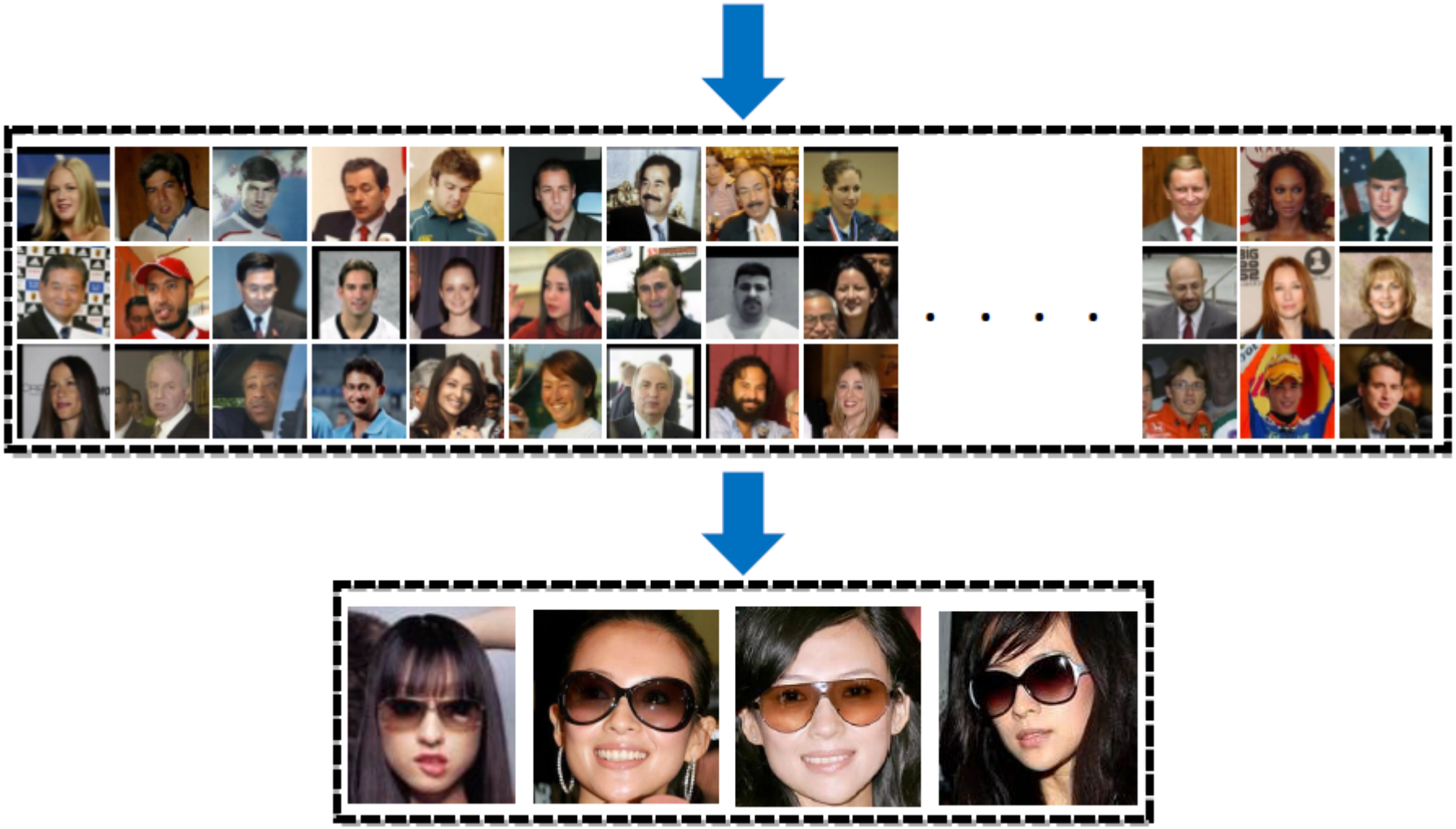
Image Ranking and Retrieval based on Multi-Attribute Queries

B. Siddiquie, R. S. Feris, L. S. Davis

Attribute Based Image Retrieval

"Young Asian woman wearing sunglasses"



Attributes

Physical traits or characteristics of a person

Male
Asian
Middle-Aged

Female
White
Young

Object Properties

Round
White
Black

Object properties that span across object categories

Striped
Four-legged
Orange
Black

Striped
Four-legged
White
Black

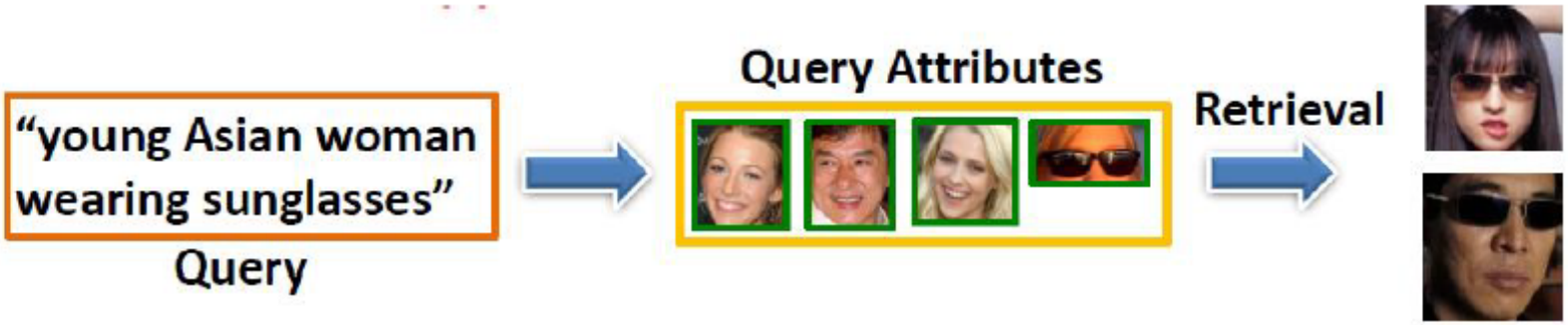
Related Work

- Several image retrieval frameworks which are based on searching a text base query in a given corpus,use a two stage progress
  - First, images are retrieved based purely on textual features
    - Tags (e.g in Flickers)
    - Html meta-data
  - Second step is to run a classifier, trained based on visual features, to filter and rerank the results from the first stage.
- The problem is; textual information in first stage is not available always. For example: survilance.
- Another problem is that they learn a seperate ranking/ classification function corresponding to each query term

Related Work

- Recently, attribute based representation is used in wide area.
  - Object recognition
  - Learning model of unkown object categories and detecting unexpected attributes in known object classes.
  - Comparing faces based on facial attributes and other visual traits can improve face verification.
- Most of these works exploit the fact that attributes provide a high level representation which is compact and semantically meaningful.

