Learning about fashion from web photos

Kristen Grauman
Facebook AI Research
University of Texas at Austin
Fashion: what people wear
Fashion and culture

- Postwar simplicity
- Manufacturing fabrics
- WWII austerity
- Feminism
- Anti-war
- Women in the workplace

1920
1930
1940
1950
1970
1980

Kristen Grauman, FAIR & UT Austin
Vision + fashion: influence

Design & creativity

Business

How we shop

Kristen Grauman, FAIR & UT Austin
Vision + fashion: problems

Fashion introduces new challenges for high-level vision:

- Subtle distinctions
- Composition and compatibility
- Personalization and taste

Requires computational models for style

Kristen Grauman, FAIR & UT Austin
This talk: fashion & web photos

• Assembling fashionable outfits
  – What goes with what?
  – How to compose a wardrobe?
  – How could this outfit look better?

• Learning subtle attributes
  – How to distinguish slight differences?
How to learn visual compatibility?

Co-purchase data [McAuley 2015, Veit 2015, He 2016]

Manual curation [Li 2017, Song 2017, Han 2017]

Our goal: Unlabeled in the wild photos

Kristen Grauman, FAIR & UT Austin
Attributes as style elements

**Style:** underlying compositions of elements.

Kristen Grauman, FAIR & UT Austin
Attributes as style elements

- Material, cut, pattern
  - Fine-tune classification on ResNet50
- Color, clothing article:
  - Segmentation on DeepLab-DenseCRF

Kristen Grauman, FAIR & UT Austin
Learning styles from web photos

Unsupervised learning of a “visual style topics” with a polylingual topic model

An outfit is a mixture of (latent) styles. A style is a distribution over attributes.

Example discovered styles (dresses)

Styles we automatically discover in the **Amazon** dataset [McAuley et al. 2015]
Example discovered styles (full outfit)

Styles we automatically discover in the HipsterWars dataset [Kiapour et al]
Visual compatibility

Calculate compatibility of garments via likelihood under topic model

\[
c(o_j) := p(o_j | \mu, \Sigma, \beta)\]

An **outfit** is a mixture of (latent) **styles**. A **style** is a distribution over **attributes**.
Visual compatibility results

Most compatible

Kristen Grauman, FAIR & UT Austin
Visual compatibility results

Least compatible

Kristen Grauman, FAIR & UT Austin
BiLSTM [Han et al. 17]: unsupervised sequential model trained on Polyvore sets.

Monomer [He et al. 16]: supervised embedding trained on Amazon products co-purchase info.

Encouraging results for learning compatibility from unlabeled, full-body Web images

Kristen Grauman, FAIR & UT Austin

Hsiao & Grauman, CVPR 2018
Creating a “capsule” wardrobe

**Goal:** Select minimal set of pieces that mix and match well to create many viable outfits

Kristen Grauman, FAIR & UT Austin
Creating a “capsule” wardrobe

Pose as subset selection problem
set of garments = argmax compatibility + versatility

Capsule pieces

Outfit #1  Outfit #2  Outfit #3  Outfit #4  Outfit #5

Kristen Grauman, FAIR & UT Austin

Hsiao & Grauman, CVPR 2018
Capsule via subset selection

Pose as *subset selection* problem

\[ \text{set of garments} = \arg \max \text{compatibility} + \text{versatility} \]

- Capsule pieces
  - \( A_{0T} \)
  - \( A_{1T} \)
  - \( A_{2T} \)

- Outfit #1
- Outfit #2
- Outfit #3
- Outfit #4

- Increase outfit compatibility
- Cover all styles (or user’s styles)

Kristen Grauman, FAIR & UT Austin
Personalized capsule example

Discover user’s style preferences from album

Kristen Grauman, FAIR & UT Austin

Hsiao & Grauman, CVPR 2018
Personalized capsule example

Discover user’s style preferences from album
This talk: fashion & web photos

• Assembling fashionable outfits
  – What goes with what?
  – How to compose a wardrobe?
  – How could this outfit look better?

