

Machine Learning for Computer Vision

MVA – ENS Cachan



Search and Vision

Iasonas Kokkinos

iasonas.kokkinos@ecp.fr

Center for Visual Computing
Ecole Centrale Paris

Galen Group
INRIA-Saclay

Perceptual Organization



Perceptual Organization



Sliding window classifiers



Sliding window classifiers



-0.2

Sliding window classifiers



...
1.5
...

Sliding window classifiers



Sliding window classifiers



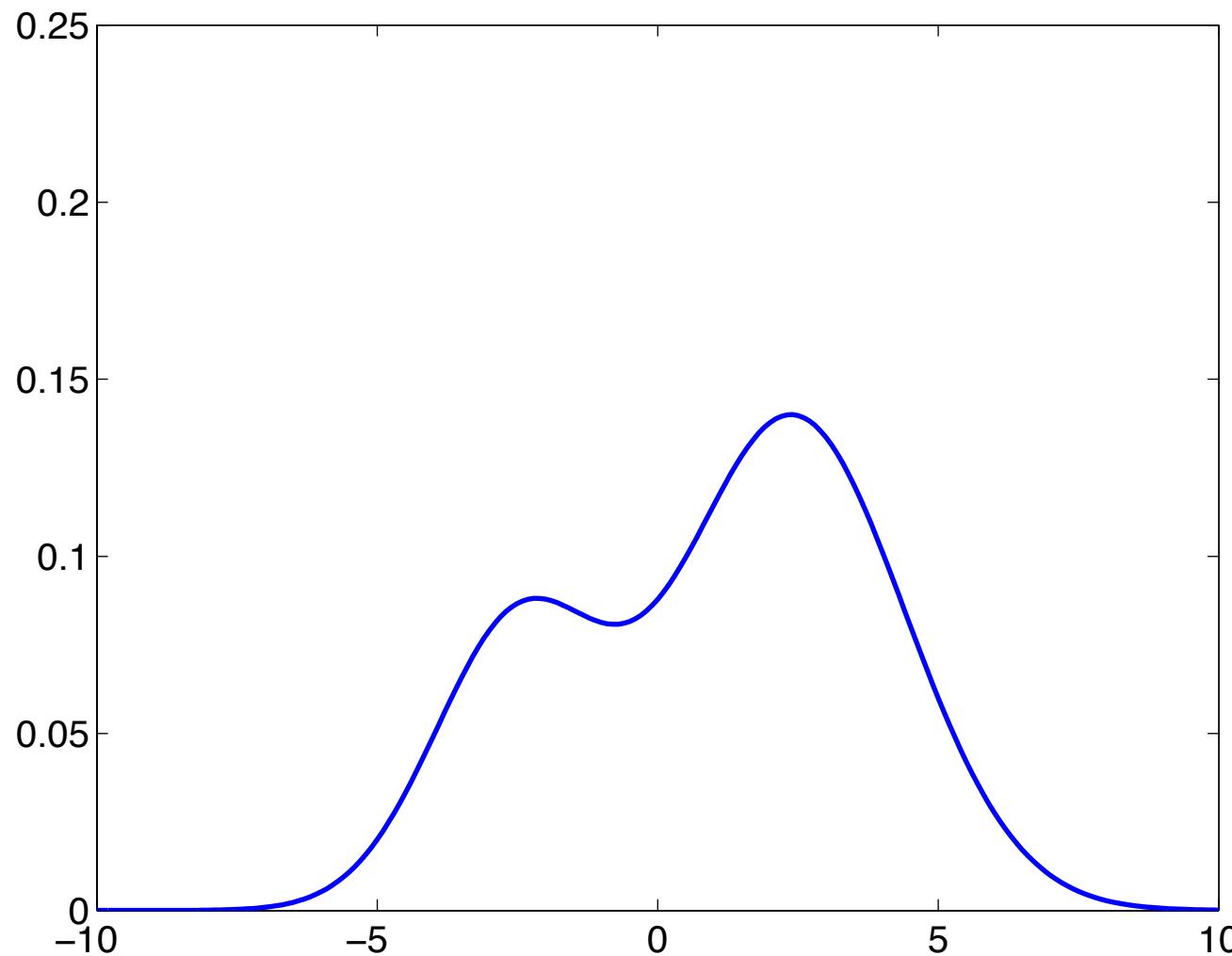
Sliding window classifiers



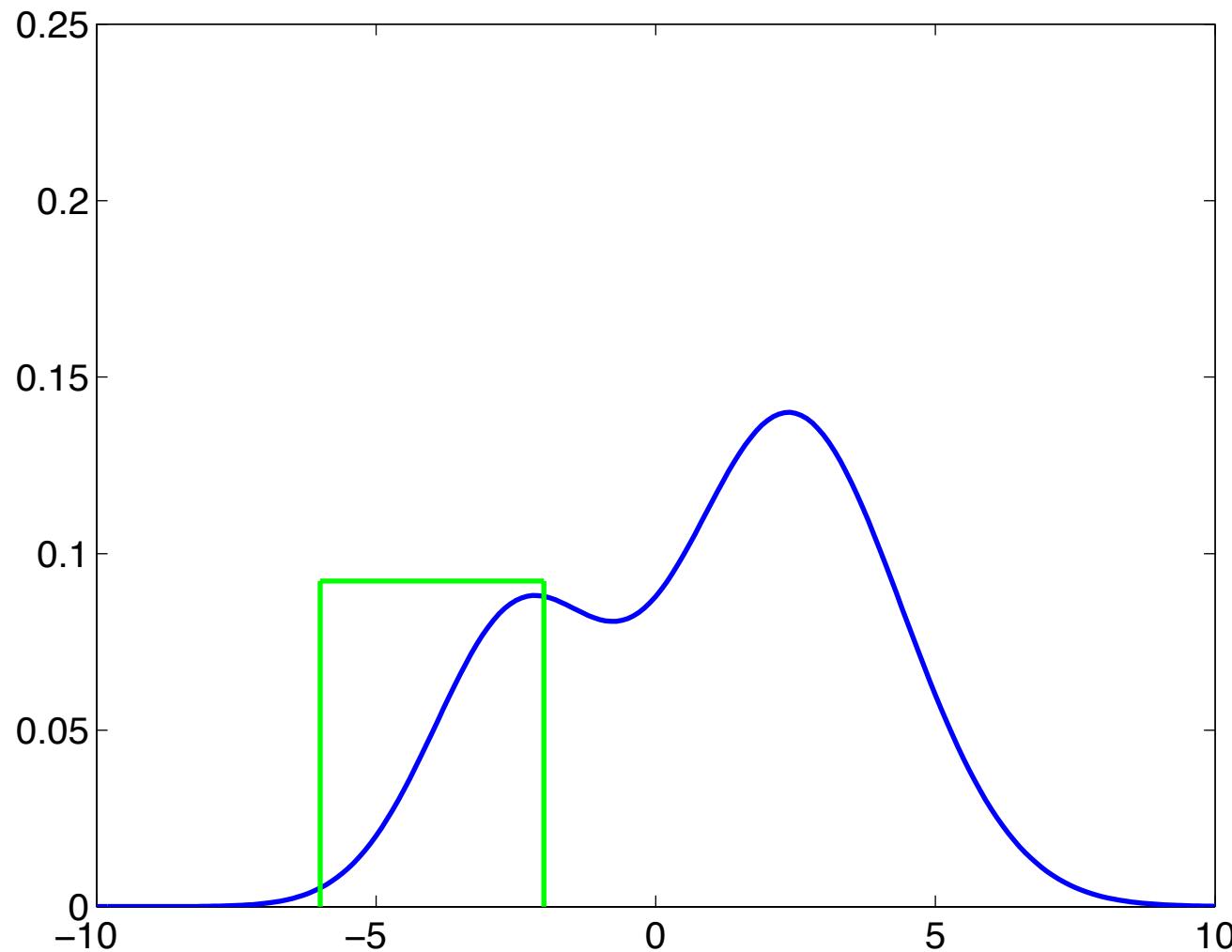
0.1
-0.2
-0.1
0.1
...
1.5
...
0.5
0.4
0.3

Branch and bound

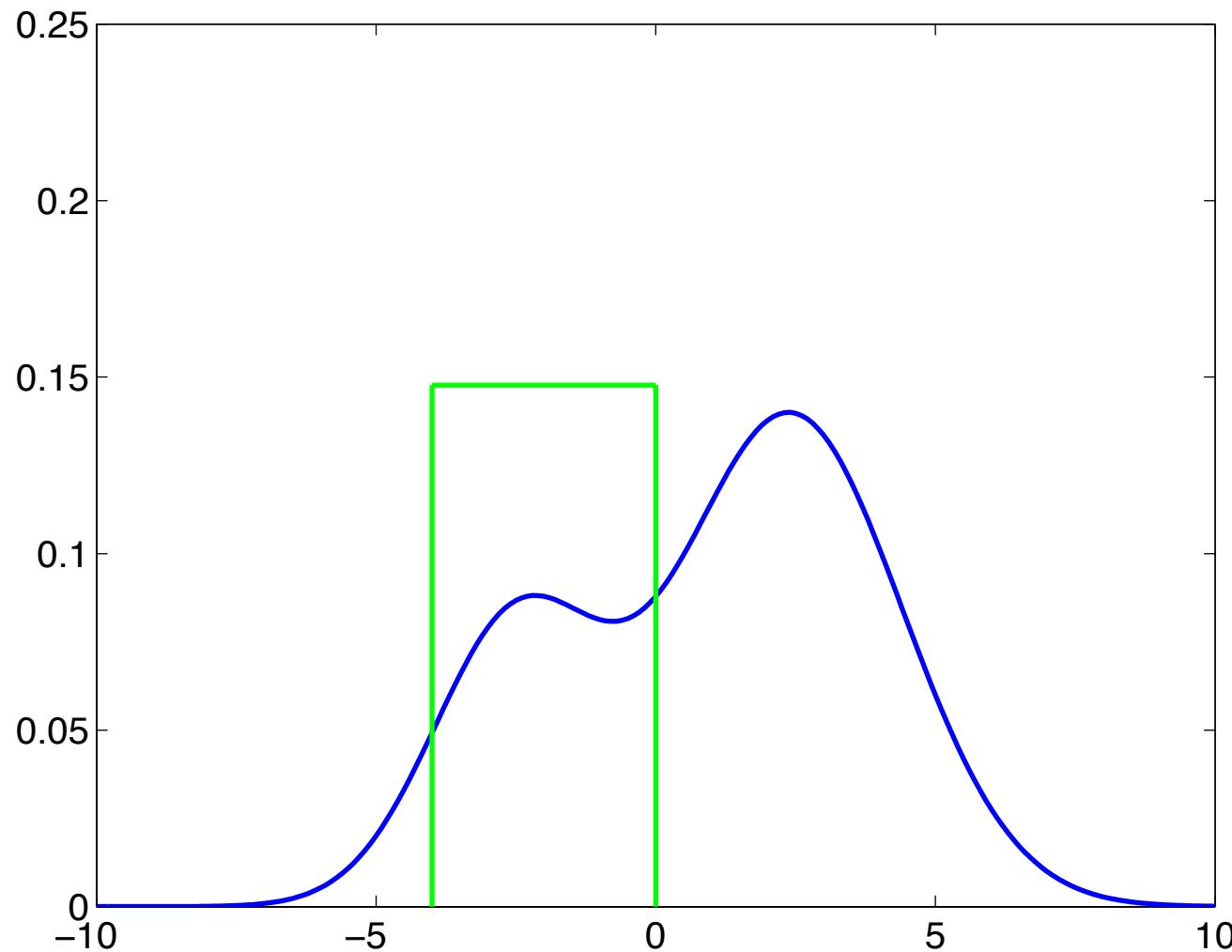
$$\max_{x \in X} f(x)$$



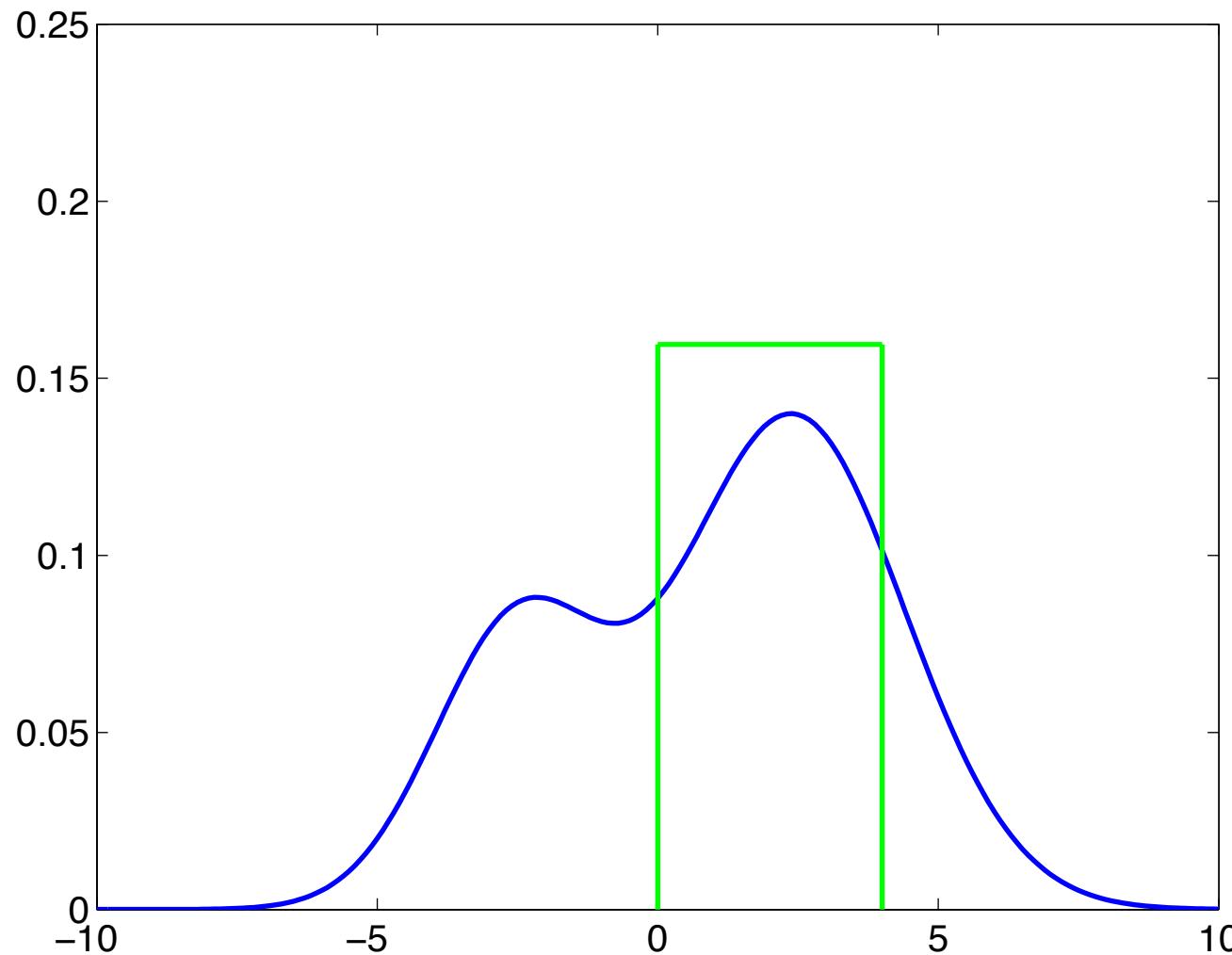
Bounding function $\bar{f}(X) \geq \max_{x \in X} f(x), \quad \bar{f}(\{x\}) = f(x)$



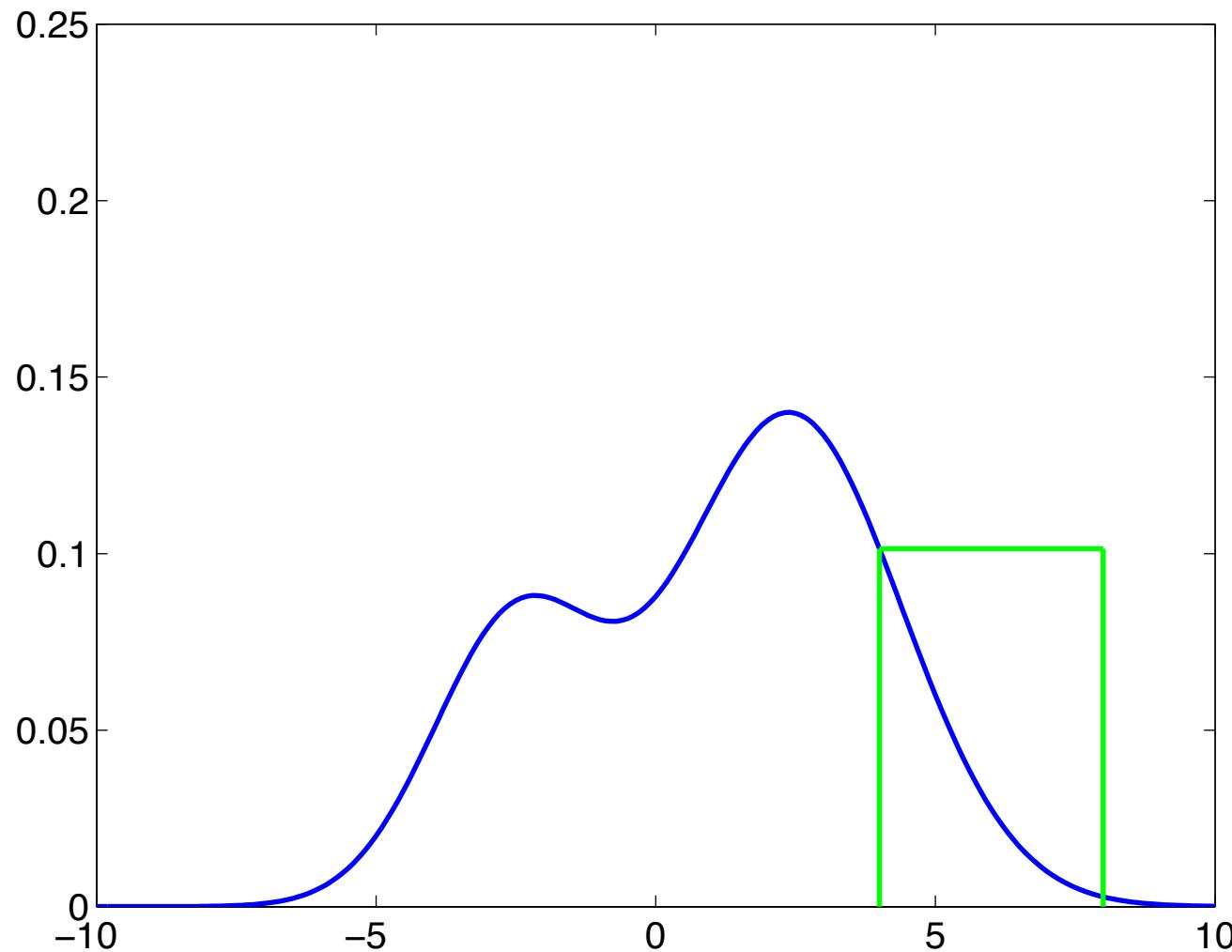
Bounding function $\bar{f}(X) \geq \max_{x \in X} f(x), \quad \bar{f}(\{x\}) = f(x)$



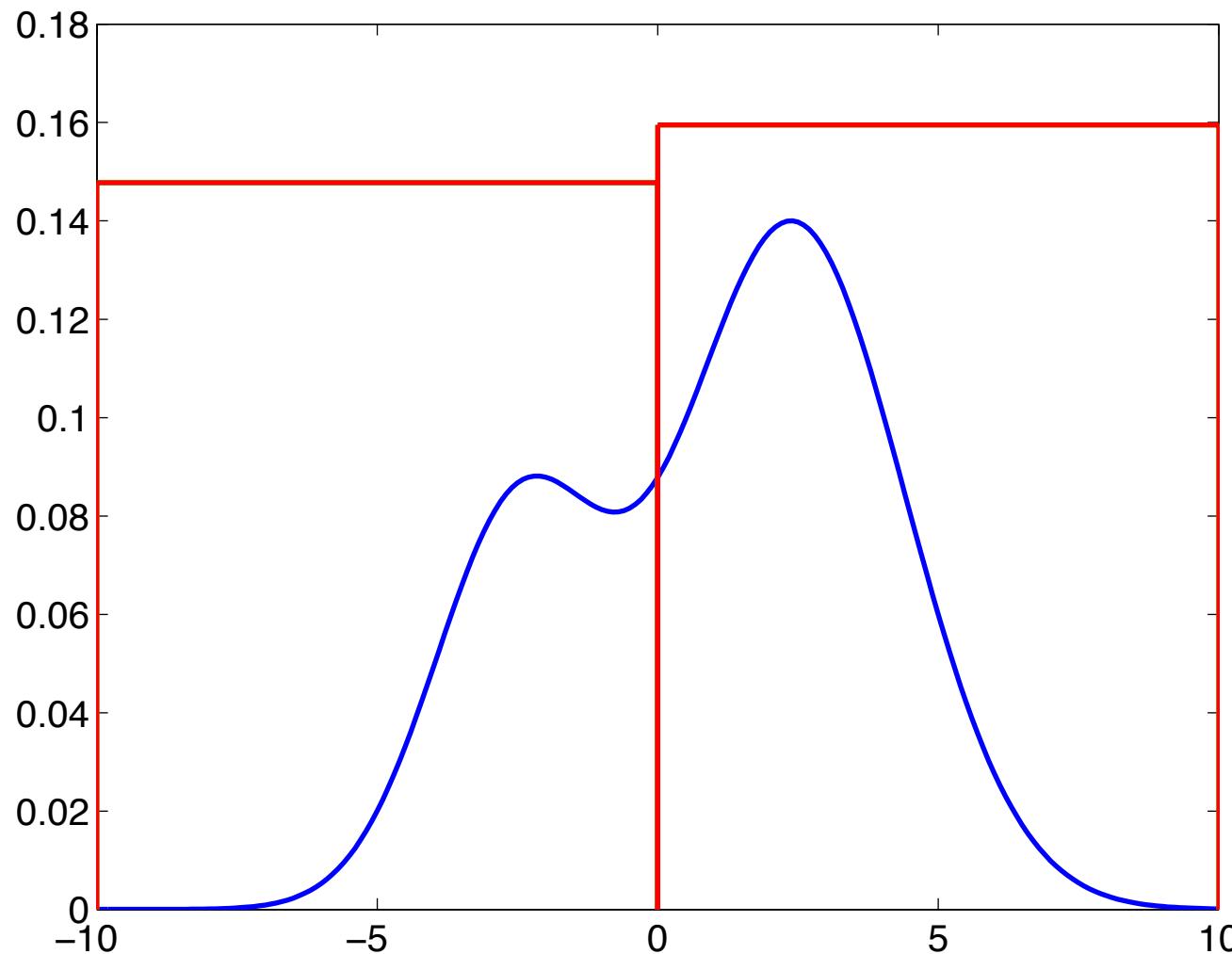
Bounding function $\bar{f}(X) \geq \max_{x \in X} f(x), \quad \bar{f}(\{x\}) = f(x)$



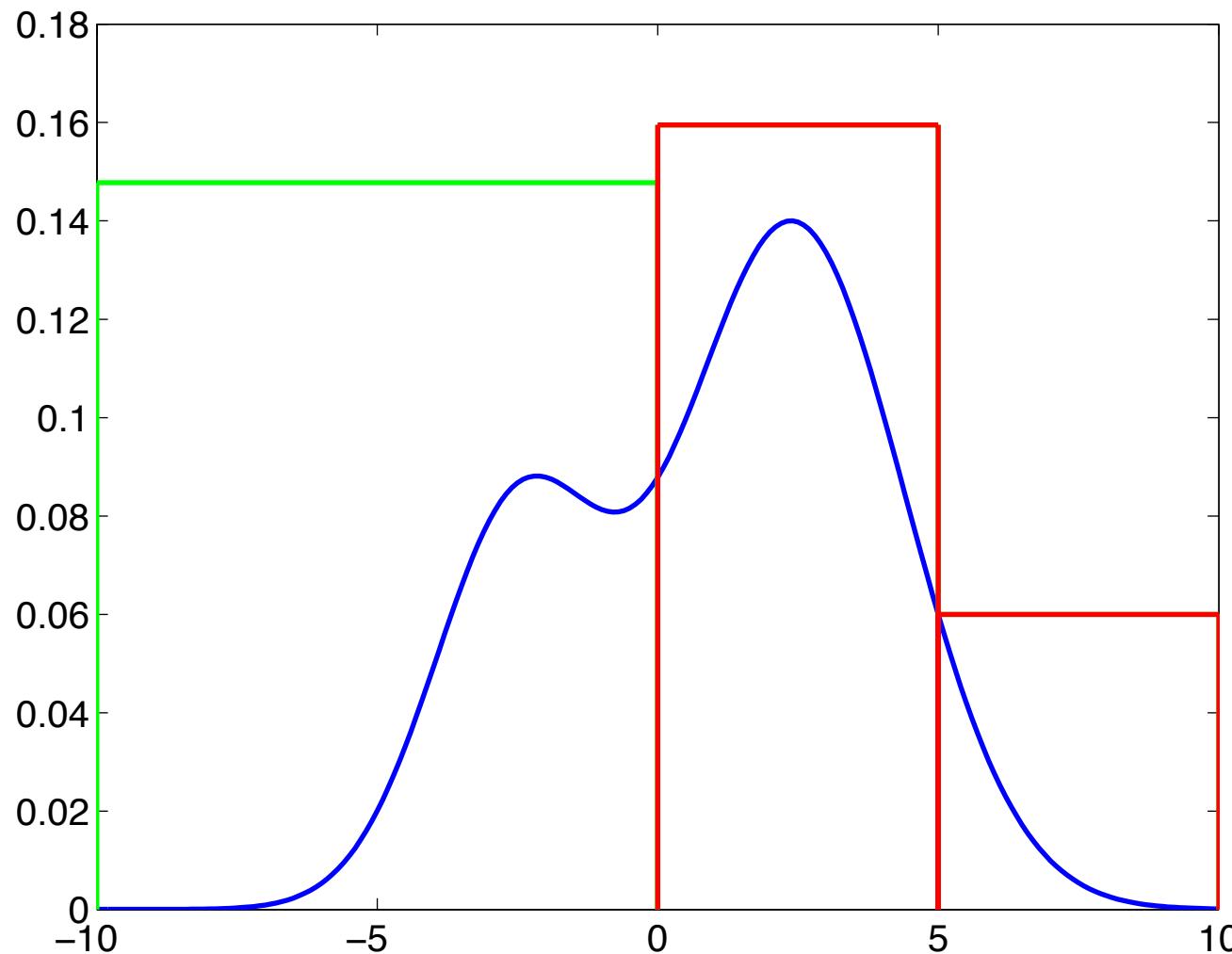
Bounding function $\bar{f}(X) \geq \max_{x \in X} f(x), \quad \bar{f}(\{x\}) = f(x)$



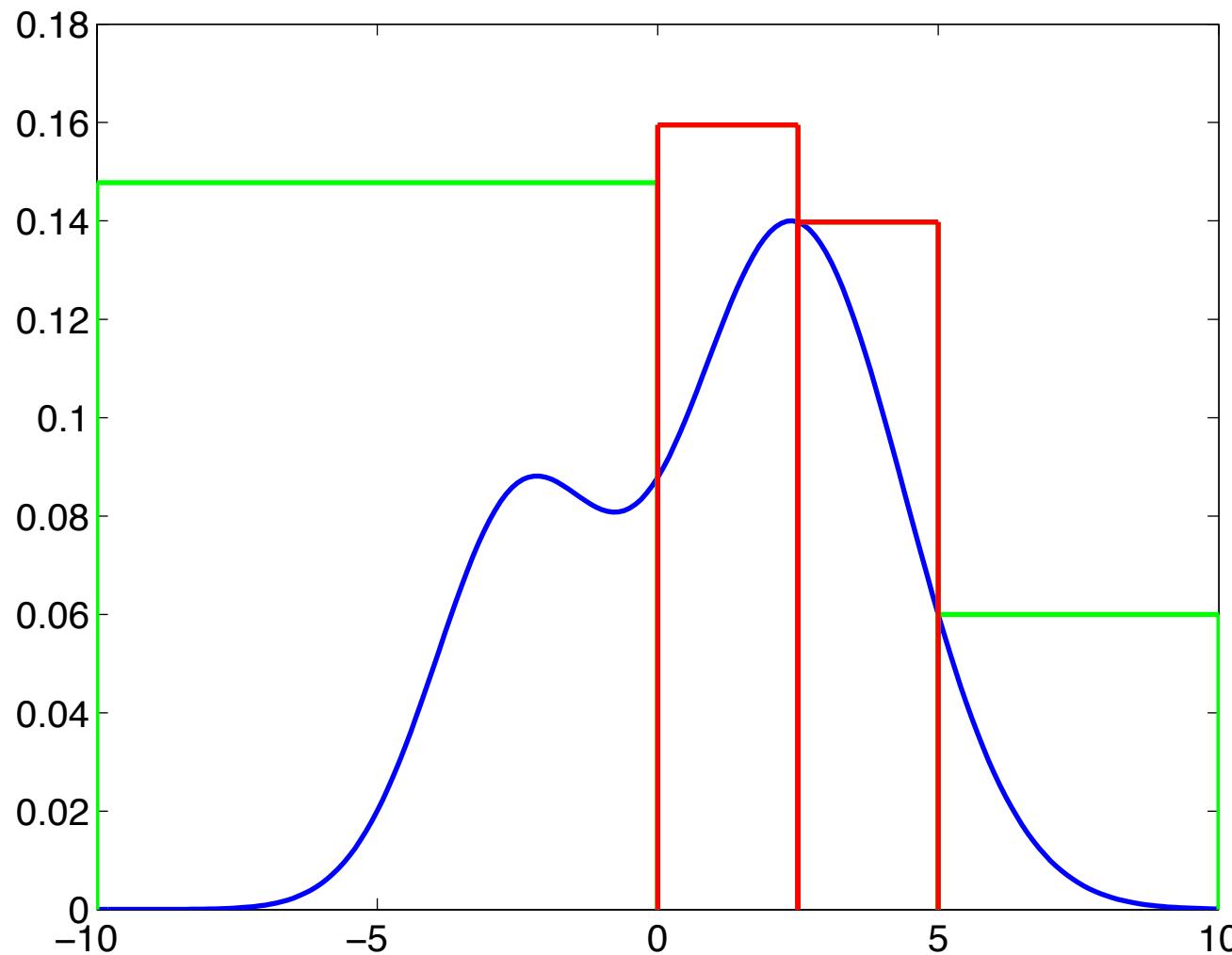
Branch-and-bound: all you need is bounds



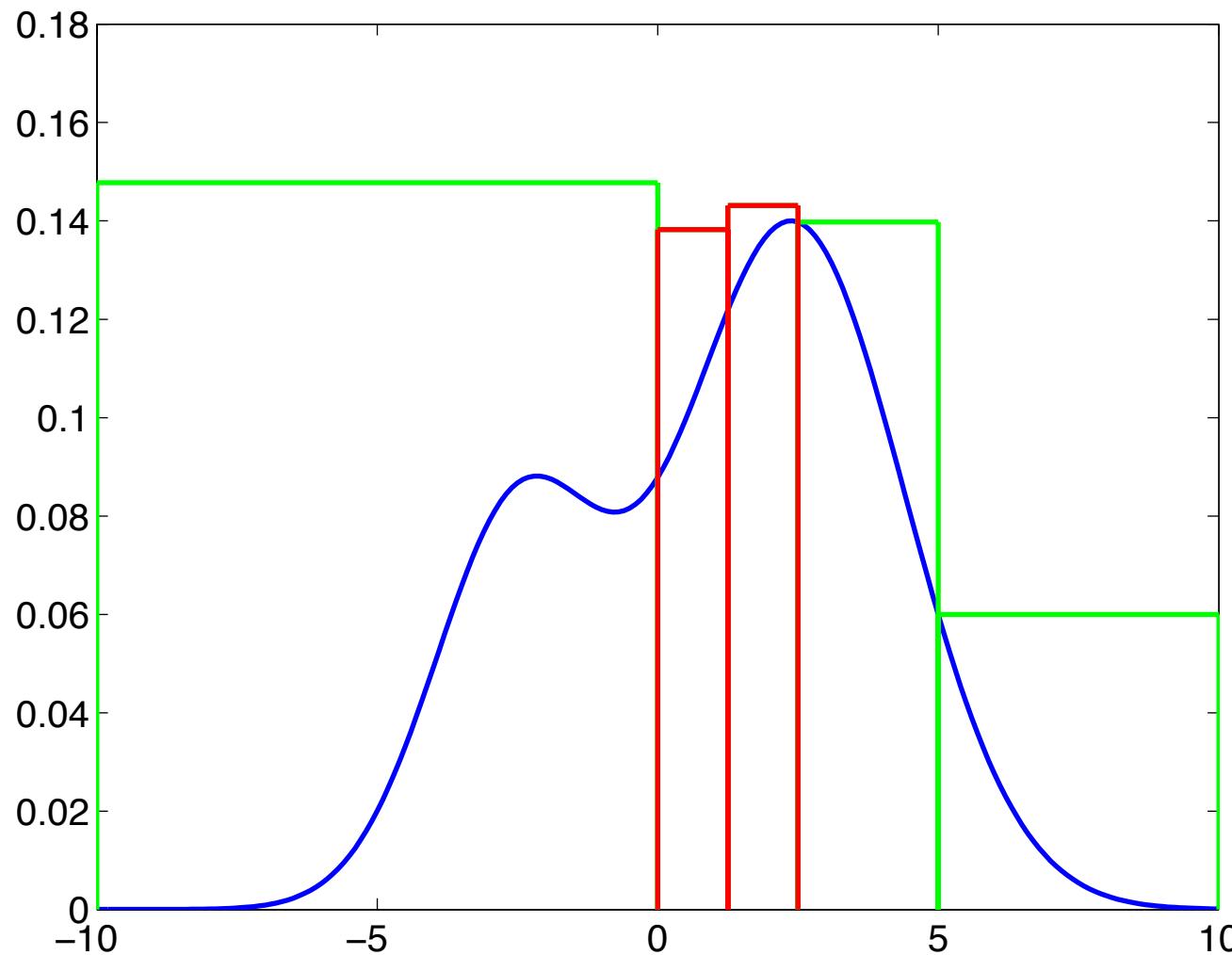
Branch-and-bound: all you need is bounds



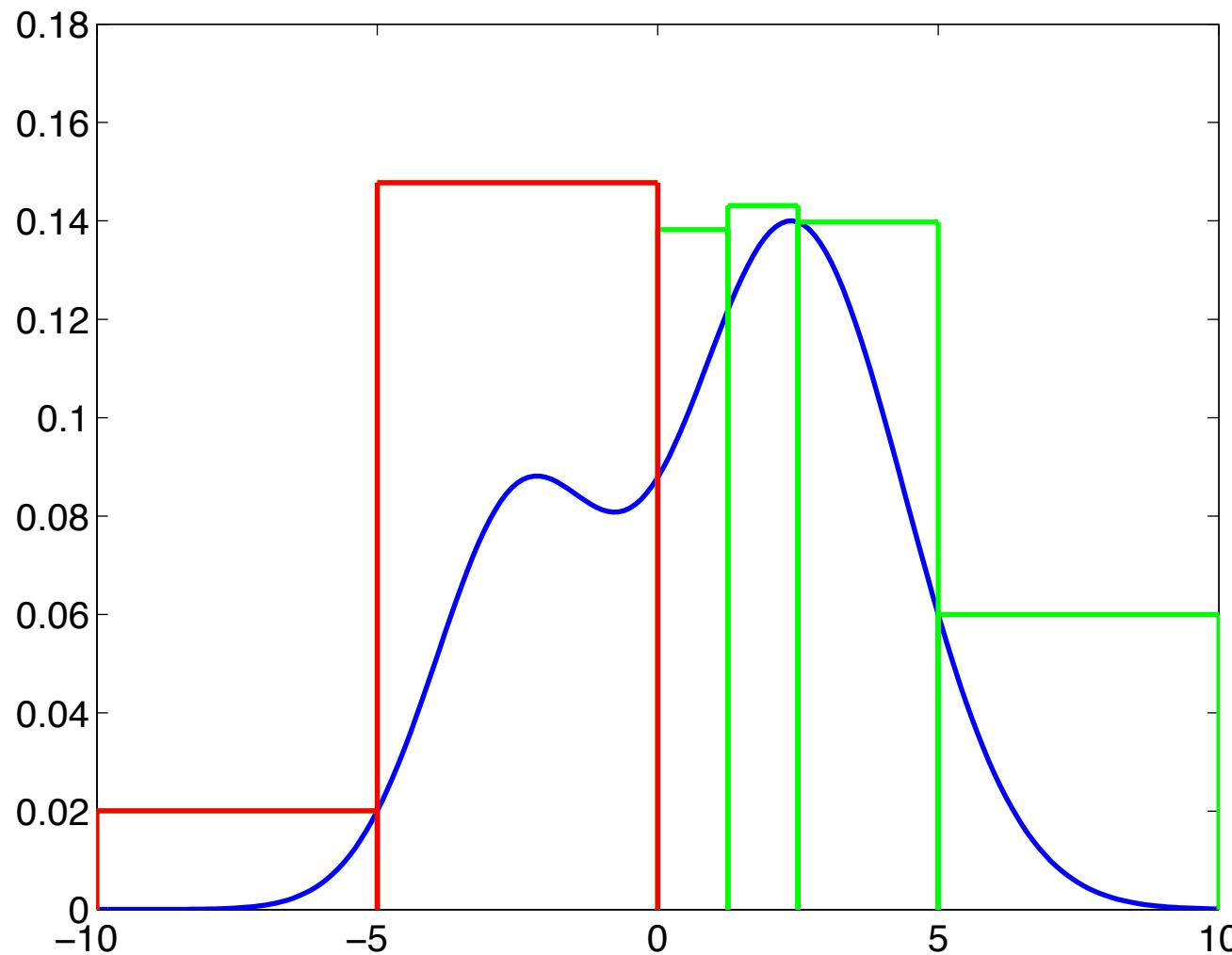
Branch-and-bound: all you need is bounds



