Motivation

- Brought the idea of scene graphs to fashion images. Helps in better understanding of fine-grained fashion data.
- A model to generate fashion scene graphs. Using object and relationship detection models, we generated new annotations for this purpose.
- Integrated the attribute detection into the scene graph model.
- Highlighted the application of SG for fashion image retrieval.

Data Annotation

We provide relation detection annotations for Fashionpedia dataset[1].

To train a SG generator, we need the following annotations:

- Fine-grained segmentation:  bounding boxes and object labels
- Relationships in the format of object, subject, predicate

Relationships for a fashion data include:

- Hierarchical. E.g. pocket belongs to shirt
- Attributes. E.g. dress is A-line
- Color. E.g. Jacket is blue

Architecture

Detection

Object Detection:  ResNet50 (1016x444-FPN)
- Weights pre-trained on VG
- Fine-tuned on Fashionpedia
SG Detection:  RelDN[2] (ResNext)
- Trained on our rel. annotations

+ subject bbox/score  + relationships  + predicate scores

Post-processing

Filter detected relationships if:
- score_{subject} * score_{predicate} < thr_{object}
- score_{subject} * score_{predicate} < thr_{attr}

+ subjectbbox/score  + relationships  + predicate scores

Application: Image Retrieval

- To rank the images for a given query image, we represent the predicted scene graph by four matrices:
  - Objects
  - Hierarchical relationships
  - Attributes
  - Colors
- Then we calculate the cosine similarity of each type between the images and the query.

Table 1. RelDN[2] on Visual Genome vs. FashionGraph on Fashionpedia

<table>
<thead>
<tr>
<th>Method</th>
<th>Dataset</th>
<th>R@20</th>
<th>R@50</th>
<th>R@100</th>
</tr>
</thead>
<tbody>
<tr>
<td>ReLDN[2]</td>
<td>Visual Genome</td>
<td>23</td>
<td>31</td>
<td>37</td>
</tr>
<tr>
<td>FashionGraph</td>
<td>Fashionpedia</td>
<td>18</td>
<td>22</td>
<td>24</td>
</tr>
</tbody>
</table>

Figure 1. Example of ground-truth annotations vs. predicted graph

```latex
Figure 2. Qualitative evaluation on fashion image retrieval
```

References
