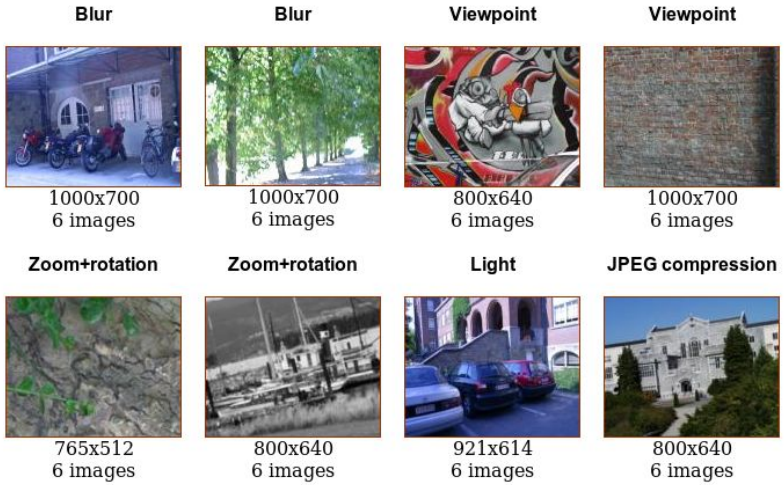
Evaluating Local Feature Descriptors

Vassileios Balntas - Imperial College London
Karel Lenc - University of Oxford

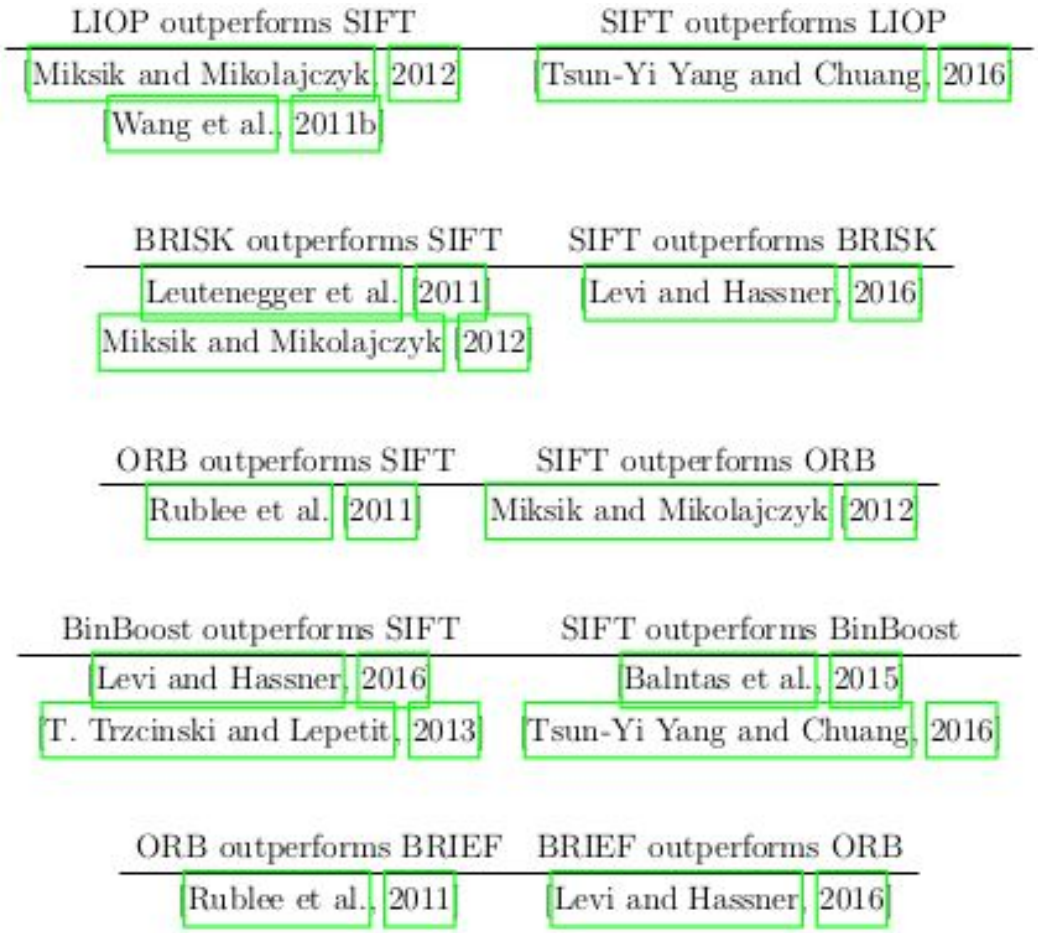
Image matching

- Measures descriptor performance in image matching task (NN matching, one2one)
- Oxford matching protocol [1]
  - precision - recall of first-nearest-neighbour matcher of descriptors from a single image pair using L2 distance metric



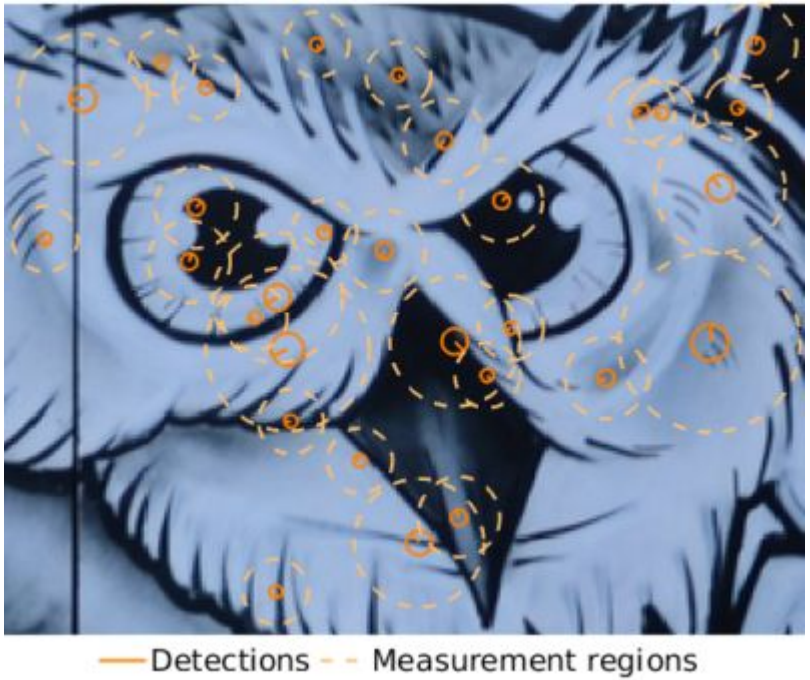
[1] Mikolajczyk, Krystian, and Cordelia Schmid. "A performance evaluation of local descriptors." *TPAMI 2005*

Inconsistency - Oxford dataset



Inconsistency in evaluation results - Oxford dataset

- no strict protocol for patch extraction and normalisation
- no strict protocol for detector configuration
- no standardised measurement region



Effect of measurement region magnification ρ

Leuven - SIFT

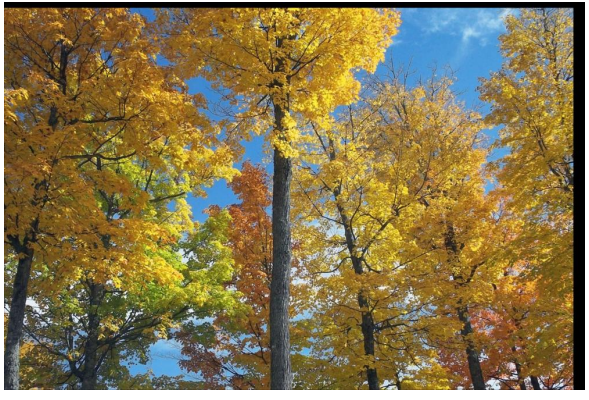
ρ	1|2	1|3	1|4	1|5	1|6
1	0.31	0.13	0.05	0.03	0.01
2	0.46	0.23	0.09	0.06	0.04
4	0.68	0.44	0.24	0.15	0.11
8	0.74	0.57	0.43	0.32	0.24
12	0.80	0.67	0.54	0.42	0.35
20	0.87	0.77	0.69	0.55	0.50

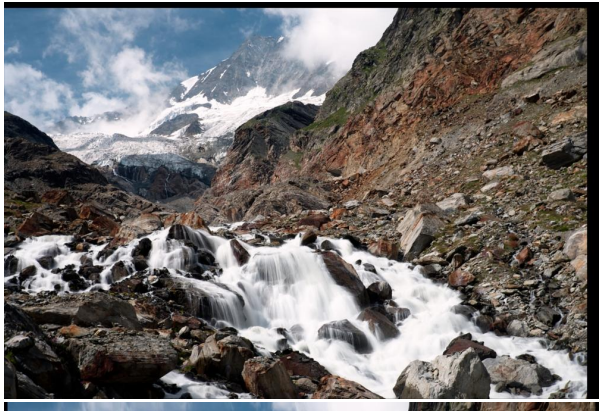
Implementation method

descr	1|2	1|3	1|4
SIFT `vl_sift`	0.47	0.40	**0.46**
SIFT `vl_covdet`	0.32	0.14	0.18

method	**paper**
`vl_sift`	***ASV*** [CVPR 2016], ***DSP-SIFT*** [CVPR 2015]
`vl_covdet`	***BinBoost*** [PAMI 2015], ***BOLD*** [CVPR 2015]

Synthesised matching dataset - Freiburg

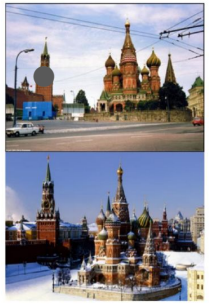




P. Fischer, A. Dosovitskiy, and T. Brox. Descriptor matching with convolutional neural networks: a comparison to SIFT ArXiv 2014

WxBS: Wide multiple Baseline Stereo - CMP

The most challenging image matching problems arise when following differences appear together: G - geometry, L - iLlumination, S - Sensor, A - Appearance,



a) WGABS (5 pairs)

b) WGSBS (5 pairs) *

c) WLABS (4 pairs)

d) WGLBS (9 pairs)

e) WGALBS (8 pairs)

*WGSBS contains image pairs of thermal camera vs visible

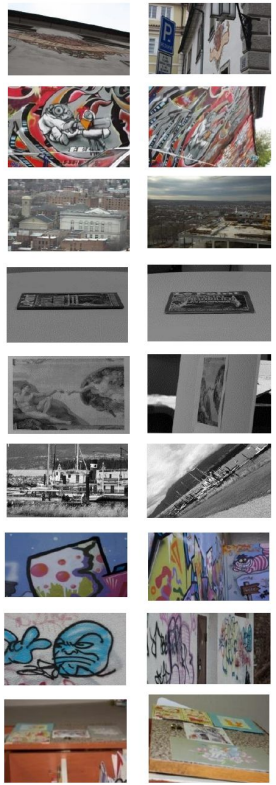
Dataset available at:
http://cmp.felk.cvut.cz/wbs/

Mishkin et.al., WxBS: Wide Baseline Stereo Generalizations, BMVC 2015

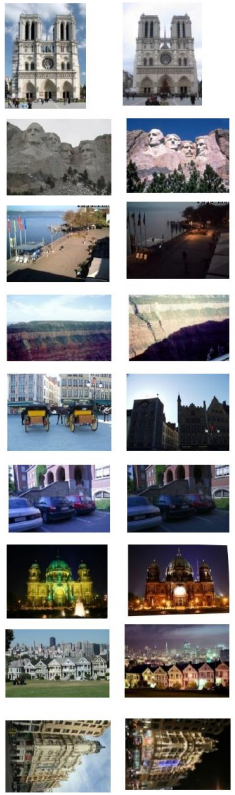
W1BS datasets - CMP

compilation of challenging matching problems from existing datasets.

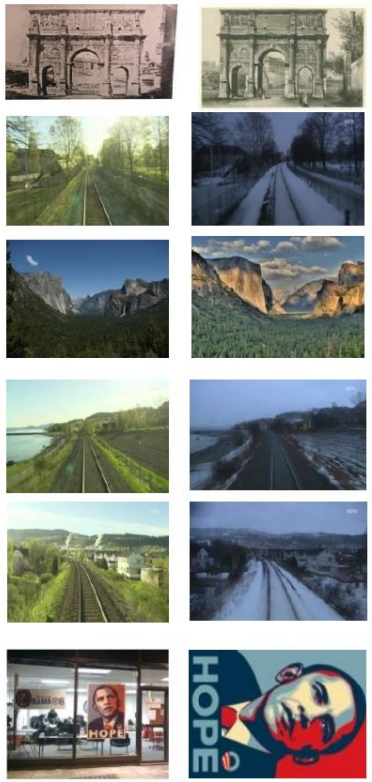
Geometry



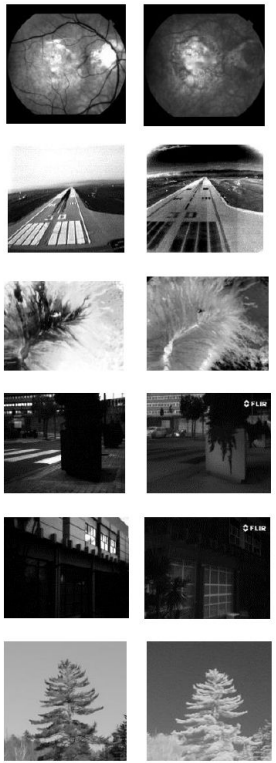
Illumination



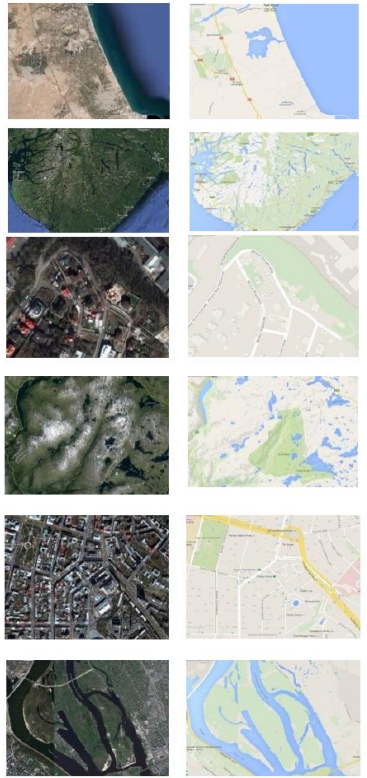
Appearance



Sensor



Map vs photo



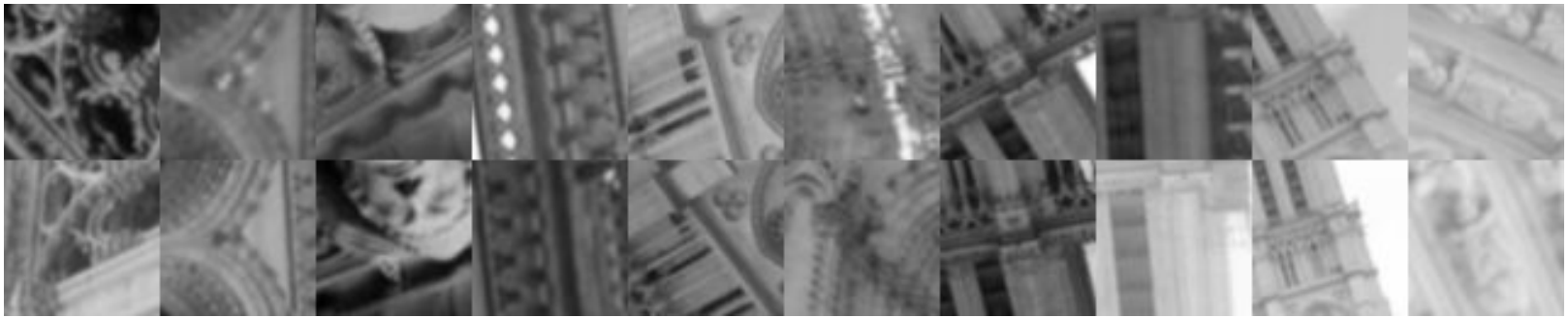
Mishkin et.al., WxBS: Wide Baseline Stereo Generalizations, BMVC 2015

