Multi-Level Active Prediction of Useful Image Annotations

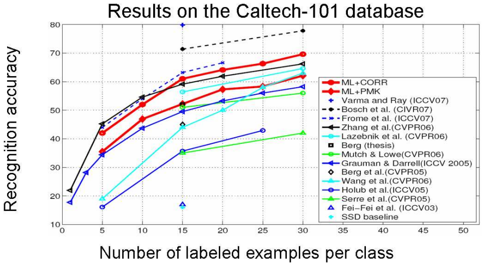
Sudheendra Vijayanarasimhan and Kristen Grauman

Department of Computer Sciences
University of Texas at Austin
Austin, Texas 78712

(svnaras,grauman)@cs.utexas.edu

Introduction

Visual category recognition is a vital thread in Computer Vision



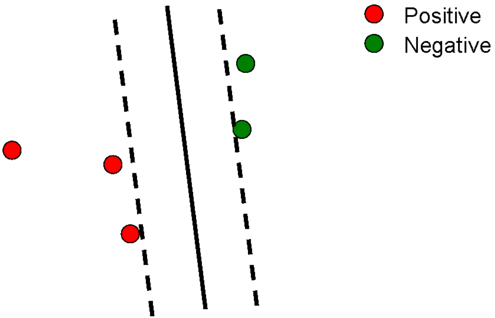
Often methods are most reliable when large training sets are available, but these are expensive to obtain.

Related Work

- Recent work considers various ways to reduce the amount of supervision required:
  - Weakly supervised category learning
    [Weber et al. 2000, Fergus et al. 2003]
  - Unsupervised category discovery
    [Sivic et al. 2005, Quelhas et al. 2005, Grauman & Darrell 2006, Liu & Chen 2006, Dueck & Frey 2007]
  - Share features, transfer learning
    [Murphy et al. 2003, Fei-Fei et al. 2003, Bart & Ullman 2005]
  - Leverage Web image search
    [Fergus et al. 2004, 2005, Li et al. 2007, Schroff et al. 2007, Vijayanarasimhan & Grauman 2008]
- Facilitate labeling process with good interfaces:
  - LabelMe [Russell et al. 2005]
  - Computer games [von Ahn & Dabbish 2004]
  - Distributed architectures [Steinbach et al. 2007]

Active Learning

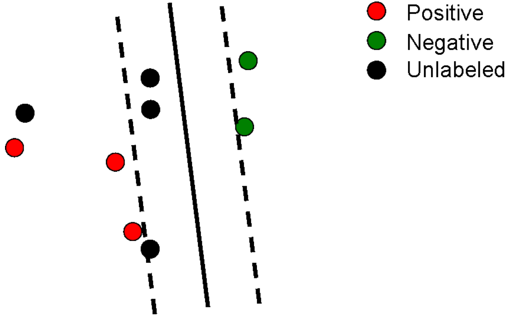
Traditional active learning reduces supervision by obtaining labels for the most informative or uncertain examples first.



[Mackay 1992, Freund et al. 1997, Tong & Koller 2001, Lindenbaum et al. 2004, Kapoor et al. 2007 ...]

Active Learning

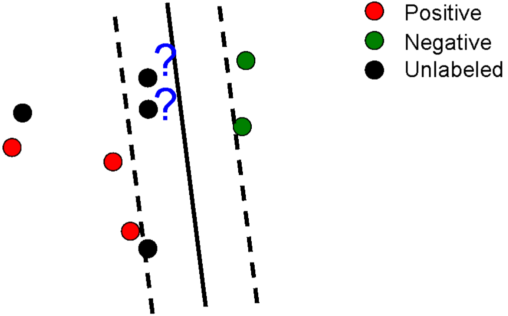
Traditional active learning reduces supervision by obtaining labels for the most informative or uncertain examples first.



[Mackay 1992, Freund et al. 1997, Tong & Koller 2001, Lindenbaum et al. 2004, Kapoor et al. 2007 ...]

Active Learning

Traditional active learning reduces supervision by obtaining labels for the most informative or uncertain examples first.



[Mackay 1992, Freund et al. 1997, Tong & Koller 2001, Lindenbaum et al. 2004, Kapoor et al. 2007 ...]

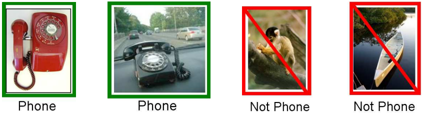
Problem

But in visual category learning, annotations can occur at multiple levels

Problem

But in visual category learning, annotations can occur at multiple levels

- **Weak labels:** informing about presence of an object



Problem

But in visual category learning, annotations can occur at multiple levels

- **Weak labels:** informing about presence of an object

Phone

Phone

Not Phone

Not Phone

- **Strong labels:** outlines demarking the object

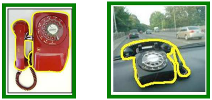
Problem

But in visual category learning, annotations can occur at multiple levels

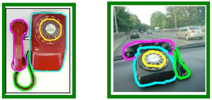
- **Weak labels:** informing about presence of an object

Phone Phone Not Phone Not Phone

- **Strong labels:** outlines demarking the object



- **Stronger labels:** informing about labels of parts of objects



Problem

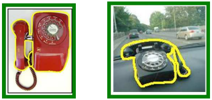
But in visual category learning, annotations can occur at multiple levels

Less expensive to obtain

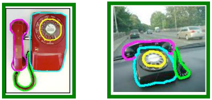
- **Weak labels:** informing about presence of an object

