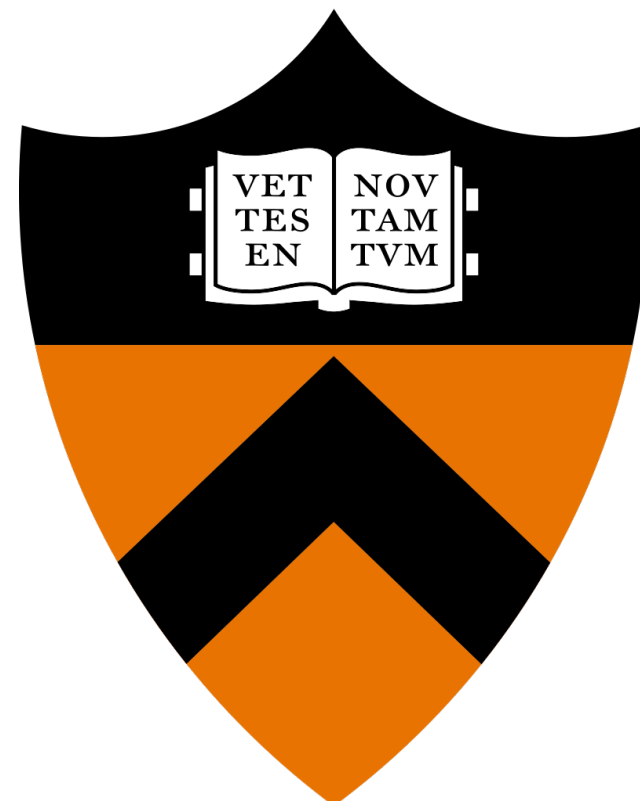


What is Machine Learning?

Ryan P. Adams
Princeton University

@ryan_p_adams
<http://lips.cs.princeton.edu>



“the study of computer programs that
improve automatically with experience”

– Tom Mitchell, 1998

What is Machine Learning?

Prediction [supervised learning]

- ▶ What movie will you like?
- ▶ What is the biological effect of this molecule?
- ▶ What will be China's 2018 GDP?

Finding Structure in Data [unsupervised learning]

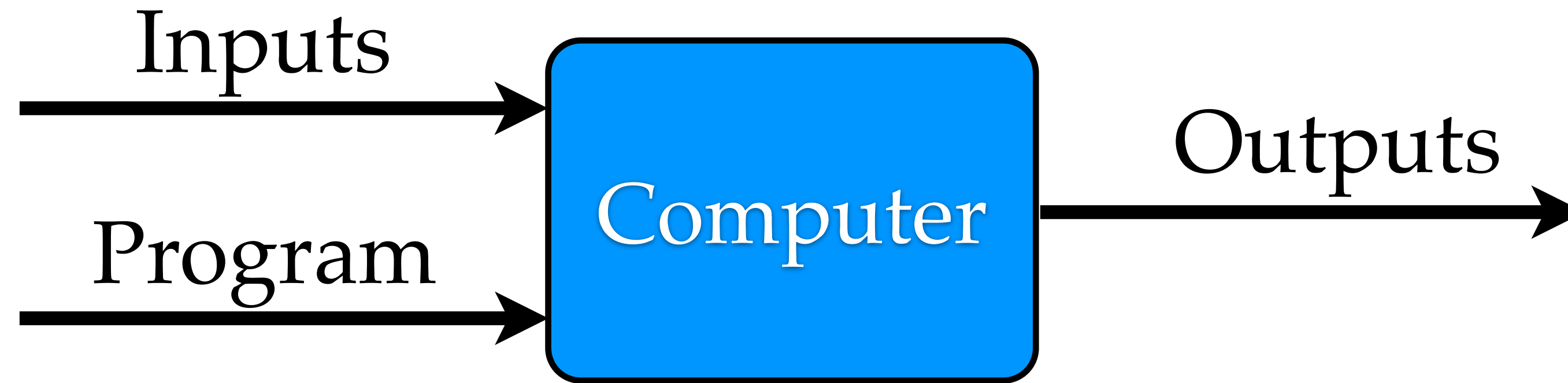
- ▶ What genes are relevant to this disease?
- ▶ Is there a low-dimensional representation of these data?
- ▶ Make me more objects like this one.

Decision Making [reinforcement learning]

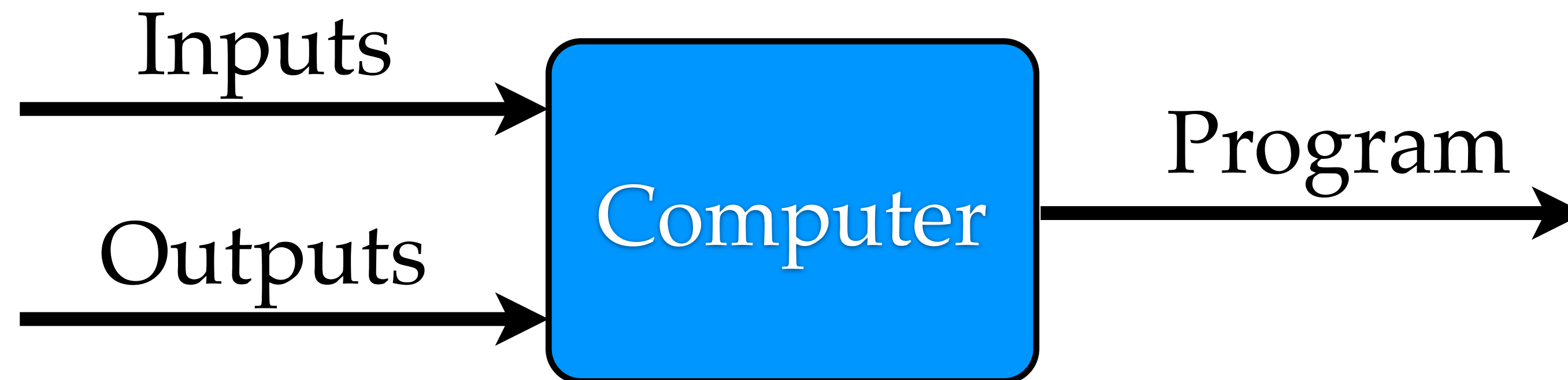
- ▶ Should we offer credit to this person?
- ▶ Should we perform this medical treatment?
- ▶ Should the car stop at this intersection?

What is Machine Learning?

Traditional Programming



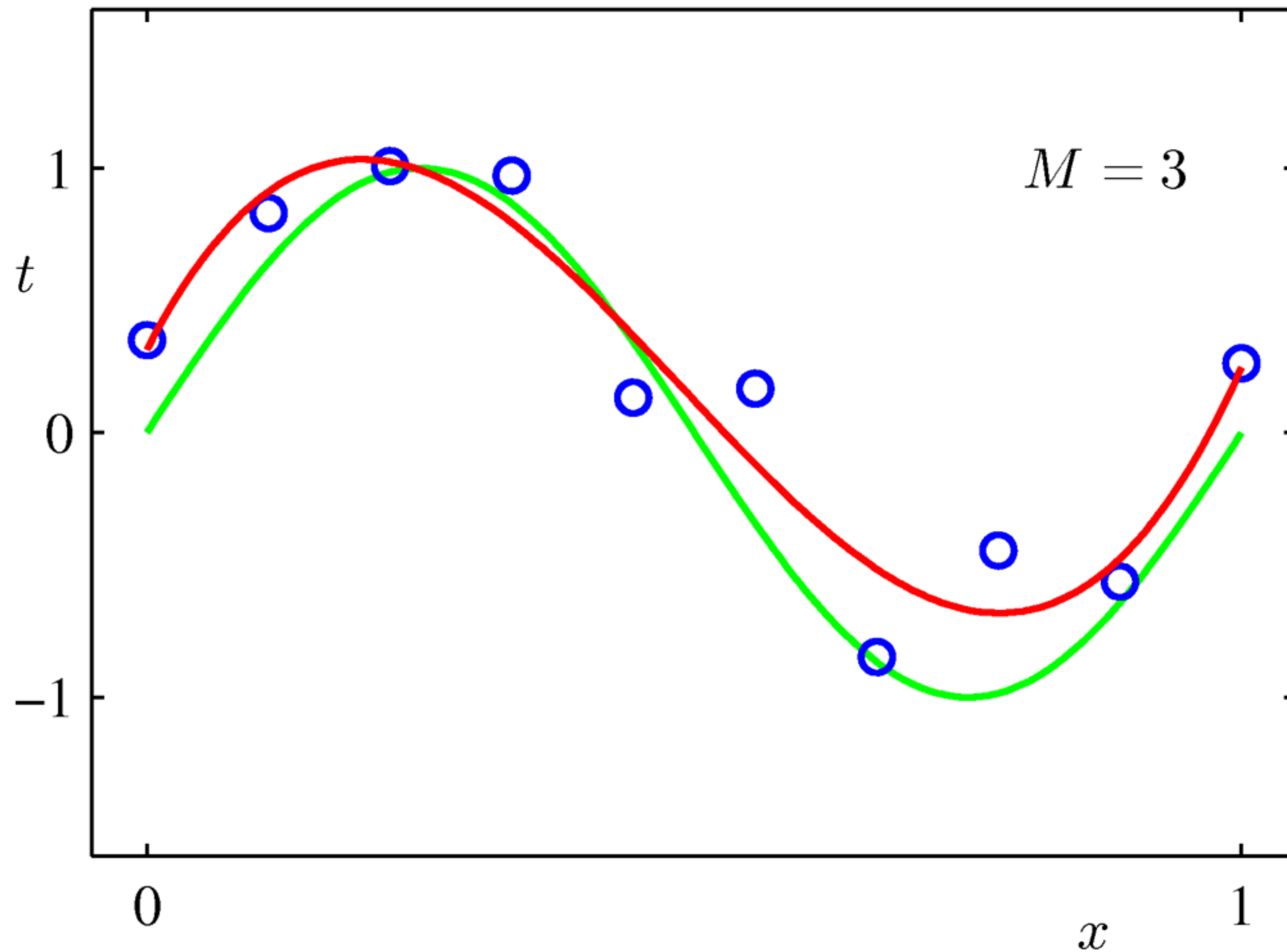
Machine Learning



Supervised Learning

Using data to build a function approximation.

Supervised Learning & Function Approximation



Supervised Learning & Function Approximation

Bad function $f\left(\text{2}\right) = 9$ $f\left(\text{50}\right) = 7$...

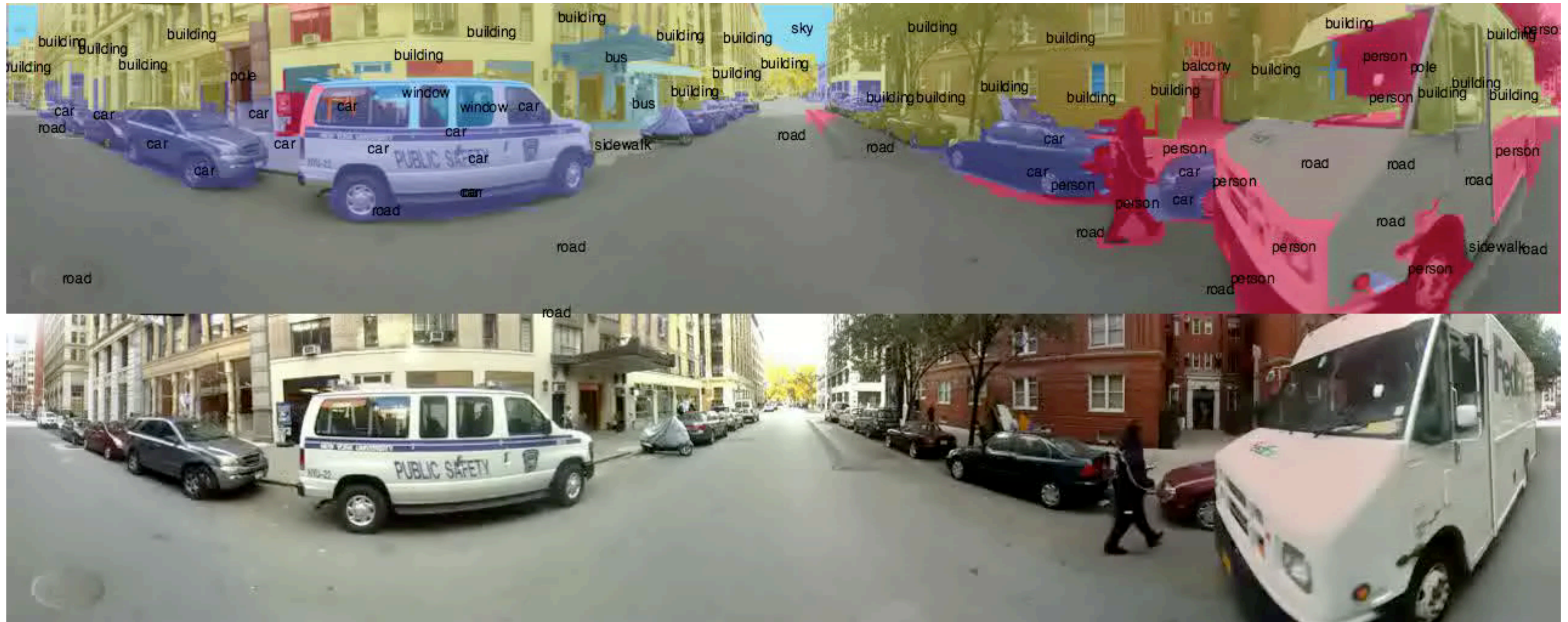
OK function $f\left(\text{3}\right) = 3$ $f\left(\text{18}\right) = 5$...

Good function $f\left(\text{18}\right) = 8$ $f\left(\text{19}\right) = 9$...

Detecting Faces



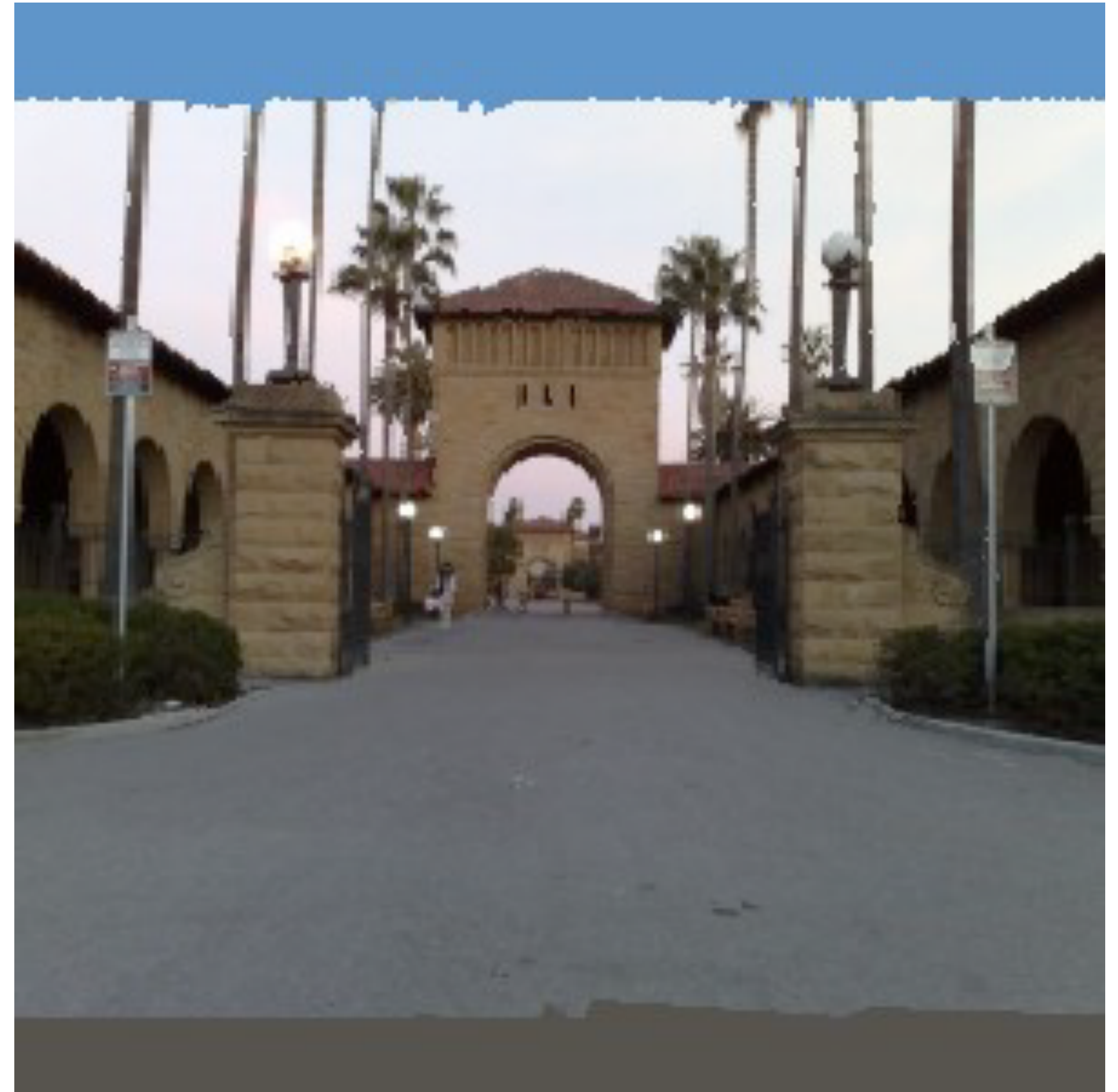
Scene Parsing



Depth Map from a Single Image



Used a neural network to predict the output of a laser depth scanner.



Supervised Learning of Natural Language

- ▶ **Input:**

- ▶ Profits soared at Boeing Co., easily topping forecasts on Wall Street, as their CEO Alan Mulally announced first quarter results.

- ▶ **Output:**

- ▶ Profits /N soared /V at /P Boeing /N Co. /N, easily /ADV topping /V forecasts /N on /P Wall /N Street /N, as /P their /POSS CEO /N Alan /N Mulally /N announced /V first /ADJ quarter /N results /N.

- ▶

Image Captioning

Describes without errors



A person riding a motorcycle on a dirt road.

Describes with minor errors



Two dogs play in the grass.

Somewhat related to the image



A skateboarder does a trick on a ramp.



A group of young people playing a game of frisbee.



Two hockey players are fighting over the puck.



