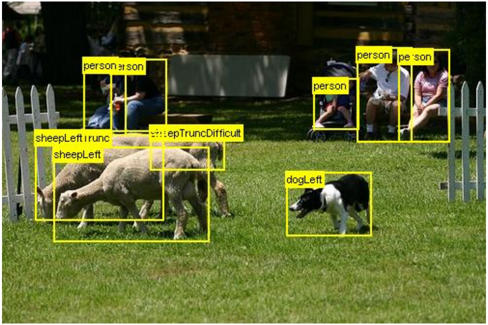
Large-Scale Live Active Learning: Training Object Detectors with Crawled Data and Crowds

Sudheendra Vijayanarasimhan
Kristen Grauman

Department of Computer Science
University of Texas at Austin
Austin, Texas

Object Detection



Challenge: Best results require large amount of cleanly labeled training examples.

Ways to Reduce Effort

Active learning

- minimize effort by focusing label requests on the most informative examples

  [Kapoor et al. ICCV 2007, Qi et al. CVPR 2008, Vijayanarasimhan et al. CVPR 2009, Joshi et al. CVPR 2009, Siddique et al. CVPR 2010]

Ways to Reduce Effort

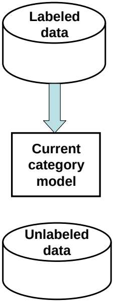
Active learning

- minimize effort by focusing label requests on the most informative examples

  [Kapoor et al. ICCV 2007, Qi et al. CVPR 2008, Vijayanarasimhan et al. CVPR 2009, Joshi et al. CVPR 2009, Siddique et al. CVPR 2010]

Ways to Reduce Effort

Active learning

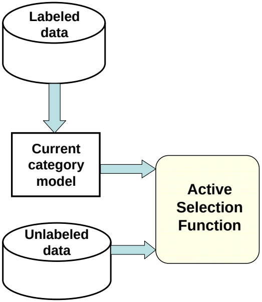


- minimize effort by focusing label requests on the most informative examples

  [Kapoor et al. ICCV 2007, Qi et al. CVPR 2008, Vijayanarasimhan et al. CVPR 2009, Joshi et al. CVPR 2009, Siddique et al. CVPR 2010]

Ways to Reduce Effort

Active learning

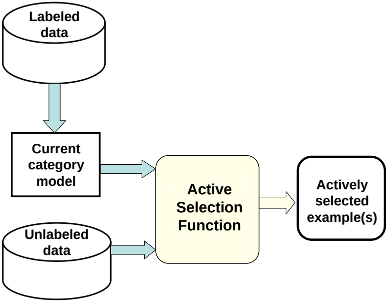


- minimize effort by focusing label requests on the most informative examples

  [Kapoor et al. ICCV 2007, Qi et al. CVPR 2008, Vijayanarasimhan et al. CVPR 2009, Joshi et al. CVPR 2009, Siddique et al. CVPR 2010]

Ways to Reduce Effort

Active learning

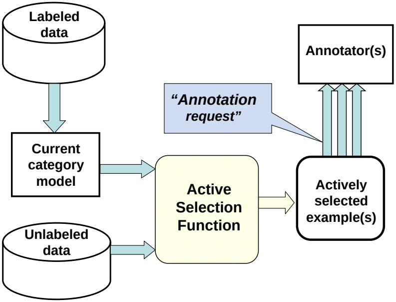


- minimize effort by focusing label requests on the most informative examples

  [Kapoor et al. ICCV 2007, Qi et al. CVPR 2008, Vijayanarasimhan et al. CVPR 2009, Joshi et al. CVPR 2009, Siddique et al. CVPR 2010]

Ways to Reduce Effort

Active learning

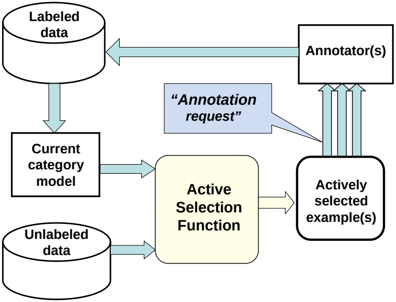


- minimize effort by focusing label requests on the most informative examples

  [Kapoor et al. ICCV 2007, Qi et al. CVPR 2008, Vijayanarasimhan et al. CVPR 2009, Joshi et al. CVPR 2009, Siddique et al. CVPR 2010]

Ways to Reduce Effort

Active learning

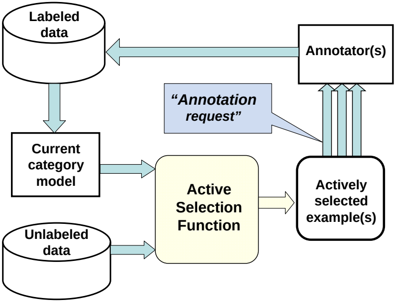


- minimize effort by focusing label requests on the most informative examples

  [Kapoor et al. ICCV 2007, Qi et al. CVPR 2008, Vijayanarasimhan et al. CVPR 2009, Joshi et al. CVPR 2009, Siddique et al. CVPR 2010]

