

Context Embedding Networks

K. Kim, O. Mac Aodha, P. Perona
 {khkimm, macaodha, perona}@caltech.edu

California Institute
 of Technology



GOALS

- Crowd-source the task of clustering an image dataset (e.g. on Amazon MTurk)
- Learn the different **clustering criteria** that workers use
- Learn a low-dimensional **embedding** of the image dataset

CHALLENGES & APPROACH

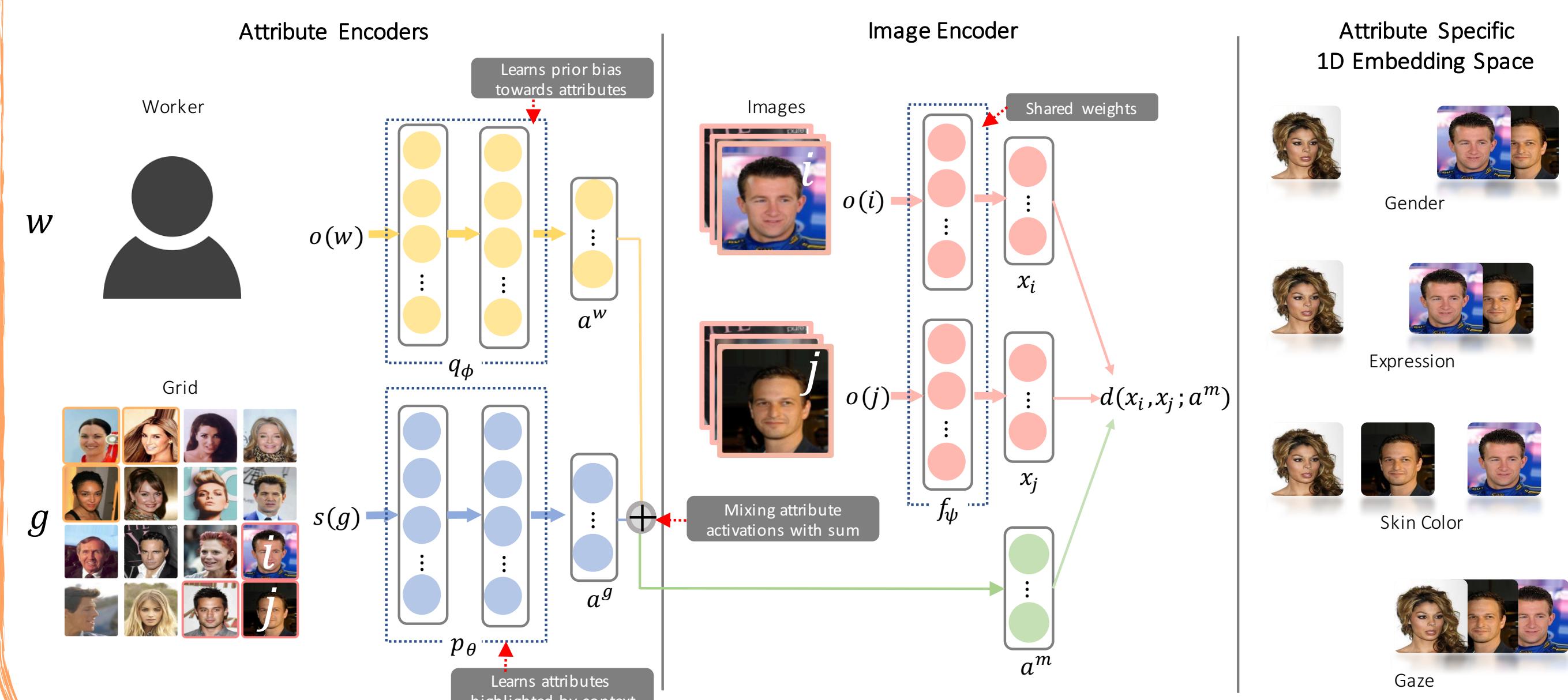
Challenges:

