R2D2: Repeatable and Reliable Detector and Descriptor

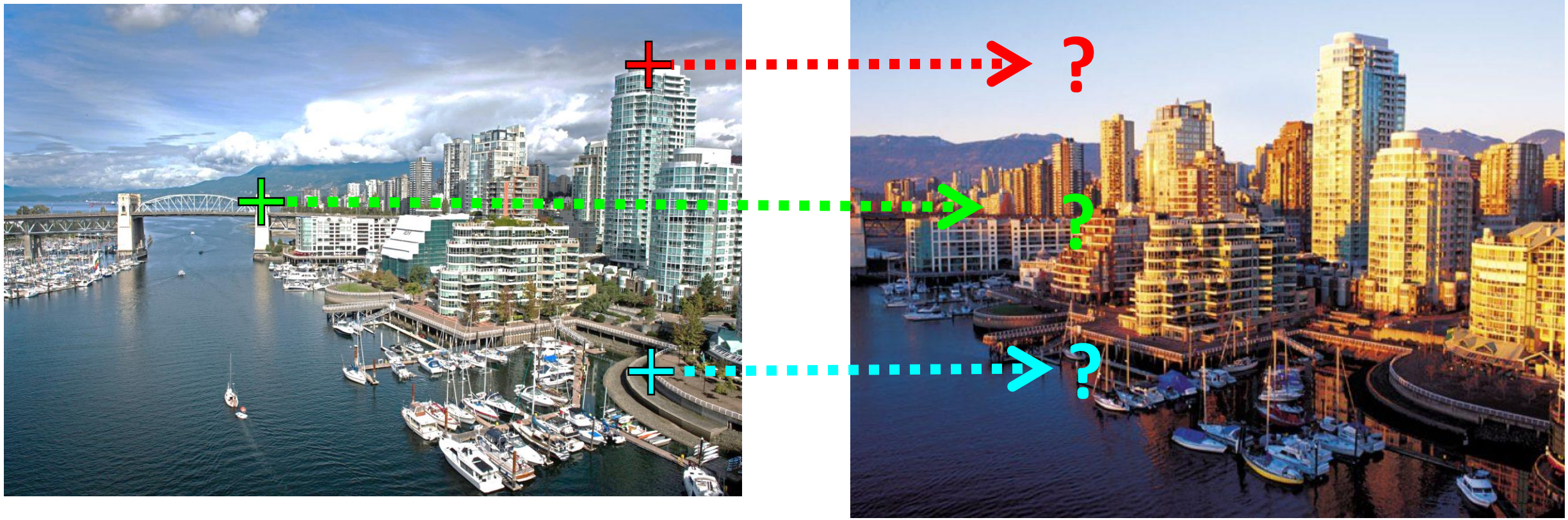
Jérôme Revaud
Philippe Weinzaepfel
César De Souza
Martin Humenberger

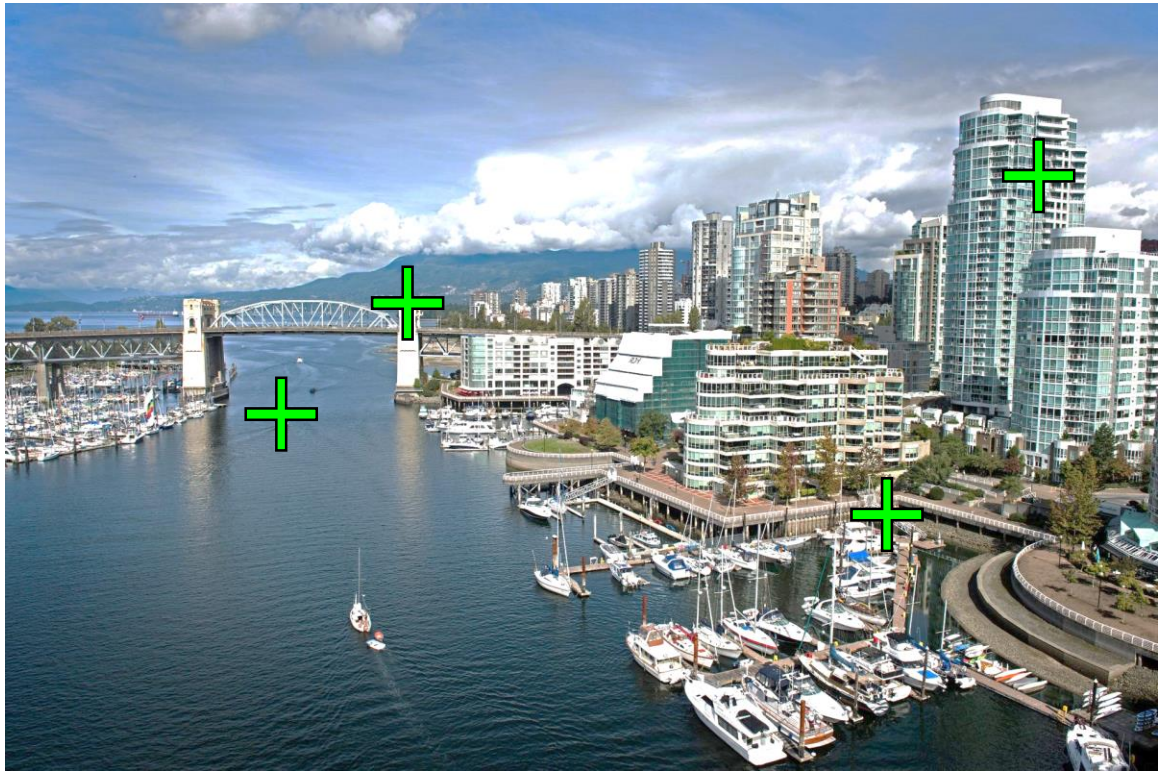
NAVER LABS Europe

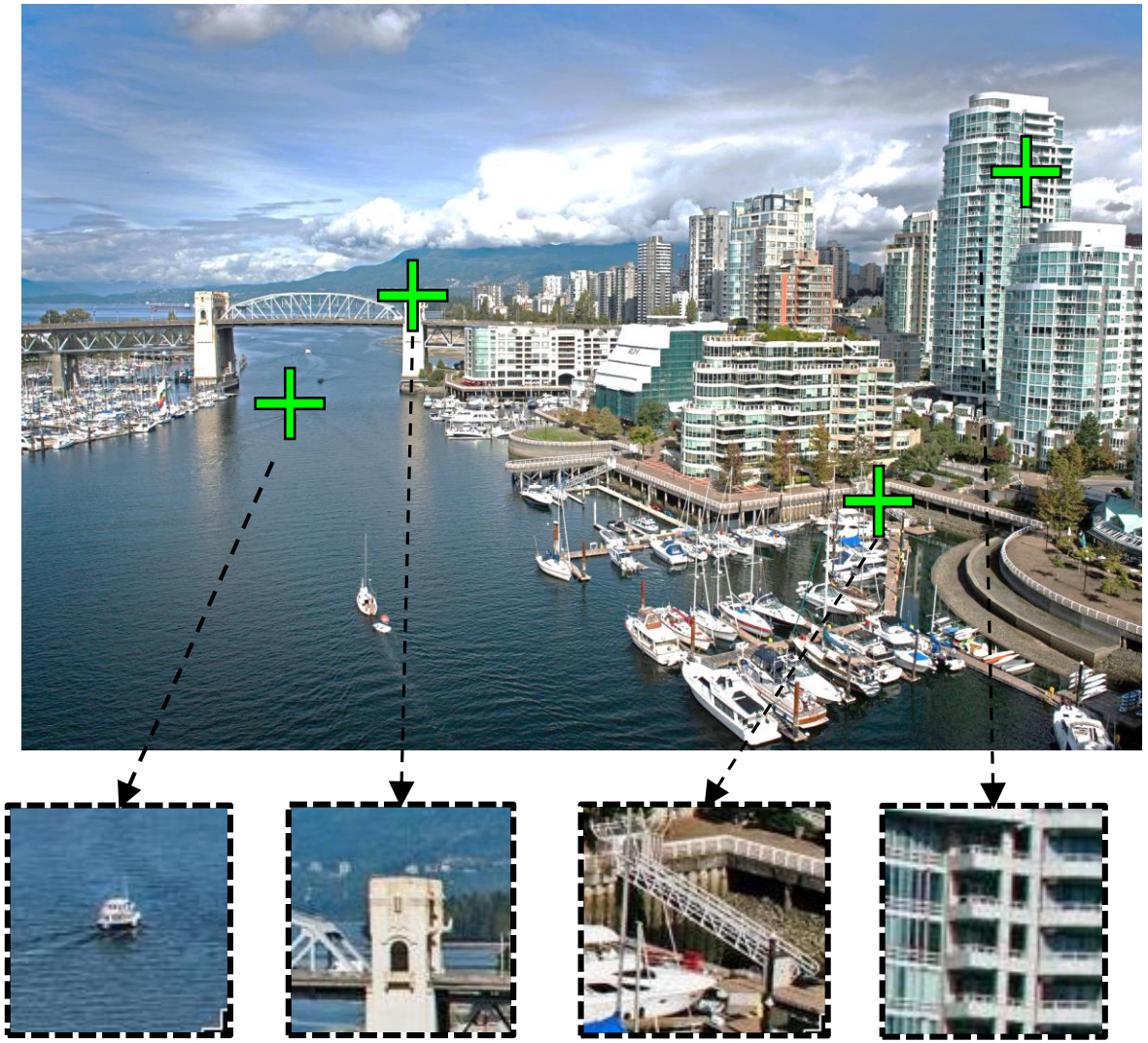
Outline

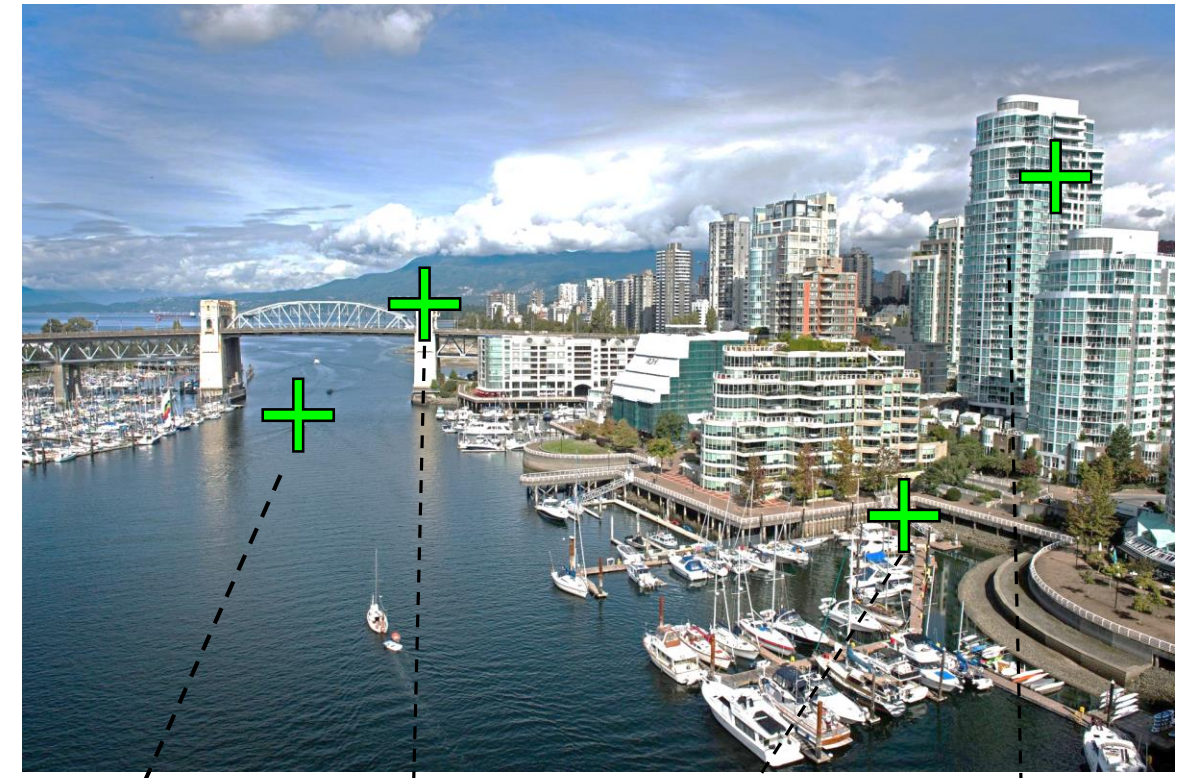
- Introduction
  - Existing methods
  - Limitations
- Proposed approach
  - Architecture
  - Training and losses
- Experimental results
  - State-of-the-art matching and localization performance

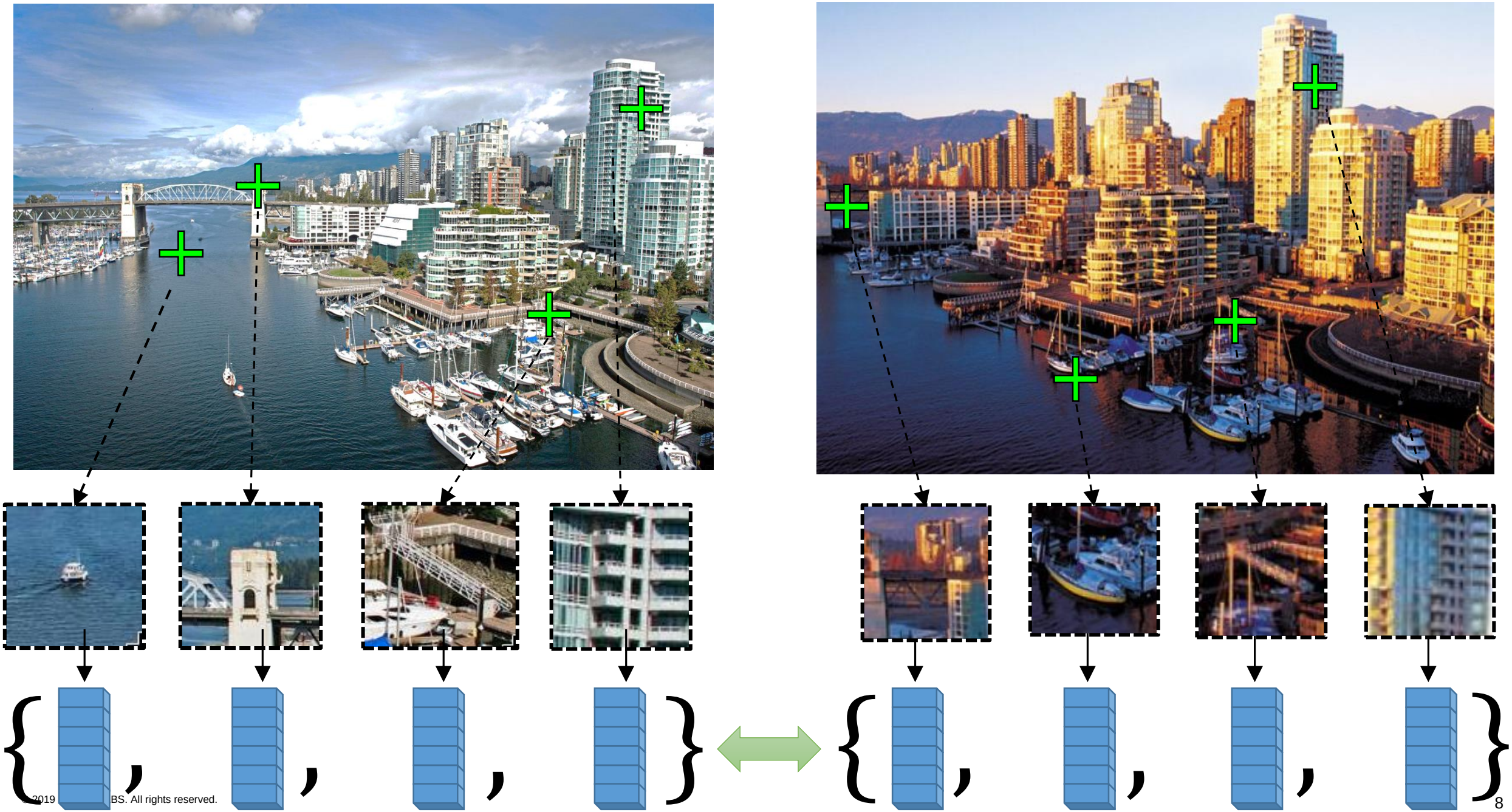
Introduction

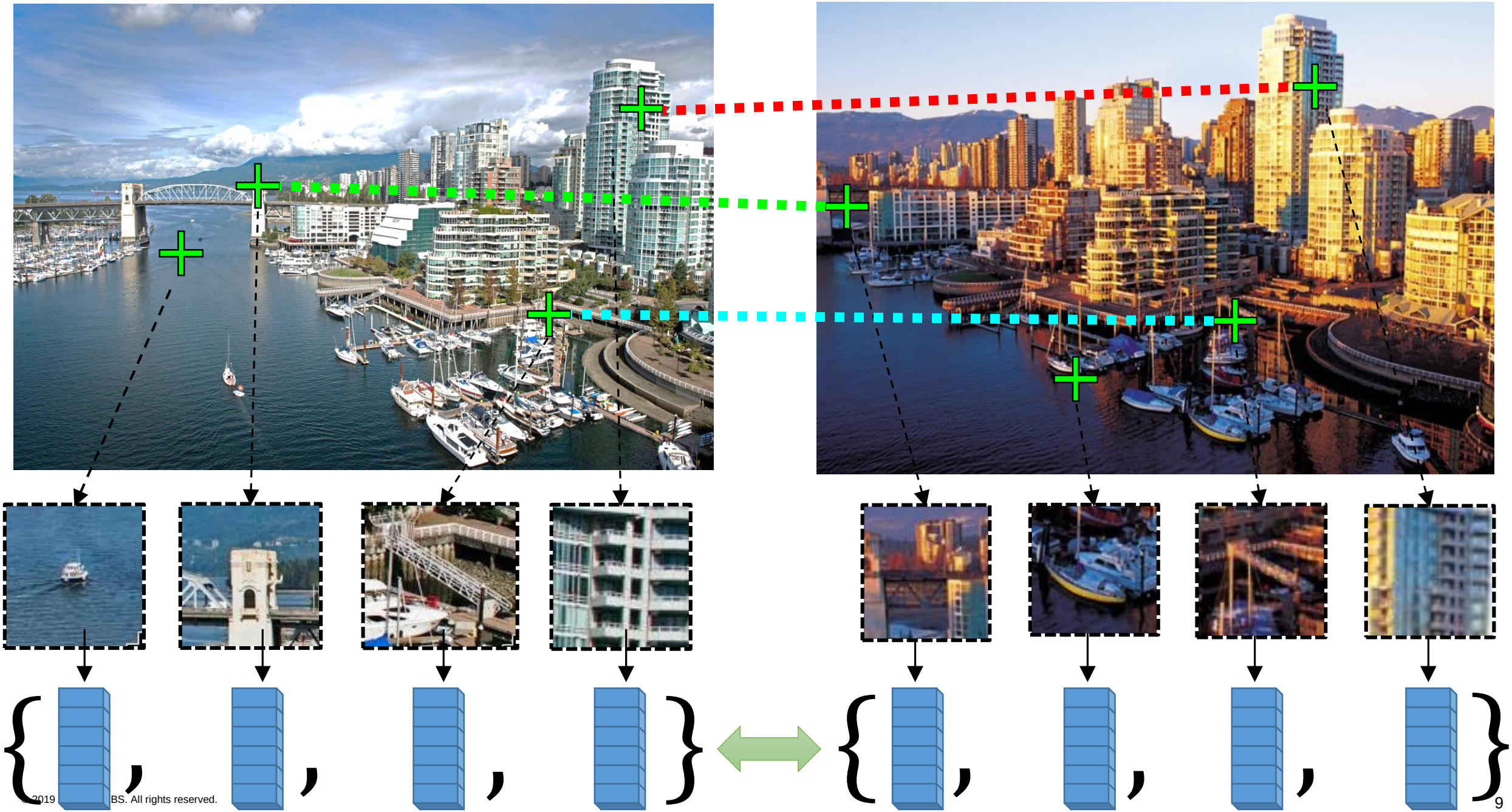








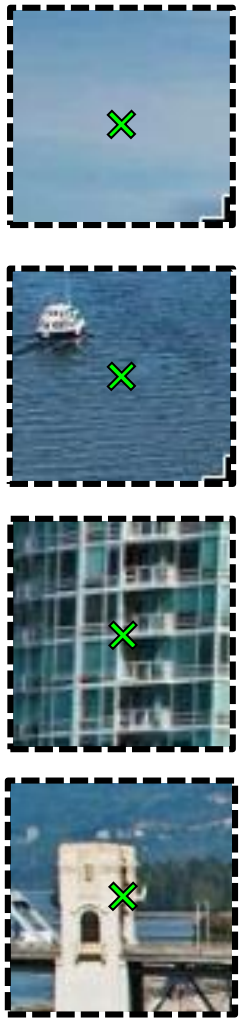




Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

What is a good keypoint?



