

Generalized Multiple-Instance-Learning and its applications in biomedical imaging

Jan Hering

Center for Machine Perception (CMP),
Katedra kybernetiky, FEL, ČVUT

21.12.2017



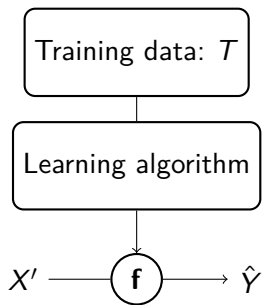
**FACULTY
OF ELECTRICAL
ENGINEERING
CTU IN PRAGUE**



Overview

- Standard MIL
 - formulation and algorithms
- Generalized MIL
 - formulation and algorithms
 - applications
- MIL in deep learning

Supervised learning



Input Data

$$X = \{x_1, x_2, \dots, x_N\}, x_i \in \mathcal{X}$$

$$Y = \{y_1, y_2, \dots, y_N\}, y_i \in \mathcal{Y}$$

$$T = \{(x_i, y_i)\}$$

$$\mathcal{Y} = \{-1, +1\} \text{ (binary case)}$$

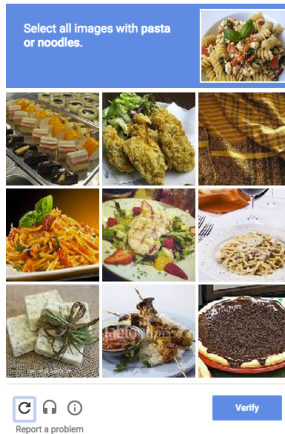
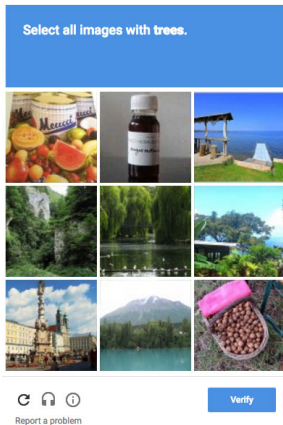
Classifier

$$f: \mathcal{X} \rightarrow \mathcal{Y}$$

- limited by (pixel-wise) label availability

Get more labels...

- hire a crowd, f.i. Google's re-CAPTCHA



TO COMPLETE YOUR REGISTRATION, PLEASE TELL US
WHETHER OR NOT THIS IMAGE CONTAINS A STOP SIGN:



NO

YES

ANSWER QUICKLY—OUR SELF-DRIVING
CAR IS ALMOST AT THE INTERSECTION.

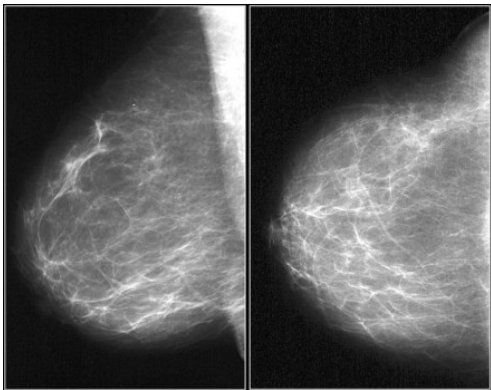
SO MUCH OF "AI" IS JUST FIGURING OUT WAYS
TO OFFLOAD WORK ONTO RANDOM STRANGERS.

Get more labels...

CT Lung



Mammography



In clinical routine:

- labels available on patient-level
 - healthy/diseased
 - diagnostic staging
- detailed annotations time consuming
- requires a medical expert

General applications:

- molecule polymorphism (MUSK) ¹
- content-based image retrieval (semantic annotation)

→ **weakly-supervised learning**

¹Dietterich et al.

