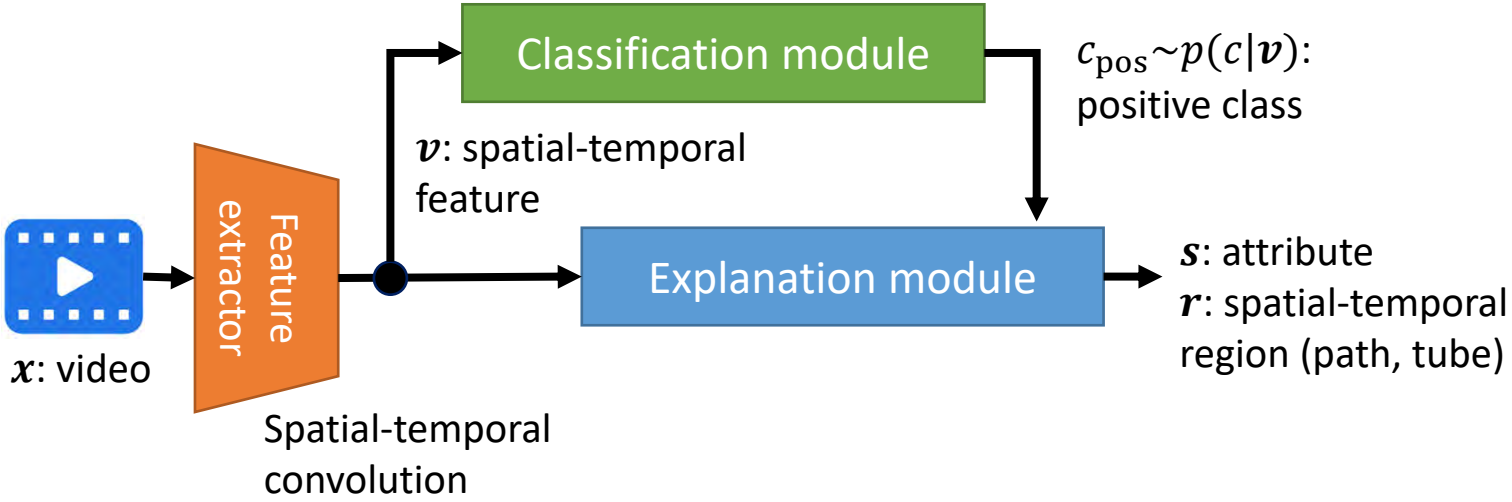
# Multimodal Explanations by Predicting Counterfactuality

Atsushi Kanehira, Kentaro Takemoto, Sho Inayoshi, Tatsuya Harada. CVPR2019 (oral)

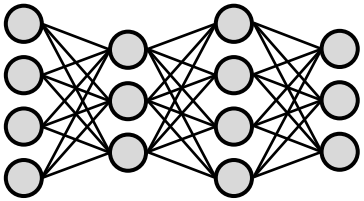
RopeClimbing not Salsaspin because Body Motion is Vertical Up

Biking not Skateboarding because Posture is Sitting



# 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

# Datasets

- ImageNet ReaL
  - Lucas Beyer, Olivier J. Hénaff, Alexander Kolesnikov, Xiaohua Zhai, Aäron van den Oord. Are we done with ImageNet?. arXiv:2006.07159
  - 1.2 million images, 1000 categories

- The JFT-300M
  - Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting unreasonable effectiveness of data in deep learning era. ICCV2017
  - 300M images, on average each image has 1.26 labels.
  - 18291 categories: e.g., 1165 type of animals and 5720 types of vehicles are labeled in the dataset.

- GraspNet-1Billion
  - Hao-Shu Fang, Chenxi Wang, Minghao Gou. Cewu Lu. GraspNet-1Billion: A Large-Scale Benchmark. CVPR2020
  - 97,280 RGB-D images, one billion grasp poses

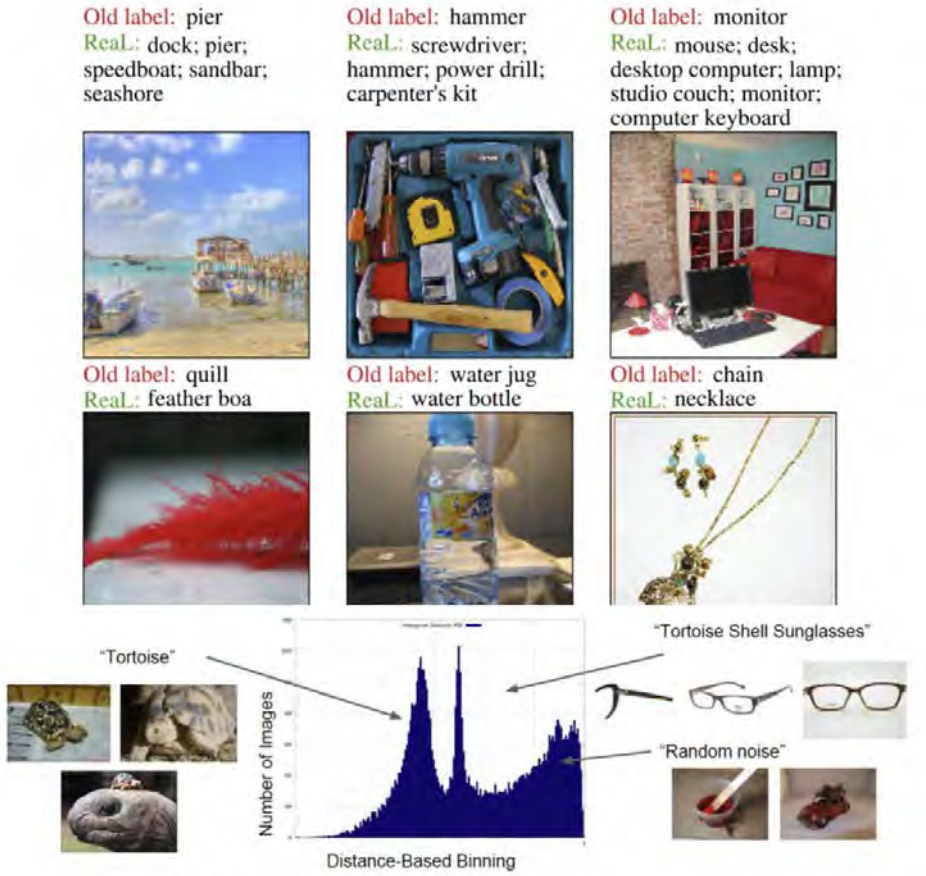
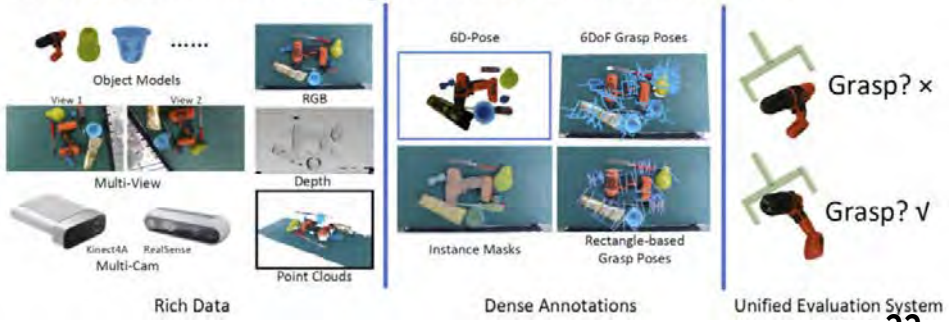
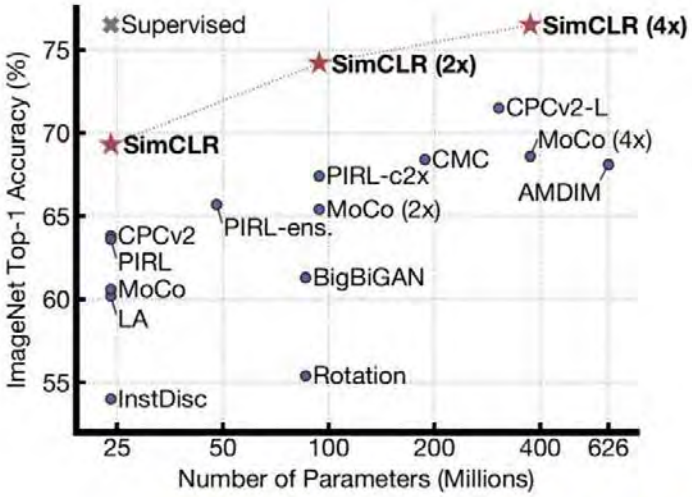


