Domain Adaptation and Few-shot Learning ACCV2018 WORKSHOP

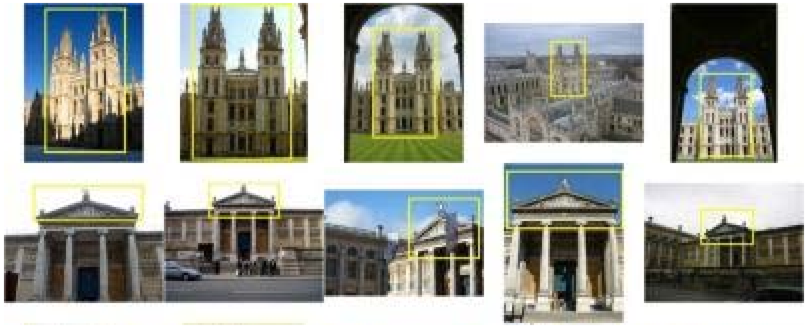
Unsupervised Feature Adaptation for Image Retrieval via Diffusion Process

Lei Wang
School of Computing and Information Technology
University of Wollongong, Australia

02-Dec-2018

Outline

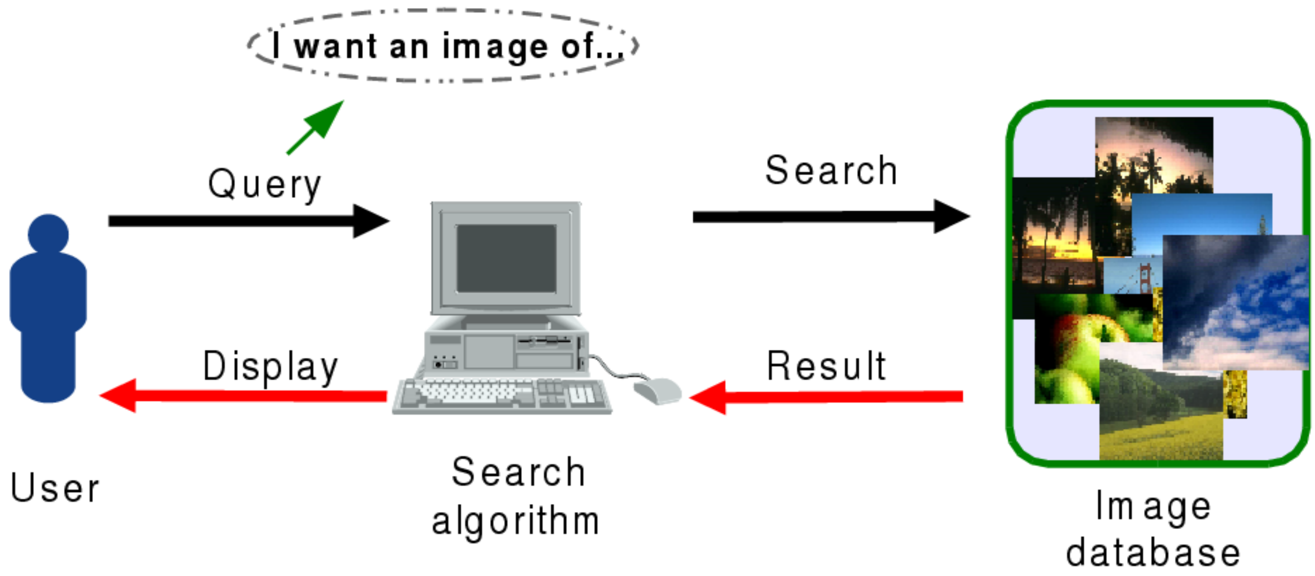
- **Content-based Image Retrieval**
- A Domain Adaptation Perspective
- Feature Adaptation for CBIR
  - Diffusion Process in image retrieval
  - A kernel mapping view of diffusion
  - Feature Adaptation by modeling diffusion process
- Conclusion



Images courtesy of related papers and authors

Introduction

- Retrieval
  - Getting back information that has been stored in a database
- Image Retrieval



Introduction

- Content-based image retrieval
  - Human annotators are replaced by **computers**
  - Text annotations are replaced by **visual features**
  - Retrieval by the **similarity** of associated visual **features**

Drouin Post Office, front desks

Iron Ore

Fashion

Introduction

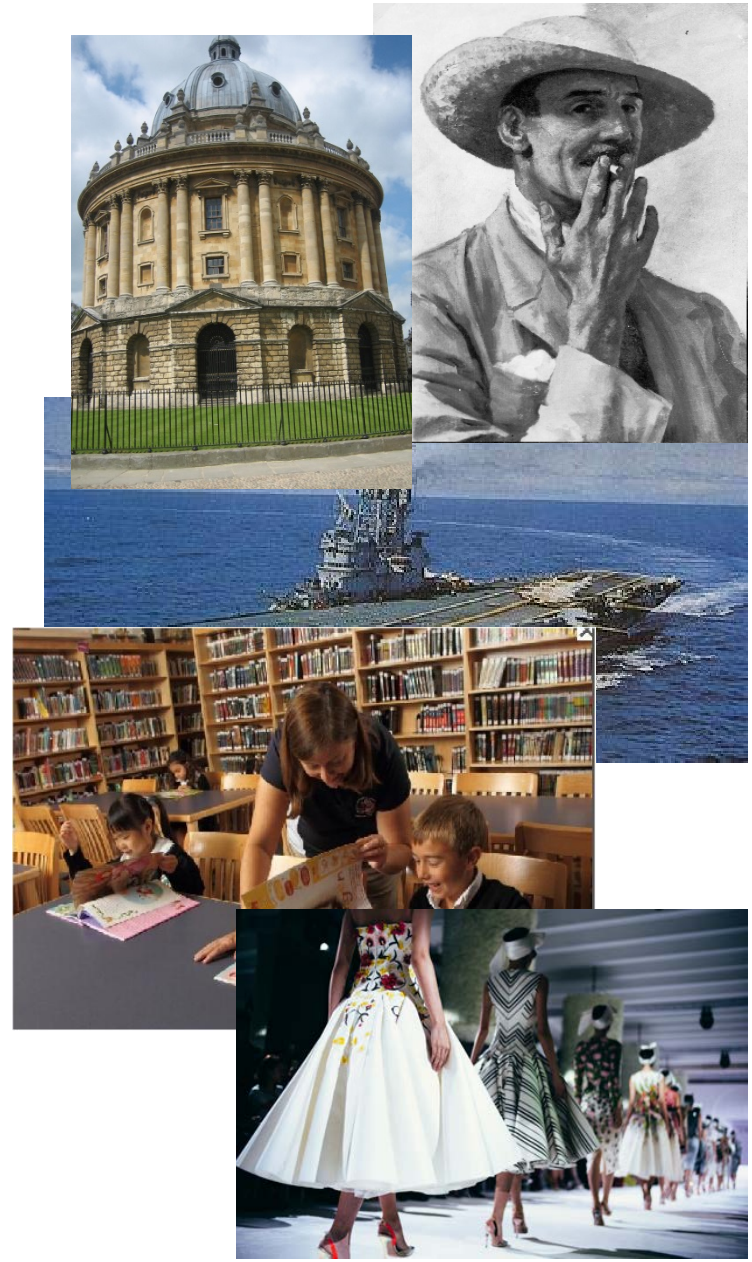
Image retrieval on the collection of National Archives of Australia

Query:

Retrieval result

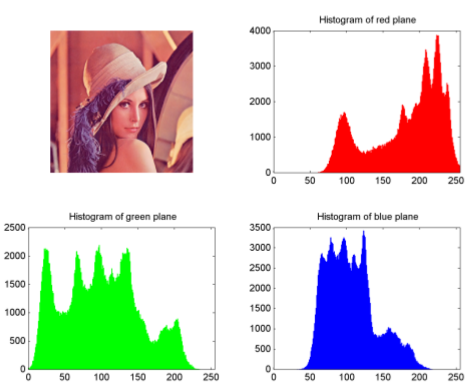
Introduction

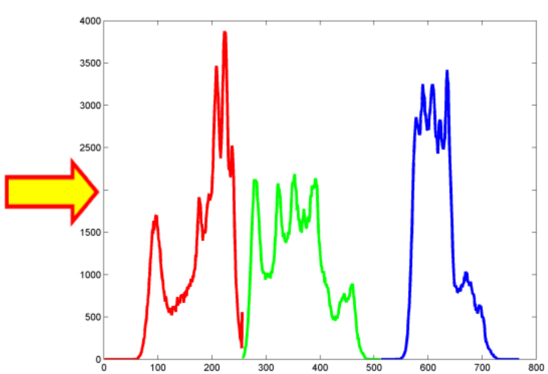
- **Applications** of CBIR
  - Scene understanding
  - Online shopping
  - Photo collection management
  - Crime investigation
  - Fashion and design
  - Localisation and navigation
  - Medical Image analysis
  - ….

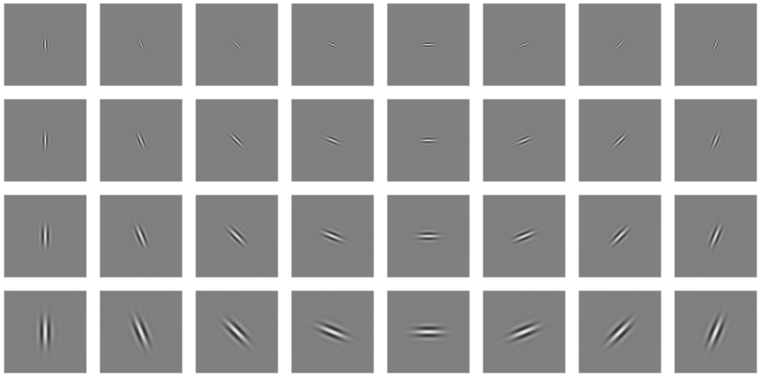


CBIR: Early days

- Hand-crafted **features**
  - Color, texture, shape, structure, etc.
  - Goal: “Invariant and discriminative”
- Similarity or distance **measure**
  - Euclidean distance, Manhattan distance, etc.
  - Specially designed measures



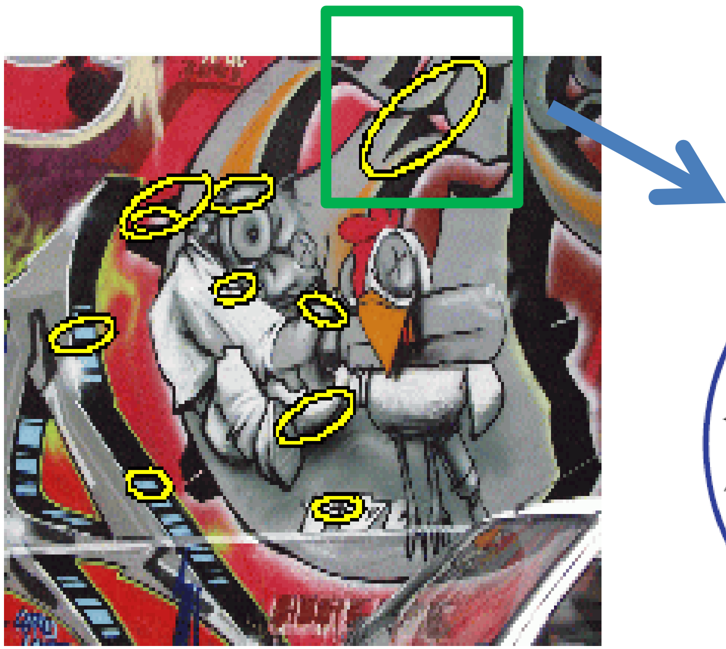




CBIR: Days of the BoF model

Local Invariant **Features**

- **SIFT,** HOG, SURF, CENTRIST, filter-based, ...
  - **Invariant** to view angle, rotation, scale, illumination, ...



