AI Meetup News

July & August 2018

Talks etc:

Romanes Lecture: The Pacification of Cyberspace

by Dr Vint Cerf - Co-inventor of the Internet

6 November 2018, 17:45 (Tuesday, 5th week, Michaelmas 2018)
Sheldonian Theatre
Booking url: http://www.ox.ac.uk/news-and-events/The-University-Year/romanes-lecture
Cost: Free of charge
Audience: Public

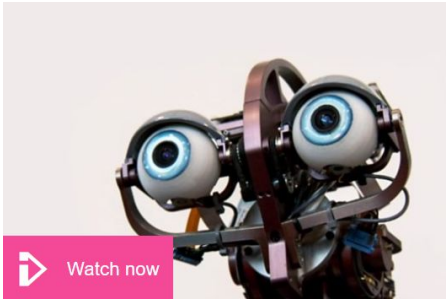
The lecture will discuss how to pacify the relatively lawless environment of the internet, while preserving the utility of its openness to creative innovation and technological revolution.

The AI Conference

https://conferences.oreilly.com/artificial-intelligence/
8-11 Oct 2018
London

TV & Radio:

The Horizon Guide to AI - https://www.bbc.co.uk/programmes/b0bhwhw3



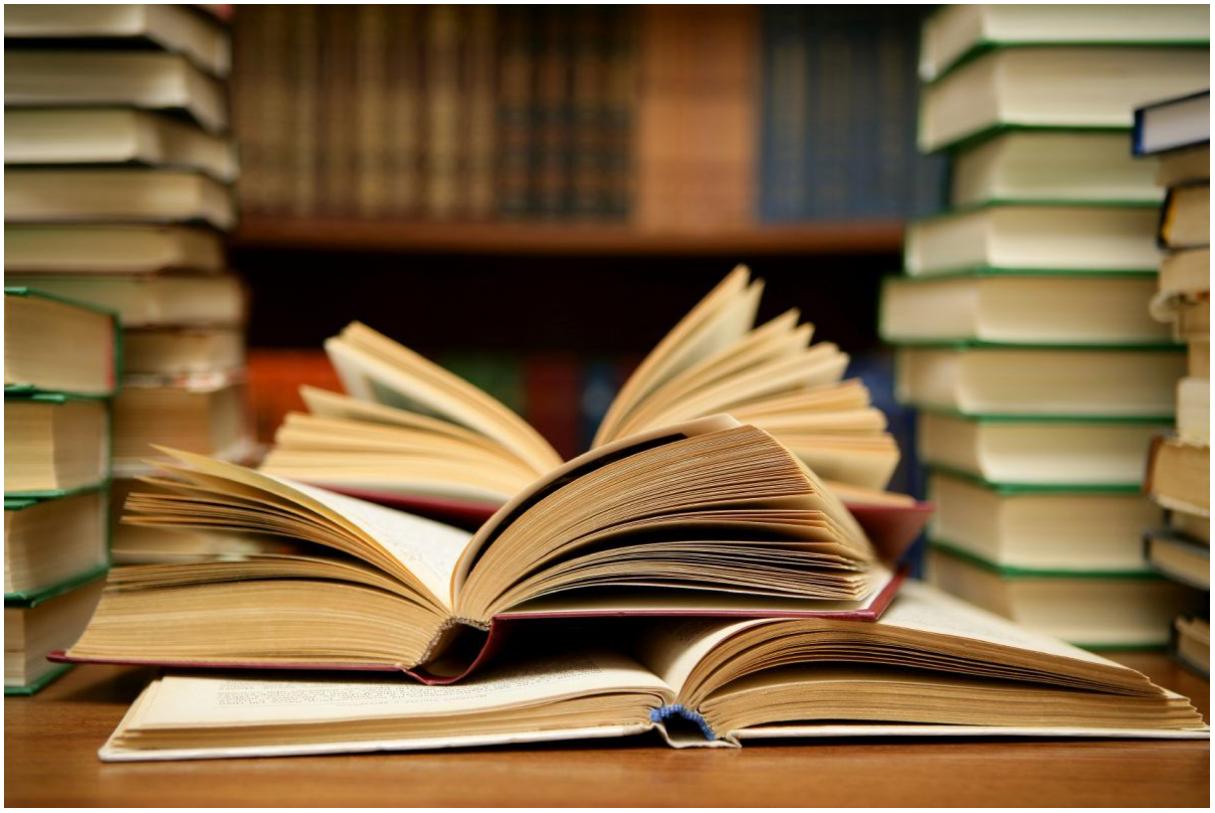
As the Horizon archive shows, throughout each decade once we have learnt to live with the new emerging technology of the time, the pattern begins again. We become once more optimistic, before becoming fearful of it.

AI & Morality in the 21st Century: https://www.bbc.co.uk/programmes/p06k4ws1

DeepMind papers at the International Conference on Machine Learning (ICML) 2018

DeepMind https://deepmind.com/blog/deepmind-papers-icml-2018/

Lots of interesting papers!



The Pentagon plans to spend $2 billion to put more artificial intelligence into its weaponry

https://www.theverge.com/2018/9/8/17833160/pentagon-darpa-artificial-intelligence-ai-investment

Officials say they want computers to be capable of **explaining their decisions** to military commanders

Turning more military analytical work – and potentially some key decision-making – over to computers and algorithms installed in weapons capable of acting violently against humans is controversial.

Google had been leading the Project Maven project for the department, but after an organized protest by Google employees who didn't want to work on software that could help pick out targets for the military to kill, the company said in June it would discontinue its work after its current contract expires.

While Maven and other AI initiatives have helped Pentagon weapons systems become better at recognizing targets and doing things like flying drones more effectively, fielding computer-driven systems that take lethal action on their own hasn't been approved to date.

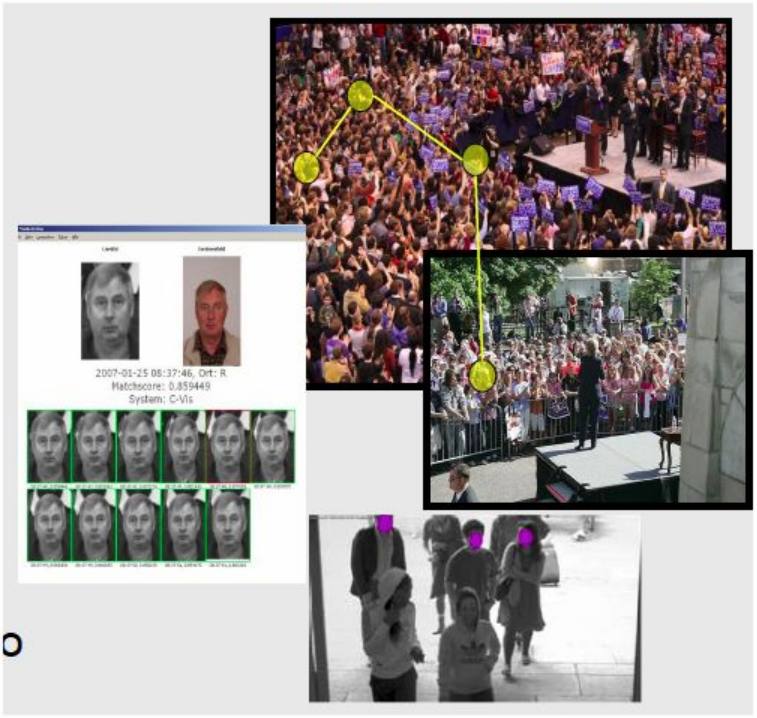
Face Off: Law Enforcement Use of Face Recognition Technology

https://www.eff.org/wp/law-enforcement-use-face-recognition

[The Electronic Frontier Foundation is the leading nonprofit organization defending civil liberties in the digital world]

“This white paper takes a broad look at the problems with law enforcement use of face recognition technology in the United States.

