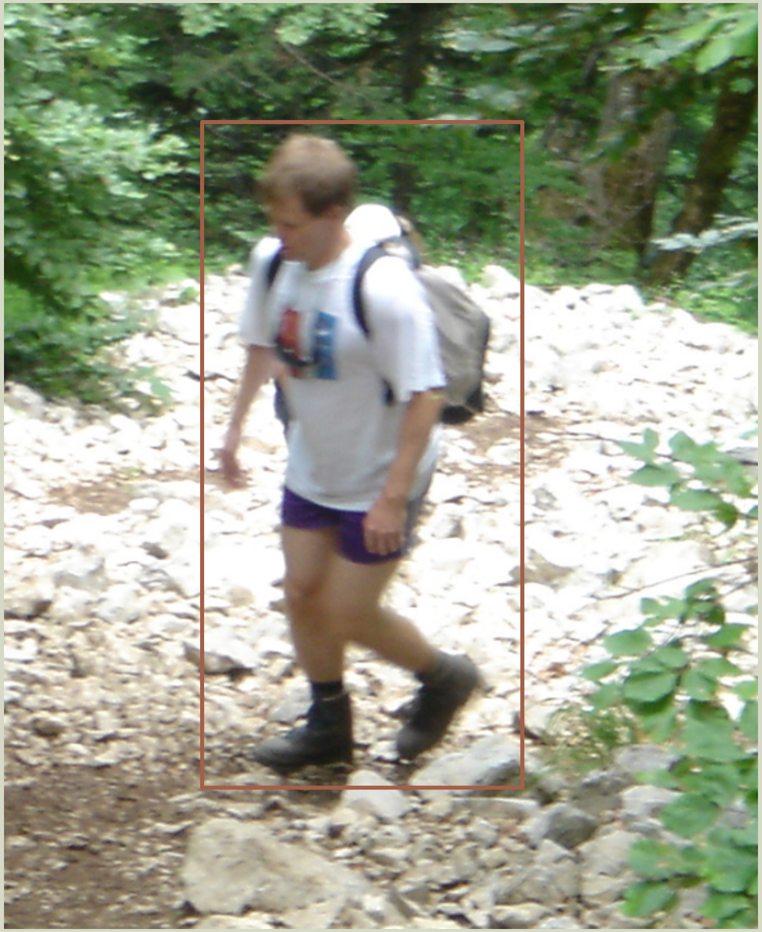
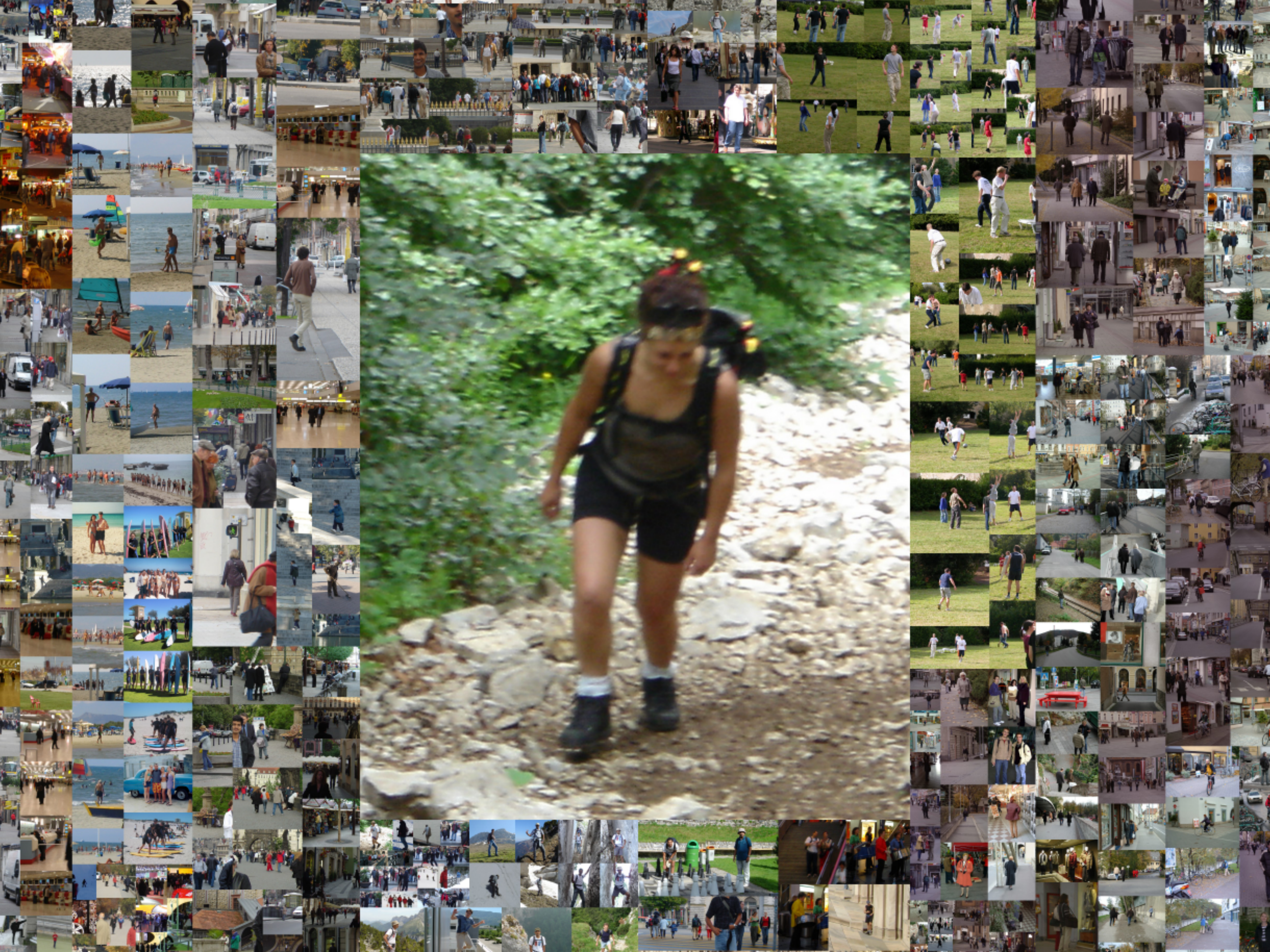
HISTOGRAMS OF ORIENTED GRADIENTS FOR HUMAN DETECTION (NAVNEET DALAL AND BILL TRIGGS)