http://www.robots.ox.ac.uk/~vgg/software/

SIFT (Scale Invariant Feature Transform

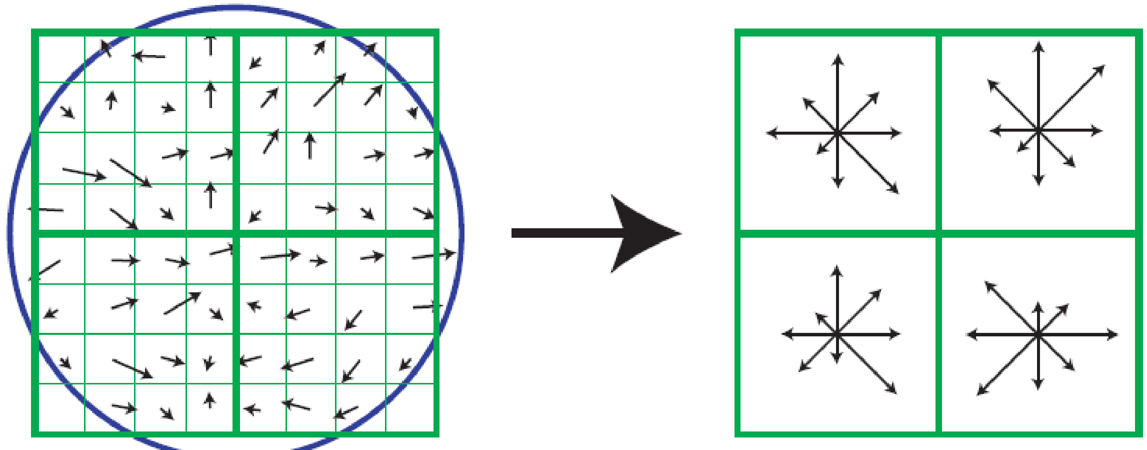


Image courtesy of David Lowe, IJCV04

CBIR: Days of the BoF model

11 — Local Soft-assignment Coding & Mix-order pooling (Liu, ICCV11); Comparative Study on BoF model (Chatfield, BMVC, 2011);

10 — Locality-constrained Linear Coding for BoF (Wang, CVPR10); Coding & pooling scheme comparison (Boureau, CVPR10);

09 — Sparse coding for BoF (Yang, CVPR09) Local Coordinate Coding (Yu, NIPS09)

08 — Kernel Codebook (van Gemert, ECCV08); In Defense of Nearest Neighbor Classifier (Boiman, CVPR08)

07 — Comparative Study (Zhang, IJCV07); Coding with Fisher Kernels (Perronnin, CVPR07)

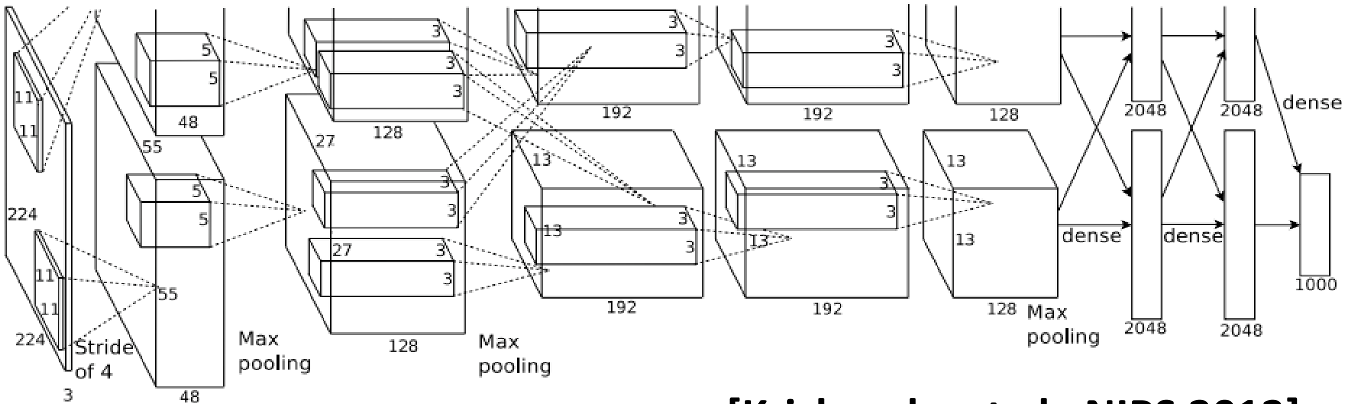
06 — Vocabulary tree (Nister, CVPR06); Randomized Clustering Forests (Moosmann, NIPS06); Spatial Pyramid Matching (Lazebnik, CVPR06)

05 — Pyramid Match Kernel (Grauman, ICCV05); Dense sampling (Jurie, ICCV05); Compact Codebook (Winn, ICCV05)

03 — Video Google (Sivic, CVPR03); Bag-of-keypoints (Csurka, SLCV@ECCV04)

99 — Local Invariant Features, such as SIFT (Lowe, ICCV99)

CNNs: ImageNet Breakthrough



[Krizhevsky et al. NIPS 2012]

- Krizhevsky et al. win 2012 ImageNet classification with a **much bigger ConvNet**
    - **deeper**: 7 stages vs 3 before
    - **larger**: 60 million parameters vs 1 million before
    - **16.4%** error (top-5) vs Next best 26.2% error

- This was made possible by:
    - **fast hardware**: GPU-optimized code
    - **big dataset**: 1.2 million images vs thousands before
    - **better regularization**: dropout et al.

Image courtesy of Deng et al.

CBIR: Era of Deep Learning

12 ← Image Classification with DCNN (Krizhevsky, NIPS12)

13 ← Some papers appeared on Arxiv

14
CNN Features off-the-shelf (Razavian, CVPRW14); Neural codes (Babenko, ECCV14); Deep ranking (Wang, CVPR14); Multi-scale orderless pooling (Gong, ECCV14); Encoding High Dimensional Local Features (Liu, NIPS14); Survey: Deep learning for CBIR (Wan, ACMMM14); …

15
Deep filter banks (Cimpoi, CVPR15); Exploiting Local Features from DNN (Ng, CVPRW15); SPoC (Babenko, ICCV15); MatchNet (Han, CVPR15); …

16
R-MAC (Tolias, ICLR16); CNN IR Learns from BoW (Radenovic, ECCV16); CroW (Kalantidis, ECCVW16); Where to focus (Cao, 2016); NetVLAD (Arandjelovic, CVPR16); …

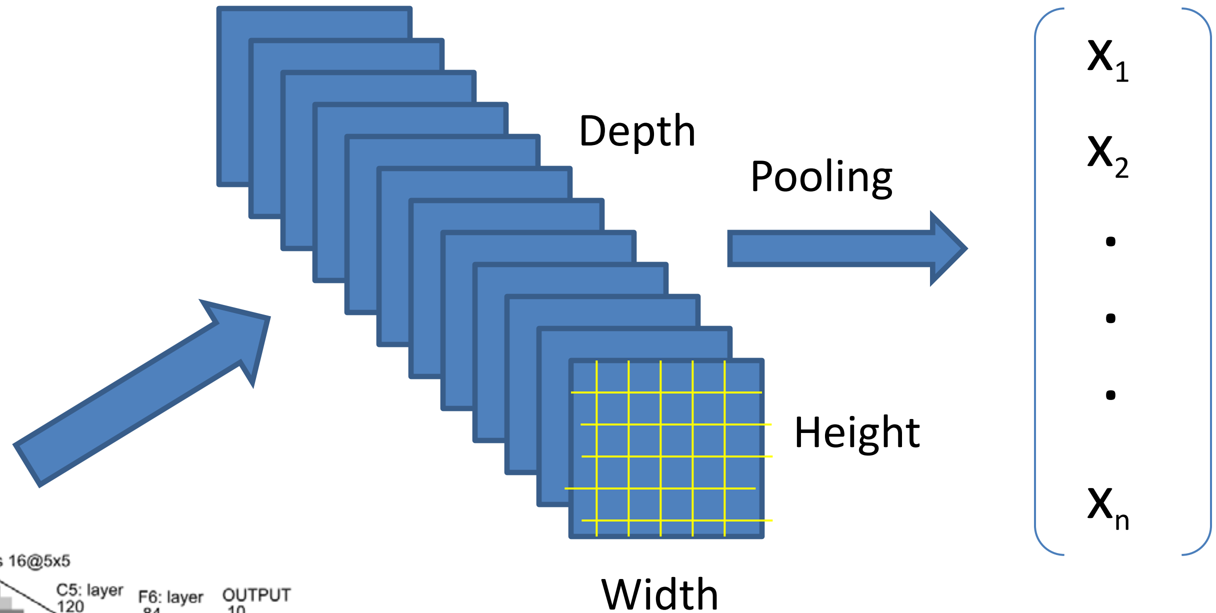
17
Efficient diffusion on Region Manifolds (Iscen, CVPR17); Large-Scale Image Retrieval with Attentive Deep Local Features (Noh, ICCV17); Ensemble Diffusion for Retrieval (Bai, ICCV17); …

18
Deep Shape Matching (Radenovic, ECCV18); Fast spectral ranking for similarity search (Iscen, CVPR18); Mining on manifolds (Iscen, CVPR18); SIFT meets CNN: A decade survey of instance retrieval (Zheng, TPAMI, 2018),…

