Personalised Fashion Recommendation Using Deep Learning

Muthusamy Chelliah
Flipkart
muthusamy.c@flipkart.com

Shamik Sural
IIT Kharagpur
shamik@cse.iitkgp.ac.in

Omprakash Sonie
Flipkart
omprakash.s@flipkart.com

Agenda

1. Introduction (5 min.s)
2. Style Detection (30 min.s)
3. Recommendation (30 min.s)
4. Outfit compatibility (15 min.s)
5. Q&A (10 min.s)

Introduction

Clothing indicates traits

Profession

Wealth

Occasion

Outing

Religion

weather

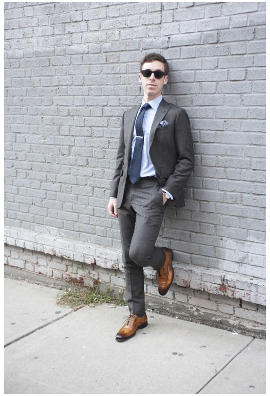
ethnicity

place

Flipkart

Socio-identity

Banker/wingtips



Chic

New york upper eastsider

Clues about fashion sense/tribe

Goth

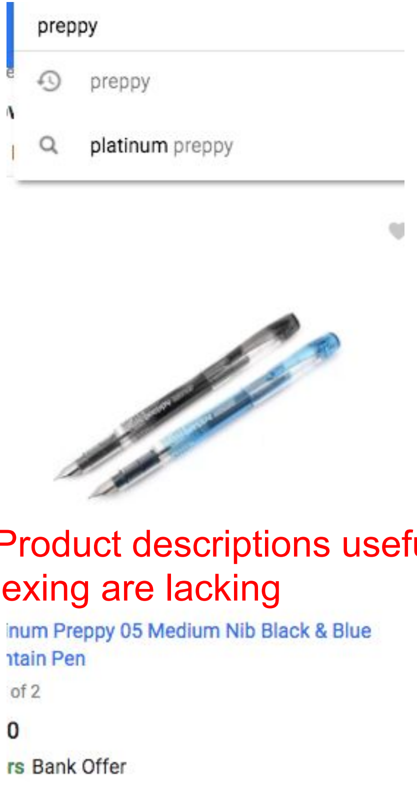
Hipster/flannel-jeans
Black-rimmed glasses

Preppy

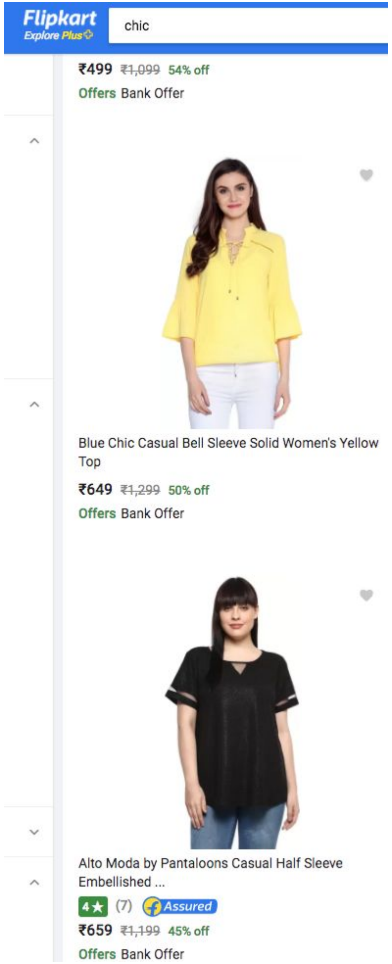
Flipkart

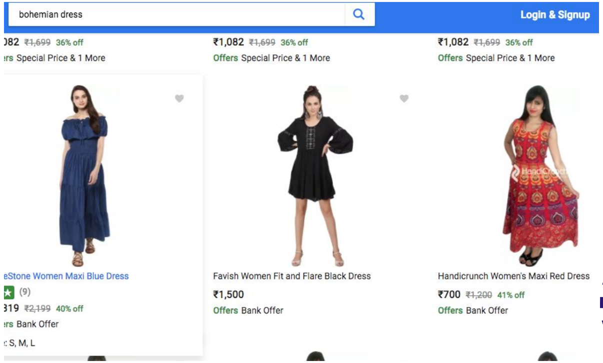
Style search

1. Relevance is visual
2. Limiting textual queries



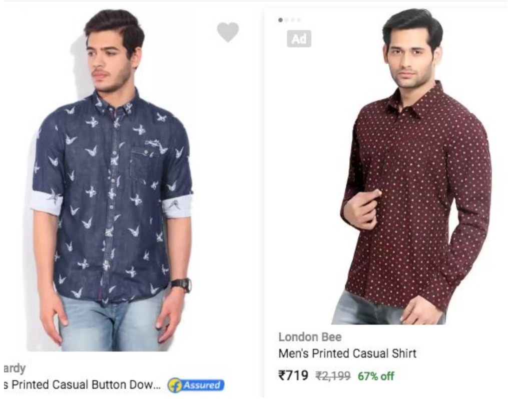
3. Product descriptions useful for indexing are lacking





Style variation - components

Day casual

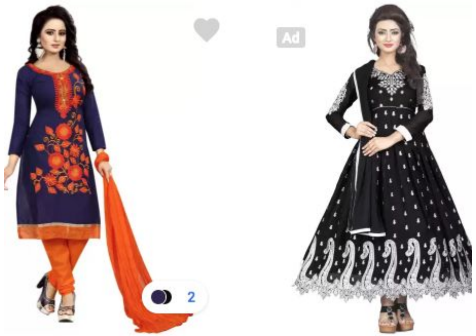


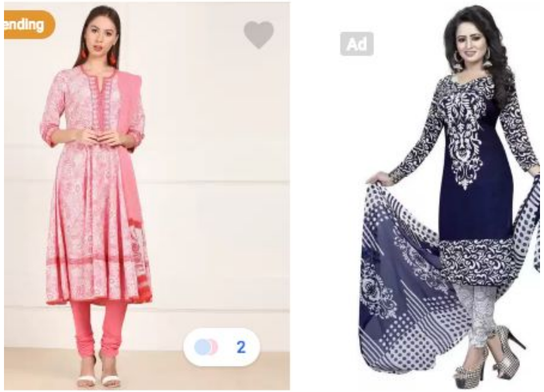
Night formal

Checkered

Striped

Festive & party



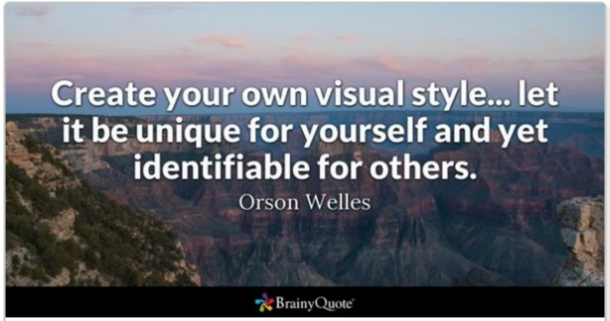


Printed

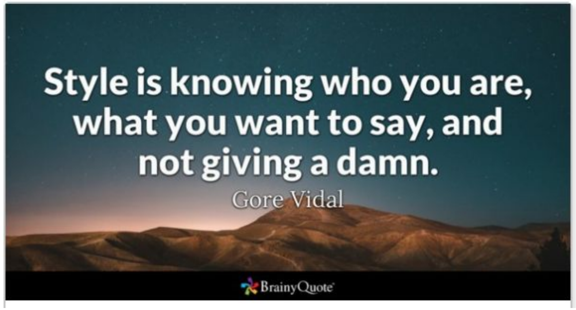
What is style?

Quality
Simple,
Personal
Brevity
Eternal

What is style? (contd.)



Identity
Comfort
Persistence
Freedom



Flipkart

Visual search

Query photo on left retrieves similar clothing on right

- Contextual link between garments/ body parts (e.g., shoe-foot)

Garment types

- Shoe, socks, belt, hat, cape
- **Romper**, vest, blazer, **wedge**
- Leggings, **jumper**

Variation

- Configuration, layering, appearance

[Berg 12] Parsing clothing in fashion photographs, CVPR

Visual recommendation

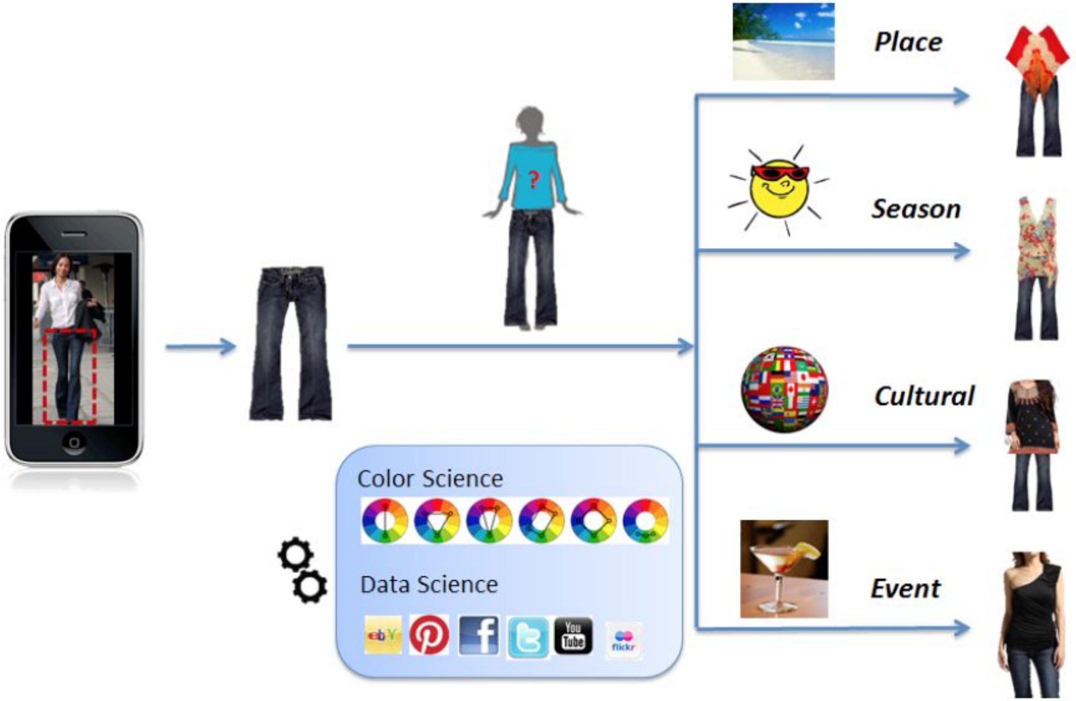
Matching/complementary items

Item representation as visual features

Inter-dependency between features of various items

Fashion-136K dataset

Annotations of brand, accessories, demographics



[Piramuthu 13] Style finder: Fine-grained clothing style detection and retrieval, CVPR workshops

Flipkart

Style detection

Computer Vision based

Style recognition

- Identify style for each image
  - how strongly displayed
- tournament -based game
  - Crowd-sources reliable ratings
- predict clothing elements
  - Discriminative for a style
  - Belonging in an outfit

[Berg '14] Hipster wars: discovering elements of fashion styles, ECCV

Style labeling

Chictopia/ Fashionista - dataset with 158K photos

Pictures uploaded to draw attention

Display range of accessories

noisy/incomplete comments/ links

Help extract info. about photos

Groundtruth annotation

14 body parts

Labels on super-pixel regions

- HEATHER GRAY DOUBLE BREASTED VINTAGE SWEATER
- CREAM FOREVER 21 SOCKS
- BROWN SEYCHELLES BOOTS
- CREAM SELF MADE POSTLAPSARIA DRESS
- BROWN BELT
- BLACK COACH BAG

- ROMANTIC // KNEE HIGHS // VINTAGE // BOOTS // DRESS // CREAM // COMFY // CARDIGAN // SELF MADE // AWESOME BOOTS // DOLL // WARM // ADORABLE // COLD WEATHER // CUUUTEE

- HOLIDAY SWEATER STYLE FOR BRUNCH IN FALL 2010

Flipkart

[Berg 12] Parsing clothing in fashion photographs, Tamara L. Berg et al., CVPR

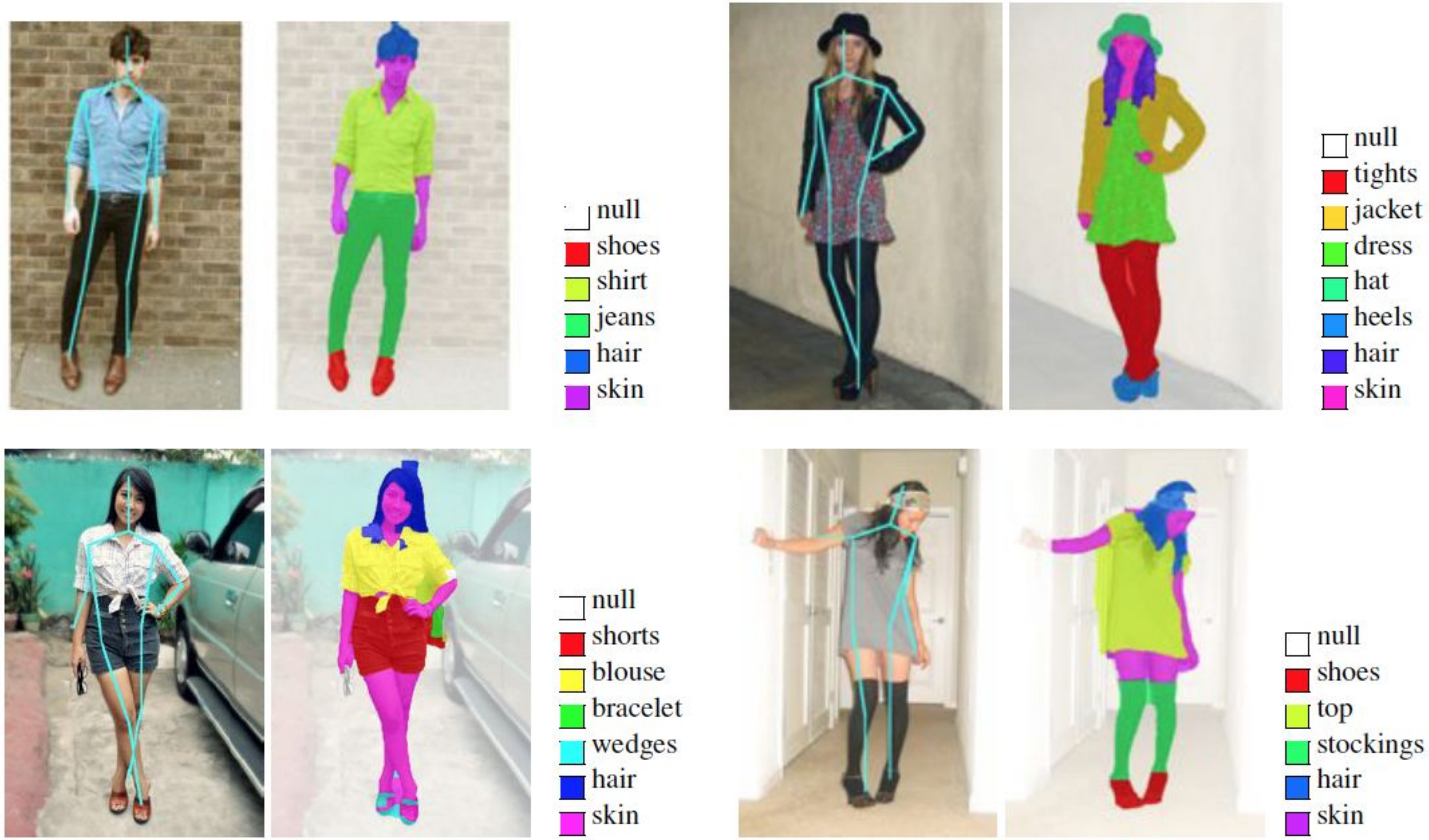
Parsing clothing

Handles partially occluded shoes/small hats

Clean background/ distinctive regions

Images segmented into superpixels

Labels predicted in a CRF model



[Berg 12] Parsing clothing in fashion photographs, CVPR

Flipkart

Outfit retrieval

Query with groundtruth annotation

Retrieved image with tags

Combine parsing

- Pre-trained global models
- Local models learned on-the-fly
- Parse mask predictions transferred

