VALSE 2016 武汉
视觉与学习青年学者研讨会

Tutorial Talk

Multi-Instance Learning and Its Applications in Computer Vision

Min-Ling Zhang (张敏灵)
PALM Lab, School of CSE,
Southeast University
Email: zhangml@seu.edu.cn

Xinggang Wang (王兴刚)
MC Lab, School of EIC,
Huazhong Univ. of Sci. & Tech.
Email: xgwang@hust.edu.cn

April 22, Wuhan

Outline

- Multi-Instance Learning (MIL)
- Representative Algorithms
- Related Topics
- Applications in Computer Vision
- Conclusion

Outline

- Multi-Instance Learning (MIL)
- Representative Algorithms
- Related Topics
- Applications in Computer Vision
- Conclusion

Origin of MIL

Originated from the research on *drug activity prediction*

Seminal paper [Dietterich et al., AIJ'97]

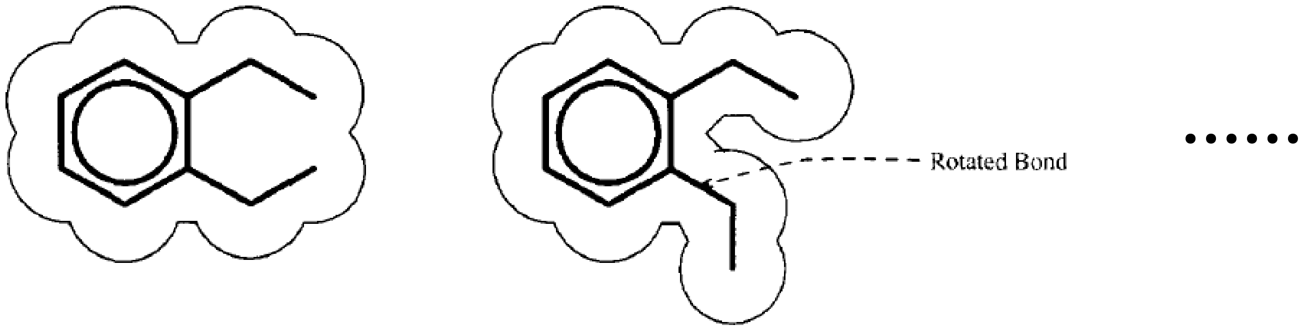
Dietterich T G, Lathrop R H, Lozano-Pérez T. Solving the multiple instance learning problem with axis-parallel rectangles. **Artificial Intelligence**, 1997, 89(1-2): 31-71.

Problem Scenario

- Drugs are small molecules working by binding to the target area
- For molecules qualified to make the drug, one of its shapes could tightly bind to the target area

Origin of MIL (Cont.)

A molecule may have many alternative shapes



Reprinted from [Dietterich et al., AIJ'97]

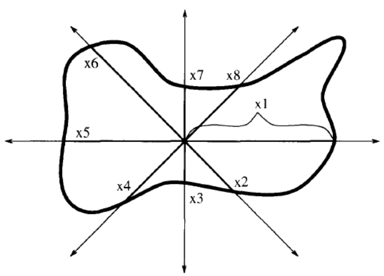
The Difficulty

Biochemists know that whether a molecule is qualified or not, but do not know which shape responses for the qualification

Origin of MIL (Cont.)

shape ⟺ instance

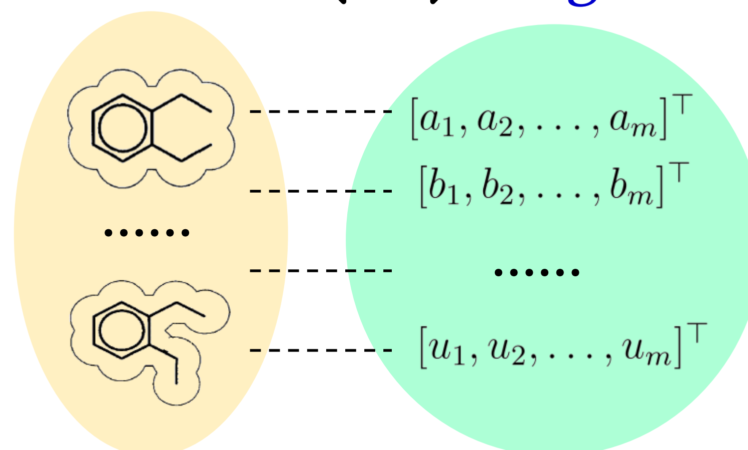
(feature vector)



$[x_1, x_2, \ldots, x_m]^\top$

Reprinted from
[Dietterich et al., AIJ'97]

molecule ⟺ bag



A molecule corresponds to a **bag** of instances

- A bag is positive if it contains at least one positive instance; otherwise it is negative
- The labels of the training bags are known
- The labels of the instances in the training bags are unknown

Formal Definition of MIL

Settings $\mathcal{X}$: feature space $\mathbb{R}^d$; $\mathcal{Y}$: label space $\{-1,+1\}$

Inputs multi-instance training set $\mathcal{D} = \{(X_i, y_i) \mid 1 \leq i \leq N\}$

$X_i = \{\boldsymbol{x}_{i1} \ldots, \boldsymbol{x}_{ij}, \ldots, \boldsymbol{x}_{i,n_i}\} \subseteq \mathcal{X}$ is a **bag** of instances

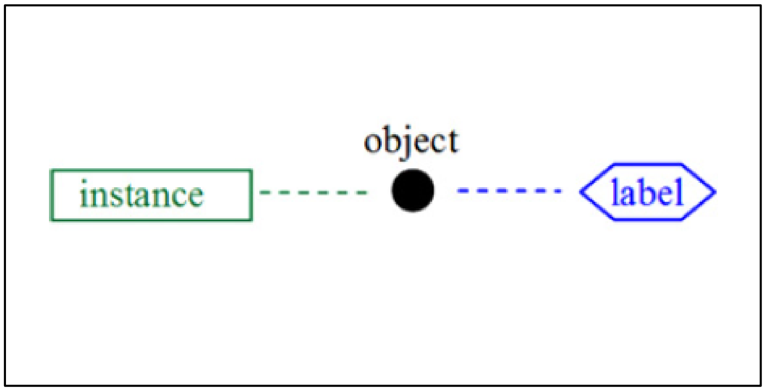
- n_i is the number of instances in X_i
- $\boldsymbol{x}_{ij} \in \mathcal{X}$ is an instance $[x_{ij1}, \ldots, x_{ijl}, \ldots, x_{ijd}]^\top$

$y_i \in \mathcal{Y}$ is the label of X_i

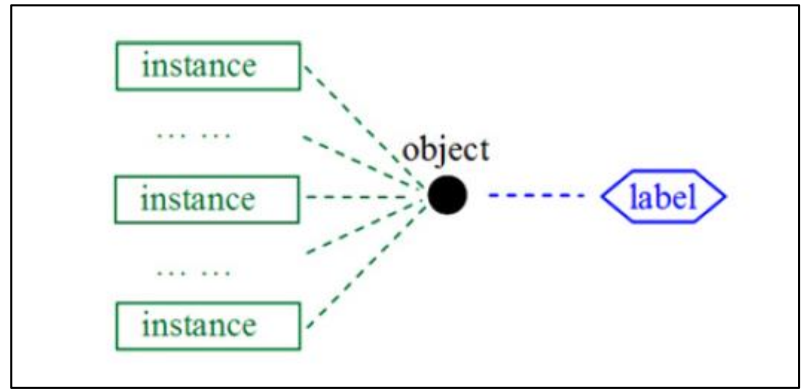
X_i is a positive bag (thus $y_i = +1$) if there exists $g \in \{1, \ldots, n_i\}$, $\boldsymbol{x}_{ig}$ is positive. Yet the value of index g is unknown

Outputs multi-instance predictive function: $f : 2^{\mathcal{X}} \mapsto \mathcal{Y}$

Comparison between MIL and Traditional Supervised Learning



Traditional (single-instance) learning

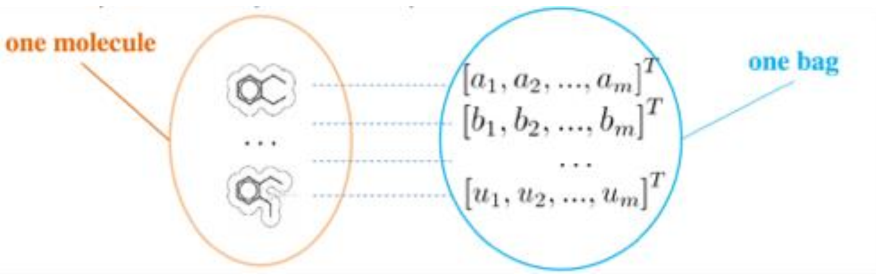


Multi-instance learning

Reprinted from [Zhou et al., AIJ'12]

The Power of MIL

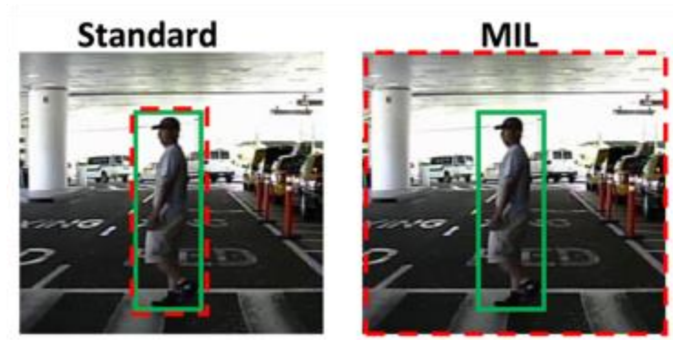
In many real-world applications, **bag representation** is a natural choice



Drug activity prediction

Image categorization

Text categorization



Pedestrian detection

Outline

- Multi-Instance Learning (MIL)
- Representative Algorithms
- Related Topics
- Applications in Computer Vision
- Conclusion

A Brief Categorization [Amores, AIJ'13]

