Segmentation-aware Deformable Part Models

Eduard Trulls[1], Stavros Tsogkas[2], Iasonas Kokkinos[2], Alberto Sanfeliu[1], Francesc Moreno-Noguer[1]

[1] Institut de Robòtica i Informàtica Industrial, Barcelona, Spain / [2] Center for Visual Computing, Ecole Centrale de Paris/INRIA-Saclay, France

CONTRIBUTIONS

1. We present a method to build **fast, soft segmentation masks** from **SLIC superpixels.**
2. We build **background-invariant features for DPM**, combining *segmentation* and *detection*. Increase of **1.7% AP** over PASCAL VOC 2007.

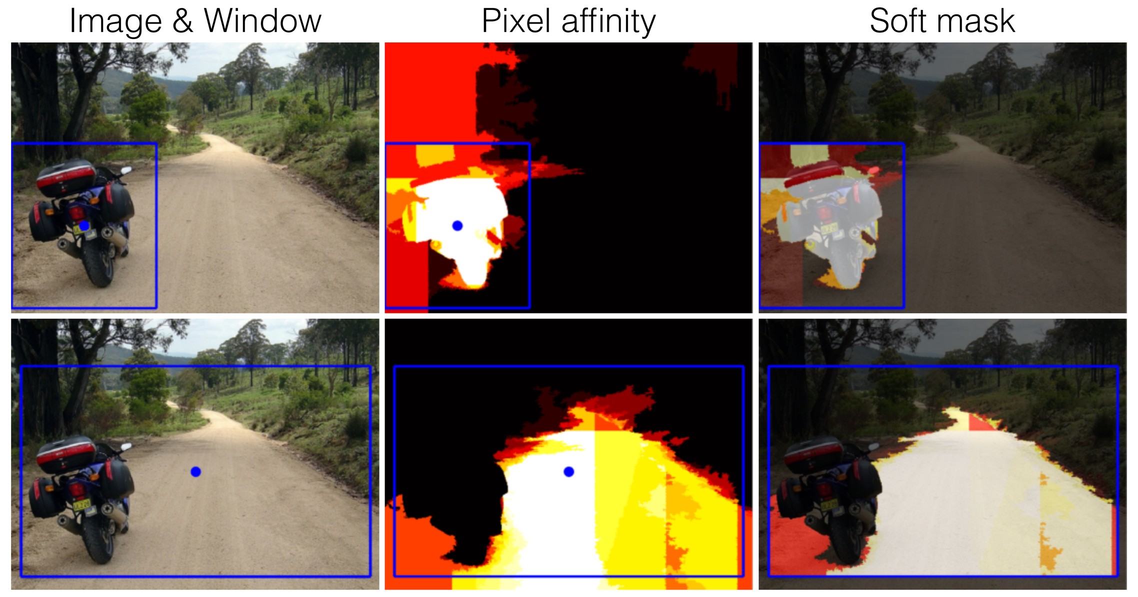
KEY FEATURES

- **Simple:** two parameters, adjusted per class.
- **Fast:** ~2.1 sec. for 266K iterations (root filter only). Convolution w/ filters takes longer: ~5 sec.
- **Generic:** we show how to extend it to dense SIFT.

Code is available: https://github.com/etrulls

BUILDING THE MASKS

We compute multi-scale SLICs *once*. The masks are computed for **every object hypothesis.** This is **fast**, i.e. **can be repeated at every displacement/scale.**



SEGMENTATION-AWARE DPM

DPMs represent objects as star-shaped graphical models of a **root** and **deformable parts**. Score is given by the match with the template filters (unary) and the deformation of the parts (pairwise).

$$S(x_0, x_1, \ldots, x_n) = \underbrace{\sum_{p=0}^{n} \langle w_p, G(x_p) \rangle}_{\text{unary match}} + \underbrace{\sum_{p=1}^{n} D_p(x_p, x_0)}_{\text{part deformation}}$$

Goal: better unary costs. **Method:** split features into figure/ground channels. **Features:**

$$G^{seg}[i] = [G[i], G^{+}[i], G^{-}[i], M[i]]$$

For both **root and parts.** Cross-validate λ per class. We build upon the code of Felzenszwalb et al [2].

