Personalized Capsule Wardrobe Creation with Garment and User Modeling

Xue Dong
Shandong University
dongxue.sdu@gmail.com

Xuemeng Song*
Shandong University
sxmustc@gmail.com

Fuli Feng
National University of Singapore
fulifeng93@gmail.com

Peiguang Jing
Tianjin University
pgjing@tju.edu.cn

Xin-Shun Xu
Shandong University
xuxinshun@sdu.edu.cn

Liqiang Nie*
Shandong University
nieliqiang@gmail.com

ABSTRACT

Recent years have witnessed a growing trend of building the capsule wardrobe by minimizing and diversifying the garments in their messy wardrobes. Thanks to the recent advances in multimedia techniques, many researches have promoted the automatic creation of capsule wardrobes by the garment modeling. Nevertheless, most capsule wardrobes generated by existing methods fail to consider the user profile, including the user preferences, body shapes and consumption habits, which indeed largely affects the wardrobe creation. To this end, we introduce a combinatorial optimization-based personalized capsule wardrobe creation framework, named PCW-DC, which jointly integrates both garment modeling (*i.e.,* wardrobe compatibility) and user modeling (*i.e.,* preferences, body shapes). To justify our model, we construct a dataset, named bodyFashion, which consists of 116, 532 user-item purchase records on Amazon involving 11,784 users and 75,695 fashion items. Extensive experiments on bodyFashion have demonstrated the effectiveness of our proposed model. As a byproduct, we have released the codes and the data to facilitate the research community.

CCS CONCEPTS

• **Information systems** → **Personalization**; **Web applications**;

KEYWORDS

Fashion Analysis; Compatibility Learning; User Modeling

ACM Reference Format:

Xue Dong, Xuemeng Song, Fuli Feng, Peiguang Jing, Xin-Shun Xu, and Liqiang Nie. 2019. Personalized Capsule Wardrobe Creation with Garment and User Modeling. In *Proceedings of Proceedings of the 27th ACM International Conference on Multimedia (MM '19).* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3343031.3350905

* Xuemeng Song and Liqiang Nie are corresponding authors.

MM '19, October 21–25, 2019, Nice, France

ACM ISBN 978-1-4503-6889-6/19/10…$15.00
https://doi.org/10.1145/3343031.3350905

Figure 1: Illustration of the personalized capsule wardrobe (PCW) creation. Given the original wardrobe of a user, a PCW is created by adding (green box) and deleting (red box) some garments.

1 INTRODUCTION

Capsule wardrobe (CW) is a minimum collection of garments (*e.g.,* clothes and shoes), with diverse combinations to inspire people to pair up various compatible outfits [1]. Apparently, the capsule wardrobe plays a crucial role in people's daily life by saving time and money spent on dressing appropriately [2]. In practice, capsule wardrobes are usually created by fashion experts through manually selecting garments and evaluating the potential outfits. To relieve the burden of labor cost, recent researches in multimedia have generated reasonable CWs by *garment modeling* (*i.e.,* analyzing the garment-garment compatibility) based on the visual appearances and textual descriptions of fashion items [1, 3].

However, the CW generated by existing approaches may be unsuitable for individual users because of their distinct demographics, preferences, body shapes and consumption habits. For example, the appearance of an outfit could largely depend on whether it is suitable for the *user body shape* [4, 5]. Therefore, lacking the modeling of the user body shape may result in inappropriate outfits for the target user. Moreover, apart from the garment compatibility, whether an outfit is suitable also highly depends on the *user preference* [6]. As such, in addition to the traditional garment modeling, we argue the necessity of *user modeling* (*i.e.,* analyzing the user-garment compatibility) to evaluate the potential outfits for the automatic CW creation. Furthermore, pursuing the practical value, we propose the *personalized capsule wardrobe (PCW)* — a

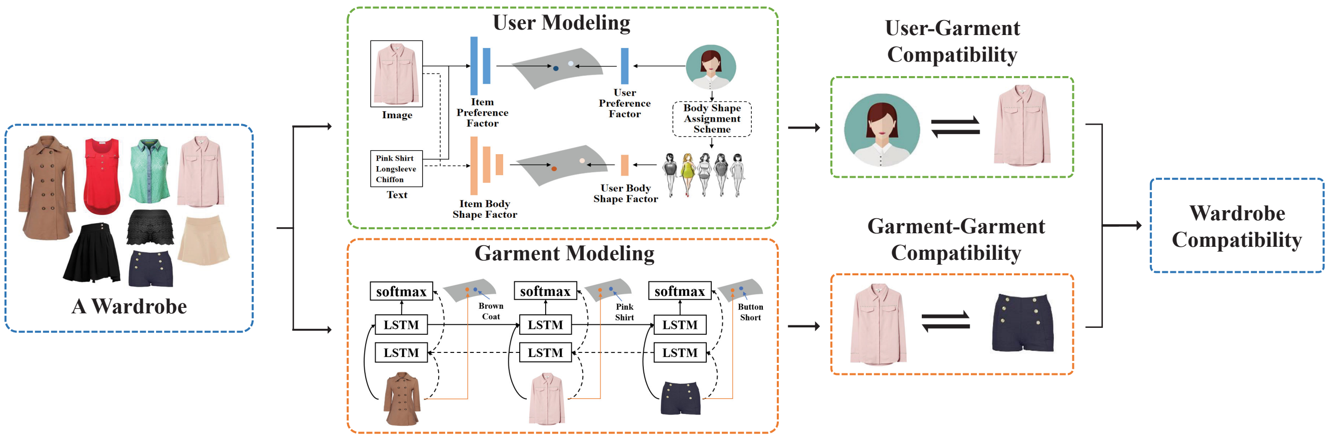

Figure 2: Schematic illustration of the scoring model, consisting of the user modeling and garment modeling.

collection of garments subject to creating both compatible and suitable outfits for the user.

