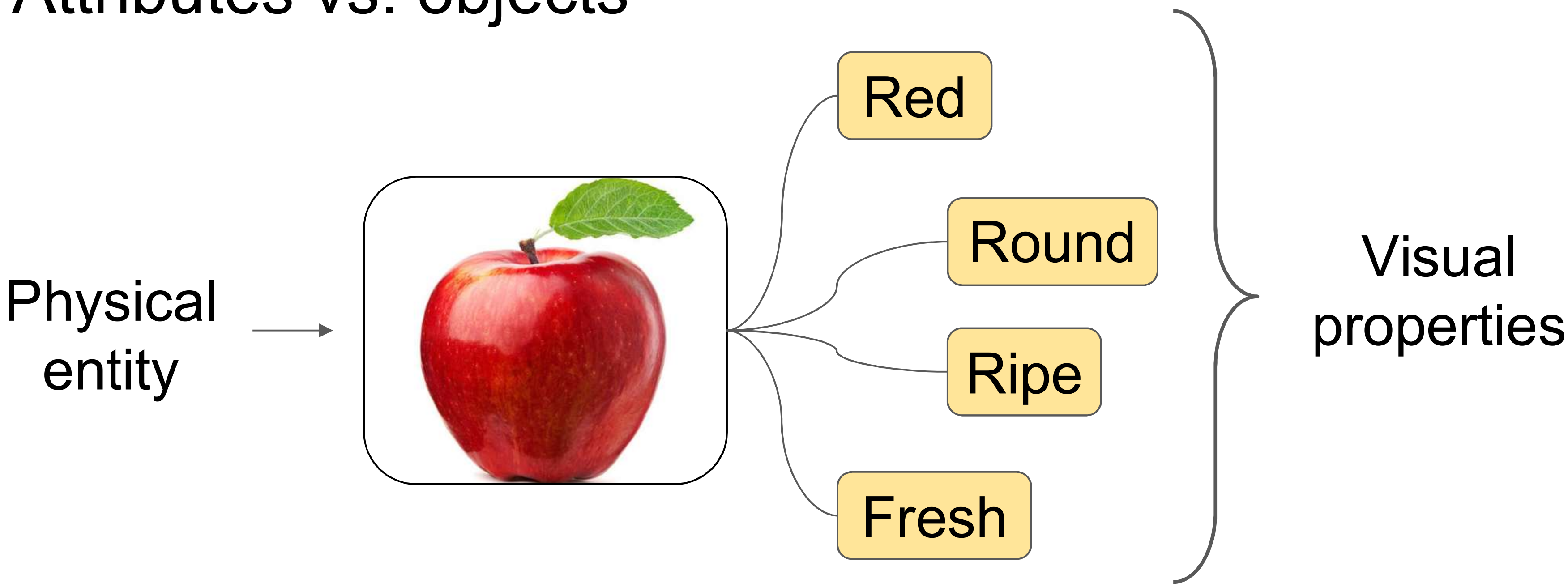
The language of visual attributes

Kristen Grauman
Facebook AI Research
University of Texas at Austin

Attributes vs. objects



Value of attributes

"Find a more formal shoe"

Visual search

Zebras have stripes and four legs…

Zero-shot learning

A lone cow grazes in a green pasture.

Image/video description

What color is the beak?

Interactive recognition

[Ferrari & Zisserman 2007, Kumar et al. 2008, Farhadi et al. 2009, Lampert et al. 2009, Wang & Mori 2010, Berg et al. 2010, Parikh & Grauman 2011, Branson et al. 2010, Kovashka et al. 2012, Kulkarni et al. 2011, Wang et al. 2016, Liu et al. 2015, Singh et al. 2016, …]

The language of visual attributes

- Attributes as operators

  Attributes:adjectives that *modify* objects:nouns

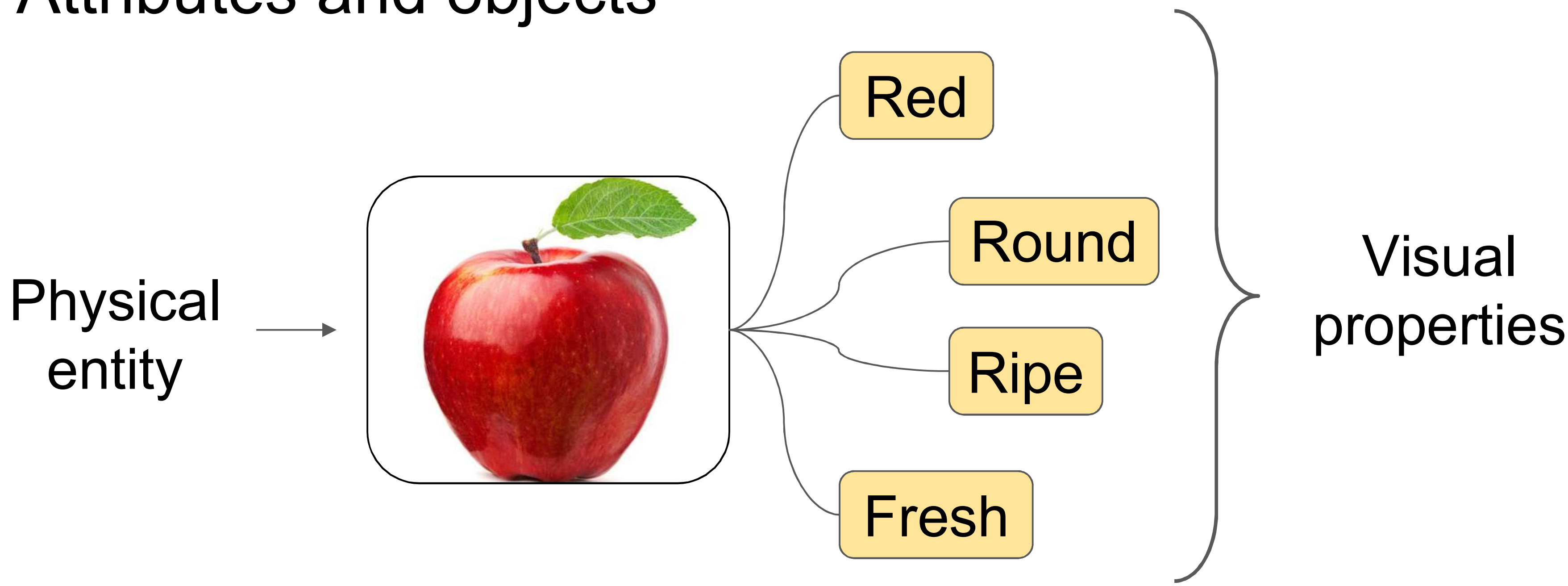
- Attributes for comparisons

  Relative differences that people first describe

- Attributes for visual styles

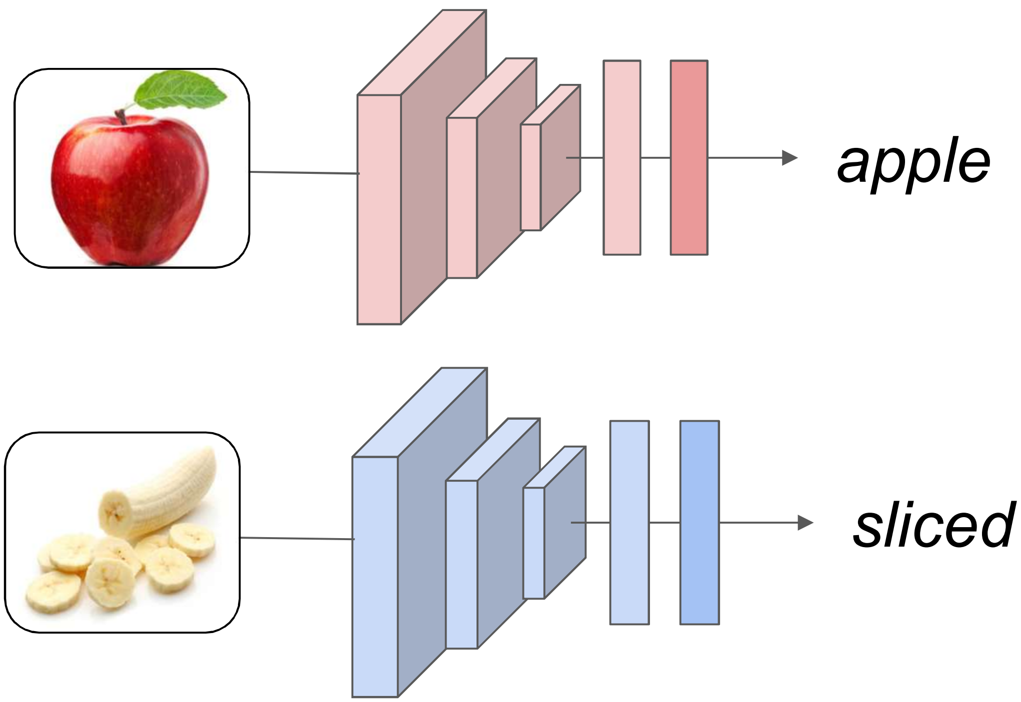
  Semantic topic models for data-driven styles

Attributes and objects



Attributes and objects are fundamentally different

Attribute and Object Representations

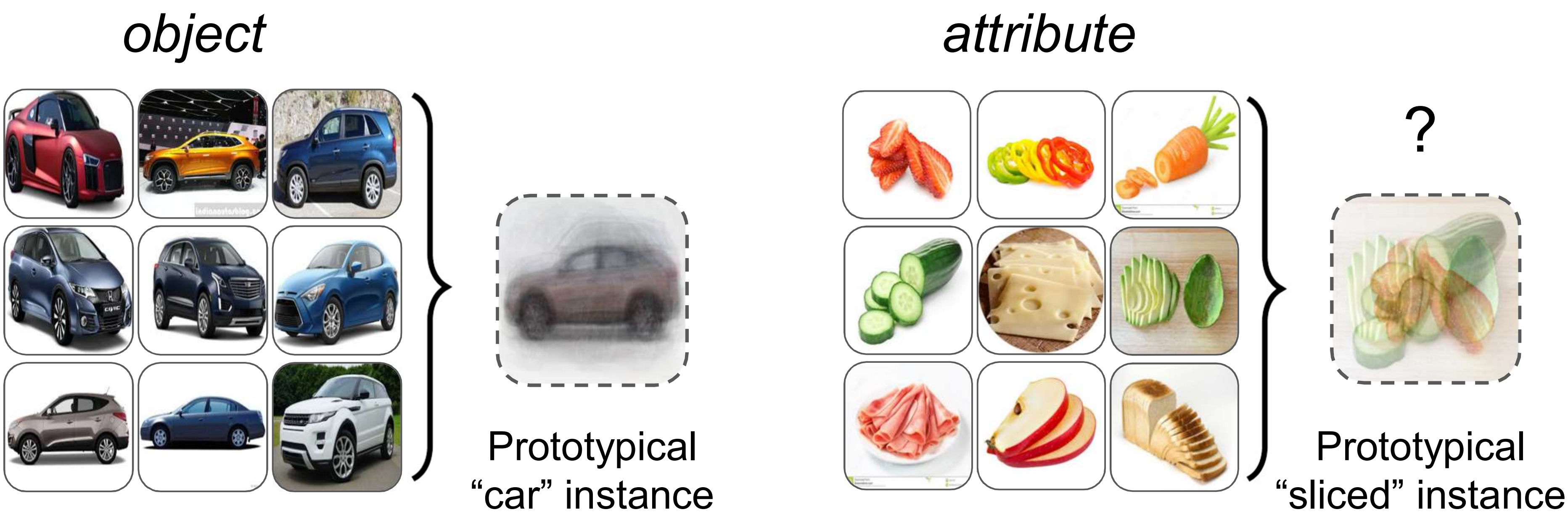


Yet status quo treats attributes and objects the same...