Failure causes:

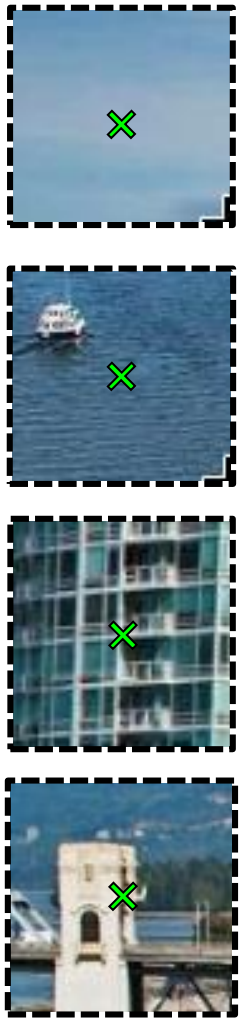
- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

What is a good keypoint?

Repeatable?



What is a good keypoint?

Repeatable?

Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

What is a good keypoint?

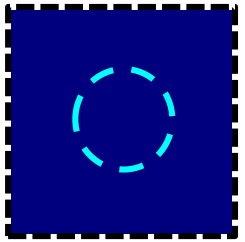
Repeatable?

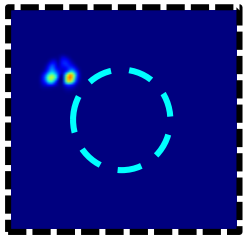
Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

What is a good keypoint?

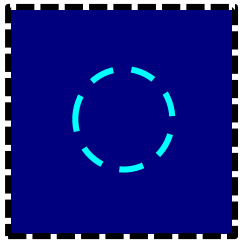
Repeatable?

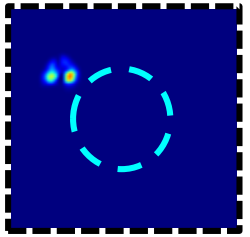




What is a good keypoint?

Repeatable?





Failure causes:

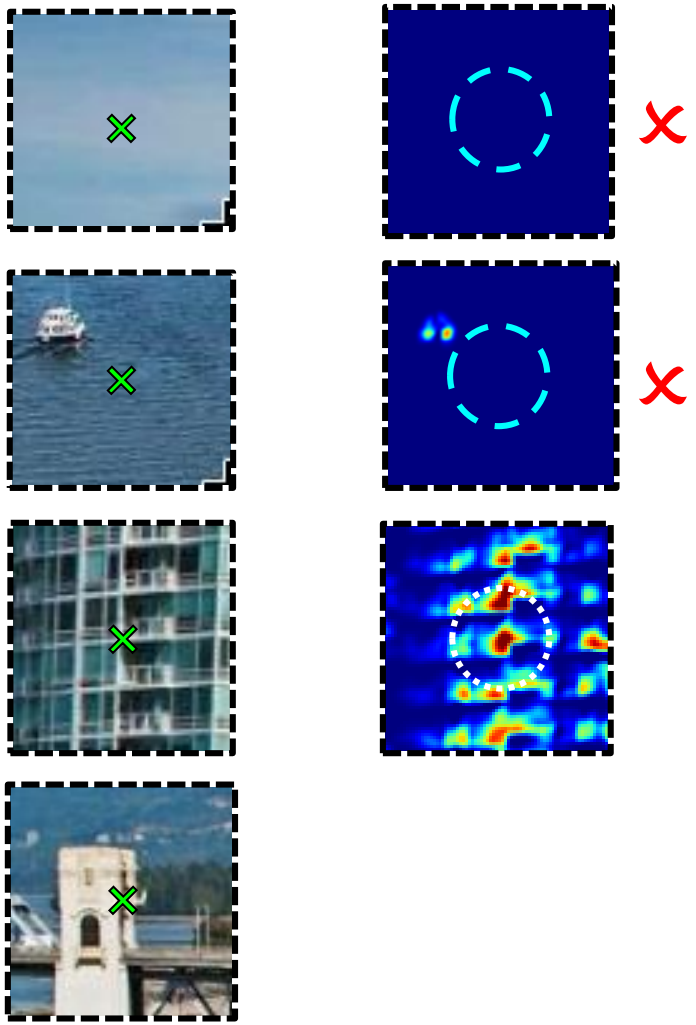
- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

What is a good keypoint?

Repeatable?

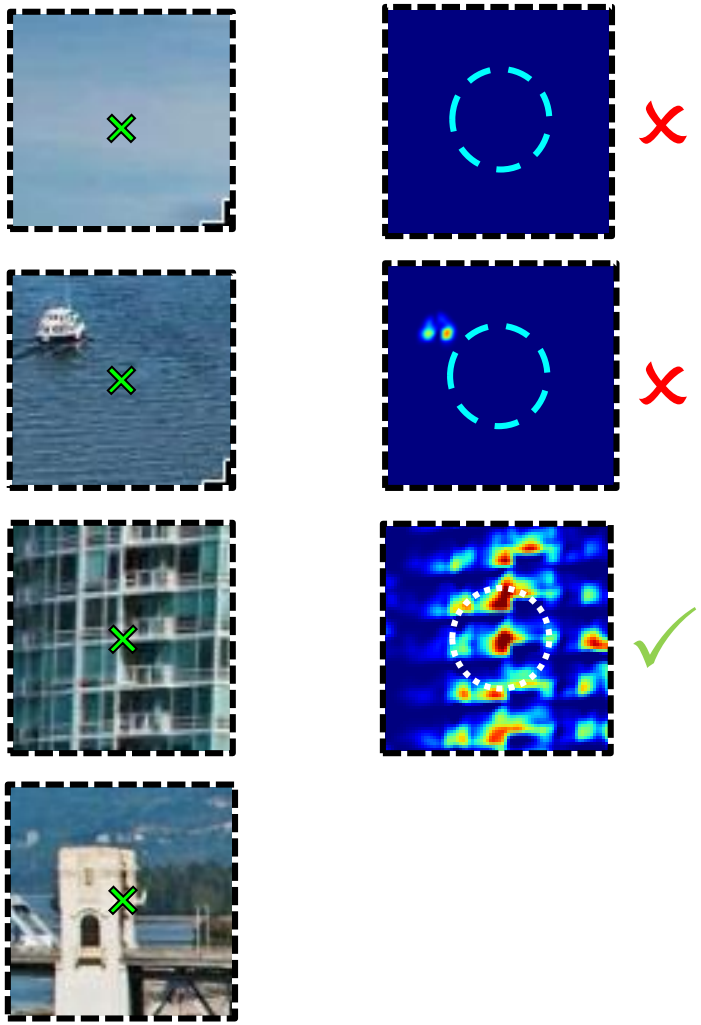


Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

What is a good keypoint?

Repeatable?

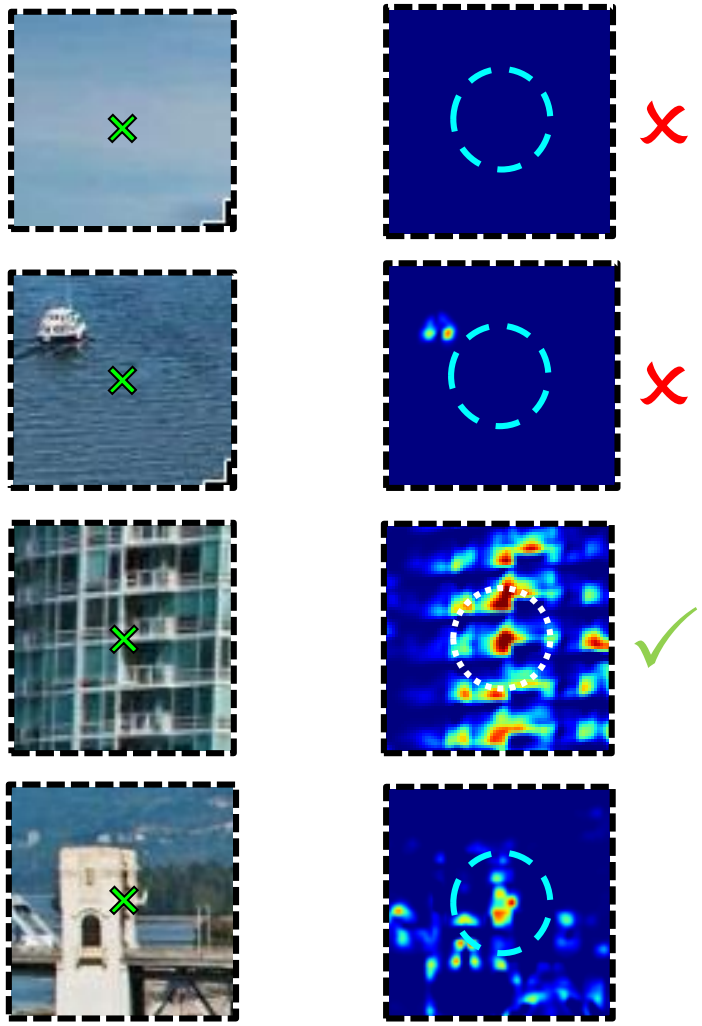


Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

What is a good keypoint?

Repeatable?

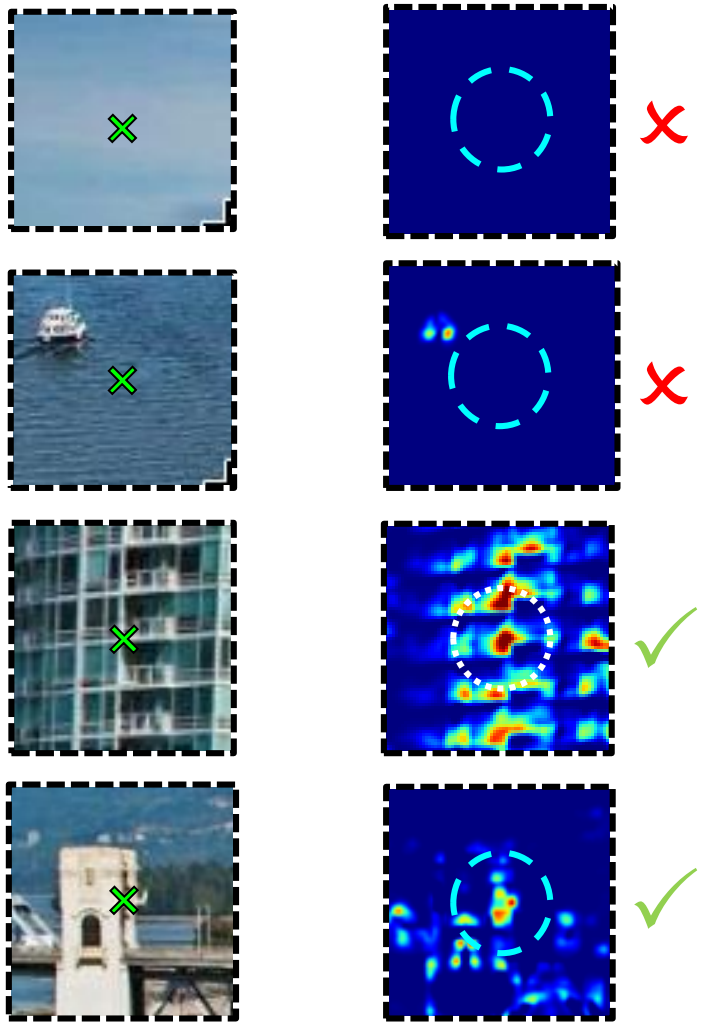


Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

What is a good keypoint?

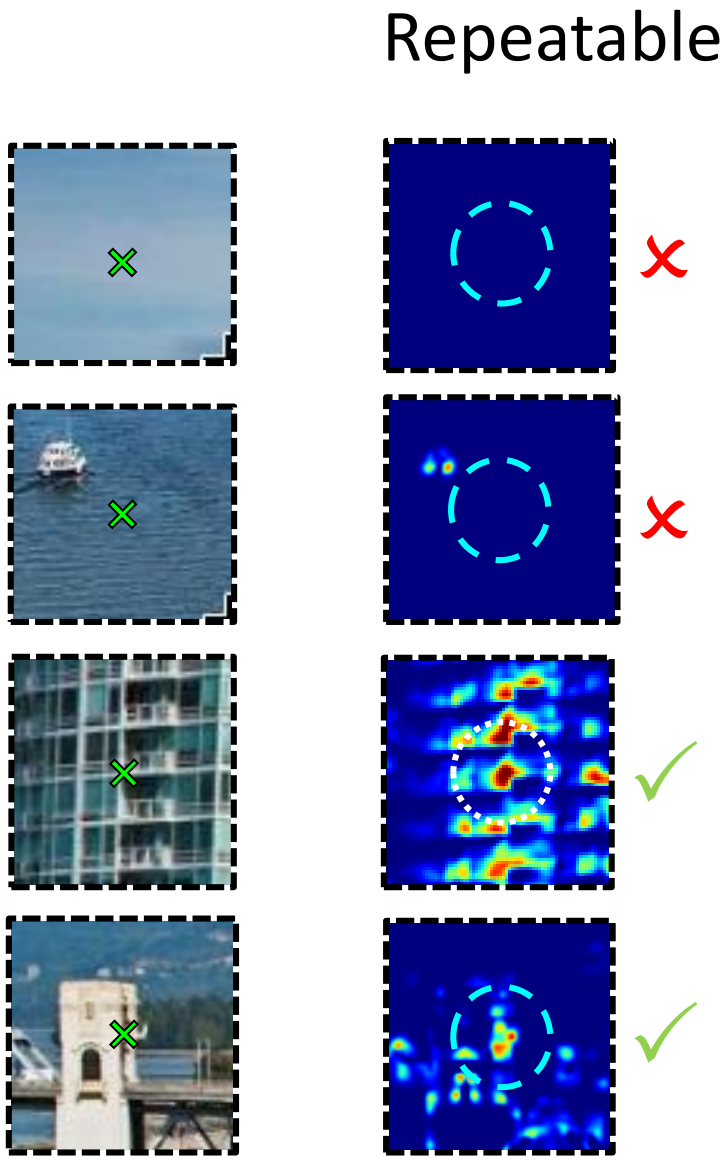
Repeatable?



Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

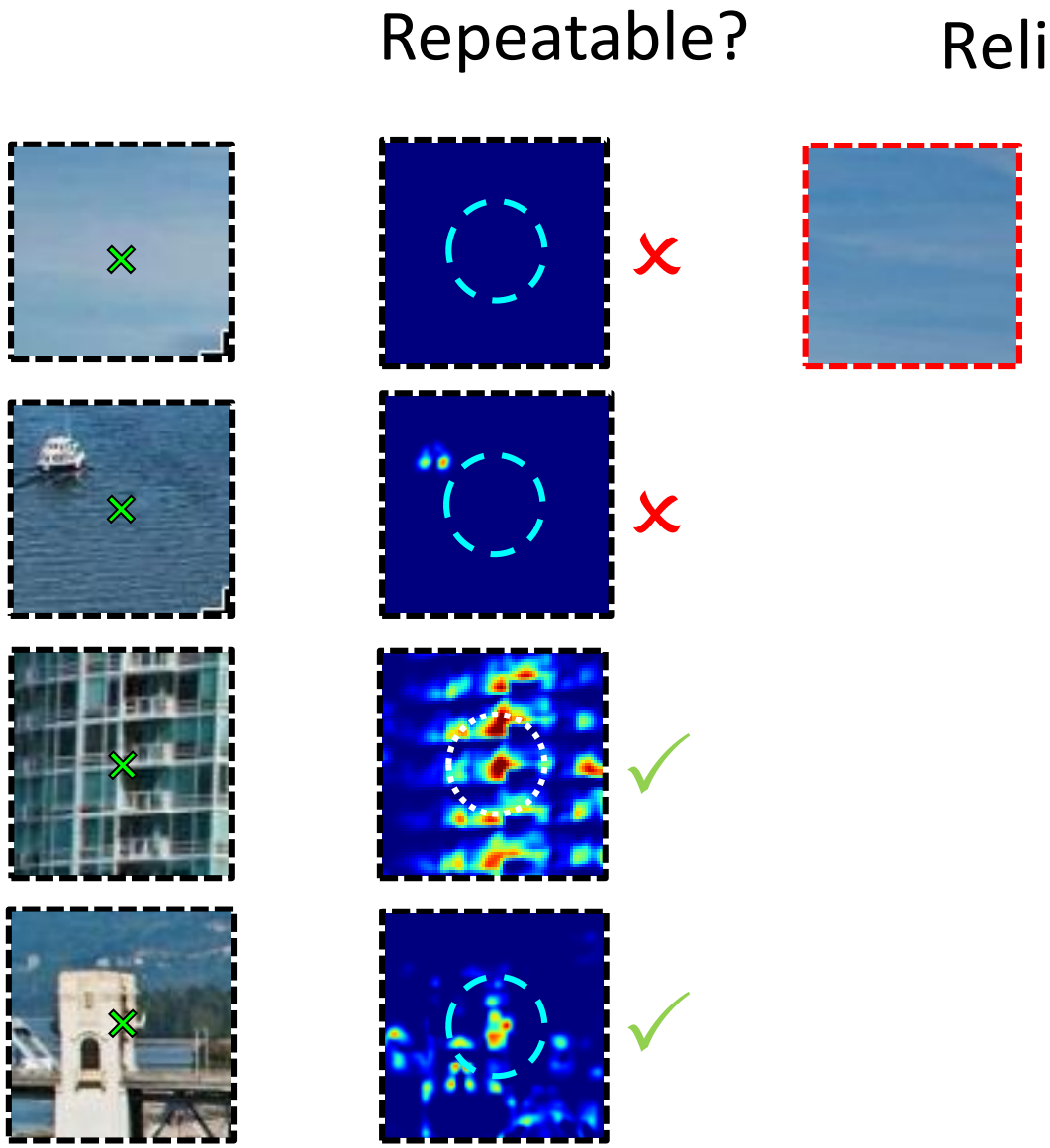
What is a good keypoint?



Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

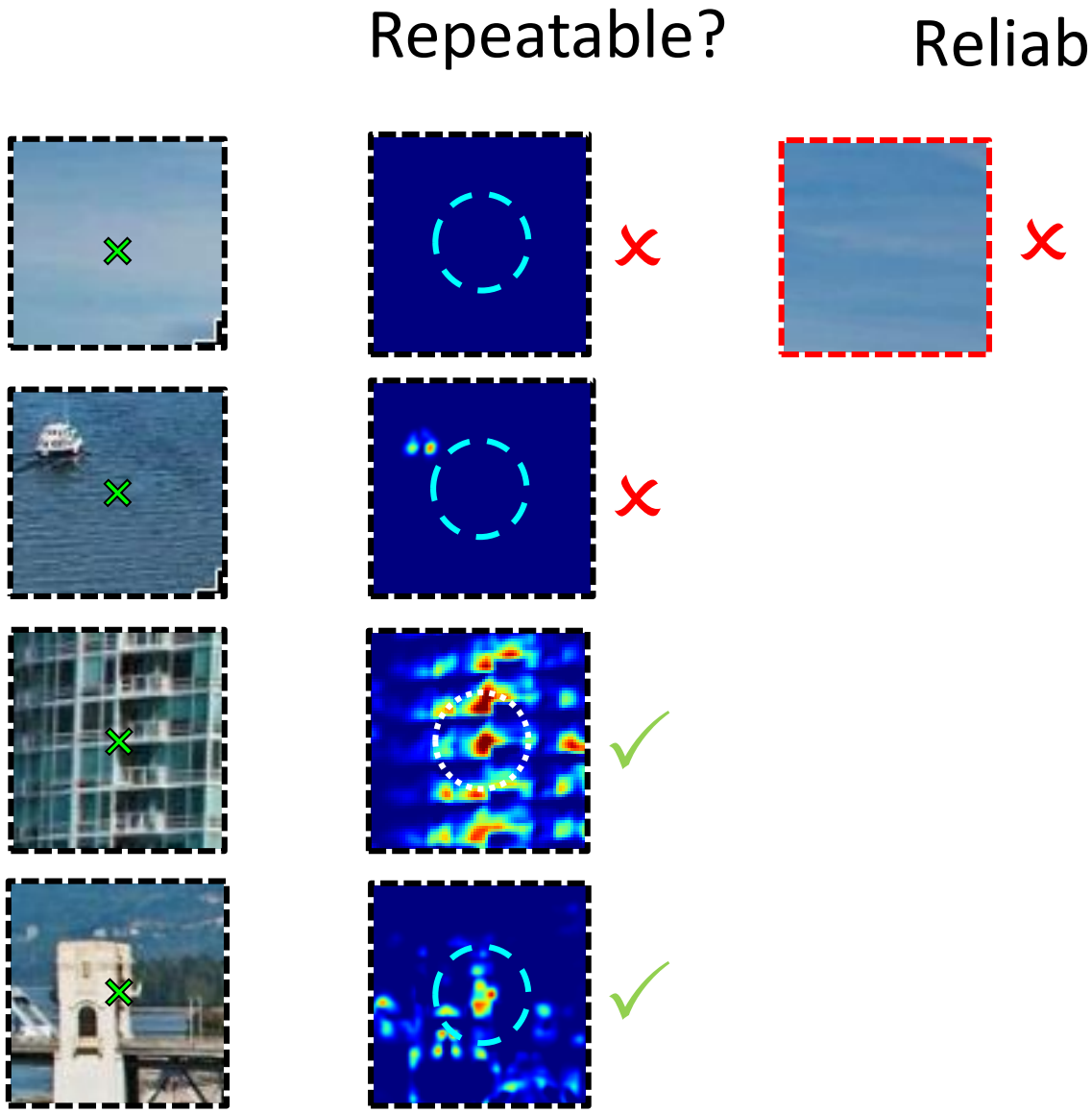
What is a good keypoint?

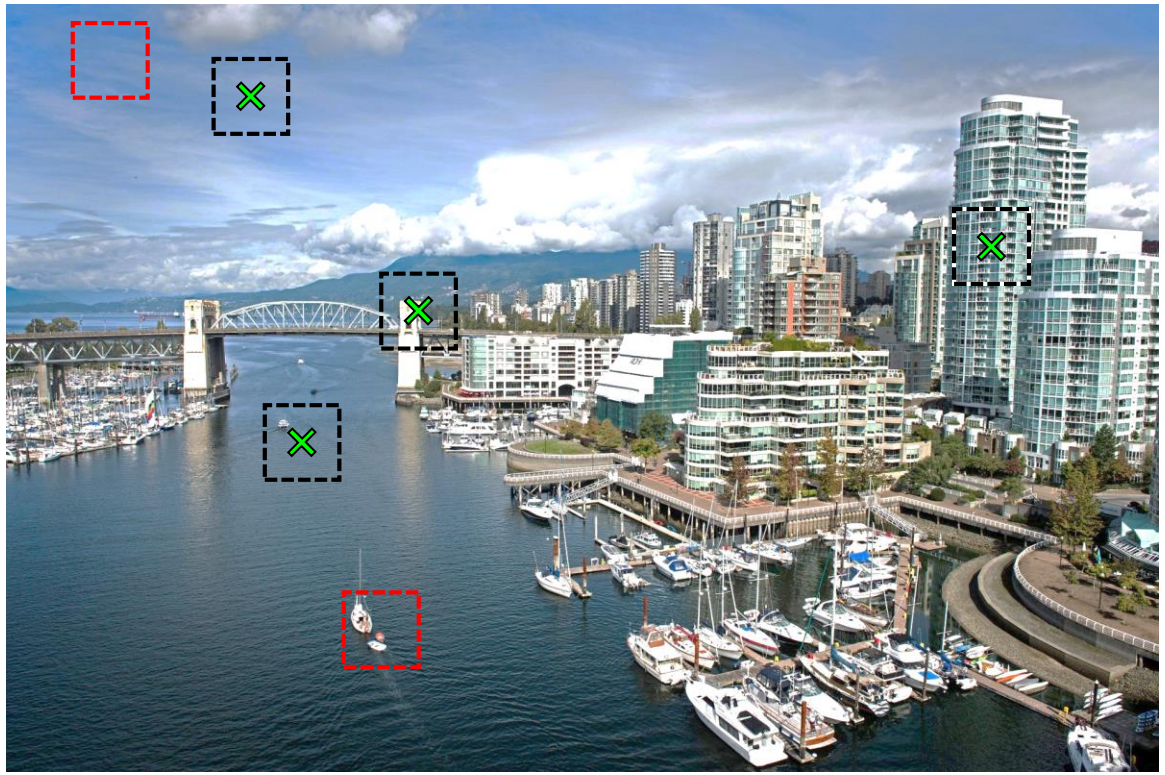


Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

What is a good keypoint?

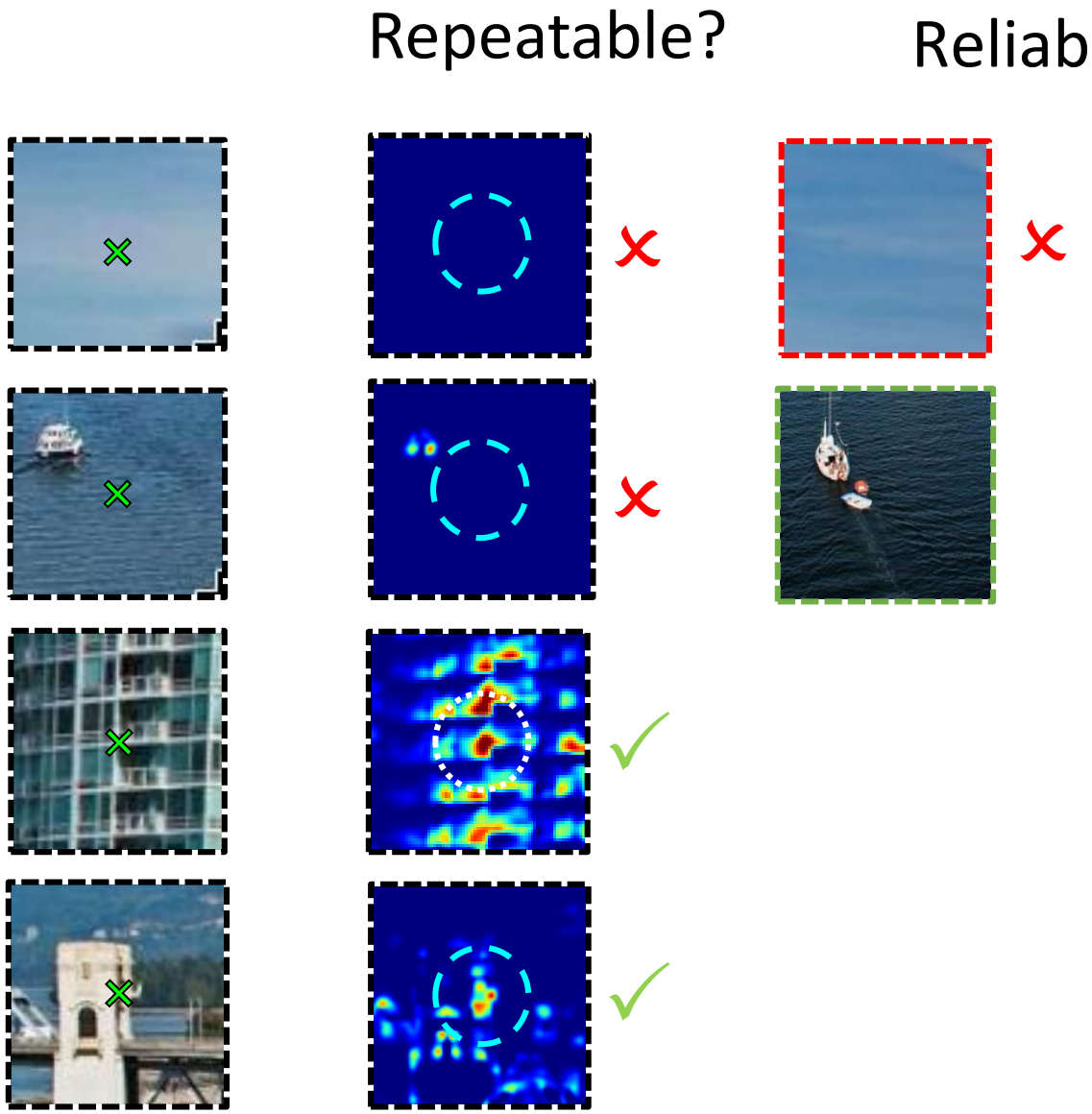


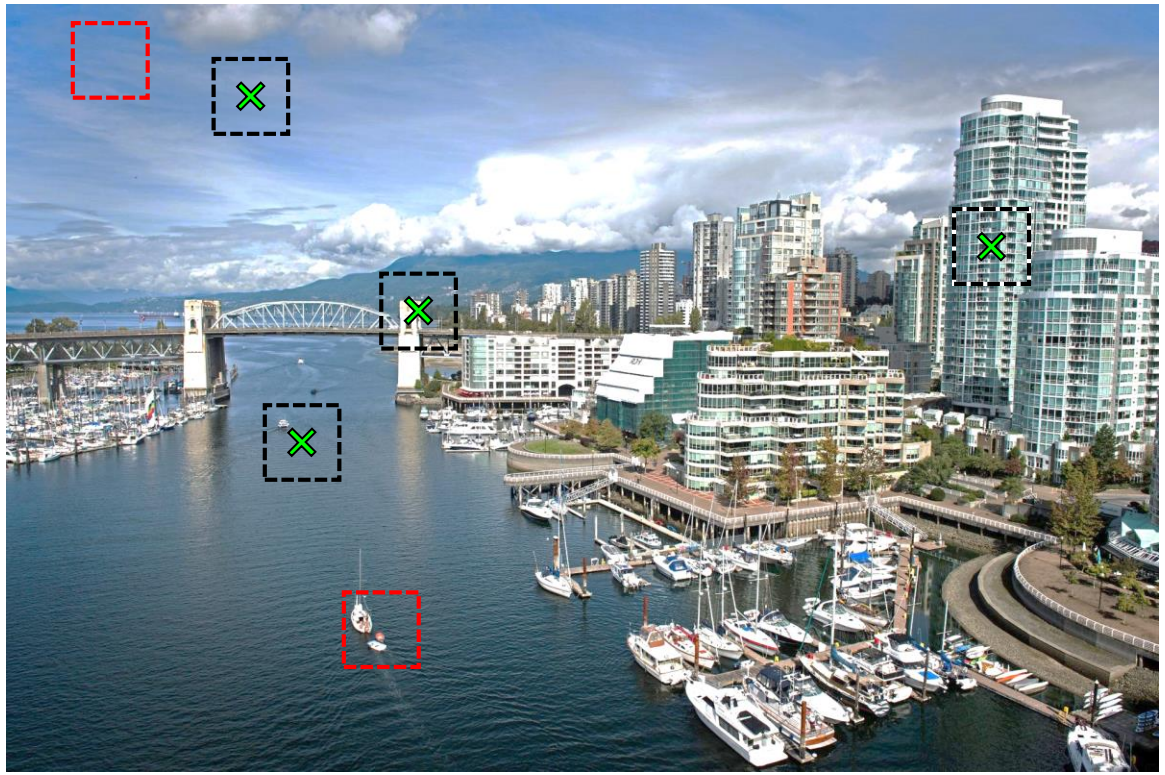


Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

What is a good keypoint?

