In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2020

ViBE: Dressing for Diverse Body Shapes

Wei-Lin Hsiao[1,2] Kristen Grauman[1,2]

[1]The University of Texas at Austin [2] Facebook AI Research

Abstract

Body shape plays an important role in determining what garments will best suit a given person, yet today's clothing recommendation methods take a "one shape fits all" approach. These body-agnostic vision methods and datasets are a barrier to inclusion, ill-equipped to provide good suggestions for diverse body shapes. We introduce ViBE, *a VIsual Body-aware Embedding that captures clothing's affinity with different body shapes. Given an image of a person, the proposed embedding identifies garments that will flatter her specific body shape. We show how to learn the embedding from an online catalog displaying fashion models of various shapes and sizes wearing the products, and we devise a method to explain the algorithm's suggestions for well-fitting garments. We apply our approach to a dataset of diverse subjects, and demonstrate its strong advantages over status quo body-agnostic recommendation, both according to automated metrics and human opinion.*

1. Introduction

Research in computer vision is poised to transform the world of consumer fashion. Exciting recent advances can link street photos to catalogs [47, 54], recommend garments to complete a look [25, 33, 34, 40, 73, 76], discover styles and trends [3, 32, 57], and search based on subtle visual properties [22, 46]. All such directions promise to augment and accelerate the clothing shopping experience, providing consumers with personalized recommendations and putting a content-based index of products at their fingertips.

However, when it comes to body shape, state-of-the-art recommendation methods falsely assume a "one shape fits all" approach. Despite the fact that the same garment will flatter different bodies differently, existing methods *neglect the significance of an individual's body shape when estimating the relevance of a given garment or outfit.* This limitation stems from two key factors. First, current large-scale datasets are heavily biased to a narrow set of body shapes[1]—typically thin and tall, owing to the fashionista or celebrity photos from which they are drawn [26, 51, 55, 66, 84] (see Fig. 1). This restricts everything learned downstream, including the extent of bodies considered for virtual try-on [26, 63, 79]. Second, prior methods to gauge clothing compatibility often learn from co-purchase patterns [25, 76, 77] or occasion-based rules [40, 53], divorced from any statistics on body shape.

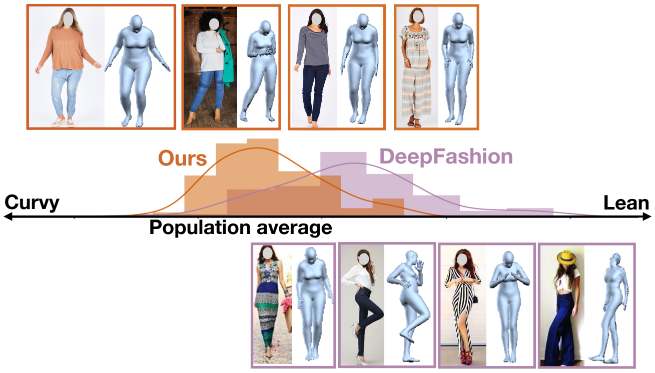


Figure 1: Trained largely from images of slender fashionistas and celebrities (bottom row), existing methods ignore body shape's effect on clothing recommendation and exclude much of the spectrum of real body shapes. Our proposed embedding considers diverse body shapes (top row) and learns which garments flatter which across the spectrum of the real population. Histogram plots the distribution of the second principal component of SMPL [56] (known to capture weight [31, 69]) for the dataset we collected (orange) and DeepFashion [55] (purple).

Body-agnostic vision methods and datasets are thus a barrier to diversity and inclusion. Meanwhile, aspects of fit and cut are paramount to what continues to separate the shopping experience in the physical world from that of the virtual (online) world. It is well-known that a majority of today's online shopping returns stem from problems with fit [61], and being unable to imagine how a garment would complement one's body can prevent a shopper from making the purchase altogether.

To overcome this barrier, we propose ViBE, a VIsual Body-aware Embedding that captures clothing's affinity with different body shapes. The learned embedding maps a given body shape and its most complementary garments close together. To train the model, we explore a novel source of Web photo data containing fashion models of di-

[1] not to mention skin tone, age, gender, and other demographic factors

verse body shapes. Each model appears in only a subset of all catalog items, and these pairings serve as implicit positive examples for body-garment compatibility.

Having learned these compatibilities, our approach can retrieve body-aware garment recommendations for a new body shape—a task we show is handled poorly by existing body-agnostic models, and is simply impossible for traditional recommendation systems facing a cold start. Furthermore, we show how to visualize what the embedding has learned, by highlighting what properties (sleeve length, fabric, cut, etc.) or localized regions (*e.g.*, neck, waist, straps areas) in a garment are most suitable for a given body shape.

We demonstrate our approach on a new body-diverse dataset spanning thousands of garments. With both quantitative metrics and human subject evaluations, we show the clear advantage of modeling body shape's interaction with clothing to provide accurate recommendations.

2. Related Work

Fashion styles and compatibility Early work on computer vision for fashion addresses recognition problems, like matching items seen on the street to a catalog [47, 54], searching for products [22, 46, 86], or parsing an outfit into garments [17, 51, 83, 87]. Beyond recognition, recent work explores models for *compatibility* that score garments for their mutual affinity [24,33,34,36,73,76,77]. Styles—meta-patterns in what people wear—can be learned from images, often with visual attributes [**?**,3,32,43,57], and Web photos with timestamps and social media "likes" can help model the relative popularity of trends [50, 74]. Unlike our approach, none of the above models account for the influence of body shape on garment compatibility or style.

Fashion image datasets Celebrities [30, 51], fashionista social media influencers [43, 52, 74, 83, 84], and catalog models [18, 26, 55, 66] are all natural sources of data for computer vision datasets studying fashion. However, these sources inject bias into the body shapes (and other demographics) represented, which can be useful for some applications but limiting for others. Some recent dataset efforts leverage social media and photo sharing platforms like Instagram and Flickr which may access a more inclusive sample of people [42, 57], but their results do not address body shape. We explore a new rich online catalog dataset comprised of models of diverse body shape.

Virtual try on and clothing retargeting Virtual try-on entails visualizing a source garment on a target human subject, as if the person were actually wearing it. Current methods estimate garment draping on a 3D body scan [20,48,62,68], retarget styles for people in 2D images or video [4, 5, 7, 85], or render a virtual try-on with sophisticated image generation methods [23, 26, 63, 79]. While existing methods display a garment on a person, they do not infer whether the garment flatters the body or not. Furthermore, in practice, vision-based results are limited to a narrow set of body shapes (typically tall and thin as in Fig. 1) due to the implicit bias of existing datasets discussed above.

Body and garment shape estimation Estimating people and clothing's 3D geometry from 2D RGB images has a long history in graphics, broadly categorizable into body only [8, 38, 89], garment only [11, 37, 82, 88], joint [49, 59, 67], and simultaneous but separate estimations [4, 5, 7, 85]. In this work, we integrate two body-based models to estimate a user's body shape from images. However, different from any of the above, our approach goes beyond estimating body shape to learn the affinity between human body shape and well-fitting garments.

Sizing clothing While most prior work recommends clothing based on an individual's purchase history [28,35,39,77] or inferred style model [33, 40, 53], limited prior work explores product *size recommendation* [14, 21, 41, 58, 72]. Given a product and the purchase history of a user, these methods predict whether a given size will be too large, small, or just right. Rather than predict which size of a given garment is appropriate, our goal is to infer which garments will flatter the body shape of a given user. Moreover, unlike our approach, existing methods do not consider the visual content of the garments or person [14,21,58,72]. While SizeNet [41] uses product images, the task is to predict whether the product will have fit issues in general, unconditioned on any person's body.

