

A Discriminatively Trained, Multiscale, Deformable Part Model

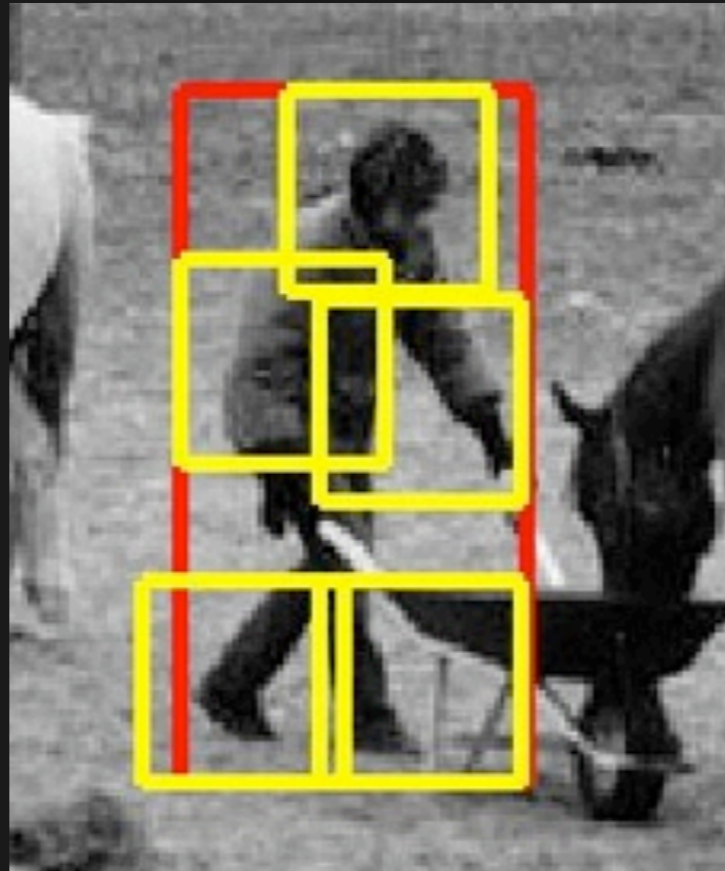
Pedro Felzenszwalb¹ & David McAllester² & Deva Ramanan^{2,3}

University of Chicago¹

Toyota Technological Institute at Chicago²

UC Irvine³

Model overview



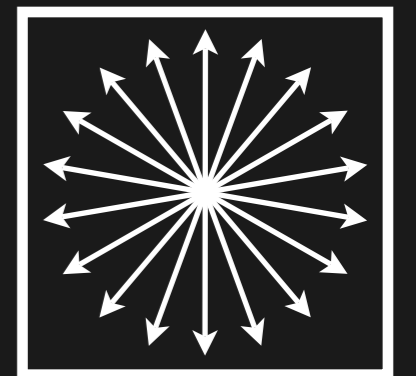
Model consists of **root filter** plus
deformable parts

We have built & tested
models for all 20 classes
(no class tuning)

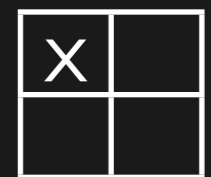
Image features - histograms of gradients

- Our implementation of DalalTriggs HOG features

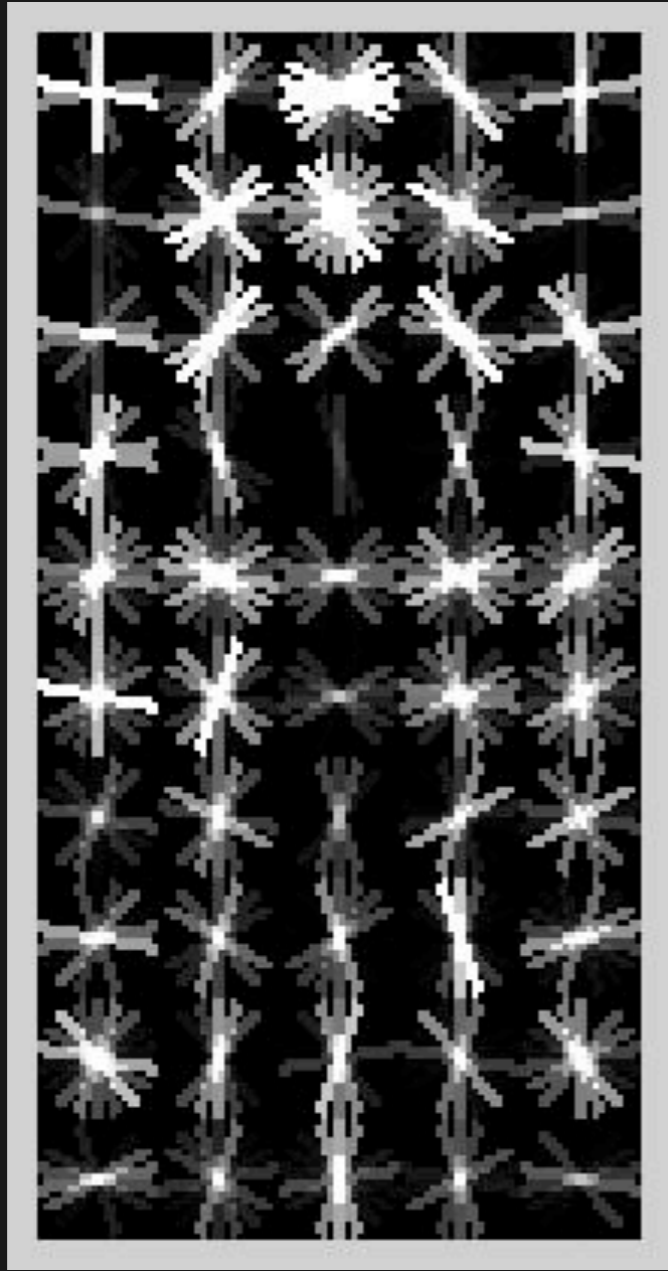
- Bin fine-scale gradients into 8x8 spatial neighborhoods and 9 orientation channels



