Layout-Graph Reasoning for Fashion Landmark Detection Supplementary Material

Weijiang Yu[1], Xiaodan Liang[1,2*], Ke Gong[1,2], Chenhan Jiang[1],
Nong Xiao[1], Liang Lin[1,2]
[1]Sun Yat-sen University, [2]DarkMatter AI Research
weijiangyu8@gmail.com, xdliang328@gmail.com, kegong936@gmail.com,
jchcyan@gmail.com, xiaon6@sysu.edu.cn, linliang@ieee.org

The supplementary material presents more details of our contributed fine-grained fashion landmark dataset (FFLD) and the details of fashion layout.

In Sec.1, we overview the character of FFLD compared with existed fashion landmark datasets, such as key-point number and clothes styles as shown in Fig.1 and Table.1. Then we specifically show more details of FFLD including image collection, image annotation (Table.2) and data statistics (Fig.3 and Fig.2).

In Sec.2, we have shown the specific hierarchical fashion layout designments of FLD, DeepFashion and FFLD.

In Sec.3, we have shown the detailed grammars of FFLD for BCRNN evaluation using same experimental setting [4].

1. Fine-grained Fashion Landmark Dataset (FFLD)

We introduce our Fine-grained Fashion Landmark Dataset, named as FFLD, a new large-scale dataset focusing on comprehensive clothes understanding to benchmark the new challenging fine-grained fashion landmark detection task, which has several appealing properties. First, it is the first and the largest fine-grained fashion landmark dataset to date, with over 200,000 diverse fashion images ranging from well-posed shop images to unconstrained consumer photos, as compared and analyzed in Table 1. Second, FFLD is annotated with rich information including 13 clothes categories and at most 32 landmark key-points, comparing with 8 key-points in DeepFashion-C[6], FLD [3] and ULD [5]. Third, the dataset contains a large percentage of consumer images, which cover a wide variety of human poses, viewpoints, and complex backgrounds. Finally, our FFLD contains a variety of real-world challenges, such as multiple people in a single image, arbitrary human poses and clothes styles. Some example images along with the annotations are shown in Fig. 1. From the detailed comparisons summarized in Table 1, we can observe that FFLD surpasses the existing datasets in terms of scale, the richness of annotations, as well as complexity.

1.1. Image Collection

The images in FFLD are collected from two representative online shopping websites, Taobao and Mogujie, which contain images taken by both the stores and consumers. Each clothing image in-shop is accompanied by several user-taken photos of exactly the same clothing item. As a result, our dataset not only covers the image distribution of professional online retailer stores, but also the other different domains such as street snapshots and selfies. We crawled the images of all categories sorted out in these two shopping websites and collected more than 250k fashion images.

1.2. Image annotation

In order to investigate the task of fine-grained fashion landmark detection, we label elaborate information for each image including clothes bounding boxes, clothes categories and landmark locations.

Category Annotation. We search the most common and favorable clothing types in the shopping websites and then classify them into 13 clothes categories according to their length and the body parts covered by the clothes, which are short-sleeved shirt, skirt, short-sleeved dress, vest dress, trousers, long-sleeved dress, long-sleeved shirt, sling dress, shorts, vest, wipe-bra dress, condole belt clothes and wipe bra. Each image received at most one category label.

Landmark Annotation. Different from existed fashion landmark datasets, we define fine-grained functional regions for 13 clothes categories as shown in Fig.1. For instance, the landmarks for short-sleeved shirt items are defined as left/middle/right collar end, left/right sleeve end,

*Corresponding Author

www.taobao.com
www.mogujie.com

Figure 1: Comparison of fashion landmark definition between existing FLD dataset [3] and our FFLD. FLD dataset only provides landmarks of three clothes items with at most 8 key-points for almost frontal viewed images. In contrast, our FFLD elaborately annotates at most 32 key-points for 13 clothes types example images and includes diverse real-world challenges such as in-shop and consumer images, arbitrary poses and multiple viewpoints.

