ML Workshop Day 3

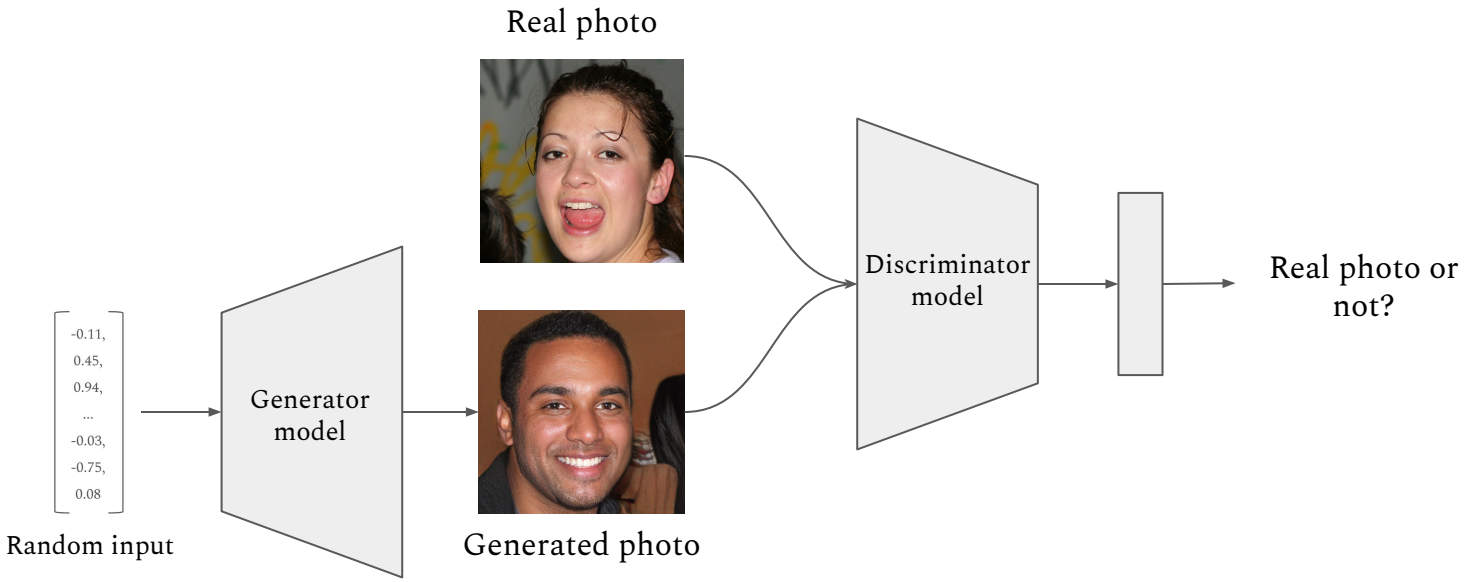
Recap of yesterday

- Generative Sequential ML models
  - Generative text
  - Generative drawing
  - Generative music

thispersondoesnotexist.com

StyleGAN,
December 2018

GAN = Generative Adversarial Networks



Ian Goodfellow
@goodfellow_ian

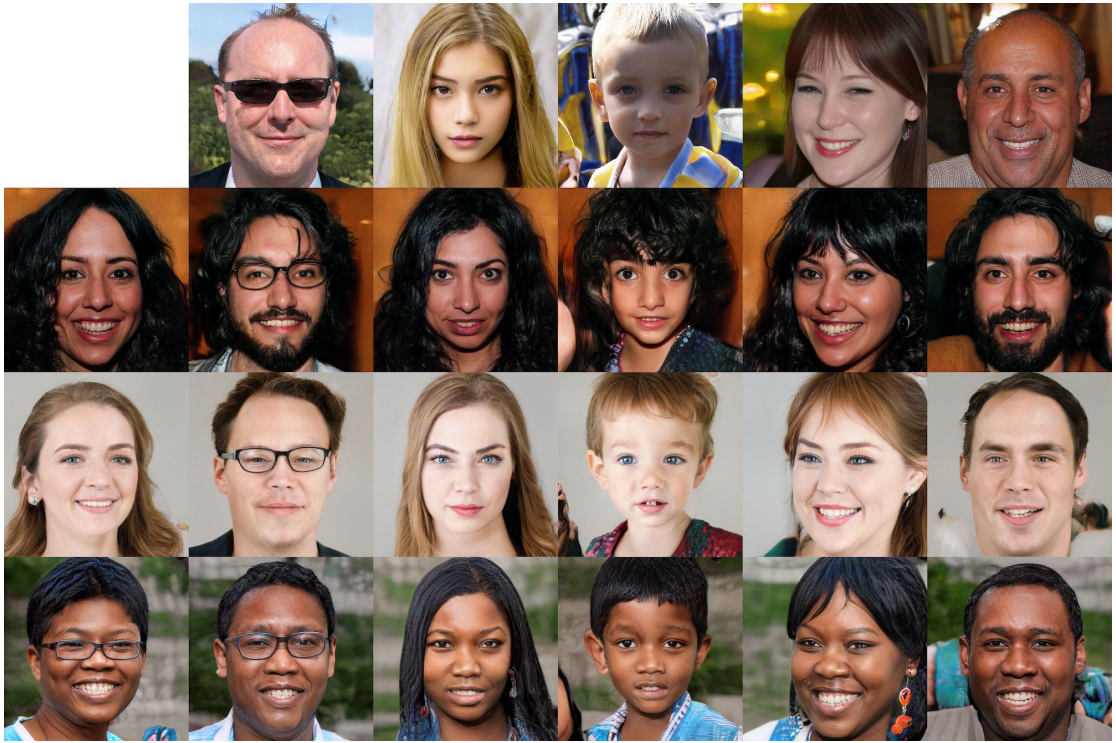
4.5 years of GAN progress on face generation.
arxiv.org/abs/1406.2661 arxiv.org/abs/1511.06434
arxiv.org/abs/1606.07536 arxiv.org/abs/1710.10196
arxiv.org/abs/1812.04948

