arXiv:2305.02307v1 [cs.CV] 3 May 2023

Fashionpedia-Taste：一个旨在解释人类时尚品味的数据集

Mengyun Shi[1] Serge Belongie[2] Claire Cardie[1]
[1]康奈尔大学 [2]哥本哈根大学



图 1：（a）可解释性与推理：我们的 Fashionpedia-taste 数据集对计算机视觉系统提出了挑战，要求其不仅预测受试者是否喜欢一张时尚图像（任务 1），还要从以下 3 个角度提供解释：局部属性（任务 2）、人类注意力（任务 3）和图注（任务 4）；（b）视觉偏好偏差：即使两个受试者喜欢同一件连衣裙，他们喜欢这件连衣裙的原因也可能完全不同（任务 2）。同样，即使两个受试者喜欢这件连衣裙的同一区域，他们喜欢该区域的原因也可能不同（任务 3 和 4）。

摘要

现有的时尚数据集没有考虑导致消费者喜欢或不喜欢某张时尚图像的多重因素。即使两位消费者喜欢同一张时尚图像，他们喜欢这张图像的原因也可能完全不同。在本文中，我们研究消费者喜欢某张特定时尚图像的原因。为此，我们引入了一个可解释性数据集 *Fashionpedia-Taste*，其中包含丰富的注释，从以下 *3* 个视角解释受试者为何喜欢或不喜欢某张时尚图像：*1*）局部属性；*2*）人类注意力；*3*）图注。此外，我们还要求受试者提供其个人属性以及对时尚的偏好，例如个性和偏好的时尚品牌。我们的数据集使研究人员能够从不同的人文视角和模态出发，构建计算模型，以全面理解和解释消费者的时尚品味

。*Fashionpedia-Taste* 已开放

位于：[1]

1. 引言

为什么用户会在社交平台上展示的时尚图像下点击“喜欢”？他们通过“喜欢”来表达自己对时尚的偏好。在时尚领域，服装是电子商务购物类别中排名第二的品类。因此，充分了解用户的时尚品味，可能会对时尚电子商务发挥重要作用。

为了提高消费者购买时尚产品的可能性，人们开发了多种推荐系统，并取得了不错的效果。然而，即使推荐系统为消费者做出了正确的推荐，这是否意味着模型真的理解了用户喜欢这张图片的原因？答——

[1]Fashionpedia 项目页面：fashionpedia.github.io/home/

Task	数据集名称
识别	iMat [3],深度 [15],服装 [29], F-MNIST [27], F-550k [11],F-128 [21], F-14 [23], 潮人 [14]
检测	ModaNet [32], 深度2 [2],主干 [19]
数据挖掘	复古 [9], 时尚 [28], 升级 [5], 潜在 [6],地理 [16] F-144k [20],F-200K [4], 街头 [17], T台 [25]
检索	DARN [10],WTBI [13],Zappos [30],深度 [15] 胶囊 [7], POG [1], VIBE [8], IQ [26]
属性定位	Fashionpedia [12]
可解释性与推理	**Fashionpedia-Taste**

表 1：与以往工作相比，Fashionpedia-Taste 是唯一一项从不同人文视角和模态研究人类时尚品味的研究。

案尚不明确。此外，即使两位消费者喜欢同一张图片，他们喜欢这张图片的原因也可能截然不同，如图所示。1据我们所知，此前尚无研究探讨过这一问题。

在本文中，我们介绍了一个可解释时尚品味数据集 Fashionpedia-Taste。该数据集要求受试者从 3 个视角为其喜欢或不喜欢某一时尚图像的原因提供解释：1）局部属性；2）人类注意力；3）图注。此外，我们还收集了受试者额外的个人偏好信息，例如购买连衣裙时偏好的裙长、个性、品牌和细粒度类别。这些信息也可能与用户对时尚图像的偏好相关。

