Dense CRF for Satellite Images

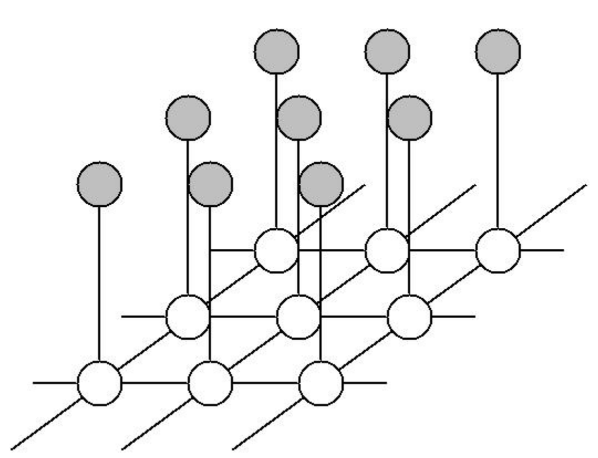
Pedro Amaral, Raphael Townshend, and Avideh Zakhor
UC Berkeley

Overview

In this project we used a densely connected graphical model to detect building rooftops in satellite imagery. Specifically, we applied a fully connected conditional random field defined on the complete set of pixels of an image as proposed by Krahenbuhl et al. [1], with unary potentials derived from Shotton et al.'s TextonBoost [2]. Inference in this model is made tractable by using mean-field approximation and using Gaussian kernels for the pairwise potentials, allowing for an efficient message-passing algorithm.

Model

Traditionally Square-Lattice Ising models have been used for image segmentation. This means we only have edges between neighboring pixels. This method fails to model longer range interactions between nodes, and leads to excessive smoothing. Some researchers have used higher order potentials to combat this problem, and others have created fully connected models on superpixels. Instead we a used dense CRF which has edges between all pixels.



Adjacency CRF

The energy for our model is:

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

Where the pairwise potential is defined as:

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

μ is the label compatibility function and the $k^{(m)}$ are the Gaussian kernels. For the label compatibility function we use the simple Potts model:

$$\mu(x_i, x_j) = [x_i \neq x_j]$$

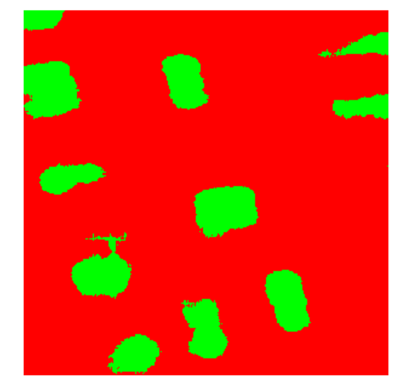
The specific kernels we use are (I is color intensity, p is position):

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

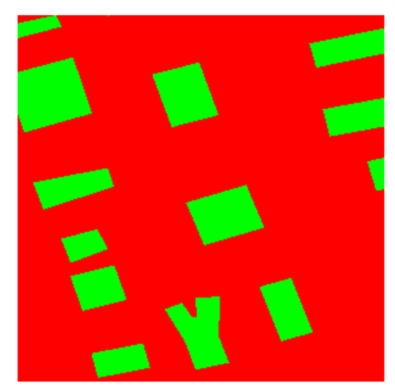
The unary potentials are learned with TextonBoost as well as HOG, color, and spatial location features as proposed by Ladicky et al. [3].

