CS6501: Deep Learning for Visual Recognition
Fairness in Computer Vision
Today’s Class

• Disparities in Face Recognition
• Disparities in Visual Recognition
• Disparities in Image Captioning
Gender Shades

Gender Shades

<table>
<thead>
<tr>
<th>Gender Classifier</th>
<th>Female Subjects Accuracy</th>
<th>Male Subjects Accuracy</th>
<th>Error Rate Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft</td>
<td>89.3%</td>
<td>97.4%</td>
<td>8.1%</td>
</tr>
<tr>
<td>FACE++</td>
<td>78.7%</td>
<td>99.3%</td>
<td>20.6%</td>
</tr>
<tr>
<td>IBM</td>
<td>79.7%</td>
<td>94.4%</td>
<td>14.7%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender Classifier</th>
<th>Darker Male</th>
<th>Darker Female</th>
<th>Lighter Male</th>
<th>Lighter Female</th>
<th>Largest Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft</td>
<td>94.0%</td>
<td>79.2%</td>
<td>100%</td>
<td>98.3%</td>
<td>20.8%</td>
</tr>
<tr>
<td>FACE++</td>
<td>99.3%</td>
<td>65.5%</td>
<td>99.2%</td>
<td>94.0%</td>
<td>33.8%</td>
</tr>
<tr>
<td>IBM</td>
<td>88.0%</td>
<td>65.3%</td>
<td>99.7%</td>
<td>92.9%</td>
<td>34.4%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Metric</th>
<th>All</th>
<th>F</th>
<th>M</th>
<th>Darker</th>
<th>Lighter</th>
<th>DF</th>
<th>DM</th>
<th>LF</th>
<th>LM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PPV(%)</td>
<td>93.7</td>
<td>89.3</td>
<td>97.4</td>
<td>87.1</td>
<td>99.3</td>
<td>79.2</td>
<td>94.0</td>
<td>98.3</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Error Rate(%)</td>
<td>6.3</td>
<td>10.7</td>
<td>2.6</td>
<td>12.9</td>
<td>0.7</td>
<td>20.8</td>
<td>6.0</td>
<td>1.7</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>TPR (%)</td>
<td>93.7</td>
<td>96.5</td>
<td>91.7</td>
<td>87.1</td>
<td>99.3</td>
<td>92.1</td>
<td>83.7</td>
<td>100</td>
<td>98.7</td>
</tr>
<tr>
<td></td>
<td>FPR (%)</td>
<td>6.3</td>
<td>8.3</td>
<td>3.5</td>
<td>12.9</td>
<td>0.7</td>
<td>16.3</td>
<td>7.9</td>
<td>1.3</td>
<td>0.0</td>
</tr>
<tr>
<td>MSFT</td>
<td>PPV(%)</td>
<td>90.0</td>
<td>78.7</td>
<td>99.3</td>
<td>83.5</td>
<td>95.3</td>
<td>65.5</td>
<td>99.3</td>
<td>94.0</td>
<td>99.2</td>
</tr>
<tr>
<td></td>
<td>Error Rate(%)</td>
<td>10.0</td>
<td>21.3</td>
<td>0.7</td>
<td>16.5</td>
<td>4.7</td>
<td>34.5</td>
<td>0.7</td>
<td>6.0</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>TPR (%)</td>
<td>90.0</td>
<td>98.9</td>
<td>85.1</td>
<td>83.5</td>
<td>95.3</td>
<td>98.8</td>
<td>76.6</td>
<td>98.9</td>
<td>92.9</td>
</tr>
<tr>
<td></td>
<td>FPR (%)</td>
<td>10.0</td>
<td>14.9</td>
<td>1.1</td>
<td>16.5</td>
<td>4.7</td>
<td>23.4</td>
<td>1.2</td>
<td>7.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Face++</td>
<td>PPV(%)</td>
<td>87.9</td>
<td>79.7</td>
<td>94.4</td>
<td>77.6</td>
<td>96.8</td>
<td>65.3</td>
<td>88.0</td>
<td>92.9</td>
<td>99.7</td>
</tr>
<tr>
<td></td>
<td>Error Rate(%)</td>
<td>12.1</td>
<td>20.3</td>
<td>5.6</td>
<td>22.4</td>
<td>3.2</td>
<td>34.7</td>
<td>12.0</td>
<td>7.1</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>TPR (%)</td>
<td>87.9</td>
<td>92.1</td>
<td>85.2</td>
<td>77.6</td>
<td>96.8</td>
<td>82.3</td>
<td>74.8</td>
<td>99.6</td>
<td>94.8</td>
</tr>
<tr>
<td></td>
<td>FPR (%)</td>
<td>12.1</td>
<td>14.8</td>
<td>7.9</td>
<td>22.4</td>
<td>3.2</td>
<td>25.2</td>
<td>17.7</td>
<td>5.20</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Both Microsoft and IBM have updated their systems since the publishing of Gender Shades, the gaps in their error rates have decreased.

