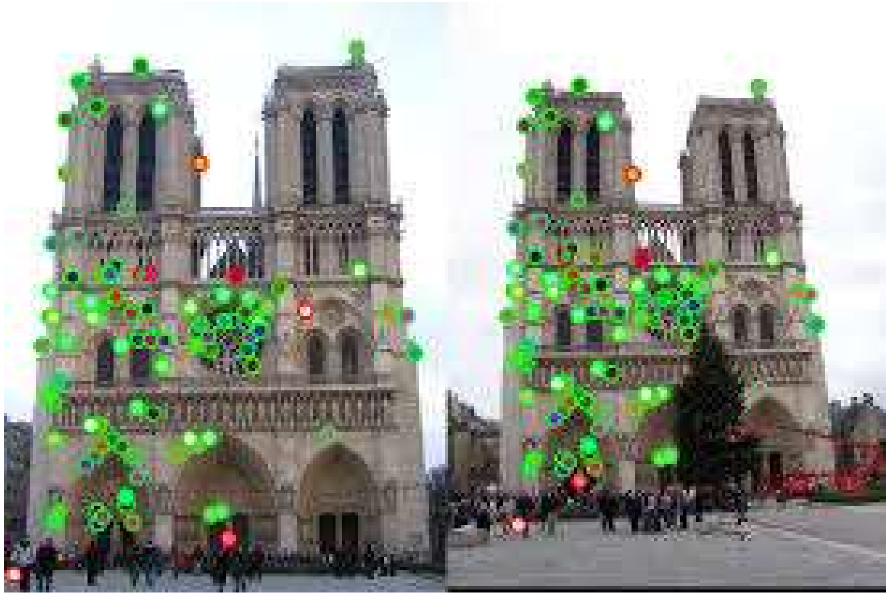
Learning visual styles

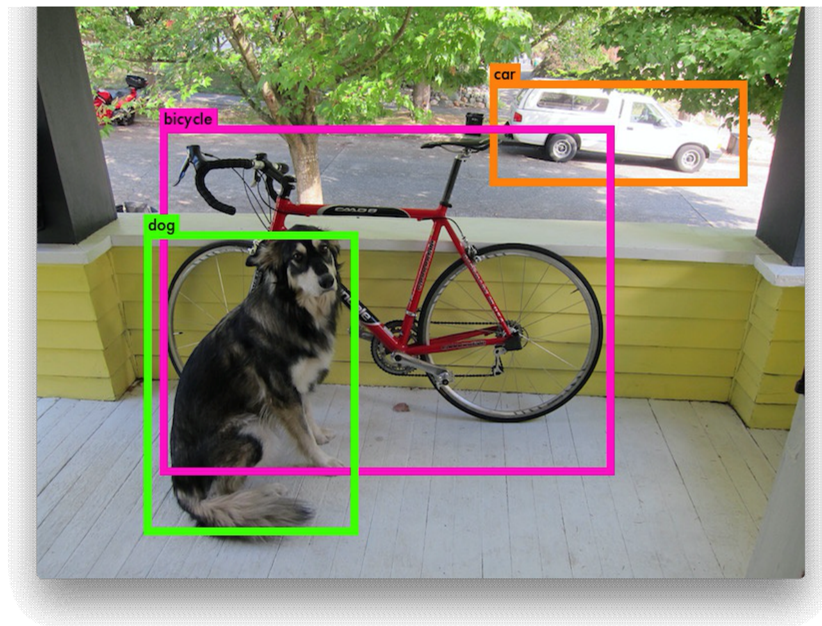
Kristen Grauman
Department of Computer Science
University of Texas at Austin

Visual recognition + fashion

Recognizing instances



Recognizing categories



Kristen Grauman, UT Austin

Visual recognition + fashion

Recognizing instances

Recognizing categories

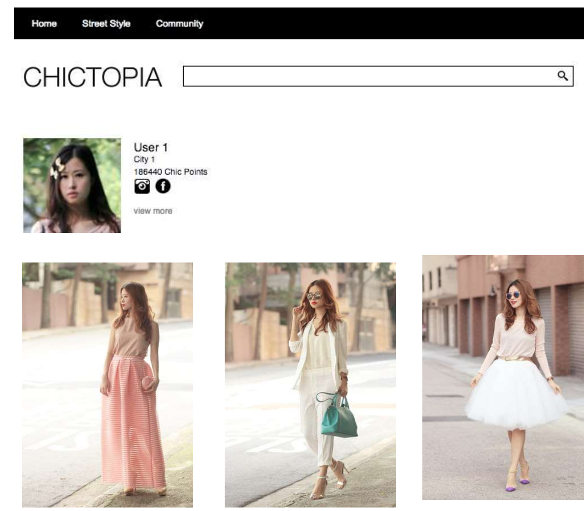
Kristen Grauman, UT Austin

Visual recognition + fashion

But fashion also introduces new challenges for high-level vision:

Subtle distinctions

Composition and compatibility



Personalization and taste

Requires computational models for *style*

Visual recognition + fashion

Many applications for learning to model style

This talk

- Subtle visual attributes
- Style discovery and forecasting
- Creating capsule wardrobes

Relative attributes

- High-level semantic properties shared by objects
- Human-understandable and machine-detectable

Smiling

???

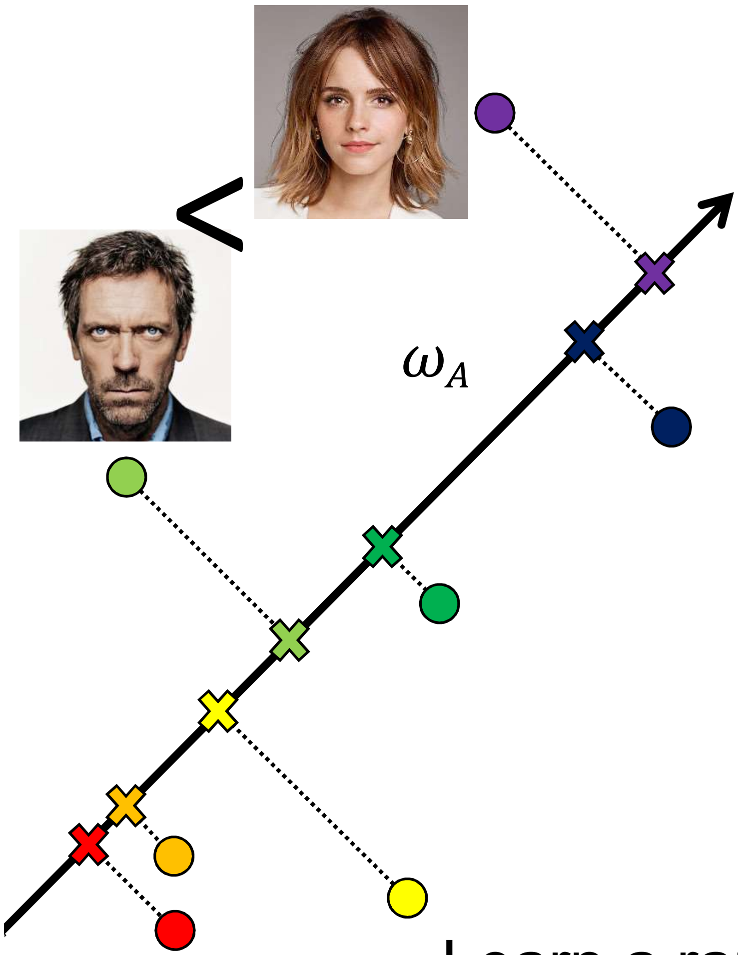
>?

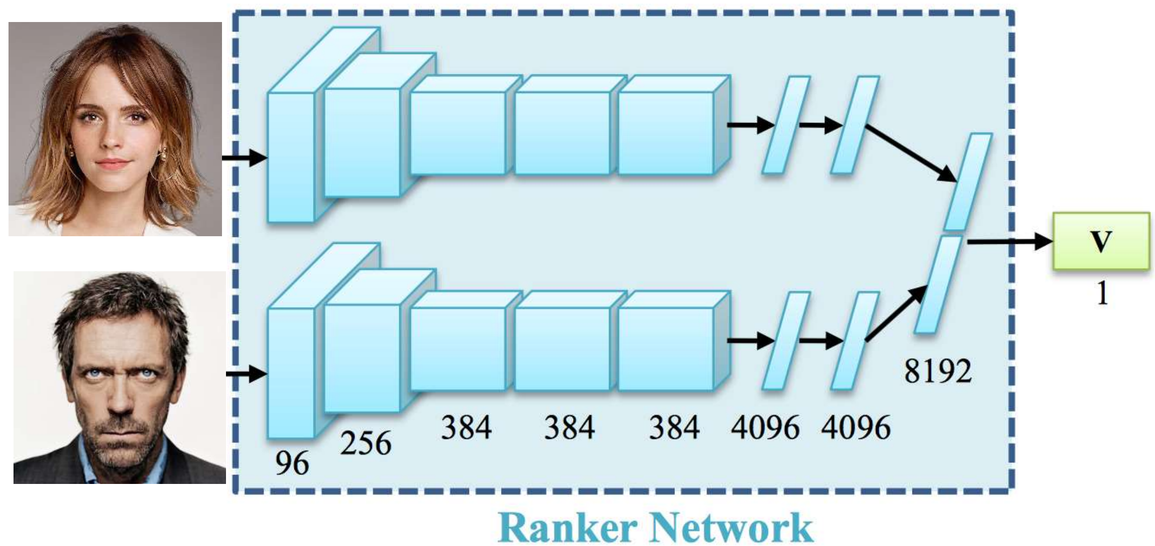
Not Smiling

[Oliva et al. 2001, Ferrari & Zisserman 2007, Kumar et al. 2008, Farhadi et al. 2009, Lampert et al. 2009, Endres et al. 2010, Wang & Mori 2010, Berg et al. 2010, Branson et al. 2010, Parikh & Grauman 2011, …]