From images to patches

sequence based

patch based

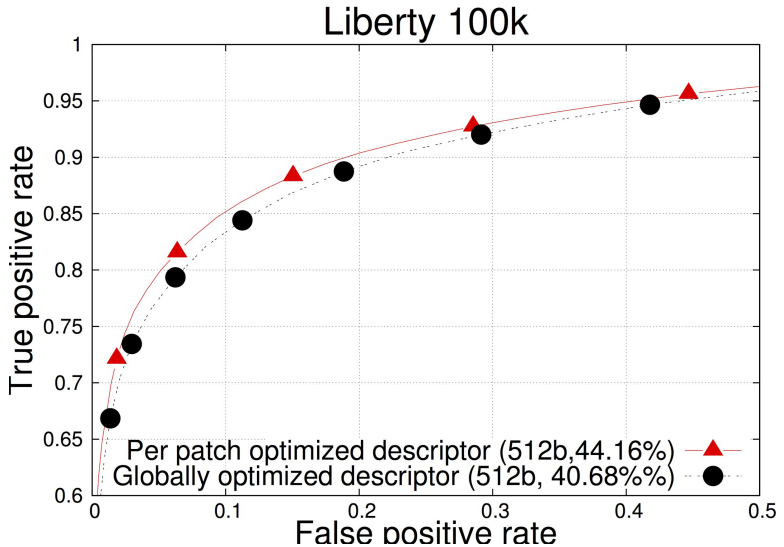
Patch datasets - Photo Tourism



S. Winder, G. Hua and M. Brown
Picking the Best DAISY
CVPR 2009

Patch pair classification

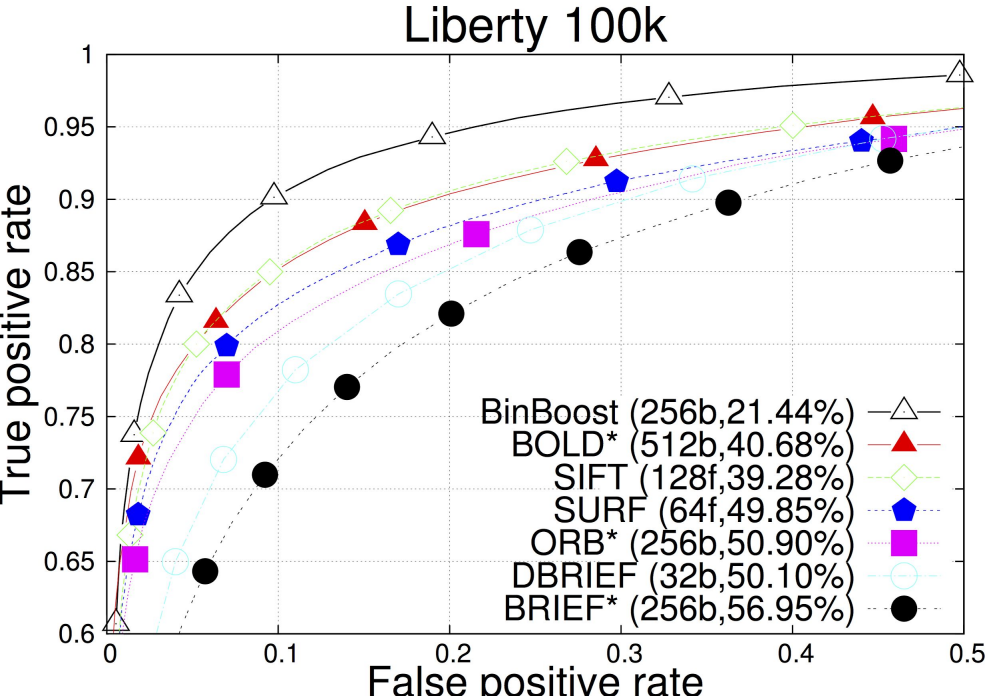
- Measures descriptor as classifier of pos and neg pairs using ROC and PR
  - Protocol similar to [1]



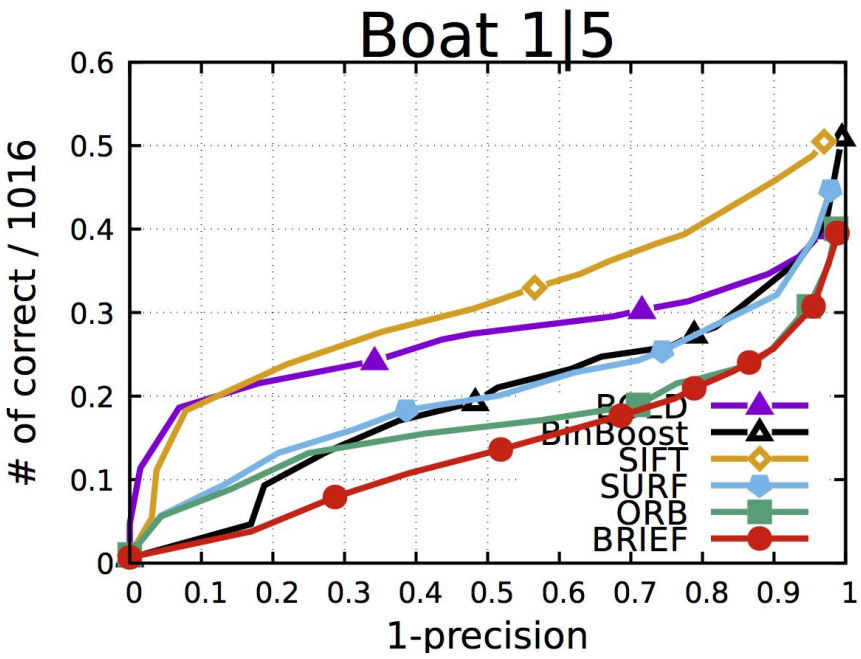
[1] S. Winder, G. Hua and M. Brown "**Picking the Best DAISY"** *CVPR 2009*

Evaluation metrics

ROC - Classification



PR - Matching



Ideal properties of a **descriptor** benchmarking method

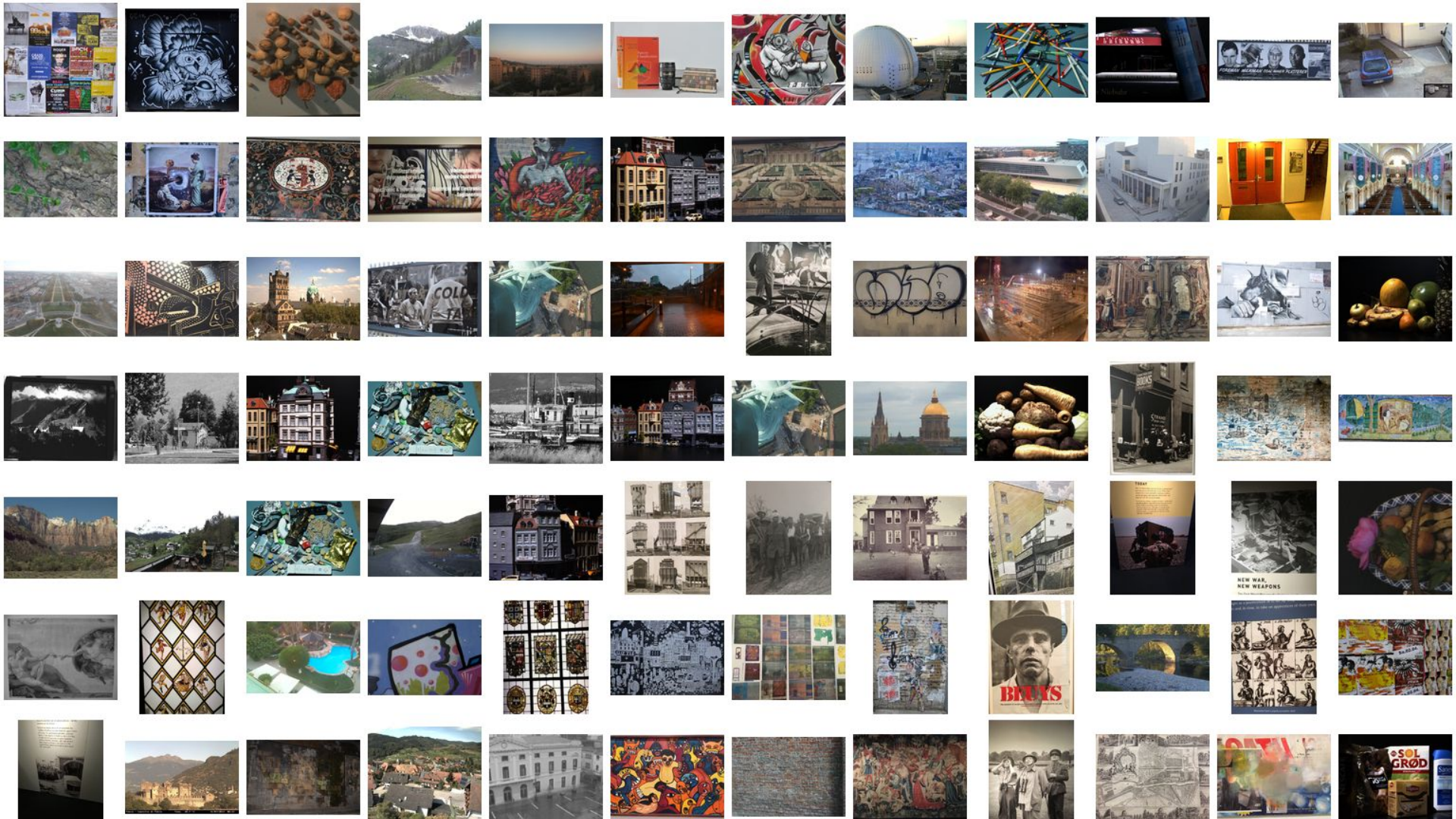
- Fixed patches
- Diverse data
- Real-world captured
- Reproducible evaluation
- Large-scale
- Support for multiple evaluation metrics

Datasets - comparison

dataset	diverse	real-world captured	large-scale	multiple ev. metrics	reproducible
Photo-Tourism		✔	✔		✔
DTU		✔	✔		
Oxford	✔	✔			
Fischer	✔				
CDVS			✔		✔
Edge Foci		✔			
Rome Patches		✔			✔
W1BS	✔	✔			
HPatch-HBench	✔	✔	✔	✔	✔

HSequences

- 116 sequences
  - 6 images per sequence
  - All images from a sequence related with **H**omography
  - Extension of the traditional Oxford Affine Dataset
- Two deformation classes
  - illumination change
  - viewpoint change (all planar scenes)


COLD
NEW WAR,
NEW WEAPONS
BEUYS
SOL
GRØD

Descriptor Baselines

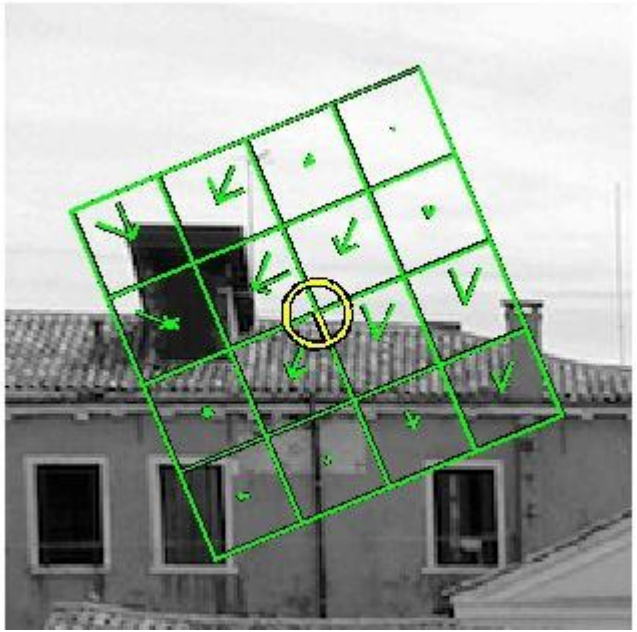
Trivial baselines

- **meanstd**: concatenate patch μ and σ (2-dim)

- **resize**: resize patch to 4x4 and and normalise to zero-mean-unit-variance (16-dim)

SIFT - [Lowe]

- rectangular grid
- pooling



vl_sift (vlfeat)

RootSIFT [Arandejelovic and Zisserman]

- Hellinger or chi-squared measures outperform Euclidean distance when comparing histograms.
- SIFT is a histogram
- Improve SIFT by altering the distance measure

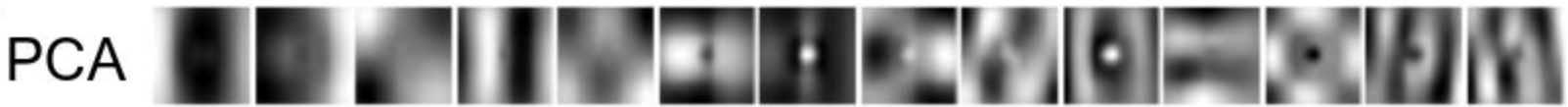
Conceptually one line of Matlab code to convert SIFT to RootSIFT

```
rootsift = sqrt( sift / sum(sift) );
```

Provided by Giorgos Tolias, CMP, Prague

RootSIFT-pca

- rootSIFT dimensionality reduction with PCA to 80D vector
- projections are learned from a large set of samples
- square rooting - powerlaw with exponent 0.5



Bursuc, Andrei, Giorgos Tolias, and Hervé Jégou. "Kernel local descriptors with implicit rotation matching." ACM, 2015.

Provided by Giorgos Tolias, CMP, Prague

Deep Learning Baselines

- deepcompare-siam
- deepcomapre-siam2stream
- deepdesc
- tfeat-margin
- tfeat-margin-star
- tfeat-ratio
- tfeat-ratio-star

trained on PhotoTourism patches

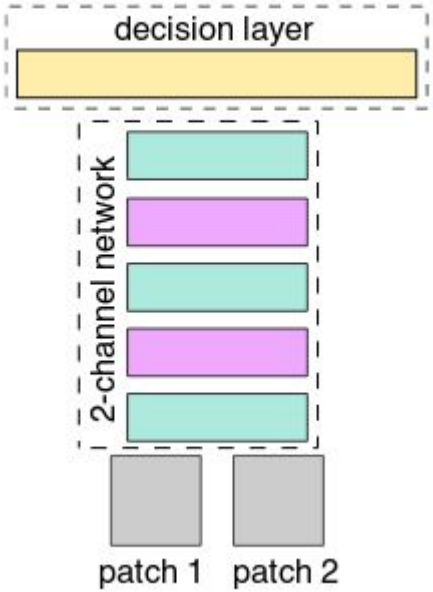
Code for extracting descriptors from bseveral aselines soon available at

`https://github.com/featw/hdescriptors`

DeepCompare-siam [Zagoruyko & Komodakis]

siamese network - contrastive loss

$$l(\mathbf{x}_1, \mathbf{x}_2) = \begin{cases} \|D(\mathbf{x}_1) - D(\mathbf{x}_2)\|_2, & p_1 = p_2 \\ \max(0, C - \|D(\mathbf{x}_1) - D(\mathbf{x}_2)\|_2), & p_1 \neq p_2 \end{cases}$$

