The Fashionpedia Ontology and Fashion Segmentation Dataset

Menglin Jia*[1] Mengyun Shi*[1] Mikhail Sirotenko*[3] Yin Cui[1,2]
Bharath Hariharan[1] Claire Cardie[1] Serge Belongie[1,2]

[1]Cornell University [2]Cornell Tech [3]Google AI

Abstract

As a step toward mapping out the visual aspects of the fashion world, we introduce the Fashionpedia ontology and fashion segmentation dataset. The Fashionpedia consists of two parts: (1) an ontology built by fashion experts containing 27 main apparel objects, 19 apparel parts, and 92 fine-grained attributes and their relationships and (2) a dataset consisting of everyday and celebrity event fashion images annotated with segmentation masks and their associated fine-grained attributes, built upon the backbone of the Fashionpedia ontology structure. The aim of our work is to cultivate research connections between the computer vision and fashion communities through the creation of a high quality dataset and associated open competitions, thereby advancing the state-of-the-art in fine-grained visual recognition for fashion and apparel.

1. Introduction

Fashion, in its various forms, influences many aspects of modern societies, having a strong financial and cultural impact. Recent breakthroughs in the field of computer vision have given rise to increased interest in the visual analysis of fashion components. A key component in these recent technological advances is the availability of large amounts of annotated training data of high-quality. Evidence of this can be seen in the engagement of the community in the COCO object recognition dataset [14] and associated challenges that have run annually from 2015 to present. One area that remains challenging for computers, however, is fine-grained visual recognition.

Recently, we have observed an increasing effort to curate datasets for fine-grained visual recognition, evolved from Caltech-UCSD Birds dataset [22] to the recent iNaturalist species classification and detection dataset [20]. The goal of this line of work is to advance the state-of-the-art in automatic image classification for large numbers of real world,

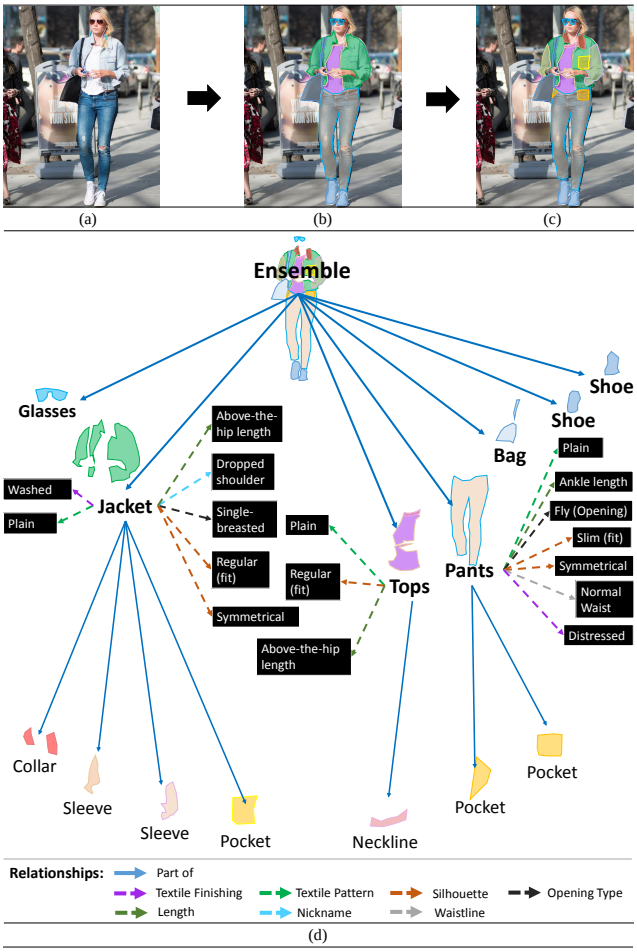

Figure 1. Overview of the Fashionpedia dataset: (a) The original image; (b) The image with main garment segmentation masks; (c) The image with both main garment and garment part segmentation masks; (d) An exploded view of the annotation diagram: the image is annotated with both segmentation masks and fine-grained attributes (black boxes)

fine-grained categories. What is missing for these datasets, however, is the capability of providing a structured repre-

*equal contribution

sentation of an image.

