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;
- Localization datasets: challenges;
- Photorealistic image stylization for local keypoint/detector learning;
- Results;
- Summary and observations.

Localization pipeline

query

image retrieval → detection & description → feature matching and filtering (2D-3D correspondences) → Camera pose estimation [1] → 6DoF

retrieved db images

R2D2 detector and descriptor [2]

[1] Schoenberger et al.: Structure-from-Motion Revisited. *CVPR* 2016
[2] Revaud et al.: R2D2: Repeatable and Reliable detector and descriptor. *NeurIPS* 2019

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;
- Random images with synthetic color augmentations?
- 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

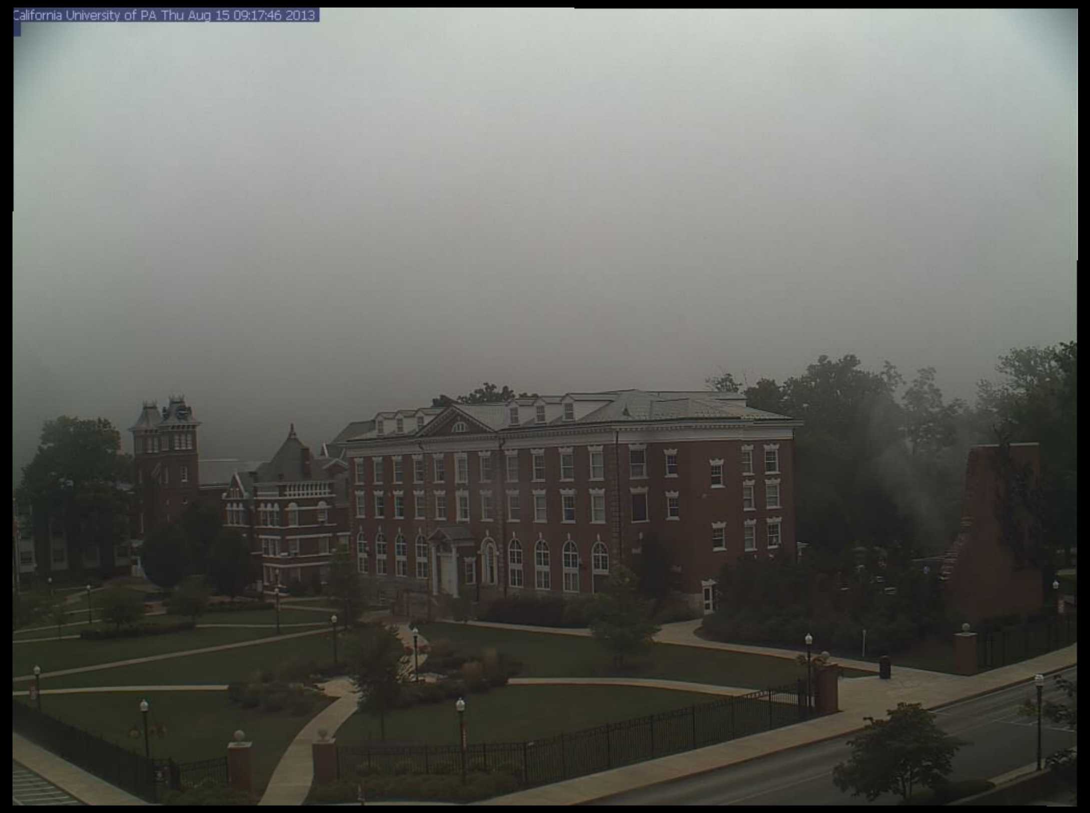
- The contributing views of AMOS Patches dataset [3] (a subset of AMOS [1, 2]):

