SIA-GCN: A Spatial Information Aware Graph Neural Network with 2D Convolutions for Hand Pose Estimation

Deying Kong, Haoyu Ma and Xiaohui Xie
University of California, Irvine
Graph Neural Networks have shown success in many application domains such as computer vision, social networks and chemistry.
Graph Convolutional Network (GCN) by Thomas Kipf

\[ H^{(l+1)} = \sigma \left( \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right) \]

- \( \tilde{A} \): Adjacency matrix with self connections
- \( \tilde{D} \): Degree matrix
- \( H^{(l)} \in \mathbb{R}^{N \times M} \): Matrix of activations in the l-th layer
- \( N \): Number of nodes in the graph
- \( M \): Length of 1-d feature at each node
- \( W^{(l)} \): Trainable weight matrix of layer l
Limitations of the vanilla GCN

- Only processes 1-d feature at each node
  \[ H^{(l+1)} = \sigma \left( \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right) \]
  \[ H^{(l)} \in \mathbb{R}^{N \times M} \]

- All nodes share the same weight matrix \( W \)

What if the feature at each node is 2-dimensional, e.g., 2D confidence maps?
Resize 2-d feature to 1-d feature?
\[ \boxed{\times} \] Would lose spatial information.

What if neighbouring nodes along different edges have different relationships?
SIA-GCN: A Spatial Information Aware Graph Neural Network with 2D Convolutions

- 2D features at each node
- 2D learnable convolution kernels along each edge
- Different 2D kernels for different edges
SIA-GCN: Propagation Rule

\[ X^{(l+1)} = \sigma \left( \hat{A} \left( \left( BX^{(l)} \right) \odot F^{(l)} \right) \right) \]

\( \mathcal{G} = (\mathcal{V}, \mathcal{E}) \) : Graph

\( \mathcal{V} = \{v_1, v_2, \ldots, v_K\} \) : The set of all nodes

\( K \) : Number of nodes in the graph

\( \mathcal{E} \) : The set of all edges

\( \odot \) : Channel-wise 2D convolutional operation

\( \mathbb{R}^{K \times h \times w} \) : Features of all nodes

\( \mathbb{R}^{|\mathcal{E}| \times h' \times w'} \) : Learnable kernels along all edges

\( \mathbb{R}^{|\mathcal{E}| \times K} \) : Broadcast matrix

\( \mathbb{R}^{K \times |\mathcal{E}|} \) : Aggregation matrix
SIA-GCN: A simple example
SIA-GCN: A simple example

Expand undirected edges to directed edges.
Add self connections

- $X_0$: 2D feature at node 0
- $F_0$: 2D convolution kernel along edge 0
We omit the superscript “$l$” in the drawing.
SIA-GCN: A simple example

Broadcast 2D features of each node to their outgoing edges

\[ X^{(l+1)} = \sigma \left( \hat{A} \left( B X^{(l)} \right) \right) \]
Perform 2D convolutions along each edge.

\[
X^{(l+1)} = \sigma \left( \hat{A} \left( (BX^{(l)}) \circledast F^{(l)} \right) \right)
\]
SIA-GCN: A simple example

\[
X^{(l+1)} = \sigma \left( \hat{A} \left( B X^{(l)} \right) \right)
\]

Information aggregation.
SIA-GCN: A simple example

\[
X^{(l+1)} = \sigma \left( \hat{A} \left( (BX^{(l)}) \odot F^{(l)} \right) \right)
\]

Information aggregation.

\[
X_{0}^{\text{new}} = \frac{1}{3} \left( X_0 \odot F_0 + X_1 \odot F_4 + X_2 \odot F_6 \right)
\]
SIA-GCN: A simple example

\[
X^{(l+1)} = \sigma(\hat{A}(BX^{(l)} \odot F^{(l)}))
\]

Information aggregation.

\[
X_0^{\text{new}} = \frac{1}{3} \left( X_0 \odot F_0 + X_1 \odot F_4 + X_2 \odot F_6 \right)
\]

\[
X_1^{\text{new}} = \frac{1}{2} \left( X_0 \odot F_3 + X_1 \odot F_1 \right)
\]
SIA-GCN: A simple example

\[ X^{(l+1)} = \sigma \left( \hat{A} \left( \left( B X^{(l)} \right)^\top F^{(l)} \right) \right) \]

Information aggregation.

\[ X_0^{\text{new}} = \frac{1}{3} \left( X_0 \circledast F_0 + X_1 \circledast F_4 + X_2 \circledast F_6 \right) \]

\[ X_1^{\text{new}} = \frac{1}{2} \left( X_0 \circledast F_3 + X_1 \circledast F_1 \right) \]

\[ X_2^{\text{new}} = \frac{1}{2} \left( X_0 \circledast F_5 + X_2 \circledast F_2 \right) \]
SIA-GCN: Application on 2D hand pose estimation

System diagram of the SiaPose, utilizing SIA-GCN.
Datasets
- CMU Panoptic Hand Dataset
- Largescale Multiview 3D Hand Pose Dataset
- MPII+NZSL Hand Dataset

SIA-GCN: Application on 2D hand pose estimation

Experiments:
- Datasets
  - CMU Panoptic Hand Dataset
  - Largescale Multiview 3D Hand Pose Dataset
  - MPII+NZSL Hand Dataset

- Baselines
  - Convolutional Pose Machine (CPM)
  - Stacked Hourglass (SHG)

- Metric
  - PCK (Percentage of Correct Keypoints):
    the percentage of detections that fall within a normalized distance of the ground truth.
SIA-GCN: Application on 2D hand pose estimation

Some results:

Table 1: SHG based SiaPose on Panoptic Dataset.

