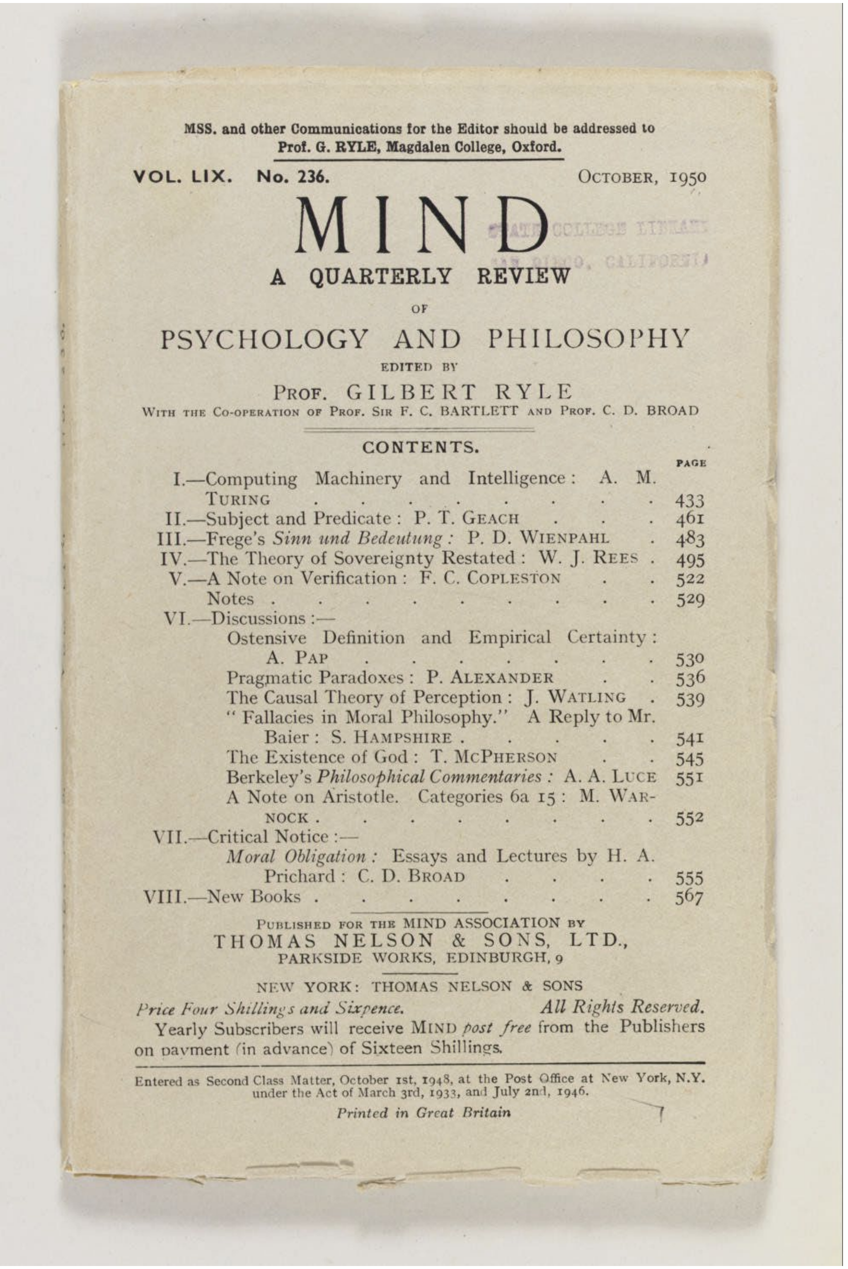
Overview of Machine Learning

CIR Community Day, 13th June 2022

David Hogg
University of Leeds



MSS. and other Communications for the Editor should be addressed to
Prof. G. RYLE, Magdalen College, Oxford.

VOL. LIX. No. 236. OCTOBER, 1950

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

EDITED BY

PROF. GILBERT RYLE

WITH THE CO-OPERATION OF PROF. SIR F. C. BARTLETT AND PROF. C. D. BROAD

CONTENTS.

PUBLISHED FOR THE MIND ASSOCIATION BY
THOMAS NELSON & SONS, LTD.,
PARKSIDE WORKS, EDINBURGH, 9

NEW YORK: THOMAS NELSON & SONS

Price Four Shillings and Sixpence.

Yearly Subscribers will receive MIND *post free* from the Publishers on payment (in advance) of Sixteen Shillings.

Entered as Second Class Matter, October 1st, 1948, at the Post Office at New York, N.Y. under the Act of March 3rd, 1933, and July 2nd, 1946.

Printed in Great Britain

A. M. Turing
Computing Machinery and Intelligence

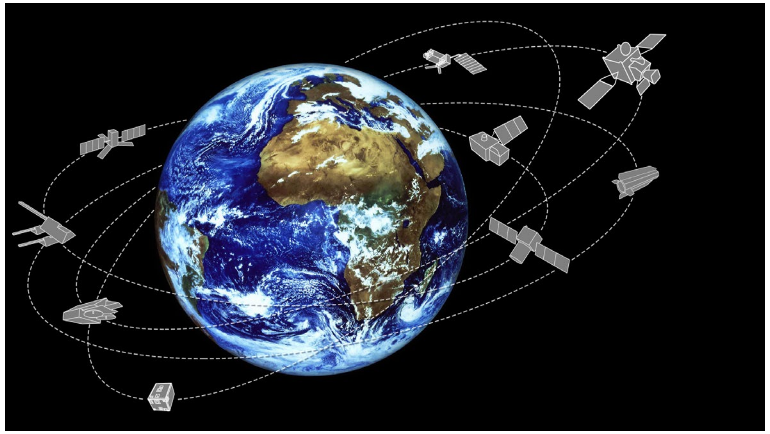
Growth in on-line data

400 million active websites

500 hours of YouTube video uploaded every minute

500,000 movies

100 terabytes of satellite images every day



DOI:10.1145/3448250

How can neural networks learn the rich internal representations required for difficult tasks such as recognizing objects or understanding language?

BY YOSHUA BENGIO, YANN LECUN, AND GEOFFREY HINTON

Deep Learning for AI

TURING LECTURE

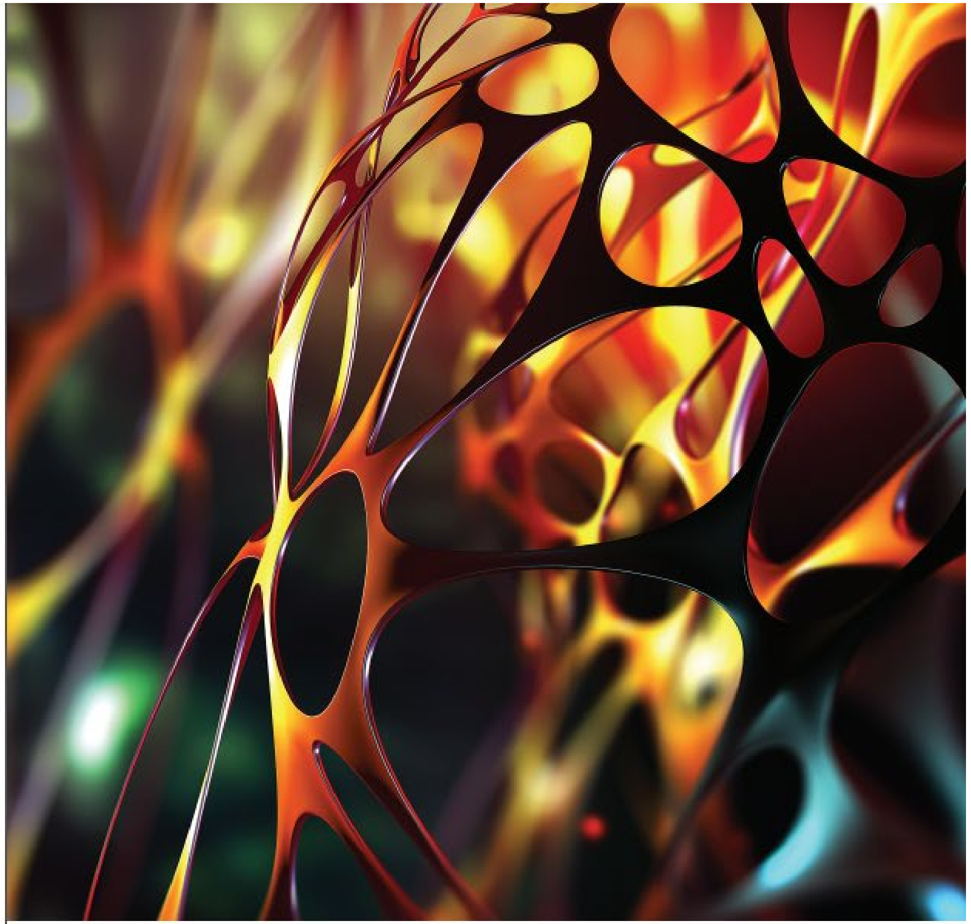
Yoshua Bengio, Yann LeCun, and Geoffrey Hinton are recipients of the 2018 ACM A.M. Turing Award for breakthroughs that have made deep neural networks a critical component of computing.

