ECCV'20 ONLINE
23-28 AUGUST 2020

Image Stylization for Robust Features

Iaroslav Melekhov   Gabriel Brostow [2]   Juho Kannala

Daniyar Turmukhambetov [2]

Aalto University
School of Science

NIANTIC LABS

Outline

- Visual localization (relative camera pose estimation);
- Localization datasets: challenges;
- Photorealistic image stylization for local keypoint/detector learning;
- Results;
- Summary and observations.

Localization pipeline

image pair

detection → description → feature matching and filtering → ego-motion estimation

R2D2 detector and descriptor [1]

Provided by the organizers [2]

[1] Revaud et al.: R2D2: Repeatable and Reliable detector and descriptor. *NeurIPS* 2019
[2] https://github.com/pmwenzel/mlad-benchmark-baselines

Problem statement: Localization datasets are challenging…

Aachen [1]: different **illumination** (day-night) conditions

SILDa [2]: different **weather** and **lighting** conditions

MLAD: direct light, greyscale images

[1] Sattler et al.: Image retrieval for image-based localization revisited. *BMVC* 2012
[2] Balntas et al.: SILDa: Scape Imperial Localisation Dataset. https://www.visuallocalization.net/, 2019

Goal

To learn local **keypoint detector** and **descriptor** which will be **robust** to different weather and lighting conditions.

Requirements to training data:

Ideally, it should be a **static scene** (no moving objects or occlusions) captured by camera under **various** **illumination**, **weather** and **season changes.**

- Structure-from-Motion (SfM) datasets;
- Webcam archives.

Method overview: stylization. Limitations

- The AMOS dataset [1, 2]:

pros:

- an outdoor publicly available dataset;
- static cameras; many cameras store images in all seasons and during the whole day;
- > 1B images.

limitations:

- the scenes are not completely static (moving objects);
- some cameras have technical issues, eg out of focus.

[1] Jacobs et al.: The global network of outdoor webcams: properties and applications. *ACM SIGSPATIAL* 2009
[2] Jacobs et al.: Consistent temporal variations in many outdoor scenes. *CVPR* 2007