<table>
<thead>
<tr>
<th></th>
<th>PCK@ 0.01</th>
<th>PCK@ 0.02</th>
<th>PCK@ 0.03</th>
<th>PCK@ 0.04</th>
<th>PCK@ 0.05</th>
<th>PCK@ 0.06</th>
<th>mPCK</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHG Baseline</td>
<td>35.85</td>
<td>71.47</td>
<td>83.15</td>
<td>88.21</td>
<td>91.10</td>
<td>92.92</td>
<td>77.12</td>
</tr>
<tr>
<td>SharedWeight GCN</td>
<td>34.76</td>
<td>69.66</td>
<td>81.33</td>
<td>86.19</td>
<td>89.14</td>
<td>90.95</td>
<td>75.34</td>
</tr>
<tr>
<td>1-head SiaPose</td>
<td>35.78</td>
<td>71.16</td>
<td>83.57</td>
<td>88.98</td>
<td>92.00</td>
<td>93.84</td>
<td>77.55</td>
</tr>
<tr>
<td>5-head SiaPose</td>
<td>37.53</td>
<td>73.07</td>
<td>84.60</td>
<td>89.51</td>
<td>92.14</td>
<td>93.85</td>
<td>78.45</td>
</tr>
<tr>
<td>10-head SiaPose</td>
<td>37.97</td>
<td>73.53</td>
<td>84.95</td>
<td>89.70</td>
<td>92.26</td>
<td>93.91</td>
<td>78.72</td>
</tr>
<tr>
<td>Improvement</td>
<td>2.12</td>
<td>2.06</td>
<td>1.80</td>
<td>1.49</td>
<td>1.16</td>
<td>0.99</td>
<td>1.60</td>
</tr>
<tr>
<td>10-head R-SiaPose</td>
<td>39.46</td>
<td>77.22</td>
<td>88.45</td>
<td>92.97</td>
<td>94.85</td>
<td>96.09</td>
<td>81.48</td>
</tr>
<tr>
<td>Improvement</td>
<td>3.61</td>
<td>5.75</td>
<td>5.30</td>
<td>4.76</td>
<td>3.75</td>
<td>3.17</td>
<td>4.36</td>
</tr>
</tbody>
</table>
SIA-GCN: Application on 2D hand pose estimation

Some results:

<table>
<thead>
<tr>
<th></th>
<th>PCK@ 0.01</th>
<th>PCK@ 0.02</th>
<th>PCK@ 0.03</th>
<th>PCK@ 0.04</th>
<th>PCK@ 0.05</th>
<th>PCK@ 0.06</th>
<th>mPCK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CMU Panoptic Hand Dataset</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-MGMN [14]</td>
<td>23.67</td>
<td>60.12</td>
<td>76.28</td>
<td>83.14</td>
<td>86.91</td>
<td>89.47</td>
<td>69.93</td>
</tr>
<tr>
<td>AGMN [13]</td>
<td>23.90</td>
<td>60.26</td>
<td>76.21</td>
<td>83.70</td>
<td>87.72</td>
<td>90.27</td>
<td>70.34</td>
</tr>
<tr>
<td>R-SiaPose (Ours)</td>
<td>24.94</td>
<td>62.08</td>
<td>77.83</td>
<td>84.91</td>
<td>88.78</td>
<td>91.34</td>
<td>71.65</td>
</tr>
<tr>
<td><strong>Large-scale Multiview 3D Hand Pose Dataset (MHP)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-MGMN [14]</td>
<td>41.51</td>
<td>85.97</td>
<td>93.71</td>
<td>96.33</td>
<td>97.51</td>
<td>98.17</td>
<td>85.53</td>
</tr>
<tr>
<td>AGMN [13]</td>
<td>41.38</td>
<td>85.67</td>
<td>93.96</td>
<td>96.61</td>
<td>97.77</td>
<td>98.42</td>
<td>85.63</td>
</tr>
<tr>
<td>R-SiaPose (Ours)</td>
<td>41.27</td>
<td>85.89</td>
<td>93.82</td>
<td>96.43</td>
<td>97.61</td>
<td>98.29</td>
<td>85.56</td>
</tr>
</tbody>
</table>
SIA-GCN: Application on 2D hand pose estimation

Qualitative results:

Qualitative results of baseline (top) and our model (bottom) on Panoptic and MPII.
We demonstrated its efficacy by
a) implementing a network for the task of hand pose estimation, and
b) achieving state-of-the-art performance.

Takeaways

• We proposed SIA-GCN, which can
  a) process graphs with 2D features at each node, and
  b) capture different spatial relationships for neighbouring nodes along different edges.

• We demonstrated its efficacy by
  a) implementing a network for the task of hand pose estimation, and
  b) achieving state-of-the-art performance.
Thanks!