FASHION MEETS COMPUTER VISION: A SURVEY

WEN-HUANG CHENG, National Chiao Tung University and National Chung Hsing University SIJIE SONG, Peking University
CHIEH-YUN CHEN, National Chiao Tung University
SHINTAMI CHUSNUL HIDAYATI, Institut Teknologi Sepuluh Nopember

JIAYING LIU*, Peking University

Presented by: Sini Suresh
24.03.2021

OVERVIEW

- The paper terms computer-vision-enabled fashion technology as intelligent fashion
- Fashion tasks:
  - Low level - pixel computation
  - Mid-level - fashion understanding
  - High-level - fashion analysis

- Fashion detection
- Fashion Analysis
- Fashion Synthesis
- Fashion Recommendation

Fig. 1. Scope of the intelligent fashion research topics covered in this survey paper.

1. FASHION DETECTION – LANDMARK DETECTION

- **Aims to predict the positions of functional keypoints defined on the clothes**
- Features extracted from the landmarks greatly facilitate fashion image analysis.
- Difference between landmark detection and human pose estimation

State of the art methods:

- Deep fashion alignment (DFA) framework
- Deep LAndmark Network (DLAN)
- Predict a confidence map of positional distributions (*i.e.*, heatmap) for each landmark.
  - For instance, "*left collar* $\boxed{\leftrightarrow}$ *left waistline* $\boxed{\leftrightarrow}$ *left hemline*"
- **Performance metrics** - Normalized error (NE) defined as the l2 distance between detected and the ground truth landmarks in the normalized coordinate space.

1. FASHION DETECTION – FASHION PARSING

- **Specific form of semantic segmentation**

State of the art methods:

- Combined the human pose estimation module and (super)pixel-level category classifier learning module to generate category classifiers
- Deformable Mixture Parsing Model (DMPM)
- CNN-based approaches capture the complex correlations between clothing appearance and structure
- Hierarchical graph was considered for human parsing tasks
- **Performance metrics -**
  - Average Pixel Accuracy (aPA) - proportion of correctly labeled pixels in the whole image, Mean Average Garment Recall (mAGR)
  - Intersection over Union (IoU), Foreground accuracy
  - Mean accuracy, average precision, average recall, average F-1 score over pixels

Fig. 3. Examples of semantic segmentation for fashion images [78].

1. FASHION DETECTION – ITEM RETRIEVAL

- **Given a fashion image query, the goal of image-based fashion item retrieval is to find similar or identical items from the gallery.**

State of the art methods:

- Unsupervised transfer learning
- Hand-crafted features
- Dual Attribute- aware Ranking Network (DARN)
- Two deep learning baseline methods, and one method aimed to learn the similarity between two different domains, street and shop domains.
- Bi-directional problem, street-to-shop and shop-to-street clothing retrieval task - deep bi-directional cross-triplet embedding algorithm
- Graph Reasoning Network
- FashionSearchNet
- Unsupervised embedding learning to train a CNN model and combined the existing retrieval methods trained on different datasets to finetune the retrieval results.
- **Performance metrics -**
  - Top-*k* retrieval accuracy,
  - Precision@*k* , Recall@*k*
  - Normalized Discounted Cumulative Gain (NDCG@*k*)
  - Mean Average Precision (MAP)

2. FASHION ANALYSIS – ATTRIBUTE RECOGNITION

- **Clothing attribute recognition is a multi-label classification problem**

State of the art methods:

- CRF-based model -
  - location-specific appearance with respect to a human body
  - compatibility of clothing items and attributes
  - trained using a max-margin learning framework
- Cross domain attribute mining
- Other methods based on:
  - attributes and the divergence of neural activations in the deep network
  - two independent deep models to perform attribute recognition
  - spatial-semantic representations for each attribute

ATTRIBUTE RECOGNITION

BENCHMARK DATASETS

Table 5. Summary of the benchmark datasets for clothing attribute recognition task.

