ModaNet: A Large-Scale Street Fashion Dataset with Polygon Annotations

Presenter: Shuai (Kyle) Zheng

Joint work with

Fan Yang  Hadi Kiapour  Robinson Piramuthu

Special acknowledgement to the original PaperDoll dataset creator: Kota Yamaguchi
Searching for an ideal dress is challenging...

a). Post the photo of the dress in Reddit

https://www.reddit.com/

b). If you are really Lucky, someone would kindly reply with an where-to-buy link

https://www.ebay.com/

What if I am looking for a similar item using key words directly...
From paper dolls to shop-the-look

Paper dolls are the game to try different clothing for children

Where-to-buy system matches the street clothing photos in online shops

image source: https://upload.wikimedia.org/wikipedia/commons/a/ab/Girlgenderrole3.jpg

PaperDoll Dataset in 2012-2014

PaperDolls dataset was created by Kota and co-authors, the goal behind creating this dataset is to tackle the clothing parsing problem using a retrieval-based approach.

This dataset has over 1 million images, collected from street fashion website, and has pixel-wise annotations for 685 images. The annotations was done using super-pixel-based tool.

Source: Kota Yamaguchi, M Hadi Kiapour, Luis E Ortiz, Tamara L Berg, "Retrieving Similar Styles to Parse Clothing", TPAMI 2014.
Virtual Recommendation Results based on PaperDoll Project

Query

Data size = 262,144

bag belt blouse bracelet cardigan dress necklace shoes skirt

Source: Kota Yamaguchi, M Hadi Kiapour, Luis E Ortiz, Tamara L Berg, "Retrieving Similar Styles to Parse Clothing", TPAMI 2014.
Where are we now with the latest deep-learning-based detection, segmentation, classification techniques for understanding the fashion products from images?
From super-pixel annotations to polygon-based annotations


Image Source: Yamaguchi et al., "Retrieving Similar Styles to Parse Clothing", TPAMI 2014.

super-pixel annotation

Polygon-based annotation used in this work
Example images and annotations in ModaNet

ModaNet dataset has provided polygon-based annotations for 55716 street fashion images in the PaperDoll dataset.

We group highly-related sub-categories to form 13 meta labels

<table>
<thead>
<tr>
<th>Meta labels</th>
<th>Label in icon</th>
<th>Categories included</th>
<th>Avg inst. size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bag</td>
<td>📦</td>
<td>bag</td>
<td>4.88%</td>
</tr>
<tr>
<td>Belt</td>
<td>🎈</td>
<td>belt</td>
<td>0.46%</td>
</tr>
<tr>
<td>Boots</td>
<td>🎈</td>
<td>boots</td>
<td>2.40%</td>
</tr>
<tr>
<td>Footwear</td>
<td>💠</td>
<td>footwear</td>
<td>0.96%</td>
</tr>
<tr>
<td>Outer</td>
<td>🚀</td>
<td>coat, jacket, suit, blazers, cardigan, sweater, jumpsuits, rompers, vest</td>
<td>7.48%</td>
</tr>
<tr>
<td>Dress</td>
<td>🛍</td>
<td>dress, t-shirt dress</td>
<td>10.49%</td>
</tr>
<tr>
<td>Sunglasses</td>
<td>👡</td>
<td>sunglasses</td>
<td>0.31%</td>
</tr>
<tr>
<td>Pants</td>
<td>🣵</td>
<td>pants, jeans, leggings</td>
<td>5.65%</td>
</tr>
<tr>
<td>Top</td>
<td>🚴♂️</td>
<td>top, blouse, t-shirt, shirt</td>
<td>4.83%</td>
</tr>
<tr>
<td>Shorts</td>
<td>🎌</td>
<td>shorts</td>
<td>2.86%</td>
</tr>
<tr>
<td>Skirt</td>
<td>🎩</td>
<td>skirt</td>
<td>6.40%</td>
</tr>
<tr>
<td>Headwear</td>
<td>🎩</td>
<td>headwear</td>
<td>1.25%</td>
</tr>
<tr>
<td>Scarf/tie</td>
<td>💍</td>
<td>scarf, tie</td>
<td>2.55%</td>
</tr>
</tbody>
</table>