Flipkart

[Berg 13] Paper doll parsing: retrieving similar styles, CVPR

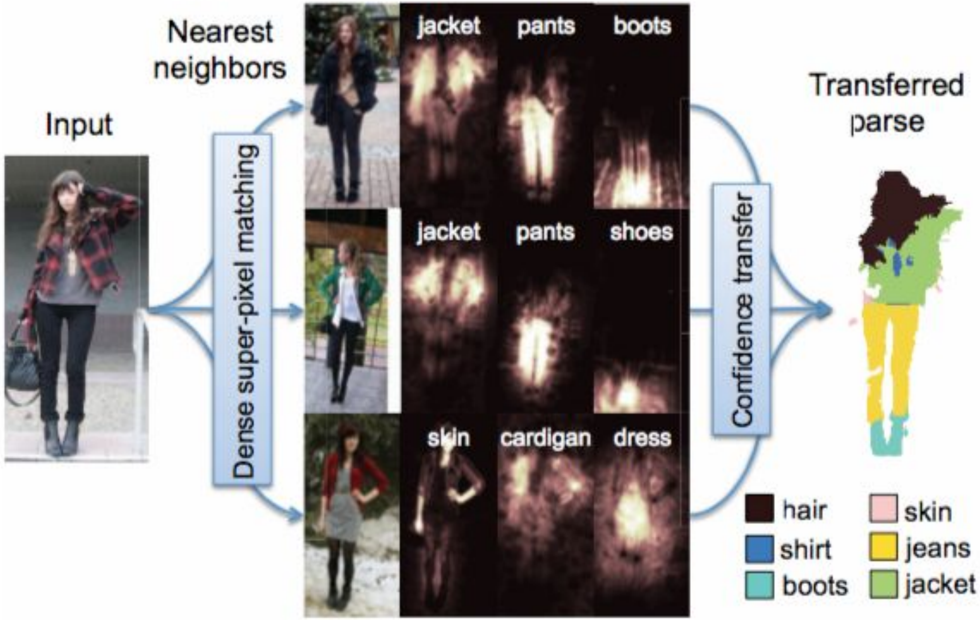
Transferred Parse

1. Global likelihood of label assignment trained for each item

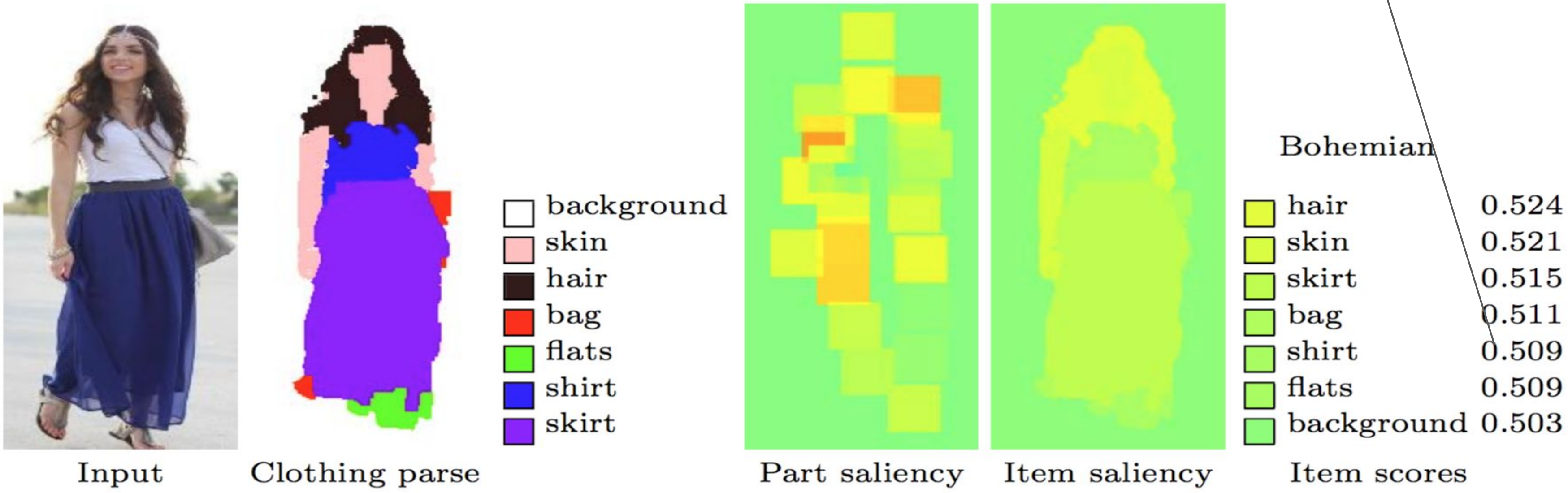
2. Logistic regression trained only on nearest-neighbor (NN) images

3. Over-segmentation of both query/retrieved images

Pixel features: BoW from RGB, Lab, MR8, gradient, HOG, boundary distance



[Berg 13] Paper doll parsing: retrieving similar styles, CVPR scores indicate impact of outfit locations on style



Clothing segmentation on all images of each style

[Berg ‘14] Hipster wars: discovering elements of fashion styles, ECCV

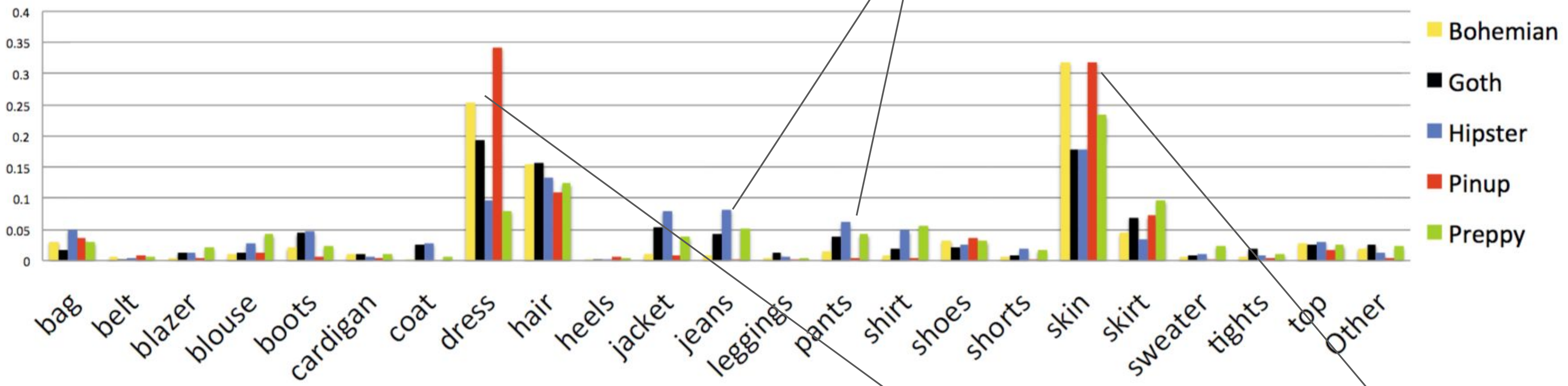
Predicted style

outfit elements ordered by indicator scores

[Berg '14] Hipster wars: discovering elements of fashion styles, ECCV

Clothing items across styles

Hipsters/preppies wear jeans and pants



garment most indicative of each style as percentage of item

Pinups/bohemians tend to wear dress

Pinups display skin

[Berg '14] Hipster wars: discovering elements of fashion styles, ECCV

Flipkart

Issues with CV-based approaches

- Similarity based on distance in image feature space
  - NOT alignment of attributes that capture high-level semantics
- Hand-crafted methods with carefully annotated data
  - NOT general-purpose to understand object relationships from large volumes on unlabeled data
- Historical user feedback not leveraged to learn personalized preferences

Big data based

Style capture

- Interface for recognizing visual elements
  - Beyond color/pattern
  - (e.g., material)
- Matching/extraction is a challenge
  - Due to deformability of clothing items
- Attribute vocabulary
  - Constructed using human annotations

[Piramuthu 13] Style finder: Fine-grained clothing style detection and retrieval, CVPR workshops

Image attribute tagging

- Helps shoppers target desired items
  - without further browsing
- Clothing styles change over time
  - and vary per season/location
- Hard to tell those shared among individual items
  - E.g., blouse, swimsuit

Image	Automatically Generated Description
(a)	Button, has-belt, asymmetrical and double-breasted, long, slim, v-neck shirt collar, round shirt collar
(b)	Wool/is woolen or felt-like, zip, loose, round collar
(c)	Has-fur, v-neck-collar
(d)	Symmetrical and single-breasted, shawl collar
(e)	Shiny, short
(f)	Has-fur, leather/is leather-like
(g)	Open, loose, round collar
(h)	Button, short, chest pocket, v-neck shirt collar

[Piramuthu 13] Style finder: Fine-grained clothing style detection and retrieval, CVPR workshops

Labeling approach

- Could afford only small datasets
  - Complex procedures are required for extraction without overfitting
- Results cannot be transferred
  - To related problems
- Annotation for modeling pair-wise distances
  - Scales with square of number of elements

[Piramuthu 13] Style finder: Fine-grained clothing style detection and retrieval, CVPR workshops

[Mcauley 15] Image-based recommendations on styles and substitutes, SIGIR

Within/across style classification

- Human notion of visual relationships, not just similarity
  - *Also viewed* data (substitute)
  - Viewed X bought Y
- Closely clustered items in style space
- Influence of appearance on desirable attributes
- Network inference
  - graphs of related images

[Mcauley 15] Image-based recommendations on styles and substitutes, SIGIR

Flipkart

Stylistic coherence

- Bought X bought Y (complement)
- Viewed X and Y, bought both

- Widely separated members
  - Project to locations that span a single k-means cluster

[Mcauley 15] Image-based recommendations on styles and substitutes, SIGIR

Uncovering book categories

- Visual features help interpret meaningful distinctions
  - Children’s, self-help, romance, graphic

[Mcauley 15] Image-based recommendations on styles and substitutes, SIGIR

Outfit recommendation

- Women’s at left and men’s at right
- First columns shows catalog item
  - E.g., hat
  - Selected randomly
- Right three columns match query item
  - Minus whichever category contains it
  - E.g., pant, shirt, shoes

Flipkart

[Mcauley 15] Image-based recommendations on styles and substitutes, SIGIR

Model: visual/relational recommendation

Represent user preference for visual appearance of one object given that of another

Learn distance metric for item compatibility

1. Feature vectors for related objects assigned a lower distance
2. Feature dimensions relevant for a particular relationship (i.e., **visual similarity** with **varying emphasis**)
3. Relationship between feature dimensions (e.g., **pants** transformed to find **compatible shoes**)
4. Product's **embedding into style-space** so related objects are closer despite visual difference
5. Dimensions of style (i.e., similar for related objects) **important to each user**

[Mcauley 15] Image-based recommendations on styles and substitutes, SIGIR

Visual evolution of styles

- Above timeline
  - Women’s sneakers
    - Three most fashionable styles (groups)
    - During each year/epoch
- Below timeline
  - Specific user’s purchase
    - Modeled as a result of style/personal factors

[Mcauley 16] Ups and downs: modeling fashion trends with one-class collaborative filtering, WWW

Flipkart

Trending styles

- On the left
  - Query image representing a resurgent style in late 2000s
- In the middle
  - Nearest images in style space
- On the right
  - Relative visual popularity
  - Normalized visual score in each epoch

[Mcauley 16] Ups and downs: modeling fashion trends with one-class collaborative filtering, WWW ‘16

Modeling fashion aware preference predictor

Evolving fashion styles

Reflected in purchase histories than ratings

Generate time-dependent, personalized ranking

of items without user's feedback

Factor in visual compatibility between user/item

Siamese CNN to learn feature transformation

from image to latent space of metric distances

[Mcauley 16] Ups and downs: modeling fashion trends with one-class collaborative filtering, WWW '16

Flipkart

Similar styles

- Cluster pair on each row
  - Items on left compatible with those on right
  - Each side represents a clothing category
- Fine-grained recognition - alternative
  - of sub-categories/ attributes
  - Slim, dark, formal pants

[Mcauley 15a] Learning visual clothing styles with dyadic co-occurrences, ICCV

Flipkart

Dissimilar styles

- incompatible clusters in each row
- Hard to collect dataset/ domain knowledge
  - Due to seasonal fashion change
- Learn feature transformation
  - From images to latent style space
  - With Siamese CNNs

[Mcauley 15a] Learning visual clothing styles with dyadic co-occurrences, ICCV

Flipkart

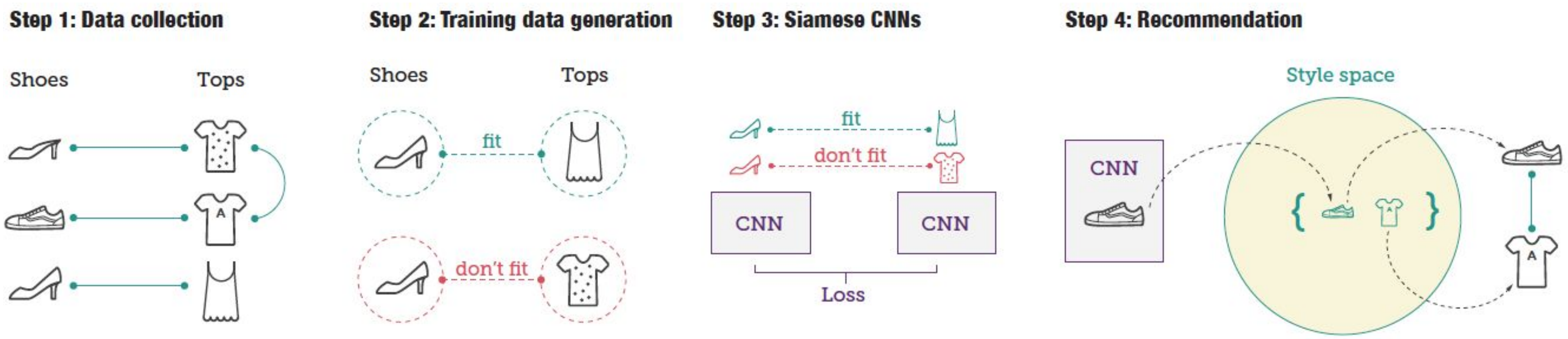
Outfit generation

- Each column is an outfit
  - generated by querying learned style space
- Indicated by green border
  - Query image
- Other items
  - Nearest neighbors

[Mcauley 15a] Learning visual clothing styles with dyadic co-occurrences, ICCV

Flipkart

Style learning with co-occurence



Links between items indicate co-occurence

Sample training pairs in different categories

Matching images are closer in style space

Retrieve structured bundles of compatible items

[Mcauley 15a] Learning visual clothing styles with dyadic co-occurrences, ICCV

Flipkart

Visual feature space issues

May fail to capture different item styles

- CNN features extracted from Caffe reference model

Metric learning not flexible to represent random item compatibility

- Only learns complex relationships through samples of positive/negative pairs

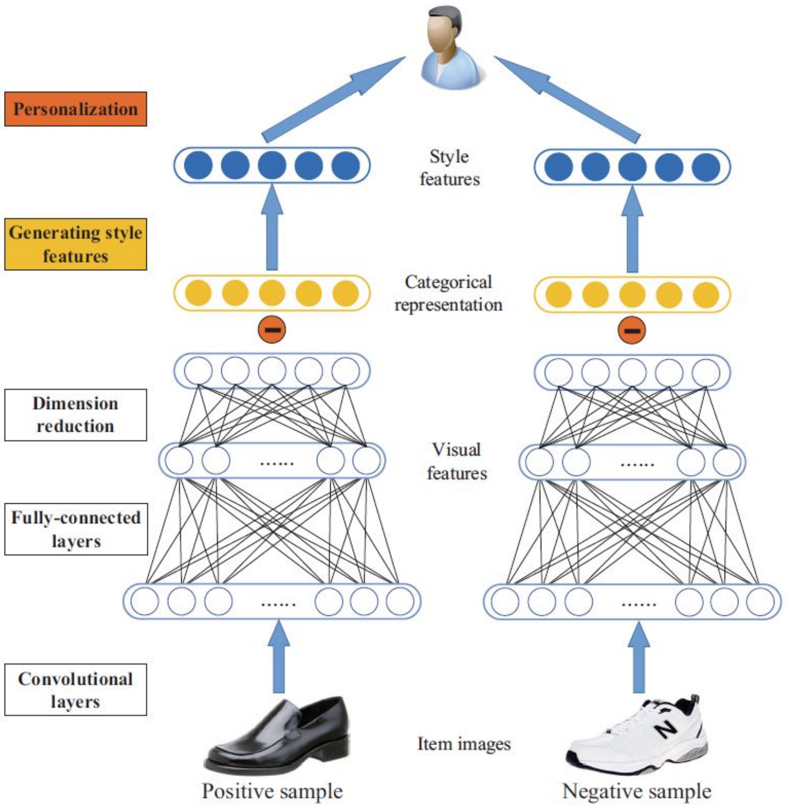
Deep learning based

Clustering by category

- Visual similarity (e.g., top, dress)
  - Style (e.g., casual, formal) not distinguished
- Similar styles bought together
  - different in visual feature space
  - E.g., suit pant vs. jeans or leather shoes