Considering the existence of garments that have already been purchased by the user, to be cost-friendly, we formulate the automatic PCW creation task as: given the original wardrobe (*i.e.,* a set of purchased garments) of a user, adding or deleting garments according to both user-garment and garment-garment compatibilities. As illustrated in Figure 1, one purchased garment and four new garments are discarded and added to the original wardrobe, respectively, to make the resulted PCW not only presents the higher garment-garment compatibility but also caters to the user's preference and body shape. In fact, this task confronts three key challenges. 1) The PCW creation is a more complex combinatorial problem as compared to the conventional CW creation, where the complex user profile derived from the original wardrobe should be taken into account. Therefore, how to adaptively build the PCW for different individuals is the major challenge. 2) As both the outfit itself and the user profile (*e.g.,* the preferences and body shapes) determine the outfit compatibility for a given user, how to accurately evaluate the compatibility of the potential outfit from both user-garment and garment-garment perspectives, poses another challenge. And 3) most existing datasets only support either the user preference modeling or user body shape modeling. Accordingly, the lack of dataset that facilitates the comprehensive user modeling constitutes a crucial challenge.

To tackle the aforementioned challenges, we propose a combinatorial optimization-based **P**ersonalized **C**apsule **W**ardrobe creation framework with **D**ual **C**ompatibility modeling, named PCW-DC. The key novelty of the proposed framework lies in the introduction of a *scoring model* that can comprehensively evaluate the compatibility of potential outfits from both user-garment and garment-garment perspectives. In particular, as illustrated in Figure 2, the scoring model consists of two key components: user modeling and garment modeling. As for the user modeling, we adopt the two most relevant aspects of the user profile: the user preference and body shape, to measure the user-garment compatibility. To tackle the heterogeneity of user aspect and garment, we learn the user-garment compatibility in the latent matching space via a cross-modal projection. Different latent spaces are associated with different user aspects to highlight their difference for the user-garment compatibility[1]. Pertaining to the garment modeling, we adopt a bidirectional LSTM to measure the compatibility among garments, which is an efficient method to assess the outfit compatibility based on the visual appearances and textual descriptions of items. Finally, the wardrobe compatibility is estimated via the linear combination of the user modeling and garment modeling.

Our main contributions can be summarized in threefold:

- We present a new combinatorial optimization-based framework to cope with the personalized capsule wardrobe creation. To the best of our knowledge, we are the first to formulate the PCW creation task with the user original wardrobe as input.
- We develop a scoring model to evaluate the wardrobe compatibility in a user-adaptive manner, which jointly models the user-garment compatibility and the garment-garment compatibility.
- We construct a real-world dataset, bodyFashion, comprising 116, 532 user-item purchase records on Amazon involving 11,784 users and 75,695 fashion items. We have released the data, codes, and involved parameter settings to facilitate other researchers[2].

2 RELATED WORK

2.1 User Modeling

User Preference Modeling. User preference modeling is gaining increasing research interest for its applications ranging from the fashion domain [7, 8] to online social networks [9, 10]. In this research line, Matrix Factorization (MF) has become a popular and effective framework [11, 12], which aims to uncover the latent user/item factors that affect people's preference behavior. For example, Hu et al. [11] first associated different "confidence levels" to the positive and non-observed user feedback and then perform the factorization over the user-item rating matrix. Noticing that previous works just regarded the missing feedback as negative one and failed to directly optimize the model for ranking, Rendle et al. [13] proposed a generalized Bayesian Personalized Ranking

[1]Note that other user aspects, like the age and occupation, could be easily incorporated in the similar manner.
[2]https://dxresearch.wixsite.com/pcw-dc.

(BPR) framework, where the user-specific order of two items is exploited by the Bayesian analysis. Thereafter, due to its great success, several extension efforts on BPR have been put forward in fashion domain. For example, He et al. [7] introduced the Visual Bayesian Personalized Ranking (VBPR), where the latent visual factor is incorporated to model the user's preference on the visual appearance of fashion items. Meanwhile, Yu et al. [14] presented a dynamic collaborative filtering model with the BPR optimization criterion, where the user aesthetics are exploited. Differently, in this work, we further explore the the textual cues (*i.e.*, descriptions and categories) of items to comprehensively model the user preference.

Body Shape Modeling. In a sense, body shape plays an important role in fashion analysis, as people with different body shapes tend to go with different types of items. In fact, several pioneer research efforts have been dedicated to the user body shape modeling. For example, Sattar et al. [5] first leveraged the fashion photos of users to estimate their body shapes with a multi-photo body model. Despite its great success, the loose garments in fashion photos used in this work may hide the real body shapes of users and thus make the modeling results less accurate. Meanwhile, Hidayati et al. [4] designed a clustering-based body shape assignment scheme where the body measurements of celebrities are studied. One problem this work suffers from is that the body shapes of celebrities tend to be too perfect to represent that of ordinary people, making the proposed method less practical in the real world. Beyond the existing approaches, we introduce a novel body shape assignment scheme targeting the body shape modeling for ordinary people.

2.2 Garment Modeling

Due to its pivotal role in fashion analysis, recently, several efforts have been made to study the compatibility among fashion items. For example, the authors in [15] and [16] studied the Amazon co-purchase data to model the human sense regarding the relationships between fashion items. Nevertheless, the co-purchased relation could be a weak and noisy proxy for the garment compatibility measuring, as the items purchased together can be incompatible. Accordingly, Song et al. [17] collected the outfit dataset from Polyvore and based on that introduced a content-based neural framework for the compatibility modeling between fashion items. Meanwhile, Li et al. [18] and Chen et al. [19] studied the outfit compatibility modeling that involves multiple items with the dataset collected from online fashion websites. Besides, several axillary information, such as the item category [20, 21], aesthetic characteristics [22] and domain knowledge [23, 24], has been explored to promote the performance. Recently, to enhance the practical value, there is also a growing trend to make the compatibility more interpretable, where the attention mechanism [25, 26] and interpretable feature learning [27, 28] have been explored. Noticing that existing methods mainly focus on the supervised learning and may suffer from the unreliability of the negative example sampling, several efforts [1, 3] have been made to found the latent distribution of well-matched outfits with only the positive examples. For example, Han et al. [3] regarded each outfit as an ordered sequence and utilized a bidirectional LSTM to model the outfit compatibility. Despite the great progress in garment compatibility modeling, the user factor has remained largely untapped.

3 PCW-DC

This section details the proposed PCW-DC. We first formulate the research problem and then detail the two key components of the scoring model: user modeling and garment modeling, based on which we can perform the PCW creation.