CBIR: Era of Deep Learning

From **hand-crafted** features to **automatically learned** ones

Pre-trained CNN models



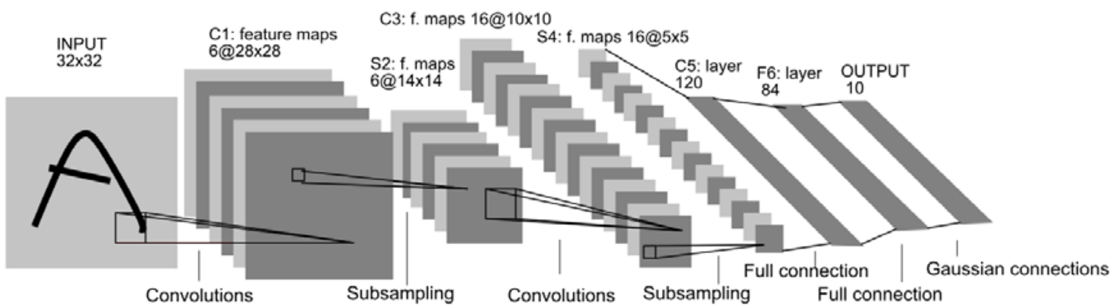
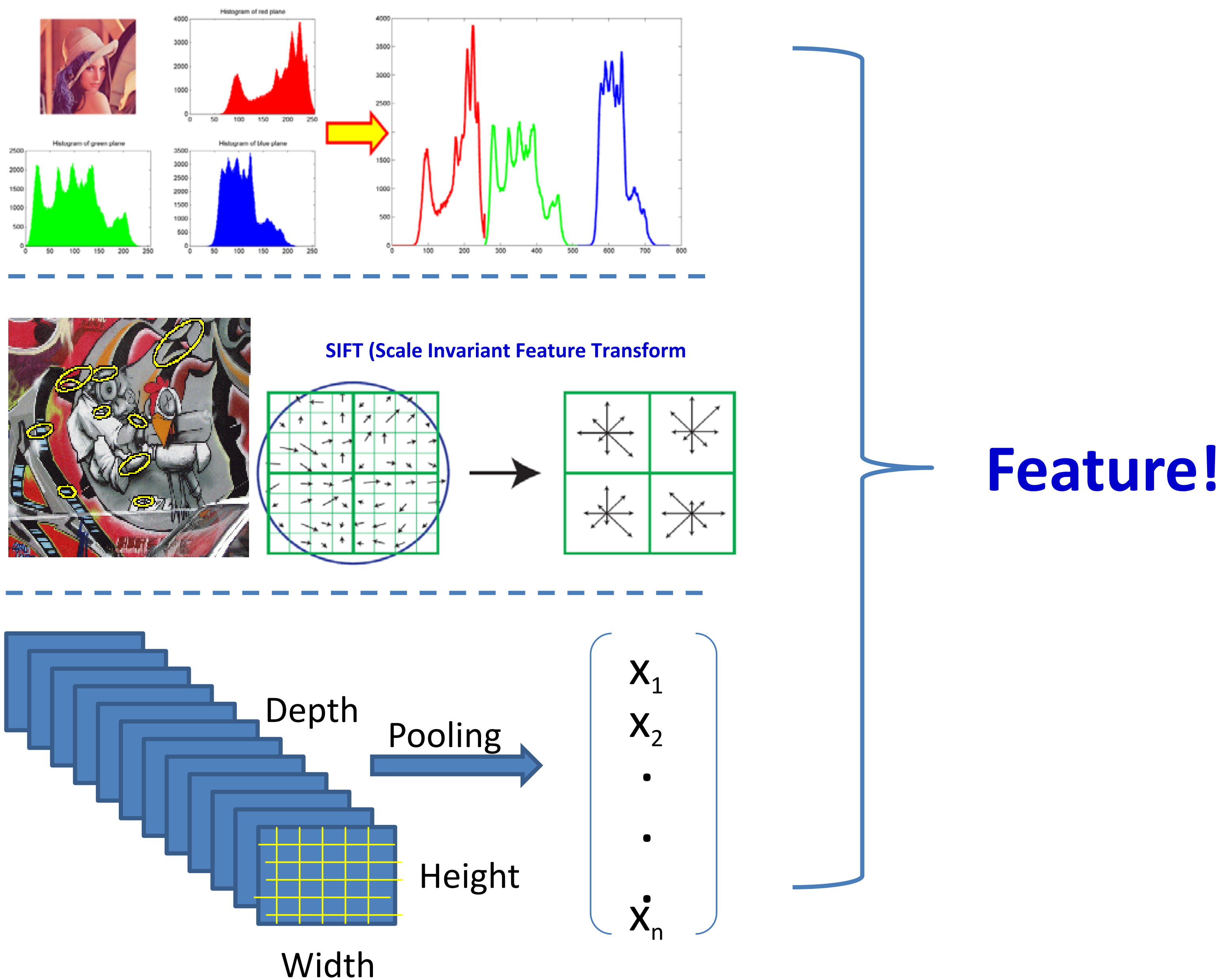


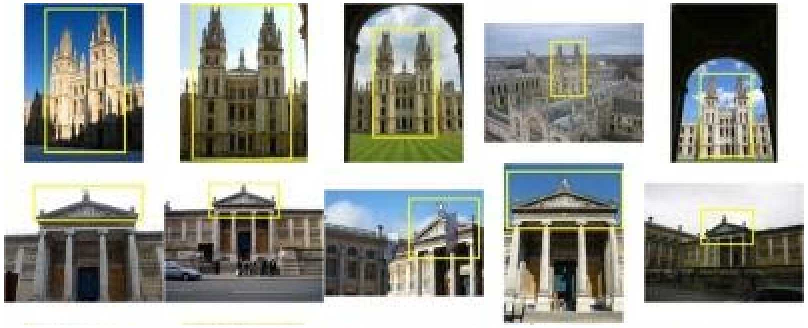
Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

CBIR: Era of Deep Learning



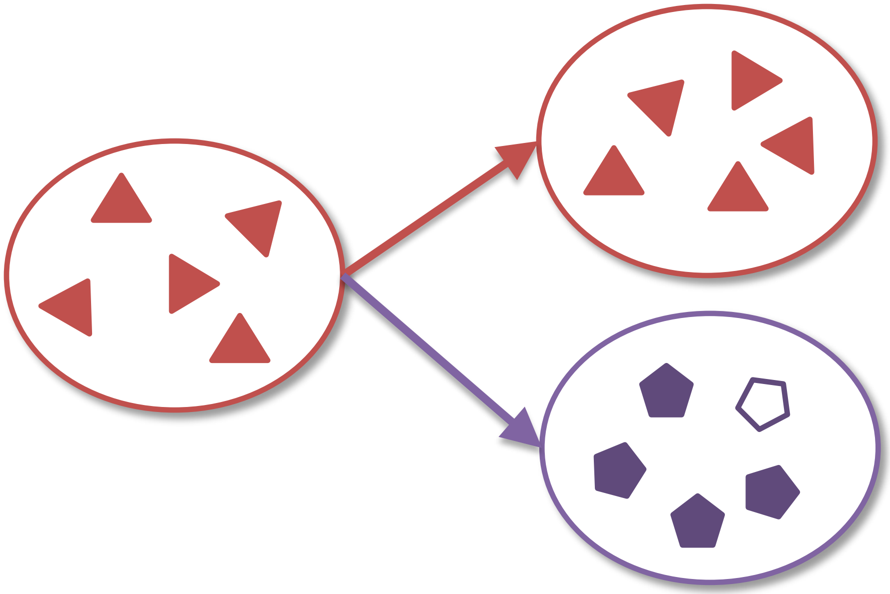
Outline

- Content-based Image Retrieval
- **A Domain Adaptation Perspective**
- Feature Adaptation for CBIR
  - Diffusion Process in image retrieval
  - A kernel mapping view of diffusion
  - Feature Adaptation by modeling diffusion process
- Conclusion



Images courtesy of related papers and authors

Domain Adaptation

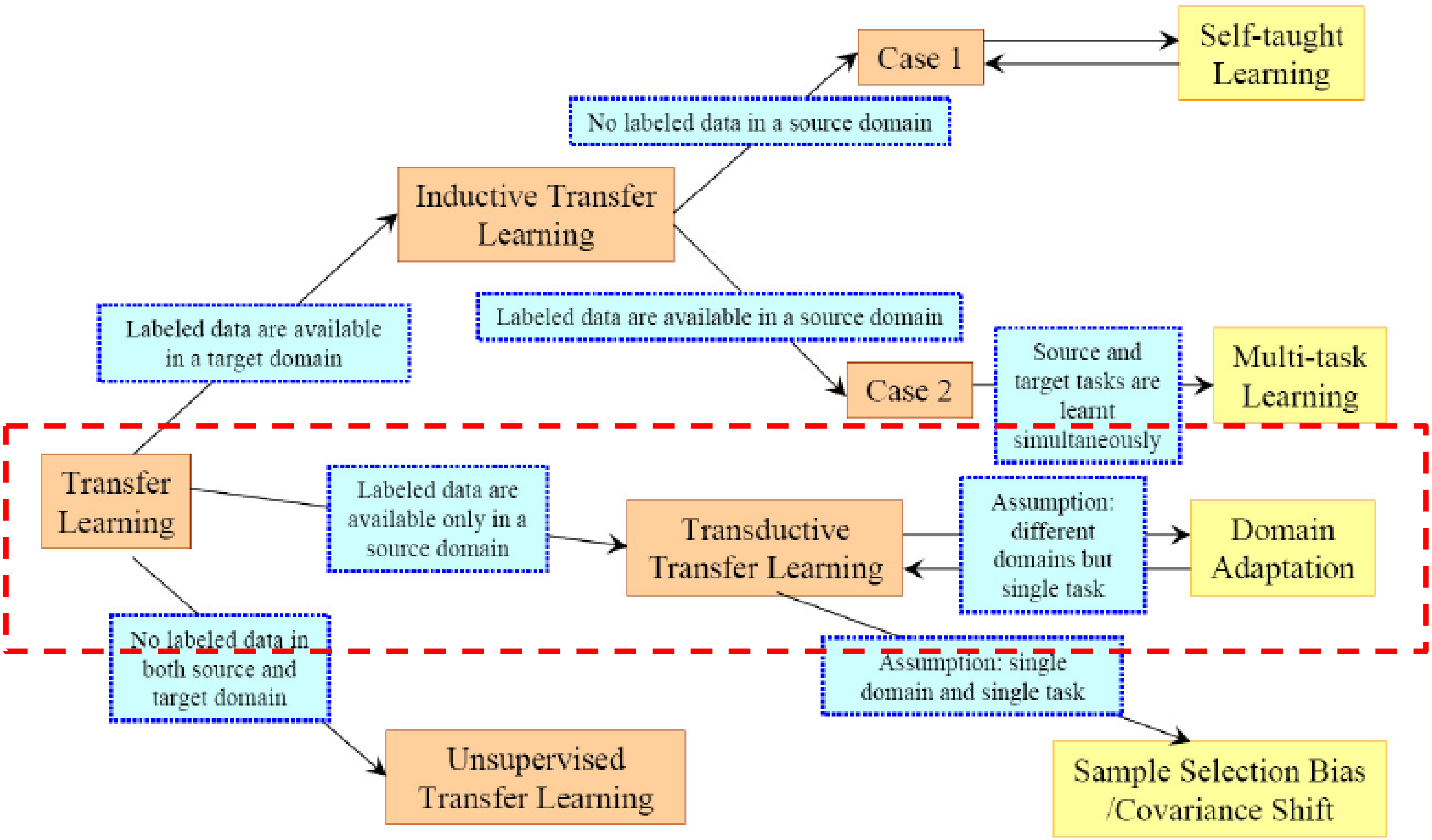


Training and test data
are from the same domain

Training and test data
are from **different domains**

- **Domain:** Probability distribution in a data space
- **Task:** Classification, regression, clustering, retrieval, etc.
- **Aim:** Improve **target task** in the target domain with knowledge from source domain and task

