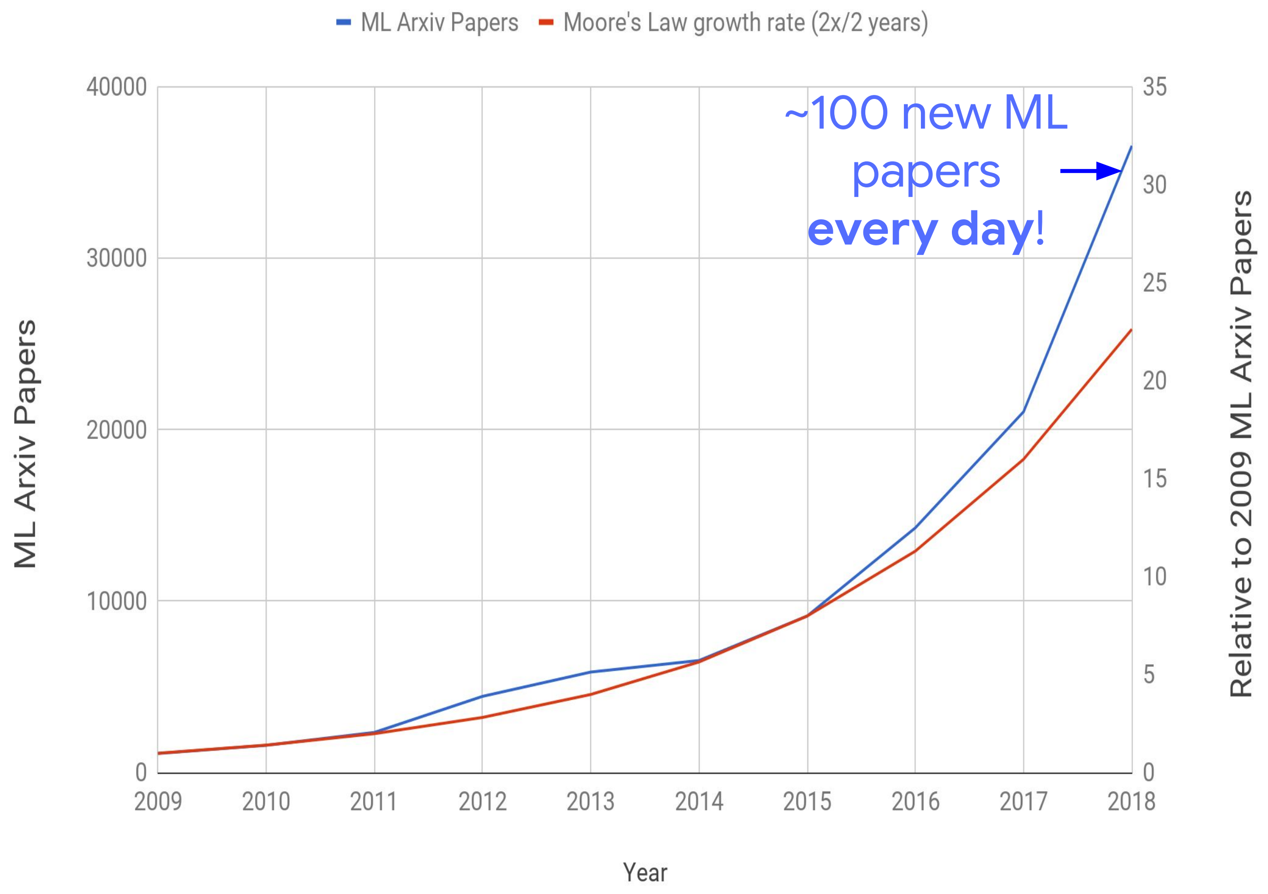
Google AI

Deep Learning to Solve Challenging Problems

Jeff Dean
Google Research & Machine Intelligence
@JeffDean
ai.google/research/people/jeff

Presenting the work of **many** people at Google

Machine Learning Arxiv Papers per Year



Deep Learning

Modern Reincarnation of Artificial Neural Networks

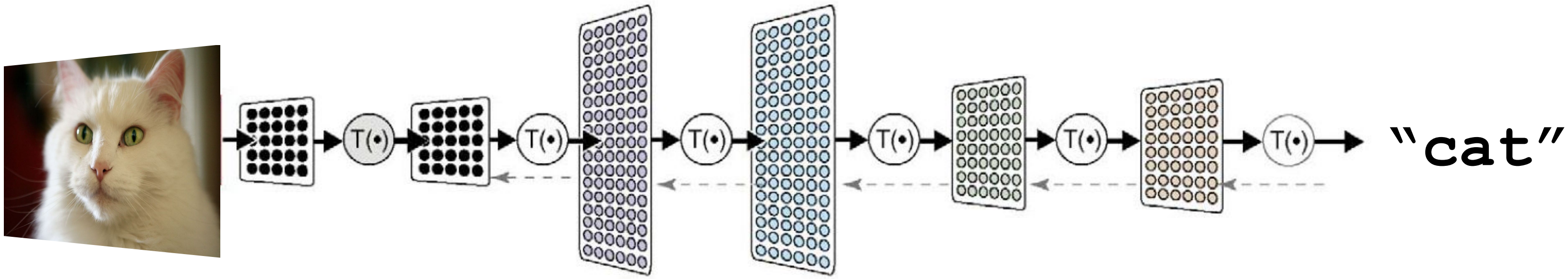
Collection of simple trainable mathematical units, organized in layers, that work together to solve complicated tasks

What's New

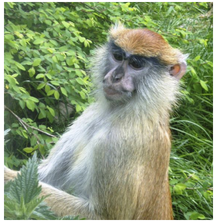
new network architectures,
new training math, **scale**

Key Benefit

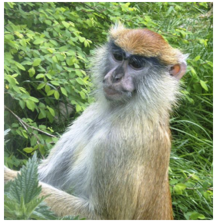
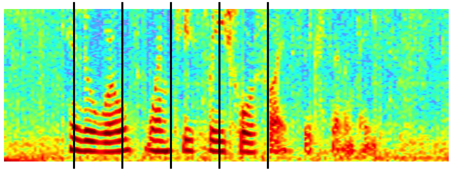
Learns features from raw, heterogeneous, noisy data
No explicit feature engineering required



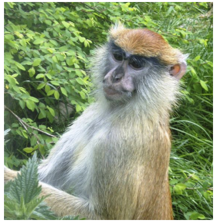
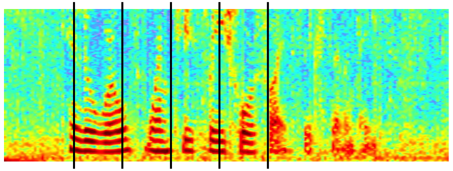
Functions a Deep Neural Network Can Learn

	input	output
Pixels:		`“hussar monkey”`

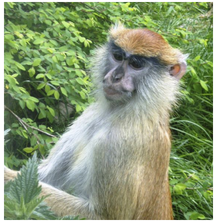
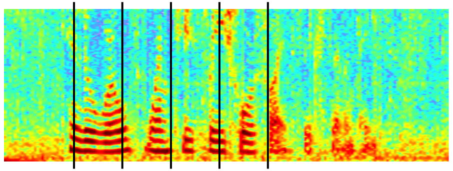
Functions a Deep Neural Network Can Learn

input	output
Pixels: 	"hussar monkey"
Audio: 	"How cold is it outside?"

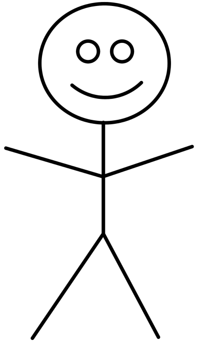
Functions a Deep Neural Network Can Learn

input	output
Pixels: 	"hussar monkey"
Audio: 	"How cold is it outside?"
"Hello, how are you?"	"Bonjour, comment allez-vous?"

Functions a Deep Neural Network Can Learn

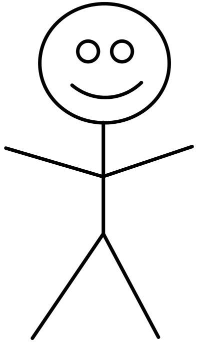
input		output
Pixels:		"hussar monkey"
Audio:		"How cold is it outside?"
"Hello, how are you?"		"Bonjour, comment allez-vous?"
Pixels:		"A blue and yellow train travelling down the tracks"

humans



5% errors

humans



5% errors

2008: U.S. National Academy of Engineering publishes

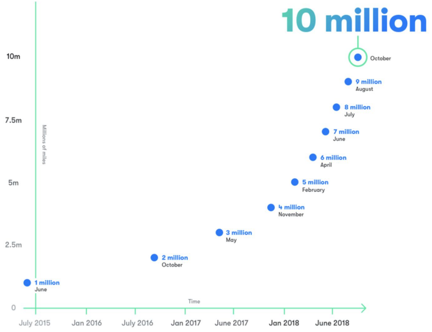
Grand Engineering Challenges for 21st Century

- Make solar energy affordable
- Provide energy from fusion
- Develop carbon sequestration methods
- Manage the nitrogen cycle
- Provide access to clean water
- **Restore & improve urban infrastructure**
- **Advance health informatics**
- Engineer better medicines
- Reverse-engineer the brain
- Prevent nuclear terror
- Secure cyberspace
- Enhance virtual reality
- Advance personalized learning
- **Engineer the tools for scientific discovery**

www.engineeringchallenges.org/challenges.aspx

Restore & improve urban infrastructure

WAYMO



10 million miles and counting

waymo.com/ontheroad/

Robots Can Pool Their Experience

Combining Vision with Robotics

~2015: 65% grasp success rate

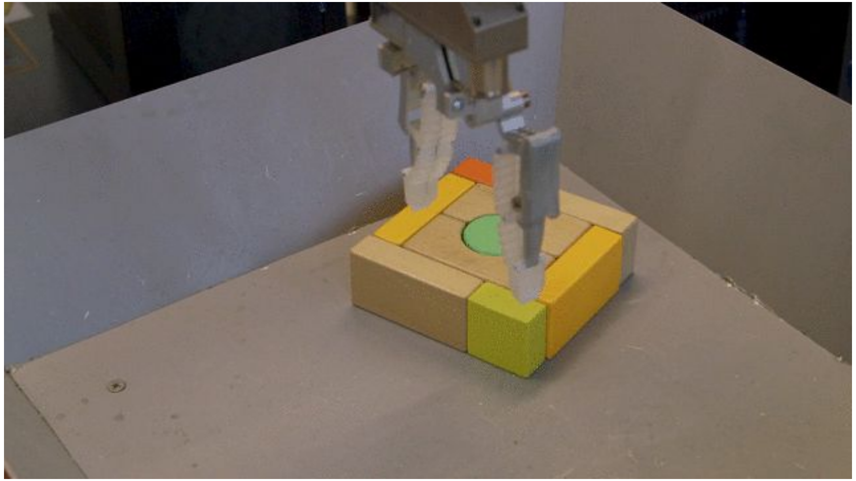
Combining Vision with Robotics

~2015: **65% grasp success rate**

2016: *"Deep Learning for Robots: Learning from Large-Scale Interaction"*, Google Research Blog, Mar. 2016