Test Results

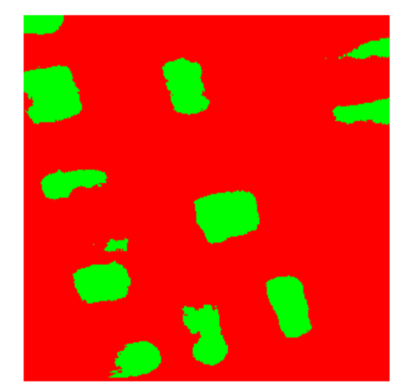
Original Image



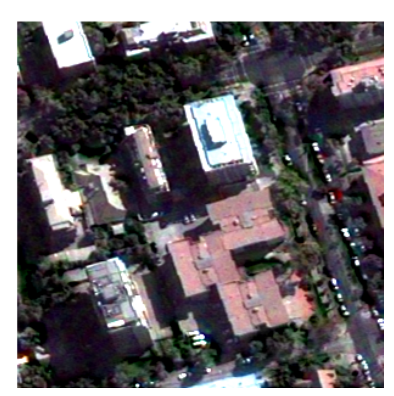
TextonBoost Output



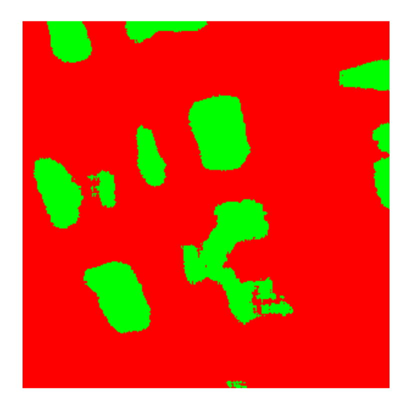
Ground Truth



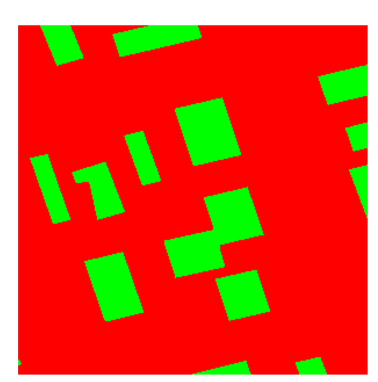
Dense CRF Output



Original Image



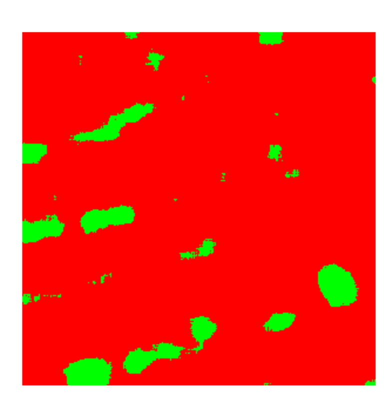
TextonBoost Output



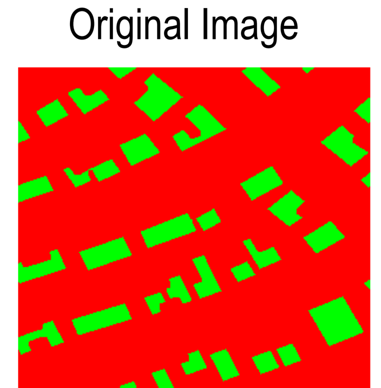
Ground Truth

Dense CRF Output

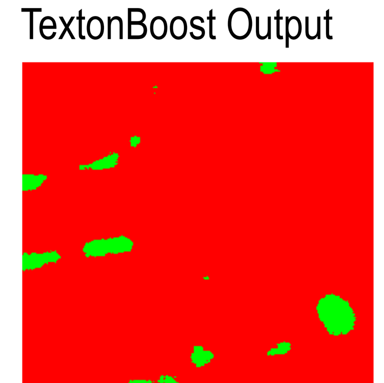
Original Image



TextonBoost Output



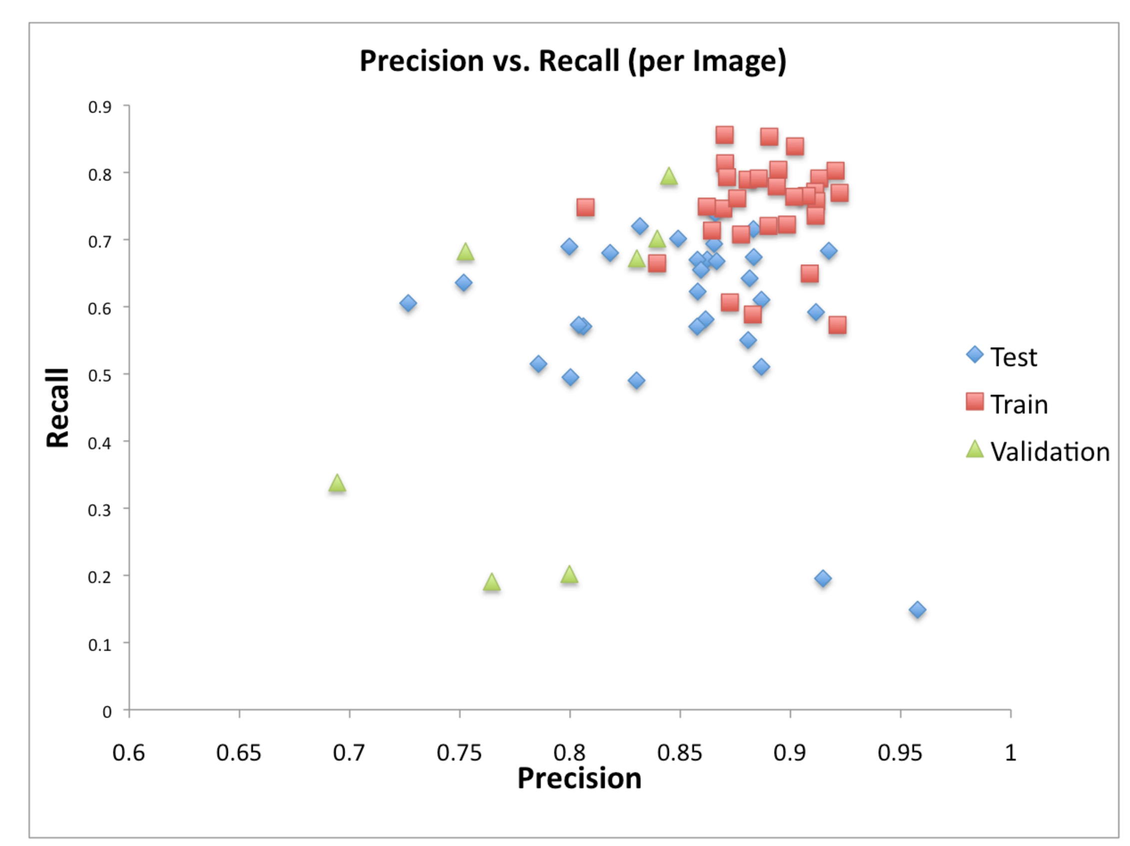
Ground Truth

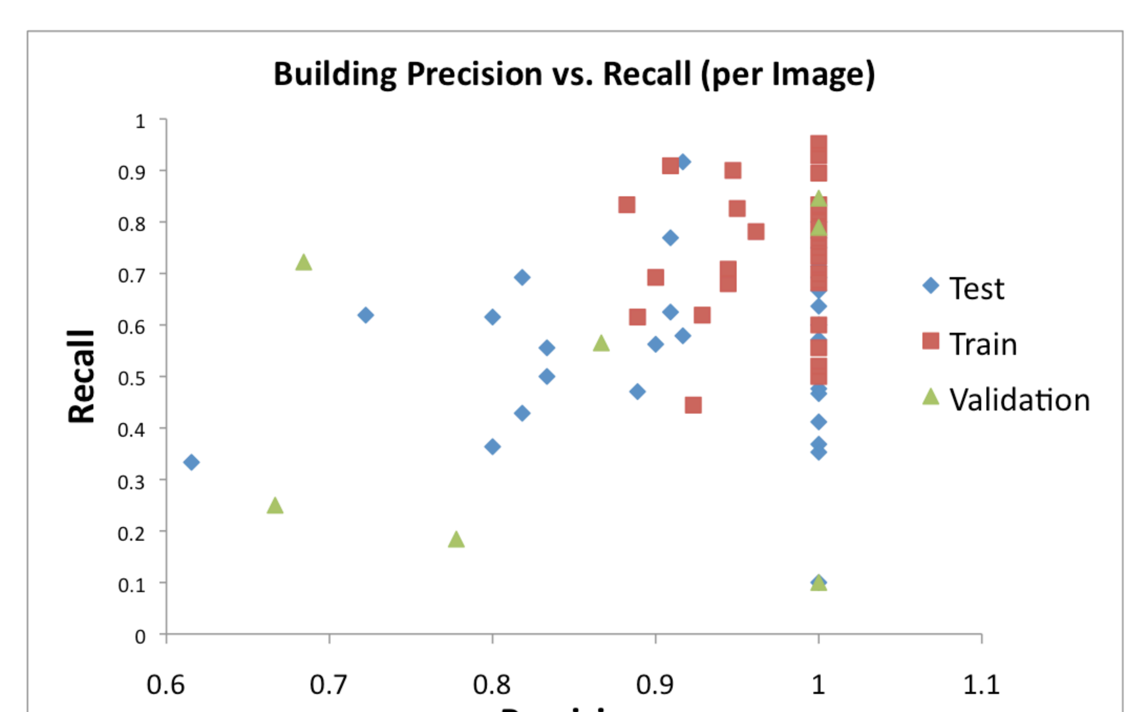


Dense CRF Output

Precision and Recall

Approach	Pix Prec	Pix Rec	Bld Prec	Bld Rec	# Bldgs
Lefevre	63.60%	79.40%	X	X	X
Muller	77.3%	79.5%	X	X	240
Persson	53.0%	93.0%	82.0%	X	17
Sirmacek	X	X	86.6%	X	177
Liu	X	X	94.5%	83.4%	277