Phone Phone Not Phone Not Phone

- **Strong labels:** outlines demarking the object



- **Stronger labels:** informing about labels of parts of objects



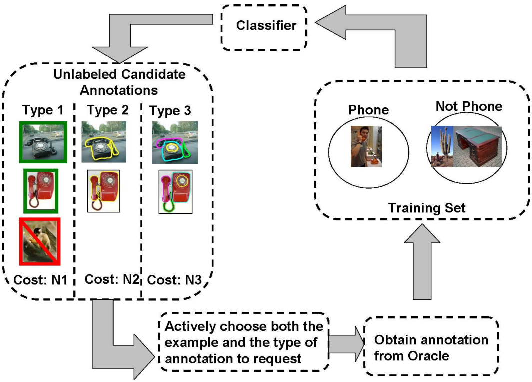
More expensive to obtain

Problem

- Strong labels provide unambiguous information but require more manual effort
- Weak labels are ambiguous but require little manual effort

How do we effectively learn from a mixture of strong and weak labels such that manual effort is reduced?

Approach: Multi-Level Active Visual Learning



- Best use of manual resources may call for combination of annotations at different levels.
- Choice must balance cost of varying annotations with their information gain.

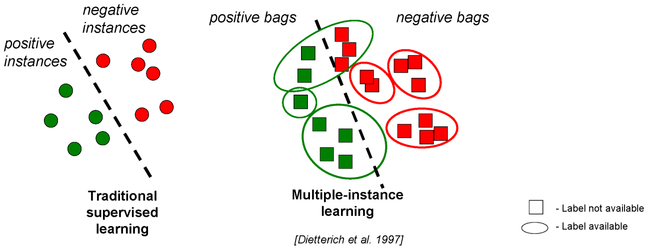
Requirements

The approach requires

- a classifier that can deal with annotations at multiple levels
- an active learning criterion to deal with
    - Multiple types of annotation queries
    - Variable cost associated with different queries

Multiple Instance learning (MIL)

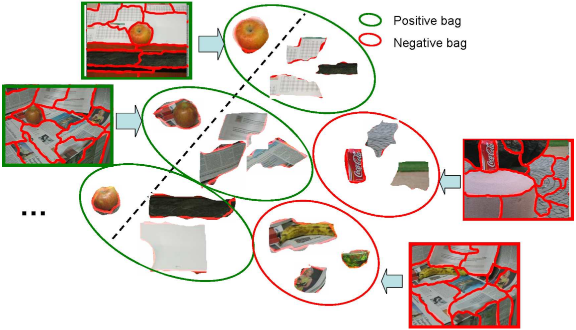
In MIL, training examples are sets (*bags*) of individual *instances*



- A *positive bag* contains at least one *positive instance*.
- A *negative bag* contains no *positive instances*.
- Labels on instances are not known.
- Learn to separate *positive bags/instances* from *negative instances*.

We use the SVM based MIL solution of Gartner et al. (2002).

MIL for visual category learning

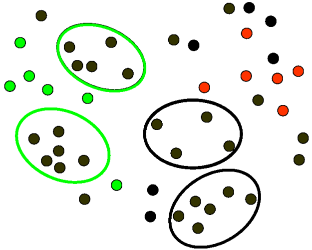


- **Postive instance:** Image segment belonging to class
- **Negative instance:** Image segment not in class
- **Positive bag:** Image containing class
- **Negative bag:** Image not containing class

[Zhang et al. (2002), Andrews et al. (2003) ...]

Multi-level Active Learning queries

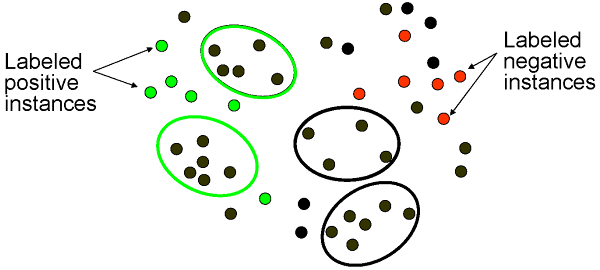
In MIL, an example can be



- **Strongly labeled:** Positive/Negative instances and Negative bags
- **Weakly Labeled:** Positive bags
- **Unlabeled:** Unlabeled instances and bags

Multi-level Active Learning queries

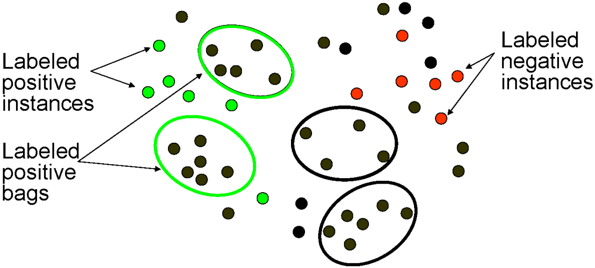
In MIL, an example can be



- **Strongly labeled:** Positive/Negative instances and Negative bags
- **Weakly Labeled:** Positive bags
- **Unlabeled:** Unlabeled instances and bags

Multi-level Active Learning queries

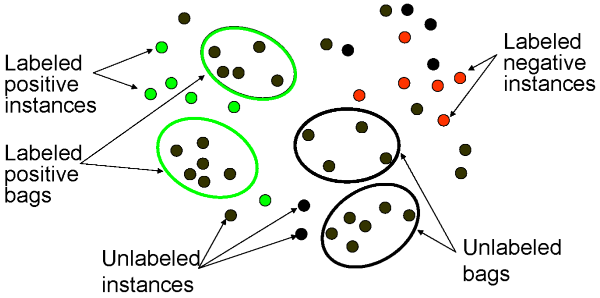
In MIL, an example can be



- **Strongly labeled:** Positive/Negative instances and Negative bags
- **Weakly Labeled:** Positive bags
- **Unlabeled:** Unlabeled instances and bags

