Cascaded Sparse Spatial Bins For Efficient And Effective
Generic Object Detection
David Novotný1, Jiří Matas2
1Visual Geometry Group, University of Oxford
2Center for Machine Perception, Czech Technical University, Prague

Introduction

- A novel object proposal method.
- Efficiency is achieved by the use of spatial bin pooling in a novel combination with sparsity-inducing group normalized SVM.
- Boundary Edge Vector (BEV), a new HoG-like “objectness” descriptor, proposed.
- State-of-the-art results on VOC07 and ILSVRC13 achieved.

Overview

- Window scoring method.
- Each bounding box described by “objectness” features pooled from learned spatial bins.
- Train time: \( \ell_1/\ell_2 \) normalized SVM automatically selects the set of relevant spatial bins.
- The pooled features are scored by a two-stage SVM cascade.

Objectness features:

- CNN-SPP: CNN features obtained by spatial pyramid pooling [1].
- EdgeBoxes score (EB): The score by which EdgeBoxes rank proposals [2].
- Boundary Edge Vector (BEV): A novel HoG-like edge statistic.

Two-stage Cascade

Initialization:

Stage 1: Reduces the pool of boxes from 100K to 10K

- For each box:
  1. Pool CNN-SPP features from 3 selected spatial bins.
  2. Append EdgeBoxes score.
  3. Score with SVM.
  4. Based on SVM score keep top 10K boxes.

Stage 2: Reduces the pool of boxes from 10K to the final requested size

- For each box kept after stage 1:
  1. Pool BEV features from \([1, 3, 11]\) selected spatial bins.
  2. Pool CNN-SPP features from \([3, 4, 3]\) selected spatial bins.
  3. Append EdgeBoxes score.
  4. Score with SVM.
  5. Apply non-maximum suppression to obtain requested number of boxes.

Parameters \( \ell_1 \) and \( \ell_2 \) were validated such that they give best compromise between execution speed and performance.

Spatial Bin Selection by \( \ell_1/\ell_2 \) Normalized SVM

- The set of spatial bins for pooling CNN-SPP and BEV features is learned automatically.
- Group sparsity inducing \( \ell_1/\ell_2 \) SVM selects groups of dimensions that correspond to relevant spatial bins.
- Significantly speeds-up “objectness” feature extraction with negligible performance decrease.
- Group sizes: CNN-SPP ... 256 (\# of conv5 filters), BEV ... 4 (\# of orientation bins).

Boundary Edge Vector (BEV)

- A novel HoG-like edge statistic.
- Reuses the EdgeBoxes structured edge detector output — almost no additional cost.
- Quantizes edges based on their orientation and pools their intensities from selected spatial bins:

For each edge orientation:

- Score with SVM.

Learned template:

Excitation template for edge orientation

Note: BEV weights are learned to be complementary to the EdgeBoxes score and CNN-SPP descriptors.

CNN-SPP Features [1]

Box descriptors formed by spatial pyramid pooling of CNN conv5 features:

Overall-Recall Curves

Proposed methods (solid lines in plots):
- SSPB (Sparse SpAtial Bins) Basic method.
- SSPB0: SSPB with non-max suppression (NMS) threshold optimized for a small number of candidates.
- SSPB+0+S: SSPB + NMS threshold optimized for a small number of candidates.

All SSPB variants are trained solely on the trainval set of VOC2007.

References


This work was sponsored by Xerox, S.A.S.

The code will be available shortly on http://cmp.felk.cvut.cz/software/SSPB