Stacked Hierarchical Labeling

Dan Munoz        Drew Bagnell        Martial Hebert
The Labeling Problem

Input

Our Predicted Labels

Sky
Tree
Bldg
Fgnd
Road
The Labeling Problem
The Labeling Problem

• Needed: better representation & interactions
  – Ohta ‘78
Using Regions

Input

Ideal Regions

Slide from T. Malisiewicz
Using Regions

Input

Actual Regions

Slide from T. Malisiewicz
Using Regions + Interactions

Image Representation

- **Big regions**
- **Small regions**

**Ideal Prob. Graphical Model**
- High-order
- Expressive interactions
Using Regions + Interactions

Image Representation

big regions

small regions

Actual PGM

• Restrictive interactions
• Still NP-hard
Learning with Approximate Inference

- PGM learning requires **exact** inference
  - Otherwise, may **diverge**  
  
*Kulesza and Pereira ’08*

Simple Random Field

Learning Path
PGM Approach
Our Approach

Sequence of simple problems

Cohen ’05, Daume III ’06
A Sequence of Simple Problems

- Training simple modules to net desired output
  - No searching in exponential space
- Not optimizing any joint distribution/energy
  - Not necessarily doing it before! *Kulesza & Pereira ‘08*

Input \( f_1 \) \( \ldots \) \( f_N \) Output

Stacked Hierarchical Labeling
Our Contribution

• An effective PGM alternative for labeling
  – Training a **hierarchical** procedure of simple problems

• Naturally analyzes multiple scales
  – Robust to imperfect segmentations

• Enables more expressive interactions
  – Beyond pair-wise smoothing
Related Work

• Learning with multi-scale configurations
  – Joint probability distribution
    Bouman ‘94, Feng ‘02, He ‘04
    Borenstein ‘04, Kumar ‘05
  – Joint score/energy
    Tu ‘03, S.C. Zhu ‘06, L. Zhu ‘08
    Munoz ‘09, Gould ‘09, Ladicky ‘09

• Mitigating the intractable joint optimization
  – Cohen ‘05, Daume III ‘06, Kou ‘07, Tu ‘08, Ross ‘10
In this work, the segmentation tree is given.

We use the technique from Arbelaez ’09.
Segmentation Tree
(Arbelaez ’09)
• Parent sees big picture
• Naturally handles scales
- Parent sees big picture
- Naturally handles scales
- Break into simple tasks
- Predict label mixtures
Handling Real Segmentation

- $f_i$ predicts **mixture** of labels for each region
Actual Predicted Mixtures

- $P(\text{Fgnd})$
- $P(\text{Building})$
- $P(\text{Tree})$

(brighter $\rightarrow$ higher probability)
Training Overview

• How to train each module $f_i$?
• How to use previous predictions?
• How to train the hierarchical sequence?
Training Overview

• How to train each module $f_i$?
• How to use previous predictions?
• How to train the hierarchical sequence?
Modeling Heterogeneous Regions

- Count **true labels** $P_r$ present in each region $r$
- Train a **model** $Q$ to match each $P_r$
  - Logistic Regression
- $\min_Q H(P,Q) \rightarrow$ **Weighted Logistic Regression**
  - Image features: texture, color, etc. (Gould ’08)
Training Overview

• How to train each module $f_i$?
• **How to use previous predictions?**
• How to train the hierarchical sequence?
Using Parent Predictions

• Use broader context in the finer regions
  – Allow finer regions access to all parent predictions

• Create & append 3 types of context features
  – Kumar ’05, Sofman ’06, Shotton ’06, Tu ’08
Parent Context

• Refining the parent
Detailed In Paper

- **Image-wise (co-occurrence)**

\[ \sum \text{regions} \]

- **Spatial Neighborhood (center-surround)**
Training Overview

• How to train each module $f_i$?
• How to use previous predictions?

• How to train the hierarchical sequence?
Approach #1

- Train each module independently
  - Use ground truth context features

- Problem: Cascades of Errors
  - Modules depend on **perfect** context features
  - Observe no mistakes during training
    → Propagate mistakes during testing
Approach #2

- **Solution**: Train in feed-forward manner
  - Viola-Jones ‘01, Kumar ‘05, Wainwright ’06, Ross ‘10
Training Feed-Forward

LogReg

$\mathbf{f}_\ell$

(Parameters)
Training Feed-Forward

A → $f_c$ → 

B → $f_c$ → 

C → $f_c$ →
Cascades of Overfitting

Solution: Stacking

- Wolpert ’92, Cohen ’05
- Similar to x-validation
- Don’t predict on data used for training
Stacking

LogReg

$f_{\ell}^A$
Stacking
Stacking

LogReg

\( f^B \)
Stacking

A → $f^A_\ell$ → 

B → $f^B_\ell$ → 

C → $f^C_\ell$ → 
Learning to Fix Mistakes

Person part of incorrect segment
Person segmented, but relies on parent
Person fixes previous mistake
Level 1/8 Predictions

Segmentation
Level 1/8 Predictions

Segmentation

P(Foreground) 15%

P(Tree) 18%

P(Building) 12%

P(Road) 31%
Level 1/8 Predictions

Current Output | Segmentation | $P(\text{Foreground})$
--- | --- | ---

$P(\text{Tree})$ | $P(\text{Building})$ | $P(\text{Road})$
--- | --- | ---
18% | 12% | 31%
Level 2/8 Predictions

Segmentation

\( P(\text{Foreground}) \)

\( P(\text{Tree}) \)  \( P(\text{Building}) \)  \( P(\text{Road}) \)
Level 2/8 Predictions

