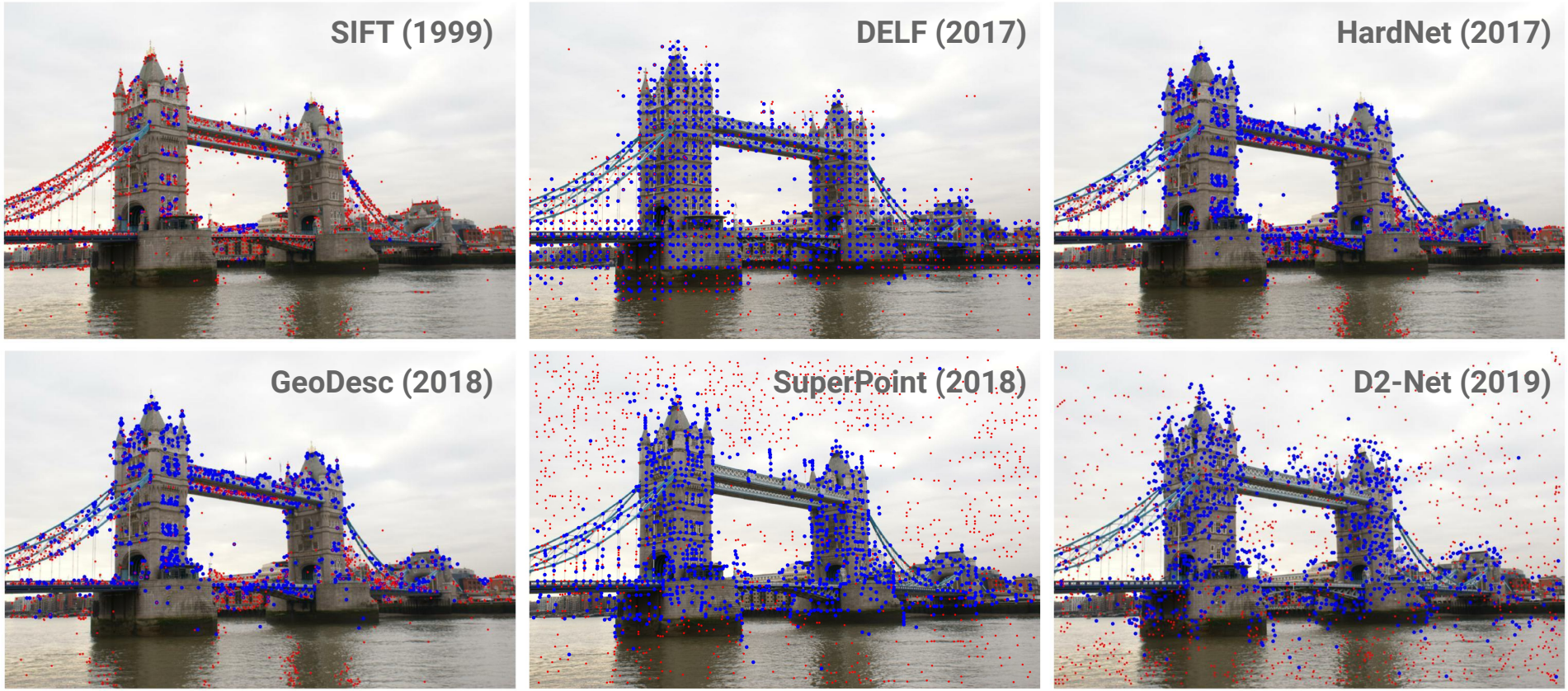


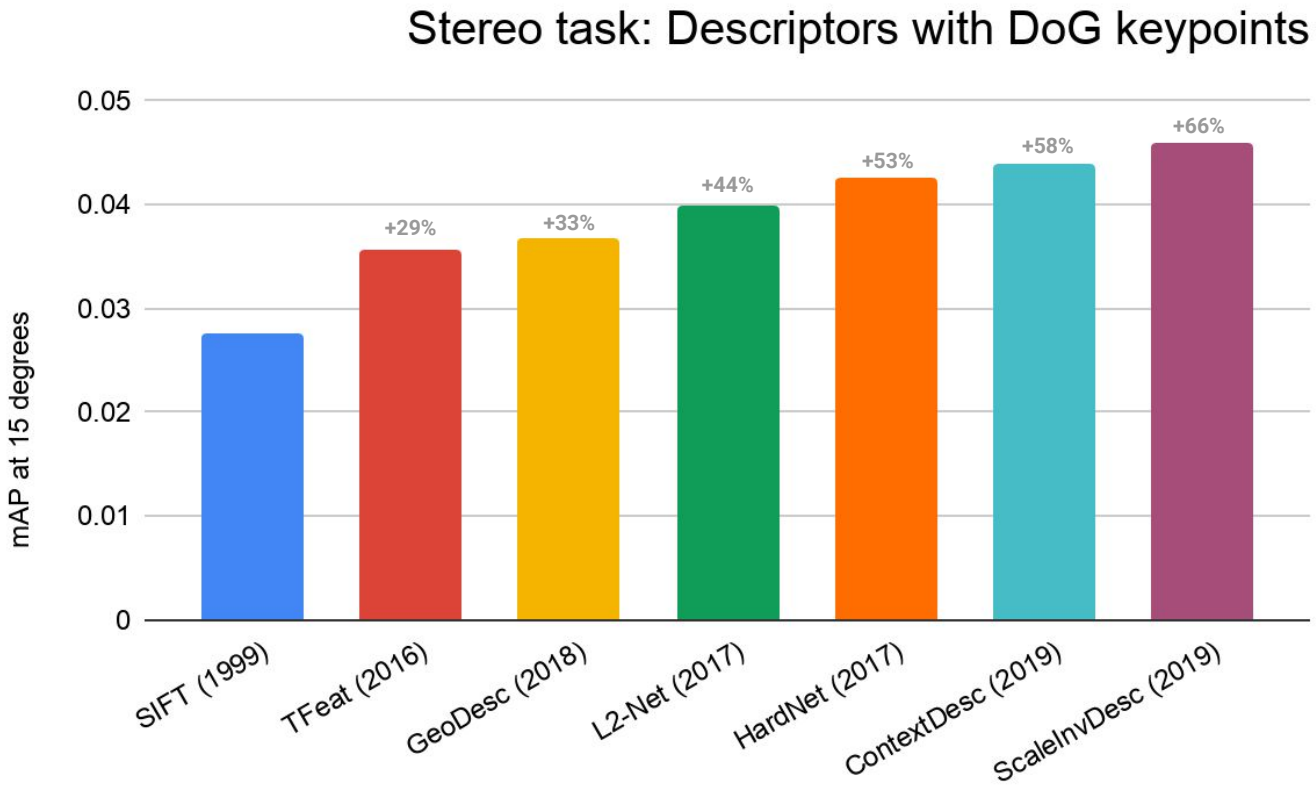
Phototourism Challenge Results

Summary

- 24 submissions + 26 baselines
  - Includes multiple settings (# keypoints, matching strategies)
  - Up to 8000 keypoints per image, broken down into categories by # of keypoints