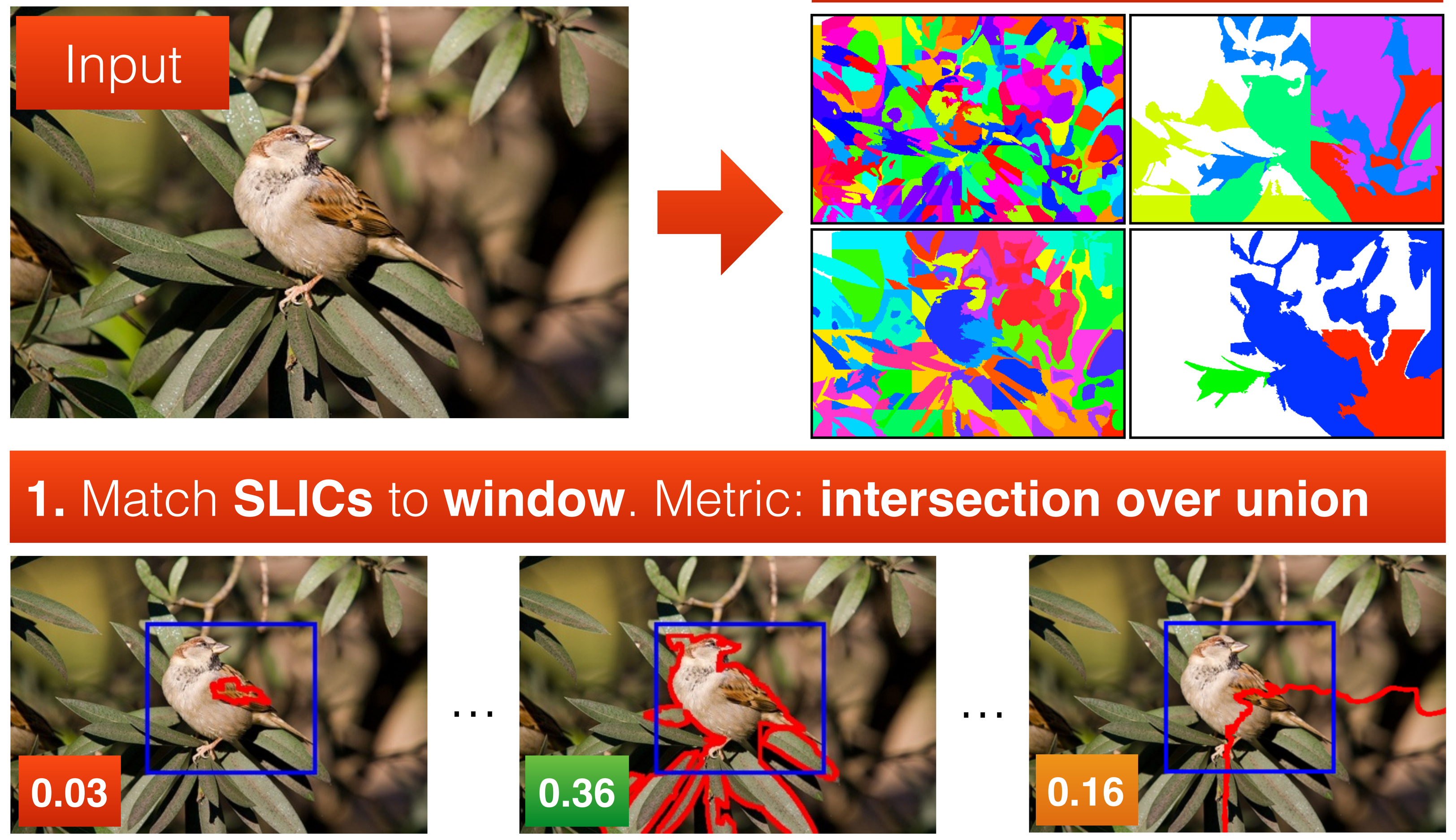
SECOND PARAMETER: ALPHA-BLENDING

Problems with some categories. Idea: **'alpha-blending.'**

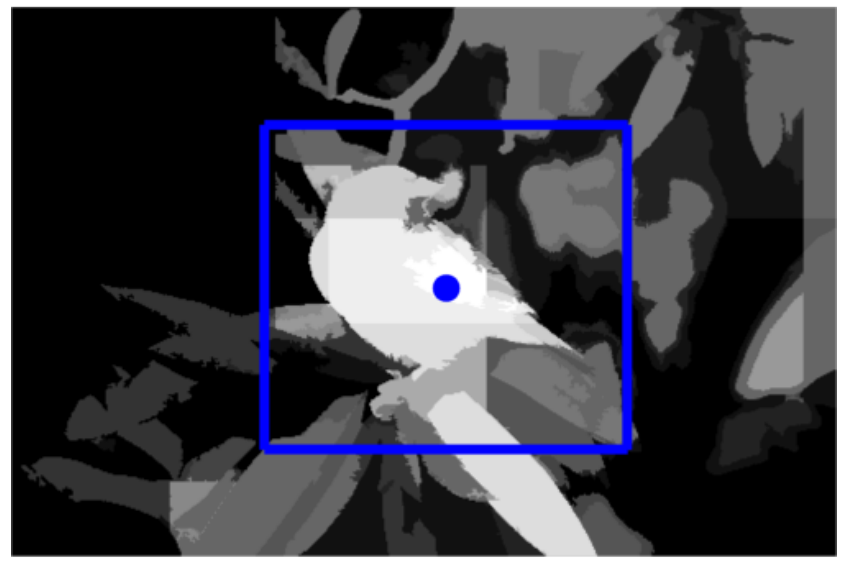
$$M_{\alpha}[i] = (1 - \alpha)1[i] + \alpha M[i]$$

i.e. α=1 our full-blown approach; α=0 segmentation-agnostic. Also cross-validated per-class, after λ.