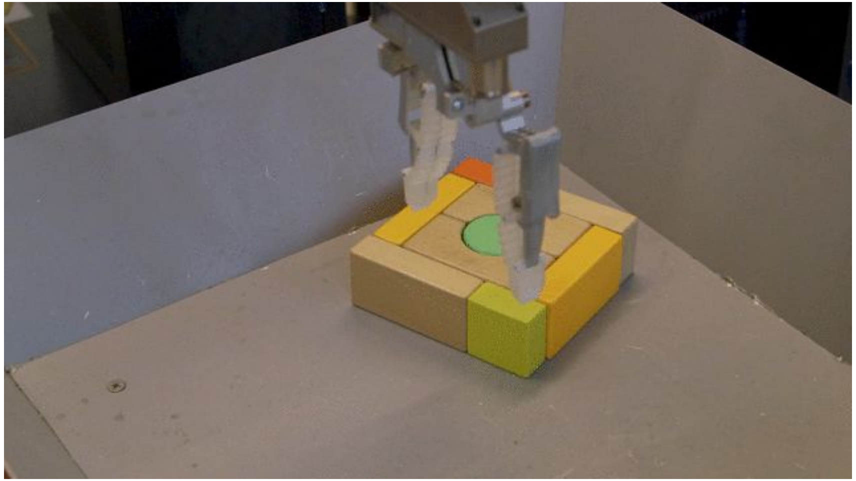
"Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection", Sergey Levine, Peter Pastor, Alex Krizhevsky, & Deirdre Quillen, arxiv.org/abs/1603.02199

78% grasp success rate



Combining Vision with Robotics

~2015: **65% grasp success rate**



2016: *“Deep Learning for Robots: Learning from Large-Scale Interaction”*, Google Research Blog, Mar. 2016

“Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection”, Sergey Levine, Peter Pastor, Alex Krizhevsky, & Deirdre Quillen, arxiv.org/abs/1603.02199

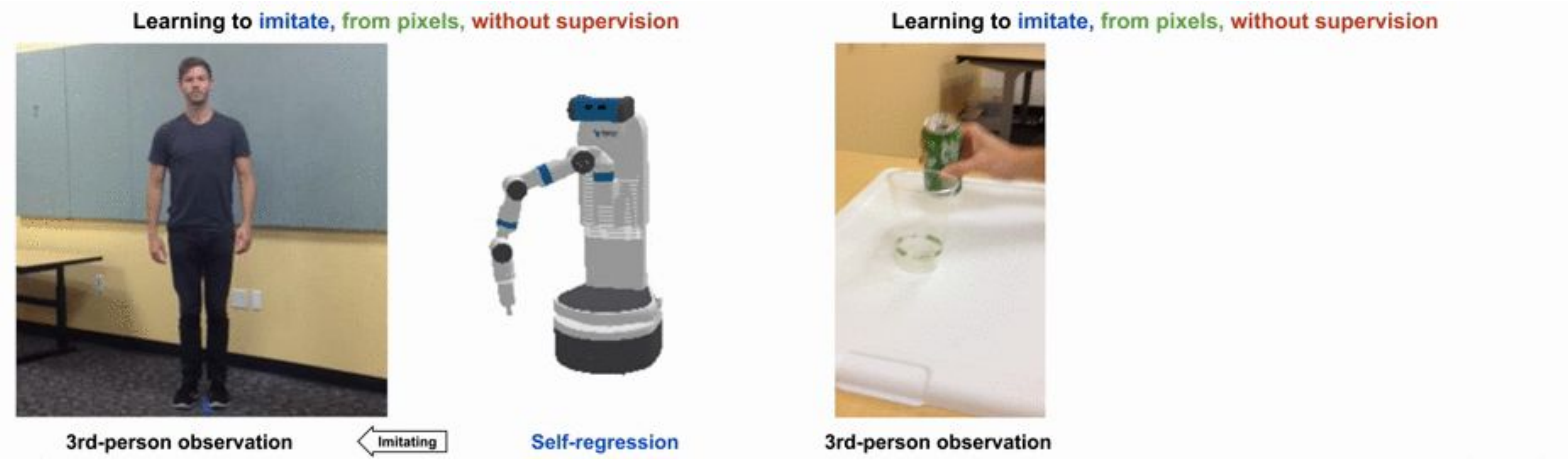
78% grasp success rate

2018: “Scalable Deep Reinforcement Learning for Robotic Manipulation”, Google AI Blog, June 2018

“*QT-Opt: Scalable Deep Reinforcement Learning for Vision-Based Robotic Manipulation*”, Kalashnikov, *et al.*, arxiv.org/abs/1806.10293

96% grasp success rate!

Self-Supervised Imitation Learning

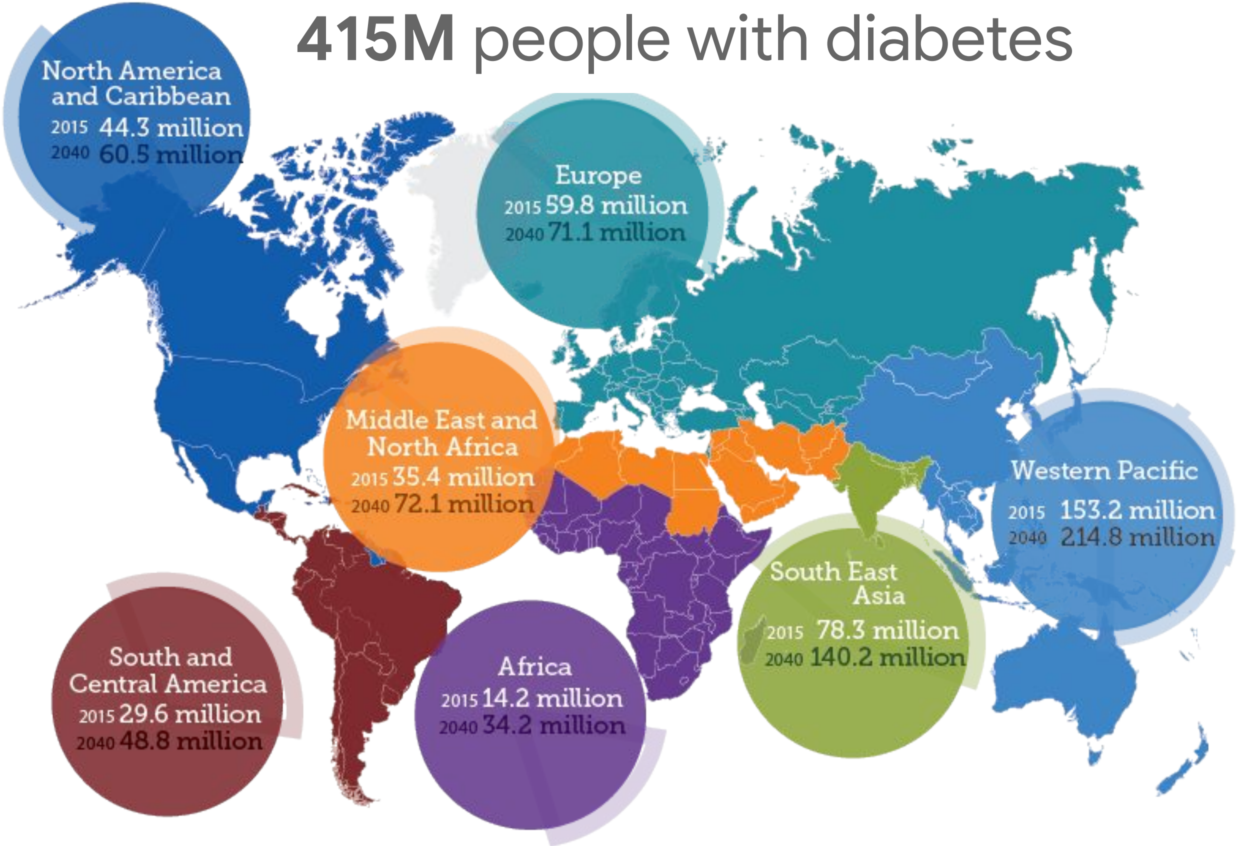


Time-Contrastive Networks: Self-Supervised Learning from Video, Pierre Sermanet, Corey Lynch, Yevgen Chebotar, Jasmine Hsu, Eric Jang, Stefan Schaal, and Sergey Levine.

