

# When Computer Vision Meets Fashion

---

Sijie Song

2018.5.31

# Main Topics in Fashion

---

## High-Level Application

Fashion Synthesis      Fashion Analysis

Recommendation

Outfit Matching      Retrieval

Style Classification

## Low-Level Understanding

Fashion Parsing

Landmark Detection

Attribute Prediction

# Researchers

---

- **Xintong Han**, Maryland University, <http://www.umiacs.umd.edu/~xintong>
- **Kota Yamaguchi**, CyberAgent Inc, <http://vision.is.tohoku.ac.jp/~kyamagu/>
- **Kristen Grauman**, University of Texas at Austin, <http://www.cs.utexas.edu/users/grauman/>
- **Edgar Simo-Serra**, Waseda University, <http://www.iri.upc.edu/people/esimo/>
- **Jia Jia**, Tsinghua University, <http://hcsi.cs.tsinghua.edu.cn/JiaJia.html>
- **Si Liu**, CAS, <http://liusi-group.com/>

# Main Topics in Fashion

## High-Level Application

Fashion Synthesis      Fashion Analysis

Recommendation

Outfit Matching      Retrieval

Style Classification

## Low-Level Understanding

Fashion Parsing

Landmark Detection

Attribute Prediction

# Fashion Parsing

---

- Pixel-level label generation

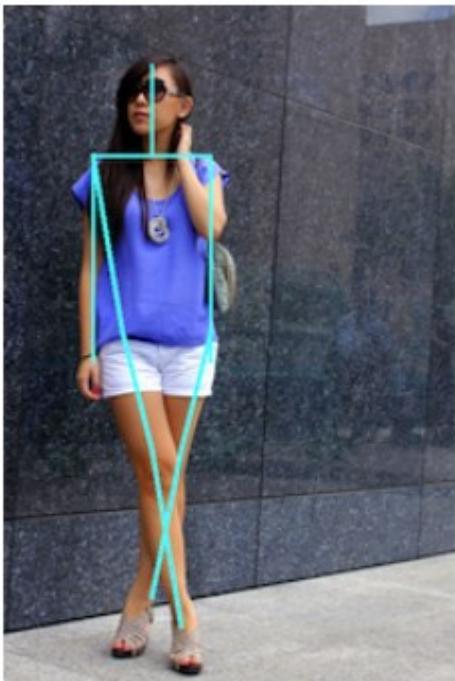


# Fashion Parsing

- **Parsing Clothing in Fashion Photographs (Kota Yamaguchi, M Hadi Kiapour, Luis E Ortiz, Tamara L Berg, CVPR 2012)**



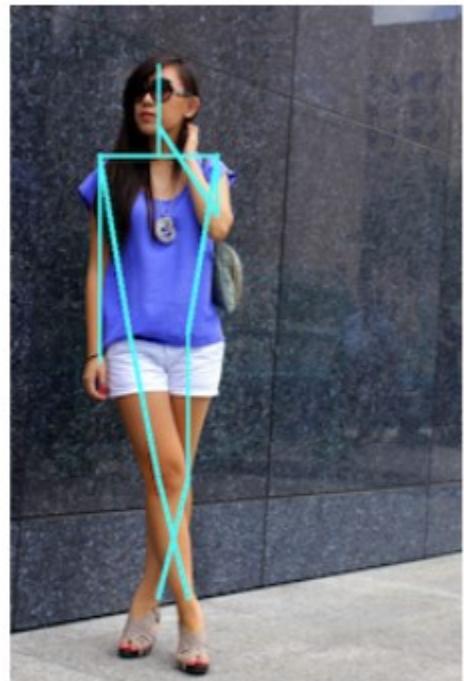
(a) Superpixels



(b) Pose estimation



(c) Predicted Clothing Parse

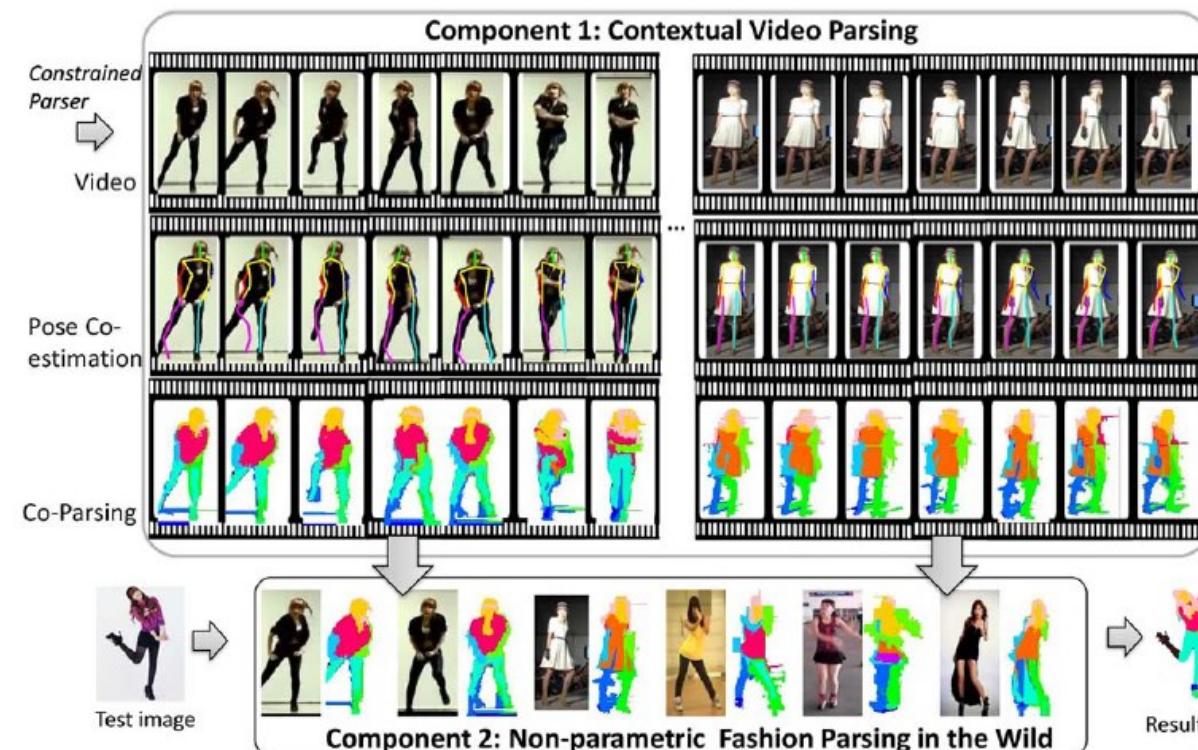


(d) Pose re-estimation

□	null
■	shorts
■	shoes
■	purse
■	top
■	necklace
■	hair
■	skin

# Fashion Parsing

- Fashion Parsing with Video Context ( Si Liu, Xiaodan Liang, Luoqi Liu, Liang Lin, Ke Lv, Shuicheng Yan, ACM MM, 2014)



# Main Topics in Fashion

## High-Level Application

Fashion Synthesis      Fashion Analysis

Recommendation

Outfit Matching      Retrieval

Style Classification

## Low-Level Understanding

Fashion Parsing

Landmark Detection

Attribute Prediction

# Landmark Detection

---

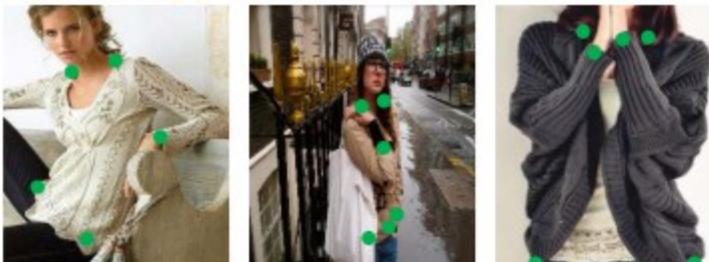
- Different from human pose joints
  - **Fashion Landmark in the Wild (Ziwei Liu, Sijie Yan, Ping Luo, Xiaogang Wang, Xiaou Tang, ECCV 2016)**



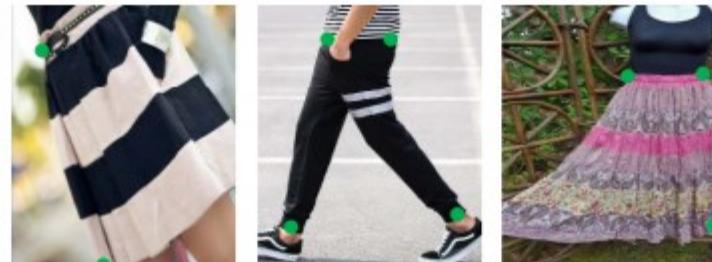
# Landmark Detection

- Upper body: left/right collar end, left/right sleeve end, left/right hem
- Lower body: left/right hem, left/right waist
- Full body: left/right collar end, left/right sleeve end, left/right waist, left/right hem

