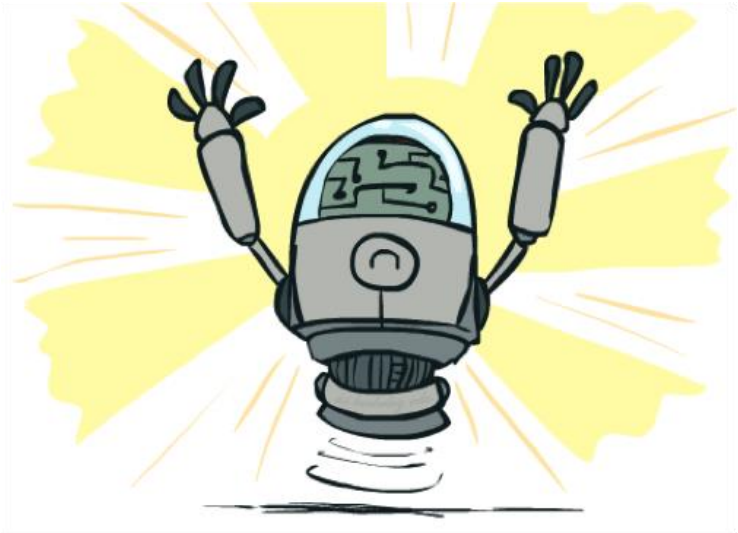
COMS W4701: Artificial Intelligence

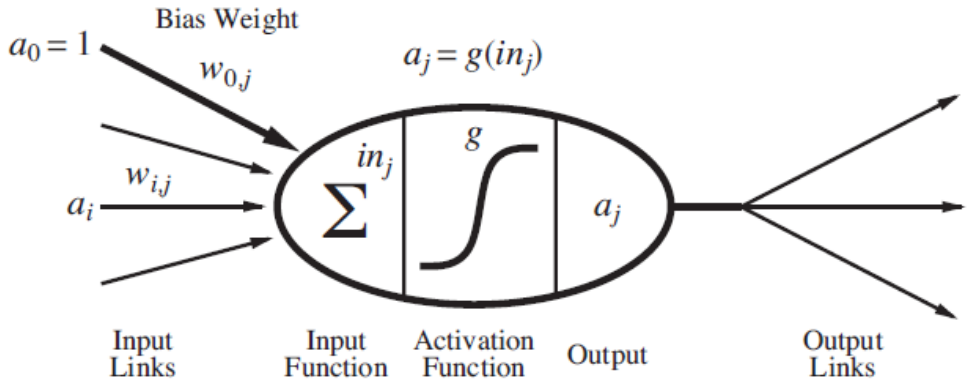
Lecture 26: Deep Learning and Future Outlook



