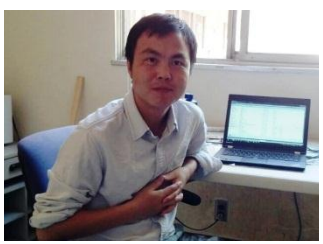
Discriminative and Generative Learning for Object Discovery

Vision And Learning SEminar (VALSE)

Xinggang Wang (王兴刚)
Huazhong University of Science and Technology

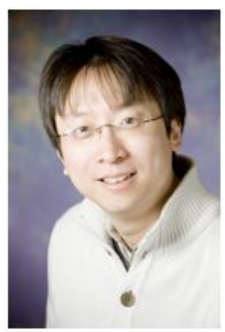
People

1. Xinggang Wang, Baoyuan Wang, Xiang Bai, Wenyu Liu, and Zhuowen Tu, Max-Margin Multiple Instance Dictionary Learning, International Conference on Machine Learning (ICML), Atlanta, June, 2013.
2. Xinggang Wang, Zhengdong Zhang, Yi Ma, Xiang Bai, Wenyu Liu, and Zhuowen Tu, Robust Subspace Discovery via Relaxed Rank Minimization, Neural Computation, Vol. 26, No. 3 April 2014, Pages 611-635.



Xinggang Wang

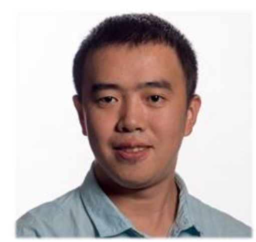
Zhuowen Tu



Yi Ma

Xiang Bai

Wenyu Liu



Zhengdong Zhang

Baoyuan Wang

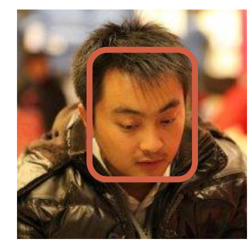
People

1. Xinggang Wang, Baoyuan Wang, Xiang Bai, Wenyu Liu, and Zhuowen Tu, Max-Margin Multiple Instance Dictionary Learning, International Conference on Machine Learning (ICML), Atlanta, June, 2013.
2. Xinggang Wang, Zhengdong Zhang, Yi Ma, Xiang Bai, Wenyu Liu, and Zhuowen Tu, Robust Subspace Discovery via Relaxed Rank Minimization, Neural Computation, Vol. 26, No. 3 April 2014, Pages 611-635.





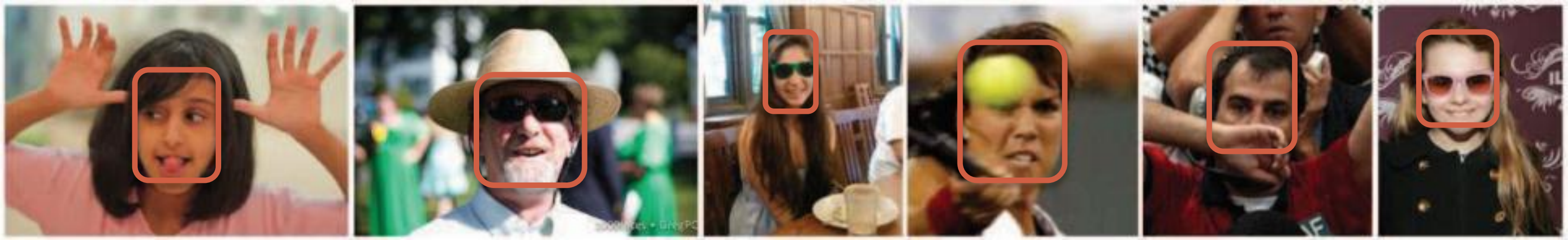




Xinggang Wang | Zhuowen Tu | Yi Ma | Xiang Bai | Wenyu Liu | Zhengdong Zhang | Baoyuan Wang

Introduction of object discovery

Introduction of object discovery



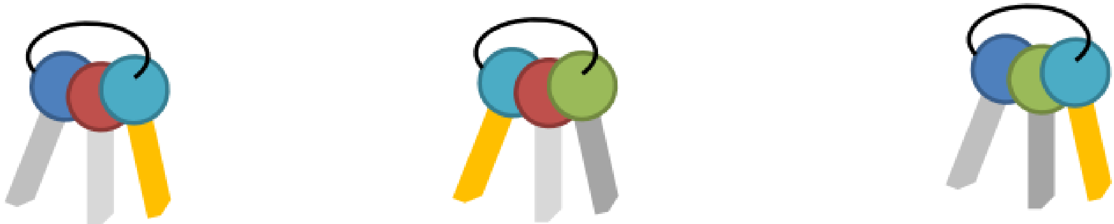
- Automatically discovery the objects from the images contain the same object.

- Discriminative approach: learning discriminative functions to classify objects and background.
- Generative approach: assume the objects lie on a low-rank subspace.

- Applications
  - Object detection in a weakly supervised way
  - Learning mid-level image representation
  - Codebook learning

Multi-Instance Learning (MIL)

刘老师的钥匙链 白老师的钥匙链 冯老师的钥匙链



刘老师**能**进入房间 白老师**能**进入房间 冯老师**不能**进入房间

正包 正包 负包

- MIL Research: mi-SVM (Andrews et al 2002), mi-Graph (Zhou et al, 2009)
- MIL for Image Object Detection: MCL (Tu et al 2008)
- MIL for Object Discovery: bMCL (Tu et al 2010)