RESEARCH ON ARTIFICIAL neural networks was motivated by the observation that human intelligence emerges from highly parallel networks of relatively simple, non-linear neurons that learn by adjusting the strengths of their connections. This observation leads to a central computational question: How is it possible for networks of this general kind to learn the complicated internal representations that are required for difficult tasks such as recognizing objects or understanding language? Deep learning seeks to answer this question by using many layers of activity vectors as representations and learning the connection strengths that give rise to these vectors by following the stochastic gradient of an objective function that measures how well the network is performing. It is very surprising that such a conceptually simple approach has proved to be so effective when applied to large training sets using huge amounts of computation and it appears that a key ingredient is depth: shallow networks simply do not work as well.

We reviewed the basic concepts and some of the breakthrough achievements of deep learning several years ago.[63] Here we briefly describe the origins of deep learning, describe a few of the more recent advances, and discuss some of the future challenges. These challenges include learning with little or no external supervision, coping with test examples that come from a different distribution than the training examples, and using the deep learning approach for tasks that humans solve by using a deliberate sequence of steps which we attend to consciously—tasks that Kahneman[56] calls *system 2* tasks as opposed to *system 1* tasks like object recognition or immediate natural language understanding, which generally feel effortless.

From Hand-Coded Symbolic Expressions to Learned Distributed Representations

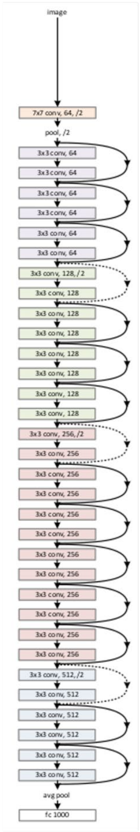
There are two quite different paradigms for AI. Put simply, the logic-inspired paradigm views sequential reasoning as the essence of intelligence and aims to implement reasoning in computers using hand-designed rules of inference that operate on hand-designed symbolic expressions that formalize knowledge. The brain-inspired paradigm views learning representations from data as the essence of intelligence and aims to implement learning by hand-designing or evolving rules for modifying the connection strengths in simulated networks of artificial neurons.



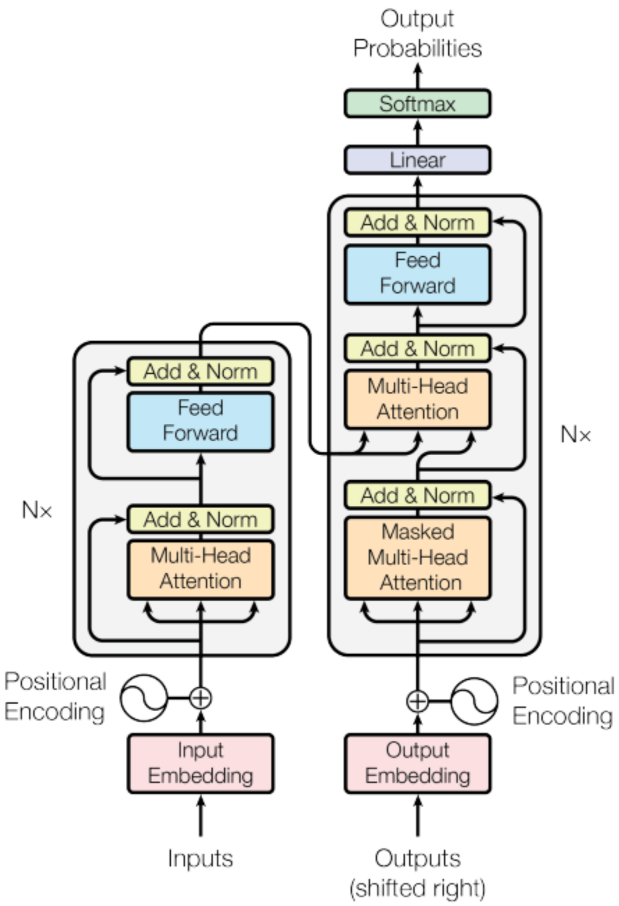
In the logic-inspired paradigm, a symbol has no meaningful internal structure: Its meaning resides in its relationships to other symbols which can be represented by a set of symbolic expressions or by a relational graph. By contrast, in the brain-inspired paradigm the external symbols that are used for communication are converted into internal vectors of neural activity and these vectors have a rich similarity structure. Activity vectors can be used to model the structure inherent in a set of symbol strings by learning appropriate activity vectors for each symbol and learning non-linear transformations that allow the activity vectors that correspond to missing elements of a symbol string to be filled in. This was first demonstrated in Rumelhart et al.[74] on toy data and then by Bengio et al.[14] on real sentences. A very impressive recent demonstration is BERT,[22] which also exploits self-attention to dynamically connect groups of units, as described later.

The main advantage of using vectors of neural activity to represent concepts and weight matrices to capture relationships between concepts is that this leads to automatic generalization. If Tuesday and Thursday are represented by very similar vectors, they will have very similar causal effects on other vectors of neural activity. This facilitates analogical reasoning and suggests that immediate, intuitive analogical reasoning is our primary mode of reasoning, with logical sequential reasoning being a much later development,[56] which we will discuss.

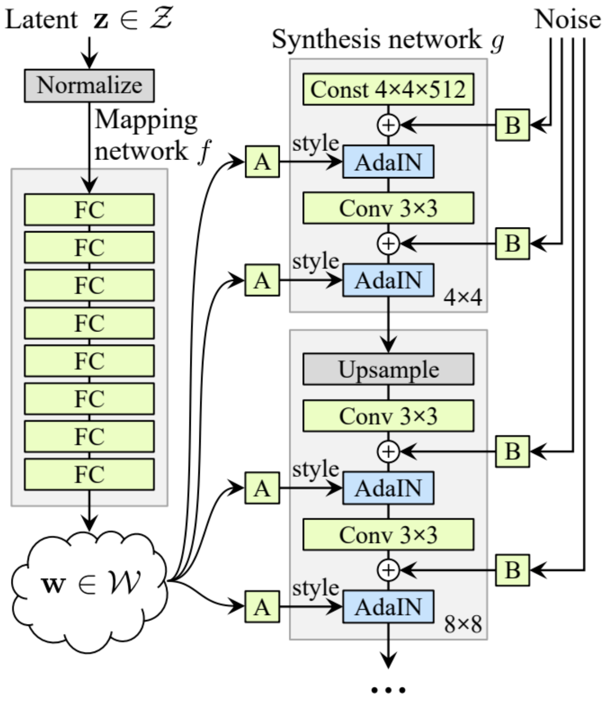
Innovation in machine learning models



ResNet

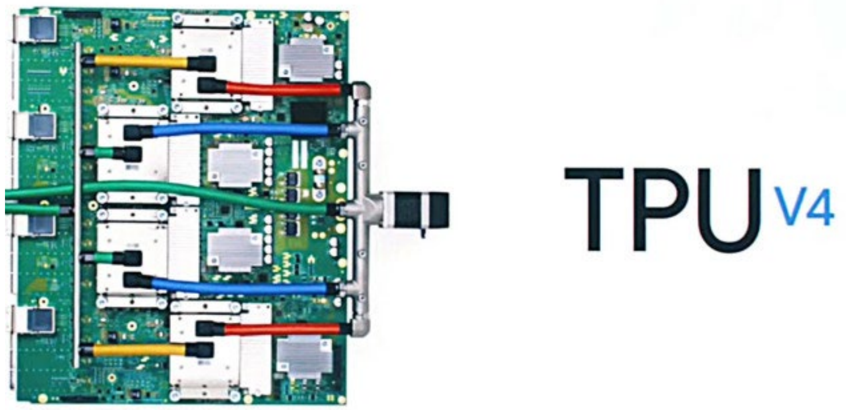


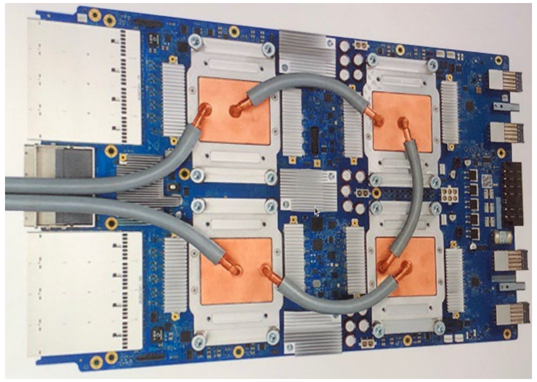
Transformer



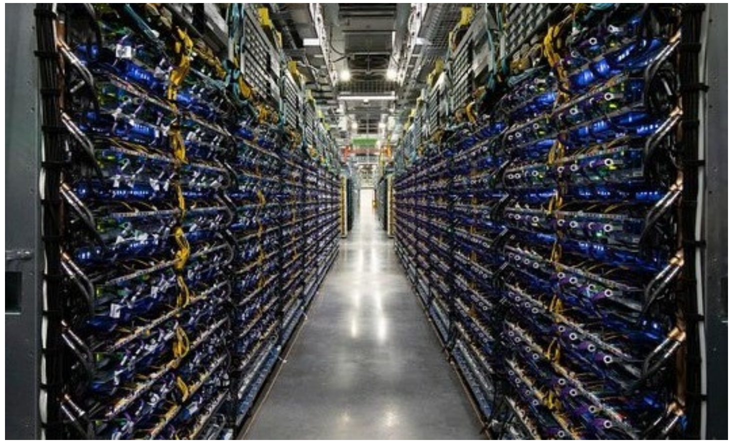
Generative network (StyleGAN2)

