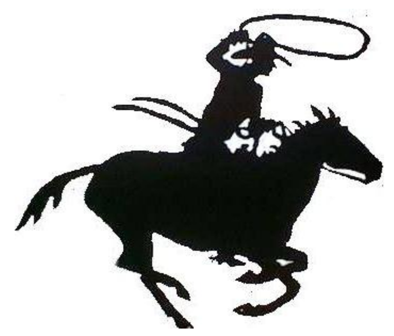
The Fastest Pedestrian Detector in the West

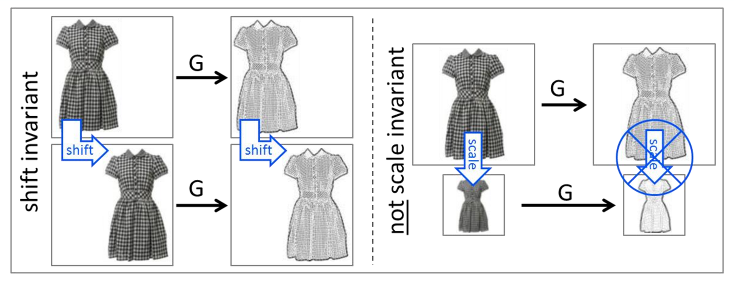


Piotr Dollár[1] **Serge Belongie[2]** **Pietro Perona[1]**

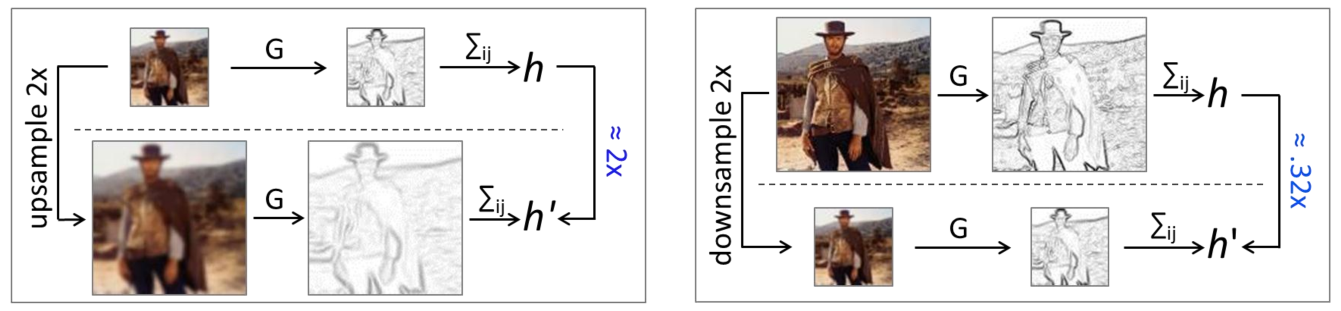
[1] Electrical Engineering, California Institute of Technology
{pdollar,perona}@caltech.edu

[2] Computer Science and Engineering University of California, San Diego
sjb@cs.ucsd.edu