Multi-Attribute Image Retrieval

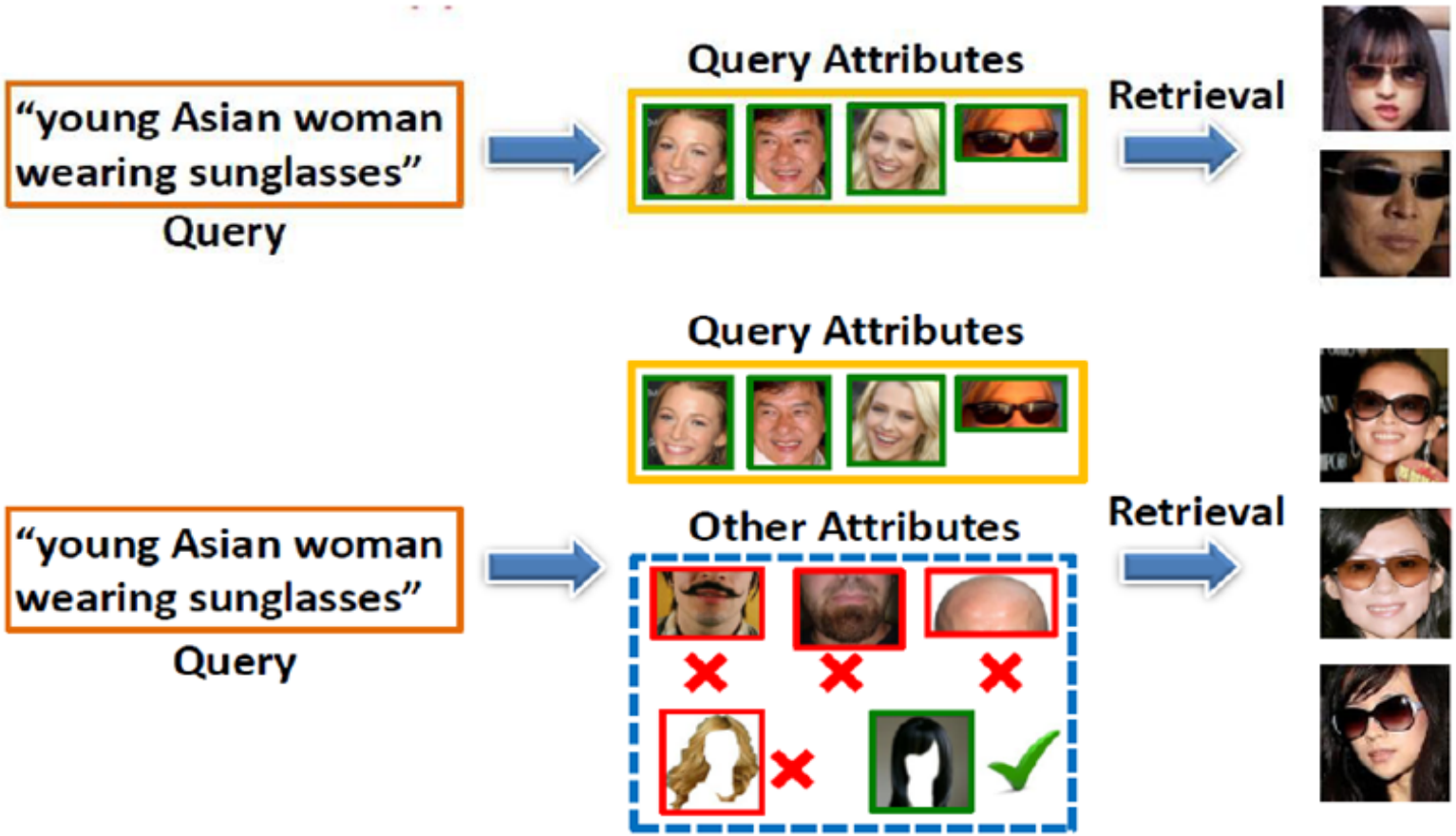


Number of possible queries is exponential

Existing Approaches

- Train independent classifiers for each attribute
- Sum up confidence scores

Multi-Attribute Image Retrieval



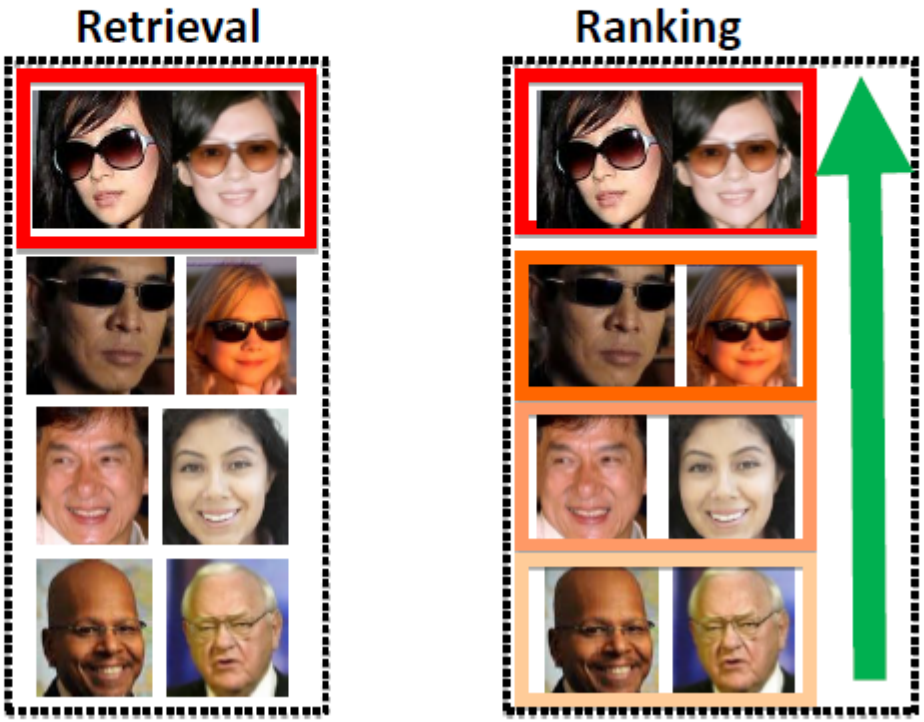
Model Correlations between attributes

- **Explicitly utilize information from non-query attributes**

Multi-Attribute Image Retrieval

Joint Ranking and Retrieval Framework

- **Retrieval:** Set of images
- **Ranking:** Ordered set of images



Multi-Attribute Image Ranking/Retrieval

We are given:

- An attribute vocabulary

$$\mathcal{X} = \{x_1, x_2, \ldots, x_K\}$$

- A set of training images

$$\mathcal{Y} = \{y_1, y_2, \ldots, y_N\}$$

- Multi-label annotation for each image

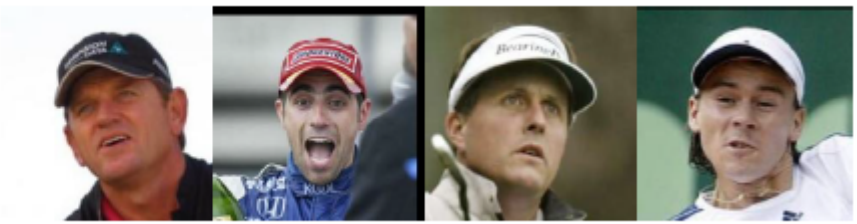
Hat
Sunglasses
Female
Asian
Young

Our goal is:

- For a multi-label query $\mathcal{Q}$, where $\mathcal{Q} \subset \mathcal{X}$

- Rank/Retrieve relevant images from a dataset

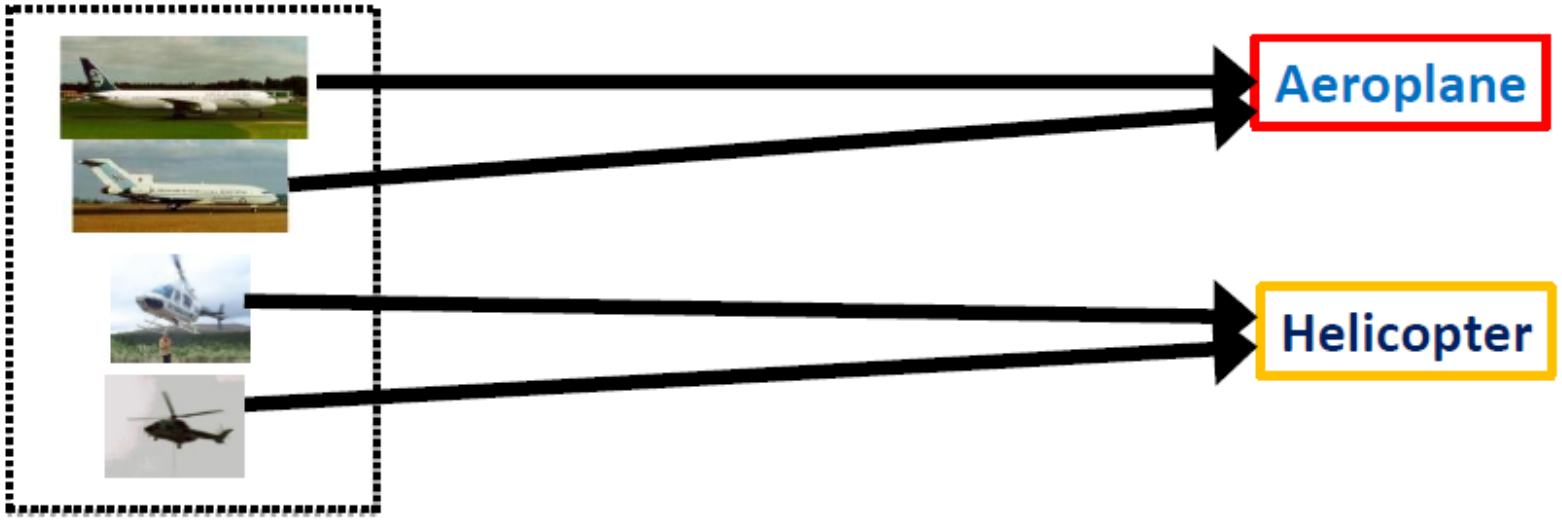
Male
Hat



Retrieval: Reverse Learning

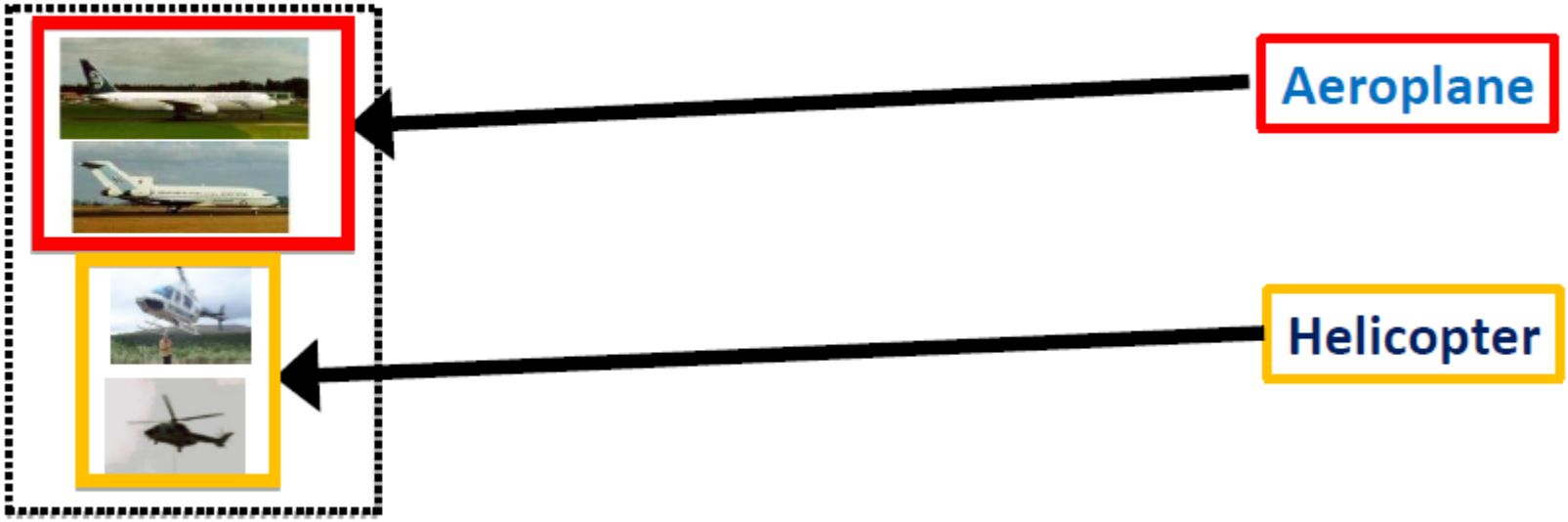
- **Reverse Multi-label Learning, Petterson and Caetano, NIPS 2010**
- **Given a label** x_i **such that** $x_i \in \mathcal{X}$
- **Predict the set of instances** $y(\subset \mathcal{Y})$ **that containing the label** x_i

Retrieval: Conventional Learning



Retrieval: Reverse Learning

- Given a label x_i such that $x_i \in \mathcal{X}$
- Predict the set of instances $y(\subset \mathcal{Y})$ that containing the label x_i



- Enables minimization of training loss based on a variety of metrics

Retrieval: Formulation

Given multi-attribute query $\mathcal{Q}$, output set of relevant images y

Learn w such that $y^* = \arg\max_{y \subset \mathcal{Y}} w^T \psi(\mathcal{Q}, y)$

where $w^T \psi(\mathcal{Q}, y) = \sum_{x_i \in \mathcal{Q}} w_i^a \Phi_a(x_i, y) + \sum_{x_i \in \mathcal{Q}} \sum_{x_j \in \mathcal{X}} w_{ij}^p \Phi_p(x_j, y)$

