CMSE 890-002: Mathematics of Deep Learning

Monday, Wednesday, and Friday, 1:50pm - 2:40pm
Engineering Building 1234
Matthew Hirn

Compressed Syllabus

- **Office hours:** MTWR, 4:00pm - 5:00pm
- **Prerequisites:**
  Calc I, II, II; linear algebra; probability; statistics
- **Nice to have, but not required:**
  Graph theory; harmonic analysis; CMSE 820; programming course in deep learning
- **Course webpage:**
  https://matthewhirn.com/teaching/spring-2020-cmse-890-002/

Grading

- Write two reports, each 4 pages + as many additional pages as needed for references
- Report can be on:
  - Topic from class
  - Topic in deep learning, but not covered in class
  - Your own research
- Due dates: February 28 and April 24
- Each report is 50% of your grade
- See the full syllabus for more details

What this course is, and is not

- This course is on the mathematical foundations of deep learning, particularly mathematical theory that gives insight into deep learning algorithm design and architecture

- Therefore this course will not teach you how to code up a neural network

- This course is not on the mathematics of non convex optimization associated with deep learning

Course Outline

1. Background on machine learning and statistical learning theory

2. Artificial neural networks and approximation theory

3. Convolutional neural networks and signal processing

4. Geometric deep learning on graphs and graph signal processing and graph isomorphism

5. Generative models and GANs and high dimensional probability

6. Other topics, possibly recurrent neural networks

Why have this course?

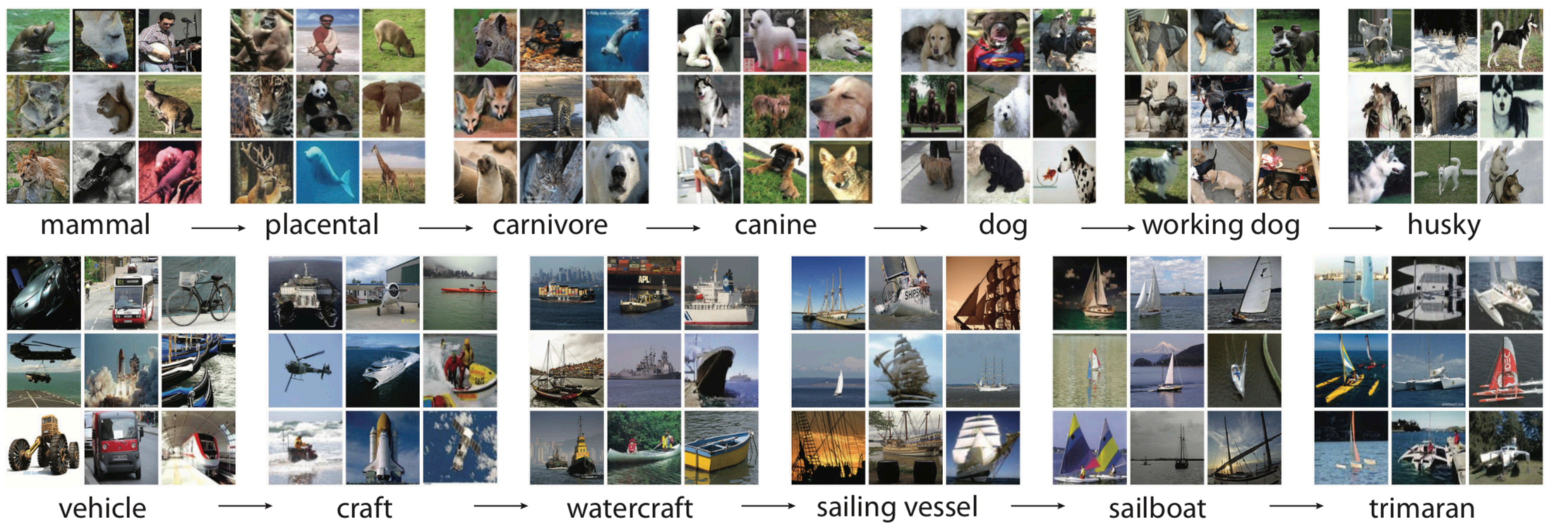
MSU already has several deep learning courses:

- CSE 891-001: Deep Learning
- CSE 891-002: Deep Learning in Biometrics
- ECE/CSE 885: Artificial Neural Networks
- ECE 802-602: Neural Networks and Deep Learning

This course will be a theoretically focused, complementary course to the ones listed above.

But why do we need theory?

Deep learning is being used in a lot of places…

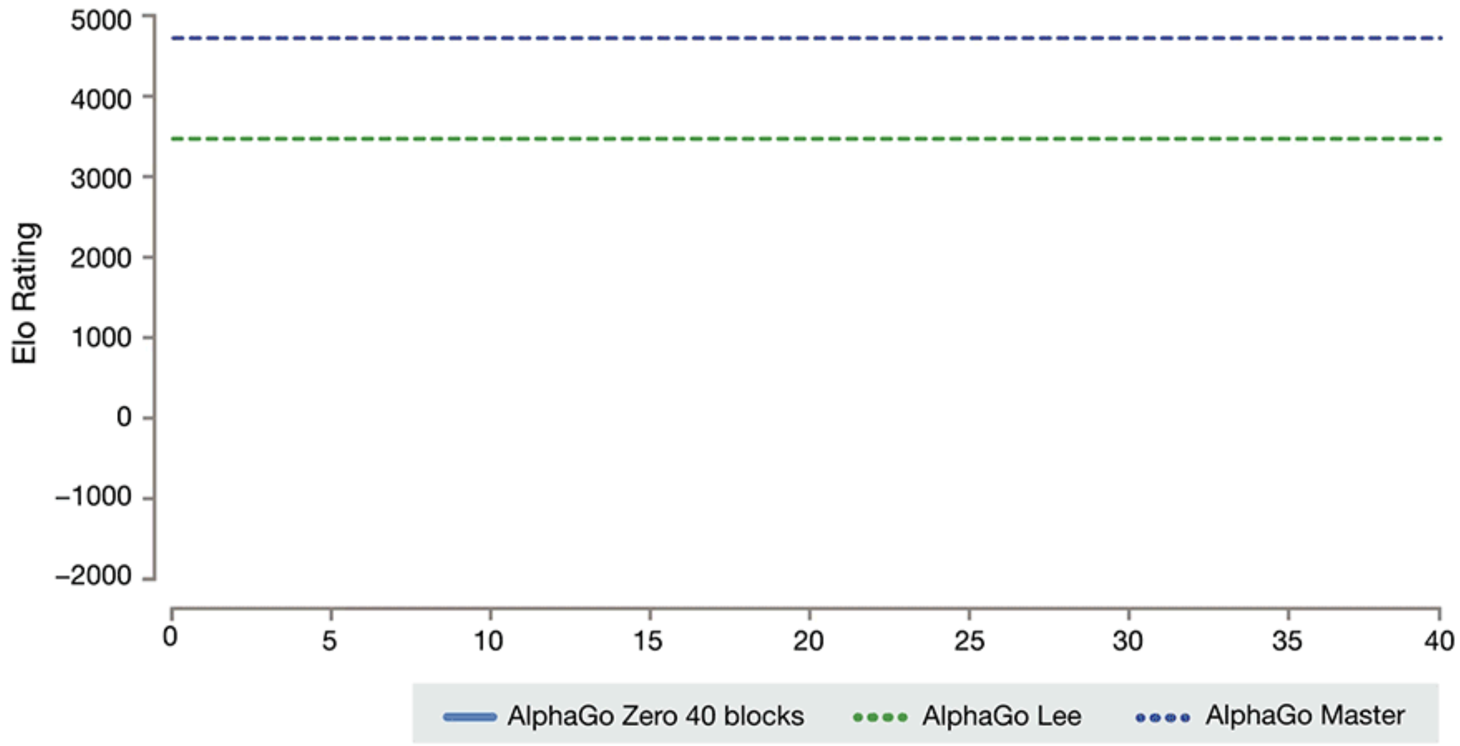


Computer vision

But why do we need theory?

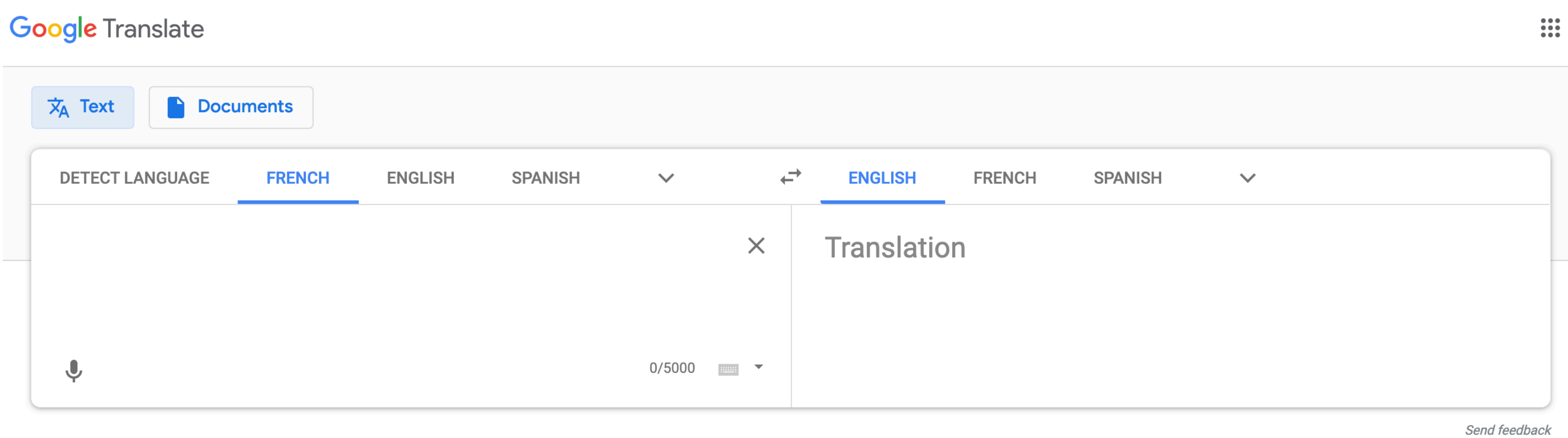
Deep learning is being used in a lot of places…

Playing games



But why do we need theory?

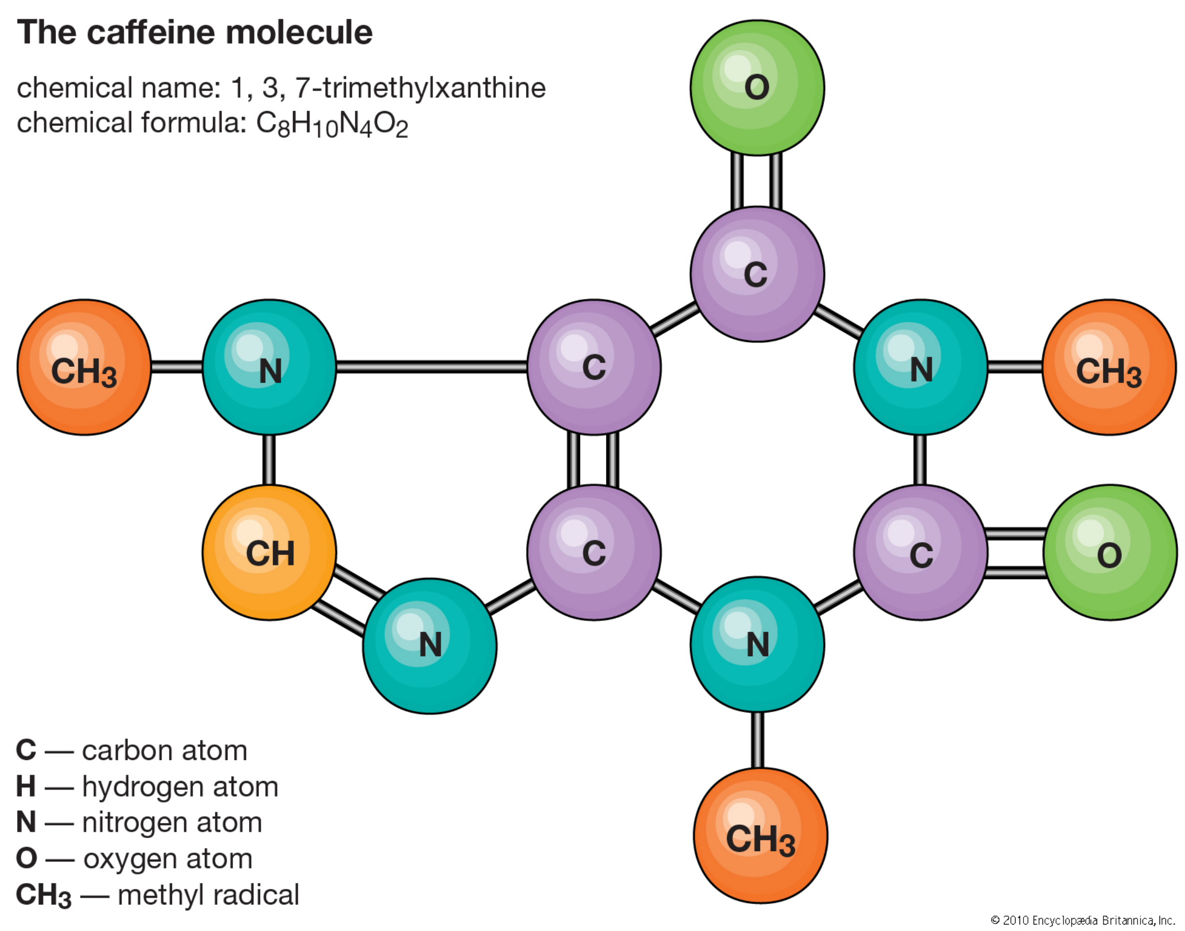
Deep learning is being used in a lot of places…



<u>Natural language processing</u>

But why do we need theory?

