Learning attributes from human gaze

Nils Murrugarra-Llerena
Adriana Kovashka

Department of Computer Science
University of Pittsburgh

IEEE 2017 Winter Conference on Applications of Computer Vision

Background

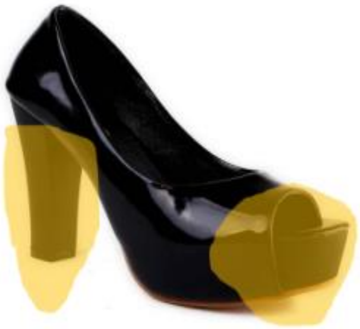
Introduction

Many **attributes** possess different **interpretations** as opposed to **objects**.

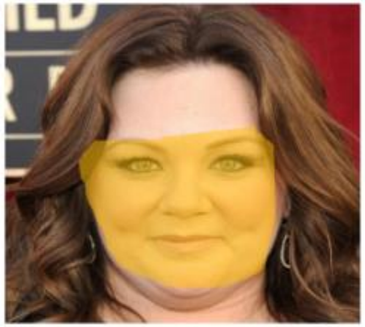
- ***boot***: most of the drawings will be **similar**
- ***formal*** or ***open shoe***: many drawings will be **different**

Motivation

- Why not integrate humans **more closely** in attribute learning?
  - using **human gaze maps**.



Q: Is it **pointy**?



Q: Is she **chubby**?

Background/Approach

Uniqueness

- Fast
- Orthogonal to DNN approaches
- Subconscious + Humans’ intuition

Data Collection

- GazePoint GP3 eye-tracker.
- 4 sub-sessions.
- **Datasets:** shoes, faces and scenes.
- Screening phase.
- Validation images.

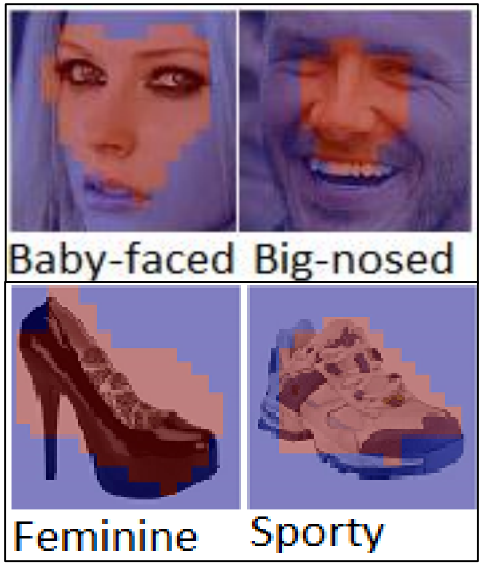
Generate gaze templates

- **ST:** Merge gaze maps from positive annotations – normalize [0, 1] – threshold with 0.1 – mask selected cell from a 15x15 grid.
- **MT:** It captures different attribute meanings using **clustering**.

Approach

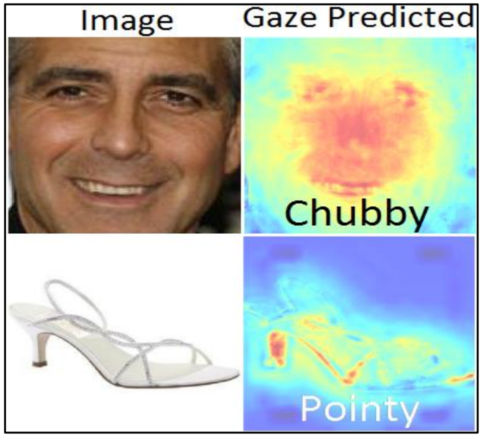
Attribute learning using fixed gaze templates

- **ST:** Mask train/test images – extract features – evaluate a classifier.
- **MT:** Similar to ST.
  - Train an individual classifier per cluster.
  - Predict a novel image as positive if at least one of the classifiers forecasts it contains the attribute.