Instructor: Tony Dear

History of Neural Networks

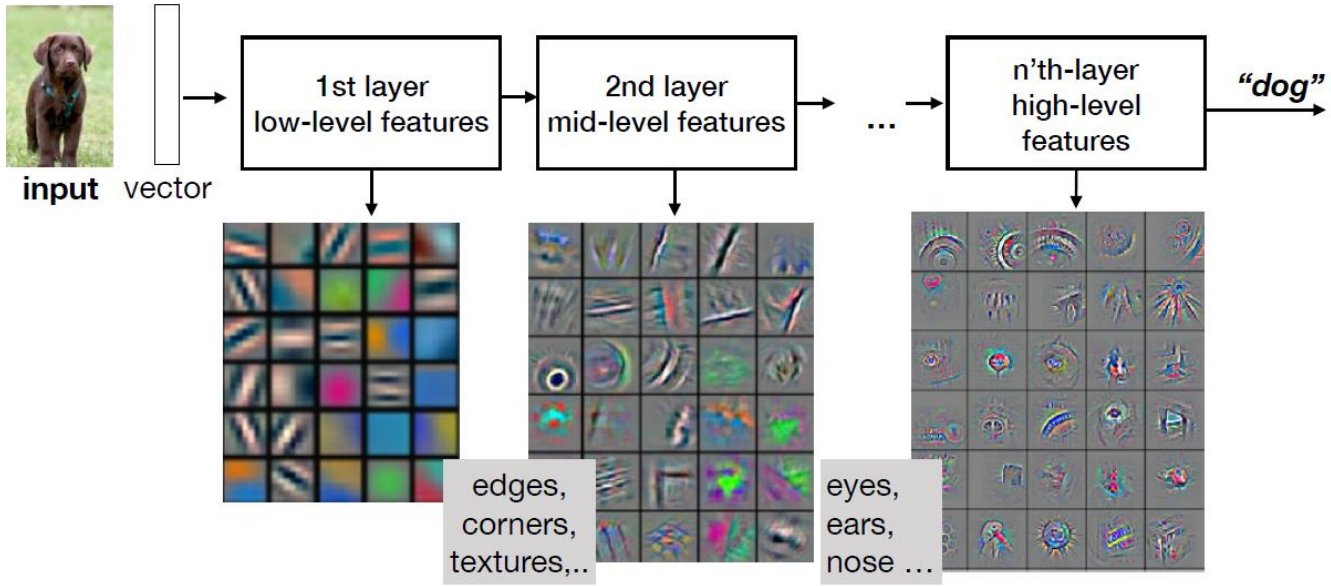
- Originally proposed to model biological neural networks (e.g., in the brain)
- Idea has been around since 1940s (McCulloch and Pitts); perceptron proposed in 1950s, backpropagation in 1970s
- However, training large networks really became feasible in the 2000s with hardware advances
- GPUs, distributed computing, etc.



Deep Neural Networks

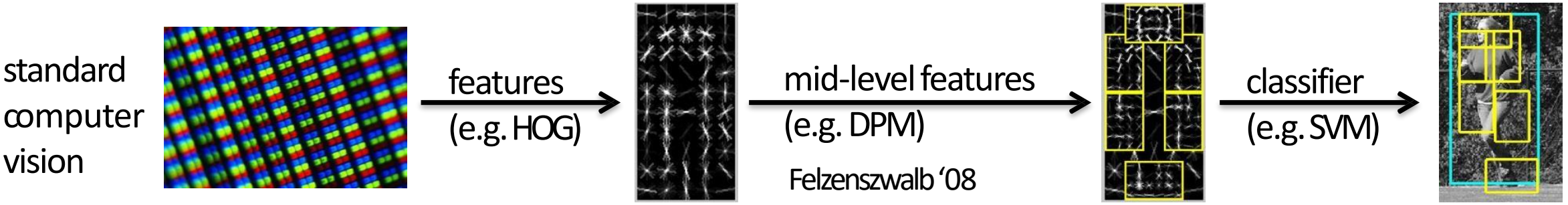
- One hidden layer is sufficient for representing any function, but…
  - We may require a lot of neurons!
  - Still have to do a lot of work with feature extraction

- Idea: Multiple hidden layers can represent a hierarchy of features and functions, from low to high level

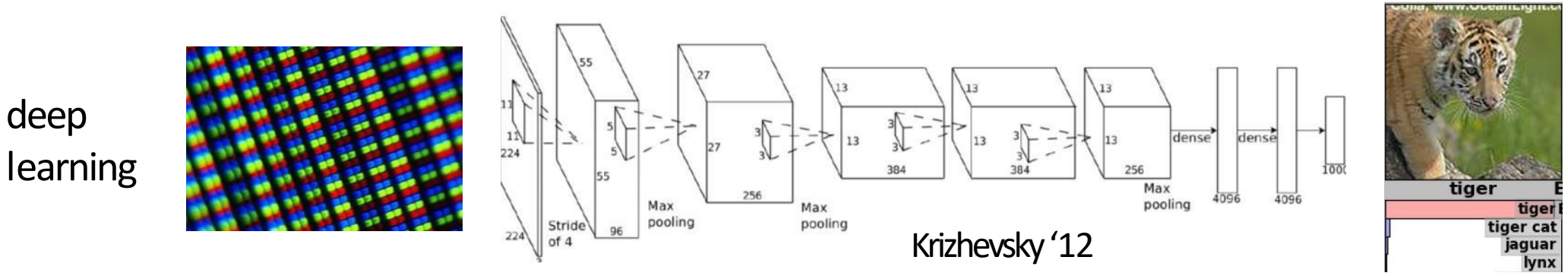


Deep Learning and Computer Vision

- Traditional classification in computer vision: Lots of different methods to extract diverse features, followed by standard ML algorithms