Asian + woman + wearing Sunglasses

$\phi_a(x$

Asian . Blonde hair
Asian . Gray hair

Asian . Black hair

Asian woman wearing Sunglasses

Retrieval: Training

Training

$$\arg\min_{w,\xi} \quad w^T w + C \sum_t \xi_t$$

$$\forall \ t \quad w^T \psi(\mathcal{Q}_t, y_t^*) - w^T \psi(\mathcal{Q}_t, y_t) \geq \boxed{\Delta(y_t^*, y_t)} - \xi_t$$

loss function

$$\boxed{\Delta(y^*, y)} = \begin{cases} 1 - \frac{y \cap y^*}{y} & \text{precision} \\ 1 - \frac{y \cap y^*}{y^*} & \text{recall} \\ 1 - \frac{y \cap y^* + \bar{y} \cap \bar{y}^*}{\mathcal{Y}} & \text{hamming loss} \end{cases}$$

Ranking

- **Given a query Q, rank documents in order of relevance**
  - **Output is an ordered set(permutation)**

- **Large Scale Datasets**
  - **Ranking is more important then retrieval**

Ranking: Formulation

Given a multi-attribute query $\mathcal{Q}$ **, generate permutation of images** z

Learn w **such that** $z^* = \arg\max_{z \in \pi(\mathcal{Y})} w^T \psi(\mathcal{Q}, z)$

where $$w^T \psi(\mathcal{Q}, z) = \sum_{x_i \in \mathcal{Q}} w_i^a \hat{\Phi}_a(x_i, z) + \sum_{x_i \in \mathcal{Q}} \sum_{x_j \in \mathcal{X}} w_{ij}^p \hat{\Phi}_p(x_j, z)$$

and

$$\hat{\Phi}_a(x_i, z) = \sum_{z_k \in z} \boxed{A(r(z_k))} \phi_a(x_i, z_k)$$

$$\hat{\Phi}_p(x_j, z) = \sum_{z_k \in z} \boxed{A(r(z_k))} \phi_p(x_j, z_k)$$

Ranking: Training

Training $\arg\min_{w,\xi} \quad w^T w + C \sum_t \xi_t$

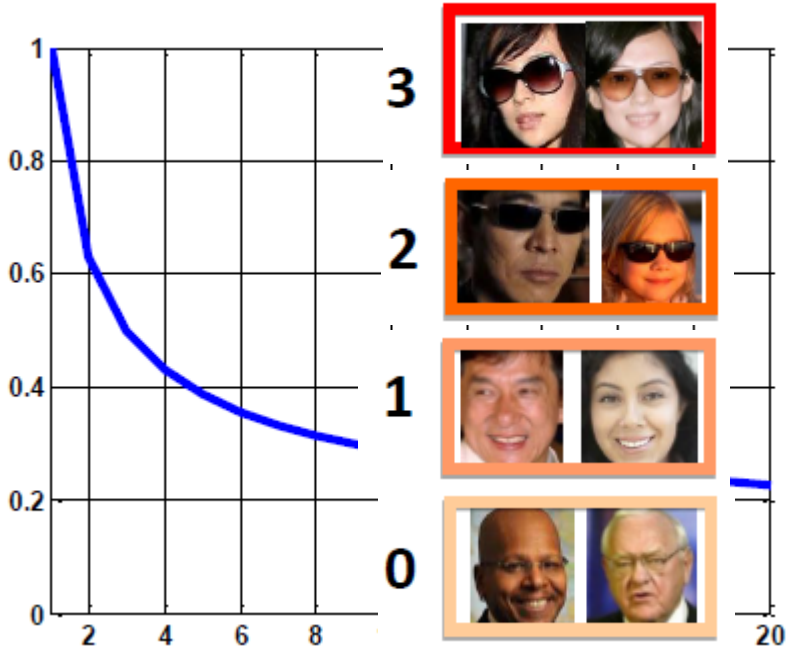
$$\forall \ t \quad w^T \psi(Q_t, z_t^*) - w^T \psi(Q_t, z_t) \geq \boxed{\Delta(z_t^*, z_t)} - \xi_t$$

loss function

$$\boxed{\Delta(z^*, z)} = 1 - \mathrm{NDCG}_k(z^*, z)$$

where

$$\boxed{NDCG_k = \frac{1}{Z} \sum_{j=1}^{k} \boxed{\frac{2^{\mathrm{rel}(j)} - 1}{\log(1+j)}}}$$