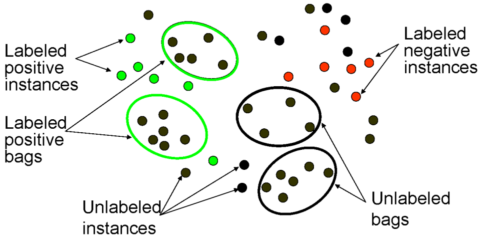
Multi-level Active Learning queries

In MIL, an example can be



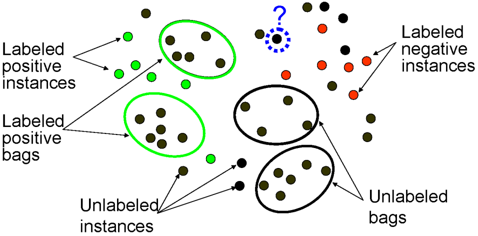
- **Strongly labeled:** Positive/Negative instances and Negative bags
- **Weakly Labeled:** Positive bags
- **Unlabeled:** Unlabeled instances and bags

Multi-level Active Learning queries



Types of queries active learner can pose

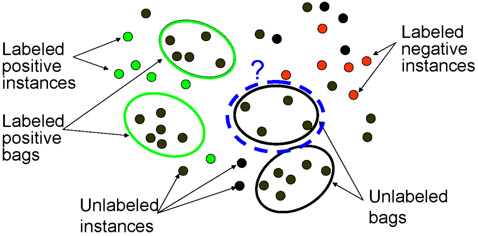
Multi-level Active Learning queries



Types of queries active learner can pose

- Label an unlabeled instance

Multi-level Active Learning queries

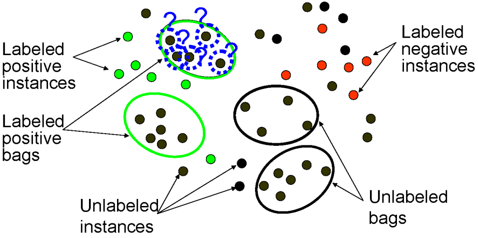


Types of queries active learner can pose

- Label an unlabeled instance

- Label an unlabeled bag

Multi-level Active Learning queries



Types of queries active learner can pose

- Label an unlabeled instance

- Label an unlabeled bag

- Label all instances within a positive bag

Possible Active Learning Strategies

- Disagreement among committee of classifiers

  [Freund et al. 1997]
- Margin-based with SVM

  [Tong & Koller 2001]
- Maximize expected information gain

  [Mackay 1992]
- Decision theoretic
  - Selective sampling [Lindenbaum et al. 2004]
  - **Value of Information [Kapoor et al. 2007]**

But all explored in the conventional single level learning setting

Decision-Theoretic Multi-level Criterion

Each candidate annotation $\mathbf{z}$ is associated with a Value of Information (VOI), defined as the total reduction in cost after annotation $\mathbf{z}$ is added to the labeled set.

$$VOI(\mathbf{z}) = T(\mathcal{X}_L, \mathcal{X}_U) - T\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}, \mathcal{X}_U \setminus \mathbf{z}\right)$$

Current dataset containing labeled examples $\mathcal{X}_L$ and unlabeled examples $\mathcal{X}_U$

Dataset after adding $\mathbf{z}$ with true label t to labeled set $\mathcal{X}_L$

$$T(\mathcal{X}_L, \mathcal{X}_U) = Risk(\mathcal{X}_L) + Risk(\mathcal{X}_U) + \sum_{X_i \in \mathcal{X}_L} \mathcal{C}(X_i)$$

Estimated risk of misclassifying labeled and unlabeled examples

Cost of obtaining labels for examples in the labeled set

Decision-Theoretic Multi-level Criterion

Each candidate annotation $\mathbf{z}$ is associated with a Value of Information (VOI), defined as the total reduction in cost after annotation $\mathbf{z}$ is added to the labeled set.

$$VOI(\mathbf{z}) = T(\mathcal{X}_L, \mathcal{X}_U) - T\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}, \mathcal{X}_U \setminus \mathbf{z}\right)$$

Current dataset containing labeled examples $\mathcal{X}_L$ and unlabeled examples $\mathcal{X}_U$

Dataset after adding $\mathbf{z}$ with true label t to labeled set $\mathcal{X}_L$

$$T(\mathcal{X}_L, \mathcal{X}_U) = Risk(\mathcal{X}_L) + Risk(\mathcal{X}_U) + \sum_{X_i \in \mathcal{X}_L} \mathcal{C}(X_i)$$

Estimated risk of misclassifying labeled and unlabeled examples

Cost of obtaining labels for examples in the labeled set

Decision-Theoretic Multi-level Criterion

Each candidate annotation $\mathbf{z}$ is associated with a Value of Information (VOI), defined as the total reduction in cost after annotation $\mathbf{z}$ is added to the labeled set.

$$VOI(\mathbf{z}) = T(\mathcal{X}_L, \mathcal{X}_U) - T\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}, \mathcal{X}_U \setminus \mathbf{z}\right)$$

Current dataset containing labeled examples $\mathcal{X}_L$ and unlabeled examples $\mathcal{X}_U$

Dataset after adding $\mathbf{z}$ with true label t to labeled set $\mathcal{X}_L$

$$T(\mathcal{X}_L, \mathcal{X}_U) = Risk(\mathcal{X}_L) + Risk(\mathcal{X}_U) + \sum_{X_i \in \mathcal{X}_L} \mathcal{C}(X_i)$$