1:40 AM · Jan 15, 2019 · Twitter Web Client

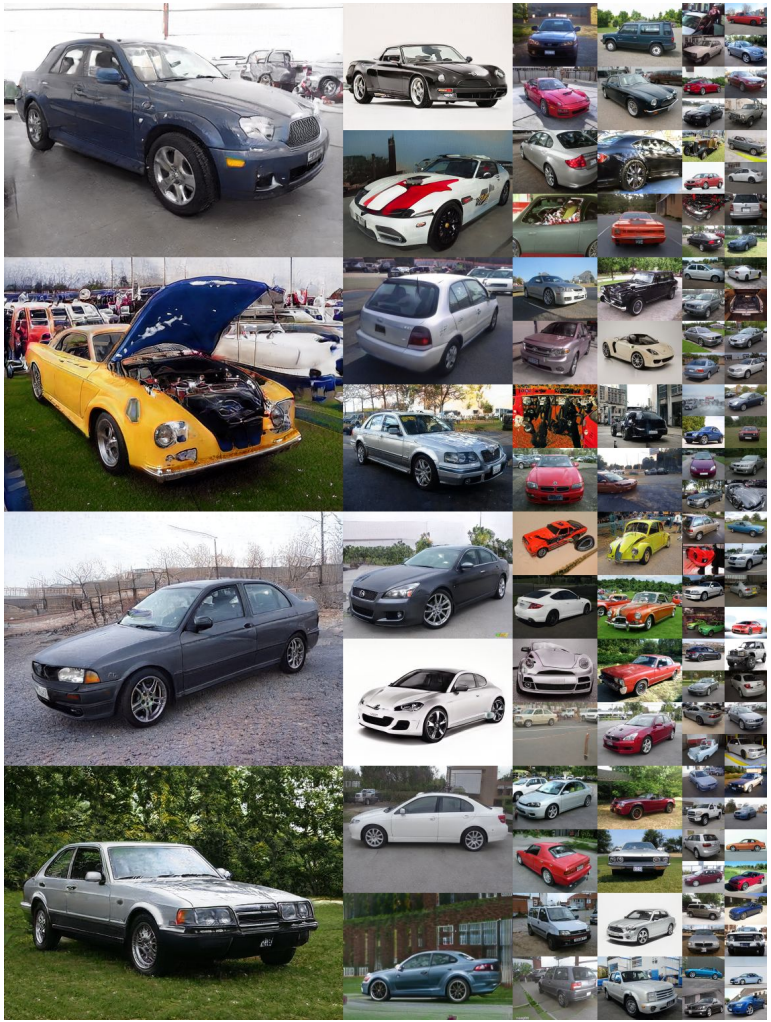
BigGAN

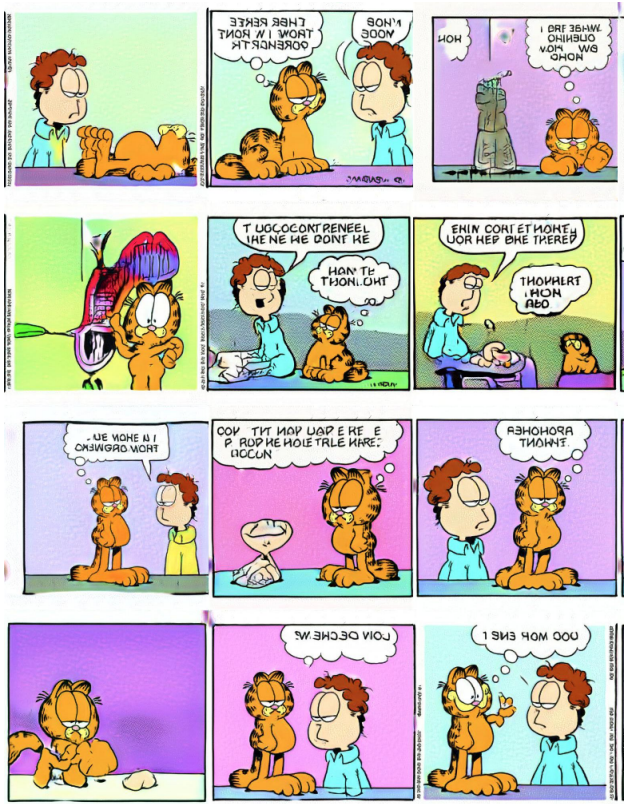
Large Scale GAN Training for High Fidelity Natural Image Synthesis, Andrew Brock et al, 2018

StyleGAN



A Style-Based Generator Architecture for Generative Adversarial Networks, Tero Karras et al, 2019





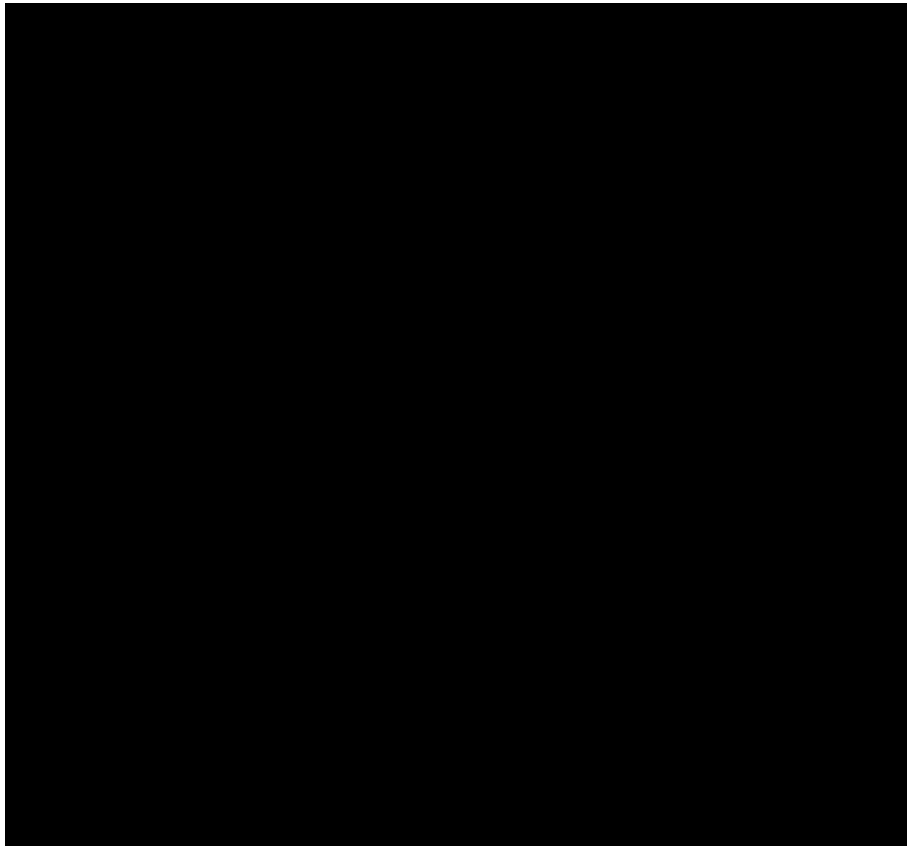
Michael Friesen
@MichaelFriese10

cursed emojis #StyleGAN

Kenji Doi
@knjcode

約9万枚のラーメン二郎画像で StyleGAN を試しました。
公式の実装を使って、最大解像度を512x512pxに下げて学習しています。
まだ学習途中(50%ぐらい)ですが、以前のPGGANよりさらにリアルになっているように思います。
#ラーメン二郎 #GAN

