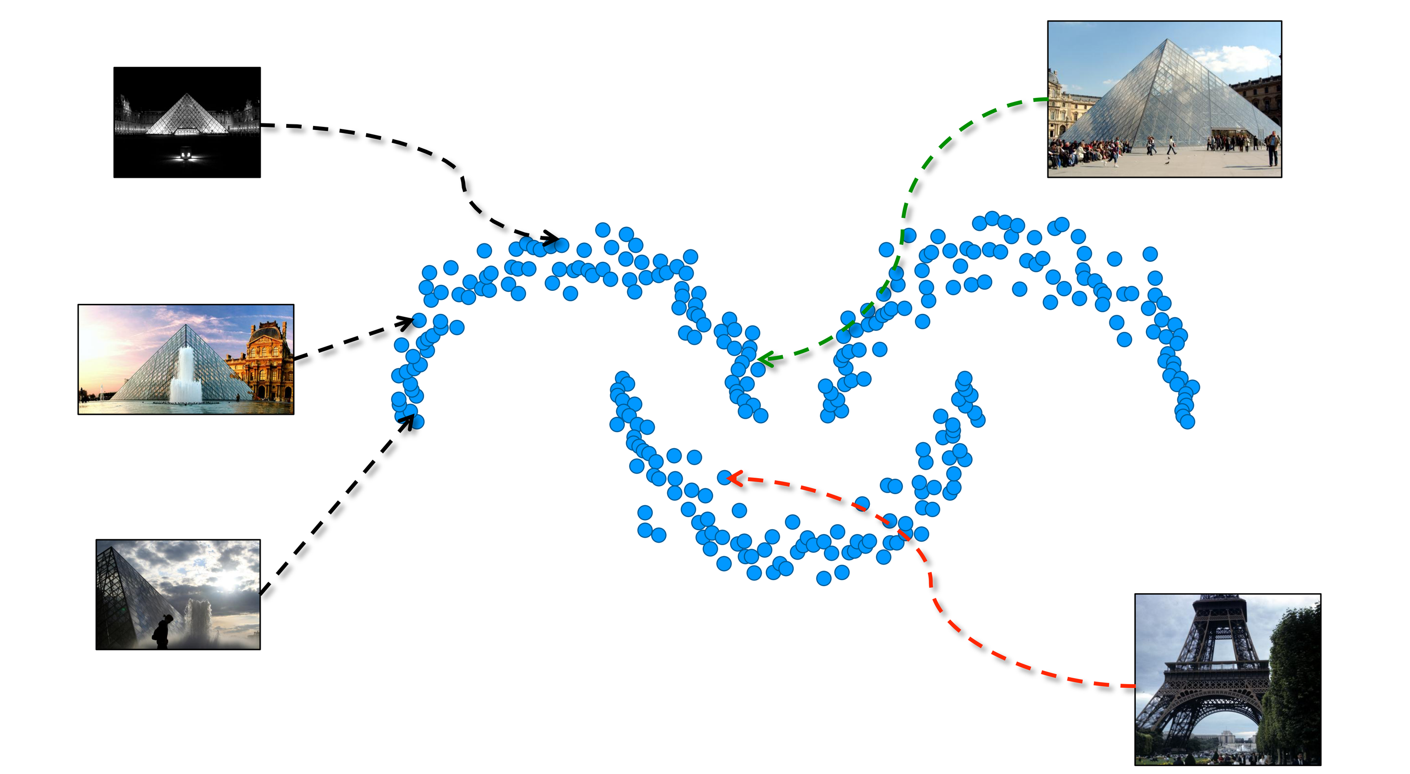
Efficient Diffusion on Region Manifolds: Recovering Small Objects with Compact CNN Representations

Ahmet Iscen[1], Giorgos Tolias[2], Yannis Avrithis[1], Teddy Furon[1], Ondřej Chum[2]