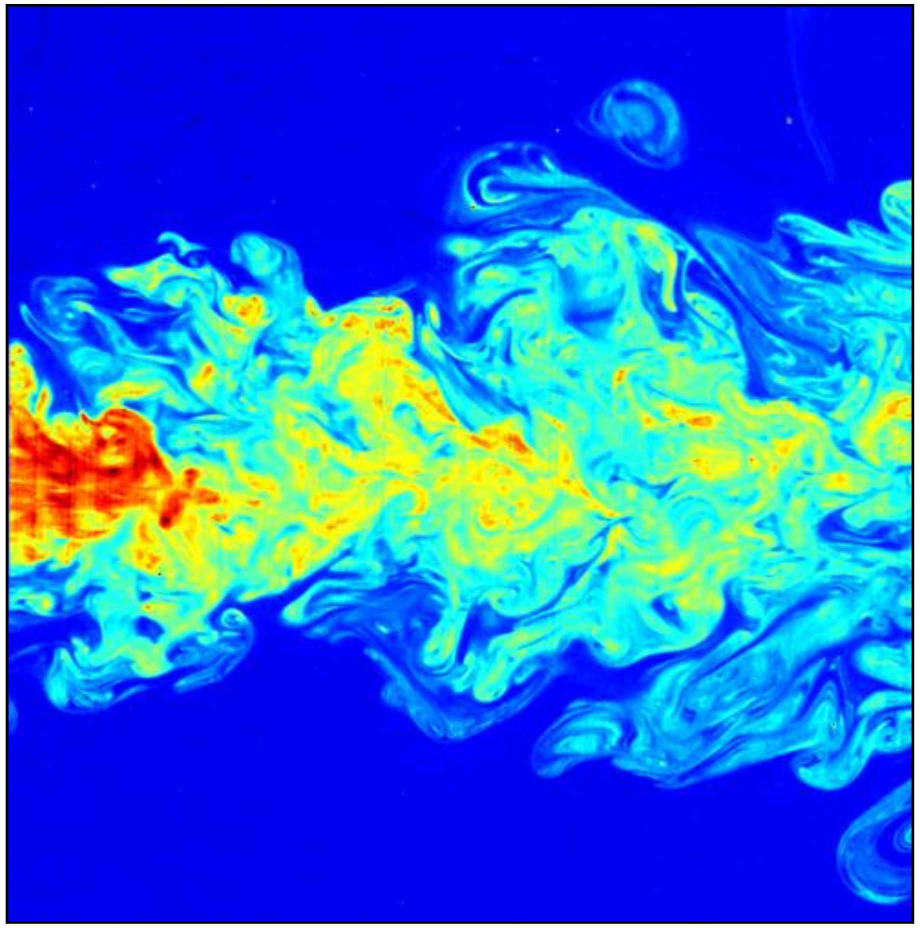
Deep learning is being used in a lot of places…
…and in many “non-traditional” machine learning areas



<u>Chemistry</u>

But why do we need theory?

Deep learning is being used in a lot of places…
…and in many “non-traditional” machine learning areas



Fluid mechanics

But why do we need theory?

Deep learning is being used in a lot of places…
…and in many “non-traditional” machine learning areas

Inverse problems

But why do we need theory?

Deep learning is being used in a lot of places…
…and in many “non-traditional” machine learning areas

AI Provides Doctors with Diagnostic Advice: How Will AI Change Future Medical Care?

FUJITSU JOURNAL

November 29, 2017

Artificial Intelligence, Healthcare

Related information

Article

Image Recognition and Supply-Demand Forecasting - Speedily Transforming AI Technologies into Businesses

Article

Technology for Aligning Multiple CT Images in One Second--Supporting Doctors to Detect Changes in Diseases

Article

AI-Based CT Image Retrieval–Retrieving Similar Cases with an 85% Accuracy Rate in One-Sixth Diagnostic Time

Article

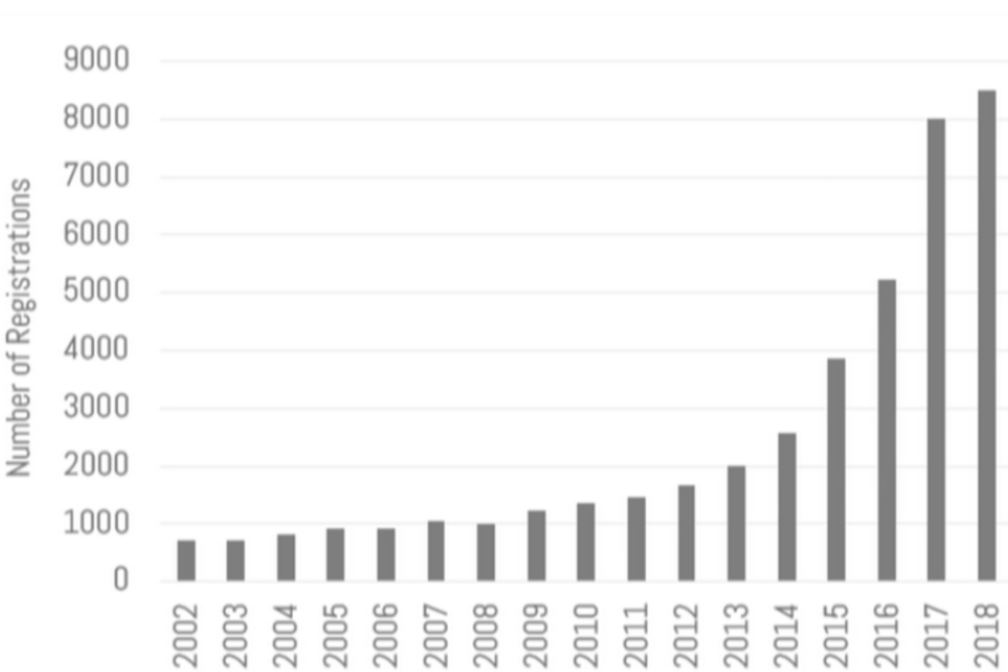
Using AI to Save the Lives of Psychiatric Patients! Fujitsu Halves the Time Required to Diagnose Patients in Joint Field Trial with San Carlos Hospital

But why do we need theory?

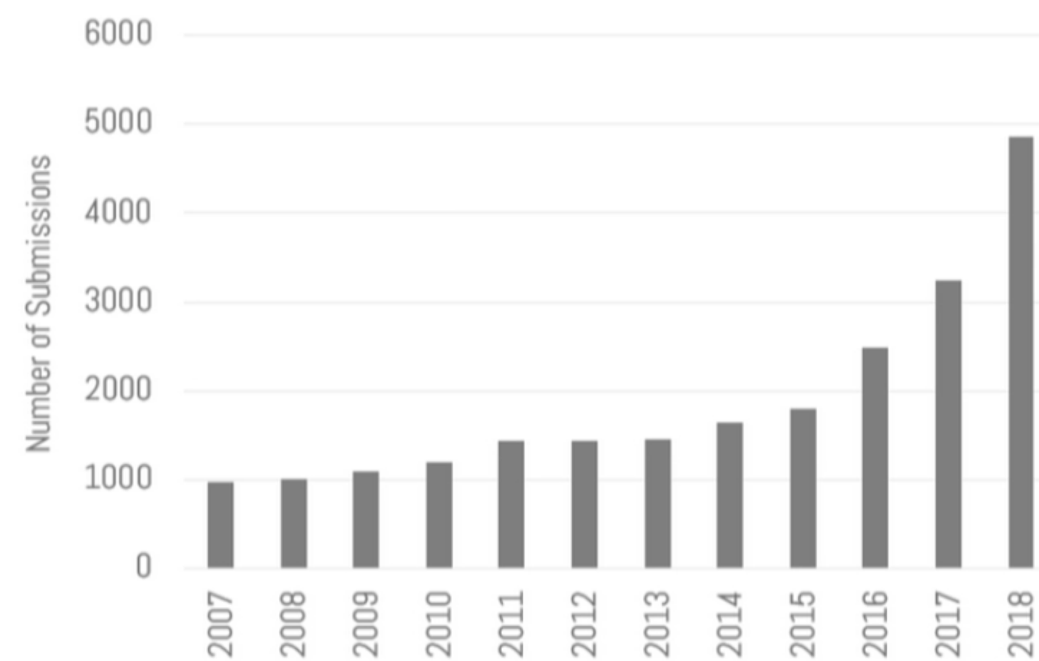
Deep learning is being used in a lot of places…
…and in many “non-traditional” machine learning areas

Self driving cars

Just how popular is deep learning?