**Upper-body Clothes**



**Lower-body Clothes**



**Full-body Clothes**



# Main Topics in Fashion

## High-Level Application

Fashion Synthesis      Fashion Analysis

Recommendation

Outfit Matching      Retrieval

Style Classification

## Low-Level Understanding

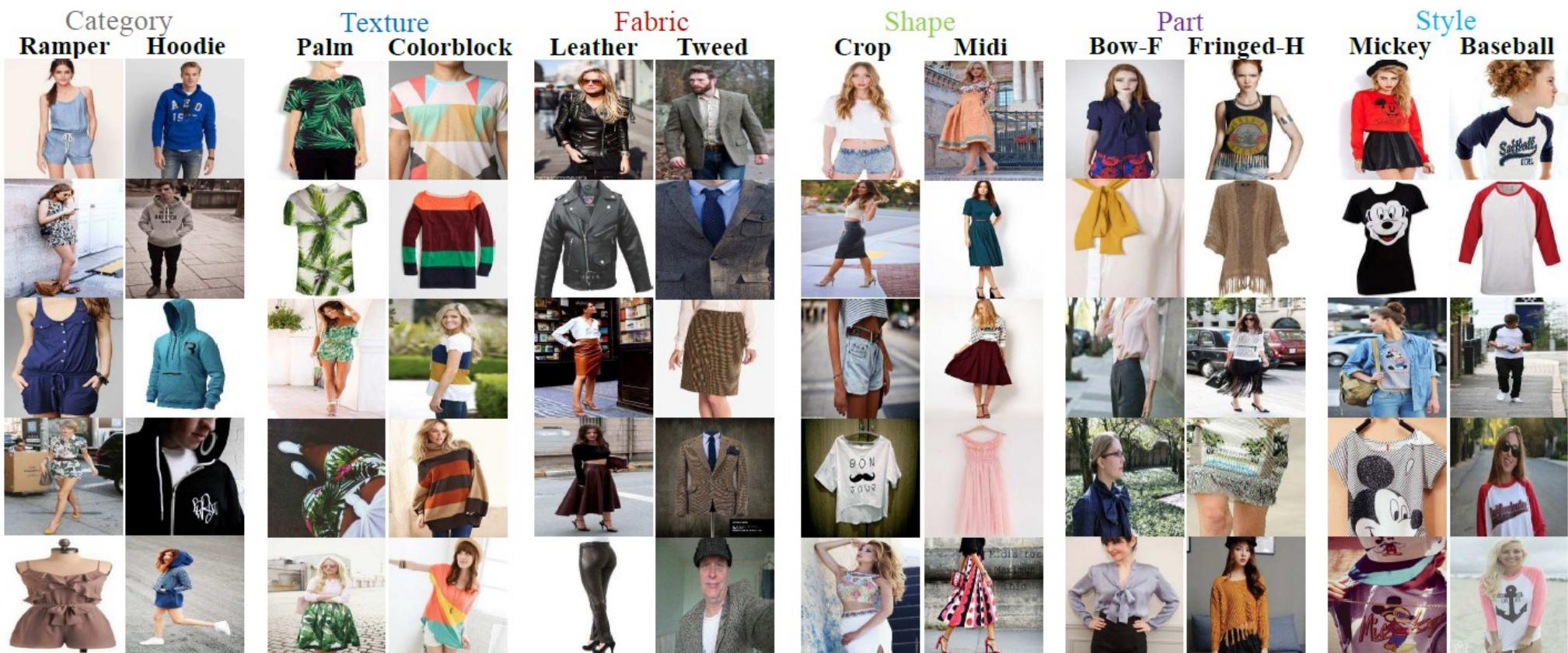
Fashion Parsing

Landmark Detection

Attribute Prediction

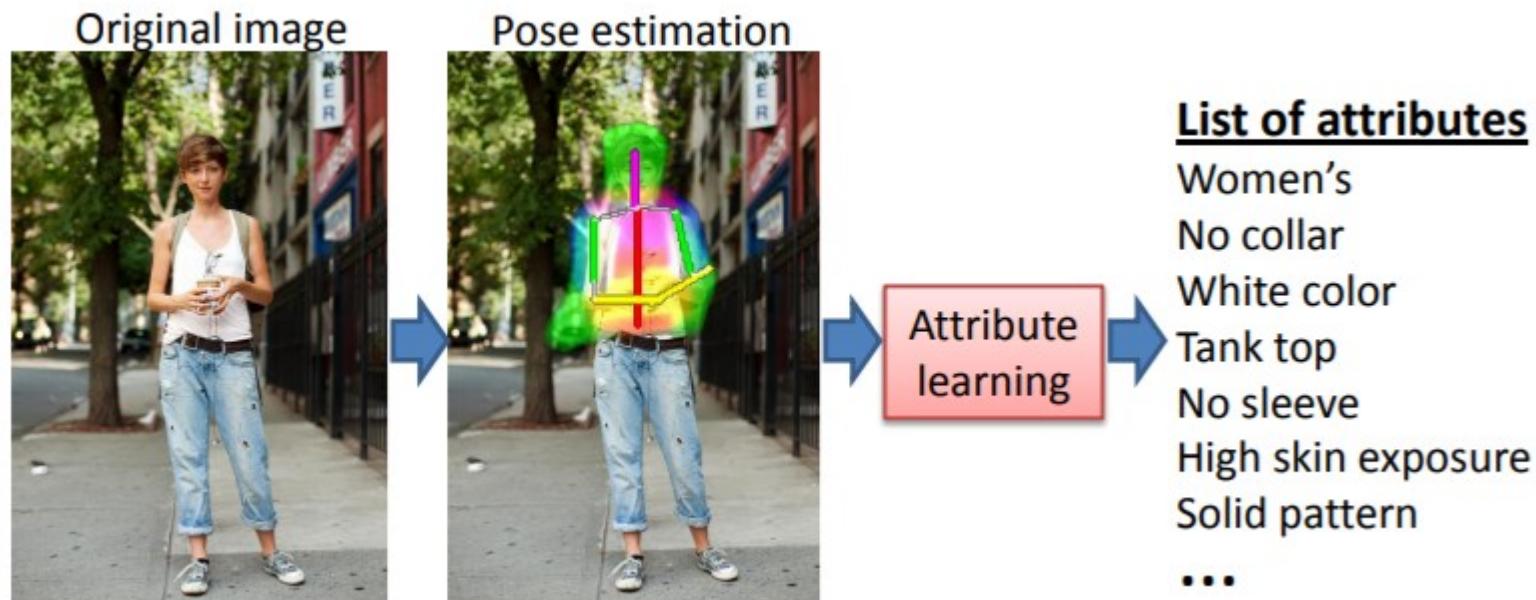
# Attribute Prediction

- Pre-defined attributes



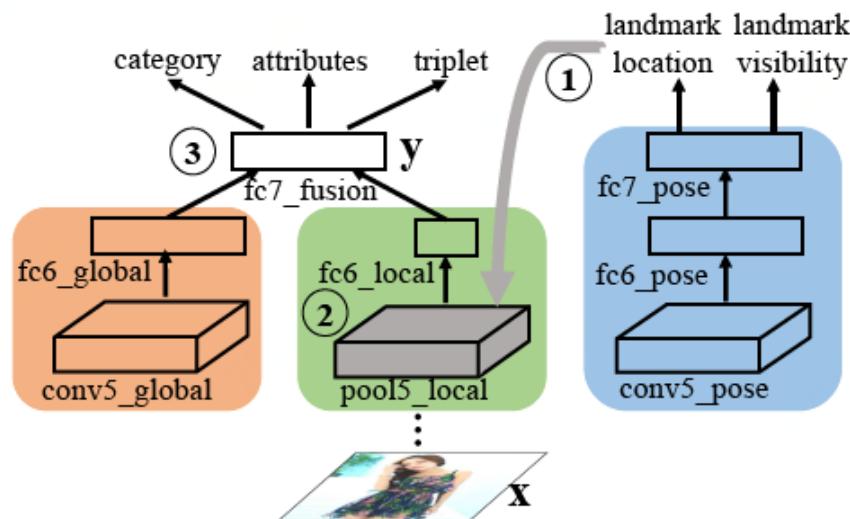
# Attribute Prediction

- Pre-defined attributes
  - Describing Clothing by Semantic Attributes (Huizhong Chen, Andrew Gallagher, and Bernd Girod, ECCV 2012)

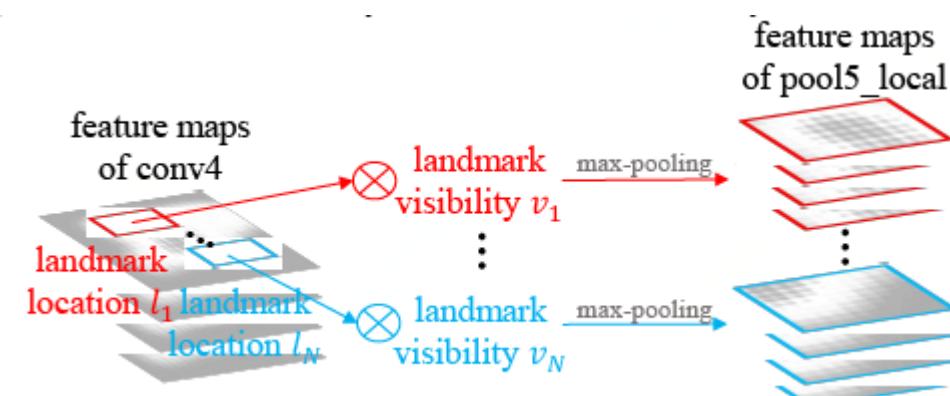


# Attribute Prediction

- Pre-defined attributes
  - DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations (Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang, CVPR 2016)



