令和2年度 第2回 人間中心のAI社会原則会議　2021/3/24

AI研究の現状について

原田 達也

東京大学 先端科学技術研究センター/理研AIP

Impact of AI in Academic

- Google Scholar Metrics
  - CVPR (Computer Vision and Pattern Recognition): 5位
  - NeurIPS (Neural Information Processing Systems): 21位
  - ICCV (International Conference on Computer Vision): 29位
  - ICML (International Conference on Machine Learning): 33位

https://scholar.google.com/citations?view_op=top_venues&hl=en

Google Scholar

Top publications

Categories

English

https://www.nature.com/

https://www.sciencemag.org/

	Publication	h5-index	h5-median
1.	Nature	376	552
2.	The New England Journal of Medicine	365	639
3.	Science	356	526
4.	The Lancet	301	493
5.	IEEE/CVF Conference on Computer Vision and Pattern Recognition	299	509
6.	Advanced Materials	273	369
7.	Nature Communications	273	366
8.	Cell	269	417
9.	Chemical Reviews	267	438
10.	Chemical Society reviews	240	368

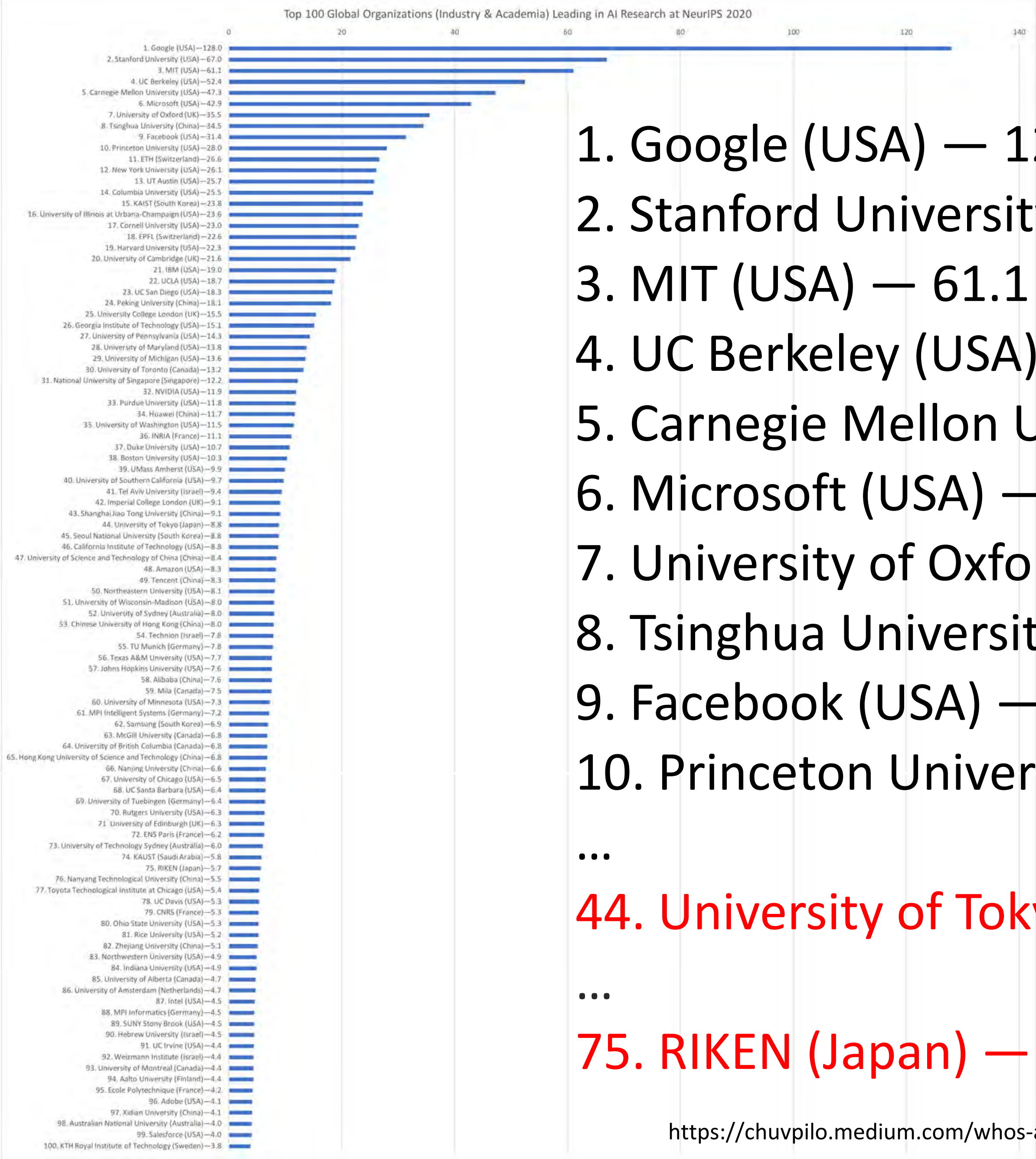
Most Influential Papers for 2020

- The most highly-cited paper of all
  - "Deep Residual Learning for Image Recognition", CVPR2016
  - From 25,256 citations in 2019 to 49,301 citations in 2020.
- "Deep learning", Nature in 2015
  - From 16,750 in 2019 to 27,375 in 2020

1. "Adam: A Method for Stochastic Optimization" (2015), International Conference on Learning Representations
2. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks" (2015), Neural Information Processing Systems
3. "Human-level control through deep reinforcement learning" (2015), Nature
4. "Attention Is All You Need" (2017), Neural Information Processing Systems
5. "The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)" (2016), JAMA
6. "limma powers differential expression analyses for RNA-sequencing and microarray studies" (2015), Nucleic Acids Research
7. "Mastering the game of Go with deep neural networks and tree search" (2016), Nature

https://www.natureindex.com/news-blog/google-scholar-reveals-most-influential-papers-research-citations-twenty-twenty#:~:text=The%20most%20highly%2Dcited%20paper,to%2049%2C301%20citations%20in%202020.

Who’s Ahead in AI Research at NeurIPS 2020?



1. Google (USA) — 128.0
2. Stanford University (USA) — 67.0
3. MIT (USA) — 61.1
4. UC Berkeley (USA) — 52.4
5. Carnegie Mellon University (USA) — 47.3
6. Microsoft (USA) — 42.9
7. University of Oxford (UK) — 35.5
8. Tsinghua University (China) — 34.5
9. Facebook (USA) — 31.4
10. Princeton University (USA) — 28.0