Table 1: Comparison among the publicly available datasets(*e.g*. MPII[1],LSP[2],DeepFashion-C[6],FLD[3],ULD[5]) for human pose estimation and fashion landmark detection. For each dataset, we report the total number of images, the separate number of images in training, validation, and test sets as well as the number of category and key-points. Note that MPII crops 40K single-person annotated images from 25K images for multiple people.

Dataset	#Total	#Train	#Validation	#Test	Categories	#Key-point
MPII	*40,000	*28,000	-	*11,000	-	16
LSP	12,000	11,000	-	1,000	-	14
DeepFashion-C	289,222	209,222	40,000	40,000	46	8
FLD	123,016	83,033	19,992	19,991	-	8
ULD	30,000	16,000	8,000	6,000	-	8
FFLD	200,000	120,000	40,000	40,000	13	32

left/right shoulder, left/right armpit, left/right chest, and left/right hem. Compared to the previous works [3, 5], we increase the number of key-points of the collar, sleeve, and bottom, and define more key-points of body joints following pose estimation [1], like shoulder and knee *et al*. The visualized differences are shown in Fig. 1. As some of the landmarks are frequently occluded in images, we also labeled the visibility (i.e. whether a landmark is occluded or not) of each landmark.

Quality Control. All images are annotated meticulously by the professional workers. We maintain data quality by manually inspecting and conduct a second-round check for annotated data. We remove the unusable images that are of low resolution, image quality, or whose dominant objects are irrelevant to clothes. In total, 200,000 clothing images are kept to construct FFLD.

1.3. Dataset Statistics

We analyze the dataset statistics of FFLD dataset in detail. First, except a few e-commerce shop images, most of the images in FFLD are taken and submitted by common customers. The percentages of images from shop and customers are presented in Fig. 3 and Fig. 2. In Fig.2, we show the percentages of different clothing types in both shop and consumer images, which displays the percentage of real situations distributing each clothes category. Compared with shop images, the consumer images is more challenge for multiple view and light, complex background and deformable clothes appearance. The statistic analysis of Fig.3 performs the diverse distribution of clothes category, which contains large percentage of consumer images in the whole dataset. So the FFLD has the most variants compared with exited fashion landmark [6, 3, 5], which is also the closest fashion landmark dataset with the real applica-