3.1 Problem Formulation

In this work, to be cost-friendly, we focus on creating a PCW based on the user's original wardrobe (*i.e.,* the set of historical purchased fashion items). Let $\mathcal{I}_u = \{i^u_{ck} | \ c = 1, \cdots, C; k = 1, \cdots, N_c\}$ be the original wardrobe of the user u, comprising a set of fashion items from C categories (e.g., the top, bottom and outer), where N_c denotes the total number of items belonging to the category c. In addition, we have a set of items $\mathcal{I} = \{i_n\}_{n=1}^N$, and each item i_n is associated with a visual image and a textual description. Our task is to generate a new personalized capsule wardrobe $\widetilde{\mathcal{I}_u}$ for the user u based on $\mathcal{I}_u$ and $\mathcal{I}$ that provides the user both compatible and suitable outfits. In a sense, we should get rid of inappropriate items from $\mathcal{I}_u$ and add proper items from $\mathcal{I}$ to maximize the user-garment and garment-garment compatibilities of the wardrobe.

Essentially, we aim to propose a comprehensive wardrobe compatibility scoring model $S(\cdot)$, based on which we can perform the PCW creation. In particular, we define $S(\cdot)$ as follows,

$$S(\mathcal{I}^*) = \alpha U(\mathcal{I}^*|\Theta_U) + (1-\alpha)G(\mathcal{I}^*|\Theta_G), \tag{1}$$

where $\mathcal{I}^*$ represents a candidate wardrobe. U and G denote the compatibility modeling from the user-garment and garment-garment perspectives, respectively. α is a trade-off parameter to balance the evaluation score of each component. Θ_U and Θ_G refer to the to-be-learned model parameters of the user modeling and garment modeling, respectively.

3.2 User Modeling

To measure the user-garment compatibility, we particularly take into account the user preferences and body shapes due to the following reasons. 1) Different people may have different preferences on fashion items because of their different ages, occupations, cultural backgrounds and even locations. And 2) the user body shape is critical for people to choose fashion items, as people in different body shapes tend to go with different items. For example, plump people may prefer items with vertical stripes to make them look slimmer. Accordingly, we define the user-garment compatibility for a candidate wardrobe $\mathcal{I}^*$ as follows,

$$U(\mathcal{I}^*|\Theta_U) = \frac{1}{|\mathcal{I}^*|}\sum_{i\in\mathcal{I}^*}(x^p_{ui} + x^s_{ui}), \tag{2}$$

where x^p_{ui} is the preference of the user u for the item i, while x^s_{ui} is the body shape compatibility between the item i and the user u.

3.2.1 User Preference Modeling. Intuitively, it is reasonable to argue that different individuals may prefer different item appearances and categories. For example, some people may prefer the white top instead of a black one, while others prefer the skirt rather than the short. In fact, user preference modeling in fashion domain has been studied by recent work [7], whereby two latent spaces are introduced to measure the user's overall preference and visual preference for a given item, respectively. However, this method overlooks the value of the item's textual context in the user preference modeling. In fact, the textual description, including the item title and category metadata, can summarize the key semantic features of items, like the style, material and category, and hence deliver important cues of the user preferences. Therefore, in this work, to comprehensively model the user preferences, we formulate x_{ui}^{p} as follows,

$$x_{ui}^{p} = \boldsymbol{\gamma}_u^T \boldsymbol{\gamma}_i + \boldsymbol{\theta}_u^T (\mathbf{W}_p[\boldsymbol{f}_i, \boldsymbol{t}_i] + \boldsymbol{\beta}_p), \tag{3}$$

where $\boldsymbol{\gamma}_u \in \mathbb{R}^K$ and $\boldsymbol{\gamma}_i \in \mathbb{R}^K$ are latent factors of the user u and the item i, respectively. $\boldsymbol{\theta}_u \in \mathbb{R}^D$ is the latent content factor of the user u. $[\boldsymbol{f}_i, \boldsymbol{t}_i]$ refers to the concatenation of item visual feature $\boldsymbol{f}_i$ and textual feature $\boldsymbol{t}_i$. $\mathbf{W}_p$ and $\boldsymbol{\beta}_p$ are parameters of the nonlinear operation that maps the item features to the latent preference space. The first and second term of the equation encode the overall preference and content preference of the user u towards the item i, respectively.

For the optimization of the user preference modeling, we adopt the Bayesian Personalized Ranking (BPR) network, which has been proven to be an effective optimization framework for the pairwise preference ranking [17]. Based on BPR, we build the following training set $\mathcal{D}_s = \{(u, i, j)\}$, where $i \in \mathcal{I}_u$ and $j \in \mathcal{I} \setminus \mathcal{I}_u$. Each triplet (u, i, j) indicates that the user u prefers the item i to the item j. Then according to [29], we have the following objective function,

$$\arg\min_{\Theta_U} \sum_{(u,i,j)\in\mathcal{D}_s} -ln(\sigma(x_{ui}^{p} - x_{uj}^{p})). \tag{4}$$

3.2.2 User Body Shape Modeling. As aforementioned, people with different body shapes would go with different types of items. As such, we assume that there should be a latent space where the compatibility between body shapes and item contents can be well captured. We first obtain the body shape for each user based on our body shape assignment scheme, which will be detailed in Section 4.2. Due to the fact that each user can be assigned with only one body shape, we represent each user with an one-hot encoding $\boldsymbol{u}_s \in \mathbb{R}^Q$, where Q is the total number of possible body shapes. And then, we attempt to learn the item embedding towards the body shape compatibility modeling.

On the one hand, the matching knowledge between items and body shapes can be explicitly affected by the item appearance. We thus employ the multi-layer perception (MLP) to map the item content to the body shape matching space. In particular, the item embedding $\boldsymbol{i}_s \in \mathbb{R}^Q$, derived from its visual and textual features, can be designed as follows,

$$\boldsymbol{i}_s = \sigma(\boldsymbol{W}_s[\boldsymbol{f}_i, \boldsymbol{t}_i] + \boldsymbol{\beta}_s), \tag{5}$$

where $[\boldsymbol{f}_i, \boldsymbol{t}_i]$ is same as that in Eqn. (3). $\boldsymbol{W}_s$ and $\boldsymbol{\beta}_s$ are the parameters of the MLP. $\sigma(x) = \frac{1}{1+exp(-x)}$ is the nonlinear activation function.

