SMART School on Computational Social and Behavioral Sciences

Paris, August 31 - September 4, 2015

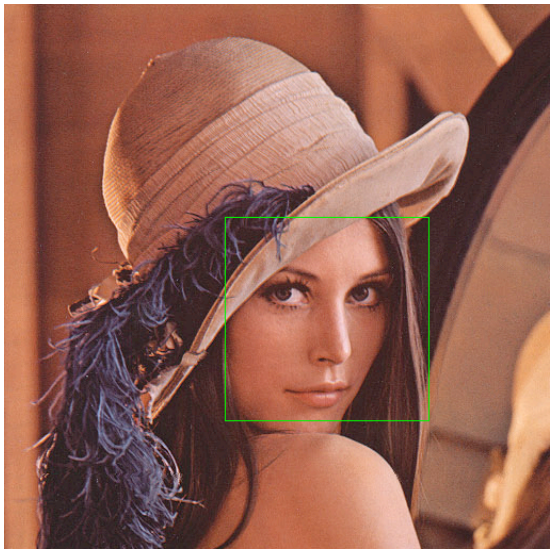
Metric Learning in Computer Vision

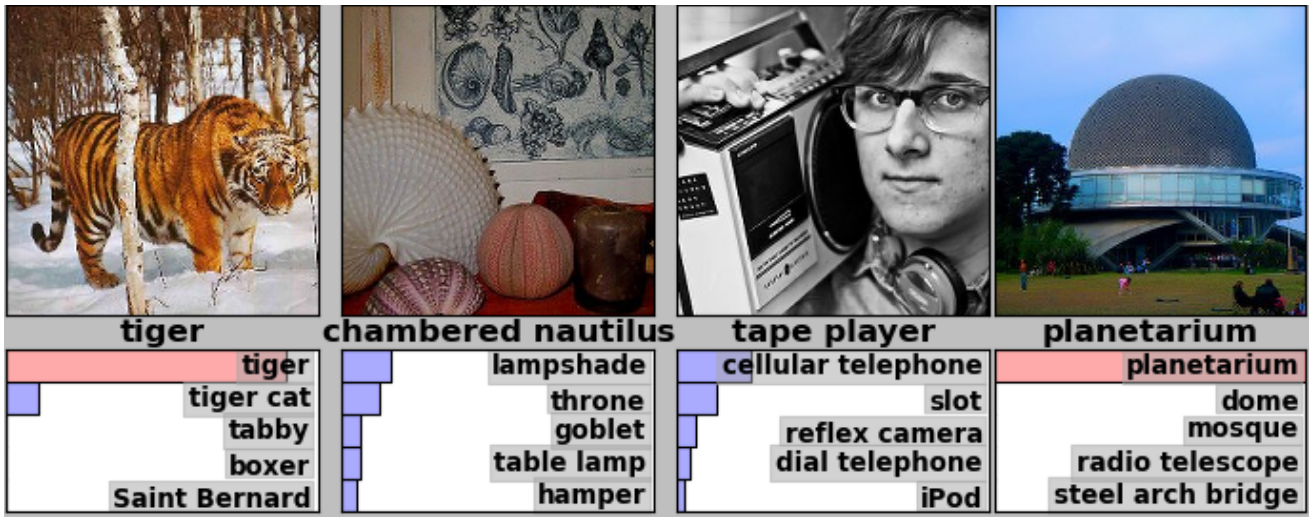
Prof. Matthieu Cord

LIP6 - Computer Science Department
UPMC PARIS 6 - Sorbonne University

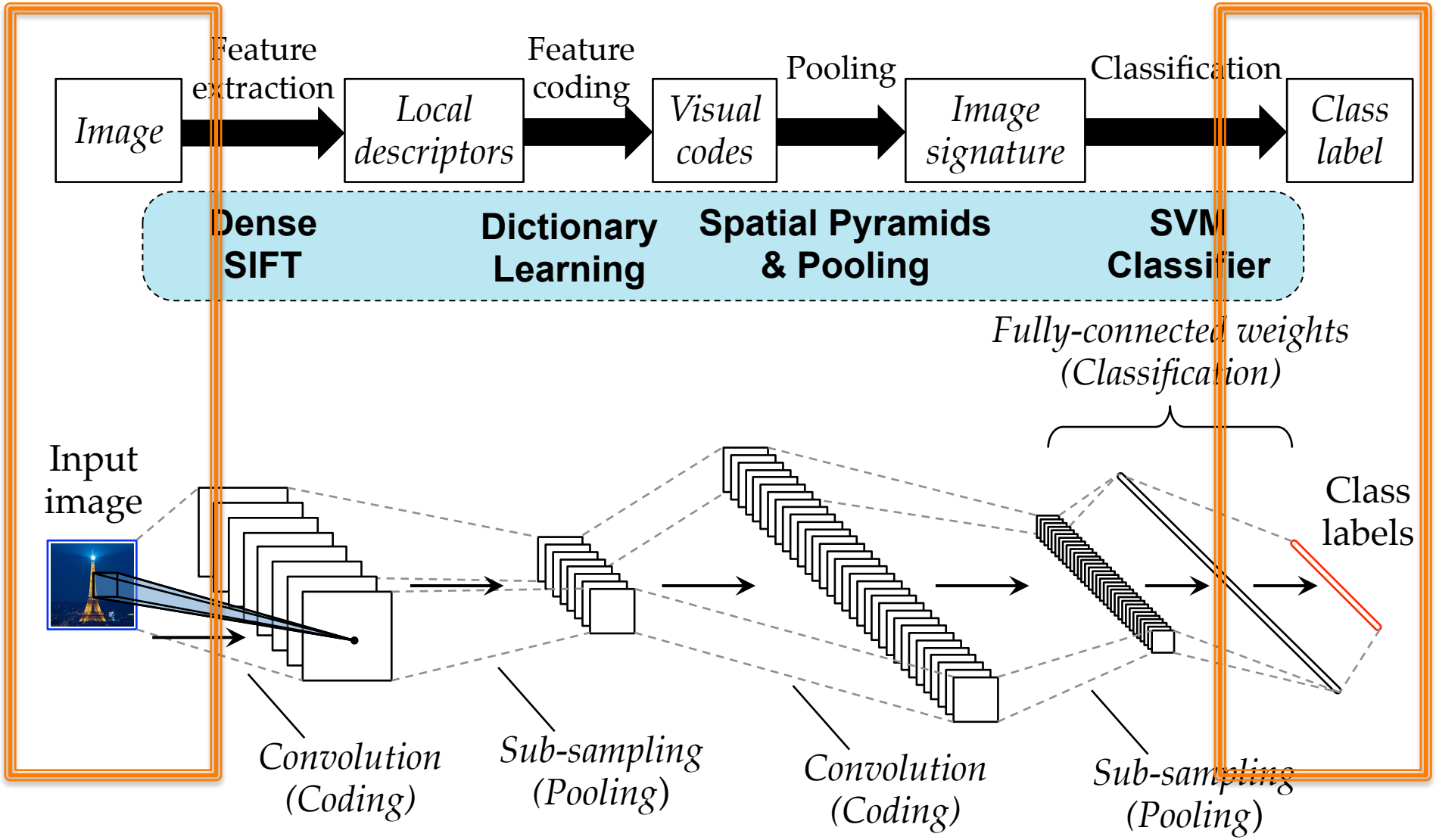
Introduction: Visual learning

- A lot of recent successful applications of Machine Learning to Visual Understanding
- Supervised classification on large dataset ImageNet [winner 2012]
  - 1M images
  - 1000 classes



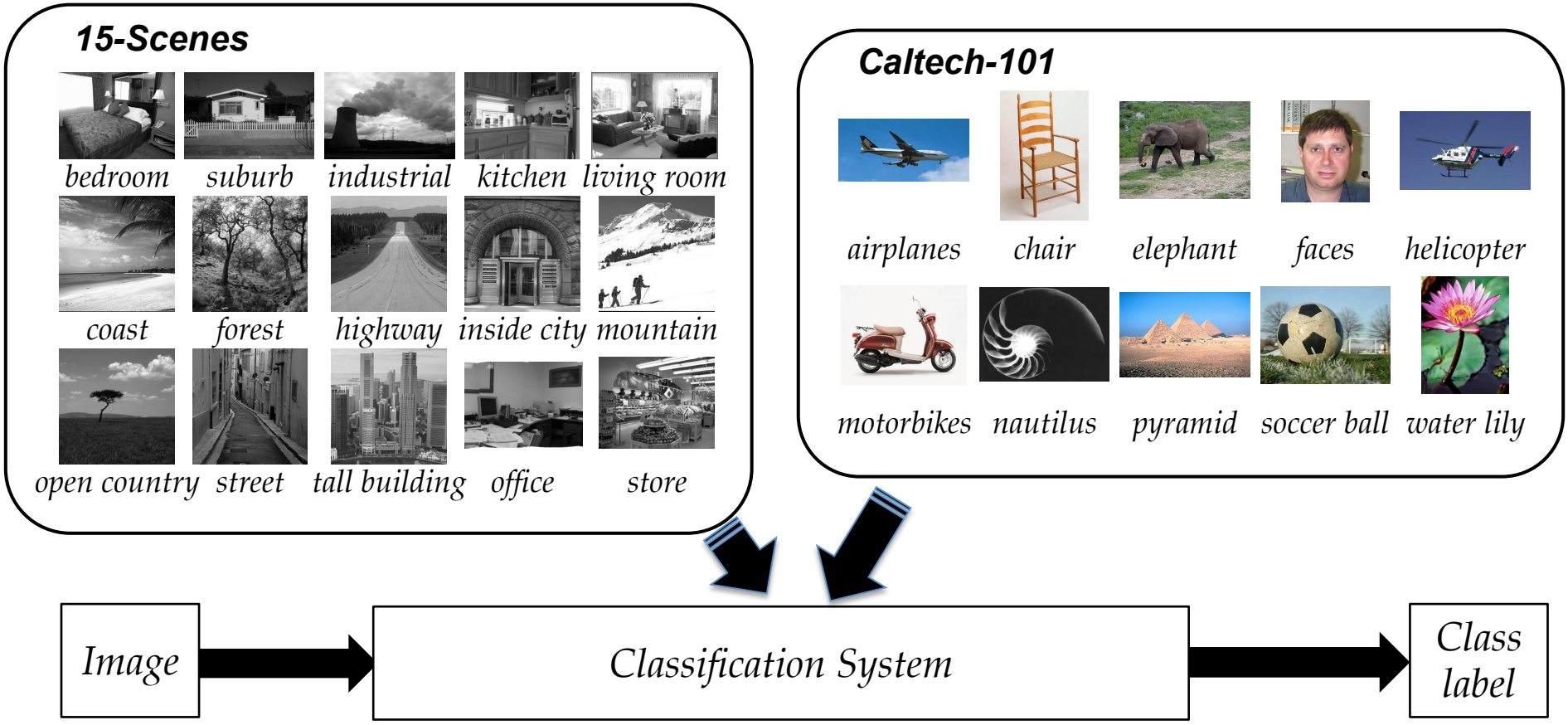


Introduction: Visual learning



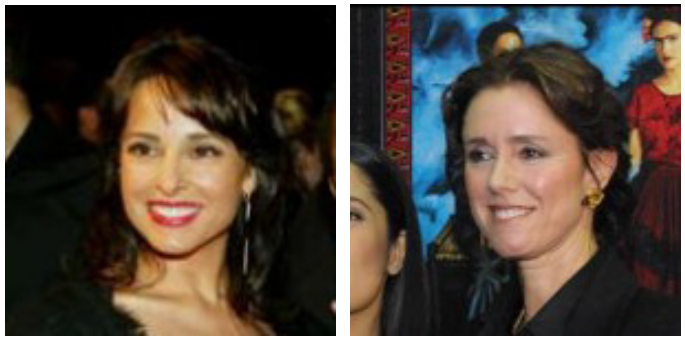
Introduction: Visual learning

- Data for training

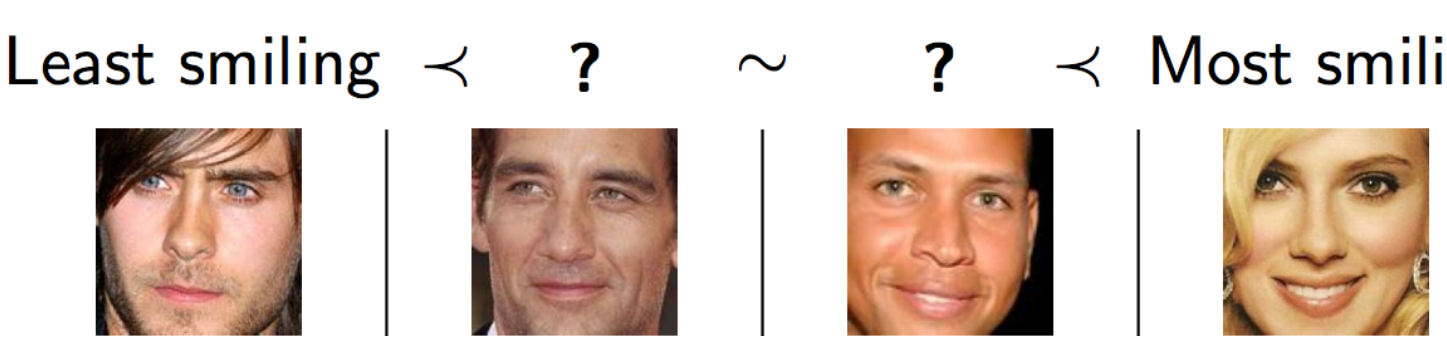


Introduction: Visual learning

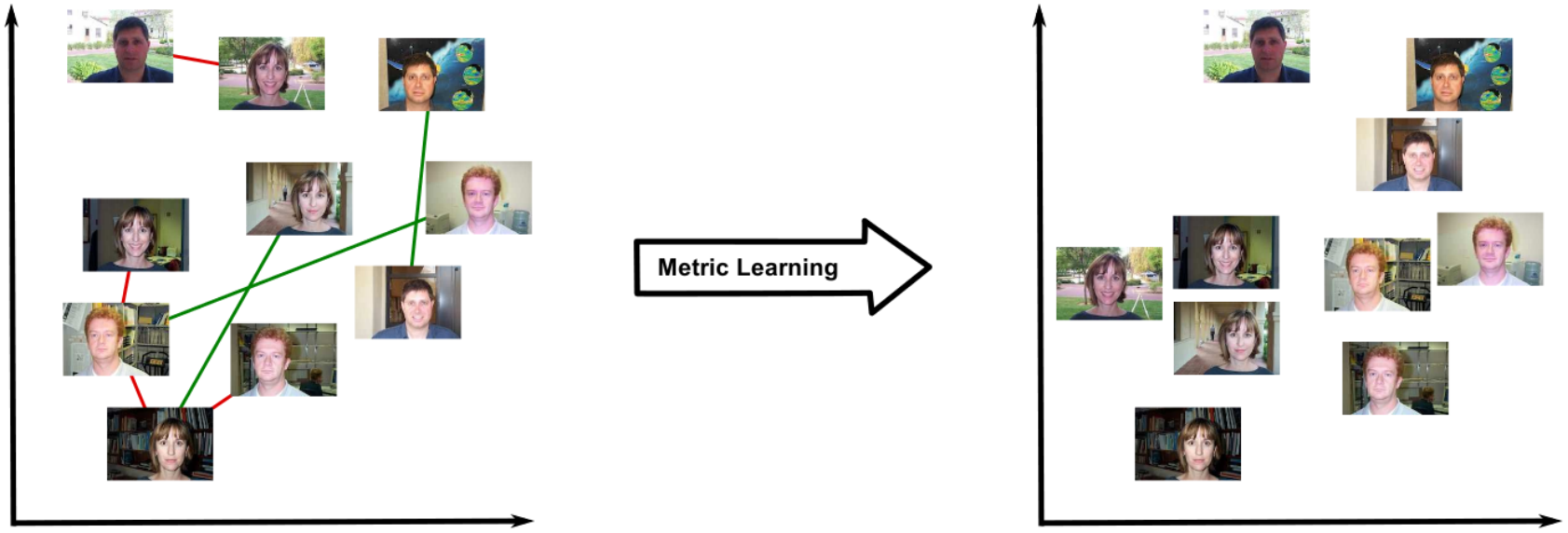
- Beyond classification image+label
- Data for training : image pairs, triplets, …
  - Pairs+label YES/NO (LFW)



  - Class information



Introduction: Metric learning for CV



Metrics in Machine Learning and Computer Vision

- Image dataset Clustering
- Information/Image retrieval
- kNN classification, Kernel methods

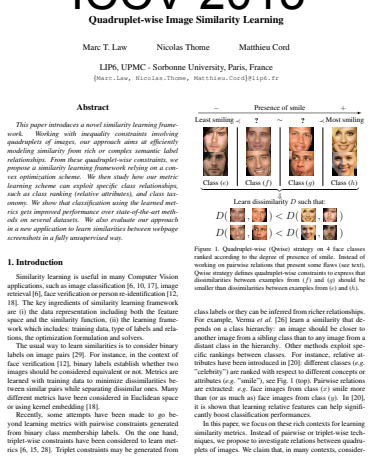
Commonly used metrics: Euclidean distance, chi2 for histograms, …

[Bellet et al., A Survey on Metric Learning for Feature Vectors and Structured Data, Tech. report, 2013]

Outline

1. Introduction
2. **Metric Learning in CV**
   - **Data and Metric models**
   - Learning schemes:
     - Constraints: Pairs, triplets …
     - Objective function: regularization, optimization …
3. Computer Vision Applications
   - Relative attribute learning
   - Web page comparison

ICCV 2013



Quadruplet-wise Image Similarity Learning

Marc T. Law Nicolas Thome Matthieu Cord

LIP6, UPMC - Sorbonne University, Paris, France

{Marc.Law, Nicolas.Thome, Matthieu.Cord}@lip6.fr

Abstract

This paper introduces a novel similarity learning framework. Working with inequality constraints involving quadruplets of images, our approach aims at efficiently modeling similarity from rich or complex semantic label relationships. From these quadruplet-wise constraints, we propose a similarity learning framework relying on a convex optimization scheme. We then study how our metric learning scheme can exploit specific class relationships, such as class ranking (relative attributes), and class taxonomy. We show that classification using the learned metrics gets improved performance over state-of-the-art methods on several datasets. We also evaluate our approach in a new application to learn similarities between webpage screenshots in a fully unsupervised way.