[1]Inria, [2]Visual Recognition Group, CTU in Prague

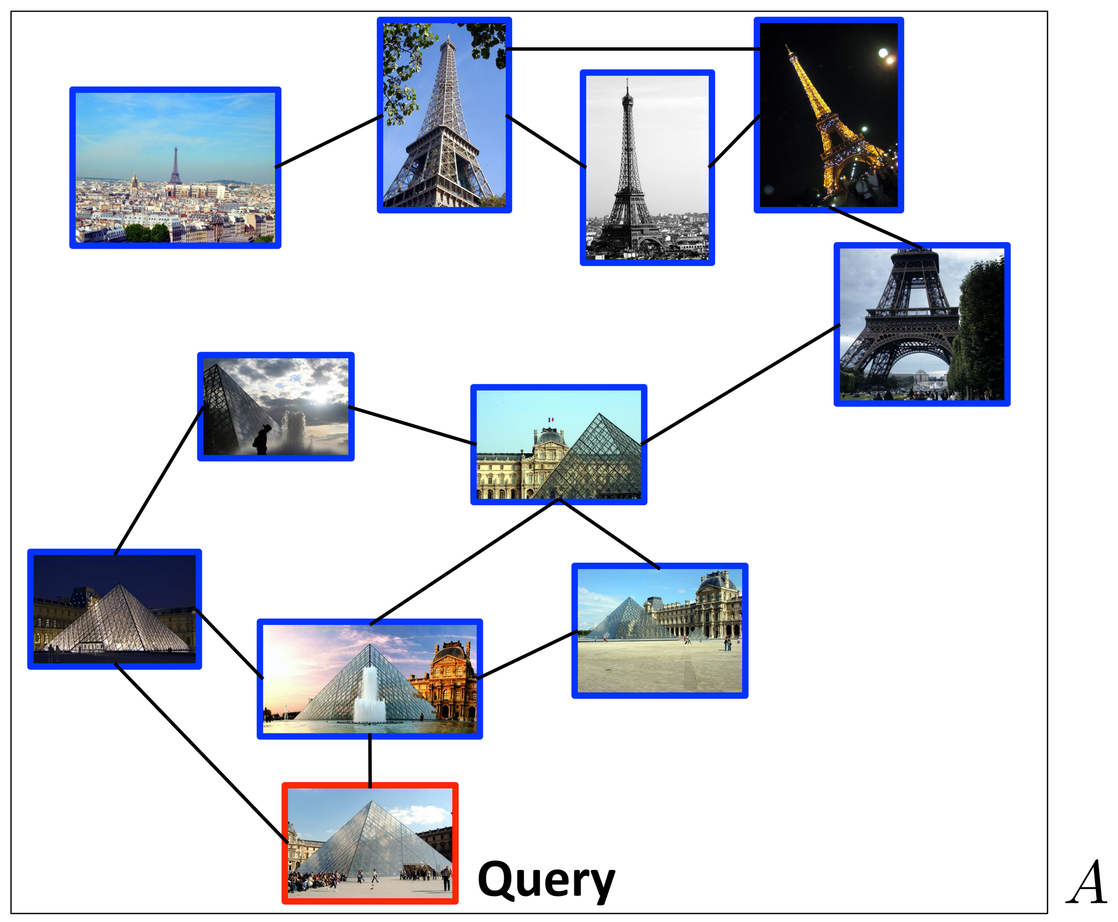
CVPR July 21-26 HONOLULU 2017

Motivation

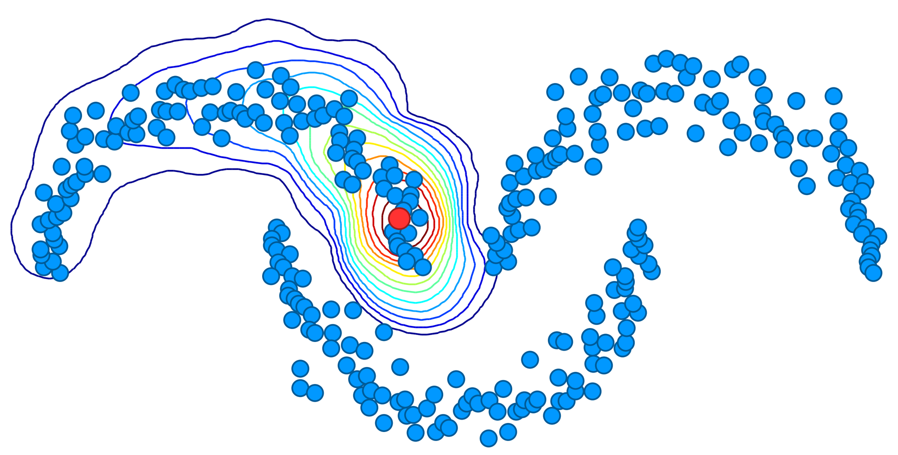


- Euclidean distance not appropriate for severe visual variations
- Solution: Ranking on manifolds via graph-based approach, i.e. diffusion [1]

Standard Diffusion [3]

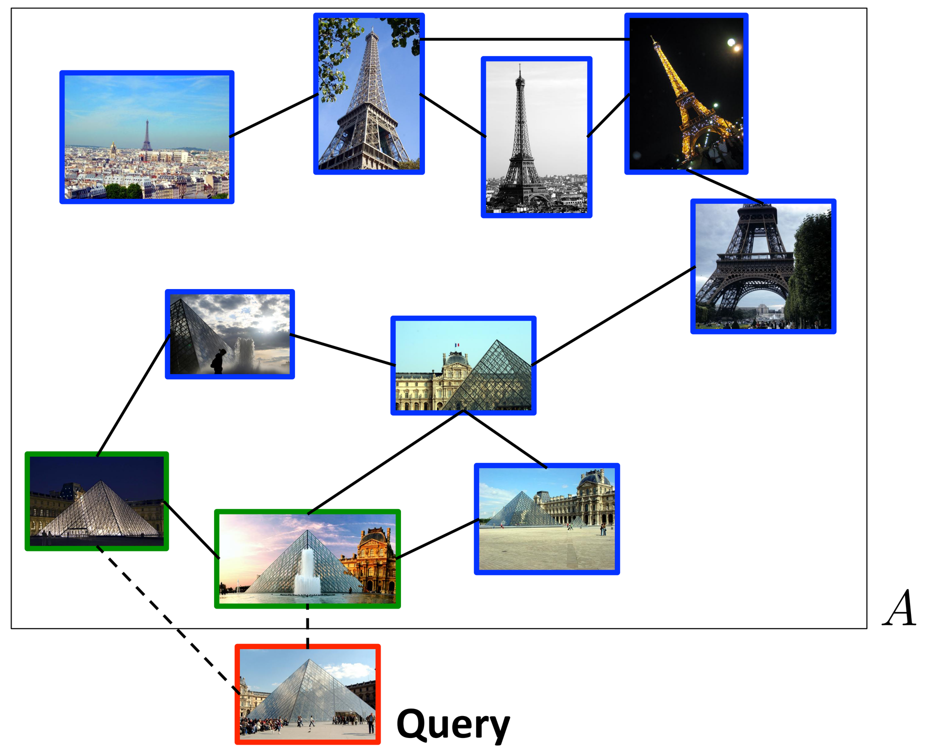


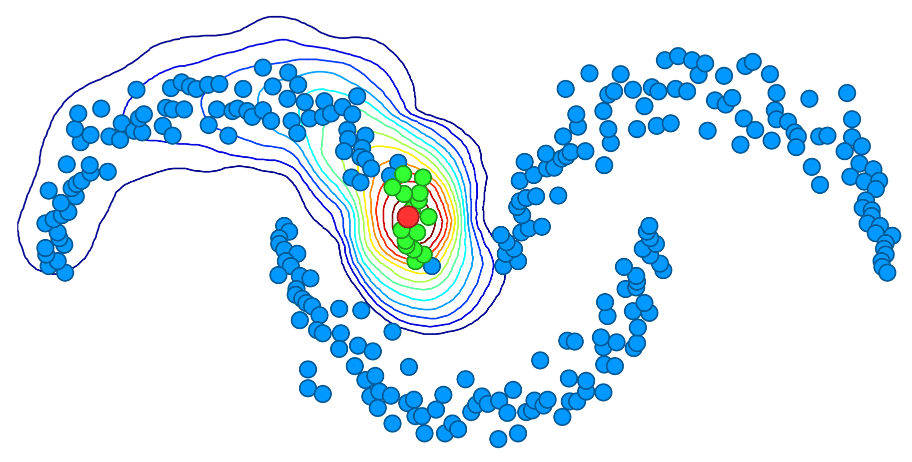
- Normalized affinity (reciprocal kNN) matrix: $S := D^{-1/2} A D^{-1/2}$
- The query is part of the graph
  $\mathbf{y} = (y_i) \in \mathbb{R}^n$, $y_i = 1$ if i-th node is a query, $y_i = 0$ otherwise
- Iterative solution preferred in prior work [1]
  $\mathbf{f}^t = \alpha S \mathbf{f}^{t-1} + (1-\alpha)\mathbf{y}$
- Closed-form solution [3] commonly avoided
  $\mathbf{f}^\star = (1-\alpha)\mathcal{L}_\alpha^{-1}\mathbf{y}$, where $\mathcal{L}_\alpha := I_n - \alpha S$