…

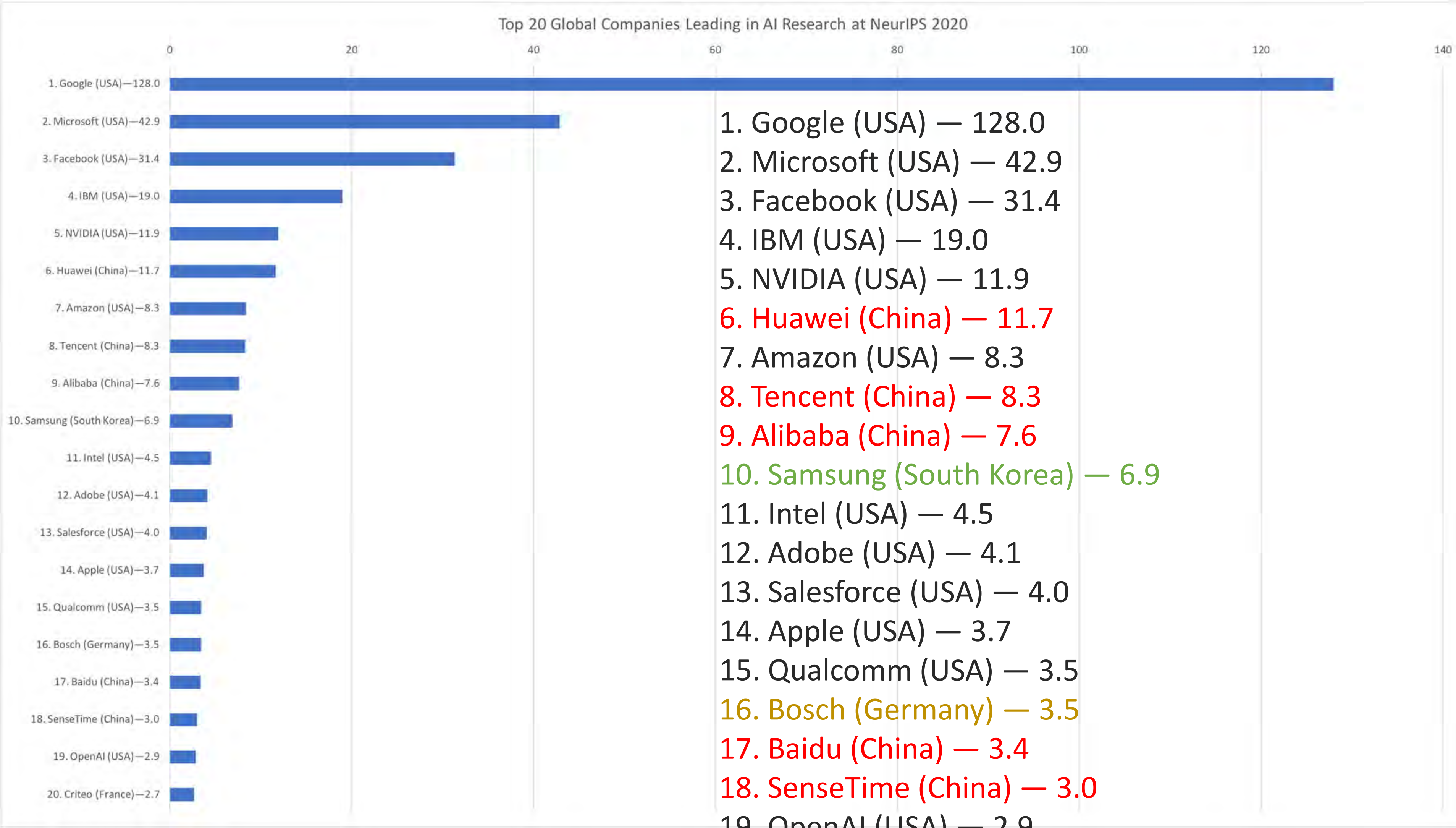
44. University of Tokyo (Japan) — 8.8

…

75. RIKEN (Japan) — 5.7

https://chuvpilo.medium.com/whos-ahead-in-ai-research-at-neurips-2020-bf2a40a54325

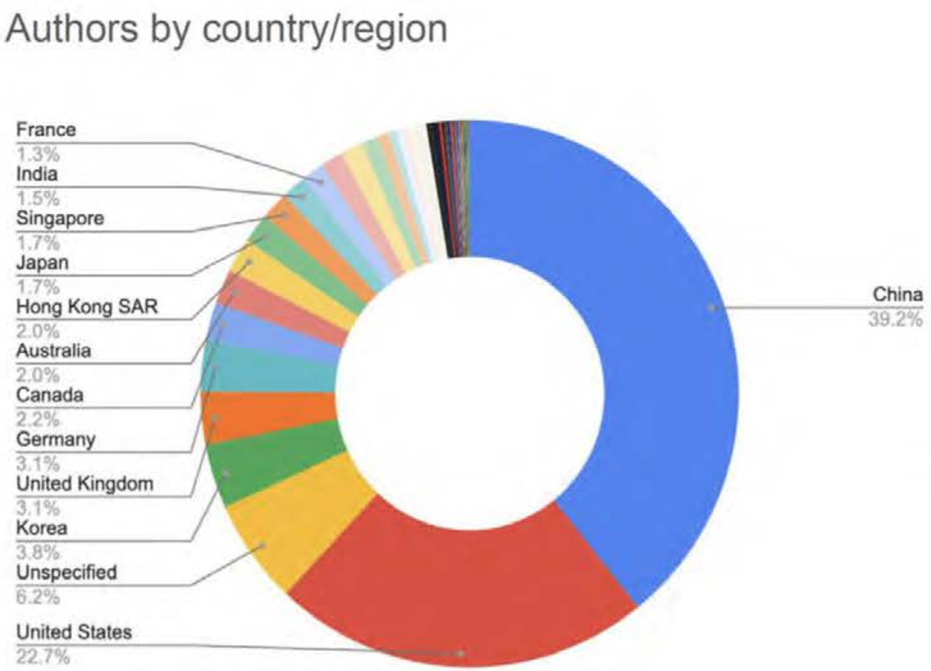
Top 20 Global Companies Leading in AI Research at NeurIPS 2020

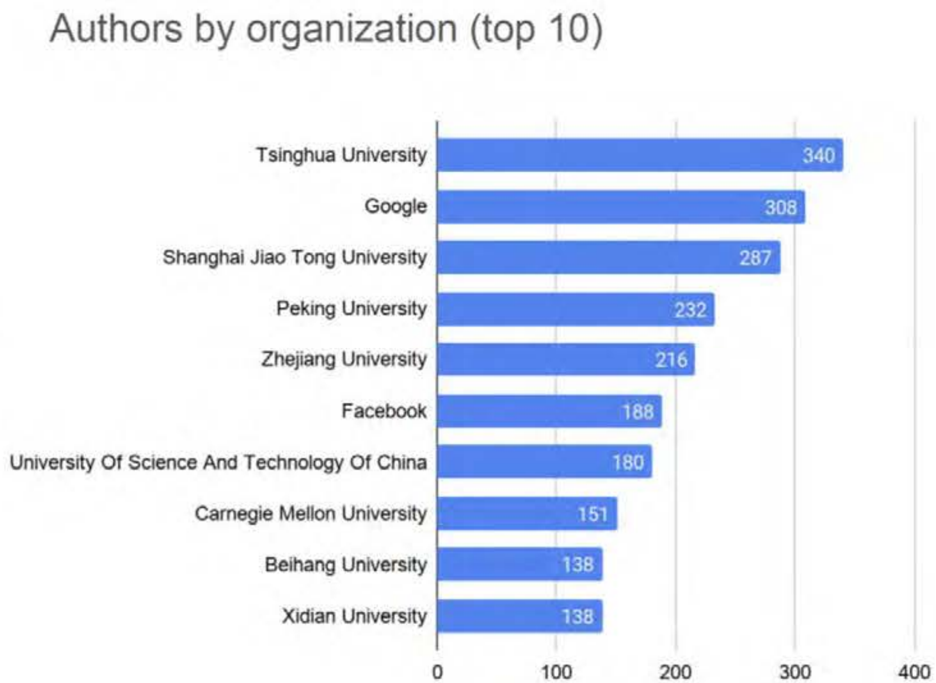


1. Google (USA) — 128.0
2. Microsoft (USA) — 42.9
3. Facebook (USA) — 31.4
4. IBM (USA) — 19.0
5. NVIDIA (USA) — 11.9
6. Huawei (China) — 11.7
7. Amazon (USA) — 8.3
8. Tencent (China) — 8.3
9. Alibaba (China) — 7.6
10. Samsung (South Korea) — 6.9
11. Intel (USA) — 4.5
12. Adobe (USA) — 4.1
13. Salesforce (USA) — 4.0
14. Apple (USA) — 3.7
15. Qualcomm (USA) — 3.5
16. Bosch (Germany) — 3.5
17. Baidu (China) — 3.4
18. SenseTime (China) — 3.0
19. OpenAI (USA) — 2.9
20. Criteo (France) — 2.7

https://chuvpilo.medium.com/whos-ahead-in-ai-research-at-neurips-2020-bf2a40a54325

Author Distribution in CVPR2020





http://cvpr2020.thecvf.com/sites/default/files/CVPR2020_opening.pdf

- 1. China: 39.2%
- 2. U.S.A.: 22.7%
- 3. unspecified: 6.2%
- 4. Korea: 3.8%
- 5. United Kingdom: 3.1%
- …
- **10. Japan: 1.7%**

- 1. 精華大学: 39.2%
- 2. Google: 22.7%
- 3. 上海交通大学: 6.2%
- 4. 北京大学: 3.8%
- 5. 浙江大学: 3.1%
- 6. Facebook
- 7. 中国科学技術大学

