Image Matching: Local Features and Beyond

CVPR 2021 Workshop
June 25, 9:00 - 13:20 MT

Vassileios Balntas (Facebook)
Vincent Lepetit (ENPC ParisTech)
Jiri Matas (CTU Prague)
Dmytro Mishkin (CTU Prague)
Johannes Schönberger (MS)
Eduard Trulls (Google)
Kwang Moo Yi (UBC)

Organizers

Vassileios Balntas	Vincent Lepetit	Jiri Matas	Dmytro Mishkin	Johannes Schönberger	Eduard Trulls	Kwang Moo Yi
Facebook	ENPC ParisTech	Czech Technical University	Czech Technical University	Microsoft	Google	University of British Columbia

Workshop website: https://image-matching-workshop.github.io

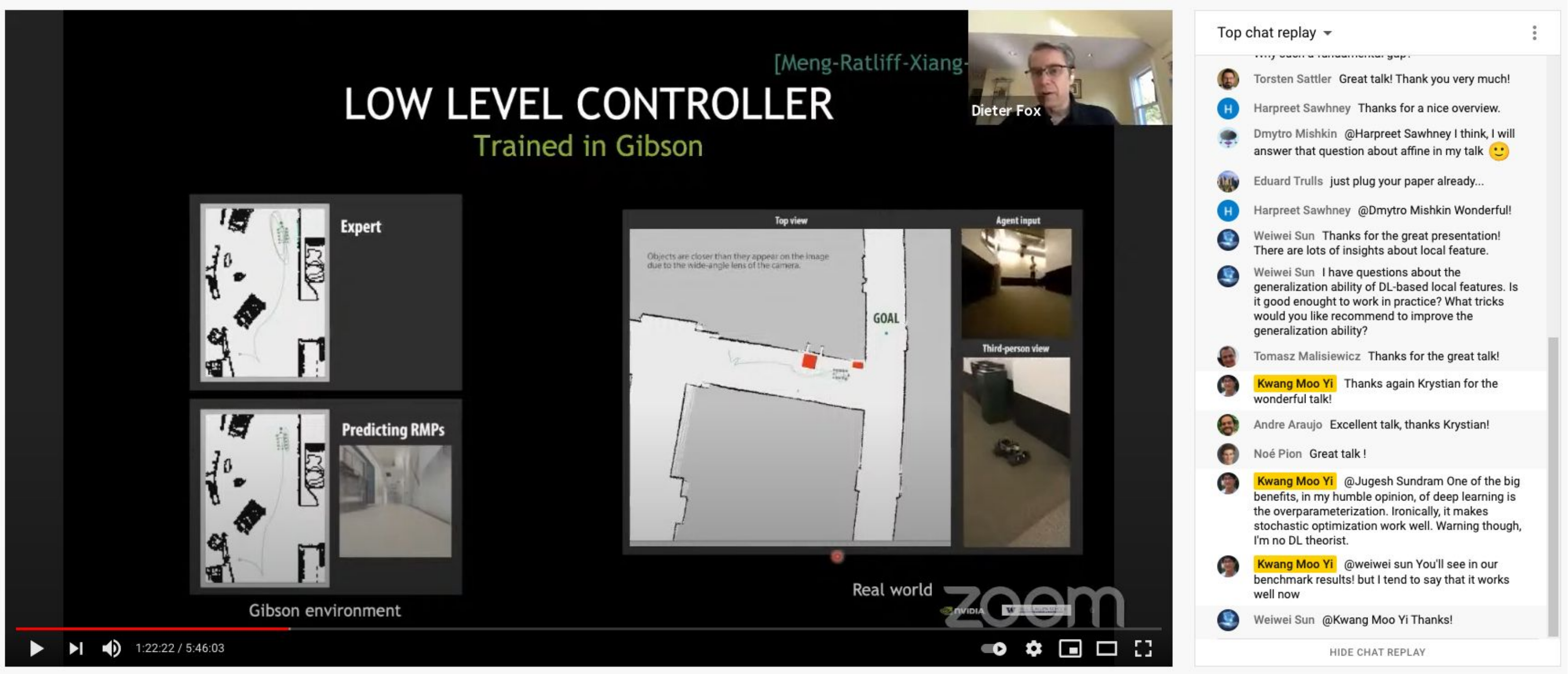
9:00 - 9:15	**Welcome session (Eduard Trulls)**
9:15 - 10:00	**Invited Talk: Davide Scaramuzza (University of Zurich/ETH Zurich)**
10:00- 10:45	**Invited Talk: Marc Pollefeys (ETH Zurich/Microsoft)**
10:45 - 11:00	***Perceptual Loss for Robust Unsupervised Homography Estimation*** Daniel Koguciuk (Advanced Research Lab, NavInfo Europe, NL)
11:00 - 11:15	***DFM: A Performance Baseline for Deep Feature Matching*** Ufuk Efe (Middle East Technical University, Ankara, Turkey)
11:15 - 11:45	**Challenge presentation**
11:45 - 12:15	**Open discussion**
12:15 - 13:35	**Challenge participant talks** 12:15-12:25: Fabio Bellavia (University of Palermo) 12:25-12:35: Prune Truong (ETH Zurich) 12:35-12:45: Jiaming Sun/Xingyi He (Zhejiang University, SenseTime Research) 12:45-12:55: Wei Jiang (University of British Columbia) 12:55-13:05: Megvii 3D 13:05-13:15: Tencent
13:15 - 13:20	**Closing**

Workshop website: https://image-matching-workshop.github.io

9:00 - 9:15	**Welcome session (Eduard Trulls)**
9:15 - 10:00	**Invited Talk: Davide Scaramuzza (University of Zurich/ETH Zurich)**
10:00- 10:45	**Invited Talk: Marc Pollefeys (ETH Zurich/Microsoft)**
10:45 - 11:00	***Perceptual Loss for Robust Unsupervised Homography Estimation*** Daniel Koguciuk (Advanced Research Lab, NavInfo Europe, NL)
11:00 - 11:15	***DFM: A Performance Baseline for Deep Feature Matching*** Ufuk Efe (Middle East Technical University, Ankara, Turkey)
11:15 - 11:45	**Challenge presentation**
11:45 - 12:15	**Open discussion**
12:15 - 13:35	**Challenge participant talks** 12:15-12:25: Fabio Bellavia (University of Palermo) 12:25-12:35: Prune Truong (ETH Zurich) 12:35-12:45: Jiaming Sun/Xingyi He (Zhejiang University, SenseTime Research) 12:45-12:55: Wei Jiang (University of British Columbia) 12:55-13:05: Megvii 3D 13:05-13:15: Tencent
13:15 - 13:20	**Closing**

Live on YouTube!

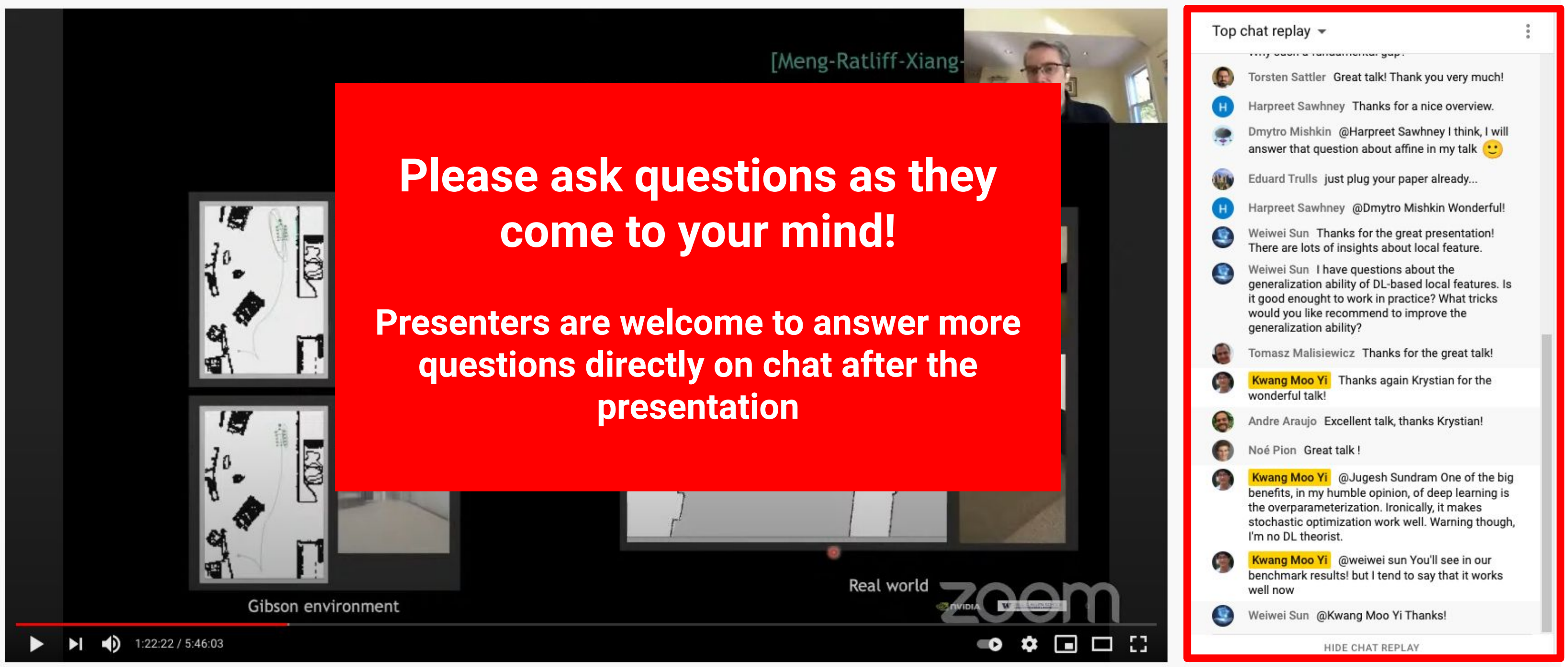
Zoom link is available on the CVPR website



Live on YouTube!

Zoom link on CVPR website

Please ask questions on chat!
The organizers will pass them on to the speakers



Please ask questions as they come to your mind!

Presenters are welcome to answer more questions directly on chat after the presentation

Focal point: Matching rigid structures

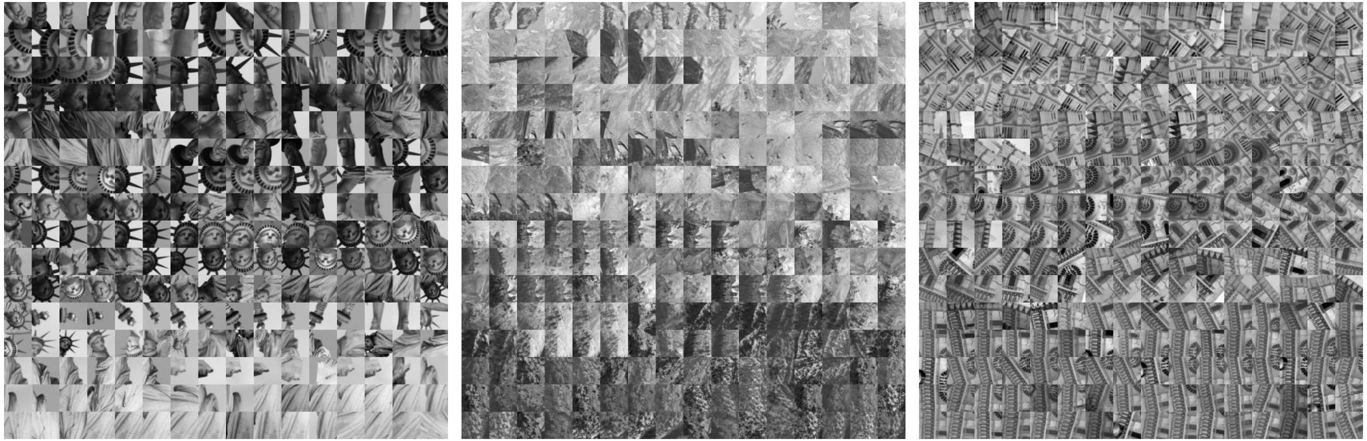
- 3D reconstruction (stereo, SfM) across baselines, time, weather, etc.
- Link in common: "***Local features..***" remain SOTA.
- "***... and beyond***"**:** but may not always be the case.

Why did we start this workshop?

- New papers come out all the time, but *what does actually work?*
- Benchmarks are often saturated, sub-optimal, biased, or de-centralized.

Why did we start this workshop?

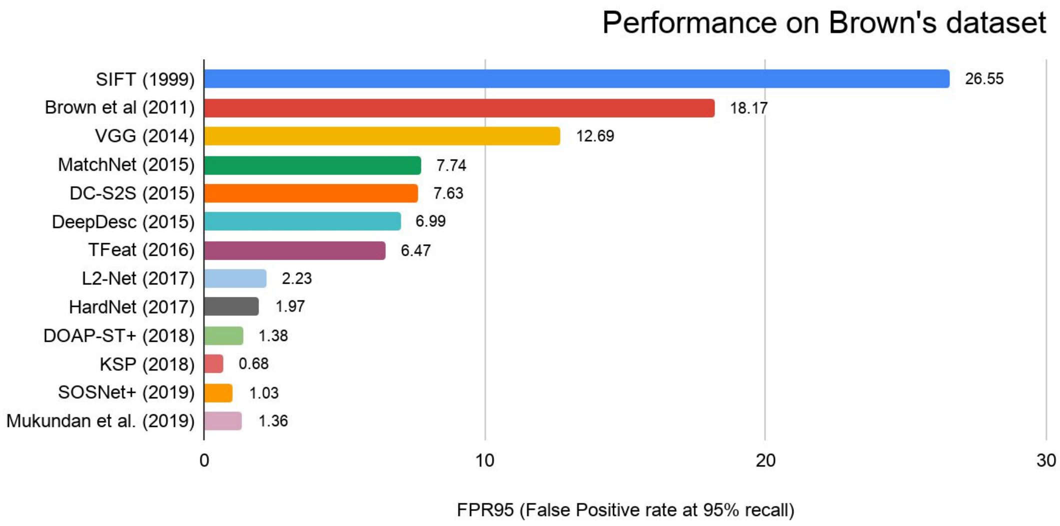
- New papers come out all the time, but *what does actually work?*
- Benchmarks are often saturated, sub-optimal, biased, or de-centralized.



Discriminative Learning of Local Image Descriptors. Brown et al., PAMI'10

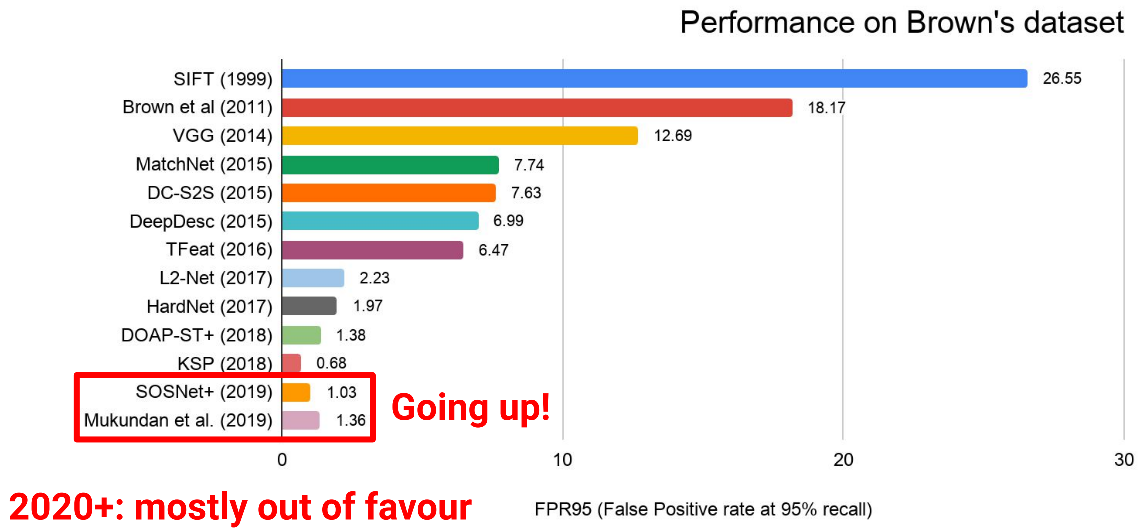
Why did we start this workshop?

- New papers come out all the time, but *what does actually work?*
- Benchmarks are often saturated, sub-optimal, biased, or de-centralized.



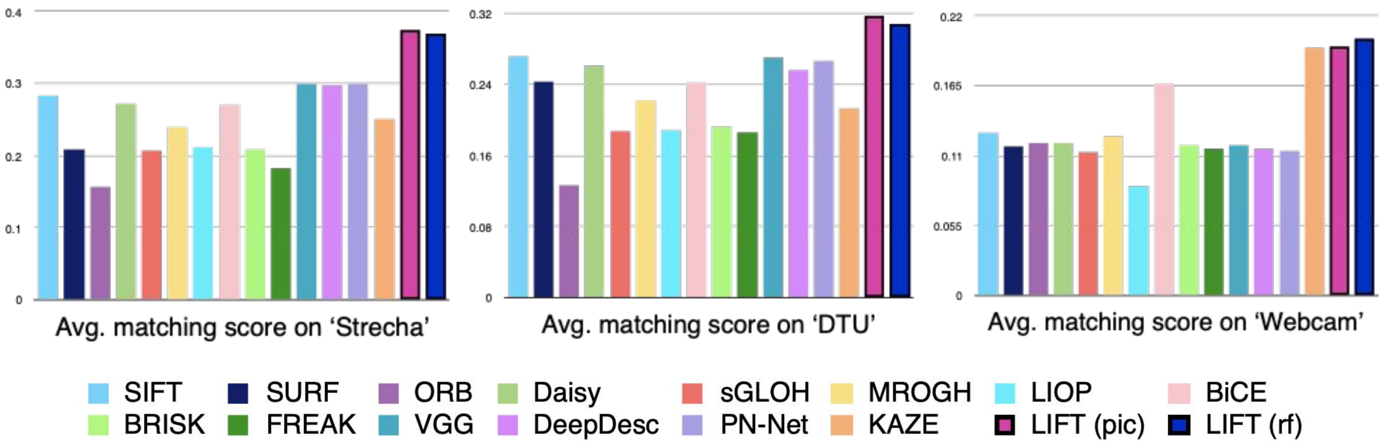
Why did we start this workshop?

- New papers come out all the time, but *what does actually work?*
- Benchmarks are often saturated, sub-optimal, biased, or de-centralized.



Why did we start this workshop?

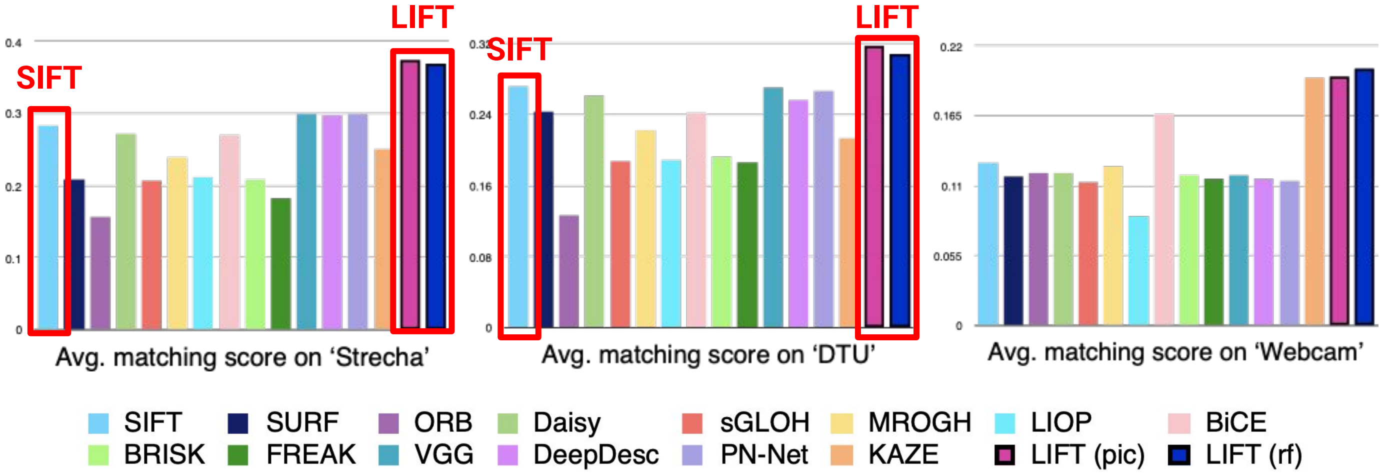
- New papers come out all the time, but *what does actually work?*
- Benchmarks are often saturated, sub-optimal, biased, or de-centralized.



LIFT: Learned Invariant Feature Transform. Yi et al., ECCV'16

Why did we start this workshop?