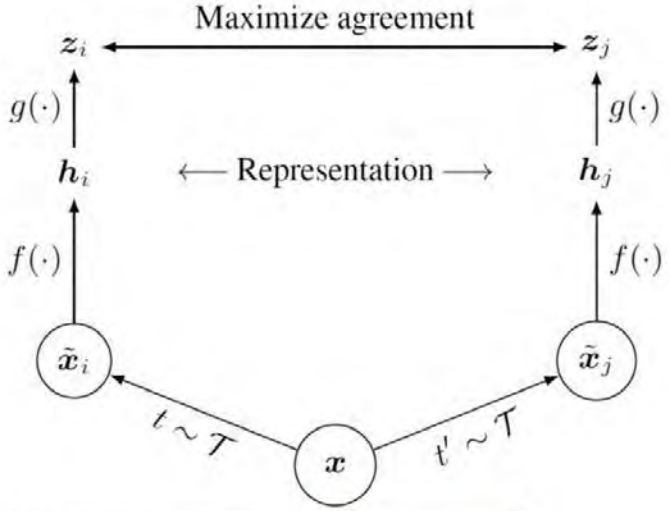

Figure 2. JFT-300M dataset can be noisy in terms of label confusion and incorrect labels. This is because labels are generated via a complex mixture of web signals, and not annotated or cleaned by humans. x-axis corresponds to the quantized distances to K-Means centroids, which are computed based on visual features.



# Self Supervised Learning (SSL)

- SimCLR (Google)
  - **Contrastive Learning**
  - Ting Chen, Simon Kornblith, Mohammad Norouzi, Geoffrey Hinton. A Simple Framework for Contrastive Learning of Visual Representations. ICML2020



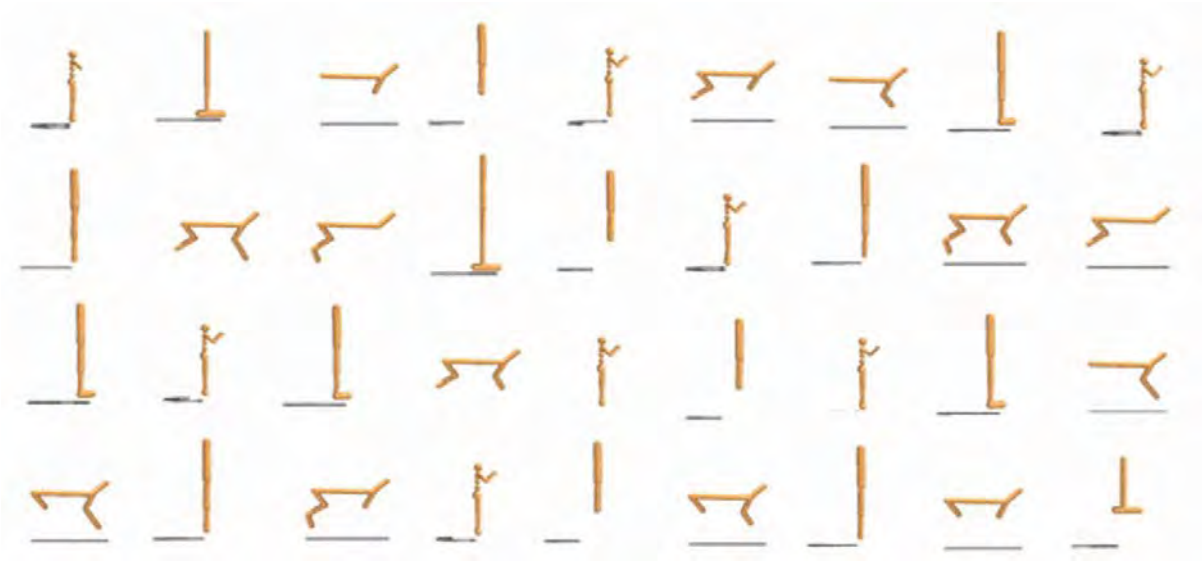

(a) Original (b) Crop and resize (c) Crop, resize (and flip) (d) Color distort. (drop) (e) Color distort. (jitter)

(f) Rotate $\{90°, 180°, 270°\}$ (g) Cutout (h) Gaussian noise (i) Gaussian blur (j) Sobel filtering

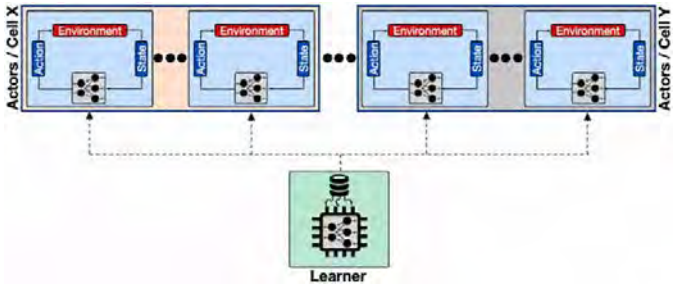
2

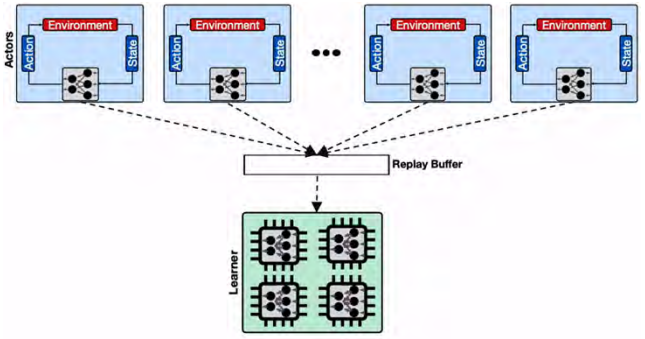
# Reinforcement Learning

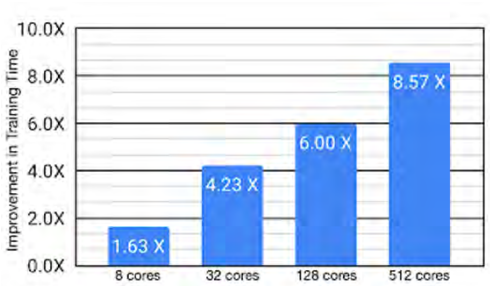
- Wenlong Huang (UC Berkeley), Igor Mordatch (Google), Deepak Pathak (CMU, FAIR). One Policy to Control Them All: Shared Modular Policies for Agent-Agnostic Control. ICML2020
  - https://wenlong.page/modular-rl/