Still a gap.
Men Also Like Shopping

Situation Recognition

- Jumping: 0.33
- Playing: 0.22
- Carrying: 0.26
- Cooking: 0.14
- Cleaning: 0.72

Commonly Uncommon: Semantic Sparsity in Situation Recognition
Situation Recognition (CVPR 2017)

Commonly Uncommon: Semantic Sparsity in Situation Recognition
Mark Yatskar, Vicente Ordonez, Luke Zettlemoyer, Ali Farhadi
Conditional Random Field (CRF) + Deep Neural Network
**Commonly Uncommon: Semantic Sparsity in Situation Recognition**
Mark Yatskar, Vicente Ordonez, Luke Zettlemoyer, Ali Farhadi
CVPR 2017

**Image Regression**

$$p(S|i; \theta) \propto \psi_v(v, i; \theta) \prod_{(e, n_e) \in R_f} \psi_e(v, e, n_e, i; \theta)$$

$$\psi_e(v, e, n_e, i; \theta) = e^{\phi_e(v, e, n_e, i, \theta)}$$

**Conditional Random Field (CRF) + Deep Neural Network**

$$\psi_v(v, i; \theta) = e^{\phi_v(v, i, \theta)}$$

**CLEAN**

- **AGENT**
  - man
- **SOURCE**
  - chimney
- **DIRT**
  - soot
- **TOOL**
  - brush
- **PLACE**
  - roof
### Situation Recognition: CVPR 2017

**Compositional Shared Learning of Underlying Concepts**

**Commonly Uncommon: Semantic Sparsity in Situation Recognition**
Mark Yatskar, Vicente Ordonez, Luke Zettlemoyer, Ali Farhadi  
**CVPR 2017**

![Query Image](http://imsitu.org/demo/)

**Recognize Situations**

<table>
<thead>
<tr>
<th>Predicted situations</th>
<th>falling</th>
<th>whipping</th>
<th>rearing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent</td>
<td>Source</td>
<td>Goal</td>
<td>Place</td>
</tr>
<tr>
<td>Person</td>
<td>Horse</td>
<td>Land</td>
<td>Outdoors</td>
</tr>
</tbody>
</table>

- **Falling**
  - Agent: person
  - Source: horse
  - Goal: land
  - Place: outdoors
  - Score: 0.58372

- **Whipping**
  - Agent: jockey
  - Item: horse
  - Tool: whip
  - Place: outdoors
  - Score: 0.10375