Ways to Reduce Effort

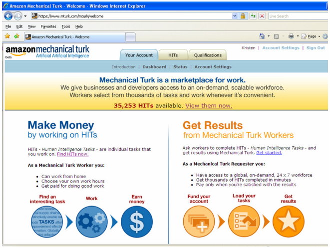
Active learning



- minimize effort by focusing label requests on the most informative examples

[Kapoor et al. ICCV 2007, Qi et al. CVPR 2008, Vijayanarasimhan et al. CVPR 2009, Joshi et al. CVPR 2009, Siddique et al. CVPR 2010]

Crowd-sourced annotations



- package annotation tasks to obtain from online human workers

[von Ahn et al. CHI 2004, Russell et al. IJCV 2007, Sorokin et al. 2008, Welinder et al. ACVHL 2010, Deng et al, CVPR 2009]

Problem

Thus far techniques are only tested in artificially controlled settings:

Problem

Thus far techniques are only tested in artificially controlled settings:

- use "sandbox" datasets - dataset's source and scope is fixed

Problem

Thus far techniques are only tested in artificially controlled settings:

- use "sandbox" datasets - dataset's source and scope is fixed
- computational cost of active selection and retraining the model generally ignored - linear/quadratic time

Problem

Thus far techniques are only tested in artificially controlled settings:

- use “sandbox” datasets - dataset’s source and scope is fixed
- computational cost of active selection and retraining the model generally ignored - linear/quadratic time
- crowd-sourced collection requires iterative fine-tuning

Goal

Goal

Take active learning and crowd-sourced annotation collection out of the "sandbox".

Goal

Goal

Take active learning and crowd-sourced annotation collection out of the "sandbox".

- break free from dataset-based learning

Goal

Goal

Take active learning and crowd-sourced annotation collection out of the "sandbox".

- break free from dataset-based learning
- collect information on the fly (no manual intervention)

Goal

Goal

Take active learning and crowd-sourced annotation collection out of the “sandbox”.

- break free from dataset-based learning
- collect information on the fly (no manual intervention)
- large-scale data

Goal

Goal

Take active learning and crowd-sourced annotation collection out of the "sandbox".

- break free from dataset-based learning
- collect information on the fly (no manual intervention)
- large-scale data

Our Approach: Live Learning

Live active learning system that autonomously builds models for object detection

Goal

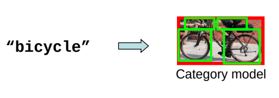
Goal

Take active learning and crowd-sourced annotation collection out of the "sandbox".

- break free from dataset-based learning
- collect information on the fly (no manual intervention)
- large-scale data

Our Approach: Live Learning

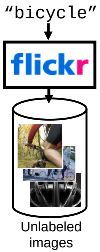
Live active learning system that autonomously builds models for object detection



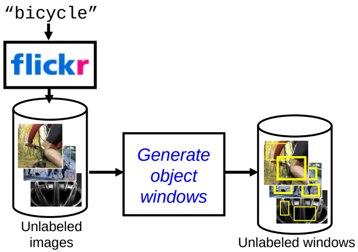
Our Approach: Live Learning

“bicycle”

