Object-class Segmentation using Higher order CRF

Ľubor Ladický[1], Pushmeet Kohli[2]
Philip Torr[1]

[1]Oxford Brookes University
[2]Microsoft Research Cambridge

ECCV 2008

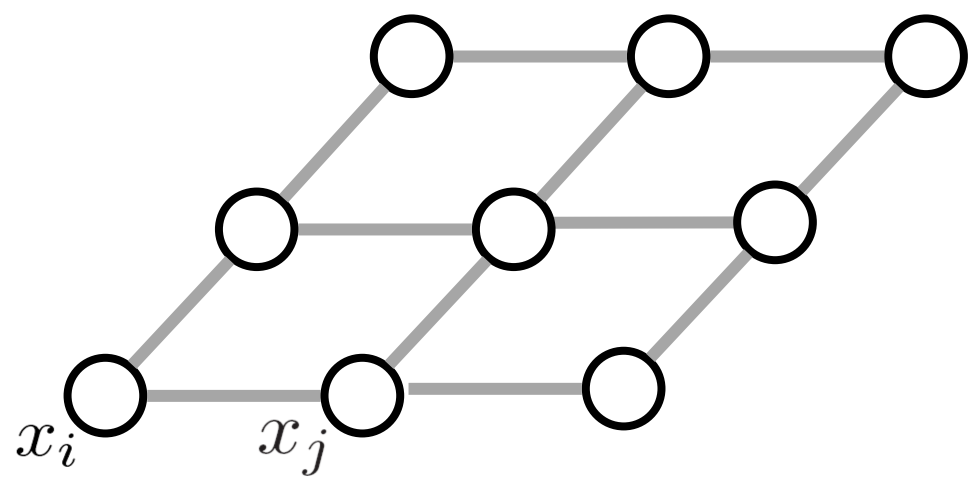
Object Segmentation using CRFs

CRF Energy

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

Labels - set of object classes
{ background, bird, car, ..}

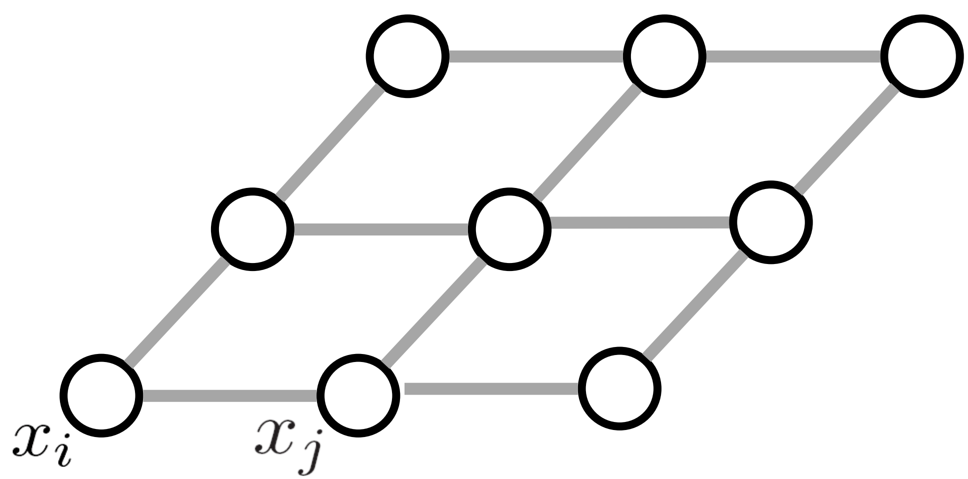
x_i - label of pixel i



Object Segmentation using CRFs

CRF Energy

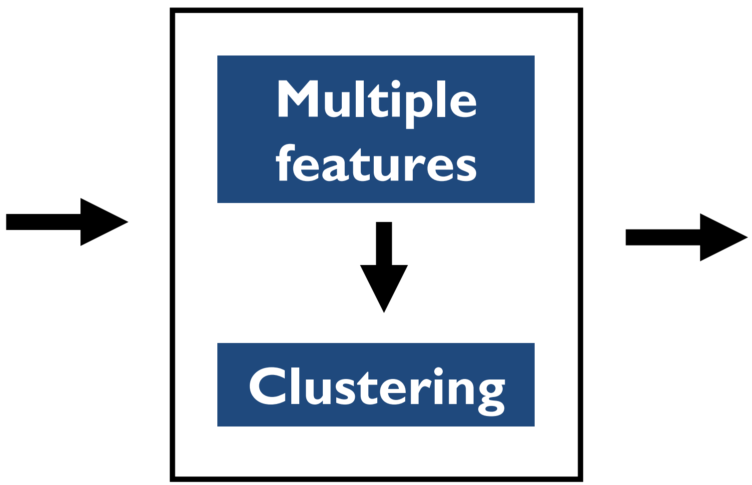
$$E(\mathbf{x}) = \underbrace{\sum_{i \in \mathcal{V}} \psi_i(x_i)}_{\text{Unary potentials based on Colour, Location and Texture features}} + \sum_{i \in \mathcal{V}, j \in \mathcal{N}_i} \psi_{ij}(x_i, x_j) + \sum_{c \in \mathcal{S}} \psi_c(\mathbf{x}_c)$$