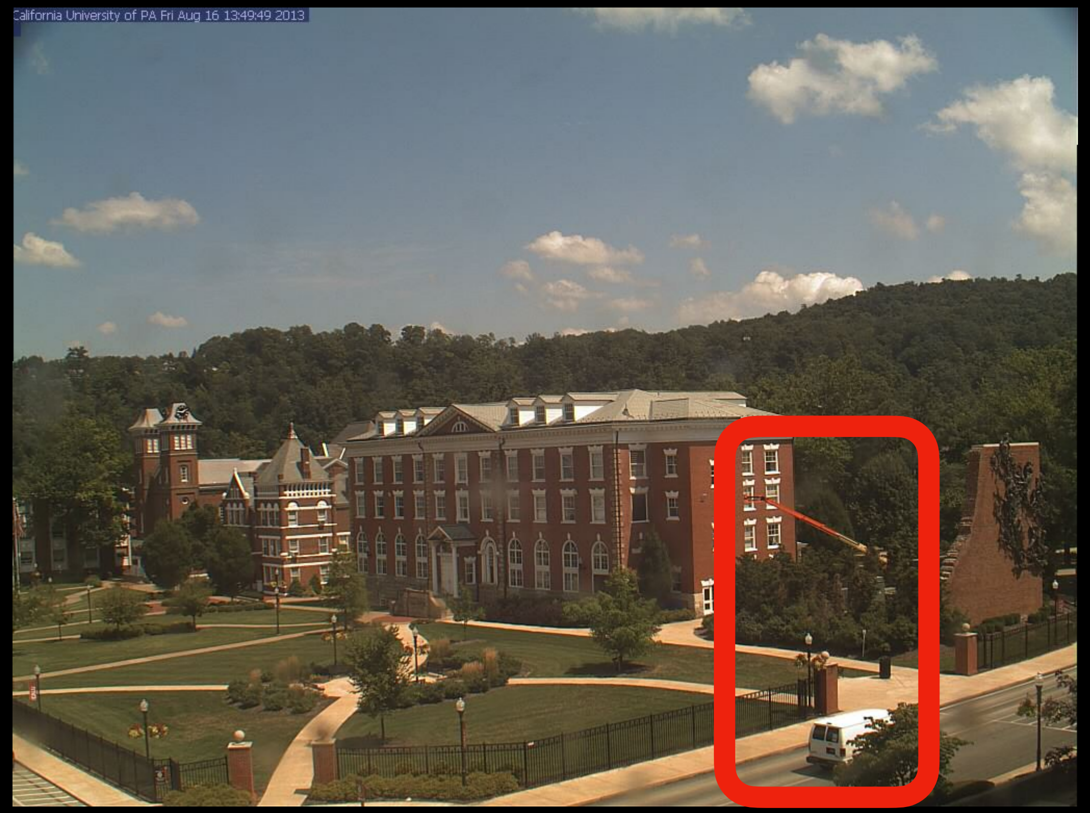
pros:

- 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

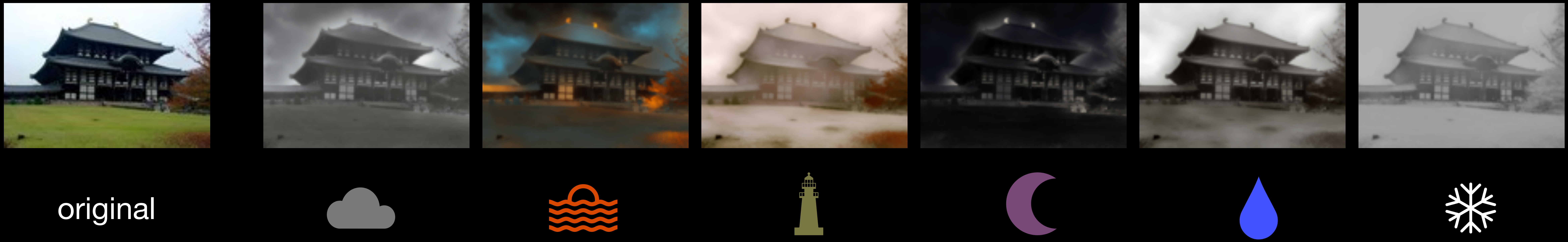
original image

stylized original image (10 versions)

...