Definition

Standard Multiple-Instance Learning (MIL)

The building blocks are *bags* $B_I, I \in \mathcal{I}$, each denotes a group of instances x_i , i.e. $B_I = \{x_i, i \in I\} \subset \mathbb{R}^m$. Furthermore, each bag is assigned a label $y_I \in \{-1, +1\}$. Instance labels $y(x_i)$ are not given. A standard MIL problem is defined if it holds:

- ① $y_I = +1 \Leftrightarrow \exists x_i, \text{ s.t. } y(x_i) = +1$ (positive identifiability)
- ② $y_I = -1 \Leftrightarrow \forall x_i \ y(x_i) = -1$ (negative exclusion)

MIL Classification

For classification, we want either

- an *instance classifier*

$$f(X) : \mathcal{X} \rightarrow \mathcal{Y}$$

- a *bag classifier*

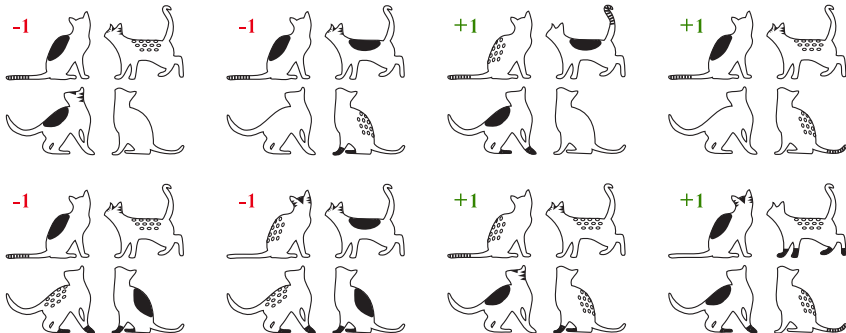
$$F(X^m) : \mathcal{X}^m \rightarrow \mathcal{Y}$$

Note

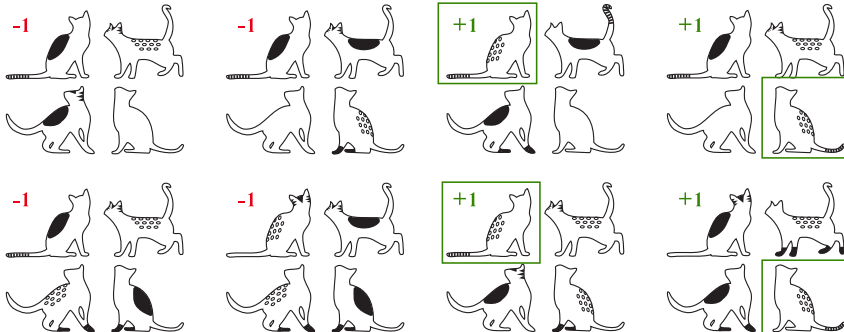
Each instance classifier $f(X)$ induces a bag classifier $F(X^m)$ by

$$F(X^m) = \max\{f(x_1), \dots, f(x_m)\},$$

MIL Example (1)



MIL Example (2)



MIL Approach Illustration

Support Vector Machine classifier:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|_2^2 + C \sum_{ij} \xi_{ij} \\ \text{s.t.} \quad & y_{ij}(w \cdot x_{ij} + b) \geq 1 - \xi_{ij} \\ & \xi_{ij} \geq 0 \end{aligned}$$

Construction of a MIL classifier:

- **MI-SVM** (bag-level)
- **mi-SVM, MI-RF**: (instance-level)
- further approaches²

²Carbonneau et al., J PatRec'2017, Review article

SVM-based MIL algorithms

Solution through iterative **heuristics**:³

MI-SVM

- ① start with bag labels
- ② while labels change:
 - select bag witness $s(j)$
 - train_SVM($X^- \cup \{s(j)\}$)

mi-SVM

- ① set $y(x_{ij}) = Y_i$
- ② while labels change:
 - train_SVM($X^- \cup X^+$)
 - ensure MIL-conditions

³Andrews et al.

