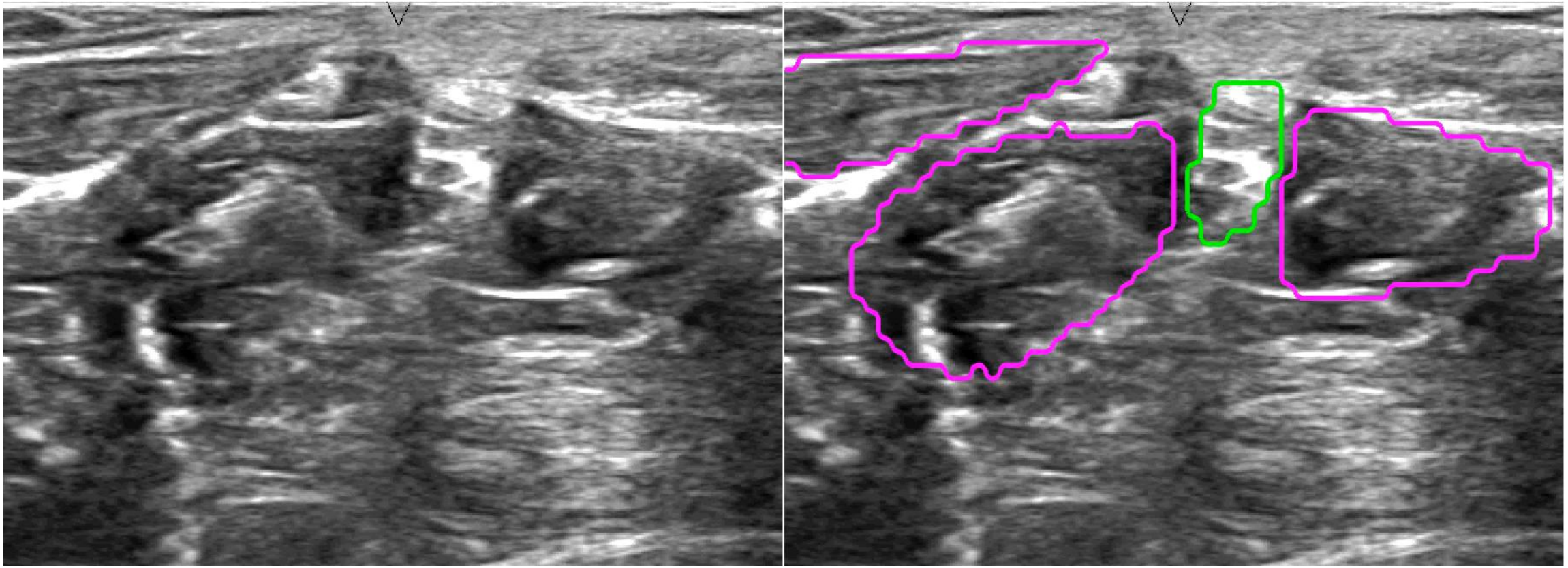
Data Science Summit 2016

San Francisco

Why do we care about AI?