  - Anonymous submissions allowed (and encouraged)

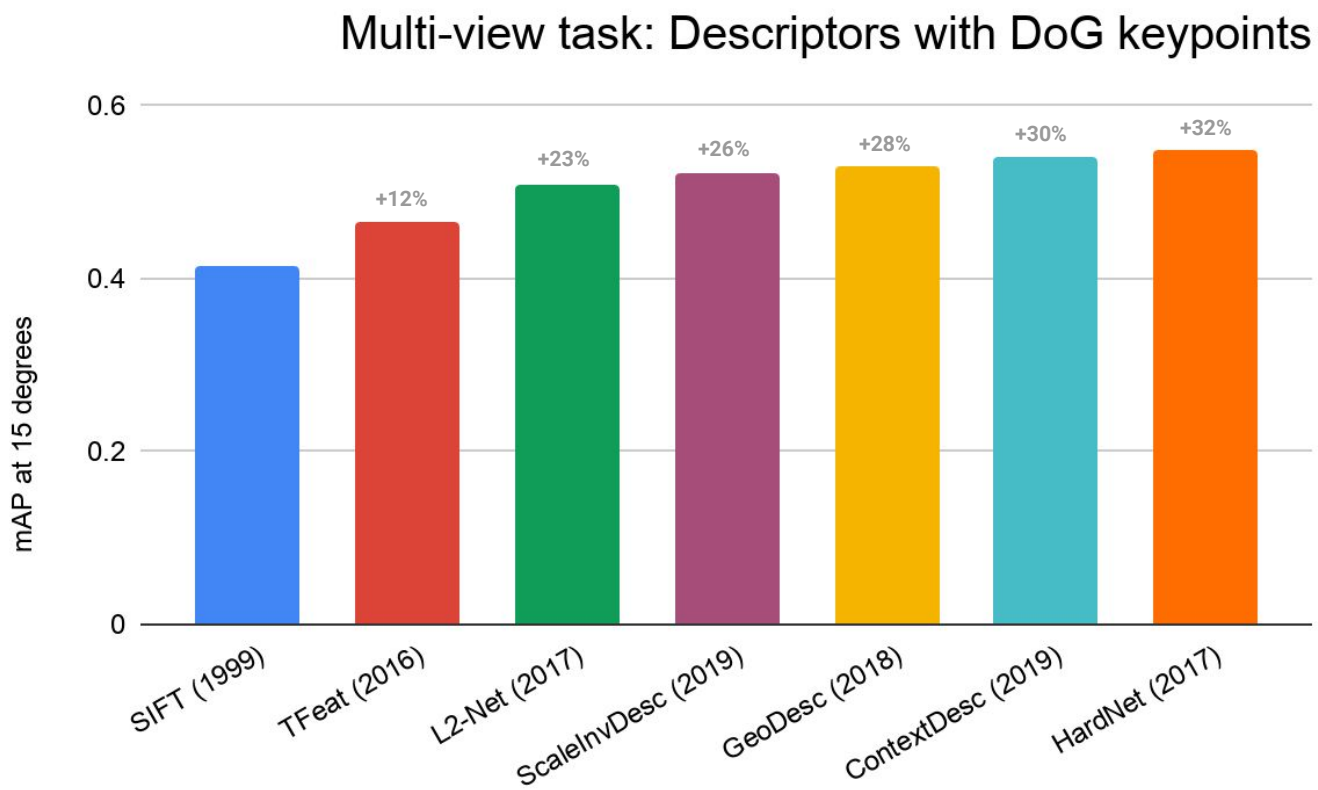
- All submissions processed for both tasks
  - We receive features and do the heavy lifting
  - Test set remains private

Improving with descriptors (stereo task)