Instance-space (IS) paradigm

Infer instance-level classifier from the training data ➔ Perform the learning process by aggregating instance-level responses

- APR [Dietterich et al., AIJ'97]
- Diverse Density [Maron & Lozano-Pérez, NIPS'97]
- MI-SVM [Andrews et al., NIPS'02]
- ……

Bag-space (BS) paradigm

Define distance function (or kernel function) over bags ➔ Perform the learning process by treating bags as a whole

- Citation-*k*NN [Wang & Zucker, ICML'00]
- MI-Kernel [Gärtner et al., ICML'02]
- MIGraph [Zhou et al., ICML'09]
- ……

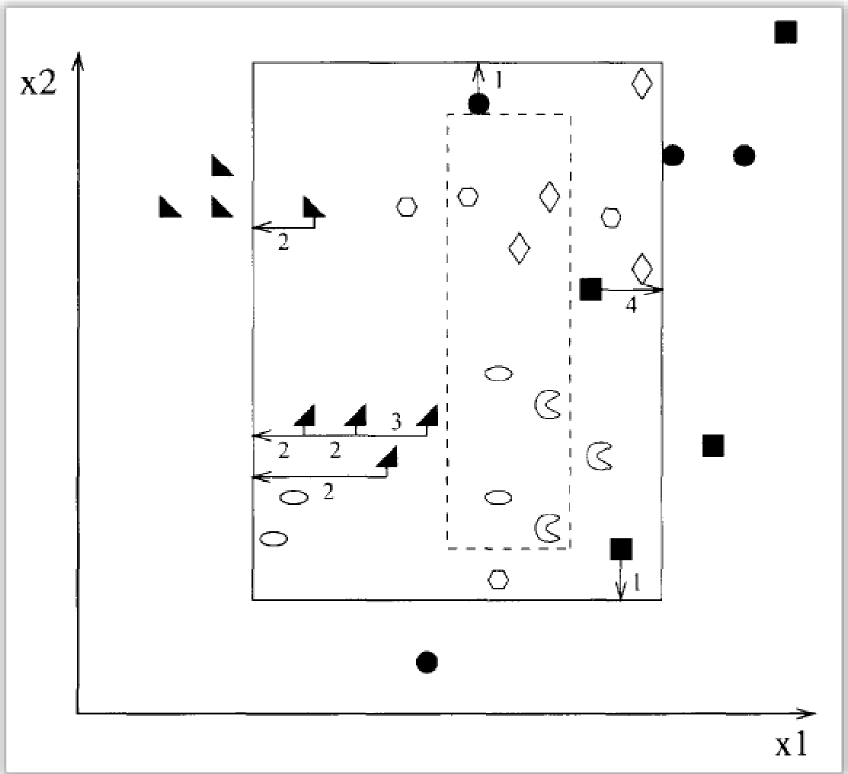
Embedded-space (ES) paradigm

Embed (map) bags into vocabulary-induced feature space➔ Perform the learning process in the embedded feature space

- MILES [Chen et al., TPAMI'06]
- CCE [Zhou & Zhang, KAIS'07]
- ……

APR [Dietterich et al., AIJ'97]

APR (Axis-Parallel Rectangles) Algorithms



Instance-level classifier

An APR which encloses at least one instance from each positive bag and no instance of any negative bag

Three APR algorithms

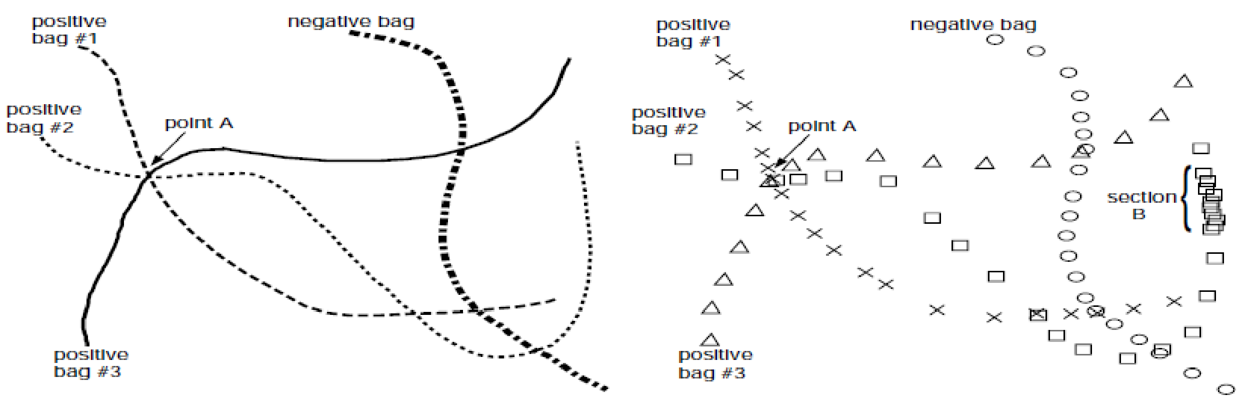
standard, *outside-in*, and *inside-out* APR algorithms

best performance: Iterated-discrim APR

on *Musk1*: 92.4%; on *Musk2*: 89.2%

Specific to the *drug activity prediction* problem

Diverse Density [Maron & Lozano-Pérez, NIPS'97]



(a) The different shapes that a molecule can take on are represented as a path. The intersection point of positive paths is where they took on the same shape.

(b) Samples taken along the paths. Section B is a high density area, but point A is a high Diverse Density area.

Figure 1: A motivating example for Diverse Density

Instance-level classifier

Target concept (point $\boldsymbol{t} \in \mathbb{R}^d$ in the feature space) which maximizes the diverse density (DD) over the training bags, i.e. being *near* to instances of different positive bags and *far* from instances of all negative bags

Diverse Density (Cont.)

Probabilistic formulation of diverse density

$$DD(\boldsymbol{t}) = \Pr(\boldsymbol{t} \mid X_1, \ldots, X_N)$$

$$\propto \Pr(X_1, \ldots, X_N \mid \boldsymbol{t})$$ [Assuming equal prior probability]

$$= \prod_{p:\, y_p=+1} \Pr(X_p \mid \boldsymbol{t}) \cdot \prod_{q:\, y_q=-1} \Pr(X_q \mid \boldsymbol{t})$$ [Assuming conditional independence among bags]

$$\propto \prod_{p:\, y_p=+1} \Pr(\boldsymbol{t} \mid X_p) \cdot \prod_{q:\, y_q=-1} \Pr(\boldsymbol{t} \mid X_q)$$ [Assuming equal prior probability]

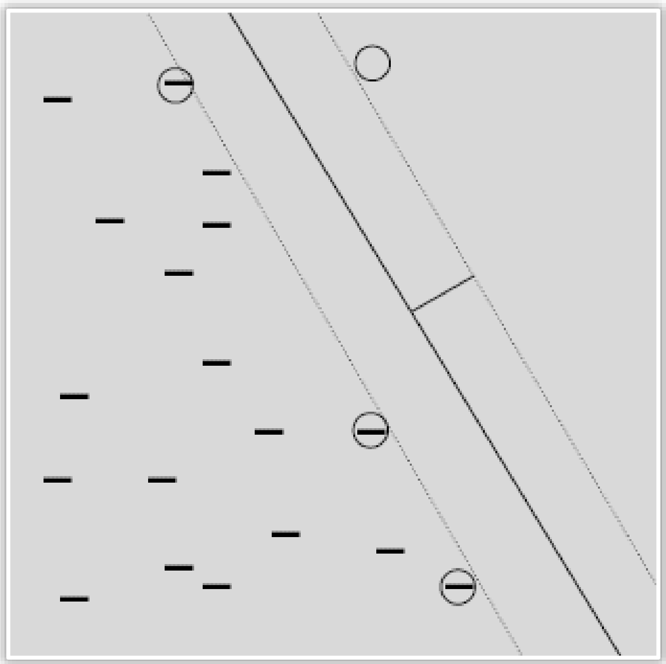
For positive bag: $\Pr(\boldsymbol{t} \mid X_p) = 1 - \prod_{j=1}^{n_p} (1 - \Pr(\boldsymbol{t} \mid \boldsymbol{x}_{pj}))$

For negative bag: $\Pr(\boldsymbol{t} \mid X_q) = \prod_{j=1}^{n_q} (1 - \Pr(\boldsymbol{t} \mid \boldsymbol{x}_{qj}))$

$$\Pr(\boldsymbol{t} \mid \boldsymbol{x}) = \exp(-||\boldsymbol{x} - \boldsymbol{t}||^2)$$

Maximize DD by gradient search

MI-SVM [Andrews et al., NIPS'02]



Instance-level classifier

A hyperplane $(\boldsymbol{w}, b)$ in $\mathbb{R}^d$ with $\boldsymbol{w} \in \mathbb{R}^d$ and $b \in \mathbb{R}$

Margin over MI bag (X_i, y_i)

$$\gamma_i = y_i \max_{1 \le j \le n_i} (\langle \boldsymbol{w}, \boldsymbol{x}_{ij} \rangle + b)$$

defined by the "most positive" ("least negative") instance

Maximum margin formulation

$$\min_{\boldsymbol{w}, b, \boldsymbol{\xi}} \ \frac{1}{2} ||\boldsymbol{w}||^2 + C \sum\nolimits_{i=1}^{N} \xi_i$$

$$\text{s.t.} \quad \forall\, i: \ y_i \max_{1 \le j \le n_i} (\langle \boldsymbol{w}, \boldsymbol{x}_{ij} \rangle + b) \ge 1 - \xi_i, \quad \xi_i \ge 0$$

MI-SVM (Cont.)

