

Formulating Structure for Vision Problems

Ziwei Liu

Department of Information Engineering
The Chinese University of Hong Kong

Appetizer

- A mathematician is a person who can find **analogies between theorems**.
- A better mathematician is one who can see **analogies between proofs**.
- And the best mathematician can notice **analogies between theories**.
- One can imagine that the ultimate mathematician is one who can see **analogies between analogies**.

Outline

- **Input Structure**: patch, image, video, multi-modality ...
- **Model Structure**: information flow + regularization
- **Target Structure**: label, sequence, mask, multi-task ...

Content

Lessons Learned (**Geometry + Semantics**)

Random Thoughts (**Inverse Thinking**)

No Equations (**Fast Forward**)

Input Structure

- Case Study I — low-level vision tasks



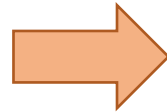
Saturating
Performance

Input Structure

- Case Study I — low-level vision tasks



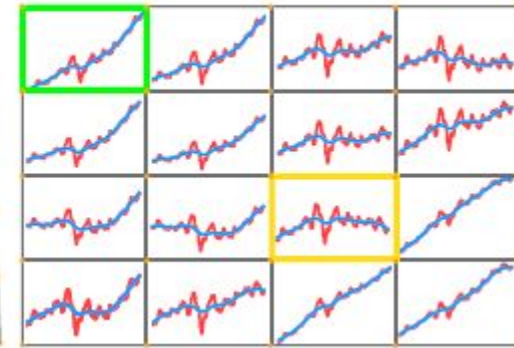
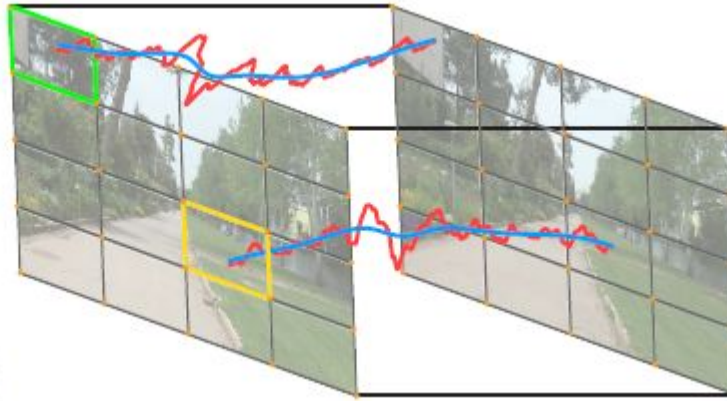
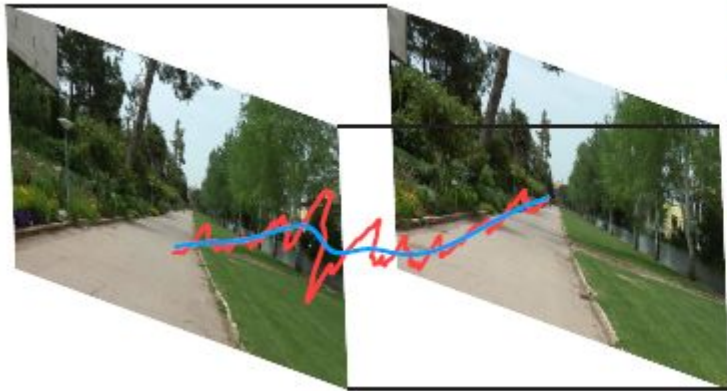
noisy burst images