1. Introduction

Similarity learning is useful in many Computer Vision applications, such as image classification [6, 10, 17], image retrieval [6], face verification or person re-identification [12, 18]. The key ingredients of similarity learning framework are (i) the data representation including both the feature space and the similarity function, (ii) the learning framework which includes: training data, type of labels and relations, the optimization formulation and solvers.

The usual way to learn similarities is to consider binary labels on image pairs [29]. For instance, in the context of face verification [12], binary labels establish whether two images should be considered equivalent or not. Metrics are learned with training data to minimize dissimilarities between similar pairs while separating dissimilar ones. Many different metrics have been considered in Euclidean space or using kernel embedding [18].

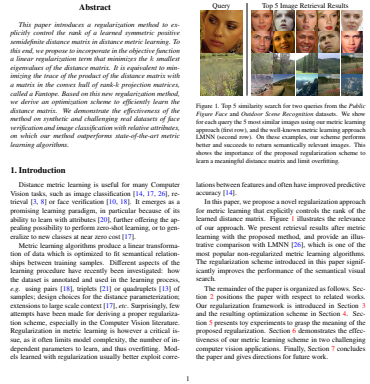
Recently, some attempts have been made to go beyond learning metrics with pairwise constraints generated from binary class membership labels. On the one hand, triplet-wise constraints have been considered to learn metrics [6, 15, 28]. Triplet constraints may be generated from class labels or they can be inferred from richer relationships. For example, Verma *et al.* [28] learn a similarity that depends on a class hierarchy: an image should be closer to another image from a sibling class than to any image from a distant class in the hierarchy. Other methods exploit specific rankings between classes. For instance, relative attributes have been introduced in [20]: different classes (*e.g.* "celebrity") are ranked with respect to different concepts or attributes (*e.g.* "smile"), see Fig. 1 (top). Pairwise relations are extracted: *e.g.* face images from class (c) smile more than (or as much as) face images from class (g). In [20], it is shown that learning relative features can help significantly boost classification performances.

In this paper, we focus on these rich contexts for learning similarity metrics. Instead of pairwise or triplet-wise techniques, we propose to investigate relations between quadruplets of images. We claim that, in many contexts, consider-

Figure 1. Quadruplet-wise (Qwise) strategy on 4 face classes ranked according to the degree of presence of smile. Instead of working on pairwise relations that present some flaws (see text), Qwise strategy defines quadruplet-wise constraints to express that dissimilarities between examples from (f) and (g) should be smaller than dissimilarities between examples from (e) and (h).

1

CVPR 2014



Fantope Regularization in Metric Learning

Marc T. Law Nicolas Thome Matthieu Cord

Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France

Abstract

This paper introduces a regularization method to explicitly control the rank of a learned symmetric positive semidefinite distance matrix in distance metric learning. To this end, we propose to incorporate in the objective function a linear regularization term that minimizes the k smallest eigenvalues of the distance matrix. It is equivalent to minimizing the trace of the product of the distance matrix with a matrix in the convex hull of rank-k projection matrices, called a Fantope. Based on this new regularization method, we derive an optimization scheme to efficiently learn the distance matrix. We demonstrate the effectiveness of the method on synthetic and challenging real datasets of face verification and image classification with relative attributes, on which our method outperforms state-of-the-art metric learning algorithms.

1. Introduction

Distance metric learning is useful for many Computer Vision tasks, such as image classification [14, 17, 26], retrieval [3, 8] or face verification [10, 18]. It emerges as a promising learning paradigm, in particular because of its ability to learn with attributes [20], further offering the appealing possibility to perform zero-shot learning, or to generalize to new classes at near zero cost [17].

Metric learning algorithms produce a linear transformation of data which is optimized to fit semantical relationships between training samples. Different aspects of the learning procedure have recently been investigated: how the dataset is annotated and used in the learning process, *e.g.* using pairs [18], triplets [21] or quadruplets [13] of samples; design choices for the distance parameterization; extensions to large scale context [17], *etc.* Surprisingly, few attempts have been made for deriving a proper regularization scheme, especially in the Computer Vision literature. Regularization in metric learning is however a critical issue, as it often limits model complexity, the number of independent parameters to learn, and thus overfitting. Models learned with regularization usually better exploit correlations between features and often have improved predictive accuracy [14].

In this paper, we propose a novel regularization approach for metric learning that explicitly controls the rank of the learned distance matrix. Figure 1 illustrates the relevance of our approach. We present retrieval results after metric learning with the proposed method, and provide an illustrative comparison with LMNN [26], which is one of the most popular non-regularized metric learning algorithms. The regularization scheme introduced in this paper significantly improves the performance of the semantical visual search.

The remainder of the paper is organized as follows. Section 2 positions the paper with respect to related works. Our regularization framework is introduced in Section 3 and the resulting optimization scheme in Section 4. Section 5 presents toy experiments to grasp the meaning of the proposed regularization. Section 6 demonstrates the effectiveness of our metric learning scheme in two challenging computer vision applications. Finally, Section 7 concludes the paper and gives directions for future work.

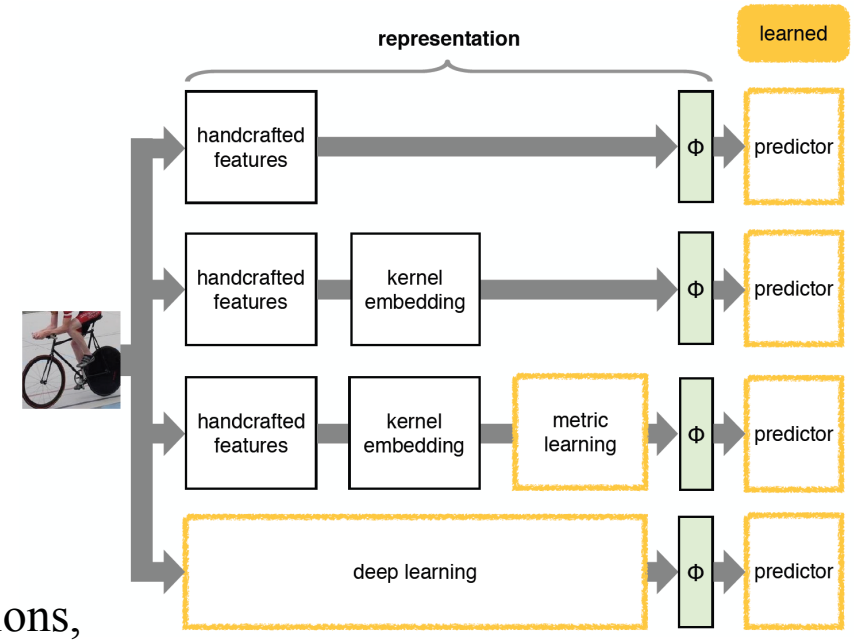
Figure 1. Top 5 similarity search for two queries from the *Public Figure Face and Outdoor Scene Recognition* datasets. We show for each query the 5 most similar images using our metric learning approach (first row), and the well-known metric learning approach LMNN (second row). On these examples, our scheme performs better and succeeds to return semantically relevant images. This shows the importance of the proposed regularization scheme to learn a meaningful distance matrix and limit overfitting.

1

Metric Learning in CV

- Key ingredients of metric/similarity learning in CV:
  - Data representation including both:
    - Feature space
      - » Bag of visual word representation (BoW)
      - » Deep features, Gist …

      IMAGE REPRESENTATION ➔ VECTOR
    - Similarity function / Metric
  - Learning framework
    - training data, type of labels and relations,
    - Optimization formulation
    - Solvers



Credit: A. Vedaldi

Metric Learning in CV

- Similarity function / Metric:

Vector representations $\mathbf{x} \in \mathbb{R}^d$ (visual BoWs, deep, ...)

Widely used approach: **Mahalanobis-like Distance Metric Learning**

$$\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^d, \mathbf{M} \in \mathbb{S}_+^d, \; D^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j) \tag{1}$$

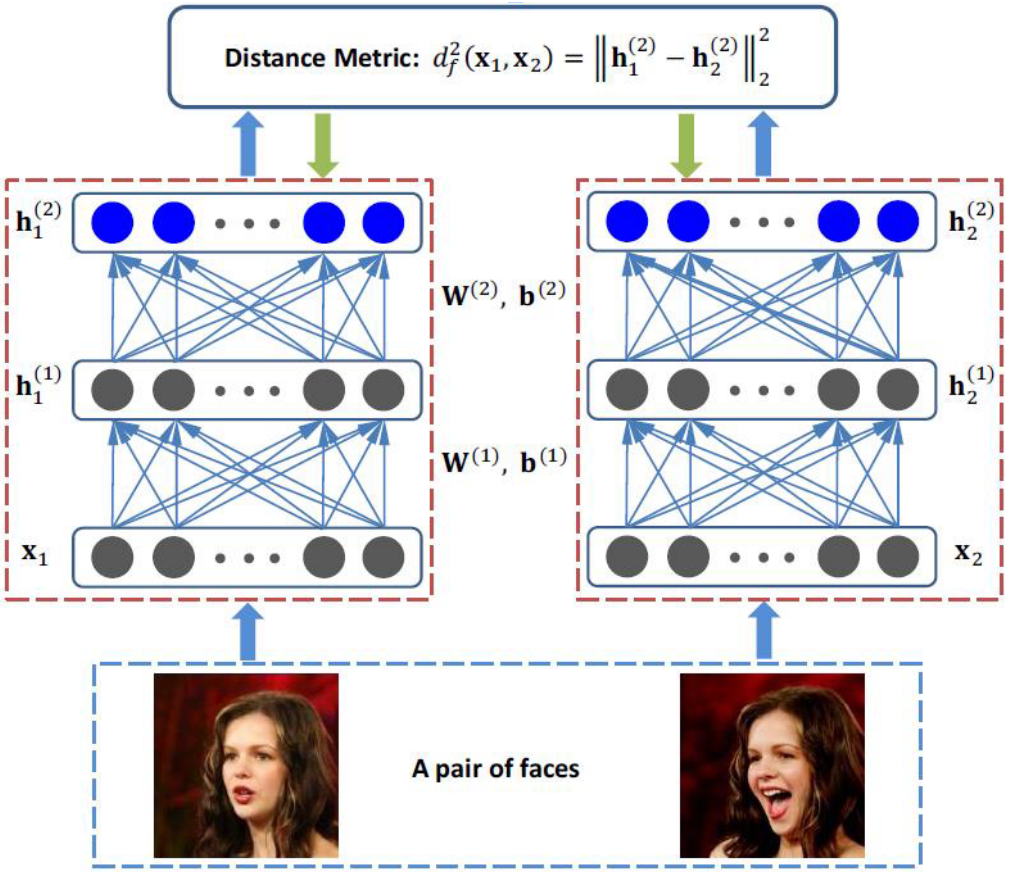
Since for all $\mathbf{M} \in \mathbb{S}_+^d$ with $\text{rank}(\mathbf{M}) = e \leq d$, there exists $\mathbf{L} \in \mathbb{R}^{e \times d}$ such that $\mathbf{M} = \mathbf{L}^\top \mathbf{L}$:

$$\begin{aligned} \mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^d, \mathbf{M} \in \mathbb{S}_+^d, \; D^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) &= (\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{L}^\top \mathbf{L} (\mathbf{x}_i - \mathbf{x}_j) \\ &= \|\mathbf{L}\mathbf{x}_i - \mathbf{L}\mathbf{x}_j\|_2^2 \end{aligned} \tag{2}$$

- All M (or L) coefficients to be learned

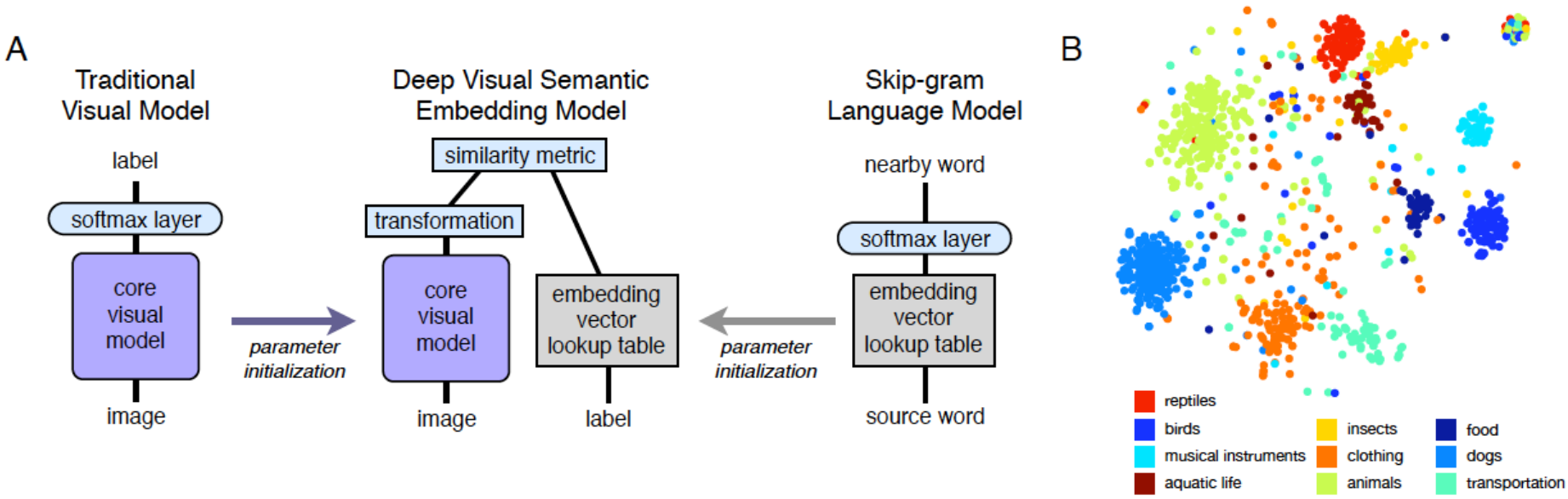
Metric Learning in CV

Non-linear extension: kernel vs deep [credit: Hu CVPR14]