- Part 1 provides an overview of the key issues with face recognition, including accuracy, security, and impact on privacy and civil rights.
- Part 2 focuses on FBI’s face recognition programs, because FBI not only manages the repository for most of the criminal data used by federal, state, local, and tribal law enforcement agencies across the United States, but also provides direct face recognition services to many of these agencies, and its systems exemplify the wider problems with face recognition.
- Part 3 looks ahead to potential future face recognition capabilities and concerns.
- Part 4 presents recommendations for policy makers on the limits and checks necessary to ensure that law enforcement use of face recognition respects civil liberties.”



AI revolution 'at risk of being stifled in UK by fear-driven backlash'... or ... “The Joy of AI?”

https://www.bbc.co.uk/programmes/b0bhwhw1

Prof Jim Al-Khalili, a physicist and the incoming president of the British Science Association, warns that without greater transparency and public engagement the full potential of AI may not be realised.

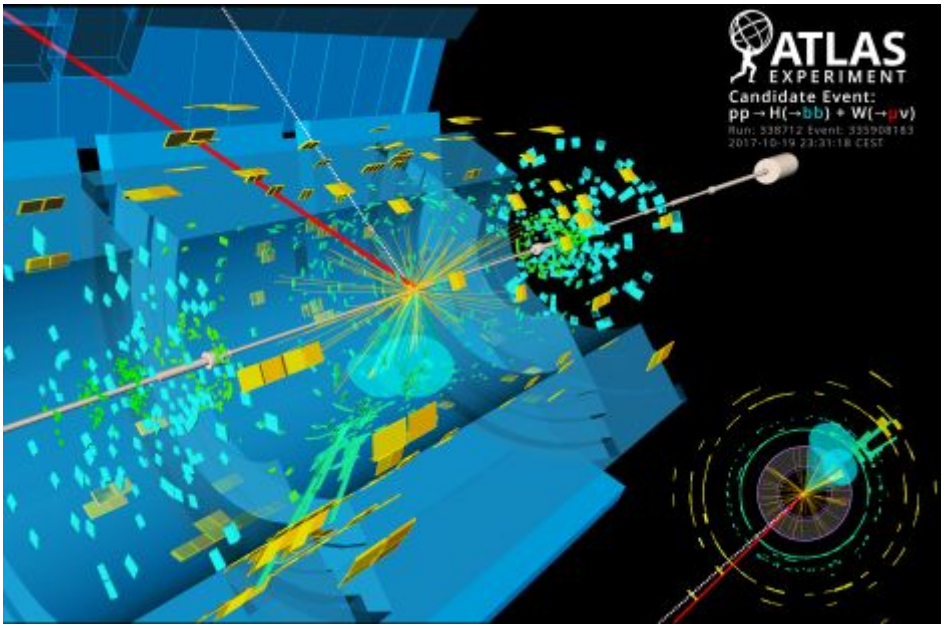
Khalili said: "There's a real danger of a public backlash against AI, potentially similar to the one we had with GM back in the early days of the millennium.

"If the public become disengaged our leaders will see it as less of a priority. Regulations will need to be in place and they may come too late. At the very least, this will result in the technology not being used to its full potential in the public sector, potentially leading to an increase in inequality in society."

Machine Learning used in the detection of Higgs decay

https://www2.physics.ox.ac.uk/news/2018/08/28/a-new-historic-milestone-in-the-study-of-the-higgs-boson

Particle physicists are celebrating that the ATLAS Collaboration at CERN's Large Hadron Collider (LHC) has – at long last – observed the Higgs boson decaying into a pair of bottom (b) quarks. This elusive interaction is predicted to make up almost 60% of the Higgs boson decays. Yet it took over seven years to accomplish this observation.

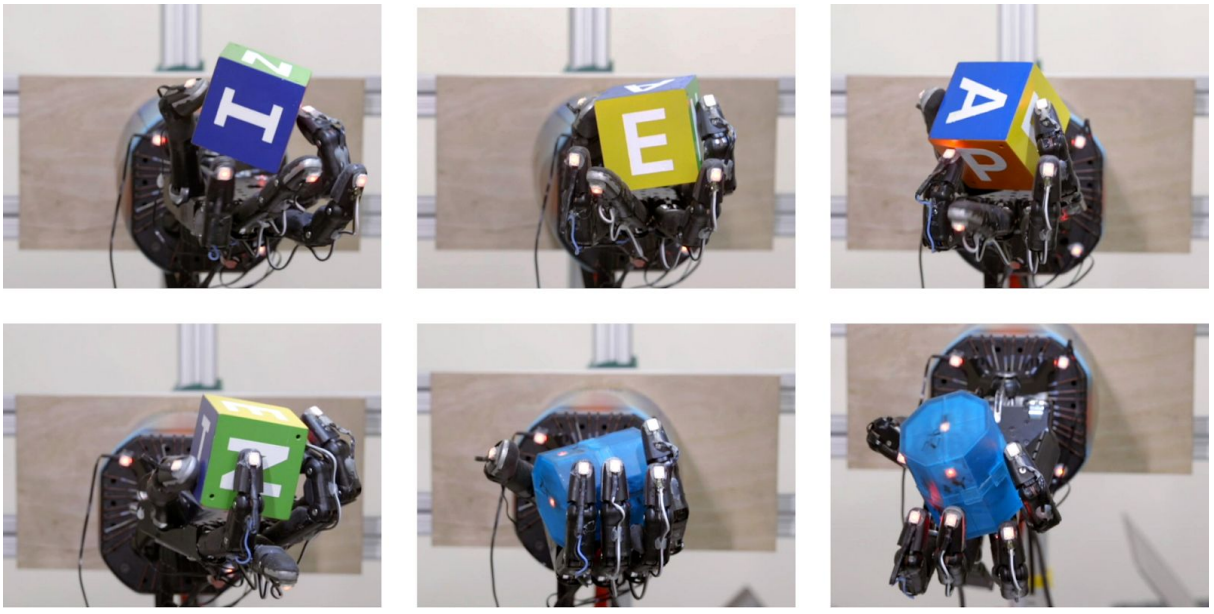


The ATLAS Collaboration first presented a preliminary result of this observation on 9 July at the 2018 International Conference on High-Energy Physics (ICHEP) in Seoul. Today, in a seminar together with the CMS Collaboration, ATLAS presented results which have been submitted for publication to Physics Letters B. They are based on combined Run-2 and Run-1 data, and **utilise machine learning technology** and new analysis techniques to reach a significance of 5.4 standard deviations. UK groups including the Universities of Birmingham, Glasgow, Liverpool, Queen Mary, Oxford and UCL have made important contributions to this historic achievement.

OpenAI: Learning Dexterity

https://blog.openai.com/learning-dexterity/

“We’ve trained a human-like robot hand to manipulate physical objects with unprecedented dexterity. Our system, called Dactyl, is **trained entirely in simulation and transfers its knowledge to reality, adapting to real-world physics** using techniques we’ve been working on for the past year. Dactyl learns from scratch using the same general-purpose reinforcement learning algorithm and code as OpenAI Five. Our results show that it’s possible to train agents in simulation and have them solve real-world tasks, without physically-accurate modeling of the world.”



Dactyl is a system for manipulating objects using a Shadow Dexterous Hand. We place an object such as a block or a prism in the palm of the hand and ask Dactyl to reposition it into a different orientation; for example, rotating the block to put a new face on top. The network observes only the coordinates of the fingertips and the images from three regular RGB cameras.

https://youtu.be/DKe8FumoD4E

Dactyl learns using **Rapid, the massively scaled implementation of Proximal Policy Optimization developed to allow OpenAI Five to solve Dota 2**. We use a different model architecture, environment, and hyperparameters than OpenAI Five does, but we use the same algorithms and training code. Rapid used 6144 CPU cores and 8 GPUs to train our policy, collecting about one hundred years of experience in 50 hours.

AI for cybersecurity is a hot new thing—and a dangerous gamble

https://www.technologyreview.com/s/611860/ai-for-cybersecurity-is-a-hot-new-thing-and-a-dangerous-gamble

“Plenty of firms are now rolling out machine-learning-based products because they feel they have to in order to get an audience with customers who have bought into the AI hype cycle. And there’s a danger that they will overlook ways in which the machine-learning algorithms could create a false sense of security.

