Multiple-Instance Learning via Embedded Instance Selection

Yixin Chen
Department of Computer Science
University of New Orleans
http://www.cs.uno.edu/~yixin

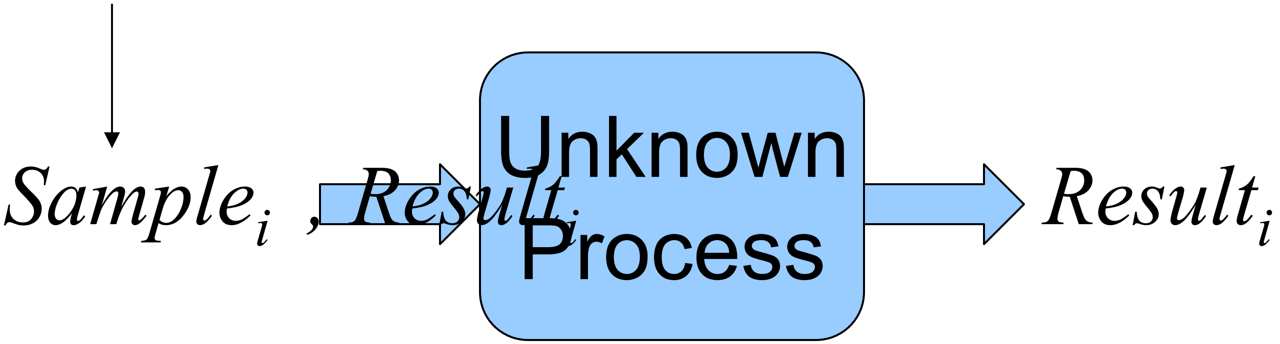
Outline

- An overview
- MIL via embedded instance selection
- Applications
  - Drug activity prediction
  - Human histological image classification
- Discussions

Supervised Learning

Fixed-length vector of attribute values

(usually called a "feature vector")



$$Result = f\,(Sample)$$

Classification problem: if $Result$ is discrete or categorical

Regression problem: if $Result$ is continuous

Multiple-Instance Learning Problem

Each instance is a fixed-length feature vector

$$Sample_i = Result_i, \begin{pmatrix} Instance_1 \\ Instance_2 \\ Instance_3 \\ \vdots \\ Instance_n \end{pmatrix} \Rightarrow \boxed{\text{Unknown Process}} \Rightarrow Label_i \; Result_i$$

Bag_i

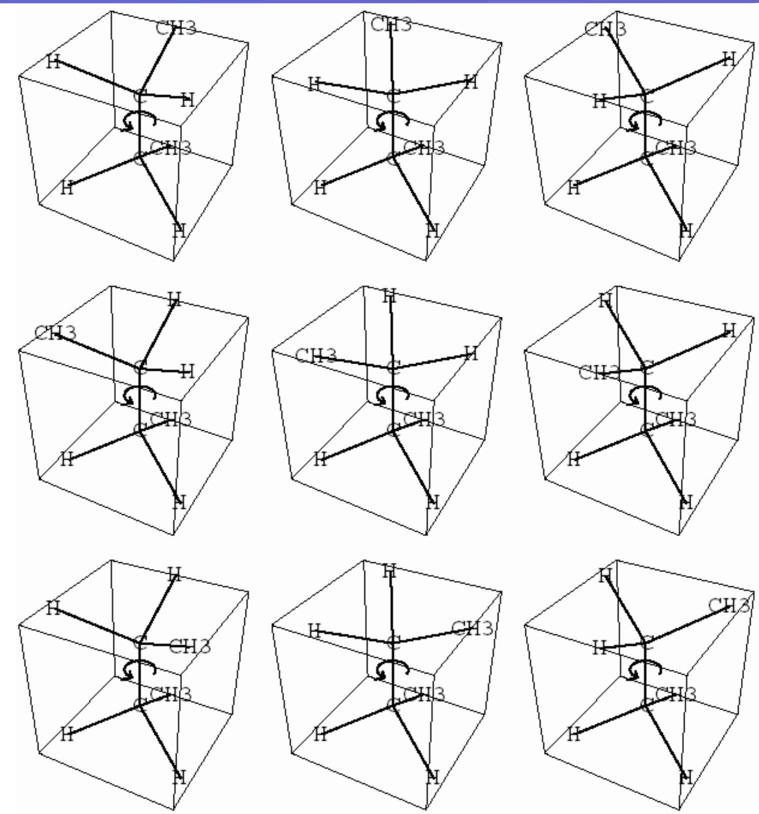
discrete or continuous

A label is associated with a bag, not the instances in the bag

Drug Activity Prediction

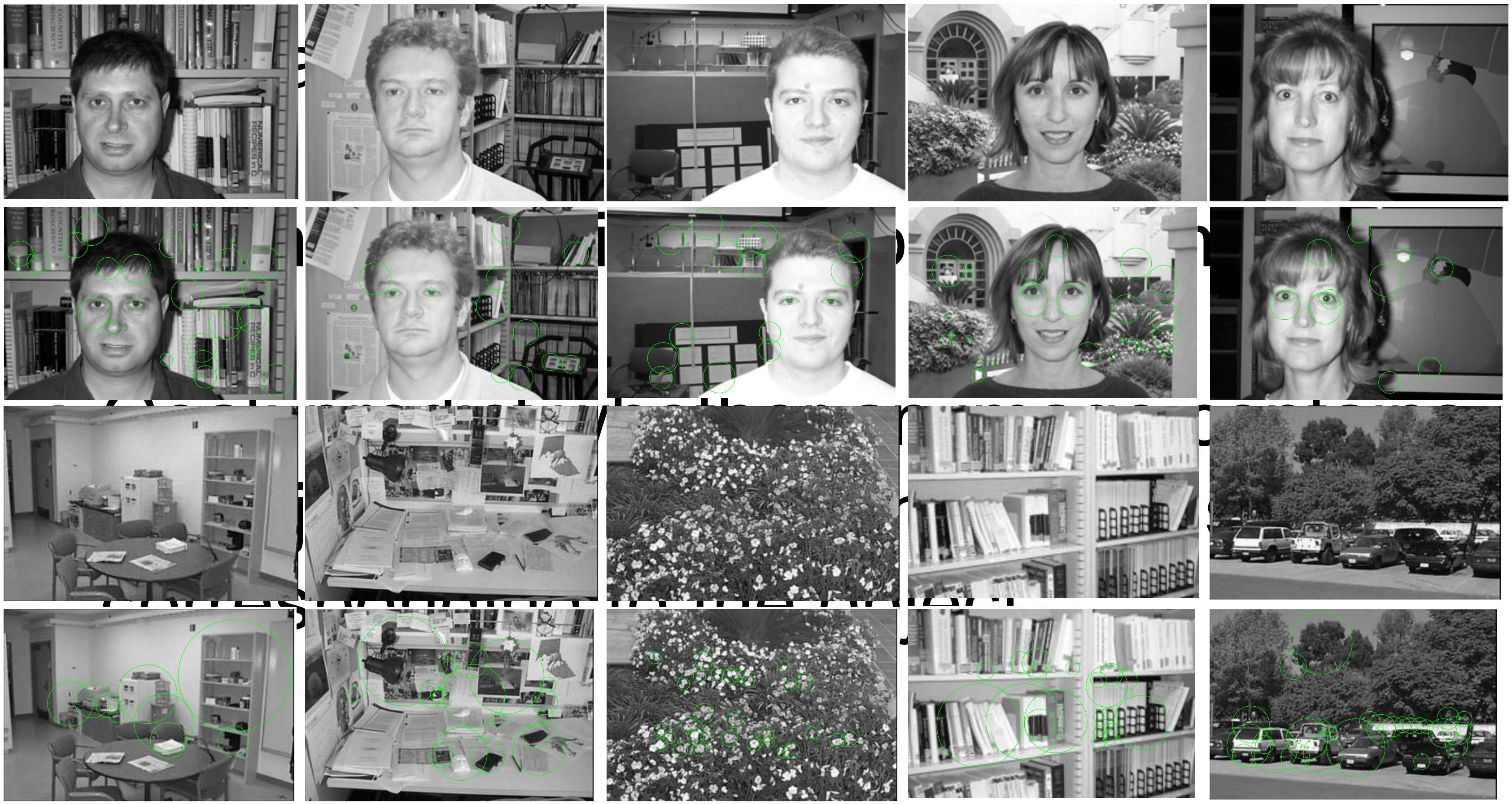
- Predict whether a candidate drug molecule will bind strongly to a target protein
- Binding strength is largely determined by the shape of drug molecules

- Bag: a molecule
- Instance: a shape conformation
- Goal: predict whether a molecule binds to a protein of interest, and find the conformation that binds



Different conformations a Butane molecule (C_4H_{10}) can take on. The molecule can rotate about the bond between the two central carbon atoms. (© 1998 by Oded Maron)