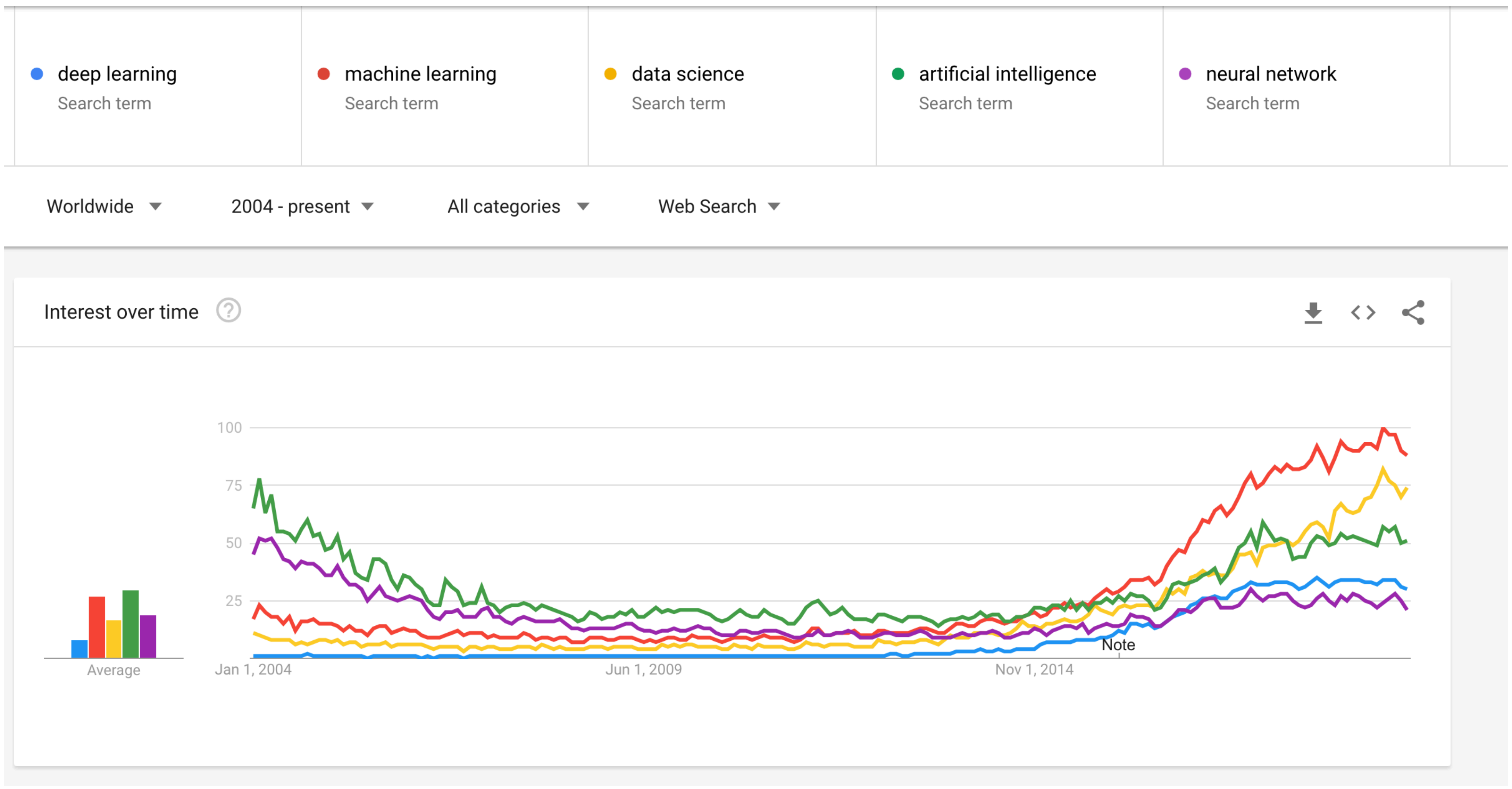


NeurIPS registrations

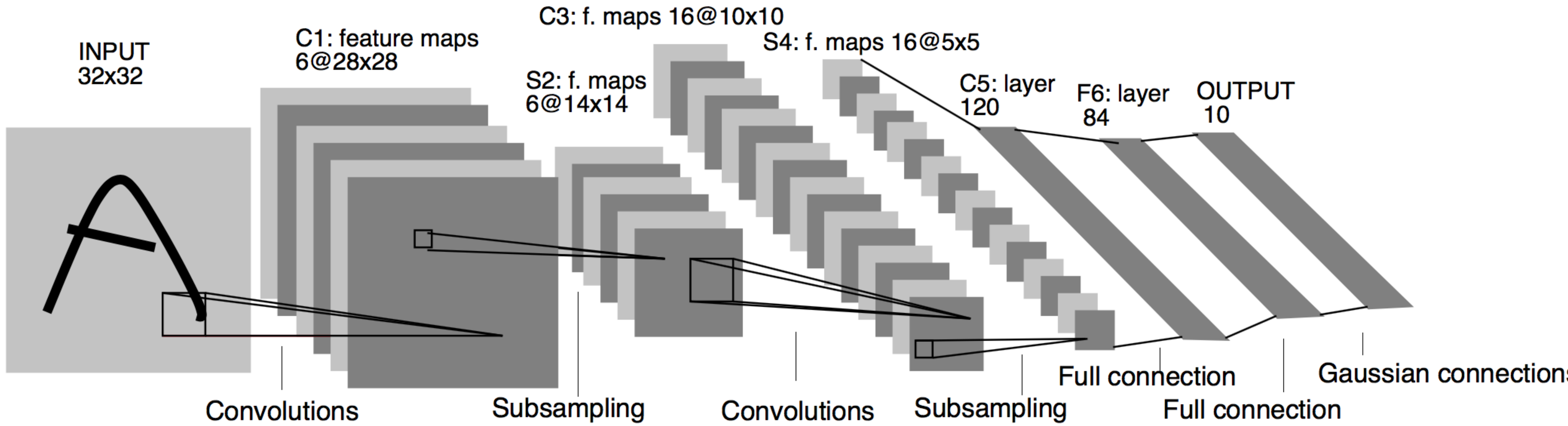


NeurIPS paper submissions

Just how popular is deep learning?



Why is deep learning so popular *now*?



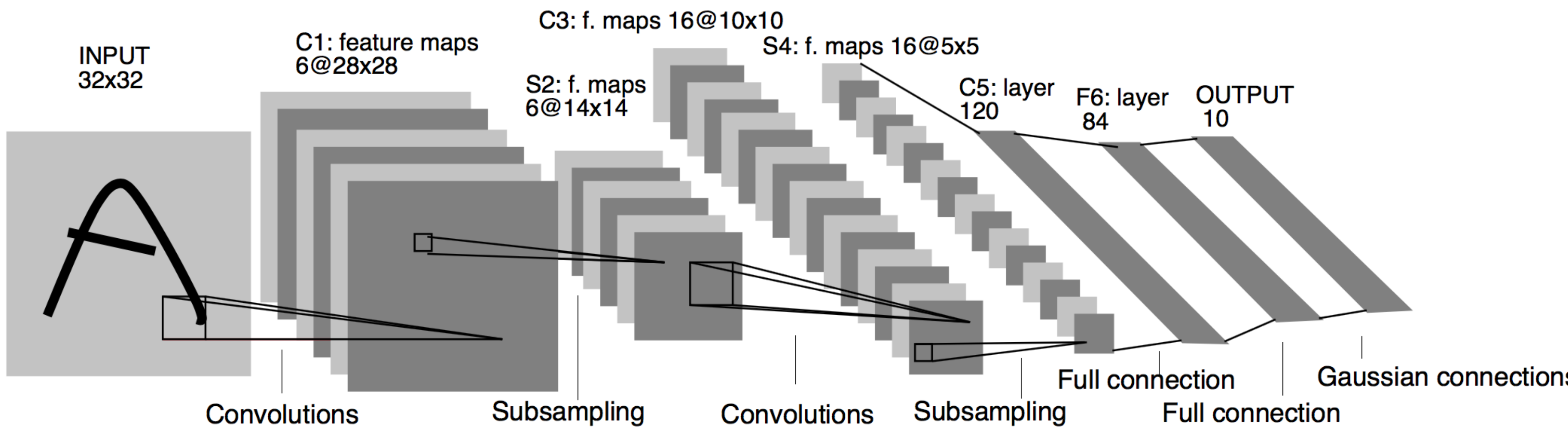
LeNet-5

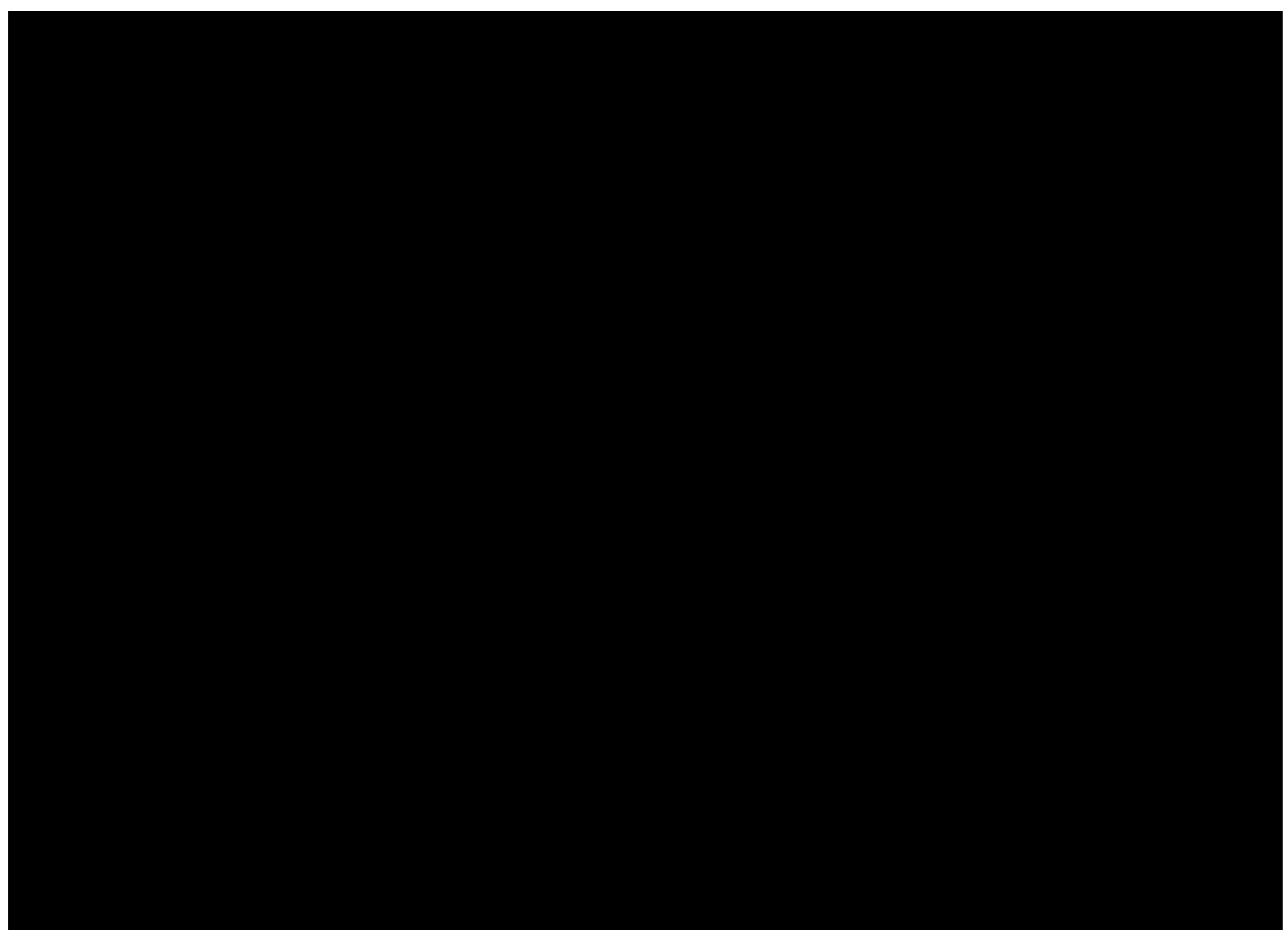
PROC. OF THE IEEE, NOVEMBER 1998 1

Gradient-Based Learning Applied to Document Recognition

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner

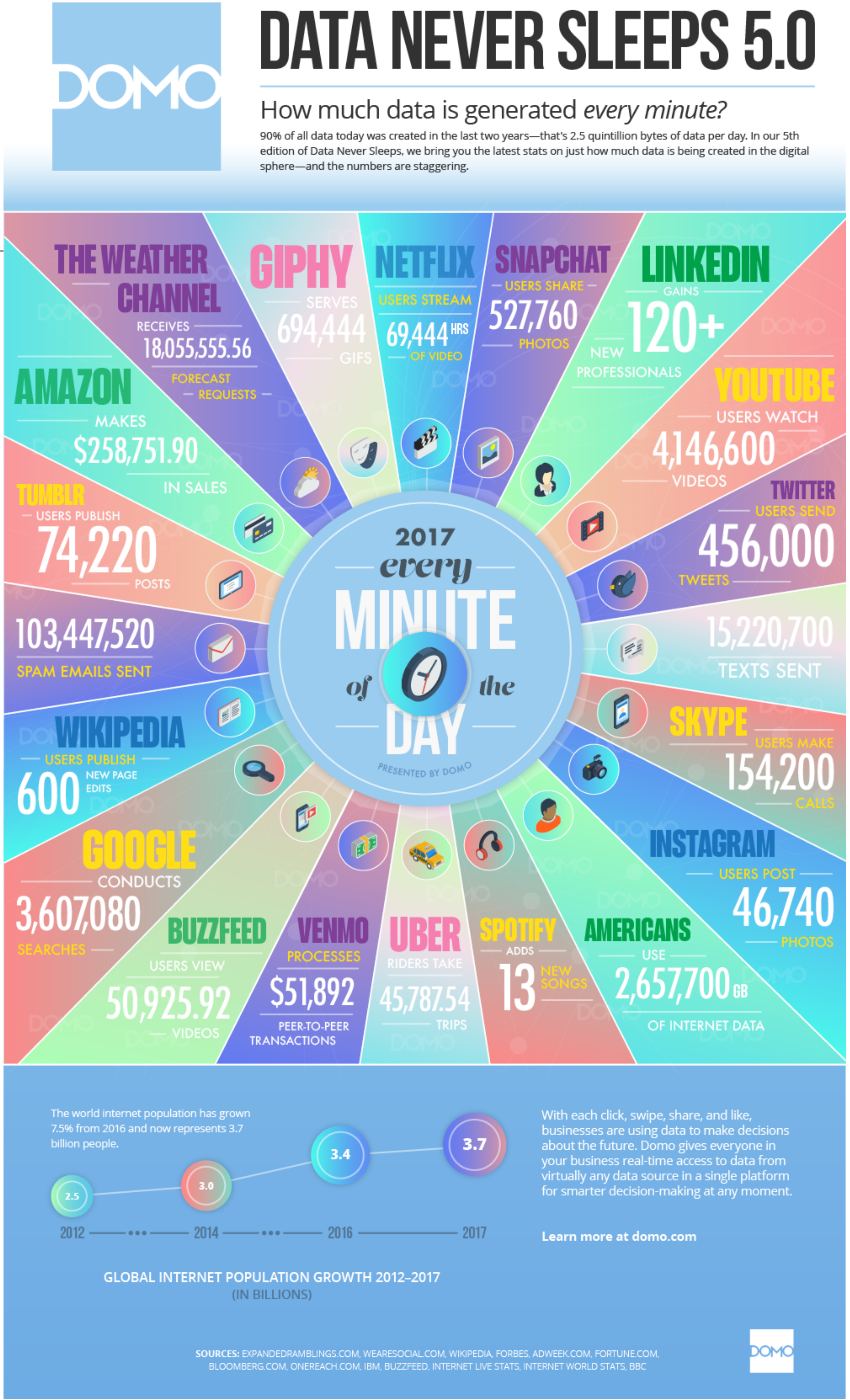
Why is deep learning so popular *now*?