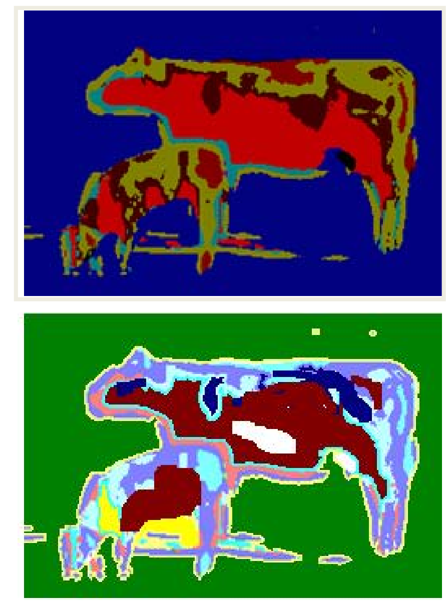


Multifeature unary potential

- Inspired by [Shotton et al. ECCV06]

Input image

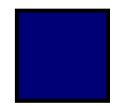
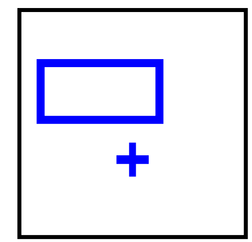




Feature maps
Colours ⇔ Cluster

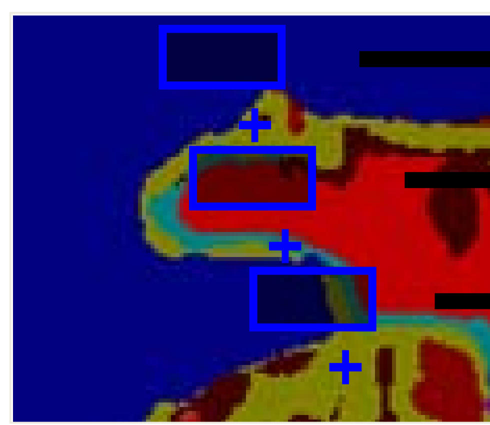
Multifeature unary potential

- Unary likelihood based on spatial configuration of features

Shape filter (texton colour location HOG ,  , )

feature type f   cluster *t*   rectangle *r*

models appearance context and shape



$v(i, \mathrm{f}, t, r) = a$

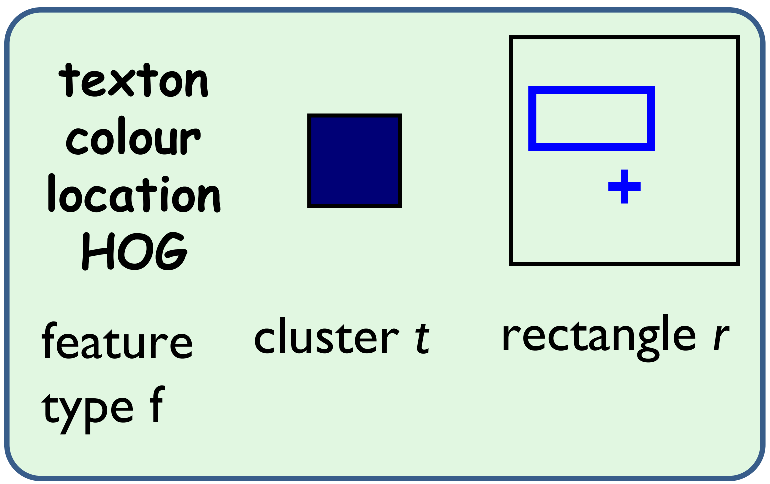
$v(i, \mathrm{f}, t, r) = 0$

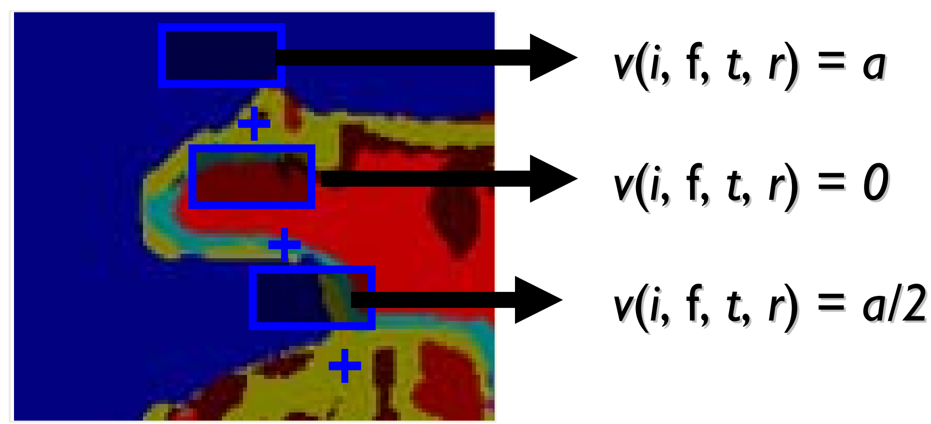
$v(i, \mathrm{f}, t, r) = a/2$

Multi-feature unary potential

Main differences [Shotton et al., ECCV06]

- Multiple features boosted together
- Gaussian distribution of rectangles
- Threshold set grows exponentially
- Memory compressions
- Weighted training samples