- PACMAN: Nerve Visualization



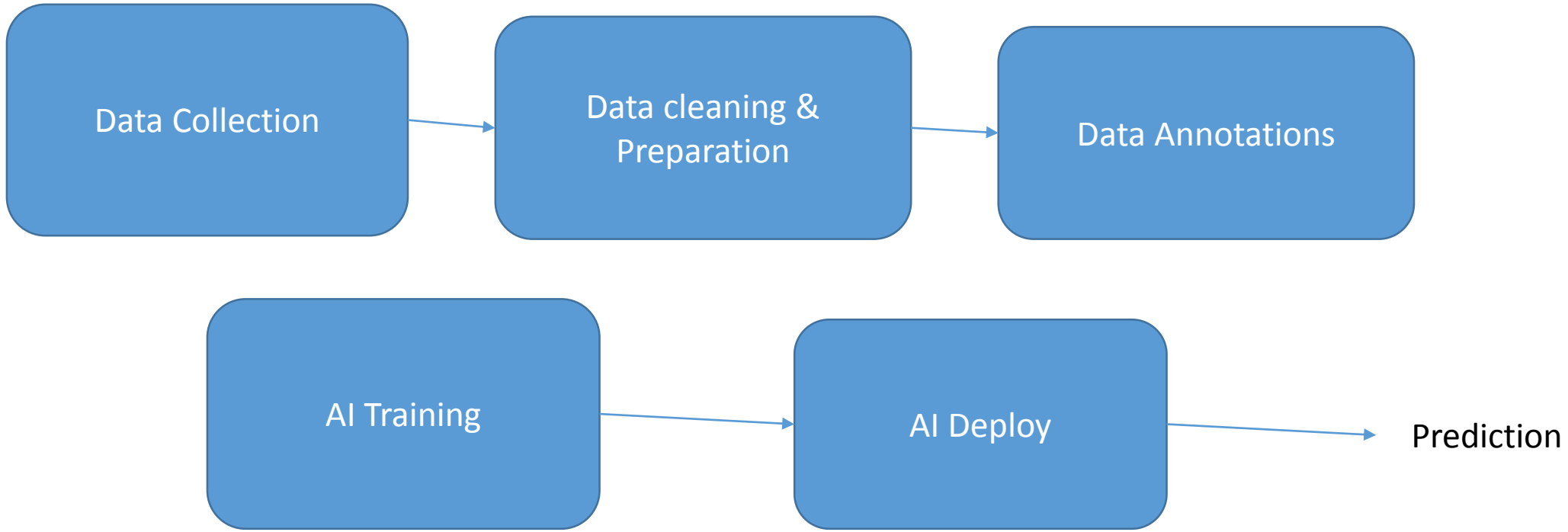
Vision

Challenges

- PACMAN goals
  - **High performance: accuracy**
  - **Low computational time**

AI Pipeline

- AI Pipeline



What is new in AI?

- Google AlphaGo AI beats GO champion
- Using AI to save energy

Overall

- Two-day conference
- Speakers
  - 50 speakers for data science summit
  - 20 speakers for startup session
- State of the art in data science and AI
- Some good ideas for our work

Companies

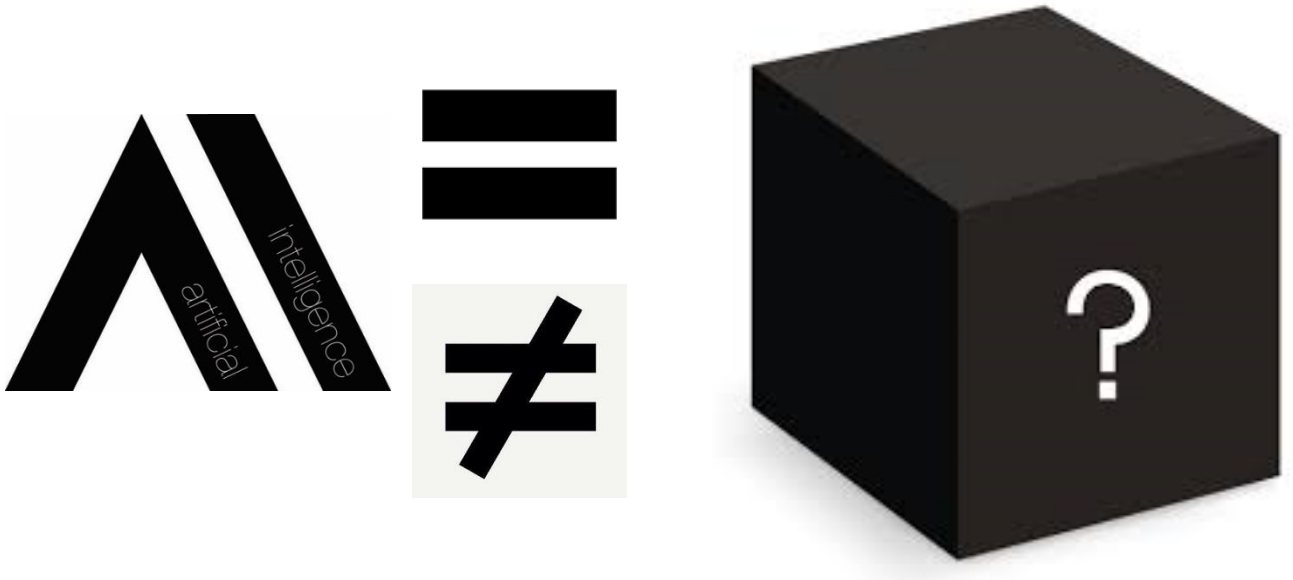
- Pinterest
- Pandora: Music Recommendation
- Concur
- Microsoft
- …