• Learning subtle attributes
  – How to distinguish slight differences?
Idea: Minimal edits for outfit improvement

[Hsiao et al. Fashion++, arXiv 2019]
Fashion++

[Hsiao et al. Fashion++, arXiv 2019]

Kristen Grauman, FAIR & UT Austin
Fashion++

Try this top!

Kristen Grauman, FAIR & UT Austin

[Hsiao et al.  Fashion++, arXiv 2019]
Our approach: Fashion++

Image generation pipeline factorizes shape and texture

Edit via activation maximization with discriminative fashionability model

Learning discriminative fashionability model

Represent an outfit as concatenation of its latent codes
Learning discriminative fashionability model

Bootstrap web photos for “negatives” to learn fashionability
Computing an outfit edit

Kristen Grauman, FAIR & UT Austin
Computing an outfit edit

Propagate gradients to iteratively update the target feature (e.g., shape)

Kristen Grauman, FAIR & UT Austin
Swap to garment from inventory.

Visualization of changed garment in new presentation.

Fashion++ minimal edits

[Hsiao et al. Fashion++, arXiv 2019]
Fashion++ minimal edits

- Change pants fit
- Change pants length
- Tuck shirt in
- Blouse shirt out

Kristen Grauman, FAIR & UT Austin

[Hsiao et al. Fashion++, arXiv 2019]
Fashion++ balances best by improving fashionability while not changing too much.
This talk: fashion & web photos

- Assembling fashionable outfits
  - What goes with what?
  - How to compose a wardrobe?
  - How could this outfit look better?

- Learning subtle attributes
  - How to distinguish slight differences?

The limits of web photos!

Kristen Grauman, FAIR & UT Austin
Fine-grained attribute comparisons

*Which is more sporty?* vs. *Which is more flirty?*

**Goal**: Subtle visual comparisons

**Challenge**: Curating training image pairs

Kristen Grauman, FAIR & UT Austin
Web photos and the streetlight effect

Kristen Grauman, FAIR & UT Austin
Idea: Semantic jitter

Overcome sparsity of available fine-grained image pairs with attribute-conditioned image generation

Images generated by Yan et al. 2016
Attribute2Image
CVAE approach
Idea: Semantic jitter

Overcome sparsity of available fine-grained image pairs with attribute-conditioned image generation

Status quo: Low-level jitter

Our idea: Semantic jitter

Yu & Grauman, ICCV 2017
Semantic jitter for attribute learning

Train rankers with both real and synthetic image pairs, test on real fine-grained pairs.

Ranking functions trained with deep spatial transformer ranking networks [Singh & Lee 2016] or Local RankSVM [Yu & Grauman 2014]

Yu & Grauman, ICCV 2017
Idea: Active training image creation

*Actively* generate best image pairs for human labeling

Kristen Grauman, FAIR & UT Austin

Yu & Grauman, CVPR 2019
Active training image creation

System “imagines” image pairs that would confuse current ranking model

Yu & Grauman, CVPR 2019
Active training image creation

Active Batch: Faces

Active Batch: Shoes

Real Pairs

Active Synthetic Pairs

Casual

Durable

Smiling

Actively curating synthetic training images → more accurate model

Kristen Grauman, FAIR & UT Austin
Summary

• Learning styles and fashionability from web photos

• New ideas and methods for:
  – Style and compatibility
  – Capsule wardrobe creation
  – Incremental outfit improvement
  – Subtle visual comparisons

Kimberly Hsiao
Aron Yu
Kristen Grauman
FAIR & UT Austin
Papers


BrowseWithMe

Make online shopping more accessible for visually impaired users

Status quo: screen reader

System: cherry red

https://www.youtube.com/watch?v=IuH5sWDSGuw&feature=youtu.be

Stangl et al. ASSETS 2018