Growth in compute-power



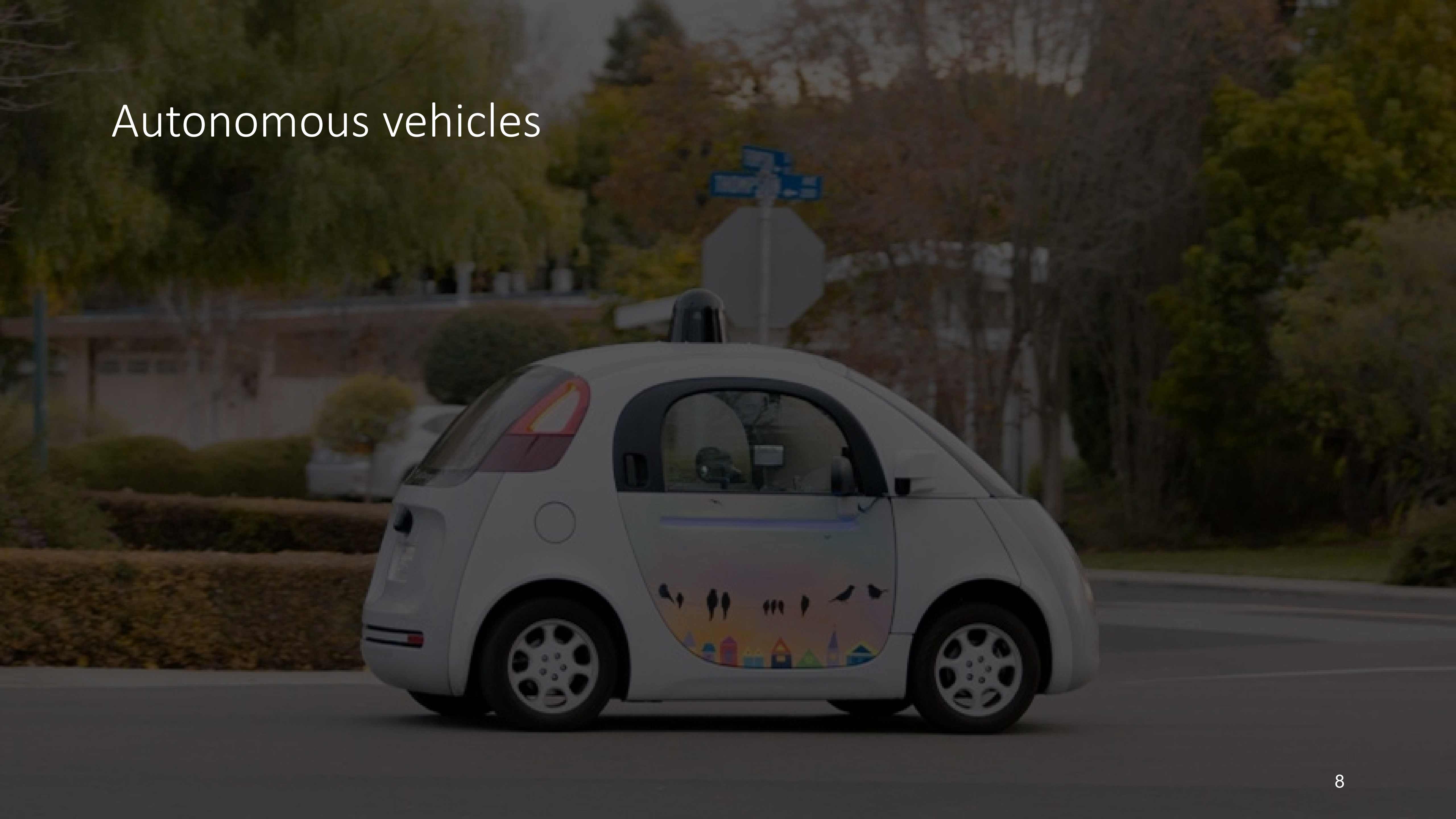


Google TPU v3
420 teraflops

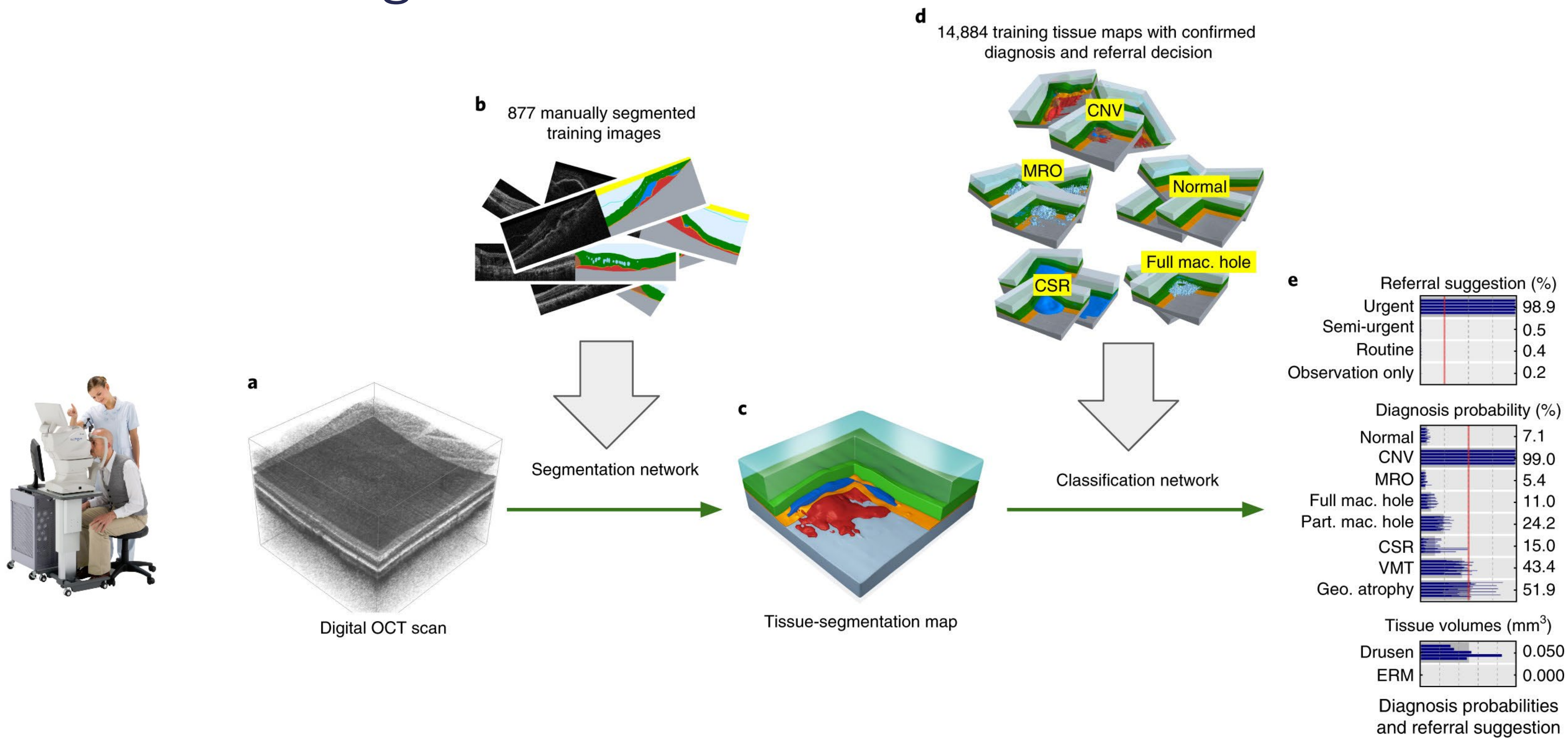


4096 x TPU v4 = 1 exaflop
(10^{18} floating pointing operations per second)

Autonomous vehicles



Medical diagnosis



De Faux et al., Clinically applicable deep learning for diagnosis and referral in retinal disease, Nature Medicine, volume 24, 2018

Stochastic processes: sequence models

$$p(\mathbf{x}_n | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n-1})$$

Text generation from language models – GPT-3

Title: Star's Tux Promise Draws Megyn Kelly's Sarcasm
Subtitle: Joaquin Phoenix pledged to not change for each awards event
Article: A year ago, Joaquin Phoenix made headlines when he appeared on the red carpet at the Golden Globes wearing a tuxedo with a paper bag over his head that read, "I am a shape-shifter. I can't change the world. I can only change myself." It was a promise to not change to fit into the Hollywood mold: "I think that's a really special thing, to not change yourself. I think it's a really special thing to say, `This is what's inside of me, I'm proud of it, and I'm not going to be ashamed because of the way that someone else thinks I should be.'" Now, it's the Oscars, and Phoenix is at it again. But this time, his publicist is saying he'll be wearing a tux no matter what. Megyn Kelly was not impressed, and she let him have it on The Tonight Show. "You know, I feel like, I feel like you could have worn the tux," she says. "But you're saying you're a shape-shifter. I don't know if you can change your tux, but you can change your mind. You can change your mind. You can change your mind." Phoenix says he did, but it didn't stick. "I was like, `Okay, I'm going to wear a tuxedo to this thing.' And then I thought, `I don't want to wear a tuxedo to this thing.'" Kelly goes on to encourage him to change his mind again, but Phoenix says it's too late: "I'm committed to wearing this."