Estimated risk of misclassifying labeled and unlabeled examples

Cost of obtaining labels for examples in the labeled set

Decision-Theoretic Multi-level Criterion

Each candidate annotation $\mathbf{z}$ is associated with a Value of Information (VOI), defined as the total reduction in cost after annotation $\mathbf{z}$ is added to the labeled set.

$$VOI(\mathbf{z}) = T(\mathcal{X}_L, \mathcal{X}_U) - T\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}, \mathcal{X}_U \setminus \mathbf{z}\right)$$

Current dataset containing labeled examples $\mathcal{X}_L$ and unlabeled examples $\mathcal{X}_U$

Dataset after adding $\mathbf{z}$ with true label t to labeled set $\mathcal{X}_L$

$$T(\mathcal{X}_L, \mathcal{X}_U) = Risk(\mathcal{X}_L) + Risk(\mathcal{X}_U) + \sum_{X_i \in \mathcal{X}_L} \mathcal{C}(X_i)$$

Estimated risk of misclassifying labeled and unlabeled examples

Cost of obtaining labels for examples in the labeled set

Decision-Theoretic Multi-level Criterion

Simplifying, the Value of Information for annotation **z** is

$$
\begin{aligned}
VOI(\mathbf{z}) &= T(\mathcal{X}_L, \mathcal{X}_U) - T\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}, \mathcal{X}_U \setminus \mathbf{z}\right) \\
&= R(\mathcal{X}_L) + R(\mathcal{X}_U) \\
&\quad - \left(R\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}\right) + R\left(\mathcal{X}_U \setminus \mathbf{z}\right)\right) \\
&\quad - \mathcal{C}(\mathbf{z})
\end{aligned}
$$

where R stands for Risk.

Risk of misclassifying examples using current classifier.

Risk of misclassifying examples after adding **z** to classifier.

Cost of obtaining annotation for **z**.

Decision-Theoretic Multi-level Criterion

Simplifying, the Value of Information for annotation $\mathbf{z}$ is

$$
\begin{aligned}
VOI(\mathbf{z}) &= T(\mathcal{X}_L, \mathcal{X}_U) - T\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}, \mathcal{X}_U \setminus \mathbf{z}\right) \\
&= \boxed{R(\mathcal{X}_L) + R(\mathcal{X}_U)} \\
&\quad - \left(R\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}\right) + R\left(\mathcal{X}_U \setminus \mathbf{z}\right)\right) \\
&\quad - \mathcal{C}(\mathbf{z})
\end{aligned}
$$

where R stands for Risk.

Risk of misclassifying examples using current classifier.

Risk of misclassifying examples after adding $\mathbf{z}$ to classifier.

Cost of obtaining annotation for $\mathbf{z}$.

Decision-Theoretic Multi-level Criterion

Simplifying, the Value of Information for annotation $\mathbf{z}$ is

$$
\begin{aligned}
VOI(\mathbf{z}) &= T(\mathcal{X}_L, \mathcal{X}_U) - T\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}, \mathcal{X}_U \setminus \mathbf{z}\right) \\
&= R(\mathcal{X}_L) + R(\mathcal{X}_U) \\
&\quad - \left(R\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}\right) + R\left(\mathcal{X}_U \setminus \mathbf{z}\right)\right) \\
&\quad - \mathcal{C}(\mathbf{z})
\end{aligned}
$$

where R stands for Risk.

Risk of misclassifying examples using current classifier.

Risk of misclassifying examples after adding $\mathbf{z}$ to classifier.

Cost of obtaining annotation for $\mathbf{z}$.

Decision-Theoretic Multi-level Criterion

Simplifying, the Value of Information for annotation **z** is

$$\begin{aligned} VOI(\mathbf{z}) &= T(\mathcal{X}_L, \mathcal{X}_U) - T\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}, \mathcal{X}_U \setminus \mathbf{z}\right) \\ &= R(\mathcal{X}_L) + R(\mathcal{X}_U) \\ &\quad - \left(R\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}\right) + R\left(\mathcal{X}_U \setminus \mathbf{z}\right)\right) \\ &\quad - \mathcal{C}(\mathbf{z}) \end{aligned}$$

where R stands for Risk.

Risk of misclassifying examples using current classifier.

Risk of misclassifying examples after adding **z** to classifier.

Cost of obtaining annotation for **z**.

Decision-Theoretic Multi-level Criterion: Risk

$$VOI(\mathbf{z}) \quad = \quad R(\mathcal{X}_L) + R(\mathcal{X}_U) - \left(R\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}\right) + R\left(\mathcal{X}_U \setminus \mathbf{z}\right)\right) - \mathcal{C}(\mathbf{z})$$

- **Labeled set** ($\mathcal{X}_L$): Consisting of positive *bags* $\mathcal{X}_p$ and negative *instances* $\mathcal{X}_n$

$$R(\mathcal{X}_L) = \sum_{X_i \in \mathcal{X}_p} r_p(1 - p(X_i)) + \sum_{x_i \in \mathcal{X}_n} r_n p(x_i),$$

Misclassification cost

Probability of misclassification

- **Unlabeled set** ($\mathcal{X}_U$):
  Similar expression for $R(\mathcal{X}_U)$, except that for unlabeled data the probability of labels must be estimated based on the current classifier's output.