METHOD OVERVIEW

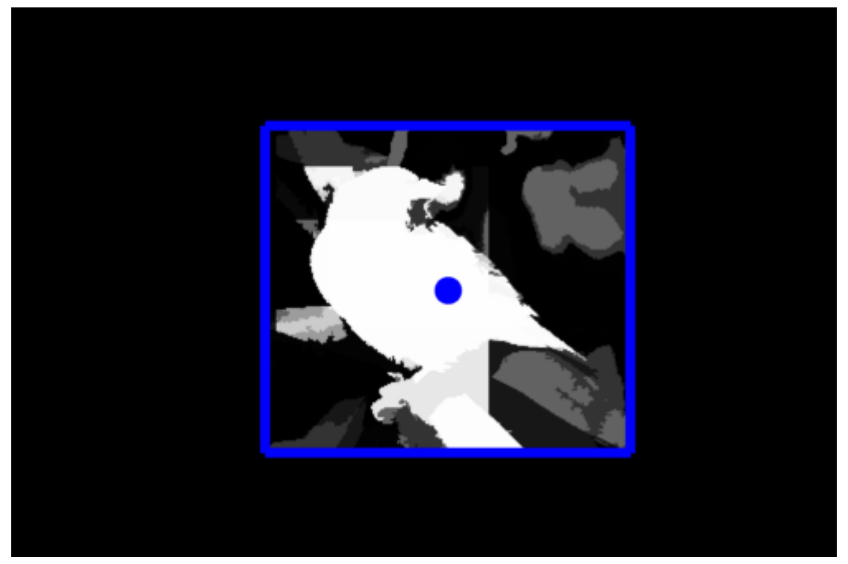


2. Affinity to center of window



Given a **bounding box hypothesis:**
1. Score the superpixels by how well they match the window.
2. Aggregate the **best-ranked superpixels** to build an **affinity measure** $f[i]$.