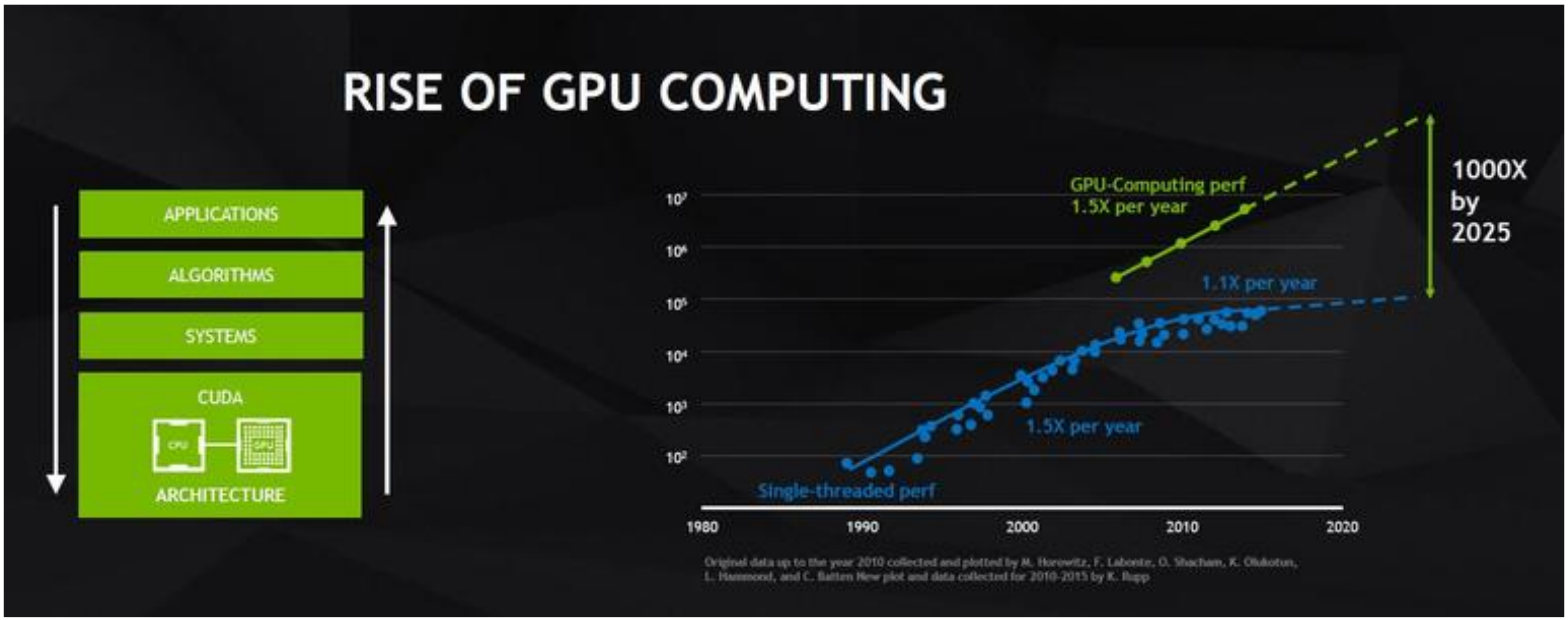


Why is deep learning so popular *now*?

Tons of data



Why is deep learning so popular *now*?



Why is deep learning so popular *now*?

Table 3: Compare AlexNet training with different approaches.

	Batch Size	Processor	GPU Interconnect	Time	Top-1 Accuracy
You et al. [40]	512	DGX-1 station	NVLink	6 hours 10 mins	58.8%
You et al. [40]	32K	CPU x 1024	-	11 mins	58.6%
Jia et al. [18]	64K	Pascal GPU x 1024	100 Gbps	5 mins	58.8%
Jia et al. [18]	64K	Pascal GPU x 2048	100 Gbps	4 mins	58.7%
This Work (DenseCommu)	**64K**	**Volta GPU x 512**	**56 Gbps**	**2.6 mins**	**58.7%**
This Work (SparseCommu)	**64K**	**Volta GPU x 512**	**56 Gbps**	**1.5 mins**	**58.2%**

Table 4: Compare ResNet-50 training with different approaches.

	Batch Size	Processor	GPU Interconnect	Time	Top-1 Accuracy
Goyal et al. [3]	8K	Pascal GPU x 256	56 Gbps	1 hour	76.3%
Smith et al. [41]	16K	Full TPU Pod	-	30 mins	76.1%
Codreanu et al. [42]	32K	KNL x 1024	-	42 mins	75.3%
You et al. [40]	32K	KNL x 2048	-	20 mins	75.4%
Akiba et al. [38]	32K	Pascal GPU x 1024	56 Gbps	15 mins	74.9%
Jia et al. [18]	64K	Pascal GPU x 1024	100 Gbps	8.7 mins	76.2%
Jia et al. [18]	64K	Pascal GPU x 2048	100 Gbps	6.6 mins	75.8%
Mikami et al. [43]	68K	Volta GPU x 2176	200 Gbps	3.7 mins	75.0%
This Work (DenseCommu)	64K	Volta GPU x 512	56 Gbps	7.3 mins	75.3%

SenseTime's innovations shortened training time and illustrated the potential of novel network optimization techniques for improving system performance. Further reductions in ImageNet/AlexNet training times are likely, due to factors such as continuing increases in GPU performance and decreasing network costs.

The paper *Optimizing Network Performance for Distributed DNN Training on GPU Clusters: ImageNet/AlexNet Training in 1.5 Minutes* is on arXiv.

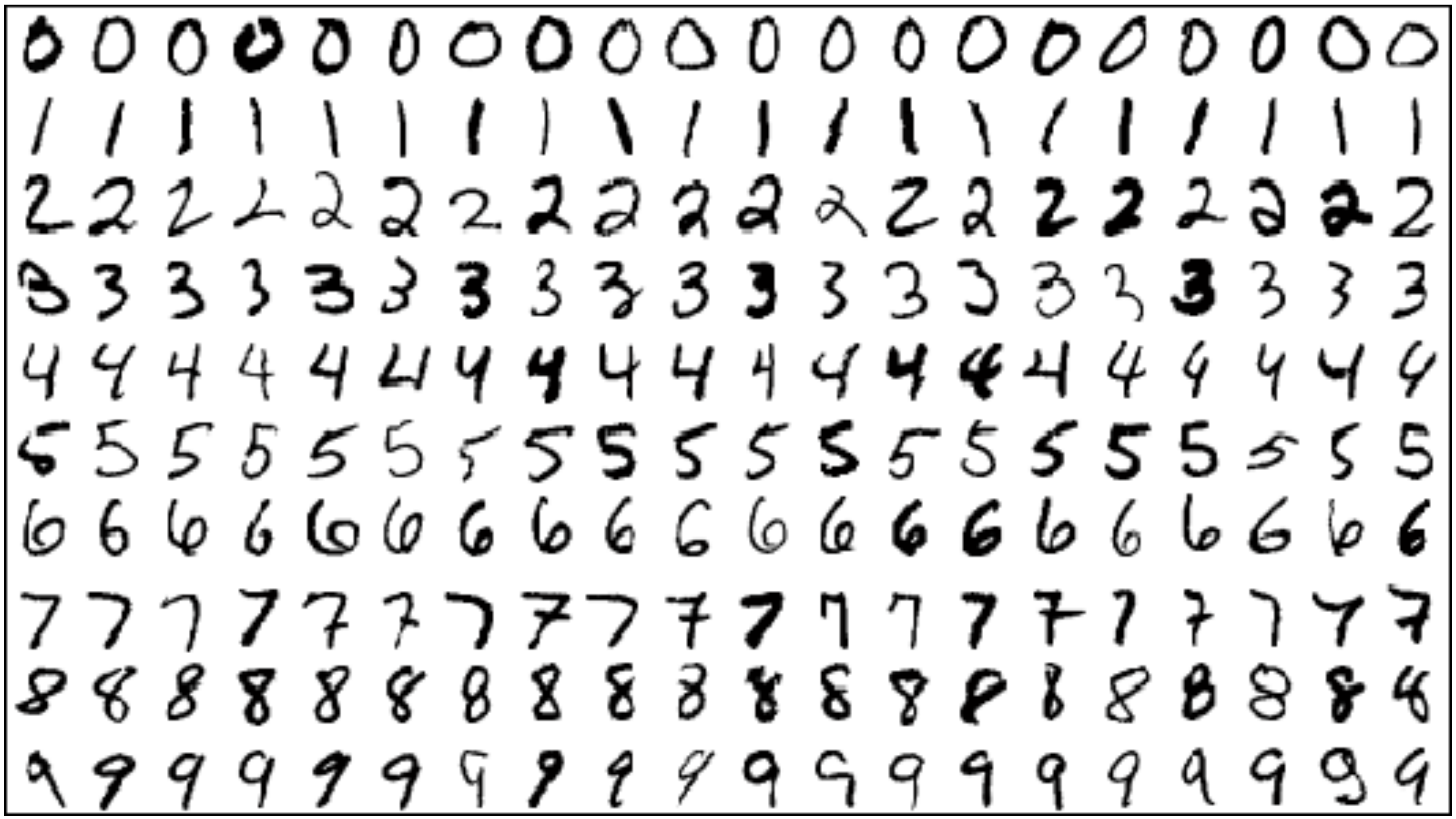
Journalist: Fangyu Cai | **Editor**: Michael Sarazen

Why is deep learning so popular *now*?

Common task framework:

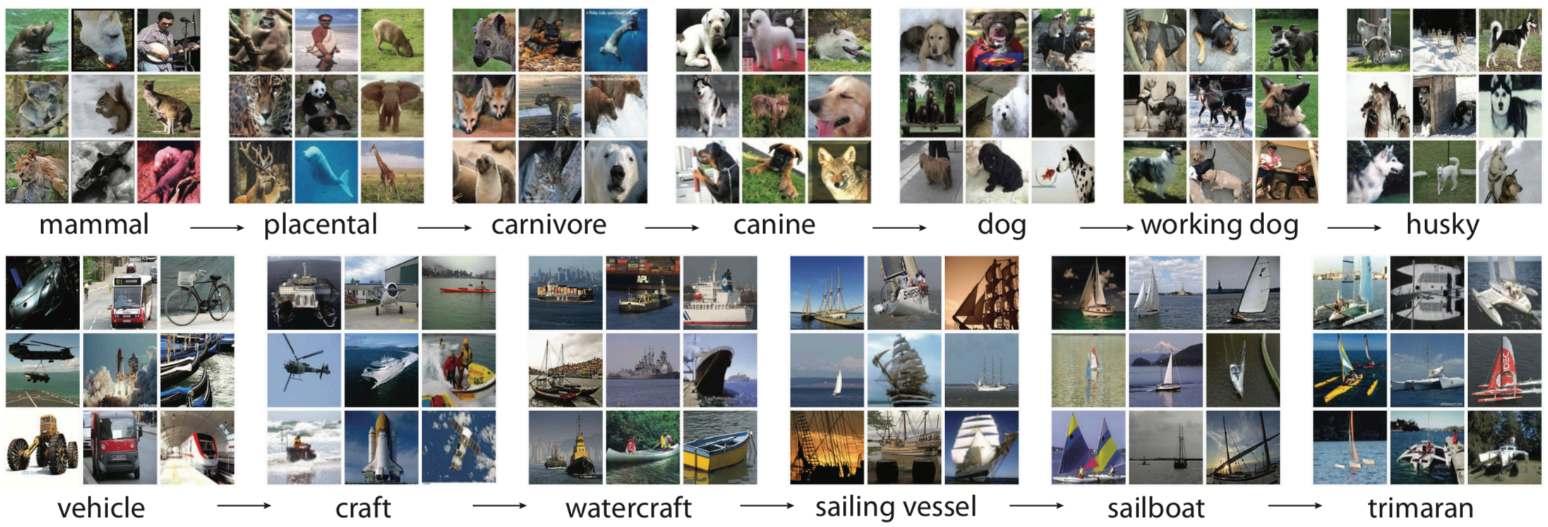
- Publicly available training/test set
- Competitors/teams with the common task of minimizing test error
- An agreed upon metric for measuring said test error

