“Hi, Magic Closet, Tell Me What to Wear!”

Si Liu[1,2], Jiashi Feng[1], Zheng Song[1], Tianzhu Zhang[3]
Hanqing Lu[2], Changsheng Xu[2], Shuicheng Yan[1]
[1] ECE Department, National University of Singapore
[2] NLPR, Institute of Automation, Chinese Academy of Science
[3] ADSC, Advanced Digital Sciences Center
{dcslius,a0066331,zheng.s,eleyans}@nus.edu.sg,{tzzhang,luhq,csxu}@nlpr.ia.ac.cn

ABSTRACT

In this paper, we aim at a practical system, magic closet, for automatic occasion-oriented clothing recommendation. Given a user-input occasion, *e.g.*, wedding, shopping or dating, magic closet intelligently suggests the most suitable clothing from the user's own clothing photo album, or automatically pairs the user-specified reference clothing (upper-body or lower-body) with the most suitable one from online shops.

Two key criteria are explicitly considered for the magic closet system. One criterion is to *wear properly*, *e.g.*, compared to suit pants, it is more decent to wear a cocktail dress for a banquet occasion. The other criterion is to *wear aesthetically*, *e.g.*, a red T-shirt matches better white pants than green pants. To narrow the semantic gap between the low-level features of clothing and the high-level occasion categories, we adopt middle-level clothing attributes (*e.g.*, clothing category, color, pattern) as a bridge. More specifically, the clothing attributes are treated as latent variables in our proposed latent Support Vector Machine (SVM) based recommendation model. The *wearing properly* criterion is described in the model through a feature-occasion potential and an attribute-occasion potential, while the *wearing aesthetically* criterion is expressed by an attribute-attribute potential. To learn a generalize-well model and comprehensively evaluate it, we collect a large clothing What-to-Wear (WoW) dataset, and thoroughly annotate the whole dataset with 7 multi-value clothing attributes and 10 occasion categories via Amazon Mechanic Turk. Extensive experiments on the WoW dataset demonstrate the effectiveness of the magic closet system for both occasion-oriented clothing recommendation and pairing.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval models; I.2.6 [**Learning**]: Knowledge acquisition

General Terms

Algorithms, Experimentation, Performance

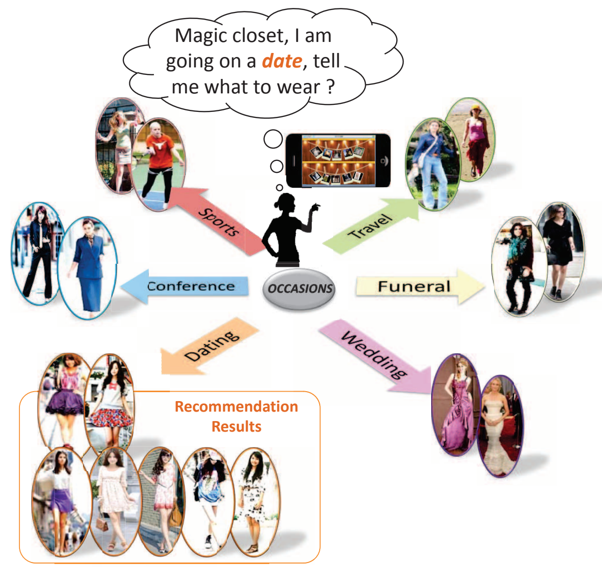


Figure 1: Scenario illustration of magic closet. A user specifies an occasion, and the most suitable clothing are recommended from the user's mobile photo album. For going on a date, the recommended clothing are sweet, have bright color and skirts are preferred. For better view, please refer to the original color pdf file.

Keywords

Clothing recommendation, Clothing pairing, Latent SVM

1. INTRODUCTION

Studies on clothing are receiving increasing interest these days mainly due to the huge market related with clothing. In China, the potential market is expected to break 20 billion US dollars[1]. Such huge market prospect greatly motivates clothing relevant research. However, related research literature is still quite limited and most of them only focus on clothing segmentation [14] and recognition [13].

In this paper, we explore a brand-new topic of occasion-oriented clothing recommendation. When people select clothing, occasion is the most important factor to consider[2]. Selecting suitable clothing for a certain occasion can reflect a person's courtesy, especially in some special occasions. For example, for an official interview, wearing a mini skirt is

[1] http://www.bizjournals.com/prnewswire/press_releases/2012/04/02/SP80325
[2] http://www.azcentral.com/style/fashion/articles/2008/10/02/20081002whattowear.html

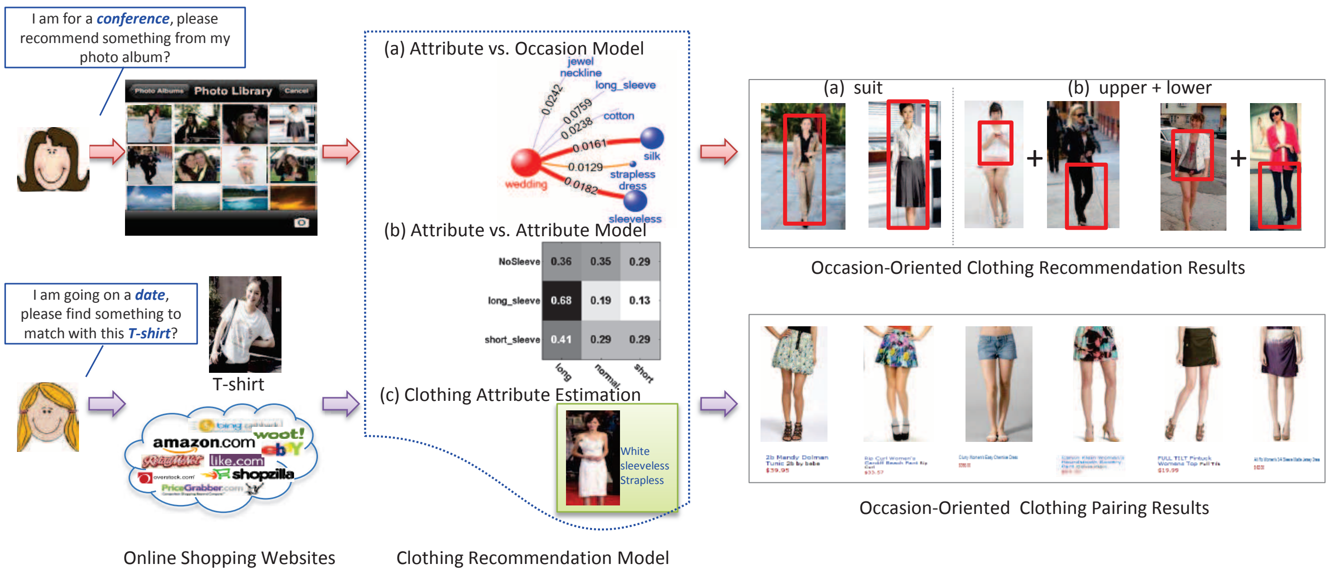


Figure 2: Two typical clothing recommendation scenarios for magic closet. (Top panel) Clothing suggestion: given an occasion, the most suitable clothing combinations from one's photo album are suggested. The suggested clothing can be one piece of suits, or two pieces of clothing. (Bottom panel) Clothing pairing: given an occasion and a reference clothing, the clothing most suitable for the occasion and most matched with the reference clothing is recommended from online shopping websites.

impolite; and it is even offensive to wear red coat when attending funeral. As shown in Figure 1, different occasions generally possess their own distinctive dressing styles. In this work, we aim to explore such clothing-occasion and clothing-clothing matching rules. And based on those mined rules we construct automatic clothing recommendation system, called magic closet, which can recommend the most suitable clothing for a user specified occasion.