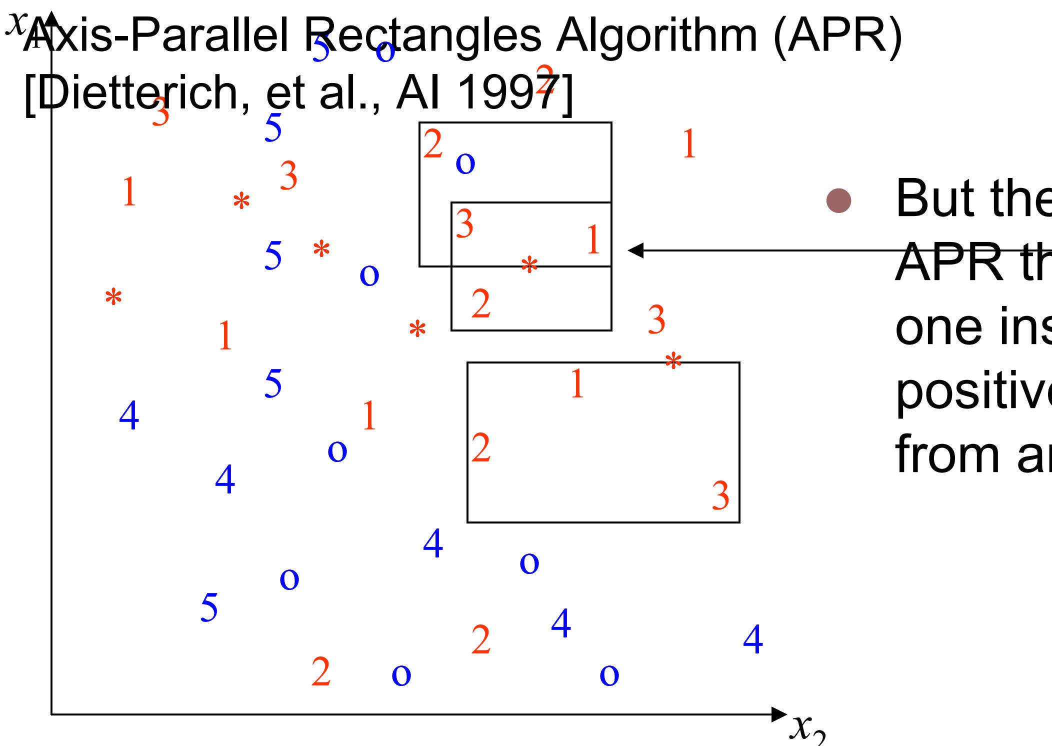
Object Recognition



Multiple-Instance Learning Models

- A bag is positive if and only if it contains at least one positive instance

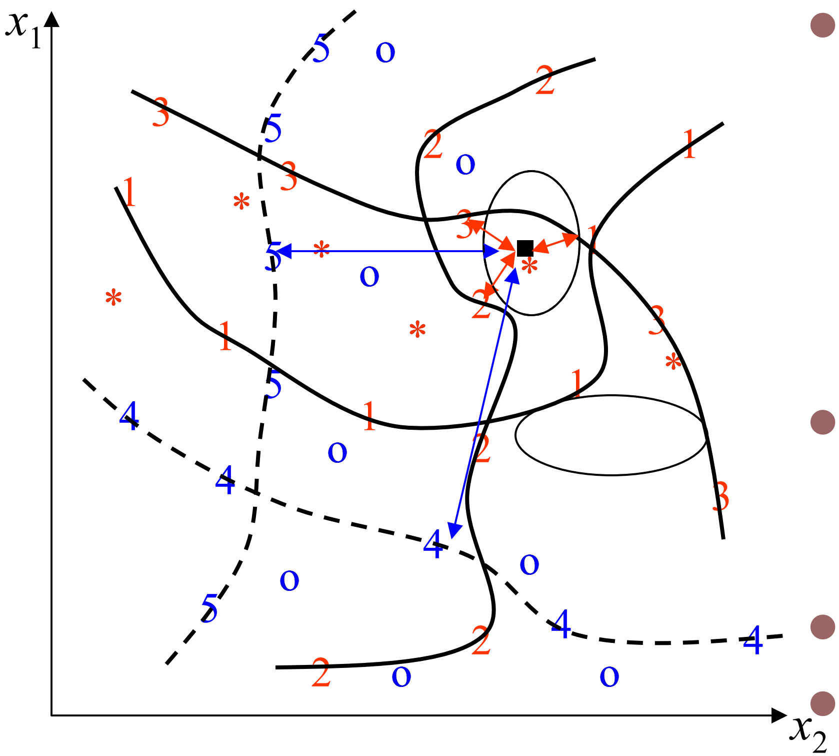
Axis-Parallel Rectangles Algorithm (APR)
[Dietterich, et al., AI 1997]



- But there may not exist an APR that contains at least one instance from each positive bag and no instance from any negative bags

Multiple-Instance Learning Models

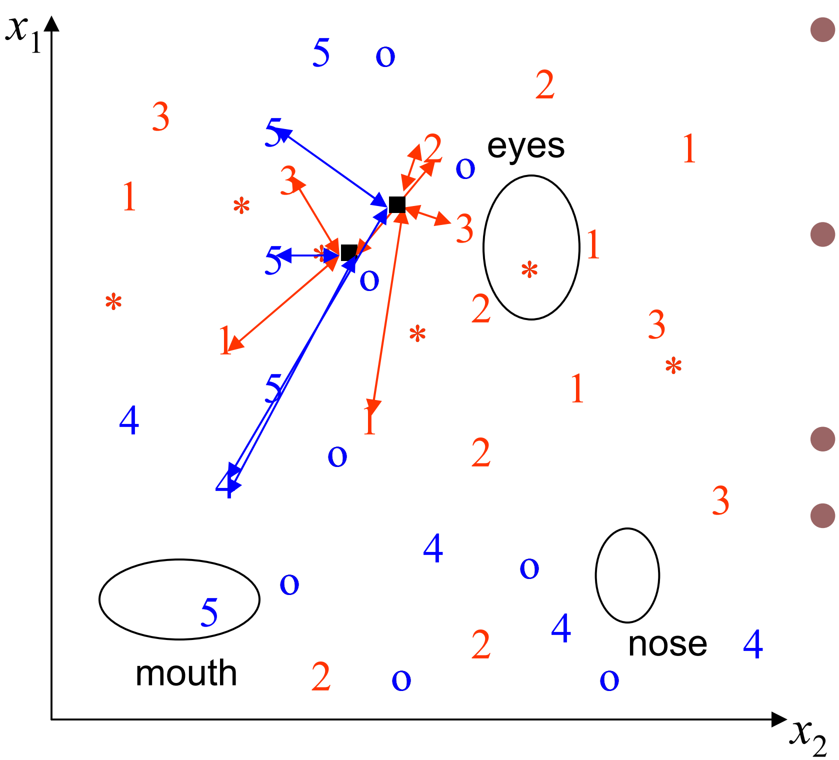
Diverse Density Algorithm (DD) [Maron and Lozano-Pérez, NIPS 1998]



- The diverse density at a location is high if the location is close to instances from different positive bags and is far way from all instances in negative bags
- Searching for an “axis-parallel ellipse” with high diverse density
- Sensitive to noise
- High computational cost

Multiple-Instance Learning Models

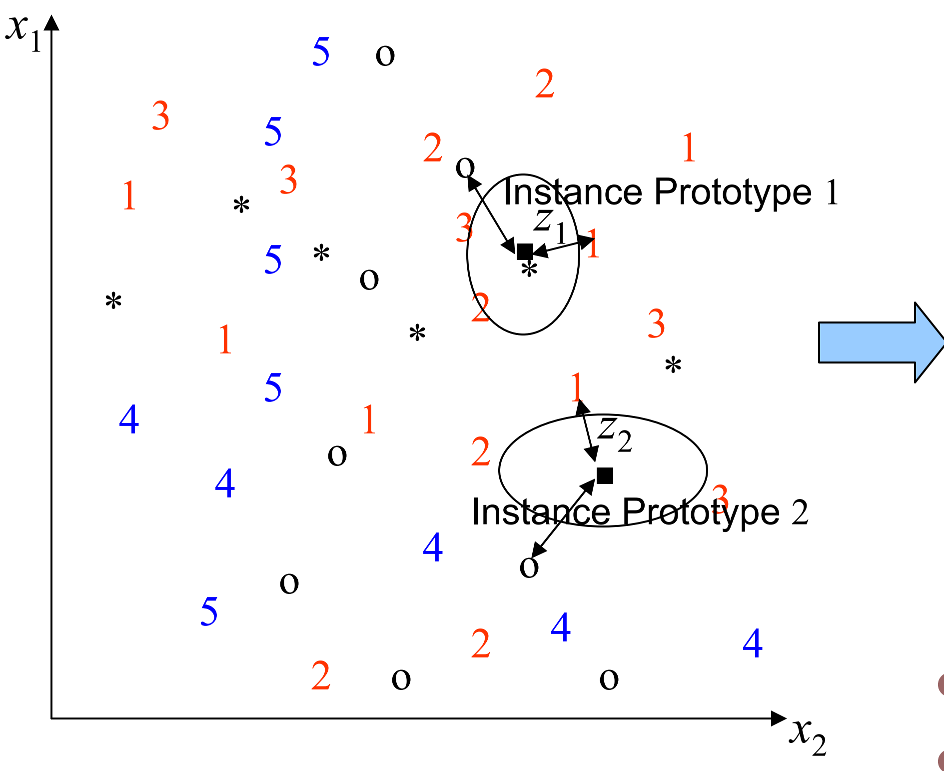
EM-DD Algorithm [Zhang and Goldman, NIPS 2001]