Heuristic EM-style optimization

E-step Identify *most positive* instance for each positive bag

$$\forall (X_i, y_i) \text{ with } y_i = +1:$$

$$g_i = \arg\max_{1 \le j \le n_i} \left(\langle \boldsymbol{w}, \boldsymbol{x}_{ij} \rangle + b \right)$$

M-step Update hyperplane based on identified instances (QP)

$$\min_{\boldsymbol{w}, b, \boldsymbol{\xi}} \ \frac{1}{2} ||\boldsymbol{w}||^2 + C \sum_{i=1}^{N} \xi_i$$

$$\forall i \text{ with } y_i = +1: \quad \langle \boldsymbol{w}, \boldsymbol{x}_{ig_i} \rangle + b \ge 1 - \xi_i$$

$$\forall i \text{ with } y_i = -1: \quad \langle \boldsymbol{w}, \boldsymbol{x}_{ij} \rangle + b \ge 1 - \xi_i \ (1 \le j \le n_i)$$

$$\forall i: \quad \xi_i \ge 0$$

Citation-kNN [Wang & Zucker, ICML'00]

Distance function between two bags $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$ and $Z = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\}$

Hausdorff distance — determined by one matching pair

$$H(X, Z) = \max\{h(X, Z), h(Z, X)\} \quad \text{where} \quad h(X, Z) = \max_{\boldsymbol{x} \in X} \min_{\boldsymbol{z} \in Z} ||\boldsymbol{x} - \boldsymbol{z}||$$

sensitive to outlying point

$$\text{replace } h(X, Z) \text{ with} : h_k(X, Z) = \mathop{k\text{th}}_{\boldsymbol{x} \in X} \min_{\boldsymbol{z} \in Z} ||\boldsymbol{x} - \boldsymbol{z}||$$ (k-th ranked distance)

k=*m*

$$h(X, Z) = h_m(X, Z)$$

k=1

$$H(X, Z) = \min_{\boldsymbol{x} \in X} \min_{\boldsymbol{z} \in Z} ||\boldsymbol{x} - \boldsymbol{z}||$$

Chamfer distance — determined by multiple matching pairs

$$C(X, Z) = \frac{1}{|X|} \sum_{\boldsymbol{x} \in X} \min_{\boldsymbol{z} \in Z} ||\boldsymbol{x} - \boldsymbol{z}|| + \frac{1}{|Z|} \sum_{\boldsymbol{z} \in Z} \min_{\boldsymbol{x} \in X} ||\boldsymbol{x} - \boldsymbol{z}||$$

Citation-kNN (Cont.)

References and citers

r-references set for Z:

$$\mathcal{R}^r(Z) = \{X \mid X \text{ is among } Z'\text{s } r \text{ nearest neighbors}\}$$

c-citers set for Z:

$$\mathcal{C}^c(Z) = \{X \mid Z \text{ is among } X'\text{s } c \text{ nearest neighbors}\}$$

In general setting, $r \neq c$

Make prediction for Z by voting among the labeled bags in $\mathcal{R}^r(Z) \bigcup \mathcal{C}^c(Z)$

MI-Kernel [Gärtner et al., ICML'02]

Kernel function between two bags $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$ and $Z = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\}$

Set kernel

$$K(X, Z) = \sum_{\boldsymbol{x} \in X} \sum_{\boldsymbol{z} \in Z} \kappa(\boldsymbol{x}, \boldsymbol{z})$$

$K(\cdot, \cdot)$ is a kernel on $2^{\mathcal{X}}$ $\Longleftrightarrow$ $\kappa(\cdot, \cdot)$ is a kernel on $\mathcal{X}$

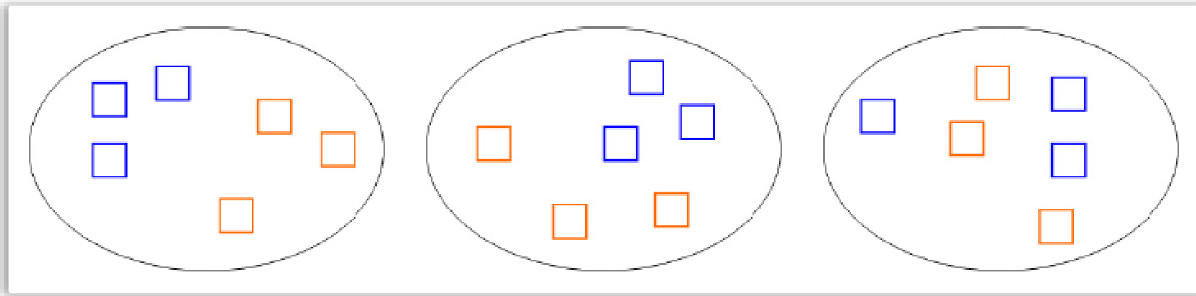
$K(X, Z) \Rightarrow D(X, Z) = \sqrt{K(X, X) - 2K(X, Z) + K(Z, Z)}$

Set kernel exploits the pairwise relations across bags, while ignores relations among instances within the bag

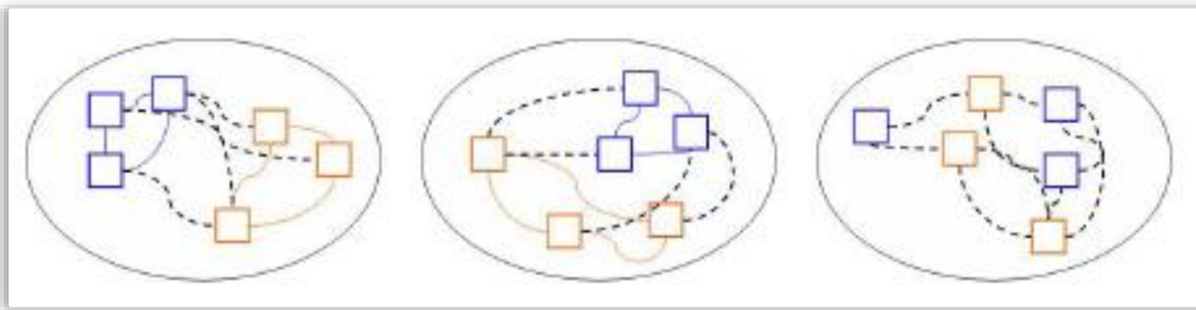
MIGraph [Zhou et al., ICML'09]

Usefulness of relation information among instances

image with six marked patches each corresponding to an instance



ignore relations among instances ➔ all bags are similar due to identical number of similar instances



consider relations among instances ➔ the first two bags are more similar than the third bag

MIGraph (Cont.)

Map each bag into a (weighted) graph

$X \Rightarrow G_X = (V_X, E_X)$

(ϵ-graph)

$$V_X = \{\boldsymbol{x}_i \mid 1 \leq i \leq m\}$$

$$E_X = \{\boldsymbol{e}_{uv} \mid \mathrm{dist}(\boldsymbol{x}_u, \boldsymbol{x}_v) \leq \epsilon\}$$

$$(\text{weight } w_{uv} \propto 1/\mathrm{dist}(\boldsymbol{x}_u, \boldsymbol{x}_v))$$

Derive the graph kernel

$$K_G(X, Z) = \boxed{\sum_{\boldsymbol{x} \in V_X} \sum_{\boldsymbol{z} \in V_Z} \kappa_{\text{node}}(\boldsymbol{x}, \boldsymbol{z})} + \boxed{\sum_{\boldsymbol{e} \in E_X} \sum_{\boldsymbol{e}' \in E_Z} \kappa_{\text{edge}}(\boldsymbol{e}, \boldsymbol{e}')}$$

MI-Kernel [Gärtner et al., ICML'02]

represent edge $\boldsymbol{e}_{uv}$ as $[d_u, p_u, d_v, p_v]^\top$

d_u : degree of $\boldsymbol{x}_u$

p_u : $w_{uv} / \sum_{\boldsymbol{e}_{uv} \in E_X} w_{uv}$

Complexity of MIGraph: $O(mn + |E_X||E_Z|)$

an efficient variant miGraph with complexity $O(mn)$ [Zhou et at., ICML'09]

MILES [Chen et al., TPAMI'06]

Vocabulary for embedding

$$V = \{\boldsymbol{x}_{ij} \mid 1 \leq i \leq N, 1 \leq j \leq n_i\}$$

$$= \{\boldsymbol{u}^k \mid 1 \leq k \leq K\}$$ (re-indexed with $K = \sum_{i=1}^{N} n_i$)

Embedding function: $M_V : 2^{\mathcal{X}} \mapsto \mathbb{R}^K$

$$M_V(X) = [a_1, \ldots, a_k, \ldots, a_K]^\top$$

$$a_k = \mathrm{sim}(X, \boldsymbol{u}^k) = \max_{\boldsymbol{x} \in X} \exp\left(-\frac{||\boldsymbol{x} - \boldsymbol{u}^k||^2}{\sigma^2}\right)$$

Generate binary training set via embedding function M_V: $D_V = \{(M_V(X_i), y_i) \mid 1 \leq i \leq N\}$

→

Induce binary classifier $h : \mathbb{R}^K \mapsto \{-1, +1\}$ based on D_V

→

Make prediction on unseen bag Z as: $h(M_V(Z))$

CCE [Zhou & Zhang, KAIS'07]

Vocabulary for embedding

$$C = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k, \ldots, \boldsymbol{c}_{\hat{K}}\}$$

(centers of $\hat{K}$ groups by clustering all the instances in V)

Embedding function: $M_C : 2^{\mathcal{X}} \mapsto \{0,1\}^{\hat{K}}$

$$M_C(X) = [a_1, \ldots, a_k, \ldots, a_{\hat{K}}]^\top$$

$a_k = 1 \text{ if } S_k \neq \emptyset; \text{ Otherwise, } a_k = 0$

(S_k : instances of X falling into the k-th group, i.e. closest to $\boldsymbol{c}_k$)

Ensemble

Generate binary training set via embedding function M_C: $D_C = \{(M_C(X_i), y_i) \mid 1 \leq i \leq N\}$

→

Induce binary classifier $h : \{0,1\}^{\hat{K}} \mapsto \{-1, +1\}$ based on D_C

→

Make prediction on unseen bag Z as: $h(M_C(Z))$

Outline

- Multi-Instance Learning (MIL)
- Representative Algorithms
- Related Topics
- Applications in Computer Vision
- Conclusion

