Deep Fashion Understanding

Ziwei Liu

Multimedia Lab, The Chinese University of Hong Kong
Human-centric Analysis
Human-centric Analysis

Person Re-identification
Human-centric Analysis

Face Recognition
Human-centric Analysis

Fashion Understanding
Overall Pipeline

Clothes Detection
Overall Pipeline

Clothes Detection → Clothes Alignment
Overall Pipeline

Clothes Recognition
Clothes Detection

A special class of general object detection

Fast R-CNN

Leverage domain knowledge

Divide & Conquer

Geometric Constraints

SSD
Clothes Alignment

A set of fashion landmarks

Collars
Cuffs
Waistlines
Hemlines
...

(a.1)  (a.2)

(a.3)  (a.4)
Clothes Alignment

More challenging than human pose estimation

Geometry

Appearance
Clothes Alignment

Reduce variations by pseudo-labels

Obtain codebook by k-means clustering in label space
Clothes Alignment

Reduce variations by pseudo-labels

(a) Deep Fashion Alignment (DFA)
Clothes Alignment

Performance

![Bar chart showing performance metrics for different poses and zoom-in levels. The x-axis represents different poses (Normal Pose, Medium Pose, Large Pose, Medium Zoom-in, Large Zoom-in). The y-axis represents detection rate. The chart compares DeepPose, DeepPose (stage 1), IDPR, DFA (stage 1), and DFA. The bars show varying detection rates for each category.]
Clothes Alignment

Relationship to multi-task learning

<table>
<thead>
<tr>
<th>Auxiliary Tasks</th>
<th>CNN</th>
<th>Cascaded CNN</th>
<th>TCDCN</th>
</tr>
</thead>
<tbody>
<tr>
<td>wearing glasses</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>smiling</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>gender</td>
<td>female</td>
<td>male</td>
<td>female</td>
</tr>
<tr>
<td>pose</td>
<td>right profile</td>
<td>frontal</td>
<td>frontal</td>
</tr>
</tbody>
</table>
Clothes Alignment

Relationship to multi-task learning

(a)  
Detection Rate

(b)  
Detection Rate

- Blue: Left Collar
- Orange: Right Collar
- Yellow: Left Sleeve
- Purple: Right Sleeve
- Green: Left Waistline
- Cyan: Right Waistline
- Red: Left Hem
- Blue: Right Hem

Normalized Dist to True Landmark

[Graphs showing detection rate versus normalized distance to true landmark for different parts of clothes.]
Clothes Alignment

More effective representation
Clothes Recognition

Think about the ultimate goal
Clothes Recognition

The interplay between identities and attributes

PID: 2000077658 (Forever21)  Ringer Tee (WOMEN)
Clothes Recognition

The interplay between identities and attributes

Attributes facilitate identification.

Identification discovers attributes.
Clothes Recognition

Attributes are noisy and imbalanced

Multi-label Ranking Loss

\[ J = \sum_{i=1}^{n} \sum_{j=1}^{c_+} \sum_{k=1}^{c_-} \max(0, 1 - f_j(x_i) + f_k(x_i)) \]
Clothes Recognition

The number of identities are huge

Positive Pair

Negative Pair

Hard Negative Mining
Clothes Recognition

In-shop Clothes Retrieval
Clothes Recognition

Consumer-to-shop Clothes Retrieval
Clothes Recognition

Further Analysis
FashionNet

End-to-end System
FashionNet

Forward Pass
FashionNet

Backward Pass
Conclusions

• Large-scale Fashion Dataset **DeepFashion**

• Clothes Alignment by **Fashion Landmarks**

• End-to-end System with **Heterogeneous Supervisions**
Future Work

• From “detection + alignment” to “parsing”

• From “pre-defined attributes” to “free-form descriptions”

• From “single clothes modeling” to “outfit understanding”
Collaborators

Sijie Yan
Shi Qiu
Ping Luo
Xiaogang Wang
Xiaoou Tang
Thanks!

DeepFashion Project by MMLAB, CUHK