本工作的目标是推动未来研究，鼓励对时尚品味可解释性研究开展更多探索，并缩小人类与机器对图像理解之间的差距。本工作的贡献包括：1）构建了一个可解释时尚品味数据集，其中包含 10,000 条表达（对图像的喜欢或不喜欢）、52,962 个选定属性、20,000 条人类注意力标注以及 20,000 条用于解释受试者时尚品味的图注，涵盖超过 1,500 张独特图像和 100 名独特受试者；2）形式化定义了一项新任务，该任务不仅要求模型预测受试者是否喜欢某一时尚图像，还要求从 3 个不同视角（局部属性、人类注意力和图注）解释其原因。

2. 相关工作

时尚数据集 以往大多数时尚数据集都聚焦于识别、检测、数据挖掘或检索任务（表 1）。在用户与时尚图像交互这一领域，Fashion IQ [26] 提供由人类撰写的图注，通过自然语言反馈区分相似的成对服装图像。ViBE [8] 则引入了一个数据集，旨在基于用户特定的体型理解其时尚偏好。Fashionpedia-Ads 研究了用户广告与时尚品味之间的相关性。与 Fashion IQ、ViBE 和 Fashionpedia-Ads 不同，我们的数据集聚焦于结合视觉和文本信号，对用户的时尚偏好进行解释和推理。时尚领域之外，与我们最相关的工作是 VCR [31], ，它要求模型正确作答，并为答案提供合理的解释。与 VCR 不同，我们的数据集要求模型通过不同模态（任务 1/2/3/4）完成更复杂的多阶段推理，以判断受试者的时尚品味，如图 1所示。

3. 受试者的时尚品味标注

受试者与标注流程我们从一所美国大学招募了100名女性受试者。用户标注流程分为两部分：1）收集受试者的基本信息（第3.1节）；2）收集受试者对给定服装长度类别的时尚品味（第3.2节）。所有受试者都必须完成这两项流程。

3.1. 基本信息调查

本调查用于收集受试者的：1）基本信息（性别、族裔、年龄）；2）个性；3）基本时尚偏好（偏好的时尚品牌、时尚属性和类别）；4）偏好的服装长度，用于确定在第3.2节所述调查中应为每位受试者分配哪种裙长的图像。

个性 与 [18], 类似，我们使用包含10道选择题的问卷来测量受试者的个性。我们收集个性数据，是因为我们想了解受试者的个性与其时尚品味之间是否存在相关性。

基本时尚偏好 我们收集用户的时尚偏好（时尚类别、时尚属性和品牌），因为我们想了解这种自我报告的时尚偏好是否与他们在第 3.2 节中测得的偏好一致。

3.2. 时尚品味调查

任务设计 在时尚品味调查中，受试者会根据其在基本信息调查（第 3.1 节）中报告的偏好连衣裙长度，获得 100 张连衣裙图像。他们需要说明自己是否喜欢这些连衣裙，并说明喜欢或不喜欢的原因。对于给定的每张连衣裙图像，他们都需要完成以下 4 项任务：

- 任务 1：判断他们是否喜欢给定的连衣裙。
- 任务2——属性选择：解释哪些方面使他们喜欢或不喜欢某件连衣裙。

• 任务3-人类注意力：指出（绘制多边形）连衣裙中让他们喜欢或不喜欢给定连衣裙的区域。

• 任务 4-文本解释：解释他们在任务 3 中标注的区域为何会让他们喜欢或不喜欢某件连衣裙。

为何设计这 4 项任务 任务 2、3 和 4 让受试者能够从 3 种不同的视角和模态来解释自己的时尚品味。任务 2 让受试者能够从细粒度属性的角度解释自己的时尚品味。不过，任务 2 可能无法捕捉到某些只能通过视觉解释的信息。任务 3 用于解决这一问题，并模拟人类注视捕捉，让受试者能够以视觉方式解释自己的时尚品味。此外，为了充分理解受试者在任务 3 中标注的区域，我们使用任务 4 让受试者进一步通过文字解释为何标注这些区域。

喜欢和不喜欢的比例失衡：我们预计，如果只询问受试者喜欢某件连衣裙的原因，数据将会严重失衡。为了解决这一问题，我们要求用户在任务 2、3 和 4 中，解释让他们喜欢或不喜欢给定图像的方面。