- Amir Yazdanbakhsh Junchao Chen Yu Zheng. Menger: Massively Large-Scale Distributed Reinforcement Learning. NeurIPS, Beyond Backpropagation Workshop, 2020 (2020)
  - https://ai.googleblog.com/2020/10/massively-large-scale-distributed.html



# How much training data?

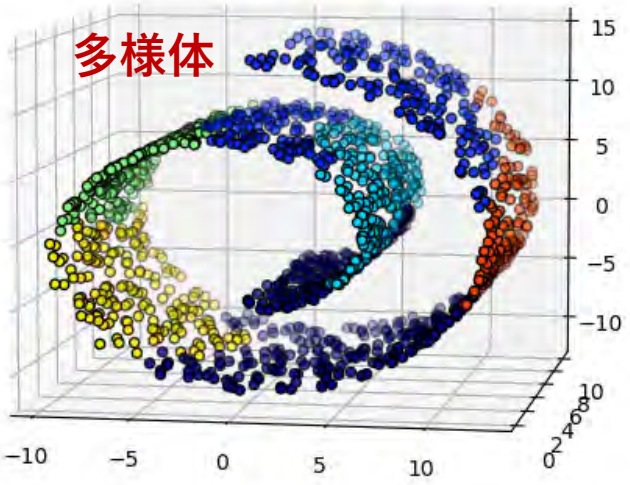
学習に必要な訓練データ数？ $N \gg d$

自然画像の次元数

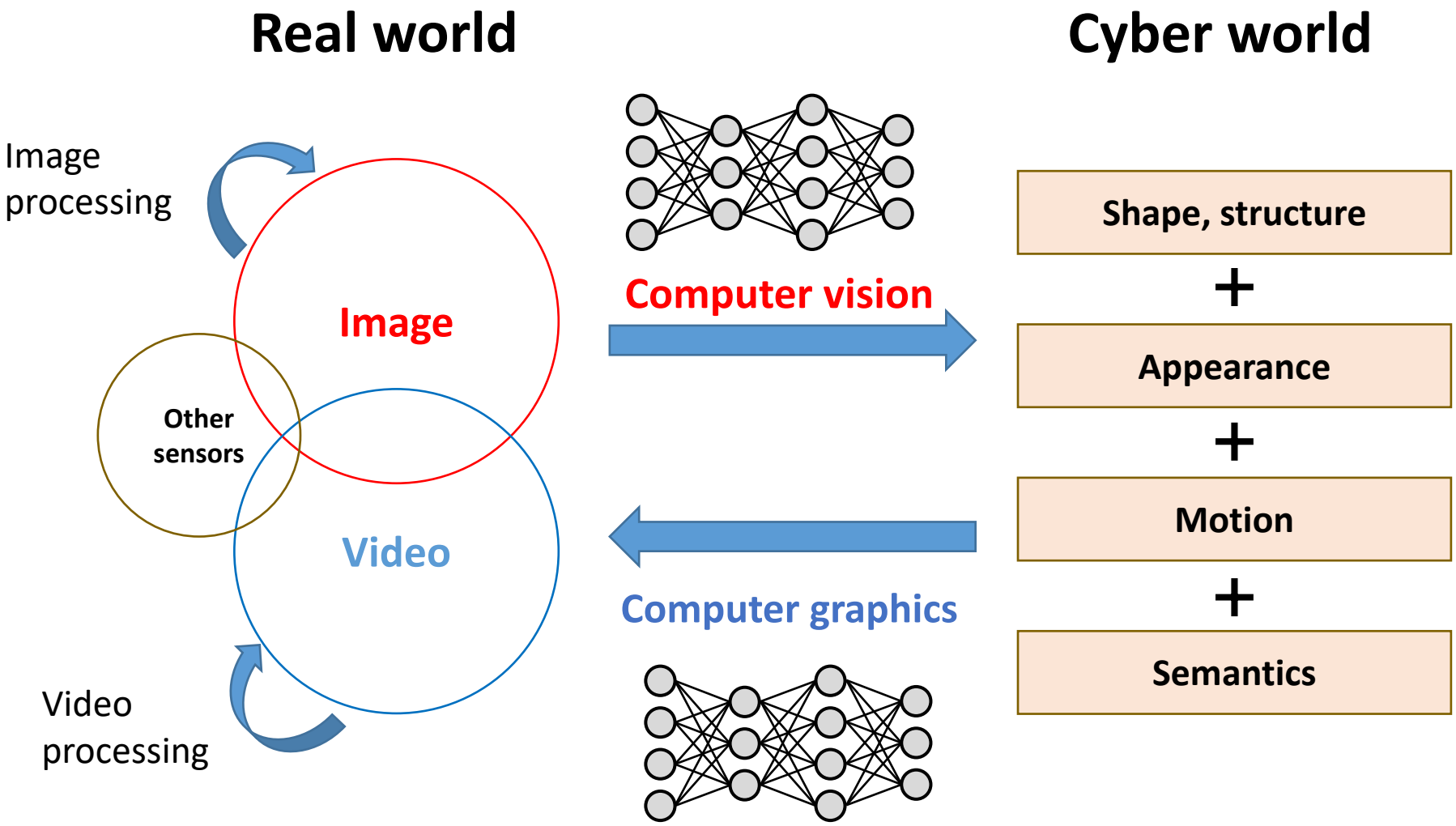

https://www.sony.jp/ichigan/products/ILCE-1/feature_1.html

5010万画素！

データの低次元の構造を捉えることで必要なデータ量を削減




# Integration of CV and CG



# Differentiable Renderer

- 3D world from 2D: fundamental problem
- Usual renderer is not differentiable!
- Our proposal: differentiable renderer

Renderer

3D shape, pose, lighting, camera, textures → 2D image

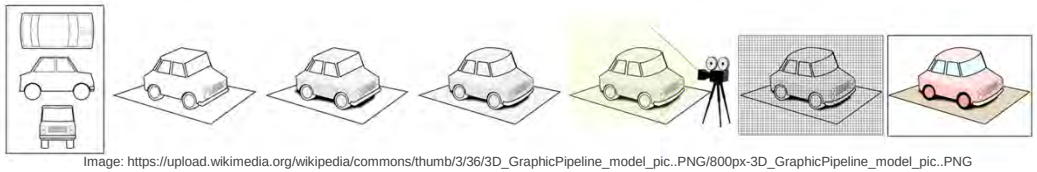
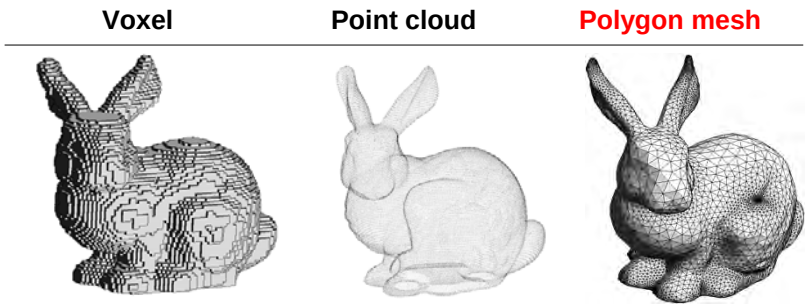