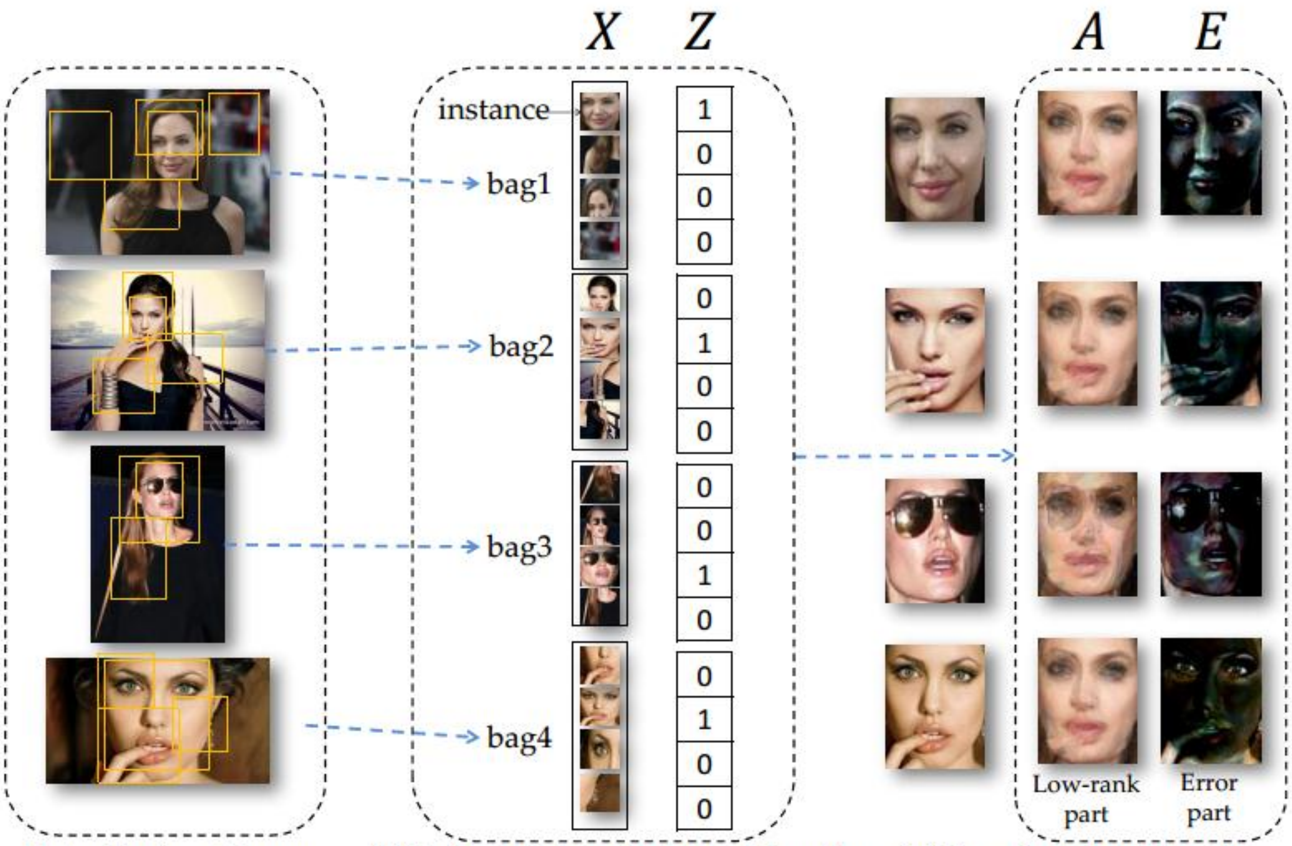
*Figure Credit: B. Babenko

Subspace learning via low rank optimization*

- Motivation
  - Modeling the objects in the same class with a low-dim linear subspace
  - Deformation and occlusion on objects are regarded as large sparse errors
  - Object locations and appearance models are learned via low rank optimization

* **Xinggang Wang**, Zhengdong Zhang, Yi Ma, Xiang Bai, Wenyu Liu, and Zhuowen Tu. “Robust Subspace Discovery via Relaxed Rank Minimization”, Neural Computation, 2013 (in press)
* Xinggang Wang, Zhengdong Zhang, Yi Ma, Xiang Bai, Wenyu Liu, and Zhuowen Tu. “OneClass Multiple Instance Learning via Robust PCA for Common Object Discovery”. ACCV, 2012

Subspace learning via low rank optimization



(a) A collection of images containing common objects with detected salient image patches.

(b) Each patch is an instance. All patches from the same image form a bag. Each patch is associated with an indicator.

(c) The selected patches, the subspace, and sparse errors.

Notations

- In a MIL framework

Instances in a bag (given): $X^{(k)} = \left[x_1^{(k)}, \ldots, x_{n_k}^{(k)}\right] \in \mathbf{R}^{d \times n_k}$

Instance labels (unknown): $Z^{(k)} = \left[z_1^{(k)}, \ldots, z_{n_k}^{(k)}\right] \in \{0,1\}^{n_k}$

All bags are positive: $\bigvee_{i=1}^{n_k} z_i^{(k)} = 1, \forall k \in [K]$

$$X = [X^{(1)}, \ldots, X^{(K)}] \in \mathbf{R}^{d \times N}$$
$$Z = [Z^{(1)}, \ldots, Z^{(K)}]$$

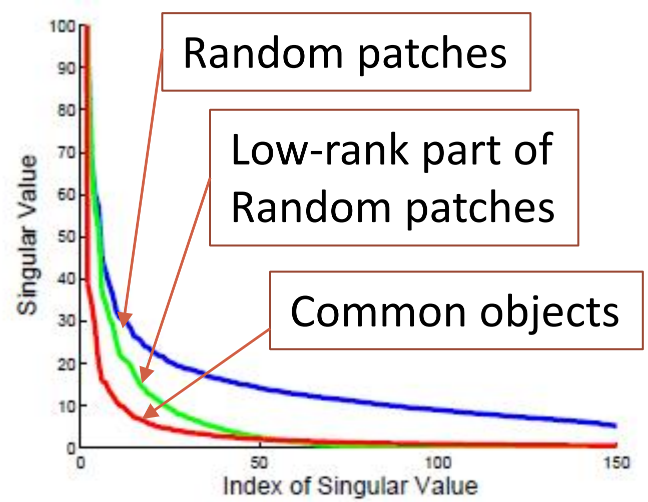
Assumptions

Different instances of the same object are highly correlated. It is reasonable to assume that such instances lie on a low-dimensional subspace $\Omega \in \boldsymbol{R}^d$.