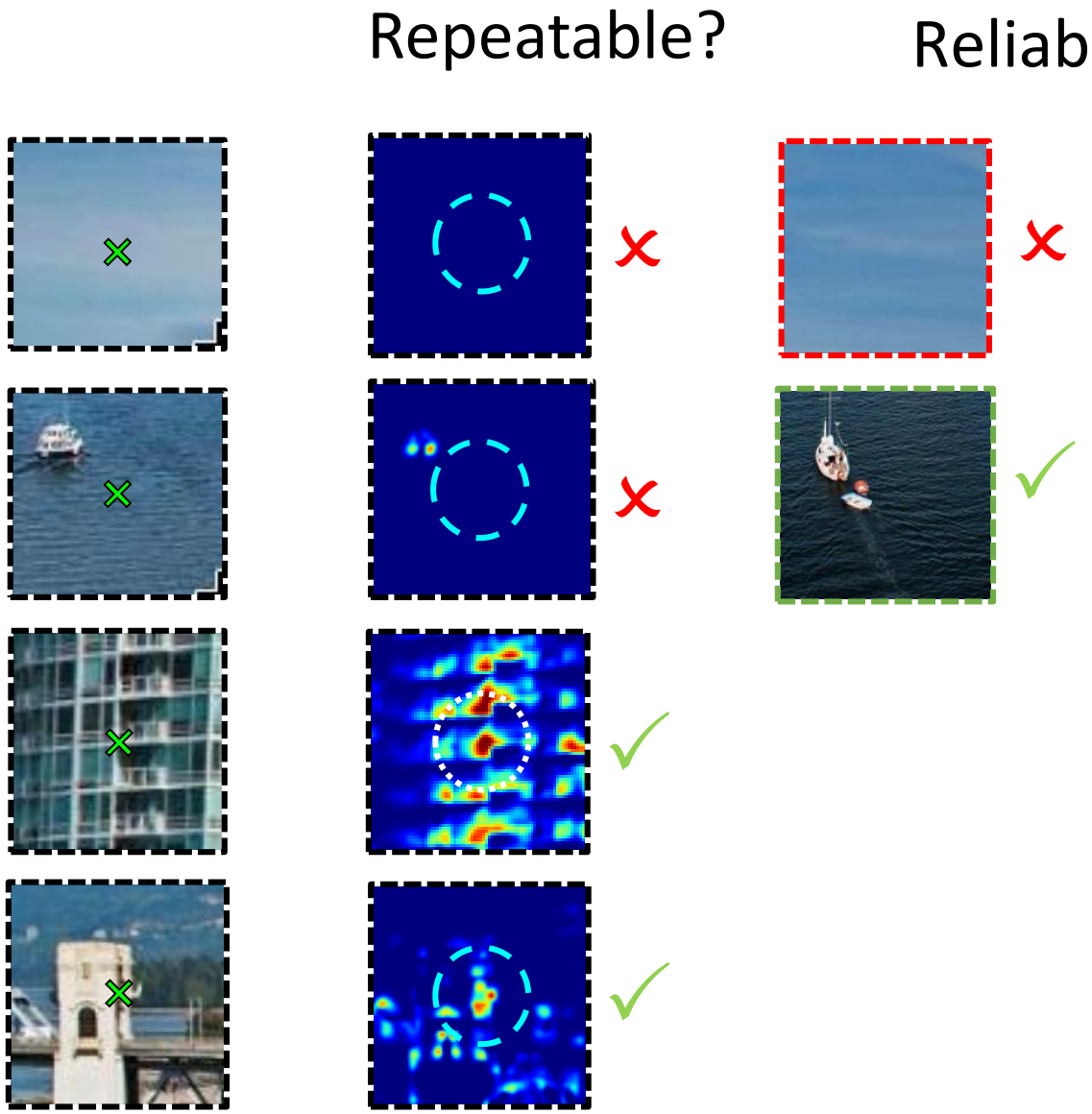


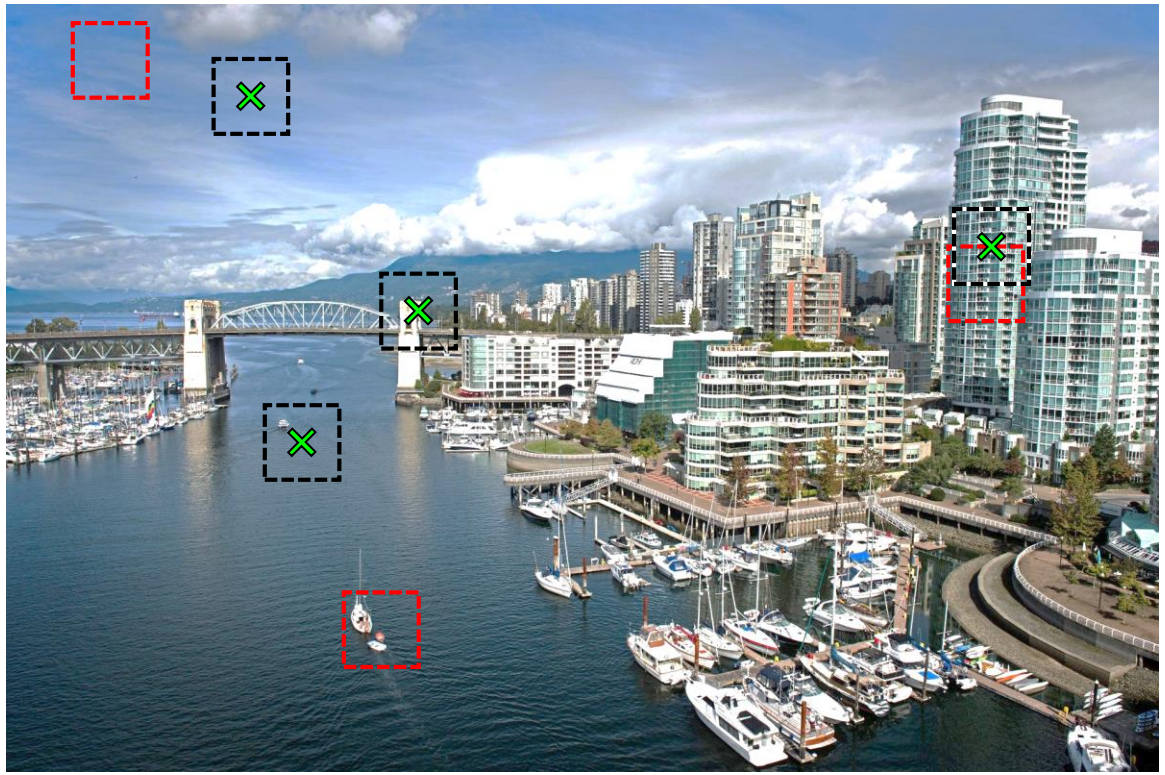


Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

What is a good keypoint?

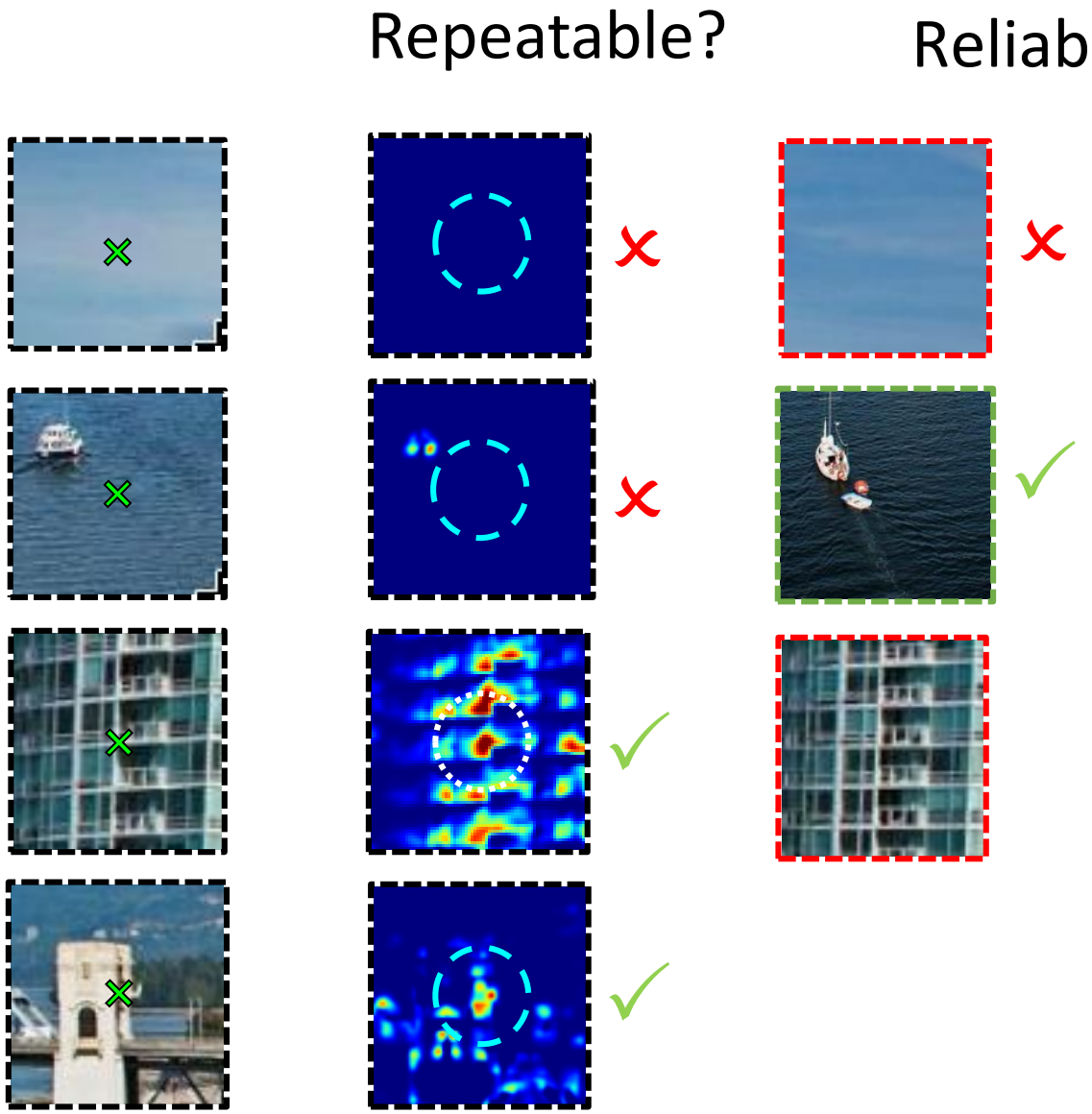


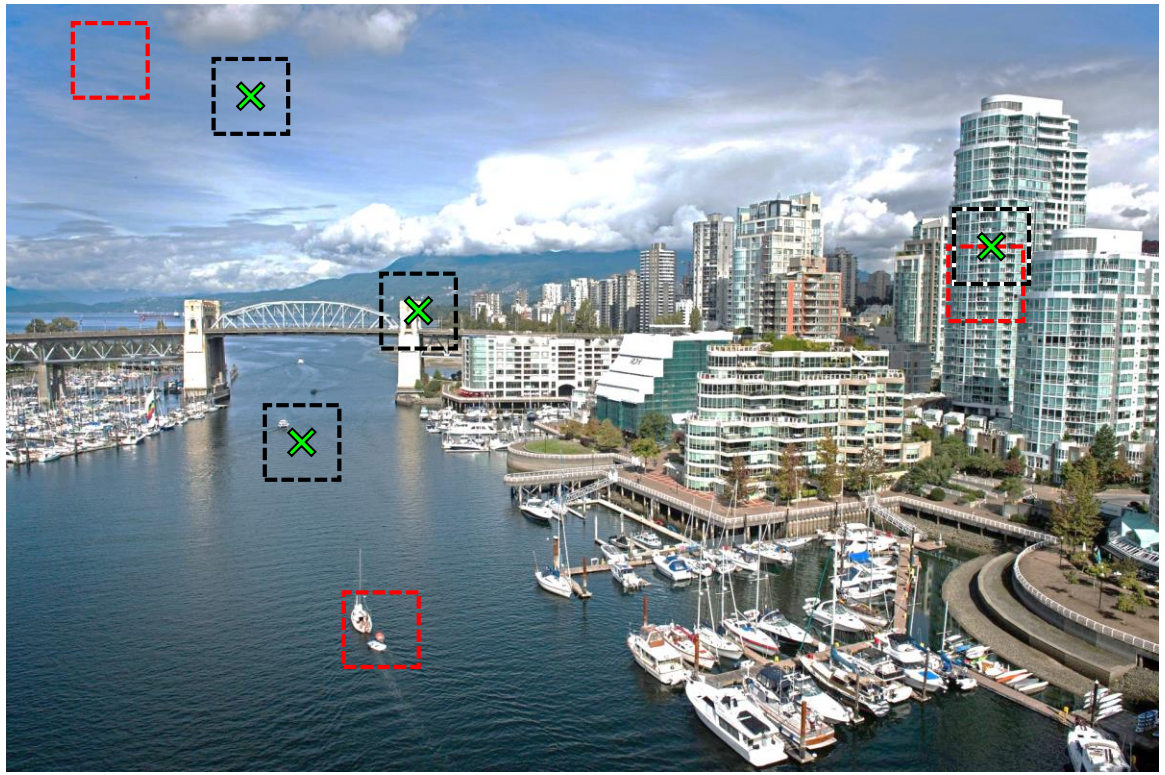


Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

What is a good keypoint?

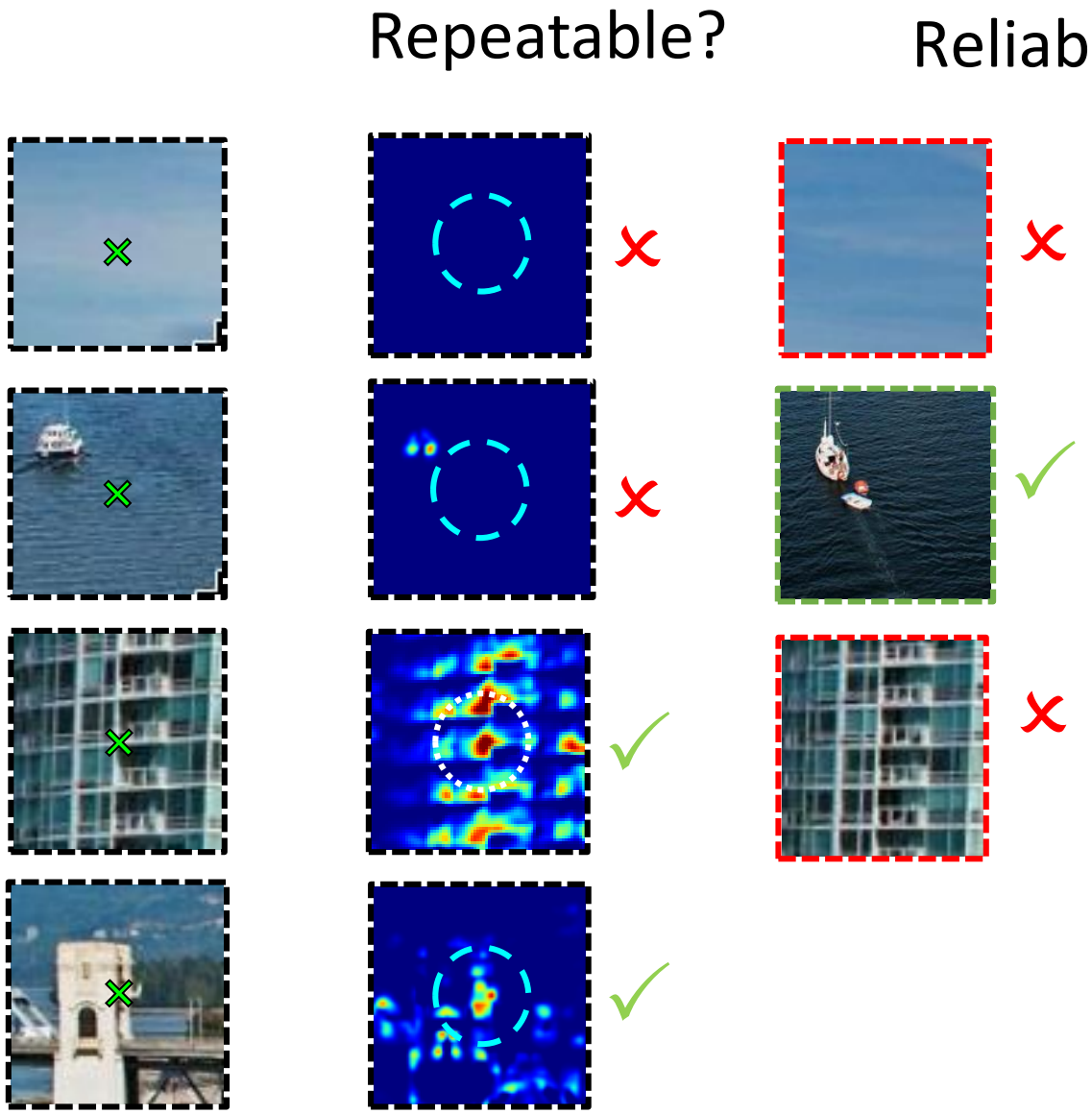


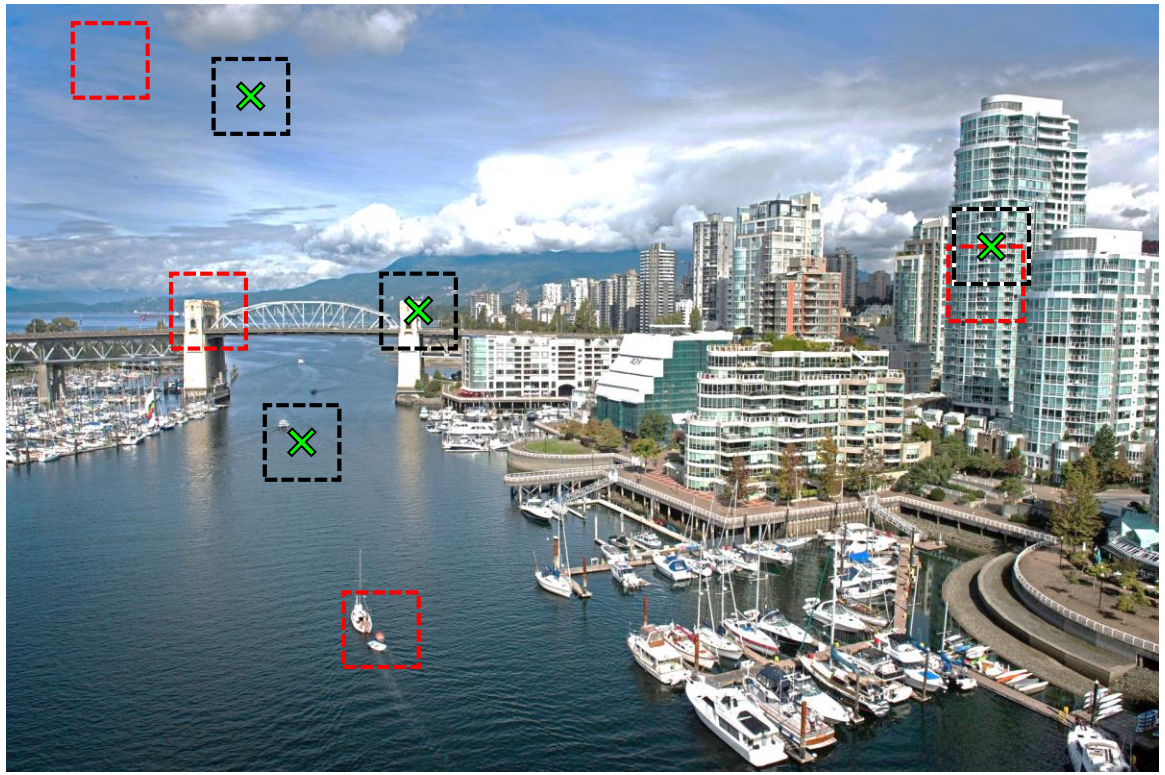


Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

What is a good keypoint?