Many products being rolled out involve “supervised learning,” which requires firms to choose and label data sets that algorithms are trained on—for instance, by tagging code that’s malware and code that is clean.

Marty says that one risk is that in rushing to get their products to market, companies use training information that hasn’t been thoroughly scrubbed of anomalous data points. That could lead to the algorithm missing some attacks. Another is that hackers who get access to a security firm’s systems could corrupt data by switching labels so that some malware examples are tagged as clean code.

The bad guys don’t even need to tamper with the data; instead, **they could work out the features of code that a model is using to flag malware and then remove these from their own malicious code so the algorithm doesn’t catch it**.”

IBM unveils Project Debater

http://www.research.ibm.com/artificial-intelligence/project-debater/

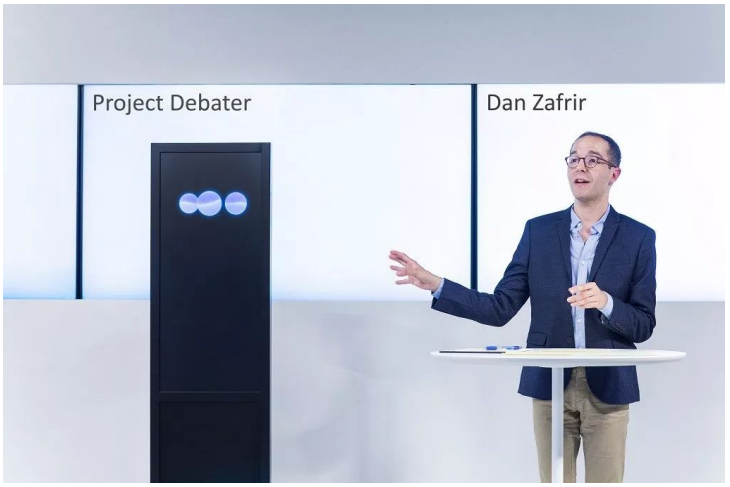
IBM has unveiled Project Debater, a conversational artificial intelligence which can hold a debate using information from journals and newspaper articles.

Project Debater began its training six years ago but it could only hold a debate with people two years ago. A demo was held for journalists at IBM's offices in San Francisco where **the AI debated issues relating to healthcare and the subsidisation of space exploration**.

"Project Debater is the first AI system that can debate humans on complex topics. Project Debater digests massive texts, constructs a well-structured speech on a given topic, delivers it with clarity and purpose, and rebuts its opponent. Eventually, Project Debater will help people reason by providing compelling, evidence-based arguments and limiting the influence of emotion, bias, or ambiguity."

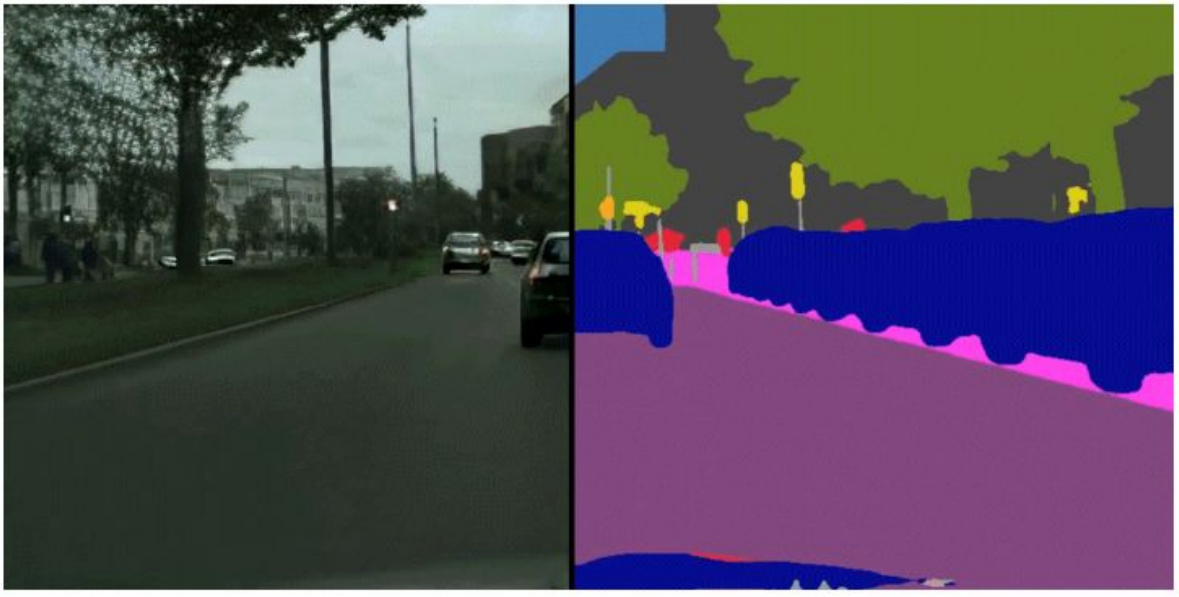
Papers/info:
http://www.research.ibm.com/artificial-intelligence/project-debater/research.html

Critical opinion:
http://theconversation.com/ibms-debating-computer-an-ai-experts-verdict-98783



Video-to-Video Synthesis

https://tcwang0509.github.io/vid2vid/



“In this paper, we propose a novel video-to-video synthesis approach under the generative adversarial learning framework. Through carefully-designed generator and discriminator architectures, coupled with a spatial-temporal adversarial objective, we achieve high-resolution, photorealistic, temporally coherent video results on a diverse set of input formats including segmentation masks, sketches, and poses. Experiments on multiple benchmarks show the advantage of our method compared to strong baselines. In particular, our model is capable of synthesizing 2K resolution videos of street scenes up to 30 seconds long, which significantly advances the state-of-the-art of video synthesis. Finally, we apply our approach to future video prediction, outperforming several state-of-the-art competing systems.”

2-minute papers:
https://youtu.be/GRQuRcpf5Gc

A major milestone for the treatment of eye disease

DeepMind - https://deepmind.com/blog/moorfields-major-milestone/

https://youtu.be/MCI0xEGvHx8?t=1m34s

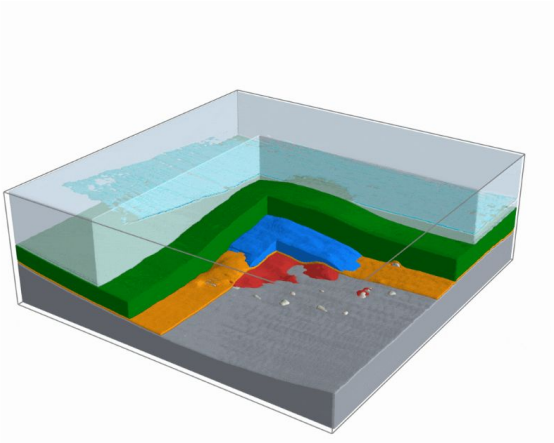
https://www.bbc.co.uk/news/health-44924948

“**Crucially, the algorithm did not miss a single urgent case**.

The results, published in the journal Nature Medicine, were described as "**jaw-dropping**" by Dr Pearse Keane, consultant ophthalmologist, who is leading the research at Moorfields Eye Hospital.”

DeepMind says: “We are delighted to announce the results of the first phase of our joint research partnership with Moorfields Eye Hospital, which could potentially transform the management of sight-threatening eye disease.

The results, published online in Nature Medicine (open access full text, see end of blog), show that our AI system can quickly interpret eye scans from routine clinical practice with unprecedented accuracy. **It can correctly recommend how patients should be referred for treatment for over 50 sight-threatening eye diseases as accurately as world-leading expert doctors.**”



“The AI technology we have developed is able to analyse OCT eye scans and correctly recommend how patients with a wide range of eye diseases, such as age-related macular degeneration, diabetic eye disease and severe myopia, should be referred for treatment. At 94% accuracy, this matches the accuracy of expert clinicians at Moorfields Eye Hospital with over 20 years’ experience in the field.