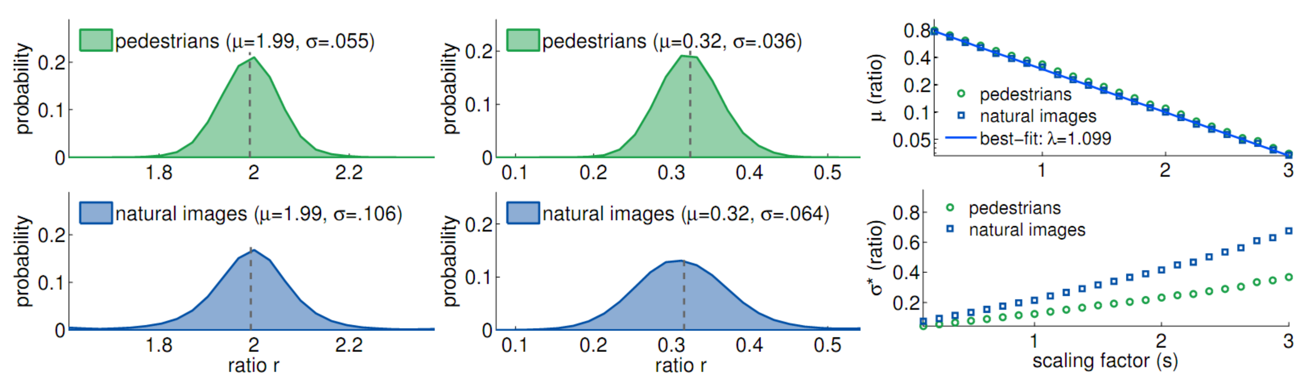
Approximating Multiscale Gradients

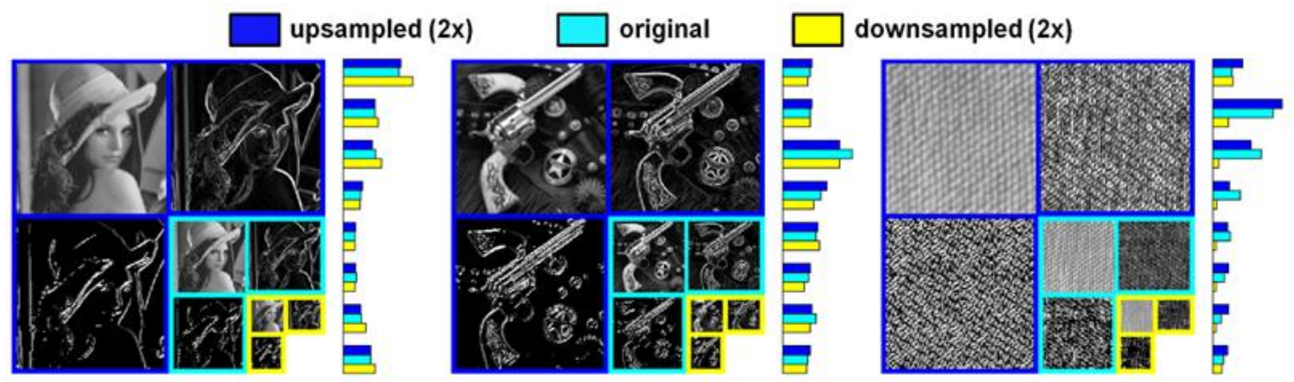


- Most features not scale invariant
- Upsampling: no information created, can predict feature response
- Downsampling: information lost, but loss consistent, can compensate!