4. 数据集分析

4.1. 用户标注分析

用户基本信息

任务1—喜欢/不喜欢我们收集了4766条喜欢和5234条不喜欢的标注，涉及1500多张独特图像和100位用户。图中展示了每位用户选择喜欢和不喜欢的频次。2，结果表明，大多数用户的喜欢与不喜欢比例相当均衡。均衡的喜欢与不喜欢比例可能有助于训练偏差较小的模型。

任务2—属性选择表2 按不同裙长细分了每位用户选择喜欢和不喜欢属性的频次。平均而言，用户选择的喜欢属性（3.9397）几乎是不喜欢属性（1.4565）的3倍。这表明，在解释自己为何喜欢或不喜欢一条连衣裙时，大多数用户倾向于选择自己喜欢的属性，而非不喜欢的属性。

表 3 展示了标注属性及其对应细粒度类别（超类别）的详细信息。「廓形」所包含的喜欢属性占比最高（21.9%）。这表明，大多数用户的时尚偏好由连衣裙的形状决定。相比之下，「纺织品整理与制造工艺」（纺织品饰面、制造技术）所包含的不喜欢属性占比最高（23.9%）。这说明，即使用户喜欢一条连衣裙的「廓形」，仍可能因为该类别而不喜欢这条连衣裙。

长度类型	选择类型	# 总计	# 平均值
所有长度	已点赞	38397	3.9397
所有长度	不喜欢的	14565	1.4565
mini	已喜欢	7991	3.89
mini	不喜欢的	3117	1.52
以上	已点赞	8001	3.63
以上	不喜欢	3076	1.39
以下	喜欢	6312	3.60
以下	不喜欢	2561	1.46
midi	喜欢	8463	4.03
midi	不喜欢	3076	1.46
maxi	喜欢	7630	4.01
maxi	不喜欢	2735	1.43

表2：任务 2—用户标注的喜欢的属性和不喜欢的属性数量。

超级类别	# 喜欢属性	频次	# 不喜欢属性	频次
连衣裙款式	2385	6.2%	1132	7.7%
廓形	8419	21.9%	1786	12.2%
纺织图案	3194	8.3 %	2031	13.9 %
纺织品饰面，制造技术	5656	14.7 %	3493	23.9 %
非纺织品类型	49	0.1 %	49	0.33 %
领口款式	6611	17.2 %	2283	15.6 %
领型	387	1 %	101	0.7 %
驳领款式	28	0.1 %	4	0.02 %
袖型	2179	5.6 %	726	4.9 %
袖长	1917	4.9 %	592	4.1 %
口袋款式	236	0.6 %	84	0.5 %
开启类型	691	1.8%	453	3.1 %
腰线	4146	10.7 %	849	5.8 %
裙长	2499	6.5 %	982	6.7 %

表 3：任务 2——用户针对每个细粒度属性类别标注的属性总数。

图3。（a）和（b）展示了用户标注的喜欢和不喜欢属性的分布。结果表明，「印花」和「正常腰线」是一些用户喜欢连衣裙的主要原因。不过，这些属性也可能导致用户不喜欢一条连衣裙，因为用户可能喜欢某种印花图案，却不喜欢另一种印花图案。为了更好地解释用户的时尚品味，需要训练一个模型，使其不仅能够理解用户对特定属性的偏好，还能理解该属性所在区域的像素级图案。为此，我们要求用户完成任务3。

