Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials

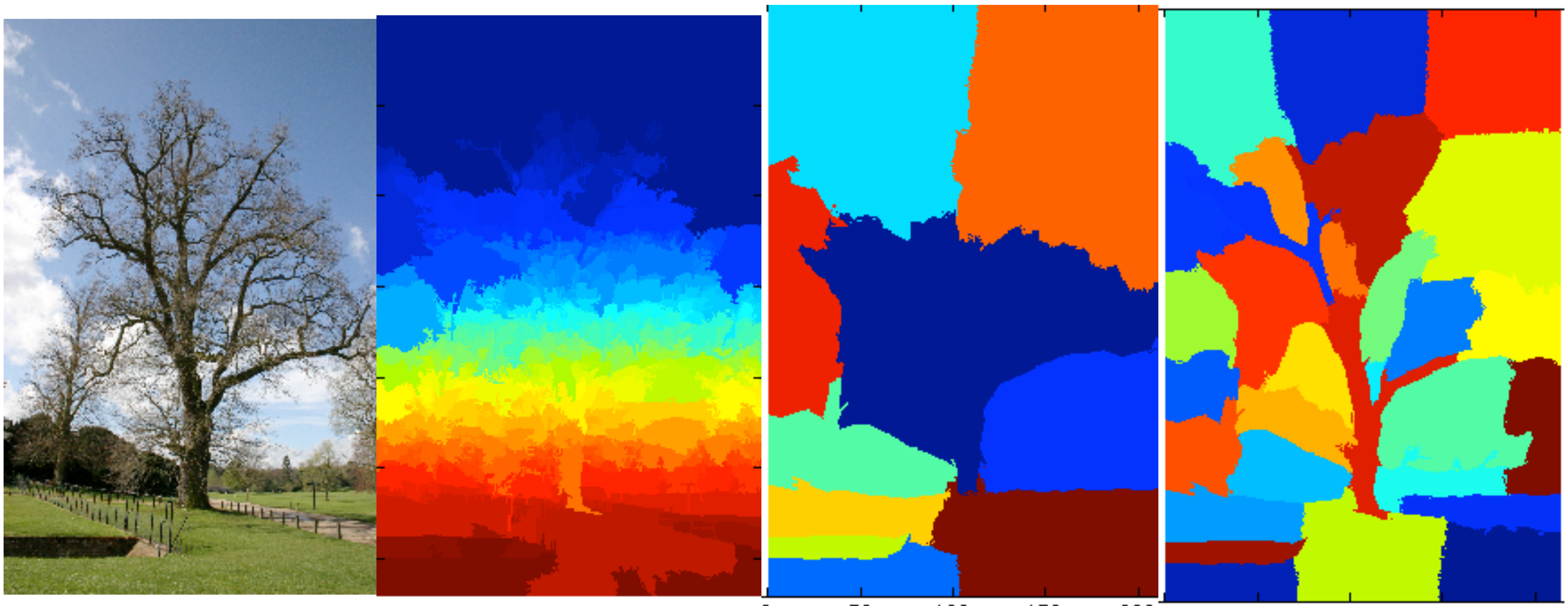
by Phillip Krahenbuhl and Vladlen Koltun

Presented by Adam Stambler

Multi-class image segmentation

Assign a class label to each pixel in the image

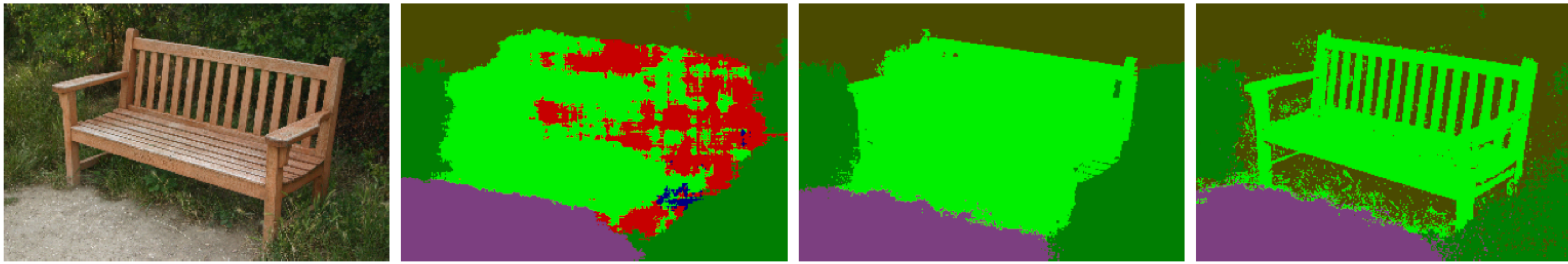
Super pixels are hard to make



meanshift ncut-10 ncut-30

- Don’t make super pixels

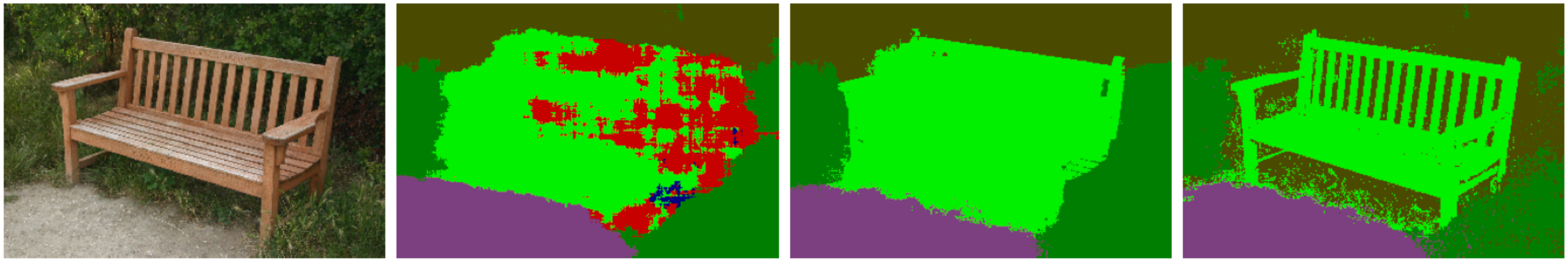
Operate On Pixels Directly



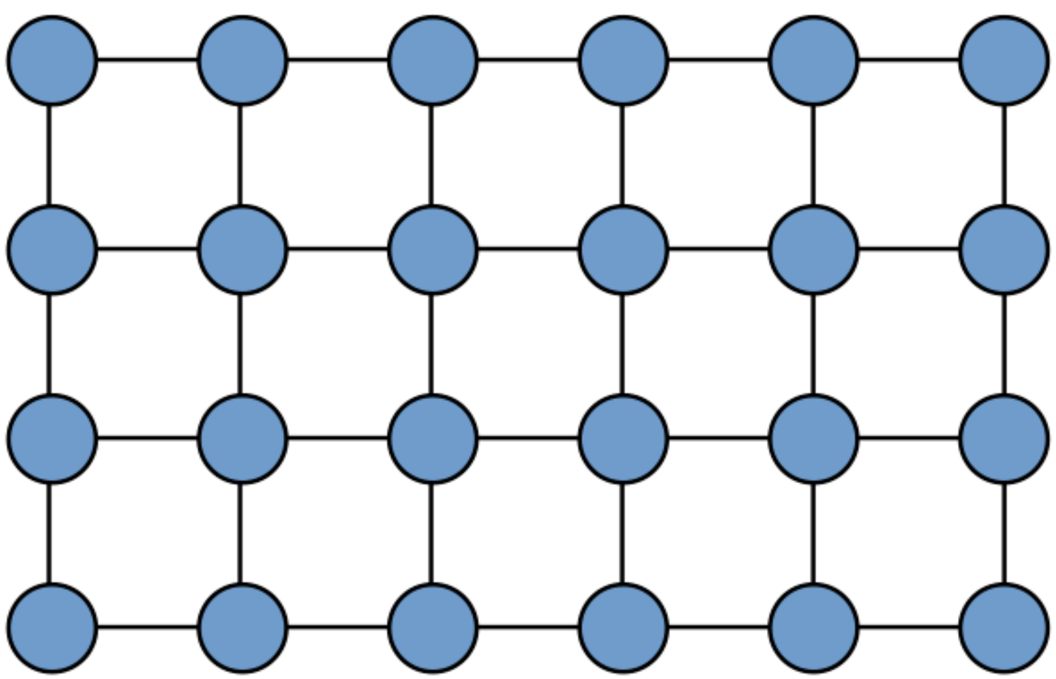
(a) Image (b) Unary classifiers (c) Robust P^n CRF (d) Fully connected CRF, MCMC inference, 36 hrs