- Normalize with respect to multiple local 16x16 regions

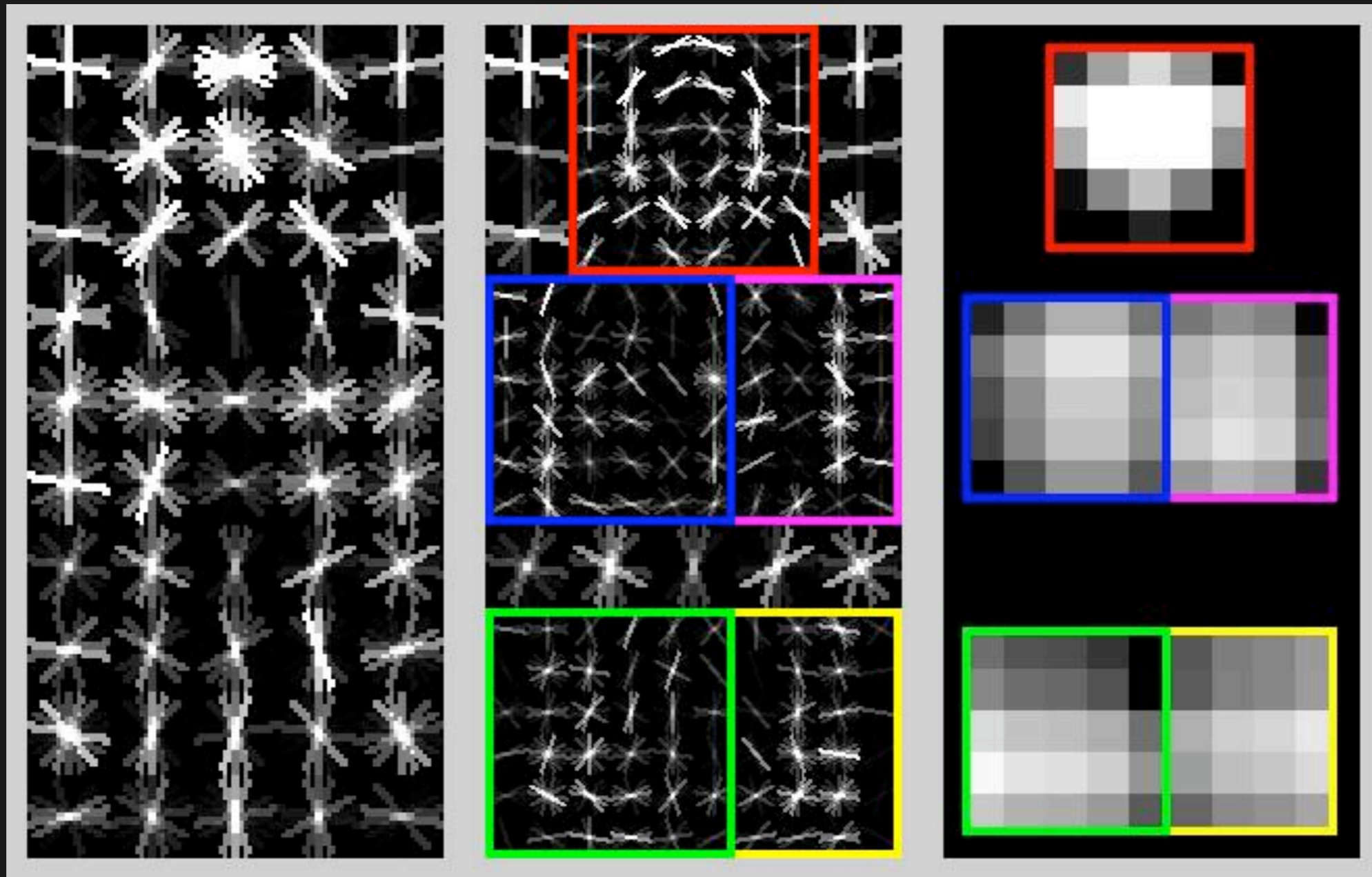


Multi-scale star model



root filter
8x8
resolution

Multi-scale star model

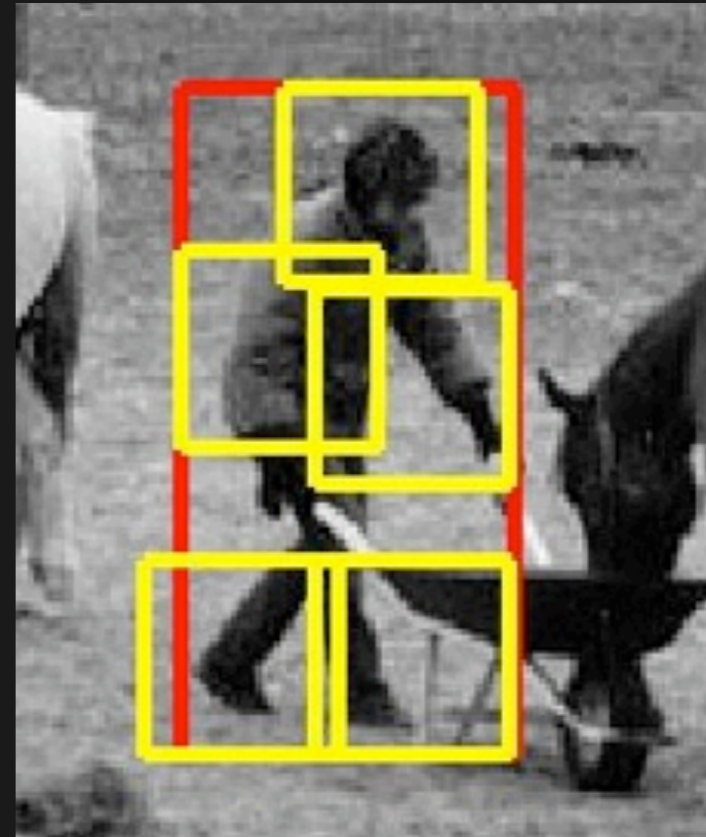


root filter
8x8
resolution

part filters
4x4
resolution

discrete
spatial model