(a) upsampling (2x) (b) downsampling (2x) (c) downsampling

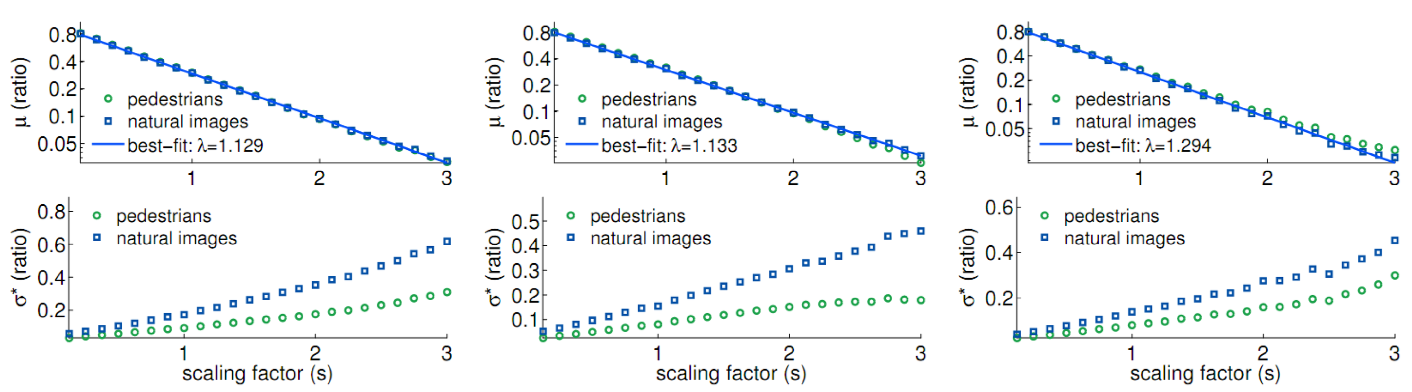




Approximating Multiscale Features

- Distribution of image statistics in natural images invariant to scale [Ruderman & Bialek 94]
- Statistics of image independent of scene area of single pixel

(a) normalized gradient (b) local standard deviation (c) HOG



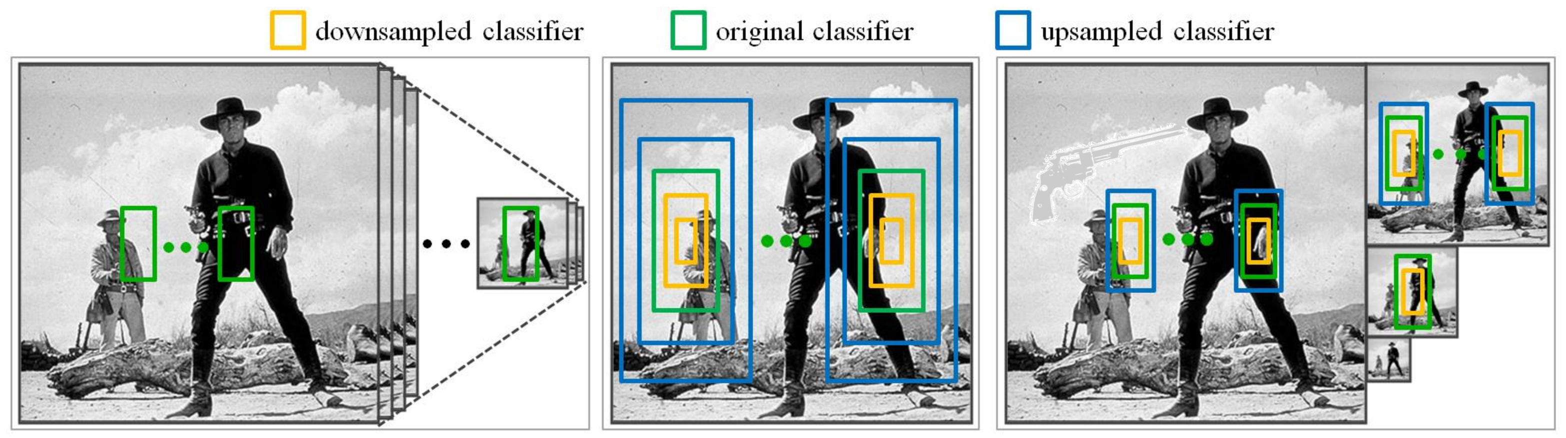
Exponential Scaling Law:

- Let $f(I,s)$ denote the channel sum computer over I after downsampling by 2^s
- Expectation of $f(I,s_1)/f(I,s_2)$ should depend only on $s_1 - s_2$
- Thus expect that: $E[f(I,s_1)/f(I,s_2)] = E[f(I,s_1)]/E[f(I,s_2)] = r(s_1 - s_2)$
- If so, then following relation must hold (derivation in paper):

$$\boxed{f(I,s) \approx f(I,0)e^{-\lambda s}}$$

- Generally applicable (should hold for most feature types)

