What's In A Name?

Huizhong Chen, Andrew C. Gallagher, Bernd Girod

Outline

- Extra background
- Cross-dataset performance
- Obscured face performance
- My face performance

Prelude: Is Implicit Egotism real?

- Implicit Egotism: Dennis is likely to be a Dentist
  - Because both start with "Den-"
  - Original studies establishing this were surprisingly small sample size
- Further investigated by Simonsohn
  - http://datacolada.org/wp-content/uploads/2015/04/Spurious-Published-JPSP.pdf
  - Control for socioeconomic status and changing demographics
  - Then, the differences are explainable
- Implicit Egotism is, then, a real effect
  - Names represent some kind of prior on a person
  - But, the effect is not necessarily psychological
- Could be studied at a larger scale
  - Not necessarily a consensus in the field as to Simonsohn vs. other views

Experiment 1: Cross-Dataset Investigation

- Validate their model on their dataset
  - Tried across a subset of 200 names of each gender
  - Performed name and gender classification
- Also tested on ~400 randomly selected IMDB-Wiki images
  - Dataset of celebrity faces and names
  - Crawled from IMDB and Wikipedia for supplementary age training
  - 500k images in total
  - Tested using only the same names as the original paper and using all names
  - Source: https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/

Name Classification Results

	Accuracy	Random Chance
Names100Dataset	68.5%	1%
ImdbWiki in domain	3%	1.2%
ImdbWiki	0%	.004%

Name Classification Observations

- The authors have overfit to their dataset
  - They do claim they train on all 80000 images
- Performance on in-domain images is on par with reported performance
- Performance on all names is poor but expected

Gender Classification Results

	Accuracy	Random Chance
Names100Dataset	86.5%	52.0%
ImdbWiki in domain	83.4%	53.2%
ImdbWiki	77.2%	51.0%