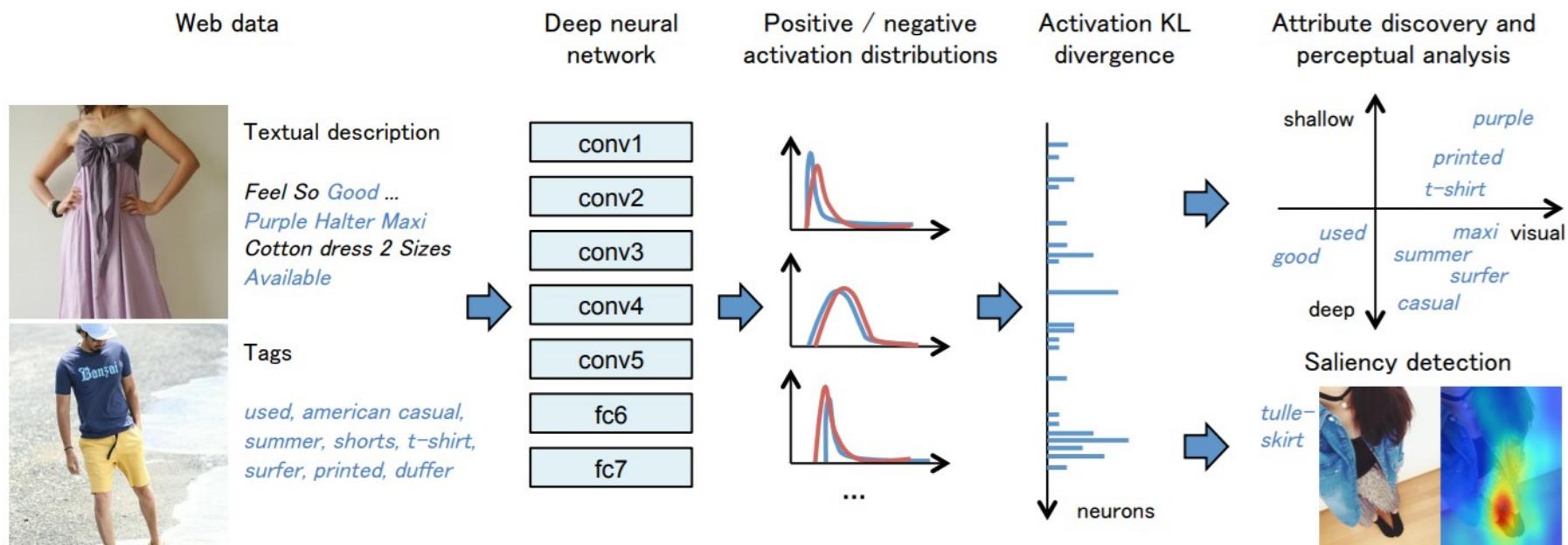
(a) Fashion-Net



(b) Landmark pooling

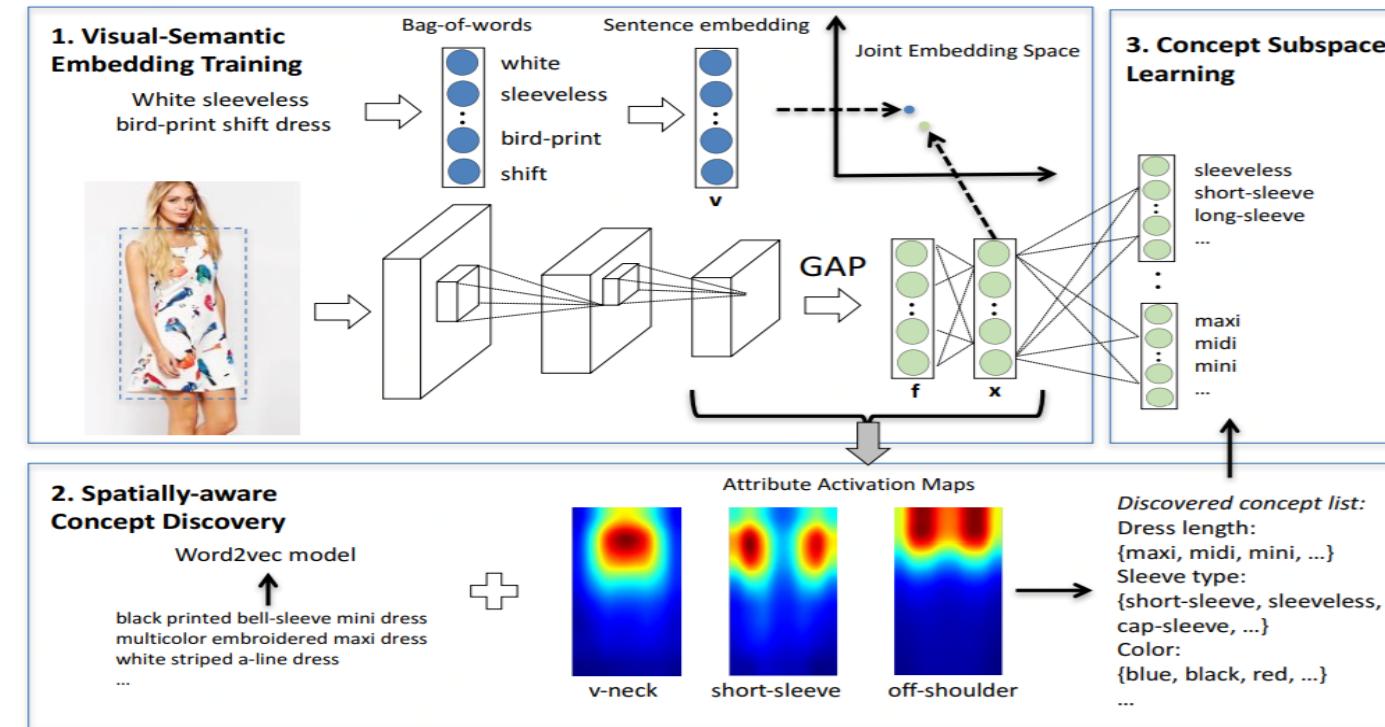
# Attribute Prediction

- Auto attribute discovery
  - Automatic Attribute Discovery with Neural Activations (Sirion Vittayakorn, Takayuki Umeda, Kazuhiko Murasaki, Kyoko Sudo, Takayuki Okatani, Kota Yamaguchi, ECCV 2016)



# Attribute Prediction

- Auto attribute discovery
  - Automatic Spatially-aware Fashion Concept Discovery (Xintong Han, Zuxuan Wu, Phoenix X. Huang, Xiao Zhang, Menglong Zhu, Yuan Li, Yang Zhao, Larry S. Davis, University of Maryland, Snap, Google, ICCV 2017)



# Main Topics in Fashion

## High-Level Application

Fashion Synthesis      Fashion Analysis

Recommendation

Outfit Matching      Retrieval

Style Classification

## Low-Level Understanding

Fashion Parsing

Landmark Detection

Attribute Prediction

# Style Recognition

---

- What clothing reveals about personal style
- Predefined styles: *Preppy, Bohemian, Pinup, Goth, Hipster*
- **Hipster Wars: Discovering Elements of Fashion Styles (M. Hadi Kiapour, Kota Yamaguchi, Tohoku University, Japan, ECCV 2014)**

Preppy      Bohemian

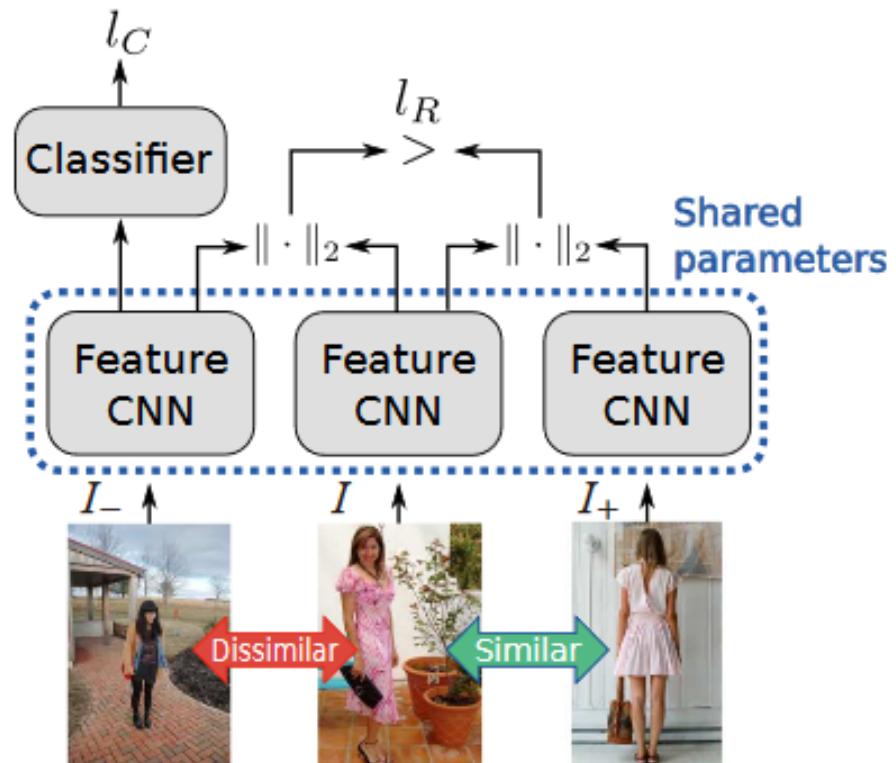


Hipster      Goth      Pinup



# Style Recognition

- **Fashion Style in 128 Floats: Joint Ranking and Classification using Weak Data for Feature Extraction (Edgar Simo-Serra and Hiroshi Ishikawa, CVPR2016)**
  - Siamese network for feature extraction
  - Classifier for style recognition



# Style Recognition

- Learning the Latent "Look": Unsupervised Discovery of a Style-Coherent Embedding from Fashion Images (Wei-Lin Hsiao, Kristen Grauman, ICCV2017)
  - Unsupervised style discovery with LDA model
  - “Word” → “Document” → “Topic” → “Attribute” → “Outfit” → “Style”

Upper	Lower	Hosiery	Global	
blue	short dr.	blue	floral	blue
purple	loose dr.		translucent	purple
white	flat dr.		dress	white
	white dr.		shoe	beige
	blue dr.		stocking	cardigan
	purple dr.		red	

 → Pinup

# Main Topics in Fashion

## High-Level Application

Fashion Synthesis

Fashion Analysis

Recommendation

Outfit Matching

Retrieval

Style Classification

## Low-Level Understanding

Fashion Parsing

Landmark Detection

Attribute Prediction

# Outfit Matching

- Pair-based matching: embedding

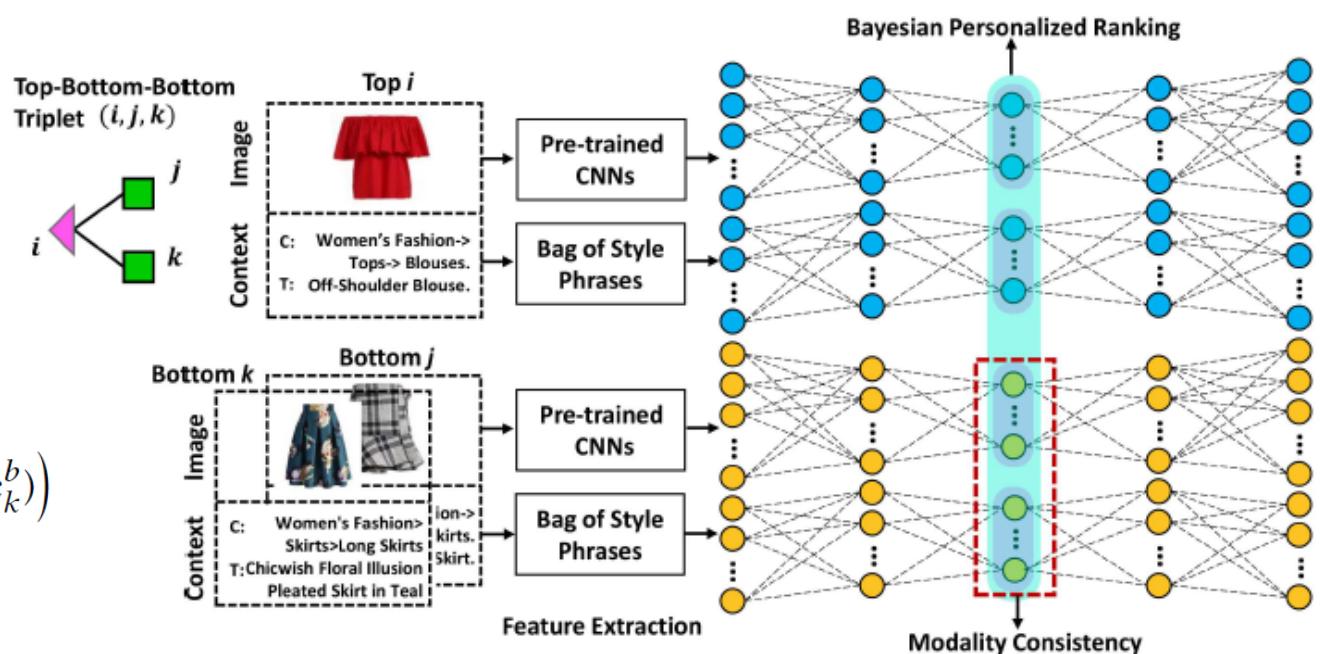
- NeuroStylist: Neural Compatibility Modeling for Clothing Matching (Xuemeng Song, Fuli Feng, Jinhuan Liu, Zekun Li, Liqiang Nie, Jun Ma, Shandong University, MM 2017)
- Bayesian personalized ranking

$$\mathcal{D}_S := \{(i, j, k) | t_i \in \mathcal{T}, b_j \in \mathcal{B}_i^+ \wedge b_k \in \mathcal{B} \setminus \mathcal{B}_i^+\}$$

$$\mathcal{L}_{bpr} = \sum_{(i, j, k) \in \mathcal{D}_S} -\ln(\sigma(m_{ij} - m_{ik}))$$