Brown, T. et al. Language models are few-shot learners, NeurIPS, 2020

Few-shot learning with GPT-3

Translate English to French,
would you like breakfast => voulez-vous un petit-dejeuner
it is raining => il pleut

GATO

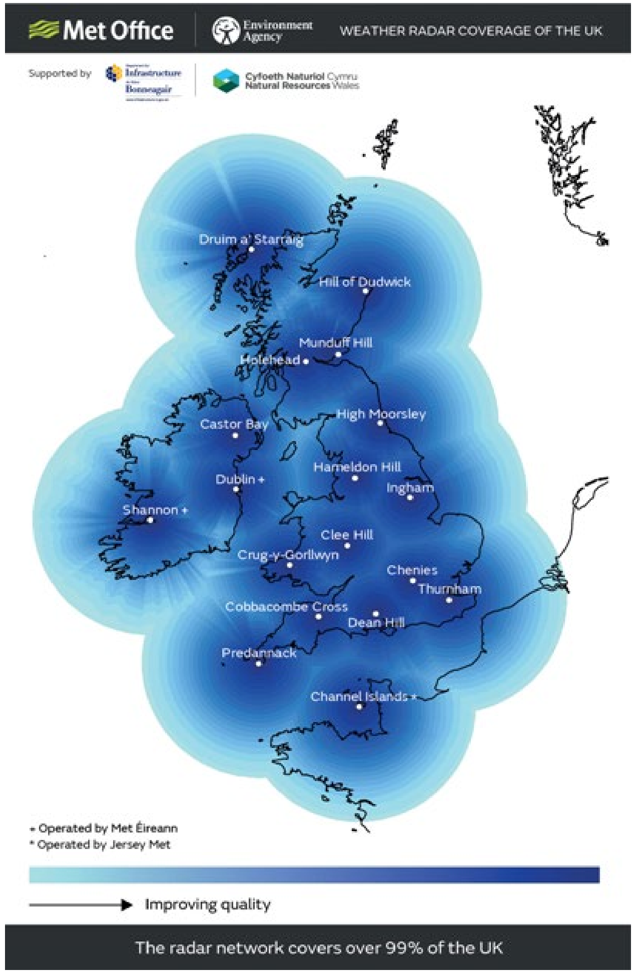
A transformer network trained with several kinds of data on multiple tasks.

Trained on several datasets/tasks:

- Text (similar to GPT-3)
- Vision+language (e.g. images and captions)
- Simulated control tasks (e.g. Atari games)
- Robotic block stacking

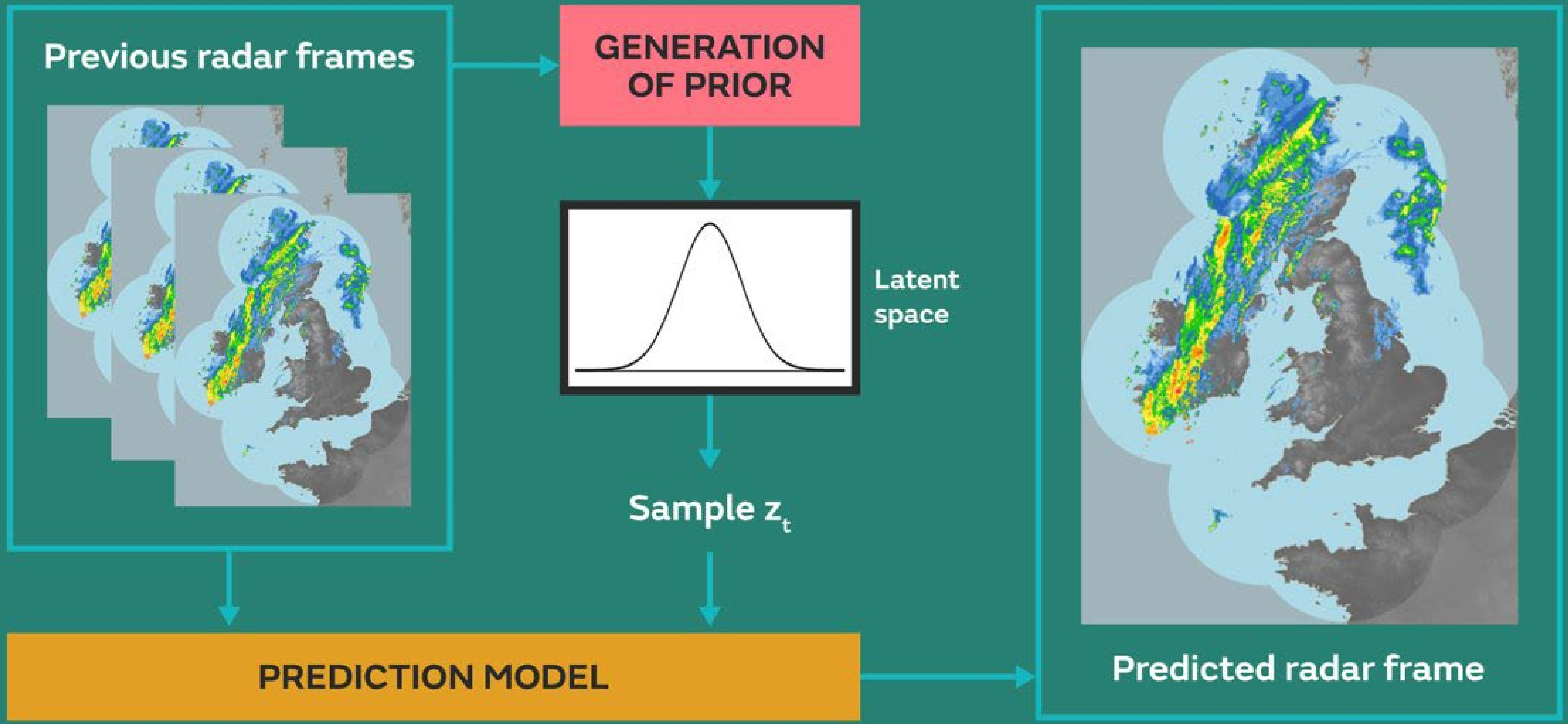
Reed et al., A Generalist Agent, May 2022, arXiv:2205.06175