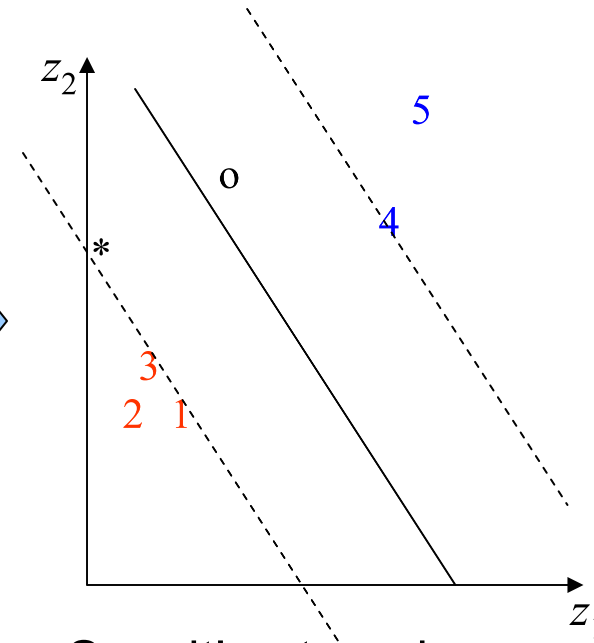


- The diverse density is approximated by the “most likely” instance in each bag
- Finding an “axis-parallel ellipse” with high diverse density
- Sensitive to noise
- Cannot learn complex concepts

Multiple-Instance Learning Models

DD-SVM Algorithm [Chen and Wang, JMLR 2004]



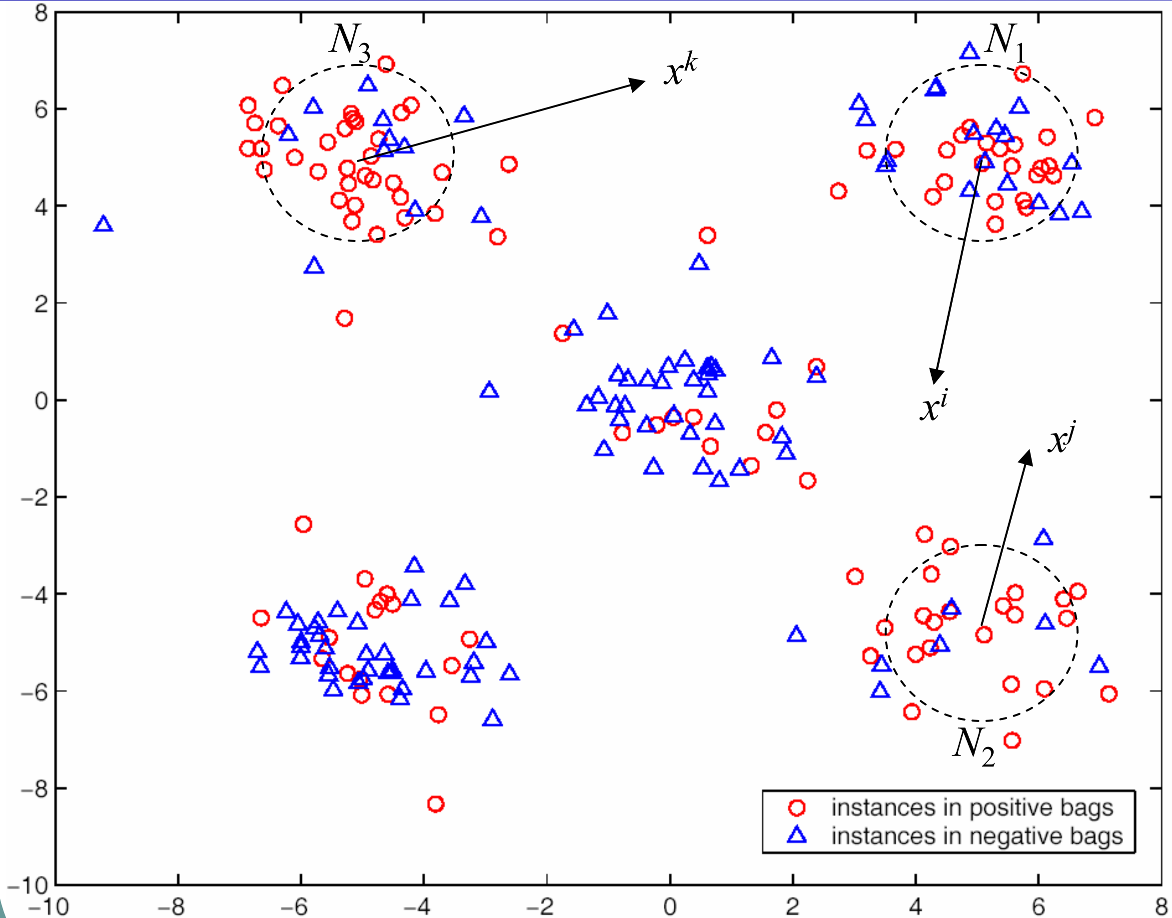


- Sensitive to noise
- Computational cost
- Instance classification

Outline

- An overview
- MIL via embedded instance selection

Motivation



20 positive bags, 20 negative bags

$N_1 \sim \mathcal{N}([5,5]^T, I)$

$N_2 \sim \mathcal{N}([5,-5]^T, I)$

$N_3 \sim \mathcal{N}([-5,5]^T, I)$

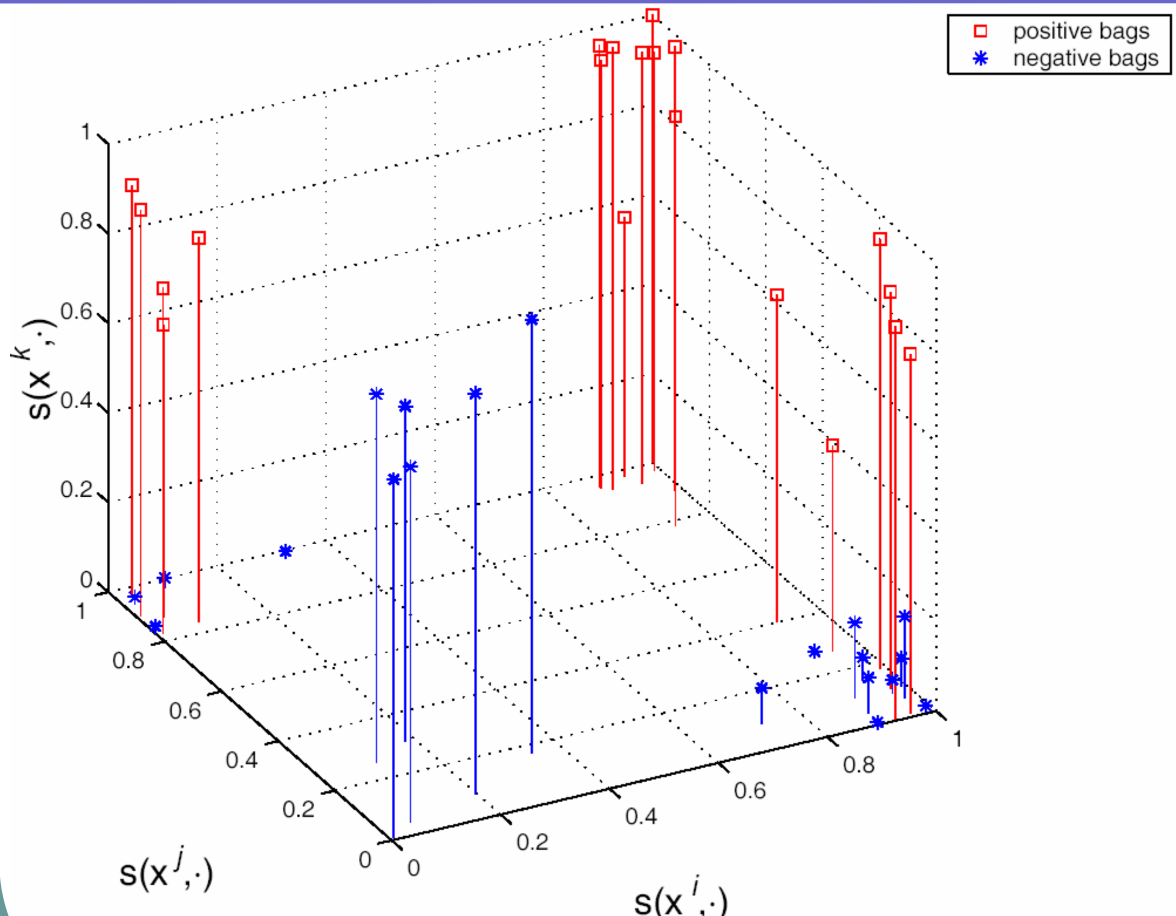
$N_4 \sim \mathcal{N}([-5,-5]^T, I)$

$N_5 \sim \mathcal{N}([0,0]^T, I)$

A bag is positive if it contains instances from at least two different distributions among N_1, N_2, and N_3

$$s(\mathbf{x}^k, \mathbf{B}_i) = \max_j \exp\left(-\frac{\|\mathbf{x}_{ij} - \mathbf{x}^k\|^2}{\sigma^2}\right)$$