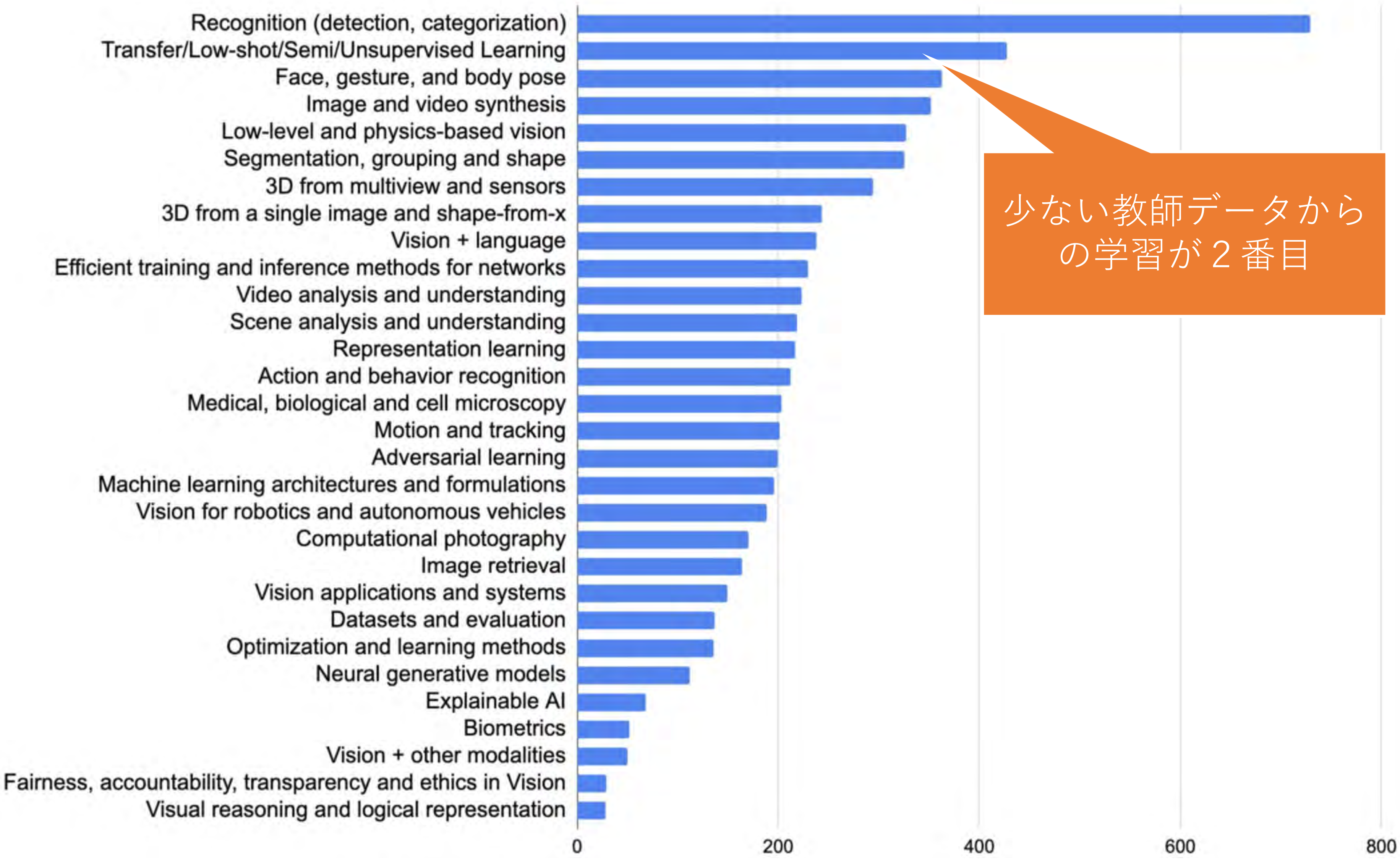
Subject Areas in CVPR2020

Main subject areas
Machine learning architectures and formulations
Explainable AI
Efficient training & inference methods
Generative models
Adversarial learning
Transfer/Low-shot/Semi/Unsupervised Learning
Recognition (detection, categorization)
Face, gesture, and body pose
Image and video synthesis
Segmentation, grouping & shape
Vision + language
3D from multiview and sensors
Low-level and physics-based vision
3D from single image and shape-from-x

Vision applications and systems
Datasets & evaluation
Optimization and learning methods
Video analysis and understanding
Biometrics
Vision for robotics or autonomous vehicles
Action recognition
Medical, biological and cell microscopy
Visual reasoning and logical representation
Image retrieval
Fairness, Accountability, Transparency and Ethics in Vision
Vision + other modalities
Motion & tracking
Representation learning
Computational photography
Scene analysis and understanding

http://cvpr2020.thecvf.com/sites/default/files/CVPR2020_opening.pdf

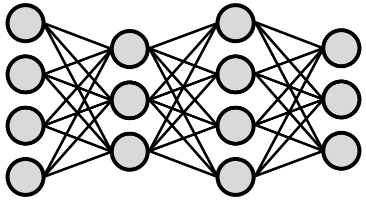
Distribution of Subject Areas in CVPR2020



http://cvpr2020.thecvf.com/sites/default/files/CVPR2020_opening.pdf

Big Wave of Artificial Intelligence

- ☐Deep Neural Networks



- ☐Large Amount of High-Quality Training Data

IMAGENET

http://www.image-net.org/

- ☐Large Amount of Computational Resource

https://blogs.nvidia.co.jp/

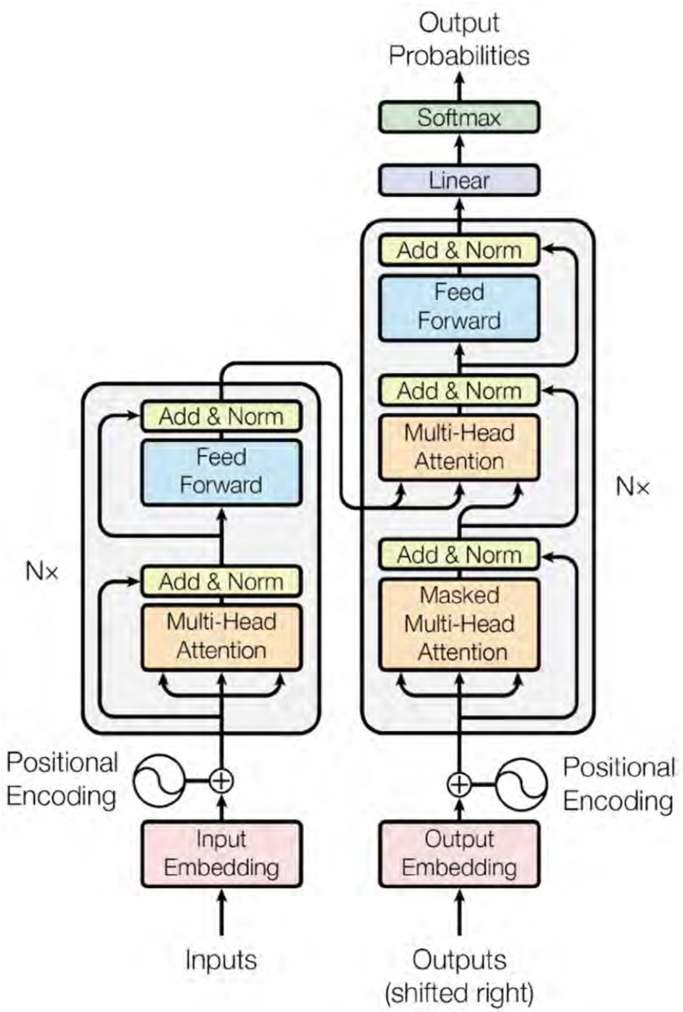
Google

Microsoft

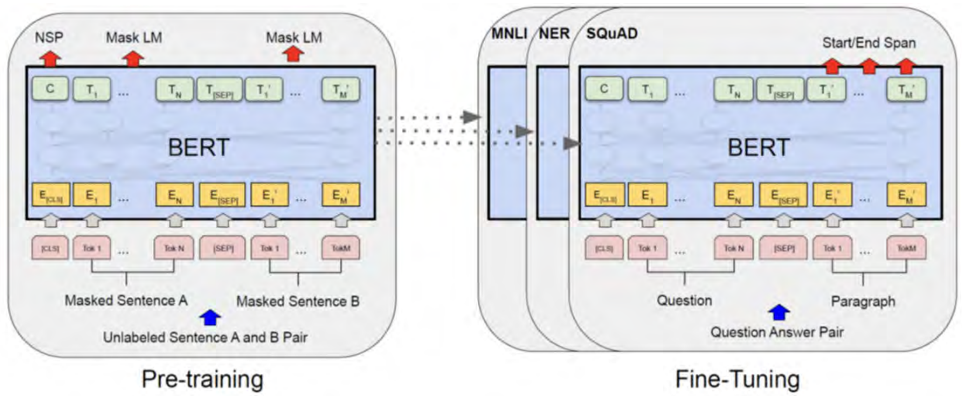
amazon.com

Big Leap in NLP!