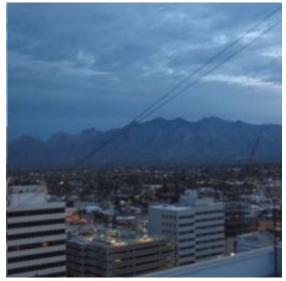
Translate Tweet

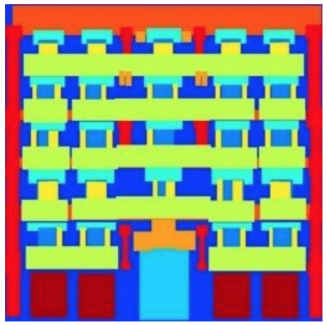


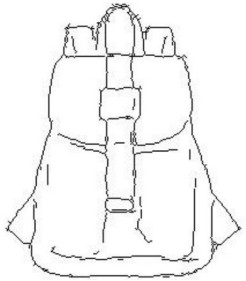
Conditional GANs

Image translation







Pix2Pix

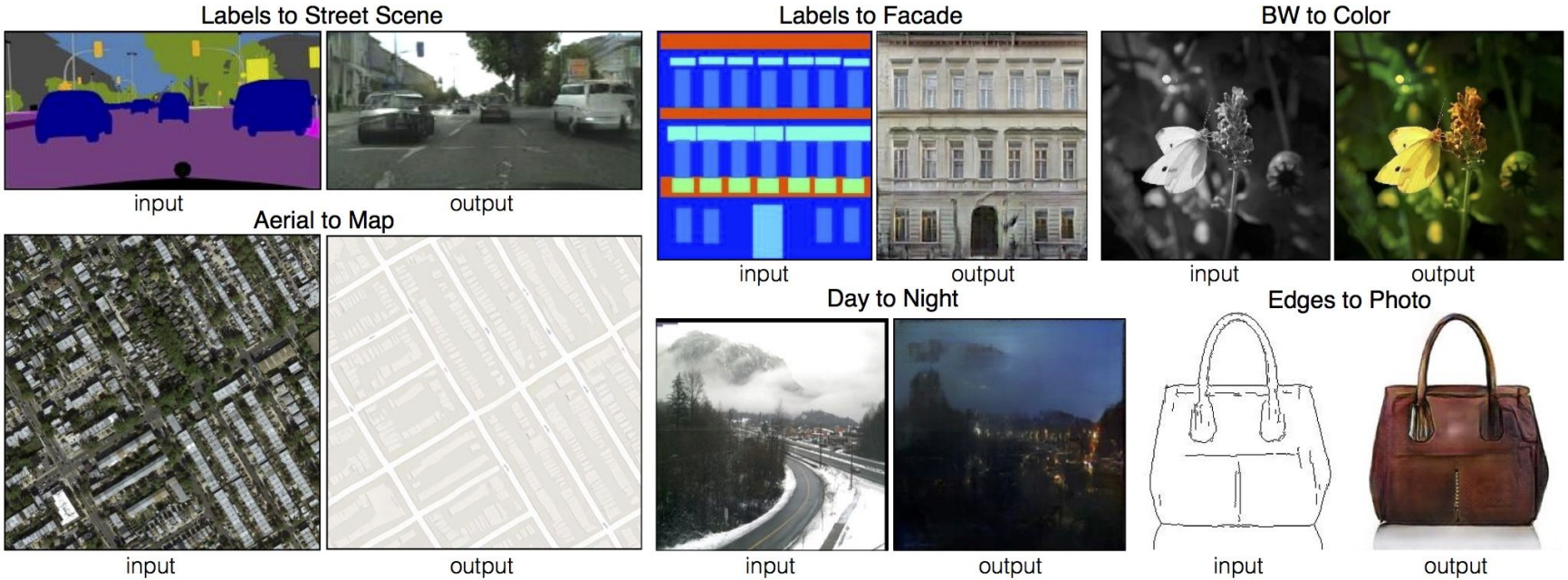
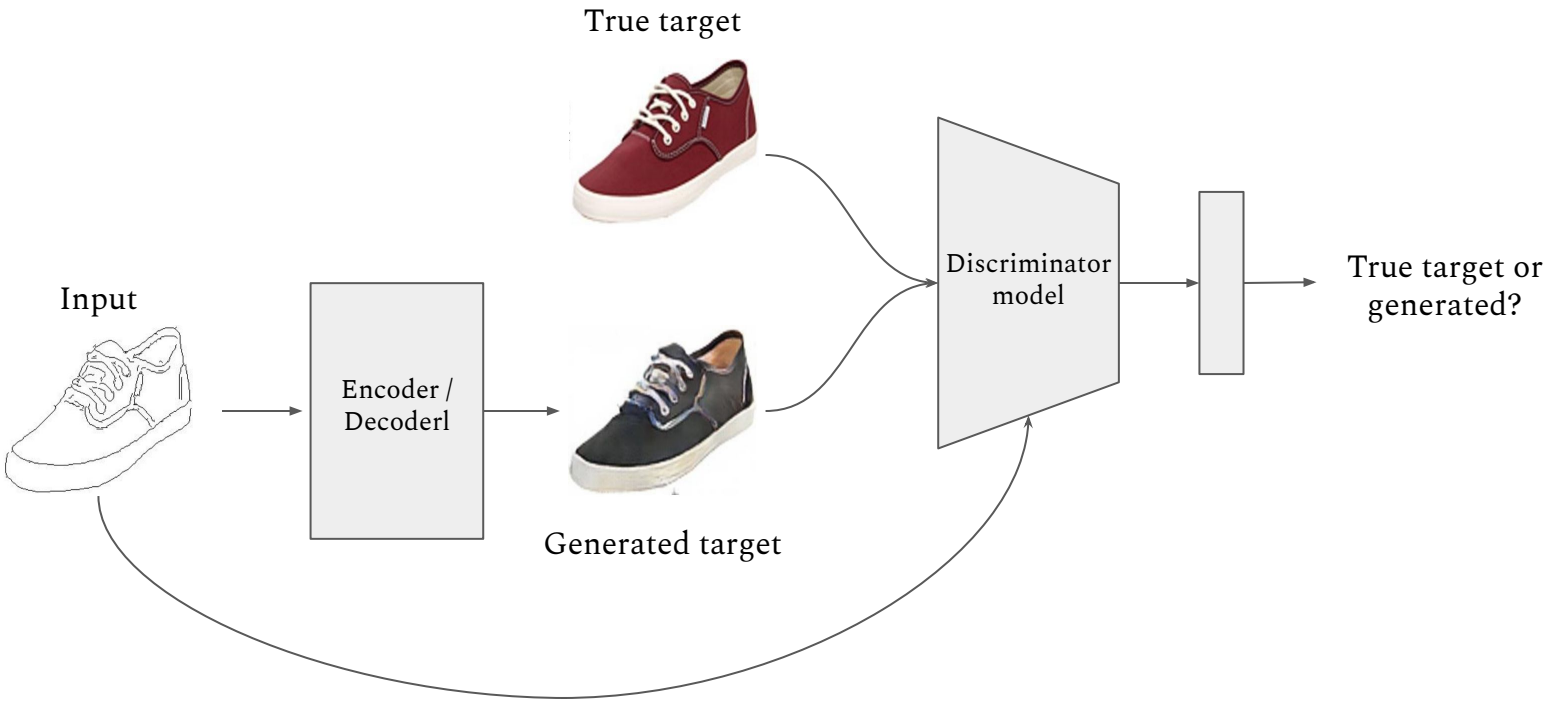
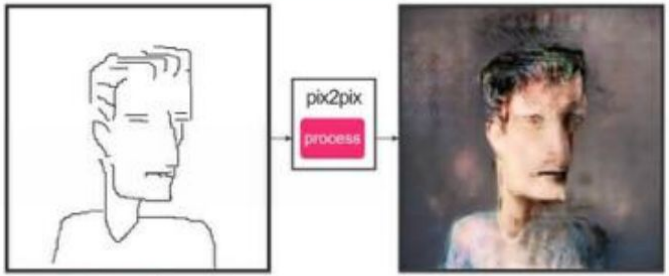


Image-to-Image Translation with Conditional Adversarial Nets, Philip Isola et al, 2017

Pix2Pix


True target
Input
Encoder / Decoderl
Generated target
Discriminator model
True target or generated?

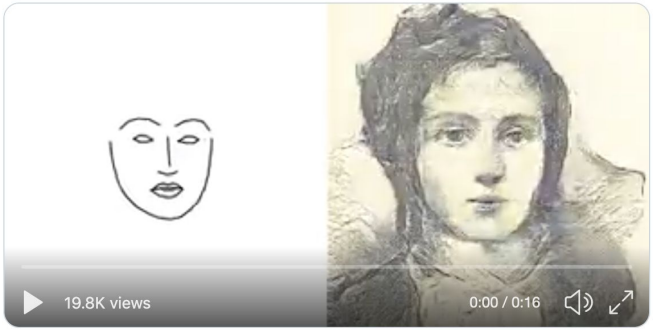
#fotogenerator



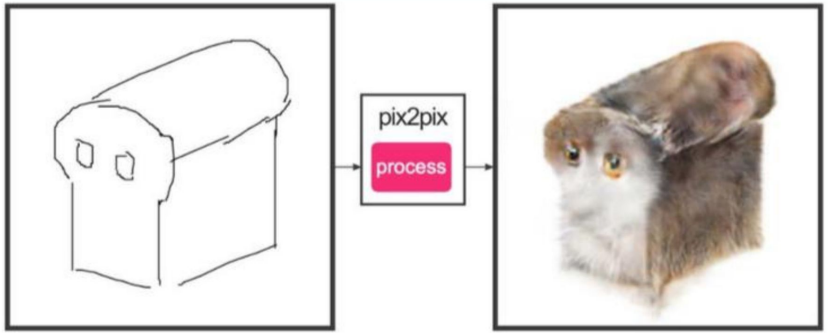
sketch by Yann LeCun

Mario Klingemann
@quasimondo

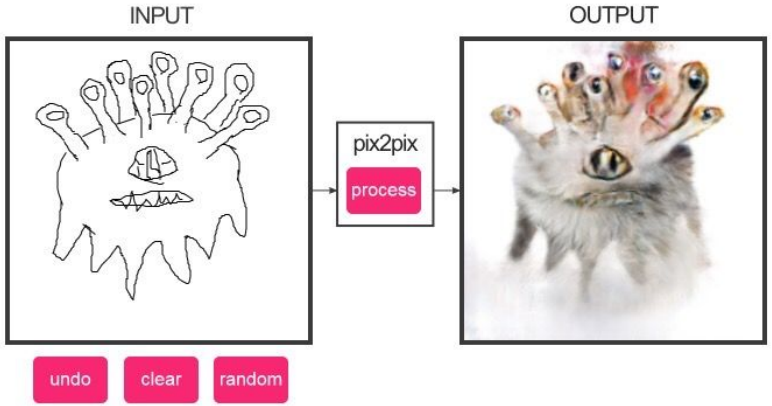
Generating faces from a sketch. I trained a pix2pix net on 1500 #bldigital faces. Left input, right output.