Magic closet mainly targets at two clothing recommendation scenarios. The first scenario is *clothing suggestion*. As shown in the top panel of Figure 2, a user specifies an occasion and the most suitable suits or two separate clothing items (such as one T-shirt and one pair of trousers) are suggested from the user's own photo album. In terms of this function, magic closet can be implemented as a mobile application. The second scenario is *clothing pairing*. As shown in the bottom panel of Figure 2, a user inputs an occasion and one reference clothing item (such as a T-shirt the user wants to pair), the most matched clothing from the online shopping website is returned (such as a skirt). The returned clothing should aesthetically pair with the reference clothing well and also together be suitable for the specified occasion. From this perspective, magic closet system can serve as a plug-in application in any online shopping websites for shopping recommendation.

We mainly consider two key principles when designing magic closet. One is *wearing properly*. Wearing properly means putting on some decent, suitable clothing, which conforms to the *dress codes*[3] and common sense. The other is *wearing aesthetically*. There are some atheistic rules which need to be followed when one pairs the upper-body coat and lower-body trousers. For example, it looks weird to wear a red coat and a green pants together.

In the model learning process, to narrow the semantic gap between the low-level visual features of clothing and the high-level occasion categories information, we propose to utilize middle-level clothing attributes. Attributes have shown to be quite useful in many computer vision tasks [3, 2, 10, 16]. Here we define 7 multi-value clothing attributes, including the category attribute (*e.g.*, "jeans", "skirts") and detail attributes, describing certain properties of clothing (*e.g.*, color, pattern).

We propose to learn the clothing recommendation model through a unified latent Support Vector Machine (SVM) framework. The model integrates four potentials: 1) visual features vs. attribute, 2) visual features vs. occasion, 3) attributes vs. occasion, and 4) attribute vs. attribute. Here the first three concern about clothing-occasion matching and the last one describes the clothing-clothing matching. Embedding these matching rules into the latent SVM model explicitly ensures that the recommended clothing satisfies the requirement of *wearing properly* and *wearing aesthetically* simultaneously.

Few existing works target at the clothing recommendation task. Some online websites [4] and pioneering works [20] can support the service of recommending the most suitable clothing for an occasion. However, their works are mainly based on the commonsense reasoning technology. As for the pairing, Daniel Cohen-Or *et al.* [7] studied the problem on how different colors can be paired together to produce a harmonious feeling. Similar idea was also explored in [6]. However, few works investigate the problem of clothing pairing. Iwata *et al.* [15] construct a fashion coordinates system, but the important occasion factors are not considered. To our best knowledge, magic closet is the first one to investigate the task of occasion-oriented clothing recommendation and clothing pairing, which mines the matching rules among more semantic attributes from real images automatically.

Main contributions of this work can be summarized as:

[3] Dress codes are written and unwritten rules with regards to clothing.

[4] http://www.dresscodeguide.com/

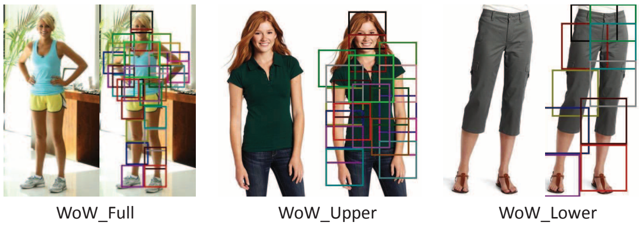


Figure 3: Detected upper/lower-body parts from WoW_Full, WoW_Upper and WoW_Lower subsets.

- It is the first time to explore two interesting practical problems: 1) how to automatically suggest the most suitable clothing for an occasions from one's photo album, and 2) how to retrieve clothing from online shopping websites to pair with a user specified clothing and also suitable for a specified occasion.
- We construct a large dataset, called What-to-Wear (WoW), containing 24,417 clothing. The collected WoW dataset contains complete attribute and occasion annotations. It is the largest fully annotated clothing dataset to date.
- We adopt a latent SVM based framework to learn the occasion-oriented clothing recommendation model and simultaneously mine the clothing matching rules, which has shown to be quite effective through comprehensive evaluations.

2. THE WOW DATASET

There are several existing clothing datasets, but none of them are suitable for evaluating the occasion-oriented clothing recommendation tasks. For example, in the dataset collected by Yang *et al.* [13], human in the images are not clear enough for detailed clothing attribute estimation. Another dataset constructed by Bourde *et al.* [5] only contains 5 attributes related with clothing. Here, we construct a new dataset specific for the occasion-oriented clothing recommendation, named as "What-to-Wear" (WoW) dataset.

2.1 Clothing Image Collection

The WoW dataset is collected from both online shopping websites and popular photo sharing websites (*e.g.*, Flickr.com), by using queries such as "street shot", "shopping girls", etc, similar with previous online dataset collection methods [18]. And those images in the photo sharing websites which are favored by lots of users are downloaded to guarantee their aestheticness. The well-trained human upper-body and lower-body detectors [25] are both applied on all these images and only the high-confidence detection outputs are retained. Thus, the possible false alarms from background are removed. Several exemplar human body detection results are shown in Figure 3.

The collected WoW dataset is split into three subsets based on the detection results. The first subset WoW_Full includes 9,469 images containing visible full-body. The second subset, denoted as WoW_Upper, contains 8,421 images with only upper-body, such as T-shirts, Fashion hoodies. And the 6,527 images containing lower-body clothing, such as Jeans and Skirts, are put into WoW_Lower.

According to different source of data, WoW_Upper is further split into two subsets, one is WoW_Upper_DP where all the images are Daily Photos (DP), which are crawled from popular photo sharing websites, while the other one is WoW_Upper_OS, the photos of which are crawled from Online Shopping (OS) websites. Similarly, both WoW_Lower and WoW_Full subsets are further split into DP and OS subsets in the same way.

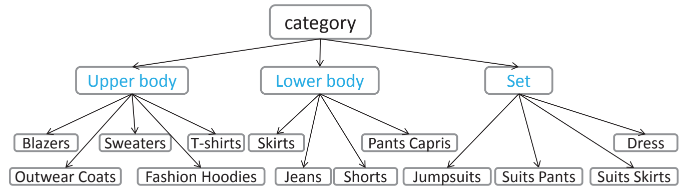


Figure 4: Clothing category attributes. All the attributes are organized in a tree structure and only the leaf nodes are considered in this work.

	Attribute Name	Attibute Values			Attribute Name	Attibute Values	
GLOBAL	**Color**	Red Black Green Blue	Orange White Brown Purple	Yellow Gray Multi-color	**Material**	Cotton Silk Denim	Chiffon Woolen Leather
	Pattern	Vertical Drawing	Plaid Plain	Horizontal Floral print			
UPPER	**Collar**	Strapless Jewel	V-shape Round	One-shoulder Shirt collar	**Sleeve**	Long Sleeveless	Shot
LOWER	**Length**	Long	Median	Short			

Figure 5: Detail attributes definition.

2.2 Attribute and Occasion Annotation

As aforementioned, two types of attributes are used in this paper. The first type is the category attribute, which is treated as one multi-value attribute. And its values are shown in Figure 4. The category values are defined manually based on a comprehensive study of many online shopping websites, and organized in a tree structure. Only the leaf nodes in the tree correspond to category values, and in total we have 13 categories. The second type of attributes are detail attributes, which are manually selected according to previous research [19]. These attributes describe different properties of the clothing and can be further summarized into three classes, *i.e.*, global, upper-body and lower-body attributes as shown in Figure 5. In this work, we consider 10 common occasions, which are summarized from fashion websites and shown in Figure 6.