任务3-人类注意力 表 4 展示了用户针对各种裙长标注的人类注意力总数。不同裙长对应的注意力标注数量分布均匀。

任务 4-任务3的文本解释 任务4包含11.3

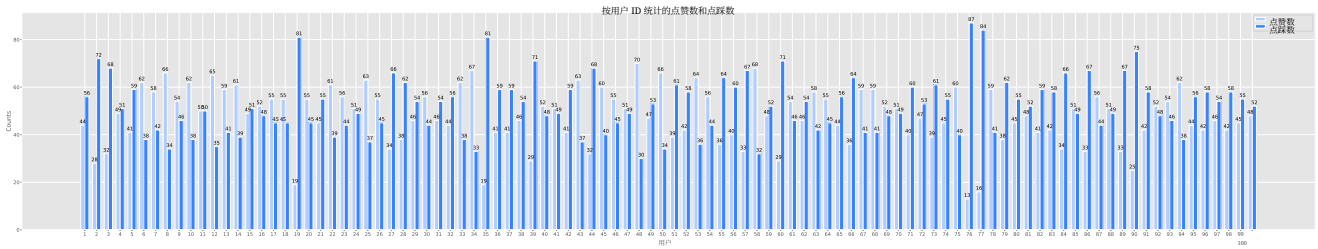
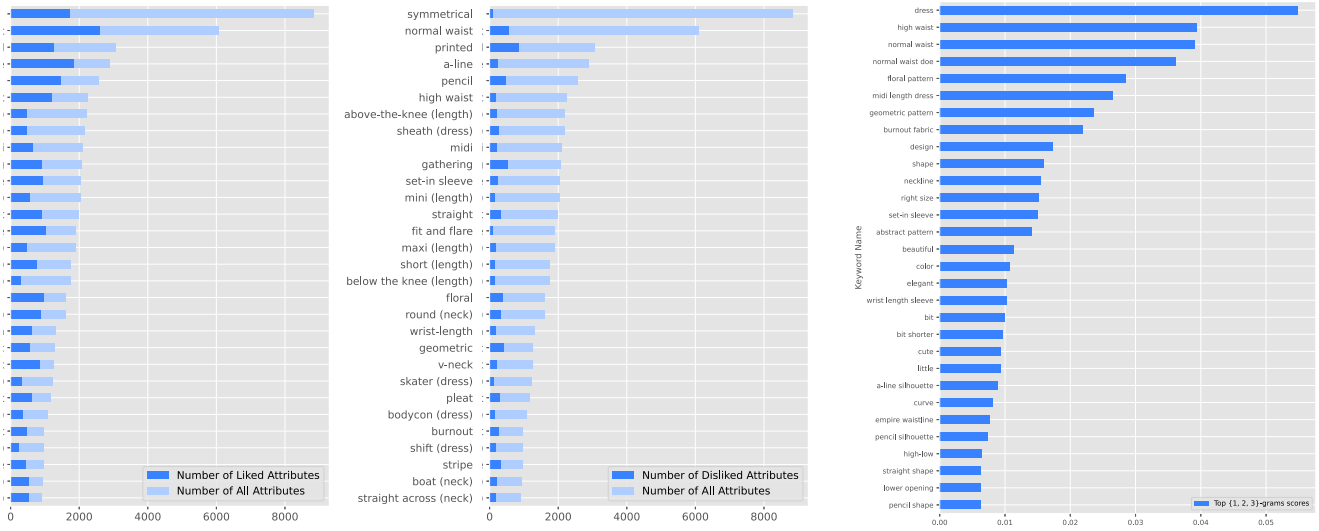


图 2：任务 1——用户的点赞数和不喜欢数。



(a) 喜欢的属性分布。 (b) 被点踩属性分布。 (c) 词数统计：1 元、2 元和 3 元组的数量。

图 3：任务 2——受试者标注的喜欢和不喜欢的属性分布；任务 4——词数统计

长度类型	# 总注视次数
所有长度	20000
mini	4100
以上	4400
以下	3500
midi	4200
maxi	3800

表4：任务3——用户对每种连衣裙长度标注的人类注意力总次数。

每位用户平均使用的词语数，以及每个子任务平均使用的5.6个词语（任务4.1用于解释喜欢的内容，任务4.2用于解释不喜欢的内容）。

词数统计：我们使用 Textacy 中的 SGRank [24]来计算词语的出现频率。图 3.(c) 展示了我们数据集中出现频率最高的 1 元、2 元和 3 元词组。与腰线（高腰线、正常腰线和高腰线）及图案（碎花、几何图案、抽象图案）相关的词语出现频率较高，用户在任务3中用这些词语解释吸引他们注意力的因素。

语言统计：我们使用 spaCy 的词性（POS）标注 [22] 为任务4中标注的图注中的名词、专有名词和形容词进行标注。表 5 展示了按词性分类的最高频唯一词语数量。我们发现，出现频率最高的普通名词更多与对连衣裙的高层次描述相关，例如领口、图案和腰线。相比之下，出现频率最高的专有名词更多与连衣裙的细节描述相关，例如贴花、珠饰和彼得潘领。这体现了我们数据集在语言上的多样性。