On the other hand, the matching knowledge can be implicitly conveyed by the user's historical reviews on their purchased items, as users tend to purchase items that highlight their figure strength and hide the shortcomings. Accordingly, we define the item referenced embedding $\boldsymbol{i}_s^* \in \mathbb{R}^Q$ as follows,

$$\boldsymbol{i}_s^* = softmax(\sum_{u\in\mathcal{U}_i} \boldsymbol{u}_s), \tag{6}$$

Algorithm 1 Personalized capsule wardrobe creation algorithm

Input: User original wardrobe $\mathcal{I}_u = \{i_{ck}^u\}$;
Max and min number of item in categories N_{max} and N_{min}.
Initialize $\mathcal{I}^0 \leftarrow \mathcal{I}_u$; $break = 0$.
repeat
  if $\exists N_c \notin [N_{min}, N_{max}]$ **then**
    if $N_c > N_{max}$ **then**
      del = $\arg\max_{i_{ck}\in\mathcal{I}^{i-1}} S(\mathcal{I}^{i-1} \setminus i_{ck})$
      $\mathcal{I}^i \leftarrow \mathcal{I}^{i-1} \setminus$ del
    else
      add = $\arg\max_{i_c\in\mathcal{I}} S(\mathcal{I}^{i-1} \cup i_c)$
      $\mathcal{I}^i \leftarrow \mathcal{I}^{i-1} \cup$ add
    end if
  else if $\exists i \in \mathcal{I}^{i-1}\ s.t.\ S(\mathcal{I}^{i-1} \setminus i) - S(\mathcal{I}^{i-1}) > 0$ **then**
    $\mathcal{I}^i \leftarrow \mathcal{I}^{i-1} \setminus i$
  else
    $break = 1$
  end if
until $break == 1$
Output: User personalized capsule wardrobe $\widetilde{\mathcal{I}_u}$.

where $\mathcal{U}_i$ denotes the set of users who bought the item i. $softmax(x) = \frac{exp(x_i)}{\sum_{k=1}^{K} exp(x_k)}$ is a normalized exponential function. Ultimately, we argue that the matching knowledge obtained from item contents and the historical reviews should be consistent, that is, the item embedding $\boldsymbol{i}_s$ and item referenced embedding $\boldsymbol{i}_s^*$ should be close. Consequently, we reach the following objective function for the body shape modeling,

$$\arg\min_{\Theta_U} ||\boldsymbol{i}_s^* - \boldsymbol{i}_s||^2, \tag{7}$$

where $||\cdot||^2$ is the Euclidean distance. Based on the well-trained model, the body shape compatibility x_{ui}^s between the item i and the user u can be calculated as follows,

$$x_{ui}^s = \boldsymbol{u}_s^T \boldsymbol{i}_s. \tag{8}$$

3.3 Garment Modeling

The garment-garment compatibility is another key factor affecting the PCW creation. To facilitate users to compose proper outfits, it is natural to expect that the complementary fashion items (*e.g.,* the top, bottom and outer) in a PCW should share high compatibility and go well with each other. Towards this end, we define the garment-garment compatibility of one wardrobe $G(\mathcal{I}^*)$ as the average compatibility of the set of all potential outfits[3] O^* that can be generated from the wardrobe $\mathcal{I}^*$. Formally, we have,

$$G(\mathcal{I}^*|\Theta_G) = \frac{1}{|O^*|} \sum_{o_i\in O^*} cmp(o_i), \tag{9}$$

where o_i is the i-th outfit, and $cmp(\cdot)$ refers to the outfit compatibility. To measure $cmp(o_i)$, we adopt the compatibility indicator in [3], where each outfit is treated as a sequence of items and each

[3]Here we only consider the following three mainstream outfit patterns: top plus bottom, top plus bottom plus outer, and one-piece plus outer.

Figure 3: An example of the user's Amazon purchase history.

item is regarded as a time step input of a bidirectional LSTM. In particular, $cmp(o_i)$ can be computed as follows,

$$\begin{cases} cmp(o_i) = E_f(o_i;\theta_f) + E_b(o_i;\theta_b), \\ E_f(o_i;\theta_f) = -\frac{1}{N}\sum_{t=1}^{N} logPr(o_{i,t+1}|o_{i,1},...,o_{i,t};\theta_f), \\ E_b(o_i;\theta_b) = -\frac{1}{N}\sum_{t=N-1}^{0} logPr(o_{i,t}|o_{i,N},...,o_{i,t+1};\theta_b), \end{cases} \tag{10}$$

where $Pr(\cdot|\cdot)$ stands for the conditional probability. $E_f(o_i;\theta_f)$ and $E_b(o_i;\theta_b)$ refer to the forward and backward probability that the outfit o_i would be a compatible one.

3.4 PCW Creation

Based on the user modeling and garment modeling that enable us to comprehensively measure the overall compatibility of a given wardrobe, we can now proceed to present our framework for the automatic PCW creation. In particular, we cast the PCW creation as a combinatorial optimization problem and propose a heuristic PCW creation method, which is summarized in Algorithm 1. The underlying philosophy is to delete items from the original wardrobe that can degrade the overall wardrobe compatibility and add candidate items that can improve the compatibility.

Considering the practical situation, we first set the maximum and minimum numbers of items for each category in a wardrobe. For simplicity, here we uniformly set that as N_{max} and N_{min} for all categories. At each iteration, we first check whether the number of items of each category (*i.e.,* N_c) in a wardrobe has reached the pre-assigned maximum and minimum number (i.e., N_{max} and N_{min}). If $N_c < N_{min}$ ($N_c > N_{max}$), the algorithm would add (delete) one item of the category c that maximizes (maximally hurts) the overall wardrobe compatibility according to our wardrobe compatibility scoring model $S(\cdot)$. Otherwise, if $N_{min} \leq N_c \leq N_{max}$, the algorithm would check if there is an existing (unsuitable or redundant) item deteriorating the compatibility and removing which would boost the wardrobe compatibility. If yes, the item will be deleted. In the light of this, this operation will adaptively adjust the number of items of each category, making the final PCW meeting the user's preferences over different item categories.

4 DATASET

In this section, we first introduce our bodyFashion dataset and then particularly present a body shape assignment scheme towards the ground truth construction for the user modeling.

4.1 Dataset Construction

In reality, it is intractable to collect a comprehensive dataset that can fully support the personalized capsule wardrobe creation. Therefore, in this work, we employ two datasets for the user modeling and garment modeling, respectively.

Table 1: Women garment sizes and their corresponding body measurements (in inch) provided by Amazon.

