
University of Colorado
Boulder

Deep Learning Applications for Computer Vision

Lecture 19: Deep Learning Networks – other considerations

The Data

- How do we build a good data set?

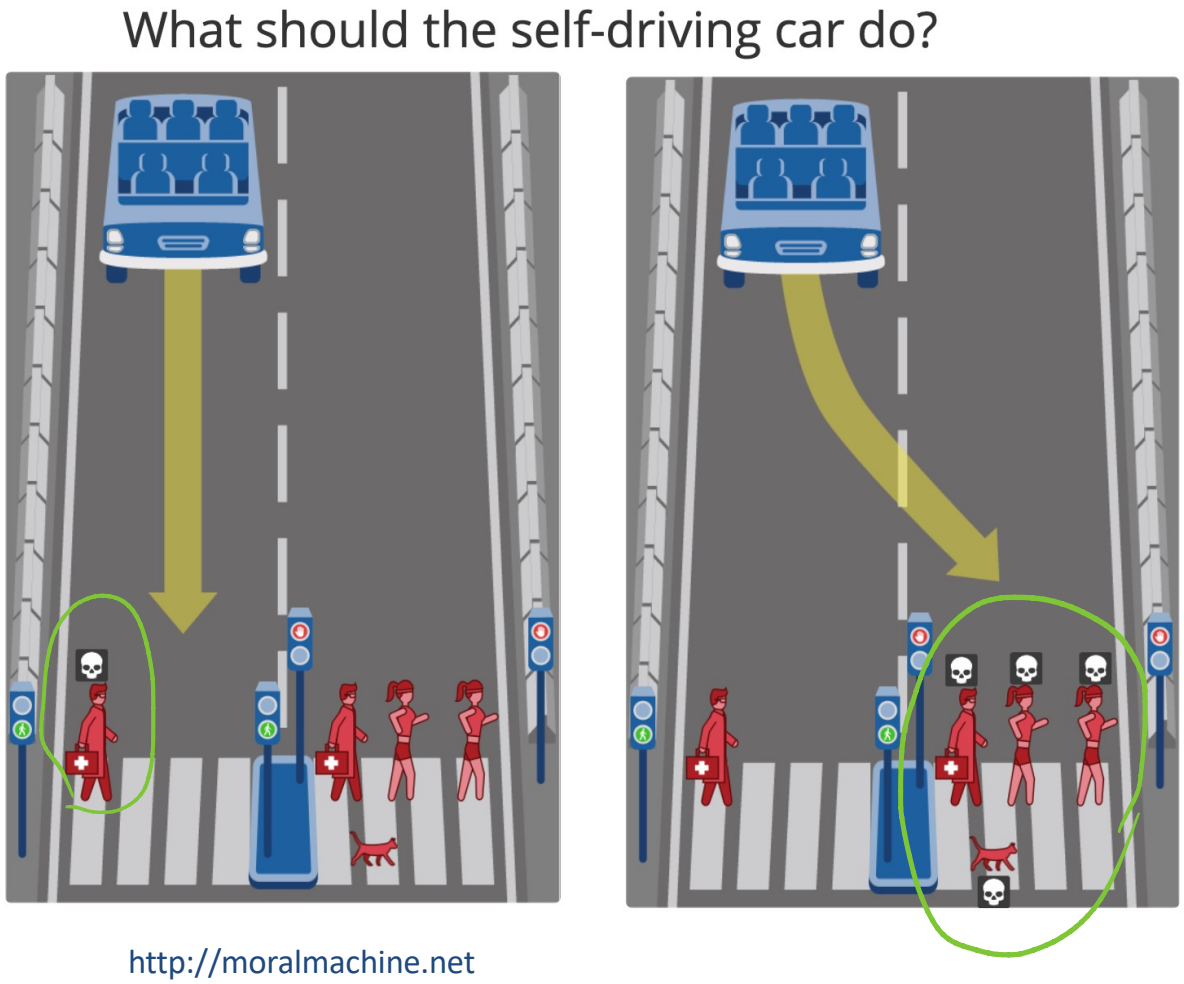
1. How much data is needed?
2. Quality training data
   - acquiring
   - labeling
   - preprocessing
3. Ethical considerations

How much data is needed?

- Dataset must represent the "reality"