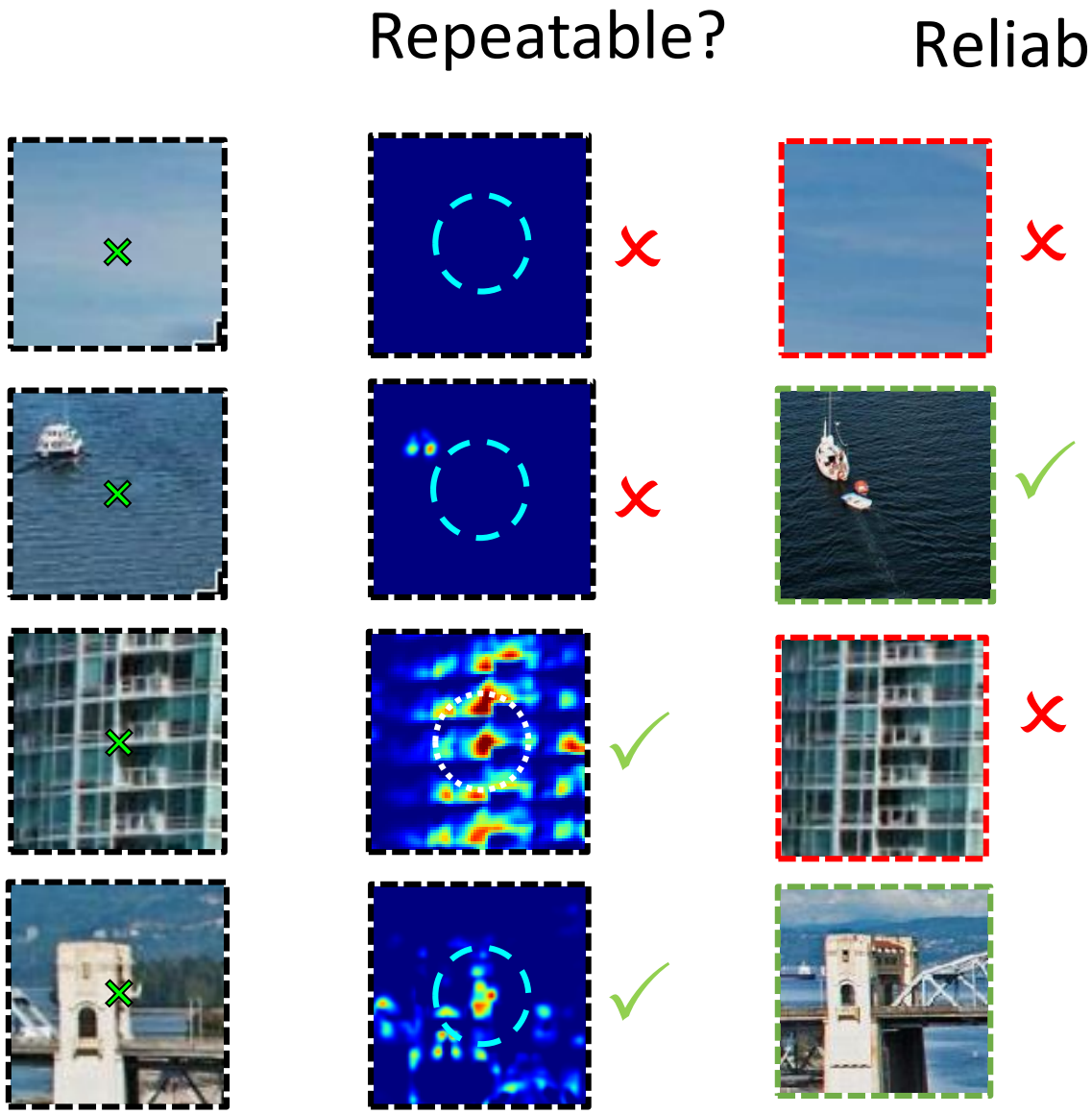


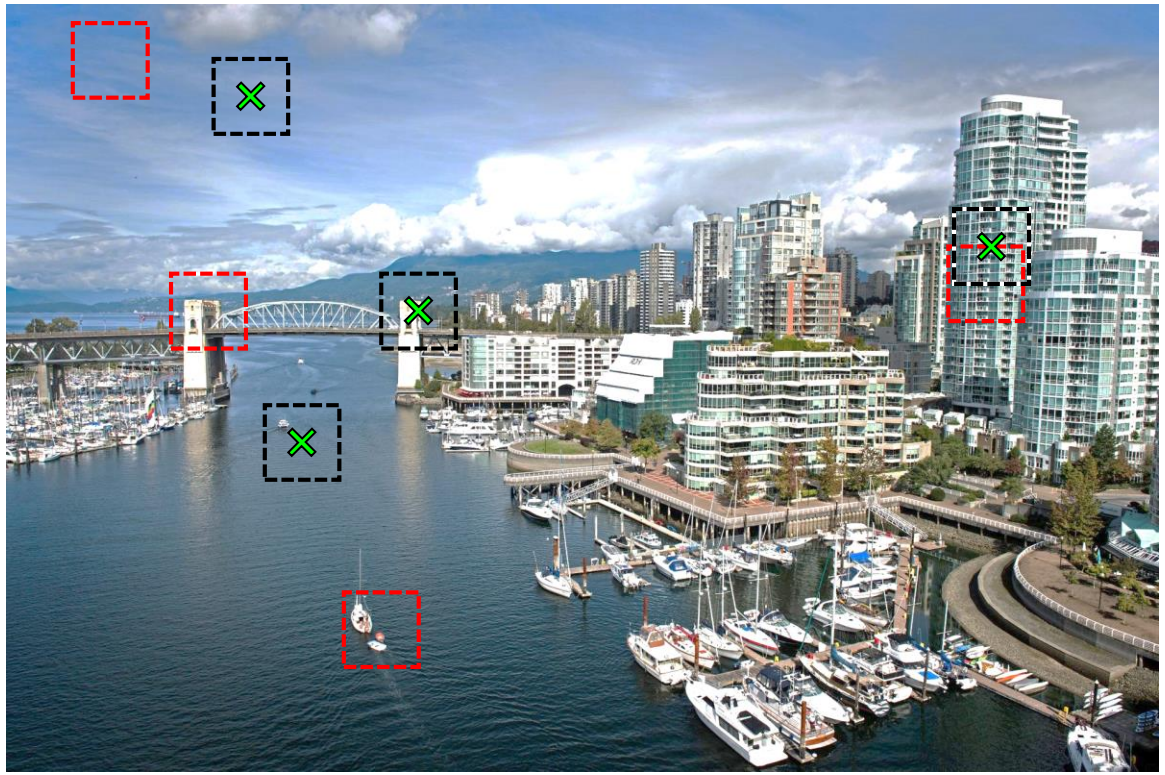


Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

What is a good keypoint?

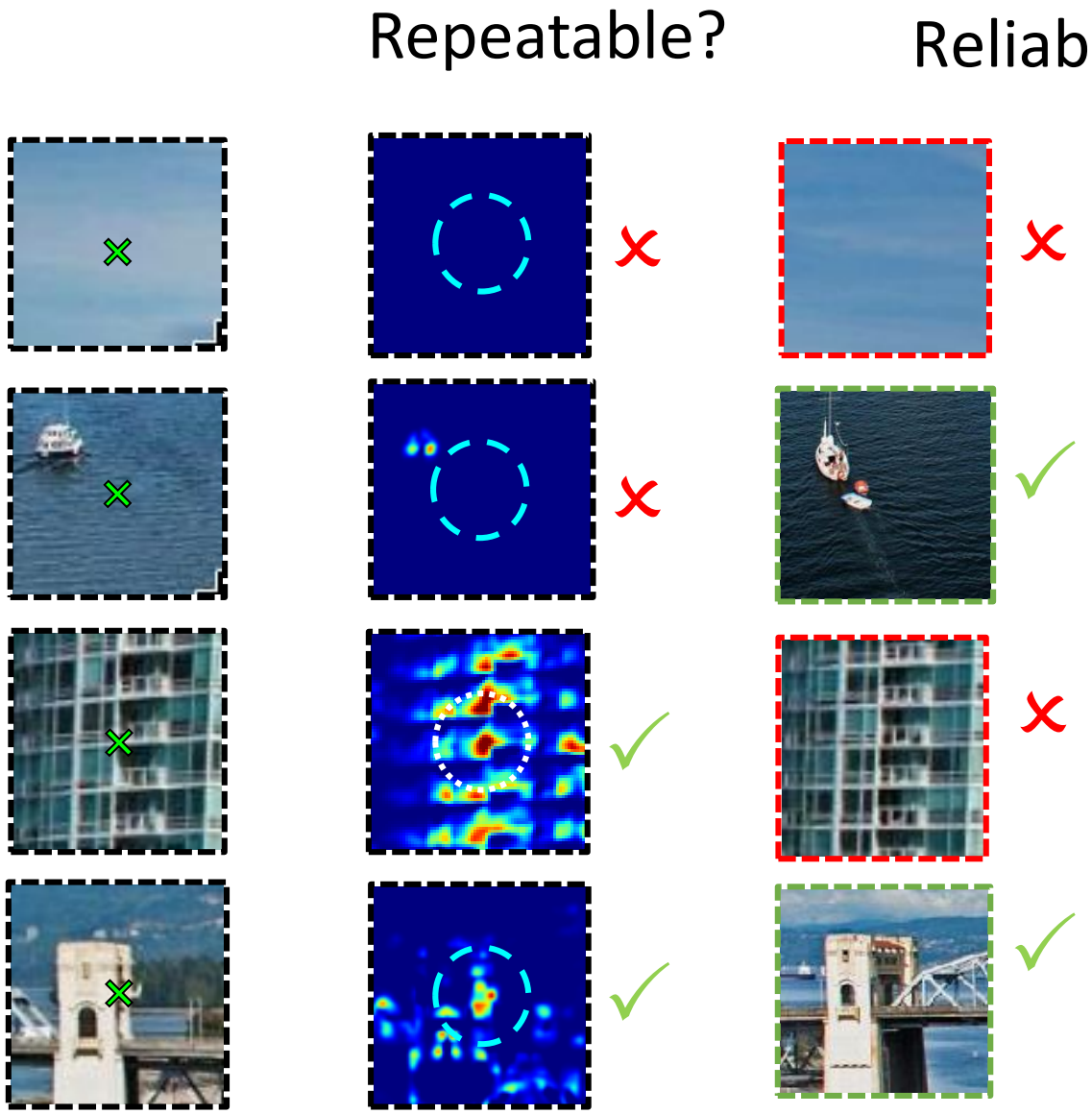




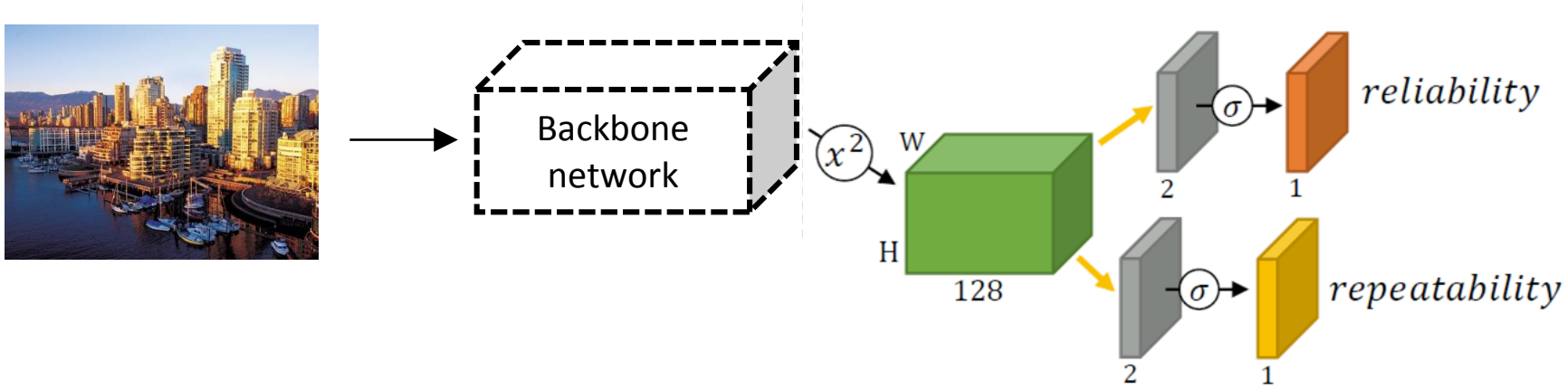
Failure causes:

- The keypoint detector only focuses on *repeatable* locations
- But repeatable locations are not necessarily *reliable* for matching.

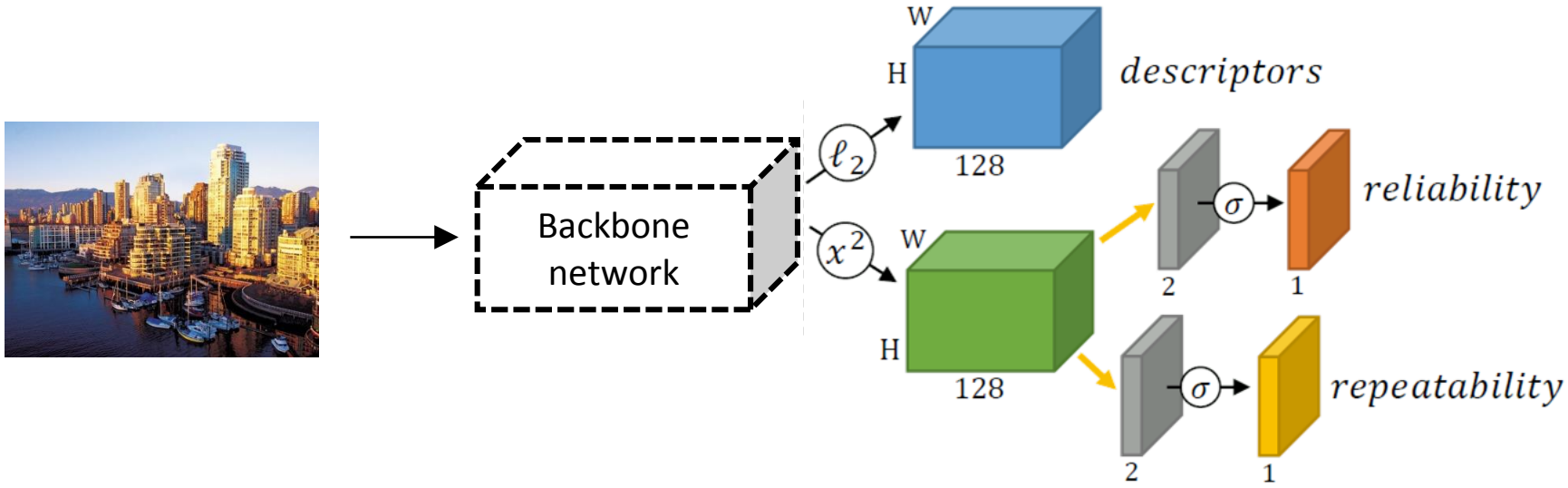
What is a good keypoint?



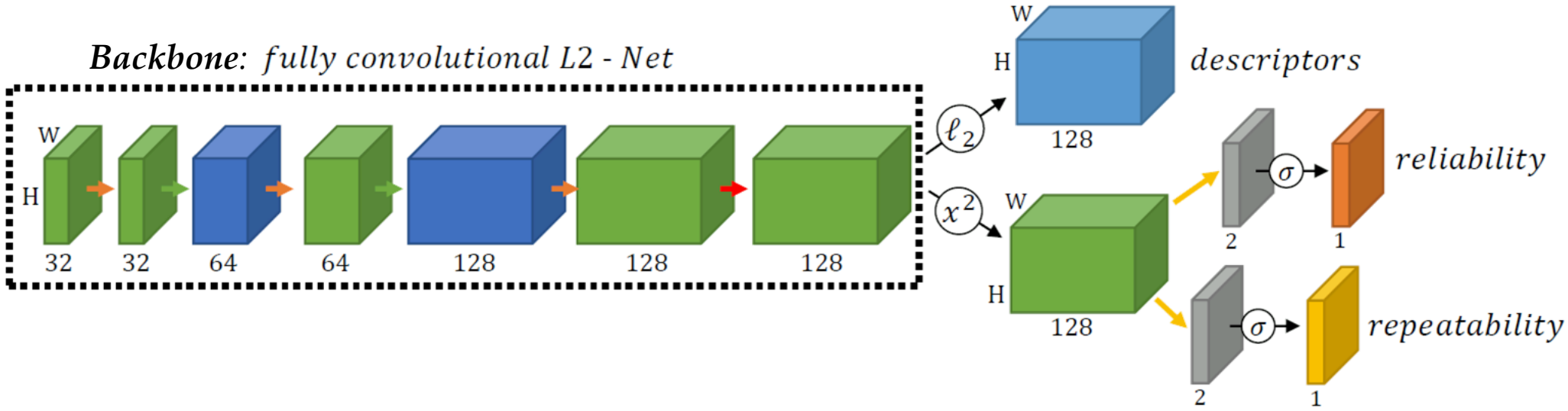
Proposed architecture



Proposed architecture



Proposed architecture

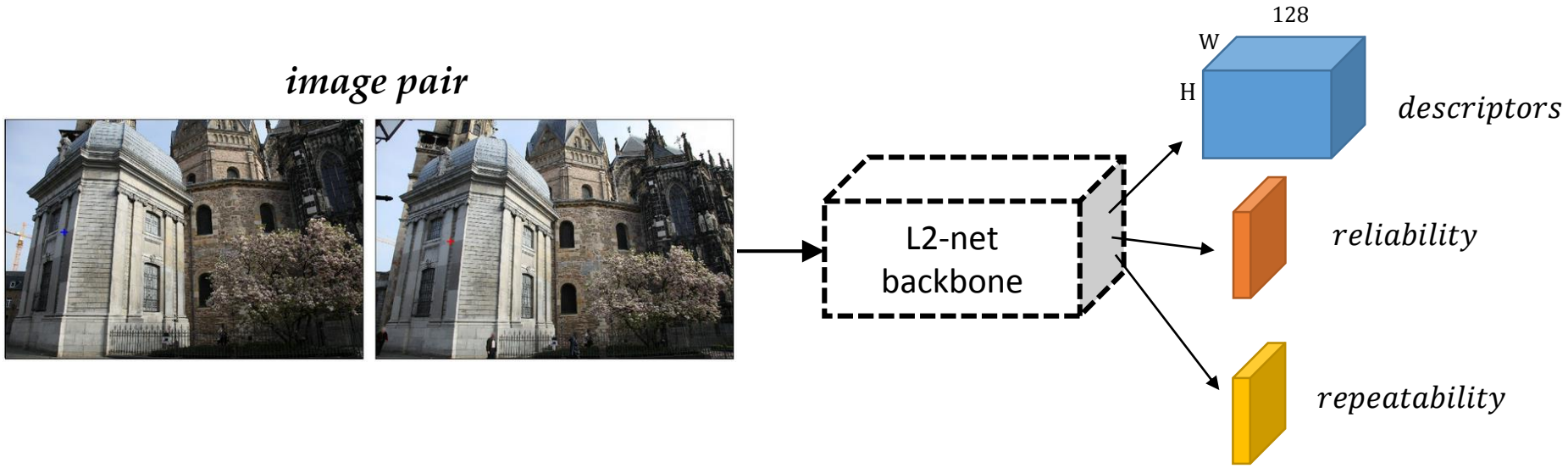


L2-Net: Deep learning of discriminative patch descriptor in euclidean space. Y. Tian, B. Fan, and F. Wu. CVPR, 2017.