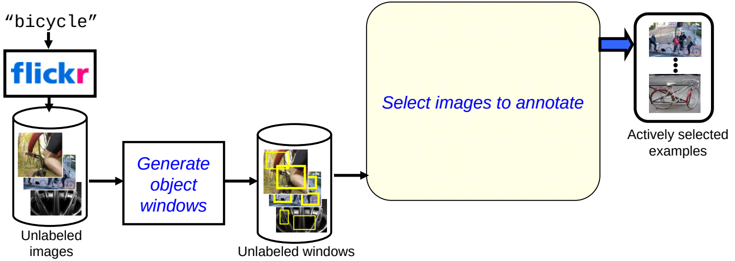
Our Approach: Live Learning



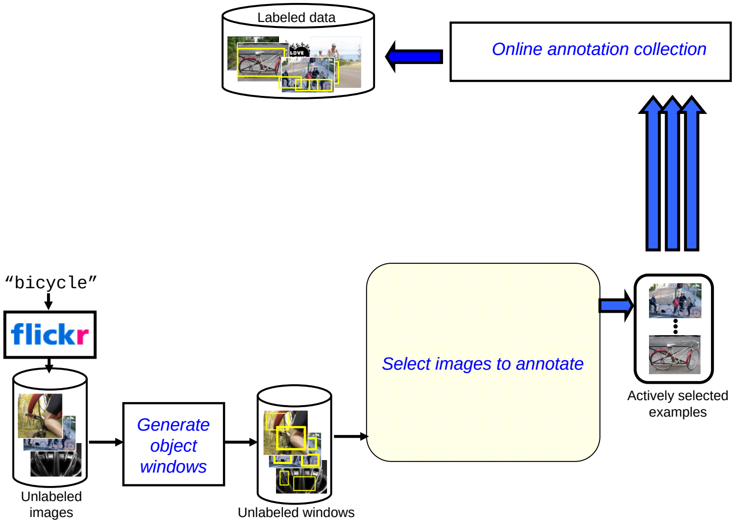
Our Approach: Live Learning



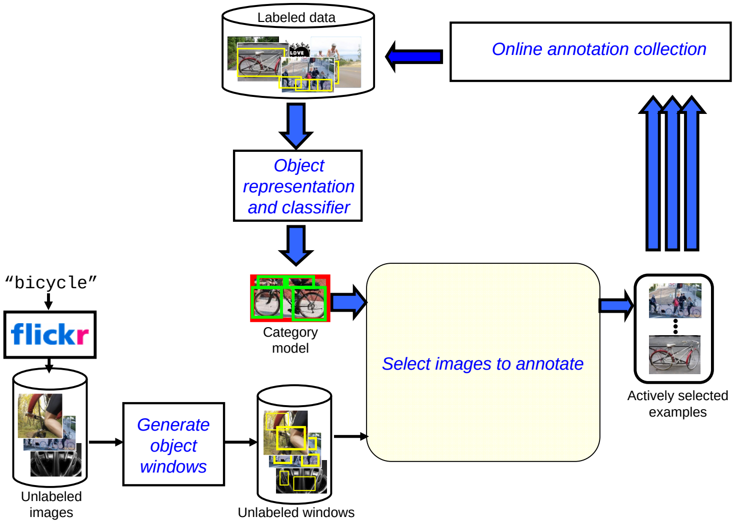
Our Approach: Live Learning



Our Approach: Live Learning



Our Approach: Live Learning



Main Contributions

- Linear classification
  part-based linear detector based on non-linear feature coding

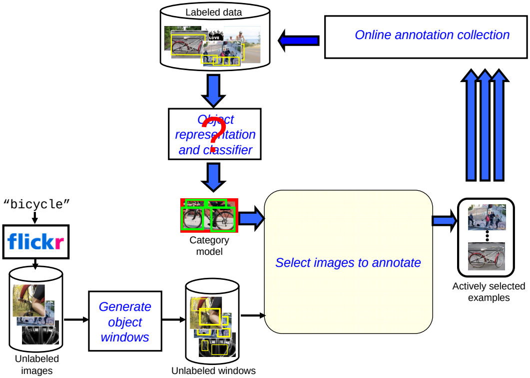
Main Contributions

- Linear classification
  part-based linear detector based on non-linear feature coding
- Large-scale active selection
  sub-linear time hashing scheme for efficiently selecting uncertain examples [Jain, Vijayanarasimhan & Grauman, NIPS 2010]

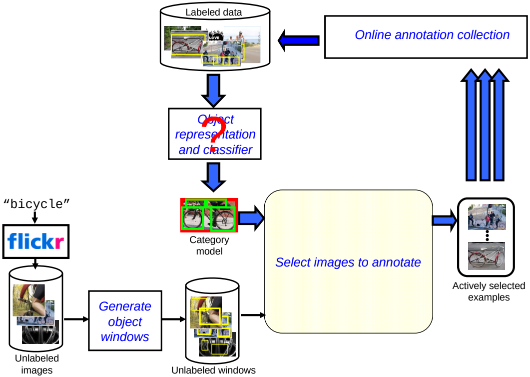
Main Contributions

- Linear classification
  part-based linear detector based on non-linear feature coding
- Large-scale active selection
  sub-linear time hashing scheme for efficiently selecting uncertain examples [Jain, Vijayanarasimhan & Grauman, NIPS 2010]
- Live learning results
  for active *detection* of unprecedented scale and autonomy for the first time

Outline


Labeled data
Online annotation collection
Object representation and classifier
?
"bicycle"
flickr
Category model
Select images to annotate
Actively selected examples
Unlabeled images
Generate object windows
Unlabeled windows

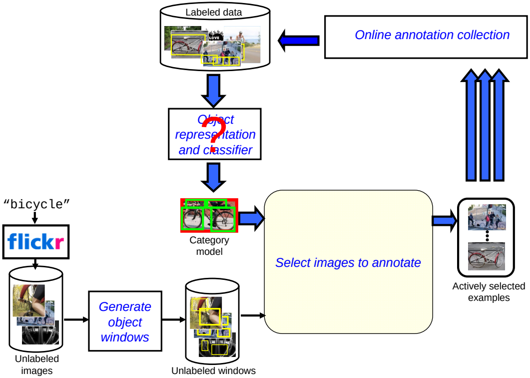
Outline



Linear classification

- fast/incremental training using linear SVM

Outline



Linear classification

- fast/incremental training using linear SVM
- efficient active selection using our hyperplane hash functions

