Classification aided two stage localization
Hedi Harzallah, Cordelia Schmid, Frédéric Jurie, Adrien Gaidon

To cite this version:

HAL Id: inria-00548652
https://inria.hal.science/inria-00548652
Submitted on 20 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Classification aided two stage localization

Hedi Harzallah, Cordelia Schmid, Frederic Jurie, Adrien Gaidon

INRIA, LEAR project team

PASCAL VOC2008 challenge workshop
Introduction

- Detection task on PASCAL VOC2008 challenge
- Method with sliding windows (Each window is classified as containing or not the targeted object)
- Learn a classifier by providing positive and negative examples
Training outline

Training windows

Shape

Filtering classifier: Linear
Training outline

Training windows

Shape
- Filtering classifier : Linear

Appearance
- Scoring classifier : Non linear
Generating training windows

Training windows

- Shape
  - Filtering classifier: Linear

- Appearance
  - Scoring classifier: Non-linear
Generating training windows

- Adding positive training examples by shifting and scaling the original annotations [Laptev06]
- Negative examples randomly extracted from background
- Training an initial classifier
- Retraining 4 times by adding false positives

Examples of false positives
Image representation

Training windows

Shape

Filtering classifier : Linear

Appearance

Scoring classifier : Non linear
Image representation: Histogram-of-gradients (HOG)

- Tiling optimized per class (around 100 overlapping tiles)
- Computed with integral histograms
- With 16 orientations
**Image representation: Dense SIFT**

- Computed over dense patches (shift step 6 pixels, scale step 1.2)
- Discretized into 100 visual words using k-means
- Used as BOW with a spatial pyramid [Lazebnik06]
Learning procedure

Training windows

Shape

Filtering classifier: Linear

Appearance

Scoring classifier: Non linear
Learning procedure

- Training one classifier per view (Side, Front/Rear, Unspecified)
- Linear SVM classifier
  - HOG only (combining with SIFT gives minor gain at high cost)
- Non Linear SVM classifier
  - We use non linear $X^2$ kernel SVM [Zhang et al 2007]
  - Training with:
    - Examples used in the linear case (positives + shifted positives + hard false positives)
    - Additional random 70K negative examples
Testing outline

Test images

Generate hypotheses
- Sliding windows
- HOG
- Apply filtering classifier

Score hypotheses
- HOG + SIFT
- Apply scoring classifier

Non maxima suppression

Reweighting using classification
Evaluation of linear/non linear SVM

• Using HOG only to learn the non linear classifier
• Linear classifier used not only for filtering but also for scoring

<table>
<thead>
<tr>
<th></th>
<th>Linear</th>
<th>Linear + X²</th>
</tr>
</thead>
<tbody>
<tr>
<td>All classes</td>
<td>0.139</td>
<td>0.220</td>
</tr>
<tr>
<td>aeroplane</td>
<td>0.039</td>
<td>0.184</td>
</tr>
<tr>
<td>horse</td>
<td>0.249</td>
<td>0.435</td>
</tr>
<tr>
<td>diningtable</td>
<td>0.096</td>
<td>0.108</td>
</tr>
<tr>
<td>pottedplant</td>
<td>0.100</td>
<td>0.118</td>
</tr>
</tbody>
</table>
Evaluation of linear/non linear SVM
Evaluation of descriptors

- Scoring classifiers learned with different features
- Applied on the same hypotheses

<table>
<thead>
<tr>
<th>All classes</th>
<th>HOG</th>
<th>SIFT</th>
<th>HOG+SIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>aeroplane</td>
<td>0.184</td>
<td>0.298</td>
<td>0.338</td>
</tr>
<tr>
<td>car</td>
<td>0.475</td>
<td>0.425</td>
<td>0.511</td>
</tr>
<tr>
<td>train</td>
<td>0.318</td>
<td>0.344</td>
<td>0.291</td>
</tr>
<tr>
<td>bus</td>
<td>0.432</td>
<td>0.397</td>
<td>0.423</td>
</tr>
</tbody>
</table>
Evaluation of descriptors

Precision recall curve for the class car

- HOG + SIFT
- HOG
- SIFT
Combining localization and image classification

- Provides contextual information
- Results are more reliable in image classification
- Transform scores into probabilities
- New score = \( P(\text{det}) \times P(\text{cls}) \)
Influence of the use of image classification score

- We use the Lear_flat submission [Gaidon and Marszalek]

<table>
<thead>
<tr>
<th></th>
<th>HOG+SIFT</th>
<th>HOG+SIFT plusclass</th>
</tr>
</thead>
<tbody>
<tr>
<td>All classes</td>
<td>0.264</td>
<td>0.290</td>
</tr>
<tr>
<td>cow</td>
<td>0.240</td>
<td>0.309</td>
</tr>
<tr>
<td>sheep</td>
<td>0.212</td>
<td>0.273</td>
</tr>
<tr>
<td>car</td>
<td>0.511</td>
<td>0.518</td>
</tr>
<tr>
<td>motorbike</td>
<td>0.417</td>
<td>0.427</td>
</tr>
</tbody>
</table>
Example of results: cars

Top true positives

First false positives

Missed
Example of results: cow

Top true positives

First false positives

Missed
Example of results: chair

Top true positives

First false positives

Missed
Summary

• Two stage classification (hypotheses prediction, hypotheses verification)
• Image representation combination
• Reweighting with classification score
• Worse performance on some articulated classes (part models more suited)
• Outperform other competitors on most of the rigid classes
Thank you