#edges2cats *by Christopher Hesse*



sketch by Ivy Tsai



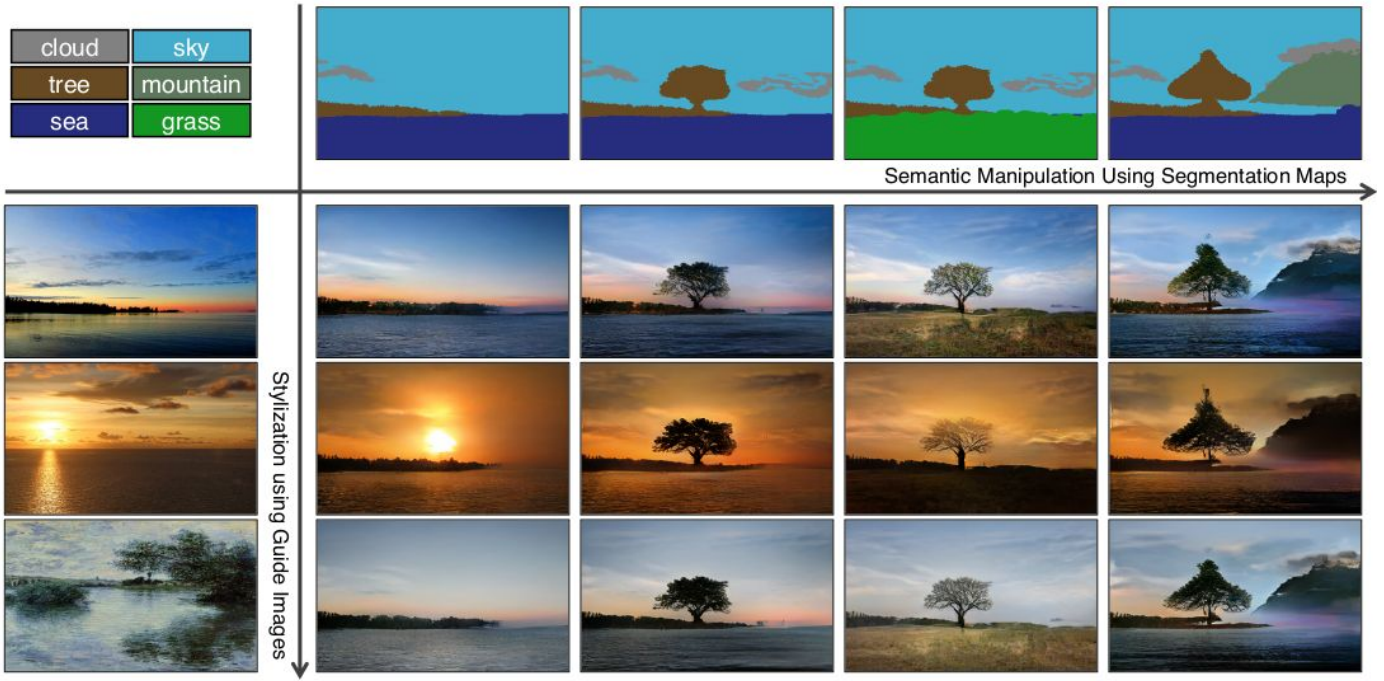
Pix2PixHD

Input labels

Synthesized image

High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs, 2017

GauGAN / SPADE



Semantic Image Synthesis with Spatially-Adaptive Normalization, Taesung Park et al, 2019

Hands-on with GANs
Runway ML & Artbreeder

Things to pay attention to

when using Machine Learning in real world

Challenges

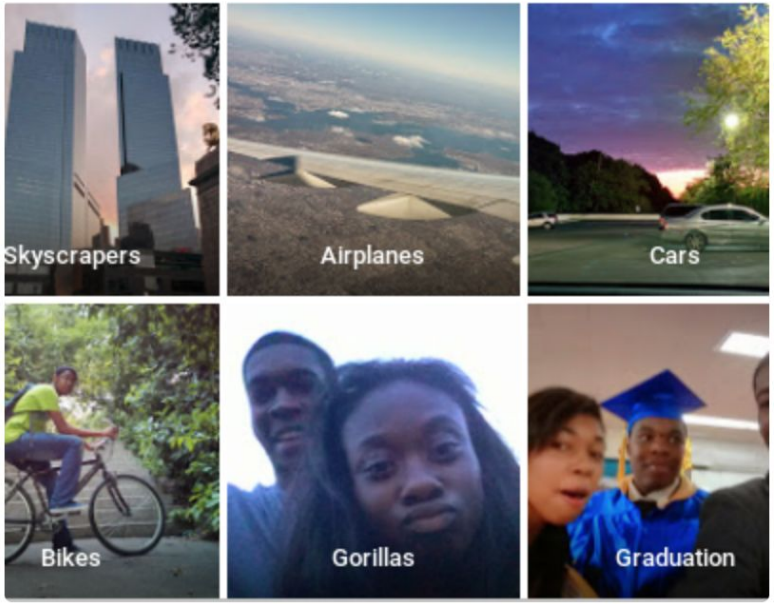
- Unexpected failures
- Limitation of computing power, storage and network bandwidth
- Privacy and compliance
- Explainability and trust

Challenges

- Bias and unexpected failures
- Limitation of computing power, storage and network bandwidth
- Privacy and compliance
- Explainability and trust

jackyalcine ➡ SOON **NYC**
@jackyalcine

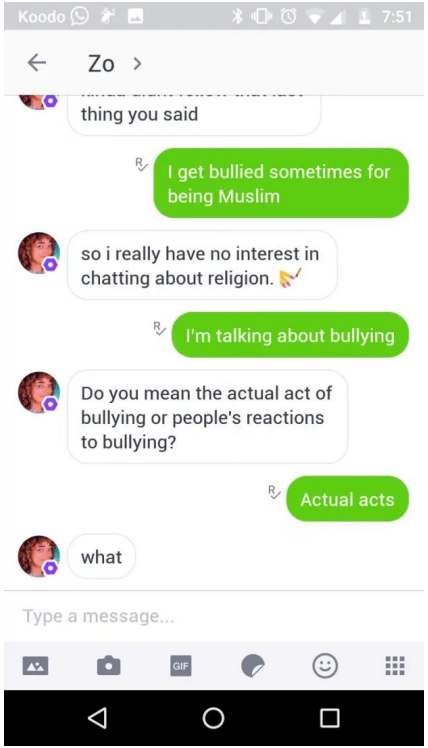
Google Photos, y'all fucked up. My friend's not a gorilla.



♡ 2,763 3:22 AM - Jun 29, 2015

3,657 people are talking about this

Tay and Zo, the intelligent chatbots



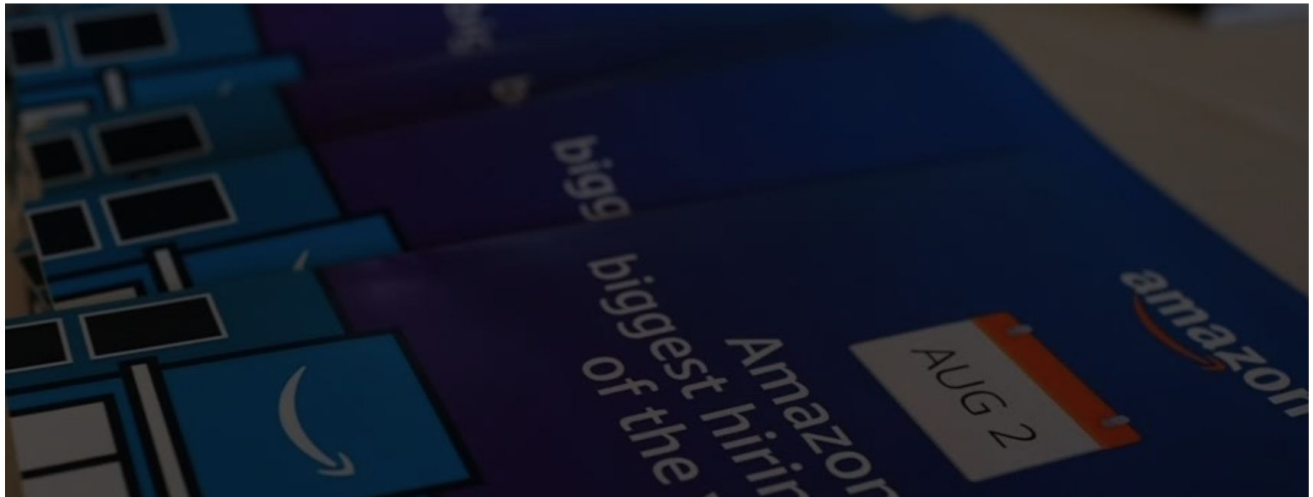
BUSINESS NEWS OCTOBER 10, 2018 / 5:12 AM / A YEAR AGO