- Visual semantic embedding

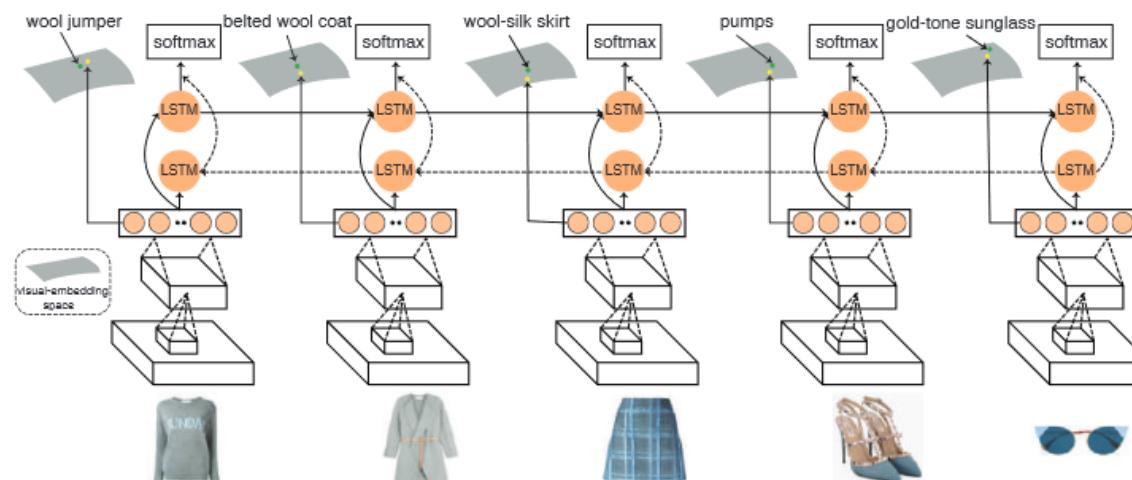
$$\sum_{(i, j, k) \in \mathcal{D}_S} \left( \mathcal{L}_{mod}(\tilde{\mathbf{v}}_i^t, \tilde{\mathbf{c}}_i^t) + \mathcal{L}_{mod}(\tilde{\mathbf{v}}_j^b, \tilde{\mathbf{c}}_j^b) + \mathcal{L}_{mod}(\tilde{\mathbf{v}}_k^b, \tilde{\mathbf{c}}_k^b) \right)$$



# Outfit Matching

- Set-based matching: total probability
  - Learning Fashion Compatibility with Bidirectional LSTMs (Xintong Han, Zuxuan Wu, Yu-Gang Jiang, Larry S. Davis, MM 2017)
  - Ideas borrowed from image caption

$$E_f(\mathbf{F}; \Theta_f) = -\frac{1}{N} \sum_{t=1}^N \log Pr(\mathbf{x}_{t+1} | \mathbf{x}_1, \dots, \mathbf{x}_t; \Theta_f),$$



# Main Topics in Fashion

## High-Level Application

Fashion Synthesis

Fashion Analysis

Recommendation

Outfit Matching

Retrieval

Style Classification

## Low-Level Understanding

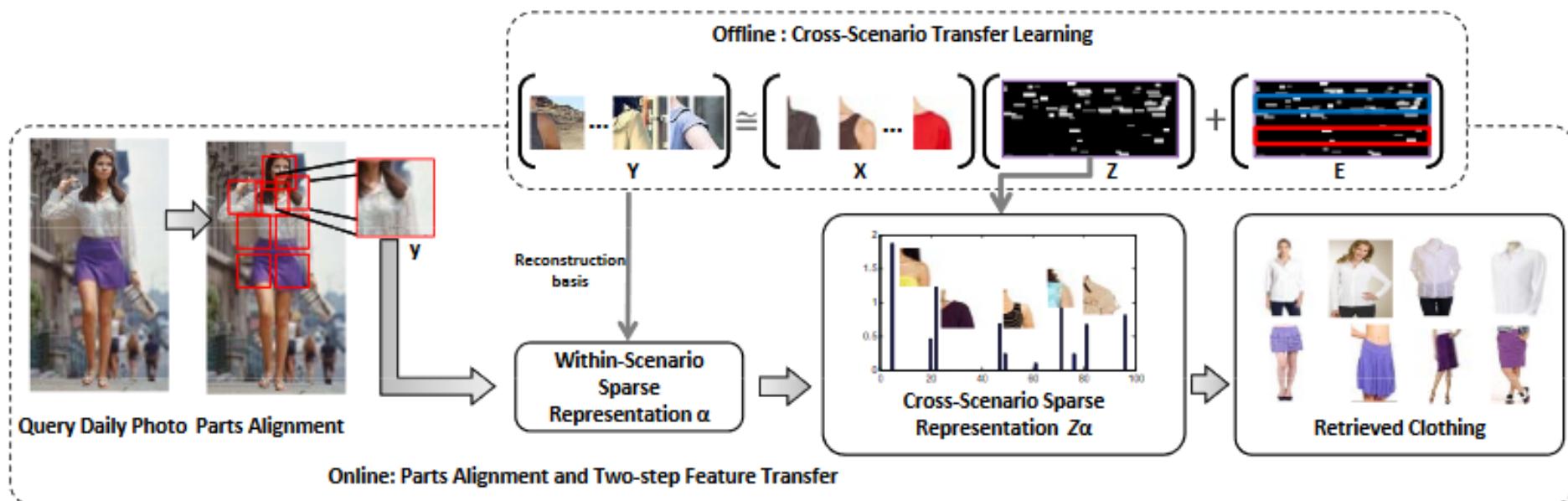
Fashion Parsing

Landmark Detection

Attribute Prediction

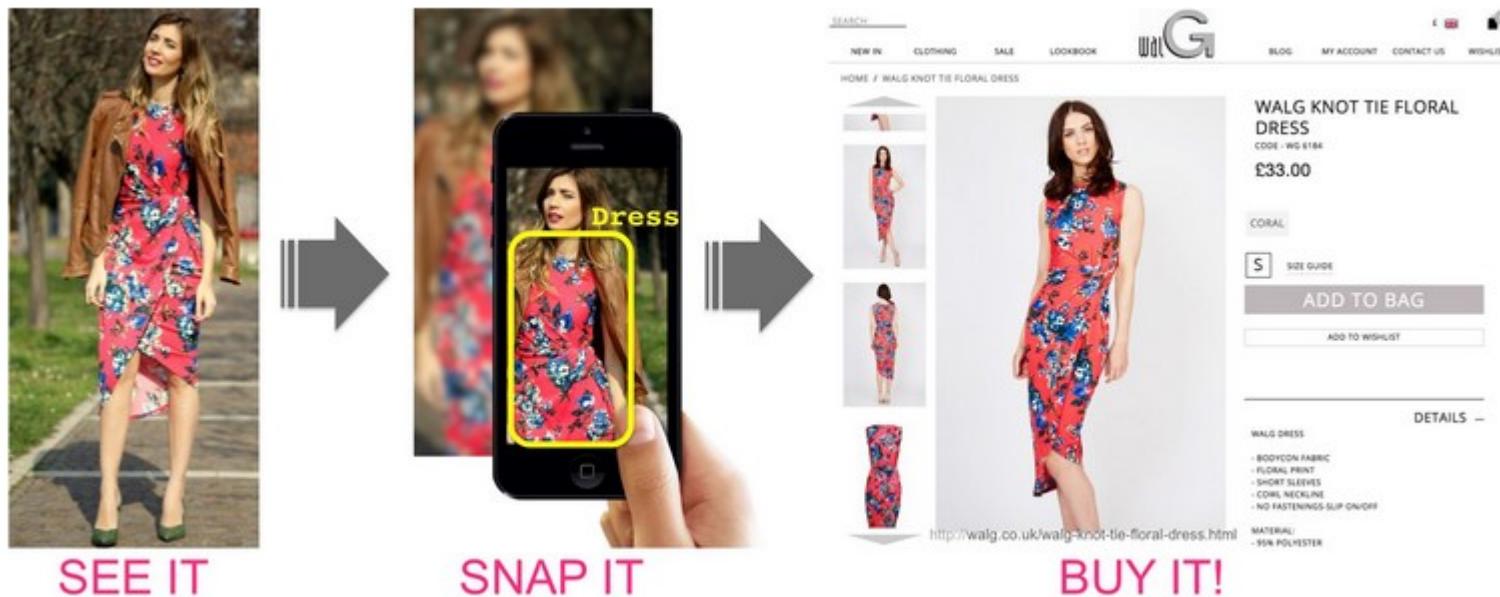
# Retrieval

- Similar retrieval
  - Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set (Si Liu, Zheng Song, Guangcan Liu, Changsheng Xu, Hanqing Lu, Shuicheng Yan, CVPR 2012)