Decision-Theoretic Multi-level Criterion: Risk

$$VOI(\mathbf{z}) \quad = \quad R(\mathcal{X}_L) + R(\mathcal{X}_U) - \left(R\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}\right) + R\left(\mathcal{X}_U \setminus \mathbf{z}\right)\right) - \mathcal{C}(\mathbf{z})$$

- **Labeled set** ($\mathcal{X}_L$): Consisting of positive *bags* $\mathcal{X}_p$ and negative *instances* $\mathcal{X}_n$

$$R(\mathcal{X}_L) = \sum_{X_i \in \mathcal{X}_p} \boxed{r_p}(1 - p(X_i)) + \sum_{x_i \in \mathcal{X}_n} \boxed{r_n} p(x_i),$$

Misclassification cost

Probability of misclassification

- **Unlabeled set** ($\mathcal{X}_U$):
  Similar expression for $R(\mathcal{X}_U)$, except that for unlabeled data the probability of labels must be estimated based on the current classifier's output.

Decision-Theoretic Multi-level Criterion: Risk

$$VOI(\mathbf{z}) \quad = \quad \boxed{R(\mathcal{X}_L) + R(\mathcal{X}_U)} - \left(R\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}\right) + R\left(\mathcal{X}_U \setminus \mathbf{z}\right)\right) - \mathcal{C}(\mathbf{z})$$

- **Labeled set** ($\mathcal{X}_L$): Consisting of positive *bags* $\mathcal{X}_p$ and negative *instances* $\mathcal{X}_n$

$$R(\mathcal{X}_L) = \sum_{X_i \in \mathcal{X}_p} \boxed{r_p} \boxed{(1 - p(X_i))} + \sum_{x_i \in \mathcal{X}_n} \boxed{r_n} \boxed{p(x_i)},$$

Misclassification cost

Probability of misclassification

- **Unlabeled set** ($\mathcal{X}_U$):
  Similar expression for $R(\mathcal{X}_U)$, except that for unlabeled data the probability of labels must be estimated based on the current classifier's output.

Decision-Theoretic Multi-level Criterion: Risk

$$VOI(\mathbf{z}) = \boxed{R(\mathcal{X}_L) + R(\mathcal{X}_U)} - \left(R\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}\right) + R\left(\mathcal{X}_U \setminus \mathbf{z}\right)\right) - \mathcal{C}(\mathbf{z})$$

- **Labeled set** ($\mathcal{X}_L$): Consisting of positive *bags* $\mathcal{X}_p$ and negative *instances* $\mathcal{X}_n$

$$R(\mathcal{X}_L) = \sum_{X_i \in \mathcal{X}_p} \boxed{r_p} \boxed{(1 - p(X_i))} + \sum_{x_i \in \mathcal{X}_n} \boxed{r_n} \boxed{p(x_i)},$$

Misclassification cost

Probability of misclassification

- **Unlabeled set** ($\mathcal{X}_U$):
  Similar expression for $R(\mathcal{X}_U)$, except that for unlabeled data the probability of labels must be estimated based on the current classifier's output.

Decision-Theoretic Multi-level Criterion: Expected Risk

$$VOI(\mathbf{z}) \quad = \quad R(\mathcal{X}_L) + R(\mathcal{X}_U) - \boxed{\left(R\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}\right) + R\left(\mathcal{X}_U \setminus \mathbf{z}\right)\right)} - \mathcal{C}(\mathbf{z})$$

Risk after adding annotation $\mathbf{z}$ is not directly computable since $\mathbf{z}$ is unlabeled.
We approximate this using the expected value of the risk:

$$\begin{aligned} R(\mathcal{X}_L \cup \mathbf{z}^{(t)}) + R(\mathcal{X}_U \setminus \mathbf{z}) &\approx E[R(\mathcal{X}_L \cup \mathbf{z}^{(t)}) + R(\mathcal{X}_U \setminus \mathbf{z})] \\ &= \mathbb{E} \end{aligned}$$

$$\mathbb{E} \quad = \quad \sum_{\ell \in \mathbb{L}} \left(R(\mathcal{X}_L \cup \mathbf{z}^{(\ell)}) + R(\mathcal{X}_U \setminus \mathbf{z}) \right) p(\ell | \mathbf{z})$$

$\mathbb{L}$ is the set of all possible labels that example $\mathbf{z}$ can take.

Decision-Theoretic Multi-level criterion: Expected Risk

$$VOI(\mathbf{z}) = R(\mathcal{X}_L) + R(\mathcal{X}_U) - \left(R\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}\right) + R\left(\mathcal{X}_U \setminus \mathbf{z}\right)\right) - \mathcal{C}(\mathbf{z})$$

- **if z is an unlabeled instance or bag:** $\mathbb{L} = \{+1, -1\}$

$$\begin{aligned} \mathbb{E} &= \left(R\left(\mathcal{X}_L \cup \mathbf{z}^{(+1)}\right) + R\left(\mathcal{X}_U \setminus \mathbf{z}\right)\right) p(\mathbf{z}) \\ &+ \left(R\left(\mathcal{X}_L \cup \mathbf{z}^{(-1)}\right) + R\left(\mathcal{X}_U \setminus \mathbf{z}\right)\right) (1 - p(\mathbf{z})) \end{aligned}$$

- $p(\mathbf{z})$ is obtained using a probabilistic for the SVM desicion value using a sigmoid function.