AI Transparency

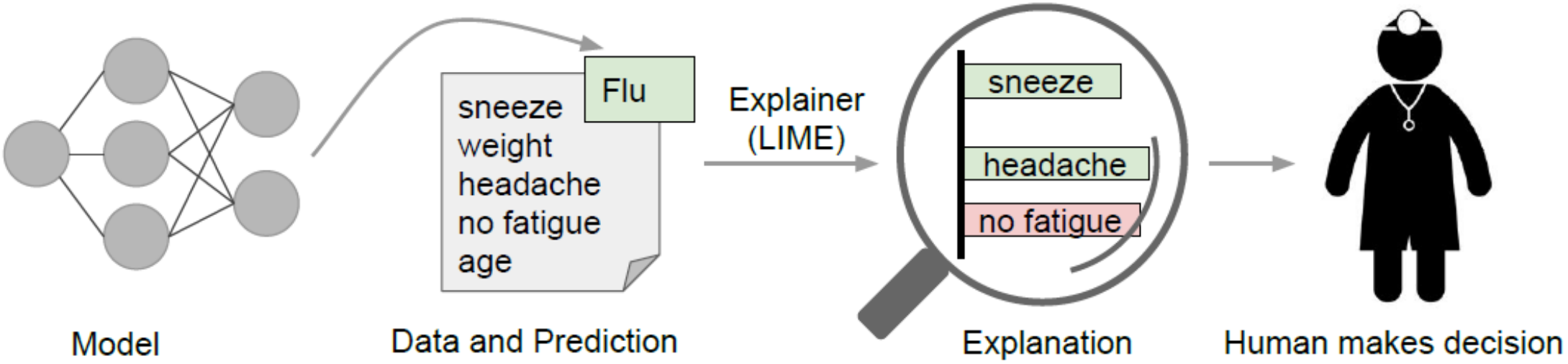
- Keynote: *Carlos Guestrin, Associate Professor, University of Washington*

Why Should I Trust You?

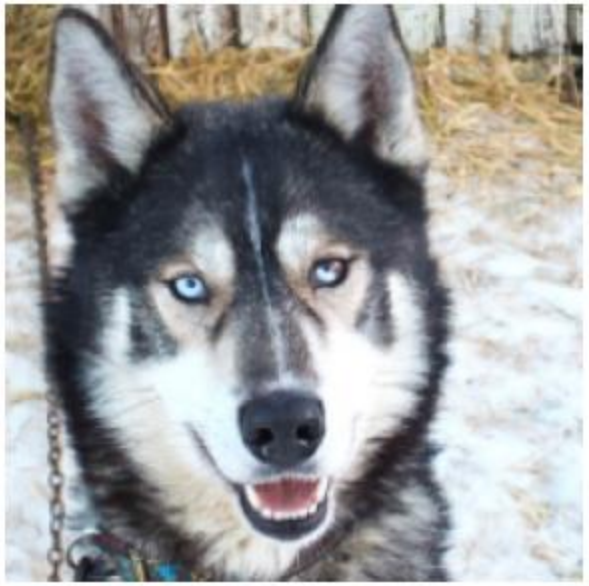
- Transparency
  - Trusting a prediction
  - Trusting a model



