Dynamic Routing Between Capsules

PRESENTATION BY: KEVIN DUARTE
Motivation

• Max Pooling
  • Reduces spatial size – loses information

• CNNs do not implicitly model entities

• Capsule Networks add extra structure
  • Models entities/features as vectors
  • Routing-by-agreement
Issues with CNNs

Computer Graphics

https://www.slideshare.net/aureliengeron/introduction-to-capsule-networks-capsnets
Inverse Graphics

https://www.slideshare.net/aureliengeron/introduction-to-capsule-networks-capsnets
Capsules

https://www.slideshare.net/aureliengeron/introduction-to-capsule-networks-capsnets
Traditional Neuron

Given input scalars $x_i$

$$a_j = \sum_i x_i w_i + b$$

$$h_j = f(a_j)$$

Output is a scalar $h_j$
Capsules

Given input vectors $u_i$

$$\hat{u}_{j|i} = W_{ij} u_i$$

$$s_j = \sum_i c_{ij} \hat{u}_{j|i}$$

$$v_j = \frac{|s_j|^2}{1+|s_j|^2} \frac{s_j}{|s_j|}$$

Output is a vector $v_j$
<table>
<thead>
<tr>
<th>Operation</th>
<th>Capsule</th>
<th>Traditional Neuron</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input from low-level capsule/neuron</td>
<td>$\text{vector}(u_i)$</td>
<td>$\text{scalar}(x_i)$</td>
</tr>
<tr>
<td>Affine Transform</td>
<td>$\hat{u}_{j</td>
<td>i} = W_{ij}u_i$</td>
</tr>
<tr>
<td>Weighting</td>
<td>$s_j = \sum_i c_{ij}\hat{u}_{j</td>
<td>i}$</td>
</tr>
<tr>
<td>Sum</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>Nonlinear Activation</td>
<td>$v_j = \frac{</td>
<td></td>
</tr>
<tr>
<td>Output</td>
<td>$\text{vector}(v_j)$</td>
<td>$\text{scalar}(h_j)$</td>
</tr>
</tbody>
</table>
Routing-by-Agreement

**Capsule J**

- Decrease $c_{ij}$

**Capsule K**

- Increase $c_{ik}$

Routing of output from one lower level capsule i
Routing Algorithm

Procedure 1 Routing algorithm.

1: procedure ROUTING($\hat{u}_{j|i}$, $r$, $l$)
2: for all capsule $i$ in layer $l$ and capsule $j$ in layer $(l + 1)$: $b_{ij} \leftarrow 0.$
3: for $r$ iterations do
4: for all capsule $i$ in layer $l$: $c_i \leftarrow \text{softmax}(b_i)$  \hspace{1cm} \triangleright \text{softmax computes Eq. 3}
5: for all capsule $j$ in layer $(l + 1)$: $s_j \leftarrow \sum_i c_{ij} \hat{u}_{j|i}$
6: for all capsule $j$ in layer $(l + 1)$: $v_j \leftarrow \text{squash}(s_j)$  \hspace{1cm} \triangleright \text{squash computes Eq. 1}
7: for all capsule $i$ in layer $l$ and capsule $j$ in layer $(l + 1)$: $b_{ij} \leftarrow b_{ij} + \hat{u}_{j|i}.v_j$

return $v_j$

Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic Routing Between Capsules."
CapsNet Architecture

Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic Routing Between Capsules."
Conv1

- Input: 28x28x1 image
- Conv1: Convolutional Layer
  - ReLU activation
  - 9x9 Filter
  - Stride of 1
  - 256 feature maps
  - Result has shape (20, 20, 256)
PrimaryCaps

- Add another Convolutional Layer
  - ReLU activation
  - 9x9 filter
  - Stride of 2
  - $(32 \times 8) = 256$ feature maps
  - Result has shape (6, 6, 256)