MIL Random Forests

- use **deterministic annealing**(DA) with random forests (RF)⁴
- DA-formulation of the loss function:

$$\mathcal{L}(\mathbf{F}, \hat{p}) = \sum_{i,j} \sum_k \hat{p}(k|x_i^j) \ell(\mathbf{F}_k(x_i^j)) + T \sum_i \mathcal{H}(\hat{p}_i).$$

- ① find minimal \hat{p} for fixed confidences $F_k(x_i)$
- ② **re-train** the *RF* with sample distribution \hat{p} (ensure at least one positive instance in positive bags prior to training phase)

⁴Leistner et al., ECCV'2010

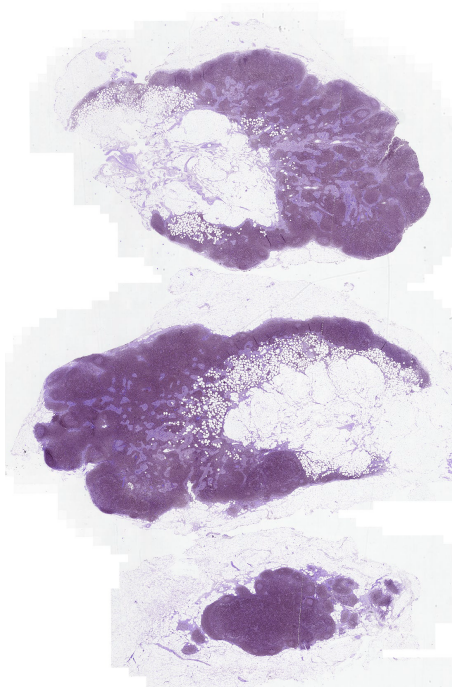
Example I

Detection of bone-marrow infiltrations in low-dose CT images of femurs. Scans of both femurs are provided, labels are available only at patient-level.



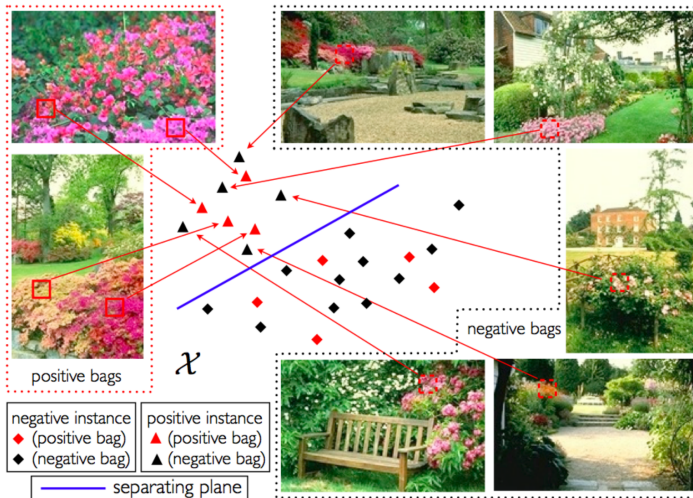
Example II

Detect tumorous
histo-pathology sections. The
tumor annotations may not
be complete or may be
missing completely (i.e. only
image-level labels).



Weak labels

Concept (**flowers**) present also in negative bags:



Standard MIL

Weak points of standard MIL

- **robustness** – impact of a single false detection
- **weak labels** – other concepts present in the input image
- **counting** – the number of positive instances is decisive ("traffic jam")

Definition

Standard Multiple-Instance Learning (MIL)

The building blocks are *bags* $B_I, I \in \mathcal{I}$, each denotes a group of instances x_i , i.e. $B_I = \{x_i, i \in I\} \subset \mathbb{R}^m$. Furthermore, each bag is assigned a label $y_I \in \{-1, +1\}$. Instance labels $y(x_i)$ are not given. A standard MIL problem is defined if it holds:

- 1 $y_I = +1 \Leftrightarrow \sum_{i \in I} \mathbb{I}[y(x_i) > 0] \geq 1$ (positive identifiability)
- 2 $y_I = -1 \Leftrightarrow \sum_{i \in I} \mathbb{I}[y(x_i) > 0] < 1$ (negative exclusion)

Definition

Generalized Multiple-Instance Learning (gMIL)

The building blocks are *bags* $B_I, I \in \mathcal{I}$, each denotes a group of instances x_i , i.e. $B_I = \{x_i, i \in I\} \subset \mathbb{R}^m$. Furthermore, each bag is assigned a label $y_I \in \{-1, +1\}$. Instance labels $y(x_i)$ are not given.