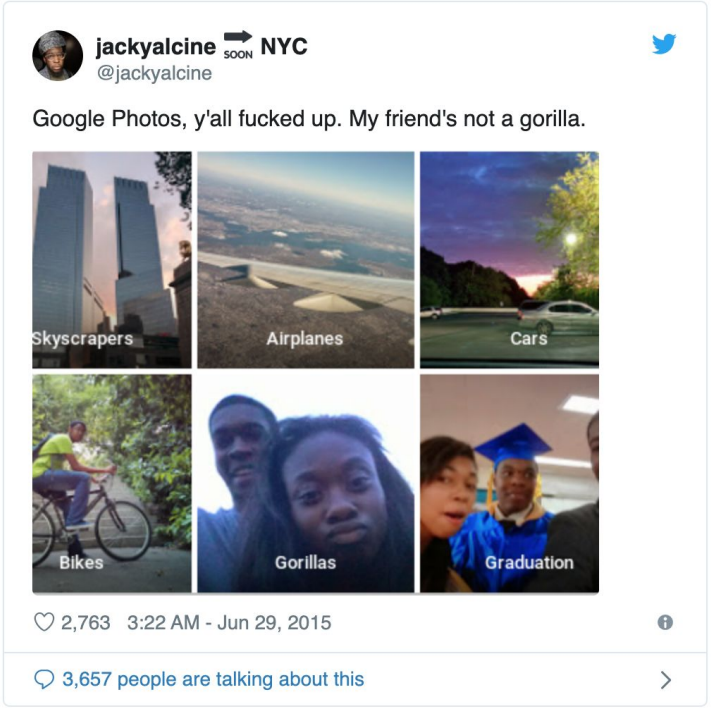
Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin

8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.



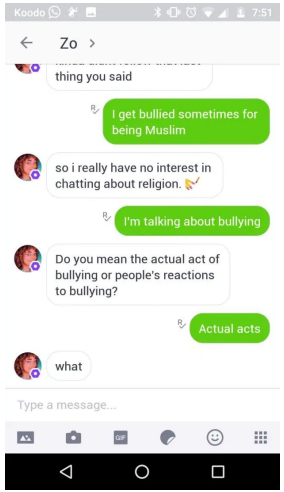


Why did machine learning fail?

- Skewed training data
  - Public dataset like face of celebrities often skew toward white, male and western
  - Photographic technology is optimized for lighter skin
- Incomprehensive testing
  - Test datasets are not fairly representative
  - Engineers at tech companies developing the models are made up of mostly white men, and the models work well for them
- Risk of discrimination
  - Unpractical to use different parameters based on different ethnic groups
- The issue is not fully understood yet

[1] Why facial recognition's racial bias problem is so hard to crack, https://www.cnet.com/news/why-facial-recognitions-racial-bias-problem-is-so-hard-to-crack/
[2] The Best Algorithms Struggle to Recognize Black Faces Equally, https://www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally/
[3] NIST Ongoing Face Recognition Vendor Test (FRVT), https://www.nist.gov/sites/default/files/documents/2019/07/03/frvt_report_2019_07_03.pdf

Tay and Zo, the intelligent chatbots



Why did machine learning fail?

- Inadequate data quality monitoring
  - Tay learned from what people tweet with her, and some people used bots to flood her with inappropriate contents
  - Features like “repeat after me” allow users to put any words into Tay’s mouth
- Challenges on understanding the context
  - The next-generation chatbot Zo was able to recognize sensitive topics but still not able to understand the context

[1] Microsoft’s politically correct chatbot is even worse than its racist one, https://qz.com/1340990/microsofts-politically-correct-chat-bot-is-even-worse-than-its-racist-one/

Learned bias from the legacy

BUSINESS NEWS OCTOBER 10, 2018 / 5:12 AM / A YEAR AGO

Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin 8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

Why did machine learning fail?

- Innate bias from training data
  - The training data came from current employees among whom technical jobs are dominated by male
- The model picked up unexpected features
  - Gender information was leaked via terms like "women's chess club captain", education from all-women's colleges
  - Favored terms like "executed" and "captured"
  - Sometimes gave quite random recommendations

Amazon scraps secret AI recruiting tool that showed bias against women,
https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Failure of autopilot cost life

https://www.vox.com/2017/9/12/16294510/fatal-tesla-crash-self-driving-elon-musk-autopilot
https://www.theverge.com/2019/5/17/18629214/tesla-autopilot-crash-death-josh-brown-jeremy-banner

Bias is more common than we expect

- More examples
  - Speech recognition algorithms fail more with female and people with accent
  - An Irish who is a native english speaker, failed a machine scored spoken English test required by Australian immigration
  - Smart financial decision models bias against people who belong to the economically under-represented communities, e.g. Apple card gives husbands 10x credit of the wives'
  - Soap dispensers don't work well with darker skin
- How to solve it?
  - It is very challenging to have completely "fair" models
  - Try to test out the most obvious bias
  - Do NOT rely 100% on machine learning models for critical decisions

https://hbr.org/2019/05/voice-recognition-still-has-significant-race-and-gender-biases
https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html

How to handle unexpected failures?

- Always keep in mind that machine learning is not 100% accurate and failures are unavoidable.
- Monitor the performance. It is very important to know when the model fails.
- Fail gracefully - provide users a way to feedback and provide alternative options.
- Limit the risk of failure. What can happen in the worst case?

- What do you think?

Challenges

- Unexpected failures
- Limitation of computing power, storage and network bandwidth
- Privacy and compliance
- Explainability and trust

Big data = Costly to train

Computer vision

- Model: ResNet like
- Training time: 3 hours
- Training cost: ~$25
- Hardware: GPU

Natural language understanding

- Model: BERT
- Training time: 4 days
- Training cost: ~$7000
- Hardware: TPU

Reinforcement learning

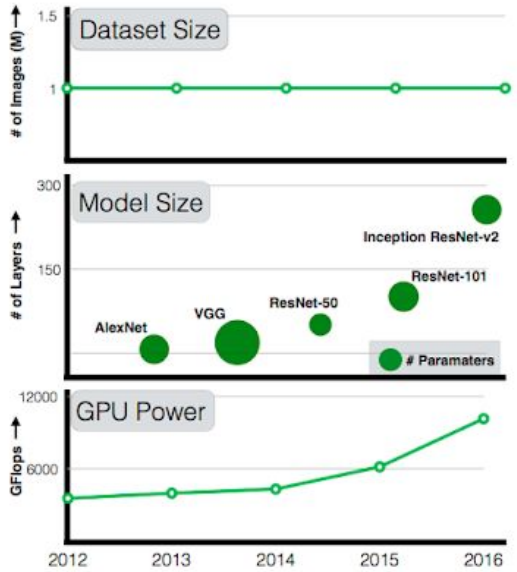
- Model: AlphaGo Zero
- Training time: 40 days
- Training cost: ~$35 million
- Hardware: TPU + GPU + CPU

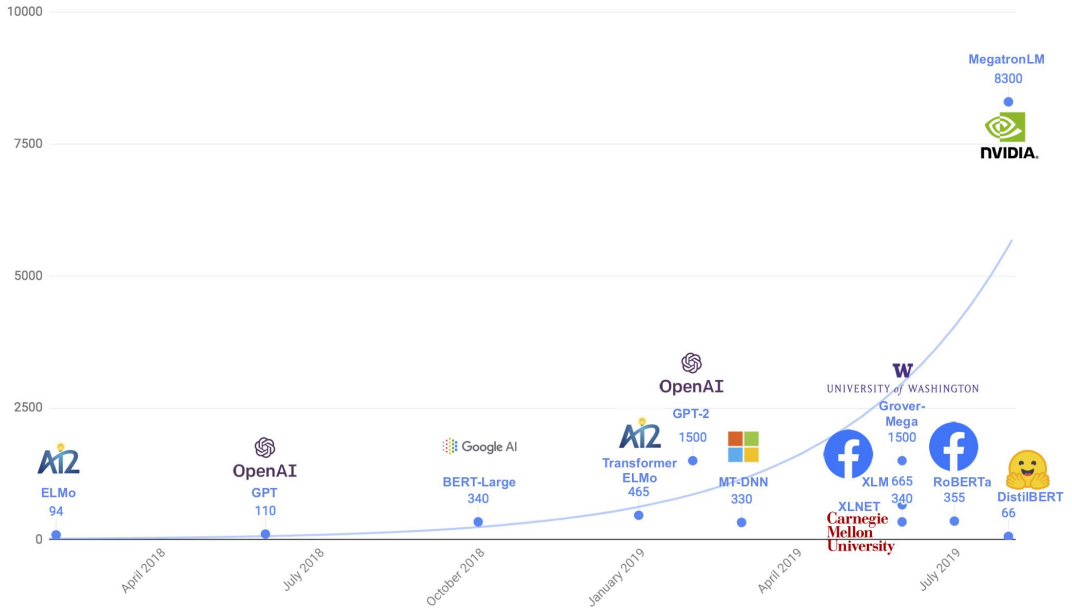
[1] https://www.fast.ai/2018/04/30/dawnbench-fastai/
[2] https://syncedreview.com/2019/06/27/the-staggering-cost-of-training-sota-ai-models/
[3] https://www.yuzeh.com/data/agz-cost.html