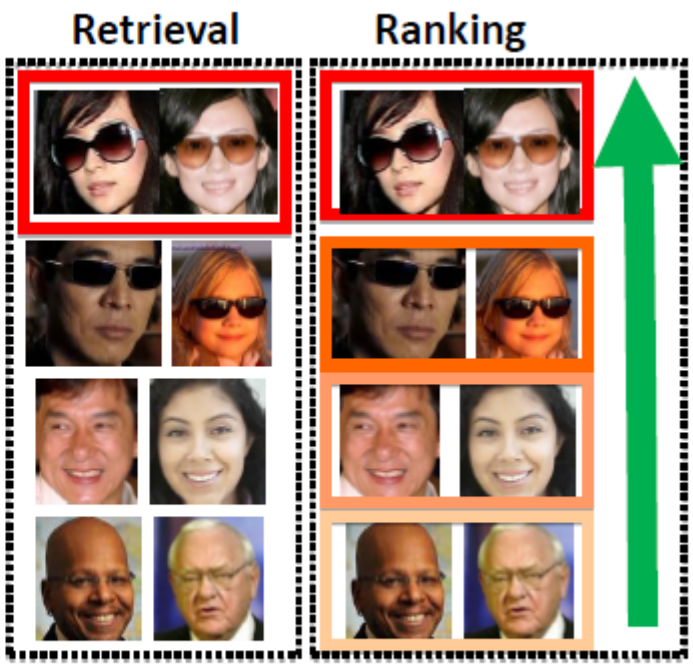


Asian + woman + wearing Sunglasses

Multi-Attribute Ranking & Retrieval

Ranking and Retrieval

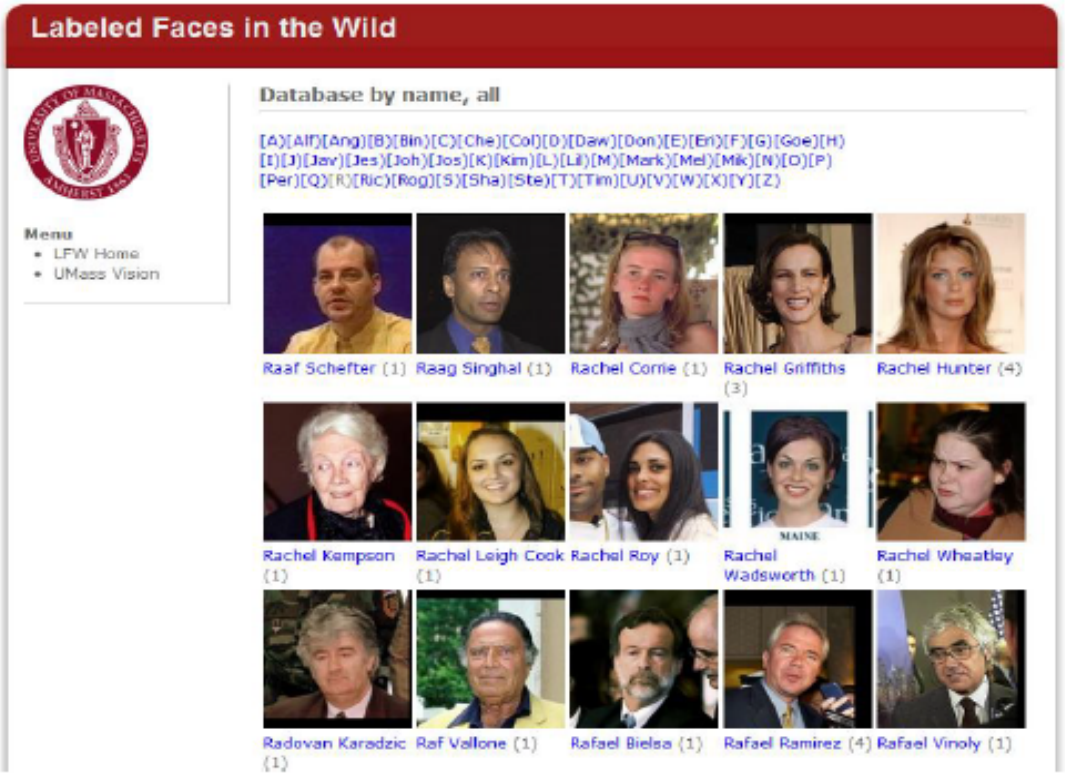
- Typically treated as separate problems



Structured Formulation

- Optimize the same model according to different performance measures

Labeled Faces in the Wild(LFW) dataset



Attribute Annotation

- **9992 images**
- **27 attributes**

LFW : Attributes

Race
- Asian
- Black
- White
- Indian

Age
- Kid
- Youth
- Middle-Aged
- Senior

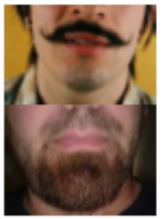
Gender
- Sex

Other
- Hat
- Lipstick
- Visible Forehead

Hair Color
- Black Hair
- Blonde Hair
- Brown Hair
- Gray Hair

Hairstyle
- Long Hair
- Short Hair
- Bangs
- Bald

Facial Hair
- Mustache
- Beard
- Goatee
- No Beard



Eyewear
- Sunglasses
- Eyeglasses
- No Eyewear

LFW: Feature Extraction

Features

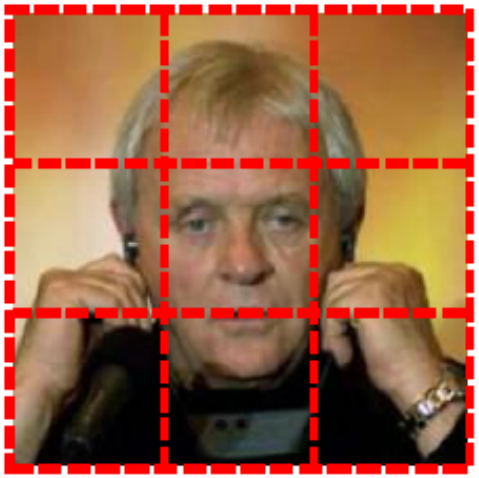
Color
- Color Histograms
- Color Correlograms
- Color Moments
- Color Wavelet

