Deep Learning

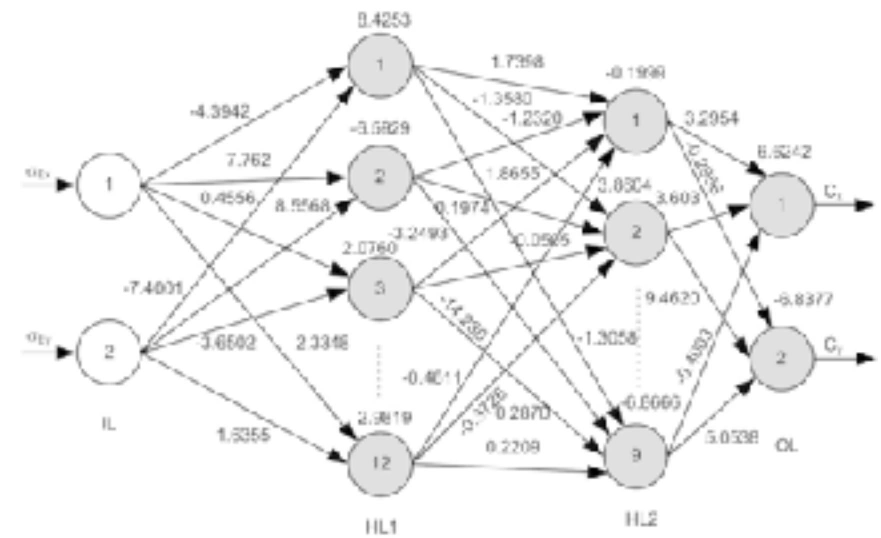
Didier Guillevic
didier.guillevic.net

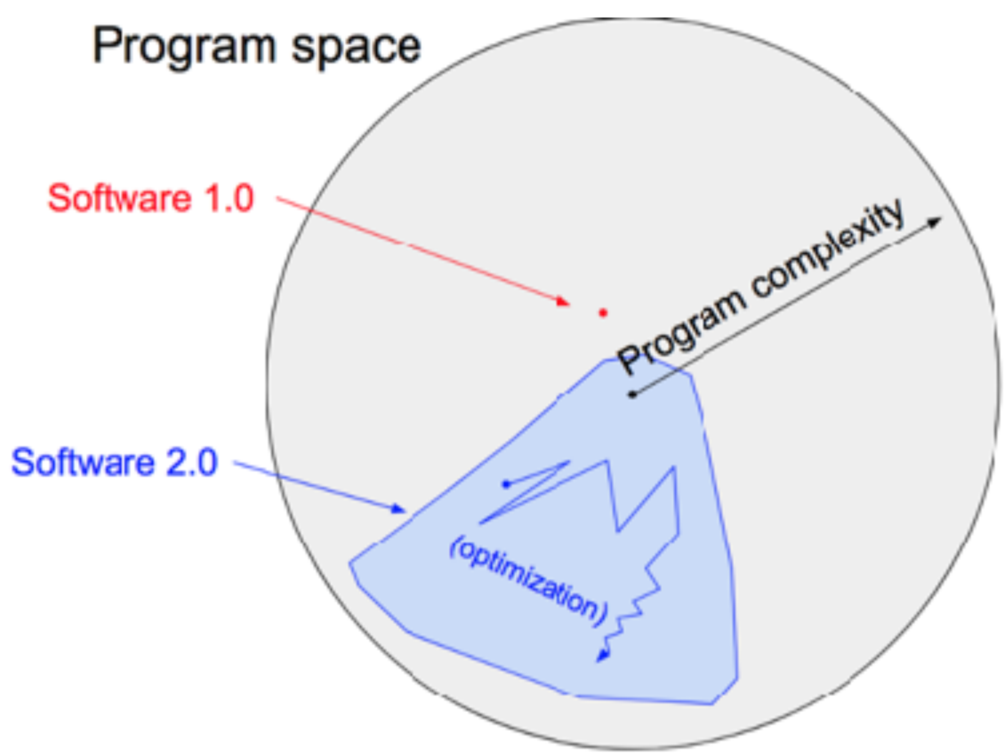
Deep Learning - Software 2.0

Andrej Karpathy (2017-11)

Software 1.0

Software 2.0





Software 2.0 - Applications

Andrej Karpathy (2017-11)