A bag with face image and random patches

Low-rank part

Error part



Subspace discovery via low-rank optimization

- The objective function

$$
\begin{aligned}
&\min_{A,E,Z} \operatorname{rank}(A) + \gamma \|E\|_0 \\
\text{s.t.} \quad & X \operatorname{diag}(Z) = A + E, \ \forall k \in [K] \ \bigvee_{i=1}^{n_k} z_i^k = 1
\end{aligned}
$$

- A is the low-rank part
- E is sparse error,
- $diag(Z)$ is $N \times N$ block-diagonal matrix with K blocks $\{diag(Z^{(k)})\}$

Subspace discovery via low-rank optimization

- Convex surrogates of A and E:

$$\begin{aligned} &\min_{A,E,Z} \operatorname{rank}(A) + \gamma \|E\|_0 \\ \text{s.t.} \quad & X\operatorname{diag}(Z) = A + E,\ \forall k \in [K]\ \bigvee_{i=1}^{n_k} z_i^k = 1 \end{aligned}$$

$$\Downarrow$$

$$\begin{aligned} &\min_{A,E,Z} \|A\|_* + \lambda \|E\|_1 \\ \text{s.t.} \quad & X\operatorname{diag}(Z) = A + E,\ \forall k \in [K]\ \bigvee_{i=1}^{n_k} z_i^{(k)} = 1 \end{aligned}$$

Subspace discovery via low-rank optimization

- A Naive Iterative Solution (NIM):

- Given A and E, estimate Z according to reconstruction error of instance

$$e_i^{(k)} = \min_w \|Aw - x_i^{(k)}\|_1$$

- Given Z, optimize A and E using RPCA method [1]

[1] E. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3), May 2011.

Subspace discovery via low-rank optimization

- Relaxing Z

$$\min_{A,E,Z} \|A\|_* + \lambda \|E\|_1$$
$$\text{s.t.} \quad X\mathrm{diag}(Z) = A + E,\ \forall k \in [K]\ \bigvee_{i=1}^{n_k} z_i^{(k)} = 1$$
$$\Downarrow$$
$$\min_{A,E,Z} \|A\|_* + \lambda \|E\|_1,$$
$$\text{s.t.} \quad X\mathrm{diag}(Z) = A + E,\quad \forall k \in [K], \mathbf{1}^T Z^{(k)} = 1$$

***Theorem 1:* If none of the columns of X is zero, the optimal solution Z^* is always non-negative.**

Subspace discovery via low-rank optimization

- Solution via Inexact ALM (IALM), also called as Alternating Direction Method of Multipliers (ADMM)

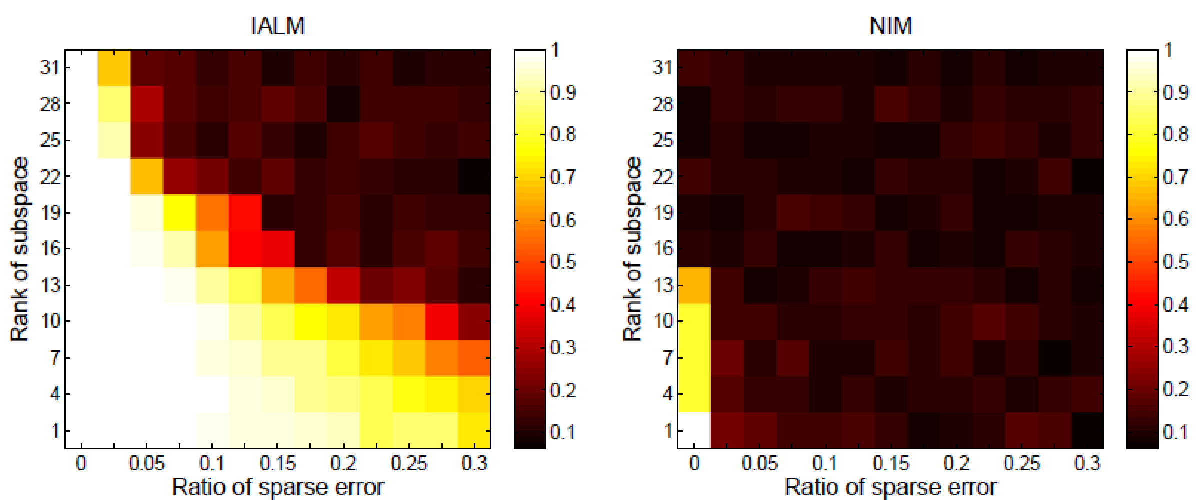
The Augmented Lagrangian function:

$$
\begin{aligned}
&L(A, E, Z, Y_0, Y_1, \ldots, Y_K) \doteq \|A\|_* + \lambda \|E\|_1 \\
&+ \langle Y_0, X\operatorname{diag}(Z) - A - E \rangle + \frac{\mu}{2} \|X\operatorname{diag}(Z) - A - E\|_F^2 \\
&+ \sum_{k=1}^{K} \left(\langle Y_k, \mathbf{1}^T Z^{(k)} - 1 \rangle + \frac{\mu}{2} \|\mathbf{1}^T Z^{(k)} - 1\|_F^2 \right).
\end{aligned}
$$

Simulations

- Robust subspace learning simulation

  In this experiment, we generate synthetic data with 50 bags; in each bag there are 10 instances which include 1 positive instance and 9 negative instances; dimension of instance is d = 500.