Parikh & Grauman, ICCV 2011
Singh & Lee, ECCV 2016

Relative attributes



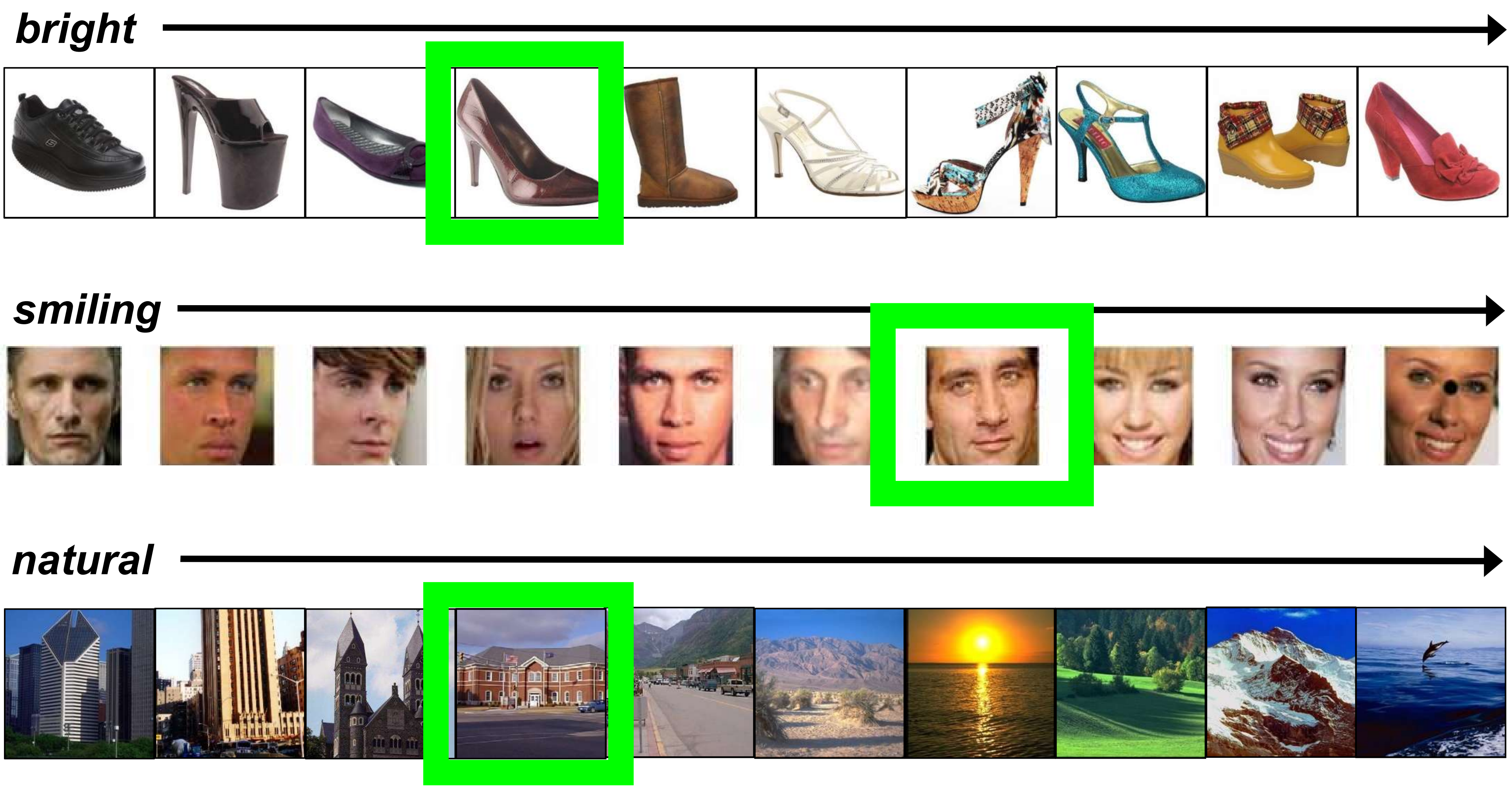


Learn a ranking function per attribute

Parikh & Grauman, ICCV 2011
Singh & Lee, ECCV 2016

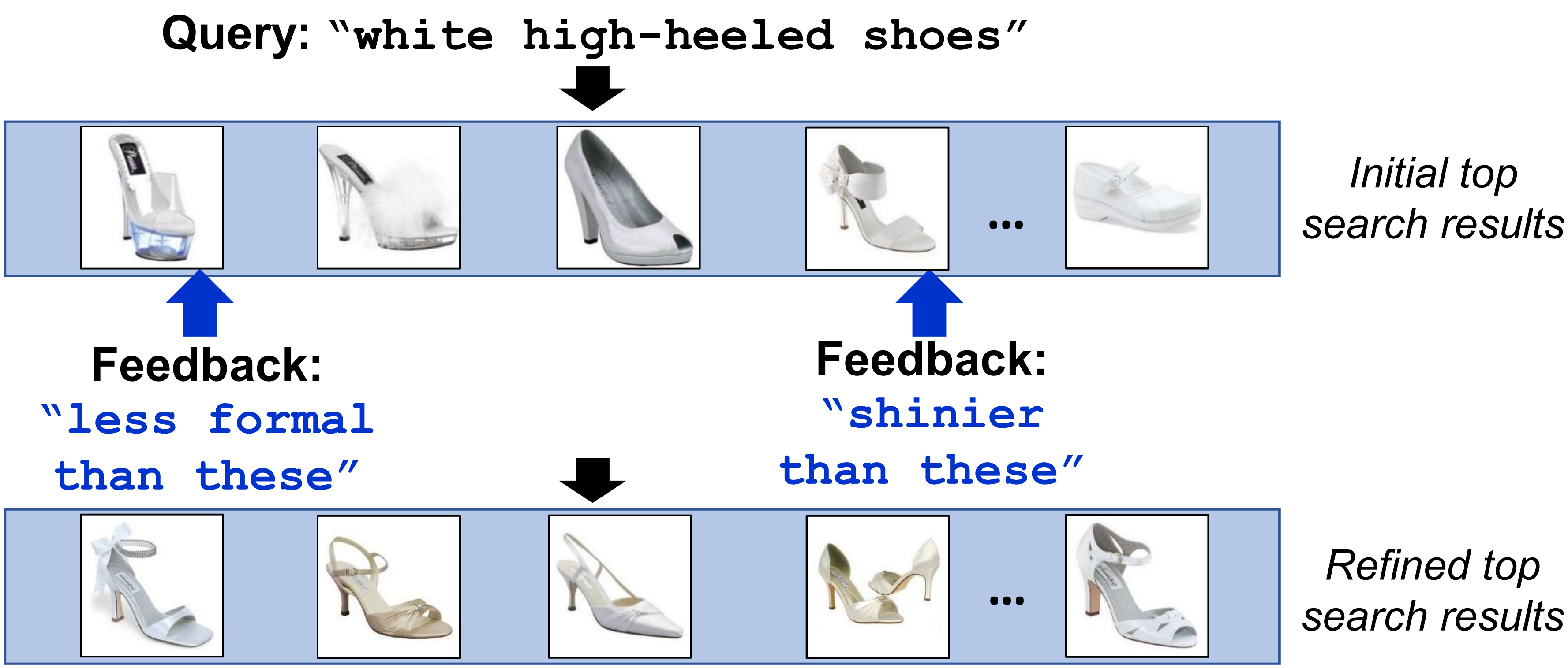
Relative attributes

Now we can compare images by attribute's "strength"



[Parikh & Grauman, ICCV 2011]

WhittleSearch: Relative attribute feedback



Whittle away irrelevant images via precise semantic feedback

[Kovashka, Parikh, and Grauman, CVPR 2012, IJCV 2015]

Challenge: fine-grained comparisons

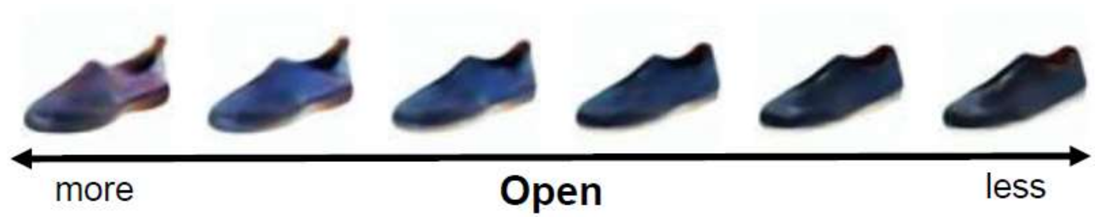
Which is more sporty?

Sparsity of supervision problem:

1. Label availability: lots of possible pairs.
2. Image availability: subtleties hard to curate.

Idea: Semantic jitter

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

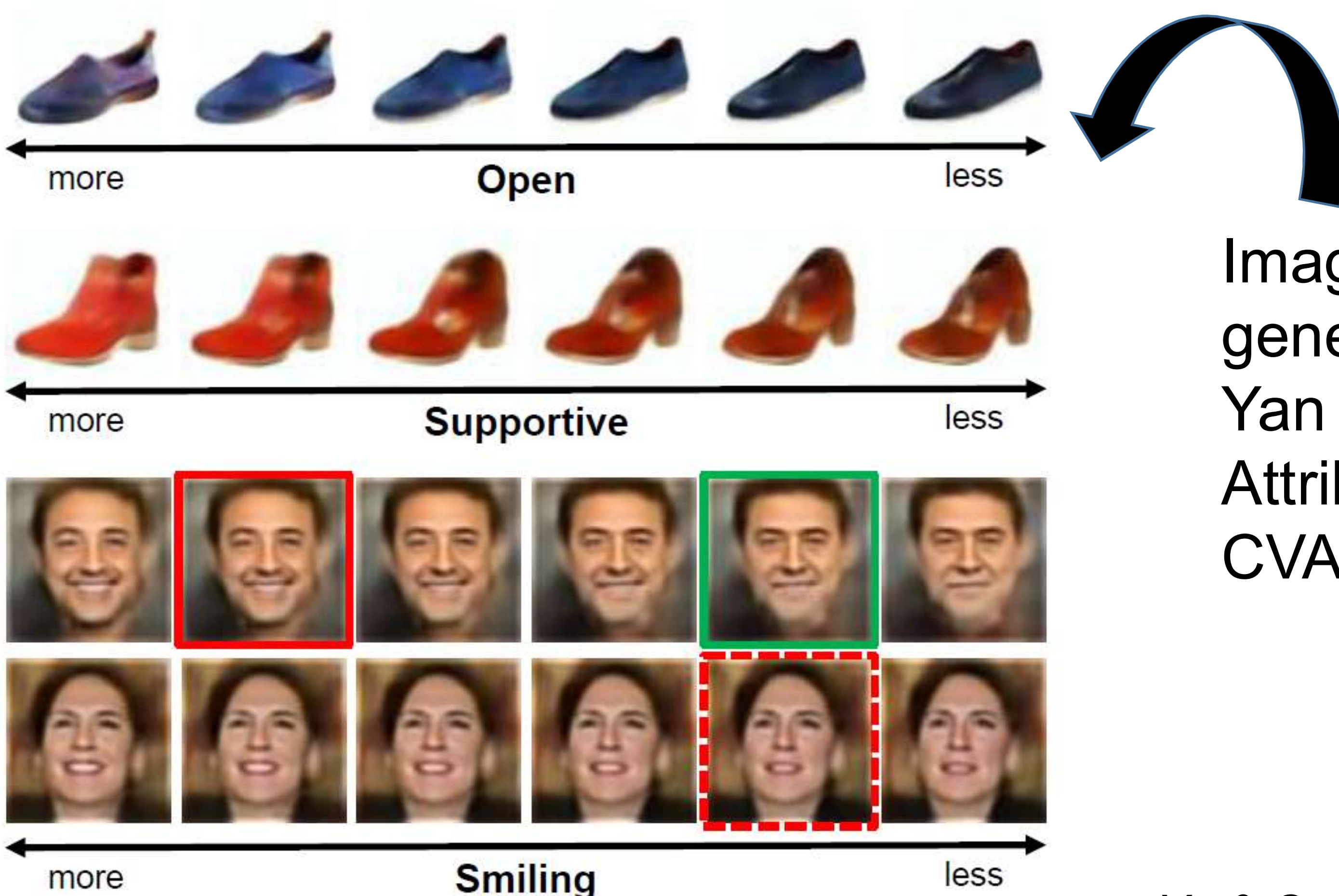


Images generated by Yan et al. 2016 Attribute2Image CVAE approach

Yu & Grauman, ICCV 2017

Idea: Semantic jitter

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



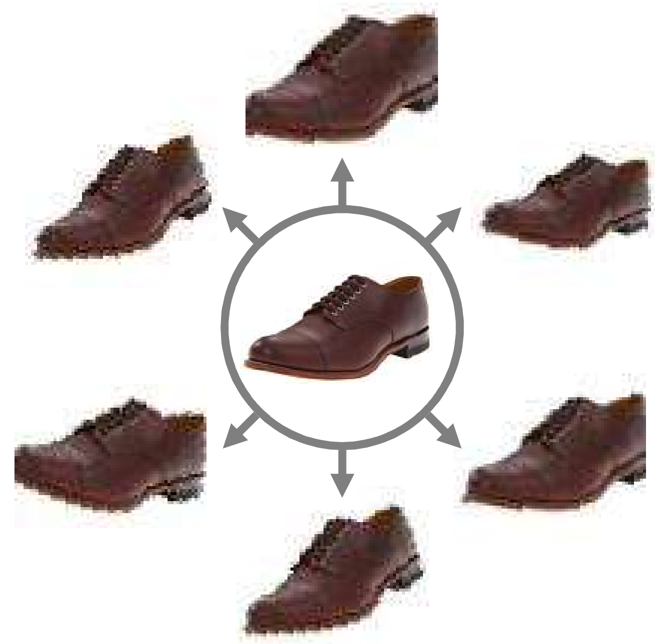
Images generated by Yan et al. 2016 Attribute2Image CVAE approach

Yu & Grauman, ICCV 2017

Idea: Semantic jitter

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

vs.



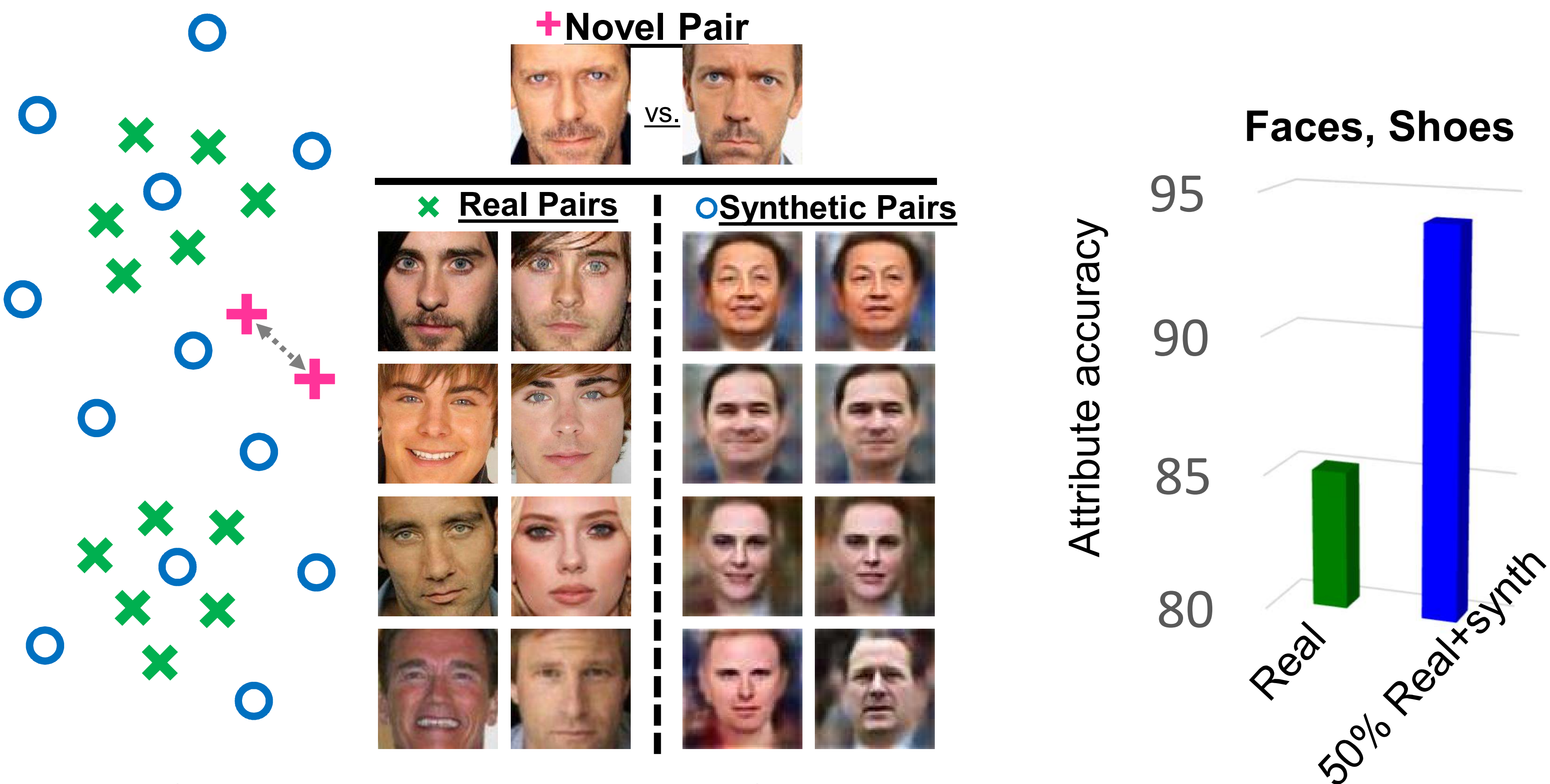
Our idea: Semantic jitter

Status quo: Low-level 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

Semantic jitter for attribute learning

		Open	Sporty	Comfort
Zap50K-1	RelAttr [Parikh 2011]	88.33	89.33	91.33
	FG-LP [Yu 2014]	90.67	91.33	93.67
	DeepSTN [Singh 2016]	93.00	93.67	94.33
	DSynth-Auto (Ours)	**95.00**	**96.33**	**95.00**
Zap50K-2	RelAttr [Parikh 2011]	60.36	65.65	62.82
	FG-LP [Yu 2014]	69.36	66.39	63.84
	DeepSTN [Singh 2016]	70.73	67.49	66.09
	DSynth-Auto (Ours)	**72.18**	**68.70**	**67.72**

- State-of-the-art fine-grained comparisons
- All models trained on 64x64 images

UT Zappos-50K dataset

Yu & Grauman, ICCV 2017

Challenge: Which attributes matter?