A bag is said to be ζ -positive, if it holds

- 1 $y_I = +1 \Leftrightarrow \sum_{i \in I} \mathbb{I}[y(x_i) > 0] \geq \zeta$ (positive identifiability)
- 2 $y_I = -1 \Leftrightarrow \sum_{i \in I} \mathbb{I}[y(x_i) > 0] < \zeta$ (negative exclusion)

gMIL Algorithms

Common principle: With an instance-classifier C

- While labels change
 - ① predict labels with C
 - ② ensure **at least** k positive instance in each positive bag
 - ③ retrain C with current labels $y(x_{ij})$
- k -mi-SVM
- k -MI-SVM
 - k -top-MI-SVM⁵
- k -MIL-RF

First approach: introduce a (hyper-)parameter k

Q: How to **learn** the threshold k (resp. ζ).

⁵Li and Vasconcelos, CVPR'2015

gMIL Algorithms (cont'd)

While labels change:

- 1 predict labels with C
- 2 $\hat{k} \leftarrow \text{get_k_hat}(X^-, X^+)$
- 3 ensure **at least** \hat{k} positive instance in each positive bag
- 4 retrain C with current labels $y(x_{ij})$

get_k_hat

estimate count pos. instances in pos. and neg. bags, \hat{k} their average

optimize take \hat{k} that minimizes bag-level classification error

Experiments

Applied to detection of multiple myeloma infiltrations (Example I) ⁶.

Classification features

Split femur into segments l, t, ϕ and compute intensity-based features. Each instance is then $\mathbf{x}_i = [l, t, \phi, \mu, \sigma, h_1, \dots, h_{N_h}]$.

		k	mi-SVM ⁸	MI-SVM ⁸	MIL-ARF ⁹
AUC	1		0.747 ± 0.065	0.751 ± 0.053	0.854 ± 0.055
	ζ_{est}		$0.867 \pm 0.049^\ddagger$	$0.774 \pm 0.068^\dagger$	$0.874 \pm 0.048^\dagger$
	ζ_{opt}		$0.823 \pm 0.066^\ddagger$	0.764 ± 0.058	$0.876 \pm 0.048^\ddagger$

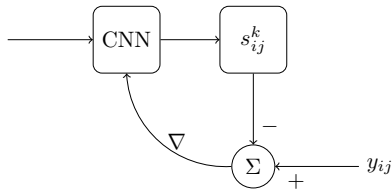
⁶Hering et al., SPIE'2018

(Convolutional) Neural Networks

Weakly supervised networks

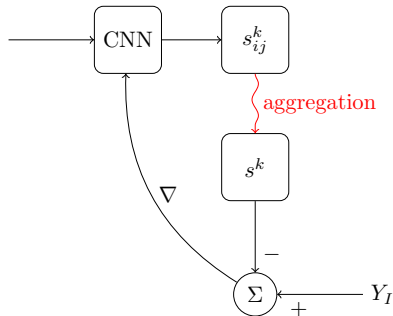
CNN backpropagation

feature map pixels: $s_{i,j}^k$

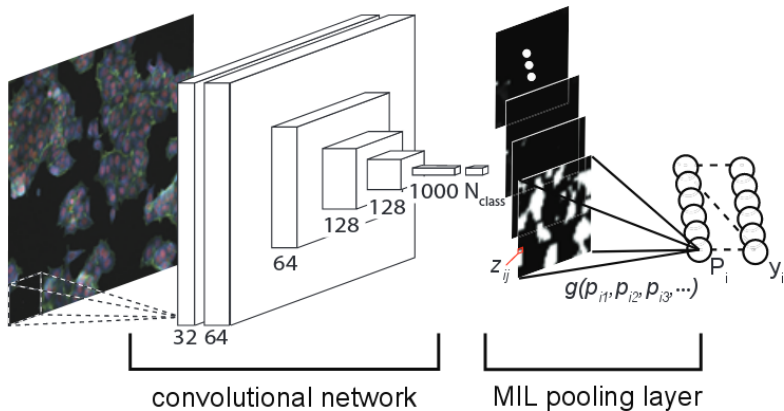


MIL-CNN

need to aggregate $s_{i,j}^k$



MIL-CNN



MIL-CNN Aggregation

- **global sum**

$$s^k = \sum_{i,j} s_{ij}^k$$

- **global max pooling**⁸

$$s^k = \max_{i,j} s_{ij}^k$$

- **LSE** (log-sum-exp)⁹

$$s^k = \frac{1}{r} \log \left[\frac{1}{wh} \sum \exp(r \cdot s_{ij}^k) \right]$$