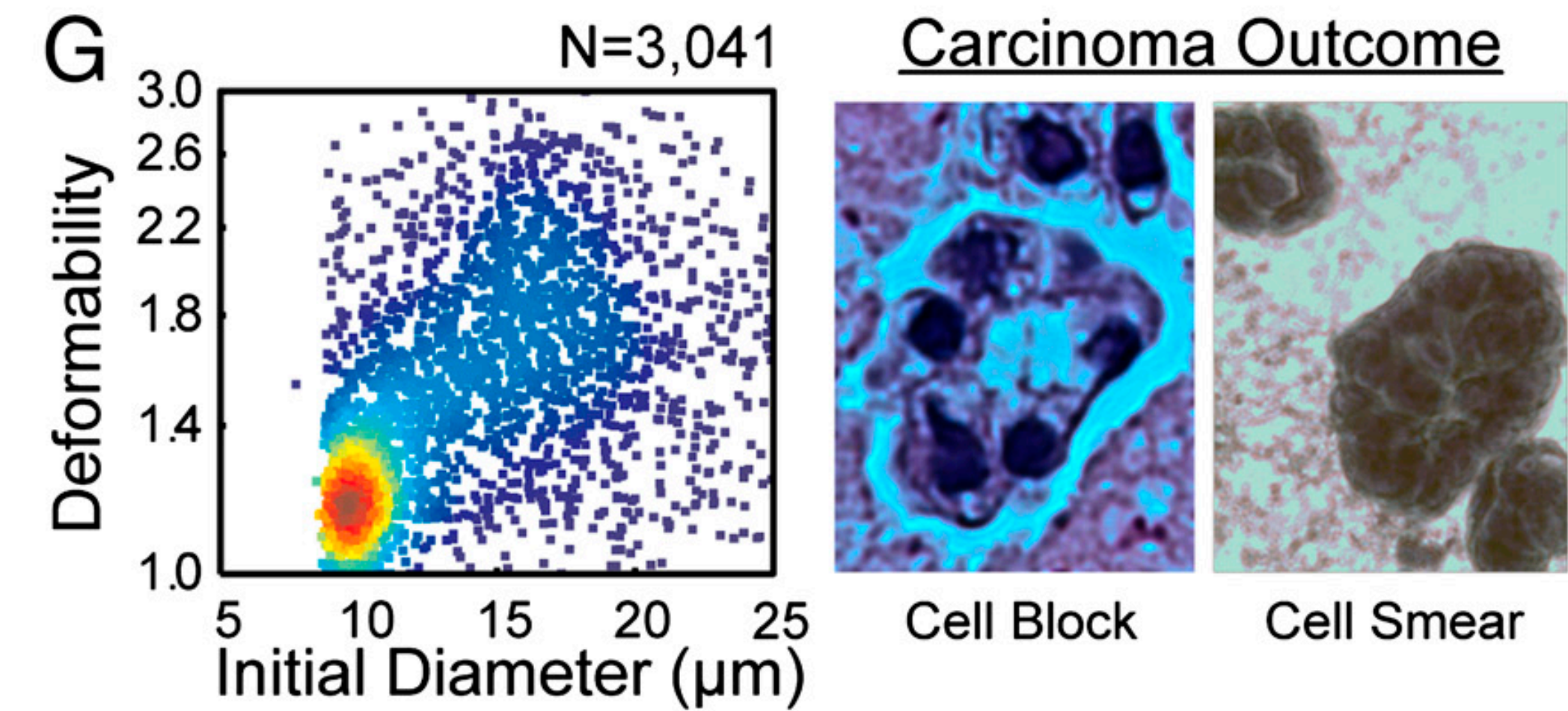
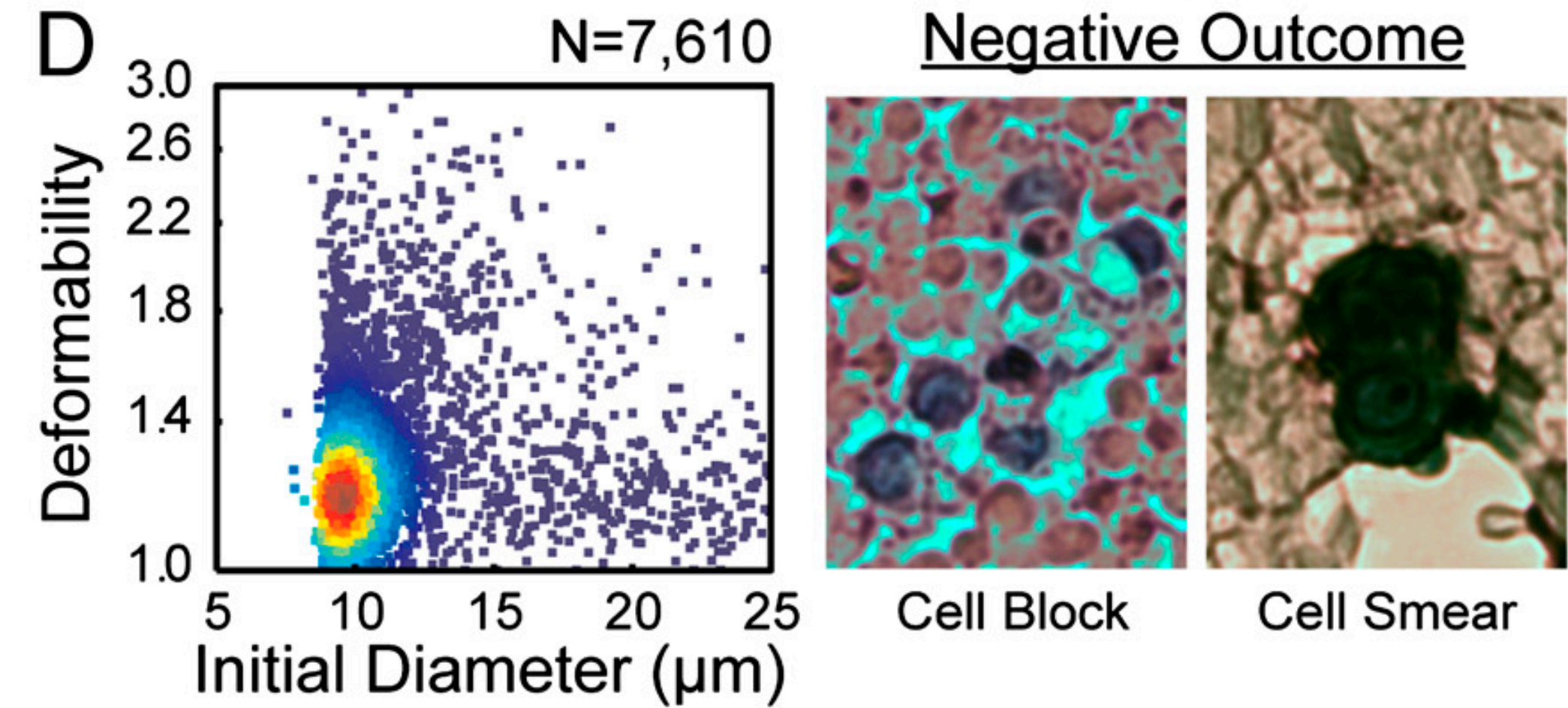
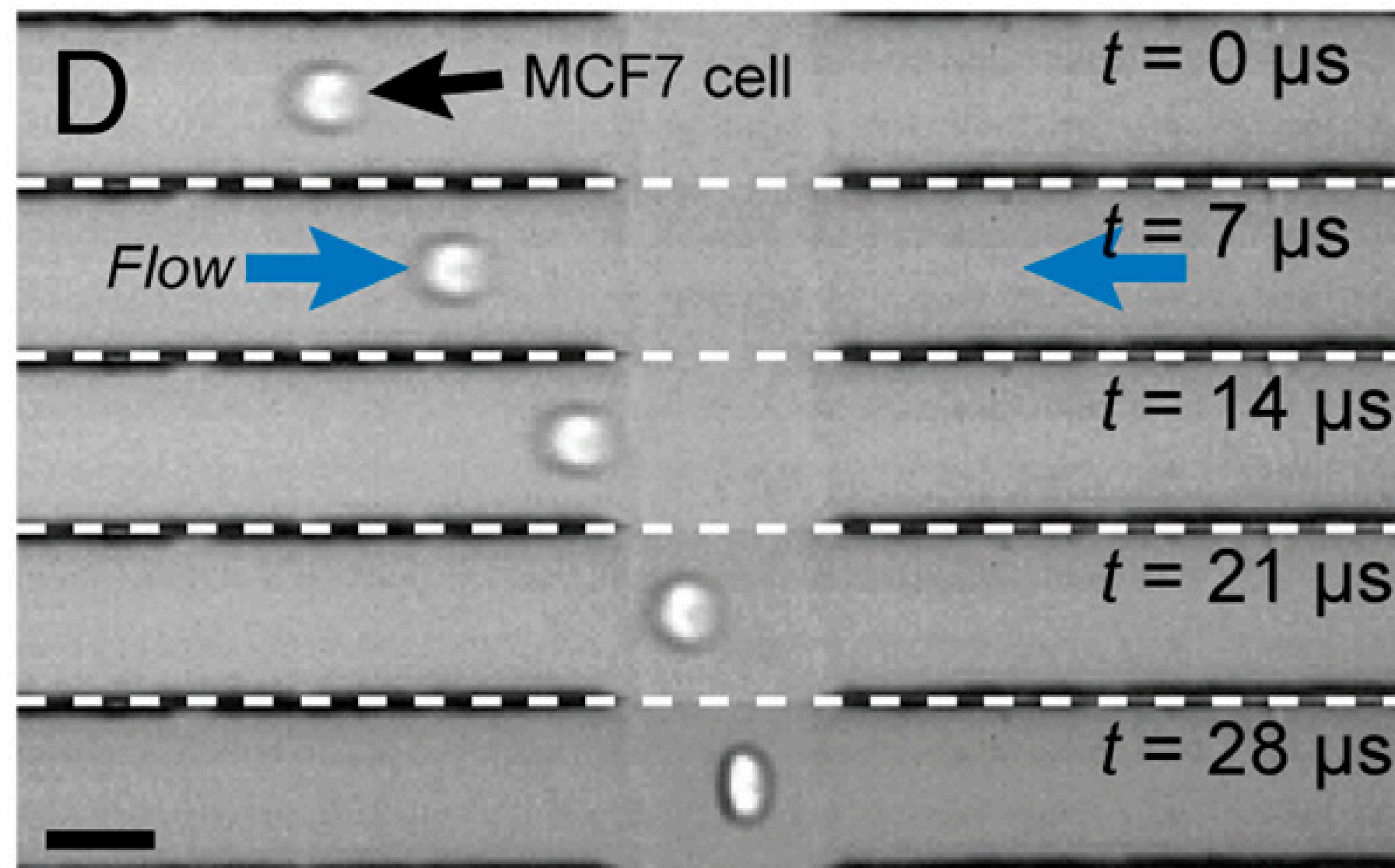
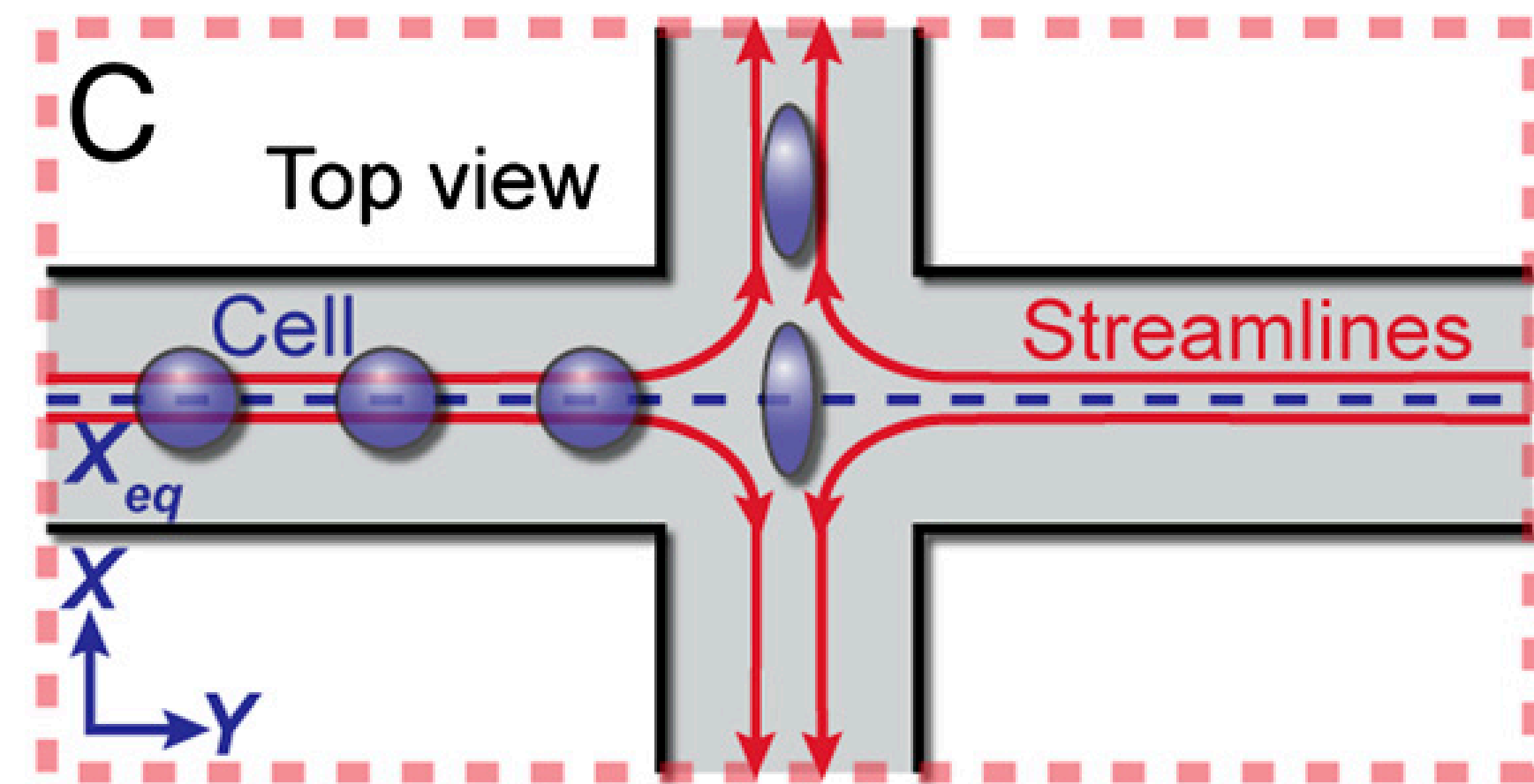
A little girl in a pink hat is blowing bubbles.

Movie Recommendations

- ▶ October 2006: Netflix released user ratings data spanning six years
 - ▶ [anon ID, date, title, year, rating]
 - ▶ Some noise added
 - ▶ 100M+ ratings
 - ▶ 480K+ users
 - ▶ 17K+ movies
- ▶ Objective: perform better than Netflix's internal algorithm in terms of RMSE on held-out test data.
- ▶ \$1M Grand Prize



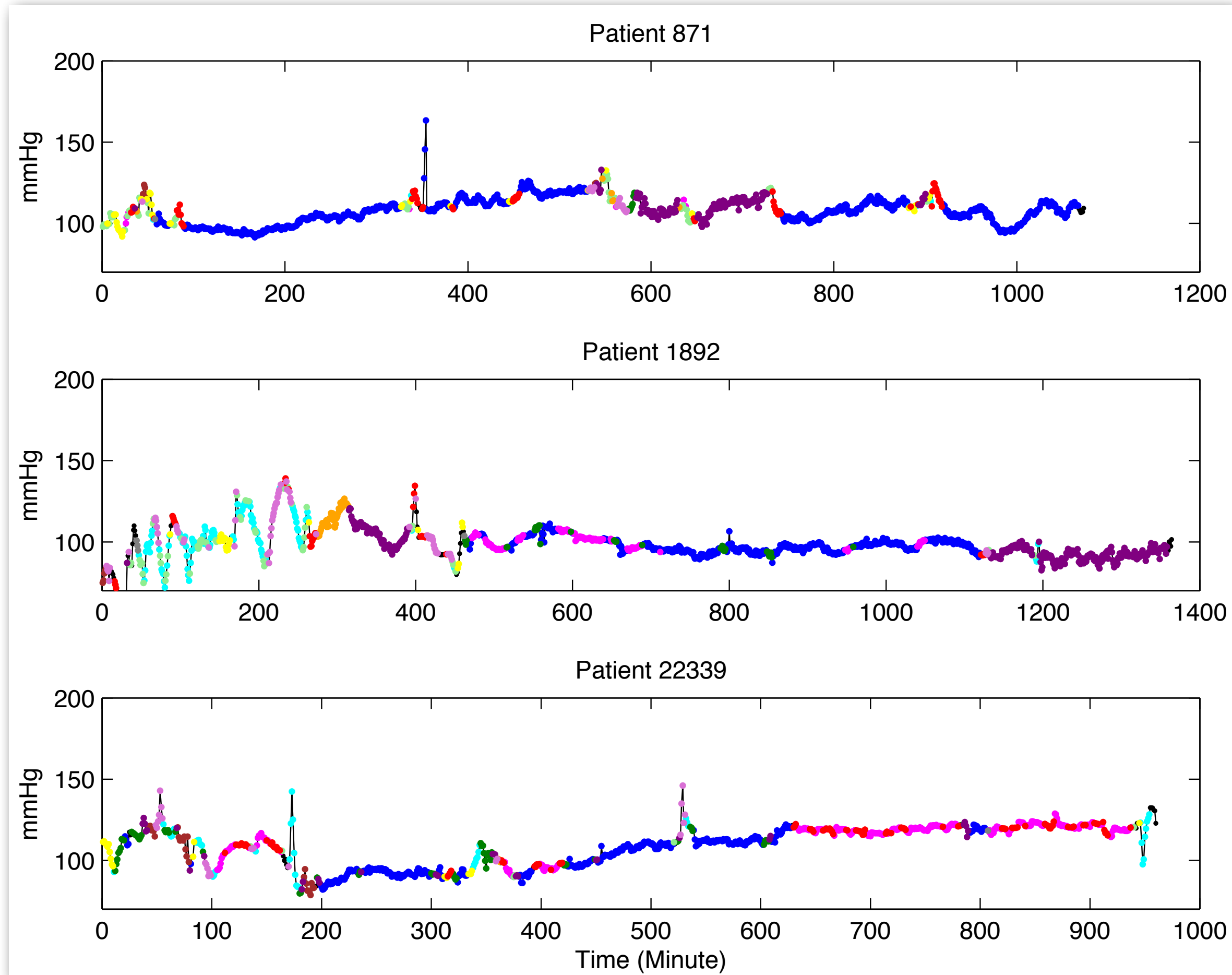
Identifying Cancer Cells



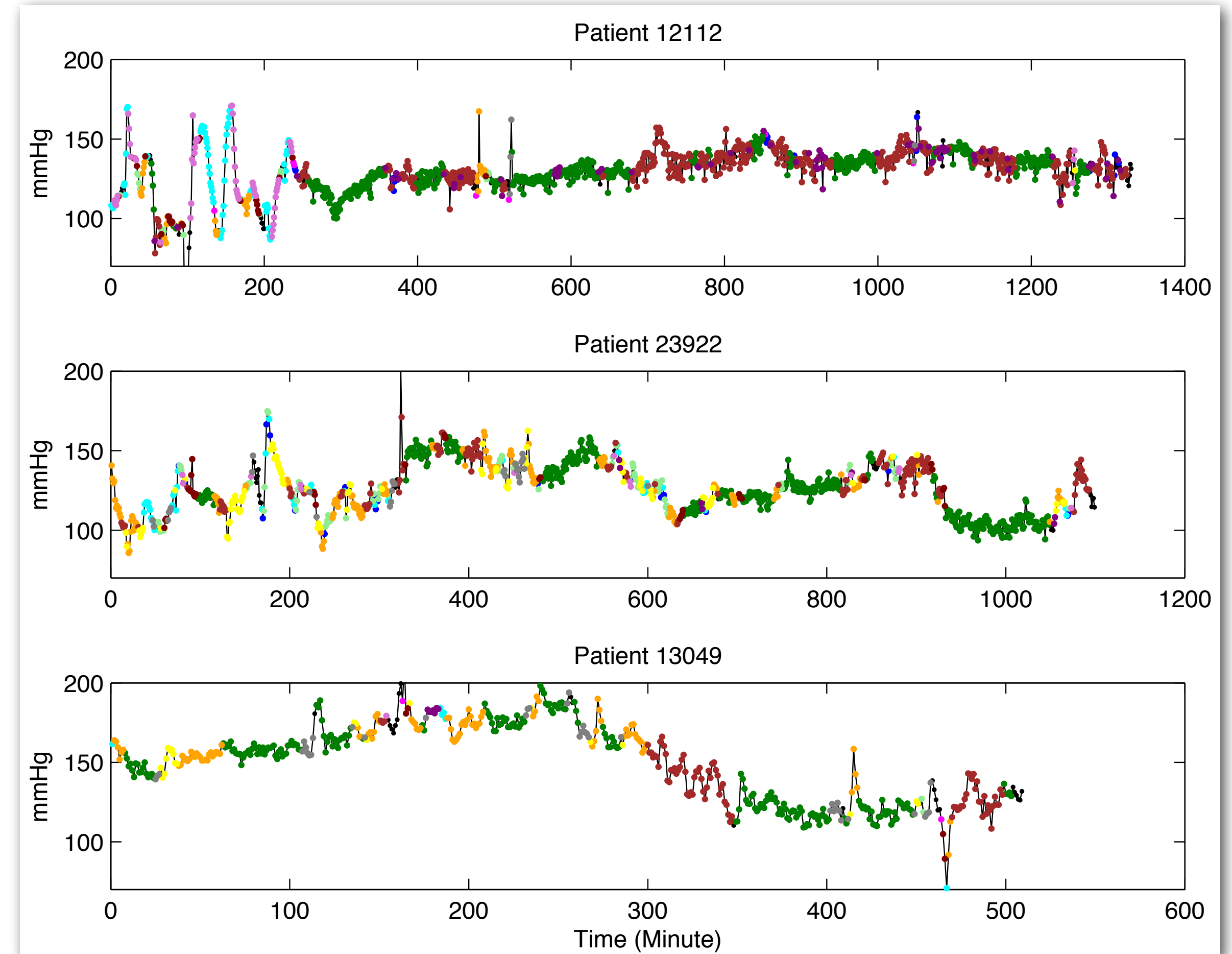
Label cancer cells from non-invasive microfluidic flow measurements.