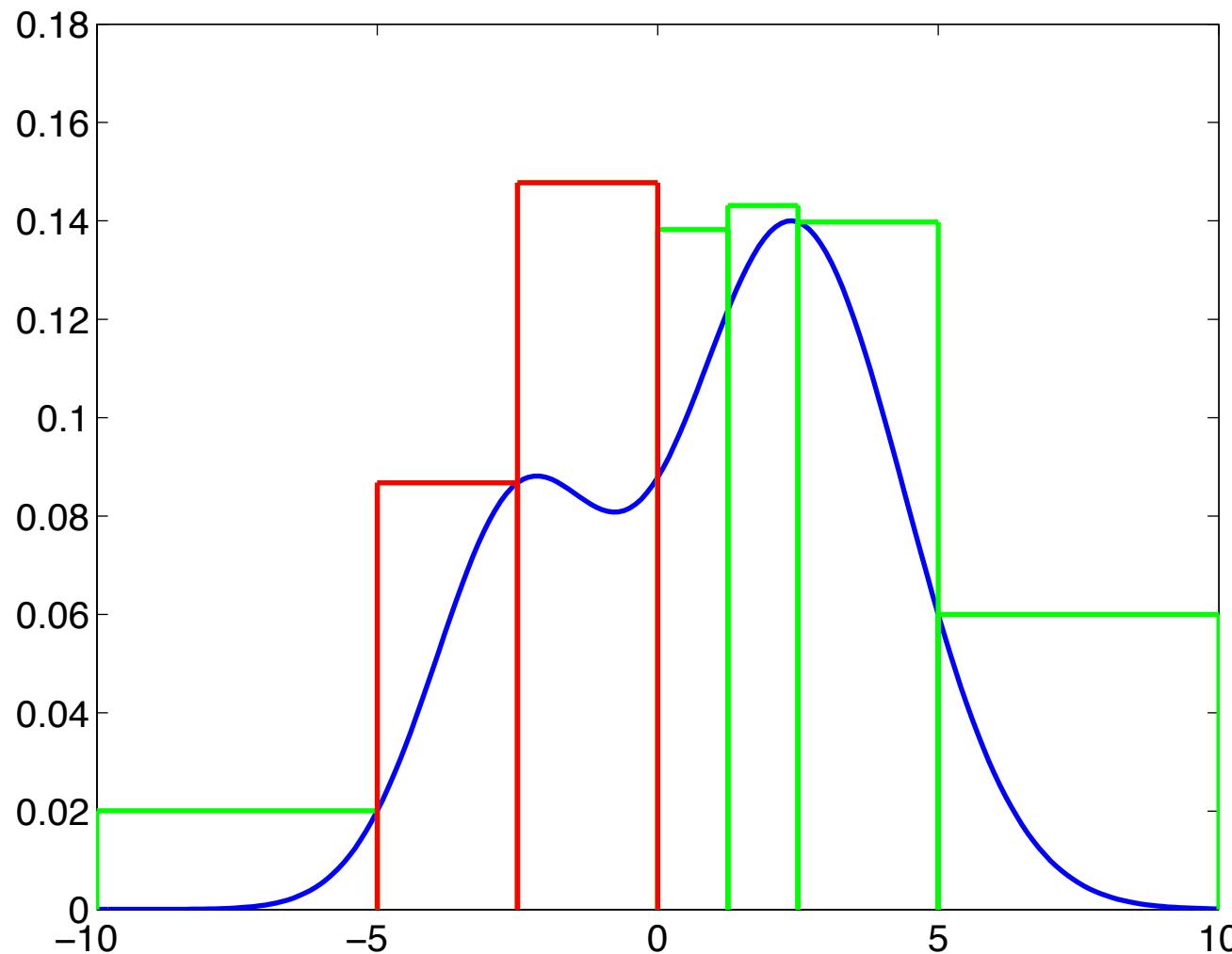
Branch-and-bound: all you need is bounds



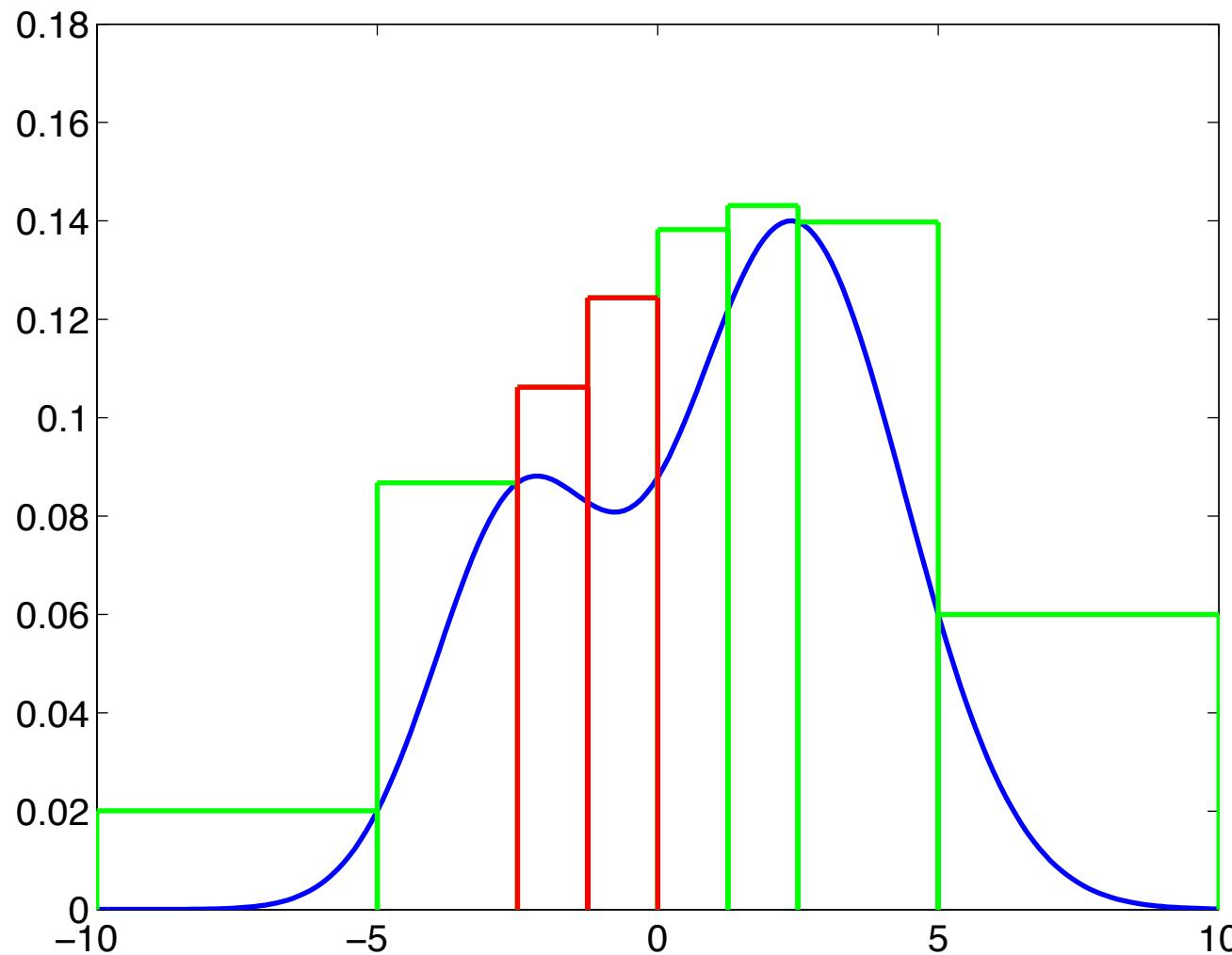
Branch-and-bound: all you need is bounds



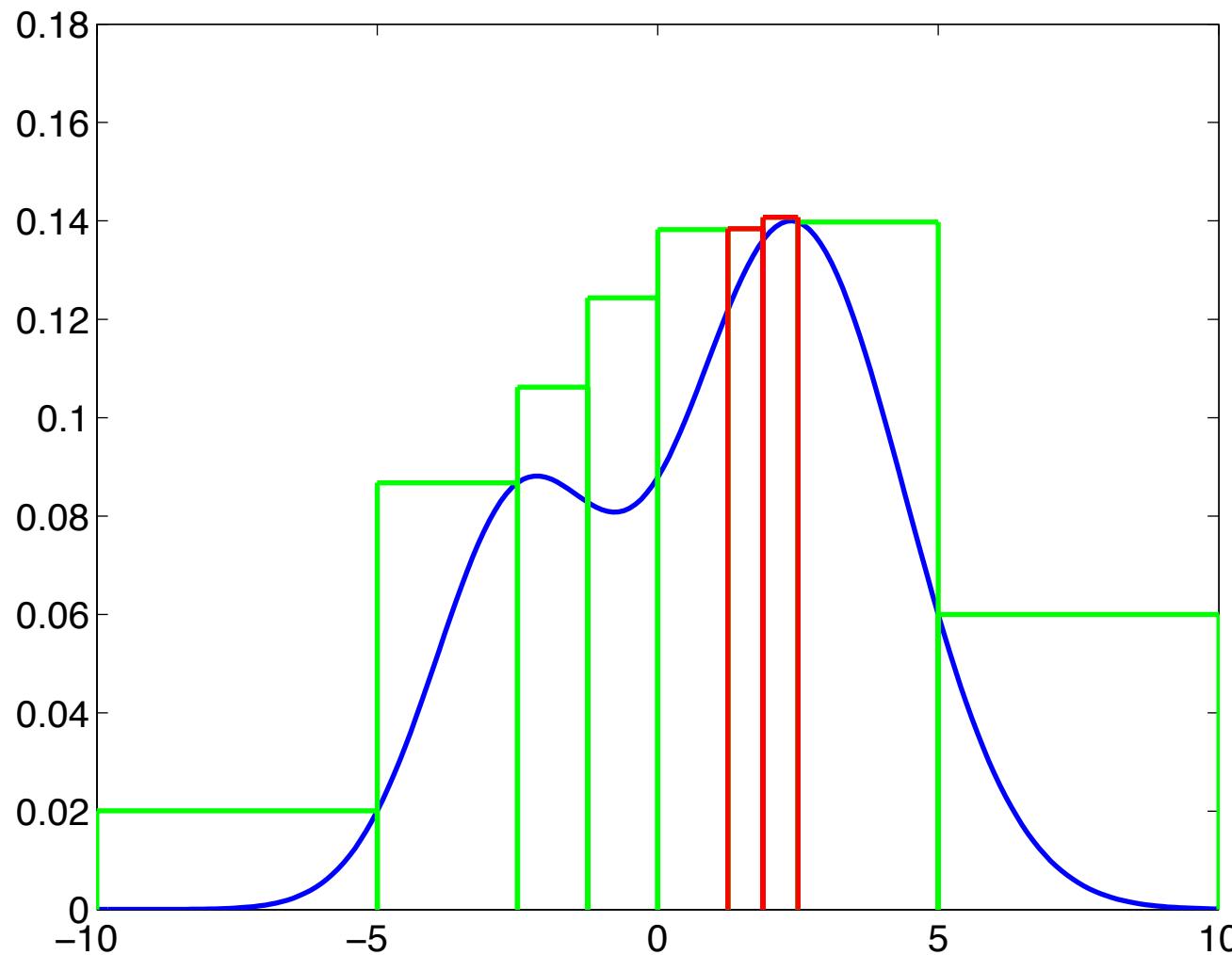
Branch-and-bound: all you need is bounds



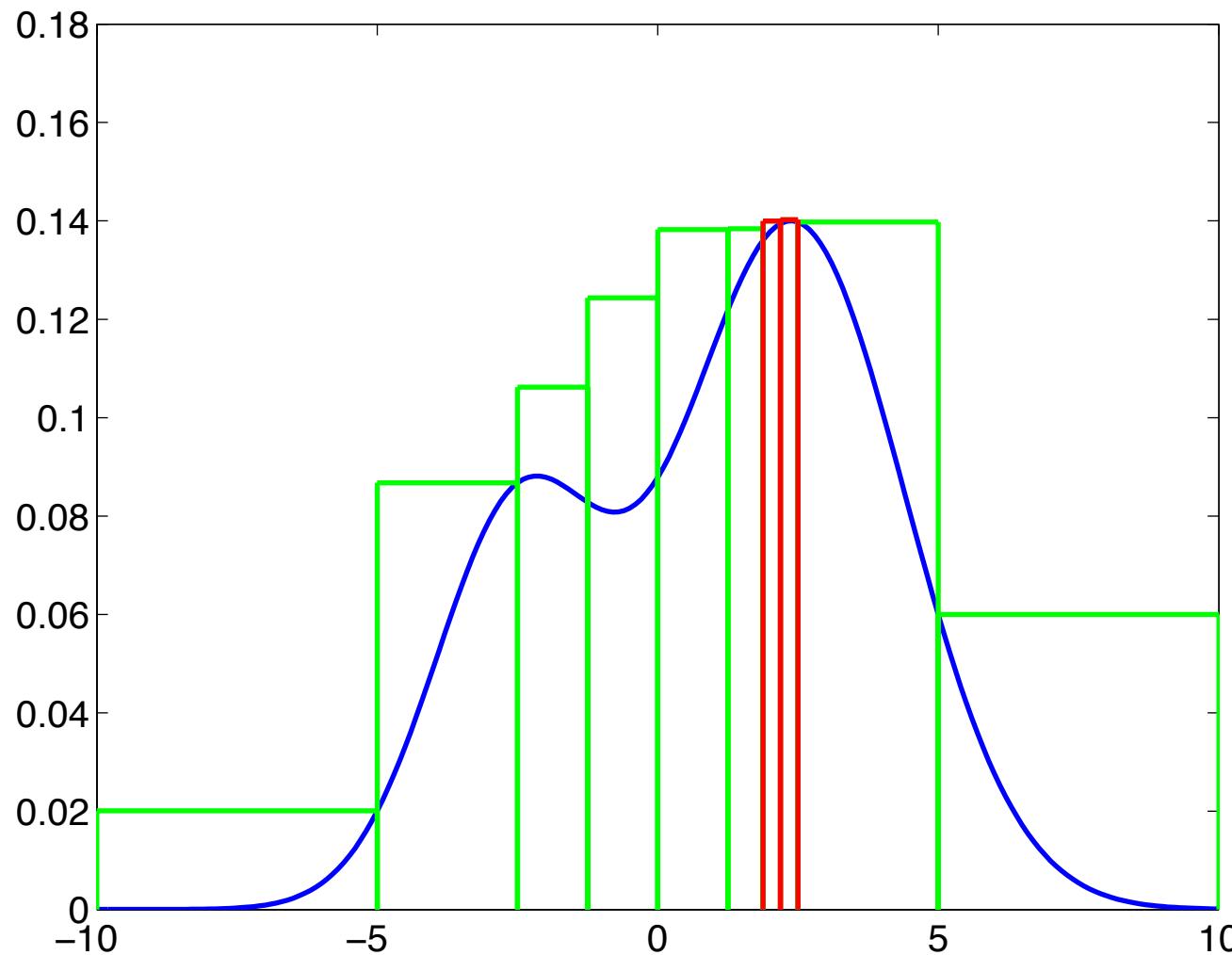
Branch-and-bound: all you need is bounds



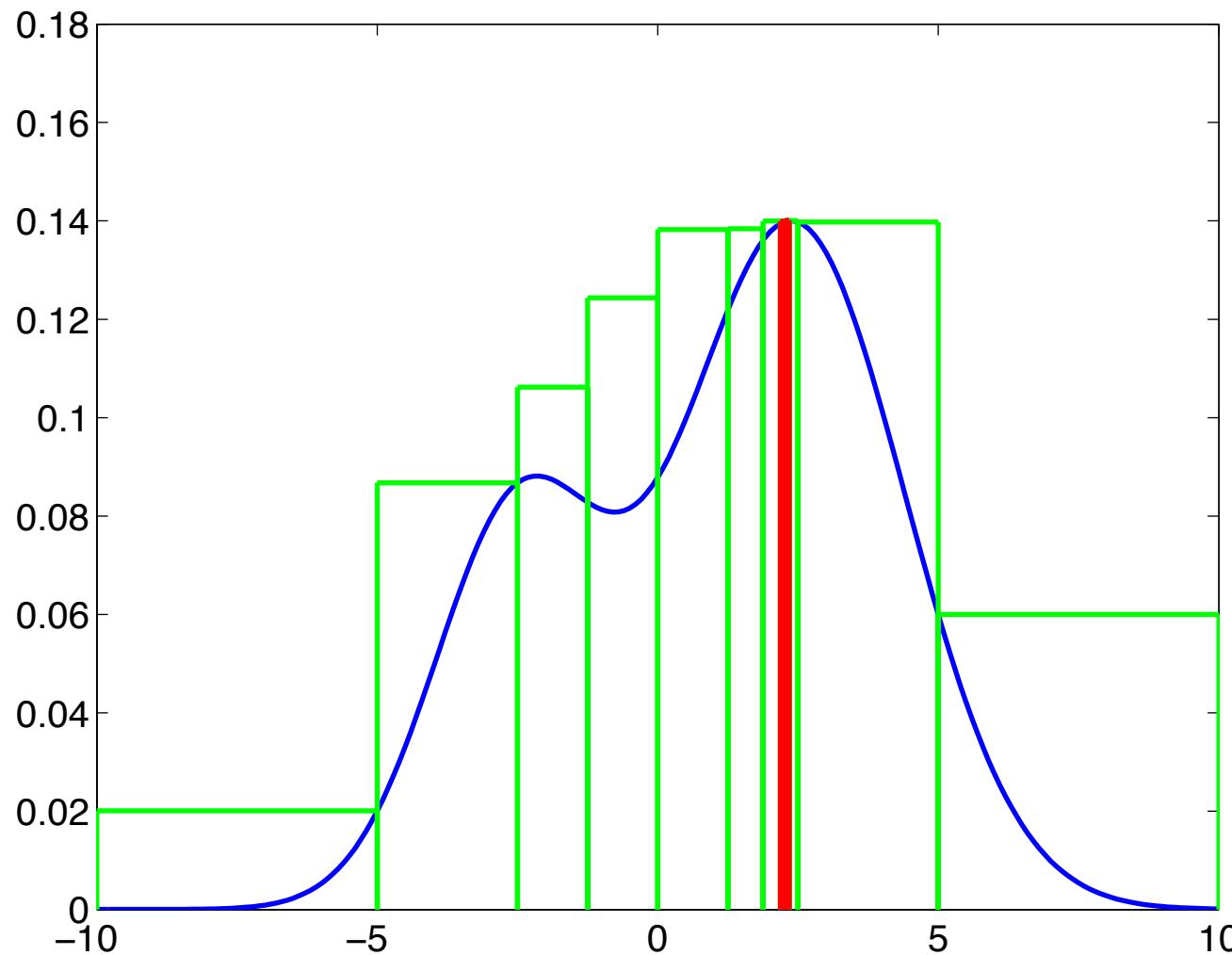
Branch-and-bound: all you need is bounds



Branch-and-bound: all you need is bounds



Branch-and-bound: all you need is bounds

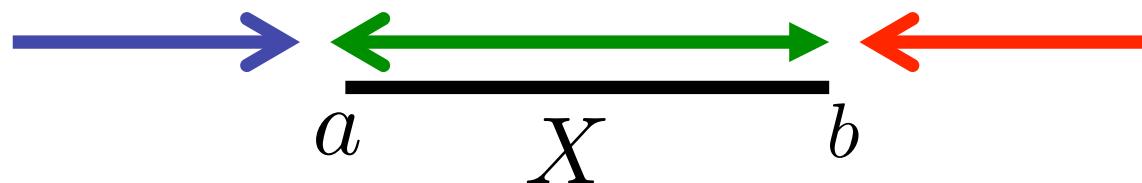


Bounding a mixture-of-gaussians

- Property: $\max_{x \in X} h(x) + g(x) \leq \max_{x \in X} h(x) + \max_{x \in X} g(x)$
- Function: $f(x) = \pi_1 N(x; \mu_1, \sigma_1) + \pi_2 N(x; \mu_2, \sigma_2)$

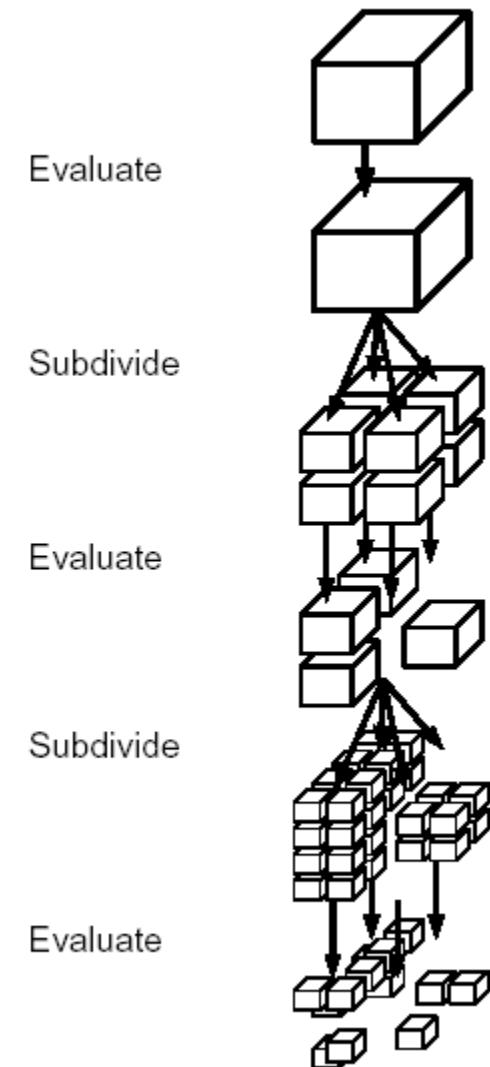
$$\begin{aligned} \max_{x \in X} f(x) &\leq \max_{x \in X} [\pi_1 N(x; \mu_1, \sigma_1)] + \max_{x \in X} [\pi_2 N(x; \mu_2, \sigma_2)] \\ &= \pi_1 N(d(X, \mu_1, \sigma_1); 0, 1) + \pi_2 N(d(X, \mu_2, \sigma_2); 0, 1) \end{aligned}$$

$$d(X, \mu, \sigma) = \begin{cases} 0 & a \leq \mu \leq b \\ \frac{1}{\sigma^2}(\mu - a)^2 & \mu \leq a \\ \frac{1}{\sigma^2}(\mu - b)^2 & b < \mu \end{cases}$$



Branch and Bound Illustration

- Guaranteed (or admissible) search heuristic
 - Bound on how good answer could be in unexplored region
 - Cannot miss an answer
 - In worst case won't rule anything out
- In practice rule out vast majority of transformations
 - Can use even simpler tests than computing distance at cell center



Branch-and-Bound -1: Bag-of-words models

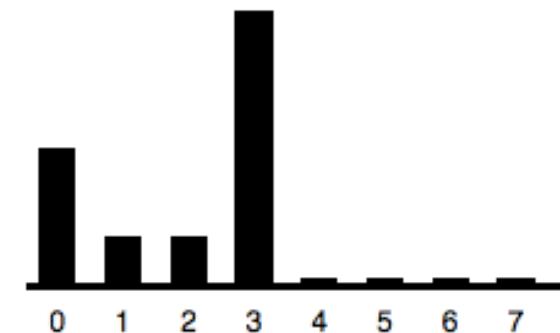
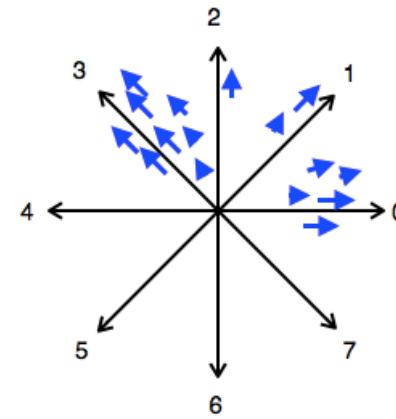
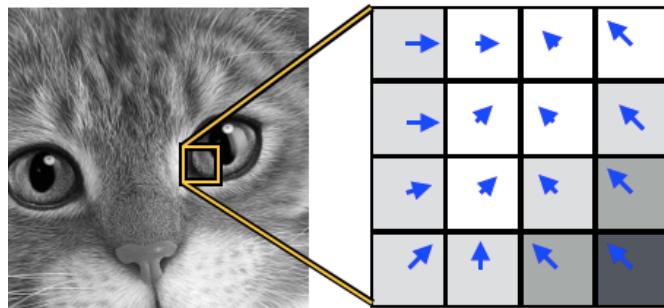
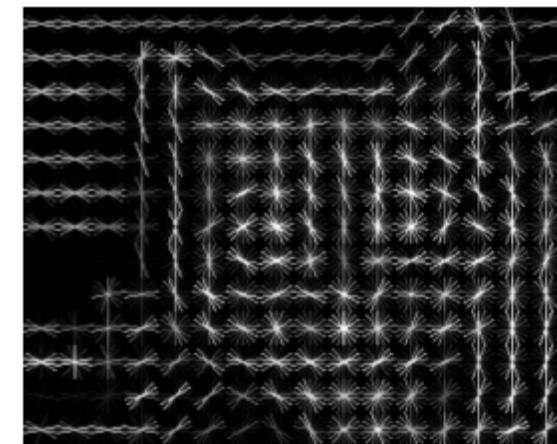
Beyond Sliding Windows: Object Localization by
Efficient Subwindow Search

Christoph H. Lampert[†], Matthew B. Blaschko[†], & Thomas Hofmann[‡]



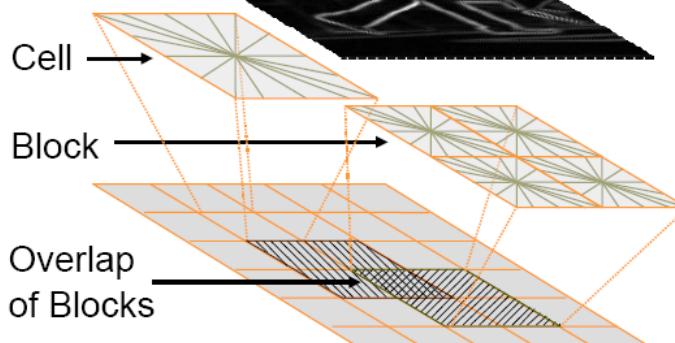
CVPR 2008 best paper award

Histogram of Gradient (HOG)/SIFT Features



Dalal and Triggs, ICCV 2005

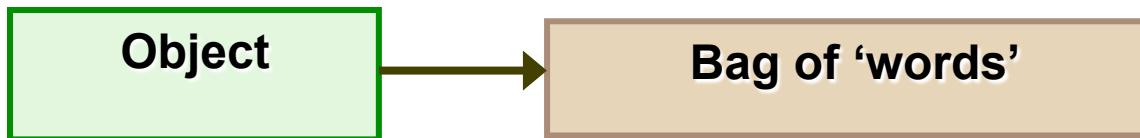
- Histogram of Oriented Gradient (HOG) features
- Highly accurate detection using linear SVM



Feature vector $f = [\dots, \dots, \dots]$

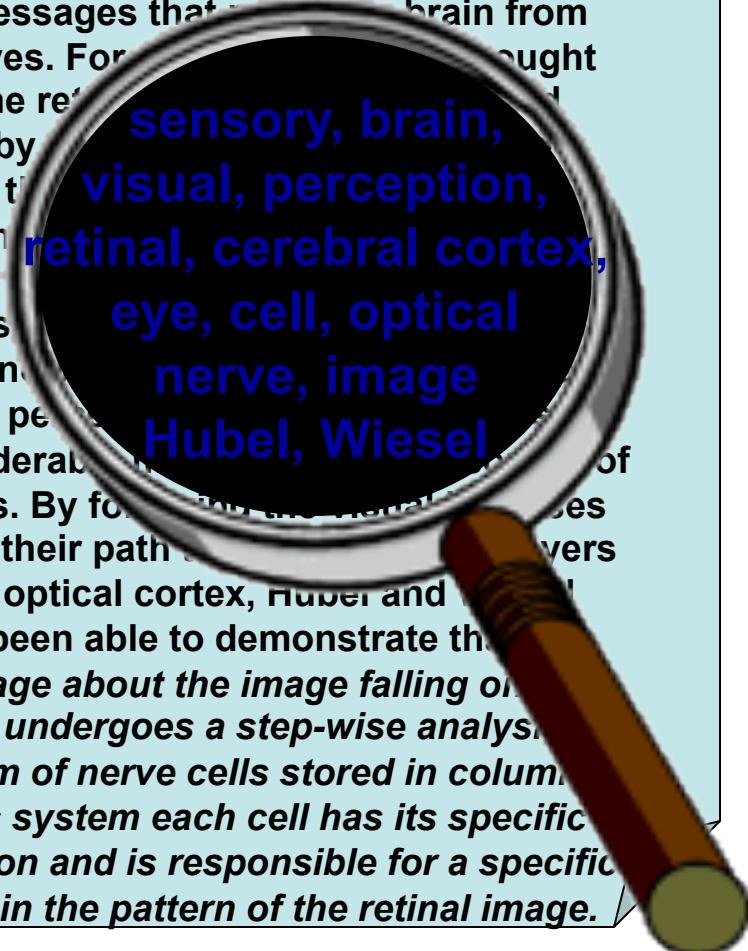


Bag-of-word models



Analogy to documents

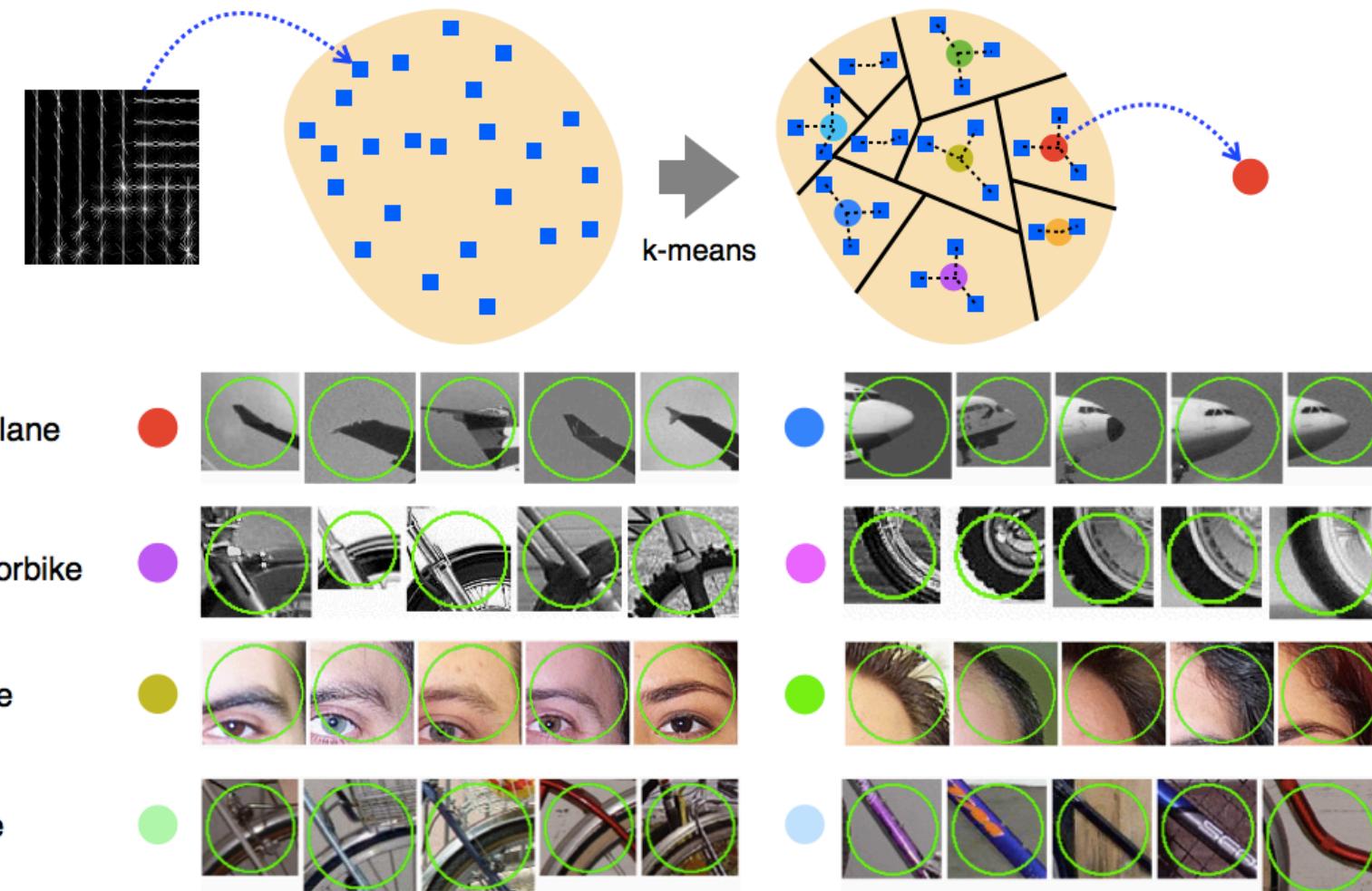
Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that reach the brain from our eyes. For a long time it was thought that the retina sent a single point by point message to the brain; the screen image of the distant object now known to be a complex visual perception. Considerable work has been done on the visual system of the eye and brain. By following the visual messages along their path through the various layers of the optical cortex, Hubel and Wiesel have been able to demonstrate that the message about the image falling on the retina undergoes a step-wise analysis. A system of nerve cells stored in columns. In this system each cell has its specific function and is responsible for a specific detail in the pattern of the retinal image.



China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be created by a predicted 30% jump in exports to \$750bn, with a 18% rise in imports. The figures are likely to be revised upwards. China has long complained of unfair trading practices underpinning its trade surplus. Zhou Xiaochuan, governor of the central bank, said the country needed to encourage more domestic demand so as to reduce the value of the yuan against the dollar. The US has already permitted it to trade within a narrow band, but the US wants the yuan to be allowed to trade freely. However, Beijing has made it clear that it will take its time and tread carefully before allowing the yuan to rise further in value.