AI Transparency



Transparency Example



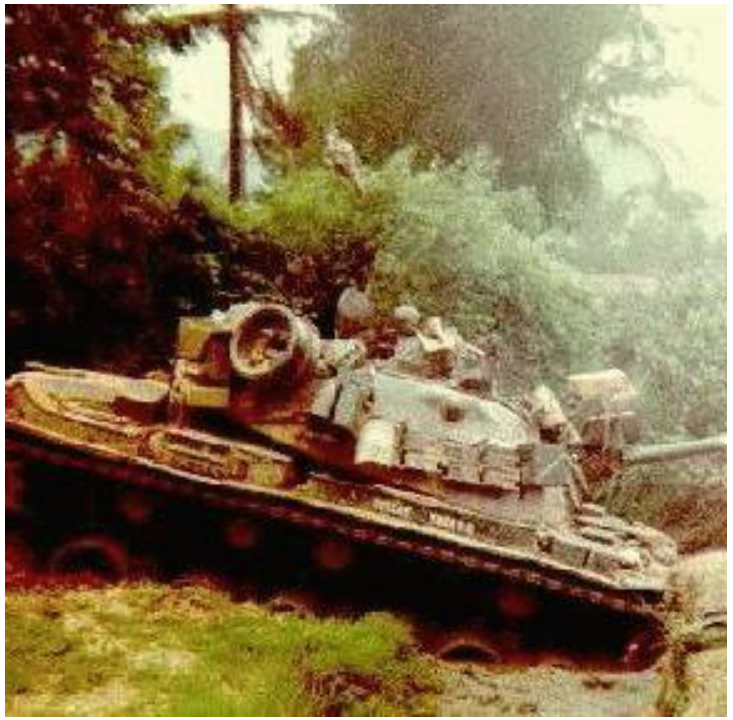
(a) Husky classified as wolf

(b) Explanation

Figure 11: Raw data and explanation of a bad model's prediction in the "Husky vs Wolf" task.

Lazy AI!

- Using of the color of the sky



Human vs. Machine

- Do Humans and AI Look at the Same Regions?

Human Attention

- Question: How many players are visible in the image?

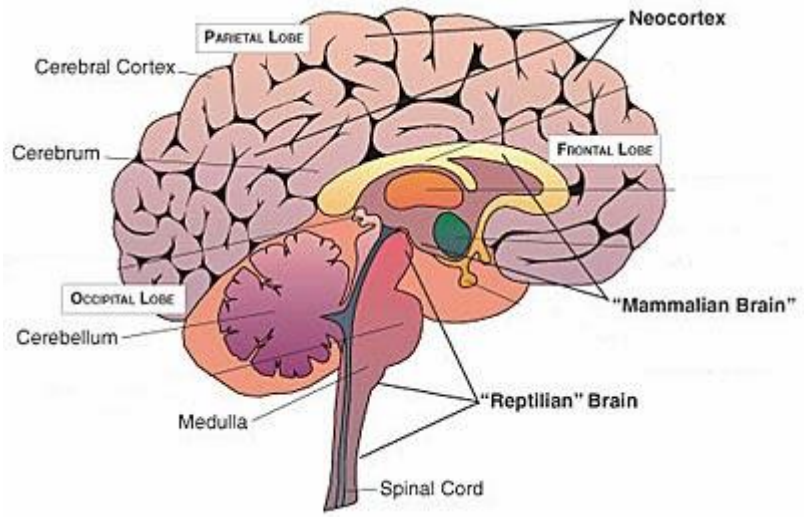
Human vs Machine

- What are they doing?

- Current AI models do not seem to be looking at the same regions as humans.

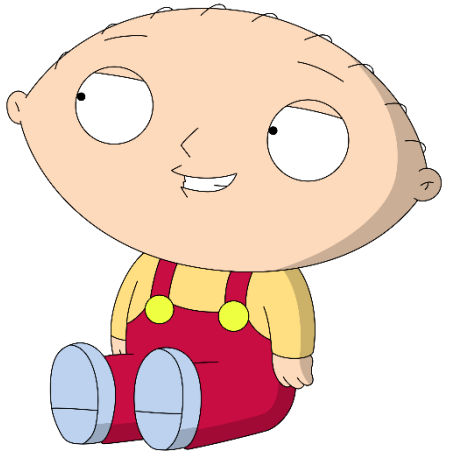
Talks

- Numenta
- The path to machine intelligence
- Neocortex
- Unsupervised
- Unlabeled data



AI Paradigms

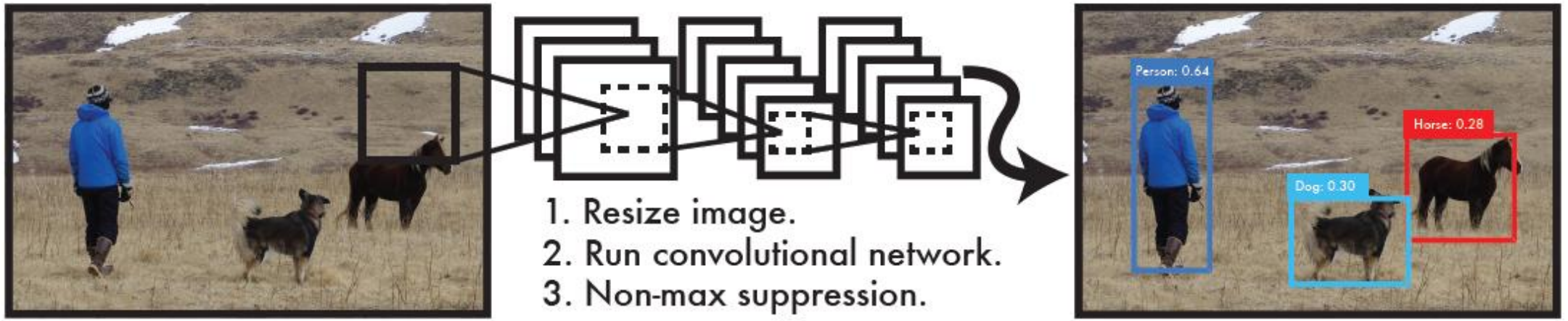
Supervised	Unsupervised
Require lots of examples with labels Most of current AI	Less examples, no/less labels Future AI