Motivation



20 positive bags and 20 negative bags in the new feature space

- Embedding of bags
- Bags can be separated by a hyperplane
- Find the “right” embedding and the classifier

MILES: Multiple-Instance Learning via Embedded Instance Selection

- Instance-based feature mapping

$$s(\mathbf{x}^k, \mathbf{B}_i) = \max_j \ \exp\left(-\frac{\|\mathbf{x}_{ij} - \mathbf{x}^k\|^2}{\sigma^2}\right)$$

$$\mathbf{m}(\mathbf{B}_i) = \left[s(\mathbf{x}^1, \mathbf{B}_i), s(\mathbf{x}^2, \mathbf{B}_i), \cdots, s(\mathbf{x}^n, \mathbf{B}_i)\right]^T$$

- Joint feature selection and classification

$$y = \text{sign}\left(\mathbf{w}^T\mathbf{m} + b\right)$$

Minimizing a regularized training error

$$\lambda P[\cdot] + error \longrightarrow \text{1-norm SVM}$$

1-norm of $\mathbf{w}$ Hinge loss function

Outline

- An overview
- MIL via embedded instance selection
- Applications
  - Drug activity prediction
  - Human histological image classification

Drug Activity Prediction

- MUSK1 and MUSK2 benchmark data sets
  - A bag represents a molecule
  - An instance represents a low-energy conformation of the molecule (166 features)

	# of bags	# of instances/ bag	# of positive bags
Musk 1	92	5.17	47
Musk 2	102	64.69	39

Prediction Accuracy

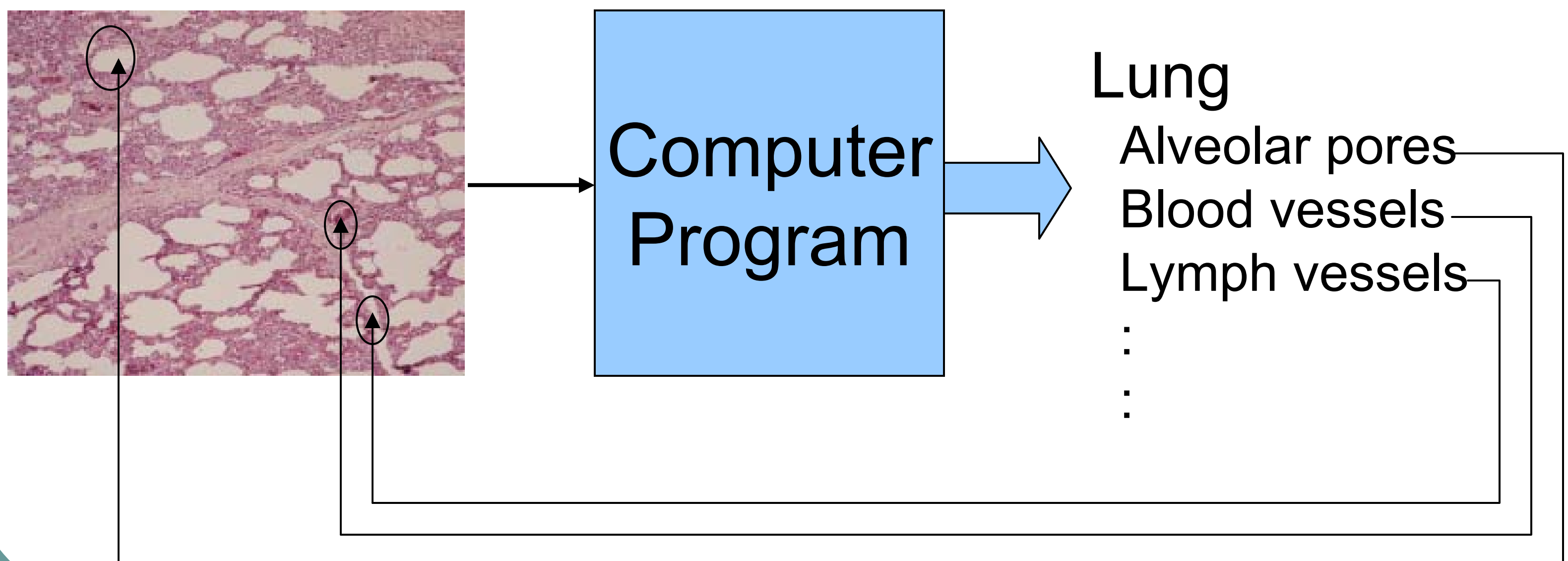
Algorithms	MUSK1	MUSK2	Type of Testing
MILES	86.3 : [84.9, 87.7]	87.7 : [86.3, 89.1]	10-fold cross-validation
	87.0	**93.1**	Leave-one-out test
APR [18]	92.4	89.2	10-fold cross-validation
Bagging-APR [57]	**92.8**	**93.1**	10-fold cross-validation
Bayesian-kNN [49]	90.2	82.4	Leave-one-out Test
Citation-kNN [49]	92.4	86.3	Leave-one-out Test
DD [33]	88.9	82.5	10-fold cross-validation
DD-SVM [16]	85.8	91.3	10-fold cross-validation
EM-DD [56]	84.8	84.9	10-fold cross-validation
mi-SVM [2]	87.4	83.6	10-fold cross-validation
MI-SVM [2]	77.9	84.3	10-fold cross-validation
MI-NN [42]	88.0	82.0	10-fold cross-validation
Multinst [4]	76.7 : [73.6, 79.8]	84.0 : [81.4, 86.6]	10-fold cross-validation
RELIC [44]	83.7	87.3	10-fold cross-validation

Computation Time

- Training time
  - SunFire V800z, Solaris, P4 1.9GHz CPU
  - 10 fold cross-validation
  - MILES: 6 seconds (MUSK1), 72 seconds (MUSK2)
  - DD-SVM: 500 minutes (MUSK1), 1500 minutes (MUSK2)

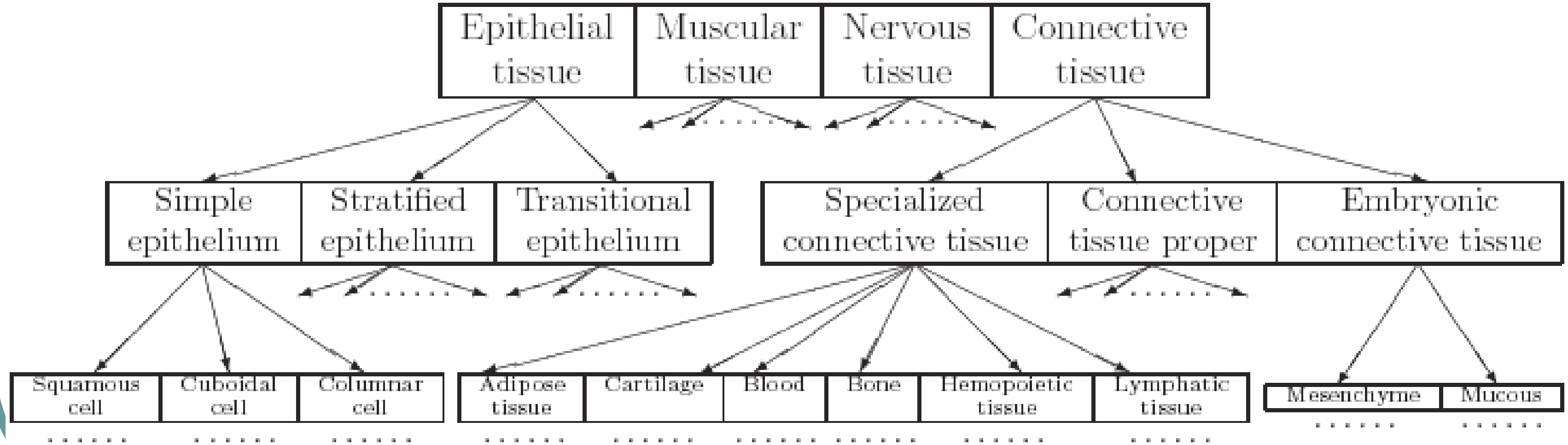
Histological Image Classification

- Why do we choose histological images
- Automatic interpretation of histological images