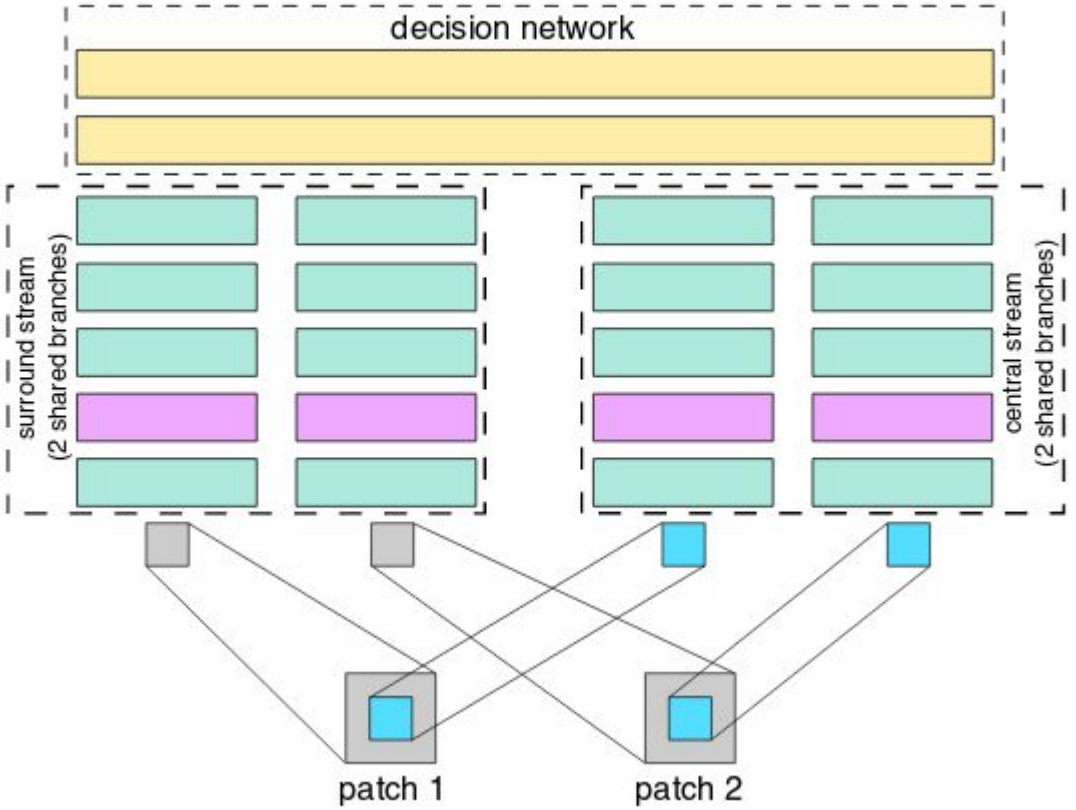


https://github.com/szagoruyko/cvpr15deepcompare

CVPR 2015

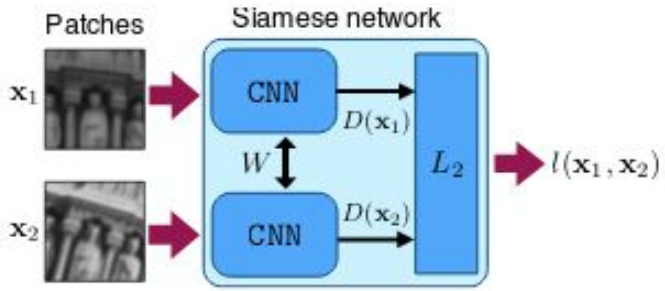
DeepCompare-siam 2stream

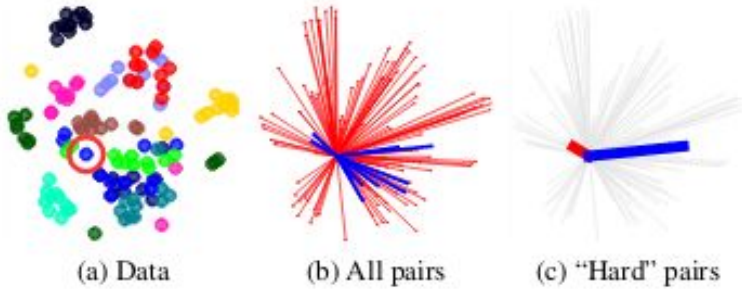
- two scales from each patch
- multiscale has advantage (DSP-SIFT)
- 2*dims as single scale



DeepDesc [Simo-Sera, Trulls et. al.]

siamese - contrastive loss - in-batch hard negative mining

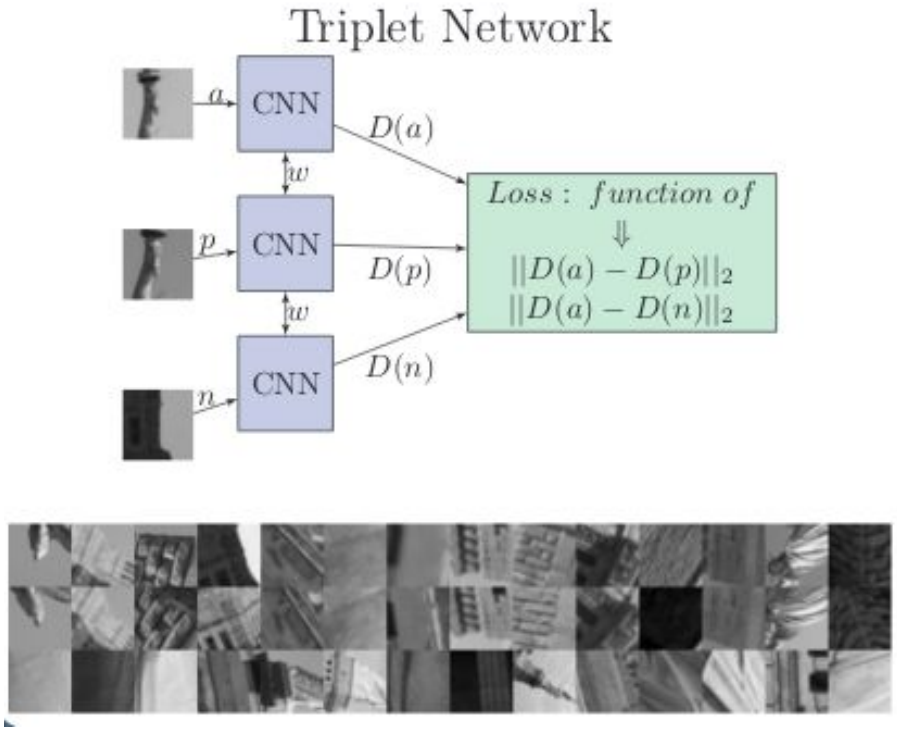


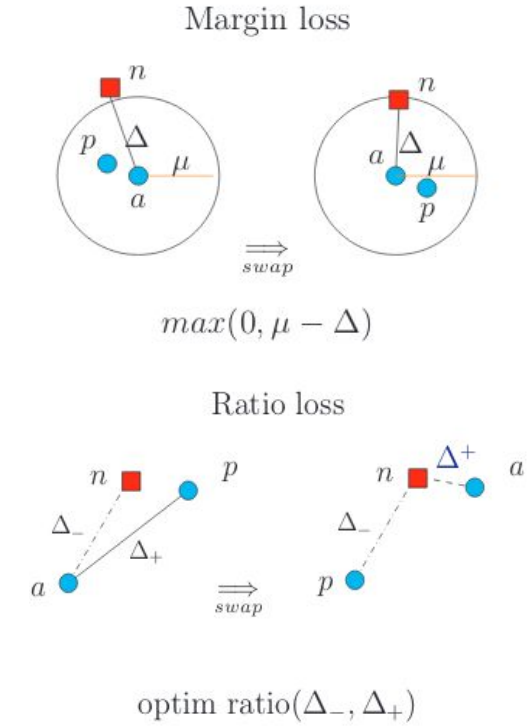


https://github.com/etrulls/deepdesc-release

arxiv 2015, ICCV 2015

TFeat / pnnet [Balntas et.al]





https://github.com/vbalnt/tfeat

arxiv 2016, BMVC 2016

HPatches

Homography Patches

https://github.com/featw/hpatches

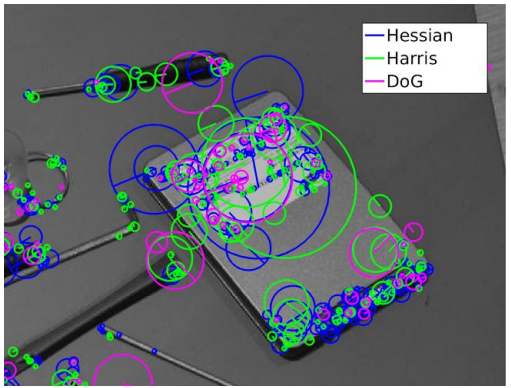
HPatches Dataset

- 6 patches per local feature cluster
- For the purpose of the challenge, a random draw of **76 train and 40 test**
- No test set obfuscation, test set released 2 weeks before submission deadline
- **Protocol after this workshop - cross validation, no single train/test split**
    - **Test labels will be released**

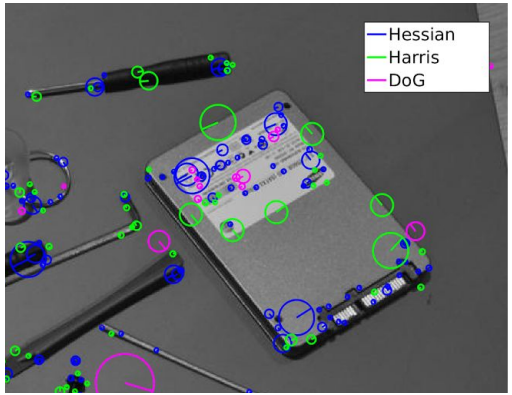
HPatches - patch normalisation process

- Local features detected in the reference image using Hessian, Harris and DoG detector
- In order to prevent duplicate regions, a random subset of patches selected s.t. max overlap is 50%
- At most 3000 feature per image selected

Detections



After NMS



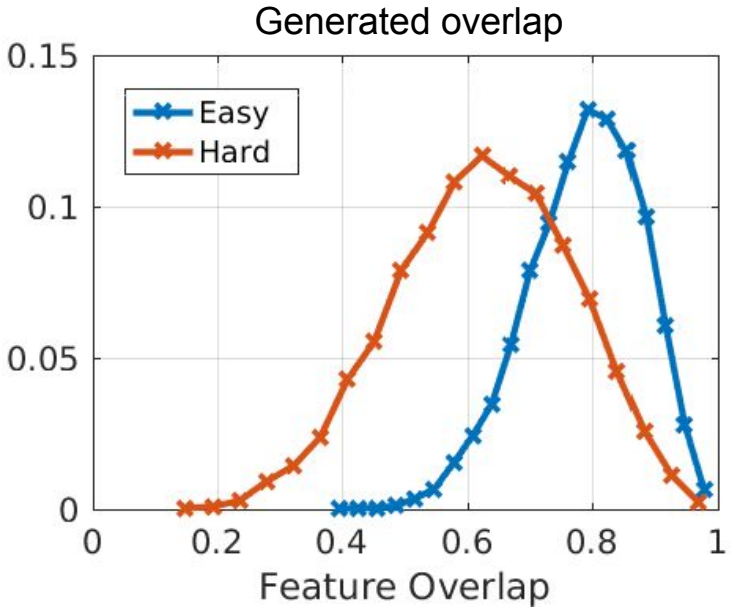
Detection Error

- Using detections from multiple detectors
- What if feature is not re-detected in transformed image?
- What affine adaptation algorithm should we use?

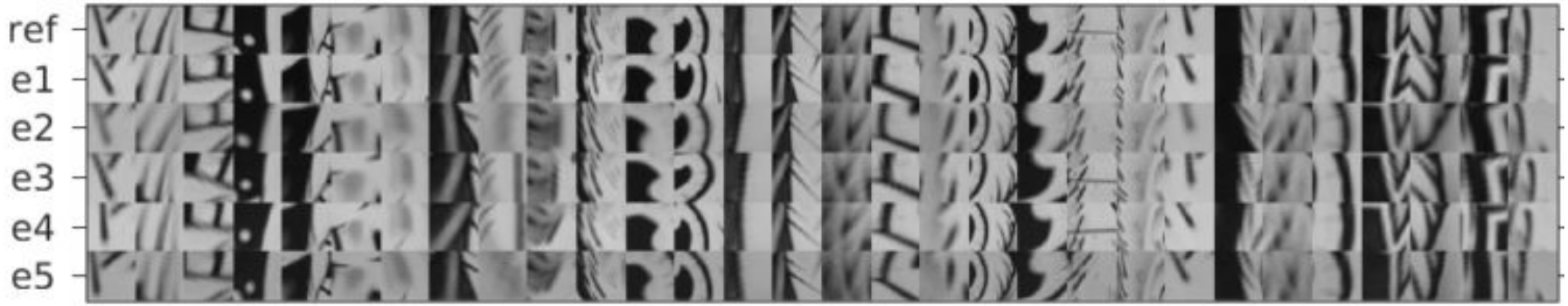
- => Controlled environment - Random geometry noise

HPatches - ‘easy’ and ‘hard’

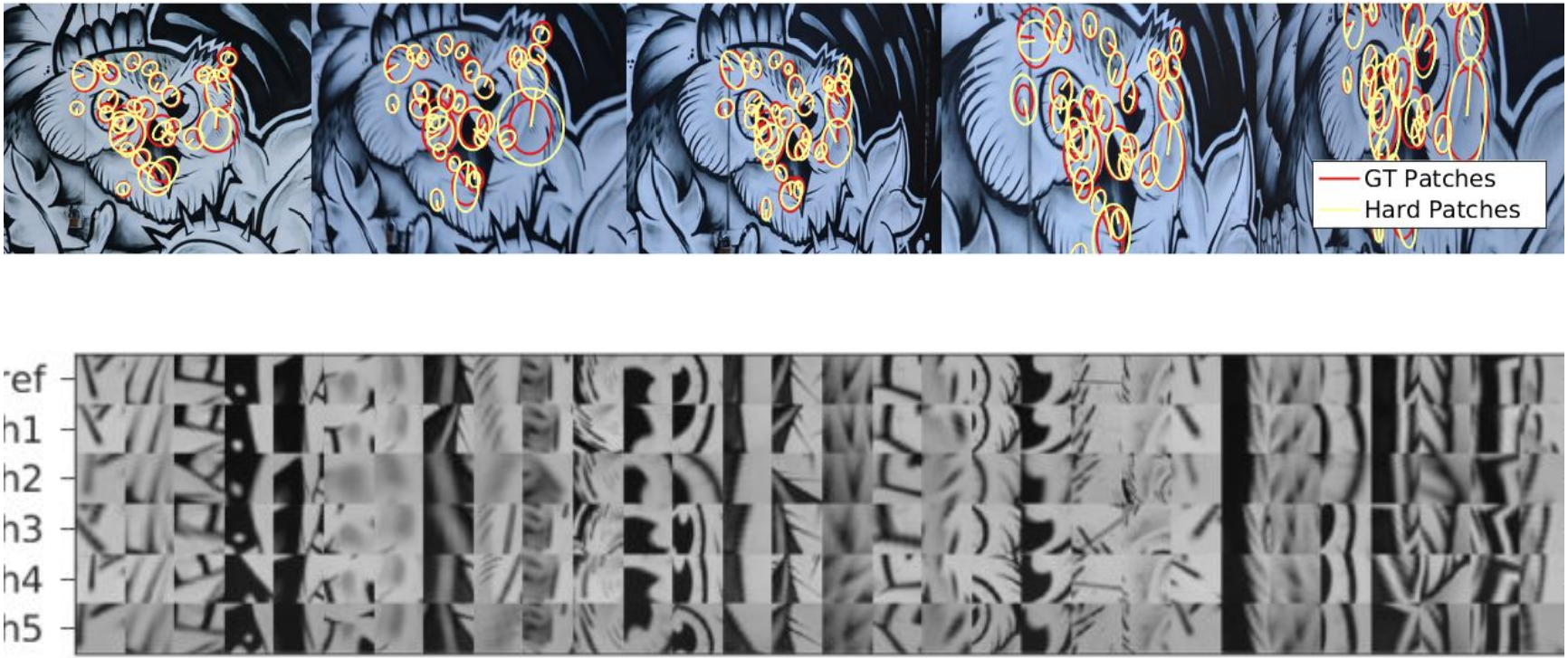
- Reference image features **reprojected with GT Homography** with additional geometric noise (translation, scale, anisotropy, rotation)
- Two distributions of the geometric noise - **easy** and **hard**
  - Allows to study geometry invariance of the descriptor
- Patches of size 65 x 65 px



Easy



Hard



Hpatches - Dataset statistics

	Train	Test
Num. of sequences	76	40
Num. of patches	581 706	353 256
AVG Patches / Image	1200	1400

HBench

Reproducible evaluation protocols for HPatches

https://github.com/featw/hbench

HBench - evaluation metrics

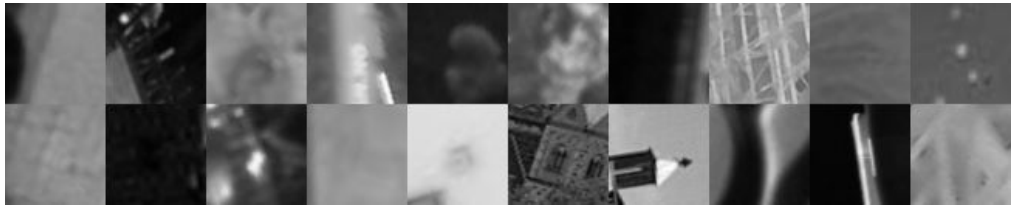
- Patch pair classification
  - Casts description to a classification problem, simple learning formulation
- Image matching
  - Evaluates the descriptor in first-nearest-neighbour image matching
- Patch retrieval
  - Measures discriminability of a patch descriptor in a large corpus

Patch pair classification

- Control of negative pairs
  - same sequence
  - different sequence
- Control of positive pairs
  - Easy detector noise
  - Hard detector noise
- => 4 Sub Tasks