a clean image

Input Structure

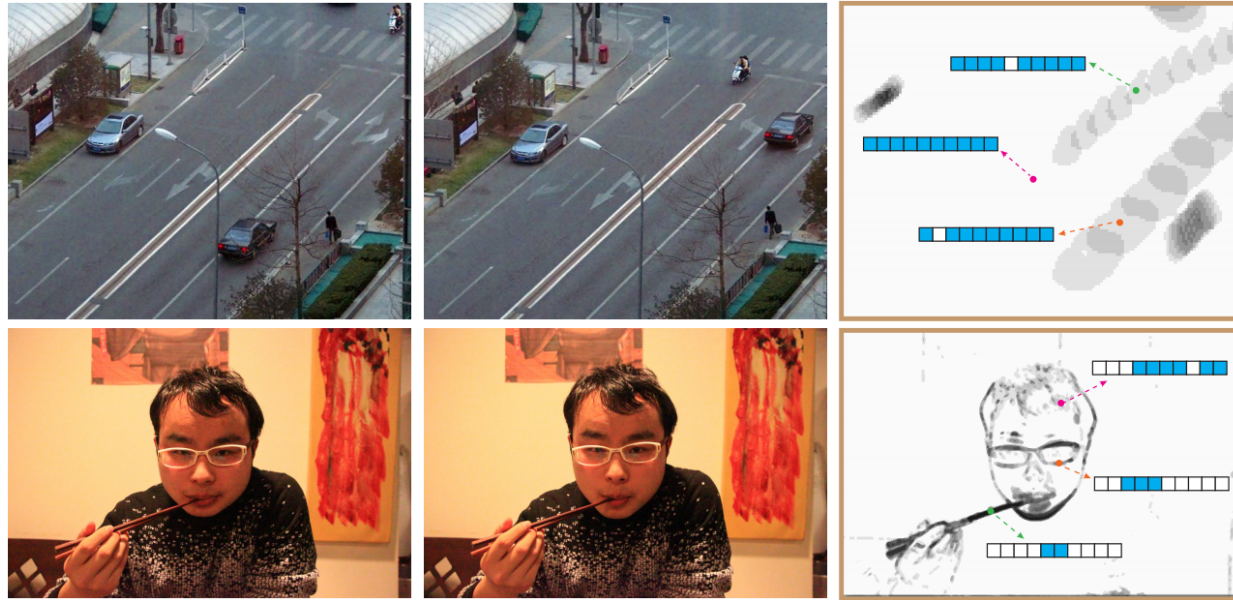
- Case Study I — low-level vision tasks



Data Alignment I:
Geometry

Input Structure

- Case Study I ——— low-level vision tasks



Data Alignment II:
Semantics

Input Structure

- Case Study I — low-level vision tasks



High Dynamic Range

Original Photo #1



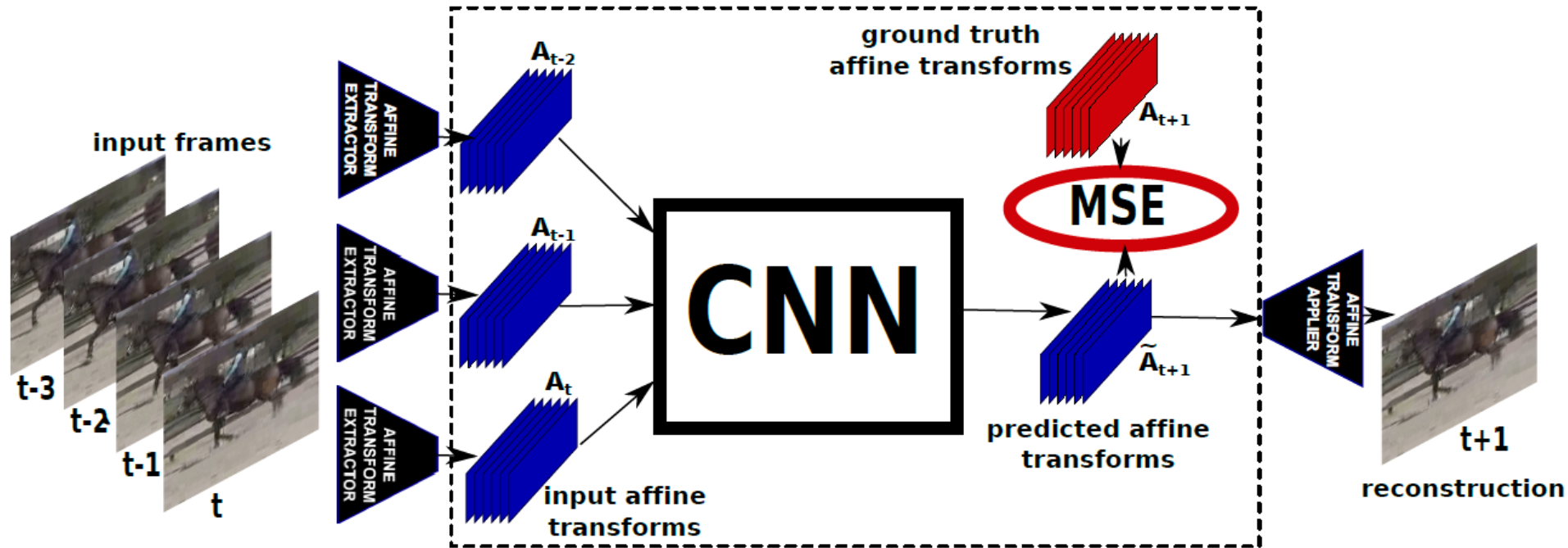
Original Photo #2



Auto Smiling

Input Structure

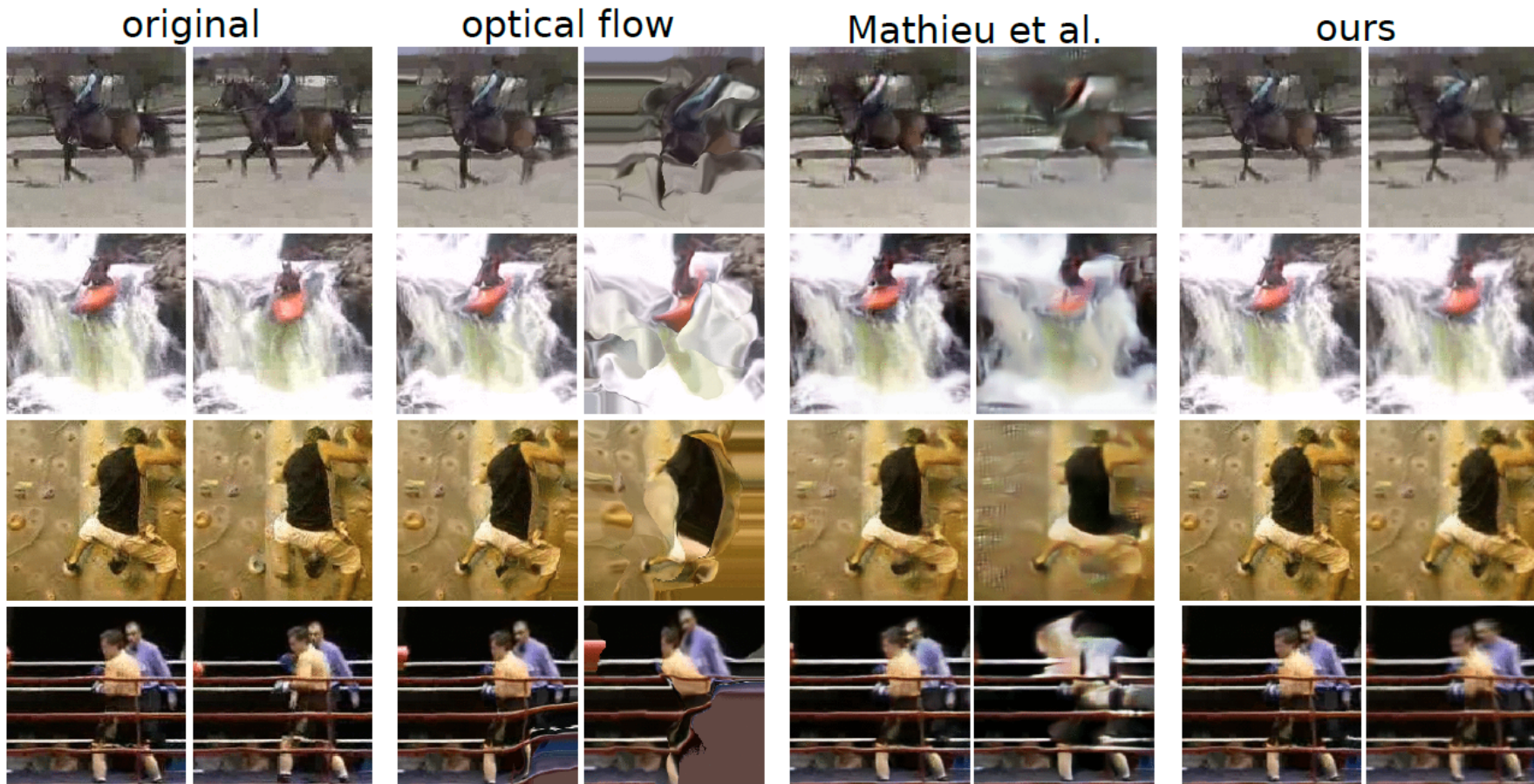
- Case Study I — low-level vision tasks



Inverse Thinking

Input Structure

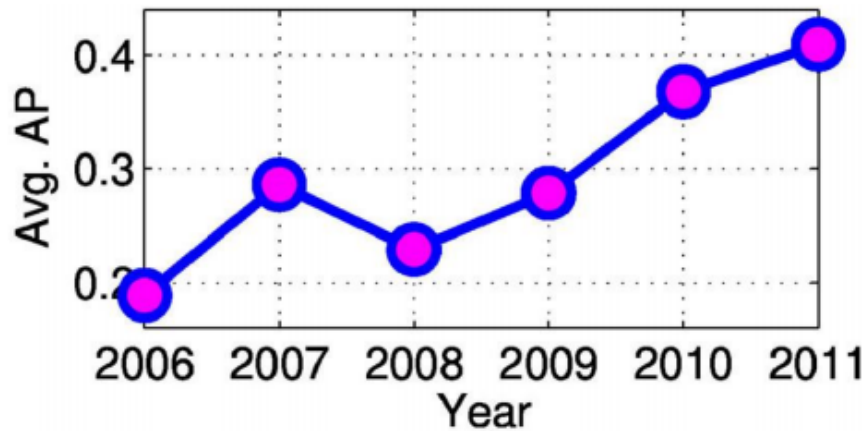
- Case Study I — low-level vision tasks



Graphics
+
Vision

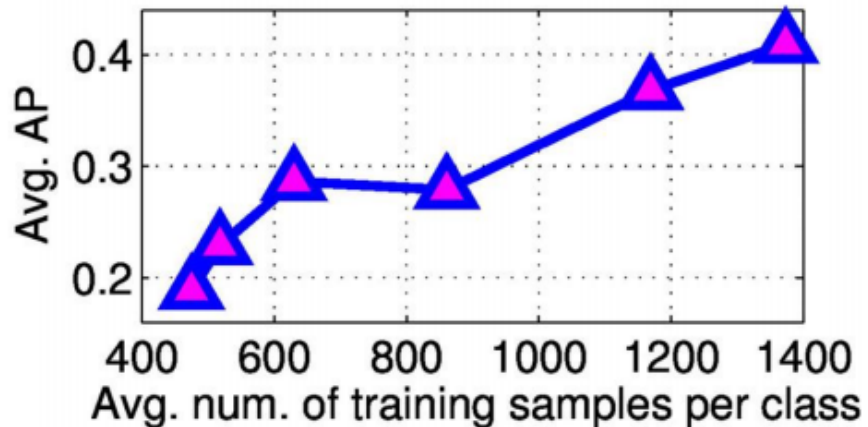
Input Structure

- Case Study II — high-level vision tasks



Data improvement?
Model improvement?

More data

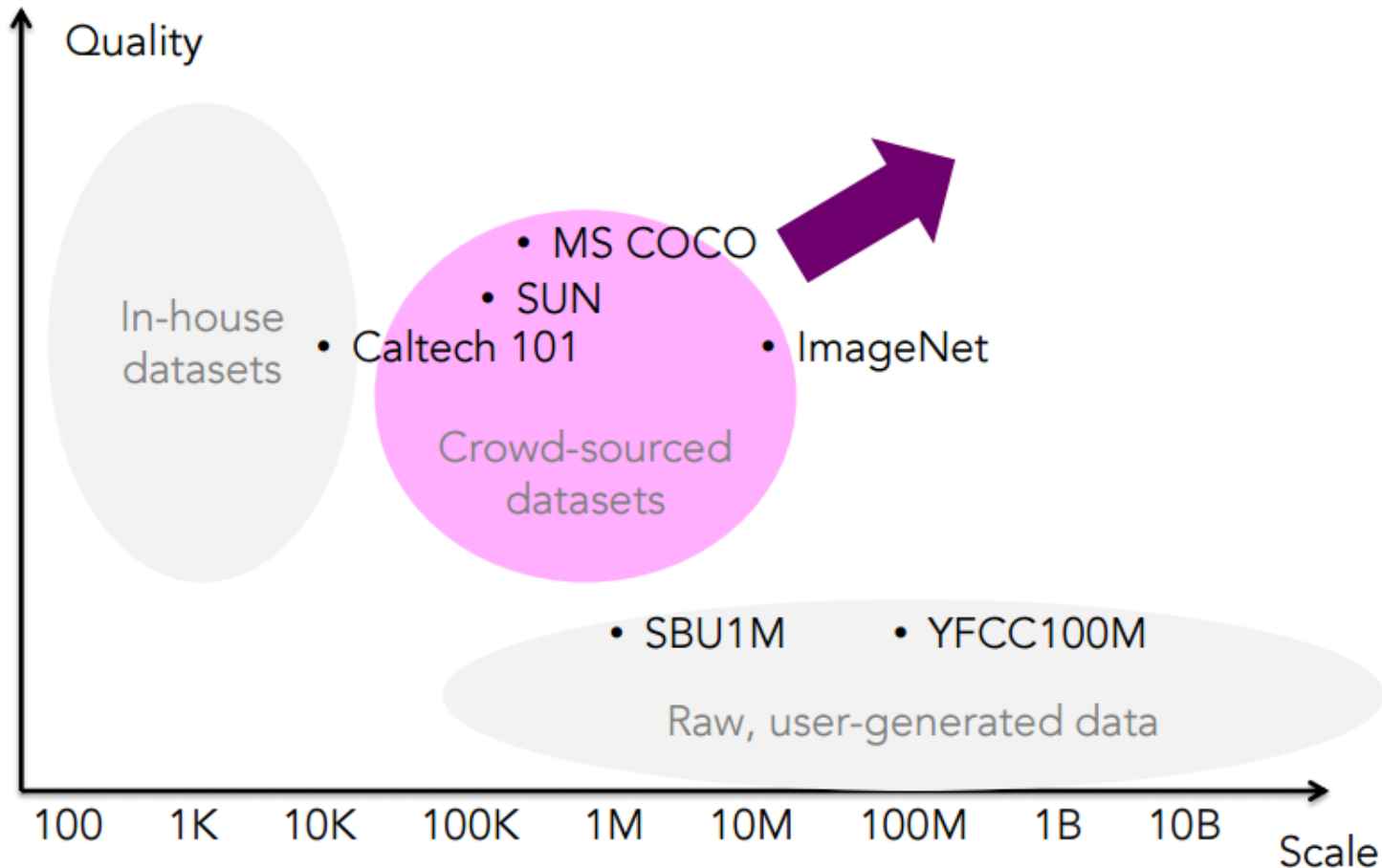


*X Zhu et al. Do We Need
More Training Data? IJCV
2015*

We need both

Input Structure

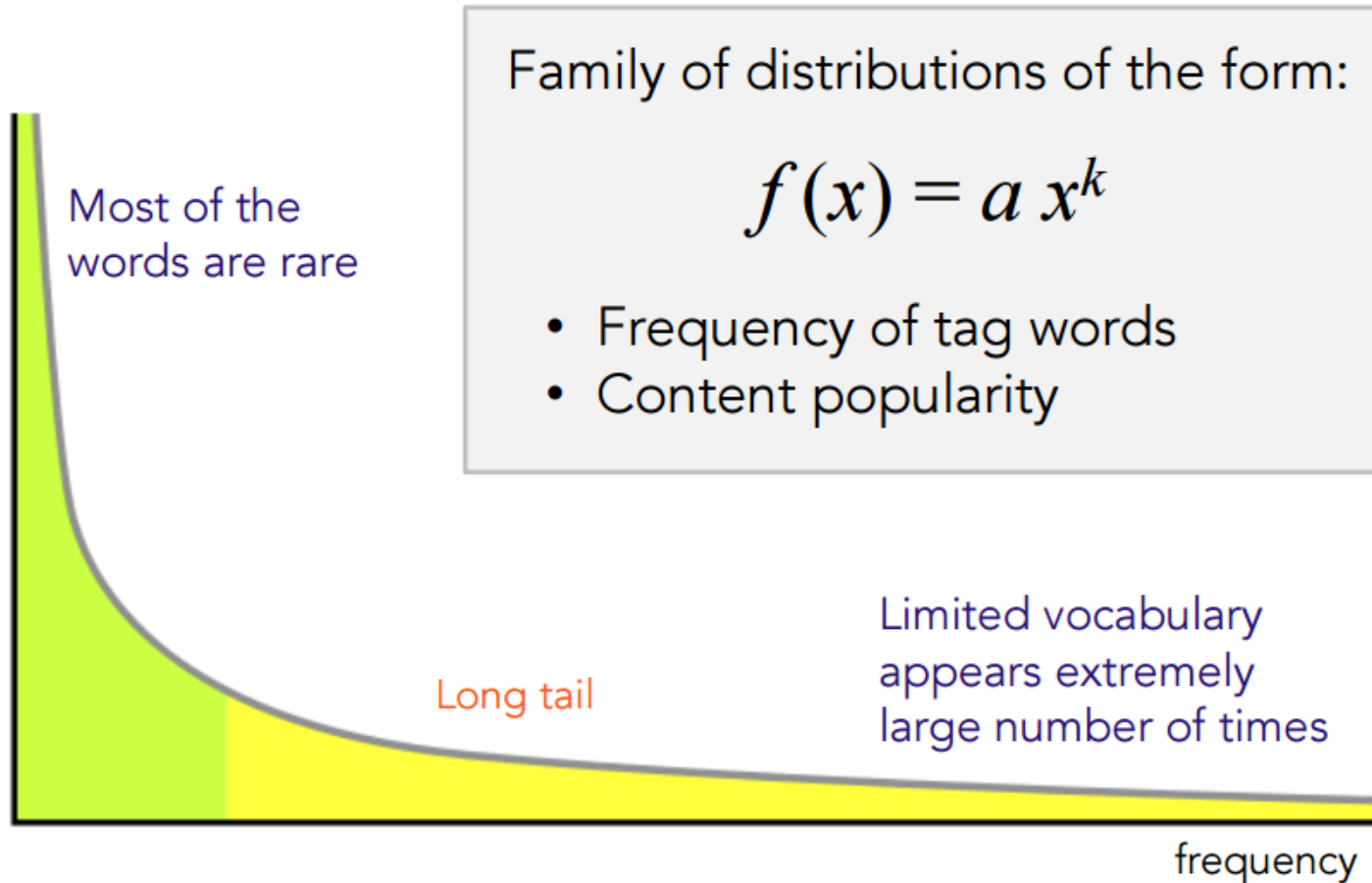
- Case Study II — high-level vision tasks



Quality v.s. scale

Input Structure

- Case Study II — high-level vision tasks



Power laws

Input Structure

- Case Study II — high-level vision tasks
 - User-generated content does not contain clean data
 - Non-visual texts / tags
 - Tags tend to have high precision, low recall
 - Frequency issue
- Hopefully, large data-size resolves issues

