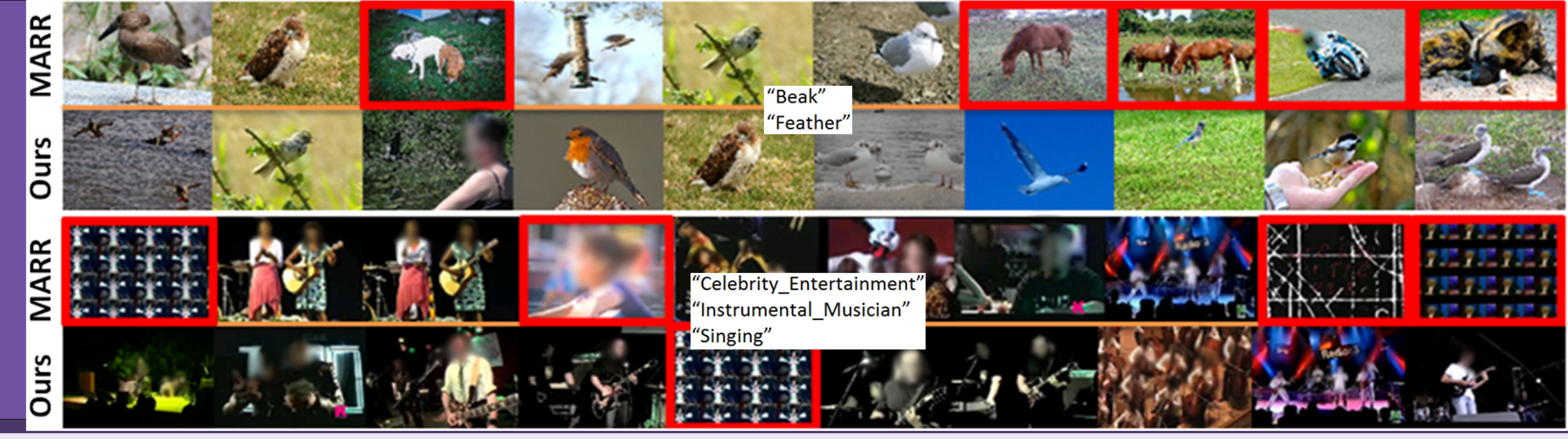
Weak Attributes for Large-Scale Image Retrieval

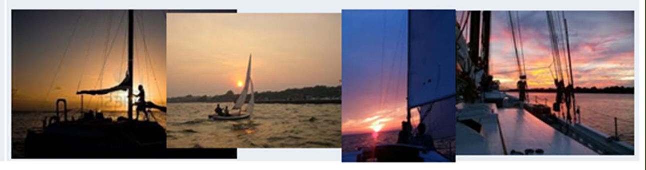
Felix X. Yu Rongrong Ji Ming-Hen Tsai Guangnan Ye Shih-Fu Chang

Columbia University

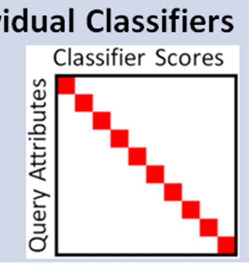
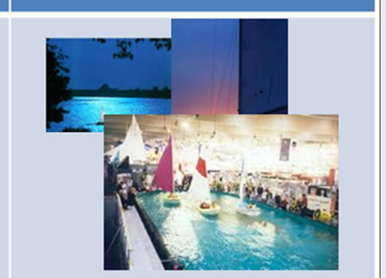
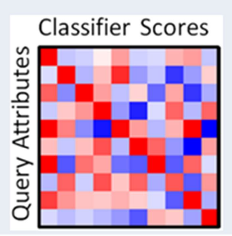
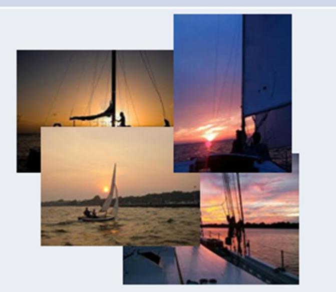


1. Multi-Attribute Image Retrieval

1.1 Problem Definition

Query	Objective
Multiple attributes	Images with ALL attributes
sailing +person +sunset	

1.2 Related Works

Method	Query	Retrieval Model	Result
Individual Classifiers 	Sailing +person +sunset	Scores of Sailing +Person +Sunset	
MARR 		Night × Indoor × Sea √ Sailboat √ Sunset √ …	