Object Representation and Classifier

Part based object representation

Object Representation and Classifier

Part based object representation

Root

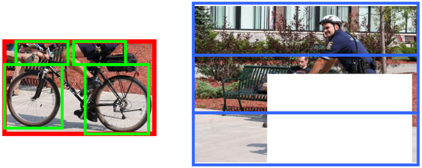
Object Representation and Classifier

Part based object representation

Root Parts

Object Representation and Classifier

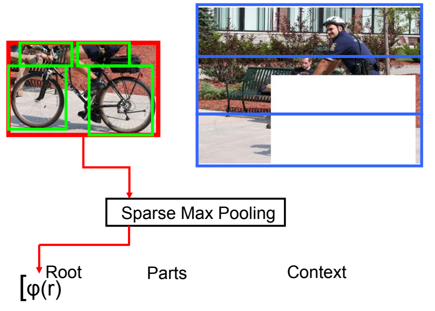
Part based object representation



Root Parts Context

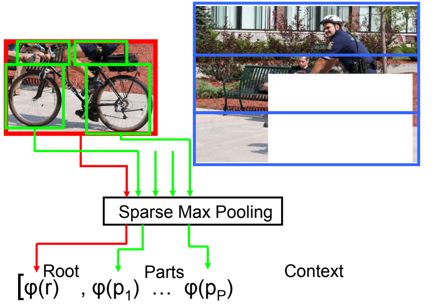
Object Representation and Classifier

Part based object representation



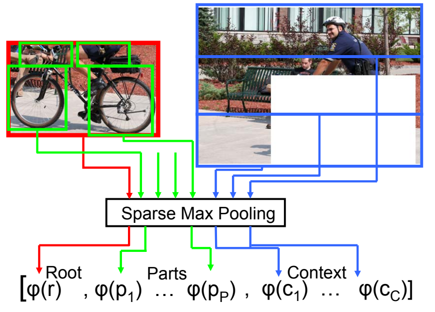
Object Representation and Classifier

Part based object representation



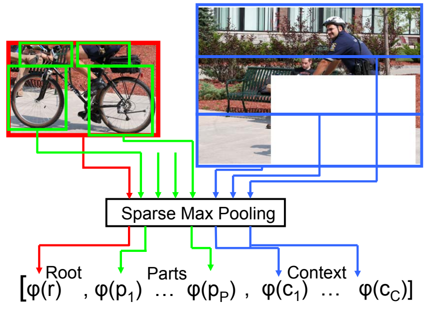
Object Representation and Classifier

Part based object representation



Object Representation and Classifier

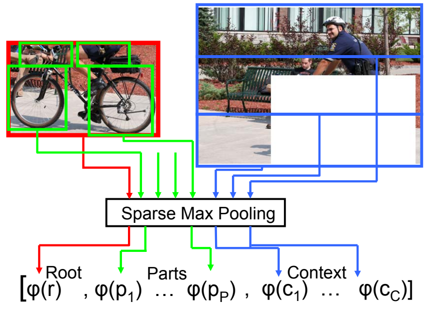
Part based object representation



Sparse Max Pooling [Yang et al. '10] – similar to bag of words:

Object Representation and Classifier

Part based object representation

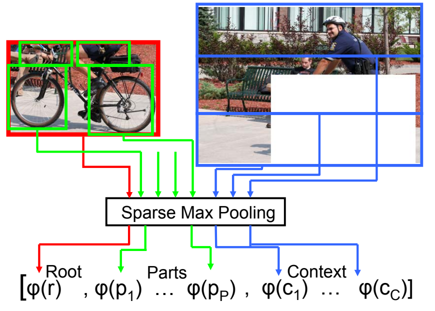


Sparse Max Pooling [Yang et al. '10] – similar to bag of words:

- sparse coding – fuller representation of original features

Object Representation and Classifier

Part based object representation

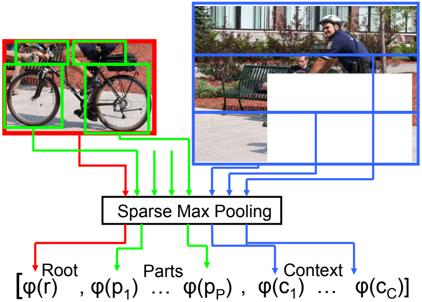


Sparse Max Pooling [Yang et al. '10] – similar to bag of words:

- sparse coding – fuller representation of original features
- max pooling – better discriminability in clutter [Boureau '10].

Object Representation and Classifier

Part based object representation



Score windows as linear sum:

$$f(O) = \mathbf{w}^T \varphi(O)$$

$$= \mathbf{w}_r \varphi(r) + \sum_{i=1}^{P} \mathbf{w}_{p_i} \varphi(p_i)$$

$$+ \sum_{i=1}^{C} \mathbf{w}_{c_i} \varphi(c_i),$$

w - linear SVM weights

Sparse Max Pooling [Yang et al. '10] – similar to bag of words:

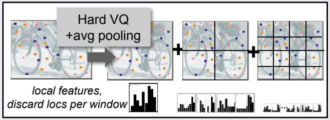
- sparse coding – fuller representation of original features
- max pooling – better discriminability in clutter [Boureau '10].

Object Representation and Classifier

Relationship to existing detection models

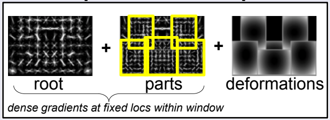
Spatial pyramid (SP)

[Lazebnik et al. '06, Vedaldi et al. '09]



Latent SVM (LSVM)

[Felzenswalb et al. '09]

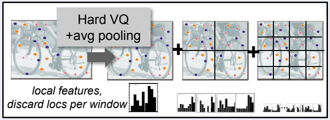


Object Representation and Classifier

Relationship to existing detection models

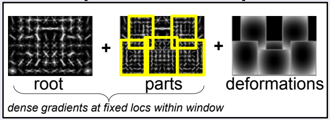
Spatial pyramid (SP)

[Lazebnik et al. '06, Vedaldi et al. '09]

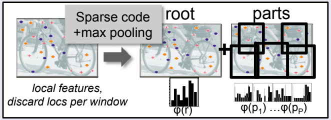


Latent SVM (LSVM)

[Felzenswalb et al. '09]



