Beyond Categories: The Visual Memex Model for Reasoning About Object Relationships

Tomasz Malisiewicz and Alexei (Alyosha) Efros

NIPS 2009

Presented at CMU’s misc-read on 11/4/2009
Overview

• Motivation
• Visual Memex Model
• Evaluation Task: Context Challenge
• Results and Comparisons to Baselines
• Examples
Understanding an Image

slide by Fei Fei, Fergus & Torralba
Object naming
Object naming / Object categorization

- sky
- building
- flag
- wall
- banner
- face
- street lamp
- bus
- cars

slide by Fei Fei, Fergus & Torralba
Vannevar Bush’ Memex

**Memex:** A device to organize concepts and ideas via a web of associations (and not indexing)

V. B. “As We May Think,” The Atlantic, 1945
Visual Memex Model

Input Image

Visual Memex

Context Edge

Similarity Edge
Learning Similarity Edges

- Per-exemplar distance function learning algorithm from CVPR 2008

<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape</td>
<td>Centered Mask</td>
<td>32x32=1024</td>
</tr>
<tr>
<td></td>
<td>BB Extent</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Pixel Area</td>
<td>1</td>
</tr>
<tr>
<td>Texture</td>
<td>Right Boundary Tex-Hist</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Top Boundary Tex-Hist</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Left Boundary Tex-Hist</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Bottom Boundary Tex-Hist</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Interior Tex-Hist</td>
<td>100</td>
</tr>
<tr>
<td>Color</td>
<td>Mean Color</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Color std</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Color Histogram</td>
<td>33</td>
</tr>
<tr>
<td>Location</td>
<td>Absolute Mask</td>
<td>8x8=64</td>
</tr>
<tr>
<td></td>
<td>Top Height</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Bot Height</td>
<td>1</td>
</tr>
</tbody>
</table>
Learning Similarity Edges

\[ f(w, b, \alpha) = \frac{\lambda}{2} \|w\|^2 + \sum_{i \in C} \alpha_i L(- (w^T d_i + b)) + \sum_{i \notin C} L(w^T d_i + b) - \sigma \|\alpha\|^2 \]
Learning Similarity Edges

\[ f(w, b, \alpha) = \frac{\lambda}{2} \|w\|^2 + \sum_{i \in C} \alpha_i L(-(w^T d_i + b)) + \sum_{i \notin C} L(w^T d_i + b) - \sigma \|\alpha\|^2 \]
Context Edges

- Compute pairwise feature for each pair of exemplars in a single image
- Overlap, Scale, Displacement, Bottom-pixel Height $f \in \mathbb{R}^{10}$
- No appearance!

$$K(f, f') = e^{-\alpha_1 \| f - f' \|^2}$$
Visual Memex Model

\[ G = (V, E_S, E_C, \{D\}, \{f\}) \]

\[ N = 87,802 \quad |E_S| = 276,782 \quad |E_C| = 989,106 \]
Evaluation Task: Context Challenge

- Compare against Category-Based models
- Perform context-only recognition “without local appearance”
- Motivated by Torralba: “How far can you go without running an object detector?”
- Our context is object-centered as opposed to Torralba’s scene-centered GIST context
Context Challenge

associations
Visual Memex Inference

- **Goal**: Create exemplar associations for hidden region given K supporting object regions

- **Approach**: Score the assignment of every exemplar to the hidden region
Visual Memex Inference

Appearance Features of supporting objects \{S_1, \ldots, S_K\}

Soft Associations between supporting objects and Memex exemplars

Exemplar-Exemplar Pairwise Potential

\[
p(e_i | A_1, \ldots, A_K, f_{i1}, \ldots, f_{iK}) \propto \prod_{j=1}^{K} \sum_{a=1}^{N} A_j^a \Psi(e_i, e_a, f_{ij})
\]
“Densification”
Potential

Visual Memex Adjacency Matrix

\[ W_{uv} = [(u, v) \in E_S] \]

\[ \log \Psi(e_i, e_j, f_{ij}) = \frac{\sum_{(u,v) \in E_C} W_{iu} W_{jv} K(f_{ij}, f_{uv})}{\sum_{(u,v) \in E_C} W_{iu} W_{jv}} \]
“Densification” Potential

Visual Memex Adjacency Matrix

\[ W_{uv} = [(u, v) \in E_S] \]

\[
\log \Psi(e_i, e_j, f_{ij}) = \frac{\sum_{(u,v) \in E_C} W_{iu} W_{jv} K(f_{ij}, f_{uv})}{\sum_{(u,v) \in E_C} W_{iu} W_{jv}}
\]

KDE with these features
Baselines

- Category-based parametric CoLA model (Serge Belongie’s Vision Group)

- Category-based Reduced Memex (categories for objects + nonparametric context model)
CoLA Baseline

\[ p(y_i | y_1, \ldots, y_K, f_{i1}, \ldots, f_{iK}, \theta) = \frac{1}{Z} \prod_{j=1}^{K} \Psi(y_i, y_j, f_{ij}, \theta) \]

\[ \log \Psi(y_i, y_j, f_{ij}, \theta) = [h(f_{ij})^T] \theta(y_i, y_j) \]
Reduced KDE Memex

- No distance function learning
- Just connect all exemplars of the same category

\[
\log \Psi(y_i, y_j, f_{ij}) = \frac{\sum_{(u,v) \in E_C} \delta_{y_i y_u} \delta_{y_j y_v} K(f_{ij}, f_{uv})}{\sum_{(u,v) \in E_C} \delta_{y_i y_u} \delta_{y_j y_v}}
\]
Results

• 200 Densely labeled LabelMe images

• Solve separate Context Challenge task for a single object while fixing supporting objects
## Results

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Per-Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual Memex</td>
<td>0.527</td>
<td>0.534</td>
</tr>
<tr>
<td>KDE Memex</td>
<td>0.430</td>
<td>0.454</td>
</tr>
<tr>
<td>CoLA</td>
<td>0.457</td>
<td>0.213</td>
</tr>
</tbody>
</table>
Per-Category Results

Context Challenge Recognition Accuracy for 30 Categories

Visual Memex
KDE
CoLA
Input Image + Hidden Region
Conclusion

• Visual Memex is a Category-free approach to Modeling Object Relationships

• Is distance function learning without labels possible? (Then the entire approach would be category-free)

• Can the Visual Memex be applied to scene parsing where only an image is given?