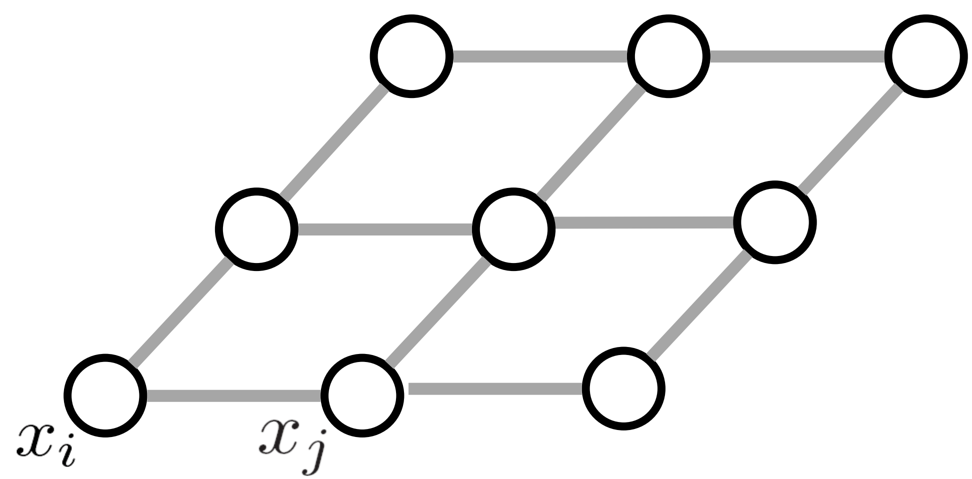


Models, appearance context, and shape

Object Segmentation using CRFs

CRF Energy

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \psi_i(x_i) + \underbrace{\sum_{i \in \mathcal{V}, j \in \mathcal{N}_i} \psi_{ij}(x_i, x_j)}_{\text{Encourages label consistency in adjacent pixels}} + \sum_{c \in \mathcal{S}} \psi_c(\mathbf{x}_c)$$



Contrast-sensitive pairwise potential

- Pairwise cost

$$\psi(x_i, x_j) = \begin{cases} 0 & \text{if } x_i = x_j, \\ g(i,j) & \text{otherwise}, \end{cases}$$

$$g(i,j) = \theta_p + \theta_v \exp(-\theta_\beta ||I_i - I_j||^2)$$

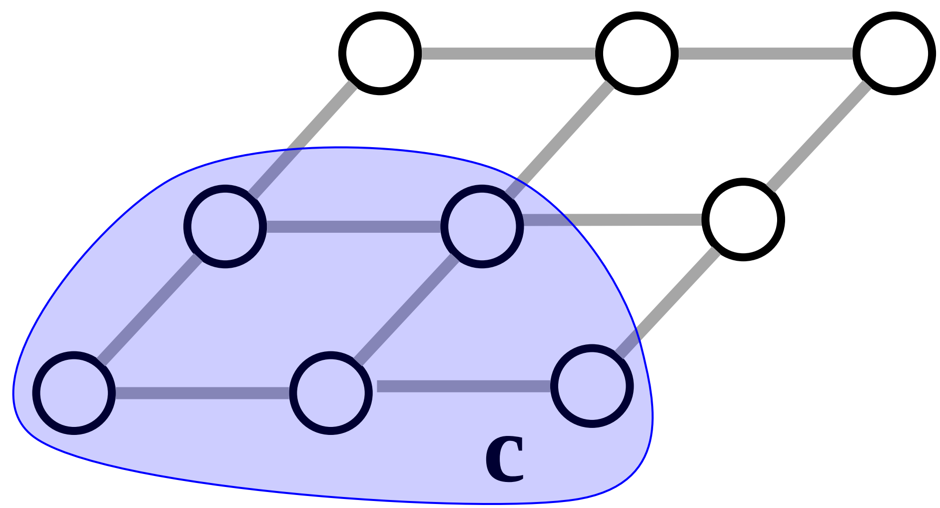
- Enforces label consistency in adjacent pixels

Higher Order CRF Model

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \psi_i(x_i) + \sum_{i \in \mathcal{V}, j \in \mathcal{N}_i} \psi_{ij}(x_i, x_j) + \underbrace{\sum_{c \in \mathcal{S}} \psi_c(\mathbf{x}_c)}_{\text{Encourages label consistency in regions}}$$

Multiple Segmentations

Comaniciu and Meer PAMI 2002
Shi and Malik PAMI 2000
Felzenszwalb and Huttenlocher IJCV 2004

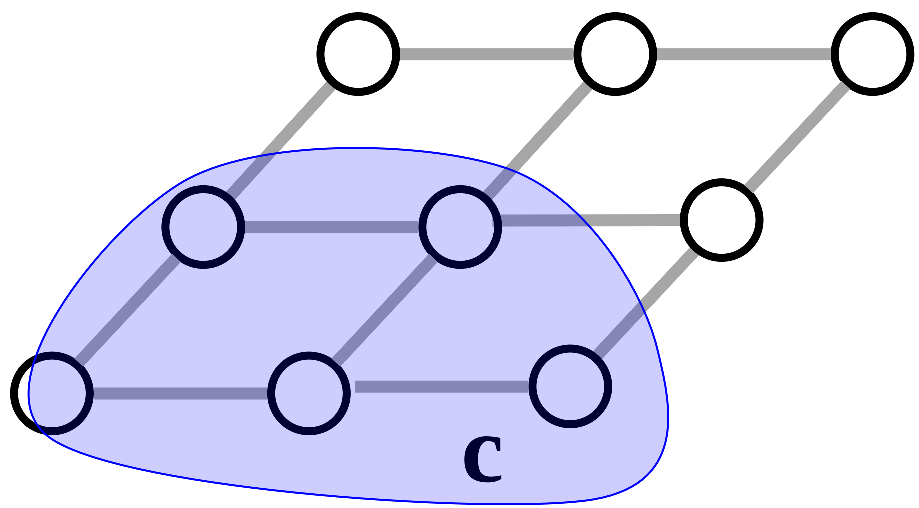


Label Consistency in Segments

- Encourages consistency within super-pixels
- Takes the form of a P^N Potts model