Precipitation nowcasting



Variational Autoencoder with latent variable

Denton and Fergus, ICML 2018

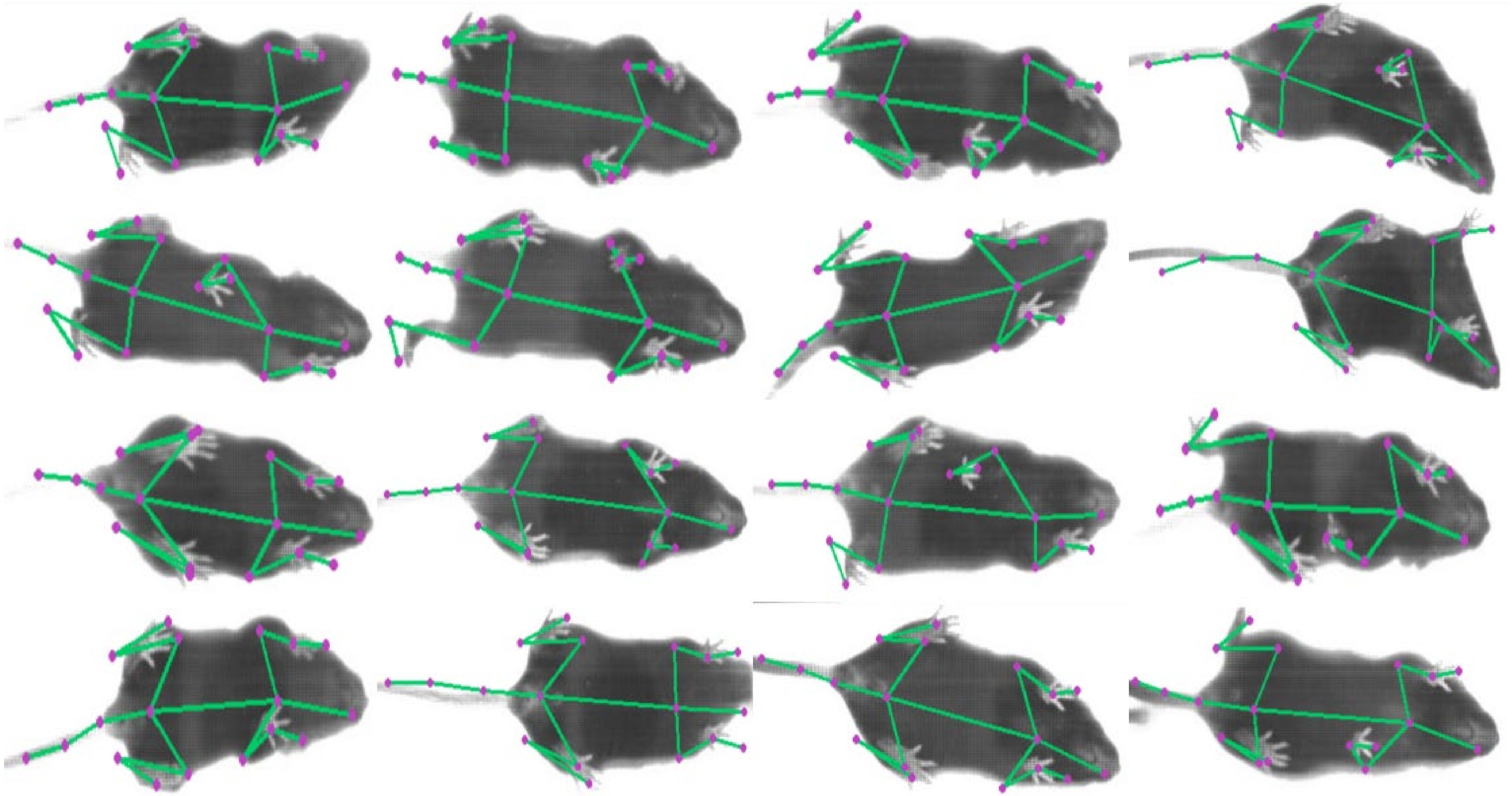


Ense

appr

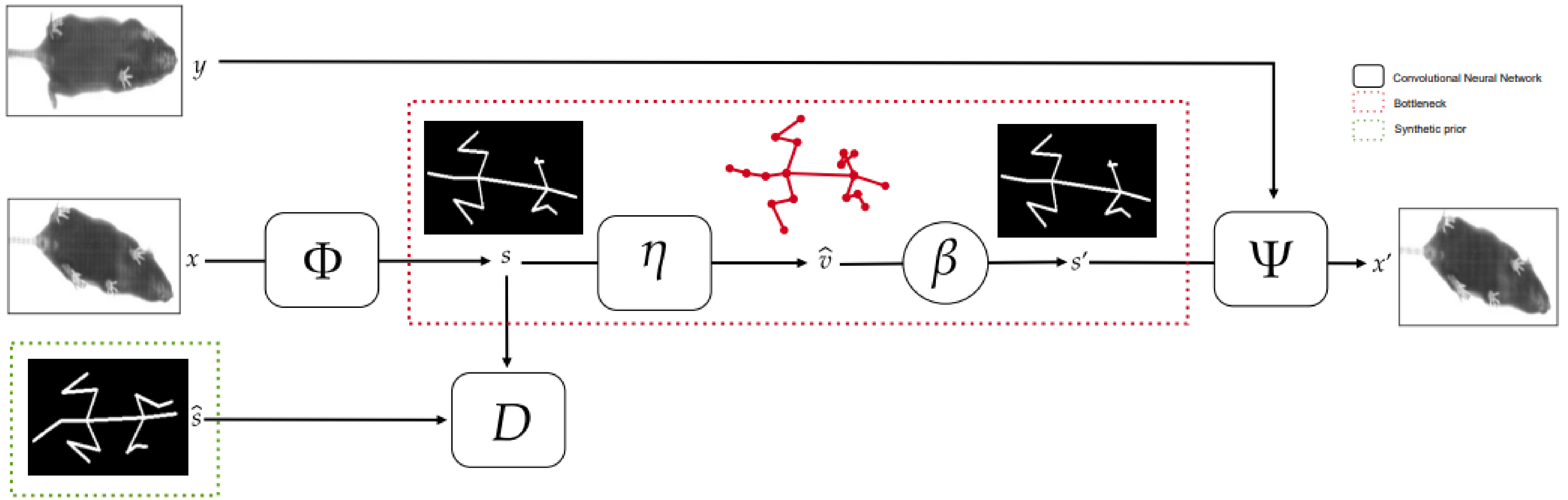
Claire Bartholomew, John Marsham, Teil Howard, Leif Denby