Clothing preference based on body shape To our knowledge, the only prior work that considers body shape's connection to clothing is the "Fashion Takes Shape" project, which studies the correlation between a subject's weight and clothing categories typically worn (*e.g.*, curvier people are more likely to wear jeans than shorts) [69], and the recommendation system of [30] that discovers which styles are dominant for which celebrity body types given their known body measurements. In contrast to either of these methods, our approach suggests specific garments conditioned on an individual's body shape. Furthermore, whereas [69] is about observing in hindsight what a collection of people wore, our approach actively makes recommendations for novel bodies and garments. Unlike [30], our method handles data beyond high-fashion celebrities and uses the inferred body shape of a person as input.

3. Approach

While the reasons for selecting clothes are complex [80], *fit* in a garment is an important factor that contributes to the confidence and comfort of the wearer. Specifically, a garment that fits a wearer well *flatters* the wearer's body. Fit is a frequent reason for whether to make an apparel purchase [6]. Searching for the right fit is time-consuming:

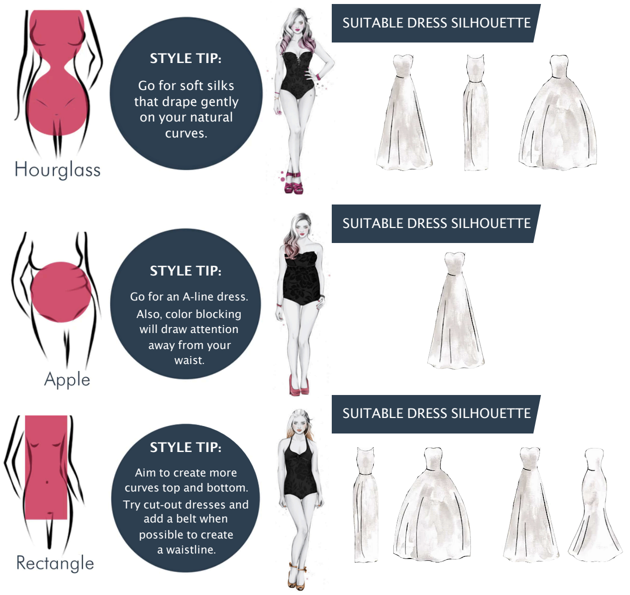

Hourglass
STYLE TIP:
Go for soft silks that drape gently on your natural curves.
SUITABLE DRESS SILHOUETTE
Apple
STYLE TIP:
Go for an A-line dress. Also, color blocking will draw attention away from your waist.
SUITABLE DRESS SILHOUETTE
Rectangle
STYLE TIP:
Aim to create more curves top and bottom. Try cut-out dresses and add a belt when possible to create a waistline.
SUITABLE DRESS SILHOUETTE

Figure 2: Example categories of body shapes, with styling tips and recommended dresses for each, according to fashion blogs [1, 2].

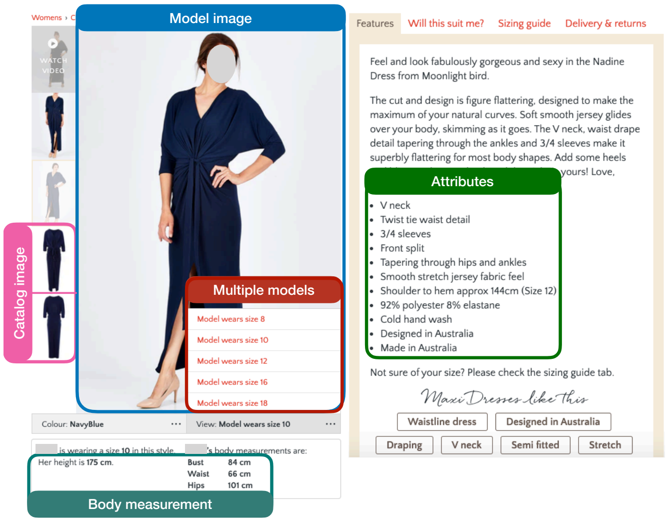
women may try on as many as 20 pairs of jeans before they find a pair that fits [64].

The 'Female Figure Identification Technique (FFIT) System' classifies the female body into 9 shapes—hourglass, rectangle, triangle, spoon, etc.—using the proportional relationships of dimensions for bust, waist, high hip, and hips [13]. No matter which body type a woman belongs to, researchers find that women participants tend to select clothes to create an hourglass look for themselves [19]. Clothing is used strategically to manage bodily appearance, so that perceived "problem areas/flaws" can be covered up, and assets are accentuated [16, 19]. Fig. 2 shows examples from fashion blogs with different styling tips and recommended dresses for different body shapes.

Our goal is to discover such strategies, by learning a body-aware embedding that recommends clothing that complements a specific body and vice versa. We first introduce a dataset and supervision paradigm that allow for learning such an embedding (Sec. 3.1, Sec. 3.2). Then we present our model (Sec. 3.3) and the representation we use for clothing and body shape (Sec. 3.4). Finally, beyond recommending garments, we show how to visualize the *strategies* learned by our model (Sec. 3.5).

3.1. A Body-Diverse Dataset

An ideal dataset for learning body-garment compatibility should meet the following properties: (1) clothed people with diverse body shapes; (2) full body photos so the body shapes can be estimated; (3) some sort of rating of whether the garment flatters the person to serve as supervision. Datasets with 3D scans of people in clothing [4, 5, 7, 65] meet (1) and (2), but are rather small and



Feel and look fabulously gorgeous and sexy in the Nadine Dress from Moonlight bird.

The cut and design is figure flattering, designed to make the maximum of your natural curves. Soft smooth jersey glides over your body, skimming as it goes. The V neck, waist drape detail tapering through the ankles and 3/4 sleeves make it superbly flattering for most body shapes. Add some heels yours! Love,

- V neck
- Twist tie waist detail
- 3/4 sleeves
- Front split
- Tapering through hips and ankles
- Smooth stretch jersey fabric feel
- Shoulder to hem approx 144cm (Size 12)
- 92% polyester 8% elastane
- Cold hand wash
- Designed in Australia
- Made in Australia

Not sure of your size? Please check the sizing guide tab.

Maxi Dresses like this

Model wears size 8
Model wears size 10
Model wears size 12
Model wears size 16
Model wears size 18

Colour: NavyBlue

View: Model wears size 10

[redacted] is wearing a size 10 in this style. [redacted]'s body measurements are:

Her height is 175 cm.

Bust	84 cm
Waist	66 cm
Hips	101 cm

Figure 3: Example page from the website where we collected our dataset. It provides the image of the model wearing the catalog item, the clean catalog photo of the garment on its own, the model's body measurements, and the item's attribute description. Each item is worn by models of multiple body shapes.

have limited clothing styles. Datasets of celebrities [30, 51], fashionistas [43, 74, 84], and catalog models [26, 55, 66] satisfy (2) and (3), but they lack body shape diversity. Datasets from social media platforms [42, 57] include more diverse body shapes (1), but are usually cluttered and show only the upper body, preventing body shape estimation.