Learning from
online content

Input Structure

- Case Study II — high-level vision tasks

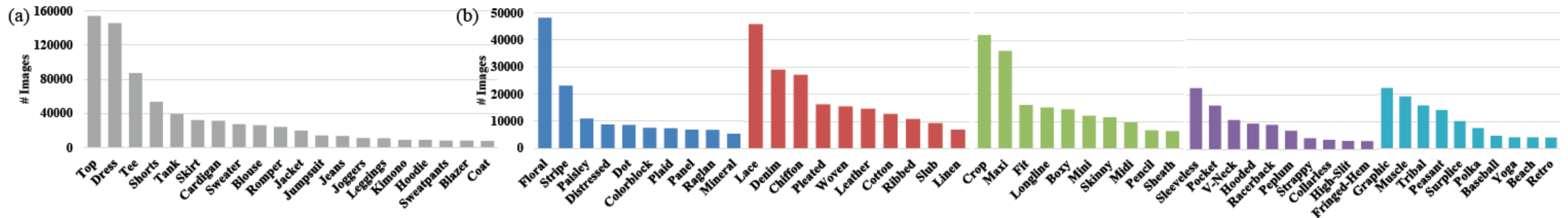


Data Alignment I:
Geometry

Data Alignment II:
Semantics

Input Structure

- Case Study II — high-level vision tasks



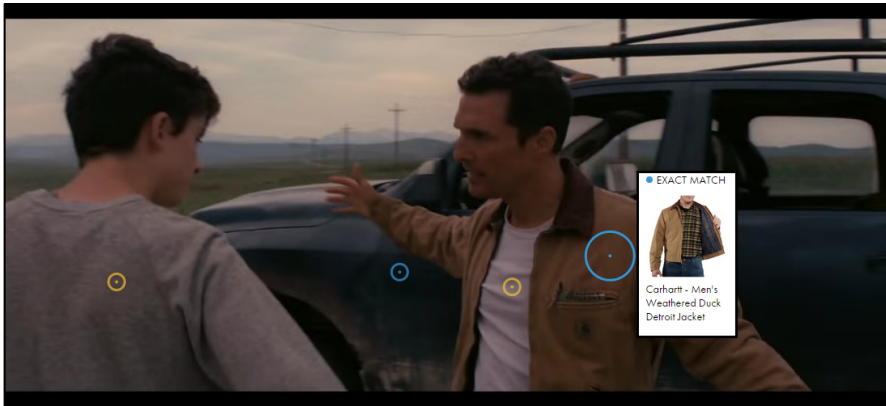
Input Structure

- Case Study II — high-level vision tasks

Has-button



Similar Style Retrieval



Cloth Spotting in Video



Street-to-shop



Fashion Assistant

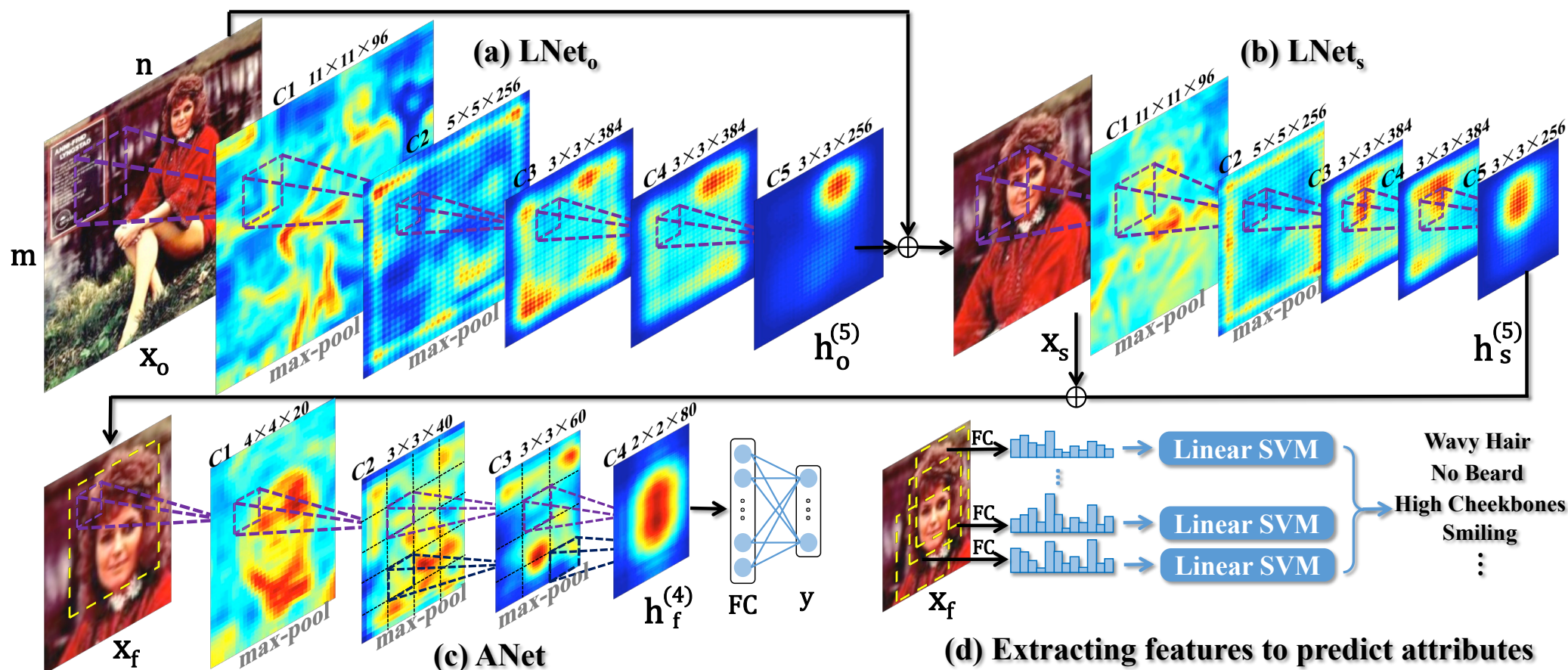
Model Structure

- Case Study — facial attributes prediction



Model Structure

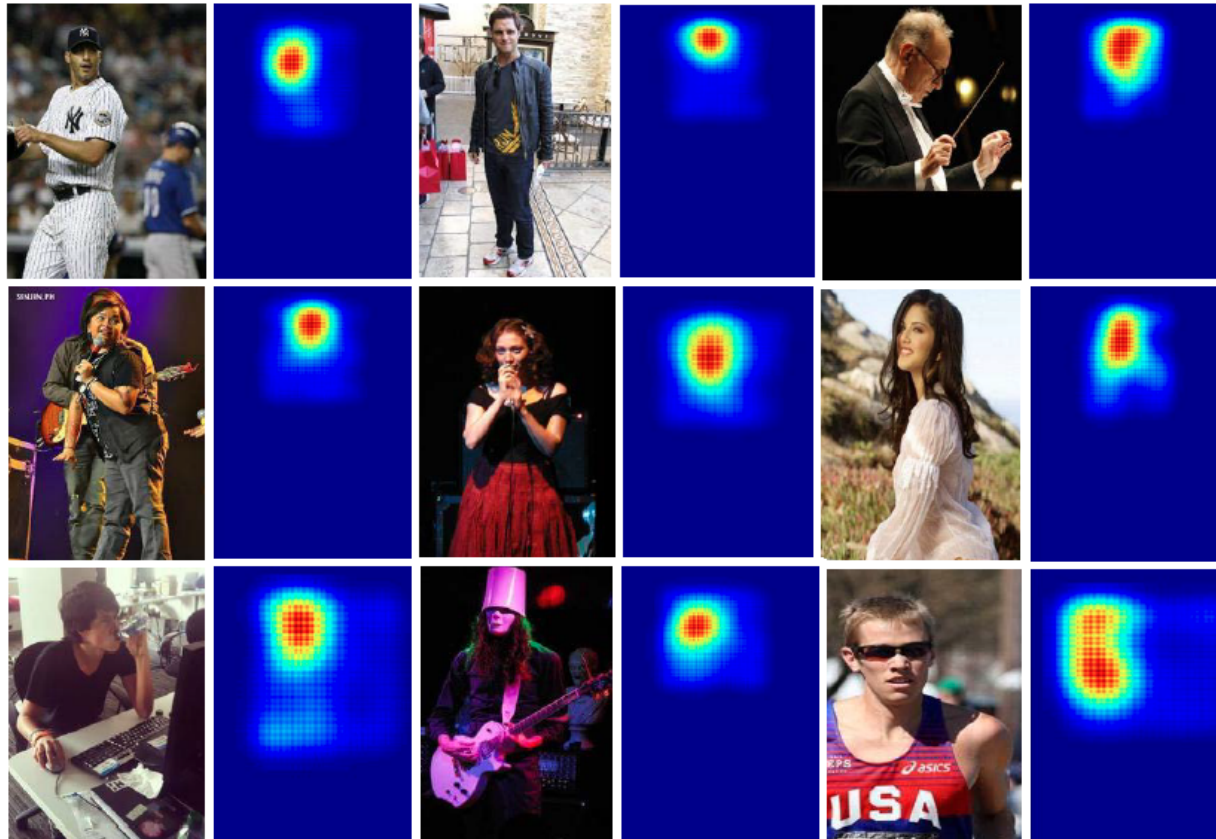
- Case Study I — facial attributes prediction