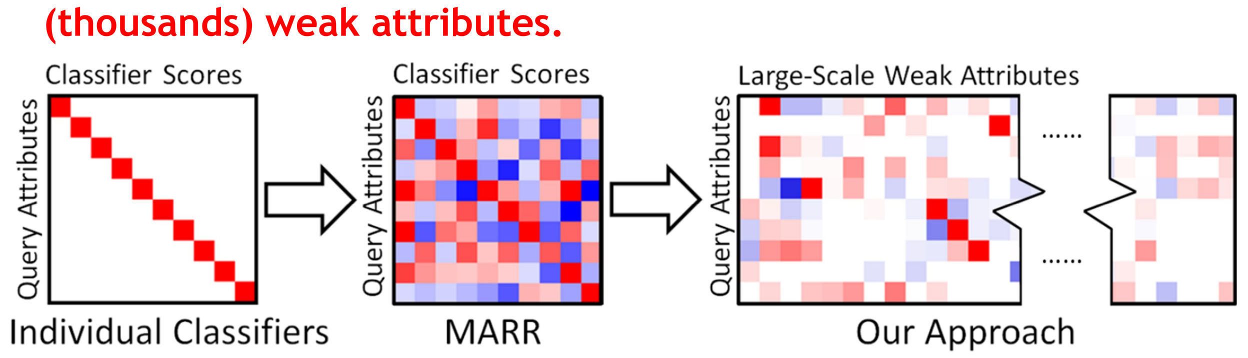
- **Individual Classifiers** [1]: Train attribute classifiers independently and then sum up the scores for multi-attribute queries.
- **MARR** [2] (Image Ranking and Retrieval based on Multi-Attribute Queries): Model dependency of query attributes.
- MARR improves over individual classifiers [2].

1.3 Problem to Solve

- MARR relies only on the pre-labeled attributes to design the dependency model.
- User labeling is a burdensome process. The small amount of attributes are far from sufficient in forming an expressive feature space.

2. Our Solution

- **Weak Attributes** are a collection of mid-level representations, which could be comprised of automatic classifier scores, distances to certain template instances, or even quantization to certain patterns derived through unsupervised learning, all of which can be easily acquired with very little or no human labor.
- **Examples:**
  Automatic classifiers (Columbia374, Classemes 2,659);
  Discriminative Attribute (A is similar to B);
  Relative Attribute (A is more natural than B);
  Topic Models (pLSA LDA);
  …
- **Build dependency model of query attributes on the large-scale (thousands) weak attributes.**



We also impose sparsity to get efficiency and avoid overfitting.

3. The Contributions

Example:
Query: "sailing"+ "person"+ "sunset"
Selected weak attributes: sailing(1.0) person(1.0) sunset(1.0) sea(0.9) sailboat(0.8) random_distance#3(0.2) discriminative#2(0.5) outdoor(0.9) indoor(-0.7) kitchen(-0.8) latent#1(-0.3) random_projection#1(0.4)…

Questions:
1. How to select a subset of weak attributes for answering a specific query?
2. How to learn the weights?

- We propose **weak attributes** that unify various kinds of mid-level image representations which can be easily acquired with **no or little human labor**.
- We apply weak attributes to image retrieval, by modeling **dependency of query attributes on weak attributes** under the framework of structural learning.
- To achieve efficiency and avoid overfitting, we propose a novel **semi-supervised graphical model** to select a subset of weak attributes adaptively for each query. This makes the proposed method applicable to large and general datasets.
- We compile the **largest multi-attribute image retrieval dataset** to date, named a-TRECVID, including 126 fully labeled query attributes and 6,000 weak attributes of 0.26 million images