Multi-Instance Regression (MIR)

In some applications, real-valued responses would be more desirable than discrete-valued ones

e.g.: *activity level* prediction vs. *active/inactive* classification

Applications of MIR

- drug activity prediction [Ray & Page, ICML'01; Amar et al., ICML'01]
- climate research [Wang et al., SDM'08]
- sentiment analysis [Pappas & Popescu-Belis, EMNLP'14]
- remote sensing [Wang et al., TGRS'12]
- ……

Solving MIR problems

IS paradigm: EM-style iterations for updating instance-level regressor

BS/ES paradigm: BS/ES methods for classification readily adaptable

MIL with Unlabeled Data

Multi-instance clustering [Zhang & Zhou, APIN'07; Zhang et al., TNN'11]

Gain insights on the distribution of multi-instance bags, and serve as a pre-processing step for multi-instance prediction (e.g. generate bag-level vocabulary for the ES paradigm)

Semi-supervised MIL [Rahmani & Goldman, ICML'06; Jia & Zhang, AAAI'08; Zeisl et al., CVPR'08]

If instances in bags were assumed as i.i.d. samples, MIL is just a special case of semi-supervised learning [Zhou & Xu, ICML'07]

instances from negative bags ⇨ labeled negative examples

instances from positive bags ⇨ unlabeled examples with positive constraints

Multi-instance active learning [Settles et al., NIPS'07; Liu et al., MMM'09]

Generalized MIL (GMIL)

Standard MIL assumption

An underlying target concept $h : \mathcal{X} \mapsto \{-1, +1\}$ which governs the prediction of class label $f(X)$ on the bag X:

$$f(X) = +1 \iff \exists \boldsymbol{x} \in X : h(\boldsymbol{x}) = +1$$

GMIL assumption [Weidmann et al., ECML'03]

A set of underlying concepts $\mathcal{H} = \{h_1, \ldots, h_r\}$ which governs the prediction of class label $f(X)$ on the bag X:

presence-based: $f(X) = +1 \iff \forall h_i \in \mathcal{H} : \Delta(X, h_i) \geq 1$

threshold-based: $f(X) = +1 \iff \forall h_i \in \mathcal{H} : \Delta(X, h_i) \geq t_i$

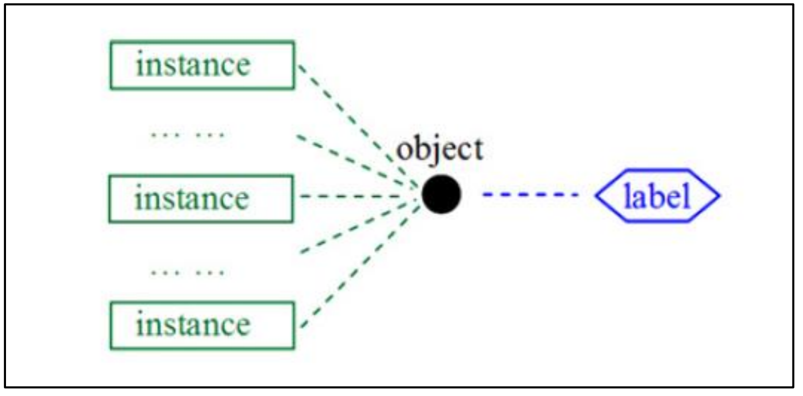
count-based: $f(X) = +1 \iff \forall h_i \in \mathcal{H} : t_i \leq \Delta(X, h_i) \leq z_i$

$\Delta(X, h_i)$: number of instances in X with concept h_i

t_i, z_i : lower and upper threshold for concept h_i

Another GMIL formulation: assuming *attraction* concepts as well as *repulsion* concepts [Scott et al., TechRep'03; Tao et al., ICML'04; Tao et al., TPAMI'08]

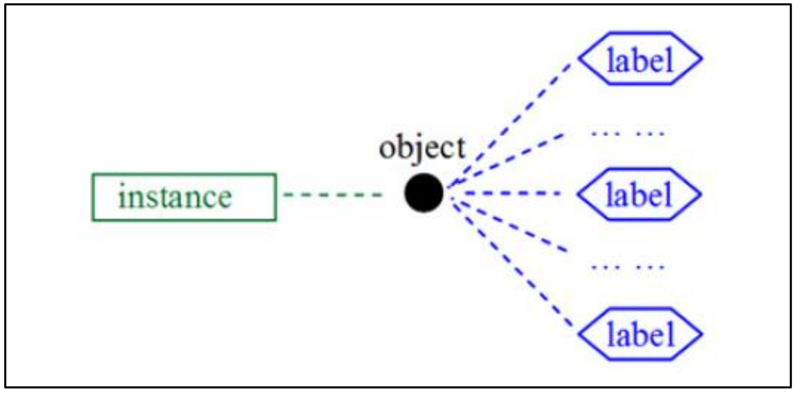
Multi-Instance Multi-Label Learning (MIML)



Multi-Instance Learning (MIL)

Input Ambiguity

object with many alternative input descriptions, i.e. instances



Multi-Label Learning (MLL)

[Zhang & Zhou, TKDE'14]

Output Ambiguity

object with many alternative input descriptions, i.e. labels

MIML (Cont.)

Real-world objects are usually inherited with input ambiguity as well as output ambiguity

[Zhou et al., AIJ'12; Zha et al., CVPR'08; Briggs et al., KDD'12; Li et al., TCBB'12; Surdeanu et al., EMNLP'12; Wu et al., TCBB'14]

An image usually contains **multiple** regions each can be represented by an instance

The image can simultaneously belong to **multiple** classes

Elephant

Lion

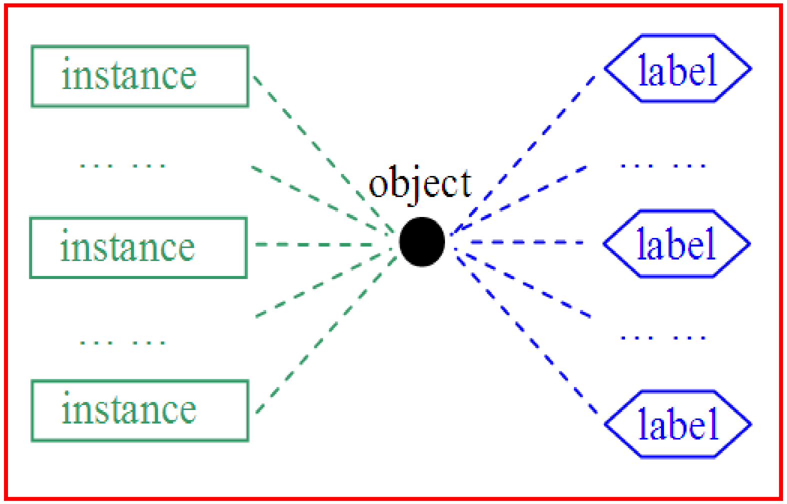
Grassland

Tropic

Africa

… …

MIML (Cont.)



MIML task:

To learn a function $f_{MIML} : 2^{\mathcal{X}} \to 2^{\mathcal{Y}}$ from a given data set $\{(X_1, Y_1), (X_2, Y_2), \cdots, (X_m, Y_m)\}$ where $X_i \subseteq \mathcal{X}$ is a set of instances $\{\boldsymbol{x}_1^{(i)}, \boldsymbol{x}_2^{(i)}, \cdots, \boldsymbol{x}_{n_i}^{(i)}\}$, $\boldsymbol{x}_j^{(i)} \in \mathcal{X}$ $(j = 1, 2, \cdots, n_i)$, and $Y_i \subseteq \mathcal{Y}$ is a set of labels $\{y_1^{(i)}, y_2^{(i)}, \cdots, y_{l_i}^{(i)}\}$ $y_k^{(i)} \in \mathcal{Y}$ $(k = 1, 2, \cdots, l_i)$.

$\mathcal{X}$ - the instance space

$\mathcal{Y}$ - the set of class labels

n_i - the number of instances in X_i

l_i - the number of labels in Y_i

More on MIML:

Zhou Z-H, Zhang M-L, Huang S-J, Li Y-F. Multi-instance multi-label learning. **Artificial Intelligence**, 2012, 176(1): 2291-2320.

Key Instance Detection (KID) for MIL

Detect key (positive) instances in the bags are desirable for a number of MIL applications

e.g. locating *region of interest* (ROIs) in CBIR [Zhou et al., AJCAI'05]

Solutions to KID in MIL

- Rely on instance-level classifier of IS methods (e.g. DD, MI-SVM)
- Design customized algorithms [Li et al., ECML'09; Liu et al., ACML'12; Kandemir & Hamprecht, UAI'14; Kotzias et al., KDD'15]

Bag Generator for MIL

Pipeline of MIL

Objects $\xrightarrow{\text{Bag generator}}$ MIL Bags $\xrightarrow{\text{MIL algorithms}}$ MIL Classifiers

Many effective MIL algorithms have been proposed and studied, e.g., the J. Amores's AIJ survey.

Few studies were about MIL bag generators although they affect the performance seriously.

Image bag generators are of particular interest in existing MIL studies

Bag Generator for MIL (Cont.)

Nine popular image bag generators [Wei & Zhou, MLJ, in press]

- Row [Maron & Ratan, ICML'98]
- SB [Maron & Ratan, ICML'98]
- SBN [Maron & Ratan, ICML'98]

Non-segmentation bag generators

- Blobworld [Carson et al., TPAMI'02]
- *k*-meansSeg [Zhang et al., ICML'02]
- WavSeg [Zhang et al., ICME'04]
- JSEG-bag [Liu et al., ICIEA'08]

Segmentation bag generators

- LBP [Ojala et al., TPAMI'02]
- SIFT [Lowe, IJCV'04]

Local descriptor-based bag generators

Bag Generator for MIL (Cont.)