Variance
reduction

Model Structure

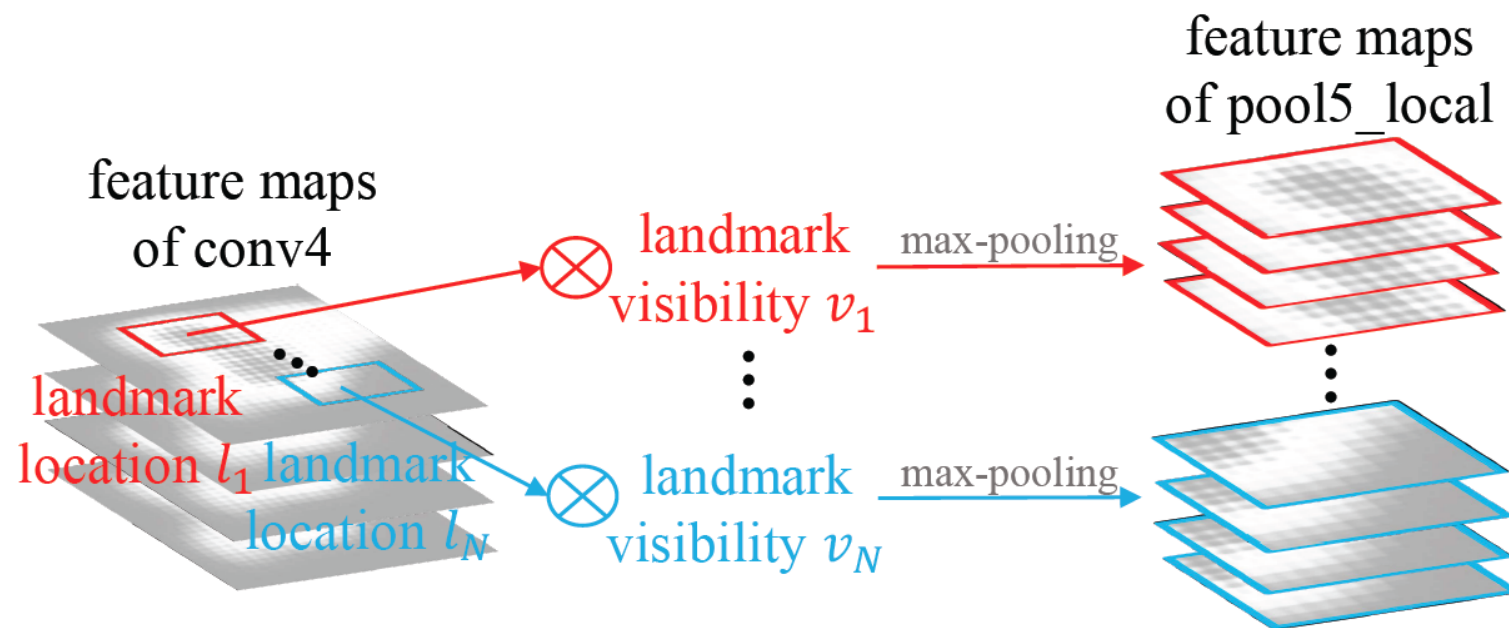
- Model Alignment I — geometry



Attention to
salient regions

Model Structure

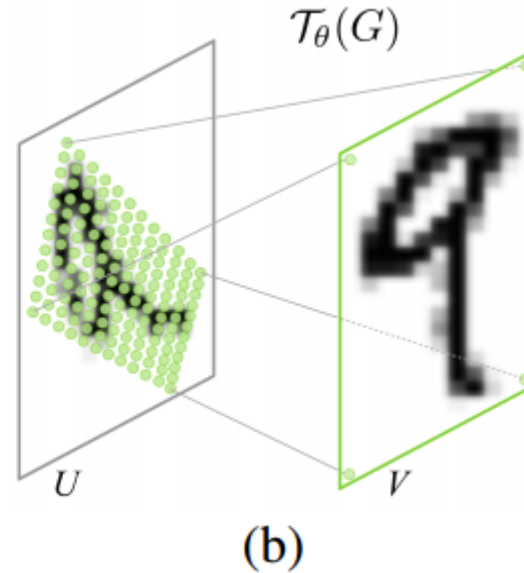
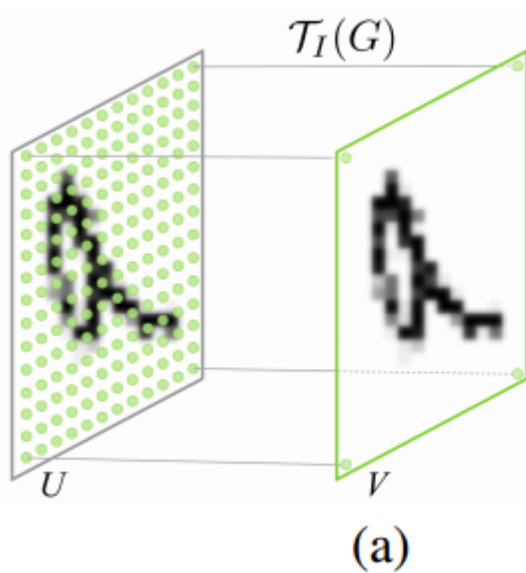
- Model Alignment I — geometry