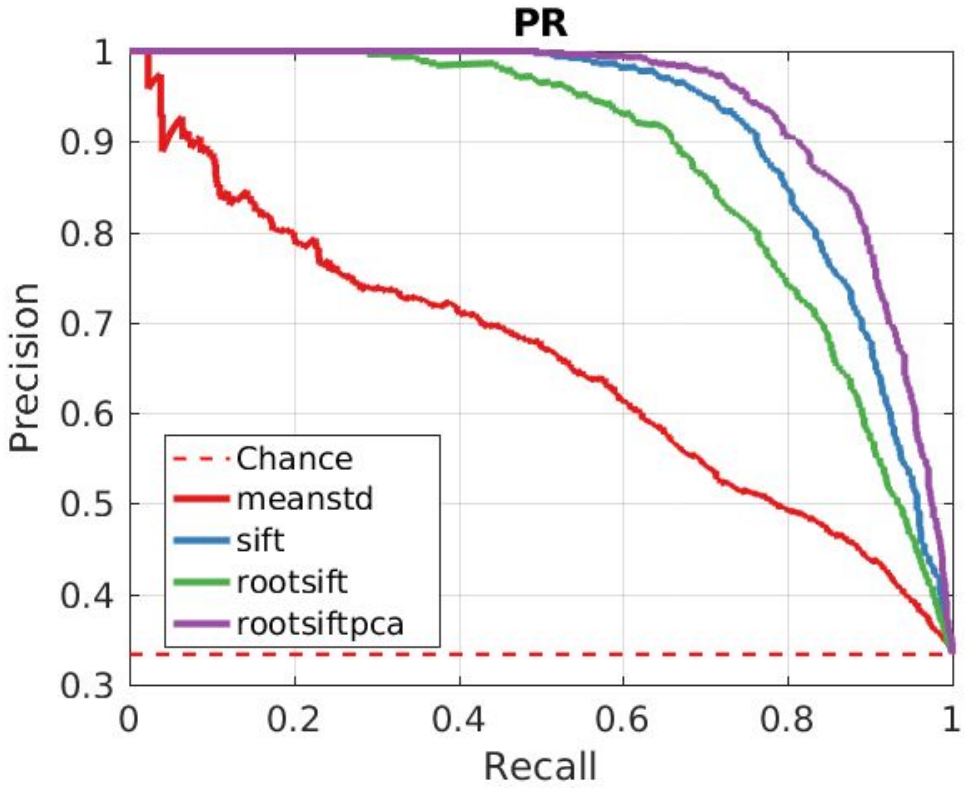
Example of Positive Pairs

Example of Negative Pairs



Patch pair classification - Overall Scores

- Distance between patches computed using L2 distance
- Final performance measured as **Average Precision**



Patch Classification Variants

- 100k Pos Pairs, 500k Neg Pairs for the test set

Detector Noise

	Easy	**Hard**
SameSeq-Negs	`easy_sameseq`	`hard_sameseq`
DiffSeq-Negs	`easy_diffseq`	`hard_diffseq`

Final score - Average AP of all 4 variants

Patch classification - baselines

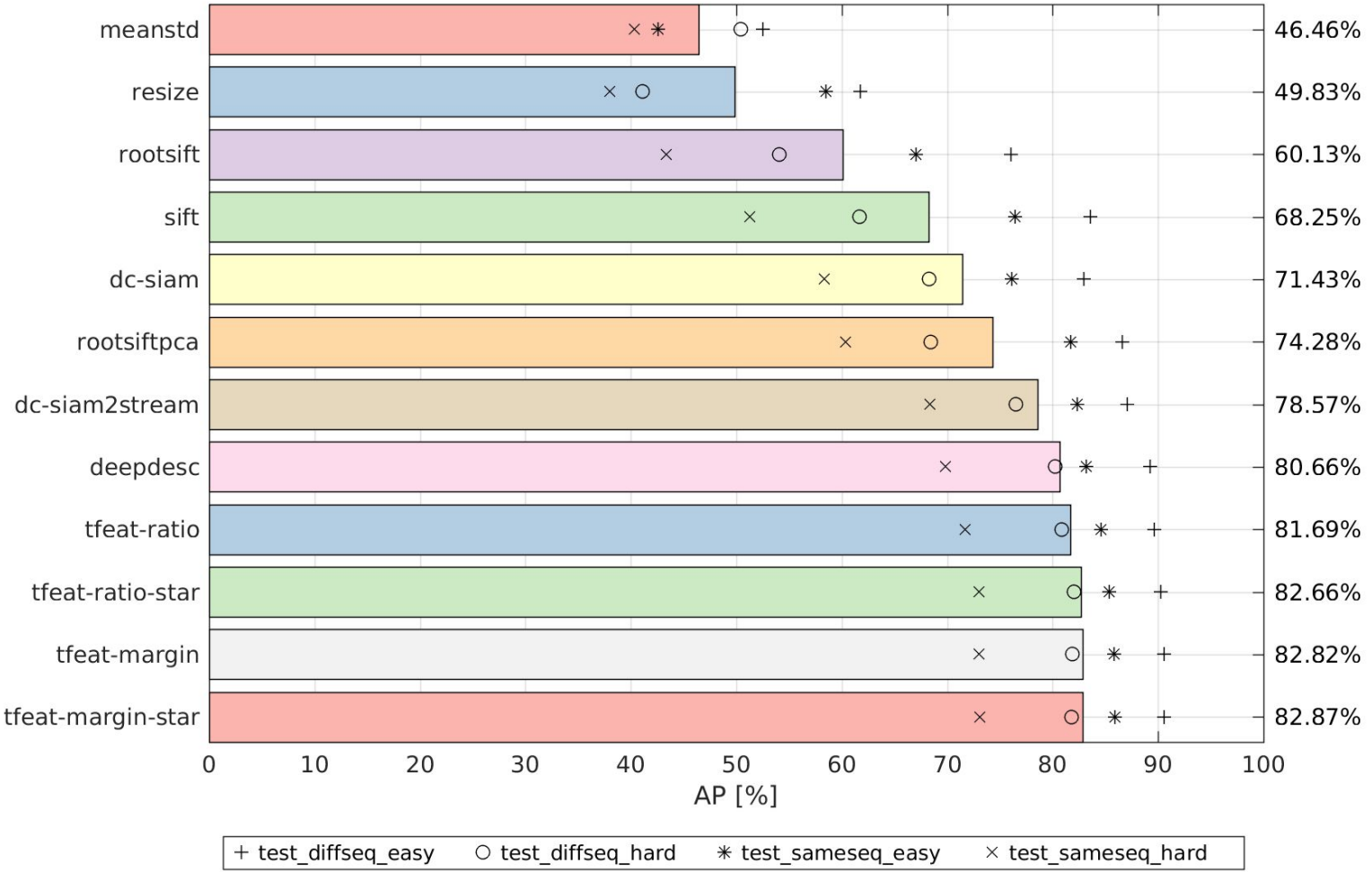
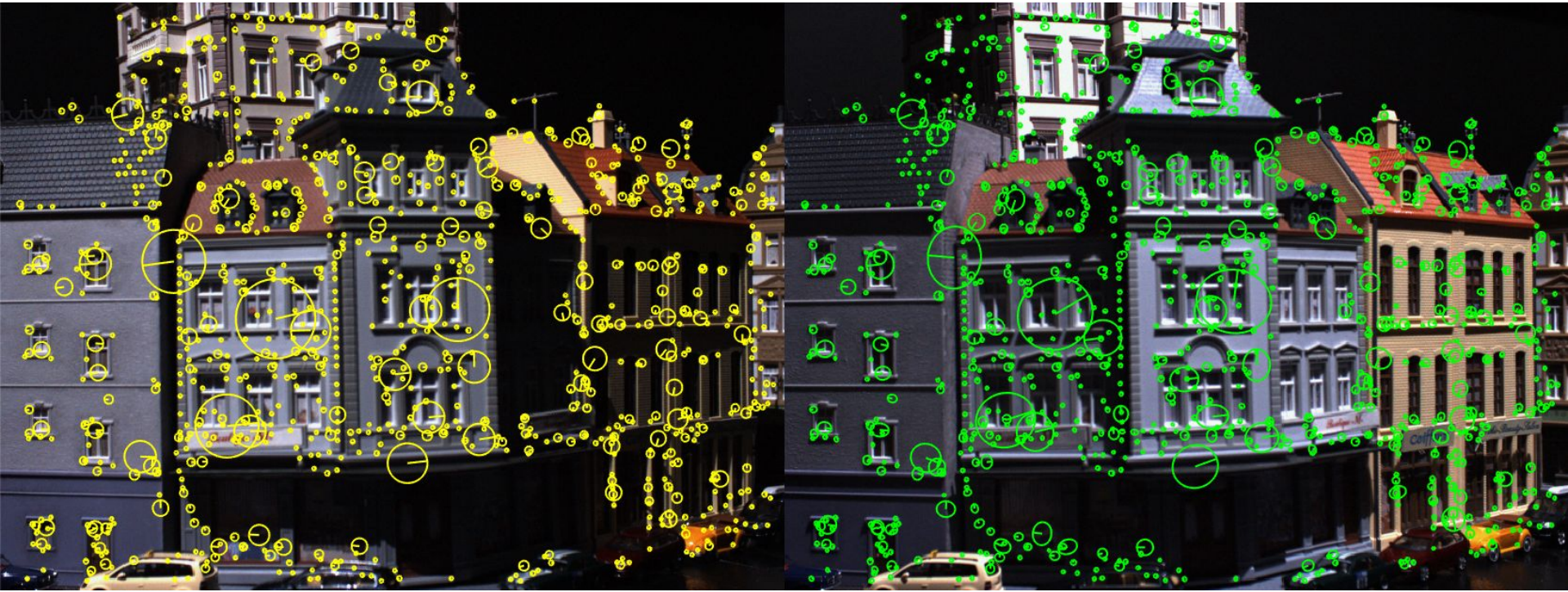


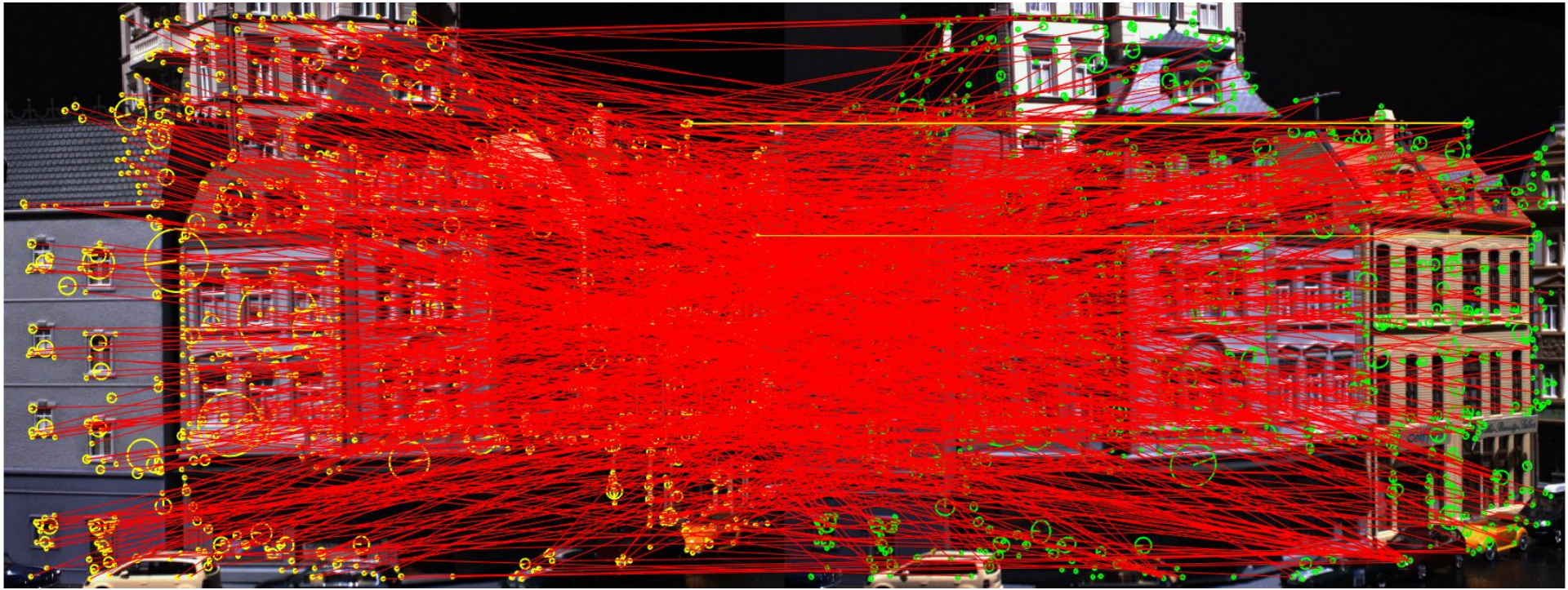
Image matching

- Measures descriptor performance in image matching task (NN matching, one2one)
- Oxford matching protocol [1]
  - precision - recall of first-nearest-neighbour matcher of descriptors from a single image pair using L2 distance metric
  - Due to ground truth reprojection, number of geometric correspondences is equal to number of matches (-> 100% recall)
  - => Performance measured as mean Average precision over all matching tasks
- 4 Variants
  - Illum / Viewpoints sequences
  - Easy / Hard Detector noise

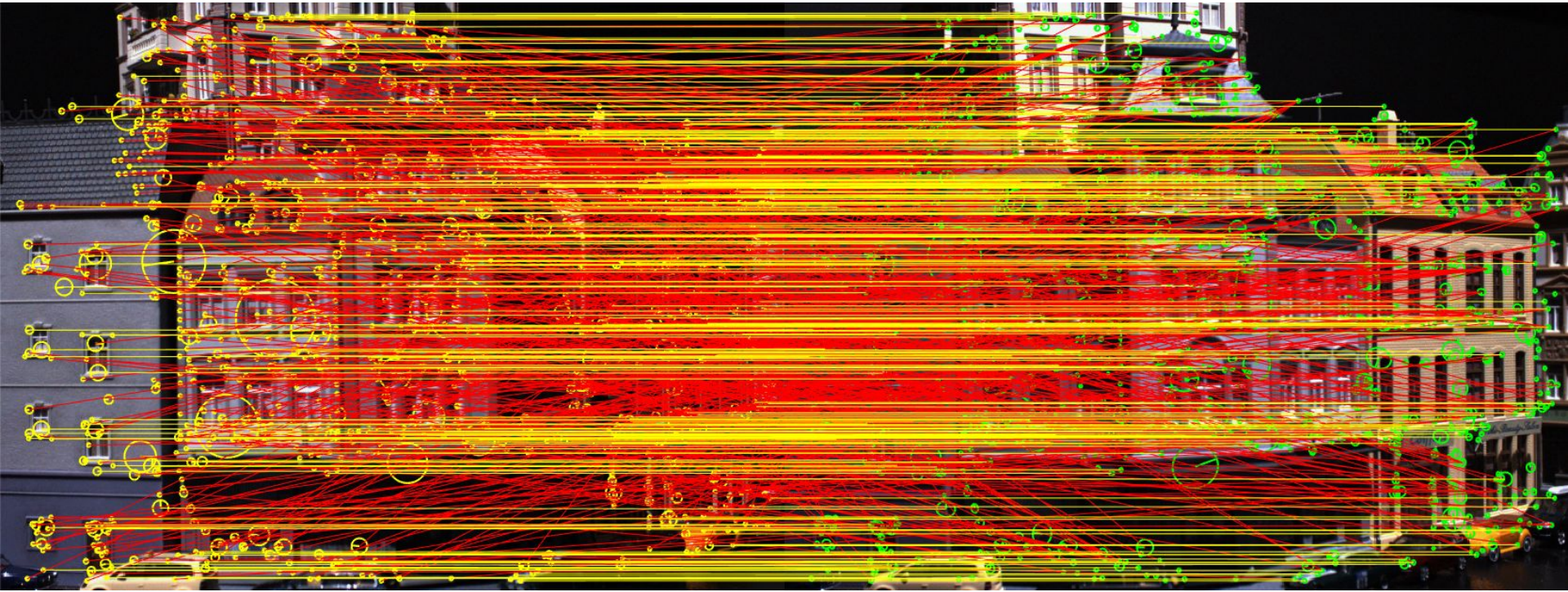
[1] Mikolajczyk, Krystian, and Cordelia Schmid. "A performance evaluation of local descriptors." *TPAMI 2005*