- Transformer
  - Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin. Attention is All you Need. NIPS2017.



- GPT (OpenAI)
  - Alec Radford, Karthik Narasimhan, Time Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. Tech. report, OpenAI, 2018.
- BERT (Google)
  - Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT 2019.
  - BERT is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers.
  - The pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks.
  - **BERT is used in the search engine of Google.**



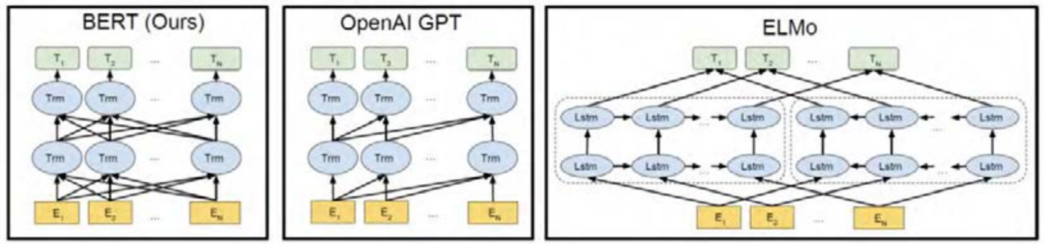


Figure 3: Differences in pre-training model architectures. BERT uses a bidirectional Transformer. OpenAI GPT uses a left-to-right Transformer. ELMo uses the concatenation of independently trained left-to-right and right-to-left LSTMs to generate features for downstream tasks. Among the three, only BERT representations are jointly conditioned on both left and right context in all layers. In addition to the architecture differences, BERT and OpenAI GPT are fine-tuning approaches, while ELMo is a feature-based approach.

GPT-3 (OpenAI)

- Tom B Brown et al. Language Models are Few-Shot Learners. NeurIPS2020. Best paper award
- Model: 175B parameters
- Dataset: 500B tokens

Manuel Araoz's Personal Website

Bio I studied Computer Science and Engineering at Instituto Tecnológico de Buenos Aires. I'm located in Buenos Aires, Argentina. My previous work is mostly about cryptocurrencies, distributed systems, machine learning, interactivity, and robotics. One of my goals is to bring new experiences to people through technology.

I cofounded and was formerly CTO at OpenZeppelin. Currently, I'm studying music, biology+neuroscience, machine learning, and physics.

Blog

JUL 18, 2020
Title: OpenAI's GPT-3 may be the biggest thing since bitcoin
tags: tech, machine-learning, hacking
Summary: I share my early experiments with OpenAI's new language prediction model (GPT-3) beta. I explain why I think GPT-3 has disruptive potential comparable to that of blockchain technology.
Full text:

GPT-3の学習には13億円必要？

Elliot Turner (Alchemy)

Elliot Turner @eturner303 · 2020年5月29日
Reading the OpenAI GPT-3 paper. Impressive performance on many few-shot language tasks. The cost to train this 175 billion parameter language model appears to be staggering: Nearly $12 million dollars in compute based on public cloud GPU/TPU cost models (200x the price of GPT-2)

OpenAI's GPT-3 may be the biggest thing since bitcoin

JUL 18, 2020

Summary: I share my early experiments with OpenAI's new language prediction model (GPT-3) beta. I explain why I think GPT-3 has disruptive potential comparable to that of blockchain technology.

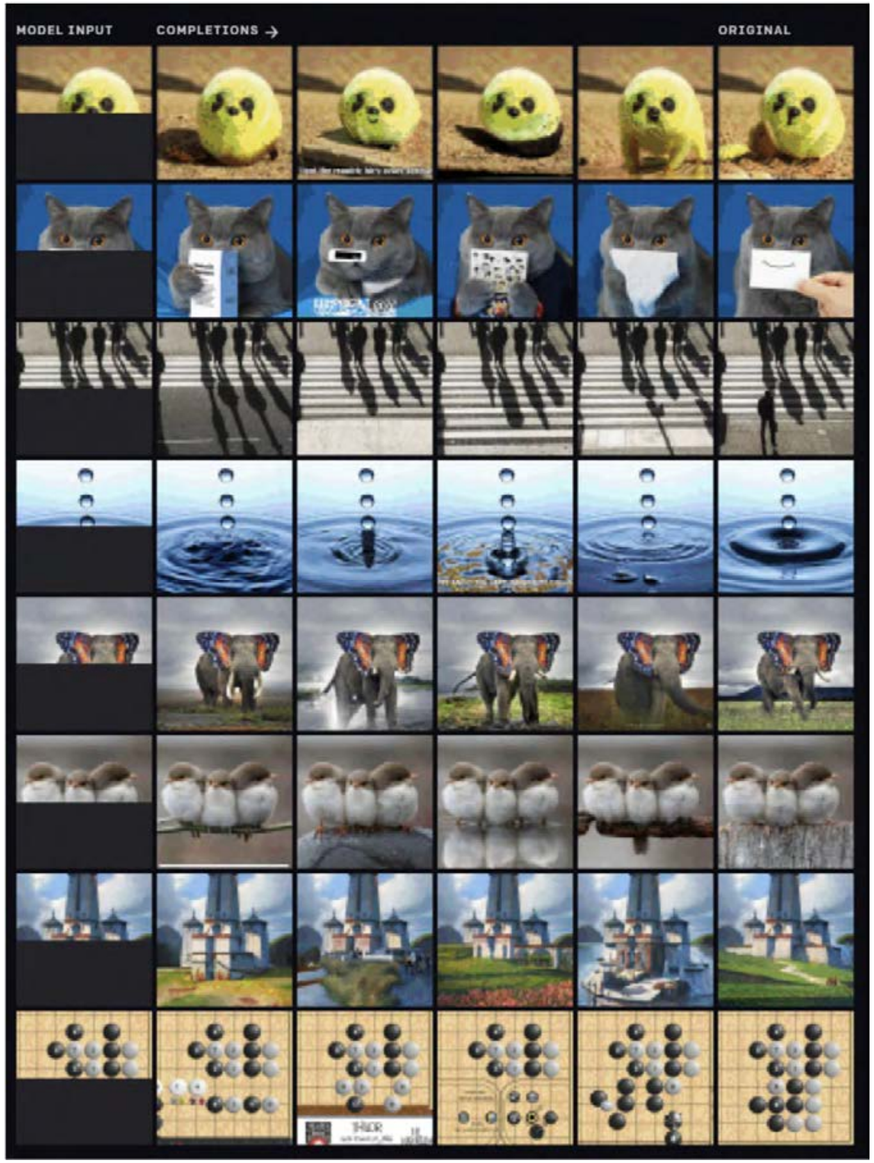
OpenAI, a non-profit artificial intelligence research company backed by Peter Thiel, Elon Musk, Reid Hoffman, Marc Benioff, Sam Altman and others, released its third generation of language prediction model (GPT-3) into the open-source wild. Language models allow computers to produce random-ish sentences of approximately the same length and grammatical structure as those in a given body of text.

In my early experiments with GPT-3 I found that GPT-3's predicted sentences, when published on the bitcointalk.org forum, attracted lots of positive attention from posters there, including suggestions that the system must have been intelligent (and/or sarcastic) and that it had found subtle patterns in their posts. I imagine that similar results can be obtained by republishing GPT-3's outputs to other message boards, blogs, and social media.