[Liu 17] DeepStyle: learning user preferences for visual recommendation, SIGIR

Flipkart



Visual minus category

- Eliminate category from item representation
- No embedding matrix per category
  - for transferring visual to style features

[Liu 17] DeepStyle: learning user preferences for visual recommendation, SIGIR

Flipkart

Clustering by style

Single category across clusters

Female items

Middle-aged women

casual

old

banquet

formal/official

Male items

[Liu 17] DeepStyle: learning user preferences for visual recommendation, SIGIR

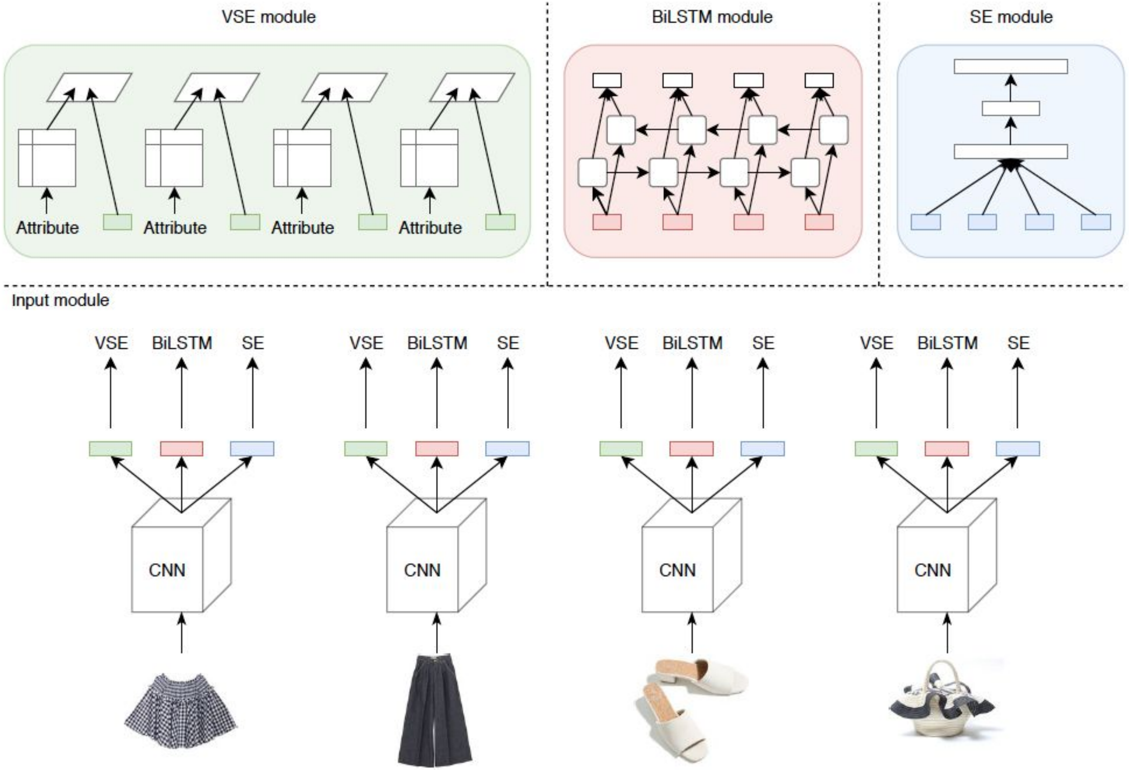
Flipkart

Outfit generation: per different styles

- Evaluate item compatibility
  - Sequence learning with local pairwise items
- Extract styles
  - which determine outfit taste

[Nakamura 18] Outfit generation via bidirectional LSTM/auto encoder, KDD, AI for Fashion WS

Outfit generation: per different styles



- Compatibility evaluated by Bidirectional LSTM
  - learning a sequence
- Style representation common to outfits extracted by autoencoder
  - Reducing item features to obtain style vector

[Nakamura 18] Outfit generation via bidirectional LSTM/auto encoder, KDD, AI for Fashion WS

Style extraction: element of a basis

- Encoded as one-hot vector in space
  - In which mixture ratio of each outfit is embedded
  - Pale tones/skirts, high contrast/sandal-jeans, dark-black/leather-metal, thick fabric/high heel-gold

[Nakamura 18] Outfit generation via bidirectional LSTM/auto encoder, KDD, AI for Fashion WS

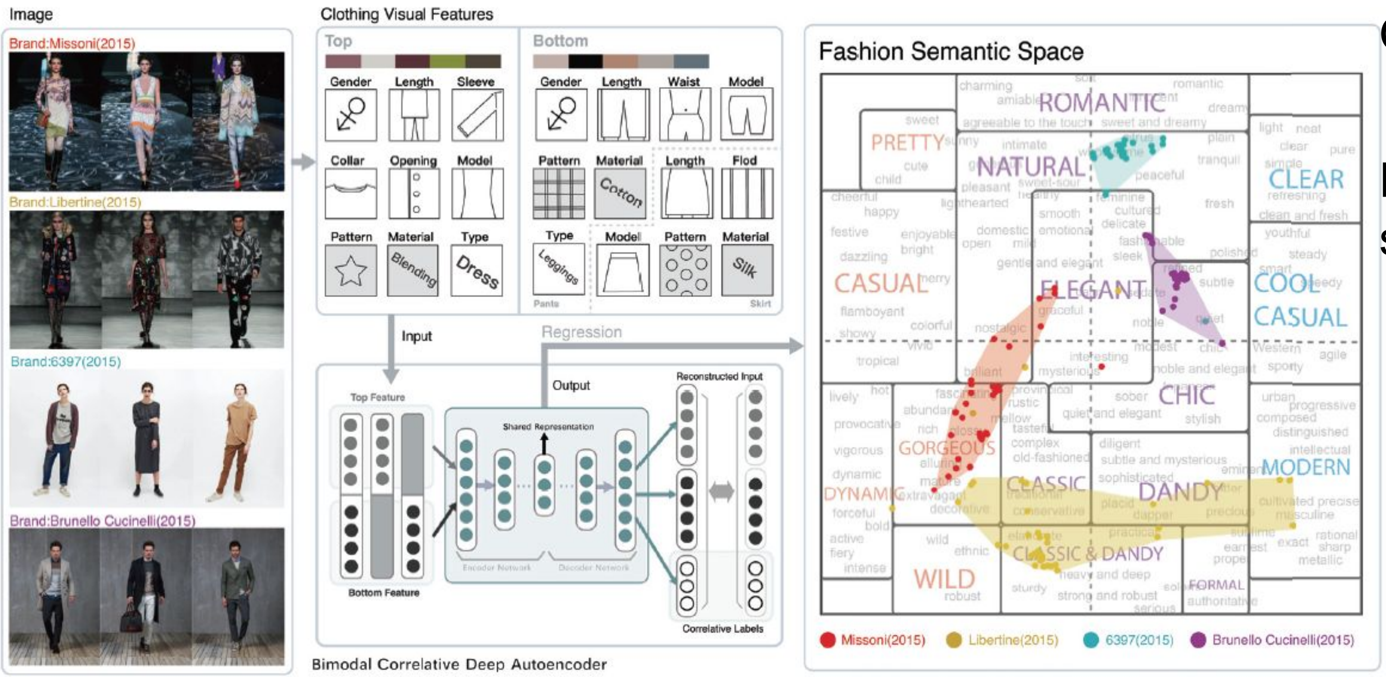
Style extraction: combination of two elements

- Mixture representation of style vectors
    - Simplify expression of complex outfits
    - Skirt + high-contrast colors
    - Skirts + dark/black items

[Nakamura 18] Outfit generation via bidirectional LSTM/auto encoder, KDD, AI for Fashion WS

Flipkart

Style understanding with multimodal deep learning



Quantitative description

Hard-soft (Y) and warm-cool (X)

Relate visual feature to style

tops/bottom

2 modals of clothing collocation

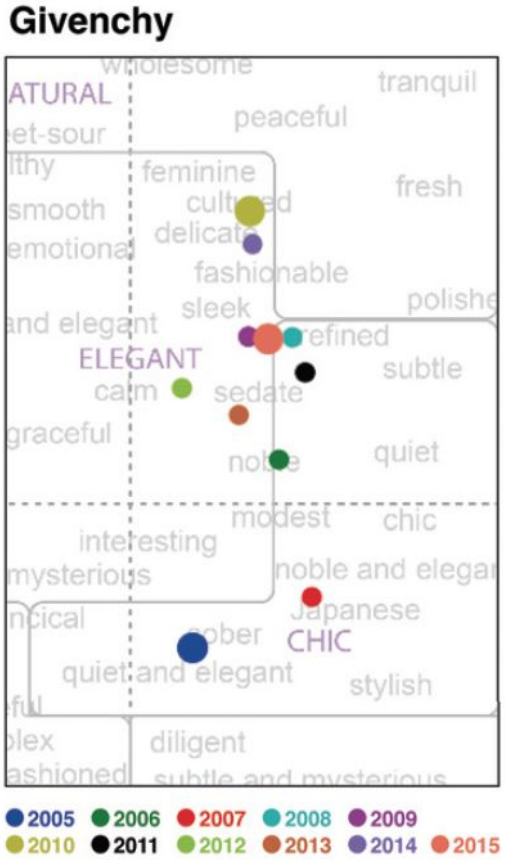
Improve feature learning

Category (suit) as correlative label

[Jia 17] Understanding clothing fashion styles: deep learning approach, AAAI

Flipkart

Fashion trend



Different years' fashion comparison of same brand

Center point of each year shows brand style has changed 2005-15

[Jia 17] Understanding clothing fashion styles: deep learning approach, AAAI

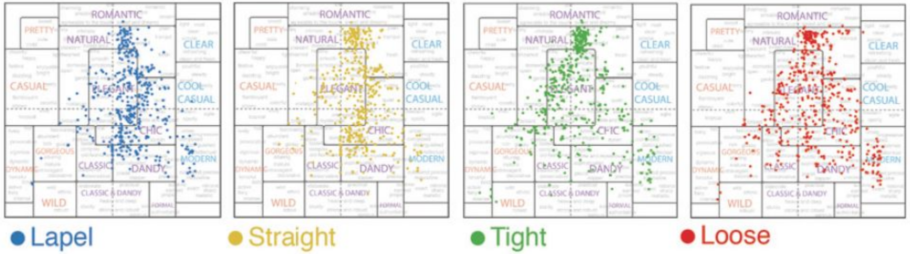
Flipkart

Feature colocation

Co-occurrence matrix of top collar shape and bottom model features

	Hoodie	Lapel	Stand	V	Fur	High	Round	Bateau
Straight	0.89	0.88	0.82	0.71	0.66	0.64	0.64	0.39
Tight	0.05	0.06	0.10	0.18	0.27	0.26	0.24	0.48
Loose	0.06	0.06	0.08	0.11	0.07	0.10	0.12	0.13

Distributions of Lapel collar and Straight/Tight/Loose pants in FSS



Samples of lapel tops matched with straight bottoms

Fall-2013-ready-to-wear Brand:Acne Studios

Spring-2014-menswear Brand:Ovadia & Sons

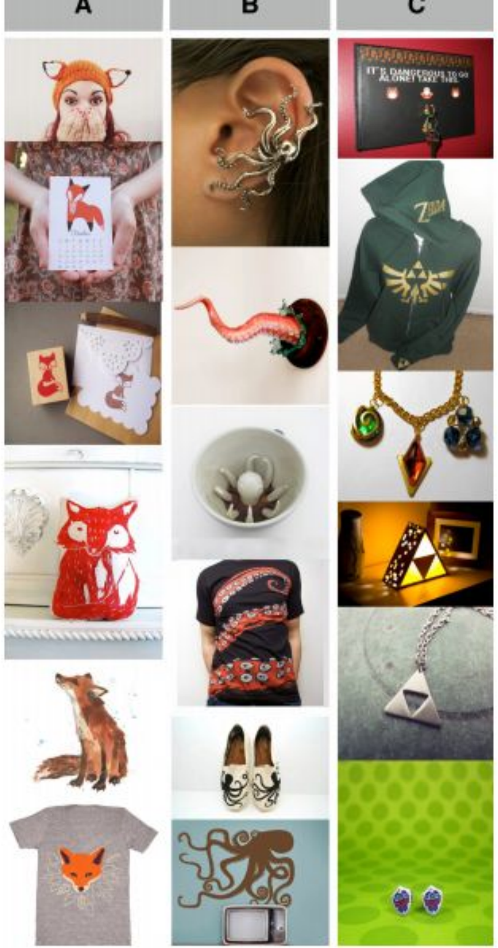
Fall-2015-ready-to-wear Brand:Hood by Air

Tokyo-spring-2016 Brand:Discovered

- **Top collar/bottom fashion**
- **Straight pants**
  - almost match every collar
  - Most similar shape with lapel
- **Tight pants**
  - fur/high/round/bateau good choice
  - Bateau better than with straight

Flipkart

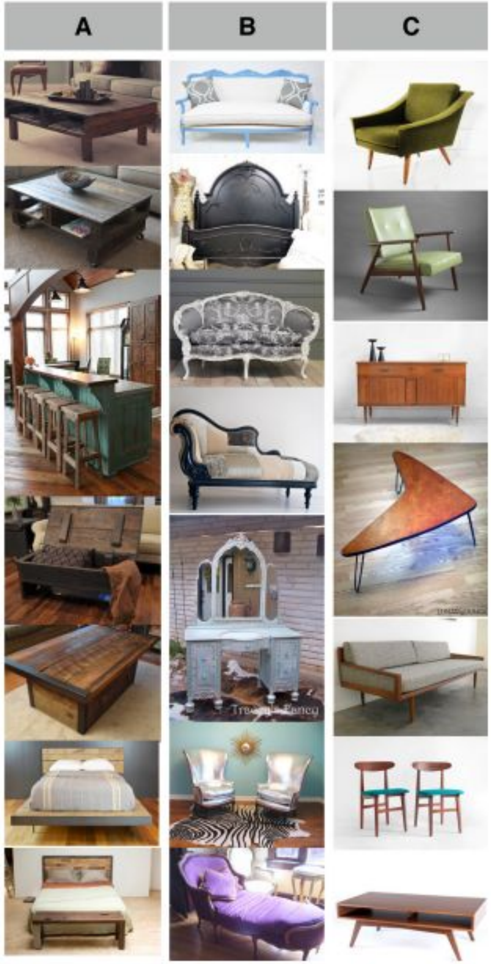
[Jia 17] Understanding clothing fashion styles: deep learning approach, AAAI



Styles across categories

- Handmade/vintage items for sale
  - Make Etsy unique
- Interests span
  - Clothing, jewelry, home decor
  - Themes: fox, cephalopod, legend of Zelda
- Detection with no image cues
  - Social-network based user behavior

[Hu 14] Style in long tail: discovering unique interests in social e-commerce, KDD

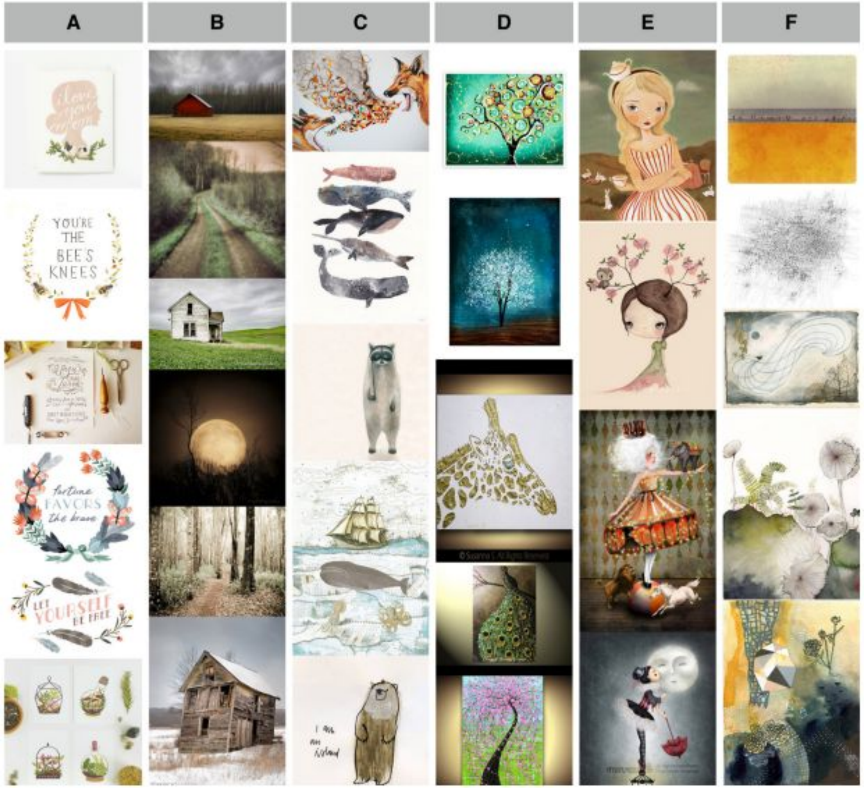


Style in furniture category

- Rustic-wooden, french-cottage, midcentury-modern (antique, industrial)
- Listing with highest weight within each topic indicative of group style
- *Favoriting* listings reliable signal for style
  - Purchase only a small, lower-priced subset
  - No overlap in vectors for users with similar interests

