Supplementary Material: Personalized Fashion Recommendation via Deep Personality Learning

Dongmei Mo[1,2]
dongmei.mo@connect.polyu.hk
Xingxing Zou[1,2]
xingxingzou@aidlab.hk
Wai Keung Wong[1,2,†]
calvin.wong@polyu.edu.hk

[1] School of Fashion and Textiles
The Hong Kong Polytechnic University
Hong Kong SAR, China

[2] The Laboratory for Artificial Intelligence in Design, Hong Kong SAR, China

Overview. The supplementary material is organized as follows: Section 1 revisits the related works; Section 2 provides more details of the SOP dataset and the experimental implementation; Section 3 introduces comparison methods and the evaluation metrics; Section 4 presents additional yet important experimental results that omitted in the main paper due to space constraints.

1 Related works

1.1 Fashion Compatibility

Fashion compatibility can be learned via pairwise compatibility or outfit compatibility learning in an end-to-end manner [16, 23, 25]. To consider the relationship between pairwise compatibility and overall compatibility, McAuley et al. proposed a single latent space to model the compatibility of different items in the same outfit [11]. The conditional similarity networks [20] [21] modeled multiple latent spaces to capture the notions of pairwise similarity for outfit compatibility learning [20]. The bidirectional LSTM was used in [3] [15] to treat the outfit as an item sequence and learn the item relationship for outfit compatibility learning. However, these methods are sensitive to item orders, which is not practical for online outfit compatibility evaluation.

Instead of just learning item-item compatibility from the aesthetics aspect, GP-BPR combined item-item and user-item interactions via general compatibility modeling and personal preference modeling for personalized clothing matching [18]. Collaborative metric learning learns a joint metric space to encode the relationships of both user-user and item-item similarity [4]. **Unlike all previous works only focusing on fashion items, we aim to dig into the relations between personality and fashion styles.**

†Corresponding author

Aspect	Attribute	occurrence	Aspect	Attribute	occurrence
body figure	triangle	10,710	skin colour	yellow	1,886
	spoon	9,058		dark	1,940
	bottom hourglass	8,913		brown	2,273
	inverted triangle	4,620	height	high	2,538
	round	10,116		middle	304
	diamond	9,872		low	1,927
	...	...			

Table 1: Physical label distribution on O4U dataset.

1.2 Fashion Personality

Fashion style analysis benefits fashion decision-making for both fashion professionals and fashion lovers. Although there are methods for fashion analysis [6] [1], the correlation between fashion style and human personality has not yet been fully explored in machine learning. Automatic personality prediction on multimodal data has attracted lots of attention in affective computing [13]. Yash et al. proposed a personality assessment method to predict user personality to release the limitation that most traditional personality learning methods heavily rely on hand-crafted and theory-based text features [12]. Theres et al. proposed a personality-based recommender systems for human-computer interaction by taking the information of personality traits to increase personal relevance and trust and acceptance [22]. Raffaele et al. explored the relationship between the personality traits and luxury brand attachment by taking a survey of around $1,500$ international luxury customers [2], and the result highlighted the positive relationship of personality learning and brand attachment.

Unlike above, we integrate automatic user personality prediction with outfit style mapping to an end-to-end deep neural model to learn the fashion personality. The outfit compatibility regarding individual physical attributes is also considered during the recommendation process. With the help of fashion personality and physical compatibility, personalized recommendation can be more effective from both psychological and physical sides.

2 More details of Dataset and Implementation

2.1 More details of the SOP dataset

The outfits of the personality-outfit mapping pairs on SOP dataset are from the O4U dataset [7] which has total $15,748$ compatible outfits with rich physical label information (briefly shown in Table 1).

2.1.1 Twitter data collection

The SOP dataset contains user personality annotations and the outfit information on physical attributes and fashion attributes. The data are collected from twitter[1] platform by
(1) Data collection. We collect relevant tweets that contains keywords or hashtag of "personality" within past six months (i.e. from 01 Jan. 2022 to 01 Jun. 2022).
(2) Data filtering. For the collected tweets, we iterate each ID to search online tweets. The users who have posted at least 3 relevant tweets are considered as qualified users and their top 100 tweets are downloaded and stored in an independent json file.

