Relatable Clothing: Detecting Visual Relationships between People and Clothing

Thomas Truong, Svetlana Yanushkevich
Department of Electrical and Computer Engineering

Abstract
Detecting visual relationships between people and clothing in an image has been a relatively unexplored problem in the field of computer vision and biometrics. The lack of readily available public dataset for "worn" and "unworn" classification has slowed the development of solutions for this problem. We present the release of the Relatable Clothing Dataset which contains 35,287 person-clothing pairs and segmentation masks for the development of "worn" and "unworn" classification models. Additionally, we propose a novel soft attention unit for performing "worn" and "unworn" classification using deep neural networks. The proposed soft attention models have an accuracy of upward 98.55±0.35% on the Relatable Clothing Dataset and demonstrate high generalizable, allowing us to classify unseen articles of clothing such as high visibility vests as "worn" or "unworn".

Relatable Clothing Dataset
- There were no datasets containing detailed instance segmentations and visual relationship classes for "worn" and "unworn" detection of clothing before Relatable Clothing.
- Relatable Clothing is a modified version of DeepFashion2 [1].
  - Mask R-CNN trained with Open Images V6 [2,3] to provide person instance segmentations.
  - "unworn" samples added to images with "worn" samples.
  - 29,852 person-clothing pairs for training
  - 5705 person-clothing pairs partitioned into 10 folds for validation and testing.

Soft-Attention Unit
- A trainable unit which guides the "attention" of the network to the areas containing masks.
- Input subject-clothing pair masks are added together and then concatenated before being resized.
- The concatenated masks are then passed through a convolutional filter block to produce the soft-attention unit output.
- The soft-attention unit output is added to the output of the 3x3 convolutional layer of each bottleneck unit in ResNet [4].

Applications
As an early proof of concept, we test our best detector on unseen personal protective equipment samples from the Personal Protective Equipment dataset from Benedetto et al. [5].

Table 2: ResNet101V2 predictions on unseen personal protective equipment images.

<table>
<thead>
<tr>
<th>Class</th>
<th>Soft Attention Unit</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>Specificity (%)</th>
<th>F1 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vest</td>
<td>98.55 ± 0.35</td>
<td>99.16 ± 0.35</td>
<td>98.79 ± 0.35</td>
<td>96.76 ± 0.35</td>
<td>98.37 ± 0.35</td>
</tr>
<tr>
<td>Mask</td>
<td>98.24 ± 0.39</td>
<td>99.16 ± 0.35</td>
<td>98.79 ± 0.35</td>
<td>96.76 ± 0.35</td>
<td>98.37 ± 0.35</td>
</tr>
</tbody>
</table>

Results
Table 1: Summary of results using the proposed soft-attention unit on the Relatable Clothing dataset validation set.

Figure 1: Example samples from the Relatable Clothing Dataset. From left to right: a) the person mask, b) the image, c) the first "unworn" article of clothing, d) the first "worn" article of clothing, e) the second "worn" article of clothing

Figure 2: Proposed Soft-Attention unit used for detecting "worn" and "unworn" articles of clothing given a subject-clothing pair.

Conclusions and Future Work
We proposed a novel soft attention unit for detecting worn and unworn clothing given person and clothing mask pairs. In addition, we release a dataset for "worn" and "unworn" clothing detection titled Relatable Clothing. We achieve the best performance using a ResNet101V2 backbone with 33 of our proposed soft attention units. This model achieved 98.55±0.35% accuracy, 98.58±0.65% specificity, and 98.84±0.29% F1 score on the Relatable Clothing Dataset. We also present some promising use cases in workplace safety through qualitative results on a small high-visibility vest dataset. In our future works we plan on expanding the proposed network to do end-to-end object detection and visual relationship detection. This involves detecting all person-clothing pairs in a given image and classifying "worn" and "unworn" for each pair.

References