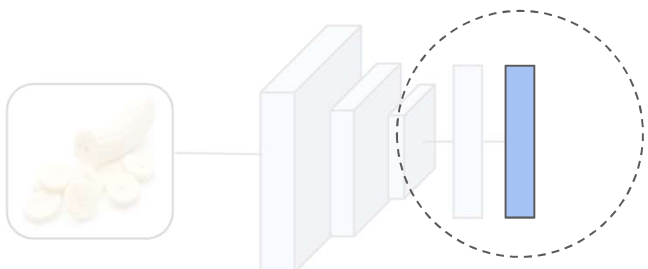
As latent vector encodings

e.g., Wang CVPR16, Liu CVPR15, Singh ECCV16, Lu CVPR17, Su ECCV16,…

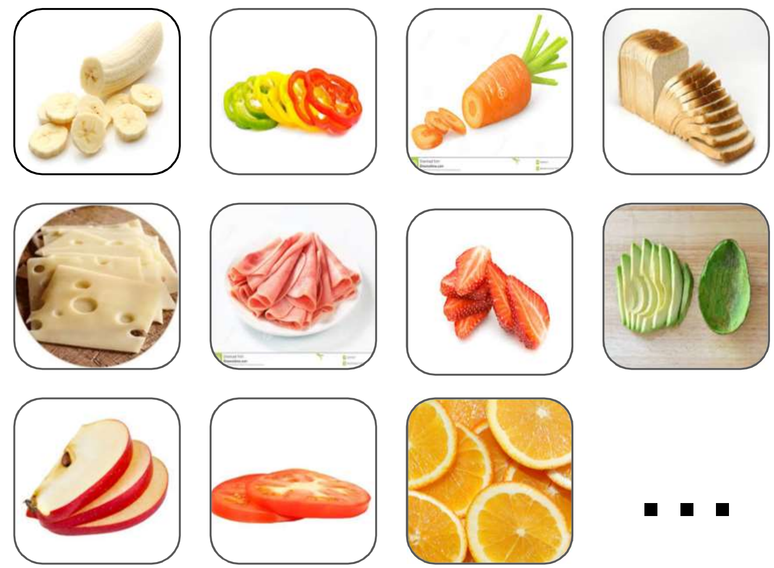
Attribute vs. Object Representations



Challenges for the status quo approach

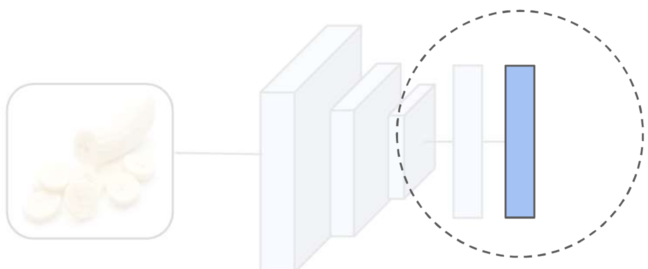


Object-agnostic
attribute representation



Has to capture interactions with **every** object

Challenges for the status quo approach



Object-agnostic
attribute representation

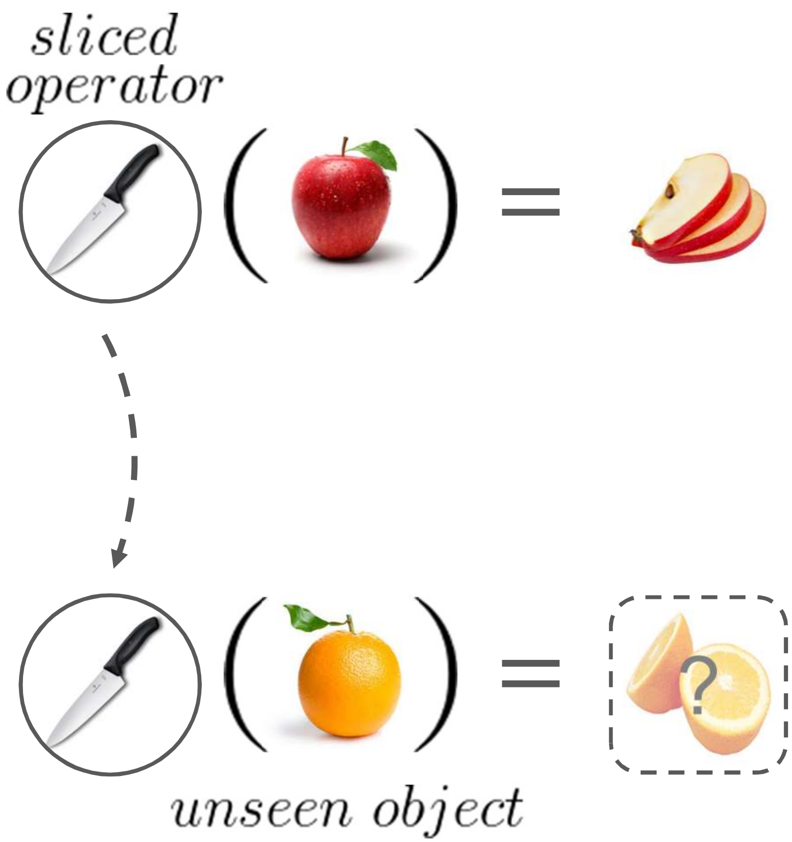
vs.

Old car

Old man

Has to capture attributes' distinct manifestations

Our idea – Attributes as operators

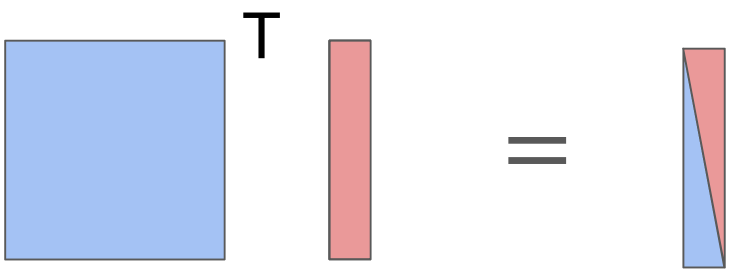


Attributes are ***operators*** that transform object encodings

[Nagarajan & Grauman, ECCV 2018]

Our idea – Attributes as operators

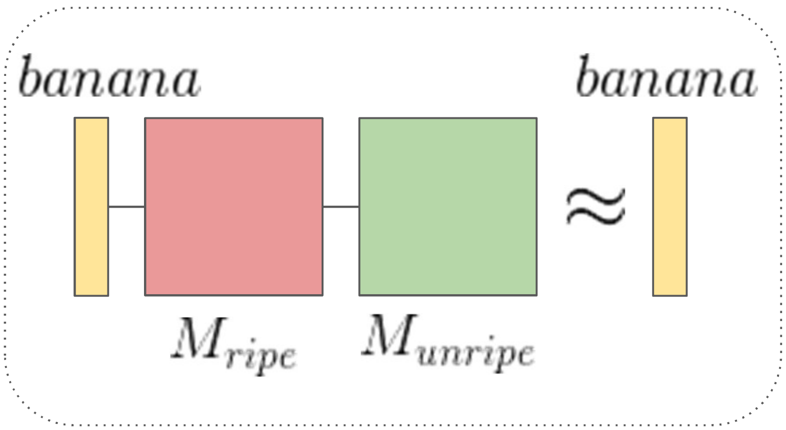
Objects are **vectors**
Attributes are **operators**



Composition is:
an **attribute operator**
transforming an **object vector**

[Nagarajan & Grauman, ECCV 2018]

Linguistically inspired regularizers

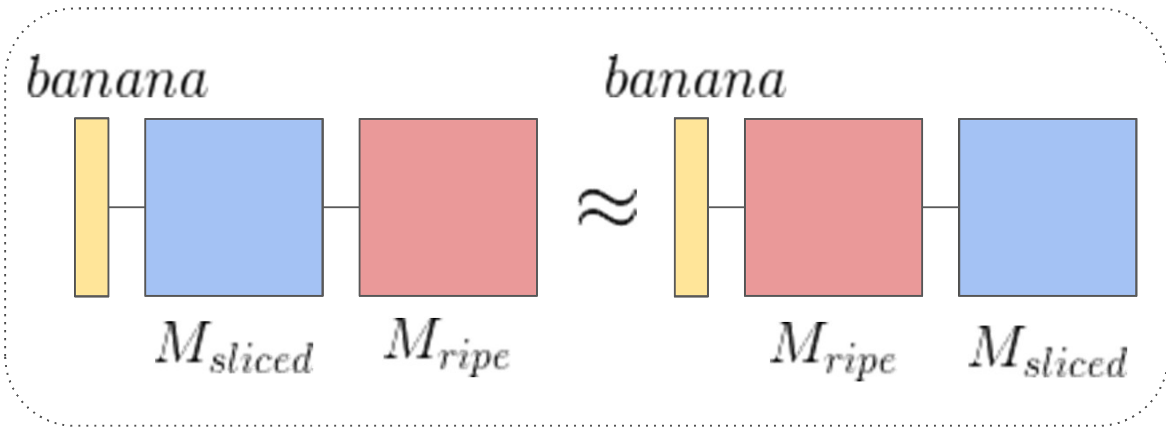


Antonym-consistency:

"Unripe should **undo** the effect of ripe"

[Nagarajan & Grauman, ECCV 2018]

Linguistically inspired regularizers

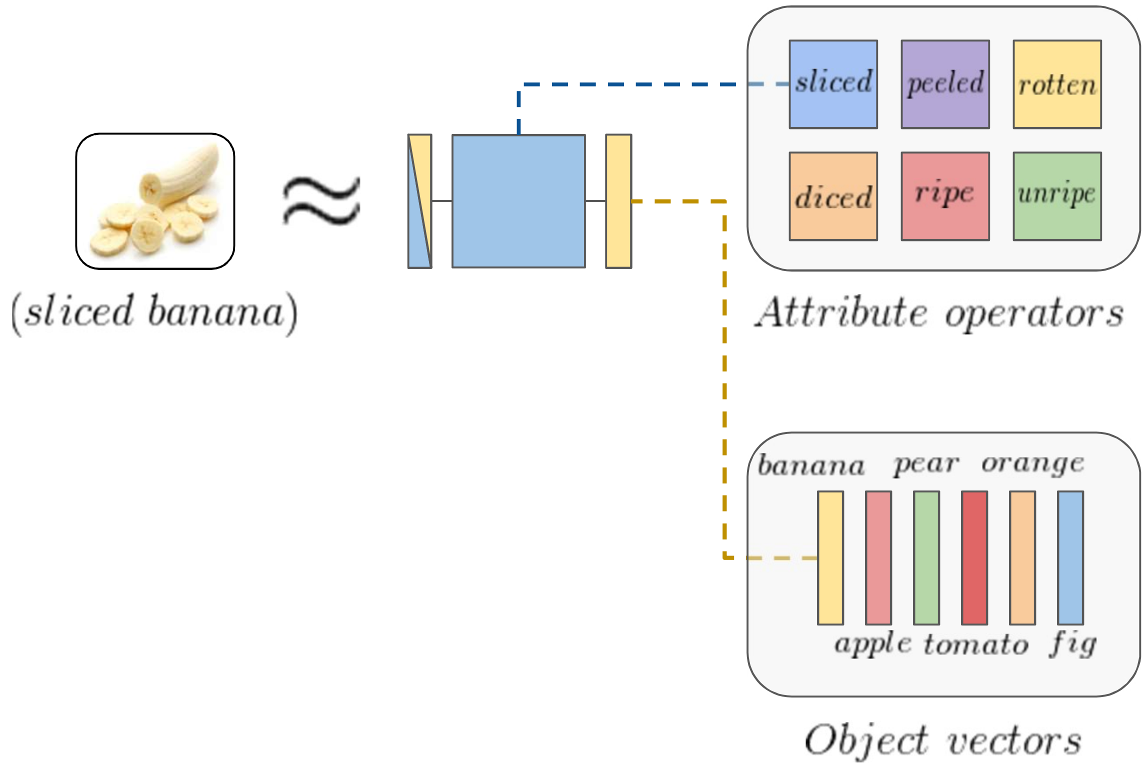


Attribute commutation:

Attribute effects should **stack**.

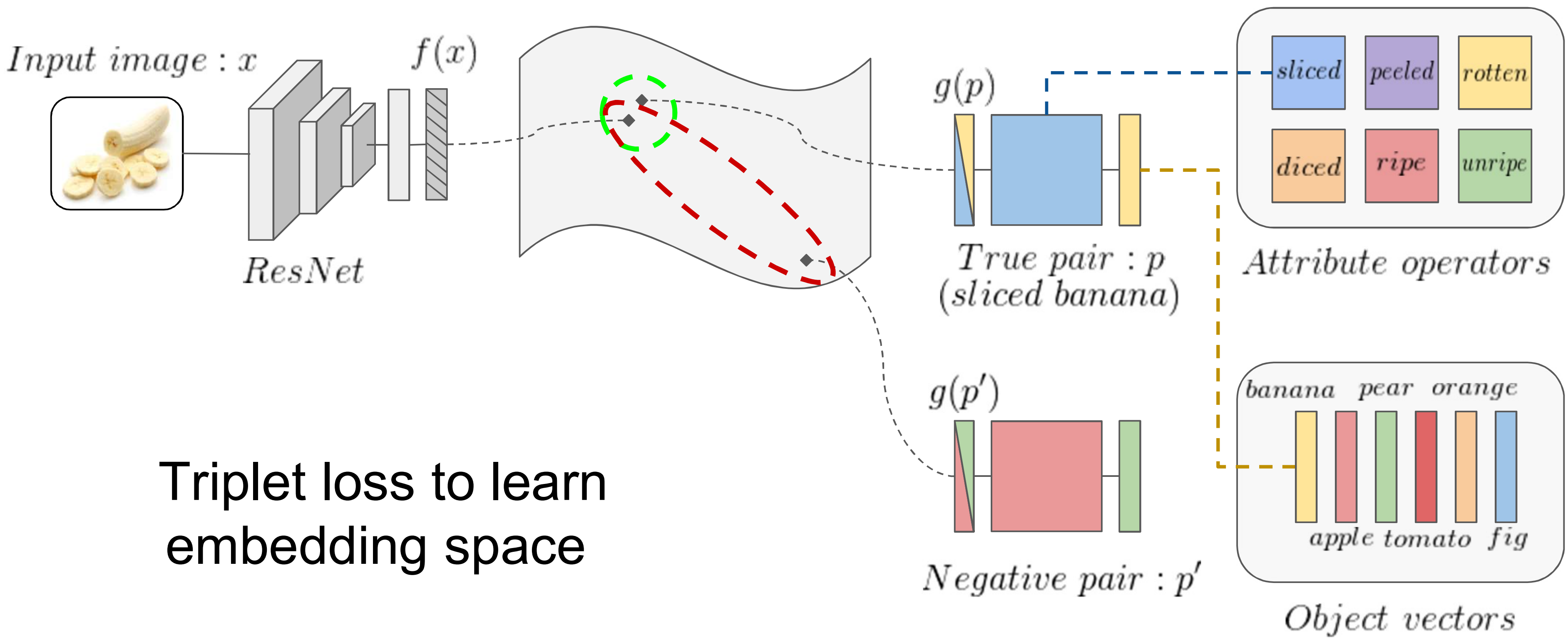
[Nagarajan & Grauman, ECCV 2018]

Learning attribute operators



[Nagarajan & Grauman, ECCV 2018]

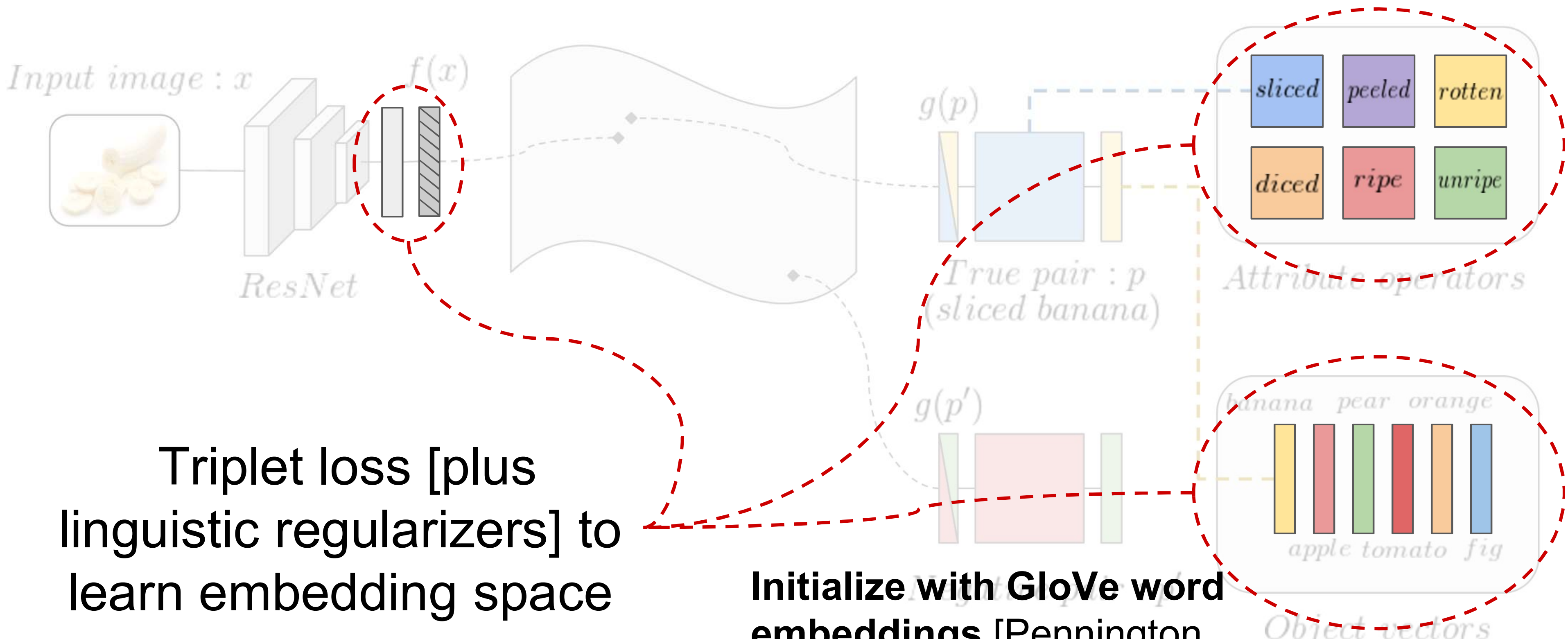
Learning attribute operators



Triplet loss to learn embedding space

[Nagarajan & Grauman, ECCV 2018]