Benediktsson	80.0%	X	X	X	8952
Shorter	78.7%	51.6%	55.4%	48.2%	2643
Ours	**85.0%**	**58.0%**	**92.0%**	**56.0%**	**1600**

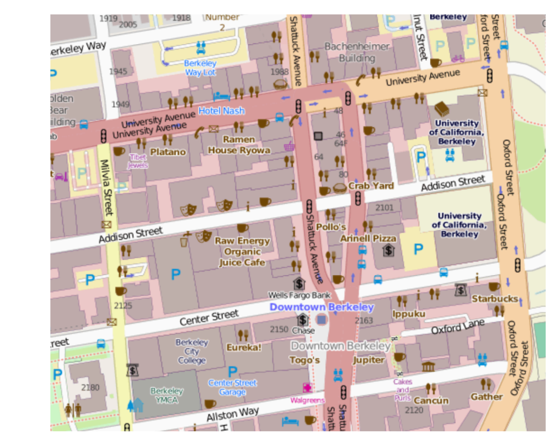




Dataset

Due to the lack of standardized datasets for building segmentation, we collected our own dataset using OpenStreetMap. Due to limitations in satellite geolocation accuracy, a limited amount of post-processing (30 minutes) was needed to allow for pixel-level accuracy. There are 1600 buildings labeled, along with road network data.

Dataset Sample



OpenStreetMap

Future Work

- Use Krahenbuhl et al.'s efficient learning algoritms to learn the parameters of the dense CRF
- Integrate height and road network data into the model
- Enforce rectangularity constraints on building detections
- Automate correction of ground truth data collection so that OpenStreetMap can be harnessed automatically and the dataset size increased

Conclusion

Densely connected CRF models allow for modeling of long range interactions and combat the problem of excessive smoothing. Inference can be efficiently performed on this model via the use of mean field inference and Gaussian kernels. Our work shows that this model can perform well in the field of remote sensing, and more specifically in the problem of building detection.

References

1. V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," NIPS, 2011.
2. J. Shotton, J. M. Winn, C. Rother, and A. Criminisi. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. IJCV, 2009.
3. L. Ladicky, C. Russell, P. Kohli, and P. H. S. Torr. Associative hierarchical crfs for object class image segmentation. In Proc. ICCV, 2009.