Contributions

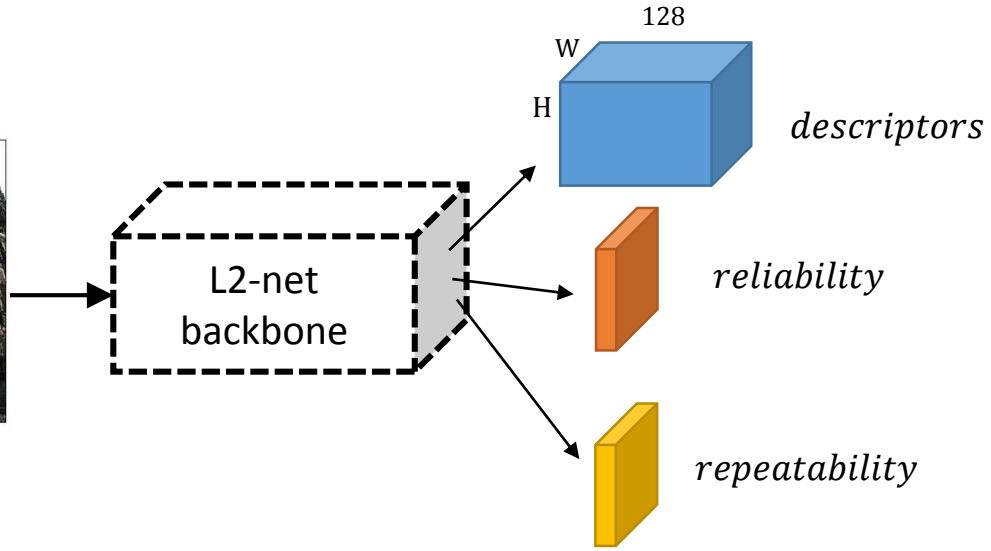
- We introduce keypoint reliability
  - “Is this keypoint good for matching?”
  - Jointly predicted along with repeatability
- Novel training scheme
  - Two novel losses
  - Training from scratch, without annotations, no bias
- State-of-the-art results
  - Matching & visual localization
  - Even when training without annotations

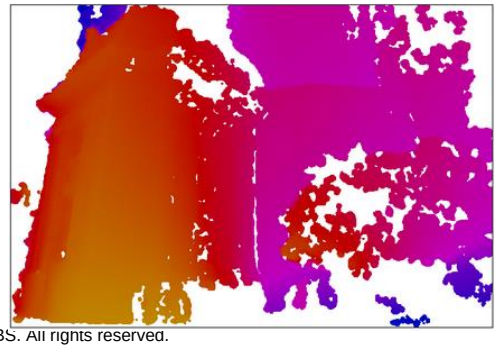
Training the network



Training the network

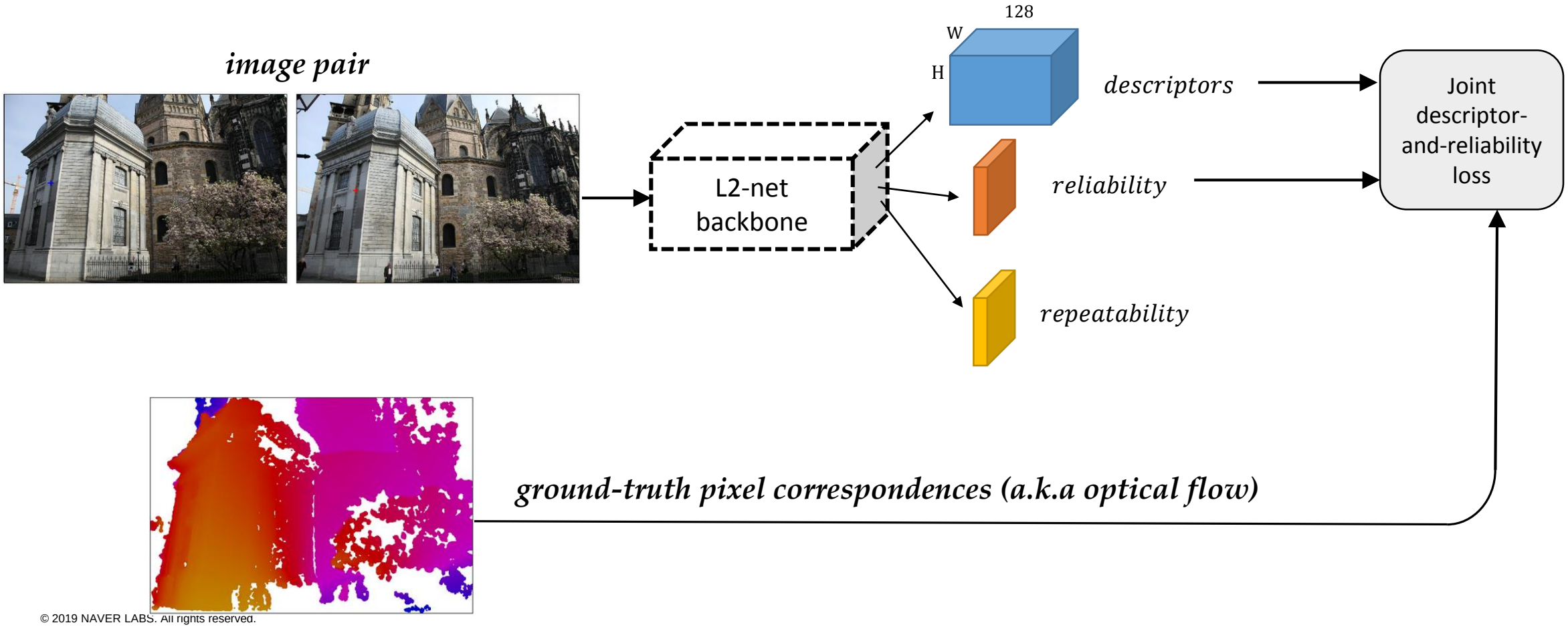
image pair



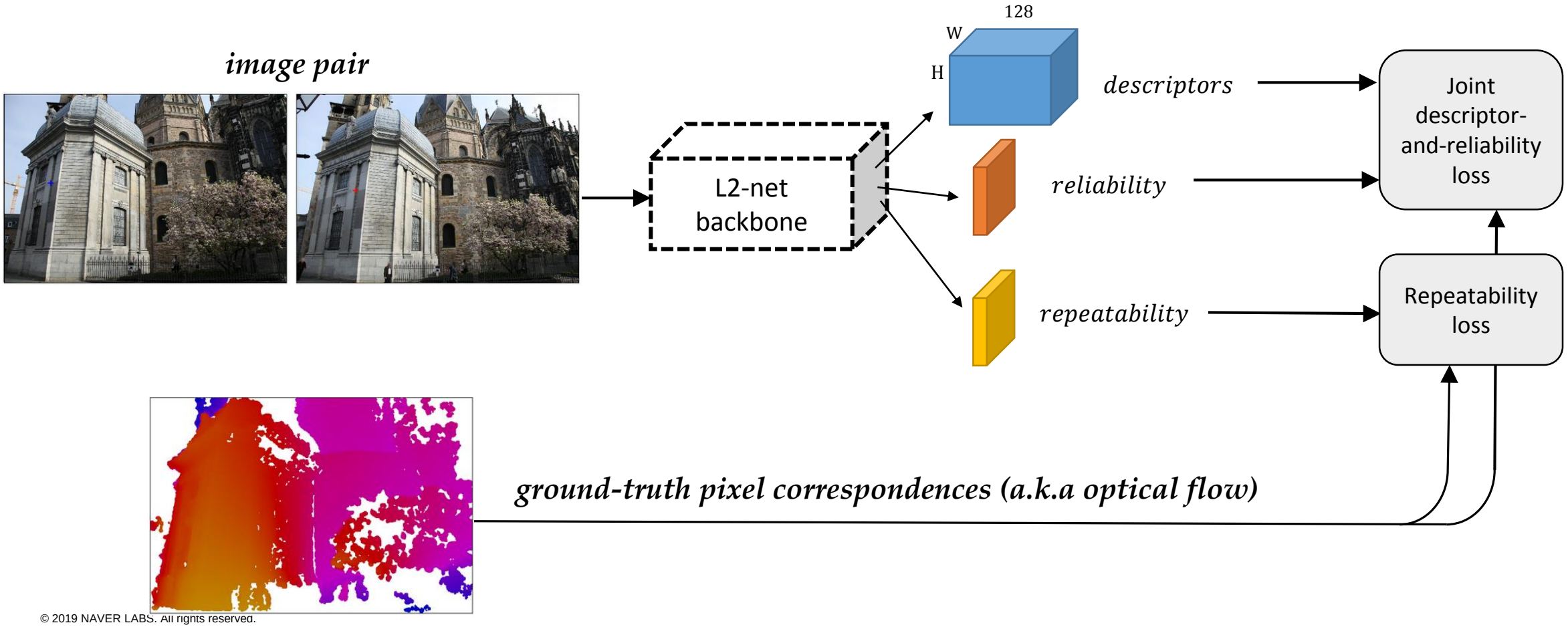


ground-truth pixel correspondences (a.k.a optical flow)

Training the network



Training the network



Joint descriptor & reliability loss

- Based on the differentiable AP loss
  - originally proposed by He et al. [1]

[1] K. He, Y. Lu, and S. Sclaroff. Local descriptors optimized for average precision. In CVPR, 2018.

Joint descriptor & reliability loss

- Based on the differentiable AP loss
  - originally proposed by He et al. [1]
- Given a query descriptor p_{ij} from image I_1
  - We compare it to **all** descriptors in image I_2:
    - 1 positive, and **many** negatives
  - We compute the AP = $\widetilde{AP}(p_{ij})$

[1] K. He, Y. Lu, and S. Sclaroff. Local descriptors optimized for average precision. In CVPR, 2018.

Joint descriptor & reliability loss

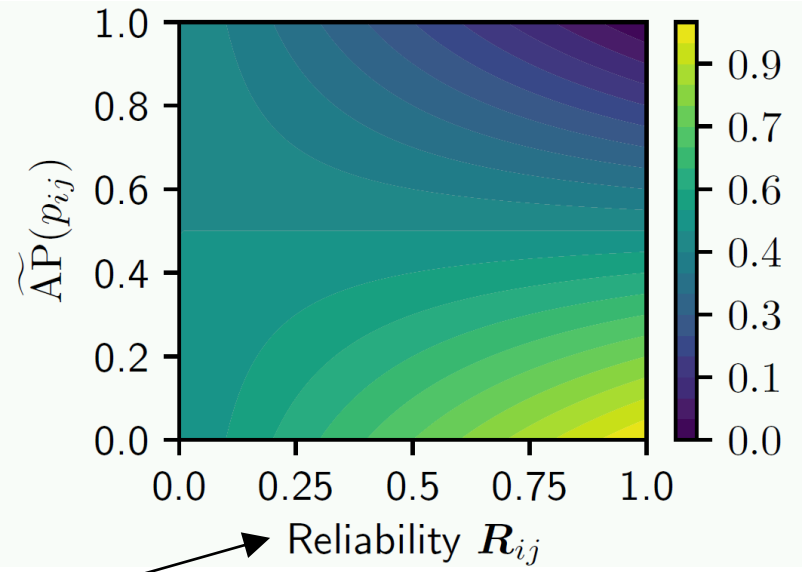
- Based on the differentiable AP loss
  - originally proposed by He et al. [1]
- Given a query descriptor p_{ij} from image I_1
  - We compare it to **all** descriptors in image I_2:
    - 1 positive, and **many** negatives
  - We compute the AP = $\widetilde{AP}(p_{ij})$

- Modified to not waste efforts on bad regions
  - We estimate the reliability at $p_{ij} = R_{ij}$
  - Many regions can't be matched (empty, 1-d pattern, repetitive…)
  - For these region, reliability is low ➔ the loss is almost flat

[1] K. He, Y. Lu, and S. Sclaroff. Local descriptors optimized for average precision. In CVPR, 2018.

Joint descriptor & reliability loss

- Based on the differentiable AP loss
  - originally proposed by He et al. [1]
- Given a query descriptor p_{ij} from image I_1
  - We compare it to **all** descriptors in image I_2:
    - 1 positive, and **many** negatives
  - We compute the AP = $\widetilde{AP}(p_{ij})$

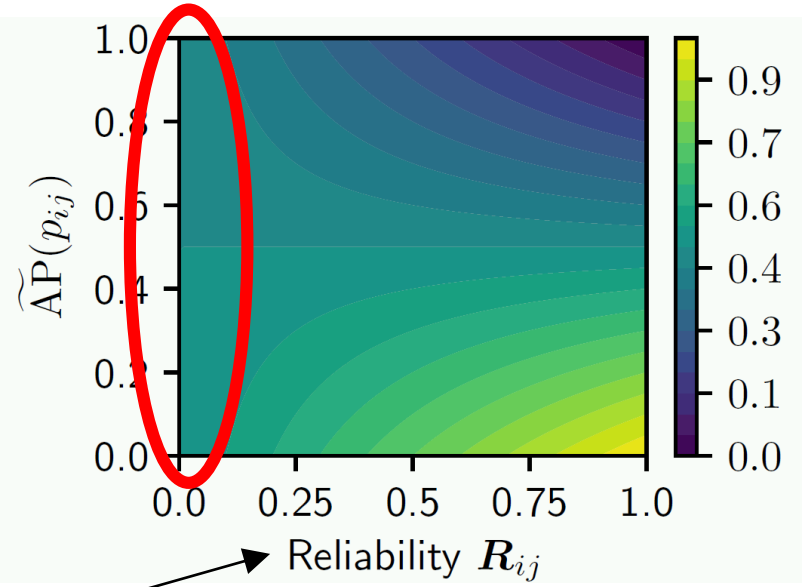


- Modified to not waste efforts on bad regions
  - We estimate the reliability at p_{ij} = R_{ij}
  - Many regions can't be matched (empty, 1-d pattern, repetitive…)
  - For these region, reliability is low ➔ the loss is almost flat

[1] K. He, Y. Lu, and S. Sclaroff. Local descriptors optimized for average precision. In CVPR, 2018.

Joint descriptor & reliability loss

- Based on the differentiable AP loss
  - originally proposed by He et al. [1]
- Given a query descriptor p_{ij} from image I_1
  - We compare it to **all** descriptors in image I_2:
    - 1 positive, and **many** negatives
  - We compute the AP = $\widetilde{AP}(p_{ij})$

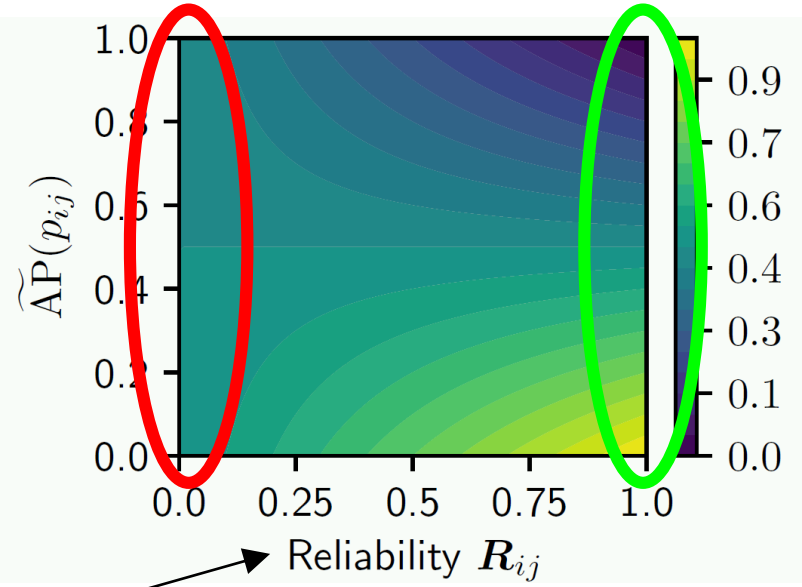


- Modified to not waste efforts on bad regions
  - We estimate the reliability at p_{ij} = R_{ij}
  - Many regions can't be matched (empty, 1-d pattern, repetitive…)
  - For these region, reliability is low ➔ the loss is almost flat

[1] K. He, Y. Lu, and S. Sclaroff. Local descriptors optimized for average precision. In CVPR, 2018.

Joint descriptor & reliability loss

- Based on the differentiable AP loss
  - originally proposed by He et al. [1]
- Given a query descriptor p_{ij} from image I_1
  - We compare it to **all** descriptors in image I_2:
    - 1 positive, and **many** negatives
  - We compute the AP = $\widetilde{AP}(p_{ij})$



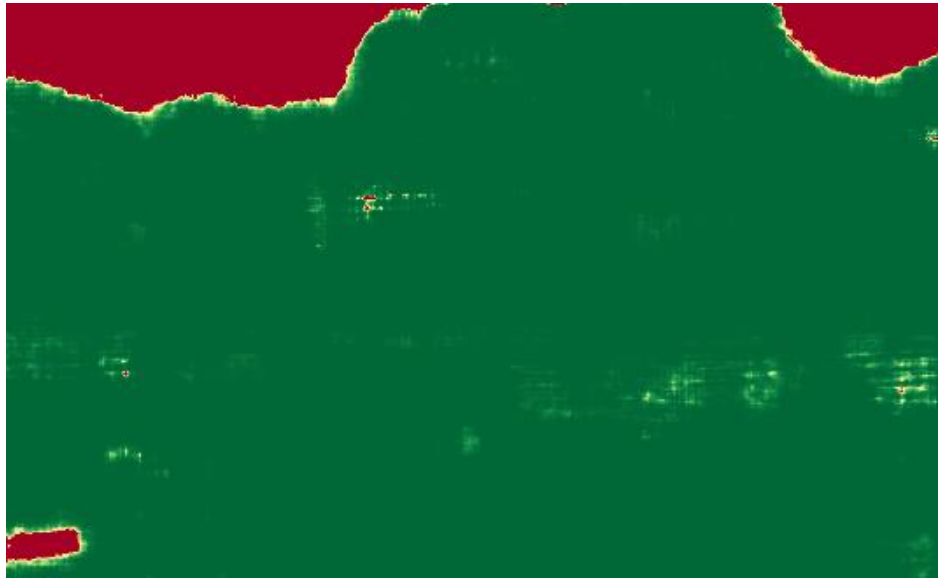
- Modified to not waste efforts on bad regions
  - We estimate the reliability at p_{ij} = R_{ij}
  - Many regions can't be matched (empty, 1-d pattern, repetitive…)
  - For these region, reliability is low ➔ the loss is almost flat

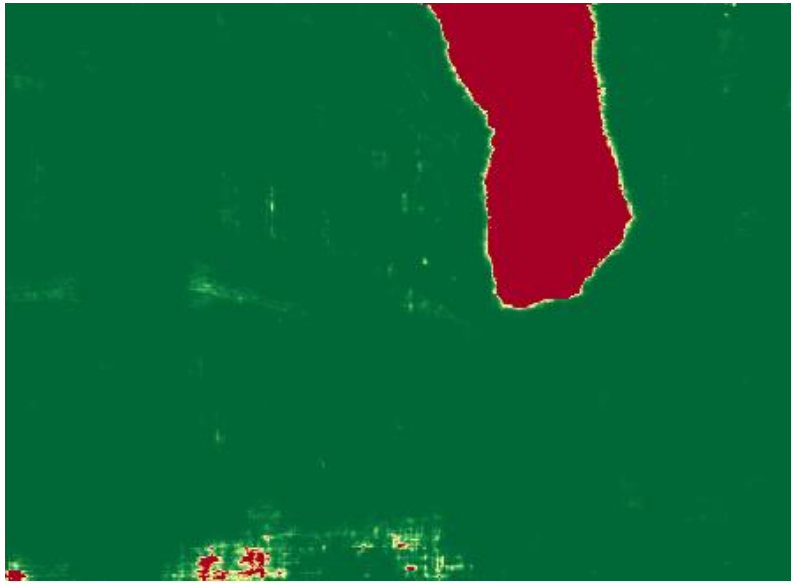
[1] K. He, Y. Lu, and S. Sclaroff. Local descriptors optimized for average precision. In CVPR, 2018.

