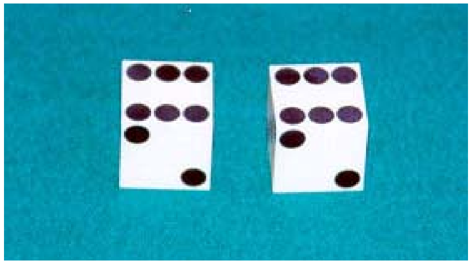
Machine Learning

Structured Models:

Hidden Markov Models versus Conditional Random Fields



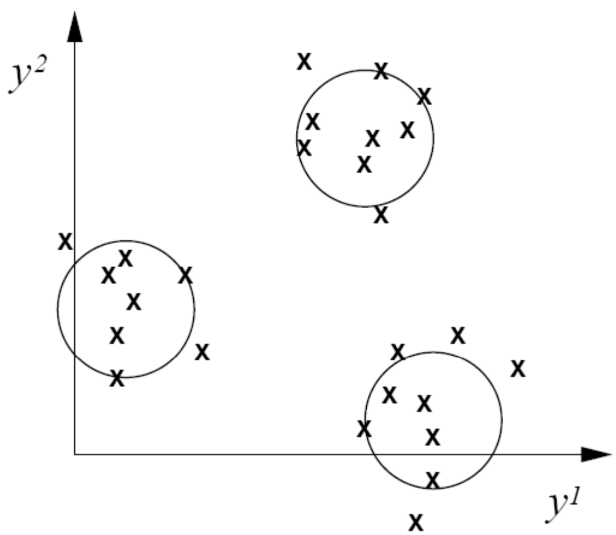
Eric Xing

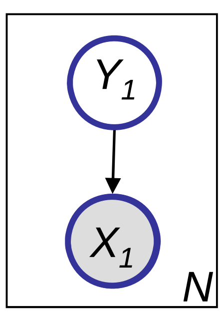
Lecture 13, August 15, 2010

Reading:

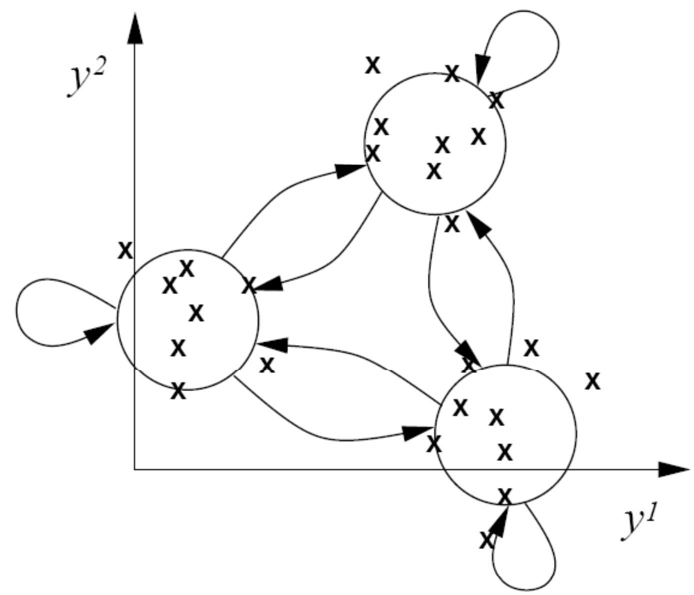
From static to dynamic mixture models

Static mixture





Dynamic mixture



The underlying source:

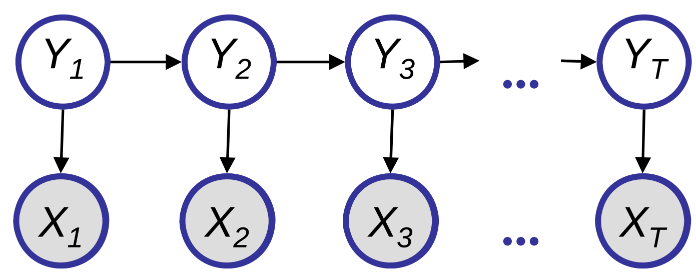
Speech signal,

dice,

The sequence:

Phonemes,

sequence of rolls,



Hidden Markov Model

- Observation space

  Alphabetic set: $\mathbb{C} = \{c_1, c_2, \cdots, c_K\}$

  Euclidean space: $\mathbb{R}^d$

- Index set of hidden states

$$\mathbb{I} = \{1, 2, \cdots, M\}$$

- Transition probabilities between any two states

$$p(y_t^j = 1 \mid y_{t-1}^i = 1) = a_{i,j},$$

or
$$p(y_t \mid y_{t-1}^i = 1) \sim \text{Multinomial}(a_{i,1}, a_{i,2}, \ldots, a_{i,M}), \forall i \in \mathbb{I}.$$

- Start probabilities

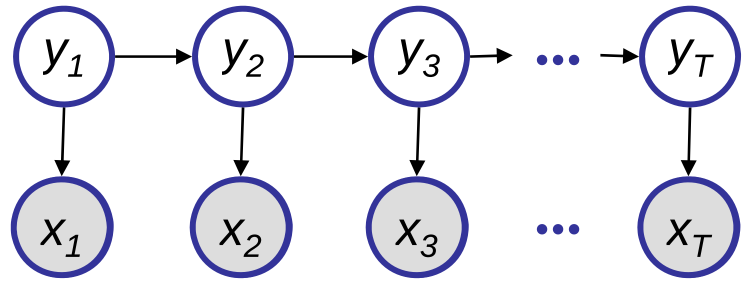
$$p(y_1) \sim \text{Multinomial}(\pi_1, \pi_2, \ldots, \pi_M).$$

- Emission probabilities associated with each state

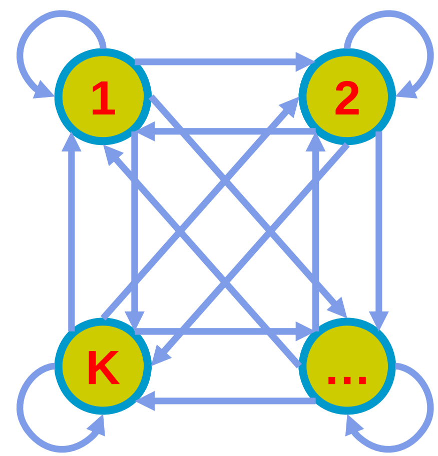
$$p(x_t \mid y_t^i = 1) \sim \text{Multinomial}(b_{i,1}, b_{i,2}, \ldots, b_{i,K}), \forall i \in \mathbb{I}.$$

or in general:

$$p(x_t \mid y_t^i = 1) \sim f(\cdot \mid \theta_i), \forall i \in \mathbb{I}.$$



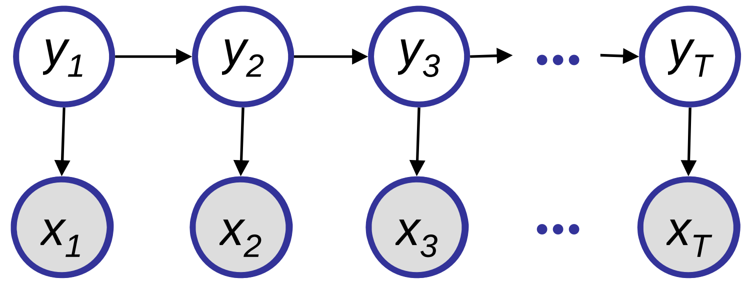
Graphical model



State automata

Probability of a Parse

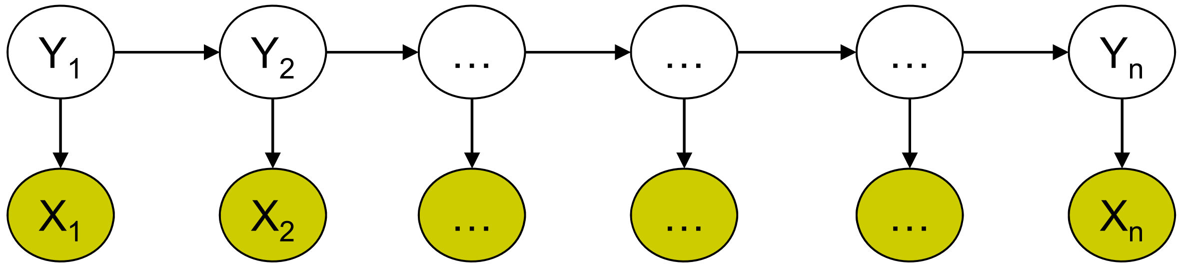
- Given a sequence $\mathbf{x} = x_1 \ldots\ldots x_T$
  and a parse $\mathbf{y} = y_1, \ldots\ldots, y_T$,
- To find how likely is the parse:
  (given our HMM and the sequence)



$$p(\mathbf{x}, \mathbf{y}) = p(x_1 \ldots\ldots x_T, y_1, \ldots\ldots, y_T) \qquad \text{(Joint probability)}$$
$$= p(y_1)\, p(x_1 \mid y_1)\, p(y_2 \mid y_1)\, p(x_2 \mid y_2) \ldots p(y_T \mid y_{T-1})\, p(x_T \mid y_T)$$
$$= p(y_1)\, P(y_2 \mid y_1) \ldots p(y_T \mid y_{T-1}) \times p(x_1 \mid y_1)\, p(x_2 \mid y_2) \ldots p(x_T \mid y_T)$$

- Marginal probability: $p(\mathbf{x}) = \sum_{\mathbf{y}} p(\mathbf{x}, \mathbf{y}) = \sum_{y_1} \sum_{y_2} \cdots \sum_{y_N} \pi_{y_1} \prod_{t=2}^{T} a_{y_{t-1}, y_t} \prod_{t=1}^{T} p(x_t \mid y_t)$
- Posterior probability: $p(\mathbf{y} \mid \mathbf{x}) = p(\mathbf{x}, \mathbf{y}) / p(\mathbf{x})$

Shortcomings of Hidden Markov Model

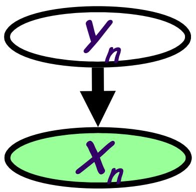


- HMM models capture dependences between each state and only its corresponding observation
  - NLP example: In a sentence segmentation task, each segmental state may depend not just on a single word (and the adjacent segmental stages), but also on the (non-local) features of the whole line such as line length, indentation, amount of white space, etc.
- Mismatch between learning objective function and prediction objective function
  - HMM learns a joint distribution of states and observations P(**Y**, **X**), but in a prediction task, we need the conditional probability P(**Y**|**X**)

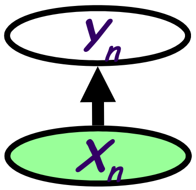
Recall Generative vs. Discriminative Classifiers

- Goal: Wish to learn f: X $\rightarrow$ Y, e.g., P(Y|X)

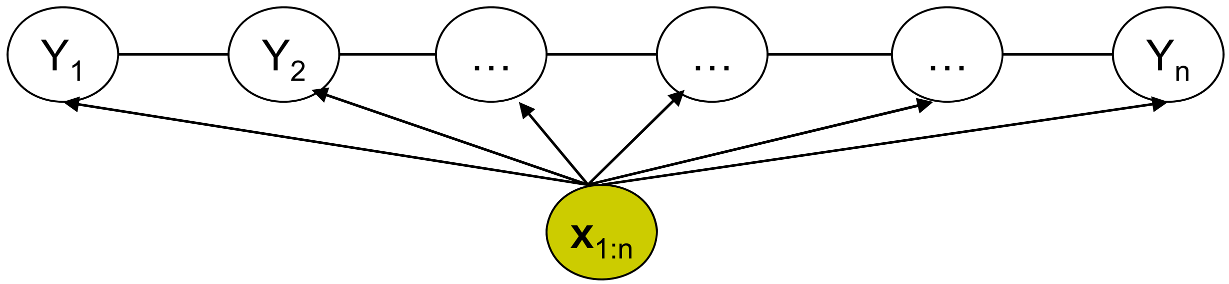
- Generative classifiers (e.g., Naïve Bayes):
  - Assume some functional form for P(X|Y), P(Y)
    This is a '***generative***' model of the data!
  - Estimate parameters of P(X|Y), P(Y) directly from training data
  - Use Bayes rule to calculate P(Y|X= x)



- Discriminative classifiers (e.g., logistic regression)
  - Directly assume some functional form for P(Y|X)
    This is a '***discriminative***' model of the data!
  - Estimate parameters of P(Y|X) directly from training data



Structured Conditional Models

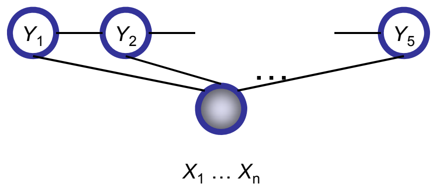


- Conditional probability P(label sequence $\mathbf{y}$ | observation sequence $\mathbf{x}$) rather than joint probability $P(\mathbf{y}, \mathbf{x})$
  - Specify the probability of possible label sequences given an observation sequence
- Allow arbitrary, non-independent features on the observation sequence $\mathbf{X}$
- The probability of a transition between labels may depend on past and future observations
- Relax strong independence assumptions in generative models