Common Task Framework

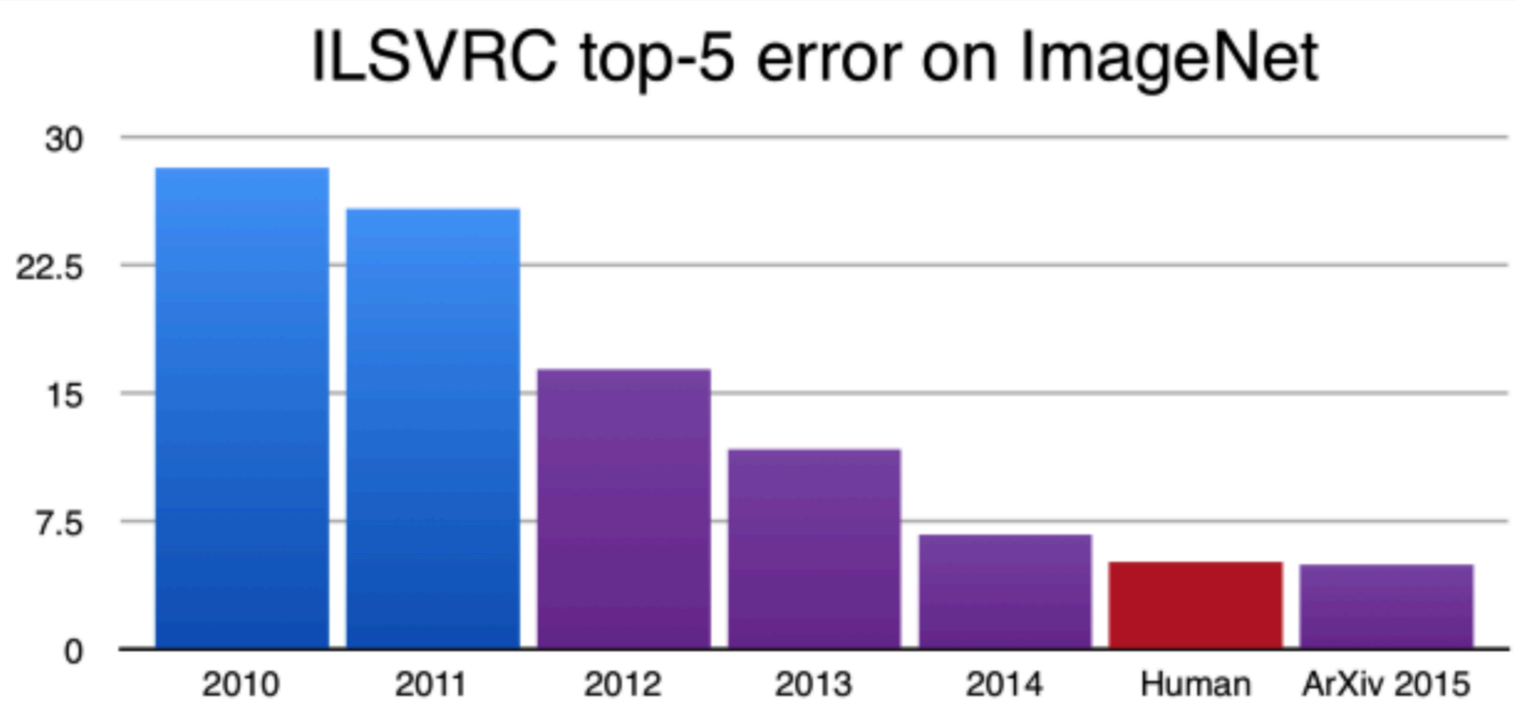


MNIST

Common Task Framework

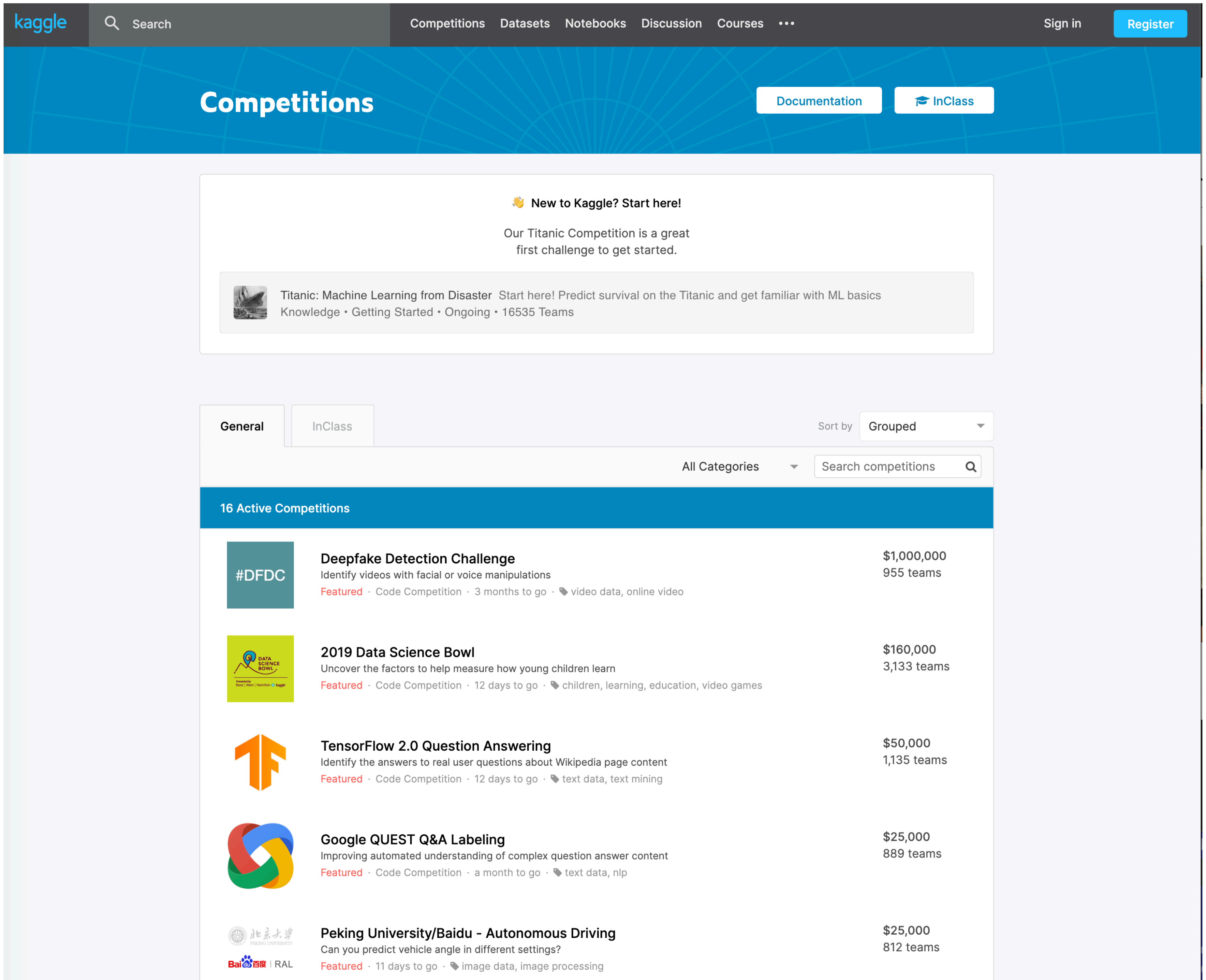


ImageNet



Common Task Framework

Common Task Framework



Common Task Framework Workflow

- Setup the challenge locally: training data, test data, metric for success
- Develop a model, train the model, evaluate on the test set
- Tweak the model, train again, test again
- Repeat as many times as needed until “state of the art”

Deep Learning Criticism?

This approach is not without its critics…

SIAM NEWS MAY 2017

Research | May 01, 2017

Print

Deep, Deep Trouble

Deep Learning's Impact on Image Processing, Mathematics, and Humanity

By Michael Elad

I am really confused. I keep changing my opinion on a daily basis, and I cannot seem to settle on one solid view of this puzzle. No, I am not talking about world politics or the current U.S. president, but rather something far more critical to humankind, and more specifically to our existence and work as engineers and researchers. I am talking about...**deep learning**.

While you might find the above statement rather bombastic and overstated, deep learning indeed raises several critical questions we must address. In the following paragraphs, I hope to expose one key conflict related to the emergence of this field, which is relevant to researchers in the image processing community.

Deep Learning Criticism?

This approach is not without its critics...

AI researchers allege that machine learning is alchemy

"Ali Rahimi, a researcher in artificial intelligence (AI) at Google in San Francisco, California, took a swipe at his field last December–and received a 40-second ovation for it. Speaking at an AI conference, Rahimi charged that machine learning algorithms, in which computers learn through trial and error, have become a form of "alchemy." Researchers, he said, do not know why some algorithms work and others don't, nor do they have rigorous criteria for choosing one AI architecture over another..."

"For example, he says, they adopt pet methods to tune their AIs' "learning rates" how much an algorithm corrects itself after each mistake–without understanding why one is better than others. In other cases, AI researchers training their algorithms are simply stumbling in the dark. For example, they implement what's called "stochastic gradient descent" in order to optimize an algorithm's parameters for the lowest possible failure rate. Yet despite thousands of academic papers on the subject, and countless ways of applying the method, the process still relies on trial and error..."

Science (May 2018)

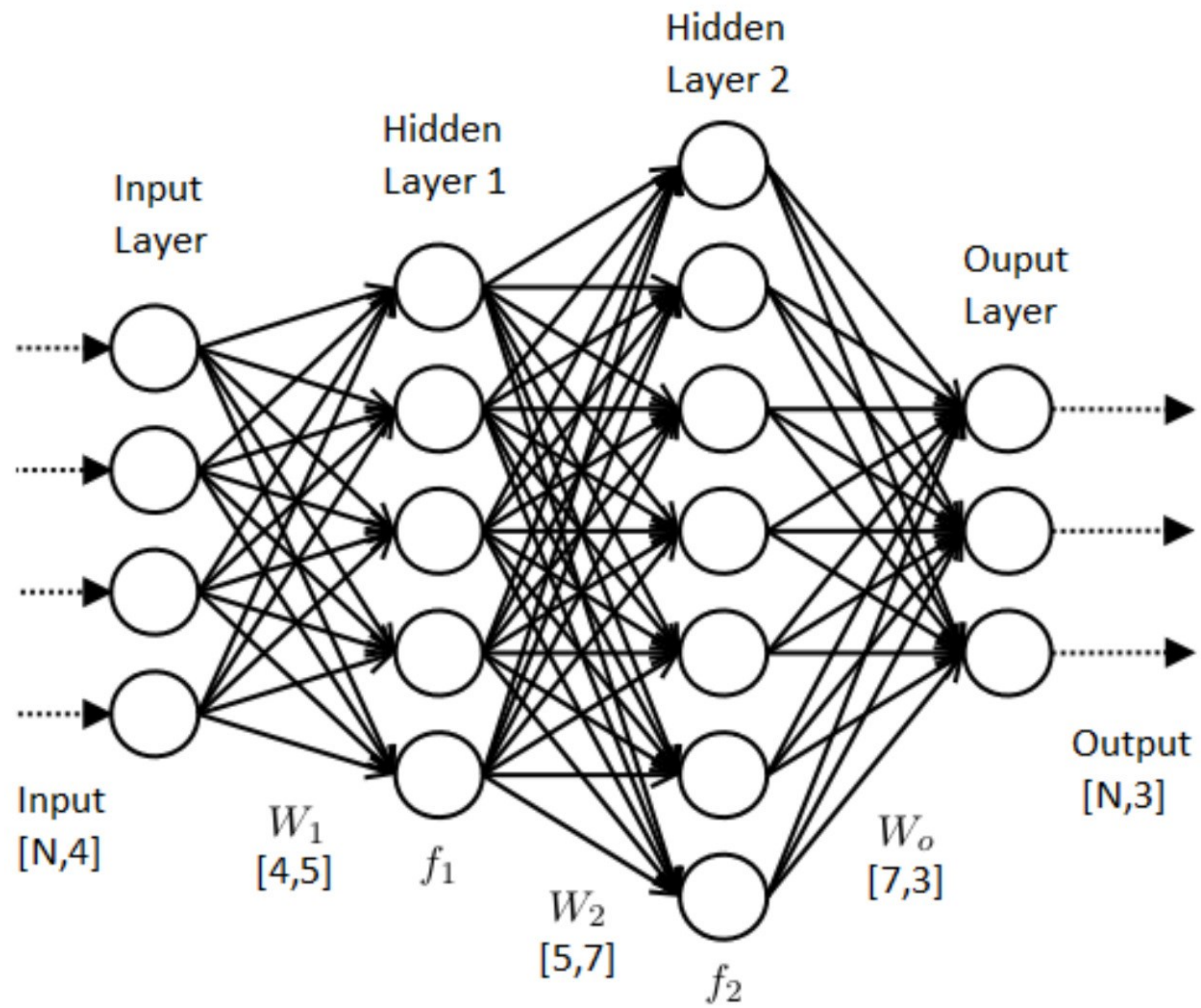
Current State of Affairs

- Deep learning has been incredibly successful on empirical tasks and model development
- But the training/testing process relies significantly on trial and error
- Training can be unpredictable
- Still not too much understanding and interpreting how and why deep networks arrive at their results… but progress is being made, and in some sense that is what this course is about

What are Deep Networks?