- With deep learning:

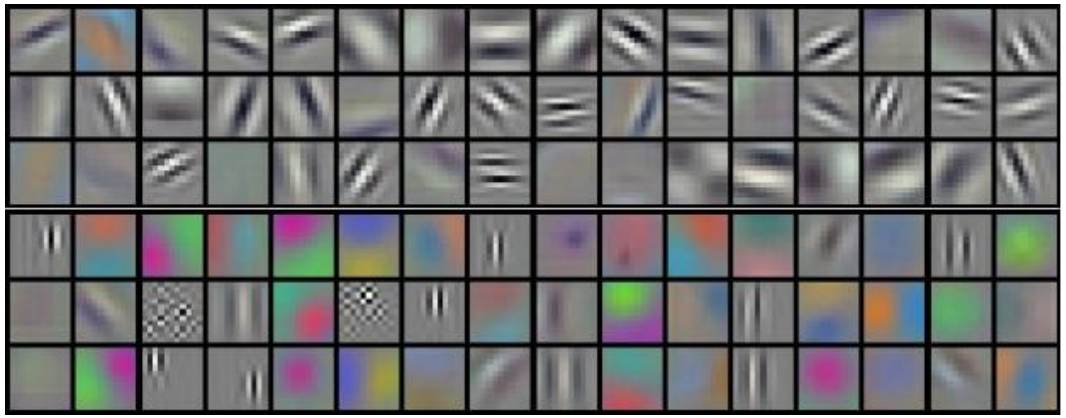


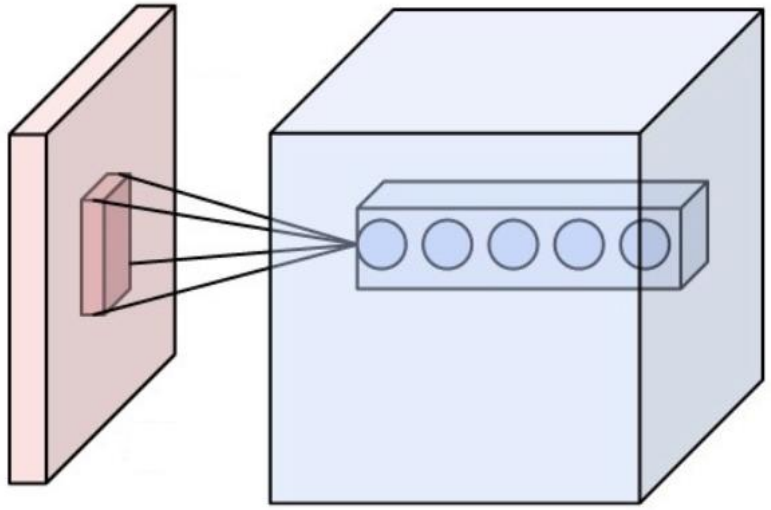
S. Levine

Convolutional Neural Nets (CNNs)

- Large images with many pixels will require lots of neurons and connections
- Idea: Images typically contain **spatial structure and transformations**