Rafael Cosman

Tao Wang

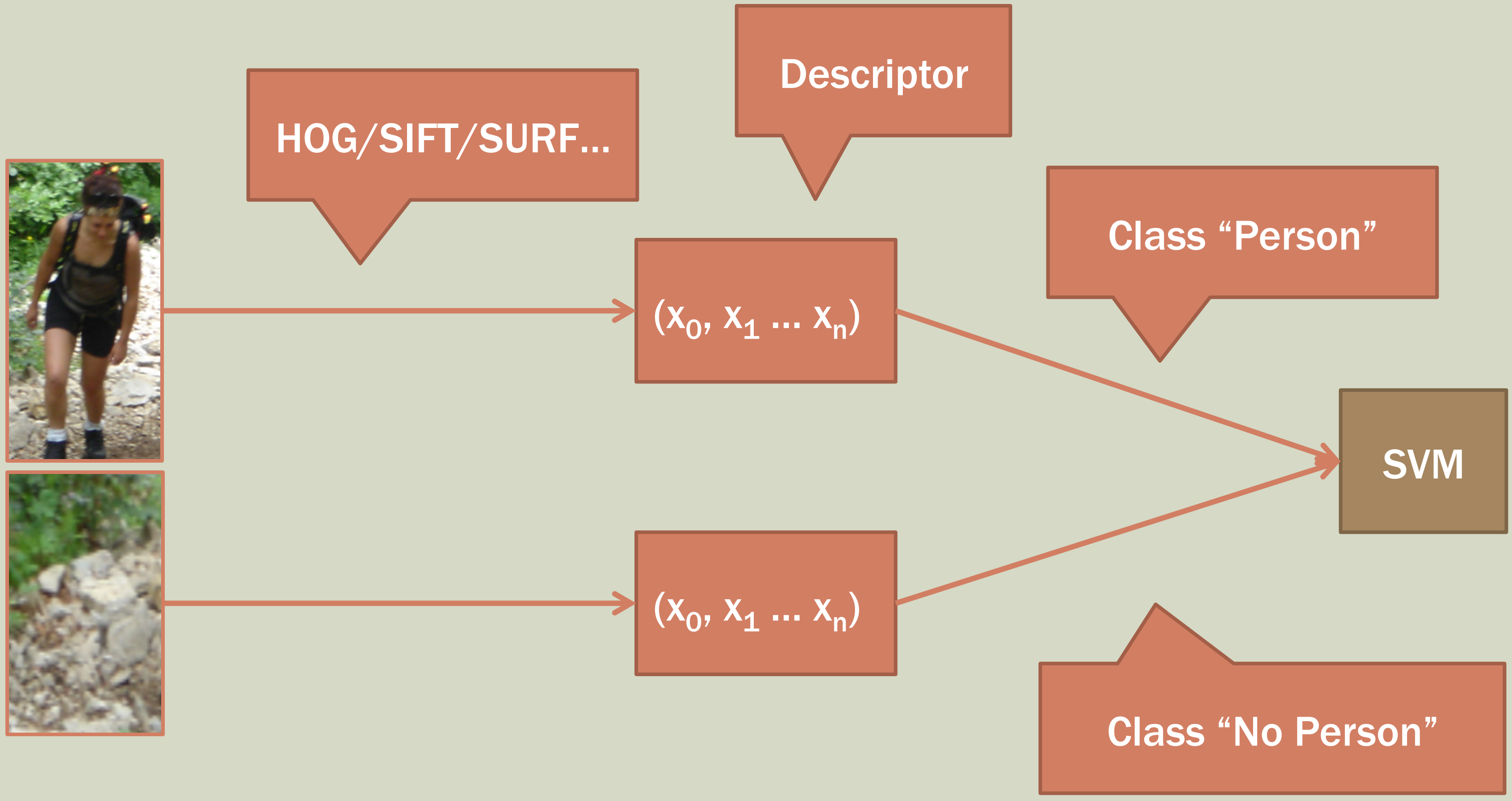
HUMAN DETECTION



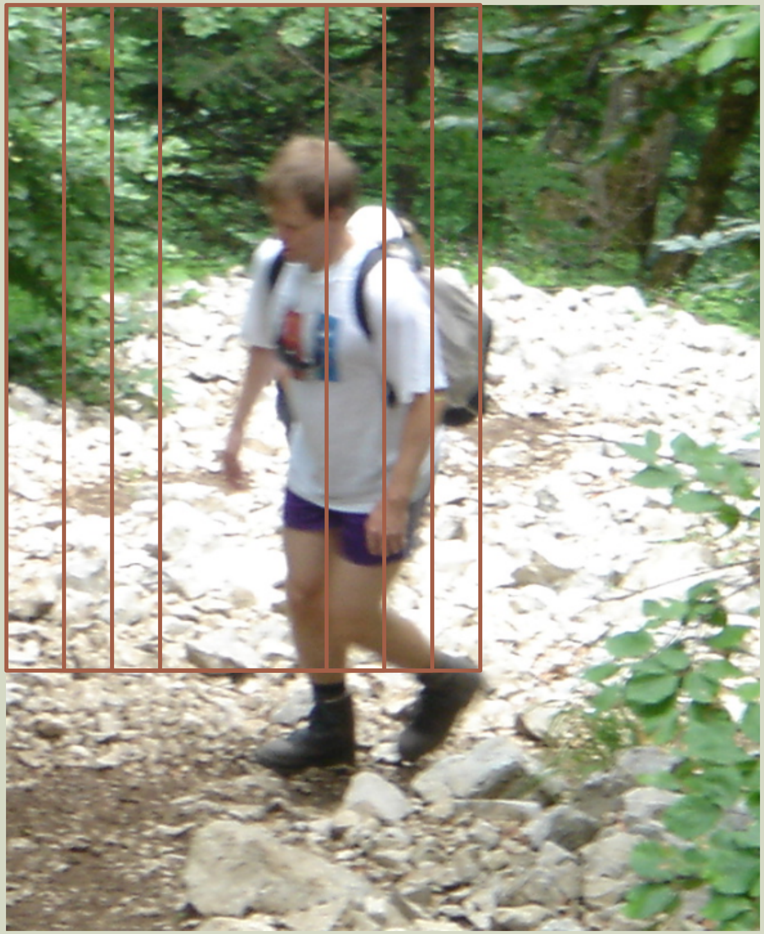



Gösser

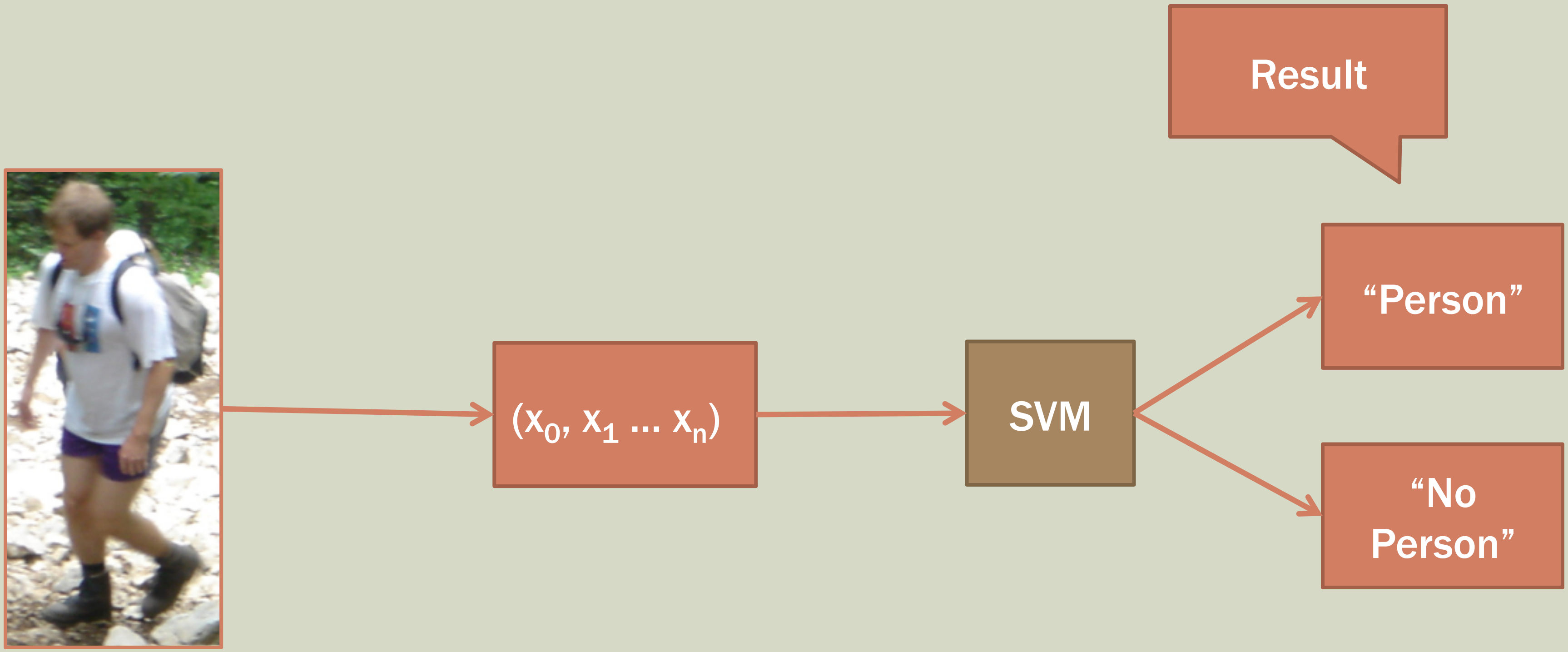
TRAINING SVM



TESTING: SCANNING WINDOW

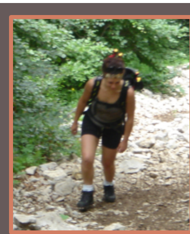


TESTING SVM



MOTIVATION

- Very simple to implement
- Performs as well or better than many descriptors
- Cited over 5000 times
  - Basis for Deformable Parts Model



HOG

$(x_0, x_1 \ldots x_n)$

Image → Gradients → Cells → Blocks → Descriptor

Image → **Gradients** → Cells → Blocks → Descriptor

- Convolve the image with discrete derivative mask
  - [-1, 0, 1]
  - $[-1, 0, 1]^T$