Decision-Theoretic Multi-level criterion: Expected Risk

$$VOI(\mathbf{z}) \quad = \quad R(\mathcal{X}_L) + R(\mathcal{X}_U) - \left(R\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}\right) + R\left(\mathcal{X}_U \setminus \mathbf{z}\right)\right) - \mathcal{C}(\mathbf{z})$$

- **if $\mathbf{z} = \{z_1, z_2, ...z_M\}$ is a positive bag:** $\mathbb{L} = \{+1, -1\}^M$

  We compute expected cost using Gibbs sampling:

Decision-Theoretic Multi-level criterion: Expected Risk

$$VOI(\mathbf{z}) \quad = \quad R(\mathcal{X}_L) + R(\mathcal{X}_U) - \left(R\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}\right) + R\left(\mathcal{X}_U \setminus \mathbf{z}\right)\right) - \mathcal{C}(\mathbf{z})$$

- **if $\mathbf{z} = \{z_1, z_2, ...z_M\}$ is a positive bag:** $\mathbb{L} = \{+1, -1\}^M$

  We compute expected cost using Gibbs sampling:

  - Starting with a random sample $I^1 = \{a_1^1, a_2^1...a_M^1\}$ we generate S samples from the joint distribution of the M instances

$$a_j^k \sim p(z_j | a_1^k, ...a_{j-1}^k, a_{j+1}^{k-1}, ...a_M^{k-1})$$

Decision-Theoretic Multi-level criterion: Expected Risk

$$VOI(\mathbf{z}) = R(\mathcal{X}_L) + R(\mathcal{X}_U) - \boxed{\left(R\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}\right) + R\left(\mathcal{X}_U \setminus \mathbf{z}\right)\right)} - \mathcal{C}(\mathbf{z})$$

- **if $\mathbf{z} = \{z_1, z_2, ... z_M\}$ is a positive bag:** $\mathbb{L} = \{+1, -1\}^M$

  We compute expected cost using Gibbs sampling:

  - Starting with a random sample $I^1 = \{a_1^1, a_2^1 ... a_M^1\}$ we generate S samples from the joint distribution of the M instances

    $$a_j^k \sim p(z_j | a_1^k, ... a_{j-1}^k, a_{j+1}^{k-1}, ... a_M^{k-1})$$

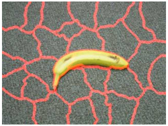
  - Compute expected value over the generated samples

$$\mathbb{E} = \frac{1}{S} \sum_{k=1}^{S} (R\left(\mathcal{X}_L \cup \{z_1^{(a_1)_k}, \ldots, z_M^{(a_M)_k}\}\right) + R\left(\mathcal{X}_U \setminus \{z_1, z_2, ..., z_M\}\right))$$

Decision-Theoretic Multi-level criterion: Cost

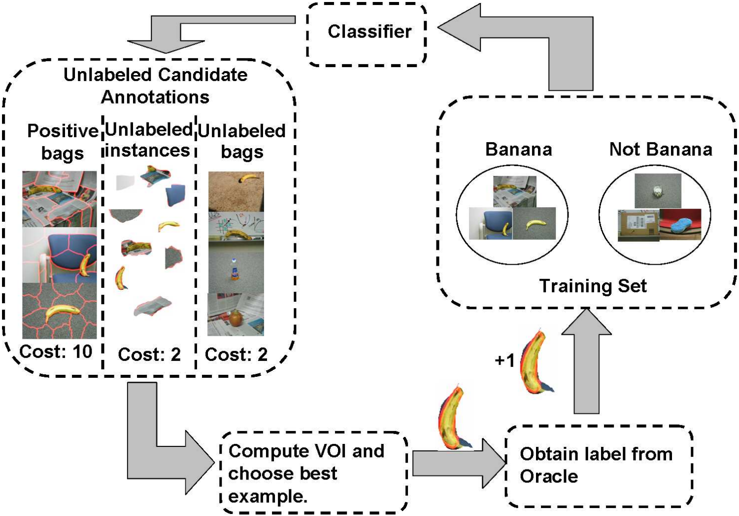
$$VOI(\mathbf{z}) = R(\mathcal{X}_L) + R(\mathcal{X}_U) - \left(R\left(\mathcal{X}_L \cup \mathbf{z}^{(t)}\right) + R\left(\mathcal{X}_U \setminus \mathbf{z}\right)\right) - \boxed{\mathcal{C}(\mathbf{z})}$$

User experiment to determine cost of each type of annotation.
Cost measured in terms of time required to obtain annotation.



Task	**Time (secs)**
click on all segments containing 'banana'	10
label a segment	2
label the image	2

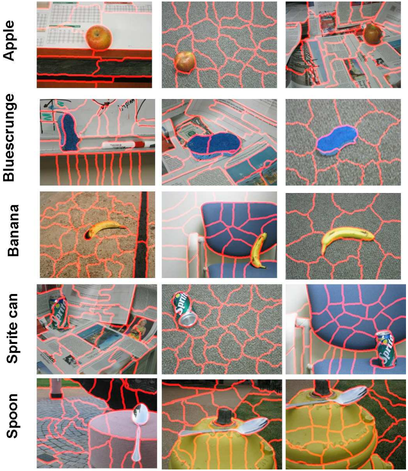
Summary of algorithm



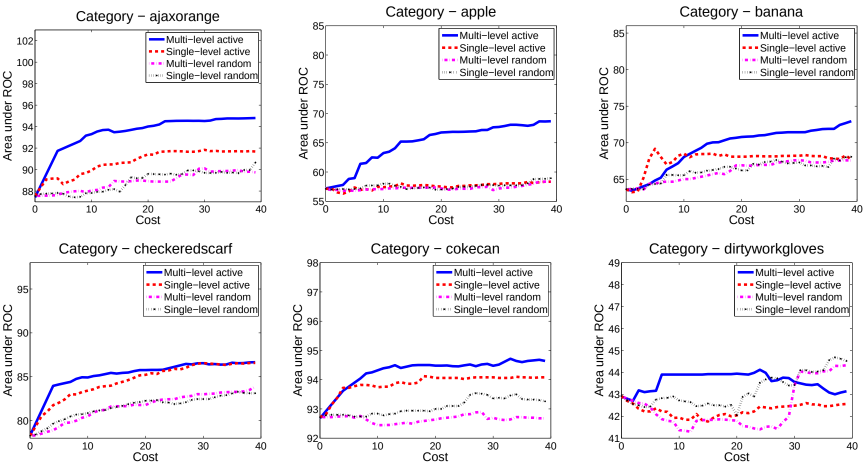
Results: SIVAL dataset

SIVAL dataset [Settles et al. 2008]