An understanding of the fashion world requires that we complement computers' ability to not only detect objects and attributes but also understand the relationships and interactions between them. In light of this, we introduce the Fashionpedia ontology and image dataset with the aim of training and benchmarking the computer vision models for a more comprehensive understanding of fashion.

The contributions of this work are:

- A fashion ontology informed by product descriptions from the internet and built by fashion experts. Our unified ontology captures the complex structure of fashion objects and ambiguity in descriptions obtained from the web, containing 46 apparel objects (27 main apparel objects and 19 apparel parts), and 92 fine-grained attributes in total.
- A dataset with a total of around 50K clothing images in daily-life, celebrity events, and online shopping annotated by both crowd workers for segmentation masks and fashion experts for fine-grained attributes. The current version of the dataset has 10K images labeled with both segmentation masks and fine-grained attributes, and the rest 40K labeled with segmentation masks only.
- We introduce a novel fine-grained segmentation task and the associated competition [1] by joining forces between the fashion and computer vision communities. The proposed task unifies visual categorization and segmentation of rich apparel attributes, which we believe is an important step toward structural understanding of fashion in real-world applications.

2. Related Work

Table 1 summarizes the comparison among different datasets with clothing category and attribute labels. Our dataset distinguishes itself in the following three aspects:

- **Exhaustive annotation of segmentation masks**: Existing fashion datasets [5, 28] offer segmentation masks for the main garment (*e.g.*, jacket, coat, dress) and the accessory categories (*e.g.*, bag, shoe). The smaller garment objects such as collars and pockets are not annotated. However, these small objects could be valuable for the real world applications such as searching for a specific collar shape during online-shopping. Our datasets are not only annotated with the segmentation masks for a total of 27 main garments and accessory categories, but also 19 garment parts (*e.g.*, collar, sleeve, pocket, zipper, embroidery).
- **Localized attributes**: The fine-grained attributes from existing datasets [15, 9, 27] tend to be noisy, mainly because the annotations are collected by crawling fashion product images associated with attribute-level descriptions directly from large online shopping websites. Unlike these datasets, the fine-grained attributes of our datasets are annotated manually by fashion experts. Furthermore, to the best of our knowledge, our dataset is the first one annotated with localized attributes – fashion experts are asked to annotate the fine-grained attributes associated with the segmentation masks labeled by the crowdworkers. Localized attributes could potentially help computational models detect and understand attributes more accurately.
- **Fine categorization**: Previous study on the attribute categorization suffers from several issues including: (1) repeated attributes belonging to the same category (*e.g.*, zip, zipped and zipper) [15, 8]; (2) only containing basic level categorization (object recognition) and lack of fine categorization (or "subordinate categorization") [5, 28, 11, 21, 25, 24, 12, 18, 2, 19, 10, 6, 23]. (3) Lack of fashion taxonomies with the needs of real-world applications for the fashion industry, possibly due to the research gap in fashion design and computer vision. To better facilitate research in the areas of fashion and computer vision, our proposed ontology is built and verified by fashion experts based on four sources: (1) world-leading e-commerce fashion websites (*e.g.*, ZARA, H&M, Gap, Uniqlo, Forever21); (2) luxury fashion brands (*e.g.*, Prada, Chanel, Gucci); (3) trend forecasting companies (*e.g.*, WGSN); (4) academic resources [4, 1].

[1] Kaggle competition website: https://www.kaggle.com/c/imaterialist-fashion-2019-FGVC6

3. Dataset Specification and Collection

3.1. Fashionpedia ontology and data representation

The Fashionpedia ontology relies on the notions of object (similar to "item" in Wikidata and "object" in Visual Genome [13]) and statement. Objects represent common items in apparels. Statements describe detailed characteristics of an object and consist of a relationship (similar to "property" in Wikidata) and an attribute (similar to "value" in Wikidata). For example, we can add a relationship to specify the silhouette of a garment by associating an attribute for the garment silhouette; or we can assign a material type relationship to a button object by specifying a material attribute. In this section, we break down each component of the Fashionpedia ontology (Figure 2) and explain how a large-scale fashion ontology can be built upon the backbone of the Fashionpedia ontology structure.