Domain	Software 1.0	Software 2.0
Visual Recognition	Engineered features with a bit of machine learning	Searching the space of Convolutional Neural Network architectures
Speech Recognition	Preprocessing, gaussian mixture models, hidden markov models	Almost entirely neural networks
Speech Synthesis	Various sticking mechanisms	Large Convolutional Neural Networks (e.g. WaveNet)
Machine Translation	Phrase-based statistical techniques	Neural networks (supervised and unsupervised)
Games	Hand-coded Go playing programs	AlphaGo Zero: only uses the game's set of rules and learns strategies by itself
Databases	Indices using B-trees or Hash indexes	Coming: Learned Index Structures (using deep learning models) - https://arxiv.org/abs/1712.01208

Software 2.0 - Benefits

Andrej Karpathy (2017-11)

Software 1.0: production-level C++ code base

Software 2.0: Convolutional Neural Network

Homogeneous computation	Software 1.0: instruction set of classical software (heterogenous and complex) Software 2.0: matrix multiplication (and thresholding at zero)
Simple to bake in silicon	Easier to make custom ASICs Small inexpensive chips could come with a pre-trained ConvNet
Constant running time	C++ code could have unintended infinite loop Forward pass of a neural network takes exactly the same amount of FLOPS Zero variability
Constant memory use	No dynamically allocated memory as in C++ Little possibility of memory leaks
Highly portable	Sequence of matrix multiplies is easy to run-on arbitrary computational configurations (not the same for binaries)
Very agile	C++: non-trivial to make a system run twice as fast Sofware 2.0: remove half of the nodes and re-train to get twice the speed OR make program work better by adding more channels (or averaging models)
Melding modules together	Software 1.0: communicates through public functions, APIs, … Software 2.0: two modules trained separately can be connected and the weights adjusted by back-propagation

Software 2.0 - Limitations

Andrej Karpathy (2017-11)

Explainability

- Hard to tell how a large network works
- Choice of using:
  - a 90% accurate model that we understand OR
  - a 99% accurate model that we don't understand

Can fail in unintuitive and embarrassing ways

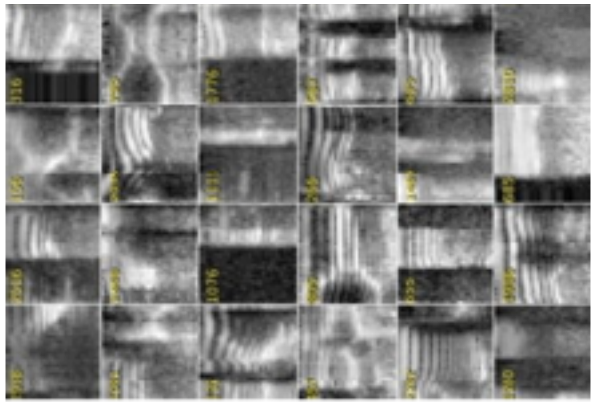
- Silently failing: adopting biases from the training data
  - Kaggle / Google: Inclusive image challenge 2018-09
- Adversarial examples, attacks