4. How to select a subset of weak attributes for answering a specific query?

4.1 Maximizing mutual information

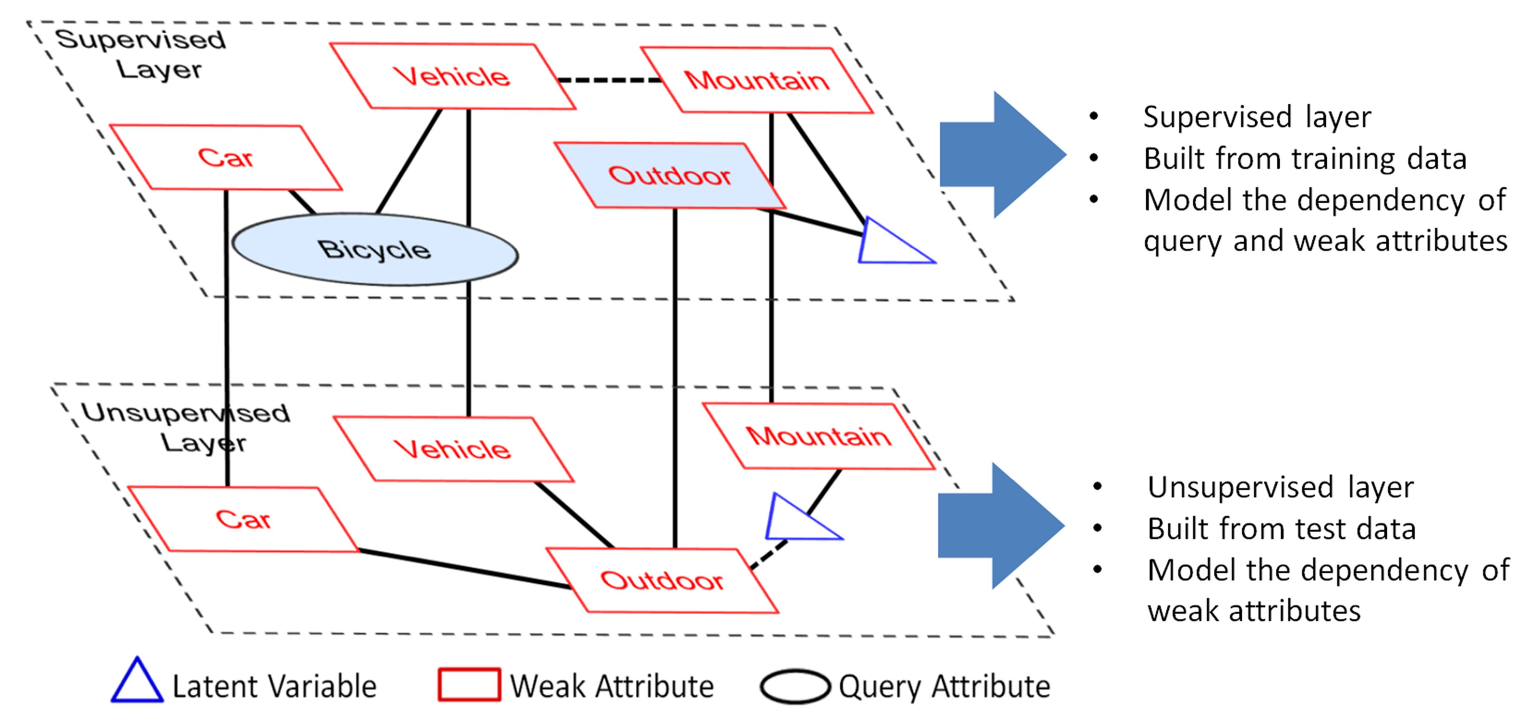
$$\max_{X_Q \subset \mathcal{X}} I(Q; X_Q) \quad s.t. \quad |X_Q| = k$$

Where Q is a multi-attribute query, and X_Q is a subset of weak attributes.

- After some simplification, the above problem can be transformed to model $P(Q|x_i)$ for every x_i (weak attribute).
- $P(Q|x_i)$ plays a key role in bridging weak attributes to multi-attribute query.
- Modeling $P(Q|x_i)$ based on training data only may bias the model, since in large-scale image retrieval, the number of training images is always limited.
- Solution: the semi-supervised graphical model with the alternating inference algorithm.

4.2 The Semi-supervised Graphical Model

A two-layer tree graphical model, which can capture the high-order dependency structure of query and weak attributes with consideration of both training/test data.



4.3 Alternating Inference

- To get $P(Q|x_i)$ from semi-supervised graphical model.
- Inference in each layer is by belief propagation.

Algorithm 1 Alternating Inference

Given a query $Q \subset \mathcal{Q}$, compute $P(Q|x_i = 1)$ (without loss of generality), $\forall x_i \in \mathcal{X}$
while Not convergent (the change of $P_{sup}(x)$ is large) **do**
  Inference on the unsupervised graph to get its marginal distribution $P_{unsup}(x|x_i = 1), \forall x \in \mathcal{X}$.
  Update the margin of $x \in \mathcal{X}$ of the supervised graph $P_{sup}(x) \leftarrow P_{unsup}(x|x_i = 1)$.
  Inference on the supervised graph, to get updated $P_{sup}(x), \forall x \in \mathcal{X}$.
  Update the margin of $x \in \mathcal{X}$ of the unsupervised graph: $P_{unsup}(x) \leftarrow P_{sup}(x)$.
end while
Compute joint distribution $P_{sup}(Q)$ in the supervised graph as output $P(Q|x_i = 1)$.