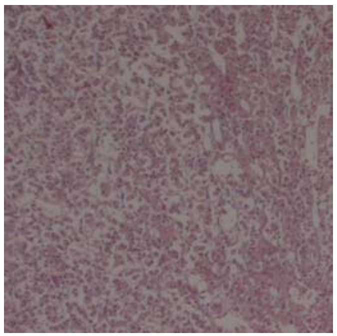
Overview

- Two research problems
  - Classification
    - Identifying the organ or part of the body
  - Annotation

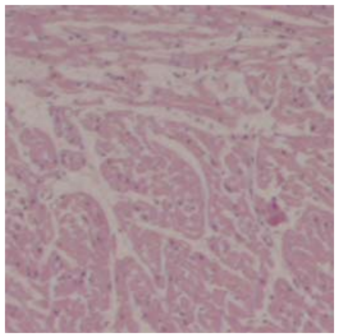


Extracting Imagery Features

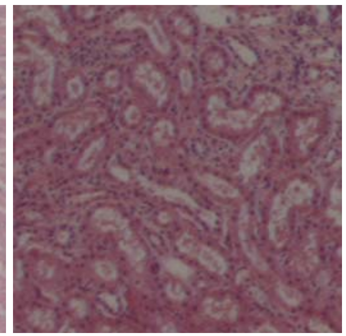
- What features to look for?



