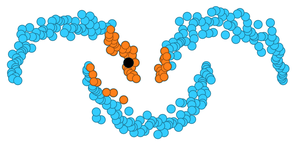
Mining on Manifolds: Metric Learning without Labels

Ahmet Iscen[1], Giorgos Tolias[1], Yannis Avrithis[2], Ondřej Chum[1]
[1]Visual Recognition Group, CTU in Prague, [2]Inria

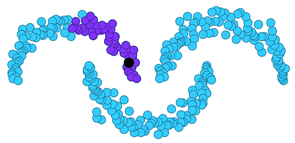
CVPR 2018 SALT LAKE CITY • JUNE 18-22

Goal: unsupervised mining of training samples

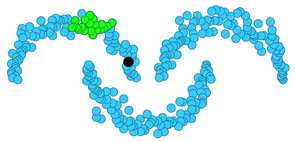
Learn descriptors that perform as manifold search



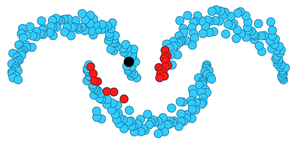
Euclidean NN (orange): $E(\mathbf{y})$



Manifold NN (purple): $M(\mathbf{y})$



Hard positives (green): $S^+ = M(\mathbf{y}) \setminus E(\mathbf{y})$



Hard negatives (red): $S^- = E(\mathbf{y}) \setminus M(\mathbf{y})$

Mining on Manifolds (MoM): Approach

1. Input: unordered collection of n images
2. Construct affinity matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$
3. Compute stationary prob. distribution π of random walk on A.
4. Anchor selection: local maxima of π on A.
5. For each anchor $\mathbf{y}$
   - Top-k manifold ($M(\mathbf{y})$) and Euclidean ($E(\mathbf{y})$) neighbors.
   - Choose positive images:

$$S^+(\mathbf{y}) = \{\mathbf{x} \in X : \mathbf{x} \in M(\mathbf{y}) \setminus E(\mathbf{y})\}$$

   - Choose negative images:

$$S^-(\mathbf{y}) = \{\mathbf{x} \in X : \mathbf{x} \in E(\mathbf{y}) \setminus M(\mathbf{y})\}.$$

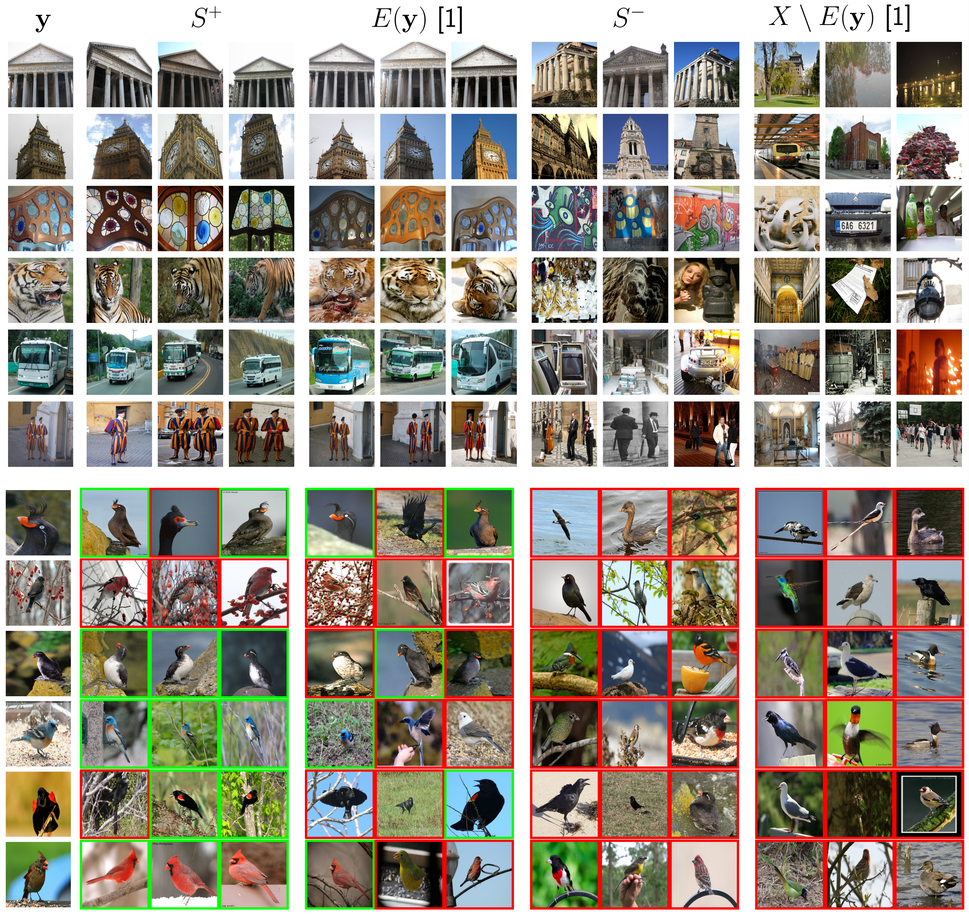
see poster 2778 for manifold search, *a.k.a diffusion*