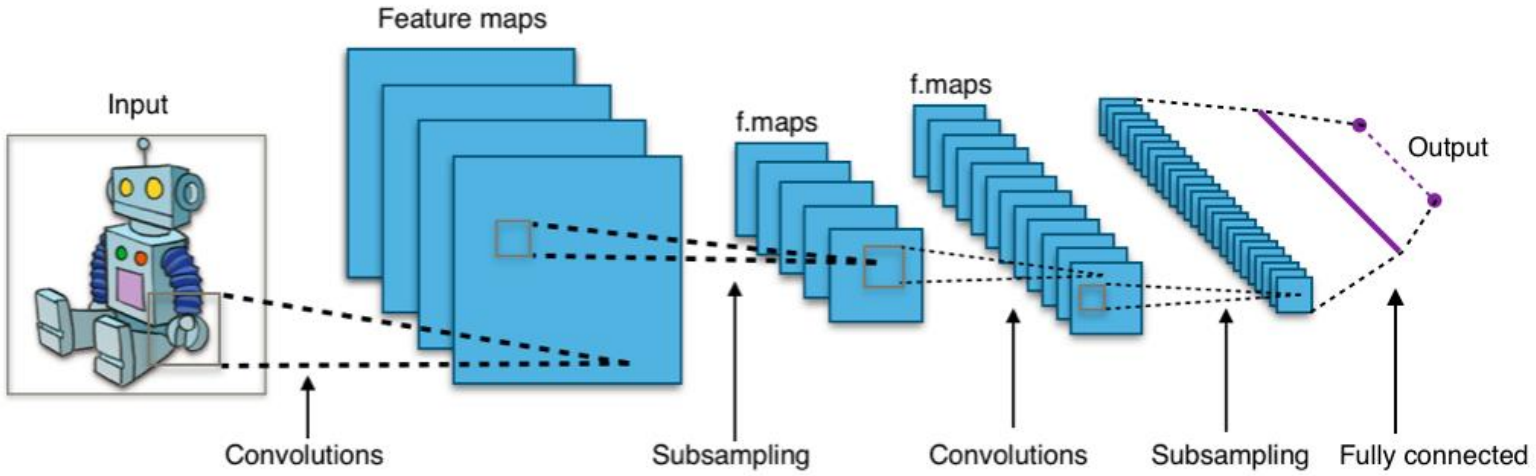
- **Convolution**: Slide filters across local groups of pixels to extract features
  - Computed as dot product between filter and pixels
  - Filter activates when feature is spatially present





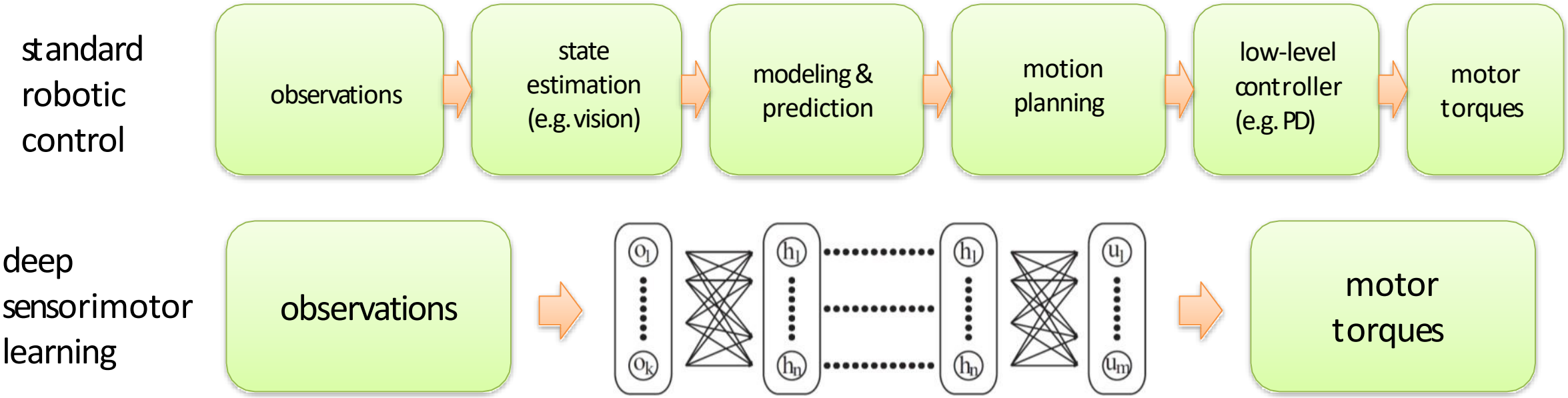
Convolutional Neural Nets (CNNs)

- We can stack convolutional layers together to extract different features
- Also important to have **pooling** (down-sampling) layers to downsize feature maps
  - Relative locations of features more important than exact locations
  - Reduce spatial complexity and overfitting
- Output layer is fully connected, while hidden layers are locally connected



Deep Learning for Robot Control

- Deep learning is redefining the extent of the "functions" that we typically learn for different tasks, e.g. robot control
- Instead of using separate traditional algorithms for subtasks, DL often combines multiple subtasks together into a unified pipeline



Other ML Tasks

- Neural networks are very flexible—can go beyond supervised learning!