Predicting Mortality in the ICU

Survived

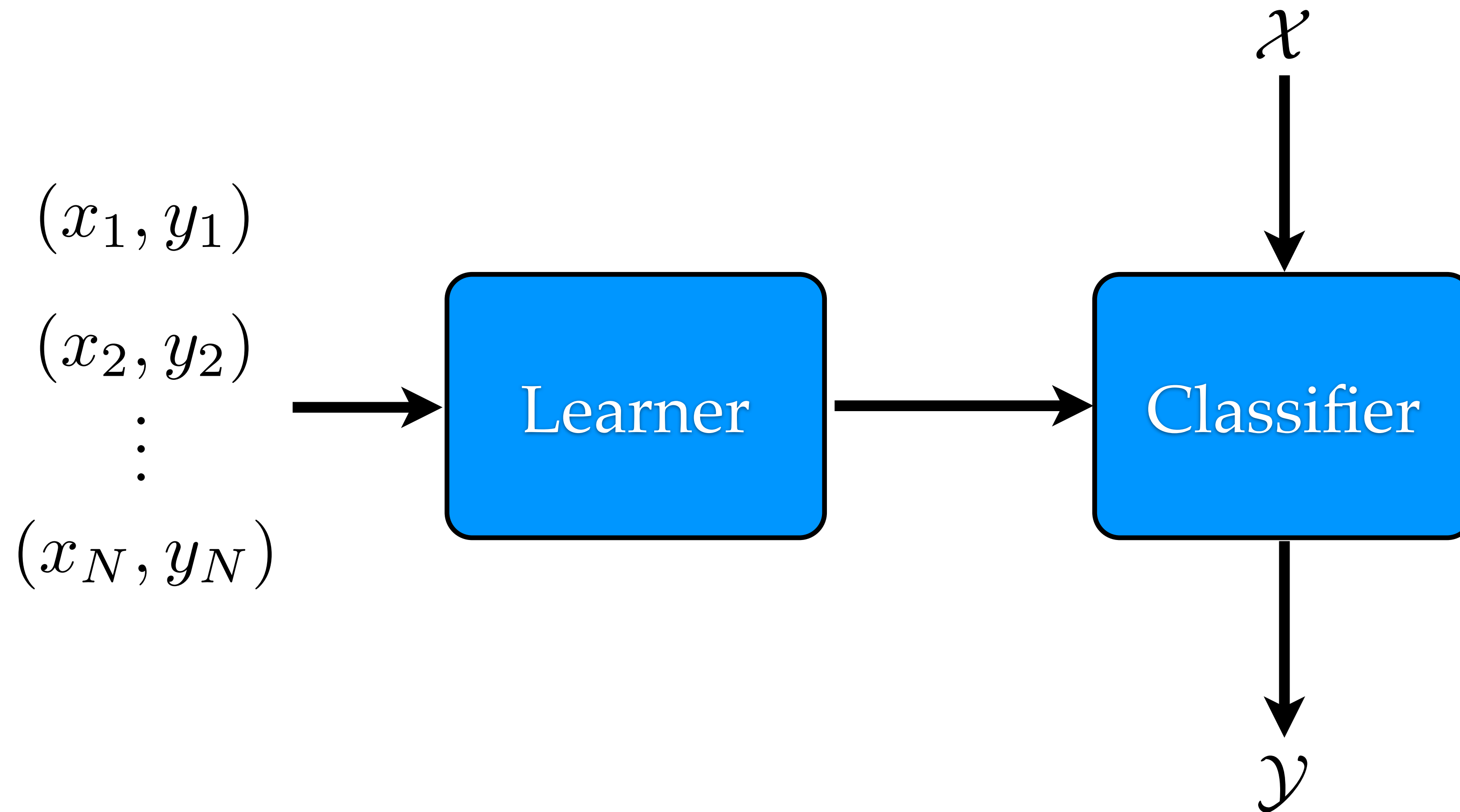


Expired



Use physiological measurements to predict whether a patient in intensive care will survive.

Supervised Learning

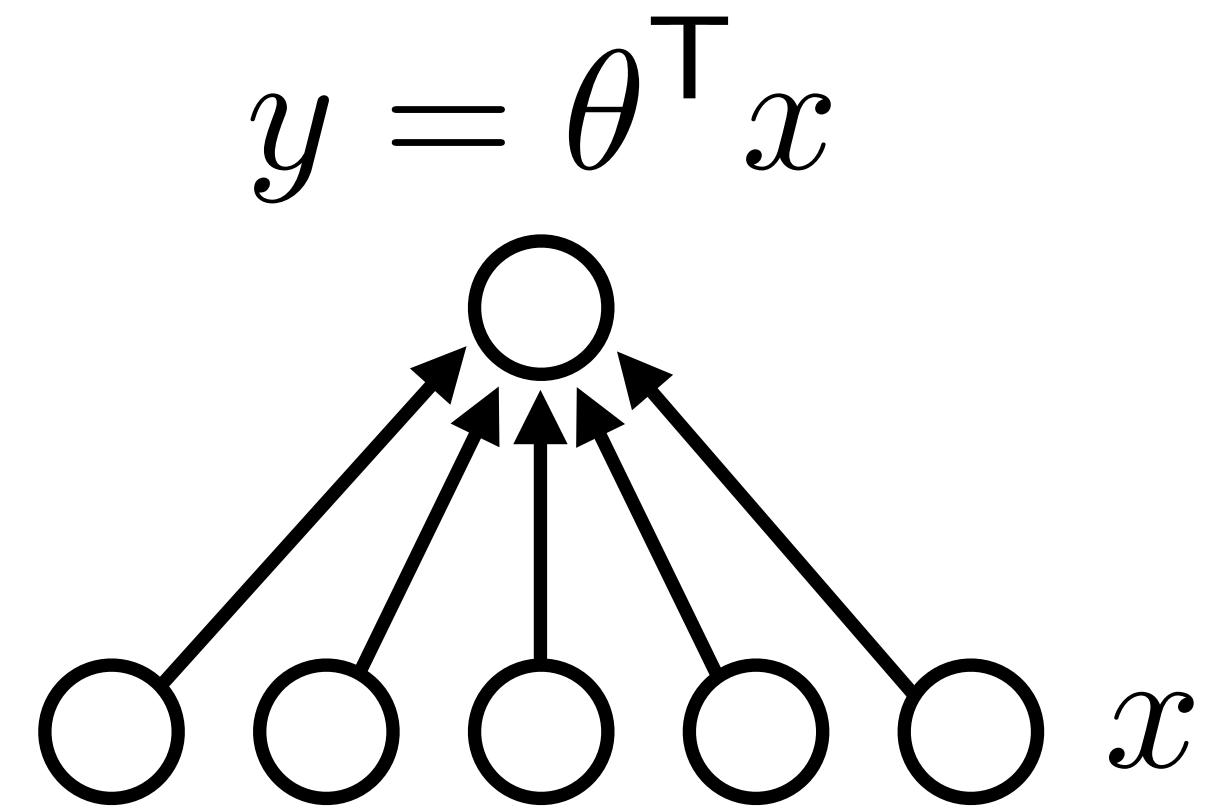


Deep Learning

It's just adaptive basis function regression.

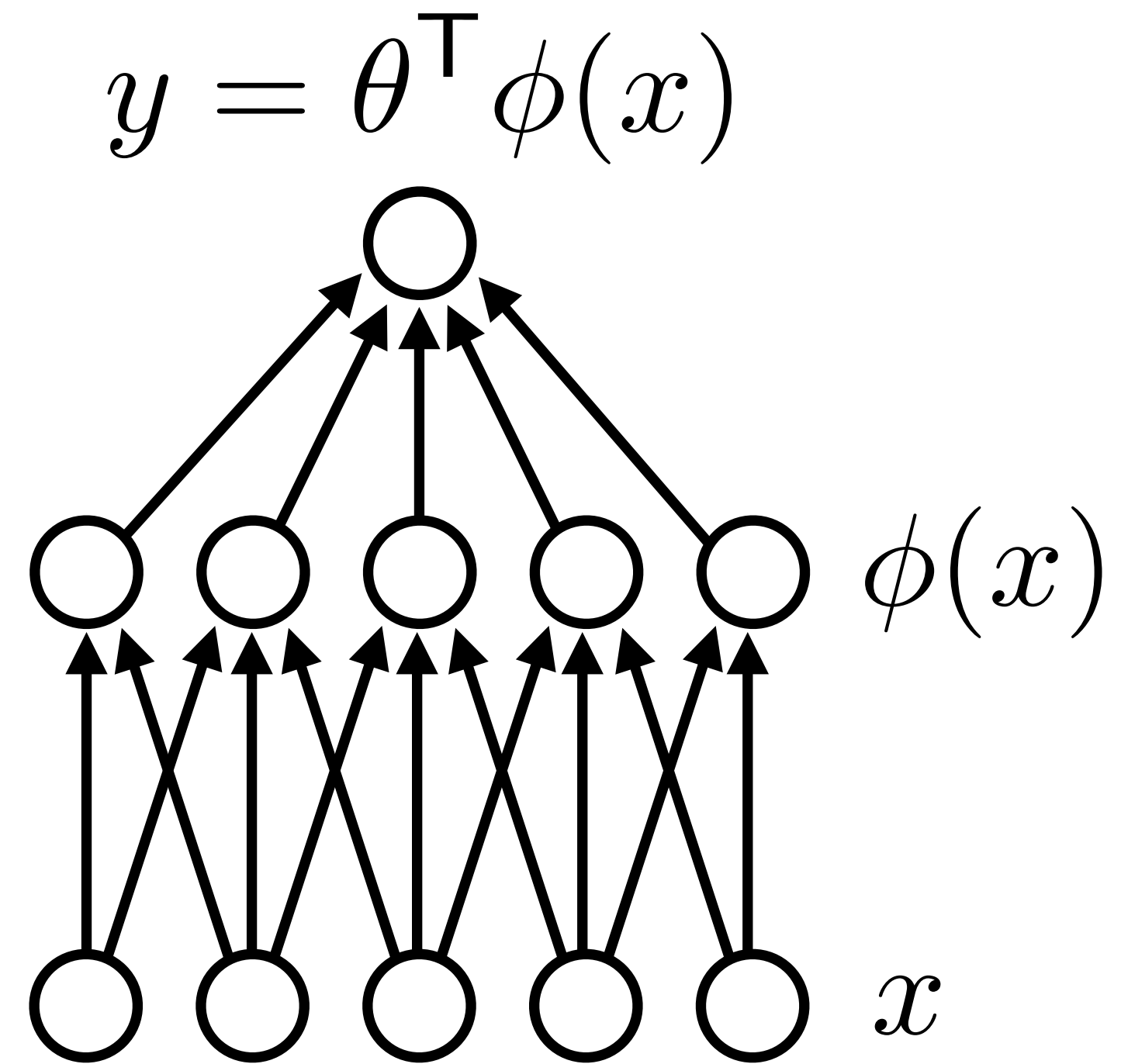
Neural networks for statisticians and economists

- ▶ This is function approximation.
- ▶ Regression, but fit the basis too.
- ▶ Adaptive basis function GLM
- ▶ Common bases: linear map then element-wise tanh or rectified linear
- ▶ Typically estimated via MAP
- ▶ “Deep” = composition of bases
- ▶ Automatic differentiation is key



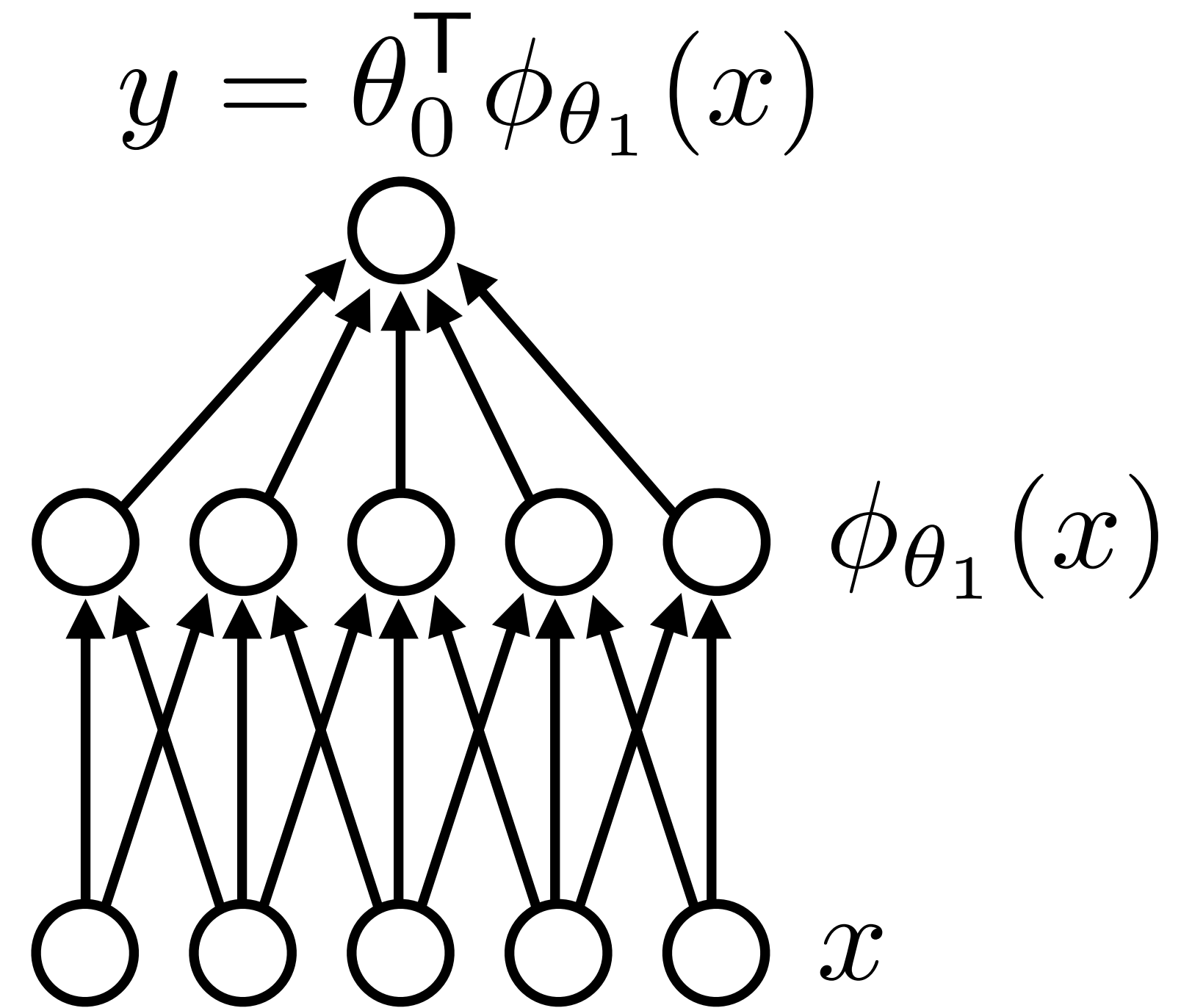
Neural networks for statisticians and economists

- ▶ This is function approximation.
- ▶ Regression, but fit the basis too.
- ▶ Adaptive basis function GLM
- ▶ Common bases: linear map then element-wise tanh or rectified linear
- ▶ Typically estimated via MAP
- ▶ “Deep” = composition of bases
- ▶ Automatic differentiation is key



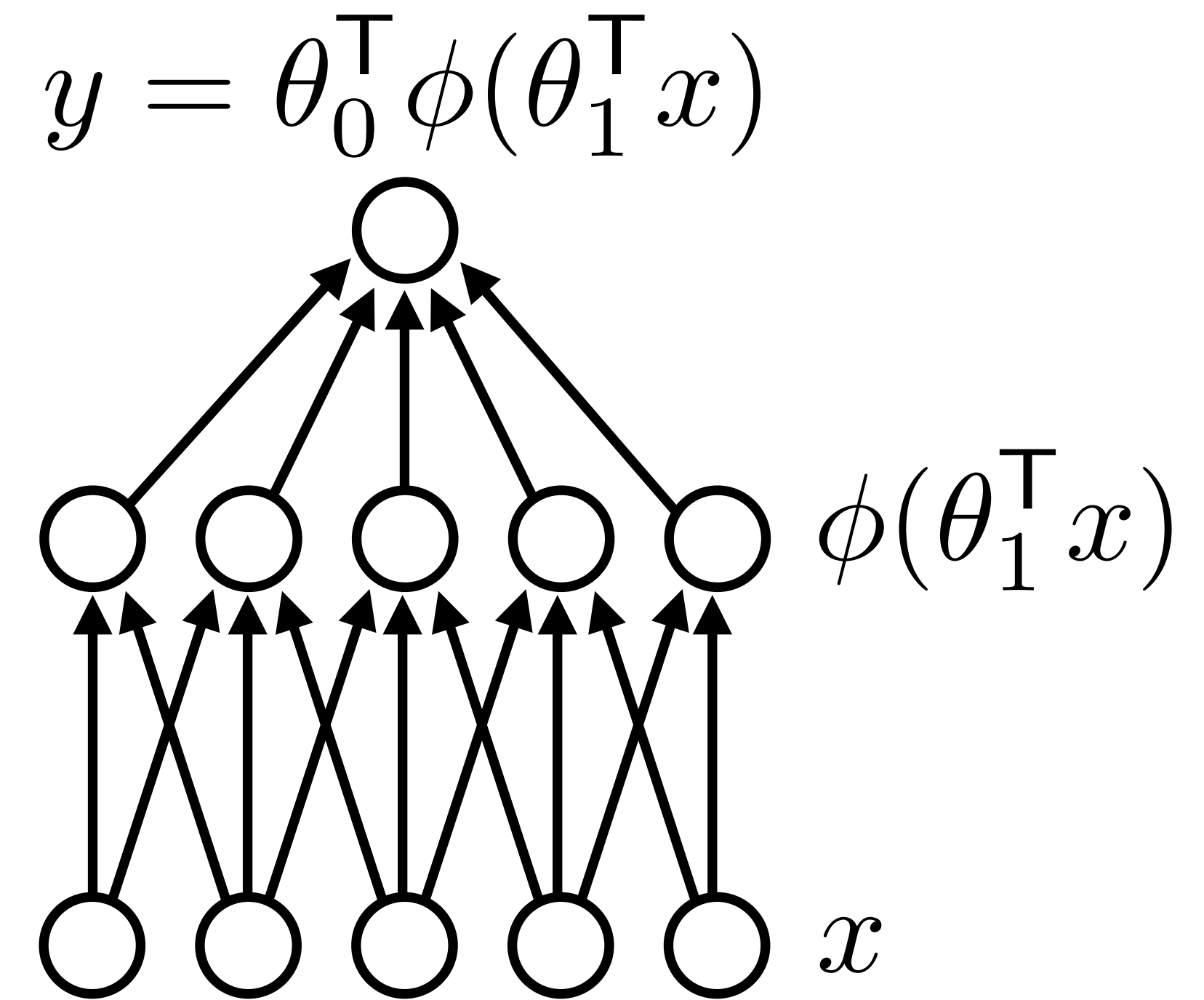
Neural networks for statisticians and economists

- ▶ This is function approximation.
- ▶ Regression, but fit the basis too.
- ▶ Adaptive basis function GLM
- ▶ Common bases: linear map then element-wise tanh or rectified linear
- ▶ Typically estimated via MAP
- ▶ “Deep” = composition of bases
- ▶ Automatic differentiation is key