Ours

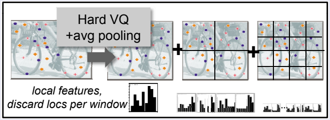


Object Representation and Classifier

Relationship to existing detection models

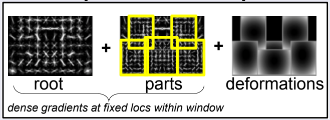
Spatial pyramid (SP)

[Lazebnik et al. '06, Vedaldi et al. '09]

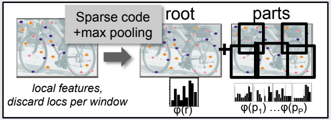


Latent SVM (LSVM)

[Felzenswalb et al. '09]



Ours



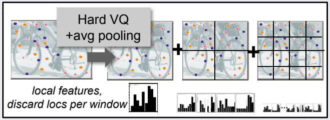
- faster training (linear SVM)

Object Representation and Classifier

Relationship to existing detection models

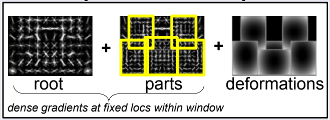
Spatial pyramid (SP)

[Lazebnik et al. '06, Vedaldi et al. '09]

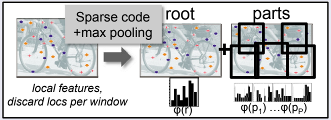


Latent SVM (LSVM)

[Felzenswalb et al. '09]

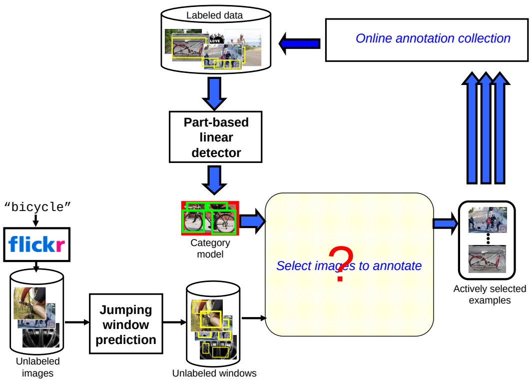


Ours

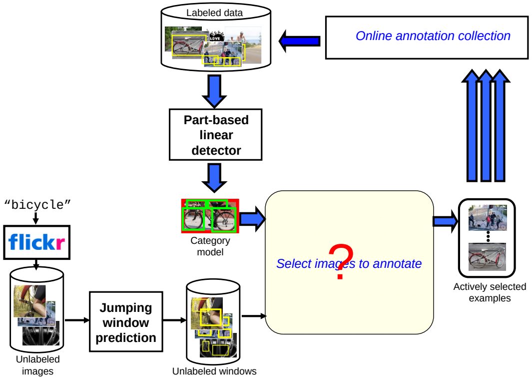


- faster training (linear SVM)
- results comparable to non-linear detectors

Selecting Images to Annotate



Selecting Images to Annotate

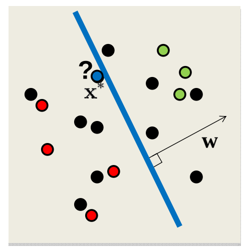


Efficient active selection

- select most useful windows from millions

Active Selection of Object Windows

SVM margin criterion for active selection



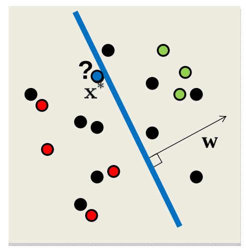
Select point nearest to hyperplane decision boundary for labeling.

$$\mathbf{x}^* = \underset{\mathbf{x}_i \in \mathcal{U}}{\operatorname{argmin}} |\mathbf{w}^T \mathbf{x}_i|$$

[Tong & Koller, 2000; Schohn & Cohen, 2000; Campbell et al. 2000]

Active Selection of Object Windows

SVM margin criterion for active selection



Select point nearest to hyperplane decision boundary for labeling.

$$\mathbf{x}^* = \underset{\mathbf{x}_i \in \mathcal{U}}{\operatorname{argmin}} |\mathbf{w}^T \mathbf{x}_i|$$

[Tong & Koller, 2000; Schohn & Cohen, 2000; Campbell et al. 2000]

Problem: With massive unlabeled pool, cannot afford exhaustive linear scan to make selection.

Active Selection of Object Windows

Sub-linear time selection through hyperplane hashing

[Jain, Vijayanarasimhan and Grauman, NIPS 2010]

- hash function $h(.)$ - high probability of collision when $\varphi(O)$ close to $\mathbf{w}$

Active Selection of Object Windows

Sub-linear time selection through hyperplane hashing

[Jain, Vijayanarasimhan and Grauman, NIPS 2010]

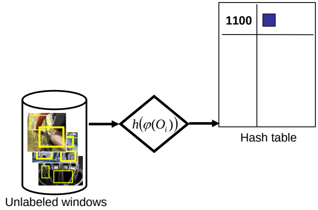
- hash function $h(.)$ - high probability of collision when $\varphi(O)$ close to $\mathbf{w}$

Active Selection of Object Windows

Sub-linear time selection through hyperplane hashing

[Jain, Vijayanarasimhan and Grauman, NIPS 2010]

- hash function $h(.)$ - high probability of collision when $\varphi(O)$ close to $\mathbf{w}$
- preprocess - hash unlabeled windows into table