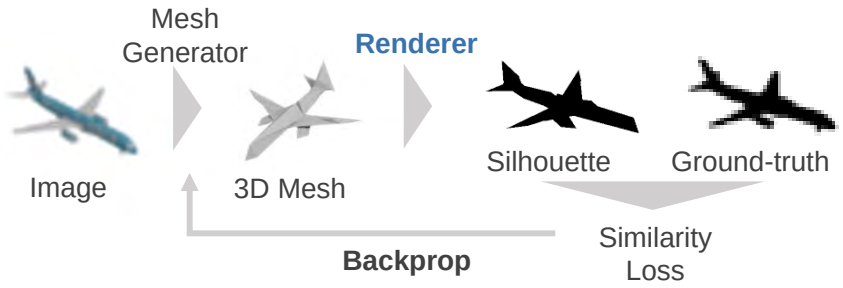
Image: https://upload.wikimedia.org/wikipedia/commons/thumb/3/36/3D_GraphicPipeline_model_pic..PNG/800px-3D_GraphicPipeline_model_pic..PNG

Hiroharu Kato, Yoshitaka Ushiku, Tatsuya Harada.
Neural 3D Mesh Renderer. CVPR2018



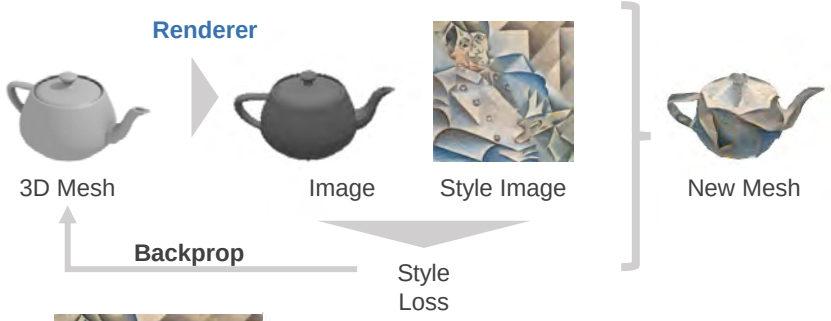

http://cse.iitkgp.ac.in/~pb/research/3dpoly/3dpoly.html
http://waldyrious.net/learning-holography/pb-cgh-formulas.xhtm
http://www.cs.mun.ca/~omeruvia/philosophy/images/BunnyWire.gif

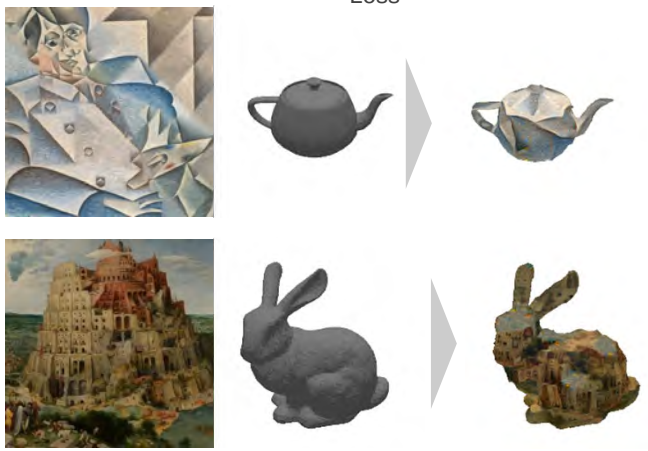
3D mesh reconstruction from single image





3D-to-2D image style transfer, 3D DeepDream



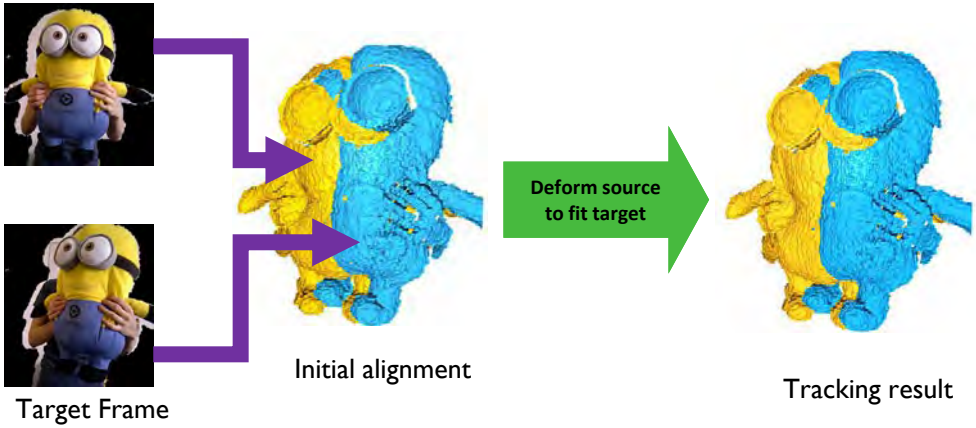
**Bridging gap between Computer Graphics and Machine Learning!**

# Non-Rigid Tracking and Mapping

Li, Bozic, Zhang, Ji, Harada & Niessner , CVPR2020 oral

Li, Zhang, Nakamura & Harada. IROS2020

- Find the deformation that best align two frames



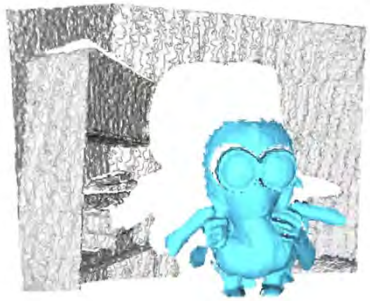

Result on **Fr3/walking_xyz** sequence


SplitFusion (Ours)