- Each worker sees a **small subset** of images
- Context** may bias workers
- Workers** may have inherent biases

Approach:

- Jointly learn a worker model, a context model, and an image embedding from pairwise similarity labels
- Learn without knowledge of clustering criteria

MODEL



Pairwise Similarity Labels

w : worker index, g : grid index, i_g, j_g : image index within grid g
 l : pairwise similarity label = +1(same group) or 0(different group)

$$\mathcal{D} = \{(w, g, i_g, j_g, l) | g = 1, \dots, G\}$$

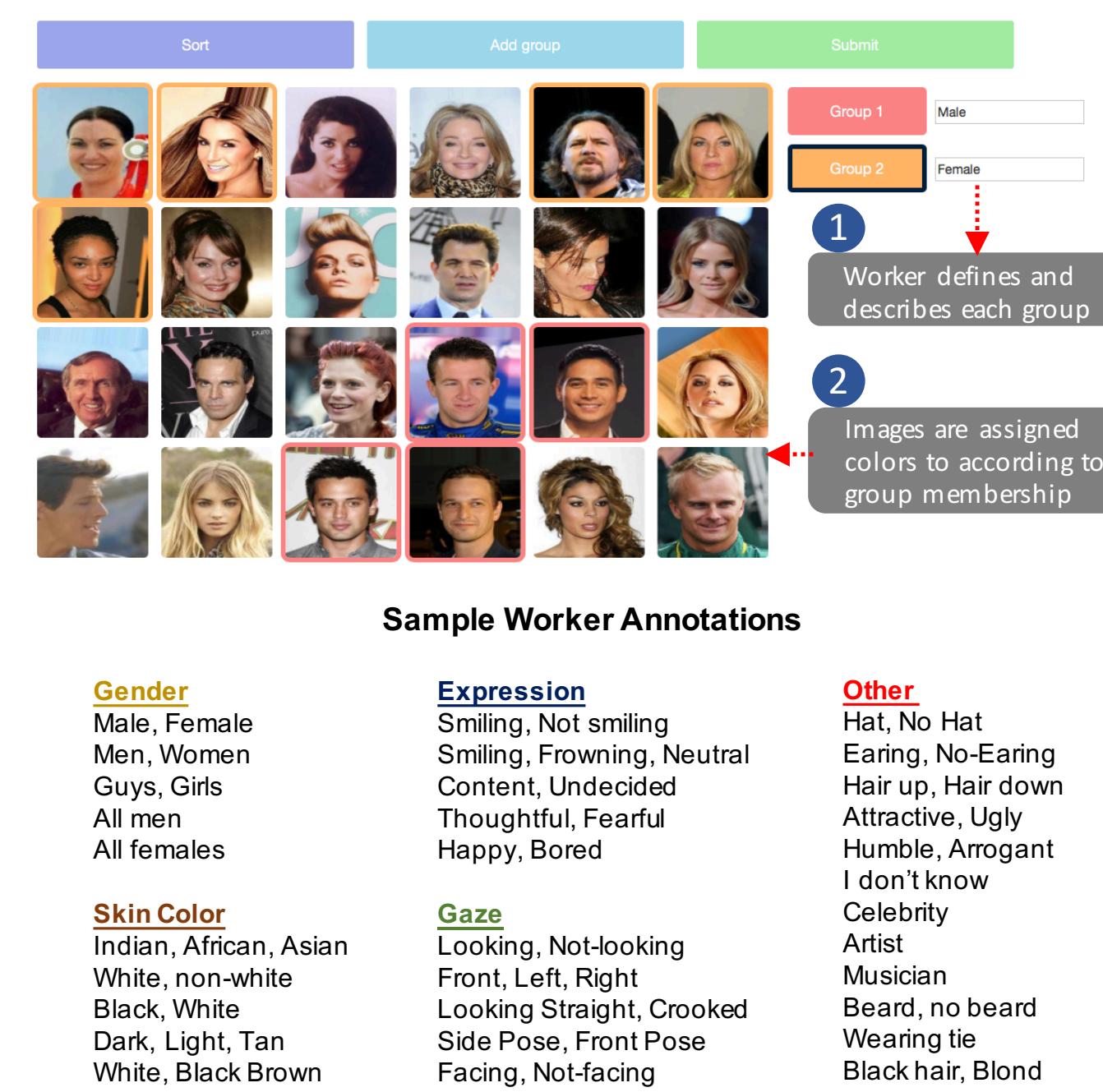
Embedding Distance

$$d(x_i, x_j; a^m) = \|a^m \cdot (x_i - x_j)\|_2 = \|(p_\theta(g) + q_\phi(w)) \cdot (f_\psi(i) - f_\psi(j))\|_2$$

Loss Function

$$L_{CEN}(x_i, x_j; a) = \gamma l \max\{0, d(x_i, x_j; a) - \xi_p\} + (1 - l) \max\{0, \xi_n - d(x_i, x_j)\} + \lambda_1 \|a\|_1 + \lambda_2 \|x_i\|_2 + \lambda_3 \|x_j\|_2.$$