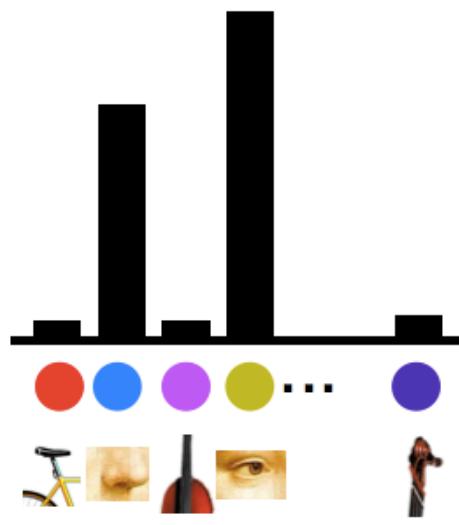


Feature quantization



Visual words representation

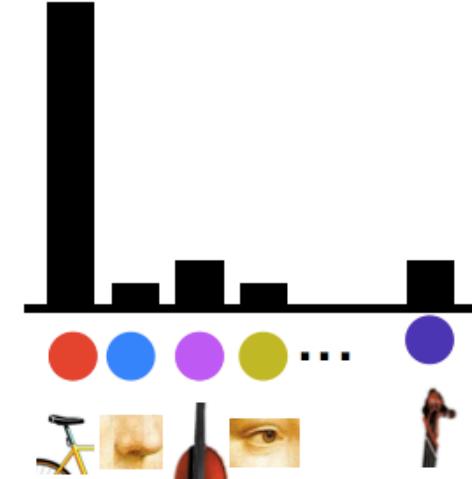
Visual words represent “iconic” image fragments



person



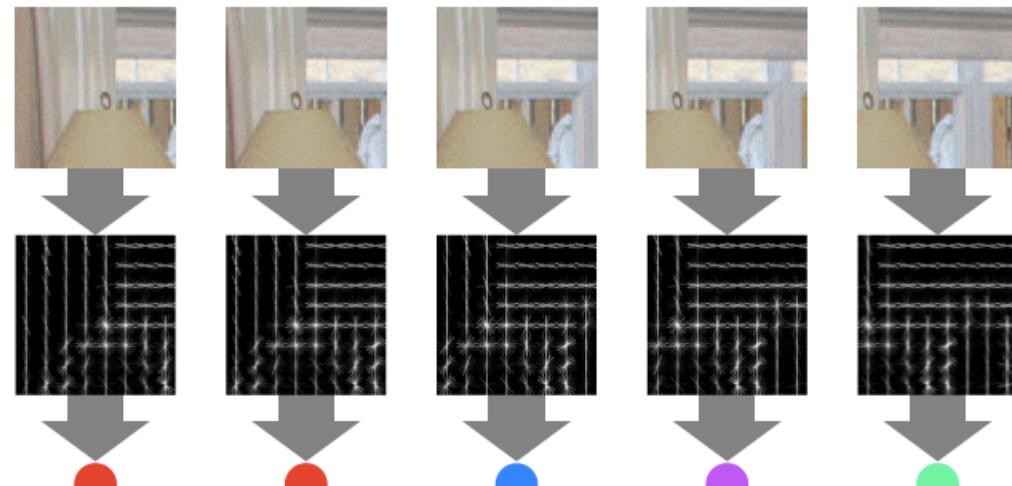
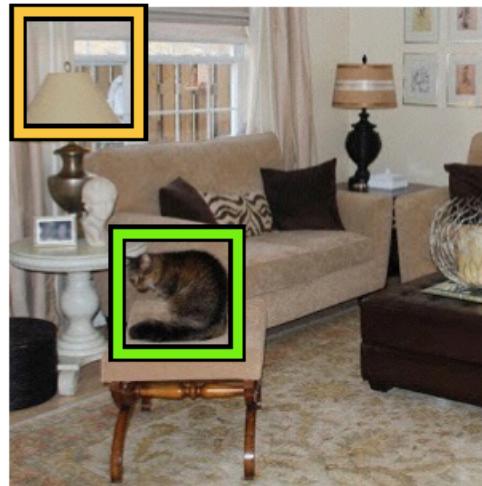
musical instrument



bike

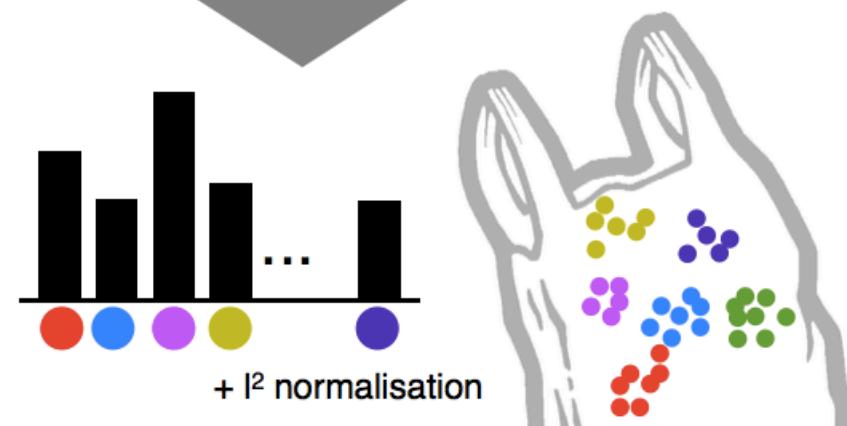
Bag-of-word features

[Sivic & Zisserman 2003, Csurka *et al.* 2004, Nowak *et al.* 2006]



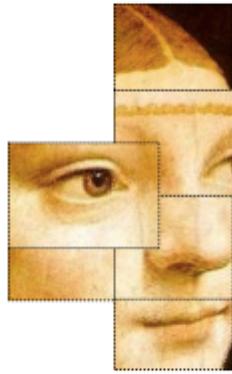
BoVW construction

1. Extract local descriptor densely
 2. Quantise descriptors
 3. Form histogram
- 2. Discards spatial information**



Bag of words limitations

image



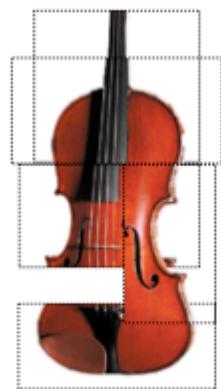
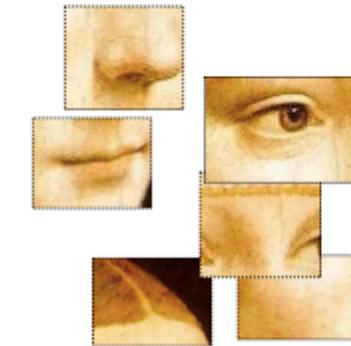
=

plausible deformation

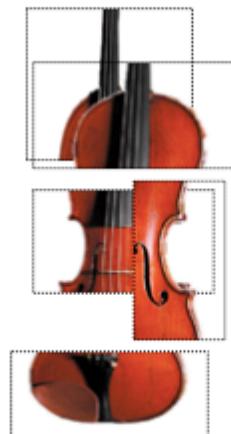


=

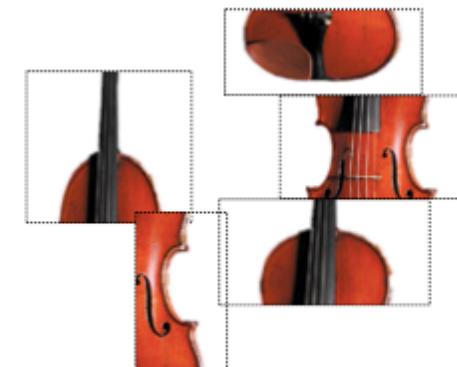
implausible deformation



=



=



Spatially-sensitive bag-of-words

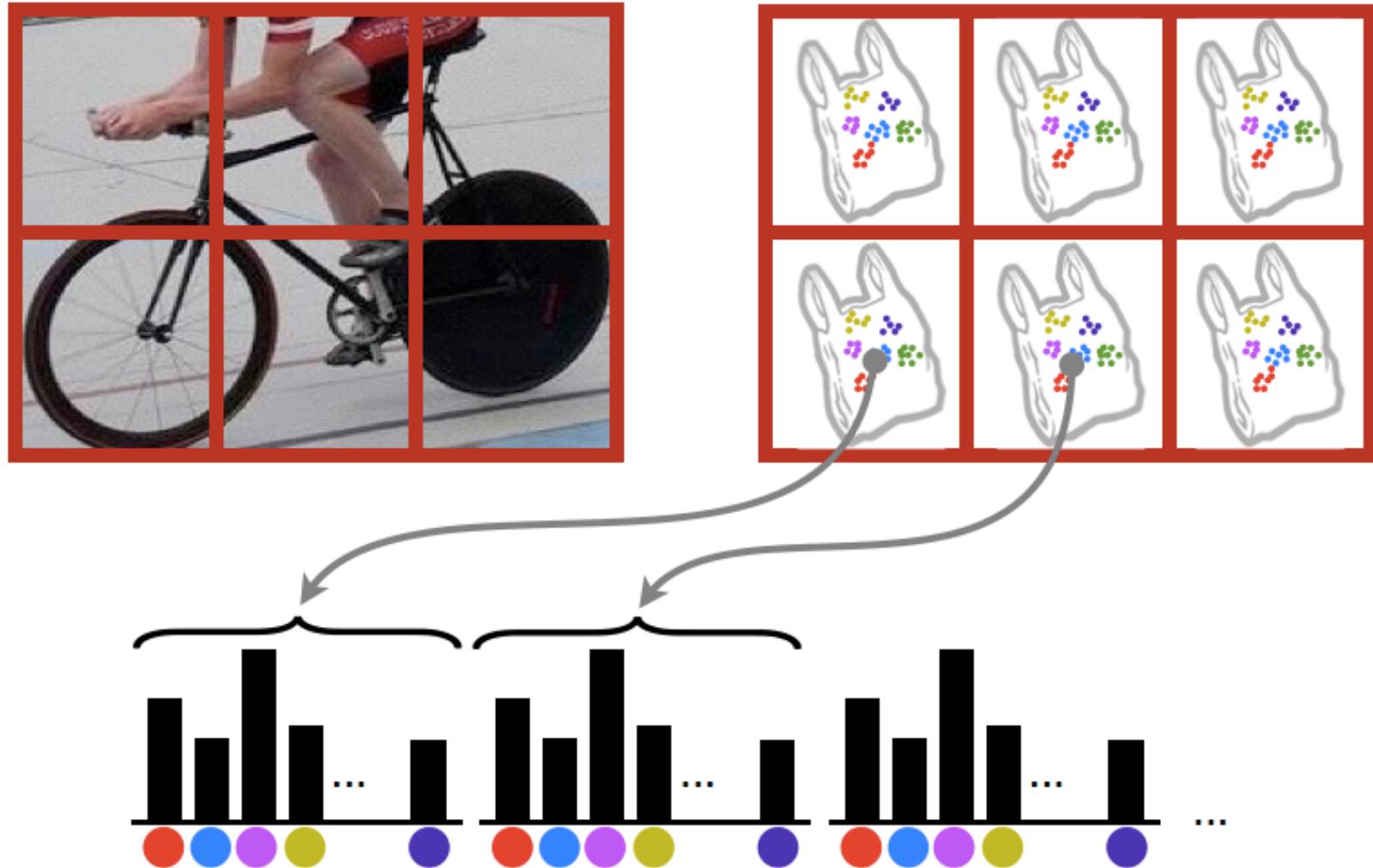
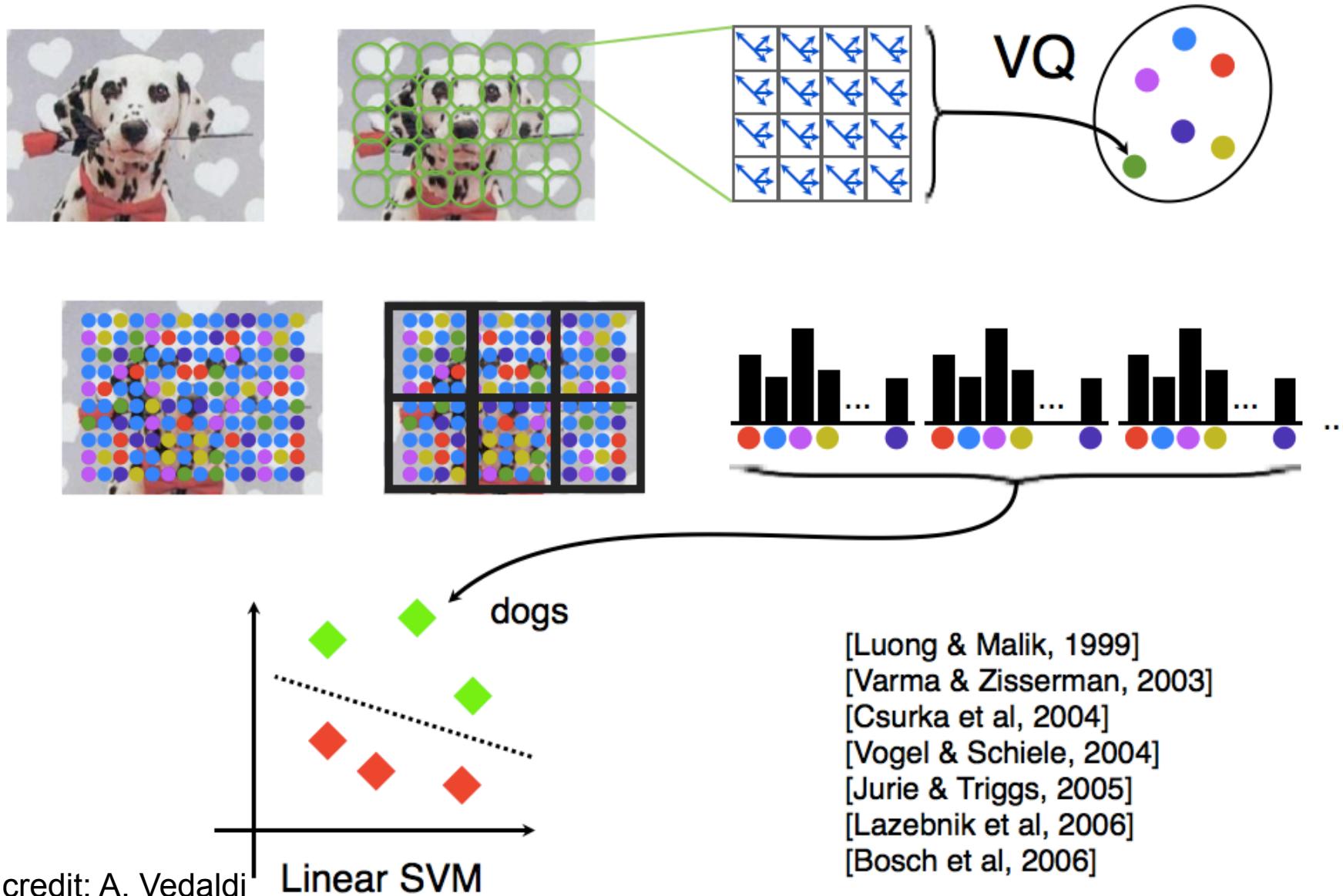
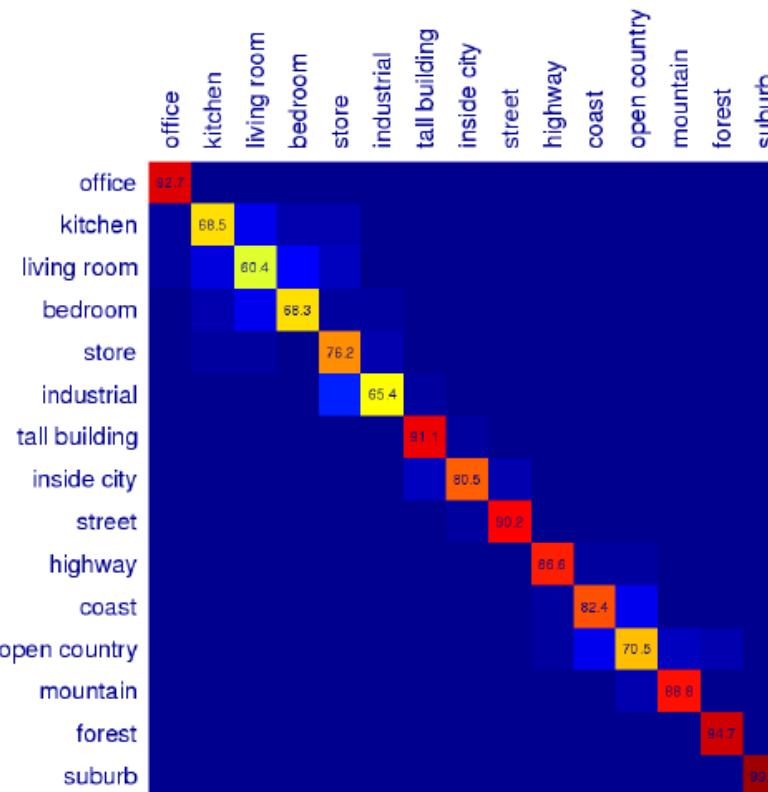


Image classification in a nutshell



Scene category confusions



Difficult indoor images



kitchen



living room



bedroom

Sliding window classifiers

approach: sliding window classifier

- evaluate classifier at candidate regions in an image - $\text{argmax}_{B \in \mathcal{B}} f_I(B)$
- for a 640×480 pixel image, there are over *10 billion* possible regions to evaluate

sample a subset of regions to evaluate

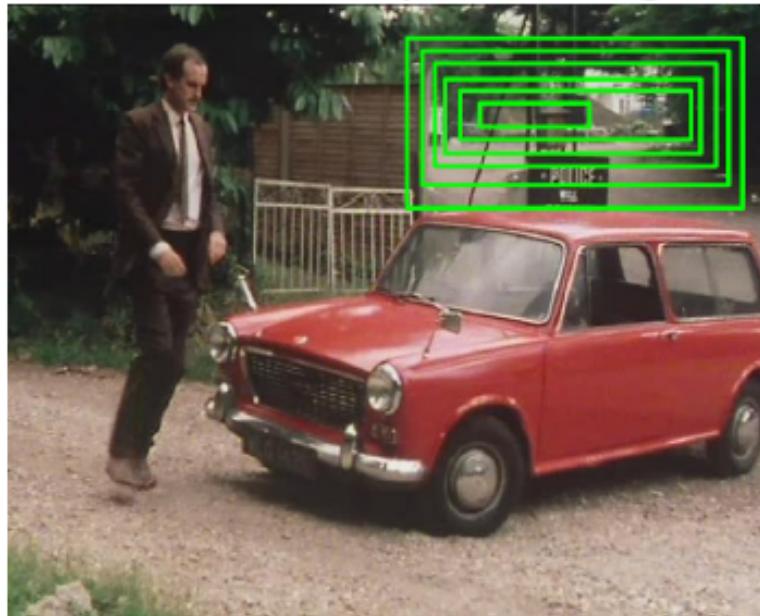
- scale
- aspect ratio
- grid size



Beyond sliding windows

Problem: Exhaustive evaluation of $\text{argmax}_{B \in \mathcal{B}} f_I(B)$ is too slow.

Solution: Use the problem's *geometric structure*.

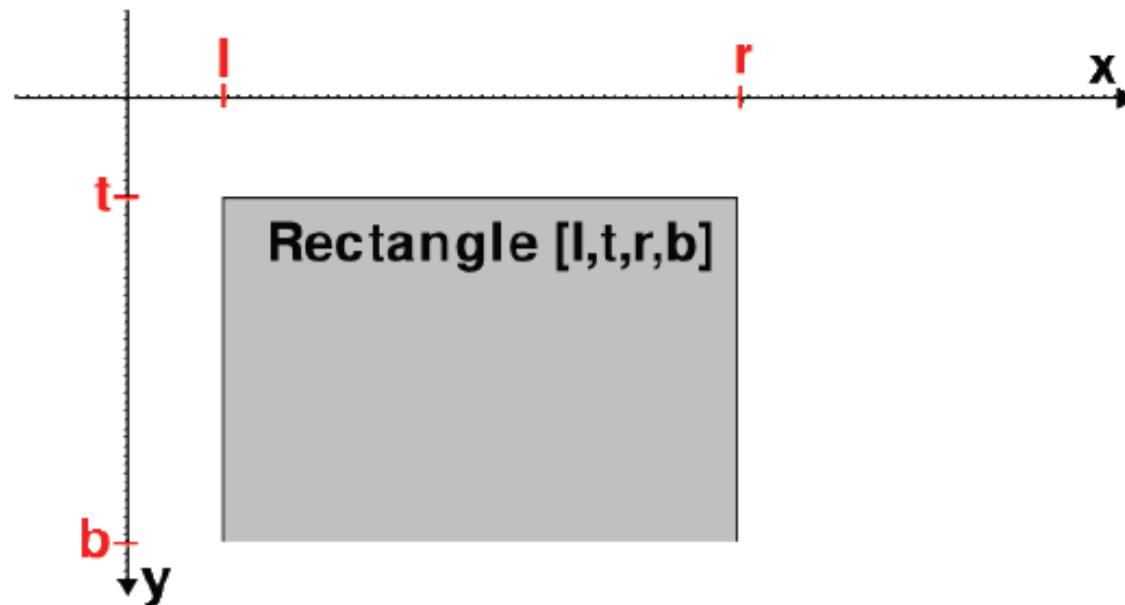


- Similar boxes have similar scores.
- Calculate scores for *sets of boxes* jointly (upper bound).
- If no element can contain the object, discard the set.
- Else, split the set into smaller parts and re-check, etc.

⇒ efficient branch & bound algorithm

Parameterization of solution

- low dimensional parametrization of bounding box
(left, top, right, bottom)



Parameterization of solution interval

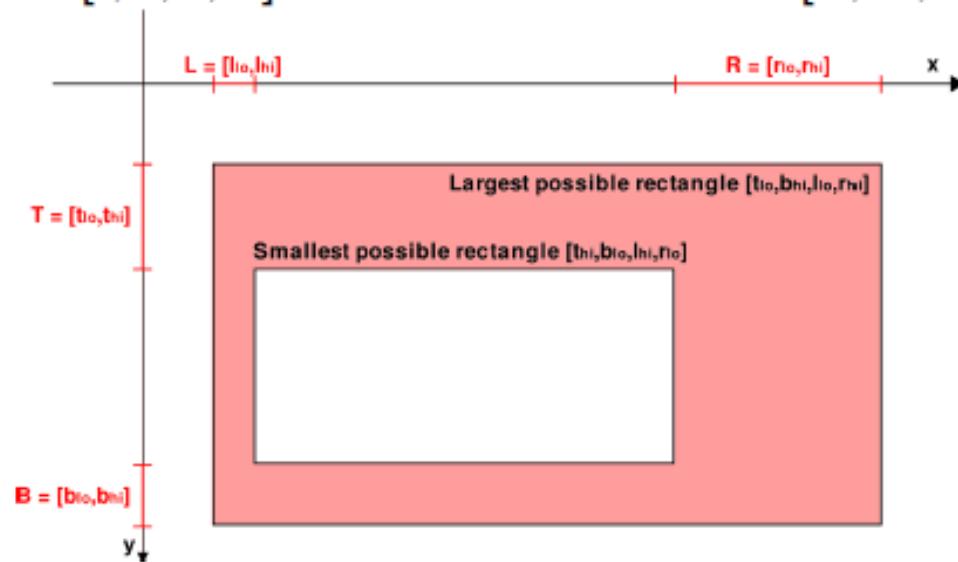
- Instead of four numbers $[l, t, r, b]$, store four intervals $[L, T, R, B]$:

$$L = [l_{lo}, l_{hi}]$$

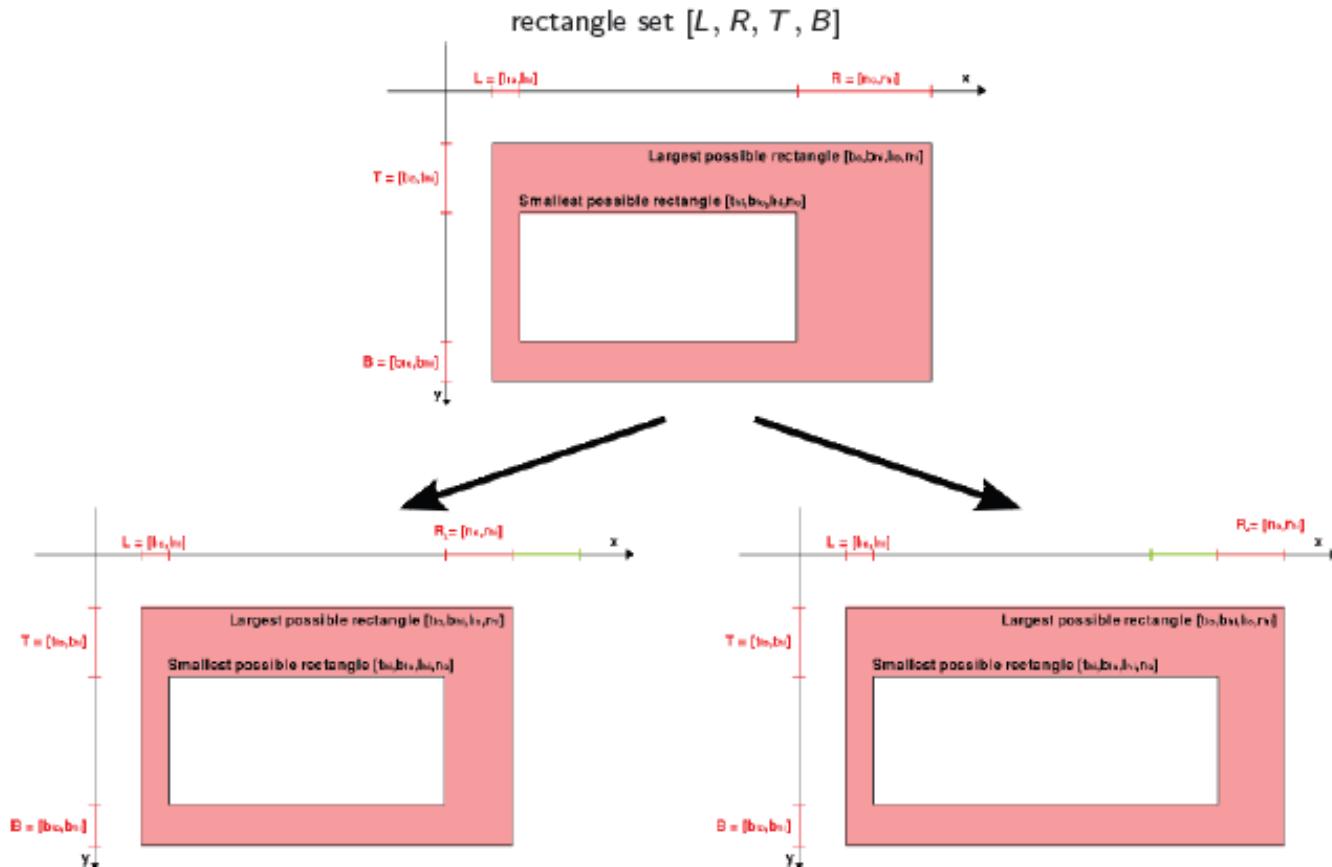
$$T = [t_{lo}, t_{hi}]$$

$$R = [r_{lo}, r_{hi}]$$

$$B = [b_{lo}, b_{hi}]$$



Branching of solution interval



$[L, R_1, T, B]$ with $R_1 := [r_{lo}, \lfloor \frac{r_{lo} + r_{hi}}{2} \rfloor]$

$[L, R_2, T, B]$ with $R_2 := [\lfloor \frac{r_{lo} + r_{hi}}{2} \rfloor + 1, r_{hi}]$

Bounding a solution interval

We have to construct $f^{upper} : \{ \text{set of boxes} \} \rightarrow \mathbb{R}$ such that

- i) $f^{upper}(\mathcal{B}) \geq \max_{B \in \mathcal{B}} f(B),$
- ii) $f^{upper}(\mathcal{B}) = f(B), \quad \text{if } \mathcal{B} = \{B\}.$

$$f(B) = \sum_j \alpha_j \langle h^B, h^j \rangle \quad h^B \text{ the histogram of the box } B.$$

$$= \sum_j \alpha_j \sum_k h_k^B h_k^j = \sum_k h_k^B w_k, \quad \text{for } w_k = \sum_j \alpha_j h_k^j$$

$$= \sum_{x_i \in B} w_{c_i}, \quad c_i \text{ the cluster ID of the feature } x_i$$

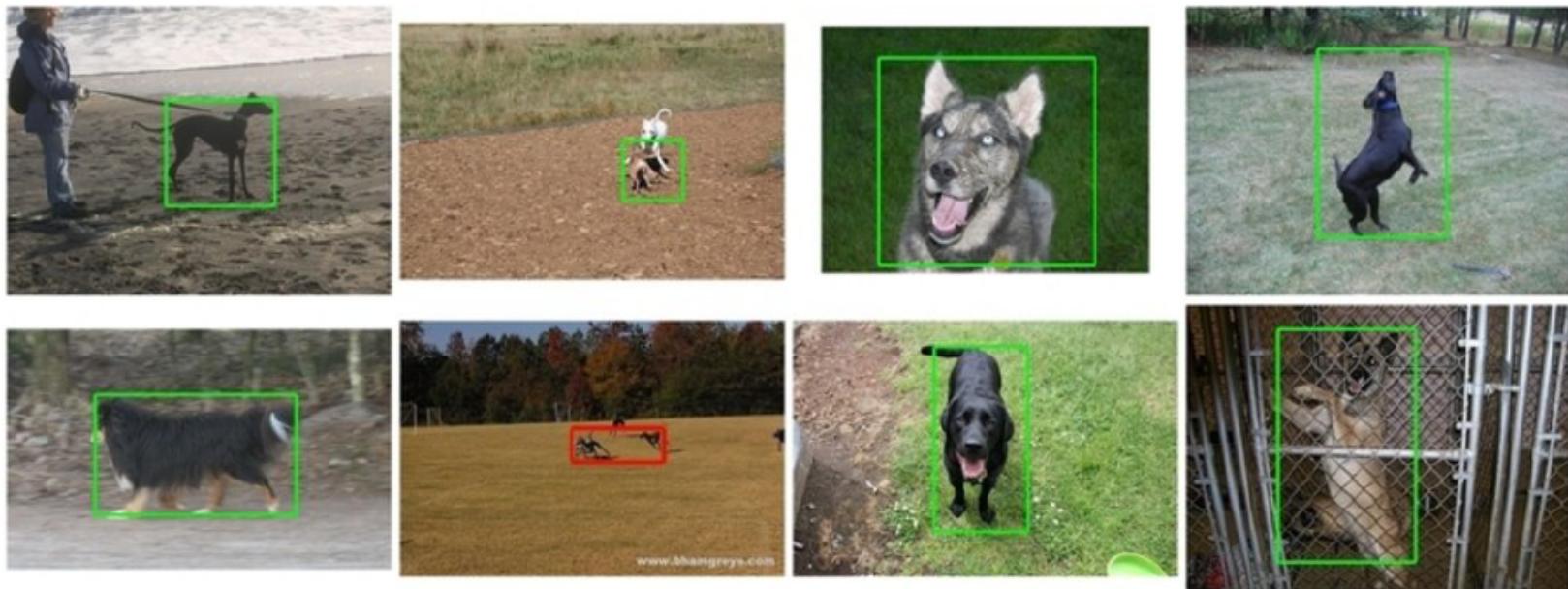
$$\text{Set } f^+(B) = \sum_{x_i \in B} [w_i]_+, \quad f^-(B) = \sum_{x_i \in B} [w_i]_-.$$

Set $B^{max} := \text{largest box in } \mathcal{B}, \quad B^{min} := \text{smallest box in } \mathcal{B}.$

$$f^{upper}(\mathcal{B}) := f^+(B^{max}) + f^-(B^{min}) \quad \text{fulfills i) and ii).}$$

Evaluating $f^{upper}(\mathcal{B})$ has same complexity as $f(B)!$

- High localization quality: first place in 5 of 20 categories.
- High speed: $\approx 40ms$ per image (excl. feature extraction)



Example detections on VOC 2007 dog.

Branch-and-bound localization allows efficient extensions:

- Multi-Class Object Localization:

$$(B, C)^{\text{opt}} = \underset{B \in \mathcal{B}, C \in \mathcal{C}}{\operatorname{argmax}} f_I^C(B)$$

finds best object class $C \in \mathcal{C}$.

- Localized retrieval from image databases or videos

$$(I, B)^{\text{opt}} = \underset{B \in \mathcal{B}, I \in \mathcal{D}}{\operatorname{argmax}} f_I(B)$$

find best image I in database \mathcal{D} .

Runtime is *sublinear* in $|\mathcal{C}|$ and $|\mathcal{D}|$.



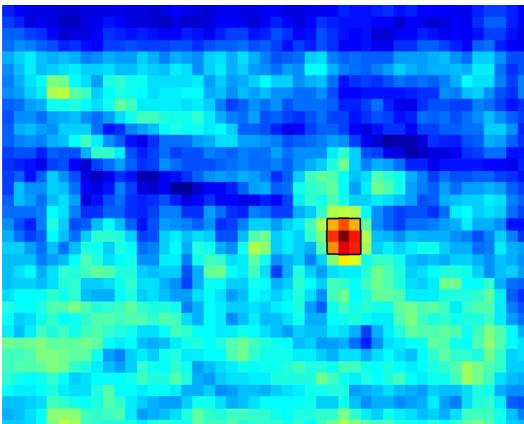
Nearest Neighbor query for *Red Wings* Logo in 10,000 video keyframes in "Ferris Buellers Day Off"

Branch-and-Bound 2: Deformable Part Models

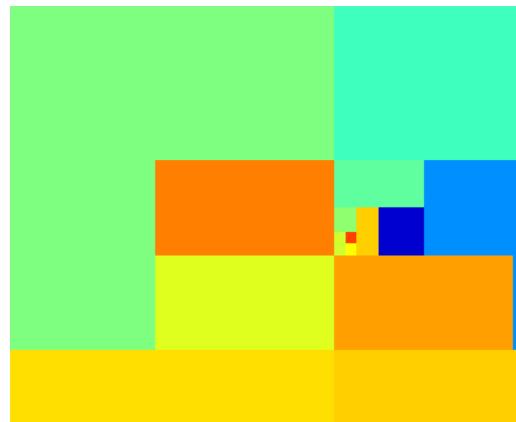
Input & Detection result



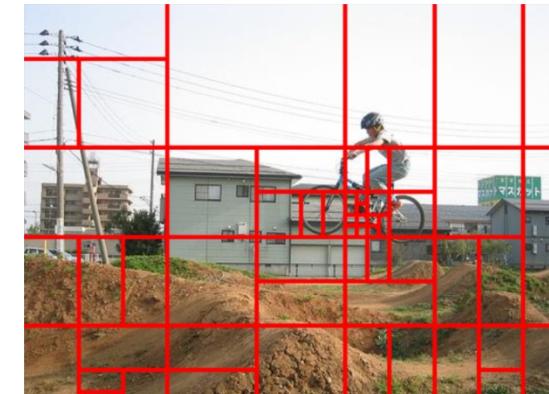
Detector score $S(x)$



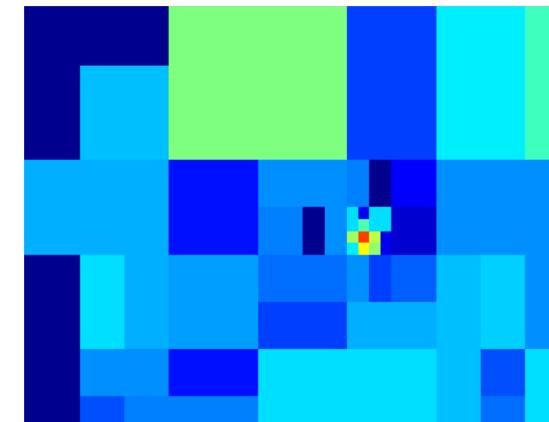
BB for $\arg \max_x S(x)$



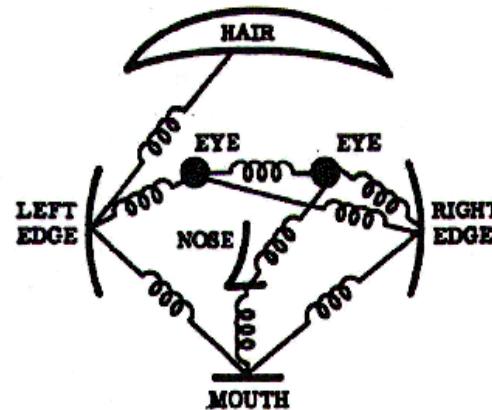
Best-case: $O(P \log N)$



BB for $S(x) \geq -1$



Object Detection with Deformable Part Models



P.F. Felzenszwalb, and D. Huttenlocher, *Pictorial Structures for Object Recognition*, IJCV 2005

P.F. Felzenszwalb, and D. Huttenlocher, *Distance Transforms of Sampled Functions*, 2004

P.F. Felzenszwalb, R. B. Girshick, and D. A. McAllester, *Cascade object detection with DPMs* CVPR 2010

Deformable Part Models (DPMs)

$$\mathbf{S}(\mathbf{x}) = \sum_{p=1}^P U_p(x_p) + B_p(x, x_p)$$

Local appearance

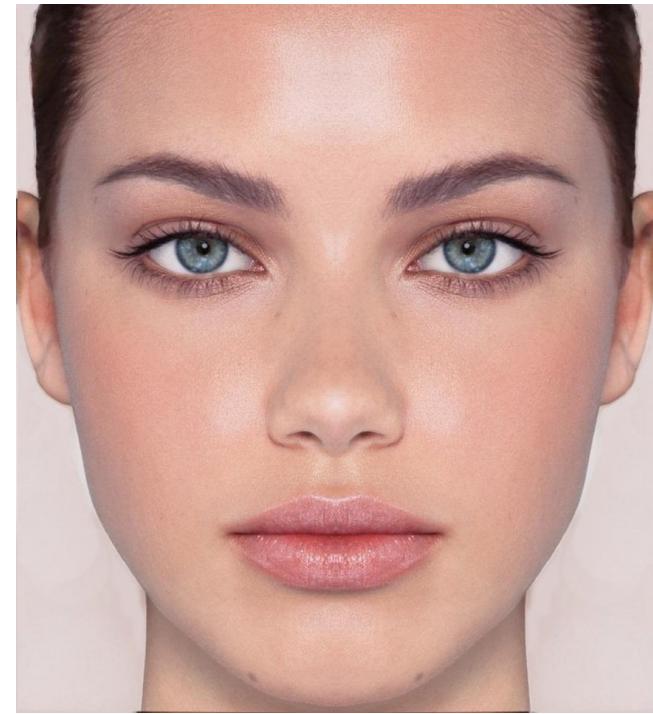
$$U_p(x_p) = \langle w_p, H(x_p) \rangle$$

Pairwise compatibility

$$B_p(x, x_p) = -(h - h_p - \hat{h}_p)^2 \eta - (v - v_p - \hat{v}_p)^2 \nu$$

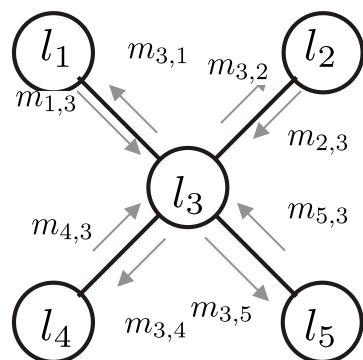
Object score

$$S(x) = \sum_{p=1}^P \max_{x'} [U_p(x') + B_p(x, x')]$$



$$\mu_{i \rightarrow j}(X_j) = \max_{X_i} \Phi_i(X_i) \Psi_{i,j}(X_i, X_j) \prod_{k \in \mathcal{N}(i) \setminus j} \mu_{k \rightarrow j}(X_i)$$

$$B_j(X_j) = \max P(X_j)$$



$$\Phi_1(l)$$



$$\Phi_2(l)$$



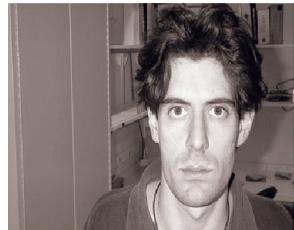
$$\Phi_3(l)$$



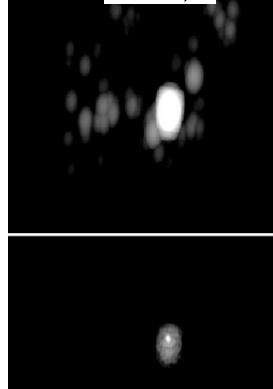
$$\Phi_4(l)$$



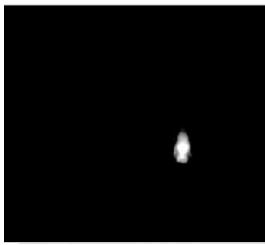
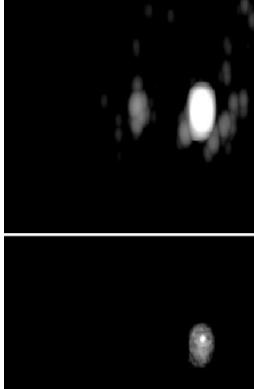
$$\Phi_5(l)$$



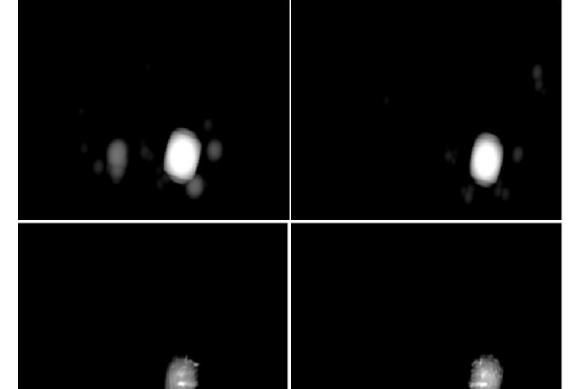
$$m_{1,3}$$



$$m_{2,3}$$



$$m_{3,4}$$



$$B_3(l)$$

$$B_1(l)$$

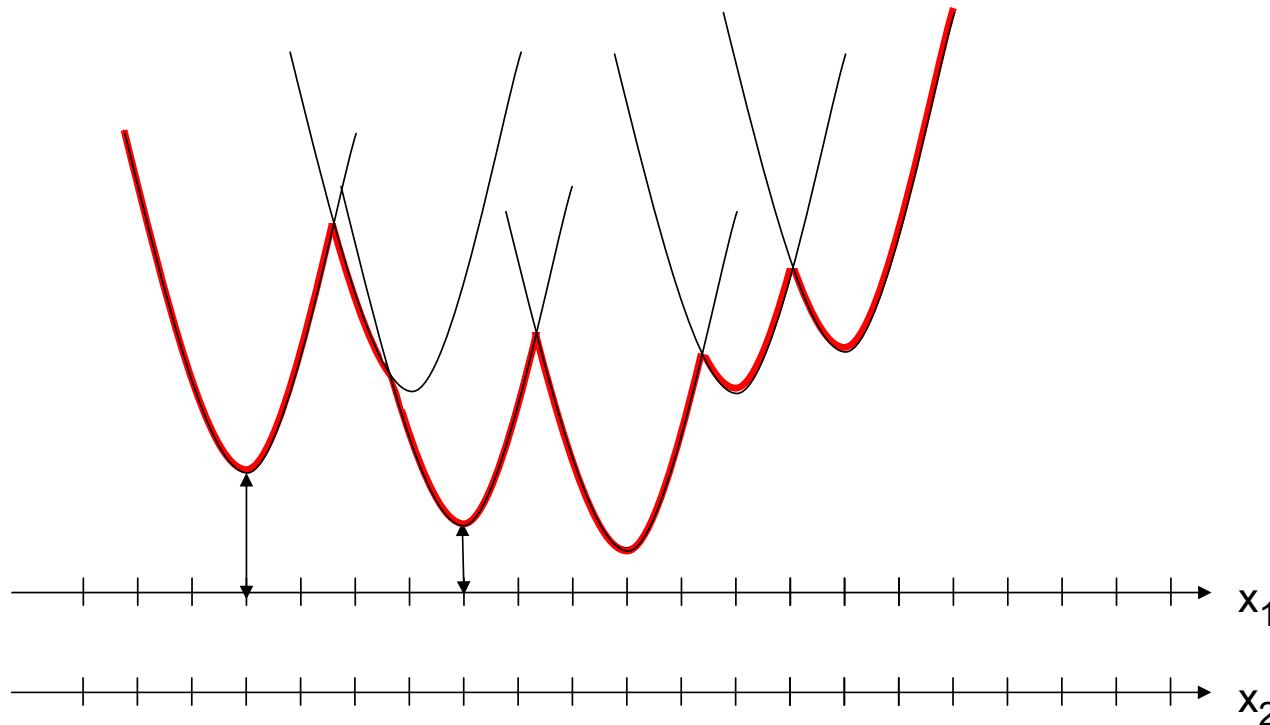
$$B_2(l)$$

$$B_4(l)$$

$$B_5(l)$$

GDT- 1D

Plot $\min_{x_1} \{m_1(x_1) + \phi(x_1, x_2)\}$ as function of x_2

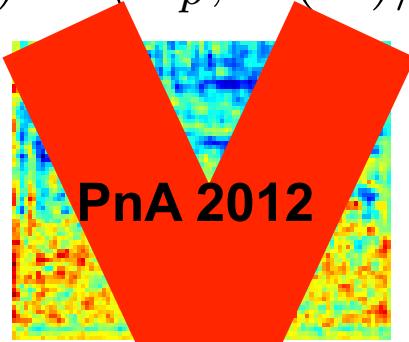


For each x_2

- Finding min over x_1 is equivalent finding minimum over set of offset parabolas
- Lower envelope computed in $O(h)$ rather than $O(h^2)$ via gen. distance transform