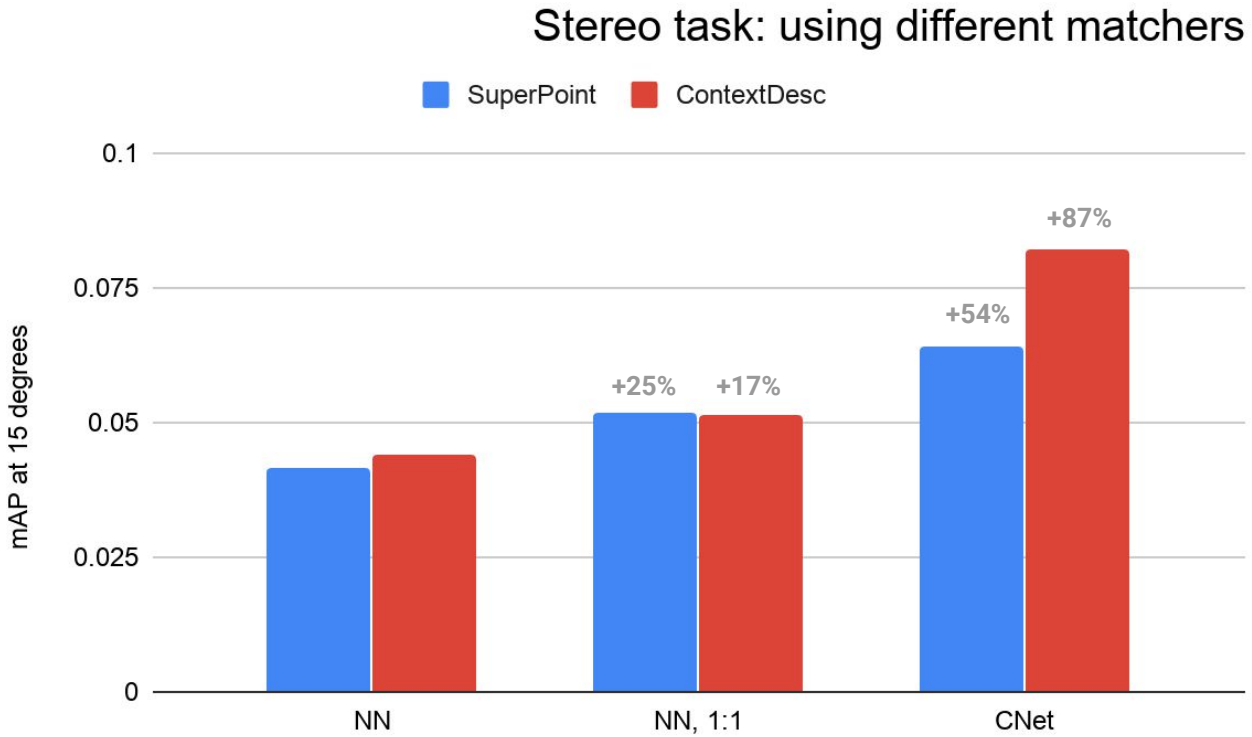
Full results: https://image-matching-workshop.github.io/leaderboard

Improving with descriptors (multi-view task)



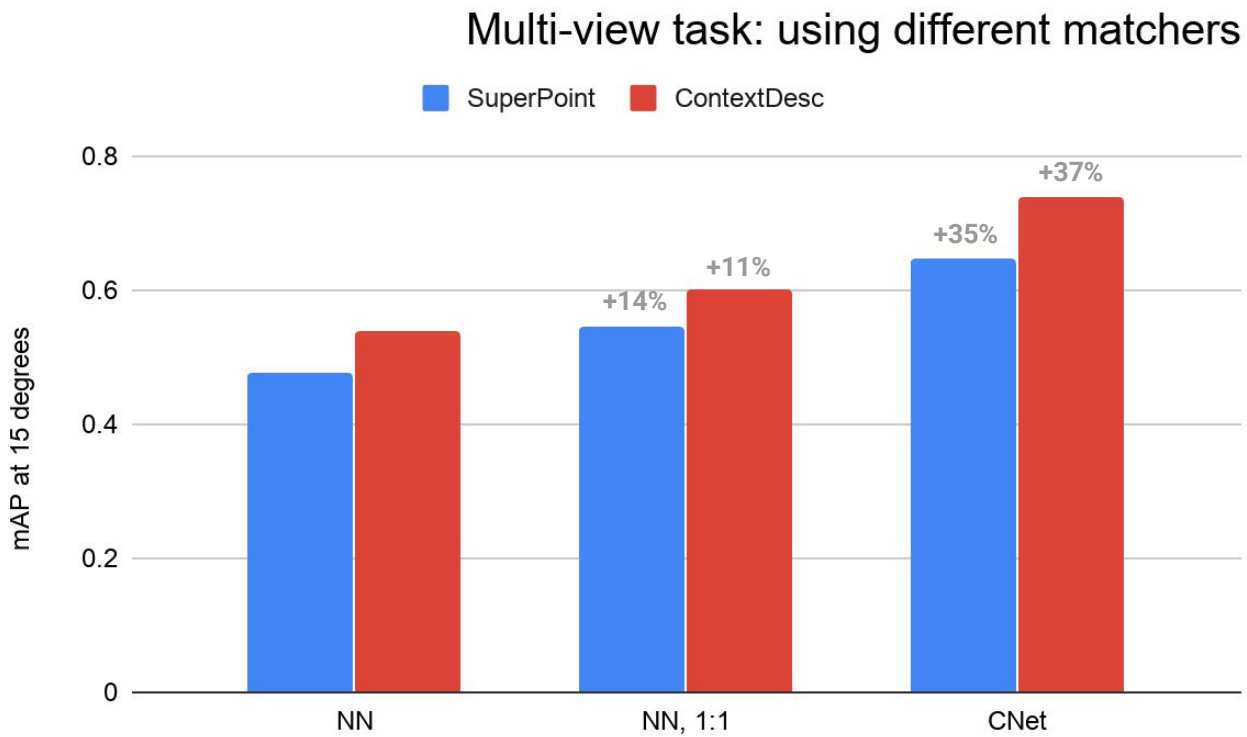
Full results: https://image-matching-workshop.github.io/leaderboard