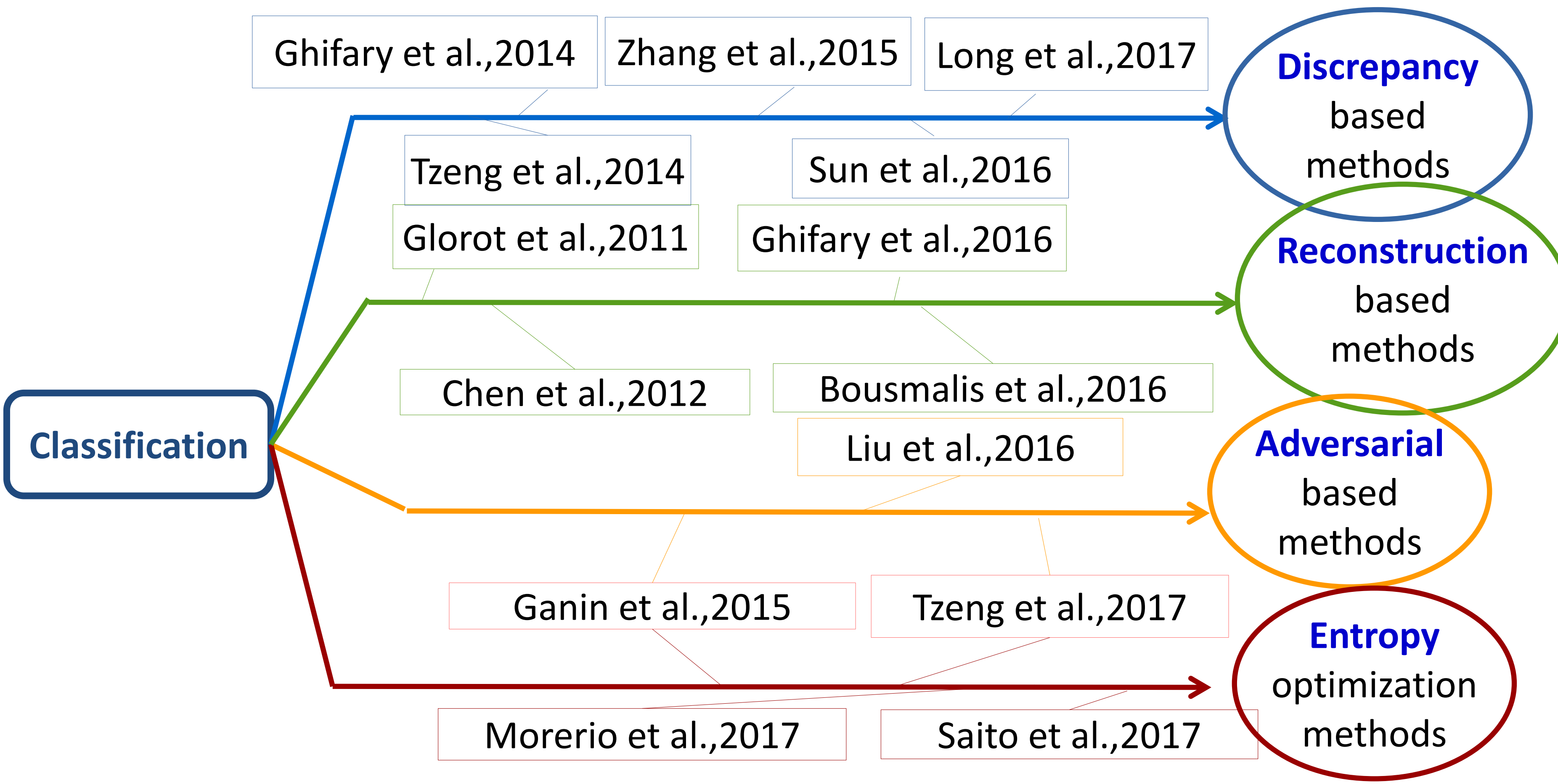
Domain Adaptation



From Sinno Jialin Pan and Qiang Yang. "A survey on transfer learning." IEEE TKDE, 22(10), October 2010, Pages 1345-1359 .

Domain Adaptation

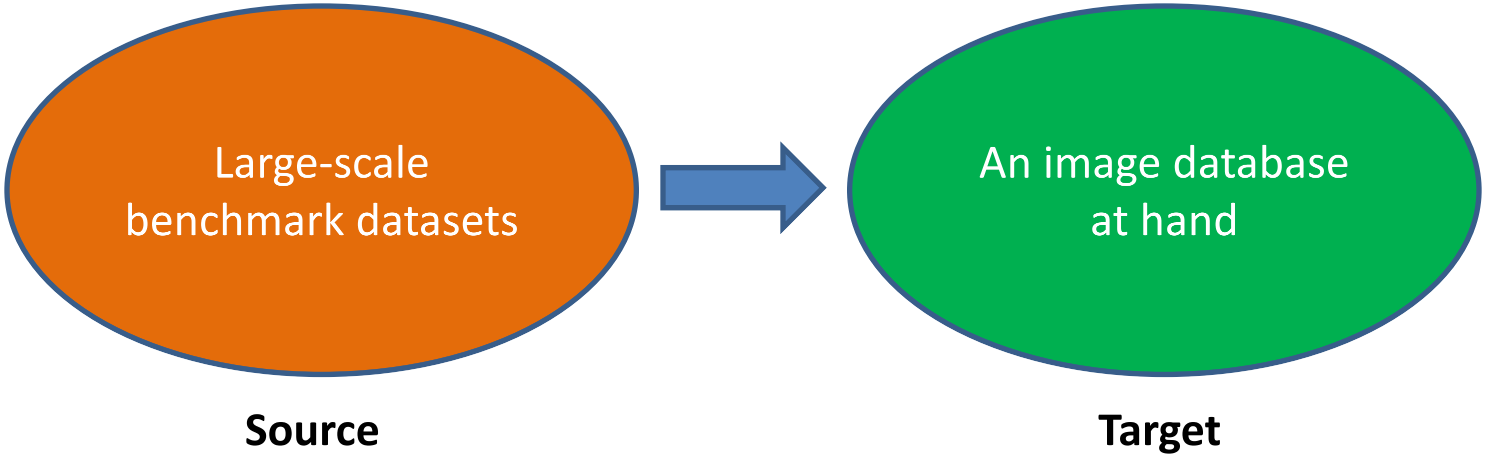
- **Current research** (on image classification)
  - Focus on classification tasks for both source and target



Domain Adaptation

CBIR: A Perspective from Domain Adaptation

- **Domain shift** (p(X) changes)
  - Large-scale benchmark dataset for pre-trained models
  - Image database at hand has a different distribution
- **Task shift** (p(Y|X) changes)
  - (supervised) classification to (unsupervised) retrieval



Domain Adaptation

CBIR: A Perspective from Domain Adaptation

- Currently, use the CNN feature as it is
- Or, fine-tune CNN network
  - Collecting **extra supervision information** for the image database
- Domain adaptation is **not** sufficiently considered

Domain Adaptation

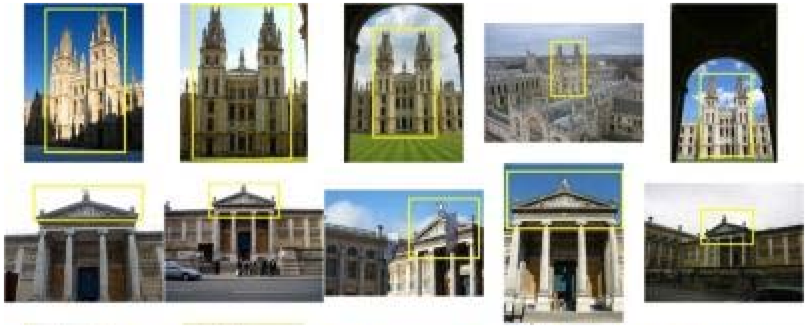
CBIR: A Perspective from Domain Adaptation

- Currently, use the CNN feature as it is
- Or, fine-tune CNN network
  - Collecting **extra supervision information** for the image database
- Domain adaptation is **not** sufficiently considered

So, can we exploit the intrinsic information of an image database to make CNN features adapted to the database?

Outline

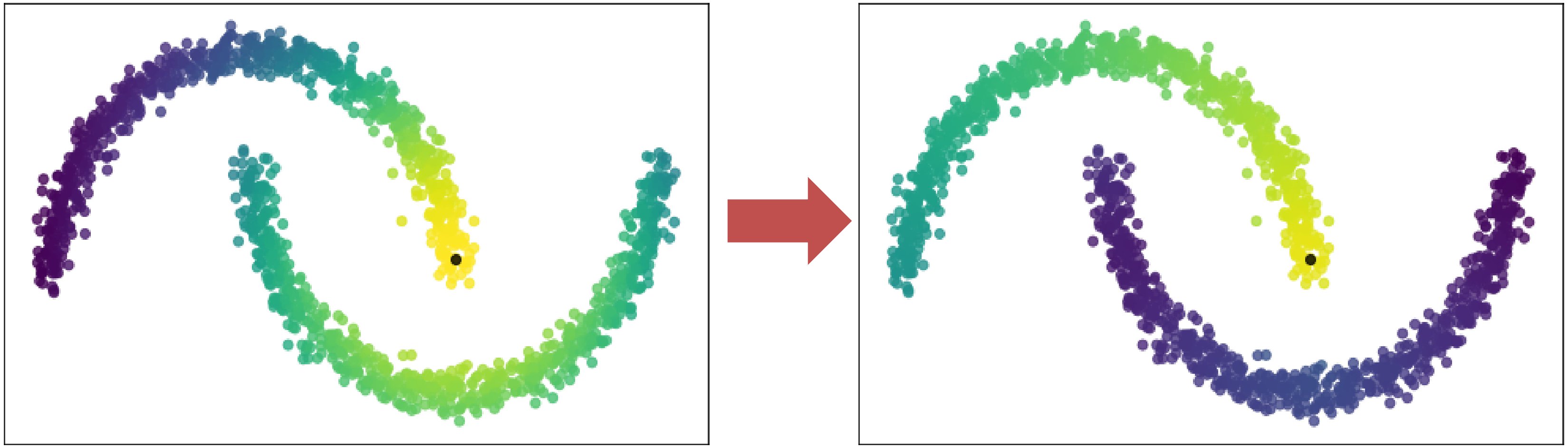
- Content-based Image Retrieval
- A Domain Adaptation Perspective
- **Feature Adaptation for CBIR**
  - **Diffusion Process in image retrieval**
  - A kernel mapping view of diffusion
  - Feature Adaptation by modeling diffusion process
- Conclusion



Images courtesy of related papers and authors

Diffusion Process

- Capture the intrinsic **manifold** structure of data
- Long been used for image retrieval (*)
  - Better evaluate image similarity
  - Unsupervised learning



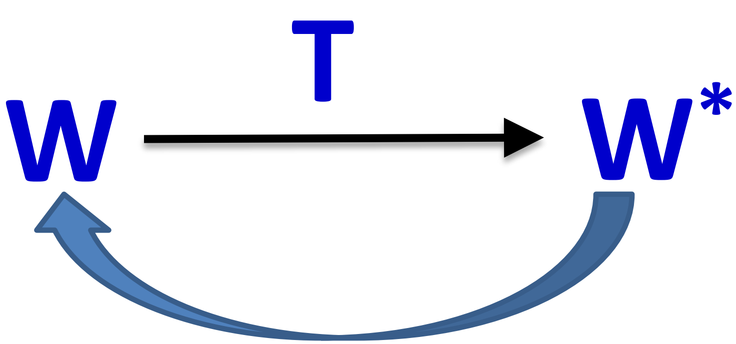
Euclidean distance / Cosine similarity

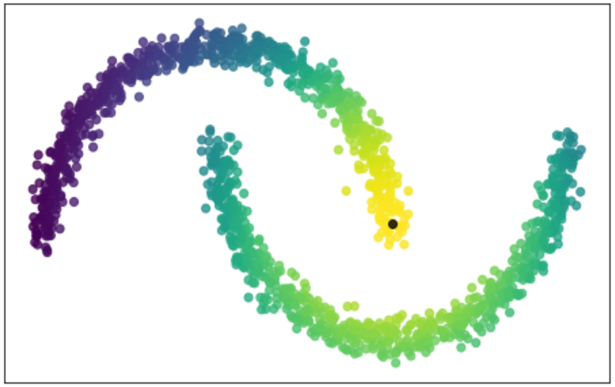
Similarity obtained after **diffusion**

(*) M. Donoser and H. Bischof, "Diffusion Processes for Retrieval Revisited," *2013 IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, 2013, pp. 1320-1327.