3. Soft mask for this window

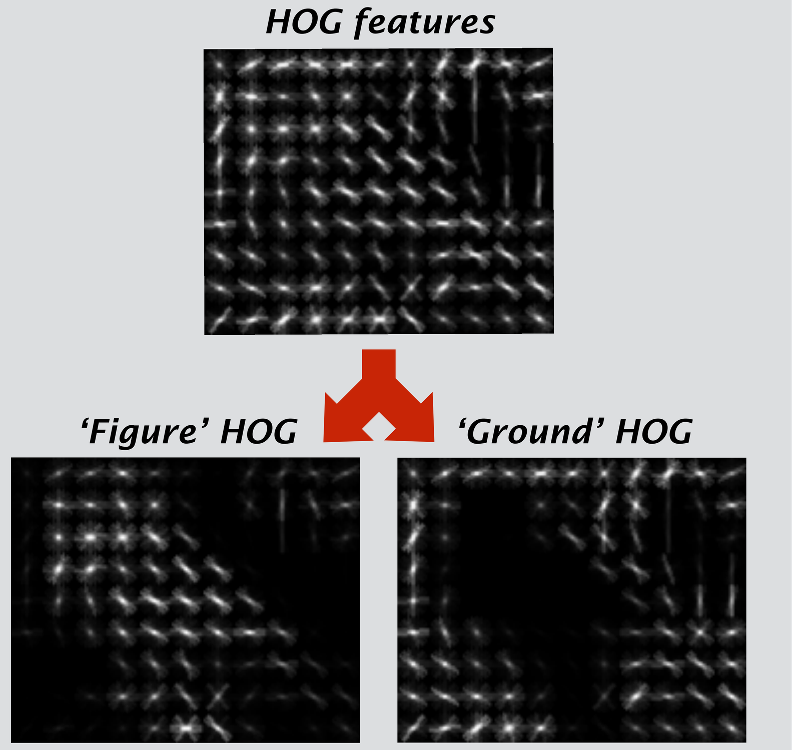


Build a **soft mask** from the affinity:

$$M[i] = 1 - \frac{1}{1 + \exp\left(-\frac{10}{1-\lambda}(f[i] - \lambda)\right)}$$

Where λ determines the 'hardness'.

Goal: **background-invariant features** for DPM. Practical solution: split HOG into **figure/ground channels.**



$G^{+}[i] = M[i] \cdot G[i]$ $\quad$ $G^{-}[i] = (1 - M[i]) \cdot G[i]$

RESULTS: DPM

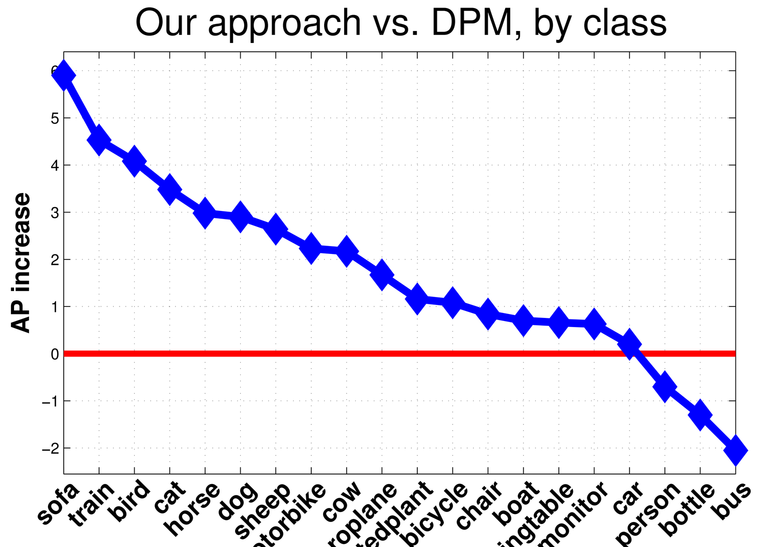
We use DPM [5] as a baseline. **Average increase of 1.7% AP** on PASCAL VOC 2007.