Size	Bust	Waist	Hip
S	34	26	36.5
M	36	28	38.5
L	38.5	30.5	41

As for the user modeling, although McAuley et al. [15] has introduced a public large-scale Amazon dataset for personalized fashion recommendation tasks, it fails to incorporate the user body shape data, making the dataset unsuitable for our comprehensive PCW creation. Fortunately, we noticed that the user purchase history, especially the size of purchased fashion items, as shown in Figure 3, conveys important cues of the user's body shape. Accordingly, we constructed our own dataset, named bodyFashion, by collecting the user purchase histories from Amazon. In particular, we first collected a set of popular fashion items from Amazon, and based on the item comments we tracked a set of Amazon users. For each user, we crawled his/her latest (at most 100) historical purchase records and only retained the fashion related ones. In order to guarantee the dataset quality, we screened out users with less than 6 fashion purchase records, and then obtained 116,532 user-item records involving 11,784 users and 75,695 fashion items. Each item comprises its image, title and category metadata. Both purchase sizes and ratings are available for each user-item record. Pertaining to the garment modeling, we adopt the public Polyvore dataset [3], comprising 21,889 outfits with 164,379 fashion items.

4.2 Body Shape Assignment Scheme

Different from previous studies that represent user body shapes with complex body features, we resort to the three most essential body measurements[4]: bust girth, waist girth and hip girth. These measurements can be easily derived from the average garment size of one's purchase history. Specifically, due to the different nature of these three body measurements, we adopt the size of tops to capture the bust girth of the user, and that of bottoms to determine one's waist girth and hip girth. Table 1 exhibits the correspondence between the women garment sizes and body

[4]https://www.iso.org/standard/65246.html

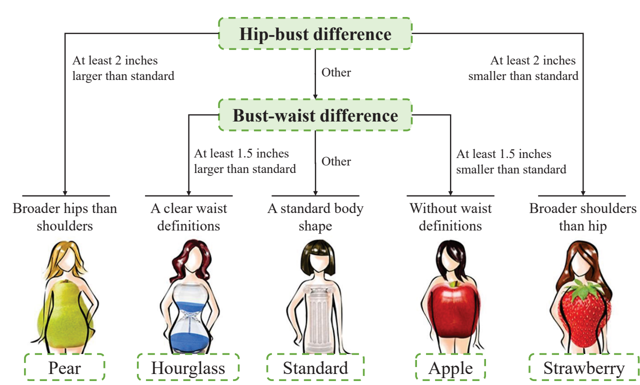

Figure 4: Body shape assignment scheme.

Table 2: Performance comparison among different methods.

	SR	AI	AD
POP	22.78%	0.45%	-3.18%
RAND	22.78%	0.51%	-3.01%
ISBM	27.22%	0.70%	-2.49%
CWC [1]	33.33%	0.93%	-2.09%
DCF-A [14]	36.67%	0.97%	-1.85%
ExDCF-A	73.33%	2.67%	-0.44%
PCW-DC	**80.56%**	**4.46%**	**-0.36%**

measurements provided by Amazon. Moreover, we adopt the hip-bust and bust-waist differences as the key indicators to distinguish body shapes. The underlying philosophy is that the hip-bust difference can directly reflect the relationship between one's upper and lower bodies, while the bust-waist difference can intuitively capture one's waist characteristic.

To guide the body shape assignment, we first seek the standard hip-bust and bust-waist differences as the reference. As listed in Table 1, the hip-bust and bust-waist differences are invariant across different garment sizes, which thus propels us to set the standard hip-bust and bust-waist differences as 2.5 inches and 8 inches, respectively. By comparing with the standard reference, we define five common body shapes with their intuitive appearances: 1) pear shape; 2) hourglass shape; 3) rectangle shape; 4) apple shape; and 5) strawberry shape. The detailed derivation rules are illustrated in Figure 4. Here we take the pear shape as an example, while the others can be derived in the similar manner. We define that if the user's hip-bust difference is larger than the standard one with at least 2 inches (*i.e.,* with a plump lower body), then the user's body shape belongs to the pear shape.

5 EXPERIMENTS

To evaluate the proposed method, we conducted extensive experiments on bodyFashion by answering the following questions:

- Does the PCW-DC outperform the state-of-the-art baselines?
- How do the user modeling and garment modeling affect the PCW creation?
- Does the body shape modeling help people with suitable dress?

5.1 Experiment Settings

Here, we will introduce the visual and textual representation extraction for fashion items, and the parameter settings for experiments.
Visual Representation. Convolutional networks have shown great success in various computer vision tasks, ranging from the image classification [30] to the item retrieval [31]. In particular, we chose the pre-trained ConvNet provided by [32], which consists of 16 convolutional layers followed by 3 fully connected layers. Accordingly, we fed each fashion item image to the ConvNet, and obtained a 4,096-D vector as the visual representation.
Textual Representation. To extract meaningful cues from the textual description, we focused on the descriptive words towards the attribute characterization of the item. Taking the 1,000 descriptive attributes defined in Deepfashion [33] as the reference, we obtained 393 descriptive words from our bodyFashion dataset. Considering the lack of colour attribute, we complemented the vocabulary with extra 127 colour-related descriptive words, leading to the final vocabulary consisting of 520 words. Regarding the category metadata, we re-summarized 9 categories from the raw data: *outer*, *short tops*, *long tops*, *short bottoms*, *long bottoms*, *skirts*, *suits*, *dresses* and *shoes*, where extremely fine-grained categories, like pants and leggings, are merged. Ultimately, based on the bag-of-word scheme, we represented the textural modality of each item as a 529-D vector.
Parameter Setting. To control the total number of items in the wardrobe, the maximum and minimum numbers of items in each category (*i.e.,* N_{max} and N_{min}) are set to 5 and 3, respectively. We adopted the grid search strategy to determine the optimal value for the regularization parameter (*i.e.,* α) in the range of $[0, 1]$ with a step of 0.1. The numbers of hidden units for the user preference modeling, user body shape modeling and garment modeling are set as 224 (*i.e.,* $K = 64$ and $D = 160$ in Eqn.(3)), 512 (*i.e.,* 256 for both visual and textual representation) and 512, respectively. As aforementioned, the garment modeling is trained on the dataset in [3], while the user modeling is learned by our bodyFashion. We then randomly sampled 180 users from bodyFashion as the testing set, on which the model performance is reported.

5.2 On Model Comparison (RQ1)

To evaluate the proposed PCW creation scheme, we chose the following baselines.