Metric Learning in CV

- One step further: heterogeneous object deep embedding and metric learning



DeVISE system (google NIPS 2013)

Outline

1. Introduction
2. **Metric Learning in CV**
   - Data and Metric models
   - **Learning schemes:**
     - **Constraints: Pairs, triplets …**
     - Objective function: regularization, optimization …
3. Computer Vision Applications

Metric Learning in CV

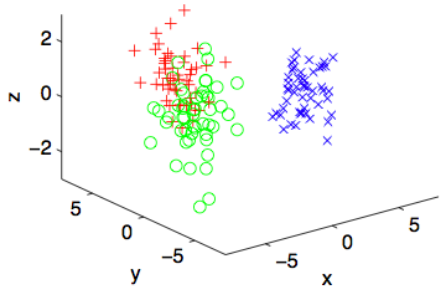
- PairWise Constraints for learning

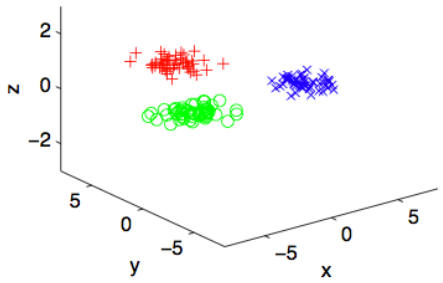
Metric Learning in CV

- Learning scheme for pairwise constraints:

  Xing et al: *Distance metric learning, with application to clustering with side-information, NIPS 2002*

$$\min_{\mathbf{M}\in\mathbb{S}_+^d} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{S}} D^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j) \quad s.t. \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{D}} \sqrt{D^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j)} \geq 1$$

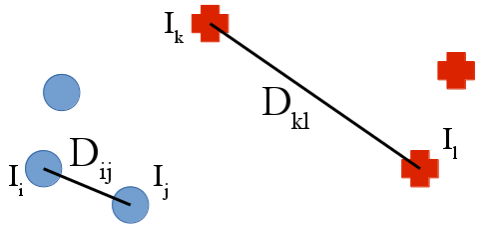




Metric Learning in CV

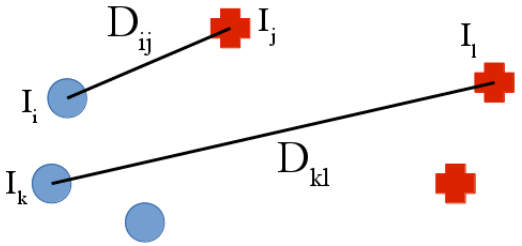
- What are the pairs in S and D ? All consistent ?

S: ok  very far : 



- Mono-modality as underlying hypothesis

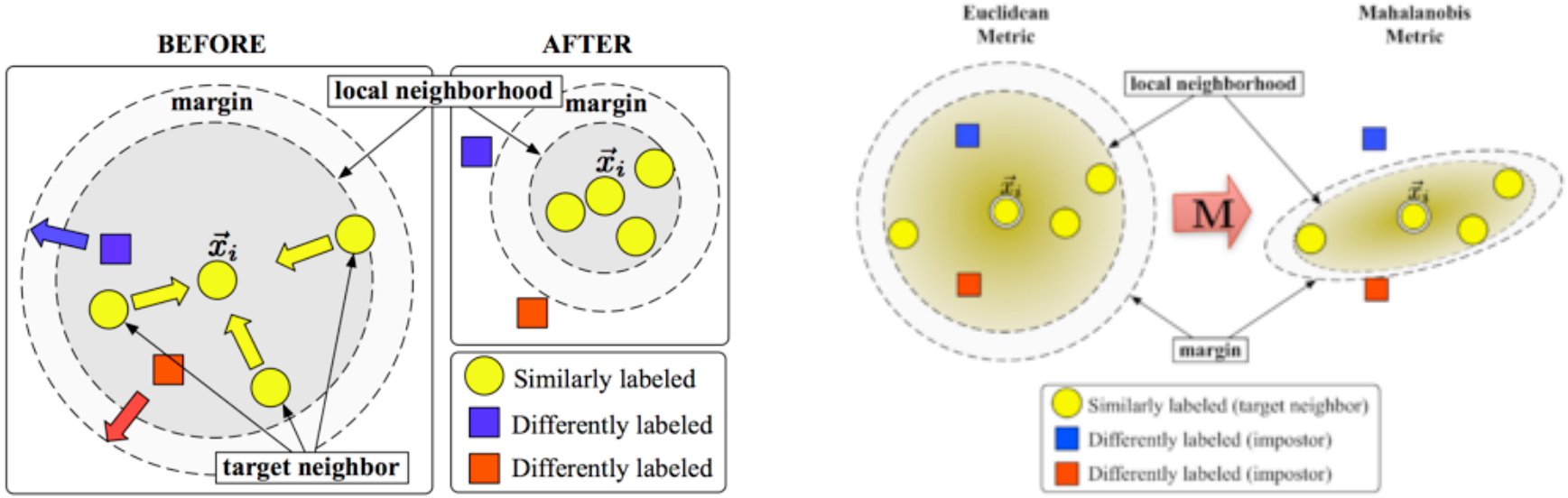
D: sometimes not far 



=> Important trick: getting training pairs using neighbor selection

Metric Learning in CV

- Triplet constraints for learning:
- The most used scheme: [Weinberger LMNN NIPS06]



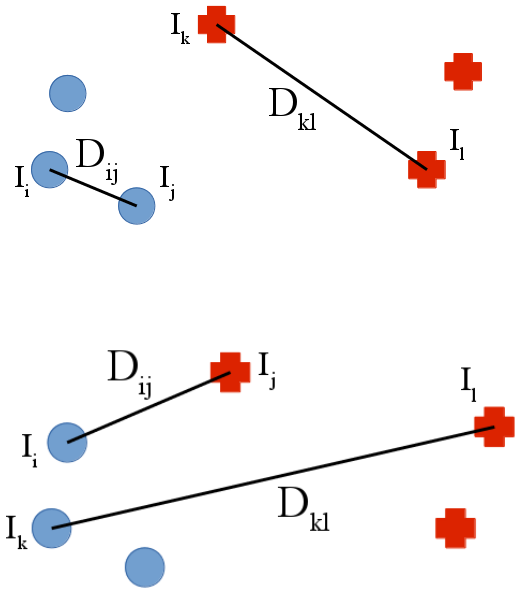
$$\min_{\mathbf{M}\in\mathbb{S}_+^d} \sum_{(\mathbf{x}_i,\mathbf{x}_i^+)\in\mathcal{S}} D^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_i^+)$$

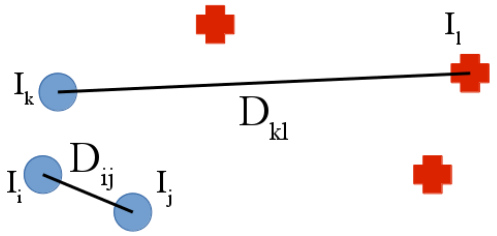
$$\text{s.t.}\quad \forall(\mathbf{x}_i,\mathbf{x}_i^+,\mathbf{x}_i^-)\in\mathcal{T}, D^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_i^-)\geq\delta+D^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_i^+)$$

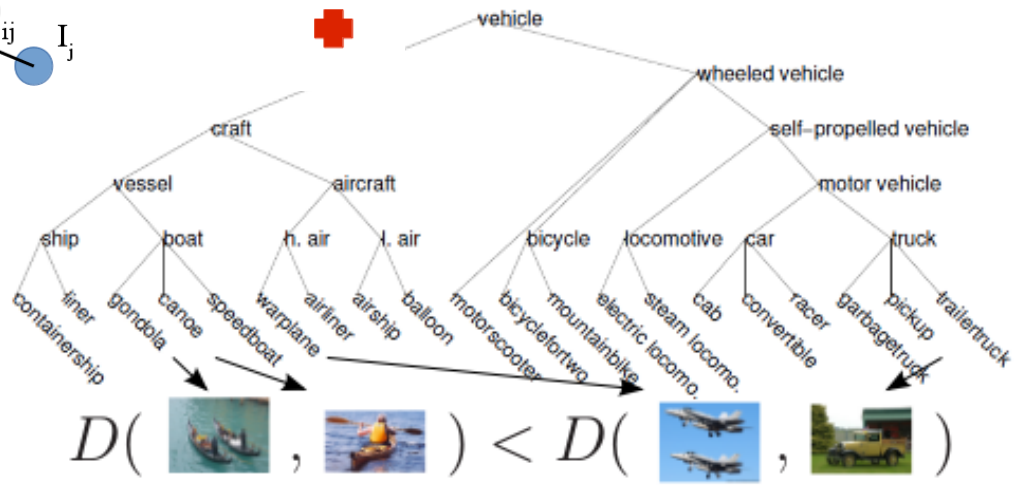
Metric Learning in CV

- Quadruplet-Wise constraints: [Cord ICCV 2013]
  - Generalizing pairs-wise (and triplets), more flexible and expressive
  - Margin-based strategy, not always selecting all constraints

$$\forall q = (\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k, \mathbf{x}_l) \in \mathcal{N},\ D^2(\mathbf{x}_i, \mathbf{x}_j) + \delta_q \leq D^2(\mathbf{x}_k, \mathbf{x}_l)$$

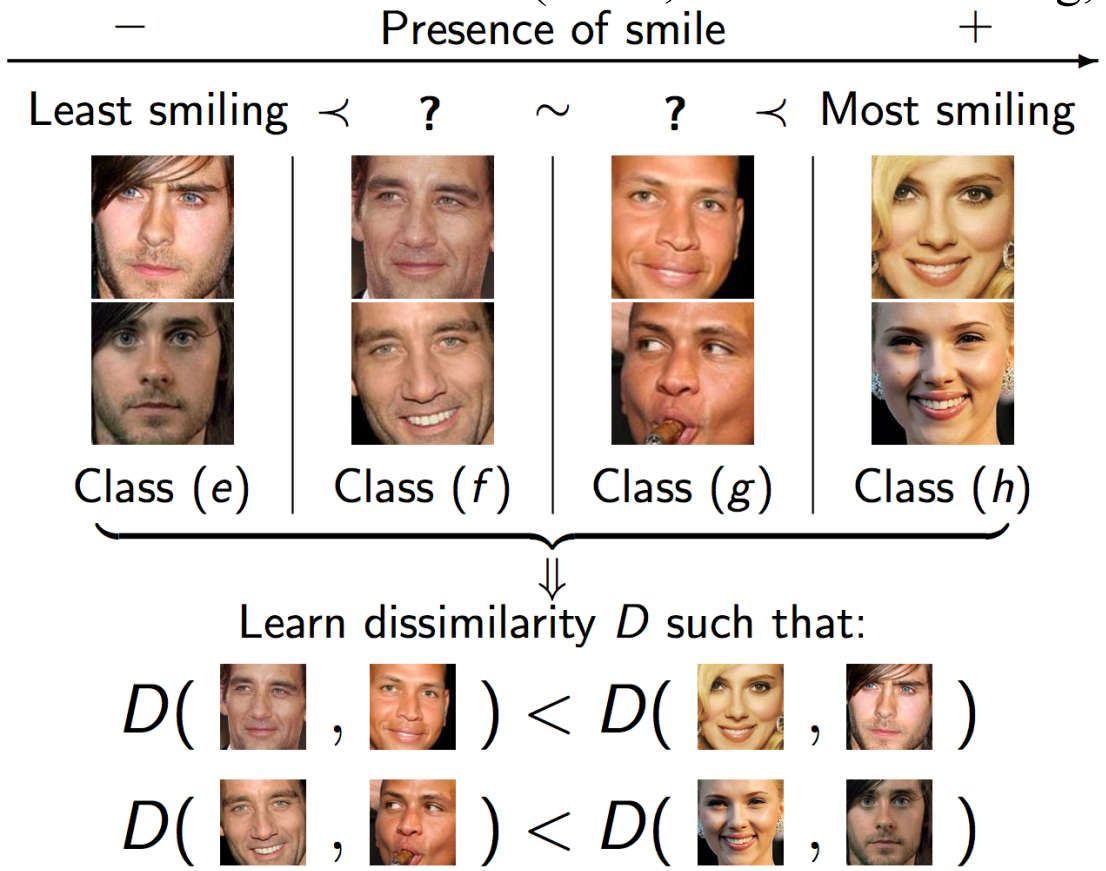






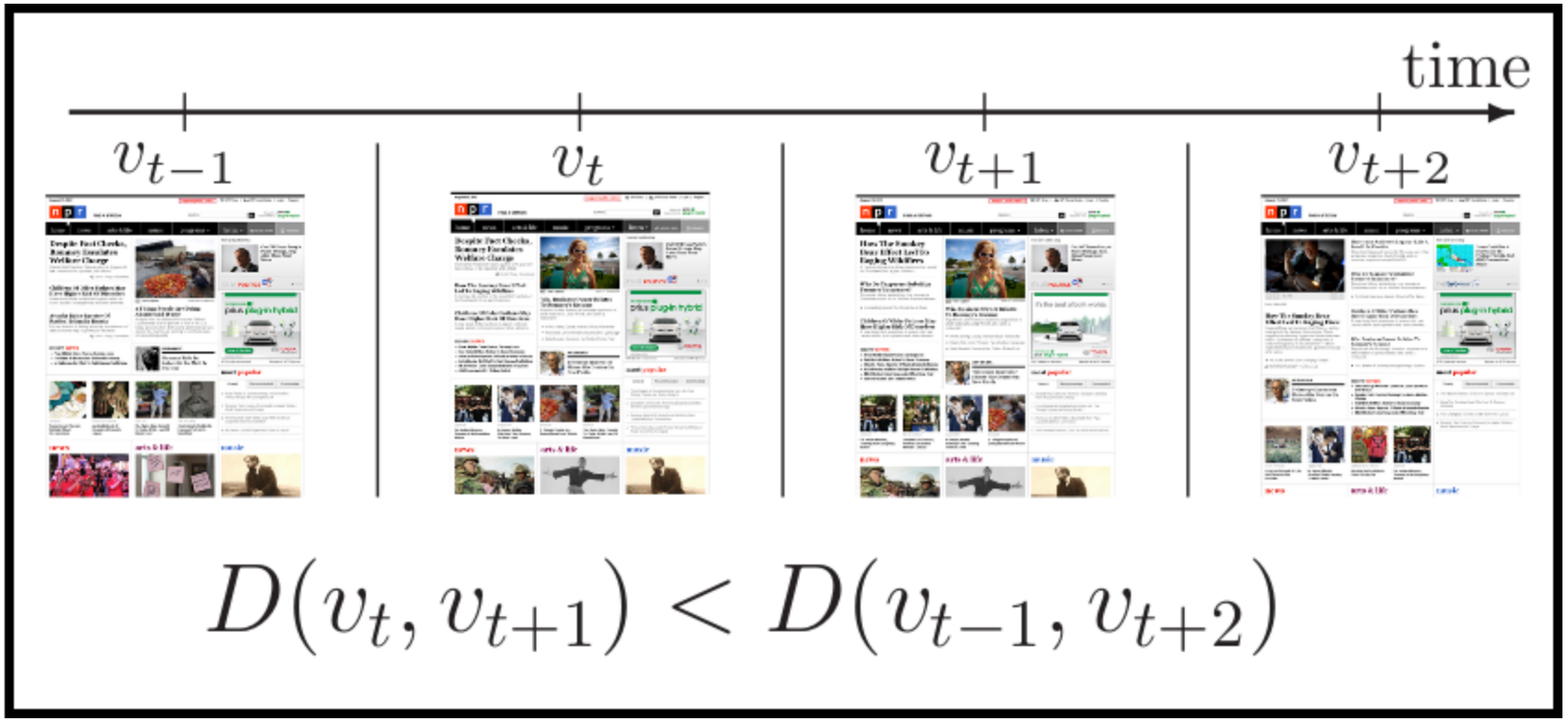
Metric Learning in CV

- Application 1: learning relative attributes
  - Supervision based on attributes (smile, masculine looking, …)



