Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials

Philipp Krähenbühl Vladlen Koltun
philkr@stanford.edu vladlen@stanford.edu

Stanford University

Overview

Most state-of-the-art techniques for multi-class image segmentation and labeling use conditional random fields defined over pixels or image regions. While region-level models often feature dense pairwise connectivity, pixel-level models are considerably larger and have only permitted sparse graph structures. In this paper, we consider **fully connected CRF models** defined on the complete set of **pixels** in an image. The resulting graphs have billions of edges, making traditional inference algorithms impractical. Our main contribution is a **highly efficient approximate inference algorithm for fully connected CRF models** in which the pairwise edge potentials are defined by a linear combination of Gaussian kernels. Our experiments demonstrate that dense connectivity at the pixel level substantially improves segmentation and labeling accuracy.

Model

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

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 $\mathbf{f}_i$

Multi-class image segmentation

Find a pixel level class labeling for an image

TextonBoost [3]

$\psi_u(x_i)$ learned from data

Color sensitive model (position $\mathbf{p}_i$ and color $\mathbf{c}_i$)

$$\psi_p(x_i, x_j) = \mu(x_i, x_j)\left(w^{(1)} \exp(-\frac{|\mathbf{p}_i - \mathbf{p}_j|}{2\theta_\alpha^2} - \frac{|I_i - I_j|}{2\theta_\beta^2}) + w^{(2)} \exp(-\frac{|\mathbf{p}_i - \mathbf{p}_j|}{2\theta_\gamma^2}) \right)$$

- Potts model $\mu(x_i, x_j) = 1_{[x_i \neq x_j]}$
- Semi-metric model: $\mu(x_i, x_j)$ learned from data

Mean-field approximation

Find the most likely assignment (MAP)

$$\hat{x} = \operatorname*{argmax}_{\mathbf{x}} P(\mathbf{x}) \qquad \text{where} \qquad P(\mathbf{x}) = \exp(-E(\mathbf{x}))$$

Mean-field approximation

- Find $Q(\mathbf{x}) = \prod_i Q(x_i)$ close to $P(\mathbf{x})$ using KL-divergence $D(Q\|P)$
- $\hat{x}_i \approx \operatorname{argmax}_{x_i} Q(x_i)$

Initialize Q $\quad\triangleright$ $Q_i(x_i) \leftarrow \frac{1}{Z_i}\exp\{-\phi_u(x_i)\}$
while not converged **do**
  $\tilde{Q}_i^{(m)}(l) \leftarrow \sum_{j \neq i} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) Q_j(l)$ for all m $\quad\triangleright$ **Message passing**
  $\hat{Q}_i(x_i) \leftarrow \sum_{l \in \mathcal{L}} \mu^{(m)}(x_i, l) \sum_m w^{(m)} \tilde{Q}_i^{(m)}(l)$ $\quad\triangleright$ **Compatibility transform**
  $Q_i(x_i) \leftarrow \exp\{-\psi_u(x_i) - \hat{Q}_i(x_i)\}$ $\quad\triangleright$ **Local update**
  normalize $Q_i(x_i)$
end while

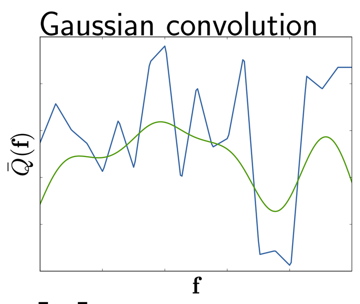
Message passing using filtering

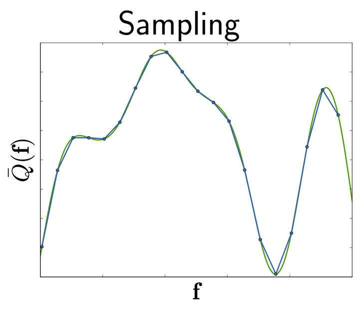
Update all $\tilde{Q}_i^{(m)}(l)$ simultaneously

$$\tilde{Q}_i^{(m)}(l) = \sum_{j \neq i} \exp\left(-\frac{1}{2}(\mathbf{f}_i^{(m)} - \mathbf{f}_j^{(m)})^2\right) Q_j(l)$$