  [Kohli *et al*. CVPR 2007]

$$\psi_c^p(\mathbf{x}_c) = \begin{cases} 0 & \text{if } \; x_i = l_k, \forall i \in c \\ f(|c|) & \text{otherwise} \end{cases}$$

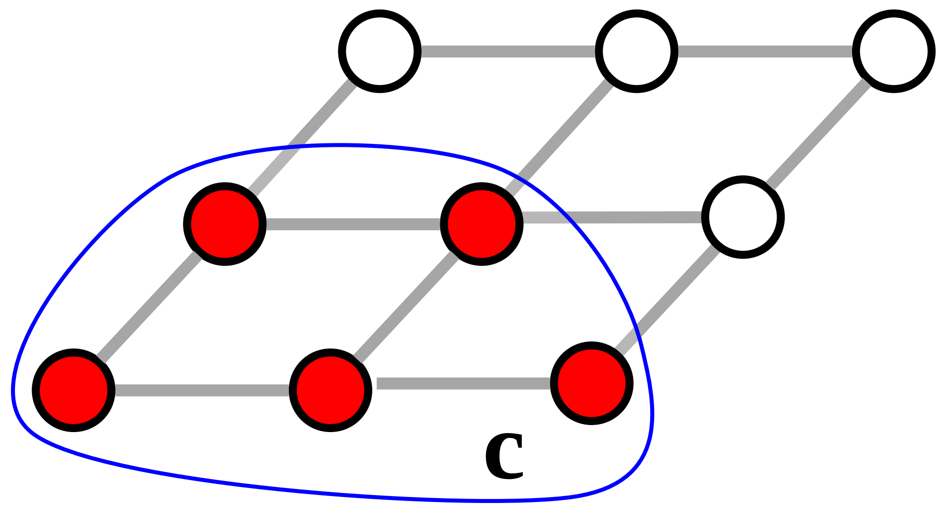


Label Consistency in Segments

- Encourages consistency within super-pixels
- Takes the form of a P^N Potts model

  [Kohli *et al*. CVPR 2007]

$$\psi_c^p(\mathbf{x}_c) = \begin{cases} 0 & \text{if } \; x_i = l_k, \forall i \in c \\ f(|c|) & \text{otherwise} \end{cases}$$



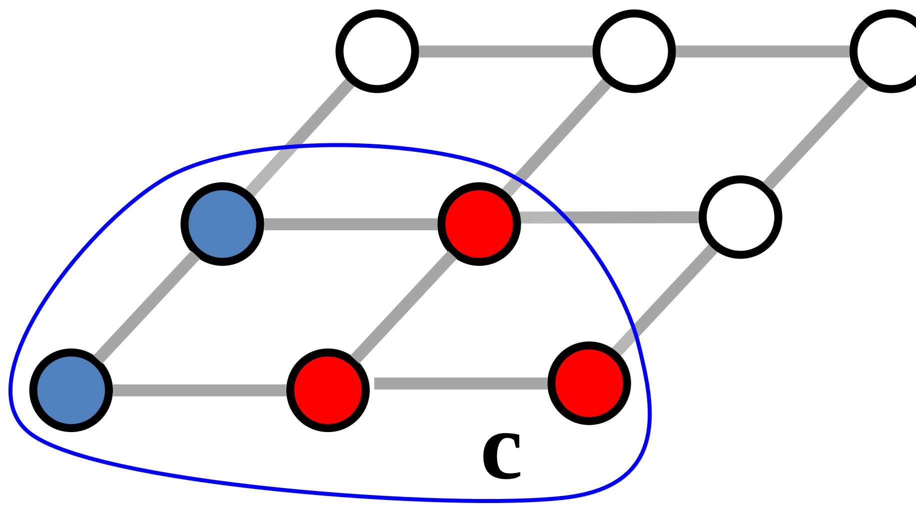
Cost: 0

Label Consistency in Segments

- Encourages consistency within super-pixels
- Takes the form of a P^N Potts model

[Kohli *et al*. CVPR 2007]

$$\psi_c^p(\mathbf{x}_c) = \begin{cases} 0 & \text{if } \; x_i = l_k, \forall i \in c \\ f(|c|) & \text{otherwise} \end{cases}$$