Object detection with Deformable Part Models (DPMs)

$$U_p(x') = \langle \mathbf{w}_p, \mathbf{H}(x') \rangle$$



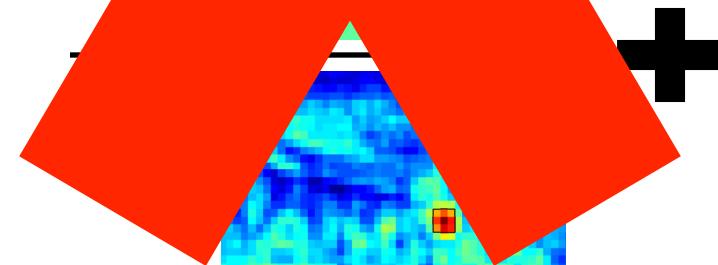
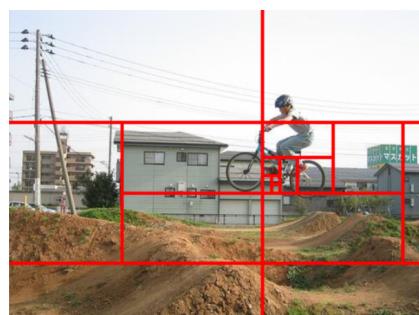
$$\max_{x'} [U_p(x') + B_p(x, x')]$$

$$p = 1$$

⋮

DTBB, NIPS 11

$$p = P$$



$$S(x) = \sum_{p=1}^P \max_{x'} [U_p(x') + B_p(x, x')]$$

Object Detection problem

$$S(x) = \sum_{p=1}^P \max_{x'} [U_p(x') + B_p(x, x')]$$

Threshold-based detection: recover all locations above threshold

$$M^\theta = \{x : S(x) \geq \theta\}$$

1-best detection: recover top-scoring candidate

$$M^* = \{\arg \max_x S(x)\}$$

Naïve implementation (Max-Product) : $O(P|x|^2)$

Generalized Distance Transforms (GDT): $O(P|x|)$

Our work (best-case): $O(|M|P \log_2 |x|)$

Key idea: avoid the exact evaluation of $S(x)$

Previous works on acceleration

Sparse image representations:

Y. Chen, L. Zhu, A. Yuille. Rapid inference on a novel and/or graph for object detection, NIPS 2007

I. Kokkinos and A. Yuille. HOP: Hierarchical Object Parsing, CVPR, 2009

Dense image representations:

B. Sapp, A. Toshev, and B. Taskar. Cascaded models for articulated pose estimation, ECCV, 2010

M. Pedersoli, A. Vedaldi, and J. Gonzalez. A coarse-to-fine approach for object detection, CVPR 2011

P F Felzenszwalb R B Girshick and D A McAllester Cascade object detection with DPMs CVPR 2010

‘Monolithic’ object models:

F. Fleuret and D. Geman. Coarse-to-fine face detection, IJCV 2001

P. Viola and M. Jones. Robust Real-time Object Detection, IJCV 2004

C. Lampert , M. Blaschko and T. Hoffman. Beyond Sliding Windows- ESS, CVPR 2008

T. Breuel. Fast recognition using adaptive subdivisions of transformation space, CVPR 1992

Deformable Part Models (DPMs)

$$\mathbf{S}(\mathbf{x}) = \sum_{p=1}^P U_p(x_p) + B_p(x, x_p)$$

Local appearance

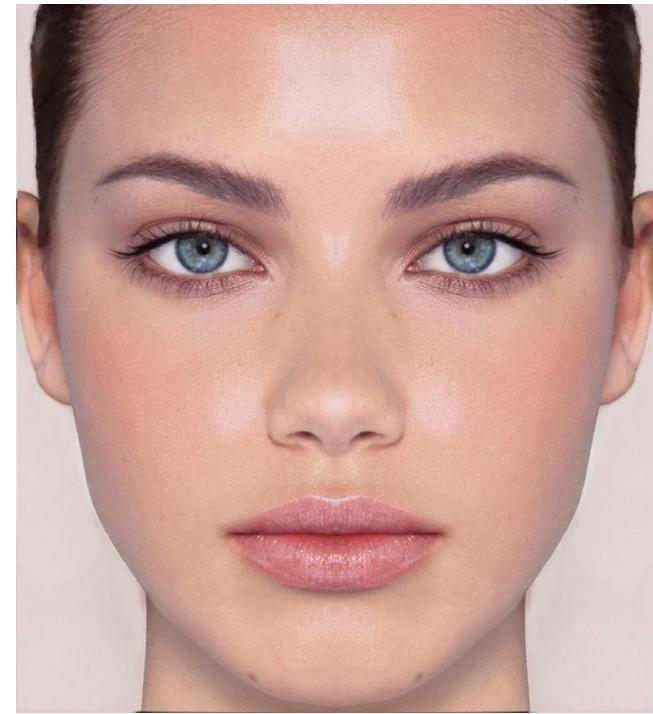
$$U_p(x_p) = \langle w_p, H(x_p) \rangle$$

Pairwise compatibility

$$B_p(x, x_p) = -(h - h_p - \hat{h}_p)^2 \eta - (v - v_p - \hat{v}_p)^2 \nu$$

Object score

$$S(x) = \sum_{p=1}^P \max_{x'} [U_p(x') + B_p(x, x')]$$



Bounding the DPM cost function

Wanted: $\max_{x \in X} S(x) \leq \bar{S}(X)$

$$S(x) = \sum_{p=1}^P m_p(x) \quad m_p(x) = \max_{x'} [U_p(x') + B_p(x, x')]$$

Property: $\max_{x \in X} h(x) + g(x) \leq \max_{x \in X} h(x) + \max_{x \in X} g(x)$

$$\max_{x \in X} \sum_{p=1}^P m_p(x) \leq \sum_{p=1}^P \max_{x \in X} m_p(x)$$



to-do

$$\max_{x \in X} S(x) \leq \sum_{p=1}^P \bar{m}_p(X) \doteq \bar{S}(X)$$

Bounding-based detection for DPMs

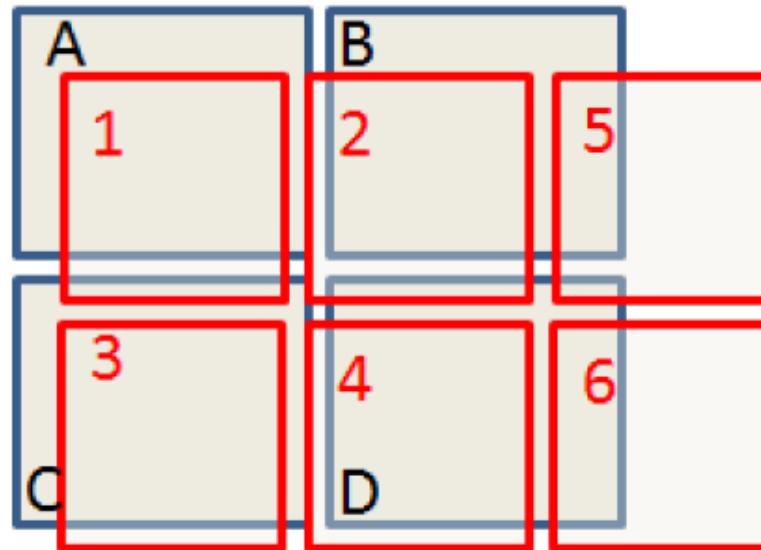
$$\begin{aligned}
 S(x) &= \sum_{p=1}^P \max_{x'} [U_p(x') + B_p(x, x')] \\
 \max_{x \in X} S(x) &= \max_{x \in X} \sum_{p=1}^P \max_{x'} [U_p(x') + B_p(x, x')] \\
 &\leq \sum_{p=1}^P \max_{x \in X} \max_{x'} [U_p(x') + B_p(x, x')] \\
 &\leq \sum_{p=1}^P \left[\max_{x'} U_p(x') + \max_{x \in X} \max_{x'} B_p(x, x') \right]
 \end{aligned}$$

Fine-to-Coarse
all intervals in $O(2N)$ **Analytic: constant-time**

Breaking up the message bound

$$m(X) \doteq \max_{x \in X} \max_{x' \in X'} U(x') + B(x', x)$$

Partition domains: $X = \cup_{d \in D} X_d$, $X' = \cup_{s \in S} X_s$



Define: $\mu_d^s \doteq \max_{x \in X_d} \max_{x' \in X_s} U(x') + B(x', x)$

$$m(X) = \max_d \max_s \mu_d^s$$

Bounding the source-to-domain contributions

$$\mu_d^s \doteq \max_{x \in X_d} \max_{x' \in X_s} U(x') + B(x', x)$$

Upper bound: $\mu_d^s \leq \max_{x' \in X_s} U(x') + \max_{x \in X_d} \max_{x' \in X_s} B(x', x)$

Intuitively: take best source point, put it at best possible source location

Quantities involved: $\bar{u}^s \doteq \max_{x' \in X_s} U(x')$, $\bar{g}_d^s \doteq \max_{x \in X_d} \max_{x' \in X_s} B(x', x)$

**Fine-to-coarse -
O(N)**

**Analytic
expression**

Pruning over combinations of s and d: Dual Trees

A.G. Gray and A.W. Moore. Nonparametric density estimation: Toward computational tractability. ICDM 2003

Analytical expressions for geometric bounds

Pairwise term: $\mathcal{G}_{x,x'} = -H(h - h')^2 - V(v - v')^2$

$$\begin{aligned}\mathcal{G}_{\bar{d},\bar{s}} &\stackrel{\cdot}{=} \max_{x \in X_d} \max_{x' \in X_s} \mathcal{G}_{x,x'} \\ &= \max_{h \in X_d^h} \max_{h' \in X_s^h} -H(h - h')^2 + \max_{v \in X_d^v} \max_{v' \in X_s^v} -H(v - v')^2 \\ &= -Hh_{\underline{d},\underline{s}}^2 - Vv_{\underline{d},\underline{s}}^2, \quad \text{where} \\ h_{\underline{d},\underline{s}} &\stackrel{\cdot}{=} \min_{h \in X_d^h} \min_{h' \in X_s^h} |h - h'|, \quad v_{\underline{d},\underline{s}} \stackrel{\cdot}{=} \min_{v \in X_d^v} \min_{v' \in X_s^v} |v - v'|\end{aligned}$$

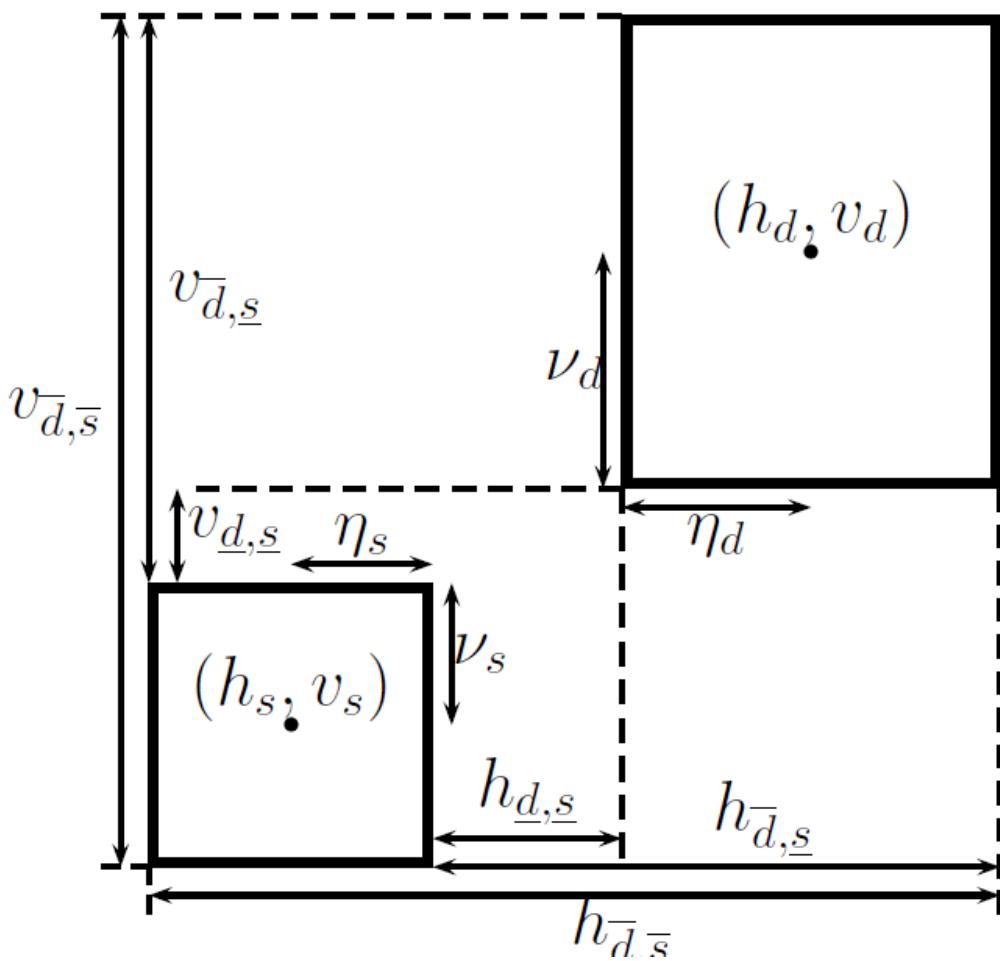
$$\mathcal{G}_{\underline{d},\bar{s}} \stackrel{\cdot}{=} \min_{x \in X_d} \max_{x' \in X_s} \mathcal{G}_{x,x'} = -Hh_{\bar{d},\underline{s}}^2 - Vv_{\bar{d},\underline{s}}^2, \quad \text{where}$$

$$h_{\bar{d},\underline{s}} \stackrel{\cdot}{=} \max_{h \in X_d^h} \min_{h' \in X_s^h} |h - h'|, \quad v_{\bar{d},\underline{s}} \stackrel{\cdot}{=} \max_{v \in X_d^v} \min_{v' \in X_s^v} |v - v'|\quad$$

$$\mathcal{G}_{\underline{d},\bar{s}} \stackrel{\cdot}{=} \min_{x \in X_d} \min_{x' \in X_s} \mathcal{G}_{x,x'} = -Hh_{\underline{d},\bar{s}}^2 - Vv_{\underline{d},\bar{s}}^2, \quad \text{where}$$

$$h_{\bar{d},\bar{s}} \stackrel{\cdot}{=} \max_{h \in X_d^h} \max_{h' \in X_s^h} |h - h'|, \quad v_{\bar{d},\bar{s}} \stackrel{\cdot}{=} \max_{v \in X_d^v} \max_{v' \in X_s^v} |v - v'|\quad$$

Exploit rectangular domain



$$\begin{aligned}
 h_{\bar{d},\bar{s}} &= (h_d + \eta_d) - (h_s + \eta_s) \\
 h_{\bar{d},s} &= (h_d - \eta_d) - (h_s + \eta_s) \\
 h_{d,\underline{s}} &= (h_d + \eta_d) - (h_s - \eta_s) \\
 v_{\bar{d},\bar{s}} &= (v_d + \nu_d) - (v_s + \nu_s) \\
 v_{\bar{d},s} &= (v_d - \nu_d) - (v_s + \nu_s) \\
 v_{d,\underline{s}} &= (v_d + \nu_d) - (v_s - \nu_s)
 \end{aligned}$$

Done so far

$$m(X) \doteq \max_{x \in X} \max_{x' \in X'} U(x') + B(x', x)$$

$$X = \cup_{d \in D} X_d, \quad X' = \cup_{s \in S} X_s$$

$$\mu_d^s \doteq \max_{x \in X_d} \max_{x' \in X_s} U(x') + B(x', x)$$

$$m(X) = \max_d \max_s \mu_d^s$$

$$\overline{m}(X) = \max_d \max_s \overline{\mu}_d^s$$

Large intervals: loose bounds

Small intervals: tight bounds, but maximization over multiple terms

$$|S| = \frac{|X|}{|X_s|}$$

Need to prune

Dual Trees

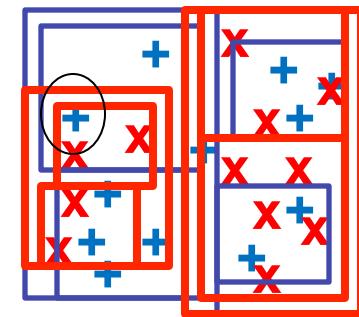
Data structure developed for Kernel Density estimation

$$P(x_j) = \sum_{i=1}^N w_i K(x_j, x_i), \quad x_i \in X_S, \quad x_j \in X_D \quad j = 1, \dots, M$$

Naïve solution: $O(|X_S||X_D|)$

Smarter: KD-tree for source points

Smartest: do this in batch mode for domain points



A.G. Gray and A.W. Moore. Nonparametric density estimation: Toward computational tractability. ICDM 2003

A. T. Ihler, E. B. Sudderth, W. T. Freeman, and A. S. Willsky. Efficient multiscale sampling from products of gaussian mixtures (NBP). In *NIPS*, 2003.

Dual Recursion

Source-/ Domain-tree: KD-trees for source/domain points, respectively.

Goal: keep manageable the computation of $\overline{m}(X) = \max_d \max_s \overline{\mu}_d^s$

Key idea: entertain list of s that can possibly contribute to any d

$$\mathcal{S}_d = \{s_i\}$$

Originally: d is Domain-tree root, its supporter is Source-tree root

$$\mathcal{S}_0 = \{s_0\}$$

When splitting (branching) d, split also its supporters

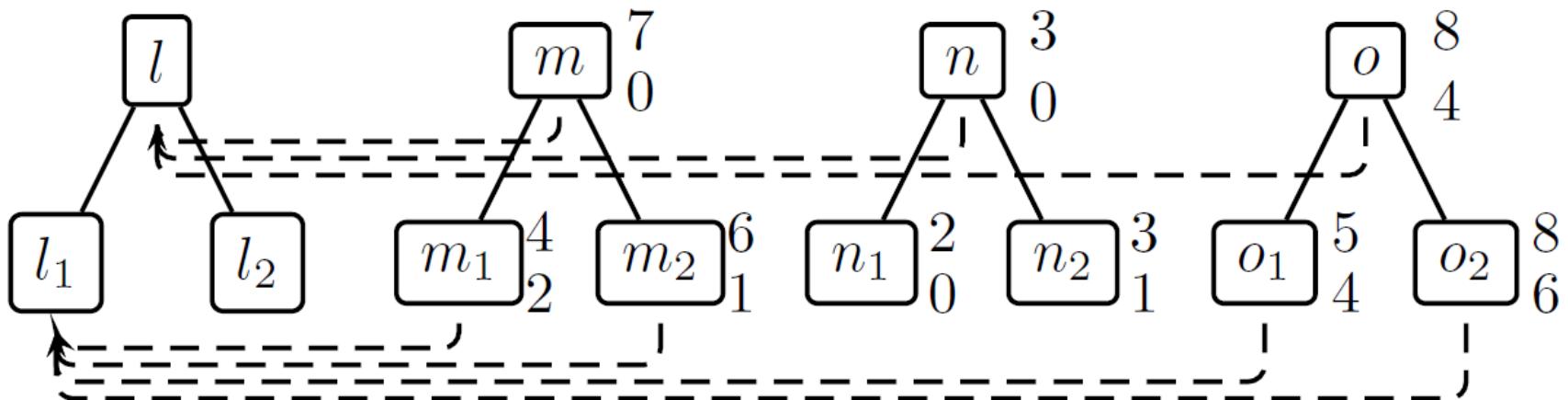
Prune supporters based on upper and lower bounds

Supporter pruning

$\mathcal{S}_l = \{m, n, o\}$ Source nodes m; n; o ‘support’ domain node l

Recurse: $\mathcal{S}_{ch(l)} = \text{Prune} [ch(m) \cup ch(n) \cup ch(o)]$

Pruning criterion: $\bar{\mu}_d^l < \max_{j \in \mathcal{S}_d} \underline{\lambda}_d^j$



DTBB demonstration



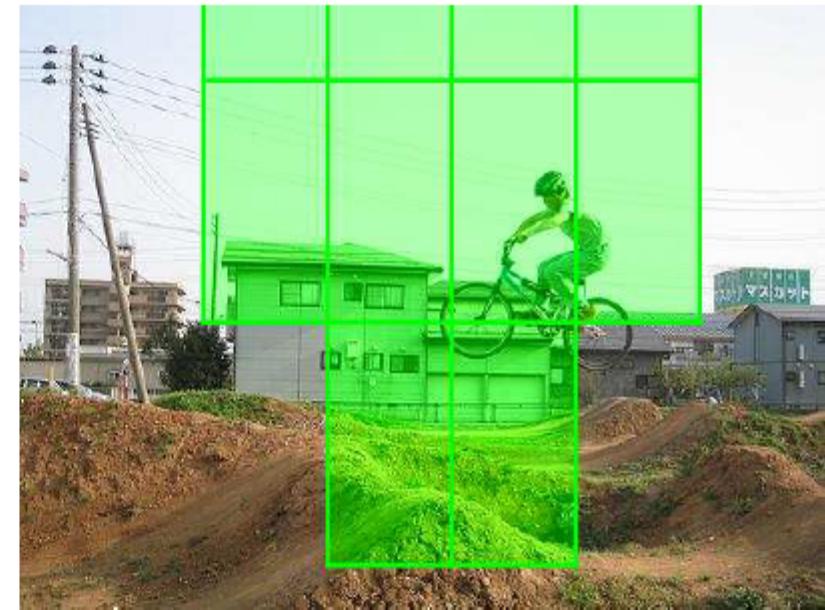
DTBB demonstration



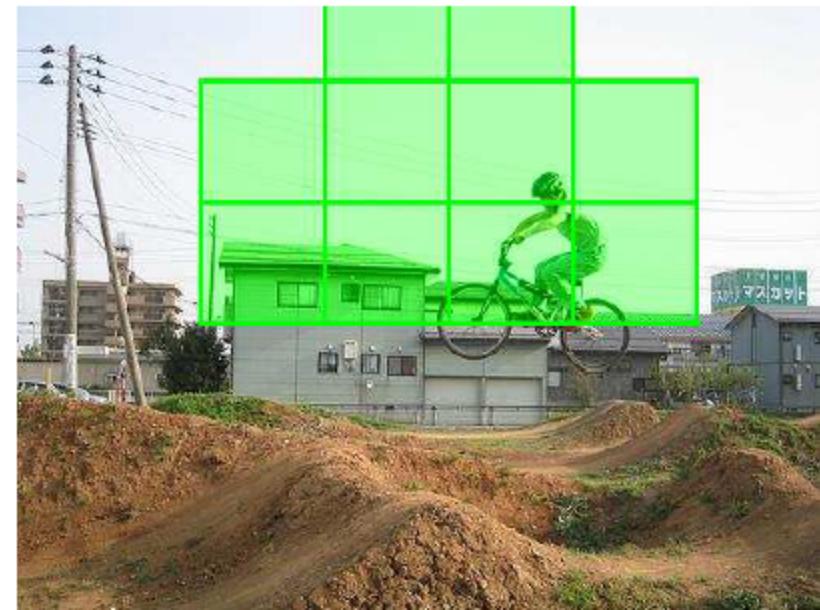
DTBB demonstration



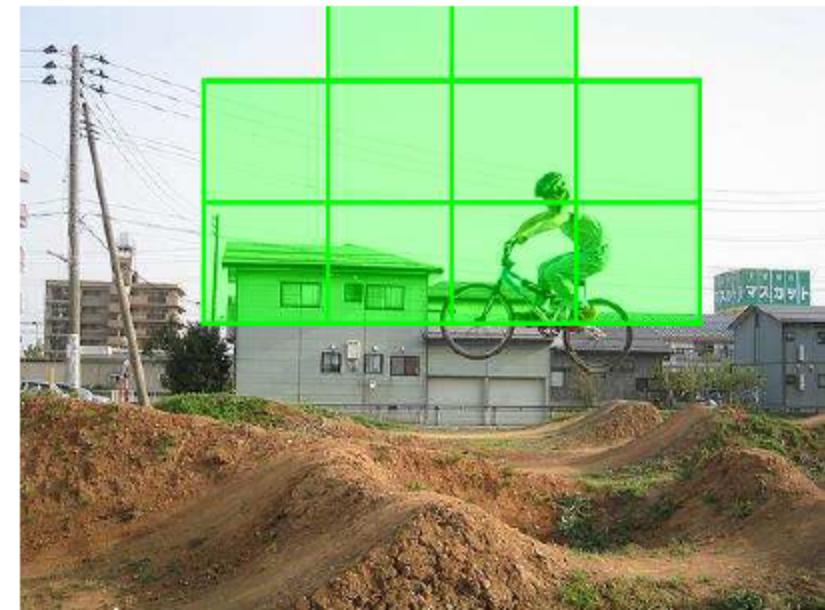
DTBB demonstration



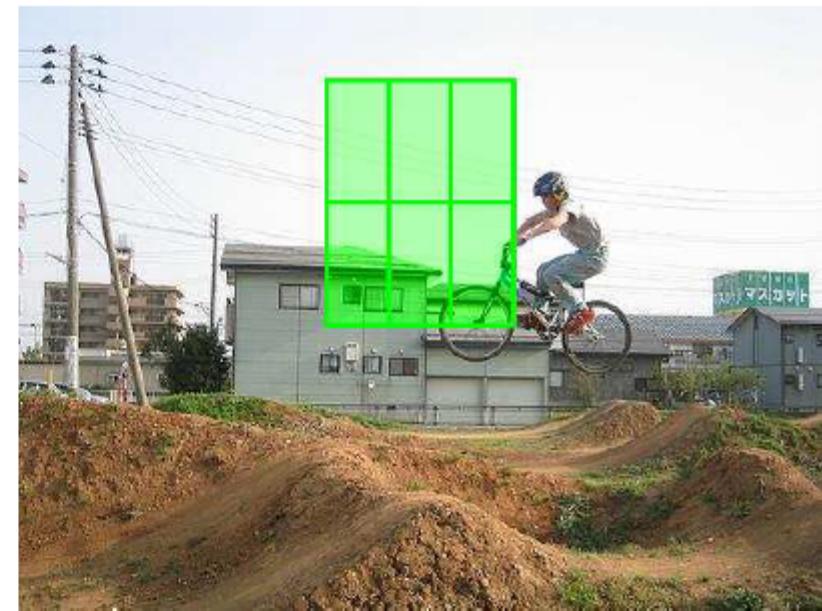
DTBB demonstration



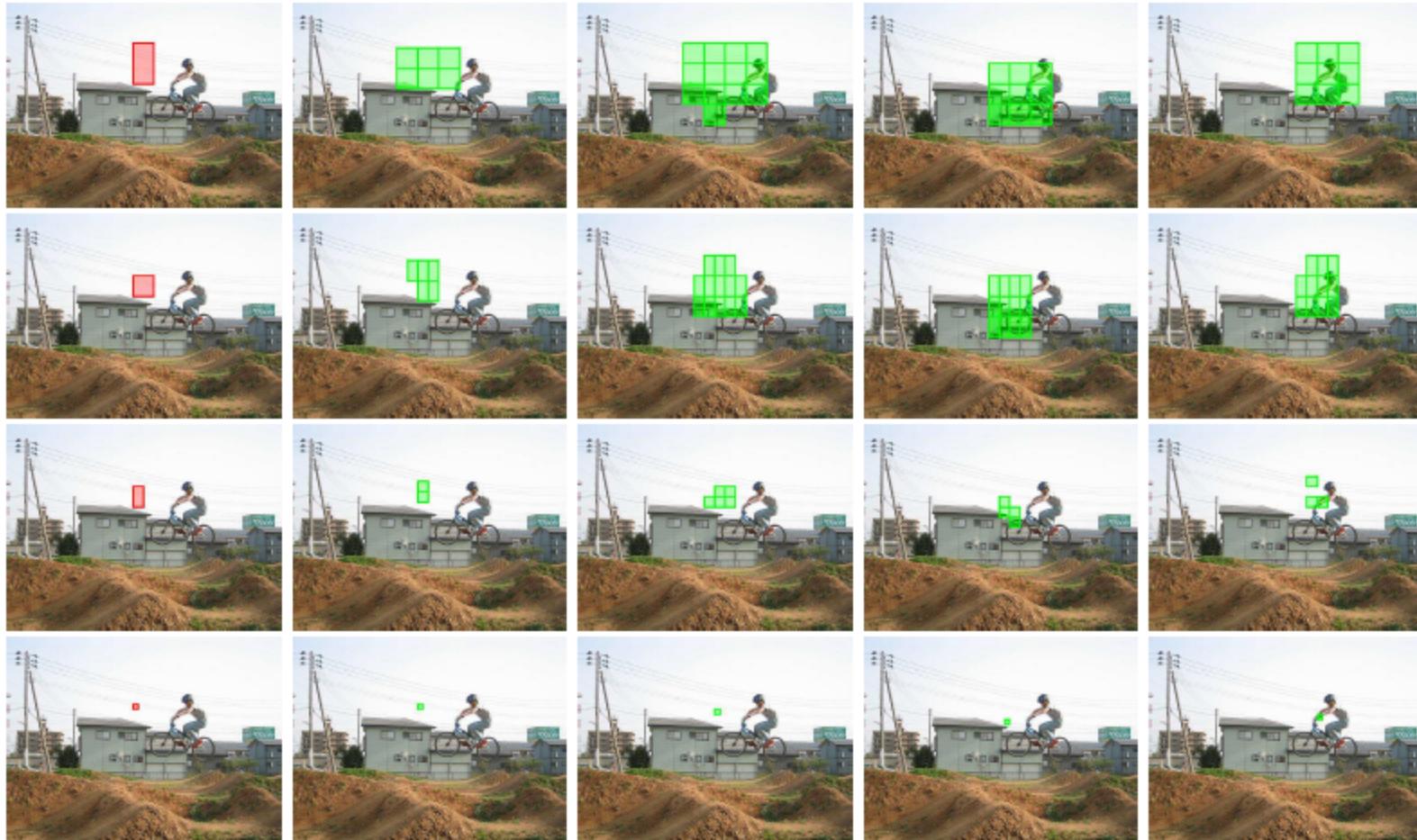
DTBB demonstration



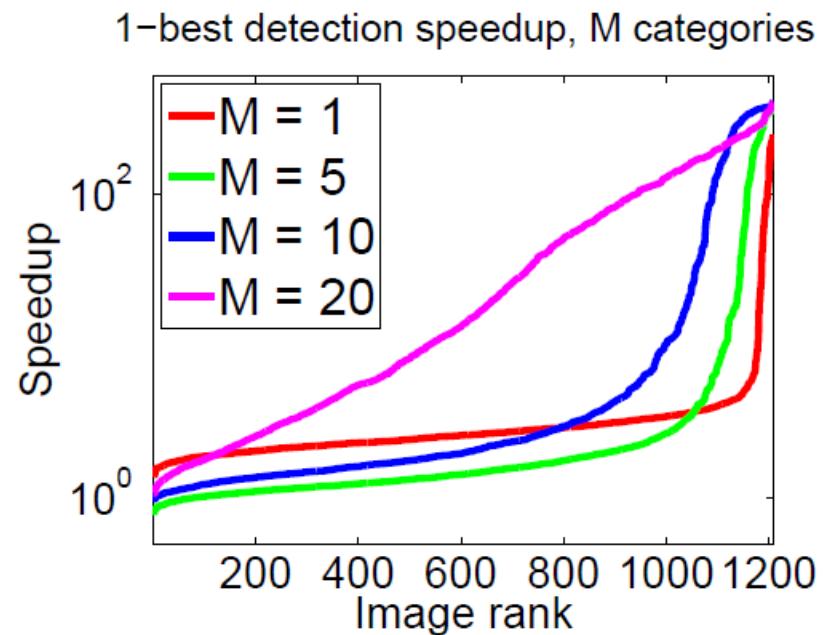
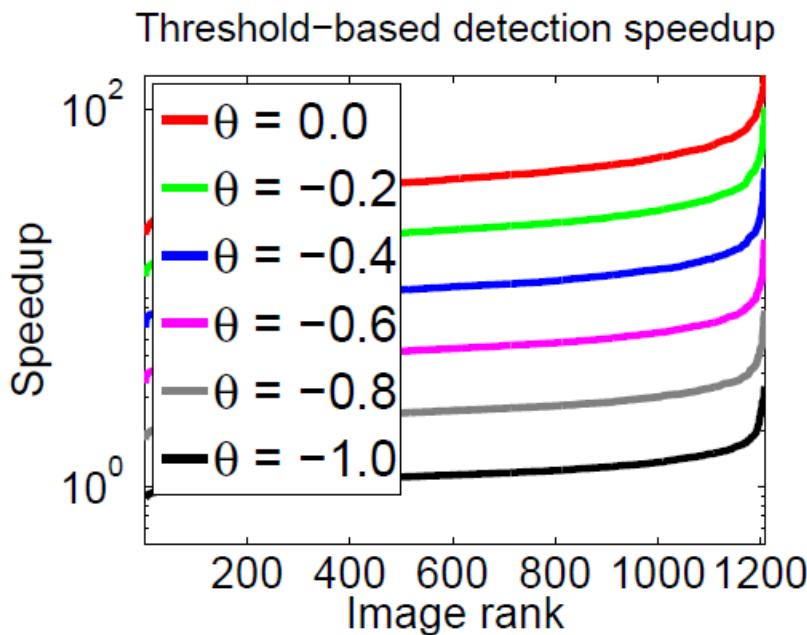
DTBB demonstration



Supporter pruning demonstration



Results on Pascal VOC



1200 images, 20 Categories per image

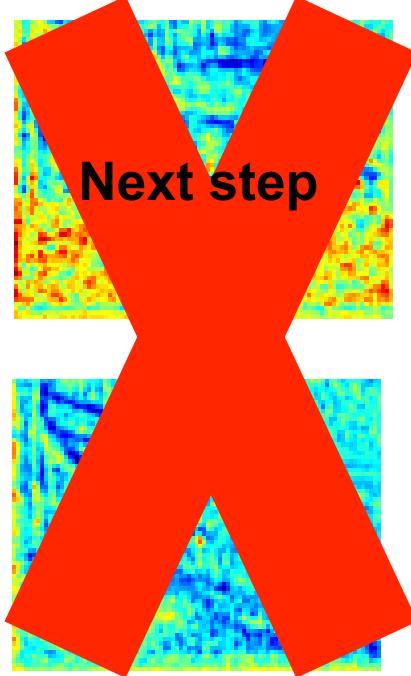
Results on Pascal VOC

	Our algorithm	[4]
Unary terms	13.20 ± 1.49	159.41 ± 15.82
KD-trees	1.72 ± 0.21	0.00 ± 0.00
Detection, $\theta = 0.0$	0.25 ± 0.07	10.74 ± 1.02
Detection, $\theta = -.2$	0.47 ± 0.12	10.74 ± 1.02
Detection, $\theta = -.4$	0.93 ± 0.22	10.74 ± 1.02
Detection, $\theta = -.6$	1.95 ± 0.42	10.74 ± 1.02
Detection, $\theta = -.8$	4.17 ± 0.84	10.74 ± 1.02
Detection, $\theta = -1$	9.14 ± 1.79	10.74 ± 1.02
Detection, 1-best	0.41 ± 0.08	10.74 ± 1.02
Detection, 5-best	0.47 ± 0.09	10.74 ± 1.02
Detection, 10-best	0.48 ± 0.10	10.74 ± 1.02

1200 images, 20 Categories per image

Object detection with Deformable Part Models (DPMs)

$$U_p(x') = \langle \mathbf{w}_p, \mathbf{H}(x') \rangle$$



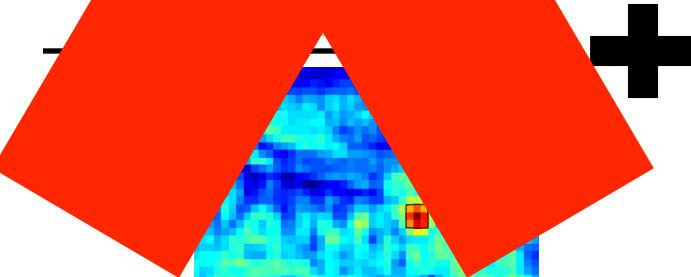
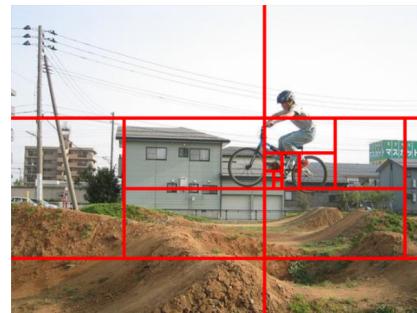
$$\max_{x'} [U_p(x') + B_p(x, x')]$$

$$p = 1$$

⋮

$$p = P$$

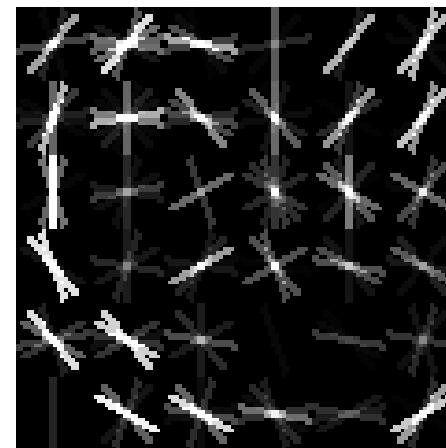
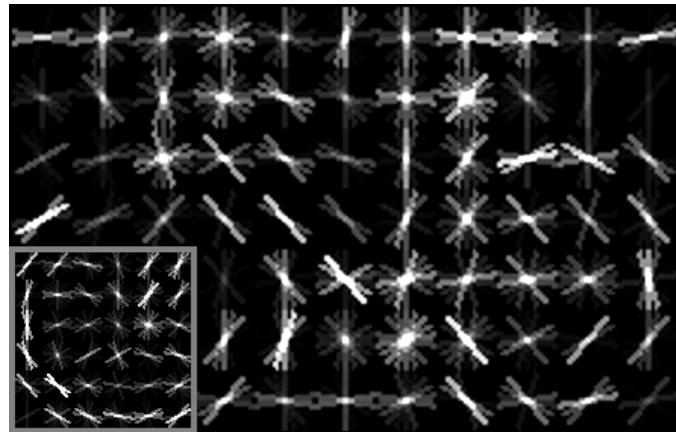
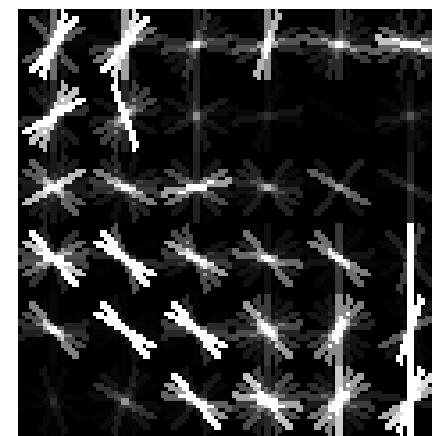
DTBB, NIPS 11



$$S(x) = \sum_{p=1}^P \max_{x'} [U_p(x') + B_p(x, x')]$$

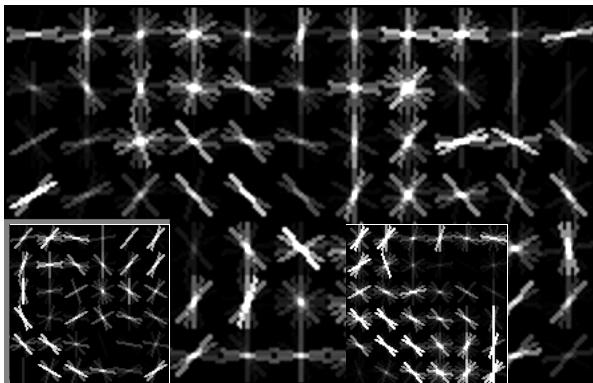
- I. Kokkinos, Bounding part scores for rapid detection, 2012
- I. Kokkinos, Shufflets, ICCV 2013