⁸Oquab et al., CVPR'2015

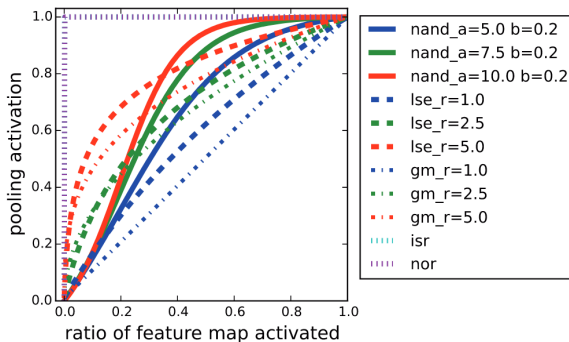
⁹Pinheiro et al., CVPR'2015

MIL-CNN Aggregation II

- **Noisy-AND**¹⁰

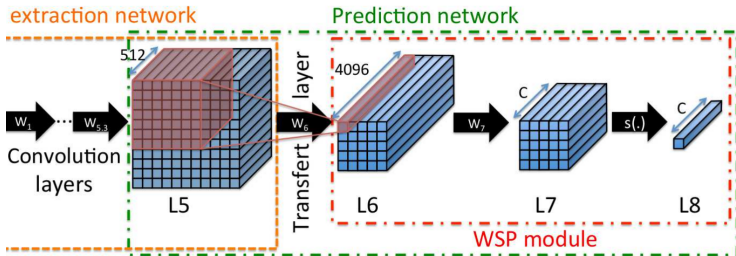
$$s = \frac{\sigma(a(\bar{s}_{ij} - b_i)) - \sigma(-ab_i)}{\sigma(a(1 - b_i)) - \sigma(-ab_i)}$$

activation once the mean of the instance probabilities \bar{s}_{ij} surpasses a threshold.



gMIL-CNN Aggregation

Two-stage network: *feature extraction + weakly-sup. module*¹¹



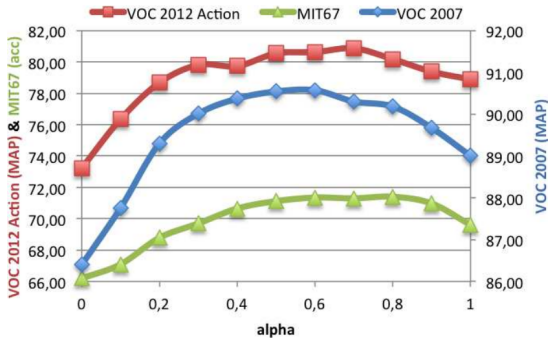
- select *top-k* and *m-low* instances

$$\mathbf{L8}(c) = s_{top}(\mathbf{L7}(c)) + s_{low}(\mathbf{L7}(c)) = \sum_{t_c^*=1}^k \mathbf{l}_{t_c^*}^7 + \sum_{l_c^*=1}^m \mathbf{l}_{l_c^*}^7$$

gMIL-CNN Aggregation

- weight the contribution of *low* instances¹²

$$s^c = \max_{\mathbf{h} \in \mathcal{H}_{k^+}} \frac{1}{k^+} \sum_{i,j} h_{i,j} \bar{z}_{i,j}^c + \alpha \left(\min_{\mathbf{h} \in \mathcal{H}_{k^-}} \frac{1}{k^-} \sum_{i,j} h_{i,j} \bar{z}_{i,j}^c \right)$$



Summary

- generalization approaches in both traditional and deep learners
- gMIL formulation better suited in multiple scenarios

gMIL - Addressed issues

- **robustness** – impact of a single false detection
- **weak labels** – multiple concepts are present
- **counting** – the number of positive instances is decisive

Thank you for your attention!