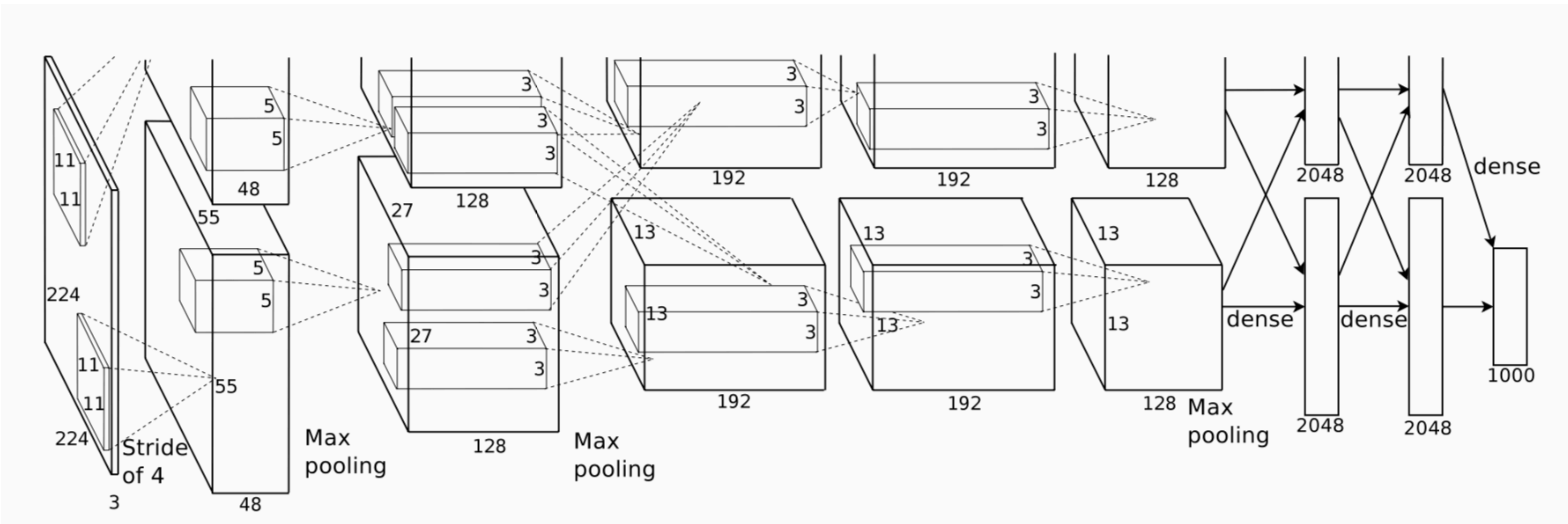
- Compositional models, consisting of alternating linear and nonlinear transforms
- Linear part learned
- Nonlinear parts and architecture specified by user
- Lots of options, but can learn complex functions!

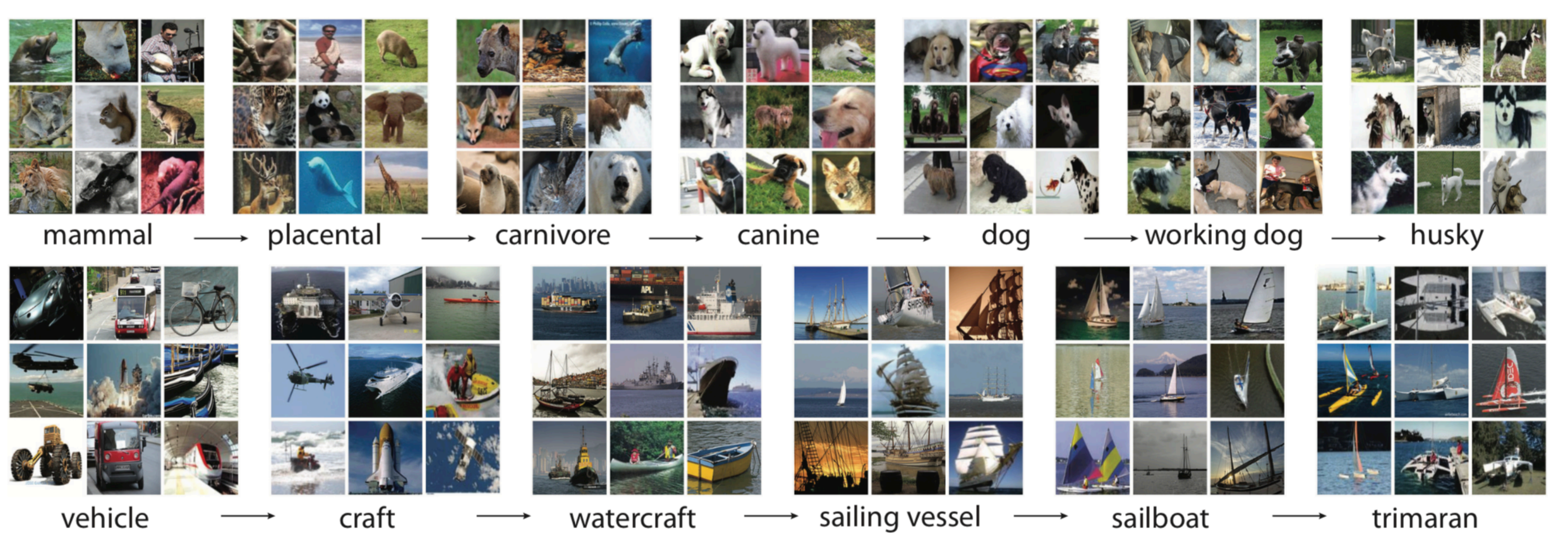
Artificial Neural Networks



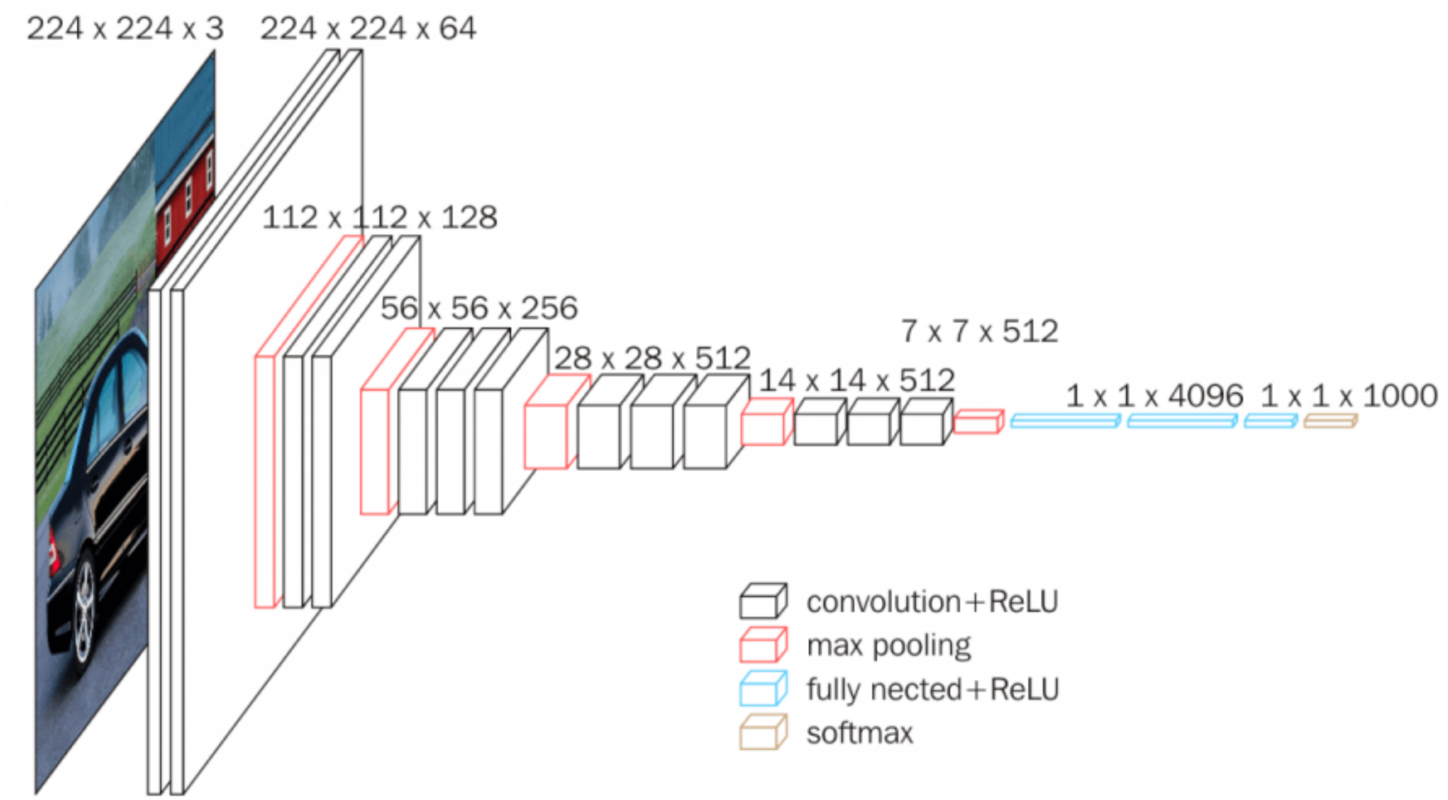
Theory: Universal approximations of complex functions