Formal model

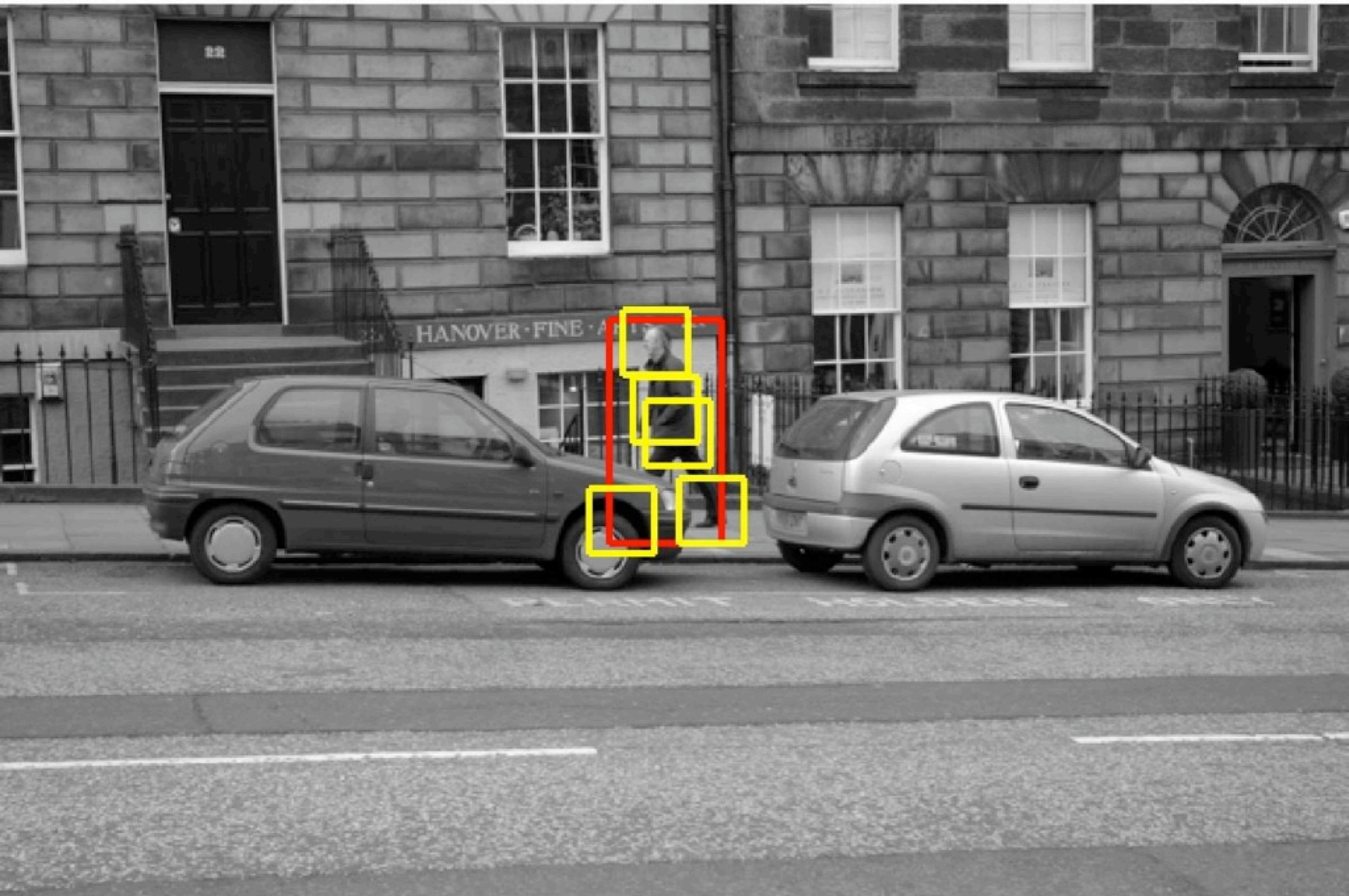


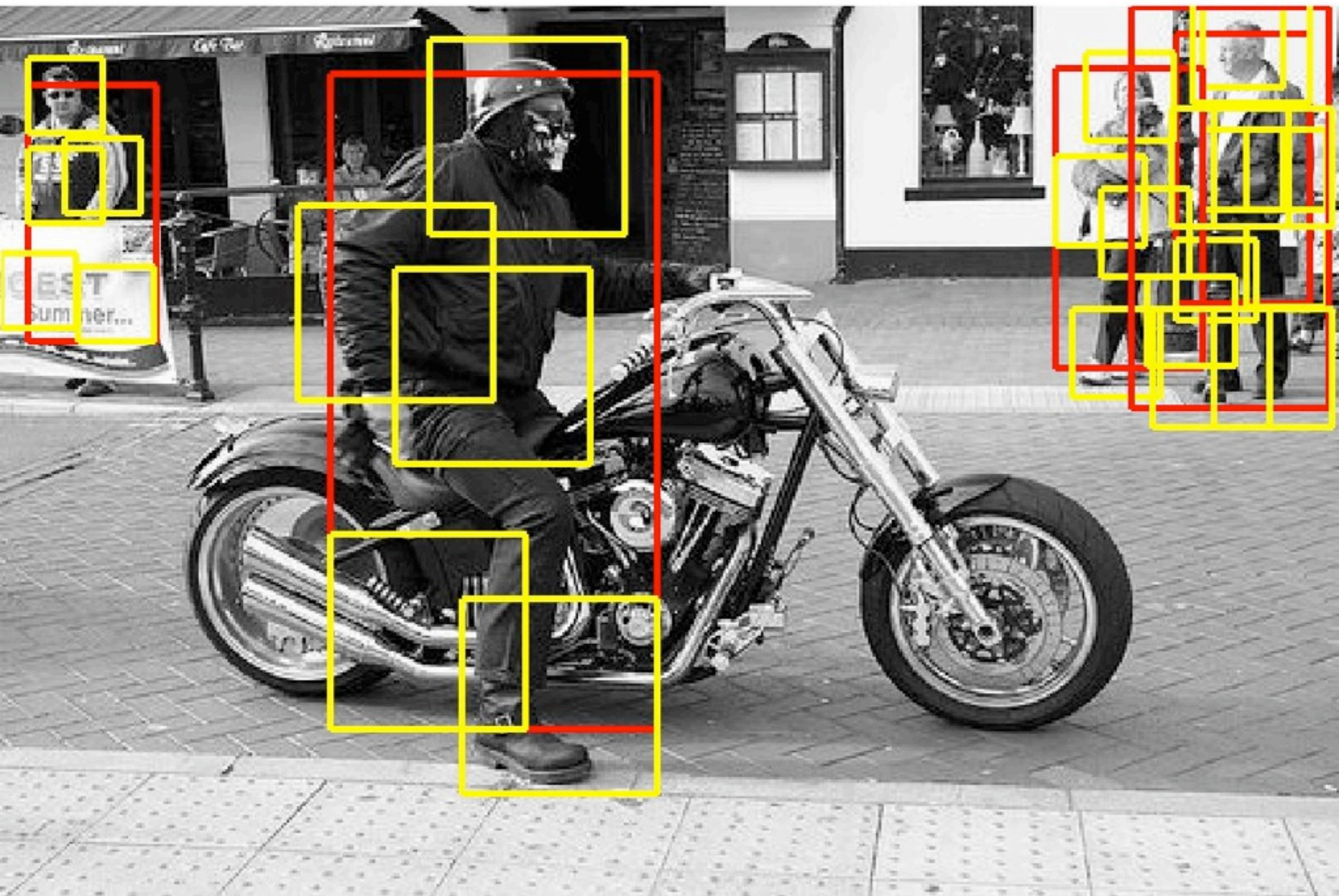
$$f_w(x) = w \cdot \Phi(x)$$

$$f_w(x) = \max_z w \cdot \Phi(x, z)$$

Z = vector of part offsets

$\Phi(x, z)$ = vector of HOG features (from root filter & appropriate part sub-windows) and part offsets