5. How to learn the weights? (Retrieval Model)

- We use structural SVM as the retrieval model.
- The advantages of structural SVM:
  - Joint optimization of all kinds of queries
  - Direct optimization of different loss functions

5.1 Retrieval

$$Y^* = \arg\max_{Y \subset \mathcal{Y}} \mathbf{w}^T \psi(Q, Y), \quad \mathbf{w}^T \psi(Q, Y) = \sum_{q_i \in Q} \sum_{x_j \in X_Q} w_{ij} \sum_{y_k \in Y} \phi(x_j, y_k).$$

- $\phi(x_j, y_k)$ is the value of weak attribute x_j of image y_k.

5.2 Training

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

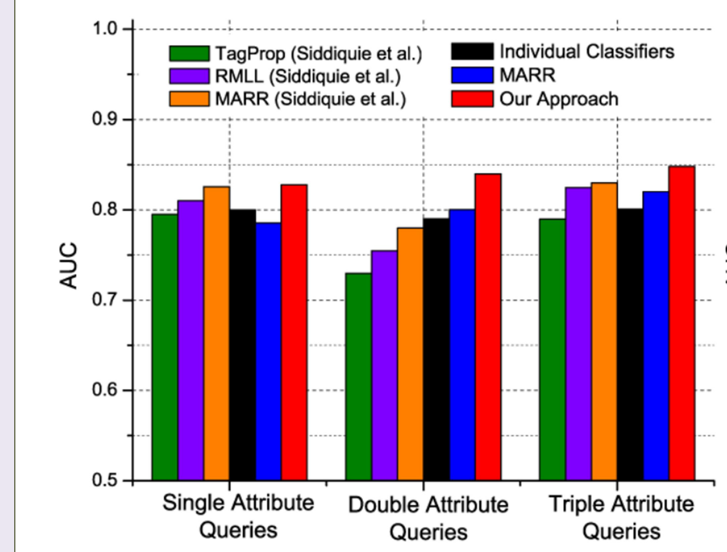
$$\forall t \quad \mathbf{w}^T (\psi(Q_t, Y_t^*) - \psi(Q_t, Y_t)) \geq \triangle(Y_t^*, Y_t) - \xi_t$$

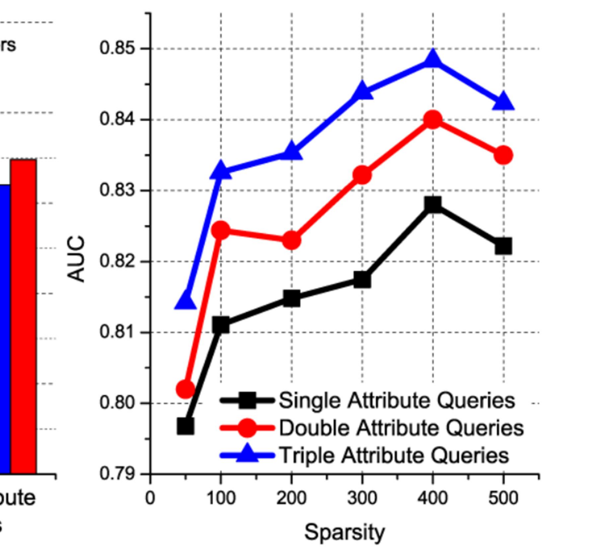
- $\triangle(Y_t^*, Y_t)$ is the loss function, which can be Hamming loss, precision, recall etc.

5. Experiments (more experiments in Technical Report [4])

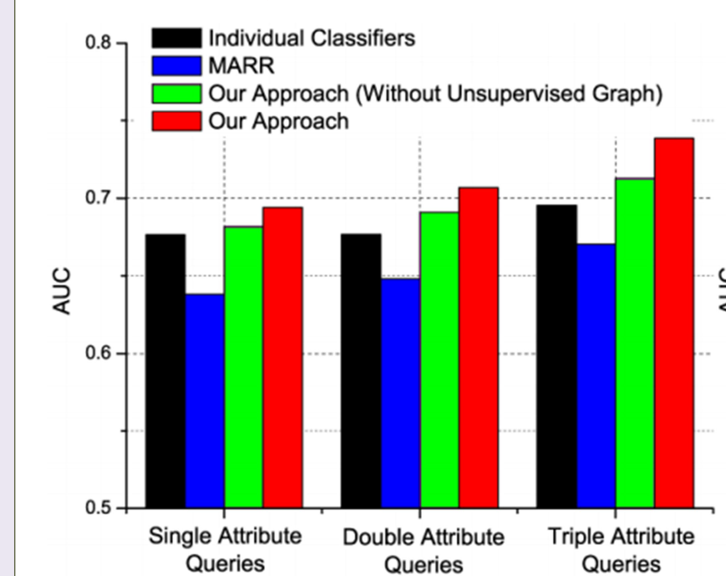
- Weak attributes used: query attribute classifiers, discriminative attributes, classemes[3], random image distances, latent variables.

