Improved Weakly-Supervised Object Localization Using Hide-and-Seek Based Deep Neural Network

Gautam Pradeep
Mentor: Krishna Kumar Singh
Faculty Advisor: Prof. Yong Jae Lee
EnvironMentors-AggieMentors
University of California Davis / Mira Loma High School

Methodology and Stages of Experimentation

Phase 1: Class Activation Map Generation

Major Stages:
1. Develop socket connections between client and server sides to send files via a specific port channel.
2. Create a web application to upload an image to send for further tasks.
3. Utilize Hide-and-Seek model on the server to localize and classify the object to the area of interest in an input image.
   - Allow option for user to run Hide-and-Seek network or AlexNet network.
   - Allow user to choose a specific class for which to generate a CAM.
4. Generate CAM image and send CAM along with classification results back to client through HTML browser to display a list of top predictions.

Phase 1 Qualitative Results

CAM is generated during test time.

Original Image
AlexNet-HaS
AlexNet-GAP

Original Image
AlexNet-HaS
AlexNet-GAP

Data

<table>
<thead>
<tr>
<th>Method</th>
<th>GT-known Loc</th>
<th>Top-1 Loc</th>
<th>Top-1 Clin</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet-GAP</td>
<td>42.59</td>
<td>36.25</td>
<td>30.23</td>
</tr>
<tr>
<td>AlexNet-HaS</td>
<td>57.86</td>
<td>36.77</td>
<td>25.97</td>
</tr>
<tr>
<td>AlexNet-HaS-52</td>
<td>87.75</td>
<td>37.33</td>
<td>27.94</td>
</tr>
<tr>
<td>AlexNet-HaS-44</td>
<td>58.33</td>
<td>37.34</td>
<td>38.18</td>
</tr>
<tr>
<td>AlexNet-HaS-50</td>
<td>58.48</td>
<td>37.34</td>
<td>38.13</td>
</tr>
<tr>
<td>AlexNet-HaS-Mixed</td>
<td>58.68</td>
<td>37.68</td>
<td>39.68</td>
</tr>
<tr>
<td>GoogLeNet-GAP</td>
<td>58.17</td>
<td>43.62</td>
<td>31.86</td>
</tr>
<tr>
<td>GoogLeNet-HaS-36</td>
<td>59.83</td>
<td>44.62</td>
<td>30.49</td>
</tr>
<tr>
<td>GoogLeNet-HaS-44</td>
<td>60.29</td>
<td>45.23</td>
<td>30.70</td>
</tr>
<tr>
<td>GoogLeNet-HaS-54</td>
<td>60.11</td>
<td>46.75</td>
<td>30.34</td>
</tr>
<tr>
<td>GoogLeNet-HaS-Mixed</td>
<td>59.97</td>
<td>47.90</td>
<td>30.37</td>
</tr>
</tbody>
</table>

Conclusion and Future Work

• By analyzing the failure cases, we can see that in certain angles and images, the H-S model fails to localize most of the object, which would result in sub-optimal/bounding box predictions.
• In the use of Class Activation Maps and the H-S deep neural networks, we were able to visualize the localization performance of different models as well as better understand the improvement in performance.
• By enhancing the network, H-S offers many new opportunities to utilize weak supervision to train networks with enormous datasets (millions of images) with less expense on time and human effort.
• After setting the H-S model into PyTorch and allowing users to input data for training, it is clear that the program developed can assist programmers around the world to apply their own data sets to a state-of-the-art network.
• Soon to be made publicly available for use to enhance the accuracy and overall performance of detection/classification and localization.
• We plan to apply the web application to access a webcam for continuously taking images, which are used as frames, to form a near-real-time CAM generator.