Robust Image Retrieval-based Visual Localization using Kapture

Martin Humenberger
Yohann Cabon
Nicolas Guerin
Julien Morat
Jérôme Revaud
Philippe Rerole
Noé Pion
Cesar de Souza
Vincent Leroy
Gabriela Csurka

17th June 2020

NAVER LABS is an ambient intelligence technology company owned by NAVER Corporation, Korea's leading internet content services company.

NAVER LABS Europe is the biggest industrial research lab in artificial intelligence in France.

Leader

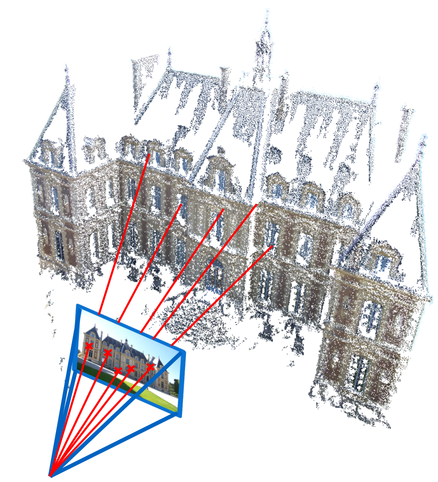
Martin Humenberger | Group Lead & Senior Scientist

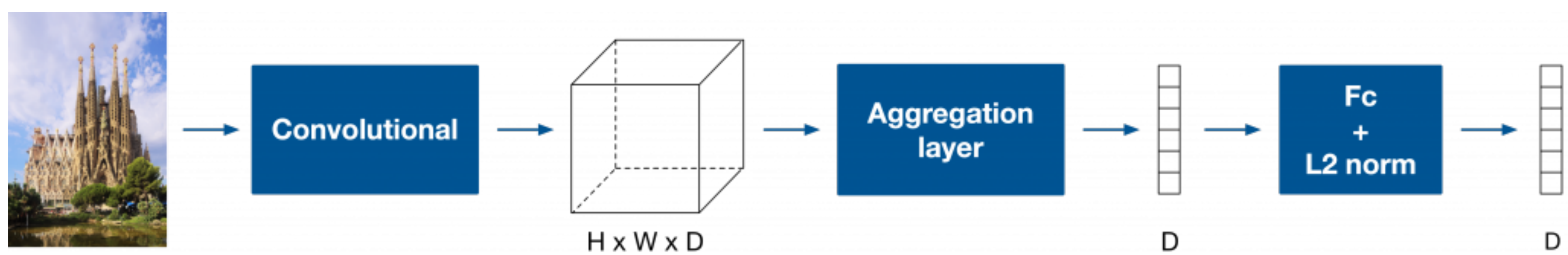
NAVER LABS Europe

NAVER LABS Europe (2017-)
Austrian Institute of Technology
NASA Jet Propulsion Lab
PhD, Vienna University of Technology

Outline

- Visual localization
- Our method (KAPTURE-R2D2-APGeM)
- APGeM-based image retrieval
- R2D2 for local feature matching
- Results
- Kapture





Kapture

Visual Localization

GPS accuracy sometimes not enough.

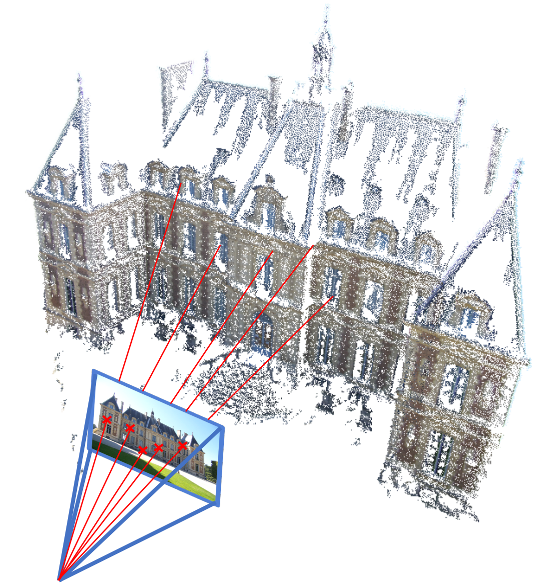
E.g. for precise robot navigation or augmented reality.

Position from GPS

Château de Sceaux

Goal: Use an image to estimate the **<u>precise</u>** position of the camera within a given area (map).

Overview of Methods



Structure-based methods	Active Search [1] OpenMVG [2]	+ -	Perform very well on most datasets -> high accuracy More difficult for very large environments (memory and processing time)
Image retrieval-based methods	IR IBL revisited [3] HF-Net [4]	+ -	Improve speed and robustness for large scale settings Quality heavily relies on image retrieval
Camera pose regression methods	PoseNet [5]	+ -	Interesting approach because no 3D maps are needed and it is data driven (can be trained for certain challenges) Low accuracy
Scene coordinate regression methods	SCR Forests [6] DSAC++ [7]	+ -	Accurate in small scale settings Does not yet work in large scale environments

[1] T. Sattler et al., Improving Image-Based Localization by Active Correspondence Search, ECCV 2012

[2] P. Moulon, OpenMVG: http://github.com/openMVG/openMVG

[3] T. Sattler et al., Image Retrieval for Image-Based Localization Revisited, BMVC 2012

[4] Sarlin et al., From Coarse to Fine: Robust Hierarchical Localization at Large Scale, CVPR 2019