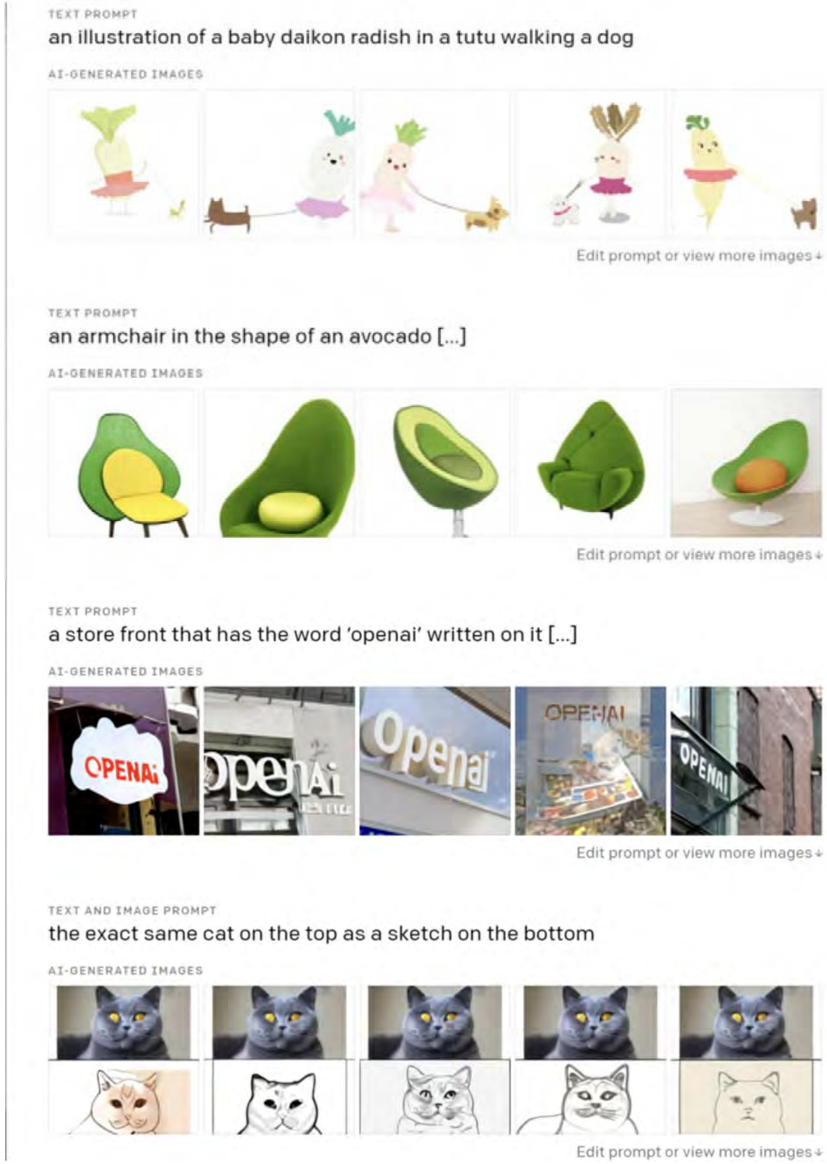
I predict that, unlike its two predecessors (PTB and OpenAI GPT-2), OpenAI GPT-3 will eventually be widely used to pretend the author of a text is a person of interest, with unpredictable and amusing effects on various communities. I further predict that this will spark a creative gold rush among talented amateurs to train similar models and adapt them to a variety of purposes, including: mock news, "researched journalism", advertising, politics, and propaganda.

GPT-3 for Visual Tasks (OpenAI)

- Image GPT (OpenAI)
  - Mark Chen, Alec Radford, Rewon Child, Jeff Wu, Heewoo Jun, David Luan, Ilya Sutskever. Generative Pretraining from Pixels. ICML2020

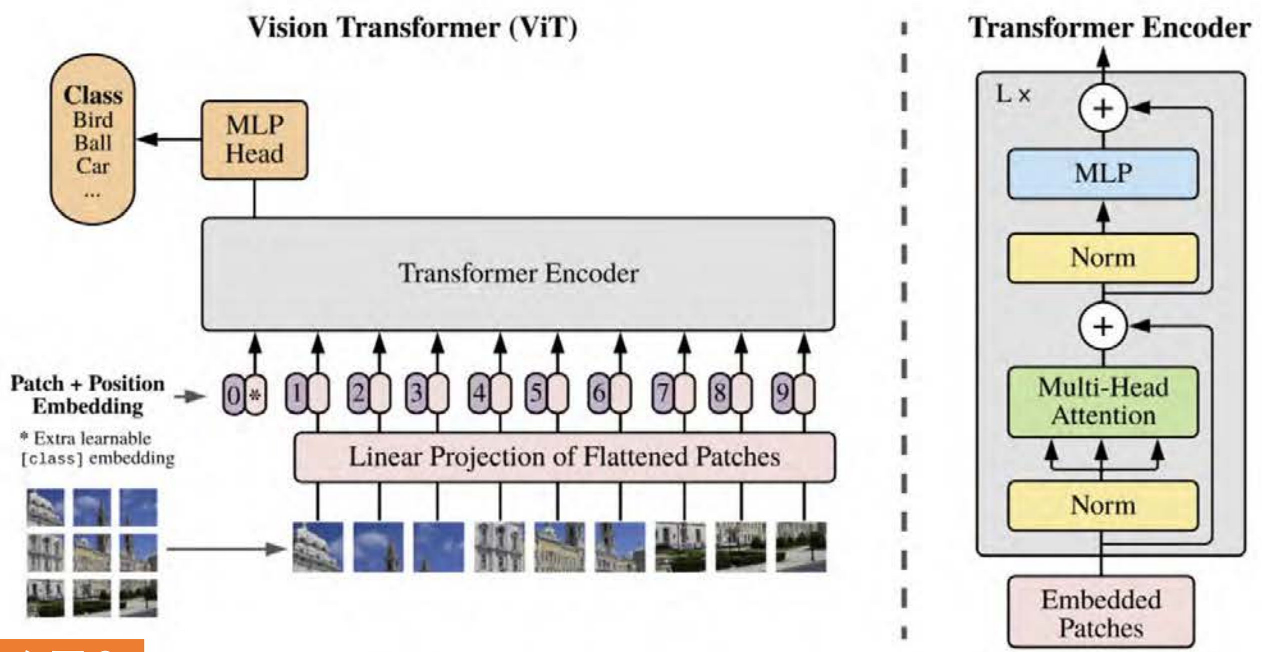


- DALL·E (OpenAI)
  - Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, Ilya Sutskever. Zero-Shot Text-to-Image Generation. arXiv:2102.12092



Vision Transformer (Google)

- Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. ICLR2021
- This reliance on CNNs is not necessary and a pure transformer applied directly to sequences of image patches can perform very well on image classification tasks.
- Dataset: JFT with 18k classes and 303M high-resolution images