Toy with 2D database points, query point, and iso-contours for manifold similarity

Diffusion for unseen queries

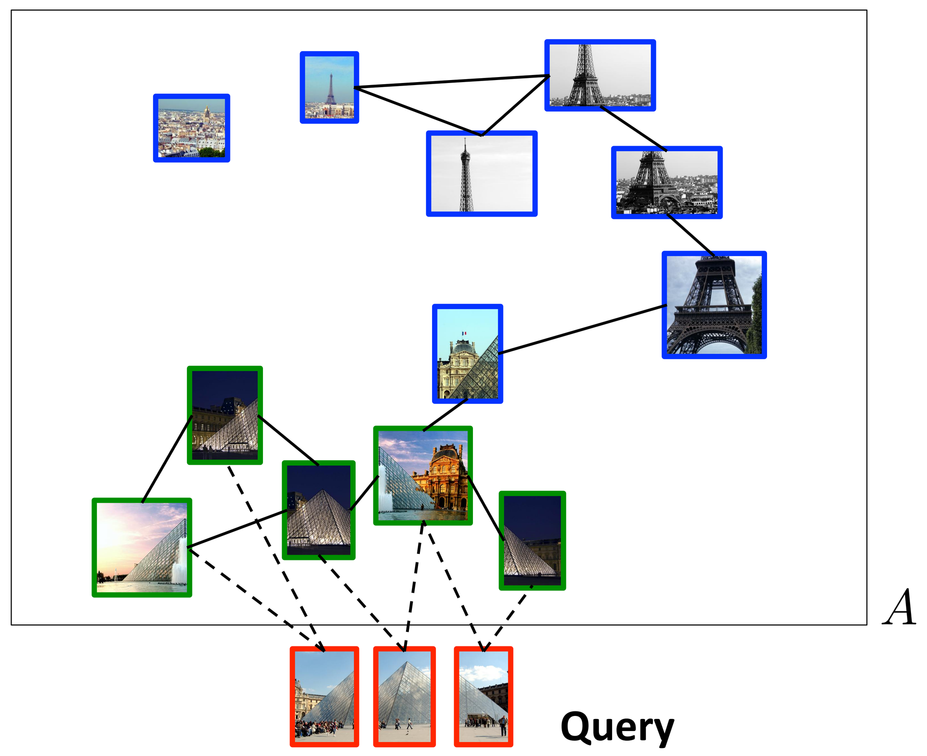


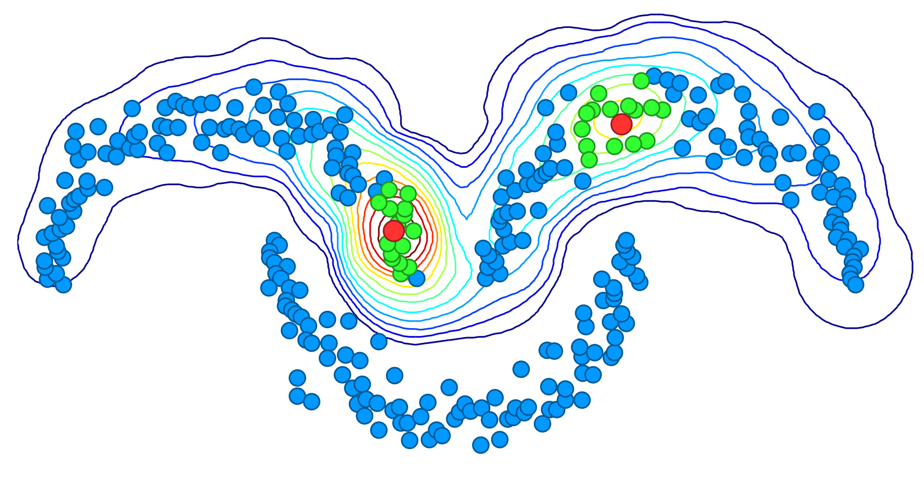


Toy with 2D database points, query point, kNN of the query, and iso-contours for manifold similarity

- Image retrieval with unseen queries: not part of the graph
- **Contribution**: Instead of searching for the query, search for its neighbors:
  $y_i = 1$ (or equal to similarity) if i-th node is a kNN of the query, $y_i = 0$ otherwise

Regional diffusion





Toy with multiple query points

- Global descriptors not effective for small objects, occlusion.
- Represent images by uniformly sampled overlapping regions [2]: each image represented by m vectors
- **Contribution**: Diffusion with regions as nodes, multiple regional queries issued with the cost of one