Neural networks for statisticians and economists

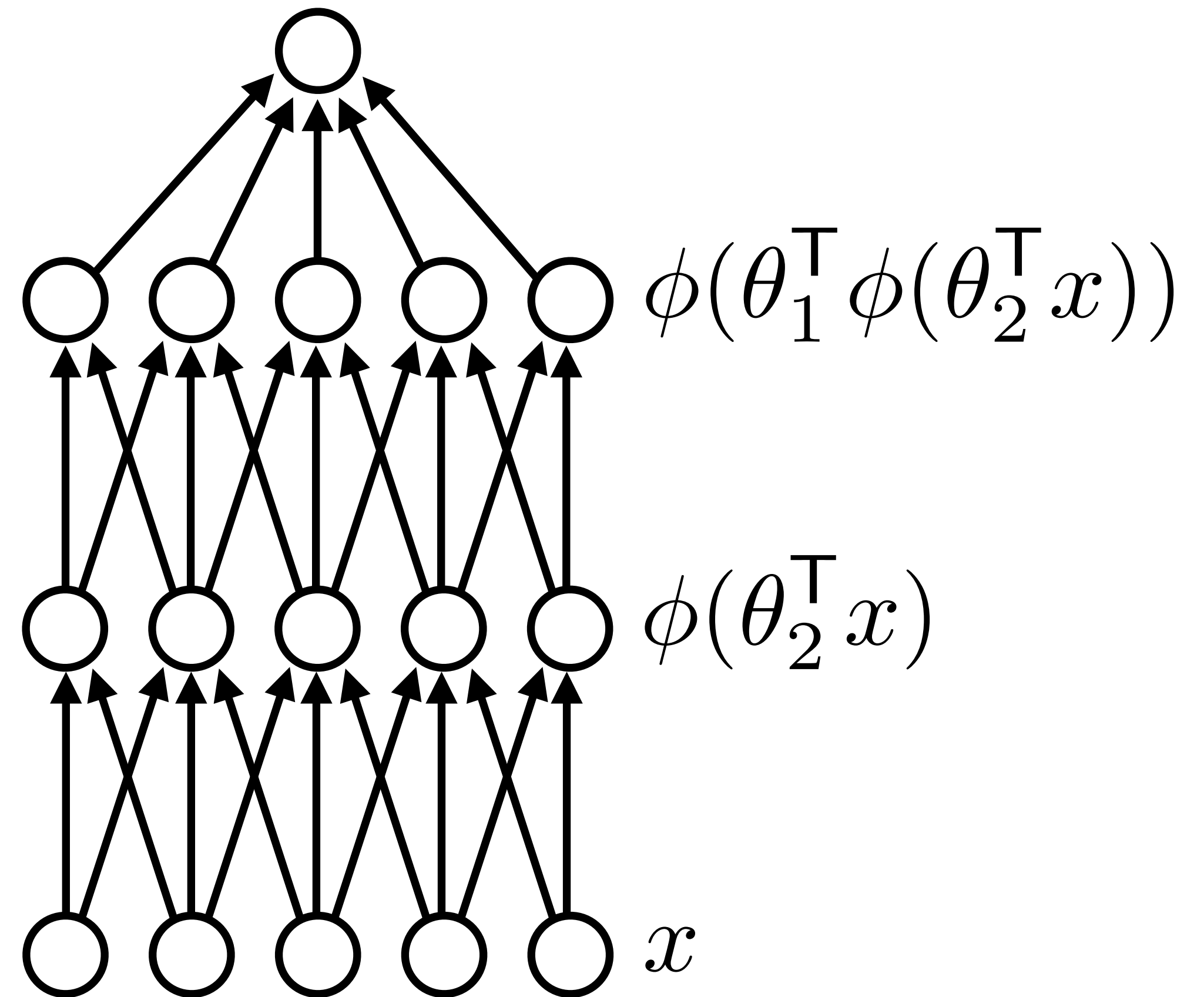
- ▶ This is function approximation.
- ▶ Regression, but fit the basis too.
- ▶ Adaptive basis function GLM
- ▶ Common bases: linear map then element-wise tanh or rectified linear
- ▶ Typically estimated via MAP
- ▶ “Deep” = composition of bases
- ▶ Automatic differentiation is key



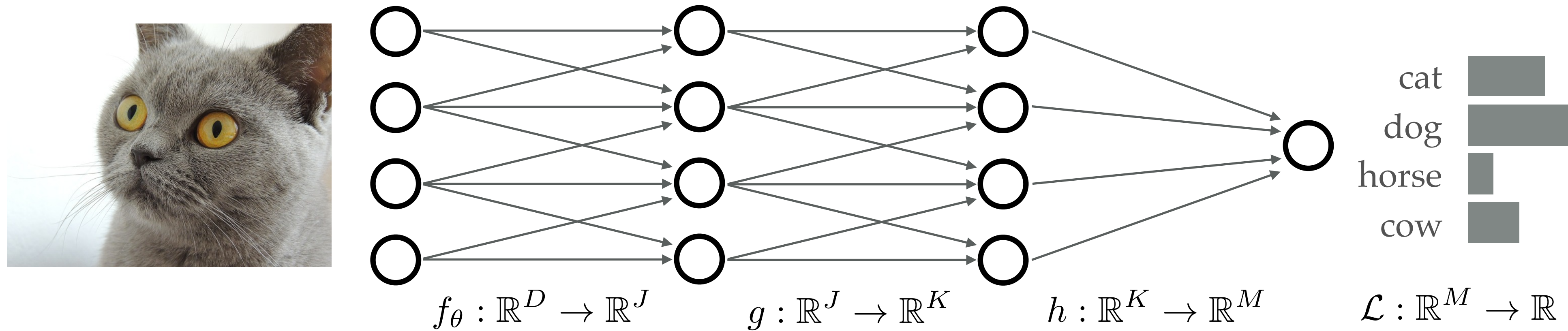
Neural networks for statisticians and economists

- ▶ This is function approximation.
- ▶ Regression, but fit the basis too.
- ▶ Adaptive basis function GLM
- ▶ Common bases: linear map then element-wise tanh or rectified linear
- ▶ Typically estimated via MAP
- ▶ “Deep” = composition of bases
- ▶ Automatic differentiation is key

$$y = \theta_0^\top \phi(\theta_1^\top \phi(\theta_2^\top x))$$



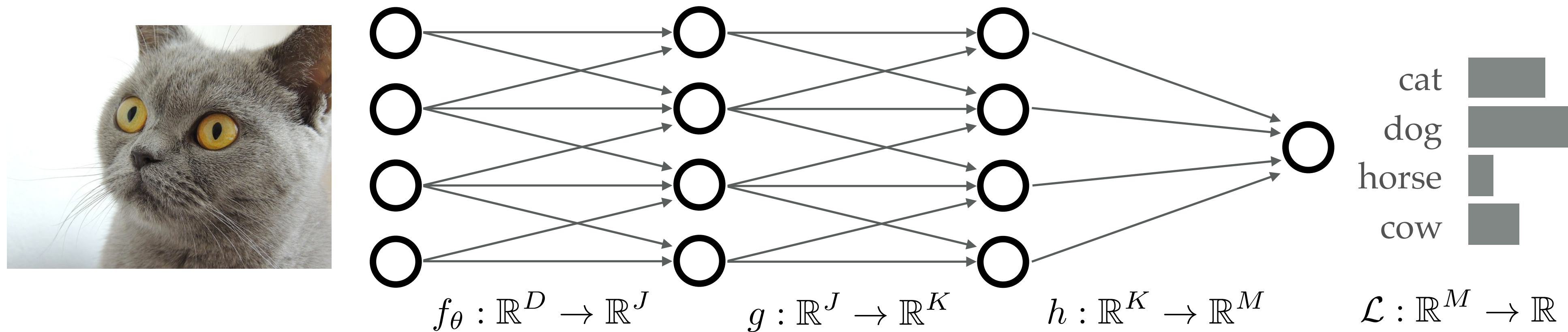
Associativity and the Chain Rule



$$\frac{\partial}{\partial \theta} \mathcal{L}(h(g(f_\theta(x)))) = \begin{bmatrix} \frac{d\mathbf{f}}{d\theta} \end{bmatrix} \begin{bmatrix} \frac{d\mathbf{g}}{d\mathbf{f}} \end{bmatrix} \begin{bmatrix} \frac{d\mathbf{h}}{d\mathbf{g}} \end{bmatrix} \begin{bmatrix} \frac{d\mathcal{L}}{d\mathbf{h}} \end{bmatrix}$$

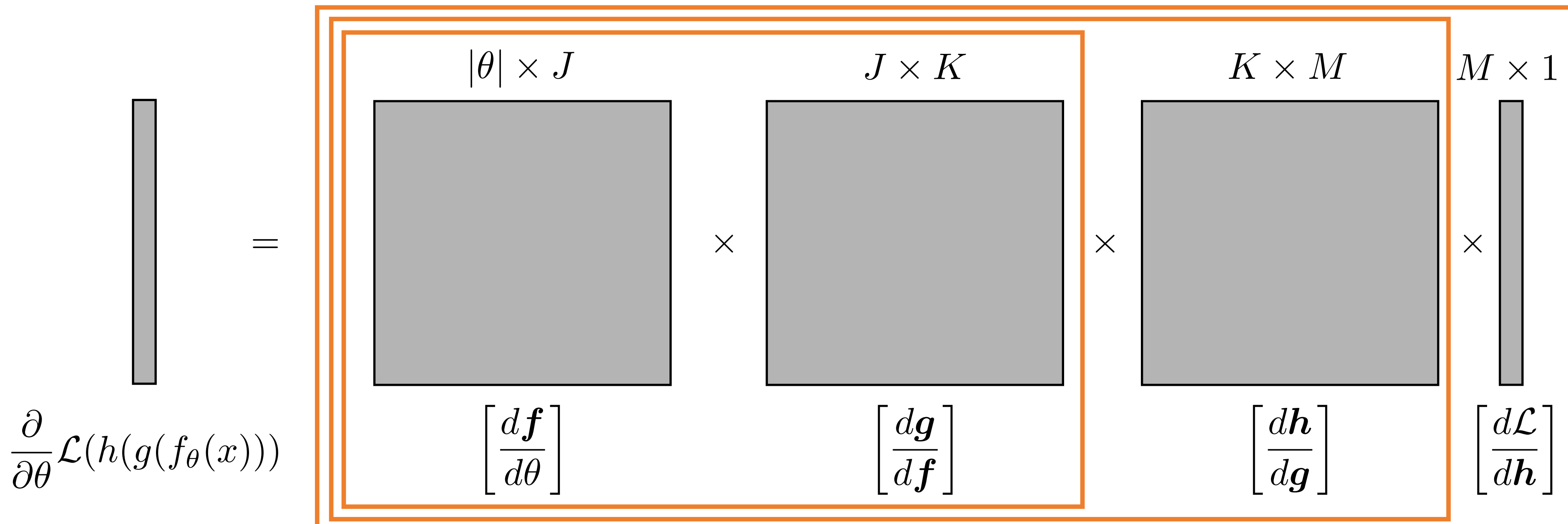
$$\frac{\partial}{\partial \theta} \mathcal{L}(h(g(f_\theta(x)))) = \begin{matrix} & |\theta| \times J & & J \times K & & K \times M & & M \times 1 \\ \begin{bmatrix} \frac{d\mathbf{f}}{d\theta} \end{bmatrix} & \times & \begin{bmatrix} \frac{d\mathbf{g}}{d\mathbf{f}} \end{bmatrix} & \times & \begin{bmatrix} \frac{d\mathbf{h}}{d\mathbf{g}} \end{bmatrix} & \times & \begin{bmatrix} \frac{d\mathcal{L}}{d\mathbf{h}} \end{bmatrix} \end{matrix}$$

Forward Mode Automatic Differentiation

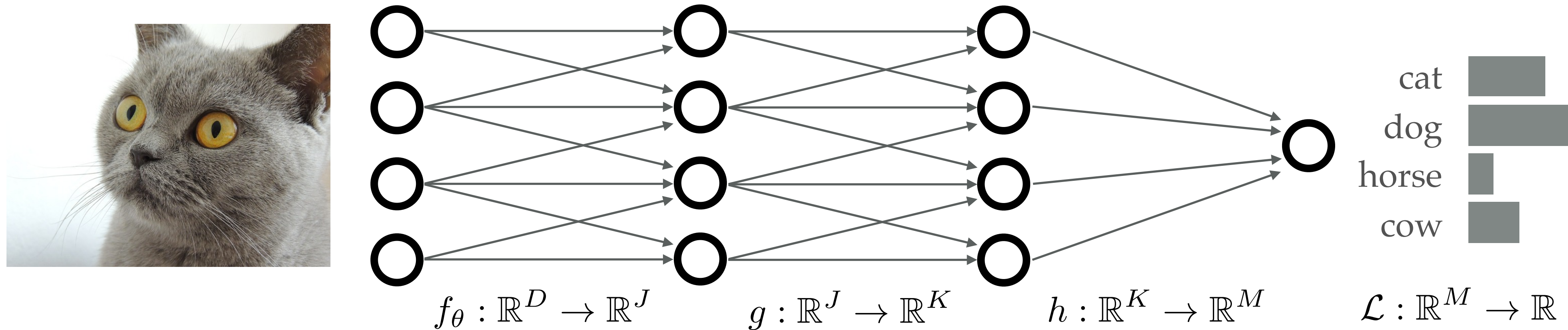


Left to right:

$$\frac{\partial}{\partial \theta} \mathcal{L}(h(g(f_\theta(x)))) = \begin{bmatrix} \frac{df}{d\theta} \\ \frac{dg}{df} \end{bmatrix} \begin{bmatrix} \frac{dh}{dg} \\ \frac{d\mathcal{L}}{dh} \end{bmatrix}$$

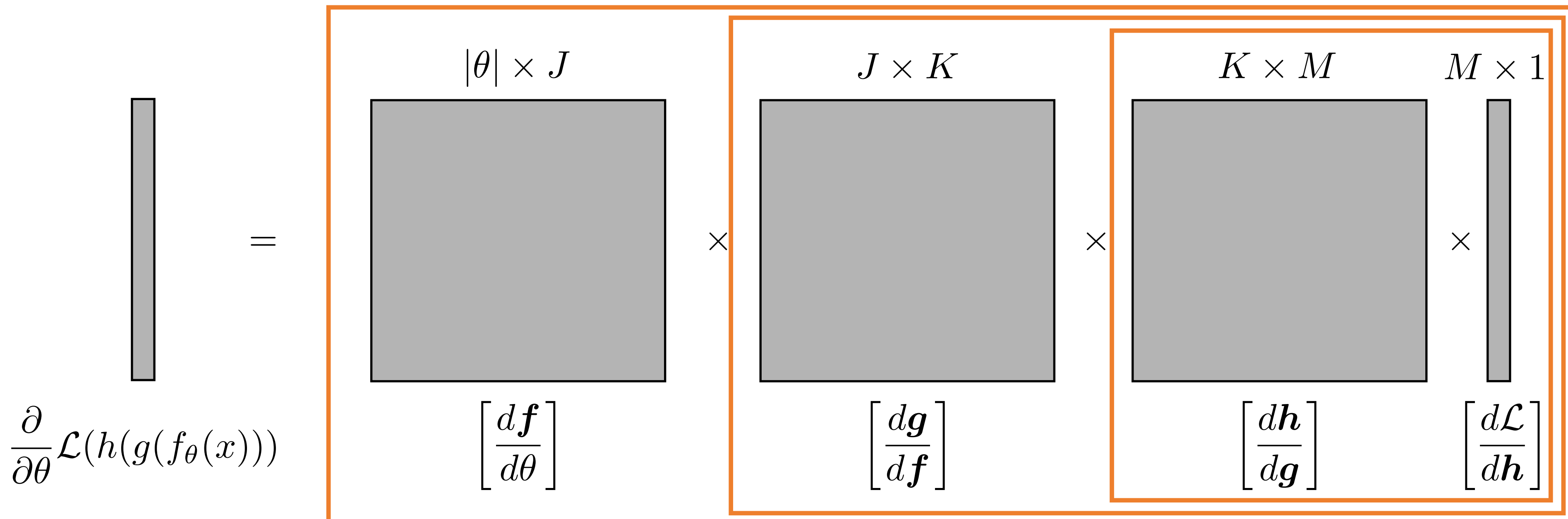


“Backpropagation”: Reverse Mode Automatic Differentiation



Right to left:

$$\frac{\partial}{\partial \theta} \mathcal{L}(h(g(f_\theta(x)))) = \frac{df}{d\theta} \frac{dg}{df} \frac{dh}{dg} \frac{d\mathcal{L}}{dh}$$



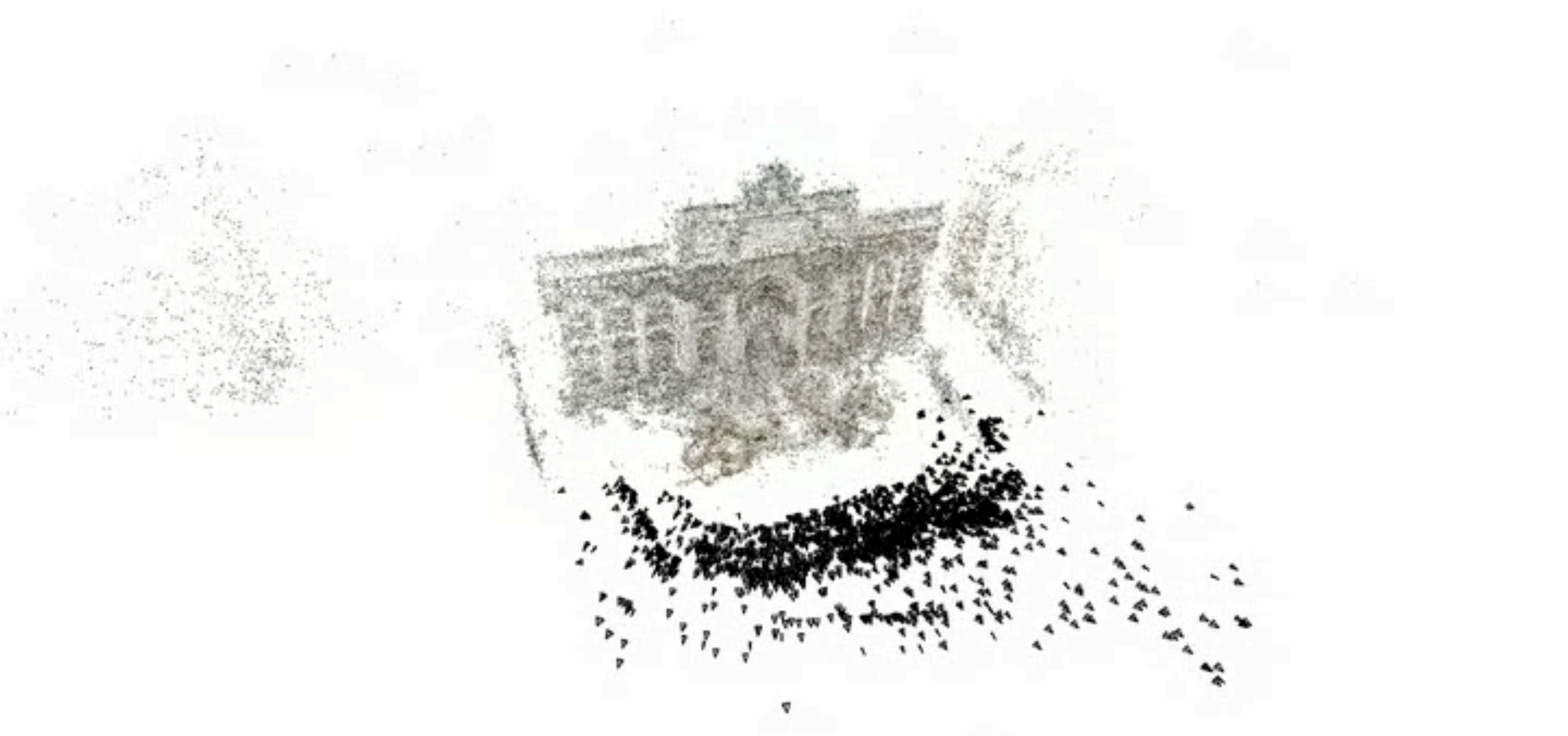
Unsupervised Learning

Identifying structure in data without known ground truth.