Flipkart

[Hu 14] Style in long tail: discovering unique interests in social e-commerce, KDD



Item ranking by style in art category

- **Visualizing helps understand different user-interest clusters**
  - Botanical-hand-lettered
  - Haunting landscape photography
  - Whimsical animal illustrations
  - Abstract paintings
  - Fairytale doll prints
  - Whimsical painting

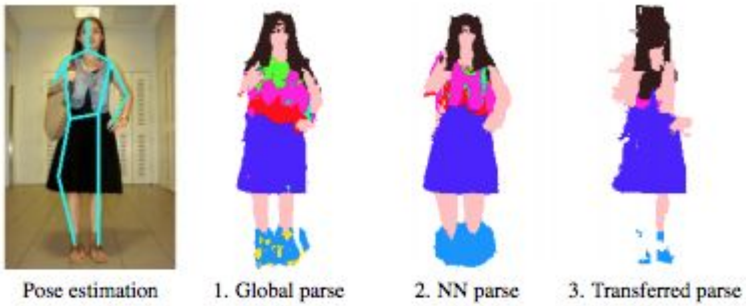
[Hu 14] Style in long tail: discovering unique interests in social e-commerce, KDD

Flipkart

Summary - style detection

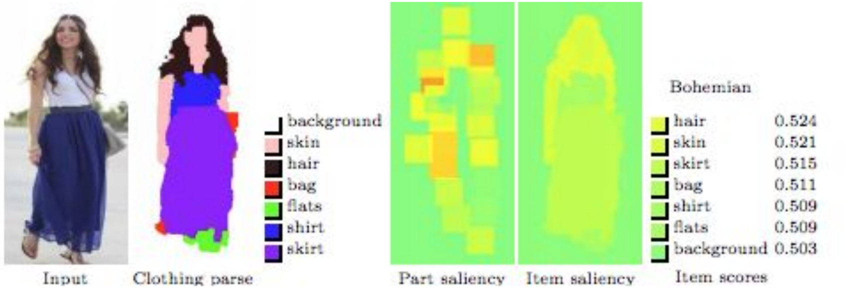
CV-based approaches

[Berg 12] parsing with metadata/garment tags and unconstrained label sets



[Berg 13] parsing based on nearest-neighbor style retrieval

[Berg 14] estimate degree of and item which indicates outfit style



Flipkart

Big data based approaches

[Mcauley 15a] Visual relationship is beyond similarity - complementarity too

[Piramuthu 13] attribute-tag enabled clothing search with style and description generation

Attribute type	Attributes
Material	Fur, Denim, Leather/is leather-like, Shiny, Wool/is woolen or felt-like
Fastener	Zip, Button, Open, Has belt
Fastener style	Symmetrical (single-breasted), Asymmetrical (single-breasted), Asymmetrical (double-breasted)
Length	Short, Medium, Long
Cut	Fitted, Loose
Pocket	Chest pocket, Side pocket
Collar	V-neck collar, Round collar, Turtle neck, V-neck shirt collar, Round shirt collar, Notched collar, Shawl collar, Peak collar

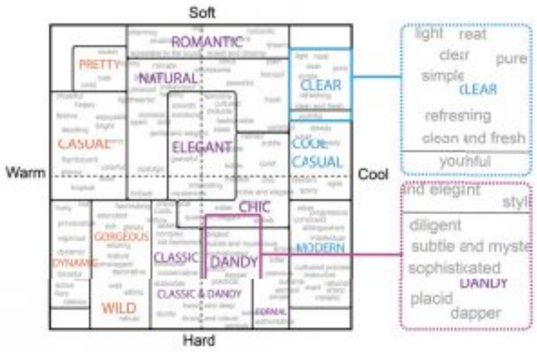
[Mcauley 16] Capture temporal drifts of fashion/personal tastes

[Mcauley 15a] recover style space for clothing items from co-occurence information and category label

Flipkart

Deep learning approaches

[Liu 17] Style obtained subtracting category from visual features



[Ma 17] fashion semantic space

[Nakamura 18] outfit sequence/style learnt together for typical generation

Identify style icon :-)







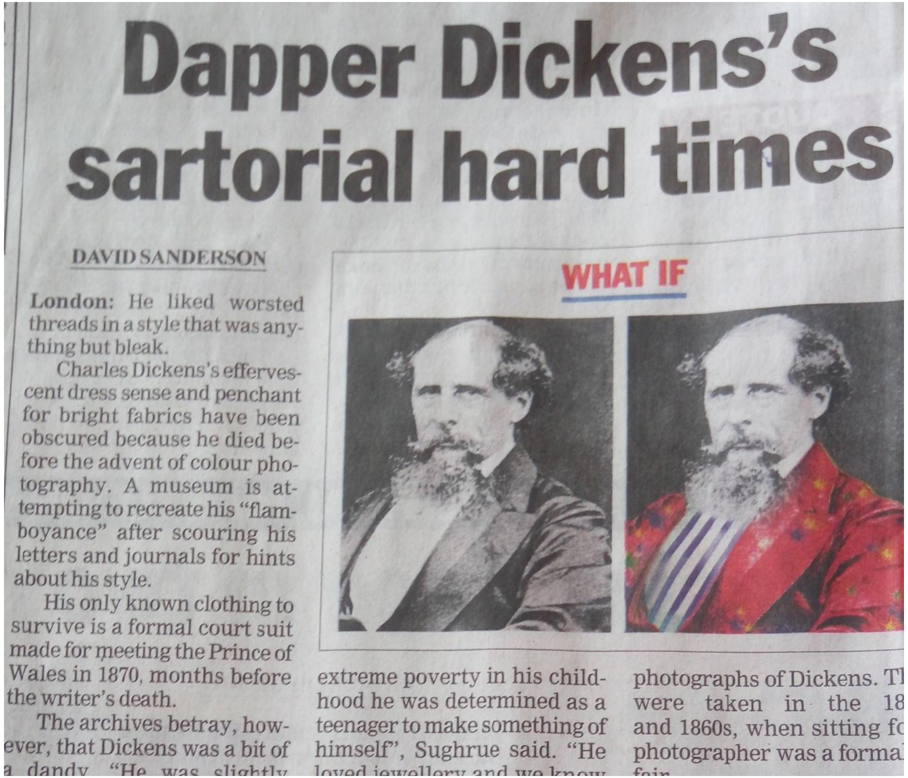





Flipkart

Fashion Recommendation

Prelude



Dapper Dickens's sartorial hard times

DAVID SANDERSON

London: He liked worsted threads in a style that was anything but bleak.

Charles Dickens's effervescent dress sense and penchant for bright fabrics have been obscured because he died before the advent of colour photography. A museum is attempting to recreate his "flamboyance" after scouring his letters and journals for hints about his style.

His only known clothing to survive is a formal court suit made for meeting the Prince of Wales in 1870, months before the writer's death.

The archives betray, however, that Dickens was a bit of a dandy. "He was slightly

WHAT IF

extreme poverty in his childhood he was determined as a teenager to make something of himself", Sughrue said. "He loved jewellery and we know

photographs of Dickens. Th were taken in the 18 and 1860s, when sitting fo photographer was a formal

Is there a Working Definition?

- Hardly any general consensus on the term “Fashion Recommendation”
- Various types of possible recommendations
- One size fits all?
- Factors that impact individual needs and comfort levels
- "Aap ruchi Khana, Par ruchi parna"... Eat so as to please yourself, Dress so as to not displease others...

What people have done so far

- Hi, Magic Closet, Tell Me What to Wear!, ACM MM 2012
- Large Scale Visual Recommendations From Street Fashion Images, KDD 2014
- Image-based Recommendations on Styles and Substitutes, SIGIR 2015
- Collaborative Fashion Recommendation: A Functional Tensor Factorization Approach, ACM MM 2015.
- Brain-Inspired Deep Networks for Image Aesthetics Assessment, Michigan Law Review 2016
- An Approach for Clothing Recommendation Based on Multiple Image Attributes, WAIM 2016

What people have done so far

- Visually-Aware Fashion Recommendation and Design with Generative Image Models,. ICDM 2017
- Learning Disentangled Multimodal Representations for the Fashion Domain, WACV 2018
- Styling with Attention to Details, arXiv 2018
- Aesthetic-based Clothing Recommendation, WWW 2018.
- DeepWear: a Case Study of Collaborative Design between Human and Artificial Intelligence, TEI 2018
- One for the Road: Recommending Male Street Attire, PAKDD 2018

Grouping of Existing Work

- Flavor of the Problem
  - Individual recommendation
  - How you want to look – Individualistic/Go with the crowd
  - Image generation
  - Fashion ensembles
  - Taking care of aesthetics
  - Use of basic color, texture features
  - Streetwear or special occasion
  - Handling occasion, season and even time of the day

Grouping of existing Work

- Tools used
  - Collaborative filtering
  - Tensor factorization in various forms
  - Bayesian Personalized Ranking
  - CNN
  - Siamese Networks
  - GAN

Identified Primary Challenges

- Every individual potentially has a different perception of fashion. How can that be captured?
- Does an individual want to look different or go with the crowd or make a fashion statement?
- Does the user have a budget in mind? Do we consider it implicitly or explicitly?
- Do we recommend a single item or an outfit?
- What are the optimization criteria?
- Cold start problem
- What else?

A Brief Look into a Couple of Specific Approaches

- One for the Road: Recommending Male Street Attire: Banerjee et al., PAKDD 2018
- Visually-Aware Fashion Recommendation and Design with Generative Image Models, Kang et al., ICDM 2017

One for the Road: Recommending Male Street Attire: Banerjee et al., PAKDD 2018

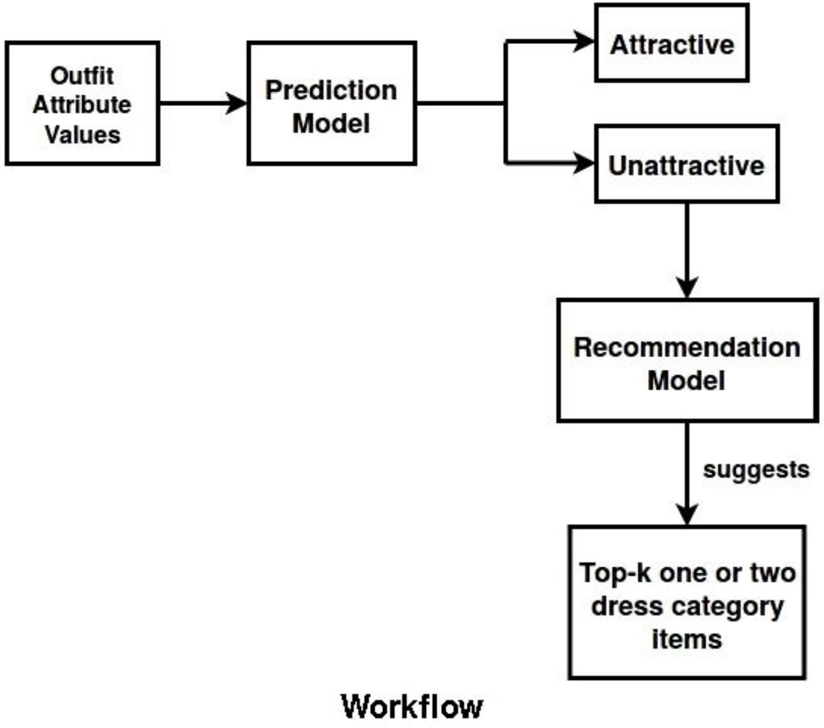
- Women are generally believed to dominate the fashion world
- Generation Y male is highly fashion conscious
- Male street fashion has become a major part of the growing fashion industry
- Efficient recommendation frameworks required to cater to male fashion needs

Female Clothing Fashion

Male Clothing Fashion

Images courtesy: Banerjee et al., PAKDD 2018

One for the Road: Recommending Male Street Attire: Banerjee et al., PAKDD 2018



Images courtesy: Banerjee et al., PAKDD 2018

One for the Road: Recommending Male Street Attire: Banerjee et al., PAKDD 2018

Collected street photos of men by crawling online fashion articles, and social media sites[1]

- Cropped pictures of individual men (having full frontal view) with dimension set as 400×255 pixels
- Pixelated faces (excluding sunglasses or caps/hats), and background set as white (gray in some cases)

(a) (b)

Sample Image: a) Before, and b) After Preprocessing

Note: After preprocessing, 824 images were shortlisted.

[1] www.fashionbeans.com, www.facebook.com, www.instagram.com

Images courtesy: Banerjee et al., PAKDD 2018

One for the Road: Recommending Male Street Attire: Banerjee et al., PAKDD 2018

Images courtesy: Banerjee et al., PAKDD 2018

One for the Road: Recommending Male Street Attire: Banerjee et al., PAKDD 2018

Survey 1

Each outfit tagged either **attractive** or **unattractive**

- Attractive - 409
- Unattrative - 415

Survey 2

Categorized each outfit into three **outfit types**:

- Conformative - 138
- Individualistic - 273
- Average - 413

http://debopriyo.epizy.com/annotation/attractive_outfit_slide_1.php

One for the Road: Recommending Male Street Attire: Banerjee et al., PAKDD 2018

Feature Identification

- 33 categorical, 11 binary predictor attributes, and binary target variable (attractive=1/unattractive=0)
- 25 discriminative predictor variables were obtained after feature selection process[a,b]

[a] M. A. Hall (1999). *Correlation-based Feature Selection for Machine Learning.*

[b] R. Kohavi and G. H. John (1997). *Wrappers for feature subset selection.* Artificial Intelligence, 97(1): 273-324.

Chosen predictors: **body structure**, **age**, **F**[color1, type], **L**[material, shape, color2], $\mathbf{TB_1}$[color2, type, patterns], $\mathbf{TB_2}$[color2, type, patterns], **TC**[color1, color2, type, patterns], **head**[color2, type], **suit**[type], **accessories**[scarf, bag, back pack, tie], **outfit type**

Models used for Predicting Attractive Outfits

- Random Forest
- Support Vector Machine

One for the Road: Recommending Male Street Attire: Banerjee et al., PAKDD 2018

Considering recommendation is required for **feet**:

Target Node

Here a target node would comprise of **F** and its corresponding attributes (**F**[color1, type]).

Non-Target Node

In this case, non-target nodes will be of the form **Z** = {**X**, **Y**}, where **X** = {*age* or *body structure*}, and
Y is either one of the attribute in {**L**[material], $\mathbf{TB_1}$[type], $\mathbf{TB_2}$[type], **TC**[type]} or all of them taken together.

One for the Road: Recommending Male Street Attire: Banerjee et al., PAKDD 2018

- MalOutRec has been evaluated with the set (PR_{uat}) of outfits predicted as unattractive
- For every outfit $o_i \in PR_{uat}$, MalOutRec suggests top-k (k=3, 5, and 10) set of items from one or two dress categories
- o_i is modified to $\hat{o}_i$ with the recommended items
- Prediction model then verified whether $\hat{o}_i$ looked attractive or not

Images courtesy: Banerjee et al., PAKDD 2018

One for the Road: Recommending Male Street Attire: Banerjee et al., PAKDD 2018

Hit Rate or HR@k[4] has been used to evaluate MalOutRec

$$Hit\ Rate = \frac{\#hits}{N}$$

where #hits is the number of cases when at least one of the top-*k* recommended items transform the unattractive outfit into attractive, and N is the cardinality of the test set.

Input Outfit o_i

Recommended top-3 feet items

[4] D. Lee, and et al. (2010). *Exploiting Contextual Information from Event Logs for Personalized Recommendation.* In Computer and Information Science, Studies in Computational Intelligence, pages 121-139.