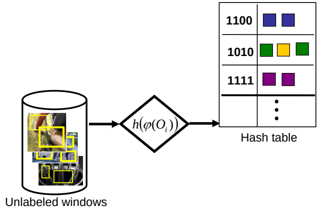


Active Selection of Object Windows

Sub-linear time selection through hyperplane hashing

[Jain, Vijayanarasimhan and Grauman, NIPS 2010]

- hash function $h(.)$ - high probability of collision when $\varphi(O)$ close to $\mathbf{w}$
- preprocess - hash unlabeled windows into table

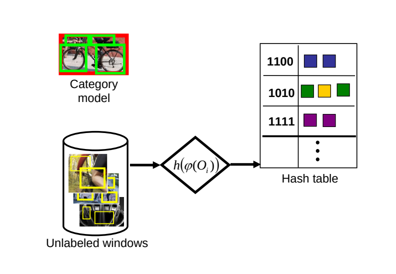


Active Selection of Object Windows

Sub-linear time selection through hyperplane hashing

[Jain, Vijayanarasimhan and Grauman, NIPS 2010]

- hash function $h(.)$ - high probability of collision when $\varphi(O)$ close to **w**
- preprocess - hash unlabeled windows into table
- active learning loop - hash classifier **w** and retrieve examples

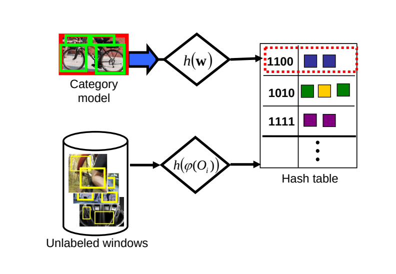


Active Selection of Object Windows

Sub-linear time selection through hyperplane hashing

[Jain, Vijayanarasimhan and Grauman, NIPS 2010]

- hash function $h(.)$ - high probability of collision when $\varphi(O)$ close to $\mathbf{w}$
- preprocess - hash unlabeled windows into table
- active learning loop - hash classifier $\mathbf{w}$ and retrieve examples

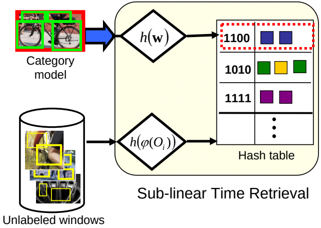


Active Selection of Object Windows

Sub-linear time selection through hyperplane hashing

[Jain, Vijayanarasimhan and Grauman, NIPS 2010]

- hash function $h(.)$ - high probability of collision when $\varphi(O)$ close to $\mathbf{w}$
- preprocess - hash unlabeled windows into table
- active learning loop - hash classifier $\mathbf{w}$ and retrieve examples

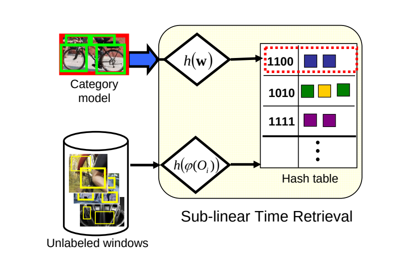


Active Selection of Object Windows

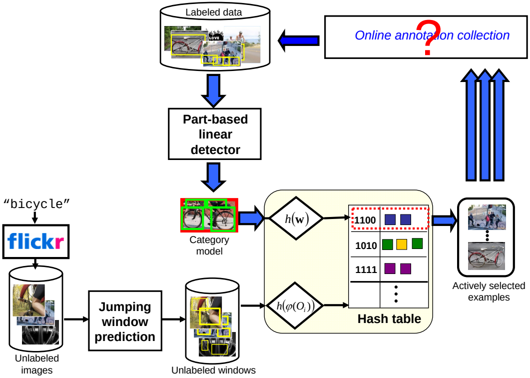
Sub-linear time selection through hyperplane hashing

[Jain, Vijayanarasimhan and Grauman, NIPS 2010]

- hash function $h(.)$ - high probability of collision when $\varphi(O)$ close to **w**
- preprocess - hash unlabeled windows into table
- active learning loop - hash classifier **w** and retrieve examples
- evaluate $\sim 10^3$ windows vs. $\sim 10^6$ for exhaustive