Conditional Distribution

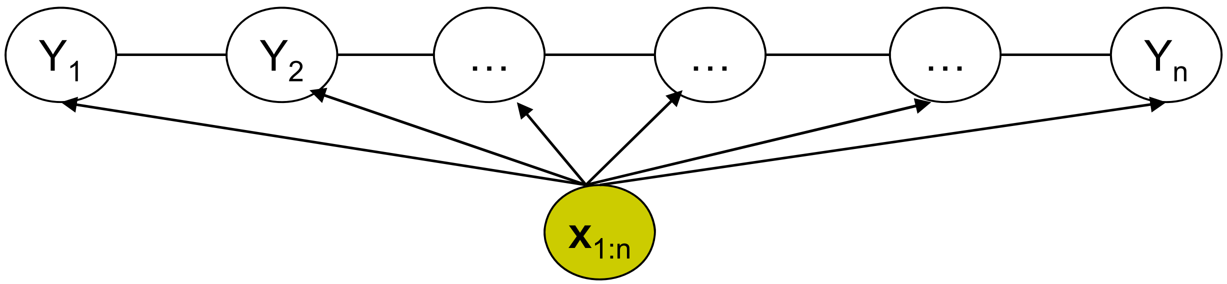
- If the graph $G = (V, E)$ of $\mathbf{Y}$ is a tree, the conditional distribution over the label sequence $\mathbf{Y} = \mathbf{y}$, given $\mathbf{X} = \mathbf{x}$, by the Hammersley Clifford theorem of random fields is:

$$p_\theta(\mathrm{y} \mid \mathrm{x}) \propto \exp\left(\sum_{e \in E, k} \lambda_k f_k(e, \mathrm{y}|_e, \mathrm{x}) + \sum_{v \in V, k} \mu_k g_k(v, \mathrm{y}|_v, \mathrm{x}) \right)$$



  - x is a data sequence
  - y is a label sequence
  - *v* is a vertex from vertex set V = set of label random variables
  - *e* is an edge from edge set E over V
  - f_k and g_k are given and fixed. g_k is a Boolean vertex feature; f_k is a Boolean edge feature
  - *k* is the number of features
  - $\theta = (\lambda_1, \lambda_2, \cdots, \lambda_n; \mu_1, \mu_2, \cdots, \mu_n); \lambda_k$ and μ_k are parameters to be estimated
  - $\mathrm{y}|_e$ is the set of components of y defined by edge *e*
  - $\mathrm{y}|_v$ is the set of components of y defined by vertex *v*

Conditional Random Fields

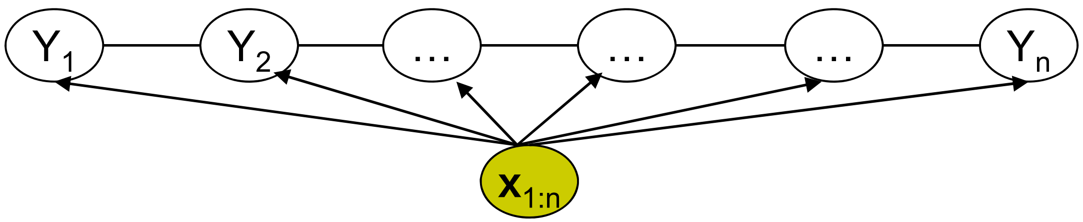


$$P(\mathbf{y}_{1:n}|\mathbf{x}_{1:n}) = \frac{1}{Z(\mathbf{x}_{1:n})}\prod_{i=1}^{n}\phi(y_i, y_{i-1}, \mathbf{x}_{1:n}) = \frac{1}{Z(\mathbf{x}_{1:n}, \mathbf{w})}\prod_{i=1}^{n}\exp(\mathbf{w}^T\mathbf{f}(y_i, y_{i-1}, \mathbf{x}_{1:n}))$$

- CRF is a partially directed model
  - Discriminative model
  - Usage of global normalizer Z(**x**)
  - Models the dependence between each state and the entire observation sequence

Conditional Random Fields

- General parametric form:

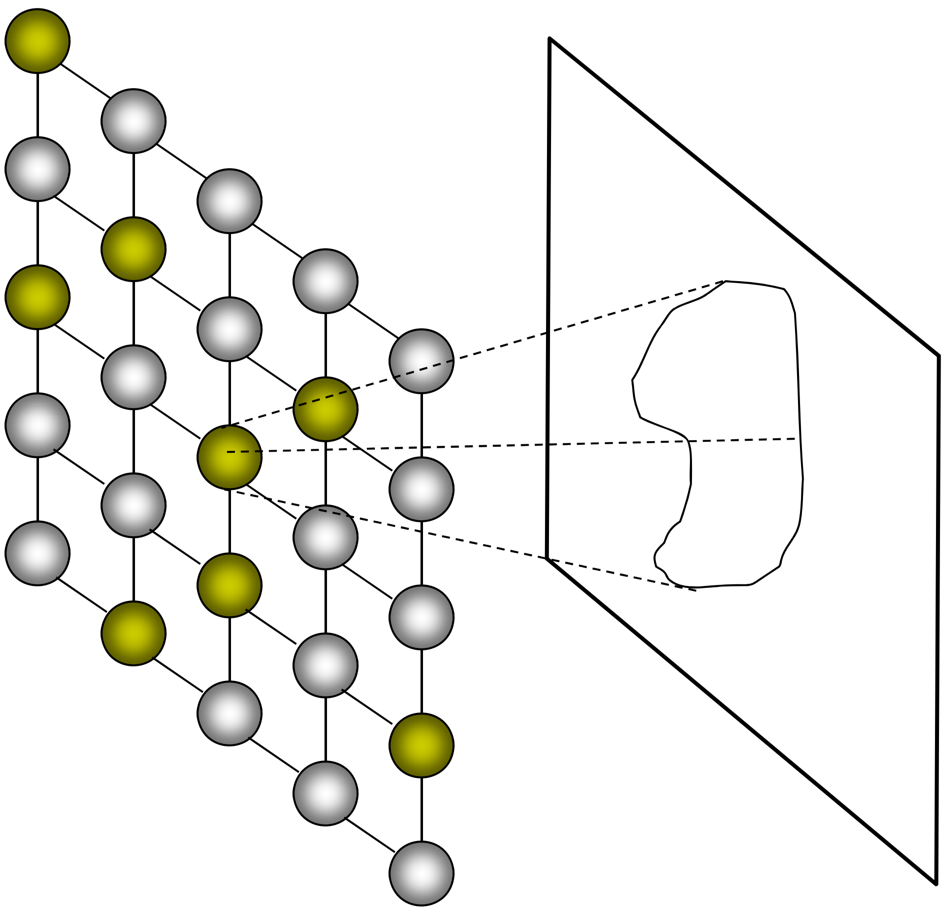


$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x},\lambda,\mu)}\exp\Big(\sum_{i=1}^{n}\big(\sum_{k}\lambda_k f_k(y_i,y_{i-1},\mathbf{x}) + \sum_{l}\mu_l g_l(y_i,\mathbf{x})\big)\Big)$$

$$= \frac{1}{Z(\mathbf{x},\lambda,\mu)}\exp\Big(\sum_{i=1}^{n}\big(\lambda^T\mathbf{f}(y_i,y_{i-1},\mathbf{x}) + \mu^T\mathbf{g}(y_i,\mathbf{x})\big)\Big)$$

$$\text{where } Z(\mathbf{x},\lambda,\mu) = \sum_{\mathbf{y}}\exp\Big(\sum_{i=1}^{n}\big(\lambda^T\mathbf{f}(y_i,y_{i-1},\mathbf{x}) + \mu^T\mathbf{g}(y_i,\mathbf{x})\big)\Big)$$

Conditional Random Fields



$$p_\theta(y \mid x) = \frac{1}{Z(\theta, x)} \exp\left\{\sum_c \theta_c f_c(x, y_c)\right\}$$

- Allow arbitrary dependencies on input
- Clique dependencies on labels
- Use approximate inference for general graphs

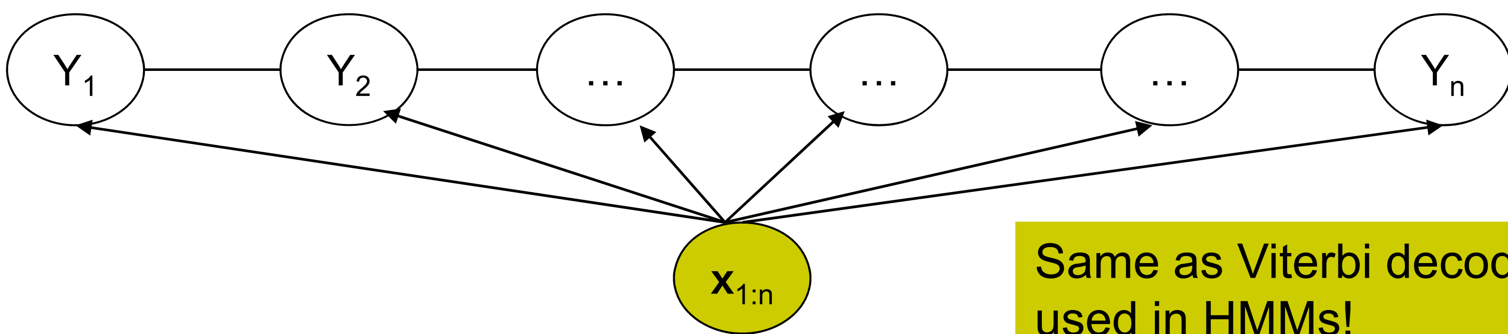
CRFs: Inference

- Given CRF parameters λ and μ, find the $\mathbf{y}^*$ that maximizes P(**y**|**x**)

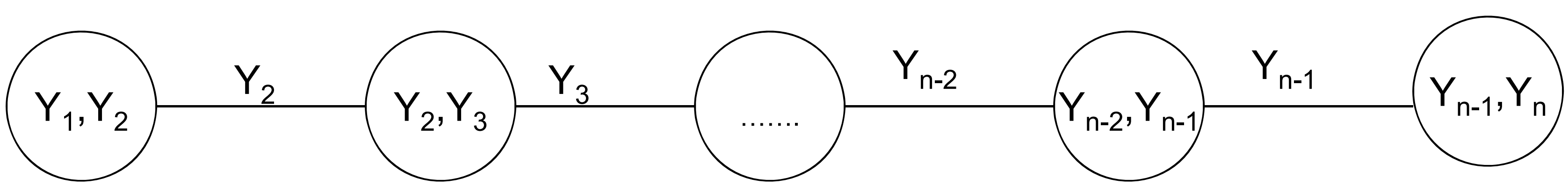
$$\mathbf{y}^* \quad = \quad \arg\max_{\mathbf{y}} \exp(\sum_{i=1}^{n}(\lambda^T \mathbf{f}(y_i, y_{i-1}, \mathbf{x}) + \mu^T \mathbf{g}(y_i, \mathbf{x})))$$

  - Can ignore Z(**x**) because it is not a function of **y**

- Run the max-product algorithm on the junction-tree of CRF:



Same as Viterbi decoding used in HMMs!



CRF learning

- Given $\{(\mathbf{x}_d, \mathbf{y}_d)\}_{d=1}^N$, find λ^*, μ^* such that

$$
\begin{aligned}
\lambda*, \mu* &= \arg\max_{\lambda,\mu} L(\lambda, \mu) = \arg\max_{\lambda,\mu} \prod_{d=1}^{N} P(\mathbf{y}_d | \mathbf{x}_d, \lambda, \mu) \\
&= \arg\max_{\lambda,\mu} \prod_{d=1}^{N} \frac{1}{Z(\mathbf{x}_d, \lambda, \mu)} \exp(\sum_{i=1}^{n} (\lambda^T \mathbf{f}(y_{d,i}, y_{d,i-1}, \mathbf{x}_d) + \mu^T \mathbf{g}(y_{d,i}, \mathbf{x}_d))) \\
&= \arg\max_{\lambda,\mu} \sum_{d=1}^{N} (\sum_{i=1}^{n} (\lambda^T \mathbf{f}(y_{d,i}, y_{d,i-1}, \mathbf{x}_d) + \mu^T \mathbf{g}(y_{d,i}, \mathbf{x}_d)) - \log Z(\mathbf{x}_d, \lambda, \mu))
\end{aligned}
$$

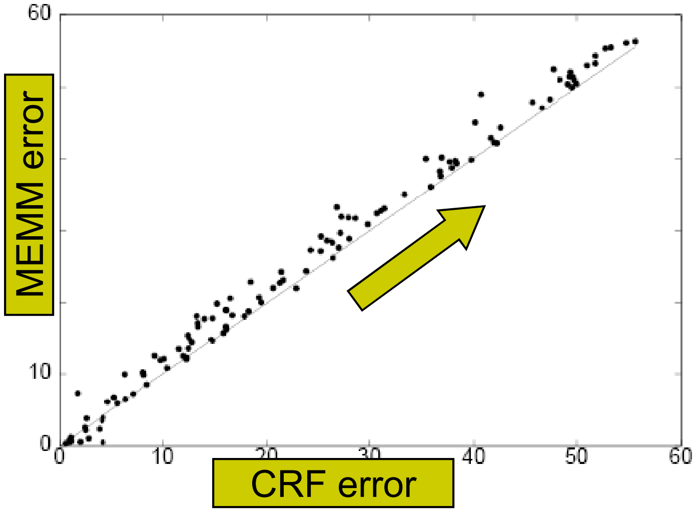
- Computing the gradient w.r.t λ:

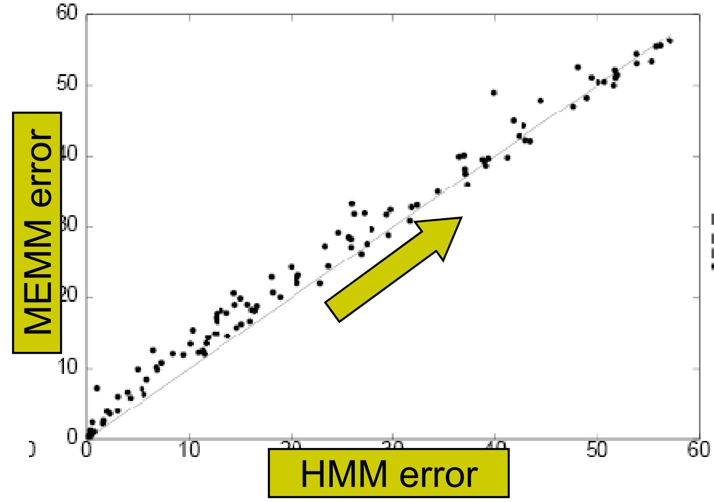
Gradient of the log-partition function in an exponential family is the expectation of the sufficient statistics.