Left shoe is _____ than right shoe:

Less colorful

Less comfortable

More rugged

More shiny

Less feminine

More stylish

More formal

Idea: Prominent relative attributes

Infer which comparisons are perceptually salient

Left shoe is _____ than right shoe:

Less colorful
Less comfortable
More rugged
More shiny
Less feminine
More stylish
More formal

Prominent Differences

More formal
More shiny
Less comfortable
Less feminine
Less colorful
More rugged
More stylish

Chen & Grauman, CVPR 2018

Approach: What causes prominence?

Prominent Difference:

- Large difference in attribute strength:


Colorful

- Unusual and uncommon attribute occurrences:


Visible Forehead

- Absence of other noticeable differences:

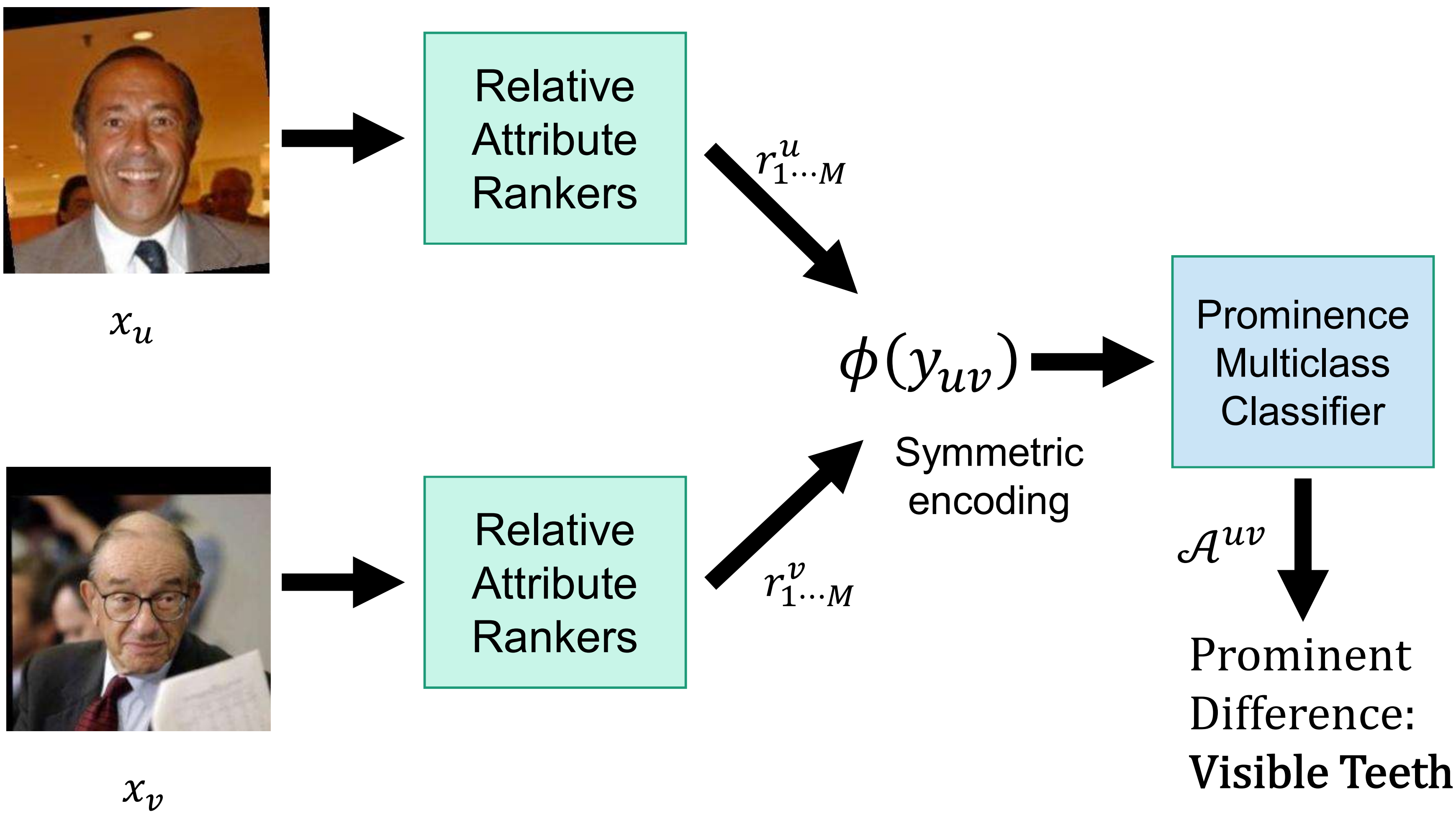

Dark Hair

In general: Interactions between all the relative attributes in an image pair cause prominent differences.

Chen & Grauman, CVPR 2018

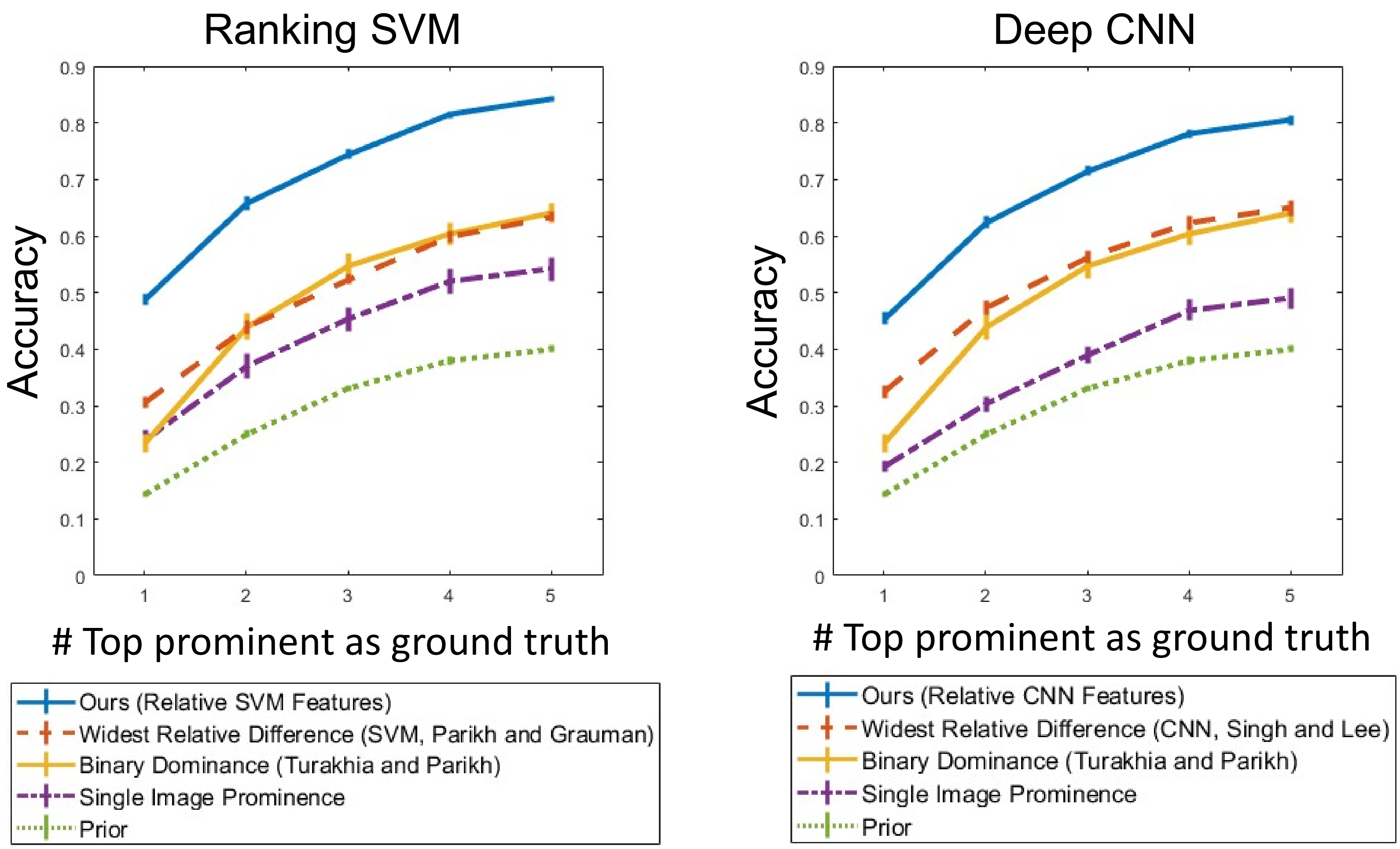
Approach: Predicting prominent differences

input: $y_{uv} = (x_u, x_v)$



Chen & Grauman, CVPR 2018

Results: Prominent differences



Chen & Grauman, CVPR 2018

Results: Prominent differences

(a) **colorful** (>),
sporty, comfortable

(b) **sporty** (>),
colorful, comfortable

(c) **tall** (<),
colorful, sporty

(d) **shiny** (>),
feminine, colorful

(e) **rugged** (<),
tall, feminine

(f) **feminine** (>),
comfortable, shiny

(j) **masculine** (>),
smiling, visible teeth

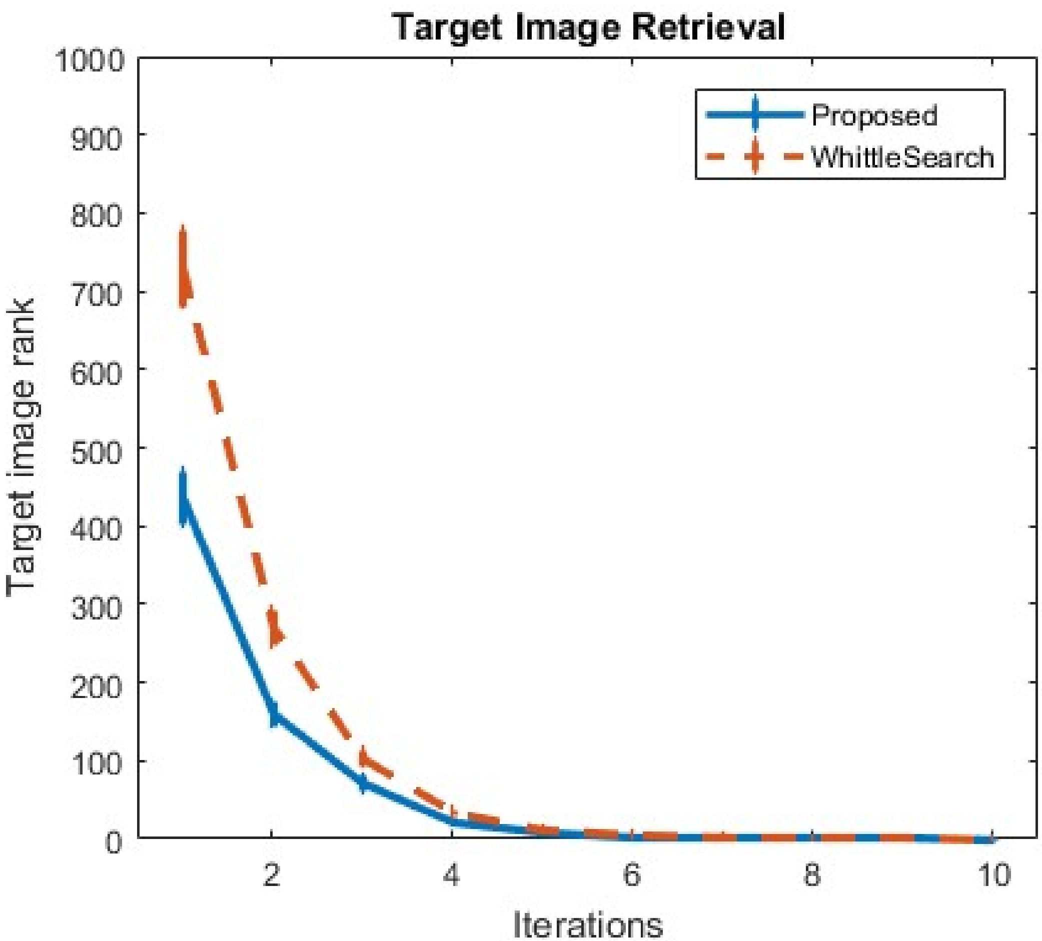
(k) **bald head** (<),
dark hair, visible teeth

(l) **dark hair** (<),
mouth open, smiling

(Top 3 prominent differences for each pair)

Chen & Grauman, CVPR 2018

Prominent differences: impact on visual search



Faster retrieval of user’s target image without using any additional user feedback.