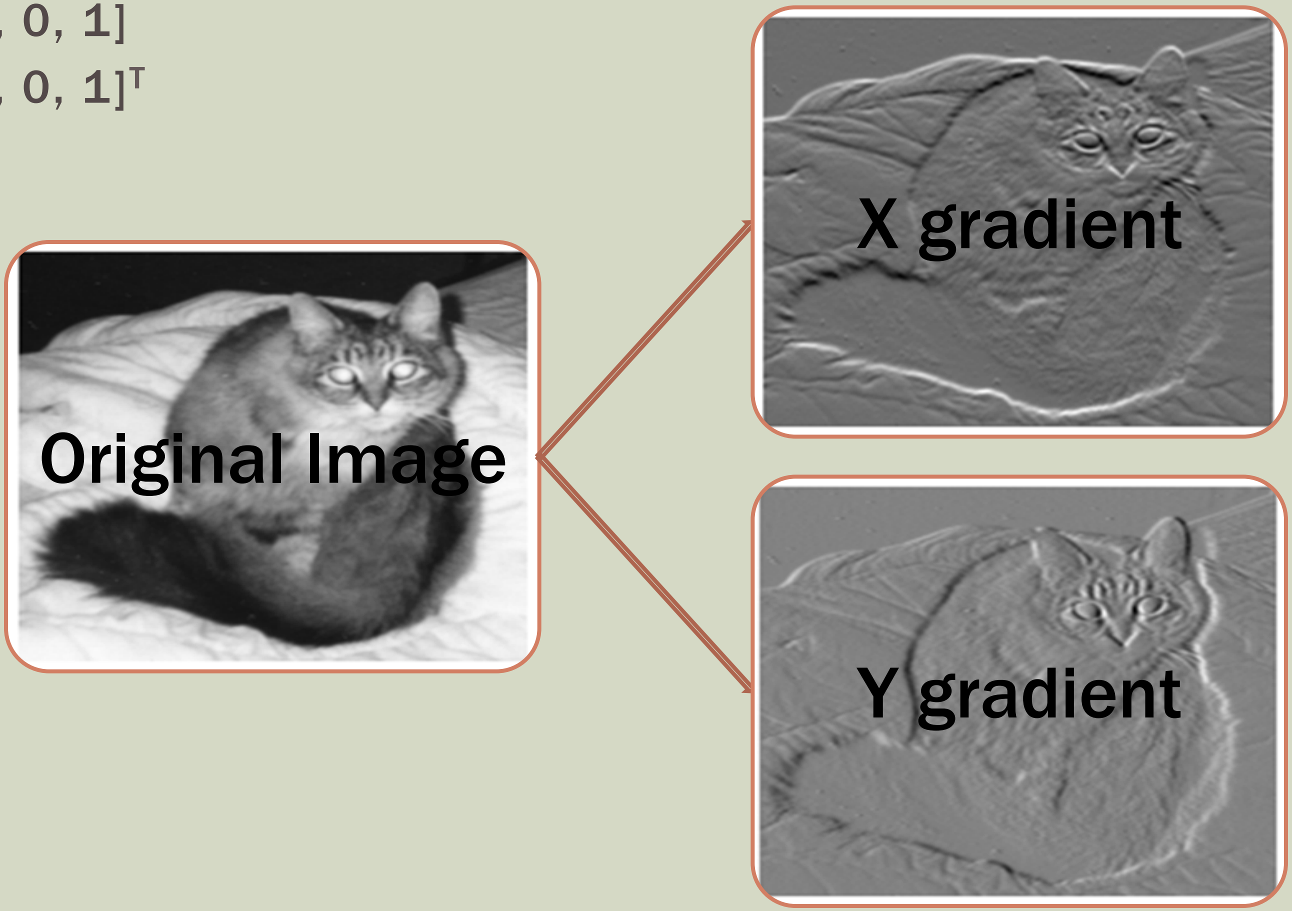
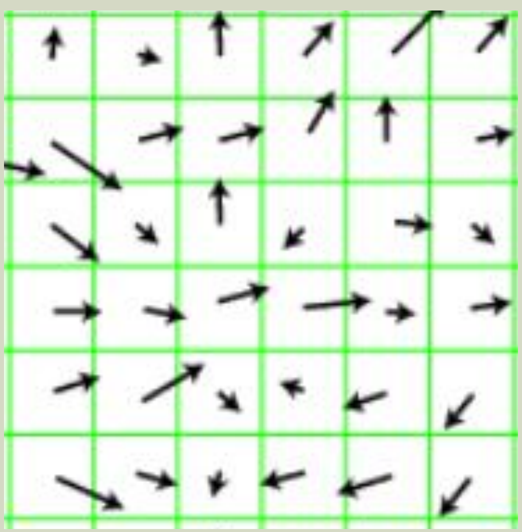
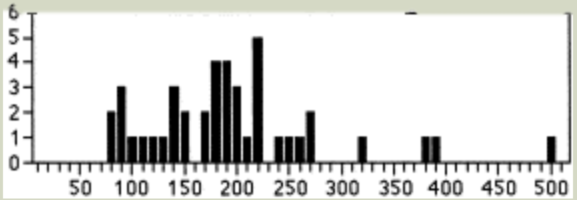
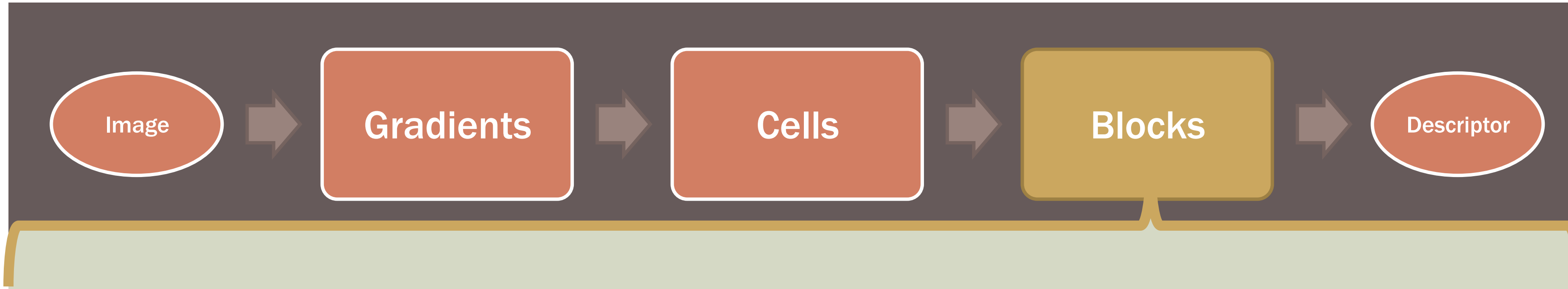


