

Multiscale Conditional Random Fields for Image Labeling

Xuming He, Richard Zemel and Miguel A. Carreira-Perpinan
Department of Computer Science
University of Toronto

Introduction

- Image labeling
 - Classifying every image patch into a finite set of classes



- Typical issues
 - Using local image features
 - Capturing structures in labels

Random Field Framework

- Generative Markov Random Fields (MRFs)

Label (\mathbf{L})

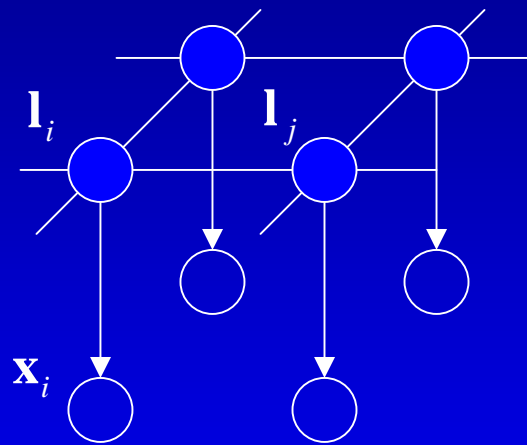
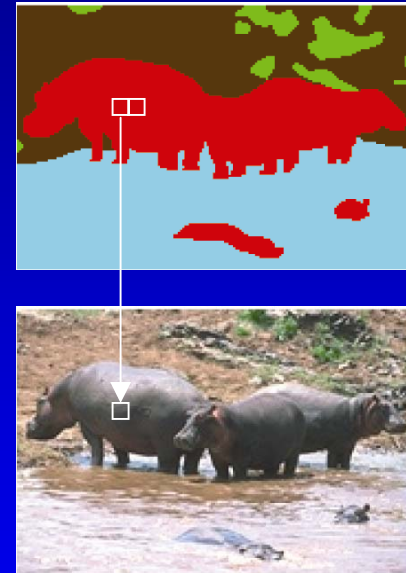


Image (\mathbf{X})



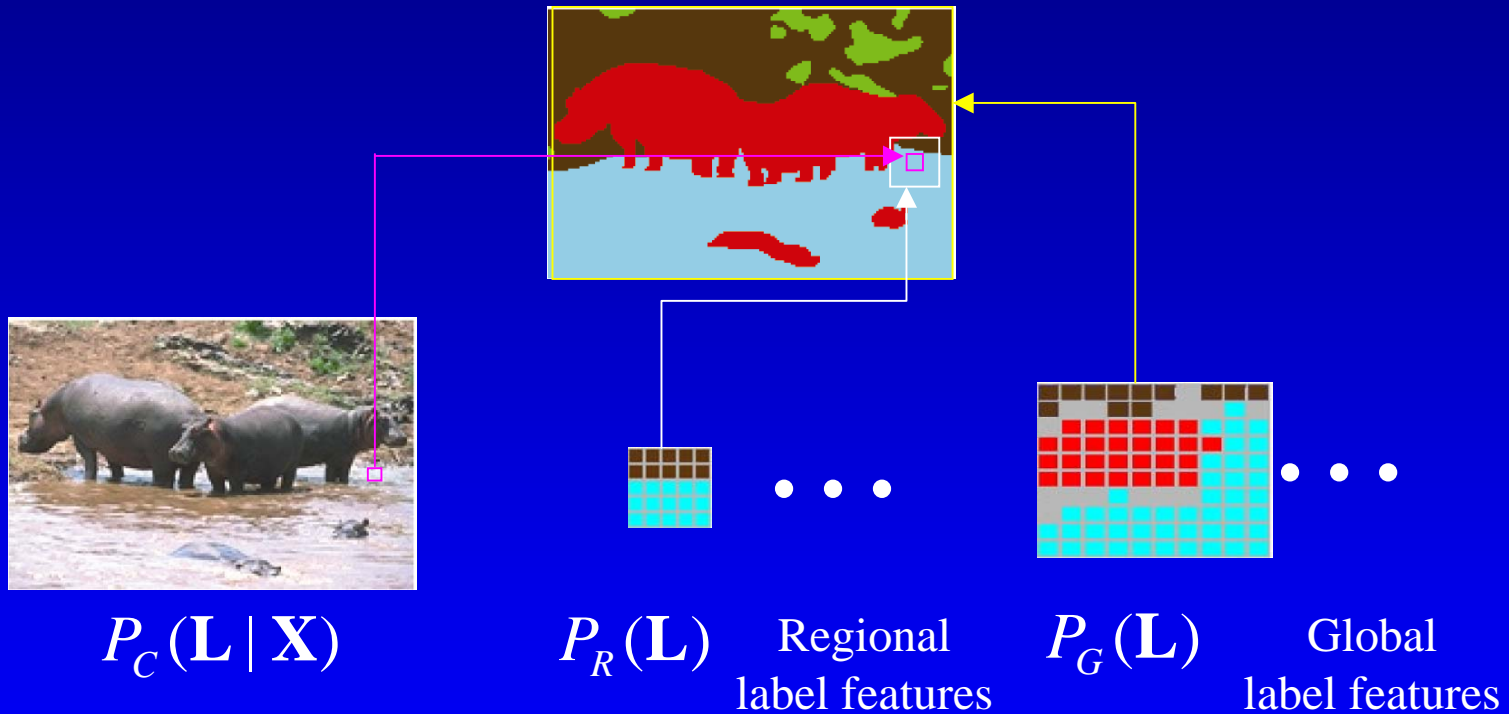
- Modeling:
$$P(\mathbf{L}, \mathbf{X}) = \prod_i P(\mathbf{x}_i | \mathbf{l}_i) \frac{1}{Z} \prod_{i,j} q(\mathbf{l}_i, \mathbf{l}_j)$$
- Labeling:
$$P(\mathbf{L} | \mathbf{X}) = P(\mathbf{L}, \mathbf{X}) / P(\mathbf{X})$$