KinectFusion


DynamicFusion

Result on **Minion** sequence

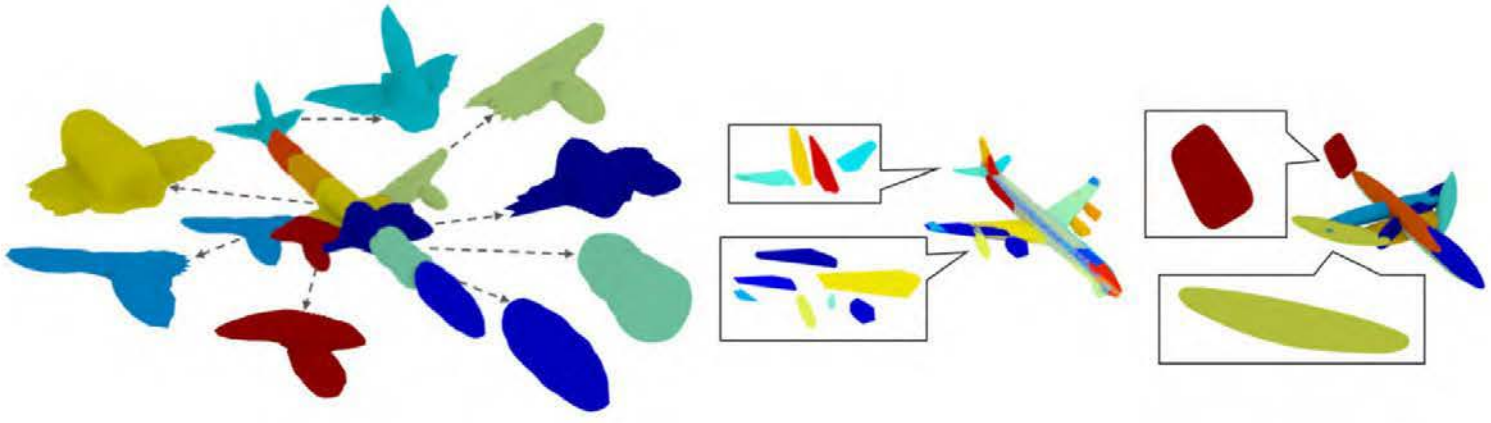


Result on **Selfie** sequence

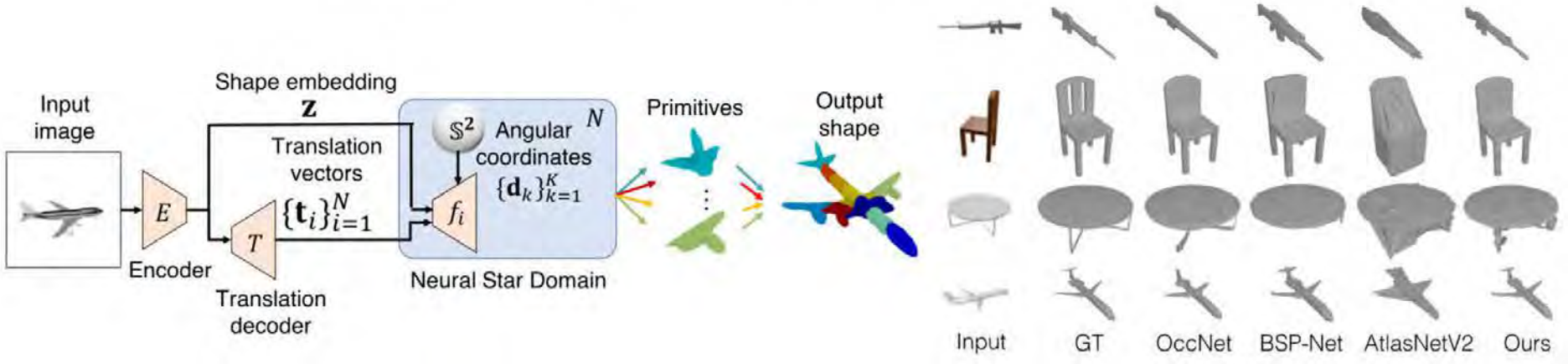
# 3-dimensional Shape Primitive Representation

Kawana, Mukuta & Harada. NeurIPS2020

- We propose Neural Star Domains as a new primitive representation of 3D shapes.
- NSD can learn analytic and semantic partial shapes.
- NSD is a new representation method that integrates implicit and explicit representations of 3D shapes, which can be used for both learning and inference.

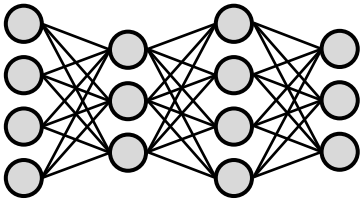


Neural Star Domain (ours) Convex (BSP-Net) Superquadrics



# 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

# Supercomputers

TOP 10 Sites for November 2020

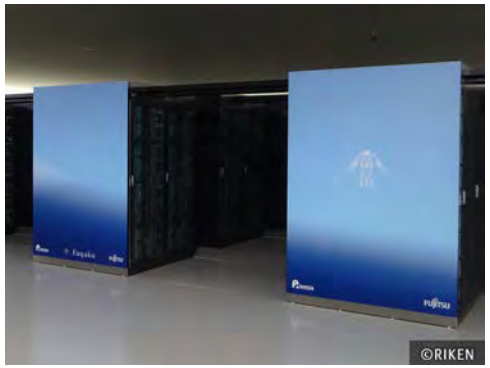
For more information about the sites and systems in the list, click on the links or view the complete list.

1-100 101-200 201-300 301-400 401-500

| Rank | System | Cores | Rmax (TFlop/s) | Rpeak (TFlop/s) | Power (kW) |
|---|---|---|---|---|---|
| 1 | **Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, **Fujitsu**<br>RIKEN Center for Computational Science<br>Japan | 7,630,848 | 442,010.0 | 537,212.0 | 29,899 |
| 2 | **Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, **IBM**<br>DOE/SC/Oak Ridge National Laboratory<br>United States | 2,414,592 | 148,600.0 | 200,794.9 | 10,096 |
| 3 | **Sierra** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, **IBM / NVIDIA / Mellanox**<br>DOE/NNSA/LLNL<br>United States | 1,572,480 | 94,640.0 | 125,712.0 | 7,438 |
| 4 | **Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway, **NRCPC**<br>National Supercomputing Center in Wuxi<br>China | 10,649,600 | 93,014.6 | 125,435.9 | 15,371 |
| 5 | **Selene** - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, **Nvidia**<br>NVIDIA Corporation<br>United States | 555,520 | 63,460.0 | 79,215.0 | 2,646 |
| 6 | **Tianhe-2A** - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000, **NUDT**<br>National Super Computer Center in Guangzhou<br>China | 4,981,760 | 61,444.5 | 100,678.7 | 18,482 |

537 PFLOPS

https://www.top500.org/lists/top500/2020/11/




富岳

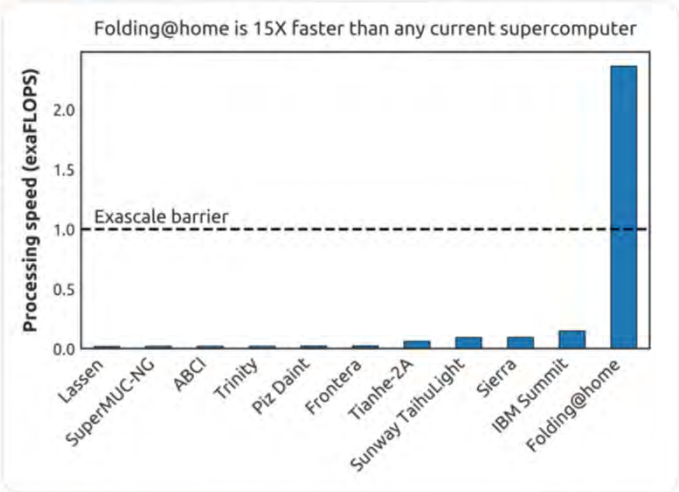
https://www.fujitsu.com/jp/about/businesspolicy/tech/fugaku/

**Folding@home**
@foldingathome

2.4 exaFLOPS!

With our collective power, we are now at ~2.4 exaFLOPS (faster than the top 500 supercomputers combined)! We complement supercomputers like IBM Summit, which runs short calculations using 1000s of GPUs at once, by spreading longer calculations around the world in smaller chunks!