Name	Apparel Category Annotation Type			Fine-Grained Attribute Annotation Type		
	Classification	BBox	Segmentation	Unlocalized	Localized	Fine Categorization
Ups and Downs [7]	MG					
Fashion550k [10]	MG, A					
Fashion-MNIST [23]	MG					
Clothing Parsing [25]	MG, A					
Chic or Social [24]	MG, A					
Hipster [12]	MG, A, S					
Runway2Realway [21]			MG, A			
ModaNet [28]		MG, A	MG, A			
Deepfashion2 [5]		MG	MG			
UT Zappos50K [26]	A			✓		
Fashion200K [6]	MG			✓		
Fashion Style-128 Floats [18]	S			✓		
Fashion144k [17]	MG, A			✓		
FashionStyle14 [19]	S			✓		
Main Product Detection [27]		MG		✓		
StreetStyle-27K [16]				✓		✓
UT-latent look [8]	MG, S			✓		✓
FashionAI [3]	MG, GP, A			✓		✓
Apparel classification-Style [2]		MG		✓		✓
DARN [9]		MG		✓		✓
WTBI [11]		MG, A		✓		✓
Deepfashion [15]	S	MG		✓		✓
Fashionpedia		**MG, GP, A**	**MG, GP, A**		✓	✓

Table 1. Comparison of Fashion Datasets (MG = Main Garment, GP = Garment Part, A = Accessory, S = Style).

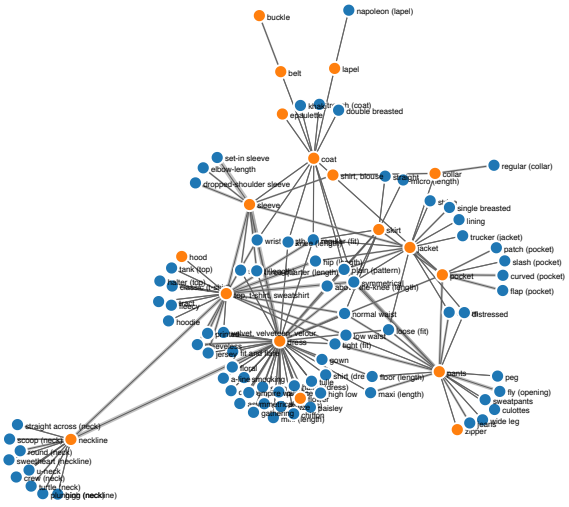

Figure 2. The visualization of the Fashionpedia ontology (based on 20 image samples).

3.1.1 Main garments, and garment parts, accessories and their segmentation masks

In the Fashionpedia dataset, all images were annotated with main garments and each main garment were also annotated with its garment parts. For example, general garment types such as jacket, dress, pants are considered as main garments. These garments also consist of several garment parts such as collars, sleeves, pockets, buttons, and embroideries. Main garments are divided into three main categories: outerwear, intimate and accessories. Garment parts also have different types: garment main parts (*e.g.*, collars, sleeves), bra parts, closures (*e.g.*, button, zipper) and decorations (*e.g.*, embroidery, ruffle). In the current version of Fashionpedia, each image consists of an average of 1 person, 3 main garments, 3 accessories,and 12 garment parts, each delineated by a tight segmentation mask (Figure 1 (b-c)). Furthermore, each object is canonicalized to a synset ID in our Fashionpedia ontology (Figure 2).

3.1.2 Fine-grained attributes

Each main garment and garment part were associated with apparel attributes (Figure 1 (d)). For example, "button" is the part of the main garment "jacket"; "Jacket" can be linked with the silhouette attribute "symmetrical"; Garment part "button" could contain attribute "metal" with relationship of material. Each image in Fashionpedia has an average of 16 attributes. As with main garments and garment parts, we canonicalize all attributes to our Fashionpedia ontology.

