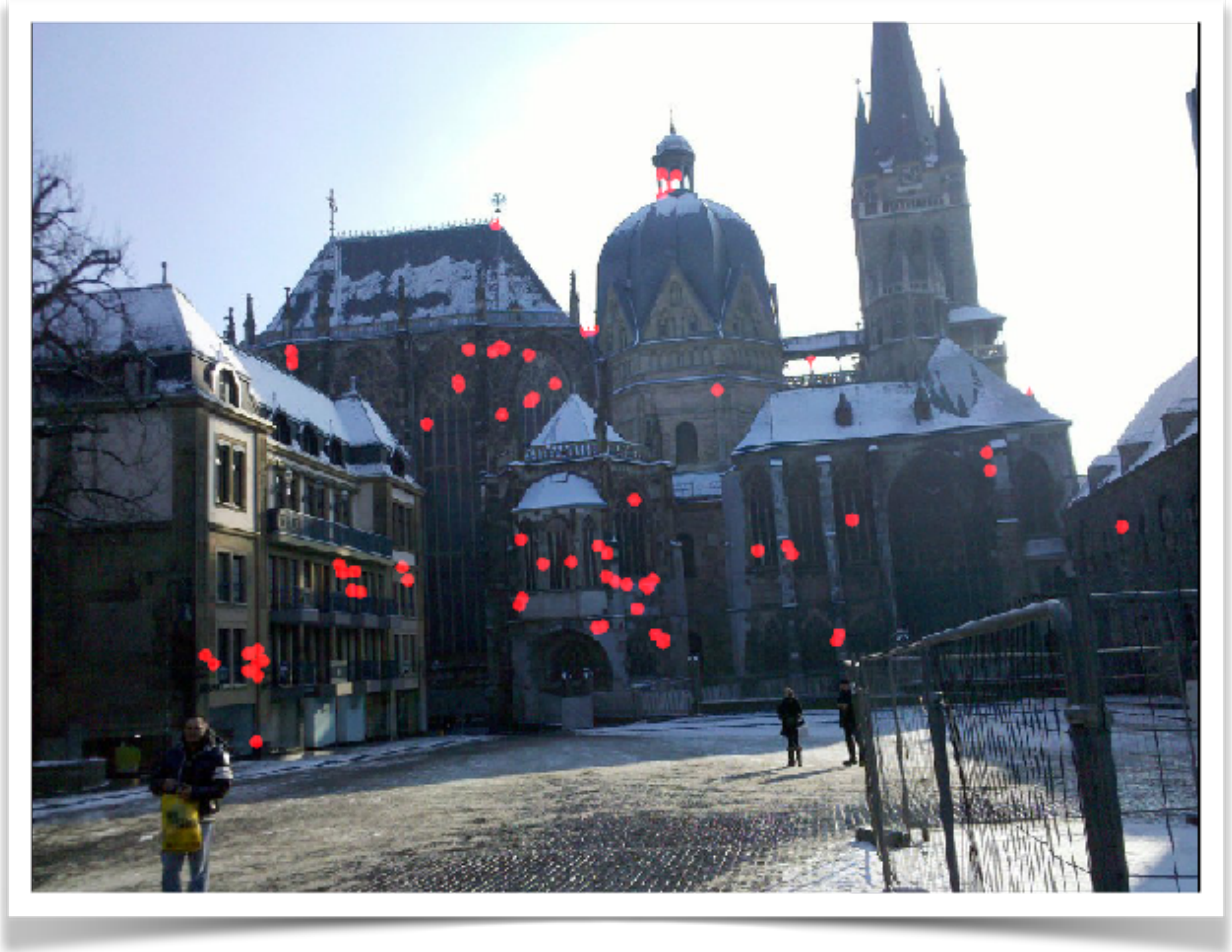
Classic Localization Pipeline

Extract Local Features

Establish 2D-3D Matches ← **nearest neighbor search**

Estimate Camera Pose ← **well-understood problem**

Classic Localization Pipeline



Extract Local Features

Establish 2D-3D Matches ← **nearest neighbor search**

Estimate Camera Pose ← **well-understood problem**

Overview

I. CNNs for Visual Localization

II. CNNs for Feature Detection & Description

Learning Feature Detectors

- What are properties of a good feature detector?

Learning Feature Detectors

- What are properties of a good feature detector?
- Repeatability, stability, viewpoint invariance

Learning Feature Detectors

- What are properties of a good feature detector?
  - Repeatability, stability, viewpoint invariance
  - Fire at “interesting regions” suitable for matching

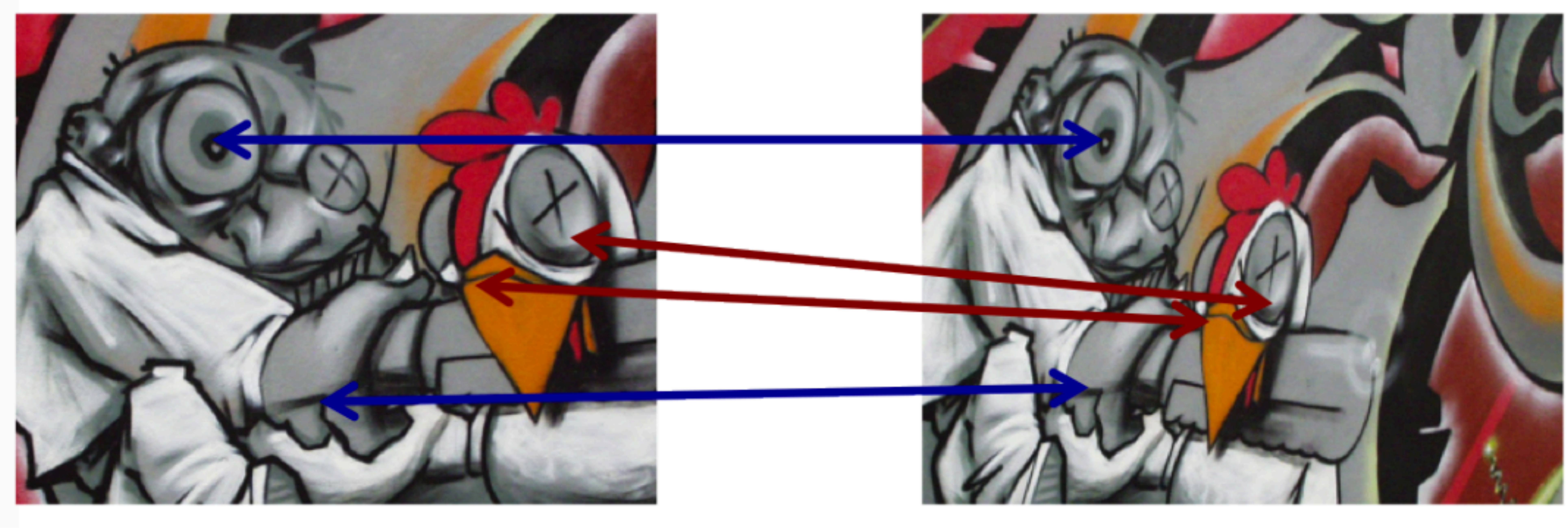
Learning Feature Detectors

- What are properties of a good feature detector?
  - Repeatability, stability, viewpoint invariance
  - Fire at “interesting regions” suitable for matching
- How to model this mathematically?

Learning Feature Detectors

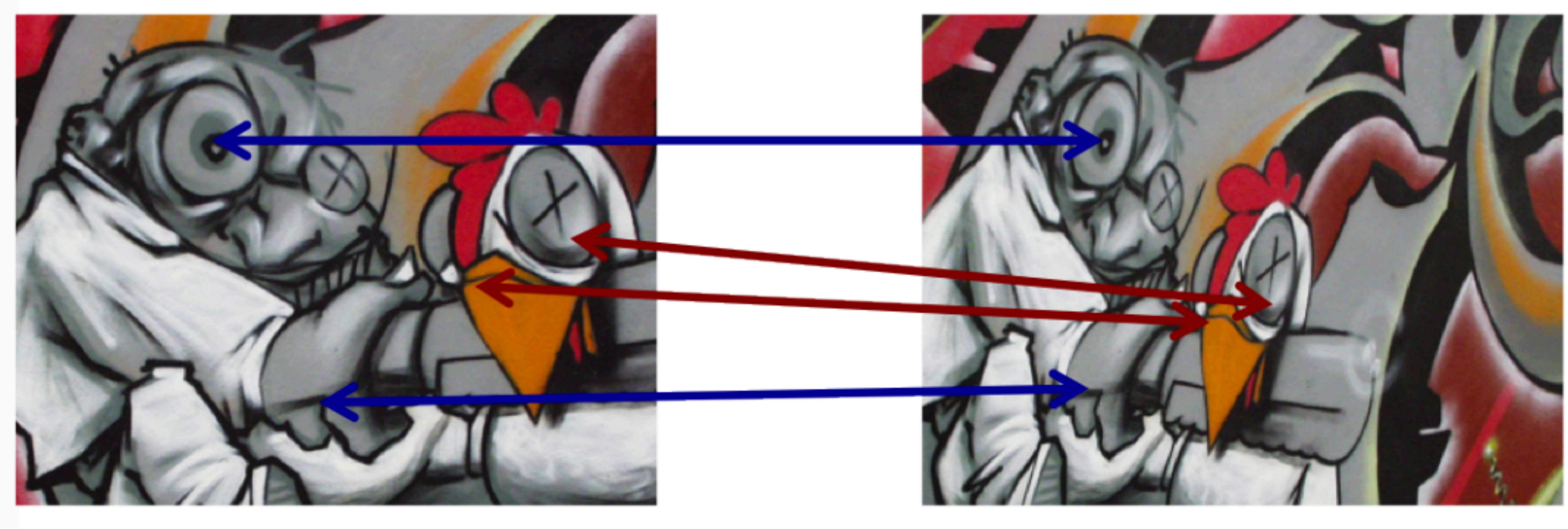
- What are properties of a good feature detector?
  - Repeatability, stability, viewpoint invariance
  - Fire at “interesting regions” suitable for matching
- How to model this mathematically?
- How to train a detector from scratch without any bias to existing solutions?

