Learning to Refine Object Segments
Pedro O. Pinheiro*, Tsung-Yi Lin*, Ronan Collobert, Piotr Dollár

Introduction
Object Recognition Tasks
(a) image classification
(b) detection with boxes
(c) semantic segmentation
(d) detection with segments

DeepMask for Object Proposals (NIPS15)
patch
DeepMask
mask
score
detect mask/center
image
masks
scores
sliding window

Limitations
- purely feedforward model
- output masks are coarse
- unoptimized architecture

Why Top Down Refinement?
- bottom-up network generates coarse semantic mask
- lower layers → spatial information (pixel-level information)
- higher layers → semantic information (object-level information)
- top-down refinement sharpens output using lower-layer features

Refinement Method
(a) feedforward
(b) feedforward + skip
(c) proposed network
(d) refinement module

Feedforward Model
- goal: optimize speed & accuracy
- reduce input size (224x224→160x160) | reduce feature dimension (512→128)
- ResNet trunk | maximize sharing of computation between branches

Learning Mask Refinement

Experiments
Object Proposal Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Box Proposals</th>
<th>Segmentation Proposals</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR@10</td>
<td>AR@100</td>
<td>AR@100</td>
</tr>
<tr>
<td></td>
<td>AUC</td>
<td>AR@100 AUC</td>
</tr>
<tr>
<td>DeepMask</td>
<td>15.3</td>
<td>21.3</td>
</tr>
<tr>
<td>Geometric</td>
<td>15.9</td>
<td>22.5</td>
</tr>
<tr>
<td>RPN</td>
<td>16.0</td>
<td>23.0</td>
</tr>
<tr>
<td>MCG</td>
<td>16.1</td>
<td>24.2</td>
</tr>
</tbody>
</table>

Object Detection Results

<table>
<thead>
<tr>
<th>Method</th>
<th>AP 50</th>
<th>AP 75</th>
<th>AP 90</th>
<th>AP 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet</td>
<td>21.5</td>
<td>35.9</td>
<td>42.5</td>
<td>55.1</td>
</tr>
<tr>
<td>SharpMask+VGG</td>
<td>21.1</td>
<td>42.1</td>
<td>44.9</td>
<td>49.9</td>
</tr>
<tr>
<td>SharpMask+MPN</td>
<td>22.3</td>
<td>42.9</td>
<td>45.1</td>
<td>49.3</td>
</tr>
<tr>
<td>ION</td>
<td>22.0</td>
<td>41.0</td>
<td>43.0</td>
<td>47.0</td>
</tr>
</tbody>
</table>

Summary
- novel architecture for object instance segmentation
- state of the art proposal and detection results
- simple, fast, general network for pixel-labeling tasks

pedro@opinheiro.com | tl483@cornell.edu