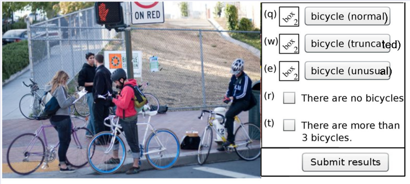
Online Annotation Collection



- on the fly
- reliable annotations without pruning

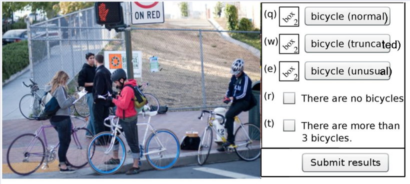
Online Annotation Collection

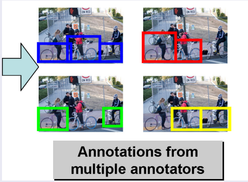
Mechanical Turk Interface



Online Annotation Collection

Mechanical Turk Interface

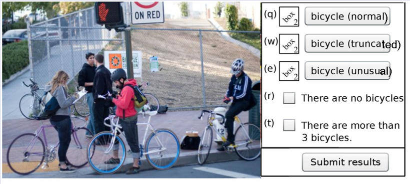


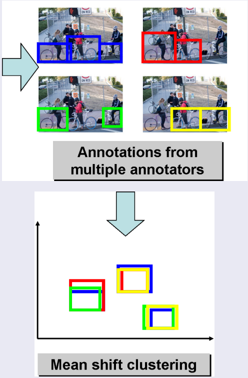


- post same image to multiple (5-10) annotators

Online Annotation Collection

Mechanical Turk Interface

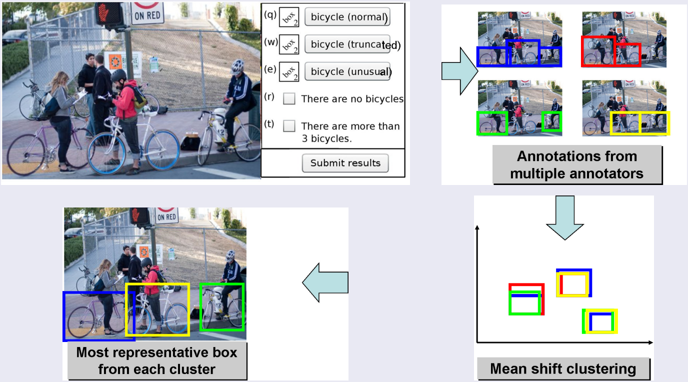




- post same image to multiple (5-10) annotators
- cluster all bounding boxes to obtain consensus

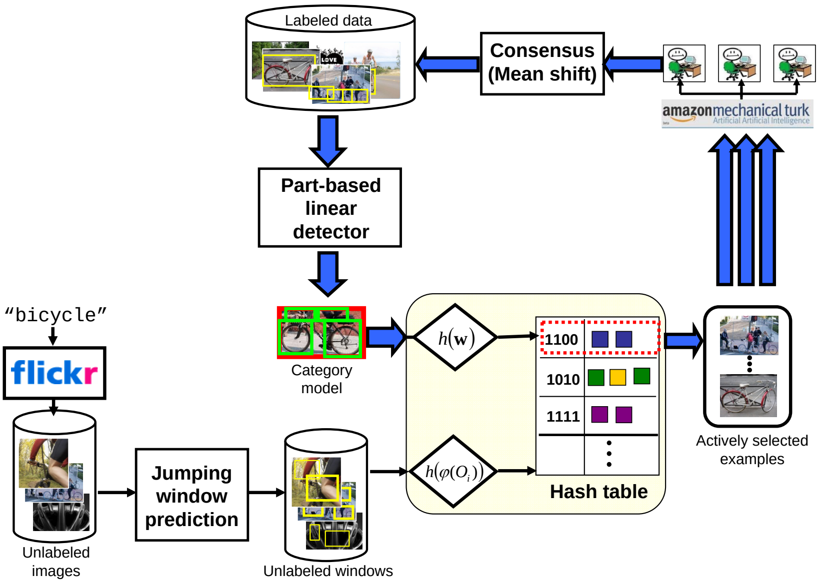
Online Annotation Collection

Mechanical Turk Interface



- post same image to multiple (5-10) annotators
- cluster all bounding boxes to obtain consensus

Summary: Live Learning



Results

PASCAL 2007 challenge

- 20 different objects under changes in viewpoint, scale, and background clutter.
- $\sim$ 5000 training and test examples
- given an image detect all objects

Results

PASCAL 2007 challenge

- 20 different objects under changes in viewpoint, scale, and background clutter.
- $\sim$ 5000 training and test examples
- given an image detect all objects

Live learning on Flickr

- 6 of the most challenging PASCAL objects
- New Flickr testset

Results

PASCAL 2007 challenge

- 20 different objects under changes in viewpoint, scale, and background clutter.
- $\sim$ 5000 training and test examples
- given an image detect all objects

Live learning on Flickr

- 6 of the most challenging PASCAL objects
- New Flickr testset

Features

- 30,000 SIFT features densely extracted
- 60,000 words with hierarchical kmeans
- sparse coding using LLC [Yang et al. '10]

Results

PASCAL 2007 challenge

- 20 different objects under changes in viewpoint, scale, and background clutter.
- ~ 5000 training and test examples
- given an image detect all objects

Live learning on Flickr

- 6 of the most challenging PASCAL objects
- New Flickr testset

Features

- 30,000 SIFT features densely extracted
- 60,000 words with hierarchical kmeans
- sparse coding using LLC [Yang et al. '10]

Implementation

- 12 parts from LSVM detector
- 100 images per active iteration

Sandbox Results (PASCAL 2007)

Comparison to state-of-art

	aero.	cat	dog	sheep	sofa	train	bicyc.	bird	boat	bottl	bus		Mean
BoF SP	30.4	17.7	18.0	19.1	14.7	35.7	43.1	6.9	3.5	10.8	35.8		23.0
Ours	**48.4**	**30.7**	**21.8**	**28.8**	**33.0**	**47.7**	48.3	14.1	13.6	15.3	43.9	$\cdots$	30.5

- part-based, single feature representation, linear model

Sandbox Results (PASCAL 2007)