[1] https://twitter.com

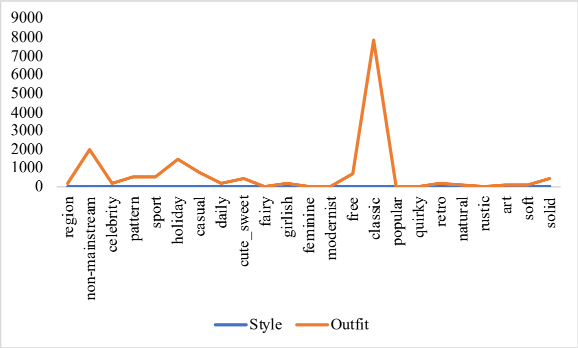


Figure 1: Outfit style distribution on the O4U dataset.

(3) Data formatting. After filtering the data, we have total 969 qualified users, and each user is represented with a user index (i.e. from 1 to 969) for user privacy protection. The tweets of each user will be fed to a psycholinguistic learning model for personality prediction.

2.1.2 User personality prediction

The Bert-based model is used to tokenize the word features with a dimensionality of 768 for all the tweets. Then the personality prediction model [12] pre-trained on Kaggle[2] personality dataset is used to predict the MBTI personality of each user. MBTI is a self-report inventory that describes a person's personality from 16 types of binary combinations on four dimensions: introversion versus extraversion, sensing versus intuiting, thinking versus feeling, and judging versus perceiving [14][12]. Introduction about different MBTI personality types can be found from the official website[3].

2.1.3 Fashion style prediction.

We predict fashion styles for the outfits and explore personality-style mapping for personalized recommendations. The model [9] predicts the style for each item in the outfit, and the overall outfit style is set as the item's style with the most significant score. There is a total of 230 styles on DeepFashion [9]. We summarize similar styles, remove the styles that are irrelevant to personality, and finally conclude the styles into 23 groups. The prediction distribution of the outfits towards different fashion styles is shown in Figure 1.

2.2 Implementation Details

The P-Net and the comparison methods use the same backbone, i.e., Resnet18 that pre-trained on ImageNet dataset, with an image size of $3 \times 224 \times 224$ and output dimension of $d = 512$. The Transformer comprises 4 attention heads with 3 layers. The optimizer of the proposed method is SGD with weight decay of 4×10^{-4} and learning rate $lr = 0.002$. The batch size is 32, and the maximum training epoch is 50 for all methods, and the training will be stopped when the validation accuracy stops increasing. The parameters γ and λ of the

[2]https://www.kaggle.com/datasets/datasnaek/mbti-type
[3]https://www.16personalities.com/personality-types

proposed method are set as 0.01 while the parameters of the comparison methods are set as default as indicated by the original papers. The comparison methods are trained end-to-end for solving the metric learning and multi-label classification problems. We test each checkpoint and report the best results on the testing set. The proposed method is implemented with PyTorch [17], and the code will be released for reproducible research.

3 Comparison Methods and Evaluation Metrics

3.1 Comparison Methods

Several baselines and state-of-the-art methods are used in the experiment for a fair comparison with the proposed method. The classification baseline is Resnet18, while the baseline of metric learning is Conditional Similarity Networks (CSN) [21], which learns different pairwise similarities in distinct subspaces (conditions) for specific similarity comparison. The type-aware method considers the specific type-aware similarity of fashion items in an outfit to capture the stylistic relationship for outfit compatibility learning [20]. SCE-Net captures similarity among different data pairs without requiring explicit condition labels to release the limitation of the generalization capability of the CSN-based methods [19]. Multi-layered Comparison Network (MCN) takes image representations at different layers of the Resnet to learn pairwise compatibility and interpret the prediction with potential aspects from multiple layers. The outfit compatibility score is computed based on pairwise compatibility, and the most problematic items that affect the compatibility can be tracked via backpropagation gradients [23]. Transformer-based Dual Relation Graph network (TDRG) constructs a structural relation graph with a cross-scale transformer-based architecture and a semantic relation graph with explicit semantic-aware constraints and combines the two graphs for effective multi-label classification [24]. Learnable Personalized Anchor Embedding (LPAE) utilizes a stacked self-attention mechanism to weight the importance of different fashion items and learn compact item representations in the outfit space. A set of anchors characterize the user embeddings, and the compatibility among different users and outfits is computed with matrix-vector multiplication in the user space [10].