Learning Feature Detectors



[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

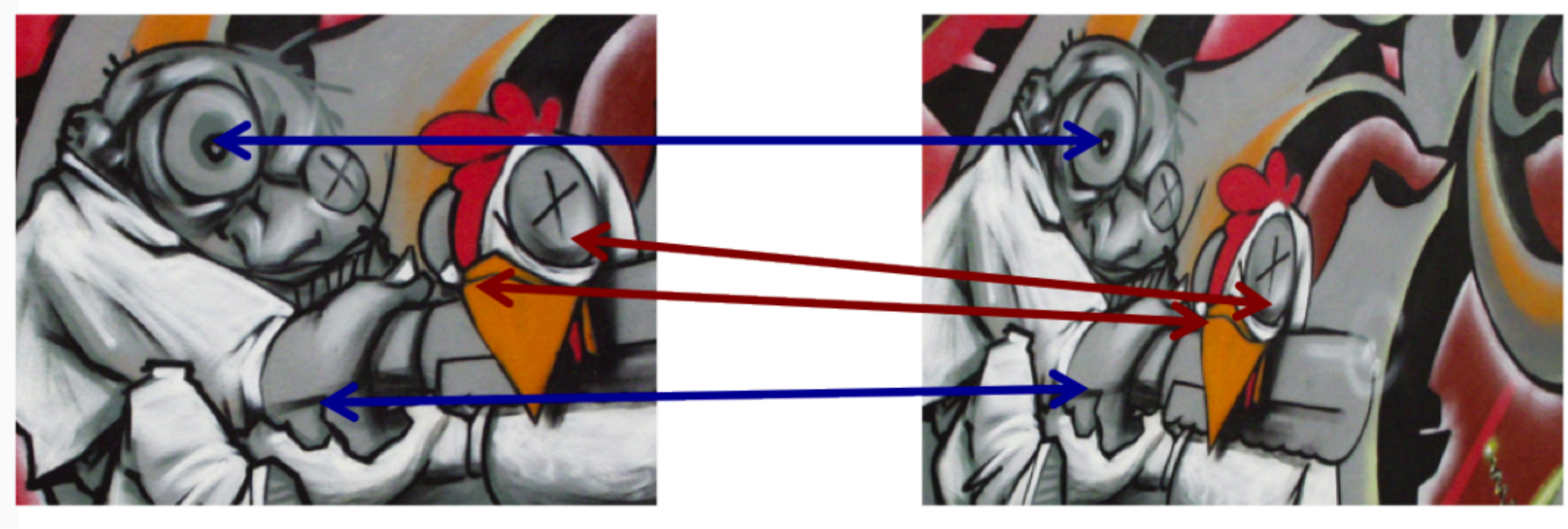
Learning Feature Detectors



- Learn function $\mathtt{H(x|w)}\colon \mathtt{R}^2 \rightarrow \mathtt{[-1, 1]}$ with parameters $\mathtt{w}$

[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

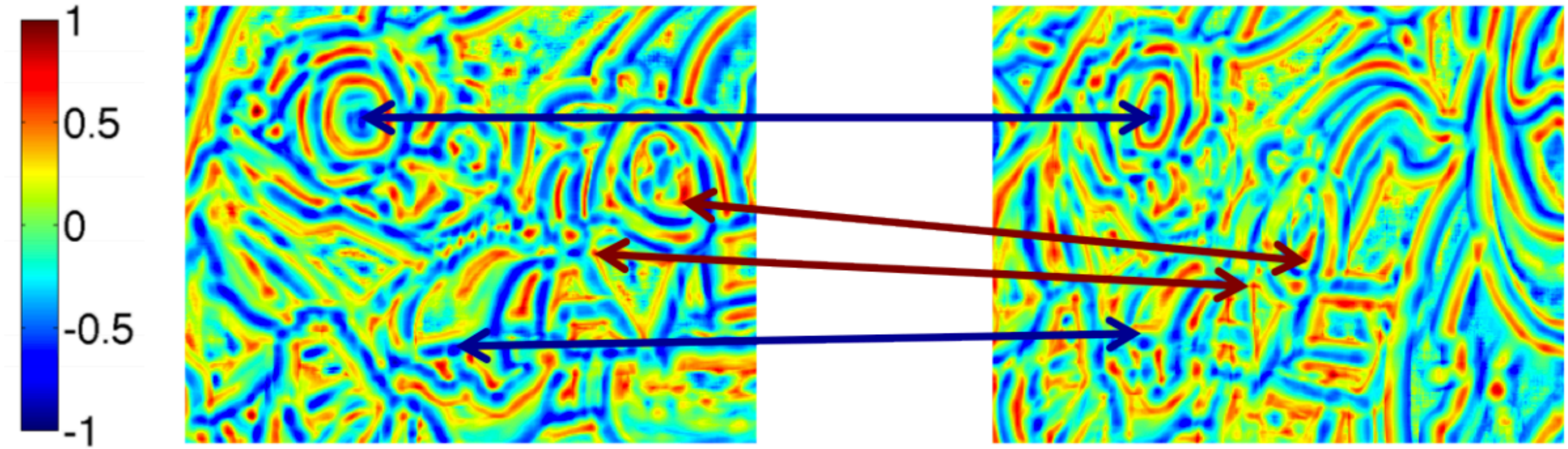
Learning Feature Detectors



- Learn function $\mathtt{H(x|w)}$: $\mathtt{R}^2 \rightarrow \mathtt{[-1, 1]}$ with parameters $\mathtt{w}$
- Interesting points are close to $\mathtt{-1}$ or $\mathtt{1}$

[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

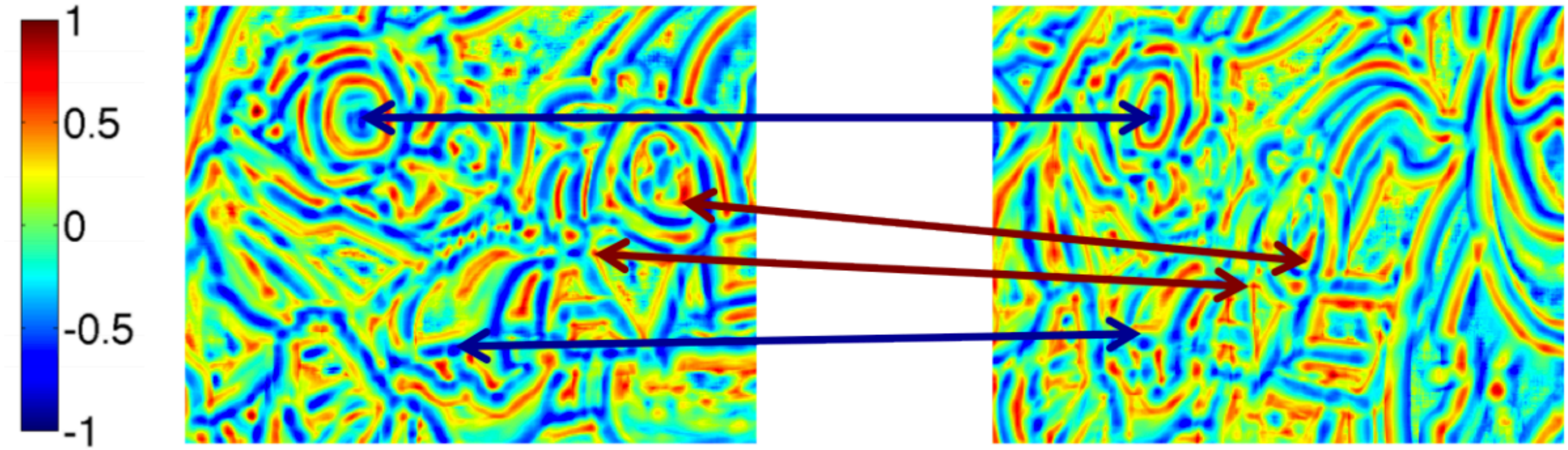
Learning Feature Detectors



- Learn function $\mathtt{H(x|w)}: \mathbb{R}^2 \rightarrow \mathtt{[-1, 1]}$ with parameters $\mathtt{w}$
- Interesting points are close to `-1` or `1`

[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

Learning Feature Detectors



- Learn function $\mathtt{H(x|w)}: \mathtt{R}^2 \rightarrow \mathtt{[-1, 1]}$ with parameters $\mathtt{w}$
- Interesting points are close to $\mathtt{-1}$ or $\mathtt{1}$
- Repeatability = consistent ranking under transformations

[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

Learning to Rank

- Learn consistent ranking $\mathtt{H(x|w)}$:

[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

Learning to Rank

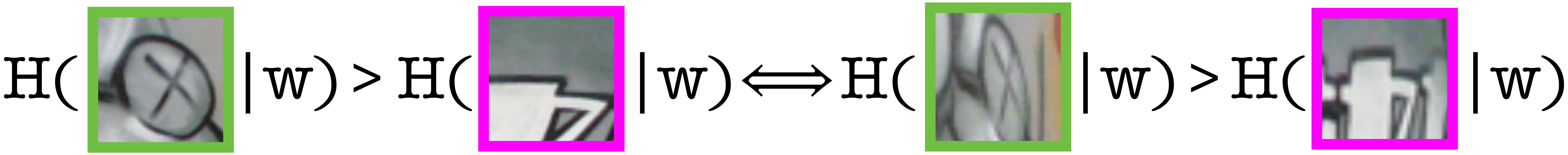
- Learn consistent ranking $\mathtt{H(x|w)}$:

$\mathtt{H(}$  $\mathtt{|w) > H(}$  $\mathtt{|w)}$

[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

Learning to Rank

- Learn consistent ranking $H(x|w)$:



[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

Learning to Rank

- Learn consistent ranking $\mathtt{H(x|w)}$:

$\mathtt{H(}$  $\mathtt{|w)}$ $\mathtt{H(}$  $\mathtt{|w)}$ $\mathtt{H(}$  $\mathtt{|w)}$ $\mathtt{H(}$  $\mathtt{|w)}$

[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

Learning to Rank

- Learn consistent ranking $\mathtt{H(x|w)}$:

(H( |w) - H( |w))   H( |w)   H( |w)

[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

Learning to Rank

- Learn consistent ranking $\mathtt{H(x|w)}$:


$\mathtt{(H(}$  $\mathtt{|w) - H(}$  $\mathtt{|w))}$ $\mathtt{(H(}$  $\mathtt{|w) - H(}$  $\mathtt{|w))}$

[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

Learning to Rank

- Learn consistent ranking $\mathtt{H(x|w)}$:

(H( |w) - H( |w)) * (H( |w) - H( |w))

[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

Learning to Rank

- Learn consistent ranking $\mathtt{H(x|w)}$:

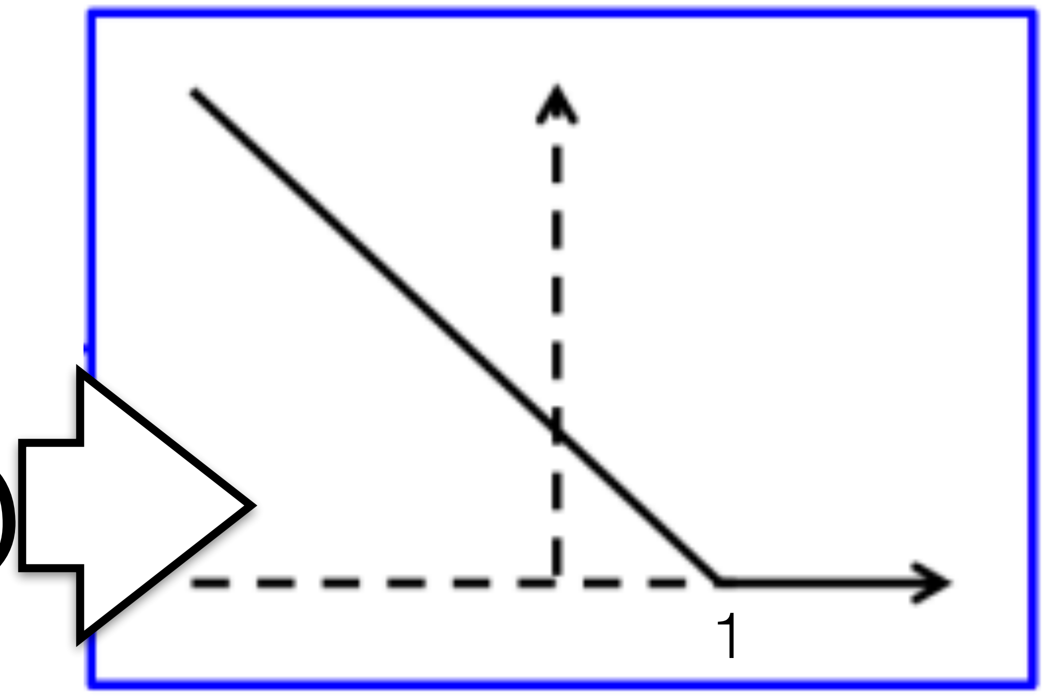
(H( |w) - H( |w)) * (H( |w) - H( |w)) > 0

[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

Learning to Rank

- Learn consistent ranking $\mathtt{H(x|w)}$:

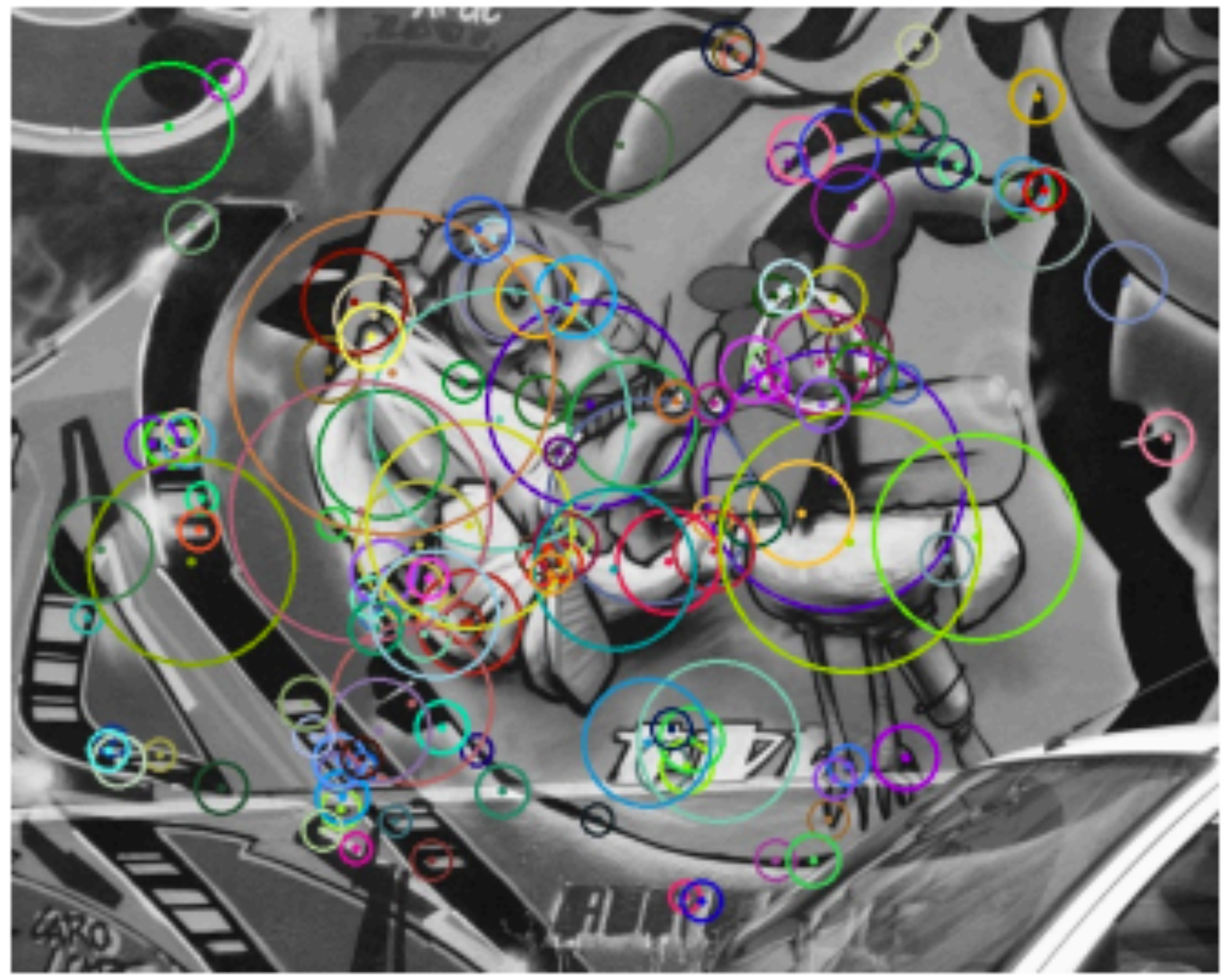

(H( |w) - H( |w)) * (H( |w) - H( |w))