Improving with matching (stereo task)



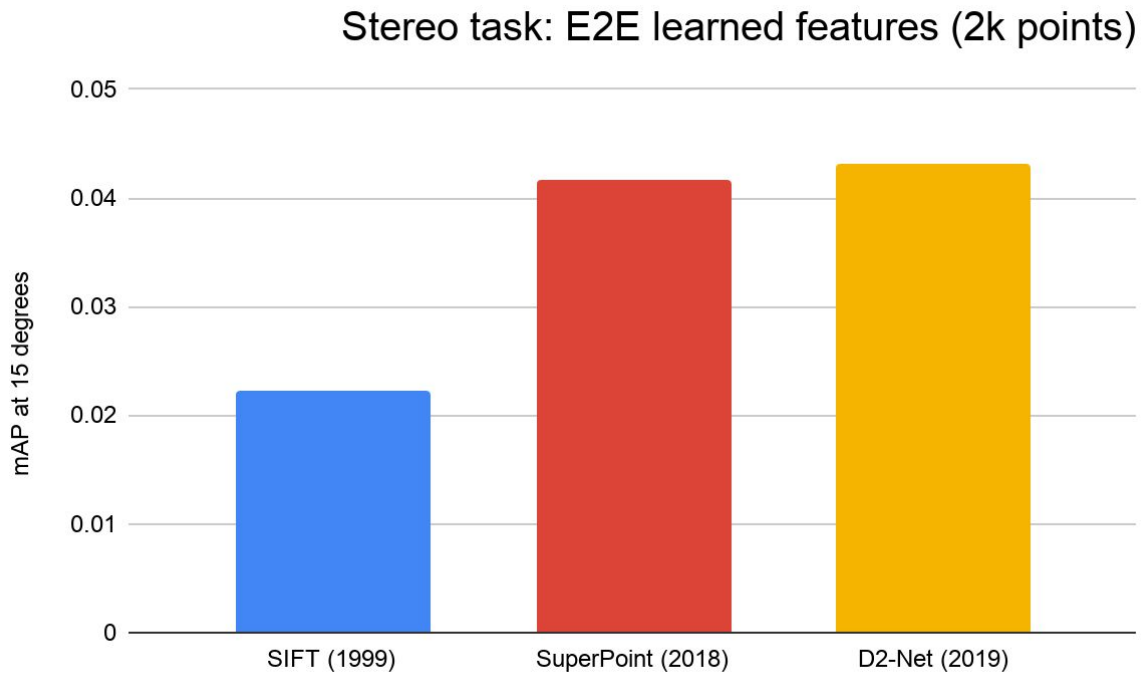
SuperPoint: Self-Supervised Interest Point Detection and Description. DeTone et al., 2018.
ContextDesc: Local Descriptor Augmentation with Cross-Modality Context. Luo et al., CVPR'19
Learning to Find Good Correspondences. Yi et al., CVPR'18

Improving with matching (multi-view task)



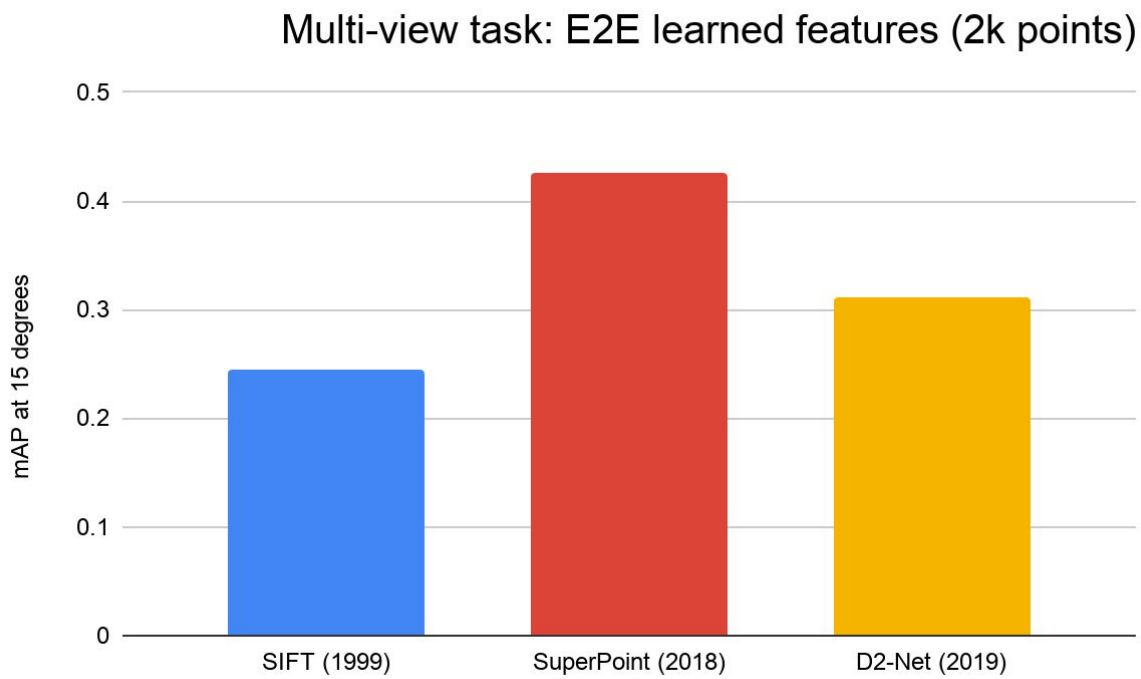
SuperPoint: Self-Supervised Interest Point Detection and Description. DeTone et al., 2018.
ContextDesc: Local Descriptor Augmentation with Cross-Modality Context. Luo et al., CVPR'19
Learning to Find Good Correspondences. Yi et al., CVPR'18