Cost: $f(|c|)$

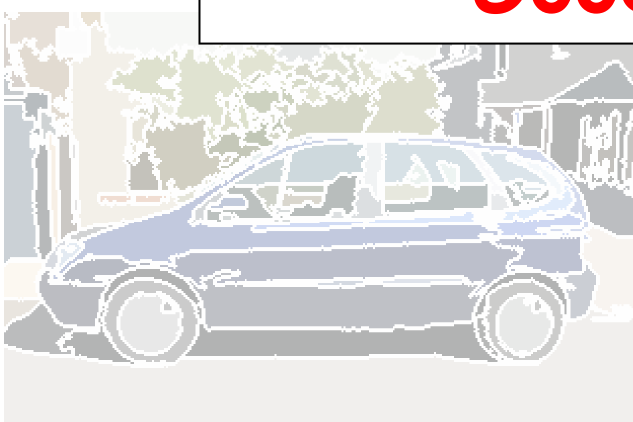
Label Consistency in Segments

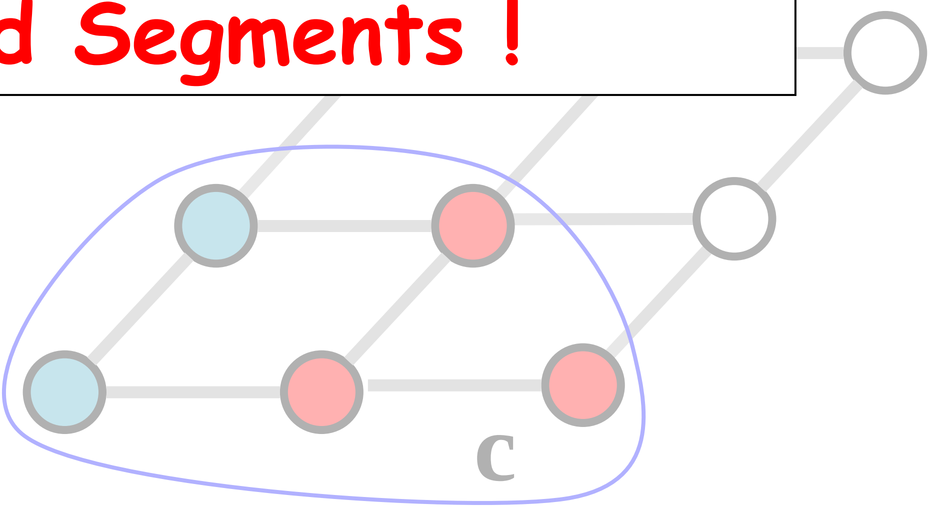
- Encourages consistency within super-pixels
- Takes the form of a P^N Potts model

  [Kohli *et al*. CVPR 2007]

$$\quad \begin{cases} 0 & \text{if } x_i = l_k, \forall i \in c \end{cases}$$

Does not distinguish between Good/Bad Segments !





Quality based Label Consistency

Label inconsistency cost depends on segment quality

$$\psi_c^v(\mathbf{x}_c) = \begin{cases} 0 & \text{if } x_i = l_k, \forall i \in c \\ \theta_p + \theta_v G(c) & \text{otherwise} \end{cases}$$

Quality based Label Consistency

Label inconsistency cost depends on segment quality

$$\psi_c^v(\mathbf{x}_c) = \begin{cases} 0 & \text{if } x_i = l_k, \forall i \in c \\ \theta_p + \theta_v G(c) & \text{otherwise} \end{cases}$$

How to measure quality G(c)?

[Ren and Malik ICCV03, Rabinovich et al. ICCV07, many others]

- Colour and Texture Similarity
- Contour Energy

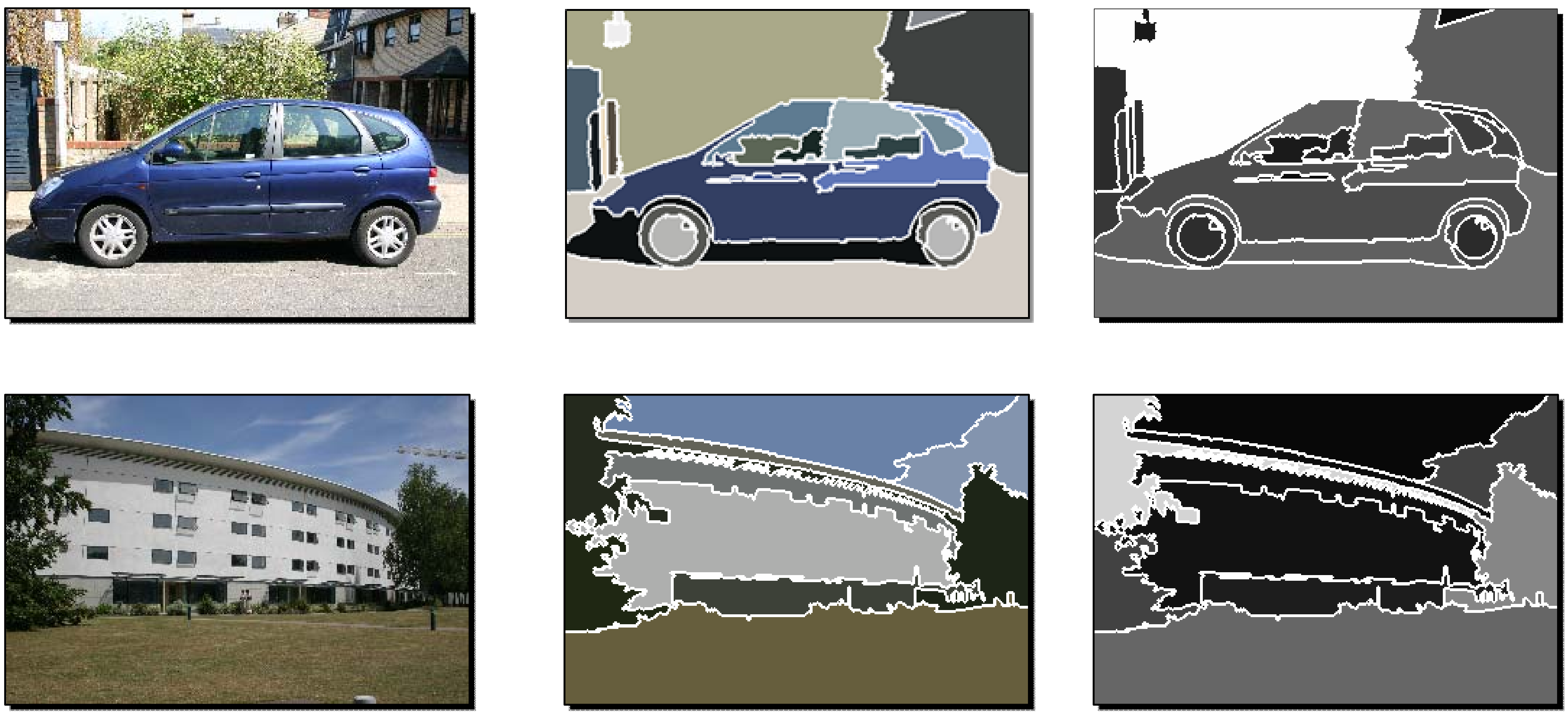
Measure quality from variance in feature responses