Web page/temporal info for ML

- Application 2:
  - Fully unsupervised ML, but temporal information available
  - Constraints by comparing screenshots of successive webpage versions



Outline

1. Introduction
2. **Metric Learning in CV**
   - Data and Metric models
   - **Learning schemes:**
     - Constraints: Pairs, triplets …
     - **Objective function: regularization, optimization …**
3. Computer Vision Applications
   - Relative attribute learning
   - Web page comparison

Metric Learning in CV

To summarize constraints with $D^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{M}(\mathbf{x}_i - \mathbf{x}_j)$:

- **Pairs:**
  $\mathcal{N} = \mathcal{S} \cup \mathcal{D} \Longrightarrow \begin{cases} \forall (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S} & D^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) < 1 \\ \forall (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{D} & D^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) > 1 \end{cases}$

- **Triplets:**
  $\mathcal{N} = \{(\mathbf{x}_i, \mathbf{x}_i^+, \mathbf{x}_i^-)\}_{i=1}^N \Longrightarrow \forall (\mathbf{x}_i, \mathbf{x}_i^+, \mathbf{x}_i^-) \in \mathcal{N},\ D^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_i^+) + \delta \leq D^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_i^-)$

- **Quadruplets:**
  $\mathcal{N} = \{q = (\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k, \mathbf{x}_l)\} \Longrightarrow \forall q \in \mathcal{N},\ D^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) + \delta_q \leq D^2_{\mathbf{M}}(\mathbf{x}_k, \mathbf{x}_l)$

Optimization scheme:

$$\min_{\mathbf{M} \in \mathbb{S}^d_+} \mu R(\mathbf{M}) + \ell(\mathbf{M}, \mathcal{N})$$

With $R(\mathbf{M})$: regularizer and $\ell(\mathbf{M}, \mathcal{N})$ loss over set of constraints $\mathcal{N}$

Metric Learning in CV

(Large margin) **optimization**:

- Qwise optimization framework with hinge loss function

$$\min_{\mathbf{M}\in\mathbb{S}_+^d} \mu R(\mathbf{M}) + \sum_{q\in\mathcal{N}} \xi_q$$

$$\text{s.t. } \forall q = (\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k, \mathbf{x}_l) \in \mathcal{N}, D^2_{\mathbf{M}}(\mathbf{x}_k, \mathbf{x}_l) \geq D^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) + \delta_q - \xi_q$$
$$\forall q \in \mathcal{N}, \xi_q \geq 0$$

  - $R(\mathbf{M})$: regularization term
  - μ : trade-off between fitting and regularization.

- Triplet optim:

$$\min_{\mathbf{M}\in\mathbb{S}_+^d} \sum_{(\mathbf{x}_i,\mathbf{x}_i^+)\in\mathcal{S}} D^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_i^+) + \sum_{(\mathbf{x}_i,\mathbf{x}_i^+,\mathbf{x}_i^-)\in\mathcal{T}} \xi_i$$

$$\text{s.t. } \forall(\mathbf{x}_i, \mathbf{x}_i^+, \mathbf{x}_i^-) \in \mathcal{T}, D^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_i^-) \geq 1 + D^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_i^+) - \xi_i$$

Metric Learning in CV

- How to define/choose the regularization R(M) in the objective function:

$$\min_{\mathbf{M} \in \mathbb{S}_+^d} \mu R(\mathbf{M}) + \ell(\mathbf{M}, \mathcal{N})$$

- Regularization term to express *prior*, to control complexity …

$$D^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j)$$

- For CV application, looking for Low rank solution:
  - Controlling overfitting
  - Sparsity of the singular values
  - Exploiting correlation between features
  - Fast/efficient solution

Metric Learning in CV

Formulation of $R(\mathbf{M})$ $\qquad D^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{M}(\mathbf{x}_i - \mathbf{x}_j)$

- Frobenius norm $R(\mathbf{M}) = \|\mathbf{M}\|_F^2 = \sum M_{ij}^2$
  - matrix analog of the standard ℓ_2 regularizer in SVM
  - does not promote low-rank solutions
  - useful when $\mathbf{M}$ is a diagonal matrix

- Schultz, Learning a Distance Metric from Relative Comparisons, NIPS 2003

- Nuclear norm regularization $R(\mathbf{M}) = \|\mathbf{M}\|_* = \mathrm{tr}(\mathbf{M})$:
  - rank NP-hard to optimize
  - convex envelope of $\mathrm{rank}(\mathbf{M})$ on the set $\{\mathbf{M} \in \mathbb{R}^{d \times d} : \|\mathbf{M}\| \leq 1\}$
  - ℓ_1 norm of vector of singular values $\sigma(\mathbf{M})$

- McFee, Metric Learning to Rank, ICML 2010

Metric Learning in CV

- Fantope regularization [Cord CVPR 2014]:

  - Explicit control of the rank of M

    By noting, $\forall \mathbf{M} \in \mathbb{S}_+^d, R(\mathbf{M})$: sum of the k smallest eigenvalues of $\mathbf{M}$

    $$R(\mathbf{M}) = 0 \Longleftrightarrow \text{rank}(\mathbf{M}) \leq d - k$$

  - Reformulation

    $$\min_{\mathbf{M} \in \mathbb{S}_+^d} \mu R(\mathbf{M}) + \ell(\mathbf{M}, \mathcal{N}) \Longrightarrow \min_{\mathbf{M} \in \mathbb{S}_+^d} \mu \langle \mathbf{W}, \mathbf{M} \rangle + \ell(\mathbf{M}, \mathcal{N})$$

    with $\mathbf{W}$ rank-k projector on the eigenvectors of $\mathbf{M}$ with k smallest eigenvalues

Metric Learning in CV

Construction of W

- $\mathbf{M} = \mathbf{V}_\mathbf{M}\text{Diag}(\lambda(\mathbf{M}))\mathbf{V}_\mathbf{M}^\top$ eigendecomposition of $\mathbf{M} \in \mathbb{S}_+^d$, $\mathbf{V}_\mathbf{M}$ orthogonal matrix
- We construct $\mathbf{w} = (w_1, \ldots, w_d)^\top \in \mathbb{R}^d$:

$$w_i = \begin{cases} 0 \text{ if } 1 \leq i \leq d-k \text{ (the first } d-k \text{ elements)} \\ 1 \text{ if } d-k+1 \leq i \leq d \text{ (the last } k \text{ elements)} \end{cases}$$

$$\mathbf{W} = \mathbf{V}_\mathbf{M}\text{Diag}(\mathbf{w})\mathbf{V}_\mathbf{M}^\top \tag{1}$$

$$\min_{\mathbf{M}\in\mathbb{S}_+^d} \mu R(\mathbf{M}) + \ell(\mathbf{M}, \mathcal{N}) \Longrightarrow \min_{\mathbf{M}\in\mathbb{S}_+^d} \mu\langle \mathbf{W}, \mathbf{M}\rangle + \ell(\mathbf{M}, \mathcal{N}) \text{ s.t. } \mathbf{W} = \mathbf{V}_\mathbf{M}\text{Diag}(\mathbf{w})\mathbf{V}_\mathbf{M}^\top$$

Metric Learning in CV

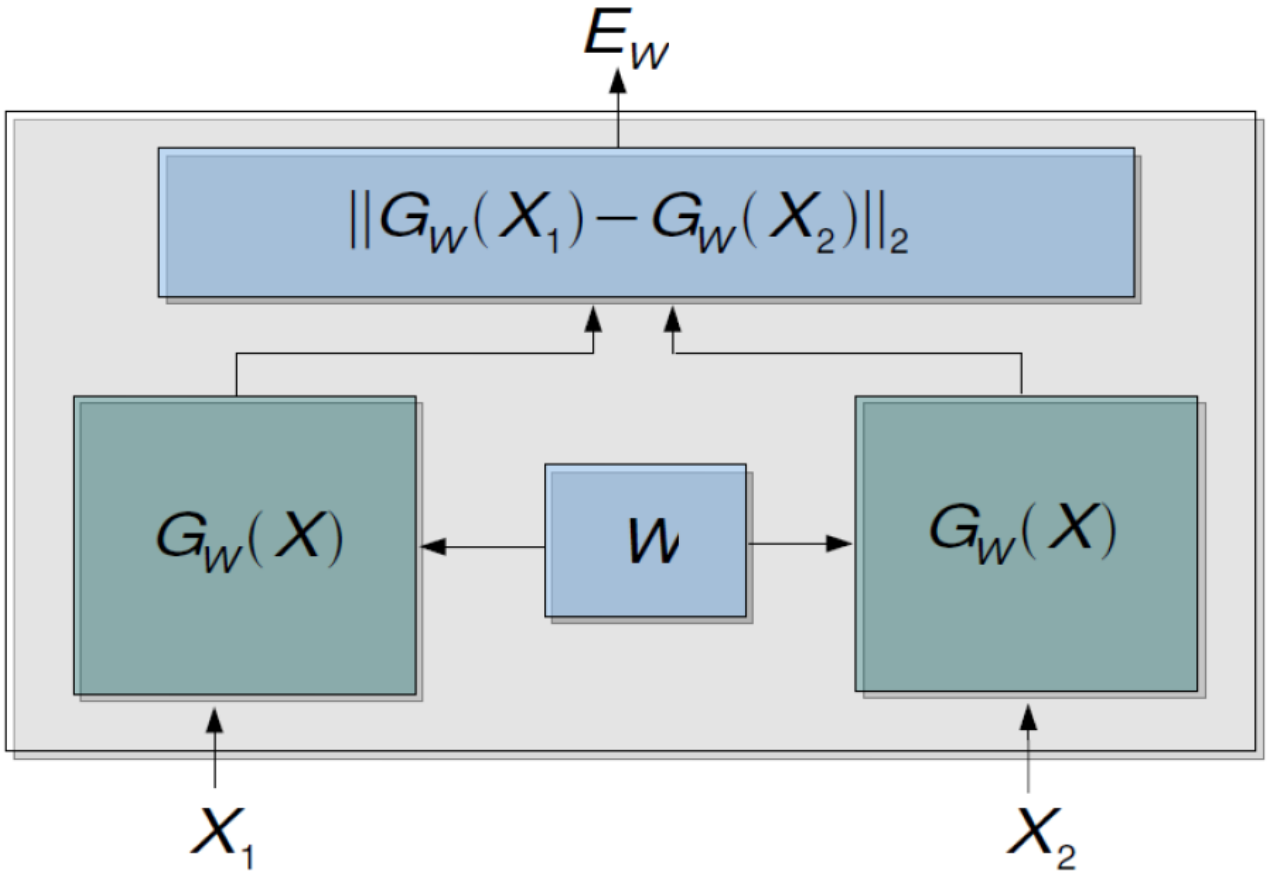
- Algorithm: alternating optimization procedure

Input: Training constraints $\mathcal{N}$, hyper-parameter μ and step size $\eta > 0$
Output: $\mathbf{M} \in \mathbb{S}_+^d$
Initialize $\mathbf{M}^1 \in \mathbb{S}_+^d$, iteration $n = 1$
Repeat until convergence

1. $\mathbf{W}^n \leftarrow \mathbf{V}_{\mathbf{M}^n} \mathrm{Diag}(\mathbf{w}) \mathbf{V}_{\mathbf{M}^n}^\top$
2. Fix $\mathbf{W}^n$ in Eq. (1)
3. $\mathbf{W}^n \in \partial(\langle \mathbf{W}^n, \mathbf{M}^n \rangle)$
4. $\mathbf{G}^n \in \partial \ell(\mathbf{M}^n, \mathcal{N})$
5. $\mathbf{M}^{n+1} \leftarrow \Pi_{\mathbb{S}_+^d} (\mathbf{M}^n - \eta(\mu \mathbf{W}^n + \mathbf{G}^n))$
6. $n \leftarrow n + 1$

Metric Learning in CV

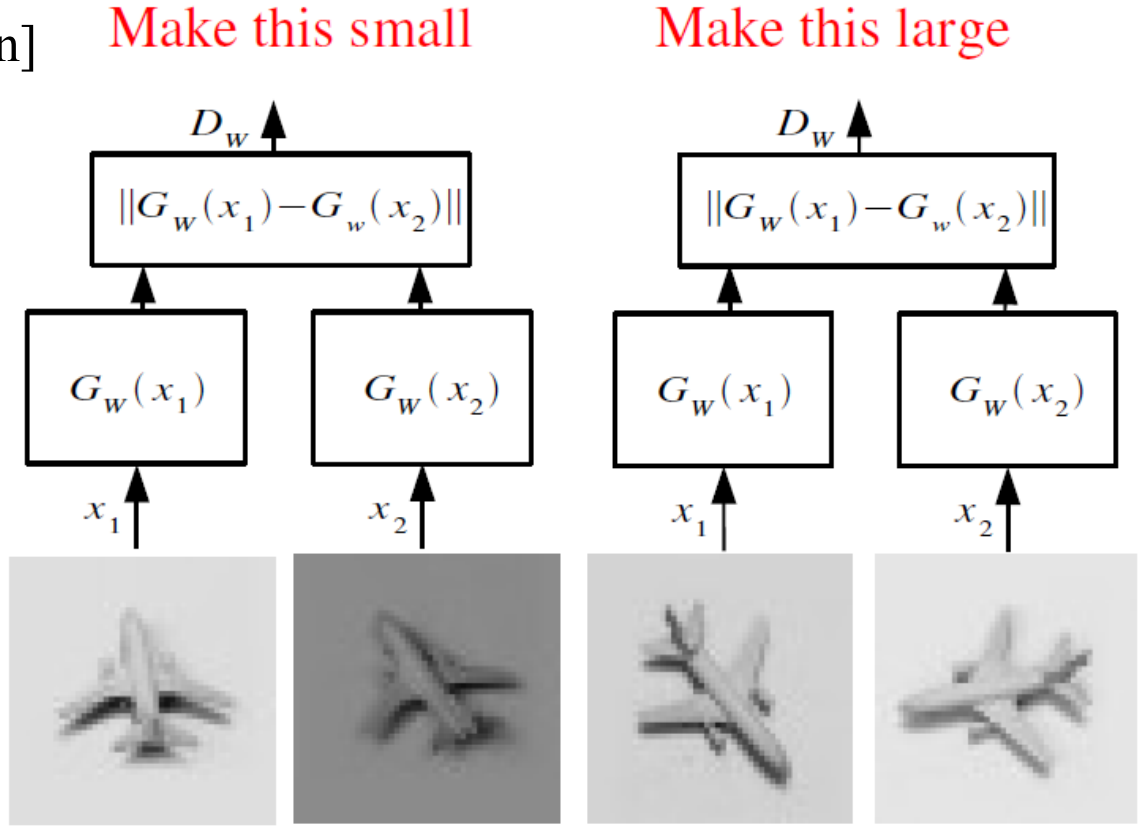
- Deep metric learning optimization
  - Siamese Architecture [LeCun]



Metric Learning in CV

- Deep metric learning optimization

[credit: Y. LeCun]



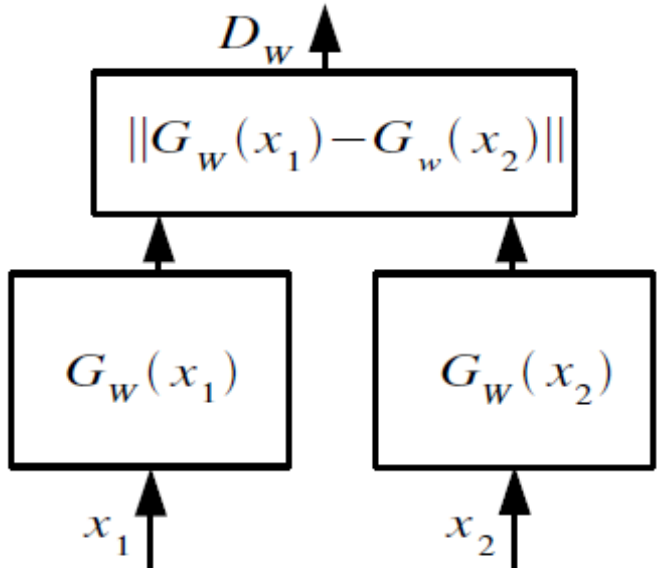
Metric Learning in CV

- Deep metric learning optimization

[credit: Y. LeCun CVPR 06]