5. 结论

在这项工作中，我们研究了时尚产品图像中的人类品味问题。我们介绍了一个可解释的时尚品味数据集 Fashionpedia-Taste，旨在从三个角度理解时尚品味：1）局部属性；2）人类注意力；3）图注。本工作的目标是推动未来研究，鼓励对时尚品味可解释性研究开展更多探索，并缩小人类与机器在图像理解方面的差距。

词性类别	Word
Noun	连衣裙、颈部、长度、领口、设计、形状、图案、腰部，颜色，袖子，曲线，腰线，面料，裙子
专有名词	长款，贴花，珠饰，和服，彼得潘领，褶裥，分层式挂脖式，蝙蝠袖，开衩，帐篷式，条纹，镂空，猎豹纹
Adj	优雅的，美丽的，好看的，可爱的，高的，直的，完美的碎花，宽松，优美，性感，合身的，迷人的，有魅力的

表5：任务4-语言统计：按词性划分的唯一词语数量。

参考文献

[1] Wen Chen、Pipei Huang、Jiaming Xu、Xin Guo、Cheng Guo、Fei Sun、Chao Li、Andreas Pfadler、Huan Zhao 和 Bin-qiang Zhao。《Pog：面向阿里巴巴 ifashion 时尚推荐的个性化服装搭配生成》。载于第 25 届 ACM SIGKDD 国际知识发现与数据挖掘会议论文集，第 2662–2670 页，2019 年。2[2] Yuying Ge、Ruimao Zhang、Xiaogang Wang、Xiaoou Tang 和 Ping Luo。《Deepfashion2：用于服装图像检测、姿态估计、分割和重新识别的通用基准》。载于 CVPR，2019 年。2[3] Sheng Guo、Weilin Huang、Xiao Zhang、Prasanna Srikhanta、Yin Cui、Yuan Li、Hartwig Adam、Matthew R Scott 和 Serge Belongie。《iMaterialist 时尚属性数据集》。载于 ICCV Workshops，2019 年。2[4] Xintong Han、Zuxuan Wu、Phoenix X. Huang、Xiao Zhang、Menglong Zhu、Yuan Li、Yang Zhao 和 Larry S. Davis。《自动化的空间感知时尚概念发现》。载于 ICCV，2017 年。2[5] Ruining He 和 Julian McAuley。《潮起潮落：使用单类协同过滤建模时尚趋势的视觉演变》。载于 WWW，2016 年。2[6] Wei-Lin Hsiao 和 Kristen Grauman。《学习潜在的“外观”：从时尚图像中无监督发现风格一致的嵌入》。载于 ICCV，2017 年。2[7] Wei-Lin Hsiao 和 Kristen Grauman。《从时尚图像创建胶囊衣橱》。载于 IEEE 计算机视觉与模式识别会议论文集，第 7161–7170, 2018 – 2 页。Wei-Lin Hsiao 和 Kristen Grauman。《ViBE：为不同体型进行穿衣搭配》。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第 11059–11069 页，2020 年。2[9] Wei-Lin Hsiao 和 Kristen Grauman。《从文化到服装：从百年的时尚图像中发现世界事件》。载于 IEEE/CVF 国际计算机视觉会议论文集，第 1066–1075 页，2021 年。2[10] Junshi Huang、Rogerio Feris、Qiang Chen 和 Shuicheng Yan。《利用双属性感知排序网络进行跨域图像检索》。载于 ICCV，2015 年。2[11] Naoto Inoue、Edgar Simo-Serra、Toshihiko Yamasaki 和 Hiroshi Ishikawa。《最少人类监督下的多标签时尚图像分类》。载于 ICCV，2017 年。2[12] Menglin Jia、Mengyun Shi、Mikhail Sirotenko、Yin Cui、Claire Cardie、Bharath Hariharan、Hartwig Adam 和 Serge Belongie. Fashionpedia：本体、分割与属性定位数据集。载于《欧洲计算机视觉会议论文集》，第316–332页。Springer，2020年。2[13] M. H. Kiapour、X. Han、S. Lazebnik、A. C. Berg 和 T. L. Berg。《去哪儿买：在线商店中的街头服饰照片匹配》。载于 ICCV，2015年。