This alone is hugely promising, but we have also designed our AI technology so that it’s able to explain to clinicians how it arrives at its recommendations, rather than simply presenting them in isolation like a “black box”. This includes information about the features of eye disease the AI technology has identified on the OCT scan, so doctors and eyecare professionals can gain insight into its “thinking”, as well as the level of confidence the system has in its recommendations, in the form of a percentage.”

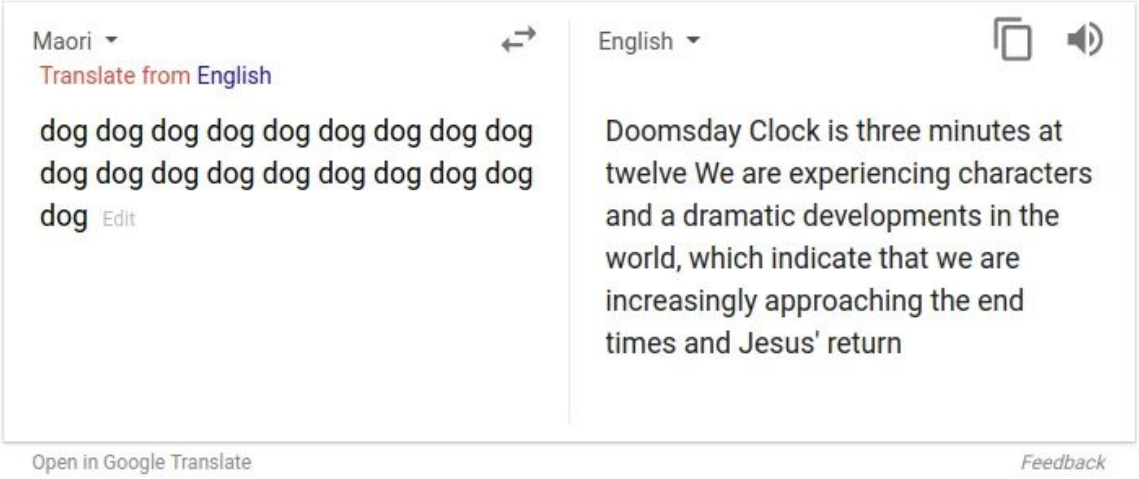
Super SloMo: High Quality Estimation of Multiple Intermediate Frames for Video Interpolation

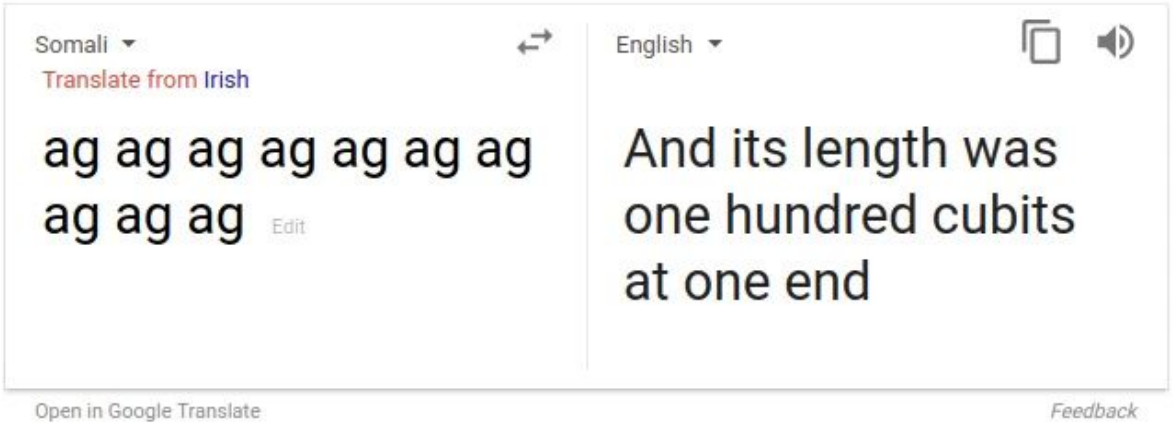
Nvidia: https://people.cs.umass.edu/~hzjiang//projects/superslomo/

“We have proposed an **end-to-end trainable CNN** that can produce as many intermediate video frames as needed between two input images. We first use a flow computation CNN to estimate the bidirectional optical flow between the two input frames, and the two flow fields are linearly fused to approximate the intermediate optical flow fields. We then use a flow interpolation CNN to refine the approximated flow fields and predict soft visibility maps for interpolation. We use more than 1.1K 240-fps video clips to train our network to predict seven intermediate frames. Ablation studies on separate validation sets demonstrate the benefit of flow interpolation and visibility map. Our multi-frame approach consistently outperforms state-of-the-art single frame methods on the Middlebury, UCF101, slowflow, and high-framerate Sintel datasets. For the unsupervised learning of optical flow, our network outperforms the recent DVF method [15] on the KITTI 2012 benchmark.”

Is Google Translate Spitting Out Sinister Religious Prophecies?

https://motherboard.vice.com/en_us/article/j5npeg/why-is-google-translate-spitting-out-sinister-religious-prophecies



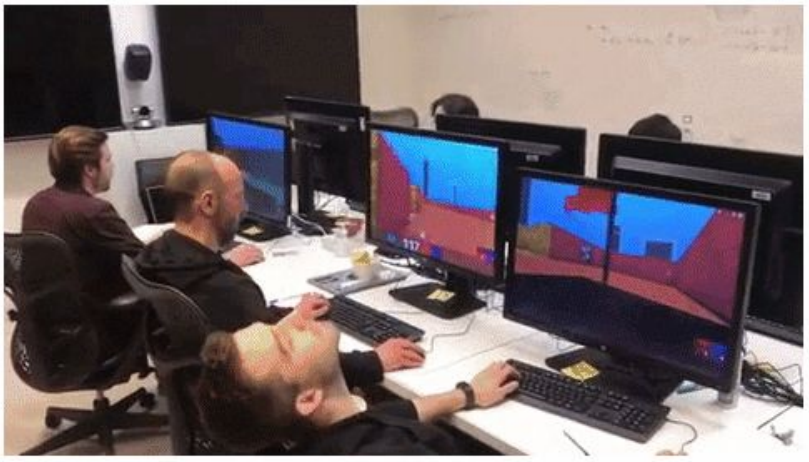


“Rush agreed that if Google is using the Bible to train its model, it could explain some of the strange outputs. Indeed, several of the bizarre translations in Somali resemble specific passages from the Old Testament. Exodus 27:18 references a hundred cubits, and several verses including Numbers 3:18 discuss the sons of Gershon.”

Capture the Flag: the emergence of complex cooperative agents

DeepMind - https://deepmind.com/blog/capture-the-flag/

Through new developments in reinforcement learning, our agents have achieved **human-level performance in Quake III Arena Capture the Flag**, a complex multi-agent environment and one of the canonical 3D first-person multiplayer games. These agents demonstrate the ability to team up with both artificial agents and human players.



An early test tournament with humans playing CTF with and against trained agents and other humans.

Our agents must learn from scratch how to see, act, cooperate, and compete in unseen environments, all from a single reinforcement signal per match: whether their team won or not. This is a challenging learning problem, and its solution is based on three general ideas for reinforcement learning:

- Rather than training a single agent, we train a population of agents, which learn by playing with each other, providing a diversity of teammates and opponents.
- Each agent in the population learns its own internal reward signal, which allows agents to generate their own internal goals, such as capturing a flag. A two-tier optimisation process optimises agents' internal rewards directly for winning, and uses reinforcement learning on the internal rewards to learn the agents' policies.
- Agents operate at two timescales, fast and slow, which improves their ability to use memory and generate consistent action sequences.