Initial top search results

Refined top search results

Leverage prominence to better focus search results

Chen & Grauman, CVPR 2018

This talk

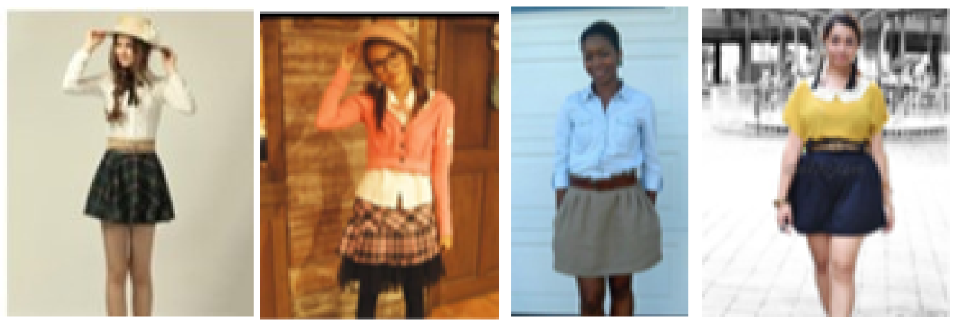
- Subtle visual attributes
- Style discovery and forecasting
- Creating capsule wardrobes

From items to styles

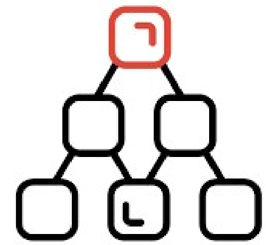
From items to styles

Requires a representation of *visual style*

CNN image similarity



stylistic similarity?

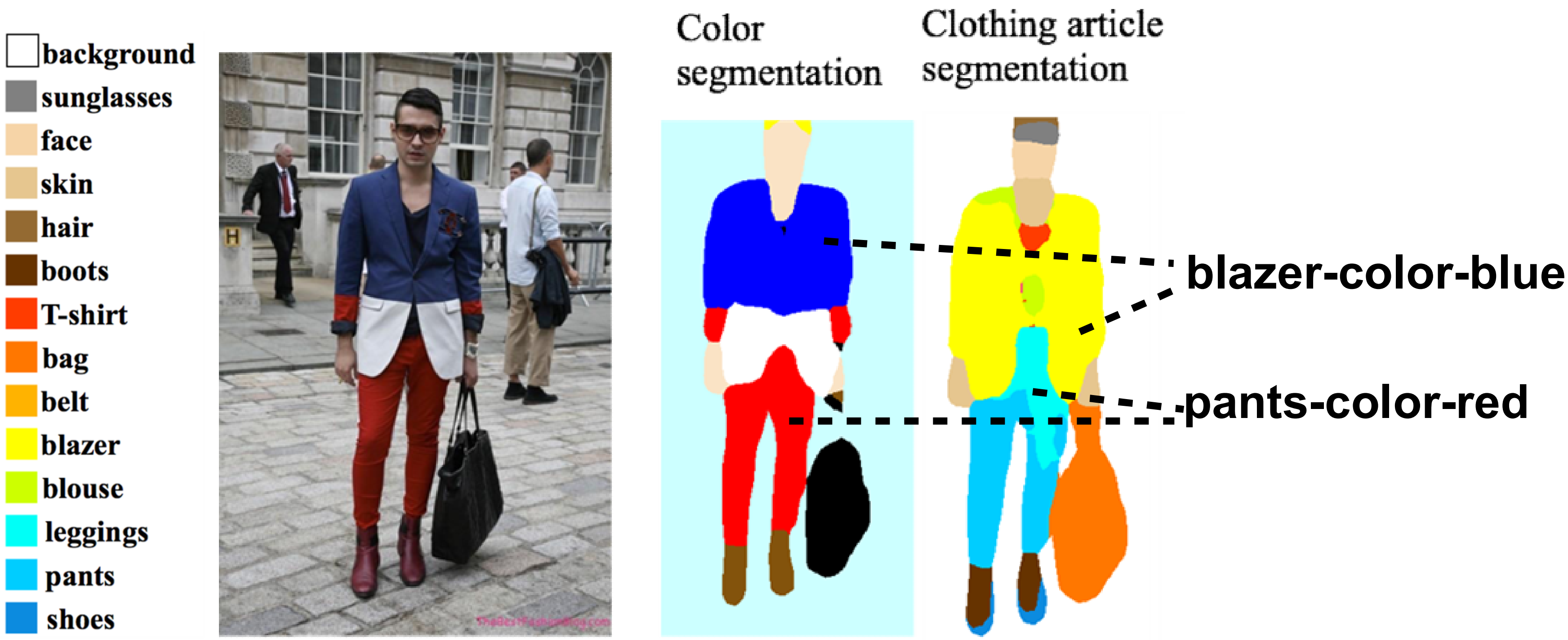


manually defined style labels

Challenges:

- Same “look” manifests in different garments
- Emerges organically and evolves over time
- Soft boundaries

Detect localized attributes



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

Topic models: Inspiration from text

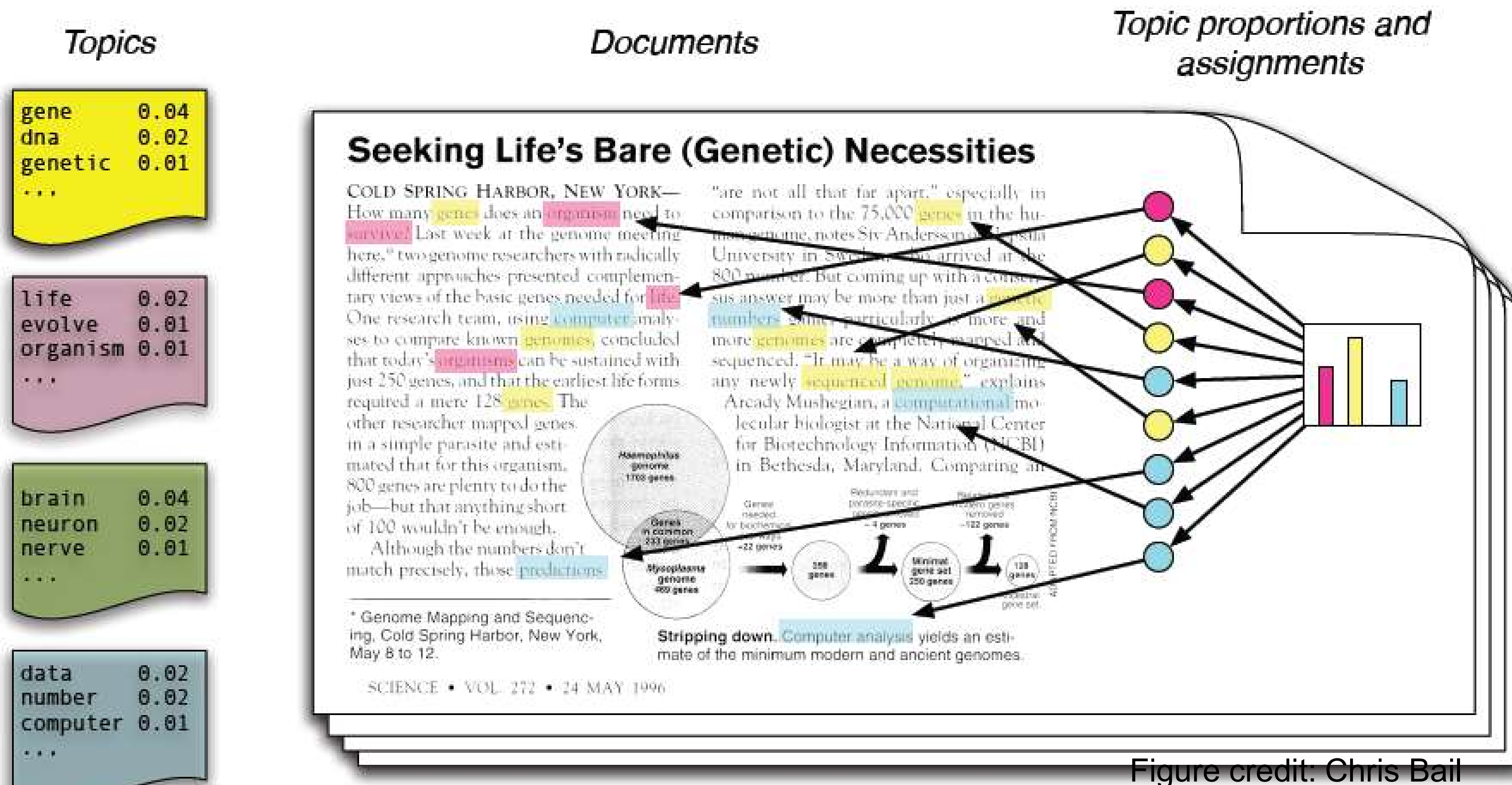
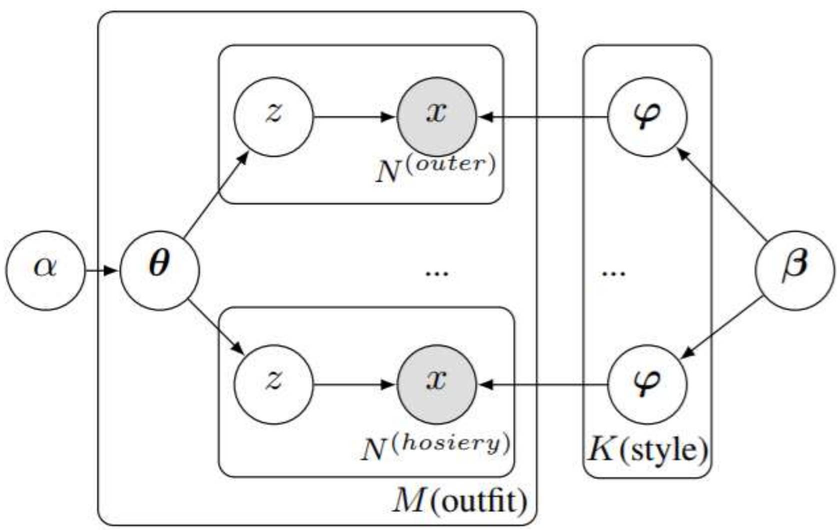


Figure credit: Chris Bail

Topic models, e.g., Latent Dirichlet Allocation (LDA)

Idea: Discovering visual styles

Unsupervised learning of a style-coherent embedding with a **polylingual topic model**



outer layer
outer_color_orange
outer_color_white
outer_pattern_printed
outer_decoration_button
outer_sleeve_long
outer_length_short
outer_front_open
⋮
upper
shirt_color_white
shirt_pattern_plain
shirt_sleeve_short

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

Mimno et al. "Polylingual topic models." EMNLP 2009.

Hsiao & Grauman, ICCV 2017

Example discovered styles (dresses)

Styles we automatically discover in the **Amazon** dataset [McAuley et al. 2015]

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]

Style discovery accuracy

How well do our discovered styles align
with human-perceived styles?

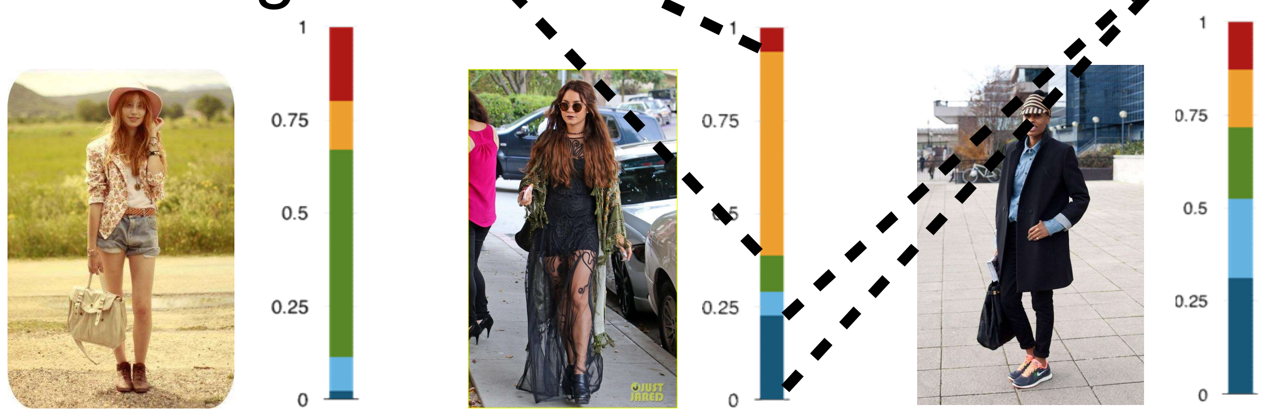
	HipsterWars		DeepFashion	
	Avg. max AP	NMI	Avg. max AP	NMI
StyleNet [33]	0.39	0.20	0.0501	0.0011
ResNet [12]	0.30	0.16	0.0524	0.0004
Attributes	0.28 / 0.32	0.19 / 0.28	0.0560 / 0.1294	0.0017 / 0.0082
PolyLDA	0.50 / **0.53**	0.21 / **0.31**	0.0407 / **0.1762**	0.0006 / **0.0227**

Attributes and PolyLDA show result if using either predicted
attributes (first) or ground truth attributes (second).