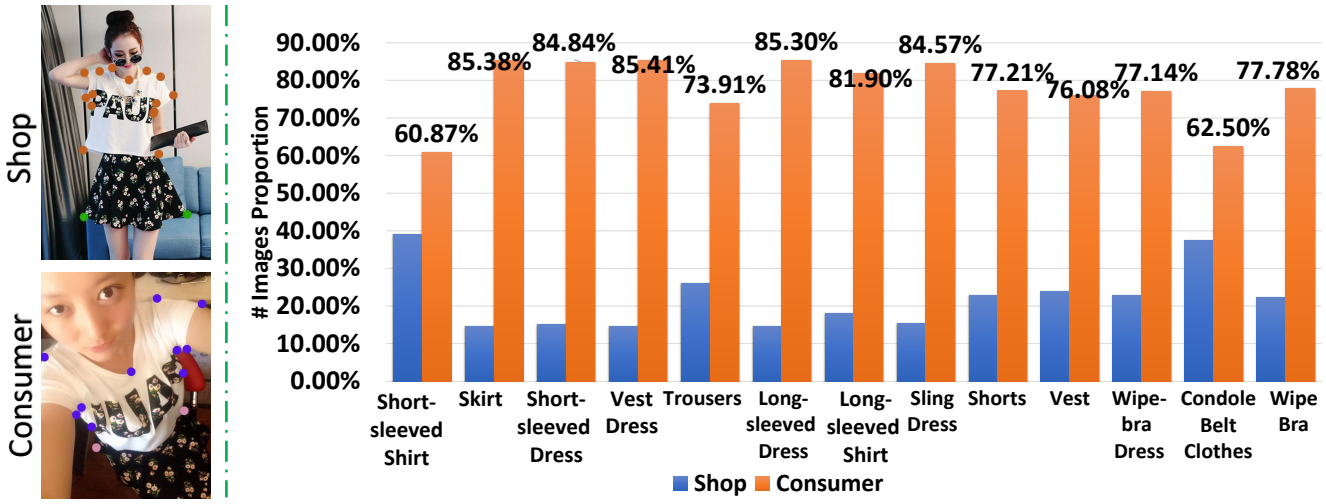

Figure 2: Percentages of different clothing types in both shop and consumer images on FFLD.

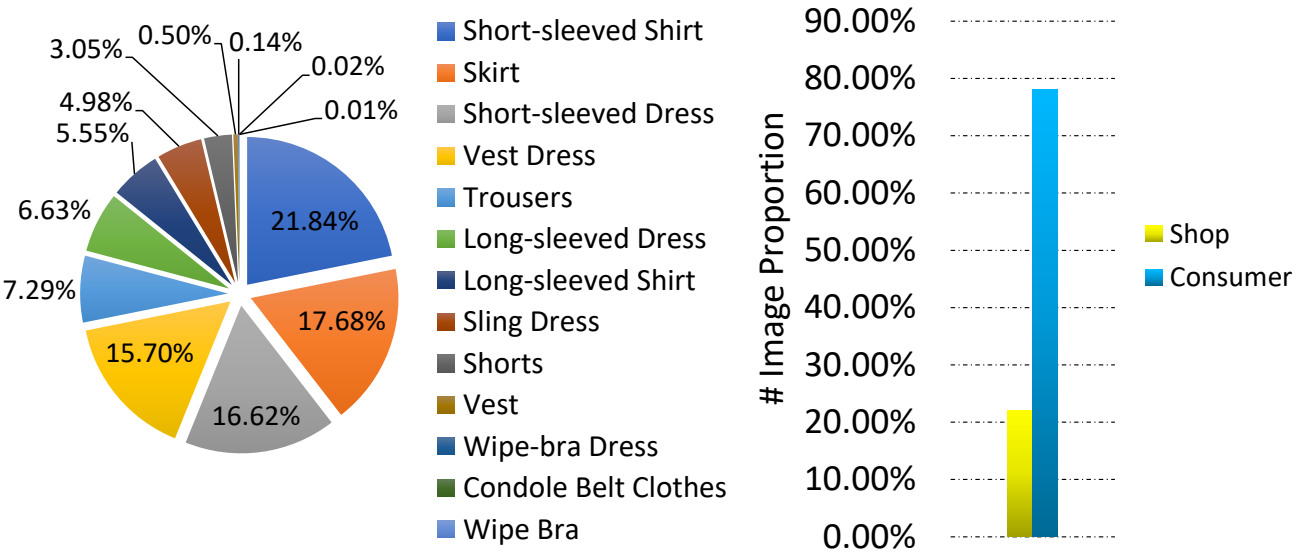

Figure 3: Percentages of distinct clothing types and the image proportions of shop and consumer images in our FFLD.

tion.

2. Layout of Fashion Node

We show layout of fashion nodes in this section. There are four hierarchies of fashion layout, including landmark nodes, clothes-part nodes, body-part nodes and root node.

2.1. Fashion Layout of FLD & DeepFashion

"r.collar"to mean right collar, and "l.collar"to mean leaf collar. "↔"meas connection between two nodes.

Layout of landmark nodes:

r.collar ↔ *l.collar*,

r.waistline ↔ *l.waistline*,

r.sleeve ↔ *l.sleeve*,

r.hem ↔ *l.hem*,

r.collar ↔ *r.waistline* ↔ *r.sleeve* ↔ *r.hem*,

l.collar ↔ *l.waistline* ↔ *l.sleeve* ↔ *l.hem*,

Layout of clothes-part nodes:

collar ↔ *sleeve* ↔ *waistline* ↔ *collar*,

waistline ↔ *hem*,

Layout of body-part nodes:

upper-body ↔ *lower-body*,

Layout of root node:

whole body (self connection)

Table 2: The definition of clothes category landmark. We defined 13 clothes categories with at most 32 landmark key points according to their length and the body parts covered by the clothes. We define a set of key points on the structures of clothes and human body. collar(3) indicates that there are three key points on the collar.