Diffusion Process

- **Initial affinity matrix W** **(W_0=A)**
  - Similarity between each pair of images
- **Transition matrix T** (of random walk)
  - Probability for walking from one node to another
- **Performing diffusion**
  - **Iteratively** update **W** through **T**
- **New affinity matrix W^***
  - Similarity scores





Diffusion Process

Method	Abbr.	Initialization $\mathbf{W}^0$	Transition $\mathbf{T}$	Diffusion
Global PageRank [17]	GPR	$\mathbf{u}$	$\mathbf{P}$	$\mathbf{f}_{t+1} = \mathbf{f}_t \, \mathbf{T}$
Personalized PageRank [17]	PPR	$\mathbf{u}$	$\mathbf{P}$	$\mathbf{f}_{t+1} = \alpha \, \mathbf{f}_t \, \mathbf{T} + (1-\alpha) \, \mathbf{y}$
Ranking on Manifolds [23]	ROM	$\mathbf{u}$	$\mathbf{P}_{\mathrm{NC}}$	$\mathbf{f}_{t+1} = \alpha \, \mathbf{f}_t \, \mathbf{T} + (1-\alpha) \, \mathbf{y}$
Label Propagation [24]	LP	$\mathbf{y}$	$\mathbf{P}$	$\mathbf{f}_{t+1} = \mathbf{f}_t \, \mathbf{T}$ and $f(i) = 1$
Graph Transduction [2]	GT	$\mathbf{y}$	$\mathbf{P}$	$\mathbf{f}_{t+1} = \mathbf{f}_t \, \mathbf{T}$ and $f(i) = 1$
Locally Constrained DP [21]	LCDP	$\mathbf{A}$	$\mathbf{P}_{\mathrm{kNN}}$	$\mathbf{W}_{t+1} = \mathbf{T} \, \mathbf{W}_t \, \mathbf{T}^T$
Tensor Graph Diffusion [22]	TGD	$\mathbf{A}$	$\mathbf{P}_{\mathrm{DS}}$	$\mathbf{W}_{t+1} = \mathbf{T} \, \mathbf{W}_t \, \mathbf{T}^T + \mathbf{I}$
Shortest Path Propagation [20]	SPP	$\mathbf{y}$	$\mathbf{P}_{\mathrm{SP}}$	$\mathbf{f}_{t+1} = \mathbf{f}_t \, \mathbf{T}$
Self Smoothing Operator [8]	SSO	$\mathbf{A}$	$\mathbf{P}$	$\mathbf{W}_{t+1} = \mathbf{W}_t \, \mathbf{T}$
Self Diffusion [19]	SD	$\mathbf{A}$	$\mathbf{P}$	$\mathbf{W}_{t+1} = \mathbf{W}_t \, \mathbf{T} + \mathbf{I}$
Replicator Dynamics [18]	RD	$\mathbf{u}$	$\mathbf{A}$	$\mathbf{f}_{t+1} = \mathbf{f}_t \odot \mathbf{T} \, \mathbf{f}_t$ and $\mathbf{f}_{t+1} = \mathbf{f}_{t+1} / \lvert \mathbf{f}_{t+1} \rvert$
Power Iteration Clustering [11]	PIC	$\mathbf{s}$	$\mathbf{P}$	$\mathbf{f}_{t+1} = \mathbf{T} \, \mathbf{f}_t$ and $\mathbf{f}_{t+1} = \mathbf{f}_{t+1} / \lvert \mathbf{f}_{t+1} \rvert$
Authority Shift Clustering [3]	ASC	$\mathbf{P}_{\mathrm{PPR}}$	$\mathbf{P}_{\mathrm{PPR}}$	$\mathbf{W}_{t+1} = \mathbf{W}_t \, \mathbf{T}$

- Computational **efficiency**
  - A direct matrix inversion
  - An iterative method
  - Graph sparsification
  - Conjugate gradient (Iscen et al. CVPR2017)

Table from M. Donoser and H. Bischof, "Diffusion Processes for Retrieval Revisited," 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, 2013, pp. 1320-1327.

Diffusion Process

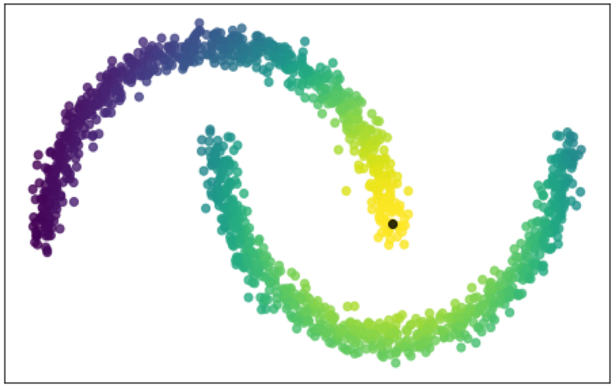
The performance of diffusion-based image retrieval in the recent literature (*)

Method	$m \times d$	INSTRE	Oxf5k	Oxf105k	Par6k	Par106k
Global descriptors - nearest neighbor search						
CroW [30]†	512	-	68.2	63.2	79.8	71.0
R-MAC [43]	512	47.7	77.7	70.1	84.1	76.8
R-MAC [19]	2,048	62.6	83.9	80.8	93.8	89.9
NetVLAD [1]†	4,096	-	71.6	-	79.7	-
Global descriptors - query expansion						
R-MAC [43]+AQE [8]	512	57.3	85.4	79.7	88.4	83.5
R-MAC [43]+SCSM [48]	512	60.1	85.3	80.5	89.4	84.5
R-MAC [43]+HN [42]	512	64.7	79.9	-	92.0	-
Global diffusion	512	70.3	85.7	82.7	94.1	92.5
R-MAC [19]+AQE [8]	2,048	70.5	89.6	88.3	95.3	92.7
R-MAC [19]+SCSM [48]	2,048	71.4	89.1	87.3	95.4	92.5
Global diffusion	2,048	80.5	87.1	87.4	96.5	95.4

(* Table from Iscen et al., Efficient Diffusion on Region Manifolds: Recovering Small Objects with Compact CNN Representations, CVPR2017)

Diffusion Process

- **Why bother feature adaptation?**
  - From the perspective of domain adaptation
  - The issue of diffusion-based image retrieval
    - Have to maintain a **large** affinity matrix **W**
    - Need to **update W** with newly inserted images
    - Need to perform **online** diffusion for retrieval
- **After feature adaptation**
  - A simple **Euclidean** search
  - No need to store **W**
  - No need to update (partially)
  - No need online diffusion

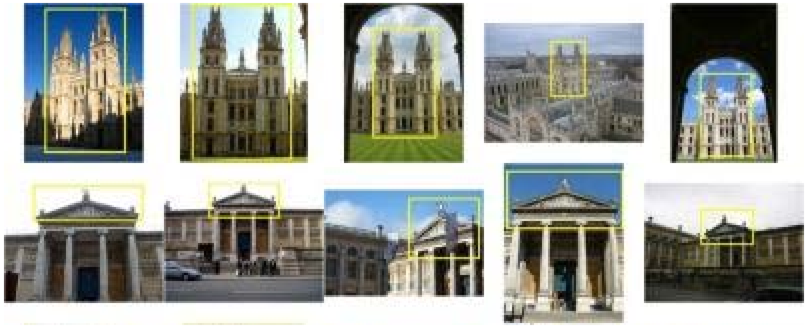


Diffusion Process

- **Why bother feature adaptation?**
  - From the perspective of domain adaptation
  - The issue of diffusion-based image retrieval
    - Have to maintain a **large** affinity matrix **W**
    - Need to **update W** with newly inserted images
    - Need to perform **online** diffusion for retrieval
- **After feature adaptation**
  - An **unsupervised** learning framework to **bootstrap** image retrieval
- Two related work
  - Iscen et. al, CVPR18a; Iscen et. al, CVPR18b

Outline

- Content-based Image Retrieval
- A Domain Adaptation Perspective
- Feature Adaptation for CBIR
  - Diffusion Process in image retrieval
  - **A kernel mapping view of diffusion**
  - Feature Adaptation by modeling diffusion process
- Conclusion



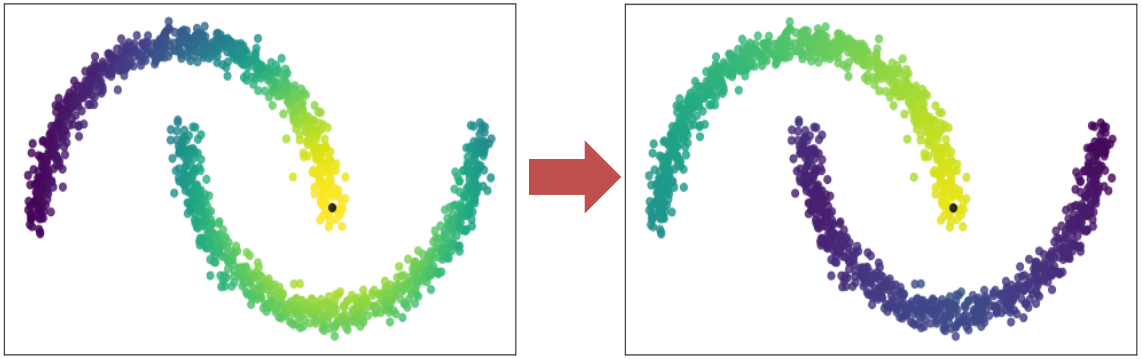
Images courtesy of related papers and authors

Kernel mapping view

- **Feature adaptation is a mapping**

$$\mathbf{f} \xrightarrow{\phi} \mathbf{f}'$$

- **Conceptually, is there such a ϕ w.r.t. diffusion?**
  - Yes, a diffusion process essentially evaluates image similarity with a new kernel
  - ϕ: the implicit, nonlinear kernel-induced mapping



Kernel mapping view

- **A typical diffusion scheme (LCDP)**

$$\mathbf{W}_{t+1} = \mathbf{T}\mathbf{W}_t\mathbf{T}^\top$$

Initial affinity matrix

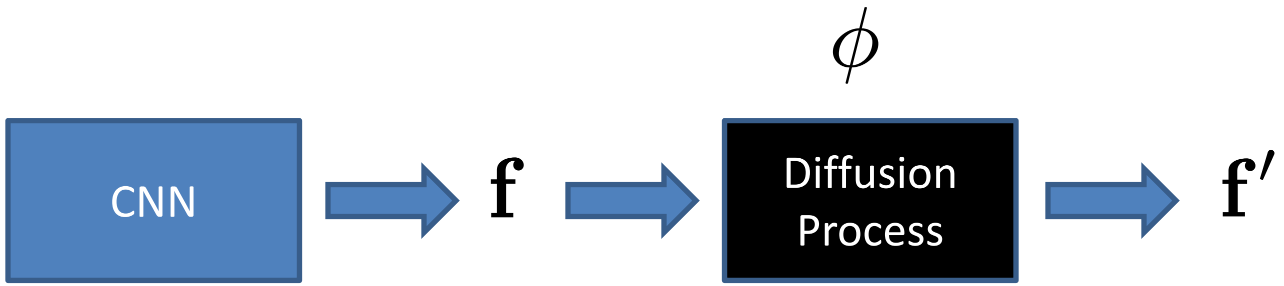
$$\mathbf{W}_{t+1} = \mathbf{T}^{t+1}\mathbf{W}_0(\mathbf{T}^{t+1})^\top = \mathbf{T}^{t+1}\mathbf{A}(\mathbf{T}^{t+1})^\top$$

$$\underline{\kappa(\mathbf{x}_i, \mathbf{x}_j | \mathbf{A})} \triangleq \mathbf{W}_{t+1}(i,j) = \mathbf{T}^{t+1}(i,:)\mathbf{A}(\mathbf{T}^{t+1}(j,:))^\top$$

