Segmentation Driven Object Detection with Fisher Vectors

Ramazan Gokberk Cinbis Jakob Verbeek Cordelia Schmid
LEAR Team - INRIA Grenoble, LJK

Contributions

- Use unsupervised segmentation to improve the features for object detection.
- Explore *Improved Fisher Vectors* [1] for object detection.

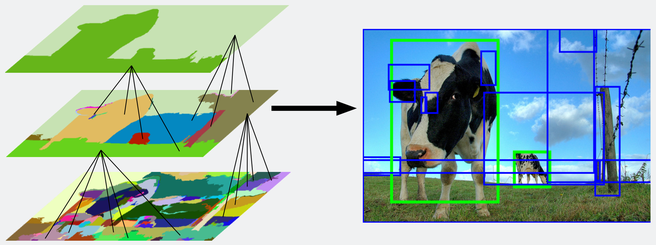
Results

- Improve state-of-the-art on PASCAL VOC 2007 and 2010 datasets.
- Mask-based descriptors contribute consistently ∼2% mAP points.

Candidate windows

Hierarchical segmentation-based candidate window generation [2]:

- Partition image into superpixels.
- Group superpixels into a segmentation tree.
- Get 8 segmentations, using 4 color spaces and 2 scale parameters
- Bounding boxes of the segments are used as the candidate windows.
- Results in ∼1500 windows per image.



Feature extraction

Local Descriptors

- SIFT & color local descriptors, each PCA-ed to D=64 dimensions.
- Dense multi-scale grid: 16 scales, 12 $\times$ 12 patches at the smallest scale.

Fisher Vectors (FV)

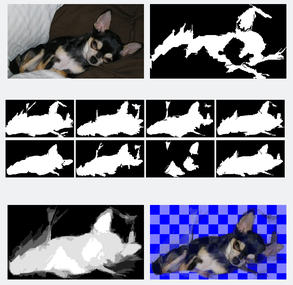
- Extends bag-of-words histograms by adding 1st and 2nd order moments.
- We use Gaussian mixtures with K=64 centers.
- We use hard-assignment to speed-up feature pooling.

Window Descriptors

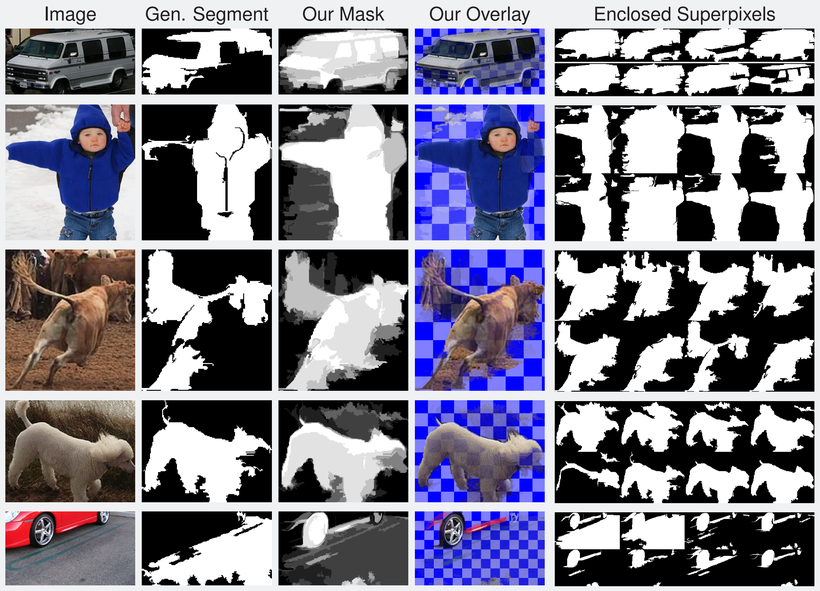
- Weight SIFTs in FV equally or according to our mask (see next column).
- Apply the power and ℓ_2 normalization per FV [1].
- $1 \times 1 + 4 \times 4$ grid over the window (SIFT-only).
- Capture global scene context using a FV over the entire image.
- Each FV for each cell results in a $K(2D+1)$ dimensional vector.
- Concatenate all the FVs, total dimensionality ∼300.000

Segmentation mask generation

- The segments used to generate these candidate windows [2] generally do not provide good object segmentations.
- Instead, suppress superpixels that straddle the window boundary, that are likely to contain background. Related to objectness measure [4].
- Use the mask averaged over eight binary masks to weight the contribution of local features in the window descriptor.