Part score computation

 $\mathbf{w}[y]$  $\mathbf{h}[x + y]$

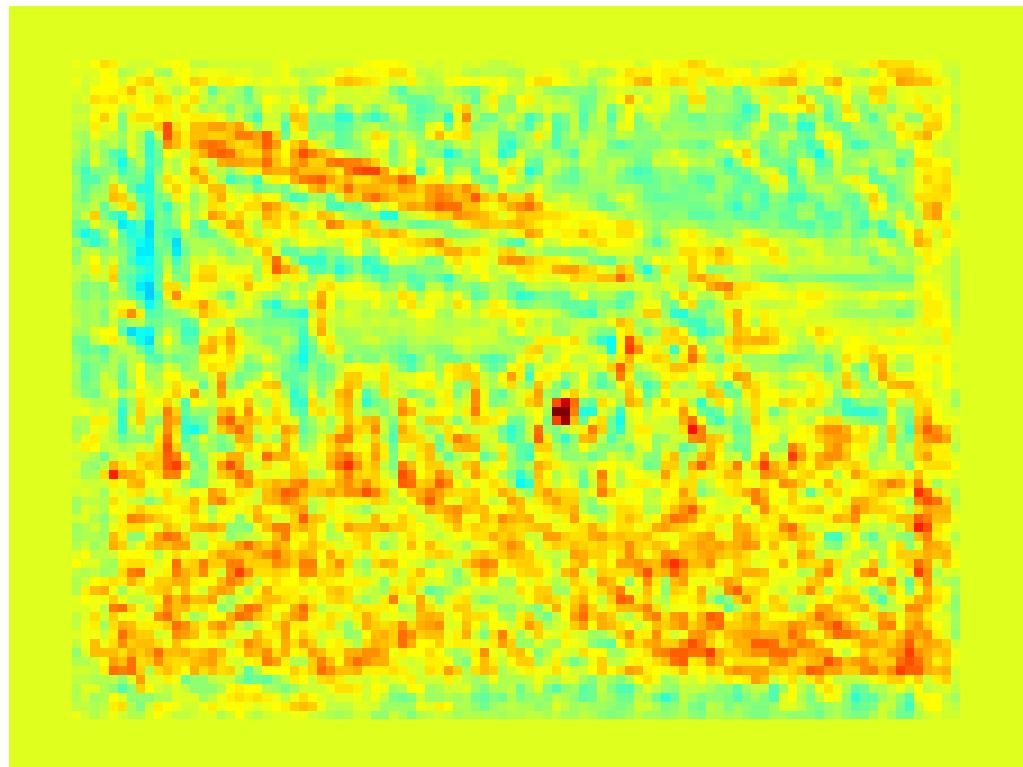
$$s[x] = \sum_y \langle \mathbf{h}[x + y], \mathbf{w}[y] \rangle$$

Part score computation



$$\mathbf{w}_p \quad \mathbf{h}_x \\ U_p(x) = \langle \boxed{\mathbf{w}_p}, \boxed{\mathbf{h}_x} \rangle$$

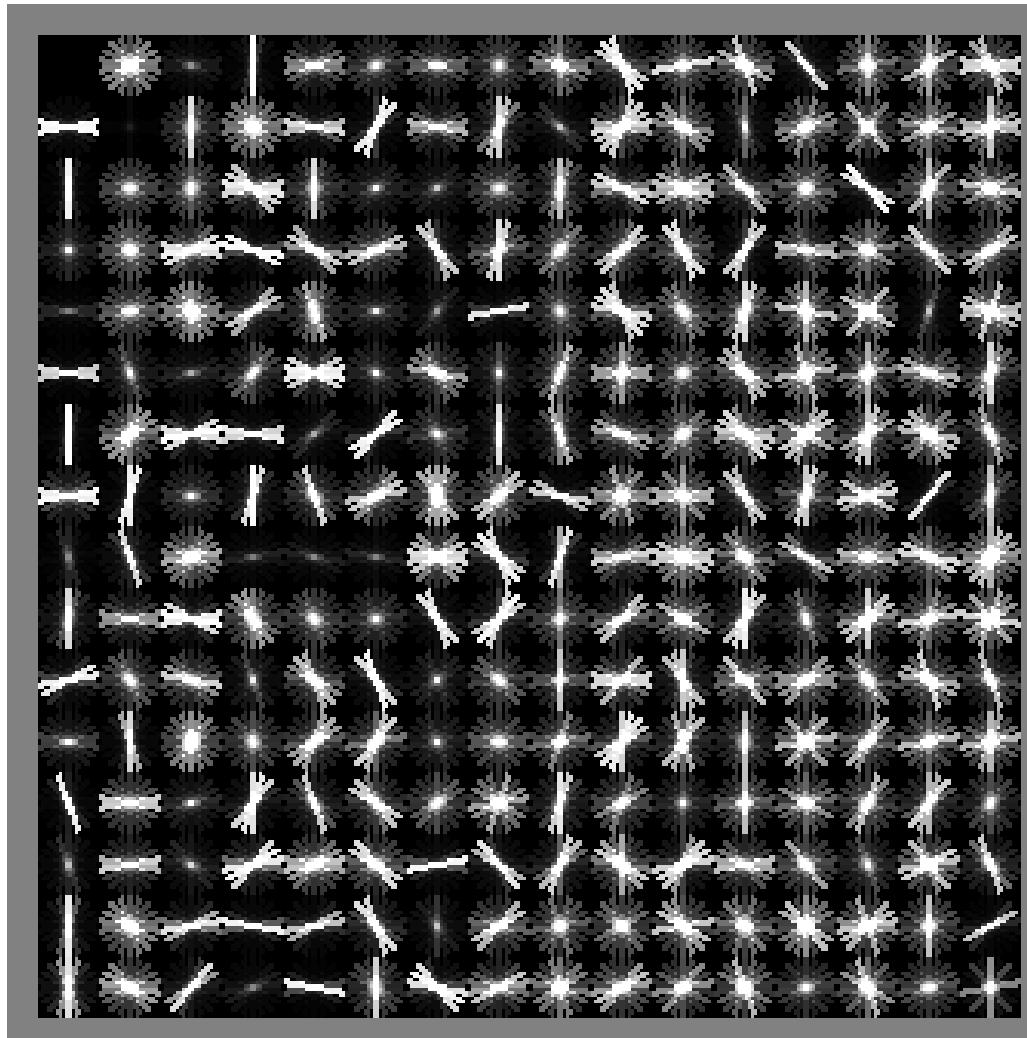
$$\mathbf{w}_p, \mathbf{h}_x \in R^D \quad D = 6 \times 6 \times 32$$



- A. Vedaldi and A. Zisserman, Sparse Kernel Maps and Faster Product Quantization Learning, CVPR 2012
- I. Kokkinos, Bounding part scores for rapid object detection, PnA workshop, ECCV 2012
- T. Dean, M. Ruzon, M. Segal, J. Shlens, S. Vijayanarasimhan, J. Yagnik, 'Fast, Accurate Detection of 100,000 Object Classes on a Single Machine' CVPR 2013
- H. Pirsiavash and D. Ramanan, Steerable Part Models, CVPR 2012
- H.O. Song, S. Zickler, T. Althoff, R. Girschick, M. Fritz, C. Geyer, P. Felzenszwalb, T. Darrell, Sparselet Models for Efficient Multiclass Object Detection, ECCV 2012
- I. Kokkinos, Shufflets, ICCV 2013

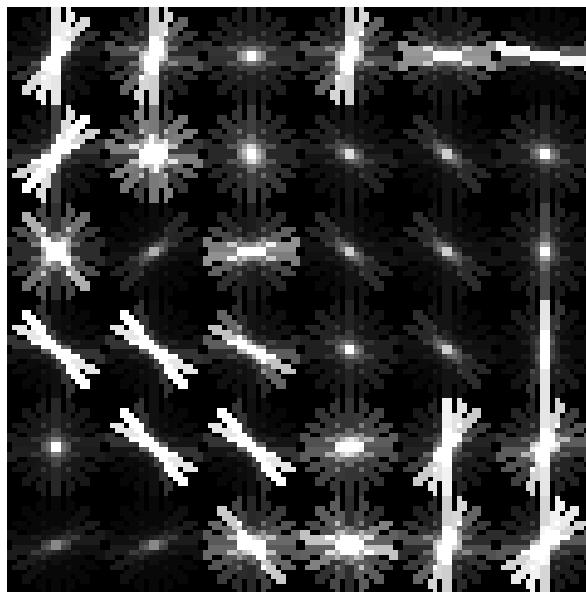
HOG cell quantization: visual ‘letters’

$$\mathcal{C} = \{C_1, \dots, C_{256}\}$$



HOG feature quantization

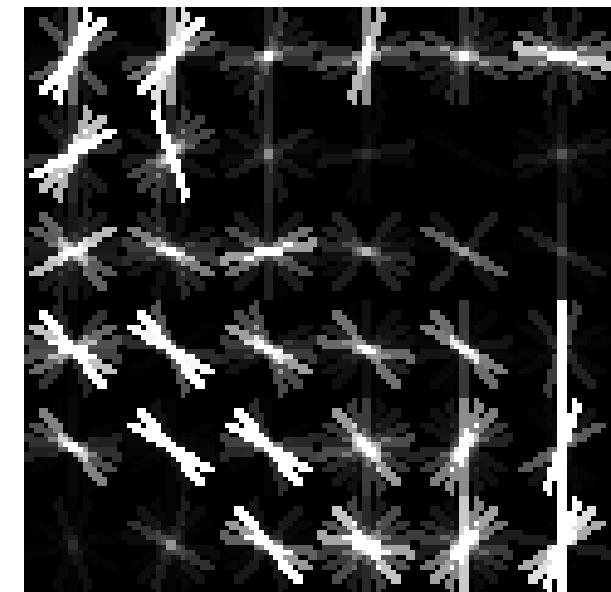
HOG detail



Codebook indices

60	199	39	199	25	62
121	143	132	45	129	85
209	70	64	129	129	117
210	210	200	85	129	118
3	210	210	20	185	115
63	63	186	242	199	155

Quantized HOG



$$\mathbf{h}[x]$$

$$i[x] = \arg \min_k d(\mathbf{h}[x], C_k)$$

$$\hat{\mathbf{h}}[x] = C_{i[x]}$$

Efficient inner product approximation

$$\left\langle \begin{array}{c} \text{[Image]} \\ , \end{array} \right\rangle \approx \left\langle \begin{array}{c} \text{[Image]} \\ , \end{array} \right\rangle$$

$$s[x] \simeq \hat{s}[x]$$

$$\sum_y \langle \mathbf{h}[x+y], \mathbf{w}[y] \rangle \simeq \sum_y \langle \hat{\mathbf{h}}[x+y], \mathbf{w}[y] \rangle$$

Efficient inner product approximation

$$\langle \begin{matrix} \text{[Image]} \\ , \end{matrix} \begin{matrix} \text{[Image]} \\ , \end{matrix} \rangle \simeq \langle \begin{matrix} \text{[Image]} \\ , \end{matrix} \begin{matrix} \text{[Image]} \\ , \end{matrix} \rangle$$

$$\langle \mathbf{h}[x + y], \mathbf{w}[y] \rangle \simeq \langle \hat{\mathbf{h}}[x + y], \mathbf{w}[y] \rangle$$

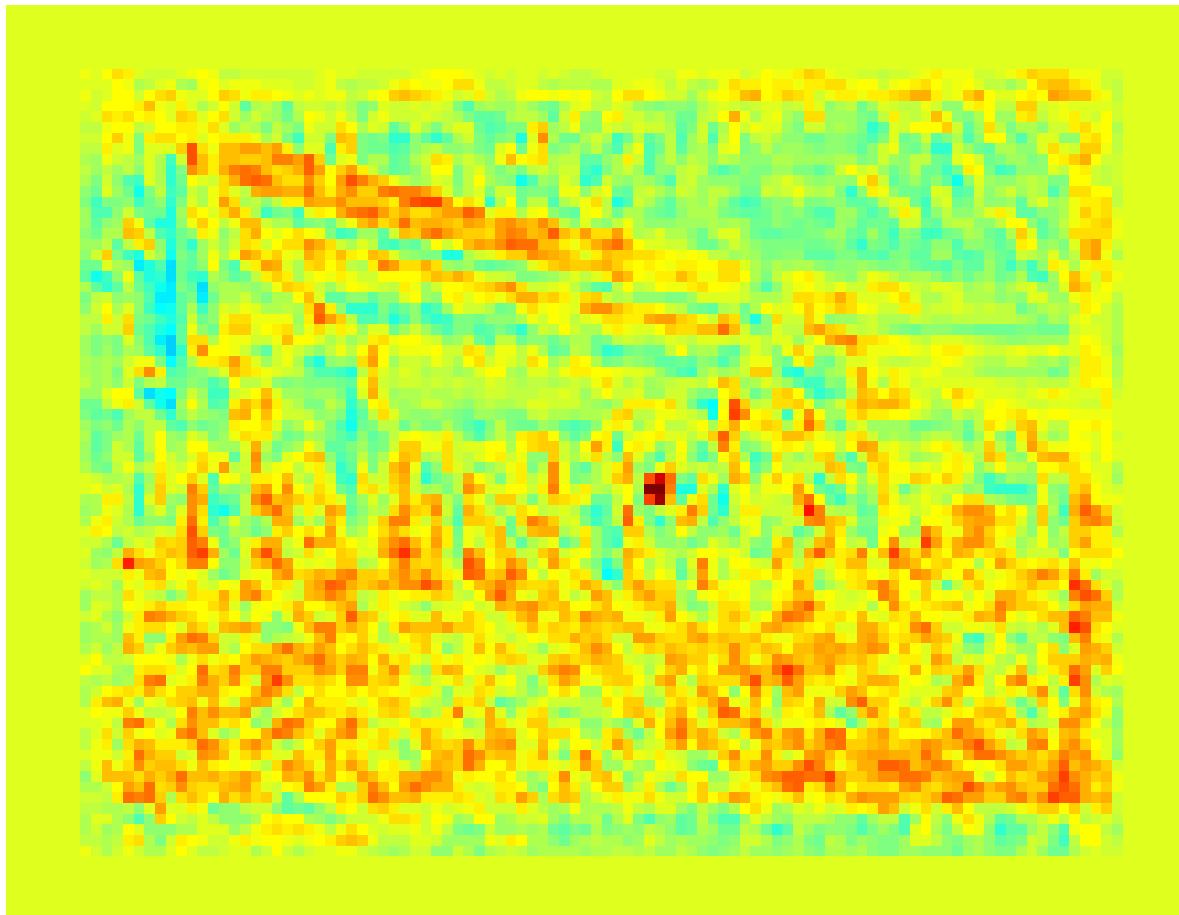
$$= \langle C_{I[x+y]}, \mathbf{w}[y] \rangle$$

$$= \Pi[I[x+y], y]$$

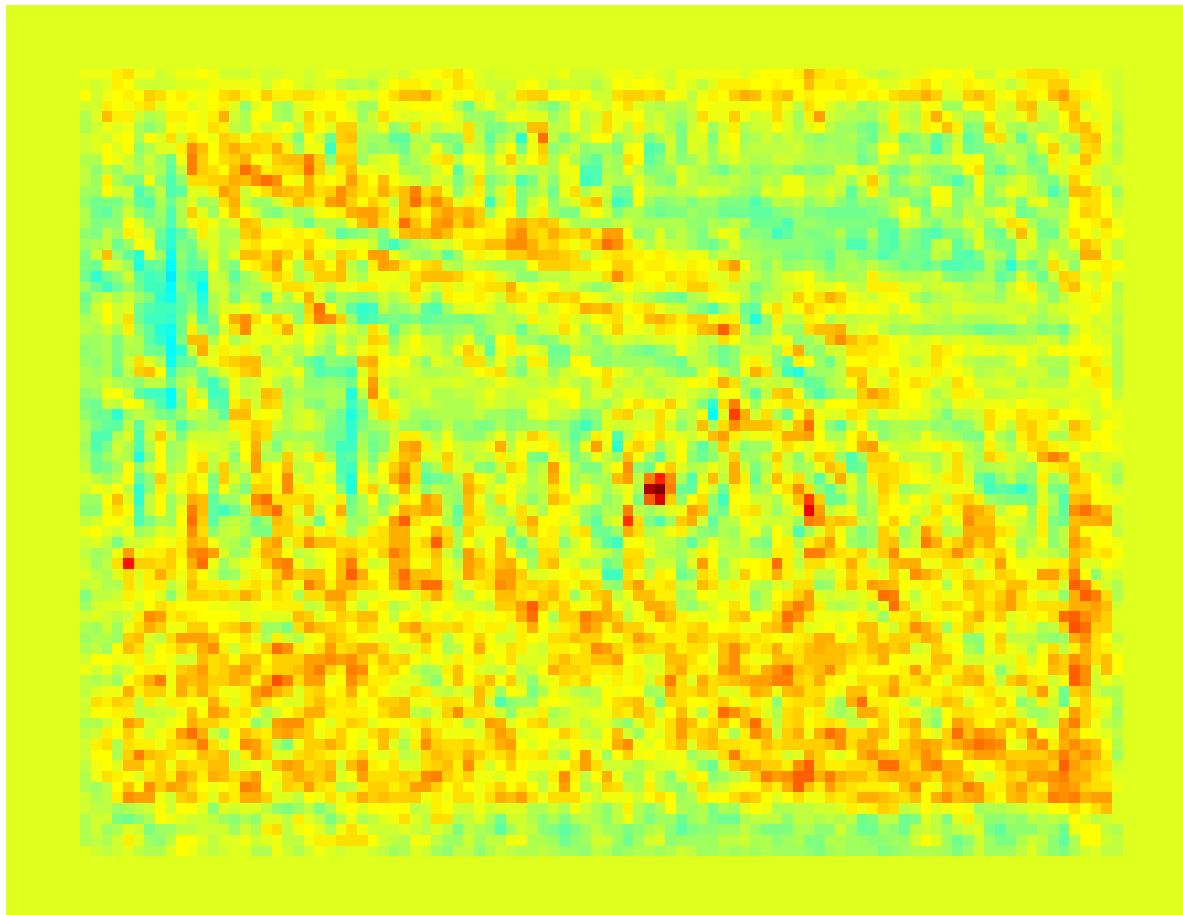
$$\Pi[k, y] = \langle C_k, \mathbf{w}_y \rangle$$

$$s[x] \simeq \hat{s}[x] = \sum_y \Pi[I[x+y], y]$$

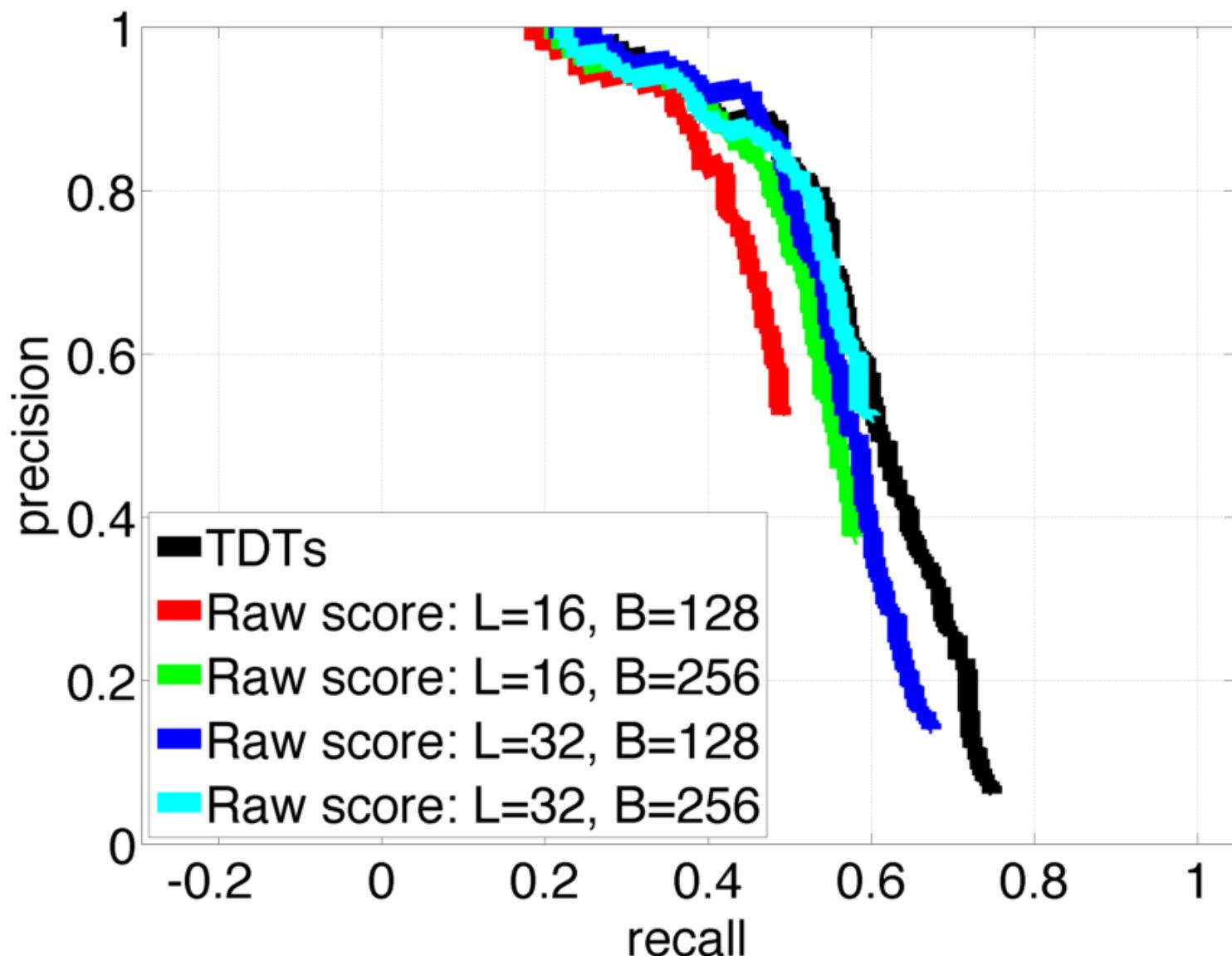
Lookup-based estimate demonstration: $s[x]$



Lookup-based estimate demonstration: $\hat{s}[x]$



Impact of approximation on detection performance



Part-level approximation error

$$\begin{aligned}\epsilon \stackrel{\bullet}{=} s - \hat{s} &= \left\langle \begin{array}{c} \text{[Image of a person with many small arrows]} \\ \text{[Image of a person with many small arrows]} \\ \text{[Image of a person with many small arrows]} \end{array} - \right. \\ &\quad \left. \left\langle \begin{array}{c} \text{[Image of a person with many small arrows]} \\ \text{[Image of a person with many small arrows]} \\ \text{[Image of a person with many small arrows]} \end{array} , \begin{array}{c} \text{[Image of a person with many small arrows]} \\ \text{[Image of a person with many small arrows]} \\ \text{[Image of a person with many small arrows]} \end{array} \right\rangle \right\rangle \\ &= \sum_y \underbrace{\langle \mathbf{h}[y] - \hat{\mathbf{h}}[y], \mathbf{w}[y] \rangle}_{e[y]}\end{aligned}$$

Cell-level approximation error

$$\begin{aligned} e[y] &= \langle \mathbf{h}[y] - \hat{\mathbf{h}}[y], \mathbf{w}[y] \rangle = \langle \text{[image]} - \text{[image]}, \text{[image]} \rangle \\ &= \langle \mathbf{e}[y], \mathbf{w}[y] \rangle \\ &= \sum_{f=1}^{32} \mathbf{e}_y[f] \mathbf{w}_y[f] \end{aligned}$$

How to bound this error?

Markov's inequality

$$P(x > a) = \int_{x>a} p(x)dx$$

$$[x > a] = \begin{cases} 0 & x \leq a \\ 1 & x > a \end{cases} \quad [x > a] \leq \frac{x}{a} \quad \forall x$$

$$P(x > a) = \int_x [x > a]p(x)dx \leq \int_x \frac{x}{a}p(x)dt = \frac{E\{x\}}{a}$$

$$P(x > a) \leq \frac{E\{x\}}{a}$$

Chebyshev's inequality

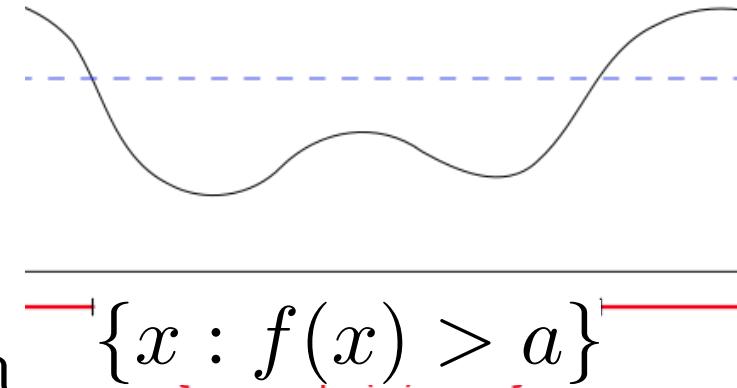
Markov: $P(x > a) \leq \frac{E\{x\}}{a}$

$x' = f(x)$: $P(f(x) > a) = \frac{E\{f(x)\}}{a}$

$x' = x^2$: $P(x^2 > a) = \frac{E\{x^2\}}{a}$

$$P(x^2 > a^2) = \frac{E\{x^2\}}{a^2}$$

$$P(|x| > a) = \frac{E\{x^2\}}{a^2}$$



Chebyshev inequality

For any zero-mean random variable, and any value of α :

$$P(|X| > \alpha) \leq \frac{E\{X^2\}}{\alpha^2}$$

Equivalently, with probability of error smaller than p_e :

$$X \in \left[-\sqrt{\frac{E\{X^2\}}{p_e}}, \sqrt{\frac{E\{X^2\}}{p_e}} \right]$$

Chebyshev-based bounds

Lookup-based approximation:

$$s[x] \simeq \hat{s}[x] = \sum_y \Pi[I[x+y], y]$$

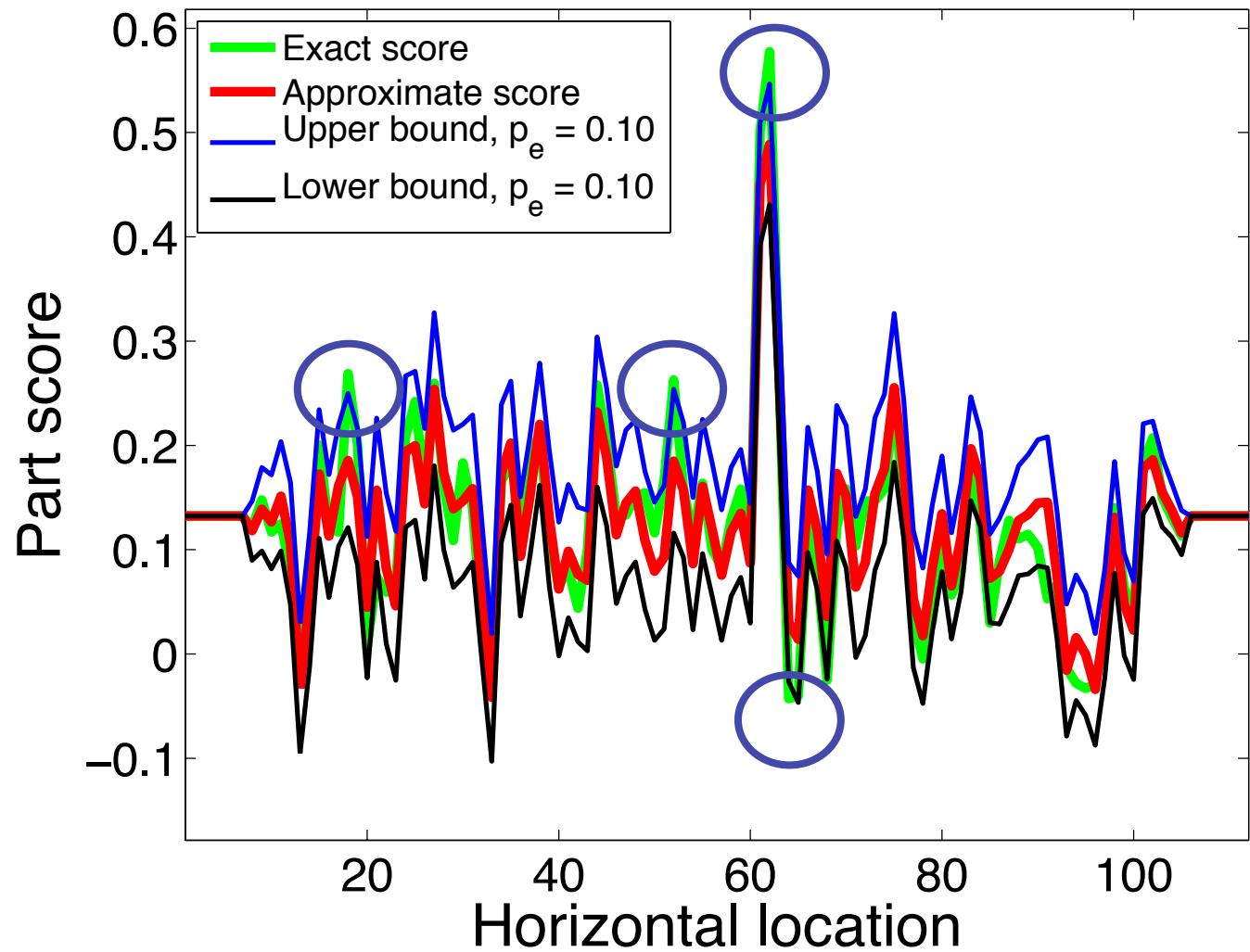
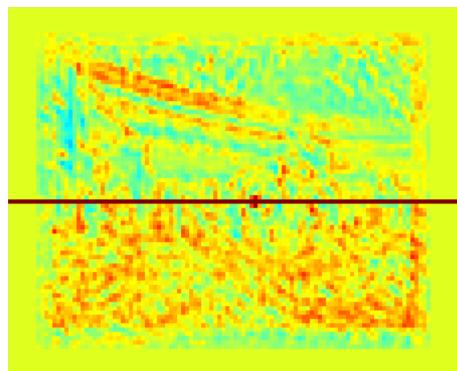
With probability of error at most p_e :

$$\underline{s}[x] \leq s[x] \leq \bar{s}[x]$$

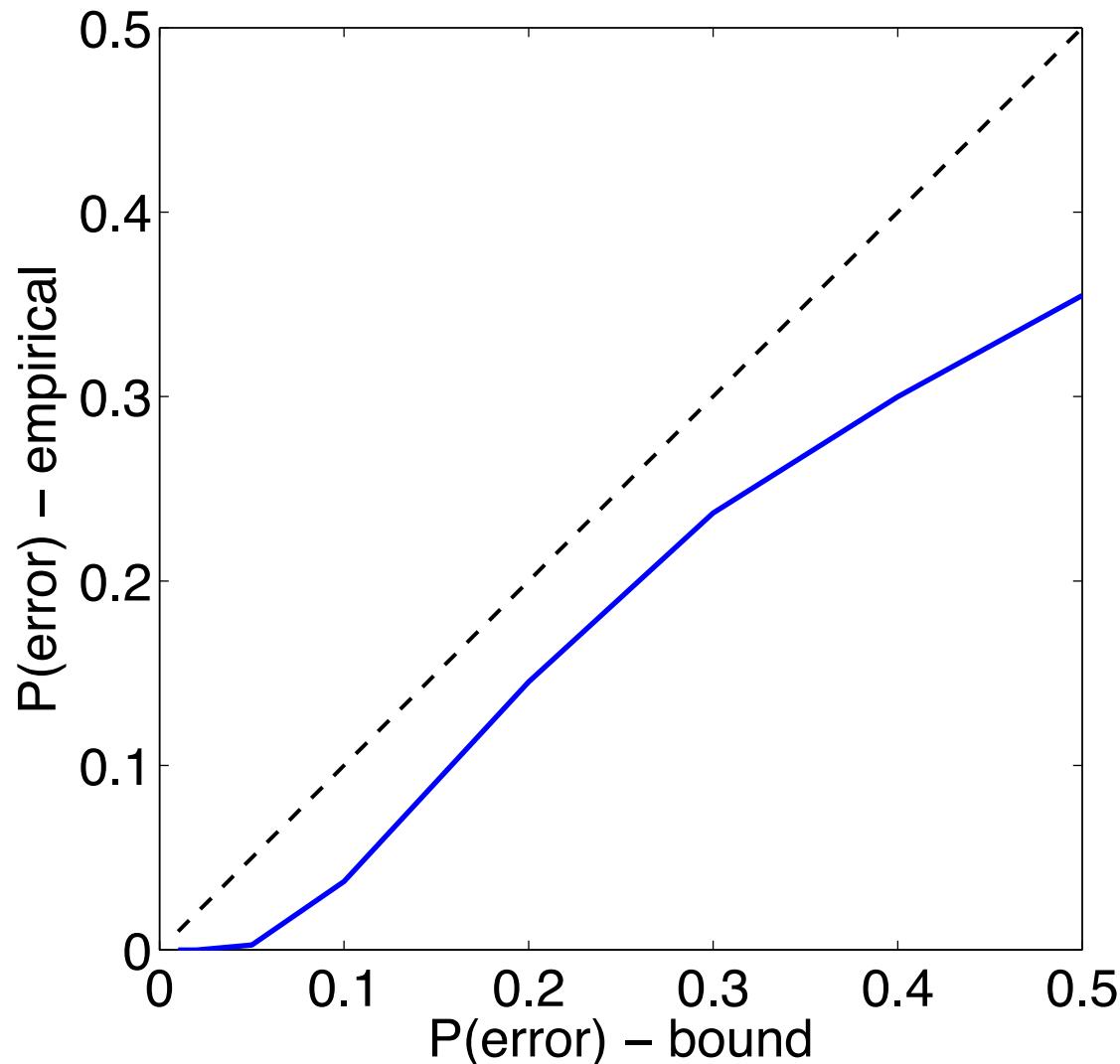
$$\underline{s}[x] = \hat{s}[x] - \sqrt{\frac{\sum_y \|\mathbf{w}[y]\|^2 \|\mathbf{e}[x+y]\|^2}{p_e F}}$$

$$\bar{s}[x] = \hat{s}[x] + \sqrt{\frac{\sum_y \|\mathbf{w}[y]\|^2 \|\mathbf{e}[x+y]\|^2}{p_e F}}$$

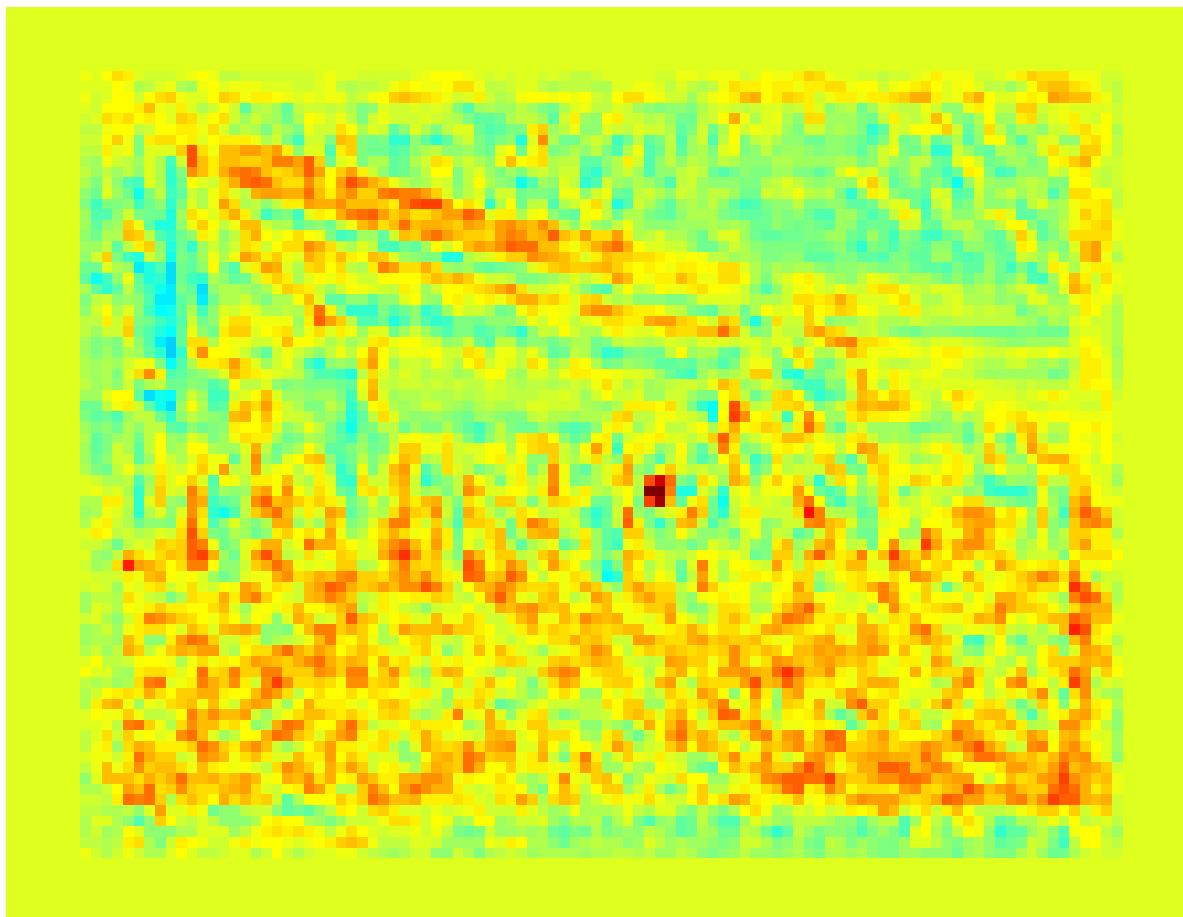
Bound demonstration for varying confidence