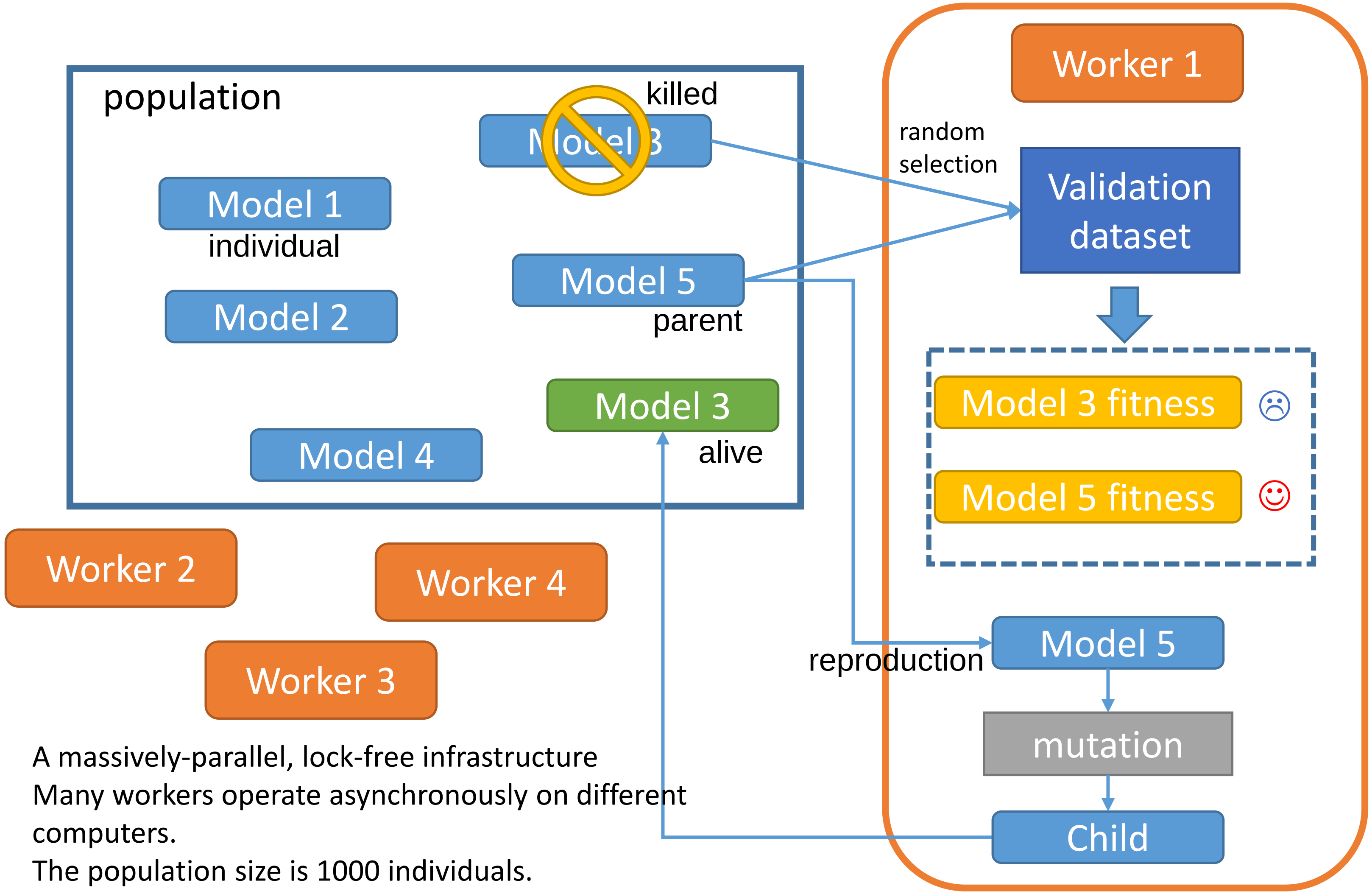
学習には600万円必要？

	Ours-JFT (ViT-H/14)	Ours-JFT (ViT-L/16)	Ours-I21k (ViT-L/16)	BiT-L (ResNet152x4)	Noisy Student (EfficientNet-L2)
ImageNet	**88.55** ± 0.04	87.76 ± 0.03	85.30 ± 0.02	87.54 ± 0.02	88.4/88.5*
ImageNet ReaL	**90.72** ± 0.05	90.54 ± 0.03	88.62 ± 0.05	90.54	90.55
CIFAR-10	**99.50** ± 0.06	99.42 ± 0.03	99.15 ± 0.03	99.37 ± 0.06	—
CIFAR-100	**94.55** ± 0.04	93.90 ± 0.05	93.25 ± 0.05	93.51 ± 0.08	—
Oxford-IIIT Pets	**97.56** ± 0.03	97.32 ± 0.11	94.67 ± 0.15	96.62 ± 0.23	
Oxford Flowers-102	99.68 ± 0.02	**99.74** ± 0.00	99.61 ± 0.02	99.63 ± 0.03	
VTAB (19 tasks)	**77.63** ± 0.23	76.28 ± 0.46	72.72 ± 0.21	76.29 ± 1.70	
TPUv3-core-days	2.5k	0.68k	0.23k	9.9k	12.3k

TPU タイプ (v3)	v3 コア	合計メモリ	オンデマンド料金 (米ドル)	プリエンプティブル料金 (米ドル)
v3-8	8	128 GiB	$8.00/時間	$2.40/時間

Network Architecture Search (NAS)

Esteban Real, Sherry Moore, Andrew Selle, Saurabh Saxena, Yutaka Leon Suematsu, Jie Tan, Quoc Le, Alex Kurakin. Large-Scale Evolution of Image Classifiers. ICML, 2017.



- A massively-parallel, lock-free infrastructure
- Many workers operate asynchronously on different computers.
- The population size is 1000 individuals.
- The number of workers is always 1/4 of the population size.

Experimental Results

Esteban Real, Sherry Moore, Andrew Selle, Saurabh Saxena, Yutaka Leon Suematsu, Jie Tan, Quoc Le, Alex Kurakin. Large-Scale Evolution of Image Classifiers. ICML, 2017.

