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

- Don’t make super pixels
Operate On Pixels Directly

- Pixel wise classification - texture/local shape features

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

- Each Clique Modeled as Gibbs Distribution

\[
\text{Pr}(x|D) = \frac{1}{Z} \exp \left( - \sum_{c \in C} \psi_c(x_c) \right),
\]

- Unary Potentials and Pairwise potential

\[
E(x) = \sum_i \psi_u(x_i) + \sum_i \sum_{j \in N_i} \psi_p(x_i, x_j) + \text{(optional higher order terms)}
\]

- Maximum-a-posteriori solutions are NP-Hard

- Message Passing algorithms: belief propagation

- Move Making Algorithms: \( \alpha \)-expansion, \( \alpha \beta \)-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

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

- 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

\[ E(x) = \sum_i \psi(u(x_i)) + \sum_i \sum_{j \in N_i} \psi(p(x_i, x_j)) \]

- **Efficient inference**
  - 1 second for 50,000 variables
- **Limited expressive power**
- **Only local interactions**
- **Excessive smoothing of object boundaries**
  - Shrinking bias
Adjacency CRF models

\[ E(x) = \sum_i \psi_u(x_i) + \sum_i \sum_{j \in N_i} \psi_p(x_i, x_j) \]

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

\[ E(x) = \sum_i \psi_u(x_i) + \sum_i \sum_{j \in N_i} \psi_p(x_i, x_j) \]

- **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

\[ E(x) = \sum_{i \in V} \psi_i(x_i) + \sum_{(i,j) \in E} \psi_{ij}(x_i, x_j) + \sum_{c \in S} \psi_c(x_c), \]

  unary  pairwise  super-pixel

- Super pixel term also models consistency within super pixel
Model definition

\[ E(x) = \sum_i \psi_u(x_i) + \sum_i \sum_{j>i} \psi_p(x_i, x_j) \]

- **Unary term** \( \psi_u(x_i) \)
- **Pairwise term** \( \psi_p(x_i, x_j) \)

**Gaussian edge potentials**

\[ \psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^{K} w^{(m)} k^{(m)}(f_i, f_j) \]

- **Label compatibility function** \( \mu \)
- **Linear combination of Gaussian kernels**

\[ k^{(m)}(f_i, f_j) = \exp\left(-\frac{1}{2}(f_i - f_j)\Sigma^{(m)}(f_i - f_j)\right) \]

- **Arbitrary feature space** \( f_i \)
Model definition

\[ E(x) = \sum_i \psi_u(x_i) + \sum_i \sum_{j>i} \psi_p(x_i, x_j) \]

Gaussian edge potentials

\[ \psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^{K} w^{(m)} k^{(m)}(f_i, f_j) \]

- Label compatibility function \( \mu \)
- Linear combination of Gaussian kernels

\[ k^{(m)}(f_i, f_j) = \exp\left(-\frac{1}{2}(f_i - f_j)\Sigma^{(m)}(f_i - f_j)\right) \]

- Arbitrary feature space \( f_i \)
Model definition

\[ E(x) = \sum_i \psi_u(x_i) + \sum_i \sum_{j>i} \psi_p(x_i, x_j) \]

Gaussian edge potentials

\[ \psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^K w^{(m)} k^{(m)}(f_i, f_j) \]

- Label compatibility function \( \mu \)
- Linear combination of Gaussian kernels

\[ k^{(m)}(f_i, f_j) = \exp\left(-\frac{1}{2}(f_i - f_j)\Sigma^{(m)}(f_i - f_j)\right) \]

- Arbitrary feature space \( f_i \)
Model definition

\[ E(x) = \sum_i \psi_u(x_i) + \sum_i \sum_{j > i} \psi_p(x_i, x_j) \]

Gaussian edge potentials

\[ \psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^K w^{(m)} k^{(m)}(f_i, f_j) \]

- Label compatibility function \( \mu \)
- Linear combination of Gaussian kernels

\[ k^{(m)}(f_i, f_j) = \exp\left(-\frac{1}{2}(f_i - f_j)\Sigma^{(m)}(f_i - f_j)\right) \]

- Arbitrary feature space \( f_i \)
Model definition

\[ E(\mathbf{x}) = \sum_i \psi_u(x_i) + \sum_i \sum_{j>i} \psi_p(x_i, x_j) \]

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 \( \mu \)
- Linear combination of Gaussian kernels

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

- Arbitrary feature space \( \mathbf{f}_i \)

Convolution is key to efficiency
Detailed model definition

\[ k(f_i, f_j) = w^{(1)} \exp \left( -\frac{|p_i - p_j|^2}{2\theta^2_\alpha} - \frac{|I_i - I_j|^2}{2\theta^2_\beta} \right) + w^{(2)} \exp \left( -\frac{|p_i - p_j|^2}{2\theta^2_\gamma} \right) \]

- Appearance kernel
- 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
\[ k(\mathbf{f}_i, \mathbf{f}_j) = w^{(1)} \exp \left( -\frac{|p_i - p_j|^2}{2\theta^2_\alpha} - \frac{|I_i - I_j|^2}{2\theta^2_\beta} \right) + w^{(2)} \exp \left( -\frac{|p_i - p_j|^2}{2\theta^2_\gamma} \right) \]

appearance kernel

smoothness kernel
\[ k(f_i, f_j) = w^{(1)} \exp \left( -\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2} \right) - w^{(2)} \exp \left( -\frac{|p_i - p_j|^2}{2\theta_\gamma^2} \right) \]

appearance kernel

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

<table>
<thead>
<tr>
<th>Method</th>
<th>Time</th>
<th>Global</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unary</td>
<td>-</td>
<td>84.0</td>
<td>76.6</td>
</tr>
<tr>
<td>Grid CRF</td>
<td>1s</td>
<td>84.6</td>
<td>77.2</td>
</tr>
<tr>
<td>FC CRF</td>
<td>0.2s</td>
<td>86.0</td>
<td>78.3</td>
</tr>
</tbody>
</table>
Results: PASCAL VOC 2010

PASCAL VOC 2010 dataset
- 1928 images
- 20 classes + background

<table>
<thead>
<tr>
<th></th>
<th>Time</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unary</td>
<td>-</td>
<td>27.6</td>
</tr>
<tr>
<td>Grid CRF</td>
<td>2.5s</td>
<td>28.3</td>
</tr>
<tr>
<td>FC Potts</td>
<td>0.5s</td>
<td>29.1</td>
</tr>
<tr>
<td>FC label comp</td>
<td>0.5s</td>
<td>30.2</td>
</tr>
</tbody>
</table>
Really Cool

- 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