3.2 Tasks and Evaluation Metrics

This paper focuses on personalized recommendation via personality-style learning and physical compatibility evaluation, which can be considered as metric learning and multi-label classification problems. Metric learning is expected to learn a practical function that can map the high-dimensional features to a low-dimensional space with similar samples lying closer than the dissimilar ones. The performance of metric learning can be evaluated with Accuracy (ACC) based on the positive-negative pair comparison and the Normalized Discounted Cumulative Gain (NDCG) [5]. The NDCG definition can be found in [5] and the NDCG in this paper is the mean NDCG over all users. The widely-used evaluation metrics for the multi-label classification problem is mean average precision (mAP), average per-class precision (CP), recall (CR), F1 (CF1), and the average overall precision (OP), recall (OR), F1 (OF1) [8].

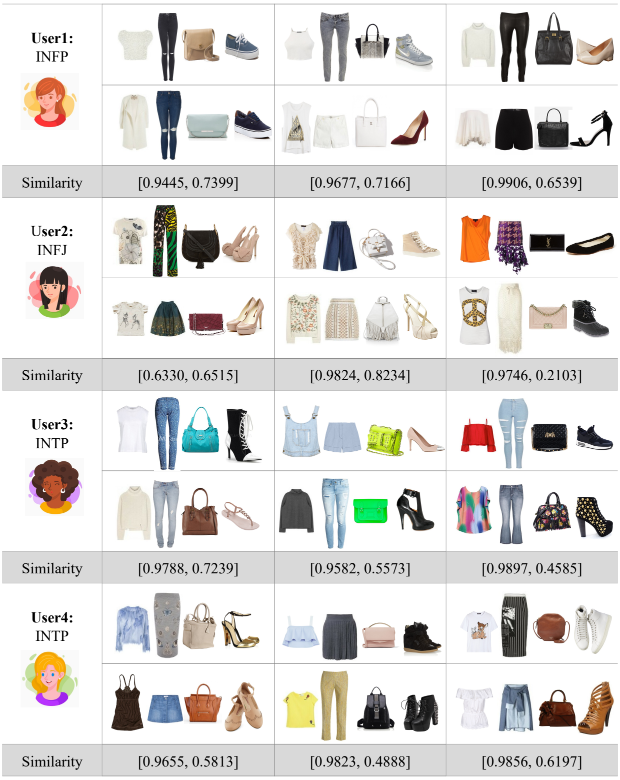


Figure 2: Recommendation of P-Net based on testing (the top row) and train (the bottom row) sets.

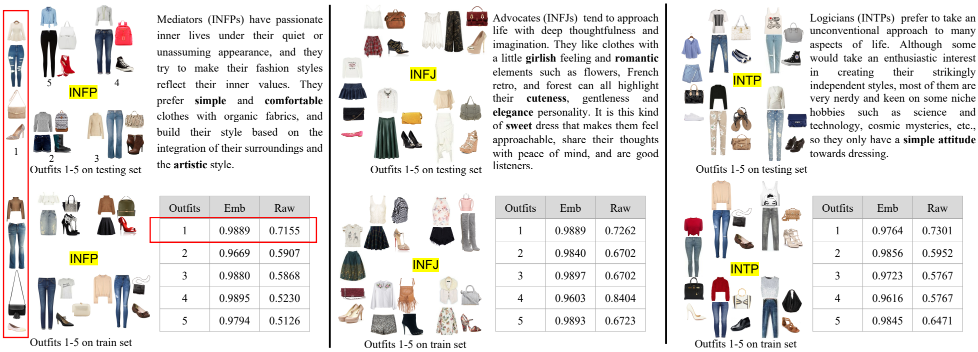
4 More Experimental Results of P-Net

4.1 How is P-Net for individual fashion personality learning?

Figure 2 shows the recommended outfits for different users from the testing and the train sets. The similarity score between the recommended testing and training outfits that are computed with the feature embeddings and raw images is also presented. Figure 2 indicates that the recommendations from the testing and train sets usually have similar colour combinations, patterns and styles, which shows that the trained model can learn the potential preference of different users and discover similar outfits from the unseen testing set. Additionally, users 1 and 2 with different personalities would have different preference as user 1 prefers simple and solid colours while user 2 likes outfits with cute prints or multiple colours. This indicates that P-Net can distinguish the fashion personality of users with different personalities.