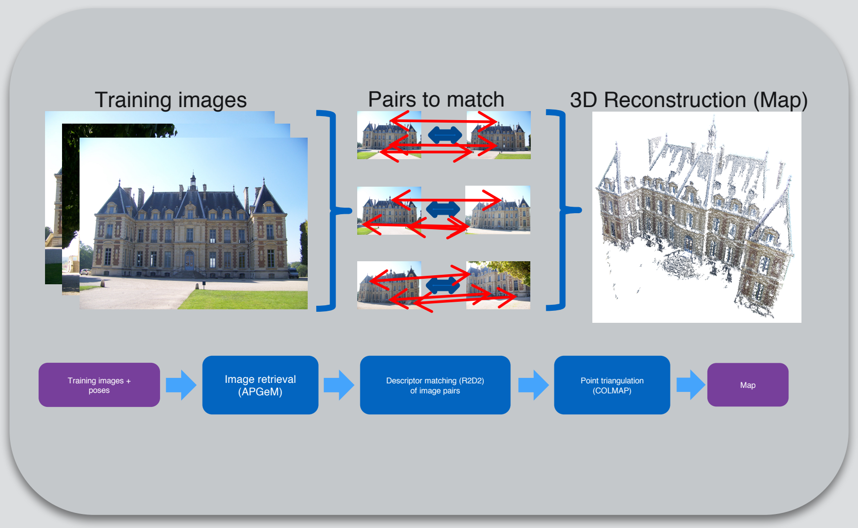
[5] A. Kendall et al., PoseNet: http://mi.eng.cam.ac.uk/projects/relocalisation/, ICCV 2015

[6] J. Shotton et al., Scene Coordinate Regression Forests for Camera Relocalization in RGB-D Images, CVPR 2013

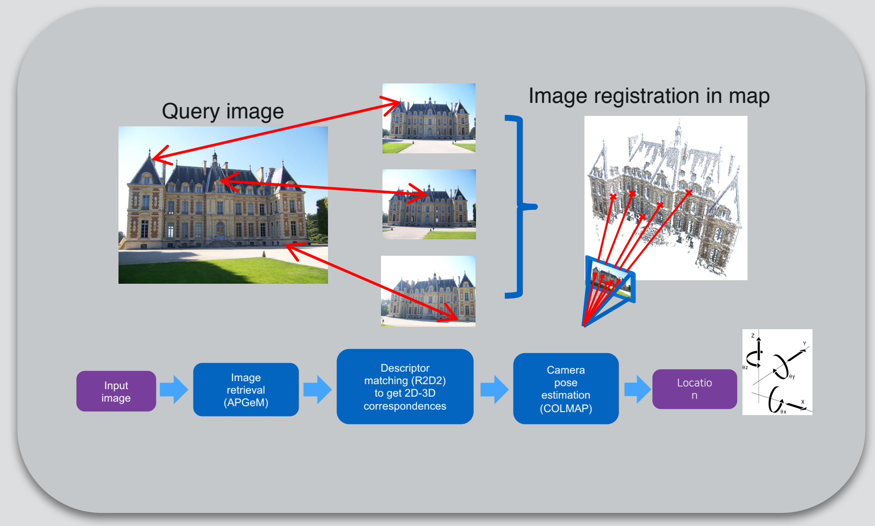
[7] E. Brachmann et al., Learning Less is More – 6D Camera Localization via 3D Surface Regression, CVPR 2018

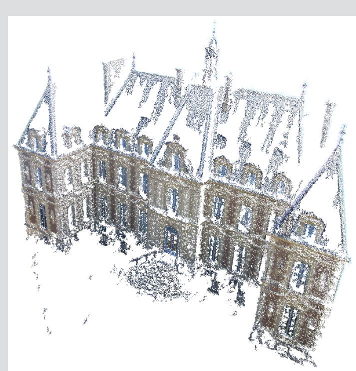
Our method for visual localization

Mapping



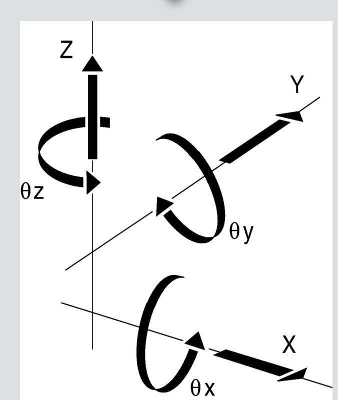
Localization





3D Map

Kapture

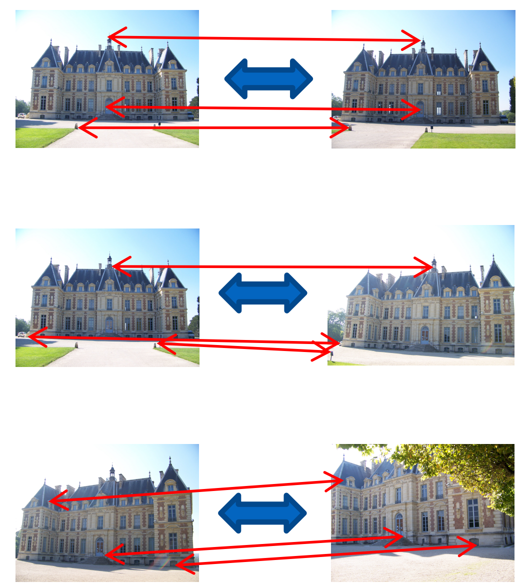


6DOF Pose

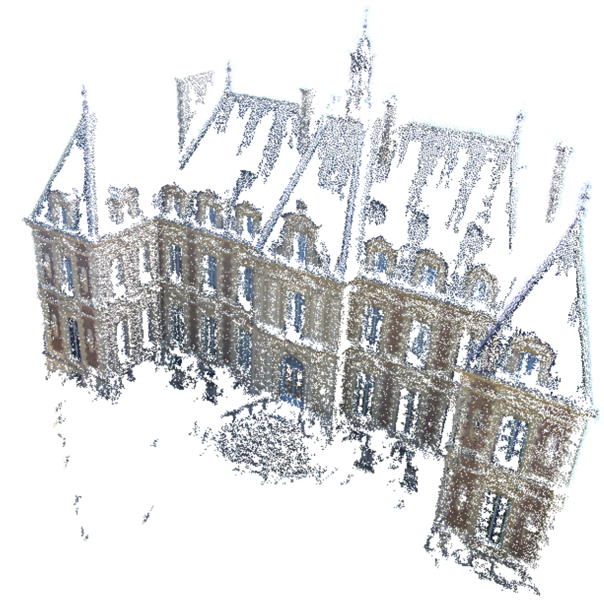
Mapping

Training images

Pairs to match



3D reconstruction (map)