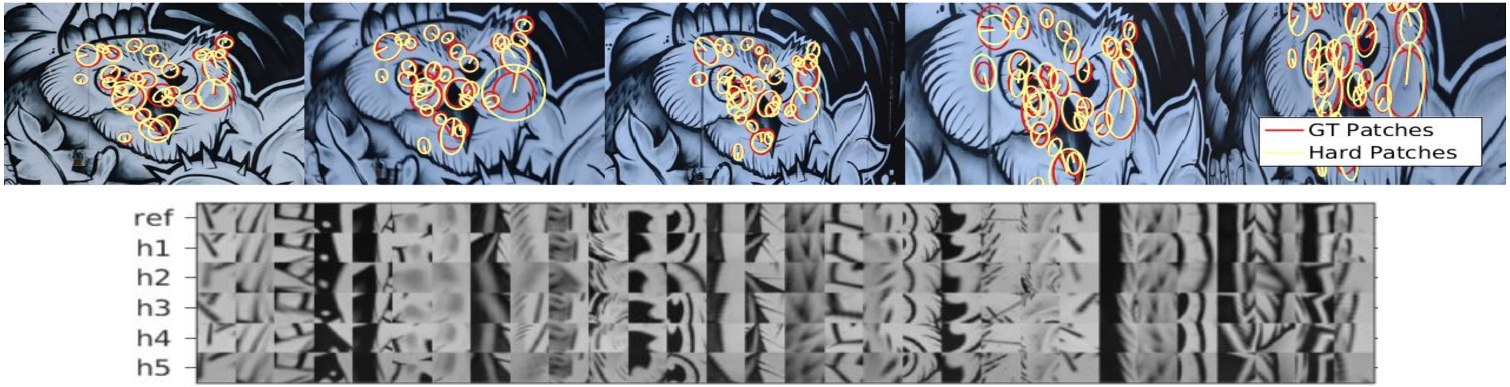
- New papers come out all the time, but *what does actually work?*
- Benchmarks are often saturated, sub-optimal, biased, or de-centralized.



LIFT: Learned Invariant Feature Transform. Yi et al., ECCV'16

Why did we start this workshop?

- New papers come out all the time, but *what does actually work?*
- Benchmarks are often saturated, sub-optimal, biased, or de-centralized.

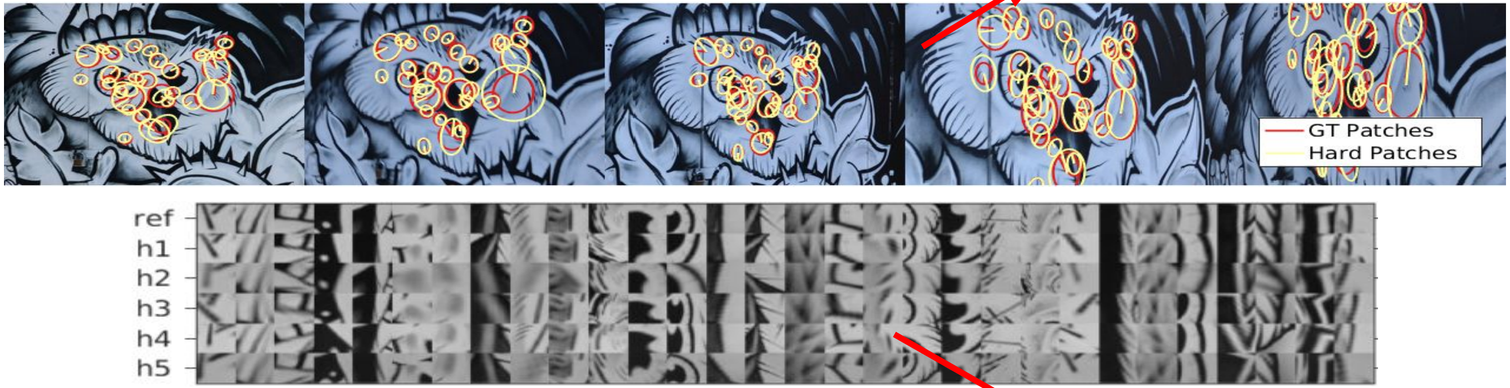


HPatches: A benchmark and evaluation of handcrafted and learned local descriptors. V. Balntas et al., CVPR'17
Source: github.com/hpatches/hpatches-dataset

Why did we start this workshop?

- New papers come out all the time, but *what does actually work?*
- Benchmarks are often saturated, sub-optimal, biased, or de-centralized.

Only perspective/illumination changes

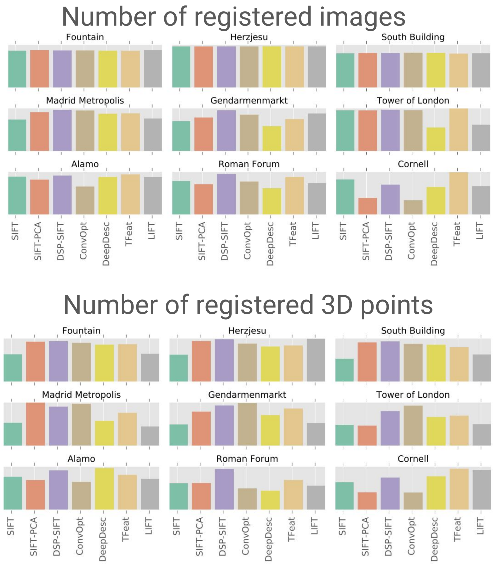


Patch-based/intermediate metrics

[HPatches: A benchmark and evaluation of handcrafted and learned local descriptors](https://github.com/hpatches/hpatches-dataset). V. Balntas et al., CVPR'17
Source: github.com/hpatches/hpatches-dataset

Why did we start this workshop?

- New papers come out all the time, but *what does actually work?*
- Benchmarks are often saturated, sub-optimal, biased, or de-centralized.



[Comparative Evaluation of Hand-Crafted and Learned Local Features](https://github.com/ahojnnes/local-feature-evaluation).
Schönberger et al., CVPR'17.
Source: github.com/ahojnnes/local-feature-evaluation

Why did we start this workshop?

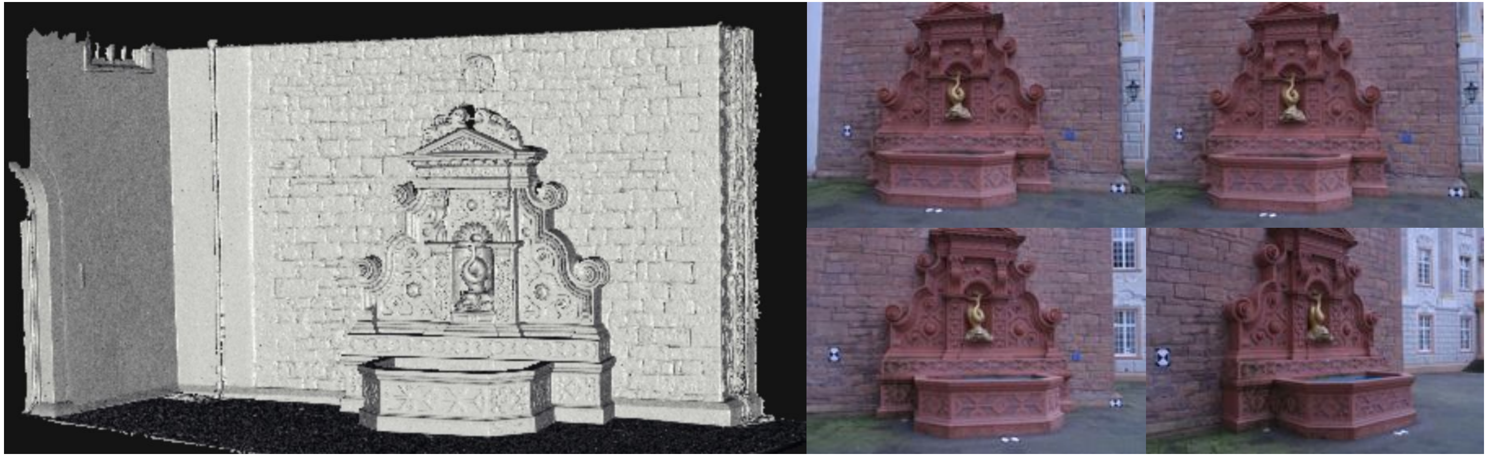
Large-scale, but no Ground Truth ⇒ Intermediate metrics

- New papers come out all the time, but *what does actually work?*
- Benchmarks are often saturated, sub-optimal, biased, or de-centralized.

[Comparative Evaluation of Hand-Crafted and Learned Local Features](https://github.com/ahojnnes/local-feature-evaluation).
Schönberger et al., CVPR'17.
Source: github.com/ahojnnes/local-feature-evaluation

Why did we start this workshop?

- New papers come out all the time, but *what does actually work?*
- Benchmarks are often saturated, sub-optimal, biased, or de-centralized.

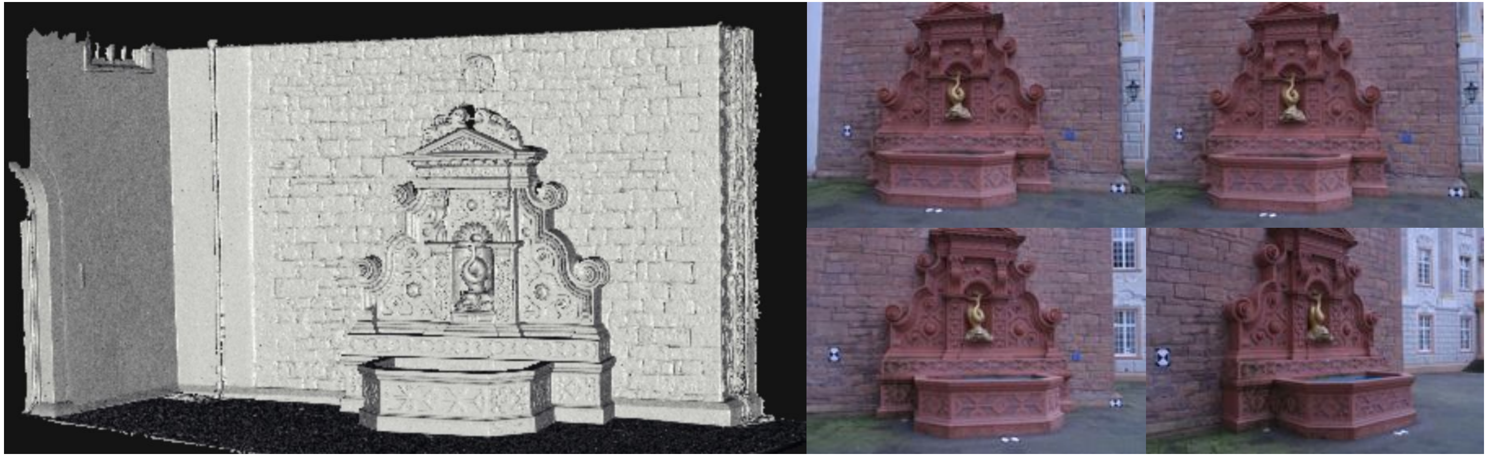


On benchmarking camera calibration and multi-view stereo for high resolution imagery. Strecha et al., CVPR'08.

Why did we start this workshop?

2-3 scenes, <100 images

- New papers come out all the time, but *what does actually work?*
- Benchmarks are often saturated, sub-optimal, biased, or de-centralized.



On benchmarking camera calibration and multi-view stereo for high resolution imagery. Strecha et al., CVPR'08.

IMW 2021: Leaderboard

Current version: 4fa67519 (2021-06-24, 18:01 UTC)

Summary:

- The challenge features three dataset with two tracks each: **stereo** and **multi-view** (see this page for details).
  - Phototourism dataset: unlimited keypoints (8k), restricted keypoints (2k)
  - PragueParks dataset: unlimited keypoints (8k), restricted keypoints (2k)
  - GoogleUrban dataset: unlimited keypoints (8k), restricted keypoints (2k)
- Performance is **averaged by rank across all datasets and tasks** using mean Average Accuracy (mAA) at a 10-degree error threshold.
- Submissions are broken down into **categories** by **number of features:** up to 2048 keypoints ("restricted") and 8000 keypoints ("unlimited").
- ~~Descriptors must have a maximum size of 512 bytes (128f). Submissions using larger descriptors will not be processed.~~**May 25, 2021: You may now use descriptors of any size.**
- Categories are non-exclusive: submissions on the "restricted" category compete with the "unlimited" category, as they are a subset of it.

Please note that this is a static website: you may want to force a reload if it does not update properly.

Leaders: Unlimited keypoints category

	Phototourism		PragueParks		GoogleUrban		Combined
Method	**Stereo**	**Multiview**	**Stereo**	**Multiview**	**Stereo**	**Multiview**	**Avg. Rank**
#1: sp_disk_scale_8k	0.63975 Rank: 1	0.78564 Rank: 1	0.80700 Rank: 2	0.49878 Rank: 6	0.43952 Rank: 1	0.33734 Rank: 8	3.17
#2: mss_scale_adapt_f_8k	0.60357 Rank: 8	0.77994 Rank: 7	0.79766 Rank: 3	0.50230 Rank: 2	0.41212 Rank: 3	0.32932 Rank: 19	7.00
#3: mss_scale_8k	0.60357 Rank: 8	0.78290 Rank: 2	0.79766 Rank: 3	0.50499 Rank: 1	0.41212 Rank: 3	0.32472 Rank: 26	7.17
#4: ss-dpth	0.59698 Rank: 9	0.78169 Rank: 4	0.75562 Rank: 18	0.49106 Rank: 19	0.41076 Rank: 5	0.34053 Rank: 4	9.83
#5: ss-unc-yt	0.59614 Rank: 10	0.78224 Rank: 3	0.72704 Rank: 36	0.50130 Rank: 4	0.40856 Rank: 6	0.33532 Rank: 11	11.67

Leaders: Restricted keypoints category

	Phototourism		PragueParks		GoogleUrban		Combined
Method	**Stereo**	**Multiview**	**Stereo**	**Multiview**	**Stereo**	**Multiview**	**Avg. Rank**
#1: ss-dpth	0.59698 Rank: 1	0.78169 Rank: 2	0.75562 Rank: 15	0.49106 Rank: 16	0.41076 Rank: 3	0.34053 Rank: 4	6.83
#2: ss-unc-yt	0.59614 Rank: 2	0.78224 Rank: 1	0.72704 Rank: 27	0.50130 Rank: 2	0.40856 Rank: 4	0.33532 Rank: 10	7.67
#3: mssscalev2	0.59205 Rank: 7	0.77662 Rank: 10	0.77377 Rank: 5	0.50092 Rank: 3	0.40769 Rank: 9	0.32729 Rank: 18	8.67
#4: ss-two-stg	0.59173 Rank: 8	0.77978 Rank: 5	0.75870 Rank: 14	0.48749 Rank: 25	0.40777 Rank: 8	0.33969 Rank: 5	10.83
#5: mss_orien	0.59211 Rank: 5	0.77621 Rank: 11	0.77654 Rank: 3	0.49760 Rank: 4	0.40210 Rank: 17	0.32353 Rank: 25	10.83

Phototourism: unlimited keypoints

Note: entries with the same multi-view configuration may seem duplicated. This is normal: performance is averaged across tasks.

Show 10 entries

Search:

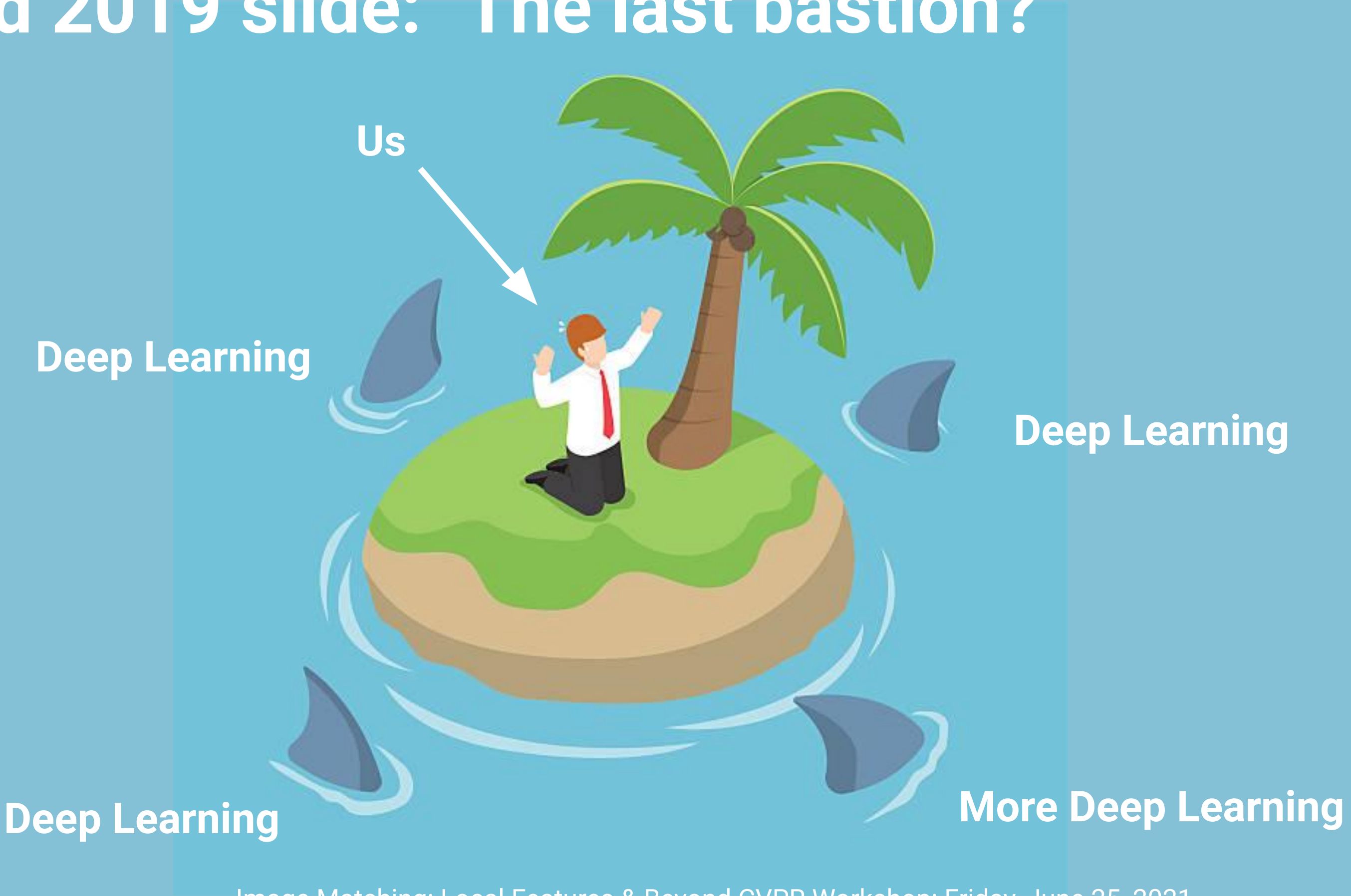
Stereo	Multiview	Avg.

Solution: open challenge!

- Workshop has invited talks and papers, but centered on the challenge
- Show that proper evaluation is key → IJCV'20 paper (arxiv/2003.01587)
  - Further discussion at 11:15!
- Focus on where ***theory*** meets ***practice***
- Meeting point for domain experts in order to figure out the SOTA

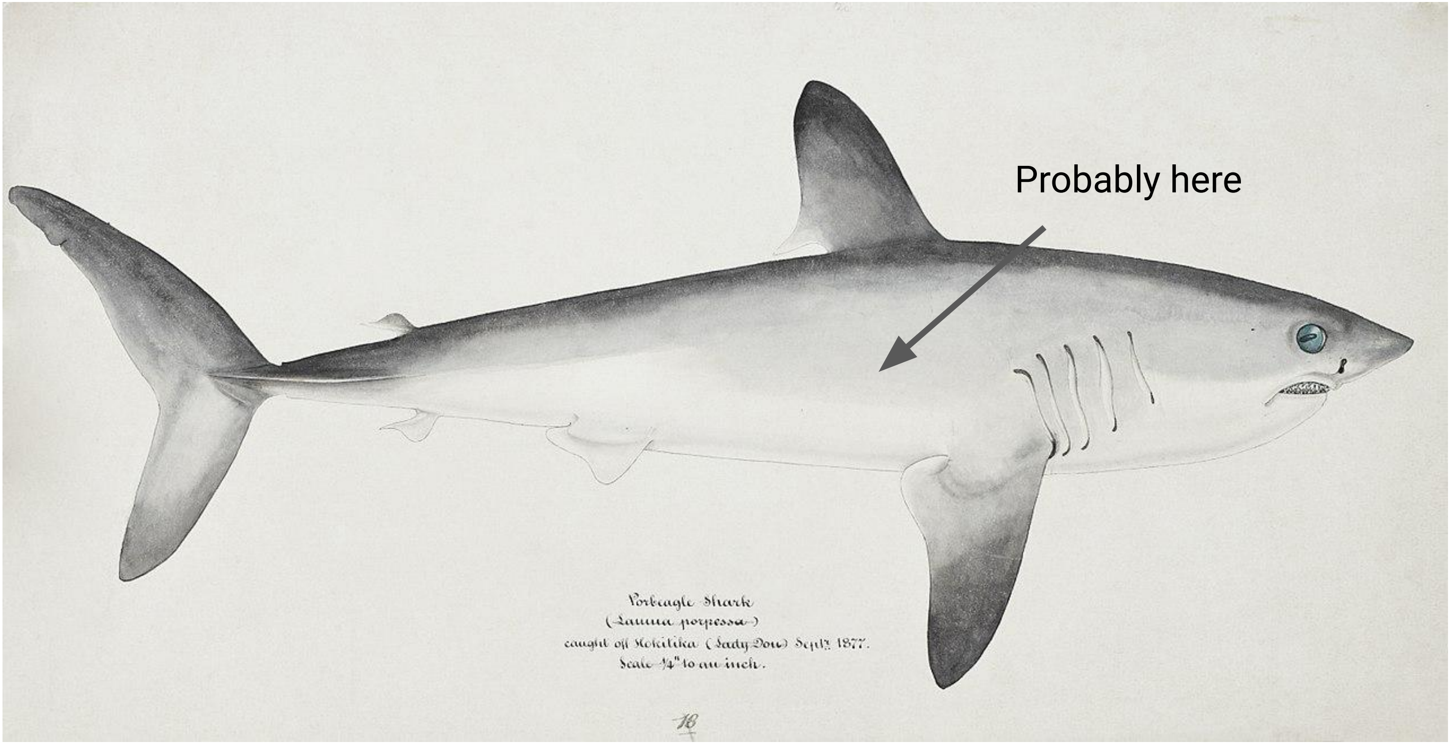
The old 2019 slide: "The last bastion?"