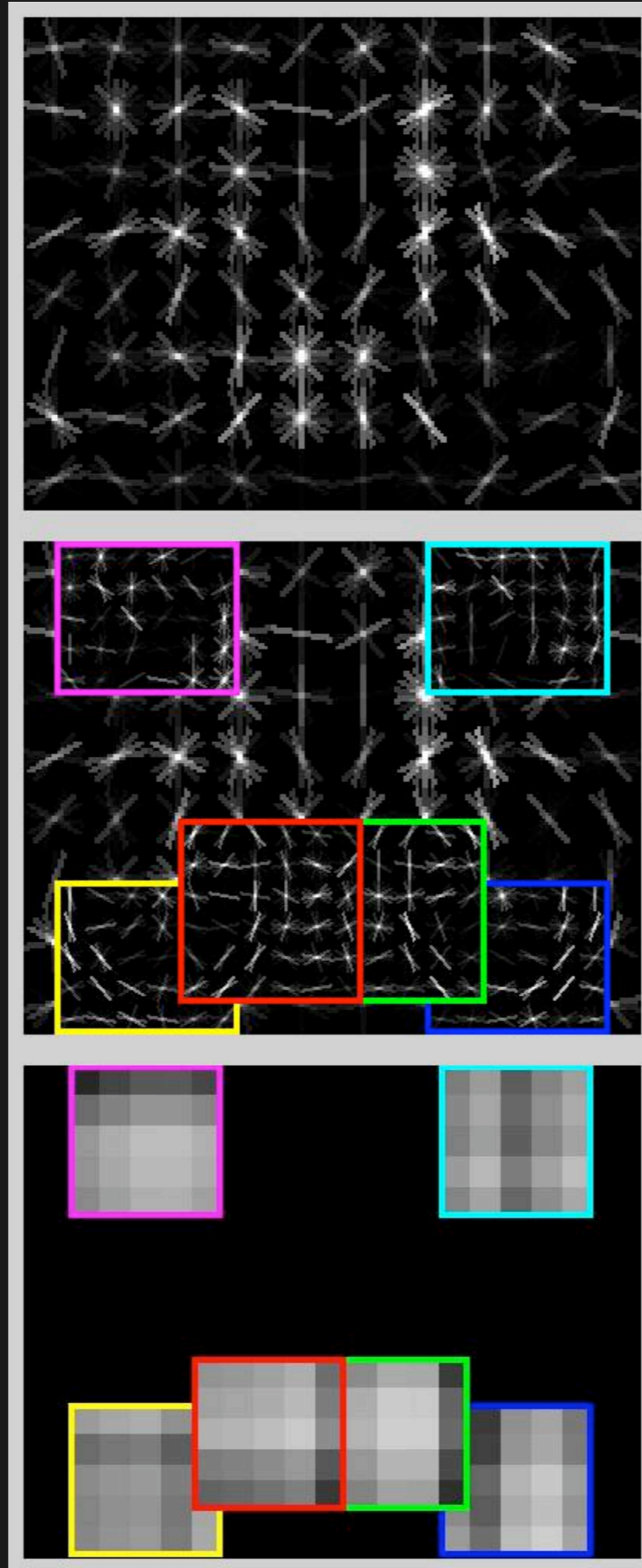




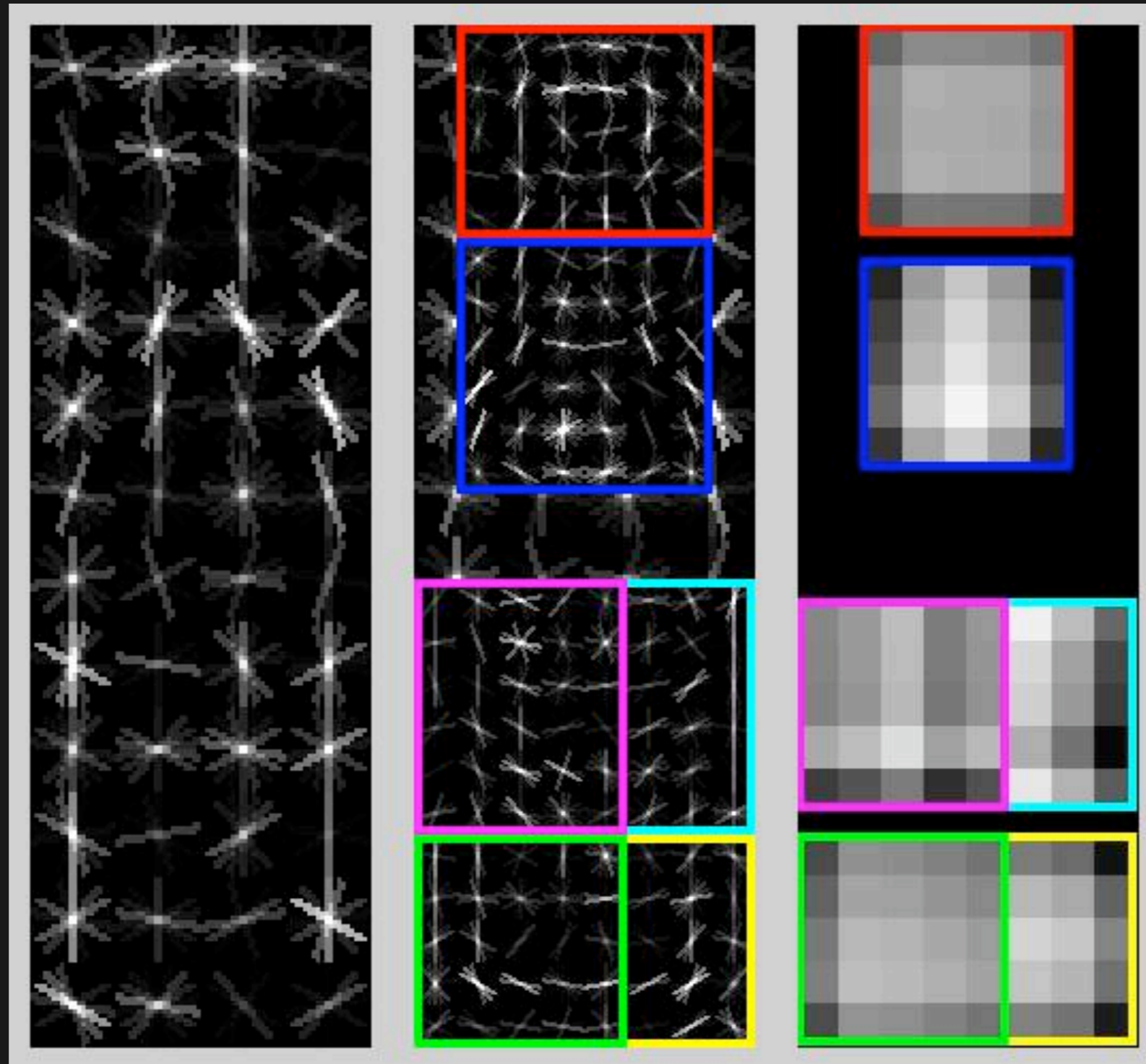
Some stats

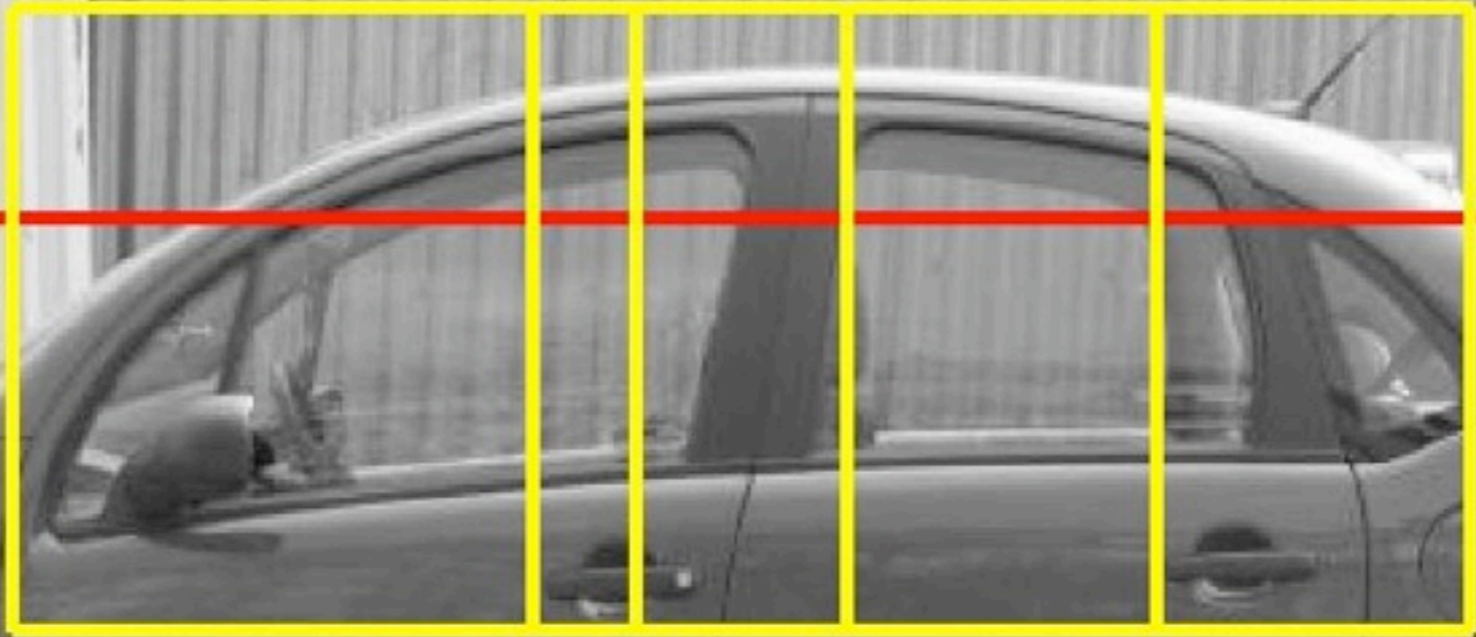
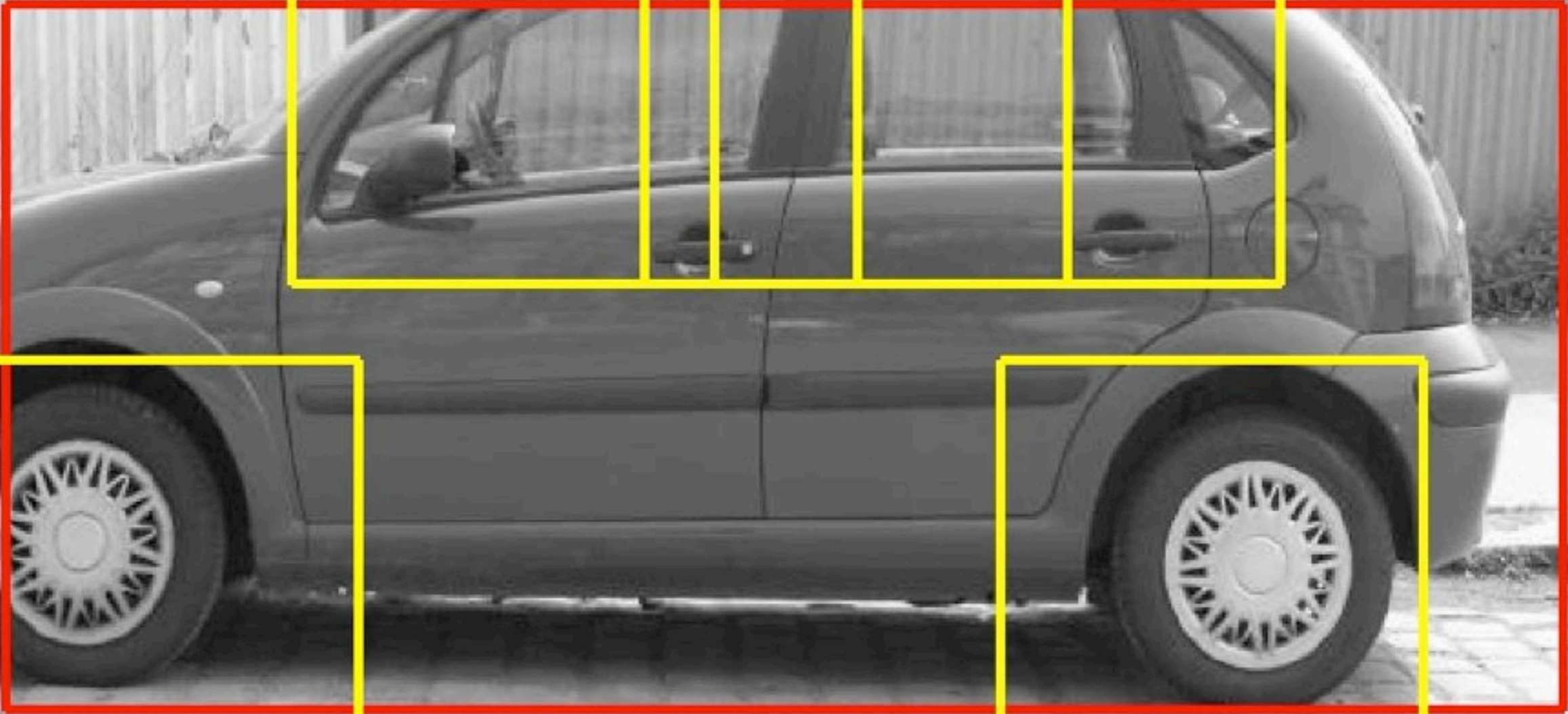
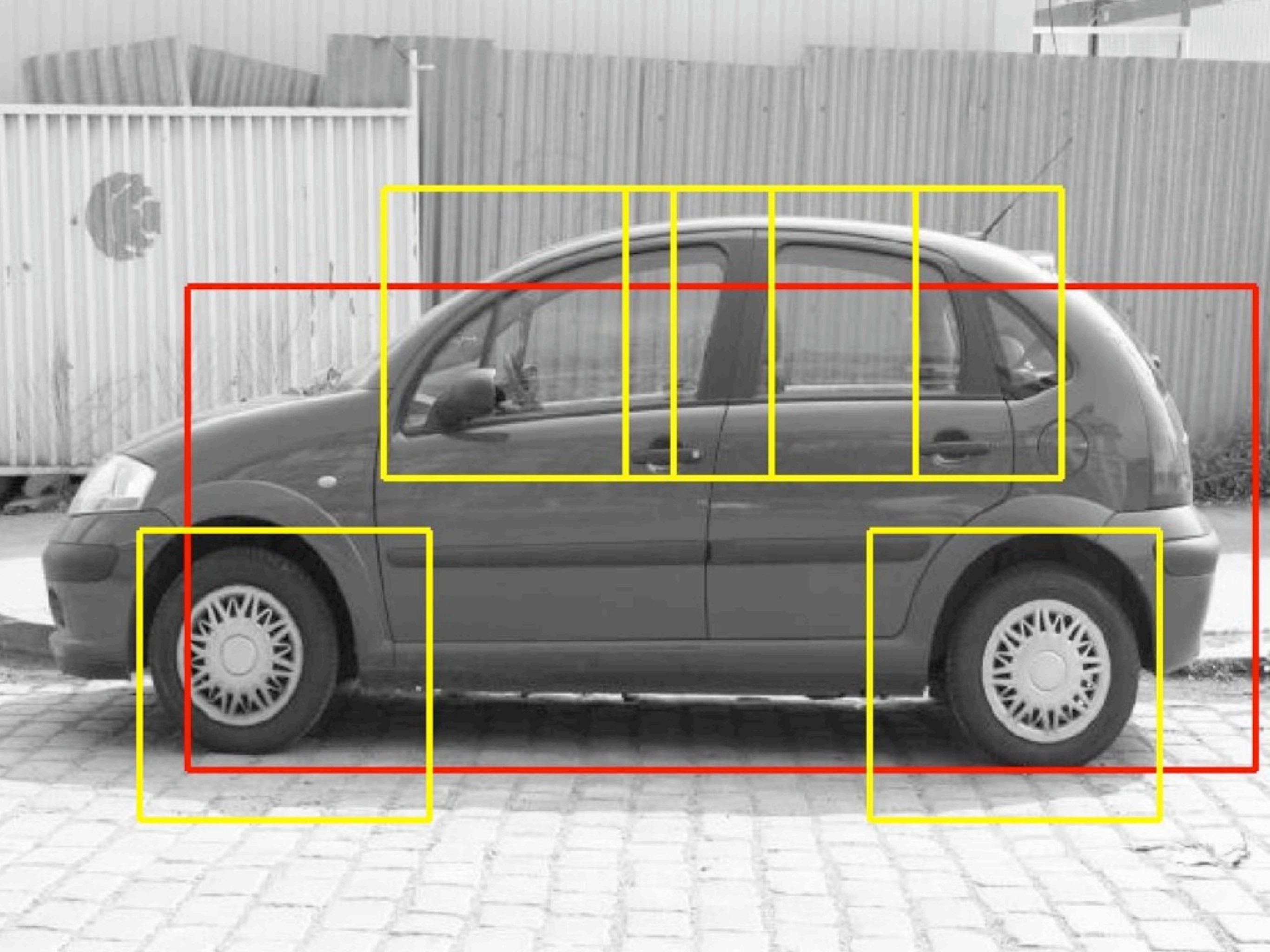
- We search over scales with an image pyramid (1.05 scaling)
- We search over 8-pixel strides for root filter, 4-pixel strides for part filter
- Training time: 3-4 hours per class using 1 cpu, including learning part models automatically
- Testing time: 2 seconds per image per model