Pool features from
salient regions

Model Structure

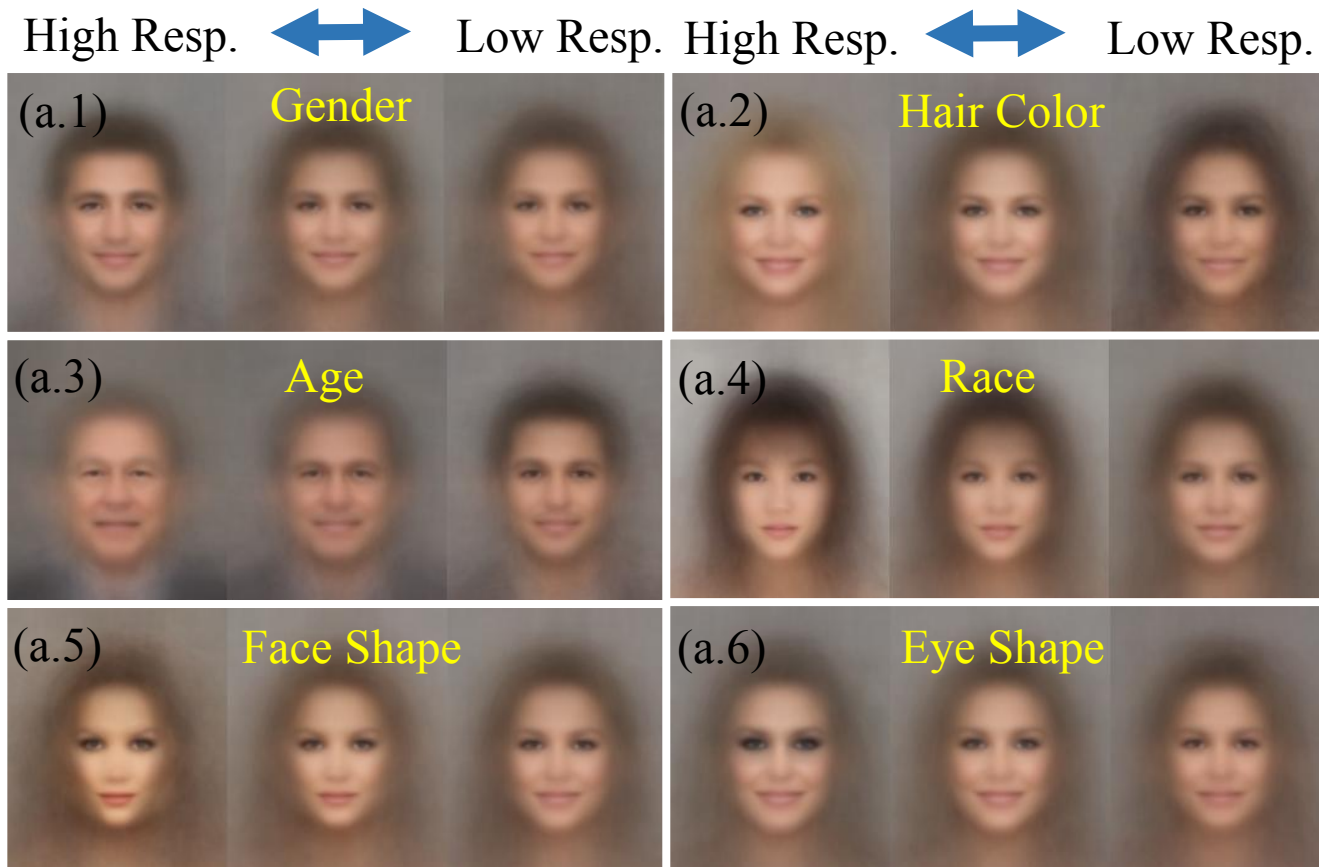
- Model Alignment I — geometry



Transform features
to canonical position

Model Structure

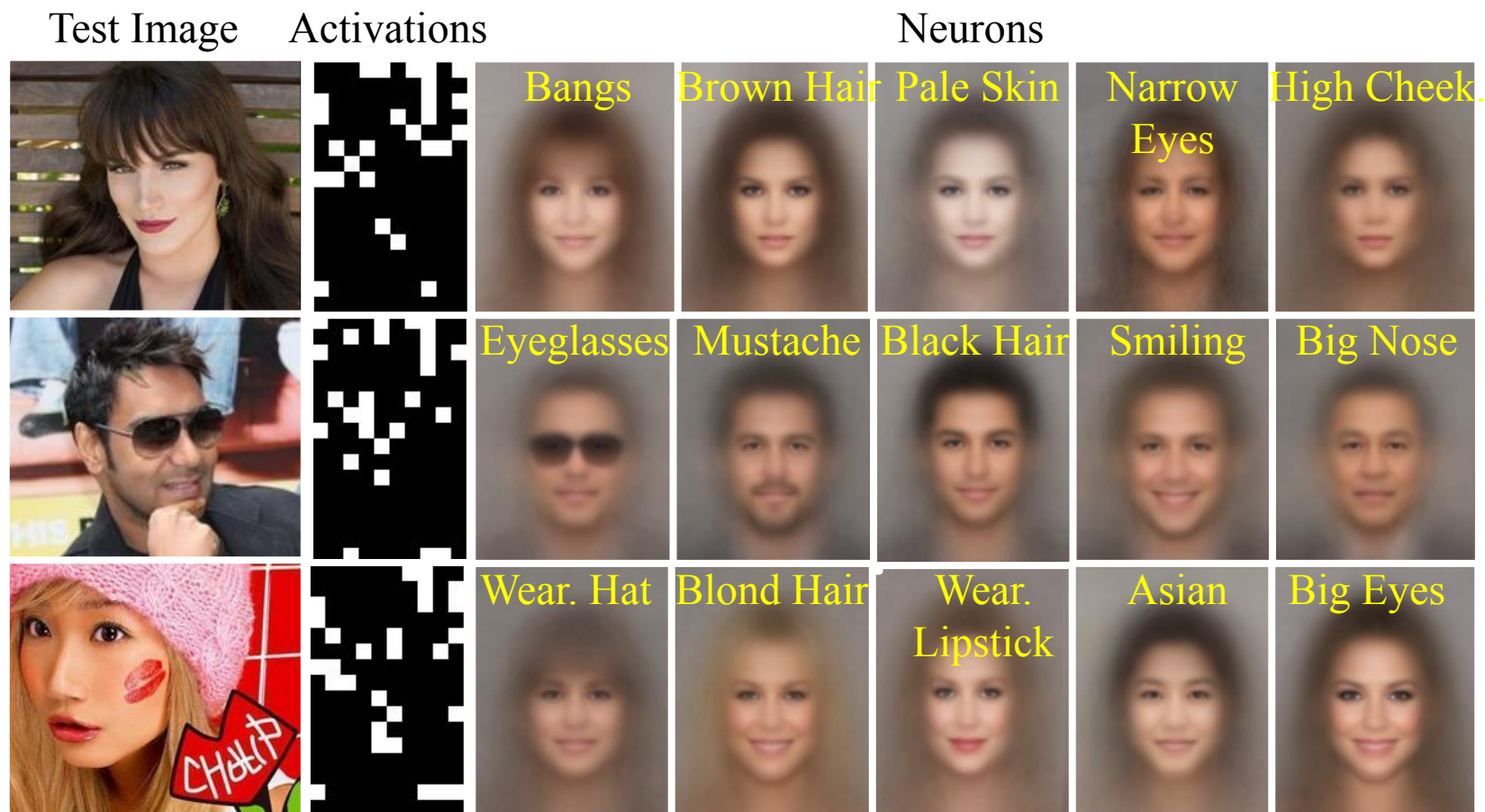
- Model Alignment II — semantics



Abstract useful
concepts

Model Structure

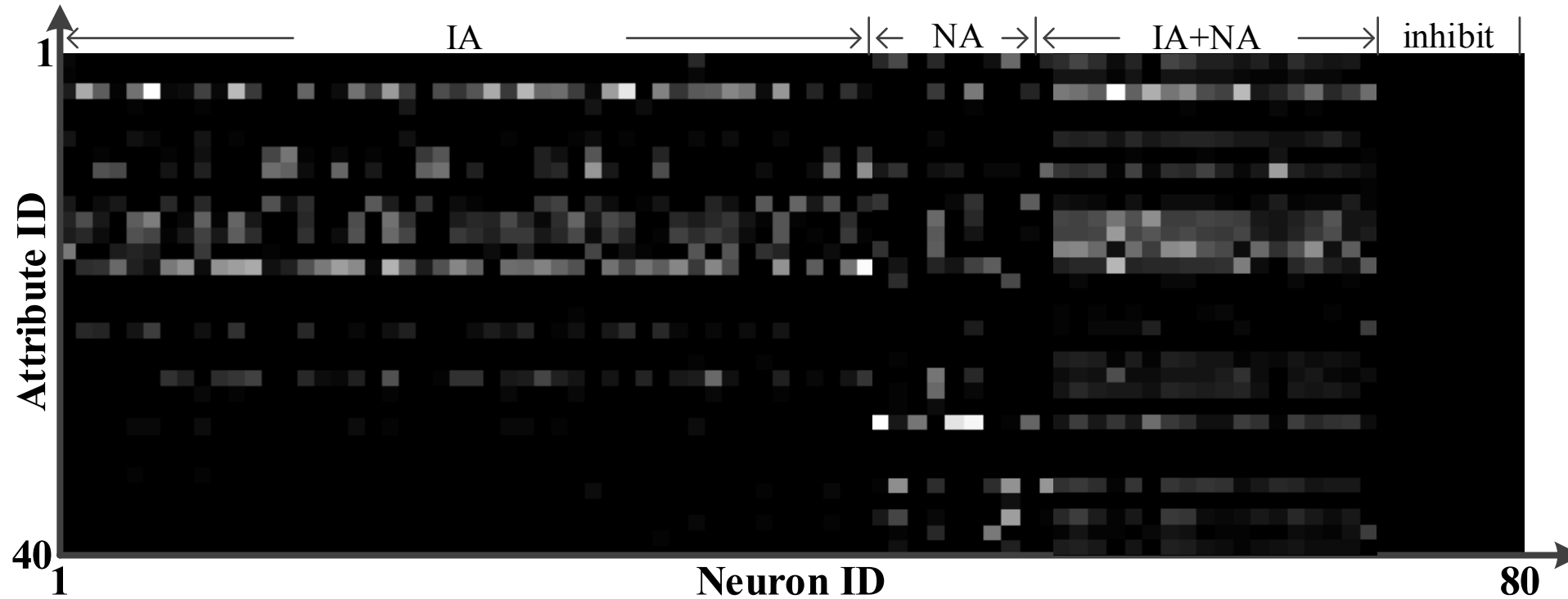
- Model Alignment II — semantics



Combine to
generalize

Model Structure

- Model Alignment II — semantics

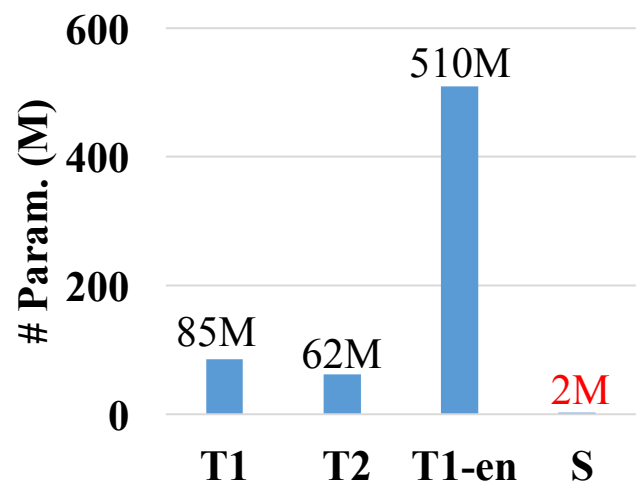


Squeeze to
compress

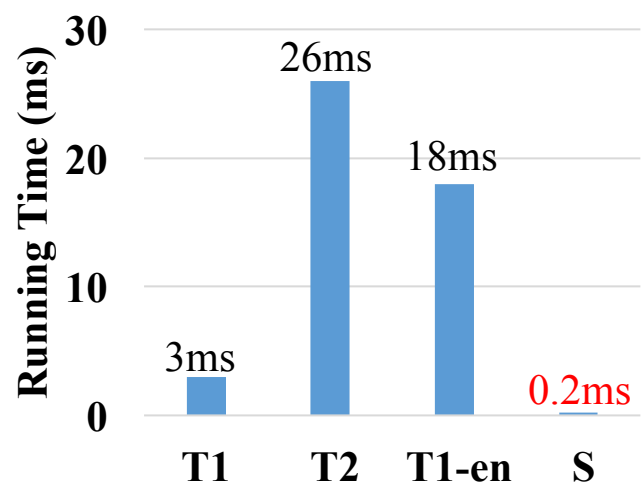
Inverse Thinking

Model Structure

- Model Alignment II — semantics



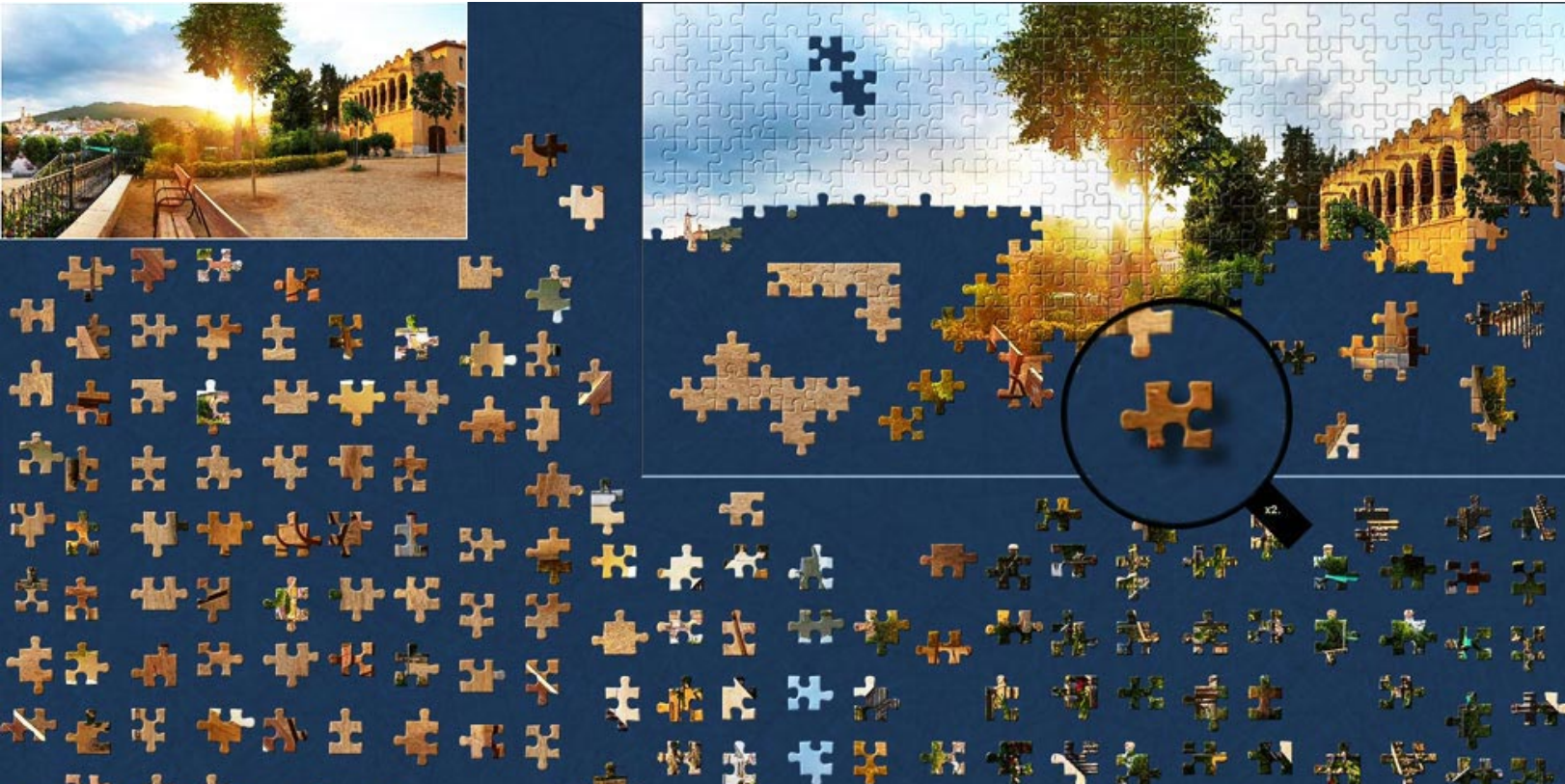
Model size comparisons



Running time comparisons

Target Structure

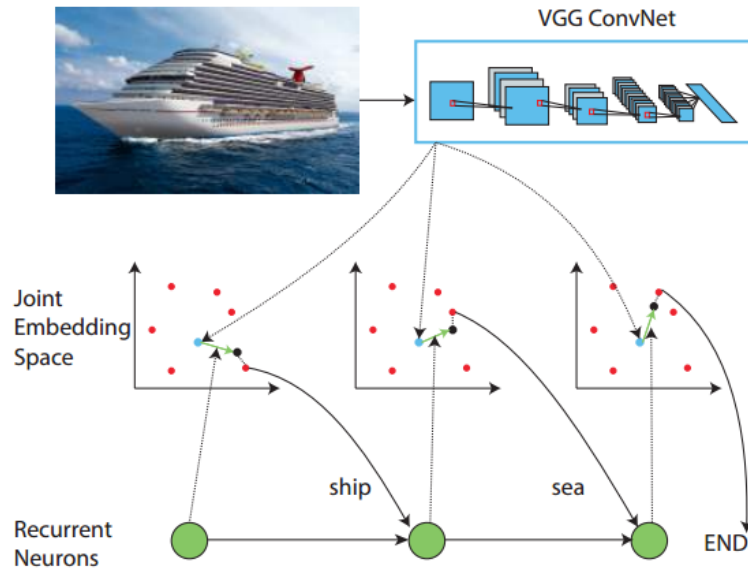