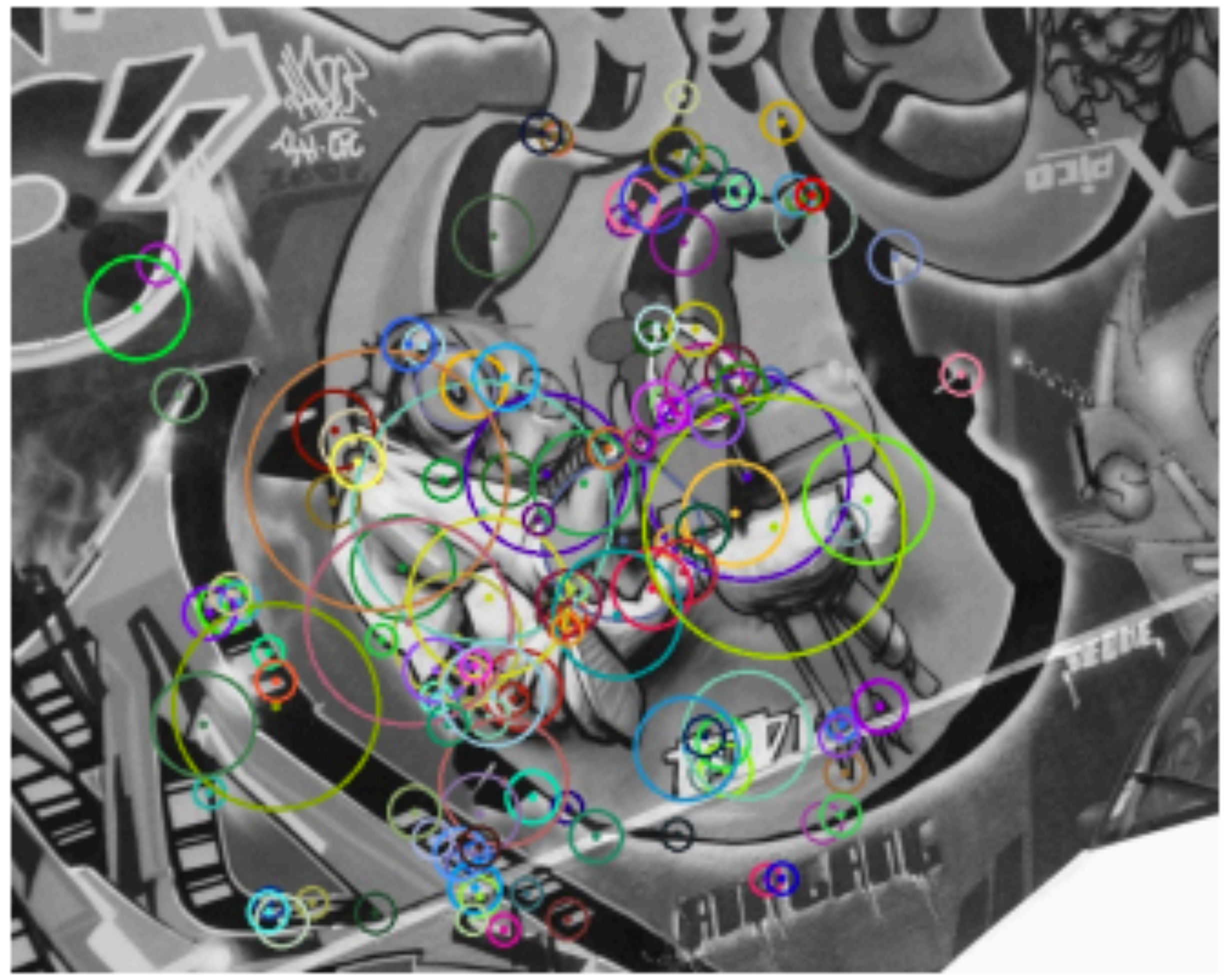


Hinge loss

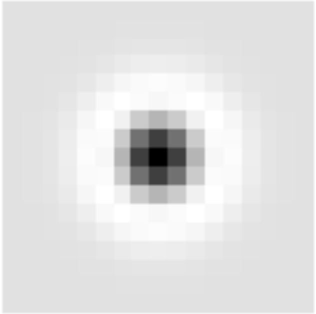
[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