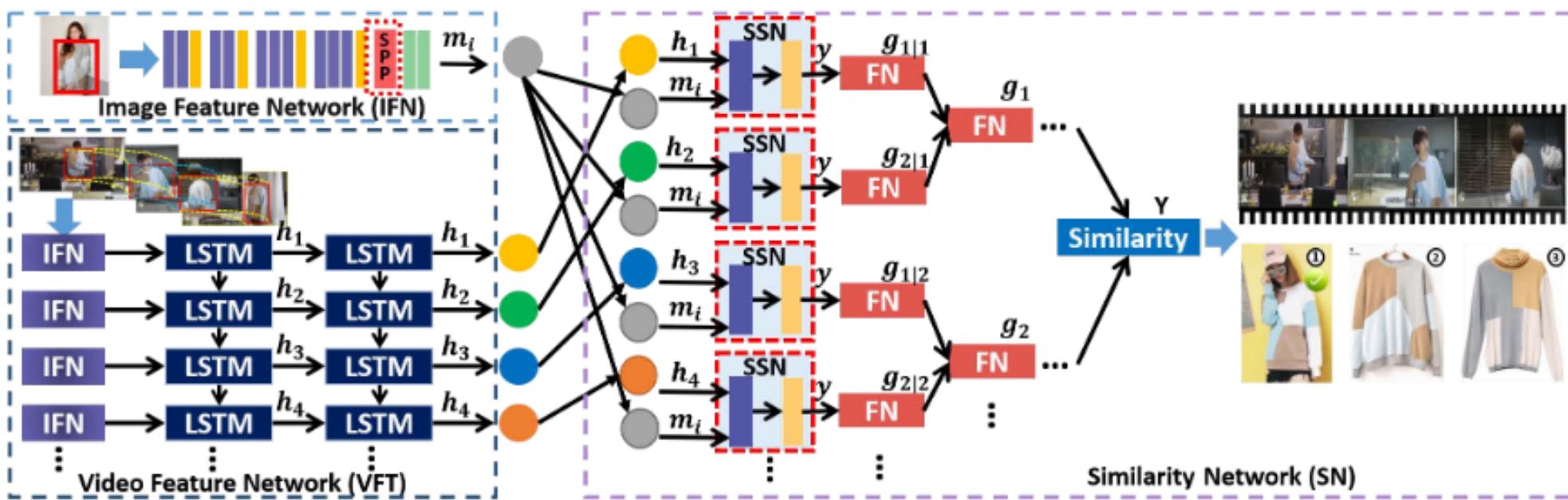
# Retrieval

- Exact retrieval: Image
  - Where to Buy It: Matching Street Clothing Photos in Online Shops (M. Hadi Kiapour, Xufeng Han, Svetlana Lazebnik, Alexander C. Berg, Tamara L. Berg, CVPR 2015)



# Retrieval

- Exact retrieval: Video
  - Video2Shop: Exact Matching Clothes in Videos to Online Shopping Images (Zhi-Qi Cheng, Xiao Wu, Yang Liu, Xian-Sheng Hua, CVPR 2017)



# Main Topics in Fashion

## High-Level Application

Fashion Synthesis      Fashion Analysis

Recommendation

Outfit Matching      Retrieval

Style Classification

## Low-Level Understanding

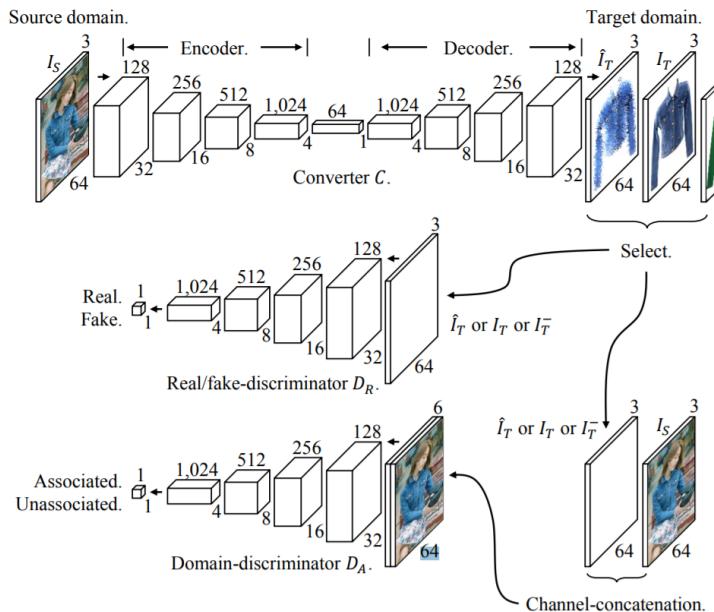
Fashion Parsing

Landmark Detection

Attribute Prediction

# Fashion Synthesis

- Pixel-Level Domain Transfer (Donggeun Yoo, Namil Kim, Sunggyun Park, Anthony S. Paek, In So Kweon, KAIST, ECCV 2016)
  - First attempt on fashion domain transfer with GAN
  - Paired data in low resolution (64x64)



# Fashion Synthesis

- Pixel-Level Domain Transfer (**Donggeun Yoo, Namil Kim, Sunggyun Park, Anthony S. Paek, In So Kweon, KAIST, ECCV 2016**)
  - First attempt on fashion domain transfer with GAN
  - Paired data in low resolution (64x64)

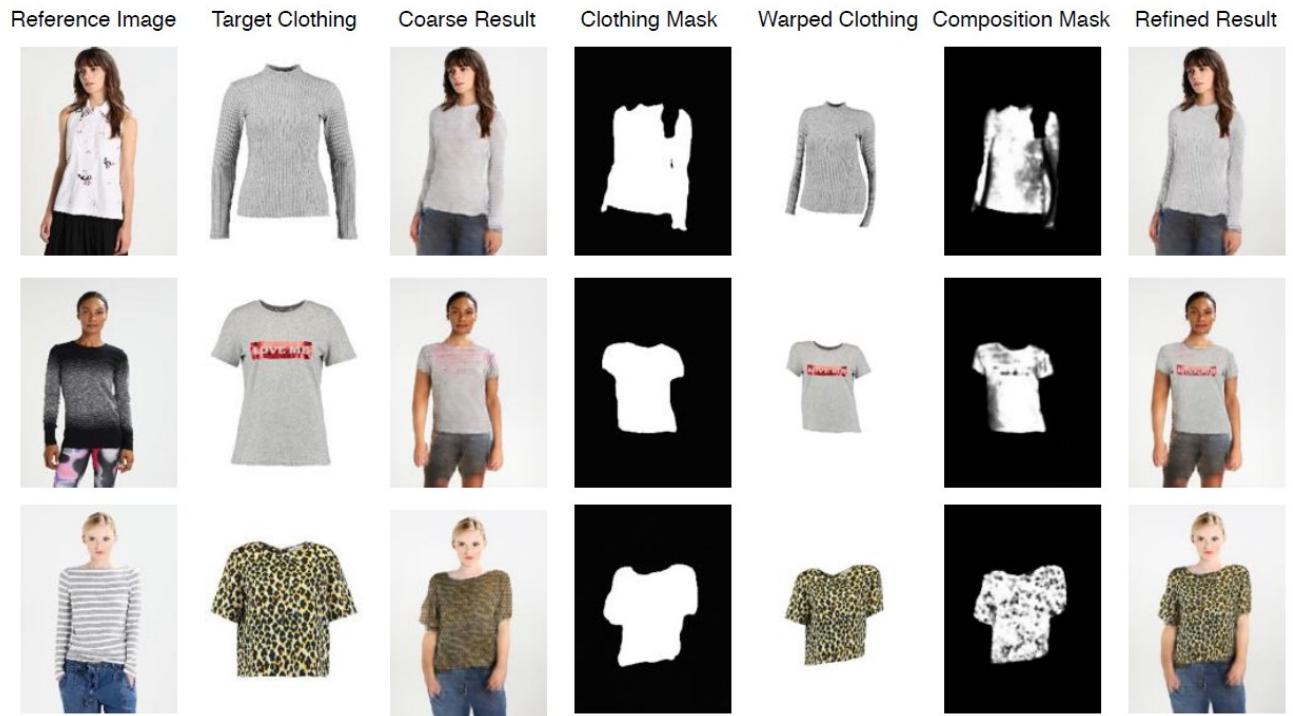
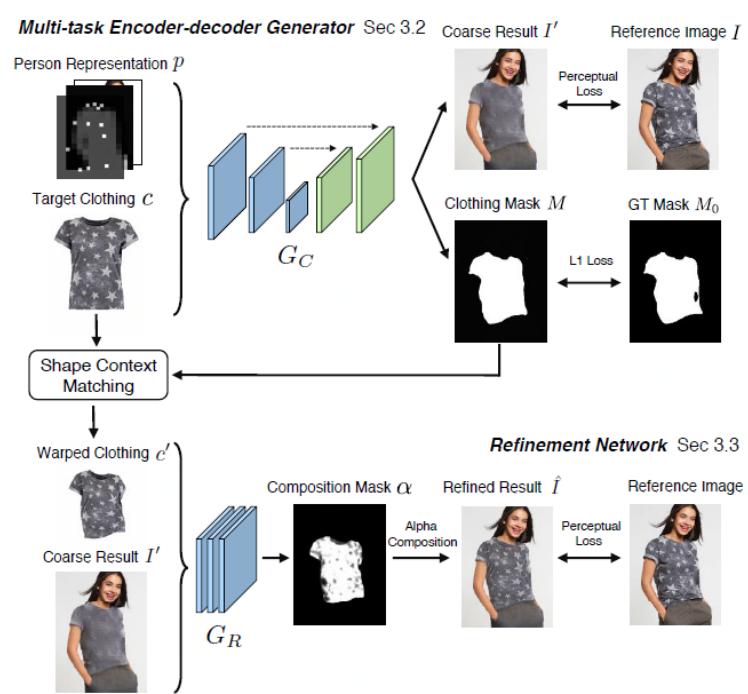


(1) Human to product

(2) Product to human

# Fashion Synthesis

- **VITON: An Image-based Virtual Try-on Network (Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, Larry S. Davis, CVPR 2018)**
  - Paired data with masks
  - Multi stages: coarse-to-fine





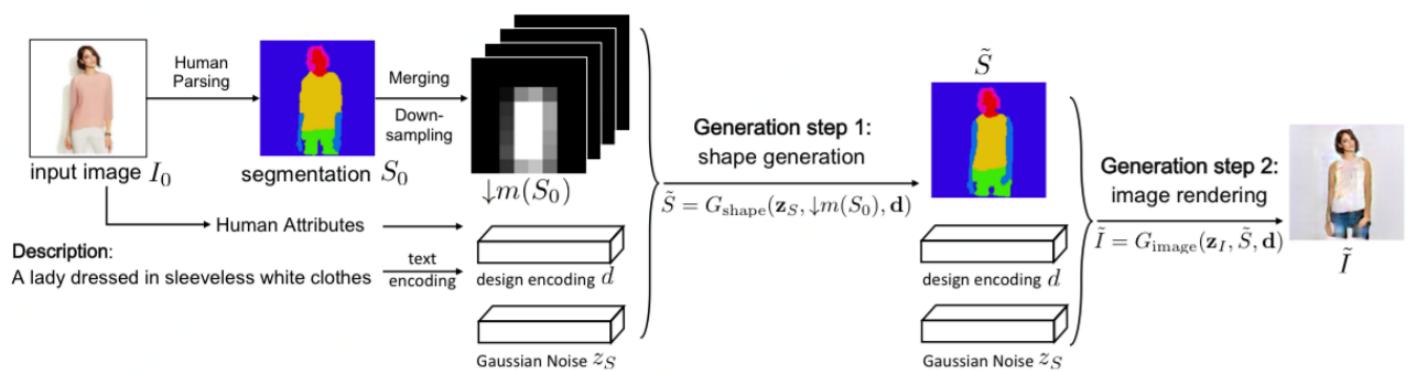
PRGAN: L. Ma, X. Jia, Q. Sun, B. Schiele, T. Tuytelaars, and L. Van Gool. Pose guided person image generation. In NIPS, 2017

