Multiple Component Learning for Object Detection

Piotr Dollár[1,2] Boris Babenko[2] Pietro Perona[1] Serge Belongie[1,2] Zhuowen Tu[3]

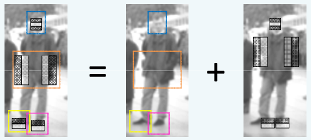
[1] Electrical Engineering, California Institute of Technology
pdollar@caltech.edu

[2] Computer Science and Engineering, University of California, San Diego
{bbabenko,sjb}@cs.ucsd.edu

[3] Lab of Neuro Imaging, University of California, Los Angeles
Zhuowen.tu@loni.ucla.edu

Goals

1. Learn part-based classifier with weak supervision (object labels provided, but no part labels)
2. Part models are classifiers from rich hypothesis class (rather than Gaussian distributions, templates, etc.)
3. No complex inference since model is discriminative



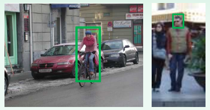
MCL = part-based + learned representation

Overview

(1) *Learning* a single part

Weakly supervised learning

- Object location in positive images unknown
- Developed for learning *objects*, use for *parts*
- We use Multiple Instance Learning (MIL)



(2) Learning *diverse* parts

What prevents learning same part repeatedly?

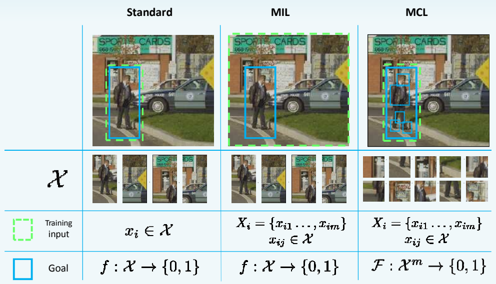
- Different *weighting* of data
- Not all parts expressed in all images

(3) *Combining* part detectors

Boosting offers way of combining multiple diverse classifiers

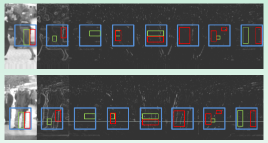
- Train one weak (part) classifier using MIL
- Re-weight samples according to current error
- Repeat until training error sufficiently low

Set Learning



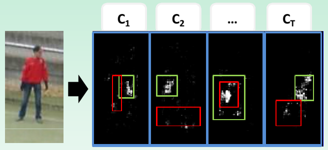
Pedestrian Detection

Low-level features



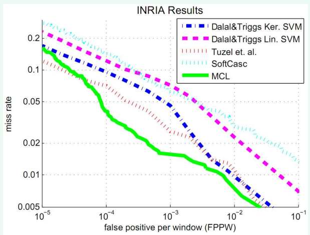
- Features for component classifiers
- Haars over multiple channels: gray (1), grad (1), quantized grad (6)

Incorporating Spatial Relations

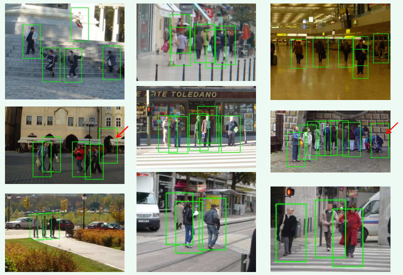


- Features for overall classifier
- Densely compute component responses C_i
- Final classifier retrained with Haars over C_i

Results on INRIA Data



We achieve state of the art results.

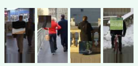


Single Scale Detection Results

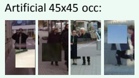
Learned components (first 5)

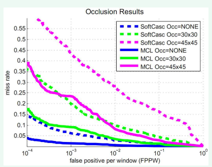
Robustness to Occlusion

Artificial 30x30 occ:



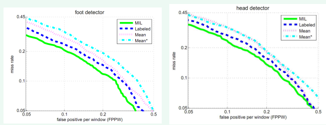
Artificial 45x45 occ:





MCL degrades gracefully w occ.

Role of Alignment



- Aligned data → higher performance
- Can't simultaneously align articulated object without part based model

Multiple Component Learning (MCL)

Derivation

- Boosting
  - Input: *N* training examples (x_i, y_i) with $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$
  - Combines *T* weak classifiers $h_t(x)$ to learn strong classifier: $H(x) = sign(\sum_{t=1}^{T} \alpha_t h_t(x))$
  - Excellent generalization and strong theoretical foundation
- No assumptions about input space
  - Only requires weak classifiers for arbitrary $\mathcal{X}$: $h_t : \mathcal{X} \to \{-1,1\}$
  - For example, can use $\mathcal{X}^m$ (sets) in place of $\mathcal{X}$
- Can therefore use MIL to train weak classifier
  - MIL learns a function $\hat{F} : \mathcal{X}^m \to [0,1]$
  - Specifically, learns *f* in $\hat{F}(X_i) = \mathrm{softmax}_j(f(x_{ij}))$
  - Only need to adjust MIL to take weights

Algorithm

Given: N labeled training examples (X_i, y_i) with $y_i \in \{-1,1\}$ and $X_i = \{x_{i1}, \ldots, x_{im}\}$, and initial distr. of weights $D_1(i) = \frac{1}{N}$ over the examples.

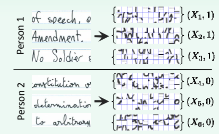
For $t = 1, \ldots, T$:

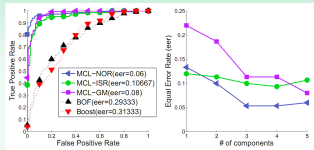
- Train a MIL classifier $\hat{F}_t : \mathcal{X}^m \to [0,1]$ using distribution D_t. Let $\hat{F}'_t(X_i) = (2 \times \mathbf{1}[\hat{F}_t(X_i) > th] - 1)$, where $th = .5$ or is chosen to minimize ϵ_t.
- Calculate error of $\hat{F}'_t$: $\epsilon_t = \sum_{i=1}^{N} D_t(i)\mathbf{1}(y_i \neq \hat{F}'_t(X_i))$.
- Set $\alpha_t = -\frac{1}{2}\log(\epsilon_t/(1-\epsilon_t))$.
- Set $D_{t+1}(i) = D_t(i)\exp\left(-\alpha_t y_i \hat{F}'_t(X_i)\right)/Z_t$, where $Z_t = 2\sqrt{\epsilon_t(1-\epsilon_t)}$ is a normalization factor.

Output the MCL classifier: $\mathcal{F}(X) = \mathrm{sign}\left(\sum_{t=1}^{T} \alpha_t \hat{F}'_t(X)\right)$

Other Applications

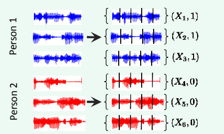
Writer Identification

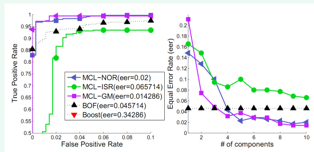




- Text independent handwriting identification
- 2 people, 2 pages of text each
- Haar features on 25x25 patches

Speaker Identification





- Content independent speaker identification
- John Kerry vs. George W. Bush, 2004
- Standard MFCC features

Summary

Advantages:

1. General notion of parts (components)
2. Component learning weakly supervised
3. Principled, general algorithm
4. State of the art results with simple features

Disadvantages:

1. Large amount of data needed
2. Evaluating all components slow (currently working on improving speed)