Texture
- Wavelet Texture
- LBP Histogram
- LBP PCA

Skin Information
- Skin Bitmap
- Skin Color
- Spatial Skin

Shape
- Edge Histogram
- Shape Moments
- SIFTogram

Features: Spatial Configurations



Center

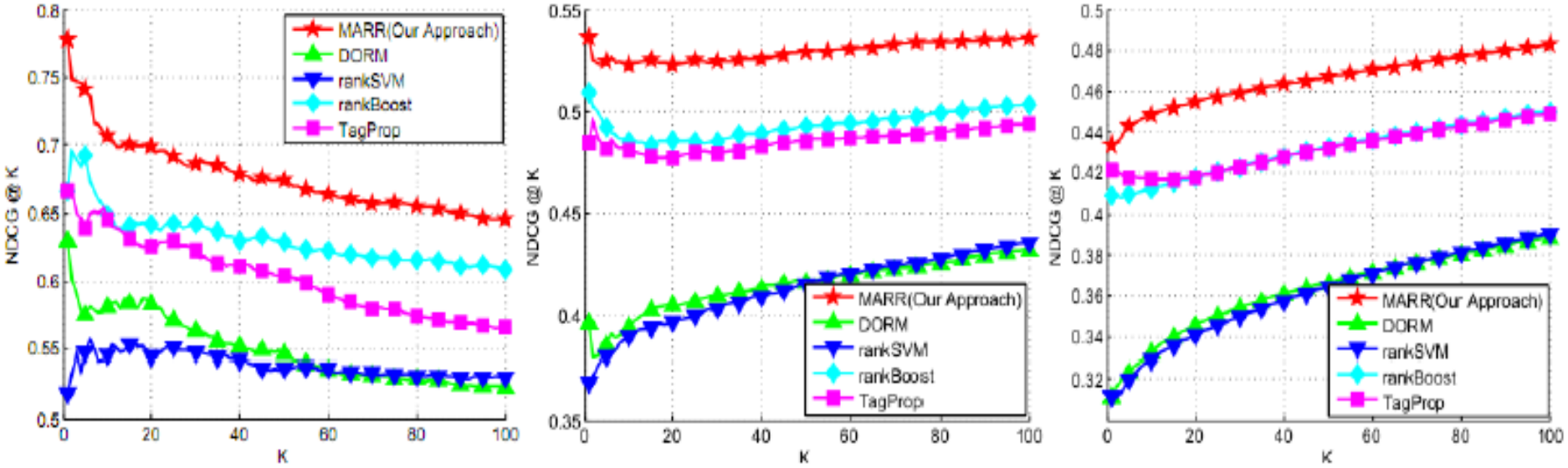
Global

Horizontal Parts

Vertical Parts

Layout

LFW: Quantitative Results

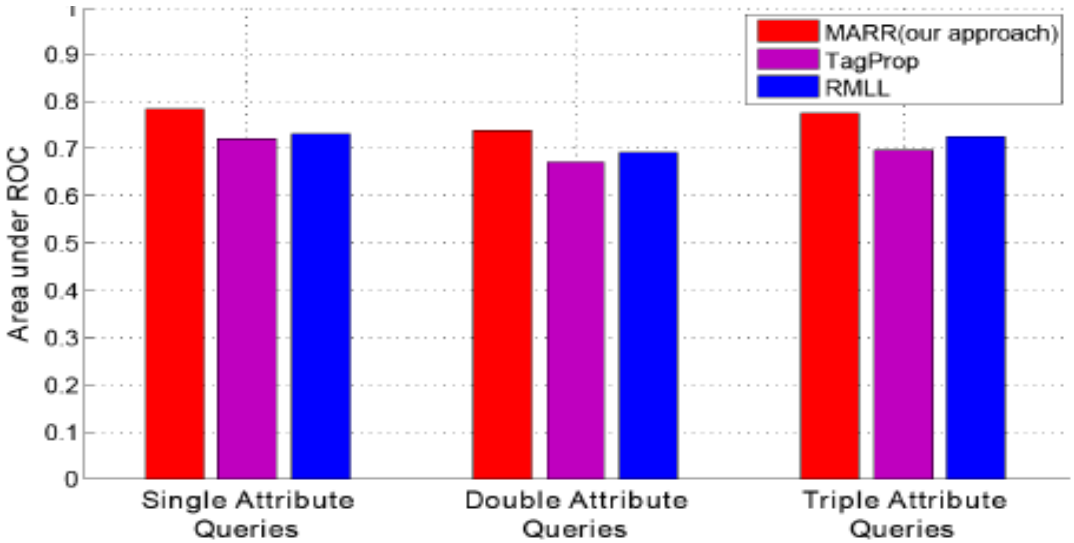


(a) Single Attribute Queries (b) Double Attribute Queries (c) Triple Attribute Queries

Results

- **rankBoost is the 2nd best**
- **Performance gain**
  - Single Attribute Queries: 8.9% improvement in NDCG@10
  - Double Attribute Queries: 7.7% improvement in NDCG@10
  - Triple Attribute Queries: 8.8% improvement in NDCG@10

LFW: Quantitative Results



Retrieval

- **Baselines**
  - Reverse Multi-Label Learning, (J. Petterson and T. Caetano, NIPS 2010)
  - TagProp, (M. Guillaumin, T. Mensink, J. Verbeek, C. Schmid, ICCV 2009)
- **mean Area under ROC**
- **Performance gain**
  - ~5% w.r.t. RMLL
  - ~7% w.r.t. TagProp

LFW: Analysis

Commonly Co-Occuring Attribute Pairs

-(Gray-Hair, Senior)

-(Middle-aged,Eyeglasses)

Mutually Exclusive Attribute Pairs

-(White,Asian)

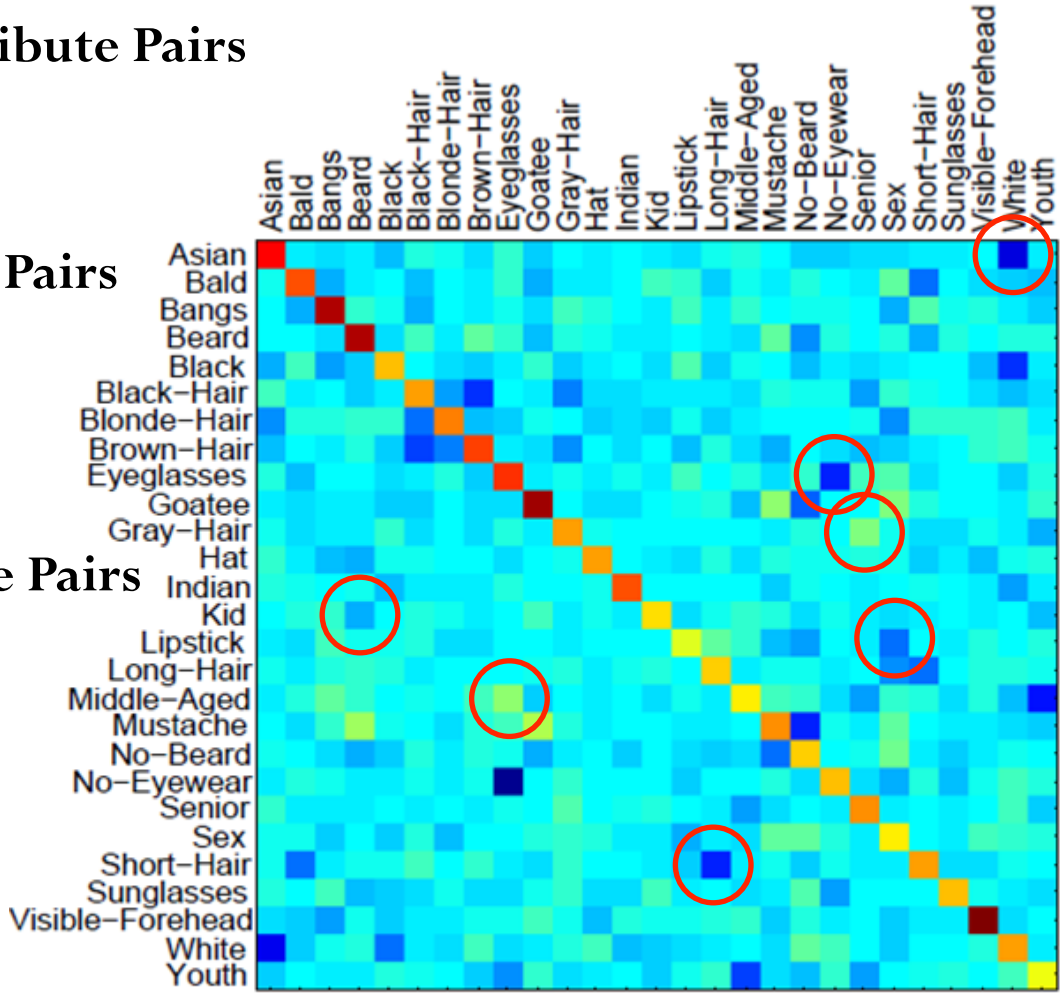
-(Eyeglasses, no-eyewear)

-(Short-hair, Long hair)