The old 2019 slide: "The last bastion?"



And then 2020...

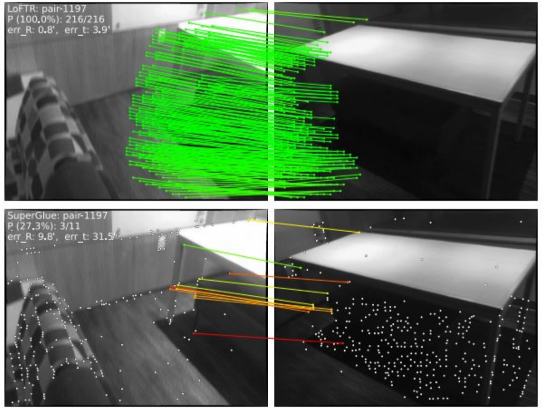
Where are we in 2021?



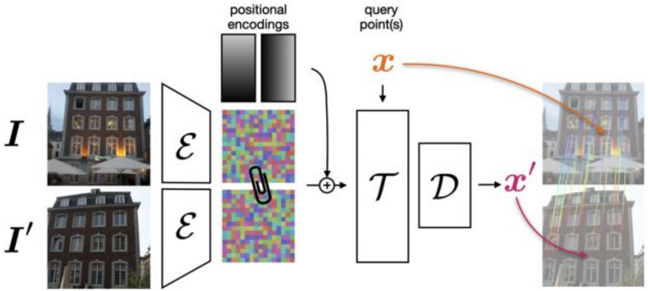
Where are we in 2021?

- **2019:** First version of the workshop and challenge (IMC + SILDa)
    - Winners used learned patch descriptors (ContextDesc, HardNet, etc) + CNe matching
- **2020:** Open-sourced benchmark codebase
    - Many top performers were "papers" (SuperGlue, AdaLAM, DISK)
- **2021:** Two new datasets and a new challenge (More at 12:15+)
    - IMC: PhotoTourism, **PragueParks, GoogleUrban**
    - Synthetic dataset: **SimLocMatch**
    - Top performers have more "engineering"
- What about **2022?** Open discussion at 11:45
    - What can we do better? What do we need to remain relevant?

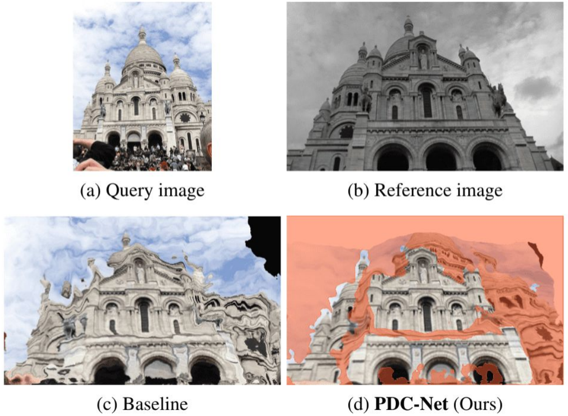
But we are moving away from local features...



"LoFTR: Detector-Free Local Feature Matching with Transformers", Sun et al (CVPR'21)



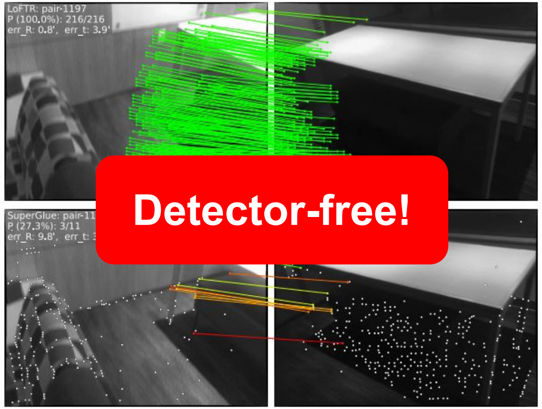
"COTR: Correspondence Transformer for Matching Across Images", Jiang et al (arxiv'21)



"Learning Accurate Dense Correspondences and When to Trust Them", Truong et al (CVPR'21)

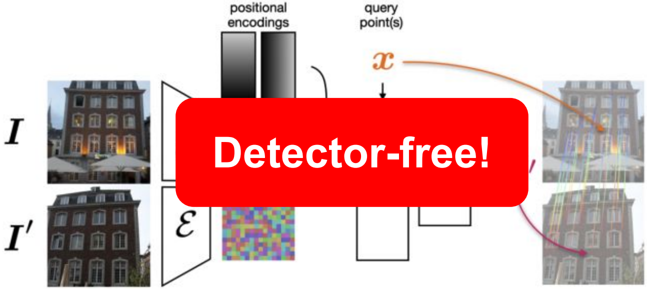
But we are moving away from local features...

Talk at 12:35!



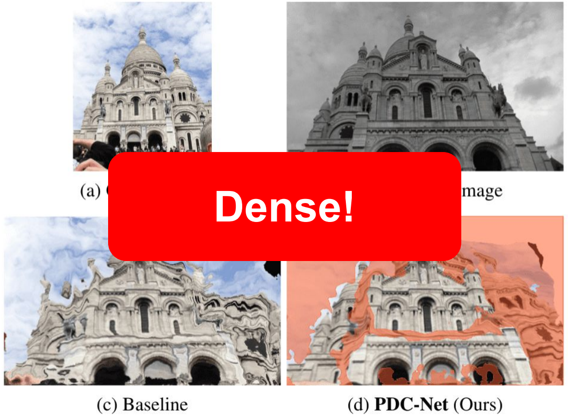
"LoFTR: Detector-Free Local Feature Matching with Transformers", Sun et al (CVPR'21)

Talk at 12:45!



"COTR: Correspondence Transformer for Matching Across Images", Jiang et al (arxiv'21)

Talk at 12:25!



"Learning Accurate Dense Correspondences and When to Trust Them", Truong et al (CVPR'21)

Workshop website: https://image-matching-workshop.github.io

9:00 - 9:15	**Welcome session (Eduard Trulls)**
9:15 - 10:00	**Invited Talk: Davide Scaramuzza (University of Zurich/ETH Zurich)**
10:00- 10:45	**Invited Talk: Marc Pollefeys (ETH Zurich/Microsoft)**
10:45 - 11:00	***Perceptual Loss for Robust Unsupervised Homography Estimation*** Daniel Koguciuk (Advanced Research Lab, NavInfo Europe, NL)
11:00 - 11:15	***DFM: A Performance Baseline for Deep Feature Matching*** Ufuk Efe (Middle East Technical University, Ankara, Turkey)
11:15 - 11:45	**Challenge presentation**
11:45 - 12:15	**Open discussion**
12:15 - 13:35	**Challenge participant talks** 12:15-12:25: Fabio Bellavia (University of Palermo) 12:25-12:35: Prune Truong (ETH Zurich) 12:35-12:45: Jiaming Sun/Xingyi He (Zhejiang University, SenseTime Research) 12:45-12:55: Wei Jiang (University of British Columbia) 12:55-13:05: Megvii 3D 13:05-13:15: Tencent
13:15 - 13:20	**Closing**

(Keynote/paper talks)

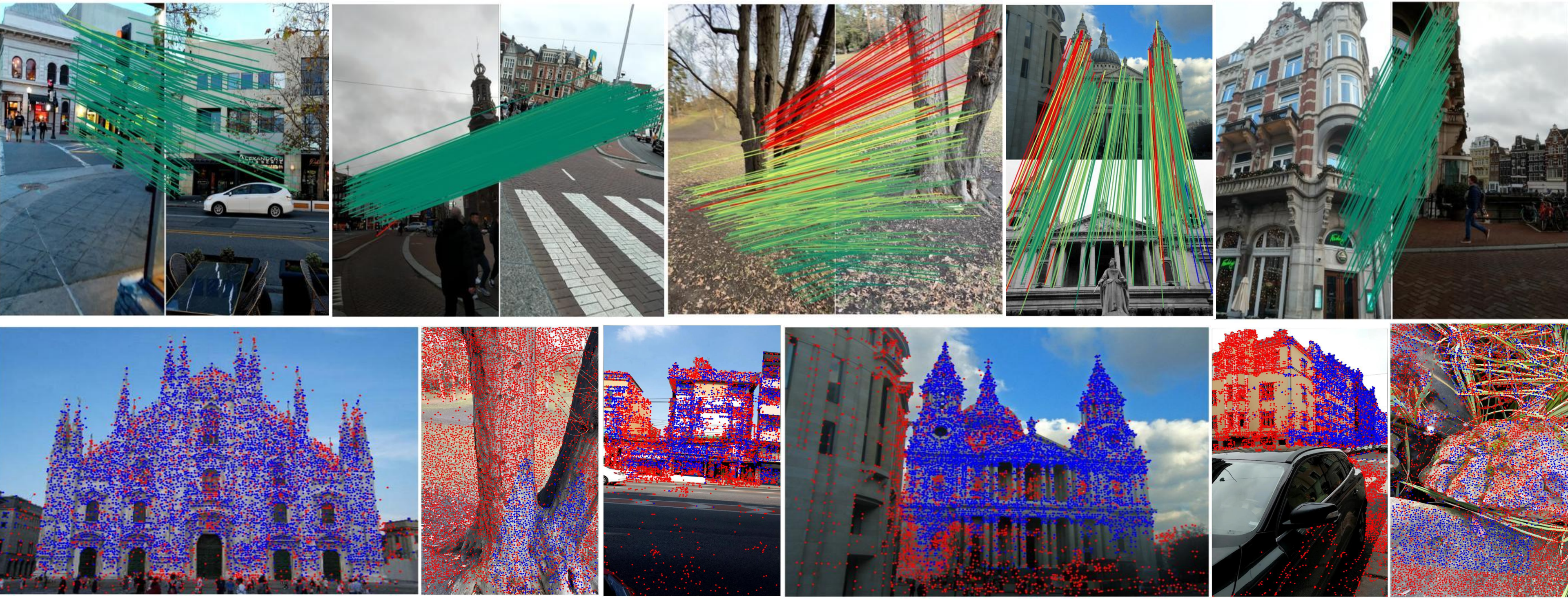
Workshop website: https://image-matching-workshop.github.io

9:00 - 9:15	**Welcome session (Eduard Trulls)**
9:15 - 10:00	**Invited Talk: Davide Scaramuzza (University of Zurich/ETH Zurich)**
10:00- 10:45	**Invited Talk: Marc Pollefeys (ETH Zurich/Microsoft)**
10:45 - 11:00	***Perceptual Loss for Robust Unsupervised Homography Estimation*** Daniel Koguciuk (Advanced Research Lab, NavInfo Europe, NL)
11:00 - 11:15	***DFM: A Performance Baseline for Deep Feature Matching*** Ufuk Efe (Middle East Technical University, Ankara, Turkey)
11:15 - 11:45	**Challenge presentation**
11:45 - 12:15	**Open discussion**
12:15 - 13:35	**Challenge participant talks** 12:15-12:25: Fabio Bellavia (University of Palermo) 12:25-12:35: Prune Truong (ETH Zurich) 12:35-12:45: Jiaming Sun/Xingyi He (Zhejiang University, SenseTime Research) 12:45-12:55: Wei Jiang (University of British Columbia) 12:55-13:05: Megvii 3D 13:05-13:15: Tencent
13:15 - 13:20	**Closing**

Outline

- The Image Matching Challenge
  - (Re-Re-)Introducing the Image Matching Benchmark
  - The PhotoTourism dataset (2019+)
  - The PragueParks dataset (2021)
  - The GoogleUrban dataset (2021)
  - The 2021 Image Matching Challenge results
- SimLocMatch
  - Motivation
  - Description
  - Roadmap for the future
  - The 2021 SimLocMatch Image-Matching Challenge Results

(Re-Re-)Introducing the Image Matching Benchmark

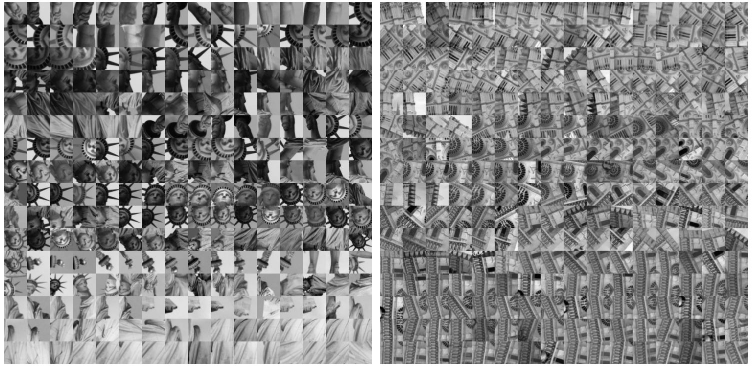


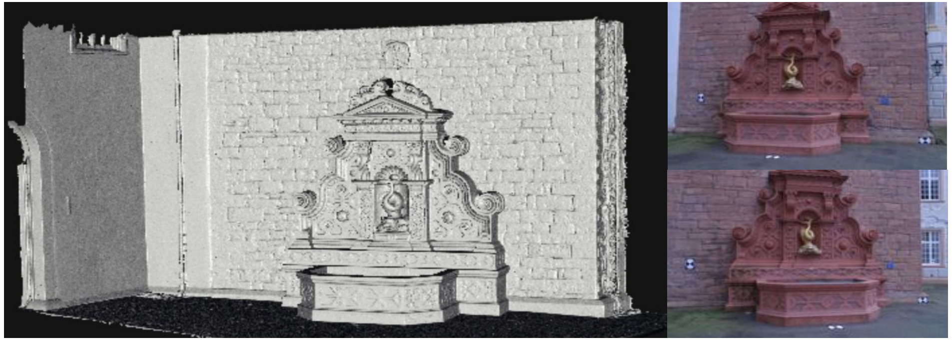
How good is

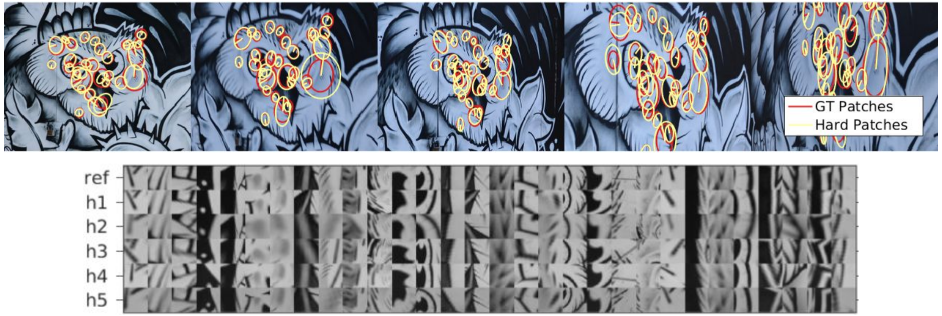
<insert-your-favorite-method-here>

in practice?

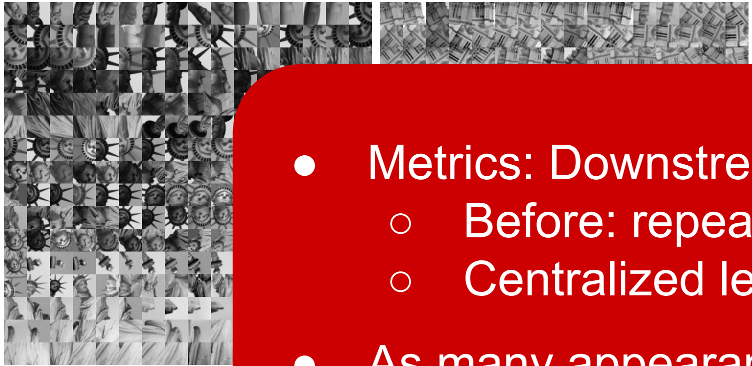
How can we do better?

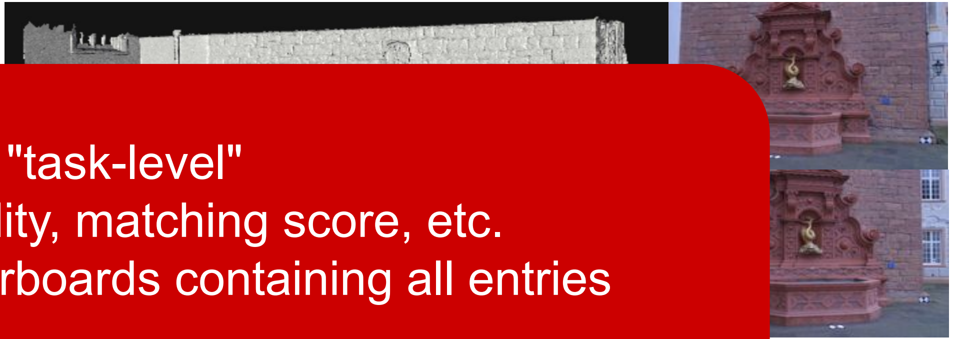






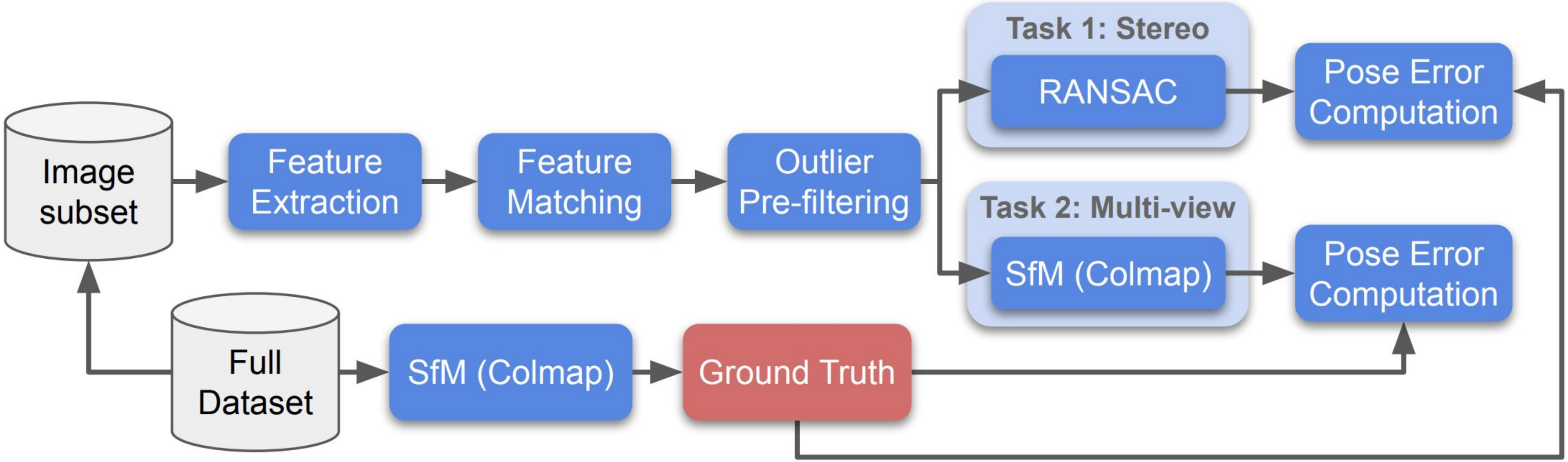
How can we do better?