See arxiv.org/abs/1704.06888 and sermanet.github.io/imitate

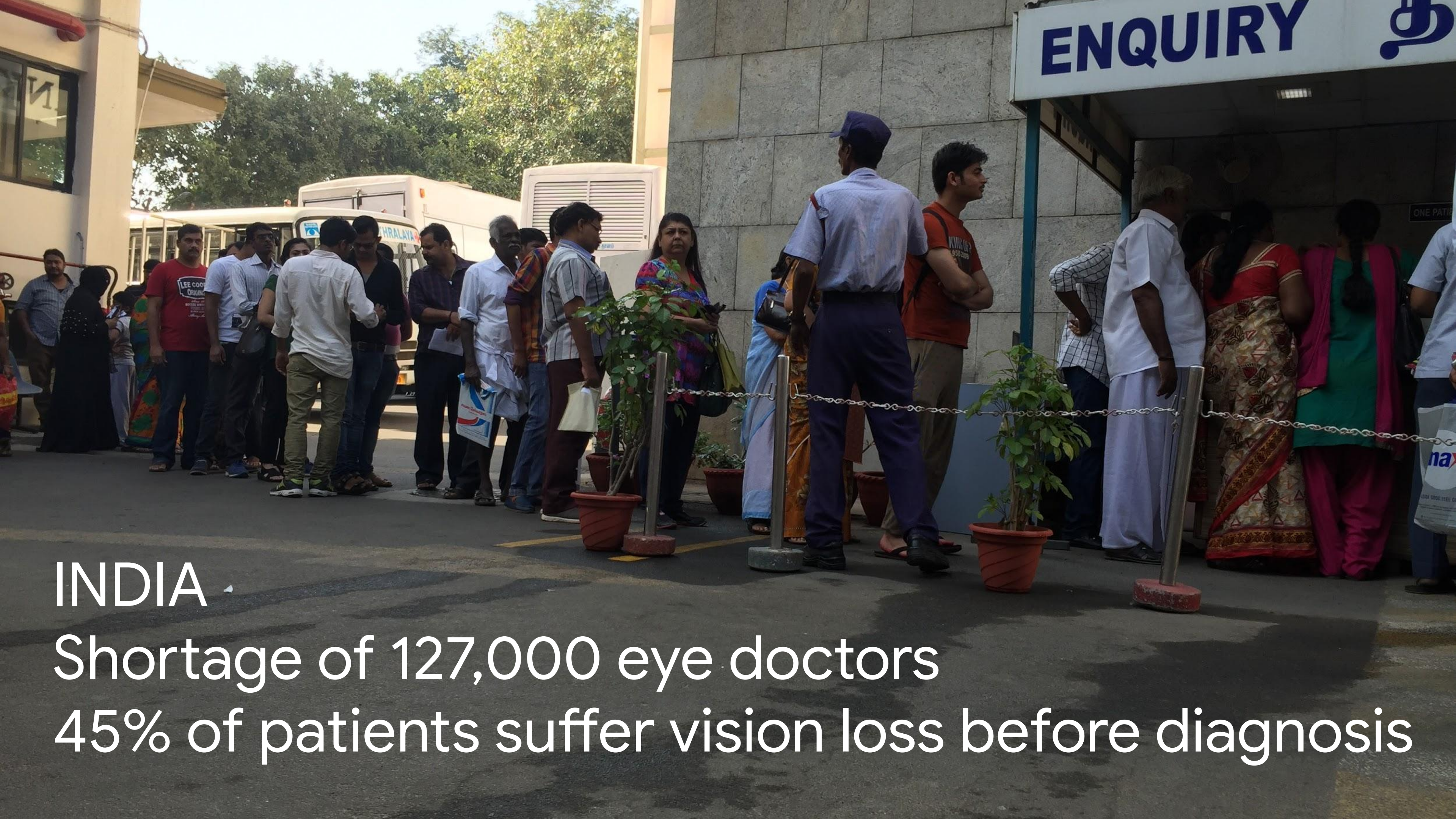
Advance health informatics

Diabetic retinopathy: fastest growing cause of blindness



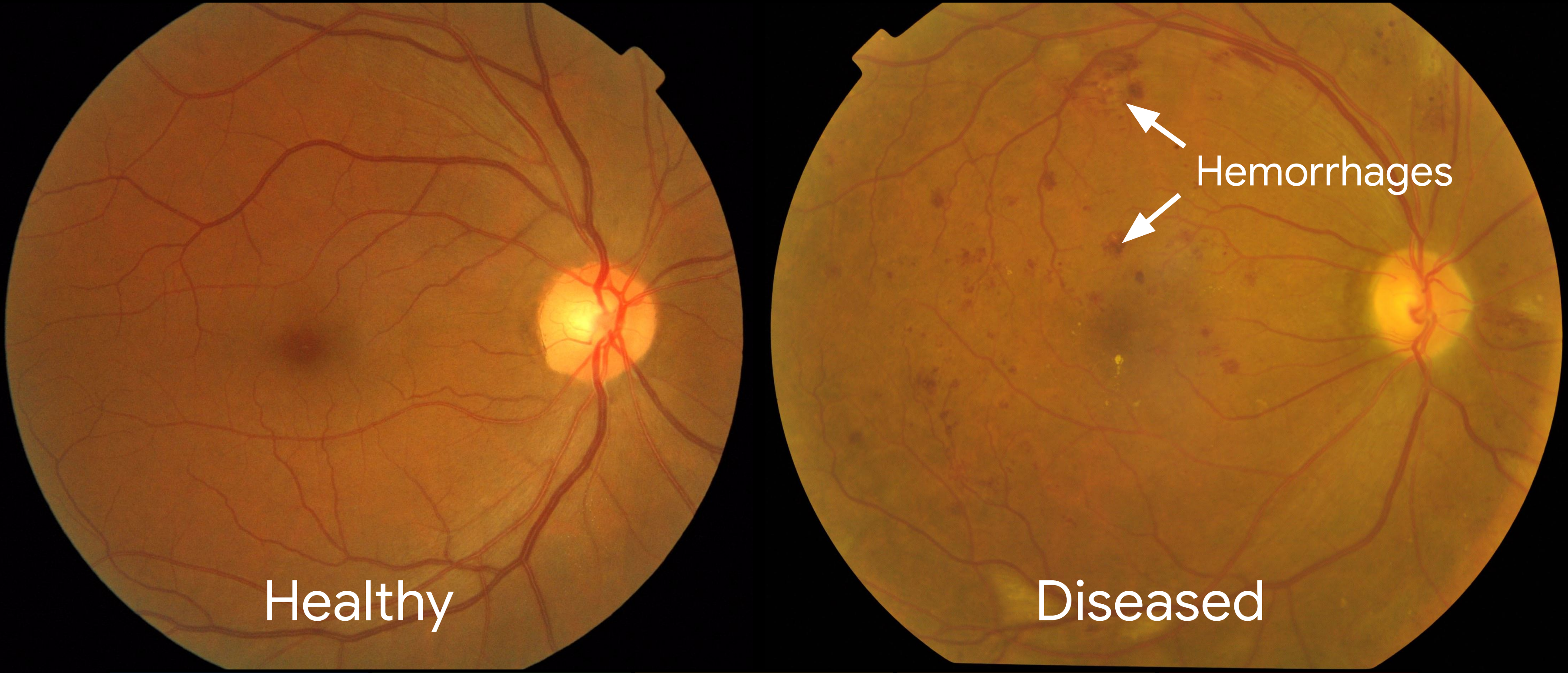
Regular screening is key to preventing blindness

=



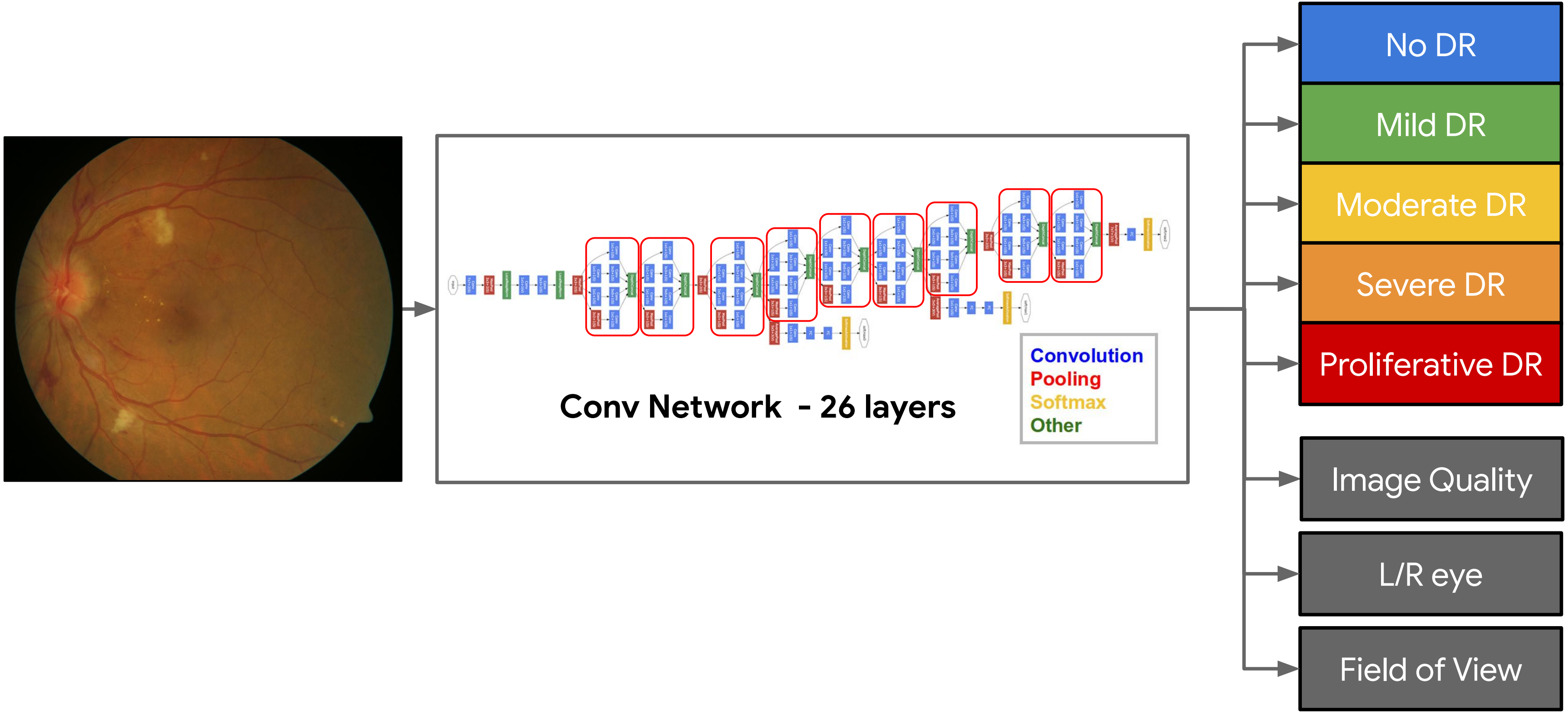
INDIA
Shortage of 127,000 eye doctors
45% of patients suffer vision loss before diagnosis

How DR is Diagnosed: Retinal Fundus Images



No DR	Mild DR	Moderate DR	Severe DR	Proliferative DR

Adapt deep neural network to read fundus images



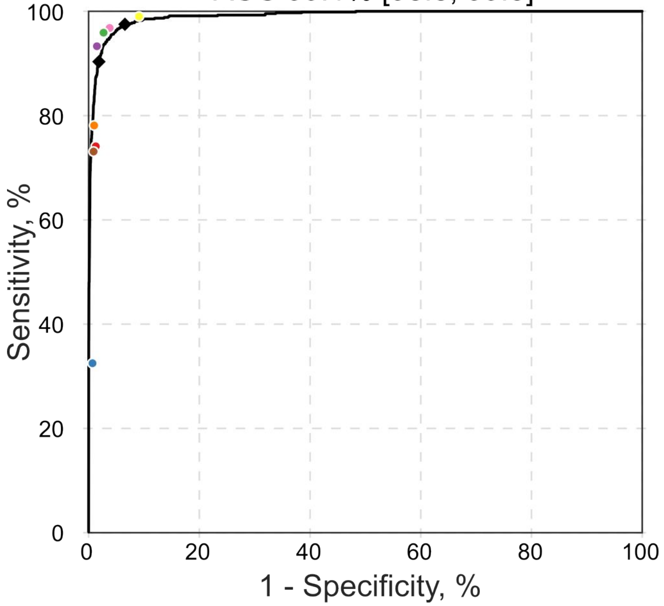
JAMA The Journal of the American Medical Association

JAMA | Original Investigation | INNOVATIONS IN HEALTH CARE DELIVERY

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

n = 9,963 images

AUC 99.1% [98.8, 99.3]



F-score

0.95	0.91
Algorithm	Ophthalmologist (median)

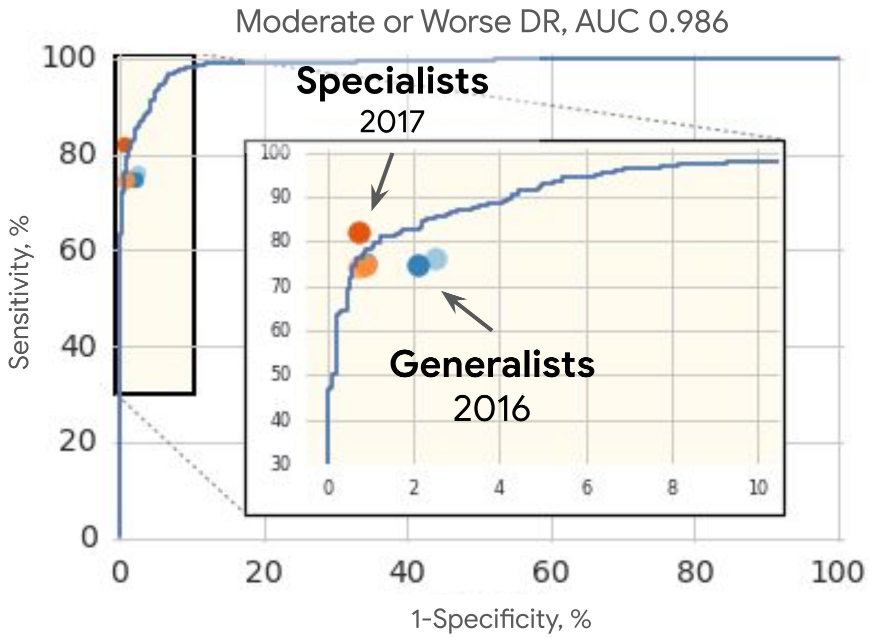
"The study by Gulshan and colleagues **truly represents the brave new world in medicine.**"