SBN (Single Blob with Neighbors)

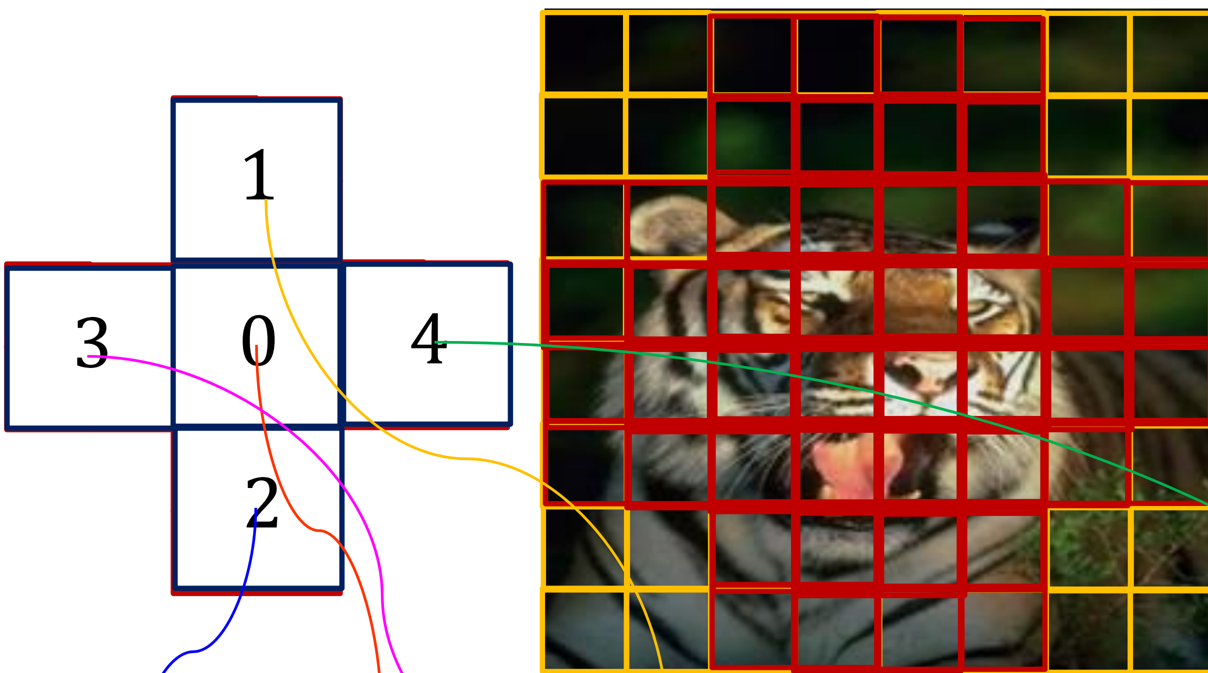


Image (bag)

filtered and resized to 8x8

Blob: 2x2 patch

Instance

one blob with four neighboring blobs

$$instance = < r_{i0}, g_{i0}, b_{i0}, r_{i0} - r_{i1}, g_{i0} - g_{i1}, b_{i0} - b_{i1},$$
$$r_{i0} - r_{i2}, g_{i0} - g_{i2}, b_{i0} - b_{i2}, r_{i0} - r_{i3}, g_{i0} - g_{i3}, b_{i0} - b_{i3}, r_{i0} - r_{i4}, g_{i0} - g_{i4}, b_{i0} - b_{i4} >$$

Bag Generator for MIL (Cont.)

Blobworld

Components of the L*a*b* color space

pixel features

extract

features

$$\{< L_i^*, a_i^*, b_i^*, a_i, c_i, p_i, x_i, y_i >|\ i = 1, 2, ..., \#\ of\ pixels\}$$

The texture features

The pixel position

group

features

image regions

describe

regions

regions features (instances)

Bag Generator for MIL (Cont.)

LBP (Local Binary Patterns)

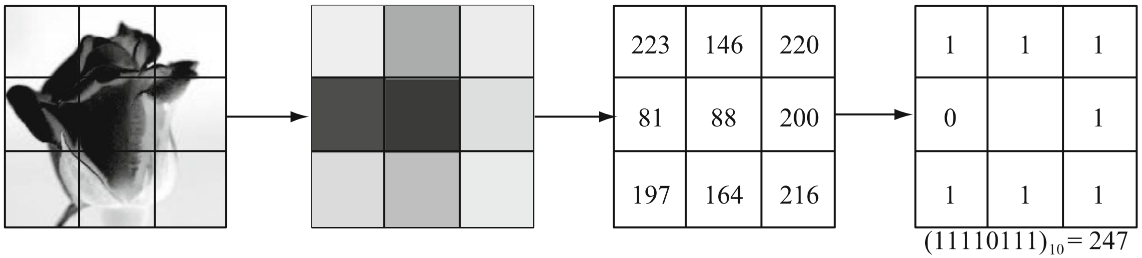


Image (bag)

35 sliding windows

Instance

58-dimensional LBP descriptors

SIFT (Scale Invariant Feature Transform)

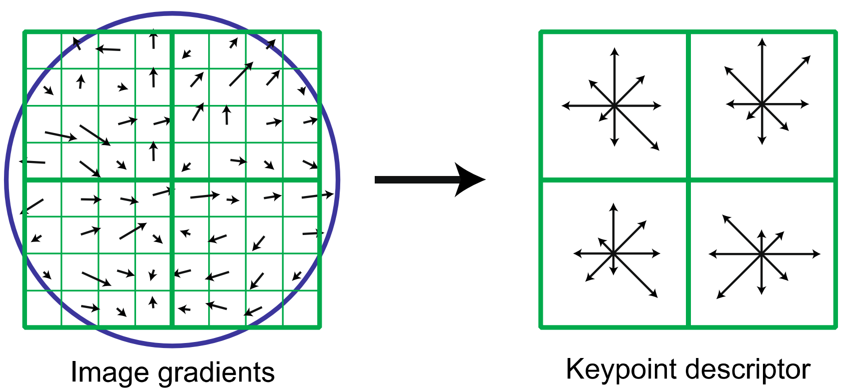


Image (bag)

40 keypoints

Instance

128-dimensional SIFT descriptors

Bag Generator for MIL (Cont.)

Recent extensive empirical study [Wei & Zhou, MLJ, in press]

6,923 configurations of experiments

[9 bag generators; 7 MIL algorithms; 4 patch sizes; 43 data sets]

Two significant new observations

- Bag generators with a dense sampling strategy (SB, SBN, LBP) perform better than those with other strategies
- The standard MIL assumption of learning algorithms is not suitable for image classification tasks

Code (image bag generators for MIL)

http://lamda.nju.edu.cn/code_MIL-BG.ashx

Scalable MIL

The limitation of MIL

An undesired outcome

most MIL algorithms are usually time-consuming and incapable of handling large scale MIL problems

Scalable MIL (Cont.)

The real world MIL applications

Millions of images

Millions of genes

And other millions of complex objects or examples …

Scalable MIL (Cont.)

Recent efforts towards scalable MIL [Wei et al., TNNLS, in press]

miFV (MIL based on the **Fisher Vector** representation)

Gradient vector for a sample of observations (i.e. bag)

[Jaakkola and Haussler, NIPS'99]

$$G_\lambda^S = \nabla_\lambda \log p(S|\lambda)$$

A sample of observations

Parameters

A probability density function

The dimensionality of G_λ^S only depends on the number of generative parameters in p, rather than on the sample size

Scalable MIL (Cont.)

Recent efforts towards scalable MIL [Wei et al., TNNLS, in press]

miFV (MIL based on the **Fisher Vector** representation)

Fisher vector representation

[Sánchez et al., IJCV'13]

$$\mathcal{K}_{FK}(S_1, S_2) = {G_\lambda^{S_1}}^\top \underbrace{F_\lambda^{-1}}_{\text{Fisher information matrix}} G_\lambda^{S_2}$$

Scalable MIL (Cont.)

Recent efforts towards scalable MIL [Wei et al., TNNLS, in press]

miFV (MIL based on the **Fisher Vector** representation)

Fisher vector representation

[Sánchez et al., IJCV'13]

$$\mathcal{K}_{FK}(S_1, S_2) = {G_\lambda^{S_1}}^\top F_\lambda^{-1} G_\lambda^{S_2}$$

Cholesky decomposition

$$\mathcal{K}_{FK}(S_1, S_2) = {G_\lambda^{S_1}}^\top L_\lambda^\top L_\lambda G_\lambda^{S_2}$$

Scalable MIL (Cont.)

Recent efforts towards scalable MIL [Wei et al., TNNLS, in press]

miFV (MIL based on the **Fisher Vector** representation)

Fisher vector representation

[Sánchez et al., IJCV'13]

$$\mathcal{K}_{FK}(S_1, S_2) = G_\lambda^{S_1\,\top} F_\lambda^{-1} G_\lambda^{S_2}$$

Cholesky decomposition

$$\mathcal{K}_{FK}(S_1, S_2) = \boldsymbol{f}_\lambda^{S_1\,\top} \boldsymbol{f}_\lambda^{S_2} L_\lambda G_\lambda^{S_2}$$

Fisher Vector

$$\boldsymbol{f}_\lambda^S = L_\lambda G_\lambda^S = L_\lambda \Delta_\lambda \log p(S|\lambda)$$

Nonlinear kernel machine with $\mathcal{K}_{FK}(\cdot,\cdot)$

Linear kernel machine with feature vector $\boldsymbol{f}_\lambda^S$

Scalable MIL (Cont.)