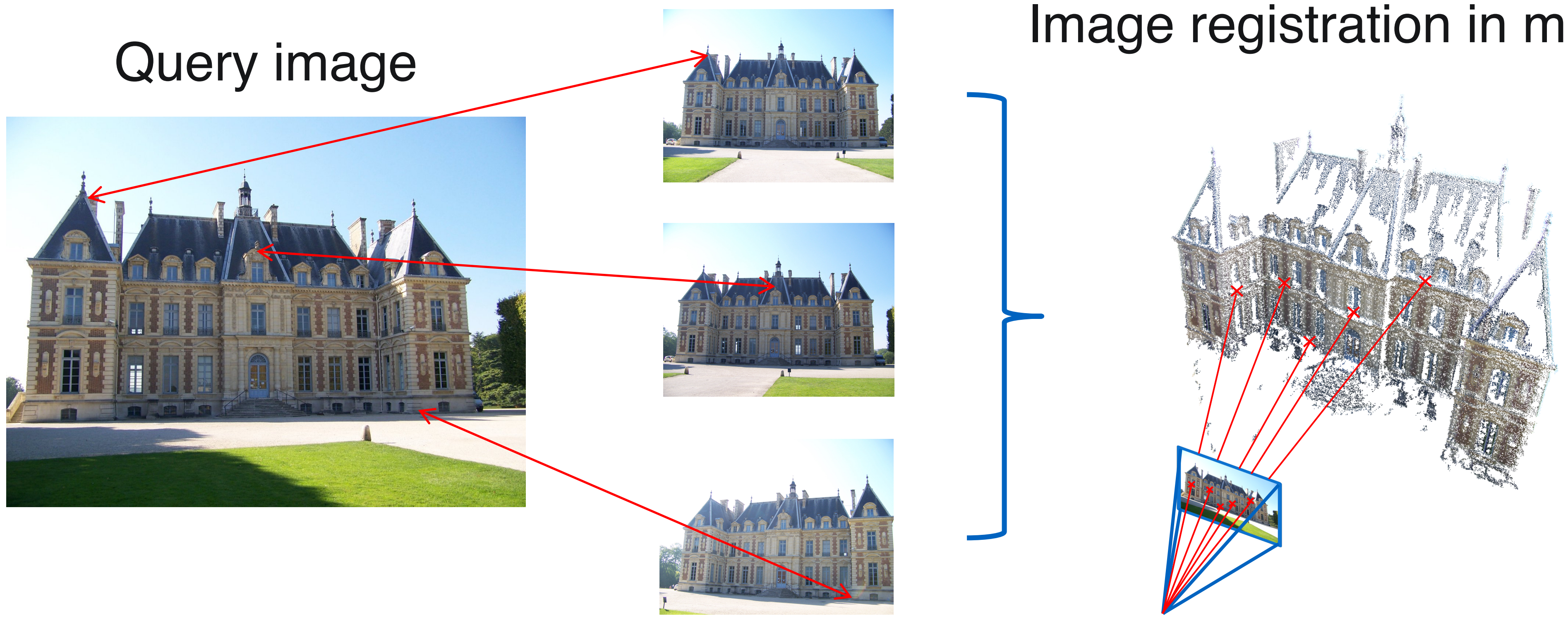
Training images + poses → Image retrieval (APGeM) → Descriptor matching (R2D2) of image pairs → Point triangulation (COLMAP) → Map

https://colmap.github.io

Localization

Query image

Image registration in map



Input image → Image retrieval (APGeM) → Descriptor matching (R2D2) to get 2D-3D correspondences → Camera pose estimation (COLMAP) → Location