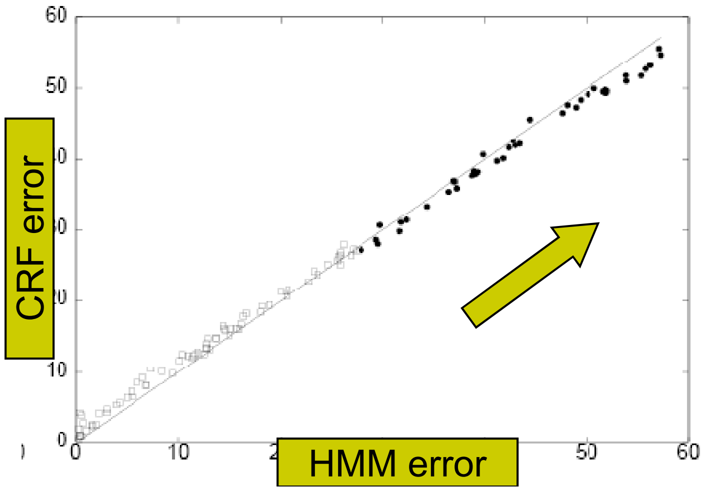
$$
\nabla_\lambda L(\lambda, \mu) = \sum_{d=1}^{N} (\sum_{i=1}^{n} \mathbf{f}(y_{d,i}, y_{d,i-1}, \mathbf{x}_d) - \sum_{\mathbf{y}} (P(\mathbf{y} | \mathbf{x_d}) \sum_{i=1}^{n} \mathbf{f}(y_{d,i}, y_{d,i-1}, \mathbf{x}_d)))
$$

CRFs: some empirical results

- Comparison of error rates on synthetic data







Data is increasingly higher order in the direction of arrow

CRFs achieve the lowest error rate for higher order data

CRFs: some empirical results

- Parts of Speech tagging

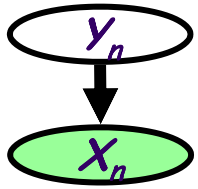
model	*error*	*oov error*
HMM	5.69%	45.99%
MEMM	6.37%	54.61%
CRF	5.55%	48.05%
MEMM^+	4.81%	26.99%
CRF^+	4.27%	23.76%

$^+$Using spelling features

- Using same set of features: HMM >=< CRF > MEMM
- Using additional overlapping features: CRF^+ > MEMM^+ >> HMM

Summary

- Conditional Random Fields is a discriminative Structured Input Output model!

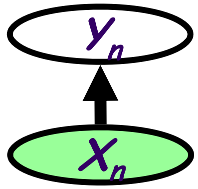


- HMM is a generative structured I/O model

- Complementary strength and weakness:

  1.
  2.
  3.

  …



Condition Random Fields: applications in vision

L. Fei-Fei
Computer Science Dept.
Stanford University

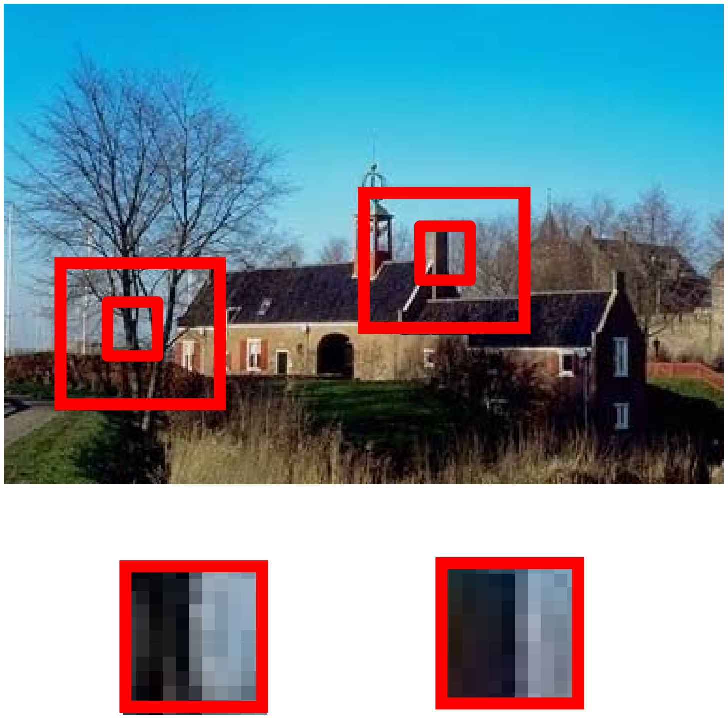
Machine learning in computer vision

- Aug 15, Lecture 13: Conditional Random Field
  - Applications in computer vision
  - Image labeling, segmentation, object recognition & image annotation

Image labeling
(slides courtesy to Sanjiv Kumar (Google))

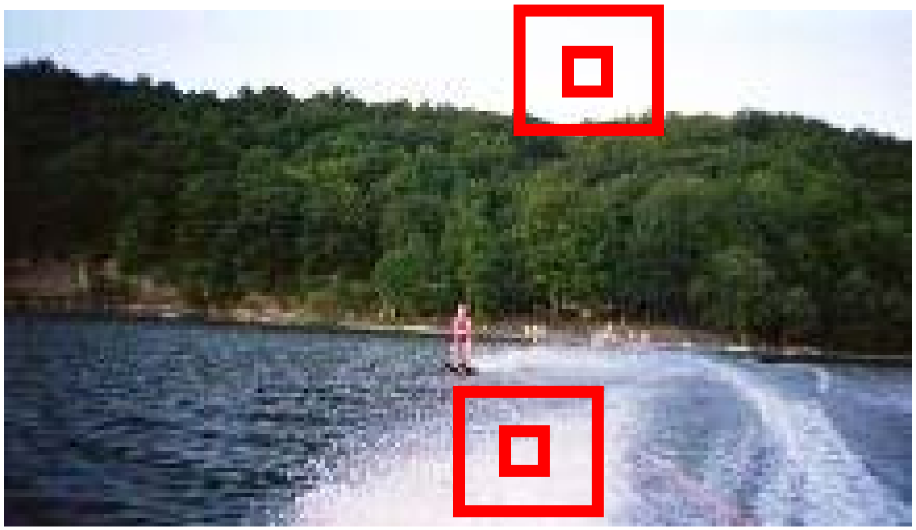
Reference:
Sanjiv Kumar & Martial Hebert, Discriminative random fields: a discriminative framework for contextual interaction in classification. ICCV, 2003

Context helps visual recognition



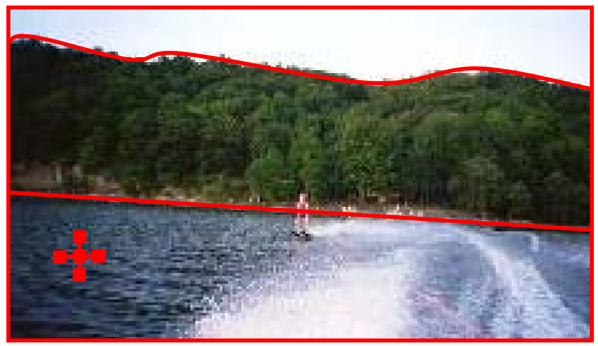
Context from Larger Neighborhoods !

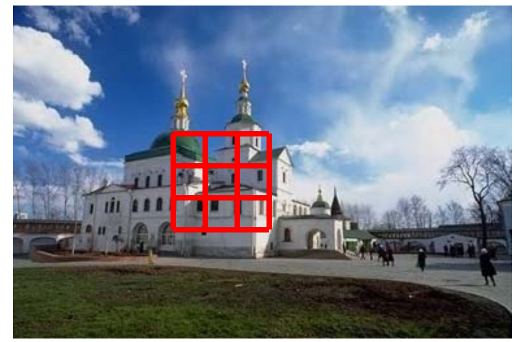
Context helps visual recognition

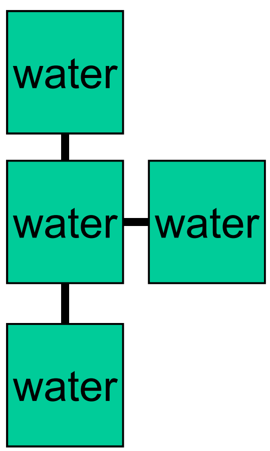


Context from Whole Image !

