Exemplar-SVMs: Visual Object Detection, Label Transfer and Image Retrieval

Tomasz Malisiewicz
(Massachusetts Institute of Technology)

Joint work with:
Abhinav Shrivastava, Abhinav Gupta, and Alexei (Alyosha) Efros
(Carnegie Mellon University)
Talk Overview

- Visual Object Detection
- Exemplar-SVM Learning
- Understanding Exemplar-SVMs
- Experimental Results
- PASCAL VOC Object Detection
- Label Transfer
- Cross-domain Image Retrieval
- Concluding remarks and take-home lessons
Object Detectors
Object Detectors

- Histogram of Oriented Gradients features computed across a multiscale pyramid

Dalal et al 2005

Image  HOG
Object Detectors

Dalal et al 2005

- Histogram of Oriented Gradients features computed across a multiscale pyramid
- Linear SVMs for learning
Object Detectors

• Histogram of Oriented Gradients features computed across a multiscale pyramid

• Linear SVMs for learning

• A large dataset such as PASCAL VOC (Everingham et al 2010)
Object Detectors
Exemplar-SVMs

- Best of both worlds:
  - Effectiveness of discriminatively-trained object detectors
  - Explicit correspondence of Nearest Neighbor approaches

Exemplar-SVMs

- Because each Exemplar-SVM is defined by a **single** positive instance, we can use different features for each exemplar.
Because each Exemplar-SVM is defined by a **single** positive instance, we can use different features for each exemplar.

Apply each Exemplar-SVM to test image in a sliding-window fashion.

- 7x4 HOG
- 4x8 HOG
Exemplar-SVMs

Exemplar E’s Objective Function:

$$\Omega_E(w, b) = \|w\|^2 + C_1 h(w^T x_E + b) + C_2 \sum_{x \in \mathcal{N}_E} h(-w^T x - b)$$

$$h(x) = \max(1-x,0) \text{ “hinge-loss”}$$
Exemplar-SVMs

Exemplar E’s Objective Function:

$$\Omega_E(w, b) = \|w\|^2 + C_1 h(w^T x_E + b) + C_2 \sum_{x \in N_E} h(-w^T x - b)$$

$$h(x) = \max(1-x, 0) \text{ "hinge-loss"}$$

Exemplar represented by \(~100\) HOG Cells (\(\sim 3,000\)D features)
Exemplar-SVMs

Exemplar E’s Objective Function:

$$\Omega_E(w, b) = ||w||^2 + C_1 h(w^T x_E + b) + C_2 \sum_{x \in \mathcal{N}_E} h(-w^T x - b)$$

$$h(x) = \max(1-x,0) \text{ “hinge-loss”}$$

Exemplar represented by ~100 HOG Cells (~3,000D features)

Windows from images not containing any in-class instances (2,000 images x 10,000 windows per image = 20M negatives)
Embarrassingly Parallel

- Each exemplar performs its own hard negative mining
- Solve many convex learning problems
- Parallel training on cluster
Visualizing Exemplar-SVMs
Visualizing Exemplar-SVMs

Exemplar-SVMs

Top Detections in Test Set
Understanding Exemplar-SVMs

Traditional NN vs. Local Distance Function vs. Exemplar-SVM

Frome et al, NIPS 2006
Understanding Exemplar-SVMs

*Learned Distance Function
Understanding Exemplar-SVMs

![Exemplar w](image)

Top 6 Detections from Testset

*Learned Distance Function*
Understanding Exemplar-SVMs

Exemplar w Top 6 Detections from Testset

Exemplar-SVM

*Learned Distance Function
Understanding Exemplar-SVMs

*Learned Distance Function
Ensemble of Exemplar-SVMs
Ensemble of Exemplar-SVMs

Platt Calibration
(Platt 1999)

Exemplars Compete
Ensemble of Exemplar-SVMs

Platt Calibration
(Platt 1999)

Exemplars Compete

Learning Exemplar Co-occurrence Matrix

Exemplars are Combined
Object Category Detection

mAP averaged across 20 object categories on the PASCAL VOC 2007 object detection task