We use Amazon Mechanical Turk [5] to annotate clothing attributes and occasion categories of the whole WoW dataset. Five annotators are assigned for each annotation task. A label is considered as groundtruth if at least more than half of the annotators agree on it. Note that each clothing may belong to different occasions, so we use multiple options for occasion annotation. Table 1 shows the distribution of each attribute. Some exemplar images of different occasions are shown Figure 6. The distribution on different occasions are shown in Figure 7.

3. HUMAN BODY PARSING AND FEATURE EXTRACTION

3.1 Human Body Parsing

Previous research on human attribute analysis tends to first align human parts [5, 21] due to the large pose varia-

[5]https://www.mturk.com/mturk/

Figure 6: Exemplar images in the What-to-Wear (WoW) dataset. For each occasion, four exemplar images are shown.

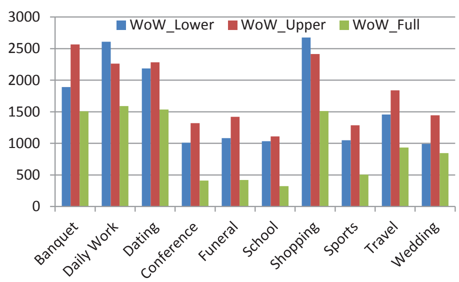


Figure 7: The image number distribution of different occasions in the WoW dataset.

tions and background noises in daily photos. The part-based detection [25, 4] is proven to be able to assist in matching human parts and thus facilitate the appearance modeling for attribute classification. In this paper, we follow this line and design a part-based detection scheme for the clothing recommendation task. We use the annotated key points in human photos provided in [4] and train one human upper-body detector and one human lower-body detector [25]. Figure 3 shows several human detection results, which also demonstrate the necessity of human parts alignment.

3.2 Feature Extraction

Following [21, 5] , we extract 5 types of features from the 20 upper-body parts and 10 lower-body parts. The features include Histograms of Oriented Gradient (HOG) [8], Local Binary Pattern (LBP) [1], color moment, color histogram and skin descriptor [5]. More specifically, each human part is first partitioned into several smaller, spatially evenly distributed regular blocks. Features extracted from all the blocks are finally concatenated into a $28,770$ dimensional feature vector to represent a human part. The block based features can roughly preserve relative position information inside each human part.

Table 1: Number of labels for each upper/lower-body attribute. Here "all" counts all the labeled clothing.

UPPER	color	pattern	material	collar	sleeve	all
	11971	14192	10082	14743	17615	17890
LOWER	color	pattern	material	length	all	
	12410	14329	9108	14826	15996	

4. CLOTHING RECOMMENDATION VIA LATENT SVM BASED MODEL

As aforementioned, we aim at an occasion-oriented clothing recommendation system, magic closet. It takes photos of user's own clothing and a specified occasion as the inputs and automatically suggests the most suitable clothing for the occasion from the provided photos or retrieves the clothing from online shops which pair with a reference clothing well. In particular, magic closet is based on a latent SVM model, which is learned from the visual features of the given clothing photos with full annotations on occasions and attributes. The model describes the matching rules among visual features, clothing attributes (treated as latent variables) and occasions. Based on this model, we can obtain occasion-oriented rank of the clothing photos and aesthetic scores of the candidate clothing pairs. In this way, magic closet provides suitable clothing recommendations. In this section, we first formally describe the clothing recommendation model constructed based on latent SVM. Then we introduce the model learning and inference process for clothing recommendation. We use the same latent SVM as the work in [24], but here we would like to emphasize that each potential function in the adopted latent SVM is defined specifically for the clothing recommendation task.

4.1 The Latent SVM Based Model

Given a set of clothing photos with full annotations of occasions and attributes, we aim to construct a recommendation model to capture the underlying rules for the clothing and occasion matching. In this work, instead of directly predicting the occasion based on the low-level visual features from the whole body, we use several aforementioned clothing attributes to represent the middle-level information. Meanwhile, we also aim to explore the matching rules of upper and lower-body clothing in the recommendation model.

A training clothing image is denoted as a tuple $(\mathbf{x}, \mathbf{a}_u, \mathbf{a}_l, \mathbf{o})$. Here $\mathbf{x}$ is the visual features from the whole body clothing, which is formed by directly concatenating the upper-body clothing feature $\mathbf{x}_u$ and lower-body clothing feature $\mathbf{x}_l$, namely $\mathbf{x} = [\mathbf{x}_u; \mathbf{x}_l]$. The occasion categories of the clothing are represented by $\mathbf{o} \subset \mathcal{O}$, where $\mathcal{O}$ denotes the finite occasion category set. Note that each clothing may have multiple occasion category labels. The attributes of the upper-body clothing are denoted by a vector $\mathbf{a}_u = [a_1^u, \ldots, a_{K_u}^u]^T$, where K_u is the number of attributes considered for the upper-body clothing. Each attribute describes certain characteristic of the upper-body clothing, *e.g.*, color, collar. Similarly, the attributes of the lower-body clothing are denoted as a vector $\mathbf{a}_l = [a_1^l, \ldots, a_{K_l}^l]^T$. All the attributes considered in this work are listed in Figure 4 and Figure 5. We denote the attribute set for the upper-body and lower-body as $\mathcal{A}^u$ and $\mathcal{A}^l$ respectively. Note that each attribute is multi-valued and we represent each attribute by a multi-dimensional binary value vector in the model learning process. For example, the attribute "color" has 11 different values, *e.g.*, red, orange, etc. Then we represent the "color" attribute by an 11-dimensional vector with each element corresponding to one specific type of color.

Given N training examples $\{(\mathbf{x}^{(n)}, \mathbf{a}_u^{(n)}, \mathbf{a}_l^{(n)}, \mathbf{o}^{(n)})\}_{n=1}^N$, our goal is to learn a model that can be used to recommend the most suitable clothing for a given occasion label $o \in \mathcal{O}$, which considers clothing-occasion and clothing-clothing matching simultaneously. Formally speaking, we are interested in learning a scoring function $f_\mathbf{w} : \mathcal{X} \times \mathcal{O} \rightarrow \mathbb{R}$, over an image $\mathbf{x}$ and a user specified occasion label o, where $\mathbf{w}$ are the parameters of $f_\mathbf{w}$. Here $\mathcal{X}$ denotes the clothing image space. During testing, $f_\mathbf{w}$ can be used to suggest the most suitable clothing $\mathbf{x}^*$ from $\mathcal{X}^t$ (candidate clothing repository) for the given occasion o as $\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathcal{X}^t} f_\mathbf{w}(\mathbf{x}, o)$. While for the clothing pairing recommendation, given specified lower-body clothing $\mathbf{x}_l$, $f_\mathbf{w}$ can select the most suitable upper-body clothing $\mathbf{x}_u^*$ as $\mathbf{x}_u^* = \arg\max_{\mathbf{x}_u\in\mathcal{X}_u^t} f_\mathbf{w}([\mathbf{x}_u; \mathbf{x}_l], o)$, where $\mathcal{X}_u^t$ denotes candidate upper-body clothing repository. For the lower-body clothing pairing, it works similarly.