Detection Results



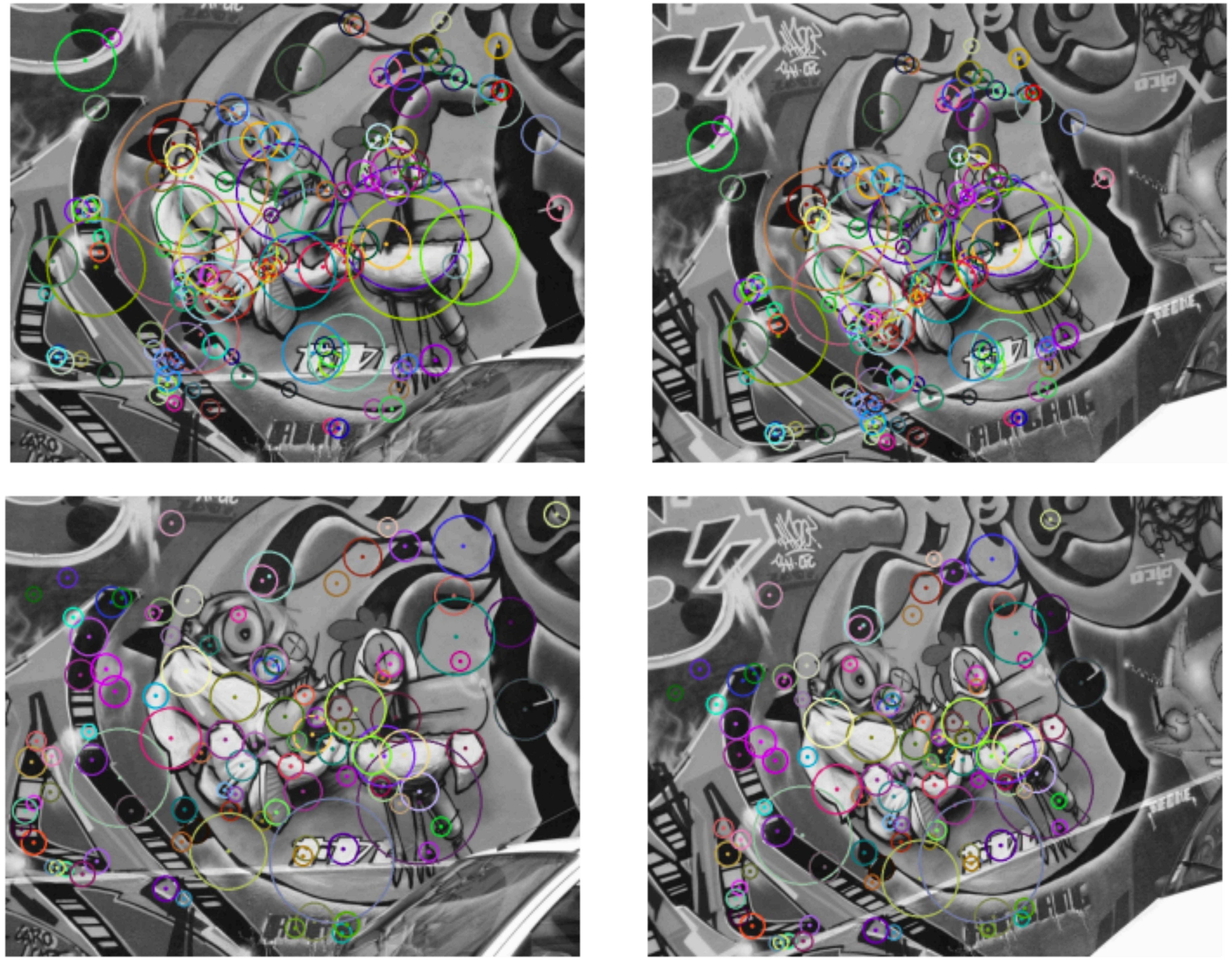


Difference-of-Gaussians

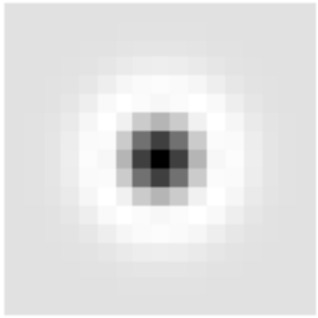


[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

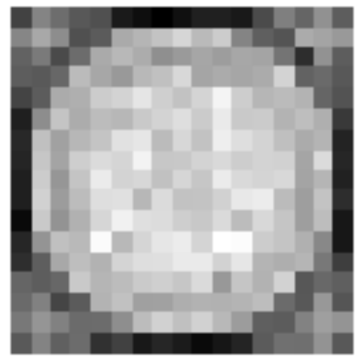
Detection Results



Difference-of-Gaussians

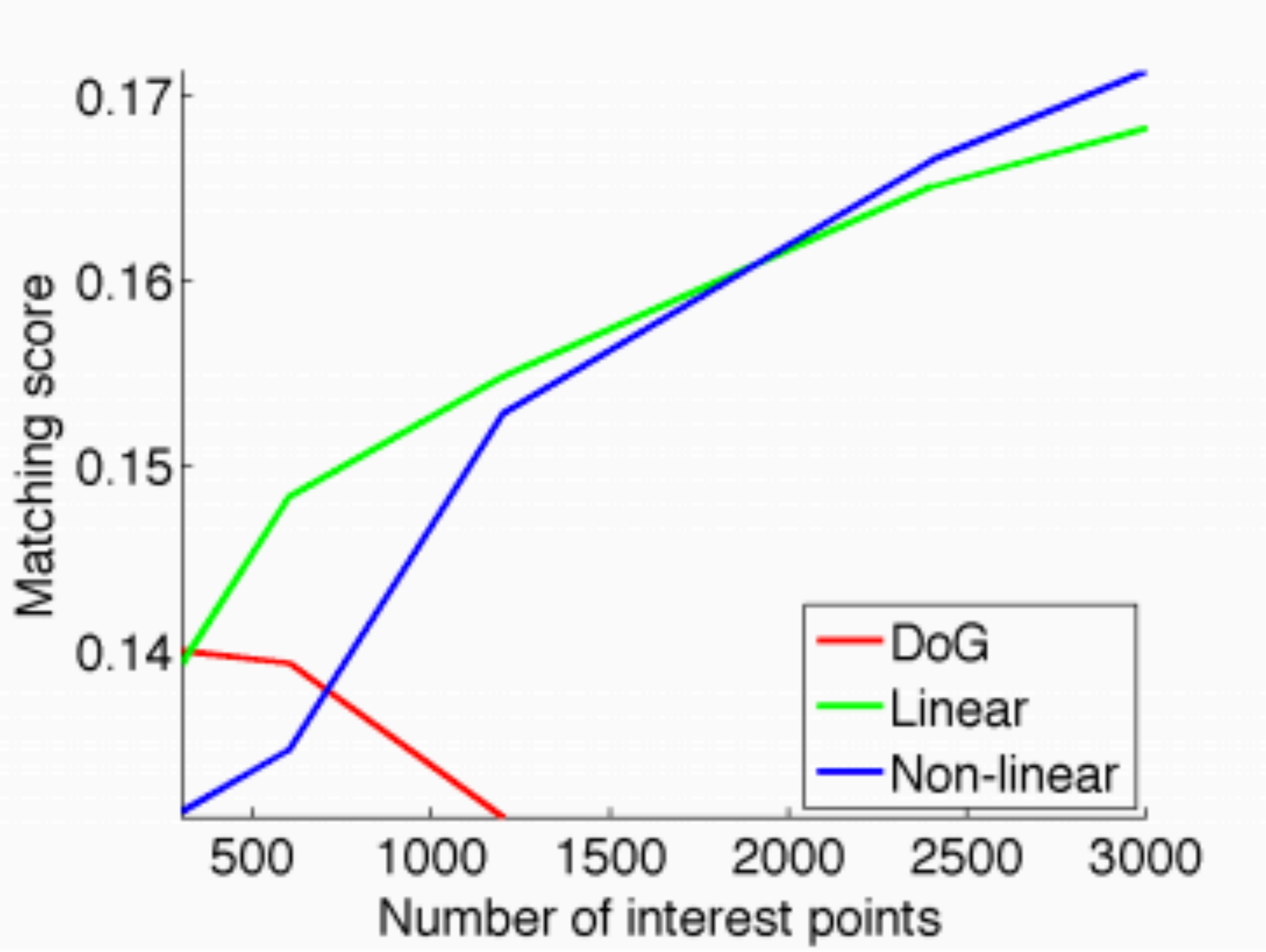


ours

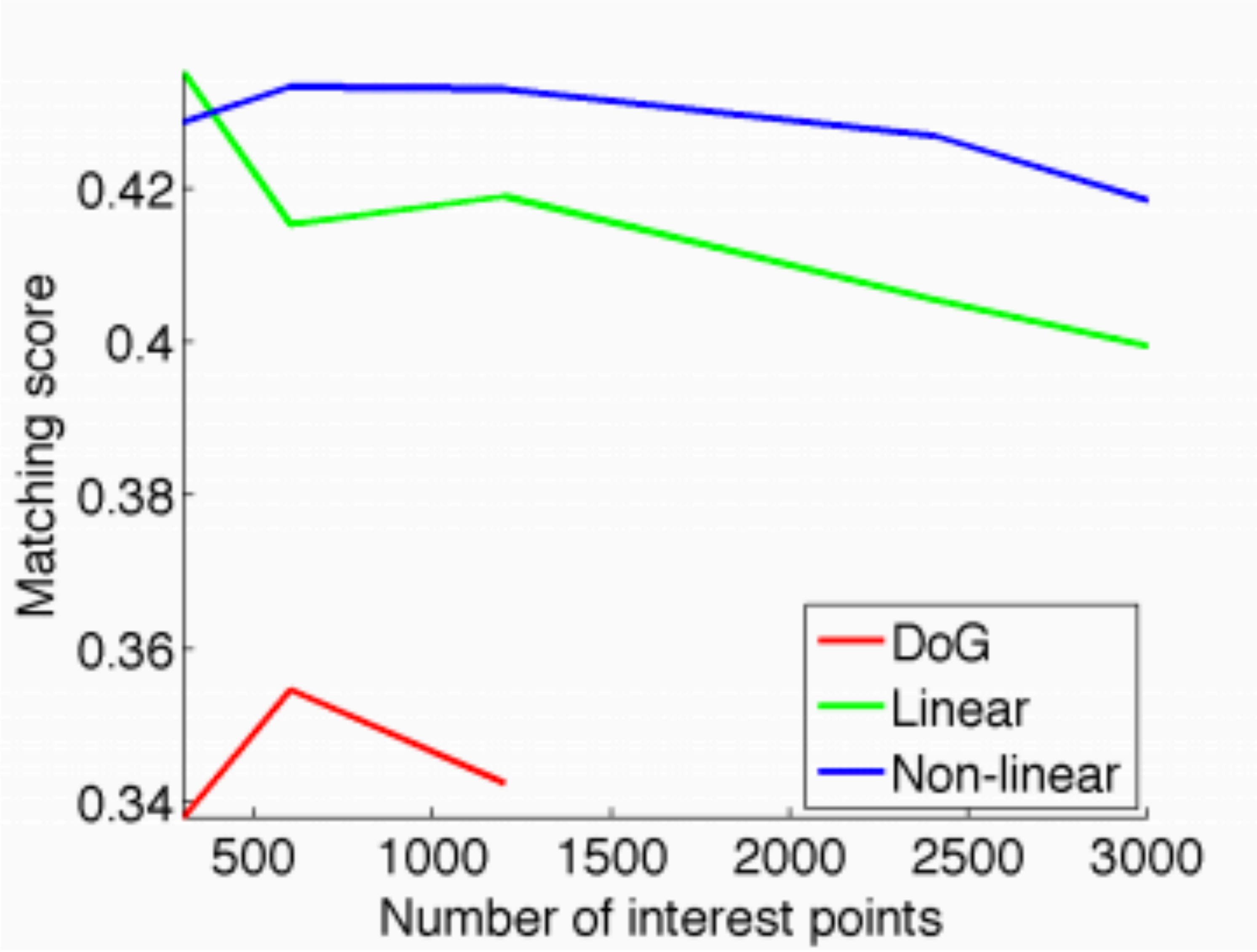


[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

Matching Results



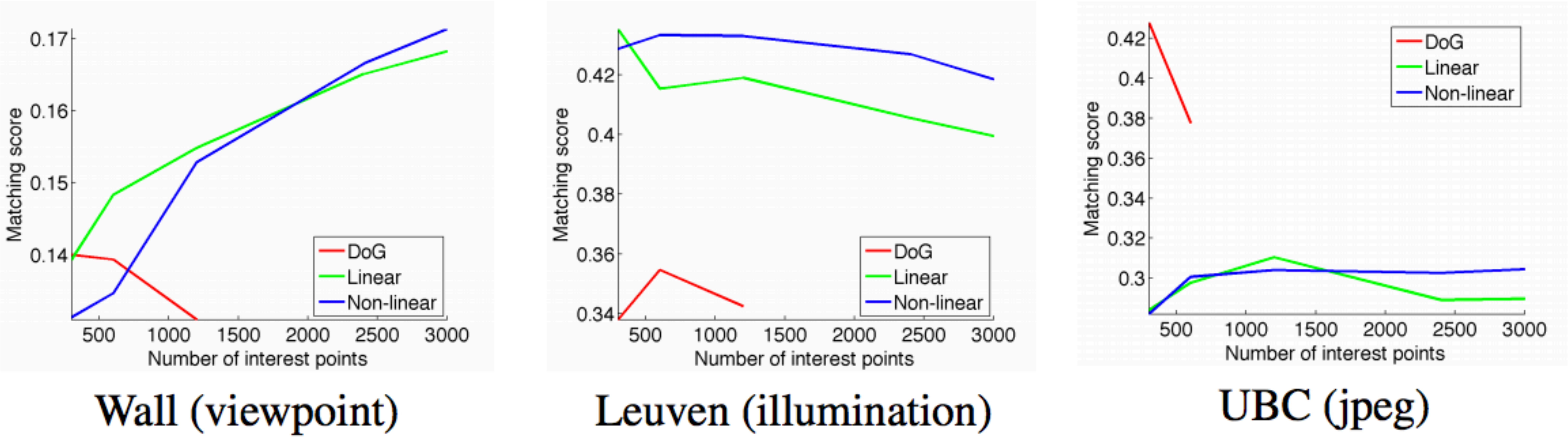
Wall (viewpoint)



Leuven (illumination)

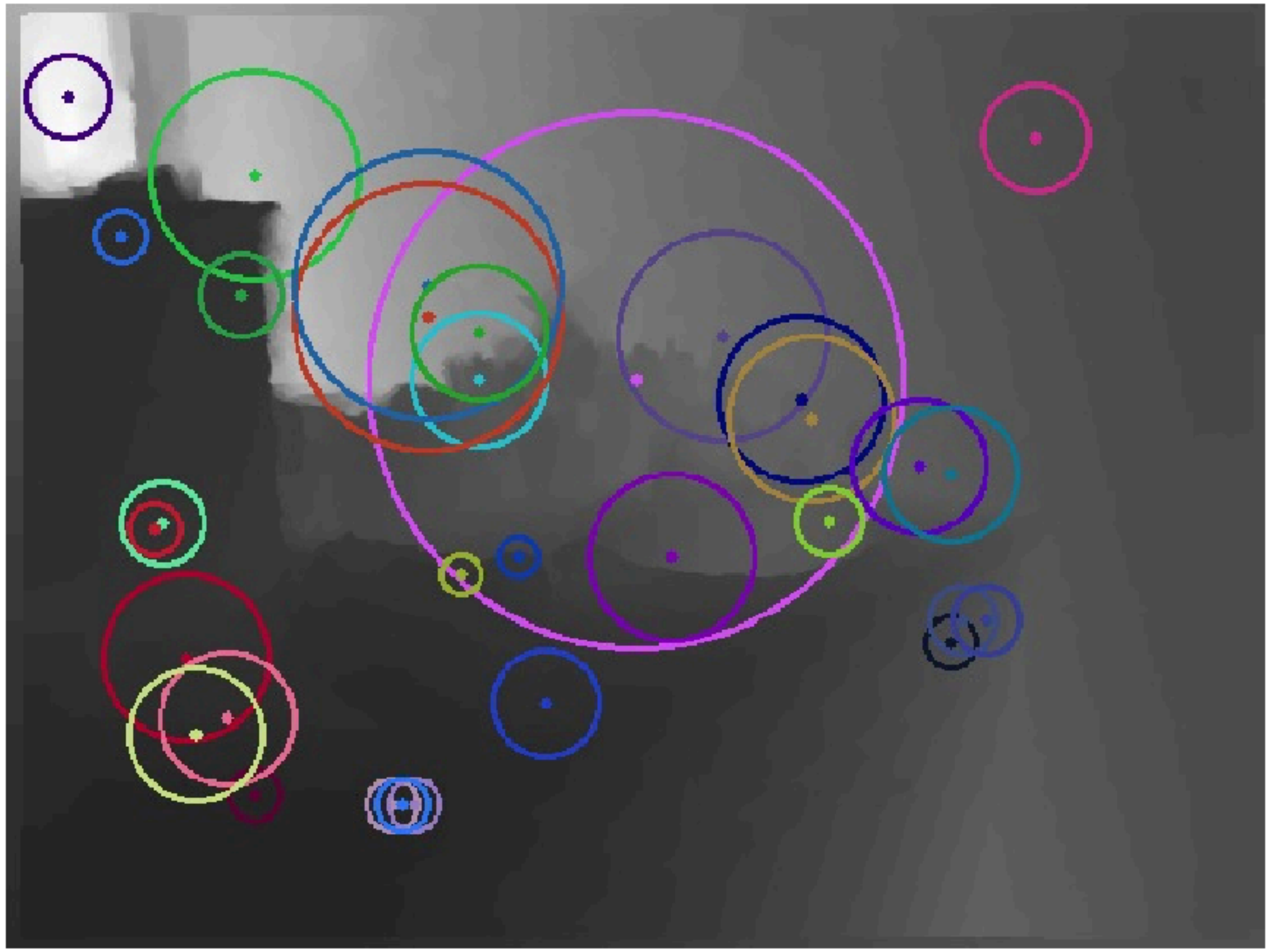
[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