End-to-end pipelines



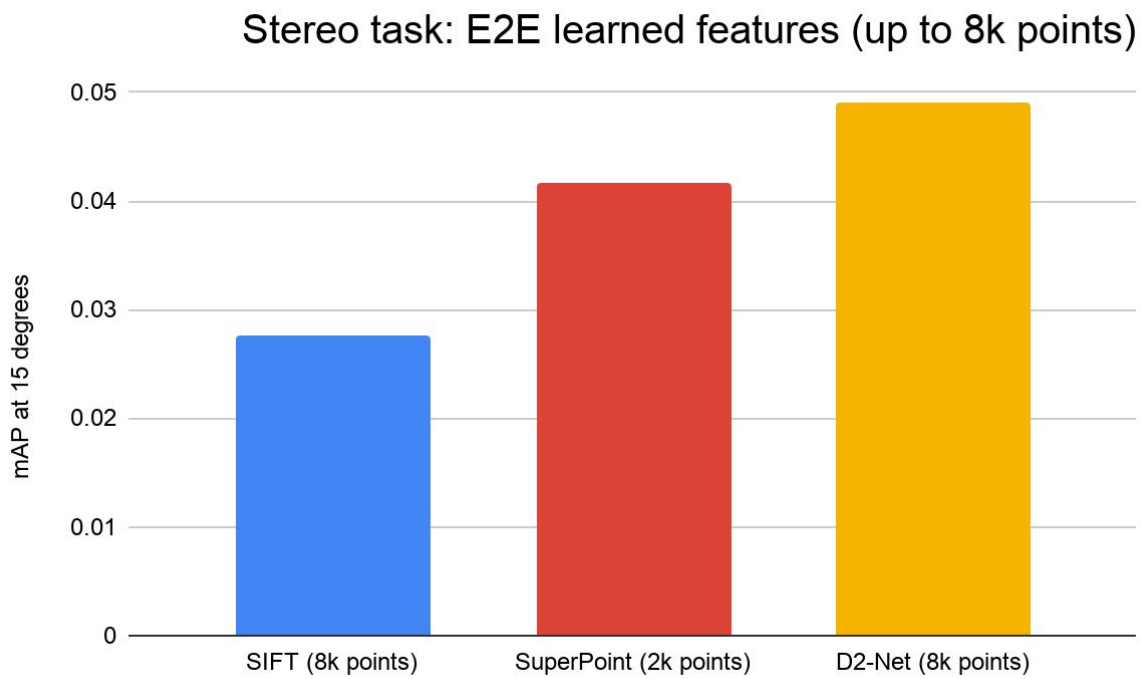
SuperPoint: Self-Supervised Interest Point Detection and Description. DeTone et al., 2018.
D2-Net: A Trainable CNN for Joint Detection and Description of Local Features. Dusmanu et al., CVPR'19

End-to-end pipelines



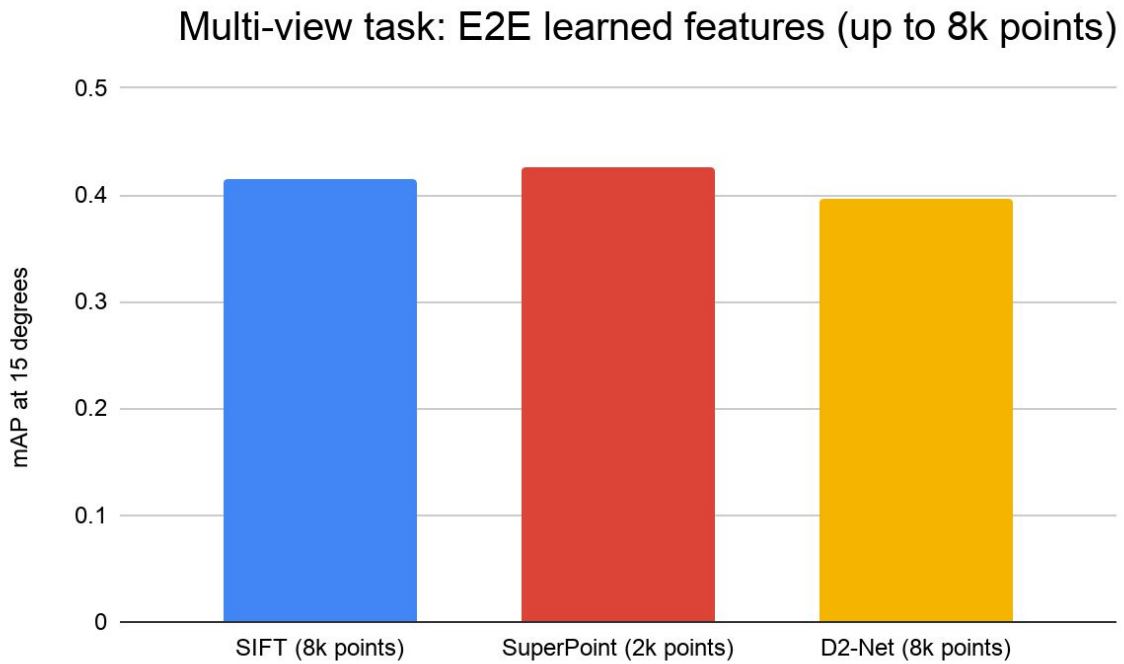
SuperPoint: Self-Supervised Interest Point Detection and Description. DeTone et al., 2018.
D2-Net: A Trainable CNN for Joint Detection and Description of Local Features. Dusmanu et al., CVPR'19