Style-coherent embedding

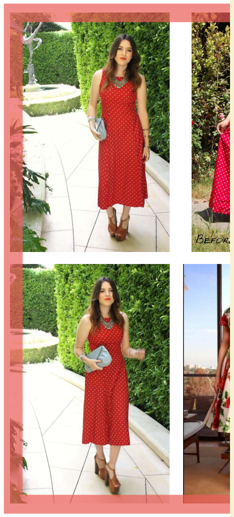
Discovered latent styles (topics)

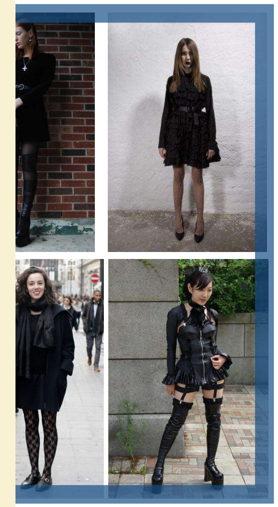
Image embedding



Style-coherent embedding

Discovered latent styles (topics)

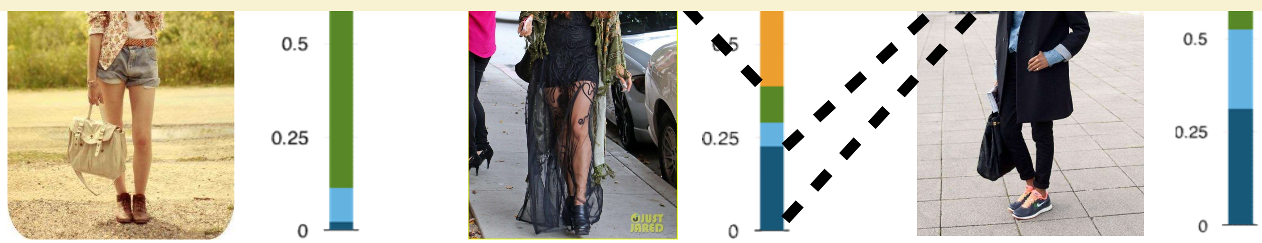




Leverage this embedding for

1) Style browsing
2) Style mixing
3) Style summarization
4) Style forecasting

Image



Style browsing results

query

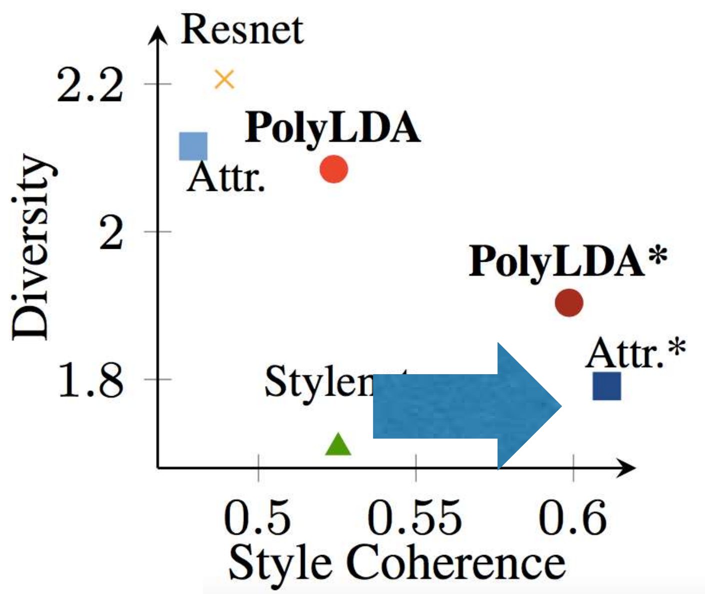
Similar in CNN space

vs.

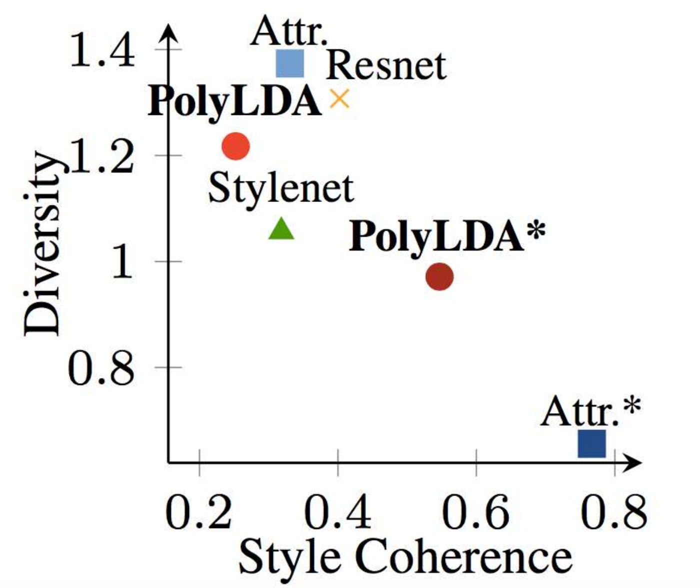
Similar in style space (ours)

Maintain **style coherence** while also permitting diversity

Style browsing results



HipsterWars dataset
[Kiapour ECCV 2014]



DeepFashion dataset
[Liu CVPR 2016]