Motivation

- Capturing context in different scales
 - Local constraints vs. Global configurations
 - Representing context dependency
- Building conditional model
 - Saving modeling resources
 - Training model discriminatively
- Learning context from data
 - Adapting context representation to images and labels

Overview

- Multiscale Conditional Random Fields (mCRF)

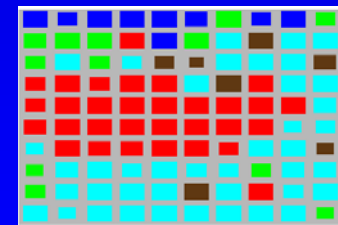
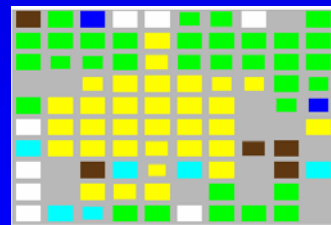
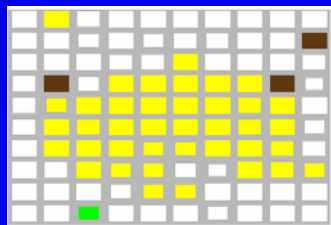
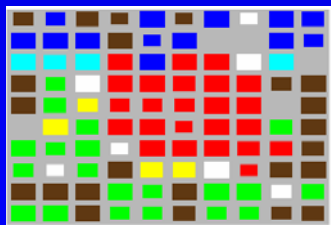


$$P(\mathbf{L} | \mathbf{X}) = \frac{1}{Z} P_C(\mathbf{L} | \mathbf{X}) P_R(\mathbf{L}) P_G(\mathbf{L})$$