- **Diffusion process uses a "context-aware" kernel**

“Modeling” diffusion process

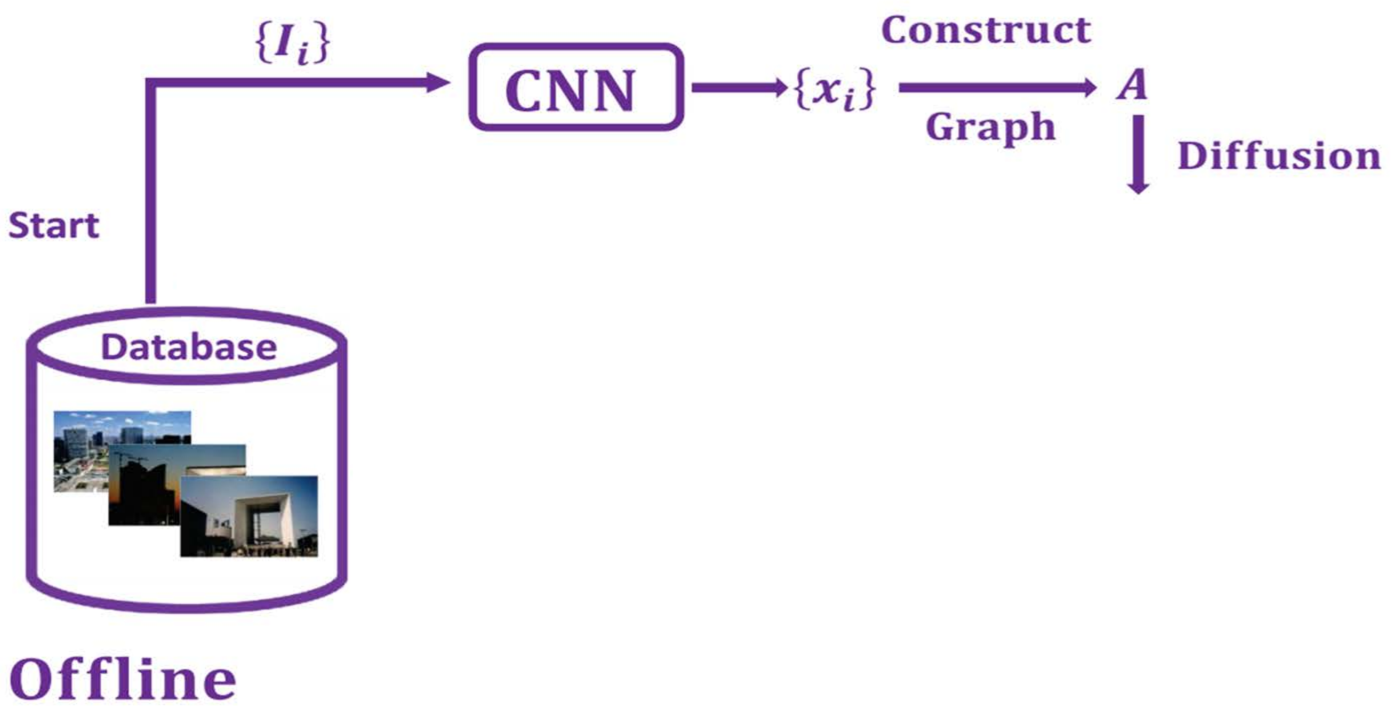
- **Learning ϕ by treating diffusion process as a “black box”**



- **Implemented by making A approach $\mathbf{W}^*$**
  - **Value-based approximation**
    - **Inner product or Euclidean distance (scale issue)**
  - **Rank-based approximation (good for image retrieval)**