Fast Accurate Multiscale Detection



(a) dense image pyramid (b) classifier pyramid (c) proposed hybrid approach

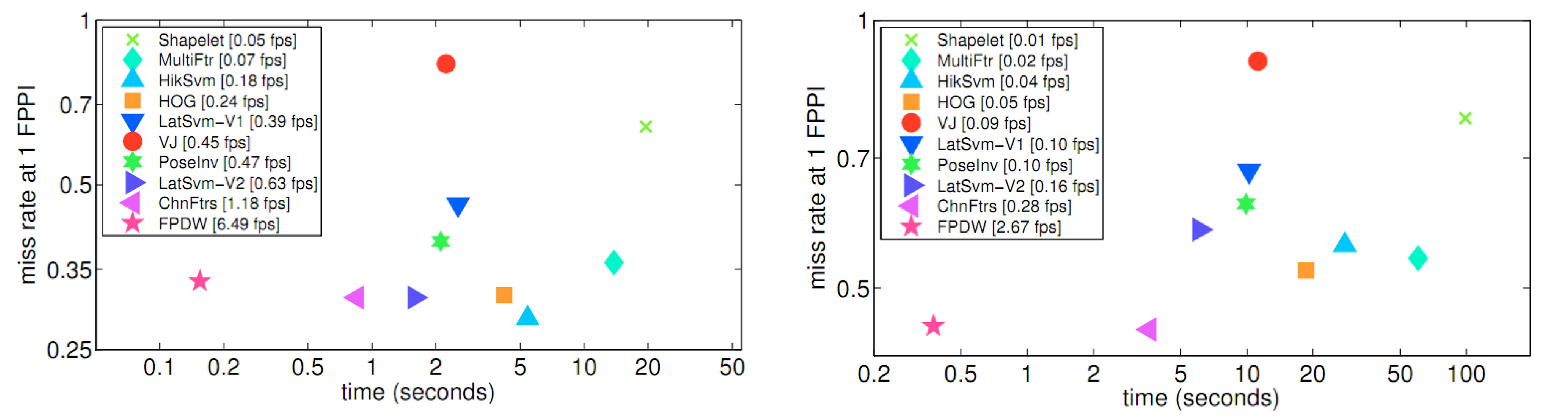
Approach:

- ***Approximate features at nearby scales***
- Coarsely sampled image pyramid (octave step)
- Classifier pyramid within each octave

Benefits:

- Speed of classifier pyramid
- Accuracy of image pyramid
- General applicability

(a) Caltech Pedestrian Data (peds. ≥ 100 pixels) (b) Caltech Pedestrian Data (peds. ≥ 50 pixels))