Reliability maps

Image

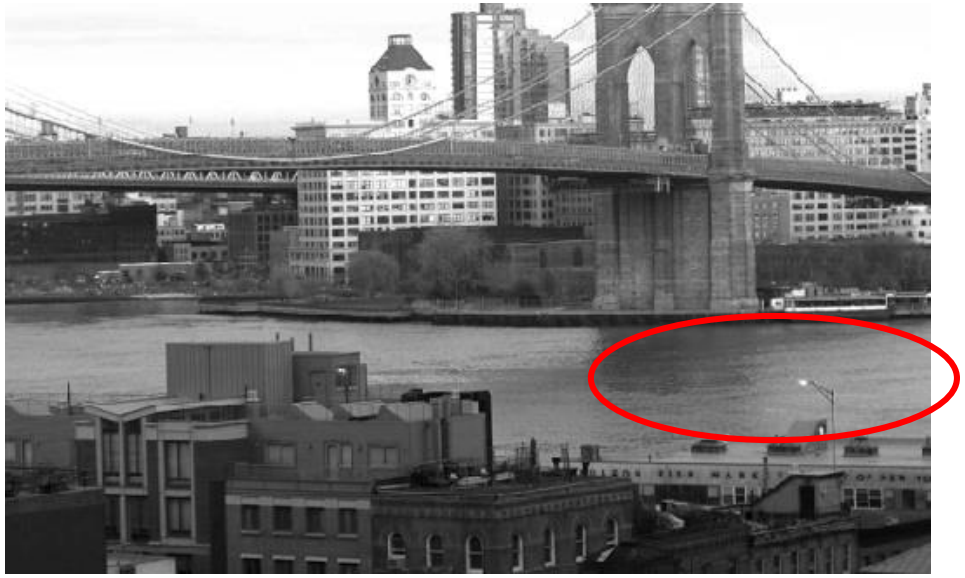
Predicted reliability



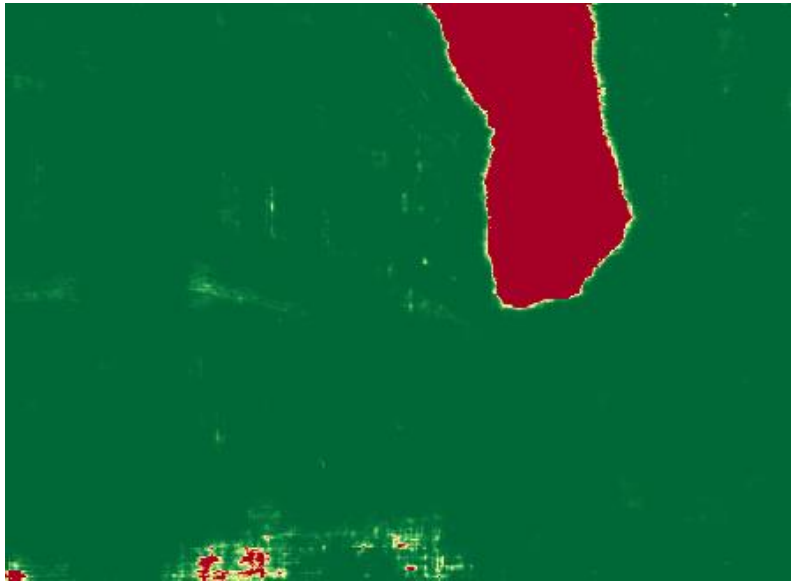


Reliability maps

Image

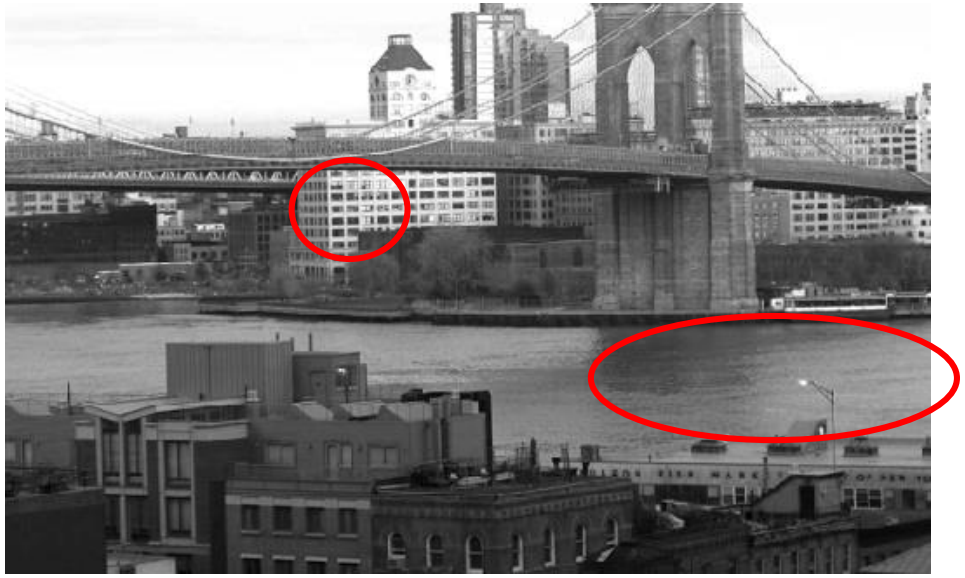


Predicted reliability

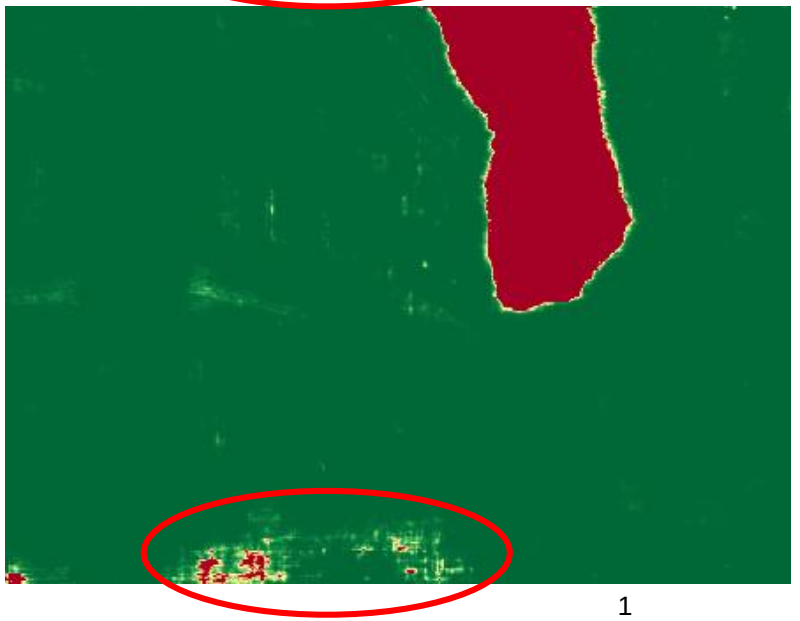


Reliability maps

Image

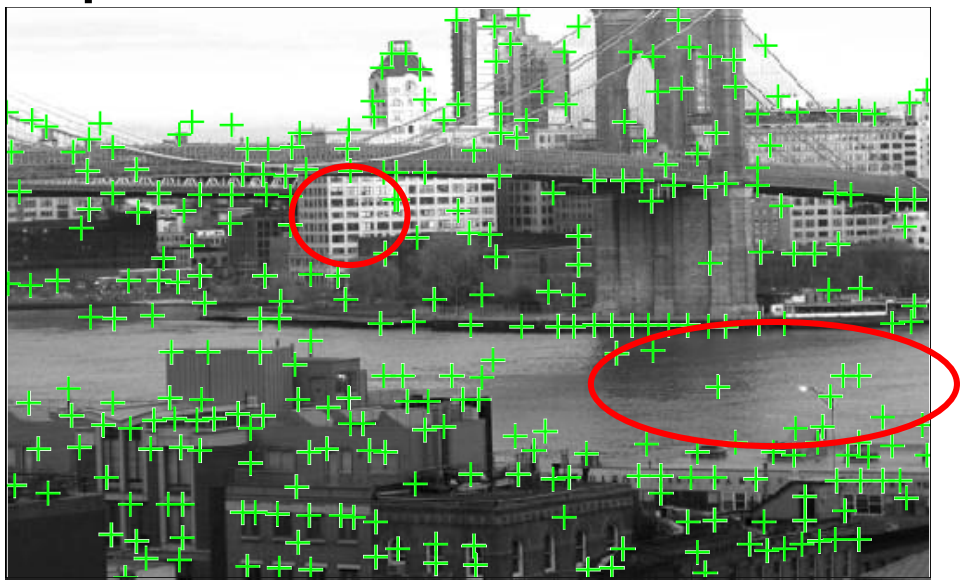


Predicted reliability



Reliability maps

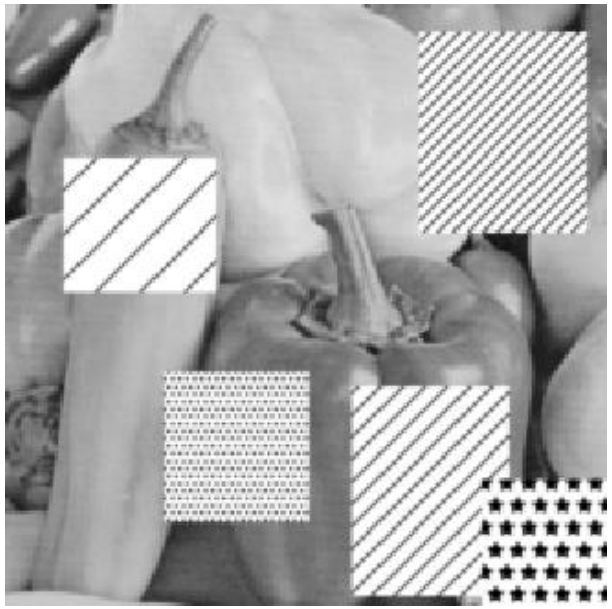
Image



Predicted reliability

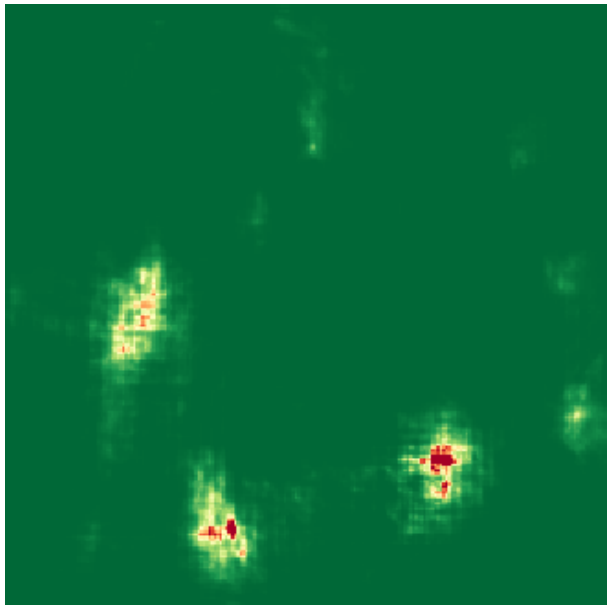
Reliability maps

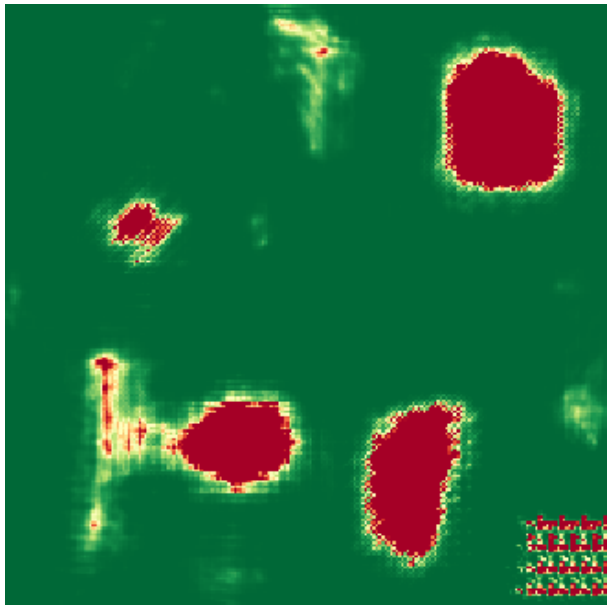
Image



← Same with repetitive patterns (unseen at training)

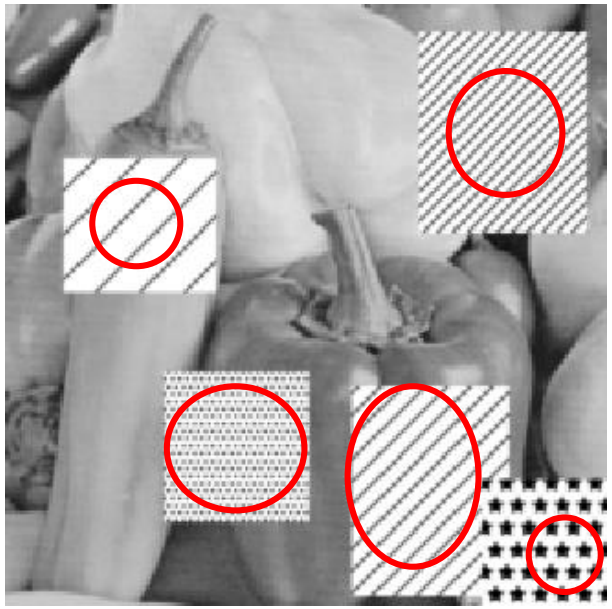
Predicted reliability





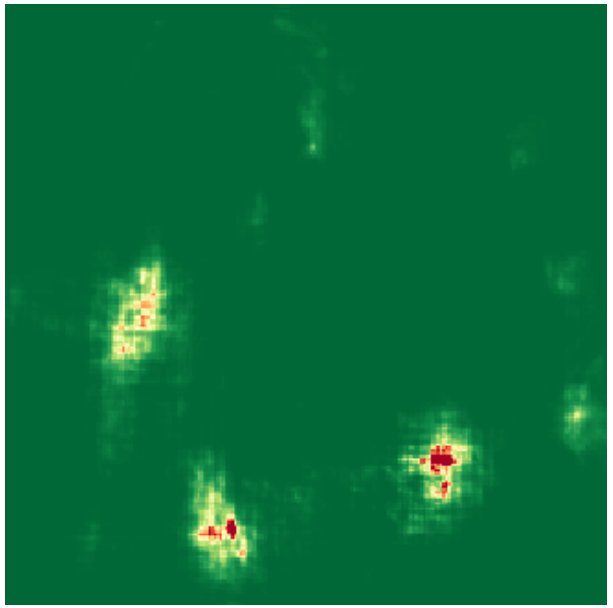
Reliability maps

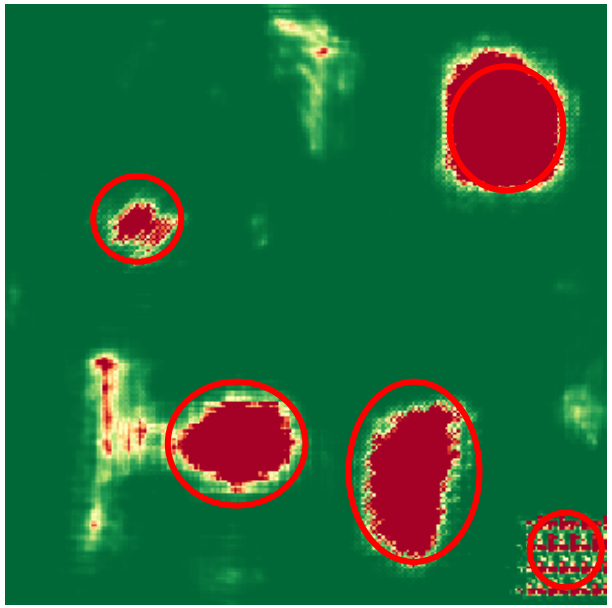
Image



← Same with repetitive patterns (unseen at training)