To overcome the above limitations, we collect a dataset from an online shopping website called Birdsnest.[2] Birdsnest provides a wide range of sizes (8 to 18 in Australian measurements) in most styles. Fig. 3 shows an example catalog page. It contains the front and back views of the garment, the image of the fashion model wearing the item, her body measurements, and an attribute-like textual descriptions of the item. Most importantly, each item is worn by a variety of models of different body shapes. We collect two categories of items, 958 dresses and 999 tops, spanning 68 fashion models in total. While our approach is not specific to women, since the site has only women's clothing, our current study is focused accordingly. This data provides us with properties (1) and (2). We next explain how we obtain positive and negative examples from it, property (3).

3.2. Implicit Rating from Catalog Fashion Models

Fashion models wearing a specific catalog item can safely be assumed to have body shapes that are flattered by that garment. Thus, the catalog offers implicit positive body-garment pairings. How do we get negatives? An intuitive way would be to assume that all unobserved body-garment pairings from the dataset are negatives. However,

[2] https://www.birdsnest.com.au/

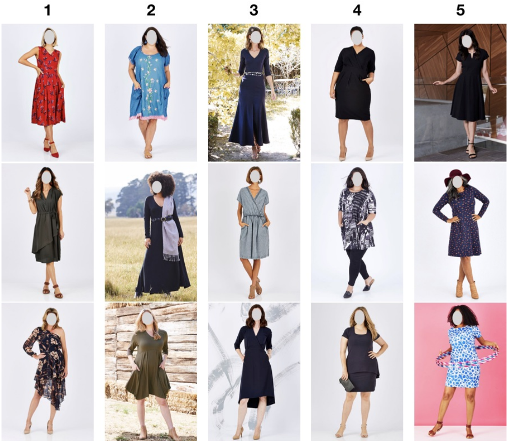


Figure 4: Columns show bodies sampled from the five discovered body types for dresses (see Supp. for tops). Each type roughly maps to 1) average, 2) curvy, 3) slender, 4) tall and curvy, 5) petite.

about 50% of the dresses are worn by only 1 or 2 distinct bodies (3% of the models), suggesting that many positive pairings are not observed.

Instead, we propose to propagate missing positives between similar body shapes. Our assumption is that if two body shapes are very similar, clothing that flatters one will likely flatter the other. To this end, we use k-means [78] clustering (on features defined in Sec. 3.4) to quantize the body shapes in our dataset into five types. Fig. 4 shows bodies sampled from each cluster. We propagate positive clothing pairs from each model observed wearing a garment to all other bodies of her type. Since most of the garments are worn by multiple models, and thus possibly multiple types, we define negative clothing for a type by pairing bodies in that type with clothing *never* worn by any body in that type.

With this label propagation, most dresses are worn by 2 distinct body *types*, which is about 40% of the bodies in the dataset, largely decreasing the probability of missing true positives. To validate our label propagation procedure with ground truth, we conduct a user study explicitly asking human judges on Mechanical Turk whether each pair of bodies in the same cluster could wear similar clothing, and whether pairs in different clusters could. Their answers agreed with the propagated labels 81% and 63% of the time for the two respective cases (see Supp. for details).

3.3. Training a Visual Body-Aware Embedding

Now having the dataset with all the desired properties, we introduce our VIsual Body-aware Embedding, ViBE, that captures clothing's affinity with body shapes. In an ideal embedding, nearest neighbors are always relevant instances, while irrelevant instances are separated by a large margin. This goal is achieved by correctly ranking all triplets, where each triplet consists of an anchor z_a, a positive z_p that is relevant to z_a, and a negative z_n that is not relevant to z_a. The embedding should rank the positive closer to the anchor than the negative, $D(z_a, z_p) < D(z_a, z_n)$ (with $D(.,.)$ denoting Euclidean distance). A margin-based loss [81] optimizes for this ranking:

$$\mathcal{L}(z_a, z_p, z_n) := (D(z_a, z_p) - \alpha_p)_+ + (\alpha_n - D(z_a, z_n))_+$$

where α_p, α_n is the margin for positive and negative pairs respectively, and the subscript $+$ denotes $\max(0, \cdot)$. We constrain the embedding to live on the d-dimensional hypersphere for training stability, following [70].

In our joint embedding ViBE, we have two kinds of triplets, one between bodies and clothing, and one between bodies and bodies. So our final loss combines two instances of the margin-based loss:

$$\mathcal{L} = \mathcal{L}_{body,cloth} + \mathcal{L}_{body,body}. \tag{1}$$

Let f_{cloth}, f_{body} be the respective functions that map instances of clothing x_g and body shape x_b to points in ViBE. For the triplet in our body-clothing loss $\mathcal{L}_{body,cloth}$, z_a is a mapped body instance $f_{body}({x_b}^a)$, z_p is a compatible clothing item $f_{cloth}({x_g}^p)$, and z_n is an incompatible clothing item $f_{cloth}({x_g}^n)$. This loss aims to map body shapes near their compatible clothing items.

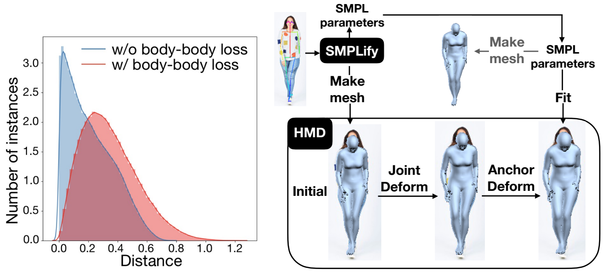
We introduce the body-body loss $\mathcal{L}_{body,body}$ to facilitate training stability. Recall that each garment could be compatible with multiple bodies. By simply pulling these shared clothing items closer to all their compatible bodies, all clothing worn on those bodies would also become close to each other, making the embedding at risk of model collapse (see Fig. 5a, blue plot). Hence, we introduce an additional constraint on triplets of bodies: z_a is again a mapped body instance $f_{body}({x_b}^a)$, z_p is now a body $f_{body}({x_b}^p)$ that belongs to the same type (*i.e.*, cluster) as ${x_b}^a$, and z_n is a body $f_{body}({x_b}^n)$ from a different type. This body-body loss $\mathcal{L}_{body,body}$ explicitly distinguishes similar bodies from dissimilar ones. Fig. 5a plots the distribution of pairwise clothing distances with and without this additional constraint, showing that this second loss effectively alleviates the model collapse issue.

We stress that the quantization for body types (Sec. 3.2) is solely for propagating labels to form the training triplets. When learning and applying the embedding itself, we operate in a continuous space for the body representation. That is, a new image is mapped to *individualized* recommendations potentially unique to that image, *not* a batch of recommendations common to all bodies within a type.

3.4. Clothing and Body Shape Features

Having defined the embedding's objective, now we describe the input features x_b and x_g for bodies and garments.

For clothing, we have the front and back view images of the catalog item (without a body) and its textual description.



(a) (b)

Figure 5: Left (a): Distribution of pairwise distances between clothing items with (red) and without (blue) the proposed body-body triplet loss. Without it, clothing embeddings are very concentrated and have close to 0 distance, causing instability in training. Right (b): Human body shape estimation stages.