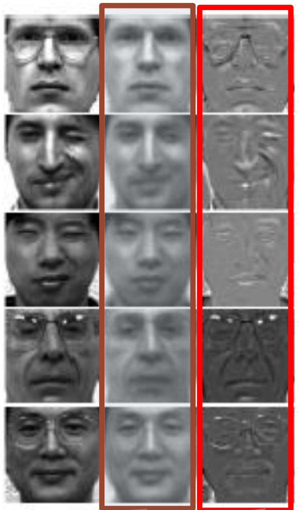
Working range of IALM and NIM: Precision of the recovered indicators when the sparsity level and the rank vary for both IALM (left) and NIM (right).

Simulations

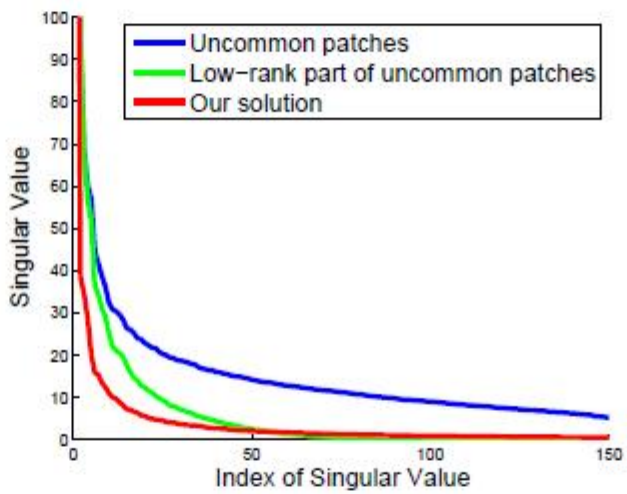
- Aligned face discovery among random image patches

The 165 face from Yale dataset images are in 165 bags; other than the face image, in each bag, there are 9 image patches from PASCAL dataset

A bag with face image and random patches



Low-rank part Error part



Accuracies of IALM and NIM (randomly initialized) are 99.5 ±0.5% and 77.8 ±3.5%

Experiments

- Object discovery on FDDB subset

It contains 440 face images from FDDB dataset [1]

METHOD	Average precision
Saliency detection (SD)	0.148
bMCL [2]	0.619
NIM-SD	0.671
NIM-RAND	0.669
IALM	**0.745**

[1] V. Jain and E. Learned-Miller. Fddb: A benchmark for face detection in nconstrained settings. Technical Report UM-CS-2010-009, University of Massachusetts, Amherst, 2010.
[2] J. Zhu, J. Wu, Y. Wei, E. Chang, and Z. Tu. Unsupervised object class discovery via saliency-guided multiple class learning. In IEEE Conference on Computer Vision and Pattern Recognition, 2012.

Experiments

- Object discovery on FDDB subset

Face discovery results. The windows with maximal score given by SD (Feng et al., 2011) (in cyan), bMCL (Zhu et al., 2012) (in green), NIM-SD (in blue) and IALM (in red) on FDDB subset (Jain & Learned-Miller, 2010).

Experiments

- PASCAL dataset

	PASCAL06-		PASCAL07-	
Method	6×2	all	6×2	all
Pandey & Lazebnik (2011)	N/A	N/A	**61**	**30**
Deselaers et al. (2012)	**64**	**49**	50	28
Chum & Zisserman (2007)	45	34	33	19
Russell et al. (2006)	28	27	22	14
ESS (Lampert et al., 2009)	24	21	27	14
IALM	57	43	40	27

Object discovery performance evaluated by CorLoc on PASCAL 2006 and 2007 data sets. We follow the setting in Deselaser et al. (2012).