DARTS: Deceiving Autonomous Cars with Toxic Signs - Chawin Sitawarin et al. 2018

Deep Learning: State of the Art

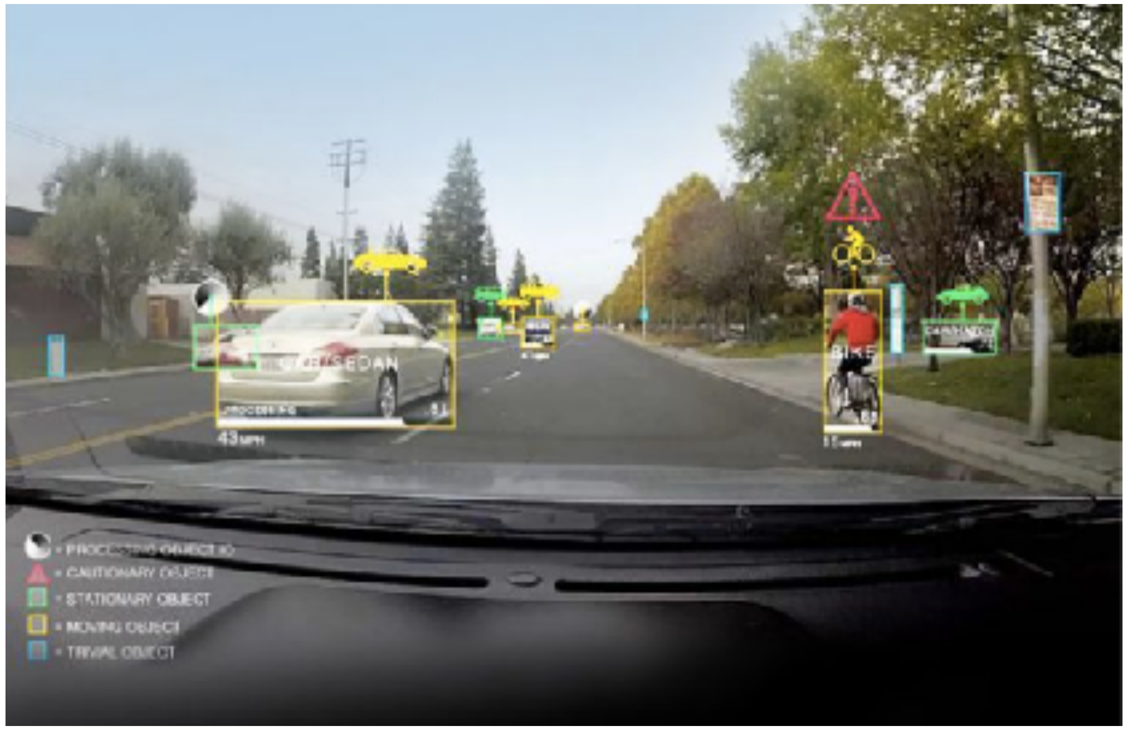
1. Speech processing

e.g. speech recognition, text to speech

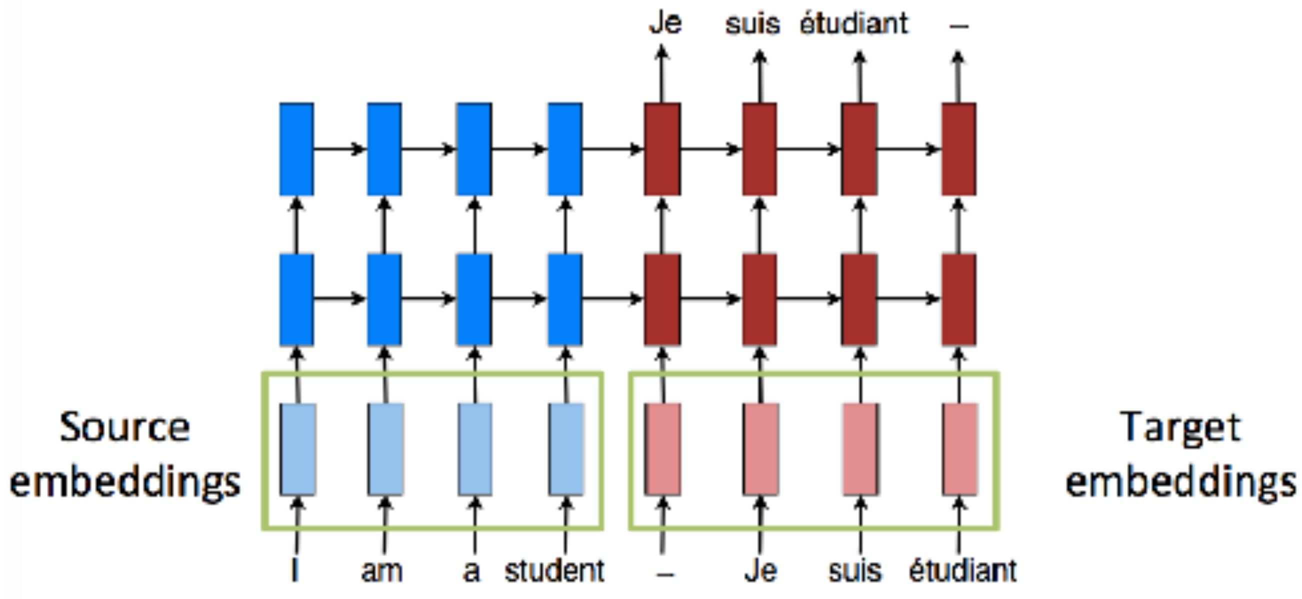


2. Image / video processing

e.g. self-driving cars



3. Natural Language Processing



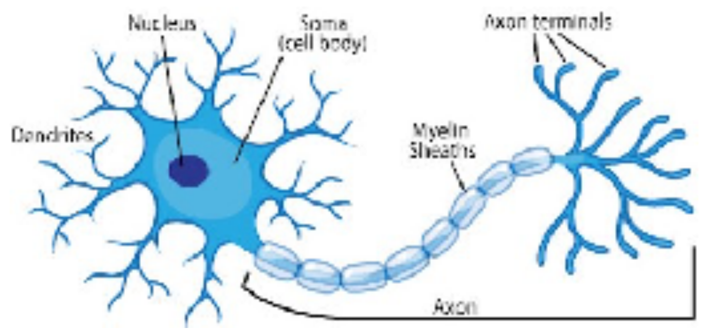
[tensorflow.org]