Meanstd



SIFT



RootSIFT-PCA

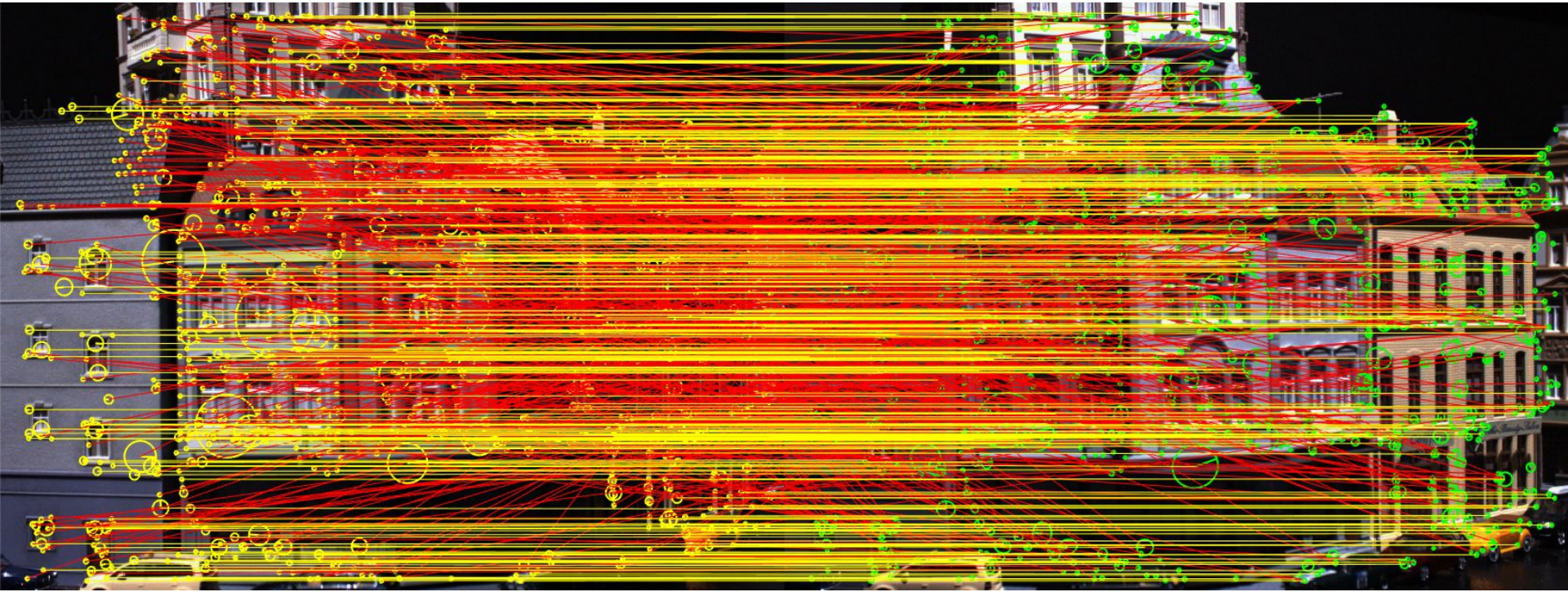


Image matching - Overall Scores

- Final performance measured as an mean Average Precision over multiple image pairs

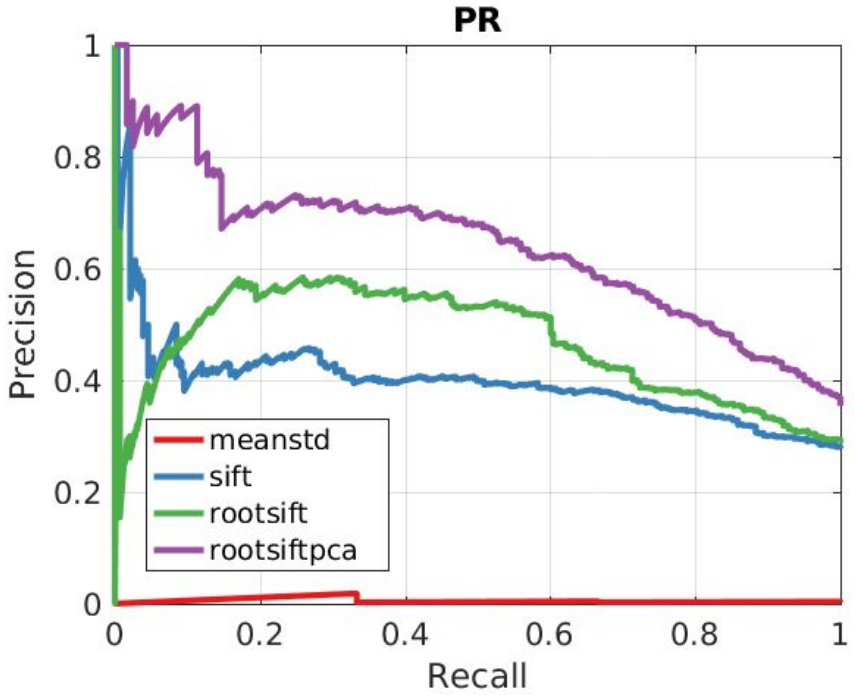


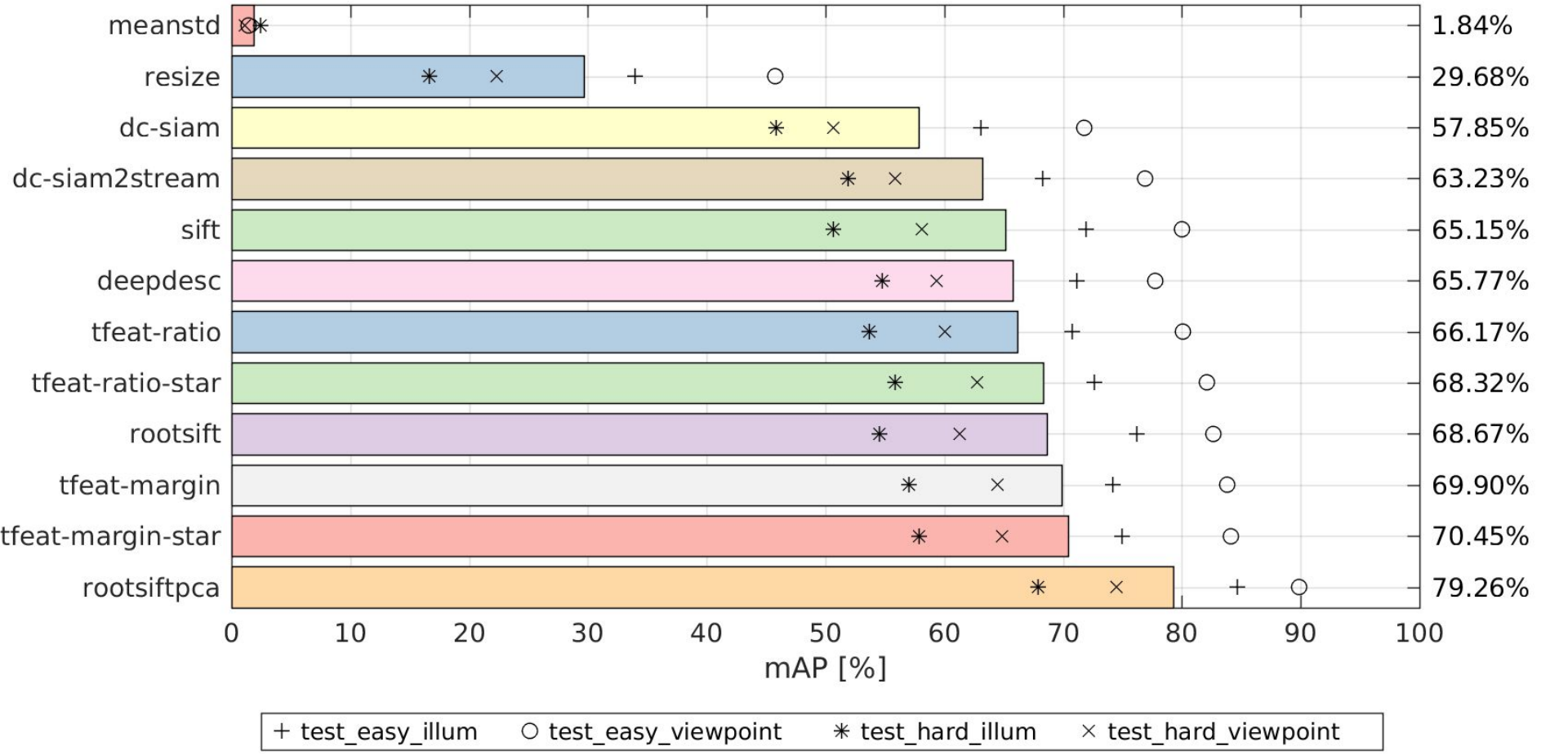
Image Matching Variants

Detector Noise

	Easy	**Hard**
Illumination (100 tasks)	`easy_illum`	`hard_illum`
Viewpoint (100 tasks)	`easy_viewpoint`	`hard_viewpoint`

Final score - Average AP of all 4 variants

Image matching - baselines



Patch Retrieval

- Retrieval from a large corpus of local feature descriptors, one2many task
- For each query descriptor, compute the AP for the top-50 closest features using L2 metric
  - Each query is part of a cluster of 6 descriptors
- Evaluation protocol similar to [1] and [2]
- 4 Variants
  - Size of descriptor pool (5 sequences, 40 sequences), multiple draws
  - Detector noise (Easy / Hard)

[1] Philbin, J. , Chum, O. , Isard, M. , Sivic, J. and Zisserman, A. Object retrieval with large vocabularies and fast spatial matching, CVPR 2007
[2] M. Paulin, M. Douze, Z. Harchaoui, J. Mairal, F. Perronnin and C. Schmid: Local Convolutional Features with Unsupervised Training for Image Retrieval, ICCV 2015

Example MeanSTD

Query

#1

#2

...

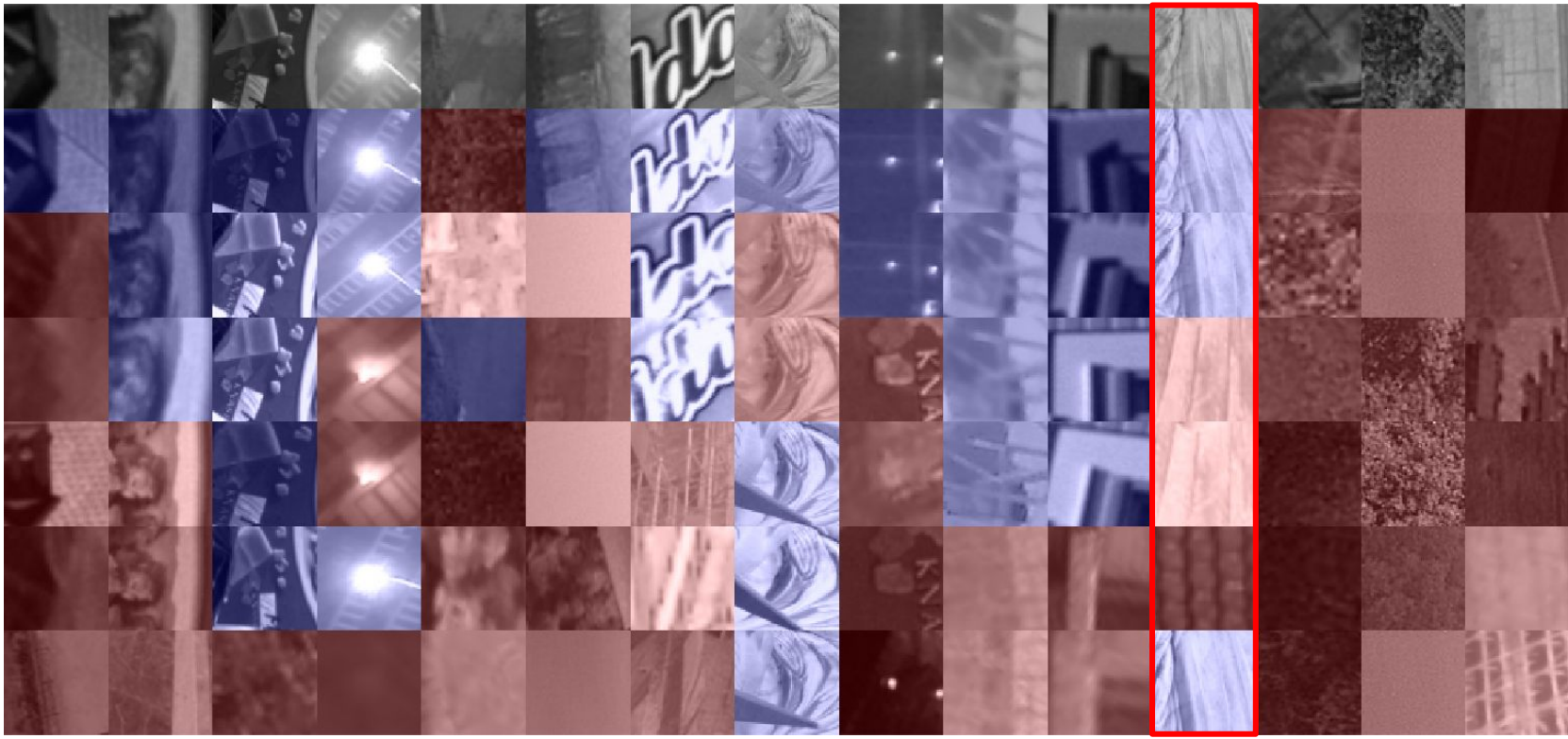
Example SIFT

Query

#1

#2

...



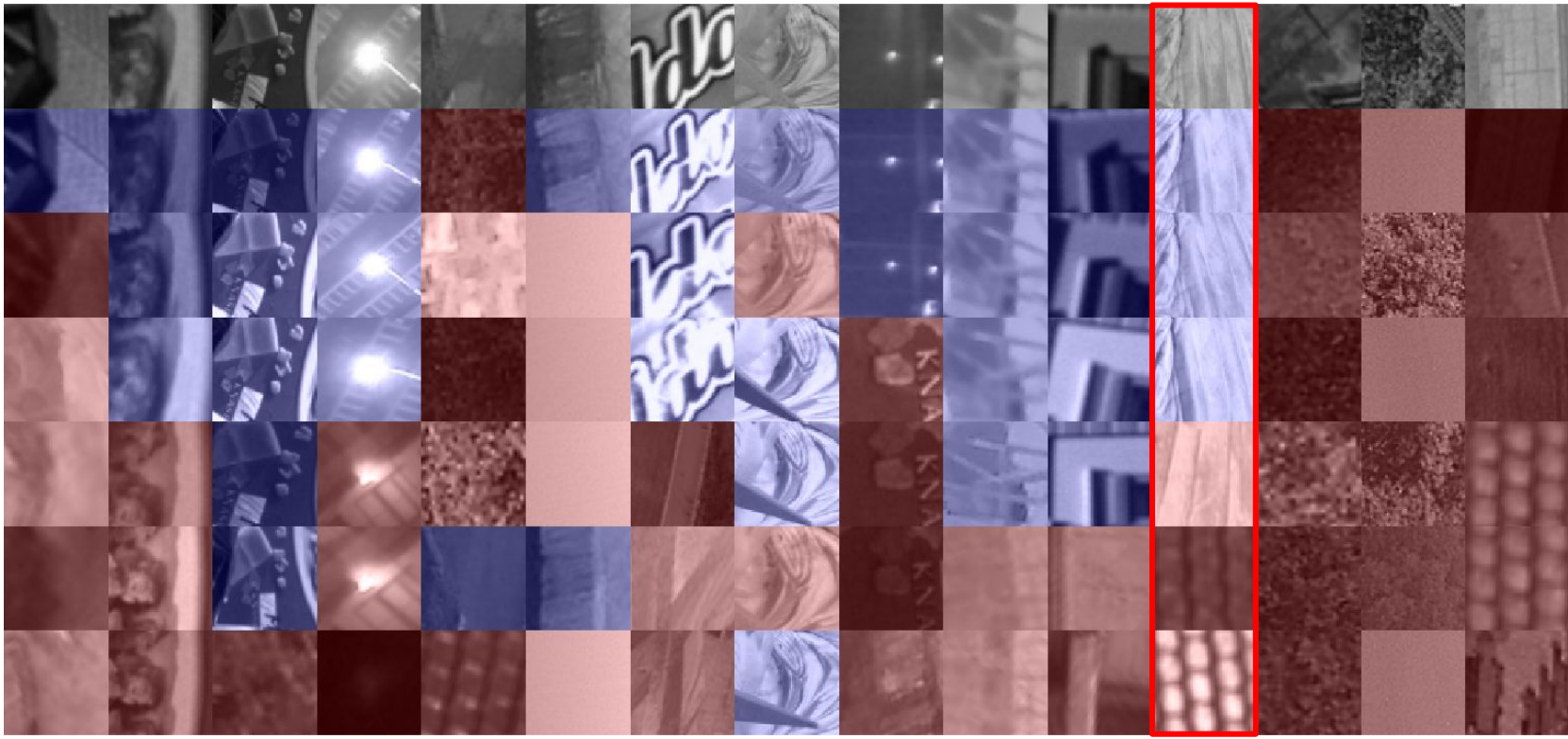
Example RootSIFT

Query

#1

#2

...