ツイートを翻訳




午前4:16 · 2020年4月14日 · Twitter Web App

https://twitter.com/foldingathome/status/1249778379634675712

# Progress in Computer System

- High performance edge devices

https://www.apple.com

2.6 TFLOPS (FP32)?


100,000 yen!

https://www.apple.com

800 GFLOPS (FP32)?

https://ja.wikipedia.org/wiki/地球シミュレータ

Earth Simulator:
122.4 TFLOPS, Jun 2009

- Limitation of Moore's Law?
  - 2016: 10 nm process
  - 2021: 5 nm process ?
  - Atom: 0.1 nm


https://en.wikipedia.org/wiki/Moore%27s_law

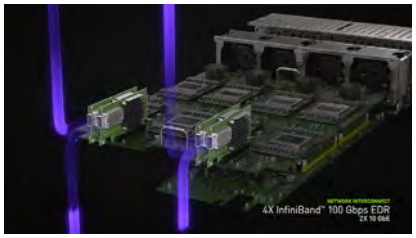
ILSVRC2010

| Team | Flat Error |
|---|---|
| 1) NEC-UIUC | 0.28 |
| 2) XRCE | 0.37 |
| 3) ISI (ours) The Univ. of Tokyo | 0.45 |

We use only one 8-cores workstation for both training and recognition except feature extractions.

72 GFLOPS? (double)

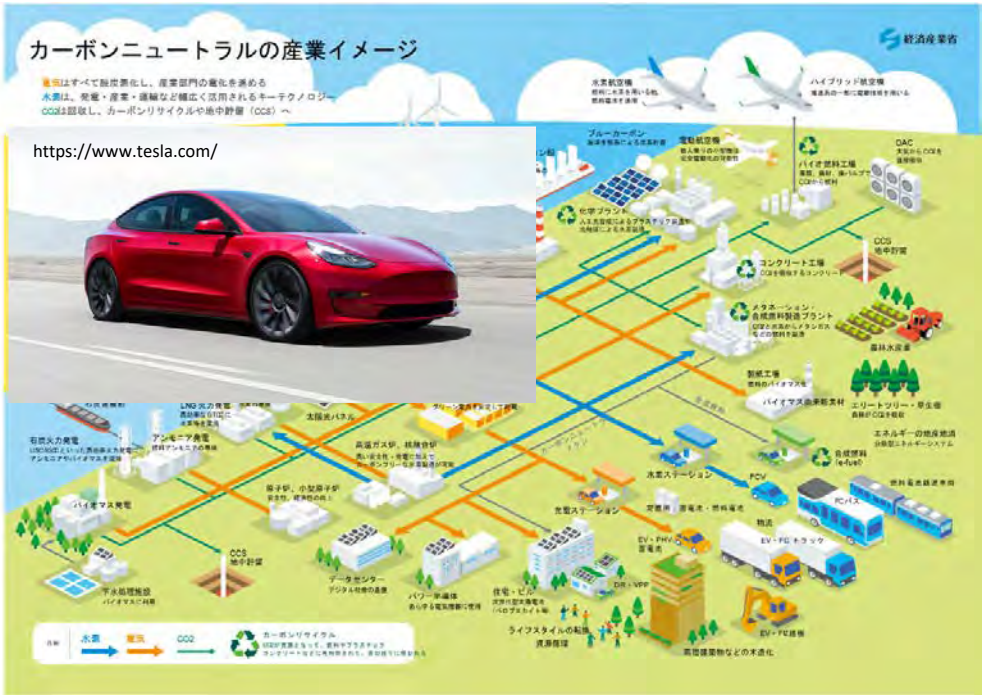
- High speed wireless networks
  - 4G: 1Gbps （2010年くらい）
  - 5G: 20Gbps （2020年くらい）
  - 6G: 100Gbps? （2030年くらい？）




Nvidia DGX-1: 100 Gbps EDR, 2016

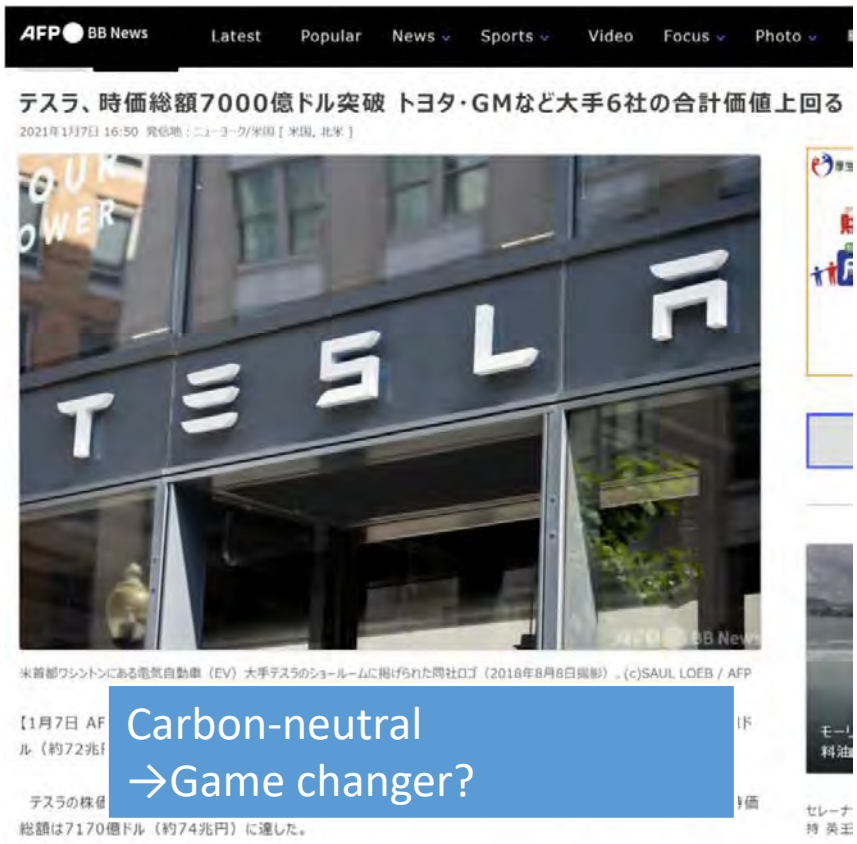
- Next Generations
  - Quantum Computing
  - Neuromorphic Spiking Neural Networks

# Carbon-neutral?




https://www.meti.go.jp/press/2020/12/20201225012/20201225012.html

https://www.afpbb.com/




Carbon-neutral
→Game changer?

## Tim Cook, Apple CEO

https://9to5mac.com/2021/01/28/tim-cook-talks-about-apples-commitment-to-privacy-and-endorses-gdpr-in-new-interview/

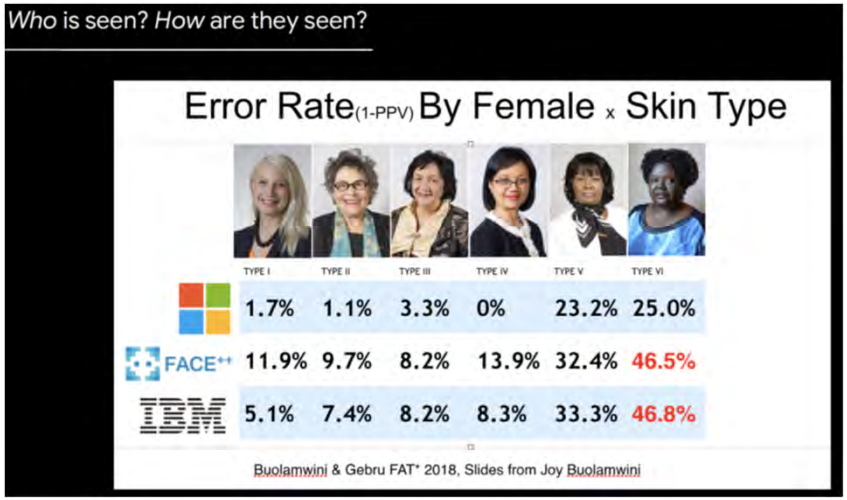
"Cook said privacy has become as much of a concern as climate change, as both things will have a profound impact on future generations. "