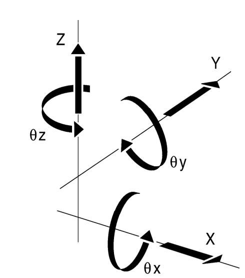
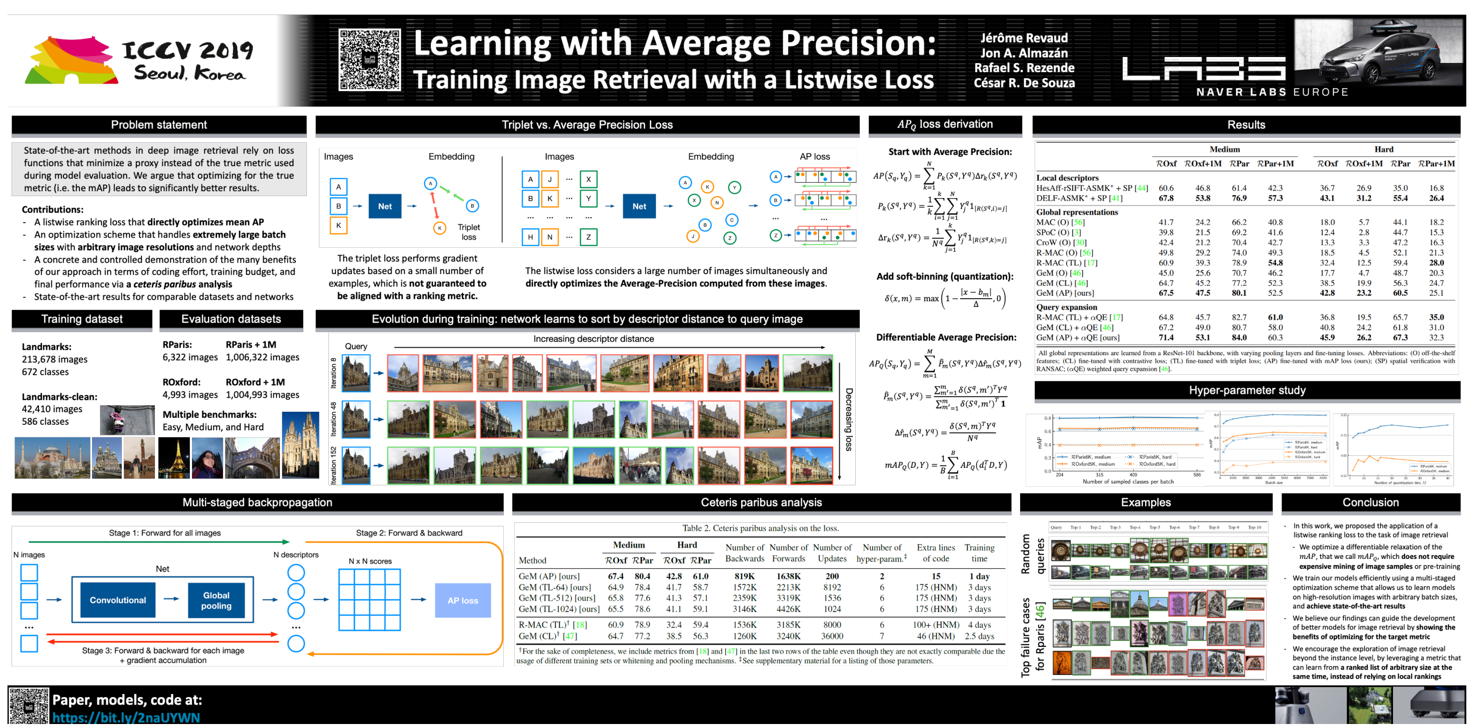


Image Retrieval – APGeM
Average Precision Generalized Mean Pooling

Learning with Average Precision: Training Image Retrieval with a Listwise Loss

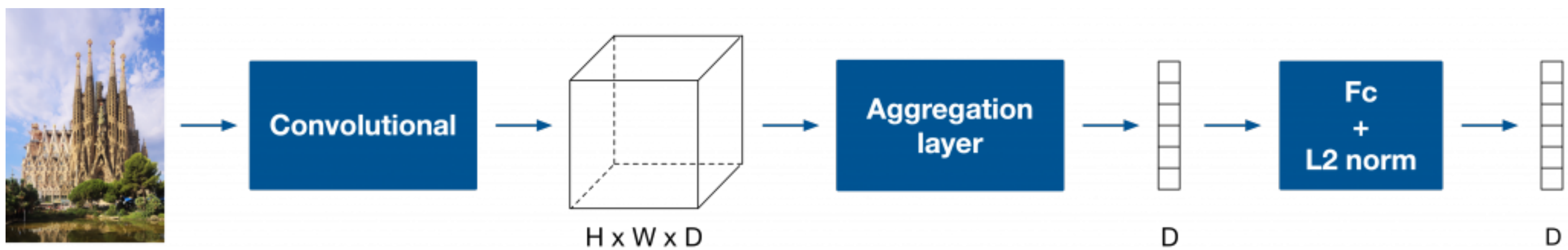
Jérôme Revaud, Jon A. Almazán, Rafael S. Rezende, Cesar De Souza

ICCV 2019 (poster)

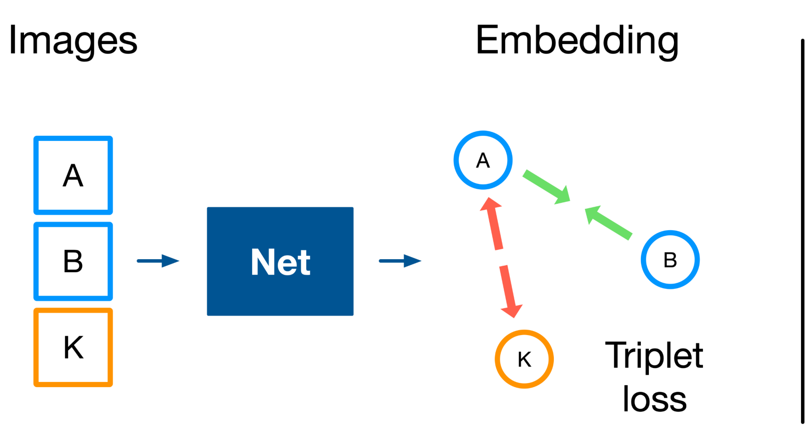


Code, paper, models:
https://europe.naverlabs.com/Research/Computer-Vision/Learning-Visual-Representations/Deep-Image-Retrieval/