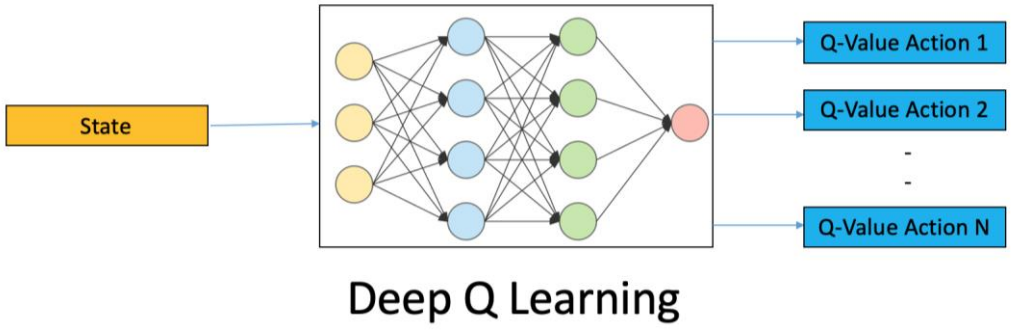
- **Unsupervised learning** and clustering
- Autoencoders: Learn codings for data compression (e.g., by ignoring noise)
- Self-organizing maps for dimensionality reduction of input sapce

- **Reinforcement learning**
- In many problems, we cannot explicitly store and update Q-values
- Neural networks can be used and updated as **Q-functions**

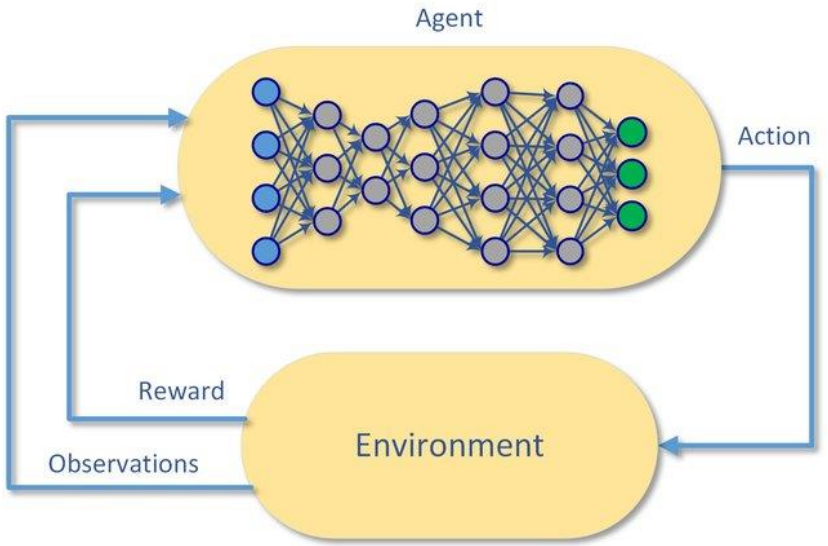
Deep Reinforcement Learning

- Recall basic idea of RL: Given state (and maybe action), compute the *value* of the state(-action) and ultimately use the values to find a policy