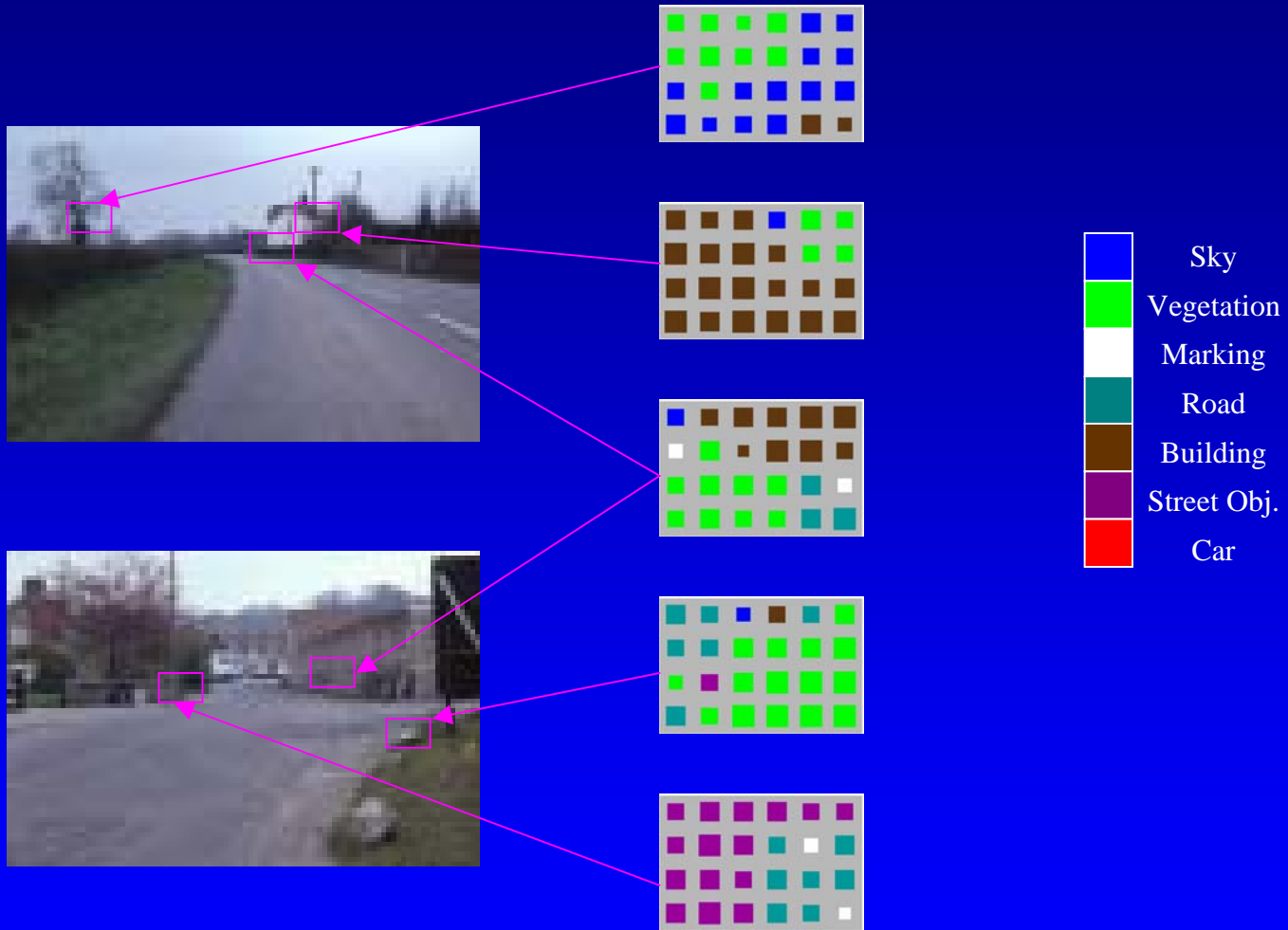
Examples of Global Features



- Rhino/hippo
- Polar bear
- Water
- Snow
- Vegetation
- Ground
- Sky



Examples of Local Features



Labeling Images

- Inferring Labels \mathbf{L} from Image \mathbf{X}
- Mode of Posterior Marginals Criterion

$$\mathbf{l}_i^* = \arg \max_{\mathbf{l}_i} P(\mathbf{l}_i | \mathbf{X})$$

- Approximate Inference
 - Block Gibbs sampling
 - Comparable to other methods: Loopy BP, Mean Field.