- Dependencies Among Target



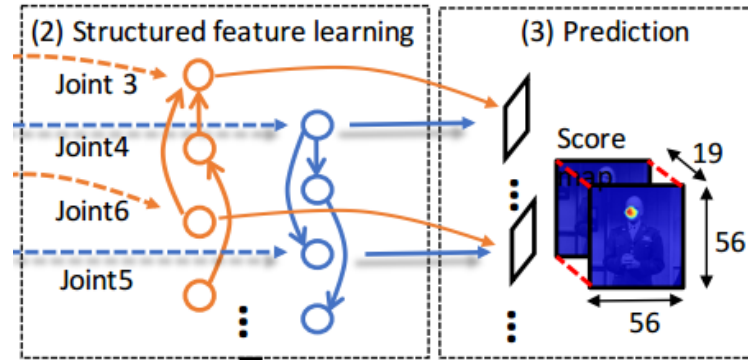
Jigsaw Puzzles

Target Structure

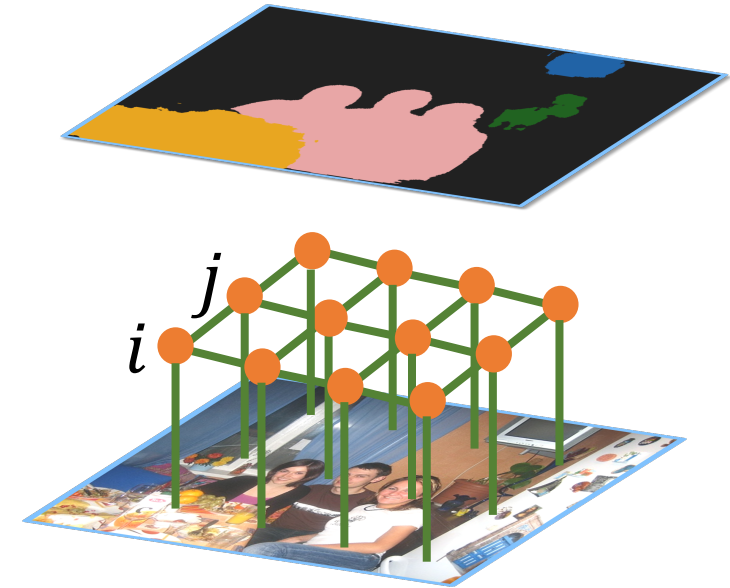
- Message Passing



Classification



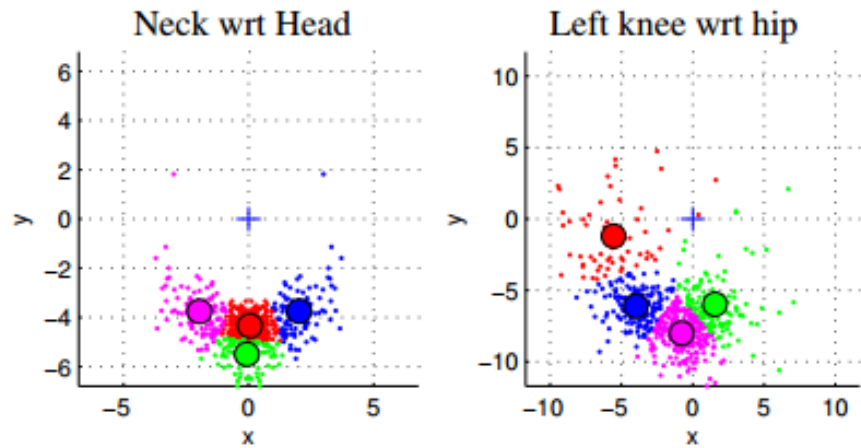
Localization



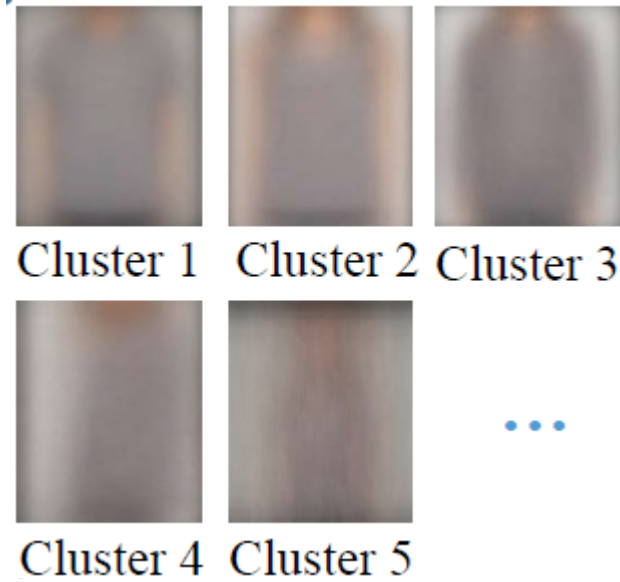
Segmentation

Target Structure

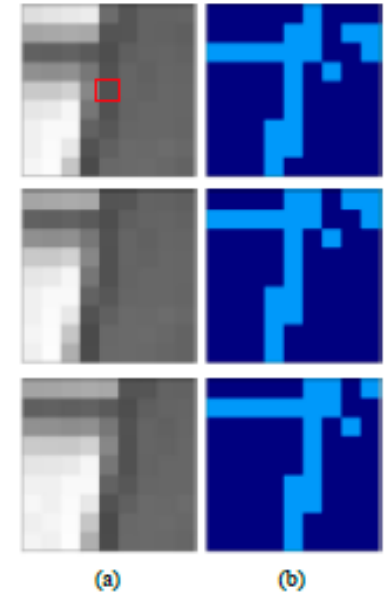
- Target Alignment I — geometry



Location



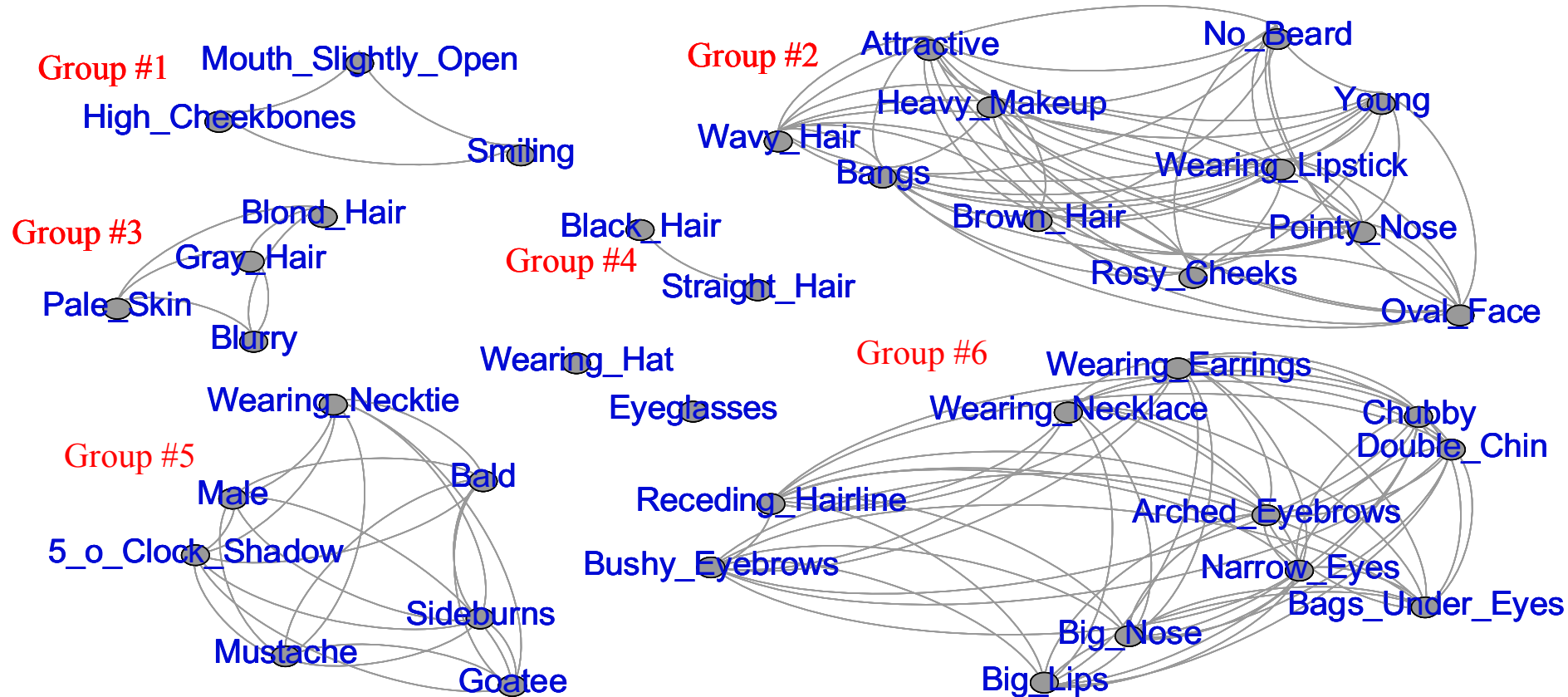
Shape



Appearance

Target Structure

- Target Alignment II — semantics



Hierarchy

Co-occurring

Exclusive

Unrelated

Target Structure

- Case Study I — semantic segmentation



Target Structure

- Case Study I — semantic segmentation

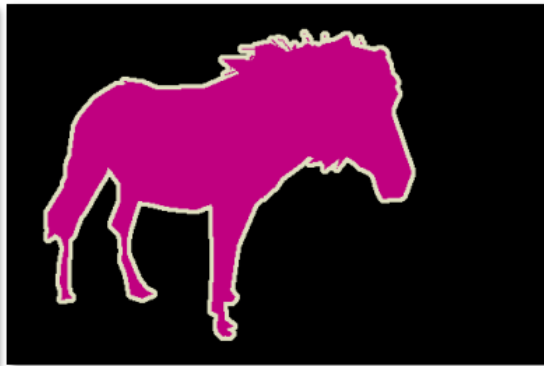


Target Structure

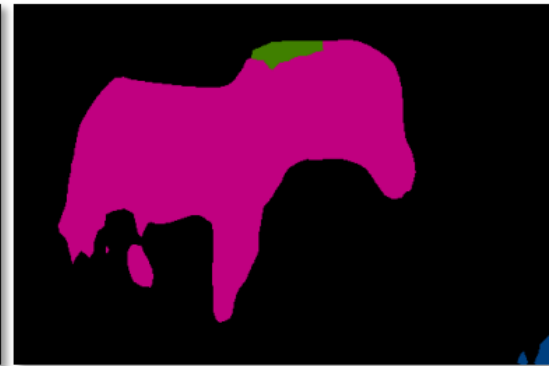