Comparison to state-of-art

	aero.	cat	dog	sheep	sofa	train	bicyc.	bird	boat	bottl	bus		Mean
BoF SP	30.4	17.7	18.0	19.1	14.7	35.7	43.1	6.9	3.5	10.8	35.8		23.0
Ours	**48.4**	**30.7**	**21.8**	**28.8**	**33.0**	**47.7**	48.3	14.1	13.6	15.3	43.9	...	30.5
LSVM+HOG[1]	32.8	21.3	8.8	16.2	24.4	39.2	**56.8**	2.5	**16.8**	**28.5**	39.7		29.1
SP+MKL[2]	37.6	30.0	21.5	23.9	28.5	45.3	47.8	**15.3**	15.3	21.9	**50.7**		**32.1**

[1] [Felzenszwalb et al. '09] [2] [Vedaldi et al. '09]

- part-based, single feature representation, linear model
- competitive with state-of-art (better for 6 classes)

Live Learning Results

Live learning tested on PASCAL testset

Live Learning Results

Live learning tested on PASCAL testset

	bird	boat	dog	potted plant	sheep	chair
Ours	**15.8**	**18.9**	**25.3**	11.6	**28.4**	9.1
Previous best	15.3	16.8	21.5	**14.6**	23.9	**17.9**

Significant improvements in state-of-art on challenging categories

Live Learning Results

Live learning tested on PASCAL testset

	bird	boat	dog	potted plant	sheep	chair
Ours	**15.8**	**18.9**	**25.3**	11.6	**28.4**	9.1
Previous best	15.3	16.8	21.5	**14.6**	23.9	**17.9**

Significant improvements in state-of-art on challenging categories

Computation Time

	Active selection	Training	Detection per image
Ours + active	10 mins	5 mins	150 secs
LSVM [Felzenszwalb et al. 2009]	3 hours	4 hours	2 secs
SP+MKL [Vedaldi et al. 2009]	93 hours	> 2 days	67 secs

Our approach's efficiency makes live learning feasible.

Live Learning Results

Live learning tested on Flickr testset

General test set of web images

Live Learning Results

Live learning tested on Flickr testset

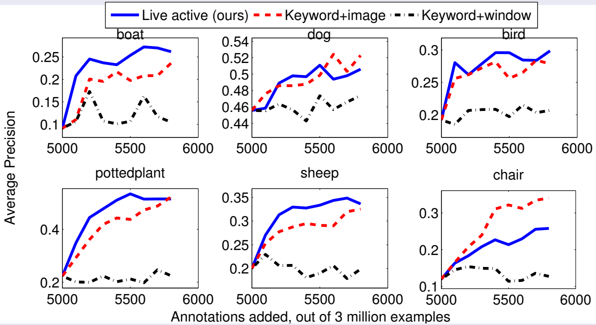
General test set of web images

- **keyword+image:** randomly select a keyword filtered image to get bounding box
- **keyword+window:** randomly select a window within a keyword filtered image and obtain binary label

Live Learning Results

Live learning tested on Flickr testset

General test set of web images

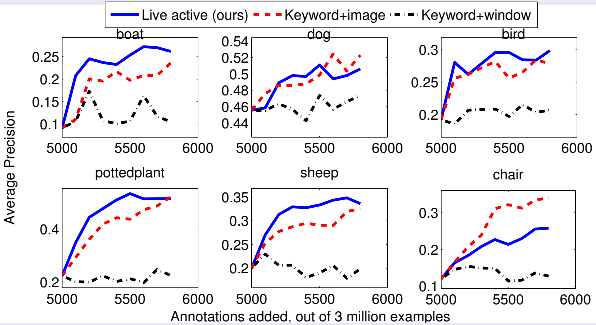


- **keyword+image:** randomly select a keyword filtered image to get bounding box
- **keyword+window:** randomly select a window within a keyword filtered image and obtain binary label

Live Learning Results

Live learning tested on Flickr testset

General test set of web images



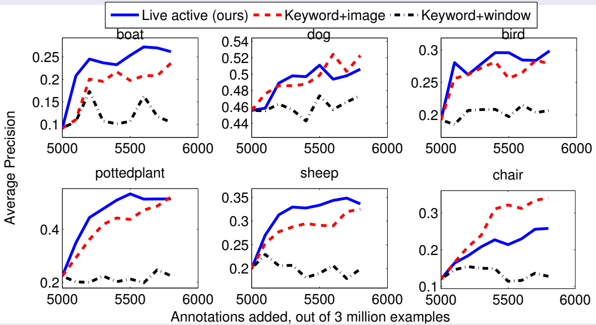
- **keyword+image:** randomly select a keyword filtered image to get bounding box
- **keyword+window:** randomly select a window within a keyword filtered image and obtain binary label

- dramatic improvements for most categories

Live Learning Results

Live learning tested on Flickr testset

General test set of web images



- **keyword+image:** randomly select a keyword filtered image to get bounding box
- **keyword+window:** randomly select a window within a keyword filtered image and obtain binary label

- dramatic improvements for most categories
- outperforms status quo approach of learning

Conclusions

- autonomous online learning - break-free from sandbox learning
- no intervention in example/annotation selection or pruning
- obtains results better than state-of-art on challenging PASCAL dataset
- largest scale active learning results to our knowledge