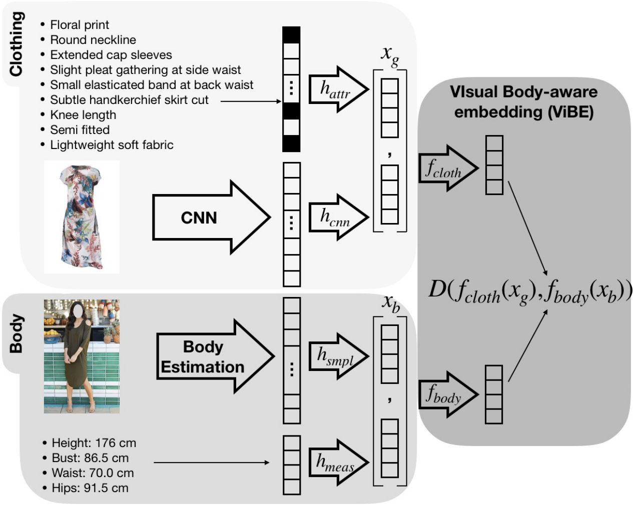


Figure 6: Overview of our visual body-aware embedding (ViBE). We use mined attributes with CNN features for clothing, and estimated SMPL [56] parameters and vital statistics for body shape (Sec. 3.4). Following learned projections, they are mapped into the joint embedding that measures body-clothing affinities (Sec. 3.3).

We use a ResNet-50 [27] pretrained on ImageNet [12] to extract visual features from the catalog images, which captures the overall color, pattern, and silhouette of the clothing. We mine the top frequent words in all descriptions for all catalog entries to build a vocabulary of attributes, and obtain an array of binary attributes for each garment, which captures localized and subtle properties such as specific necklines, sleeve cuts, and fabric.

For body shape, we have images of the fashion models and their measurements for height, bust, waist, and hips, the so called *vital statistics*. We concatenate the vital statistics in a 4D array and standardize them. However, the girths and lengths of limbs, the shoulder width, and many other characteristics of the body shape are not captured by the vital statistics, but are visible in the fashion models' images. Thus, we estimate a 3D human body model from each image to capture these fine-grained shape cues.

To obtain 3D shape estimates, we devise a hybrid approach built from two existing methods, outlined in Fig. 5b. Following the basic strategy of HMD [89], we estimate an initial 3D mesh, and then stage-wise update the 3D mesh by projecting it back to 2D and deforming it to fit the silhouette of the human in the RGB image. However, the initial 3D mesh that HMD is built on, *i.e.*, HMR [38], only supports gender-neutral body shapes. Hence we use SMPLify [8], which does support female bodies, to create the initial mesh.[3] We then deform the mesh with HMD.

Finally, rather than return the mesh itself—whose high-dimensionality presents an obstacle for data efficient embedding learning— we optimize for a compact set of body shape model parameters that best fits the mesh. In particular, we fit SMPL [56] to the mesh and use its first 10 principal components as our final 3D body representation. These dimensions roughly capture weight, waist height, masculine/feminine characteristics, etc. [31, 75]. When multiple images (up to 6) for a fashion model are available, we process all of them, and take the median per dimension.

In summary, for clothing, we accompany mined attributes (64 and 100 attributes for dresses and tops respectively) with CNN features (2048-D); for body shape, we accompany estimated 3D parameters (10-D) with vital statistics (4-D). Each is first reduced into a lower dimensional space with learned projection functions (h_{attr}, h_{cnn}, h_{smpl}, h_{meas}). Then the reduced attribute and CNN features are concatenated as the representation x_g for clothing, and the reduced SMPL and vital features are concatenated as the representation x_b for body shape. Both are forwarded into the joint embedding (defined in Sec. 3.3) by f_{cloth} and f_{body} to measure their affinity. Fig. 6 overviews the entire procedure. See Supp. for architecture details.

3.5. Recommendations and Explanation

After learning our embedding, we make clothing recommendations for a new person by retrieving the garments closest to her body shape in this space. In addition, we propose an automatic approach to convey the underlying strategy learned by our model. The output should be general enough for users to apply to future clothing selections, in the spirit of the expert advice as shown in Fig. 2—e.g., the styling tip for *apple* body shape is to wear A-line dresses—but potentially even more tailored to the individual body.

To achieve this, we visualize the embedding's learned decision with separate classifiers (cf. Fig. 10). We first map a subject's body shape into the learned embedding, and take

[3] We apply OpenPose [9] to the RGB images to obtain the 2D joint positions required by SMPLify. We could not directly use the SMPLify estimated bodies because only their pose is accurate but not their shape.

		Dresses					Tops				
	type	1	2	3	4	5	1	2	3	4	5
Train	body	18	7	11	4	6	19	4	8	15	6
	clothing	587	481	301	165	167	498	202	481	493	232
Test	body	5	2	3	2	2	5	2	3	4	2
	clothing	149	126	76	42	34	115	54	115	129	58

Table 1: Dataset statistics: number of garments and fashion models for each clustered type.

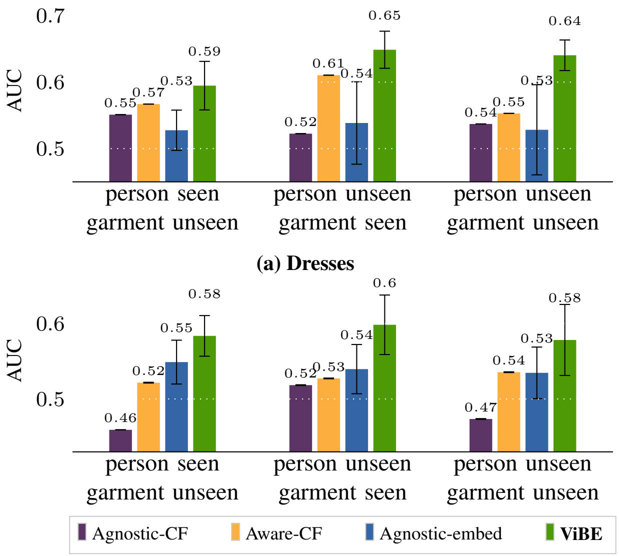
the closest and furthest 400 clothing items as the most and least suitable garments for this subject. We then train binary classifiers to predict whether a clothing item is suitable for this subject. By training a linear classifier over the attribute features of the clothing, the high and low weights reveal the most and least suitable attributes for this subject. By training a classifier over the CNN features of the clothing, we can apply CNN visualization techniques [15, 60, 71] (we use [60]) to localize important regions (as heatmaps) that activate the positive or negative prediction.

4. Experiments

We now evaluate our body-aware embedding with both quantitative evaluation and user studies.

Experiment setup. Using the process described in Sec. 3.1 and Sec. 3.2, we collect two sets of data, one for dresses and one for tops, and train separately on each for all models. To propagate positive labels, we cluster the body shapes to $k = 5$ types. We find the cluster corresponding to an *average* body type is the largest, while *tall and curvy* is the smallest. To prevent the largest cluster's bodies from dominating the evaluation, we randomly hold out 20%, or at least two bodies, for each cluster to comprise the test set. For clothing, we randomly hold out 20% of positive clothing for each cluster. Tab. 1 summarizes the dataset breakdown.

Baselines. Since no prior work tackles this problem, we develop baselines based on problems most related to ours: user-item recommendation and garment compatibility modeling. Suggesting clothing to flatter a body shape can be treated as a recommendation problem, where people are users and clothing are items. We compare with two standard recommendation methods: (1) body-AGNOSTIC-CF: a vanilla collaborative filtering (CF) model that uses neither users' nor items' content; and (2) body-AWARE-CF: a hybrid CF model that uses the body features and clothing visual features as content ("side information" [10]). Both use a popular matrix completion [45] algorithm [29]. In addition, we compare to a (3) body-AGNOSTIC-EMBEDDING that uses the exact same features and models as our body-AWARE-EMBEDDING (ViBE), but—as done implicitly by current methods—is only trained on bodies of the same type, limiting body shape diversity.[4] It uses all bodies and clothing in the largest cluster (*average* body type), since results for the baseline were best on this type. This baseline resembles current embeddings for garment compatibility [28, 76, 77], by changing garment type to body shape.