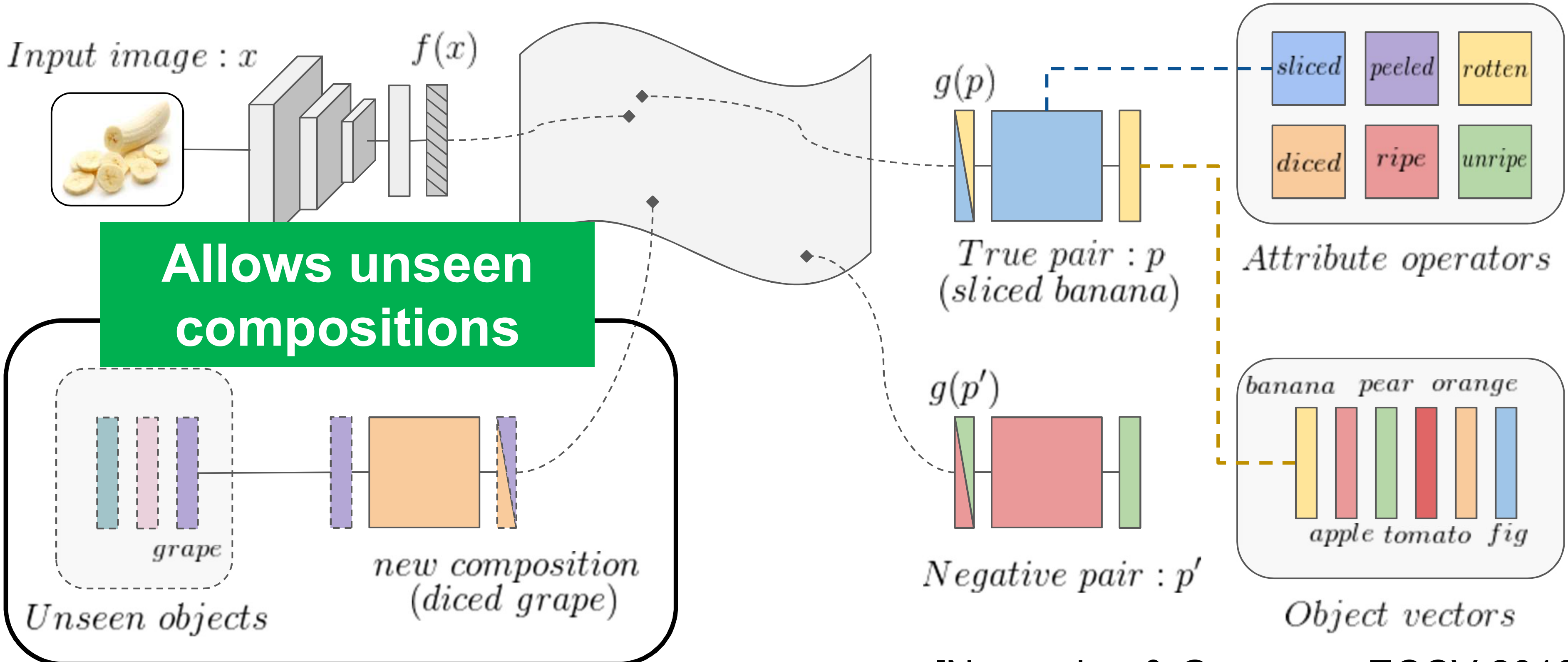
Learning attribute operators



Triplet loss [plus linguistic regularizers] to learn embedding space

Initialize with GloVe word embeddings [Pennington et al. EMNLP 2014]

Learning attribute operators



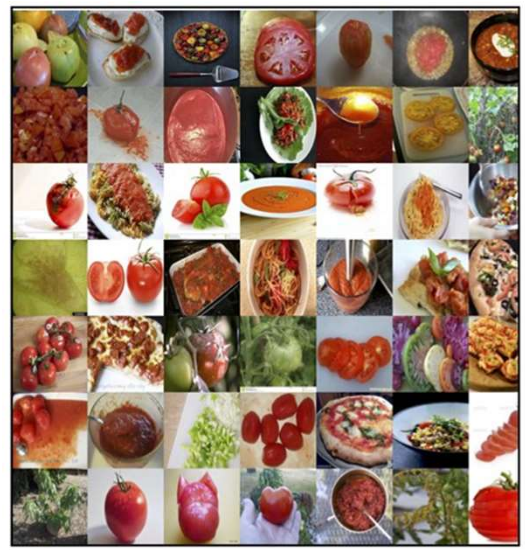
[Nagarajan & Grauman, ECCV 2018]

Evaluation

UT-Zappos 50k
(Yu & Grauman, CVPR 14)

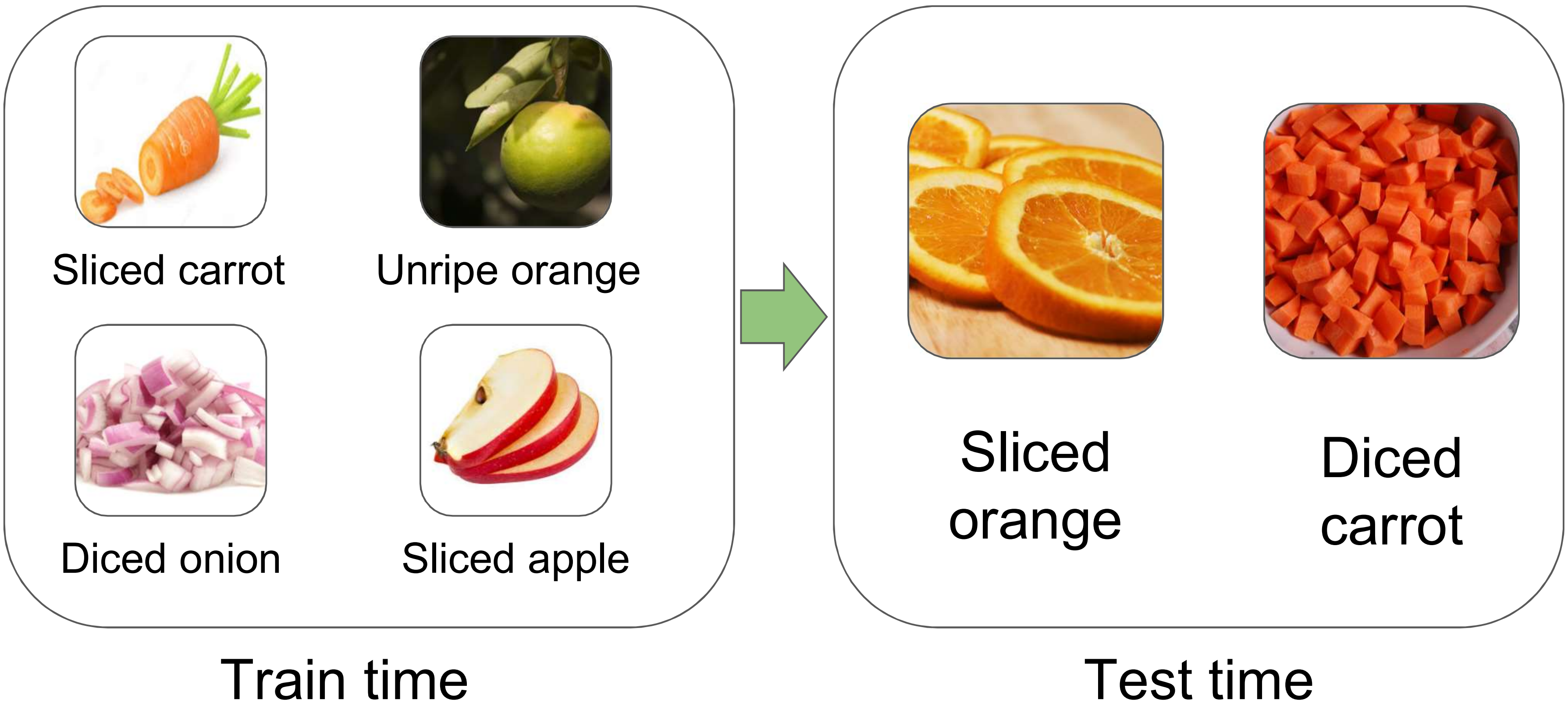
16 attributes x 12 objects

MIT States
(Isola et al., CVPR 15)

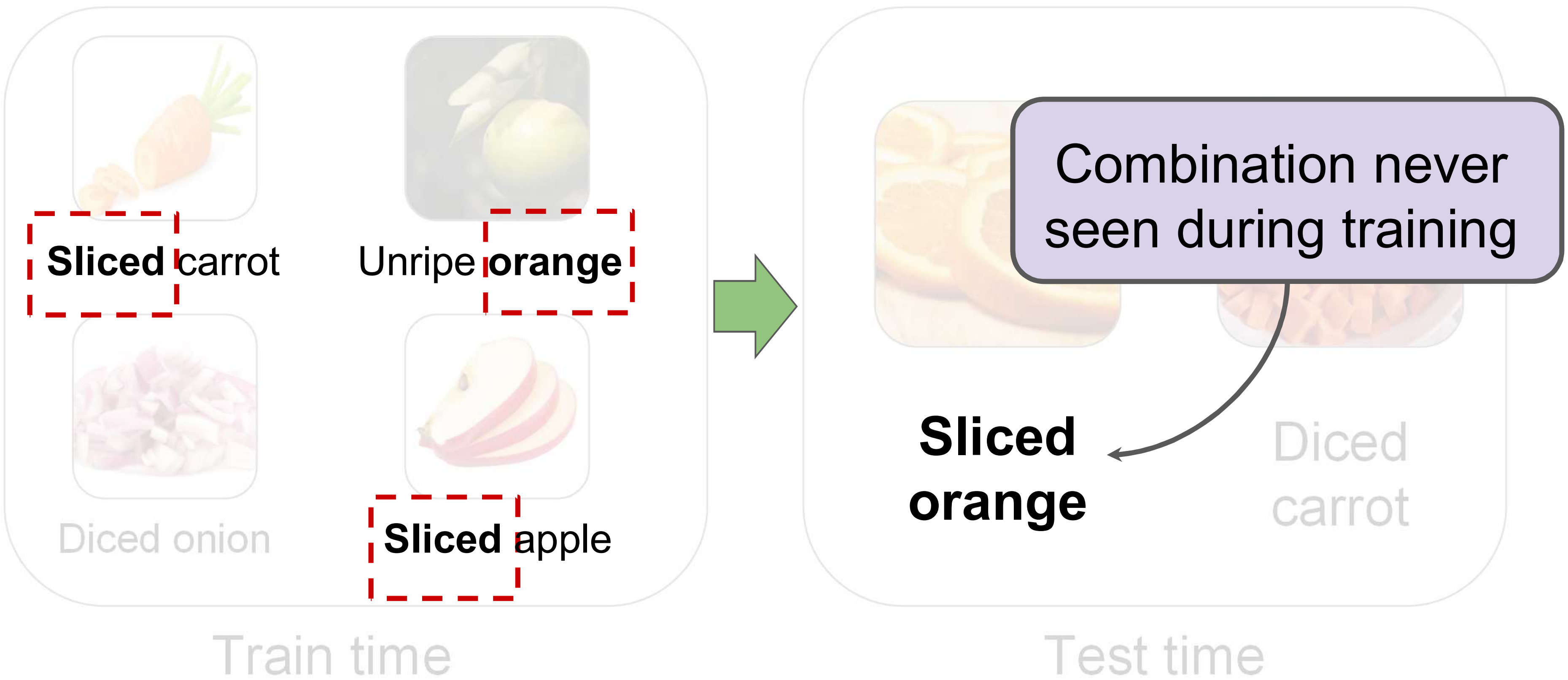


115 attributes x 245 objects

Evaluating our composition model



Evaluating our composition model



Results – Attribute+object composition recognition

MIT States: 6% increase in open world (3% h-mean)