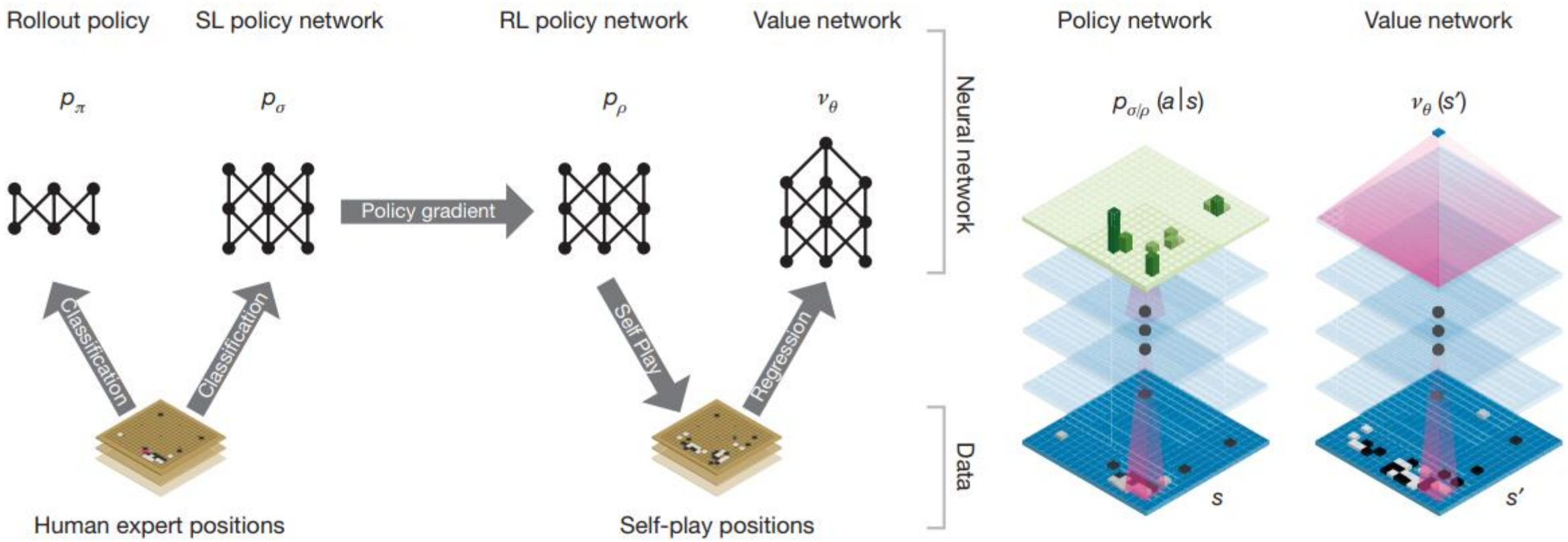
- These are just functions: State(-action) -> values, state -> policy
- Functions can be learned and approximated by neural networks!