Efficiently computed using a cross-bilateral filter [2, 1]

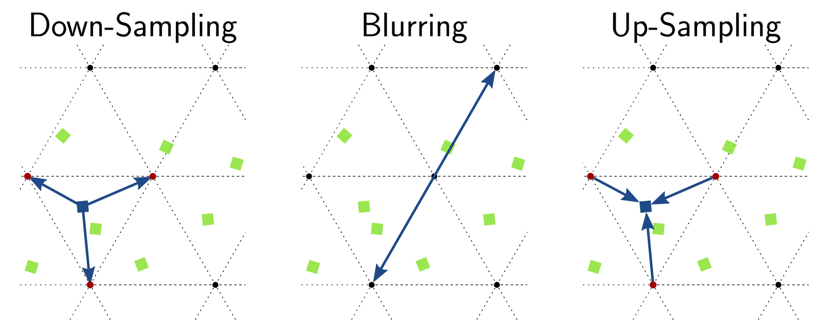
- Gaussian is band-limiting $\tilde{Q}_i^{(m)}(l)$
- $\tilde{Q}_i^{(m)}(l)$ is smooth
- well reconstructed by sparse samples





Evaluated using the permutohedral lattice [1]

- Down-sample $Q_j(l)$ in high dimensional space
- Compute Gaussian convolution on samples
- Up-sample $\tilde{Q}_i^{(m)}(l)$

