ML/Deep learning, 2018

Lyle Ungar

University of Pennsylvania

Deep learning thus far

- is data hungry

- is shallow and has limited capacity for transfer

- has no natural way to deal with hierarchical structure

- has struggled with open-ended inference

- is not sufficiently transparent

- has not been well integrated with prior knowledge

- cannot inherently distinguish causation from correlation

- presumes a largely stable world

- works well as an approximation, but often cannot be trusted

- is difficult to engineer

Gary Marcus

Hierarchical structure

The teenager who previously crossed the Atlantic set a record for flying around the world.

Dialog has structure at many time scales.

As do images, movies, animals, people, companies,…

Common sense

◆ Who is taller, Prince William or his baby son Prince George?

◆ Can you make a salad out of a polyester shirt?

◆ If you stick a pin into a carrot, does it make a hole in the carrot or in the pin?

Gary Marcus

Instability

Peyton Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by **John Elway**, who **led** the Broncos to victory in **Super Bowl XXXIII** at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV.

Google translate is very clever

◆ **the bat** 球棒

◆ **the bat ate** 蝙蝠吃了

but still unreliable

◆ the bat   el murciélago

◆ the bat ate   el bate comió

What's hot in ML?

- **Solving problems**
  - Chatbots, cancer diagnosis; Game playing with deep-RL
- *ICML 2018*
  - *… Circumventing Defenses to Adversarial Examples*
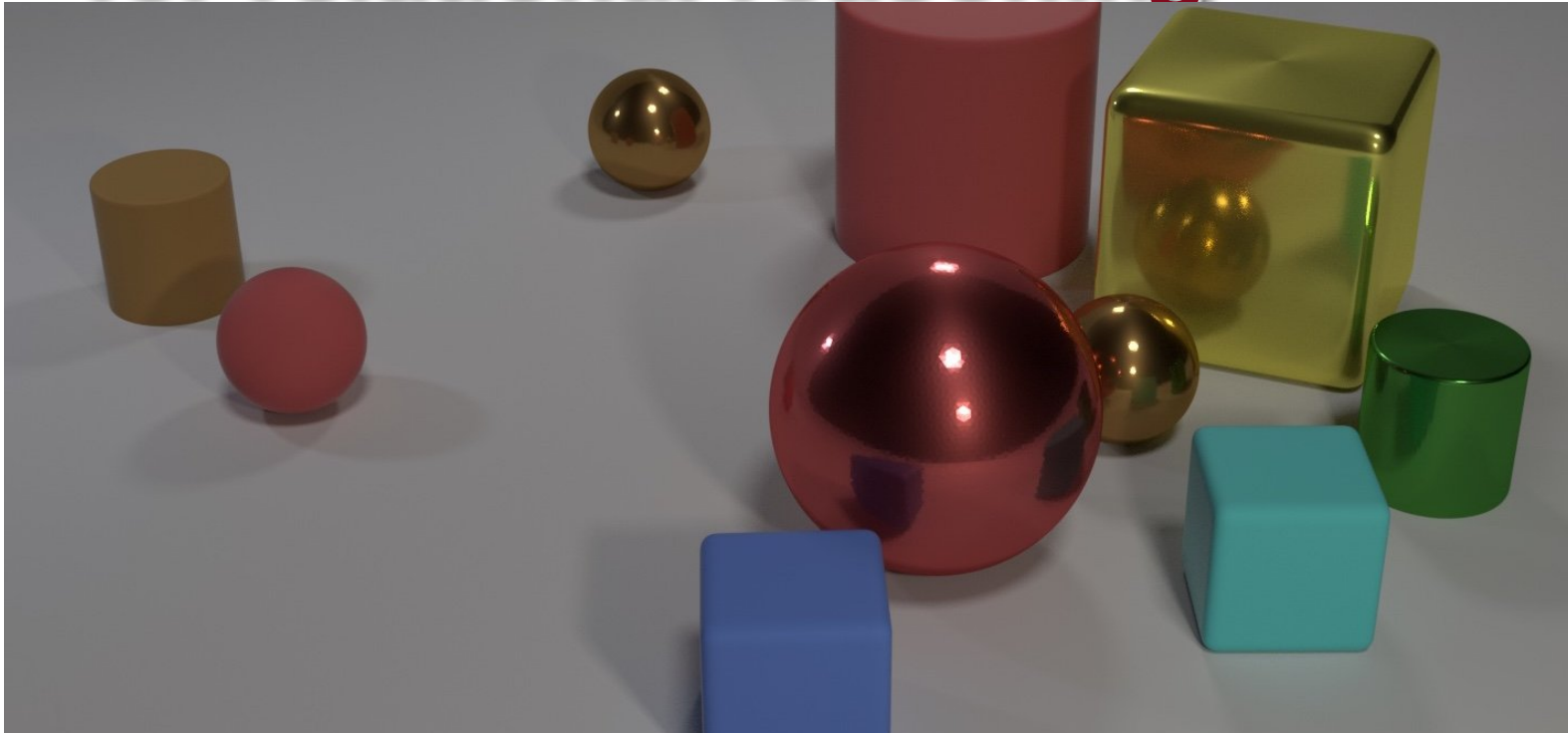 - *Delayed Impact of Fair Machine Learning*
- **NIPS 2018: music generation**
  https://magenta.tensorflow.org/demos/performance_rnn/index.html - 2|2,0,1,0,1,1,0,1,0,1,0,1|1,1,1,1,1,1,1,1,1,1,1,1|1,1,1,1,1,1,1,1,1,1,1,1|false
- **GANS (Generative Adversarial Networks)**
- **Why gradient descent does so well (theory!)**
- **AutoML**