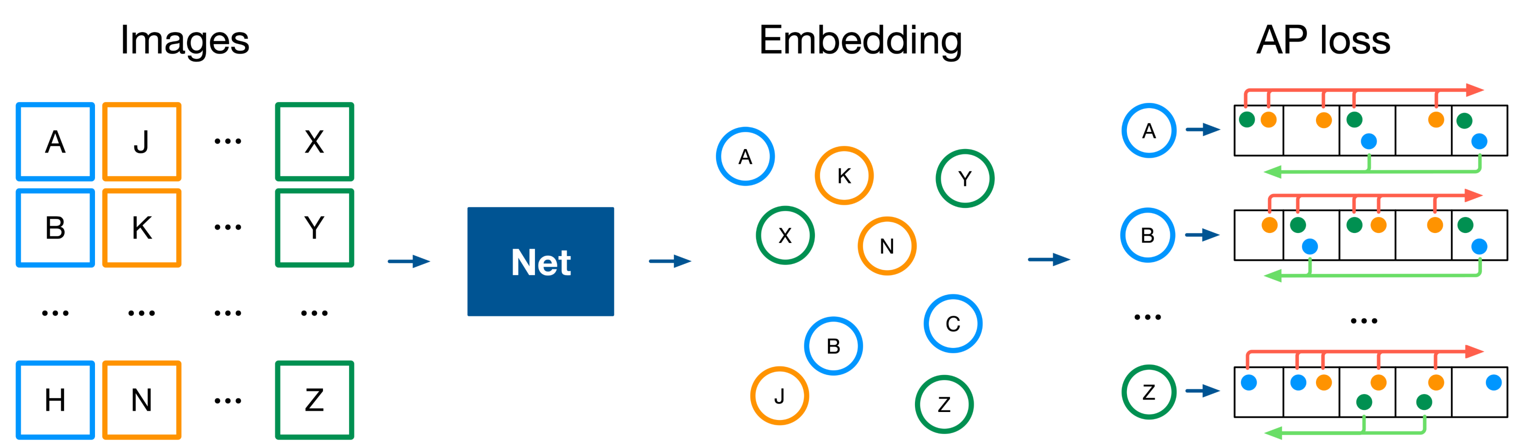
Image Retrieval - APGeM



Instead of minimizing a proxy (e.g. triplet loss), APGeM uses a listwise ranking loss that directly optimizes the true metric, the mean average precision (AP).



The triplet loss performs gradient updates based on a small number of examples, which is **not guaranteed to be aligned with a ranking metric.**



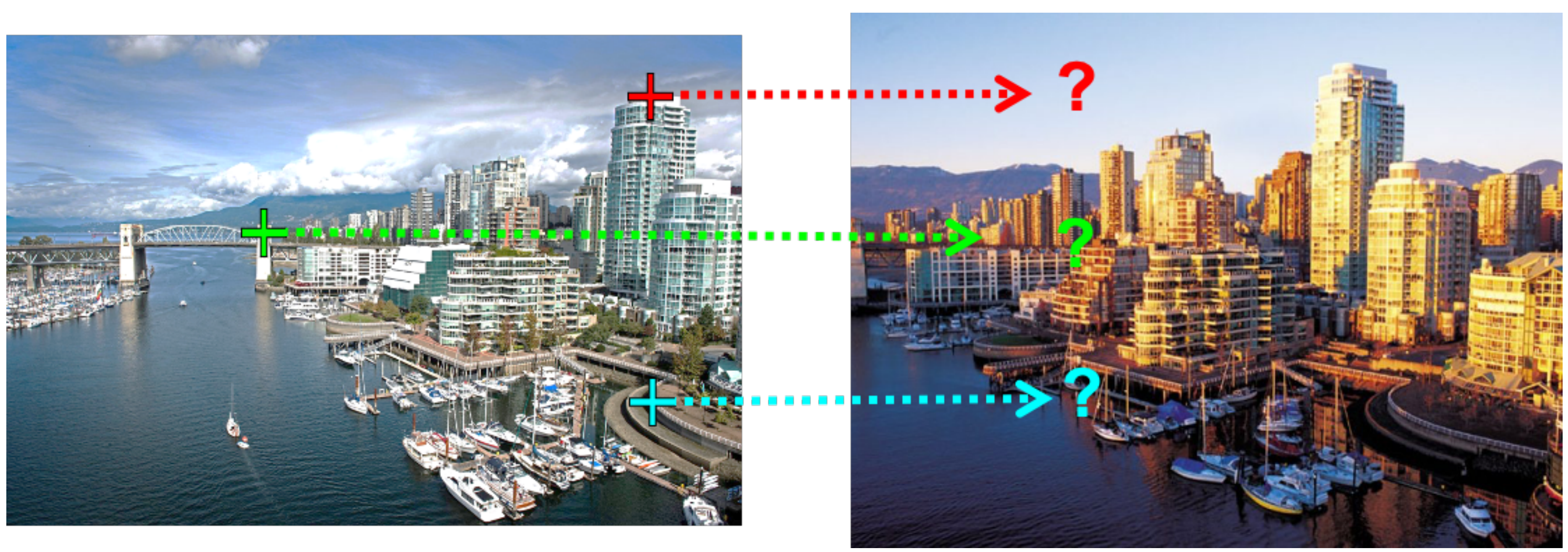
The listwise loss considers a large number of images simultaneously and **directly optimizes the Average-Precision computed from these images**.