Draw This

https://danmacnish.com/2018/07/01/draw-this/

“Instant cameras are a classic that still delights today. There is something eternally amusing about a physical, unique image, that is uniquely different to digital. Playing with neural networks for object recognition one day, I wondered if I could take the concept of an instant camera one step further, and ask the camera to re-interpret the image, printing out a cartoon instead of a faithful photograph.”

Machine Theory of Mind

DeepMind - https://arxiv.org/abs/1802.07740

Theory of mind (ToM; Premack & Woodruff, 1978) broadly refers to humans' ability to represent the mental states of others, including their desires, beliefs, and intentions. We propose to train a machine to build such models too. **We design a Theory of Mind neural network -- a ToMnet -- which uses meta-learning to build models of the agents it encounters, from observations of their behaviour alone**. Through this process, it acquires a strong prior model for agents' behaviour, as well as the ability to bootstrap to richer predictions about agents' characteristics and mental states using only a small number of behavioural observations. **We apply the ToMnet to agents behaving in simple gridworld environments, showing that it learns to model random, algorithmic, and deep reinforcement learning agents from varied populations**, and that it passes classic ToM tasks such as the "Sally-Anne" test (Wimmer & Perner, 1983; Baron-Cohen et al., 1985) of recognising that others can hold false beliefs about the world. We argue that this system -- which autonomously learns how to model other agents in its world -- is an important step forward for developing multi-agent AI systems, for building intermediating technology for machine-human interaction, and for advancing the progress on interpretable AI.

AI drone pilots will challenge humans in competition sponsored by Lockheed Martin

https://www.theverge.com/2018/9/5/17822384/ai-autonomous-drone-quadcopter-pilot-drl-drone-racing-league-2019

“The Drone Racing League (DRL), which is one of the foremost organizations trying to turn drone racing into the next NASCAR, announced a new competition for teams to develop AI pilots for its aircraft.

Up for grabs is more than $2 million in prizes, with a one-off $200,000 reward for the first AI team to beat a professional human pilot.”

https://youtu.be/SrqrGweKQAU?t=54s

Evolutionary algorithm outperforms deep-learning machines at video games

https://www.technologyreview.com/s/611568/evolutionary-algorithm-outperforms-deep-learning-machines-at-video-games

“Neural networks, after all, have begun to outperform humans in tasks such as object and face recognition and in games such as chess, Go, and various arcade video games.

Not quite. An entirely different type of computing has the potential to be significantly more powerful than neural networks and deep learning. This technique is based on the process that created the human brain—evolution. In other words, a sequence of iterative change and selection that produced the most complex and capable machines known to humankind—the eye, the wing, the brain, and so on. The power of evolution is a wonder to behold.”

Discussion of “**Evolving simple programs for playing Atari games**” - https://arxiv.org/abs/1806.05695

“Cartesian Genetic Programming (CGP) has previously shown capabilities in image processing tasks by evolving programs with a function set specialized for computer vision. A similar approach can be applied to Atari playing. Programs are evolved using mixed type CGP with a function set suited for matrix operations, including image processing, but allowing for controller behavior to emerge. While the programs are relatively small, many controllers are competitive with state of the art methods for the Atari benchmark set and require less training time. By evaluating the programs of the best evolved individuals, simple but effective strategies can be found.”

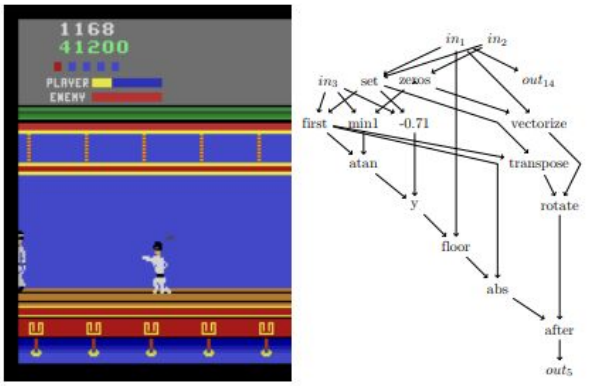
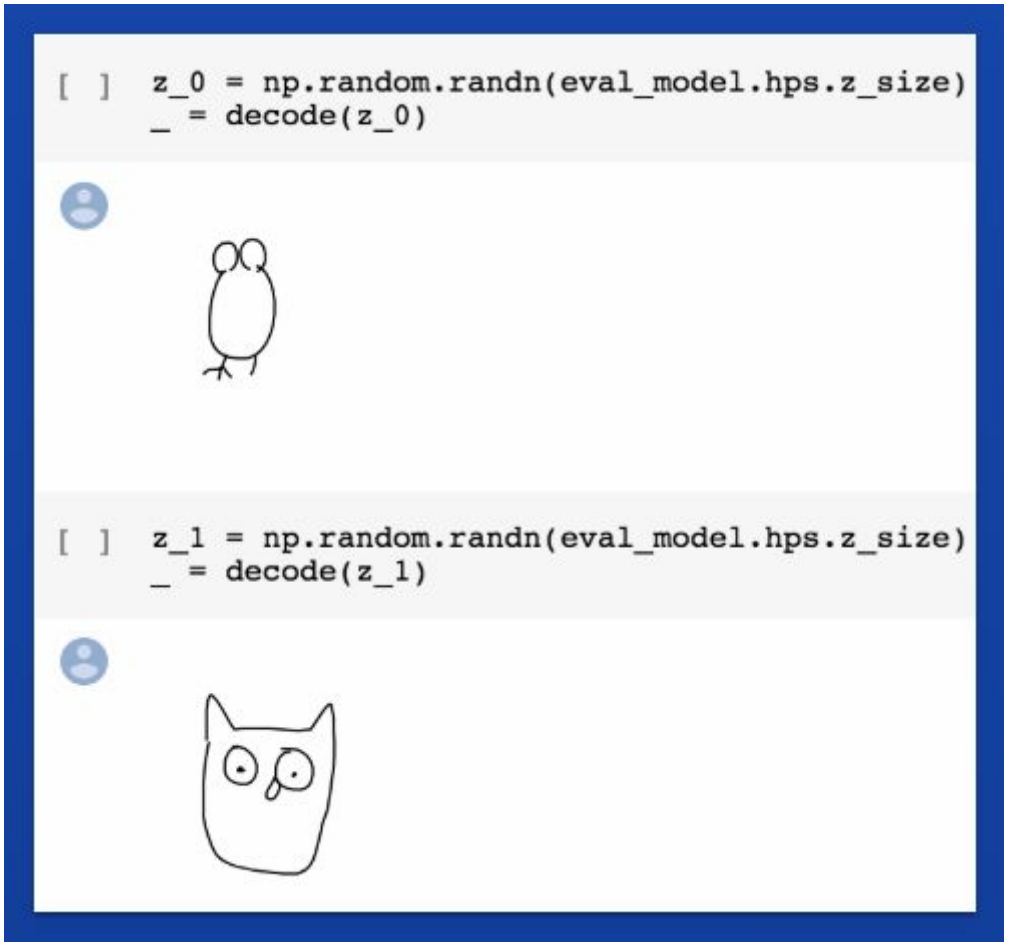


Figure 2: The Kung-Fu Master crouching approach and the functional graph of the player. Outputs which are never activated, and the computational graph leading to them, are omitted for clarity.