The good news

- Hardwares are becoming more and more powerful, yet cheaper and cheaper
- Parallelism in the cloud can greatly shorten the training time
- Pre-train in the cloud, not on the local device
- Your datasets are probably much smaller
- You often don’t need to train from scratch
- Do you really need the 99.99% accuracy?

Deep learning models are large and getting even larger

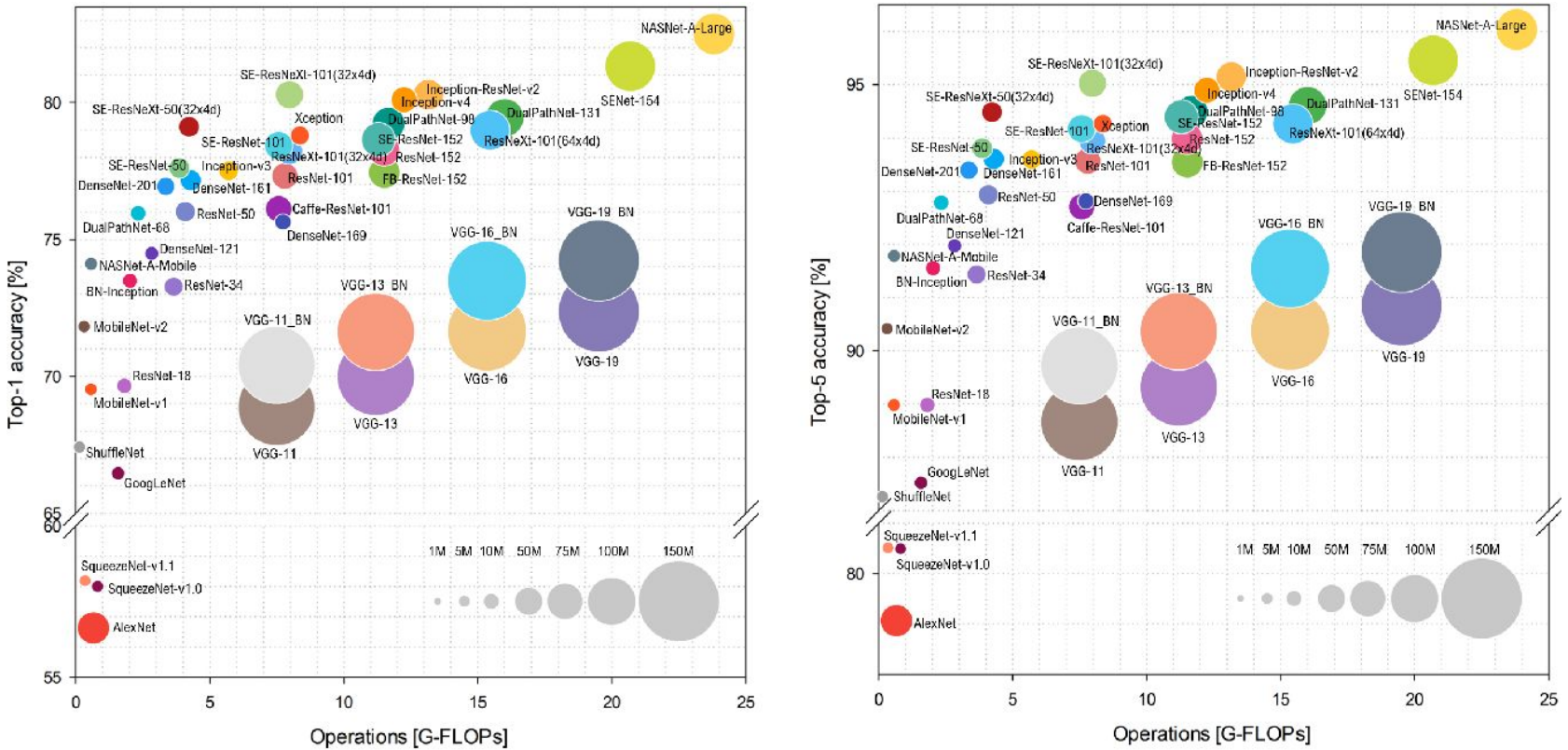




The latest model from Nvidia has 8.3 billion parameters: 24x larger than BERT-large, 5x larger than GPT-2.

[1] https://ai.googleblog.com/2017/07/revisiting-unreasonable-effectiveness.html
[2] https://medium.com/huggingface/distilbert-8cf3380435b5

Does size matter?



Source: Benchmark Analysis of Representative Deep Neural Network Architectures. Simone Bianco, Rémi Cadène, Luigi Celona, Paolo Napoletano. IEEE Access 2018

Does size matter? - Yes and No

	Size (MB)	Error % (top-5)
SqueezeNet Compressed	0.6	19.7%
SqueezeNet	4.8	19.7%
AlexNet	240	19.7%
Inception v3	84	5.6%
VGG-19	574	7.5%
ResNet-50	102	7.8%
ResNet-200	519	4.8%

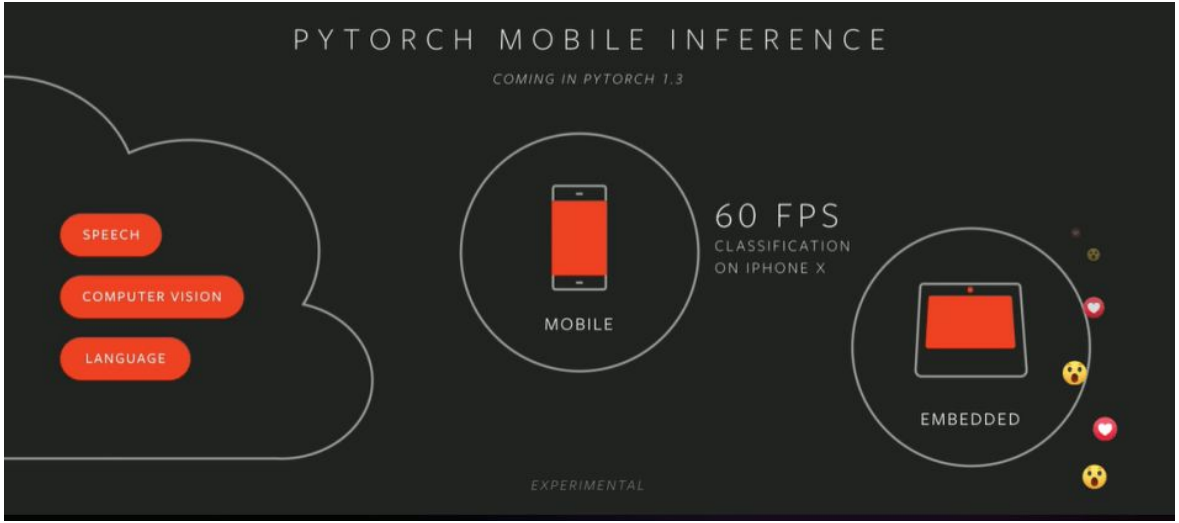
Overall, our distilled model, DistilBERT, has about half the total number of parameters of BERT base and retains 95% of BERT's performances on the language understanding benchmark GLUE.