Icon image source: https://www.iconfinder.com/search/?q=image
# Comparison between ModaNet and others similar datasets

<table>
<thead>
<tr>
<th></th>
<th>DeepFashion</th>
<th>CFPD</th>
<th>CCP</th>
<th>Fashionista</th>
<th>HPW</th>
<th>ModaNet</th>
</tr>
</thead>
<tbody>
<tr>
<td># of images</td>
<td>800,000</td>
<td>2,682</td>
<td>1,004</td>
<td>685</td>
<td>1,833</td>
<td>55,176</td>
</tr>
<tr>
<td># of categories</td>
<td>50</td>
<td>19</td>
<td>56</td>
<td>53</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Pixel annotation</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Bounding box</td>
<td>landmarks</td>
<td>✓*</td>
<td>✓*</td>
<td>✓*</td>
<td>✓*</td>
<td>✓</td>
</tr>
<tr>
<td>Polygon</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table source is from this work Shuai Zheng, Fan Yang, M. Hadi Kiapour, Robinson Piramuthu. ModaNet: A Large-Scale Street Fashion Dataset with Polygon Annotations. ACM Multimedia, 2018.
Statistics in ModaNet

Number of instances per category

Number of items vs number of images

Figure source is from this work Shuai Zheng, Fan Yang, M. Hadi Kiapour, Robinson Piramuthu. ModaNet: A Large-Scale Street Fashion Dataset with Polygon Annotations. ACM Multimedia, 2018.
Figure source is from this work Shuai Zheng, Fan Yang, M. Hadi Kiapour, Robinson Piramuthu. ModaNet: A Large-Scale Street Fashion Dataset with Polygon Annotations. ACM Multimedia, 2018.
Object Detection for Street Fashion Images Using ModaNet
The landscape of state-of-the-art object detectors

SSD

YOLO (SSD-like detector)

Faster RCNN

Image credit: Speed/accuracy trade-offs for modern convolutional object detectors, CVPR 2017
Object detection performance on ModaNet

Figure 4: Performance comparison of Faster RCNN (left), SSD (middle) and YOLO (right). Best view in color.

- **TP** = True positive
- **TN** = True negative
- **FP** = False positive
- **FN** = False negative

**Precision**

\[ \text{Precision} = \frac{TP}{TP + FP} \]

**Recall**

\[ \text{Recall} = \frac{TP}{TP + FN} \]

**F1 Score**

\[ F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \]
Qualitative Object Detection Results in ModaNet

Object Detection failure cases in ModaNet

Semantic Image Segmentation for Street Fashion Images Using ModaNet
Evolution of Semantic Segmentation with/without CRF

A standard metric that is used to evaluate the performance of semantic segmentation algorithms is Mean IoU (Intersection Over Union), where IoU is defined as:

$$\text{IOU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{A_{\text{pred}} \cap A_{\text{true}}}{A_{\text{pred}} \cup A_{\text{true}}}$$

Table source is from this work Shuai Zheng, Fan Yang, M. Hadi Kiapour, Robinson Piramuthu. ModaNet: A Large-Scale Street Fashion Dataset with Polygon Annotations. ACM Multimedia, 2018.
Results of different approaches

Table source is from this work Shuai Zheng, Fan Yang, M. Hadi Kiapour, Robinson Piramuthu. ModaNet: A Large-Scale Street Fashion Dataset with Polygon Annotations. ACM Multimedia, 2018.
Qualitative Semantic Segmentation Results in ModaNet

Semantic Segmentation failure cases in ModaNet

What you can do with the semantic segmentation immediately

We can predict the color attribute for the fashion products based on the semantic image segmentation results.

Other potential tasks for street fashion images using ModaNet

Polygon Prediction from an single image figure is from Acuna et al., Polygon RNN++, CVPR 2018.

Instance segmentation from an single image figure is from He et al., Mask R-CNN, ICCV 2017.
Thank you

Please check out the ModaNet benchmark hosted on EvalAI, and also checkout the released annotations.