- GDPR: The European Union (EU) General Data Protection Regulation
- CCPA: California Consumer Privacy Act of 2018

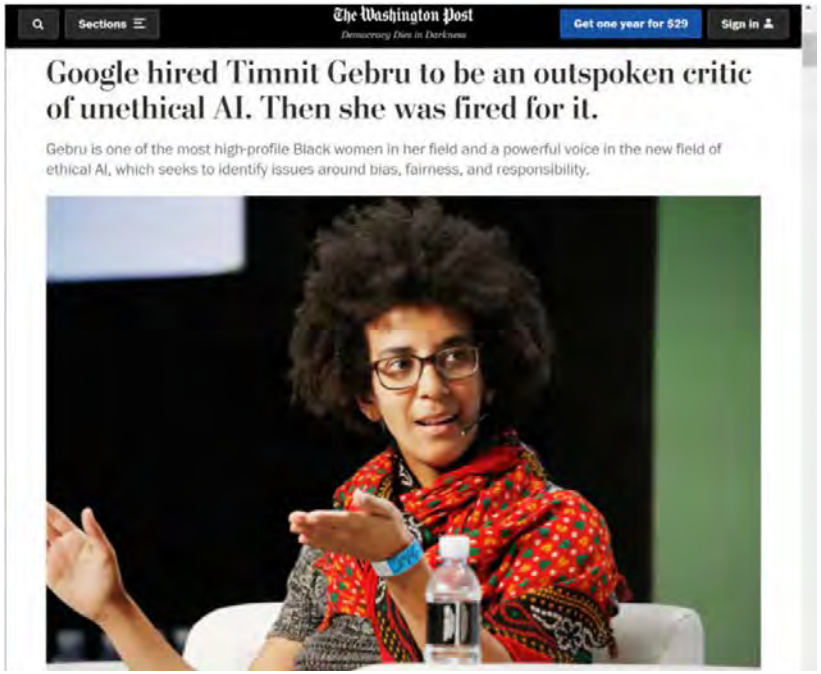
Privacy issues
→Game changer?

# Fairness Accountability Transparency and Ethics (FATE)

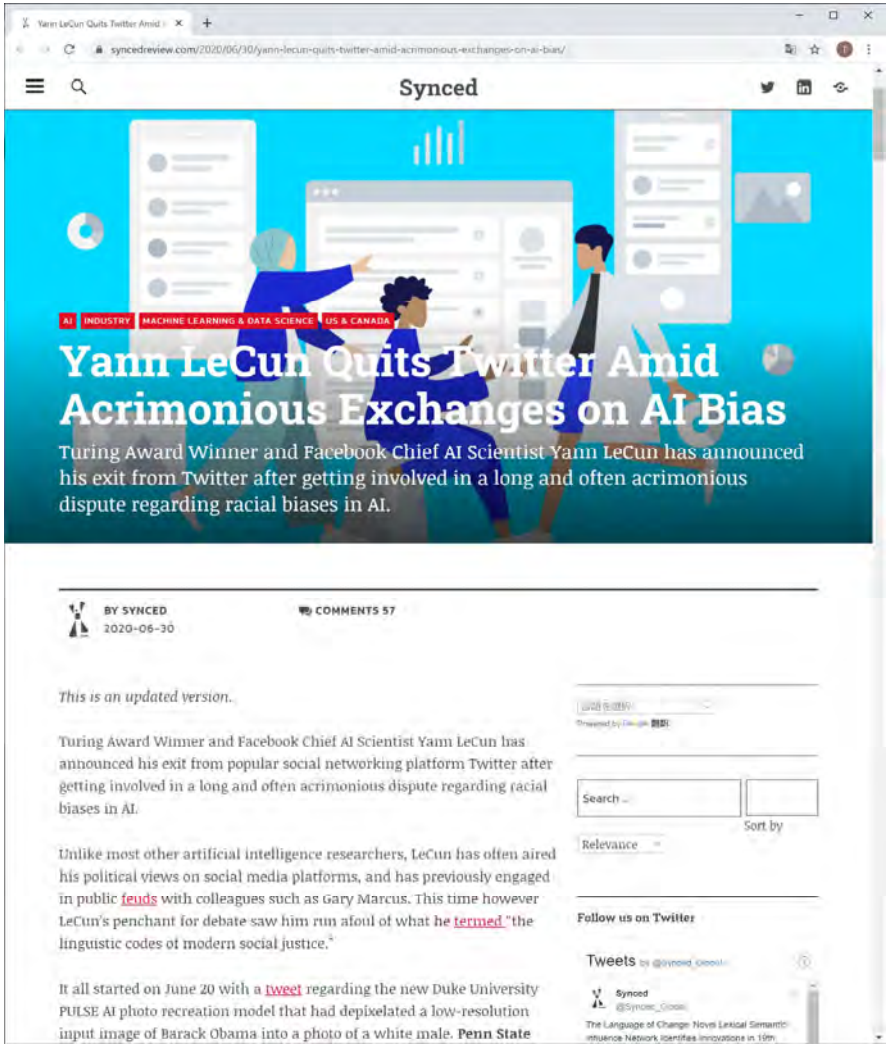
- Tutorial on Fairness Accountability Transparency and Ethics in Computer Vision at CVPR 2020
  - Computer vision in practice: who is benefiting and who is being harmed?
  - Speaker: Timnit Gebru




https://drive.google.com/file/d/1rcG8KVmjRUWWNSg-R6cTBlAScP9UkCJp/view




https://www.washingtonpost.com/technology/2020/12/23/google-timnit-gebru-ai-ethics/