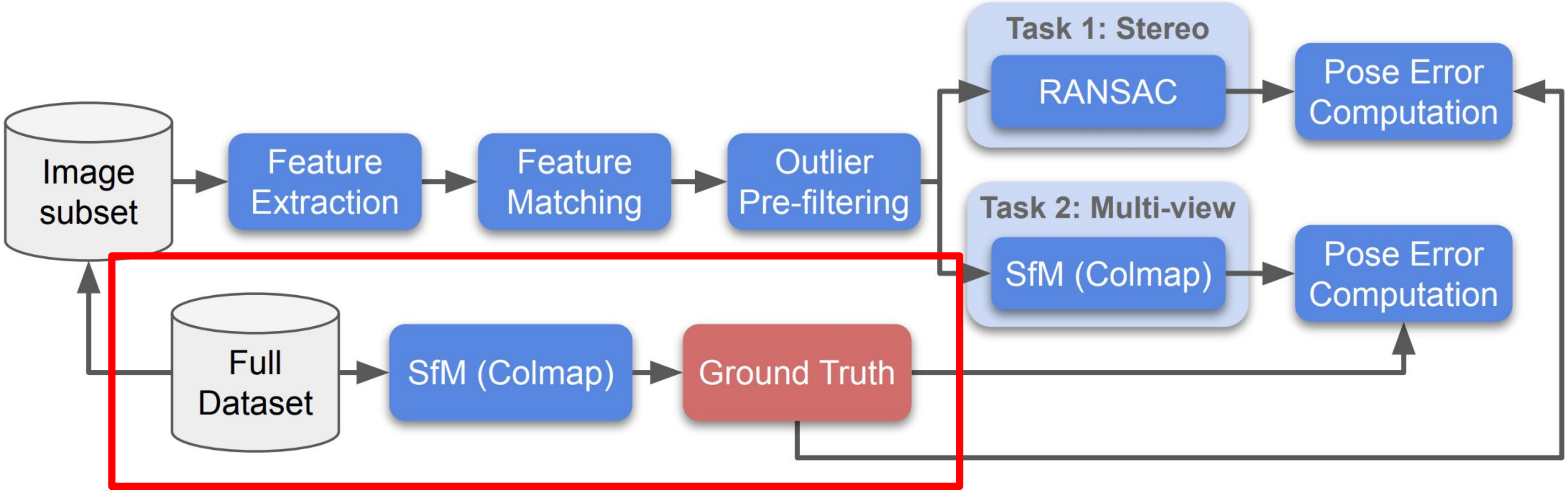


- Metrics: Downstream, "task-level"
  - Before: repeatability, matching score, etc.
  - Centralized leaderboards containing all entries
- As many appearance changes as possible
  - Viewpoint, illumination, cameras, etc.
- Scale: as large as possible
  - But we cannot benchmark large-scale SfM

The IMC pipeline



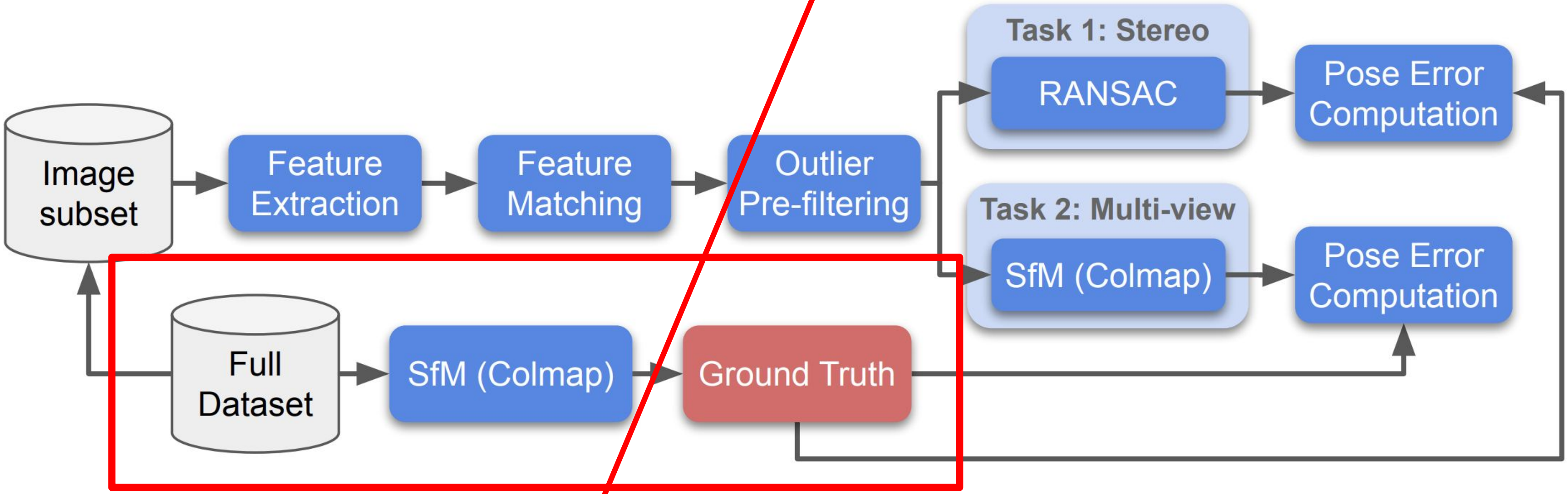
The IMC pipeline



Key insight #1: Ground Truth (pose) comes from off-the-shelf, large-scale SfM (100s~1000s of images). For evaluation we use much smaller and thus harder subsets (2~25 images).

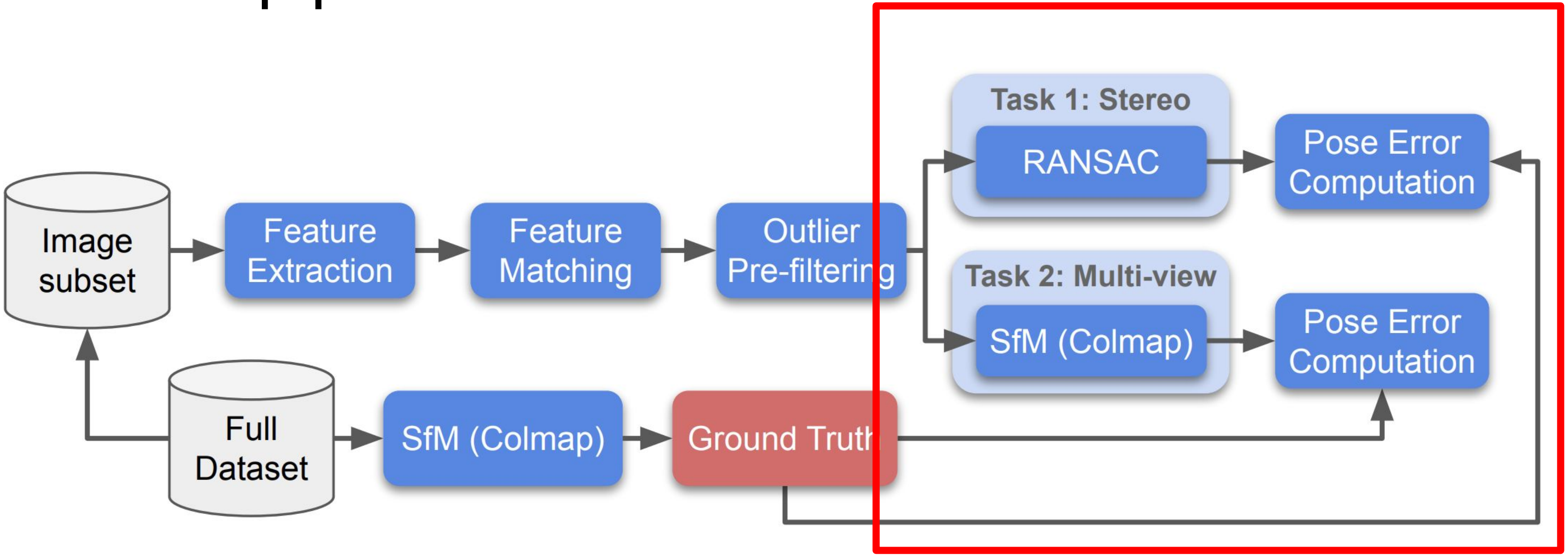
The IMC pipeline

This is different for the 2021 datasets, but the same principle holds



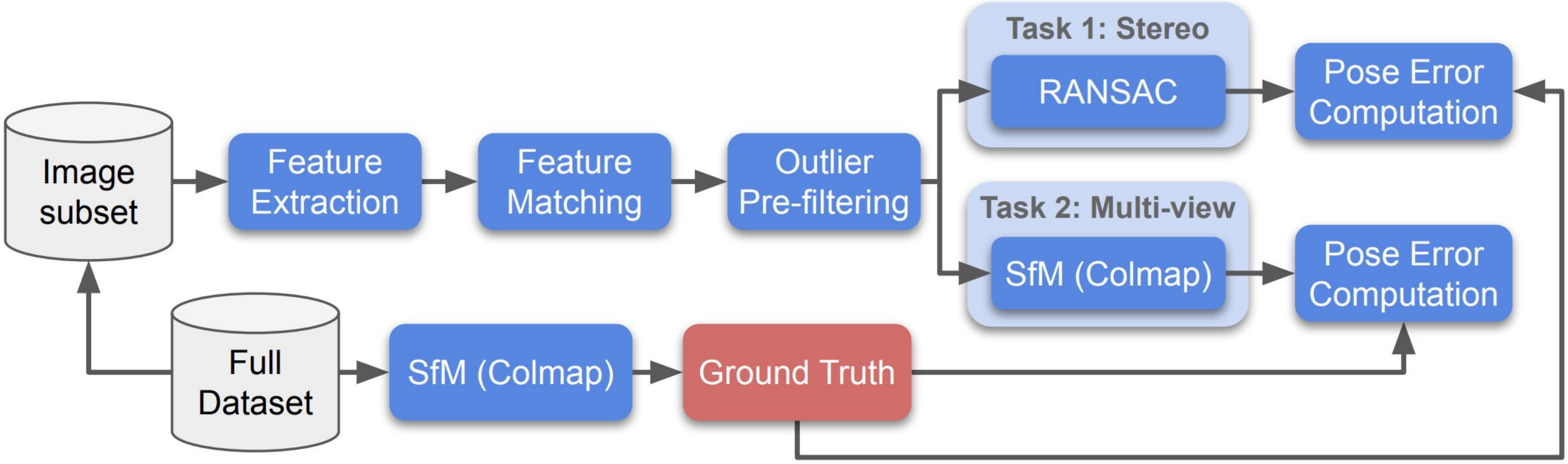
Key insight #1: Ground Truth (pose) comes from off-the-shelf, large-scale SfM (100s~1000s of images). For evaluation we use much smaller and thus harder subsets (2~25 images).

The IMC pipeline



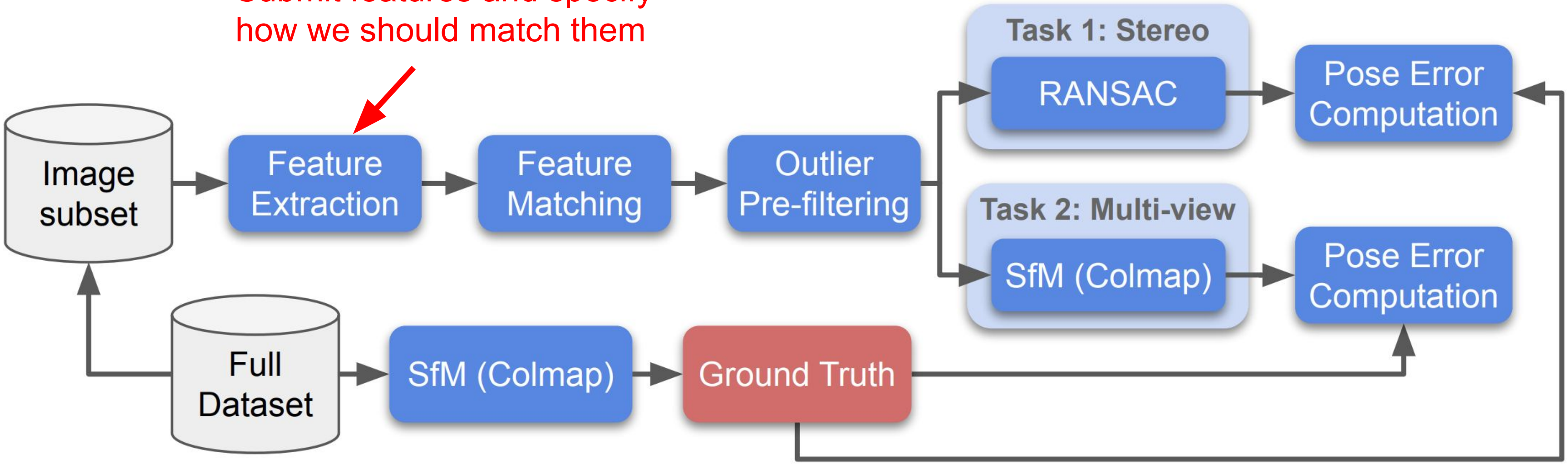
Key insight #2: Evaluation happens *downstream*.
Nothing is measured *by itself*.

Submitting your method

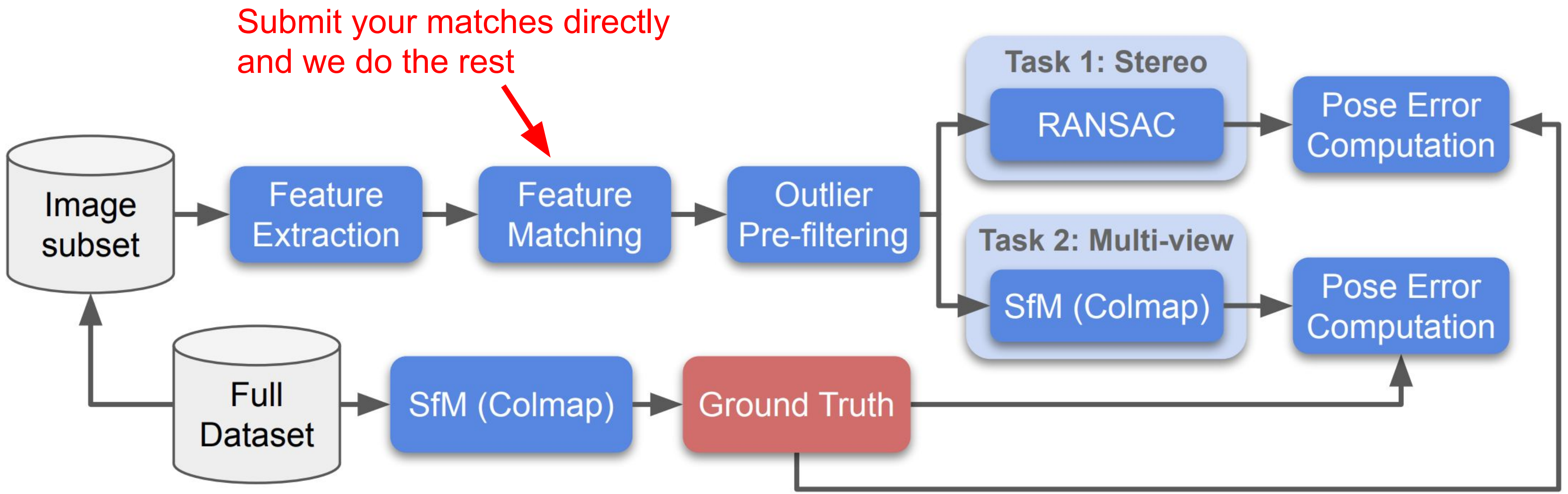


Submitting your method

Submit features and specify how we should match them



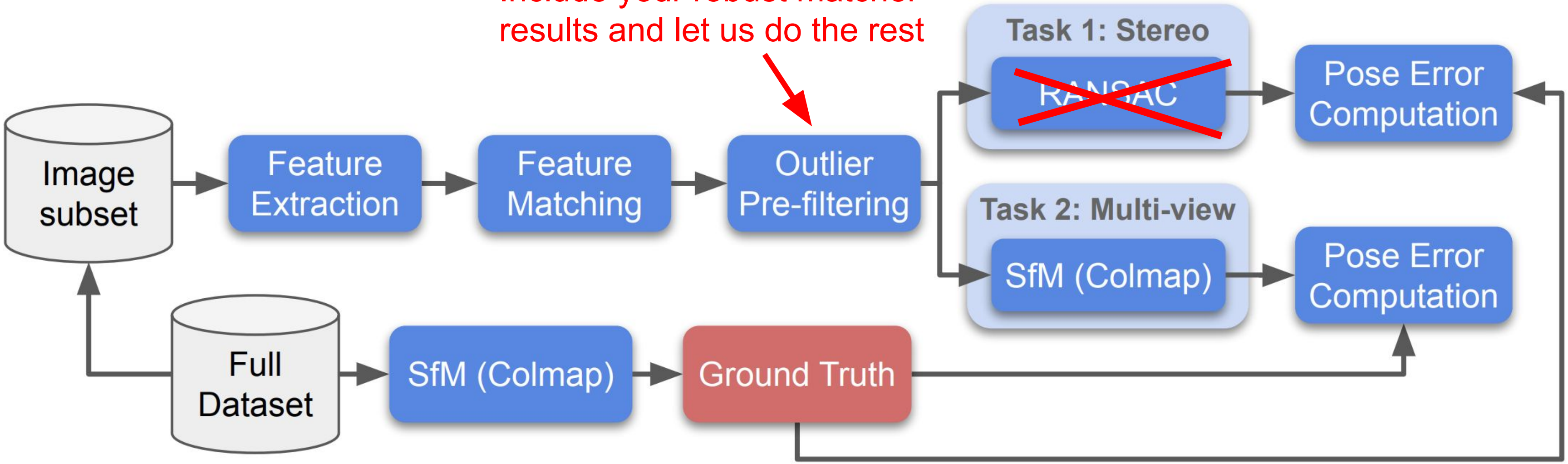
Submitting your method



Submitting your method

You can even
disable RANSAC

Include your robust matcher
results and let us do the rest



Metrics

Method	NF	Stereo NI	Stereo Rep. (3 pix.)	Stereo MS (3 pix.)	Stereo mAA (at 10°)	Multiview NM	Multiview NL	Multiview TL	Multiview ATE	Multiview mAA (at 10°)	Avg. mAA (at 10°)
Submission ID: 00024 SIFT (OpenCV), DEGENSAC Size: 512 bytes. Matches: built-in	7860.73	238.66 Rank: 116/127	0.472 Rank: 81/127	0.824 Rank: 85/127	0.45426 (±0.00097) Rank: 99/127	418.86 Rank: 78/127	3515.63 Rank: 87/127	4.001 Rank: 117/127	0.502 Rank: 109/127	0.60193 (±0.00185) Rank: 110/127	0.52810 Rank: 102/127
Submission ID: 00010 AKAZE (OpenCV), DEGENSAC Size: 61 bytes. Matches: built-in	7857.11	246.74 Rank: 115/127	0.553 Rank: 9/127	0.735 Rank: 111/127	0.30717 (±0.00122) Rank: 114/127	479.55 Rank: 65/127	2778.68 Rank: 114/127	3.393 Rank: 124/127	0.737 Rank: 126/127	0.36048 (±0.00382) Rank: 125/127	0.33383 Rank: 116/127

Mean Average Accuracy (mAA): average ratio of correct estimates under varying thresholds up to 10 degrees (considering max(R, T))

Metrics

		Stereo				Multiview					Avg.
Method	NF	NI	Rep. (3 pix.)	MS (3 pix.)	mAA (at 10°)	NM	NL	TL	ATE	mAA (at 10°)	mAA (at 10°)
Submission ID: **00024** SIFT (OpenCV), DEGENSAC Size: **512** bytes. Matches: **built-in**	7860.73	238.66 Rank: 116/127	0.472 Rank: 81/127	0.824 Rank: 85/127	0.45426 (±0.00097) Rank: 99/127	418.86 Rank: 78/127	3515.63 Rank: 87/127	4.001 Rank: 117/127	0.502 Rank: 109/127	0.60193 (±0.00185) Rank: 110/127	0.52810 Rank: 102/127
Submission ID: **00010** AKAZE (OpenCV), DEGENSAC Size: **61** bytes. Matches: **built-in**	7857.11	246.74 Rank: 115/127	0.553 Rank: 9/127	0.735 Rank: 111/127	0.30717 (±0.00122) Rank: 114/127	479.55 Rank: 65/127	2778.68 Rank: 114/127	3.393 Rank: 124/127	0.737 Rank: 126/127	0.36048 (±0.00382) Rank: 125/127	0.33383 Rank: 116/127

Matching score and repeatability thresholding at 3 pixels, using depth projection (if depth is available)

Metrics

		Stereo				Multiview					Avg.
Method	NF	NI	Rep. (3 pix.)	MS (3 pix.)	mAA (at 10°)	NM	NL	TL	ATE	mAA (at 10°)	mAA (at 10°)
Submission ID: 00024 SIFT (OpenCV), DEGENSAC Size: 512 bytes. Matches: built-in	7860.73	238.66 Rank: 116/127	0.472 Rank: 81/127	0.824 Rank: 85/127	0.45426 (±0.00097) Rank: 99/127	418.86 Rank: 78/127	3515.63 Rank: 87/127	4.001 Rank: 117/127	0.502 Rank: 109/127	0.60193 (±0.00185) Rank: 110/127	0.52810 Rank: 102/127
Submission ID: 00010 AKAZE (OpenCV), DEGENSAC Size: 61 bytes. Matches: built-in	7857.11	246.74 Rank: 115/127	0.553 Rank: 9/127	0.735 Rank: 111/127	0.30717 (±0.00122) Rank: 114/127	479.55 Rank: 65/127	2778.68 Rank: 114/127	3.393 Rank: 124/127	0.737 Rank: 126/127	0.36048 (±0.00382) Rank: 125/127	0.33383 Rank: 116/127

Other standard metrics for Multi-view

Metrics

		Stereo				Multiview					Avg.
Method	NF	NI	Rep. (3 pix.)	MS (3 pix.)	mAA (at 10^{o})	NM	NL	TL	ATE	mAA (at 10^{0})	mAA (at 10^{0})
Submission ID: 00024 SIFT (OpenCV), DEGENSAC Size: 512 bytes. Matches: built-in	7860.73	238.66 Rank: 116/127	0.472 Rank: 81/127	0.824 Rank: 85/127	0.45426 (±0.00097) Rank: 99/127	418.86 Rank: 78/127	3515.63 Rank: 87/127	4.001 Rank: 117/127	0.502 Rank: 109/127	0.60193 (±0.00185) Rank: 110/127	0.52810 Rank: 102/127
Submission ID: 00010 AKAZE (OpenCV), DEGENSAC Size: 61 bytes. Matches: built-in	7857.11	246.74 Rank: 115/127	0.553 Rank: 9/127	0.735 Rank: 111/127	0.30717 (±0.00122) Rank: 114/127	479.55 Rank: 65/127	2778.68 Rank: 114/127	3.393 Rank: 124/127	0.737 Rank: 126/127	0.36048 (±0.00382) Rank: 125/127	0.33383 Rank: 116/127

We also use mAA for multiview using all pairs of images in each reconstruction.