Method overview: training pipeline

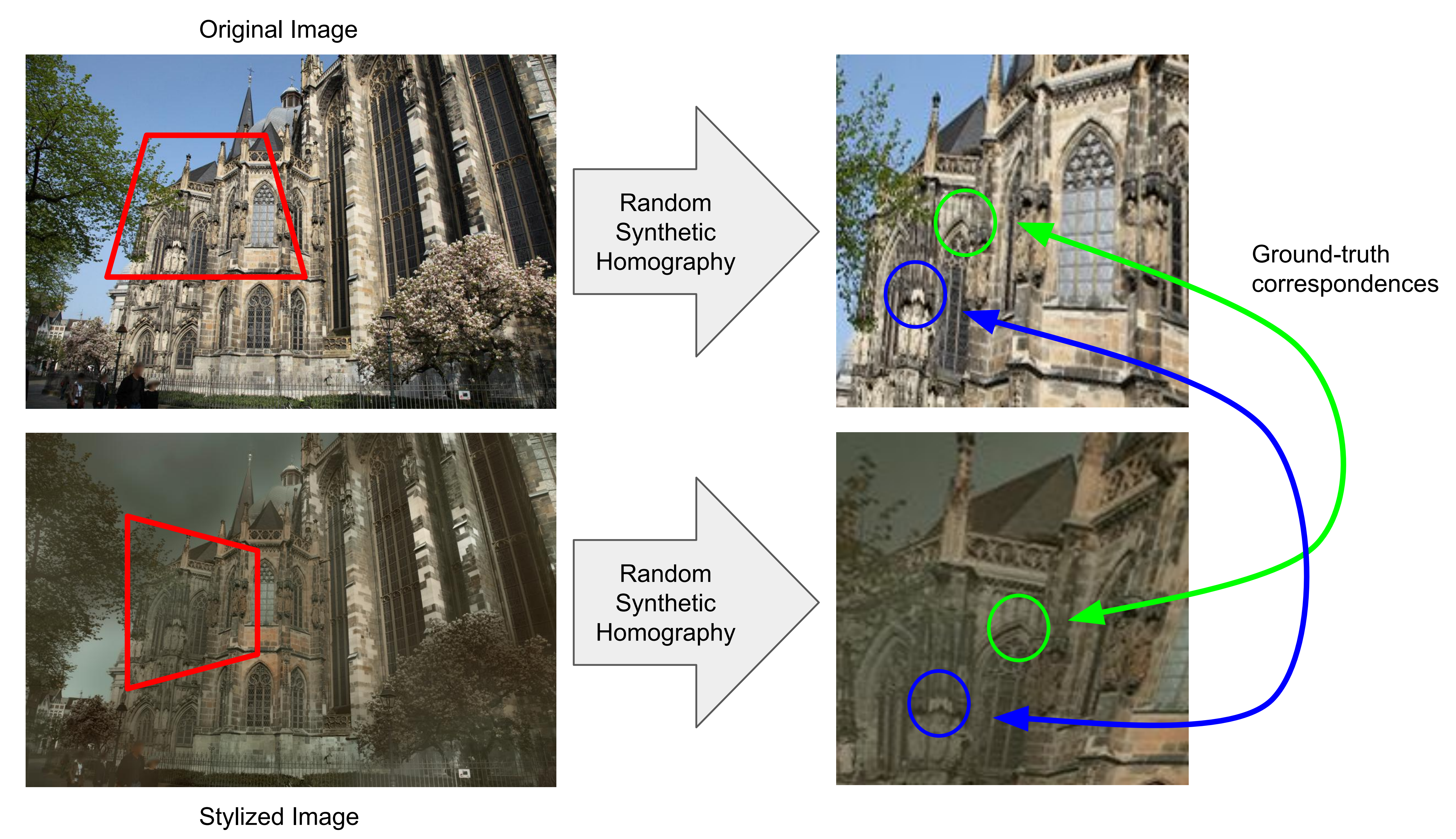
- Using Phototourism dataset:
  - From each training scene (16 scenes), we randomly sample **300** images;
  - For each image, we generate **60** stylized images, i.e 10 for each of 6 styles.
- For stylization, we use **FastPhotoStyle** [1] by NVIDIA;



[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

Relocalization: Color Augmentation vs Stylization

We use synthetic color augmentations: Gaussian blur, noise together with augmentation in brightness, contrast, saturation, and hue [5]: CA

	Aachen v1.1 [1]		RobotCar Seasons [2, 3]		SILDa [4]		
	day	night	day-overcast	other	evening	snow	night
ISRF-5k-**P**	87.9 / 94.7 / 98.5	68.1 / 81.7 / 94.8	56.3 / 80.7 / 95.9	17.5 / 31.9 / 42.4	31.6 / 65.2 / 85.1	0.3 / 12.2 / 64.4	28.7 / 53.4 / 78.4
ISRF-5k-**P-CA**	87.4 / 94.7 / 98.4	66.5 / 83.8 / 95.8	56.3 / 80.5 / 95.3	20.8 / 38.4 / 50.2	31.7 / 65.2 / 84.9	1.0 / 12.7 / 64.4	28.8 / 53.5 / 78.6
ISRF-5k-**P-S**	87.9 / 94.5 / 98.3	72.3 / 88.0 / 97.4	56.4 / 80.5 / 94.9	21.6 / 40.2 / 53.0	31.9 / 65.2 / 87.7	2.9 / 14.9 / 67.8	30.5 / 53.8 / 78.8

[1] Zhang et al.: Reference pose generation for visual localization via learned features and view synthesis. arXiv preprint. *arXiv:2005.05179* 2020
[2] Maddern et al.: 1 year, 1000 km: The Oxford robotcar dataset. *IJRR* 2017
[3] Sattler et al.: Benchmarking 6DoF outdoor visual localization in changing conditions. *CVPR* 2018
[4] Balntas et al.: SILDa: Scape Imperial Localization Dataset. https://www.visuallocalization.net/ 2019
[5] Buslaev et al.: Albumentations: fast and flexible image augmentations. *Information* 2020