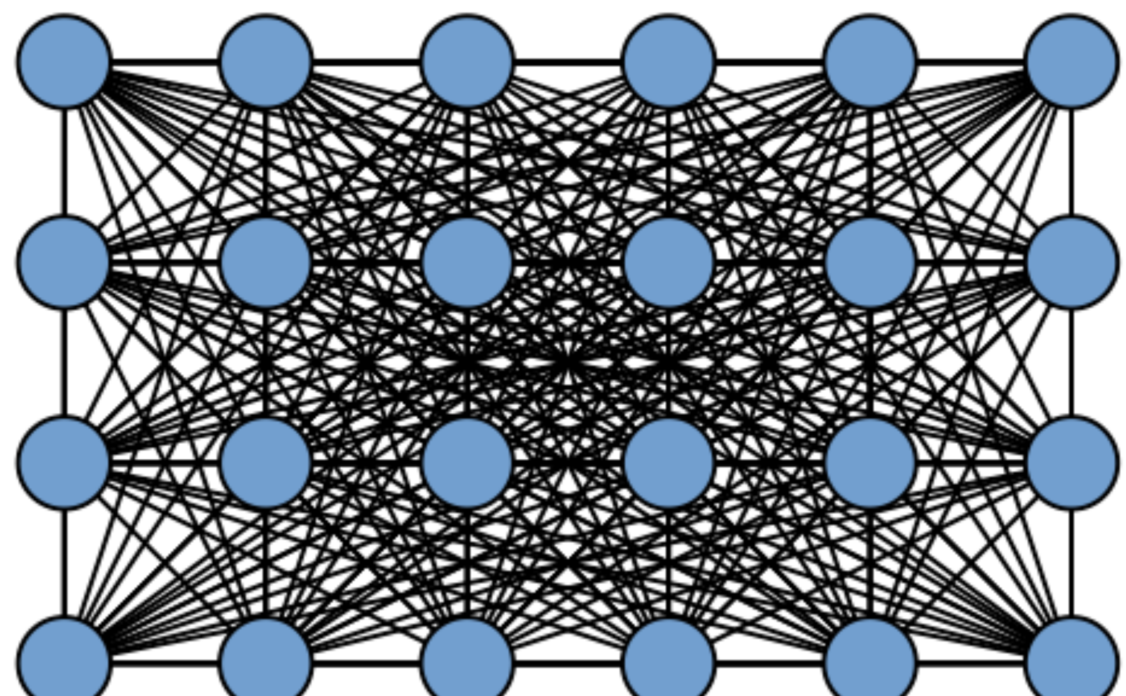
- Pixel wise classification- texture/local shape features

Consistency with MRF/CRF



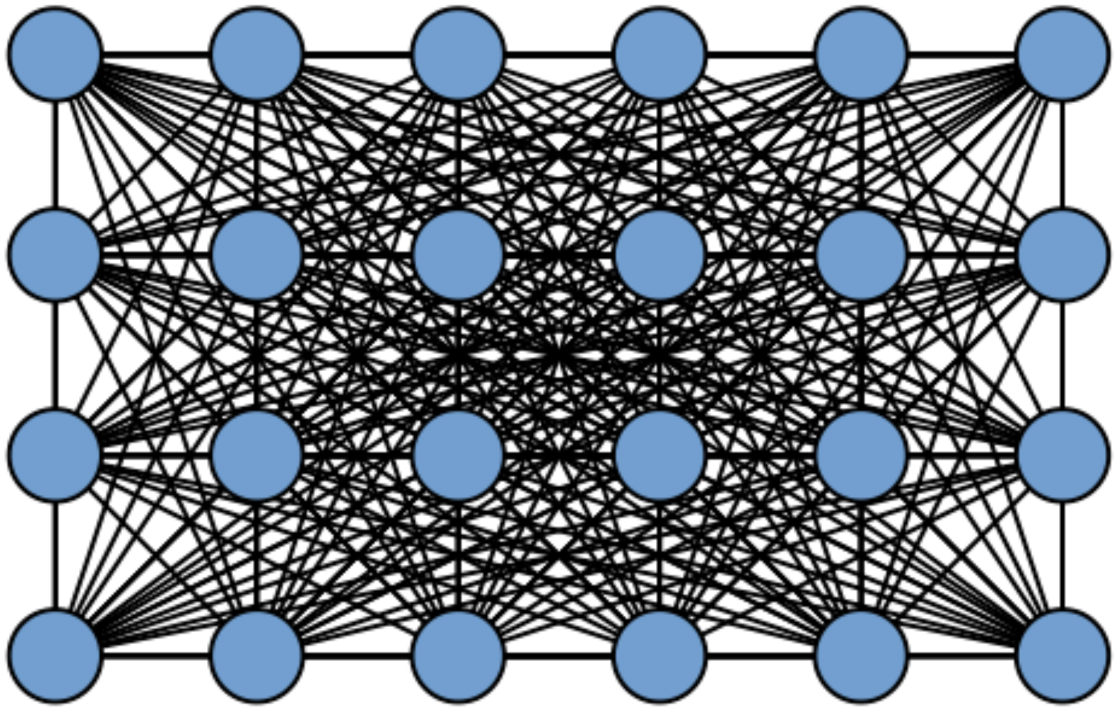
(a) Image (b) Unary classifiers (c) Robust P^n CRF (d) Fully connected CRF, MCMC inference, 36 hrs





36

hours!

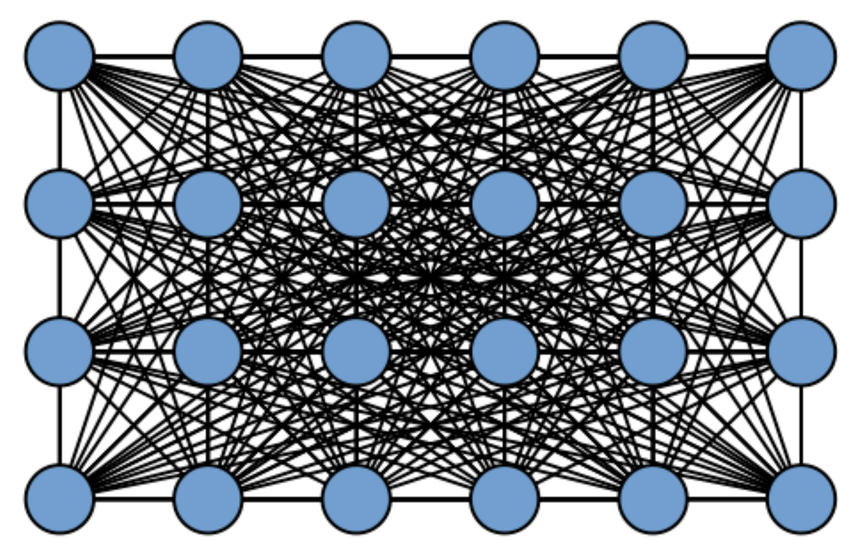


Efficient CRF's results:

(a) Image (b) Unary classifiers (c) Robust P^n CRF (d) Fully connected CRF, MCMC inference, 36 hrs (e) Fully connected CRF, our approach, 0.2 seconds

Figure 1: Pixel-level classification with a fully connected CRF. (a) Input image from the MSRC-21 dataset. (b)

0.2 s



Solving MRFs and CRFs

- Each Clique Modeled as Gibbs Distribution

$$\Pr(\mathbf{x}|\mathbf{D}) = \frac{1}{Z}\exp\Big(-\sum_{c\in\mathcal{C}}\psi_c(\mathbf{x}_c)\Big),$$

  - Unary Potentials and Pairwise potential

$$E(\mathbf{x}) = \sum_i \underbrace{\psi_u(x_i)}_{\text{unary term}} + \sum_i \sum_{j\in\mathcal{N}_i} \underbrace{\psi_p(x_i, x_j)}_{\text{pairwise term}}$$

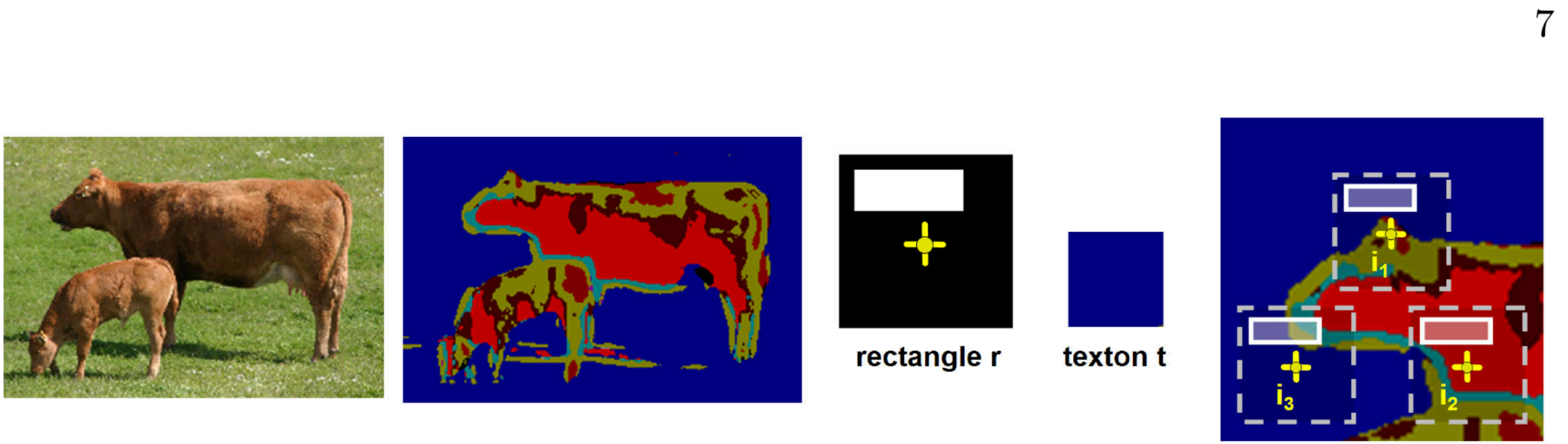
+ (optional higher order terms)

- Maximum-a-posteriori solutions are NP-Hard
  - Message Passing algorithms : belief propagation
  - Move Making Algorithms : α-expansion , αβ-swap