Metrics

Method	NF	Stereo: NI	Stereo: Rep. (3 pix.)	Stereo: MS (3 pix.)	Stereo: mAA (at 10°)	Multiview: NM	Multiview: NL	Multiview: TL	Multiview: ATE	Multiview: mAA (at 10°)	Avg.: mAA (at 10°)
Submission ID: 00024 SIFT (OpenCV), DEGENSAC Size: 512 bytes. Matches: built-in	7860.73	238.66 Rank: 116/127	0.472 Rank: 81/127	0.824 Rank: 85/127	0.45426 (±0.00097) Rank: 99/127	418.86 Rank: 78/127	3515.63 Rank: 87/127	4.001 Rank: 117/127	0.502 Rank: 109/127	0.60193 (±0.00185) Rank: 110/127	0.52810 Rank: 102/127
Submission ID: 00010 AKAZE (OpenCV), DEGENSAC Size: 61 bytes. Matches: built-in	7857.11	246.74 Rank: 115/127	0.553 Rank: 9/127	0.735 Rank: 111/127	0.30717 (±0.00122) Rank: 114/127	479.55 Rank: 65/127	2778.68 Rank: 114/127	3.393 Rank: 124/127	0.737 Rank: 126/127	0.36048 (±0.00382) Rank: 125/127	0.33383 Rank: 116/127

One number to rule them all…
And in the darkness evaluate them

How to use it (for validation)

- Python codebase with simple requirements
  - Benchmark repository: https://github.com/ubc-vision/image-matching-benchmark
- Input: Local features are directly embedded (OpenCV) or imported (the rest)
  - Baselines repository: https://github.com/ubc-vision/image-matching-benchmark-baselines
    - No changes since last year, though!
  - Robust matchers are embedded with python (PyRANSAC) or OpenCV
  - SOTA RANSACs now in OpenCV 4.5! https://opencv.org/evaluating-opencvs-new-ransacs
- Parallelized via a job scheduler: SLURM (Compute Canada)
  - Can be run single-threaded for validation
  - Still pretty heavy! Every dataset runs stereo ~1000x, and SfM ~100x.

How to use it (for validation)

1: Configure it (and import features/matches)

```
{
    "config": {
        "config_common": {
            "descriptor": "hardnet64-train-all-l2-val-14000",
            "keypoint": "sift8k",
            "num_keypoints": 8000,
            "json_label": "sid-00611-sift8k_8000_hardnet64-train-all-l2-val-14000"
        },
        "metadata": {
            "publish_anonymously": true,
            "contact_email": "stliwenbin@gmail.com",
            "authors": "Ximin Zheng, Sheng He, Hualong Shi",
            "link_to_website": "",
            "method_name": "[sid:00611] sift and hardnet64 train scale(12)",
            "link_to_pdf": "",
            "method_description": "SIFT with 8000 keypoints(scale 12), hardnet64 with
128 descriptors(trained with l2 loss and step 14000), FLANN disabled"
        },
        "config_phototourism_stereo": {
            "use_custom_matches": false,
            "matcher": {
                "num_nn": 1,
                "symmetric": {
                    "reduce": "both",
                    "enabled": true
                },
                "filtering": {
                    "type": "snn_ratio_pairwise",
                    "threshold": 0.9
                },
                "distance": "L2",
                "method": "nn",
                "flann": false
            },
            "geom": {
                "degeneracy_check": true,
                "max_iter": 100000,
```

Step 2: Run it… and wait

```
python run.py --json_method=<config_file>.json
```

Step 3: Profit!

How to submit to the challenge

Home Leaderboard Data Benchmark Submit News SimLocMatch Link to IMC2020 Website

Challenge submissions

The submission website is password-protected to prevent abuse. Please contact the organizers at image-matching@googlegroups.com for the password (please account for short delays in answering and uploading close the deadline). Please upload the results as a zip or tarball containing the JSON file and your features/matches, if applicable. You can also check the status of your submission via the status tracking spreadsheet.

Please always run our validation script to ensure your submission is in proper format. We also have a general tutorial on how to use our benchmark and create submission file and a tutorial specific for custom matcher, please have a look if you have trouble on creating submissions.

- Submission link
- Submission status
- Submission spec LaTeX kit

Challenge categories

Submissions are broken down into two categories by **number of keypoints:** we consider a "restricted" budget of 2048 features, and an "unlimited" budget (capped to 8000 features per image for practical reasons). In previous editions we also broke down submissions by **descriptor size**, ~~but nearly all participants opted for 128-dimensional floating-point descriptors (float32), which is the maximum size allowed this year.~~ **May 25, 2021: We have removed this rule. You may use descriptors of any size. If you use descriptors larger than 128D, we ask that you submit custom matches instead of using built-in matchers: you may use the benchmark to obtain them, but they need to be in the submission — this is required in order to keep our compute budget in order. You are still required to submit descriptor files. If your method does not use descriptors at all, you may leave these files empty. If in doubt, please reach out to us.**

Submission format

Submissions should come in the form zip files containing keypoints, descriptors, for every dataset and scene, and a single JSON file with metadata and settings. Matches can be provided, or generated by the benchmark. If provided, we require separate files for stereo and multiview (the optimal settings typically vary across tasks — even if they are not, you must provide two files). The datasets are labeled by the benchmark as `"phototourism"`, `"pragueparks"`, and `"googleurban"`. For example:

```
$ ls my_submission
config.json  googleurban  phototourism  pragueparks

$ ls my_submission/pragueparks
lizard  pond  tree_new

$ ls my_submission/pragueparks/lizard
descriptors.h5  keypoints.h5  matches_stereo.h5  matches_multiview.h5
```

Please note that we do not allow combining different methods for local feature extraction and matching in a single submission. For instance, you may not use HardNet descriptors on the PhotoTourism dataset and SuperPoint on the PragueParks dataset, or RANSAC on one dataset and SuperGlue on another,

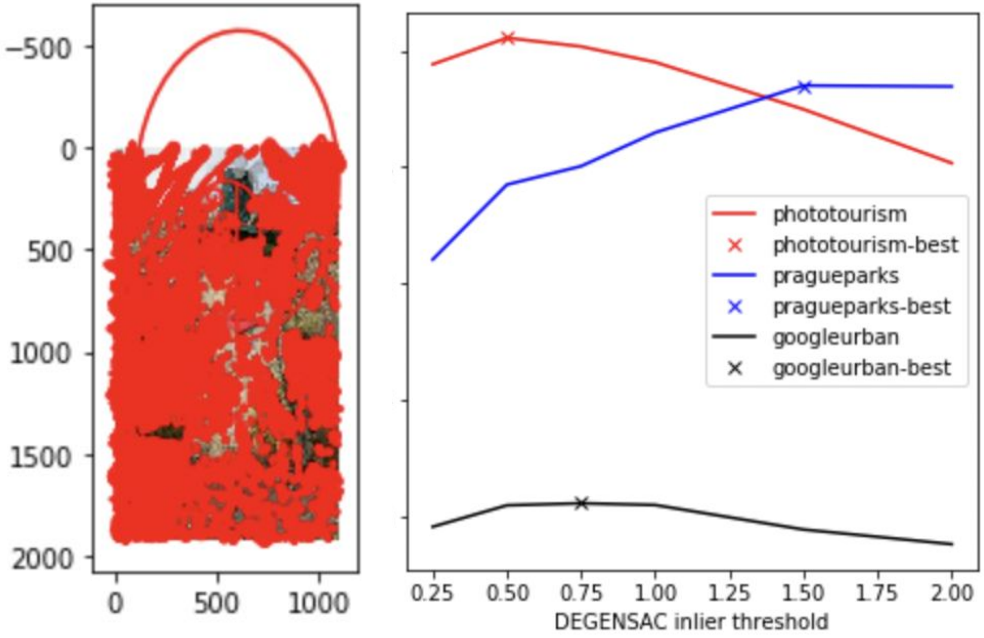
- Upload server is password-protected
  - Contact us for the password
- You must provide:
  - A configuration file
  - Features and, optionally, matches
- Validate your submissions
  - https://github.com/ubc-vision/image-matching-benchmark/blob/master/submission_validator.py
- Submission rules
  - https://www.cs.ubc.ca/research/image-matching-challenge/2021/submit/
- Tutorial
  - https://ducha-aiki.github.io/wide-baseline-stereo-blog/2021/05/27/submitting-to-IMC2021-with-custom-matcher.html

How to submit to the challenge: tutorial

1. Extract features/matches
2. Create a config.json file
3. Tune matching/RANSAC based on the validation set
4. Check the submission with the validator-script
5. Upload 2-10 Gb to the website

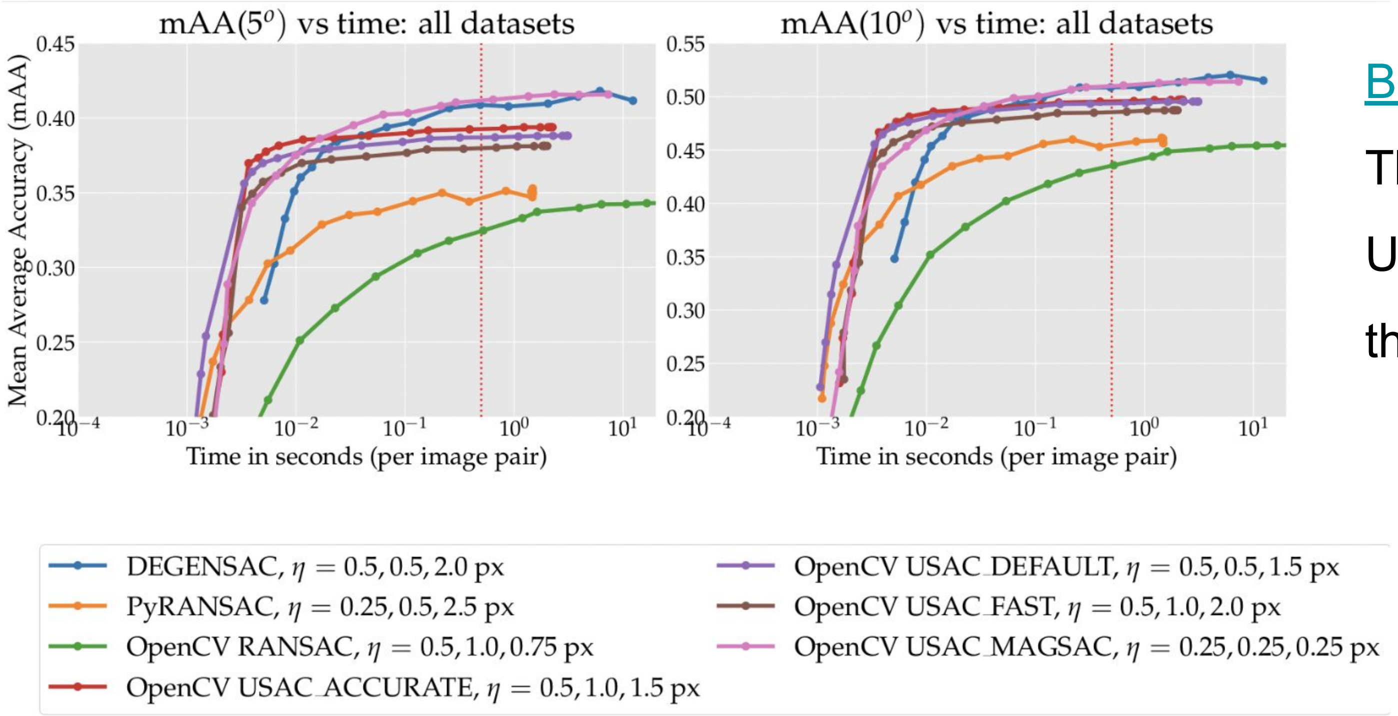
compute canada

https://ducha-aiki.github.io/wide-baseline-stereo-blog/2021/05/12/submitting-to-IMC2021-step-by-step.html



Checkout new OpenCV RANSACs, they are great!

- They are added to the benchmark (use them for the future submissions)

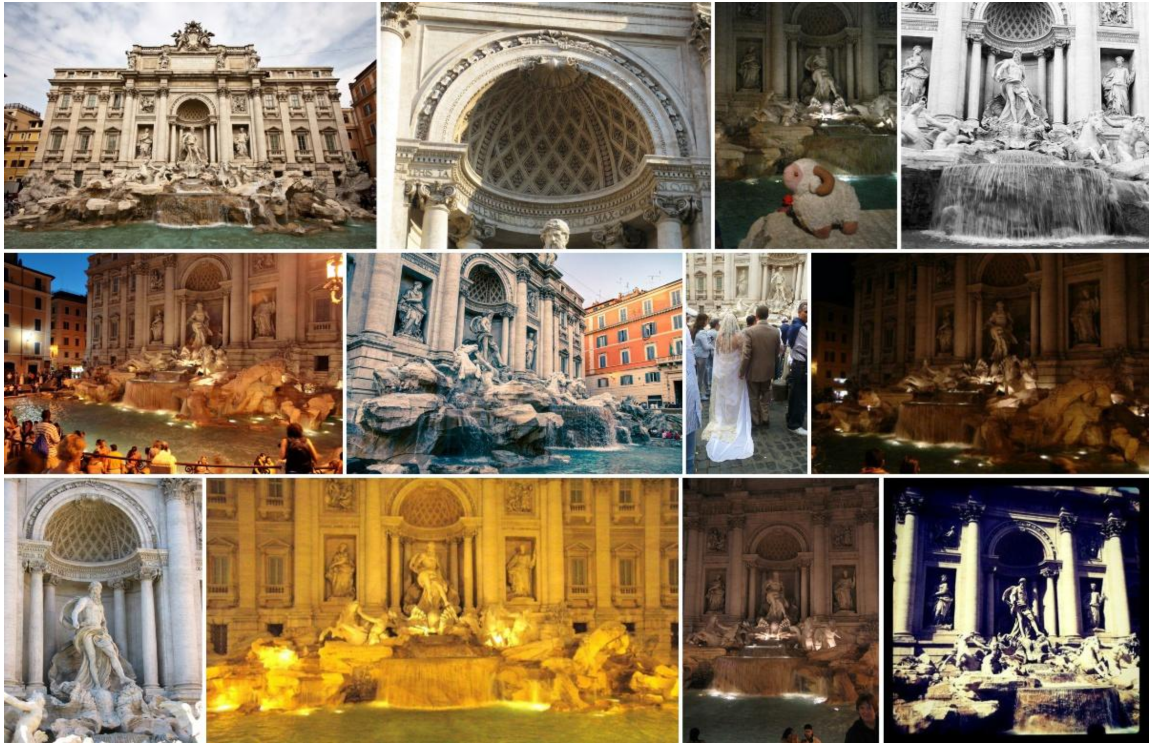


[Benchmark is here]

Tl;dr: use USAC_MAGSAC with th=0.25 for all datasets

The PhotoTourism Dataset

PT dataset: Ground Truth from large-scale SfM



PT Dataset: Training data

Training scene	Num. images	Num. 3D points
Brandenburg Gate	1363	100040
Buckingham Palace	1676	234052
Colosseum Exterior	2063	259807
Grand Place Brussels	1083	229788
Hagia Sophia Interior	888	235541
Notre Dame Front Facade	3765	488895
Palace of Westminster	983	115868
Pantheon Exterior	1401	166923
Prague Old Town Square	2316	558600
Reichstag	75	17823
Sacre Coeur	1179	140659
Saint Peter's Square	2504	232329
Taj Mahal	1312	94121
Temple Nara Japan	904	92131
Trevi Fountain	3191	580673
Westminster Abbey	1061	198222
Total	25.6k	3.7M

- We provide 25k registered images for training
- However, you can use anything else! (As long as it does not overlap)

PT Dataset: Test data

Test scenes	Num. images	Num. 3D points
British Museum	660	73569
Florence Cathedral Side	108	44143
Lincoln Memorial Statue	850	58661
London Bridge	629	72235
Milan Cathedral	124	33905
Mount Rushmore	138	45350
Piazza San Marco	249	95895
Sagrada Familia	401	120723
Saint Paul's Cathedral	615	98872
Total	4107	696k

- 9 different scenes
- Over 4k images in total, from which we subsample 100-image subsets, which are given to participants
- Valid pairs are determined with a simple visibility check
- For SfM, random bags of images are subsampled to form test subsets (5, 10, or 25 images at a time)

PT dataset: Ground truth from large-scale SfM

PT dataset: *Can you call this "ground truth"?*

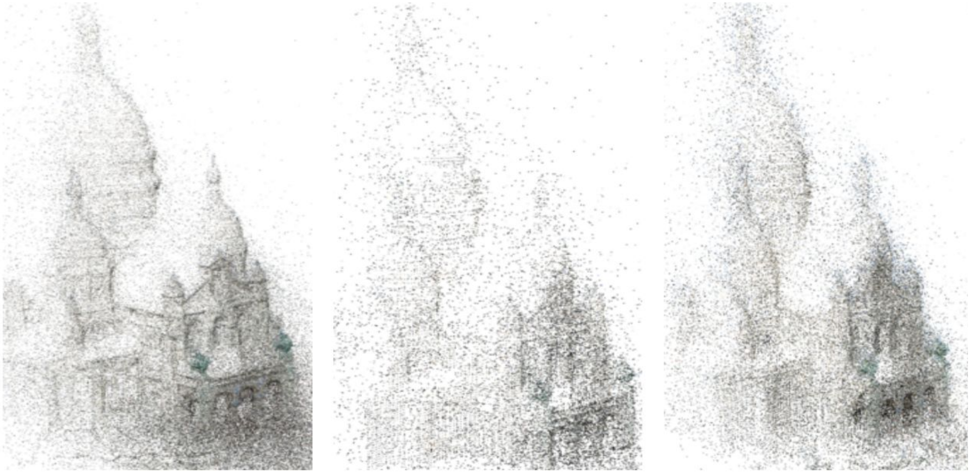
Feature used	Number of images			
	100 vs. all	200 vs. all	400 vs. all	800 vs. all
SIFT	0.46° / 0.13°	0.42° / 0.11°	0.32° / 0.08°	0.39° / 0.08°
SuperPoint	2.09° / 1.57°	2.09° / 1.54°	1.87° / 1.21°	2.53° / 0.53°
R2D2	0.41° / 0.14°	0.29° / 0.09°	0.28° / 0.09°	0.21° / 0.06°

(Mean / median)

- We reconstruct a scene (Sacre Coeur) while adding images to it
- Pose converges as more images are used for reconstruction (but are quite stable at 100-200 already)
- Small pose differences by swapping the features
- Further "sanity checks" by the organizers: misregistered images have been removed

SIFT: Distinctive Image Features from Scale-Invariant Keypoints. David G. Lowe. IJCV, 20(2):91–110, November 2004.
SuperPoint: Self-Supervised Interest Point Detection and Description. DeTone et al., CVPR'18
R2D2: Reliable and Repeatable Detector and Descriptor. J Revaud et al., NeurIPS'19

PT dataset: *Are you biased towards SIFT/COLMAP?*



(a) SIFT (b) SuperPoint (c) R2D2

- It doesn't matter. The reconstructions may look quite different, we only need *good poses*
- Are they good? We compare the reconstructions with SIFT vs two other methods and observe that the poses are similar across different methods

Reference	Compared	
	SuperPoint	R2D2
SIFT	2.06° / 1.57°	0.42° / 0.14°

- Better features/matchers might register more images, but this is not our focus (yet)

Curious? More results in the IJCV paper

https://arxiv.org/abs/2003.01587

Image Matching Across Wide Baselines: From Paper to Practice

Yuhe Jin · Dmytro Mishkin · Anastasiia Mishchuk · Jiri Matas · Pascal Fua · Kwang Moo Yi · Eduard Trulls

arXiv:2003.01587v2 [cs.CV] 8 May 2020

Received: date / Accepted: date

Abstract We introduce a comprehensive benchmark for local features and robust estimation algorithms, focusing on the downstream task – the accuracy of the reconstructed camera pose – as our primary metric. Our pipeline's modular structure allows us to easily integrate, configure, and combine different methods and heuristics. We demonstrate this by embedding dozens of popular algorithms and evaluating them, from seminal works to the cutting edge of machine learning research. We show that with proper settings, classical solutions may still outperform the *perceived* state of the art.

Besides establishing the *actual* state of the art, the experiments conducted in this paper reveal unexpected properties of Structure from Motion (SfM) pipelines that can

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant "Deep Visual Geometry Machines" (RGPIN-2018-03788), by systems supplied by Compute Canada, and by Google's Visual Positioning Service. DM and JM were supported by OP VVV funded project CZ.02.1.01/0.0/0.0/16 019/0000765 "Research Center for Informatics". DM was also supported by CTU student grant SGS17/185/OHK3/3T/13 and by the Austrian Ministry for Transport, Innovation and Technology, the Federal Ministry for Digital and Economic Affairs, and the Province of Upper Austria in the frame of the COMET center SCCH. AM was supported by the Swiss National Science Foundation.