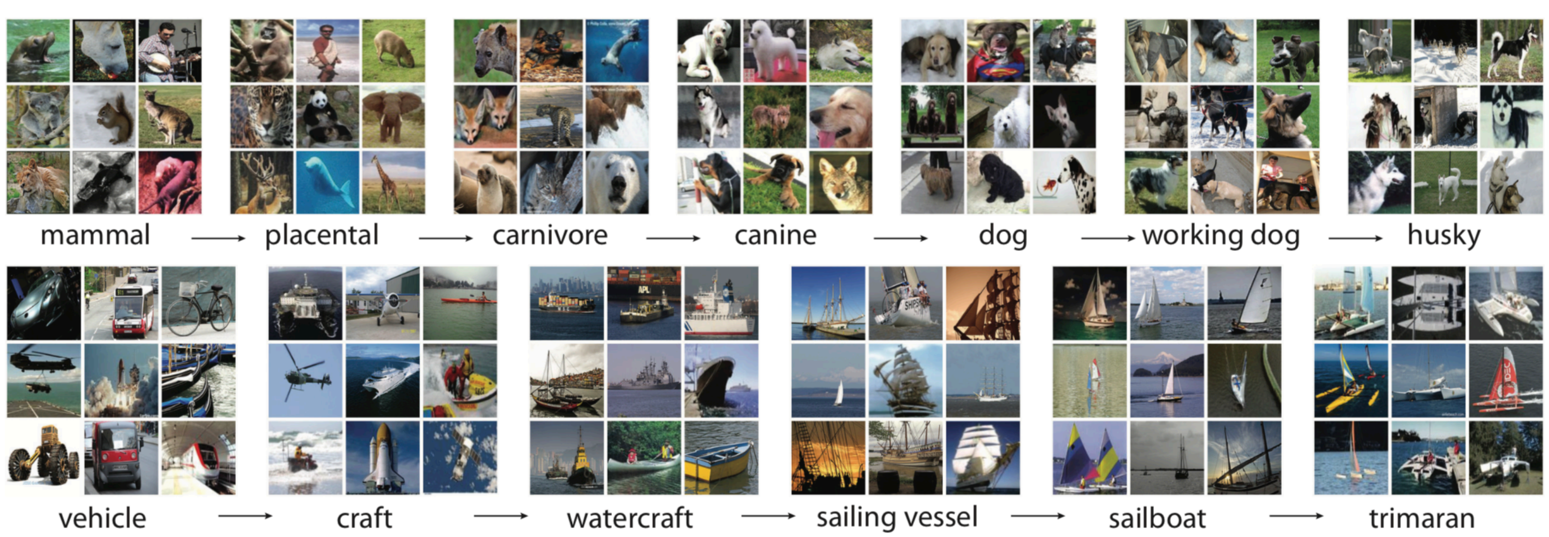
Convolutional Neural Networks



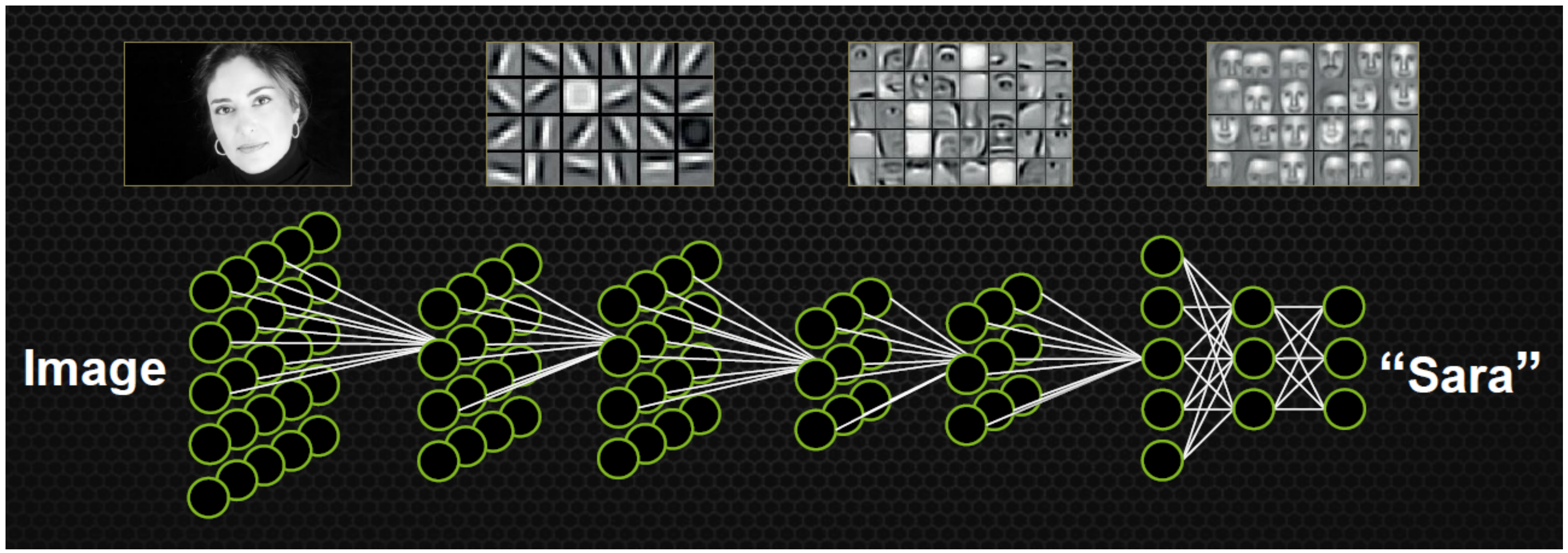


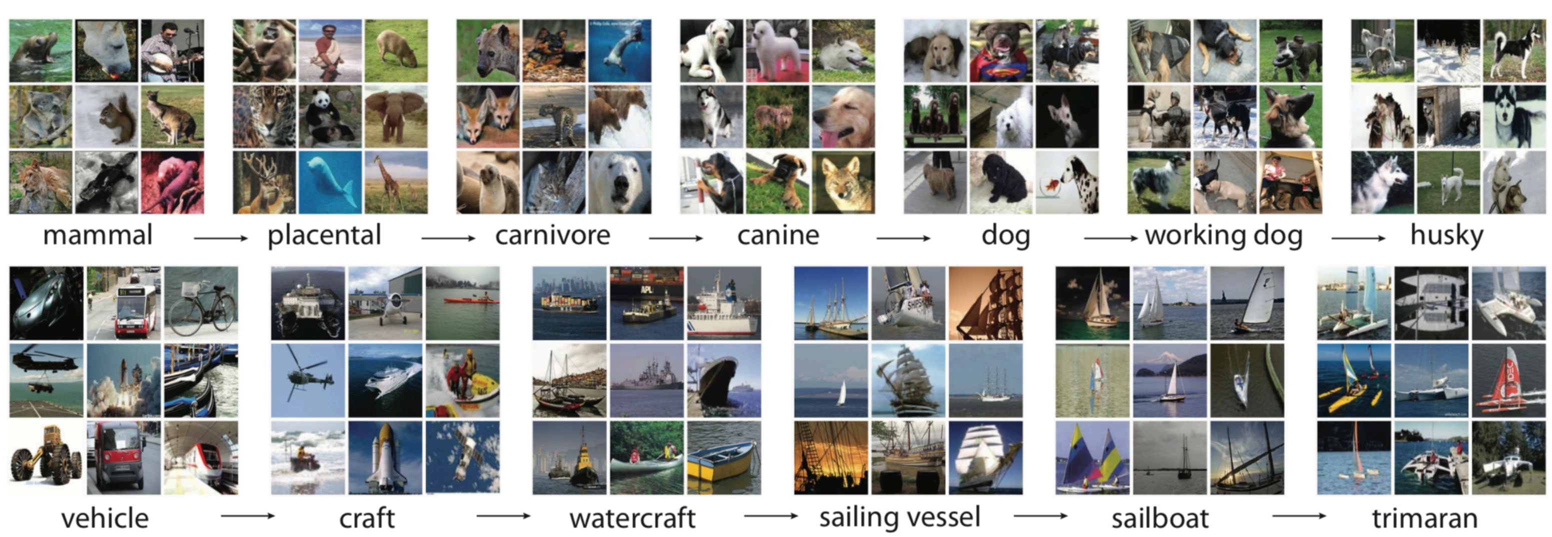
Convolutional Neural Networks



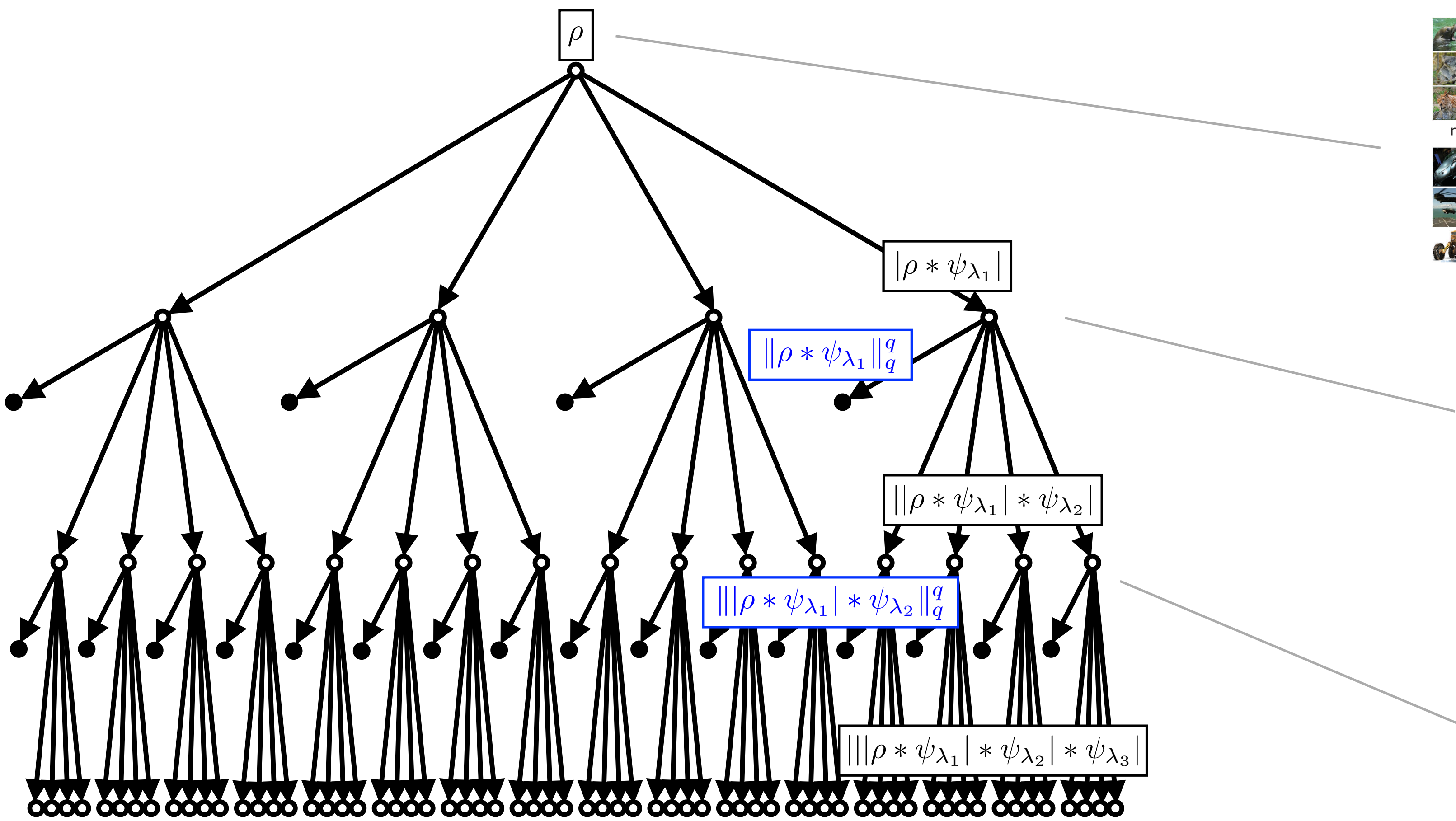


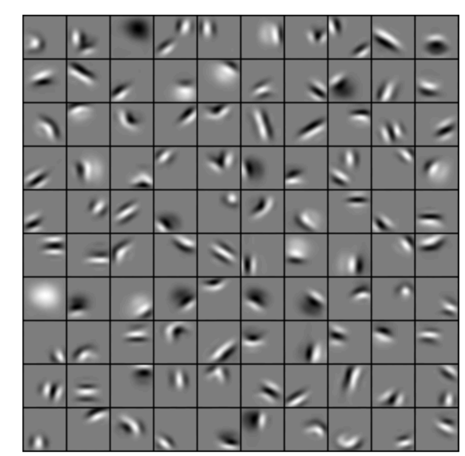
Convolutional Neural Networks



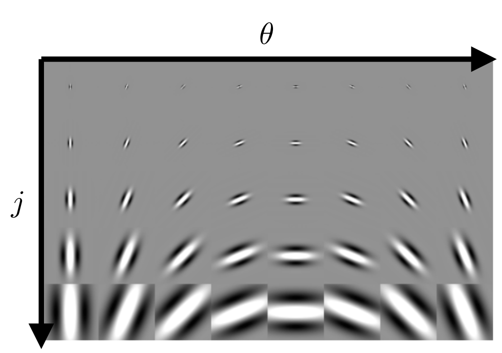


Convolutional Neural Networks

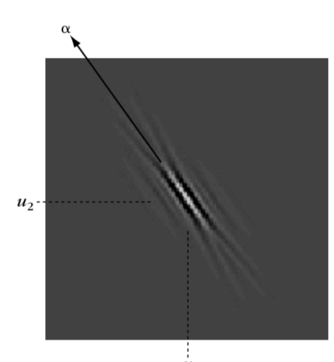




Machine learned filters at each layer (again similar to wavelets!)