Talks

- Kaggle: Data science competitions
- Anthony Goldbloom*, co-founder: Kaggle*
- Dominating methods
- New Features

Real-Time Object Detection

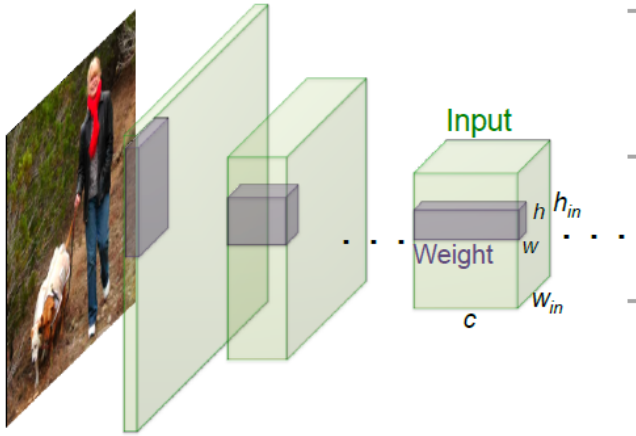


Real time Demo

- http://pjreddie.com/darknet/yolo/
- 30 frames/sec

Image Classification Using Binary Weights

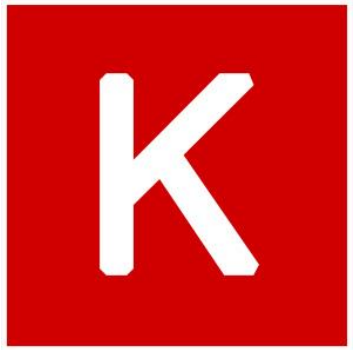
- 3.04342423*2.0242352423



	Network Variations	Operations used in Convolution	Memory Saving (Inference)	Time Saving on CPU (Inference)	Accuracy on ImageNet (AlexNet)
Standard Convolution	Real-Value Inputs 0.11 -0.21 ... -0.34 / -0.25 0.61 ... 0.52; Real-Value Weights 0.12 -1.2 ... 0.41 / -0.2 0.5 ... 0.68	+ , − , ×	1x	1x	%56.7
Binary Weight	Real-Value Inputs 0.11 -0.21 ... -0.34 / -0.25 0.61 ... 0.52; Binary Weights 1 -1 ... 1 / -1 1 ... 1	+ , −	~32x	~2x	%53.8
BinaryWeight Binary Input (**XNOR-Net**)	Binary Inputs 1 -1 ... -1 / -1 1 ... 1; Binary Weights 1 -1 ... 1 / -1 1 ... 1	XNOR , bitcount	~32x	~58x	%44.2

Other Talks

- *Jeff Dean, Google*
- *Tensorflow*, Deep Learning Library



Startups

- Domino data lab
- Github for data science
- More iterations. Better collaboration. Greater return on data.

Startups

- Crowdflower
- Machine Learning Data Pipeline
  - Data collection
  - Annotation
  - Cleaning

Startups

- Spare 5
- Training data for machine learning
- Clean labeled data at scale
- Datasets to train and test machine learning models.

spare 5

Looking Ahead

- We have way to go

Looking Ahead

“What if a cure for an intractable cancer is hidden within the tedious reports on thousands of clinical studies? In 20 years’ time, AI will be able to read — and more importantly, understand — scientific text. These AI readers will be able to ***connect the dots*** between disparate studies to identify novel hypotheses and to suggest experiments which would otherwise be missed. AI-based discovery engines will help find the answers to science's thorniest problems.”

Oren Etzioni, professor of Computer Science and CEO of the Allen Institute for Artificial Intelligence

Takeaways ☺

- Crowded: 1300 attendees
- Lectures location: quite workout
- Lots of speakers
- Progress in AI field
- A few good ideas for PACMAN