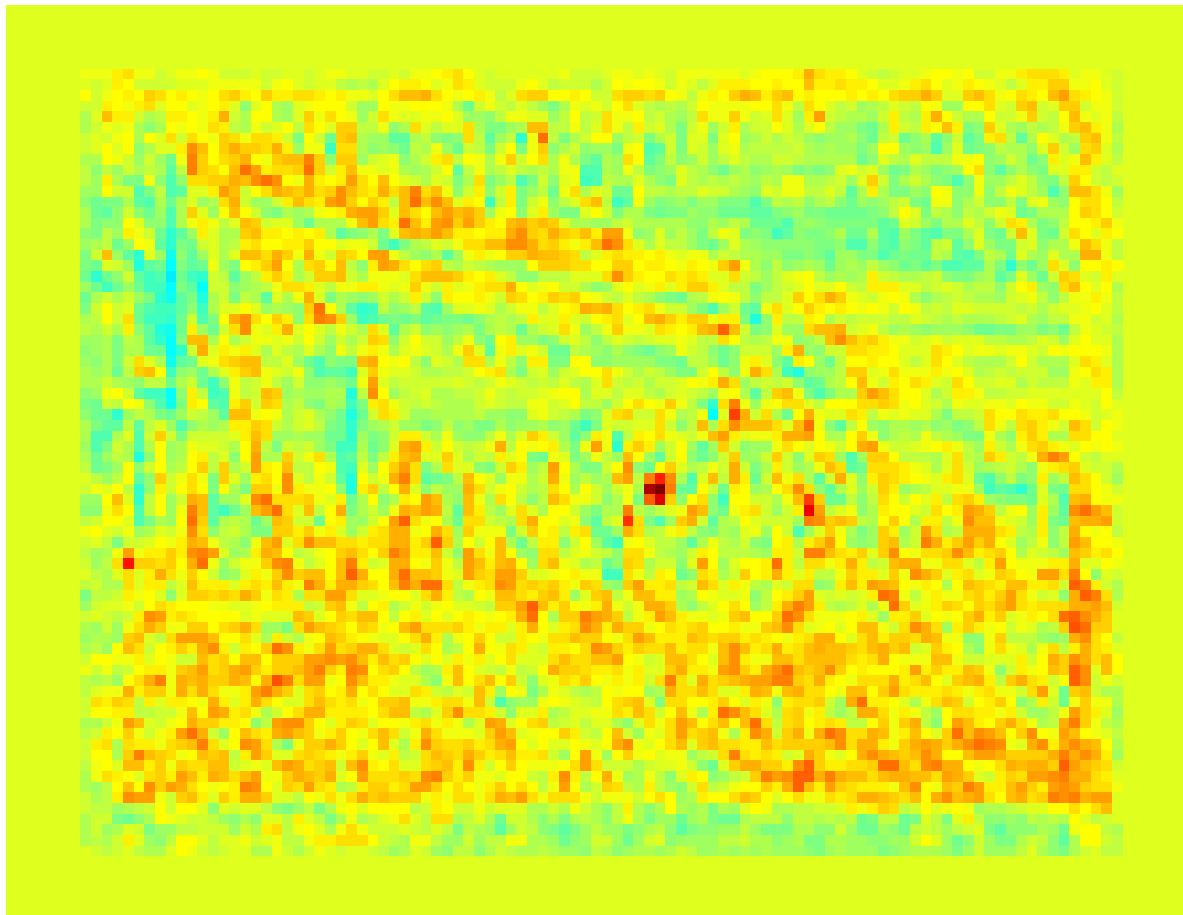
Bound tightness



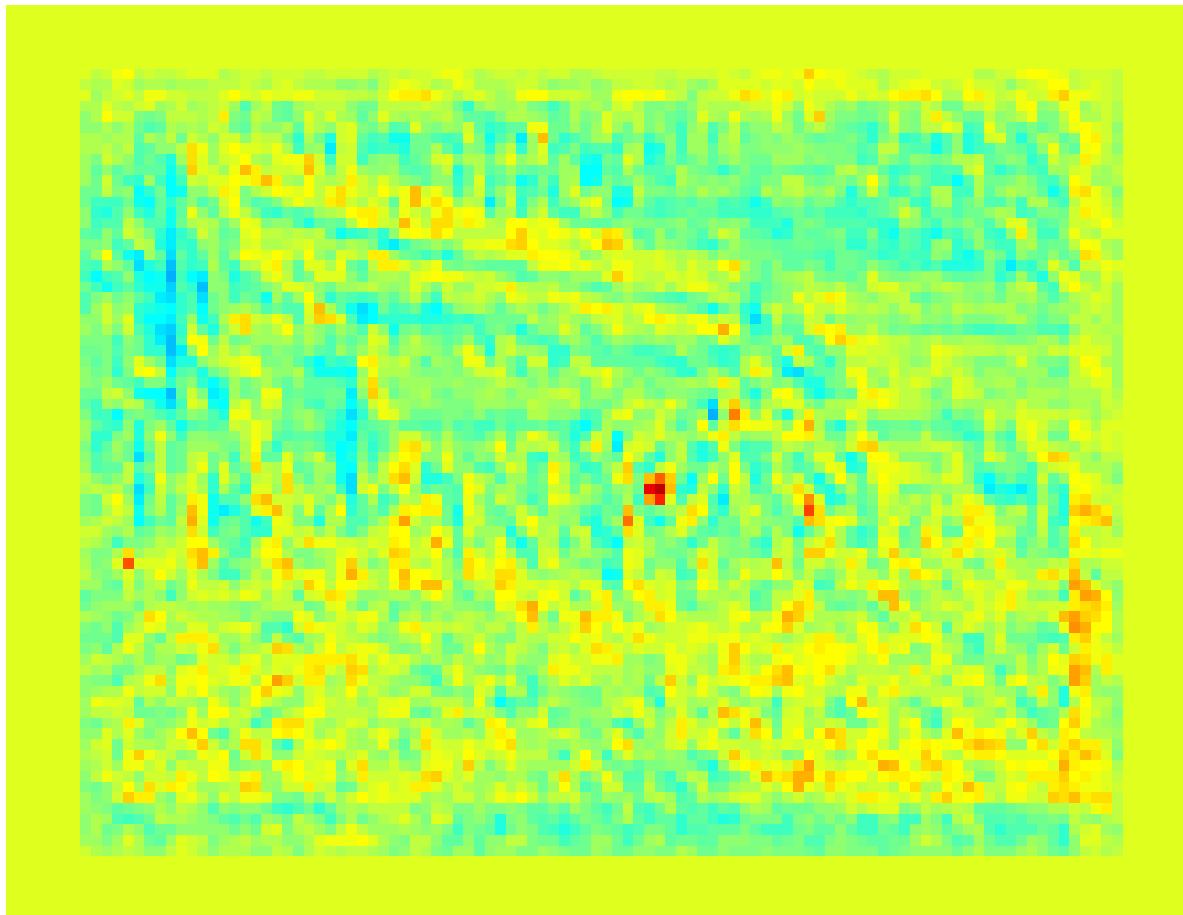
Bound demonstration: $s[x]$



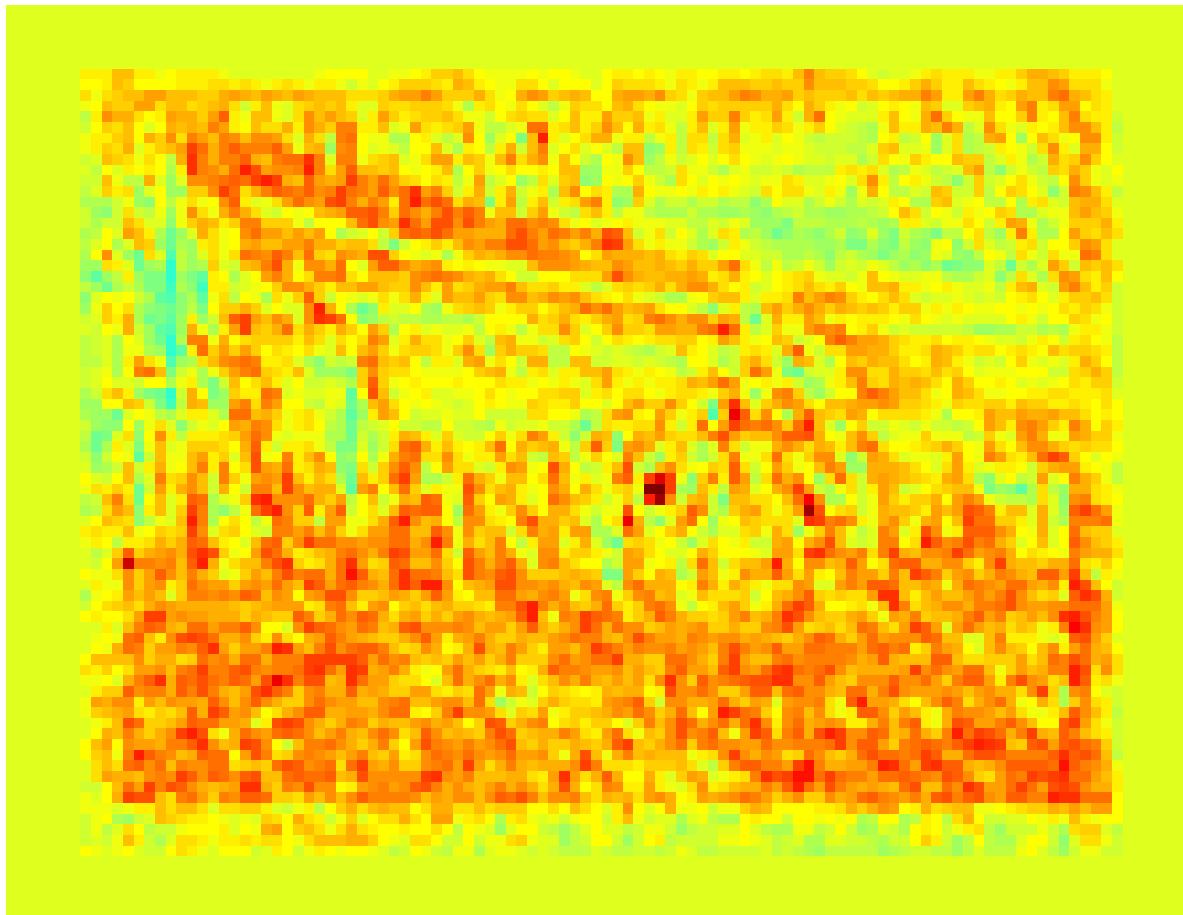
Bound demonstration: $\hat{s}[x]$



Bound demonstration: $\underline{s}[x]$, $p_e = .05$



Bound demonstration: $\bar{s}[x]$, $p_e = .05$

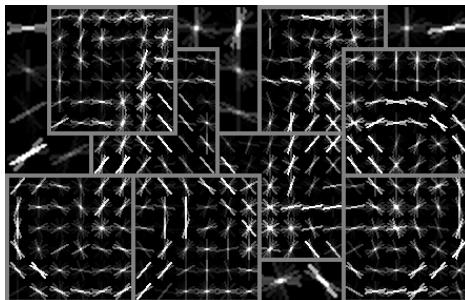


Accelerating detection with DPMs

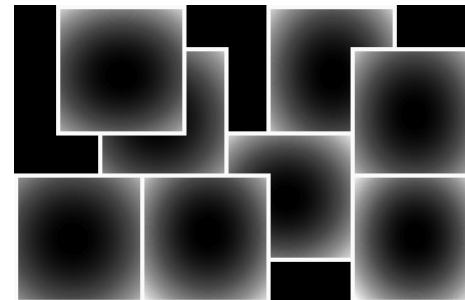
This work



w_1

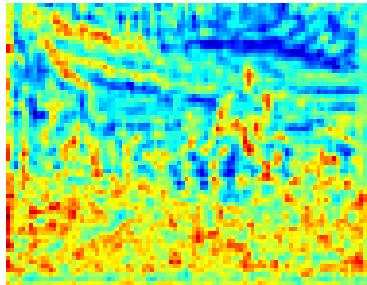


$w_2 \dots w_P$



$B_2 \dots B_P$

$$U_p(x) = \langle w_p, H(x) \rangle$$



⋮ GDT

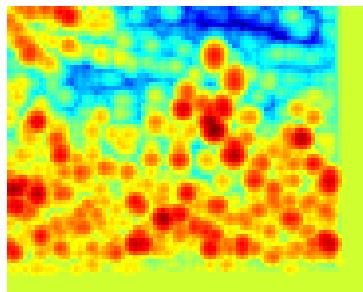
$$p = 3$$

⋮

$$p = 5$$

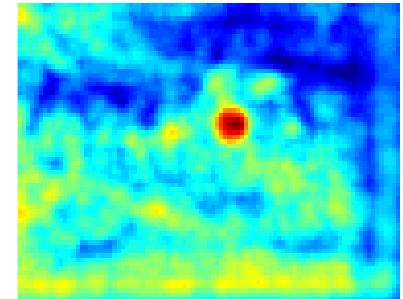
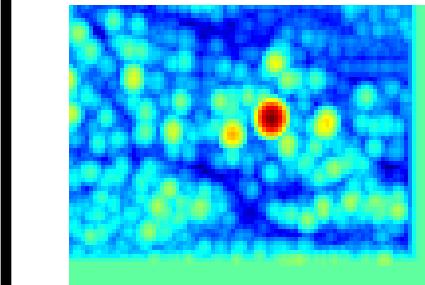
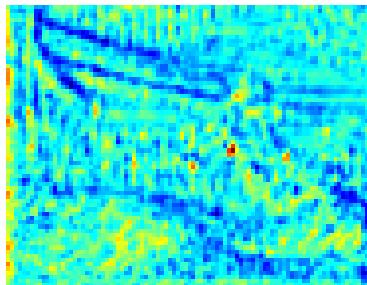
⋮

$$\mu_p(x) = \max_{x'} [U_p(x') + B_p(x, x')]$$



DTBB, NIPS 2011

$$S(x) = \sum_{p=1}^P \mu_p(x)$$



Bounding-based detection for DPMs, revisited

$$S(x) = \sum_{p=1}^P \max_{x'} [U_p(x') + B_p(x, x')]$$

$$\max_{x \in X} S(x) = \max_{x \in X} \sum_{p=1}^P \max_{x'} [U_p(x') + B_p(x, x')]$$

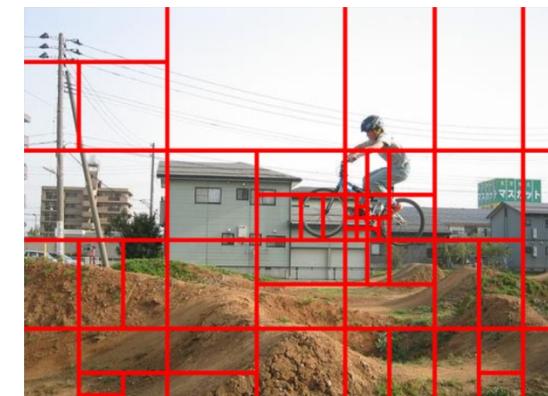
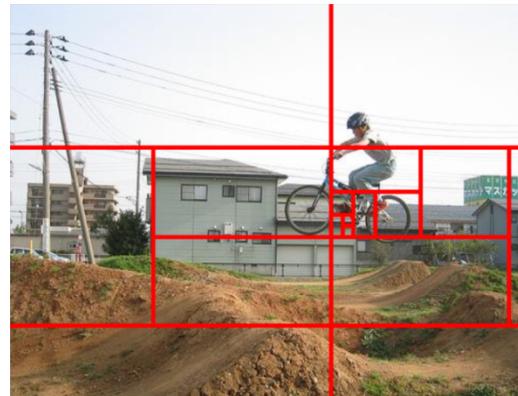
$$\leq \sum_{p=1}^P \max_{x \in X} \max_{x'} [U_p(x') + B_p(x, x')]$$

$$\leq \sum_{p=1}^P \left[\max_{x'} U_p(x') + \max_{x \in X} \max_{x'} B_p(x, x') \right]$$

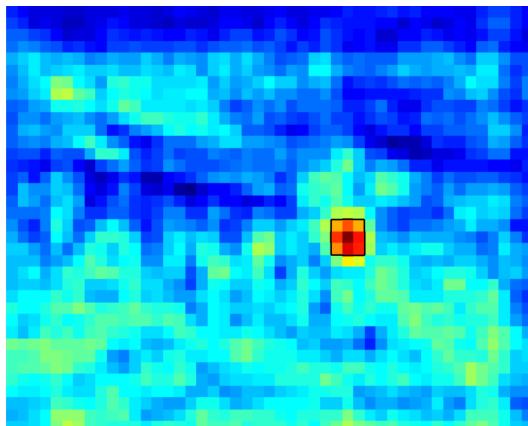
$$\leq \sum_{p=1}^P \left[\max_{x'} \bar{U}_p(x') + \max_{x \in X} \max_{x'} B_p(x, x') \right]$$

Dual-Tree Brand-and-Bound

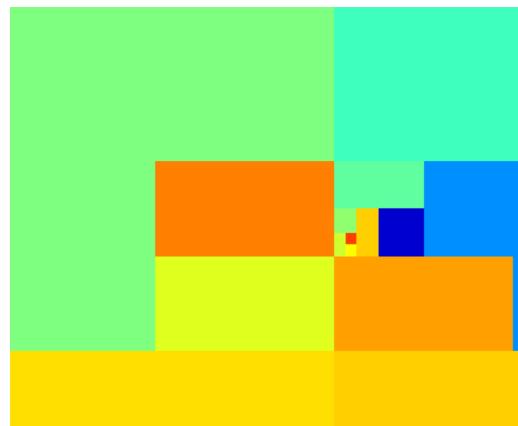
Input & Detection result



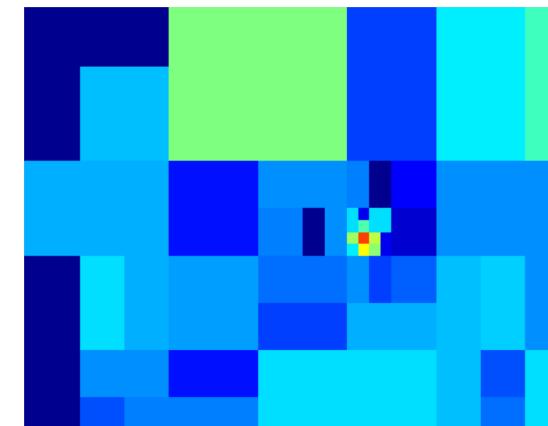
Detector score $S(x)$



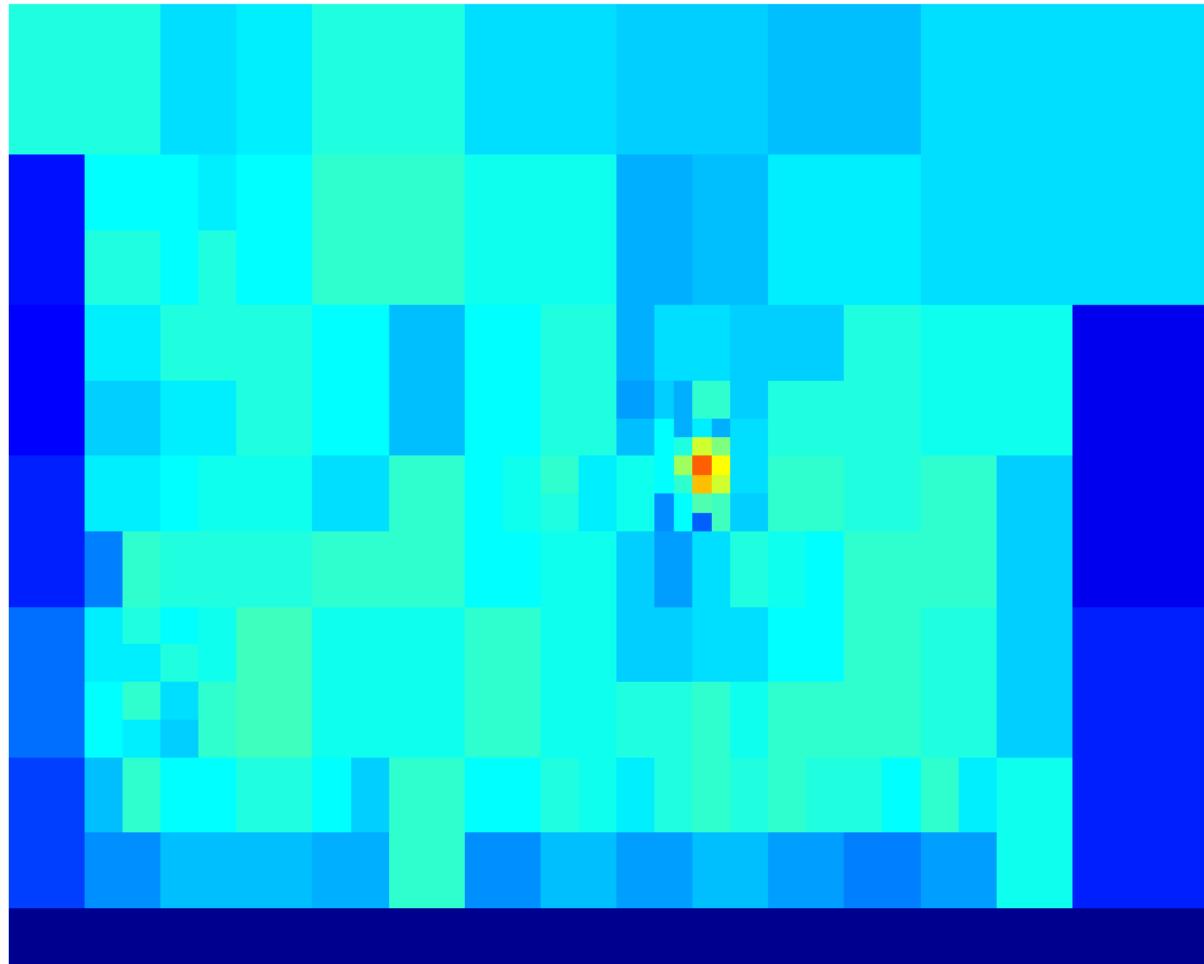
BB for $\arg \max_x S(x)$



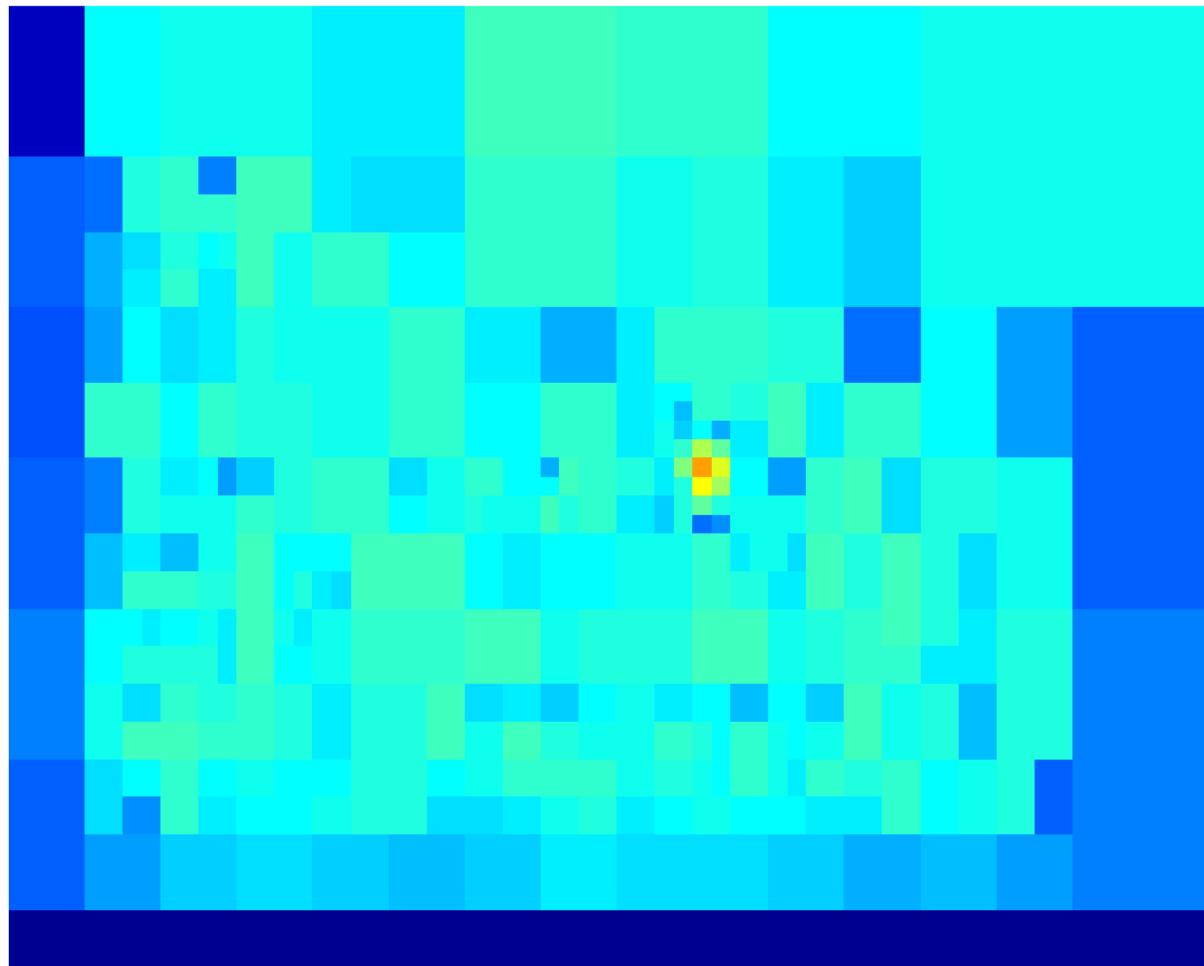
BB for $S(x) \geq -1$



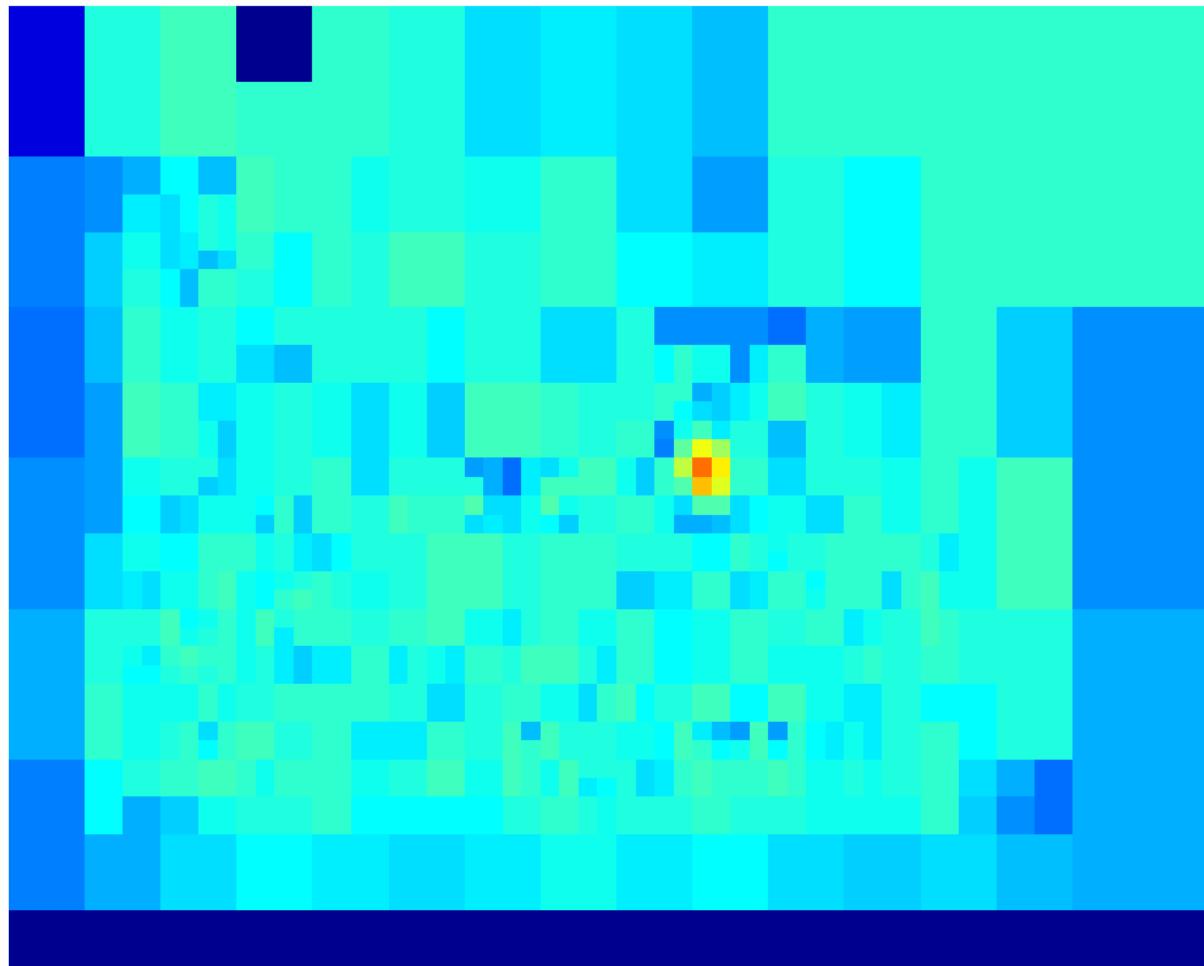
DTBB results: exact part scores



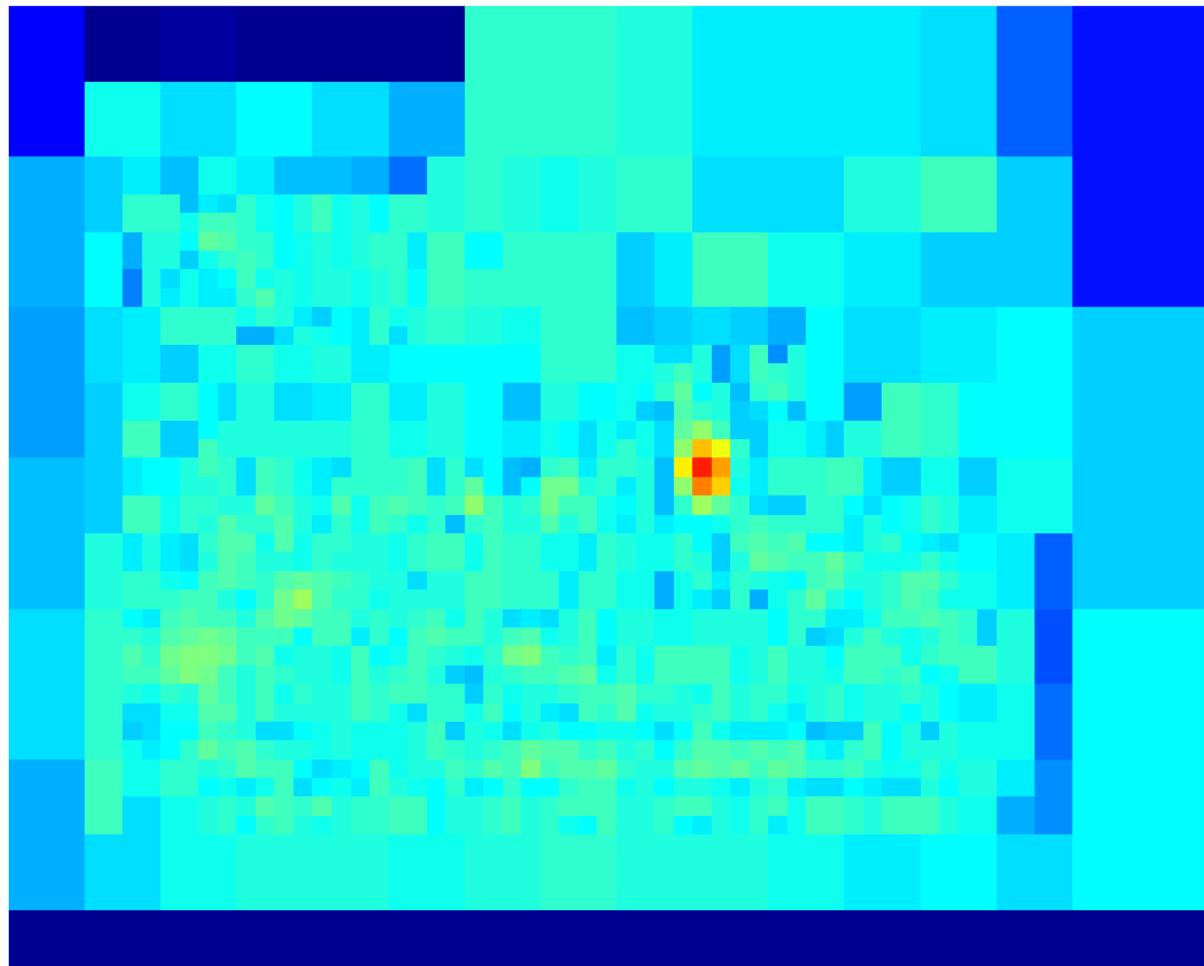
DTBB results, part score bounds @ $p_e = 0.2$



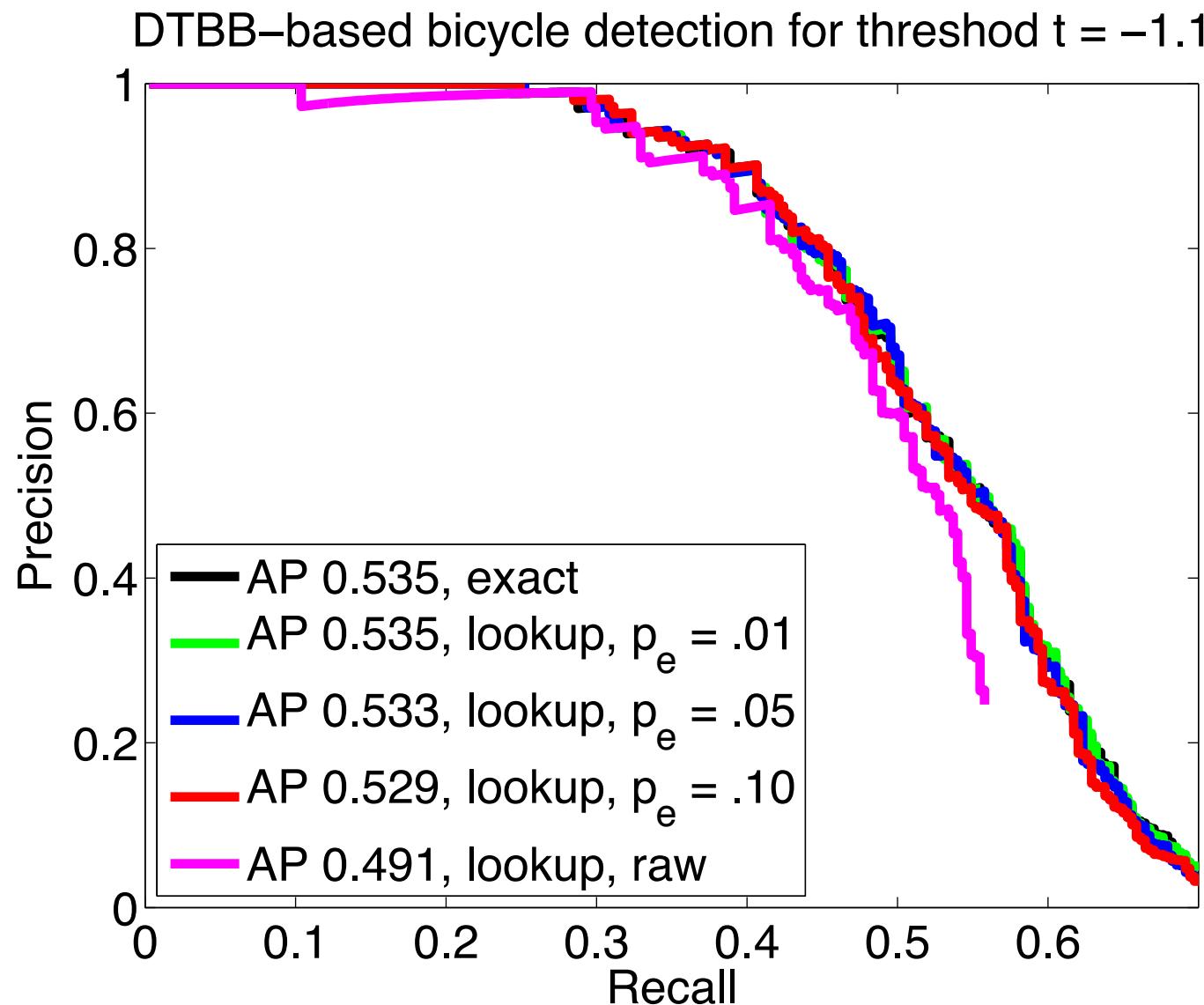
DTBB results, part score bounds @ $p_e = 0.1$