CAGAN: N. Jetchev and U. Bergmann. The conditional analogy gan: Swapping fashion articles on people images. In ICCVW, 2017

CRN: Q. Chen and V. Koltun. Photographic image synthesis with cascaded refinement networks. In ICCV, 2017

# Fashion Synthesis

- Be Your Own Prada: Fashion Synthesis with Structural Coherence (Shizhan Zhu, Sanja Fidler, Raquel Urtasun, Dahua Lin, Chen Change Loy, ICCV 2017)
  - Image generation conditioned on text



# Other Fancy Works

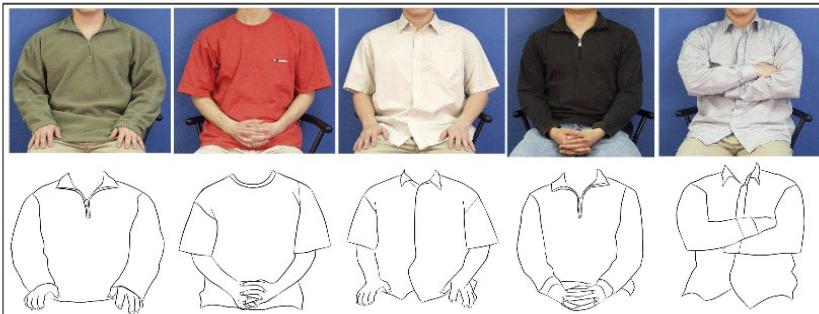


Figure 9. More recognition results of upper body with clothes. The results are the composite graphical templates inferred from the images.

Image to Sketch(Chen et al. CVPR 2006)



Sketch to Image (Isola et al. CVPR 2017)



Fashion Style Generator (Jiang et al. IJCAI 2017)



Clothing Generation  
(Lassner et al. ICCV 2017)

# Main Topics in Fashion

## High-Level Application

Fashion Synthesis      Fashion Analysis

Recommendation

Outfit Matching      Retrieval

Style Classification

## Low-Level Understanding

Fashion Parsing

Landmark Detection

Attribute Prediction

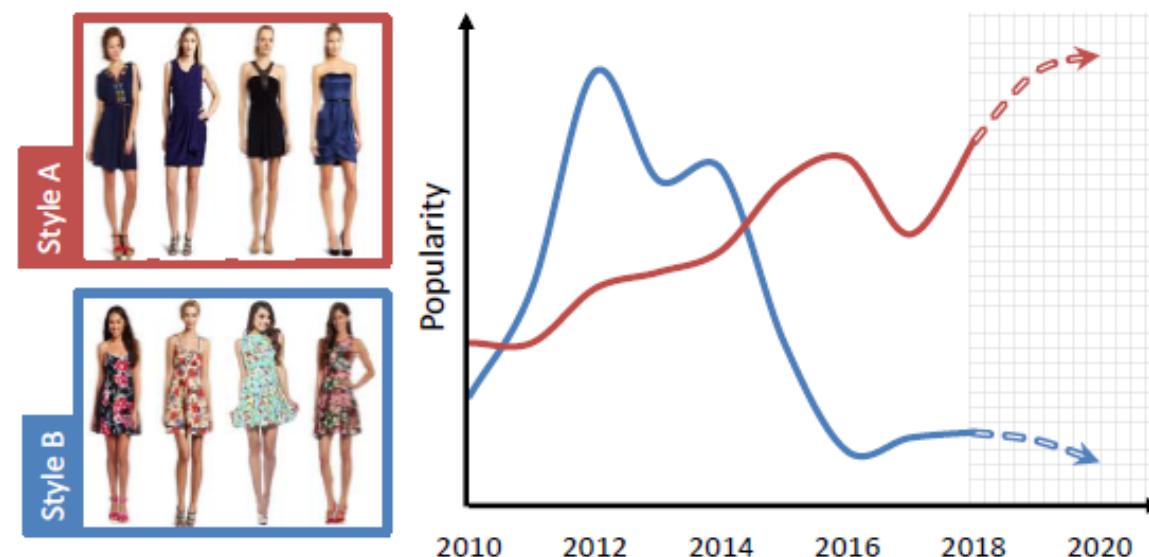
# Fashion Trend Analysis

- Spatial domain
  - **Fashion World Map: Understanding Cities Through Streetwear Fashion (Yu-Ting Chang, Wen-Haung Cheng, Bo Wu, Kai-Lung Hua, MM 2017)**

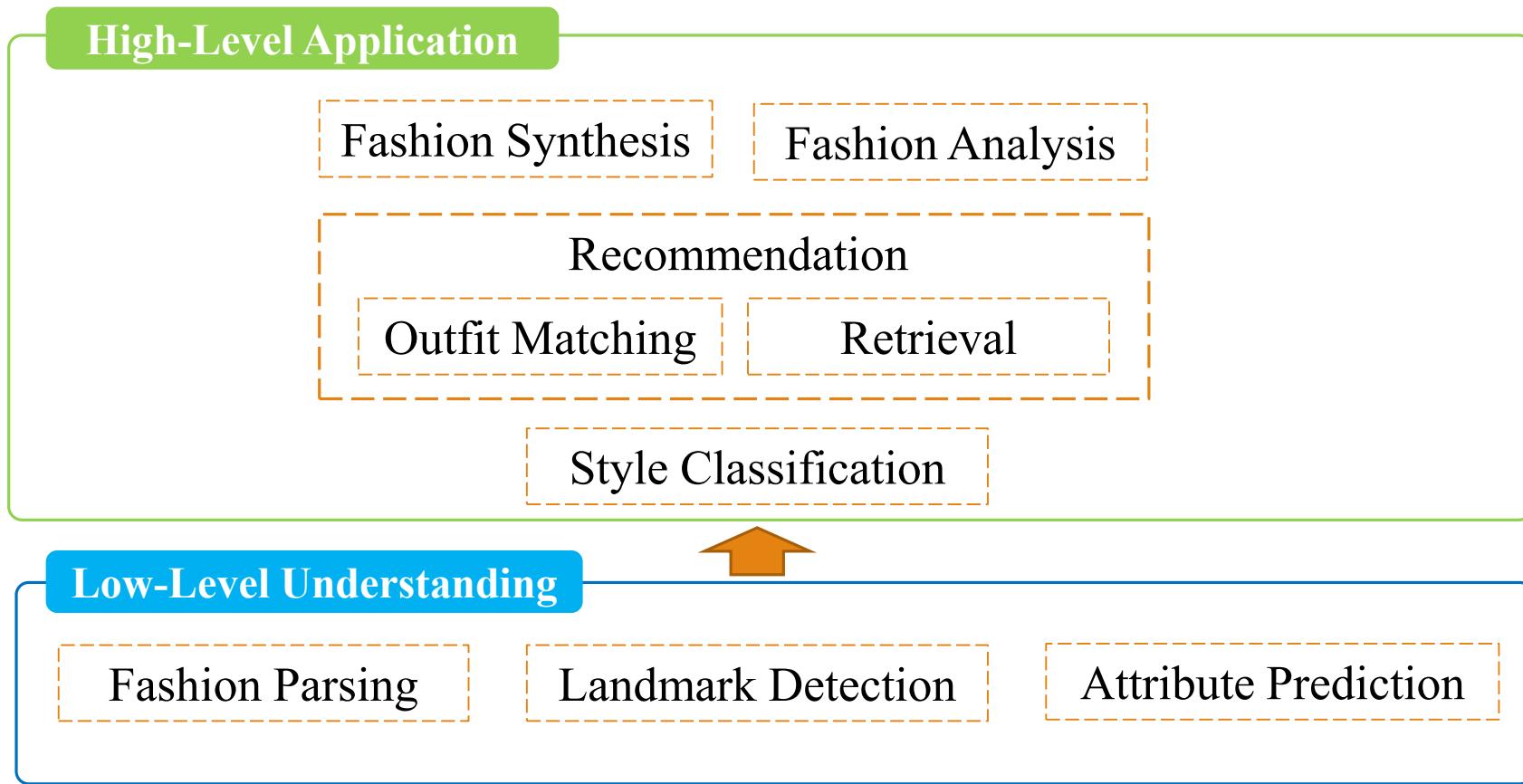


# Fashion Trend Analysis

- Temporal domain
  - Fashion Forward: Forecasting Visual Style in Fashion (Ziad Al-Halah, Rainer Stiefelhagen, Kristen Grauman, ICCV 2017)