- **Rearing**
  - Agent: horse
  - Place: grass
  - Score: 0.07997

[http://imsitu.org/demo/](http://imsitu.org/demo/)
However we kept running into this…

http://imsitu.org/demo/

Commonly Uncommon: Semantic Sparsity in Situation Recognition
However we kept running into this…

http://imsitu.org/demo/

### Commonly Uncommon: Semantic Sparsity in Situation Recognition

Mark Yatskar, Vicente Ordonez, Luke Zettlemoyer, Ali Farhadi

**CVPR 2017**

<table>
<thead>
<tr>
<th>Predicted situations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>dusting</strong></td>
</tr>
<tr>
<td>agent</td>
</tr>
<tr>
<td>woman</td>
</tr>
<tr>
<td><strong>vacuuming</strong></td>
</tr>
<tr>
<td>agent</td>
</tr>
<tr>
<td>woman</td>
</tr>
<tr>
<td><strong>cleaning</strong></td>
</tr>
<tr>
<td>agent</td>
</tr>
<tr>
<td>woman</td>
</tr>
</tbody>
</table>
Key Finding: Models Amplify Biases in the Dataset
Key Finding: Models Amplify Biases in the Dataset

Dataset? -> Model?

Images of People Cooking
Key Finding: Models Amplify Biases in the Dataset

- Dataset?
- Model?

Men Cooking: 33%  
Women Cooking: 66%
Key Finding: Models Amplify Biases in the Dataset

Dataset?

Model?

Men Cooking: 33%  Women Cooking: 66%  Test Images
Key Finding: Models Amplify Biases in the Dataset

Dataset?

Model?

Men Cooking: 33%    Women Cooking: 66%

Men Cooking: 16%    Women Cooking: 84%
Contributions

High dataset gender bias
38% (objects) 47% (events) exhibit strong bias

Models amplify existing gender bias
~70% objects and events have bias amplification

Reducing bias amplification
~50% reduction in amplification
Insignificant loss in performance

data

model

RBA

imSitu vSRL
(events)

COCO MLC
(objects)
Gender Dataset Bias

- imSitu Verb
- COCO Noun

% of items

Unbiased Gender Ratio

Female bias

Male bias

coaching

lecturing

repairing

cooking

washing

braiding

shopping

imSitu Verb

COCO Noun
Gender Dataset Bias

- imSitu Verb
- COCO Noun

% of items

Female bias

Male bias

Unbiased Gender Ratio

- surfboard
- ski
- skateboard
- refrigerator
- fork
- bed
Men Also Like Shopping: Reducing Gender Bias Amplification Using Corpus Level Constraints

Dataset

Men Cooking: 33%
Women Cooking: 66%

Model

Men Cooking: 16%
Women Cooking: 84%

Model*

*Our solution: RBA: Optimize for accuracy but also to match data distribution.
Men Also Like Shopping: Reducing Gender Bias Amplification Using Corpus Level Constraints

Dataset

Men Cooking: 33%
Women Cooking: 66%

Model

Men Cooking: 16%
Women Cooking: 84%

Model*

Men Cooking: 20%
Women Cooking: 80%

*Our solution: RBA: Optimize for accuracy but also to match data distribution.
Reducing Bias Amplification (RBA)

Integer Linear Program

\[ \sum_{i} \max_{y_i} s(y_i, \text{image}) \]

\[ \forall \text{ points } \mid \text{Training Ratio - Predicted Ratio} \mid f(y_1 \ldots y_n) \leq \text{margin} \]

Lagrangian Relaxation

Sontag et al., 2011; Rush and Collins, 2012; Chang and Collins, 2011; Peng et al., 2015, Chang et al., 2013; Dalvi, 2015
Women Also Snowboard

Wrong

Right for the Right Reasons

Right for the Wrong Reasons

Wrong

Right for the Right Reasons

Right for the Wrong Reasons

Baseline:
A man sitting at a desk with a laptop computer.

Our Model:
A woman sitting in front of a laptop computer.

Baseline:
A man holding a tennis racquet on a tennis court.

Our Model:
A man holding a tennis racquet on a tennis court.

Women Also Snowboard

Women Also Snowboard

Women Also Snowboard

Women Also Snowboard

Final Advice: Accountability

• i.e. Detecting Criminality from Faces:

i.e. Detecting Sexual Orientation from Faces:

Recommended Reading:

Do algorithms reveal sexual orientation or just expose our stereotypes?
Questions?