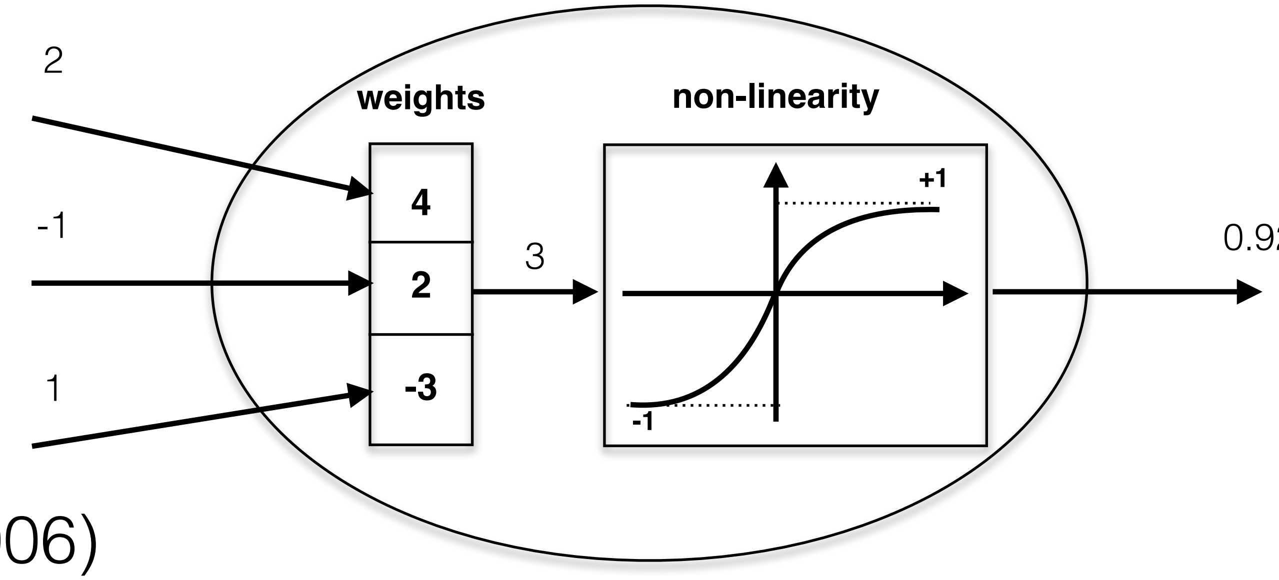
e.g. Neural Machine Translation

The Great A.I. Awakening (NY Times 2016)