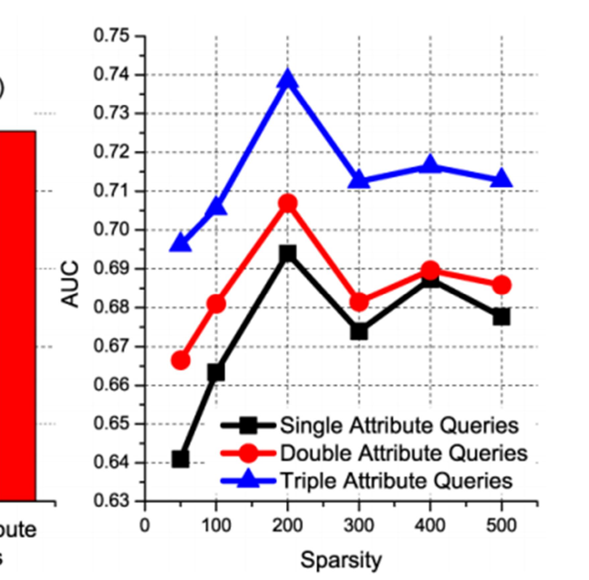
5.1 a-PASCAL





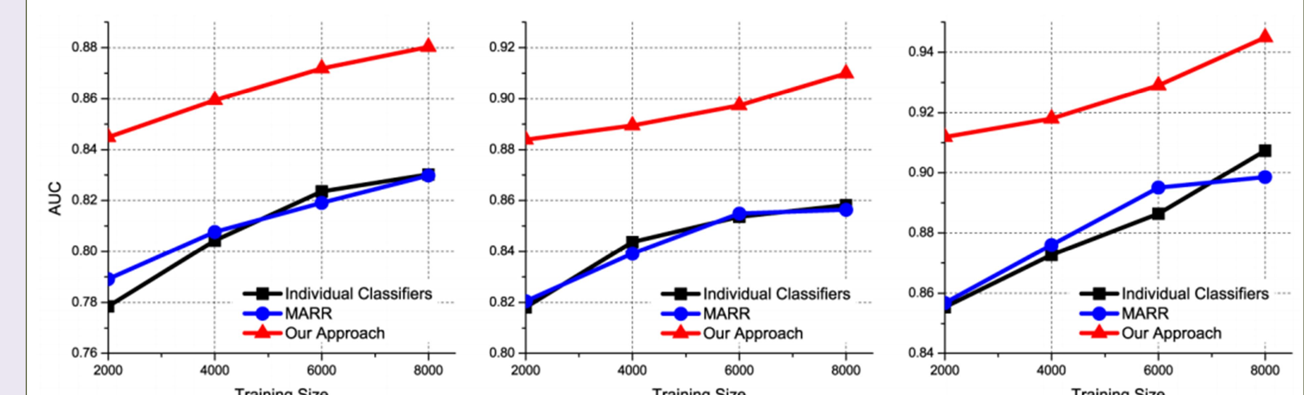
5.2 a-Yahoo (cross-dataset, train on a-PASCAL and test on a-Yahoo)





5.3 a-TRECVID

- Compiled from TRECVID 2011 Semantic Indexing Competition.
- 126 labeled attributes, 6,000 weak attributes, 0.26 million images.
- Data/code: http://www.ee.columbia.edu/dvmm/a-TRECVID/



5.4 Contributions (weights) of different types of weak attributes

	Query Att Classifier	"Classemes"	Disriminative Att	Others
a-PASCAL	10.49%(64)	25.87%(2659)	48.73%(1350)	14.91%(927)
a-Yahoo	3.23%(64)	74.30%(2659)	17.23%(1350)	5.24%(927)
a-TRECVID	20.09%(126)	70.83%(2659)	3.84%(2000)	5.24%(1215)

Reference

[1] N. Kumar, A. Berg, P. Belhumeur, and S. Nayar. A search engine for large collections of images with faces. In ECCV, 2008.
[2] B. Siddiquie, R. Feris, and L. Davis. Image ranking and retrieval based on multi-attribute queries. In CVPR, 2011.
[3] L. Torresani, M. Szummer, and A. Fitzgibbon. Efficient object category recognition using classemes. In ECCV, 2010.
[4] F. Yu, R. Ji, M.-H. Tsai, G. Ye, and S.-F. Chang. Experiments of image retrieval using weak attributes. Columbia University Computer Science Department Technical Report # CUCS 005-12, 2012.