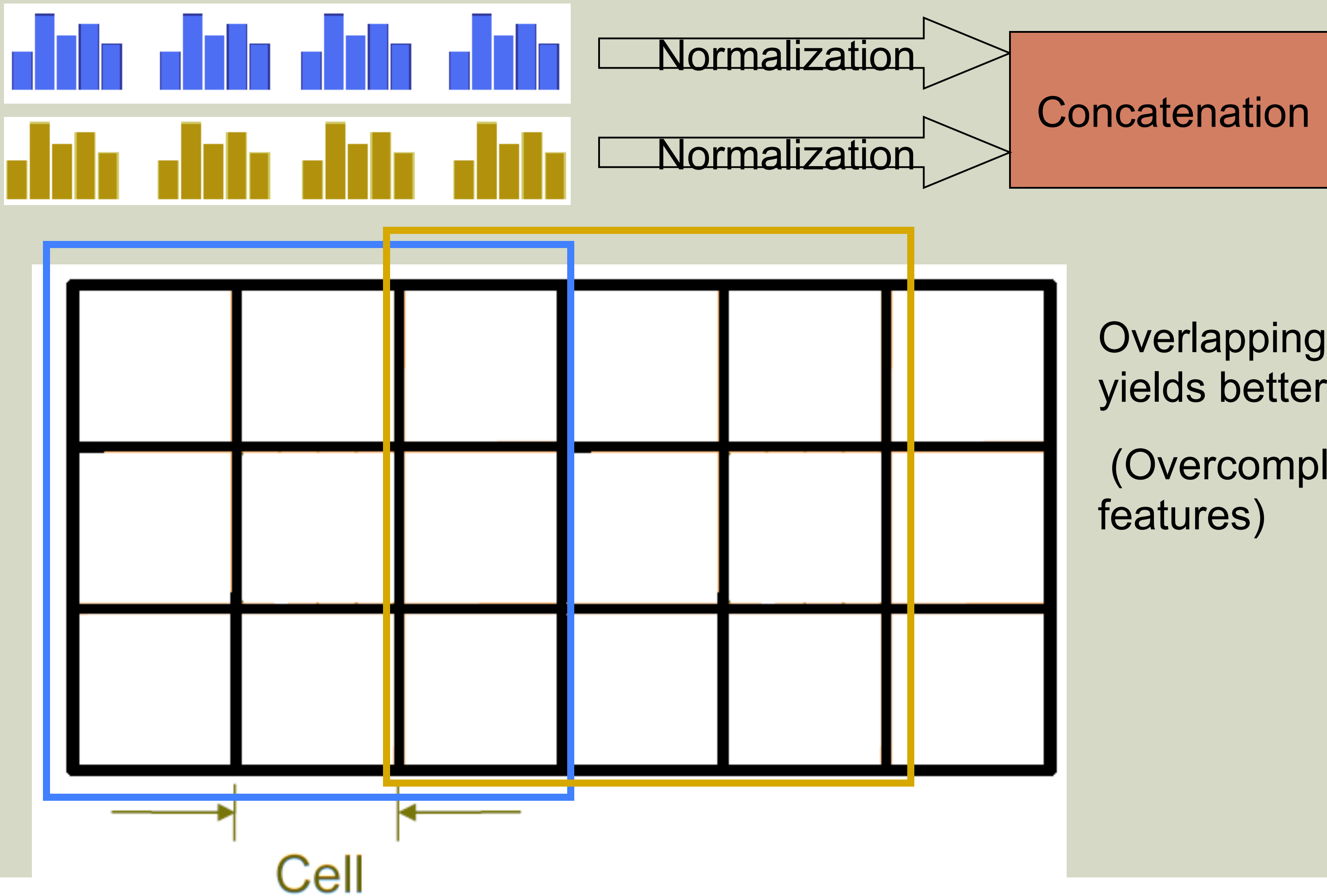
Image → Gradients → **Cells** → Blocks → Descriptor

- Now we count up the gradient angles in 8x8 cells
  - Vote weight = magnitude = $\sqrt{dx^2 + dy^2}$
  - Who you vote for ~ angle = $\arctan(dy/dx)$





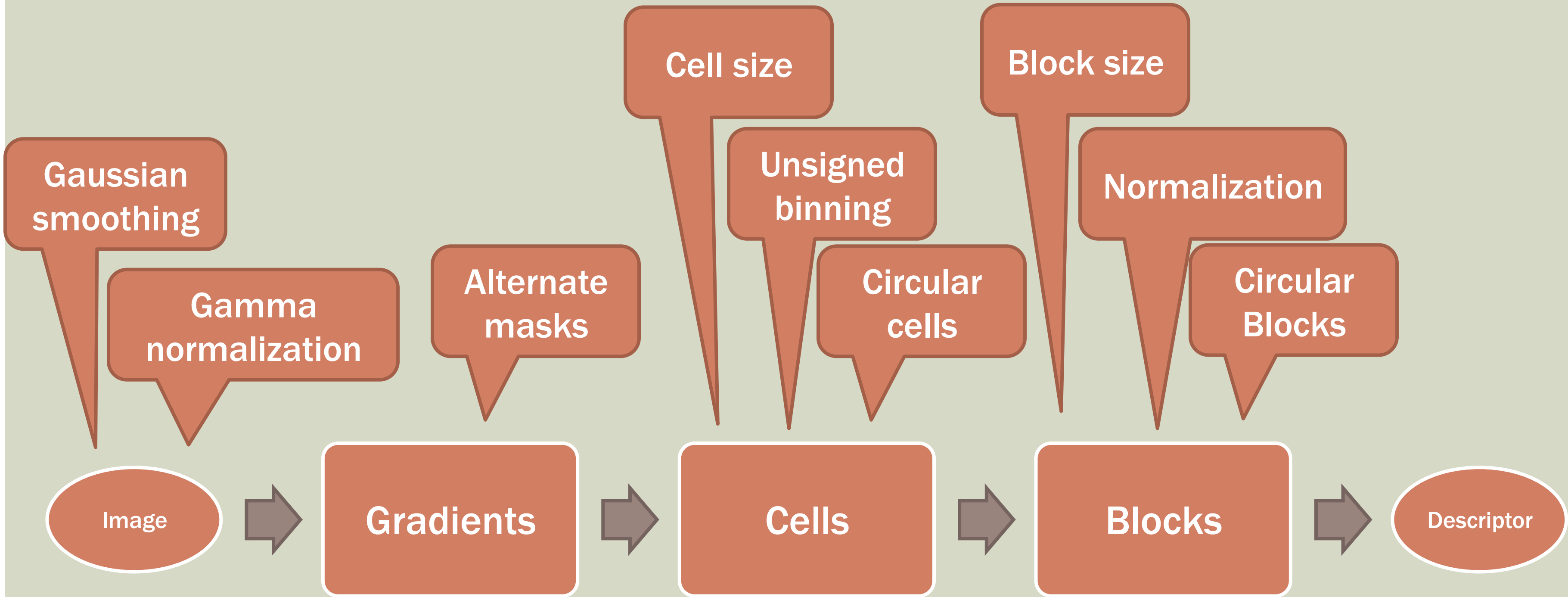




Overlapping blocks yields better results!

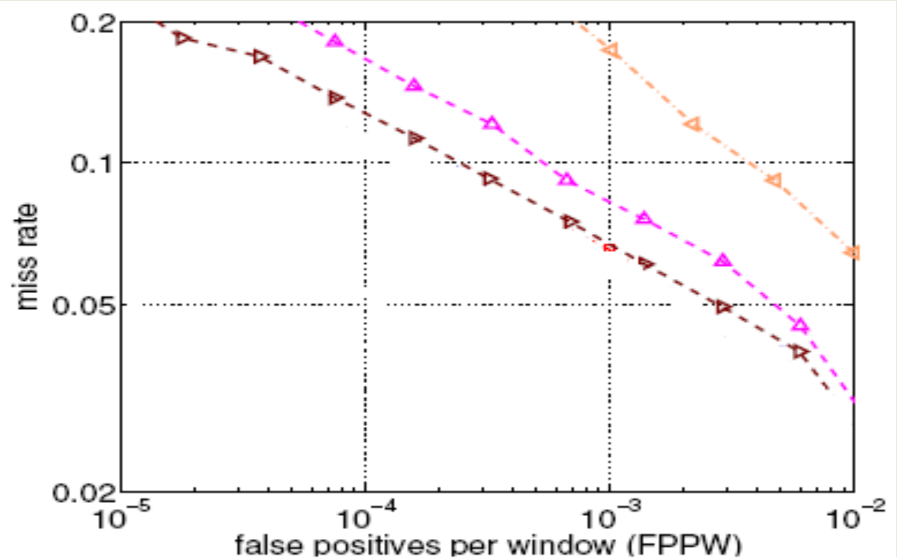
(Overcomplete features)

VARIATIONS


Gaussian smoothing
Gamma normalization
Alternate masks
Cell size
Unsigned binning
Circular cells
Block size
Normalization
Circular Blocks
Image
Gradients
Cells
Blocks
Descriptor

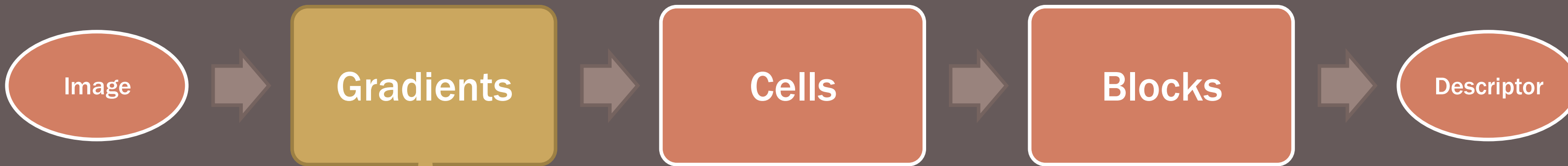
EVALUATION METRIC

- Detection task, so a single accuracy value does not make sense
  - Low threshold – Low miss rate, but many false positives
  - High threshold – Few false positives, but misses a lot
- **Detection Error Tradeoff (DET)** Curves
  - Vertical Axis: Miss Rate = 1-Recall = (1 – True Pos / Total Pos in GT)
  - Horizontal Axis: False Positive Per Window (FPPW)