Images courtesy: Banerjee et al., PAKDD 2018

Visually-aware Fashion Recommendation and Design with Generative image Models

- Joint Image Representation and Recommendation
- New Design using GAN

Images courtesy: Kang et al., ICDM 2017

Visually-aware Fashion Recommendation and Design with Generative image Models

- The work has primarily two parts
  - Visual Recommendation – A combination of Siamese CNN and Bayesian Personalized Ranking
  - Image generation using GAN
- GAN
  - One component trained to generate images
  - Another component to distinguish generated images from real images
  - Idea is to ensure new images are generated having distribution similar to database images
- Used Amazon Dataset as well as data crawled from Tradesy.com

Visually-aware Fashion Recommendation and Design with Generative image Models

Generated image samples and their ℓ_1 nearest neighbors in the dataset (ℓ_1 and ℓ_2 nearest neighbors were equivalent in almost all cases). Note that all images are rescaled to be square during preprocessing.

Images courtesy: Kang et al., ICDM 2017

Visually-aware Fashion Recommendation and Design with Generative image Models

Top-3 results from the dataset and GAN. Each row is a separate retrieval/optimization process for a given user and product category. Left: real images; right: synthetic images. The values shown are preference scores for each image.

Images courtesy: Kang et al., ICDM 2017

Points to Ponder

- Application domain
  - Who gets benefited?
- How to specify the input
- What do you get as output
- Usability and Flexibility
- Level of personalization
- What do Gen Z people need?

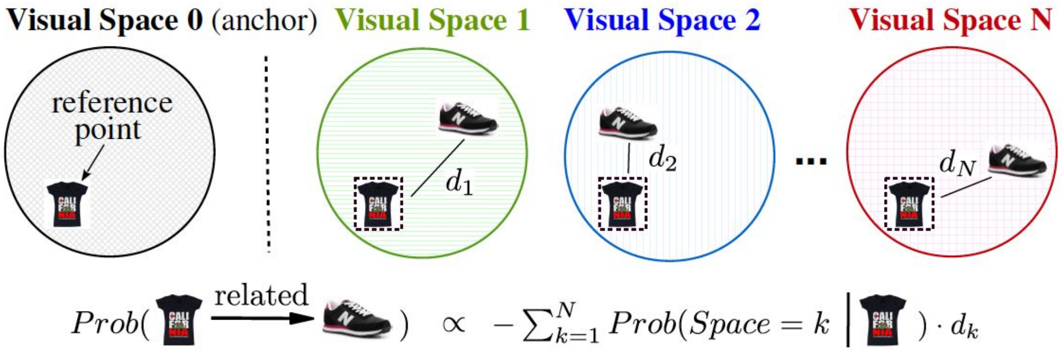
Outfit Compatibility

Complex/heterogenous relationships

- Compatible: Jeans and t-shirts, laptop and charger
  - Similar in someway
  - Systematically different
    - Dimensions not to be ignored
- Huge amount of relationships
  - High dimensional
  - Semantically complicated features
  - Transitivity: iphone, surface, ipad
  - Shirts of same color/brand, pants, shoes
- Learn relatedness as mixture of competing notions
  - Independent of category tree existene

[Mcauley 16] Learning Compatibility Across Categories for Heterogeneous Item Recommendation, ICDM

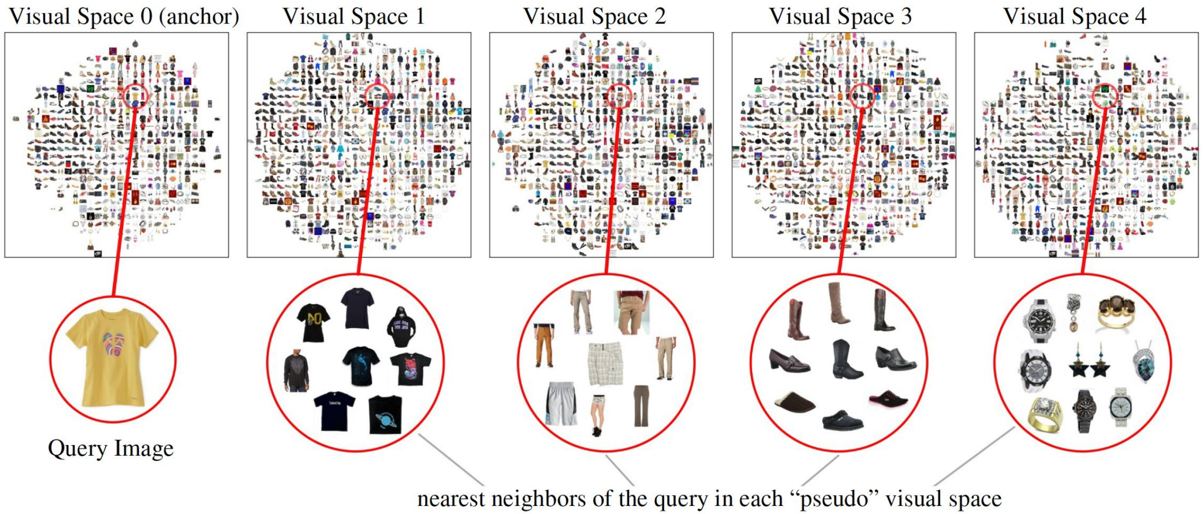
Monomer: Mixtures of Non-Metric Embeddings for Recommendation



- The query item (a t-shirt) is embedded into visual space 0
- The potential match (a shoe) is embedded to N visual spaces
- Compute Euclidean distance between the pair
- Model the relative importance of the different components w.r.t. the given query (use mixtures-of-experts framework)

[Mcauley 16] Learning Compatibility Across Categories for Heterogeneous Item Recommendation, ICDM

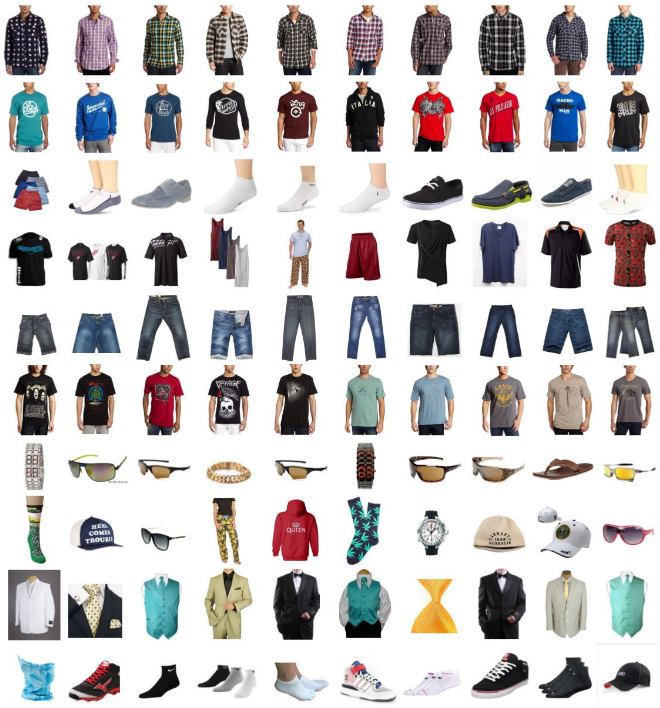
Visualisation of Monomer:



- Women's Clothing for 'also bought' prediction
- Each visual space is demonstrated by a 2-d grid view
- Recommend the nearest neighbors of query based on the associated 'reasons' learned from data (e.g., color, pattern, style).
- Each visual space exhibits different category 'clusters' at the query image's location
- Recommend diverse sets of items from the most closely-related categories

[Mcauley 16] Learning Compatibility Across Categories for Heterogeneous Item Recommendation, ICDM

Visual space learned on Men's Clothing



- Demonstration of the 10 visual dimensions of one (randomly selected) visual space learned
- Each row shows the top ranked items for a particular dimension
  - Fine-grained (1 - plaid, 5 - jeans)
  - Human notion (2 - casual, 9 - formal)
  - Subtle (2,6)

[Mcauley 16] Learning Compatibility Across Categories for Heterogeneous Item Recommendation, ICDM

Comparison with Monomer and State of Art

- With Low-rank Mahalanobis Transform (LMT)
- Left: query images
- Above line: nearest neighbors retrieved by LMT
- Below line: nearest neighbors retrieved by Monomer

[Mcauley 16] Learning Compatibility Across Categories for Heterogeneous Item Recommendation, ICDM

Flipkart

Outfit Composition with accumulated knowledge

(a) Composition1. (b) Composition2. (c) Composition3.

- Neural networks for fashion item representation suffers from poor interpretability
- Use rich valuable knowledge (rules) to boost performance
  - Tank tops with shorts - not dress
  - Silk tops avoid knit bottoms

[Song 18] Neural Compatibility Modeling with Attentive Knowledge Distillation, SIGIR

Complementary clothing matching

(a) Example1. (b) Example2.

Rule confidence on different item pairs varies. Both satisfy the rule "stripe tops can go with stripe bottoms"- but (a) is better

E

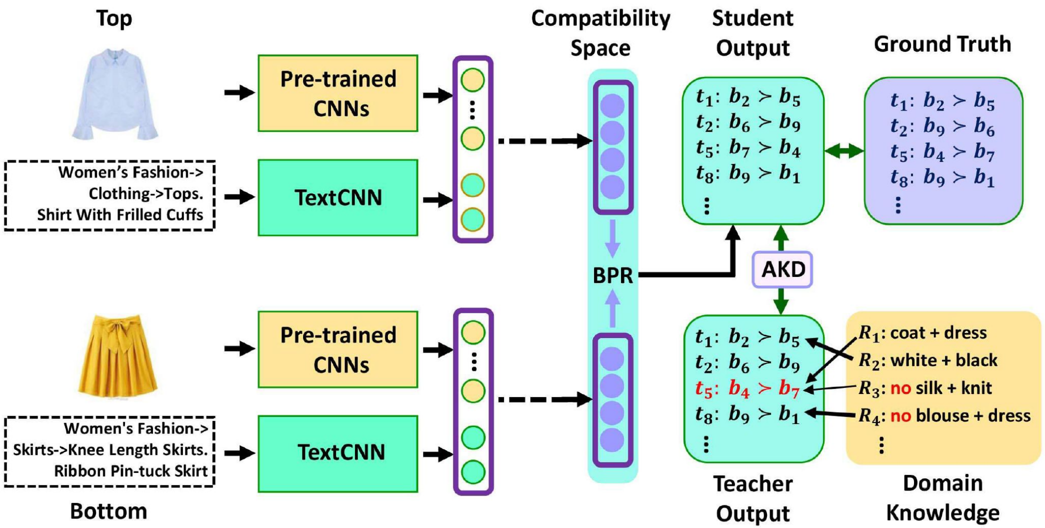
ncode fashion domain knowledge to traditional neural networks (using attentive knowledge distillation scheme)

Flexibly assign rule confidence (using attention mechanism)

Flipkart

[Song 18] Neural Compatibility Modeling with Attentive Knowledge Distillation, SIGIR

AKD-DBPR : Attentive Knowledge Distillation



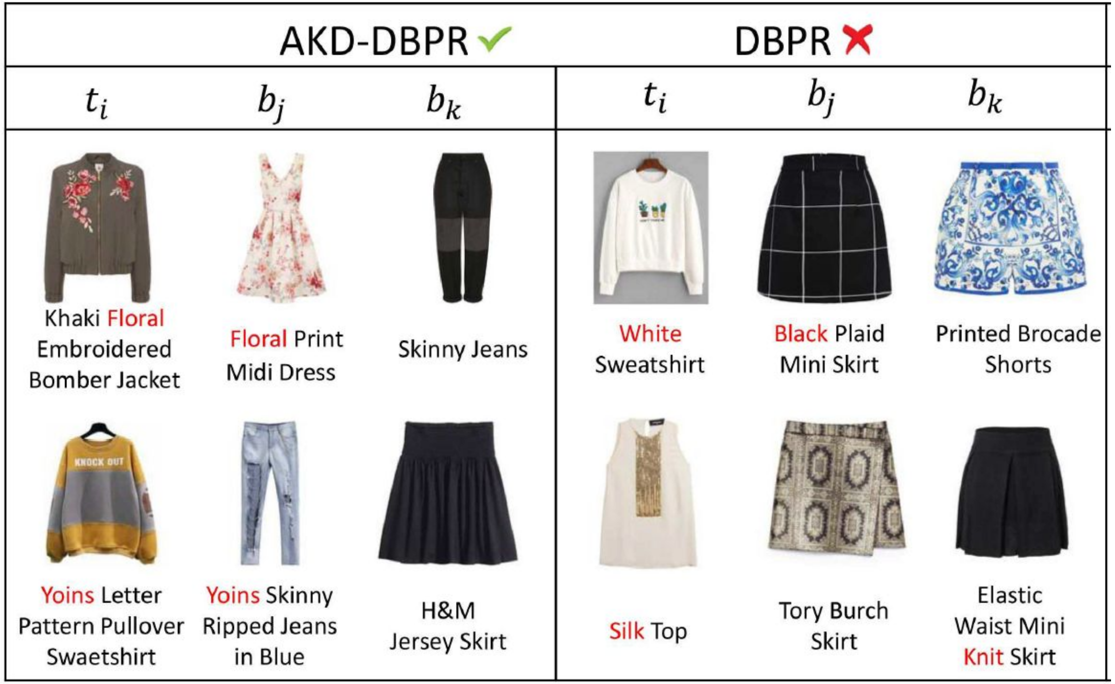
- *t_i: top, b_j : bottom, "$>$": pair-wise preference.*
- *"->" denotes the category hierarchy.*
- *Width of arrows originated from rules refers to confidence.*

- Learn from both specific data samples and general domain knowledge
- Adopt teacher-student scheme to incorporate domain knowledge (as a teacher) and enhance performance of neural networks (as a student).
- Student network, consisting of dual-path neural networks, aims to learn latent compatibility space
- Implicit preference among items can be modeled via Bayesian Personalized Ranking.
- Teacher network encodes domain knowledge and guides student network via attentive knowledge distillation.

Flipkart

[Song 18] Neural Compatibility Modeling with Attentive Knowledge Distillation, SIGIR

AKD-DBPR : Baseline comparison



- AKD-DBPR performs better than DBPR (baseline) in cases when the given two bottoms b_j and b_k both seem to be visually compatible to the top *ti*.

[Song 18] Neural Compatibility Modeling with Attentive Knowledge Distillation, SIGIR

Rule Confidence

Activated Rules	λ
No silk + knit	0.22
No silk + chiffon	0.78

Activated Rules	λ
No blouse + dress	0.66
No silk + knit	0.34

Activated Rules	λ
Tank top + short	0.63
Stripe + stripe	0.37

- Different confidence assigned for same rule
  - No silk + knit
- Category rules assigned higher confidence
  - E.g., Blouse, short
  - This attribute valued more
- Visual signals override fuzzy rules
  - Stripe + stripe

[Song 18] Neural Compatibility Modeling with Attentive Knowledge Distillation, SIGIR

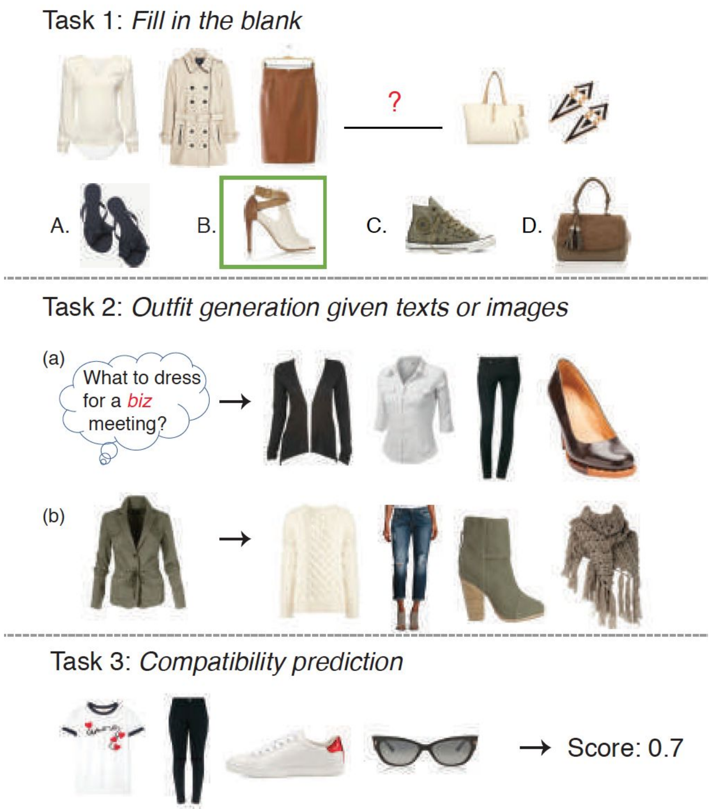
Ranking Results

Testing Tops

- bottoms highlighted in red boxes are positive ones
- First example activates rules
  - “floral + floral” and
  - “coat + dress”,
- Second example triggers rule
  - “white + black”
- Which may contribute to good performance of AKD-DBPR