- Case Study I — semantic segmentation



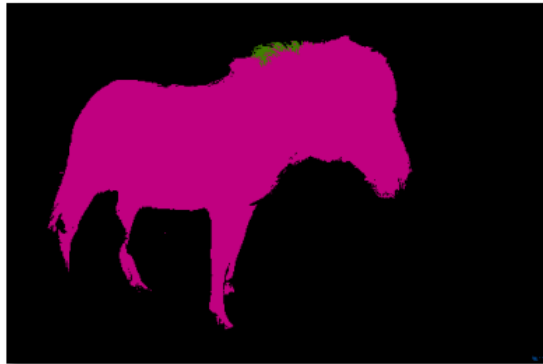
(a) Original Image



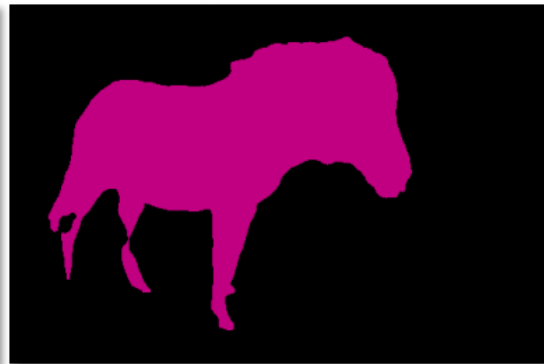
(b) Ground Truth



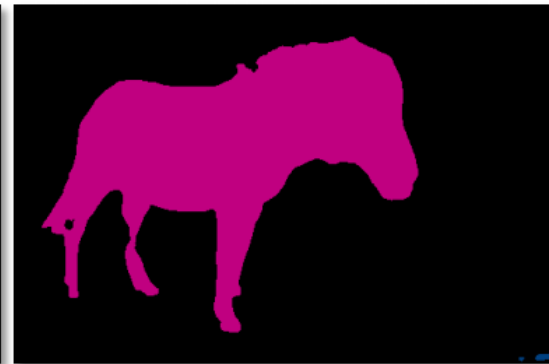
(c) Unary Term



(d) +Triple Penalty



(e) +Label Contexts

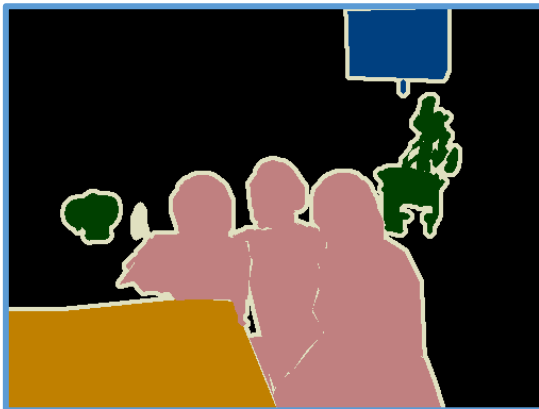


(f) +Joint Tuning

Target Structure



Original Image



Ground Truth



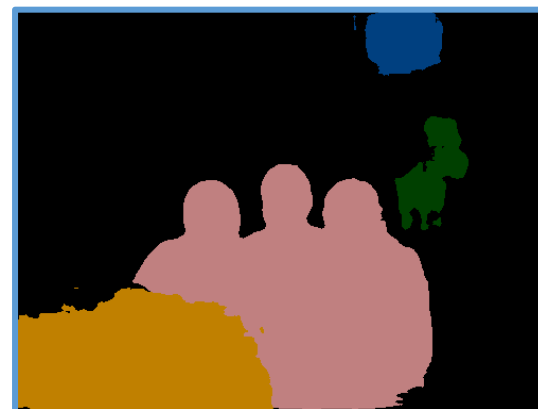
Unary Term



Triple Penalty

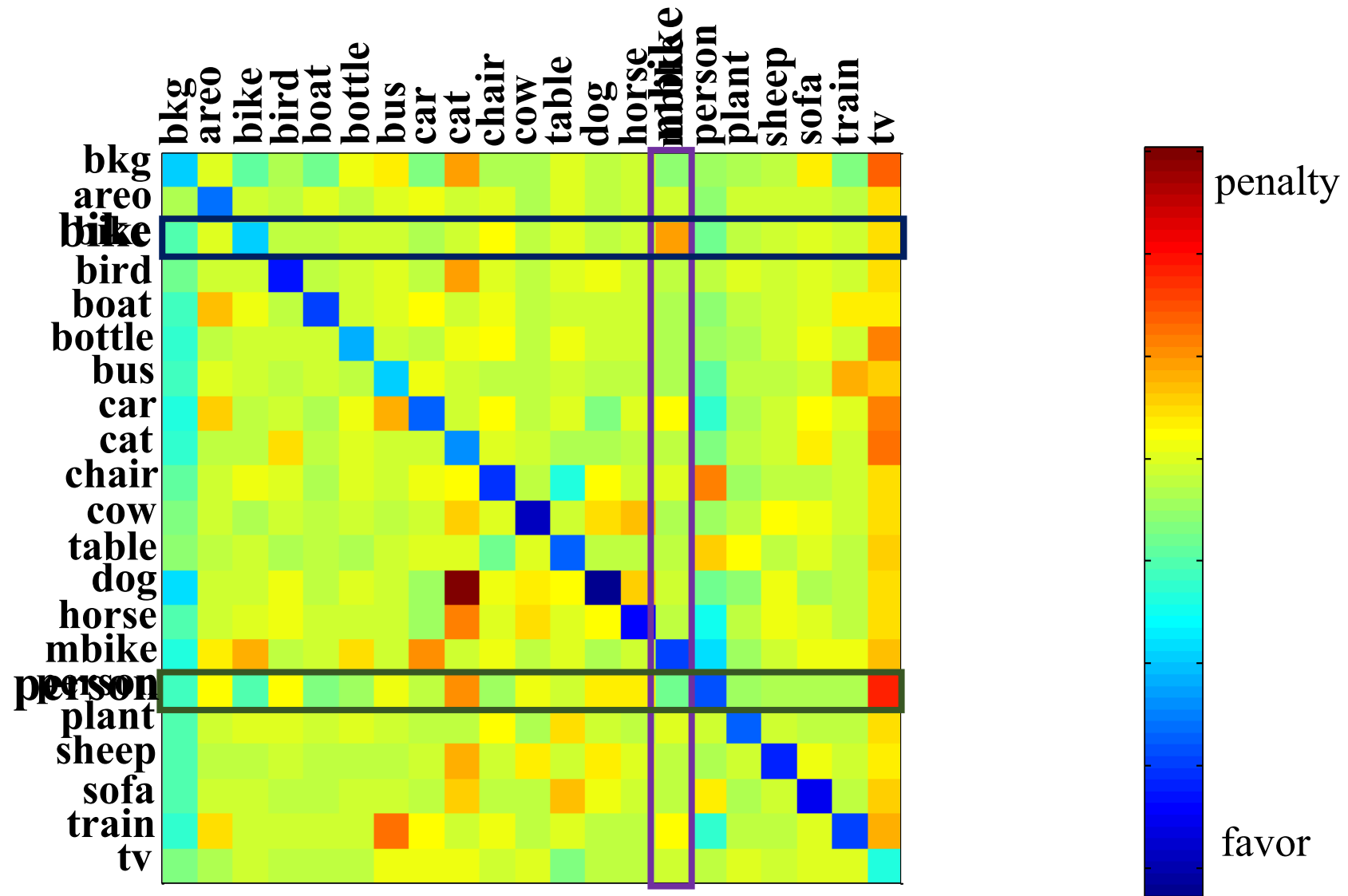


Label Contexts

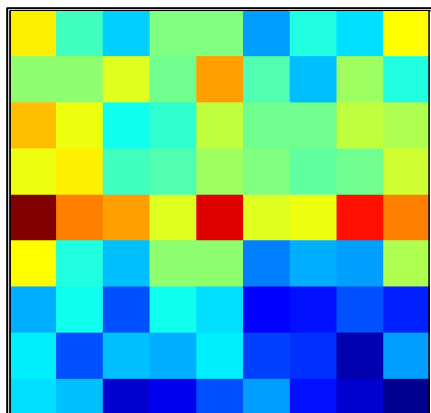


Joint Tuning

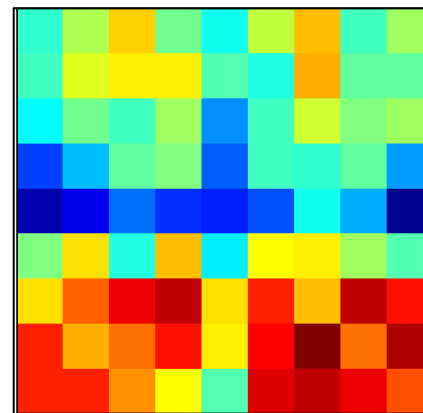
Target Structure



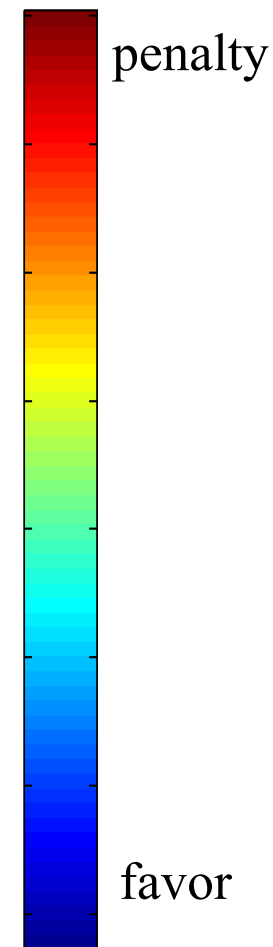
Target Structure



person : mbike

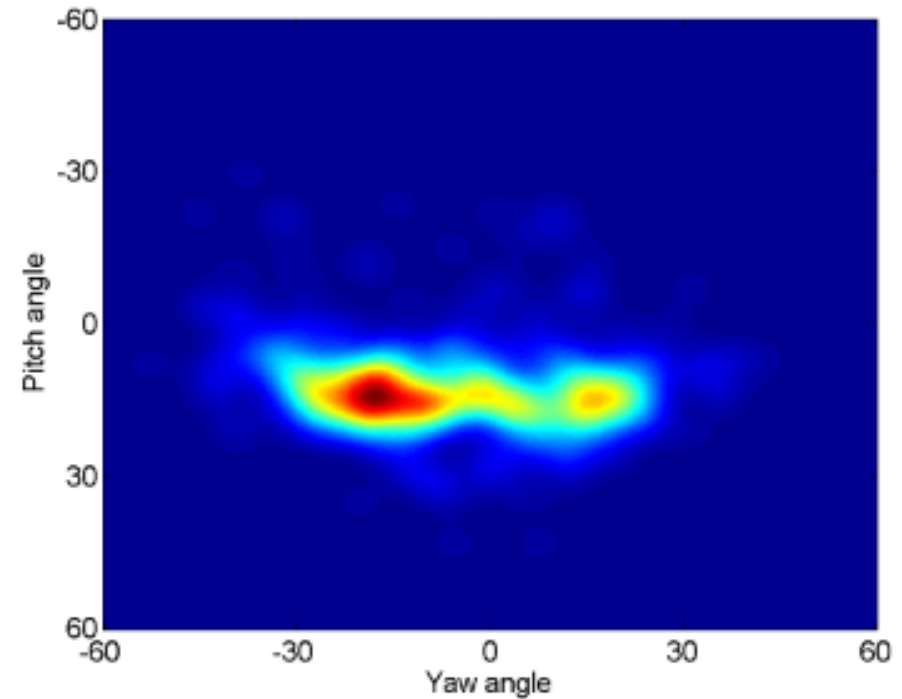
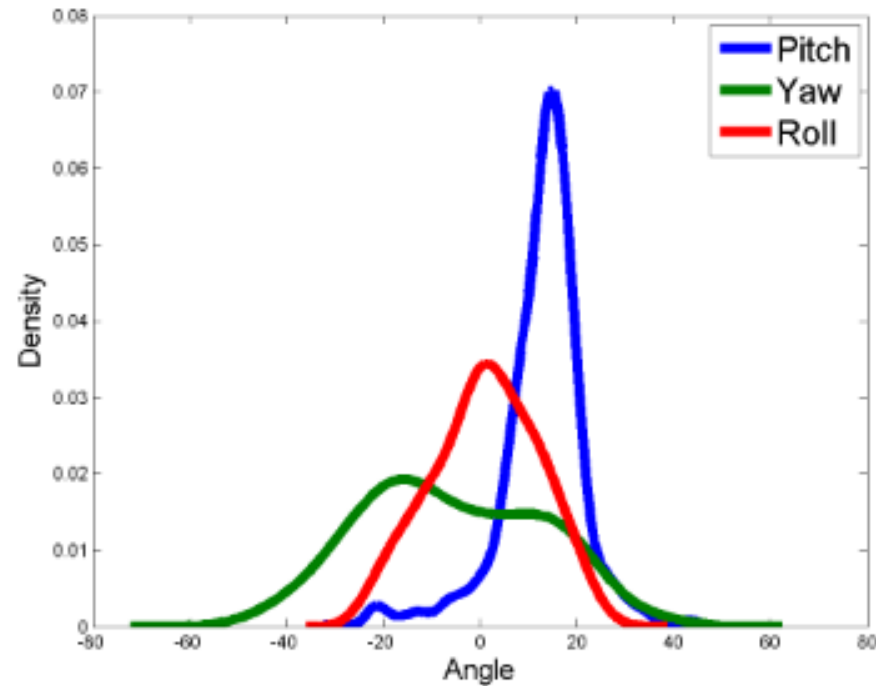