Dr. Andrew Beam, Dr. Isaac Kohane
Harvard Medical School

"Google just published this paper in JAMA (impact factor 37) [...] **It actually lives up to the hype.**"

Dr. Luke Oakden-Rayner
University of Adelaide

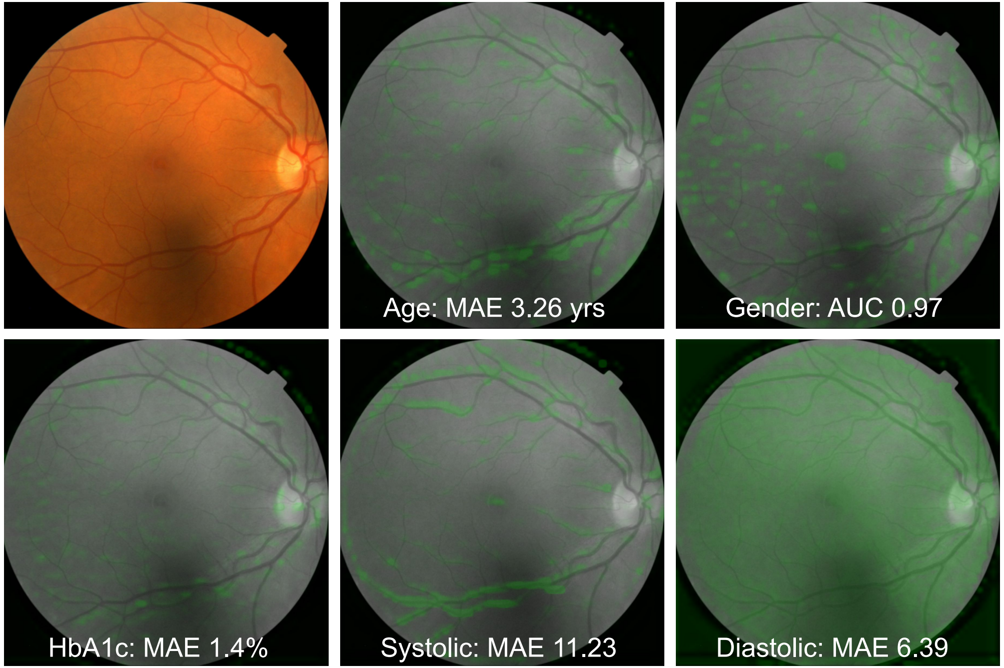
2016 - On Par with General Ophthalmologists
2017 - On Par with Retinal Specialist Ophthalmologists



	Weighted Kappa
Ophthalmologists Individual	0.80-0.84
Algorithm	0.84
Retinal Specialists Individual	0.82-0.91

Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. J. Krause, *et al.*, *Ophthalmology*, doi.org/10.1016/j.ophtha.2018.01.034

Completely new, novel scientific discoveries



Predicting things that doctors can't predict from imaging

—

Potential as a new biomarker

Preliminary 5-yr MACE AUC: 0.7

—

Can we predict cardiovascular risk? If so, this is a very nice non-invasive way of doing so

Can we also predict treatment response?

R. Poplin, A. Varadarajan *et al.* Predicting Cardiovascular Risk Factors from Retinal Fundus Photographs using Deep Learning. *Nature Biomedical Engineering, 2018.*

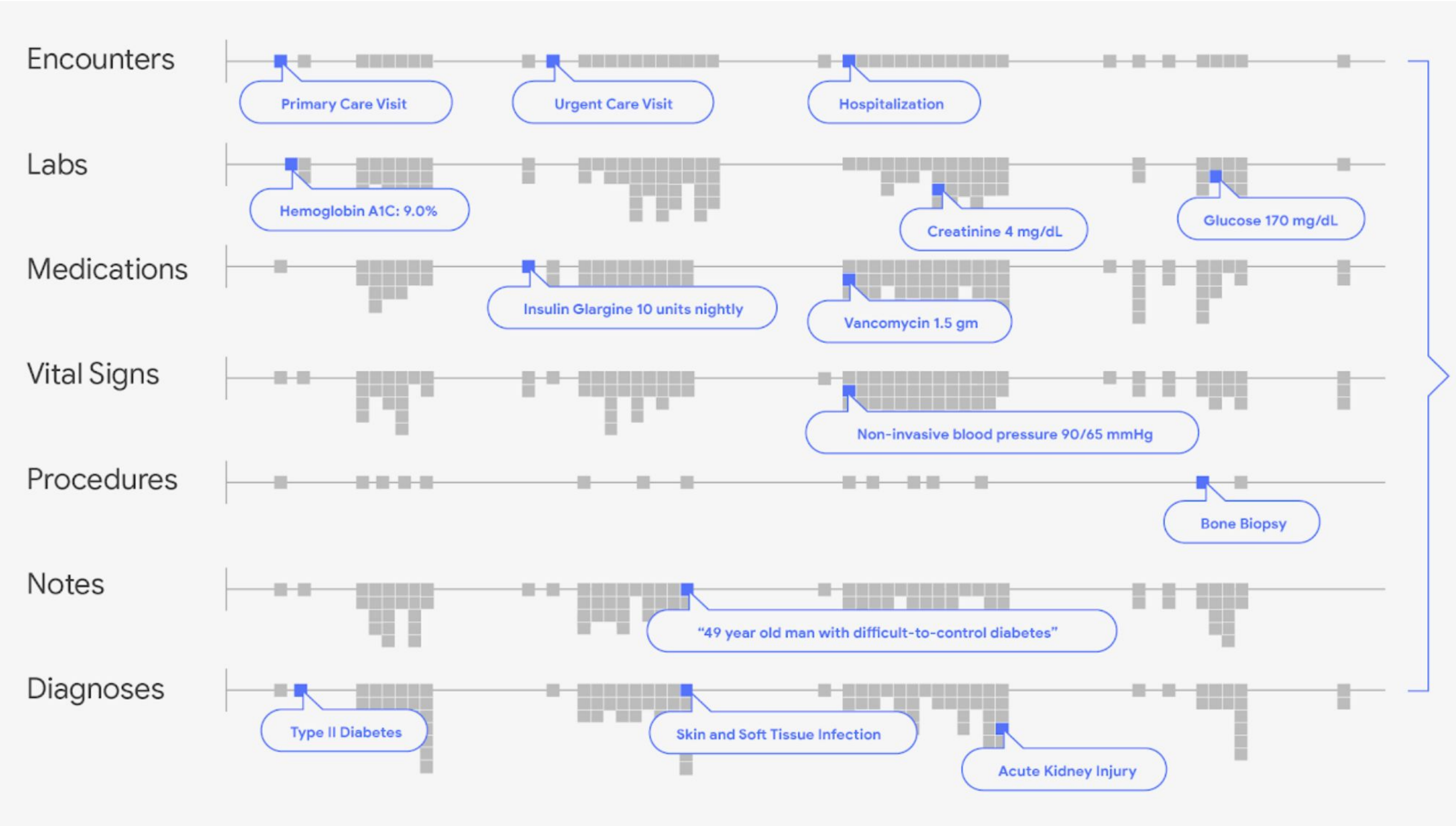
Predictive tasks for healthcare

Given a patient's electronic medical record data, **can we predict the future** and therefore deliver better care?

Deep learning methods for sequential prediction are becoming extremely good

e.g. recent improvements in Google Translation

Could you read the entire medical record?



Admitted to hospital?

Total medical expense?

Likely to develop diabetes?

Likely diagnoses?

Medications to consider?

npj | Digital Medicine

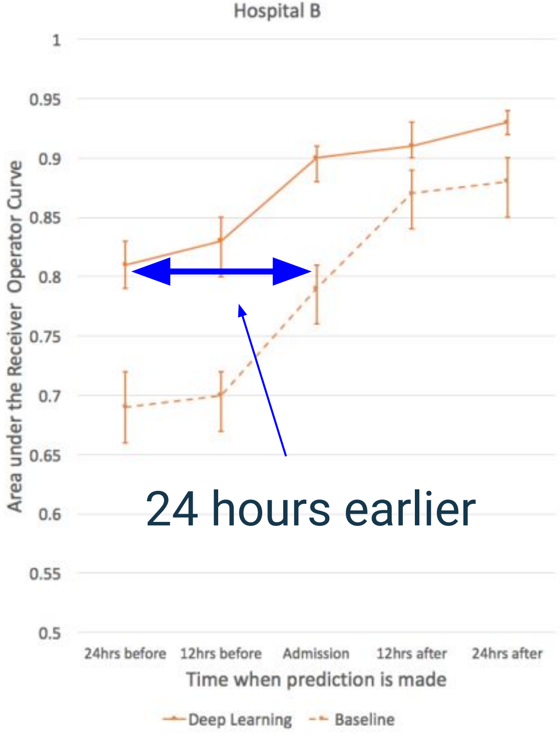
www.nature.com/npjdigitalmed

Mortality Risk Prediction Accuracy

ARTICLE OPEN

Scalable and accurate deep learning with electronic health records

Alvin Rajkomar[1,2], Eyal Oren[1], Kai Chen[1], Andrew M. Dai[1], Nissan Hajaj[1], Michaela Hardt[1], Peter J. Liu[1], Xiaobing Liu[1], Jake Marcus[1], Mimi Sun[1], Patrik Sundberg[1], Hector Yee[1], Kun Zhang[1], Yi Zhang[1], Gerardo Flores[1], Gavin E. Duggan[1], Jamie Irvine[1], Quoc Le[1], Kurt Litsch[1], Alexander Mossin[1], Justin Tansuwan[1], De Wang[1], James Wexler[1], Jimbo Wilson[1], Dana Ludwig[2], Samuel L. Volchenboum[3], Katherine Chou[1], Michael Pearson[1], Srinivasan Madabushi[1], Nigam H. Shah[4], Atul J. Butte[2], Michael D. Howell[1], Claire Cui[1], Greg S. Corrado[1] and Jeffrey Dean[1]



THE UNIVERSITY OF CHICAGO MEDICINE

UCSF University of California San Francisco

Stanford HEALTH CARE

Google

https://arxiv.org/abs/1801.07860 and https://www.nature.com/articles/s41746-018-0029-1