Y. Jin, K.M. Yi

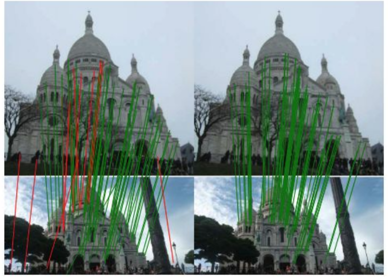


Fig. 1 Every paper claims to outperform the state of the art. Is this possible, or an artifact of insufficient validation? On the left, we show stereo matches obtained with **D2-Net** (2019) [38], a state-of-the-art local feature, using OpenCV RANSAC with its default settings. We color the inliers in green if they are correct and in red otherwise. On the right, we show **SIFT** (1999) [55] with a carefully tuned MAGSAC [32] – notice how the latter performs much better. This illustrates our take-home message: to correctly evaluate a method's performance, it needs to be embedded within the pipeline used to solve a given problem, and the different components in said pipeline need to be tuned carefully and jointly, which requires engineering and domain expertise. We fill this need with a new, modular benchmark for sparse image matching, incorporating dozens of built-in methods.

be exploited to help improve their performance, for both

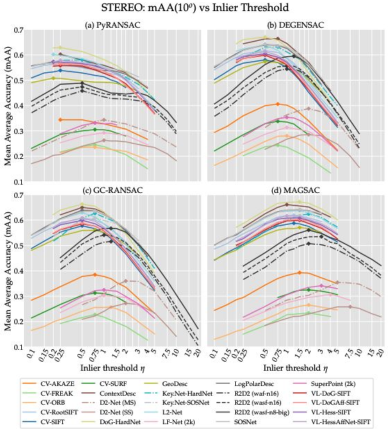


Fig. 10 Validation – Inlier threshold for RANSAC, η. We determine η for each combination, using 8k features (2k for LF-Net and SuperPoint) with the "both" matching strategy and a reasonable value for the ratio test. Optimal parameters (diamonds) are listed in the Section 7.

PyRANSAC. MAGSAC gives the best results for this experiment, closely followed by DEGENSAC. We patch OpenCV to increase the limit of iterations, which was hardcoded to $\Gamma = 1000$; this patch is now integrated into OpenCV. This increases performance by 10-15% relative, within our budget. However, PyRANSAC is significantly better than OpenCV version even with this patch, so we use it as our "vanilla" RANSAC instead. The sklearn implementation is too slow for practical use.

We find that, in general, default settings can be woefully inadequate. For example, OpenCV recommends $\tau = 0.99$

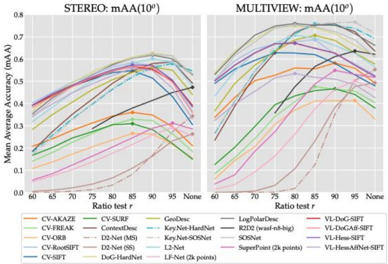


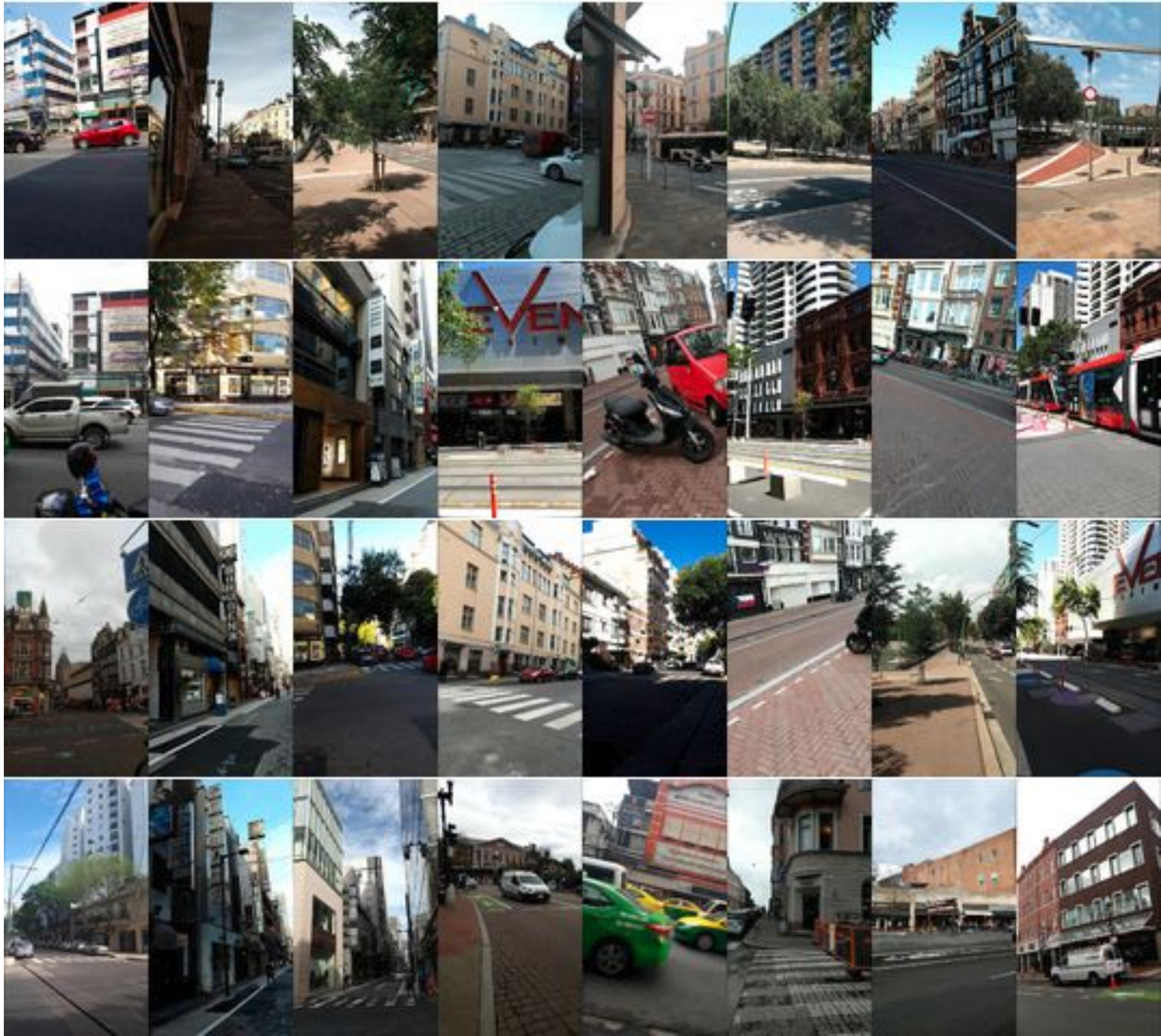
Fig. 11 Validation – Optimal ratio test r for matching with "both". We evaluate bidirectional matching with the "both" strategy (the best one), and different ratio test thresholds r, for each feature type. We use 8k features (2k for SuperPoint and LF-Net). For stereo, we use PyRANSAC.

test with the threshold recommended by the authors of each feature, or a reasonable value if no recommendation exists, and the "both" matching strategy – this cuts down on the number of outliers.

5.3 Ratio test: One feature at a time

Having "frozen" RANSAC, we turn to the feature matcher – note that it comes *before* RANSAC, but it cannot be evaluated in isolation. We select PyRANSAC as a "baseline" RANSAC and evaluate different ratio test thresholds, separately for the stereo and multiview tasks. For this experiment, we use 8k features with all methods, except for those which cannot work on this regime – SuperPoint and LF-Net. This choice will be substantiated in Section 5.4. We report the results for bidirectional matching with the "both" strat-

(New) The GoogleUrban Dataset



The GoogleUrban dataset

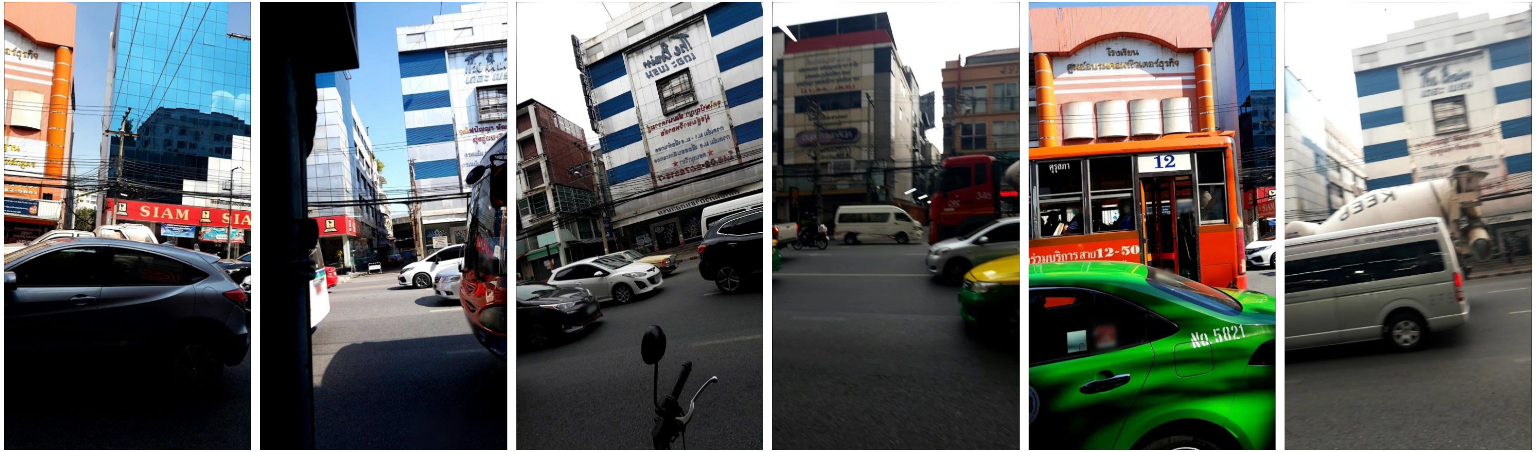
- ~1500 images from video sequences captured with a phone
- Images posed with internal systems at Google
  - No SfM, unlike PhotoTourism/PragueParks
  - Focus: close-up façades, no "touristic" landmarks
- Blurred faces and license plates automatically, followed by manual inspection
- Released with a restricted license: please delete by tomorrow!
  - We plan to use similar images in future editions

Tokyo

Data from 20 cities across 4 continents (validation/test)

Mountain View

Bangkok



More difficult than previous datasets

PhotoTourism

Method	NF	Stereo NI	Stereo Rep. (3 pix.)	Stereo MS (3 pix.)	Stereo mAA (at 10°)	Multiview NM	Multiview NL	Multiview TL	Multiview ATE	Multiview mAA (at 10°)	Avg. mAA (at 10°)
Submission ID: e9043677 sp_disk_scale_8k Size: 0 bytes. Matches: custom	7941.60	1707.18 Rank: 2/35	0.580 Rank: 13/35	0.842 Rank: 10/35	0.63975 (±0.00000) Rank: 1/35	1739.70 Rank: 2/35	8924.36 Rank: 1/35	5.365 Rank: 7/35	0.365 Rank: 5/35	0.78564 (±0.00000) Rank: 1/35	0.71269 Rank: 1/35

PragueParks

Method	NF	Stereo NI	Stereo Rep. (3 pix.)	Stereo MS (3 pix.)	Stereo mAA (at 10°)	Multiview NM	Multiview NL	Multiview TL	Multiview ATE	Multiview mAA (at 10°)	Avg. mAA (at 10°)
Submission ID: e9043677 sp_disk_scale_8k Size: 0 bytes. Matches: custom	7437.97	1214.98 Rank: 1/35	0.096 Rank: 29/35	0.027 Rank: 20/35	0.80700 (±0.00000) Rank: 1/35	1383.42 Rank: 4/35	3658.74 Rank: 2/35	3.217 Rank: 5/35	6.962 Rank: 30/35	0.49878 (±0.00000) Rank: 3/35	0.65289 Rank: 1/35

GoogleUrban

Method	NF	Stereo NI	Stereo Rep. (3 pix.)	Stereo MS (3 pix.)	Stereo mAA (at 10°)	Multiview NM	Multiview NL	Multiview TL	Multiview ATE	Multiview mAA (at 10°)	Avg. mAA (at 10°)
Submission ID: e9043677 sp_disk_scale_8k Size: 0 bytes. Matches: custom	7890.64	746.07 Rank: 6/35	N/A Rank: —/35	N/A Rank: —/35	0.43952 (±0.00000) Rank: 1/35	477.34 Rank: 6/35	3842.69 Rank: 6/35	3.620 Rank: 4/35	20.984 Rank: 1/35	0.33734 (±0.00000) Rank: 1/35	0.38843 Rank: 1/35

More difficult than previous datasets

PhotoTourism

Method	NF	Stereo NI	Stereo Rep. (3 pix.)	Stereo MS (3 pix.)	Stereo mAA (at 10°)	Multiview NM	Multiview NL	Multiview TL	Multiview ATE	Multiview mAA (at 10°)	Avg. mAA (at 10°)
Submission ID: **e9043677** sp_disk_scale_8k Size: **0** bytes. Matches: **custom**	7941.60	1707.18 Rank: 2/35	0.580 Rank: 13/35	0.842 Rank: 10/35	0.63975 (±0.00000) Rank: 1/35	1739.70 Rank: 2/35	8924.36 Rank: 1/35	5.365 Rank: 7/35	0.365 Rank: 5/35	0.78564 (±0.00000) Rank: 1/35	0.71269 Rank: 1/35

PragueParks

Method	NF	Stereo NI	Stereo Rep. (3 pix.)	Stereo MS (3 pix.)	Stereo mAA (at 10°)	Multiview NM	Multiview NL	Multiview TL	Multiview ATE	Multiview mAA (at 10°)	Avg. mAA (at 10°)
Submission ID: **e9043677** sp_disk_scale_8k Size: **0** bytes. Matches: **custom**	7437.97	1214.98 Rank: 1/35	0.096 Rank: 29/35	0.027 Rank: 20/35	0.80700 (±0.00000) Rank: 1/35	1383.42 Rank: 4/35	3658.74 Rank: 2/35	3.217 Rank: 5/35	6.962 Rank: 30/35	0.49878 (±0.00000) Rank: 3/35	0.65289 Rank: 1/35

GoogleUrban

Method	NF	Stereo NI	Stereo Rep. (3 pix.)	Stereo MS (3 pix.)	Stereo mAA (at 10°)	Multiview NM	Multiview NL	Multiview TL	Multiview ATE	Multiview mAA (at 10°)	Avg. mAA (at 10°)
Submission ID: **e9043677** sp_disk_scale_8k Size: **0** bytes. Matches: **custom**	7890.64	746.07 Rank: 6/35	N/A Rank: —/35	N/A Rank: —/35	0.43952 (±0.00000) Rank: 1/35	477.34 Rank: 6/35	3842.69 Rank: 6/35	3.620 Rank: 4/35	20.984 Rank: 1/35	0.33734 (±0.00000) Rank: 1/35	0.38843 Rank: 1/35

More difficult than previous datasets

PhotoTourism

Method	NF	NI	Rep. (3 pix.)	MS (3 pix.)	Stereo mAA (at 10°)	NM	NL	TL	ATE	Multiview mAA (at 10°)	Avg. mAA (at 10°)
Submission ID: **e9043677** sp_disk_scale_8k Size: **0** bytes. Matches: **custom**	7941.60	1707.18 Rank: 2/35	0.580 Rank: 13/35	0.842 Rank: 10/35	0.63975 (±0.00000) Rank: 1/35	1739.70 Rank: 2/35	8924.36 Rank: 1/35	5.365 Rank: 7/35	0.365 Rank: 5/35	0.78564 (±0.00000) Rank: 1/35	0.71269 Rank: 1/35

PragueParks

Method	NF	NI	Rep. (3 pix.)	MS (3 pix.)	Stereo mAA (at 10°)	NM	NL	TL	ATE	Multiview mAA (at 10°)	Avg. mAA (at 10°)
Submission ID: **e9043677** sp_disk_scale_8k Size: **0** bytes. Matches: **custom**	7437.97	1214.98 Rank: 1/35	0.096 Rank: 29/35	0.027 Rank: 20/35	0.80700 (±0.00000) Rank: 1/35	1383.42 Rank: 4/35	3658.74 Rank: 2/35	3.217 Rank: 5/35	6.962 Rank: 30/35	0.49878 (±0.00000) Rank: 3/35	0.65289 Rank: 1/35

GoogleUrban

Method	NF	NI	Rep. (3 pix.)	MS (3 pix.)	Stereo mAA (at 10°)	NM	NL	TL	ATE	Multiview mAA (at 10°)	Avg. mAA (at 10°)
Submission ID: **e9043677** sp_disk_scale_8k Size: **0** bytes. Matches: **custom**	7890.64	746.07 Rank: 6/35	N/A Rank: —/35	N/A Rank: —/35	0.43952 (±0.00000) Rank: 1/35	477.34 Rank: 6/35	3842.69 Rank: 6/35	3.620 Rank: 4/35	20.984 Rank: 1/35	0.33734 (±0.00000) Rank: 1/35	0.38843 Rank: 1/35

Multiview harder than stereo

(New) The PragueParks Dataset

PragueParks GT generation

- Data: captured with iPhone 11 video (stabilized) → images (24 fps)
- Ground truth: reconstructed by RealityCapture: commercial 3d reconstruction software
  - Benefit over COLMAP: 100x faster, reconstruction in hours instead of weeks
- Test data: sample less frequently: 24fps -> ~1 fps. The scripts for the data creation are open-sourced:
  - https://github.com/ducha-aiki/creating-data-for-imc
- Plans for next year? more aggressive sampling, also day-vs-night matching

Analyzing the IMC'21 results

		Stereo				Multiview					Avg.
Method	NF	NI	Rep. (3 pix.)	MS (3 pix.)	mAA (at 10°)	NM	NL	TL	ATE	mAA (at 10°)	mAA (at 10°)
Submission ID: **00591** Guided-Hardnet-Epoch2 Size: **512** bytes. Matches: **custom**	7829.63	761.99 Rank: 13/127	0.486 Rank: 44/127	0.823 Rank: 88/127	0.61081 (±0.00000) Rank: 1/127	785.57 Rank: 23/127	6330.70 Rank: 7/127	4.680 Rank: 7/127	0.358 Rank: 2/127	0.78288 (±0.00094) Rank: 2/127	0.69684 Rank: 1/127
Submission ID: **00624** Guided-HardNet-OANet Size: **512** bytes. Matches: **custom**	7829.63	765.34 Rank: 12/127	0.486 Rank: 44/127	0.820 Rank: 90/127	0.60261 (±0.00000) Rank: 2/127	788.49 Rank: 21/127	6346.62 Rank: 6/127	4.682 Rank: 6/127	0.355 Rank: 1/127	0.78550 (±0.00157) Rank: 1/127	0.69405 Rank: 2/127
Submission ID: **00590** Guided-HardNet-epoch4 Size: **512** bytes. Matches: **custom**	7829.63	586.24 Rank: 34/127	0.486 Rank: 44/127	0.875 Rank: 7/127	0.59919 (±0.00000) Rank: 3/127	604.81 Rank: 33/127	5062.27 Rank: 22/127	4.710 Rank: 3/127	0.369 Rank: 7/127	0.76219 (±0.00253) Rank: 7/127	0.68069 Rank: 3/127
Submission ID: **00610** Hardnet-Upright-AdaLAM Size: **512** bytes. Matches: **custom**	6556.61	627.71 Rank: 19/127	0.442 Rank: 114/127	0.828 Rank: 79/127	0.58300 (±0.00000) Rank: 12/127	645.47 Rank: 27/127	5074.91 Rank: 21/127	4.575 Rank: 30/127	0.361 Rank: 3/127	0.77056 (±0.00064) Rank: 3/127	0.67678 Rank: 4/127
Submission ID: **00614** ContextDesc Upright [illegible] Size: **512** bytes. Matches: **custom**	[illegible]	[illegible]	[illegible]	[illegible]	[illegible] Rank: 17/127	[illegible]	[illegible]	[illegible]	[illegible]	0.77041 (±0.00298) Rank: 5/127	0.67433 Rank: 5/127
Submission ID: **00600** ContextDesc Upright + Mutual Che... Size: **512** bytes. Matches: **custom**	7830.09	624.55 Rank: 20/127	0.487 Rank: 25/127	0.847 Rank: 46/127	0.57344 (±0.00000) Rank: 19/127	644.39 Rank: 28/127	5427.23 Rank: 18/127	4.700 Rank: 4/127	0.368 Rank: 5/127	0.77043 (±0.00133) Rank: 4/127	0.67194 Rank: 6/127
Submission ID: **00567** Guided-HardNet32-v1-lib-qht-p Size: **512** bytes. Matches: **custom**	7829.63	520.40 Rank: 48/127	0.486 Rank: 44/127	0.875 Rank: 8/127	0.58509 (±0.00000) Rank: 10/127	536.93 Rank: 45/127	4685.16 Rank: 31/127	4.645 Rank: 12/127	0.376 Rank: 9/127	0.75702 (±0.00159) Rank: 12/127	0.67105 Rank: 7/127
Submission ID: **00611** sift and hardnet64 train scale(1... Size: **512** bytes. Matches: **built-in**	7830.09	622.13 Rank: 22/127	0.486 Rank: 32/127	0.871 Rank: 15/127	0.58870 (±0.00041) Rank: 5/127	899.14 Rank: 14/127	6086.16 Rank: 12/127	4.647 Rank: 10/127	0.386 Rank: 15/127	0.75127 (±0.00234) Rank: 14/127	0.66999 Rank: 8/127
Submission ID: **00568** Guided-SOSNet-lib-p Size: **512** bytes. Matches: **custom**	7829.63	508.45 Rank: 51/127	0.486 Rank: 44/127	0.874 Rank: 10/127	0.57982 (±0.00000) Rank: 16/127	524.66 Rank: 48/127	4618.86 Rank: 32/127	4.632 Rank: 16/127	0.369 Rank: 6/127	0.75888 (±0.00437) Rank: 10/127	0.66935 Rank: 9/127
Submission ID: **00613** HardNet64-data-aug-sort-51 Size: **512** bytes. Matches: **built-in**	7830.09	624.09 Rank: 21/127	0.486 Rank: 32/127	0.870 Rank: 16/127	0.58727 (±0.00076) Rank: 8/127	964.80 Rank: 9/127	6350.68 Rank: 5/127	4.644 Rank: 14/127	0.383 Rank: 13/127	0.74952 (±0.00162) Rank: 16/127	0.66839 Rank: 10/127