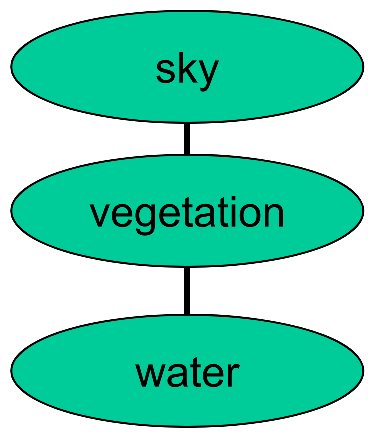
Contextual Interactions



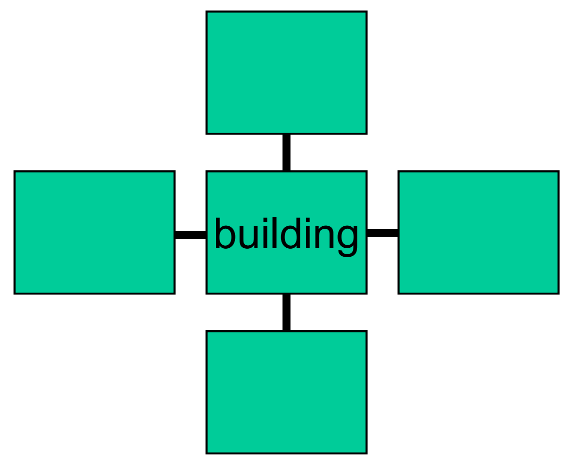




Pixel-Pixel

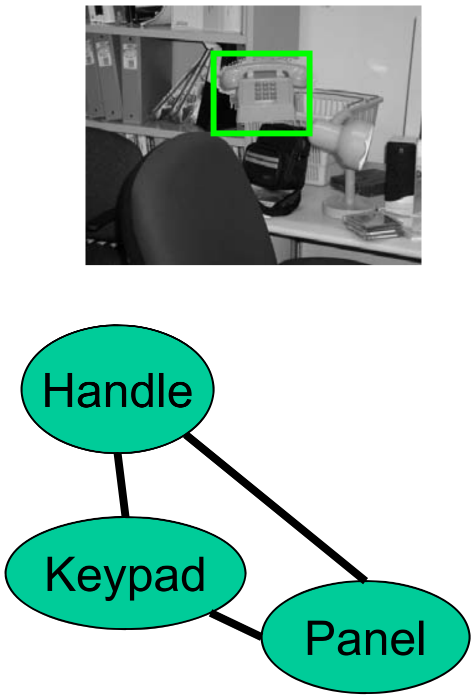


Region-Region

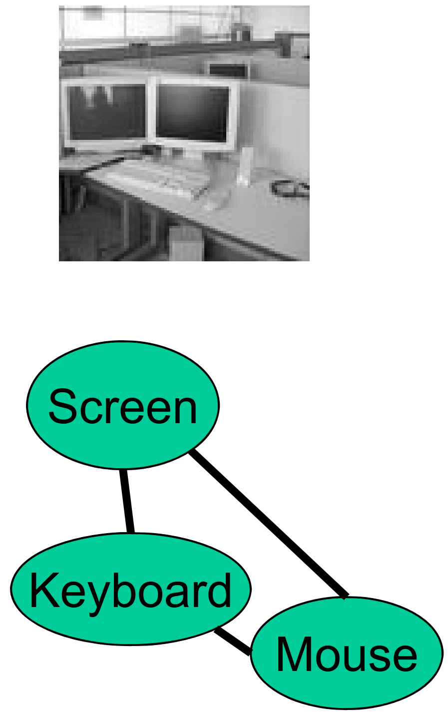


Block-Block

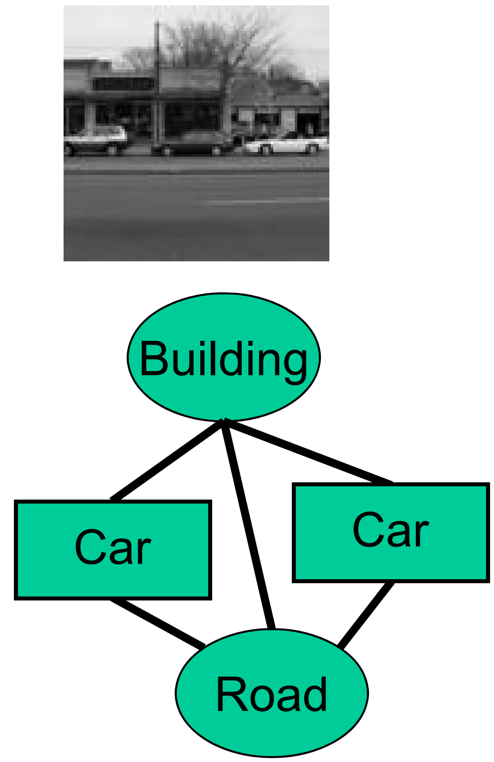
Contextual Interactions



Part-Part



Object-Object



Region-Object

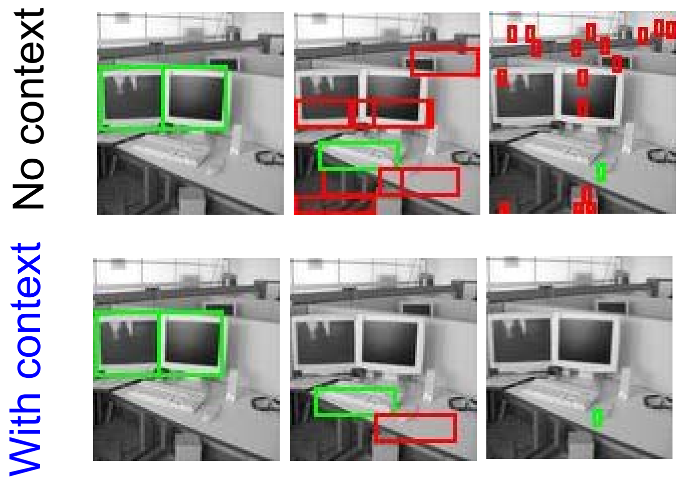
Modeling Contextual Interactions

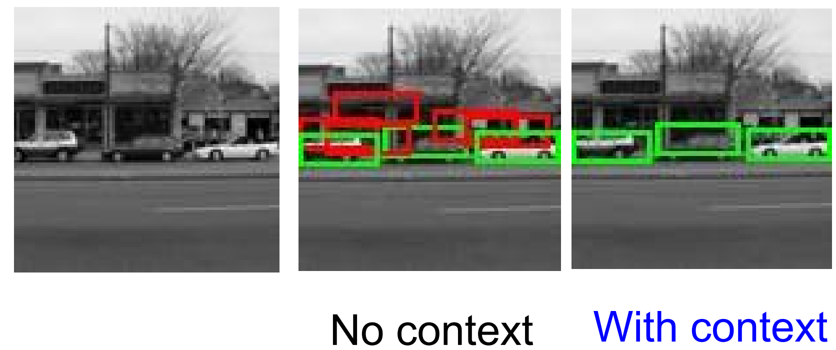
- **Framework to learn all relevant contextual interactions in a single model automatically from training data**

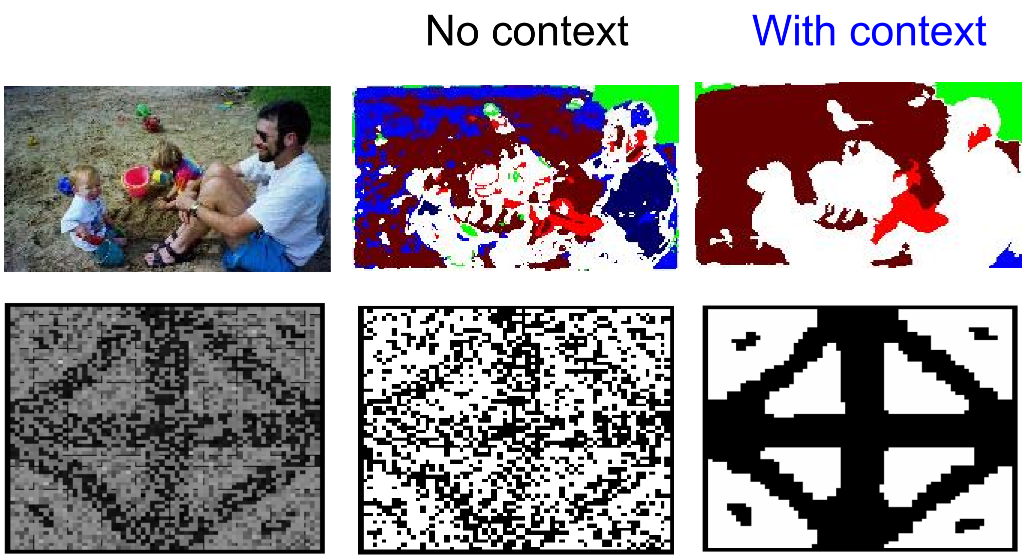
- **Probabilistic models**
  - **Principled way to deal with ambiguities**

- **Graphical models**
  - **Powerful framework for ensuring global consistency using relatively local constraints**

Undirected graphs = Random Fields


No context
With context


No context
With context


No context
With context

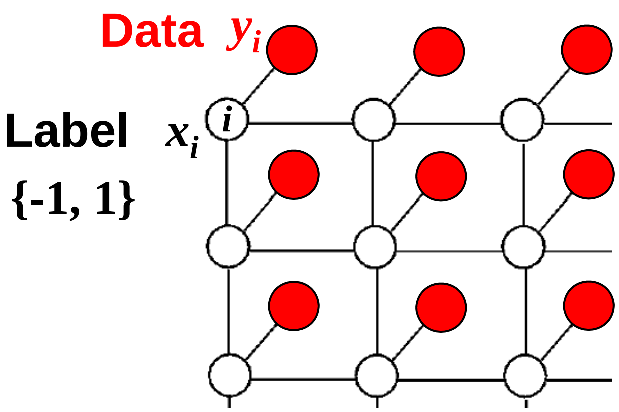

No context
With context

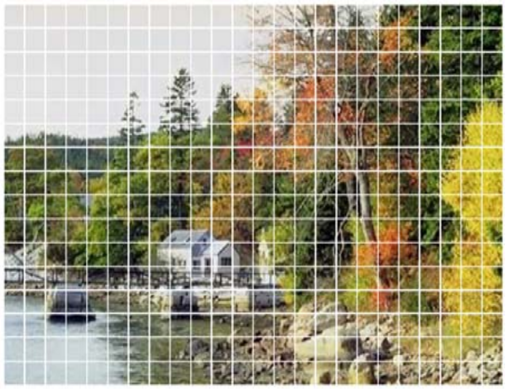
[Kumar ’05]

Image Labeling Outline

- **Background**
  - **Markov Random Fields (MRFs)**
- **Conditional Random Fields (CRFs)**
- **Multiclass and Hierarchical Interactions**
- **Hidden Conditional Random Fields**
- **Semi-supervised Learning**
- **Open Issues**

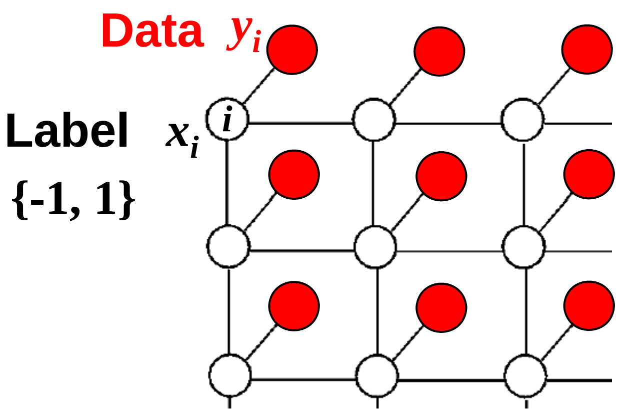
Markov Random Field (MRF)

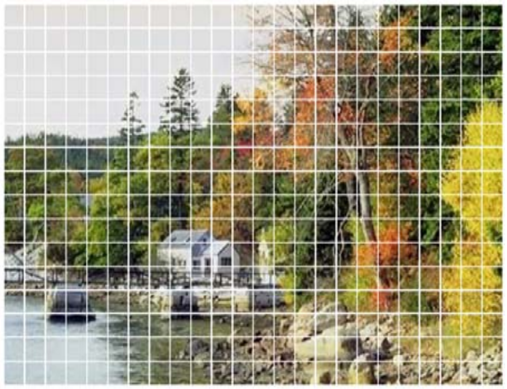




[Geman & Geman ’84]

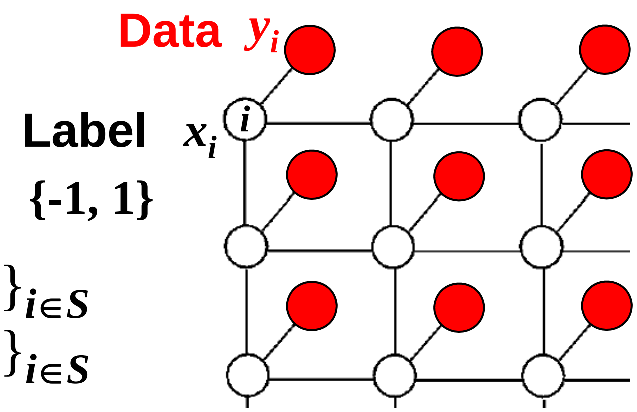
Markov Random Field (MRF)





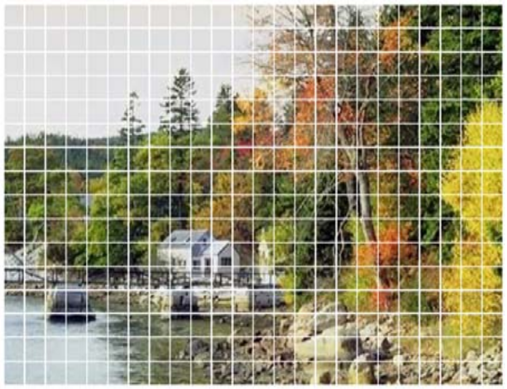
[Geman & Geman ’84]

Markov Random Field (MRF)



$x = \{x_i\}_{i \in S}$

$y = \{y_i\}_{i \in S}$



Generative Framework

$$\underbrace{P(x|y)}_{\text{We want}} \propto P(x,y) = \underbrace{P(y|x)}_{\text{Observation Model}} \underbrace{P(x)}_{\text{Prior Model (MRF)}}$$

[Geman & Geman '84]

Markov Random Field (MRF)

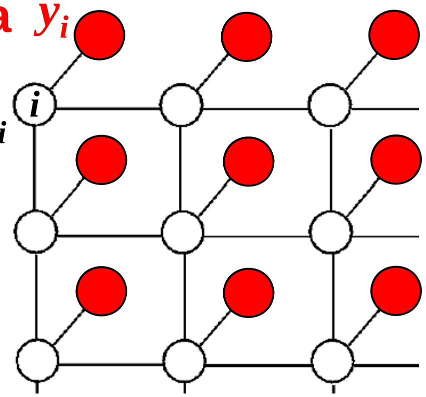
Data y_i

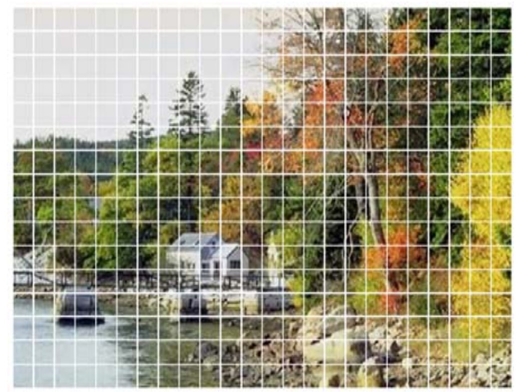
Label x_i

{-1, 1}

$x = \{x_i\}_{i \in S}$

$y = \{y_i\}_{i \in S}$





Generative Framework

$$\underbrace{P(x|y)}_{\text{We want}} \propto P(x, y) = \underbrace{P(y|x)}_{\text{Observation Model}} \underbrace{P(x)}_{\text{Prior Model (MRF)}}$$

$$P(y|x) \approx \prod_{i \in S} P(y_i|x_i)$$

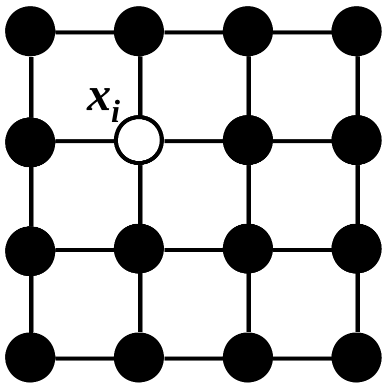
Too restrictive !