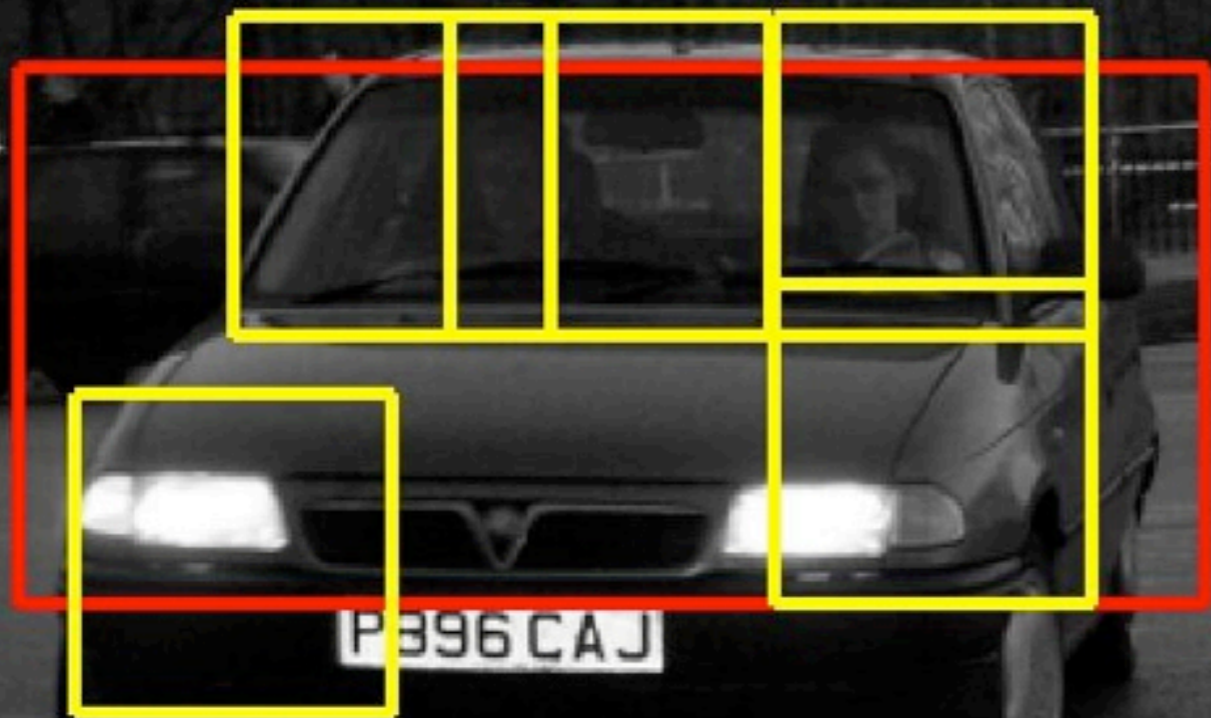
Example models



Example models







P896 CAJ

Latent SVMs

$$f_w(x) = \max_z w \cdot \Phi(x, z)$$

Assume we are given positive and negative training windows $\{x_i\}$

$$w^* = \arg \min_w \lambda \|w\|^2 + \dots$$

$$\sum_{i \in pos} \max(0, 1 - f_w(x_i)) + \sum_{i \in neg} \max(0, 1 + f_w(x_i))$$

If $f()$ is linear classifier, this is a standard SVM (convex)

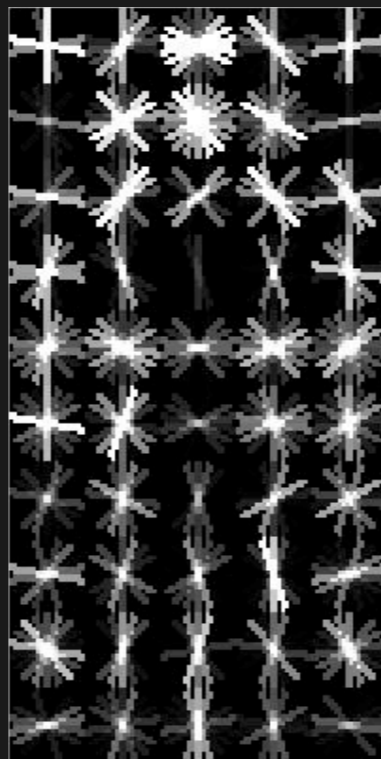
If $f()$ is an arbitrary classifier, this (in general) is not convex