Misclassified Faces from Names100Dataset

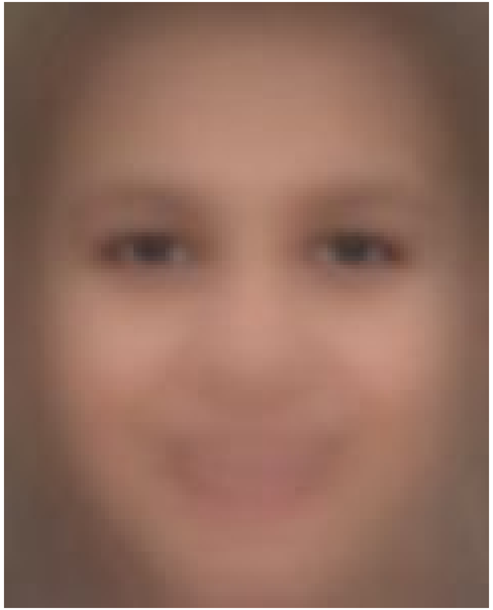
Misclassified Faces from IMDBWiki

Gender Classification Qualitative Observations

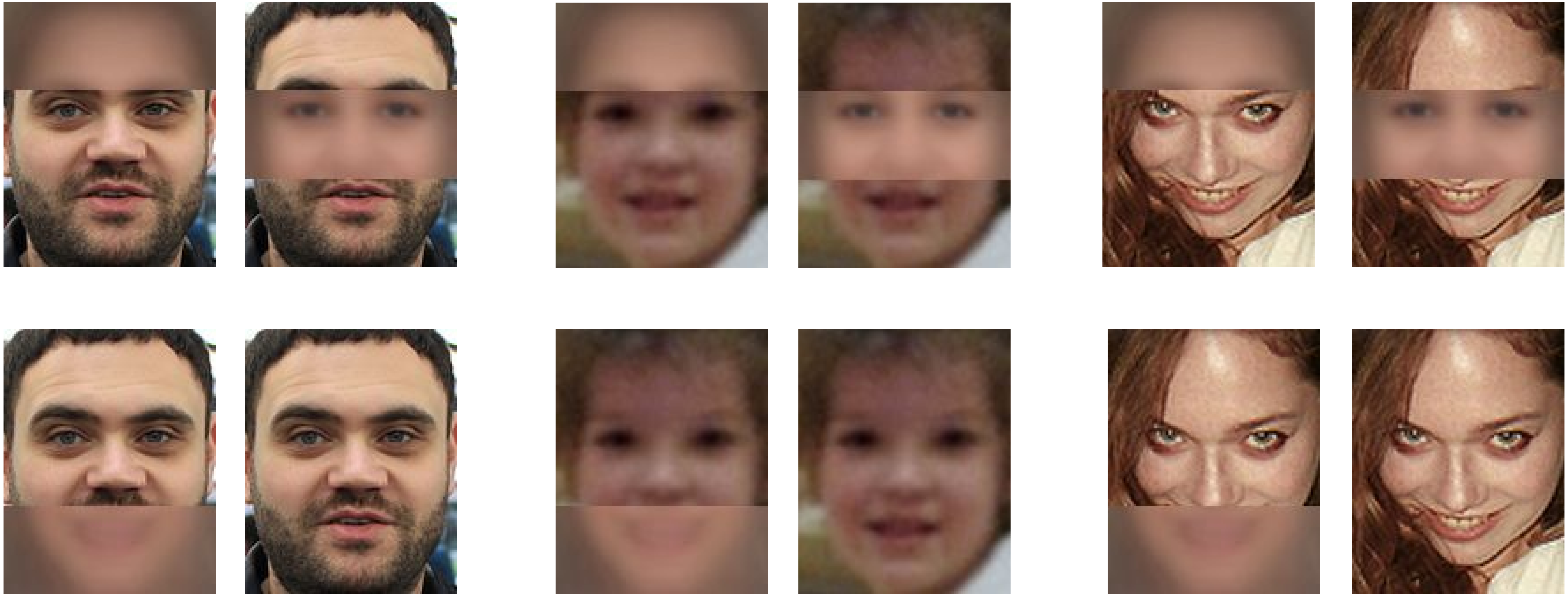
- Results on IMDBWiki are impressive
  - Both in and out of domain
- This model struggles with children
- The Names100Dataset is not well annotated
- Gains might be made by:
  - Training on different data
  - Pruning the Names 100 dataset for better annotations
  - Breaking the task into children / adults

Experiment 2: What's in a Face?

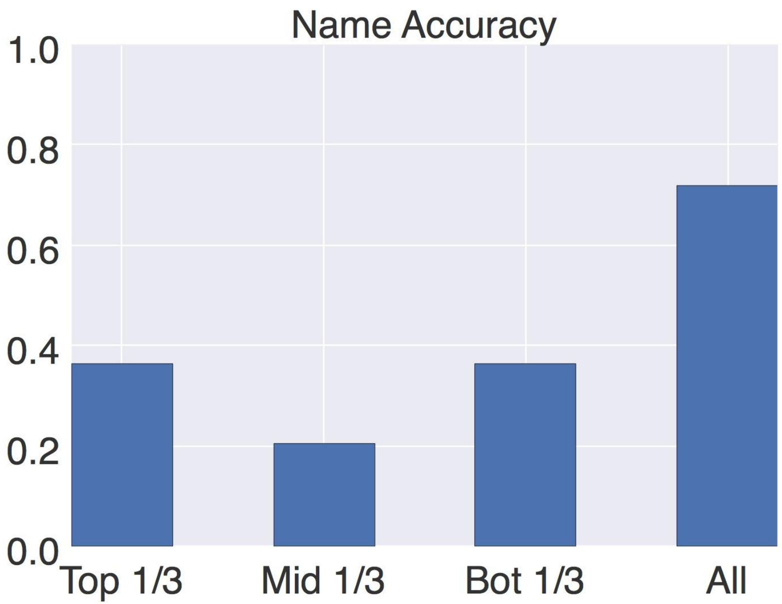
- Using Names100Dataset
- Replaced the top, bottom, and middle third with the average across all images of both genders
- Then ran tests for name & gender classification for both datasets
- What part of a face does the classifier rely on?