UT-Zap: 14% increase in open world (12% h-mean)

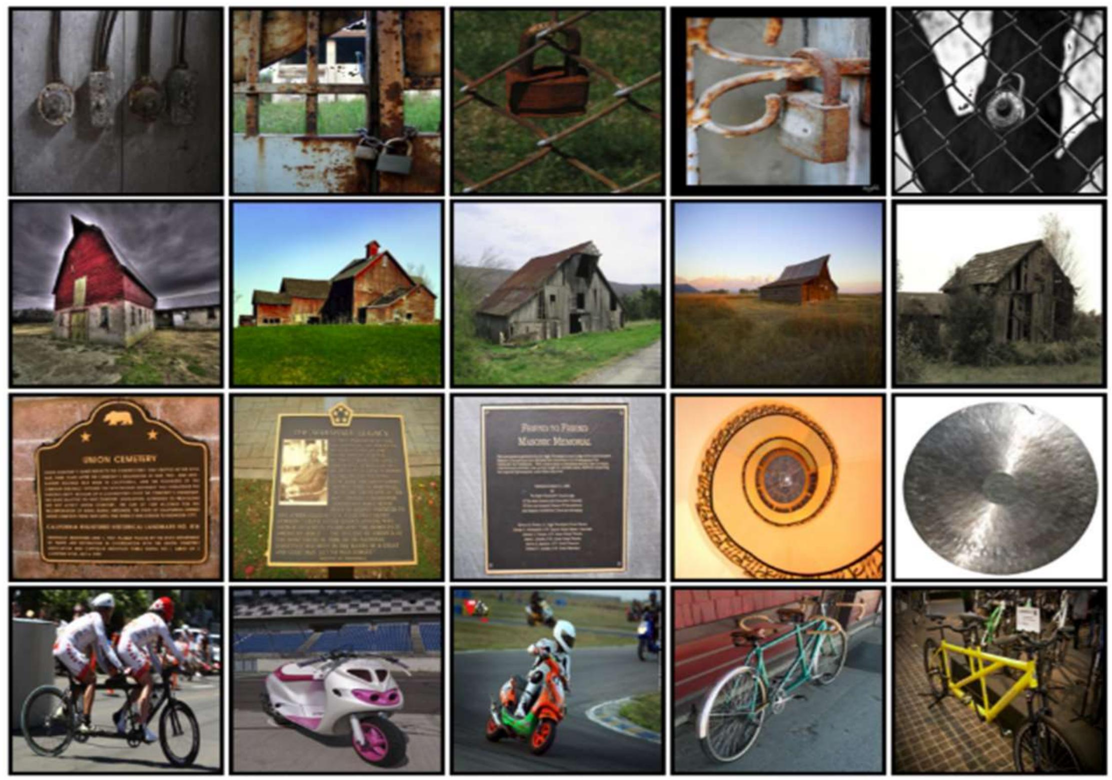
	MIT-States		
	closed	open	h-mean
CHANCE	0.1	0.05	0.1
VISPROD(SVM)	11.1	2.4	3.9
VISPROD(NN)	13.9	2.8	4.7
ANALOGOUSATTR#	1.4	0.2	0.4
REDWINE*	12.5	3.1	5.0
LABELEMBED	13.4	3.3	5.3
LABELEMBED+	**14.8**	5.7	8.2
OURS	12.0	**11.4**	**11.7**

*Misra et al. CVPR 2017
#Chen & Grauman CVPR 2014

[Nagarajan & Grauman, ECCV 2018]

Results - Retrieving unseen (unseen) compositions

Rusty Lock



query

Nearest Images in ImageNet

The language of visual attributes

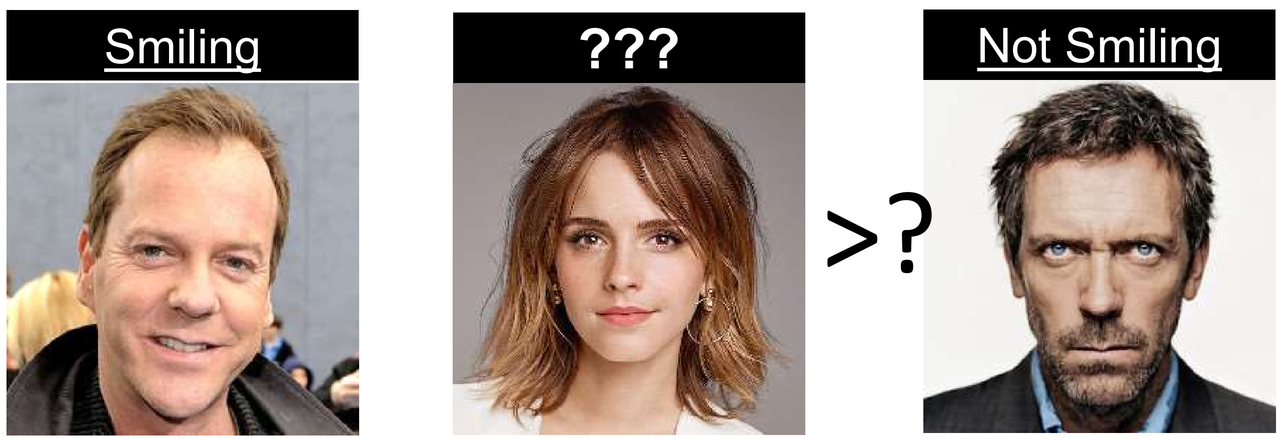
- Attributes as operators

  Attributes:adjectives that *modify* objects:nouns
- Attributes for comparisons

  Relative differences that people first describe
- Attributes for visual styles

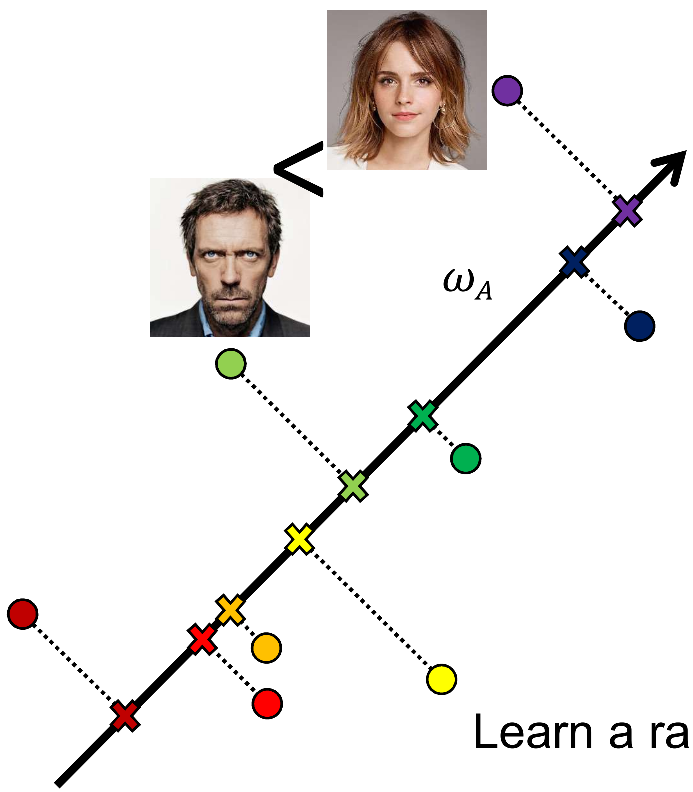
  Semantic topic models for data-driven styles