If $f()$ is convex in w , the training objective is 'semi-convex'

(Instance of LeCun's Energy Based Model)

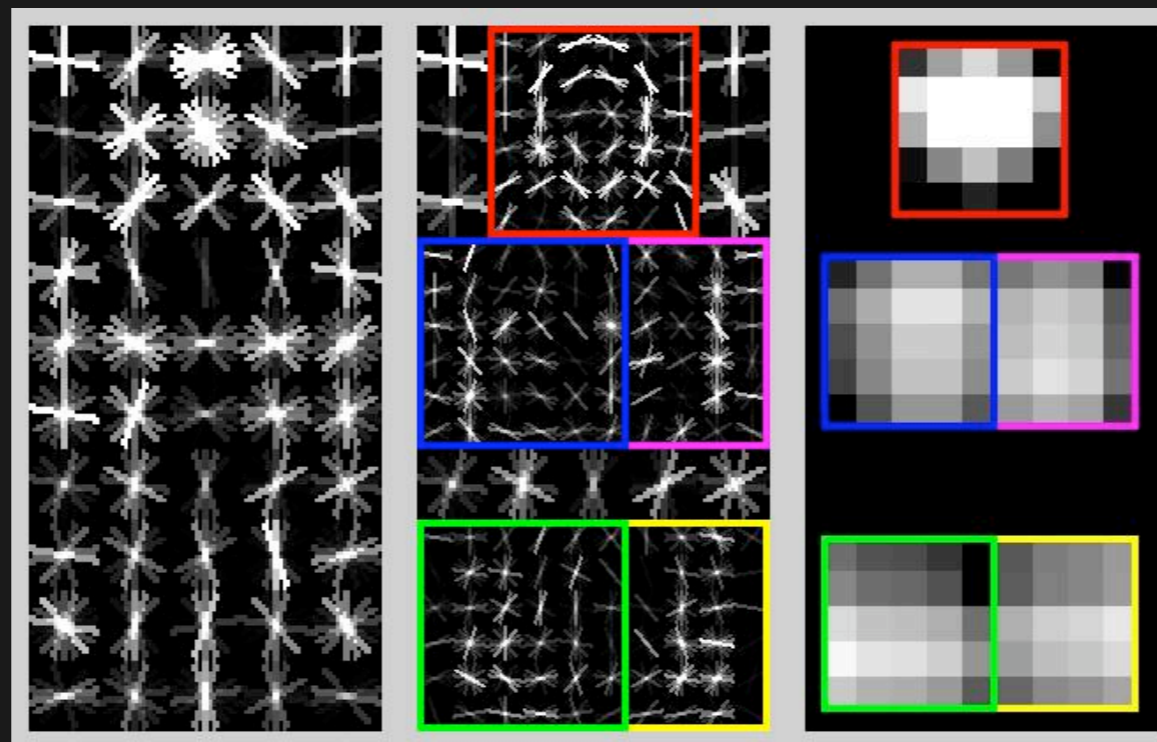
Root filter initialization

- We select the aspect and size by a heuristic tuned on 2006 data (use most common aspect and smallest area $> 80\%$ of training bounding boxes)
- Train a root filter with SVM-light: use non-truncated positives (warped to fixed aspect & size) and random negatives



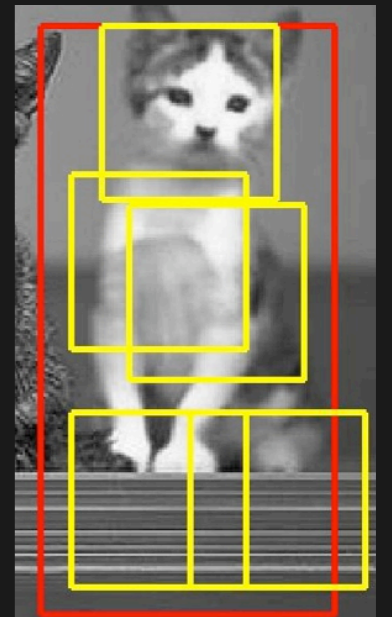
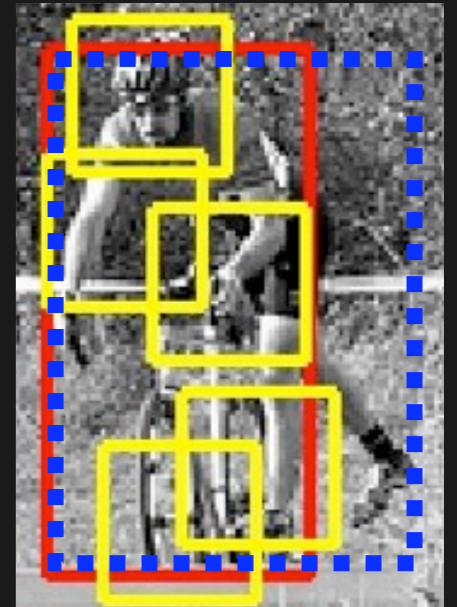
Part filter initialization

- Look for regions in root filter with lots of positive energy - part filter initialized to subwindow doubled in resolution
- Spatial model allows for a bounded offset from original anchor point - discrete deformation cost initialized to 0 everywhere