- 25 different classes
- 1500 images
- **Positive instance:** segment containing class
  Positive bag: image containing class
  Negative bag: images of all other classes
- Each segment represented by color and texture around 20-30 regions per image

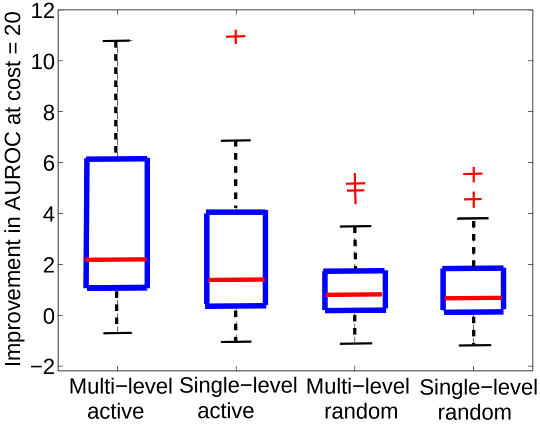


Results: SIVAL dataset



Sample learning curves per class, each averaged over five trials.
Multi-level active selection performs the best for most classes.

Results: SIVAL dataset



Summary of the average improvement over all categories at a cost of 20 units

Results: SIVAL dataset

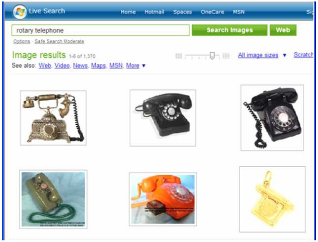
Cost	Gain over Random (%)	
	Our Approach	**[Settles et al.]**
10	372	117
20	176	112
50	81	52

Comparison with Settles et al. 2008 on the SIVAL data, as measured by the average improvement in the AUROC over the initial model for increasing labeling cost values.

Scenario 2: MIL for learning from keyword searches

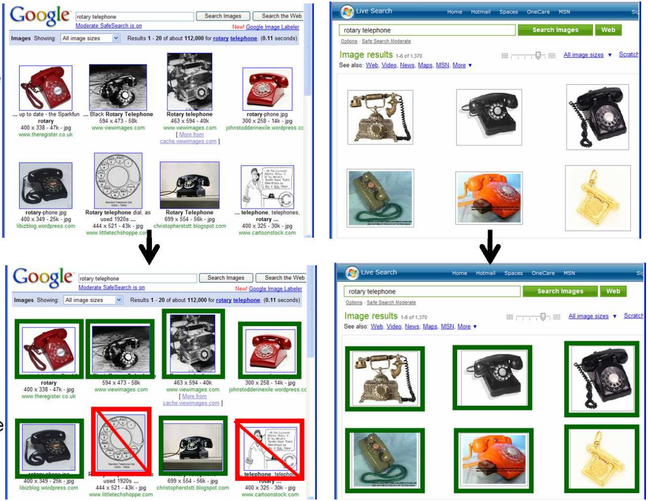
Less expensive to obtain

More expensive to obtain



Scenario 2: MIL for learning from keyword searches

Less expensive to obtain



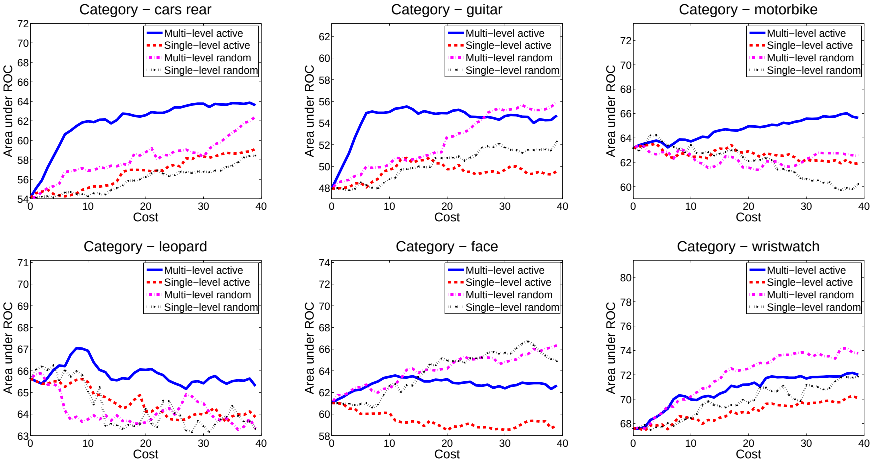
More expensive to obtain

Results: Google dataset

Google dataset [Fergus et al. 2005]

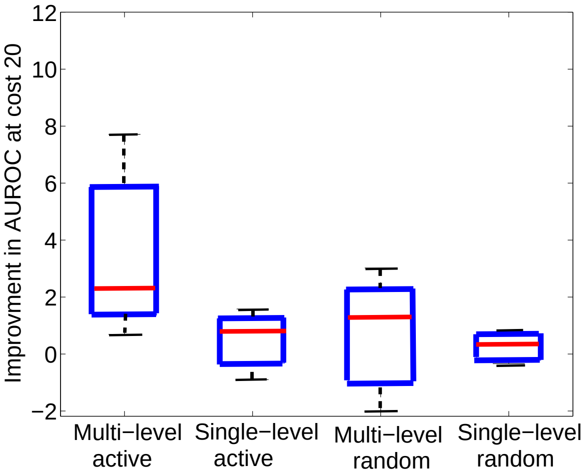
- 7 different classes
- 500-700 images per class
- **Positive instance:** image containing class
  Positive bag: set of images returned by keyword search for class
  Negative bag: images of all other classes
- Each image represented using bag of words of SIFT features on 4 different keypoints

Results: Google dataset



Learning curves for all categories in the Google dataset for the four methods.