3d Structure from Image Collections



3d Structure from Image Collections

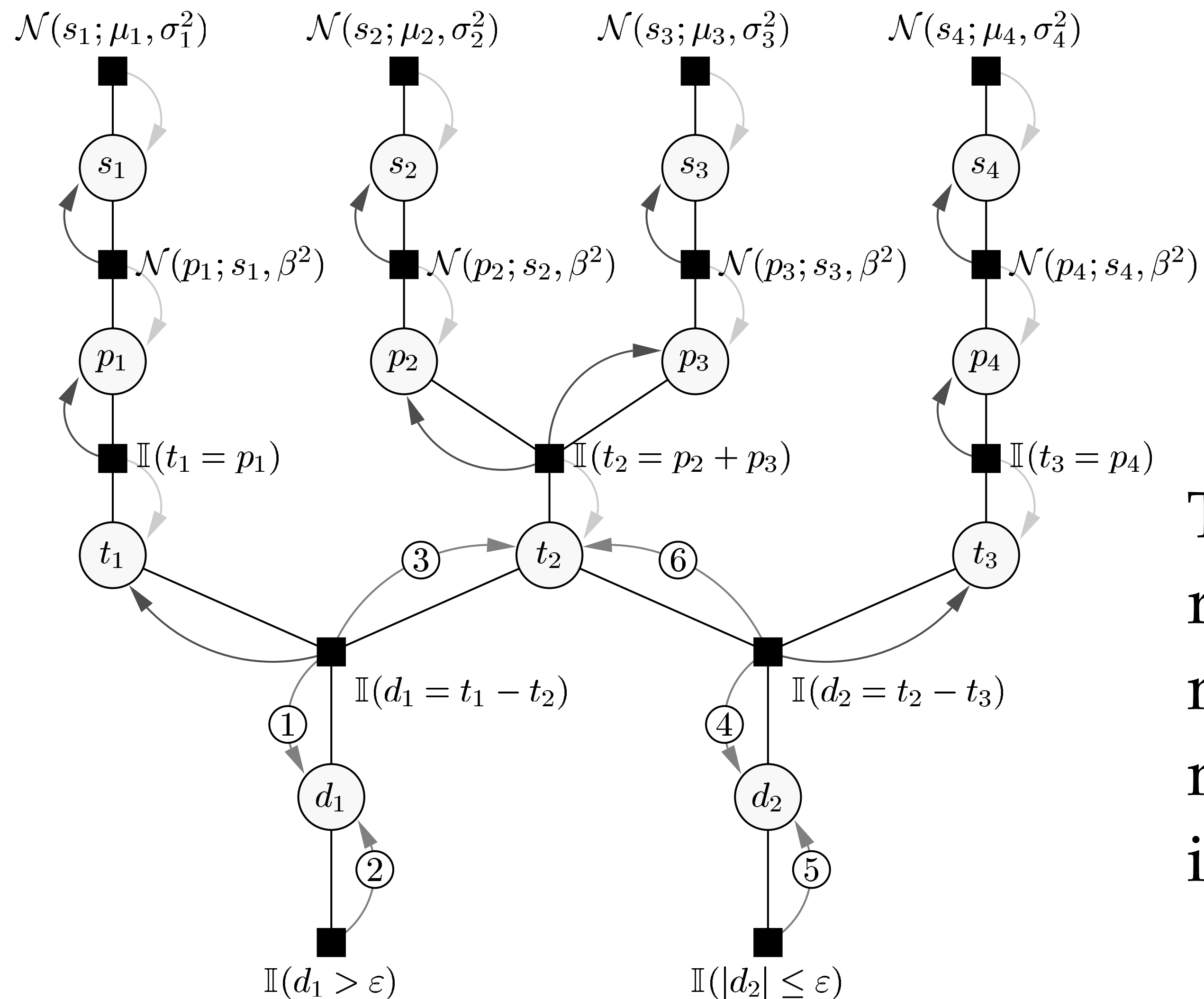


Agarwal, Snavely, Simon, Seitz, Szelisky (ICCV 2009)

3d Structure from Image Collections



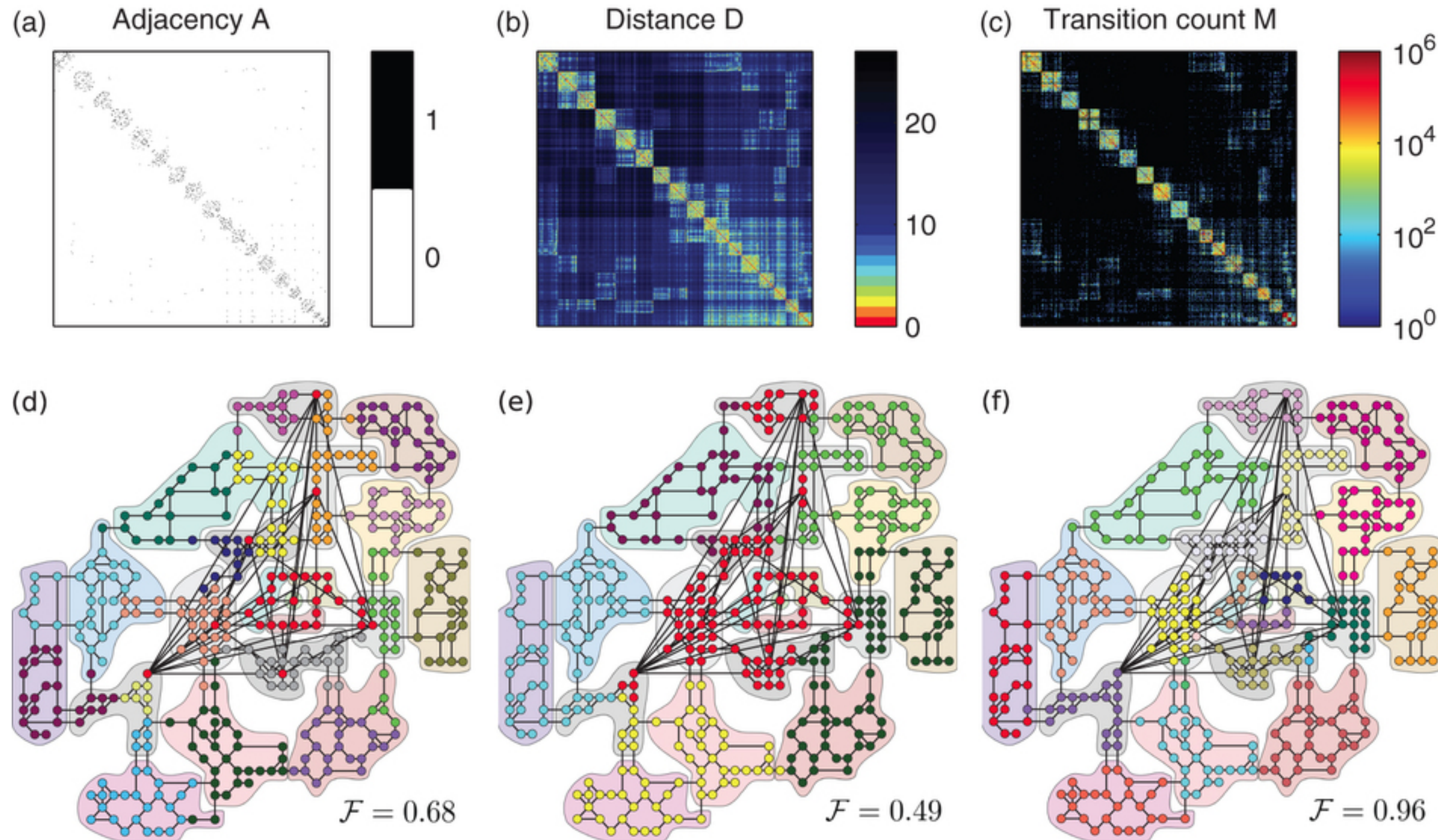
Ranking Online Gamers



XBOX 360™

The TrueSkill system for Xbox rankings uses an ML probabilistic model to estimate the abilities of millions of gamers and match them into teams.

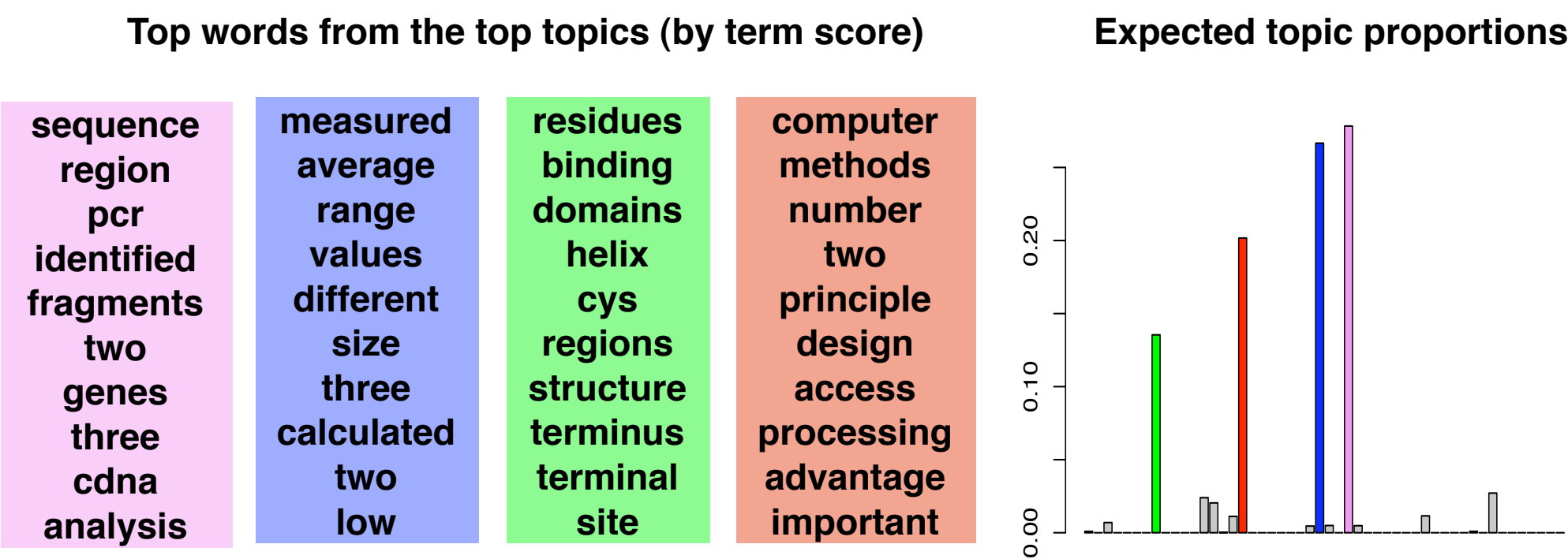
Finding Communities in a Social Network



Extracting Topics from Text

Chance and Statistical Significance in Protein and DNA Sequence Analysis

Samuel Karlin and Volker Brendel



Abstract with the most likely topic assignments

Statistical approaches help in the determination of significant configurations in protein and nucleic acid sequence data. Three recent statistical methods are discussed: (i) score-based sequence analysis that provides a means for characterizing anomalies in local sequence text and for evaluating sequence comparisons; (ii) quantile distributions of amino acid usage that reveal general compositional biases in proteins and evolutionary relations; and (iii) *r*-scan statistics that can be applied to the analysis of spacings of sequence markers.

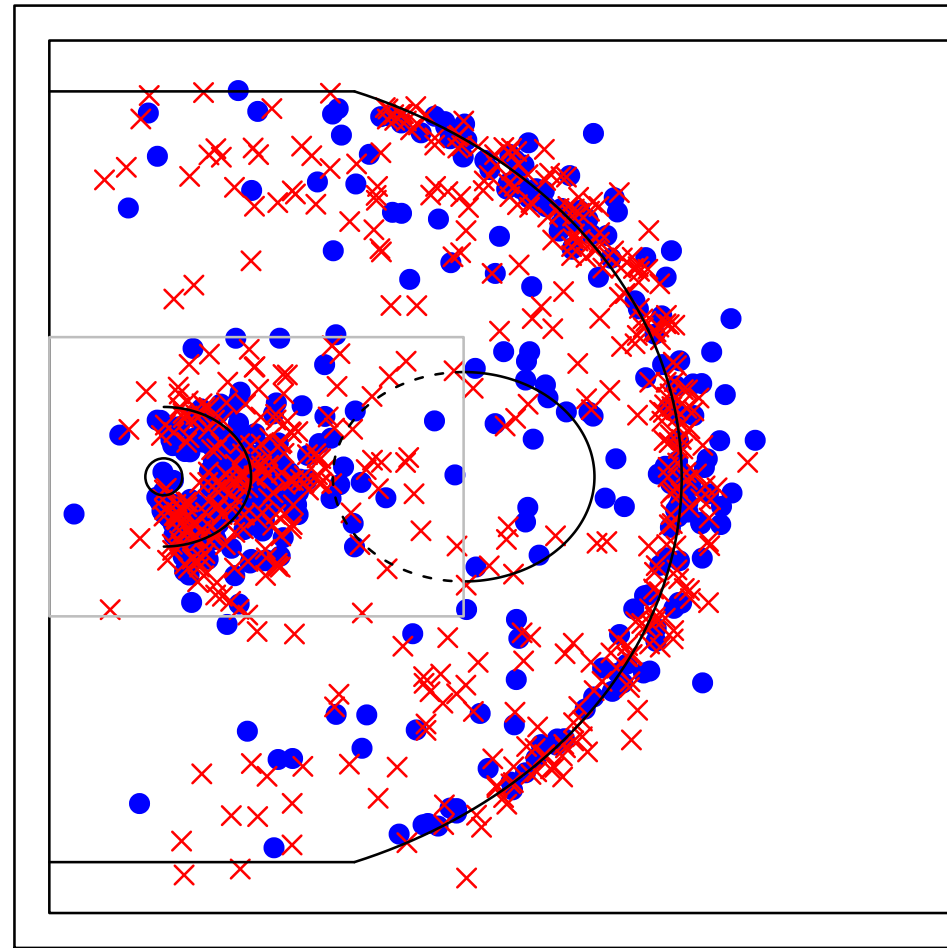
Top Ten Similar Documents

- Exhaustive Matching of the Entire Protein Sequence Database
- How Big Is the Universe of Exons?
- Counting and Discounting the Universe of Exons
- Detecting Subtle Sequence Signals: A Gibbs Sampling Strategy for Multiple Alignment
- Ancient Conserved Regions in New Gene Sequences and the Protein Databases
- A Method to Identify Protein Sequences that Fold into a Known Three- Dimensional Structure
- Testing the Exon Theory of Genes: The Evidence from Protein Structure
- Predicting Coiled Coils from Protein Sequences
- Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology

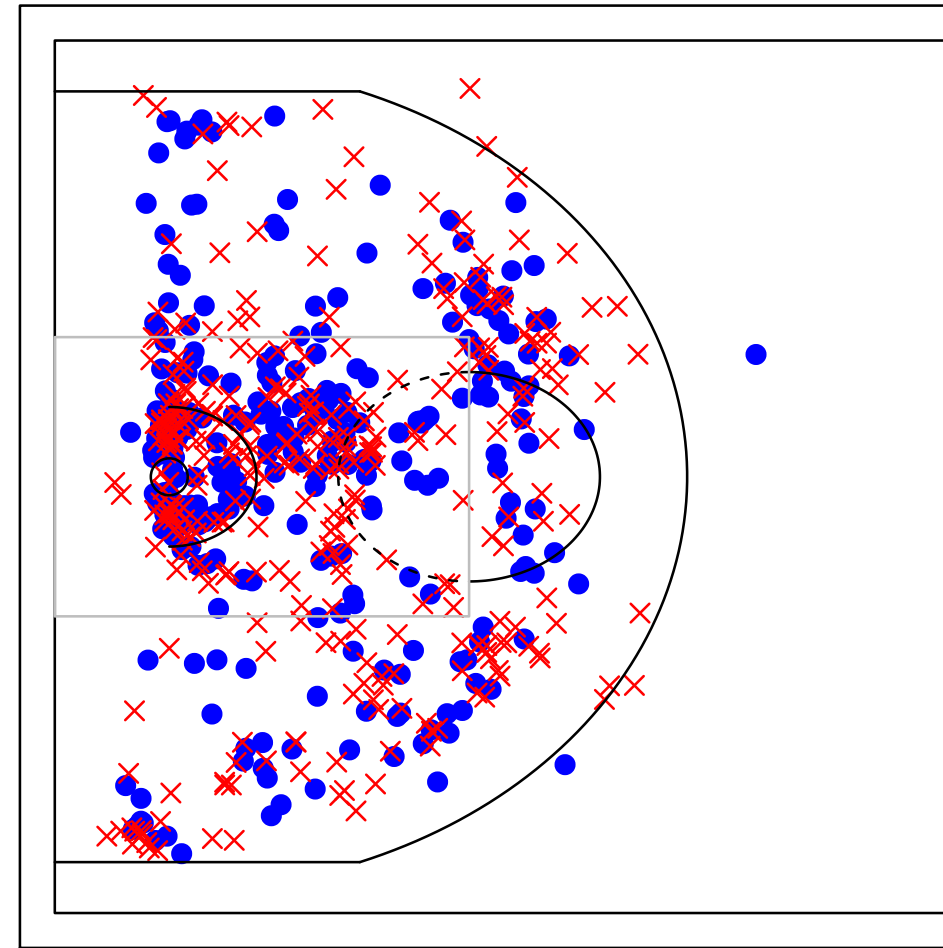
Figure by David Blei

Modeling Sports Data

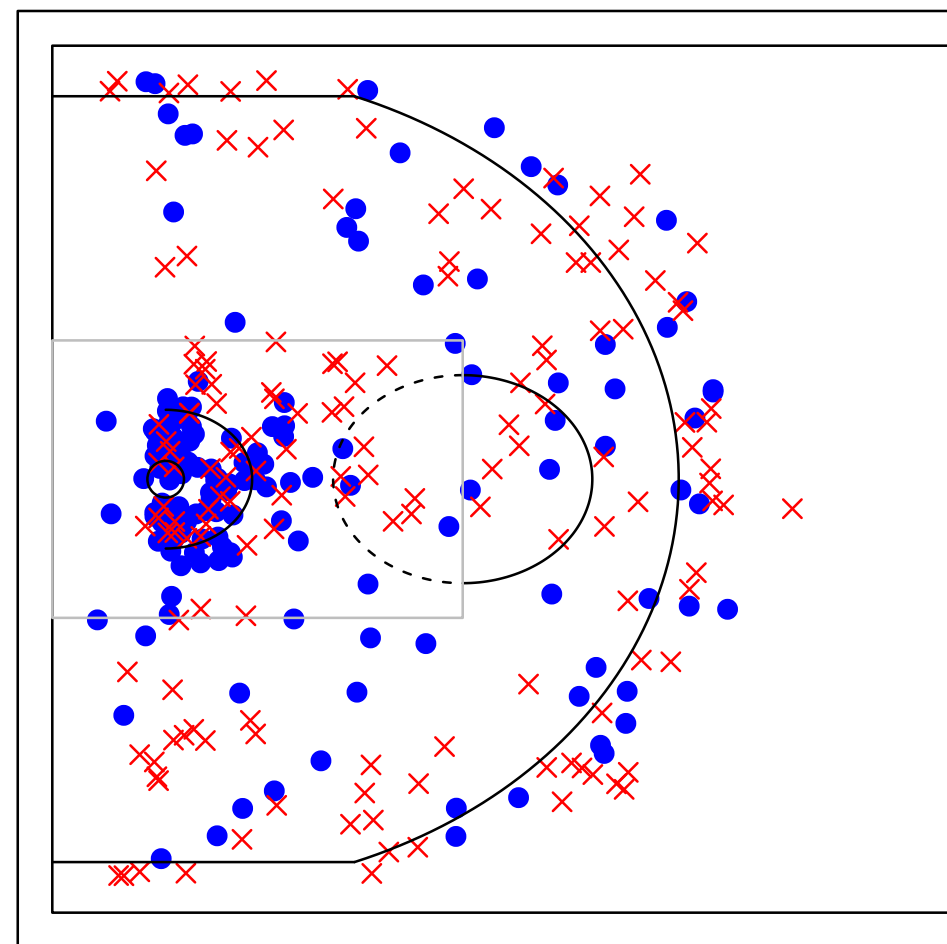
James Harden (965 shots)