2[14] M Hadi Kiapour、Kota Yamaguchi、Alexander C Berg 和 Tamara L Berg。《潮人之战：发现时尚风格元素》。载于 ECCV，2014年。2[15] Ziwei Liu、Ping Luo、Shi Qiu、Xiaogang Wang 和 Xiaoou Tang。《DeepFashion：通过丰富标注提升服装识别与检索的鲁棒性》。载于 CVPR，2016年。2[16]Utkarsh Mall、Kevin Matzen、Bharath Hariharan、Noah Snavely 和 Kavita Bala。《GeoStyle：发现时尚趋势与事件》。载于 IEEE/CVF国际计算机视觉会议论文集，第411–420页，2019年。2 – .[17] Kevin Matzen、Kavita Bala 和 Noah Snavely。《StreetStyle：从数百万张照片中探索全球服装风格》。arXiv预印本 arXiv:1706.01869，2017年。2[18] Nils Murrugarra-Llerena 和 Adriana Kovashka。《面向检索的跨模态个性化》。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第6429–6438页，2019年。2[19] Antonio Rubio、LongLong Yu、Edgar Simo-Serra 和 Francesc Moreno-Noguer。《用于时尚主商品检测的多模态嵌入》。载于 ICCV，2017年。2[20] Edgar Simo-Serra、Sanja Fidler、Francesc Moreno-Noguer 和 Raquel Urtasun。《时尚中的神经美学：时尚感知度建模》。载于 CVPR，2015年。2[21] Edgar Simo-Serra 和 Hiroshi Ishikawa。《128维浮点数中的时尚风格：利用弱标注数据进行联合排序与分类以提取特征》。载于 CVPR，2016年。2[22] spaCy。工业级自然语言处理。https://spacy.io/，2022年。4[23] Moeko Takagi、Edgar Simo-Serra、Satoshi Iizuka 和 Hiroshi Ishikawa。《何为一种风格：时尚预测的实验分析》。载于 ICCV，2017年。2[24] Textacy。《Textacy：spaCy之前与之后的自然语言处理》。https://textacy.readthedocs.io/en/latest/，2022年。4[25] S. Vittayakorn、K. Yamaguchi、A. C. Berg 和 T. L. Berg。《从秀场到街头：时尚的视觉分析》。载于 WACV，2015年。2[26] Hui Wu、Yupeng Gao、Xiaoxiao Guo、Ziad Al-Halah、Steven Rennie、Kristen Grauman 和 Rogerio Feris。《Fashion IQ：一个旨在通过自然语言反馈检索图像的新数据集》。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第11307–11317页，2021年。2[27] Han Xiao、Kashif Rasul 和 Roland Vollgraf。《Fashion-MNIST：用于机器学习算法基准测试的新型图像数据集》。arXiv预印本 arXiv:1708.07747，2017年。2[28] Kota Yamaguchi、Tamara L Berg 和 Luis E Ortiz。《时尚还是社交：在线时尚网络中的视觉人气分析》。载于 ACM MM，2014年。2

[29] Kota Yamaguchi、M Hadi Kiapour、Luis E Ortiz 和 Tamara L Berg。时尚照片中的服装解析。载于 *CVPR*，2012年。 2[30] Aron Yu 和 Kristen Grauman。语义抖动：通过合成图像进行视觉比较的密集监督。载于 *ICCV*，2017年。 2[31] Rowan Zellers、Yonatan Bisk、Ali Farhadi 和 Yejin Choi。从识别到认知：视觉常识推理。载于 *IEEE/CVF*计算机视觉与模式识别会议论文集，第6720–6731页，2019年2– 。[32] Shuai Zheng、Fan Yang、M Hadi Kiapour 和 Robinson Piramuthu。ModaNet：一个带有多边形标注的大规模街头时尚数据集。载于*ACM MM*，2018年。 2