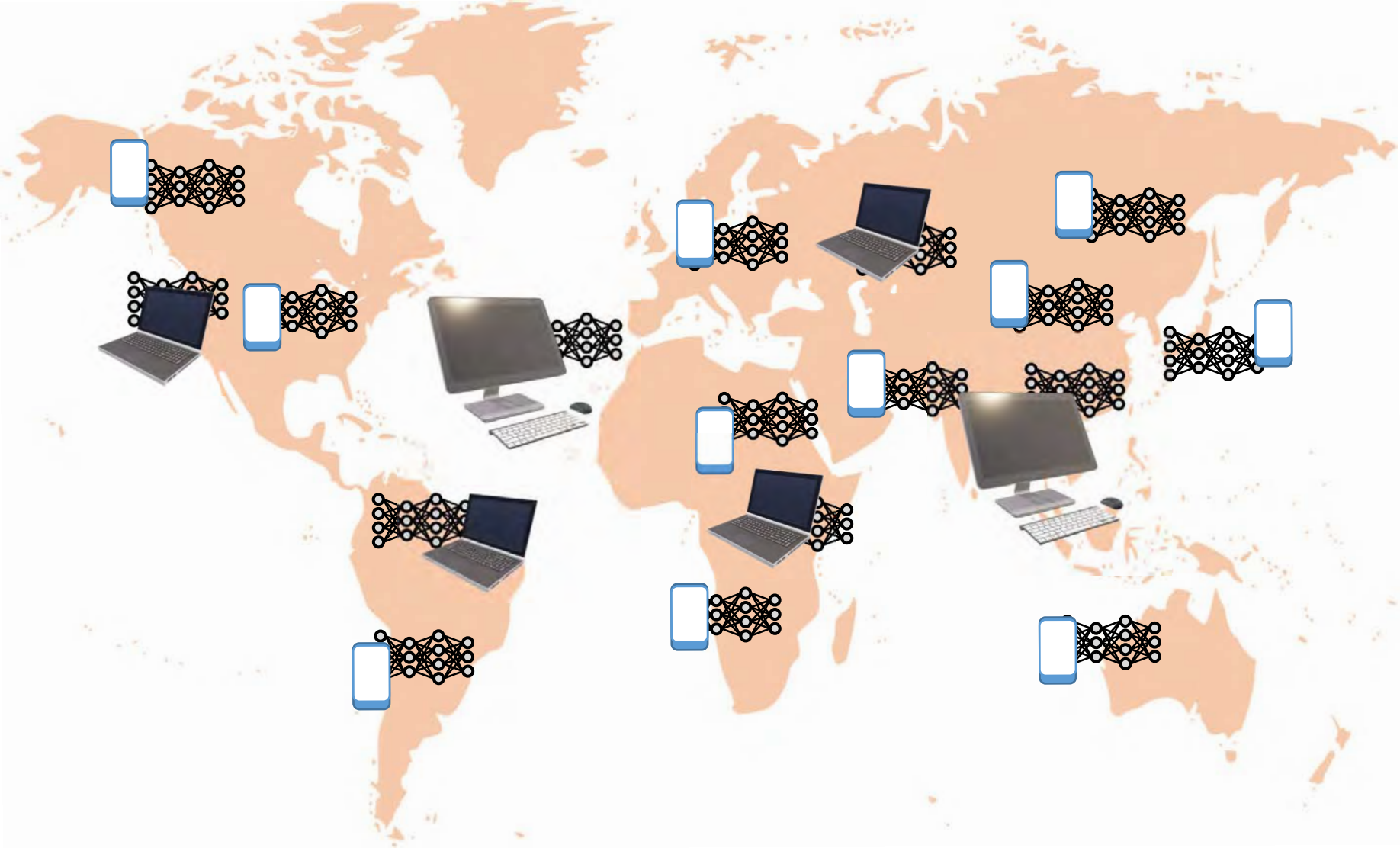

https://syncedreview.com/2020/06/30/yann-lecun-quits-twitter-amid-acrimonious-exchanges-on-ai-bias/

# Fairness in Dataset

- ImageNet
  - Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy
  - http://image-net.org/update-sep-17-2019


- TinyImages: https://groups.csail.mit.edu/vision/TinyImages/
  - June 29th, 2020
  - It has been brought to our attention [1] that the **Tiny Images dataset contains some derogatory terms as categories and offensive images.** This was a consequence of the automated data collection procedure that relied on nouns from WordNet. We are greatly concerned by this and apologize to those who may have been affected.
  - The dataset is too large (80 million images) and the images are so small (32 x 32 pixels) that it can be difficult for people to visually recognize its content. Therefore**, manual inspection, even if feasible, will not guarantee that offensive images can be completely removed.**
  - **We therefore have decided to formally withdraw the dataset.** It has been taken offline and it will not be put back online. We ask the community to refrain from using it in future and also delete any existing copies of the dataset that may have been downloaded.

[1] Large image datasets: A pyrrhic win for computer vision?, anonymous authors, OpenReview Preprint, 2020.

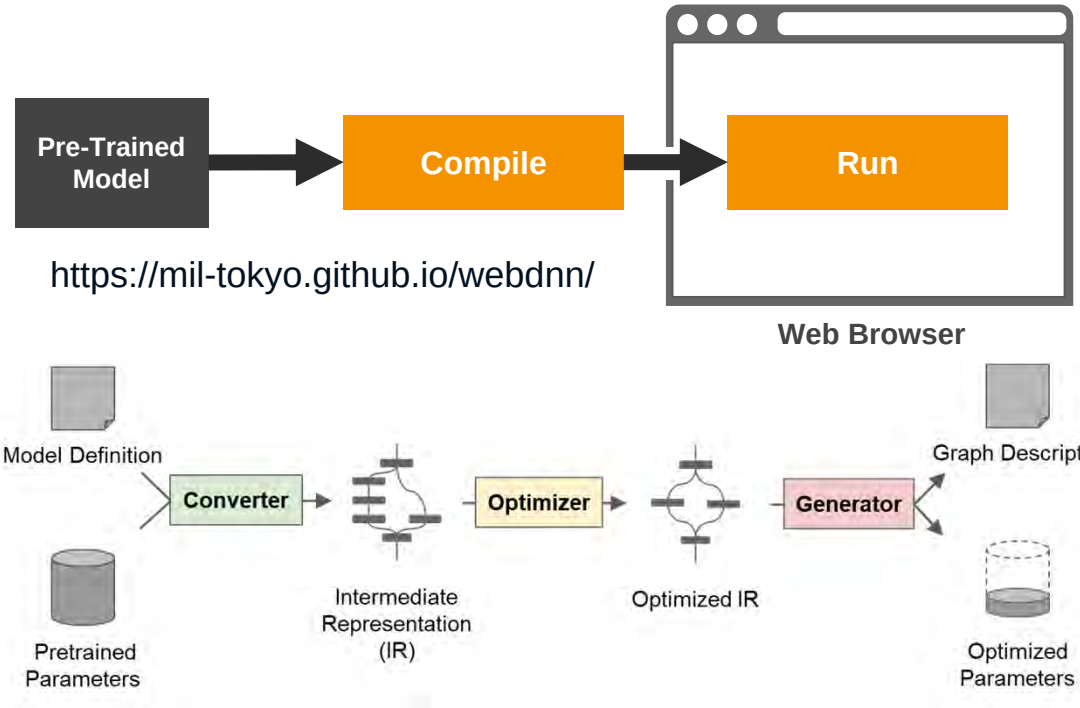


# WebDNN: Fastest DNN Framework on Web Browser

Hidaka, Kikura, Ushiku & Harada (ACMMM2017)
Hidaka, Miura & Harada (ICLRW2017)

**ACM Multimedia 2017 Honorable Mention Open source software Award**

No need to install any applications and libraries in your smartphone and laptop





- WebDNN compiles and optimizes pretrained model to execute on web browser.
- Tensorflow, Keras model, Caffe model, pytorch model, Chainer chain are supported.
- Open Neural Network Exchange (ONNX) is partially supported.

# 次世代のAI？

- 個別最適化だけではなく，全体最適化に目を向けることも重要.

- プラットフォーム非依存（Platform independent）
  - OS（windows, macos, linux, iOS, Android,,,）非依存
  - ハードウェア（CPU, GPU, Quantum Computer, Neuromorphic,,,）非依存

- 分散計算（Distributed computation）

- 非同期計算（Asynchronous computation）

- プライバシ保護（Privacy-aware）

- 分散データ（Distributed data）

- 多様性（Diversity）

- 連合学習（Federated learning）