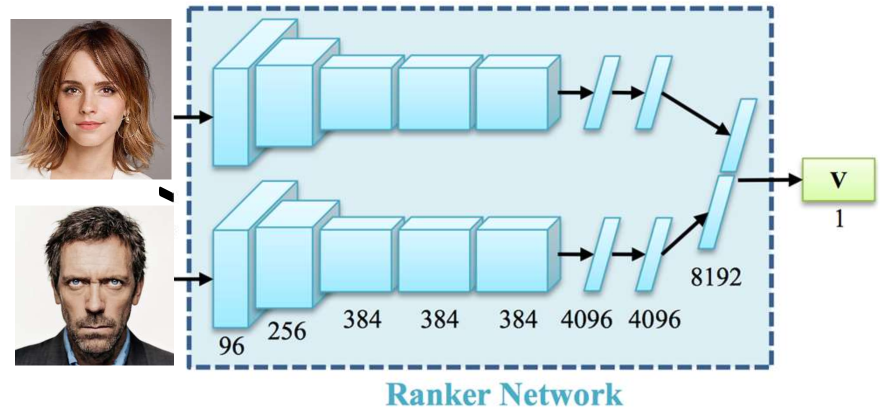
Relative attributes



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

Relative attributes



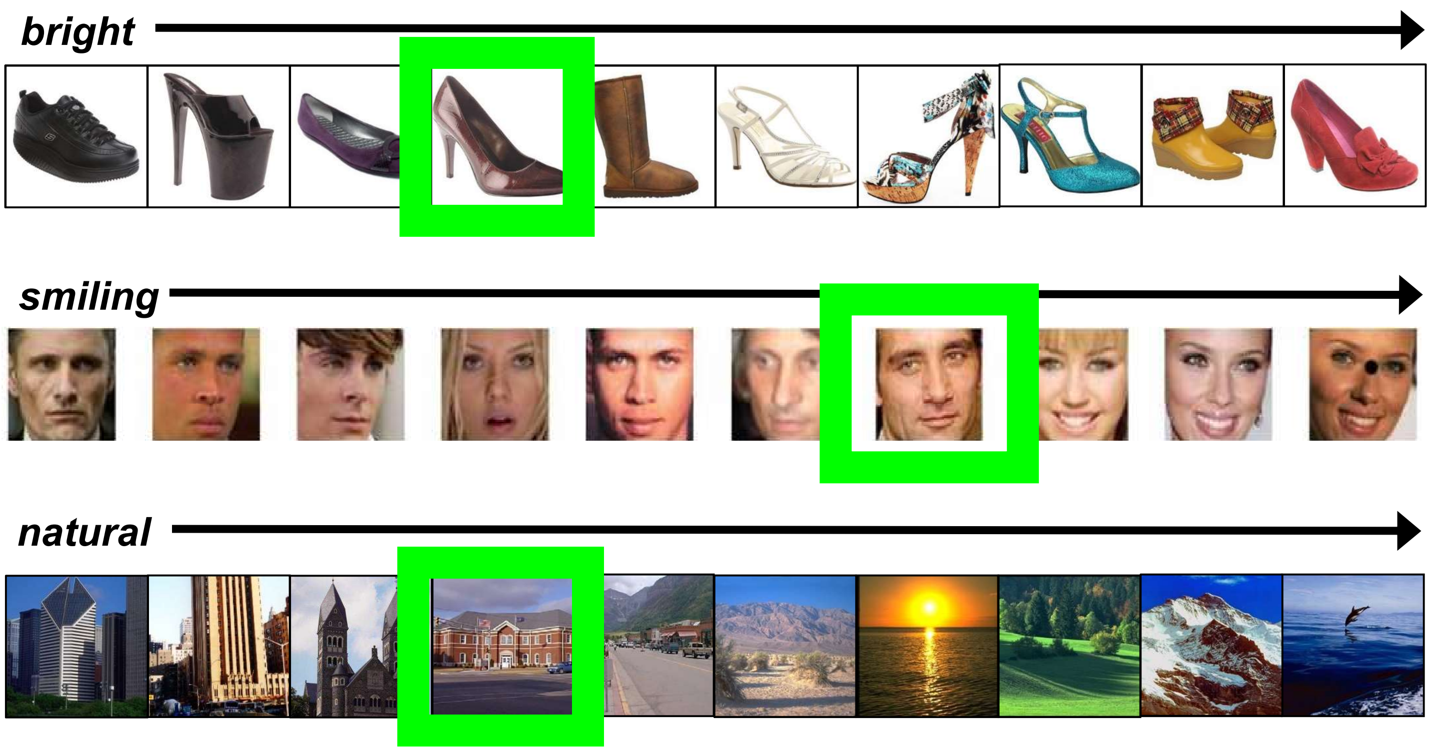


Learn a ranking function per attribute

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

Relative attributes

Compare images by an attribute's "strength"



[Parikh & Grauman, ICCV 2011]

Challenge #1: fine-grained comparisons

Which is more sporty?

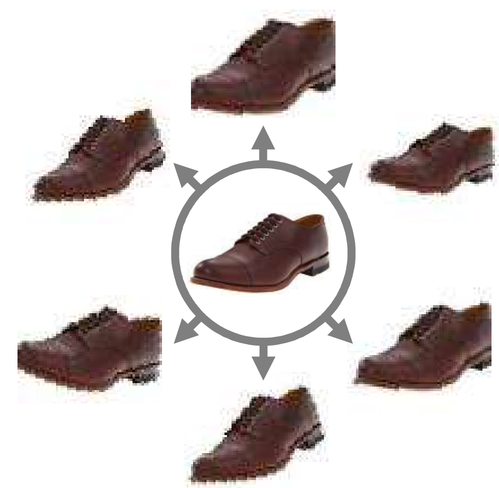
Sparsity of supervision problem:

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

vs.



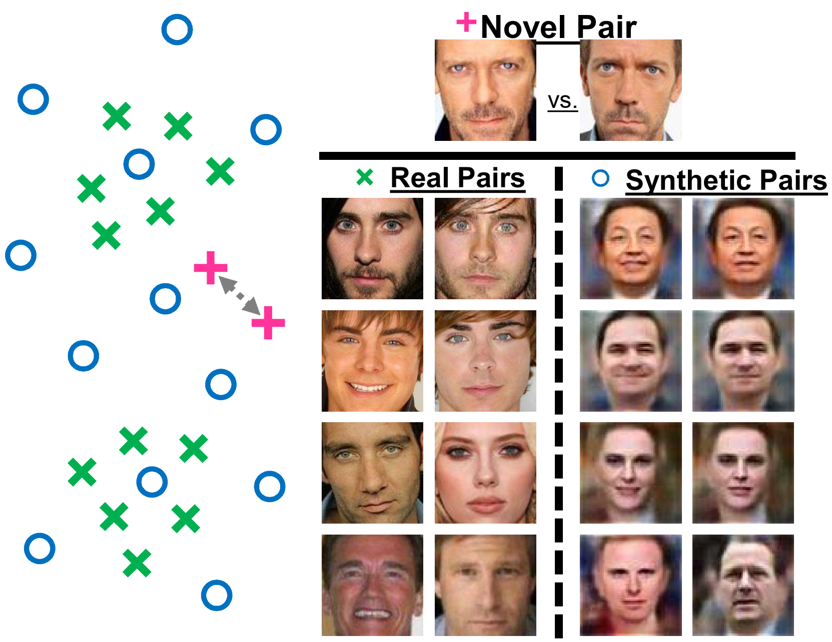
Our idea:
Semantic jitter

Status quo:
Low-level jitter

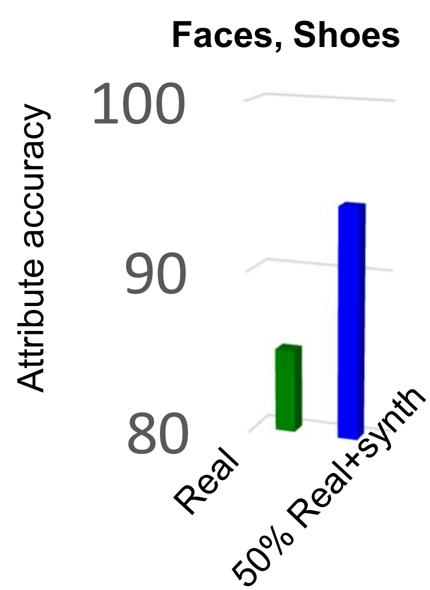
Yu & Grauman, ICCV 2017

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

Challenge #2: 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

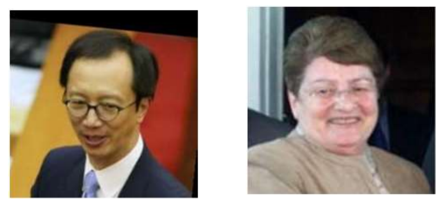
Approach: What causes prominence?

Prominent Difference:

- Large difference in attribute strength:

Colorful

- Unusual and uncommon attribute occurrences:



Visible Forehead

- Absence of other noticeable differences:

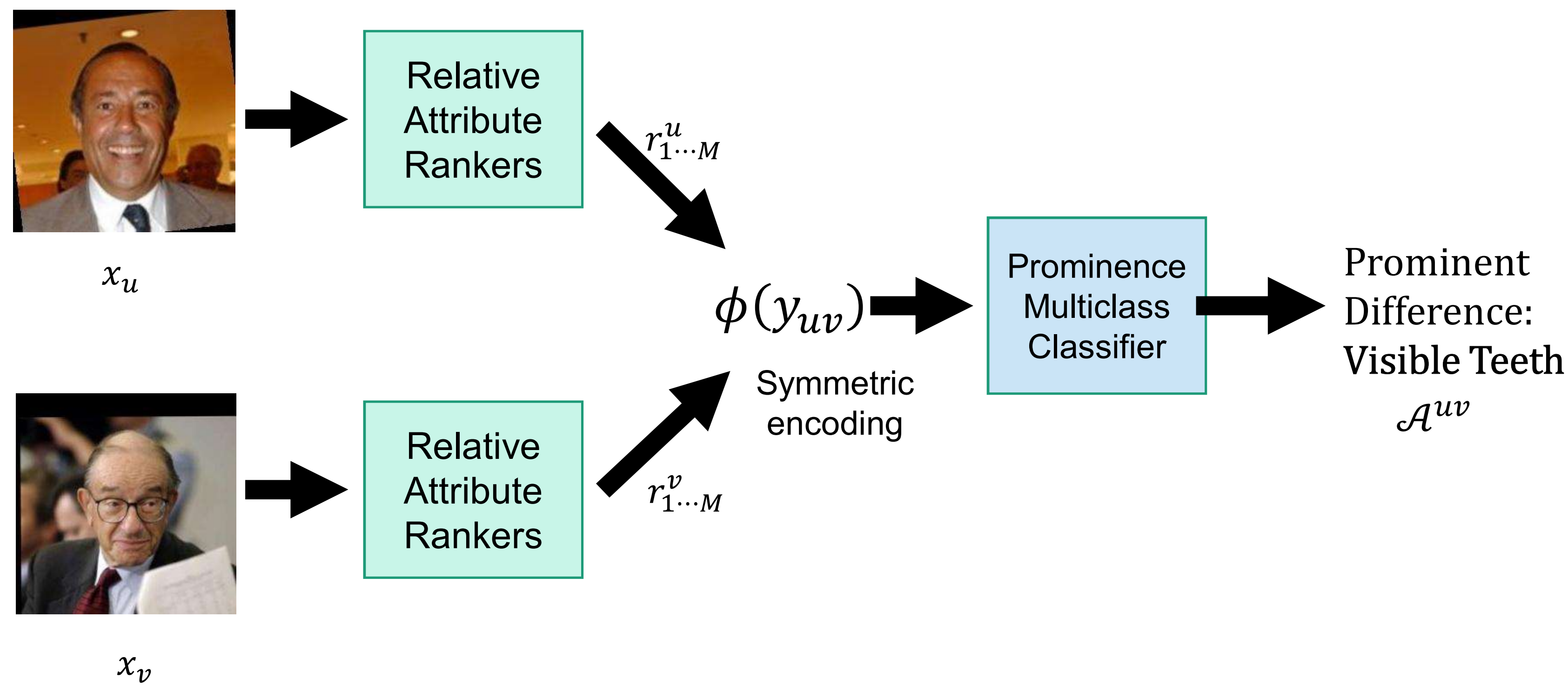
Dark Hair

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

Chen & Grauman, CVPR 2018

Approach: Predicting prominent differences

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

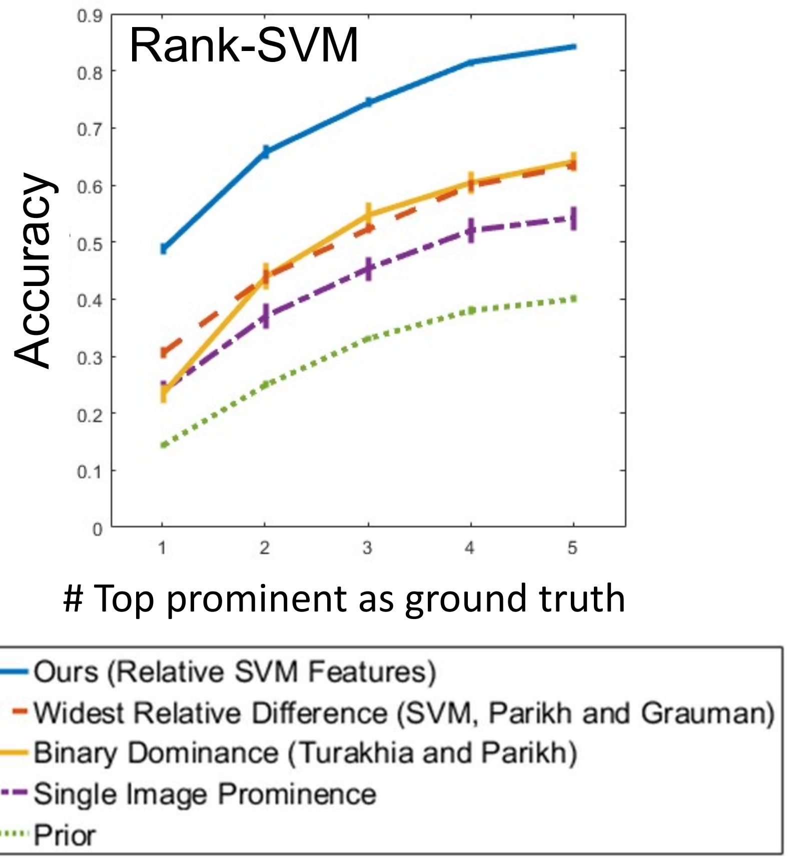


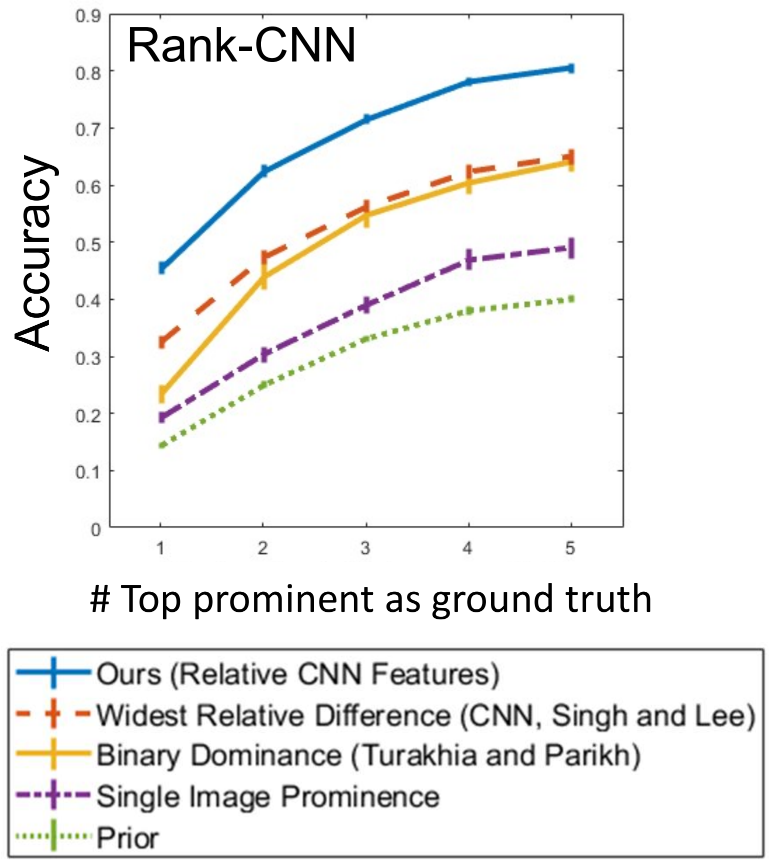
Chen & Grauman, CVPR 2018

Results: Prominent differences

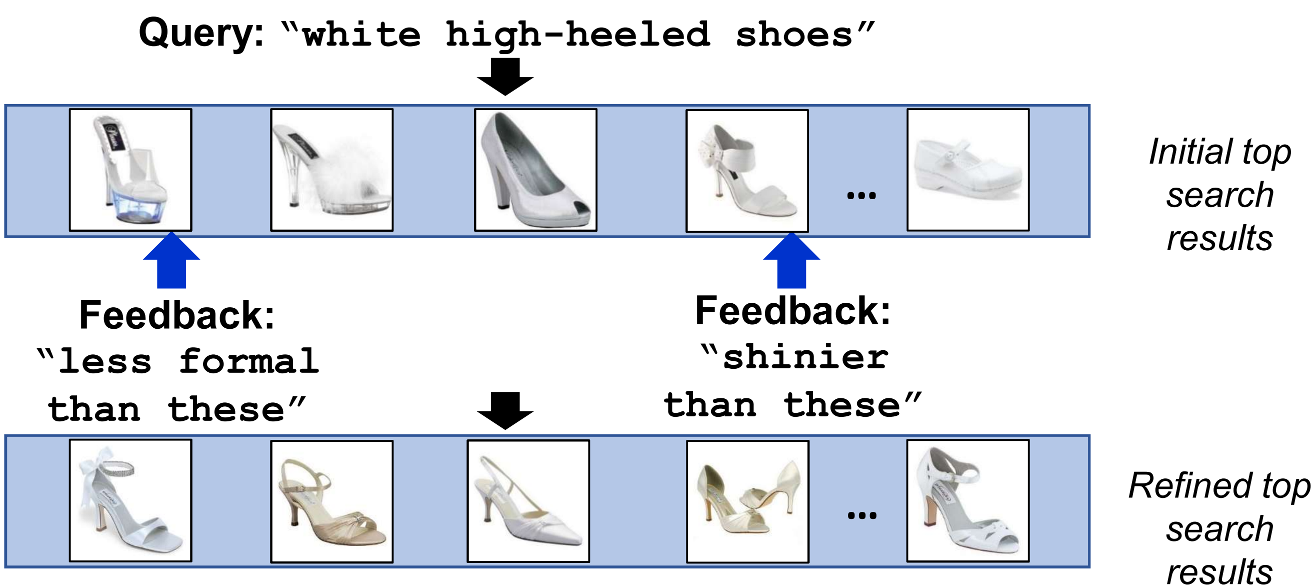
(Top 3 prominent differences for each pair)

Results: Prominent differences





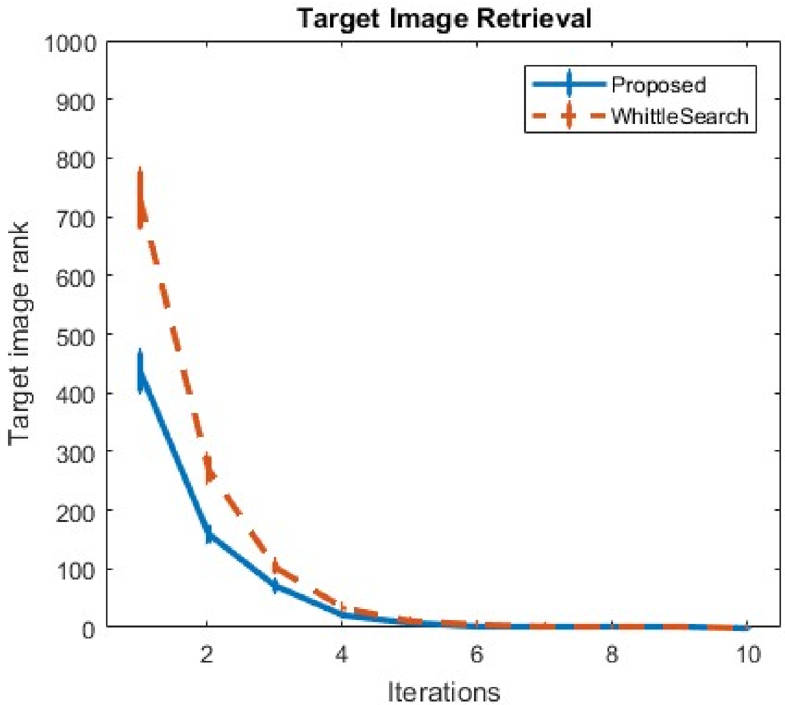
Prominent differences: impact on visual search



Leverage prominence to better focus search results

Chen & Grauman, CVPR 2018

Prominent differences: impact on visual search



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

Leverage prominence to better focus search results

Chen & Grauman, CVPR 2018

From items to styles

The language of visual attributes

- Attributes as operators

  Attributes:adjectives that *modify* objects:nouns

- Attributes for comparisons

  Relative differences that people first describe

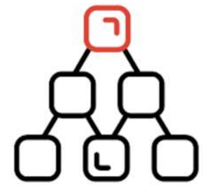
- Attributes for visual styles

  Semantic topic models for data-driven styles

How to represent visual *style*?

CNN image similarity

stylistic similarity?