In this work, we propose to introduce the middle-level clothing attributes to more accurately explore and describe the clothing matching rules. Here, the attributes are treated as latent variables. And inspired by the latent Support Vector Machine (SVM) [11], we assume $f_\mathbf{w}(\mathbf{x}, o)$ takes the following form: $f_\mathbf{w}(\mathbf{x}, o) = \max_{\mathbf{a}_u,\mathbf{a}_l} \mathbf{w}^T\Phi(\mathbf{x}, \mathbf{a}_u, \mathbf{a}_l, o)$, where $\Phi(\mathbf{x}, \mathbf{a}_u, \mathbf{a}_l, o)$ is a feature vector depending on the images $\mathbf{x}$, the attributes $\mathbf{a}_u$,$\mathbf{a}_l$ and occasion label o. And $\mathbf{w}$ is the model parameter.

To explore the matching rules of upper and lower-body clothing, we need to consider certain dependencies between attributes of upper-body and those of lower-body (a_j^u, a_k^l) in the model $f_\mathbf{w}(\mathbf{x}, o)$. For example, the attributes a_j^u and a_k^l may correspond to the "color" attribute of upper and lower-body clothing, and might take the value of "red" and "green" respectively. Then their values are highly exclusive, since generally the "red" does not match "green" and they should not be recommended as a suitable pair. Here we assume that only the values of the same attribute are correlated (*e.g.*, upper "red" and lower "green"). This assumption is reasonable since which color of the upper-body clothing is generally independent of the material of the lower-body clothing. We use a directed graph $G = (\mathcal{V}, \mathcal{E})$, which we call the attribute relation graph, to represent these dependency relations of the attribute pairs of upper-body and lower-body clothing. A vertex $j \in \mathcal{V}$ in the graph G corresponds to one attribute with certain value, and an edge $(j, k) \in \mathcal{E}$ indicates that attributes a_j^u and a_k^l have certain dependency (highly correlated or exclusive).

We define the recommendation function as follows:

$$\begin{aligned} & \mathbf{w}^T\Phi(\mathbf{x}, \mathbf{a}_u, \mathbf{a}_l, o) \\ = \ & \mathbf{w}_o^T\phi(\mathbf{x}, o) + \sum_{j\in\mathcal{A}^u\cup\mathcal{A}^l} \mathbf{w}_{a_j}^T\varphi(\mathbf{x}, a_j) \qquad (1) \\ & + \sum_{j\in\mathcal{A}^u\cup\mathcal{A}^l} \mathbf{w}_{o,a_j}^T\omega(a_j, o) + \sum_{(j,k)\in\mathcal{E}} \mathbf{w}_{j,k}^T\psi(a_j^u, a_k^l). \end{aligned}$$

In this model, the parameter vector $\mathbf{w}$ is the concatenation of the parameters in all the factors. The model presented in Eqn. (1) simultaneously considers the dependencies among visual features, attributes and occasions. In particular, its first term predicts occasion from visual features; the second term describes the relationship between visual features and attributes; the third term captures the relationship between attributes and occasion; and the last term expresses the dependencies between the attributes of upper and lower-body clothing. Instead of predicting the occasion from visual features or attributes directly, we mine much richer matching rules among them explicitly. The impacts of different relationships on the matching score in Eqn. (1) are automatically determined in the learning process, therefore, the four relationships are not treated equally. The details of functions $\phi(\cdot)$, $\varphi(\cdot)$, $\omega(\cdot)$ and $\psi(\cdot)$ and the parameters $\mathbf{w}_o$, $\mathbf{w}_{a_j}$, $\mathbf{w}_{o,a_j}$ and $\mathbf{w}_{j,k}$ are described in Section 4.2.

In the model learning process, we learn the model parameter $\mathbf{w}$ of the discriminative function $f_\mathbf{w}(\cdot)$ that tends to assign the highest score to the most suitable clothing for a given occasion o and considers the upper and lower-body clothing matching simultaneously. Though the groundtruth attributes of the training samples are provided, we treat them as latent variables in the model learning process. The reasons are two-fold: 1) the groundtruth attributes of the testing samples are not provided; 2) according to [23], treating the attributes of training samples as latent variables can yield more discriminate scoring function and thus better performance. Note that the information of training samples attributes are well utilized in designing potential functions.

In this work, we adopt the latent SVM formulation to learn the model as follows [11]:

$$\begin{aligned} & \min_{\mathbf{w},\xi} \beta\|\mathbf{w}\|^2 + \sum_{n=1}^N \xi^{(n)} \\ & \text{s.t.} \ \max_{\mathbf{a}_u,\mathbf{a}_l} \mathbf{w}^T\Phi(\mathbf{x}^{(n)}, \mathbf{a}_u, \mathbf{a}_l, \mathbf{o}^{(n)}) - \max_{\mathbf{a}_u,\mathbf{a}_l} \mathbf{w}^T\Phi(\mathbf{x}^{(n)}, \mathbf{a}_u, \mathbf{a}_l, o) \\ & \qquad \geq \Delta(o, \mathbf{o}^{(n)}) - \xi^{(n)}, \forall n, \forall o \in \mathcal{O}, \qquad (2) \end{aligned}$$

where β is the trade-off parameter controlling the amount of regularization, and $\xi^{(n)}$ is the slack variable for the n-th training sample to handle the soft margin. Such an objective function requires that the score of clothing for a suitable occasion should be much higher than for a non-suitable occasion. $\Delta(o, \mathbf{o}^{(n)})$ is a loss function defined as:

$$\Delta_{0/1}(\mathbf{o}^{(n)}, o) = \begin{cases} 1 & \text{if } o \notin \mathbf{o}^{(n)} \\ 0 & \text{otherwise} \end{cases}$$

In Eqn. (2) we aim to learn a discriminative occasion-wise scoring function on each pair of clothing (more specifically, on their features and inferred attributes) such that the scoring function can rank clothing correctly by maximizing the score difference between suitable ones and unsuitable ones for the interest occasion.

4.2 Potential Function Design for Clothing Recommendation

Here, we describe the details of the potential functions used in Eqn. (1) for clothing recommendation.

4.2.1 Feature vs. Occasion Potential $\mathbf{w}_o^T\phi(\mathbf{x}, o)$

This potential is a standard linear model for occasion prediction, which is only based on low-level visual features and does not consider clothing attributes. Here $\phi(\mathbf{x}, o)$ represents a certain mapping of the feature vector $\mathbf{x}$ extracted from the clothing and the mapping result depends on the occasion label o. In this work, rather than keeping $\phi(\mathbf{x}, o)$ as a high dimensional vector as in traditional implementation [22], we follow the strategy in [24] and simply represent $\phi(\mathbf{x}, o)$ as the score of the pre-trained multi-class linear SVM. In the SVM training process, we ignore the attributes annotated in the training samples and train a multi-class occasion classifier from $\{\mathbf{x}^{(n)}, o^{(n)}\}_{n=1}^{N}$. Then we define the mapping function $\phi(\cdot)$ such that the mapped feature $\phi(\mathbf{x}, o) \in \mathbb{R}^{|\mathcal{O}|}$ has only one non-zero element whose value is equal to the SVM score of the occasion specified by o. And the dimension of the mapped feature $\phi(\mathbf{x}, o)$ is the same as the number of occasions. In this model, the parameter $\mathbf{w}_o$ re-weights the output scores from pre-trained multi-class SVM to guarantee more accurate occasion predictions.