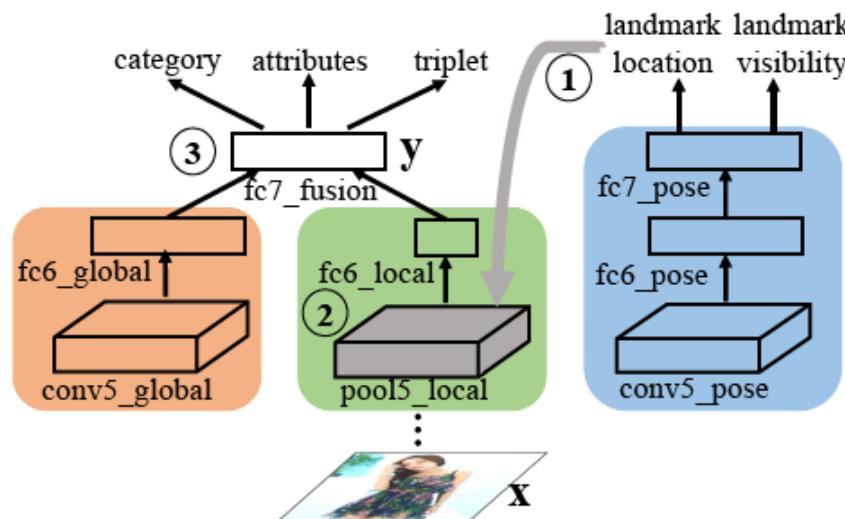
# Main Topics in Fashion



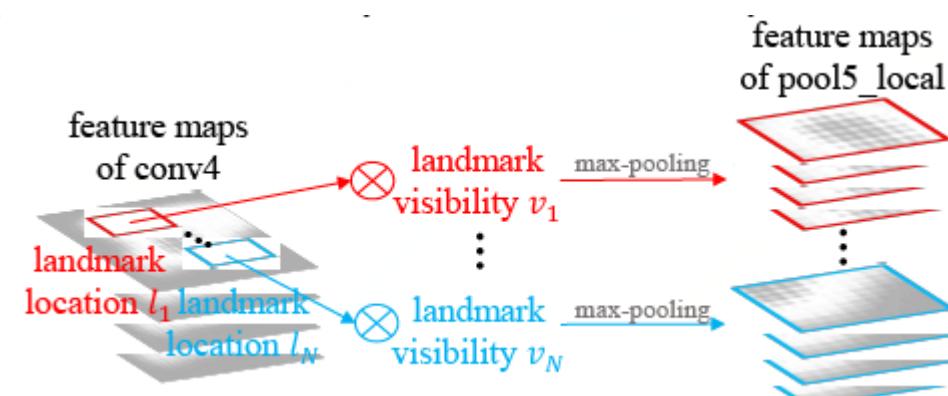
# Insights from DeepFashion

- Fashion-Net

- DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations (Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang, CVPR 2016)



(a) Fashion-Net



(b) Landmark pooling

# Insights from DeepFashion

- Category classification & Attribute prediction
  - Replace landmark with poselets: 6% ~ 9% drop
  - Training with 100, 500, 1000 attributes: Increasing consistently

	Category		Texture		Fabric		Shape		Part		Style		All	
	top-3	top-5												
WTBI [3]	43.73	66.26	24.21	32.65	25.38	36.06	23.39	31.26	26.31	33.24	49.85	58.68	27.46	35.37
DARN [10]	59.48	79.58	36.15	48.15	36.64	48.52	35.89	46.93	39.17	50.14	66.11	71.36	42.35	51.95
FashionNet+100	47.38	70.57	28.05	36.59	29.12	40.58	28.51	36.51	31.65	38.53	53.92	62.47	31.58	39.06
FashionNet+500	57.44	77.39	34.73	46.35	34.47	46.60	33.61	44.57	38.48	49.01	63.48	67.94	38.94	49.71
FashionNet+Joints [34]	72.30	81.52	35.92	48.73	38.21	49.04	37.59	47.73	40.21	51.81	64.91	73.14	43.14	52.33
FashionNet+Poselets [34]	75.34	84.87	36.85	49.11	38.88	49.48	38.19	47.09	41.60	52.85	64.84	73.03	43.57	52.65
FashionNet (Ours)	<b>82.58</b>	<b>90.17</b>	<b>37.46</b>	<b>49.52</b>	<b>39.30</b>	<b>49.84</b>	<b>39.47</b>	<b>48.59</b>	<b>44.13</b>	<b>54.02</b>	<b>66.43</b>	<b>73.16</b>	<b>45.52</b>	<b>54.61</b>

Table 2. Performance of category classification and attribute prediction.

# Insights from DeepFashion

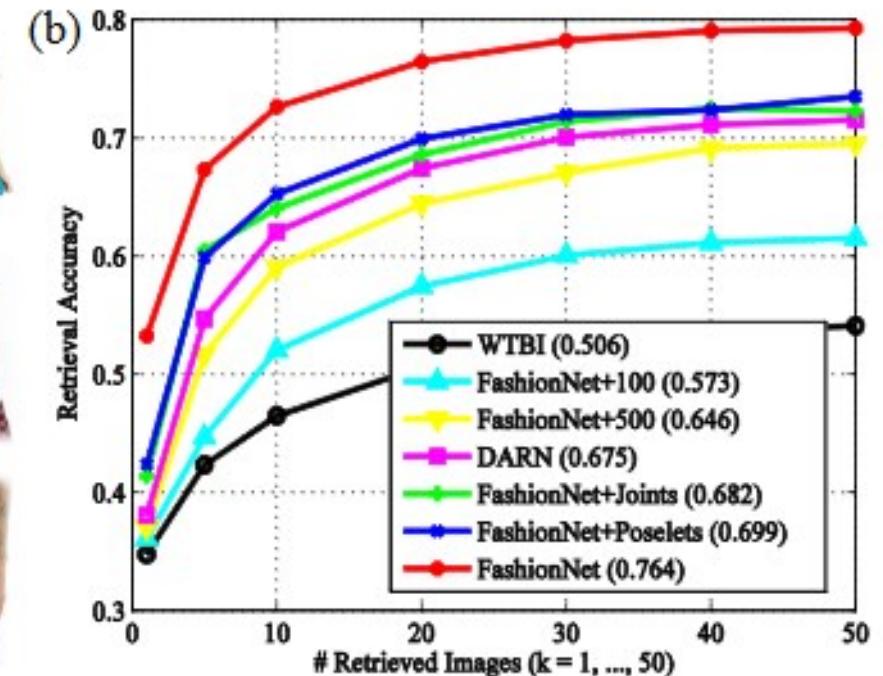
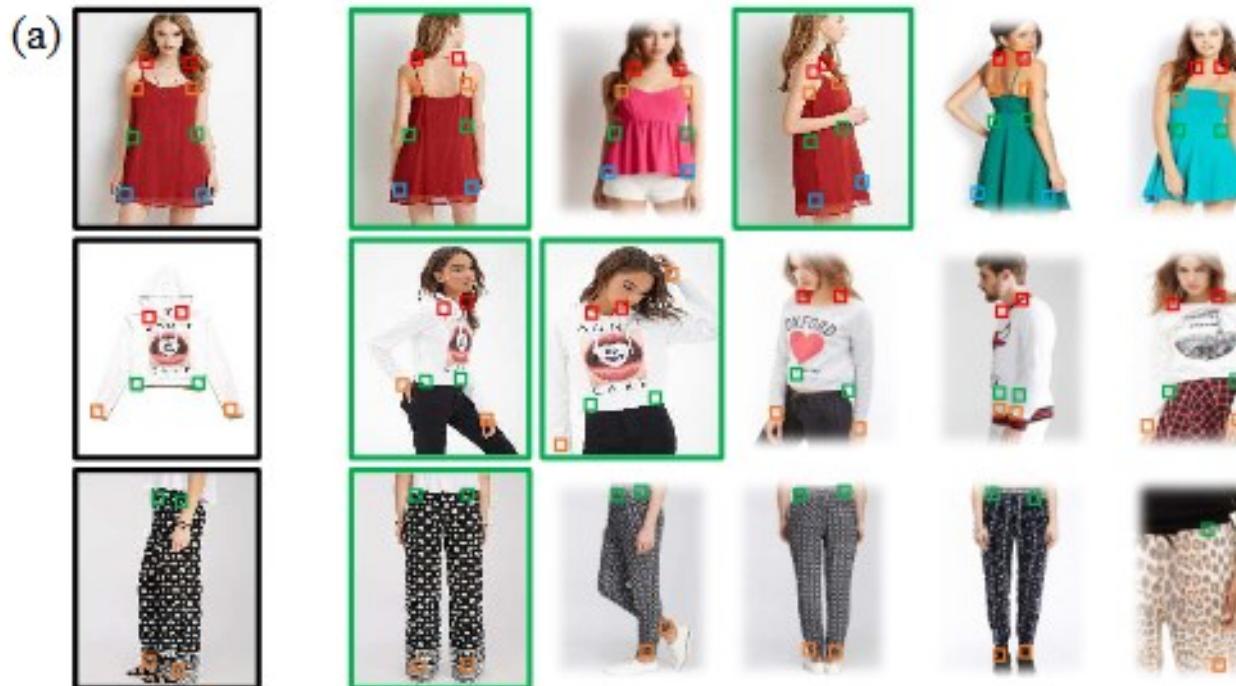
- Category classification & Attribute prediction
  - Replace landmark with poselets: 6% ~ 9% drop
  - Training with 100, 500, 1000 attributes: Increasing consistently

	Category		Texture		Fabric		Shape		Part		Style		All	
	top-3	top-5												
WTBI [3]	43.73	66.26	24.21	32.65	25.38	36.06	23.39	31.26	26.31	33.24	49.85	58.68	27.46	35.37
DARN [10]	59.48	79.58	36.15	48.15	36.64	48.52	35.89	46.93	39.17	50.14	66.11	71.36	42.35	51.95
FashionNet+100	47.38	70.57	28.05	36.59	29.12	40.58	28.51	36.51	31.65	38.53	53.92	62.47	31.58	39.06
FashionNet+500	57.44	77.39	34.73	46.35	34.47	46.60	33.61	44.57	38.48	49.01	63.48	67.94	38.94	49.71
FashionNet+Joints [34]	72.30	81.52	35.92	48.73	38.21	49.04	37.59	47.73	40.21	51.81	64.91	73.14	43.14	52.33
FashionNet+Poselets [34]	75.34	84.87	36.85	49.11	38.88	49.48	38.19	47.09	41.60	52.85	64.84	73.03	43.57	52.65
FashionNet (Ours)	<b>82.58</b>	<b>90.17</b>	<b>37.46</b>	<b>49.52</b>	<b>39.30</b>	<b>49.84</b>	<b>39.47</b>	<b>48.59</b>	<b>44.13</b>	<b>54.02</b>	<b>66.43</b>	<b>73.16</b>	<b>45.52</b>	<b>54.61</b>

Table 2. Performance of category classification and attribute prediction.