Dataset	Clothes type	Points	Clothes Part
Fashion Landmark Dataset(FLD)	No define	8 point	collar(2), sleeve(2), hem(2), bottom(2)
Fine-grained Fashion Landmark Dataset(FFLD)	Short-sleeved Shirt	15 point	collar(3), sleeve(4), shoulder(2), armpit(2), chest(2), hem(2)
	Skirt	4 point	lower head(2), bottom(2)
	Short-sleeved Dress	17 point	sleeve(4), collar(3), shoulder(2), chest(2), armpit(2), waistline(2), bottom(2)
	Vest Dress	11 point	collar(3), shoulder(2), chest(2), waistline(2), bottom(2)
	Trousers	11 point	waistline(2), bottom(4), knee(4), crotch(1)
	Long-sleeved Dress	21 point	sleeve(4), collar(3), shoulder(2), chest(2), armpit(2), elbow(4), waistline(2), bottom(2)
	Long-sleeved Shirt	19 point	sleeve(4), collar(3), shoulder(2), chest(2), armpit(2), elbow(4), hem(2)
	Sling Dress	11 point	collar(3), shoulder(2), chest(2), waistline(2), bottom(2)
	Shorts	7 point	waistline(2), bottom(4), crotch(1)
	Vest	9 point	collar(3), shoulder(2), chest(2), hem(2)
	Wipe-bra Dress	9 point	collar(3), chest(2), waistline(2), bottom(2)
	Condole Belt Clothes	9 point	collar(3), shoulder(2), chest(2), hem(2)
	Wipe Bra	7 point	collar(3), chest(2), hem(2)

2.2. Fashion Layout of FFLD

"c.collar"to mean center collar, "r.e.elbow"to mean right external elbow, "r.i.elbow"to mean right internal elbow. "↔"means connection between two nodes.

Layout of landmark nodes:

r.collar ↔ *c.collar* ↔ *l.collar* ↔ *r.collar*,
r.shoulder ↔ *l.shoulder*,
r.armpit ↔ *l.armpit*,
r.e.elbow ↔ *r.i.elbow*,
r.e.sleeve ↔ *r.i.sleeve*,
l.e.elbow ↔ *l.i.elbow*,
l.e.sleeve ↔ *l.i.sleeve*,
r.chest ↔ *l.chest*,
r.waistline ↔ *l.waistline*,
r.hem ↔ *l.hem*,
r.collar ↔ *r.shoulder* ↔ *r.e.elbow* ↔ *r.e.sleeve* ↔ *r.i.sleeve* ↔ *r.i.elbow* ↔ *r.armpit* ↔ *r.chest* ↔ *r.waistline* ↔ *r.hem*,
l.collar ↔ *l.shoulder* ↔ *l.e.elbow* ↔ *l.e.sleeve* ↔ *l.i.sleeve* ↔ *l.i.elbow* ↔ *l.armpit* ↔ *l.chest* ↔ *l.waistline* ↔ *l.hem*,
r.lower head ↔ *l.lower head* ↔ *crotch* ↔ *r.lower head*,
r.i.knee ↔ *r.e.knee*,
l.i.knee ↔ *l.e.knee*,
r.e.bottom ↔ *r.i.bottom*,
l.e.bottom ↔ *l.i.bottom*,
r.lower head ↔ *r.e.knee* ↔ *r.e.bottom* ↔ *r.lower head*,