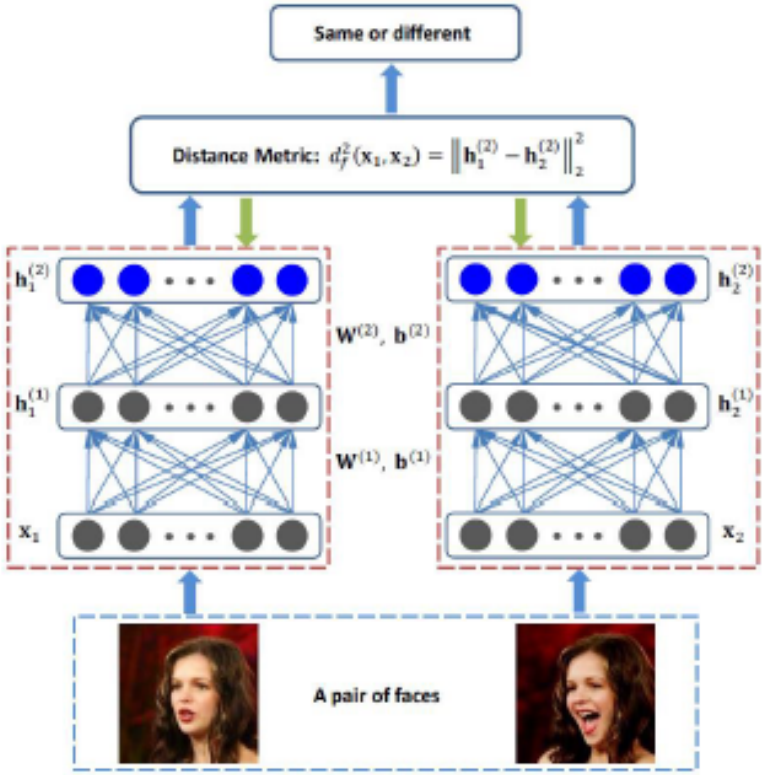
The exact loss function is

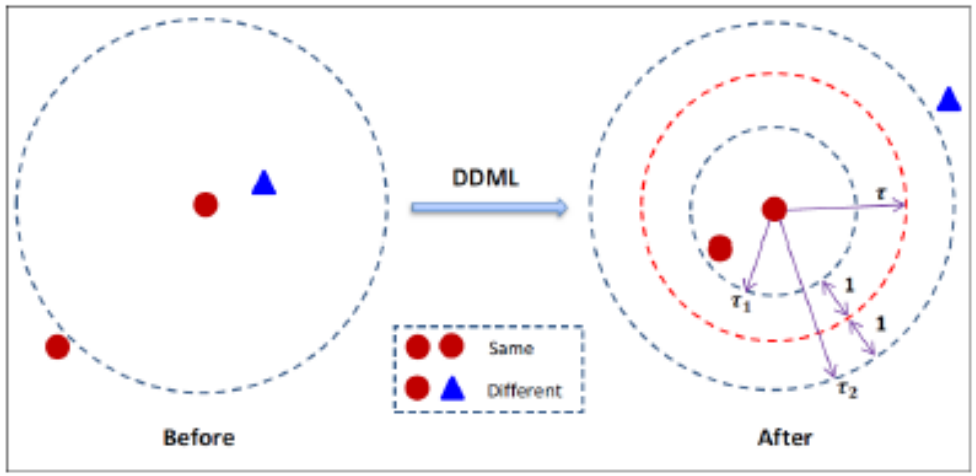
$$L(W, Y, \vec{X}_1, \vec{X}_2) =$$
$$(1 - Y)\frac{1}{2}(D_W)^2 + (Y)\frac{1}{2}\{max(0, m - D_W)\}$$



Metric Learning in CV

Siamese Network for paiwise comparison: DDML approach





Intuitive illustration of the proposed DDML method

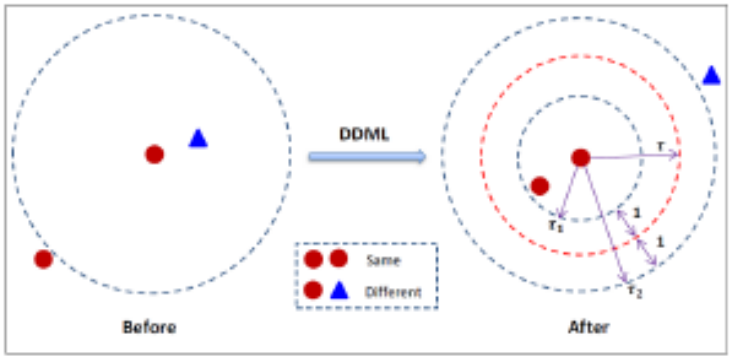
Junlin Hu, etc. Discriminative Deep Metric Learning for Face Verification in theWild, CVPR 2014

Metric Learning in CV

DDML optimization

$$d_f^2(x_i, x_j) < \tau - 1, l_{ij} = 1$$
$$d_f^2(x_i, x_j) > \tau + 1, l_{ij} = -1$$

$$\ell_{ij}\left(\tau - d_f^2(\mathbf{x}_i, \mathbf{x}_j)\right) > 1$$

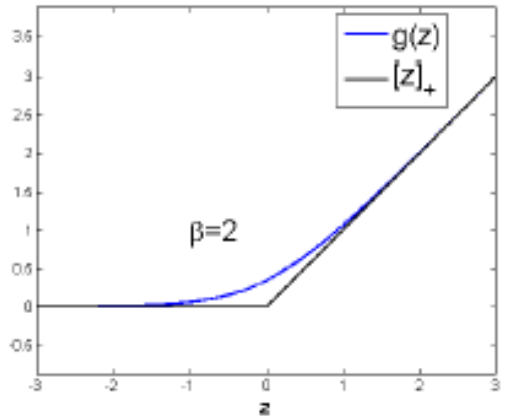


Intuitive illustration of the proposed DDML method

DDML as the following optimization problem:

$$\begin{aligned}\arg\min_f J &= J_1 + J_2 \\ &= \frac{1}{2}\sum_{i,j} g\left(1 - \ell_{ij}\left(\tau - d_f^2(\mathbf{x}_i, \mathbf{x}_j)\right)\right) \\ &+ \frac{\lambda}{2}\sum_{m=1}^{M}\left(\left\|\mathbf{W}^{(m)}\right\|_F^2 + \left\|\mathbf{b}^{(m)}\right\|_2^2\right)\end{aligned}$$

where $g(z) = \frac{1}{\beta}\log\left(1+\exp(\beta z)\right)$ is the generalized logistic loss function [25], which is a smoothed approximation of the hinge loss function $[z]_+ = \max(z, 0)$



Junlin Hu, etc. Discriminative Deep Metric Learning for Face Verification in theWild, CVPR 2014

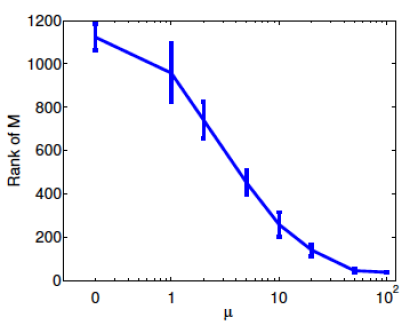
Outline

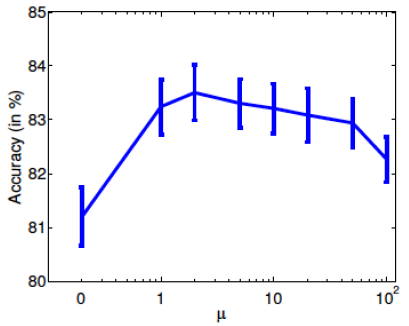
1. Introduction
2. **Metric Learning in CV**
   - Data and Metric models
   - Learning schemes:
     - Constraints: Pairs, triplets …
     - Objective function: regularization, optimization …
   - **Results**
3. Computer Vision Applications

Results on face verification pb

2 images => same face ?
Labeled Faces in the Wild (LFW)-- 27 SIFT descriptors concatenated
10-fold Cross Validation (600 pairs per fold)

Method	Accuracy (in %)
ITML	76.2 ± 0.5
LDML	77.5 ± 0.5
PCCA	82.2 ± 0.4
Fantope	$\mathbf{83.5 \pm 0.5}$

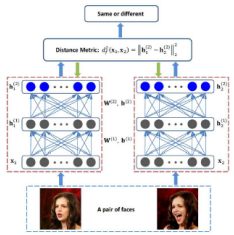




About 15% better with metric learning

Results on face verification pb

Performances of deep DDML on LFW (more features): 90.68%



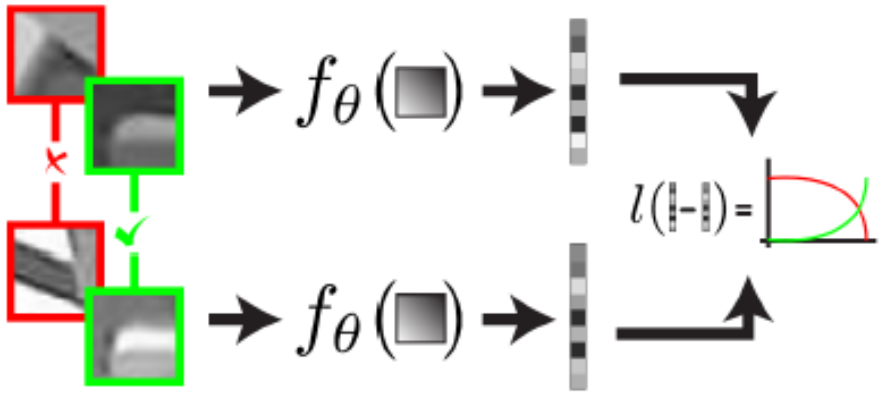
Recent extensions of deep archi (extra data, diff protocol):

Method	Accuracy (%)	No. of points	No. of images	Feature dimension
Joint Bayesian [8]	92.42 (o)	5	99,773	2000 × 4
ConvNet-RBM [31]	92.52 (o)	3	87,628	N/A
CMD+SLBP [17]	92.58 (u)	3	N/A	2302
Fisher vector faces [29]	93.03 (u)	9	N/A	128 × 2
Tom-vs-Pete classifiers [2]	93.30 (o+r)	95	20,639	5000
High-dim LBP [9]	95.17 (o)	27	99,773	2000
TL Joint Bayesian [6]	96.33 (o+u)	27	99,773	2000
DeepFace [32]	97.25 (o+u)	6 + 67	4,400,000 + 3,000,000	4096 × 4
DeepID on CelebFaces	**96.05** (o)	5	87,628	150
DeepID on CelebFaces+	**97.20** (o)	5	202,599	150
DeepID on CelebFaces+ & TL	**97.45** (o+u)	5	202,599	150

Other appli:
People verification

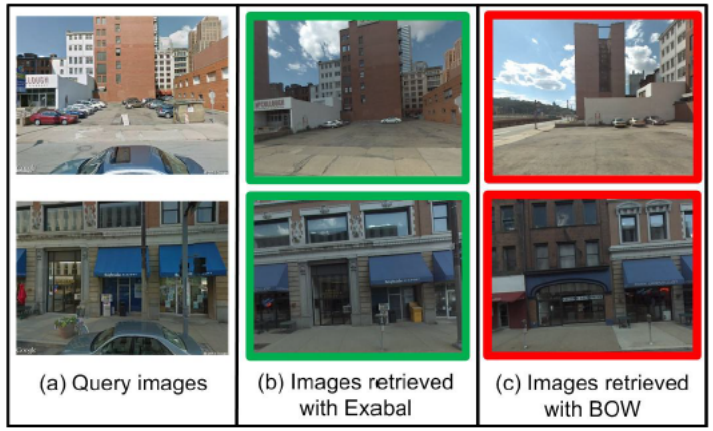
Results: feature learning

Application of deep Siamese Nets for Learning Image Descriptors



Using the contrastive cost function

$$l_{\boldsymbol{\theta}}\left(\mathrm{y}_i, \mathrm{y}_j\right) = \begin{cases} s_{ij} d_{ij}^2, & \text{if matching} \\ \max\left(1.0 - d_{ij}^2,\ 0\right), & \text{if non-matching} \end{cases}$$



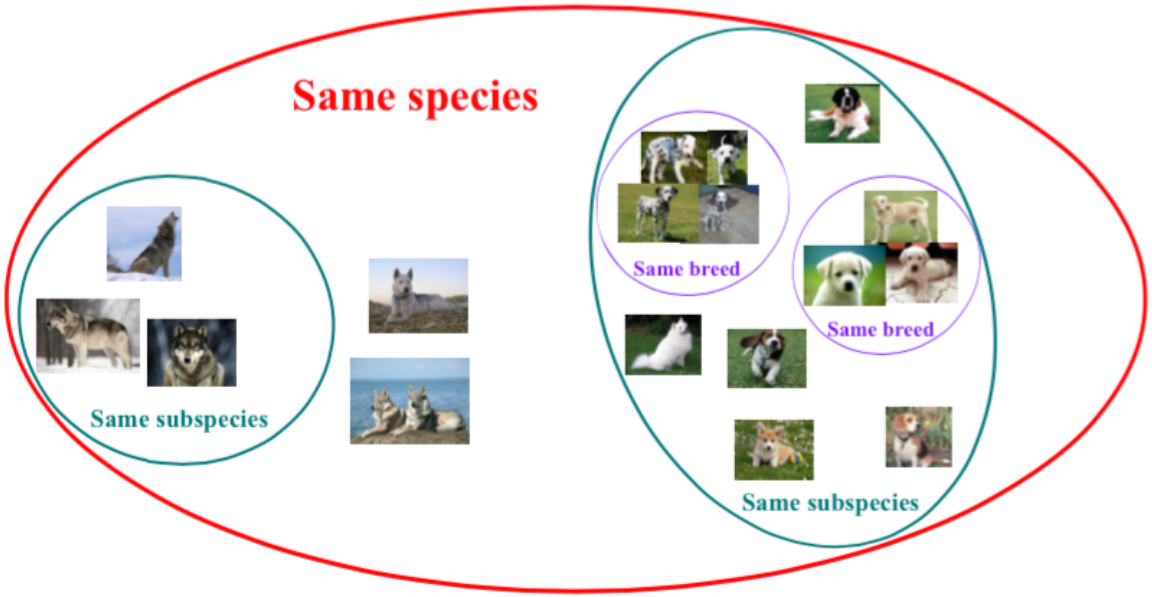
ICIP 2015

Nicholas Carlevaris-Bianco and Ryan M. Eustice, Learning visual feature descriptors for dynamic lighting conditions. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2014

Many different contexts producing training data

Results: Hierarchical Classification

Rich relationships in taxonomies can be described with relative distances
Information richer that "is similar" or "is dissimilar"
Different levels of similarity

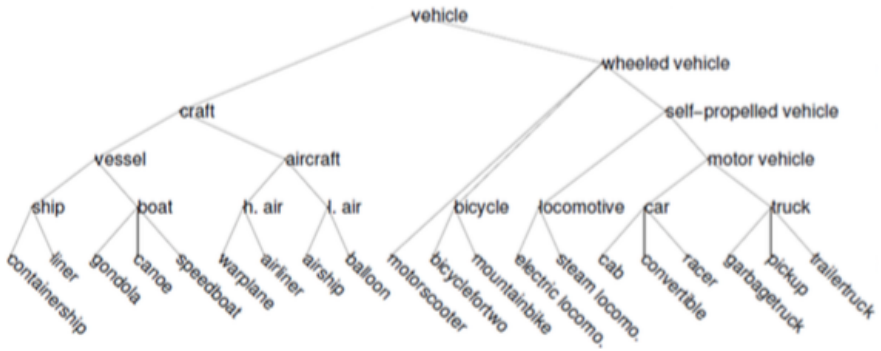


Learn dissimilarity D such that:

$D($  ,  $)$ $<$ $D($  ,  $)$

$D($  ,  $)$ $<$ $D($  ,  $)$

Results: Hierarchical Classification



$D($  ,  $) < D($  ,  $)$

(airliner, airliner) $<$ (airliner, warplane)

$D($  ,  $) < D($  ,  $)$

(canoe, speedboat) $<$ (canoe, pickup)

Two types of constraints:

- same class vs sibling classes
- sibling classes vs more distant classes

ImageNet Subtree Dataset	Verma *et al.*, 2012	Qwise
Amphibian	41%	**43.5%**
Fish	39%	**41%**
Fruit	**23.5%**	21.1%
Furniture	46%	**48.8%**
Geological Formation	52.5%	**56.1%**
Musical Instrument	32.5%	**32.9%**
Reptile	22%	**23.0%**
Tool	**29.5%**	26.4%
Vehicle	27%	**34.7%**
Average Accuracy	34.8%	**36.4%**

- Verma *et al.*: complex model that learns many adhoc metrics

Outline

1. Introduction
2. Metric Learning
3. **Computer Vision Applications**
   - Relative attribute learning
   - Web page comparison

CV app: Scarlett and others

- Best Paper (Marr Prize) at ICCV 2011:

  Relative attributes,

  D. Parikh (TTI Chicago) and

  K. Grauman (Texas Univ)

To appear, Proceedings of the International Conference on Computer Vision (ICCV), 2011.