[Geman & Geman '84]

Markov Random Field (MRF)

Label Prior $P(x)$ is modeled as MRF

Markovianity $P(x_i | x_{S\setminus\{i\}}) = P(x_i | x_{N_i})$



Markov Random Field (MRF)

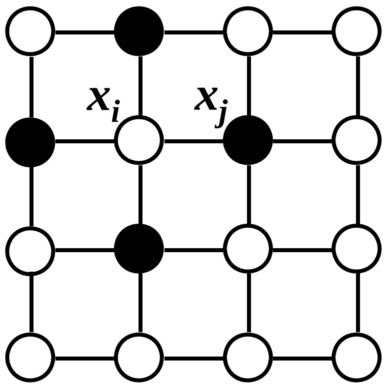
Label Prior $P(x)$ is modeled as MRF

Markovianity $P(x_i \mid x_{S\setminus\{i\}}) = P(x_i \mid x_{N_i})$

Positivity $P(x) > 0 \quad \forall\, x$

$$P(x) \propto \prod_{e\in E} \Psi_e(x_i, x_j)$$

[Hammersley & Clifford '71]

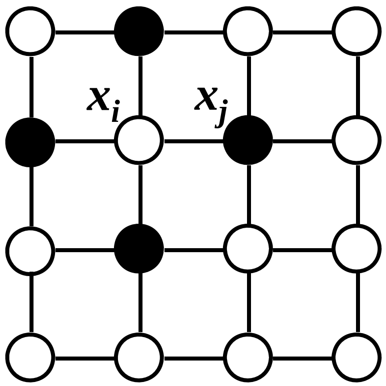


Markov Random Field (MRF)

Label Prior $P(x)$ is modeled as MRF

Markovianity $P(x_i | x_{S\backslash\{i\}}) = P(x_i | x_{N_i})$

Positivity $P(x) > 0 \quad \forall\, x$



$$P(x) \propto \prod_{e \in E} \Psi_e(x_i, x_j)$$

[Hammersley & Clifford '71]

Ising Model $P(x) \propto \exp(\sum_{i \in S} \sum_{j \in N_i} \beta\, x_i x_j) \quad \beta > 0$

[Ising '25]

Markov Random Field (MRF)

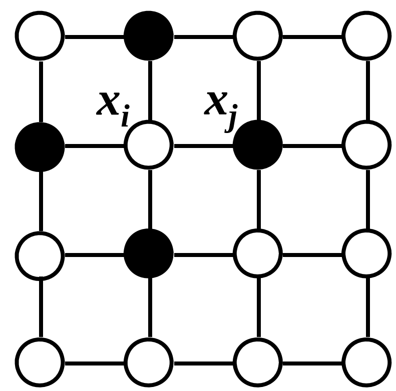
Label Prior $P(x)$ is modeled as MRF

Markovianity $P(x_i | x_{S\backslash\{i\}}) = P(x_i | x_{N_i})$

Positivity $P(x) > 0 \quad \forall\, x$

$$P(x) \propto \prod_{e \in E} \Psi_e(x_i, x_j)$$

[Hammersley & Clifford '71]



{-1, 1}

Ising Model $P(x) \propto \exp(\sum_{i \in S} \sum_{j \in N_i} \beta\, x_i x_j) \quad \beta > 0$

[Ising '25]

Traditional MRF

$$P(x|y) = \frac{1}{Z} \exp\left(\sum_{i \in S} \log P(y_i | x_i) + \sum_{i \in S} \sum_{j \in N_i} \beta\, x_i x_j \right)$$

Generative vs. Discriminative

We want : $P(x|y)$

- **Generative Framework**
  - **Models** $P(x,y)$ **to get** $P(x|y)$
  - **Implicit modeling of observations**
- **Discriminative Framework**
  - **Models conditional distribution** $P(x|y)$ **directly**

Conditional Random Field (CRF)

[Lafferty et al. '01]

- **Conditional distribution $P(x|y)$ is modeled as MRF**

$$P(x_i | x_{S\backslash\{i\}}, y) = P(x_i | x_{N_i}, y)$$

$$P(x|y) > 0 \quad \forall\, x$$

$$P(x|y) \propto \prod_{e \in E} \Psi_e(x_i, x_j, y)$$

[Hammersley & Clifford '71]

- Segmentation and labeling of 1D text sequences

Conditional Random Field (CRF)

- **Graphs on images with loops**
  - **Grid or irregular topology**
- **Potentials using arbitrary discriminative classifiers**
  - **Discriminative Random Field (DRF)**

[Kumar & Hebert, ICCV '03]

Conditional Random Field (CRF)

- Graphs on images with loops
  - Grid or irregular topology
- Potentials using arbitrary discriminative classifiers
- If only unary and pairwise potentials are non-zero

$$\boldsymbol{P}(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{\boldsymbol{Z}}\exp\left(\underbrace{\sum_{i \in S} \boldsymbol{A}_i(\boldsymbol{x}_i, \boldsymbol{y})}_{\text{Association Potential}} + \sum_{i \in S}\sum_{j \in N_i} \underbrace{\boldsymbol{I}_{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{y})}_{\text{Interaction Potential}}\right)$$

[Kumar & Hebert, ICCV'03]

Comparison with MRF framework

Traditional MRF

$$P(x|y) = \frac{1}{Z}\exp\left(\sum_{i\in S}\log P(y_i|x_i) + \sum_{i\in S}\sum_{j\in N_i}\beta\, x_i x_j\right)$$

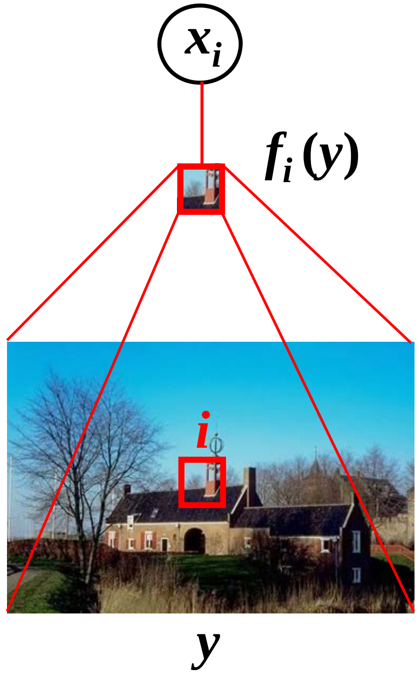
CRF

$$P(x|y) = \frac{1}{Z}\exp\left(\sum_{i\in S}A(x_i, y) + \sum_{i\in S}\sum_{j\in N_i}I(x_i, x_j, y)\right)$$

Data from multiple sites

Data-dependent label interactions

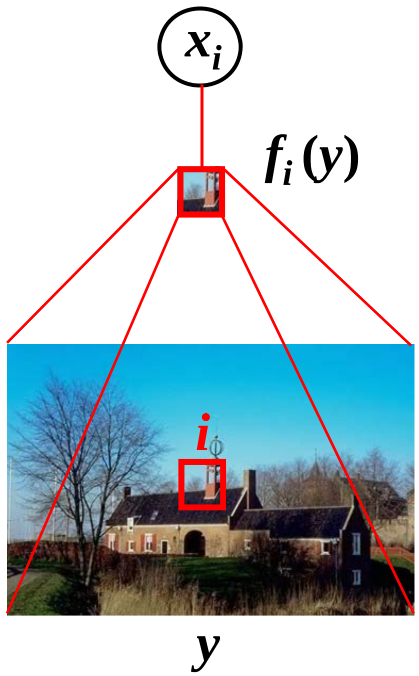
Association Potential $A(x_i, y)$



Association Potential $A(x_i, y)$

discriminative classifier

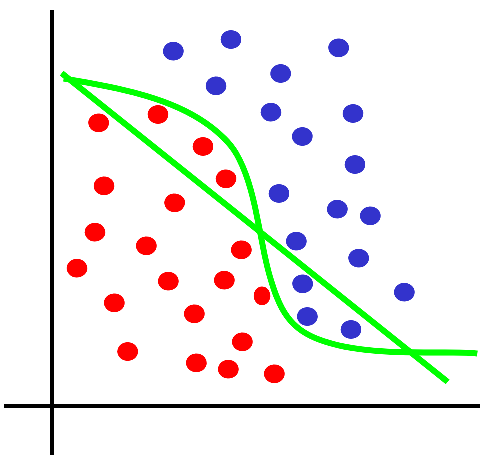
$$A(x_i, y) = \log P(x_i | f_i(y))$$

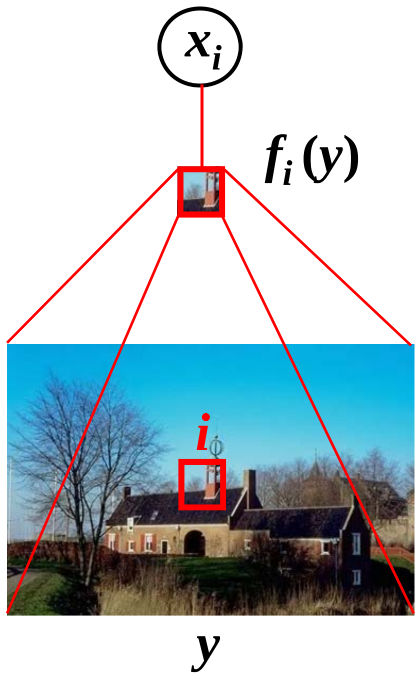


Association Potential $A(x_i, y)$

discriminative classifier

$$A(x_i, y) = \log P(x_i \mid f_i(y))$$
$$= \log \sigma(x_i w^{\mathrm{T}} f_i(y))$$





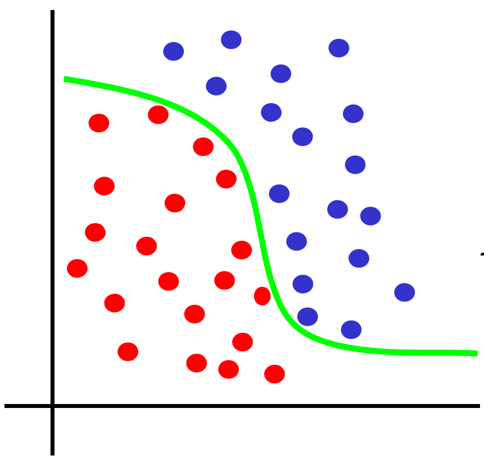
Other classifier choices: [Szummer et al. '04] [He et al. '04][Torralba et al. '05]

Association Potential $A(x_i, y)$

discriminative classifier

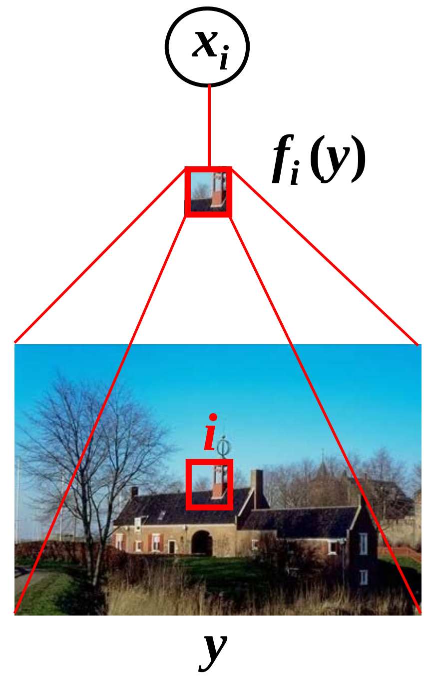
$$A(x_i, y) = \log P(x_i | f_i(y))$$

$$= \log \sigma(x_i w^{\mathrm{T}} f_i(y))$$



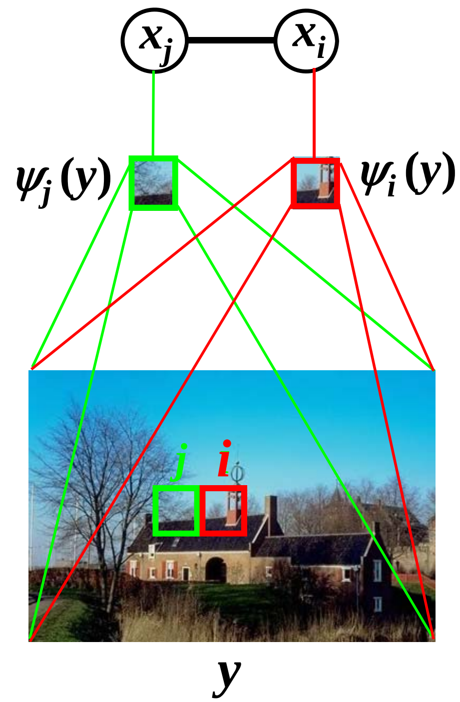
$$f_i(y) \xrightarrow{\phi(.)} \phi_i(y)$$

$$A(x_i, y) = \log \sigma(x_i w^{\mathrm{T}} \phi_i(y))$$



Other classifier choices: **[Szummer et al. '04] [He et al. '04][Torralba et al. '05]**

