DisPositioNet: Disentangled Pose and Identity in Semantic Image Manipulation
Azade Farshad\textsuperscript{1,2}, Yousef Yeganeh\textsuperscript{1}, Helisa Dhamo\textsuperscript{3}, Federico Tombari\textsuperscript{1,4}, and Nassir Navab\textsuperscript{1}
\textsuperscript{1}Computer Aided Medical Procedures, Technical University of Munich, Germany
\textsuperscript{2}Munich Center for Machine Learning (MCML)
\textsuperscript{3}Huawei Noah’s Ark Lab, UK
\textsuperscript{4}Google, Zurich

Problem Statement

Problem: Semantic Image Manipulation using Scene Graphs

Current Solutions

- Learn Entangled Features for Pose and Appearance
- Complicated and Unpredictable Image Manipulation

Our Contributions:

- Self-Supervised Disentanglement of Pose and Appearance
- Disentangled Scene Graph Neural Network
- Higher Diversity in Image Manipulation
- Outperforming SOTA in Image Generation from Scene Graphs

Experiments and Results

![Figure 2: A comparison of our method against SIMSG [1] on VG dataset.](image)

<table>
<thead>
<tr>
<th>Method</th>
<th>Decoder</th>
<th>All pixels</th>
<th>Roll only</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISG [2]</td>
<td>Pic2pixHD</td>
<td>46.44</td>
<td>28.10</td>
</tr>
<tr>
<td>SIMSG [1]</td>
<td>SPADE</td>
<td>41.88</td>
<td>34.89</td>
</tr>
<tr>
<td>DispositioNet (Ours)</td>
<td>SPADE</td>
<td>41.62</td>
<td>35.30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Roll only</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISG [2]</td>
<td>N/A</td>
</tr>
<tr>
<td>SIMSG [1]</td>
<td>N/A</td>
</tr>
<tr>
<td>DispositioNet (Ours)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

References