Seedbank - Collection of interactive machine learning examples

http://tools.google.com/seedbank/

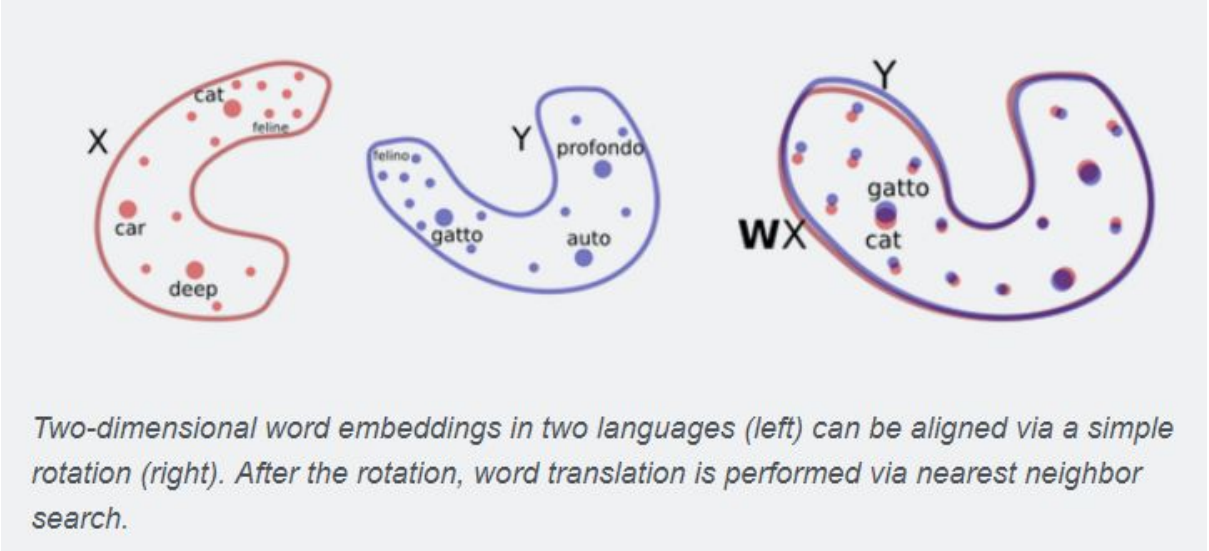
“Seedbank is a registry and search engine for Colab(oratory) notebooks for and around machine learning for rapid exploration and learning. You can browse the site and use, experiment, fork Colab notebooks. The forked notebooks will be stored in your Google Drive and you can share them just like any other Google Docs.”



Unsupervised machine translation: A novel approach to provide fast, accurate translations for more languages

Facebook:
https://code.fb.com/ai-research/unsupervised-machine-translation-a-novel-approach-to-provide-fast-accurate-translations-for-more-languages

“ Research we are presenting at EMNLP 2018 outlines our recent accomplishments with that task. Our new approach provides a dramatic improvement over previous state-of-the-art unsupervised approaches and is equivalent to supervised approaches trained with nearly 100,000 reference translations. To give some idea of the level of advancement, an improvement of 1 BLEU point (a common metric for judging the accuracy of MT) is considered a remarkable achievement in this field; our methods showed an improvement of more than 10 BLEU points.”



Two-dimensional word embeddings in two languages (left) can be aligned via a simple rotation (right). After the rotation, word translation is performed via nearest neighbor search.

Simple guide to Neural Arithmetic Logic Units (NALU): Explanation, Intuition and Code

https://medium.com/mlreview/simple-guide-to-neural-arithmetic-logic-units-nalu-explanation-intuition-and-code-64bc22605712

“The research engineers at DeepMind including well known AI researcher and author of the book Grokking Deep Learning, Andrew Trask have published an impressive paper on a neural network model that can learn simple to complex numerical functions with great extrapolation (generalisation) ability.

In this post I will explain NALU, its architecture, its components and significance over traditional neural networks. The primary intention behind this post is to provide simple and intuitive explanation of NALU (both with concepts and code) which can be comprehended by researchers, engineers and students who have a limited knowledge of neural networks and deep learning.

I believe NALU is a modern breakthrough in AI and specifically in neural networks that seems very promising. They can open doors to many applications which seem to be difficult for standard NNs to deal with.”

Original DeepMind paper: https://arxiv.org/pdf/1808.00508/

Graph networks as learnable physics engines for inference and control

DeepMind - https://arxiv.org/abs/1806.01242

“Flexible, learnable physics models that can predict simulated and real physical systems” - i.e. Emulation not simulation.

“Here we introduce a new class of learnable models - based on graph networks (a class of neural network) - which implement an inductive bias for object- and relation-centric representations of complex, dynamical systems. Our results show that as a forward model, our approach supports accurate predictions from real and simulated data, and surprisingly strong and efficient generalization, across eight distinct physical systems which we varied parametrically and structurally.

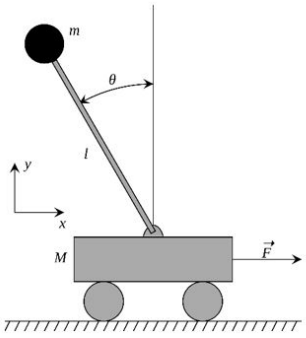
Our models are also differentiable, and support online planning via gradient-based trajectory optimization, as well as offline policy optimization.”

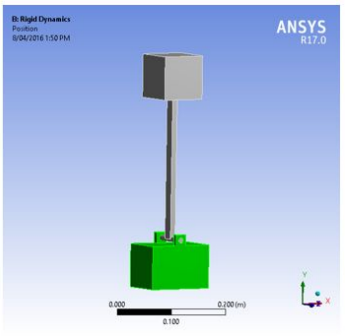
https://m.facebook.com/story.php?story_fbid=429607650887089&id=118896271958230&_rdr - 53:30

https://drive.google.com/file/d/1-R4qC7NHacsDlkjhnwr3w4Gf_89AfCQw/view

Traditional simulation of physical systems

Traditionally physical systems are simulated in a stepwise manner. Given the state of the system, the equations of motion are used to predict the state in the near future:





- State of the system at time t: All the object positions, orientations, velocities, angular velocities
- External controls (forces, motor torques)
- Constraints on the system:
    - Newton-Euler laws of motion (F = ma etc)
    - No penetration at points of contact
    - Hinges maintain velocity & position on the connected objects etc
- Assemble this into a big LCP equation
- Solve to get d(state)/dt
- Integrate to get state(t + dt)

Timestep dt has to be small in order for the system to remain approximately linear (and stable) - e.g. 1ms.

Simulations run at around 1000x realtime for moderately complex scenes (e.g. those involved in robotics).

Off-the-shelf physics engines - e.g. MuJoCo (popular in robotics), PhysX (games) etc.

What’s wrong with traditional simulation?!

- Each step is somewhat “slow”
- It takes a lot of steps to simulate significant amount of time.
- Differentiation (i.e. calculate the change in the simulation for a change in a robot’s controls) is possible - but typically only by making lots of evaluations.

Robotics - Model Predictive Control

- You know the robot’s current state
- You have some goal - perhaps:
    - A target state
    - A trajectory
- Some way to measure how well the system reaches the goal (cost function)
    - e.g. distance of end effector from the target state
    - Sum of the squared deviations of a robot’s parts from an animation
- You want either:
    - A plan of controls that will minimise the cost function (open loop control)
    - A controller that will minimise the cost function (adapts to perturbations/errors)

MPC takes a model of the system and allows you to make that plan/controller.

Sampling based MPC:

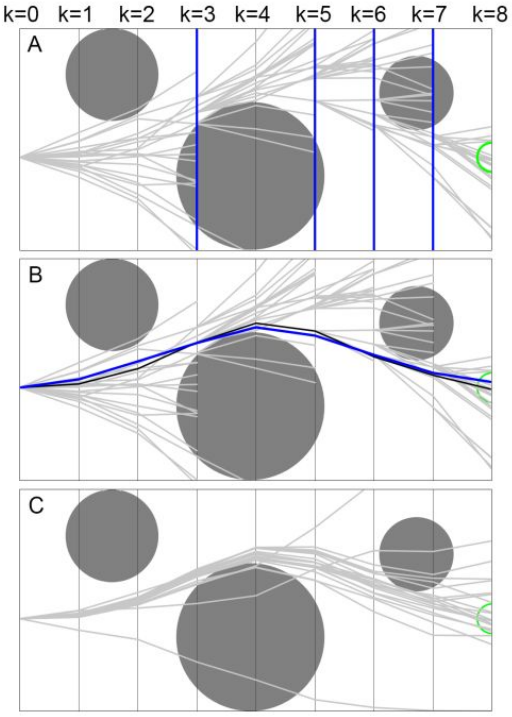
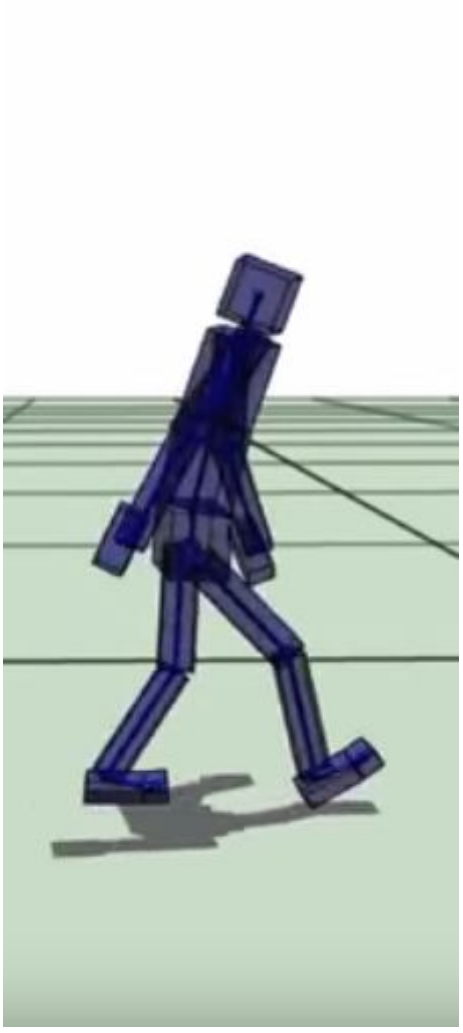


Figure 2: *Our method illustrated in a case where the controlled object moves from left to right at a constant velocity (i.e., k denotes both time and horizontal position), vertical acceleration is controlled, and the green circle denotes the target. Figure A shows the trajectories sampled during the forward pass. The vertical blue lines indicate time instants of resampling. In Figure B the black line indicates the optimal trajectory found in the forward pass and the blue line indicates the results of a backwards local refinement pass. Figure C shows the trajectories sampled by our method in the next iteration, using the previous trajectories as a prior.*



Sequence is:

- Simulate a range of control plans
- Measure how good the results are.
- Pick the best + perturb or somehow come up with refined control plans
- Do it again... and again... and again...

Can “invent” complex solutions to problems - all driven by minimising a cost function: https://youtu.be/dsxbKSG7_bs

Derivative based MPC

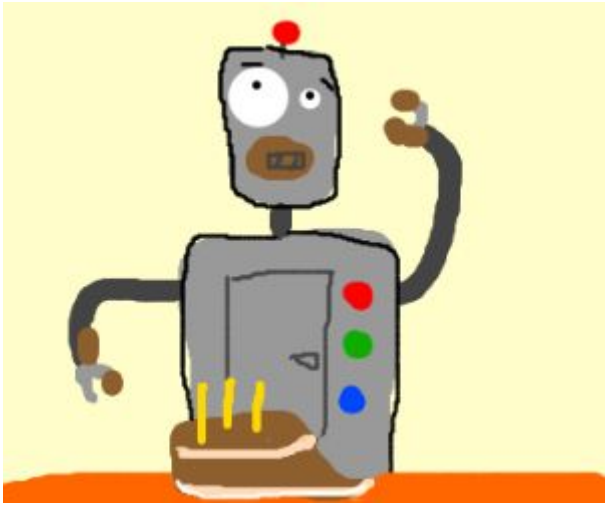
- Start with a trial solution (i.e. control plan)
- Simulate the system...
- Linearise the result
- Calculate response to perturbations (take derivatives)
- Adjust the trial solution so it reduces the cost
- Do it again... and again... and again...

https://homes.cs.washington.edu/~todorov/media/TassaIROS12.mp4

Slow!

We want:

- Big simulation steps (> 1ms)
- Cheap simulation steps (no complex maths)
- Easy gradients



Emulation not simulation

Train a neural network to predict the next state from the current one..

“This work represents the first general purpose, learnable physics engine that can handle complex, 3D physical systems. Unlike classic physics engines, our model has no specific a priori knowledge of physical laws, but instead leverages its object- and relation-centric inductive bias to learn to approximate them via supervised training on current-state/next-state pairs.

The forwards (simulation) model is based on treating physical systems as graphs and learning about them using GNs.

Why better?

- Evaluating each step is quick (forward pass through NN)
- Steps can be much bigger - don’t need so many
- Easily differentiable (good for control)

Training the model (simulation)

Our experiments involved seven actuated MuJoCo simulations environments. We generated training data by applying simulated random controls to the system and recording the state transitions (i.e. how the state changes as time increase by 1ms)

We also trained models from recorded trajectories of a real JACO robotic arm under human control during a stacking task.

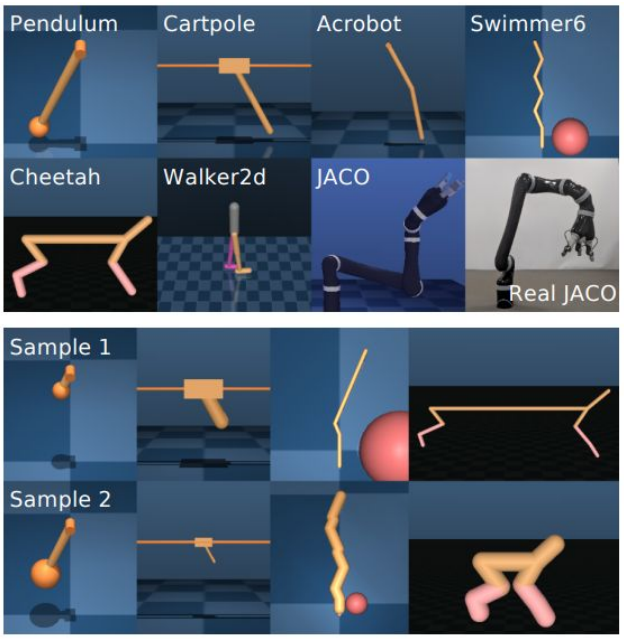
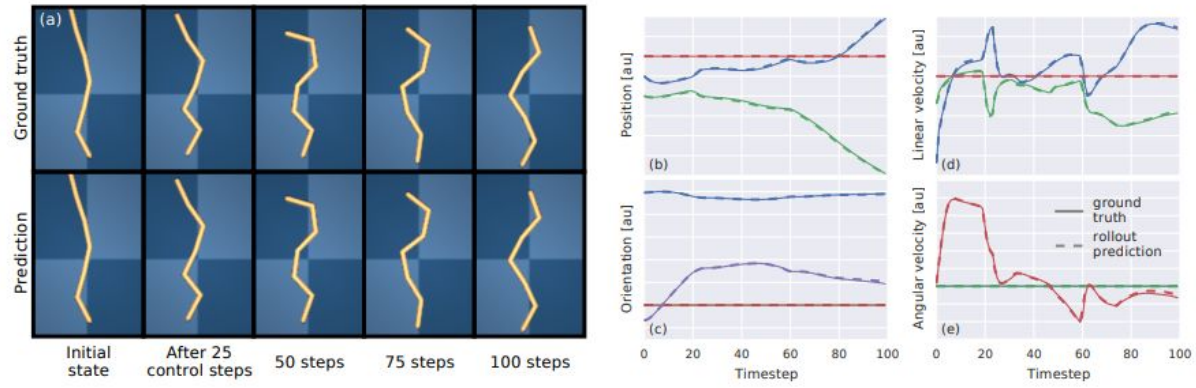


Figure 1. (Top) Our experimental physical systems. (Bottom) Samples of parametrized versions of these systems (see videos: link).

Results: Emulation vs Simulation (dashed line on top of solid!):



Results: Robot control

Imitate full pose:
https://drive.google.com/file/d/1-R4qC7NHacsDlkjhnwr3w4Gf_89AfCQw/view

Imitate hand (has to make up the rest):
https://drive.google.com/file/d/1nbb-yobuX2sFyAWJM_TvOI3usc6DbVVo/view

Creatures:
https://drive.google.com/file/d/1xZme1bxvUWQeb9fWFelECR7YKUXur4XU/view

Hmmmm

But is it new?