4.2 How is P-Net for fashion style learning respecting different MBTI types?

We explore the relationships between different personality types and fashion styles by comparing the text description of the personality types and the visual characteristics of the recommended outfits. Figure 3 shows 5 recommended outfits from the testing and train sets as



Outfits	Emb	Raw
1	0.9889	0.7155
2	0.9669	0.5907
3	0.9880	0.5868
4	0.9895	0.5230
5	0.9794	0.5126

Outfits	Emb	Raw
1	0.9889	0.7262
2	0.9840	0.6702
3	0.9897	0.6702
4	0.9603	0.8404
5	0.9893	0.6723

Outfits	Emb	Raw
1	0.9764	0.7301
2	0.9856	0.5952
3	0.9723	0.5767
4	0.9616	0.5767
5	0.9845	0.6471

Figure 3: Recommendation samples for different personalities on testing (the top row) and training (the bottom row) sets. The texts explain the fashion styles corresponding to different personalities.

examples, and we can know that:

(1) The fashion styles of outfits from testing and train sets are consistent, which indicates that the common similarity of different users from the same personalities is captured by the proposed P-Net.

(2) Compared with the recommendations in Figure 2, the recommendations for a specific personality type instead of an individual present visual variety in a wider range, which is reasonable as the recommended choice is made on the average preference of all users.

(3) The tables in Figure 3 show that the similarity of feature embeddings from testing and train sets tends to be higher than that of the raw images. For example, the first outfits from the top and bottom row on the left column have an embedding similarity of 0.9889 while the raw similarity is 0.7155. The potential reason is that P-Net can capture the abstract features like outfit style to represent the images and balance the effect of the low-level features like colour, pattern and texture for similarity comparison, while the raw similarity focuses more on the low-level features. Thus, for any outfits from a similar style, the embedding similarity can be different from the raw similarity.

4.3 Can P-Net discover the potential between personalities and styles?

Figure 4 shows the similarity scores of the personalities and the fashion styles from two evaluation metrics: pairwise distance (the top colorbar) and Cosine similarity (the bottom colorbar), while Figure 5 shows the Cosine similarity of the pairwise personalities. Figure 4 further verifies that INFP presents strong preference on the styles of natural and art, which is consistent with the personality analysis in Figure 3. Both INFP and INTP would pick clothes that are casual, natural and comfortable without considering details that make them quirky while other personalities like ENFP and ENFJ would pay more attention to this style to reflect their attitudes. Figure 5 indicates that the personalities of INTP, INTJ, INFP and INFJ have close relations. Specifically, INFP and INTP are the closest, which provides us inspiration to learn fashion personality from the users who belong to a close personality.

4.4 How is the performance based on larger network?

Tables 2 and 3 show the performance of all methods based on Resnet50 for outfit-style mapping learning and outfit physical attribute prediction. It is interesting that the P-Net achieves

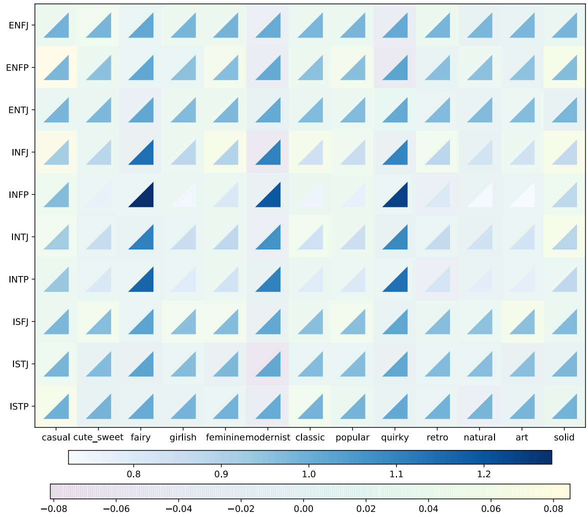


Figure 4: The visualization of the similarity of the MBTI personalities and fashion styles. The first and second colorbars are corresponding to pairwise distance and Cosine similarity, respectively.