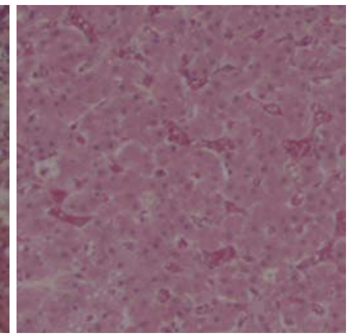
Adrenals



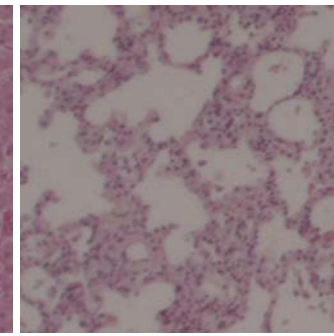
Heart



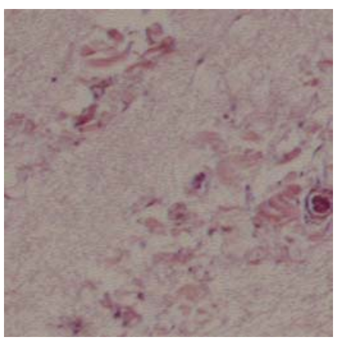
Kidney



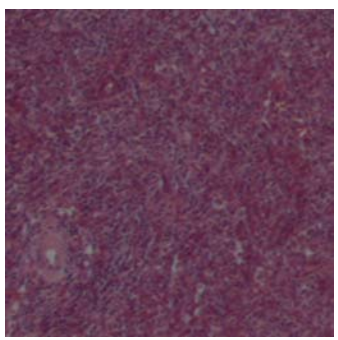
Liver



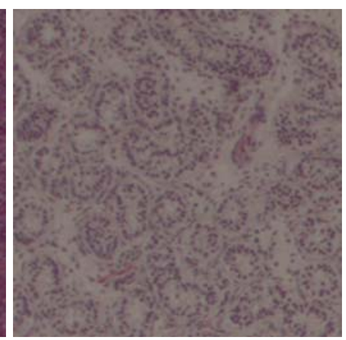
Lung



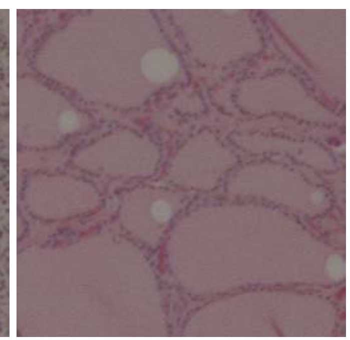
Pancreas



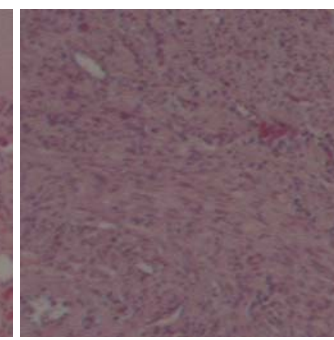
Spleen



Testis



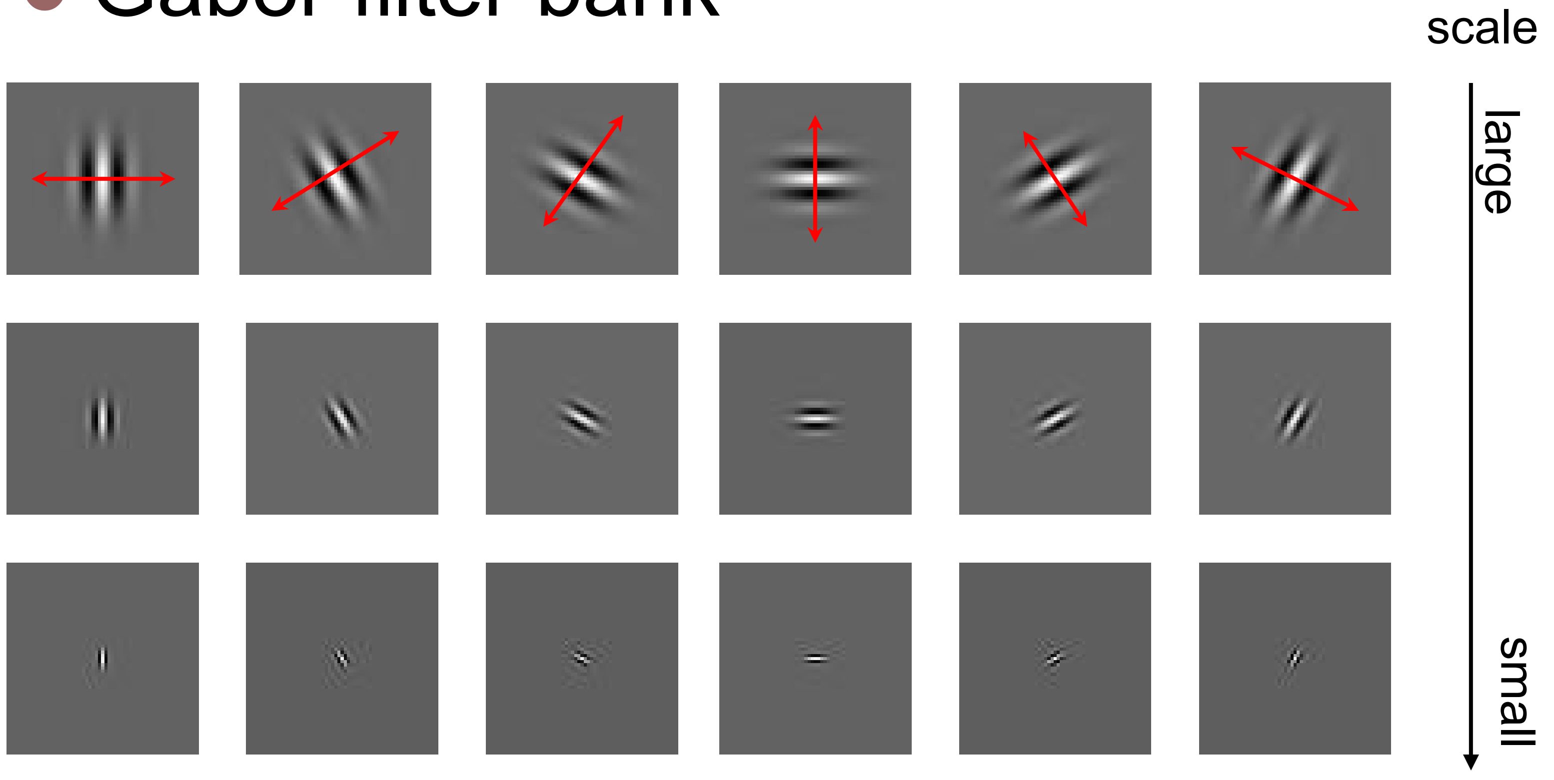
Thyroid



Uterus

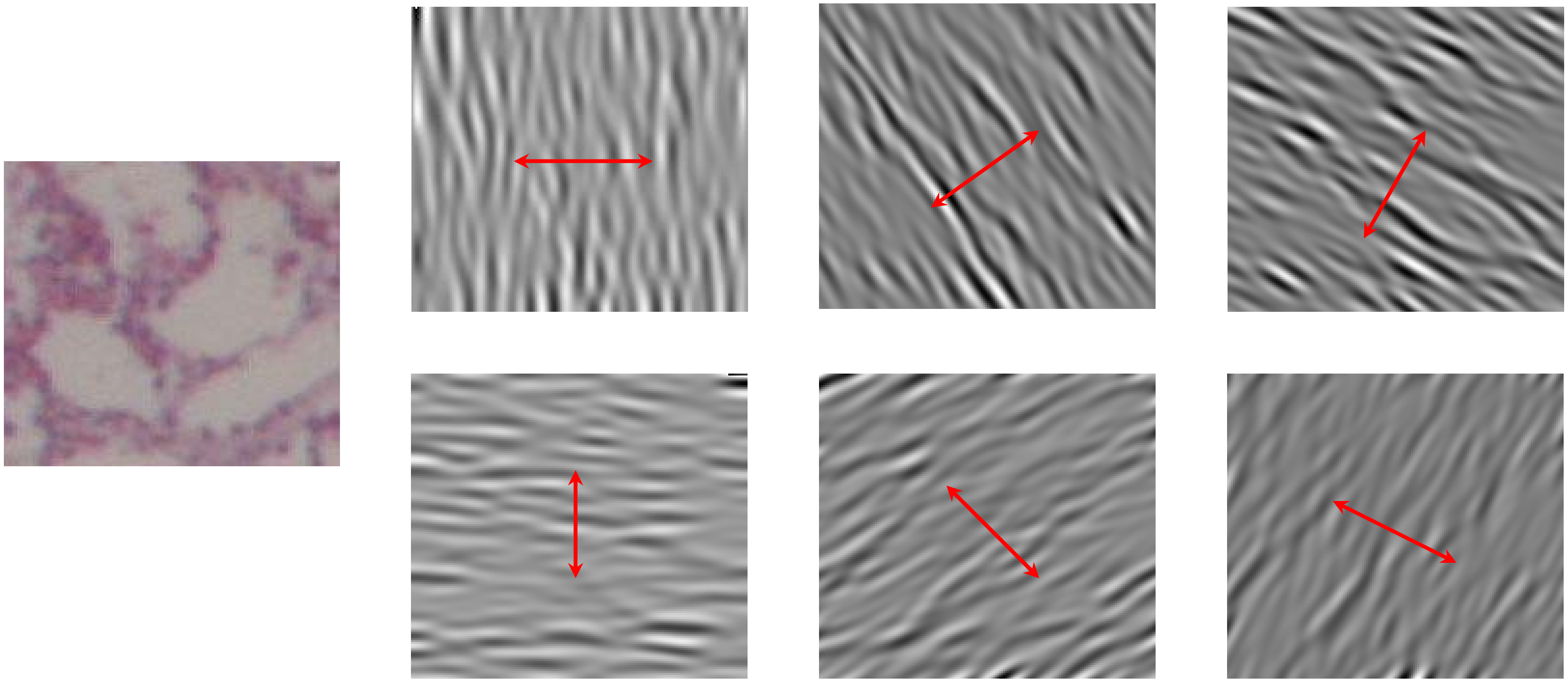
Extracting Imagery Features

- Gabor filter bank



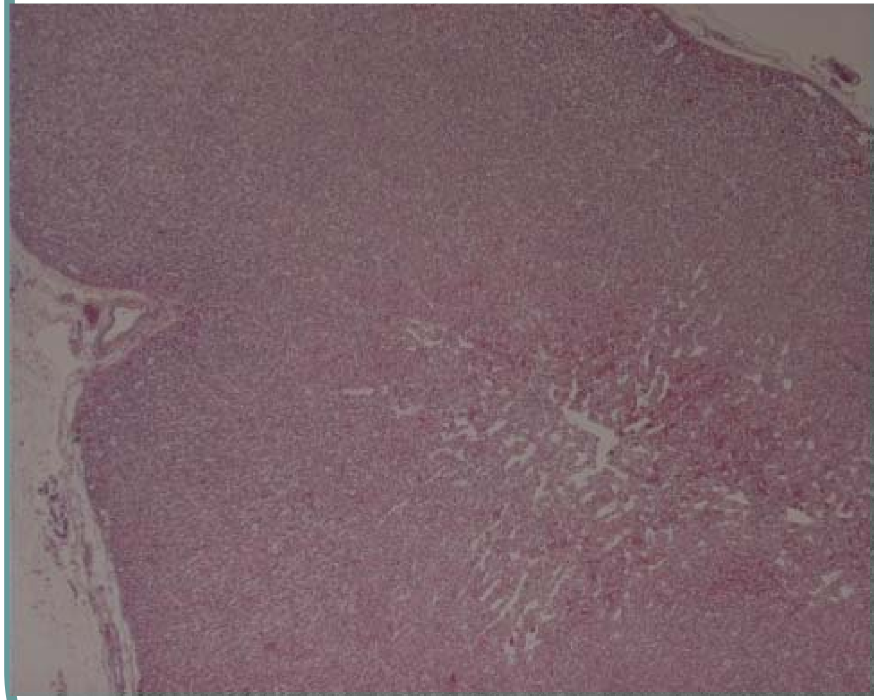
Extracting Imagery Features

- Example

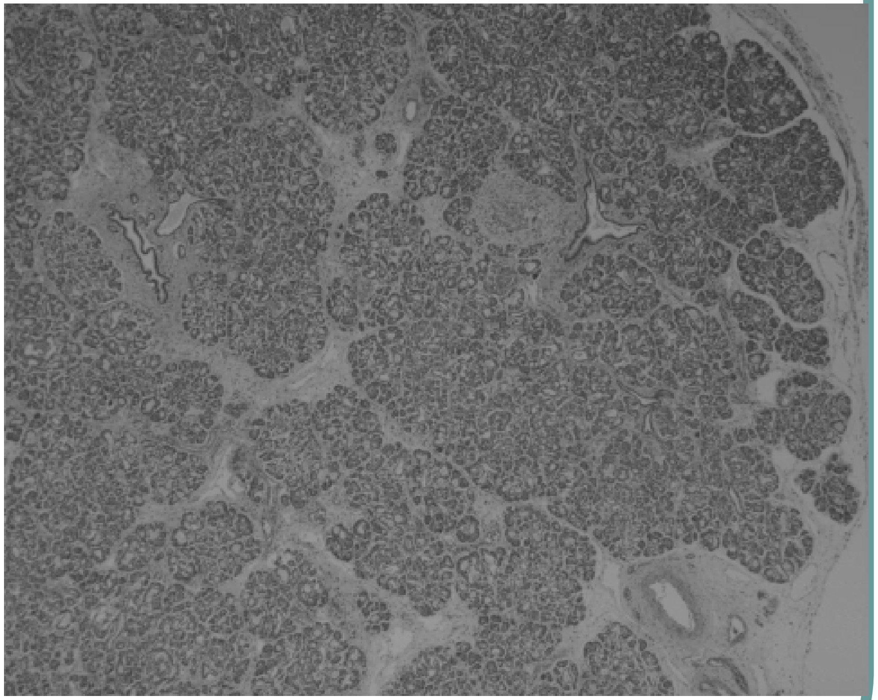


Extracting Imagery Features

- Color or grayscale?



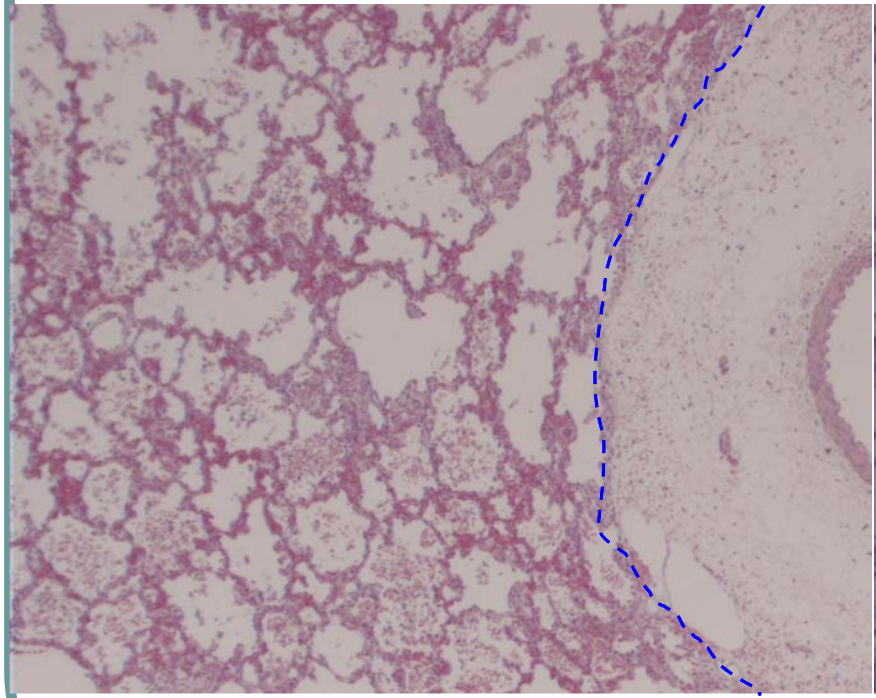
Adrenal (H&E staining)