Current Output  Segmentation  P(Foreground)

P(Tree)  P(Building)  P(Road)
Level 3/8 Predictions

Current Output  Segmentation  P(Foreground)

P(Tree)  P(Building)  P(Road)
Level 5/8 Predictions

Current Output  Segmentation  P(Foreground)

P(Tree)  P(Building)  P(Road)
Level 6/8 Predictions

Current Output  Segmentation  P(Foreground)

P(Tree)  P(Building)  P(Road)
Level 7/8 Predictions

Current Output  Segmentation  P(Foreground)

P(Tree)  P(Building)  P(Road)
Level 8/8 Predictions

Current Output | Segmentation | P(Foreground)

P(Tree) | P(Building) | P(Road)
Level 1/8 Predictions

Current Output  Segmentation  P(Foreground)

P(Tree)  P(Building)  P(Road)
Level 2/8 Predictions

Current Output | Segmentation | P(Foreground)

P(Tree) | P(Building) | P(Road)
Level 3/8 Predictions

Current Output  Segmentation  P(Foreground)

P(Tree)  P(Building)  P(Road)
Level 4/8 Predictions

Current Output  Segmentation  P(Foreground)

P(Tree)  P(Building)  P(Road)
Level 5/8 Predictions

Current Output  Segmentation  P(Foreground)

P(Tree)  P(Building)  P(Road)
Level 6/8 Predictions

Current Output  Segmentation  P(Foreground)

P(Tree)  P(Building)  P(Road)
Level 7/8 Predictions

Current Output  Segmentation  P(Foreground)

P(Tree)  P(Building)  P(Road)
Level 8/8 Predictions

Current Output  Segmentation  P(Foreground)

P(Tree)  P(Building)  P(Road)
Stanford Background Dataset

- 8 Classes
- 715 Images

<table>
<thead>
<tr>
<th>Method</th>
<th>Avg Class Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gould ICCV ‘09</td>
<td>65.5</td>
</tr>
<tr>
<td>LogReg (Baseline)</td>
<td>58.0</td>
</tr>
<tr>
<td>SHL (Proposed)</td>
<td>66.2</td>
</tr>
</tbody>
</table>

- Inference time
  - Segmentation & image features held constant

<table>
<thead>
<tr>
<th>Method</th>
<th>sec/image</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gould ICCV ‘09</td>
<td>30 - 600</td>
</tr>
<tr>
<td>SHL (Proposed)</td>
<td>10 - 12</td>
</tr>
</tbody>
</table>
**MSRC-21**

- 21 Classes
- 591 Images

<table>
<thead>
<tr>
<th>Method</th>
<th>Avg Class Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gould IJCV ‘08</td>
<td>64</td>
</tr>
<tr>
<td>LogReg (Baseline)</td>
<td>60</td>
</tr>
<tr>
<td>SHL (Proposed)</td>
<td>71</td>
</tr>
<tr>
<td>Lim ICCV’09</td>
<td>67</td>
</tr>
<tr>
<td>Tu PAMI’09</td>
<td>69</td>
</tr>
<tr>
<td>Zhu NIPS’08</td>
<td>74</td>
</tr>
<tr>
<td>Ladicky ICCV ‘09</td>
<td>75</td>
</tr>
</tbody>
</table>
## MSRC-21

- 21 Classes
- 591 Images

<table>
<thead>
<tr>
<th>Method</th>
<th>Avg Class Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gould IJCV ‘08</td>
<td>64</td>
</tr>
<tr>
<td>LogReg (Baseline)</td>
<td>60</td>
</tr>
<tr>
<td><strong>SHL (Proposed)</strong></td>
<td><strong>71</strong></td>
</tr>
<tr>
<td>Lim ICCV’09</td>
<td>67</td>
</tr>
<tr>
<td>Tu PAMI’09</td>
<td>69</td>
</tr>
<tr>
<td>Zhu NIPS’08</td>
<td>74</td>
</tr>
<tr>
<td>Ladicky ICCV ‘09</td>
<td>75</td>
</tr>
<tr>
<td>LogReg (Baseline)</td>
<td>69</td>
</tr>
<tr>
<td><strong>SHL (Proposed)</strong></td>
<td><strong>75</strong></td>
</tr>
</tbody>
</table>
Ongoing Work

Labeling 3-D Point Clouds with Xuehan Xiong
Conclusion

• An effective structured prediction alternative
  – High performance with no graphical model

• Beyond site-wise representations
  – Robust to imperfect segmentations & multiple scales

• Prediction is a series of simple problems
  – Stacked to avoid cascading errors and overfitting
Thank You

• Acknowledgements
  – QinetiQ North America Robotics Fellowship
  – ONR MURI: Reasoning in Reduced Information Spaces
  – Reviewers, S. Ross, A. Grubb, B. Becker, J.-F. Lalonde

• Questions?
Image-wise

\[ \sum \text{regions} \]
Spatial neighborhood
Interactions

• Described in this talk

\[ \mathcal{R}_{A,\ell-1} \xrightarrow{b_{A,\ell-1}} \mathcal{R}_{A,\ell} \xrightarrow{b_{A,\ell}} \mathcal{R}_{A,\ell+1} \]

• Described in the paper

\[ \mathcal{R}_{A,\ell-1} \xrightarrow{b_{A,\ell-1}} \mathcal{R}_{A,\ell} \xrightarrow{\tilde{b}_{A,\ell}} \mathcal{R}_{A,\ell} \xrightarrow{b_{A,\ell}} \mathcal{R}_{A,\ell+1} \]
SHL vs. M3N