Interaction Potential $I(x_i, x_j, y)$

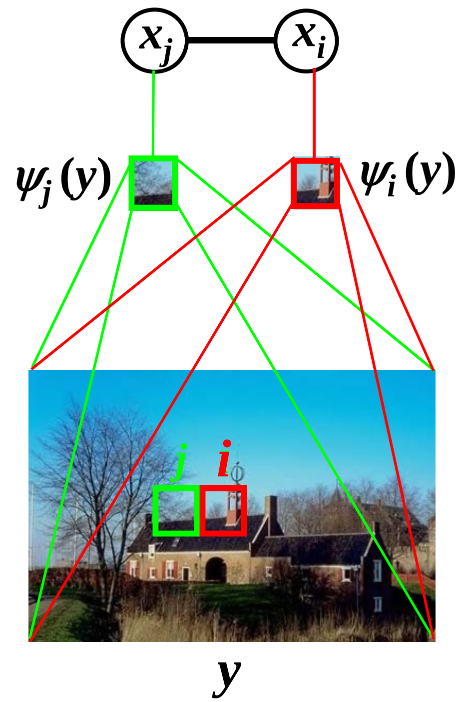


[Kumar & Hebert, NIPS '04]

Interaction Potential $I(x_i, x_j, y)$

pairwise discriminative classifier

$$I(x_i, x_j, y) = \log P(x_i, x_j | \psi_i, \psi_j)$$
$$= x_i x_j \underbrace{v^T \mu_{ij}(y)}_{\beta}$$



[Kumar & Hebert, NIPS '04]

Learning and Inference

Given input image y and $P(x|y)$, get the optimal labels x

Inference Methods

Parameter Learning Methods	MAP (min-cut)	MPM (BP)
SPA (min-cut)	5.82	19.19
MMA (BP)	26.53	5.70
Contrastive Divergence	8.88	6.29
Pseudo-Likelihood	17.69	7.31

Pixelwise error (%) on 200 test images

Learning	⟷ coupling	Inference
SPA (min-cut)	⟷	MAP (min-cut)
MMA (BP)	⟷	MPM (BP)

[Kumar et al. EMMCVPR '05]

Man-Made Structure Detection

- Training Set – 130 images (256x384 pixels)
- Test Set – 108 images

Scale variations | **Illumination variations** | **Pose variations** | **Non-linear structures** | **Negative samples**

- Gradient magnitude and orientation features

14 dim ⟹ (quadratic kernel) 119 dim

[Kumar & Hebert CVPR ‘03]

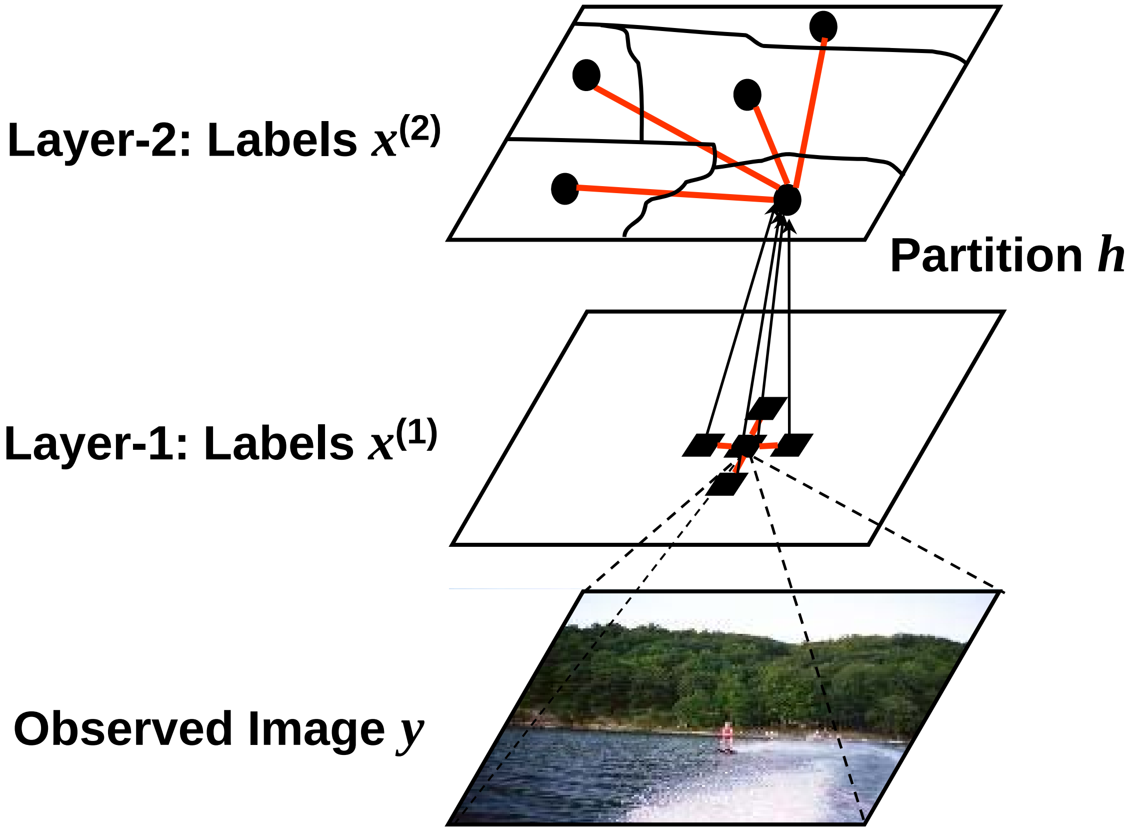
Traditional MRF | **CRF**

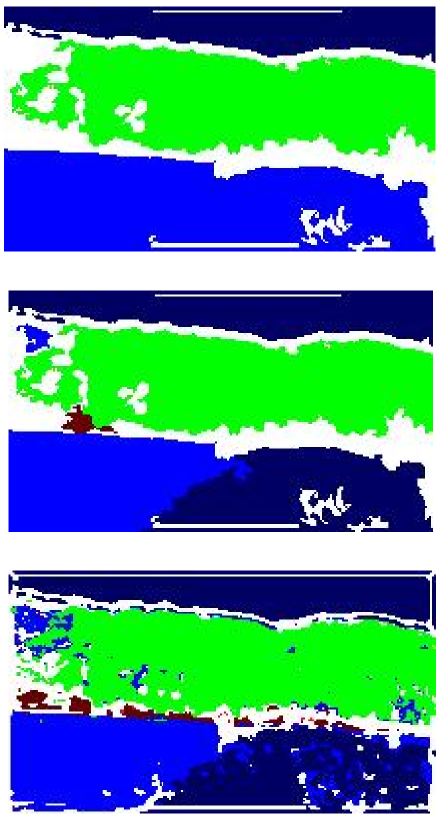
[Kumar & Hebert ICCV ‘03]

Real-time Structure Detection

[Collins et al. '04]

Hierarchical Interactions

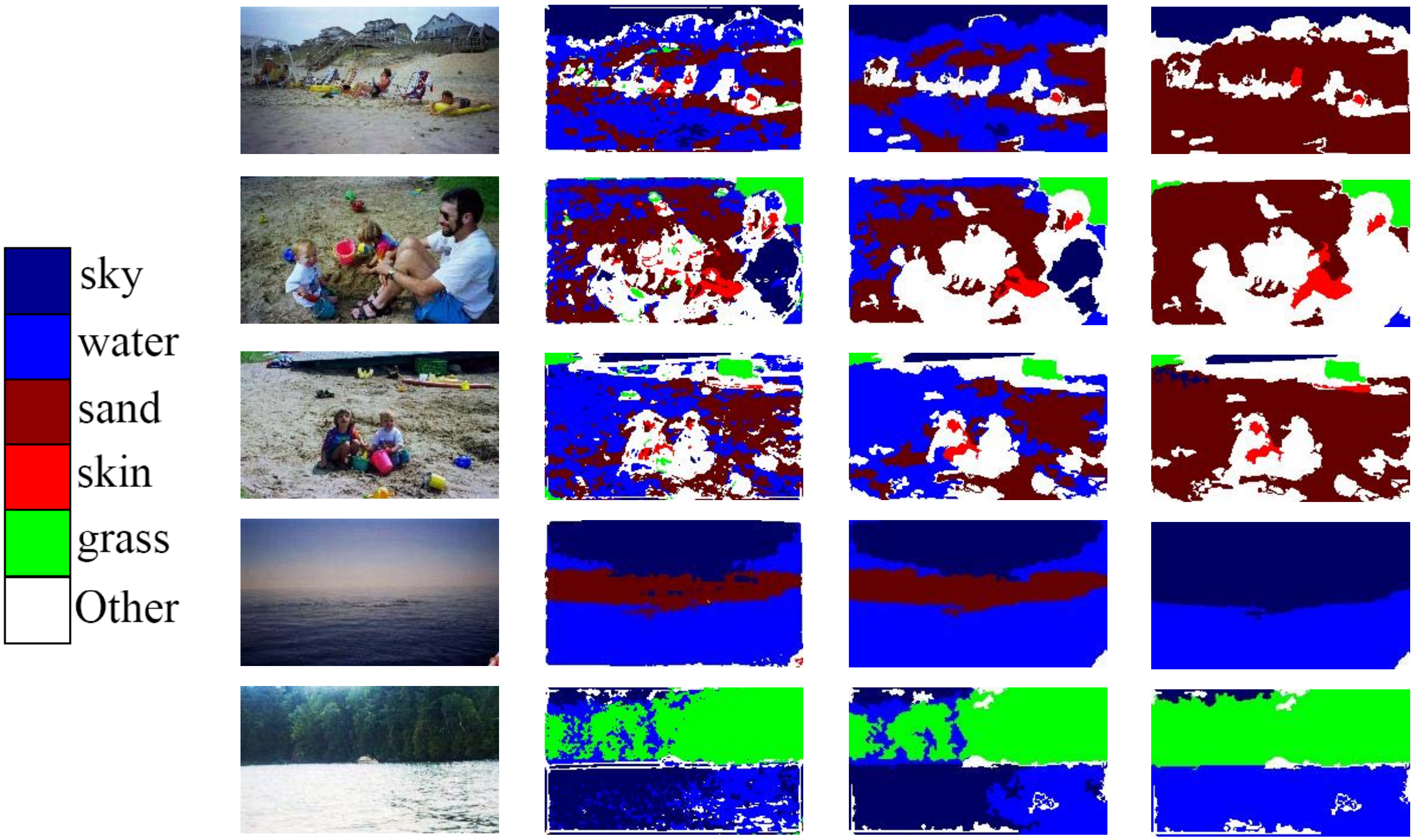




Classification

[Kumar & Hebert, ICCV '05]

Region Classification



Input image

Softmax (no context)
62.3 %

Layer-1 (label smoothing only)
63.8 %

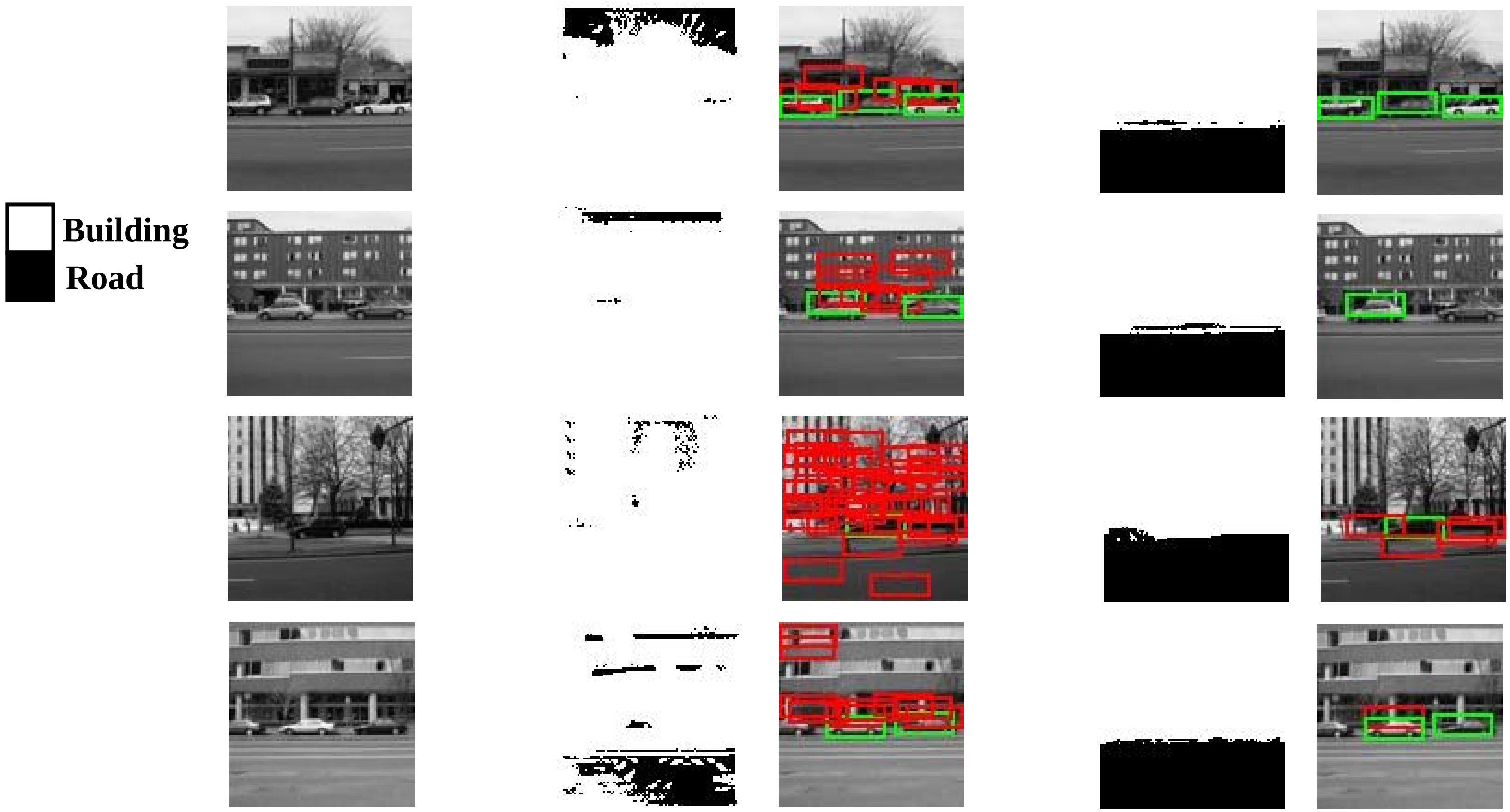
Layer-2 (full model)
74 % (~ 2 Sec)