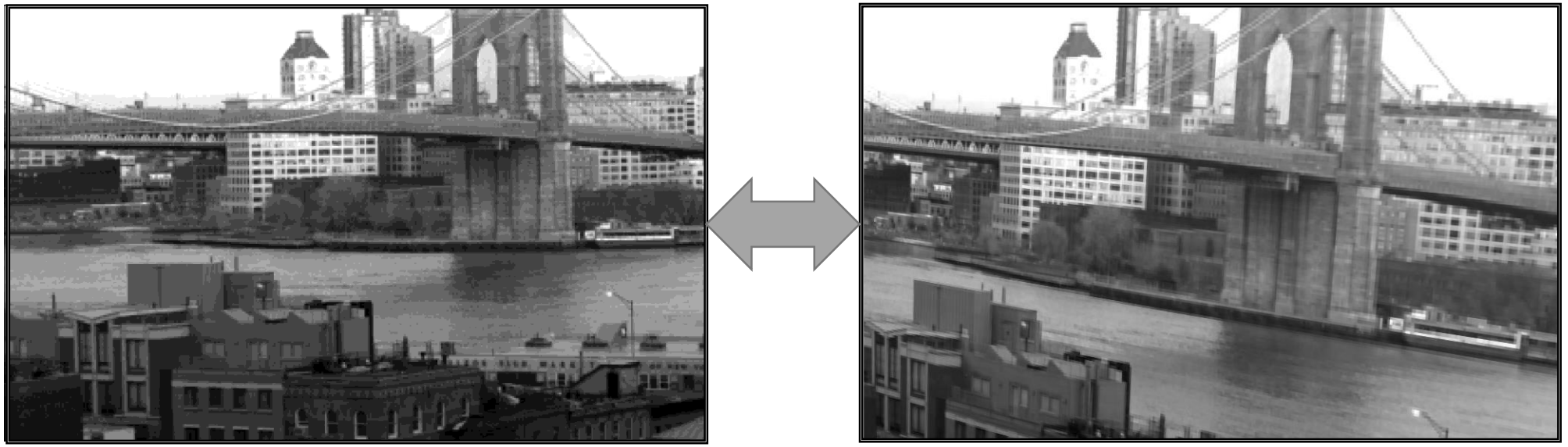
Predicted reliability





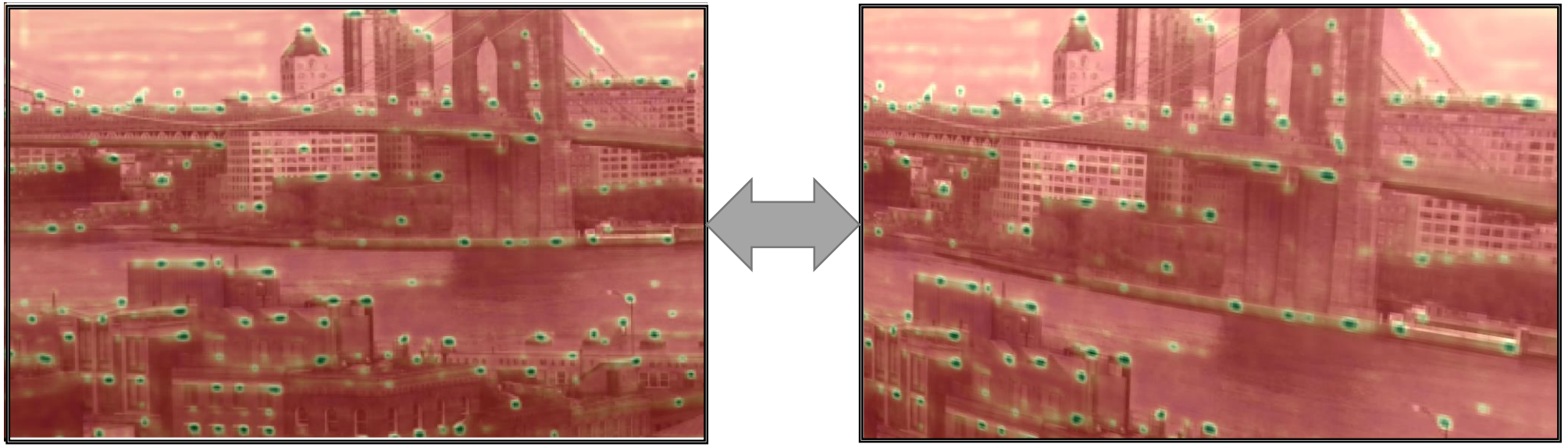
Repeatability loss

- Self-supervised loss
- Key idea:
  - Repeatability maps for an image pairs should be correlated
  - We directly maximize the cosine similarity
  - Locally rather than globally



Repeatability loss

- Self-supervised loss
- Key idea:
  - Repeatibility maps for an image pairs should be correlated
  - We directly maximize the cosine similarity
  - Locally rather than globally

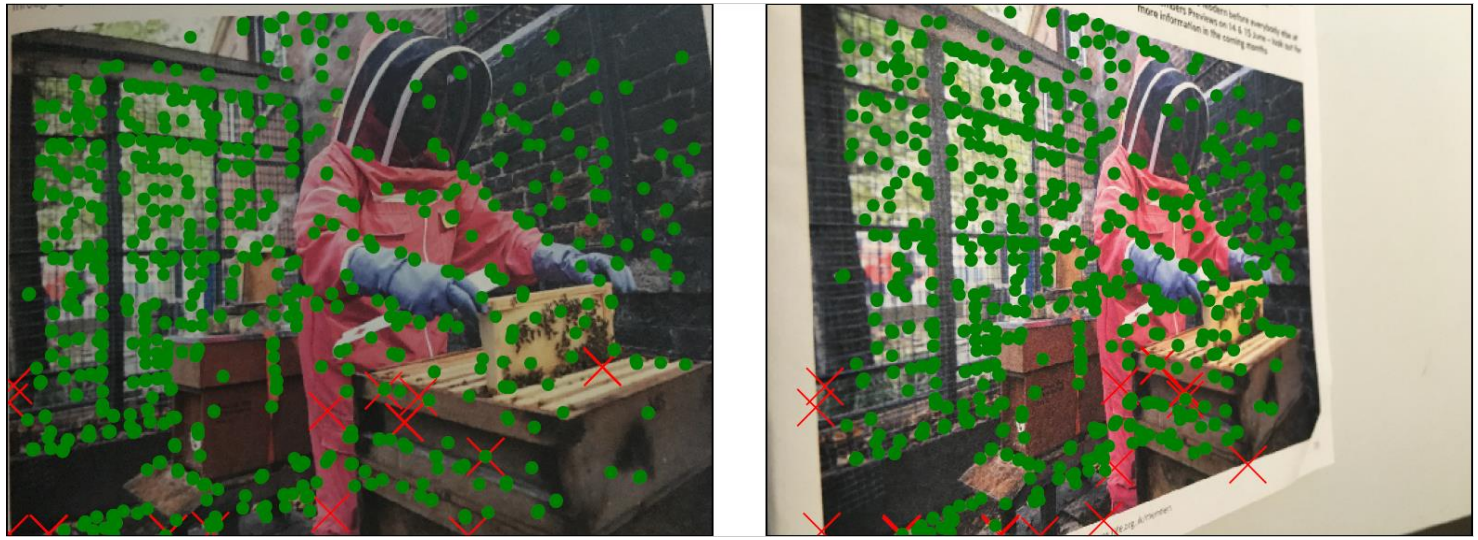


Feature matching experiments

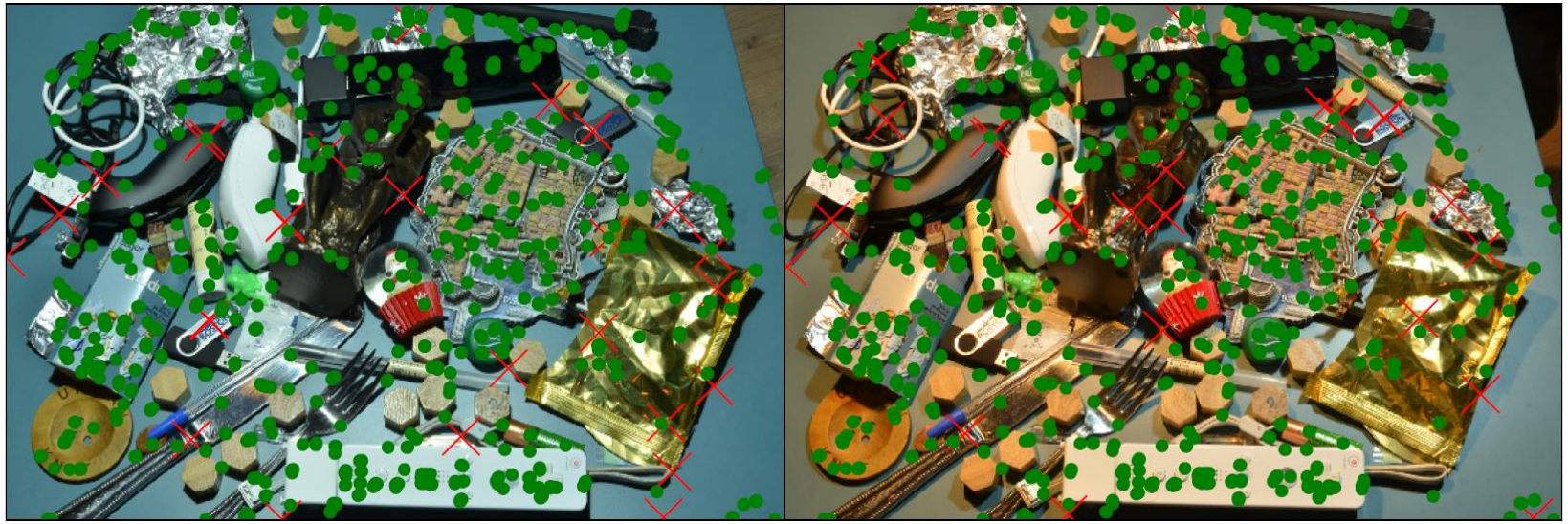
- HPatches dataset:
  - 116 sequences of 6 images = 696 images
  - Viewpoint changes: 59 / Illumination changes: 57
- Evaluation metric: *Mean Matching Accuracy (MMA)*
  - average percentage of correct matches

Feature matching experiments

Viewpoint change:



Illumination change:



Feature matching experiments

- Ablation study on the losses:

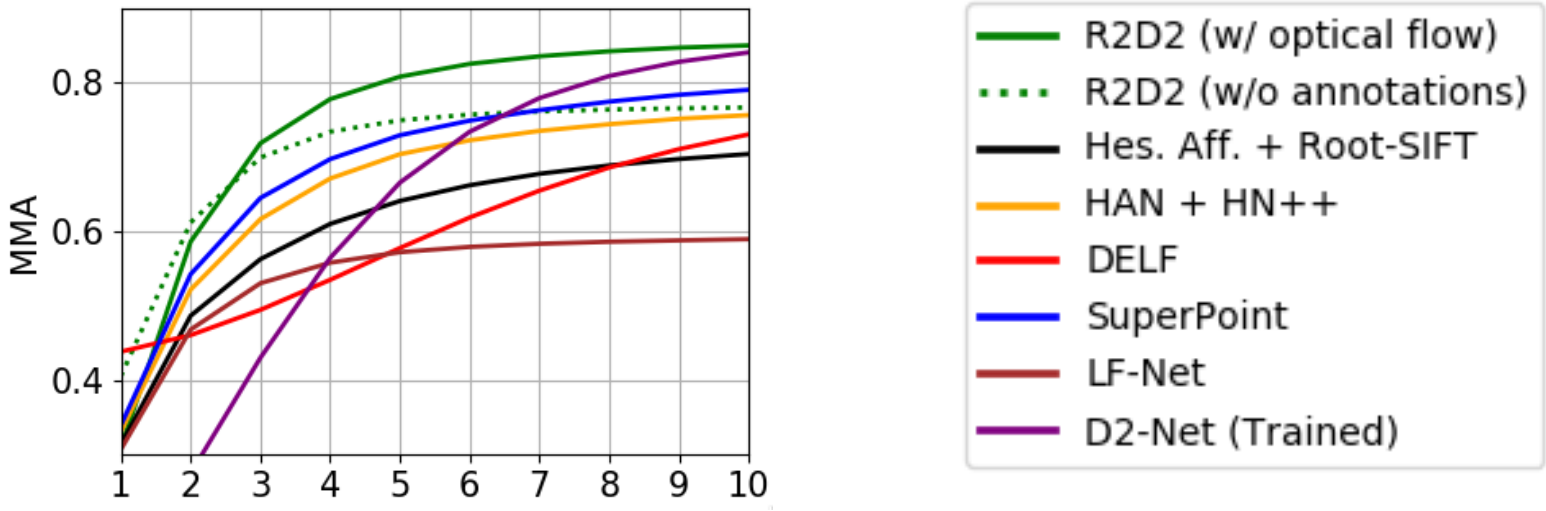
Repeatability	Reliability	MMA@3
	✓	0.588 ± 0.010
✓		0.639 ± 0.034
✓	✓	$\mathbf{0.688 \pm 0.009}$

Feature matching experiments

- Ablation study on the losses:

Repeatability	Reliability	MMA@3
	✓	0.588 ± 0.010
✓		0.639 ± 0.034
✓	✓	$\mathbf{0.688 \pm 0.009}$

- Comparison with the state of the art:



Visual localization experiments

- Aachen day-night benchmark [1]
  - 4328 daytime training images
  - 98 night-time queries
  - Evaluation metric:

    Percentages of successfully localized images within 3 error thresholds

[1] Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions. T. Sattler, W. Maddern, C. Toft, A. Torii, L. Hammarstrand, E. Stenborg, D. Safari, M. Okutomi, M. Pollefeys, J. Sivic, F. Kahl, and T. Pajdla. CVPR, 2018.

Visual localization experiments

- Aachen day-night benchmark [1]
  - 4328 daytime training images
  - 98 night-time queries
  - Evaluation metric:

    Percentages of successfully localized images within 3 error thresholds

- Local feature visual localization challenge at CVPR'19:

Method	#weights	#dim	#kpts	0.5m, 2°	1m, 5°	5m, 10°
RootSIFT	-	128	11K	33.7	52.0	65.3
HAN+HN	2 M	128	11K	37.8	54.1	75.5
SuperPoint	1.3 M	256	7K	42.8	57.1	75.5
DELF (new)	9 M	1024	11K	39.8	61.2	85.7
D2-Net	15 M	512	19K	44.9	**66.3**	**88.8**
R2D2 (ours)	1.0 M	128	10K	**45.9**	**66.3**	**88.8**

[1] Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions. T. Sattler, W. Maddern, C. Toft, A. Torii, L. Hammarstrand, E. Stenborg, D. Safari, M. Okutomi, M. Pollefeys, J. Sivic, F. Kahl, and T. Pajdla. CVPR, 2018.

Visual localization experiments

- Aachen day-night benchmark [1]
  - 4328 daytime training images
  - 98 night-time queries
  - Evaluation metric:

    Percentages of successfully localized images within 3 error thresholds

- Local feature visual localization challenge at CVPR'19:

Method	#weights	#dim	#kpts	0.5m, 2°	1m, 5°	5m, 10°
RootSIFT	-	128	11K	33.7	52.0	65.3
HAN+HN	2 M	128	11K	37.8	54.1	75.5
SuperPoint	1.3 M	256	7K	42.8	57.1	75.5
DELF (new)	9 M	1024	11K	39.8	61.2	85.7
D2-Net	15 M	512	19K	44.9	**66.3**	**88.8**
R2D2 (ours)	1.0 M	128	10K	**45.9**	**66.3**	**88.8**

[1] Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions. T. Sattler, W. Maddern, C. Toft, A. Torii, L. Hammarstrand, E. Stenborg, D. Safari, M. Okutomi, M. Pollefeys, J. Sivic, F. Kahl, and T. Pajdla. CVPR, 2018.

Visual localization experiments

- Aachen day-night benchmark [1]
  - 4328 daytime training images
  - 98 night-time queries
  - Evaluation metric:

    Percentages of successfully localized images within 3 error thresholds

- Local feature visual localization challenge at CVPR'19:

Method	#weights	#dim	#kpts	0.5m, 2°	1m, 5°	5m, 10°
RootSIFT	-	128	11K	33.7	52.0	65.3
HAN+HN	2 M	128	11K	37.8	54.1	75.5
SuperPoint	1.3 M	256	7K	42.8	57.1	75.5
DELF (new)	9 M	1024	11K	39.8	61.2	85.7
D2-Net	15 M	512	19K	44.9	**66.3**	**88.8**
R2D2 (ours)	1.0 M	128	10K	**45.9**	**66.3**	**88.8**

[1] Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions. T. Sattler, W. Maddern, C. Toft, A. Torii, L. Hammarstrand, E. Stenborg, D. Safari, M. Okutomi, M. Pollefeys, J. Sivic, F. Kahl, and T. Pajdla. CVPR, 2018.

Visual localization experiments

- Aachen day-night benchmark [1]
  - 4328 daytime training images
  - 98 night-time queries
  - Evaluation metric:

    Percentages of successfully localized images within 3 error thresholds

- Local feature visual localization challenge at CVPR'19:

Method	#weights	#dim	#kpts	0.5m, 2°	1m, 5°	5m, 10°
RootSIFT	-	128	11K	33.7	52.0	65.3
HAN+HN	2 M	128	11K	37.8	54.1	75.5
SuperPoint	1.3 M	256	7K	42.8	57.1	75.5
DELF (new)	9 M	1024	11K	39.8	61.2	85.7
D2-Net	15 M	512	19K	44.9	**66.3**	**88.8**
R2D2 (ours)	1.0 M	128	10K	**45.9**	**66.3**	**88.8**

[1] Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions. T. Sattler, W. Maddern, C. Toft, A. Torii, L. Hammarstrand, E. Stenborg, D. Safari, M. Okutomi, M. Pollefeys, J. Sivic, F. Kahl, and T. Pajdla. CVPR, 2018.

Visual localization experiments

- Aachen day-night benchmark [1]
  - 4328 daytime training images
  - 98 night-time queries
  - Evaluation metric:

    Percentages of successfully localized images within 3 error thresholds

- Local feature visual localization challenge at CVPR'19:

Method	#weights	#dim	#kpts	0.5m, 2°	1m, 5°	5m, 10°
RootSIFT	-	128	11K	33.7	52.0	65.3
HAN+HN	2 M	128	11K	37.8	54.1	75.5
SuperPoint	1.3 M	256	7K	42.8	57.1	75.5
DELF (new)	9 M	1024	11K	39.8	61.2	85.7
D2-Net	15 M	512	19K	44.9	**66.3**	**88.8**
R2D2 (ours)	1.0 M	128	10K	**45.9**	**66.3**	**88.8**

[1] Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions. T. Sattler, W. Maddern, C. Toft, A. Torii, L. Hammarstrand, E. Stenborg, D. Safari, M. Okutomi, M. Pollefeys, J. Sivic, F. Kahl, and T. Pajdla. CVPR, 2018.

Conclusion

- Come to our poster #XXX!
- The code is online at https://github.com/naver/r2d2