Dataset name		Publish time	# of photos	# of categories	# of attributes	Key features	Sources
Clothing Attributes [14]		2012	1,856	7	26	Annotated with 23 binary-class attributes and 3 multi-class attributes	Thesartorialist.com, Flickr.com
Hidayati *et al.* [56]		2012	1,077	8	5	Annotated with clothing categories	Online shopping sites
UT-Zap50K shoe [218]		2014	50,025	N/A	4	Shoe images annotated with associated metadata (shoe type, materials, gender, manufacturer, *etc.*)	Zappos.com
Chen *et al.* [20]	Online-data	2015	341,021	15	67	Each attribute has 1000+ images	Online shopping sites
	Street-data-a		685	N/A	N/A	Annotated with fine-grained attributes	Fashionista [206]
	Street-data-b		8,000	N/A	N/A		Parsing [31]
	Street-data-c		4,200	N/A	N/A		Surveillance videos
Lu *et al.* [170]		2016	~1,1 M	N/A	N/A	Annotated with the associated tags	Taobao.com
WIDER Attribute [101]		2016	13,789	N/A	N/A	Annotated with 14 human attribute labels and 30 event class labels	The 50574 WIDER images [200]
Vittayakorn *et al.* [177]	Etsy	2016	173,175	N/A	250	Annotated with title and description of the product	Etsy.com
	Wear		212,129	N/A	250	Annotated with the associated tags	Wear.jp
DeepFashion-C [123]		2016	289,222	50	1,000	Annotated with clothing bounding box, type, category, and attributes	Online shopping sites, Google Images
Fashion200K [48]		2017	209,544	5	4,404	Annotated with product descriptions	Lyst.com
Hidayati *et al.* [61]		2018	3,250	16	12	Annotated with clothing categories	Online shopping sites
CatalogFashion-10x [8]		2019	1M	43	N/A	The categories are identical to the DeepFashion dataset [123]	Amazon.com

N/A: there is no reported information to cite

PERFORMANCE METRICS

- **Top-k accuracy**
- MAP – Mean Average Precision
- Geometric Mean

2. FASHION ANALYSIS – STYLE LEARNING

- **Challenges**:
  - how to analyze discriminative features for different styles
  - learn what style makes a trend

State-of-the-art methods

- Categories explored - hipster, bohemian, pinup, preppy, and goth.
- Joint ranking and classification framework based on the Siamese network
- Bimodal Correlative Deep Autoencoder (BCDA)
- Data-driven fashion model - adapted a natural language processing technique to learn latent fashion concepts jointly over the style and element vocabularies
- User-centric fashion information based on occasions, clothing categories, and attributes
- **Fashion trends analysis -** New York Fashion Week, Seasons, Cities, Social Events, Temporal estimation
- **Performance metrics -** precision, recall & accuracy

2. FASHION ANALYSIS – POPULARITY PREDICTION

- **Essential for both Fashion brands and Individuals**

State-of-the-art methods

- Vision based approach - quantitatively evaluate the influence of visual, textual, and social factors
- Aesthetic quality assessment
- Impact of facial detail, lighting, and color
- Fashion forecasting
  - learning a representation of fashion images
  - discovering the set of fine-grained styles
  - constructing styles' temporal trajectories
- **Performance metrics -** Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Mean Square Error (MSE), and Spearman Ranking Correlation (SRC)

3. FASHION SYNTHESIS – STYLE TRANSFER

- **Transferring an input image into a corresponding output image**

State-of-the-art methods

- Pix2pix – popular style transfer work
- **Facial make-up:**
  - image processing methods >> deep learning framework >> deep neural network based makeup recommendation model
- Challenge – facial recognition apps
- Framework to train the makeup transfer and removal networks together – BeautyGAN, BeautyGlow
- **Virtual try-on:** coarse-to-fine strategy
  - VIrtual Try-On Network (VITON) framework
  - Characteristic-Preserving Image-based Virtual Try-On Network
  - FashionGAN & M2E-TON
- **Performance metrics -** user study, inception score (IS) or structural similarity (SSIM)

3. FASHION SYNTHESIS – POSE TRANSFORMATION

Challenge - input and output are not spatially aligned.

State-of-the-art methods

- 4 modules of the framework:
  - source image segmentation,
  - spatial transformation,
  - foreground synthesis,
  - background synthesis
- **Performance metrics -**
  - inception score (IS) or structural similarity (SSIM)
  - mask-IS and mask-SSIM – to eliminate background effect

3. FASHION SYNTHESIS – PHYSICAL SIMULATION

Physical simulation works based on 3D information

State-of-the-art methods

- Pose-dependent model
- ClothCap - multi-part 3D model
- SimulCap -
  - multi-layer avatar generation
  - body tracking and cloth tracking for simulating the physical clothing-body interactions
- **Performance metrics -** vision comparison with state-of-the-art methods:

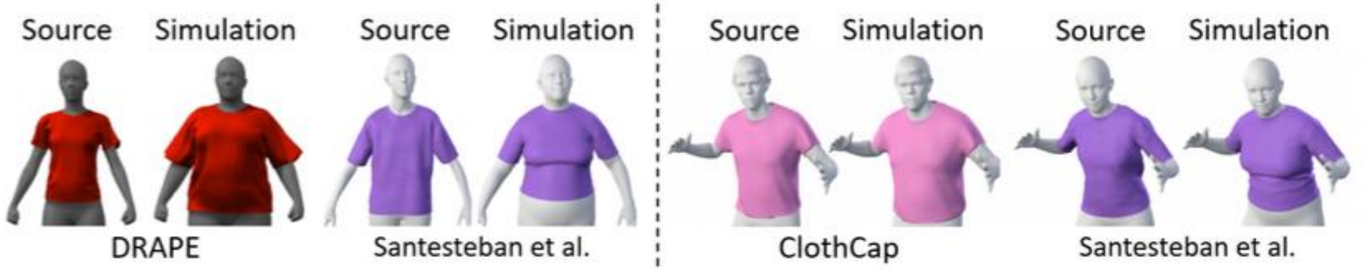

Fig. 9. Left comparison is between DRAPE [44] and Santesteban *et al.* [156], while the right one compares between ClothCap [144] and [156]. Both are given a source and simulate the physical clothing deformation in different body shapes.

4. FASHION RECOMMENDATION – FASHION COMPATIBILITY

How well items of different types can collaborate to form fashionable outfits!

Visual similarity and visual compatibility between different clothing types.

State-of-the-art methods

- Conditional Similarity Networks
- Integrated visual and contextual modalities of fashion items
- Integrated fashion domain knowledge
- Other addition – image embedding, body shape calculator
- Information about which item made the outfit incompatible.
- General compatibility modeling and personal preference modeling,
- **Performance metrics -**
  - AUC - area under the receiver operating characteristic curve - most used metric
  - AUC scores range between 0 and 1

4. FASHION RECOMMENDATION – OUTFIT MATCHING

A key to a stylish outfit lies in the matching fashion items !

Challenges:

- fashion concept is subtle and subjective
- large number of attributes for describing fashion
- notion of fashion item compatibility goes across categories and involves complex relationships

State-of-the-art methods

- Probabilistic topic model
- Occasion-oriented clothing recommendation
  - *wearing properly* and *wearing aesthetically* principles
- Co-purchase behaviour, top-bottom matching along with accesories
- IStylist - personalized clothing recommendation system.
- **Performance metrics -**
  - AUC
  - FITB (fill in the blank) accuracy
  - no unified benchmark for outfit matching

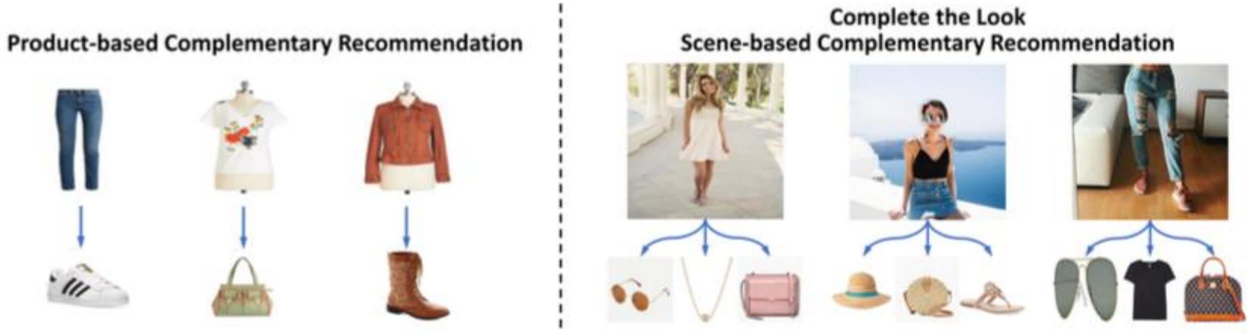

Fig. 11. A comparison between product-based and scene-based complementary recommendation [87].

4. FASHION RECOMMENDATION – HAIRSTYLE SUGGESTION

The right hairstyle can enhance the best facial features while concealing the flaws, bringing out natural beauty and style.

State-of-the-art methods

- By learning the relationship between facial shapes and successful hairstyle examples
- Beauty e-Experts system
- **Performance metrics -**
  - User study
  - NDCG - measures how close the ranking of the top-*k* recommended styles is to the optimal ranking

APPLICATIONS

- AI Chatbots - *Dior Insider, Levi's Virtual Stylist, Nike on Demand*
- Improving product discovery
- Tailor recommendation – *Stitch Fix*
- Reducing product return
- Powering productivity and creativity
- Development of cost-effective annotations approach on fashion and beauty related data are necessary
- **Challenges:**
  - System performance
  - Handling data bias
  - Variations to improve the true-positive rate while maintaining a low false-positive rate

THANK YOU!