Attribute learning using gaze prediction

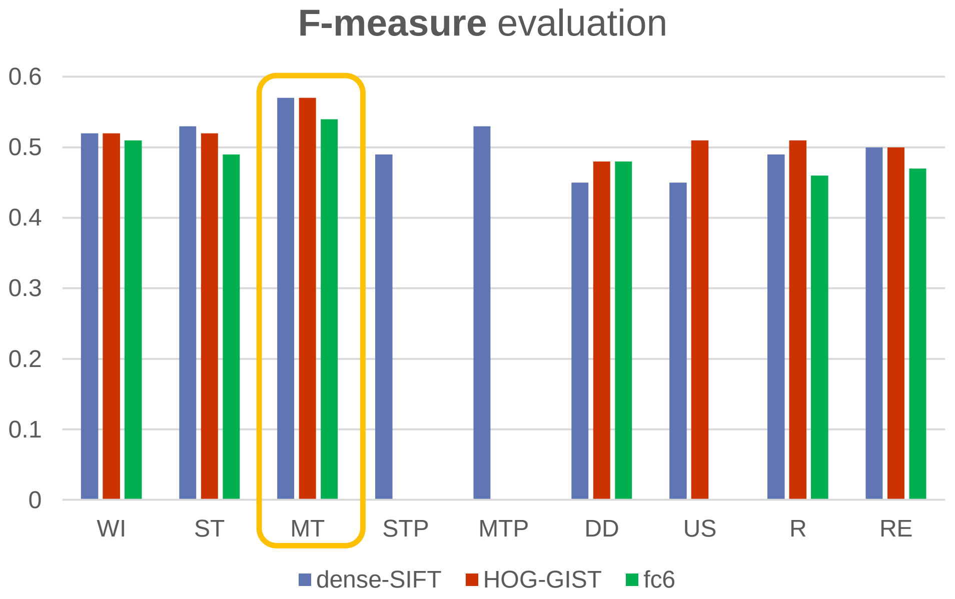
- Instead of using a fixed template - Learn a gaze predictor – predict gaze maps for novel images – **STP/MTP**



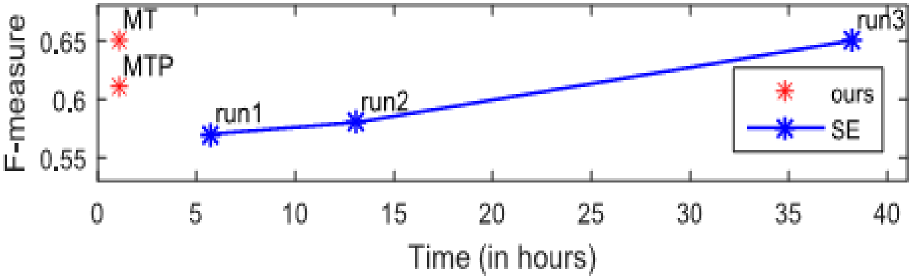
Evaluation

Baselines

- ***Whole Image (WI)***, which extracts features from the whole image without a mask.
- ***Data-Driven (DD)***, which uses a binary mask created from an L1-regularizer over features extracted on a grid.
- ***Unsupervised Saliency (US)***, which uses a binary mask from a state-of-the-art saliency predictor **(Huang et al, ICCV 2015)**.
- ***Random grid (R)***, which employs a random binary mask from a 15x15 grid.
- ***Random Ensemble grid (RE)***, which creates an ensemble of *R*.



Evaluation



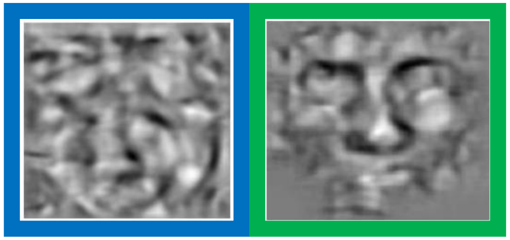
Comparison with **Spatial Extent (SE)** method **(Xiao and Lee, ICCV 2015)**

Adaptation to scenes attributes

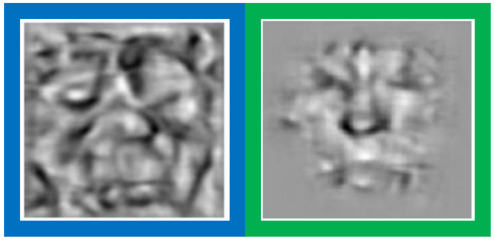
Attribute	Objects	Attribute	Objects
climbing	mountain, sky, tree, trees, building	sunny	sky, tree, building, grass, trees
open area	sky, trees, grass, road, tree	driving	sky, road, tree, trees, building

Applications

Visualizing attribute models



Baby-faced attribute



Big-nosed attribute

Whole Image
Gaze Template

Finding schools of thought

We improve schools of thought using gaze.

Original	Gaze-based
0.37	**0.40**

F-measure

Further discussion

See you at **poster #6**

References

1. X. Huang, C. Shen, X. Boix, and Q. Zhao. Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks. In ICCV, 2015
2. F. Xiao and Y. J. Lee. Discovering the spatial extent of relative attributes. In ICCV, 2015.