Method overview: stylization. Limitations

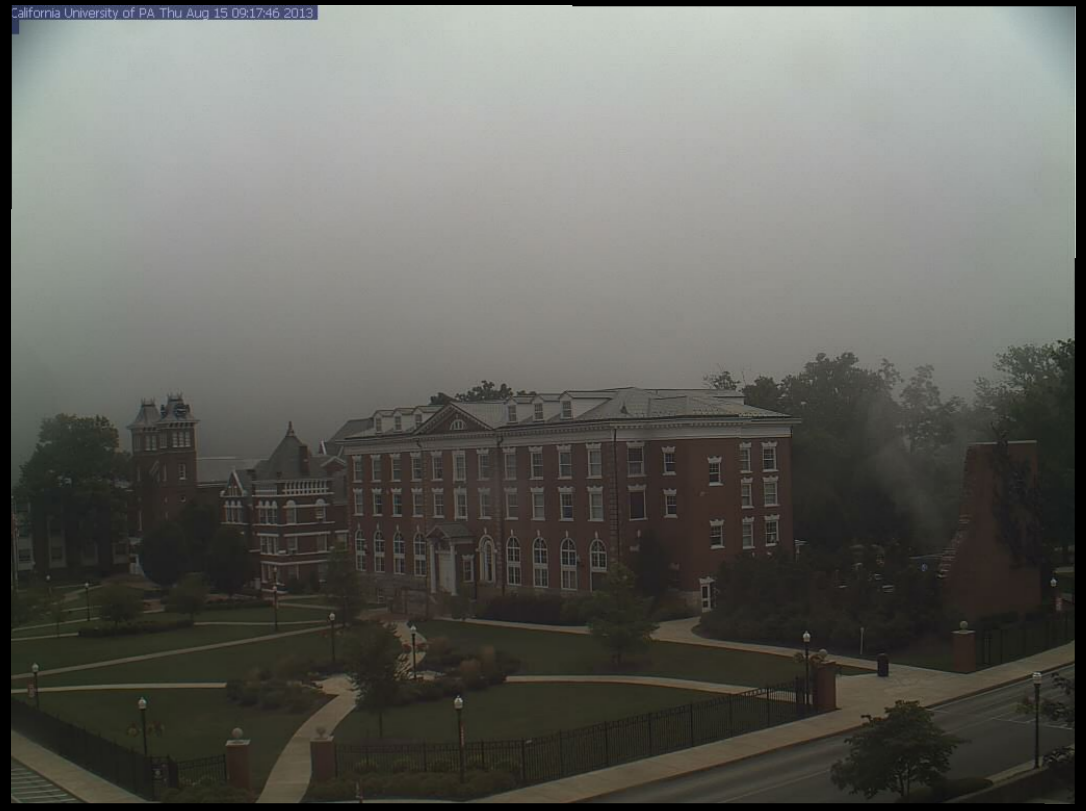
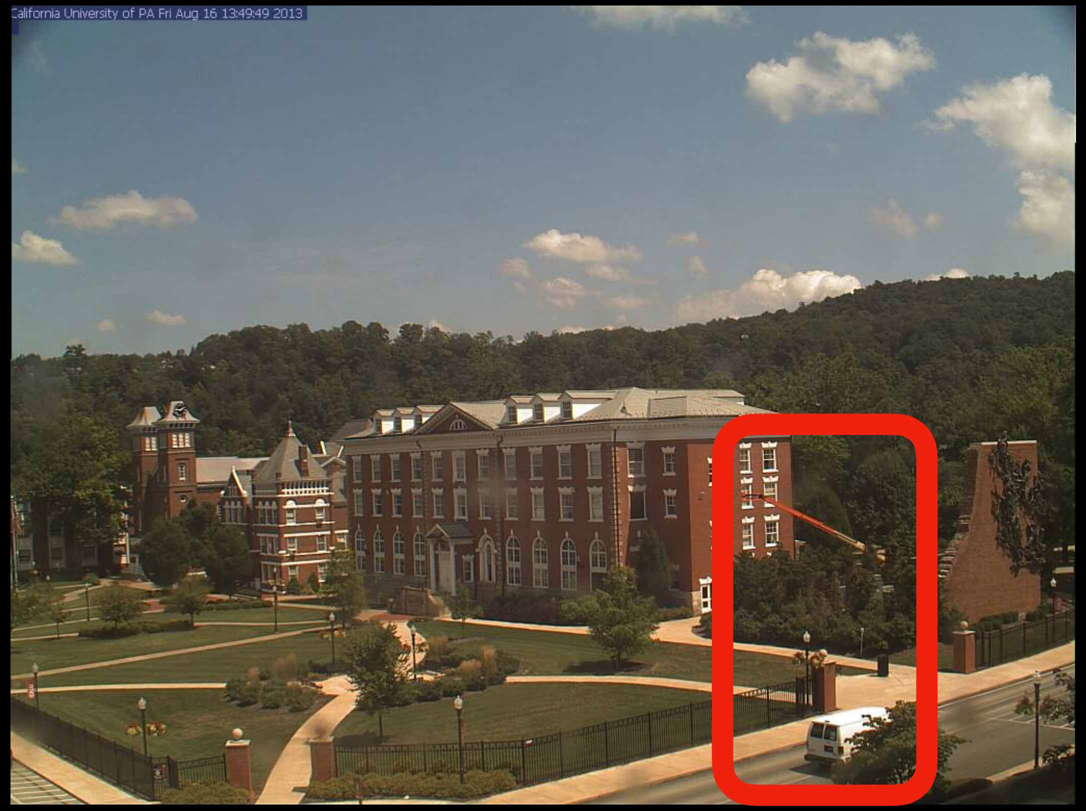
- The contributing views of AMOS Patches dataset [3] (a subset of AMOS [1, 2]):

pros:

- Overcome the limitations of AMOS (moving objects, blurry images, etc.) in a semi-automatic way [3];
- 27 scenes (50 images per scene) captured by static cameras exhibiting different weather and lighting conditions.

limitations:

- The filtering process is not perfect

[1] Jacobs et al.: The global network of outdoor webcams: properties and applications. *ACM SIGSPATIAL* 2009
[2] Jacobs et al.: Consistent temporal variations in many outdoor scenes. *CVPR* 2007
[3] Pultar et al.: Leveraging Outdoor Webcams for Local Descriptor Learning. *CVWW* 2019

Method overview: stylization. The proposed idea

- We use the contributing views of AMOS Patches dataset [3];
- The following **6** styles have been considered:

cloudy, dusk, mist, night, rainy, snow

- For each style, **10** images have been sampled manually;
- We post-process the images by removing watermarks and timestamps.