Local Features – R2D2

R2D2: Repeatable and Reliable Detector and Descriptor

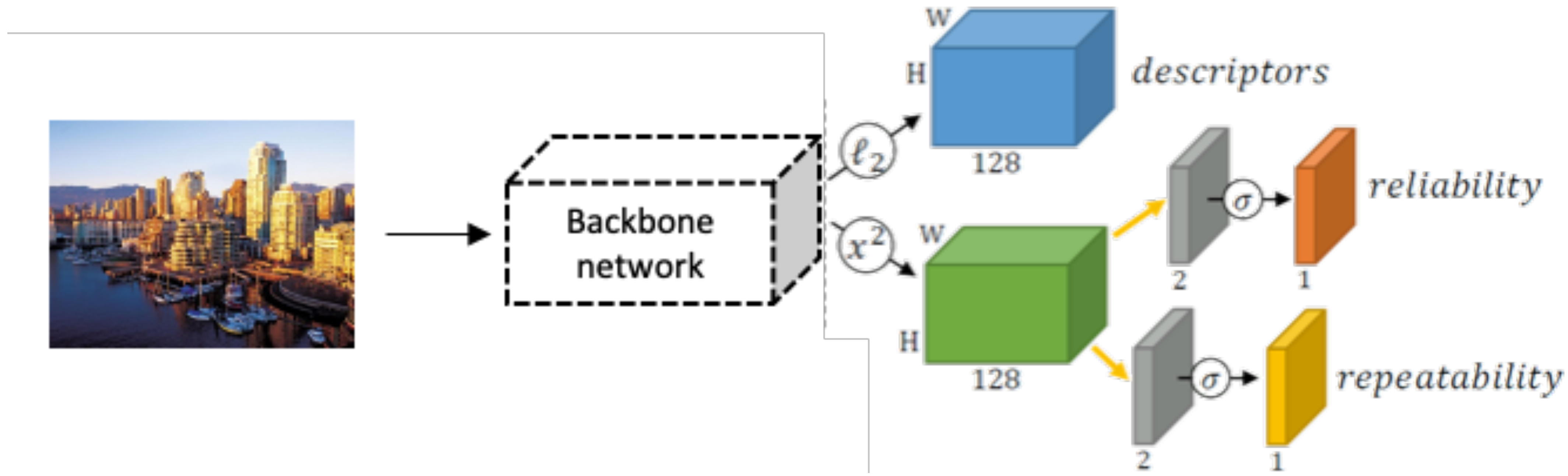
Jérôme Revaud, Philippe Weinzaepfel, Cesar De Souza, Martin Humenberger

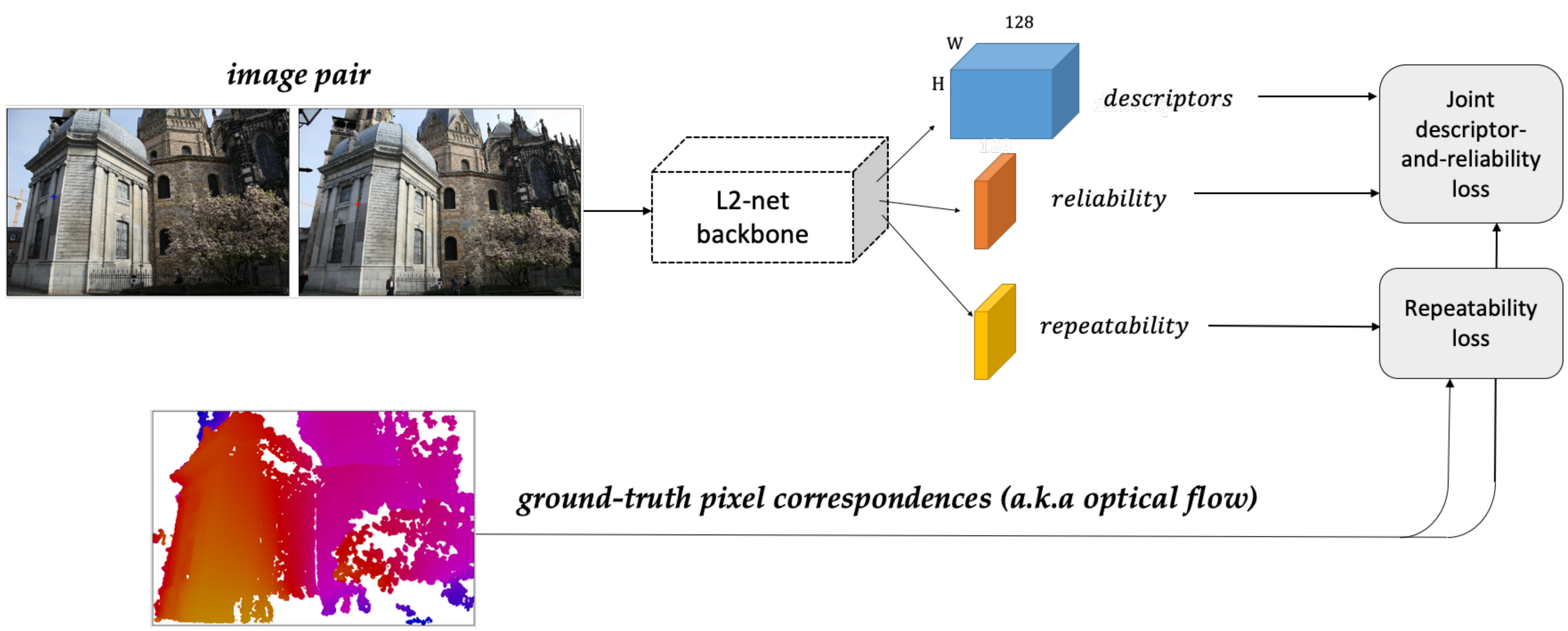
NeurIPS 2019 (oral)



Code, paper, models: https://github.com/naver/r2d2

Local Features – R2D2





cosim

mAP

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

This year, VisLocOdomMapCVPR2020 (Joint Workshop on Long-Term Visual Localization, Visual Odometry and Geometric and Learning-based SLAM at the 2020 Conference on Computer Vision and Pattern Recognition) issued three visual localization challenges to advance research on the topic.