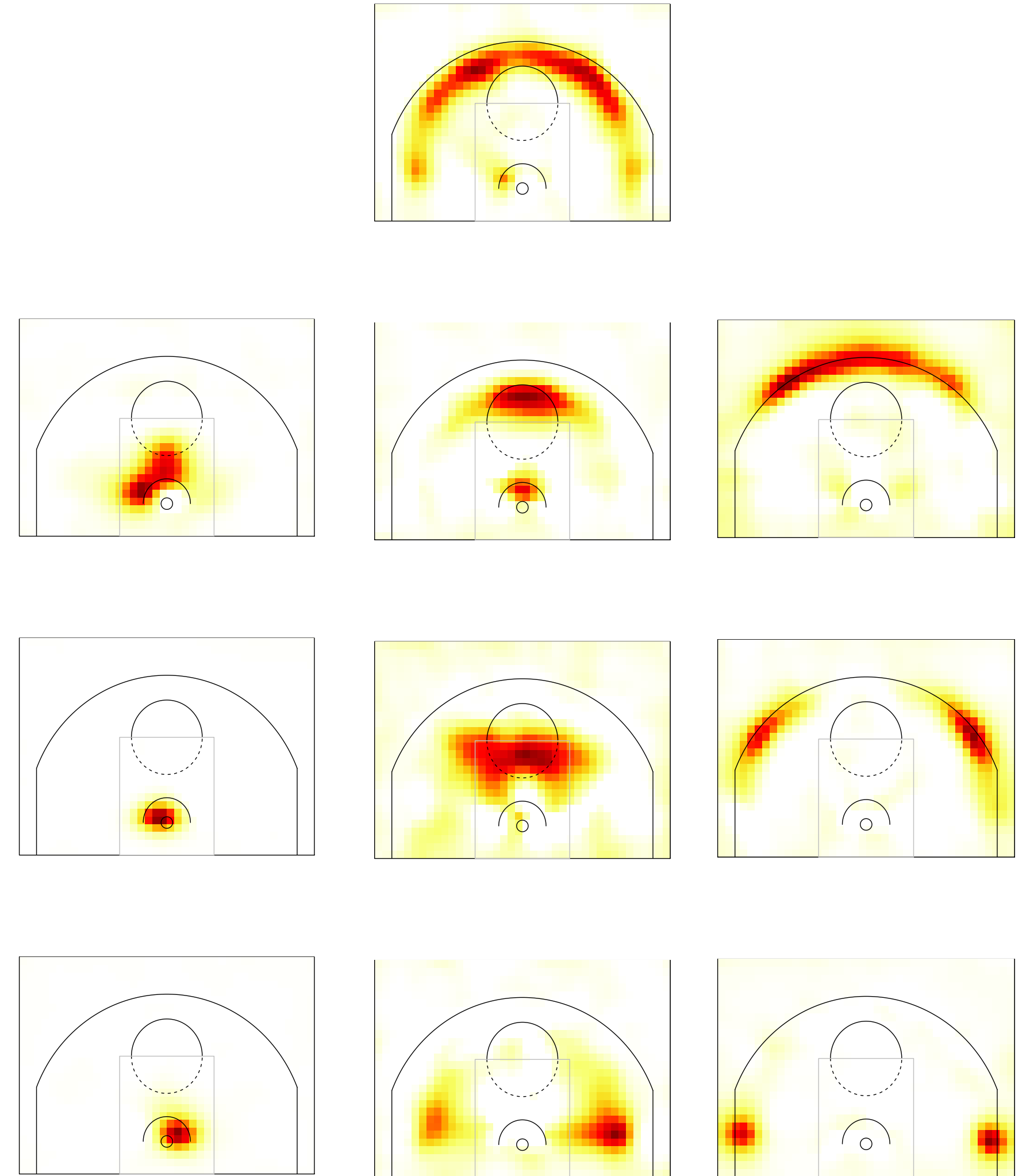
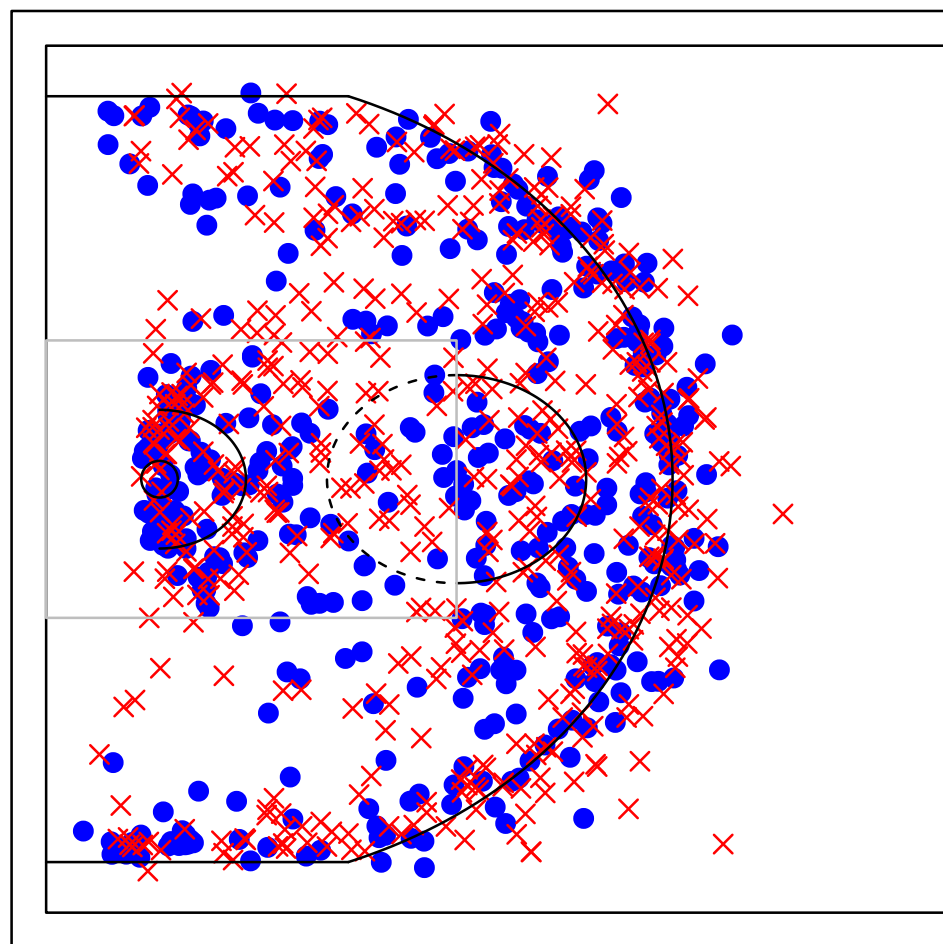
Tony Parker (692 shots)



LeBron James (315 shots)

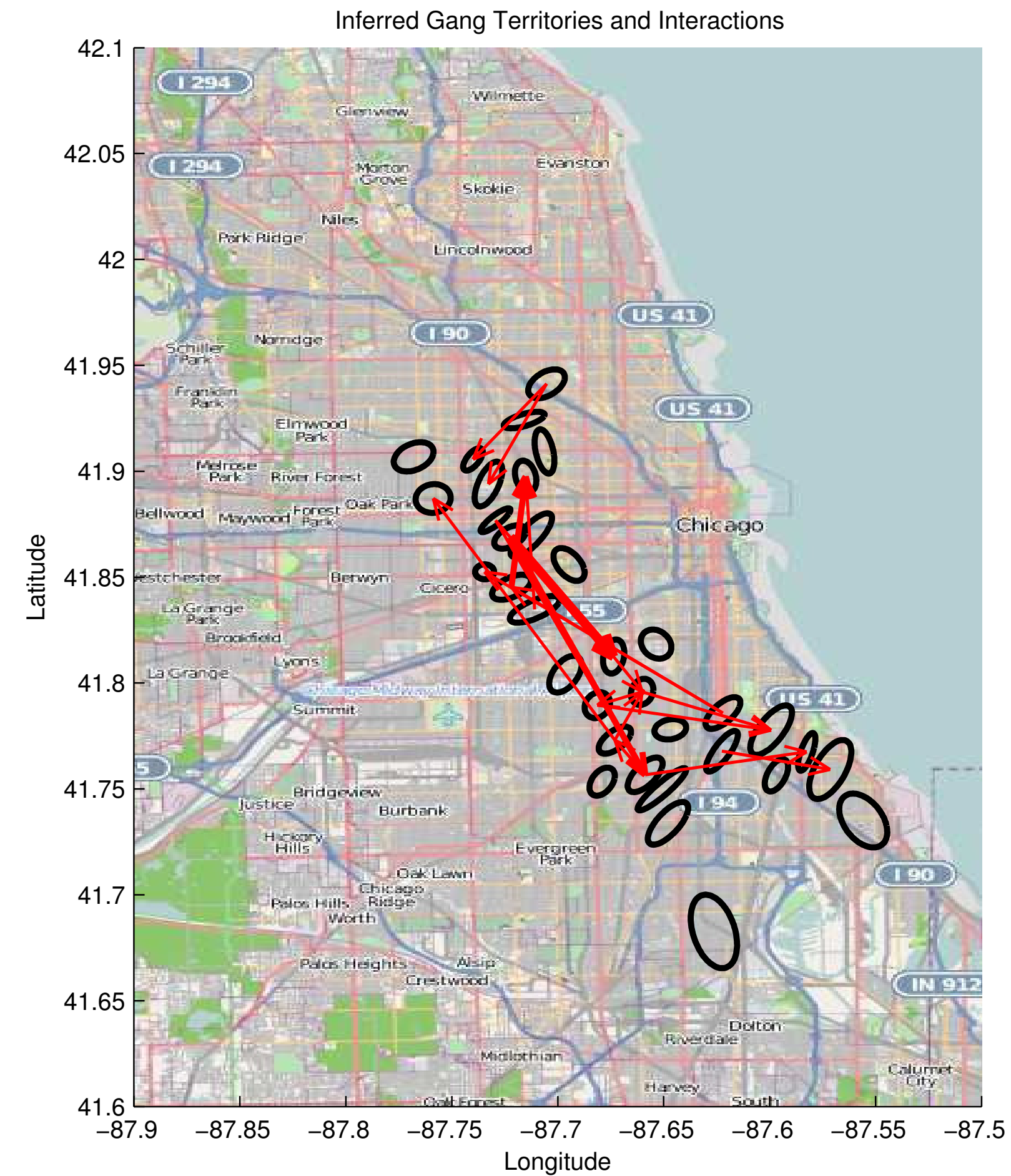
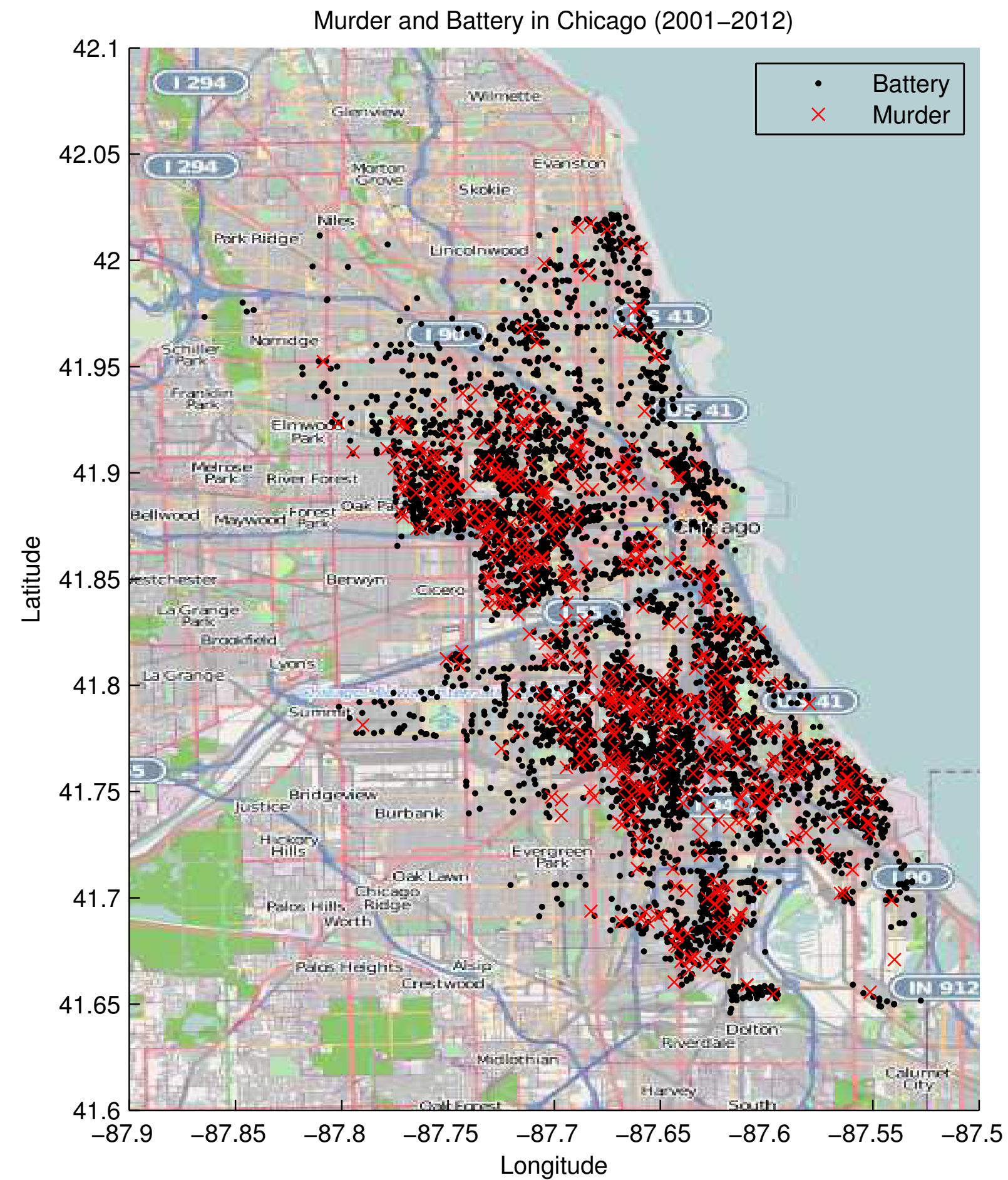


Stephen Curry (940 shots)



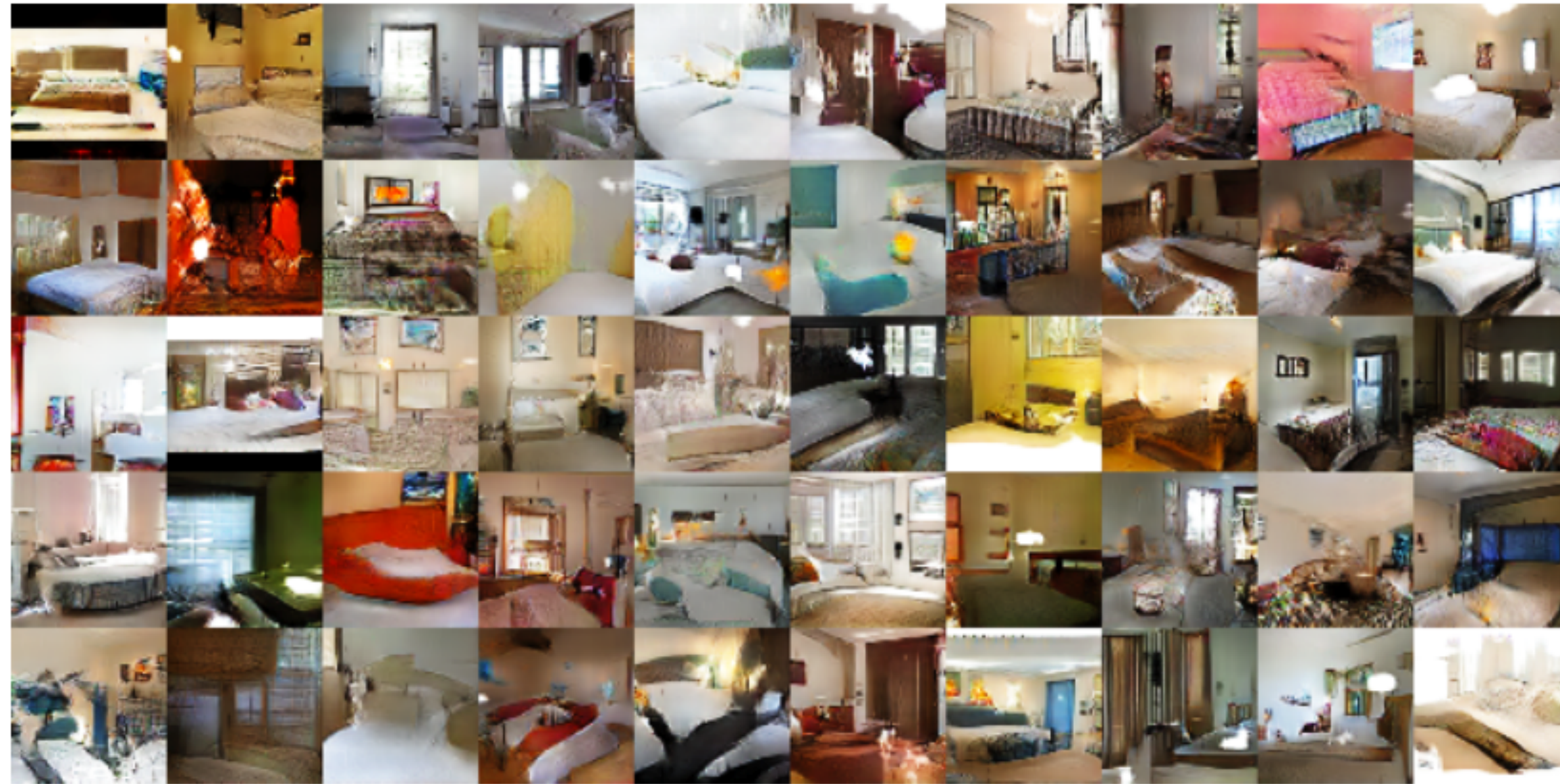
Low-dimensional structure in NBA shooting patterns