Graph connections

1. Adjacent pixels are connected
    - Textonboost CRF approach
2. Adjacent pixels are connected + super-pixels consistent
    - Robust Pn CRF
3. All pixels are fully connected
    - Efficient CRF (this paper)

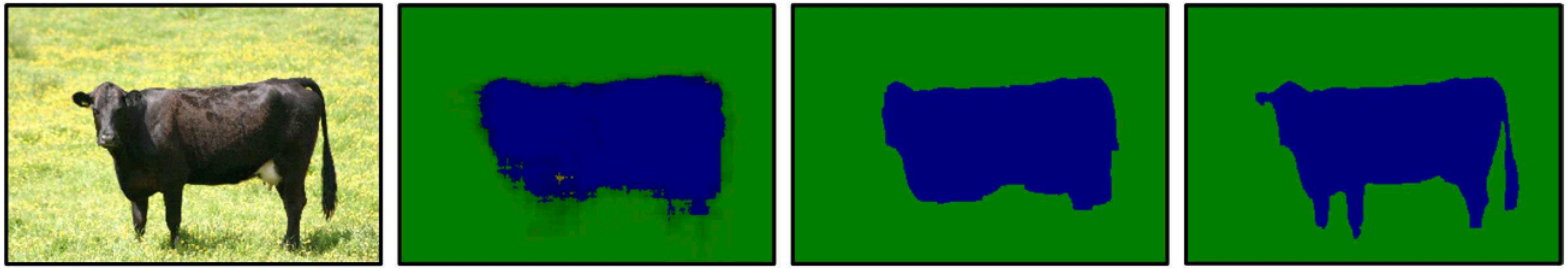
Unary Potential : Texton Boost



(a) Input image (b) Texton map (c) Feature pair = (r,t) (d) Superimposed rectangles

- Responsible for most of the accuracy in all of the papers

Texton Boosting

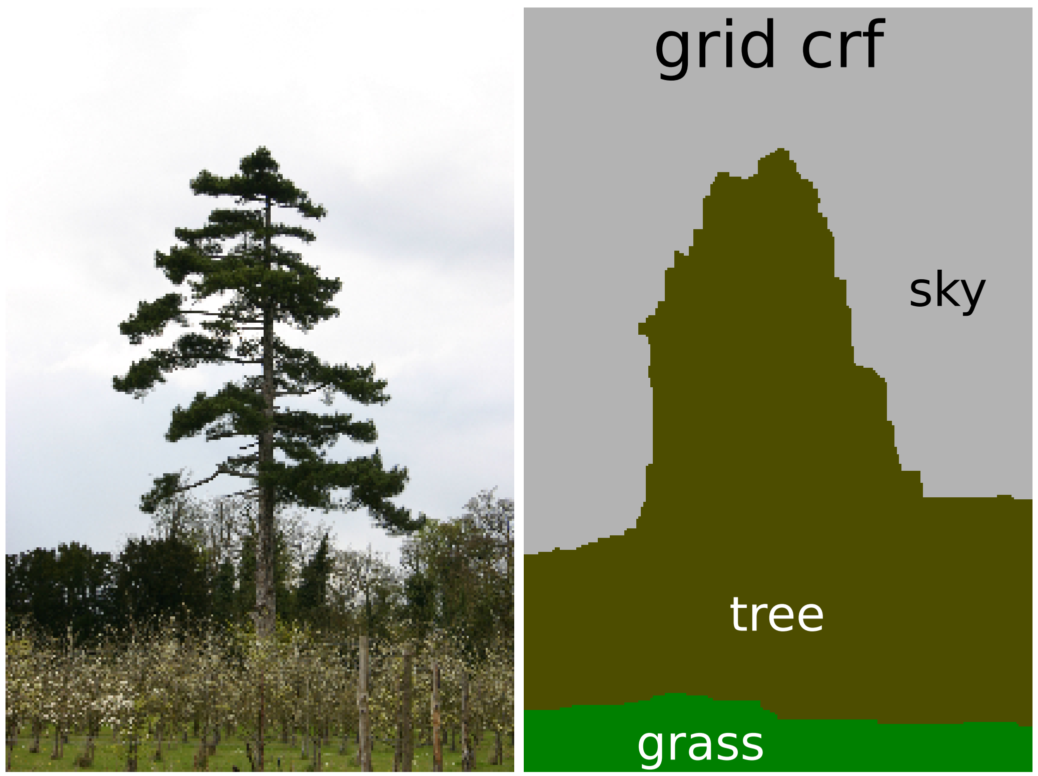


(a) (b) 69.6% just pixel (c) 70.3% crf-no color (d) 72.2% crf-full

- TextonBoost: Joint Appearance, Shape and Context Modeling for Multi-Class Object Recognition and Segmentation
    - Each pixel is only connected to its adjacent neighbors
    - Jointly model the texture and shape a single feature

Adjacency CRF models

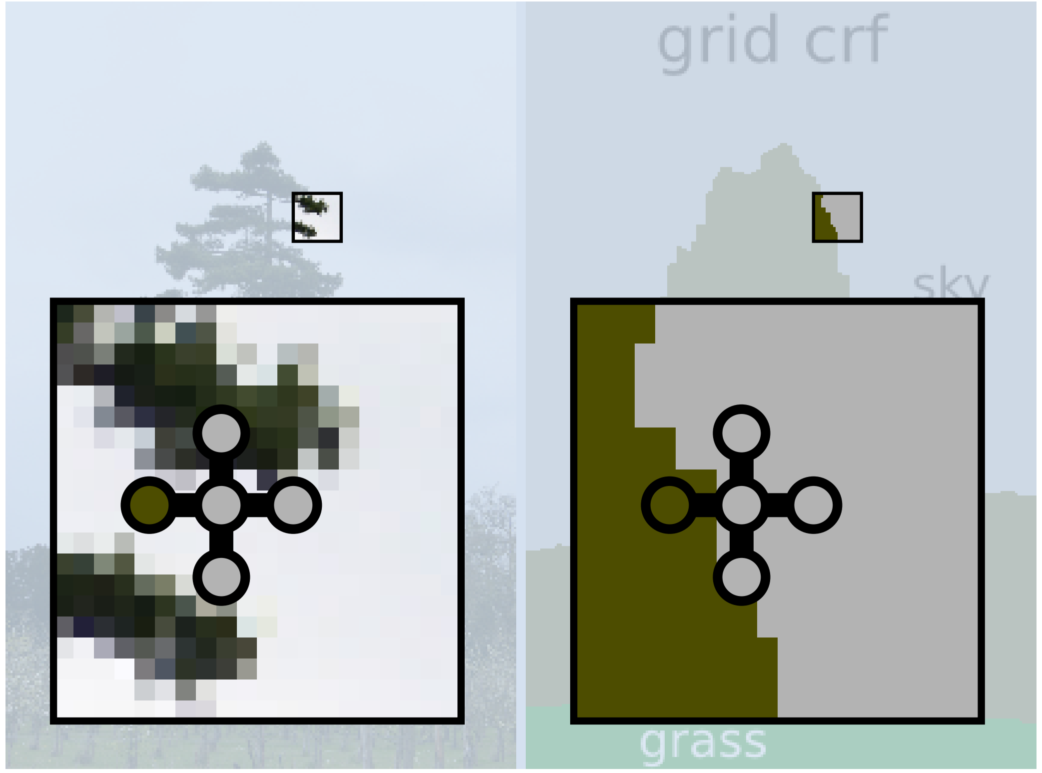
$$E(\mathbf{x}) = \sum_i \underbrace{\psi_u(x_i)}_{\text{unary term}} + \sum_i \sum_{j \in \mathcal{N}_i} \underbrace{\psi_p(x_i, x_j)}_{\text{pairwise term}}$$



- Efficient inference
  - 1 second for 50′000 variables
- Limited expressive power
- Only local interactions
- Excessive smoothing of object boundaries
  - Shrinking bias

Adjacency CRF models

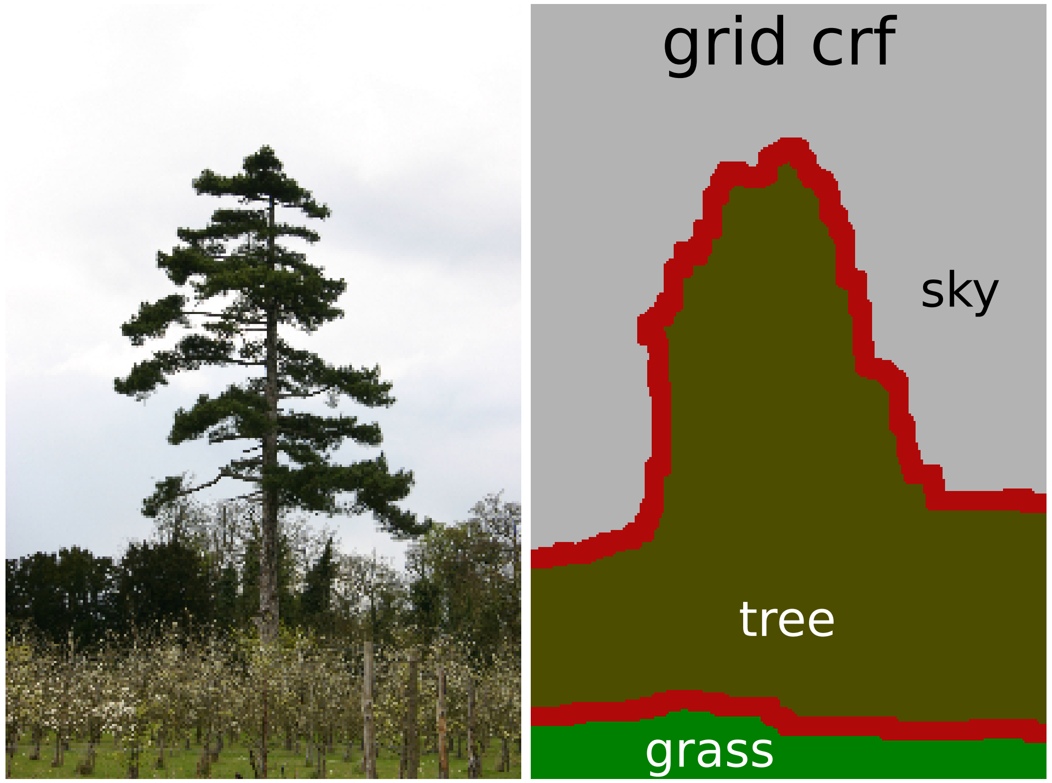
$$E(\mathbf{x}) = \sum_i \underbrace{\psi_u(x_i)}_{\text{unary term}} + \sum_i \sum_{j \in \mathcal{N}_i} \underbrace{\psi_p(x_i, x_j)}_{\text{pairwise term}}$$