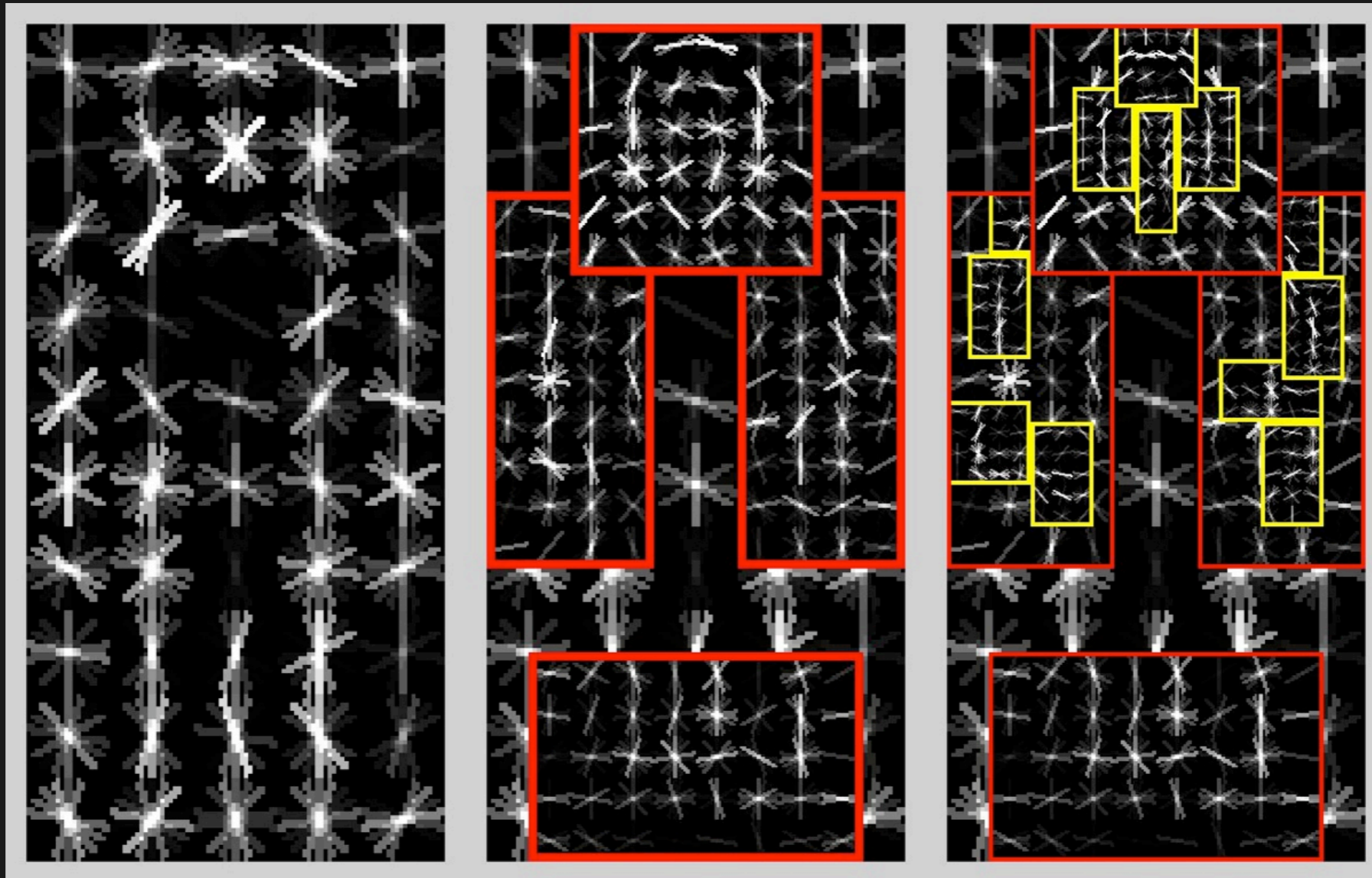


Model update

- Update positives
 - Apply current detector over all positions & scales
 - Find best-scoring $\Phi(x_i, z_i)$ that overlaps $> 50\%$ with ground truth positive bounding box
 - Allows for automatic adjustment of b. box
- Collect negative $\Phi(x_i, z_i)$'s by finding high-scoring detections, cycling through negative training images
- Use $\Phi(x_i, z_i)$'s to train a new detector (w) with SVM-light (Joachims)
- Repeat update 10 times

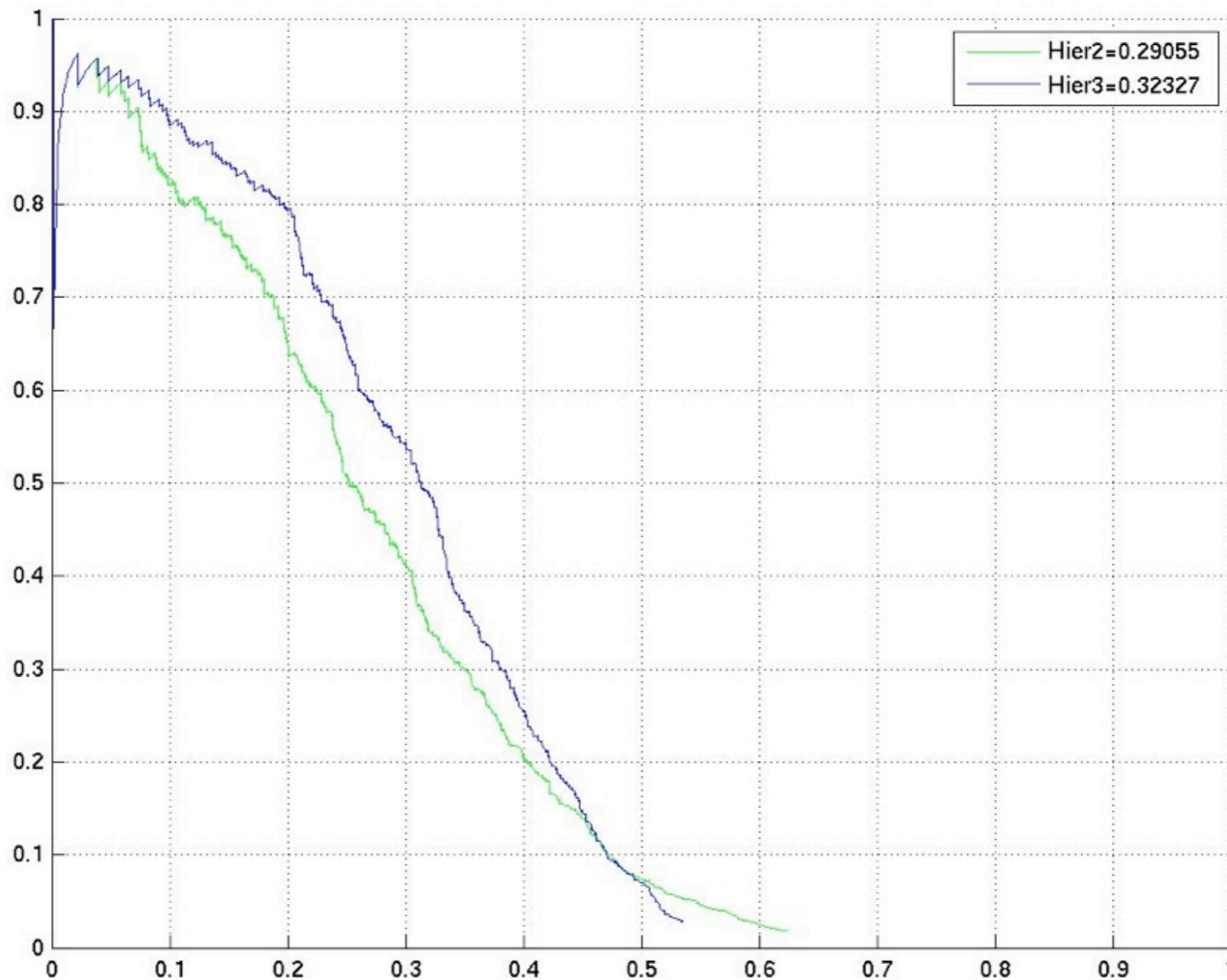


Ongoing work



Hierarchically learn parts

Ongoing work



Improves performance by 10% (AP of .29 vs .32)
Training & testing on Person2006

Conclusions

Deformable part model
Histograms-of-gradient features
Multi-scale / hierarchical
Discriminatively-trained