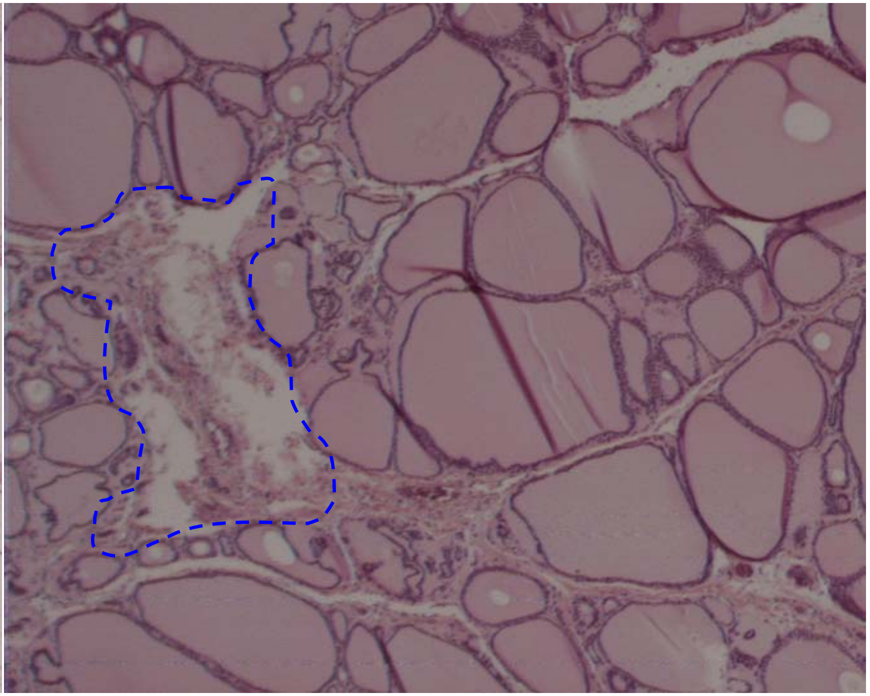
Pancreas (H&E staining)

Extracting Imagery Features

- Texture inhomogeneity

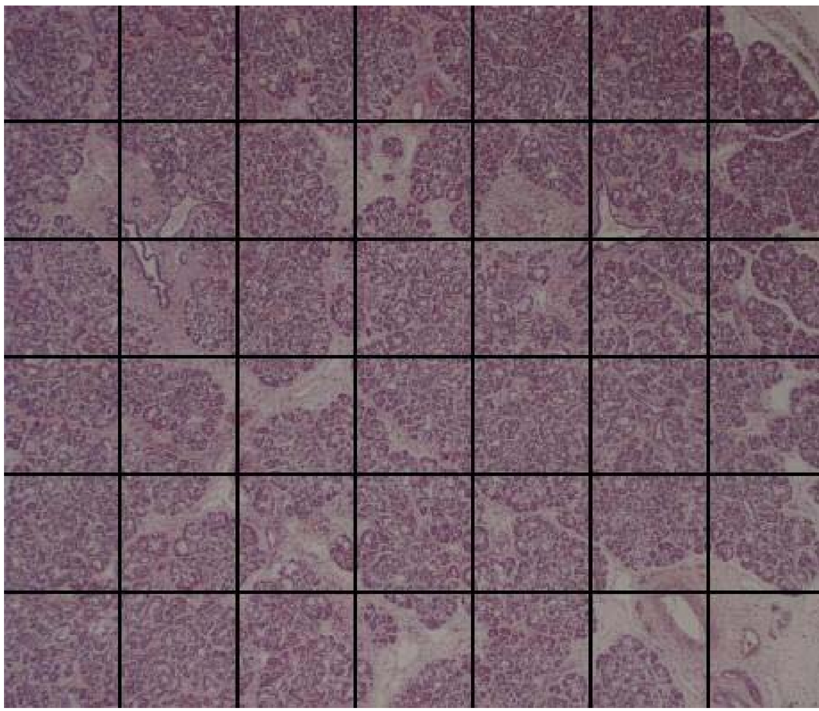


Lung



Thyroid

Extracting Imagery Features



3x6 Gabor Filter Bank

$$\left\{ \vec{f}_i \in R^{36} : i = 1,\ldots,42 \right\}$$

6x7 blocks

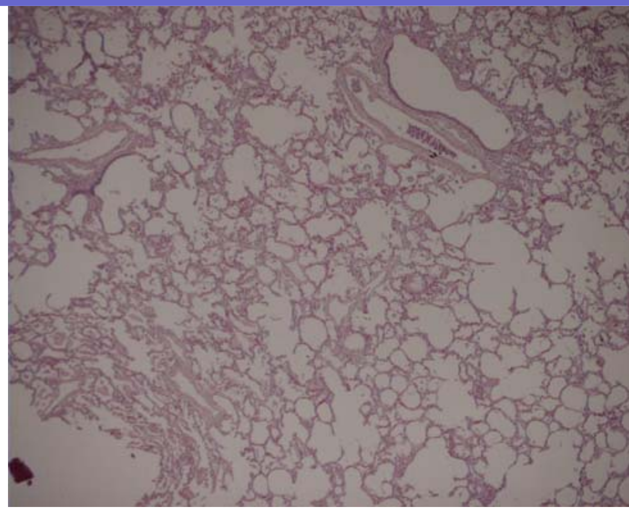
Experimental Evaluation

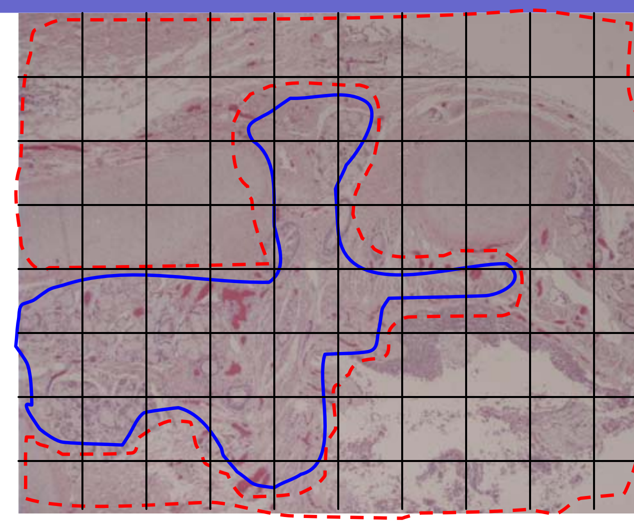
- H&E stained, 40x
- 3112 images
- Size 1536x1920
- Block size 64x64
  720 blocks
- Gabor filter bank

Category ID	Category Name	Number of Images
C1	Adrenals	100
C2	Heart	465
C3	Kidney	80
C4	Liver	428
C5	Lung	1152
C6	Pancreas	480
C7	Spleen	72
C8	Testis	100
C9	Thyroid	156
C10	Uterus	88