End-to-end pipelines



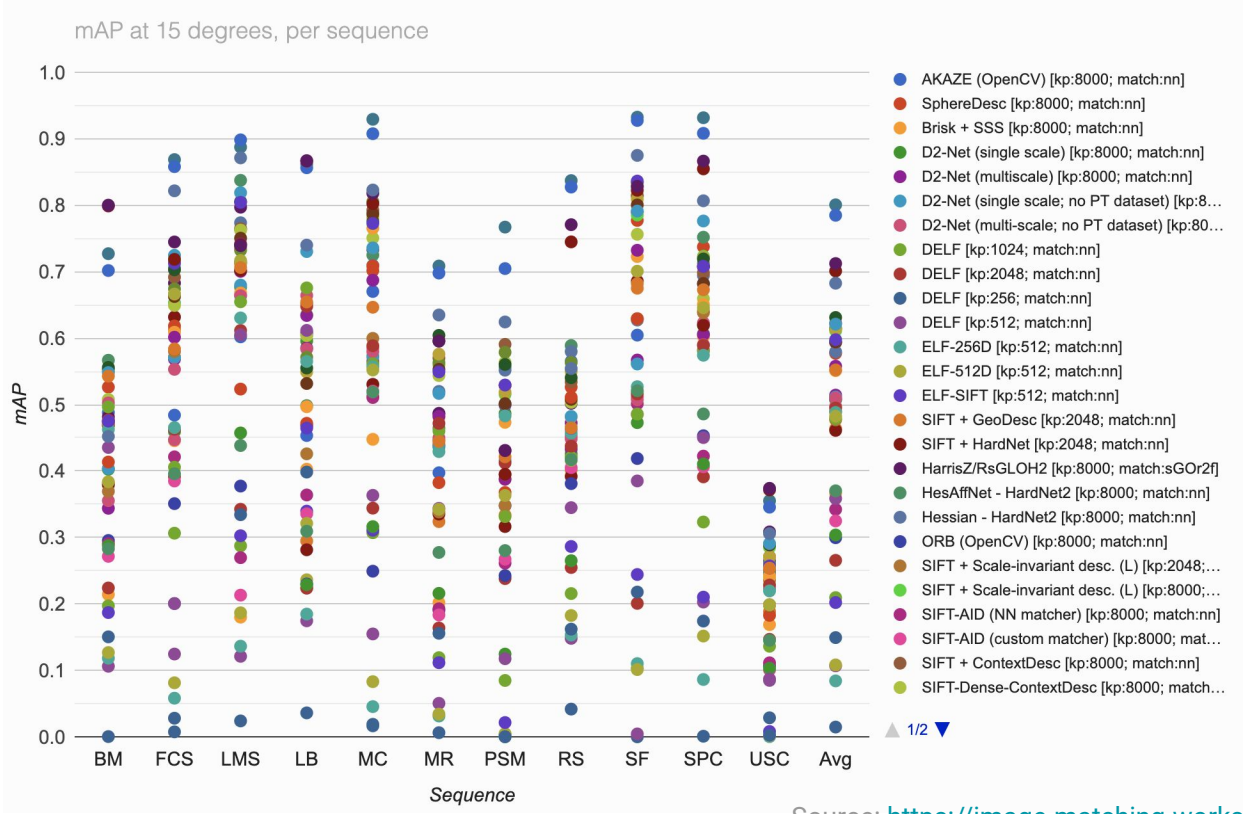
SuperPoint: Self-Supervised Interest Point Detection and Description. DeTone et al., 2018.
D2-Net: A Trainable CNN for Joint Detection and Description of Local Features. Dusmanu et al., CVPR'19