Spatiotemporal Patterns on Networks

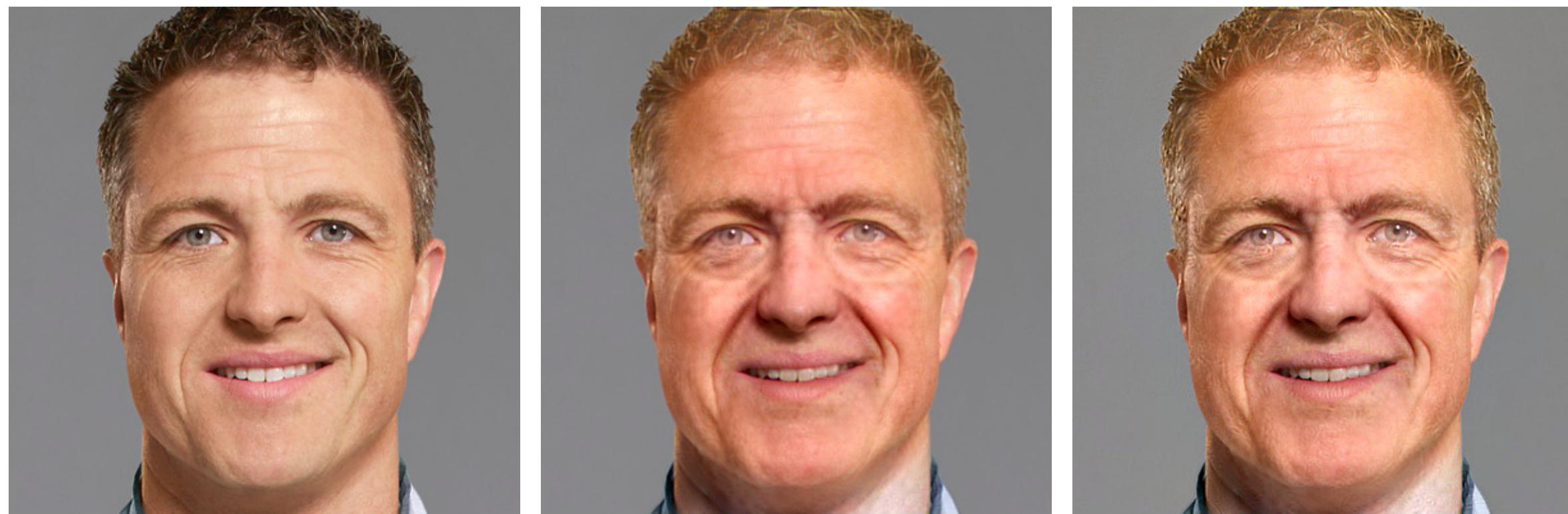


Modeling network-driven patterns of violence in Chicago

Generative Models for Synthesizing Data



Radford et al. 2016: Bedroom Scenes



Original

DFI

DFI + Artifact Removal

Upchurch et al. 2016: Aging Faces



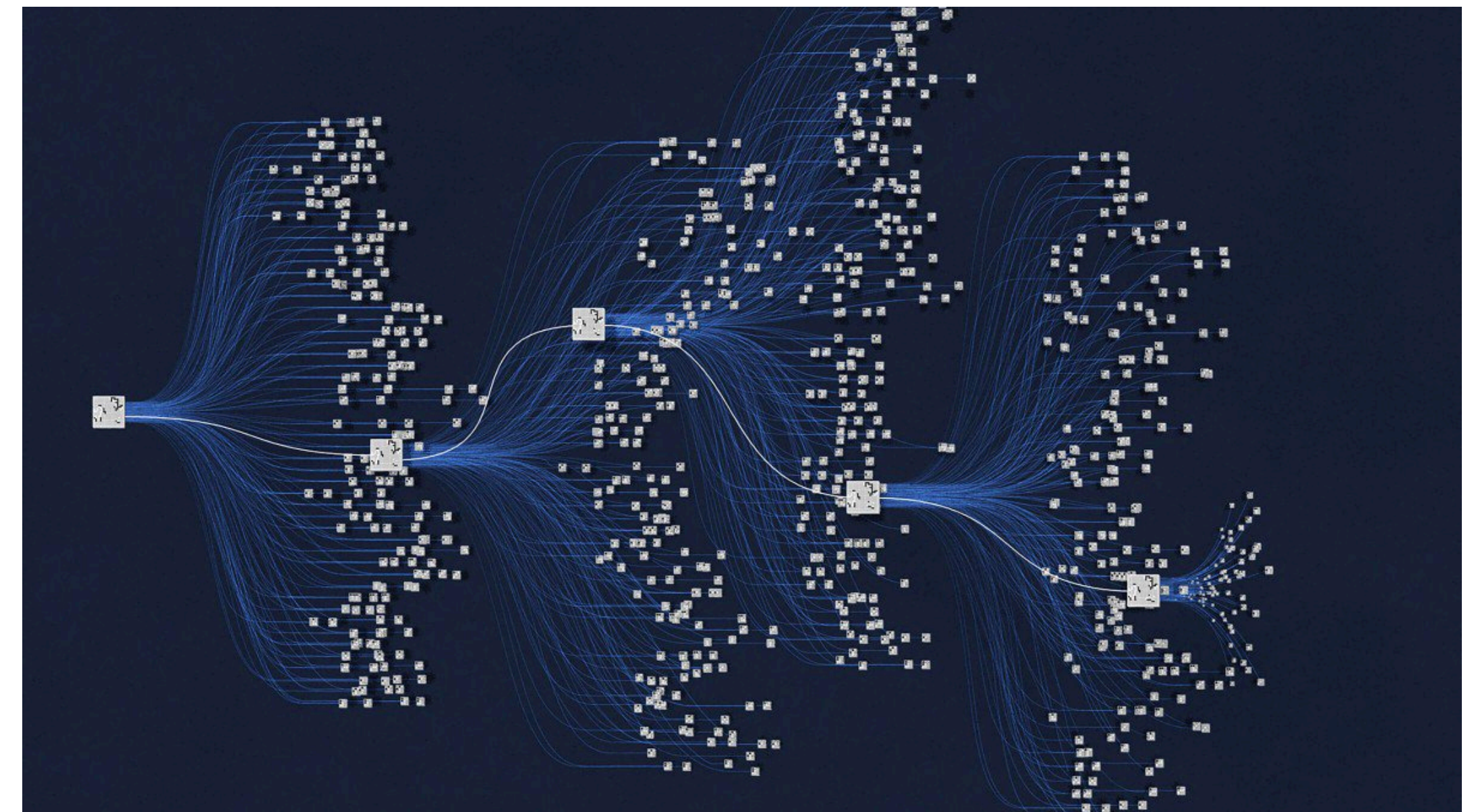
Radford et al. 2016: Faces

Reinforcement Learning

Interacting with an environment to maximize reward.

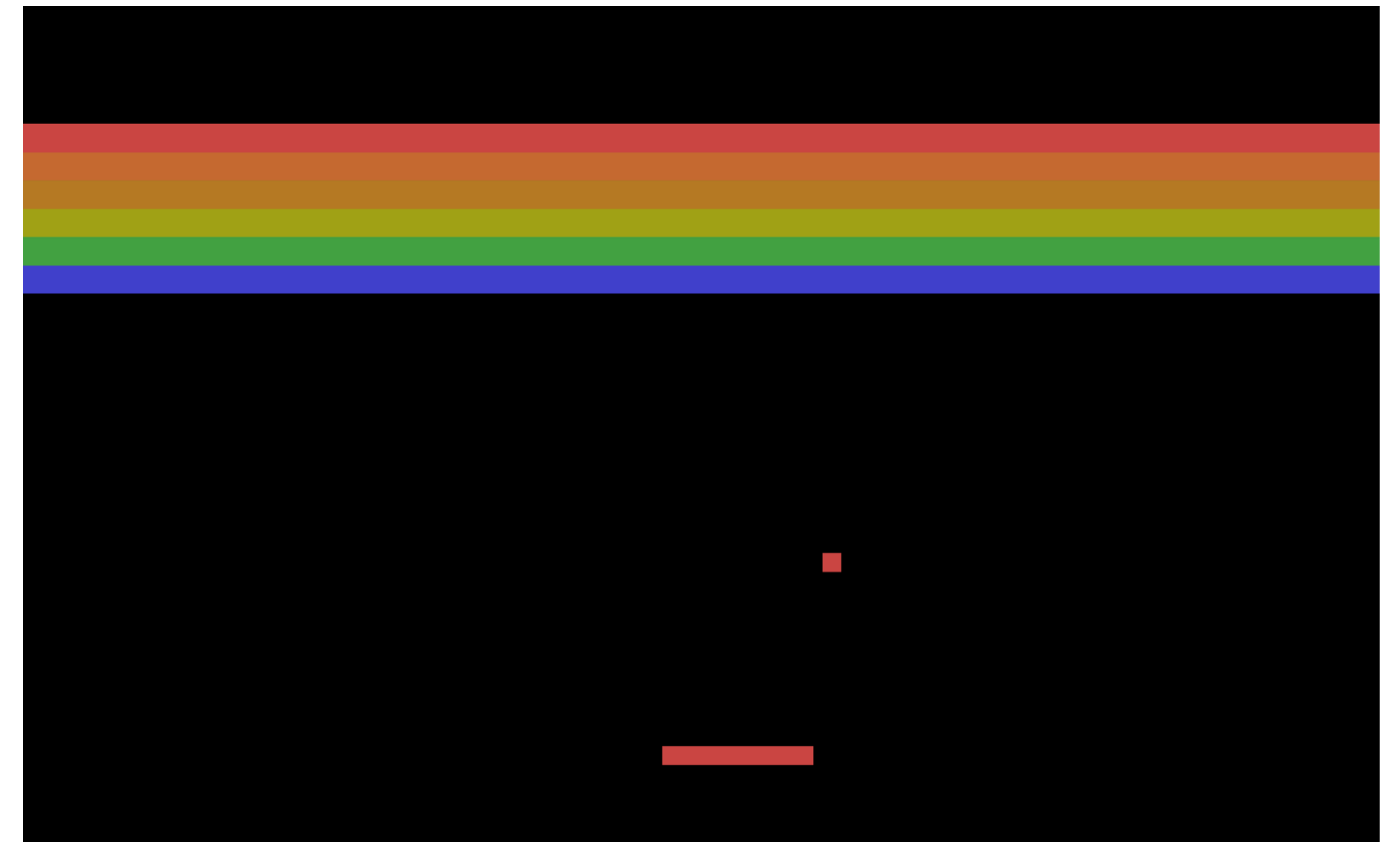
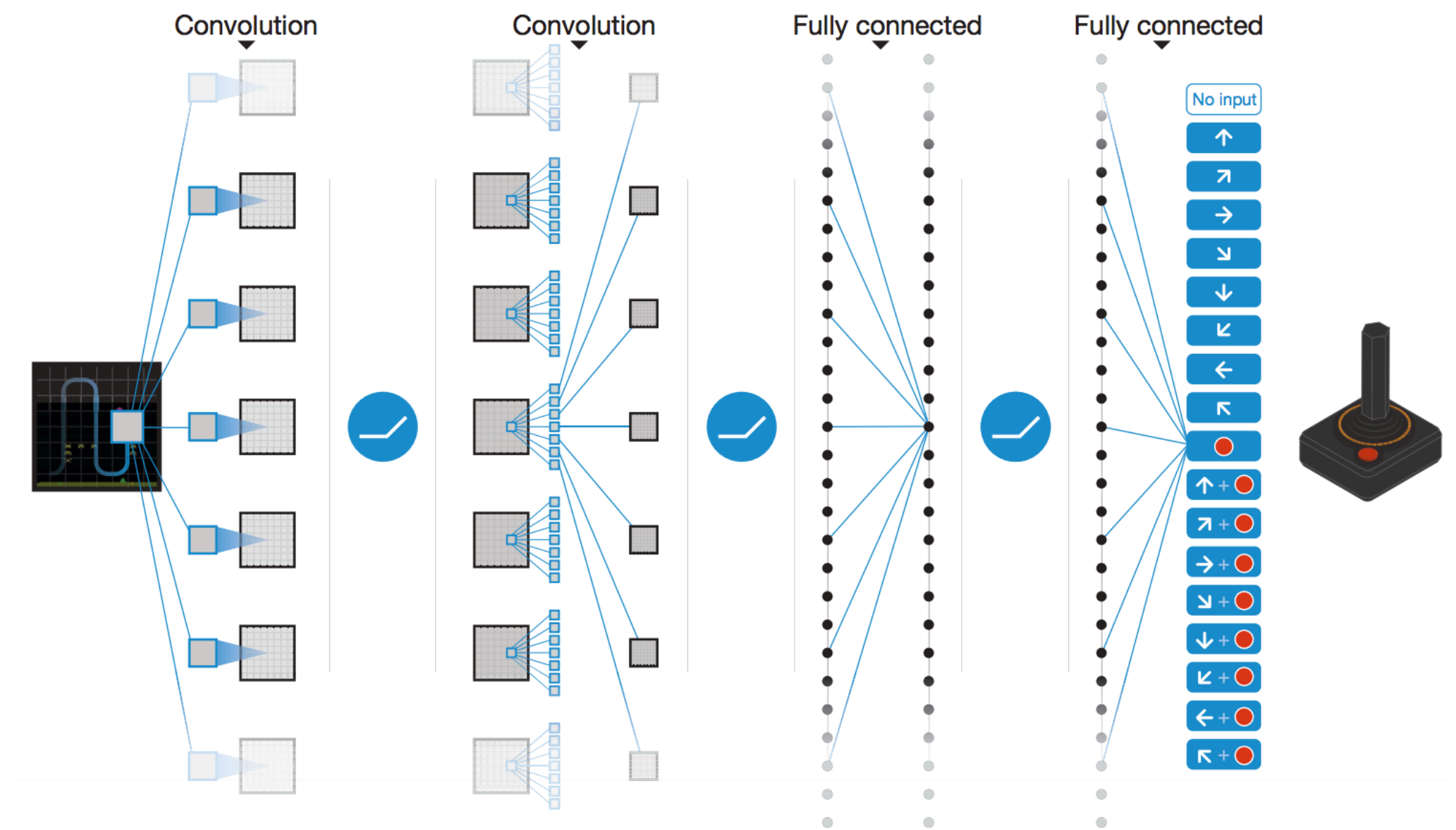
Learning to Plan and Make Decisions: Go

- ▶ Go is a much harder game for computers to solve than chess.
- ▶ Has been a “holy grail” with solutions thought to be decades away.
- ▶ DeepMind’s AlphaGo surprised the world by beating a champion.
- ▶ Bespoke approach based on recognizing board positions, tree search, expert data, lots of self play.



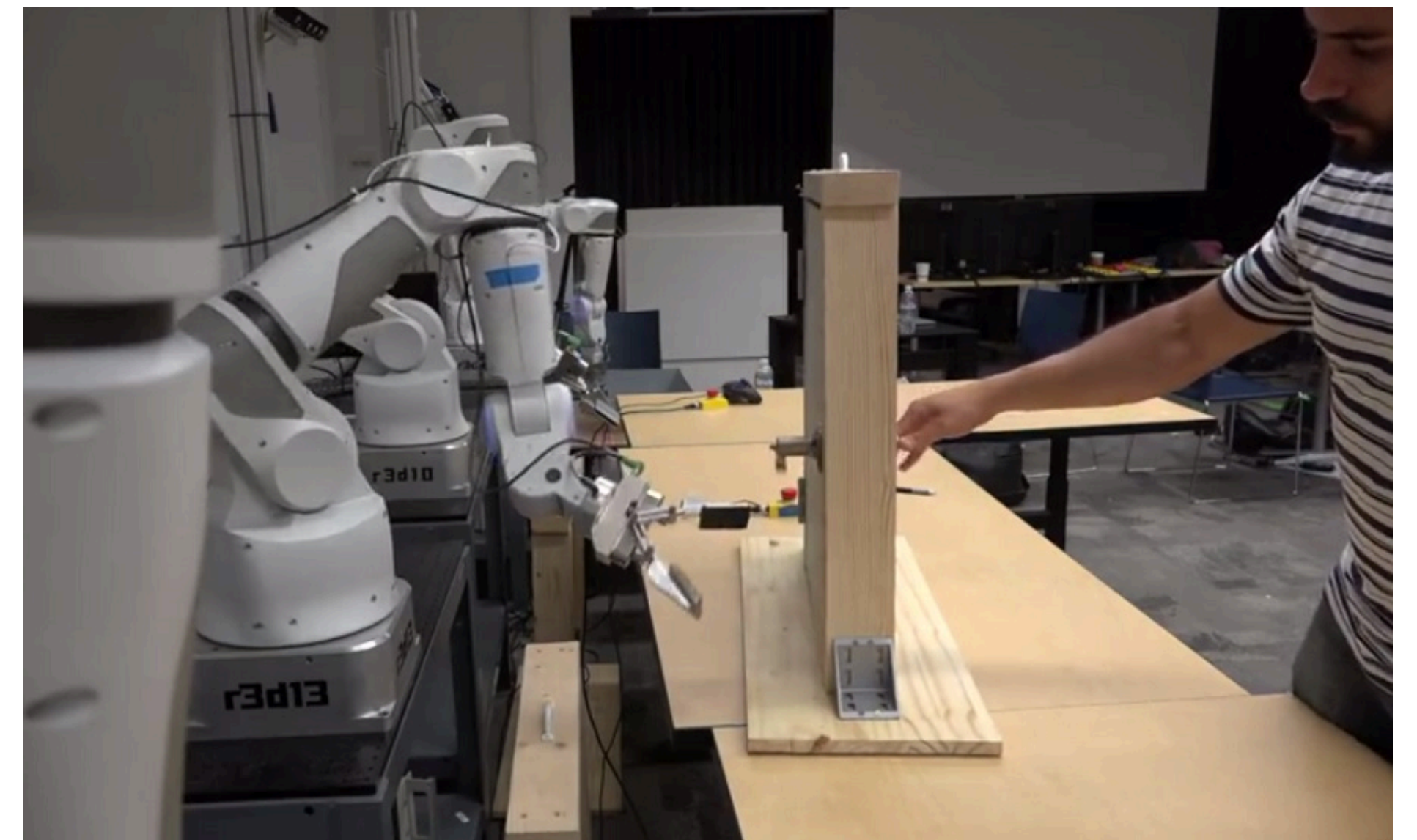
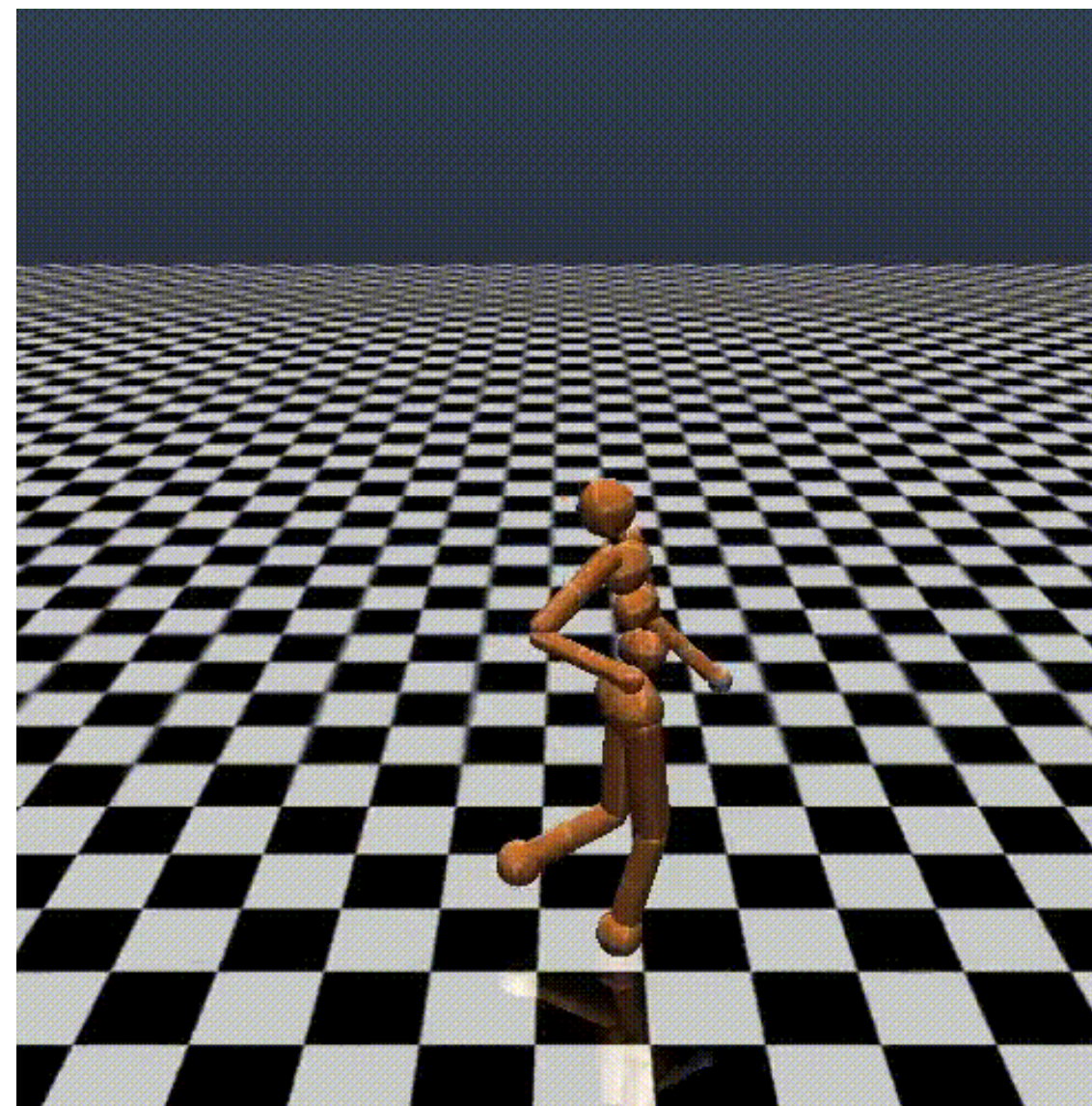
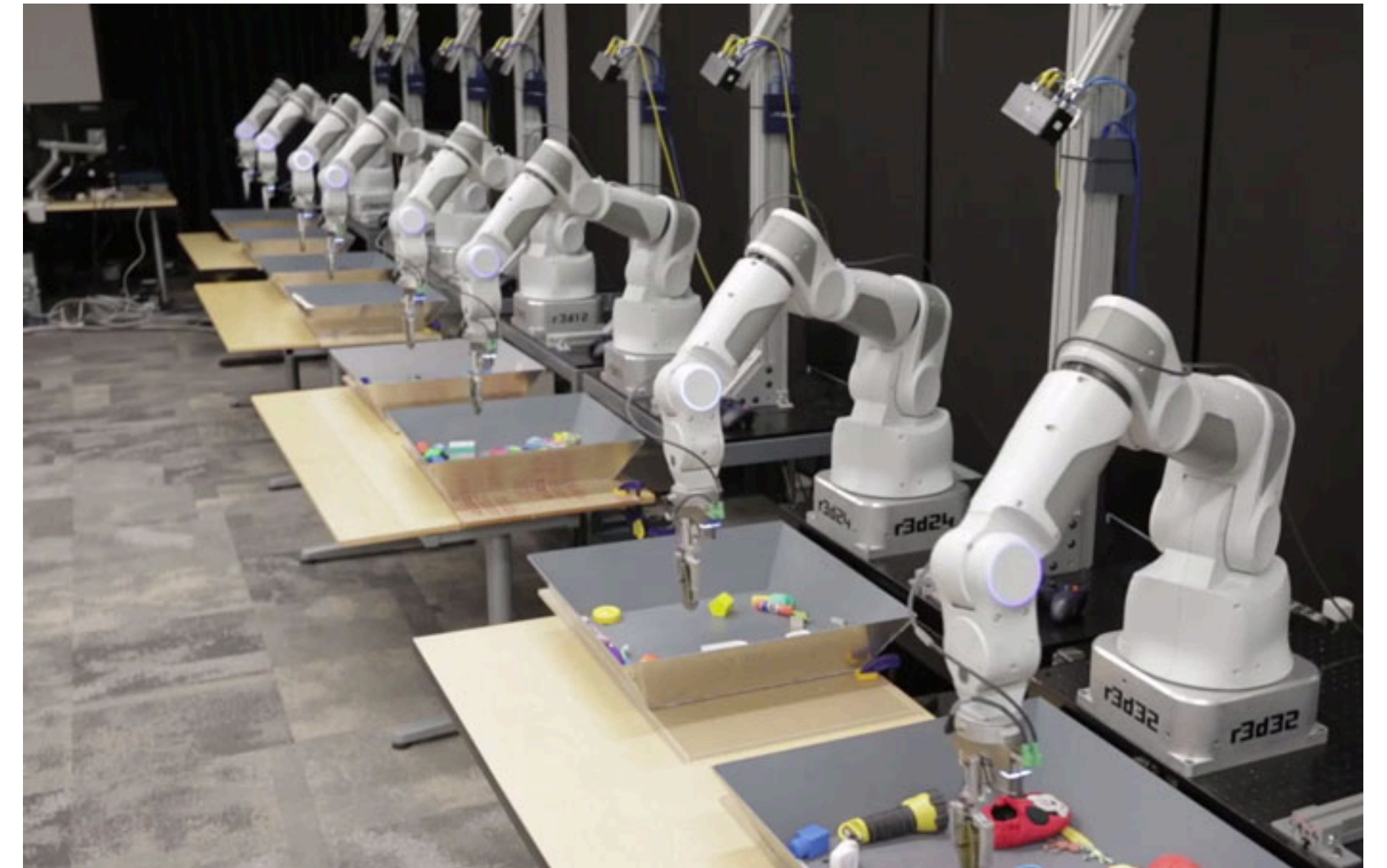
Learning to Plan and Make Decisions: Atari