Maintain **style coherence** while also permitting diversity

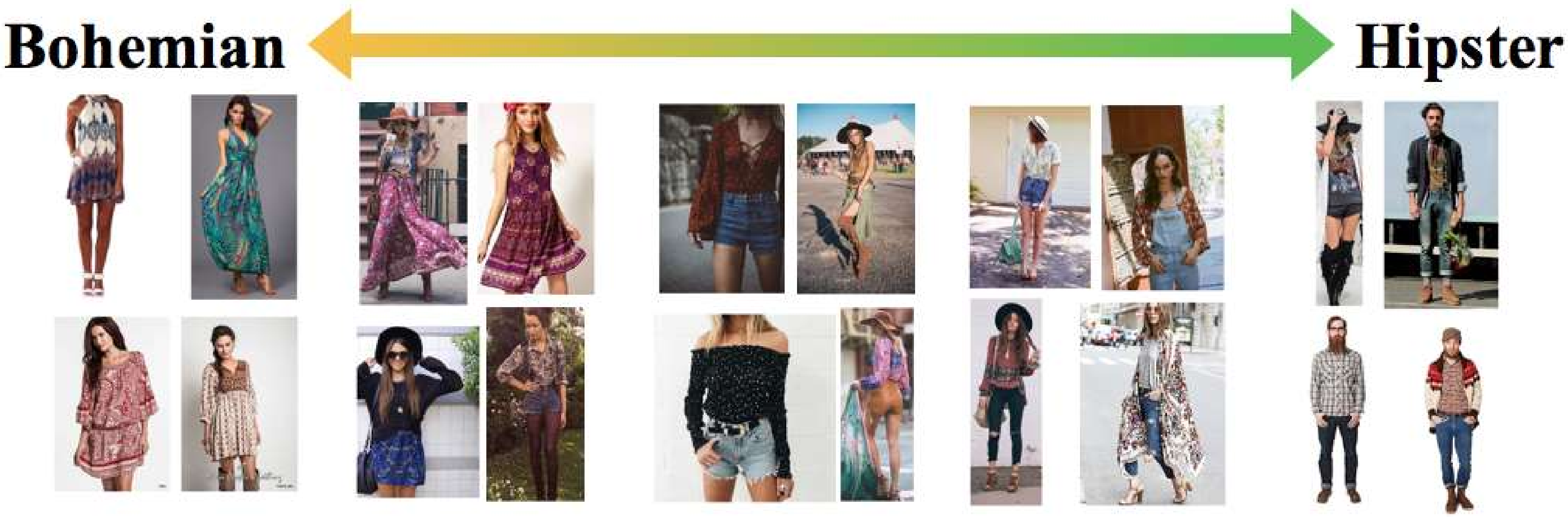
Mixing styles

Our embedding naturally facilitates browsing for mixes of user-selected styles

Hsiao & Grauman, ICCV 2017

Mixing styles

Our embedding naturally facilitates browsing for mixes of user-selected styles

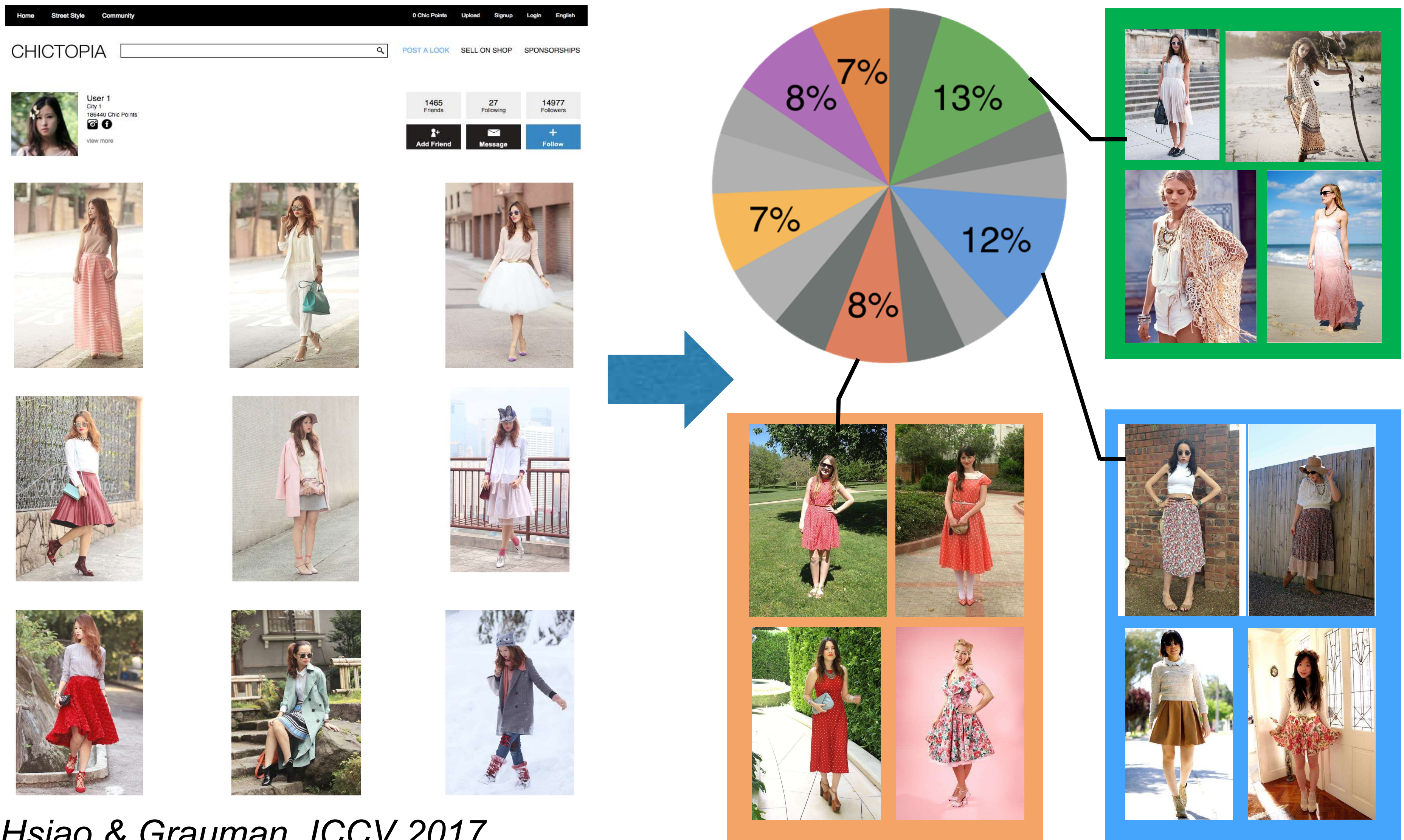


Hsiao & Grauman, ICCV 2017

Style summarization

Given a gallery of photos

Summarize by dominant styles

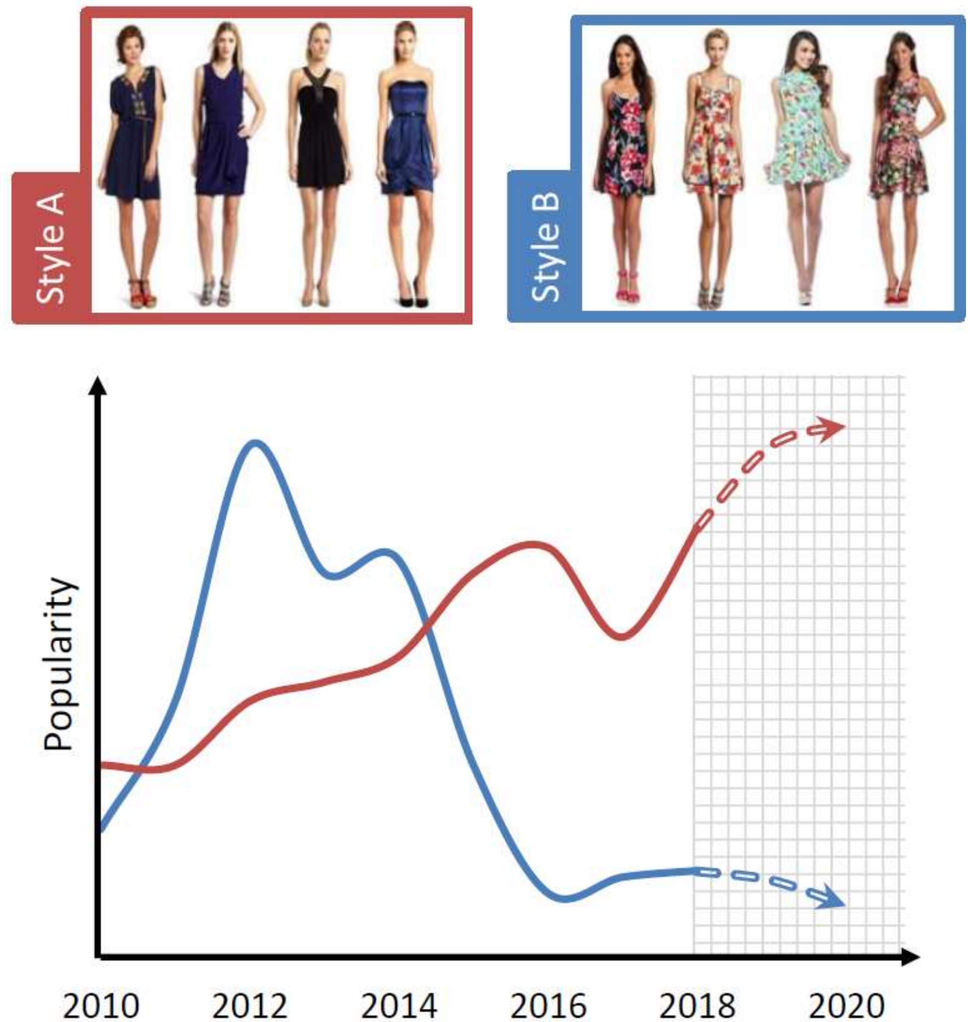


Hsiao & Grauman, ICCV 2017

Style forecasting

Can we predict the future popularity of styles?

1. Visual style discovery
2. Construct style temporal trajectory
3. **Forecast future trend**
4. Style description via signature attributes



Al-Halah et al., ICCV 2017

Amazon dataset

[McAuley et al. SIGIR 2015]

- Dresses, Tops & Tees and Shirts -- over 6 years
- 80,000 items and 210,000 transactions

Text
Women's Stripe Scoop Tunic Tank, Coral, Large

Tags
- Women
- Clothing
- Tops & Tees
- Tanks & Camis

Text
The Big Bang Theory DC Comics Slim-Fit T-Shirt

Tags
- Men
- Clothing
- T-Shirts

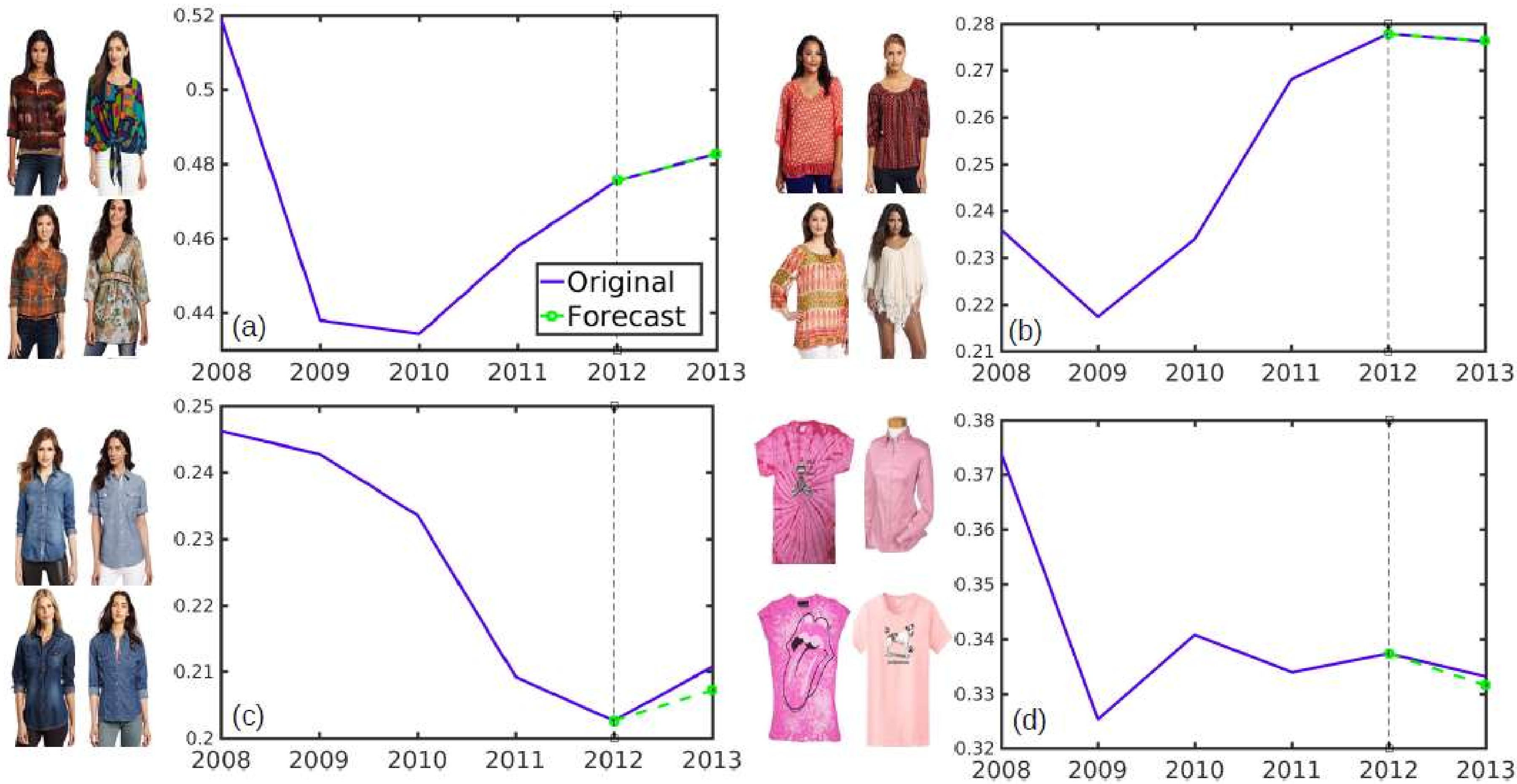
Text
Amanda Uprichard Women's Kiana Dress, Royal, Small

Tags
- Women
- Clothing
- Dresses
- Night Out & Cocktail
- Women's Luxury Brands

Visual trend forecasting

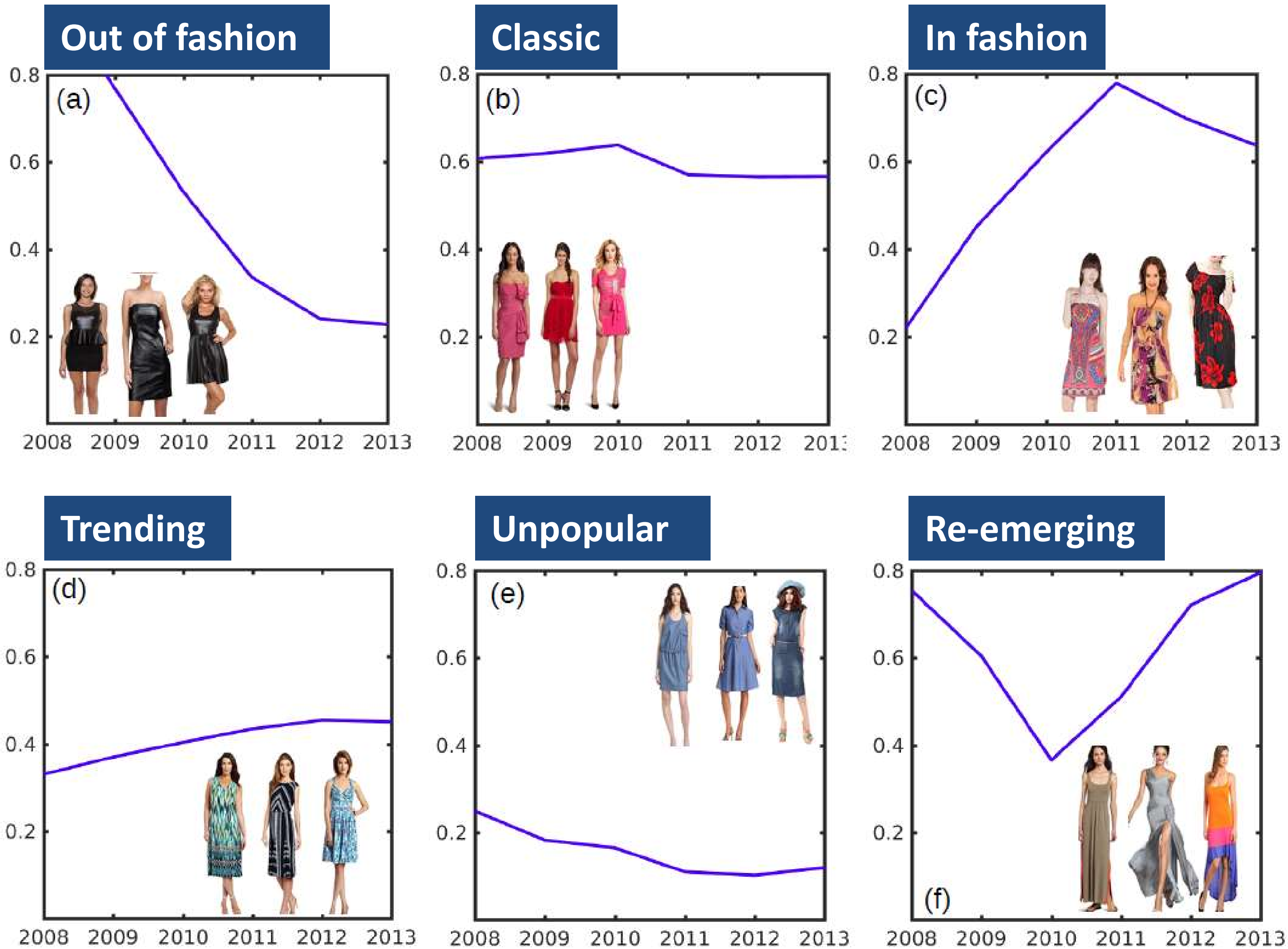
We predict the future popularity of each style

Amazon dataset [McAuley et al. SIGIR 2015]



Al-Halah et al., ICCV 2017

Lifecycle of a visual style



Al-Halah et al., ICCV 2017

Interpretable forecasts

What kind of fabric, texture, color will be popular next year?

(a) Texture

(b) Shape

This talk

- Subtle visual attributes
- Style discovery and forecasting
- Creating capsule wardrobes

Creating a "capsule" wardrobe

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

Creating a "capsule" wardrobe

Capsule pieces

Outfit #1 Outfit #2

Incompatible outfits!

Creating a “capsule” wardrobe

Capsule pieces

Outfit #1

Outfit #2

Outfit #3

All too similar…

Creating a “capsule” wardrobe

Q1: How to learn visual compatibility?

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

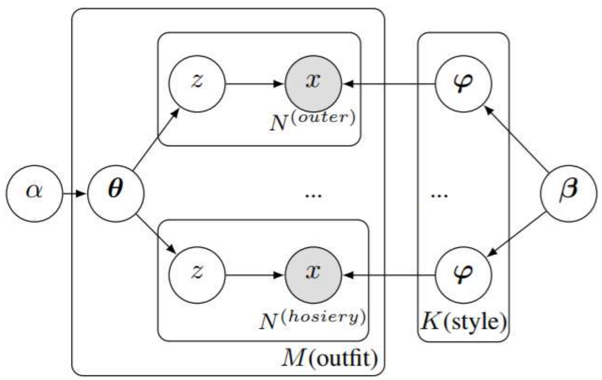
Manual curation
[Li 2017, Song 2017, Han 2017]

Supervised

Unlabeled in the wild photos?