1. visual localization for autonomous vehicles
2. visual localization for handheld devices
3. local features for long-term localization

We're proud to announce that the entry of our team at NAVER LABS Europe performed extremely well – **ranking first in challenge 1**, **fourth in challenge 2**, and **second in challenge 3**.

https://europe.naverlabs.com/blog/one-method-one-pipeline-naver-labs-europe-ranks-high-across-three-visual-localization-challenges-at-cvpr-2020/

Results on autonomous driving datasets from https://www.visuallocalization.net

RobotCar Seasons

Rank 1 / 47

https://data.ciirc.cvut.cz/public/projects/2020VisualLocalization/RobotCar-Seasons/

Method	day all	night all
KAPTURE-R2D2-APGeM	**55.1 / 82.1 / 97.3**	**28.8 / 58.8 / 89.4**
Visual Localization Using Dense Semantic 3D Map And Hybrid Features	54.6 / 81.9 / 96.9	14.8 / 33.0 / 51.3
RT_AP_IR+CRBNet	55.3 / 81.8 / 98.5	11.5 / 26.5 / 39.2
SIFT+IR50+SIG+FM+R70_85+MUL+RE3+GPNP+MERGE	57.2 / 81.5 / 97.4	9.3 / 30.1 / 53.3
Geometric Prior Guided Camera Localization	57.3 / 81.7 / 97.6	8.0 / 21.6 / 40.7

Extended CMU

Rank 4 / 30

https://data.ciirc.cvut.cz/public/projects/2020VisualLocalization/Extended-CMU-Seasons/

Method	urban	suburban	park
PFSL FGSN 200 classes (sequential localization)	94.1 / 99.3 / 100.0	100.0 / 100.0 / 100.0	97.6 / 99.9 / 99.9
PFSL Cityscapes classes (sequential localization)	88.1 / 93.8 / 100.0	97.4 / 99.2 / 100.0	81.8 / 97.2 / 99.9
DenseVLAD & D2-Net	94.0 / 97.7 / 99.1	93.0 / 95.7 / 98.3	89.2 / 93.2 / 95.0
KAPTURE-R2D2-APGeM	**96.7 / 98.9 / 99.7**	**94.4 / 96.8 / 99.2**	**83.6 / 89.0 / 95.5**
Hierarchical-Localization (multi-camera when available)	91.6 / 96.4 / 99.1	84.7 / 91.5 / 98.6	69.3 / 77.8 / 90.5

SILDa Weather and Time of Day dataset

Rank 3 / 10

https://medium.com/scape-technologies/silda-a-multi-task-dataset-for-evaluating-visual-localization-7fc6c2c56c74

Method	evening	snow	night
NetVLAD (top-50) & D2-Net - multi-scale	29.6 / 67.8 / 94.8	6.0 / 16.4 / 72.3	25.6 / 51.6 / 79.9
NetVLAD (top-50) & D2-Net - single-scale	30.0 / 67.8 / 94.4	1.4 / 10.8 / 67.5	25.7 / 51.9 / 79.9
KAPTURE-R2D2-APGeM	**31.9 / 66.6 / 92.5**	**0.5 / 5.8 / 89.2**	**30.5 / 54.2 / 78.5**
NetVLAD (top-20) & D2-Net - single-scale	28.1 / 66.8 / 93.5	1.7 / 11.1 / 67.3	24.4 / 49.0 / 75.7

Results on other datasets from https://www.visuallocalization.net

Aachen Day-Night

Rank 4 / 55

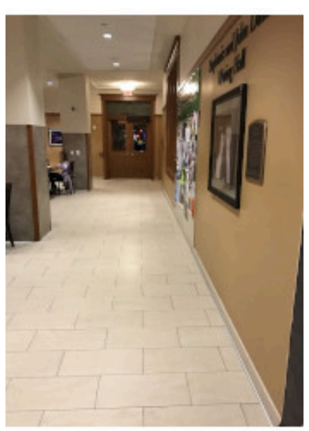
https://data.ciirc.cvut.cz/public/projects/2020VisualLocalization/Aachen-Day-Night/

Method	day	night
ONavi	85.7 / 93.7 / 98.9	48.0 / 71.4 / 88.8
Hierarchical-Localization + SuperGlue	89.6 / 96.1 / 98.8	44.9 / 71.4 / 88.8
Visual Localization Using Dense Semantic 3D Map And Hybrid Features	90.3 / 95.5 / 97.9	44.9 / 67.3 / 87.8
KAPTURE-R2D2-APGeM	**88.7 / 95.8 / 98.8**	**44.9 / 62.2 / 85.7**
rkpd2m_5k	87.7 / 93.7 / 97.0	42.9 / 66.3 / 85.7

Inloc (we did not follow the Inloc pipeline)

Rank 28 / 35

http://www.ok.sc.e.titech.ac.jp/INLOC/



Method	duc1	duc2
perl-nvsg+rf	48.5 / 70.7 / 80.8	56.5 / 75.6 / 84.0
perl-nvsg+srf	50.5 / 71.7 / 80.3	55.0 / 74.0 / 81.7
perl-nvsg	50.0 / 69.7 / 78.3	54.2 / 72.5 / 80.2
Hierarchical-Localization + SuperGlue	49.0 / 69.2 / 79.8	53.4 / 77.1 / 80.9
ONavi	41.9 / 68.2 / 84.3	50.4 / 76.3 / 80.2
KAPTURE-R2D2-APGeM	**21.7 / 37.4 / 54.5**	**23.7 / 41.2 / 54.2**