- Efficient inference
  - 1 second for 50′000 variables
- Limited expressive power
- Only local interactions
- Excessive smoothing of object boundaries
  - Shrinking bias

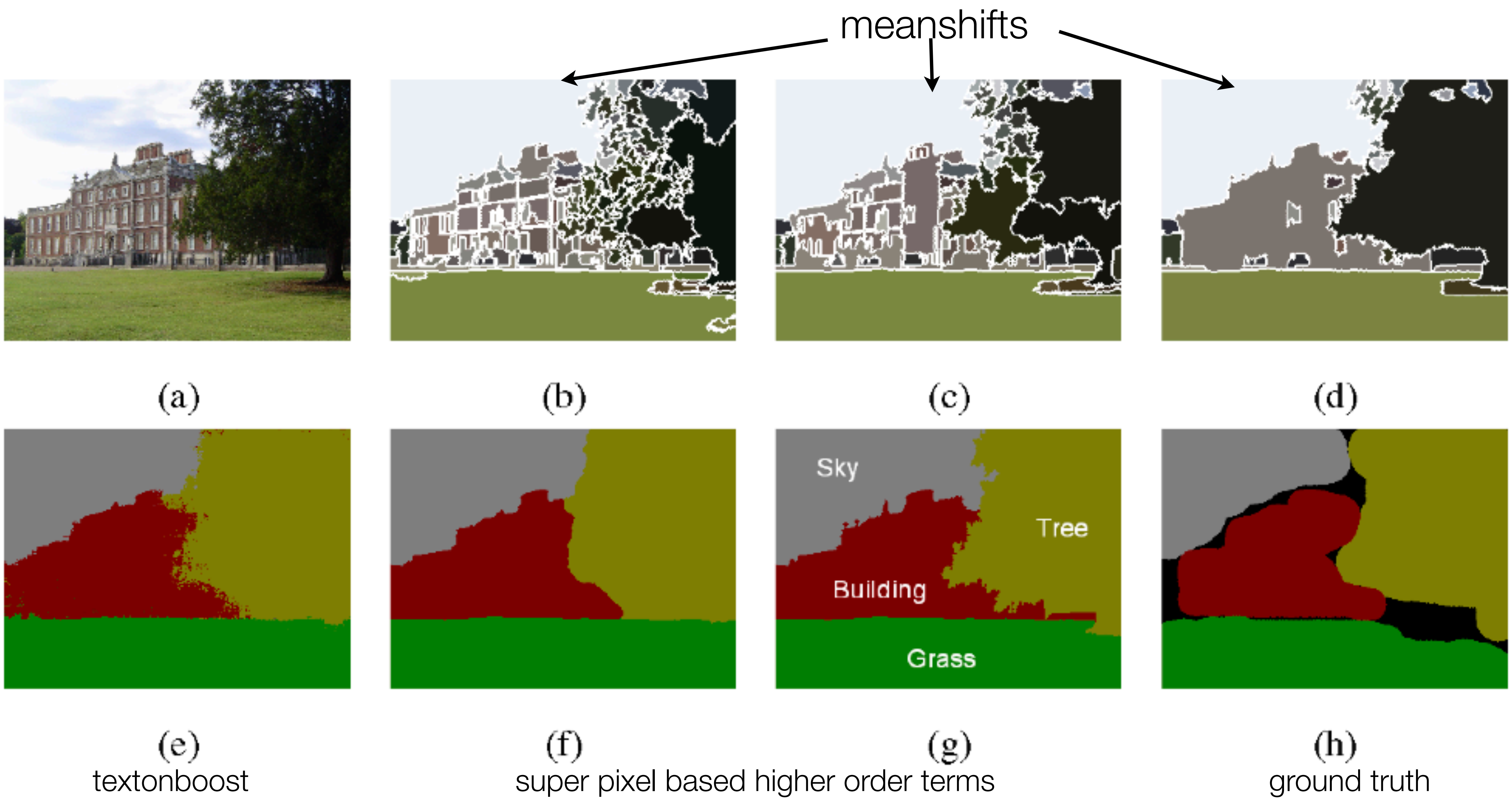
Adjacency CRF models

$$E(\mathbf{x}) = \sum_i \underbrace{\psi_u(x_i)}_{\text{unary term}} + \sum_i \sum_{j \in \mathcal{N}_i} \underbrace{\psi_p(x_i, x_j)}_{\text{pairwise term}}$$



- Efficient inference
  - 1 second for $50'000$ variables
- Limited expressive power
- Only local interactions
- Excessive smoothing of object boundaries
  - Shrinking bias

Operate On Pixels + Super-pixels



- “Robust Higher Order Potentials for Enforcing Label Consistency”- Koli et al.

Operate on Super-pixels + Pixels

- higher order potentials defined on super pixels to enforce regional consistency
  - soft label constraints using super-pixel consistency potentials

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \psi_i(x_i) + \sum_{(i,j) \in \mathcal{E}} \psi_{ij}(x_i, x_j) + \sum_{c \in \mathcal{S}} \psi_c(\mathbf{x}_c),$$

unary pairwise super-pixel

  - Super pixel term also models consistency within super pixel

Model definition

$$E(\mathbf{x}) = \sum_i \underbrace{\psi_u(x_i)}_{\text{unary term}} + \sum_i \sum_{j>i} \underbrace{\psi_p(x_i, x_j)}_{\text{pairwise term}}$$

Gaussian edge potentials

$$\psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^{K} w^{(m)} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j)$$

- Label compatibility function μ
- Linear combination of Gaussian kernels

$$k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) = \exp(-\frac{1}{2}(\mathbf{f}_i - \mathbf{f}_j)\Sigma^{(m)}(\mathbf{f}_i - \mathbf{f}_j))$$

- Arbitrary feature space $\mathbf{f}_i$

Model definition

$$E(\mathbf{x}) = \sum_i \underbrace{\psi_u(x_i)}_{\text{unary term}} + \sum_i \sum_{j>i} \underbrace{\psi_p(x_i, x_j)}_{\text{pairwise term}}$$

Gaussian edge potentials

$$\psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^{K} w^{(m)} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j)$$

- Label compatibility function μ
- Linear combination of Gaussian kernels

$$k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) = \exp(-\frac{1}{2}(\mathbf{f}_i - \mathbf{f}_j)\Sigma^{(m)}(\mathbf{f}_i - \mathbf{f}_j))$$

- Arbitrary feature space $\mathbf{f}_i$

Model definition

$$E(\mathbf{x}) = \sum_i \underbrace{\psi_u(x_i)}_{\text{unary term}} + \sum_i \sum_{j>i} \underbrace{\psi_p(x_i, x_j)}_{\text{pairwise term}}$$

Gaussian edge potentials

$$\psi_p(x_i, x_j) = \boxed{\mu(x_i, x_j)} \sum_{m=1}^{K} w^{(m)} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j)$$

- Label compatibility function μ
- Linear combination of Gaussian kernels

$$k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) = \exp(-\frac{1}{2}(\mathbf{f}_i - \mathbf{f}_j)\Sigma^{(m)}(\mathbf{f}_i - \mathbf{f}_j))$$

- Arbitrary feature space $\mathbf{f}_i$

Model definition

$$E(\mathbf{x}) = \sum_i \underbrace{\psi_u(x_i)}_{\text{unary term}} + \sum_i \sum_{j>i} \underbrace{\psi_p(x_i, x_j)}_{\text{pairwise term}}$$

Gaussian edge potentials

$$\psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^{K} w^{(m)} k^{(m)}(\boxed{\mathbf{f}_i}, \boxed{\mathbf{f}_j})$$

- Label compatibility function μ
- Linear combination of Gaussian kernels

$$k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) = \exp(-\frac{1}{2}(\mathbf{f}_i - \mathbf{f}_j)\Sigma^{(m)}(\mathbf{f}_i - \mathbf{f}_j))$$

- Arbitrary feature space $\mathbf{f}_i$

Model definition

$$E(\mathbf{x}) = \sum_i \underbrace{\psi_u(x_i)}_{\text{unary term}} + \sum_i \sum_{j>i} \underbrace{\psi_p(x_i, x_j)}_{\text{pairwise term}}$$