Matching Results



[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

Multi-Modal Features



[Savinov, Seki, Ladicky, **Sattler**, Pollefeys, Quad-networks: unsupervised learning to rank for interest point detection, CVPR 2017]

Learning Patch Descriptors

- Learn mapping from patch to descriptor in $\mathbb{R}^n$

[Schönberger, Hardmeier, **Sattler**, Pollefeys, Evaluation of Hand-Crafted and Learned Local Features. CVPR 2017]

Learning Patch Descriptors

- Learn mapping from patch to descriptor in $\mathbb{R}^n$
- Popular approach: Learning via *triplets*

[Schönberger, Hardmeier, **Sattler**, Pollefeys, Evaluation of Hand-Crafted and Learned Local Features. CVPR 2017]

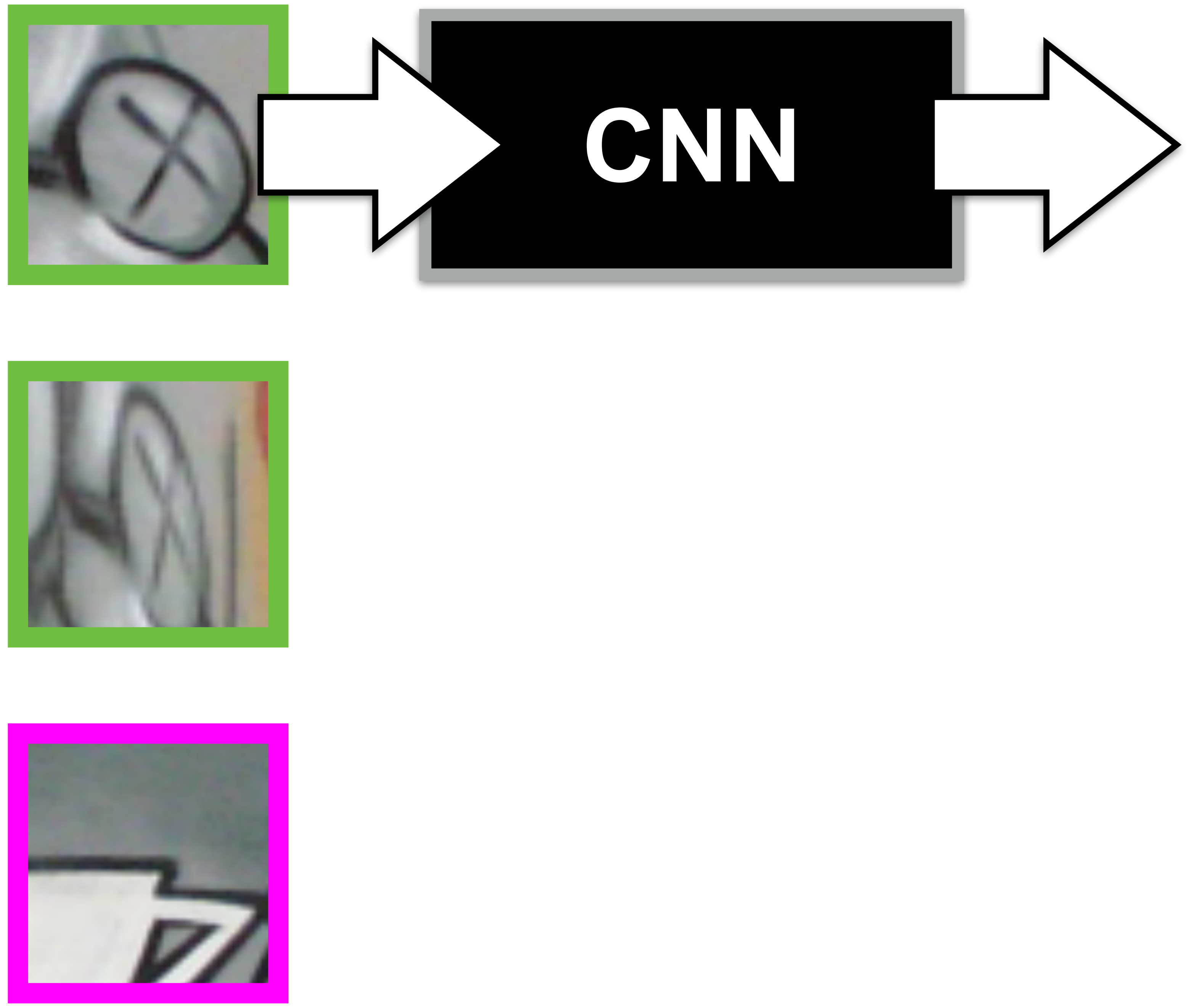
Learning Patch Descriptors

- Learn mapping from patch to descriptor in $\mathbb{R}^n$
- Popular approach: Learning via *triplets*

[Schönberger, Hardmeier, **Sattler**, Pollefeys, Evaluation of Hand-Crafted and Learned Local Features. CVPR 2017]

Learning Patch Descriptors

- Learn mapping from patch to descriptor in $\mathbb{R}^n$
- Popular approach: Learning via *triplets*

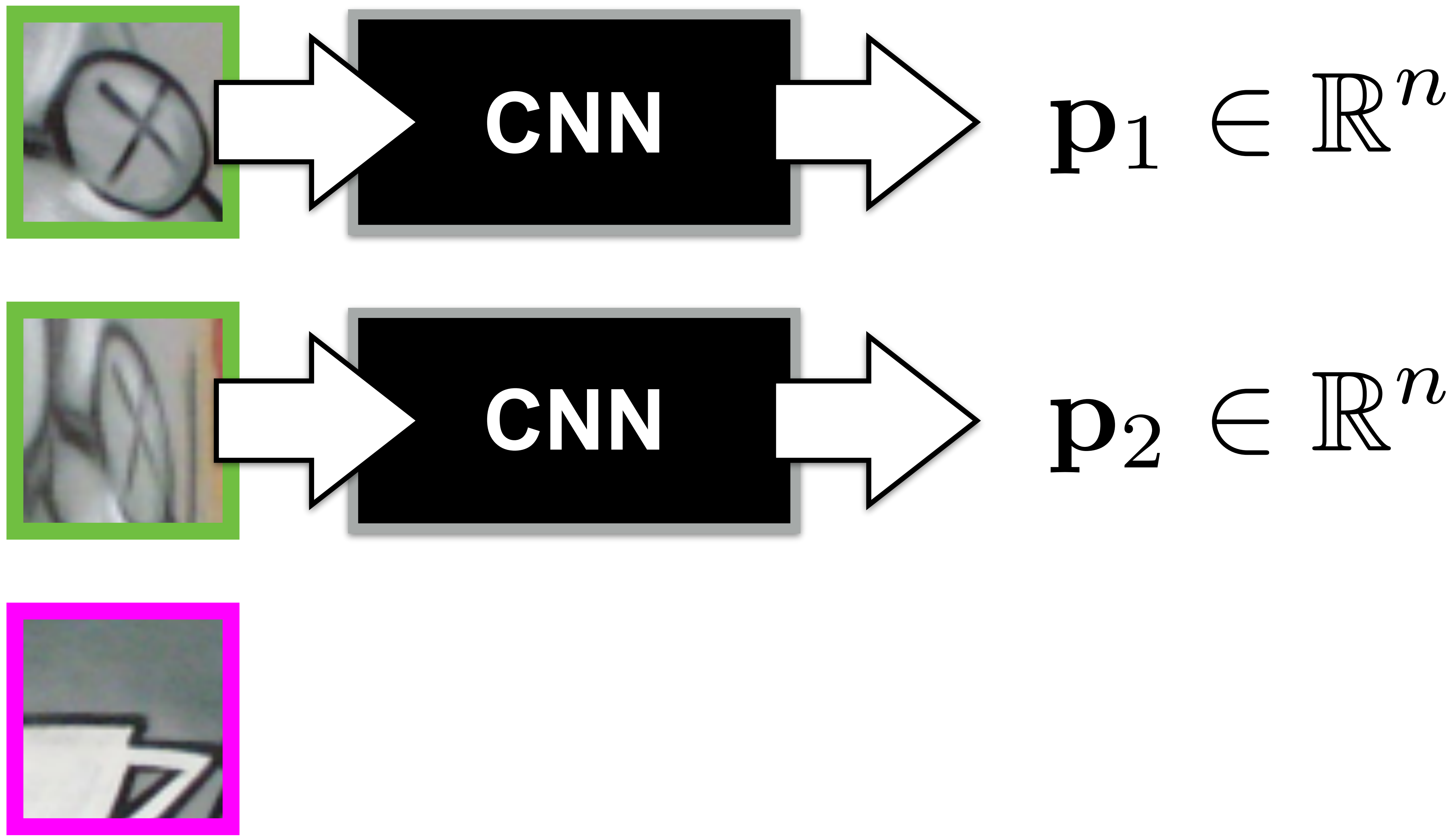


$\mathbf{p}_1 \in \mathbb{R}^n$

[Schönberger, Hardmeier, **Sattler**, Pollefeys, Evaluation of Hand-Crafted and Learned Local Features. CVPR 2017]