[Song 18] Neural Compatibility Modeling with Attentive Knowledge Distillation, SIGIR

Compatibility tasks



- Suggest an item that matches existing components in a set to form a stylish outfit (a collection of fashion items)
- Generate an outfit with multimodal (images/text) specifications from a user
- Predicting compatibility of an outfit

[Han 17] Learning Fashion Compatibility with Bidirectional LSTMs, Multimedia

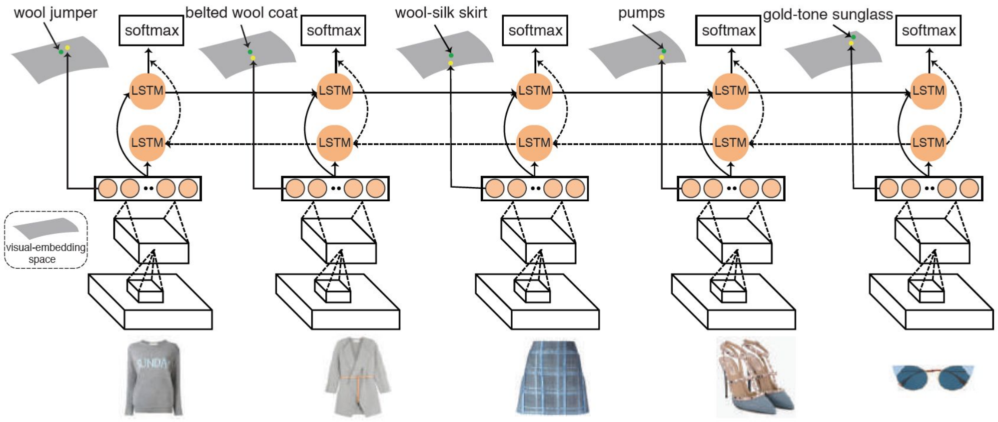
Sequence learning

- Polyvore - fashion image with corresponding descriptions
- Consider a fashion outfit to be a sequence (usually from top to bottom and then accessories) and each item in the outfit as a time step
- Learn a visual-semantic space by regressing image features with attribute and category information

Flipkart

[Han 17] Learning Fashion Compatibility with Bidirectional LSTMs, Multimedia

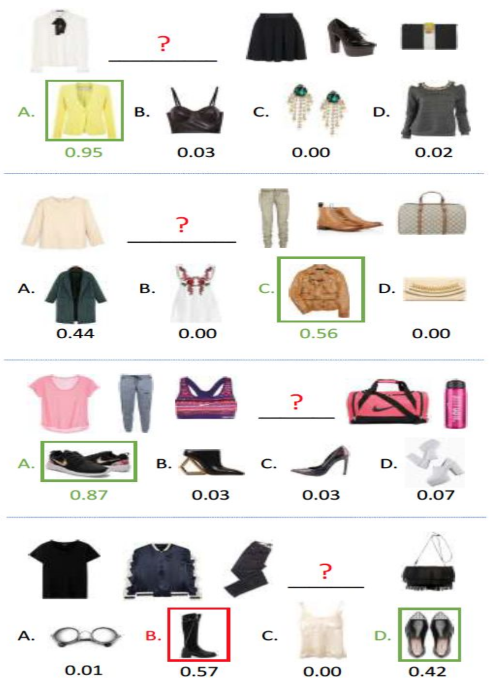
Compatibility semantics



- Treat a given outfit as a sequence of fashion items (jumper, coat, skirt, pumps, sunglasses).
- Build a bidirectional LSTM to sequentially predict next item conditioned on previously seen items in both directions.
- Given jumper and coat, predict skirt
- Embedding for regularization

[Han 17] Learning Fashion Compatibility with Bidirectional LSTMs, Multimedia

Fill in the blank task



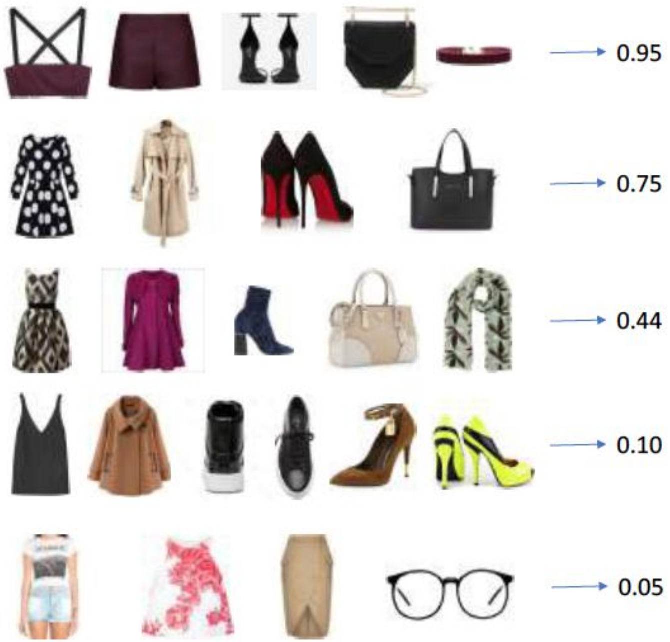
- Not only detect what kinds of fashion item is missing(e.g.,coat in 2nd example)
- But also select fashion item most compatible to query and match their style as well
  - (e.g., running shoes more compatible with sporty outfit in 3rd example)

- *Green bounding boxes indicate correct answers*
- *Red box shows a failure case*
- *Prediction score of each choice is also displayed.*

Flipkart

[Han 17] Learning Fashion Compatibility with Bidirectional LSTMs, Multimedia

Predict Outfit Compatibility

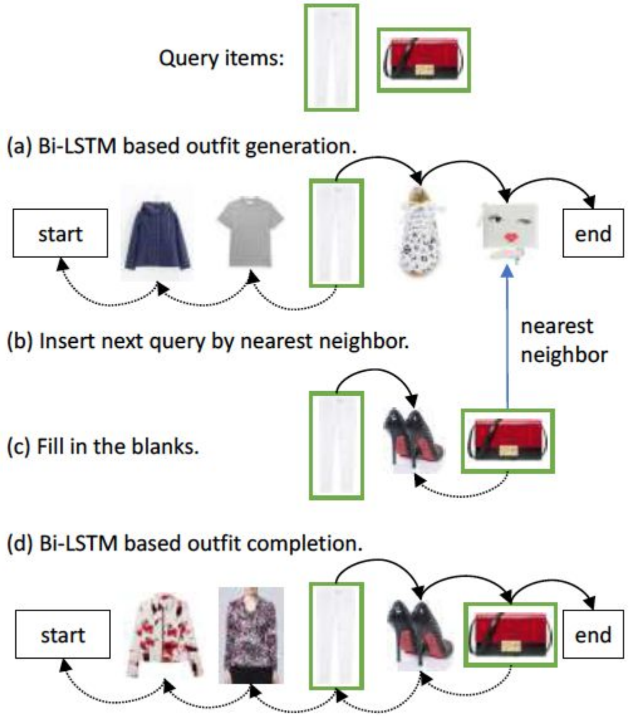


- 1st row contains purple/black items with same style and thus has a high compatibility score
- All items in 3rd row have different colors, which makes them somewhat incompatible to form an outfit
- Fourth outfit contains 4 pairs of shoes without a bottom, and last one contains two dresses but no shoes; thus they are both incompatible outfits

Flipkart

[Han 17] Learning Fashion Compatibility with Bidirectional LSTMs, Multimedia

Generate Outfits from Query Image



- a) Users provide a single image and wish to obtain an entire outfit with consistent style.
  - Run trained Bi-LSTM in two directions
- b) Users provide more than one item, utilize first item to generate an initial outfit and then find and replace the nearest neighbor of the next query item
- c) Inference in both directions when 2 items are contiguous, ensures subsequence used to generate entire outfit is visually compatible
- Items compatible with white pants, casual initially but becomes formal with shoes matching red/black handbag

Flipkart

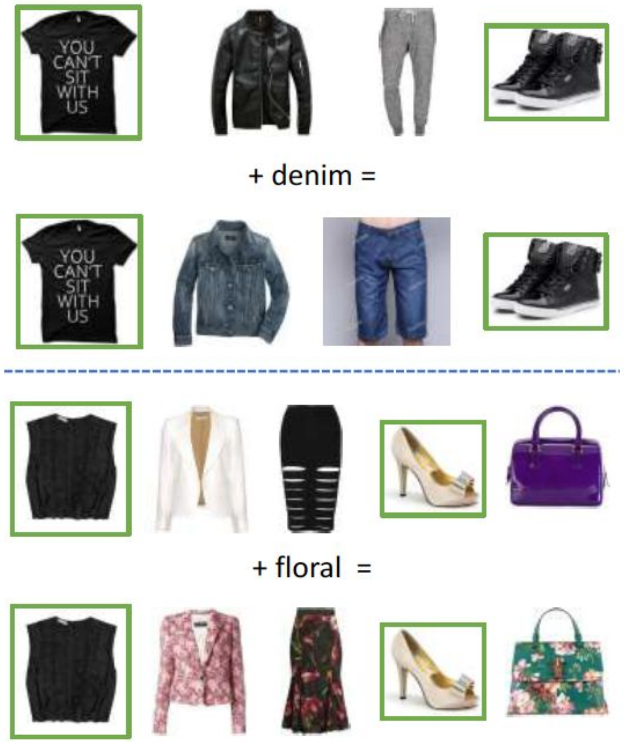
[Han 17] Learning Fashion Compatibility with Bidirectional LSTMs, Multimedia

Fashion outfit - given query items

- Each row contains a recommended outfit where query images are indicated by green boxes
- Produces visually compatible and complete outfits

[Han 17] Learning Fashion Compatibility with Bidirectional LSTMs, Multimedia

Fashion outfit - given query items and text input



- Query images are indicated by green boxes.
- Outfits on top are generated without using text input.
- When a text query is provided outfits are adjusted accordingly.

[Han 17] Learning Fashion Compatibility with Bidirectional LSTMs, Multimedia

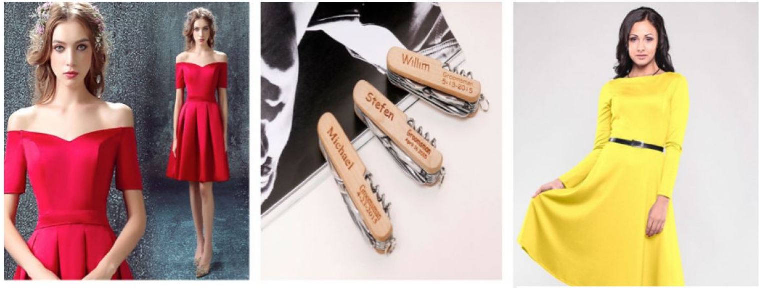
Fashion outfit - text input

- An attribute or style that all items are expected to share (e.g., denim, casual)
  - Nearest image to text query is chosen as query image
- Descriptions of fashion items generated outfit should contain (e.g., lace dress + red pump).
  - A fashion item image is retrieved using each description, and all images are treated as query images to generate the outfit

Flipkart

[Han 17] Learning Fashion Compatibility with Bidirectional LSTMs, Multimedia

Images Don’t Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank



“Red Short dress, Prom Dress, **wedding dress**, dress, …”

“Pocket Knife wedding shower ideas **wedding dresses**, beach …”

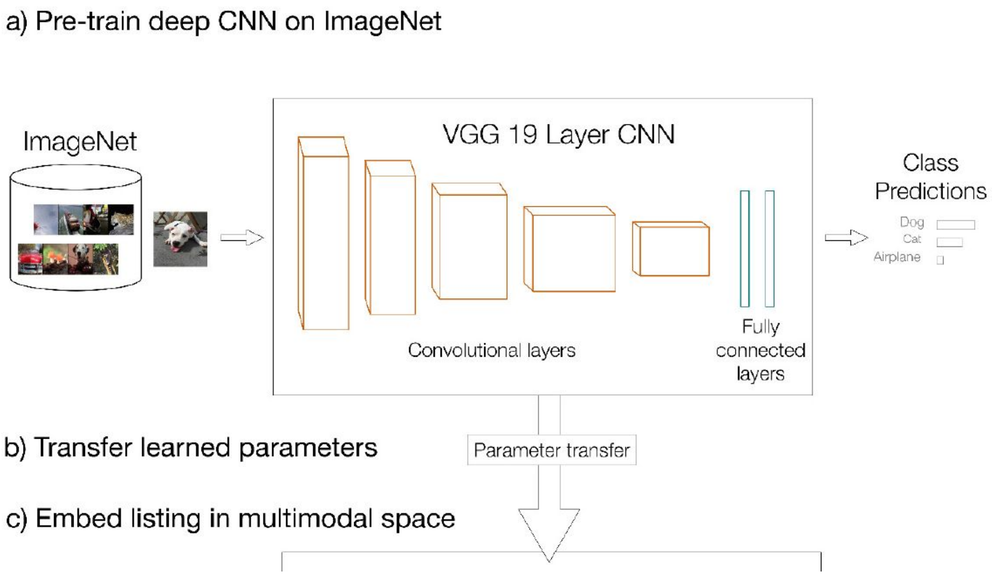
“Yellow dress. Retro dress **Wedding dress**. Flared skirt…"

Irrelevant search results for the query “wedding dress"
Even though it’s apparent in the images that these are not wedding dresses, each listing’s descriptive title contains the phrase “wedding dress", allowing it to show in search results for the query.

Flipkart

[Lynch 16] Images Don’t Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank

Images Don’t Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank

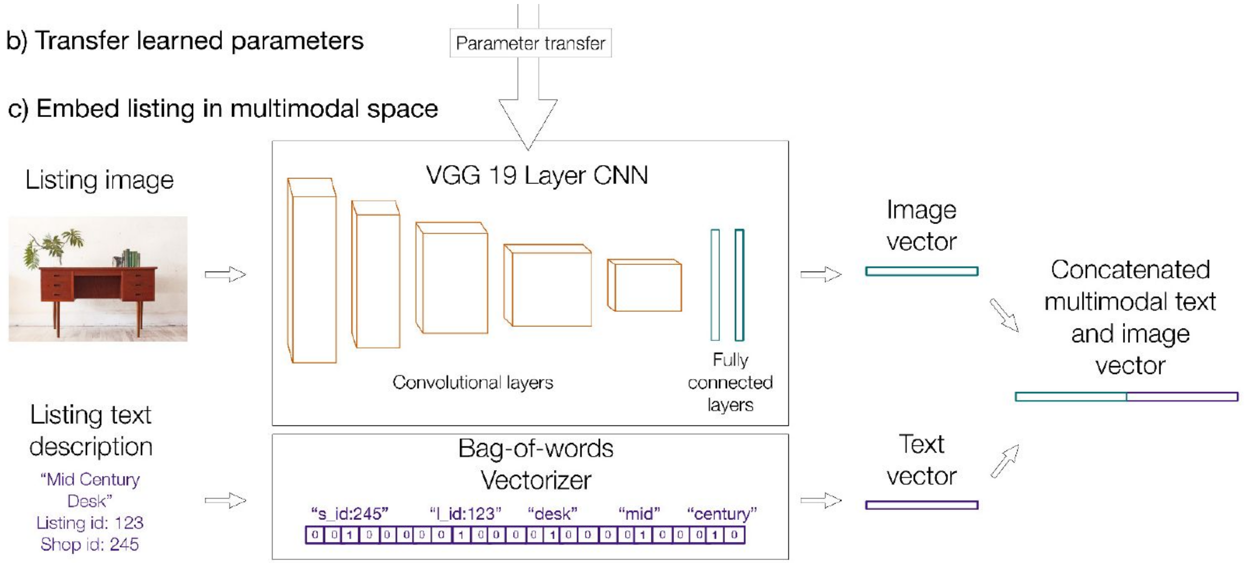


Transferring parameters of a CNN to the task of multimodal embedding:

- Utilize a pre-trained 19 layer VGG-style network that is trained on a large scale object recognition task (ImageNet challenge)
- Remove the last layer (containing scores for the different object classes) and transfer the parameters of the modified network to our task.