Findings

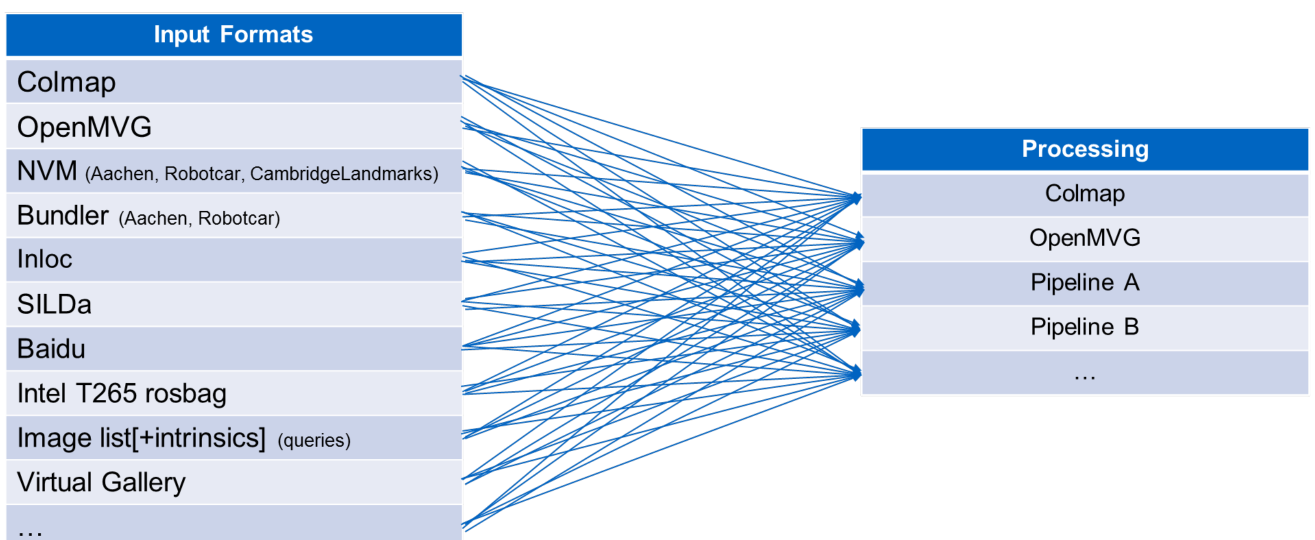
- List-wise loss is beneficial for visual localization: Can be seen in R2D2 and APGeM
- Structure-based methods provide a good way of combining machine learning with geometry to increase robustness (e.g. learned local and global features)
- We think that there is more potential in utilizing image sequences and multi-camera rigs.
- Situations where the environment dramatically changed, e.g. caused by snow, still are an interesting challenge for image retrieval.

Kapture, a file format for Visual Localization datasets

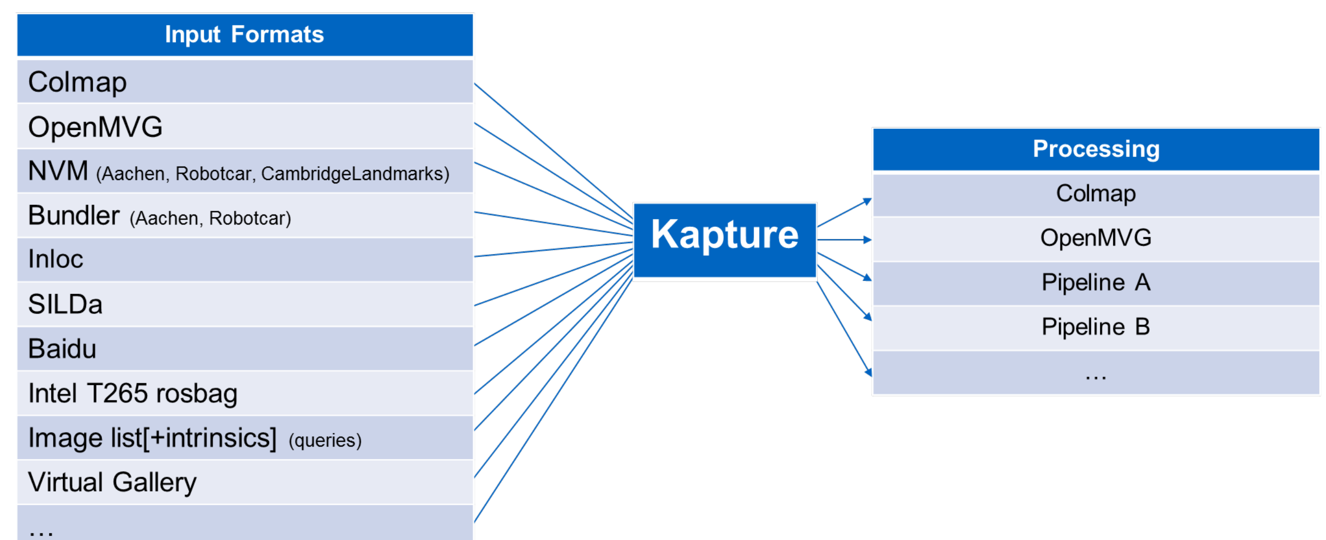
Visual Localization datasets may provide different kinds of data:

- Camera sensor data: images, timestamps, camera parameters, rig configurations
- Other sensor data: lidar, wifi, etc.
- Reconstruction data: descriptors, keypoints, matches, 3D reconstruction

Most datasets use their own data format (sometimes with different coordinate systems), making it difficult to benchmark different algorithms on many datasets.



without Kapture



with Kapture

This is why we created Kapture!

Kapture, more than a file format

Kapture is:

- A versatile and extensible format for SFM and other data.
- A set of converters between many popular formats:
  - COLMAP, OpenMVG, NVM, Bundler, rosbag, some datasets, etc.
- A set of tools, e.g. for merging datasets or evaluation and visualization of localization results
- A Python library to load/save/manipulate Kapture data, that can be used to create new converters and custom processing pipelines.

As an example, we implemented a mapping and localization pipeline based on COLMAP:

- using COLMAP SIFT features
- or using custom descriptors and matches.

More (hopefully) useful functions will follow which will complement existing tools.

Soon available on GitHub under a BSD license, with ready to use datasets!

Conclusion

- Robust visual localization using APGeM for image retrieval and R2D2 for local feature matching
- Using a single method, we report very good results on a set of quite different datasets
- Kapture: A versatile data format to facilitate future research and data processing in visual localization and SFM

- A big thank you to the organizers of this great workshop and the interesting challenges. This is an inspiration for many researchers in the field!

Links:
APGeM: https://europe.naverlabs.com/Research/Computer-Vision/Learning-Visual-Representations/Deep-Image-Retrieval/
R2D2: https://github.com/naver/r2d2
Kapture: will be released soon

Thank you