A simple neural network module for relational reasoning

There is a tiny rubber thing that is the same colour as the large cylinder; what shape is it?

Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm

Starting from random play, and given no domain knowledge except the game rules, AlphaZero achieved within 24 hours a superhuman level of play in the games of chess and shogi (Japanese chess) as well as Go, and convincingly defeated a world-champion program in each case.

Lots of fancy network structures

input

outputs
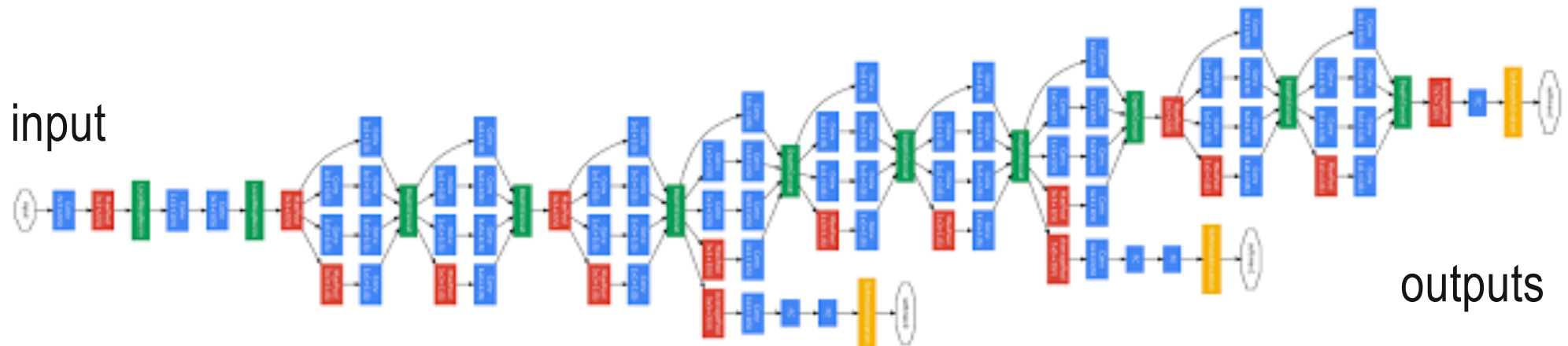
Convolutional (different sizes)
Or fully connected
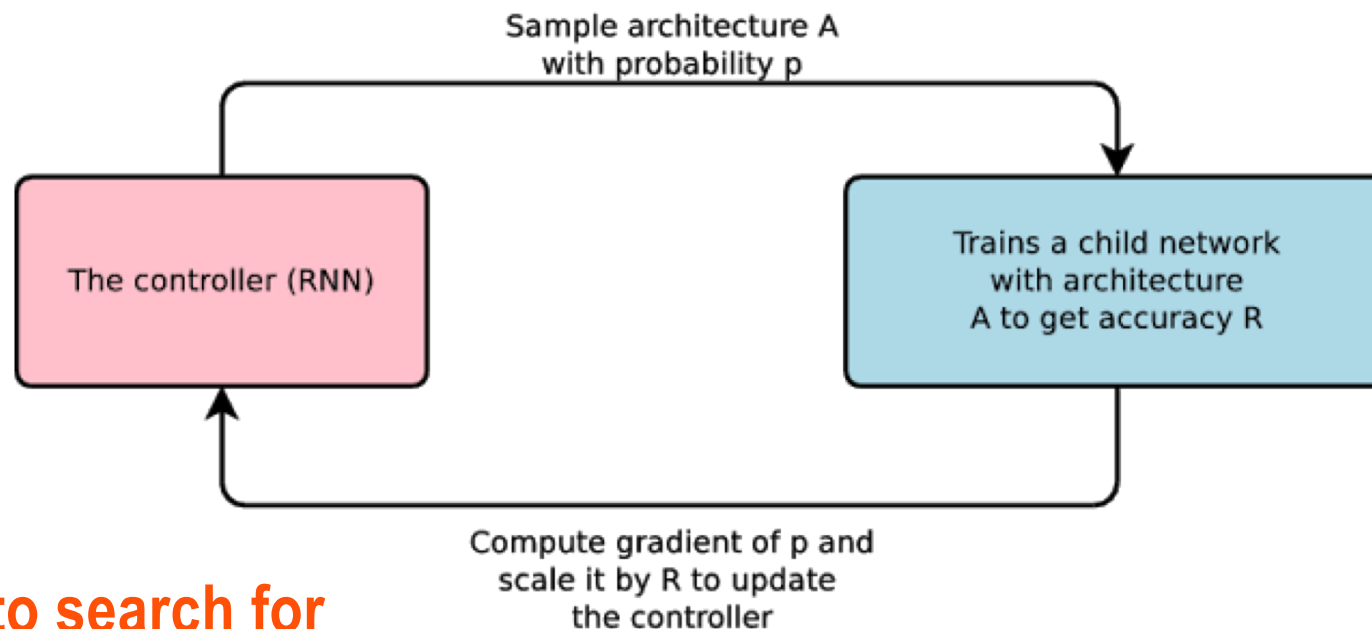**Maxpool**
Concatination
Softmax

Some layers
use dropout

googlenet

Design a really smart computer

## Using Machine Learning to Explore Neural Network Architecture

Sample architecture A
with probability p

The controller (RNN)