Anchor images

Randomly sampled anchors from the top 100 anchors according to node importance

Randomly sampled anchors from the top 1000 anchors according to node importance

Positive and negative images



Applications

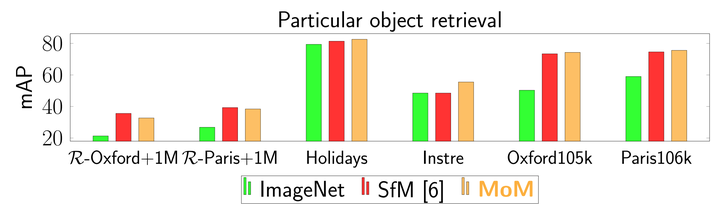
Particular object retrieval

- Training set: 1M unlabeled Flickr images
- Extract R-MAC descriptors with pre-trained network
- Mine 1k anchors
- Mine positive and negative pools per anchor (50 images per pool)
- Train MAC image descriptors as in Radenović *et al.* [6]
- Test sets: Oxford, Paris, Holidays, and Instre datasets

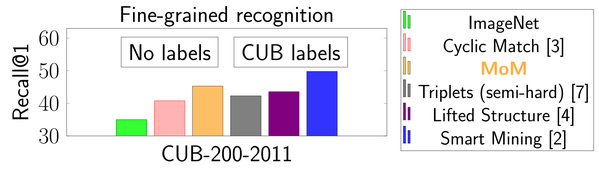
Fine-grained categorization

- Training set: images for 100 CUB200-2011 classes, as in [4] but unlabeled
- Extract R-MAC descriptors with pre-trained network
- Set each training image as anchor
- Mine positive and negative pools per anchor (50 images per pool)
- Train 64D image descriptors with triplet loss
- Test set: remaining 100 CUB200-2011 classes, standard benchmark [4]

Experiments



See paper 2730 [5] for $\mathcal{R}$-Oxford and $\mathcal{R}$-Paris benchmarks.



References

[1] R. Hadsell, S. Chopra, and Y. Lecun. Dimensionality reduction by learning an invariant mapping. In *CVPR*, 2006.
[2] B. Harwood, V. Kumar B G, G. Carneiro, I. Reid, and T. Drummond. Smart mining for deep metric learning. In *ICCV*, 2017.
[3] D. Li, W.-C. Hung, J.-B. Huang, S. Wang, N. Ahuja, and M.-H. Yang. Unsupervised visual representation learning by graph-based consistent constraints. In *ECCV*, 2016.
[4] H. Oh Song, Y. Xiang, S. Jegelka, and S. Savarese. Deep metric learning via lifted structured feature embedding. In *CVPR*, 2016.
[5] F. Radenović, A. Iscen, G. Tolias, Y. Avrithis, and O. Chum. Revisiting oxford and paris: Large-scale image retrieval benchmarking. *CVPR*, 2018.
[6] F. Radenović, G. Tolias, and O. Chum. CNN image retrieval learns from bow: Unsupervised fine-tuning with hard examples. *ECCV*, 2016.
[7] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *CVPR*, 2015.

Contact: ahmet.iscen@cmp.felk.cvut.cz