4.2.2 Feature vs. Attribute Potential $\mathbf{w}_{a_j}^T \varphi(\mathbf{x}, a_j)$

This potential is a standard linear model trained to predict the value of the j-th attribute given a visual feature vector $\mathbf{x}$. Similar to the potential function in feature vs. occasion model, here the potential function $\varphi(\mathbf{x}, a_j)$ also represents the confidence score from a standard multi-class linear SVM. Since there are several attributes used to describe the clothing and each attribute admits multiple values (see Figure 4 and Figure 5), we need to train multiple multi-class SVM classifiers. Each multi-class SVM corresponds to a specific attribute (*e.g.*, the j-th attribute) and is learned from $\{(\mathbf{x}^{(n)}, a_j^{(n)})\}_{n=1}^{N}$ with only considering the j-th attribute annotation. For each attribute, the dimension of the output feature from potential function $\varphi(\mathbf{x}, a_j)$ is equal to the number of different values for the attribute. For example, suppose the j-th attribute is "color", and there are 11 different values for the "color" attribute, then the dimension of the mapped feature $\varphi(\mathbf{x}, a_j)$ is 11. And only the a_j-th element of the vector $\varphi(\mathbf{x}, a_j)$ takes non-zero value, which is equal to the score assigned from multi-class SVM.

4.2.3 Attribute vs. Occasion Potential $\mathbf{w}_{o,a_j}^T \omega(a_j, o)$

This potential represents a standard linear model for occasion recognition based on only attributes. It integrates the matching rules between the attributes and occasions in its parameter $\mathbf{w}_{o,a_j}$. The value of $\omega(a_j, o)$ is equal to the co-occurrence statistics between the j-th attribute and the occasion o, which is estimated from the training samples $\{a_j^{(n)}, o^{(n)}\}_{n=1}^{N}$. The parameter $\mathbf{w}_{o,a_j}$ is also a re-weighting vector for predicting the possibility of the j-th attribute having value a_j to appear in the occasion o.

4.2.4 Attribute vs. Attribute Potential $\mathbf{w}_{j,k}^T \psi(a_j^u, a_k^l)$

This potential is used to describe the matching rules between the attributes a_j^u and a_k^l, from upper-body clothing and lower-body clothing respectively. Here $\psi(a_j^u, a_k^l)$ is a sparse vector with dimension $|\mathcal{A}_j^u| \times |\mathcal{A}_k^l|$, where $|\mathcal{A}_j^u|$ denotes the number of different values for the j-th upper-body clothing attribute and $|\mathcal{A}_k^l|$ denotes for the lower-body clothing. The value of the element in the vector $\psi(a_j^u, a_k^l)$ indexed by (a_j^u, a_k^l) is equal to the co-occurrence statistics of a_j^u and a_k^l, which approximates the probability of the clothing pairing that the j-th attribute of upper-body clothing takes value a_j^u while the k-th attribute of lower-body clothing takes value a_k^l. Note that for this model, we only consider the co-occurrence of the attributes linked in the attribute relation graph as discussed in Section 4.1.

The parameter $\mathbf{w}_{j,k}$ is a template to capture the matching degree of the considered attributes. For example, for the attribute "color", if the upper-body is red, then $\mathbf{w}_{j,k}$ may encourage the lower-body to be black by assigning their co-occurrence a higher weight while prevents it from being green by lowering their co-occurrence weight. After the model learning, the parameter $\mathbf{w}_{j,k}$ can be seen as incorporating the mined matching rules between the upper-body and lower-body clothing.

4.3 Optimization and Inference for Clothing Recommendation

4.3.1 Optimization for Model Learning

In this work, we adopt a non-convex cutting plane method proposed by [9] to solve the optimization problem in Eqn. (2) due to its ease of use. First, it is easy to show that Eqn. (2) is equivalent to $\min_{\mathbf{w}} L(\mathbf{w}) = \beta\|\mathbf{w}\|^2 + \sum_{n=1}^{N} R^n(\mathbf{w})$ where $R^n(\mathbf{w})$ is a hinge loss function defined as:

$$
\begin{aligned}
R^n(\mathbf{w}) \quad = \quad & \max_{o} \left(\Delta(o, o^{(n)}) + \max_{\mathbf{a}_u, \mathbf{a}_l} \mathbf{w}^T \Phi(\mathbf{x}^{(n)}, \mathbf{a}_u, \mathbf{a}_l, o) \right) \\
& - \max_{\mathbf{a}_u, \mathbf{a}_l} \mathbf{w}^T \Phi(\mathbf{x}^{(n)}, \mathbf{a}_u, \mathbf{a}_l, o^{(n)}). \qquad (3)
\end{aligned}
$$

The non-convex cutting plane method in [9] aims to iteratively build an increasingly accurate piecewise quadratic approximation of $L(\mathbf{w})$ based on its sub-gradient $\partial_{\mathbf{w}} L(\mathbf{w})$. Let us define:

$$
\begin{aligned}
\{\mathbf{a}_u^{(n)*}, \mathbf{a}_l^{(n)*}\} \quad &= \quad \arg\max_{\mathbf{a}_u, \mathbf{a}_l} \mathbf{w}^T \Phi(\mathbf{x}^{(n)}, \mathbf{a}_u, \mathbf{a}_l, o), \forall n, \forall o \in \mathcal{O}, \\
\{\mathbf{a}_u^{(n)}, \mathbf{a}_l^{(n)}\} \quad &= \quad \arg\max_{\mathbf{a}_u, \mathbf{a}_l} \mathbf{w}^T \Phi(\mathbf{x}^{(n)}, \mathbf{a}_u, \mathbf{a}_l, o^{(n)}), \forall n
\end{aligned}
$$

and

$$
o^{(n)*} = \arg\max_{o} \left(\Delta(o, o^{(n)}) + \mathbf{w}^T \Phi(\mathbf{x}^{(n)}, \mathbf{a}_u^{(n)}, \mathbf{a}_l^{(n)}, o) \right).
$$

It is easy to show a sub-gradient $\partial_{\mathbf{w}} L(\mathbf{w})$ can be calculated as follows:

$$
\begin{aligned}
\partial_{\mathbf{w}} L(\mathbf{w}) \quad = \quad & 2\beta \cdot \mathbf{w} + \sum_{n=1}^{N} \Phi \left(\mathbf{x}^{(n)}, \mathbf{a}_u^{(n)*}, \mathbf{a}_l^{(n)*}, o^{(n)*} \right). \\
& - \sum_{n=1}^{N} \Phi \left(\mathbf{x}^{(n)}, \mathbf{a}_u^{(n)}, \mathbf{a}_l^{(n)}, o^{(n)} \right). \qquad (4)
\end{aligned}
$$

Given the sub-gradient $\partial_{\mathbf{w}} L(\mathbf{w})$ computed by Eqn. (4), we can minimize $L(\mathbf{w})$ using the cutting plane method in [9].

4.3.2 Inference for Clothing Recommendation

After learning the model, we can use it to score any image-occasion pair $(\mathbf{x}, o)$. The score is inferred as $f_{\mathbf{w}}(\mathbf{x}, o) = \max_{\mathbf{a}_u, \mathbf{a}_l} \mathbf{w}^T \Phi(\mathbf{x}, \mathbf{a}_u, \mathbf{a}_l, o)$. Thus after specifying the occasion o, we can obtain a rank of the clothing from the user's clothing photo album. In particular, given the parameter model $\mathbf{w}$, we need to solve the following inference problem during recommendation:

$$
\{\mathbf{a}_u^*, \mathbf{a}_l^*\} = \arg\max_{\mathbf{a}_u, \mathbf{a}_l} \mathbf{w}^T \Phi(\mathbf{x}, \mathbf{a}_u, \mathbf{a}_l, o),
$$

which can be solved by linear programming since the attributes form a tree structure [23]. And then the clothing obtaining the highest score will be suggested, namely

$$
\mathbf{x}^* = \arg\max_{\mathbf{x}} \left\{ \max_{\mathbf{a}_u, \mathbf{a}_l} \mathbf{w}^T \Phi(\mathbf{x}, \mathbf{a}_u, \mathbf{a}_l, o) \right\}. \qquad (5)
$$

Usage	Image Source		Subset Name
Traing	WoW_Full_OS	WoW_Full_DP_1	WoW_Full
Exp1:Repository		WoW_Full_DP_2	
Exp2: Query	WoW_Upper_DP	WoW_Upper_OS	WoW_Upper
	⇩	⇧	
Exp2: Repository	WoW_Lower_OS	WoW_Lower_DP	WoW_Lower

Figure 8: Split of the dataset for different experiments. Arrows indicate query vs. repository relationship.