[1] Jacobs et al.: The global network of outdoor webcams: properties and applications. *ACM SIGSPATIAL* 2009

[2] Jacobs et al.: Consistent temporal variations in many outdoor scenes. *CVPR* 2007

[3] Pultar et al.: Leveraging Outdoor Webcams for Local Descriptor Learning. *CVWW* 2019

Method overview: stylization. Examples

Method overview: stylization. Examples

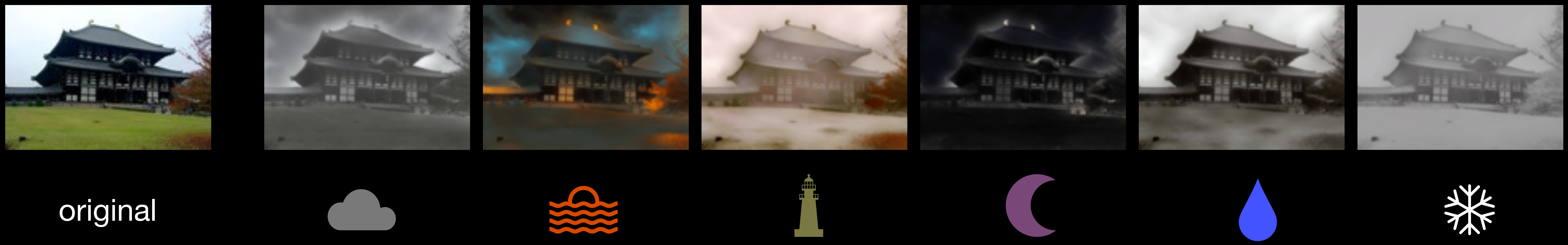
original image

stylized original image (10 versions)

…

Method overview: training pipeline

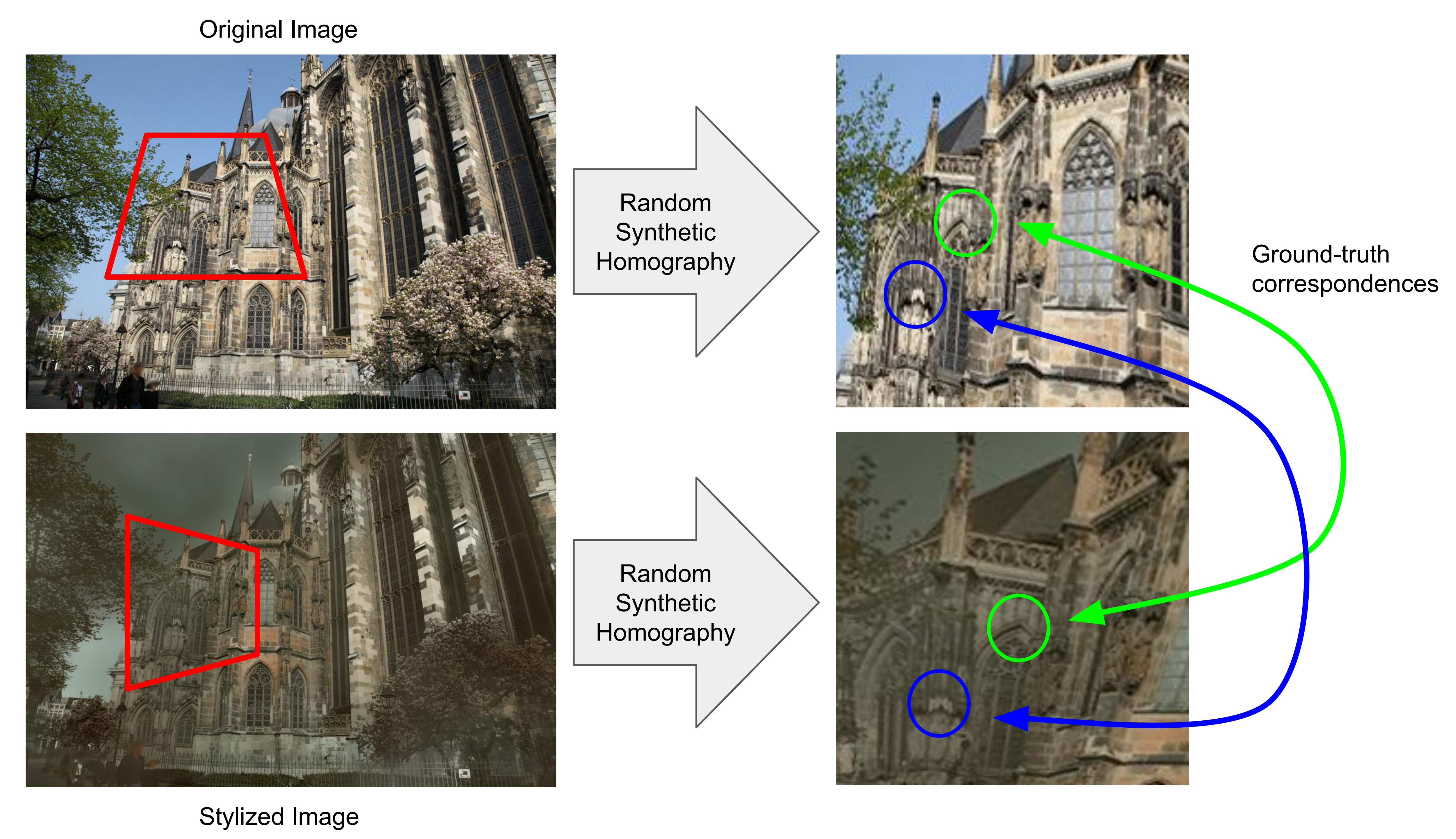
- Using Phototourism dataset:
  - From each training scene (16 scenes), we randomly sample **300** images;
  - For each image, we generate **60** stylized images, i.e 10 for each of 6 styles.
- For stylization, we use **FastPhotoStyle** [1] by NVIDIA;