Rarely Co-Occuring Attribute Pairs

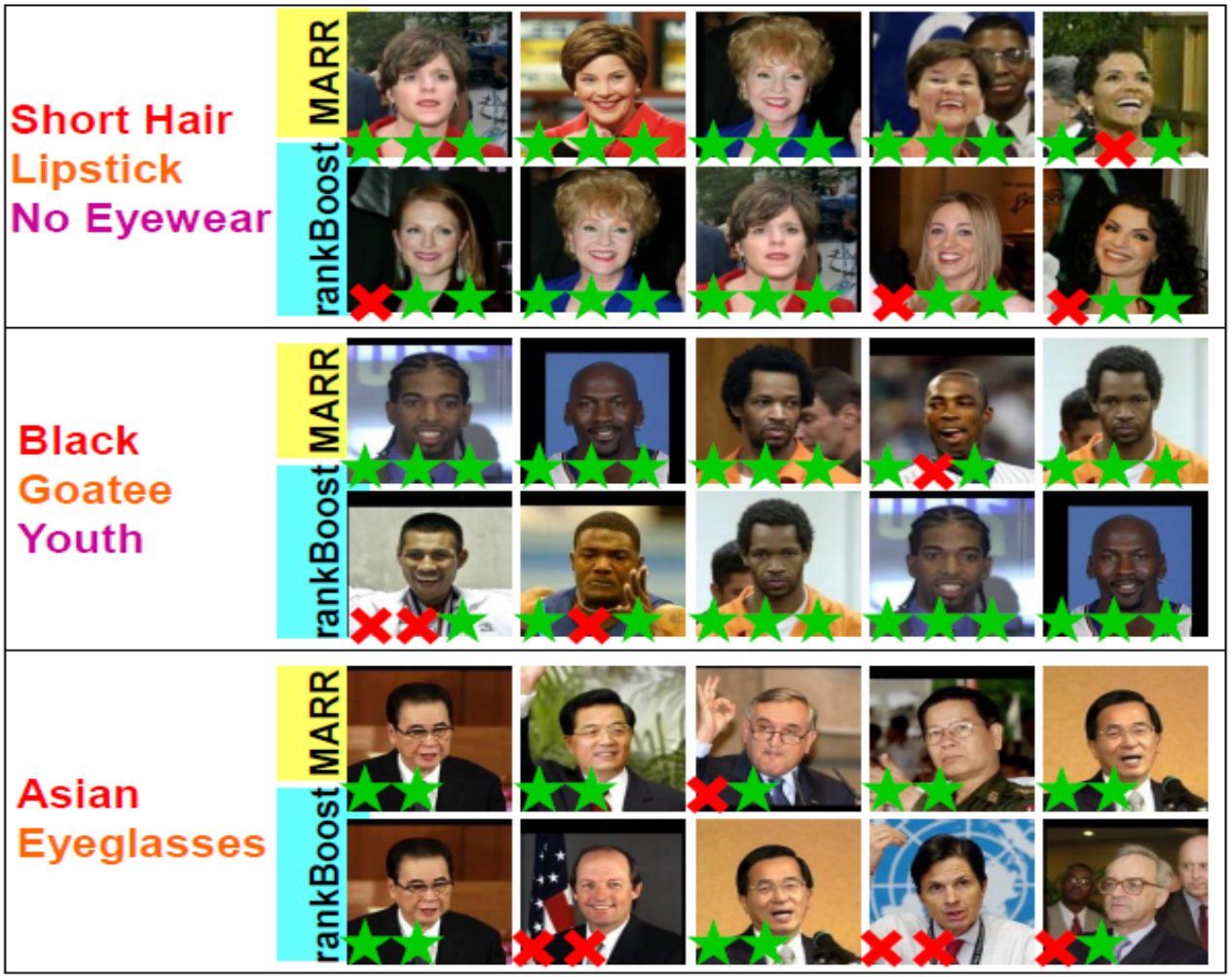
-(Lipstick, Male)

-(Kid, Beard)

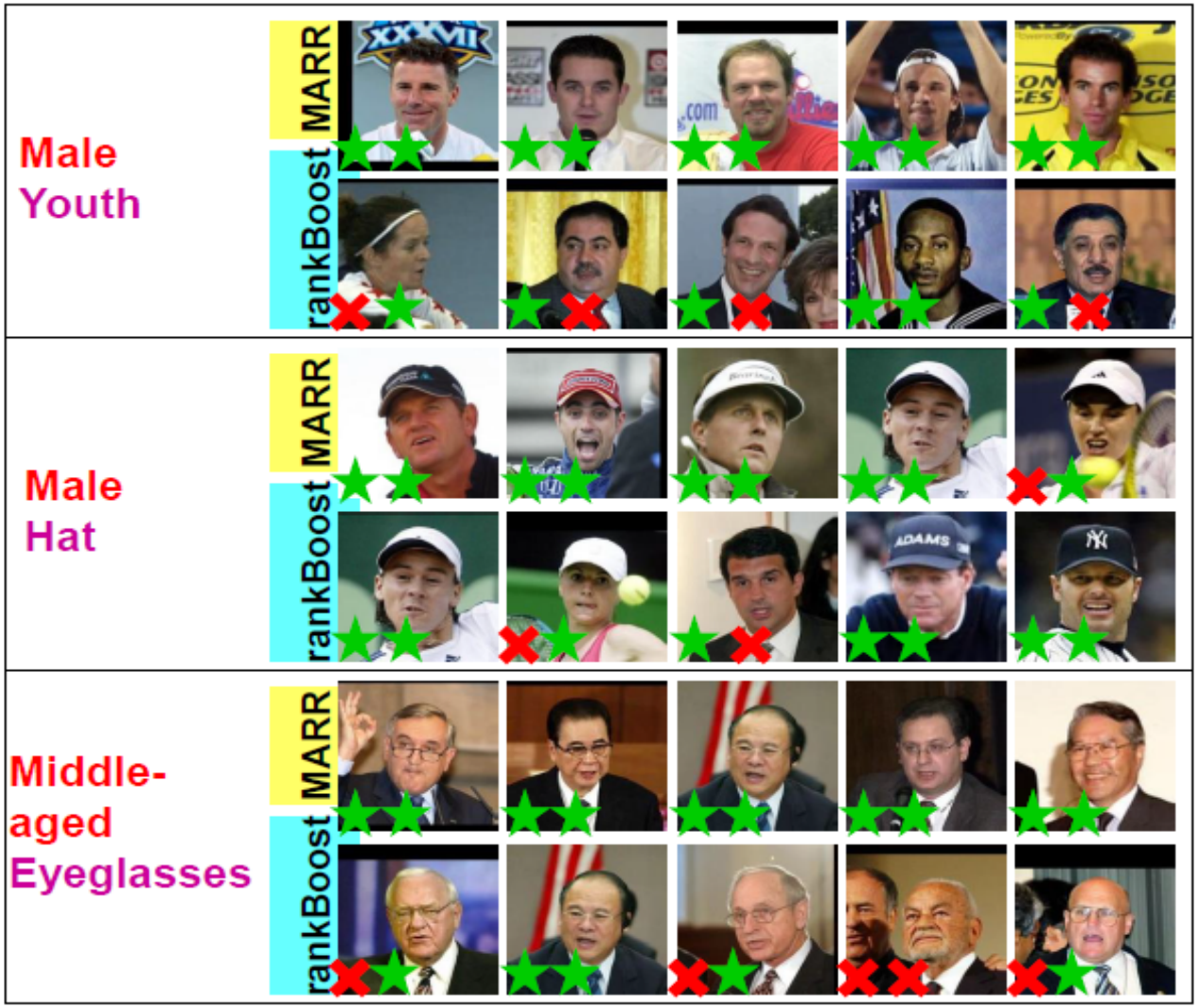


Classifier weights learnt on the LFW dataset, red and yellow indicate high values while blue and green denote low values. (best viewed in color).