$$G(c) = \exp(-\theta_\beta f_v(c))$$

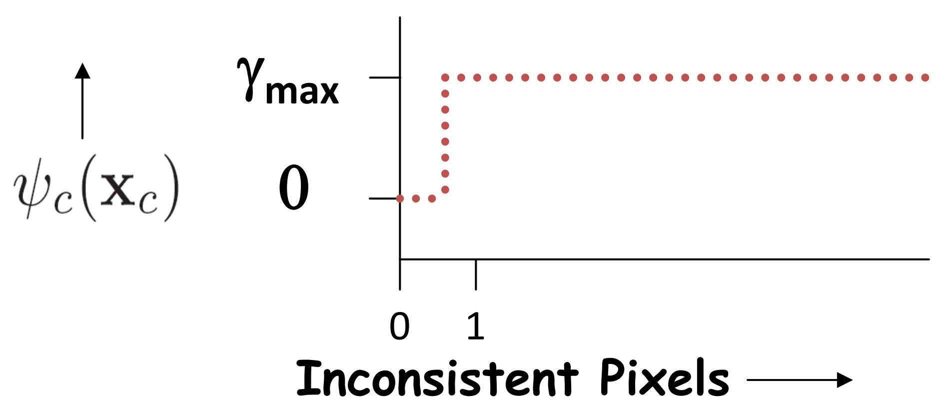
Higher order generalization of contrast-sensitive pairwise potential

Quality based Label Consistency

MSRC image	Mean shift segmentation	Segment Quality (darker is better)

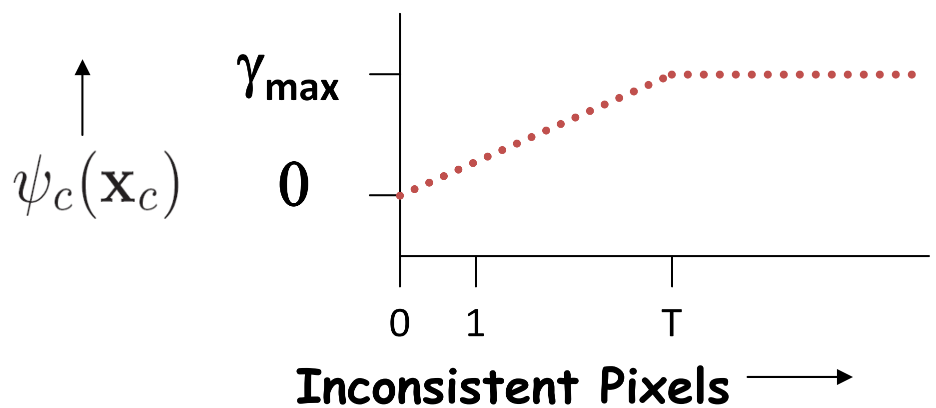


Robust Consistency Potentials



P^N Potts

Too Rigid!