Engineer the Tools of Scientific Discovery

TensorFlow

http://tensorflow.org/

and

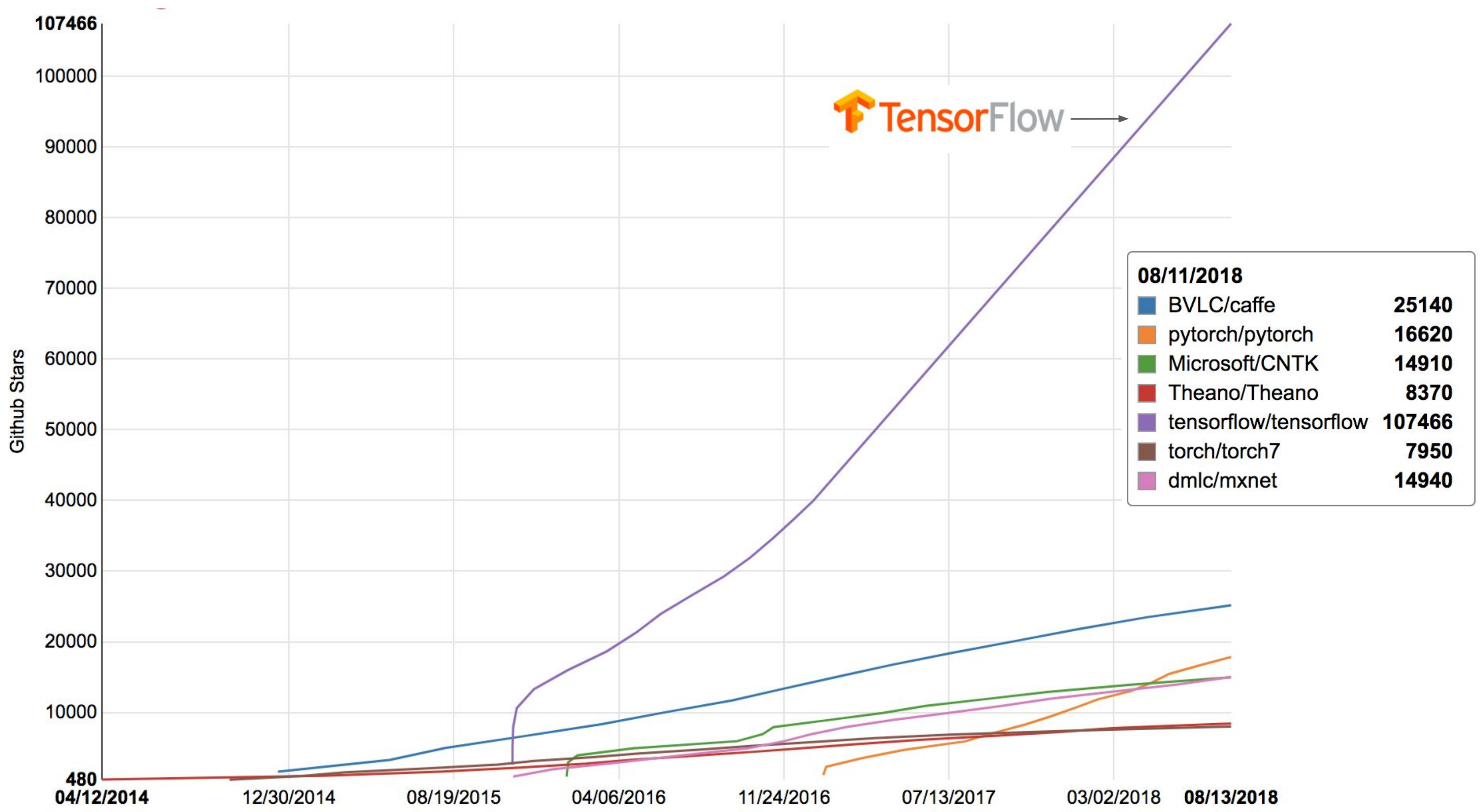
https://github.com/tensorflow/tensorflow

Open, standard software for general machine learning

Great for Deep Learning in particular

First released Nov 2015

Apache 2.0 license


TensorFlow
Github Stars
107466
100000
90000
80000
70000
60000
50000
40000
30000
20000
10000
480
04/12/2014
12/30/2014
08/19/2015
04/06/2016
11/24/2016
07/13/2017
03/02/2018
08/13/2018
08/11/2018
BVLC/caffe 25140
pytorch/pytorch 16620
Microsoft/CNTK 14910
Theano/Theano 8370
tensorflow/tensorflow 107466
torch/torch7 7950
dmlc/mxnet 14940

A vibrant **Open-Source Community**

Positive Reviews

115,000+
GitHub Stars

50,000+
GitHub repositories with 'TensorFlow' in the title

Rapid Development

1,800+
Contributors

47,000+
Commits in <40 months

Direct Engagement

10,000+
Stack Overflow questions answered

100+
Community-submitted GitHub issues responded to weekly

36,000,000+
Downloads

https://www.blog.google/topics/machine-learning/using-tensorflow-keep-farmers-happy-and-cows-healthy/

Deep Learning for Image-Based Cassava Disease Detection

Amanda Ramcharan,[1] Kelsee Baranowski,[1] Peter McCloskey,[2] Babuali Ahmed,[3] James Legg,[3] and David P. Hughes[1,4,5,*]

Penn State and International Institute of Tropical Agriculture

frontiers in Plant Science

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5663696/

AutoML: Automated machine learning
(“learning to learn”)

Current:

Solution = ML expertise + data + computation

Current:

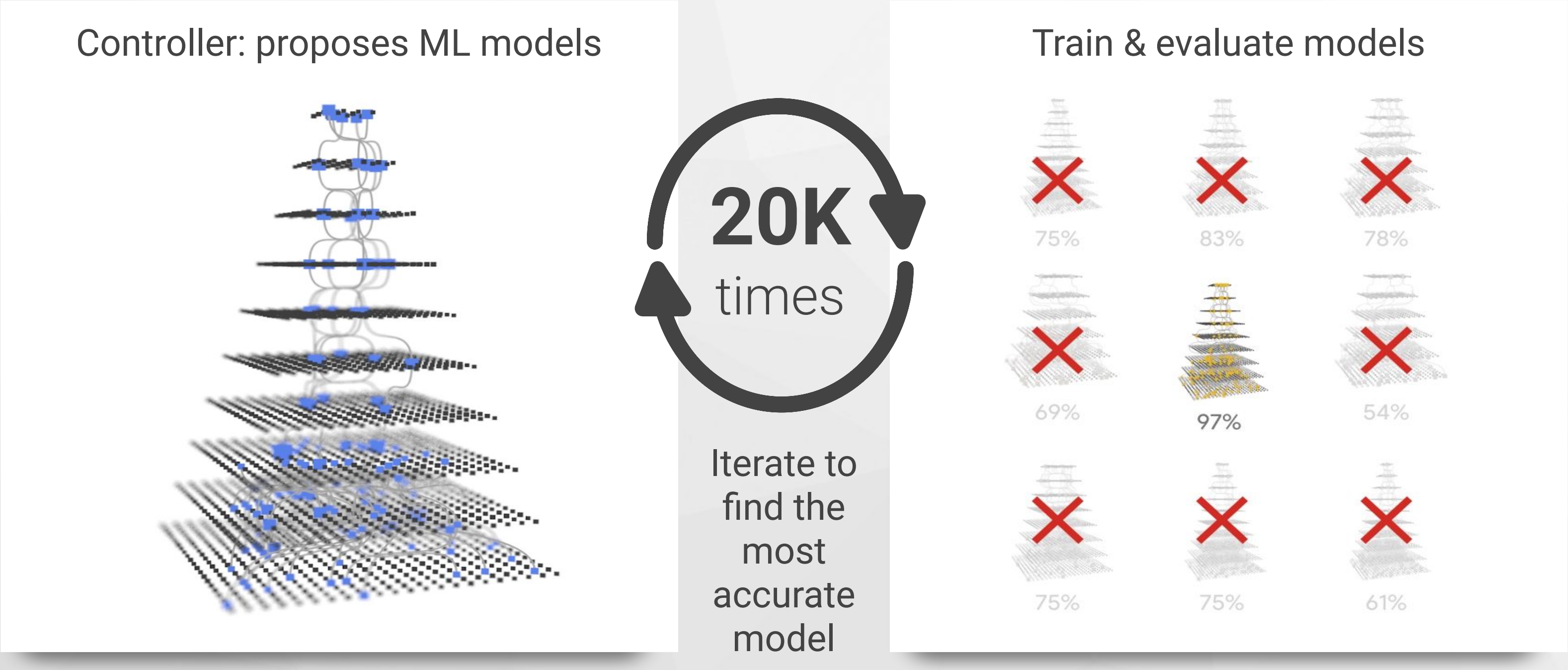
Solution = ML expertise + data + computation

Can we turn this into:

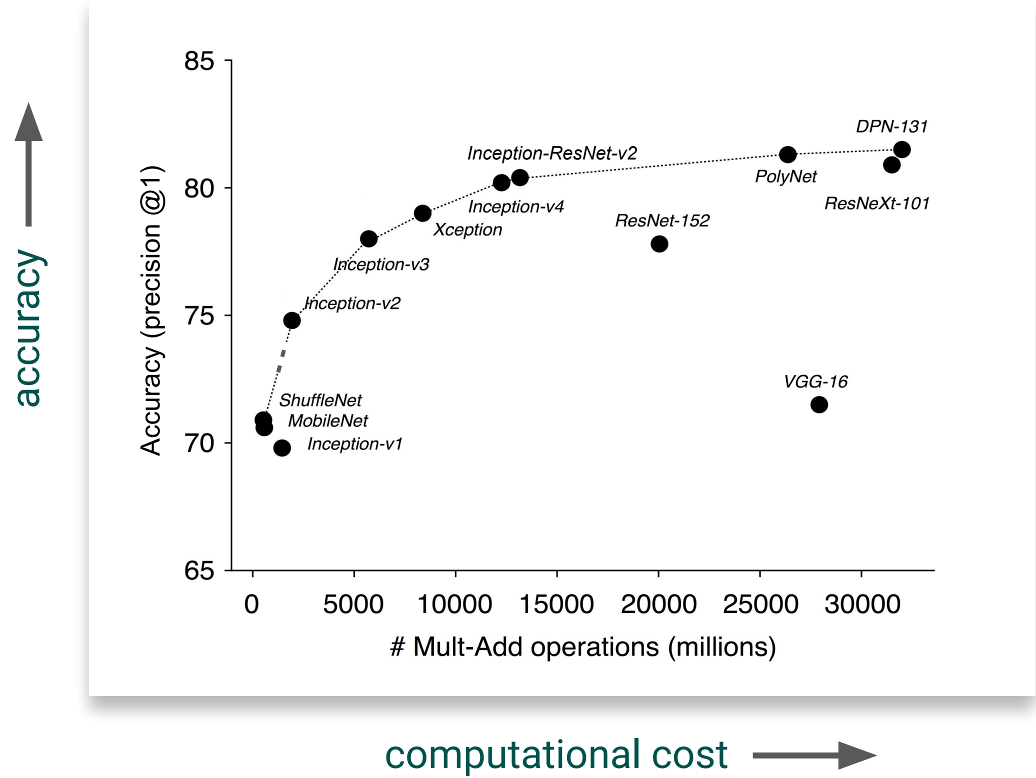
Solution = data + computation

???