Trains a child network
with architecture
A to get accuracy R

Compute gradient of p and
scale it by R to update
the controller

Use RL to search for the 'best' neural net architecture

https://research.googleblog.com/2017/05/
using-machine-learning-to-explore.html

AutoML learns network structure
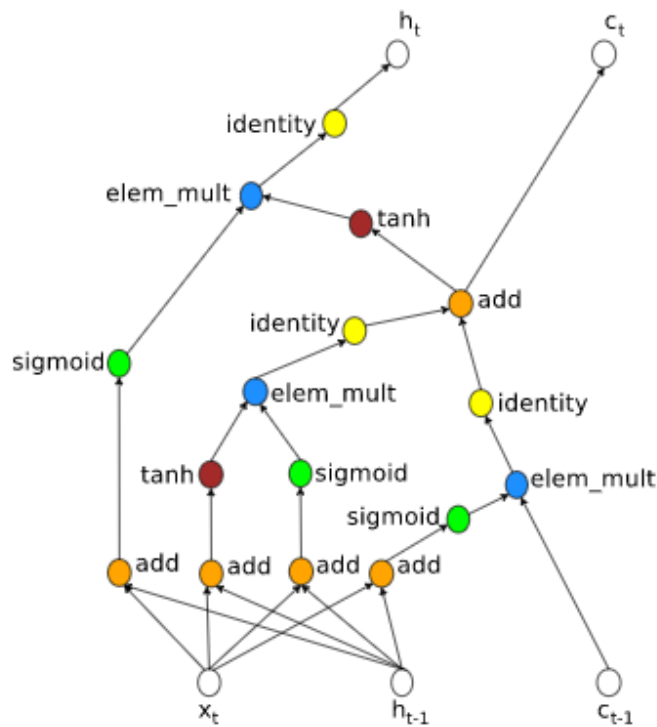
Human built   **Learned by RL**
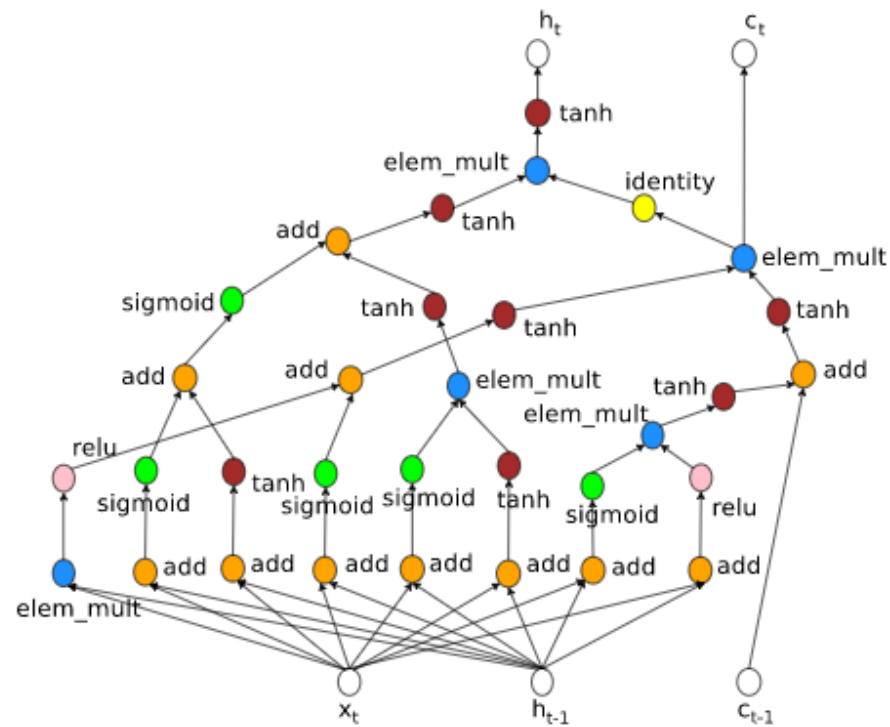
https://research.googleblog.com/2017/05/
using-machine-learning-to-explore.html

Auto-Sklearn

- ◆ **15 Classifiers**

- ◆ **14 feature preprocessing methods**

- ◆ **4 data preprocessing methods**

-> 110 hyperparameters

Combined Algorithm Selection and Hyperparameter (CASH) Optimization

Auto-Sklearn

◆ Uses **Bayesian optimization**

- Fit a (random forest) model between hyperparameters and performance and use it to find the optimum

  - speed up by discarding values that look bad on the first fold of 10-fold CV

◆ **Warmstart/Metalearning:** Start from hyperparameters that worked in the past for similar datasets.

- Based on 38 metafeatures of 140 datasets

◆ Use **Ensembles** of the 50 best classifiers considered

Auto-Sklearn

◆ **Performance (with limited CPU) was third among a large set of human competitors**

Big open directions (an opinion)

◆ **Multitask learning / domain transfer**

◆ **One shot learning**

  ● "See one do one"

◆ **Integrating deep learning with external data**

  ● e.g. results of database queries or search

◆ **Learning generalizable "deep structure"**

  ● Whatever that means?

Thank you!!!