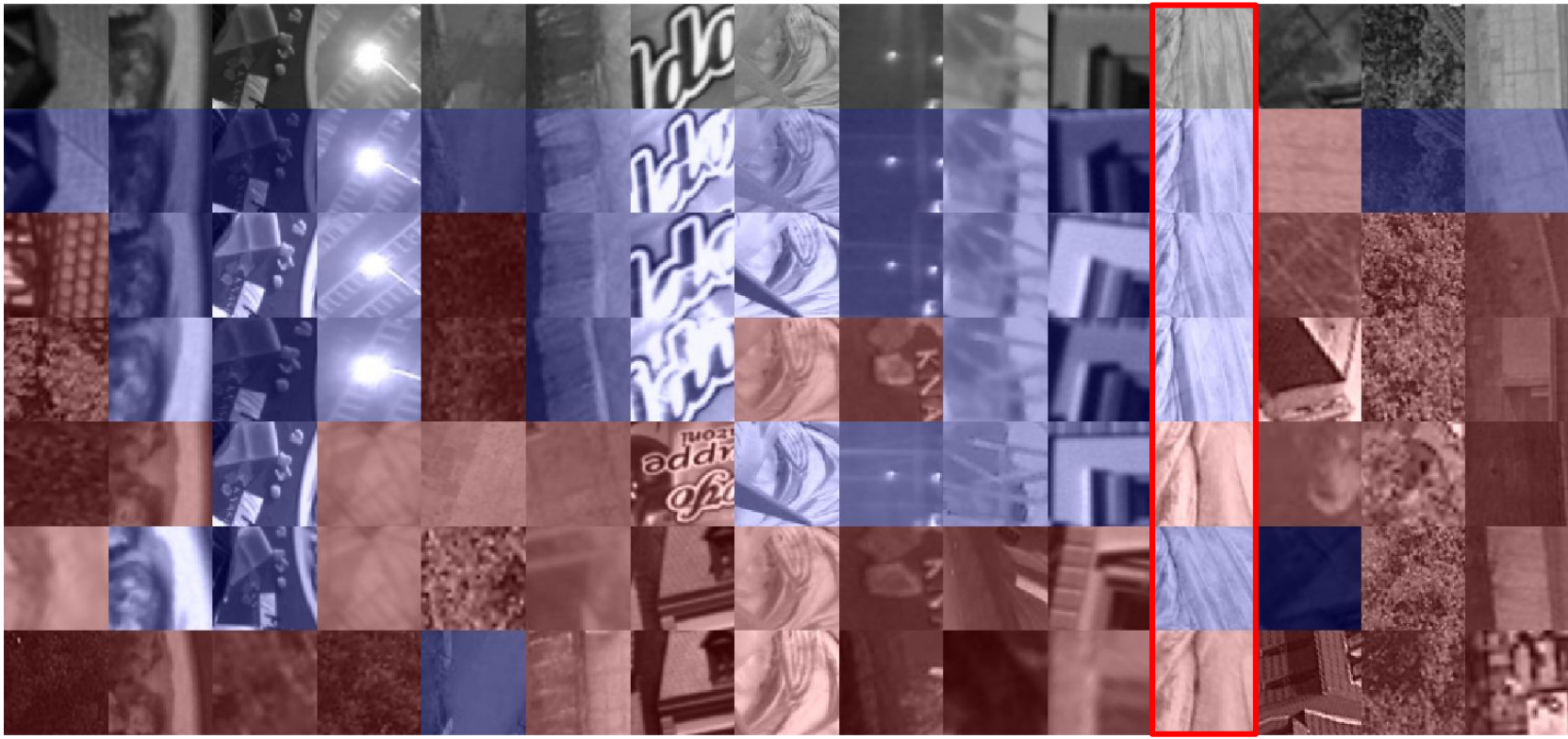
Example RootSIFT-PCA

Query

#1

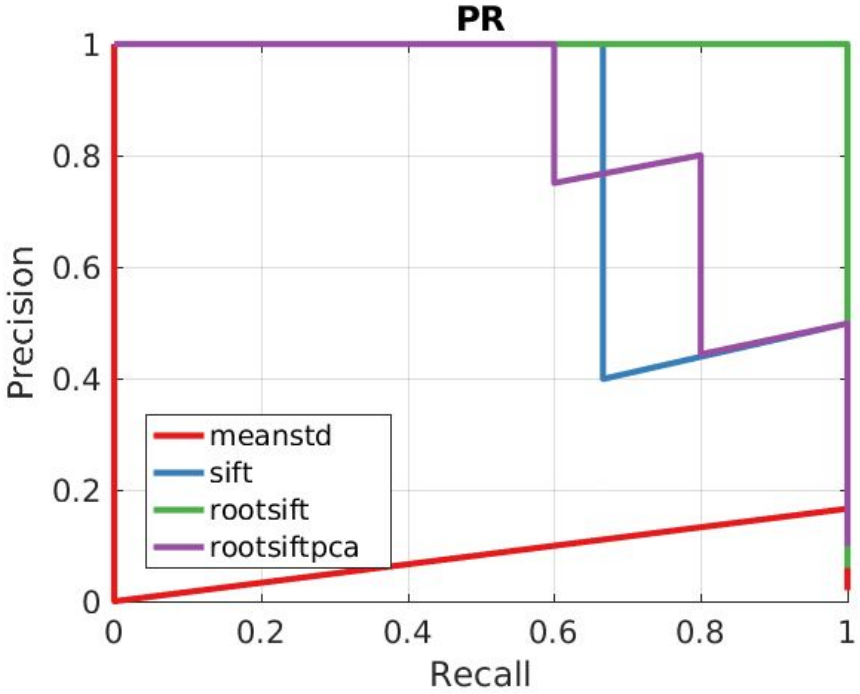
#2

...



Patch Retrieval - Overall Scores

- Final performance measured as an mean Average Precision over multiple Query descriptors
- In this case, always only 5 positives - jagged PR-curve



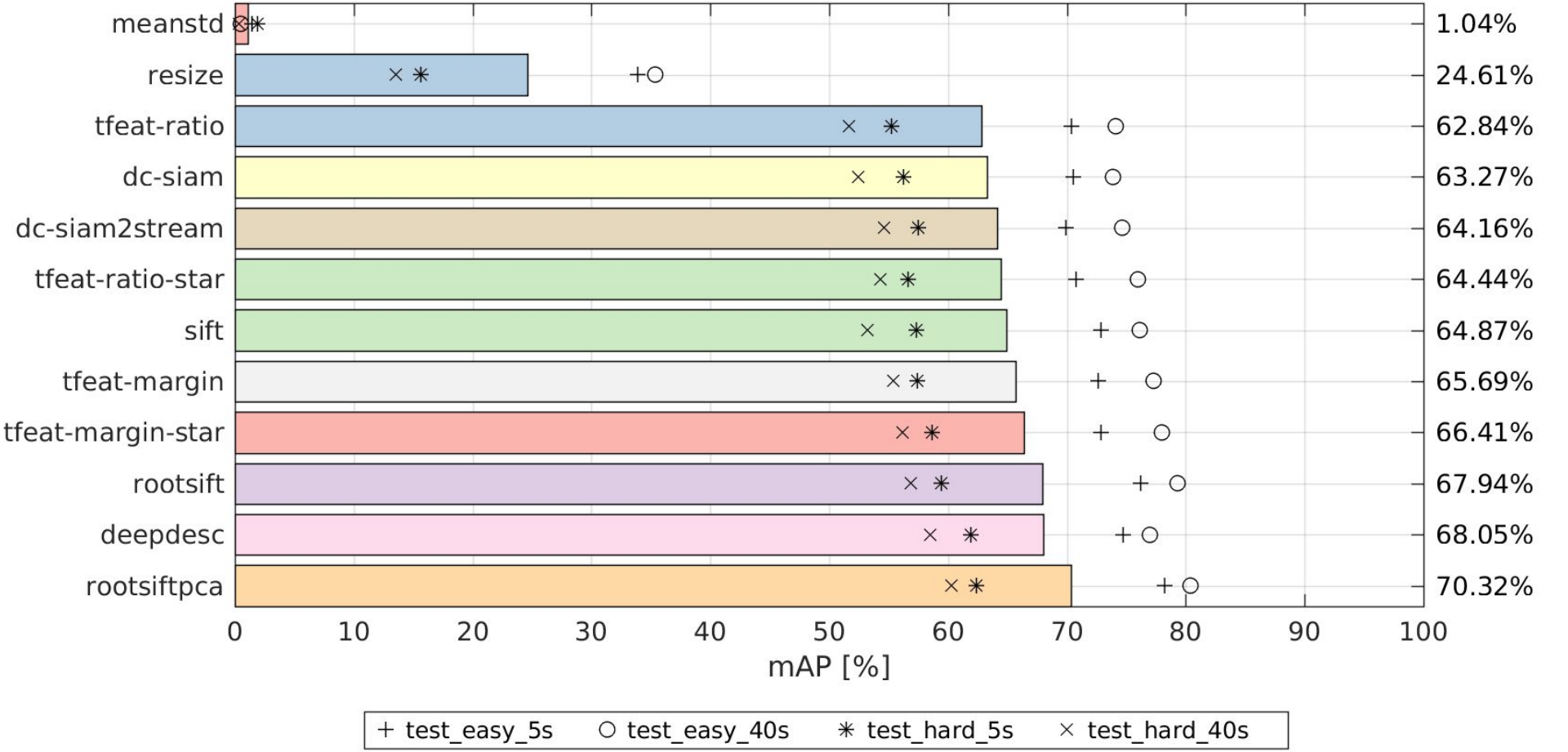
Patch Retrieval Variants

Detector Noise

	Easy	**Hard**
5 Sequences (1500 queries)	`easy_5s`	`hard_5s`
40 Sequences (4000 queries)	`easy_40s`	`hard_40s`

Final score - Average AP of all 4 variants

Patch retrieval - baselines



Cross-task ranking comparison - baselines

Rank	Classification	Matching	Retrieval - Patch
1	TFeat	RootSIFT-PCA	RootSIFT-PCA
2	DeepDesc	TFeat	DeepDesc
3	DeepCompare	RootSIFT	RootSIFT
4	RootSIFT-PCA	DeepDesc	TFeat
5	SIFT	SIFT	SIFT

Competition Results

Competing Submissions

- **zagreb-nm** - Nenad Markuš, University of Zagreb, Croatia
- **cmp-dm** - Anastasiya Mishchuk (Szkocka Research Group, Ukraine), Dmytro Mishkin, Jiri Matas (CTU Prague, Czech Republic)
- **cmp-ab** - A. Bursuc, INRIA, G. Tolias,CTU Prague, Czech Rep.
- **hannover-cl** - Chen Lin, University of Hannover, Germany
- **casia-yt** - Yurun Tian, Bin Fan and Fuchao Wu, CASIA, China

cmp-dm - Anastasiya Mishchuk (Szkocka Research Group, Ukraine), Dmytro Mishkin, Jiri Matas (CTU Prague, Czech Republic)

- cmp-dm-1: sift based / no learning
- cmp-dm-2: sift based with learning
- paper under submission

casia-yt - Yurun Tian, CASIA, China

- CNN based
- paper under submission (with Bin Fan and Fuchao Wu)

hannover-cl - Chen Lin, Univ. of Hannover

Training Architecture: Siamese CNN

Loss: Double Margin Hinge ($l_{pull} = 1$, $l_{push} = 5$)

$$Loss = \sum_{i=1}^{N} \left[y_i \cdot \max\left(0, d - l_{pull}\right)^2 + \left(1 - y_i\right) \cdot \max\left(0, l_{push} - d\right)^2 \right]$$

Training Data: 1:1 match: unmatch pairs; 1.7 M training/ 0.24M validation set

Training:

-- Mini Batch (size 500);

-- Hard Mining (factor: 4), Early Stopping

-- From Scratch only with hpatches-train dataset

-- Caffe / Geforce Titan X GPU

Trained CNN Descriptor: **128** Dimensional

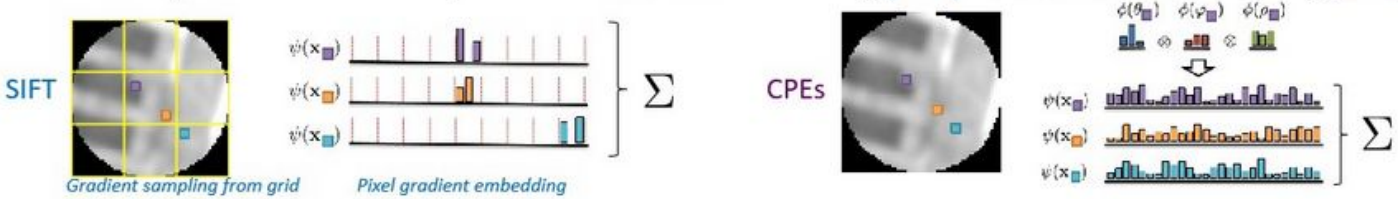
Layer	Conv. Kernel / Full Co..	Non-Lin	Pool.	Norm.	# Param
conv1	6x6x1x32, 1	TanH	Max	LRN	1.1K
conv2	5x5x32x128, 1, group: 2	TanH	Max	LRN	51.2K
conv3	4x4x128x256, 1, group: 2	TanH	Max	LRN	262 K
conv4	5x5x256x512, 1, group: 2	--	--	--	1.64M
ip1	512 --- 256	ReLU	--	--	131K
ip2	256 --- 128	--	--	--	32.8K
total					**2.1M**

cmp-ab - A. Bursuc, INRIA, G. Tolias,CTU Prague

Continuous Patch Embeddings

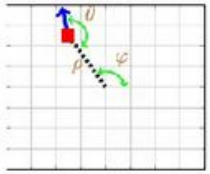
A. Bursuc (Inria), G. Tolias (CMP), H. Jégou (FAIR), O. Chum (CMP)

- Leverage strengths of SIFT and address its drawbacks
  - Patch representation: set of pixels with discrete embedding (SIFT) or continuous embedding (CPE)

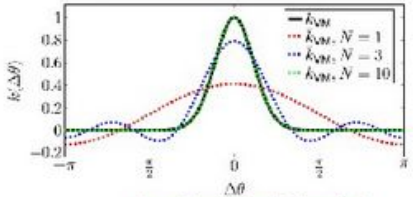


Gradient sampling from grid *Pixel gradient embedding*

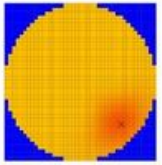
  - Pixel attributes: polar coordinates, gradient angle and magnitude
  - Feature maps for pixel attributes reproducing behavior of non-linear kernel while preserving linearity



Pixel attributes



Kernel behavior 1D (angles)



continuous spatial similarity
red: high orange: low

Kernel behavior 2D (polar coords)

- **Our submission:** Handcrafted kernel descriptor with 567 dimensions, post-processed and reduced down to 100D with supervised linear discriminant projections.