Parameter Estimation

- Supervised Learning with Dataset $\{(\mathbf{L}^t, \mathbf{X}^t), t = 1, \dots, N\}$
- Conditional Maximum Likelihood

$$\hat{\theta} = \max_{\theta} \sum_t \log(P(\mathbf{L}^t | \mathbf{X}^t; \theta))$$

- Gradient Ascent using MCMC

$$\Delta \theta_k \propto \sum_t \left\langle \frac{\partial \log h_k(\mathbf{L} | \mathbf{X}^t; \theta_k)}{\partial \theta_k} \right\rangle_{P_0(\mathbf{L} | \mathbf{X}^t)} - \left\langle \frac{\partial \log h_k(\mathbf{L} | \mathbf{X}^t; \theta_k)}{\partial \theta_k} \right\rangle_{P_{\theta}(\mathbf{L} | \mathbf{X}^t)}$$

- Supervised Contrastive Divergence

Experiment Setup

- Data set:
 - Subset of Corel database: African and Arctic wildlife natural scenes. (Training: 60, Testing: 40, 7 label classes)


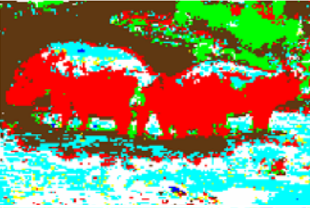

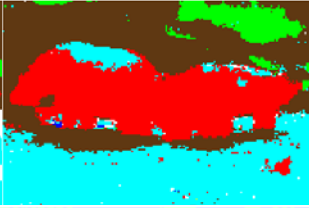












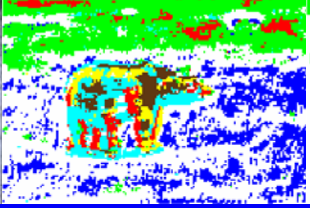

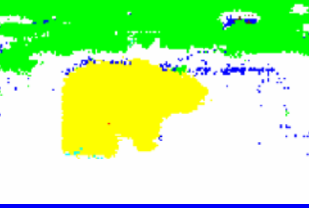



- Sowerby database: Rural and suburban scenes. (Training: 60, Testing: 44, 7 label classes)




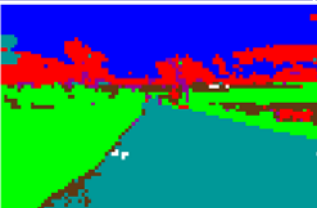
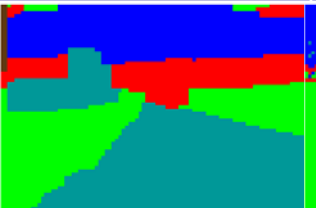
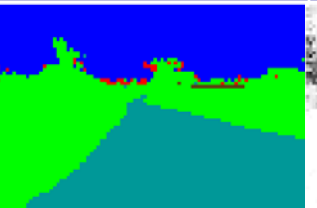







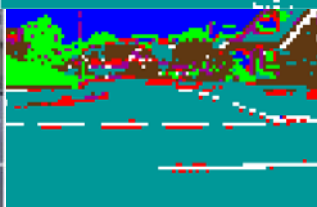
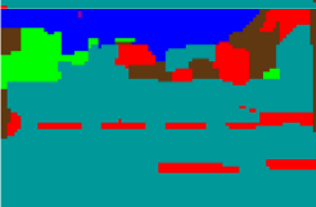
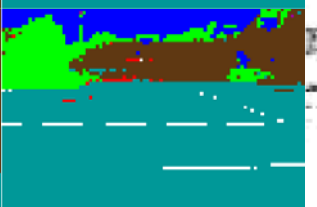

- Performance comparison:
 - Generative MRF
 - Local classifier

Results on Corel Image Data

Image	Classifier	MRF	mCRF	mCRF confidence
				
				
				
				
Classification rate	66.9%	66.2%	80.0%	-

-  Rhino/hippo
-  Polar bear
-  Water
-  Snow
-  Vegetation
-  Ground
-  Sky

Results on Sowerby Image Data

Image	Classifier	MRF	mCRF	mCRF confidence
				
				
				
Classification rate	82.4%	81.8%	89.5%	-

-  Sky
-  Vegetation
-  Marking
-  Road
-  Building
-  Street Obj.
-  Car

Summary and Future Work

- Summary
 - Multiscale representation of contextual information
 - Learning features from image data
 - Conditional RF model trained discriminatively
- Future Work
 - Comparison with tree-structured models
 - Invariance in contextual features
 - Contextual features with adaptive position and scale