- ▶ Agenda initiated by DeepMind to tackle AI via playing old Atari games.
- ▶ Intellectual concept: transfer learning across different games.
- ▶ Huge amount of experience required for competent play.
- ▶ Serious issues around reproducibility.
- ▶ Unclear how the results generalize.



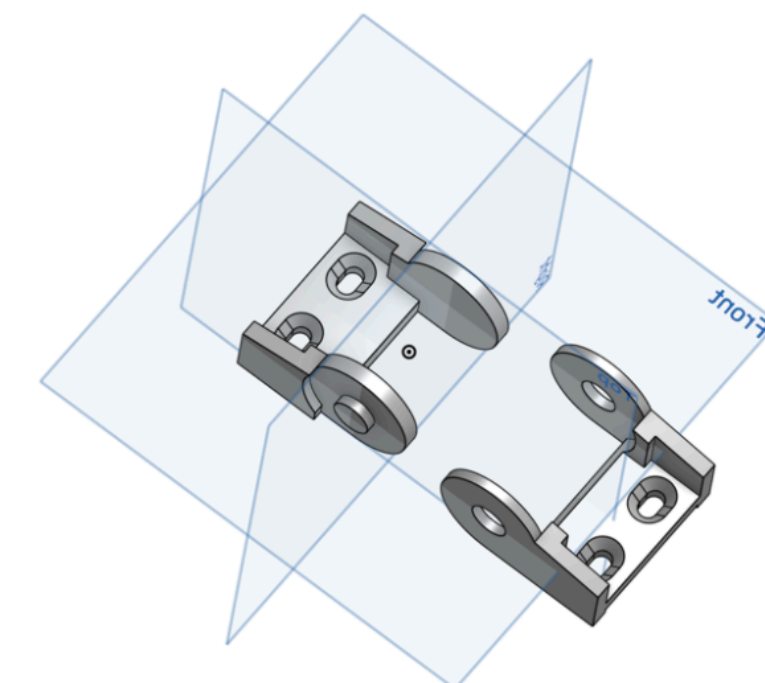
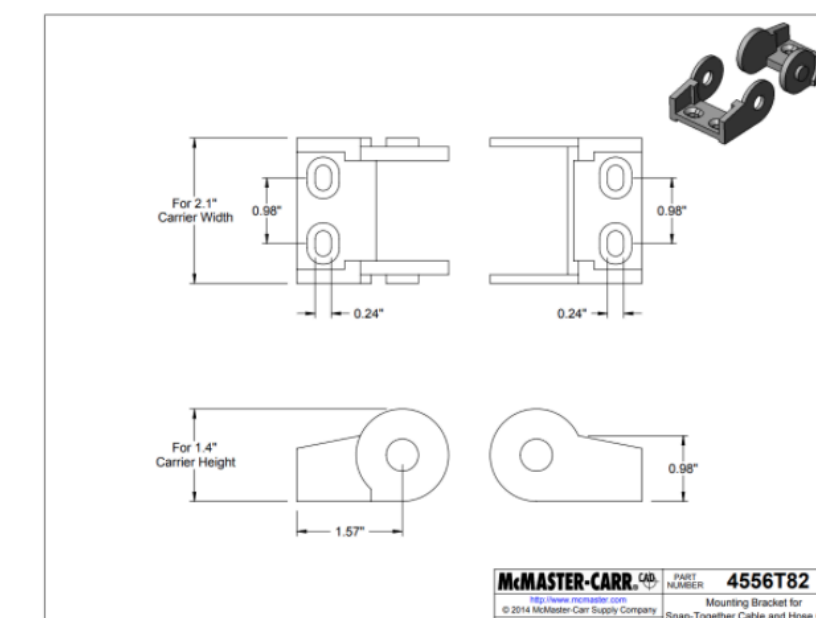
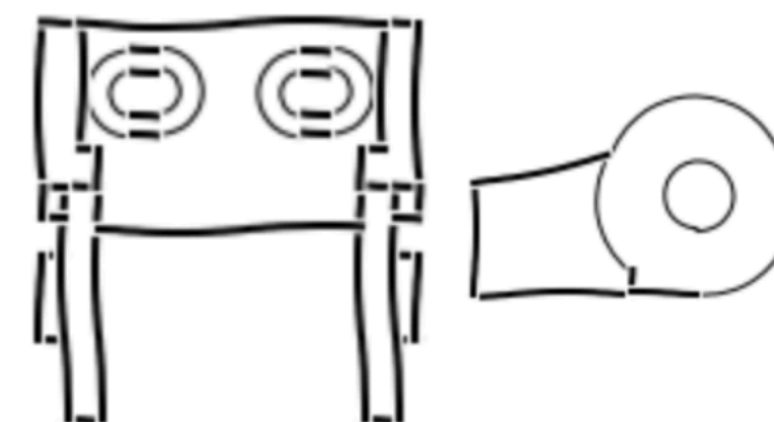
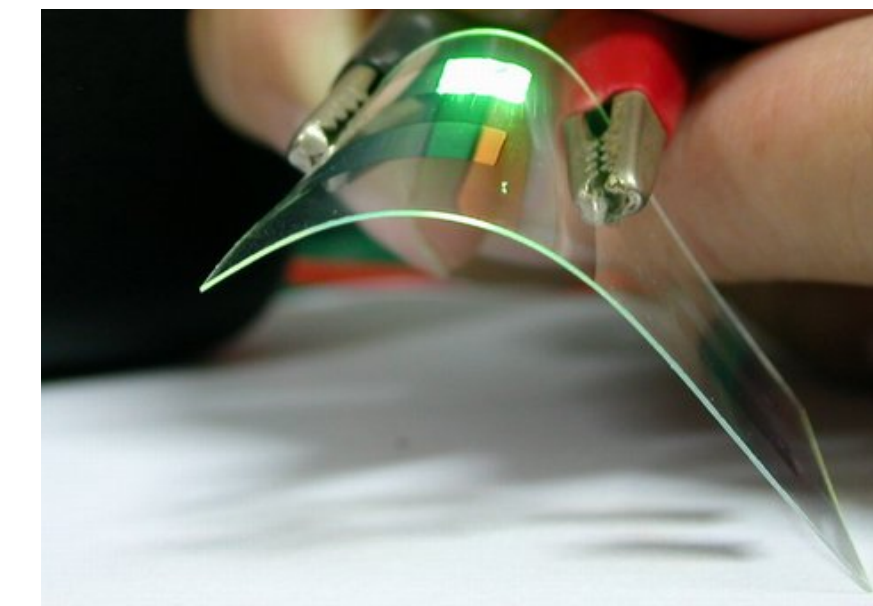
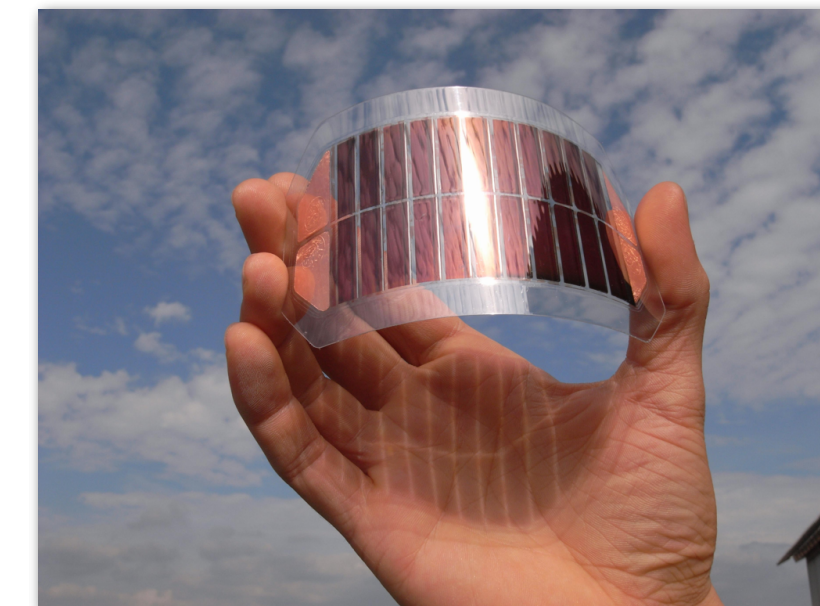
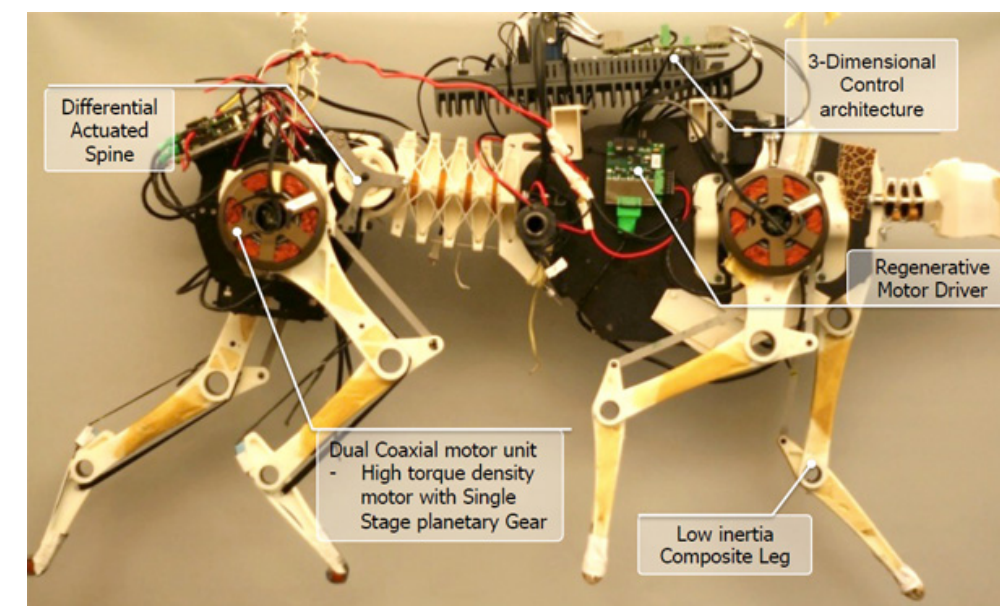
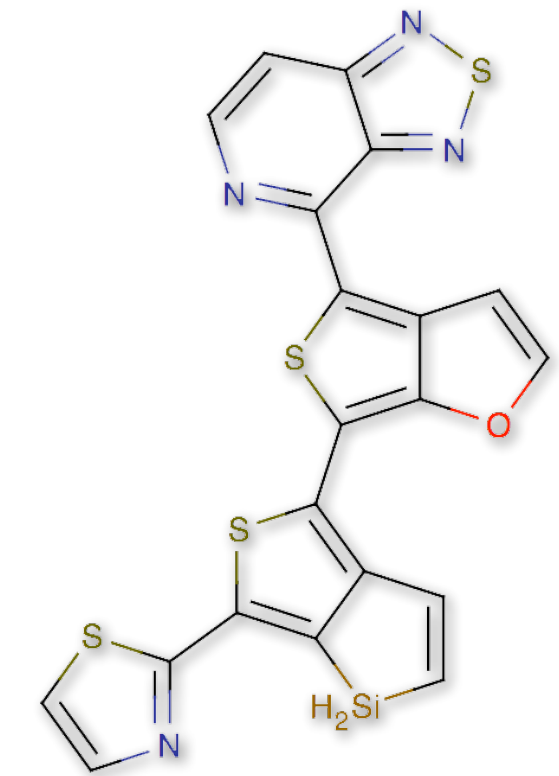
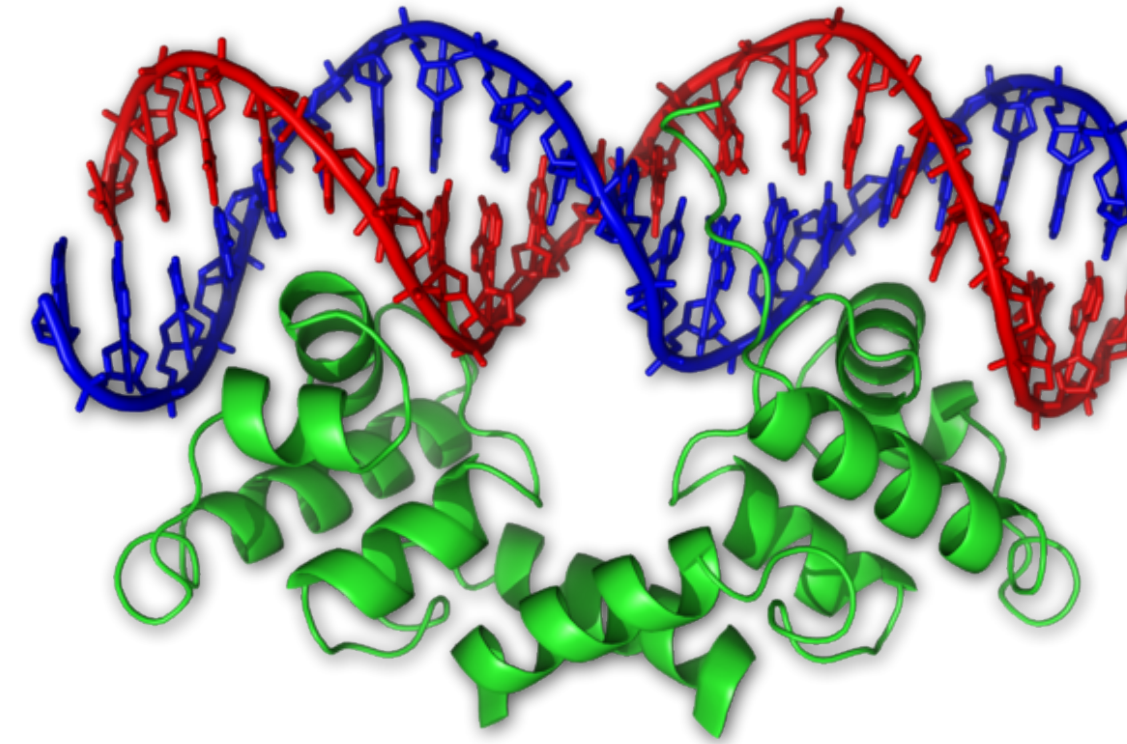
Learning to Plan and Make Decisions: Robotics

- ▶ Traditional domain for AI research
- ▶ Widely available simulators have enabled more researchers to study robotics.
- ▶ Emphasis on abstractions of high-level human tasks.
- ▶ For low-level problems still not competitive with traditional control approaches.
- ▶ Very few guarantees.



Learning to Plan and Make Decisions: Automating Design

- ▶ Iterative design is a sequential decision making process.
- ▶ Big opportunities for modeling and automation with ML.
- ▶ Successes in materials design:
 - Light emitting diodes
 - Photovoltaics
 - Fluorescent proteins
- ▶ Entering more traditional spaces in controls, mechanical & chemical engineering.



Why the Hype? What's Changed?

- ▶ Faster hardware: GPUs and ASICs
- ▶ Web-scale data
- ▶ Better tools for automatic differentiation
- ▶ Some new technical tricks
- ▶ Large investments by tech companies
- ▶ Some big problems have yielded to focused effort, e.g., AlphaGo

What About Artificial Intelligence?

Weak AI

- ▶ Designed to automate particular tasks, e.g., play Go, drive cars, recognize cats, predict ad clicks, recognize speech
- ▶ Often can meet or exceed human performance.
- ▶ Typically requires months or years of dedicated human effort.
- ▶ Deep learning has made it easier to roll some of these out.

Strong AI

- ▶ Anthropocentric: “Whatever humans can do”
- ▶ Highly unlikely without many new technical advances.
- ▶ Not somehow uniquely enabled by deep learning.