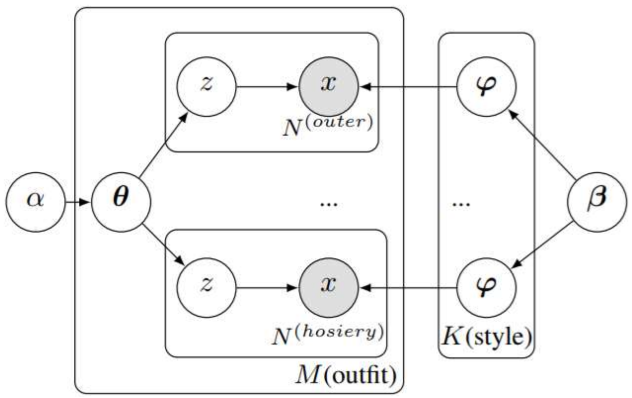
manually defined style labels

Challenges:

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

Idea: Discovering visual styles

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



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)

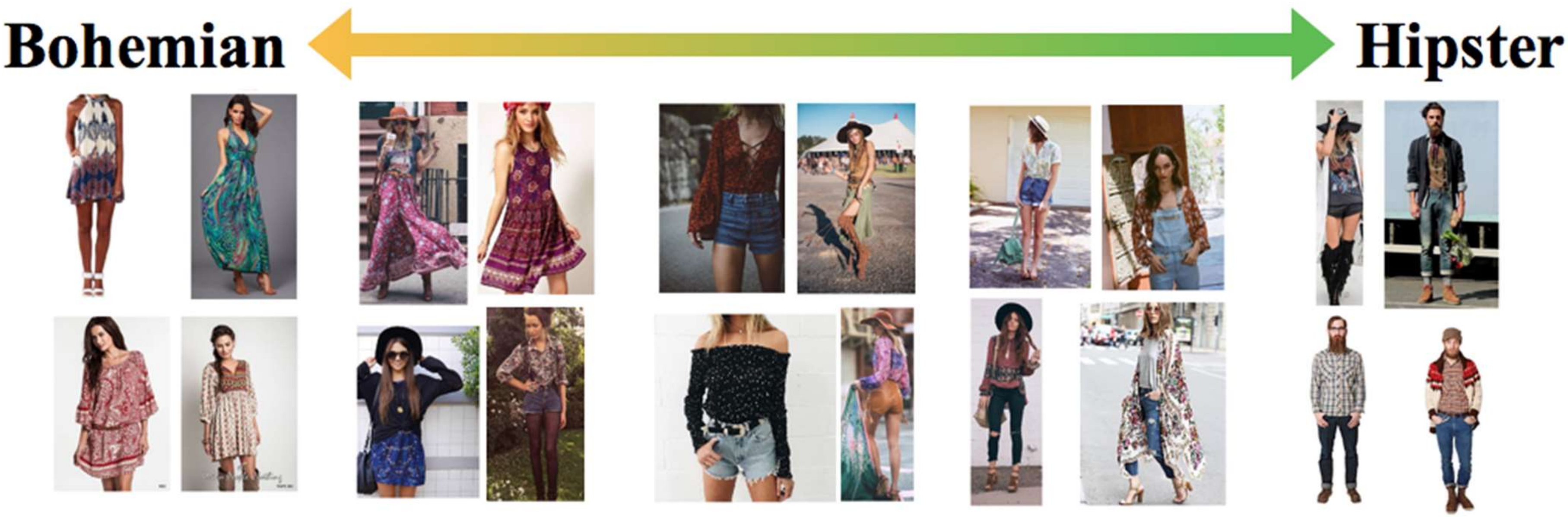
Styles we automatically discover in the **Amazon** dataset [McAuley et al. 2015]

Example discovered styles (full outfit)

Styles automatically discovered in the **HipsterWars** dataset [Kiapour et al]

Mixing styles

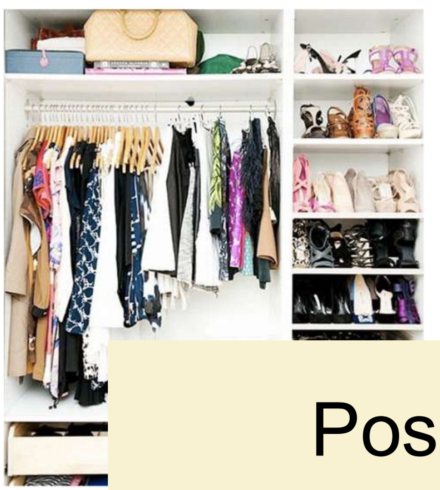
Our embedding naturally facilitates browsing for mixes of user-selected styles



Hsiao & Grauman, ICCV 2017

Creating a “capsule” wardrobe

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



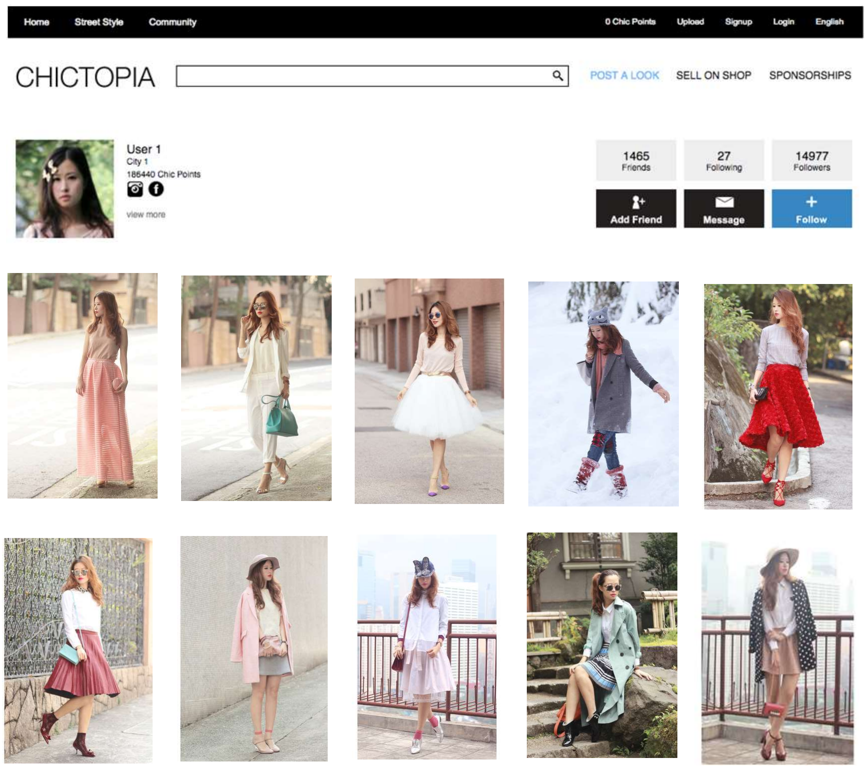
Pose as *subset selection* problem

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

Hsiao & Grauman, CVPR 2018

Creating a "capsule" wardrobe

Discover user's style preferences from album

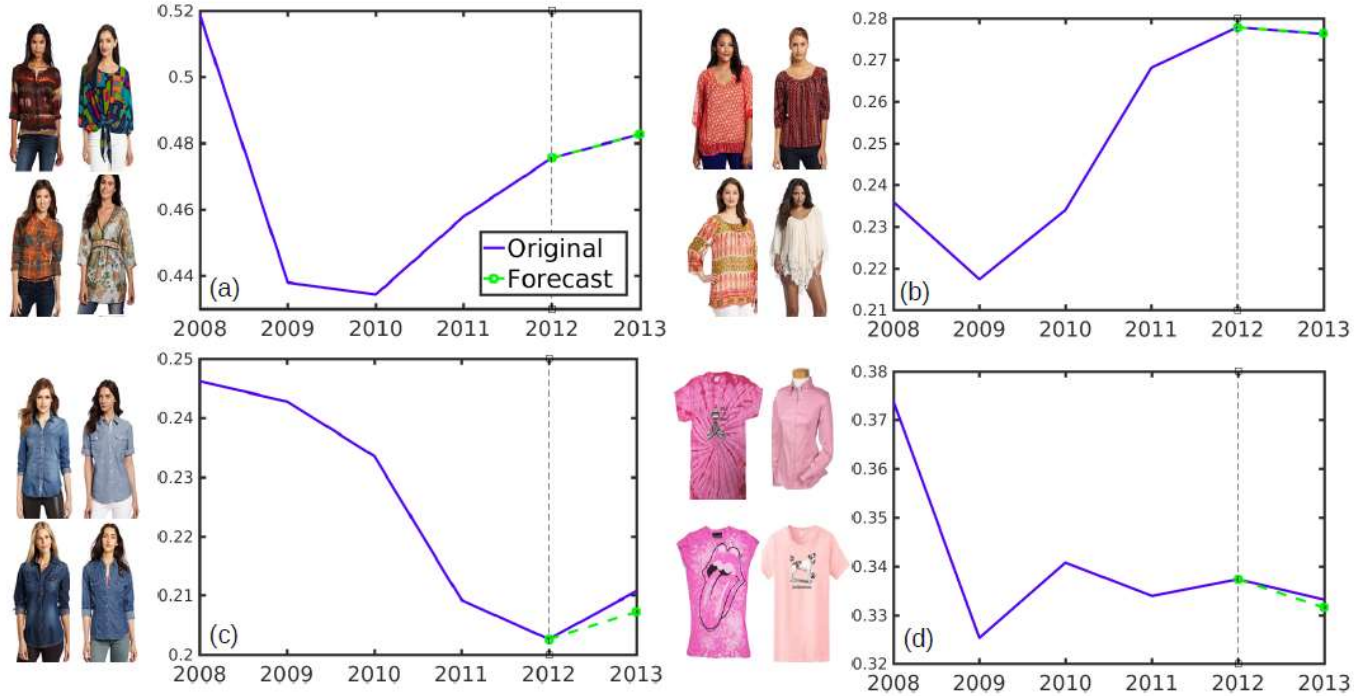


Personalized capsule

Hsiao & Grauman, CVPR 2018

Visual trend forecasting

We predict the future popularity of each style



Amazon dataset [McAuley et al. SIGIR 2015]

Al-Halah et al., ICCV 2017

Visual trend forecasting

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

(a) Texture

(b) Shape

VizWiz: Answer blind people's visual questions

[Gurari et al. CVPR 2018] Spotlight/Poster Wednesday

Is my monitor on?

Hi there can you please tell me what flavor this is?

- **Goal-oriented** visual questions
- **Conversational** language
- **Assistive** technology

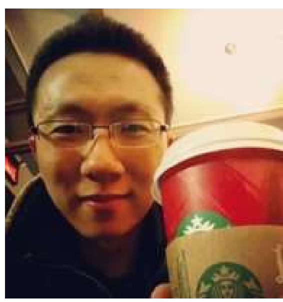
Summary: the language of visual attributes

New ideas for attributes as operators, comparisons, style basis

Applications for visual search and fashion image analysis

Tushar Nagarajan

Aron Yu



Kimberly Hsiao

Spotlight/Poster
Thursday

Ziad Al-Halah

Steven Chen

Poster
Tuesday

Papers/code

- **Attributes as Operators**. T. Nagarajan and K. Grauman. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, Sept 2018. [pdf] [supp] [code]
- **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. [pdf] [supp] [poster]
- **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. [pdf] [supp] [project page]
- **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. [pdf] [supp] [project page]
- **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. [pdf] [supp] [project page/code]
- **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. (Spotlight) [pdf]
- **VizWiz Grand Challenge: Answering Visual Questions from Blind People.** D. Gurari, Q. Li, A. Stangl, A. Guo, C. Lin, K. Grauman, J. Luo, and J. Bigham. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, June 2018. (Spotlight) [pdf]