- **POP.** We added/deleted items according to its "popularity", which is defined as the number of users that have purchased the item.
- **RAND.** We randomly added/deleted items to creat PCWs.
- **Item Similarity Based Method (ISBM).** We added/deleted an item according to its average visual similarity to each item in the user original wardrobe, measured by the inner product of their visual features extracted by ConvNet.
- **Capsule Wardrobe Creation (CWC).** Focusing on the outfit compatibility and versatility, this method [1] creates the capsule wardrobes using a topic model over the item attributes. Here, we directly adopted the descriptive words as the item attributes.
- **Dynamic Collaborative Filtering (DCF-A).** This user preference modeling approach [14] incorporates the aesthetic features to boost the performance for item recommendation. We adapted it as one baseline by dropping the time factor that is unavailable in our context and adopting the state-of-the-art aesthetic features [34].
- **ExDCF-A.** We extended DCF-A by introducing our garment modeling score to its final recommendation score of each garment for a user, where the linear fusion with equal weights is adopted.

As it is intractable to obtain the exact ground-truth, we introduced the following three metrics to softly evaluate the PCW creation: successful rate (SR), average improvement (AI) and average diminishment (AD), whose definitions are given as follows,

$$\begin{cases} SR = \frac{|\mathcal{A}|}{|\Omega|}, \\ AI = \frac{1}{|\Omega|} \sum_{\widetilde{\mathcal{I}_u} \in \mathcal{A}} S(\widetilde{\mathcal{I}_u}) - S(\mathcal{I}_u), \\ AD = \frac{1}{|\Omega|} \sum_{\widetilde{\mathcal{I}_u} \in \Omega \setminus \mathcal{A}} S(\widetilde{\mathcal{I}_u}) - S(\mathcal{I}_u), \end{cases} \tag{11}$$

where Ω is the set of testing users. $\mathcal{A} = \{\widetilde{\mathcal{I}_u} \mid S(\widetilde{\mathcal{I}_u}) - S(\mathcal{I}_u) > 0\} \subseteq \Omega$ is the set of successfully created PCWs, which are defined as those whose compatibilities, assessed by our scoring model, get improved as compared to the original ones.

Figure 5: An example of PCW creation. First line: Results of PCW created from user original wardrobe by the proposed PCW-DC method and its variants. Second line: Possible outfits provided by the created personalized capsule wardrobe.

Table 2 shows the PCW creation results of different methods. As can be seen, PCW-DC achieves the best performance with respect to all metrics, demonstrating the superiority of the proposed PCW-DC for PCW creation. In addition, methods that overlook both the garments and user profiles (i.e., POP and RAND) perform worst, while the method that considers the naive garment interaction (i.e., ISBM) promotes the performance slightly. Comparing with the above methods, DCF-A and CWC with more advanced compatibility modeling achieve the better performance. However, due to the limited modeling perspective of each method, DCF-A and CWC still suffer from the inferior performance than PCW-DC. Moreover, ExDCF-A outperforms DCF-A, which demonstrates the necessity of incorporating the user modeling to fulfill the PCW creation. To gain a better understanding, a successful example created by PCW-DC is shown in Figure 5. As can be seen, the somewhat monotonous original wardrobe has turned to be a versatile wardrobe by adding garments that share the similar style with the original wardrobe and deleting those hurt the overall wardrobe compatibility. Moreover, we found that most of the potential outfits of the final PCW are compatible, which meets the initial motivation of our PCW creation.

5.3 On Ablation Study (RQ2)

To verify the necessity of both the user modeling and garment modeling in the PCW creation, we further conducted the ablation study. In particular, we compared our framework with its two derivatives: the PCW creation taking the user modeling only (PCW-U) and that admitting the garment modeling only (PCW-G). It is worth noting that PCW-U and PCW-G can be effortlessly derived by setting $\alpha = 1$ and $\alpha = 0$ in Eqn.(1), respectively.

Table 3 illustrates the performance of the ablation study. As can be seen, PCW-DC consistently achieves the best performance over different metrics, which verifies the importance of both the user modeling and garment modeling for the personalized wardrobe creation. In addition, PCW-G outperforms PCW-U, suggesting that the garment modeling contributes more towards the overall wardrobe compatibility modeling. One possible explanation is that the garment compatibility is the main factor during the dressing as compared to the personalized factors. To gain the deep insight, we further checked the wardrobe creation results and illustrated one example in Figure 5. As we can see, PCW-G retains limited garments of the original wardrobe but incorporates many external garments in different styles from the original ones. On the contrary, PCW-U follows the user's personal taste and admits garments in similar styles, but leads to several incompatible garment pairs. Beyond that, the PCW created by PCW-DC seems to be more reasonable as it not only meets the personal preferences of the user but also maintains the high garment-garment compatibility for the wardrobe.

Table 3: Performance of different methods.

	SR	AI	AD
PCW-U	50.56%	1.63%	-1.01%
PCW-G	58.33%	3.33%	-1.64%
PCW-DC	**80.56%**	**4.46%**	**-0.36%**

5.4 On User Body Shape Modeling (RQ3)

Here we attempted to examine the performance of the user **B**ody **S**hape **M**odeling (BSM) in our framework, which is also a major contribution of this work. In particular, we first focused on checking the rationality of the body shape assignment scheme and then assessed our body shape modeling with several baselines.

5.4.1 Body Shape Assignment Scheme. To evaluate the body shape assignment scheme, we compared it with the method in [4], which learns the body shapes by clustering celebrities' body measurements. To adapt it for our context, we extracted the following features for each user in our dataset, including the 1) bust girth; 2) waist girth; 3) hip girth; 4) ratio between the bust and hip girths; 5) ratio between the waist and hip girths; 6) ratio between the bust and waist girths; 7) difference between the bust and hip girths; 8) difference between the waist and hip girths, and 9) difference between the bust and waist girths. According to [4], we conducted the affinity propagation clustering [35] and finally obtained six clusters. We visualized the body shape assignment results of both methods with the help of t-SNE [36] in Figure 6. For clear illustration, we drew two user bodies based on their body measurements with bodybuilder[5]. As can be seen, the method in [4] clusters most users into a single body shape, while checking the ground truth we found that body shapes of user1 and user2 are totally distinct. This may be attributed to the

[5]http://www.bodyvisualizer.com.