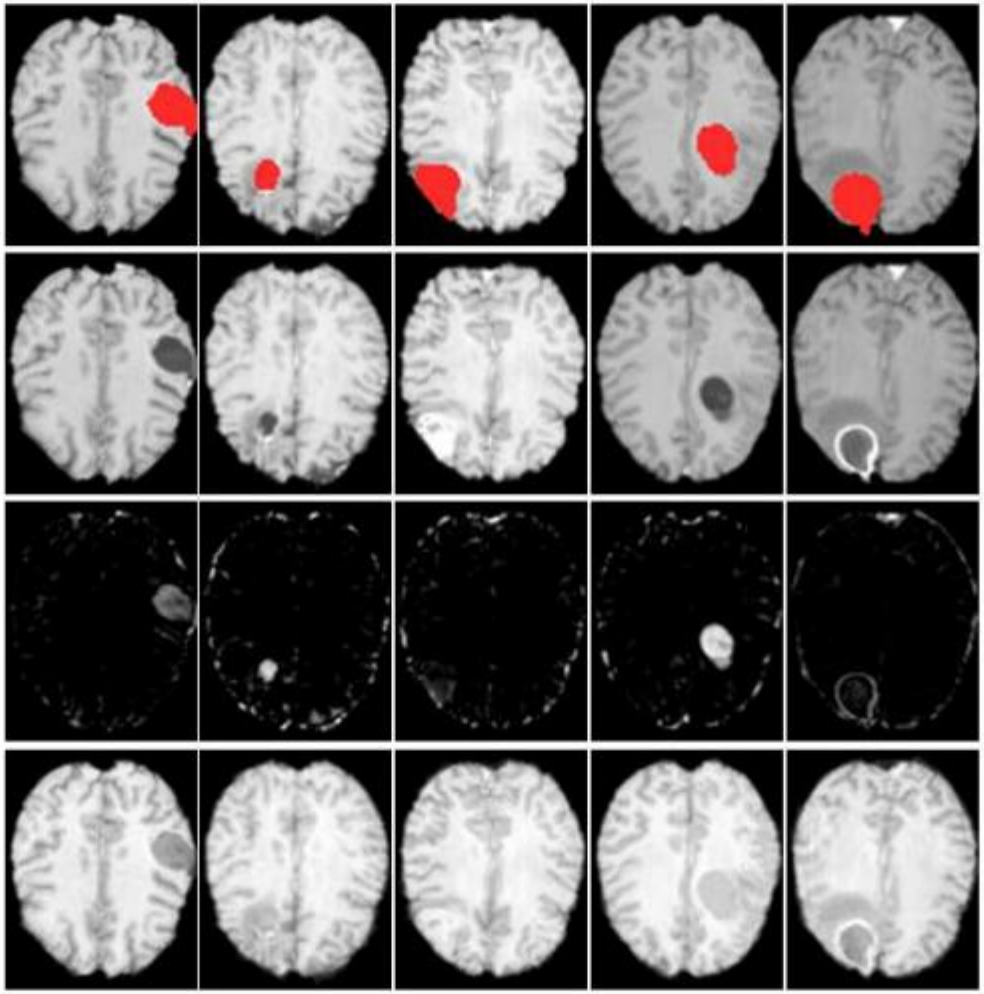
Experiments

- PASCAL dataset

Red rectangles: object discovery results of IALM (from top to bottom: aeroplane, bicycle, bus, motorbike, plotted-plants and tv-monitors) on the challenging PASCAL 2007. Green rectangles: annotated object ground-truth.

Experiments

- Tumor discovery



Experiments

- Multiple instance learning

Data sets	*Musk1*	*Musk2*	*Elephant*	*Fox*	*Tiger*	Average
MI-SVM	77.9	84.3	81.4	59.4	84.0	77.4
mi-SVM	87.4	83.6	82.0	58.2	78.9	78.0
MILES	86.3	87.7	-	-	-	-
EM-DD	84.8	84.9	78.3	56.1	72.1	75.2
PPMM Kernel	**95.6**	81.2	82.4	60.3	80.2	79.9
MIGraph	90.0±3.8	90.0±2.7	85.1±2.8	61.2±1.7	81.9±1.5	81.6
miGraph	88.9±3.3	**90.3±2.6**	**86.8±0.7**	61.6±2.8	**86.0±1.6**	**82.7**
MI-CRF	87.0	78.4	85.0	65.0	79.5	79.0
Our method	89.9±0.7	85.0±1.6	79.6±0.9	**65.4±1.2**	81.5±1.0	80.3

Conclusions

- It is robust to sparse error and overwhelming outliers.
- It has a convex solution and insensitive to initialization of the algorithm.
- We use the IALM (ADMM) algorithm to solve a combinatorial problem.
- In real applications, it is effective, but fails to get the state-of-the-art performance.

Dictionary learning Literature

It is an important component for building effective and efficient representation

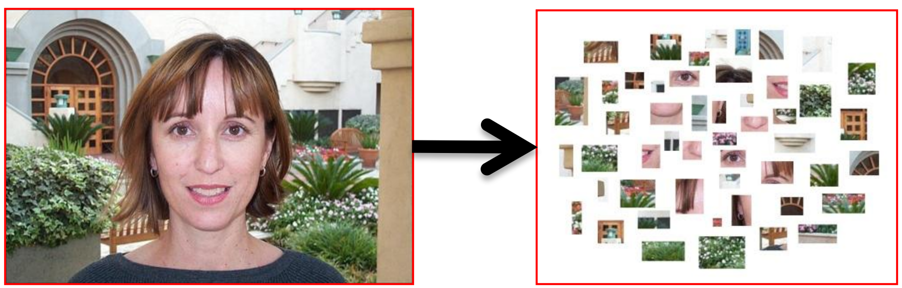
- Sparse coding (Olshausen and Field, 1996)
- Latent Dirichlet Allocation (Blei et al., 2003)
- Bag of words (Blei et al., 2003)
- Deep Belief Nets (Hinton et al., 2006)

It is effective for many machine learning problems

1. Explicit representations are often enforced;
2. dimensionality reduction is performed through quantization;
3. it facilitates hierarchical representations;
4. spatial configuration can be also imposed.

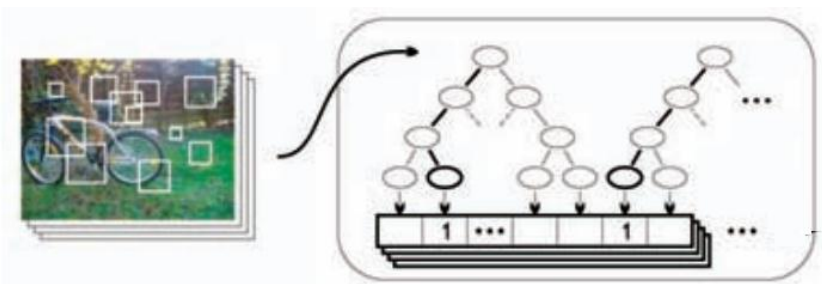
Unsupervised and supervised dictionary learning

- Unsupervised codebook learning: kmeans (Duda et al., 2000).
- Supervised codebook learning: ERC-Forests (Moosmann et al., 2008)