LFW: Qualitative Results



LFW: Qualitative Results



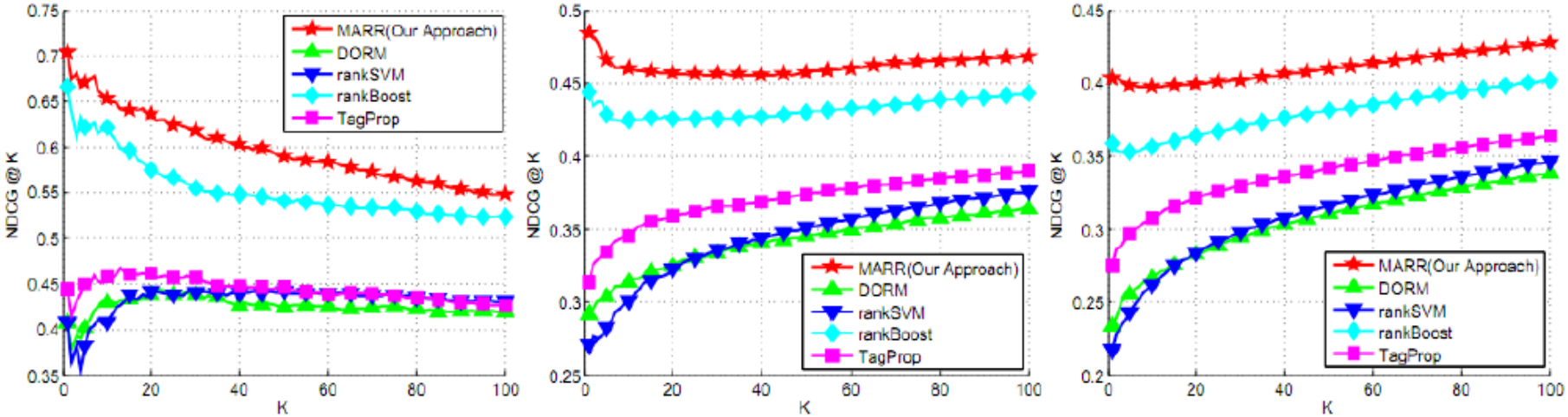
FaceTracer Dataset

(Kumar et. al, ECCV 2008)

Attribute Annotation

- 3000 images
- 27 attributes

FaceTracer: Quantitative Results



(a) Single Attribute Queries (b) Double Attribute Queries (c) Triple Attribute Queries

Results

- **rankBoost is the 2nd best**
- **Performance gain**
  - **Single Attribute Queries: 5.0% improvement in NDCG@10**
  - **Double Attribute Queries: 8.1% improvement in NDCG@10**
  - **Triple Attribute Queries: 11.6% improvement in NDCG@10**

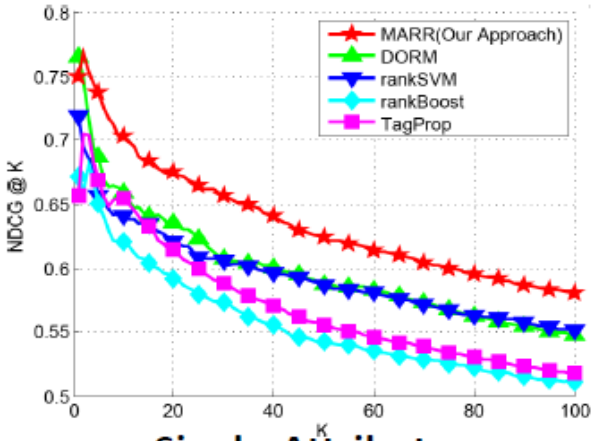
Pascal Dataset

Visual Object Classes Challenge 2008 (VOC2008)

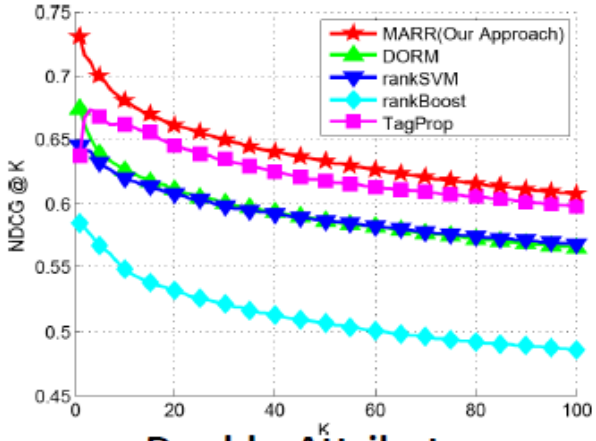
Dataset Statistics

- 12695 images (6340 train, 6355 test)
- 20 classes
  - Airplane, bicycle, bus, horse, person, ...
- 64 Attributes and Parts
  - Attributes: 2D Boxy, Round, Vertical Cylinder, Horizontal Cylinder, ...
  - Parts: Window, Headlight, Text, Leg, ...

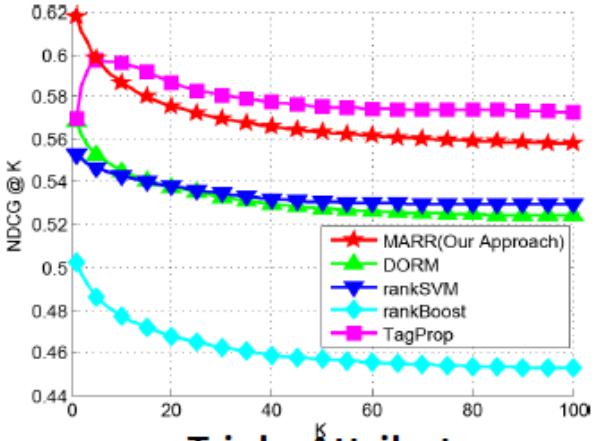
Pascal Dataset: Quantitative results



Single-Attribute Queries



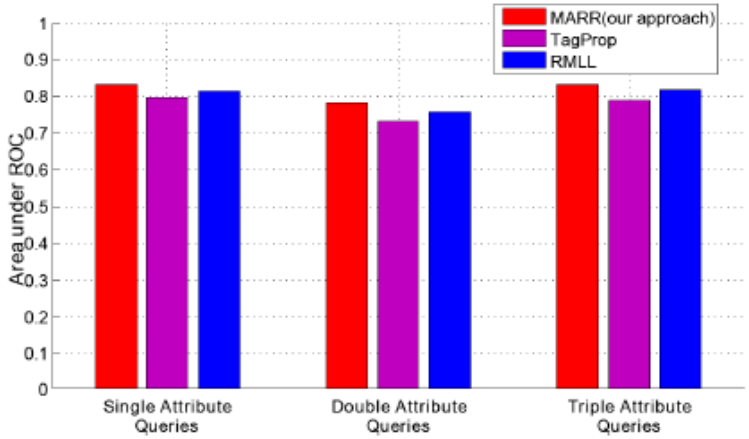
Double-Attribute Queries



Triple-Attribute Queries

Ranking

Retrieval



Contributions

- **Single framework for ranking and retrieval**
- **Supports retrieval and ranking based on multi label queries**
- **Attributes within a single object category and even across multiple object categories are independent**

References

- http://www.cs.umd.edu/~behjat/papers/CVPR11_slides.pdf
- Image Ranking and Retrieval Based on Multi-Attribute Queries, B. Siddiquie, R. Feris and L. Davis, CVPR 2011.