(a) Dresses

(b) Tops

Figure 7: Recommendation accuracy measured by AUC over all person-garment pairs. Our body-aware embedding (ViBE) performs best on all test scenarios by a clear margin.

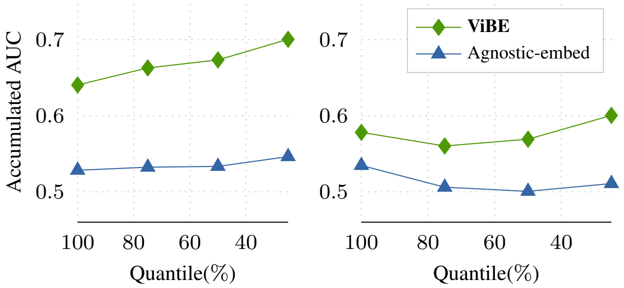
Implementation. All dimensionality reduction functions h_{attr}, h_{cnn}, h_{smpl}, h_{meas} are 2-layer MLPs, and the embedding functions f_{cloth} and f_{body} are single fully connected layers. We train the body-aware (agnostic) embeddings with Adam-optimizer [44], learning rate 0.003 (0.05), weight decay 0.01, decay the learning rate by 0.3 at epoch 100 (70) and 130 (100), and train until epoch 180 (130). See Supp. for more architecture and training details. We use the best models in quantitative evaluation for each method to run the human evaluation.

4.1. Quantitative evaluation

We compare the methods on three different recommendation cases: i) person ("user") seen but garment ("item") unseen during training, ii) garment seen but person unseen, iii) neither person nor garment seen. These scenarios capture realistic use cases, where the system must make recommendations for new bodies and/or garments. We exhaust all pairings of test bodies and clothing, and report the mean AUC with standard deviation across 10 runs.

Fig. 7a and Fig. 7b show the results. Our model outperforms all methods by a clear margin. AGNOSTIC-CF performs the worst, as all three test cases involve cold-start problems, and it can only rely on the learned bias terms. Including the person's body shape and clothing's features in the CF method (AWARE-CF) significantly boosts its performance, demonstrating the importance of this content for clothing recommendation. In general, the embedding-based

[4] The proposed body-body triplet loss is not valid for this baseline.



(a) Dresses **(b) Tops**

Figure 8: Accuracy trends as test garments are increasingly body-specific. We plot AUC from all clothing (100%) then gradually exclude body-versatile ones, until only the most body-specific (25%) are left. ViBE offers even greater improvement when clothing is body-specific (*least* body-versatile), showing recommendations for those garments only succeed if the body is taken into account.

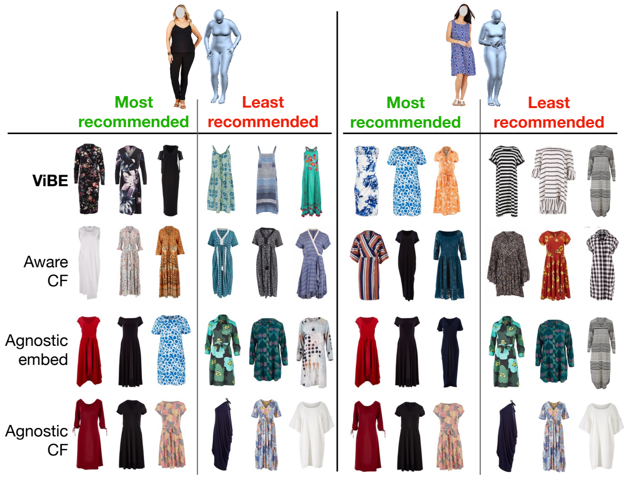


Figure 9: Example recommendations for 2 subjects by all methods. Subjects' images and their estimated body shapes are shown on the top of the tables. Each row gives one method's most and least recommended dresses. See text for discussion.

methods perform better than the CF-based methods. This suggests that clothing-body affinity is modeled better by ranking than classification; an embedding can maintain the individual idiosyncrasies of the body shapes and garments.

All methods perform better on dresses than tops. This may be due to the fact that dresses cover a larger portion of the body, and thus could be inherently more selective about which bodies are suitable. In general, the more selective or *body-specific* a garment is, the more value a body-aware recommendation system can offer; the more *body-versatile* a garment is, the less impact an intelligent recommendation can have. To quantify this trend, we evaluate the embeddings' accuracy for scenario (iii) as a function of the test garments' *versatility*, as quantified by the number of distinct body types (clusters) that wear the garment. Fig. 8 shows the results. As we focus on the body-specific garments (right hand side of plots) our body-aware embedding's gain over the body-agnostic baseline increases.

	Agnostic-CF	Aware-CF	Agnostic-embed	**ViBE**
AUC	0.51	0.52	0.55	**0.58**

Table 2: Recommendation AUC on unseen people paired with garments sampled from the entire dataset, where ground-truth labels are provided by human judges. Consistent with Fig. 7a, the proposed model outperforms all the baselines.

4.2. Example recommendations and explanations

Fig. 9 shows example recommendations for all methods on two heldout subjects: each row is a method, with most and least recommended garments. Being agnostic to body shape, AGNOSTIC-CF and AGNOSTIC-EMBEDDING make near identical recommendations for subjects with different body shapes: top recommended dresses are mostly body-versatile (captured as *popularity* by the bias term in CF based methods), while least recommended are either body-specific or less interesting, solid shift dresses. ViBE recommends knee-length, extended sleeves, or wrap dresses for curvy subjects, which flow naturally on her body, and recommends shorter dresses that fit or flare for the slender subjects, which could show off her legs.

Fig. 10 shows example explanations (cf. Sec. 3.5) for ViBE's recommendations. For a petite subject, the most suitable attributes are waistbands and empire styles that create taller looks, and embroidery and ruffles that increase volume. For a curvier subject, the most suitable attributes are extended or 3/4 sleeves that cover the arms, v-necklines that create an extended slimmer appearance, and wrap or side-splits that define waists while revealing curves around upper-legs. The heatmaps showing important regions for why a dress is suitable for the subject closely correspond to these attributes. We also take the top 10 suitable dresses and their heatmaps to generate a weighted average dress to represent the *gestalt* shape of suitable dresses for this person.

4.3. Human judged ground truth evaluation

Having quantified results against the catalog ground truth, next we solicit human opinions. We recruit 329 subjects on Mechanical Turk to judge which dresses better flatter the body shape of the test subjects. See Supp. for all user study interfaces. We first ask subjects to judge each dress as either *body-specific* or *body-versatile*. Then we randomly sample 10 to 25 pairs of clothing items that are the same type (*i.e.*, both body-specific or -versatile) for each of 14 test bodies, and for each one we ask 7 subjects to rank which dress is more suitable for the given body. We discard responses with low consensus (*i.e.*, difference of votes is less than 2), which yields 306 total pairs.