3.1.3 Relationships

There are three main types of relationships: 1) outfits to main garments, main garments to garment parts: meronymy (part-of) relationship (Figure 1 (d)); 2) main garments or garment parts to attributes: these relationships types can be garment silhouette (*e.g.*, peplum), collar nickname (*e.g.*, peter pan collars), textile type (*e.g.*, lace), textile finishing (*e.g.*, distressed), or textile-fabric patterns (*e.g.*, paisley), *etc.*; 3) within garments, garment parts or attributes: there are a maximum of four levels of Hyponymy (is-an-instance-of) relationships. For example, weft knit is an instance of knit fabric, and fleece is an instance of weft knit.

3.1.4 Apparel graphs

Integrating the main garments, garment parts, attributes and relationships, we create an apparel graph representation for each outfit in an image. Each apparel graph is a structured representation of an outfit ensemble, containing certain types of garments. Nodes in the graph represent main garments, garment parts, and attributes. Main garments and garment parts are linked to their respective attributes through different types of relationship. The relationships connecting garment objects and attributes point from the main garments to the attributes and from the garment parts to their corresponding attributes. (Figure 1 (d)) illustrates one example of the apparel graph for jacket.

3.1.5 Fashionpedia ontology

While apparel graphs are localized representations of certain outfit ensembles in fashion images, we also create a single Fashionpedia ontology (Figure 2). The Fashionpedia ontology is the union of all apparel graphs and contains entire main garments, garment parts, attributes, and relationships. By doing so, we are able to combine multiple levels of information in a more coherent way.

3.2. Images Collection

A total of 48827 images were harvested from Flickr and the free license photo websites (Unsplash, Burst by Shopify, Freestocks, Kaboompics, and Pexels). Two fashion experts were asked to verify the quality of the collected images manually. The annotation process consist of two phases, firstly, segmentation masks with apparel objects were annotated by crowd workers. Secondly, 15 fashion experts were recruited to annotate the fine grained attributes for the segmentation masks labeled at the first stage.

4. Conclusion

In this work, we propose the Fashionpedia ontology and fashion segmentation dataset. To the best of our knowledge, Fashionpedia is the first dataset that combines part-level segmentation with fine-grained attributes. The expected outcome of this project is to advance the state-of-the-art in domain-specific fine-grained visual recognition. We expect our Fashionpedia image dataset and its associated ontology will have applicability to many applications including better product recommendation for users in online shopping, enhanced visual search results, and resolving ambiguous fashion-related words for textual query. Finally, we expect that our work will act as a catalyst for increased attention to domain-specific ontology for fashion by joining forces between the fashion, computer vision, and natural language processing communities.

5. Acknowledgements

We thank Kavita Bala, Carla Gomes, Dustin Hwang, Rohun Tripathi, Omid Poursaeed, Hector Liu, and Nayanathara Palanivel for their helpful feedback and discussion in the development of Fashionpedia dataset. We also thank Zeqi Gu, Fisher Yu, Wenqi Xian, Chao Suo, Junwen Bai, Paul Upchurch, Anmol Kabra, and Brendan Rappazzo for their help developing the fine-grained attribute annotation tool.

References

[1] Bloomsbury.com. Fashion photography archive. Retrieved May 9, 2019 from https://www.bloomsbury.com/dr/digital-resources/products/fashion-photography-archive/. 2

[2] L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. Van Gool. Apparel classification with style. In *Computer Vision – ACCV 2012*, pages 321–335, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. 2, 3

[3] FashionAI. Retrieved May 9, 2019 from http://fashionai.alibaba.com/. 3

[4] Fashionary.org. Fashionpedia - the visual dictionary of fashion design. Retrieved May 9, 2019 from https://fashionary.org/products/fashionpedia. 2

[5] Y. Ge, R. Zhang, L. Wu, X. Wang, X. Tang, and P. Luo. DeepFashion2: A Versatile Benchmark for Detection, Pose Estimation, Segmentation and Re-Identification of Clothing Images. *arXiv:1901.07973 [cs]*, Jan. 2019. arXiv: 1901.07973. 2, 3

[6] X. Han, Z. Wu, P. X. Huang, X. Zhang, M. Zhu, Y. Li, Y. Zhao, and L. S. Davis. Automatic spatially-aware fashion concept discovery. In *ICCV*, 2017. 2, 3