Results: Google dataset



Summary of the average improvement over all categories at a cost of 20 units.

Conclusion

- First framework to actively learn from multi-level annotations.
- Compares different types of annotations using both information gain and cost of obtaining it.
- Results show that optimally choosing from multiple types of annotations reduces manual effort to learn accurate models.
- Applies to non-vision scenarios containing multi-level data.
  - like document classification (*bags:* documents, *instances:* passages)

Future Work

- Extend to multi-class setting.
- Reduce computational complexity.

MIL-SVM

The MIL problem can be solved using an SVM.

- Given an instance x described in some kernel embedding space as $\phi(x)$, a bag X is described by $\frac{\phi(X)}{|X|}$, where $\phi(X) = \sum_{x \in X} \phi(x)$ and $|X|$ counts the number of instances in the bag.
- This is the Normalized Set Kernel (NSK) of Gartner et al.
- Setup and solve a standard SVM using the above kernel function for bags.

$$
\begin{aligned}
\text{minimize:} \quad & \tfrac{1}{2}||w||^2 + \tfrac{C}{|\tilde{\mathcal{X}}_n|} \textstyle\sum_{x \in \tilde{\mathcal{X}}_n} \xi_x + \tfrac{C}{|\mathcal{X}_p|} \textstyle\sum_{X \in \mathcal{X}_p} \xi_X \\
\text{subject to:} \quad & w\ \phi(x) + b \leq -1 + \xi_x, \qquad \forall x \in \tilde{\mathcal{X}}_n \\
& w \tfrac{\phi(X)}{|X|} + b \geq +1 - \xi_X, \qquad \forall X \in \mathcal{X}_p \\
& \xi_x \geq 0, \xi_X \geq 0,
\end{aligned}
$$

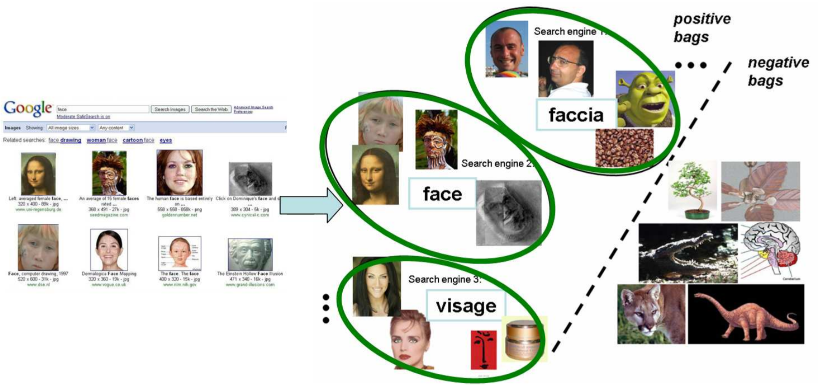
Expected Risk

- **Unlabeled set** ($\mathcal{X}_U$):
  Similar expression for $R(\mathcal{X}_U)$, except that for unlabeled data the probability of labels must be estimated based on the current classifier's output.

$$
\begin{aligned}
R(\mathcal{X}_U) &= \sum_{x_i \in \mathcal{X}_U} \boxed{r_p} \boxed{(1 - p(x_i))} \Pr(y_i = +1|x_i) \\
&+ \quad \boxed{r_n} \boxed{p(x_i)} (1 - \Pr(y_i = +1|x_i)), \\
\Pr(y = +1|x) &\approx p(x)
\end{aligned}
$$

$\Pr(y = +1|x)$ is the true probability of example x having label $+1$. We approximate this as $\Pr(y = +1|x) \approx p(x)$.

Scenario 2: MIL for learning from keyword searches



- **Postive instance:** Image belonging to class
- **Negative instance:** Image not in class
- **Positive bag:** Set of images returned by a keyword search for the class
- **Negative bag:** Set of images known to not contain the class

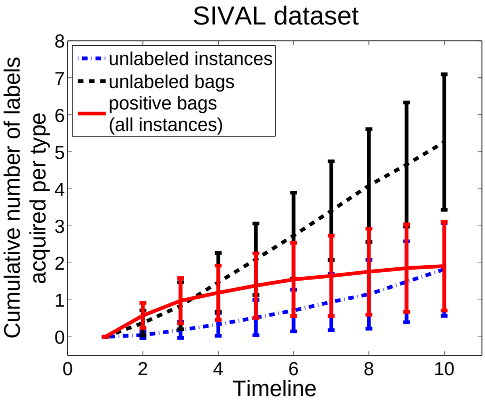
Google user experiment

Task	**Time (secs)**
click on all images containing 'airplane'	12
label an image	3

Results: SIVAL dataset

Cost	Our Approach			[Settles et al.]		
	Random	Multi-level Active	Gain over Random %	Random	MIU Active	Gain over Random%
10	+0.0051	+0.0241	372	+0.023	+0.050	117
20	+0.0130	+0.0360	176	+0.033	+0.070	112
50	+0.0274	+0.0495	81	+0.057	+0.087	52

What gets selected when?



The cumulative number of labels acquired for each type with increasing number of queries. Our method tends to request complete segmentations or image labels early on, followed by queries on unlabeled segments later on.