Relative Attributes

Devi Parikh
Toyota Technological Institute Chicago (TTIC)
dparikh@ttic.edu

Kristen Grauman
University of Texas at Austin
grauman@cs.utexas.edu

Abstract

Human-nameable visual "attributes" can benefit various recognition tasks. However, existing techniques restrict these properties to categorical labels (for example, a person is 'smiling' or not, a scene is 'dry' or not), and thus fail to capture more general semantic relationships. We propose to model relative *attributes. Given training data stating how object/scene categories relate according to different attributes, we learn a ranking function per attribute. The learned ranking functions predict the relative strength of each property in novel images. We then build a generative model over the joint space of attribute ranking outputs, and propose a novel form of zero-shot learning in which the supervisor* relates *the unseen object category to previously seen objects via attributes (for example, 'bears are furrier than giraffes'). We further show how the proposed relative attributes enable richer textual descriptions for new images, which in practice are more precise for human interpretation. We demonstrate the approach on datasets of faces and natural scenes, and show its clear advantages over traditional binary attribute prediction for these new tasks.*

1. Introduction

While traditional visual recognition approaches map low-level image features directly to object category labels, recent work proposes models using *visual attributes* [1–8]. Attributes are properties observable in images that have human-designated names (e.g., 'striped', 'four-legged'), and they are valuable as a new semantic cue in various problems. For example, researchers have shown their impact for strengthening facial verification [5], object recognition [6, 8, 16], generating descriptions of unfamiliar objects [1], and to facilitate "zero-shot" transfer learning [2], where one trains a classifier for an unseen object simply by specifying which attributes it has.

Problem: Most existing work focuses wholly on attributes as binary predicates indicating the presence (or absence) of a certain property in an image [1–8, 16]. This may suffice for part-based attributes (e.g., 'has a head') and some binary properties (e.g., 'spotted'). However, for a large variety of attributes, not only is this binary setting restrictive, but it is also unnatural. For instance, it is not clear if in Figure 1(b) Hugh Laurie is smiling or not; different people are likely to respond inconsistently in providing the presence or absence of the 'smiling' attribute for this image, or of the 'natural' attribute for Figure 1(e).

(a) Smiling (b) ? (c) Not smiling

(d) Natural (e) ? (f) Manmade

Figure 1. Binary attributes are an artificially restrictive way to describe images. While it is clear that (a) is smiling, and (c) is not, the more informative and intuitive description for (b) is via *relative* attributes: he is smiling more than (a) but less than (c). Similarly, scene (e) is less natural than (d), but more so than (f). Our main idea is to model relative attributes via learned ranking functions, and then demonstrate their impact on novel forms of zero-shot learning and generating image descriptions.

Indeed, we observe that *relative* visual properties are a semantically rich way by which humans describe and compare objects in the world. They are necessary, for instance, to refine an identifying description ("the 'rounder' pillow"; "the same except 'bluer'"), or to situate with respect to reference objects ("'brighter' than a candle; 'dimmer' than a flashlight"). Furthermore, they have potential to enhance active and interactive learning—for instance, offering a better guide for a visual search ("find me similar shoes, but 'shinier'." or "refine the retrieved images of downtown Chicago to those taken on 'sunnier' days").

Proposal: In this work, we propose to model *relative attributes*. As opposed to predicting the presence of an attribute, a relative attribute indicates the strength of an attribute in an image with respect to other images. For exam-

CV app: What are attributes?

- Mid-level concepts
  - Higher than low-level features
  - Lower than high-level categories
- Shared across categories
- Human-understandable (semantic)
- Machine-detectable (visual)

Face Tracer Image Search (Kumar 08) "Smiling Asian Men With Glasses"

Slide credit: Devi Parikh

CV app: Attribute Models

$\boldsymbol{x}_i \rightarrow$ Real value

Density,
Smiling,
….

"I am 60% sure this person is smiling"
(Binary Classifier Confidence)

"This person is smiling 60%"
(Attribute Strength)

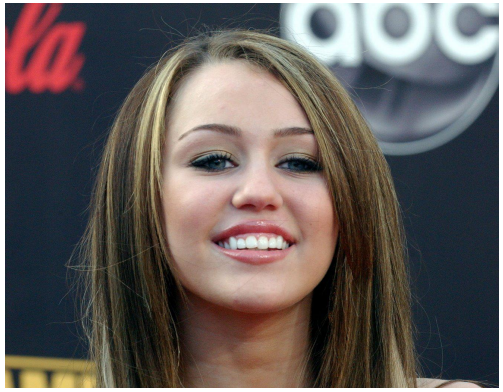
Slide credit: Devi Parikh

CV app: Relative Attributes

"Person A is smiling more than Person B"
(Relative Attribute, Parikh and Grauman ICCV 2011)

<

smiling



>

natural

- Training sets:

Attributes labeled at category level

	Binary	Relative
OSR	T I S H C O M F	
natural	0 0 0 0 1 1 1 1	T≺I∼S≺H≺C∼O∼M∼F
open	0 0 0 1 1 1 1 0	T∼F≺I∼S≺M≺H∼C∼O
perspective	1 1 1 1 0 0 0 0	O≺C≺M∼F≺H≺I≺S≺T
large-objects	1 1 1 0 0 0 0 0	F≺O∼M≺I∼S≺H∼C≺T
diagonal-plane	1 1 1 1 0 0 0 0	F≺O∼M≺C≺I∼S≺H≺T
close-depth	1 1 1 1 0 0 0 1	C≺M≺O≺T∼I∼S∼H∼F
PubFig	A C H J M S V Z	
Masculine-looking	1 1 1 1 0 0 1 1	S≺M≺Z≺V≺J≺A≺H≺C
White	0 1 1 1 1 1 1 1	A≺C≺H≺Z≺J≺S≺M≺V
Young	0 0 0 0 1 1 0 1	V≺H≺C≺J≺A≺S≺Z≺M
Smiling	1 1 1 0 1 1 0 1	J≺V≺H≺A∼C≺S∼Z≺M
Chubby	1 0 0 0 0 0 0 0	V≺J≺H≺C≺Z≺M≺S≺A
Visible-forehead	1 1 1 0 1 1 1 0	J≺Z≺M≺S≺A∼C∼H∼V
Bushy-eyebrows	0 1 0 1 0 0 0 0	M≺S≺Z≺V≺H≺A≺C≺J
Narrow-eyes	0 1 1 0 0 0 1 1	M≺J≺S≺A≺H≺C≺V≺Z
Pointy-nose	0 0 1 0 0 0 0 1	A≺C≺J∼M∼V≺S≺Z≺H
Big-lips	1 0 0 0 1 1 0 0	H≺J≺V≺Z≺C≺M≺A≺S
Round-face	1 0 0 0 1 1 0 0	H≺V≺J≺C≺Z≺A≺S≺M

Table 1. Binary and relative attribute assignments used in our experiments. Note that none of the relative orderings violate the binary memberships. The OSR dataset includes images from the following categories: coast (C), forest (F), highway (H), inside-city (I), mountain (M), open-country (O), street (S) and tall-building (T). The 8 attributes shown above are listed in [11] as the properties subjects used to organize the images.The PubFig dataset includes images of: Alex Rodriguez (A), Clive Owen (C), Hugh Laurie (H), Jared Leto (J), Miley Cyrus (M), Scarlett Johansson (S), Viggo Mortensen (V) and Zac Efron (Z). The 11 attributes shown above are a

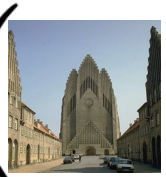
CV app: Attribute Models

- Ranking functions for relative attributes

  For each attribute a_m, open

Supervision = all pairs as:

	Binary	Relative
OSR	TI S HC OMF	
natural	0 0 0 0 1 1 1 1	T≺I∼S≺H≺C∼O∼M∼F
open	0 0 0 1 1 1 1 0	T∼F≺I∼S≺M≺H∼C∼O
perspective	1 1 1 1 0 0 0 0	O≺C≺M∼F≺H≺I≺S≺T
large-objects	1 1 1 0 0 0 0 0	F≺O∼M≺I∼S≺H∼C≺T
diagonal-plane	1 1 1 1 0 0 0 0	F≺O∼M≺C≺I∼S≺H≺T
close-depth	1 1 1 1 0 0 0 1	C≺M≺O≺T∼I∼S∼H∼F
PubFig	ACHJ MS V Z	
Masculine-looking	1 1 1 1 0 0 1 1	S≺M≺Z≺V≺J≺A≺H≺C
White	0 1 1 1 1 1 1 1	A≺C≺H≺Z≺J≺S≺M≺V
Young	0 0 0 0 1 1 0 1	V≺H≺C≺J≺A≺S≺Z≺M
Smiling	1 1 1 0 1 1 0 1	J≺V≺H≺A∼C≺S∼Z≺M
Chubby	1 0 0 0 0 0 0 0	V≺J≺H≺C≺Z≺M≺S≺A
Visible-forehead	1 1 1 0 1 1 1 0	J≺Z≺M≺S≺A∼C∼H∼V
Bushy-eyebrows	0 1 0 1 0 0 0 0	M≺S≺Z≺V≺H≺A≺C≺J
Narrow-eyes	0 1 1 0 0 0 1 1	M≺J≺S≺A≺H≺C≺V≺Z
Pointy-nose	0 0 1 0 0 0 0 1	A≺C≺J∼M∼V≺S≺Z≺H
Big-lips	1 0 0 0 1 1 0 0	H≺J≺V≺Z≺C≺M≺A≺S
Round-face	1 0 0 0 1 1 0 0	H≺V≺J≺C≺Z≺A≺S≺M

$$O_m: \{(\text{  } \succ \text{  }), \ldots\},$$

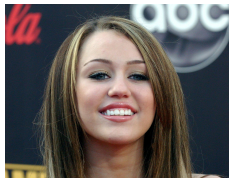
$$S_m: \{\{\text{  } \sim \text{  }\}, \ldots\}$$

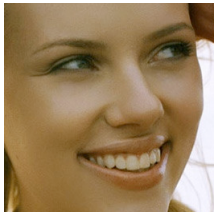
CV app: pairwise ranking

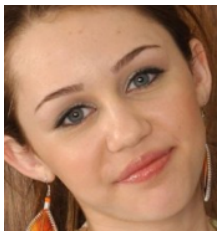
- Coarse labeling at category level => noisy pair sampling

	Binary	Relative
OSR	TI S HC OMF	
natural	0 0 0 0 1 1 1 1	T≺I∼S≺H≺C∼O∼M∼F
open	0 0 0 1 1 1 1 0	T∼F≺I∼S≺M≺H∼C∼O
perspective	1 1 1 1 0 0 0 0	O≺C≺M∼F≺H≺I≺S≺T
large-objects	1 1 1 0 0 0 0 0	F≺O∼M≺I∼S≺H∼C≺T
diagonal-plane	1 1 1 1 0 0 0 0	F≺O∼M≺C≺I∼S≺H≺T
close-depth	1 1 1 1 0 0 0 1	C≺M≺O≺T∼I∼S∼H∼F
PubFig	ACHJ MS V Z	
Masculine-looking	1 1 1 1 0 0 1 1	S≺M≺Z≺V≺J≺A≺H≺C
White	0 1 1 1 1 1 1 1	A≺C≺H≺Z≺J≺S≺M≺V
Young	0 0 0 0 1 1 0 1	V≺H≺S≺J≺A≺S≺Z≺M
Smiling	1 1 1 0 1 1 0 1	J≺V≺H≺A∼C≺S∼Z≺M
Chubby	1 0 0 0 0 0 0 0	V≺J≺H≺C≺Z≺M≺S≺A
Visible-forehead	1 1 1 0 1 1 1 0	J≺Z≺M≺S≺A∼C∼H∼V
Bushy-eyebrows	0 1 0 1 0 0 0 0	M≺S≺Z≺V≺H≺A≺C≺J
Narrow-eyes	0 1 1 0 0 0 1 1	M≺J≺S≺A≺H≺C≺V≺Z
Pointy-nose	0 0 1 0 0 0 0 1	A≺C≺J∼M∼V≺S≺Z≺H
Big-lips	1 0 0 0 1 1 0 0	H≺J≺V≺Z≺C≺M≺A≺S
Round-face	1 0 0 0 1 1 0 0	H≺V≺J≺C≺Z≺A≺S≺M