Animal tracking for Biological Sciences

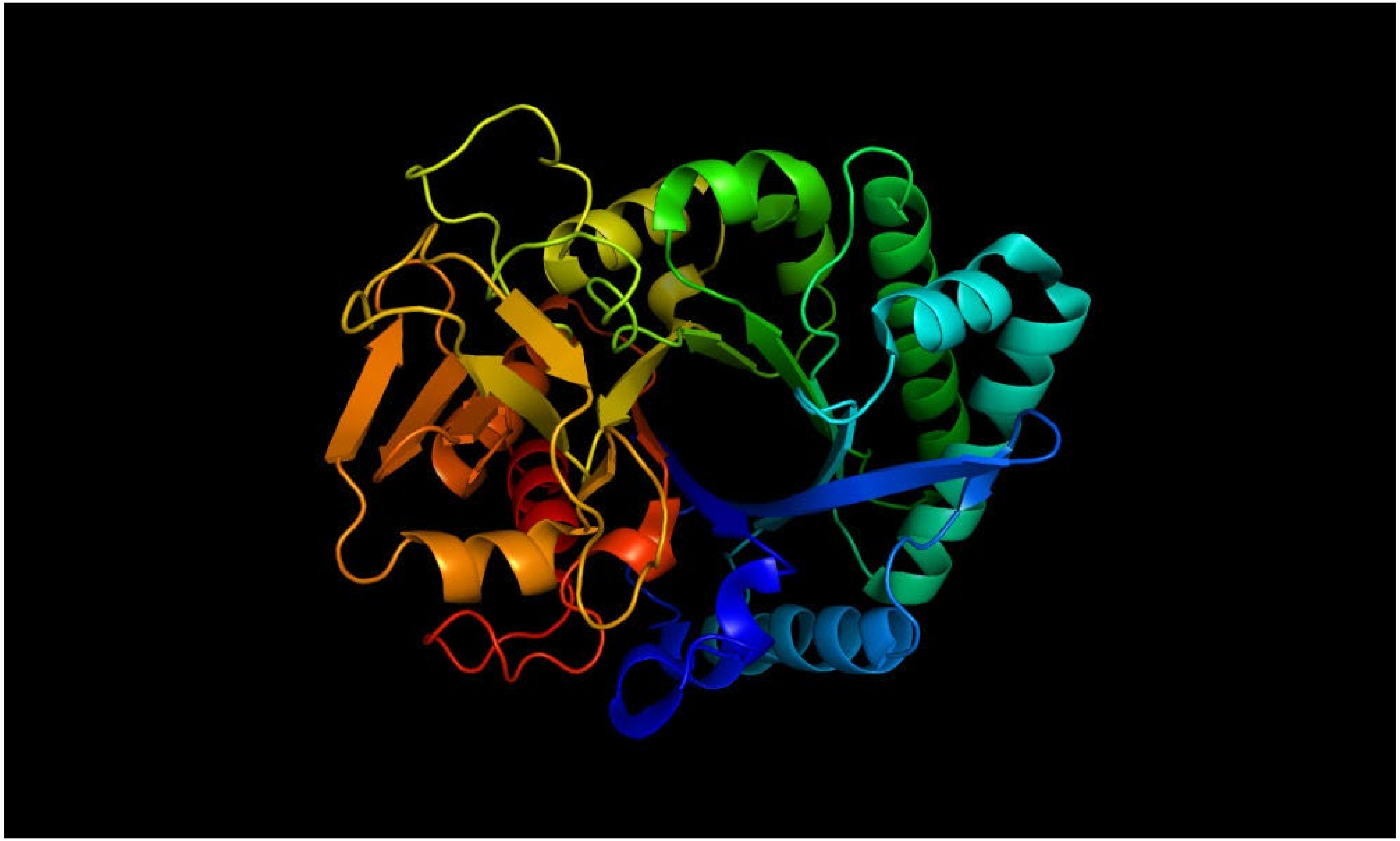


Jose Sosa Martinez

Emergent representation



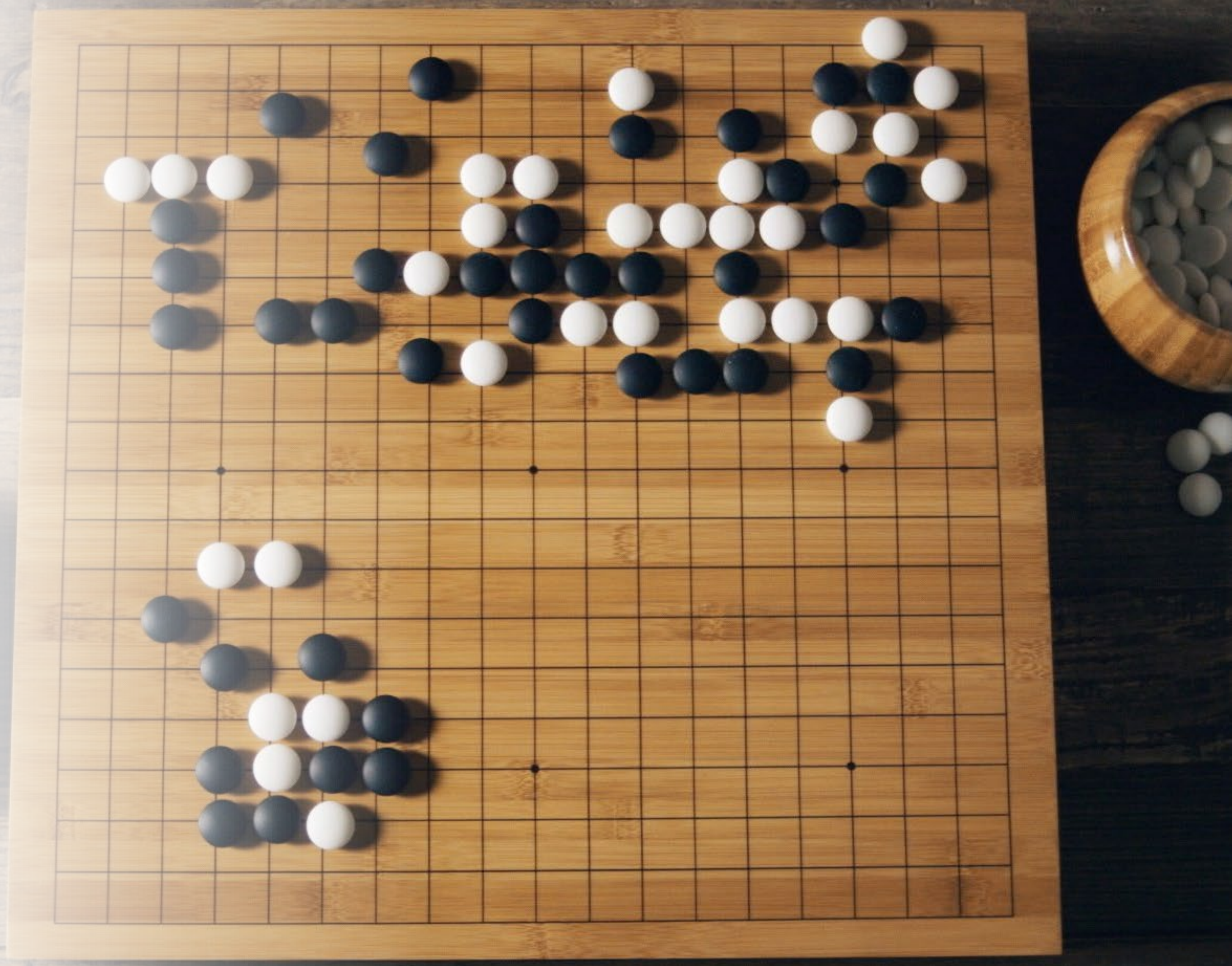
Protein 3D structure prediction from amino acid sequence



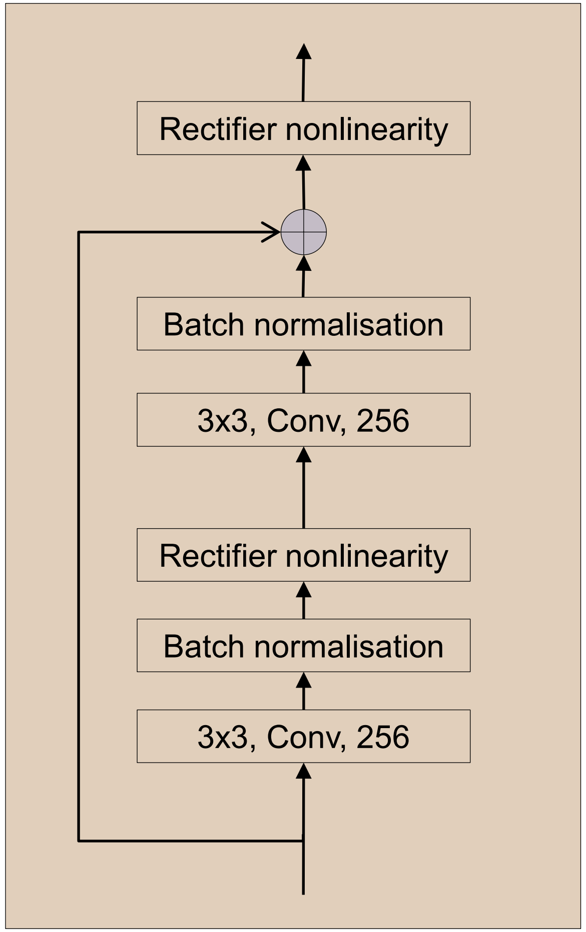
Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021)

Learning to play games through self-play

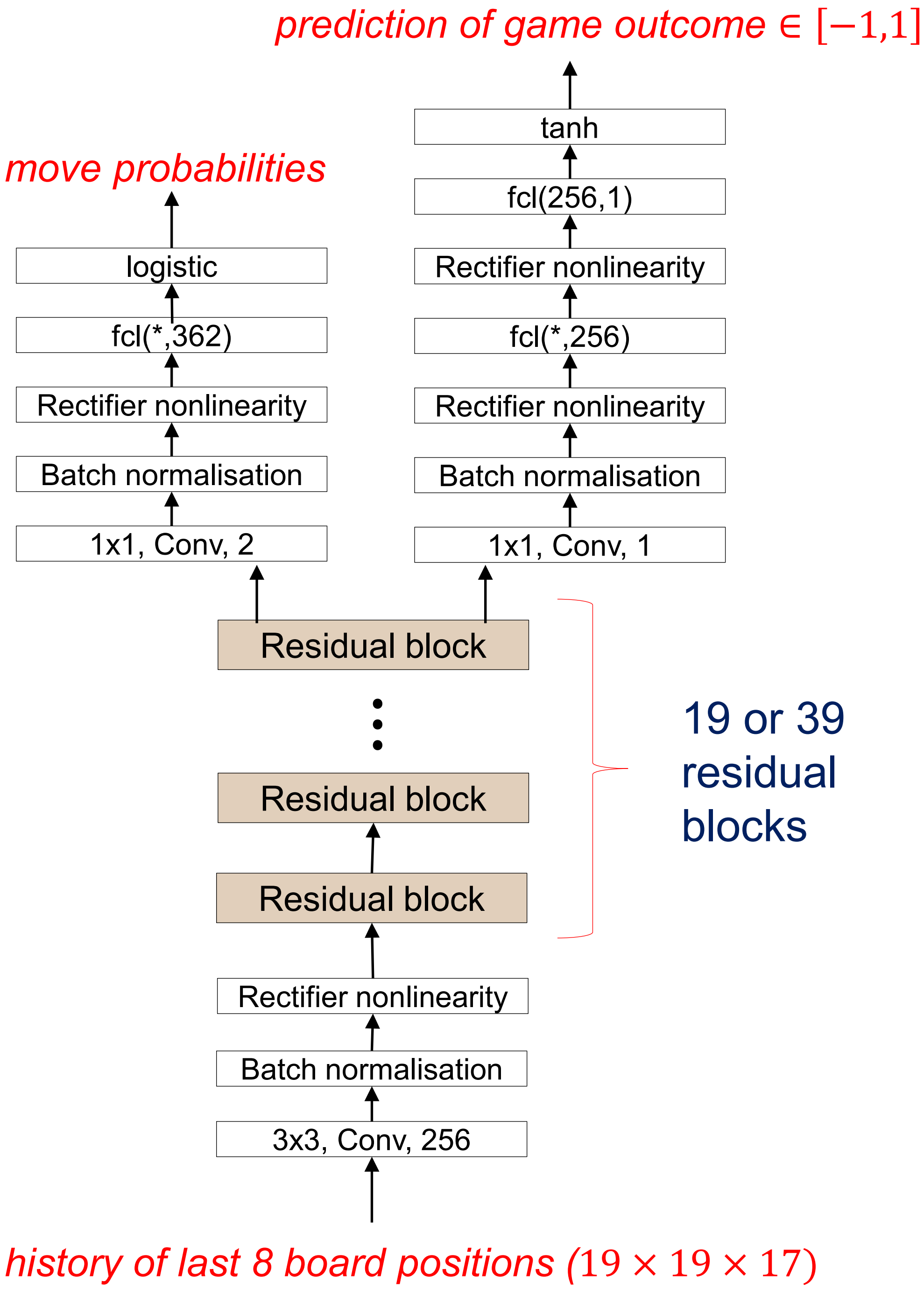
- *D. Silver, et al., Mastering the game of Go without human knowledge, Nature vol. 550, 2017.*