Learning Patch Descriptors

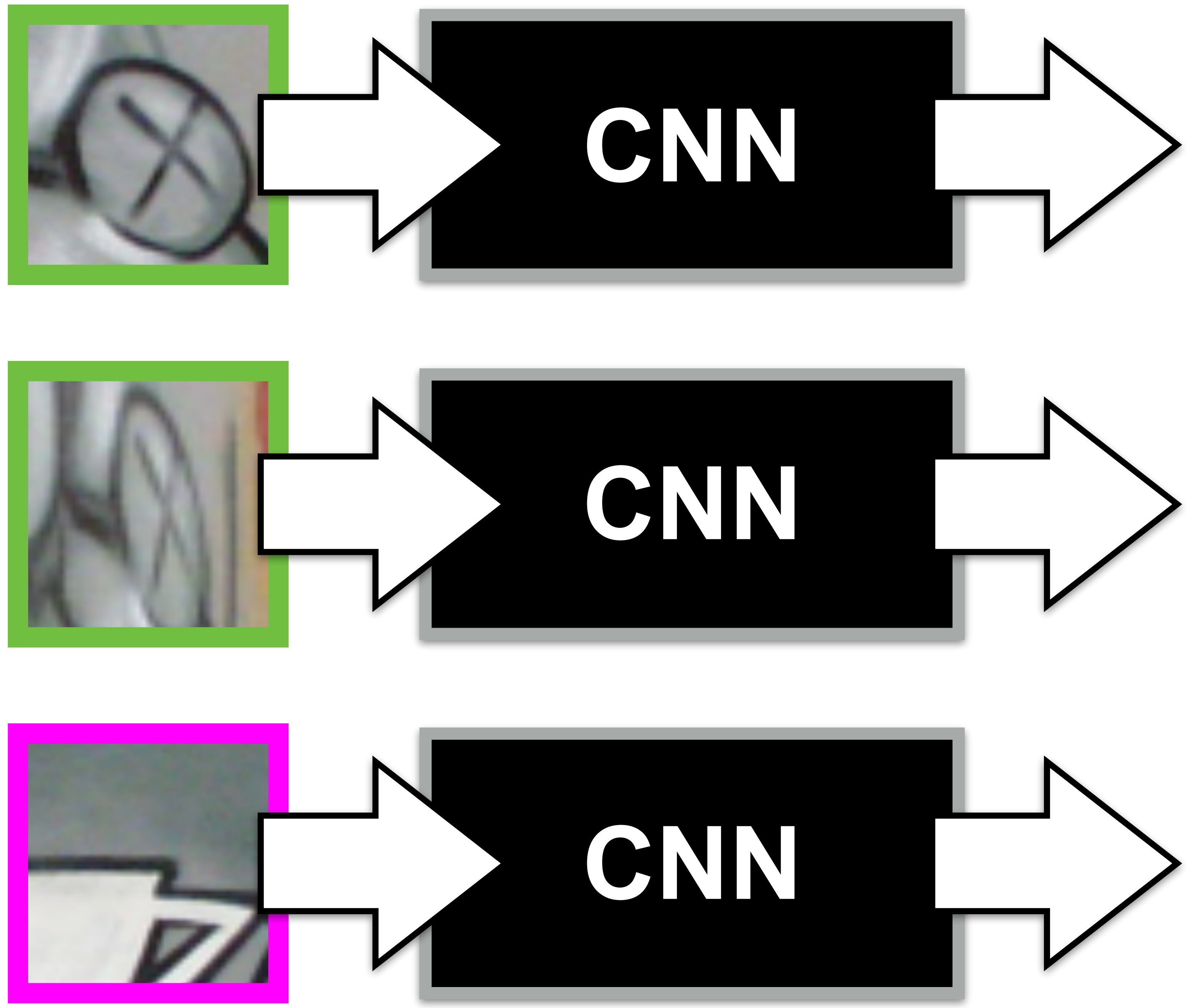
- Learn mapping from patch to descriptor in $\mathbb{R}^n$
- Popular approach: Learning via *triplets*



[Schönberger, Hardmeier, **Sattler**, Pollefeys, Evaluation of Hand-Crafted and Learned Local Features. CVPR 2017]

Learning Patch Descriptors

- Learn mapping from patch to descriptor in $\mathbb{R}^n$
- Popular approach: Learning via *triplets*



$\mathbf{p}_1 \in \mathbb{R}^n$

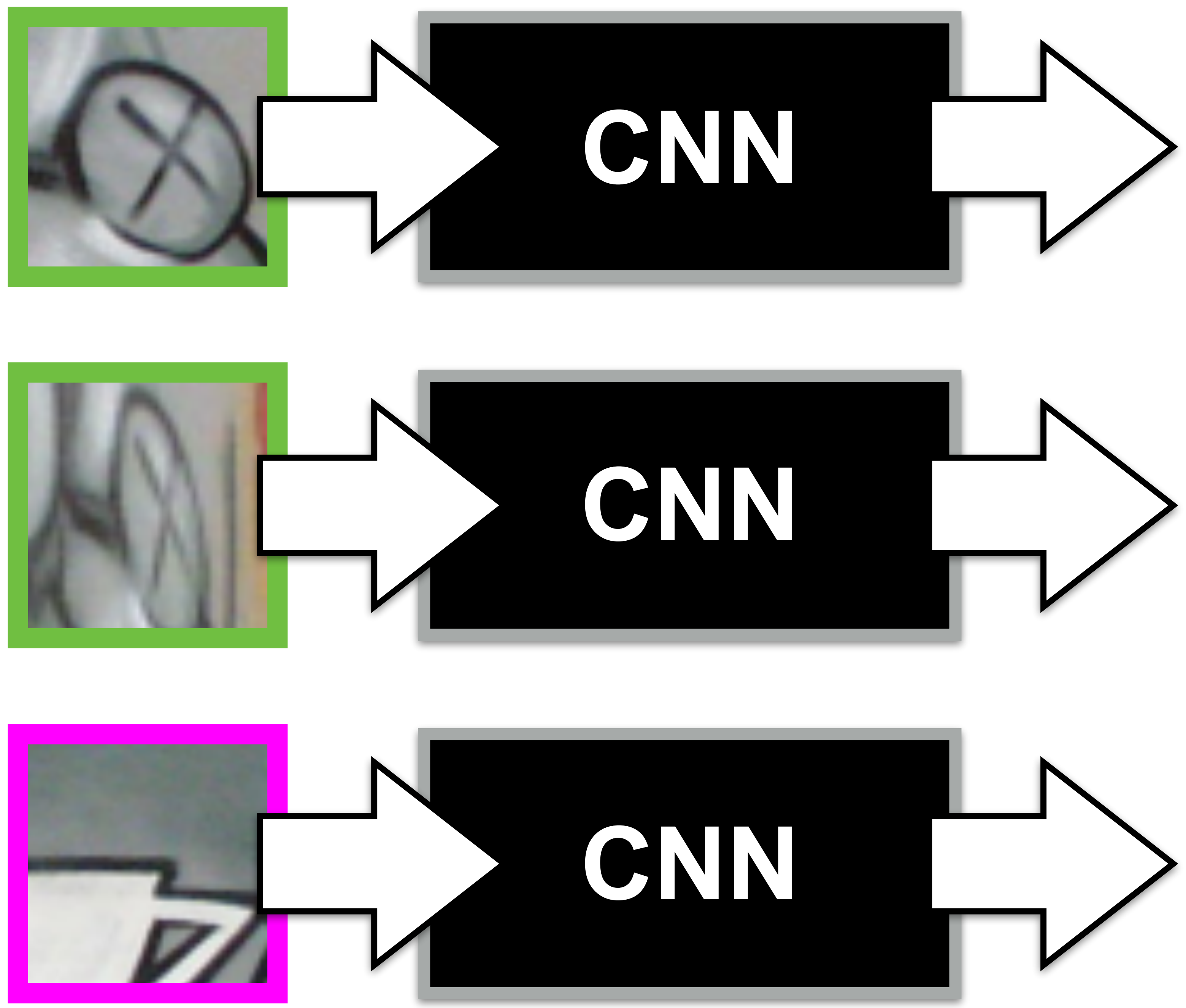
$\mathbf{p}_2 \in \mathbb{R}^n$

$\mathbf{n} \in \mathbb{R}^n$

[Schönberger, Hardmeier, **Sattler**, Pollefeys, Evaluation of Hand-Crafted and Learned Local Features. CVPR 2017]