Tab. 2 shows the results for all methods. The overall trend is consistent with the automatic evaluation in Fig-

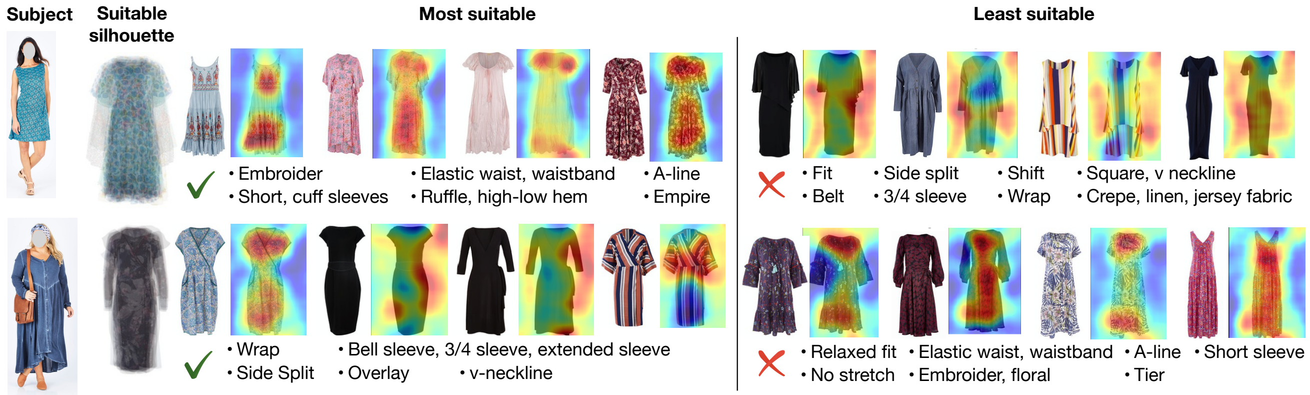


Figure 10: Example recommendations and explanations from our model: for each subject (row), we show the predicted most (left) and least (right) suitable attributes (text at the bottom) and garments, along with the garments' explanation localization maps. The "suitable silhouette" image represents the *gestalt* of the recommendation. The localization maps show where our method sees (un)suitable visual details, which agree with our method's predictions for (un)recommended attributes.

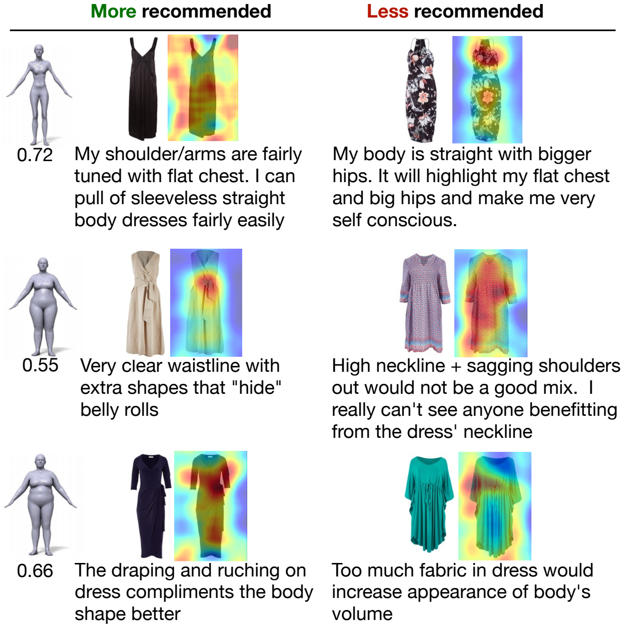

More recommended
Less recommended
0.72
My shoulder/arms are fairly tuned with flat chest. I can pull of sleeveless straight body dresses fairly easily
My body is straight with bigger hips. It will highlight my flat chest and big hips and make me very self conscious.
0.55
Very clear waistline with extra shapes that "hide" belly rolls
High neckline + sagging shoulders out would not be a good mix. I really can't see anyone benefitting from the dress' neckline
0.66
The draping and ruching on dress compliments the body shape better
Too much fabric in dress would increase appearance of body's volume

Figure 11: Examples of our model's more/less recommended dresses for users (body types selected by users; numbers shown under are AUC for each), along with the reasons why users preferred a dress or not. Our model's explanation roughly corresponds to users' reasoning: user 2 prefers a clear waistline to hide the belly, while user 1 tends to draw attention away from the chest.

ure 7a. As tops are in general less body-specific than dresses, human judges seldom reach consensus for tops, thus we did not include a human annotated benchmark for it. See Supp. for examples of Turkers' explanations for their selections. We share the collected ground truth to allow benchmarking future methods.[5]

Next we perform a second user study in which women judge which garments would best flatter their *own* body shape, since arguably each person knows her own body best. We first ask subjects to select the body shape among 15 candidates (adopted from BodyTalk [75]) that best resembles themselves, and then select which dresses they prefer to wear. We use the selected dresses as positive, unselected as negative, and evaluate our model's performance by ranking AUC. In total, 4 volunteers participated, each answered 7 to 18 different pairs of dresses, summing up to 61 pairs of dresses. Our body-aware embedding[6] achieves a mean AUC of 0.611 across all subjects, compared to 0.585 by the body-agnostic embedding (the best competing baseline).

Fig. 11 shows our method's recommendations for cases where subjects explained the garments they preferred (or not) for their own body shape. We see that our model's visual explanation roughly corresponds to subjects' own reasoning (*e.g.*, (de)emphasizing specific areas).

5. Conclusion

We explored clothing recommendations that complement an individual's body shape. We identified a novel source of Web photo data containing fashion models of diverse body shapes, and developed a body-aware embedding to capture clothing's affinity with different bodies. Through quantitative measurements and human judgments, we verified our model's effectiveness over body-agnostic models, the status quo in the literature. In future work, we plan to incorporate our body-aware embedding to address fashion styling and compatibility tasks.

Acknowledgements: We thank our human subjects: Angel, Chelsea, Cindy, Layla, MongChi, Ping, Yenyen, and our anonymous friends and volunteers from Facebook. We also thank the authors of [75] for kindly sharing their collected SMPL parameters with us. UT Austin is supported in part by NSF IIS-1514118.

[5] `http://vision.cs.utexas.edu/projects/VIBE`

[6] Since we do not have these subjects' vital statistics, we train another version of our model that uses only SMPL and CNN features.

References

[1] https://chic-by-choice.com/en/what-to-wear/best-dresses-for-your-body-type-45. 3

[2] https://www.topweddingsites.com/wedding-blog/wedding-attire/how-to-guide-finding-the-perfect-gown-for-your-body-type. 3

[3] Z. Al-Halah, R. Stiefelhagen, and K. Grauman. Fashion forward: Forecasting visual style in fashion. In *ICCV*, 2017. 1, 2

[4] Thiemo Alldieck, Marcus Magnor, Bharat Lal Bhatnagar, Christian Theobalt, and Gerard Pons-Moll. Learning to reconstruct people in clothing from a single rgb camera. In *CVPR*, 2019. 2, 3

[5] Thiemo Alldieck, Marcus Magnor, Weipeng Xu, Christian Theobalt, and Gerard Pons-Moll. Video based reconstruction of 3d people models. In *CVPR*, 2018. 2, 3

[6] Kurt Salmon Associates. Annual consumer outlook survey. *presented at a meeting of the American Apparel and Footwear Association Apparel Research Committee*, 2000. 2

[7] Bharat Lal Bhatnagar, Garvita Tiwari, Christian Theobalt, and Gerard Pons-Moll. Multi-garment net: Learning to dress 3d people from images. In *ICCV*, 2019. 2, 3