  $y_i = 1$ (or equal to similarity) if i-th node is a kNN of any query region, $y_i = 0$ otherwise

Efficient diffusion

- Iterative solution is not efficient: long to converge
- Closed-form solution $\mathbf{f}^\star = (1-\alpha)\mathcal{L}_\alpha^{-1}\mathbf{y}$ not scalable: $\mathcal{L}_\alpha^{-1}$ not sparse
- **Contribution**: Solve linear system $\mathcal{L}_\alpha \mathbf{f} = (1-\alpha)\mathbf{y}$ with conjugate gradients (CG)
- Conjugate directions with initial large step size: only a few iterations for good approximation
- We show that the iterative solution is equivalent to Jacobi solver: known as worse than CG

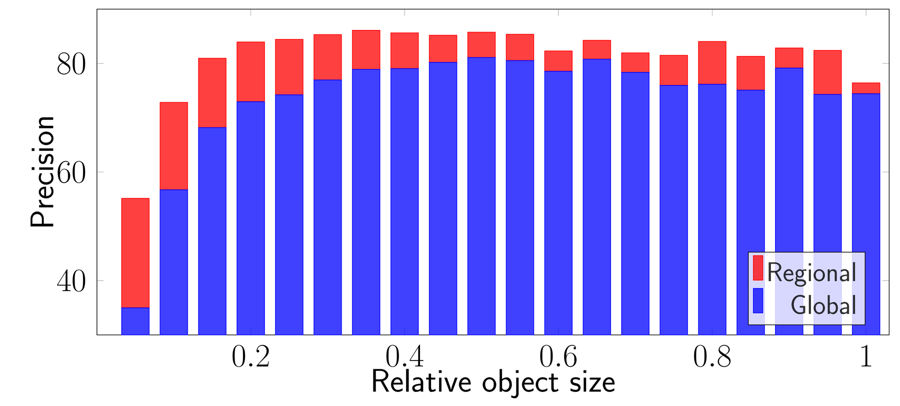
Large Scale Diffusion

- (Off-line) Reduce number of vectors: learn Gaussian Mixture Model (GMM) per image
- (Off-line) Use approximate NN-search for offline graph construction
- (On-line) Regional diffusion as re-ranking: only on top ranked images by truncated affinity matrix