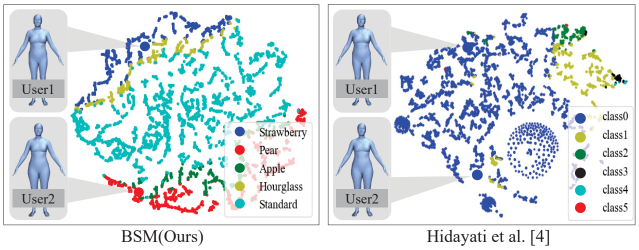

Figure 6: Visualization of body shape classification results.

dense distribution of the body measurements of ordinary people, making it inappropriate to distinguish different body shapes with the clustering method. Meanwhile, the better performance achieved by our scheme suggests that it is advisable to explicitly model the user body shape with their body measurements.

5.4.2 BSM Assessment. The user body shape modeling in our PCW-DC is designed to predict the compatibility of the garment for a given body shape. To assess the effectiveness of our BSM, we adopted the accuracy of body shape prediction as the evaluation metric, where the predicted most suitable body shape and the ground truth body shape for a given garment are compared. Due to the limited related work, we chose the following baselines:

- **Probability Model (PM).** Following the work in [4], we employed the class-conditional-probability density [37] to model the global body shape matching knowledge.
- **BSM-V.** This method is derived from our BSM model, which takes only the garment visual appearance into account to learn the garment compatibility for the body shape.
- **BSM-T.** Similar to BSM-V, we derived this method by utilizing only the garment textural descriptions.

Table 4 summarizes the performance of different approaches with different category configurations. From this table, we have the following observations. 1) Our method outperforms PM, verifying the advantages of introducing the latent matching space learning with neural networks for the body shape modeling compared to the probability model. 2) BSM-V performs better than BSM-T. This may be due to the fact that the visual appearance conveys more accurate cues regarding the body shape compatibility than the textural description. And 3) our model achieves better performance than both BSM-V and BSM-T. This suggests that although textural information may deliver less significant cues than visual images for the body shape prediction, it can still boost the performance with descriptive words, like “high-waist” and “tight”.

To obtain more deep insights, we further investigated the most suitable and unsuitable garments for each body shape. Without loss of generality, we only considered the dresses for illustration. In particular, we fed dresses to the BSM network and obtained their latent embeddings, based on which we can derive their compatibilities for each body shape. Figure 7 shows three suitable and unsuitable dresses for different body shapes. It can be seen that both suitable and unsuitable garments for each body shape share certain latent garment features. For example, people in the strawberry shape are more suitable to dresses with a broad or deep neckline, while those in the pear shape would be better to wear dresses with umbrella-shaped hemlines instead of the tight ones. These observations do give plausible suggestions to help people dress properly.

Table 4: Performance of different approaches in the user body shape modeling.

	Outer	Top	Bottom	Suit	Total
PM [4]	**0.50**	0.38	0.45	0.40	0.43
BSM-V	0.48	0.45	0.45	0.61	0.50
BSM-T	0.33	0.34	0.40	0.53	0.40
BSM	0.49	**0.46**	**0.48**	**0.63**	**0.52**

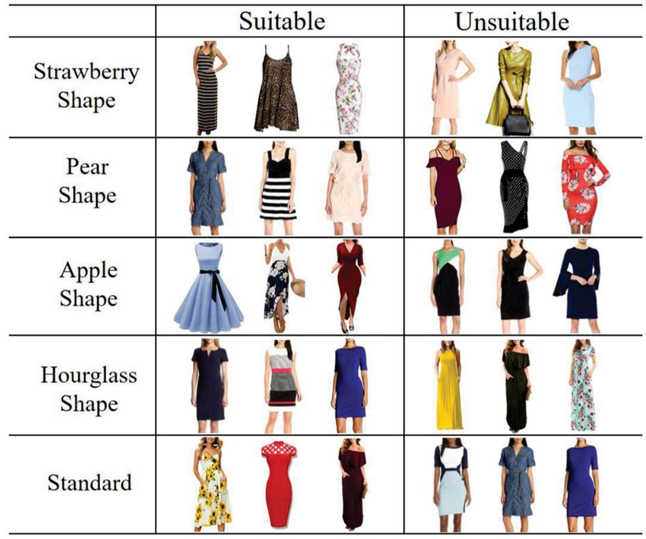

Figure 7: Suitable and unsuitable garments for different body shapes.

6 CONCLUSION AND FUTURE WORK

In this work, we studied the problem of the PCW creation based on the user’s original wardrobe. In particular, we presented a combinatorial optimization-based personalized capsule wardrobe creation framework, named PCW-DC, with dual compatibility modeling: the user modeling and garment modeling, where the user modeling explores the user preference and user body shape. In addition, we collected a large-scale dataset bodyFashion from Amazon, comprising 116,532 user-item records with 11,784 users and 75,695 fashion items. Extensive experiments have been conducted over the bodyFashion dataset, and the results demonstrate the necessity of considering both the user-garment and garment-garment compatibilities in PCW creation. Interestingly, we found that the garment-garment compatibility plays the more important role in PCW creation than the user-garment compatibility. Currently, the user modeling and garment modeling in our model are learned separately. In the future, we plan to devise an end-to-end unified scheme to boost the model performance.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China, No.: 61772310, No.: 61702300, No.: 61702302, No.: 61802231, and No. U1836216; the Project of Thousand Youth Talents 2016; the Shandong Provincial Natural Science and Foundation, No.: ZR2019JQ23, No.: ZR2019QF001; the Future Talents Research Funds of Shandong University, No.: 2018WLJH63.

REFERENCES

[1] Wei-Lin Hsiao and Kristen Grauman. Creating capsule wardrobes from fashion images. In *Conference on Computer Vision and Pattern Recognition*, pages 7161–7170, 2018.

[2] Ranjitha Kumar and Kristen Vaccaro. An experimentation engine for data-driven fashion systems. In *AAAI Spring Symposium Series*, 2017.

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

[4] Shintami Chusnul Hidayati, Cheng-Chun Hsu, Yu-Ting Chang, Kai-Lung Hua, Jianlong Fu, and Wen-Huang Cheng. What dress fits me best?: Fashion recommendation on the clothing style for personal body shape. In *ACM Multimedia Conference on Multimedia Conference*, pages 438–446, 2018.

[5] Hosnieh Sattar, Gerard Pons-Moll, and Mario Fritz. Fashion is taking shape: Understanding clothing preference based on body shape from online sources. In *IEEE Winter Conference on Applications of Computer Vision*, pages 968–977, 2019.

[6] Wang-Cheng Kang, Chen Fang, Zhaowen Wang, and Julian McAuley. Visually-aware fashion recommendation and design with generative image models. In *IEEE International Conference on Data Mining*, pages 207–216, 2017.