[8] Federica Bogo, Angjoo Kanazawa, Christoph Lassner, Peter Gehler, Javier Romero, and Michael J. Black. Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image. In *ECCV*, 2016. 2, 5

[9] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*, 2017. 5

[10] Tianqi Chen, Weinan Zhang, Qiuxia Lu, Kailong Chen, Zhao Zheng, and Yong Yu. Svdfeature: a toolkit for feature-based collaborative filtering. *JMLR*, 2012. 6

[11] R Daněřek, Endri Dibra, Cengiz Öztireli, Remo Ziegler, and Markus Gross. Deepgarment: 3d garment shape estimation from a single image. In *Computer Graphics Forum*, 2017. 2

[12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A Large-Scale Hierarchical Image Database. In *CVPR*, 2009. 5

[13] Priya Devarajan and Cynthia L Istook. Validation of female figure identification technique (ffit) for apparel software. *Journal of Textile and Apparel, Technology and Management*, 2004. 3

[14] Kallirroi Dogani, Matteo Tomassetti, Sofie De Cnudde, Saúl Vargas, and Ben Chamberlain. Learning embeddings for product size recommendations. In *SIGIR Workshop on ECOM*, 2018. 2

[15] Ruth C Fong and Andrea Vedaldi. Interpretable explanations of black boxes by meaningful perturbation. In *ICCV*, 2017. 6

[16] Hannah Frith and Kate Gleeson. Dressing the body: The role of clothing in sustaining body pride and managing body distress. *Qualitative Research in Psychology*, 2008. 3

[17] Cheng-Yang Fu, Tamara L. Berg, and Alexander C. Berg. Imp: Instance mask projection for high accuracy semantic segmentation of things. In *ICCV*, 2019. 2

[18] Yuying Ge, Ruimao Zhang, Lingyun Wu, Xiaogang Wang, Xiaoou Tang, and Ping Luo. A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. *CVPR*, 2019. 2

[19] Sarah Grogan, Simeon Gill, Kathryn Brownbridge, Sarah Kilgariff, and Amanda Whalley. Dress fit and body image: A thematic analysis of women's accounts during and after trying on dresses. *Body Image*, 2013. 3

[20] Peng Guan, Loretta Reiss, David A Hirshberg, Alexander Weiss, and Michael J Black. Drape: Dressing any person. *TOG*, 2012. 2

[21] Romain Guigourès, Yuen King Ho, Evgenii Koriagin, Abdul-Saboor Sheikh, Urs Bergmann, and Reza Shirvany. A hierarchical bayesian model for size recommendation in fashion. In *RecSys*, 2018. 2

[22] X. Guo, H. Wu, Y. Cheng, S. Rennie, and R. Feris. Dialog-based interactive image retrieval. In *NIPS*, 2018. 1, 2

[23] Xintong Han, Xiaojun Hu, Weilin Huang, and Matthew R. Scott. Clothflow: A flow-based model for clothed person generation. In *ICCV*, 2019. 2

[24] Xintong Han, Zuxuan Wu, Weilin Huang, Matthew R Scott, and Larry S Davis. Compatible and diverse fashion image inpainting. *ICCV*, 2019. 2

[25] Xintong Han, Zuxuan Wu, Yu-Gang Jiang, and Larry S. Davis. Learning fashion compatibility with bidirectional lstms. In *ACM MM*, 2017. 1

[26] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S Davis. Viton: An image-based virtual try-on network. In *CVPR*, 2018. 1, 2, 3

[27] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 5

[28] R. He, C. Packer, and J. McAuley. Learning compatibility across categories for heterogeneous item recommendation. In *ICDM*, 2016. 2, 6

[29] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *WWW*, 2017. 6

[30] 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 MM*, 2018. 2, 3

[31] Matthew Q Hill, Stephan Streuber, Carina A Hahn, Michael J Black, and Alice J O'Toole. Creating body shapes from verbal descriptions by linking similarity spaces. *Psychological science*, 2016. 1, 5

[32] Wei-Lin Hsiao and Kristen Grauman. Learning the latent "look": Unsupervised discovery of a style-coherent embedding from fashion images. In *ICCV*, 2017. 1, 2

[33] Wei-Lin Hsiao and Kristen Grauman. Creating capsule wardrobes from fashion images. In *CVPR*, 2018. 1, 2

[34] Wei-Lin Hsiao, Isay Katsman, Chao-Yuan Wu, Devi Parikh, and Kristen Grauman. Fashion++: Minimal edits for outfit improvement. In *ICCV*, 2019. 1, 2

[35] Yang Hu, Xi Yi, and Larry S. Davis. Collaborative fashion recommendation: A functional tensor factorization approach. In *ACM MM*, 2015. 2

[36] C. Huynh, A. Ciptadi, A. Tyagi, and A. Agrawal. Craft: Complementary recommendation by adversarial feature transform. In *ECCV Workshop on Computer Vision For Fashion, Art and Design*, 2018. 2

[37] Moon-Hwan Jeong, Dong-Hoon Han, and Hyeong-Seok Ko. Garment capture from a photograph. *Computer Animation and Virtual Worlds*, 2015. 2

[38] Angjoo Kanazawa, Michael J. Black, David W. Jacobs, and Jitendra Malik. End-to-end recovery of human shape and pose. In *CVPR*, 2018. 2, 5

[39] Wang-Cheng Kang, Chen Fang, Zhaowen Wang, and Julian McAuley. Visually-aware fashion recommendation and design with generative image models. In *ICDM*, 2017. 2

[40] Wang-Cheng Kang, Eric Kim, Jure Leskovec, Charles Rosenberg, and Julian McAuley. Complete the look: Scene-based complementary product recommendation. In *CVPR*, 2019. 1, 2

[41] Nour Karessli, Romain Guigourès, and Reza Shirvany. Sizenet: Weakly supervised learning of visual size and fit in fashion images. In *CVPR Workshop on FFSS-USAD*, 2019. 2

[42] Hirokatsu Kataoka, Yutaka Satoh, Kaori Abe, Munetaka Minoguchi, and Akio Nakamura. Ten-million-order human database for world-wide fashion culture analysis. In *CVPR Workshop on FFSS-USAD*, 2019. 2, 3

[43] M. Hadi Kiapour, K. Yamaguchi, A. Berg, and T. Berg. Hipster wars: Discovering elements of fashion styles. In *ECCV*, 2014. 2, 3

[44] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015. 6

[45] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 2009. 6

[46] A. Kovashka, D. Parikh, and K. Grauman. WhittleSearch: Interactive image search with relative attribute feedback. *IJCV*, 2015. 1, 2

[47] Zhanghui Kuang, Yiming Gao, Guanbin Li, Ping Luo, Yimin Chen, Liang Lin, and Wayne Zhang. Fashion retrieval via graph reasoning networks on a similarity pyramid. *ICCV*, 2019. 1, 2

[48] Zorah Lahner, Daniel Cremers, and Tony Tung. Deepwrinkles: Accurate and realistic clothing modeling. In *ECCV*, 2018. 2

[49] Verica Lazova, Eldar Insafutdinov, and Gerard Pons-Moll. 360-degree textures of people in clothing from a single image. *arXiv preprint arXiv:1908.07117*, 2019. 2

[50] Yuncheng Li, Liangliang Cao, Jiang Zhu, and Jiebo Luo. Mining fashion outfit composition using an end-to-end deep learning approach on set data. *Transactions on Multimedia*, 2017. 2