LTVL challenge

- Visual localization for autonomous vehicles (I) [1, 2]: Extended CMU, RobotCar (v2), SILDa
- Visual localization for handheld devices (II) [1, 2]: Aachen v1.1, InLoc
- Local feature track (III): Aachen v1.1 (night)
the list of images to be matched is provided

[1] Radenović et al.: Fine-tuning CNN Image Retrieval with No Human Annotation. *TPAMI* 2018
[2] Radenović et al.: CNN Image Retrieval Learns from BoW: Unsupervised Fine-Tuning with Hard Examples. *ECCV* 2016

LTVL challenge: track I and track II

Track I

	Extended CMU Seasons			RobotCar Seasons V2		SILDa		
	urban	suburban	park	day all	night all	evening	snow	night
HLoc-Superpoint-Superglue	98.1 / 99.8 / 99.9	98.3 / 99.5 / 100.0	94.2 / 97.1 / 98.5	63.8 / 95.0 / 100.0	45.0 / 86.2 / 94.6	35.5 / 75.0 / 97.1	0.0 / 2.4 / 86.3	31.7 / 54.4 / 81.9
KAPTURE-R2D2-FUSION	97.0 / 99.1 / 99.8	95.0 / 97.0 / 99.4	89.2 / 93.4 / 97.5	66.0 / 95.1 / 100.0	46.2 / 76.5 / 91.4	32.4 / 67.4 / 93.3	0.2 / 4.1 / 88.9	30.4 / 54.2 / 81.1
ISRF 5k	93.8 / 96.6 / 98.2	83.5 / 86.8 / 90.5	76.4 / 80.9 / 85.7	NA	NA	31.9 / 65.2 / 87.7	2.9 / 14.9 / 67.8	30.5 / 53.8 / 78.8

Track II

	Aachen v1.1		InLoc	
	day	night	duc1	duc2
HLoc-Superpoint-Superglue	89.8 / 96.1 / 99.4	77.0 / 90.6 / 100.0	49.0 / 68.7 / 80.8	53.4 / 77.1 / 82.4
RLOCS_v1.0	86.0 / 94.8 / 98.8	72.3 / 88.5 / 99.0	47.0 / 71.2 / 84.8	58.8 / 77.9 / 80.9
KAPTURE-R2D2-FUSION	90.9 / 96.7 / 99.5	78.5 / 91.1 / 98.4	41.4 / 60.1 / 73.7	47.3 / 67.2 / 73.3
ISRF 5k	87.1 / 94.7 / 98.3	74.3 / 86.9 / 97.4	39.4 / 58.1 / 70.2	41.2 / 61.1 / 69.5

The Local Feature track

Track III

Method	night
Superpoint + SuperGlue [1,2]	73.3 / 88.0 / 98.4
ISRF (ours)	**69.1 / 87.4 / 98.4**
LISRD [3] + Superpoint kpts + AdaLAM [4]	73.3 / 86.9 / 97.9
R2D2_40k	71.2 / 86.9 / 97.9
LISRD + Superpoint keypoints	72.3 / 86.4 / 97.4
FBpoint (single scale, AdaLAM)	72.8 / 85.9 / 96.3
FBpoint (single scale, mutual matcher)	71.2 / 85.9 / 95.8

[1] DeTone et al.: Superpoint: Self-supervised interest point detection and description. *CVPRW* 2018
[2] Sarlin et al.: SuperGlue: Learning Feature Matching with Graph Neural Networks. *CVPR* 2020
[3] Pautrat et al.: Online Invariance Selection for Local Feature Descriptors. *ECCV* 2020
[4] Cavalli et al.: AdaLAM: Revisiting Handcrafted Outlier Detection. preprint *arXiv:2006.04250*, 2020

The Local Feature track: Ablation study

Method	Score	Rank
ISRF, 5k, N=16, mutual matcher	67.5 / 86.9 / 95.8	6-7
ISRF, 5k, N=16, AdaLAM [1]	64.4 / 85.3 / 96.9	7
ISRF, 10k, N=16, mutual matcher	69.6 / 86.4 / 97.4	5-6
ISRF, 20k, N=8, mutual matcher	69.1 / 86.4 / 97.9	5-6
ISRF, 20k, N=8, AdaLAM	**69.1 / 87.4 / 98.4**	**2**

[1] Cavalli et al.: AdaLAM: Revisiting Handcrafted Outlier Detection. preprint *arXiv:2006.04250*, 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;
- The color augmentations at training time increases robustness to appearance changes, but it is not sufficient by itself;
- SuperGlue?;
- The technical report is available online. Pre-trained models as well as the training dataset with stylized images will be released soon.

Thank you

Questions? Feel free to contact me :)

@iMelekhov

iaroslav.melekhov@aalto.fi

Aalto University
School of Science

NIANTIC
LABS