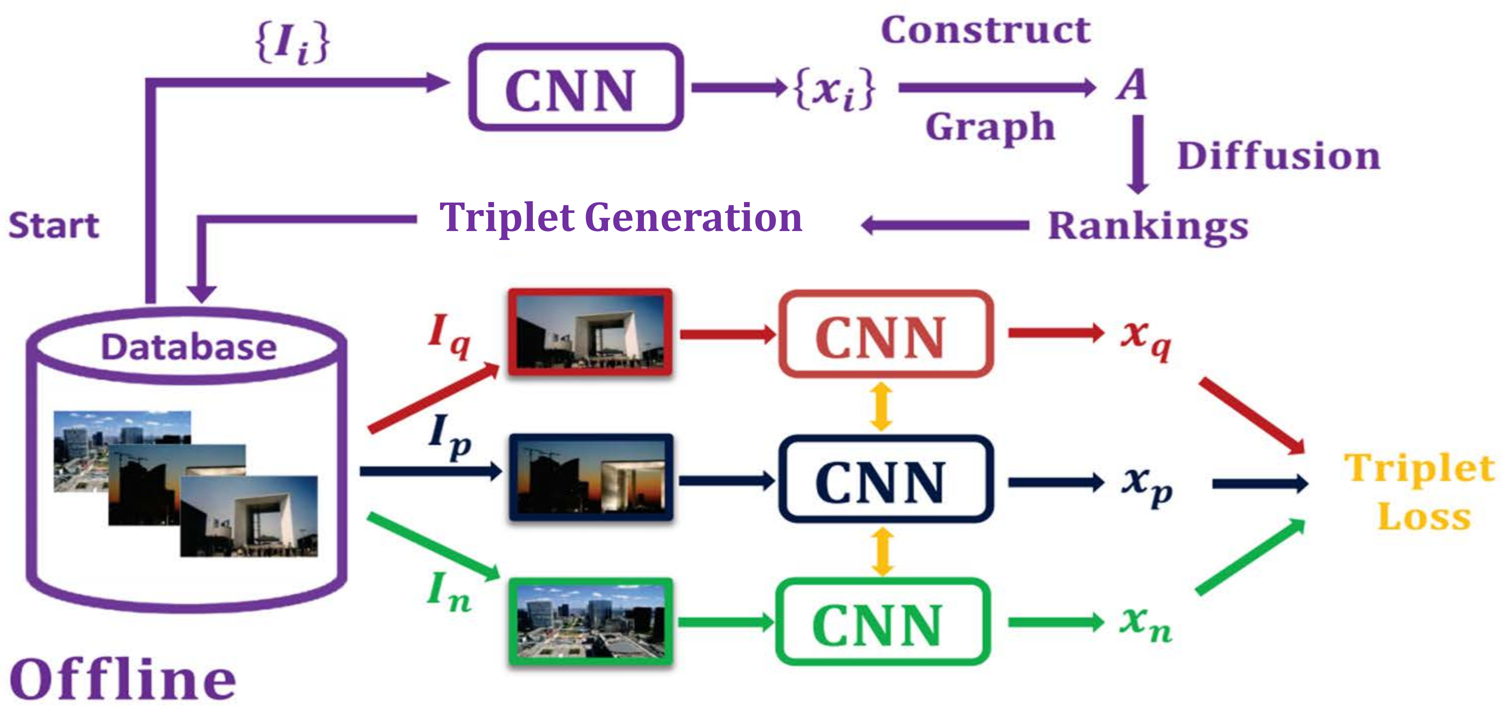
"Modeling" diffusion process

- **A deep metric learning approach**



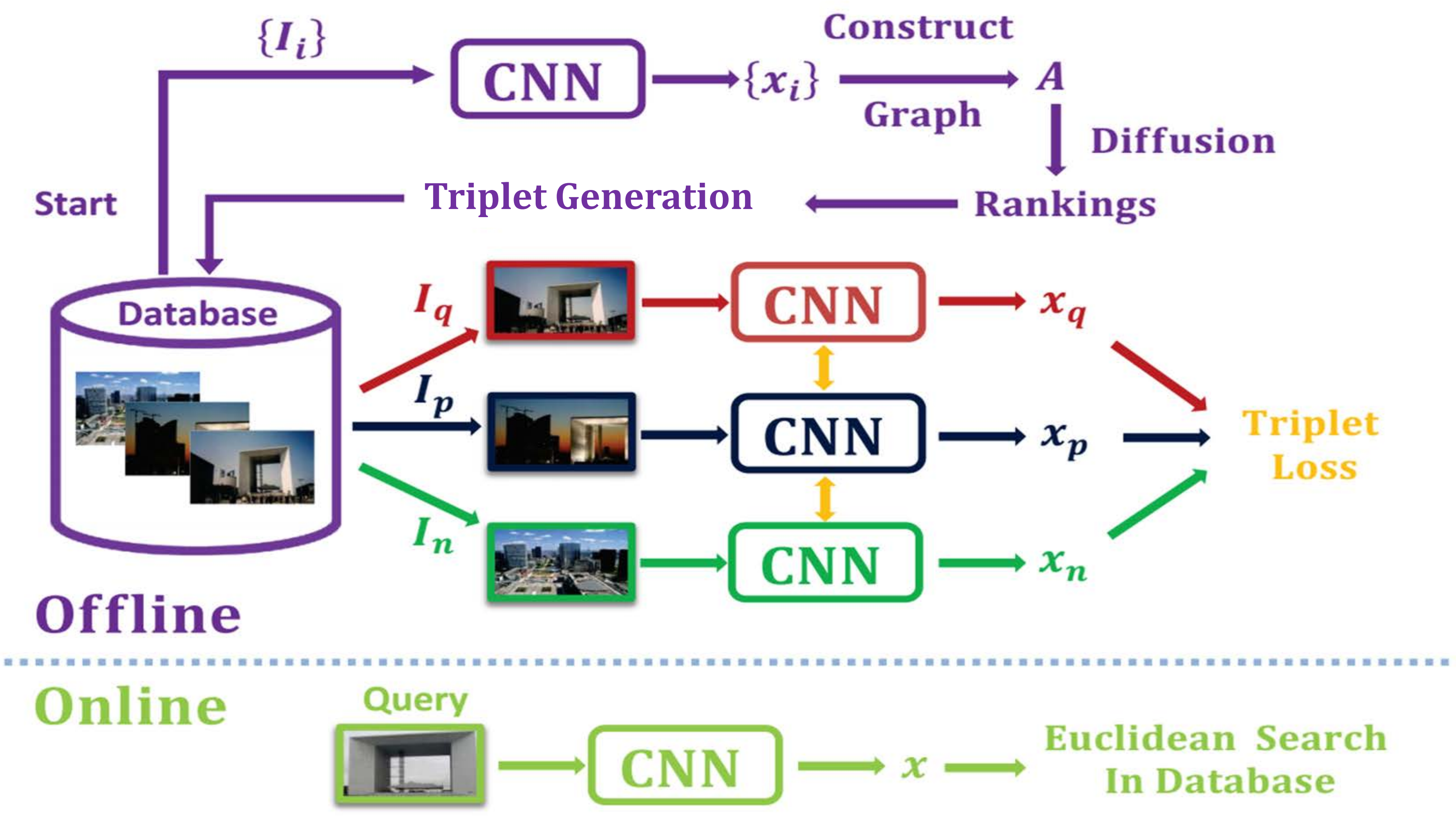
“Modeling” diffusion process

- **A deep metric learning approach**



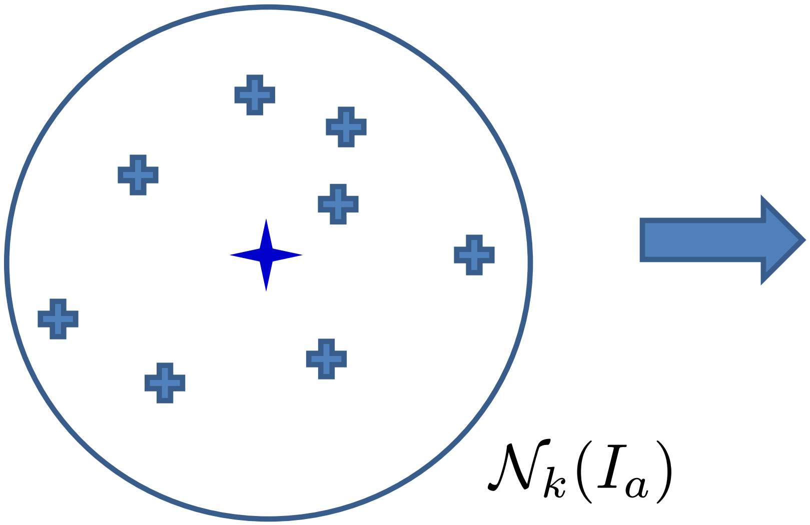
"Modeling" diffusion process

- **A deep metric learning approach**



"Modeling" diffusion process

- **Triplet generation**



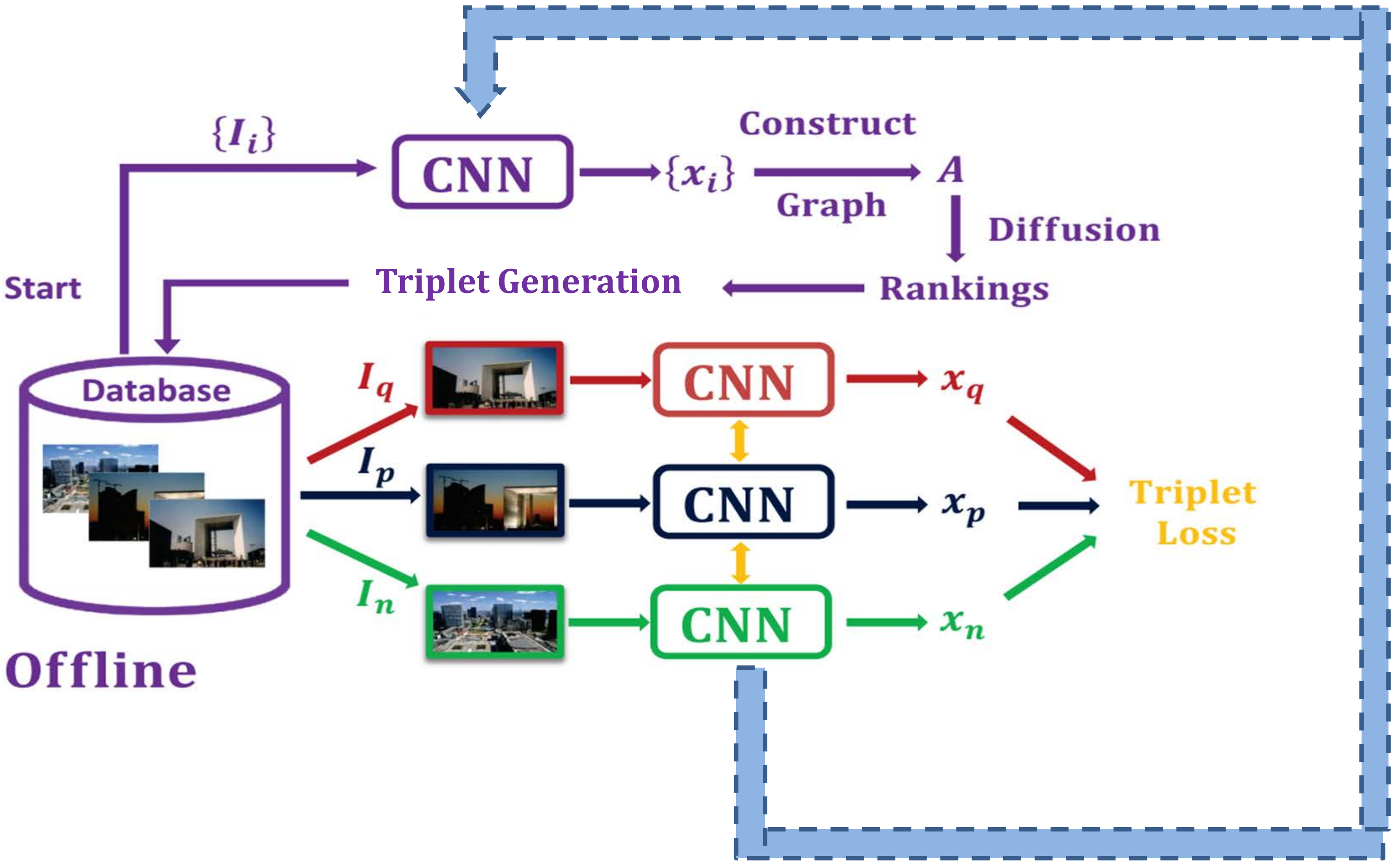
$$\mathcal{S} = \{(\overset{\text{Anchor}}{I_a}, \underset{\text{Closer}}{I_c}, \underset{\text{Farther}}{I_f})\}$$

- **Triplet loss function**

$$L = \sum_{(I_a, I_c, I_f) \in \mathcal{S}} \left[d(I_a, I_c) - d(I_a, I_f) + \frac{|r_f - r_c|}{k} m_0 \right]_+$$

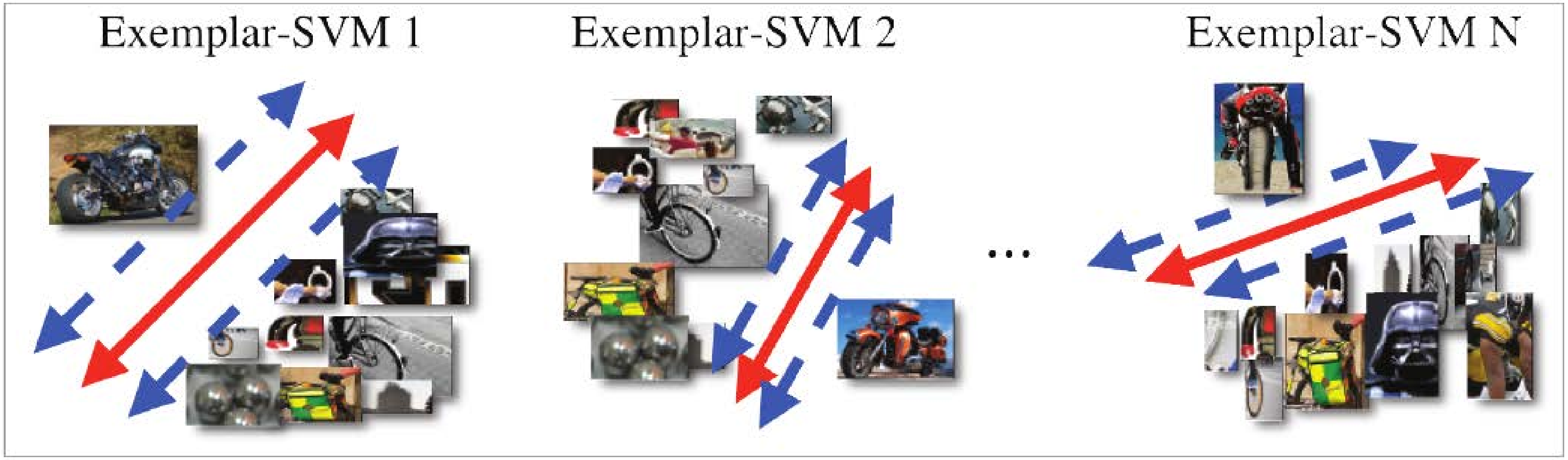
“Modeling” diffusion process

- **Finally, an unsupervised bootstrapping framework**



“Modeling” diffusion process

- **Ways other than diffusion process to utilize data distribution information**
  - Exemplar-SVM: each image in the database is used as the only positive sample to train an SVM



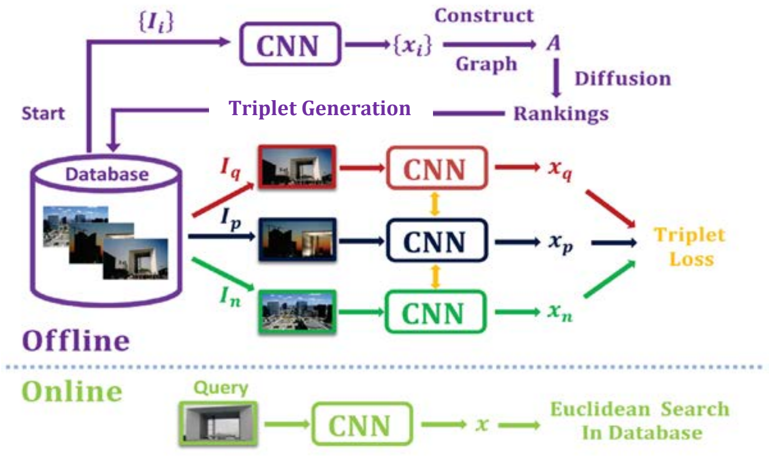
1. Ensemble of Exemplar-SVMs for Object Detection and Beyond, Malisiewicz et al. ICCV 2011
2. Instance Image Retrieval by Aggregating Sample-based Discriminative Characteristics, Zhang et al. ICMR 2018

Experimental Result

- **Datasets**
  - Oxford5k, Pairs6k, Oxford105k, Pairs106k, INSTRE, and Sculpture
  - Diffusion process is performed on gallery images only
  - Query images are exclusively reserved for evaluation
- **Experimental setting**
  - ResNet101 pre-trained on ImageNet
  - M = 0.1 and k = 300
  - R-MAC feature representation
  - LCDP diffusion process ($\mathbf{W}_{t+1} = \mathbf{T}\mathbf{W}_t\mathbf{T}^\top$)

Experimental Result

- **Five tasks**
  1. Comparison using **global** representations
  2. Comparison using **regional** representations
  3. Comparison with the state-of-the-art methods
  4. Time and memory cost
  5. Properties (Image insertion, iterative training)



Experimental Result

- Task 1: Comparison using global representations

Method (mAP)	Oxford5k	Paris6k	Oxford105k	Paris106k	INSTRE	Sculpture
R-MAC+E (global)	58.5	73.3	57.3	67.4	37.7	51.0
R-MAC+D (global)	63.1	83.5	62.0	77.0	53.0	61.6
Proposed (global)	**63.2**	**89.6**	**62.4**	**82.6**	**54.5**	**75.5**

- Task 2: Comparison using regional representations

Pairs6k dataset

Method	Cross-region Matching	Regional Diffusion	Proposed
mAP	84.4	91.8	**93.8**

Achieve **comparable or better** retrieval than the diffusion process

Experimental Result

- Task 3: Comparison with the state-of-the-art methods
  - Methods with **Euclidean** search
  - Achieve higher retrieval accuracy

Method	Dim.	Oxford5k	Paris6k	Oxford105k	Paris106k	INSTRE
Global image representation with Euclidean search						
[16]	128	43.3	-	35.3	-	-
[5]	128	55.7	-	52.3	-	-
[31]	128	59.3	59.0	-	-	-
[4]	256	53.1	-	50.1	-	-
[28]	512	66.9	83.0	61.6	75.7	-
[17]	512	68.2	79.7	63.3	71	-
[13]*	512	77.7	84.1	70.1	76.8	47.7
[15]	512	78.2	85.1	72.6	78.0	57.7
[22]	512	79.7	83.8	73.9	76.4	-
[16]	1024	56.0	-	50.2	-	-
[28]	2048	69.4	85.2	63.7	77.8	-
[11]	2048	86.1	94.5	82.8	90.6	-
[13]⋆	2048	83.9	93.8	80.8	89.9	62.6
[2]	4096	71.6	79.7	-	-	-
Our global image representation (by modelling diffusion process) + Euclidean search						
Proposed	2048	85.4	**96.3**	85.1	**94.7**	**71.7**