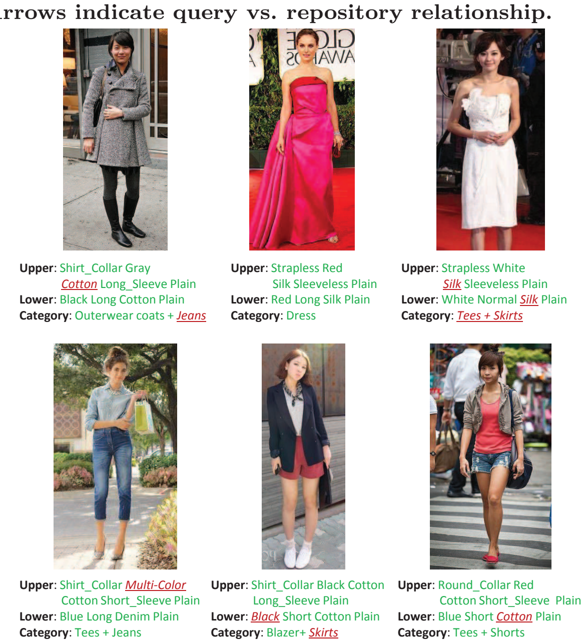


Figure 9: The estimated latent variables (clothing attributes) of the testing sample clothing. Incorrect estimations are highlighted.

Similarly, for the clothing pairing recommendation, given a specified upper-body clothing $\mathbf{x}_u$ and the occasion o, the most suitable lower-body clothing $\mathbf{x}_l^*$ is paired as:

$$\mathbf{x}_l^* = \arg\max_{\mathbf{x}_l} \left\{ \max_{\mathbf{a}_u, \mathbf{a}_l} \mathbf{w}^T \Phi\left([\mathbf{x}_u; \mathbf{x}_l], \mathbf{a}_u, \mathbf{a}_l, o\right) \right\}. \quad (6)$$

The upper-body clothing recommendation for a given lower-body clothing is conducted in the similar way.

5. EXPERIMENTS

5.1 Model Analysis

The WoW dataset is split as shown in Figure 8 for model learning and evaluation. Here we use the WoW_Full_OS and WoW_Full_DP_1 subsets to train the model and the WoW_Full_DP_2 subset for testing.

Attribute prediction: We first evaluate the performance of the latent SVM model in attribute prediction. The prediction is performed through jointly maximizing the feature vs. attribute and attribute vs. attribute potentials in Eqn. (1). Here since we do not concern the occasion, the attribute vs. occasion potential is discarded. Figure 9 shows several exemplar results of attribute prediction. From the figure, it can be seen that the latent SVM model accurately predicts most of the attributes, such as the collar, pattern and sleeve. Though some predictions are not correct, for these cases the predicted attribute values are quite similar to their groundtruth. For example, the material attribute of the

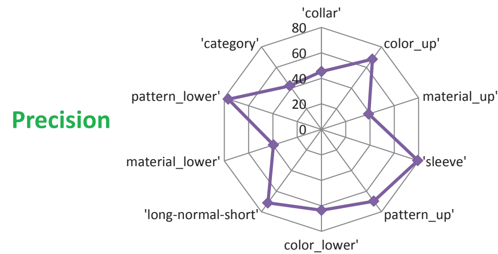


Figure 10: The precision of each attribute (%).

third clothing in the first row is predicted as "silk". This is acceptable as its visual feature is quite similar to silk. And the category attribute of this clothing is inferred as "T-shirt" (upper-body) and "skirts" (lower-body). If we treat the upper and lower-body of this clothing separately, the predicted attribute value can be considered as correct. The precision of each attribute is shown in Figure 10.

Visualization of the mined attribute-attribute matching rules: Here we show the mined matching rules among attributes in Figure 11 by visualizing the attribute vs. attribute potential parameter $\mathbf{w}_{j,k}$ in Eqn. (1). The model parameter vector is reshaped as a matrix. From the figure, we can see that the mined rules generally conform with common sense. For example, the upper "outerwear coats" matches well with lower "jeans", while does not match "shorts"; the upper clothing pattern "letter" matches lower clothing 'plain" well and does not match "letter" and "plaid".

Visualization of occasion-attribute matching rules: Similarly, in Figure 12, we visualize the attribute vs. occasion potential in the learned model to show the matching rules between occasion (wedding, funeral) and attributes (category, materials, sleeve length, color, etc.). It can be seen that many meaningful rules (consistent with common sense) are accurately captured by the model. For example, the "wedding" occasion matches "dress" best; "funeral" matches "black_upper" and "black_lower" well. These results demonstrate the effectiveness of the proposed model in matching rules mining.

5.2 Occasion-Oriented Clothing Suggestion

In this subsection, we present the experimental evaluation of the magic closet system for occasion-oriented clothing suggestion. In this scenario, given an occasion specified by a user, the task is to find the clothing which is most suitable for the occasion from the user's clothing photo album.

5.2.1 Experimental Setting

Though in practical system, all the clothing photos are from the same user, here in order to comprehensively evaluate the magic closet system for suggesting clothing with different attributes, we simulate the suggestion scenario on WoW_Full_DP dataset, which contains 6,661 images from multiple users. We evenly split the WoW_Full_DP subset into two groups as shown in Figure 8. The first half WoW_Full_DP_1 together with WoW_Full_OS (contain 2,808 images) are used for training the latent SVM based model embedded in magic closet. And the second half WoW_Full_DP_2 is used as repository for testing. Given an occasion, the clothing from the WoW_Full_DP_2 which maximizes the score function in Eqn. (5) is suggested by magic closet.

We compare the proposed magic closet with two linear SVM based models. The first baseline is a feature-occasion multi-class linear SVM which predicts occasion from visual features directly without considering attributes. After train-

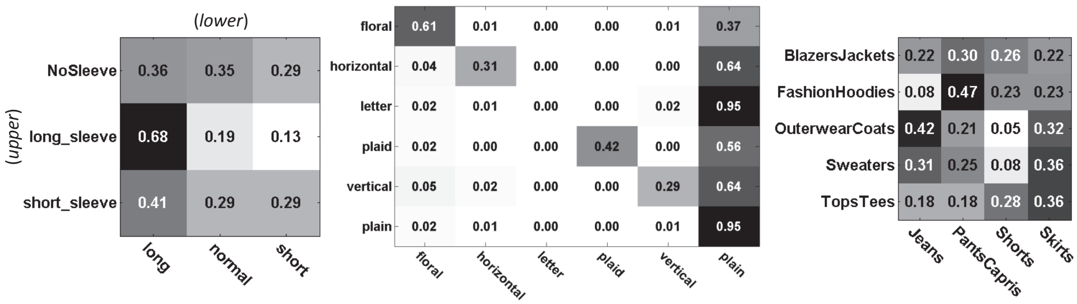


Figure 11: The mined attribute-attribute matching rules. The darker color means stronger correlation.