Style model → Visual compatibility

Gauge mutual compatibility of garments via likelihood under topic model

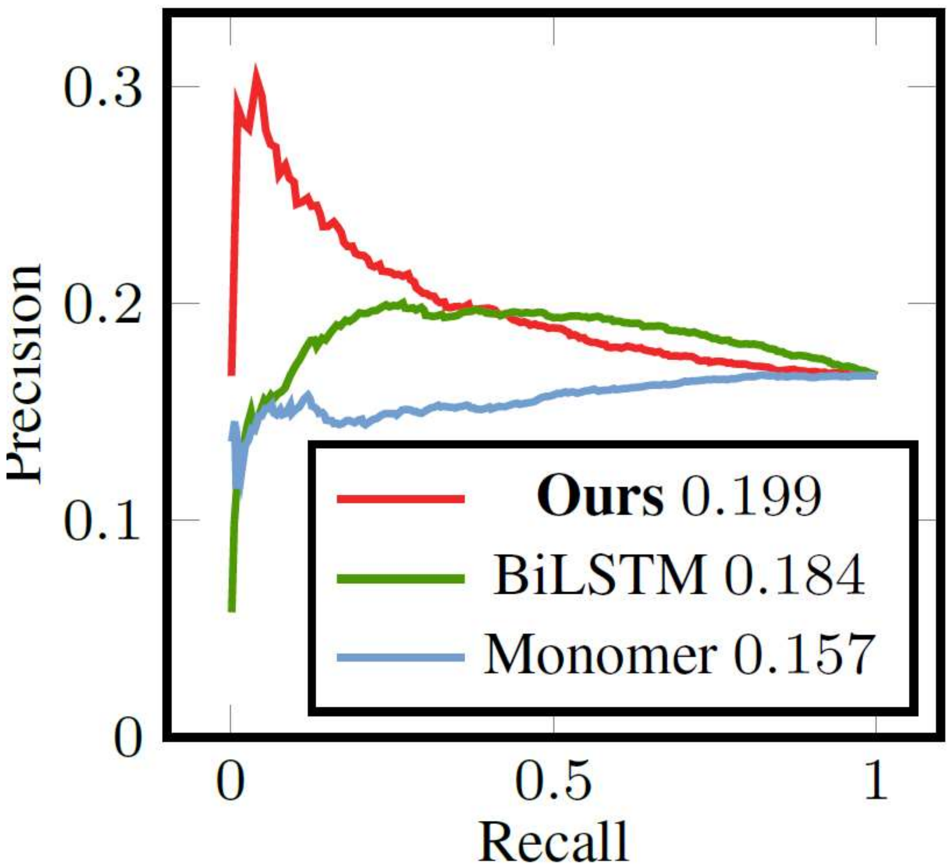


$$c(o_j) := p(o_j | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta)$$

Recall: an **outfit** is a mixture of (latent) **styles.**
A **style** is a distribution over **attributes.**

Hsiao & Grauman, CVPR 2018

Visual compatibility results



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 images

Visual compatibility results

Most compatible

Hsiao & Grauman, CVPR 2018

Visual compatibility results

Least compatible

Hsiao & Grauman, CVPR 2018

Q2: How to optimize a capsule?

Pose as *subset selection* problem

set of garments = argmax **compatibility** + **versatility**

Hsiao & Grauman, CVPR 2018

Capsule via subset selection

optimal set of composed outfits

se as *subset selection* problem

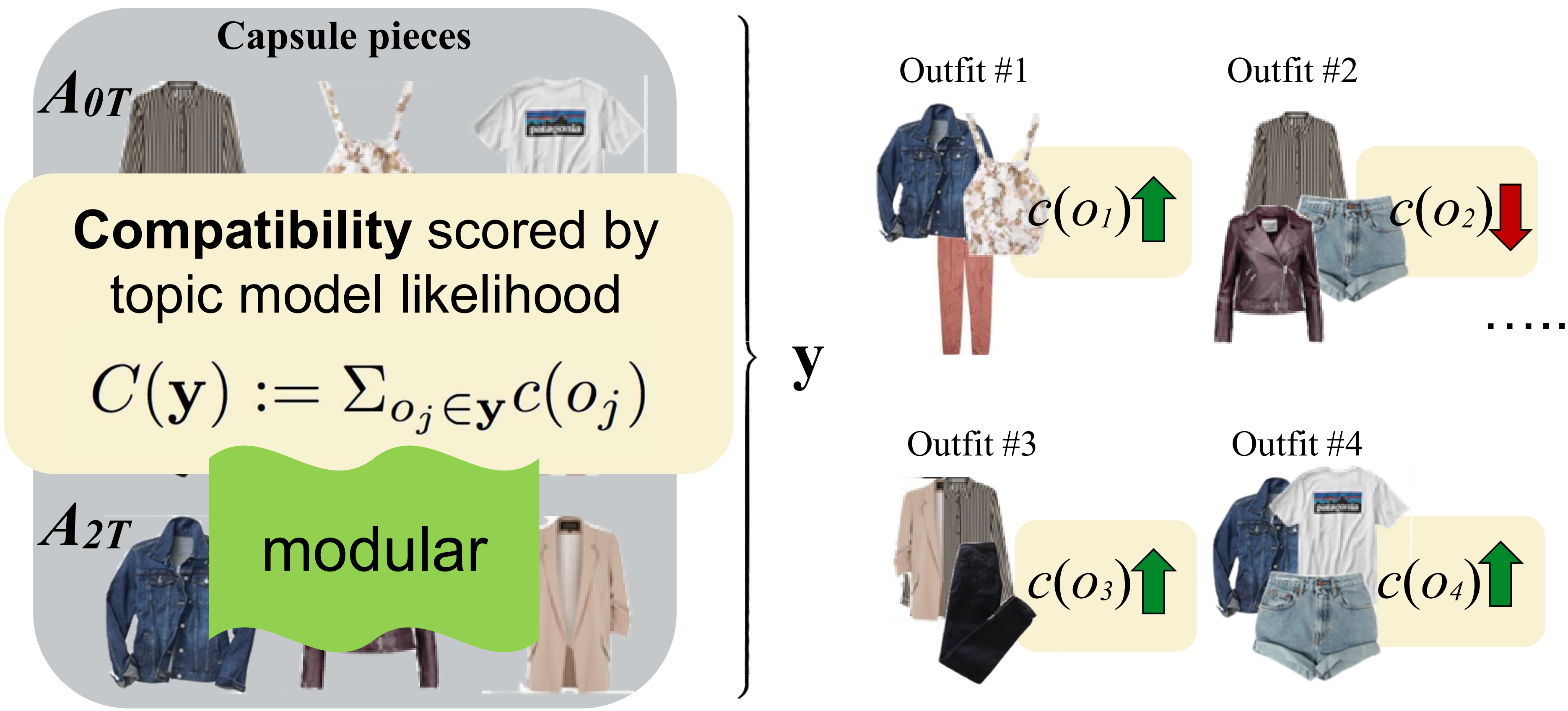
set of garments = argmax **compatibility** + **versatility**

Hsiao & Grauman, CVPR 2018

Capsule via subset selection

$$\mathbf{y}^* = \underset{\mathbf{y} \subseteq \mathcal{Y}}{\operatorname{argmax}}\, C(\mathbf{y}) + V(\mathbf{y}),$$

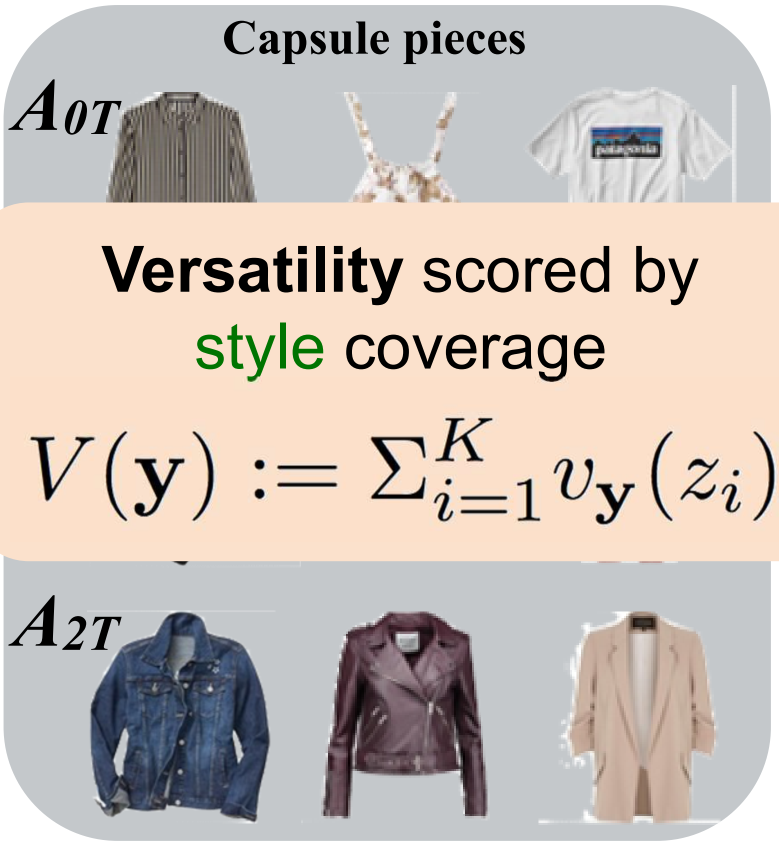
$$s.t.\ \mathbf{y} = A_{0T} \times\ A_{1T} \times \ldots \times A_{(m-1)T}$$



Capsule via subset selection

$$\mathbf{y}^* = \underset{\mathbf{y}\subseteq\mathcal{Y}}{\operatorname{argmax}}\, C(\mathbf{y}) + \boxed{V(\mathbf{y})},$$

$$s.t.\ \mathbf{y} = A_{0T} \times\ A_{1T} \times \ldots \times A_{(m-1)T}$$



Versatility scored by style coverage

$$V(\mathbf{y}) := \Sigma_{i=1}^{K} v_{\mathbf{y}}(z_i)$$

$$v_{\mathbf{y}}(z_i) = 1 - \prod_{o_j\in\mathbf{y}} (1 - P(z_i|o_j))$$

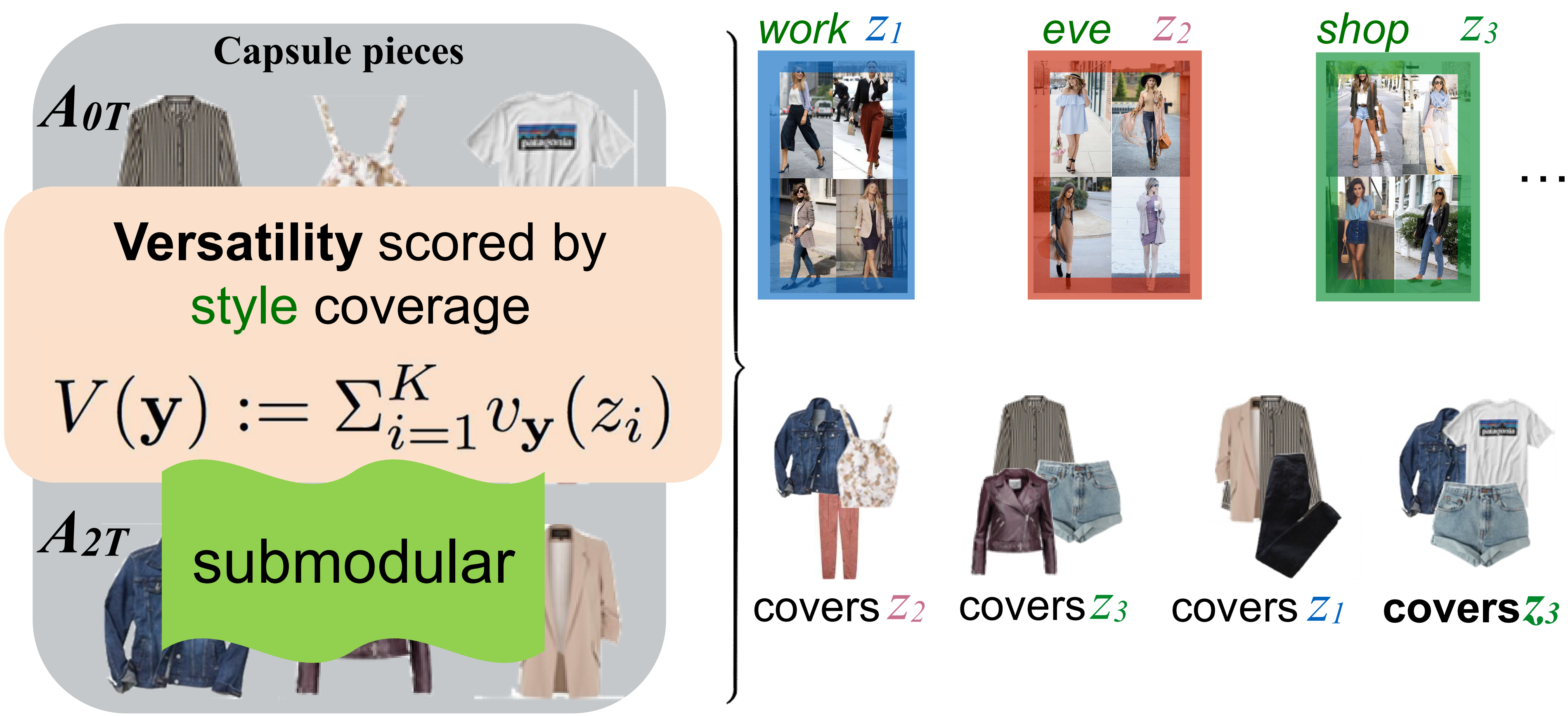
style

outfit

Capsule via subset selection

$$\mathbf{y}^* = \underset{\mathbf{y}\subseteq\mathcal{Y}}{\operatorname{argmax}}\, C(\mathbf{y}) + \boxed{V(\mathbf{y})},$$