chair : person



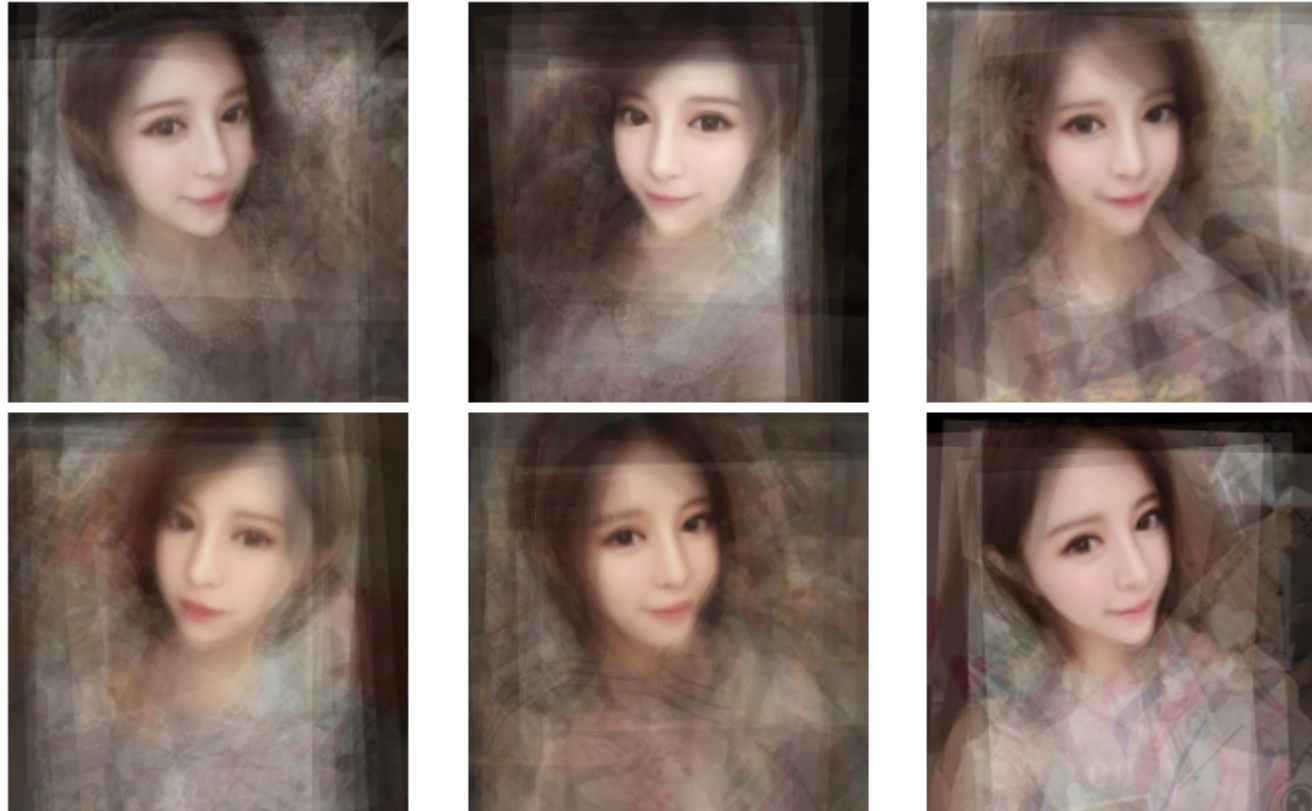
Target Structure

- Case Study II — best pose for a selfie



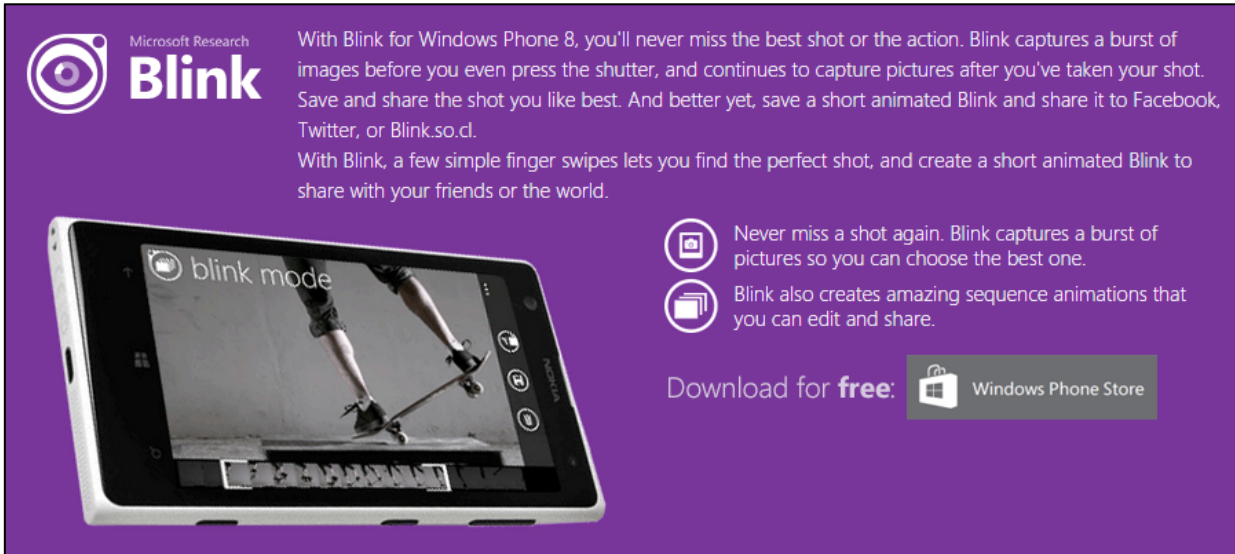
Target Structure

- Case Study II ——— best pose for a selfie



Reference

- Windows BLINK App




Microsoft Research **Blink**

With Blink for Windows Phone 8, you'll never miss the best shot or the action. Blink captures a burst of images before you even press the shutter, and continues to capture pictures after you've taken your shot. Save and share the shot you like best. And better yet, save a short animated Blink and share it to Facebook, Twitter, or Blink.so.cl.

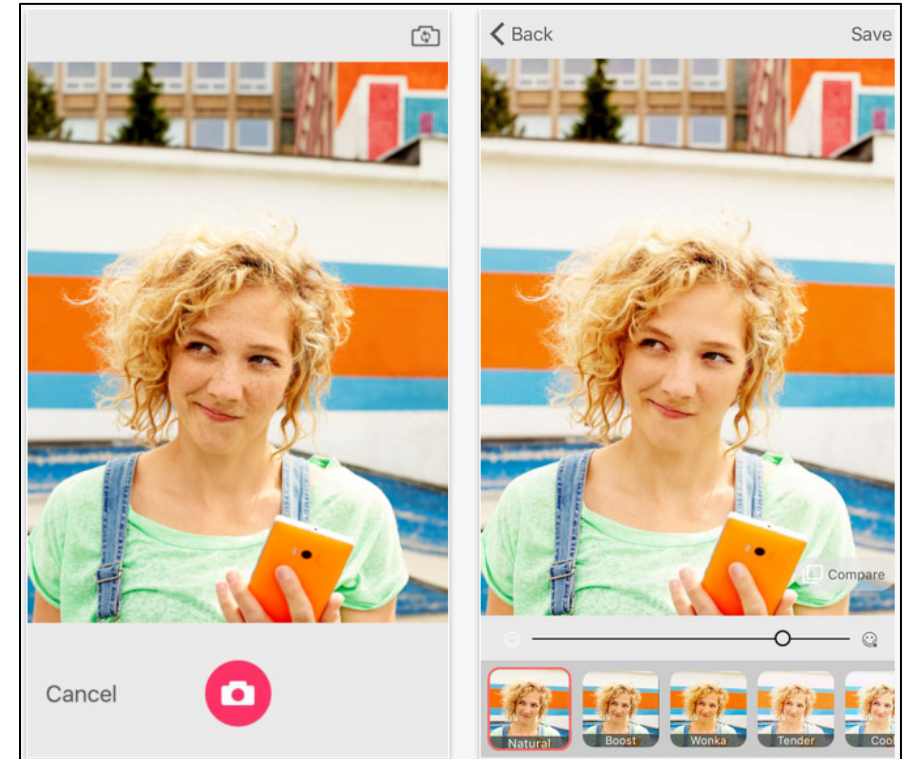
With Blink, a few simple finger swipes lets you find the perfect shot, and create a short animated Blink to share with your friends or the world.

Never miss a shot again. Blink captures a burst of pictures so you can choose the best one.

Blink also creates amazing sequence animations that you can edit and share.

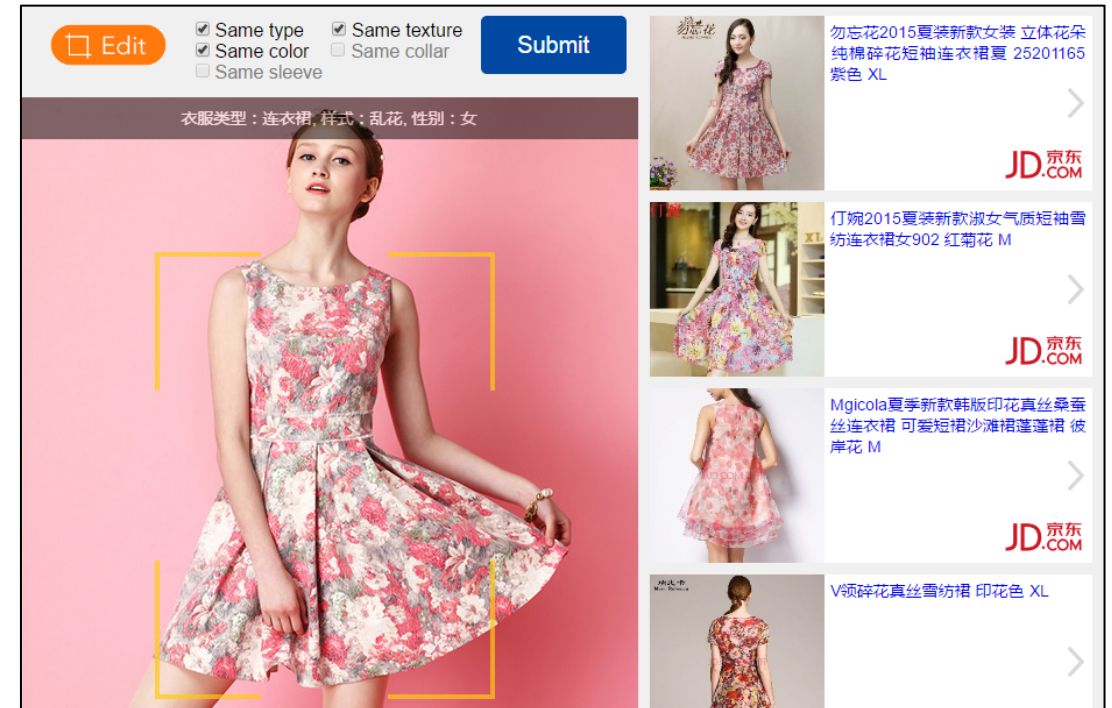
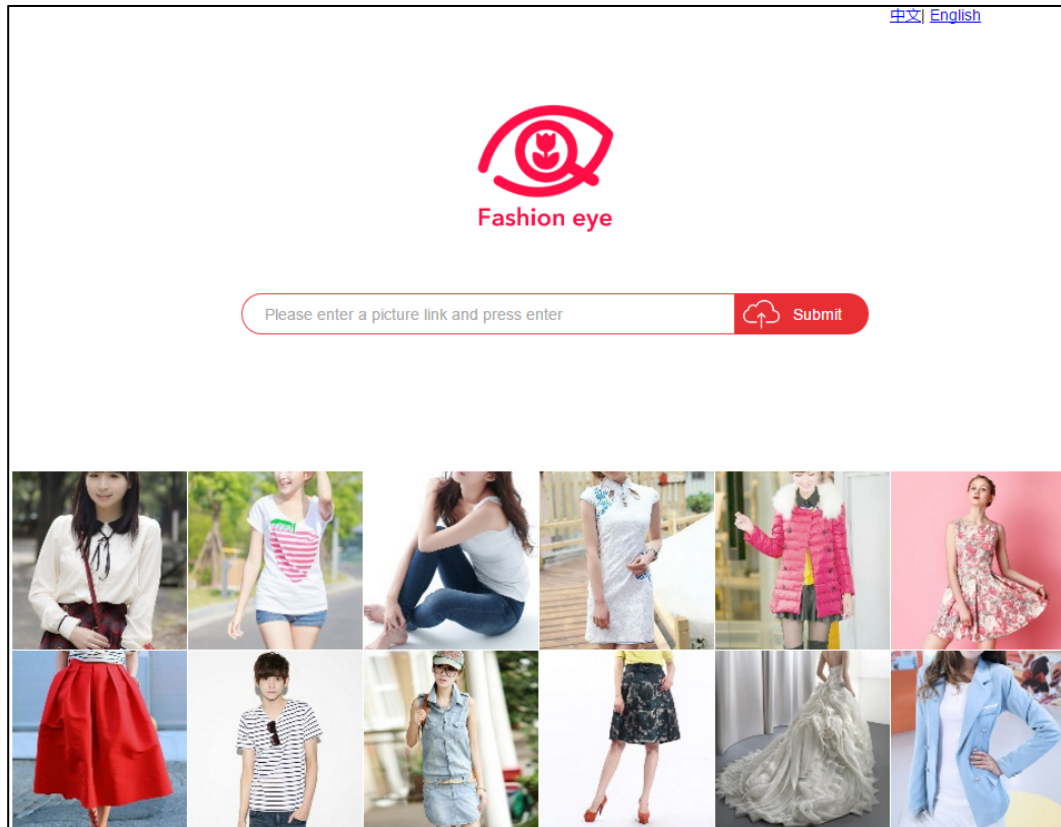
Download for **free**:  Windows Phone Store

The graphic features a purple background. On the left, a Windows Phone 8 is shown in 'blink mode', displaying a sequence of images of a person's arm. On the right, there are two circular icons: one with a camera and a burst of images, and another with a film strip. At the bottom right, there is a button to download the app for free from the Windows Phone Store.



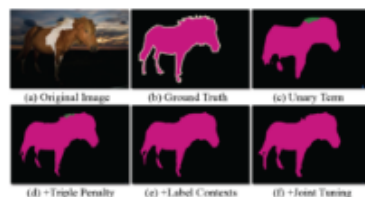
Reference

- SenseTime Fashion Eye



Reference

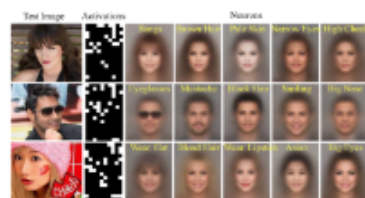
- More Details



Semantic Image Segmentation via Deep Parsing Network

Ziwei Liu*, Xiaoxiao Li*, Ping Luo, Chen Change Loy, Xiaoou Tang.
International Conference on Computer Vision (ICCV), 2015 (Oral)

[PDF](#) [Project Page](#)



Deep Learning Face Attributes in the Wild

Ziwei Liu, Ping Luo, Xiaogang Wang, Xiaoou Tang.
International Conference on Computer Vision (ICCV), 2015

[PDF](#) [Project Page](#) [Dataset](#)

Burst Images Denoising



Fast Burst Images Denoising

Ziwei Liu, Lu Yuan, Xiaoou Tang, Matt Uyttendaele, Jian Sun.
ACM Transactions on Graphics (SIGGRAPH Asia), 2014

[PDF](#) [Project Page](#) [Product Transfer](#) [iOS App](#)

Q & A