Supplementary materials for SCOUT: Self-aware Discriminant Counterfactual Explanation

Pei Wang  Nuno Vasconcelos  
Department of Electrical and Computer Engineering  
University of California, San Diego  
{pew062,nuno}@ucsd.edu

1. Comparison to attributive explanations on segmentation datasets

In the paper, we mainly showed the results on CUB200 [7] due to limited space. The results on ADE20K [8] are shown here in Table 1. The same conclusions as those in the paper can be obtained.

2. More visualization comparison to state of the art

Please see Figure 1.

3. More Visualizations of SCOUT

Please see Figure 2 on CUB200 and Figure 3 on ADE20K.

4. Implementation details

Both datasets were subject to standard normalizations. Training images were first resized to $224 \times 224$ and then randomly flipped, whereas test images were first resized to $256 \times 256$ and then center-cropped to $224 \times 224$. All images were also first converted to $[0.0, 1.0]$ from $[0, 255]$ and then normalized by subtracting the mean ($[0.471, 0.460, 0.454]$) and dividing by the standard deviation ($[0.267, 0.266, 0.271]$) of each RGB color channel. All results are presented on the standard CUB200 test set and the official validation set of ADE20K. Experiments were ran three times. Used classifiers and predictors are trained by standard strategies [3, 1, 2, 6, 4].

5. Attribute Assignment

The parts and attributes of the CUB200 dataset [7] are listed in Table 2 following [5].

References

Table 1: Comparison to attributive explanations (ADE20K): Upper: on beginners, lower: on advanced users.

<table>
<thead>
<tr>
<th>Explanation maps</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a(h_{y^e}(x))$</td>
<td>8.31(0.02)</td>
<td>15.41(0.01)</td>
<td>21.75(0.02)</td>
<td>27.64(0.03)</td>
<td>33.19(0.04)</td>
<td>21.25(0.02)</td>
</tr>
<tr>
<td>$a(h_{y^e}(x)) \cdot a(h_{y^e}(x))$</td>
<td><strong>8.39(0.04)</strong></td>
<td><strong>15.43(0.08)</strong></td>
<td>21.79(0.11)</td>
<td>27.70(0.13)</td>
<td>33.28(0.15)</td>
<td>21.32(0.10)</td>
</tr>
<tr>
<td>$a(h_{y^e}(x)) \cdot a(h_{y^e}(x)) \cdot a(s^e(x))$</td>
<td>8.30(0.05)</td>
<td>15.40(0.06)</td>
<td>21.82(0.09)</td>
<td><strong>27.81(0.11)</strong></td>
<td><strong>33.45(0.15)</strong></td>
<td><strong>21.36(0.09)</strong></td>
</tr>
<tr>
<td>$a(h_{y^e}(x)) \cdot a(h_{y^e}(x)) \cdot a(s^e(x))$</td>
<td>8.31(0.04)</td>
<td>15.39(0.06)</td>
<td><strong>21.83(0.09)</strong></td>
<td><strong>27.81(0.12)</strong></td>
<td><strong>33.45(0.14)</strong></td>
<td><strong>21.36(0.09)</strong></td>
</tr>
<tr>
<td>$a(h_{y^e}(x)) \cdot a(h_{y^e}(x)) \cdot a(s^e(x))$</td>
<td>8.35(0.02)</td>
<td>15.42(0.00)</td>
<td>21.82(0.02)</td>
<td>27.78(0.02)</td>
<td>33.38(0.03)</td>
<td>21.35(0.01)</td>
</tr>
</tbody>
</table>

Table 2: Attributes assignments on CUB200 [7]
True: Scissor tailed Flycatcher (white breast, white forehead, white crown)
Counter: Sayornis (buff breast, brown forehead, brown crown)

True: Horned Grebe (black bill, white breast)
Counter: Red Breasted Merganser (orange bill, red breast)

True: Cape May Warbler (striped back pattern, black wing, eyeline head pattern)
Counter: Yellow Warbler (solid back pattern, yellow wing, plain head pattern)

True: Rock Wren (buff forehead)
Counter: Winter Wren (brown forehead)

Ours
Goyal et al

Figure 1: Comparison of counterfactual explanations (true and counter classes shown below each example, and ground truth class-specific part attributes in parenthesis).
Figure 2: Counterfactual explanations on CUB200 (true and counter classes shown below each example, and ground truth class-specific part attributes in parenthesis).
Figure 3: Counterfactual explanations on ADE20K.