Gaussian edge potentials

$$\psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^{K} w^{(m)} k^{(m)}$$

Convolution is key to efficiency

- Label compatibility function μ
- Linear combination of Gaussian kernels

$$k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) = \exp(-\frac{1}{2}(\mathbf{f}_i - \mathbf{f}_j)\Sigma^{(m)}(\mathbf{f}_i - \mathbf{f}_j))$$

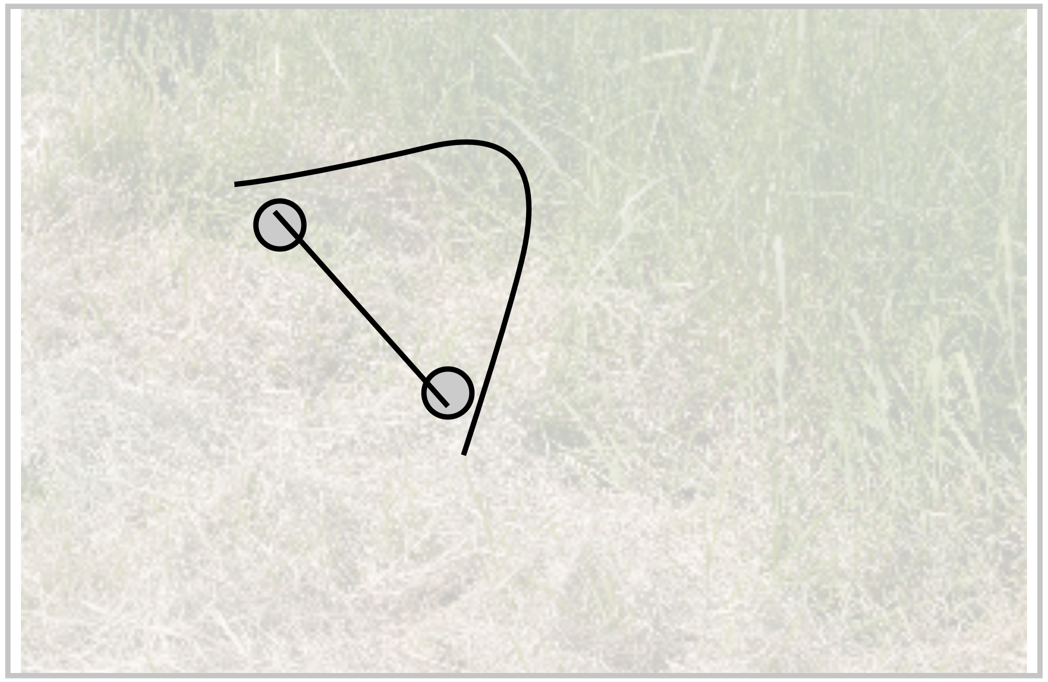
- Arbitrary feature space $\mathbf{f}_i$

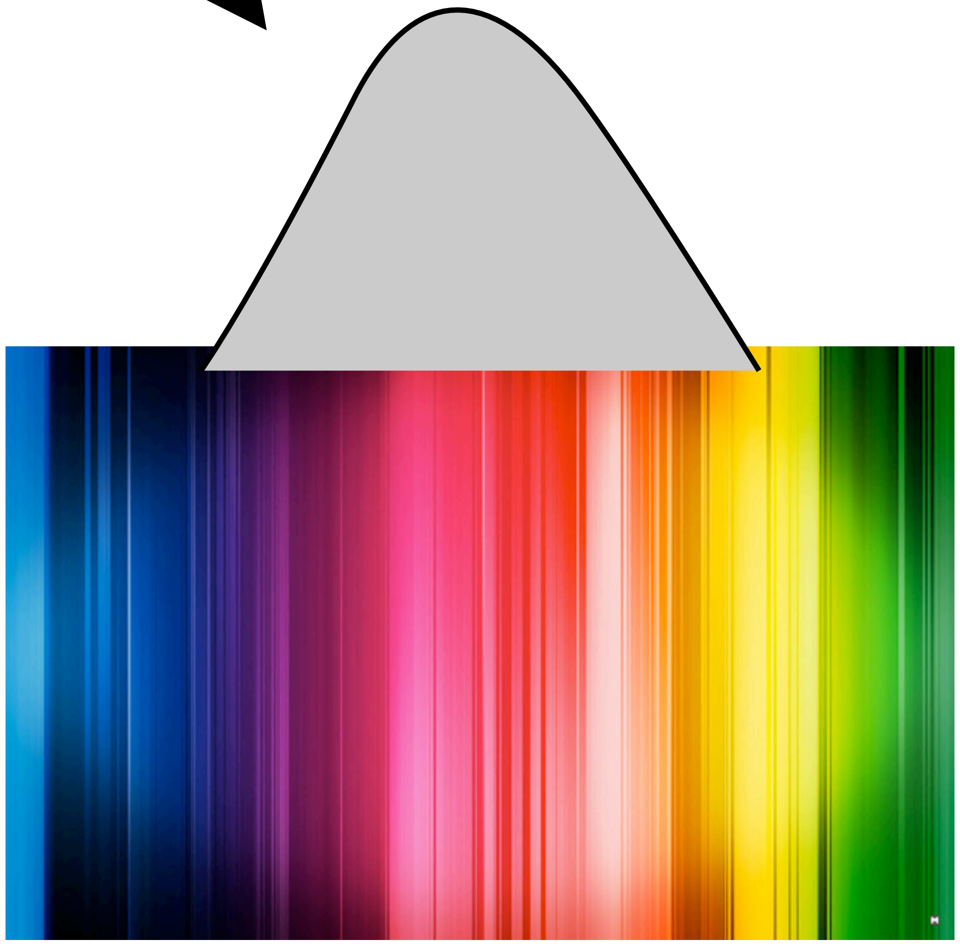
Detailed model definition

$$k(\mathbf{f}_i, \mathbf{f}_j) = w^{(1)} \underbrace{\exp\left(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}\right)}_{\text{appearance kernel}} + w^{(2)} \underbrace{\exp\left(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}\right)}_{\text{smoothness kernel}}$$

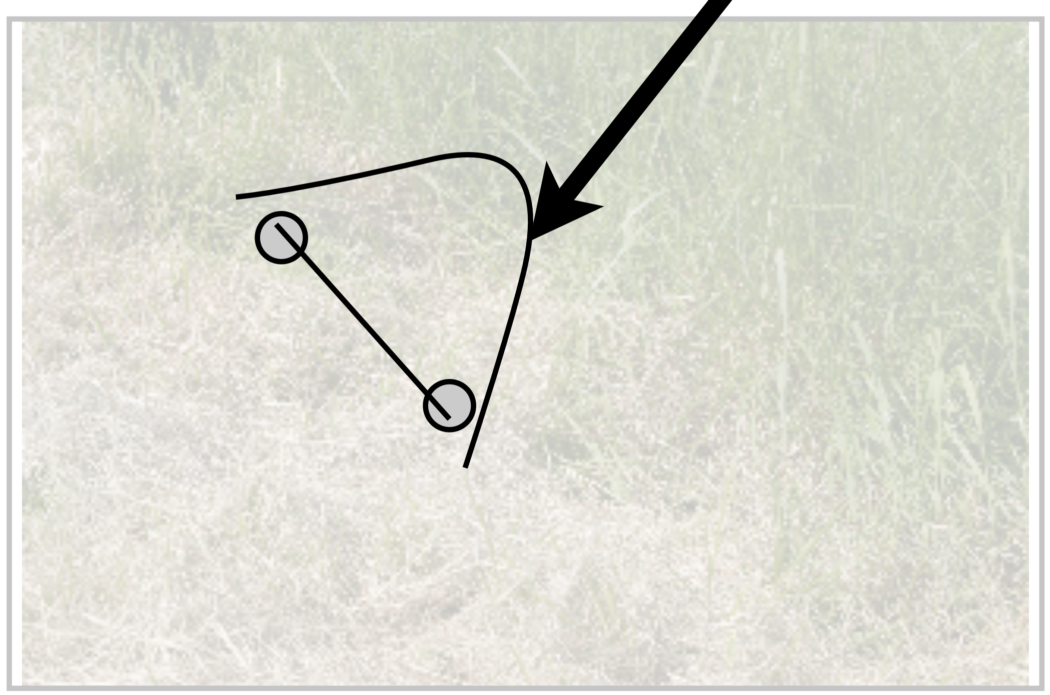
- Label compatibility
  - Potts model: $\mu(x_i, x_j) = 1_{[x_i \neq x_j]}$
  - Semi-metric model: $\mu(x_i, x_j)$ learned from data
- Appearance kernel
  - Color-sensitive model
- Local smoothness
  - Discourages pixel level noise

$$k(\mathbf{f}_i, \mathbf{f}_j) = \underbrace{w^{(1)} \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}\right)}_{\text{appearance kernel}} + \underbrace{w^{(2)} \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}\right)}_{\text{smoothness kernel}}$$





$$k(\mathbf{f}_i, \mathbf{f}_j) = \underbrace{w^{(1)} \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}\right)}_{\text{appearance kernel}} + \underbrace{w^{(2)} \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}\right)}_{\text{smoothness kernel}}$$