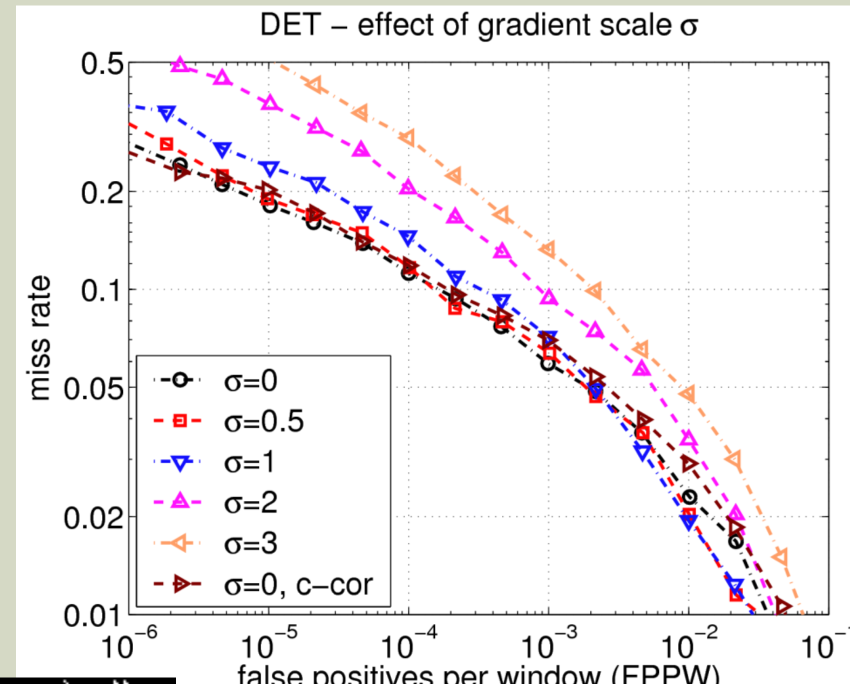


- Using Grayscale/Color
  - Compute gradients in all color channels, pick the highest one
  - Using color gives slightly better results

- Gamma/Color Normalization
  - As a preprocessing step
  - Has little effect on results



- Gaussian smoothing
  - Reduces performance
  - Fine-grained edge detection is crucial




raw image


σ>0

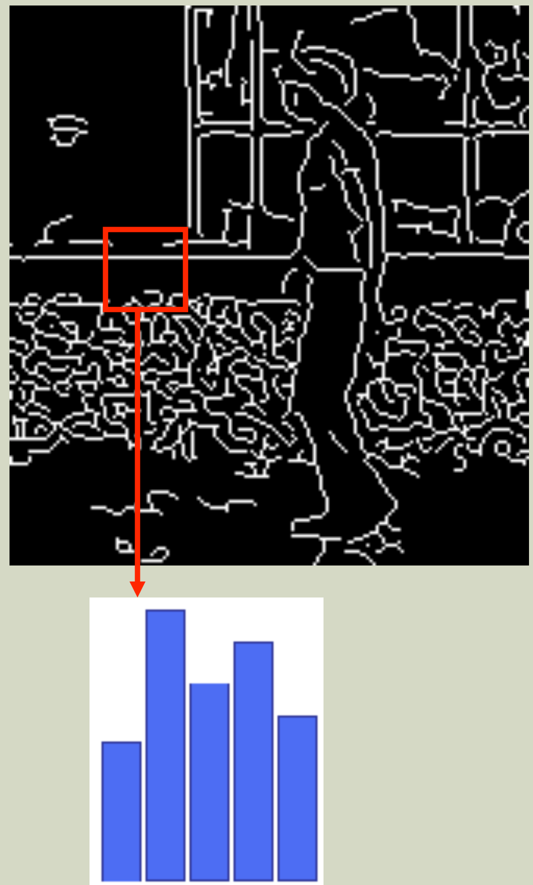

σ=0

Image → **Gradients** → Cells → Blocks → Descriptor

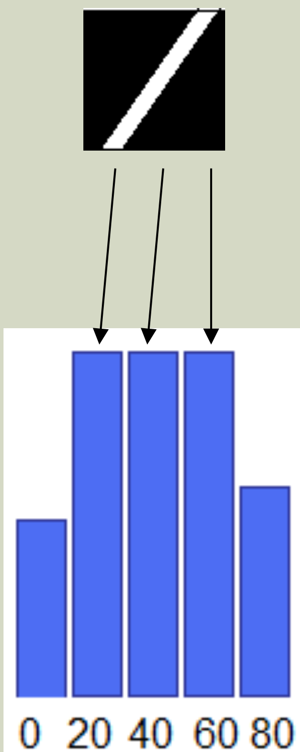
- Choice of masks
  - 1-D centered: [-1, 0, 1], $[-1, 0, 1]^T$
  - 1-D uncentered: [-1, 1], $[-1, 1]^T$
  - 1-D Cubic corrected: [1, -8, 0, 8, -1], $[1, -8, 0, 8, -1]^T$
  - 2-D Sobel mask: $\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}^T$
- 1-D centered [-1, 0, 1], $[-1, 0, 1]^T$ with no Gaussian smoothing works best

Image → Gradients → **Cells** → Blocks → Descriptor

- Rectangular vs Circular cells



- Bilinear interpolation to reduce aliasing
- Across both orientations and locations (weighted by distance in pixels)



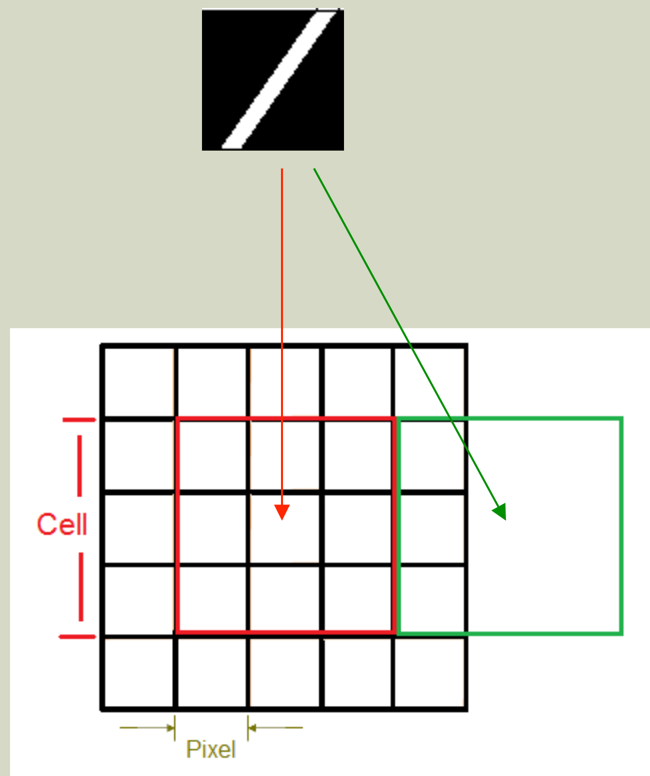
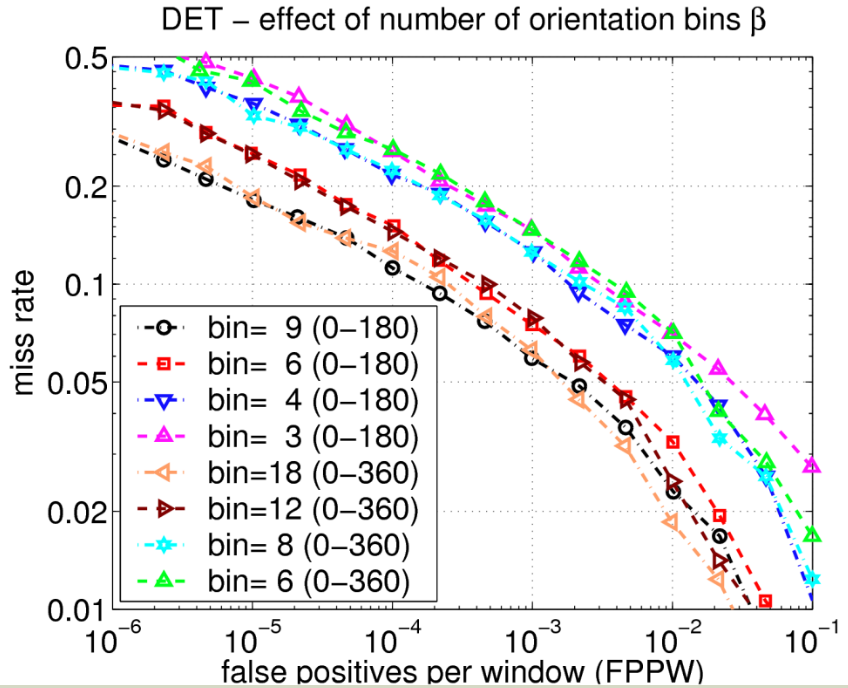