[Lynch 16] Images Don’t Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank

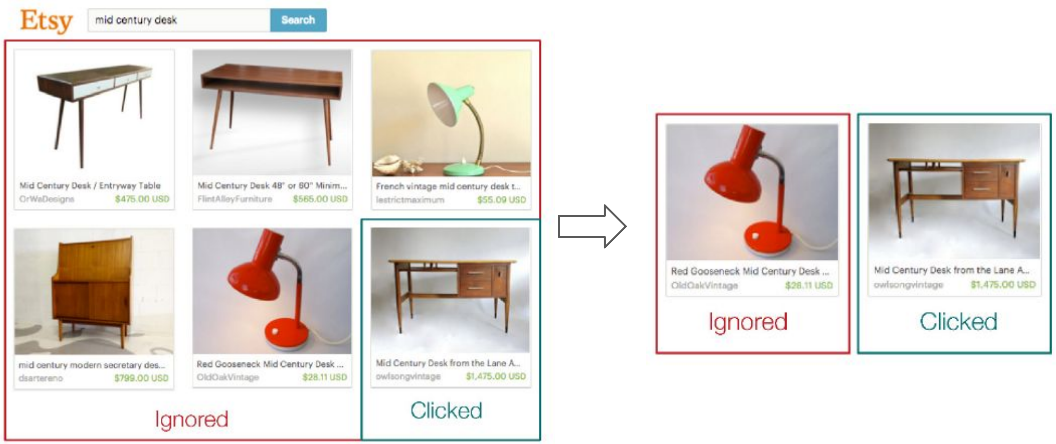
Images Don't Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank



- Use the modified network as a fixed feature extractor of high-level image content information, taking the last fully connected layer as an image embedding.
- Simultaneously embed the listing's text in a bag of words space, then concatenate the two embeddings to form a single multimodal descriptor of a listing.

[Lynch 16] Images Don't Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank

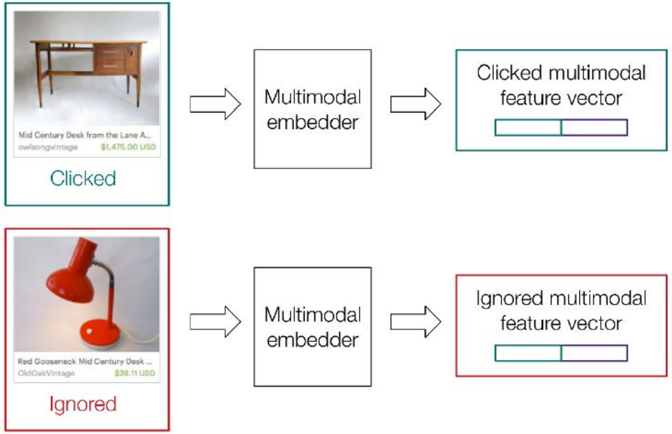
Images Don’t Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank



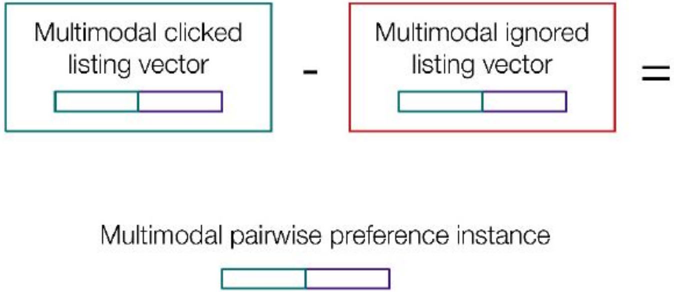
From search logs to multimodal pairwise classification instances:
- A user comes to the site, enters the search query mid century desk, is presented with some results, clicks one and ignores the others. Take the listing she clicked and an adjacent listing she ignored as an instance of implicit pairwise preference in the context of the query mid century desk forming a training triplet from (mid century desk, clicked listing, ignored listing).

[Lynch 16] Images Don’t Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank

Images Don't Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank



Embed both listings in the pair in multimodal space.



Map labeled embedded pair to a single pairwise classification instance. Flip a coin.
If heads, create a well-ordered pairwise instance (clicked vector - ignored vector) and label it +1
If tails, create a non-well-ordered pairwise instance (ignored vector - clicked vector) and label it -1

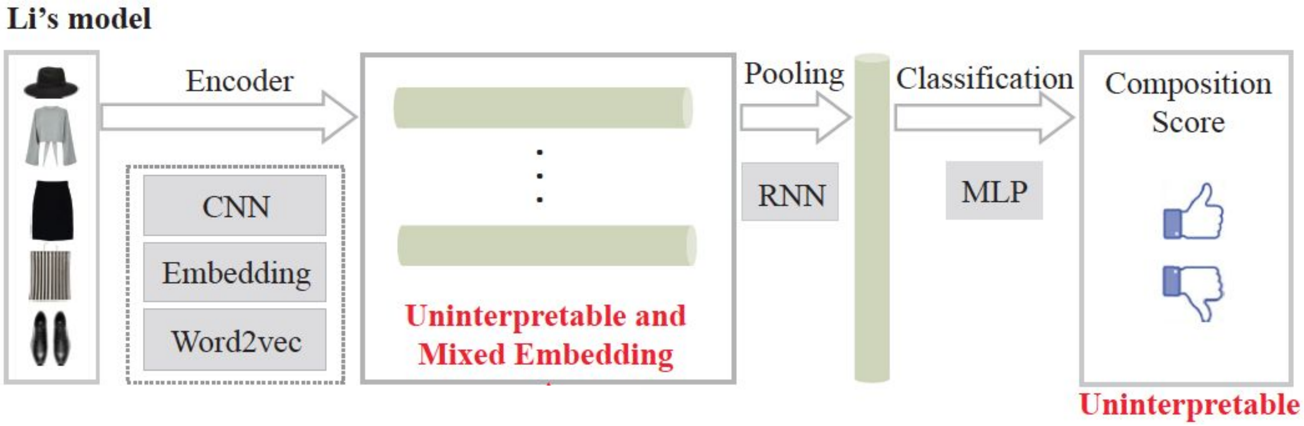
[Lynch 16] Images Don't Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank

Images Don’t Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank

Visualizing ranking changes by incorporating image information

[Lynch 16] Images Don’t Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank

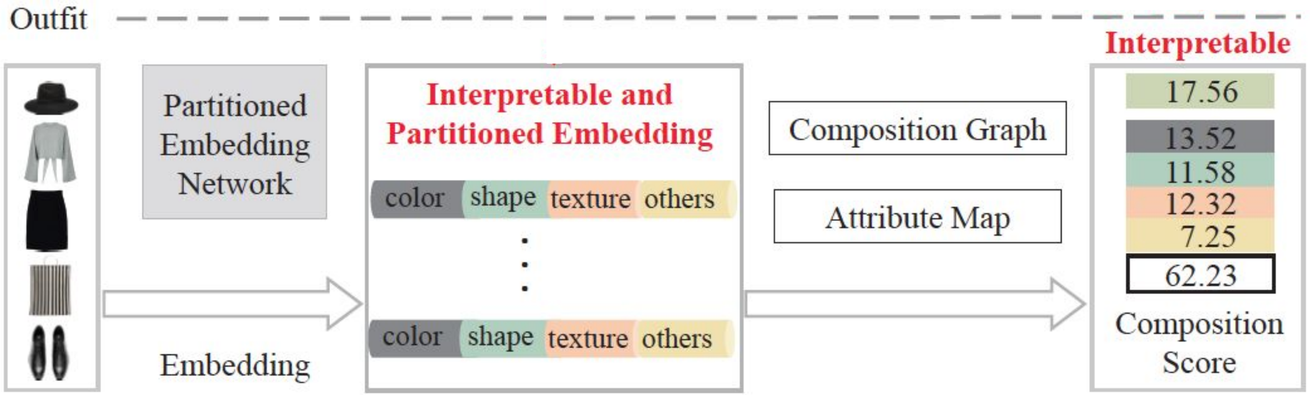
Interpretable Partitioned Embedding for Customized Multi-item Fashion Outfit Composition



- Deep learning based approaches are uninterpretable
- Cannot meet the designers, businesses and consumers' urge to comprehend the importance of different attributes in an outfit composition

[Feng 18] Interpretable Partitioned Embedding for Customized Multi-item Fashion Outfit Composition

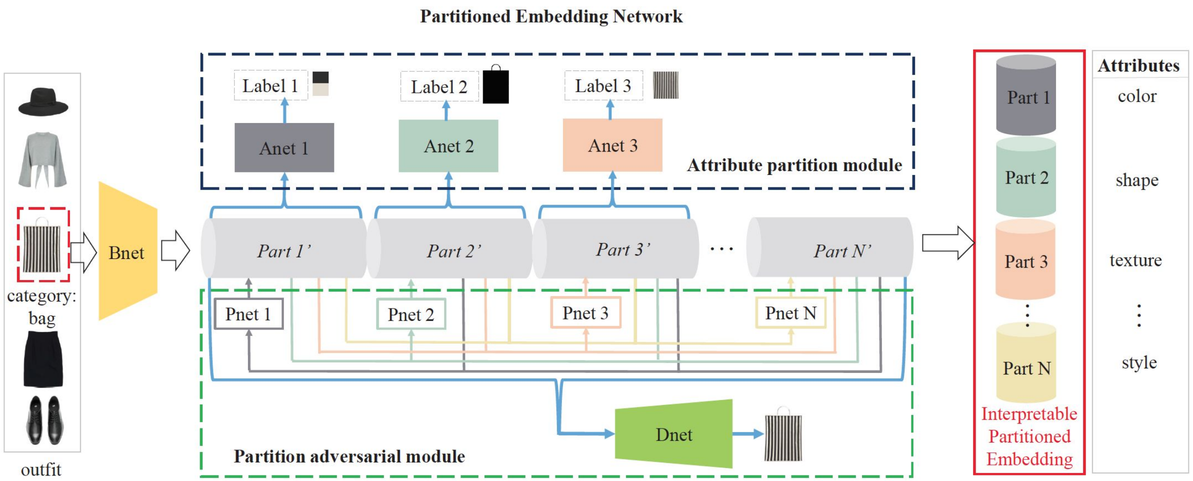
Interpretable Partitioned Embedding for Customized Multi-item Fashion Outfit Composition



- A partitioned embedding network to learn interpretable embeddings from clothing items

[Feng 18] Interpretable Partitioned Embedding for Customized Multi-item Fashion Outfit Composition

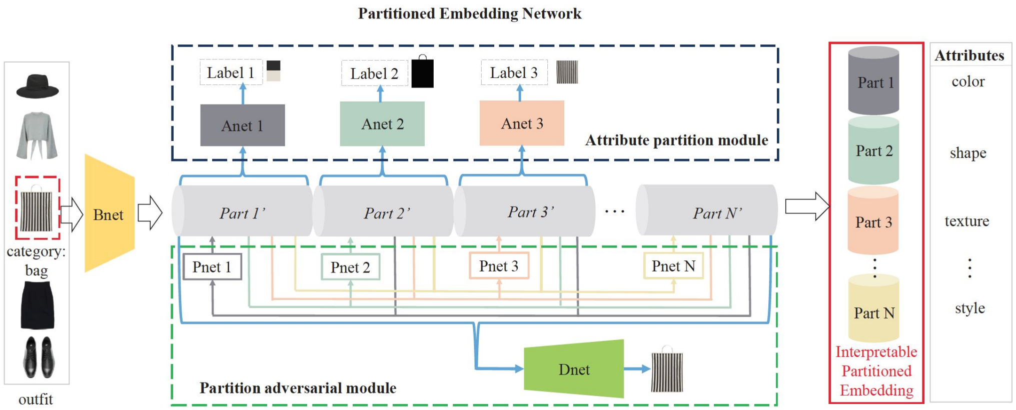
Interpretable Partitioned Embedding for Customized Multi-item Fashion Outfit Composition



- **Attribute partition module:** multiple attribute labels are adopted to ensure that different parts of the overall embedding correspond to different attributes.
- **Partition adversarial module:** Adversarial operations are adopted to achieve the independence of different parts.
- With the interpretable and partitioned embedding, construct an outfit composition graph and an attribute matching map.

[Feng 18] Interpretable Partitioned Embedding for Customized Multi-item Fashion Outfit Composition

Interpretable Partitioned Embedding for Customized Multi-item Fashion Outfit Composition

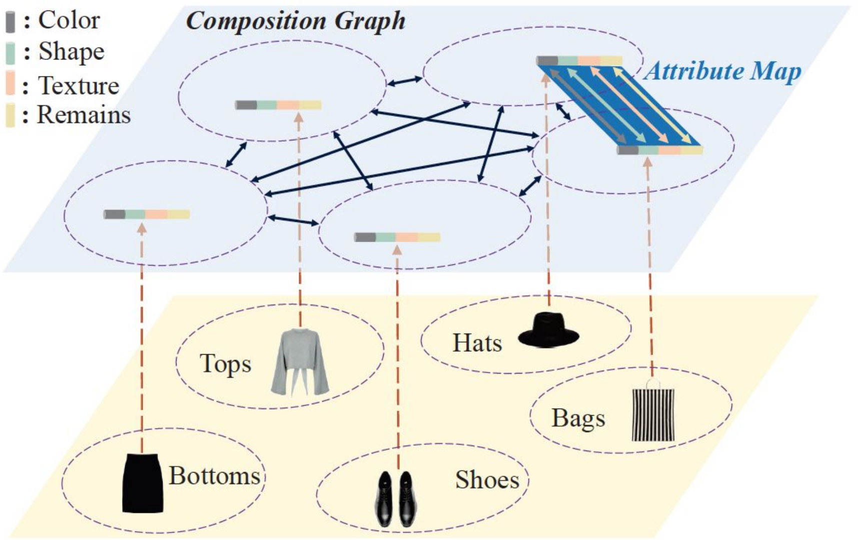


- The **attribute partition module** is composed of attribute networks Anet 1; Anet 2; Anet 3, which are used to ensure that Part 1', Part 2' and Part 3' of the overall embedding correspond to color, shape and texture attribute, respectively.
- The **partition adversarial module** includes adversarial prediction networks Pnet 1; Pnet 2; :::; Pnet N and decoder network Dnet.
- The adversarial prediction networks can ensure that different parts of the whole embedding are independent.
- The decoder network guarantees that embedding contains all useful information of original clothing item.

Flipkart

[Feng 18] Interpretable Partitioned Embedding for Customized Multi-item Fashion Outfit Composition

Interpretable Partitioned Embedding for Customized Multi-item Fashion Outfit Composition



Fashion outfit composition graph:

- In the graph, all items are classified into five classes according to the category.
- For each category, items are clustered into different cluster centers according to attributes' importance.
- With these cluster centers as vertexes, edges and weights of fashion composition graph are learned from outfit dataset. Meanwhile, an attribute map is built, which models significance of different attributes.