Message Passing

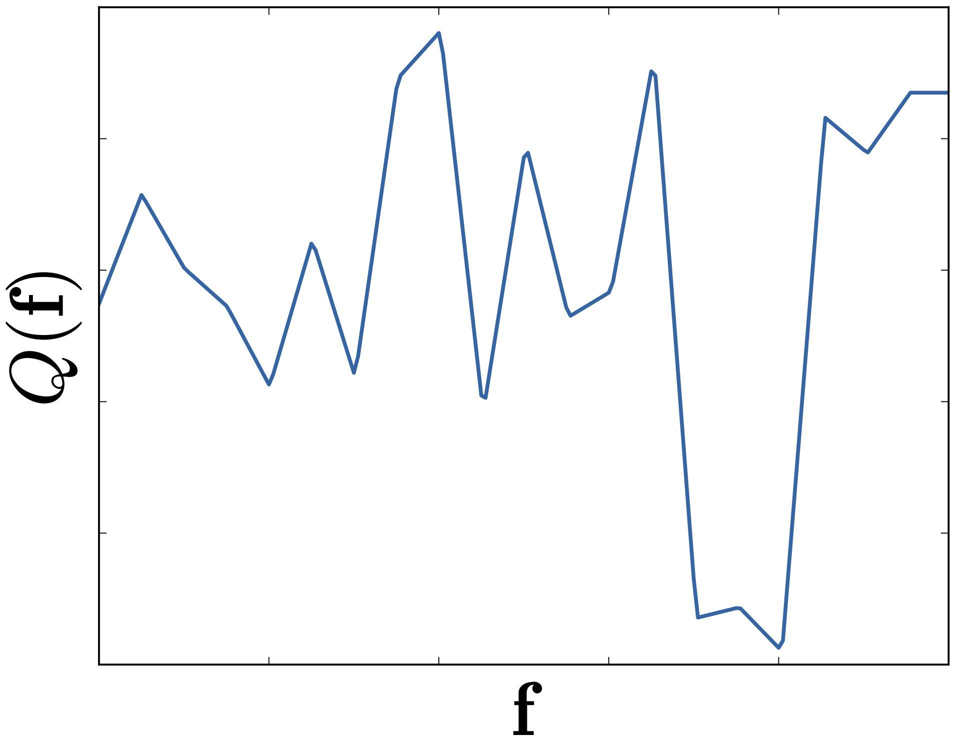
- Uses Mean Field Approximation to minimize KL-divergence
- Efficiency through signal theory low pass filtering
- Separable low-pass Gaussian filters propagate information over permutohedral lattice

Message Passing by high dimensional filtering

- Initialize graph with Unary potentials
- While not converged
    - Pass messages from each node to all other nodes
        - Messages consist of the pairwise blur weighting
    - Update node using compatibility transform

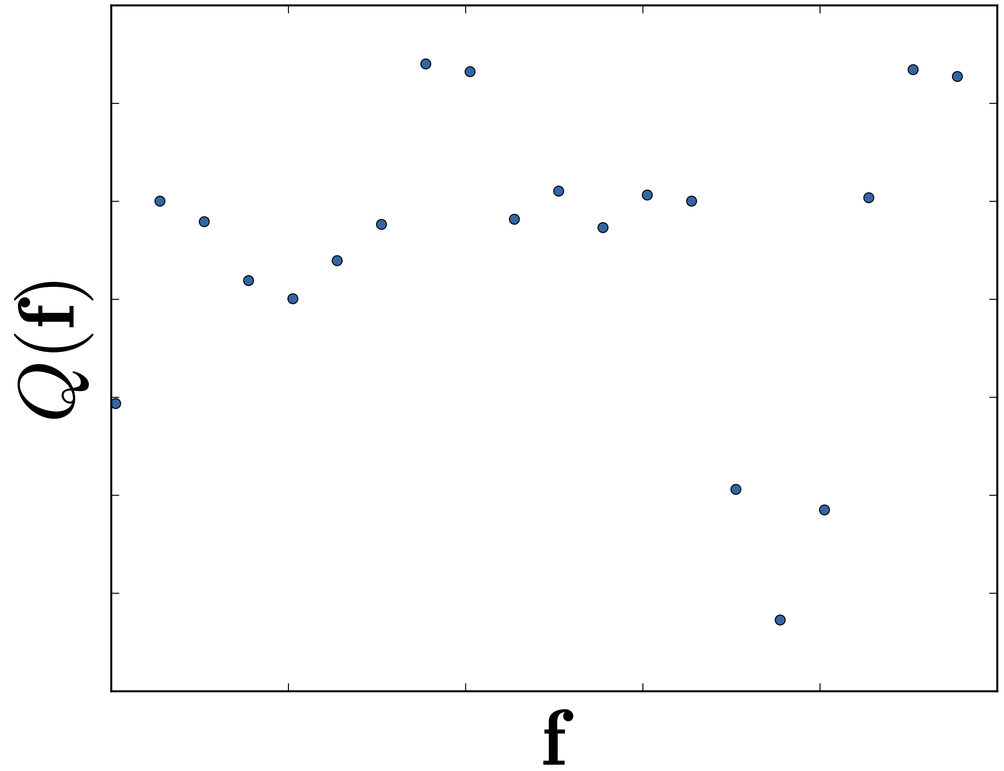
High-dimensional filtering [Paris & Durand 09]

- Downsample input signal $Q_j(I)$
- Blur the sampled signal
- Upsample to reconstruct the filtered signal $\overline{Q}_j(I)$



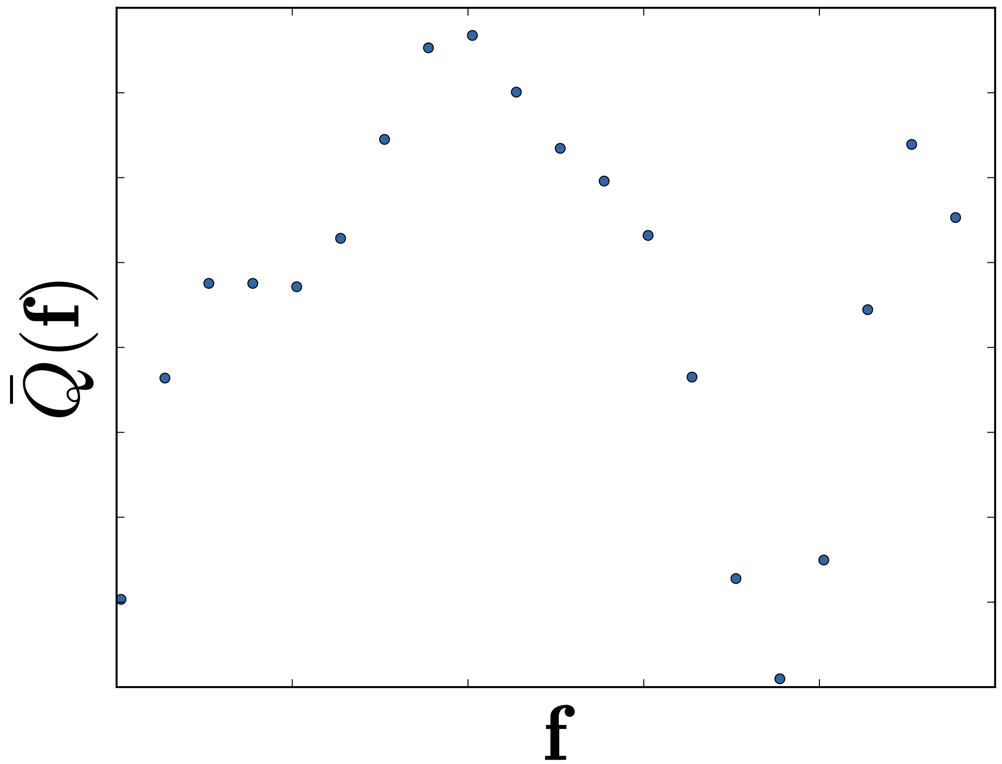
High-dimensional filtering [Paris & Durand 09]

- Downsample input signal $Q_j(l)$
- Blur the sampled signal
- Upsample to reconstruct the filtered signal $\overline{Q}_j(l)$



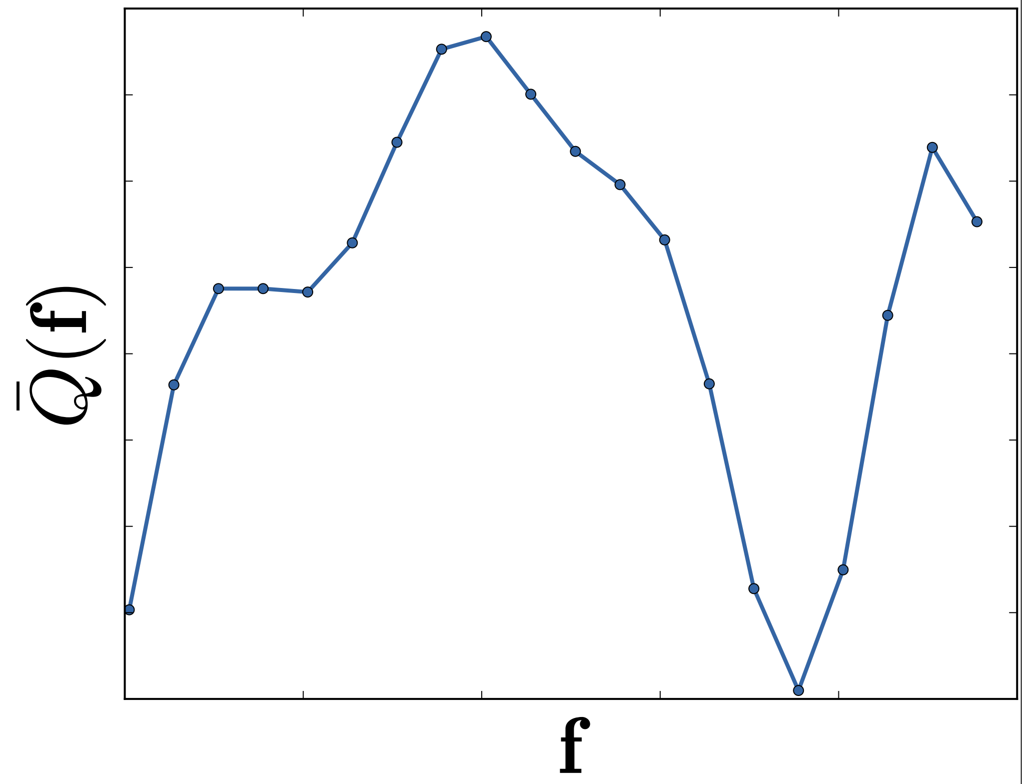
High-dimensional filtering [Paris & Durand 09]