Neural Architecture Search to find a model

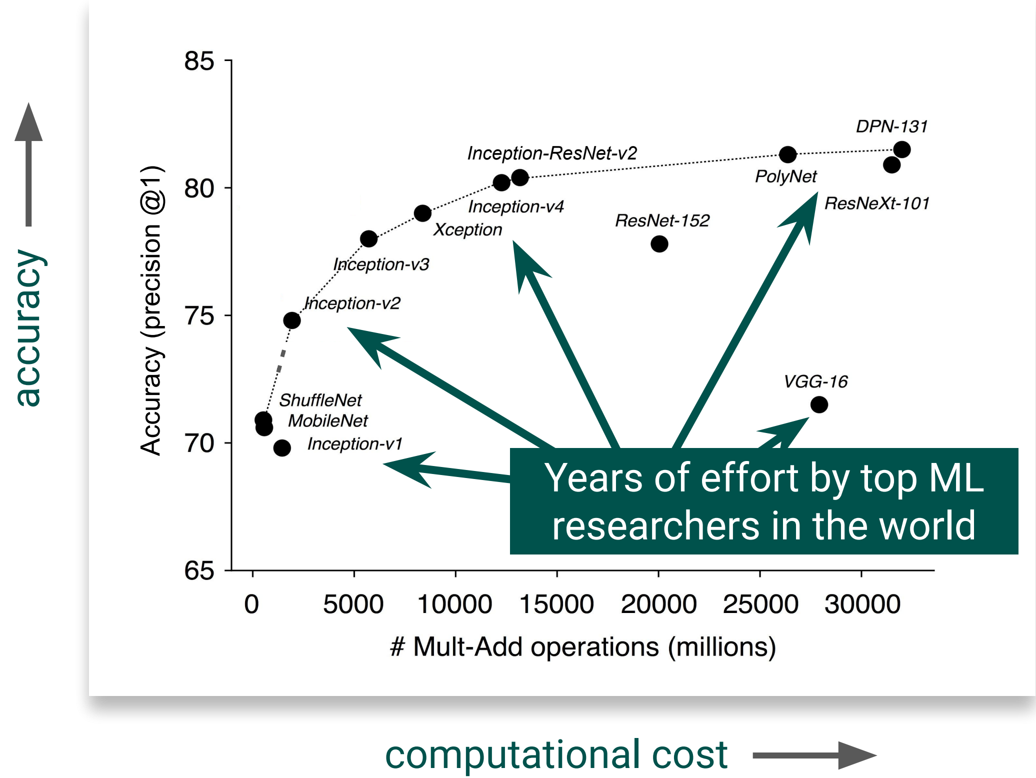


AutoML outperforms handcrafted models



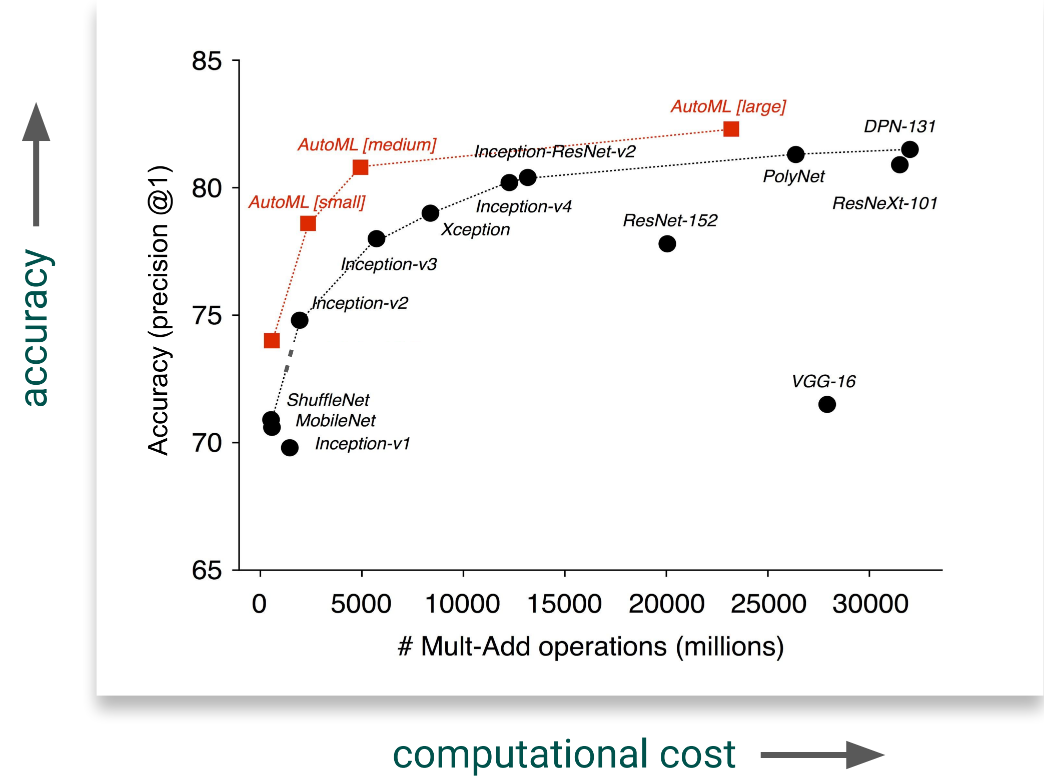
Learning Transferable Architectures for Scalable Image Recognition, Zoph et al. 2017, https://arxiv.org/abs/1707.07012

AutoML outperforms handcrafted models



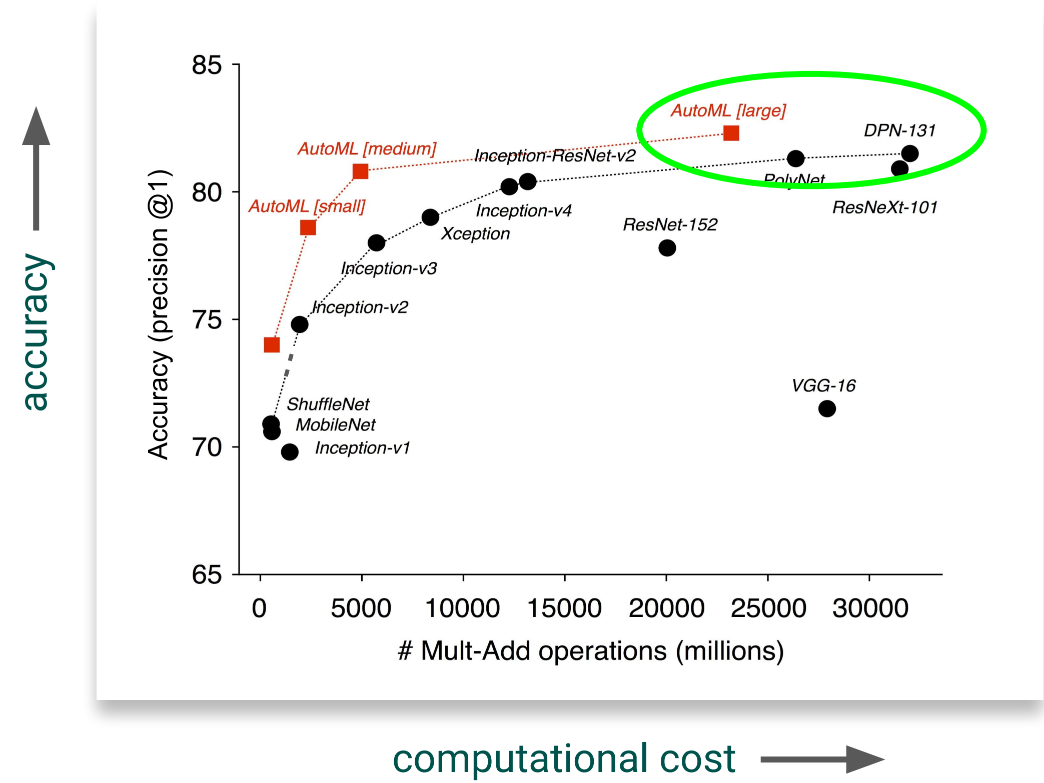
Learning Transferable Architectures for Scalable Image Recognition, Zoph et al. 2017,
https://arxiv.org/abs/1707.07012

AutoML outperforms handcrafted models



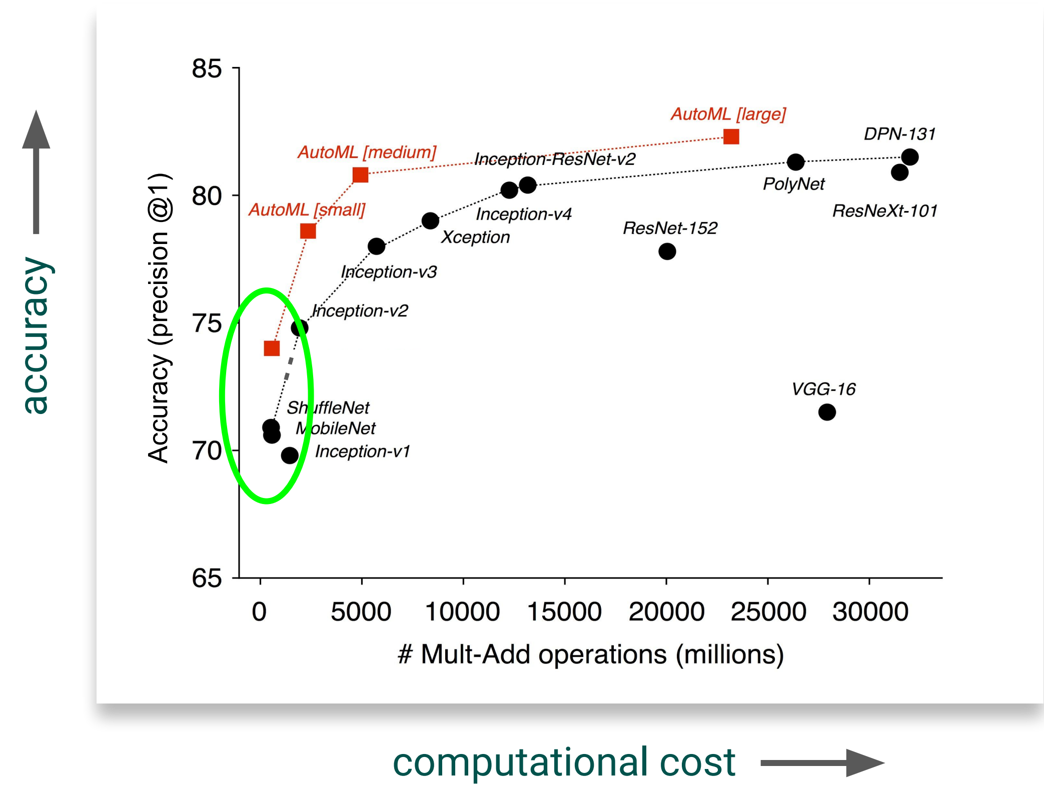
Learning Transferable Architectures for Scalable Image Recognition, Zoph et al. 2017,
https://arxiv.org/abs/1707.07012

AutoML outperforms handcrafted models



Learning Transferable Architectures for Scalable Image Recognition, Zoph et al. 2017, https://arxiv.org/abs/1707.07012

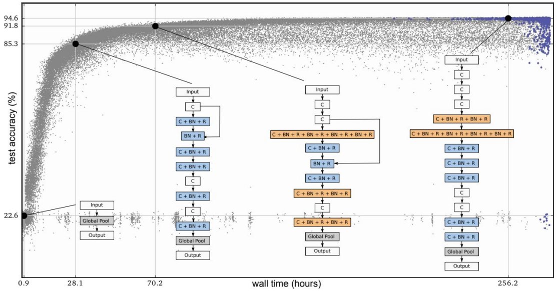
AutoML outperforms handcrafted models



Learning Transferable Architectures for Scalable Image Recognition, Zoph et al. 2017, https://arxiv.org/abs/1707.07012