1998: Neuroanimator: Fast neural network emulation and control of physics-based models

http://www.cs.toronto.edu/~fritz/absps/siggraph98.pdf

DIdn't NeuroAnimator do exactly this (in a simple way) with similar complexity models 20 years ago?! (just with fewer vowels: Grzeszczuk). It also implemented forward simulation with NNs, and demonstrated control taking advantage of differentiability... but before YouTube.

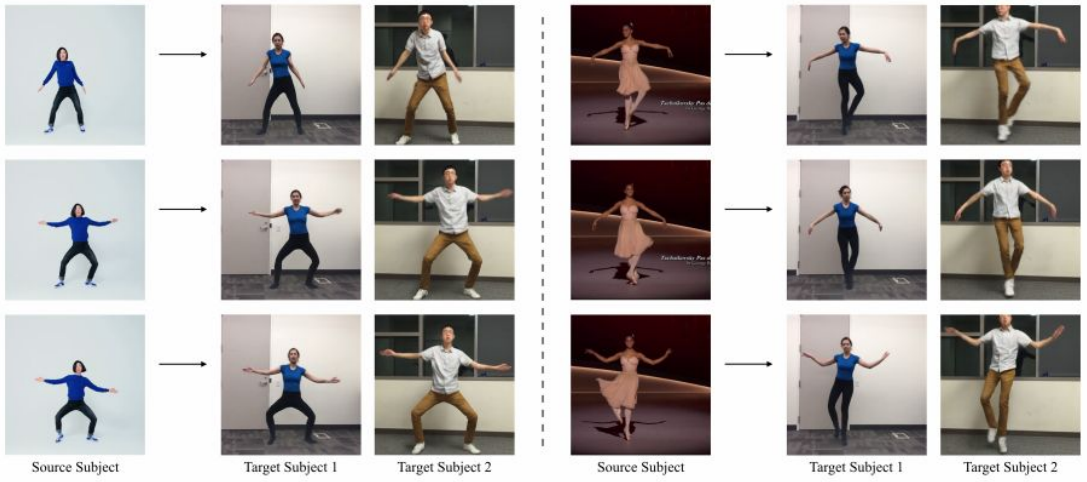
https://www.cs.auckland.ac.nz/~pat/706_99/seminars/reviews/luke.d/seminar.lhut001/sig98avi/

Everybody Dance Now

https://carolineec.github.io/everybody_dance_now/

Style transfer & pose estimation.

“This paper presents a simple method for "do as I do" motion transfer: given a source video of a person dancing we can transfer that performance to a novel (amateur) target after only a few minutes of the target subject performing standard moves. We pose this problem as a per-frame image-to-image translation with spatio-temporal smoothing. Using pose detections as an intermediate representation between source and target, we learn a mapping from pose images to a target subject's appearance.”



“they first used an algorithm to detect poses from a source video of a dancer—basically creating a moving stick figure. They also took video of the target person performing a range of motion. Then, they trained two deep learning algorithms known as Generative Adversarial Networks (GAN) to produce a whole image and to produce a sharper and more realistic face for the amateur subject. The end result is a system that generates video and convincingly maps the bodily motion of an expert dancer to a total amateur.”

2-minute papers:
https://youtu.be/cEBgi6QYDhQ

Detail

- Image-to-image transfer, from a source to target subject, via an end-to-end pixel-based pipeline.
- Want to learn a mapping between images of two individuals...
- ... but don’t have corresponding pairs of images in the same pose.

How?!

- Via intermediate representation - a keypoint based pose - i.e. stick figures.
- Condition the prediction on the previous timestep to maintain temporal smoothness
- Specialised GAN for the face.

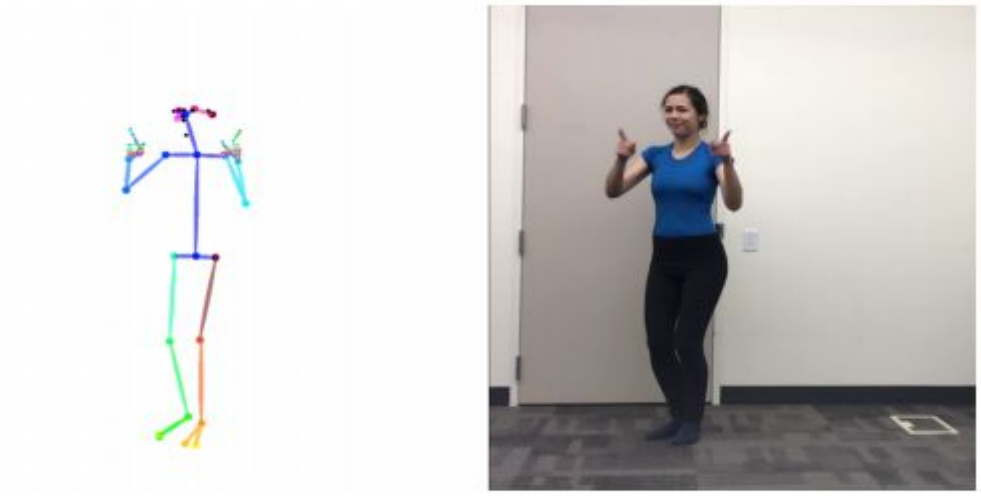
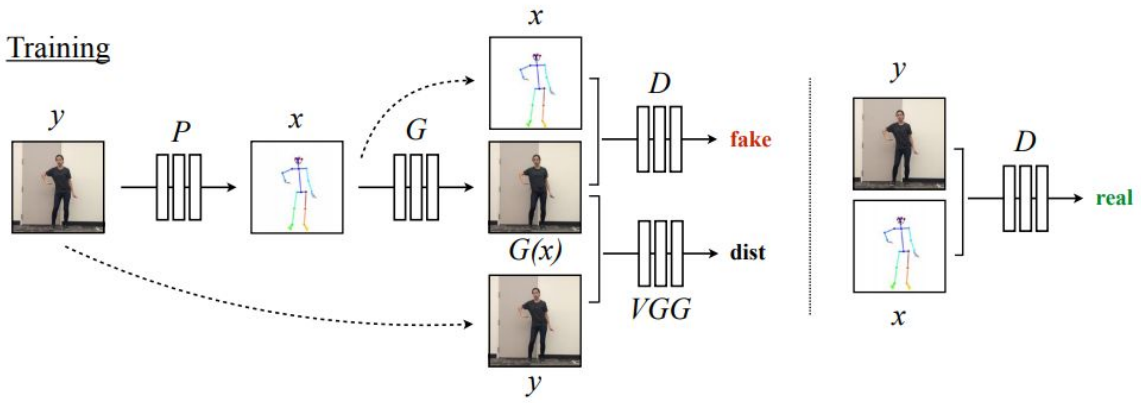


Fig. 2. Correspondence between pose stick figure and target person frame.

Training pipeline:

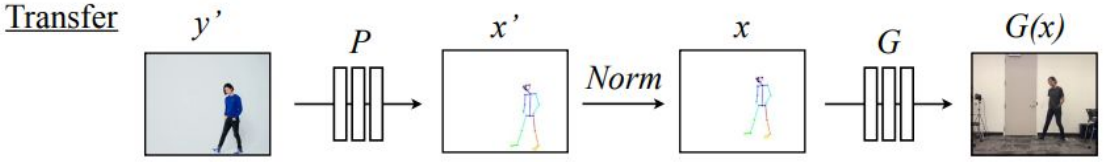
Adversarial training:

- Use **pre-trained** network to generate pose x from y
- Learn a mapping G which creates images from x
- by training it with a discriminator D that attempts to distinguish between G(x) and y (i.e. the synthesised image and the original).



Transfer Pipeline:

Pose detection (stick figure) from source -> pose normalisation -> convert pose to target



Results:

https://youtu.be/PCBTZh41Ris