Localization with Classification ConvNets
Classification ConvNets

• Classification is an easy task
  • Good performance
  • Labels are easy to obtain

• Top 5 predictions of a ConvNet is usually good
  • Multiple objects

• But can we know where the object is from just a classification ConvNet i.e. a net with just image-level labels?
Class Saliency Visualization

• “Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps” by Simonyan et al.
• Class Score, $S_c(I)$, of a ConvNet can be approximated by

$$S_c(I) \approx w^T I + b$$

where $w$ is

$$w = \frac{\partial S_c}{\partial I} \bigg|_{I_0}$$

can be found by back-propagation
Examples
Examples – Fine Grained Classification
2 New Approaches

1. “Learning Deep Features for Discriminative Localization” by Zhou et al. They call their approach Class Action Mapping (CAM)
2. “Top-down Neural Attention by Excitation Backprop” by Zhang et al.
Class Activation Mapping (1/2)

• Use Global Average Pooling (GAP) to spatially pool activations of the last convolutional layer

• Output of GAP are used as features for the last fully-connected layer

• So, important image regions is a linear combination of the convolutional feature maps and the weights of the output layers

\[ M_c(x, y) = \sum_k w^c_k f_k(x, y). \]
Class Activation Mapping (2/2)
Qualitative Examples – CAM ResNet152
Qualitative Examples – Grad ResNet152
Excitation Backprop

**Assumptions:**

- The response of the activation neural is non-negative
- An activation neuron is tuned to detect certain visual features. Its response is positively correlated to its confidence of the detection.

\[
P(a_j | a_i) = \begin{cases} 
Z_i \hat{a}_j w_{ji} & \text{if } w_{ji} \geq 0, \\
0 & \text{otherwise.}
\end{cases}
\]
Excitation Backprop

**Assumptions:**

- The response of the activation neural is non-negative.
- An activation neuron is tuned to detect certain visual features. Its response is positively correlated to its confidence of the detection.
Excitation Backprop+

Contrastive Attention

![Contrastive Attention Diagram](image)
Excitation Backprop+  

Contrastive Attention

[Diagram showing the flow of information through zebra and non-zebra classifiers]
Qualitative Examples – Excitation Backprop+
Qualitative Examples – CAM ResNet152
Qualitative Examples – Grad ResNet152

African elephant, Loxodonta africana: 0.610
Tusker: 0.346

Zebra: 0.022

Indian elephant, Elephas maximus: 0.020
Wart hog: 0.000
Excitation Backprop+ 18K Tags

Input  chair  glass  boy  woman  man  couple  father
The Pointing Game

• Task:
  • Given an image and an object category, point to the targets.

• Metric
  • Pointing accuracy.
  • Anywhere on the target is fine.

• Dataset
  • VOC (20 Categories)
  • COCO (80 Categories)
# The Pointing Game

<table>
<thead>
<tr>
<th></th>
<th>VOC07 Test (All/Diff.)</th>
<th>COCO Val. (All/Diff.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CNN-S</td>
<td>VGG16</td>
</tr>
<tr>
<td><strong>Center</strong></td>
<td>69.5/42.6</td>
<td>69.5/42.6</td>
</tr>
<tr>
<td><strong>Grad [5]</strong></td>
<td>78.6/59.8</td>
<td>76.0/56.8</td>
</tr>
<tr>
<td><strong>Deconv [6]</strong></td>
<td>73.1/45.9</td>
<td>75.5/52.8</td>
</tr>
<tr>
<td><strong>LRP [9]</strong></td>
<td>68.1/41.3</td>
<td>-</td>
</tr>
<tr>
<td><strong>CAM [8]</strong></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>MWP</strong></td>
<td>73.7/52.9</td>
<td>76.9/55.1</td>
</tr>
<tr>
<td><strong>c-MWP</strong></td>
<td><strong>78.7/61.7</strong></td>
<td><strong>80.0/66.8</strong></td>
</tr>
</tbody>
</table>