[7] R. He and J. McAuley. Ups and Downs: Modeling the Visual Evolution of Fashion Trends with One-Class Collaborative Filtering. *Proceedings of the 25th International Conference on World Wide Web - WWW '16*, pages 507–517, 2016. arXiv: 1602.01585. 3

[8] W. Hsiao and K. Grauman. Learning the latent look: Unsupervised discovery of a style-coherent embedding from fashion images. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 4213–4222, Oct 2017. 2, 3

[9] J. Huang, R. Feris, Q. Chen, and S. Yan. Cross-domain image retrieval with a dual attribute-aware ranking network. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ICCV '15, pages 1062–1070, Washington, DC, USA, 2015. IEEE Computer Society. 2, 3

[10] N. Inoue, E. Simo-Serra, T. Yamasaki, and H. Ishikawa. Multi-label fashion image classification with minimal human supervision. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 2261–2267, Oct 2017. 2, 3

[11] M. H. Kiapour, X. Han, S. Lazebnik, A. C. Berg, and T. L. Berg. Where to buy it: Matching street clothing photos in online shops. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 3343–3351, Dec 2015. 2, 3

[12] M. H. Kiapour, K. Yamaguchi, A. C. Berg, and T. L. Berg. Hipster wars: Discovering elements of fashion styles. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 472–488. Springer International Publishing, Cham, 2014. 2, 3

[13] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D. A. Shamma, M. S. Bernstein, and F.-F. Li. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations. *arXiv:1602.07332 [cs]*, Feb. 2016. arXiv: 1602.07332. 2

[14] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 1

[15] Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1096–1104, June 2016. 2, 3

[16] K. Matzen, K. Bala, and N. Snavely. StreetStyle: Exploring world-wide clothing styles from millions of photos. *arXiv preprint arXiv:1706.01869*, 2017. 3

[17] E. Simo-Serra, S. Fidler, F. Moreno-Noguer, and R. Urtasun. Neuroaesthetics in fashion: Modeling the perception of fashionability. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 869–877, June 2015. 3

[18] E. Simo-Serra and H. Ishikawa. Fashion style in 128 floats: Joint ranking and classification using weak data for feature extraction. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 298–307, June 2016. 2, 3

[19] M. Takagi, E. Simo-Serra, S. Iizuka, and H. Ishikawa. What makes a style: Experimental analysis of fashion prediction. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 2247–2253, Oct 2017. 2, 3

[20] G. Van Horn, O. Mac Aodha, Y. Song, Y. Cui, C. Sun, A. Shepard, H. Adam, P. Perona, and S. Belongie. The inaturalist species classification and detection dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8769–8778, 2018. 1

[21] S. Vittayakorn, K. Yamaguchi, A. C. Berg, and T. L. Berg. Runway to realway: Visual analysis of fashion. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 951–958, Jan 2015. 2, 3

[22] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011. 1

[23] H. Xiao, K. Rasul, and R. Vollgraf. Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. *arXiv:1708.07747 [cs, stat]*, Aug. 2017. arXiv: 1708.07747. 2, 3

[24] K. Yamaguchi, T. L. Berg, and L. E. Ortiz. Chic or social: Visual popularity analysis in online fashion networks. In *Proceedings of the 22Nd ACM International Conference on Multimedia*, MM '14, pages 773–776, New York, NY, USA, 2014. ACM. 2, 3

[25] K. Yamaguchi, M. H. Kiapour, L. E. Ortiz, and T. L. Berg. Parsing clothing in fashion photographs. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3570–3577. IEEE, 2012. 2, 3

[26] A. Yu and K. Grauman. Semantic jitter: Dense supervision for visual comparisons via synthetic images. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5570–5579, 2017. 3

[27] L. Yu, E. Simo-Serra, F. Moreno-Noguer, and A. Rubio. Multi-modal embedding for main product detection in fashion. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 2236–2242, Oct 2017. 2, 3

[28] S. Zheng, F. Yang, M. H. Kiapour, and R. Piramuthu. ModaNet: A Large-Scale Street Fashion Dataset with Polygon Annotations. *arXiv:1807.01394 [cs]*, July 2018. arXiv: 1807.01394. 2, 3