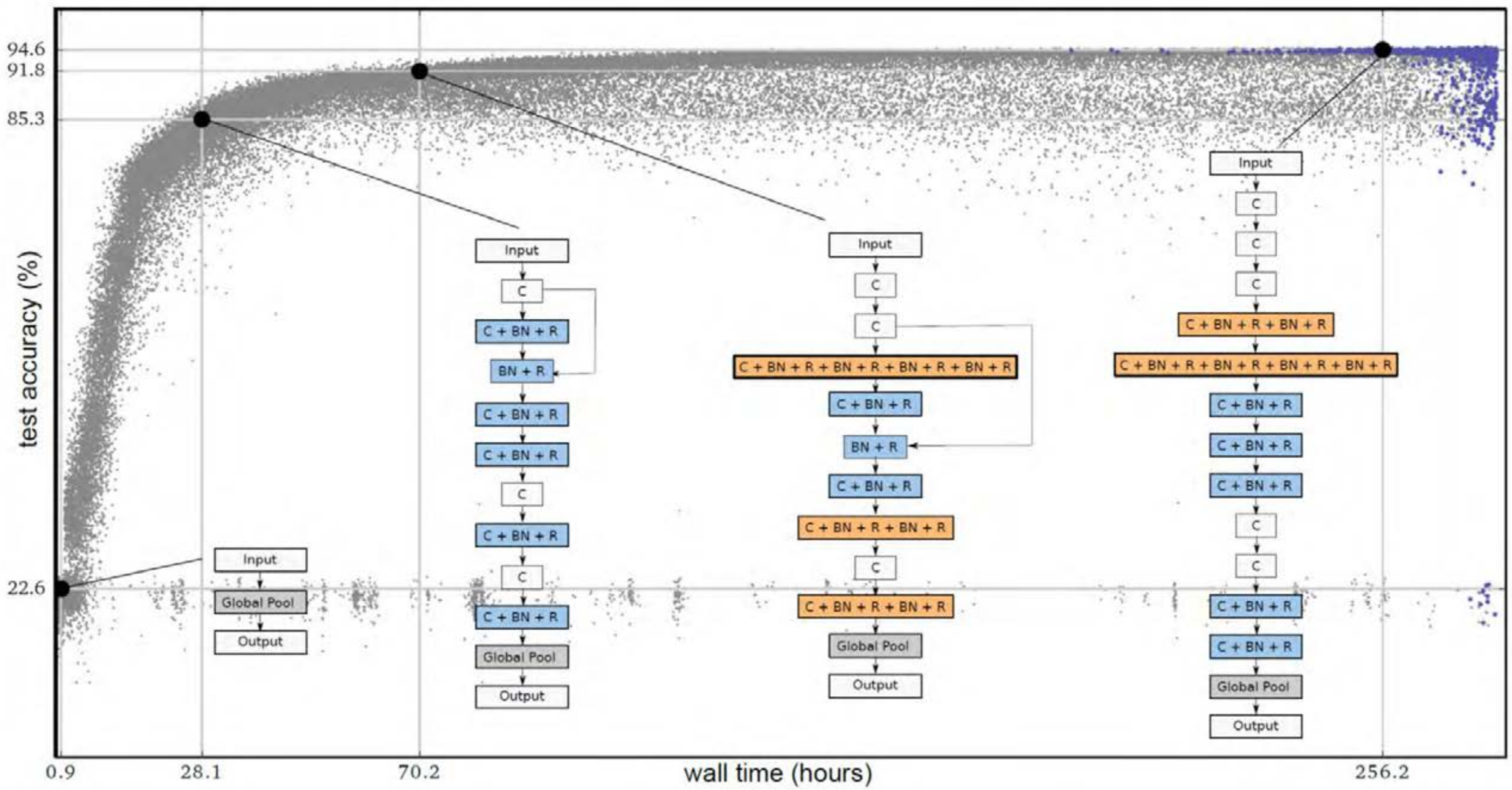
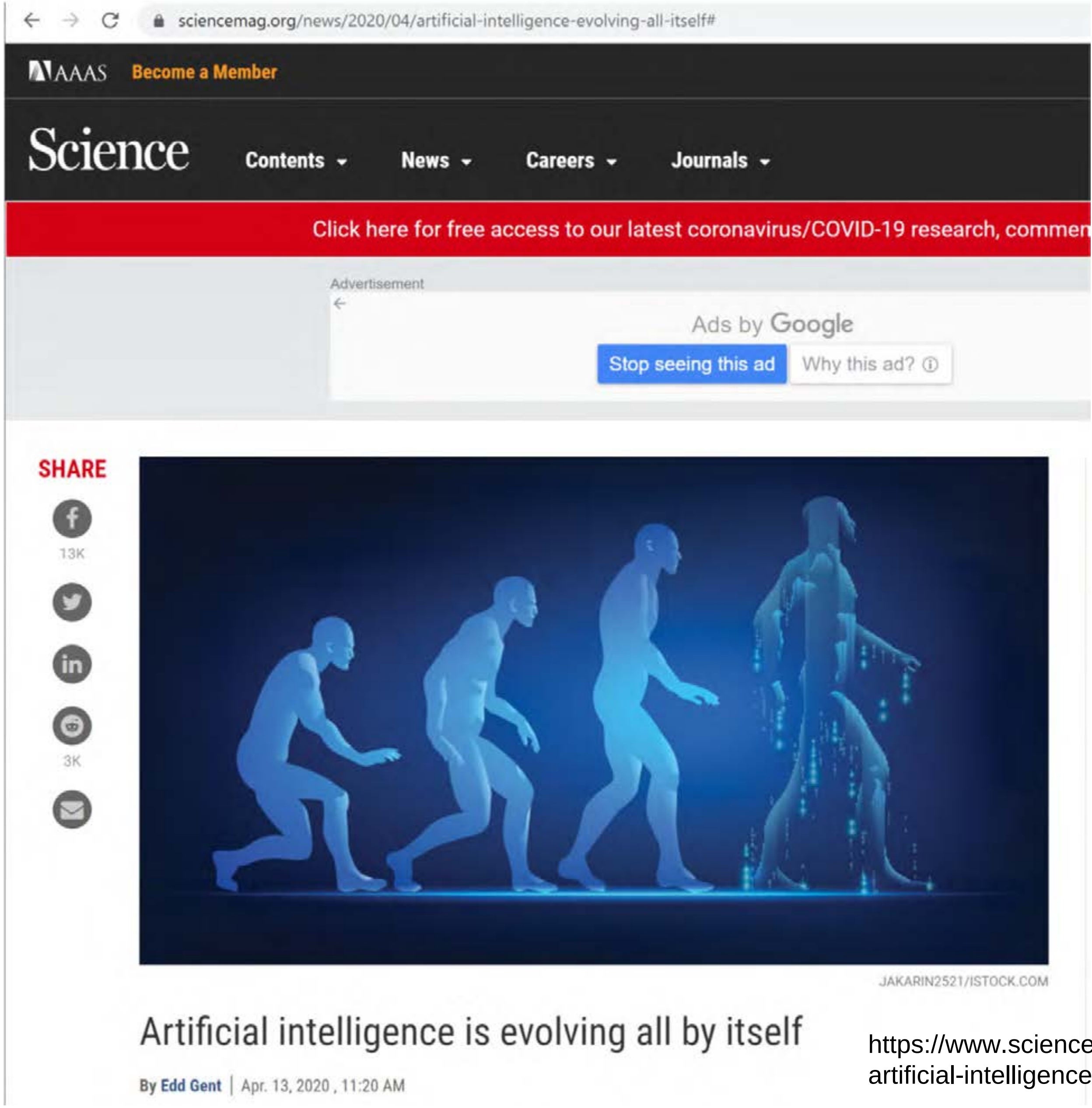


Figure 1. Progress of an evolution experiment. Each dot represents an individual in the population. Blue dots (darker, top-right) are alive. The rest have been killed. The four diagrams show examples of discovered architectures. These correspond to the best individual (right-most) and three of its ancestors. The best individual was selected by its validation accuracy. Evolution sometimes stacks convolutions without any nonlinearity in between ("C", white background), which are mathematically equivalent to a single linear operation. Unlike typical hand-designed architectures, some convolutions are followed by more than one nonlinear function ("$C+BN+R+BN+R+...$", orange background).

AutoML-Zero: Evolving Machine Learning Algorithms From Scratch

ICML2020



https://www.sciencemag.org/news/2020/04/artificial-intelligence-evolving-all-itself

Results

Esteban Real, Chen Liang, David R. So, Quoc V. Le. AutoML-Zero: Evolving Machine Learning Algorithms From Scratch. ICML2020

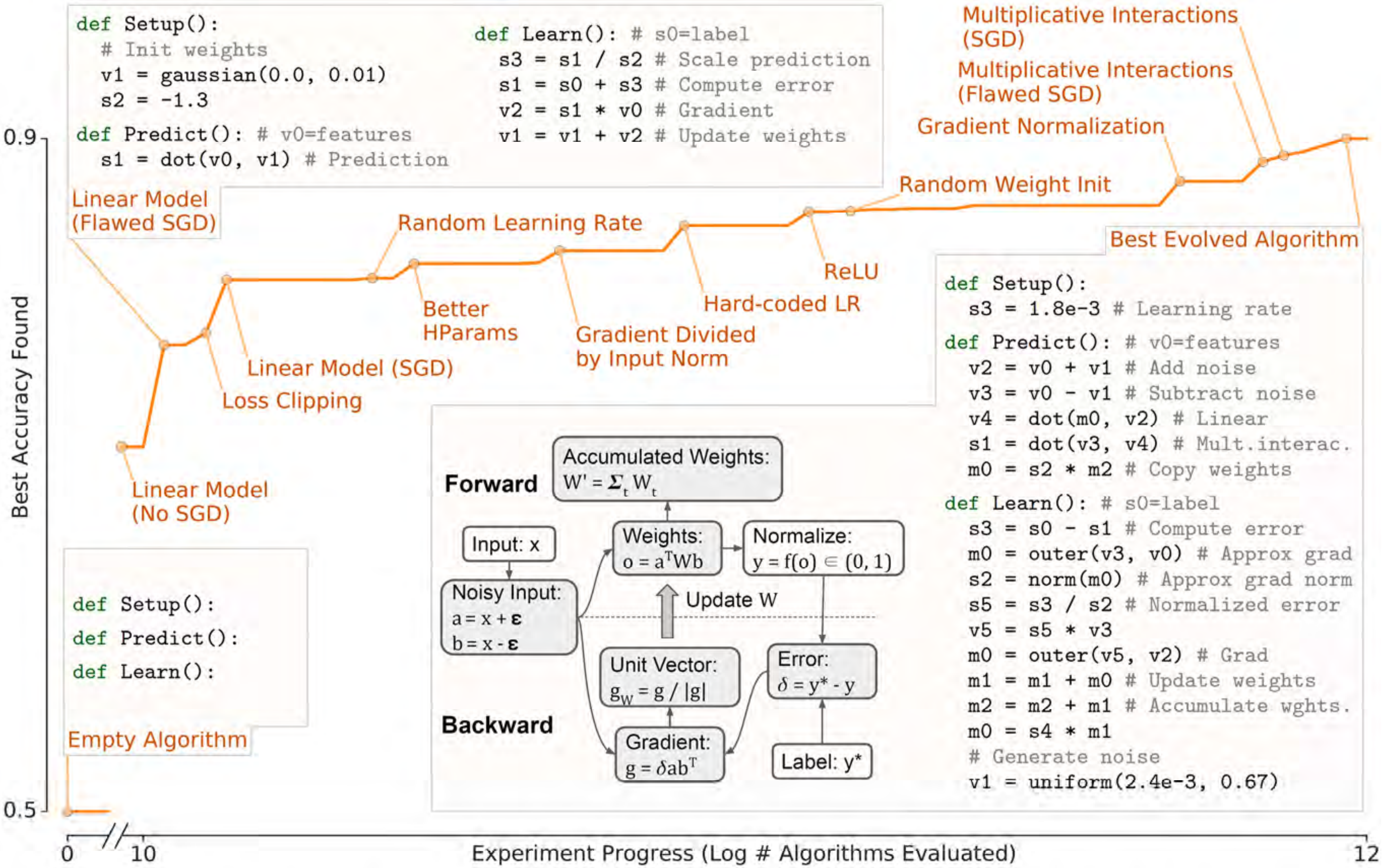
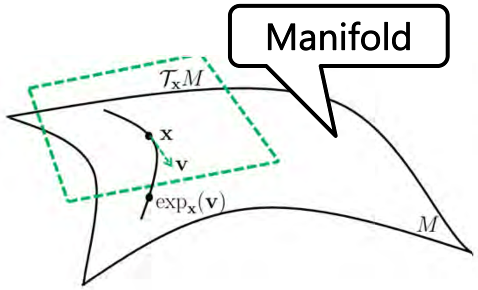