r.i.bottom ↔ *r.i.knee* ↔ *crotch* ↔ *r.i.bottom*,
l.i.bottom ↔ *l.i.knee* ↔ *crotch* ↔ *l.i.bottom*,
l.lower head ↔ *l.e.knee* ↔ *l.e.bottom* ↔ *l.lower head*,
Layout of clothes-part nodes:
collar ↔ *shoulder* ↔ *armpit* ↔ *chest* ↔ *waistline* ↔ *hem*,
waistline ↔ *armpit* ↔ *collar*,
sleeve ↔ *waistline* ↔ *elbow* ↔ *sleeve*,
elbow ↔ *sleeve* ↔ *shoulder* ↔ *elbow*,
lower head ↔ *crotch* ↔ *knee* ↔ *bottom* ↔ *lower head*,
crotch ↔ *bottom*,
knee ↔ *lower head*
Layout of body-part nodes:
upper-body ↔ *lower body*
Layout of root node:
whole body (self connection)

3. Grammars for FFLD in BRCNN

In this section, we introduce our grammar setting in FFLD following the rule from [4]. Following the grammar rule of Wang et al. [4], there are 21 fashion grammars in FFLD totally.

3.1. Symmetry grammar

l.collar ↔ *c.collar* ↔ *r.collar*
l.shoulder ↔ *r.shoulder*
l.armpit ↔ *r.armpit*
r.chest ↔ *l.chest*
r.e.elbow ↔ *r.i.elbow* ↔ *l.i.elbow* ↔ *l.e.elbow*
r.e.sleeve ↔ *r.i.sleeve* ↔ *l.i.sleeve* ↔ *l.e.sleeve*
r.waistline ↔ *l.waistline*
r.hem ↔ *l.hem*
r.lower head ↔ *crotch* ↔ *l.lower head*
r.i.knee ↔ *r.e.knee* ↔ *l.i.knee* ↔ *l.e.knee*
r.e.bottom ↔ *r.i.bottom* ↔ *l.i.bottom* ↔ *l.e.bottom*

3.2. Kinematics grammar

c.collar ↔ *r.collar* ↔ *r.shoulder* ↔ *r.chest* ↔ *r.e.elbow* ↔ *r.e.sleeve*
c.collar ↔ *l.collar* ↔ *l.shoulder* ↔ *l.chest* ↔ *l.e.elbow* ↔ *l.e.sleeve*
r.armpit ↔ *r.i.elbow* ↔ *r.i.sleeve*
l.armpit ↔ *l.i.elbow* ↔ *l.i.sleeve*
r.chest ↔ *r.waistline* ↔ *r.hem*
l.chest ↔ *l.waistline* ↔ *l.hem*
r.lower head ↔ *r.e.knee* ↔ *r.e.bottom*
l.lower head ↔ *l.e.knee* ↔ *l.e.bottom*
crotch ↔ *r.i.knee* ↔ *r.i.bottom*
crotch ↔ *l.i.knee* ↔ *l.i.bottom*

Following the grammars as above, the BCRNN [4] is evaluated on FFLD, and the results as shown in paper (Table.3).

References

[1] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In *CVPR*, June 2014.

[2] S. Johnson and M. Everingham. Clustered pose and nonlinear appearance models for human pose estimation. In *Proceedings of the British Machine Vision Conference*, 2010. doi:10.5244/C.24.12.

[3] Z. Liu, S. Yan, P. Luo, X. Wang, and X. Tang. Fashion landmark detection in the wild. In *ECCV*. Springer, 2016.

[4] W. Wang, Y. Xu, J. Shen, and S.-C. Zhu. Attentive fashion grammar network for fashion landmark detection and clothing category classification. In *CVPR*, pages 4271–4280, 2018.

[5] S. Yan, Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang. Unconstrained fashion landmark detection via hierarchical recurrent transformer networks. In *Proceedings of the 2017 ACM on Multimedia Conference*. ACM, 2017.

[6] S. Q. X. W. Ziwei Liu, Ping Luo and X. Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *CVPR*, June 2016.