<table>
<thead>
<tr>
<th>Method</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional NN + Calibration</td>
<td>0.110</td>
</tr>
<tr>
<td>Local Distance Function + Calibration</td>
<td>0.157</td>
</tr>
<tr>
<td><strong>Exemplar-SVMs + Calibration</strong></td>
<td><strong>0.198</strong></td>
</tr>
<tr>
<td><strong>Exemplar-SVMs + Co-occurrence</strong></td>
<td><strong>0.227</strong></td>
</tr>
<tr>
<td>One SVM per category (Dalal and Triggs 2005)</td>
<td>0.097</td>
</tr>
<tr>
<td>Deformable Part Model (Felzenszwalb et al 2010)</td>
<td>0.266</td>
</tr>
</tbody>
</table>
Beyond Detection: Label Transfer

Category-based Detector

Ensemble of Exemplar-SVMs
Exemplar

Detector \( w \)

Appearance
Exemplar

Detector $w$

Appearance
Exemplar

Detector $w$

Appearance

Meta-data

Person

Person
Talk Overview

• Visual Object Detection
• Exemplar-SVM Learning
• Understanding Exemplar-SVMs
• Experimental Results
• PASCAL VOC Object Detection
• Label Transfer

• Cross-domain Image Retrieval

• Concluding remarks and take-home lessons
Exemplar-SVMs

Exemplar E’s Objective Function:

\[ \Omega_E(w, b) = \|w\|^2 + C_1 h(w^T x_E + b) + C_2 \sum_{x \in \mathcal{N}_E} h(-w^T x - b) \]

\[ h(x) = \max(1-x, 0) \text{ “hinge-loss”} \]

Exemplar represented by \( \sim 100 \) HOG Cells (\( \sim 3,000 \)D features)

Windows from images not containing any in-class instances (2,000 images x 10,000 windows per image = 20M negatives)
Exemplar-SVMs

Exemplar E’s Objective Function:

\[ \Omega_E(w, b) = \|w\|^2 + C_1 h(w^T x_E + b) + C_2 \sum_{x \in N_E} h(-w^T x - b) \]

\[ h(x) = \max(1-x, 0) \text{ “hinge-loss”} \]

Exemplar represented by \(~100\) HOG Cells (\(~3,000\)D features)

Windows from images not containing any in-class instances
(\(2,000\) images x \(10,000\) windows per image = \(20\)M negatives)
## Object Category Detection

Exemplar-SVMs* = Exemplar-SVMs with random negatives

<table>
<thead>
<tr>
<th>Method</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional NN + Calibration</td>
<td>0.110</td>
</tr>
<tr>
<td>Local Distance Function + Calibration</td>
<td>0.157</td>
</tr>
<tr>
<td>Exemplar-SVMs + Calibration</td>
<td>0.198</td>
</tr>
<tr>
<td>Exemplar-SVMs + Co-occurrence</td>
<td>0.227</td>
</tr>
<tr>
<td><em><em>Exemplar-SVMs</em> + Calibration</em>*</td>
<td><strong>0.142</strong></td>
</tr>
<tr>
<td><em><em>Exemplar-SVMs</em> + Co-occurrence</em>*</td>
<td><strong>0.197</strong></td>
</tr>
</tbody>
</table>
Cross-domain Image Matching

Learn Exemplar-SVM for query image

Query Image

Negatives mined from random Flickr images
Image Retrieval

Query Image

Random Flickr Images
Image Retrieval

Query Image

Random Flickr Images
Search using Paintings

Painting

Our Approach

GIST

Bag-of-Words

Tiny Images

HOG

Our Approach
Search Using Sketches

Input Sketch

Our Approach

Tiny Images

GIST

Bag-of-Words

HOG
Sketch to Image

Input Sketch

Our Top Matches
Exemplar-SVM vs. Google
Exemplar-SVM vs. Google
Concluding Remarks

• A mixture model with N mixture components

• The positives are represented non-parametrically and the negatives are represented parametrically
Concluding Remarks

- Exemplar-SVMs can be used for detection, label transfer, as well as cross-domain image matching.
Concluding Remarks

• Exemplar-SVMs can be used for detection, label transfer, as well as cross-domain image matching.

• Goods news: Results surprisingly nice, embarrassingly parallel learning.
Concluding Remarks

- Exemplar-SVMs can be used for detection, label transfer, as well as cross-domain image matching

- Goods news: Results surprisingly nice, embarrassingly parallel learning

- Bad news: Computationally Expensive
Thank you
Thank you

Come visit poster #30 in the Informatics Forum or Google “exemplar svm” to find papers and code