- Downsample input signal $Q_j(I)$
- Blur the sampled signal
- Upsample to reconstruct the filtered signal $\overline{Q}_j(I)$



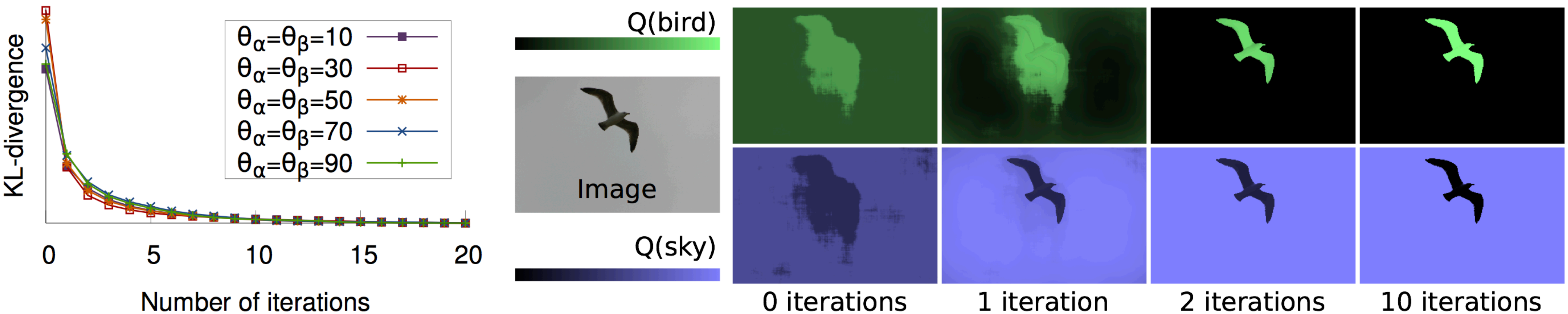
High-dimensional filtering [Paris & Durand 09]

- Downsample input signal $Q_j(I)$
- Blur the sampled signal
- Upsample to reconstruct the filtered signal $\overline{Q}_j(I)$



Permutohedral Lattice

Message Passing



(a) KL-divergence (b) Distributions $Q(X_i = \text{“bird”})$ (top) and $Q(X_i = \text{“sky”})$ (bottom)

Learned Parameters

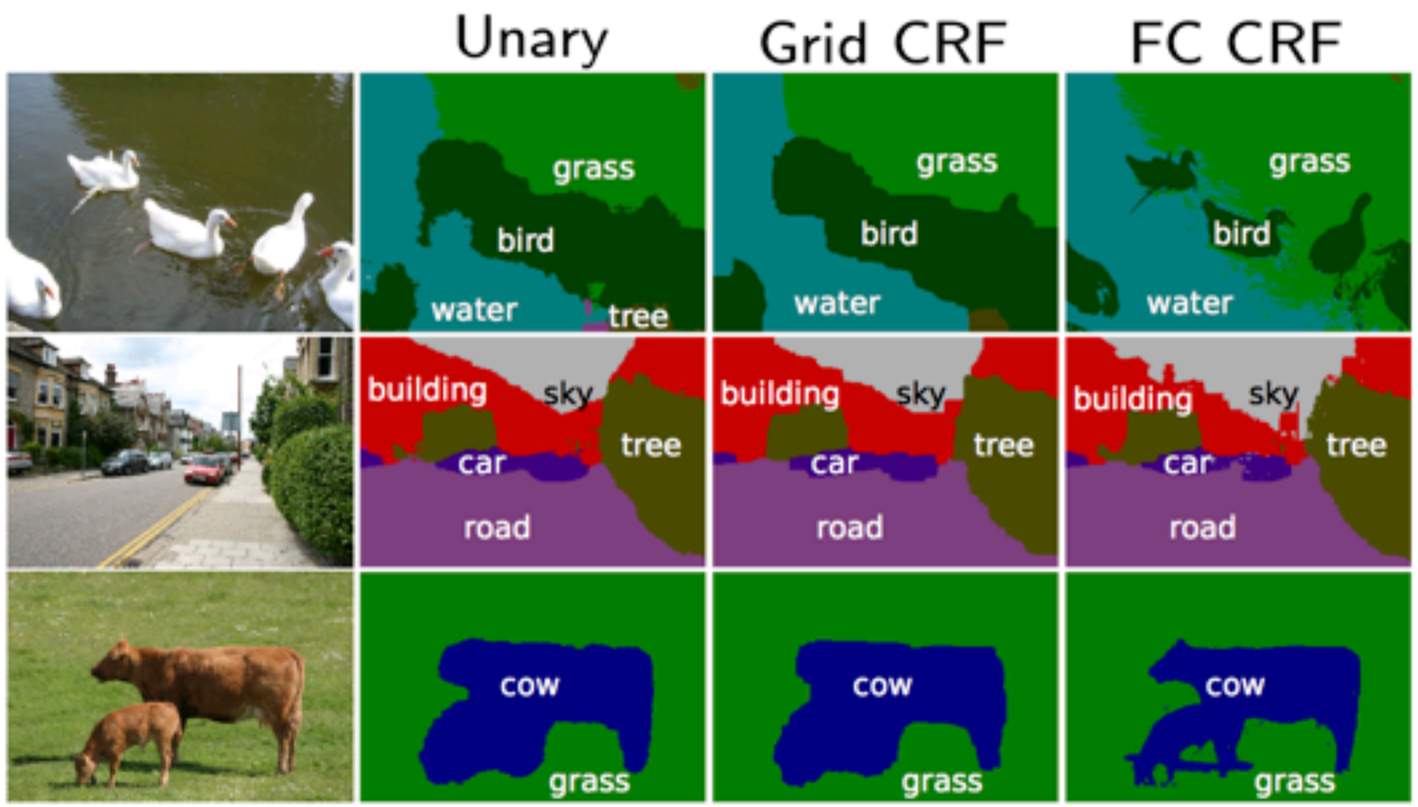
- Unary Potentials learned using Joint Boost
  - Allows classes to share boundaries and improves generalization
- Weights for pairwise filtering found via L-BGFS using expectation maximization
- Kernel Bandwidths hard to learn;
  - Grid search used to pick best value

Results: MSRC

MSRC dataset

- 591 images
- 21 classes

	Time	Global	Avg
Unary	-	84.0	76.6
Grid CRF	1s	84.6	77.2
FC CRF	**0.2s**	**86.0**	**78.3**