Scarlett Johansson *vs* Miley Cyrus

O_m: {( ≺ OK )}

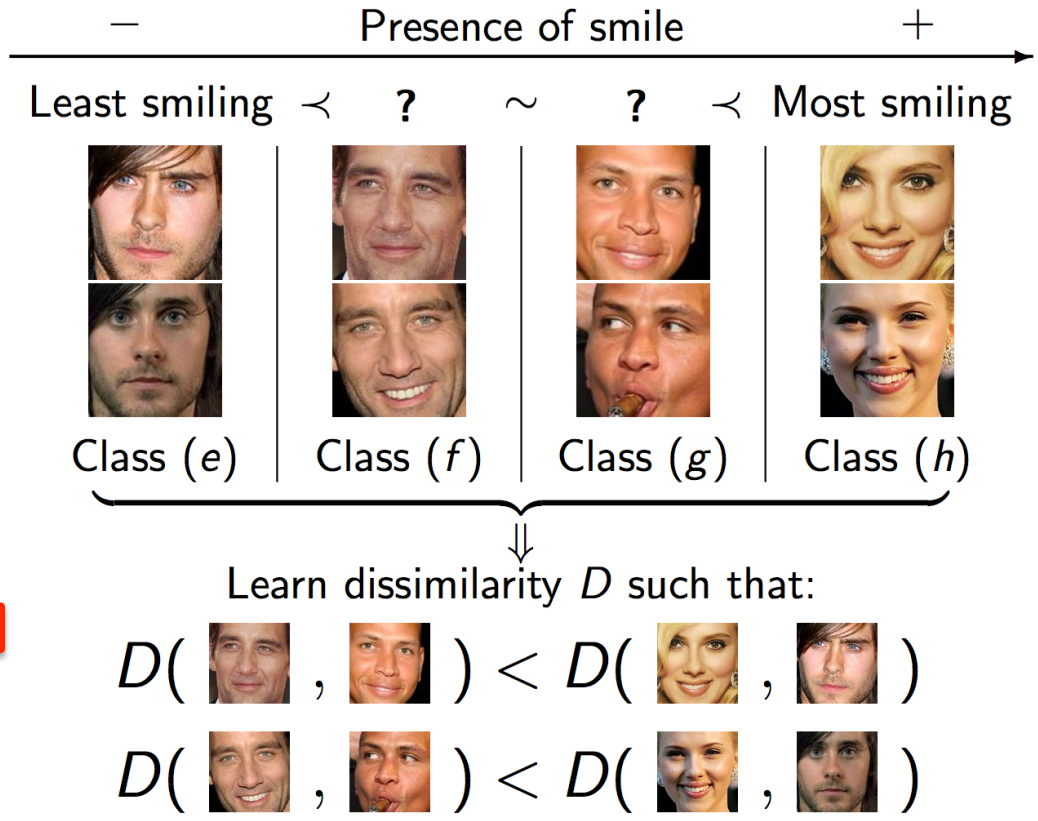
{( ≺ ? )}

{( ≺ NO )}

- Quadruplet to minimize this artefact

CV app: Quadruplet-wise ML

	Binary	Relative
OSR	TI S HC OMF	
natural	0 0 0 0 1 1 1 1	T≺I∼S≺H≺C∼O∼M∼F
open	0 0 0 1 1 1 1 0	T∼F≺I∼S≺M≺H∼C∼O
perspective	1 1 1 1 0 0 0 0	O≺C≺M∼F≺H≺I≺S≺T
large-objects	1 1 1 0 0 0 0 0	F≺O∼M≺I∼S≺H∼C≺T
diagonal-plane	1 1 1 1 0 0 0 0	F≺O∼M≺C≺I∼S≺H≺T
close-depth	1 1 1 1 0 0 0 1	C≺M≺O≺T∼I∼S∼H∼F
PubFig	ACHJ MS V Z	
Masculine-looking	1 1 1 1 0 0 1 1	S≺M≺Z≺V≺J≺A≺H≺C
White	0 1 1 1 1 1 1 1	A≺C≺H≺Z≺J≺S≺M≺V
Young	0 0 0 0 1 1 0 1	V≺H≺C≺J≺A≺S≺Z≺M
Smiling	1 1 1 0 1 1 0 1	J≺V≺H≺A∼C≺S∼Z≺M
Chubby	1 0 0 0 0 0 0 0	V≺J≺H≺C≺Z≺M≺S≺A
Visible-forehead	1 1 1 0 1 1 1 0	J≺Z≺M≺S≺A∼C∼H∼V
Bushy-eyebrows	0 1 0 1 0 0 0 0	M≺S≺Z≺V≺H≺A≺C≺J
Narrow-eyes	0 1 1 0 0 0 1 1	M≺J≺S≺A≺H≺C≺V≺Z
Pointy-nose	0 0 1 0 0 0 0 1	A≺C≺J∼M∼V≺S≺Z≺H
Big-lips	1 0 0 0 1 1 0 0	H≺J≺V≺Z≺C≺M≺A≺S
Round-face	1 0 0 0 1 1 0 0	H≺V≺J≺C≺Z≺A≺S≺M



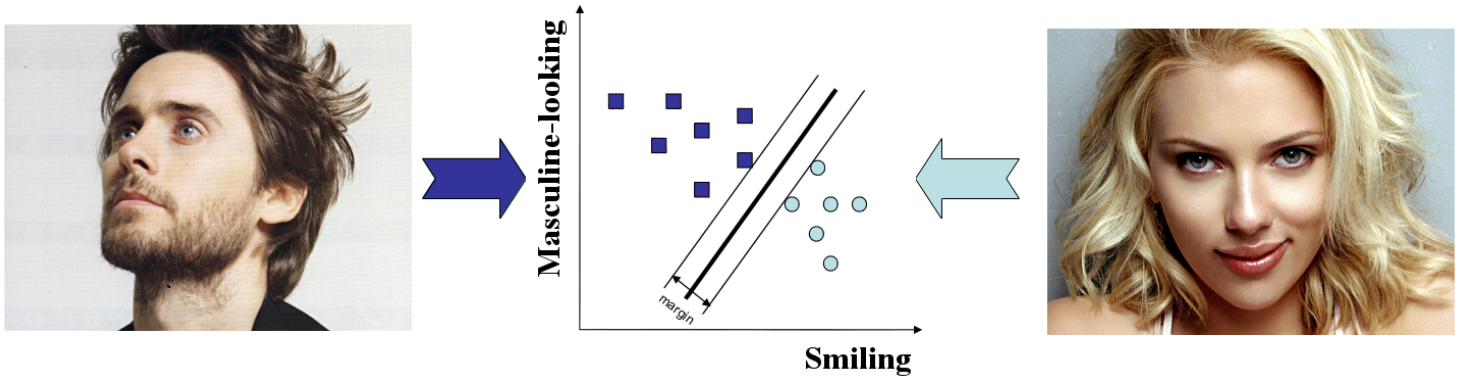
- Relative attributes => (Dis)similarity Learning under Qwise constraints

Relative attribute learning

- Learning a feature space

$$\begin{aligned} D^2_{\mathbf{M}}(p_i, p_j) &= \Phi(p_i, p_j)^\top \mathbf{M} \Phi(p_i, p_j) \\ &= (\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{L}^\top \mathbf{L} (\mathbf{x}_i - \mathbf{x}_j) \end{aligned}$$

- Corresponds to learn a linear transformation parameterized by $\mathbf{L} \in \mathbb{R}^{M \times d}$ such that $\mathbf{h}_i = \mathbf{L}\mathbf{x}_i$ where the m-th row of $\mathbf{L}$ is $\mathbf{w}_m^\top$
- Application to Actor retrieval and classification:



Relative attribute learning

$$\min_{\mathbf{w}} \mu\|\mathbf{w}\|_2^2 + \sum_{\substack{(p_i,p_j,p_k,p_l) \\ D(\text{face}_i,\text{face}_j) < D(\text{face}_k,\text{face}_l) \\ D(\text{face}_i,\text{face}_j) < D(\text{face}_k,\text{face}_l)}} \ell\left(\mathbf{w}^\top [\Psi(p_k,p_l) - \Psi(p_i,p_j)]\right)$$

- $\mathbf{x}_i \in \mathbb{R}^d$: GIST (+ color) descriptor
- $\Psi(p_i, p_j) = \mathbf{x}_i - \mathbf{x}_j$
- Relative attributes a_m for $m \in \{1, \ldots, M\}$: smiling, masculine-looking young...
- Learning a $\mathbf{w}_m$ for each attribute a_m using Qwise optimization
- Resulting in learning a linear transformation parameterized by $\mathbf{L} \in \mathbb{R}^{M \times d}$

$$\mathbf{L} = \begin{bmatrix} w_{1,1} & \cdots & w_{1,d} \\ \vdots & \vdots & \vdots \\ w_{M,1} & \cdots & w_{M,d} \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1^\top \\ \vdots \\ \mathbf{w}_M^\top \end{bmatrix}, \quad \mathbf{w}_m^\top : m\text{-th row}$$

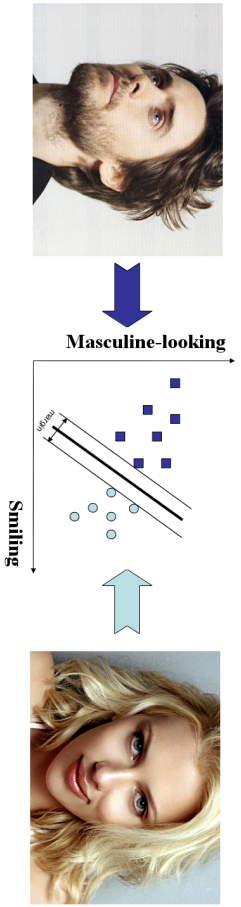
Relative attribute experiments

- Outdoor Scene Recognition OSR [Oliva 01]
- 8 classes, ~2700 images, GIST
- 6 attributes: open, natural …

- Public Figures Faces PubFig [Kumar 09]
- 8 classes, ~800 images, GIST +color
- 11 attributes: smiling, shubby …

Relative attribute experiments

- Baselines
  - RA Relative attribute method (Parikh and Grauman)
    - annotations on class relationships with pairwise constraints
  - LMNN Linear transformation learned [Wein.09]
    - class membership information used only unlike RA
  - RA + LMNN: Combination of the first two baselines
    1. Relative attribute annotations to learn attribute space
    2. Metric in attribute space with LMNN
- Qwise Method:
  - Qwise constraints generated as pairwise
  - Qwise output alone or combined Qwise + LMNN



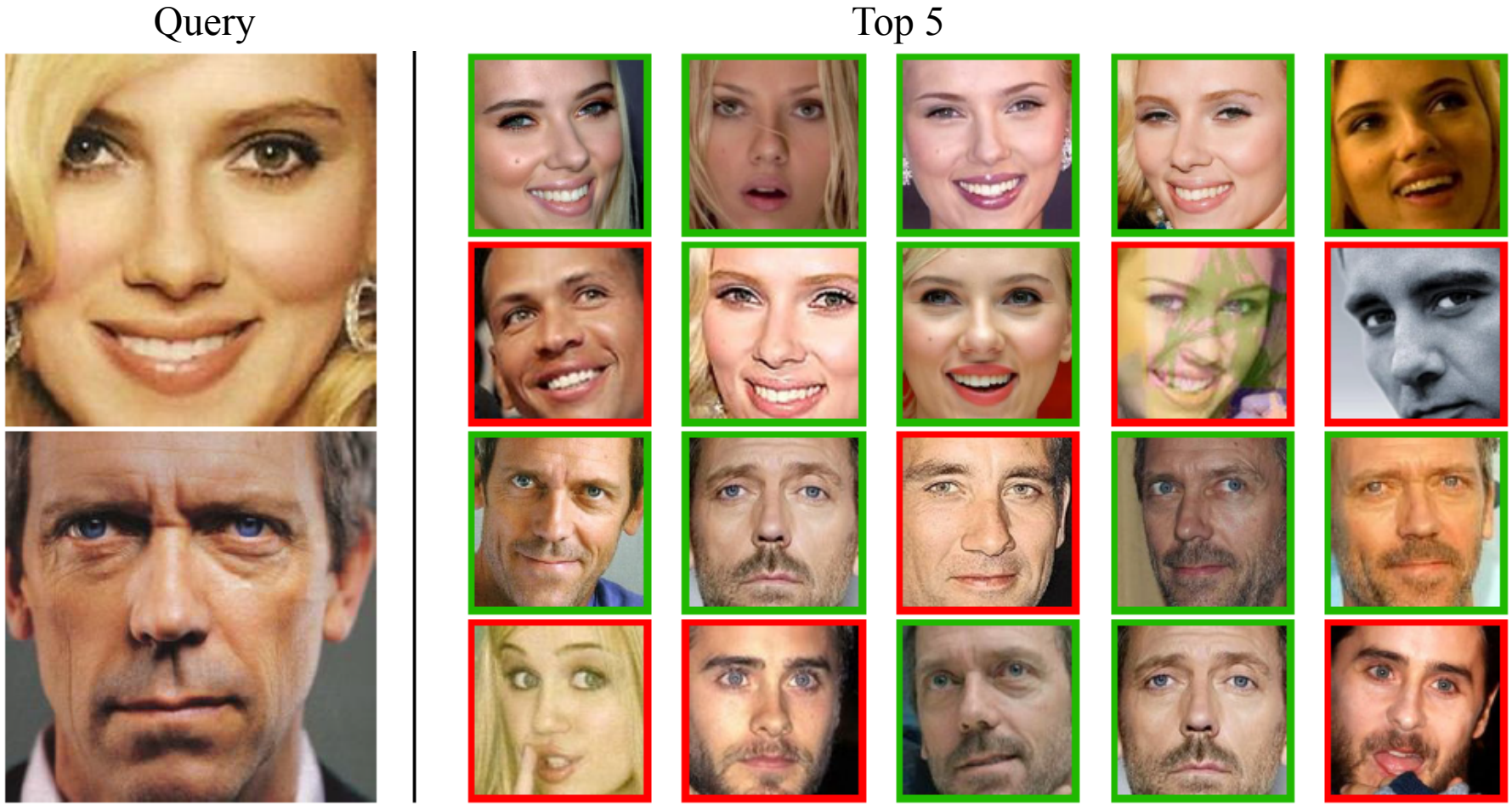
[Wein.09] K.Q. Weinberger, and L.K. Saul, Distance metric learning for large margin nearest neighbor classication, In JMLR 2009

Relative attribute experiments

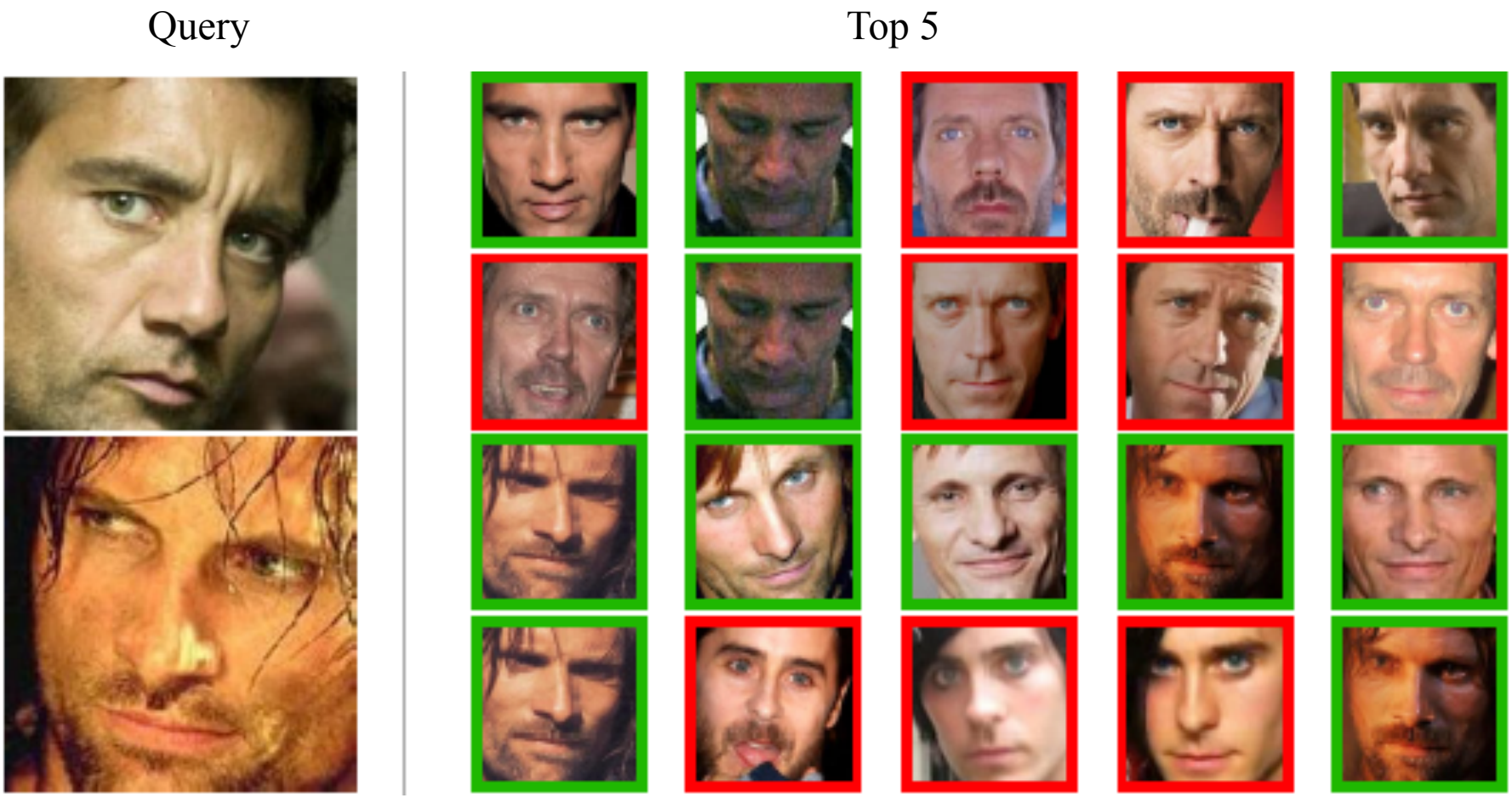
	OSR	Pubfig
Parikh's code	$71.3 \pm 1.9\%$	$71.3 \pm 2.0\%$
LMNN-G	$70.7 \pm 1.9\%$	$69.9 \pm 2.0\%$
LMNN	$71.2 \pm 2.0\%$	$71.5 \pm 1.6\%$
RA + LMNN	$71.8 \pm 1.7\%$	$74.2 \pm 1.9\%$
Qwise	$74.1 \pm 2.1\%$	$74.5 \pm 1.3\%$
Qwise + LMNN-G	$\mathbf{74.6 \pm 1.7}\%$	$76.5 \pm 1.2\%$
Qwise + LMNN	$74.3 \pm 1.9\%$	$\mathbf{77.6 \pm 2.0}\%$

Table 1: Test classification accuracies on the OSR and Pubfig datasets for different methods.