Average AP for DPM: 32.0

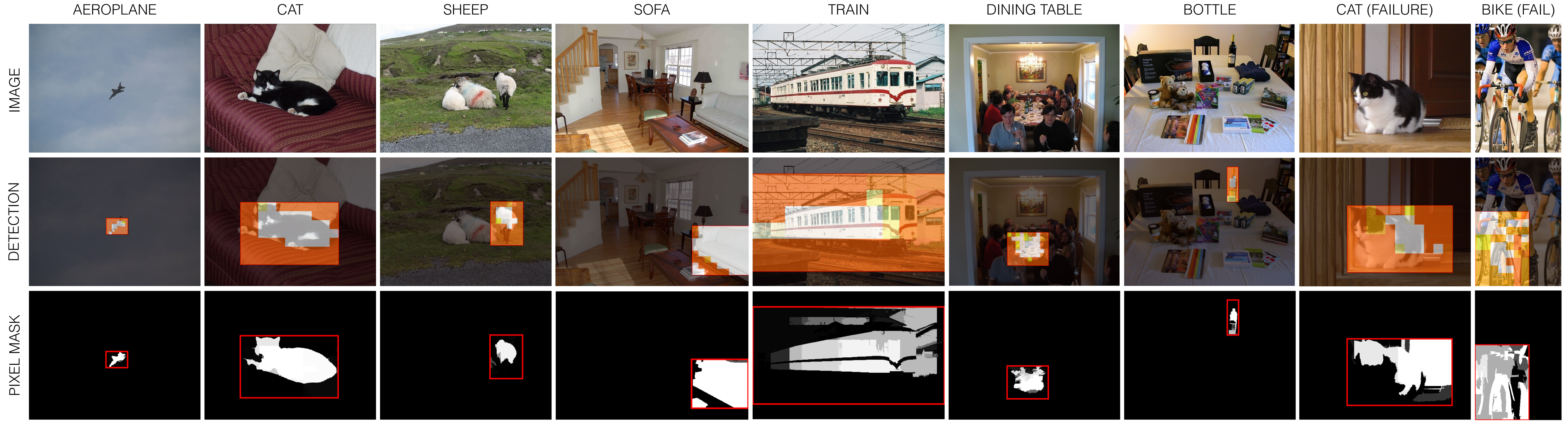
Ours: 33.7

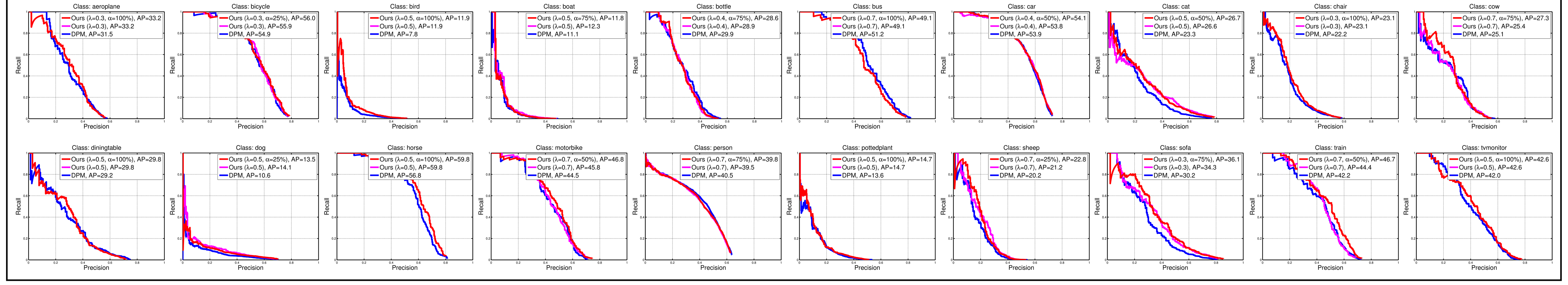
	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow
DPM	31.5	54.9	7.8	11.1	**29.9**	**51.2**	53.9	23.3	22.2	25.1
Ours	**33.2**	**56.0**	**11.9**	**11.8**	28.6	49.1	**54.1**	**26.7**	**23.1**	**27.3**
Diff.	1.7	1.1	4.1	0.7	-1.3	-2.1	0.2	3.4	0.9	2.2

	table	dog	horse	mbik	pers	plant	sheep	sofa	train	tv
DPM	29.2	10.6	56.8	44.5	**40.4**	13.6	20.2	30.2	42.2	42.0
Ours	**29.9**	**13.5**	**59.8**	**46.8**	39.8	**14.7**	**22.8**	**36.1**	**46.7**	**42.6**
Diff.	0.7	2.9	3.0	2.3	-0.6	1.1	2.6	5.9	4.5	0.6