Image → Gradients → **Cells** → Blocks → Descriptor

- # of orientation bins
  - Increasing orientation bins from 4 to 9 decreases false positives by 10 times
- Unsigned cells
  - 0-180 degrees instead of 0-360 degrees
  - Actually improves performance slightly!
    - Why?



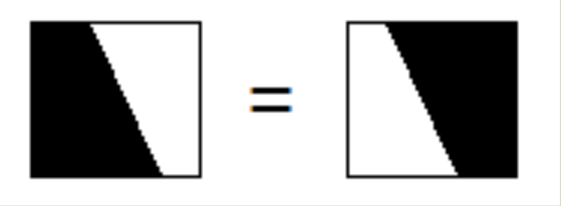
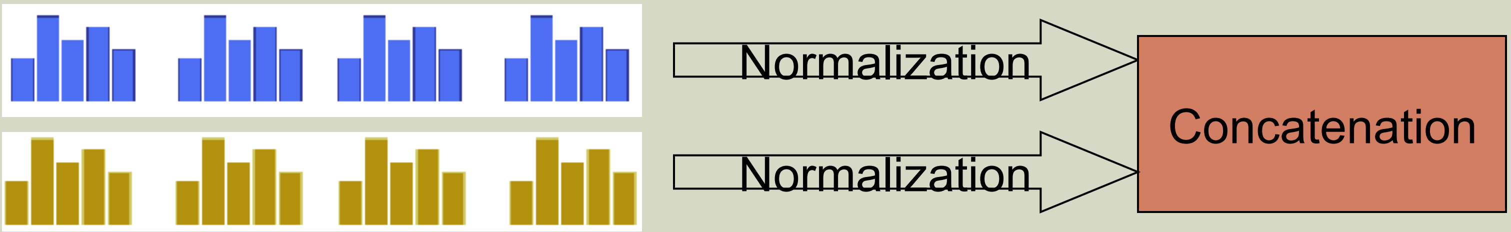


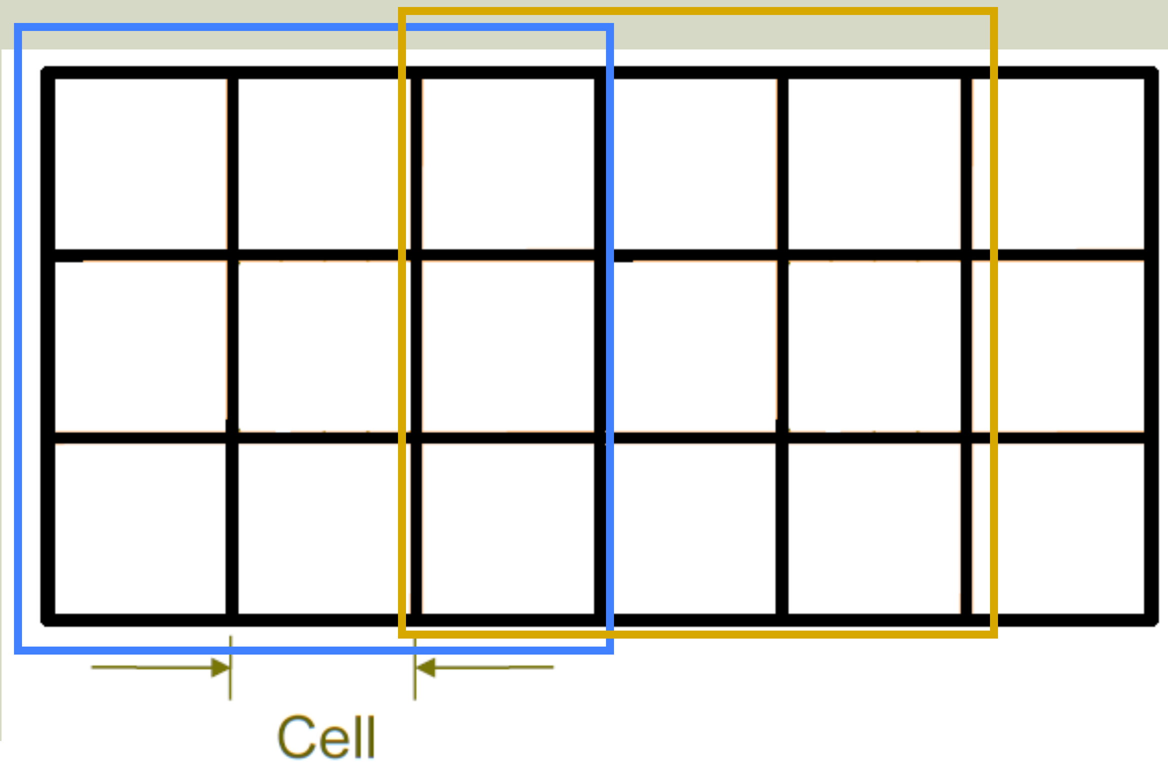
Image → Gradients → Cells → Blocks → Descriptor

- Motivation
  - Gradient magnitude lacks invariance to changes in illumination and foreground background contrast.

- Need local contrast normalization to find the “true” weight of an edge

R-HOG



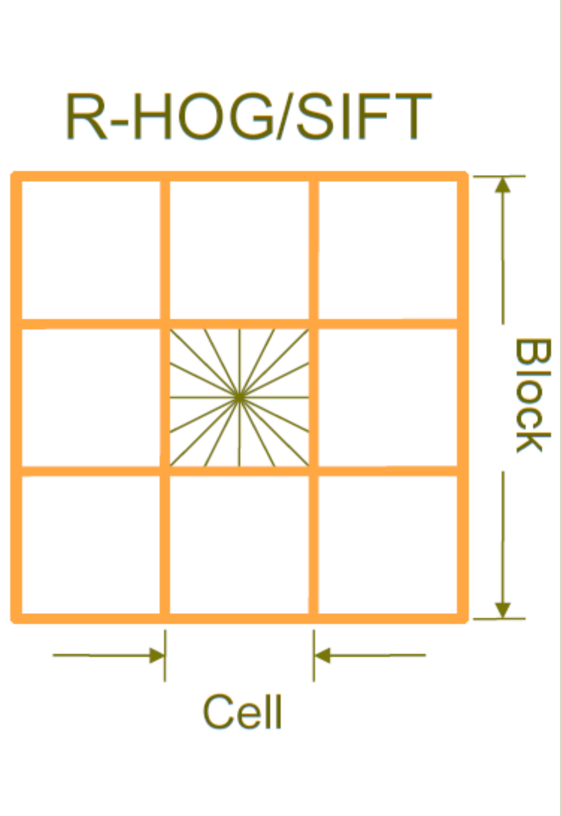


Overlapping blocks yields better results!