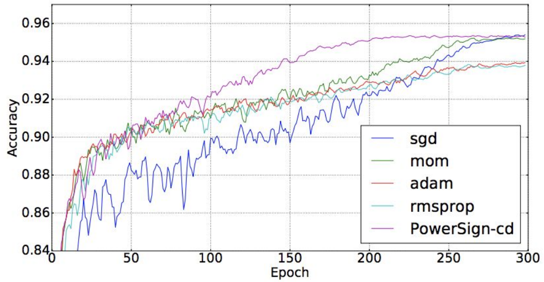
Additional Work in AutoML

Evolution for search rather than reinforcement learning:
Regularized Evolution for Image Classifier Architecture Search,
Esteban Real, Alok Aggarwal, Yanping Huang, Quoc V Le,
https://arxiv.org/abs/1802.01548



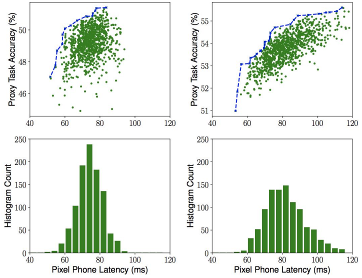
Large-Scale Evolution of Image Classifiers,
Esteban Real, Sherry Moore, Andrew Selle, Saurabh Saxena, Yutaka Leon Suematsu, Jie Tan, Quoc Le, Alex Kurakin
https://arxiv.org/abs/1703.01041

Learn the optimization update rule:
Neural Optimizer Search with Reinforcement Learning,
Irwan Bello, Barret Zoph, Vijay Vasudevan, Quoc V. Le,
https://arxiv.org/abs/1709.07417



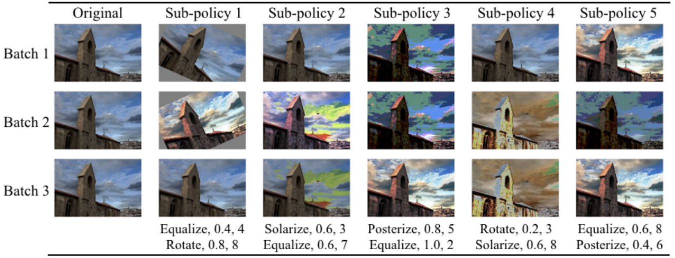
Additional Work in AutoML (cont)

Incorporate inference latency & accuracy into reward:
MnasNet: Platform-Aware Neural Architecture Search for Mobile,
Mingxing Tan, Bo Chen, Ruoming Pang, Vijay Vasudevan,
Quoc V. Le, https://arxiv.org/abs/1807.11626

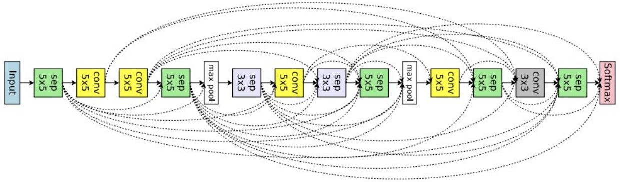


(a) $\alpha = 0, \beta = -1$ (b) $\alpha = \beta = -0.07$

Learn data augmentation policies:
AutoAugment: Learning Augmentation Policies from Data,
Ekin D. Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan,
Quoc V. Le, https://arxiv.org/abs/1805.09501



Explore many architectures simultaneously w/ parameter sharing:
Efficient Neural Architecture Search via Parameters Sharing In Deep Learning,
Hieu Pham, Melody Guan, Barret Zoph, Quoc Le, Jeff Dean
https://arxiv.org/abs/1802.03268



More computational power needed

Deep learning is transforming how we design computers

Special computation properties

reduced precision ok

about 1.2	1.21042
× about 0.6	× 0.61127
about 0.7	0.73989343

NOT

Special computation properties

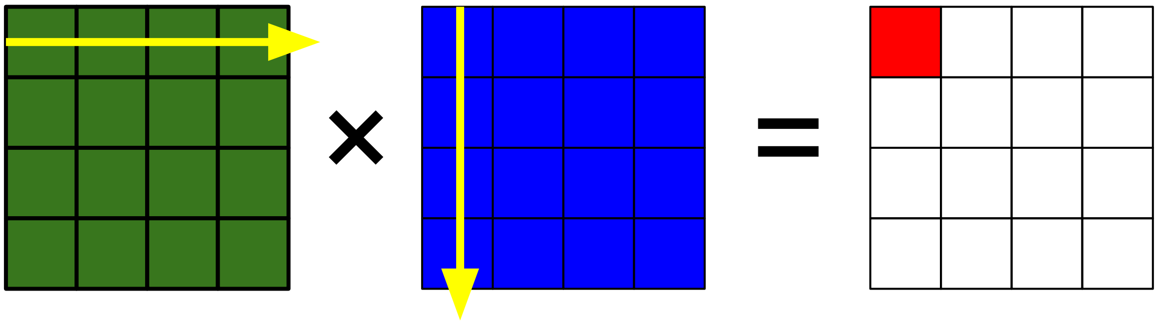
reduced precision ok

about 1.2
× about 0.6
about 0.7

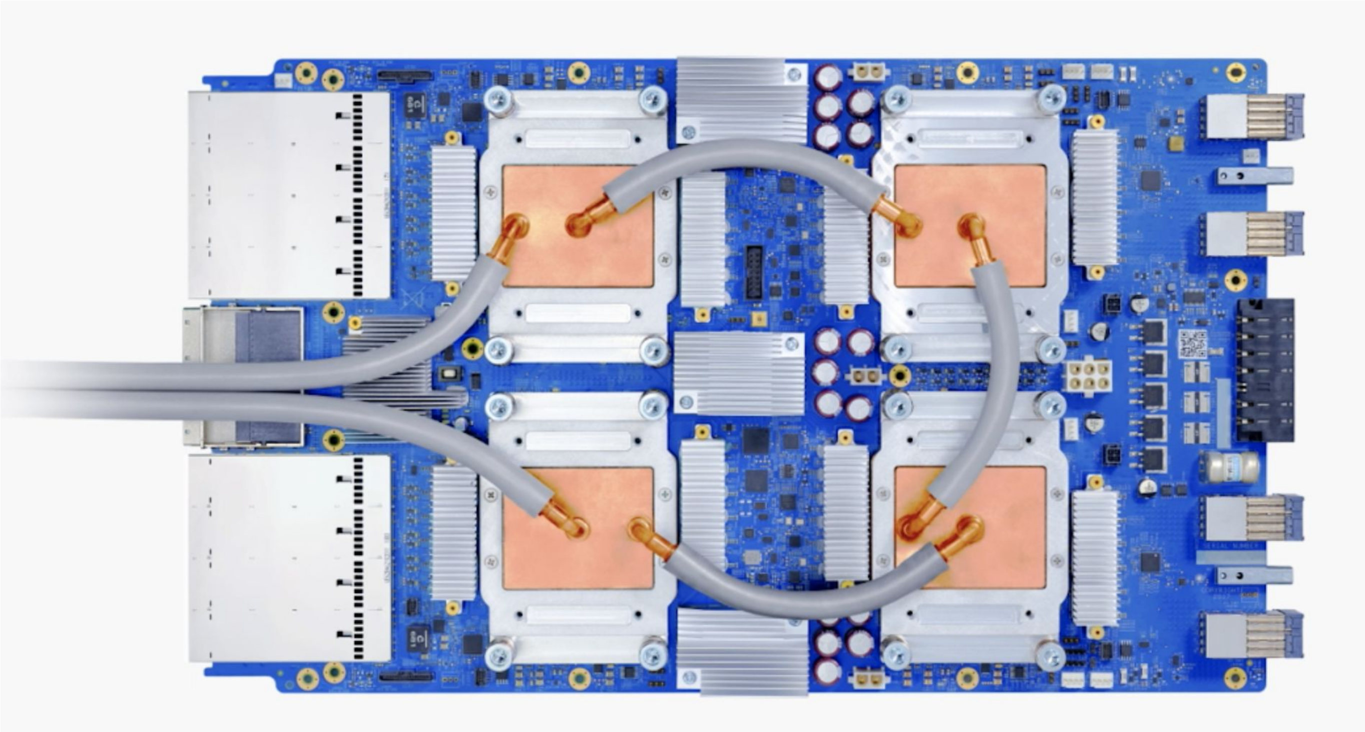
NOT

~~1.21042~~
~~× 0.61127~~
~~0.73989343~~

handful of specific operations



Tensor Processing Unit v3

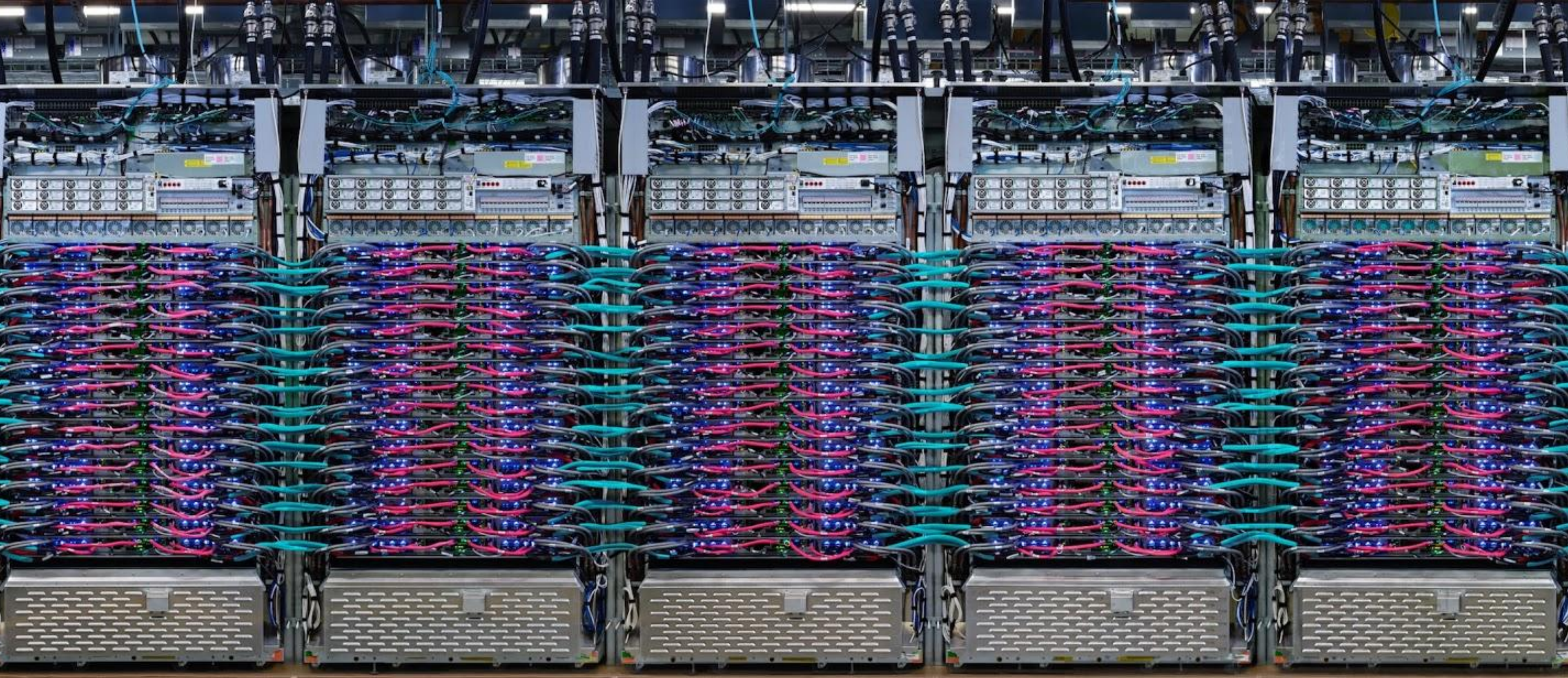


(liquid cooling!)