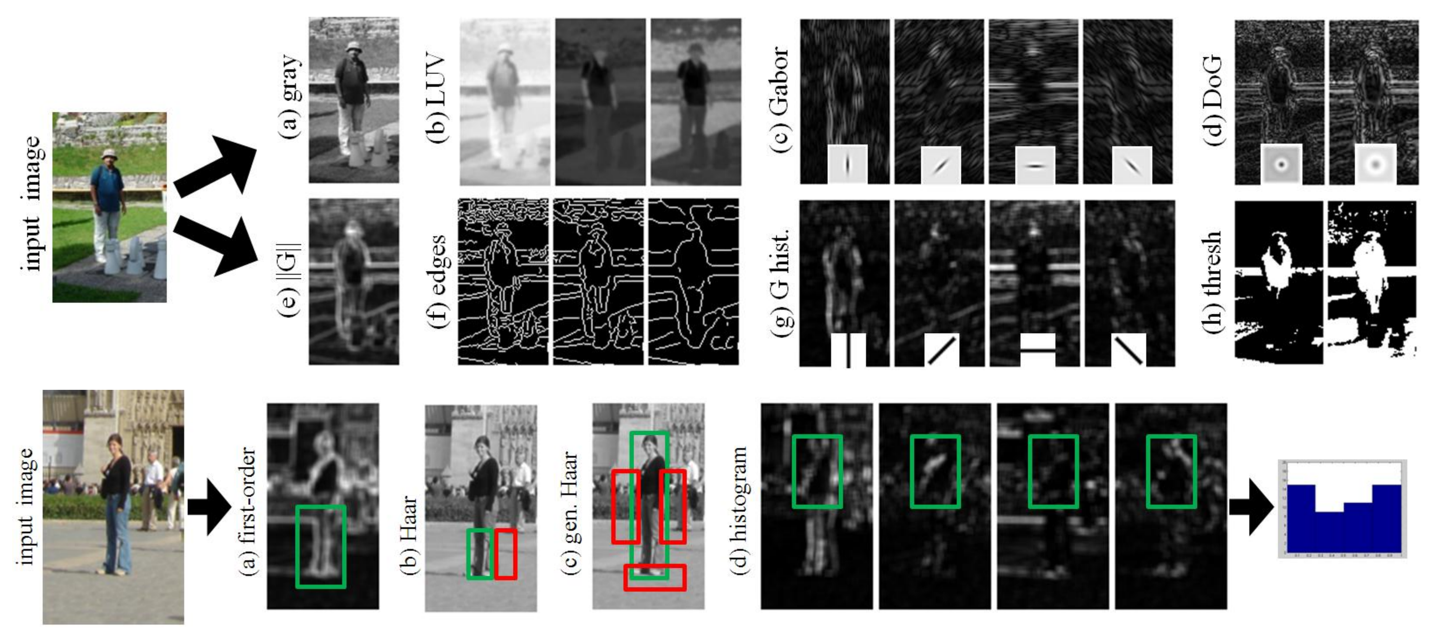


FPDW = Fastest Pedestrian Detector in the West

	Accurate	Fast	Feature Comp.
(a) Dense image pyramid	x		kmn/ln(4)
(b) Classifier pyramid		x	kn
(c) Hybrid Approach	x	x	4kn/3

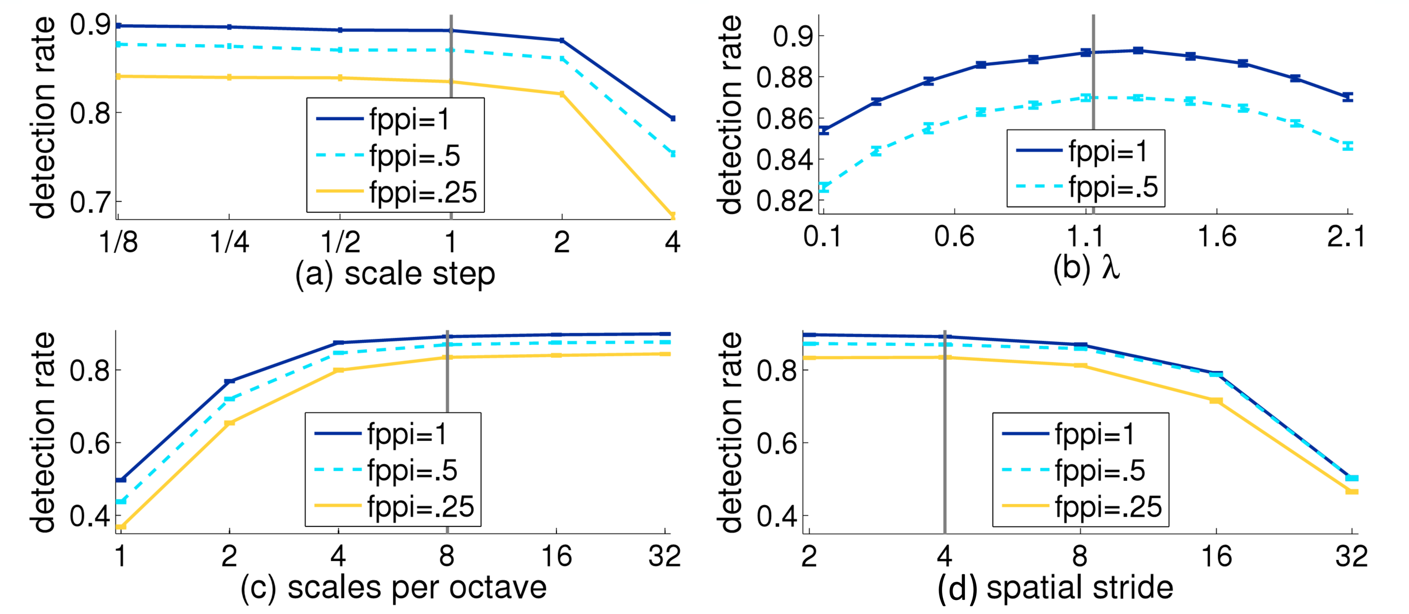
k=cost per pixel, n = # of pixels, m = # of scales / octave

