MOTIVATION

• Images contain many layers of complexity.
MOTIVATION

• Images contain many layers of complexity.

• Current issues: scalability and location of features
KEY TERMINOLOGY

- Deep Learning (DL) – a type of machine learning inspired by the human brain

Restricted Boltzmann Machine (RBM)

Deep Belief Network (DBN)
Figure 1. Convolutional RBM with probabilistic max-pooling. For simplicity, only group $k$ of the detection layer and the pooling layer are shown. The basic CRBM corresponds to a simplified structure with only visible layer and detection (hidden) layer. See text for details.
CONVOLUTION FROM VISIBLE TO DETECTION

Visible Layer → Summed Element-wise Matrix Multiplication → Detection Layer
CONVOLUTION FROM VISIBLE TO DETECTION

Visible Layer → Summed Element-wise Matrix Multiplication → Detection Layer
CONVOLUTION FROM VISIBLE TO DETECTION

Visible Layer

Summed Element-wise Matrix Multiplication

Detection Layer
MAX-POOLING

<table>
<thead>
<tr>
<th>Detection Layer</th>
<th>0</th>
<th>0.05</th>
<th>0.08</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0.23</td>
<td>0</td>
<td>0.11</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.38</td>
<td>0.43</td>
<td>0.28</td>
<td>0</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>0.24</td>
<td>0</td>
<td>0.39</td>
<td>0.38</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>0.19</td>
<td>0.08</td>
<td>0.04</td>
<td>0</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Pooling Layer
#### MAX-POOLEDING

**Detection Layer**

<table>
<thead>
<tr>
<th></th>
<th>0.17</th>
<th>0.08</th>
<th>0</th>
<th>0</th>
<th>0.23</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.12</td>
<td>0.32</td>
<td>0.15</td>
<td>0.22</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.19</td>
<td>0.24</td>
<td>0.13</td>
<td>0.24</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td><strong>0.38</strong></td>
<td>0</td>
<td>0</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>0.20</td>
<td>0.07</td>
<td>0.22</td>
<td>0</td>
<td>0.22</td>
<td></td>
</tr>
</tbody>
</table>

**Pooling Layer**

Distance: 40px

Distance: 8px
MAX-POOLING

Detection Layer

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.07</td>
<td>0.37</td>
<td>0.14</td>
<td>0.15</td>
<td>0</td>
</tr>
<tr>
<td>0.18</td>
<td>0.25</td>
<td>0.04</td>
<td>0.28</td>
<td>0.16</td>
</tr>
<tr>
<td>0.32</td>
<td>0.09</td>
<td>0</td>
<td>0.16</td>
<td>0.24</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.07</td>
<td>0.15</td>
<td>0.13</td>
</tr>
<tr>
<td>0.21</td>
<td>0.06</td>
<td>0.18</td>
<td>0.40</td>
<td>0</td>
</tr>
</tbody>
</table>

Pooling Layer

40px

8px
ABILITY TO CLASSIFY IMAGES IN SEMI-SUPERVISED SETTING

TESTED ON THE CALTECH-101 DATASET

<table>
<thead>
<tr>
<th>Table 1. Classification accuracy for the Caltech-101 data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Size</td>
</tr>
<tr>
<td>CDBN (first layer)</td>
</tr>
<tr>
<td>CDBN (first+second layers)</td>
</tr>
<tr>
<td>Raina et al. (2007)</td>
</tr>
<tr>
<td>Ranzato et al. (2007)</td>
</tr>
<tr>
<td>Mutch and Lowe (2006)</td>
</tr>
<tr>
<td>Lazebnik et al. (2006)</td>
</tr>
<tr>
<td>Zhang et al. (2006)</td>
</tr>
</tbody>
</table>
Figure 3. Columns 1-4: the second layer bases (top) and the third layer bases (bottom) learned from specific object categories. Column 5: the second layer bases (top) and the third layer bases (bottom) learned from a mixture of four object categories (faces, cars, airplanes, motorbikes).
HIERARCHICAL PROBABILISTIC INFERENCE

- Model combines bottom-up input and top-down context
- This improves representation.

Figure 6. Hierarchical probabilistic inference. For each column: (top) input image. (middle) reconstruction from the second layer units after single bottom-up pass, by projecting the second layer activations into the image space. (bottom) reconstruction from the second layer units after 20 iterations of block Gibbs sampling.
CONCLUSIONS

• The authors’ convolutional DBN was scalable to larger images.
• It could learn hierarchical representations from unlabeled images.
• Future work: use on high-dimensional, complex data