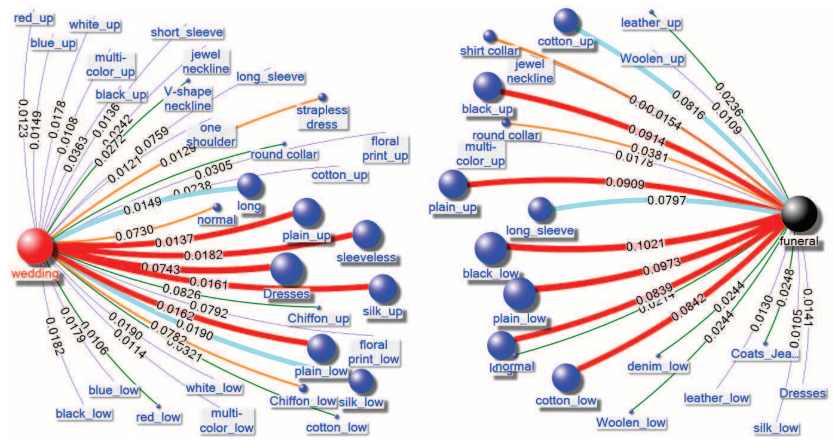


Figure 12: The mined occasion-attribute matching rules. The left is for the wedding occasion and the right is for the funeral occasion. Balls are for occasion or attributes. The size of balls and links between balls indicates the strength of their correlation.

ing based on $\{\mathbf{x}^{(n)}, \mathbf{o}^{(n)}\}_{n=1}^{N}$, given an occasion, all the clothing in the repository are ranked according to the output confidence score of the feature-occasion SVM. The second baseline feature-attribute-occasion SVM is composed of two-layer linear SVM. The first-layer SVM linearly maps visual features to attribute values, which is trained based on $\{\mathbf{x}^{(n)}, \mathbf{a}_u^{(n)}, \mathbf{a}_l^{(n)}\}_{n=1}^{N}$. Then the visual features are converted into attribute confidence score vectors via such first-layer SVM. The second-layer SVM is trained on these attribute confidence vectors to predict their occasion labels. Similar to feature-occasion SVM, all clothing in the repository are ranked based on the output of the two-layer feature-attribute-occasion SVM.

We evaluate their performance via Normalized Discounted Cumulative Gain (NDCG) [12, 17] which is commonly used to evaluate ranking systems. NDCG is defined as follows,

$$NDCG@k = \frac{1}{Z}\sum_{j=1}^{k}\frac{2^{\mathrm{rel}(j)}-1}{\log(1+j)}, \tag{7}$$

where $\mathrm{rel}(\cdot)$ is a binary value indicating whether the sample is relevant (with value 1) or irrelevant (with value 0), and Z is a constant to normalize the calculated score.

5.2.2 Results and Discussion

Quantitative evaluation results of the clothing suggestion are shown in Figure 13. From the results, we can make following observations. 1) The feature-occasion SVM consistently outperforms the feature-attribute-occasion SVM. This is because that the visual features we adopt possess relatively strong discriminative power and its high dimensionality benefits linear classifiers in classification. Compared with this, it is harder to construct linear relationship between low-dimensional attribute confidence vector and occasions. 2) The proposed latent SVM model outperforms the two baseline models significantly. This result well demonstrates the effectiveness of the proposed model in mining matching rules among features, attributes, occasions and utilizing their correlation in occasion-oriented clothing suggestion.

Some exemplar clothing suggestion results from magic closet are shown in Figure 14. In this figure, for each occasion, clothing in the repository are ranked by their scores from the latent SVM model and the top 6 suggestions are presented. From the figure, we can observe that most of the suggestions match the corresponding occasions quite well. For the conference occasion, the magic closet suggests formal clothing (most of them are the “suit pants” with “shirt collar”). For the “sports” occasion, all the top suggestions from magic closet are made from “cotton” and comfortable. Though some suggestions are not perfect, but most of their attributes conform with the underlying matching rules. For example, in the “funeral” occasion, the two incorrect suggestions are also “black” and “long”, which are distinctive characteristics of the clothing for funeral.

5.3 Occasion-Oriented Clothing Pairing

Here, we evaluate the effectiveness of magic closet in another scenario: given an occasion and one’s daily clothing as reference, the task is to find the most suitable clothing from online shopping dataset which pairs well with the reference clothing and simultaneously is suitable for the occasion.

5.3.1 Experimental Setting

To simulate this scenario, all queries (*i.e.*, reference clothing) are selected from the WoW_Upper_DP (5,921 images) and WoW_Lower_DP (2,736 images) subsets as in Figure 8.

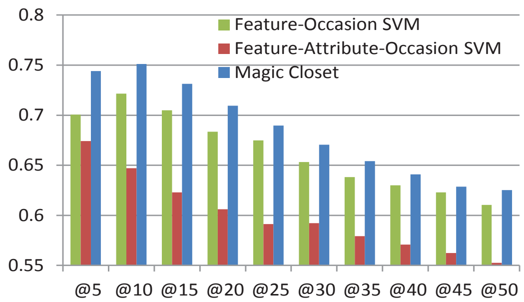


Figure 13: Comparison of magic closet with two baselines on the clothing suggestion task (NDCG vs. # returned samples).

Figure 14: Examples of the clothing recommendations. The clothing with highest scores for each occasions are shown in descending order. Incorrect recommendations are highlighted by red rectangles.

For each occasion, 20 clothing (10 upper-body and 10 lower-body) are randomly chosen as the queries. Summing up across 8 occasions[6], the total number of queries is 160.

The repository consists of clothing from online shopping dataset, including two subsets WoW_Upper_OS ($2,500$ images) and WoW_Lower_OS ($3,791$ images). We observe that many clothing are quite similar in the dataset. Therefore, we first cluster the clothing images from the two subsets based on their labeled attributes into 160 clusters, respectively. And only the most representative ones near the cluster centers are collected into the repository. This strategy ensures that the repository has wide coverage of the whole dataset. Reducing the size of repository also reduces the complexity of collecting ranking groundtruth, and is a good trade-off between an accurate evaluation and small labeling effort.

In clothing pairing, for each query of upper/lower-body clothing, we provide the rank of the candidate lower/upper-body clothing in the online shop dataset. The rank is calculated based on the pair's aesthetic score and suitableness for the specified occasion, as evaluated in Eqn. (6). To obtain the ranking ground truth of the returned clothing, we do not require labelers (40 people aging from 19 to 40) to score each candidate pair. We follow Gray et al. [13] and adopt the group-wise labeling strategy: given an occasion, we randomly show 8 clothing as a group to the labelers. So, labelers only need to rank the clothing within each group and the final rank is obtained using the method as in [13]. Such strategy can alleviate the burden of labelers significantly. Each pair is labeled at least 10 times and thus the potential inaccurate rank can be eliminated via averaging.

In the clothing pairing evaluation, we use the same two baseline models as in Section 5.2. For the proposed latent SVM based model (adopted by the magic closet), the clothing maximizing the pairing score as in Eqn. (6) is returned. And the ranking of the clothing is based on their scores. And NDCG in Eqn. (7) is also used to evaluate the performance of the magic closet and the baselines for clothing pairing. Note that here $\mathrm{ref}(\cdot)$ is the score from the labelers.

[6]The "wedding" and "banquet" occasions are not considered since the most suitable clothing are dress and pairing is not necessary.