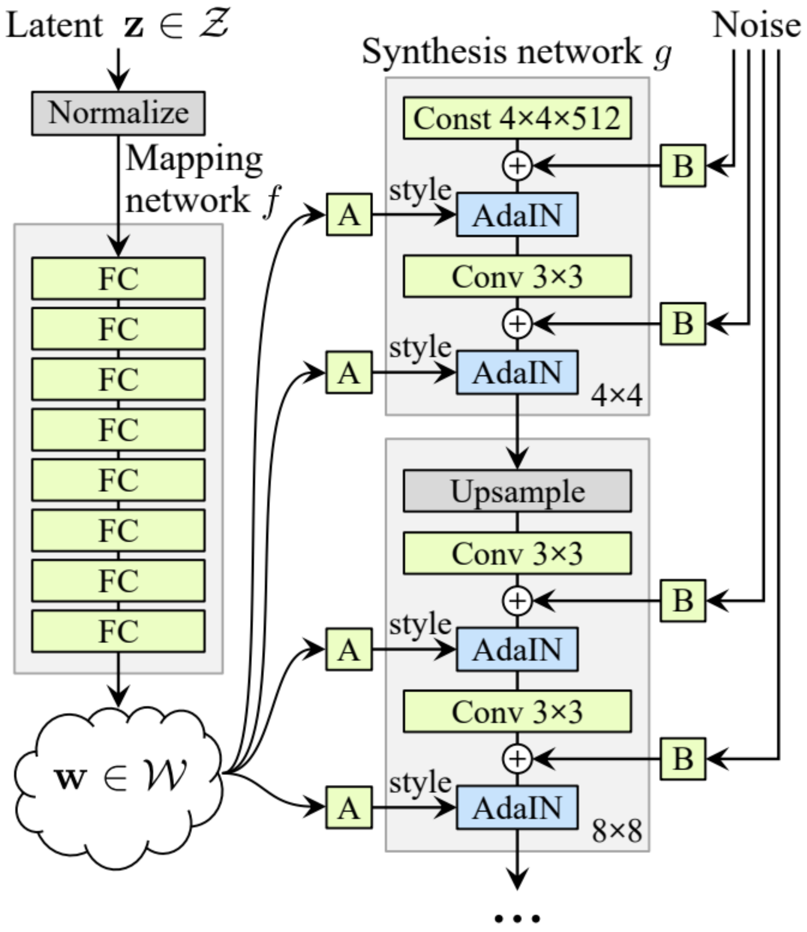
Evaluation of board positions



Residual block

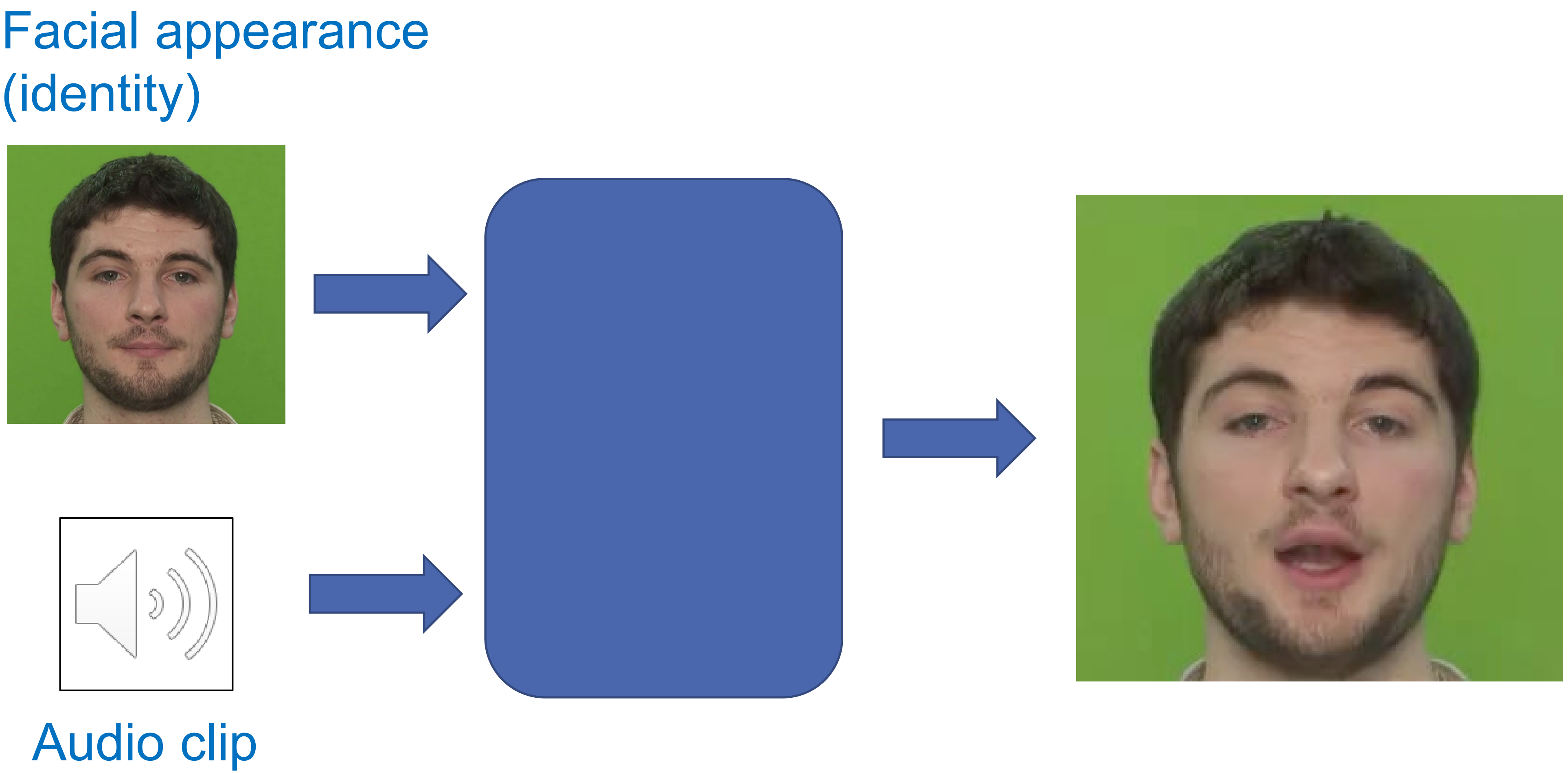


Generative Networks



Tero Karras, Samuli Laine, Timo Aila. A Style-Based Generator Architecture for Generative Adversarial Networks, CVPR 2019.