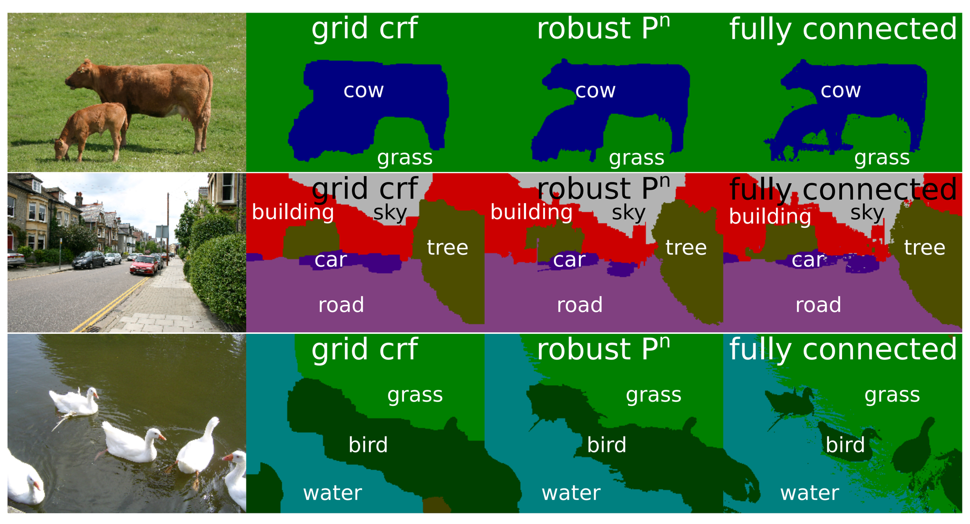


Results

MSRC dataset

- 591 images
- 21 classes

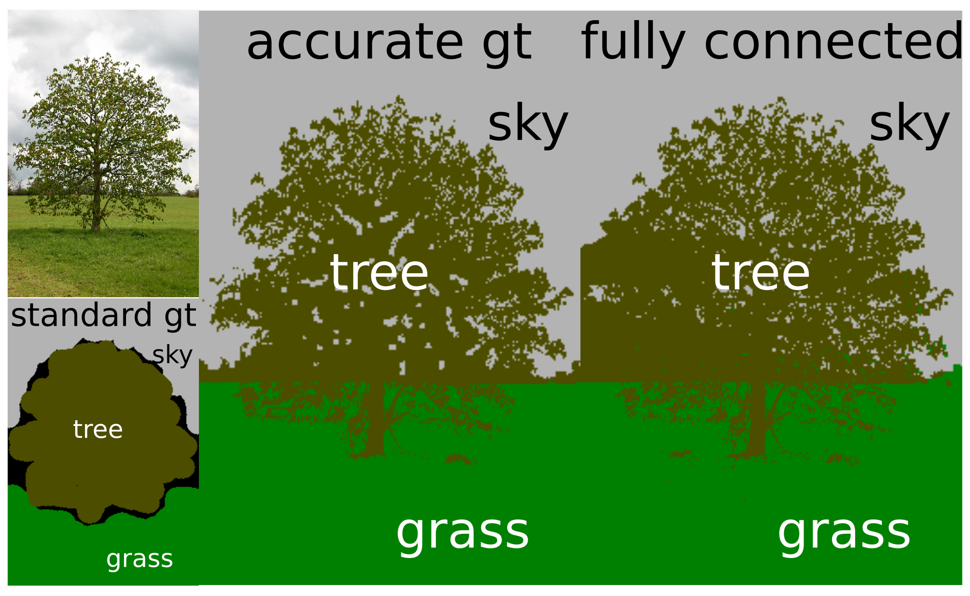
	Time	Global	Avg
Unary	-	84.0	76.6
Grid CRF	1s	84.6	77.2
Robust P^n	30s	84.9	77.5
FC CRF	**0.2s**	**86.0**	**78.3**



MSRC accurate annotations

- 94 manually annotated images
- 5-fold cross validation

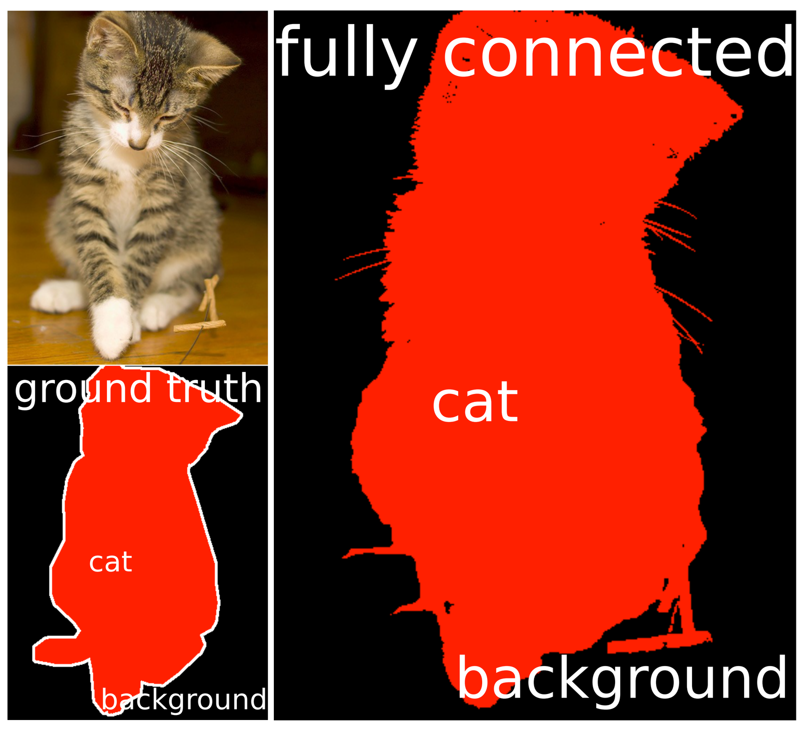
	Global	Avg
Unary	83.2 ± 1.5	80.6 ± 2.3
Grid CRF	84.8 ± 1.5	82.4 ± 1.8
Robust P^n	86.5 ± 1.0	83.1 ± 1.5
FC CRF	$\mathbf{88.2 \pm 0.7}$	$\mathbf{84.7 \pm 0.7}$



PASCAL VOC 2010 dataset

- 1928 images
- 20 classes + background
- μ learned from data

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



References

[1] Andrew Adams, Jongmin Baek, and Myers Abraham Davis.
Fast high-dimensional filtering using the permutohedral lattice.
Computer Graphics Forum, 29(2), 2010.

[2] Sylvain Paris and Frédo Durand.
A fast approximation of the bilateral filter using a signal processing approach.
IJCV, 81(1), 2009.

[3] Jamie Shotton, John M. Winn, Carsten Rother, and Antonio Criminisi.
Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context.
IJCV, 81(1), 2009.