Advantages of miFV

- Dimensionality of the FV is independent of the size of training set ➔ efficiency
- Capture high-order statistics in the FV (e.g. (co)variance of instances) ➔ effectiveness

Highly competitive performance with hundreds (or even thousands) of times faster than state-of-the-art MIL algorithms
[data sets: up to 60K+ bags with 0.68M+ instances]

miVLAD (MIL based on the **Vector of Locally Aggregated Descriptors**)

A simplified version of miFV which is more efficient

Code (miFV & miVLAD)

http://lamda.nju.edu.cn/code_SMIL.ashx

Outline

- Multi-Instance Learning (MIL)
- Representative Algorithms
- Related Topics
- Applications in Computer Vision
- Conclusion

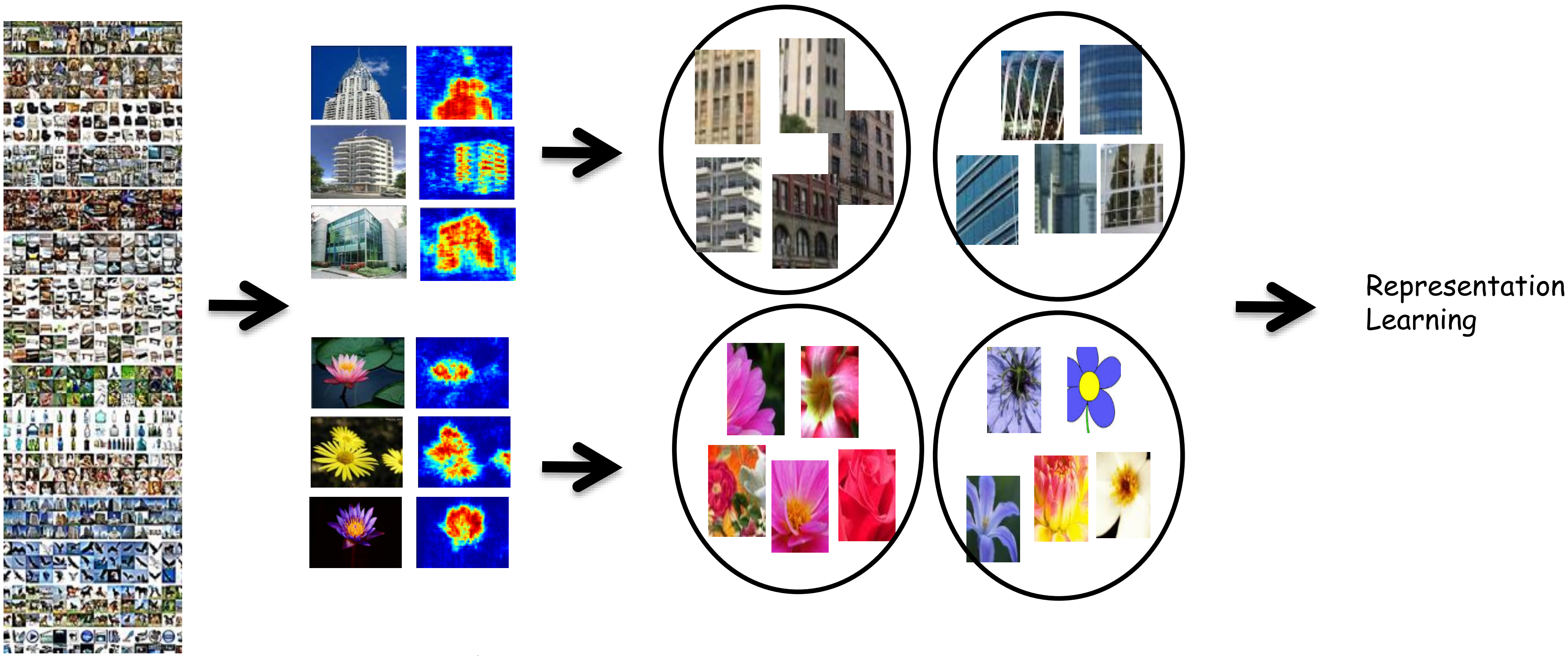
Applications in Computer Vision

- Image Classification
- Object Detection
- Object Discovery (Weakly Supervised Object Detection)
- Semantic Segmentation
- Visual Tracking

Image Classification

Mining mid-level image representation [Li et al., CVPR'13, Wang et al., ICML'13]

- Mining discriminative patterns using MIL for image representation



Step by step → End to end?

Quannan Li, Jiajun Wu, and Zhuowen Tu, "Harvesting Mid-level Visual Concepts from Large-scale Internet Images", CVPR 2013
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.

Image Classification

Multiple Instance Learning with Deep Learning [Wu et al., CVPR'15; Wei, et al., TPAMI'15]

- Learning deep feature for "instance" (image patch)
- Max/mean for multiple instance learning

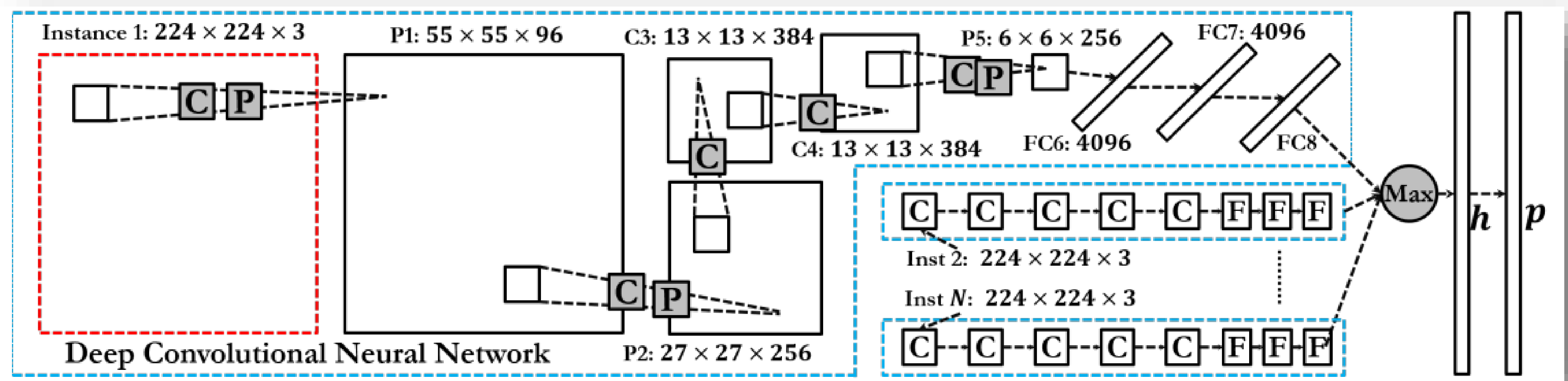
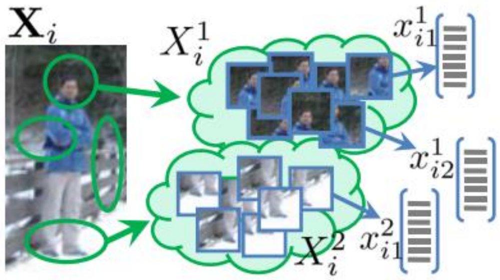


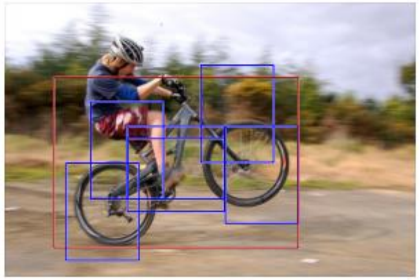
Figure from Jiajun

Jiajun Wu, Yinan Yu, Chang Huang, Kai Yu, Deep Multiple Instance Learning for Image Classification and Auto-Annotation, CVPR 2015
Yunchao Wei, Wei Xia, Min Lin, Junshi Huang, Bingbing Ni, Jian Dong, Yao Zhao, Shuicheng Yan, HCP: A Flexible CNN Framework for Multi-label Image Classification, PAMI 2015

Object Detection

Learning object detector using multiple instance learning





[Viola et. al, NIPS'06] [Dollar et. al, ECCV'08] [Felzenszwalb et. al, TPAMI'10]

Learning object detector using weak label

Mining object part from bounding-box annotation

Paul Viola, John C. Platt, and Cha Zhang, Multiple Instance Boosting for Object Detection, NIPS 2006
Piotr Dollár, Boris Babenko, Serge Belongie, Pietro Perona, and Zhuowen Tu, "Multiple Component Learning for Object Detection", ECCV 2008
Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester and Deva Ramanan, Object Detection with Discriminatively Trained Part Based Models, TPAMI 2010

Object Discovery

Big Data vs. Expensive Labels

Possible solutions to this problem: clustering based, matching based, co-segmentation based, topic model based, multi-instance learning based methods.

Solving this problem by multiple instance learning
- Image as bag, since image label is given
- Proposals (Selective Search, EdgeBox, Bing) as instances [Zhu et al, TPAMI'14]
    - Proposal descriptors: Deep CNN Features, Fisher Vectors
    - Number of proposals: ~2k (SS), ~3k (EB)

Jun-Yan Zhu, Jiajun Wu, Yan Xu, Eric Chang, and Zhuowen Tu, "Unsupervised Object Class Discovery via Saliency-Guided Multiple Class Learning", TPAMI, 2014.

Benchmark

- PASCAL VOC 2007
  - CorLoc, percentage of image with at least one correctly localized object
  - Average Precision, BBox IoU >= 0.5

	CorLoc (%) (trainval)	mAP (test)
Multi-fold MIL [Cinbis et al., TPAMI'16]	47.3	27.4
RMI-SVM [Wang et al., ICCV'15]	40.2	-
LCL [Wang et al., ECCV'14]	48.5	31.6
WSDDN [Bilen & Vedaldi, CVPR'16]	58.0	39.3
DPM-v5	-	33.7

- ImageNet [Tang et al., CVPR'14]
  - Average CorLoc of 3,624 classes and 939,542 images is **53.2%**

Ramazan Gokberk Cinbis, Jakob Verbeek, and Cordelia Schmid, Weakly Supervised Object Localization with Multi-fold Multiple Instance Learning, TPAMI 2016
Kevin Tang, Armand Joulin, Li-Jia Li, Li Fei-Fei, Co-localization in Real-World Images, CVPR 2014
C. Wang, W. Ren, K. H., and T. Tan. Weakly supervised object localization with latent category learning. ECCV 2014
Xinggang Wang, Zhuotun Zhu, Cong Yao, Xiang Bai. Relaxed Multiple-Instance SVM with Application to Object Discovery. ICCV 15
H. Bilen, A. Vedaldi. Weakly Supervised Deep Detection Networks. CVPR 2016.