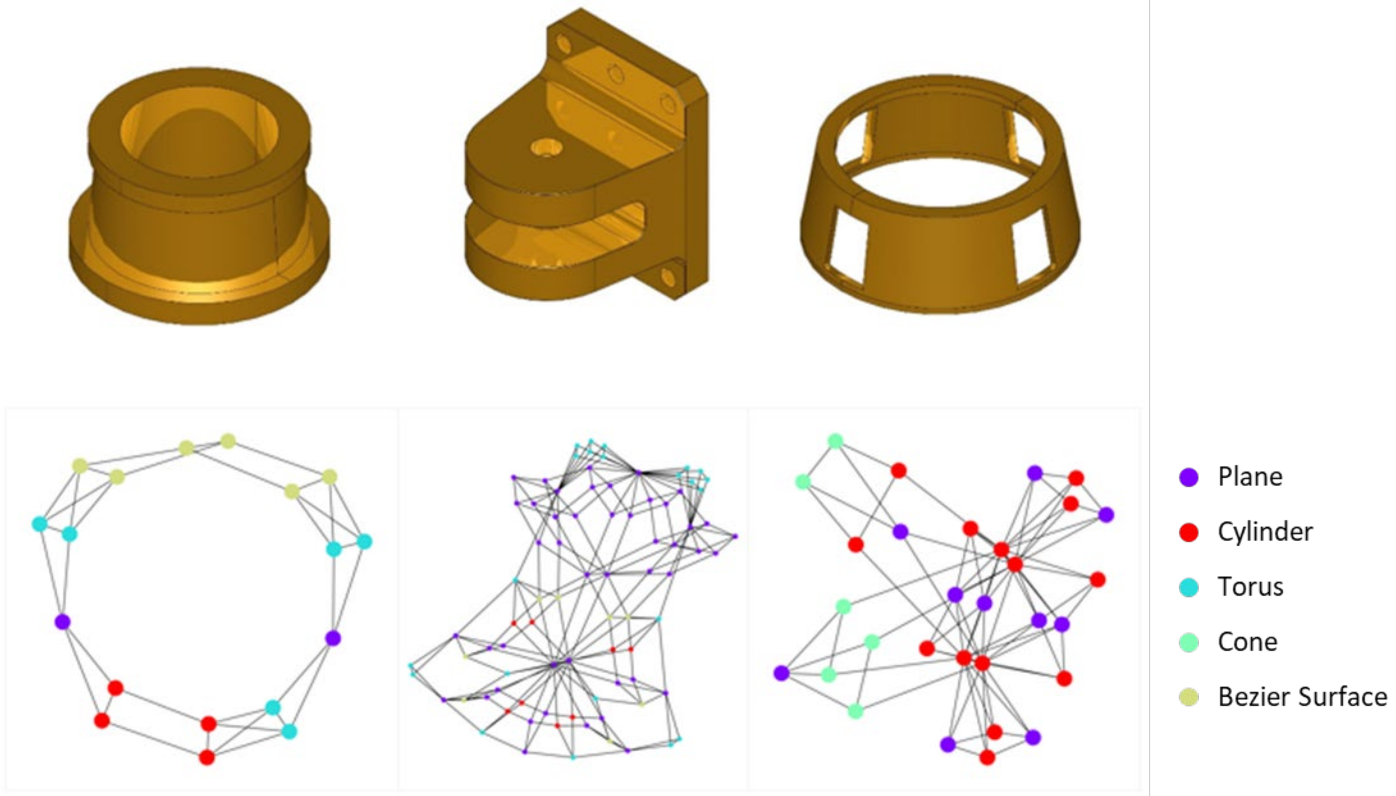
Video synthesis from audio

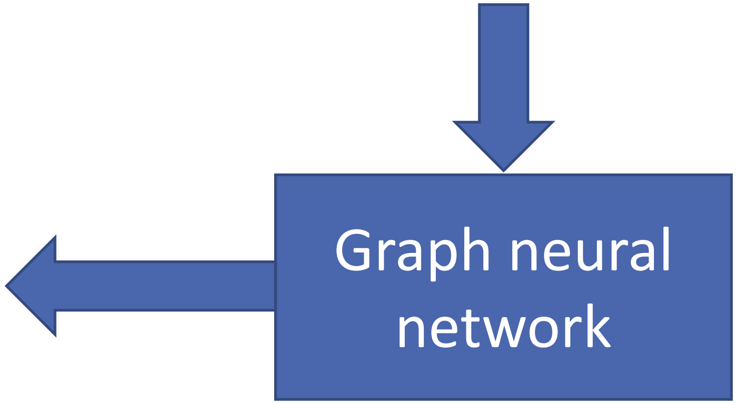


Mohammed Alghamdi, Andy Bulpitt

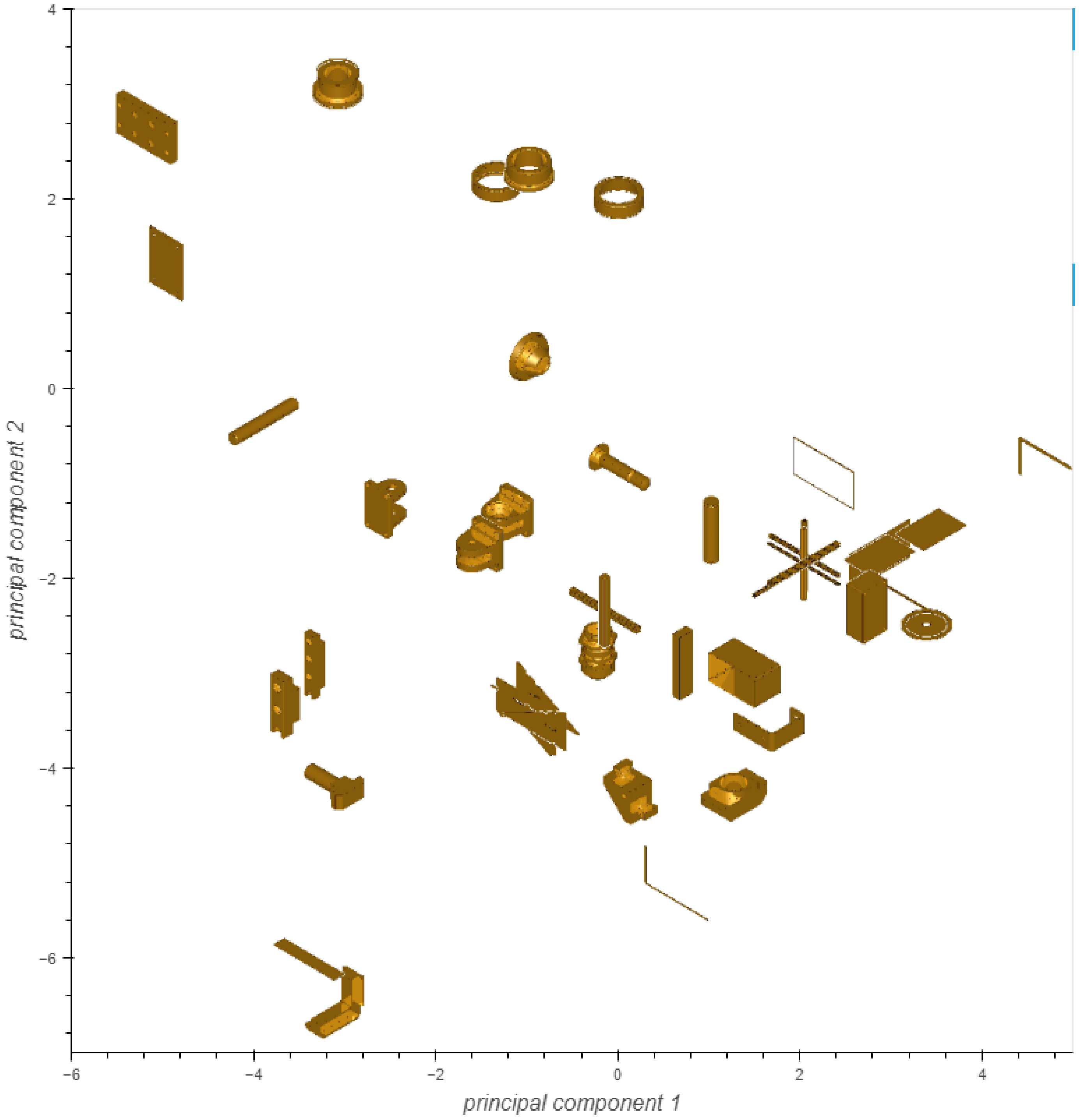
Engineering Design

Task: *finding similar parts*





Latent vector space where similar parts are close-by



Visualisation of manufactured parts in latent space

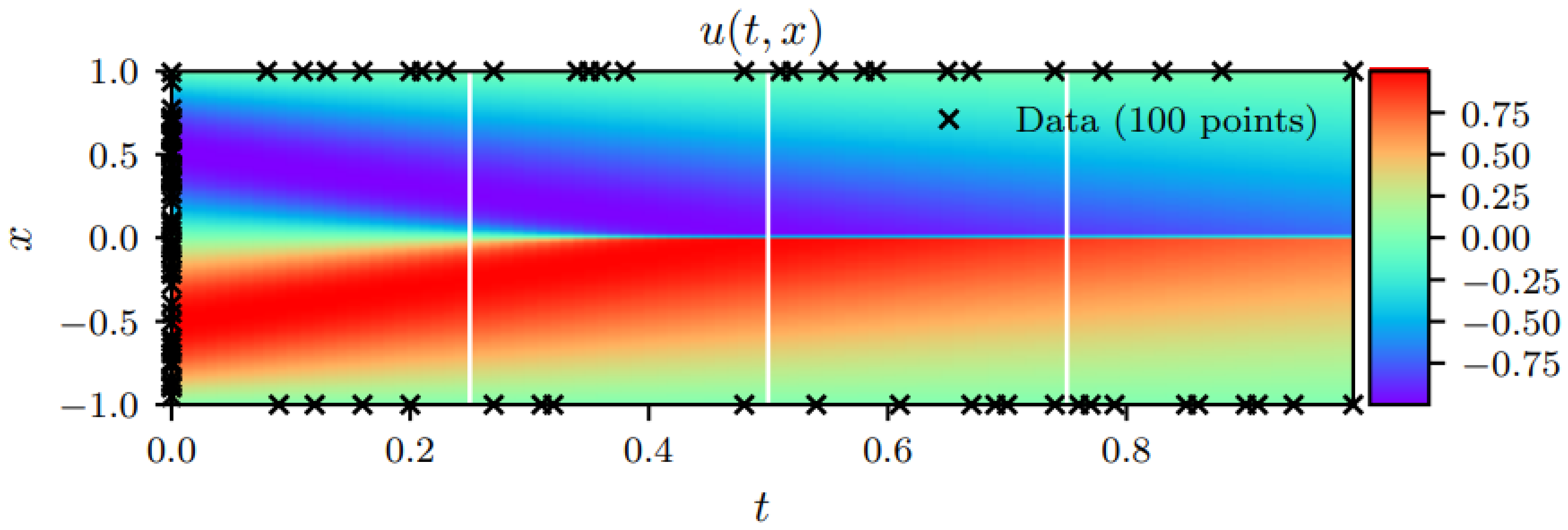
Solving PDEs

$$u_t + uu_x - (0.01/\pi)u_{xx} = 0$$

$$x \in [-1,1], \qquad t \in [0,1]$$

$$u(0,x) = -sin(\pi x)$$

$$u(t,-1) = u(t,1) = 0$$



Example with Burger's equation from: *Maziar Raissi et al., Physics Informed Deep Learning (Part 1): Data-driven Solutions of Nonlinear Partial Differential Equations, arXiv:1711.10561, 2017.*