Method	Testing100		Testing	
	ACC	NDCG	ACC	NDCG
Resnet50	0.4985	0.5522	0.498	0.5538
CSN	0.6345	0.5372	0.6327	0.5367
T-Aware	0.6601	0.533	0.6633	0.5307
SCE-Net	0.4871	0.5592	0.4872	0.5585
MCN	0.4907	0.5739	0.4918	0.5748
TDRG	0.4582	0.5440	0.4385	0.5265
LAPE	0.6695	0.752	0.6723	0.7531
Ours	**0.7404**	**0.7915**	**0.7676**	**0.8176**

Table 2: Comparison of different methods based on Resnet50 backbone.

improved prediction results for physical compatibility learning based on ResNet50, even though the ACC and NDCG results are not as high as expected. This suggests that employing a larger network, such as ResNet50, can indeed enhance performance for multi-label physical compatibility problems. This observation aligns with the general principle that deeper and more complex networks excel at capturing intricate patterns. However, it's worth noting that they may become prone to overfitting if not properly utilized, particularly in binary classification tasks.

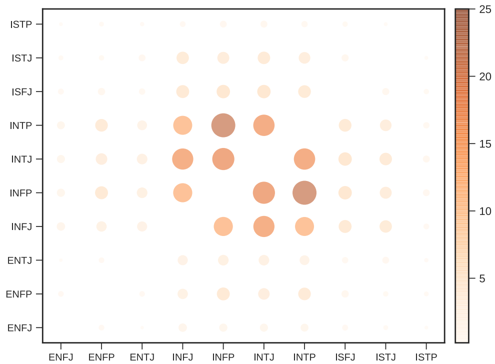


Figure 5: The visualization of the similarity among different MBTI personalities.

Method	Testing100							Testing						
	mAP	CP	CR	CF1	OP	OR	OF1	mAP	CP	CR	CF1	OP	OR	OF1
Resnet50	39.31	37.52	32.40	34.77	53.14	46.71	49.72	38.91	37.49	32.26	34.68	52.82	46.76	49.61
CSN	33.70	31.06	29.12	30.05	53.40	47.48	50.27	34.02	31.75	29.41	30.54	54.19	48.31	51.08
T-Aware	36.47	35.15	28.77	31.64	58.36	48.88	53.20	36.52	35.20	28.73	31.63	58.88	49.33	53.68
SCE-Net	37.20	37.17	32.14	34.47	56.86	50.23	53.34	36.92	36.83	32.01	34.25	57.01	50.57	53.60
MCN	37.27	31.44	29.78	30.58	59.42	50.34	54.50	37.28	32.00	29.83	30.88	59.70	50.67	54.82
TDRG	31.87	19.21	21.12	20.12	58.70	40.37	47.84	31.85	19.35	21.14	20.21	58.98	40.62	48.11
LAPE	41.07	**41.19**	29.94	34.67	61.80	46.92	53.34	40.90	**41.87**	29.85	34.85	61.90	47.17	53.54
Ours	**44.09**	37.89	**33.06**	**35.31**	**65.94**	**52.61**	**58.53**	**43.80**	38.21	**32.97**	35.40	**65.72**	**52.84**	**58.58**

Table 3: Comparison of different methods based on Resnet50 backbone on physical compatibility prediction task.

References

[1] Lavinia De Divitiis, Federico Becattini, Claudio Baecchi, and Alberto Del Bimbo. Style-based outfit recommendation. In *2021 International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–4. IEEE, 2021.

[2] Raffaele Donvito, Gaetano Aiello, Laura Grazzini, Bruno Godey, Daniele Pederzoli, Klaus-Peter Wiedmann, Chris Halliburton, Priscilla Chan, Junji Tsuchiya, Irina Ivanovna Skorobogatykh, et al. Does personality congruence explain luxury brand attachment? the results of an international research study. *Journal of Business Research*, 120:462–472, 2020.