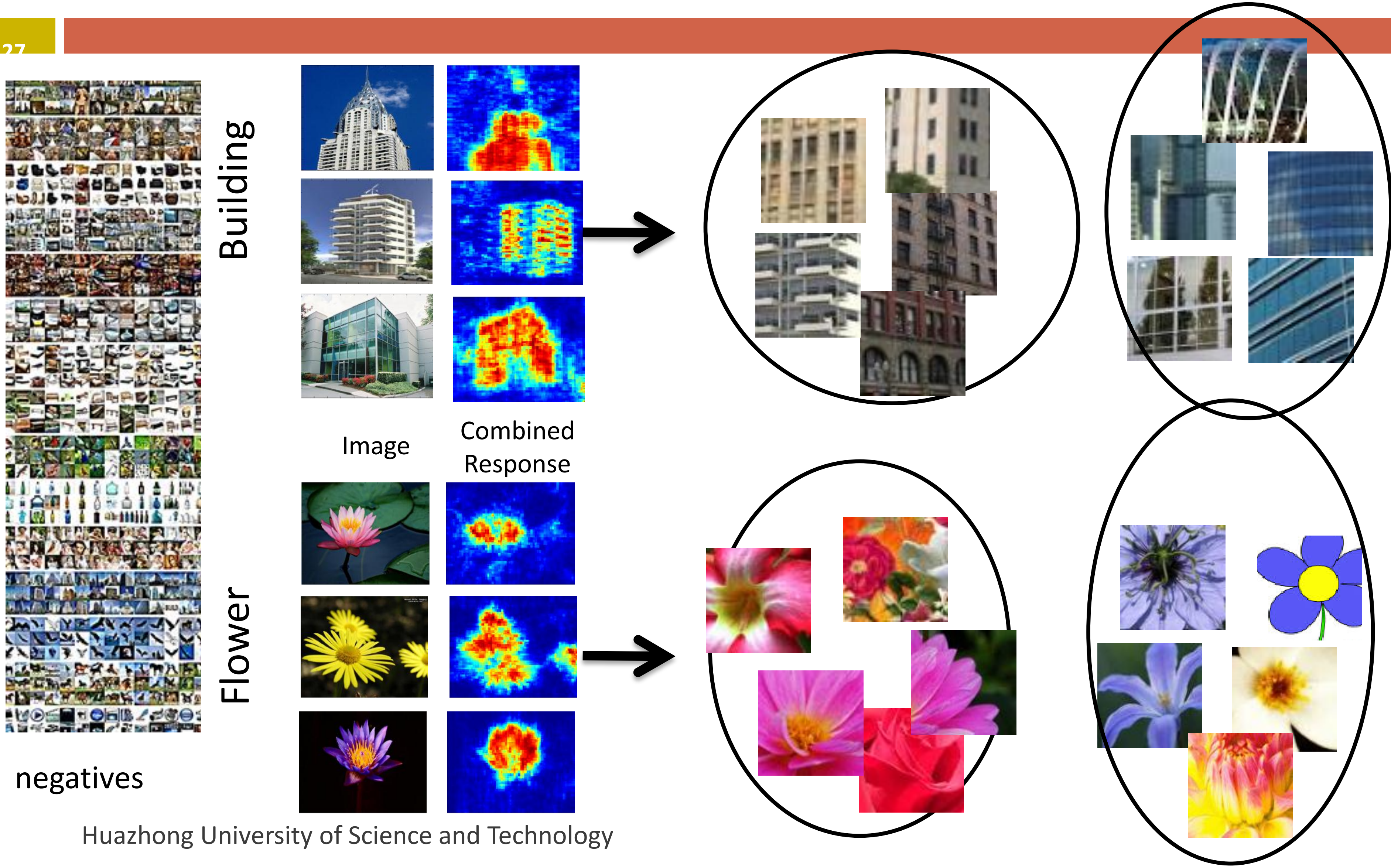
face, flowers, building

Kmeans for dictionary learning



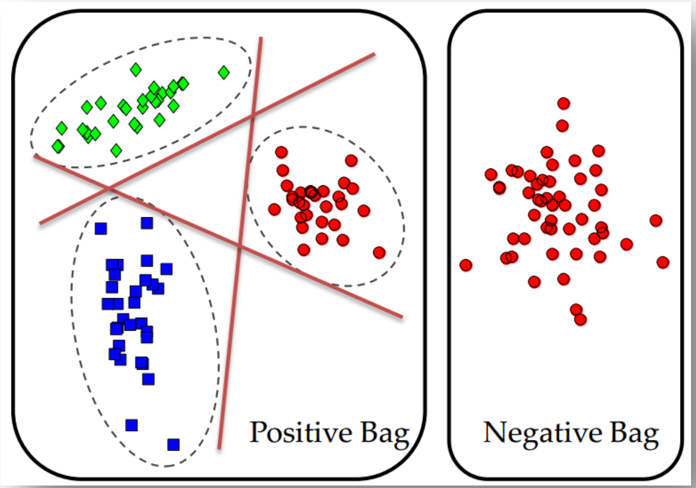
ERC-forest for dictionary learning

Weakly-supervised dictionary learning



Max-margin Multiple-instance dictionary learning (MMDL) *

- Exploring sematic information using multiple-instance learning (MIL) (Dietterich et al.,1997).
- Assuming positive instances are drawn from multiple clusters.
- Directly maximizing margins between different clusters.
- Using the cluster classifiers as the codebook.



*Xinggang Wang, Baoyuan Wang, Xiang Bai, Wenyu Liu, and Zhuowen Tu. "Max-Margin Multiple Instance Dictionary Learning", ICML, 2013

Formulation

- MIL notations:

$$bag: X_i = \{x_{i1}, \ldots x_{im}\} \qquad instance: x_{ij} \in R^{d*1}$$

$$label: Y_i \in \{0,1\} \qquad y_{ij} \in \{0,1\}$$

$$Y_i = 0, \ \ y_{ij} = 0 \qquad Y_i = 1, \ \ \forall y_{ij} = 1$$

- Generalized code (G-code):

$$linear\ classifier: f(x) = w^T x$$

$$codebook: W = [w_0, w_1, \ldots, w_K] \ \ w_k \in R^{d*1}$$

- Latent variable (indicator) for instance:

$$z_{ij} = argmax_k w_k^T x_{ij}$$

Formulation

- The objective function of MMDL:

$$\min_{W, z_{ij}} \underbrace{\sum_{k=0}^{K} \|w_k\|^2}_{regularization\ term} + \underbrace{\lambda \sum_{ij} \max(0, 1 + w_{r_{ij}}^{k} x_{ij} - w_{z_{ij}}^{k} x_{ij})}_{loss\ function}$$