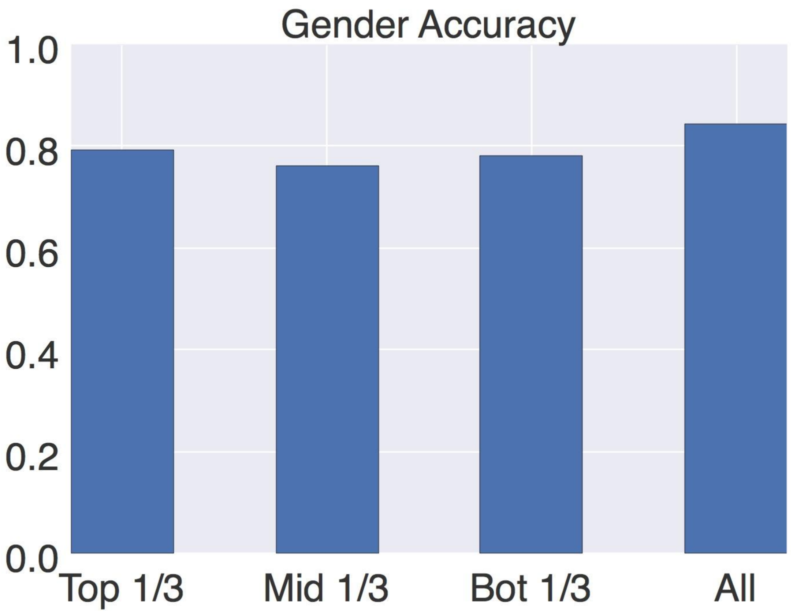


Example Averaged Images

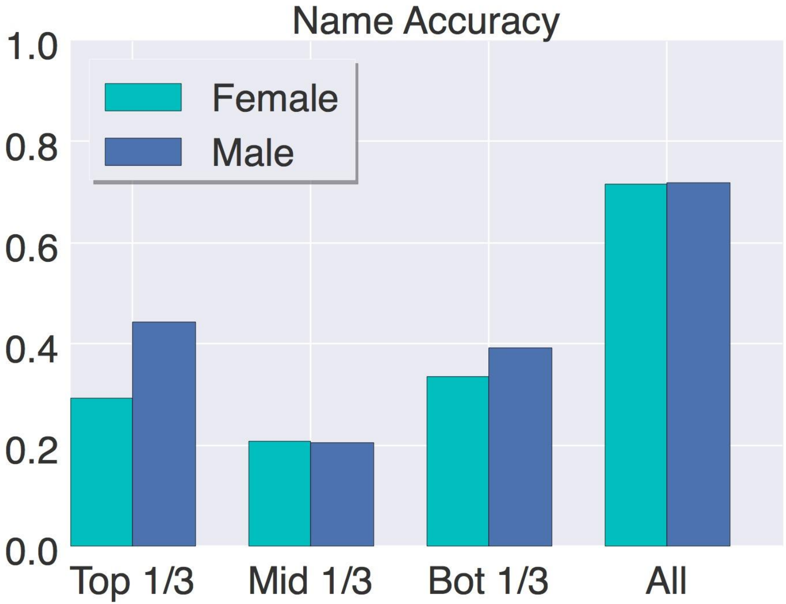


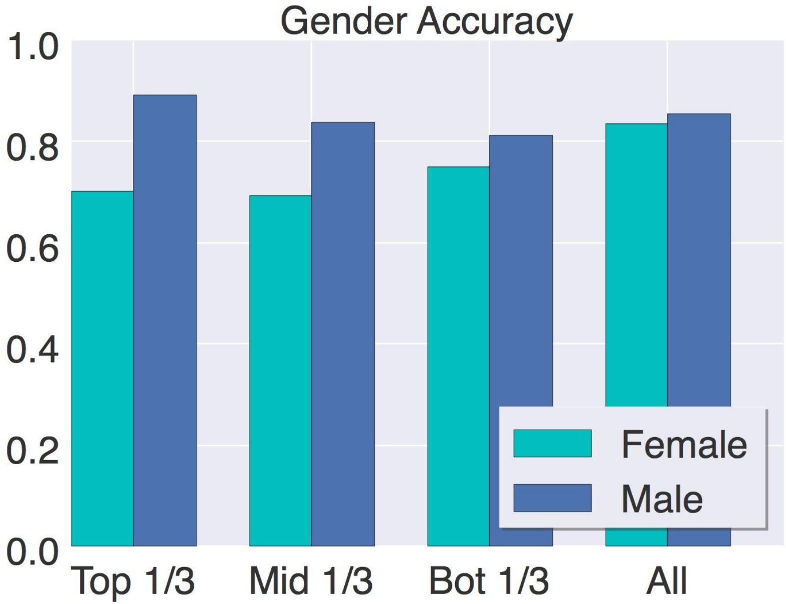
Results of Averaged Images





Results by Gender





Precision & Recall by Gender

	Female Precision	Female Recall	Male Precision	Male Recall
Top 1/3	.87	.70	.73	.89
Middle 1/3	.82	.69	.72	.83
Bottom 1/3	.81	.75	.75	.81
All	.86	.84	.83	.85

Averaged Images Summary

- Genders:
  - Baseline performance is comparable across genders
  - Blurring adversely affects female prediction more than male
- Names:
  - Performance is most adversely impacted by blurring the middle third of the face
  - Significant hit regardless

Top 3 predicted names