Supplementary Material of BubbleFormer: Bubble Diagram Generation via Dual Transformer Models

Jiahui Sun  Liping Zheng  Gaofeng Zhang  Wenming Wu
Hefei University of Technology, China
Outline

• **Network architecture**

• **Questionnaire for perceptual studies comparison to GT**
We adopt the input representation from WallPlan [Sun et al. 2022]. The bubble diagram is visually represented as a raster image. Nodes are represented as circles at their central locations with a certain radius proportional to their space sizes, and we assign pixel values of 1, 2, etc., to different node categories. Connections are visualized as line segments and assigned a pixel value of 2 for edges between the living room node and the other room node and 1 for edges between any two non-living room nodes. An example is shown below.
BubbleFormer for graphic design

NodeFormer

Positional encoding

Encoder

Decoder

Training

Testing

$\mathcal{N}(0, 1)$
Transformer architecture
BNet and Enet architecture

<table>
<thead>
<tr>
<th>Index</th>
<th>Inputs</th>
<th>Operation</th>
<th>Output Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>Multi-channels image</td>
<td>[n,120,120]</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>Conv2d(n,64,2)</td>
<td>[64,60,60]</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>Batch Normalization</td>
<td>[64,60,60]</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>ReLU</td>
<td>[64,60,60]</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>Encoder1(2 layers)</td>
<td>[64,30,30]</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>Encoder2(2 layers)</td>
<td>[128,30,30]</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>Encoder3(2 layers)</td>
<td>[128,30,30]</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>Encoder4(2 layers)</td>
<td>[128,30,30]</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>Avgpool</td>
<td>[128,30,30]</td>
</tr>
</tbody>
</table>
## Encoder module

<table>
<thead>
<tr>
<th>Module</th>
<th>Index</th>
<th>Inputs</th>
<th>Operation</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>1</td>
<td>-</td>
<td>Feature map</td>
<td>[64,30,30]</td>
</tr>
<tr>
<td>Encoder1</td>
<td>2</td>
<td>1</td>
<td>ResnetBlock(64,64,1)</td>
<td>[64,30,30]</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>ResnetBlock(64,64,1)</td>
<td>[64,30,30]</td>
</tr>
<tr>
<td>Encoder2</td>
<td>4</td>
<td>3</td>
<td>ResnetBlock(64,64,1)</td>
<td>[128,30,30]</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
<td>ResnetBlock(64,64,1)</td>
<td>[128,30,30]</td>
</tr>
<tr>
<td>Encoder3</td>
<td>6</td>
<td>5</td>
<td>ResnetBlock(64,64,1)</td>
<td>[128,30,30]</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>6</td>
<td>ResnetBlock(64,64,1)</td>
<td>[128,30,30]</td>
</tr>
<tr>
<td>Encoder4</td>
<td>8</td>
<td>7</td>
<td>ResnetBlock(64,64,1)</td>
<td>[128,30,30]</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>8</td>
<td>ResnetBlock(64,64,1)</td>
<td>[128,30,30]</td>
</tr>
</tbody>
</table>
## Hyper parameters

<table>
<thead>
<tr>
<th>Hyper parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batch size</td>
<td>4</td>
</tr>
<tr>
<td>Epoch</td>
<td>120</td>
</tr>
<tr>
<td>Optimizer</td>
<td>AdamW</td>
</tr>
<tr>
<td>Learning rate*</td>
<td>1.5e-4</td>
</tr>
<tr>
<td>Weight decay</td>
<td>1e-5</td>
</tr>
</tbody>
</table>
Outline

• Network architecture

• Questionnaire for perceptual studies comparison to GT
Based on your design experience, rate the bubble diagram under this boundary, with 1 being the lowest and 5 being the highest.
Thank you