$$s.t.\ \mathbf{y} = A_{0T} \times\ A_{1T} \times \ldots \times A_{(m-1)T}$$

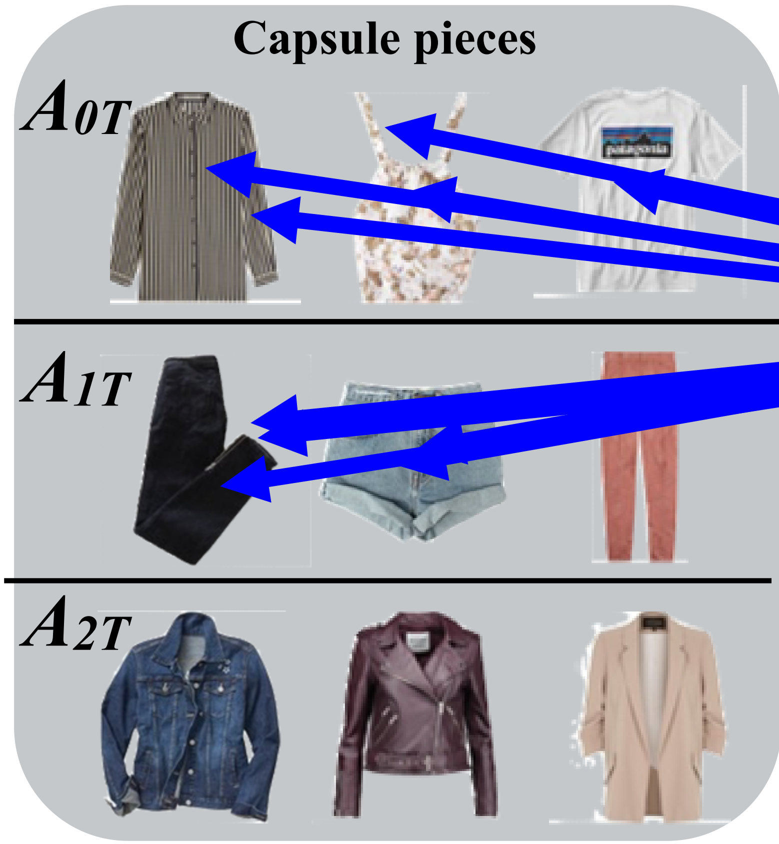


Capsule via subset selection

optimal **set of outfits**

$$\mathbf{y}^* = \underset{\mathbf{y} \subseteq \mathcal{Y}}{\operatorname{argmax}}\, C(\mathbf{y}) + V(\mathbf{y}),$$

$$s.t.\ \mathbf{y} = A_{0T} \times\ A_{1T} \times \ldots \times A_{(m-1)T}$$



$\mathbf{y}$

Compatibility scored by topic model likelihood

$C(\mathbf{y})$ ... $\in \mathbf{y}\, c(o_i)$

We ... ke solution for which ... show (sub)mod... olds.

Versa... cored by st... rage

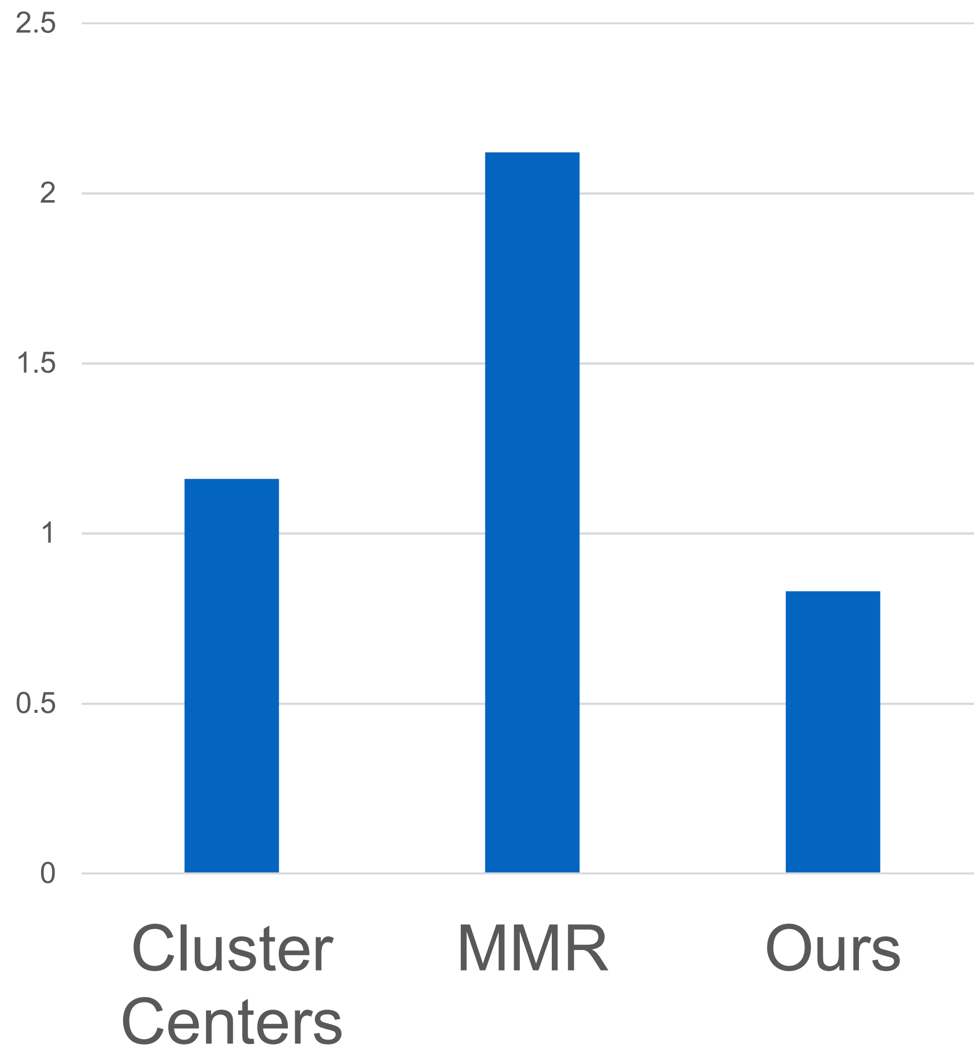
$V(\mathbf{y}) = \Sigma_{i=1}^{K} v_{\mathbf{y}}(z_i)$

But each addition is a garment!

submodular

Quantifying capsule error

Distance from "ground truth" manually curated capsules from Polyvore.com



Hsiao & Grauman, CVPR 2018

Human subject study

14 subjects, female, ages 20’s-60’s

2) Which is better *

◯ a

◯ b

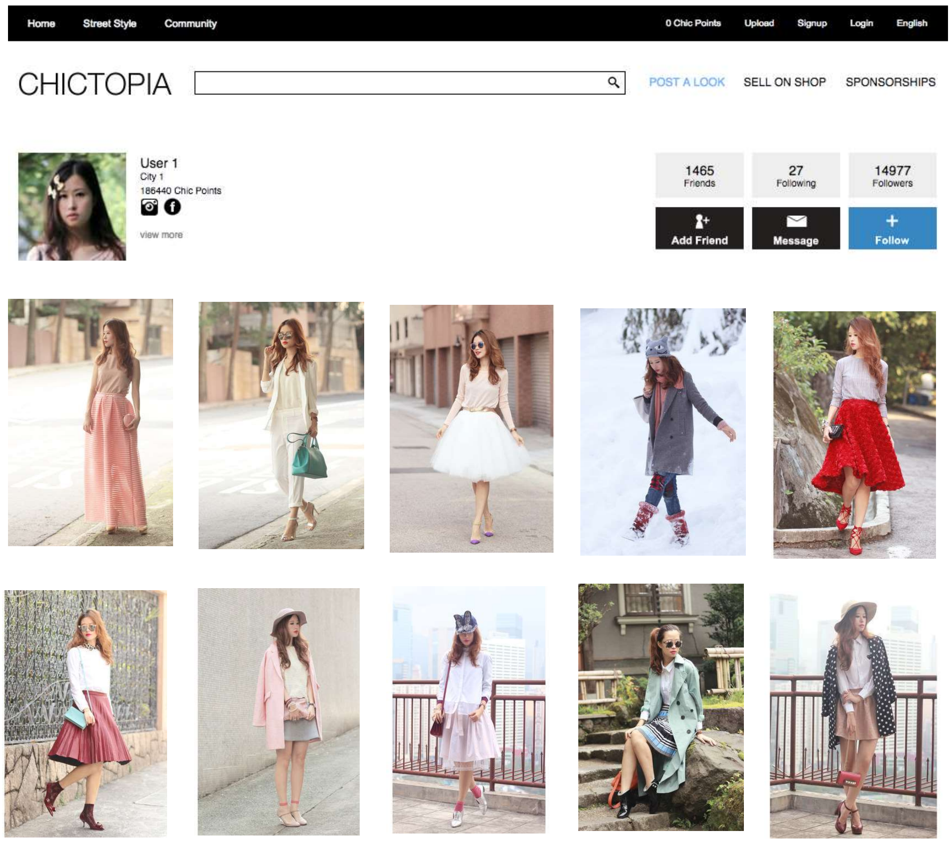
◯ EQUAL

Iterative preferred **59%** of the time
vs. naïve greedy

Hsiao & Grauman, CVPR 2018

Example personalized capsule

Discover user's style preferences from album

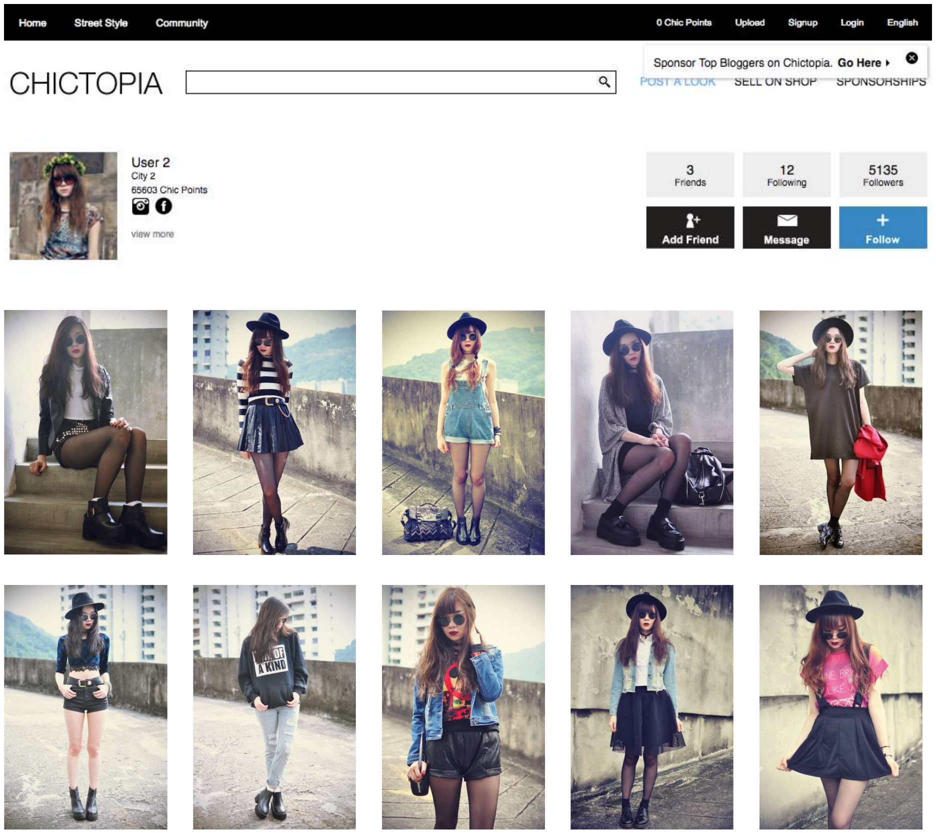


Personalized capsule

Hsiao & Grauman, CVPR 2018

Example personalized capsule

Discover user's style preferences from album



Personalized capsule

Hsiao & Grauman, CVPR 2018

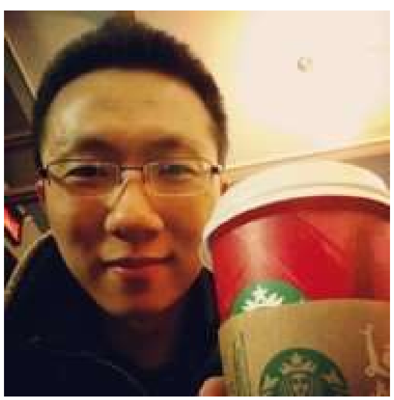
Summary

THE UNIVERSITY OF TEXAS AT AUSTIN

- Visual style introduces new problems for computer vision beyond traditional recognition

- New ideas and methods for:
  - Subtle visual comparisons
  - Style discovery and forecasting
  - Capsule wardrobe creation

Aron Yu



Kimberly Hsiao

Ziad Al-Halah

Steven Chen

Papers

- **Learning the Latent "Look": Unsupervised Discovery of a Style-Coherent Embedding from Fashion Images**. W-L. Hsiao and K. Grauman. In Proceedings of the International Conference on Computer Vision (ICCV), Venice, Italy, Oct 2017.
- **Creating Capsule Wardrobes from Fashion Images**. W-L. Hsiao and K. Grauman. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, June 2018.
- **Compare and Contrast: Learning Prominent Visual Differences**. S. Chen and K. Grauman. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, June 2018.
- **Fashion Forward: Forecasting Visual Style in Fashion**. Z. Al-Halah, R. Stiefelhagen, and K. Grauman. In Proceedings of the International Conference on Computer Vision (ICCV), Venice, Italy, Oct 2017.
- **Semantic Jitter: Dense Supervision for Visual Comparisons via Synthetic Images.** A. Yu and K. Grauman. In Proceedings of the International Conference on Computer Vision (ICCV), Venice, Italy, Oct 2017.

`Code and data:`
`http://www.cs.utexas.edu/~grauman/research/pubs.html`