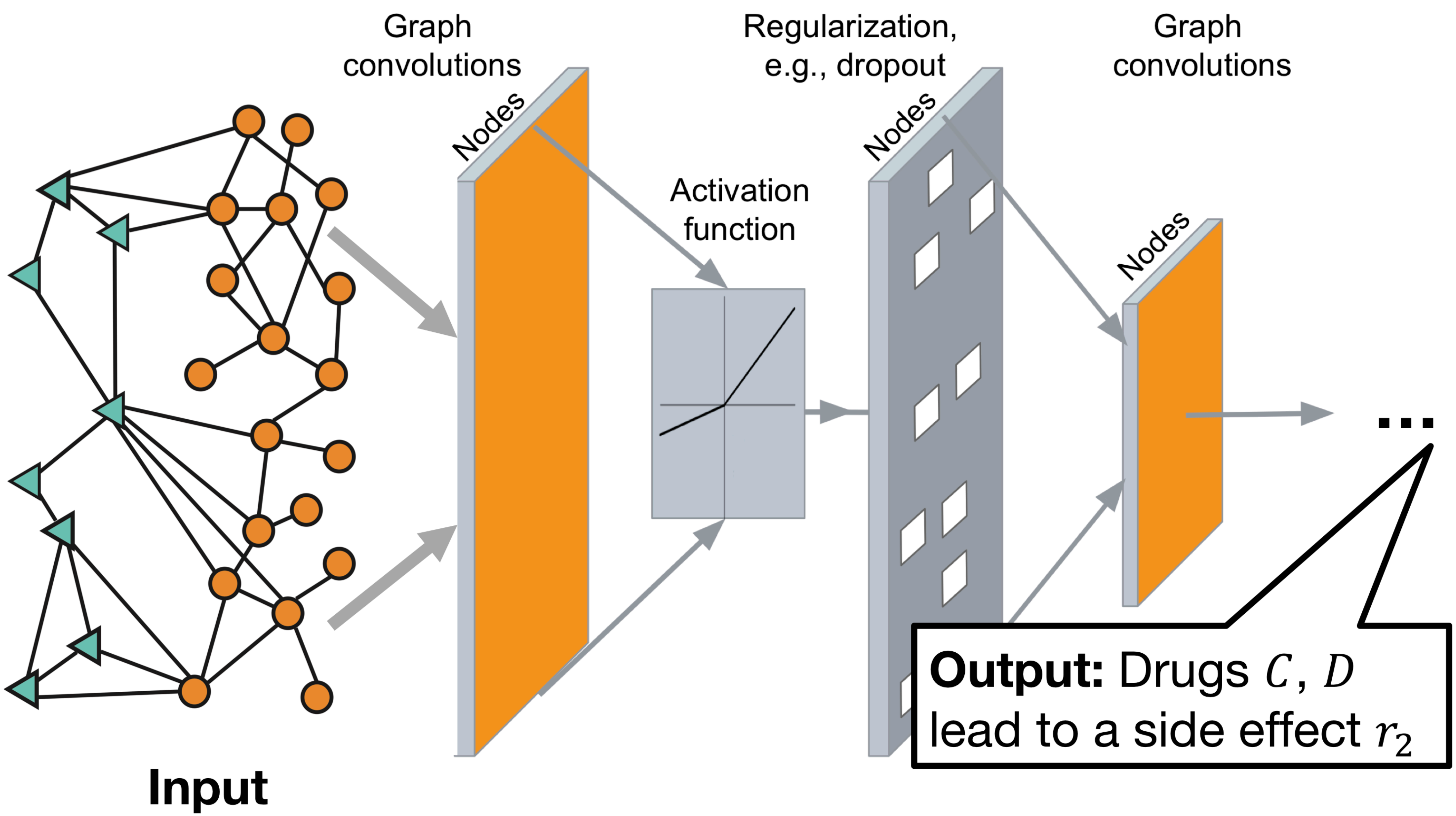


Gabor wavelets $\psi_{j,\theta}$

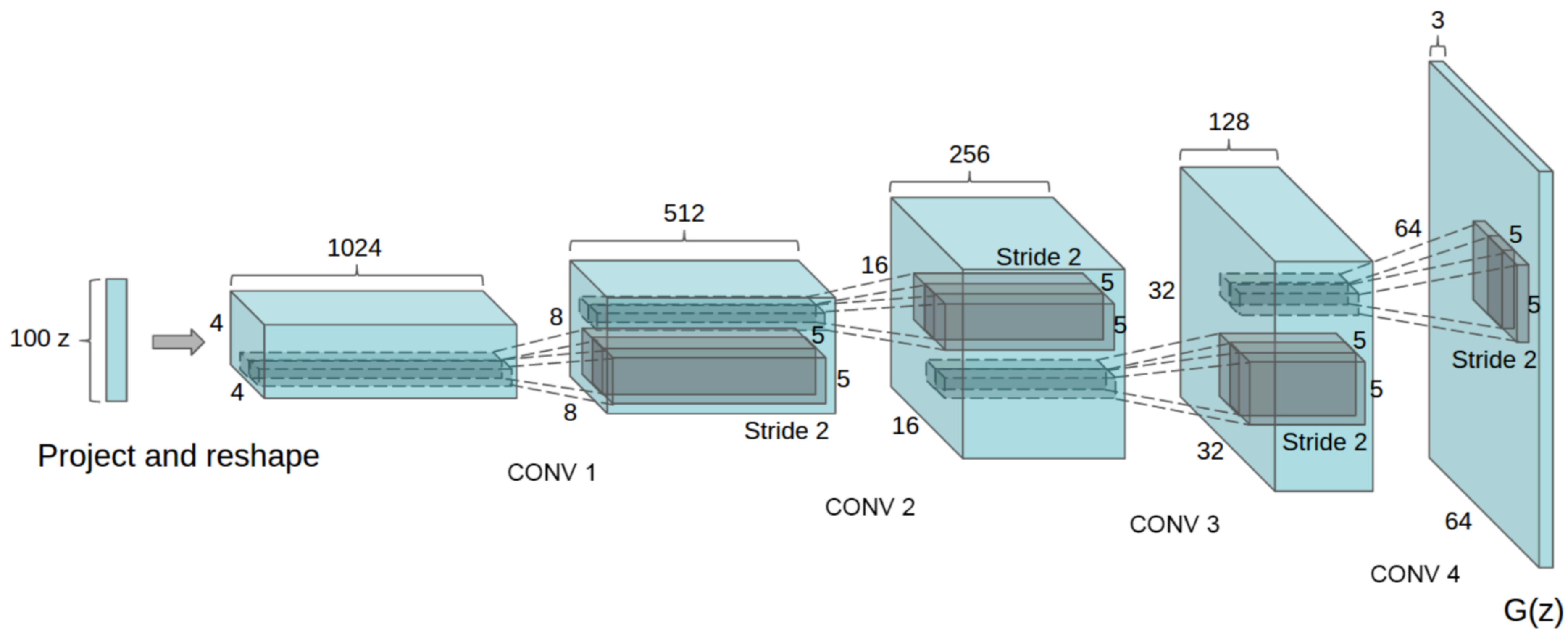


Curvelet $\psi_{j,\alpha}$

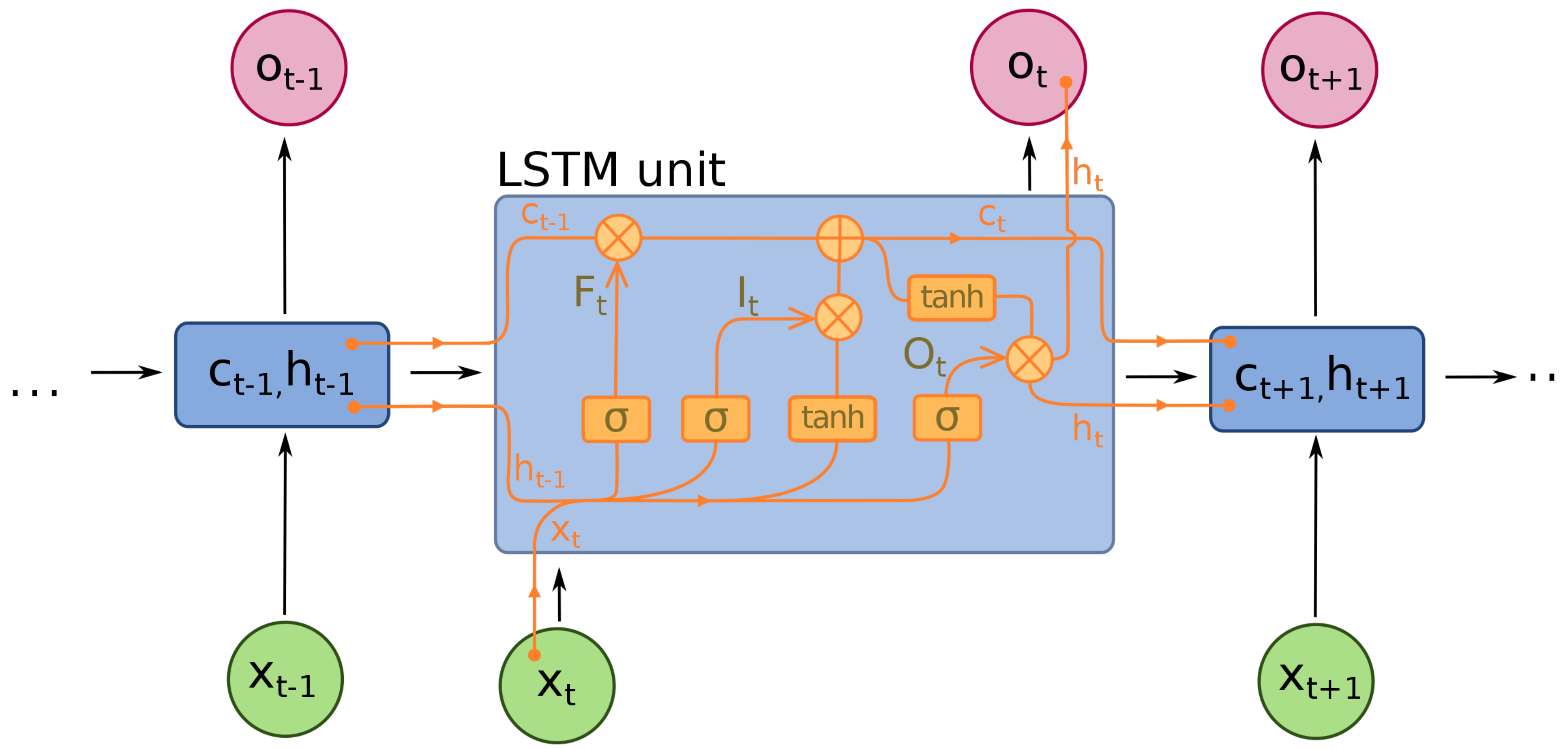
Graph Neural Networks

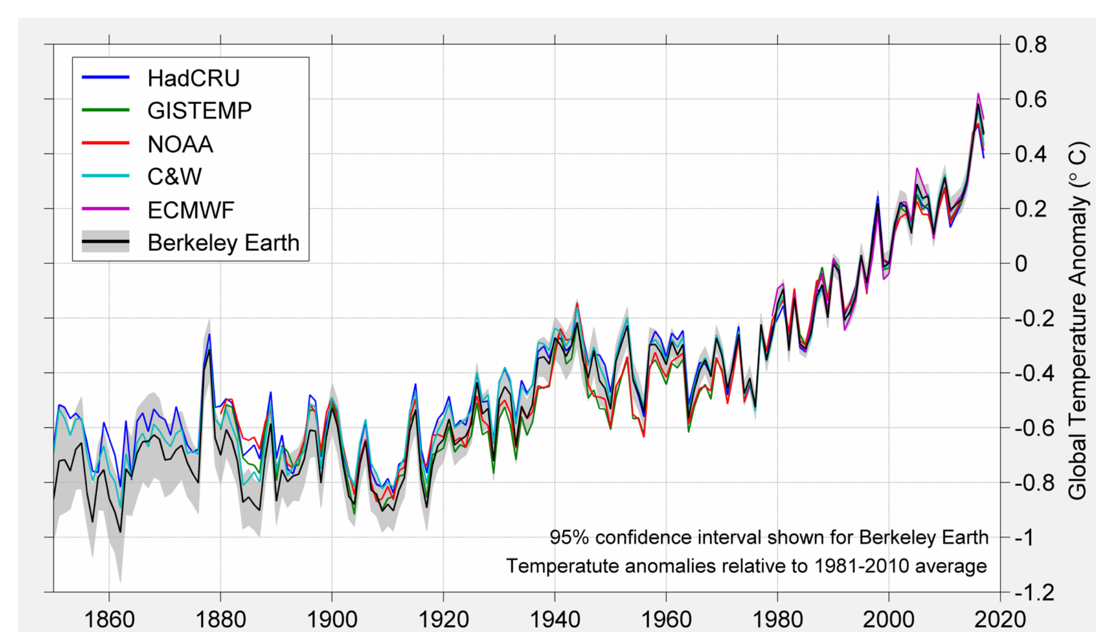


Generative Models



Recurrent Neural Networks





Questions?