Experimental Result

- Task 3: Comparison with the state-of-the-art methods
  - Methods using diffusion et. al.
  - Achieve higher computational efficiency
  - Achieve competitive retrieval accuracy

Method	Dim.	Oxford5k	Paris6k	Oxford105k	Paris106k	INSTRE
Global image representation + diffusion / query expansion / matching / verification						
[17]	-	72.2	85.5	67.8	79.7	-
[24]	-	75.2	74.1	72.9	-	-
[21]	-	81.4	80.3	76.7	-	-
[7]	-	82.7	80.5	76.7	71.0	-
[9]	-	84.3	83.4	80.2	-	-
[18]	-	84.9	82.4	79.5	77.3	-
[27]	-	86.9	85.1	85.3	-	-
[26]	-	89.4	82.8	84.0	-	-
[28]	512	77.3	86.5	73.2	79.8	-
[3]	512	79.0	85.1	-	-	-
[22]	512	84.5	86.4	80.4	79.7	-
[13]*	512	85.4	88.4	79.7	83.5	57.3
[28]	2048	78.9	89.7	75.5	85.3	-
[13]	2048	87.1	**96.5**	87.4	**95.4**	**80.5**
[11]	2048	**90.6**	96.0	**89.4**	93.2	-
[13]*	2048	**89.6**	95.3	**88.3**	92.7	70.5
[14]	2048	**87.5**	**96.4**	**87.9**	**95.3**	**80.5**
Our global image representation (by modelling diffusion process) + Euclidean search						
Proposed	2048	85.4	**96.3**	85.1	**94.7**	**71.7**

Experimental Result

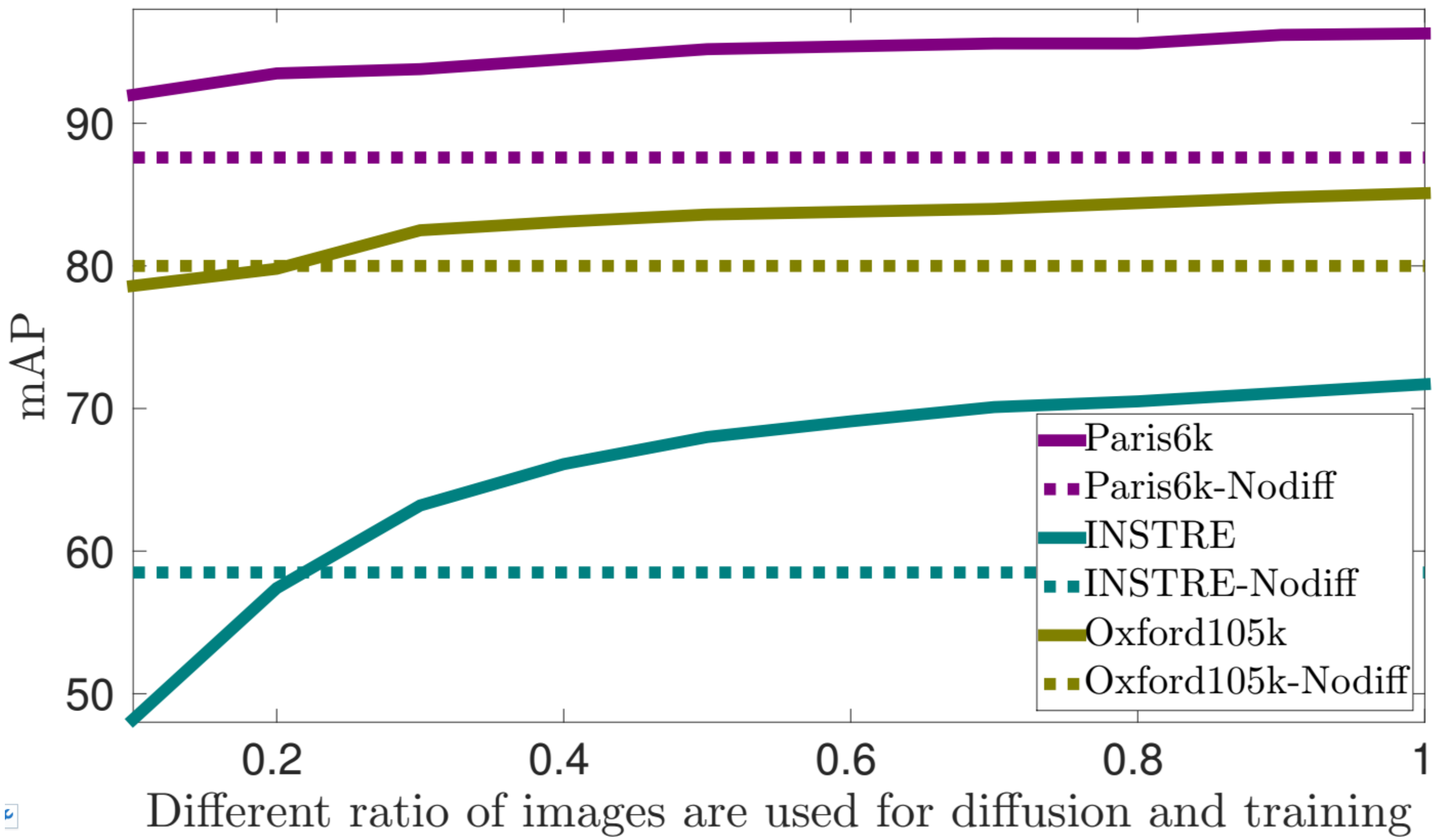
- Task 4: Time and memory cost
  - No need extra memory to store the affinity matrix A
  - Consistently faster (10 times or more) in online retrieval

Comparison of average time / memory usage (**Second / GB**) in online retrieval

	Global feature representation			Regional feature representation		
Dataset	Oxford5k	INSTRE	Oxford105k	Oxford5k	INSTRE	Oxford105k
Diff. based	0.020/0.01	0.100/0.03	2.90/0.11	0.6/0.1	2.9/0.6	13.0/2.1
Proposed	0.002/N.A.	0.011/N.A.	0.03/N.A.	0.1/N.A.	0.4/N.A.	1.43/N.A.

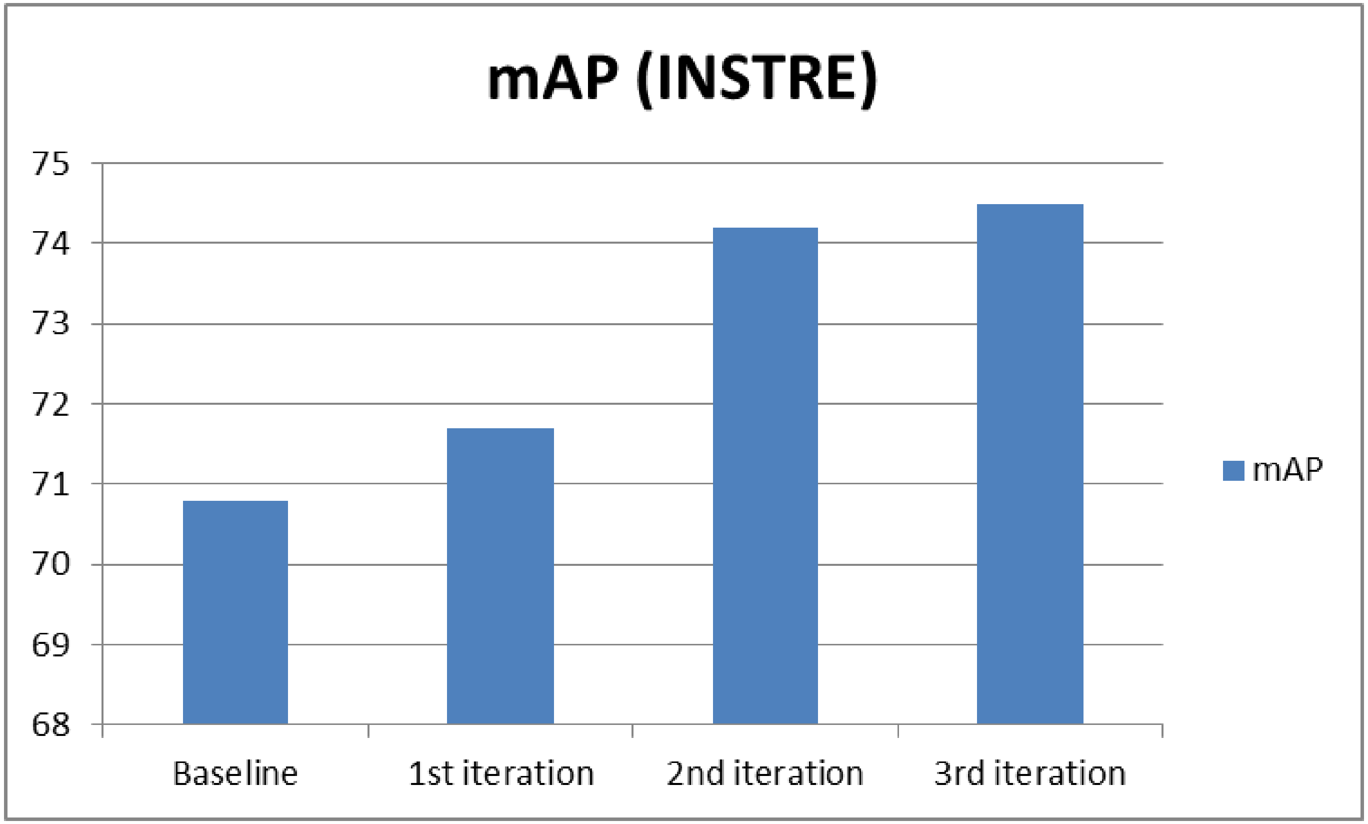
Experimental Result

- Task 5: Properties (robustness to new image insertion)
  - What if new images are inserted?
  - Do we need to redo diffusion immediately?



Experimental Result

- Task 5: Properties (Iterative training)
  - Obtain **better retrieval** by extra one or two iterations
  - A **gradual** feature adaptation



Conclusion

- **Adapt** pre-trained CNN features to new image datasets
- Utilize the unprecedented **modelling** capability of DNN
- Improve retrieval **without** using additional labels, extra information, or external datasets
- An **unsupervised** framework to bootstrap image retrieval

- But, a data-specific approach
- Not yet explicitly resolve the gap between domains
- Any more direct approach other than diffusion process
- Computational efficiency for large image databases

Key references

1. M. Donoser and H. Bischof, "Diffusion Processes for Retrieval Revisited," CVPR 2013.
2. Ahmet Iscen, Giorgos Tolias, Yannis S Avrithis, Teddy Furon, and Ondrej Chum. Efficient diffusion on region manifolds: Recovering small objects with compact CNN representations. CVPR 2017.
3. Ahmet Iscen, Giorgos Tolias, Yannis S Avrithis, and Ondrej Chum. Mining on manifolds: Metric learning without labels. CVPR 2018.
4. Yan Zhao, Lei Wang, Luping Zhou, Yinghuan Shi, Yang Gao. Modelling Diffusion Process by Deep Neural Networks for Image Retrieval. BMVC 2018.
5. Zhongyan Zhang, Lei Wang, Yang Wang, Luping Zhou, Jianjia Zhang, Fang Chen. Instance Image Retrieval by Aggregating Sample-based Discriminative Characteristics. ICMR 2018.

Q & A

Images Courtesy of Google Image