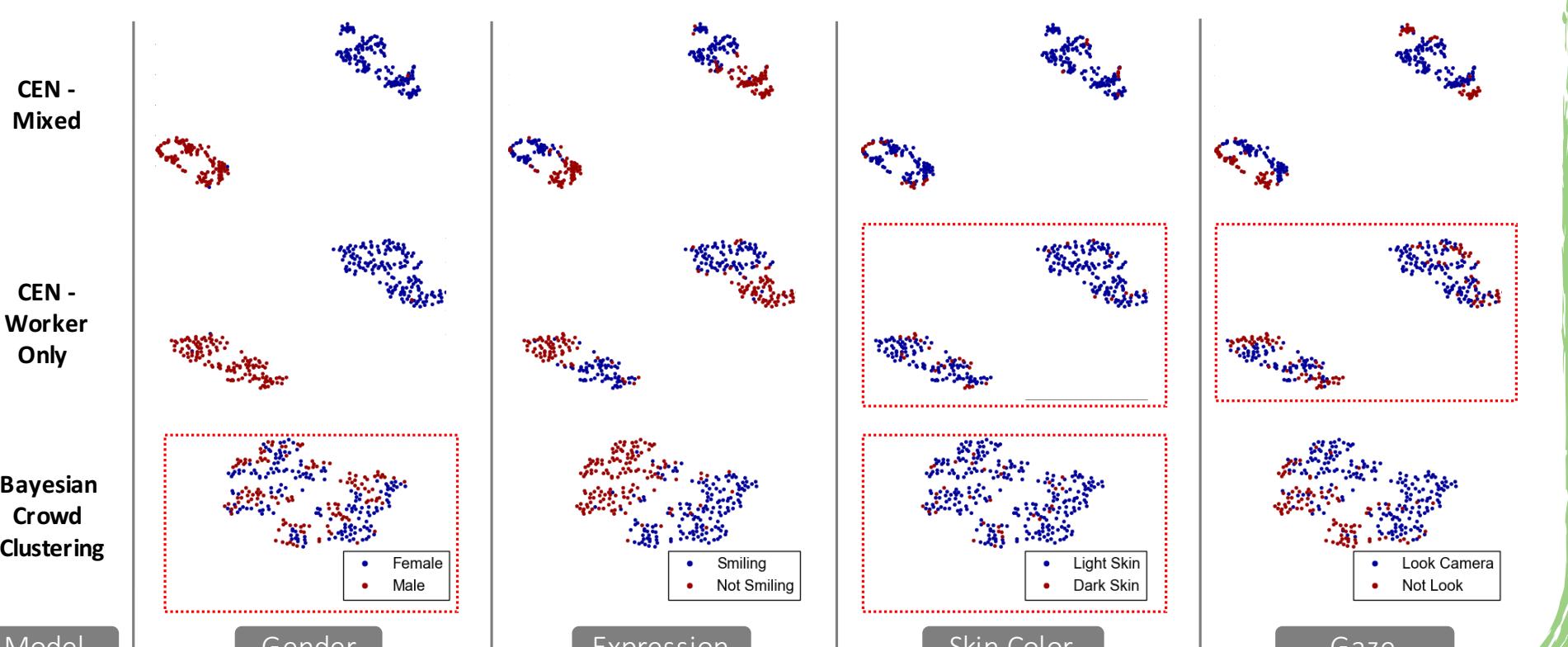
DATA COLLECTION GUI



RESULTS

- Model accurately retrieves clustering criteria (> 85%)
- Beats existing methods in heldout label prediction
- Learns interpretable embeddings
- Learns sensible worker, context models