Example Images



- Since segmentation may suppress the background even in incorrect object hypotheses, it is important to combine with features from the entire window.

Feature compression and training

- We need to apply detectors several times for hard negative mining.
- Re-extracting descriptors for ∼8 million windows at each iteration very costly.
- Instead, use data compression techniques: PQ (lossy) + Blosc (lossless).
- Modified Liblinear SVM training: Decompress individual examples on-the-fly.
- With data in RAM, detection over 5000 images in 5 mins on 35 cores.

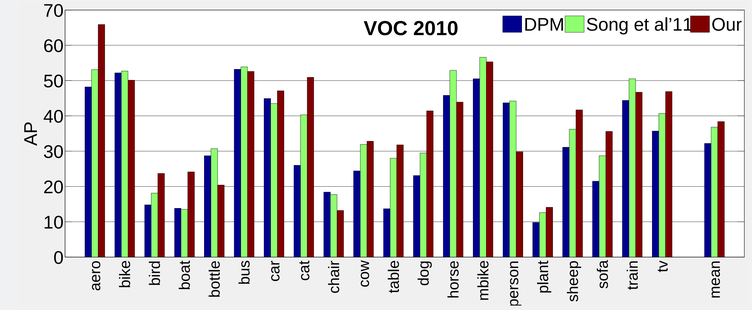
Feature evaluation

- PASCAL VOC 2007 dataset.
- S: SIFT, C: Color
- W: Window, M: Mask, F: Full Image
- Con: Contextual rescoring of [3].
- Our segmentation-based descriptors improve performance by approximately 2 mAP points.

		mAP
S	W	34.0
S	W+M	35.8
S+C	W	35.2
S+C	W+M	37.6
S+C	W+F	36.6
S+C	W+M+F	38.5
S+C	W+M+F+Con	**40.5**

Comparison with state of the art

	VOC'07	VOC'10	Description
w/o inter-class context			
Sande et al.'11	33.7	34.1	Same candidate boxes w/ bag-of-words.
Vedaldi and Zisserman'12	27.7	-	HOG detector w/ PQ encoding.
Girshick et al.'12	33.7	29.6	Deformable parts model (DPM).
Song et al.'13	34.7	29.4	Discriminatively trained and-or tree models.
Ours	38.5	35.8	Our model (S+C,W+M+F).
with inter-class context			
NLPR 2010	-	36.8	HOG+LBP detector.
Song et al.'11	37.7	36.8	Integrated detection & image classification.
Girshick et al.'12	35.4	32.2	Deformable parts model (DPM).
Chen et al.'13	38.7	-	Multi-order contextual co-occurrence.
Ours	**40.5**	**38.4**	Our model (S+C,W+M+F+Con).
uncomparable methods			
Fidler et al.'13	-	40.1	Fully-supervised segmentation model, uses additional train images.



References

[1] J. Sánchez and F. Perronnin and T. Mensink and J. Verbeek, *IJCV* 2013.
[2] K. van de Sande and J. Uijlings and T. Gevers and A. Smeulders, *ICCV* 2011.
[3] Felzenszwalb, P. F. and Girshick, R. B. and McAllester, D. and Ramanan, D., *PAMI* 2010.
[4] B. Alexe and T. Deselaers and V. Ferrari, *PAMI* 2012.

Improved detections using masks

Failure Cases