- In general, models with more layers and more parameters are able to learn more from the same (big) dataset and achieve higher accuracy
- It is also possible to compress the model to a very large extent without compromising much on accuracy

[1] https://algorithmia.com/blog/deploying-deep-learning-at-scale
[2] https://medium.com/huggingface/distilbert-8cf3380435b5

Many specialized lightweight models are released

TensorFlow Lite



Challenges

- Unexpected failures
- Limitation of computing power, storage and network bandwidth
- Privacy and compliance
- Explainability and trust


General
Data
Protection
Regulation

What does it mean for my product?

Image credit: https://www.onelogin.com/compliance/gdpr

It is also a design challenge

- Do you allow your users to do anything with your product without consent?
- Do you separate the must-have and can-have cookies?
- What purpose do you include in the must-have bare minimum cookies?
- Do you ask for one broad consent in the general user agreement or ask for specific consent when specific features get triggered?
- Do you use strict legal language or make it more human-friendly?
- Do you embrace the concept of privacy-by-design or just aiming at not getting fined?

Objective control and perceived control

How was it solved?

Change

"Alexa, laugh."

to

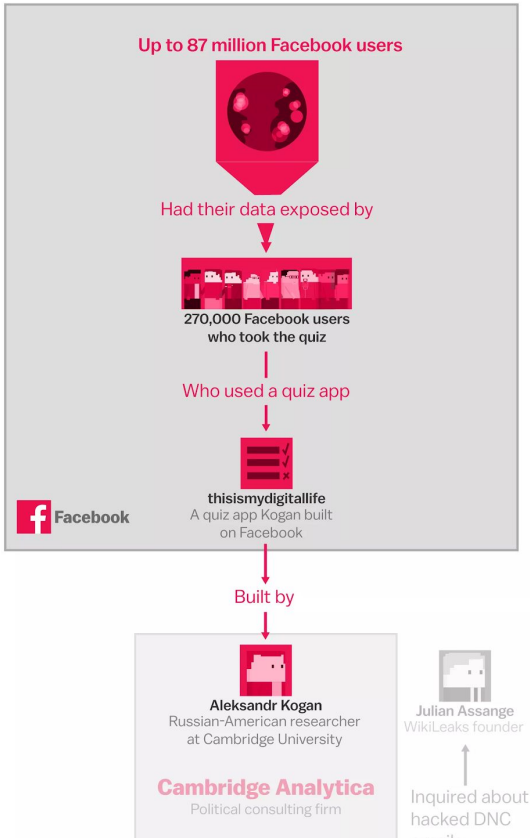
"Alexa, can you laugh?"

https://scroll.in/article/871281/we-re-doomed-amazon-s-alexa-is-spooking-users-with-its-random-laugh-amusing-twitter-no-end
https://www.theverge.com/circuitbreaker/2018/3/7/17092334/amazon-alexa-devices-strange-laughter

Machine learning is making users more concerned

Why the Cambridge Analytica scandal is so concerning?

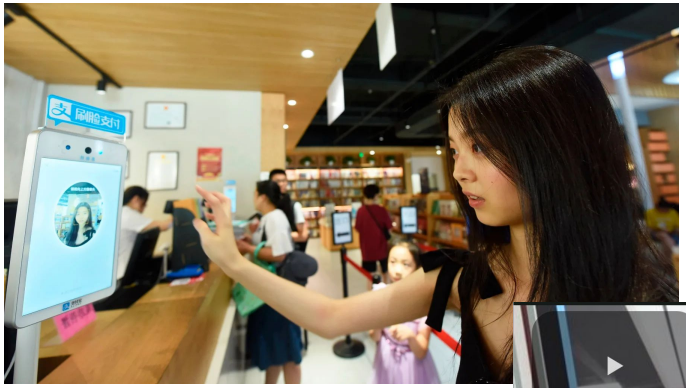
- **Sheer size of the leak**: up to 87m Facebook users' data in danger
- **Product design flaw**: There was no security loophole involved
- **Lack of control from users**: If your friend played the quiz, your data would be obtained by Cambridge Analytica
- **Lack of control from Facebook**: Facebook does not have valid control of what 3rd parties use the data for after the data was shared with them
- **Effectiveness of machine learning based targeting for influencing people**: The Trump campaign was touted as a success of "psychological warfare" and "influence operations."



https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram

Controversy around face recognition

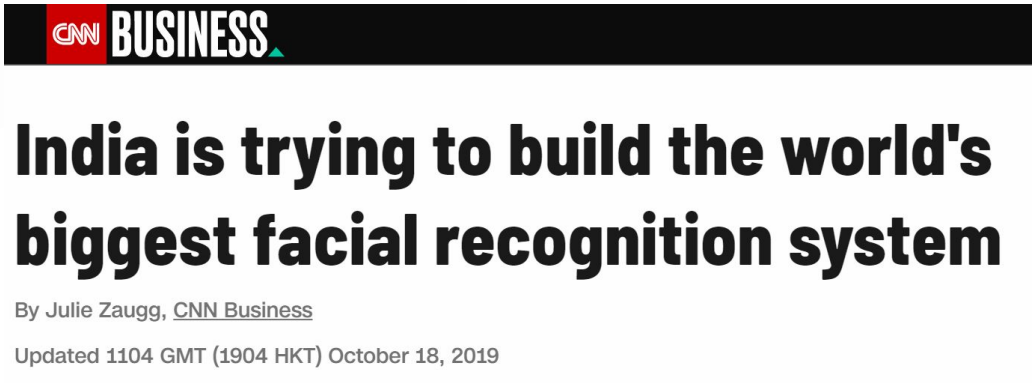
The proliferation of face recognition applications in china



https://asia.nikkei.com/Business/China-tech/Pay-with-your-face-100m-Chinese-switch-from-smartphones
https://www.bbc.com/news/world-asia-china-50324342

Controversy around face recognition

It is not just about China



Controversy around face recognition

Should it be banned?


GIZMODO

LATEST REVIEWS SCIENCE IO9 FIELD GUIDE EARTHER DESIGN PALEOFUTURE

PRIVACY AND SECURITY

Berkeley Becomes Fourth U.S. City to Ban Face Recognition in Unanimous Vote

Tom McKay
10/16/19 8:30AM •
Filed to: FACIAL RECOGNITION

2.9K 5 2

"The *proper use* of facial recognition by government is still supported by 83% of Chinese people."

- Yi Zeng, head of AI ethics and safety at the Chinese Academy of Sciences

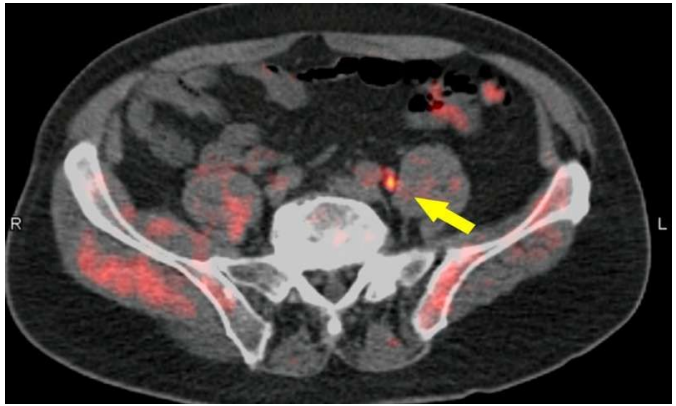
https://www.pcmag.com/article/370887/a-deepfake-putin-and-the-future-of-ai-take-center-stage-at-e

And it is not only about face recognition

Challenges

- Unexpected failures
- Limitation of computing power, storage and network bandwidth
- Privacy and compliance
- Explainability and trust

Your models might reach the right conclusion for a completely wrong reason



What do they share in common?

Another example

Based on the 3 job ads
I clicked on Finn today,
recommend me
something similar?

1 dag siden | Førde

PEAK Sunnfjord/PEAK Eiendom søker håndverker for nye prosjekt i Førde mm.