420 TeraFLOPS, 128 GB HBM

Now available as Cloud TPUv3 Beta

g.co/tpu



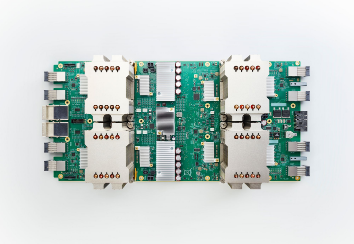
TPUv3 Pod: 256 TPUv3 devices (1024 chips)

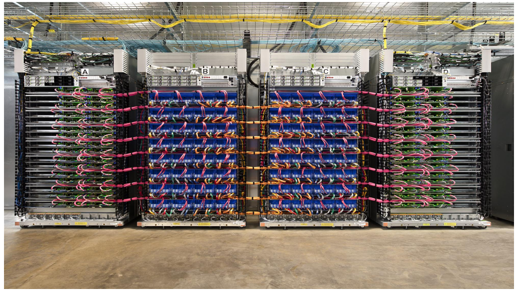
>100 Petaflops

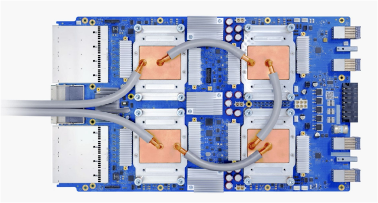
Programmed via TensorFlow

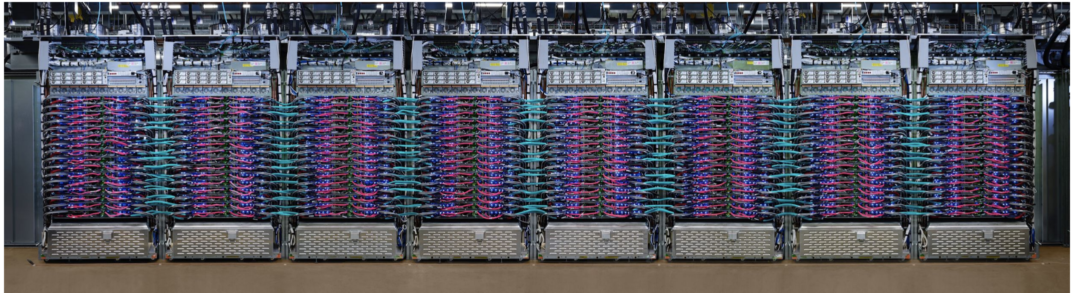
Same program will run w/only minor modifications on CPUs, GPUs, & TPUs

Same program scales via synchronous data parallelism without modification on TPU pods



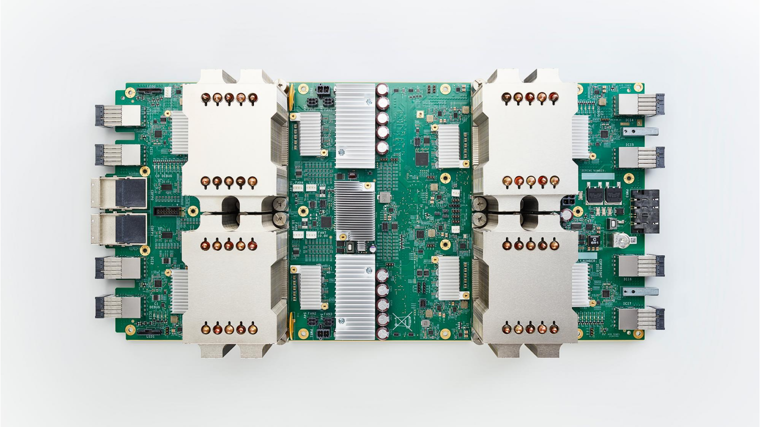






Free Cloud TPUs to support your ML research

TensorFlow
RESEARCH CLOUD

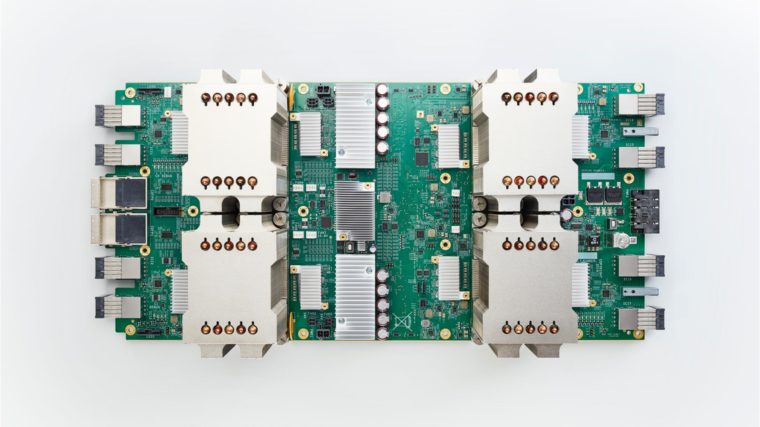


1,000 Cloud TPUs available **for free** to top researchers who are committed to **open machine learning research**

We're excited to see what researchers will do with much more computation!

Free Cloud TPUs to support your ML research

TensorFlow
RESEARCH CLOUD



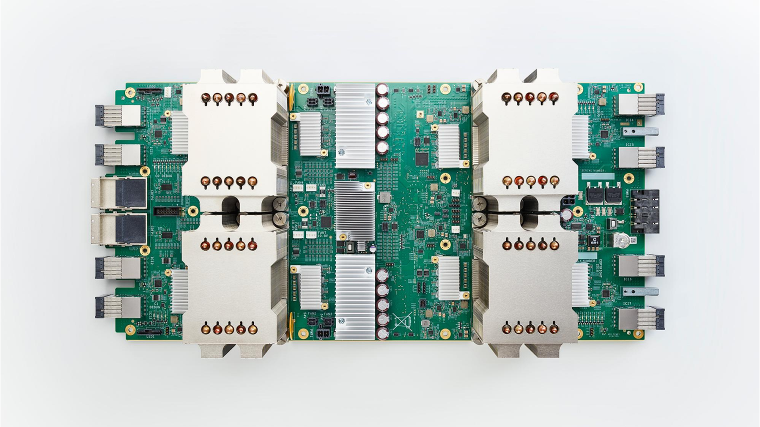
5 regular Cloud TPUs + 20 preemptible Cloud TPUs per person

Free access available for several months

(1 Cloud TPU ~= 4 V100 GPUs)

Free Cloud TPUs to support your ML research

TensorFlow
RESEARCH CLOUD



5 regular Cloud TPUs + 20 preemptible Cloud TPUs per person

Free access available for several months

(1 Cloud TPU ~= 4 V100 GPUs)

If interested, please email tfrc-support@google.com and mention this talk

Or try Cloud TPUs in your browser right now!

via colab

Start with "Hello, TPU:" https://goo.gl/Amd9qQ

(https://colab.research.google.com/notebooks/tpu.ipynb)

Then try **Keras!** https://goo.gl/dnW4wp

Thoughtful use of AI in Society

AI at Google: our principles

Sundar Pichai
CEO

Published Jun 7,

At its heart, AI is computer programming that learns and adapts. It can't solve every problem, but its potential to improve our lives is profound. At Google, we use AI to make products more useful—from email that's spam-free and easier to compose, to a digital assistant you can speak to naturally, to photos that pop the fun stuff out for you to enjoy.

Beyond our products, we're using AI to help people tackle urgent problems. A pair of high school students are building AI-powered sensors to predict the risk of wildfires. Farmers are using it to monitor the health of their herds. Doctors are starting to use AI to help diagnose cancer and prevent blindness. These clear benefits are why Google invests heavily in AI research and development, and makes AI technologies widely available to others via our tools and open-source code.

We recognize that such powerful technology raises equally powerful questions about its use. How AI is developed and used will have a significant impact on society for many years to come. As a leader in AI, we feel a deep responsibility to get this right. So today, we're announcing seven principles to guide our work going forward. These are not theoretical concepts; they are concrete standards that will actively govern our research and product development and will impact our business decisions.

1. Be socially beneficial.
2. Avoid creating or reinforcing unfair bias.
3. Be built and tested for safety.
4. Be accountable to people.
5. Incorporate privacy design principles.
6. Uphold high standards of scientific excellence.
7. Be made available for uses that accord with these principles.

https://ai.google/principles

Machine Learning Fairness

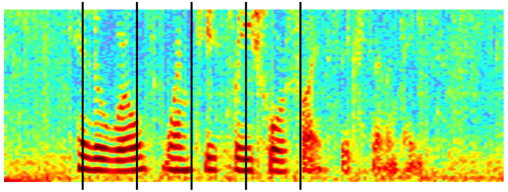
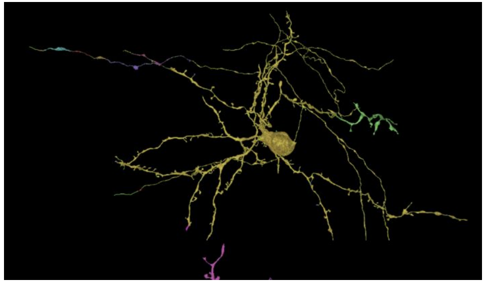
- Text Embedding Models Contain Bias. Here's Why That Matters. (Packer et al., Google 2018)
- Measuring and Mitigating Unintended Bias in Text Classification (Dixon et al., AIES 2018)
  - Exercise demonstrating Pinned AUC metric
- Mitigating Unwanted Biases with Adversarial Learning (Zhang et al., AIES 2018)
  - Exercise demonstrating Mitigating Unwanted Biases with Adversarial Learning
- Data Decisions and Theoretical Implications when Adversarially Learning Fair Representations (Beutel et al., FAT/ML 2017)
- No Classification without Representation: Assessing Geodiversity Issues in Open Data Sets for the Developing World (Shankar et al., NIPS 2017 workshop)
- Equality of Opportunity in Supervised Learning (Hardt et al., NIPS 2016)
- Satisfying Real-world Goals with Dataset Constraints (Goh et al., NIPS 2016)
- Designing Fair Auctions:
  - Fair Resource Allocation in a Volatile Marketplace (Bateni et al. EC 2016)
  - Reservation Exchange Markets for Internet Advertising (Goel et al., LIPics 2016)
- The Reel Truth: Women Aren't Seen or Heard (Geena Davis Inclusion Quotient)

https://developers.google.com/machine-learning/fairness-overview/

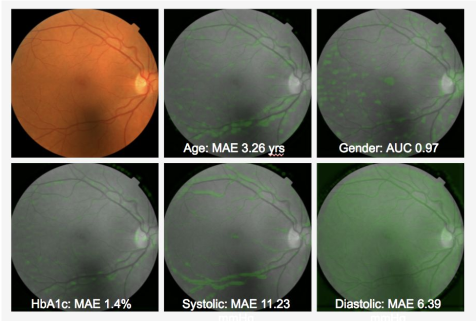
Google AI

Conclusions

Deep neural networks and machine learning are helping to make headway on some of the world's grand challenges





Thank you! More info about our work at ai.google/research
Overview of 2018 work:
ai.googleblog.com/2019/01/looking-back-at-googles-research.html