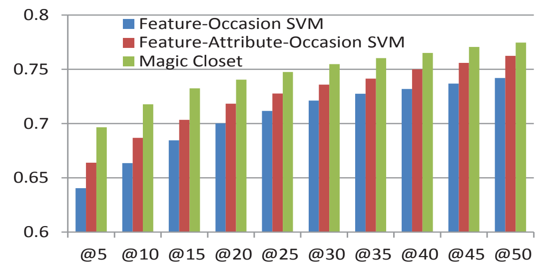
5.3.2 Results and Discussion

Figure 15 shows the NDCG value w.r.t. the increasing number of returned samples of the baseline models and magic closet system. From the figure, we can have the following observations. 1) For the two baseline methods, the feature-attribute-occasion SVM significantly performs better than the feature-occasion SVM. This is because that the feature-occasion SVM is a linear model. The calculated pairing score equals to $\mathbf{w}^T[\mathbf{x}_u;\mathbf{x}_l] = \mathbf{w}_u^T\mathbf{x}_u + \mathbf{w}_l^T\mathbf{x}_l$. The maximization of this score w.r.t. $\mathbf{x}_l$ is independent of $\mathbf{x}_u$. Therefore, in a specified occasion, for different queries, the returned results are identical. However, due to the good performance of feature-occasion SVM in occasion prediction, it can still return suitable clothing for the occasion. Thus its performance is still acceptable. While for the feature-attribute-occasion SVM, since the features are mapped to the attribute space at first, this issue is alleviated. Moreover, the attribute-based features are more robust to cross-domain variation (DP vs. OS). 2) The proposed magic closet outperforms the two baseline models. This result is as expected since magic closet can better capture matching rules among attributes and thus recommend more aesthetic clothing pairs, which has been verified partially in Section 5.2.

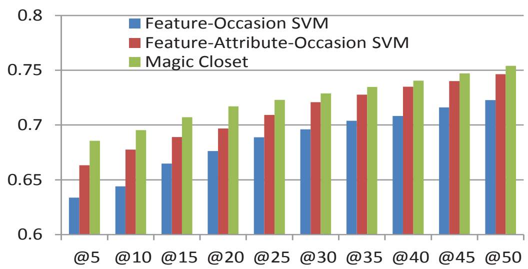
For more intuitive illustration, we give more exemplar paired clothing returned by magic closet for different queries in Figure 16. We can see that most of the returned clothing match the query and the specified occasion quite well. For the "sports" occasion, the query is "short pants", and the top 5 returned clothing all have "short sleeves" and are made of "cotton". For the second query in "sports", though the second returned result is scored lower by the labelers, it actually can be considered as pairing the query well considering its "long sleeves". For the "conference" occasions, the first query has "long sleeve" and "black" color, and the returned results also have "dark color" and are all "long pants". The paired clothing appear quite formal and are suitable for conference. While for the second query, the color is much "brighter" and long pants with brighter color are returned. We can further observe that most results are consistent with the groundtruth. These results clearly demonstrate the advantages of the magic closet in the clothing pairing scenario.

6. CONCLUSIONS

In this work, we developed a practical occasion oriented clothing recommendation and pairing system, named magic closet. Given a user specified occasion, the magic closet sys-



(a) Pairing with upper-body clothing



(b) Pairing with lower-body clothing

Figure 15: Comparison of magic closet with baselines for clothing pairing (NDCG vs. # returned samples).

tem is able to automatically recommend the most suitable clothing by considering the *wearing properly* and *wearing aesthetically* principles. We adopted a latent SVM based recommendation model to incorporate the matching rules among visual feature, attribute and occasion within a unified framework. To learn and evaluate the model, we collected a large clothing dataset with full attribute and occasion annotations. Extensive experiments were conducted on the collected dataset for the occasion-oriented clothing recommendation and clothing pairing tasks, and showed the effectiveness of the proposed model in capturing the underlying rules and recommending suitable clothing.

Note that the performance of the proposed model heavily depends on the human detection accuracy. Limited by the current performance of human detector in handling pose variance, some clothing in the user's clothing photo album may be misdetected. This issue can be further alleviated along with the development of state-of-the-art detection methods. In this work, we mainly focus on mining general rules and therefore we collect clothing from various users. In fact, the proposed framework can be personalized for specific users through taking their clothing images as training data and mining rules from them, or giving heavier weights on the training data from users'own clothing collection. In the future, we will investigate such personalization thoroughly.

Acknowledgment

This work was supported by 973 Program No. 2010CB327905, the National Science Foundation under Grant No. 60833006 & 60905008 and NExT Research Center funded under the research grant WBS. R-252-300-001-490 by MDA, Singapore.

Figure 16: Some exemplars of the paired clothing, given an occasion and one reference clothing. The most favored clothing for each occasion are shown in descending order. Their groundtruth scores are shown nearby.

7. REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *TPAMI*, 2006.

[2] T. Berg, A. Berg, and J. Shih. Automatic attribute discovery and characterization from noisy web data. In *ECCV*, 2010.

[3] M.K. Berkowitz and G.H. Haines Jr. Relative attributes. In *ICCV*, 2011.

[4] L. Bourdev, S. Maji, T. Brox, and J. Malik. Detecting people using mutually consistent poselet activations. In *ECCV*, 2010.

[5] L. Bourdev, S. Maji, and J. Malik. Describing people: A poselet-based approach to attribute classification. In *ICCV*, 2011.

[6] C. Cheng, M. Chung, M. Yu, M. Ouhyoung, H. Chu, and Y. Chuang. Chromirror: a real-time interactive mirror for chromatic and color-harmonic dressing. In *CHI*, 2008.

[7] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y. Xu. Color harmonization. In *SIGGRAPH*, 2006.

[8] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005.

[9] T. Do and T. Artières. Large margin training for hidden markov models with partially observed states. In *ICML*, 2009.

[10] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *CVPR*, 2009.

[11] P. Felzenszwalb, D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model. In *CVPR*, 2008.

[12] R. Feris and L. Davis. Image ranking and retrieval based on multi-attribute queries. In *CVPR*, 2011.

[13] D. Gray, K. Yu, W. Xu, and Y. Gong. Predicting facial beauty without landmarks. In *ECCV*, 2010.

[14] B. Hasan and D. Hogg. Segmentation using deformable spatial priors with application to clothing. In *BMVC*, 2010.

[15] T. Iwata, S. Wanatabe, and H. Sawada. Fashion coordinates recommender system using photographs from fashion magazines. In *IJCAI*, 2011.

[16] C. Lampert, H. Nickisch, and S. Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In *CVPR*, 2009.

[17] Z. Li, M. Wang, J. Liu, C. Xu, and H. Lu. News contextualization with geographic and visual information. In *MM*, 2011.

[18] B. Ni, Z. Song, and S. Yan. Web image mining towards universal age estimator. In *ACM MM*, 2009.

[19] Liu S., Song Z., Liu G.C., Xu C.S., Lu H.Q., and Yan S.C. Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set. In *CVPR*, 2012.

[20] E. Shen, H. Lieberman, and F. Lam. What am i gonna wear?: scenario-oriented recommendation. In *IUI*, 2007.

[21] Z. Song, M. Wang, X. Hua, and S. Yan. Predicting occupation via human clothing and contexts. In *ICCV*, 2011.

[22] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large margin methods for structured and interdependent output variables. *JMLR*, 2006.

[23] Y. Wang and G. Mori. Max-margin hidden conditional random fields for human action recognition. In *CVPR*, 2009.

[24] Y. Wang and G. Mori. A discriminative latent model of object classes and attributes. In *ECCV*, 2010.

[25] Y. Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In *CVPR*, 2011.