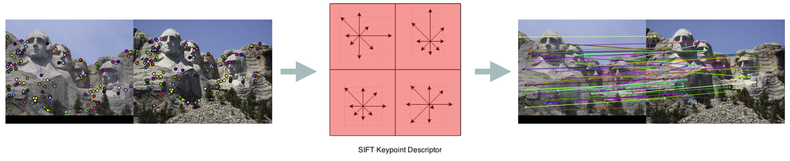
Local Descriptors Optimized for Average Precision

Kun He[1] Yan Lu[2] Stan Sclaroff[1]
[1]Boston University [2]Honda Research Institute USA

Learning Local Features

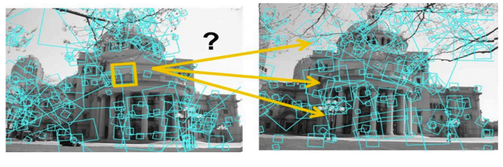
- Traditional (local feature based) computer vision pipelines:



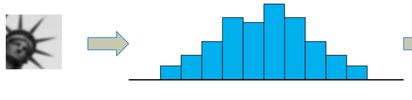
- We learn **Local Feature Descriptors** for vision pipelines
- By optimizing for the **Feature Matching** stage

Optimizing Feature Matching Performance

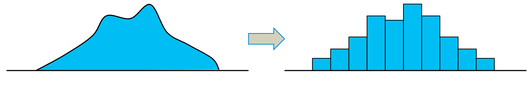
- Feature matching is nearest neighbor retrieval w/ *binary relevance*
- Common evaluation metric: Average Precision (AP)



- **Optimize AP**: [Paper 367] *Hashing as Tie-Aware Learning to Rank*

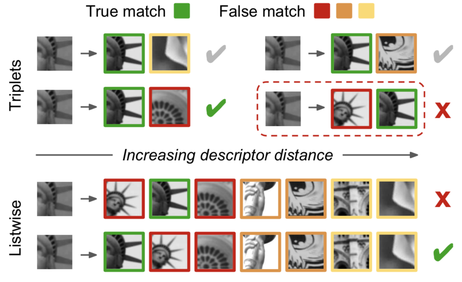


- Binary descriptors: directly reuse TALR
- Real-valued descriptors: reduce to TALR by **distance quantization**



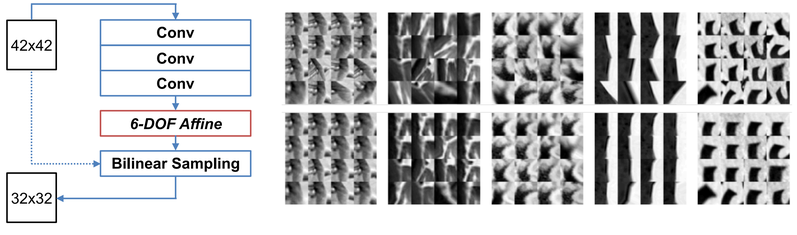
"Learning to Rank" View

- Most existing methods: local ranking with triplets
- Optimization issues (hard negative mining, sampling)
- Ours: listwise ranking
- Direct optimization, no complex heuristics

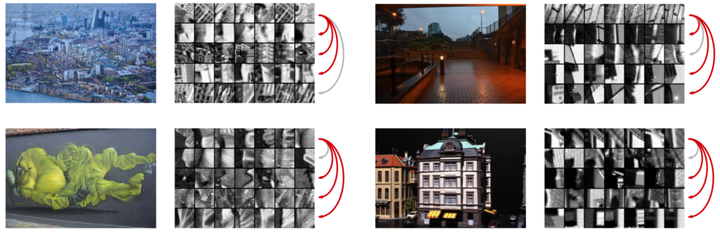


Task-Specific Improvements

- **Geometric Alignment**: Spatial Transformer module [2]



- **Label Mining** on HPatches dataset [3]
- Cluster patches to mine in-sequence hard negatives



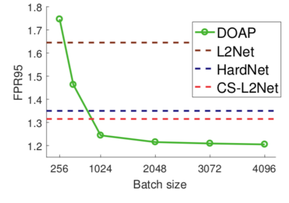
Experiments

- UBC Phototour / Brown dataset: patch verification

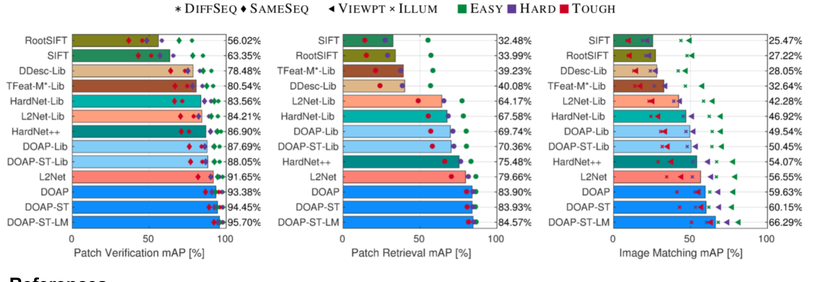
Method	Train	Notredame Yosemite		Liberty Yosemite		Liberty Notredame		FPR95
	Test	Liberty		Notredame		Yosemite		Mean
SIFT	128	29.84		22.53		27.29		26.55
MatchNet (CVPR'15)	128	7.04	11.47	3.82	5.65	11.6	8.70	8.05
TFeat-M* (BMVC'16)	128	7.39	10.31	3.06	3.80	8.06	7.24	6.64
TL-AS-GOR (ICCV'17)	128	4.80	6.45	1.95	2.38	5.40	5.15	4.36
DC-2ch2st+ (CVPR'15)	512	4.85	7.20	1.90	2.11	5.00	8.39	4.19
L2Net+ (CVPR'17)	128	2.36	4.7	0.72	1.29	2.57	1.71	2.23
HardNet+ (NIPS'17)	128	2.28	3.25	0.57	0.96	2.13	2.22	1.90
DOAP+	128	1.54	2.62	0.43	0.87	2.00	**1.21**	1.45
DOAP-ST+	128	**1.47**	**2.29**	**0.39**	**0.78**	**1.98**	1.35	**1.38**

- RomePatches [3]: patch retrieval

Method	Coverage	Dim.	Train	Test
SIFT	51x51	128	91.6	87.9
AlexNet-conv3	99x99	384	81.6	79.2
PhilippNet (arXiv'14)	64x64	512	86.1	81.4
CKN-grad (ICCV'15)	51x51	1024	92.5	88.1
DOAP	51x51	128	**95.9**	**88.4**
Binary DOAP	51x51	256	95.2	86.8



- HPatches [3]: patch verification/retrieval, image matching
  116 image sequences (76 train, 40 test), 2.5M patches



References

[1] Y. Tian, B. Fan, F. Wu. *L2-Net: Deep Learning of Discriminative Patch Descriptor in Euclidean Space*, CVPR 2017
[2] M. Jaderberg et al. *Spatial Transformer Networks*, NIPS 2015
[3] M. Paulin et al. *Local Convolutional Features with Unsupervised Training for Image Retrieval*, ICCV 2015
[4] V. Balntas*, K. Lenc*, A. Vedaldi, K. Mikolajczyk. *HPatches: A benchmark and evaluation of handcrafted and learned local descriptors*, CVPR 2017