Relaxed MIL for Object Discovery

- A low rank formulation [Wang et al., NC'13]

$$min_{A,E,Z}\|k\|_* + \gamma\|E\|_1$$
$$s.t.\ Xdiag(Z) = A + E, \forall k\epsilon[k] \bigvee_{i=1}^{n_k} z_i^{(k)} = 1$$

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

- A discriminative formulation

$$p_{ij} = \Pr(y_{ij} = 1|x_{ij}; w) = \frac{1}{1 + e^{-w^T x_{ij}}}$$

$$P_i = \Pr(Y_i = 1|X_i; w) = 1 - \prod_{j=1}^{m_i}(1 - p_{ij}) \text{ (NOR)}$$

$$\min_w \frac{\lambda}{2}\|w\|^2 + \frac{\beta}{n}\sum_{i=1}^{n}\mathcal{L}_{bag_i} + \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}\mathcal{L}_{ins_{ij}}$$

$$\mathcal{L}_{bag_i} = -\{Y_i \log(P_i) + (1 - Y_i)\log(1 - P_i)\},$$
$$\mathcal{L}_{ins_{ij}} = \max(0, [m_0 - sgn(p_{ij} - p_0)w^T x_{ij}])$$

X. Wang, Z. Zhu, C. Yao, X. Bai. Relaxed Multiple-Instance SVM with Application to Object Discovery. ICCV 15
X. Wang, Z. Zhang, Y. Ma, X. Bai, W. Liu, and Z. Tu. Robust Subspace Discovery via Relaxed Rank Minimization, Neural Computation 13

Multi-fold MIL for Object Discovery

Motivation: avoiding poor local optima in MIL

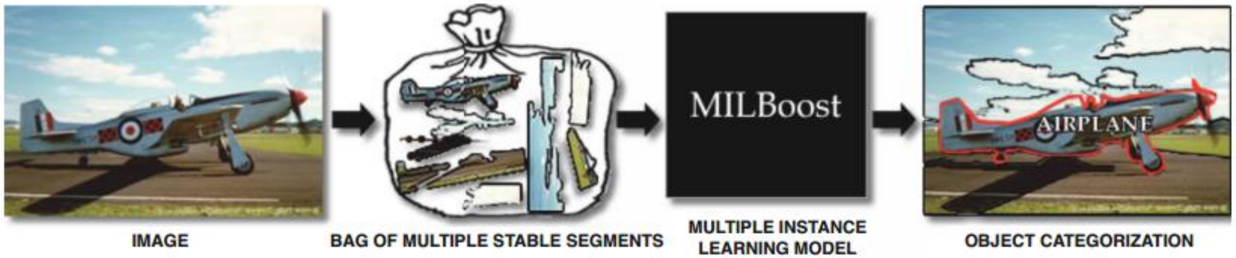
Multi-fold weakly supervised training

1) Initialization of positive and negative instances
2) For iteration t = 1 to T
   a) Divide positive images randomly into K folds
   b) For k = 1 to K
      i) Train using positive examples in all folds but k, and all negative examples
      ii) **Re-localize** positives by selecting the top scoring window in each image of fold k using this detector
   c) Train detector using re-localized positives and all negatives
   d) Add new negative windows by hard-negative mining
3) Return final detector and object windows in train data

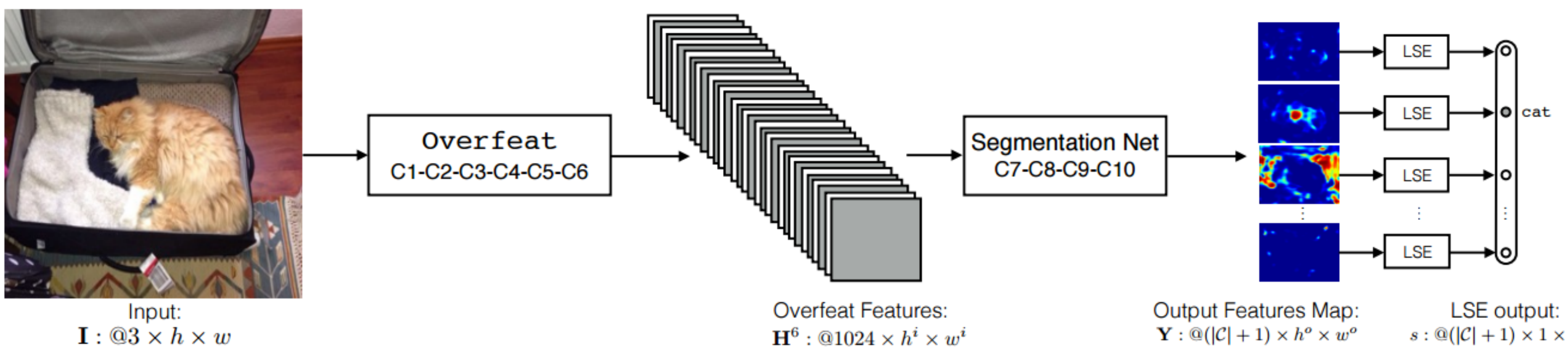
Ramazan Gokberk Cinbis, Jakob Verbeek, and Cordelia Schmid, Weakly Supervised Object Localization with Multi-fold Multiple Instance Learning, TPAMI 2016

Semantic Segmentation

- MIL-Boost Segmentation [Galleguillos et al., ECCV'08]



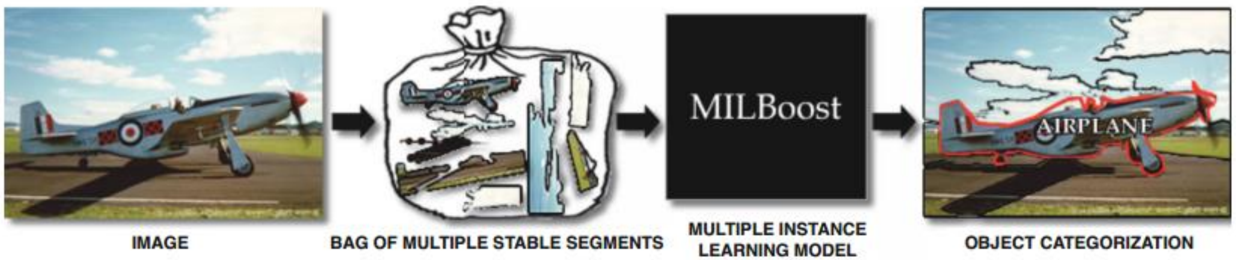
- Weakly Supervised Deep Segmentation [Pinheiro et al., CVPR'15]



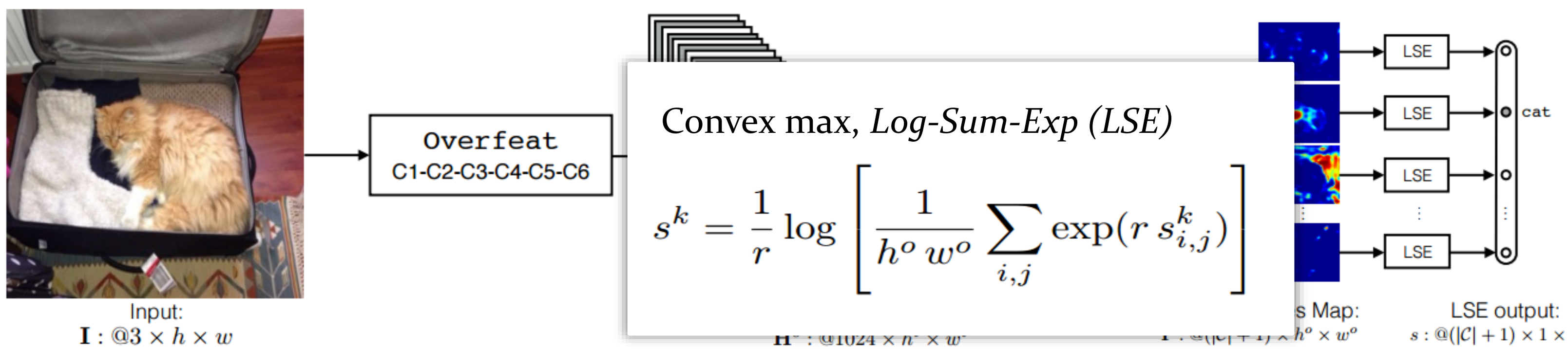
Carolina Galleguillos, Boris Babenko, Andrew Rabinovich, Serge Belongie, Weakly Supervised Object Localization with Stable Segmentations, ECCV 08
Pedro O. Pinheiro, Ronan Collobert. From Image-level to Pixel-level Labeling with Convolutional Networks. CVPR 2015

Semantic Segmentation

- MIL-Boost Segmentation [Galleguillos et al., ECCV'08]



- Weakly Supervised Deep Segmentation [Pinheiro et al., CVPR'15]



Convex max, *Log-Sum-Exp (LSE)*

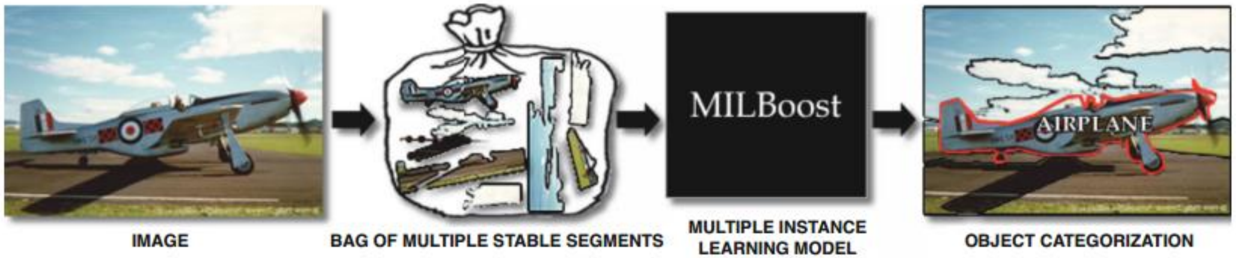
$$s^k = \frac{1}{r} \log \left[\frac{1}{h^o\, w^o} \sum_{i,j} \exp(r\, s^k_{i,j}) \right]$$