[1] Li et al.: A Closed-form Solution to Photorealistic Image Stylization. *ECCV* 2018

Method overview: training pipeline

- The keypoint detector and descriptor (we use R2D2 [1] architecture) have been trained jointly by applying **synthetic homographies** to the original image and its stylized copy;



- The network was trained for 70 epochs (each epoch consists of **3900** image pairs) with warm-up of 5 epochs and exponentially decaying LR.

[1] Revaud et al.: R2D2: Repeatable and Reliable detector and descriptor. *NeurIPS* 2019

MLAD challenge

Goal: estimate the 6DoF relative pose between individual images from a **reference** sequence to a **test** sequence

reference

test_seq0

test_seq1

Results

Translation thresholds: 0.1m, 0.2m, 0.5m

Method	Scene: `test-sequence-0`	Scene: `test-sequence-1`
Superpoint [1]	15.5 / 27.5 / 47.5	9.0 / 19.4 / 36.4
D2-Net [2]	12.5 / 29.3 / 56.7	7.5 / 21.4 / 47.7
R2D2-5k [3]	21.5 / 33.1 / 53.0	12.3 / 23.7 / 42.0
Superpoint+SuperGlue [4]	21.2 / 33.9 / **60.0**	12.4 / **26.5** / **54.4**
ISRF-5k	20.9 / 33.2 / 54.8	**13.2** / 25.3 / 45.7
ISRF-10k	**22.3** / **34.1** / 57.4	**13.2** / 26.0 / 47.8

[1] DeTone et al.: Superpoint: Self-supervised interest point detection and description. *CVPRW* 2018

[2] Dusmanu et al.: D2-Net: A Trainable CNN for Joint Detection and Description of Local Features. *CVPR* 2019

[3] Revaud et al.: R2D2: Reliable and repeatable detector and descriptor. *NeurIPS* 2019

[4] Sarlin et al.: SuperGlue: Learning Feature Matching with Graph Neural Networks. *CVPR* 2020

Summary and observations

- We investigate if image stylisation can improve robustness of local features to illumination, weather and season changes;
- Using a single method, we could get competitive results on different localization datasets, despite training without 3D correspondences;
- Outlier rejection technique [1, 2, 3] did not help:
  - best model: **seq0**: 22.3 / 34.1 / 57.4; **seq1**: 13.2 / 26.0 / 47.8
  - best model + DEGENSAC: **seq0**: 18.5 / 28.2 / 47.9; **seq1**: 10.7 / 21.1 / 38.8
- The technical report is available online. Pre-trained models as well as the training dataset with stylized images will be released soon.

[1] Chum et al.: Locally optimized RANSAC. *Pattern Recognition* 2003
[2] Chum et al.: Two-View Geometry Estimation Unaffected by a Dominant Plane. *CVPR* 2005
[3] Mishkin et al.: MODS: Fast and robust method for two-view image matching. *CVIU* 2015

Image stylization for visual localization

Second place in two image-based localization challenges:

- **Map-based localization** for Autonomous driving: **this talk**
- **Local features** for day/night matches (workshop on Long-Term Visual Localization under Challenging Conditions): **August 28th, 11:40 - 11:55 UTC+1**

Thank you

Questions? Feel free to contact me :)

@iMelekhov

iaroslav.melekhov@aalto.fi

Aalto University
School of Science

NIANTIC
LABS