End-to-end pipelines



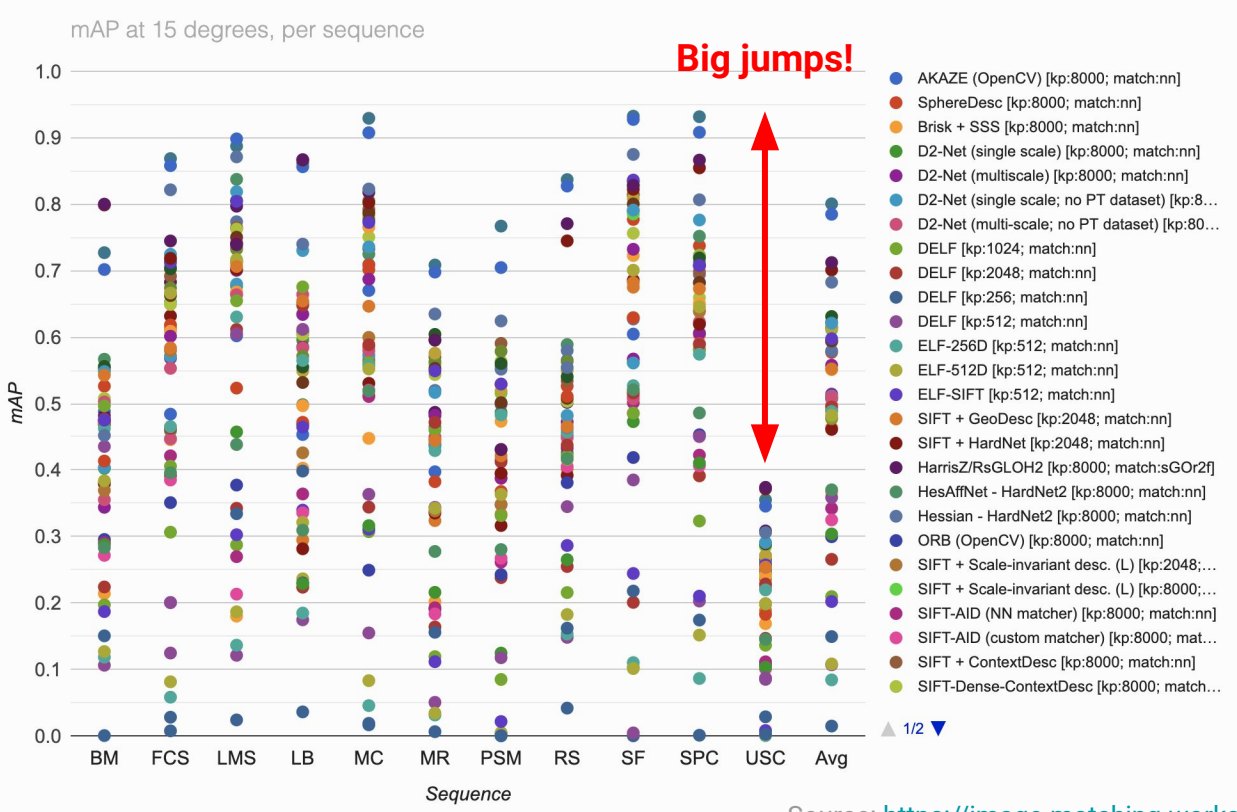
SuperPoint: Self-Supervised Interest Point Detection and Description. DeTone et al., 2018.
D2-Net: A Trainable CNN for Joint Detection and Description of Local Features. Dusmanu et al., CVPR'19

Per-sequence results (surprises!)



Source: https://image-matching-workshop.github.io/leaderboard

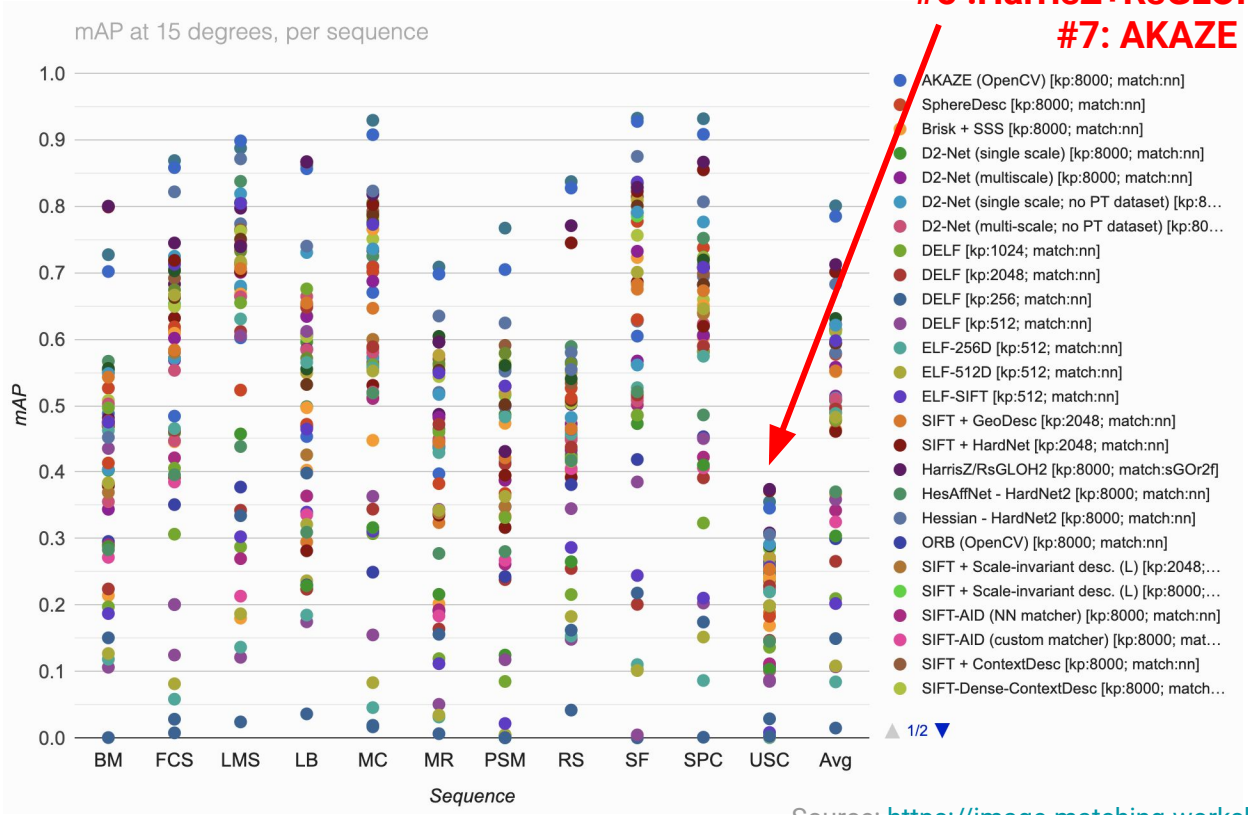
Per-sequence results (surprises!)



