Learning Canonical Representations for Scene Graph to Image Generation

Roei Herzig
TAU

Amir Bar
TAU

Huijuan Xu
Berkeley

Gal Chechik
Bar-Ilan/Nvidia

Trevor Darrell
Berkeley

Amir Globerson
TAU
Image Generation from Scene Graphs (SGs)

Goal: learn a function $F$ for SG-to-image

Input - scene graph

Output - generated image

Task proposed in: Image Generation from Scene Graphs, Johnson et al., CVPR 2018.
Limitation of Current Models

Equivalent inputs lead to different predictions

Semantically equivalent

Man → Left → Woman

Woman → Right → Man
Limitation of Current Models

Equivalent inputs lead to different predictions

Semantically equivalent

\[
\begin{align*}
\text{Man} & \rightarrow \text{Left} \rightarrow \text{Woman} \\
\text{Woman} & \rightarrow \text{Right} \rightarrow \text{Man}
\end{align*}
\]

Two different images

\[
\begin{align*}
\text{Man} & \rightarrow \text{Left} \rightarrow \text{Woman} \\
\text{Woman} & \rightarrow \text{Right} \rightarrow \text{Man}
\end{align*}
\]
Contributions

Our model learns a canonical graph representation from the data that obtains stronger invariance properties.
Contributions

Our model learns a **canonical** graph representation from the data that obtains **stronger invariance** properties.
Contributions

This leads to improved performance on large SGs, robustness to noise in the input SG, and better generalization.
High Level Architecture

Input Scene Graph

Scene Graph Canonicalization

GCN

Scene Layout
What should be the canonical form of an input scene graph?

A natural choice is the “relation-closure”, the graph containing all the possible implied edges.
We deal with two types of completion rules. Converse completion and transitive completions.
Scene Graph Canonicalization

**Goal:** given an input scene graph, compute its *relation-closure*

```
Man → Left → Woman → Left → Boy
```

```
Right

Man

Right

Left

Woman

Right

Left

Boy

Left

“relation-closure”
```
Scene Graph Canonicalization

**Step 1:** converse completions

For every edge, we complete its converse edges.
Scene Graph Canonicalization

Step 1: converse completions
Step 2: transitive completions

For every transitive relation, we construct missing transitive edges
Scene Graph Canonicalization

Step 1: converse completions
Step 2: transitive completions

This procedure produces the *relations-closure*. 
Scene Graph Canonicalization

An assumption of the SGC is that **converse and transitive completion rules are known**.

We next show how to learn this from data.

**Converse Rules**

- Man → Left → Woman
- Man → Right → Woman

**Transitive Rules**

- Man → Left → Woman → Left → Boy
- Man → Left → Boy
Weighted Scene Graph Canonicalization

**Step 1:** converse completions

For every relation pair $r$ we can learn the probability that $r'$ is converse to it.

$$p_{\text{conv}}(r'|r) = \frac{e^{\theta_{r,r'}}}{\sum_{r' \in R \cup \phi} e^{\theta_{r,r'}}}$$

To complete edges, for every edge, we sample from its corresponding converse distribution.
Weighted Scene Graph Canonicalization

**Step 2: transitive completions**

For every relation $r$, we learn the probability that it is transitive.

$$p^{\text{trans}}(r) = \sigma(\theta_r^{\text{trans}})$$

We complete transitive edges and assign them this probability.
GCN for Weighted Scene Graph

Weighted Scene graph

GCN

Predicted Layout
Generation Results

*Sg2im*

*Ours*

*Ground Truth*

*Only scene graph to layout component is compared. Layout to image component is fixed to LostGANs: Sun, W., & Wu, T. (2019). Image synthesis from reconfigurable layout and style. ICCV 2019*
Large Graph Sizes

Improved performance over packed scenes

Example #1

Example #2

Example #3
Large Graph Sizes

Improved performance over packed scenes

<table>
<thead>
<tr>
<th>Method</th>
<th>Standard</th>
<th></th>
<th></th>
<th></th>
<th>Packed</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mIOU COCO</td>
<td>R@0.3 COCO</td>
<td>R@0.5 COCO</td>
<td></td>
<td></td>
<td>mIOU COCO</td>
<td>R@0.3 COCO</td>
<td>R@0.5 COCO</td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>VG</td>
<td>VG</td>
<td></td>
<td></td>
<td>VG</td>
<td>VG</td>
<td></td>
</tr>
<tr>
<td>Sg2Im [17] 5 GCN$^9$</td>
<td></td>
<td>52.4</td>
<td>21.9</td>
<td>32.2</td>
<td>10.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sg2Im [17] 5 GCN$^{10}$</td>
<td>41.7</td>
<td>16.9</td>
<td>62.6</td>
<td>24.7</td>
<td>37.5</td>
<td>9.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sg2Im [17] 8 GCN$^{10}$</td>
<td>41.5</td>
<td><strong>18.3</strong></td>
<td>62.9</td>
<td><strong>26.2</strong></td>
<td>38.1</td>
<td>10.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sg2Im [17] 16 GCN$^{10}$</td>
<td>40.8</td>
<td>16.4</td>
<td>61.4</td>
<td>23.3</td>
<td>36.6</td>
<td>7.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSGC 5 GCN (ours)</td>
<td><strong>41.9</strong></td>
<td>18.0</td>
<td><strong>63.3</strong></td>
<td>25.9</td>
<td><strong>38.2</strong></td>
<td>10.6</td>
<td><strong>39.3</strong></td>
<td><strong>28.5</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** The table highlights improved performance over packed scenes with red circles.
Thank you!

Project page: https://roeiherz.github.io/CanonicalSq2lm/

Poster #5328