Integral Channel Features [BMVC09]



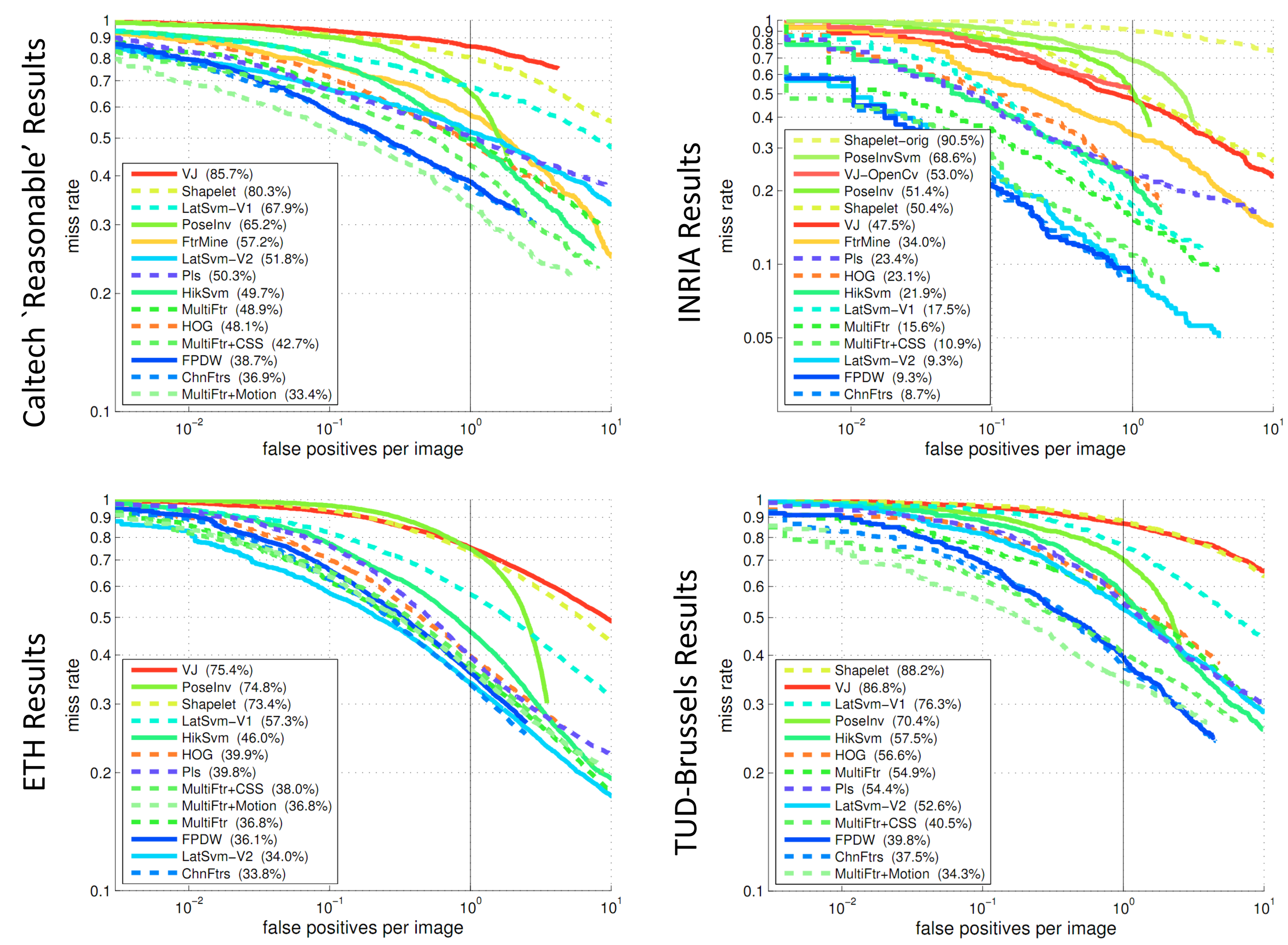
- Generate and compute features efficiently using integral images over multiple registered image channels (followed by feature selection via boosting)
- Integrate heterogeneous information, few parameters, state of the art performance, fast to compute, accurate spatial localization