Deep Q Learning

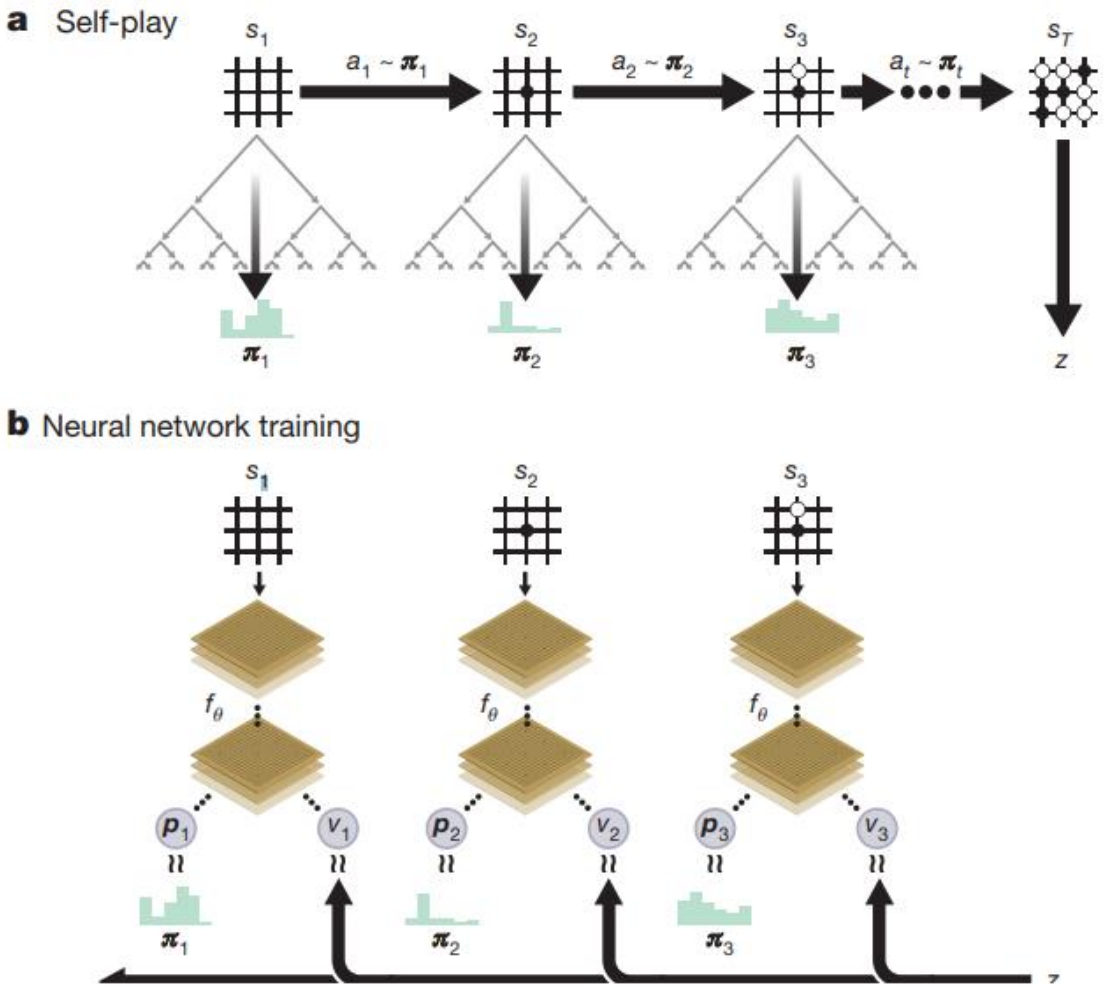


Case Study: AlphaGo



Silver et al., “Mastering the game of Go with deep neural networks and tree search”, *Nature*, 2016.

Case Study: AlphaGo Zero



Silver et al., “Mastering the game of Go without human knowledge”, *Nature*, 2017.

Case Study: AlphaZero

Silver et al., “A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play”, *Science*, 2018.

Now that you've taken 4701, whose worldview of AI's future do you agree with more?

Jack Ma ("utopian")