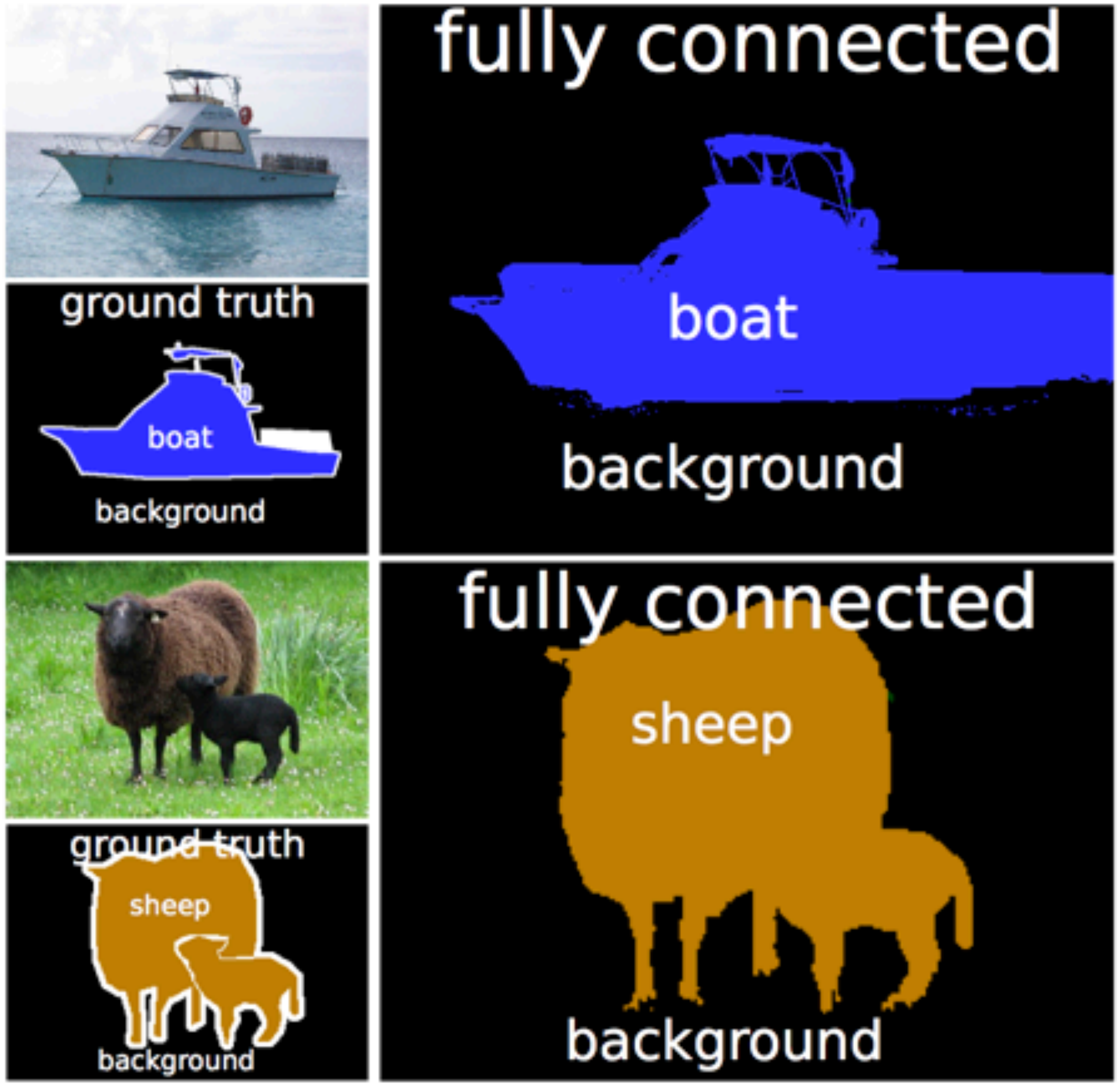


Results: PASCAL VOC 2010

PASCAL VOC 2010 dataset

- 1928 images
- 20 classes + background

	Time	Acc
Unary	-	27.6
Grid CRF	2.5s	28.3
FC Potts	0.5s	29.1
FC label comp	**0.5s**	**30.2**



Really Cool

Image | Grid CRF | Robust P^n CRF | Our approach | Accurate grou

- Fine feature segmentation in 0.2 seconds

Reported Failures

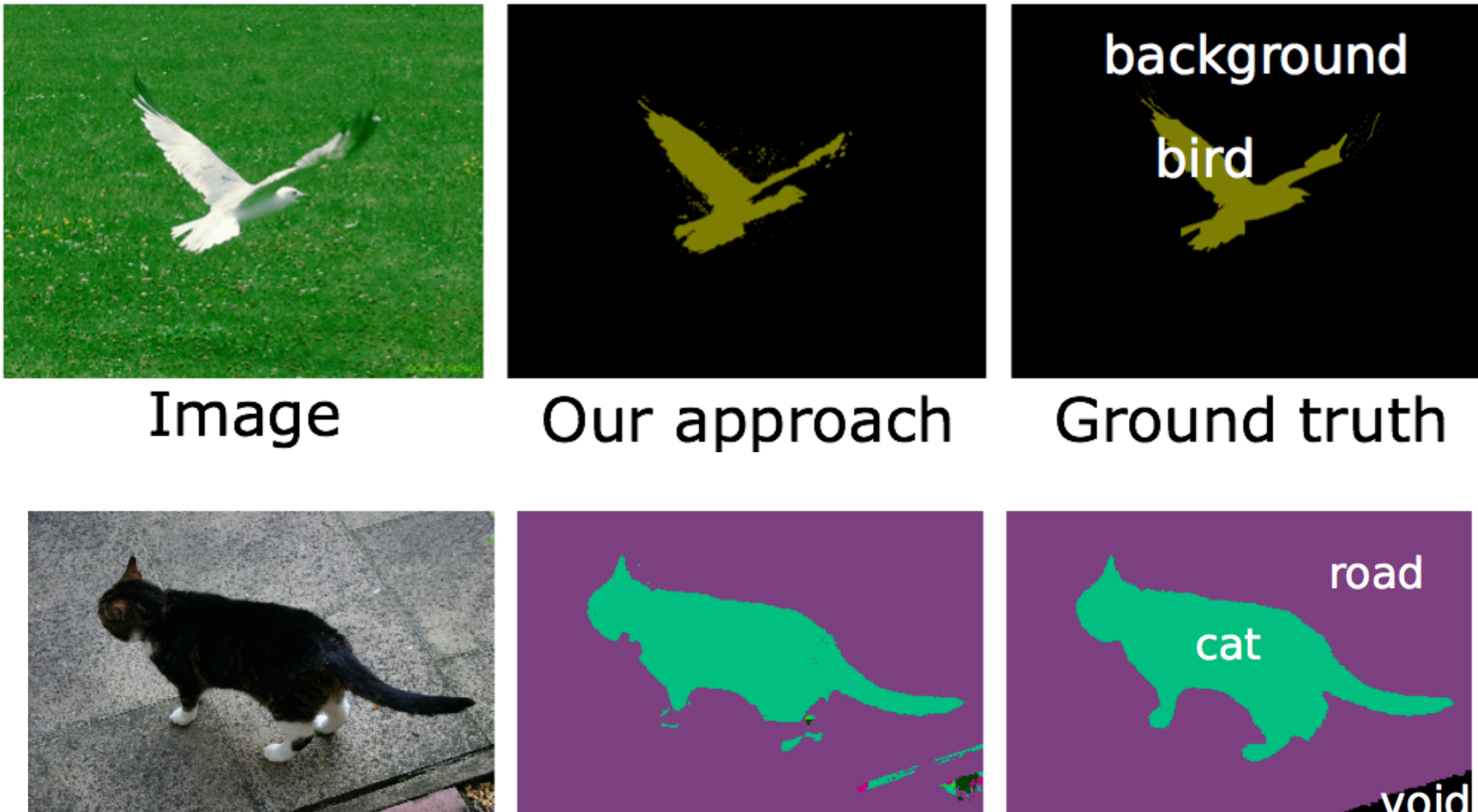
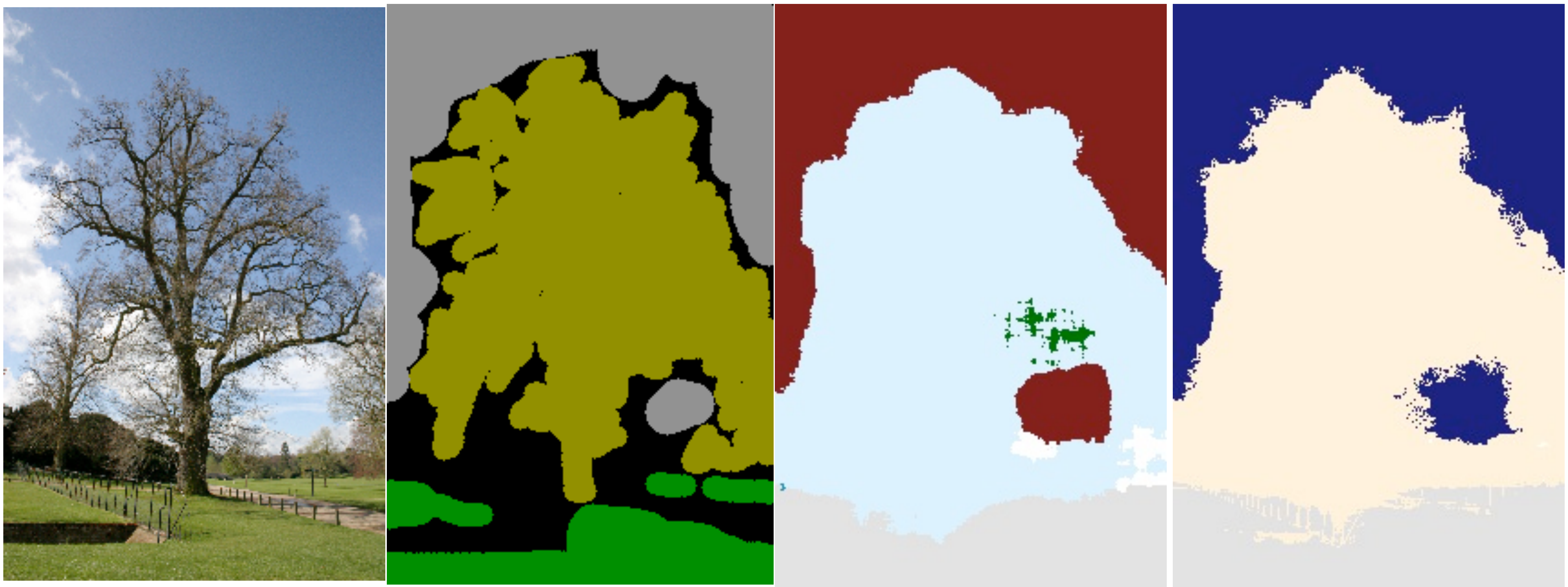


Image | Our approach | Ground truth

Image | Our approach | Ground truth

Replicating Results



ground truth unary crf

Replicating Results : overlay

Replicating Results