[Kumar & Hebert ICCV '05]

Object-Region Interactions

MIT Dataset

[Torralba et al., '05]



Input image

No Context

	Building/Road	Car
	70 %	62 %

With Context

	Building/Road	Car
	98 %	80 %

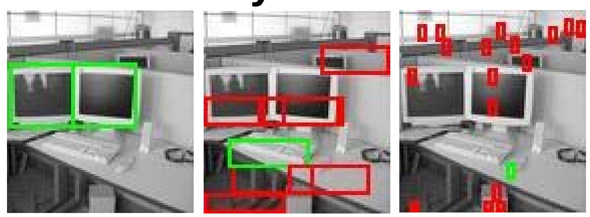
[Kumar & Hebert ICCV '05]

Object-Object Interactions

- **MIT context database: 164 images (100x100 pixels)**
- **Very small objects (8x5 pixels) → High false positives**
- **Initial object detectors trained with gentle boosting**

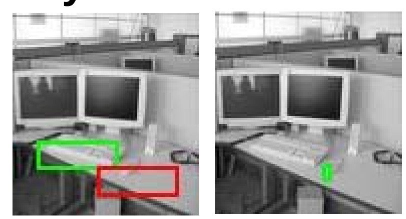
Input image

Monitor Keyboard Mouse



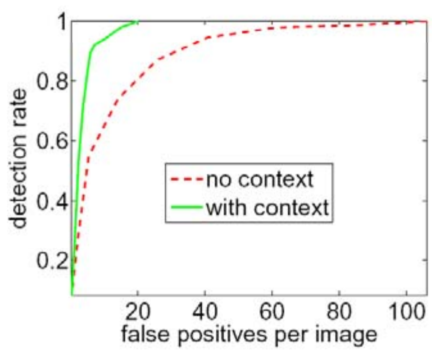
No Context

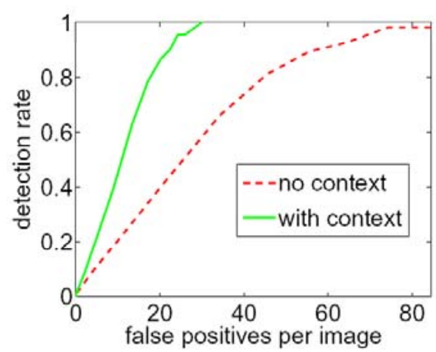
Keyboard Mouse



With Context

Keyboard





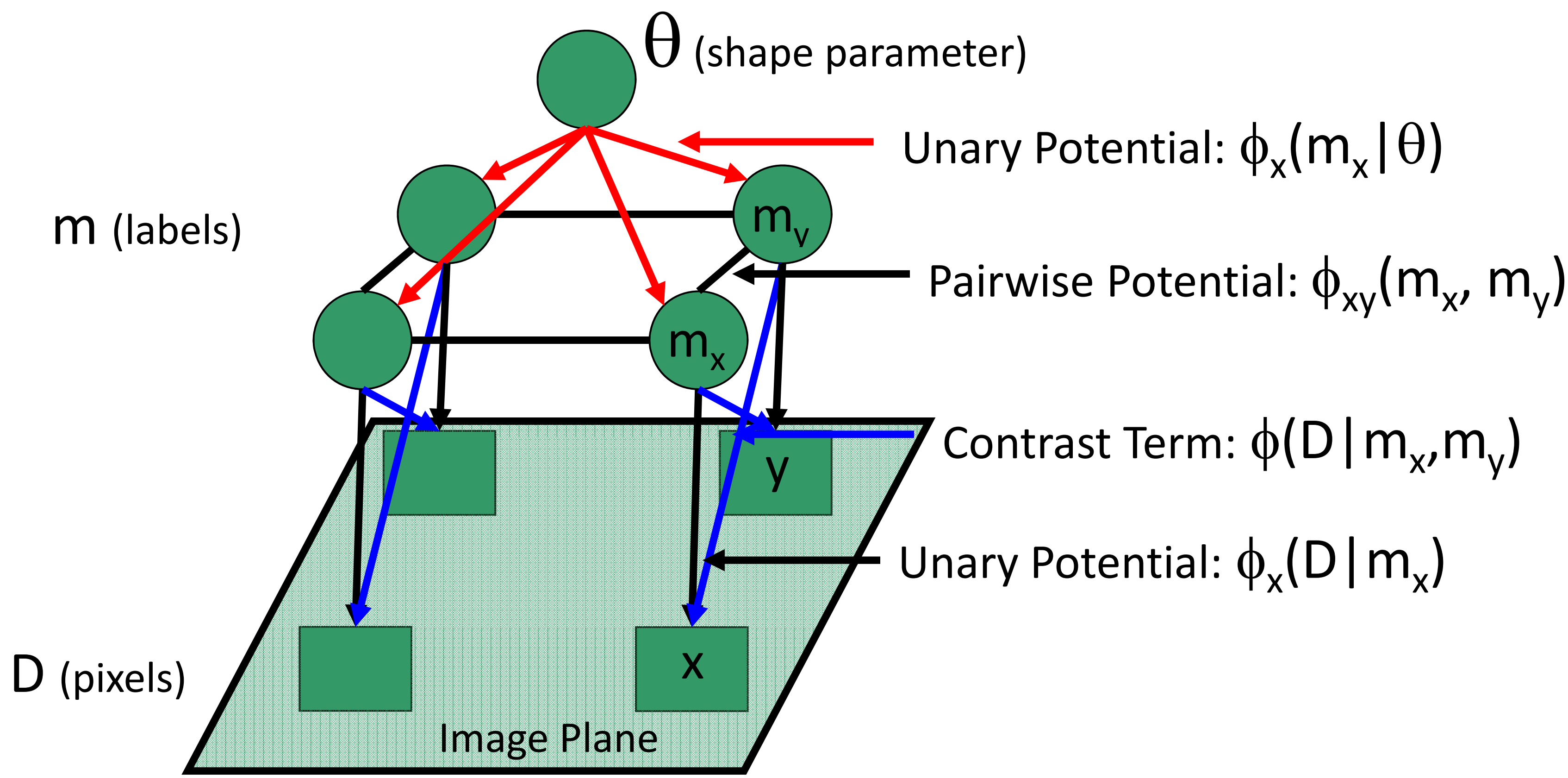
Mouse

[Kumar & Hebert ICCV '05]

Applications of Conditional Random Fields in Computer Vision

Segmentation, Object recognition, Scene recognition, Human-object interaction

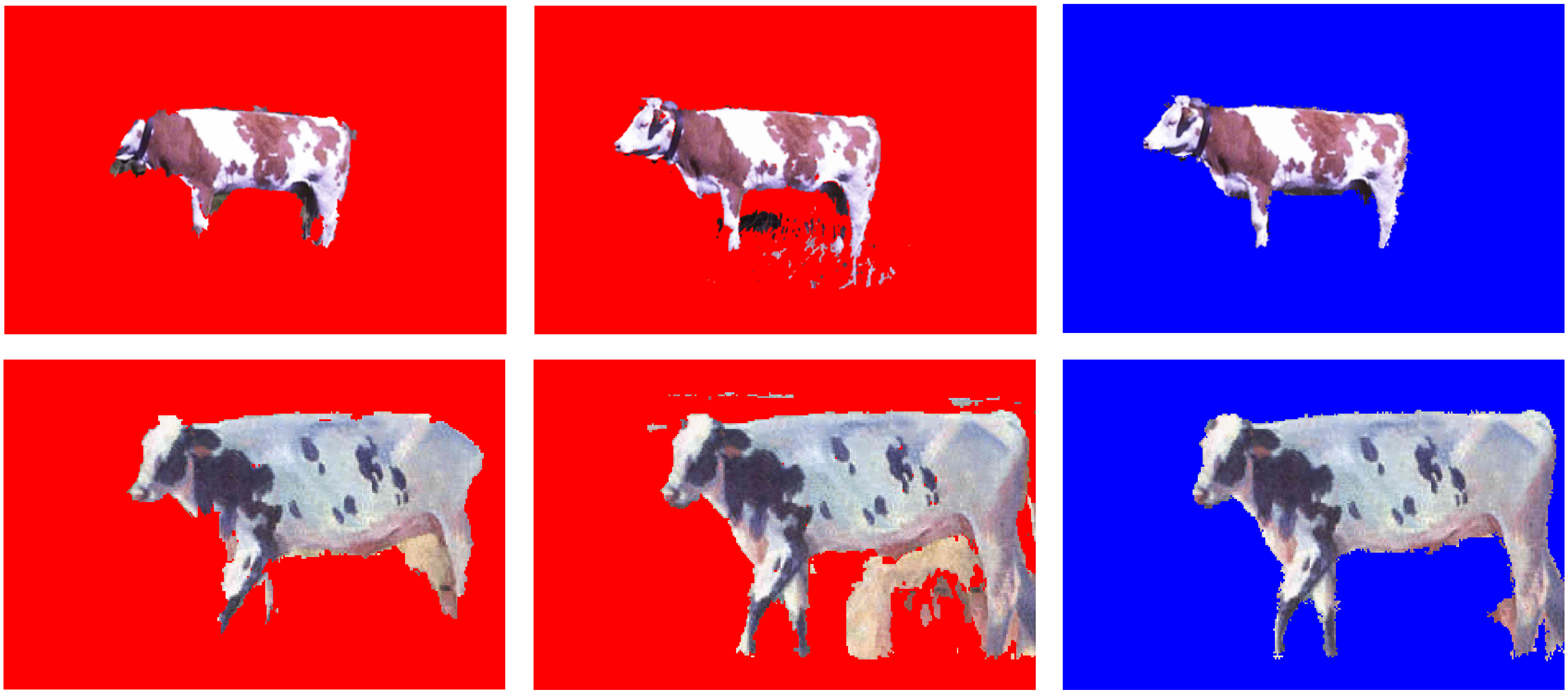
ObjCut – CRF for image segmentation



[Kumar et al, CVPR 2005]

Segmentation Results

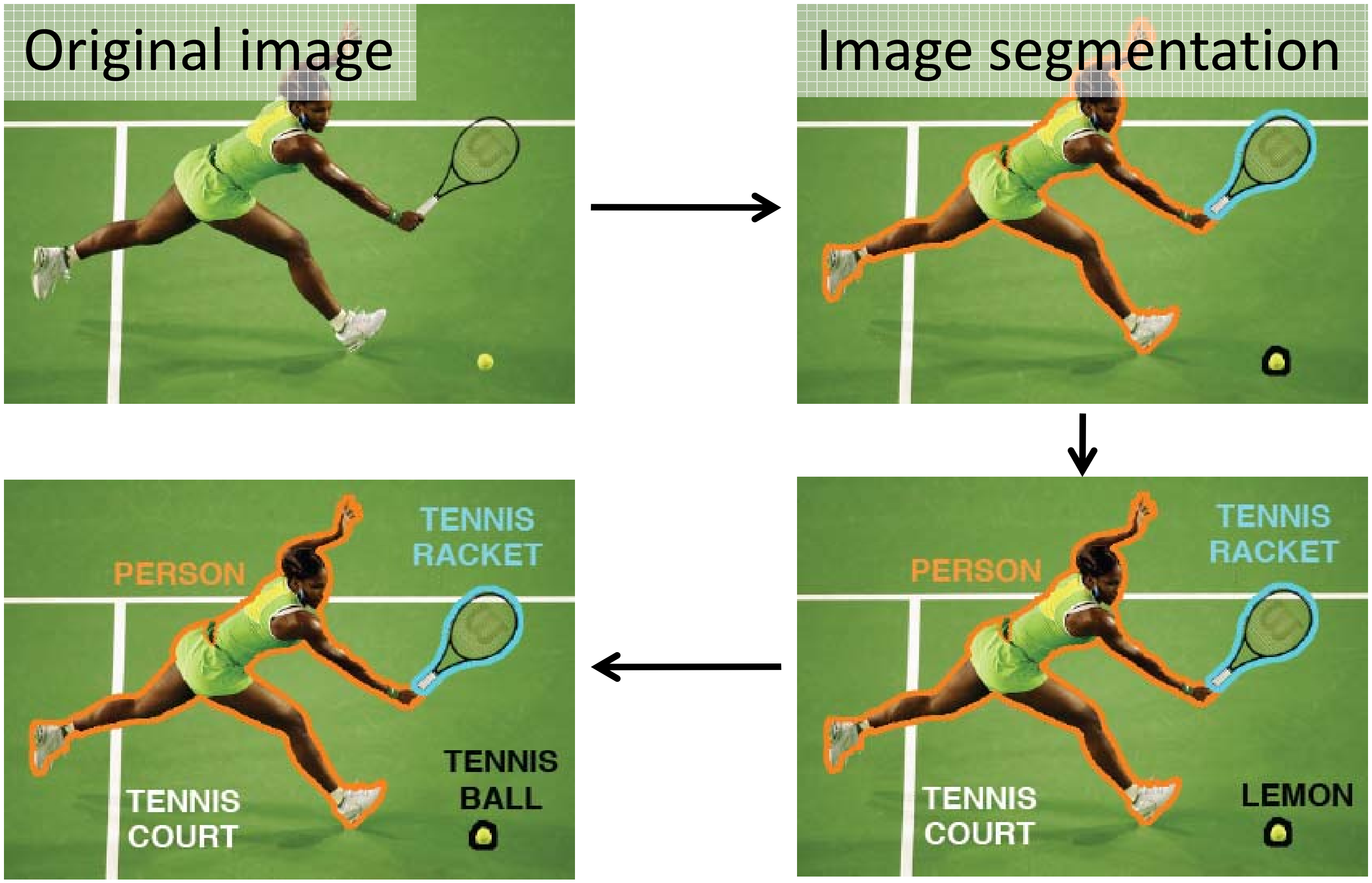
Shape only | Appearance only | Shape + Appearance