Crammer & Singer SVM

$$s.t. \quad if\ Y_i = 1, \sum_j z_{ij} > 0$$

$$if\ Y_i = 0, z_{ij} = 0$$

$$where \quad r_{ij} = argmax_{k \in \{0,\ldots,K\}, k \neq z_{ij}} w_k^T x_{ij}$$

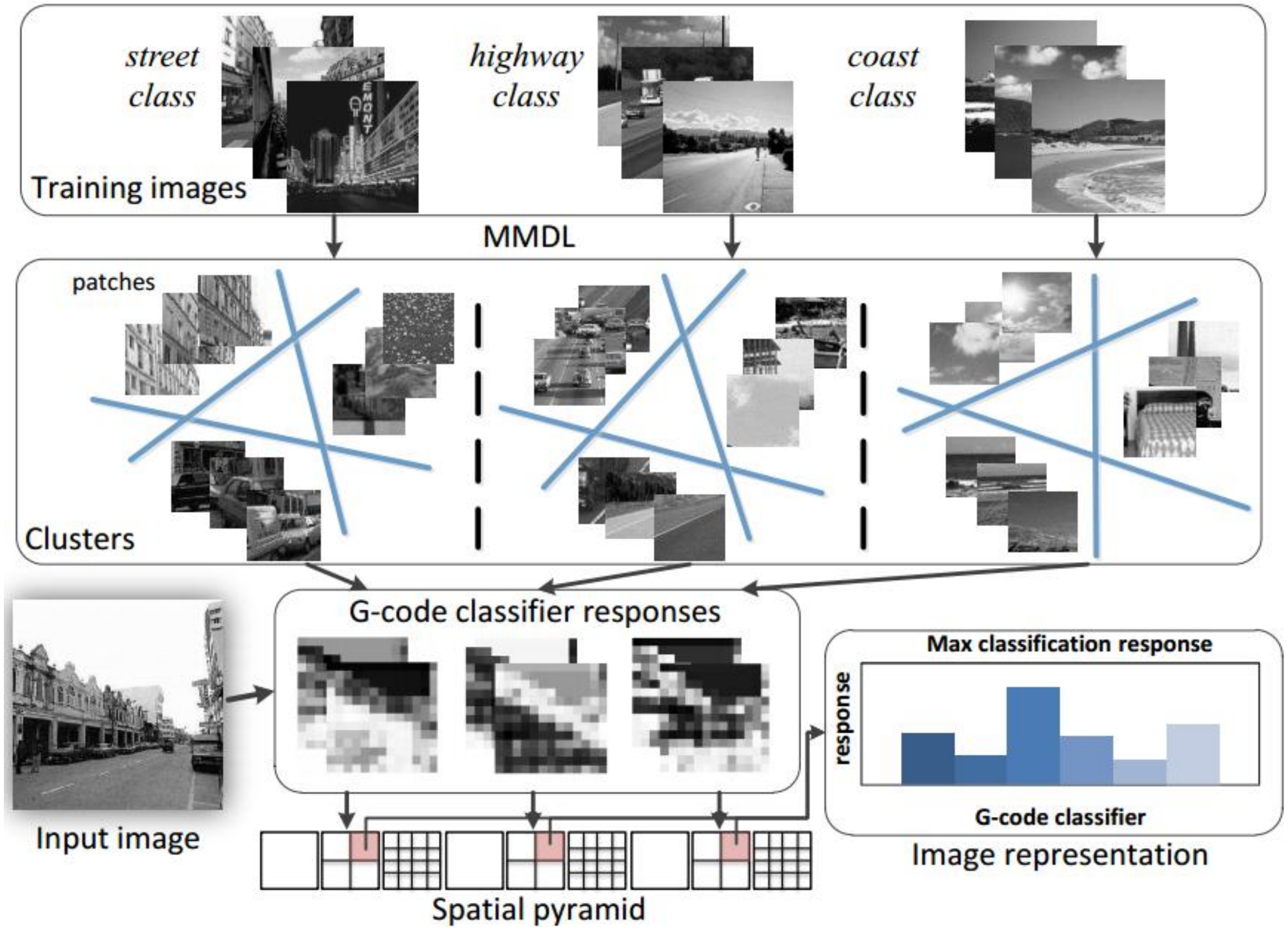
Learning Strategies

- **Initialization: k-means**
- **Iteration:**
  - Optimize W
    - Sample the instances in positive bags acceding to their "positiveness":

$$p_{ij} = sigmoid(max_{k \in \{1,\ldots,K\}}(w_k^T x_{ij} - w_0^T x_{ij})/\sigma)$$

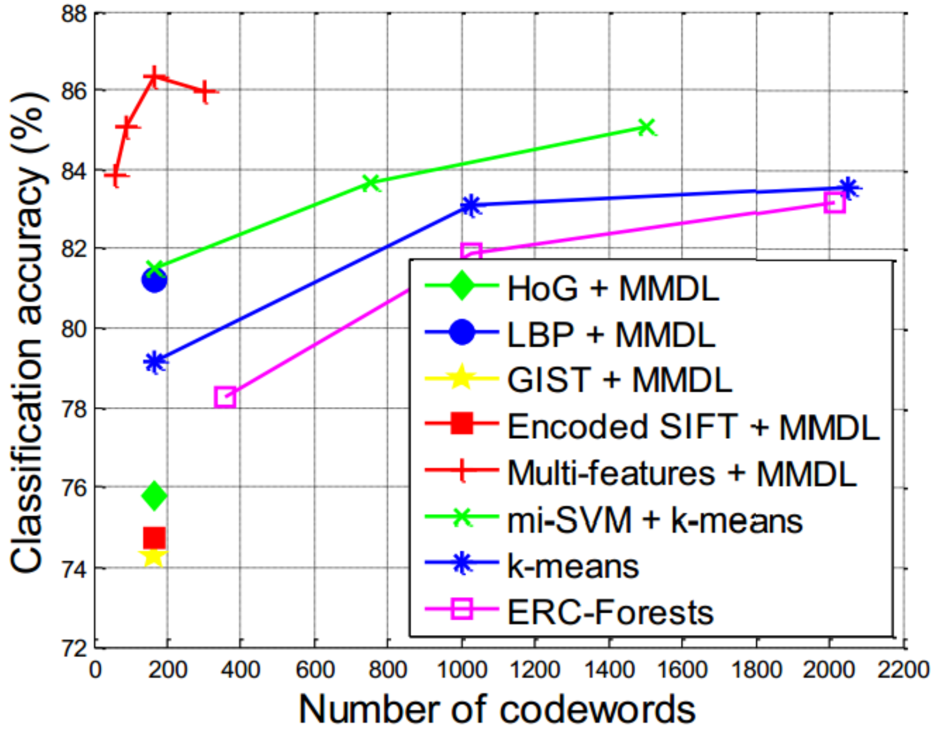
    - Train multi-class SVM
  - Optimize Z
    - Update the "positiveness" using learnt classifiers
    - Update Z by assigning each data to its "nearest" cluster