[Bursuc et al. ICMR 2015]

zagreb-nm - Nenad Markuš, University of Zagreb, Croatia

- Datasets with annotated correspondences are rare
- Learn from matching and non-matching bags of keypoints!
- A matching score between two keypoint bags, K_1 and K_2:

$$m_{e,\tau}(K_1, K_2) = \sum_{i=1}^{n} \left[\min_{j=1}^{n} d_{ij}^2 \leq \tau \right],$$

(extracted with a CNN)

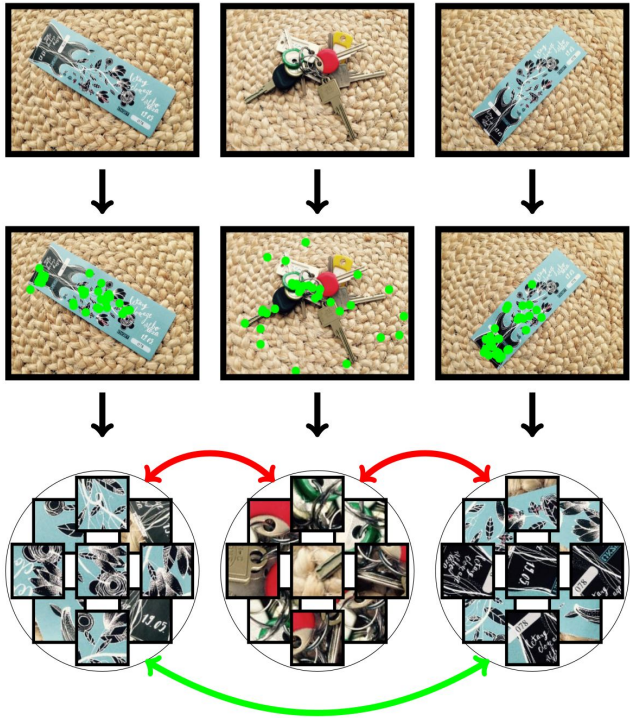
where d_{ij} is the Euclidean distance between descriptors of k_1 from K_1 and k_2 from K_2; $[x \leq \tau]$ is 1 if $x \leq \tau$ is true and 0 otherwise)

- Define a triplet loss: (+1 is regularization)

$$L = \sum \frac{1 + m_{e,\tau}(K, K^-)}{1 + m_{e,\tau}(K, K^+)}$$

- Approximate $[x \leq \tau]$ to achieve differentiability:

$$[x \leq \tau] \approx \frac{1}{1 + \exp(\beta(x - \tau))}$$



https://github.com/nenadmarkus/wlrn

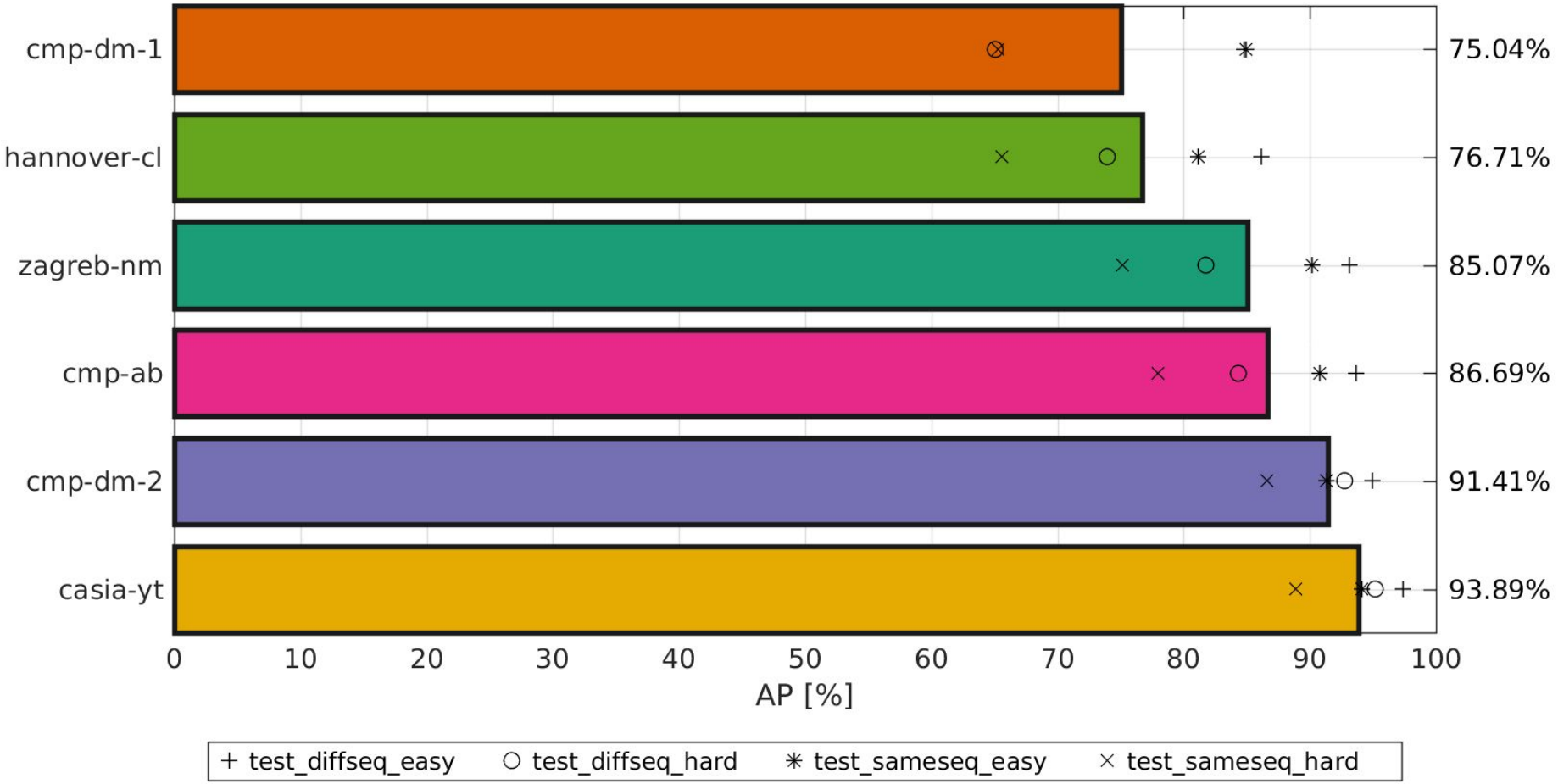
Learning local descriptors by optimising the keypoint correspondence criterion
Markus et. al. ICPR 2016.

Descriptor dimensionalities

Descriptor Name	Input patch size	Dimensionality
DC-Siam	64x64	256
DC-Siam 2stream	64x64	512
DeepDesc	64x64	128
TFeat	32x32	128
SIFT	65x65	128
zagreb-nm	**32x32**	**256**
cmp-dm	**65x65**	**128**
cmp-ab	**32x32**	**100**
hannover-cl	**65x65**	**128**
casia-yt	**32x32**	**128**