EMBEDDING COMPARISON



WORKER MODEL

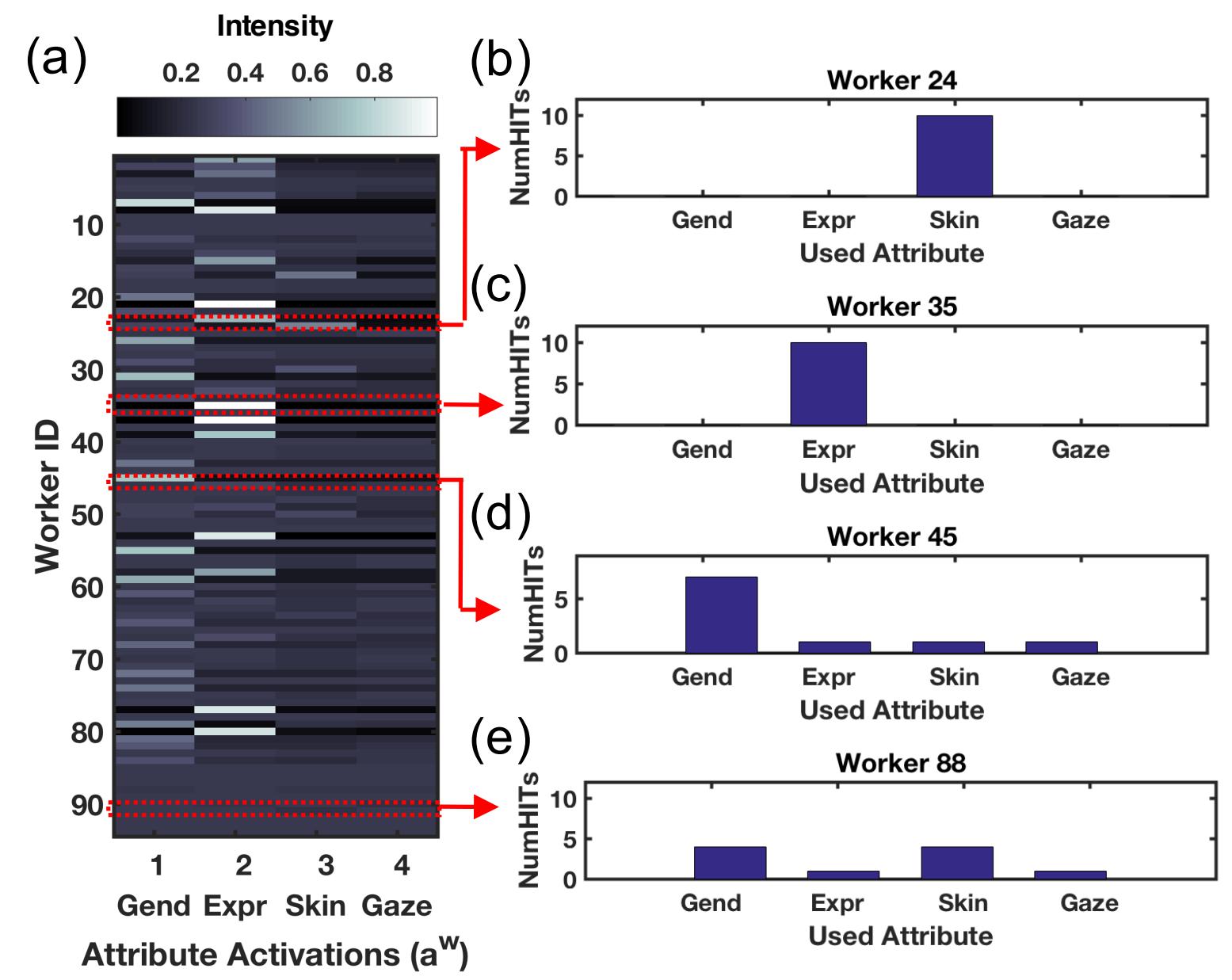


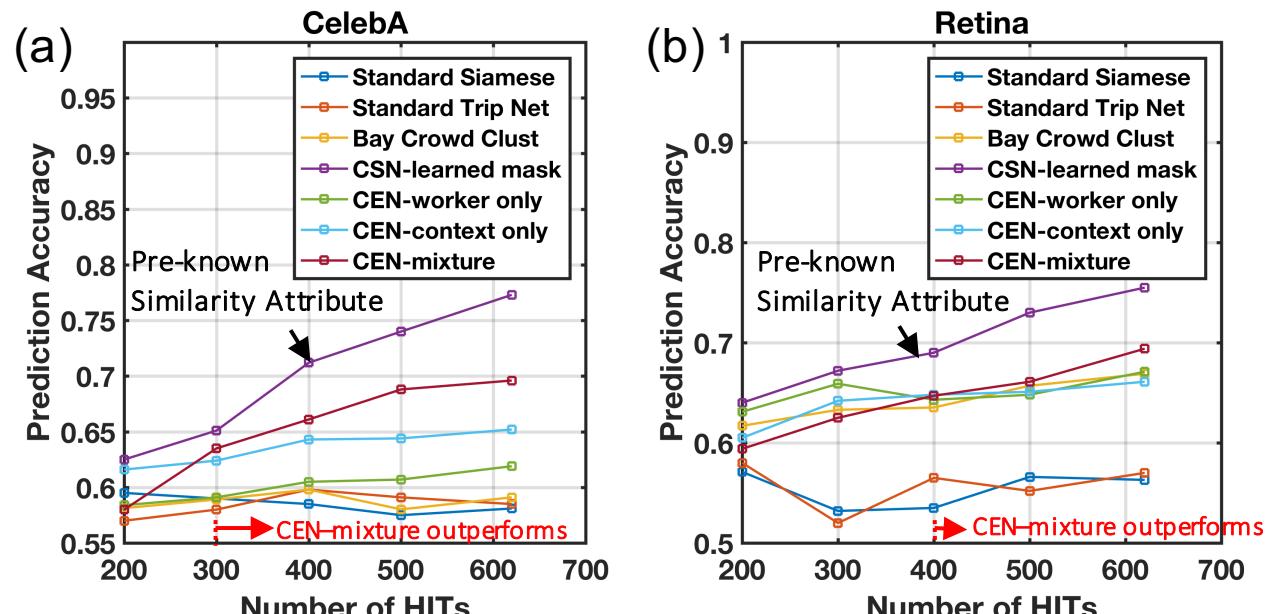
IMAGE EMBEDDING



SYNTHESIZED IMAGE GRID



HELDOUT LABEL PREDICTION



REFERENCES

- [1]. R. Gomes et al., "Crowd Clustering," In NIPS, 2011
 [2]. A. Veit et al., "Conditional Similarity Networks," In CVPR 2017