[Feng 18] Interpretable Partitioned Embedding for Customized Multi-item Fashion Outfit Composition

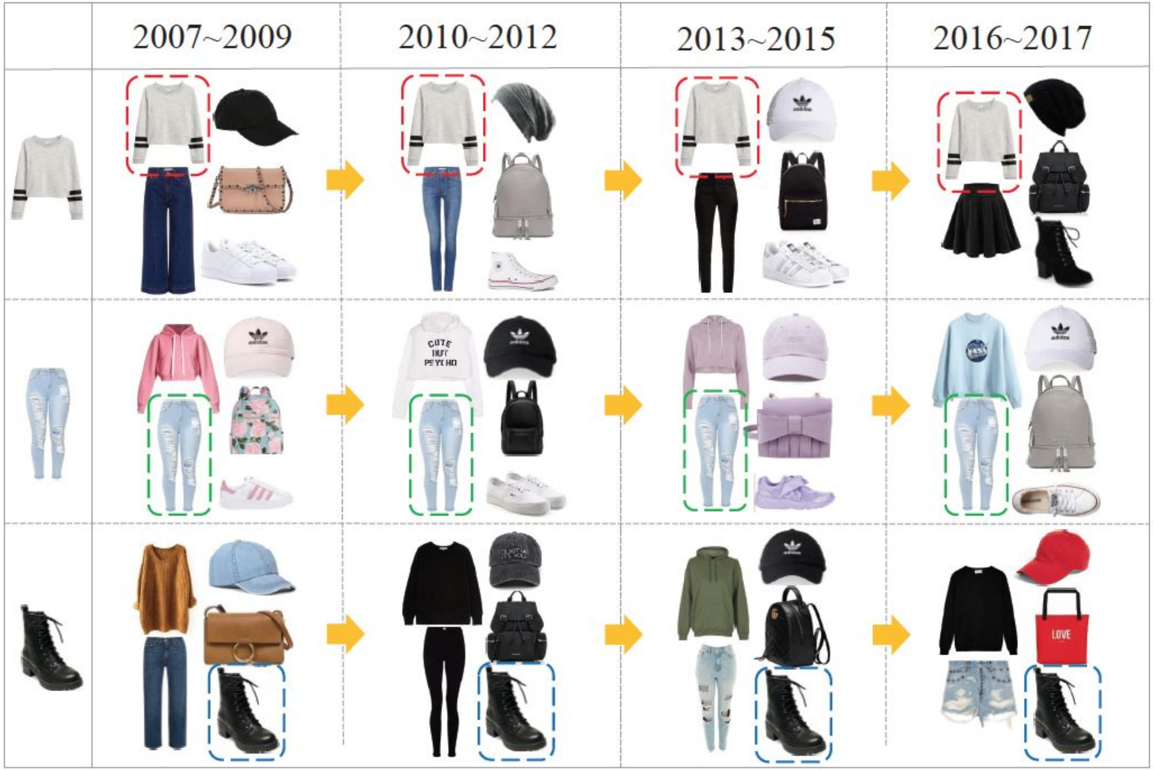
Interpretable Partitioned Embedding for Customized Multi-item Fashion Outfit Composition

Composition visual comparison among different methods

[Feng 18] Interpretable Partitioned Embedding for Customized Multi-item Fashion Outfit Composition

Flipkart

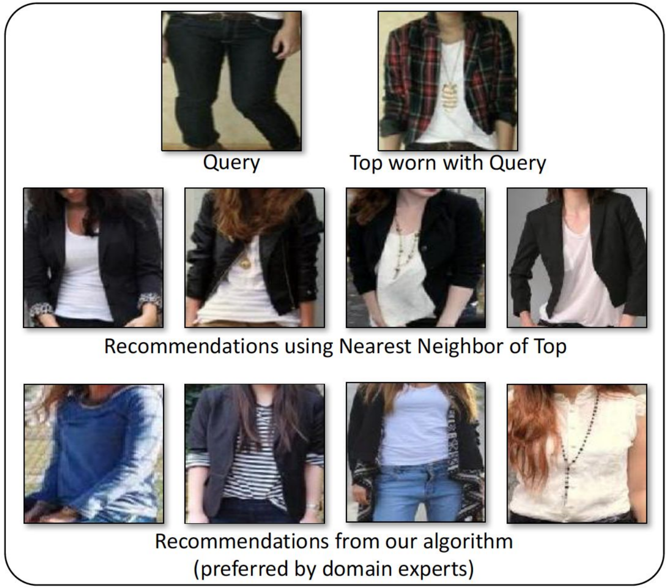
Interpretable Partitioned Embedding for Customized Multi-item Fashion Outfit Composition



Fashion trends with different years

[Feng 18] Interpretable Partitioned Embedding for Customized Multi-item Fashion Outfit Composition

CRAFT: Complementary Recommendations Using Adversarial Feature Transformer



Traditional approaches for complementary product recommendations rely on behavioral and non-visual data such as customer co-views or co-buys

Proposed approach harnesses visual cues in an unsupervised manner to learn the distribution of co-occurring complementary items in real world images

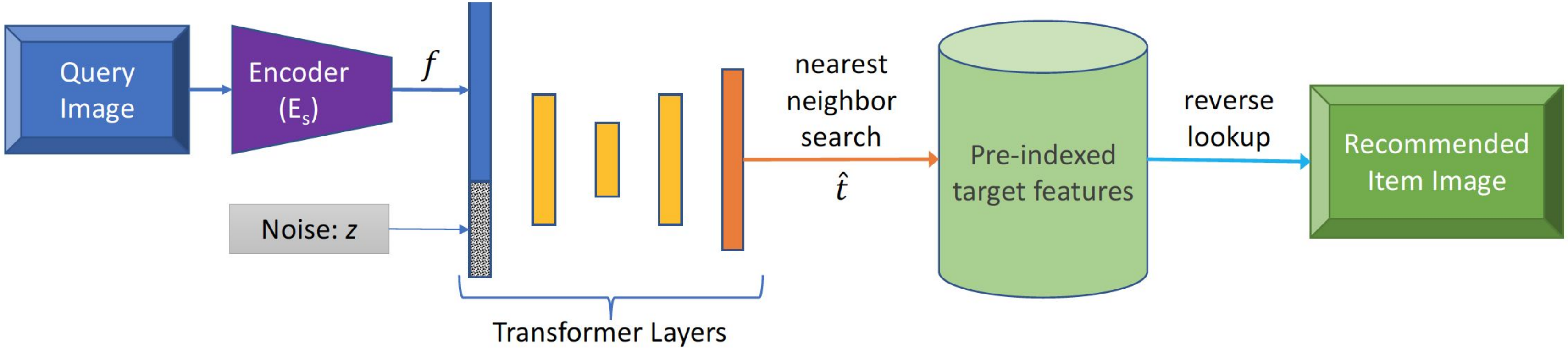
Recommending tops for a given query bottom.

Tops that are visually similar to the actual top worn with the query item are acceptable options, but lack diversity.

Proposed approach generates both complementary and diverse recommendations that are also preferred by the fashion specialists.

[Huynh 18] CRAFT: Complementary Recommendations Using Adversarial Feature Transformer

CRAFT: Complementary Recommendations Using Adversarial Feature Transformer



Generating recommendations using generative transformer network

[Huynh 18] CRAFT: Complementary Recommendations Using Adversarial Feature Transformer

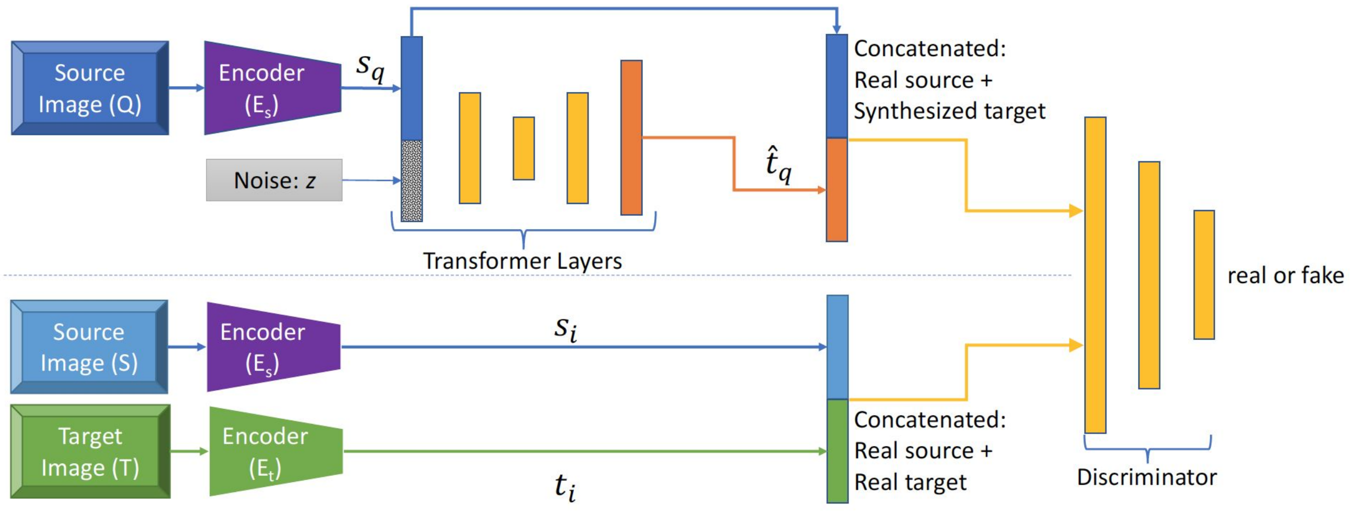
CRAFT: Complementary Recommendations Using Adversarial Feature Transformer

GAN Input	Generative (w/ random seed)	Output	Example
N/A	Yes	Image	Image Generation [7]
Image	No	Image	Image-to-Image Translation [34]
Image + Attribute	No	Image	Image Manipulation [16]
Synthetic Image	No	Image	Adding Realism [26]
Synthetic Image	Yes	Image	Adding Realism [2]
Image	No	Features	Domain Adaptation [28]
Features	Yes	Features	Ours

Similarities and differences between proposed approach and those that use adversarial loss for training

[Huynh 18] CRAFT: Complementary Recommendations Using Adversarial Feature Transformer

CRAFT: Complementary Recommendations Using Adversarial Feature Transformer

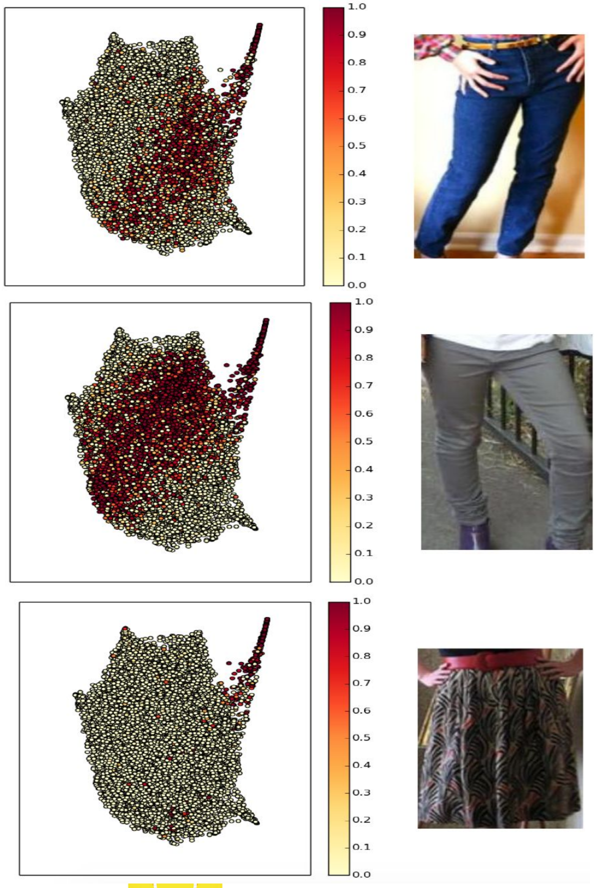


Architecture for generative complementary transformer network.

- The transformer is trained using the adversarial loss to generate the target features conditioned on the source features and a sampled noise vector

[Huynh 18] CRAFT: Complementary Recommendations Using Adversarial Feature Transformer

CRAFT: Complementary Recommendations Using Adversarial Feature Transformer



- Each row shows a 2D t-SNE embedding of all the candidate tops (left) with the corresponding query image (right).
- The colors represent the discriminator score for tops conditioned on the query (red: high score, yellow: low score).
- The discriminator is able to learn that common bottoms such as blue jeans and gray pants are compatible with a wide range of tops as compared to rarer query items such as the patterned skirt shown in the last row.

[Huynh 18] CRAFT: Complementary Recommendations Using Adversarial Feature Transformer

Flipkart

CRAFT: Complementary Recommendations Using Adversarial Feature Transformer

Complementary recommendation for a common query item (dark jeans)

[Huynh 18] CRAFT: Complementary Recommendations Using Adversarial Feature Transformer

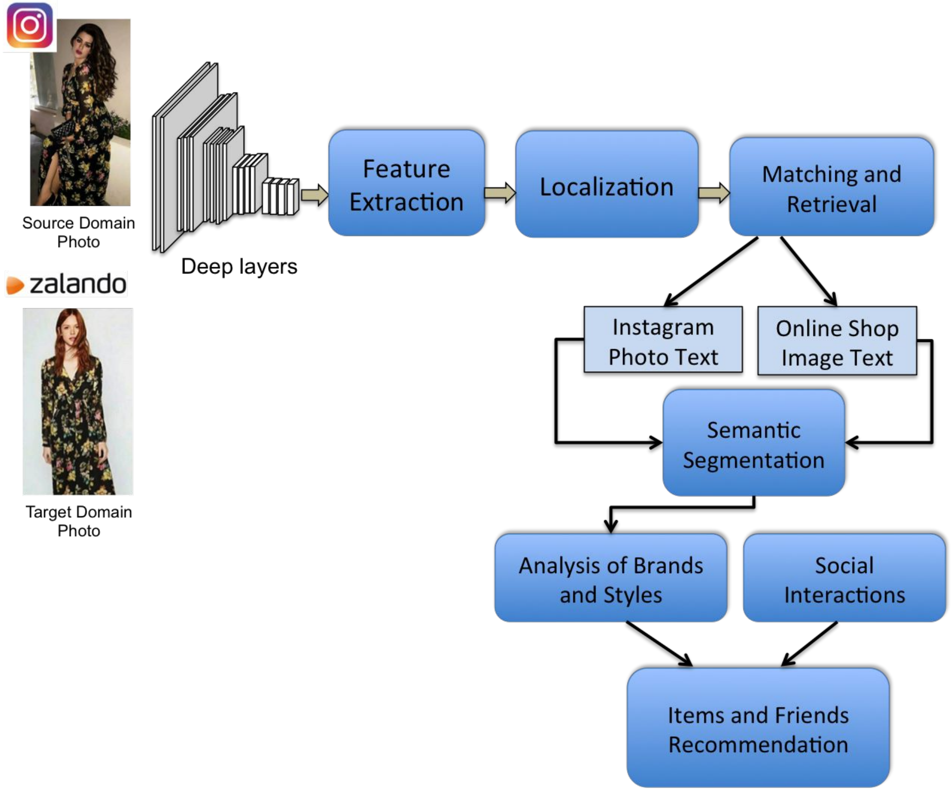
Flipkart

CRAFT: Complementary Recommendations Using Adversarial Feature Transformer

Complementary recommendation for a less common query item (pink skirt)

[Huynh 18] CRAFT: Complementary Recommendations Using Adversarial Feature Transformer

Cross Domain Recommendation



- Cross-Domain Adaptation Strategy
- Deep Pixel-Wise Segmentation
- Text Analysis for Enhanced Style Detection
- Fashion Recommendation Engine

Flipkart

[S. Jaradat 17] Deep Cross-Domain Fashion Recommendation

What Next??

The Party has just Begun

Herzlich willkommen!!

References

References

[1] M. H. Kiapour , K. Yamaguchi, A. C. Berg , and T. L. Berg. Hipster Wars: Discovering Elements of Fashion Styles, ECCV 2014.
[2] V. Jagadeesh, R. Piramuthu, A. Bhardwaj, W. Di, N. Sundaresan. Large Scale Visual Recommendations From Street Fashion Images, KDD 2014.
[3] D. Hu, R. Hall, J. Attenberg. Style in the Long Tail: Discovering Unique Interests with Latent Variable Models in Large Scale Social E-commerce. KDD 2014.
[4] R. He, J. McAuley. Ups and Downs: Modeling the Visual Evolution of Fashion Trends with One-Class Collaborative Filtering, WWW 2016.
[5] Q. Liu, S. Wu, L. Wang. DeepStyle: Learning User Preferences for Visual. SIGIR 2017.
[6] L. Dhakad, M. Das, C. Bhattacharyya, S. Datta et. al. SOPER: Discovering the Influence of Fashion and the Many Faces of User from Session Logs using Stick Breaking Process. CIKM 2017.
[7] J. McAuley, C. Targett. Image-based Recommendations on Styles and Substitutes. SIGIR 2015.
[8] Y. Hu, X. Yi, L. S. Davis. Collaborative Fashion Recommendation: A Functional Tensor Factorization Approach. ACM Multimedia 2015.
[9] W. Kang, C. Fang, Z. Wang, J. McAuley. Visually-Aware Fashion Recommendation and Design with Generative Image Models. ICDM 2017.
[10] W. Yu, H. Zhang, X. He et. al. Aesthetic-based Clothing Recommendation. WWW 2018.

References

[11] One for the Road: Recommending Male Street Attire. PAKDD 2018.
[12] R. He, C. Packer, J. McAuley. Learning Compatibility Across Categories for Heterogeneous Item Recommendation. ICDM 2016.
[13] X. Song, F. Feng, X. Han, X. Yang, W. Liu, L. Nie. Neural Compatibility Modeling with Attentive Knowledge Distillation. SIGIR 2018.
[14] X. Han, Z. Wu, Y. G. Jiang, L. S. Davis. Learning Fashion Compatibility with Bidirectional LSTMs. ACM Multimedia 2017.
[15] Z. Wang, Z. Jiang, Z. Ren, J. Tang, D. Yin. A Path-constrained Framework for Discriminating Substitutable and Complementary Products in E-commerce. WSDM 2018.
[16] C. Lynch, K. Aryafar, J. Attenberg. Images Don't Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank. KDD 2016.
[17] S. Jaradat. Deep Cross-Domain Fashion Recommendation. RecSys 2017.
[18] Z. Feng, Z. Yu, Y. Yang et. al. Interpretable Partitioned Embedding for Customized Multi-item Fashion Outfit Composition. ICMR 2018.
[19] V. Sembium, R. Rastogi, L. Tekumalla, A. Saroop. Bayesian Models for Product Size Recommendations. WWW 2018.
[20] C. P. Huynh, A. Ciptadi, A. Tyagi, A. Agarwal. CRAFT: Complementary Recommendations Using Adversarial Feature Transformer.

Thank You.