Robust

Kohli, Kumar, Torr, CVPR

Kohli, Ladicky, Torr, CVPR 2008

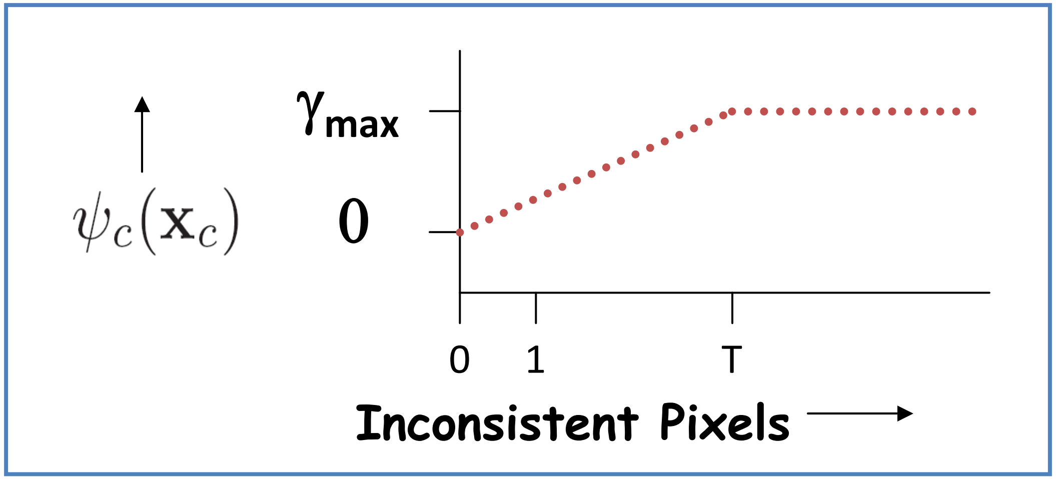
Robust Consistency Potentials

$$\psi_c(\mathbf{x}_c) = \min\{ N_i(\mathbf{x}_c)\, Q\ ,\gamma_{\max}\}$$

Number of Inconsistent Pixels

Slope

Maximum Inconsistency Cost



Robust

Robust Consistency Potentials

$$\psi_c(\mathbf{x}_c) = \min\{ N_i(\mathbf{x}_c)\, Q \;, \gamma_{\max} \}$$

Number of Inconsistent Pixels

Slope

Maximum Inconsistency Cost

- **Some pixels are less important than others (for instance, Segment boundary pixels)**

- **Weighted version of the inconsistency count**

Minimizing Higher order Energy Functions

- **Message passing is computationally expensive**
  - High runtime and space complexity - $O(L^N)$
  - L = Number of Labels, N = Size of Clique

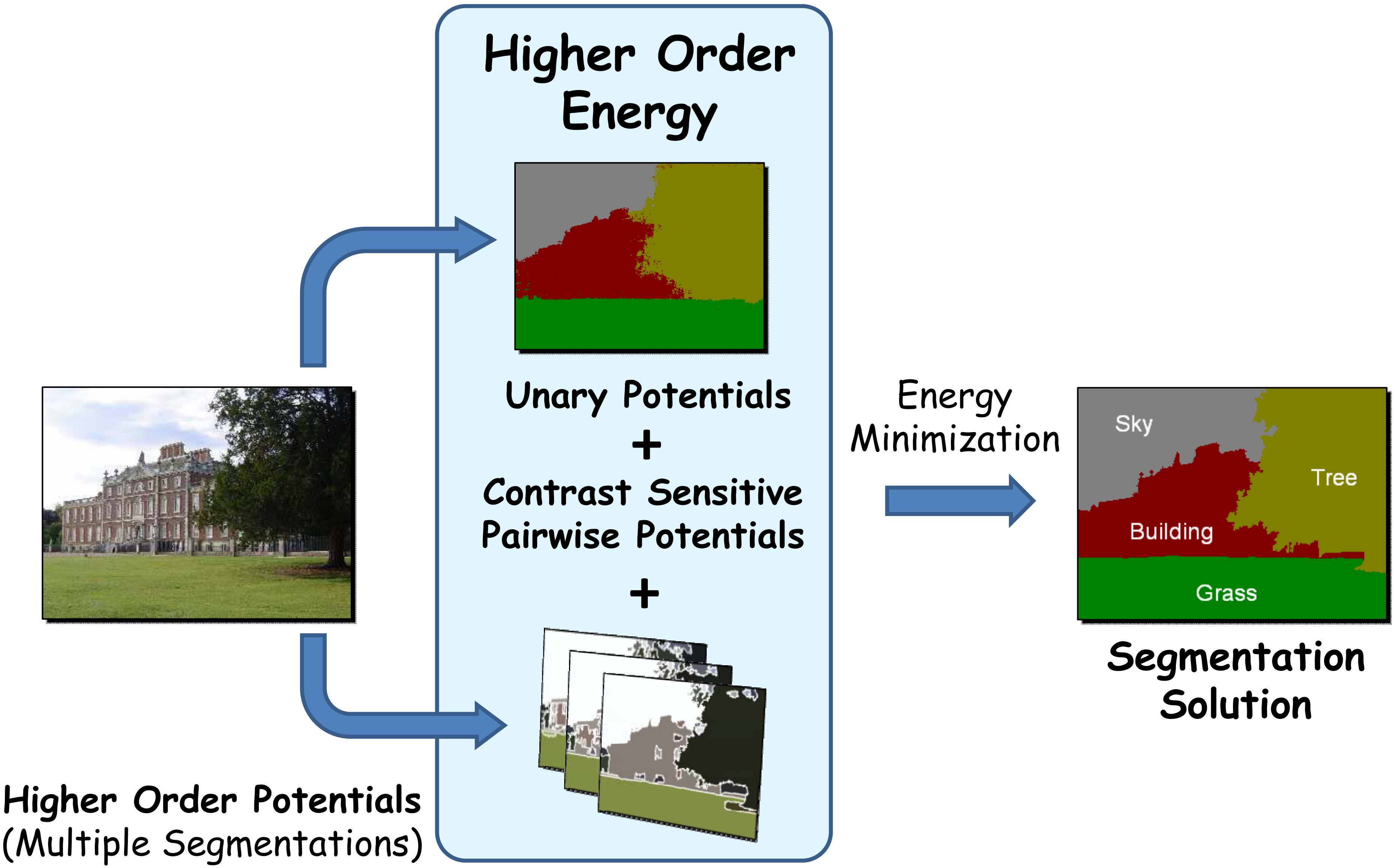
- **Efficient BP for Higher Order MRFs**
  [Lan et al. ECCV 06, Potetz CVPR 2007]
  - 2x2 clique potentials for Image Denoising
  - Take minutes per iteration (Hours to converge)