- Reshape into capsules with 8-d vectors
  - 32 channels of capsules
  - Result has shape (6, 6, 32, 8)
  - Squash the capsule vectors

- Total of $(6 \times 6 \times 32) = 1152$ different capsules
DigitCaps

- Multiply PrimaryCaps with matrices $W_{ij}$
  - $\hat{u}_{j|i} = W_{ij} u_i$
  - 8-d vectors transform into 16-d vectors

- Iterative routing algorithm
  - $s_j = \sum_i c_{ij} \hat{u}_{j|i}$
  - $v_j = \frac{||s_j||^2}{1+||s_j||^2} \frac{s_j}{||s_j||}$

- Result has shape (10, 16)
  - 10 capsules, with 16 dimensions
CapsNet Architecture - Reconstruction

Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic Routing Between Capsules."
Loss Function

• Margin Loss
  \[ L_c = T_c \max(0, m^+ - \|v_c\|)^2 + \lambda (1 - T_c) \max(0, \|v_c\| - m^-)^2 \]

• Reconstruction Loss – L2 Loss
  \[ L_r = \sum_{i=0}^{n} (y_i - h(x_i))^2 \]

• Total Loss
  \[ L = \sum_c L_c + 0.0005L_r \]
## Results on MNIST

<table>
<thead>
<tr>
<th>Method</th>
<th>Routing</th>
<th>Reconstruction</th>
<th>MNIST (%)</th>
<th>MultiMNIST (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>-</td>
<td>-</td>
<td>0.39</td>
<td>8</td>
</tr>
<tr>
<td>CapsNet</td>
<td>1</td>
<td>no</td>
<td>$0.34_{\pm0.032}$</td>
<td>-</td>
</tr>
<tr>
<td>CapsNet</td>
<td>1</td>
<td>yes</td>
<td>$0.29_{\pm0.011}$</td>
<td>7</td>
</tr>
<tr>
<td>CapsNet</td>
<td>3</td>
<td>no</td>
<td>$0.35_{\pm0.036}$</td>
<td>-</td>
</tr>
<tr>
<td>CapsNet</td>
<td>3</td>
<td>yes</td>
<td>$0.25_{\pm0.005}$</td>
<td>5</td>
</tr>
</tbody>
</table>

Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic Routing Between Capsules."
## Results on MNIST - Reconstruction

<table>
<thead>
<tr>
<th>$(l, p, r)$</th>
<th>(2, 2, 2)</th>
<th>(5, 5, 5)</th>
<th>(8, 8, 8)</th>
<th>(9, 9, 9)</th>
<th>(5, 3, 5)</th>
<th>(5, 3, 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong></td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
</tbody>
</table>
What Capsule Dimensions Represent

<table>
<thead>
<tr>
<th>Feature</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale and thickness</td>
<td>6 6 6 6 6 6 6 6 6 6</td>
</tr>
<tr>
<td>Localized part</td>
<td>6 6 6 6 6 6 6 6 6 6</td>
</tr>
<tr>
<td>Stroke thickness</td>
<td>5 5 5 5 5 5 5 5 5 5</td>
</tr>
<tr>
<td>Localized skew</td>
<td>4 4 4 4 4 4 4 4 4 4</td>
</tr>
<tr>
<td>Width and translation</td>
<td>3 3 3 3 3 3 3 3 3 3</td>
</tr>
<tr>
<td>Localized part</td>
<td>2 2 2 2 2 2 2 2 2 2</td>
</tr>
</tbody>
</table>
Results on affNIST

• More robust to small affine transformations
  • 79% on affNIST test set
  • Traditional CNN achieved 66%
Results on multiMNIST
Other Dataset

- CIFAR10: 10.6% error
  - Ensemble of 7 models
  - Not state of the art
- SVHN: 4.3% error
- smallNORB: 2.7% error

Y. LeCun, F.J. Huang, L. Bottou, “Learning Methods for Generic Object Recognition with Invariance to Pose and Lighting.”