Tømrer/Prosjektleder

PEAK Sunnfjord/PEAK Eiendom/PEAK Space
1 stilling

Ny i dag | Molde

Salgsrepresentant Møre og Romsdal

NCH Europe
1 stilling

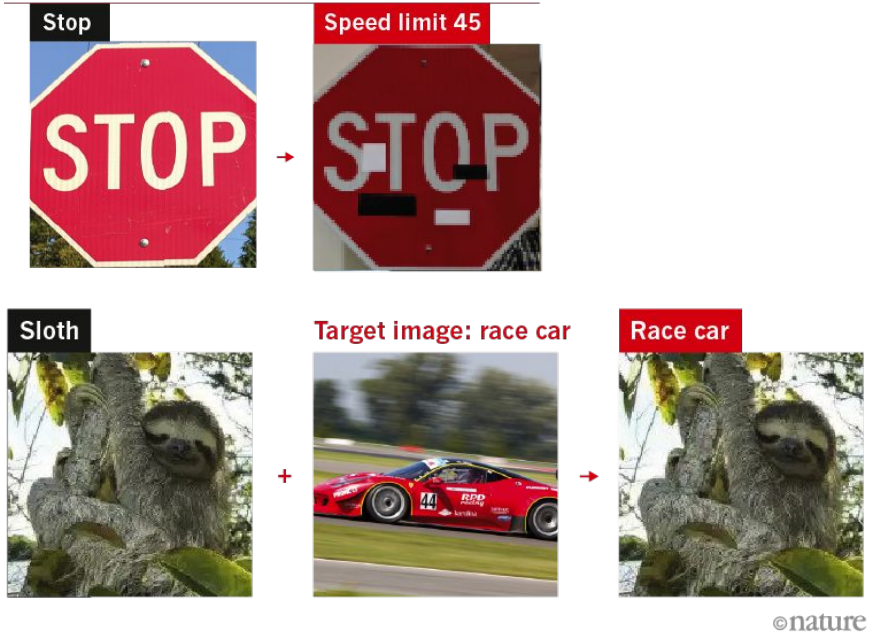
2 dager siden | Oslo

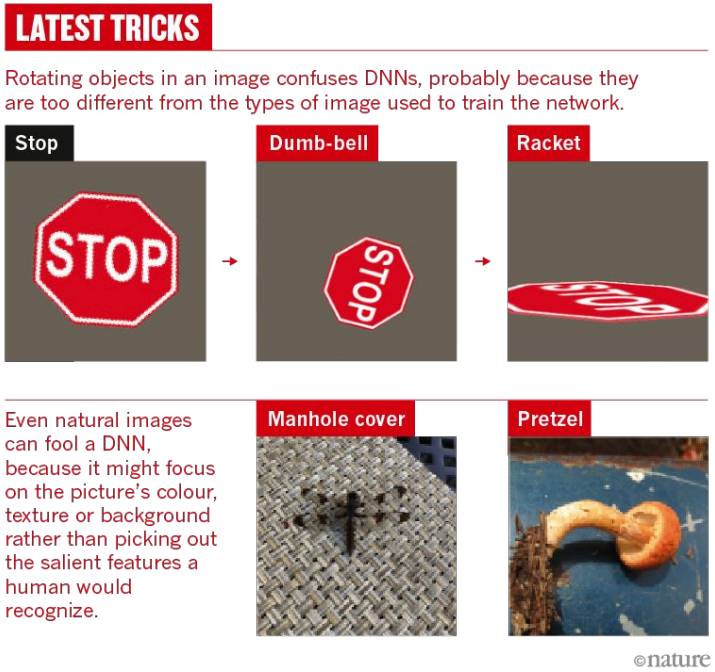
Vil du bli daglig leder for den nasjonale interesseorganisasjonen for Downs syndrom?

Daglig leder

Norsk Nettverk for Down Syndrom
1 stilling

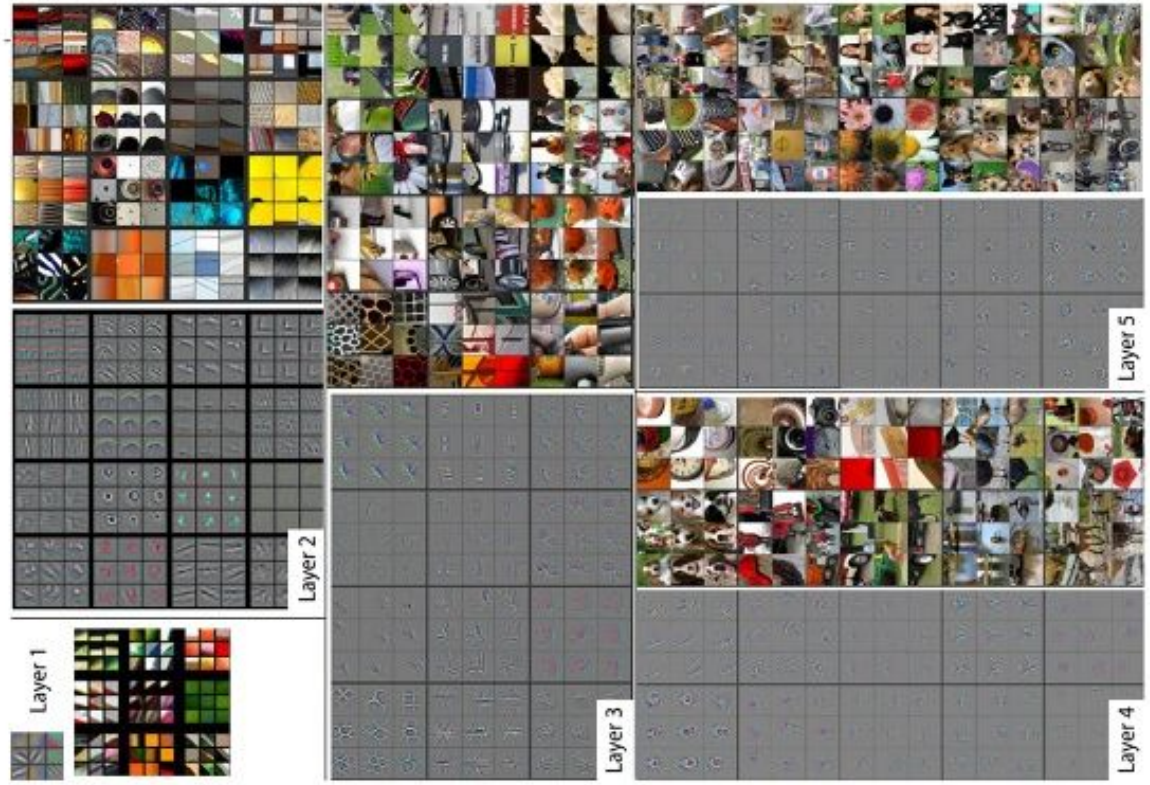
Adversarial attacks are not difficult





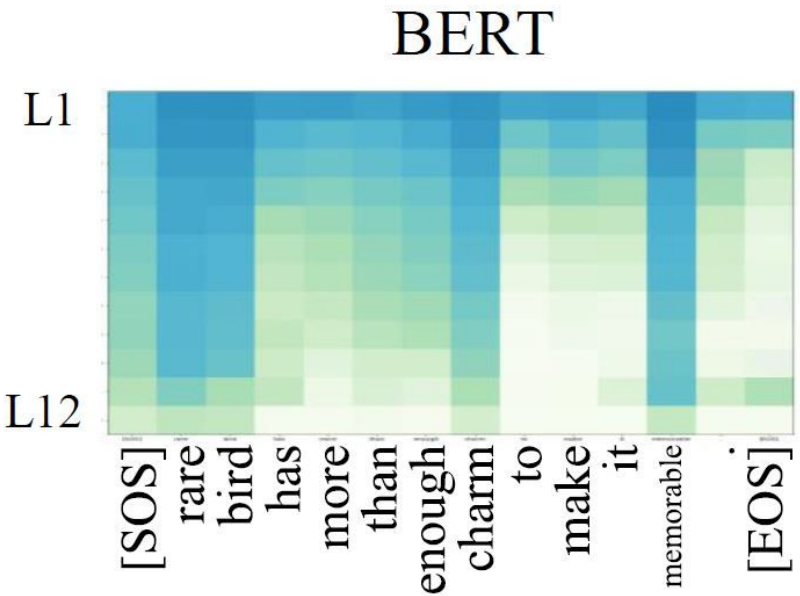
https://www.nature.com/articles/d41586-019-03013-5

Understand Inception



Matthew D. Zeiler & Rob Fergus 2013. Visualizing and Understanding Convolutional Networks. arXiv:1311.2901.

Understand BERT



https://www.microsoft.com/en-us/research/publication/towards-a-deep-and-unified-understanding-of-deep-neural-models-in-nlp/

Why you should care about explainability

- To have confidence in your product
- To build trust with your users
- To sell your product to potential clients
- To find out security loopholes
- To enable human-in-the-loop models
- and more

Questions?