[51] Xiaodan Liang, Si Liu, Xiaohui Shen, Jianchao Yang, Luoqi Liu, Jian Dong, Liang Lin, and Shuicheng Yan. Deep human parsing with active template regression. *TPAMI*, 2015. 1, 2, 3

[52] Si Liu, Jiashi Feng, Csaba Domokos, Hui Xu, Junshi Huang, Zhenzhen Hu, and Shuicheng Yan. Fashion parsing with weak color-category labels. *Transactions on Multimedia*, 2013. 2

[53] S. Liu, J. Feng, Z. Song, T. Zheng, H. Lu, C. Xu, and S. Yan. Hi, magic closet, tell me what to wear! In *ACM MM*, 2012. 1, 2

[54] Si Liu, Zheng Song, Guangcan Liu, Changsheng Xu, Hanqing Lu, and Shuicheng Yan. Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set. In *CVPR*, 2012. 1, 2

[55] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *CVPR*, 2016. 1, 2, 3

[56] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multi-person linear model. *TOG*, 2015. 1, 5

[57] Utkarsh Mall, Kevin Matzen, Bharath Hariharan, Noah Snavely, and Kavita Bala. GeoStyle: Discovering fashion trends and events. In *ICCV*, 2019. 1, 2, 3

[58] Rishabh Misra, Mengting Wan, and Julian McAuley. Decomposing fit semantics for product size recommendation in metric spaces. In *RecSys*, 2018. 2

[59] Ryota Natsume, Shunsuke Saito, Zeng Huang, Weikai Chen, Chongyang Ma, Hao Li, and Shigeo Morishima. Siclope: Silhouette-based clothed people. In *CVPR*, 2019. 2

[60] Vitali Petsiuk, Abir Das, and Kate Saenko. Rise: Randomized input sampling for explanation of black-box models. In *BMVC*, 2018. 6

[61] Gina Pisut and Lenda Jo Connell. Fit preferences of female consumers in the usa. *Journal of Fashion Marketing and Management: An International Journal*, 2007. 1

[62] Gerard Pons-Moll, Sergi Pujades, Sonny Hu, and Michael Black. Clothcap: Seamless 4d clothing capture and retargeting. *TOG*, 2017. 2

[63] Amit Raj, Patsorn Sangkloy, Huiwen Chang, James Hays, Duygu Ceylan, and Jingwan Lu. Swapnet: Image based garment transfer. In *ECCV*, 2018. 1, 2

[64] Consumer Reports. Why don't these pants fit?, 1996. 3

[65] Kathleen M Robinette, Hans Daanen, and Eric Paquet. The caesar project: a 3-d surface anthropometry survey. In *The International Conference on 3-D Digital Imaging and Modeling*. IEEE, 1999. 3

[66] Negar Rostamzadeh, Seyedarian Hosseini, Thomas Boquet, Wojciech Stokowiec, Ying Zhang, Christian Jauvin, and Chris Pal. Fashion-gen: The generative fashion dataset and challenge. *arXiv preprint arXiv:1806.08317*, 2018. 1, 2, 3

[67] Shunsuke Saito, Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, and Hao Li. Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *ICCV*, 2019. 2

[68] Igor Santesteban, Miguel A Otaduy, and Dan Casas. Learning-based animation of clothing for virtual try-on. In *Computer Graphics Forum*, 2019. 2

[69] Hosnieh Sattar, Gerard Pons-Moll, and Mario Fritz. Fashion is taking shape: Understanding clothing preference based on body shape from online sources. In *WACV*, 2019. 1, 2

[70] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *CVPR*, 2015. 4

[71] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *CVPR*, 2017. 6

[72] Abdul-Saboor Sheikh, Romain Guigourès, Evgenii Koriagin, Yuen King Ho, Reza Shirvany, Roland Vollgraf, and Urs Bergmann. A deep learning system for predicting size and fit in fashion e-commerce. In *RecSys*, 2019. 2

[73] Yong-Siang Shih, Kai-Yueh Chang, Hsuan-Tien Lin, and Min Sun. Compatibility family learning for item recommendation and generation. In *AAAI*, 2018. 1, 2

[74] Edgar Simo-Serra, Sanja Fidler, Francesc Moreno-Noguer, and Raquel Urtasun. Neuroaesthetics in Fashion: Modeling the Perception of Fashionability. In *CVPR*, 2015. 2, 3

[75] Stephan Streuber, M Alejandra Quiros-Ramirez, Matthew Q Hill, Carina A Hahn, Silvia Zuffi, Alice O'Toole, and Michael J Black. Body talk: crowdshaping realistic 3d avatars with words. *TOG*, 2016. 5, 8

[76] Mariya I. Vasileva, Bryan A. Plummer, Krishna Dusad, Shreya Rajpal, Ranjitha Kumar, and David Forsyth. Learning type-aware embeddings for fashion compatibility. In *ECCV*, 2018. 1, 2, 6

[77] Andreas Veit, Balazs Kovacs, Sean Bell, Julian McAuley, Kavita Bala, and Serge Belongie. Learning visual clothing style with heterogeneous dyadic co-occurrences. In *ICCV*, 2015. 1, 2, 6

[78] K. Wagstaff, C. Cardie, S. Rogers, and S. Schroedl. Constrained K-means Clustering with Background Knowledge. In *ICML*, 2001. 4

[79] Bochao Wang, Huabin Zheng, Xiaodan Liang, Yimin Chen, Liang Lin, and Meng Yang. Toward characteristic-preserving image-based virtual try-on network. In *ECCV*, 2018. 1, 2

[80] A. Williams. Fit of clothing related to body-image, body built and selected clothing attitudes. In *Unpublished doctoral dissertation*, 1974. 2

[81] Chao-Yuan Wu, R Manmatha, Alexander J Smola, and Philipp Krahenbuhl. Sampling matters in deep embedding learning. In *ICCV*, 2017. 4

[82] Yi Xu, Shanglin Yang, Wei Sun, Li Tan, Kefeng Li, and Hui Zhou. 3d virtual garment modeling from rgb images. *arXiv preprint arXiv:1908.00114*, 2019. 2

[83] Kota Yamaguchi, M Hadi Kiapour, and Tamara L Berg. Paper doll parsing: Retrieving similar styles to parse clothing items. In *ICCV*, 2013. 2

[84] Kota Yamaguchi, Hadi Kiapour, Luis Ortiz, and Tamara Berg. Parsing clothing in fashion photographs. In *CVPR*, 2012. 1, 2, 3

[85] Shan Yang, Zherong Pan, Tanya Amert, Ke Wang, Licheng Yu, Tamara Berg, and Ming C. Lin. Physics-inspired garment recovery from a single-view image. *TOG*, 2018. 2

[86] B. Zhao, J. Feng, X. Wu, and S. Yan. Memory-augmented attribute manipulation networks for interactive fashion search. In *CVPR*, 2017. 2

[87] Shuai Zheng, Fan Yang, M Hadi Kiapour, and Robinson Piramuthu. Modanet: A large-scale street fashion dataset with polygon annotations. In *ACM MM*, 2018. 2

[88] Bin Zhou, Xiaowu Chen, Qiang Fu, Kan Guo, and Ping Tan. Garment modeling from a single image. In *Computer graphics forum*, 2013. 2

[89] Hao Zhu, Xinxin Zuo, Sen Wang, Xun Cao, and Ruigang Yang. Detailed human shape estimation from a single image by hierarchical mesh deformation. In *CVPR*, 2019. 2, 5