DTBB results, part score bounds @ $p_e = 0.05$



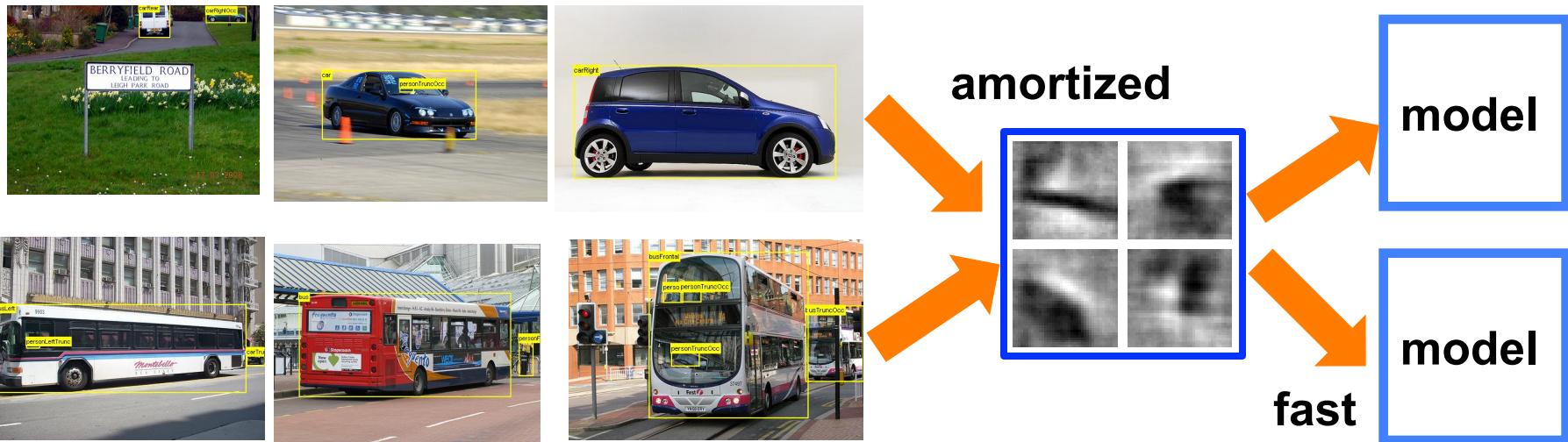
Impact on performance



Part bounds from part sharing

Pascal to Imagenet: 10K to 1000K images, 20 to 1000 categories

Sharing: perform certain operations once, reuse across all categories



D. Lowe, 'Perceptual Organization and Visual Recognition', 1985

I. Biederman, Recognition-by-components: a theory of human image understanding, 1987

S. Dickinson et. al., 3-D Shape Recovery Using Distributed Aspect Matching, 1992

A. Torralba, K. Murphy, W. Freeman, 'Sharing features: efficient boosting for multiclass object detection', CVPR 2004

A. Opelt, A. Pinz, A. Zisserman: 'Incremental learning of object detectors using a visual shape alphabet.', CVPR 2006

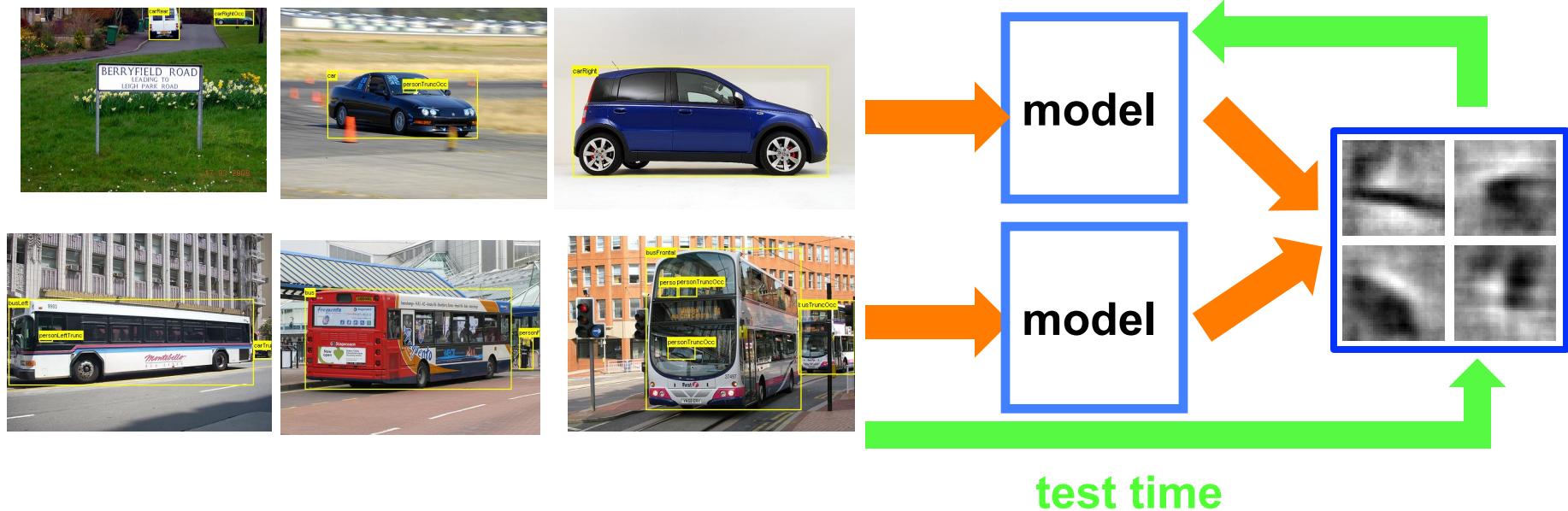
L. Zhu, Y. Chen, A. Torralba, W. Freeman, A. Yuille. Part and Appearance Sharing: RCMs for Multi-View, Multi-Object Detection', CVPR 2010

S. Fidler and A. Leonardis. Towards Scalable Representations: Learning a Hierarchy of Parts. In CVPR, 2007.

K. Mikolajczyk, B. Leibe, and B. Schiele. Multiple object class detection with a generative model. In CVPR, 2006

N. Razavi, J. Gall, and L. V. Gool. Scalable multi-class object detection. In CVPR, 2011.

Sharing pipeline, this work

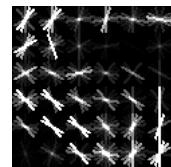
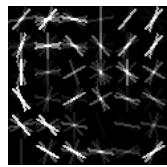


This work: decouple sharing from learning (i.e. share post-hoc)

Sharing for Deformable Part Models

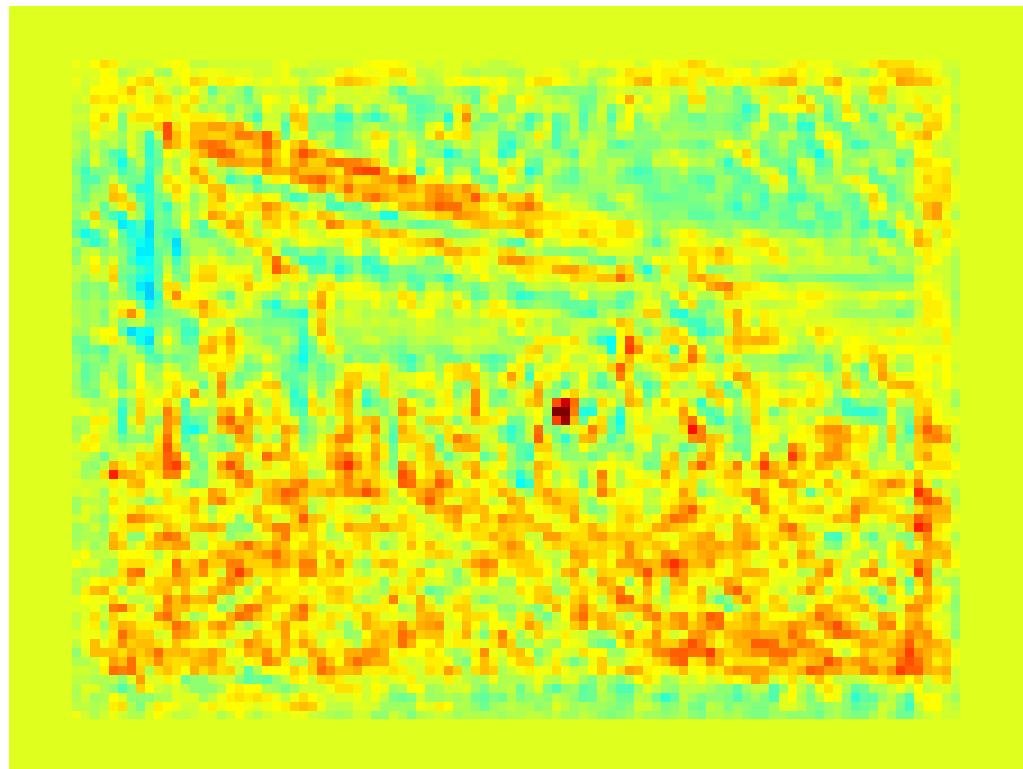
Fast and exact (cascade: 0.1-0.2 sec/category)

Part score computation



$$\mathbf{w}_p \quad \mathbf{h}_x \\ U_p(x) = \langle \boxed{\mathbf{w}_p}, \boxed{\mathbf{h}_x} \rangle$$

$\mathbf{w}_p, \mathbf{h}_x \in R^D \quad D = 6 \times 6 \times 32$



- A. Vedaldi and A. Zisserman, Sparse Kernel Maps and Faster Product Quantization Learning, CVPR 2012
- I. Kokkinos, Bounding part scores for rapid object detection, PnA workshop, ECCV 2012
- T. Dean, M. Ruzon, M. Segal, J. Shlens, S. Vijayanarasimhan, J. Yagnik, 'Fast, Accurate Detection of 100,000 Object Classes on a Single Machine' CVPR 2013
- H. Pirsiavash and D. Ramanan, Steerable Part Models, CVPR 2012
- H.O. Song, S. Zickler, T. Althoff, R. Girschick, M. Fritz, C. Geyer, P. Felzenszwalb, T. Darrell, Sparselet Models for Efficient Multiclass Object Detection, ECCV 2012
- I. Kokkinos, Shufflets, ICCV 2013

Part score approximation with a shared basis

$$s = \langle \mathbf{w}, \mathbf{h} \rangle \quad \mathbf{w}, \mathbf{h} \in R^D$$

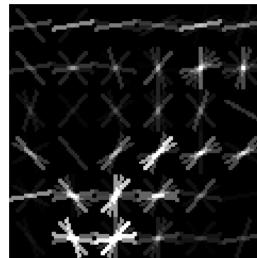
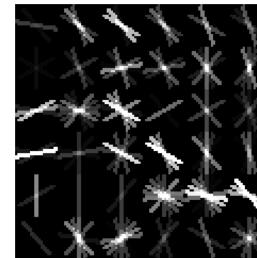
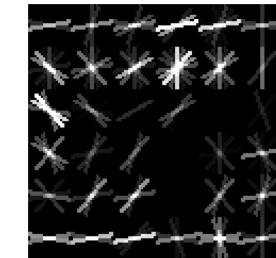
$$\mathbf{w} \simeq \hat{\mathbf{w}} = \mathbf{a}\mathbf{B} \quad \mathbf{B} = [\mathbf{b}_1 | \cdots | \mathbf{b}_K] \quad \|\mathbf{a}\|_0 = L$$

$$\mathbf{w} \simeq \hat{\mathbf{w}} = \sum_{l=1}^L \mathbf{a}_{i(l)} \mathbf{b}_{i(l)}$$

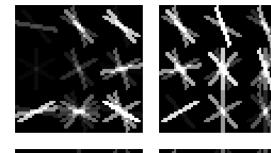
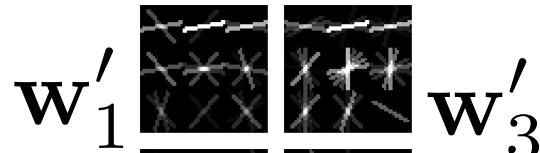
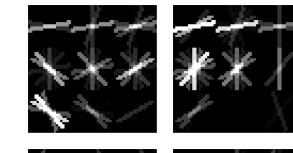
$$\langle \mathbf{w}, \mathbf{h} \rangle \simeq \sum_{l=1}^L \mathbf{a}_{i(l)} \langle \mathbf{b}_{i(l)}, \mathbf{h} \rangle$$

shared computation

Basis learning

 \mathbf{w}_1  $\dots \mathbf{w}_2$  \mathbf{w}_P

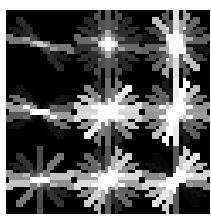
$$\min \sum_{p=1}^P \|\mathbf{w}_p - \mathbf{a}_p \mathbf{B}\|_2^2 \quad \|\mathbf{a}_p\|_0 \leq L$$

 \dots  \mathbf{w}'_{4P}

$$\min \sum_{p=1}^{4P} \|\mathbf{w}'_p - \mathbf{a}' \mathbf{B}'\|_2^2 \quad \|\mathbf{a}'_p\|_0 \leq \frac{L}{4}$$

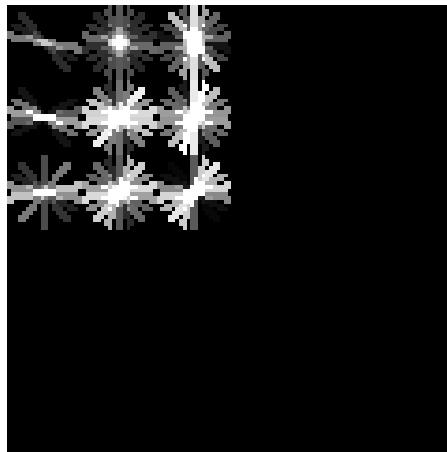
Shufflets: “Shiftable” basis

“shufflet”

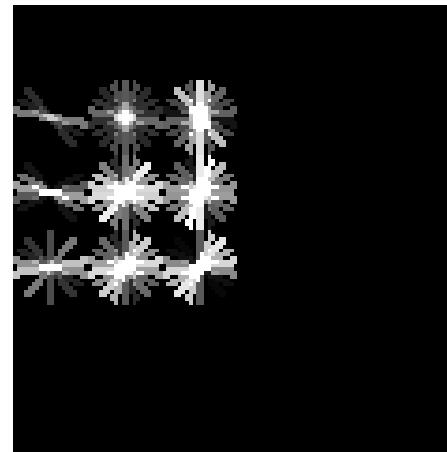


\mathbf{k}_1

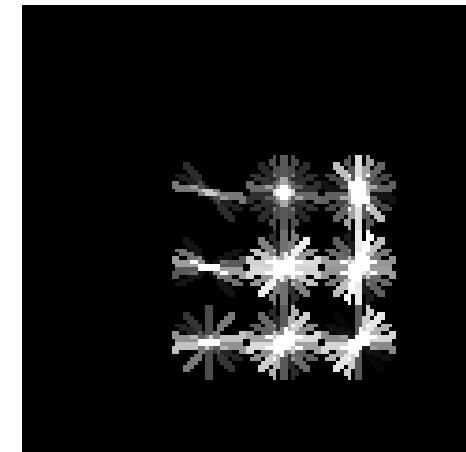
“replicas”



$\mathbf{k}_1^{0,0}$



$\mathbf{k}_1^{1,0}$

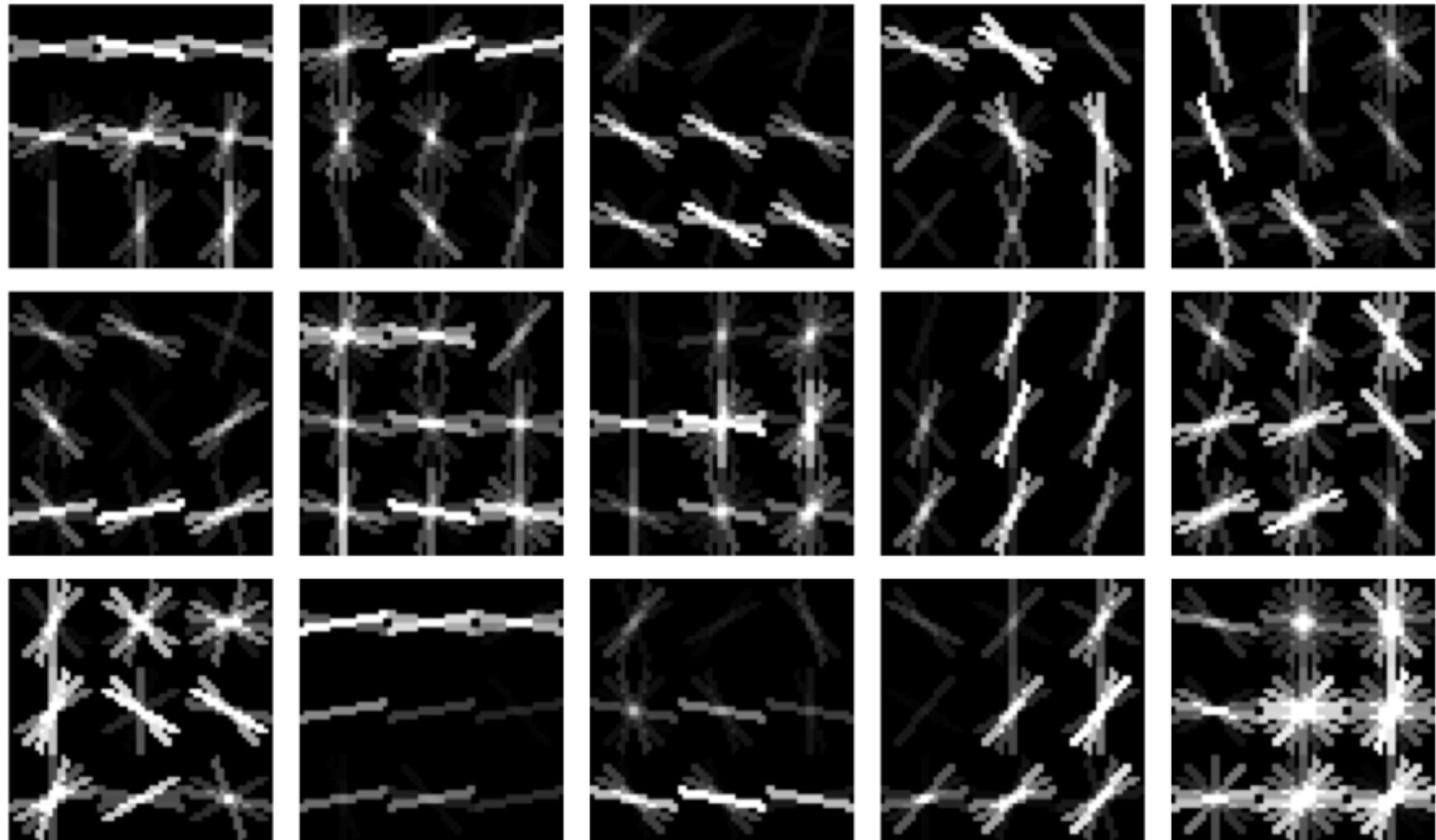


$\mathbf{k}_1^{m,n}$

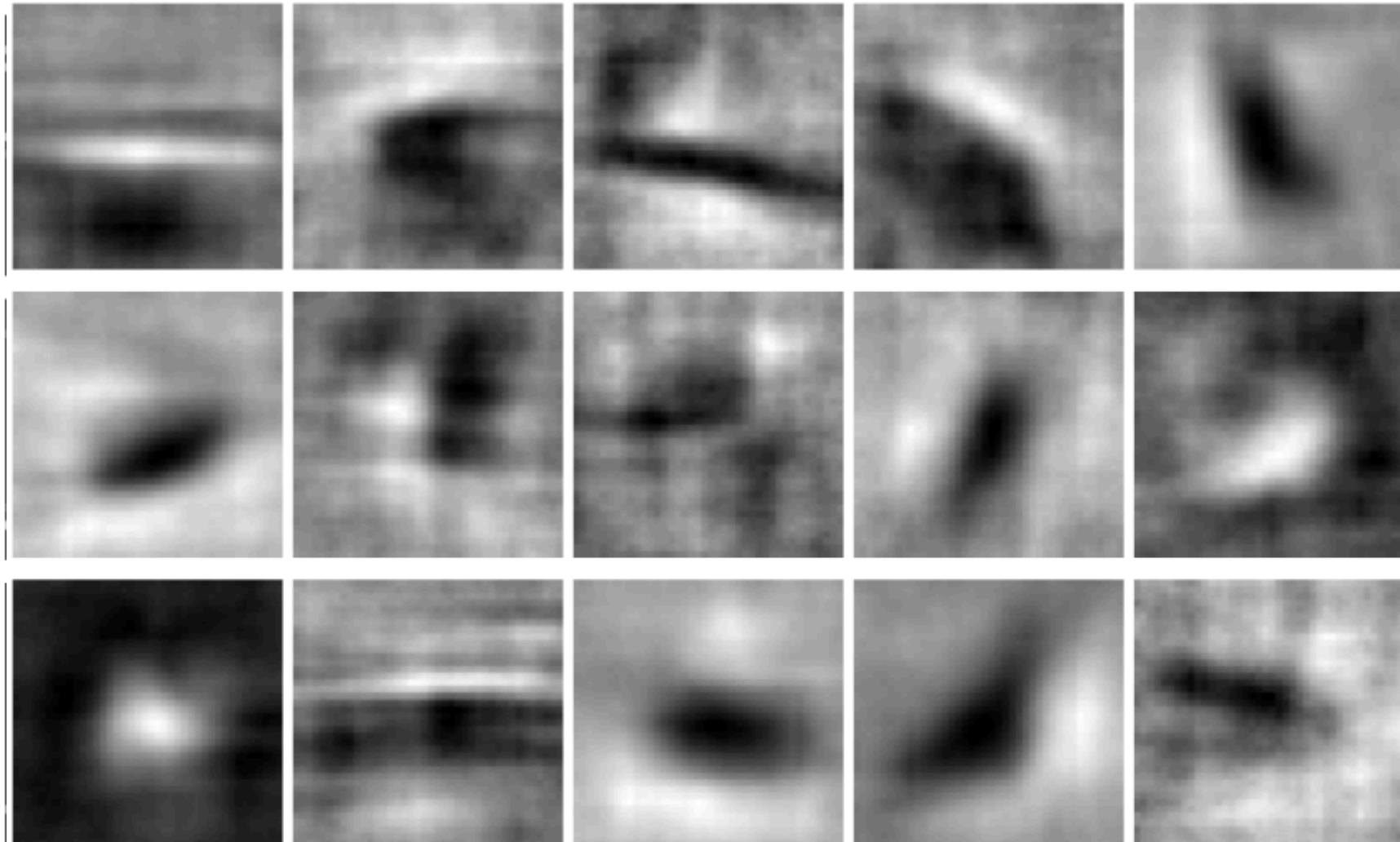
$$\min \sum_{p=1}^P \|\mathbf{w}_p - \mathbf{a}_p \mathbf{B}\|_2^2 \quad \mathbf{B} = \{\mathbf{k}_1^{0,0}, \dots, \mathbf{k}_1^{v,h}, \dots, \mathbf{k}_D^{v,h}\}$$

$$\|\mathbf{k}_d\|_2 \leq 1, \quad d = 1 \dots D \quad \|\mathbf{a}_p\|_0 \leq L$$

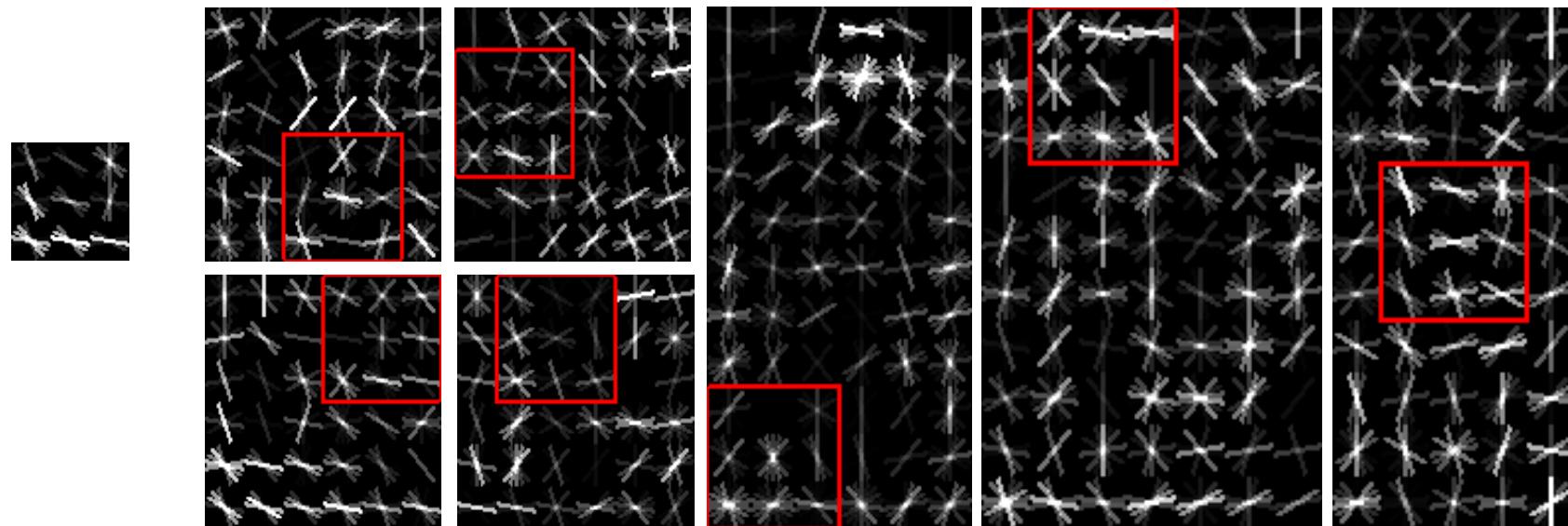
Shufflet basis (top 12 out of 128)



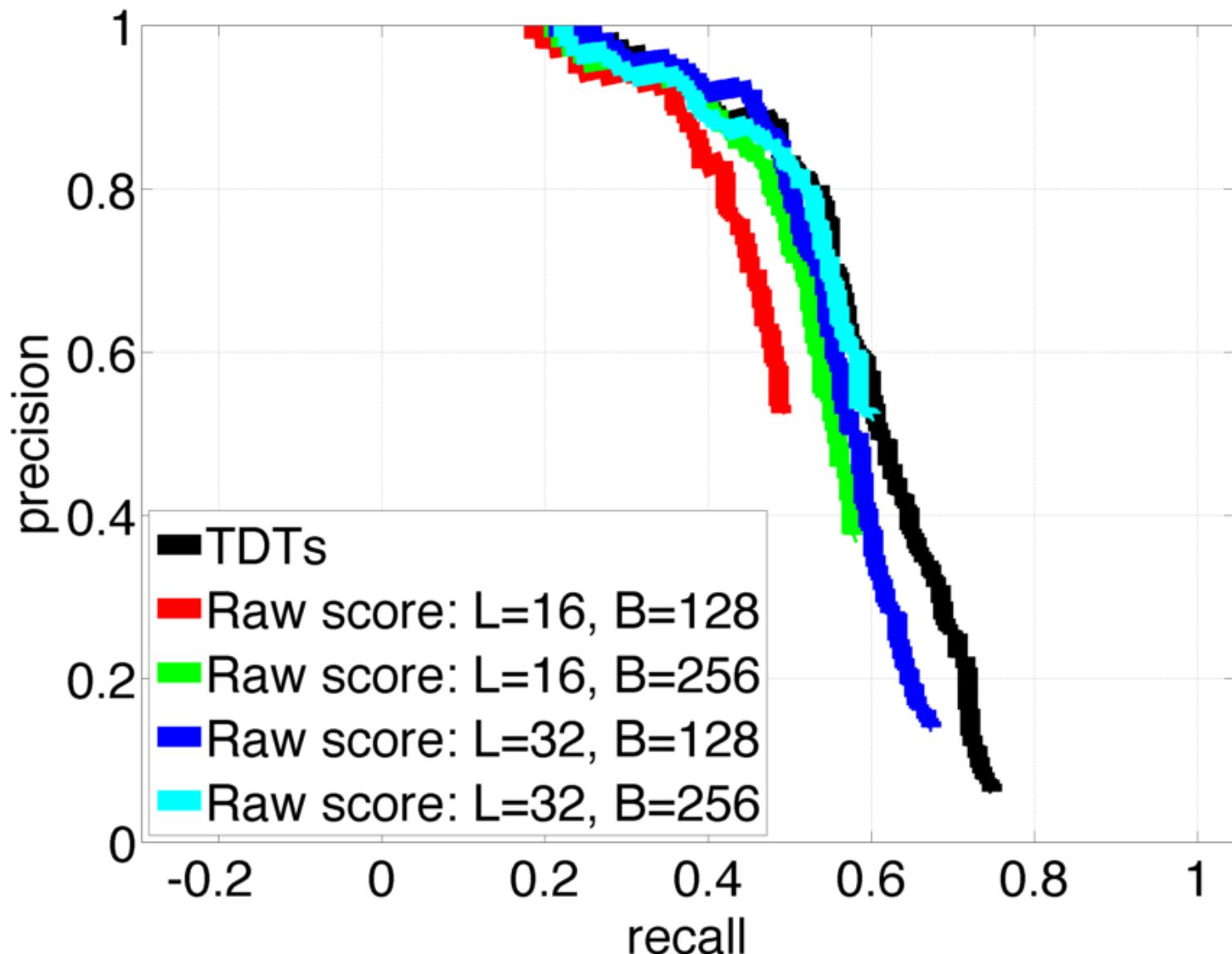
Inverted shufflet elements



The multiple lives of a shufflet



Impact of approximation on detection performance



Cost breakdown

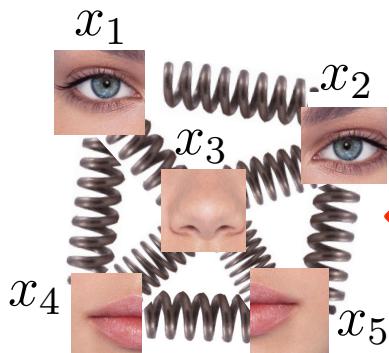
	Part scores	Part combination	Total
DPM, V5	2.26	0.41	2.67
DTBB, 2011	1.69	0.21	1.90
This work	0.41	0.54	0.95
		$\theta = -.5$	$\theta = -.5$

Cascade DPMs

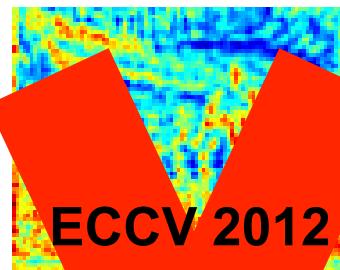
DPM, V5	0.56
Shufflets	0.23
	$\theta = -.5$

HOG pyramid	0.43
Shufflets	0.89

Work in progress: putting it all together



$$U_p(x') = \langle \mathbf{w}_p, \mathbf{H}(x') \rangle$$



ECCV 2012
ICCV 2013

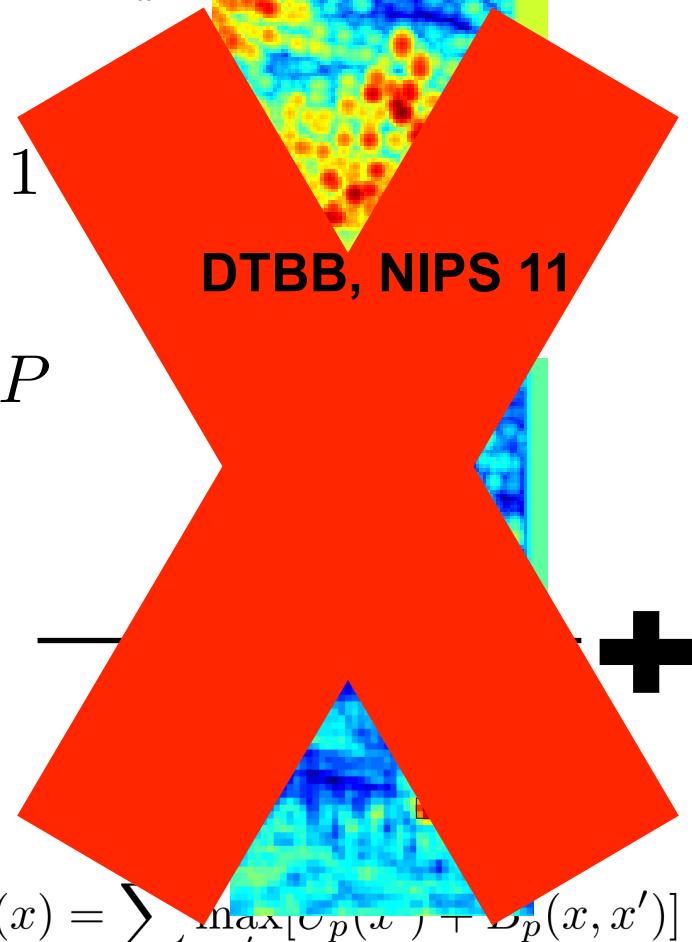
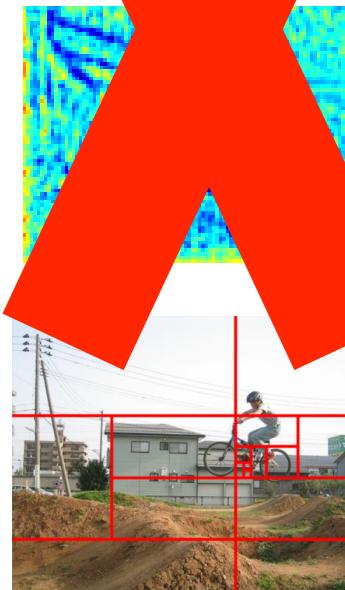
$$p = 1$$

 \vdots

$$p = P$$

$$\max_{x'} [U_p(x') + B_p(x, x')]$$

DTBB, NIPS 11



$$S(x) = \sum_{p=1}^P \max_{x'} [U_p(x') + B_p(x, x')]$$

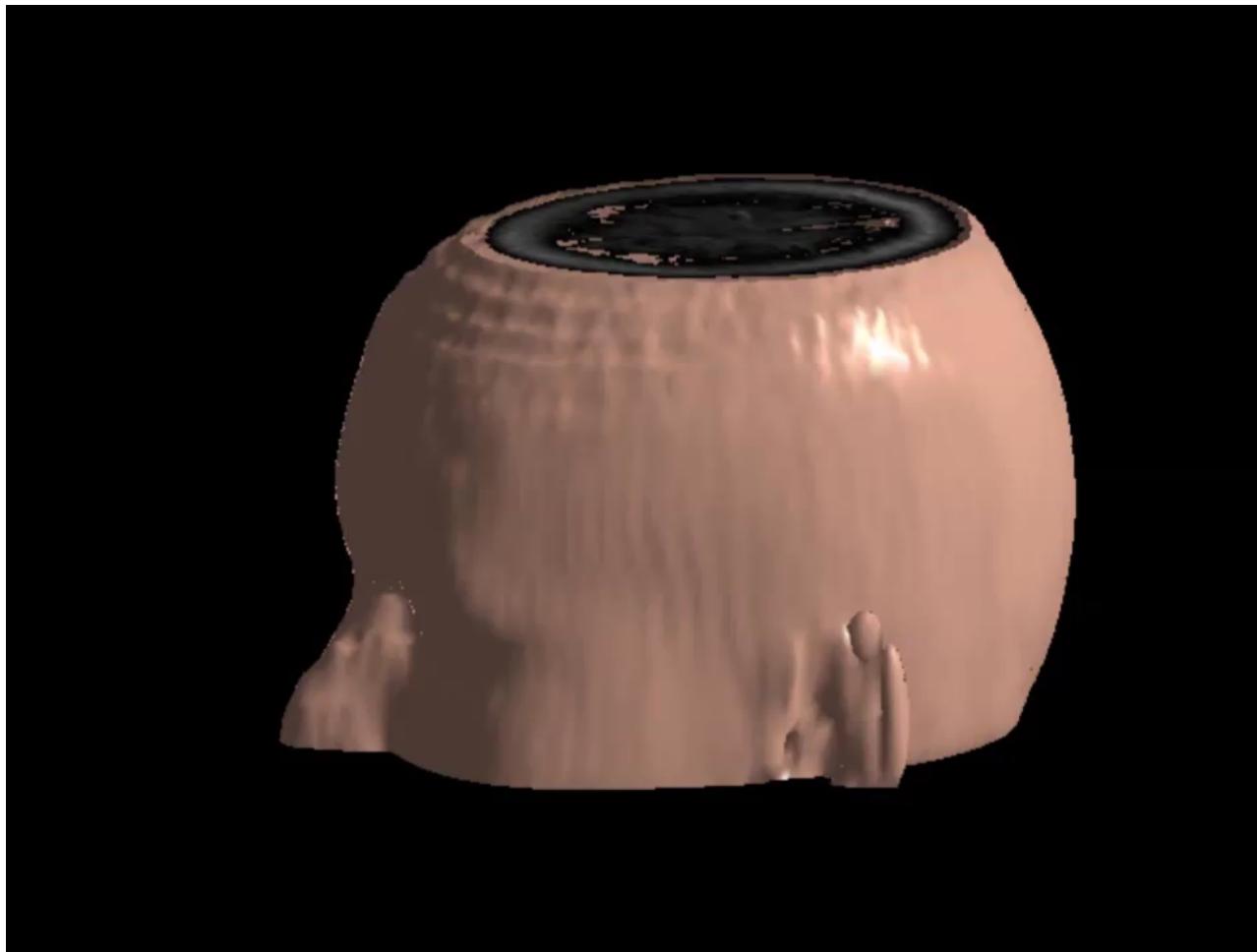
I. K., Rapid DPM detection using Dual Tree Branch-and-Bound, NIPS, 11

I. K., Bounding Part Scores for Rapid Object Detection with DPMs, ECCV12, Shufflets,

H. B. Gross and I. K., Fast and Exact: ADMM-Based Discriminative Shape Segmentation using Loopy Part Models, CVPR 14

P. Felzenszwalb and D. McAllester, The Generalized A* Architecture,, JAIR 07

More dimensions: 3D medical data



H. Boussaid

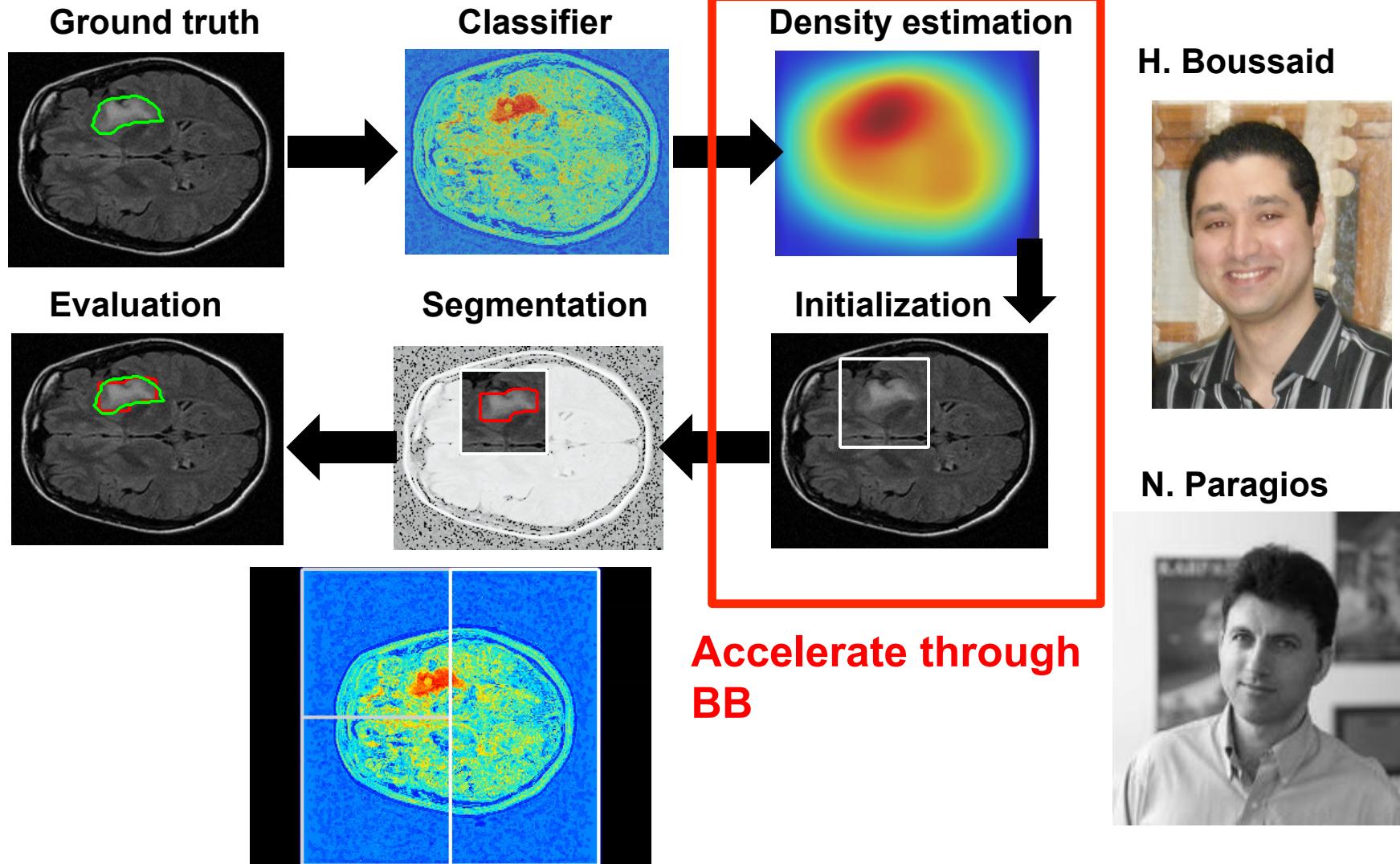


N. Paragios

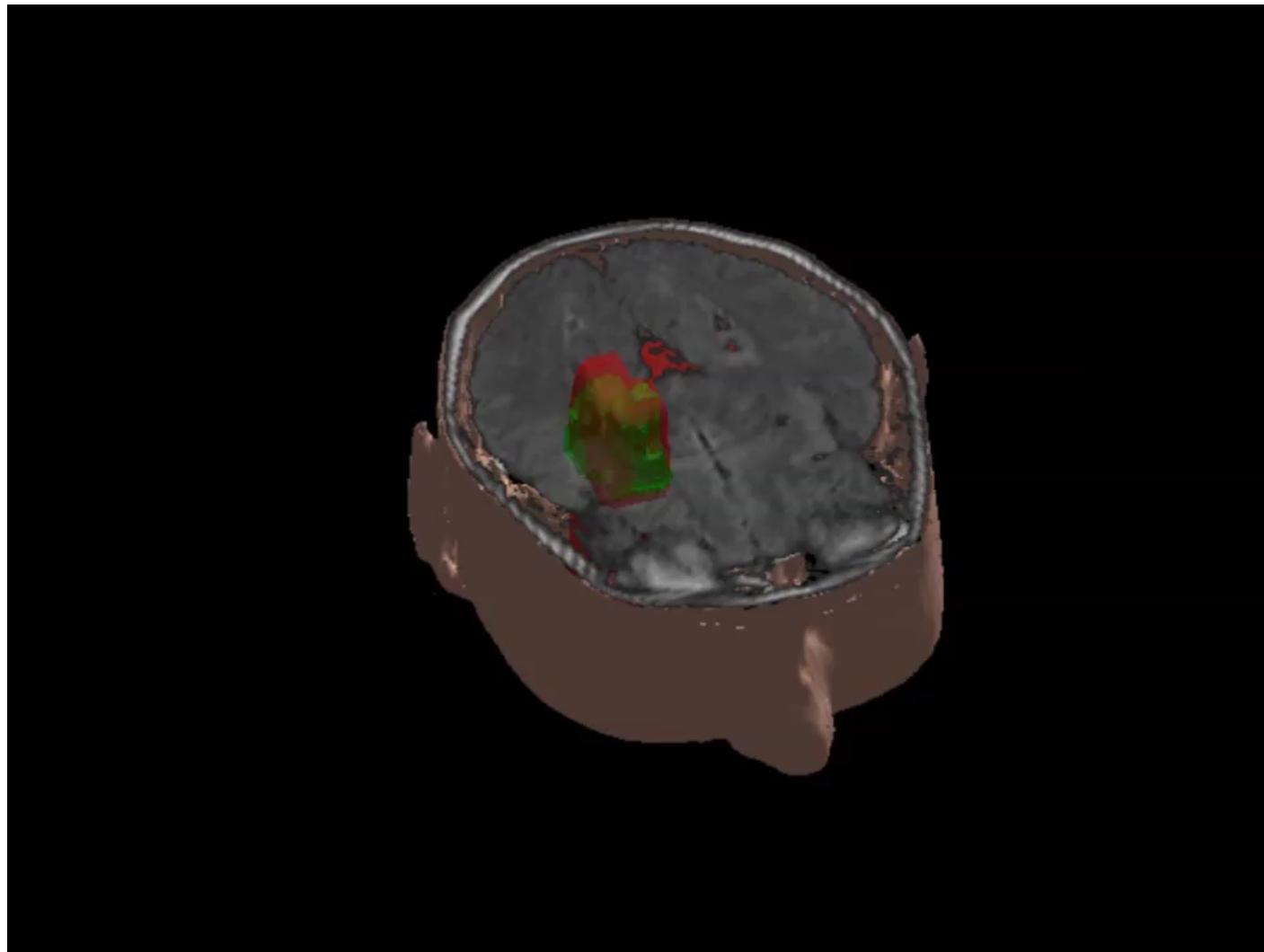


H. Boussaid, I. Kokkinos, and N. Paragios, Rapid Mode Estimation for 3D Brain MRI Tumor Segmentation, Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), 2013.