Relative attribute experiments



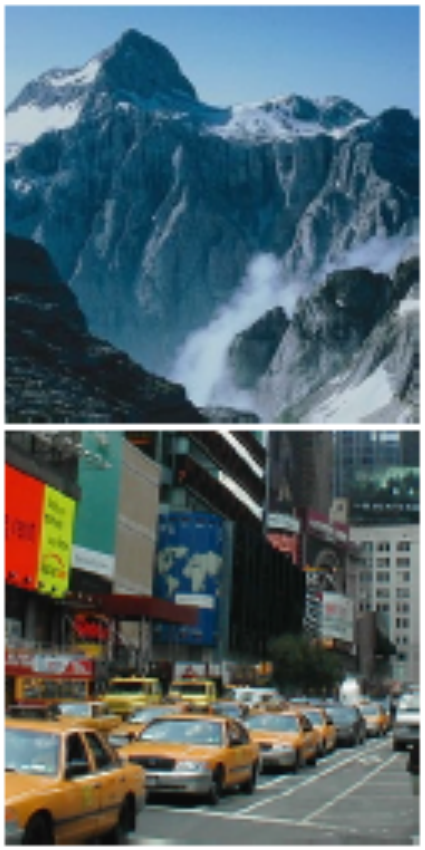
Relative attribute experiments



Relative attribute experiments

Relative attribute experiments

Query



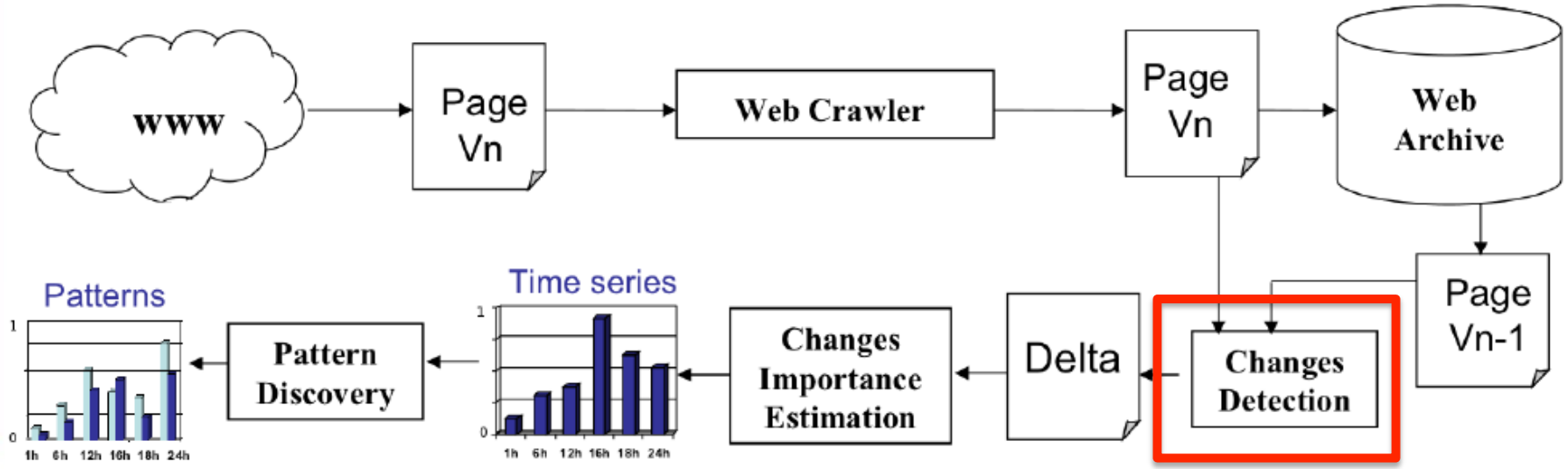
Top 5

Outline

1. Introduction
2. Metric Learning
3. Computer Vision Applications
   - Relative attribute learning
   - **Web page comparison**

Web page ML

- Context:
  - For Web crawling purpose, useful to understand the change behavior of websites over time [AWUPCP11]



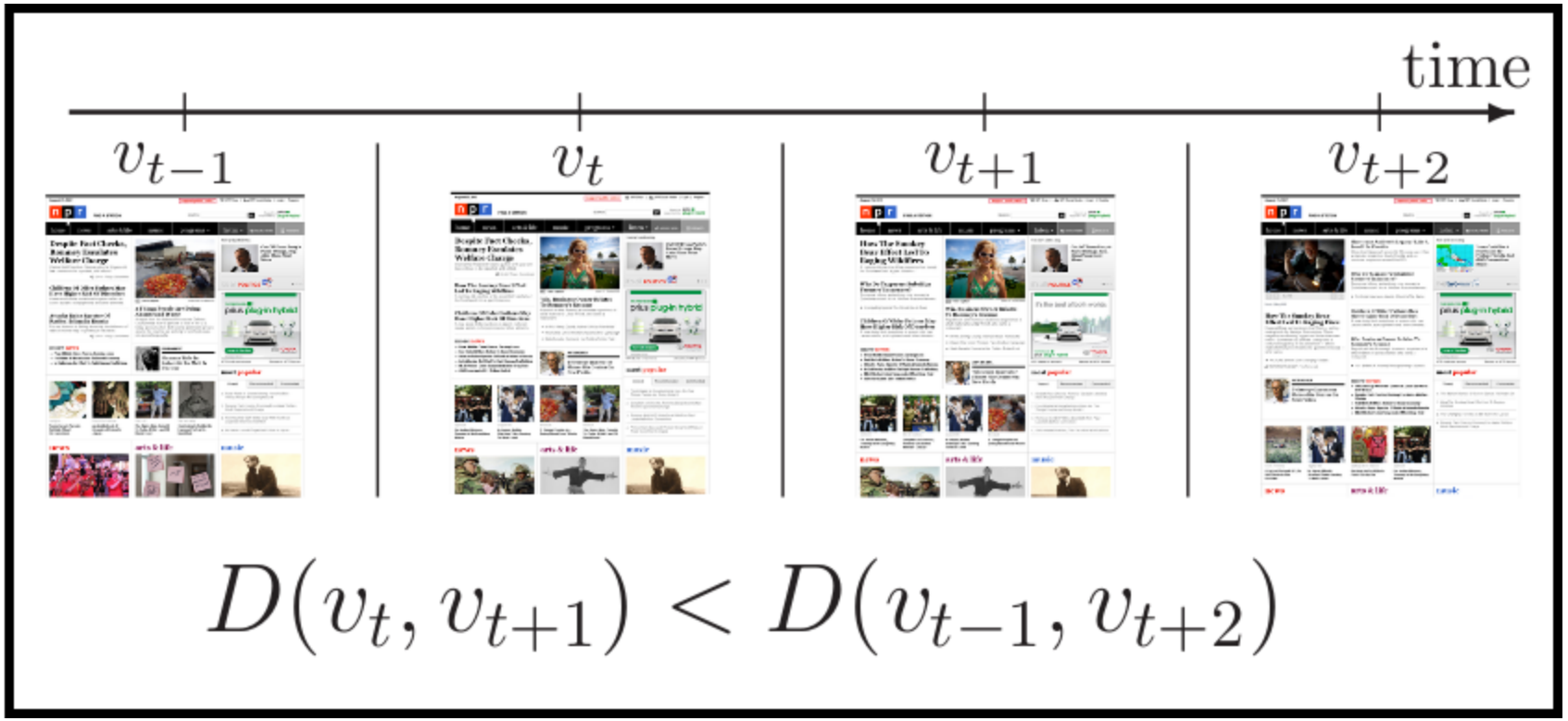
  - Significant changes between successive versions of a same webpage => revisit the page
- Web page comparison
  - Qwise to learn Web page metric and significant webpage regions

Web page ML

- Focus on news websites
  - Advertisements or menus not significant
  - News content significant
- Find a metric able to properly identify **significant** changes between webpage versions
- Localize changes inside pages [Song04]:
  - semantic spatial structure
  - significant to capture

Web page ML

- Triplet/Qwise Constraints:
  - Fully unsupervised ML, but temporal information available
  - Constraints by comparing screenshots of successive webpage versions



Web page ML

- Descriptors: GIST on m-by-m grid over screenshots
- Ψ is a m-by-m vector of Euclidean distance between blocks
- Diagonal PSD matrix: **w** represents block weights
- Optimization over **w**
  - Learning of spatial weights of webpage regions using temporal relationships
  - Automatically
    - » Discovering important change regions
    - » Ignoring menus and advertisements

Web page ML

- Evaluation and Comparison
  - Crawling 50 days Several sites CNN, NPR, BBC, …
  - Manual change detection (news updates) for GT on 5 days
  - Baselines: Euclidean Dist, LMNN
  - GIST on 10x10
  - Mean Average Precision on succ. Web page Metric scores

Site	CNN			NPR			New York Times			BBC		
Eval.	AP_S	AP_D	MAP	AP_S	AP_D	MAP	AP_S	AP_D	MAP	AP_S	AP_D	MAP
Eucl. Dist.	68.1 ±0.6	85.9 ±0.6	77.0 ±0.5	96.3 ±0.2	89.5 ±0.5	92.9 ±0.3	69.8 ±0.9	79.5 ±0.4	74.6 ±0.5	91.1 ±0.3	76.7 ±0.6	83.9 ±0.4
LMNN	78.8 ±1.9	91.7 ±1.7	85.2 ±1.8	98.0 ±0.6	92.5 ±1.1	95.2 ±0.9	83.2 ±1.4	89.1 ±2.7	86.1 ±2.0	92.5 ±0.4	**80.1 ±1.0**	**86.3 ±0.6**
Qwise	**82.7 ±4.1**	**94.6 ±1.8**	**88.6 ±2.9**	**98.6 ±0.2**	**94.3 ±0.6**	**96.5 ±0.4**	**85.5 ±5.4**	**92.3 ±4.1**	**88.9 ±4.6**	**92.8 ±0.4**	79.3 ±1.3	86.1 ±0.8

Web page ML

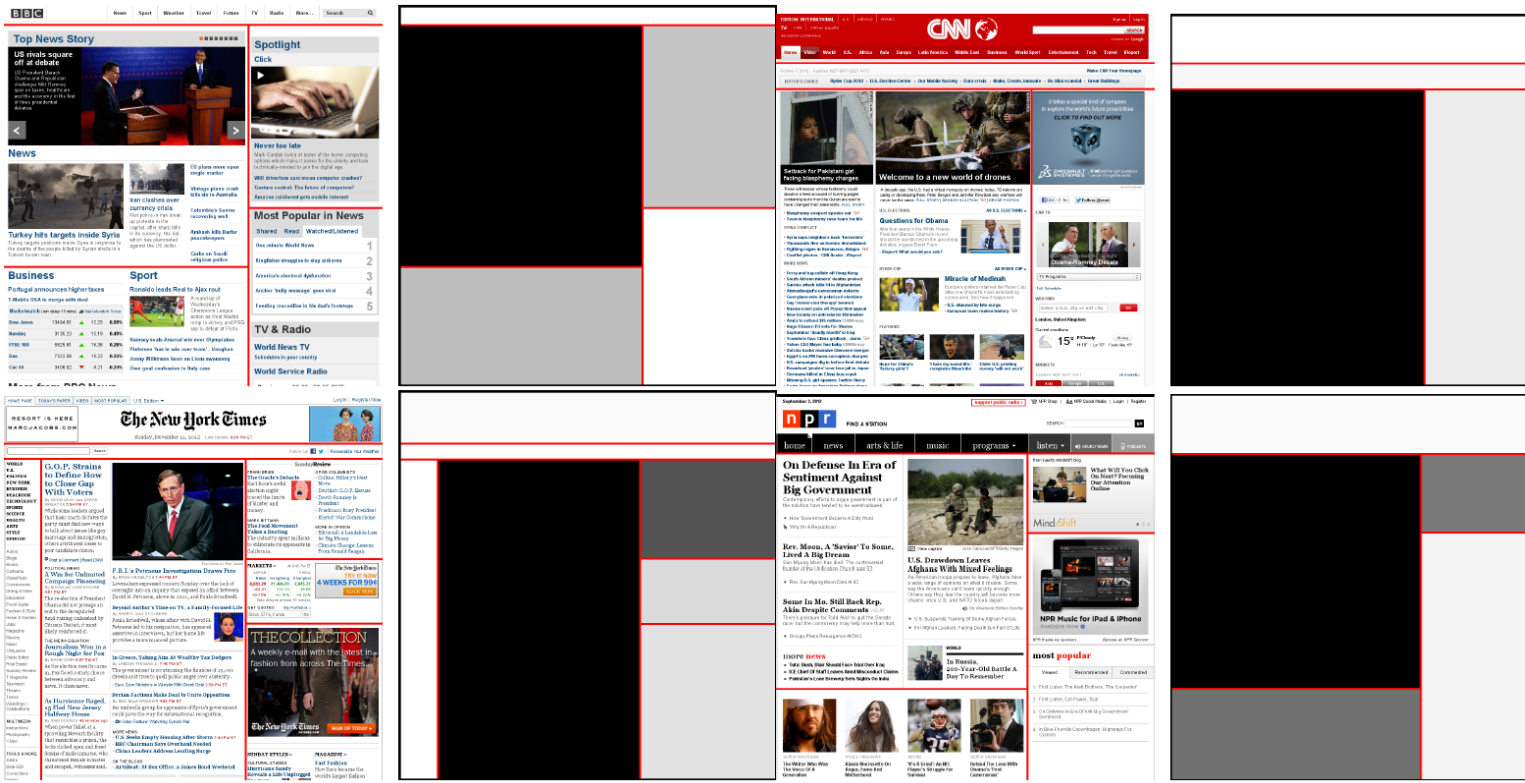

Most Popular in News
TV & Radio
Welcome to a new world of drones
The New York Times
On Defense In Era of Sentiment Against Big Government
most popular

Web page ML

- Not connected to the structural layout of the Web page

Web page ML

- Detect significant changes using the source code of pages (Segmentation) + Qwise



Conclusion

- Key issues in Metric Learning for CV:
  - Modeling: Data representation, type of metric (linear, non lin., local)
    - Connection to deep : deep features + metric learn on top
  - Learning Paradigm: unsupervised, semi-supervised, transfer**, type of constraints**
    - Temporal/spatial relationships, class relationships => rich context to learn metrics or semantic embedding
  - Optimization issues: Global/local solution, Convexity, Scalability, …
  - Learning joint embedding

Thx

Matthieu Cord
Joint work with Marc T. Law and Nicolas Thome
LIP6, Computer Science Department UPMC Paris 6 - Sorbonne University
http://webia.lip6.fr/~cord

Metric learning:

- C. LeBarz *et al* Exemplar based metric learning for robust visual localization, ICIP 2015
- M.T. Law, N. Thome and M. Cord. Fantope Regularization in Metric Learning, CVPR 2014
- M.T. Law, N. Thome and M. Cord. Quadruplet-wise Image Similarity Learning, ICCV 2013
- M.T. Law, N. Thome, S. Gancarski and M. Cord. Structural and Visual Comparisons for Web Page Archiving, ACM DocEng, 2012

Image representation:

- S. Avila, N. Thome, M. Cord, E. Valle, A. Araujo, Pooling in Image Representation: the Visual Codeword Point of View, CVIU 2012
- H. Goh, , N. Thome, M. Cord, JH. Lim, Top-Down Regularization of Deep Belief Networks, NIPS 2013

Many Codes available on demand

VISIIR PROJECT http://visiir.lip6.fr/ COOKING ANALYSIS