(Overcomplete features)

Image → Gradients → Cells → Blocks → Descriptor

R-HOG versus C-HOG



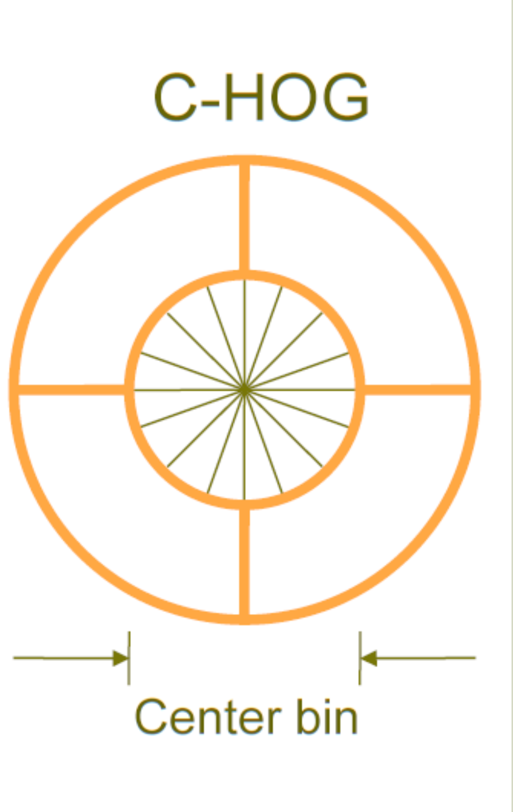
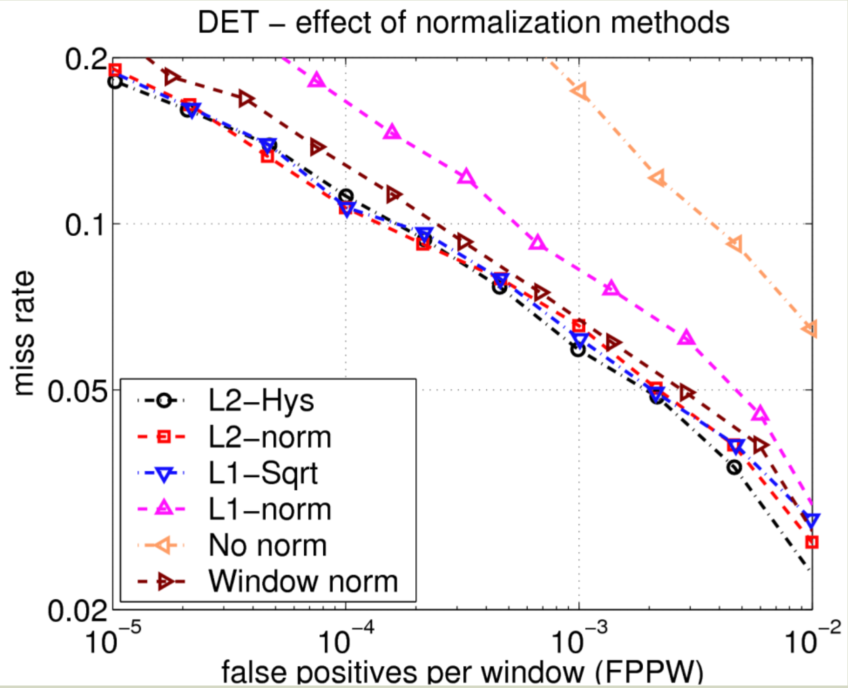
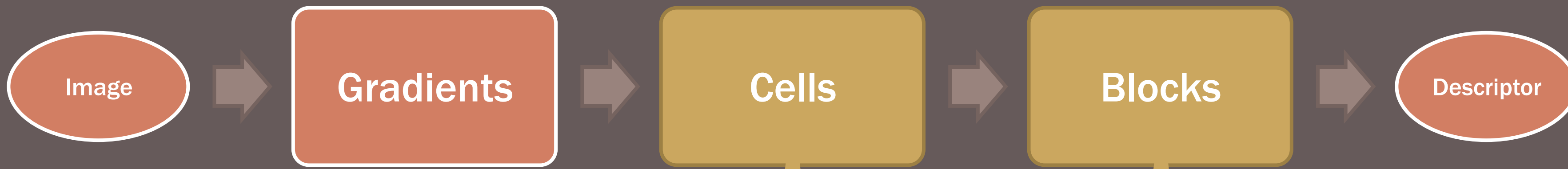


Image → Gradients → Cells → Blocks → Descriptor

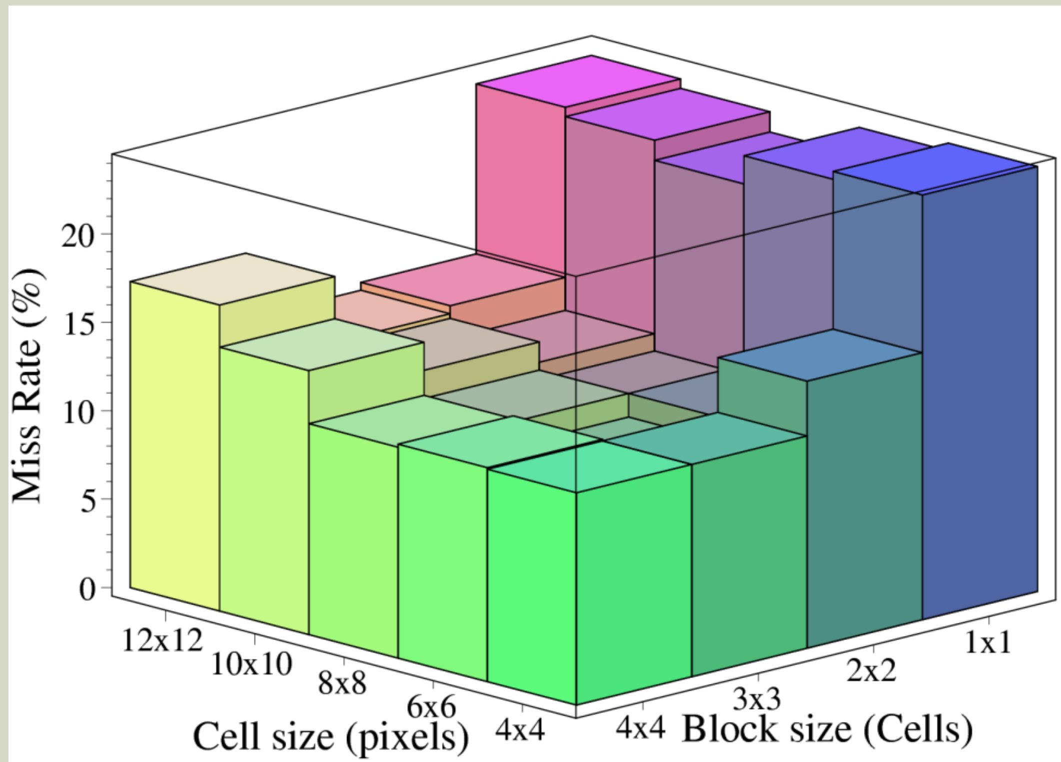
Normalizations

- L2-norm
- L2-hys: L2-norm followed by clipping (limiting the maximum values of v to 0.2) and renormalizing
- L1-norm:
- L1-sqrt:

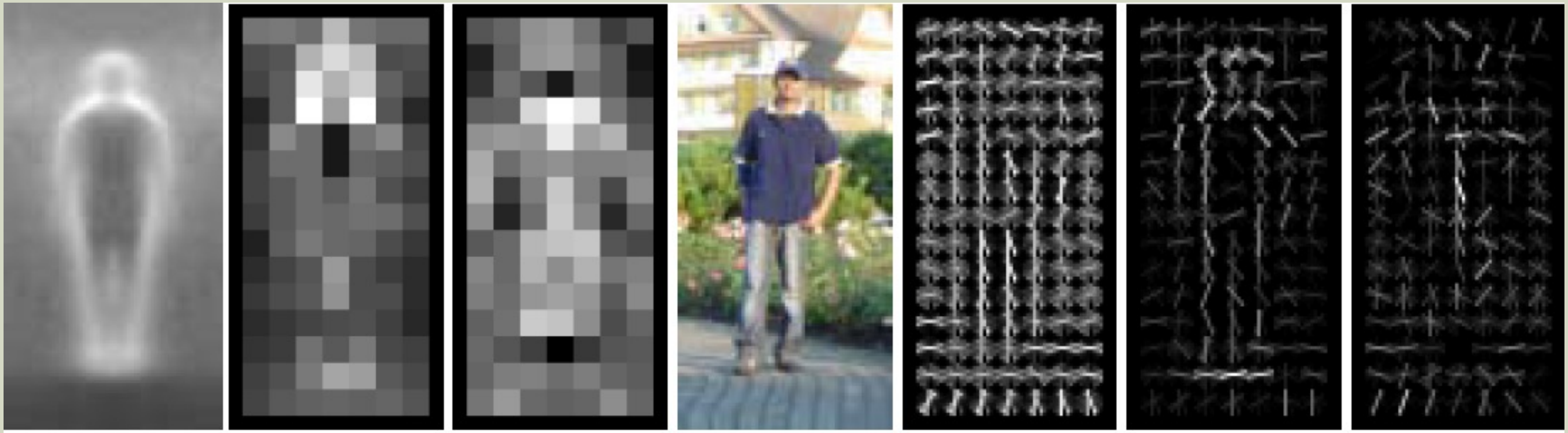




- Best cell & block sizes:
  - Cell size of 6x6
  - Block size of 3x3

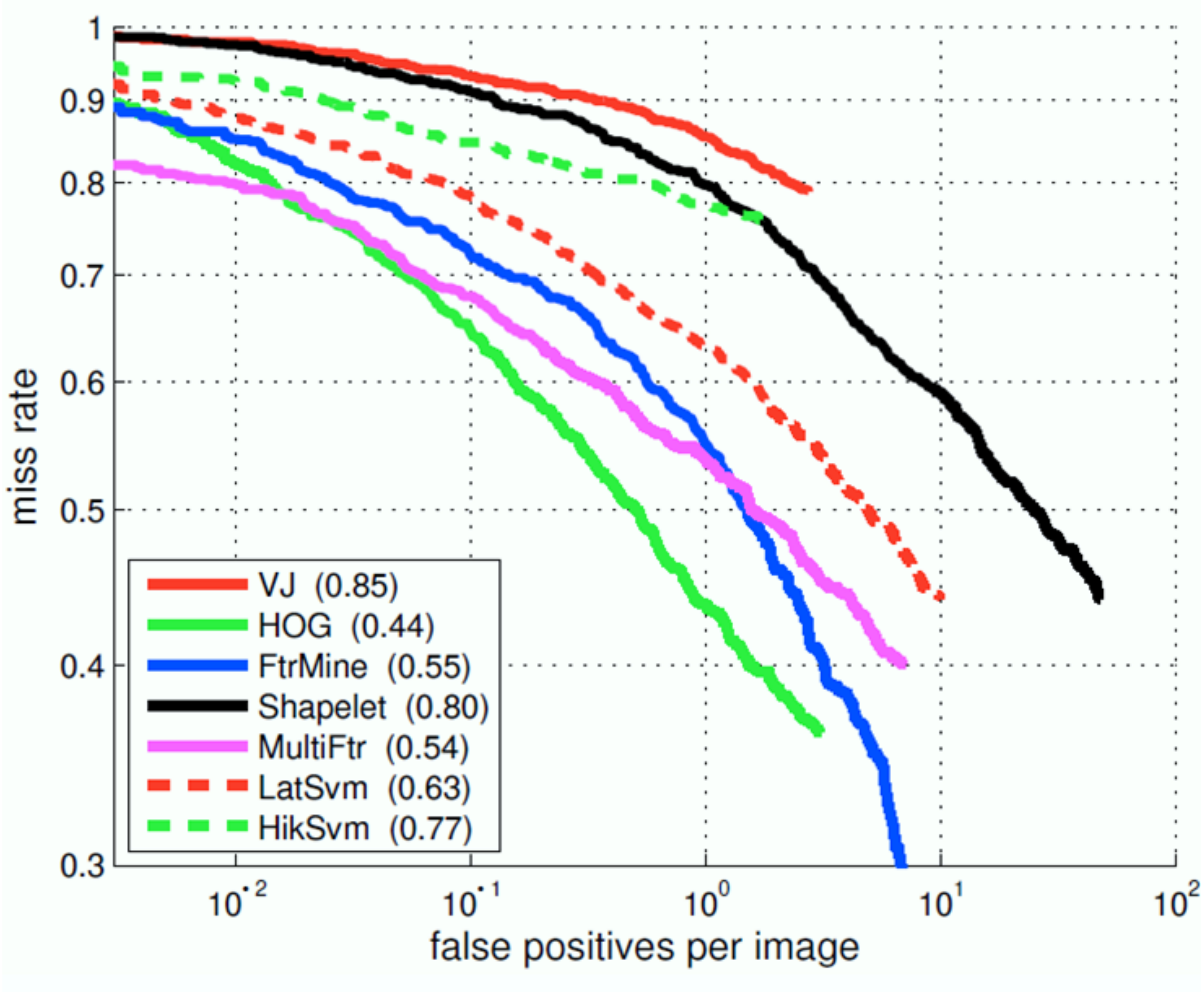


VISUALIZATION AND INSIGHTS



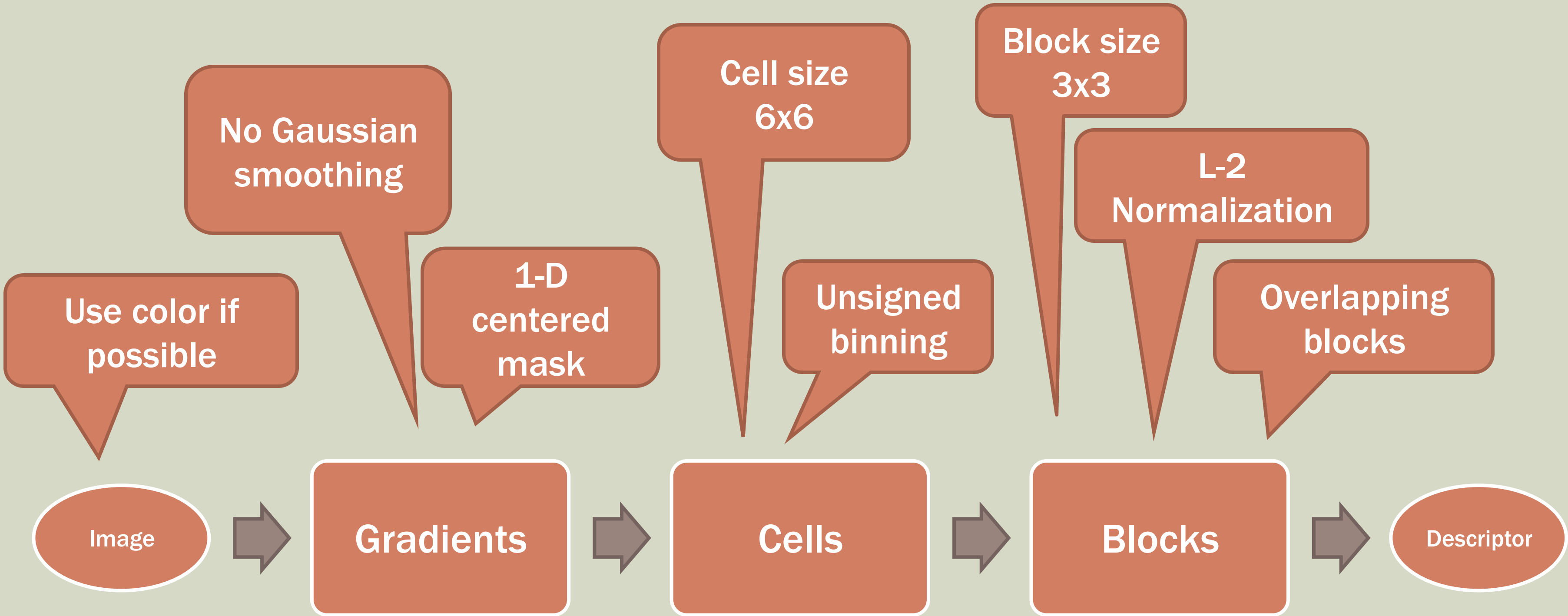
(a) (b) (c) (d) (e) (f) (g)

a. Average gradient over positive examples
b. Maximum positive SMV weight in each block
c. Maximum negative SMV weight in each block
d. A test image
e. It's R-HOG descriptor
f. R-HOG descriptor weighted by positive SVM weights
g. R-HOG descriptor weighted by negative SVM weights



HOG COMPARED TO OTHERS

BEST SETUP



Use color if possible
No Gaussian smoothing
1-D centered mask
Cell size 6x6
Unsigned binning
Block size 3x3
L-2 Normalization
Overlapping blocks
Image
Gradients
Cells
Blocks
Descriptor

RULES OF THUMB

- Abrupt edges at fine scales are essential
  - No blurring
- Local contrast normalization is essential
- Overlapping blocks w/ “redundant” information improves results significantly.
- Fine orientation quantization is more important than fine spatial orientation