Brief reminder on the rules

- Number of features
  - "Restricted": up to 2048 features per image
  - "Unrestricted": up to 8000 features per image
- 2019-2020: Descriptor size
  - "Small": up to 128 bytes (32 float32)
    - Zero submissions!
  - "Regular": up to 512 bytes (128 float32)
    - The gold standard in academia
    - **Eligible for prizes (2k and 8k)**
  - "Large": up to 2048 bytes (512 float32)
    - Only papers in this category are D2-Net and SuperPoint

Brief reminder on the rules

- Number of features
  - "Restricted": up to 2048 features per image
  - "Unrestricted": up to 8000 features per image
- 2021: Descriptor size
  - ~~"Small": up to 128 bytes (32 float32)~~
    - ~~Zero submissions!~~
  - ~~"Regular": up to 512 bytes (128 float32)~~
    - ~~The gold standard in academia~~
    - ~~**Eligible for prizes (2k and 8k)**~~
  - ~~"Large": up to 2048 bytes (512 float32)~~
    - ~~Only papers in this category are D2-Net and SuperPoint~~
  - Eliminated all restrictions in order to facilitate experimentation
    - Requires a measure of good faith from the participants

The 2021 IMC Results

The 2021 Image Matching Challenge: Highlights

- 2019: 28 submission from 13 teams
- 2020: 102 submissions from 23 teams (plus 113 baselines)
- 2021: 91 submissions from 25 teams
- Why the drop? Delays were a factor (new datasets, COVID, etc)
  - 2020 challenge: February 10, 2020 - May 31, 2020 (~16 weeks)
  - 2021 challenge: May 10, 2021 - June 12, 2021 (~5 weeks)
  - Ok, but what else? Open discussion later!
- Anecdotal observation: not *that* many papers using it
  - Favoured: Aachen@LVTL, HPatches

Winners of IMC 2021: "unlimited" keypoints

WINNER

Xiaopeng Bi, Yu Chen, Xinyang Liu, Dehao Zhang, Ran Yan, Zheng Chai, Haotian Zhang & Xiao Liu

Megvii Inc. Research 3D

RUNNER-UP

Dongli Tan, Xingyu Chen, Ruixin Zhang, Kai Zhao, Xuehui Wang, Shaoxin Li, Jilin Li, Feiyue Huang & RongRong Ji

Youtu Lab, Tencent & Institute of Artificial Intelligence, Xiamen University

Winners of IMC 2021: "restricted" keypoints

WINNER

Dongli Tan, Xingyu Chen, Ruixin Zhang, Kai Zhao, Xuehui Wang, Shaoxin Li, Jilin Li, Feiyue Huang & RongRong Ji

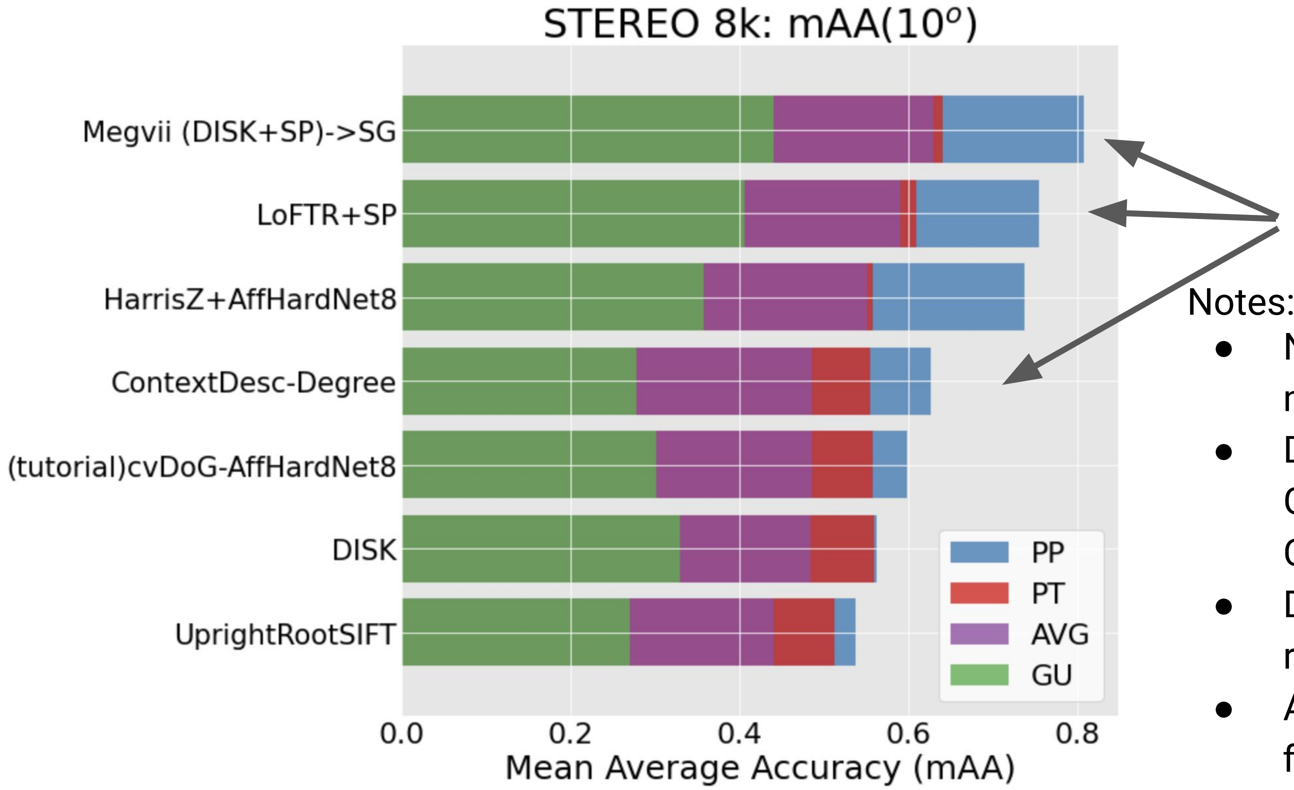
Youtu Lab, Tencent & Institute of Artificial Intelligence, Xiamen University

RUNNER-UP

Xiaopeng Bi, Yu Chen, Xinyang Liu, Dehao Zhang, Ran Yan, Zheng Chai, Haotian Zhang & Xiao Liu

Megvii Inc. Research 3D

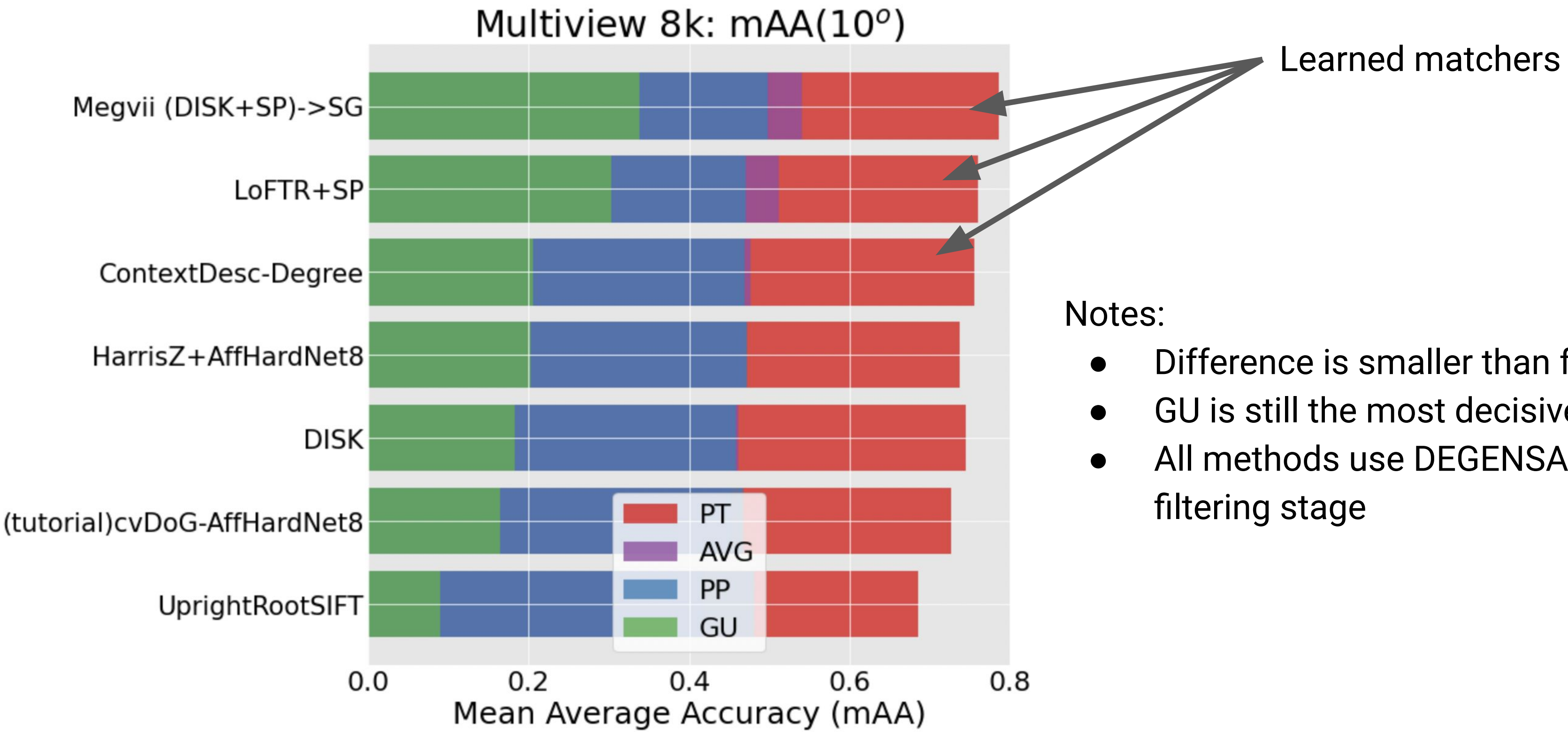
Stereo 8k category



Notes:

- New datasets (GU, PP) are more "decisive", than PT
- DoG (SIFT) performs poorly on GoogleUrban -- see tutorial, ContextDesc, RootSIFT
- DISK (trained on MegaDepth) is relatively bad in PP
- All methods use DEGENSAC as final filtering stage

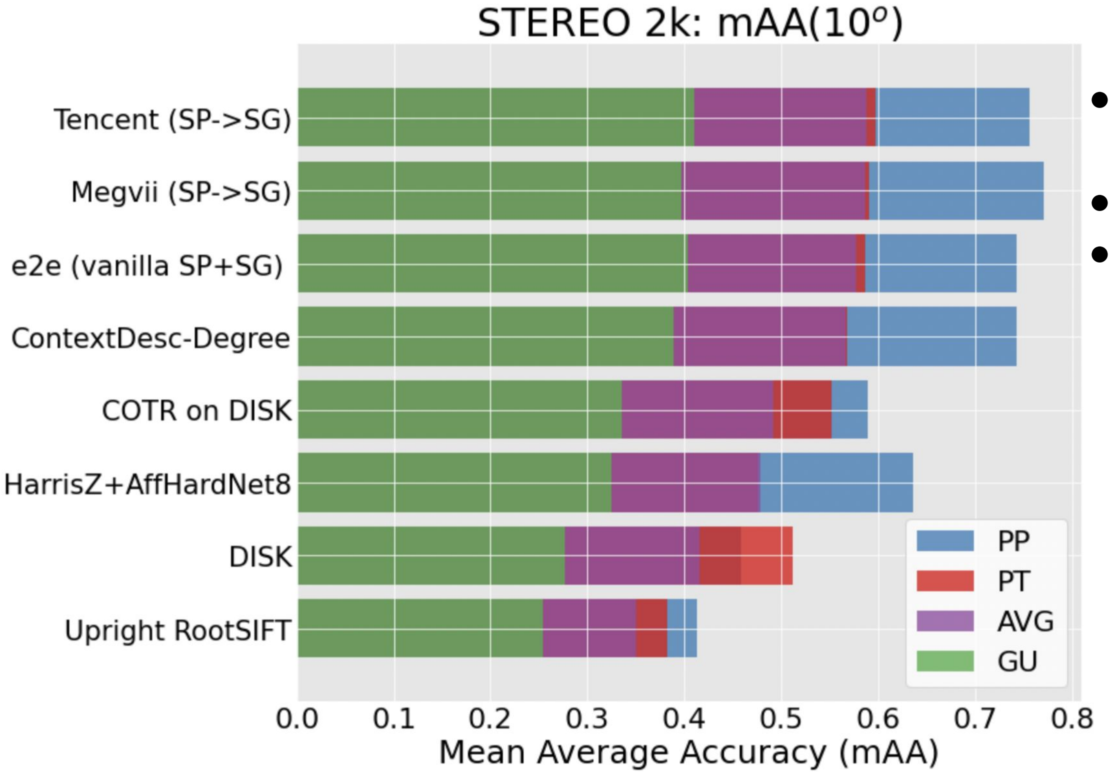
Multiview 8k category



Notes:

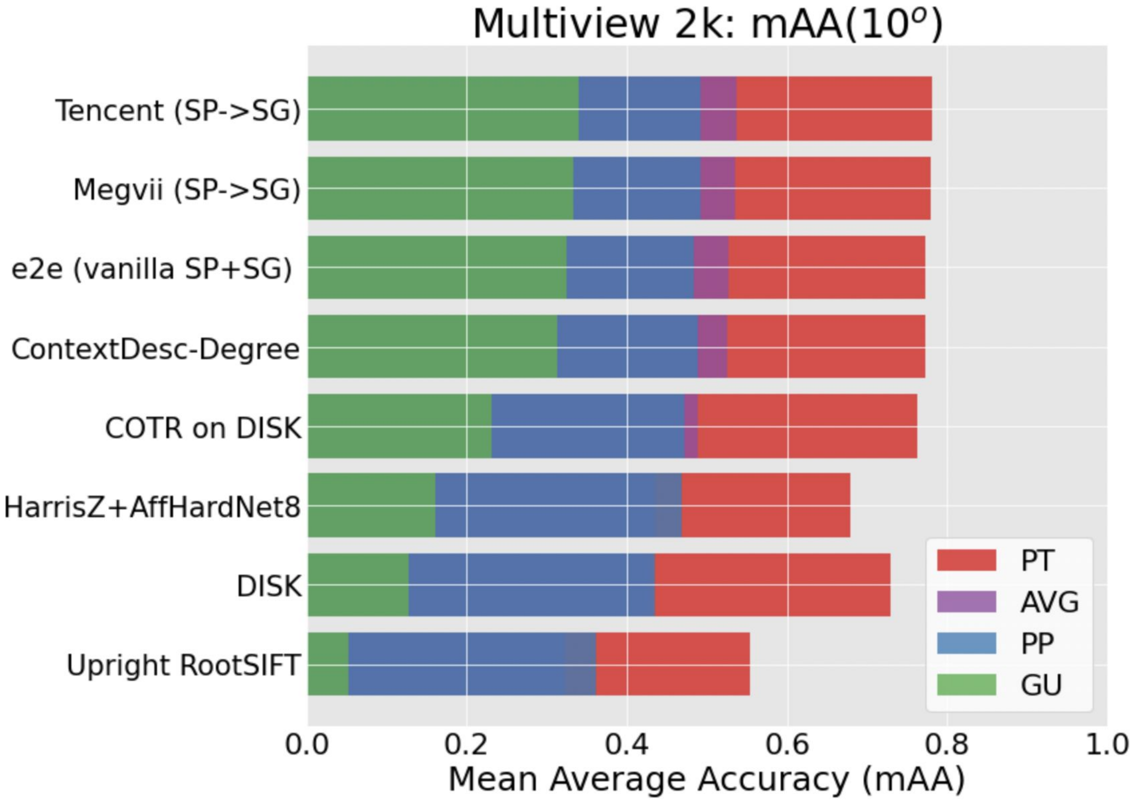
- Difference is smaller than for stereo,
- GU is still the most decisive
- All methods use DEGENSAC as final filtering stage

Stereo 2k category



- The improvement of the leaders over vanilla SuperPoint+SuperGlue is marginal
- DISK is overfit to buildings (bad on PP)
- All methods use DEGENSAC as final filtering stage

Multiview 2k category



- The improvement of the leaders over vanilla SuperPoint+SuperGlue is marginal
- RootSIFT is terrible for GU
- All methods use DEGENSAC as final filtering stage

Congratulations and thank you!

Method	NF	Stereo NI	Stereo Rep. (3 pix.)	Stereo MS (3 pix.)	Stereo mAA (at 10°)	Multiview NM	Multiview NL	Multiview TL	Multiview ATE	Multiview mAA (at 10°)	Avg. mAA (at 10°)
Submission ID: **00591** Guided-Hardnet-Epoch2 Size: **512** bytes. Matches: **custom**	7829.63	761.99 Rank: 13/127	0.486 Rank: 44/127	0.823 Rank: 88/127	0.61081 (±0.00000) Rank: 1/127	785.57 Rank: 23/127	6330.70 Rank: 7/127	4.680 Rank: 7/127	0.358 Rank: 2/127	0.78288 (±0.00094) Rank: 2/127	0.69684 Rank: 1/127
Submission ID: **00624** Guided-HardNet-OANet Size: **512** bytes. Matches: **custom**	7829.63	765.34 Rank: 12/127	0.486 Rank: 44/127	0.820 Rank: 90/127	0.60261 (±0.00000) Rank: 2/127	788.49 Rank: 21/127	6346.62 Rank: 6/127	4.682 Rank: 6/127	0.355 Rank: 1/127	0.78550 (±0.00157) Rank: 1/127	0.69405 Rank: 2/127
Submission ID: **00590** Guided-HardNet-epoch4 Size: **512** bytes. Matches: **custom**	7829.63	586.24 Rank: 34/127	0.486 Rank: 44/127	0.875 Rank: 7/127	0.59919 (±0.00000) Rank: 3/127	604.81 Rank: 33/127	5062.27 Rank: 22/127	4.710 Rank: 3/127	0.369 Rank: 7/127	0.76219 (±0.00253) Rank: 7/127	0.68069 Rank: 3/127
Submission ID: **00610** Hardnet-Upright-AdaLAM Size: **512** bytes. Matches: **custom**	6556.61	627.71 Rank: 19/127	0.442 Rank: 114/127	0.828 Rank: 79/127	0.58300 (±0.00000) Rank: 12/127	645.47 Rank: 27/127	5074.91 Rank: 21/127	4.575 Rank: 30/127	0.361 Rank: 3/127	0.77056 (±0.00064) Rank: 3/127	0.67678 Rank: 4/127
Submission ID: [illegible] ContextDesc [illegible] Size: **512** bytes. Matches: **custom**	[illegible]	[illegible]	[illegible]	[illegible]	[illegible] Rank: 17/127	[illegible]	[illegible]	[illegible]	[illegible]	[illegible] Rank: 5/127	0.67433 Rank: 5/127
Submission ID: **00600** ContextDesc Upright + Mutual Che... Size: **512** bytes. Matches: **custom**	7830.09	624.55 Rank: 20/127	0.487 Rank: 25/127	0.847 Rank: 46/127	0.57344 (±0.00000) Rank: 19/127	644.39 Rank: 28/127	5427.23 Rank: 18/127	4.700 Rank: 4/127	0.368 Rank: 5/127	0.77043 (±0.00133) Rank: 4/127	0.67194 Rank: 6/127
Submission ID: **00567** Guided-HardNet32-v1-lib-qht-p Size: **512** bytes. Matches: **custom**	7829.63	520.40 Rank: 48/127	0.486 Rank: 44/127	0.875 Rank: 8/127	0.58509 (±0.00000) Rank: 10/127	536.93 Rank: 45/127	4685.16 Rank: 31/127	4.645 Rank: 12/127	0.376 Rank: 9/127	0.75702 (±0.00159) Rank: 12/127	0.67105 Rank: 7/127
Submission ID: **00611** sift and hardnet64 train scale(1... Size: **512** bytes. Matches: **built-in**	7830.09	622.13 Rank: 22/127	0.486 Rank: 32/127	0.871 Rank: 15/127	0.58870 (±0.00041) Rank: 5/127	899.14 Rank: 14/127	6086.16 Rank: 12/127	4.647 Rank: 10/127	0.386 Rank: 15/127	0.75127 (±0.00234) Rank: 14/127	0.66999 Rank: 8/127
Submission ID: **00568** Guided-SOSNet-lib-p Size: **512** bytes. Matches: **custom**	7829.63	508.45 Rank: 51/127	0.486 Rank: 44/127	0.874 Rank: 10/127	0.57982 (±0.00000) Rank: 16/127	524.66 Rank: 48/127	4618.86 Rank: 32/127	4.632 Rank: 16/127	0.369 Rank: 6/127	0.75888 (±0.00437) Rank: 10/127	0.66935 Rank: 9/127
Submission ID: **00613** HardNet64-data-aug-sort-51 Size: **512** bytes. Matches: **built-in**	7830.09	624.09 Rank: 21/127	0.486 Rank: 32/127	0.870 Rank: 16/127	0.58727 (±0.00076) Rank: 8/127	964.80 Rank: 9/127	6350.68 Rank: 5/127	4.644 Rank: 14/127	0.383 Rank: 13/127	0.74952 (±0.00162) Rank: 16/127	0.66839 Rank: 10/127

SimLocMatch Challenge

Motivation

- Why do we need a synthetic dataset for evaluation?
  - Precise control of variation factors (lights, sun, occlusions)
  - Pixel-perfect accuracy of GT
  - Lack of introduced bias from pseudo-GT methods
  - Privacy!