3D volume segmentation - medical image analysis



3D volume segmentation - medical image analysis



H. Boussaid



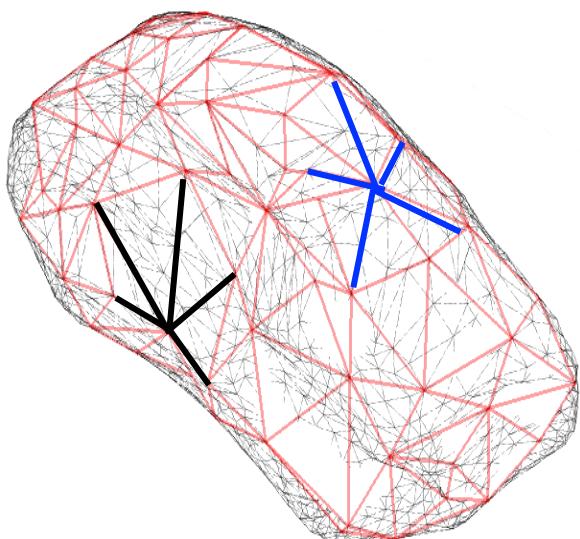
N. Paragios



H. Boussaid, I. Kokkinos, and N. Paragios, Rapid Mode Estimation for 3D Brain MRI Tumor Segmentation, Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), 2013.

Pipeline

Loopy DPM model



- decomposition of all edges of the 3d shape into star-shaped graphs
- **pairwise**: nominal displacements and variances for each edges estimated from training dataset
- **unaries**: affinity of each point of the mesh to the 2d image (invariant in Z)

Optimization:

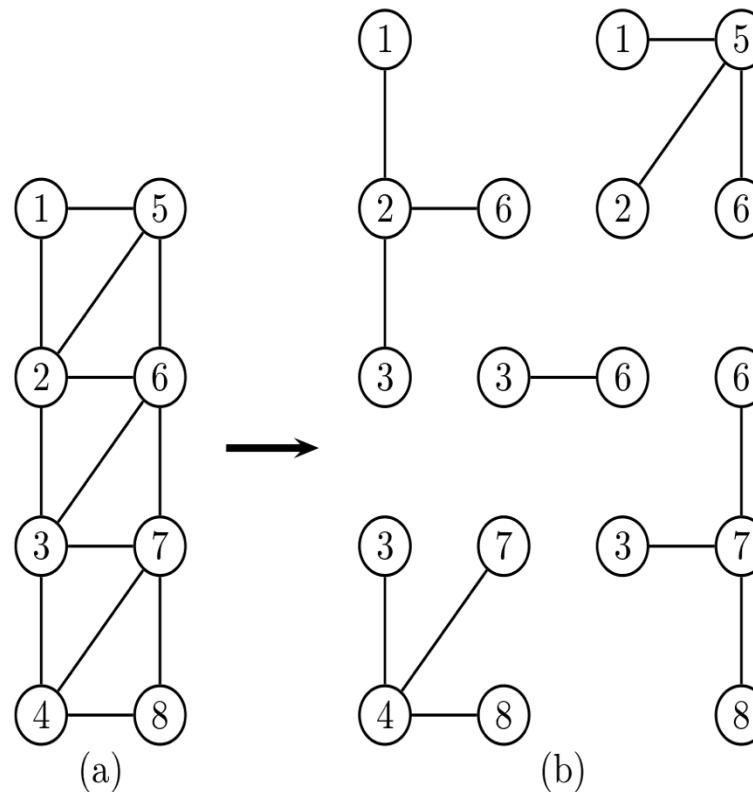
- Branch –and– bound for each slave + ADMM

Structured learning:

- first iteration: unaries only for annotated keypoints (provides ground truth)
- subsequent iterations: unaries and pairwise adjust to make the keypoints closer to the ground truth

ADMM - Branch&Bound

- In ADMM the complete graph is split into several states



3D NPMs

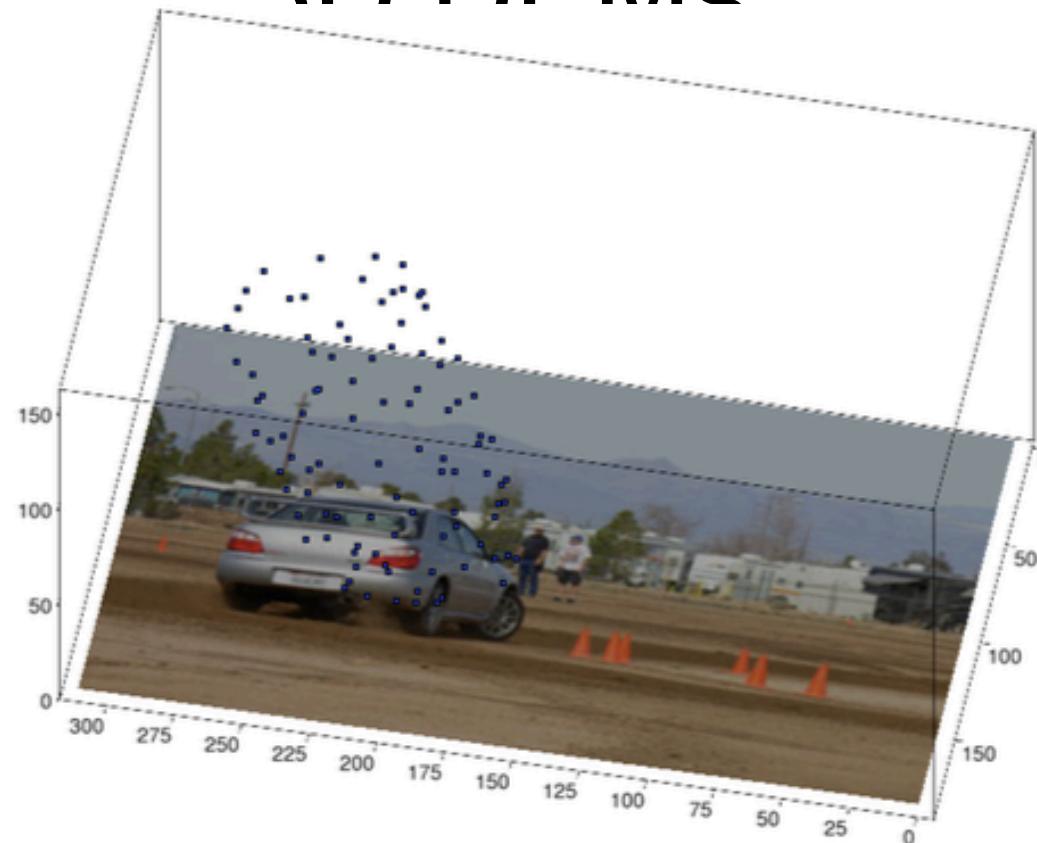


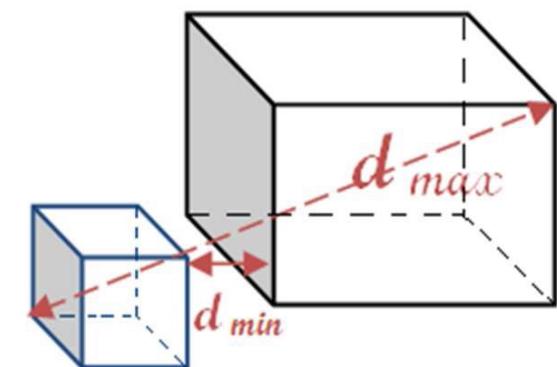
Figure 3.5: This figure shows how the algorithm reconstructs a car. The image plane containing the photo of the car is the input, the blue points represent the positions of the car model reconstructed in 3d.

image space with Branch&Bound

3D Branch & Bound

- previously: every part gets a position (width*height ~ i.e. 640*480) in the 2d image
- now: every part is assigned a 3d position (width*height*depth ~ 640*480*200)
- Inference on a single slave with 5 parts needs less than 180ms in a 100*100*200 image space with Branch&Bound
- With dynamic programming the inference takes about 1.5sec

$$\bar{\mu}_X^Y \doteq \left(\sum_{i \in Y} w_i \right) \max_{i \in Y} \max_{j \in X} K(x_i, x_j)$$



From image processing to object detection

Test-time acceleration

Convolution Theorem

$$y(t) = x(t) * h(t) \Leftrightarrow Y(\omega) = X(\omega) \cdot H(\omega)$$

Distributive property

$$x * (y + z) = x * y + x * z$$

Training-time acceleration

Shifting property

$$x(t) \xleftrightarrow{\mathcal{F}} X(\omega)$$

$$x(t - t_0) \xleftrightarrow{\mathcal{F}} X(\omega)e^{-j\omega t_0}$$

Outline

Convolution Theorem

C. Dubout and F. Fleuret. Exact Acceleration of Linear Object Detectors, ECCV 12

M. Mathieu, M. Henaff, Y. LeCun, Fast Training of Convolutional Networks through FFTs, ICLR 2013

Distributive property

H. Pirsiavash, D. Ramanan. 'Steerable Part Models' CVPR 12.

H. O. Song, et. al. Sparselet models for efficient object detection. ECCV 12.

I. Kokkinos, Shufflets: Shared Mid-level Parts for Fast Object Detection, ICCV13.

Shifting property

I. Kokkinos and A. Yuille, Scale Invariance without Scale Selection CVPR '08.

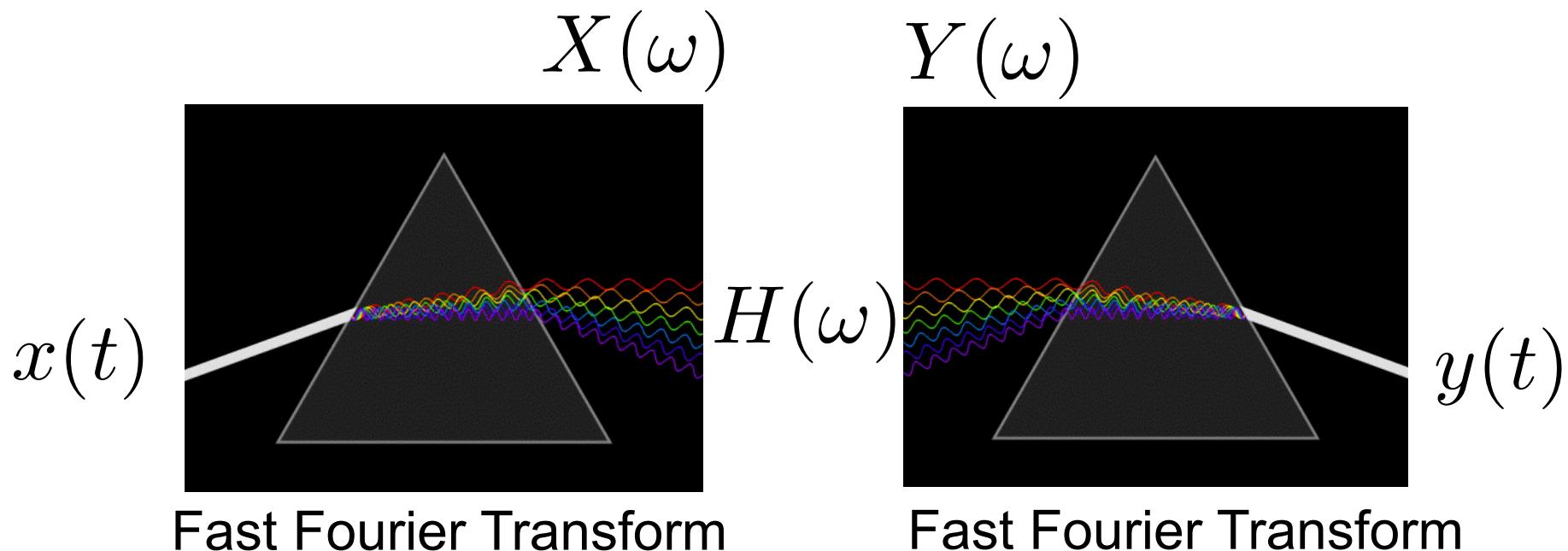
R. B. Grosse, R. Raina, H. Kwong, and A. Y. Ng: Shift-Invariant Sparse Coding for Audio Classification. UAI '07.

K. Galoogahi, T. Sim and S. Lucey, 'Multi-Channel Correlation Filters', ICCV '13

J. F. Henriques, J. Carreira, R. Caseiro, J. Batista, 'Beyond Hard Negative Mining: Efficient Detector Learning via Block-Circulant Decomposition', ICCV '13

Convolution Theorem

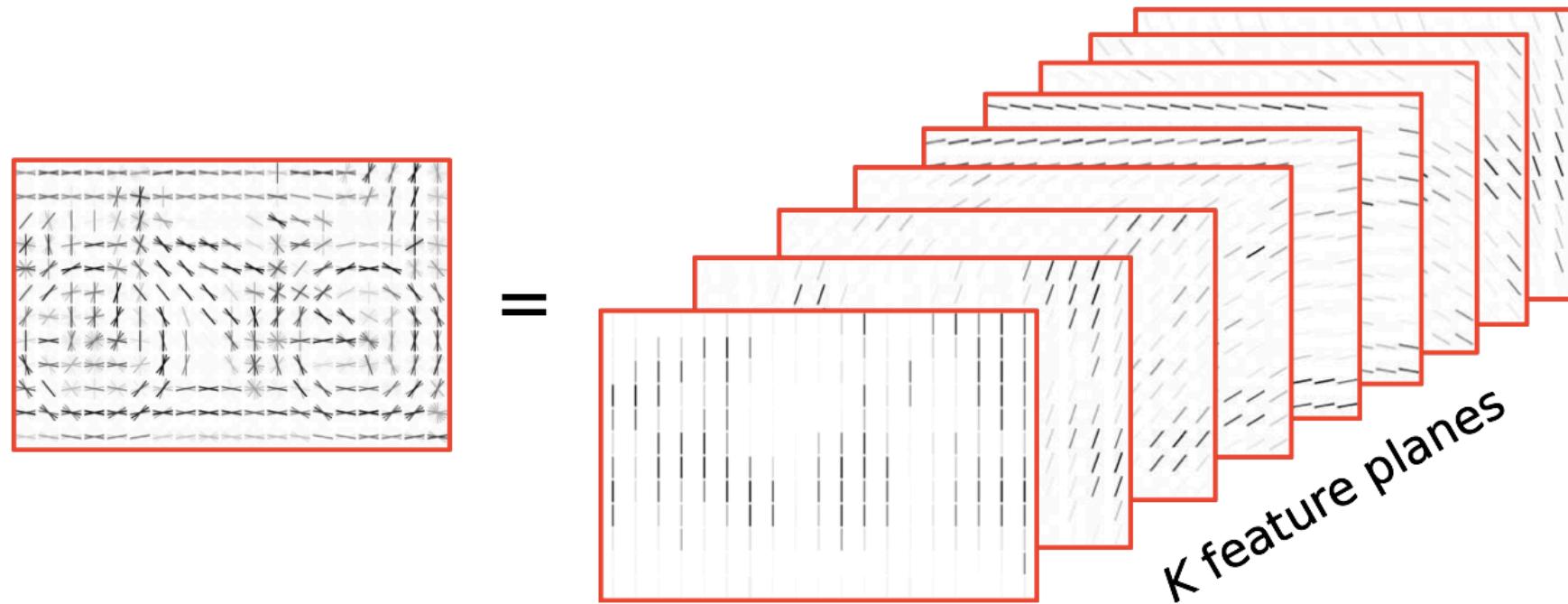
$$y(t) = x(t) * h(t) \leftrightarrow Y(\omega) = X(\omega) \cdot H(\omega)$$



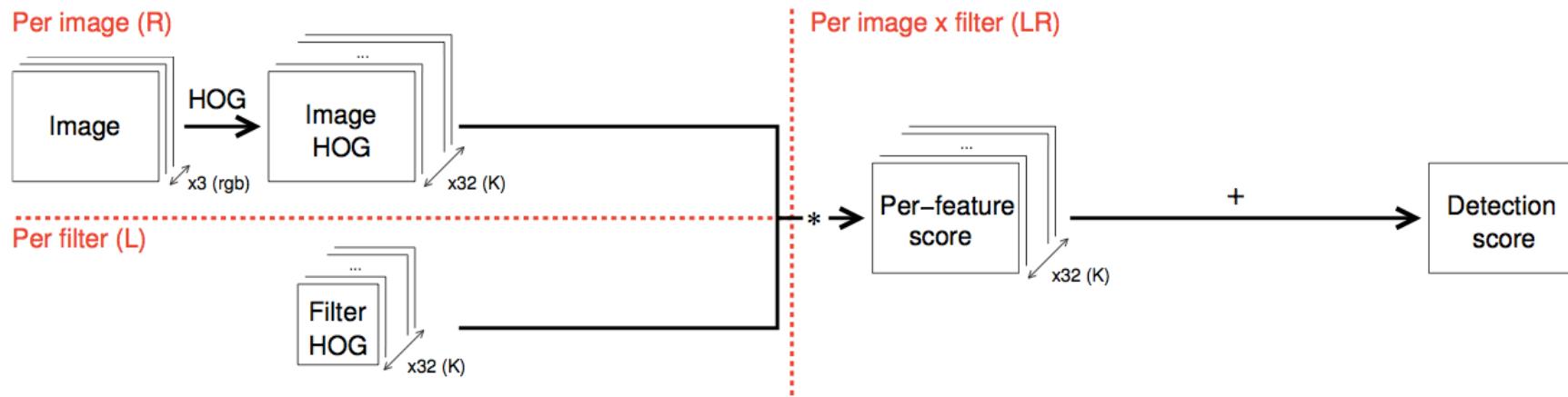
$$O(NK) \rightarrow O(N \log N)$$

Typically only interesting if $K > \log N$

HOG features = Multi-Channel images



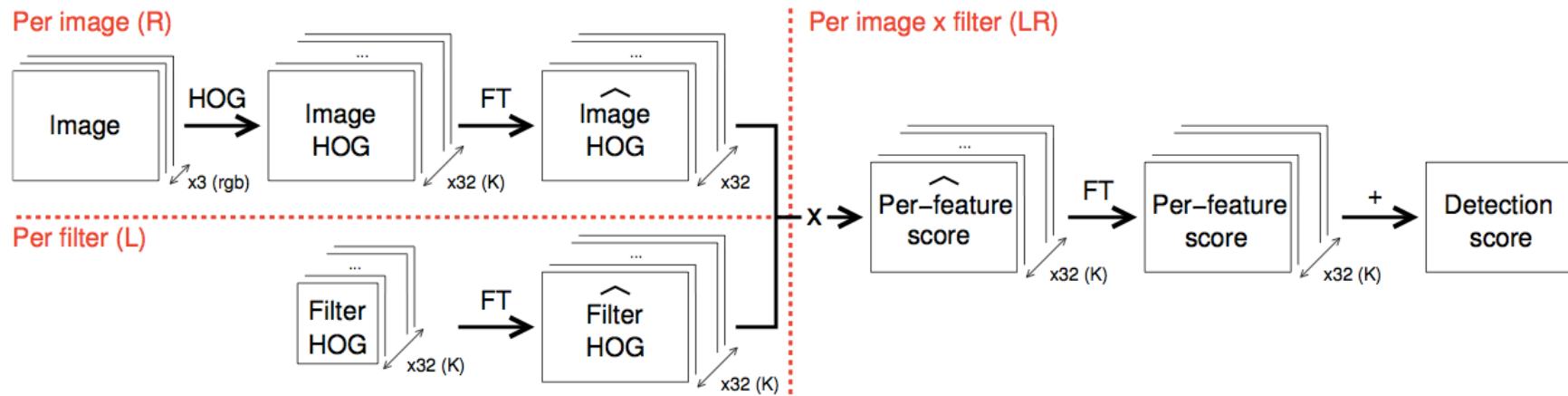
Time-Domain implementation



M points, L filters with C-wide support, K channels:

$$O(MCLK)$$

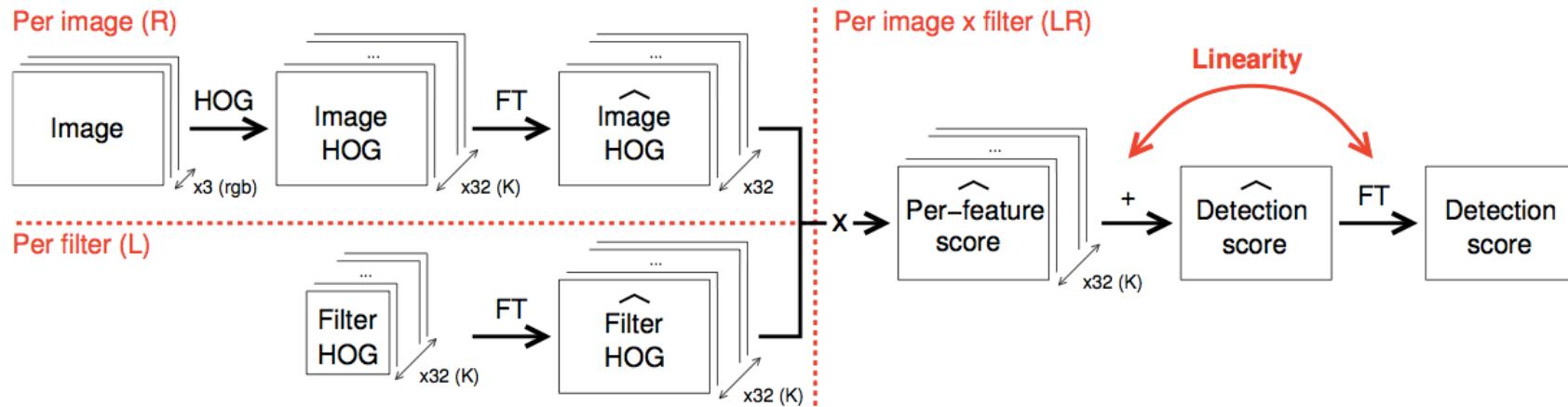
Naïve Frequency-Domain Implementation



MxM points, L CxC filters, K channels:

$$O(MK \log M) + O(KLM) + O(KLM \log M)$$

Smart Frequency-Domain Implementation



$M \times M$ points, $L \times C$ filters, K channels:

$$\begin{aligned}
 O(MK \log M) + O(KLM) + O(LM \log M) \\
 \simeq O(KLM)
 \end{aligned}$$

Outline

Convolution Theorem

C. Dubout and F. Fleuret. Exact Acceleration of Linear Object Detectors, ECCV 12

Distributive property

H. Pirsiavash, D. Ramanan. 'Steerable Part Models' CVPR 12.

H. O. Song, et. al. Sparselet models for efficient object detection. ECCV 12.

I. Kokkinos, Shufflets: Shared Mid-level Parts for Fast Object Detection, ICCV13.

Shifting property

I. Kokkinos and A. Yuille, Scale Invariance without Scale Selection CVPR '08.

R. B. Grosse, R. Raina, H. Kwong, and A. Y. Ng: Shift-Invariant Sparse Coding for Audio Classification. UAI '07.

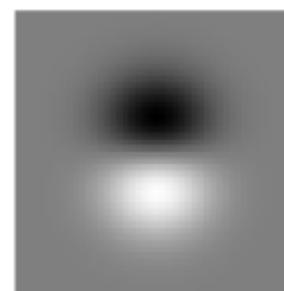
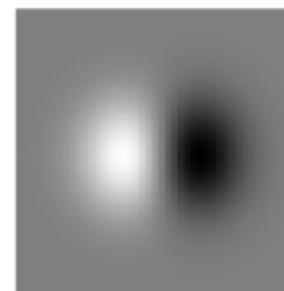
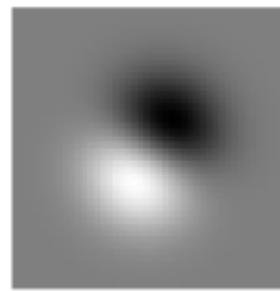
K. Galoogahi, T. Sim and S. Lucey, 'Multi-Channel Correlation Filters', ICCV '13

J. F. Henriques, J. Carreira, R. Caseiro, J. Batista, 'Beyond Hard Negative Mining: Efficient Detector Learning via Block-Circulant Decomposition', ICCV '13

Distributive property & Steerable Filters

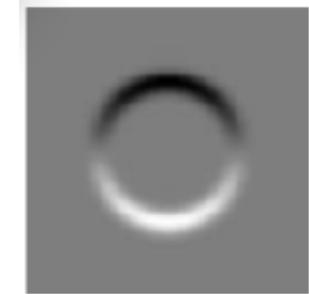
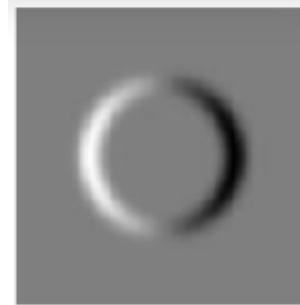
Distributive property: $f * (g + h) = f * g + f * h$

Steerable filter: $g_\theta(x, y) = \cos(\theta)g_0(x, y) + \sin(\theta)g_{\pi/2}(x, y)$



I

$$I * g_\theta = \cos(\theta)(I * g_0) + \sin(\theta)(I * g_{\pi/2})$$

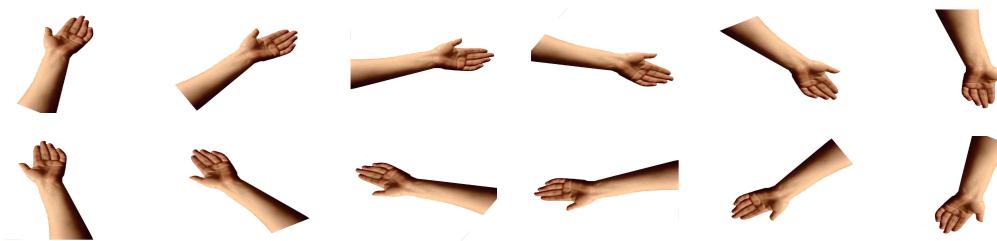


Motivation

Viewpoint mixtures in DPMs handle appearance variation discretely



Finer treatment: requires more training data, training time and test time



Idea: represent large set of mixtures by a small set of bases

Discriminative Learning of a Basis for Steering

$$S(x) = \sum_{p=1}^P \max_{x_p} \langle \mathbf{w}_p, H(x_p) \rangle + B_p(x, x_p)$$
$$\mathbf{w}_p = \sum_{i=1}^K s_{i,p} \mathbf{b}_i$$

Steering coefficients

$$L(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|_2^2 + \sum_{i=1}^N \max(0, 1 - y_i S(x_i))$$

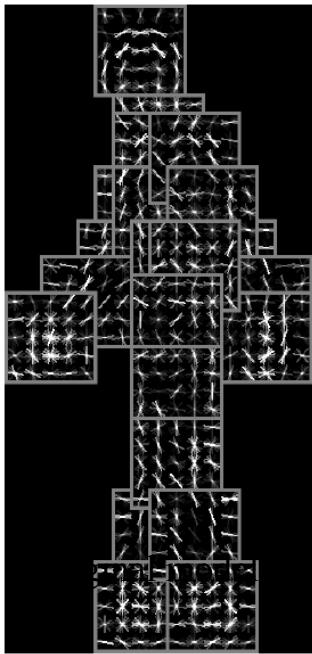
Given part positions: biconvex problem

Alternating optimization over basis and steering coefficients

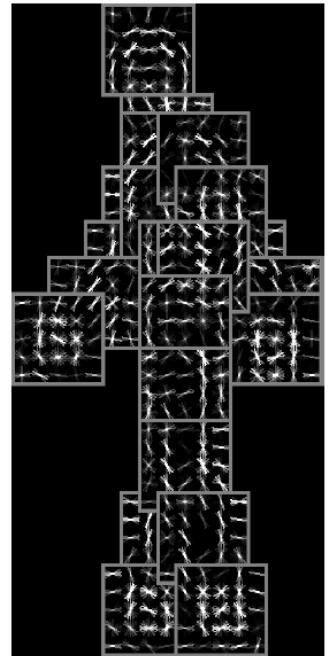
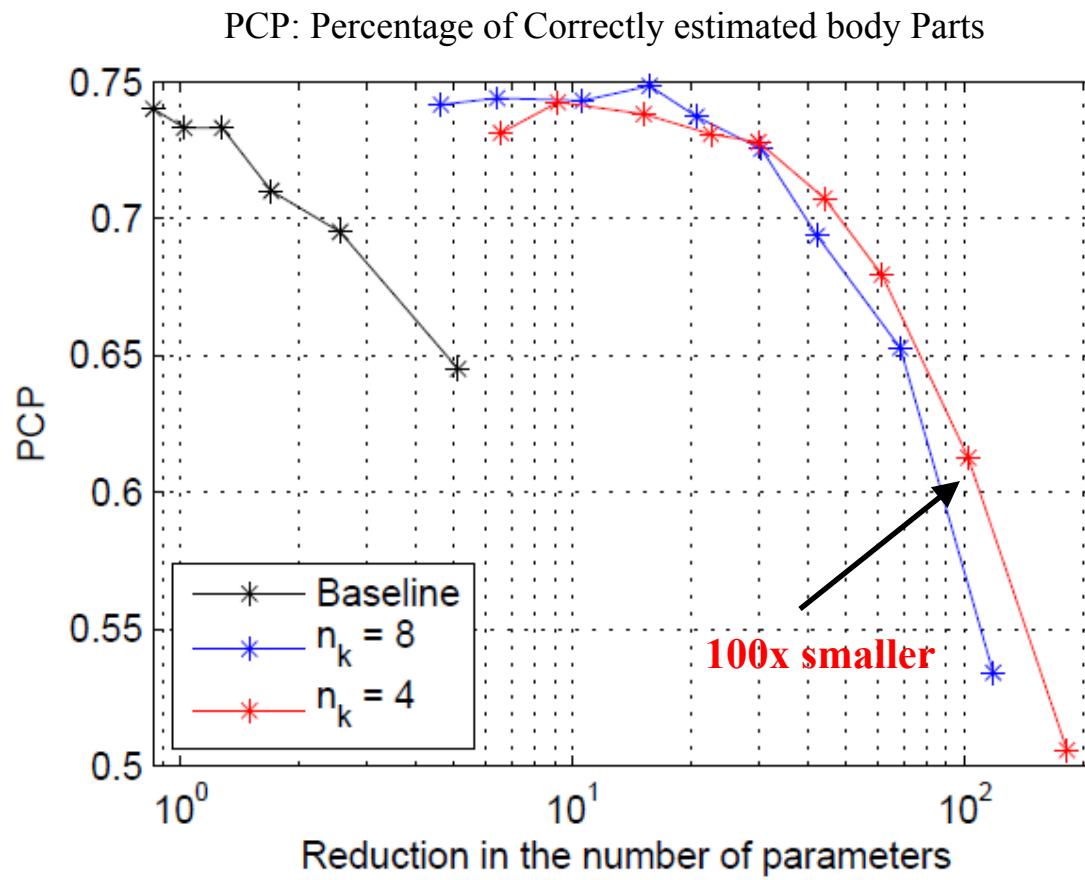
Human pose estimation

138 filters (800 dim each)

Reduction in the model size



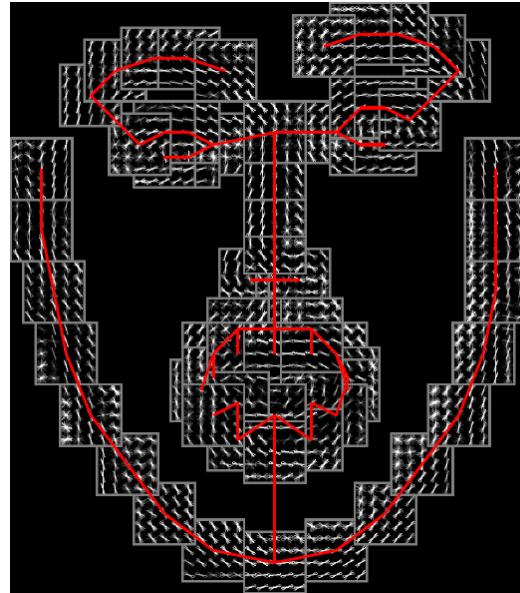
Yang, Ramanan,
CVPR' 11



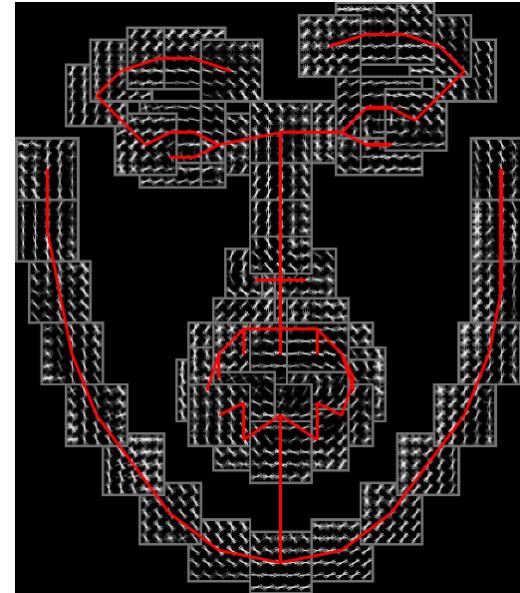
Reconstructed model
(20x smaller)

Pirsiavash & Ramanan,
“Steerable Part Models”
CVPR 2012

Face detection, pose estimation, and landmark localization



Original model

Reconstructed model
(24x smaller)

Zhu & Ramanan, CVPR' 12

Pirsiavash & Ramanan, "Steerable
Part Models" CVPR 2012