Image representation



Experiments

Features: LBP, HoG, encoded SIFT, LAB histogram, GIST.



Average classification accuracies of different methods comparison on 15 Scene over different number of codewords and different types of feature.

Experiments

- 15 Scene

Methods	Accuracy (%)	#(codewords)
Object Bank (Li et al., 2010)	80.90	2400
Lazebnik et al. 2006	81.10	200
Yang et al. 2009	80.40	1024
Kernel Desp. (Bo et al. 2010)	86.70	1000
Ours	86.35	165

- UIUC Sports dataset

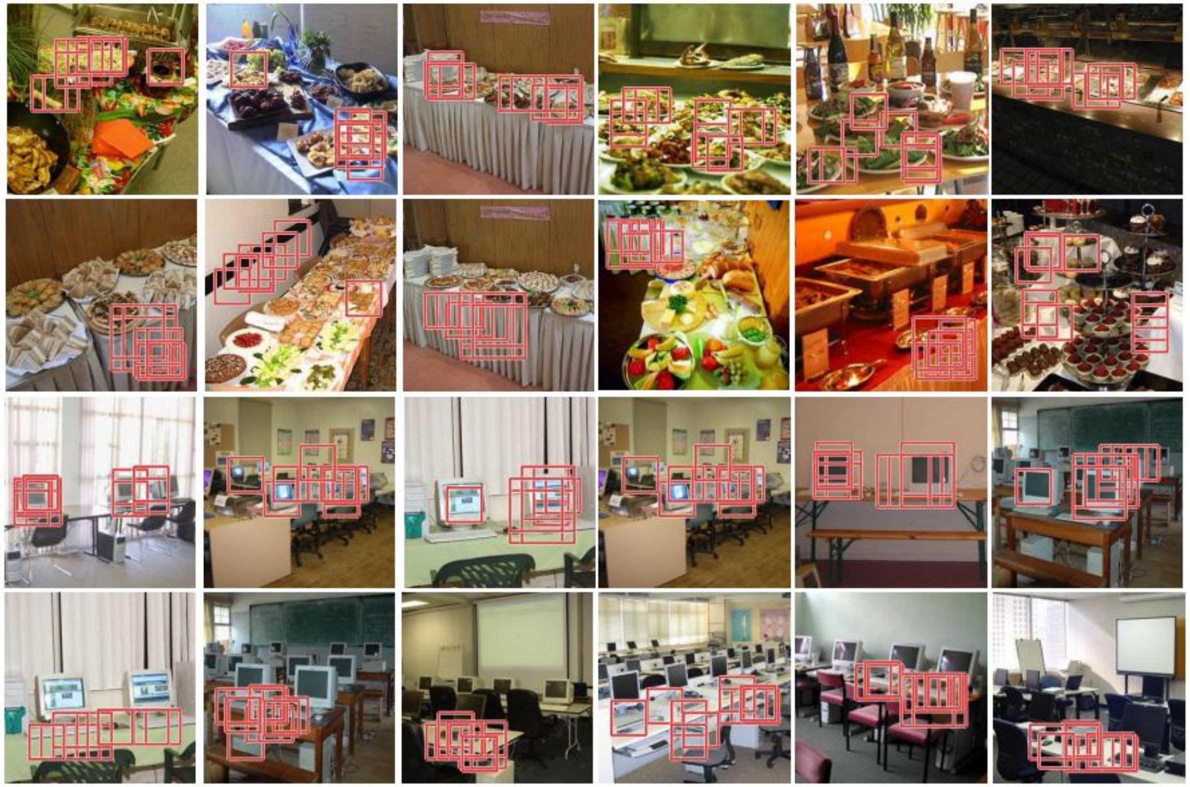
Methods	Accuracy (%)
Li & Fei-Fei, 2007	73.4
Wu& Rehg, 2009	84.3
Object Bank (Li et al., 2010)	76.3
SPMSM (Kwitt et al., 2012)	83.0
LPR (Sadeghi & Tappen, 2012)	86.25
Ours (88 codes)	88.47

- MIT 67 Indoor

Methods	Accuracy (%)
ROI+GIST (Quattoni & Torralba, 2009)	26.5
RBOW (Parizi et al., 2012)	37.93
Disc. Patches (Sigh et al., 2012)	38.1
SPMSM (Kwitt et al., 2012)	44.0
LPR (Sadeghi & Tappen, 2012)	44.84
Ours (737 codes)	50.15

Experiments

- MIT 67 Indoor



Some meaningful clusters learned by MMDL for different categories. Each row illustrates a cluster model: red rectangles shows positions of G-code classier red where SVM function value is bigger than zero.

Conclusions

- MMDL can naturally learn a metric to take the advantage of multiple features.
- The max-margin formulation leads to very compact code for image representation with the state-of-the-art image classification performance.
- The MIL strategy can learn codebook contains rich semantic information.

Thank you!

Q&A