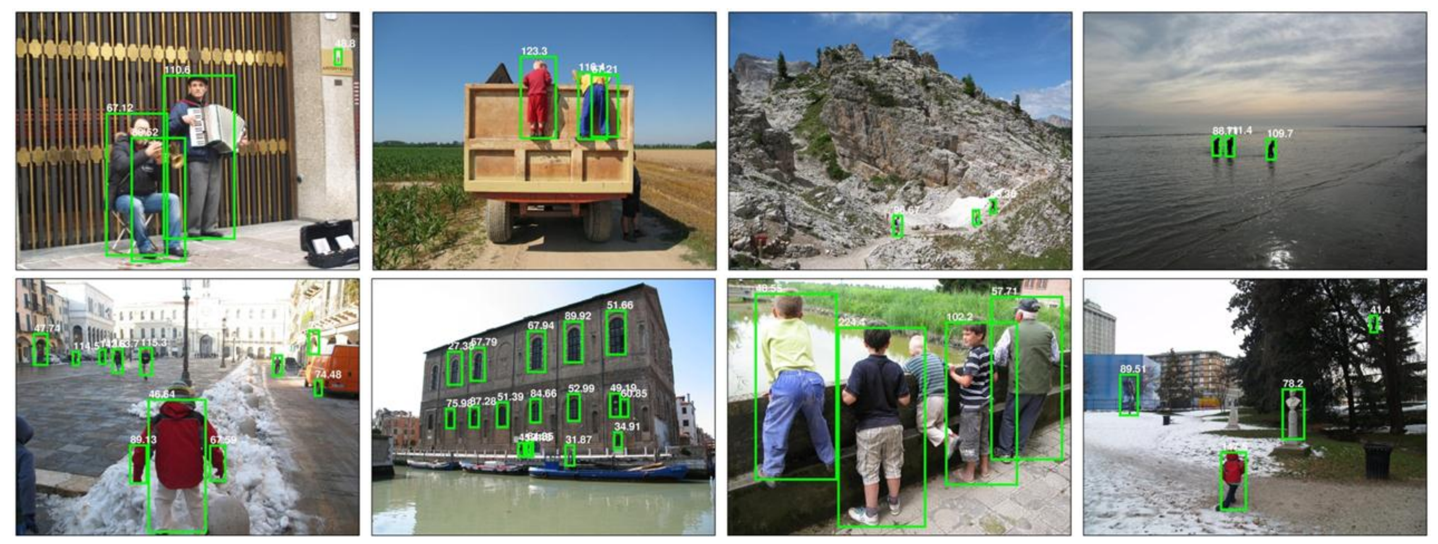
Method Parameters



Detection Performance

www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/





FPDW

- **ACCURATE:** within 1-2% of top performing algorithm
- **ROBUST:** consistent performance across datasets/scenarios
- **FAST:** 10-100x faster than competing methods