**Spatial Memory**
- Generic framework to store all kinds of instance-, pixel-, & image-level knowledge
- Models the joint layout distribution of all instances and learns de-duplication
- Efficient context reasoning with ConvNet for object detection

**Object Detection**
- Formulation
  \[ \arg\max_{\mathcal{M}} \log P(O_n|\mathcal{M}, I) = \sum_{n=1}^{N} \log P(O_n|O_{n-1}, \mathcal{M}, I) \]
- Learned Prior
- Faster RCNN (ConvNet) Results

**Memory-Based Detection**
- \[ \mathcal{M} = \log P(I_{\text{max}}|\mathcal{M}, I) \]
- Faster RCNN
- RoI Pooling
- fc8
- conv5
- m-conv5
- RoIs
- Region Proposal Network
- Memory Iterations
- Select Detection
- Input Feature

**Recall Pitfall**
1. Top RoIs 2. Guessing game: ski or snowboard?

**Qualitative Examples**
- Previous detections: current
- Previous detections: current

**COCO Results**
- Baseline: [Chen & Gupta 2017] on VGG16
- MLP: Extension of Baseline with more layers

<table>
<thead>
<tr>
<th>Method</th>
<th>AP</th>
<th>AP-5</th>
<th>AP-75</th>
<th>AP-S</th>
<th>AP-M</th>
<th>AP-L</th>
<th>AR-S</th>
<th>AR-M</th>
<th>AR-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>29.4</td>
<td>50.0</td>
<td>30.9</td>
<td>12.2</td>
<td>33.7</td>
<td>43.8</td>
<td>18.5</td>
<td>45.5</td>
<td>58.9</td>
</tr>
<tr>
<td>MLP</td>
<td>30.1</td>
<td>50.8</td>
<td>31.7</td>
<td>12.5</td>
<td>34.2</td>
<td>44.5</td>
<td>19.2</td>
<td>47.0</td>
<td>59.8</td>
</tr>
<tr>
<td>SMN</td>
<td>31.6</td>
<td>52.2</td>
<td>33.2</td>
<td>14.4</td>
<td>35.7</td>
<td>45.8</td>
<td>20.5</td>
<td>48.8</td>
<td>63.2</td>
</tr>
</tbody>
</table>

**Future Work**
- Speed-Up
- Learning the selection process with REINFORCE

**Output for De-Duplication**
- Naive Design
- Final Design
- remove ground truth during training if detected

**Approximation (Detectors Now)**
- \[ \mathcal{L} \approx \sum \log P(O_n | \mathcal{M}, I) \]