Carolina Galleguillos, Boris Babenko, Andrew Rabinovich, Serge Belongie, Weakly Supervised Object Localization with Stable Segmentations, ECCV 08

Pedro O. Pinheiro, Ronan Collobert. From Image-level to Pixel-level Labeling with Convolutional Networks. CVPR 2015

Semantic Segmentation

- MIL-Boost Segmentation [Galleguillos et al., ECCV'08]



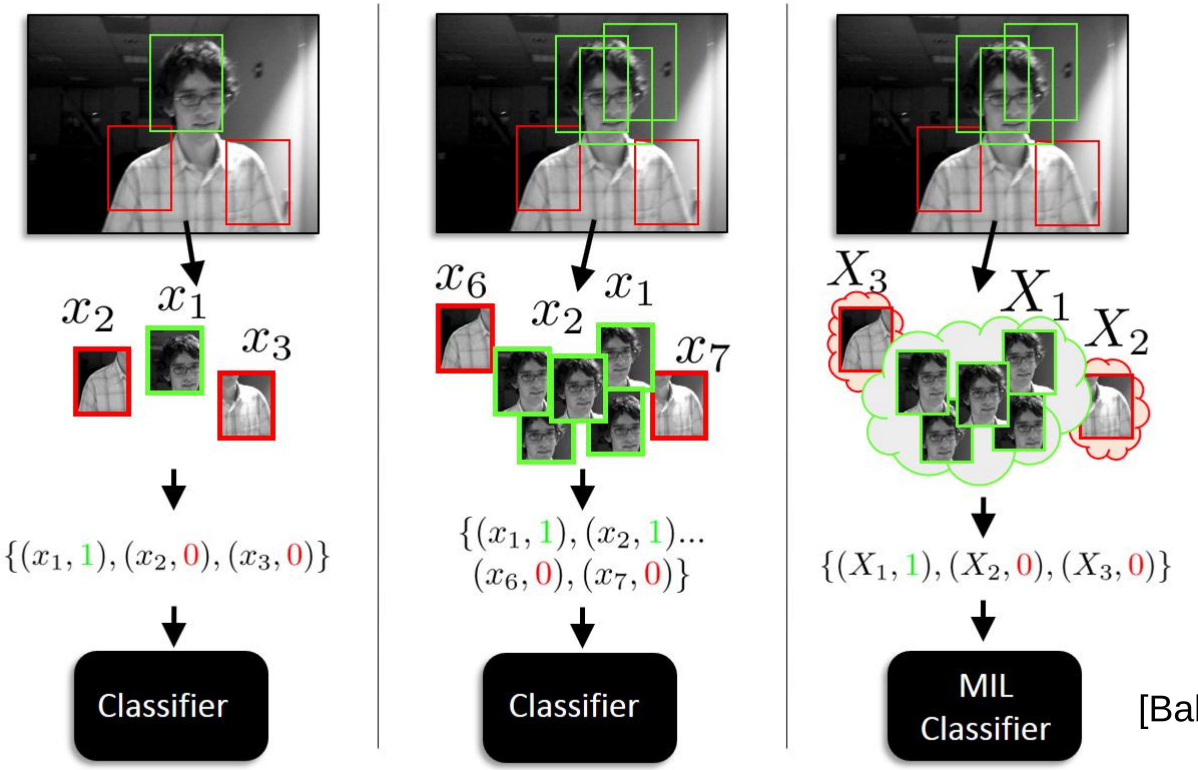
-

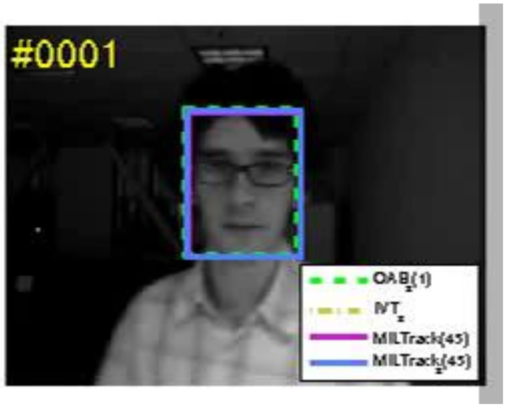
Caro
Segr
Pedr

Mir

Object Tracking

Motivation: Learning robust appearance model using MIL





[Babenko, 11]

Babenko B., Yang M., Belongie S., "Visual Tracking with Online Multiple Instance Learning", TPAMI 2011

Applications in Computer Vision

- MIL is a very effective tool for weakly supervised learning in computer vision applications.
- MIL is able to mining informative features in image.
- MIL works well with deep networks! Any differentiable MIL operator can be integrated into deep networks.

Outline

- Multi-Instance Learning (MIL)
- Representative Algorithms
- Related Topics
- Applications in Computer Vision
- Conclusion

Introductory Materials

- ✓ Zhou Z-H. Multi-instance learning: A survey. **Technical Report**, AI Lab, Department of Computer Science & Technology, Nanjing University, Nanjing, China, 2004.
- ✓ Foulds J, Frank E. A review of multi-instance learning assumptions. **Knowledge Engineering Review**, 2010, 25(1): 1-25.
- ✓ Amores J. Multiple instance classification: Review, taxonomy and comparative study. **Artificial Intelligence**, 2013, 201: 81-105.
- ✓ Ray S, Scott S D, Blockeel H. Multiple-instance learning. In: **Encyclopedia of Machine Learning and Data Mining**, Berlin: Springer, 2015.

For More Details…

Multi-instance regression

- ✓ Ray S, Page D. Multi-instance regression. In: **Proceedings of the 18th International Conference on Machine Learning (ICML'01)**, Williamstown, MA, 2001, 425-432.
- ✓ Amar R A, Dooly D R, Goldman S A, Zhang Q. Multiple-instance learning of real-valued data. In: **Proceedings of the 18th International Conference on Machine Learning (ICML'01)**, Williamstown, MA, 2001, 3-10.

Multi-instance clustering

- ✓ Zhang M-L, Zhou Z-H. Multi-instance clustering with applications to multi-instance prediction. **Applied Intelligence**, 2009, 31(1): 47-68.
- ✓ Zhang D, Wang F, Si L, Li T. Maximum margin multiple instance clustering with applications to image and text clustering. **IEEE Transactions on Neural Networks**, 2011, 22(5): 739-751.

For More Details…

Semi-supervised MIL

- ✓ Rahmani R, Goldman S A. MISSL: multiple-instance semi-supervised learning. In: **Proceedings of the 23rd International Conference on Machine Learning (ICML'06)**, Pittsburgh, MA, 2006, 705-712.
- ✓ Zhou Z-H, Xu J-M. On the relation between multi-instance learning and semi-supervised learning. In: **Proceedings of the 24th International Conference on Machine Learning (ICML'07)**, Corvallis, OR, 2007, 1167-1174.

Multi-instance active learning

- ✓ Settles B, Craven M, Ray S. Multiple-instance active learning. In: **Advances in Neural Information Processing Systems 20 (NIPS'07)**, Cambridge, MA: MIT Press, 2008, 1289-1296.

For More Details…

Generalized MIL

- ✓ Weidmann N, Frank E, Pfahringer B. A two level learning method for generealized multi-instance learning. In: **Proceedings of the 14th European Conference on Machine Learning (ECML'03)**, Cavtat-Dubrovnik, Croatia, 2003, 492-502.
- ✓ Tao Q, Scott S D, Vinodchandran N V, Osugi T T, Mueller B. Kernels for generalized multiple-instance learning. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 2008, 30(12): 2084-2098.

Multi-instance multi-label learning (MIML)

- ✓ Zhou Z-H, Zhang M-L, Huang S-J, Li Y-F. Multi-instance multi-label learning. **Artificial Intelligence**, 2012, 176(1): 2291-2320.
- ✓ Zhang M-L, Zhou Z-H. A review on multi-label learning algorithms. **IEEE Transactions on Knowledge and Data Engineering**, 2014, 26(8): 1819-1837.

For More Details

Key Instance Detection for MIL

- ✓ Li Y-F, Kwok J T, Tsang I W, Zhou Z-H. A convex method for locating regions of interest with multi-instance learning. In: **Proceedings of the 20th European Conference on Machine Learning (ECML'09)**, Bled, Slovenia, 2009, 15-30.
- ✓ Liu G, Wu J, Zhou Z-H. Key instance detection in multi-instance learning. In: **Proceedings of the 4th Asian Conference on Machine Learning (ACML'12)**, Singapore, 2012, 253-268.

Bag generators for MIL

- ✓ Wei X-S, Zhou Z-H. An empirical study on image bag generators for multi-instance learning. **Machine Learning**, in press.

Scalable MIL

- ✓ Wei X-S, Wu J, Zhou Z-H. Scalable algorithms for multi-instance learning. **IEEE Transactions on Neural Networks and Learning Systems**, in press.

Data & Code

Data for MIL

- ✓ http://www.miproblems.org/datasets/
- ✓ http://www.cs.wustl.edu/~sg/multi-inst-data/
- ✓ http://lamda.nju.edu.cn/CH.Data.ashx#data

Code for MIL

- ✓ https://weka.wikispaces.com/Multi-instance+classification [Weka MIL toolbox]
- ✓ http://prlab.tudelft.nl/david-tax/mil.html
- ✓ http://lamda.nju.edu.cn/CH.Data.ashx#code

Thanks!