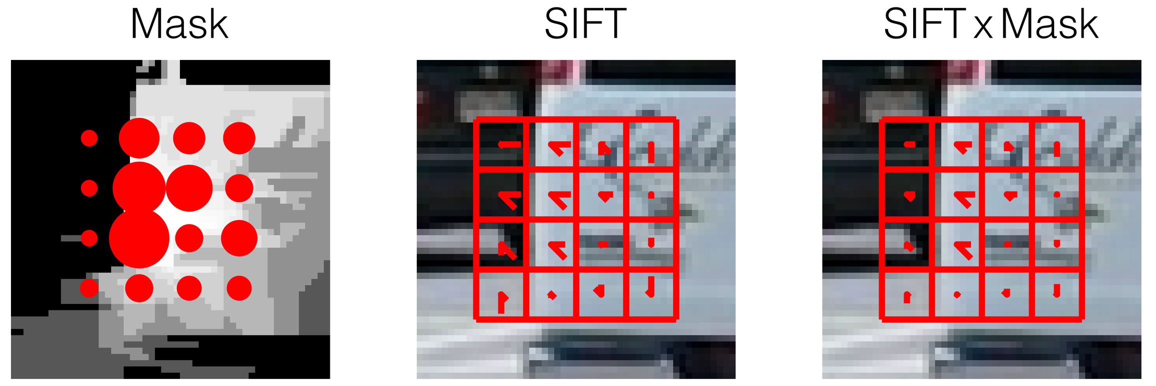
In addition to increased performance, we obtain **object segmentations** (pictured: root). Middle row: **HOG block level**. Bottom row: recomputed at the **pixel level.**



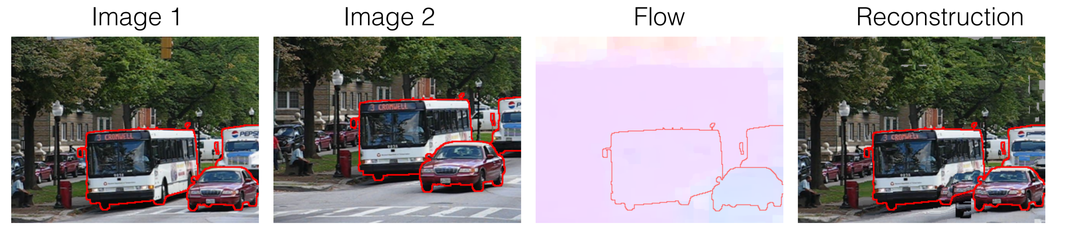


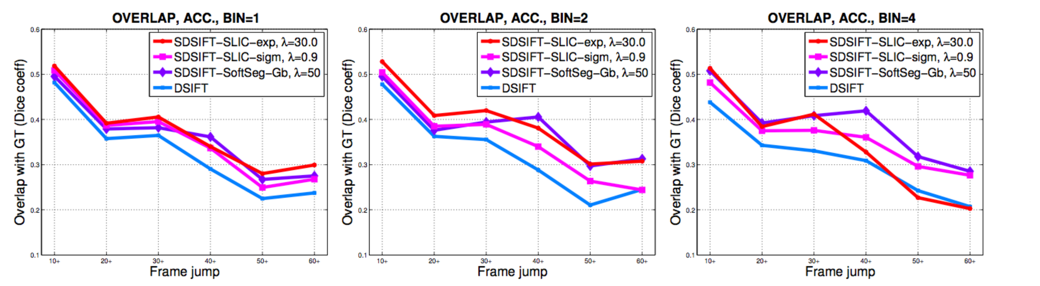
DENSE DESCRIPTORS & RESULTS

Goal is to **suppress background** structures during **descriptor construction.** How: 'gate' SIFT histogram bins with our SLIC-supported **soft segmentation masks:**



MOSEG/JHU benchmark: traffic sequences with ground truth segmentation every ~10 frames. Task: match **first & every annotated frame.** Method: SIFT-flow [3]. Metric: DICE. DSIFT vs SDSIFT-Gb [4] vs SDSIFT-SLIC (ours). Results: better than DSIFT, comparable to [4], but faster.





[1] Achanta, Shaji, Smith, Lucchi, Fua, Süsstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *PAMI*, 2012.
[2] Felzenszwalb, Girshick, McAllester, Ramanan. Object detection with discriminatively trained part based models. *PAMI*, 2010.
[3] Liu, Yuen, Torralba. SIFT flow: dense correspondence across difference scenes. *PAMI*, 2011.
[4] Trulls, Kokkinos, Sanfeliu, Moreno-Noguer. Dense segmentation-aware descriptors. *CVPR*, 2013.
[5] DPM release, v.5: http://www.cs.berkeley.edu/~rbg/latent