[7] Ruining He and Julian McAuley. VBPR: visual bayesian personalized ranking from implicit feedback. In *AAAI Conference on Artificial Intelligence*, pages 144–150, 2016.

[8] Ruining He, Chunbin Lin, Jianguo Wang, and Julian McAuley. Sherlock: Sparse hierarchical embeddings for visually-aware one-class collaborative filtering. In *International Joint Conference on Artificial Intelligence*, pages 3740–3746, 2016.

[9] Liqiang Nie, Xuemeng Song, and Tat-Seng Chua. *Learning from Multiple Social Networks*. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers, 2016.

[10] Markus Brill, Edith Elkind, Ulle Endriss, and Umberto Grandi. Pairwise diffusion of preference rankings in social networks. In *International Joint Conference on Artificial Intelligence*, pages 130–136, 2016.

[11] Yifan Hu, Yehuda Koren, and Chris Volinsky. Collaborative filtering for implicit feedback datasets. In *International Conference on Data Mining*, pages 263–272, 2008.

[12] Rong Pan, Yunhong Zhou, Bin Cao, Nathan Nan Liu, Rajan M. Lukose, Martin Scholz, and Qiang Yang. One-class collaborative filtering. In *International Conference on Data Mining*, pages 502–511, 2008.

[13] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. BPR: bayesian personalized ranking from implicit feedback. In *Conference on Uncertainty in Artificial Intelligence*, pages 452–461, 2009.

[14] Wenhui Yu, Huidi Zhang, Xiangnan He, Xu Chen, Li Xiong, and Zheng Qin. Aesthetic-based clothing recommendation. In *Proceedings of World Wide Web Conference on World Wide Web*, pages 649–658, 2018.

[15] Julian J. McAuley, Christopher Targett, Qinfeng Shi, and Anton van den Hengel. Image-based recommendations on styles and substitutes. In *ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 43–52, 2015.

[16] Andreas Veit, Balazs Kovacs, Sean Bell, Julian McAuley, Kavita Bala, and Serge J. Belongie. Learning visual clothing style with heterogeneous dyadic co-occurrences. In *IEEE International Conference on Computer Vision*, pages 4642–4650, 2015.

[17] Xuemeng Song, Fuli Feng, Jinhuan Liu, Zekun Li, Liqiang Nie, and Jun Ma. Neurostylist: Neural compatibility modeling for clothing matching. In *ACM Multimedia Conference on Multimedia Conference*, pages 753–761, 2017.

[18] Yuncheng Li, Liangliang Cao, Jiang Zhu, and Jiebo Luo. Mining fashion outfit composition using an end-to-end deep learning approach on set data. *IEEE Trans. Multimedia*, 19(8):1946–1955, 2017.

[19] Long Chen and Yuhang He. Dress fashionably: Learn fashion collocation with deep mixed-category metric learning. In *AAAI Conference on Artificial Intelligence*, pages 2103–2110, 2018.

[20] Xun Yang, Yunshan Ma, Lizi Liao, Meng Wang, and Tat-Seng Chua. Transnfcm: Translation-based neural fashion compatibility modeling. In *AAAI Conference on Artificial Intelligence*, 2019.

[21] Mariya I. Vasileva, Bryan A. Plummer, Krishna Dusad, Shreya Rajpal, Ranjitha Kumar, and David A. Forsyth. Learning type-aware embeddings for fashion compatibility. In *European Conference on Computer Vision*, pages 405–421, 2018.

[22] Yujie Lin, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Jun Ma, and Maarten de Rijke. Improving outfit recommendation with co-supervision of fashion generation. In *International World Wide Web Conference*, 2019.

[23] Xuemeng Song, Fuli Feng, Xianjing Han, Xin Yang, Wei Liu, and Liqiang Nie. Neural compatibility modeling with attentive knowledge distillation. In *International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 5–14, 2018.

[24] Zhengzhong Zhou, Xiu Di, Wei Zhou, and Liqing Zhang. Fashion sensitive clothing recommendation using hierarchical collocation model. In *ACM Multimedia Conference on Multimedia Conference*, pages 1119–1127, 2018.

[25] Wenguan Wang, Yuanlu Xu, Jianbing Shen, and Song-Chun Zhu. Attentive fashion grammar network for fashion landmark detection and clothing category classification. In *Conference on Computer Vision and Pattern Recognition*, pages 4271–4280, 2018.

[26] Jingyuan Liu and Hong Lu. Deep fashion analysis with feature map upsampling and landmark-driven attention. In *European Conference on Computer Vision Workshops*, pages 30–36, 2018.

[27] Zunlei Feng, Zhenyun Yu, Yezhou Yang, Yongcheng Jing, Junxiao Jiang, and Mingli Song. Interpretable partitioned embedding for customized multi-item fashion outfit composition. In *International Conference on Multimedia Retrieval*, pages 143–151, 2018.

[28] Ziaeefard M, Camacaro J, and Bessega C. Hierarchical feature map characterization in fashion interpretation. In *Conference on Computer and Robot Vision*, pages 88–94, 2018.

[29] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. BPR: bayesian personalized ranking from implicit feedback. In *Uncertainty in Artificial Intelligence*, pages 452–461, 2009.

[30] Xishan Zhang, Jia Jia, Ke Gao, Yongdong Zhang, Dongming Zhang, Jintao Li, and Qi Tian. Trip outfits advisor: Location-oriented clothing recommendation. *IEEE Trans. Multimedia*, 19(11):2533–2544, 2017.

[31] Shuhui Jiang, Yue Wu, and Yun Fu. Deep bi-directional cross-triplet embedding for cross-domain clothing retrieval. In *ACM Multimedia Conference on Multimedia Conference*, pages 52–56, 2016.

[32] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.

[33] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1096–1104, 2016.

[34] Hossein Talebi and Peyman Milanfar. NIMA: neural image assessment. *IEEE Trans. Image Processing*, 27(8):3998–4011, 2018.

[35] Brendan J Frey and Delbert Dueck. Clustering by passing messages between data points. *science*, 315(5814):972–976, 2007.

[36] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[37] Yingying Deng, Fan Tang, Weiming Dong, Hanxing Yao, and Bao-Gang Hu. Style-oriented representative paintings selection. In *Special Interest Group on Computer Graphics*, pages 12:1–12:2, 2017.