Retrieval of Small Objects

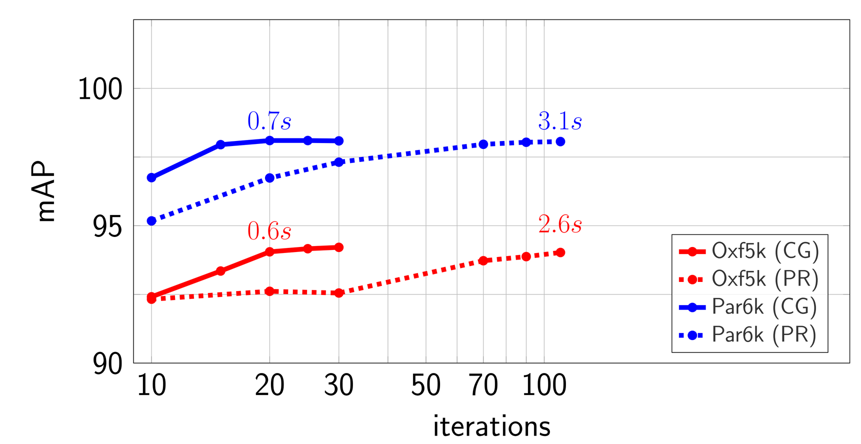
Precision at retrieved position with global → regional diffusion.

Query images (left: bounding box) and retrieved images with the largest improvement by regional diffusion.

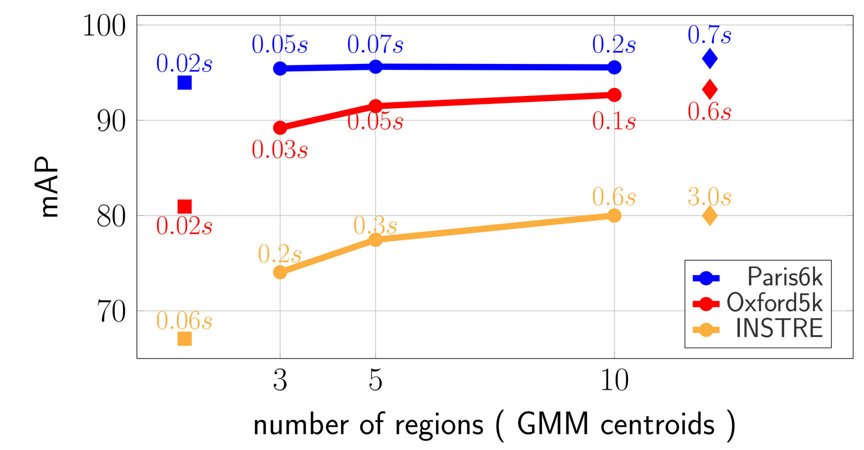


Precision at retrieved position on INSTRE dataset (averaged over positive images according relative object size)

Experiments



Speed and convergence comparison for regional diffusion between the iterative solution (PR) and ours with conjugate gradient (CG)



Performance and speed comparison vs number of vectors/image.
□: global diffusion, ◇: default grid with 21 regions per image

Method	$m \times d$	INSTRE	Oxf5k	Oxf105k	Par6k	Par106k
Regional descriptors - nearest neighbor search						
R-match [2]	21×512	55.5	81.5	76.5	86.1	79.9
R-match [2]	21×2,048	71.0	88.1	85.7	94.9	91.3
Regional descriptors - query expansion						
Hamming Query Expansion	2.4k×128	74.7	89.4†	84.0†	82.8†	-
R-match [2]+AQE	21×512	60.4	83.6	78.6	87.0	81.0
Regional diffusion⋆	5×512	77.5	91.5	84.7	95.6	93.0
Regional diffusion⋆	21×512	80.0	93.2	90.3	96.5	92.6
R-match [2]+AQE	21×2,048	77.1	91.0	89.6	95.5	92.5
Regional diffusion⋆	5×2,048	88.4	95.0	90.0	96.4	**95.8**
Regional diffusion⋆	21×2,048	**89.6**	**95.8**	**94.2**	**96.9**	95.3

References:

[1] M. Donoser and H. Bischof. Diffusion processes for retrieval revisited. In *CVPR*, 2013.

[2] A. S. Razavian, J. Sullivan, S. Carlsson, and A. Maki. Visual instance retrieval with deep convolutional networks. *ITE Transactions on Media Technology and Applications*, 4:251–258, 2016.

[3] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf. Ranking on data manifolds. In *NIPS*, 2003.

Contact: ahmet.iscen@inria.fr