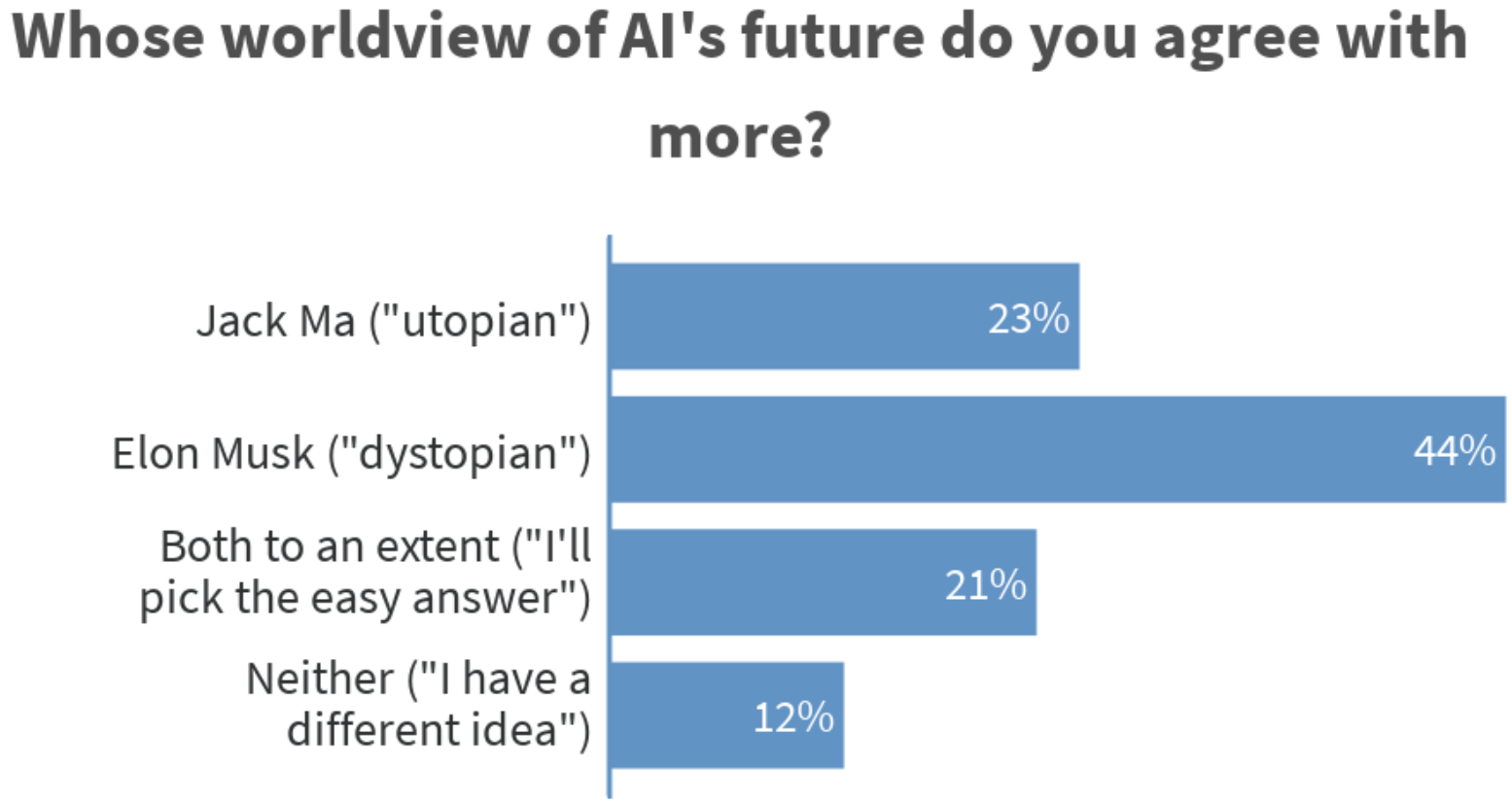
Elon Musk ("dystopian")

Both to an extent ("I'll pick the easy answer")

Neither ("I have a different idea")

Polling Results from First Class

When poll is active, respond at **PollEv.com/coms4701**
Text **COMS4701** to **22333** once to join



Philosophical Questions

- Weak AI vs strong AI
- **Weak AI**: Machines act as if they were intelligent (esp for specific tasks)

- **Strong AI**: Machines can actually think in the same sense as humans
  - What does it mean to think and have cognition?
  - Can machines ever be conscious or self-aware?

- Does it even matter?
  - Most people are happy as long as their system meets their goals

Social Questions

- Automation and job security
  - Can this be compared to previous advances in human ingenuity?
- Role of humans in an automated world
  - Do we fully incorporate AI into our lives? What about when it surpasses us?
- Unintentional and intentionally bad consequences of AI systems
  - E.g., human prejudice and biases in AI
  - Hostile entities and governments

Logistics: Final Exam

- Modified office hour schedule next week—check calendar
- Review session: Thurs Dec 12 at 5:30pm
- Practice final exam posted, solutions will be posted after review session

- What to expect: 150 minutes, 7 questions, “cumulative” material
- TWO double-sided sheets of notes allowed
- Please arrive on time—we will be doing ID check!

- 1 or 2 problems will be almost exact copies from HW1-3 and midterm
- Rest will cover material from HW4-7 and Lectures 9-24
- “Programming”-ish questions still fair game

Topics Covered

- Nondeterministic search: MDP formulation, value iteration, policy iteration
- Reinforcement learning: Model-based vs model-free, passive vs active, Q-learning
- Markov chains and HMMs: Forward and Viterbi algorithms, smoothing
- Bayes nets: D-separation, analytical inference, Monte Carlo methods
- Machine learning: Naïve Bayes, decision trees, perceptrons and kernels, knn
- GOOD LUCK ON YOUR FINALS!