Self-driving car:
- weather
- lighting

Less represented data:
↓
accuracy
↓
decisions
↓
impact

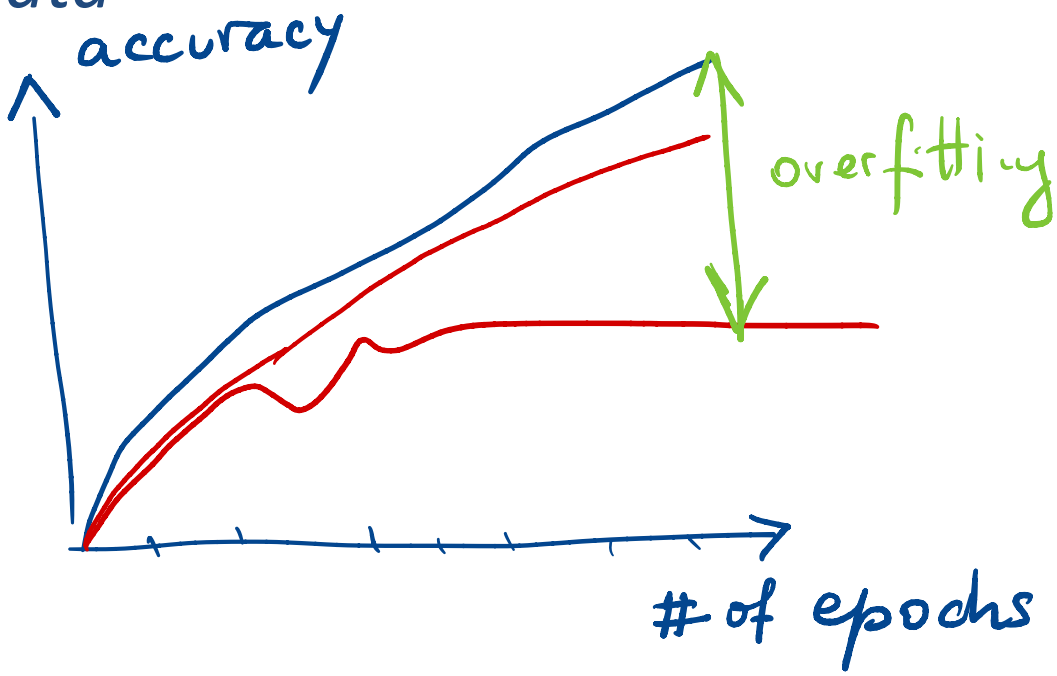


http://moralmachine.net

How much data is needed?

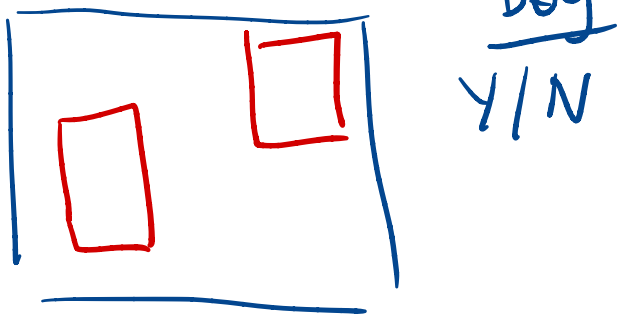
- Overfitting
  - mismatch between the model capacity and the data complexity
  - *insufficient training data*

- training data
- testing data



Labeling the Data

- Labelers' expertise:
  - Example: the Allen Institute for Brain Science
    - electron microscopy images
    - neuroscientists
- Multiple levels of labeling:
  - object occurrence, location, bounding box
  - variety of data form (images, point cloud from other sensors)



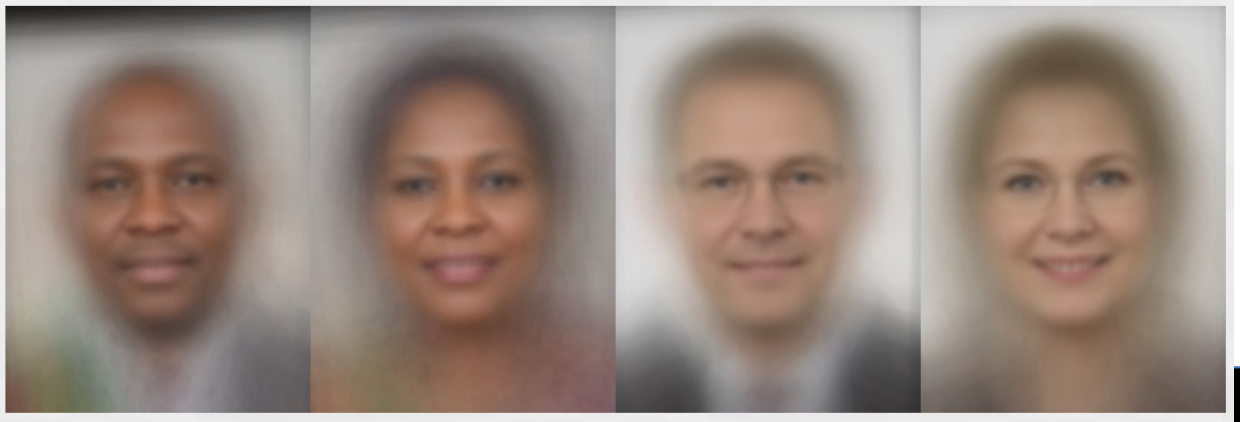
- Cultural considerations:
  - inherent biases, use of slurs

Ethical Considerations

- Dataset bias:
  - omission, underrepresentation

http://gendershades.com

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%



Ethical Considerations

- Dataset bias:
  - omission, underrepresentation
- Association bias:
  - reinforcement of human bias
- Unfair discrimination

Financial Institution → Loan
• geografic data

↳ women wear pink