Learning Patch Descriptors

- Learn mapping from patch to descriptor in $\mathbb{R}^n$
- Popular approach: Learning via *triplets*



$$\mathbf{p}_1 \in \mathbb{R}^n$$

$$\mathbf{p}_2 \in \mathbb{R}^n$$

$$\mathbf{n} \in \mathbb{R}^n$$

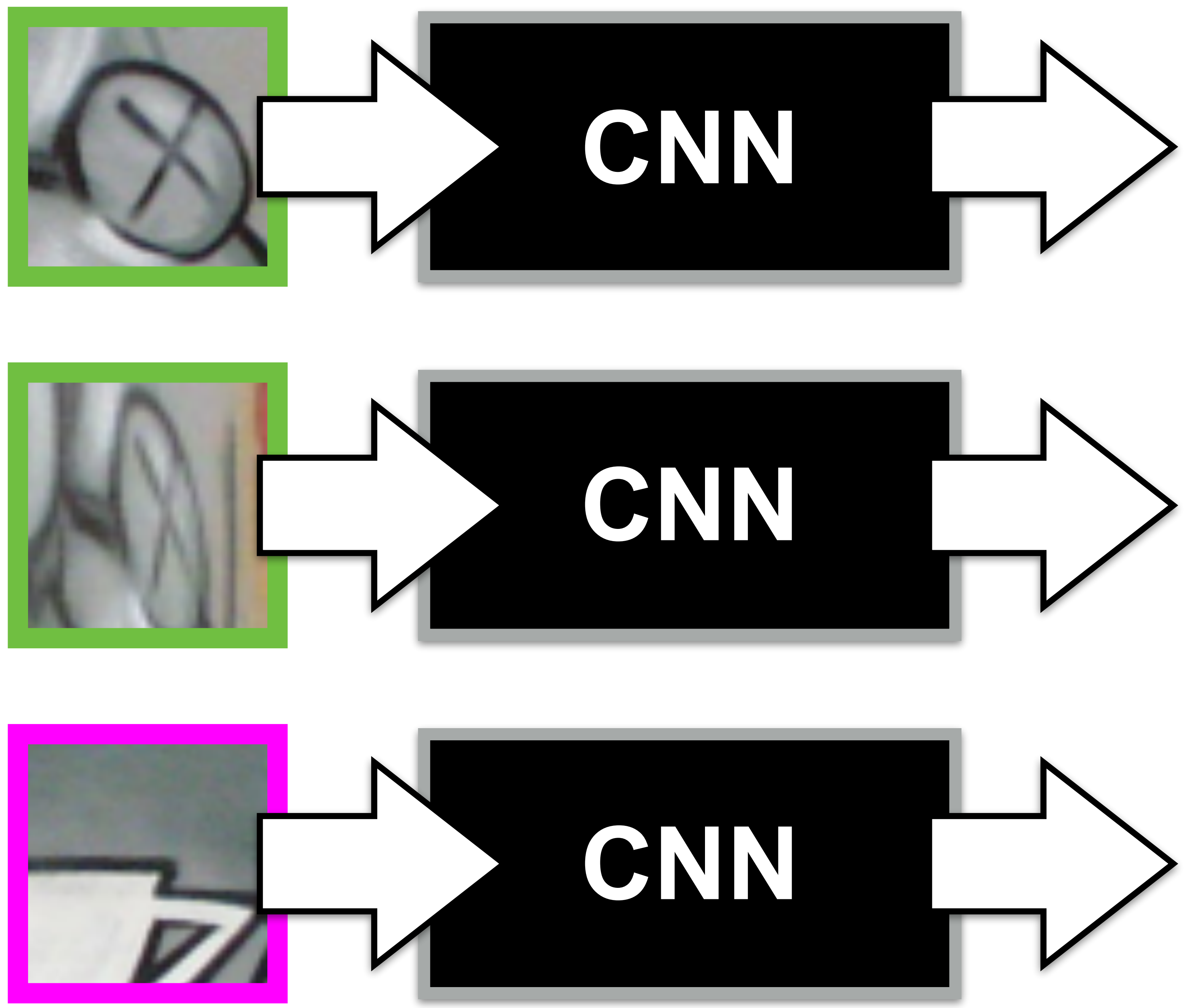
triplet loss:

$$\max(0, \eta + ||\mathbf{p}_1 - \mathbf{p}_2||_2 - ||\mathbf{n} - \mathbf{p}_2||_2)$$

[Schönberger, Hardmeier, **Sattler**, Pollefeys, Evaluation of Hand-Crafted and Learned Local Features. CVPR 2017]

Learning Patch Descriptors

- Learn mapping from patch to descriptor in $\mathbb{R}^n$
- Popular approach: Learning via *triplets*



$\mathbf{p}_1 \in \mathbb{R}^n$

$\mathbf{p}_2 \in \mathbb{R}^n$

$\mathbf{n} \in \mathbb{R}^n$

triplet loss:

$$\max(0, \eta + ||\mathbf{p}_1 - \mathbf{p}_2||_2 - ||\mathbf{n} - \mathbf{p}_2||_2)$$

margin

[Schönberger, Hardmeier, **Sattler**, Pollefeys, Evaluation of Hand-Crafted and Learned Local Features. CVPR 2017]

Hand-Crafted vs. Learned Descriptors

- Comparing learned with hand-crafted descriptors (SIFT variants)

[Schönberger, Hardmeier, **Sattler**, Pollefeys, Evaluation of Hand-Crafted and Learned Local Features. CVPR 2017]

Hand-Crafted vs. Learned Descriptors

- Comparing learned with hand-crafted descriptors (SIFT variants)
- Evaluated on Structure-from-Motion task

[Schönberger, Hardmeier, **Sattler**, Pollefeys, Evaluation of Hand-Crafted and Learned Local Features. CVPR 2017]

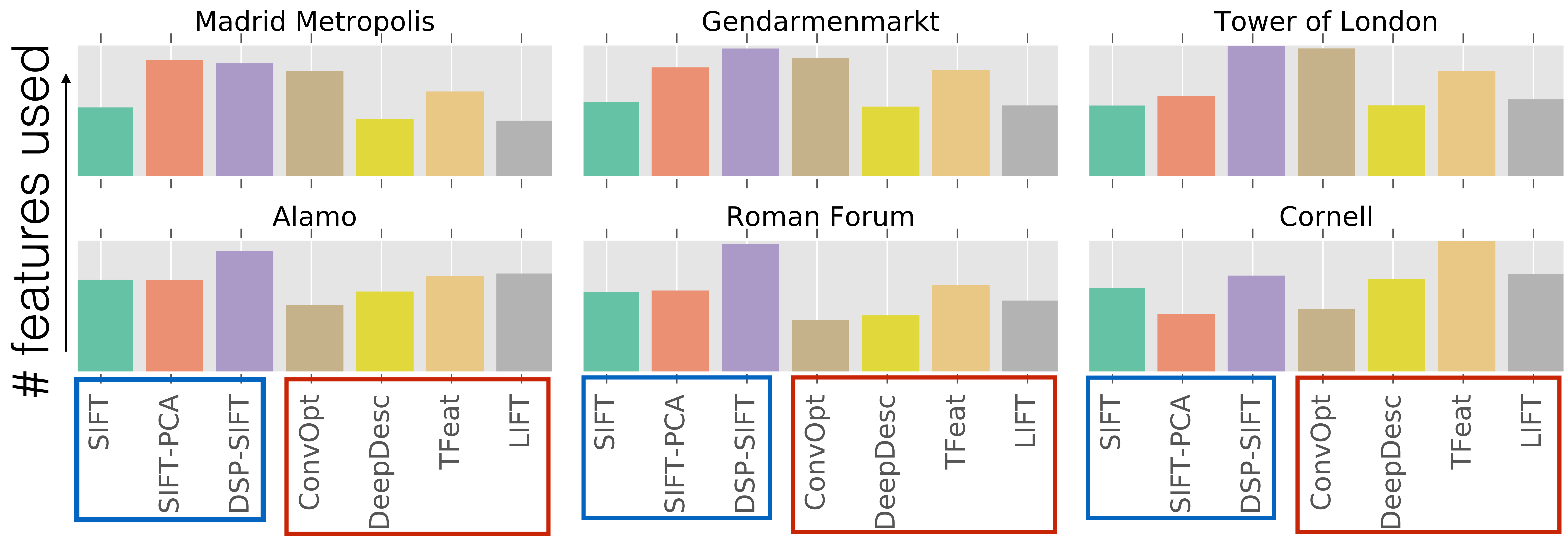
Hand-Crafted vs. Learned Descriptors

- Comparing learned with hand-crafted descriptors (SIFT variants)
- Evaluated on Structure-from-Motion task
- Measure: Number of triangulated features (higher = better)

[Schönberger, Hardmeier, **Sattler**, Pollefeys, Evaluation of Hand-Crafted and Learned Local Features. CVPR 2017]

Hand-Crafted vs. Learned Descriptors

- Comparing learned with hand-crafted descriptors (SIFT variants)
- Evaluated on Structure-from-Motion task
- Measure: Number of triangulated features (higher = better)



[Schönberger, Hardmeier, **Sattler**, Pollefeys, Evaluation of Hand-Crafted and Learned Local Features. CVPR 2017]

Three Cases

- Visual Localization: CNN-based approach clearly worse than state-of-the-art

Three Cases

- Visual Localization: CNN-based approach clearly worse than state-of-the-art
- Feature detector learning: Similar to better performance compared to state-of-the-art

Three Cases

- Visual Localization: CNN-based approach clearly worse than state-of-the-art
- Feature detector learning: Similar to better performance compared to state-of-the-art
- Feature descriptor learning: Hand-crafted descriptors perform better for wide range of scenes

Out with the Old? - Lessons Learned

- Hold off replacing everything with CNNs (at least for now)

Out with the Old? - Lessons Learned

- Hold off replacing everything with CNNs (at least for now)
- Consider using a CNN if:

Out with the Old? - Lessons Learned

- Hold off replacing everything with CNNs (at least for now)
- Consider using a CNN if:
  - Current solutions do not perform well on your task

Out with the Old? - Lessons Learned

- Hold off replacing everything with CNNs (at least for now)
- Consider using a CNN if:
  - Current solutions do not perform well on your task
  - Your task is rather specific, i.e., generalization is not important

Out with the Old? - Lessons Learned

- Hold off replacing everything with CNNs (at least for now)
- Consider using a CNN if:
  - Current solutions do not perform well on your task
  - Your task is rather specific, i.e., generalization is not important
  - You have enough training data

Out with the Old? - Lessons Learned

- Hold off replacing everything with CNNs (at least for now)
- Consider using a CNN if:
  - Current solutions do not perform well on your task
  - Your task is rather specific, i.e., generalization is not important
  - You have enough training data
- In any case: Compare against simple baselines!