# Insights from DeepFashion

- In-Shop Clothes Retrieval

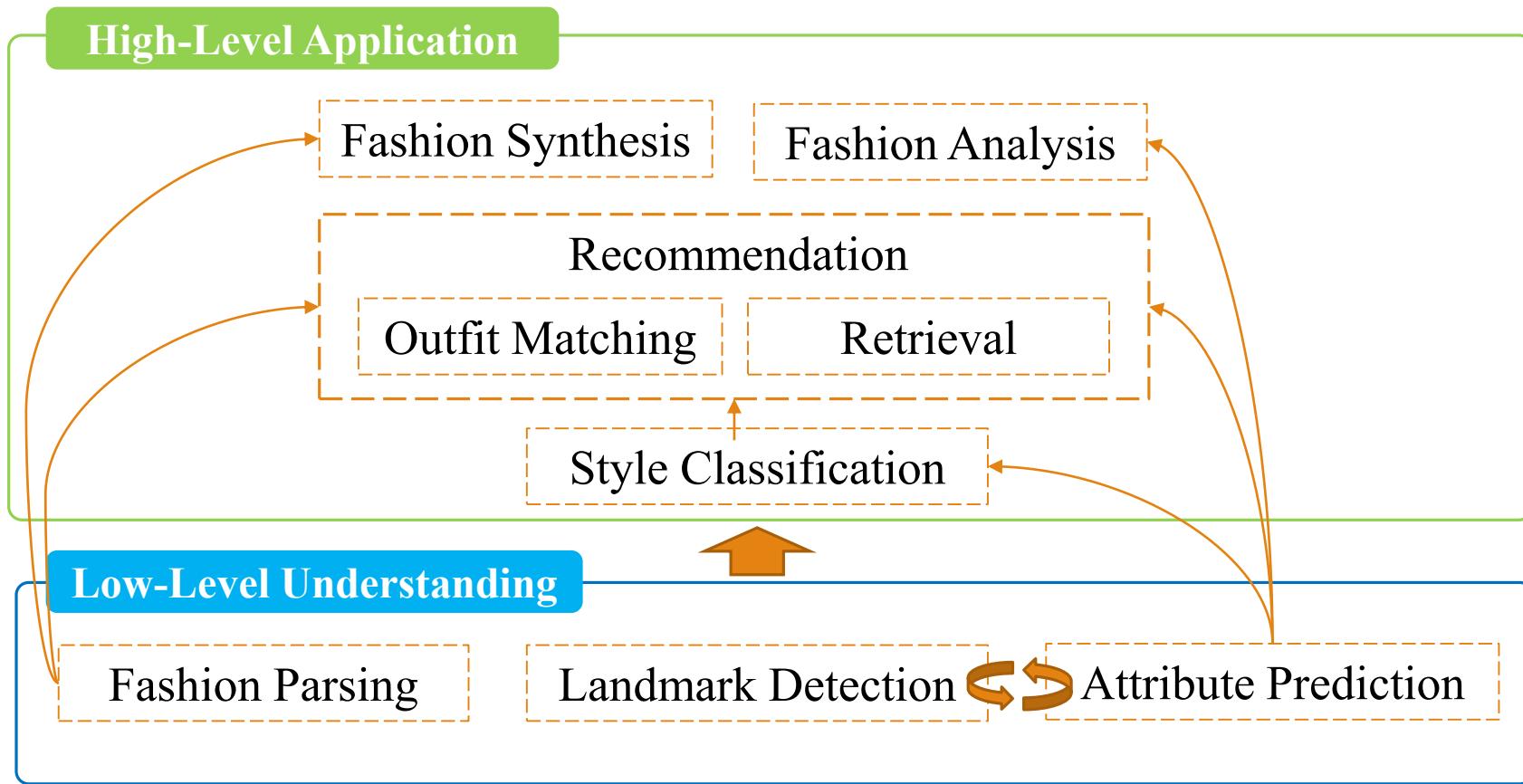


# Insights from DeepFashion

- Consumer-to-Shop Clothes Retrieval



# Main Topics in Fashion



# Our Work

---

- Outfit matching
  - A key step to fashion design
  - **Definition:** compatibility prediction

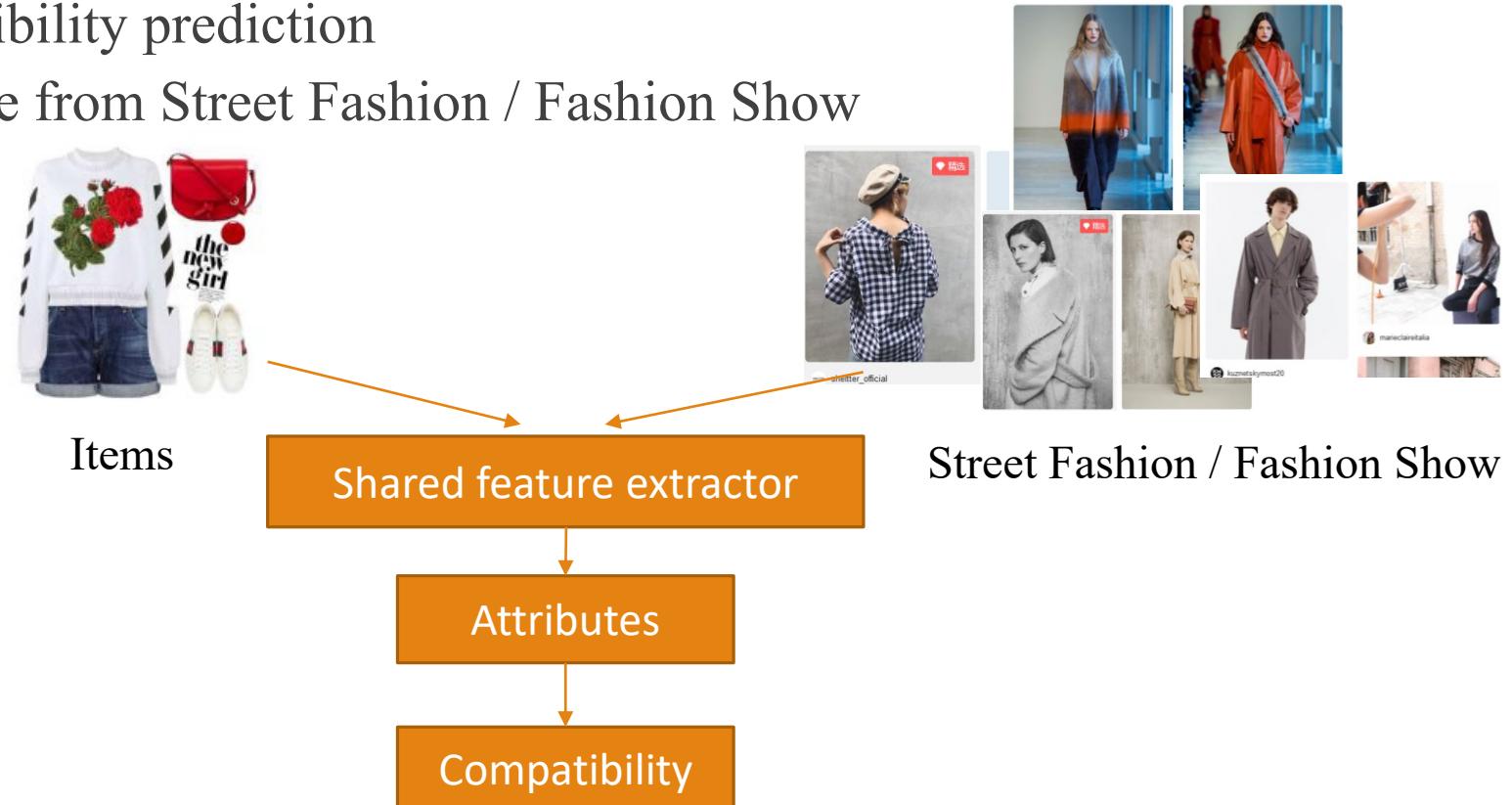


- **Application:** recommendation/generation



# Our Work

- Outfit matching
  - **Definition:** compatibility prediction
  - Learning knowledge from Street Fashion / Fashion Show
  - Cross domain
  - Sparse to attributes



# Our Work

---

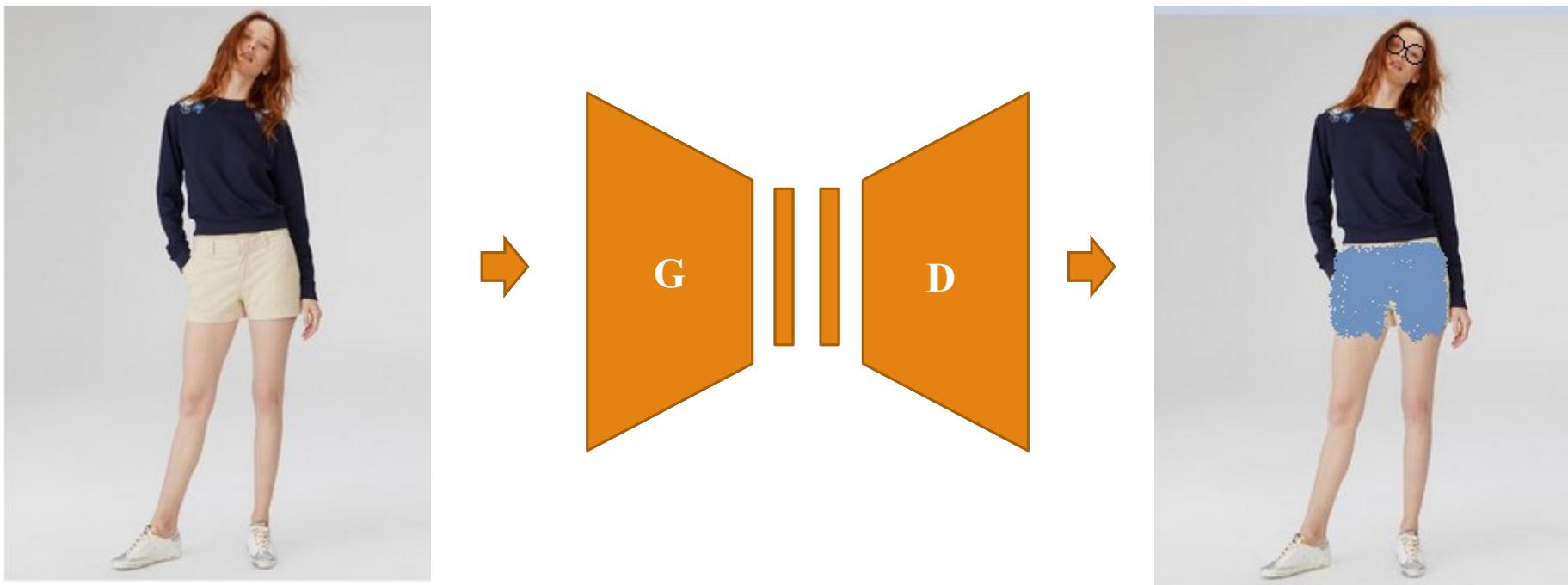
- Outfit matching
  - **Application:** recommendation
  - Decision making → Reinforcement learning
  - Challenge: huge action space



# Our Work

---

- Outfit matching
  - **Application:** generation
  - GAN: improve compatibility (side effect: D → compatibility prediction?)



# Thanks

---

## Q & A

Sijie Song