Patch pair classification

Patch classification - submissions



Patch classification - all

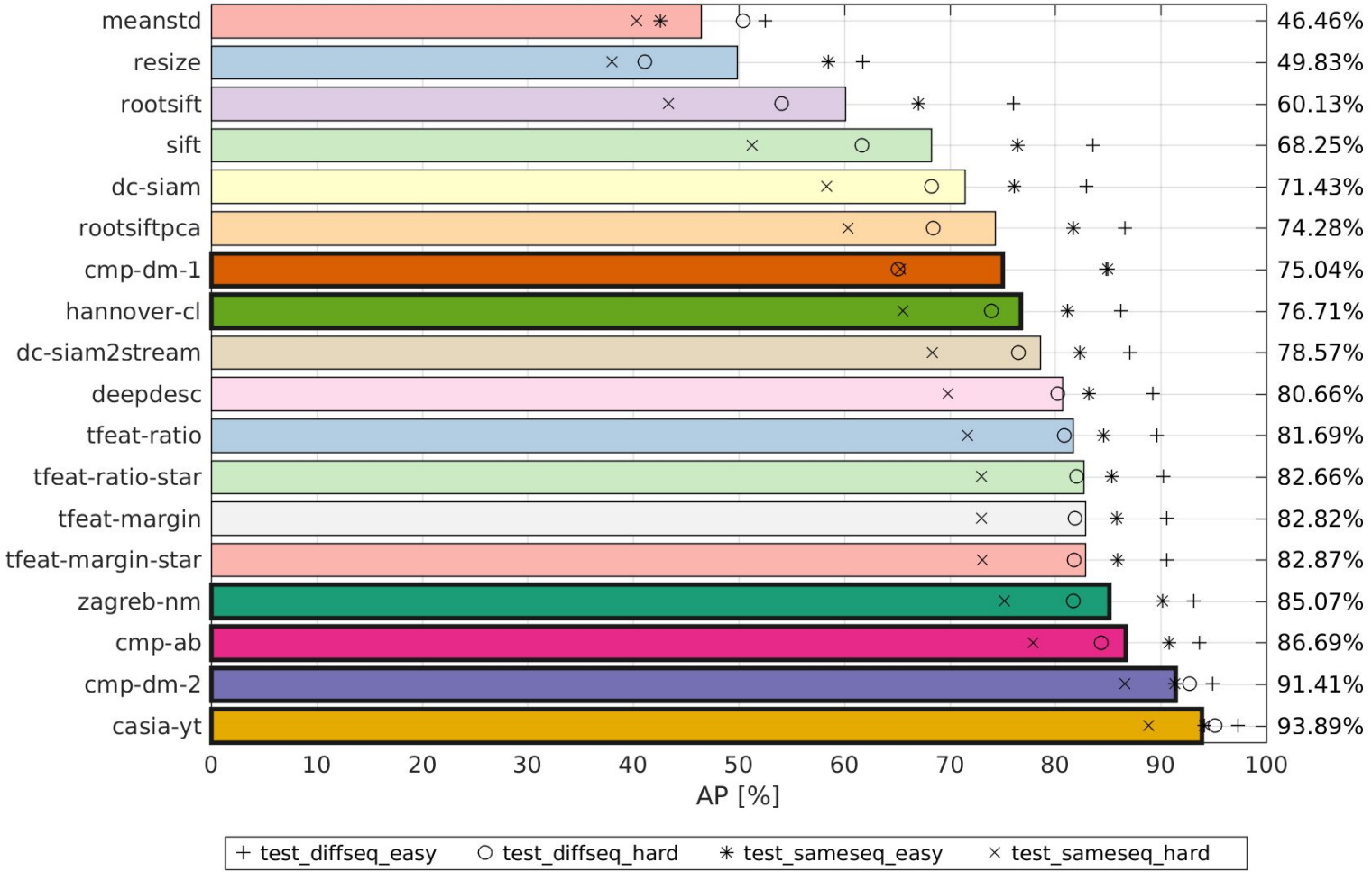


Image Matching

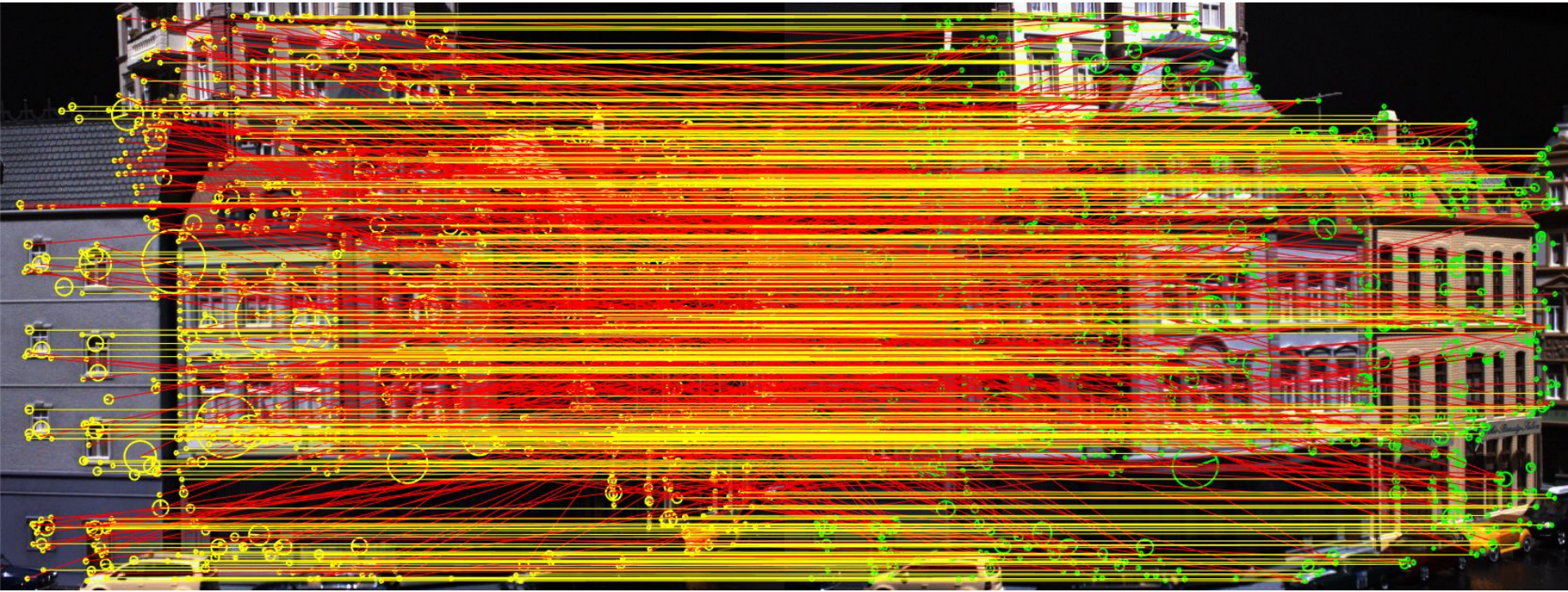


Image matching - submissions

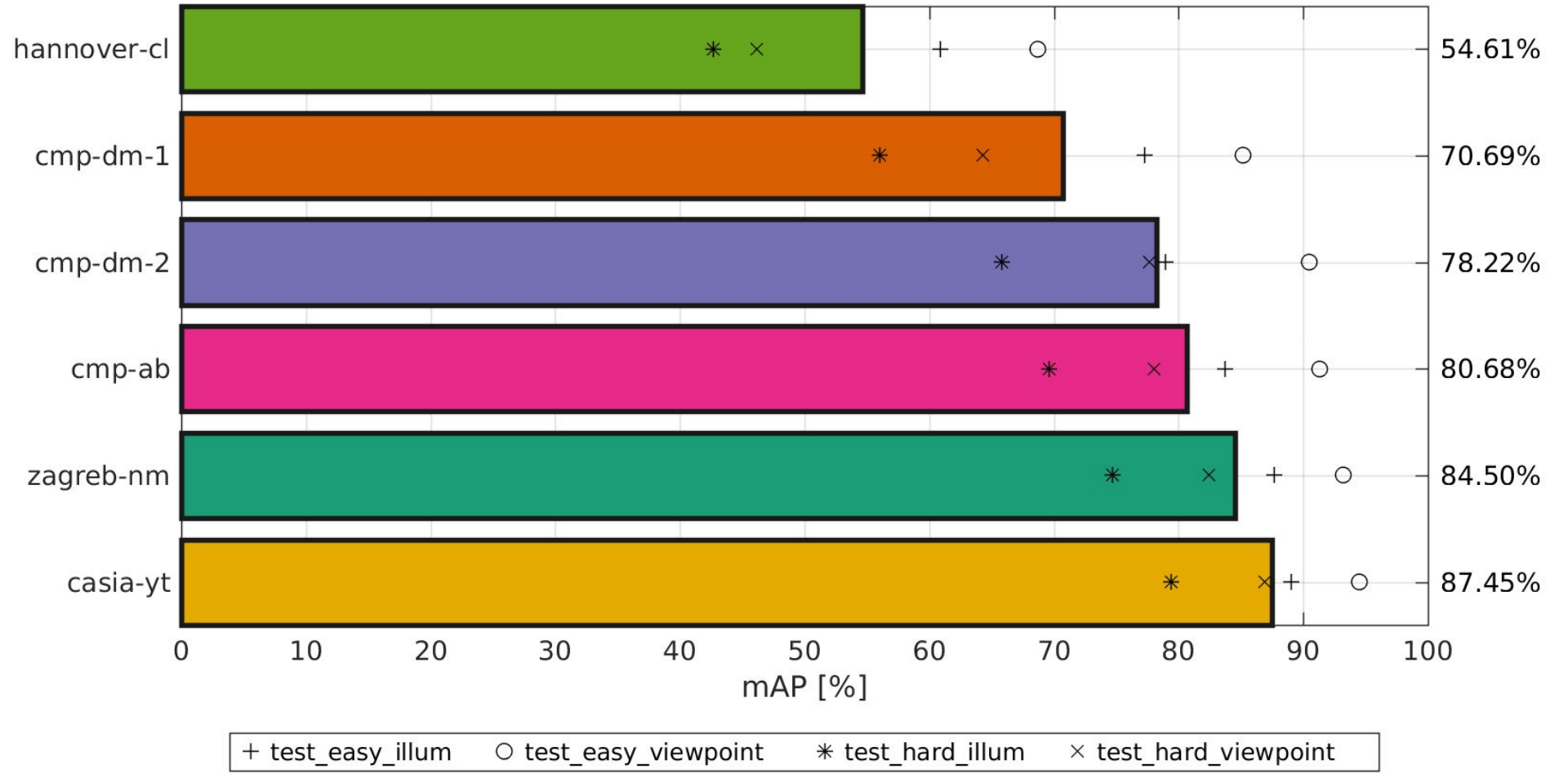
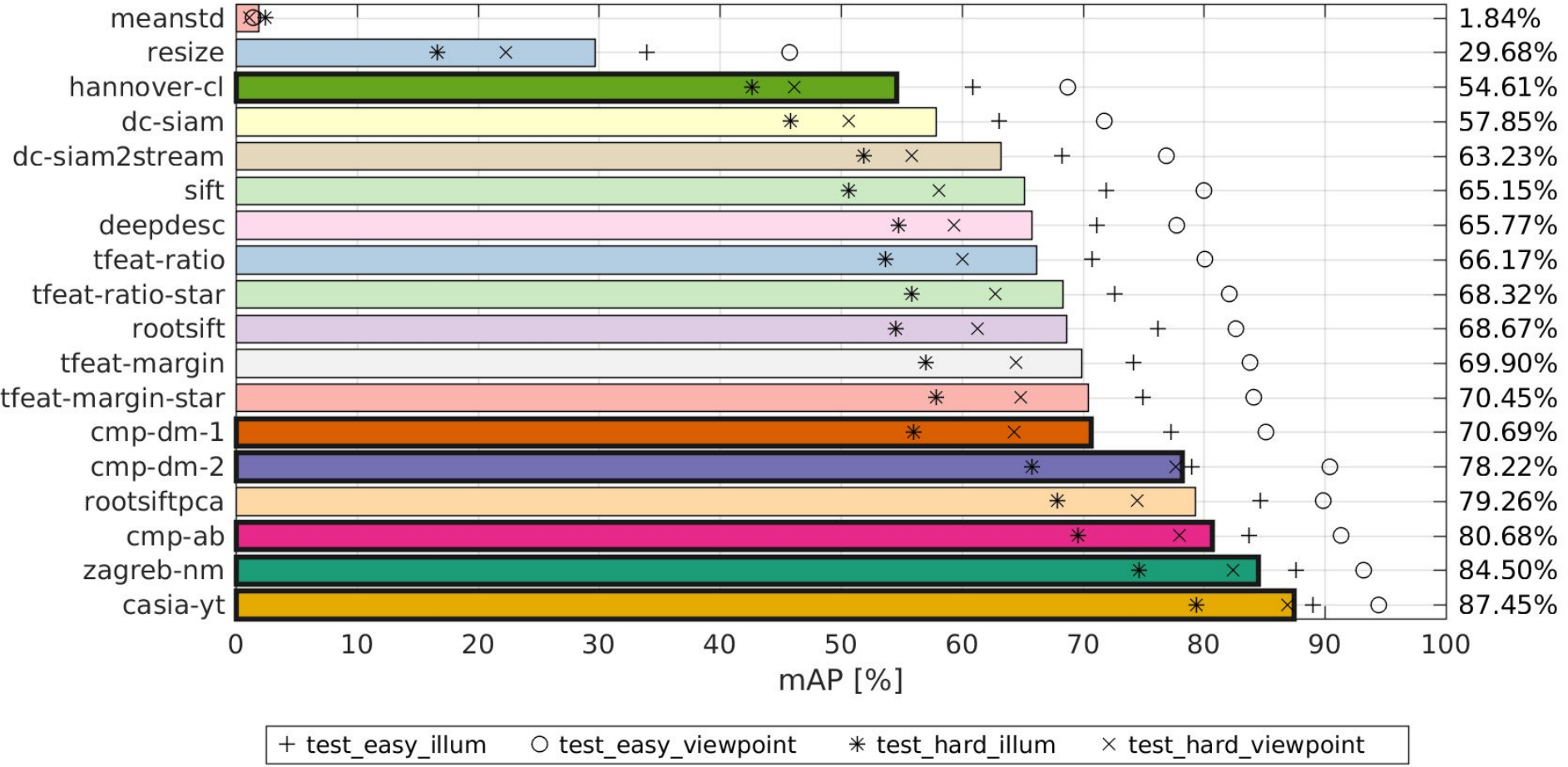
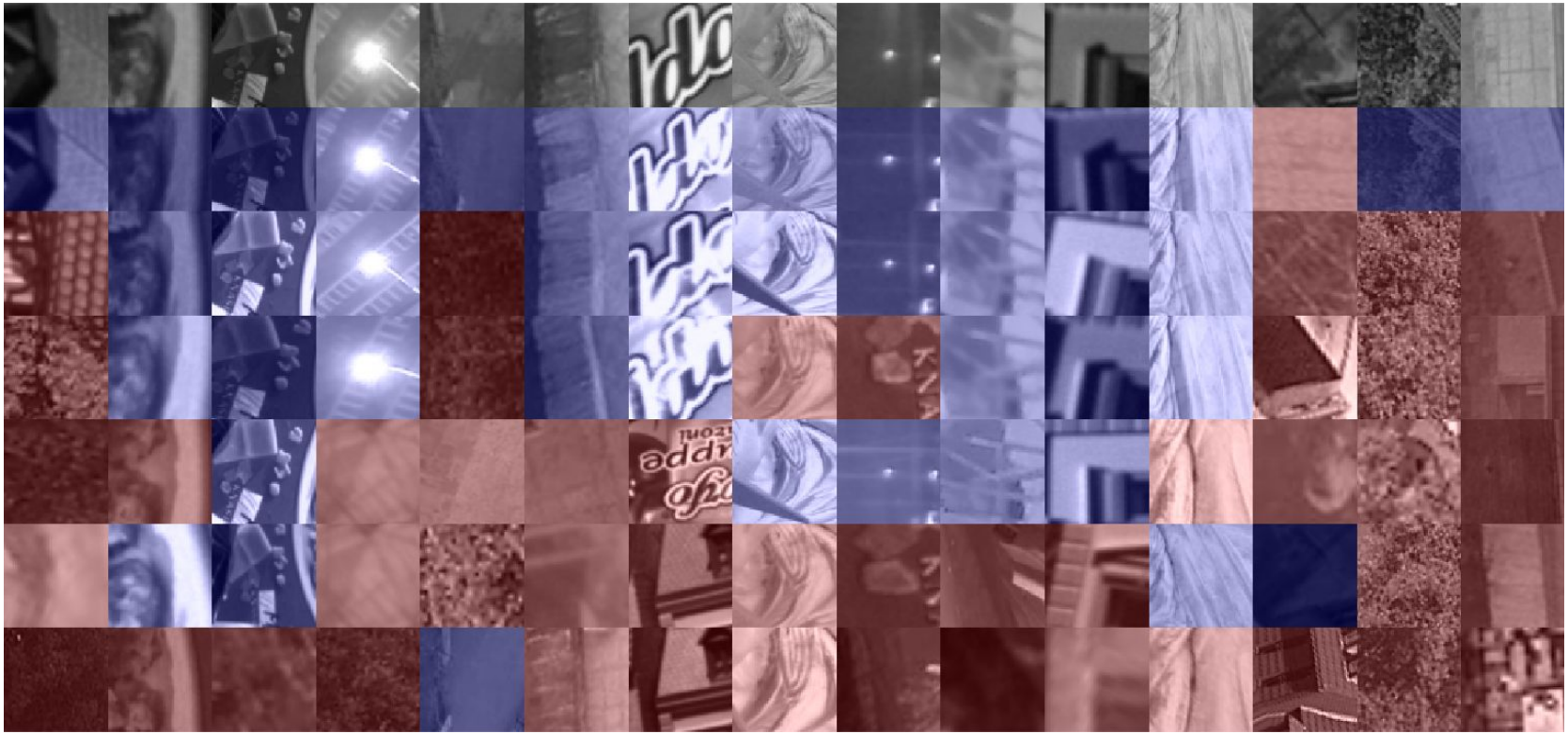


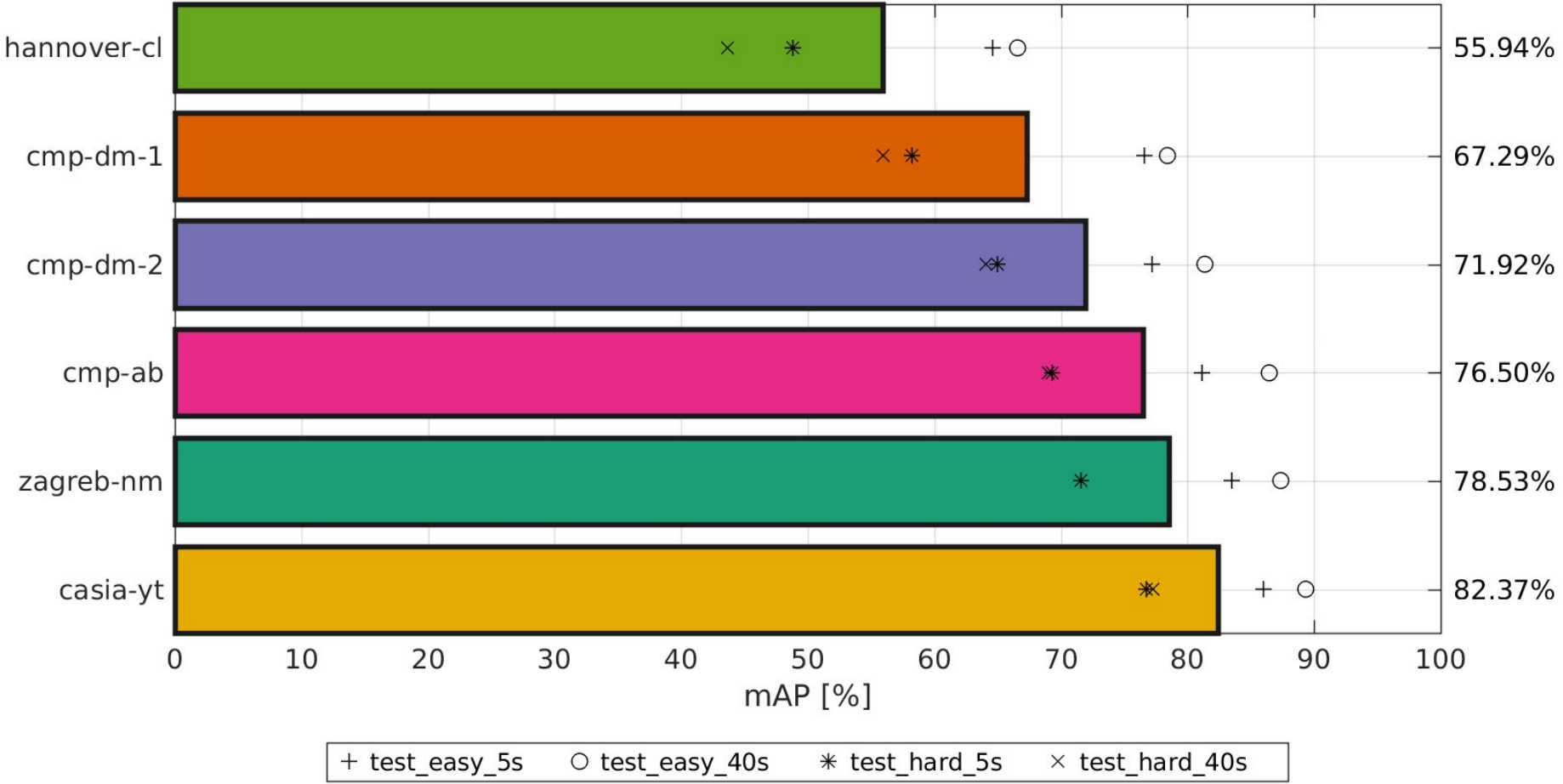
Image matching - all



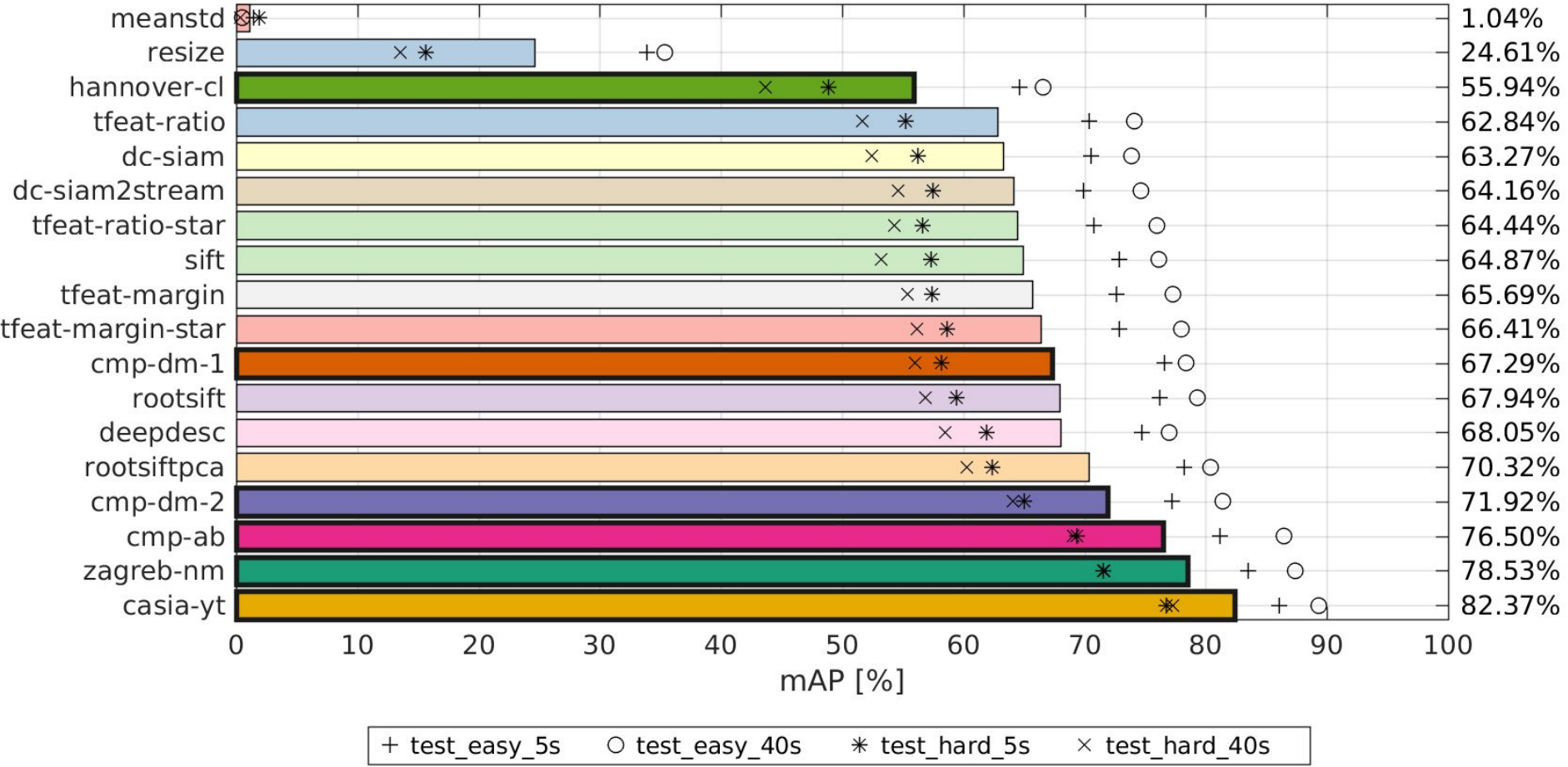
Patch Retrieval



Patch retrieval - submissions



Patch retrieval - all



Cross-task ranking comparison - submissions

Rank	Classification	Matching	Retrieval - Patch
1	casia-yt	casia-yt	casia-yt
2	cmp-dm-2	zagreb-nm	zagreb-nm
3	cmp-ab	cmp-ab	cmp-ab
4	zagreb-nm	cmp-dm-2	cmp-dm-2
5	hannover-cl	hannover-cl	hannover-cl

Cross-task ranking comparison - baselines

Rank	Classification	Matching	Retrieval - Patch
1	TFeat	RootSIFT-PCA	RootSIFT-PCA
2	DeepDesc	TFeat	DeepDesc
3	DeepCompare	RootSIFT	RootSIFT
4	RootSIFT-PCA	DeepDesc	TFeat
5	SIFT	SIFT	SIFT