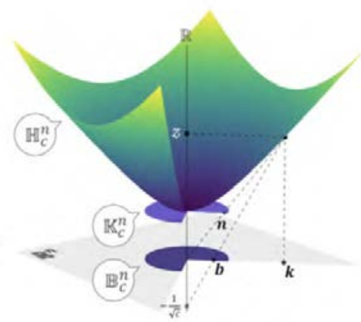
Figure 6: Progress of one evolution experiment on projected binary CIFAR-10. Callouts indicate some beneficial discoveries. We also print the code for the initial, an intermediate, and the final algorithm. The last is explained through the flow diagram. It outperforms a simple fully connected neural network on held-out test data and transfers to features 10x its size. Code notation is the same as in Figure 5.

Hyperbolic Neural Networks++ Shimizu, Mukuta, Harada. ICLR2021

- **DNNs in the hyperbolic geometry**
  - Hyperbolic space has the **exponential growing volume** and is one of the promising manifold to better process real-world **hierarchical data structure**
  - Hyperbolic version of SoftMax, FC, convolutional, and attention layers
- **Benefit of Hyperbolic Geometry**
  - Efficient for embedding **tree structures**



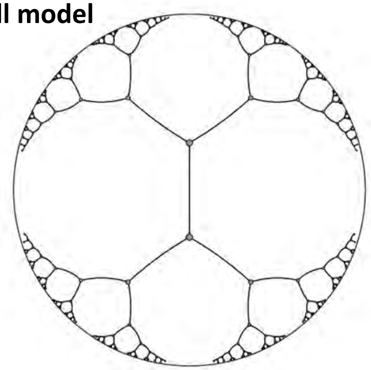
Low-Dimensional Hyperbolic Knowledge Graph Embeddings, 2020, Chami et al.



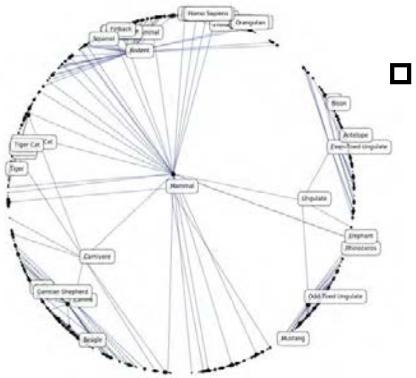
Constant Negative ($K < 0$)
Hyperbolic

2-D Poincaré ball model

M.C. Escher, Circle Limit III, 1959



Embedding of a tree in Poincare disk.



- Poincaré Embeddings for Learning Hierarchical Representations [Nickel et al., NIPS 2017]
  - Word2Vec derivative in the Poincaré ball model
  - Higher performance than Euclidean counterparts with fewer dimensions

	$\mathbb{H}_c^n$	$\mathbb{B}_c^n$
Parallel transport		✔
Vector addition		✔
Scalar multiplication		✔
Weighted centroid	✔	★
Function apply		✔
Fully-connected layer		★
Multinomial logistic regression (SoftMax)		★
RNN and GRU		✔
Split and concatenation		★
Convolutional layer		★
(Fully hyperbolic) attention mechanism		★

Poincaré multi-head attention

Table 2: Negative log-likelihood on the test set. Oracle indicates the score when estimating the test sets with the ground truth MoG parameters. For each setting, we report the 95% confidence intervals for all converged results from the five trials. The numbers in brackets indicate the diverged trials.

Model	K=4	K=5	K=6	K=7	K=8
Oracle	1.485	1.675	1.857	2.003	2.132
Set Transformer w/o LN	**1.556**$\pm$0.214 (3)	1.912$\pm$0.701 (2)	**2.032**$\pm$0.193 (3)	5.066$\pm$5.239 (3)	**2.608**$\pm$N/A (4)
Ours	**1.558**$\pm$0.008 (0)	**1.833**$\pm$0.046 (0)	**2.081**$\pm$0.036 (0)	**2.370**$\pm$0.098 (0)	**2.682**$\pm$0.164 (0)
Set Transformer w/ LN	1.558$\pm$0.032 (0)	1.776$\pm$0.030 (0)	2.046$\pm$0.030 (0)	2.297$\pm$0.047 (0)	2.519$\pm$0.020 (0)

Poincaré convolutional layers

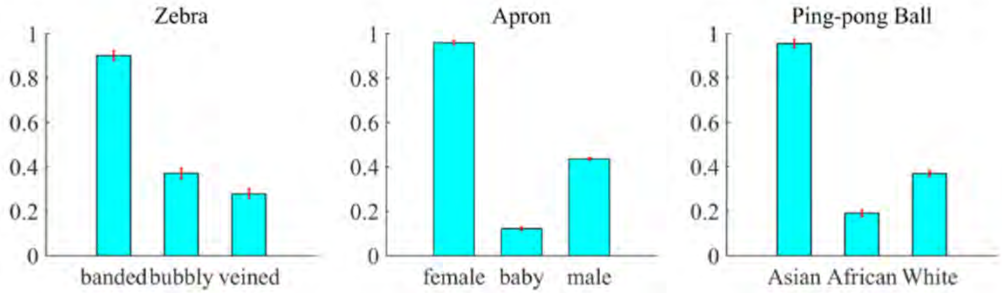
Table 3: BLEU-4 scores [33] on newstest2013. The target sentences were decoded using beam search with a beam size of five. D indicates the dimensions of token embeddings and the final MLR layer.

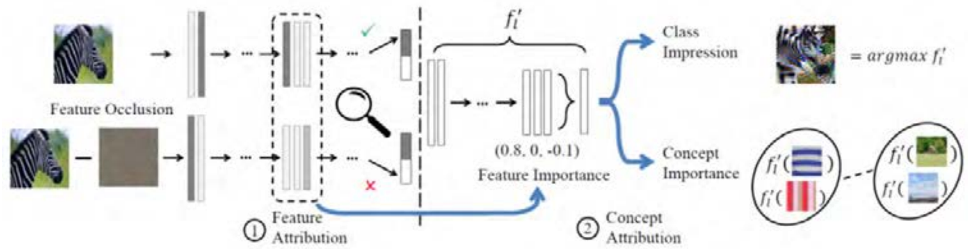
Model	D=16	D=32	D=64	D=128	D=256
ConvSeq2Seq [10]	2.68	8.43	14.92	**20.02**	**21.84**
Ours	**9.81**	**14.11**	**16.95**	19.40	21.76

Non-Euclidean Neural Networks!

Explainable AI (XAI)

- Weibin Wu, Yuxin Su, Xixian Chen, Shenglin Zhao, Irwin King, Michael R. Lyu, Yu-Wing Tai. Towards Global Explanations of Convolutional Neural Networks with Concept Attribution. CVPR2020





- Zixuan Huang, Yin Li. Interpretable and Accurate Fine-grained Recognition via Region Grouping. CVPR2020