- Potential drawback
  - Simulation vs reality gap
  - ~~Issue?~~ Opportunity!
    - Sim2Real will play an ever-increasing role in all fields - including image matching

Motivation: Pseudo vs *actual* GT

Pseudo GT

- Built Automated Methods
  (no guarantee about arbitrary pixels)
- Pose is paramount- since no pixel-level GT is available
- However pose is key in the localization task
- Forces us to use a proxy - and measure downstream tasks instead

Noname manuscript No.
(will be inserted by the editor)

Image Matching Across Wide Baselines: From Paper to Practice

Yuhe Jin · Dmytro Mishkin · Anastasiia Mishchuk · Jiri Matas · Pascal Fua · Kwang Moo Yi · Eduard Trulls

arXiv:2003.01587v5 [cs.CV] 11 Feb 2021

Received: date / Accepted: date

Abstract We introduce a comprehensive benchr
cal features and robust estimation algorithms, f
the downstream task – the accuracy of the re
camera pose – as our primary metric. Our pipelin
structure allows easy integration, configuration,
nation of different methods and heuristics. This
strated by embedding dozens of popular algo
evaluating them, from seminal works to the cutt
machine learning research. We show that with
tings, classical solutions may still outperform th
state of the art.

Besides establishing the *actual state of the a*
ducted experiments reveal unexpected propertie
ture from Motion (SfM) pipelines that can he

This work was partially supported by the Natural Scie
gineering Research Council of Canada (NSERC) Dis
"Deep Visual Geometry Machines" (RGPIN-2018-037
tems supplied by Compute Canada, and by Google's
sitioning Service. DM and JM were supported b
funded project CZ.02.1.01/0.0/0.0/16 019/0000765 "R
ter for Informatics". DM was also supported by CTU
SGS17/185/OHK3/3T/13 and by the Austrian Ministry f

Comparative Evaluation of Hand-Crafted and Learned Local Features

Johannes L. Schönberger[1] Hans Hardmeier[1] Torsten Sattler[1] Marc Pollefeys[1,2]
[1] Department of Computer Science, ETH Zürich [2] Microsoft Corp.
{jsch,harhans,sattlert,pomarc}@inf.ethz.ch

Abstract

Matching local image descriptors is a key step in many computer vision applications. For more than a decade, hand-crafted descriptors such as SIFT have been used for this task. Recently, multiple new descriptors learned from data have been proposed and shown to improve on SIFT in terms of discriminative power. This paper is dedicated to an extensive experimental evaluation of learned local features to establish a single evaluation protocol that ensures comparable results. In terms of matching performance, we evaluate the different descriptors regarding standard criteria. However, considering matching performance in isolation only provides an incomplete measure of a descriptor's quality. For example, finding additional correct matches between similar images does not necessarily lead to a better performance when trying to match images under extreme viewpoint or illumination changes. Besides pure descriptor matching, we thus also evaluate the different descriptors in the context of image-based reconstruction. This enables us to study the descriptor performance on a set of more practical criteria including image retrieval, the ability to register images under strong viewpoint and illumination changes, and the accuracy and completeness of the reconstructed cameras and scenes. To facilitate future research, the full evaluation pipeline is made publicly available.

ability of neural networks to learn feature representations from data that are superior to prior hand-crafted ones has led to significant progress in the field of computer vision, *e.g.*, in object detection and recognition [12, 23, 41]. Consequently, neural networks have also been applied to the problem of descriptor learning [3, 14, 24, 42] in order to derive more discriminative representations for local features. The resulting methods demonstrate clear improvements over standard hand-crafted representations, such as SIFT [26], SURF [4], or DAISY [46]. However, there is usually no direct comparison with more advanced hand-crafted SIFT variants such as RootSIFT [2], RootSIFT-PCA [7], or DSP-SIFT [9]. Moreover, learned descriptors are typically evaluated on the patch classification benchmark from Brown *et al.* [6]. The task measures how well a descriptor can distinguish between related and unrelated patches based on their distance in descriptor space. Yet, a better performance on this benchmark does not necessarily imply a better matching quality, as shown by Balntas *et al.* [3]. For example, pruning steps such as Lowe's ratio test [26] or mutual nearest neighbor constraints might compensate for a higher false positive matching rate in terms of descriptor distance. Similarly, reaching a better average matching performance does not automatically imply a better performance in terms of subsequent processing steps. In the context of SFM, finding additional correspondences for image pairs where SIFT already provides enough matches does not necessarily re-

Actual GT

- Built using manual annotation (HPatches)
- Built using actual model GT (simulation)
  (guarantee about arbitrary pixels)
- No pose is needed - GT available for all pixels

Motivation: Consistency of *pseudo* GT

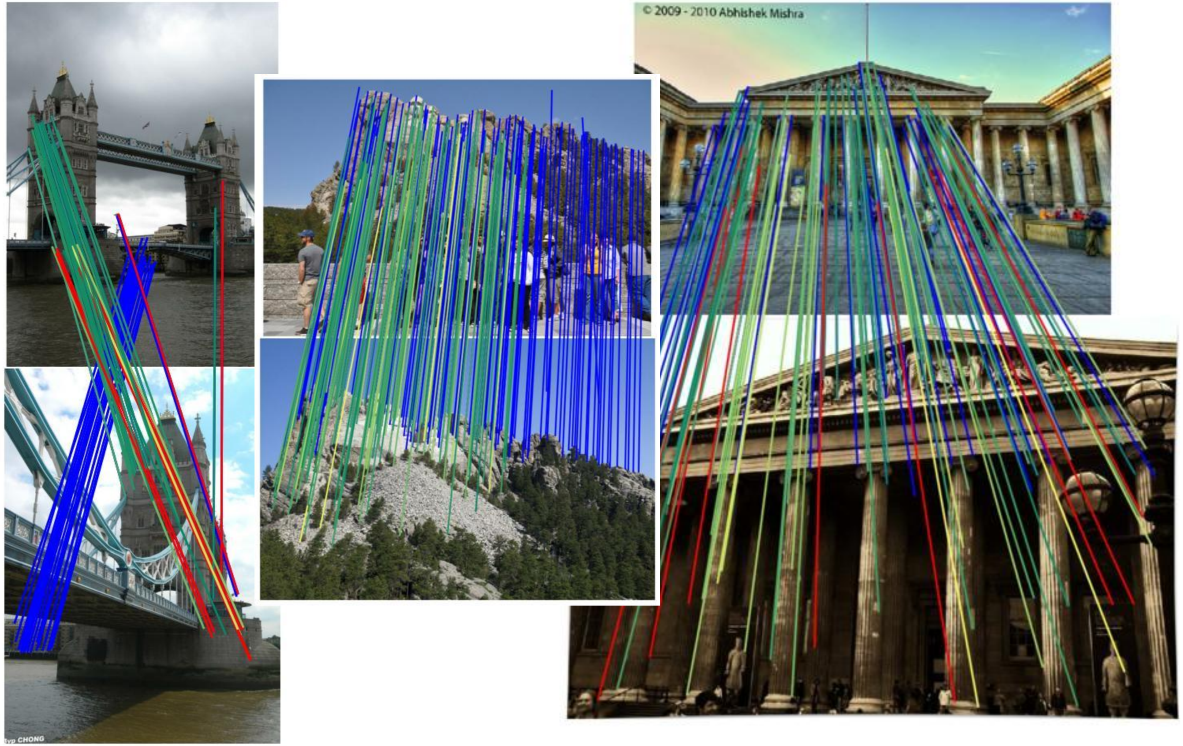


Table 1. Contradicting conclusions reported in literature while evaluating the same descriptors on the same benchmark (Oxford [22]). Rows report inconsistent evaluation results due to variations of the implicit parameters e.g. of feature detectors.

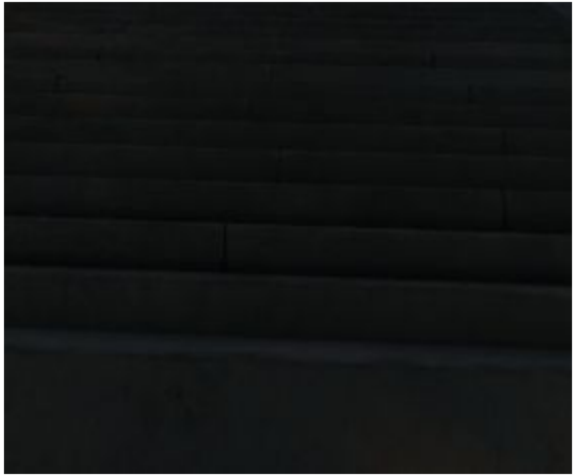
LIOP > SIFT	[24, 36]	,	SIFT > LIOP	[39]
BRISK > SIFT	[18, 24]	,	SIFT > BRISK	[19]
ORB > SIFT	[29]	,	SIFT > ORB	[24]
BINBOOST > SIFT	[19, 32]	,	SIFT > BINBOOST	[5, 39]
ORB > BRIEF	[29]	,	BRIEF > ORB	[19]

From the PhotoTourism leaderboard:

Matches for which we do not have depth estimates are drawn in blue.
Please note that the depth maps are estimates and may contain errors.

Motivation: Challenging scenes not suitable for building *pseudo* GT using SfM





SimLocMatch

Goal: Utilize 3D models + simulation to build large-scale benchmarks for image matching and visual localization.

* 7 scenes
* 80k image pairs

Details about building the datasets and generating the challenges + results are coming later this year in a technical report.

Future Roadmap

- Image Matching Challenge will be released by end of 2021
  - Large number of scenes, variations & occlusions
  - Detection & Relative Pose Estimation Tasks
  - Validation Set
  - More Metrics, More Tasks (e.g. Semantic Matching, Line/Plane Matching)

- Visual Localization Challenge
  - Different than image matching scenes to avoid overfitting
  - ICCV 2021 Visual Localization Workshop

- Matching & Localization
  - A small set of scenes will be jointly parts of both Matching+Localization challenges to facilitate interesting research on their relation

SimLocMatch: Future Research Roadmap

2019 *In this workshop, we aim to encourage novel strategies for image matching that deviate from and advance traditional formulations, with a focus on large-scale, wide-baseline matching for 3D reconstruction or pose estimation. This can be achieved by applying new technologies to sparse feature matching, or doing away with keypoints and descriptors entirely.*

SimLocMatch: Future Research Roadmap

Research Enablement Goal: Be able to facilitate “*doing away with keypoints and descriptors entirely*”

- Extremely limited keypoints (~8)
- Matching using non-point primitives instead of SfM (line matching, plane matching)
- Utilize GT semantics/geometry of scenes

SimLocMatch CVPR 2021 Challenge Winners

# Teams	# Submissions	# Public Submissions
19	174	43

SimLocMatch CVPR 2021 Challenge Winners

Final Ranking Metric for CVPR 2021: Matching Success Rate
- Given a random match **m**, probability of **m** being correct
- Incorporation of metrics such as false positives, will come later this year

SimLocMatch CVPR 2021 Challenge Winners

WINNER

Xiaopeng Bi, Yu Chen, Xinyang Liu, Dehao Zhang, Ran Yan, Zheng Chai, Haotian Zhang & Xiao Liu

Megvii Inc. Research 3D

RUNNER-UP

Jiaming Sun, Xingyi He, Zehong Shen, Yuang Wang (LoFTR)

Zhejiang University & SenseTime Research

HONORABLE MENTION

Fabio Bellavia and Dmytro Mishkin (HarrisZ+)

Università degli Studi di Palermo, Czech Technical University in Prague

(Some) Learnings from this first version

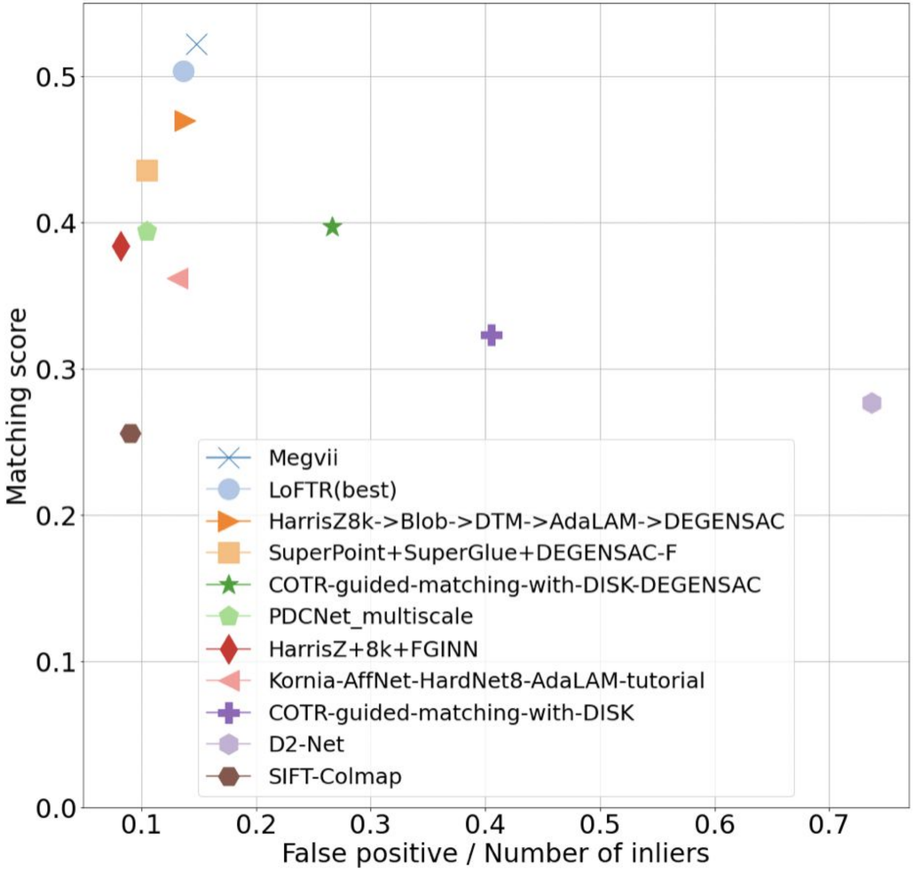
- **Negative results matter!**
  - Top performing methods (Megvii & LoFTR) have both ~10% ratio between TP and FP matches.
  - Most methods at around ~30-50%
  - Some methods are close to indistinguishable (D2Net ~90%)
- **SOTA transformer methods are better than a local features + deep learning elements based pipeline (HarrisZ*), but not hugely. Minor gains when comparing the gains of HarrisZ*** w.r.t **SIFT**.

Method	Matching Success Rate
Megvii	52.19%
LoFTR	50.36%
HarrisZ+8k->Blob->DTM->AdaLAM->DEGENSAC	46.94%
SIFT	25.59%

(Some) Learnings from this first version

Negative results matter!

- Top performing methods (Megvii & LoFTR, HarrisZ) have both ~10% ratio between TP and FP matches.
- Most methods at around ~30-50%
- Some methods are close to indistinguishable (D2Net ~70%)
- COTR w/o DEGENSAC: 40% FP,
  with DEGENSAC: 27% FP
- PDCNet: in between
- Blob-DTM-AdaLAM greatly improve matching score (HarrisZ), but not camera pose (in IMC challenge)



Public Leaderboard - Image Matching

Show 100 entries

Search:

Image Matching				
Name	# Inliers (Matching Images)	Matching Success Rate (Matching Images)	# Matches (Non-Matching Images)	Date
aaa-1000k_50_no_ms111	214.70	51.56%	21.80	2021/06/20, 16:00
aaa-1000k_80_no_ms111	275.39	51.40%	26.30	2021/06/20, 16:29
aaa-1000 _no_ms			30.01	
aaa-1000 _no_ms2			23.15	
COTR-gu ed-match g-with-Di K			43.65	
COTR-gu ed-match g-with-Di K-Degen c			15.19	
D2-Net			91.35	
DISK_12 AdaLAM_ v2_USAC MAGSAC bKBISe			14.94	
DISK_12 AdaLAM_ v2_USAC MAGSAC CSaW7			15.76	
DISK_8k daLAM_ 2_USAC MAGSAC YuvqRx			11.31	
DISK_8k v2_USAC MAGSAC_ 5XHiEj	104.32	42.55%	13.19	[illegible]
DISK_8k_c v2_USAC_ MAGSAC_ g2vY9A	147.98	39.65%	32.80	2021/06/17, 13:51
DoG+HarrisZ8k+-AdaLAM	91.32	43.72%	10.61	2021/06/13, 08:57

9:00 - 9:15	**Welcome session (Eduard Trulls)**
9:15 - 10:00	**Invited Talk: Davide Scaramuzza (University of Zurich/ETH Zurich)**
10:00- 10:45	**Invited Talk: Marc Pollefeys (ETH Zurich/Microsoft)**
10:45 - 11:00	***Perceptual Loss for Robust Unsupervised Homography Estimation*** Daniel Koguciuk (Advanced Research Lab, NavInfo Europe, NL)
11:00 - 11:15	***DFM: A Performance Baseline for Deep Feature Matching*** Ufuk Efe (Middle East Technical University, Ankara, Turkey)
11:15 - 11:45	**Challenge presentation**
11:45 - 12:15	**Open discussion**
12:15 - 13:35	**Challenge participant talks** 12:15-12:25: Fabio Bellavia (University of Palermo) 12:25-12:35: Prune Truong (ETH Zurich) 12:35-12:45: Jiaming Sun/Xingyi He (Zhejiang University, SenseTime Research) 12:45-12:55: Wei Jiang (University of British Columbia) 12:55-13:05: Megvii 3D 13:05-13:15: Tencent
13:15 - 13:20	**Closing**

or…?

The 2021 Image Matching Challenge: Highlights

- Performance is not saturated (on PhotoTourism), but most submissions were highly competitive
  - Organizers submitted fewer baselines
- Nearly all submissions used custom matchers
- More engineering rather than "ground-breaking" papers
  - 2020: SuperGlue, AdaLAM, DISK, etc (many used by top methods in 2021). To be expected?
  - Nothing fully end-to-end yet.

One caveat: The challenge that did not happen

- IMC: We extensively explored a collaboration with Kaggle
- Why? Notebook-based submissions
  - Allows for a truly private test set where "cheating" is not a factor
  - Makes categories irrelevant in favour or a fixed compute budget
- Why did it not happen?
  - Time constraints
  - Difficulty in combining both frameworks

Your input: IMC

- Ease of use?
  - Running it on your own
  - Submitting
- Other tasks?
- More/fewer data?
- Current rules (e.g. desc size)?
- Pose submissions?
- Is average rank a good way to combine?

- Why do I have to submit a PDF after the fact?
- Why does it take time to process an entry? Why can't I edit/delete?
- Is it difficult to use non-standard methods (e.g. keypoint-agnostic)?
- What do you like/dislike?
- What else would you like to see?
- Does it help you publish papers?

Your input: SimLocMatch

- What would researchers would like to see as first priority?
  - Semantics? geometry? cars + objects?
- How is the submission process?
- What other tasks would be interesting except the ones already planned (Detectors, Relative Poses)
- Evaluation server pain points

Workshop website: https://image-matching-workshop.github.io

9:00 - 9:15	**Welcome session (Eduard Trulls)**
9:15 - 10:00	**Invited Talk: Davide Scaramuzza (University of Zurich/ETH Zurich)**
10:00- 10:45	**Invited Talk: Marc Pollefeys (ETH Zurich/Microsoft)**
10:45 - 11:00	***Perceptual Loss for Robust Unsupervised Homography Estimation*** Daniel Koguciuk (Advanced Research Lab, NavInfo Europe, NL)
11:00 - 11:15	***DFM: A Performance Baseline for Deep Feature Matching*** Ufuk Efe (Middle East Technical University, Ankara, Turkey)
11:15 - 11:45	**Challenge presentation**
11:45 - 12:15	**Open discussion**
12:15 - 13:35	**Challenge participant talks** 12:15-12:25: Fabio Bellavia (University of Palermo) 12:25-12:35: Prune Truong (ETH Zurich) 12:35-12:45: Jiaming Sun/Xingyi He (Zhejiang University, SenseTime Research) 12:45-12:55: Wei Jiang (University of British Columbia) 12:55-13:05: Megvii 3D 13:05-13:15: Tencent
13:15 - 13:20	**Closing**

(Challenge talks)

Workshop website: https://image-matching-workshop.github.io

9:00 - 9:15	**Welcome session (Eduard Trulls)**
9:15 - 10:00	**Invited Talk: Davide Scaramuzza (University of Zurich/ETH Zurich)**
10:00- 10:45	**Invited Talk: Marc Pollefeys (ETH Zurich/Microsoft)**
10:45 - 11:00	***Perceptual Loss for Robust Unsupervised Homography Estimation*** Daniel Koguciuk (Advanced Research Lab, NavInfo Europe, NL)
11:00 - 11:15	***DFM: A Performance Baseline for Deep Feature Matching*** Ufuk Efe (Middle East Technical University, Ankara, Turkey)
11:15 - 11:45	**Challenge presentation**
11:45 - 12:15	**Open discussion**
12:15 - 13:35	**Challenge participant talks** 12:15-12:25: Fabio Bellavia (University of Palermo) 12:25-12:35: Prune Truong (ETH Zurich) 12:35-12:45: Jiaming Sun/Xingyi He (Zhejiang University, SenseTime Research) 12:45-12:55: Wei Jiang (University of British Columbia) 12:55-13:05: Megvii 3D 13:05-13:15: Tencent
13:15 - 13:20	**Closing**

Thanks for your attention and participation!

Last chance for questions!