Graph Cuts for Minimizing Higher order Energy Functions (Our Approach)

- Binary label problems can be solved exactly
- Can handle very high order energy functions
- Extremely efficient: computation time in the order of seconds
- Graph Cut based move making algorithm for Multilabel Functions
  [Kohli, Ladický, Torr - CVPR08]

- Only applicable to some classes of functions (described earlier)

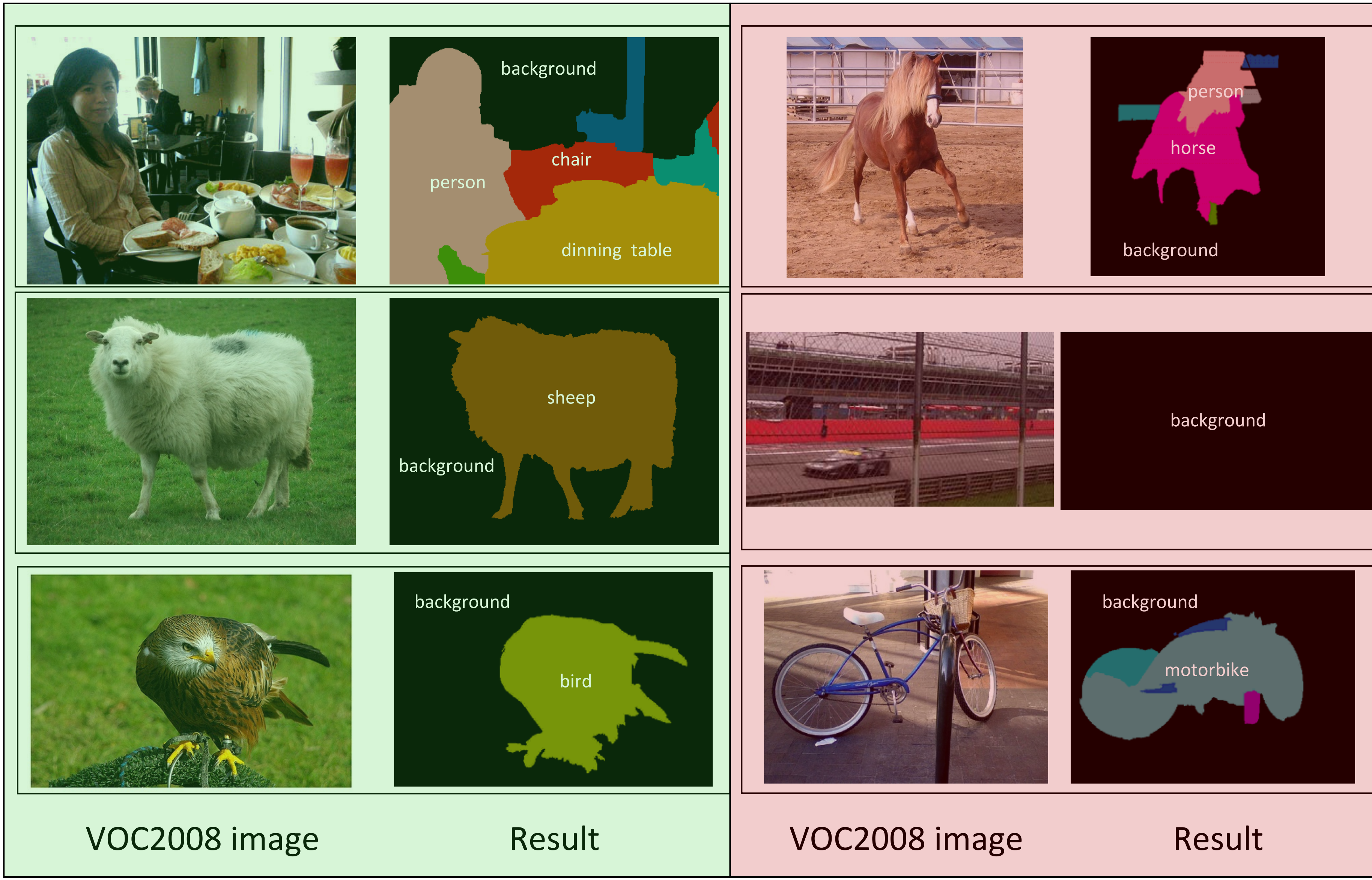
Overview of our Method



Results

aeroplane
background
person
horse
background
background
train
background
person
dinning table
background
car
background
bird
VOC2008 image
Result
VOC2008 image
Result

Results


background
chair
person
dinning table
person
horse
background
sheep
background
background
background
bird
background
motorbike
VOC2008 image
Result
VOC2008 image
Result

Conclusions

- Improvement of boosted unary potential
- Method to enforce label consistency in image regions
- Generalization of the commonly used pair-wise contrast sensitive potential
- Allows integration of pixel and region level features for image labelling problems

Conclusions

- Bottom-Up approach showed its potential
- Performance can be boosted using state-of the art descriptors and more tunning
- Future work
  - More solvable energies
  - More features
  - More tunning
- We want background classes!

Thank you

Questions ?