Source: https://image-matching-workshop.github.io/leaderboard

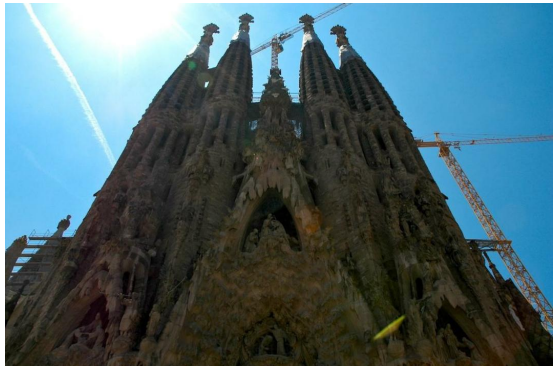
Per-sequence results (surprises!)

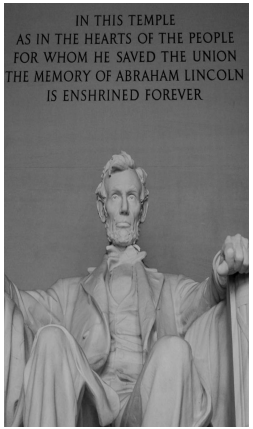
Inconsistent ranks!
#5 :HarrisZ+RsGLOH2 (link)
#7: AKAZE



Source: https://image-matching-workshop.github.io/leaderboard

Per-sequence results (surprises!)





SILDa Challenge results

Matching Scores: Results

Method Name	Matching Score *0.1*	# inliners *0.1*
Hessian-HardNet2	**36.41**	**52.28**
HesAffNet-HardNet2	**36.43**	**60.10**
mkdnet-PT	31.57	40.61
mkdnet-SILDa	23.31	45.22
ELF-SIFT	28.23	5.23
ELF-SIFT51	27.20	10.75
SOSNet	**37.12**	39.05
BRISK	29.39	42.24

Epipolar Arc Distance Statistics: Results

Method Name	mean dist. to ep. arc	median dist. to ep. arc
Hessian-HardNet2	**0.12**	**0.05**
HesAffNet-HardNet2	**0.17**	**0.08**
mkdnet-PT	0.18	0.10
mkdnet-SILDa	0.26	0.17
ELF-SIFT	0.23	0.13
ELF-SIFT51	0.25	0.15
SOSNet	0.18	0.09
BRISK	0.26	0.18

Number of image pairs with more than 8 inliers: Results

Method Name	% image pairs
Hessian-HardNet2	38.71
HesAffNet-HardNet2	55.00
mkdnet-PT	49.25
mkdnet-SILDa	91.63
ELF-SIFT	20.44
ELF-SIFT51	28.71
SOSNet	53.38
BRISK	77.80

Lessons learned

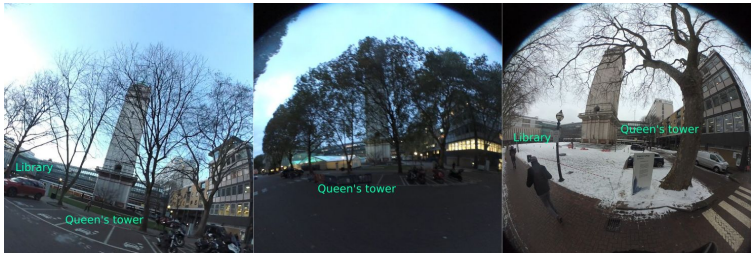
- Matching scores != camera pose
- Stereo ~= Multi-view
  - Still important (e.g. SIFT vs D2-Net)
- Matching seems very important
  - Even for bundle adjustment!

Future improvements

- Dense baselines (and submissions)
- Open-source everything
- Validation set for fast iteration
- More visualizations (matches, poses, point clouds)
- More tasks
  - Patch matching: is it a good proxy? How can we make it better?
  - Re-localization: is it feasible/desirable? (large-scale, requires finer tuning, etc)
  - Ground-to-aerial matching

Future improvements

- Dense baselines (and submissions)
- Open-source everything
- Validation set for fast iteration
- More visualizations (matches, poses, point clouds)
- More tasks
  - Patch matching: is it a good proxy? How can we make it better?
  - Re-localization: is it feasible/desirable? (large-scale, requires finer tuning, etc)
  - Ground-to-aerial matching



Challenge in the training loop

- More images would surely give improved results compared to when using only a subset
- Our challenge gt is not perfect ground truth, but provides “at least better” poses.
- We can bootstrap to keep on improving

The progressive “challenge”

- Current “ground truth” is limited
  - Based on SIFT, RANSAC, and Bundle Adjustment
  - It is the best we can do right now
  - We see the "ground truth poses" as a "higher bound"
- We can bootstrap to keep on improving
  - Can we build *better* models with *better* features and matchers?
  - What happens with this new "ground truth"?

The evolving “challenge”

- Current “ground truth” is limited
  - Based on SIFT, RANSAC, and Bundle Adjustment

- Core assumption: “SfM with more images provide better pose”
  - We see the "ground truth poses" as a "higher bound"

- “Ground truth” evolves along features and matchers

Closing ceremony

Winner of the Phototourism challenge:

- **Stereo track:** Dawei Sun, Zixin Luo, Jiahui Zhang
- **Multi-view track:** Dawei Sun, Zixin Luo, Jiahui Zhang

PRIZES SPONSORED BY:

Closing ceremony

Winners of the SILDa challenge:

Milan Pultar, Dmytro Mishkin, Jiří Matas, Visual Recognition Group, Dept. of Cybernetics, Faculty of Electrical Engineering, CTU in Prague

PRIZE SPONSORED BY:

SCAPE