Without $\phi_x(D|m_x)$ | Without $\phi_x(m_x|\theta)$

[Kumar et al, CVPR 2005]

Objects in Context – CRF for Object Recognition



Semantic context
(Interaction terms)

Segment labeling
(Unary terms)

[Rabinovich et al, CVPR 2005]

Object recognition results

Original image | Without context | With context

[Rabinovich et al, CVPR 2005]

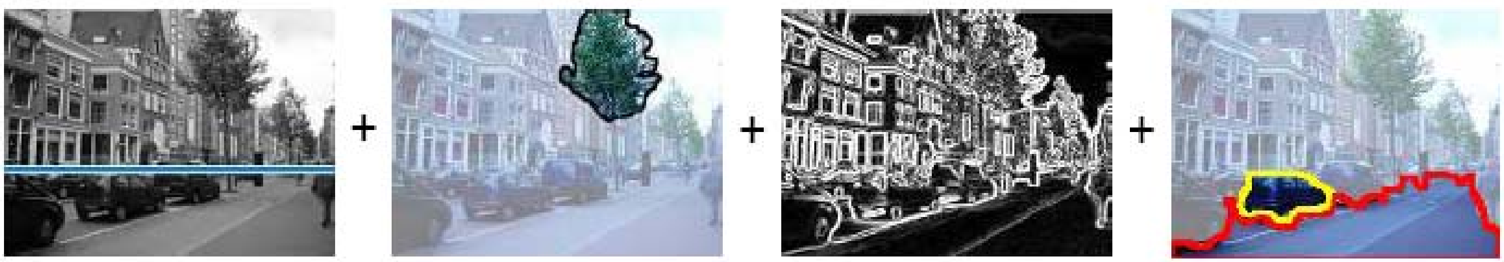
Region-based Model – CRF for Scene recognition

$$E(\mathbf{R}, \mathbf{A}, \mathbf{S}, \mathbf{G}, v^{hz}, K \mid I, \theta)$$

Variables
- α_p: pixel appearance
- $\mathbf{R}_p$: pixel-to-region correspondence
- $\mathbf{A}_r$: region appearance
- $\mathbf{S}_r$: region semantic class
- $\mathbf{G}_r$: region geometry
- $\mathbf{v}^{hz}$: location of horizon

=

$$\psi^{horizon}(v^{hz}) \quad \psi^{region}(S_r, G_r, A_r, v^{hz}) \quad \psi^{boundary}(A_r, A_s) \quad \psi^{pair}(S_r, S_s, G_r, G_s)$$



Horizon Term e.g., vanishing lines

Region Term e.g., consistent appearance and location

Boundary Term e.g., difference in color/texture between regions

Pairwise Term e.g., foreground on road

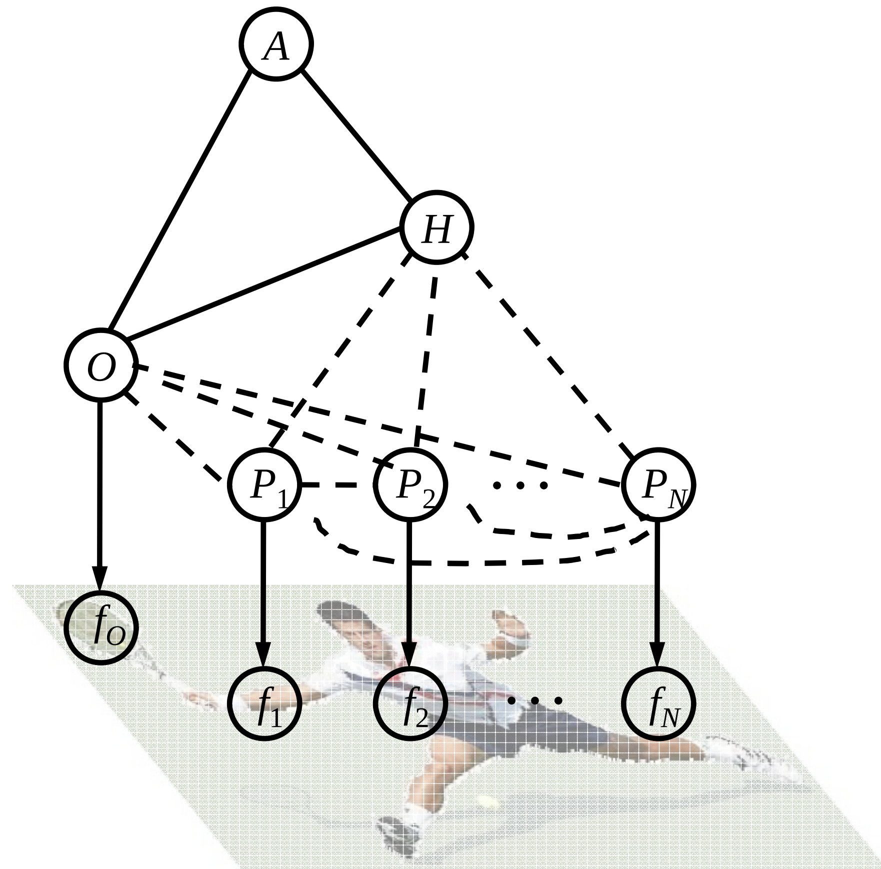
[Gould et al, CVPR 2009]

Scene recognition results

[Gould et al, CVPR 2009]

Mutual Context Model – CRF for human-object interaction

HOI activity: Tennis Forehand

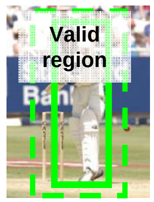


[Yao and Fei-Fei, CVPR 2010]

Object Detection Results

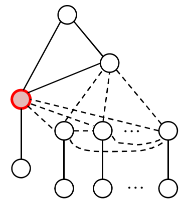
Sliding window

[Andriluka et al, 2009]



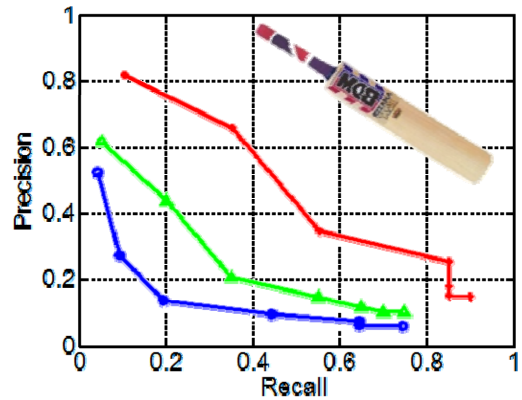
Pedestrian context

[Dalal & Triggs, 2006]

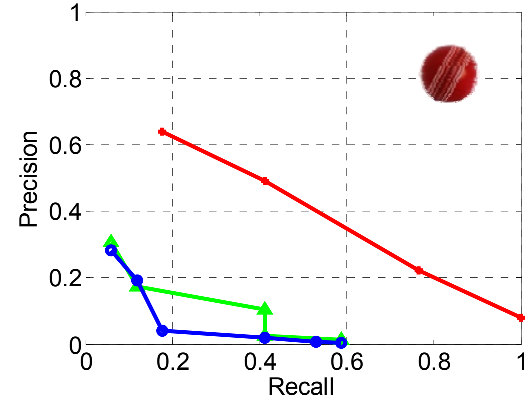


Our Method

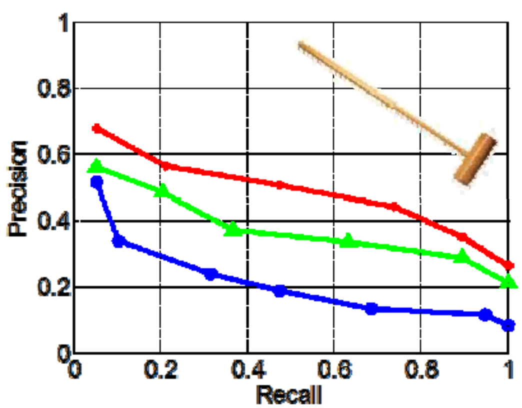
Cricket bat



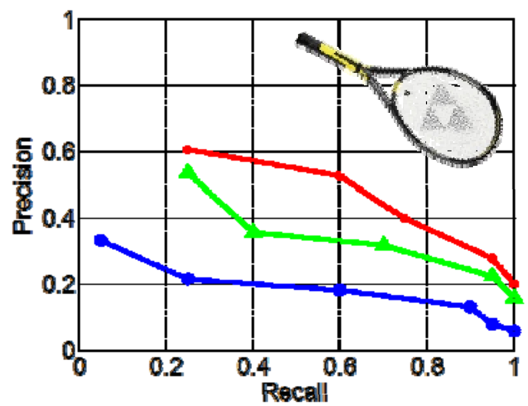
Cricket ball



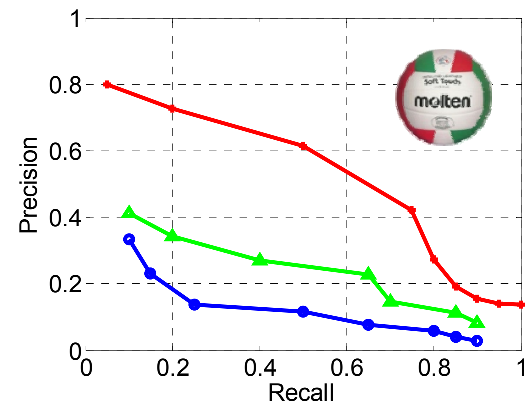
Croquet mallet



Tennis racket



Volleyball

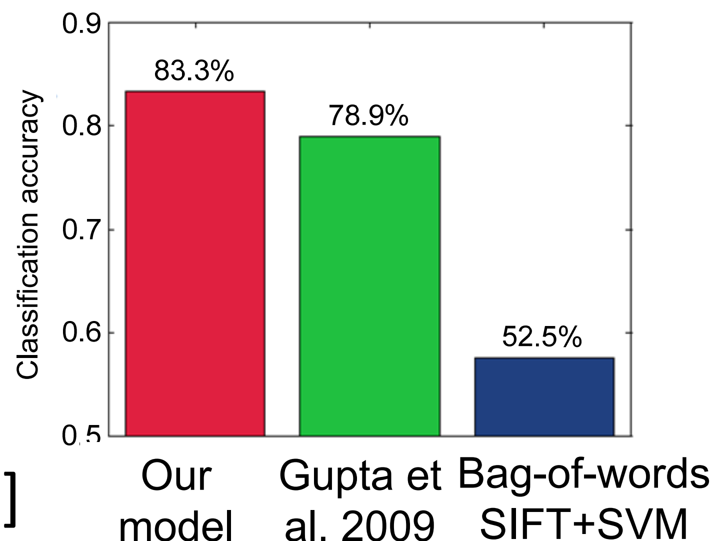


[Yao and Fei-Fei, CVPR 2010]

Human Pose Estimation Results

Method	Torso	Upper Leg		Lower Leg		Upper Arm		Lower Arm		Head
Ramanan, 2006	.52	.22	.22	.21	.28	.24	.28	.17	.14	.42
Andriluka et al, 2009	.50	.31	.30	.31	.27	.18	.19	.11	.11	.45
Our model	**.66**	**.43**	**.39**	**.44**	**.34**	**.44**	**.40**	**.27**	**.29**	**.58**

Activity Recognition Results



[Yao and Fei-Fei, CVPR 2010]

Summary

- **CRF-based discriminative models in Vision**
  - **Principled approach to model interactions at pixel, patch, region or object level for robust classification**
- **Combine local discriminative classifiers with data-dependent label interactions**
  - **Alternative to traditional MRFs**
- **Hierarchy of fields to capture different contexts**
- **Several computer vision tasks in the same framework**
  - **Denoising, region classification, texture recognition, object detection, gesture recognition,…**

Open Questions

Features ⟺ **Model** **Convolutional Networks [Lecun &Bengio '98]**

Ambiguity in recognition!