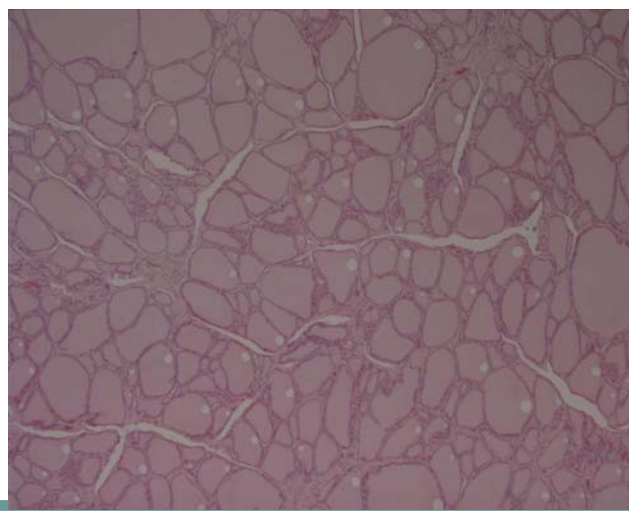
Multiple-Instance Problem

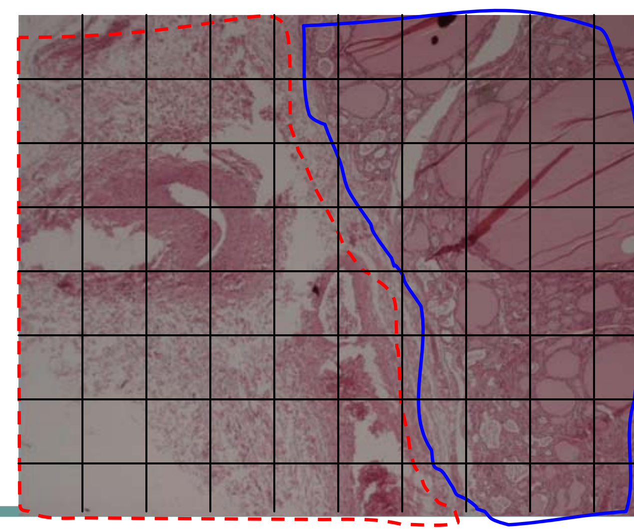
Lung





Thyroid





Performance of MILES

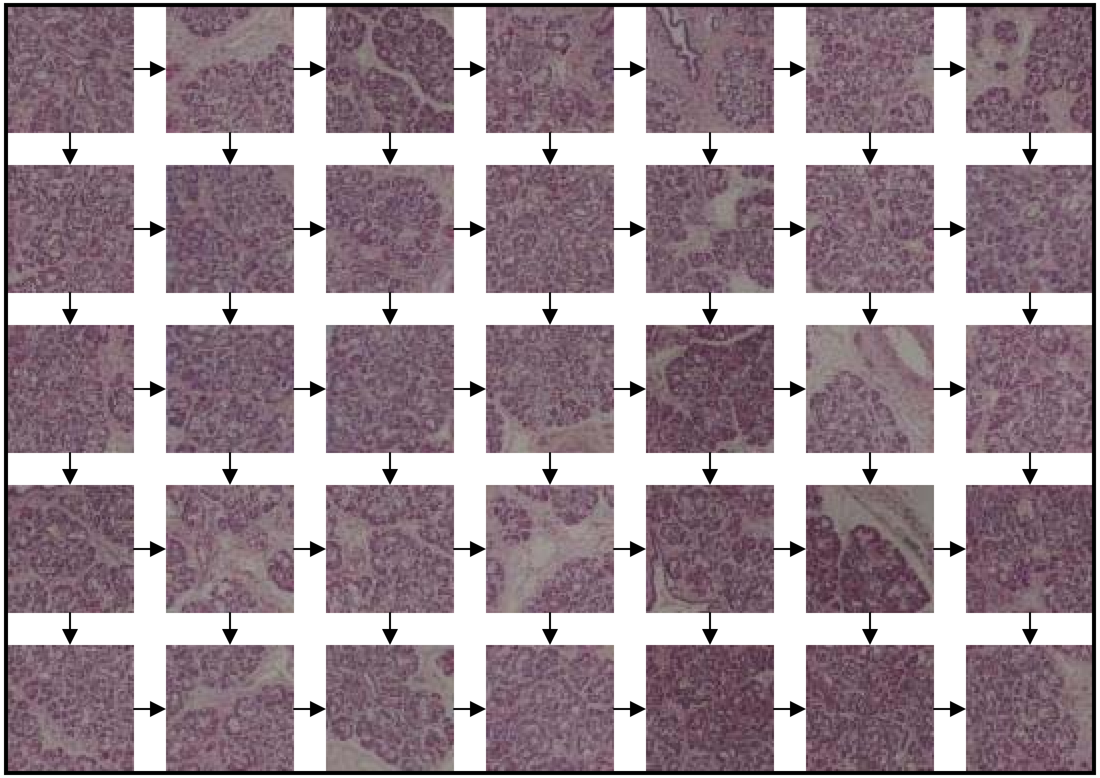
- 5-fold cross validation
  - MILES: 77.8%
  - A simple generative model: 71.5%
- Training time
  - SunFire V800z, Solaris, P4 1.9GHz CPU
  - Generative model: ≈ 5~6 hours per class
  - MILES: ≈ 0.5 hour per class

Outline

- An overview
- MIL via embedded instance selection
- Applications
  - Drug activity prediction
  - Human histological image classification
- Discussions

Future Work

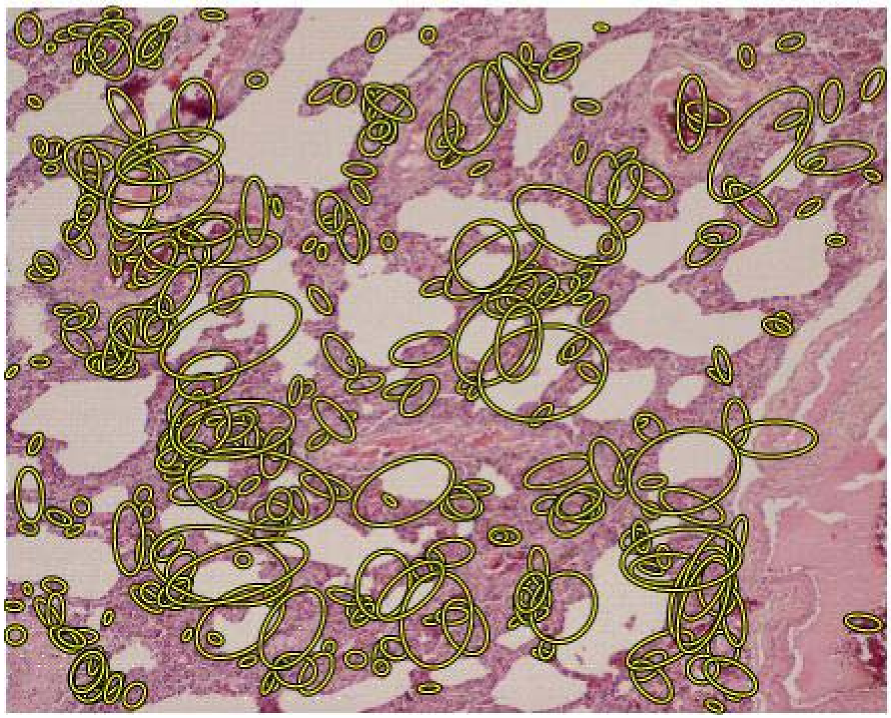
- Spatial information



Two-dimensional Hidden Markov Model

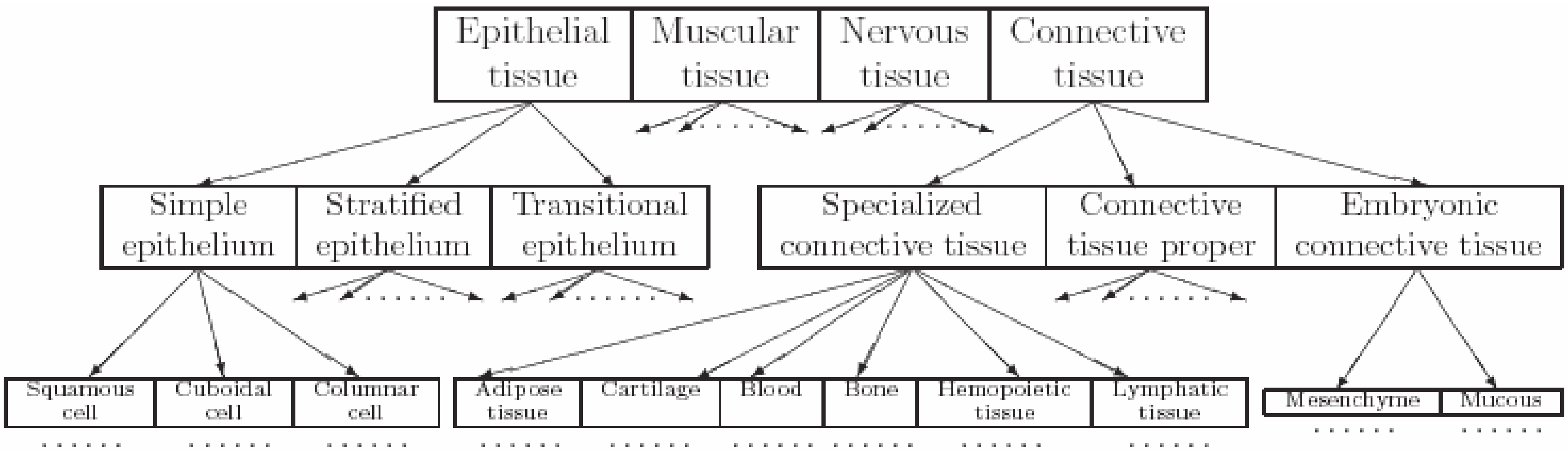
Future Work

- Bag generators
  - Affine invariant regions
  - Image segmentation



Future Work

- A larger scale
- Annotation



- Pathology images
- Content-based image retrieval

Future Work

- Storage requirement
  - A data matrix of size $(\ell^+ + \ell^-) \times n$
  - Sparseness
- MIL in a 1-class setting
  - Protein interaction inference

Supported by

- Louisiana Board of Regents RCS Grant
- NSF EPSCoR Pilot Fund
- The Research Institute for Children
- University of New Orleans

Acknowledgement

- Hernan Correa, LSUHSC
- Dehua Zhao, University of New Orleans
- Jinbo Bi, Siemens Medical Solutions
- James Z. Wang, The Pennsylvania State University
- Ya Zhang, The University of Kansas

More Information

- Papers in PDF, demonstrations, data sets, etc.

  http://www.cs.uno.edu/~yixin

  yixin@cs.uno.edu