[3] Xintong Han, Zuxuan Wu, Yu-Gang Jiang, and Larry S Davis. Learning fashion compatibility with bidirectional lstms. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 1078–1086, 2017.

[4] Cheng-Kang Hsieh, Longqi Yang, Yin Cui, Tsung-Yi Lin, Serge Belongie, and Deborah Estrin. Collaborative metric learning. In *Proceedings of the 26th international conference on world wide web*, pages 193–201, 2017.

[5] Yang Hu, Xi Yi, and Larry S Davis. Collaborative fashion recommendation: A functional tensor factorization approach. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 129–138, 2015.

[6] Youngseung Jeon, Seungwan Jin, and Kyungsik Han. Fancy: human-centered, deep learning-

based framework for fashion style analysis. In *Proceedings of the Web Conference 2021*, pages 2367–2378, 2021.

[7] Waikeung Wong Kaicheng Pang, Xingxing Zou. Dress well via fashion cognitive learning. *arXiv preprint arXiv:2208.00639*, 2022.

[8] Jack Lanchantin, Tianlu Wang, Vicente Ordonez, and Yanjun Qi. General multi-label image classification with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16478–16488, 2021.

[9] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1096–1104, 2016.

[10] Zhi Lu, Yang Hu, Yan Chen, and Bing Zeng. Personalized outfit recommendation with learnable anchors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12722–12731, 2021.

[11] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, pages 43–52, 2015.

[12] Yash Mehta, Samin Fatehi, Amirmohammad Kazameini, Clemens Stachl, Erik Cambria, and Sauleh Eetemadi. Bottom-up and top-down: Predicting personality with psycholinguistic and language model features. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 1184–1189. IEEE, 2020.

[13] Yash Mehta, Navonil Majumder, Alexander Gelbukh, and Erik Cambria. Recent trends in deep learning based personality detection. *Artificial Intelligence Review*, 53(4):2313–2339, 2020.

[14] Isabel Briggs Myers and Linda K Kirby. Introduction to type: a guide to understanding your results on the myers-briggs type indicator, palo alto, 1998.

[15] Takuma Nakamura and Ryosuke Goto. Outfit generation and style extraction via bidirectional lstm and autoencoder. *arXiv preprint arXiv:1807.03133*, 2018.

[16] Kaicheng Pang, Xingxing Zou, and Waikeung Wong. Dress well via fashion cognitive learning. *arXiv preprint arXiv:2208.00639*, 2022.

[17] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[18] Xuemeng Song, Xianjing Han, Yunkai Li, Jingyuan Chen, Xin-Shun Xu, and Liqiang Nie. Gp-bpr: Personalized compatibility modeling for clothing matching. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 320–328, 2019.

[19] Reuben Tan, Mariya I Vasileva, Kate Saenko, and Bryan A Plummer. Learning similarity conditions without explicit supervision. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10373–10382, 2019.

[20] Mariya I Vasileva, Bryan A Plummer, Krishna Dusad, Shreya Rajpal, Ranjitha Kumar, and David Forsyth. Learning type-aware embeddings for fashion compatibility. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 390–405, 2018.

[21] Andreas Veit, Serge Belongie, and Theofanis Karaletsos. Conditional similarity networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 830–838, 2017.

[22] Sarah Theres Völkel, Ramona Schödel, Daniel Buschek, Clemens Stachl, Quay Au, Bernd Bischl, Markus Bühner, and Heinrich Hussmann. Opportunities and challenges of utilizing personality traits for personalization in hci. *Personalized human–computer interaction*, page 31, 2019.

[23] Xin Wang, Bo Wu, and Yueqi Zhong. Outfit compatibility prediction and diagnosis with multi-layered comparison network. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 329–337, 2019.

[24] Jiawei Zhao, Ke Yan, Yifan Zhao, Xiaowei Guo, Feiyue Huang, and Jia Li. Transformer-based dual relation graph for multi-label image recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 163–172, 2021.

[25] Xingxing Zou, Kaicheng Pang, Wen Zhang, and Waikeung Wong. How good is aesthetic ability of a fashion model? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21200–21209, June 2022.