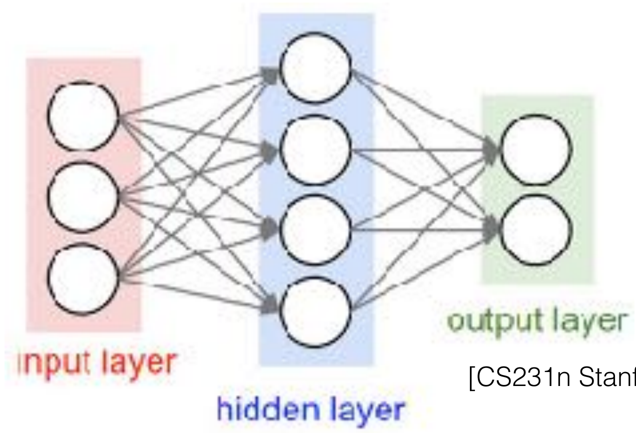
Deep Neural Networks

Neuron



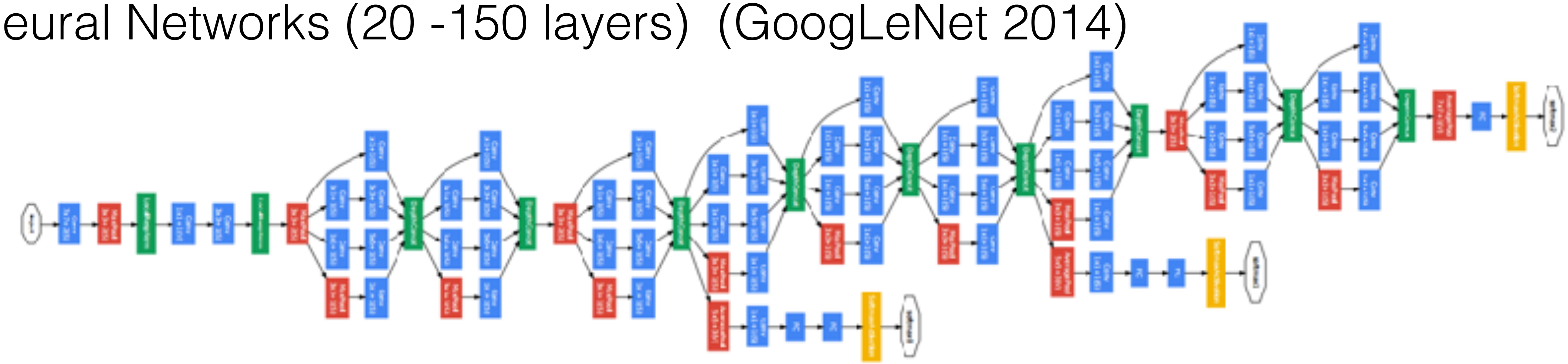


Neural Networks (2 - 5 layers) (until 2006)



[CS231n Stanford]

Deep Neural Networks (20 -150 layers) (GoogLeNet 2014)

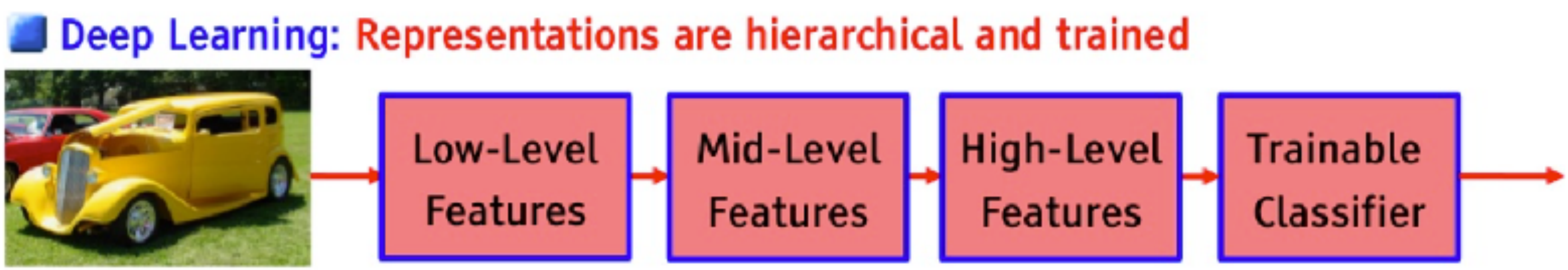


Change in paradigm

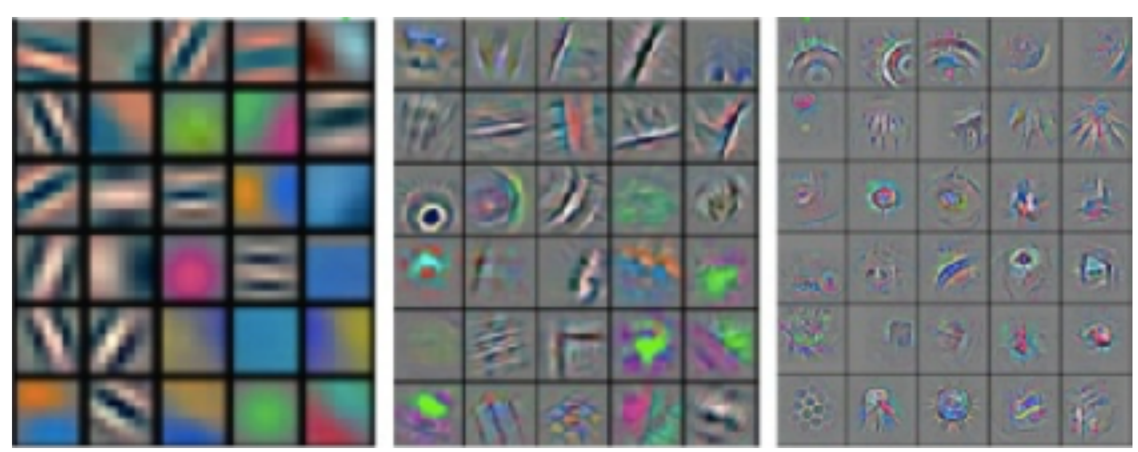
Before: Handcrafting features: domain experts: 10s of years

[Yann Le Cun]

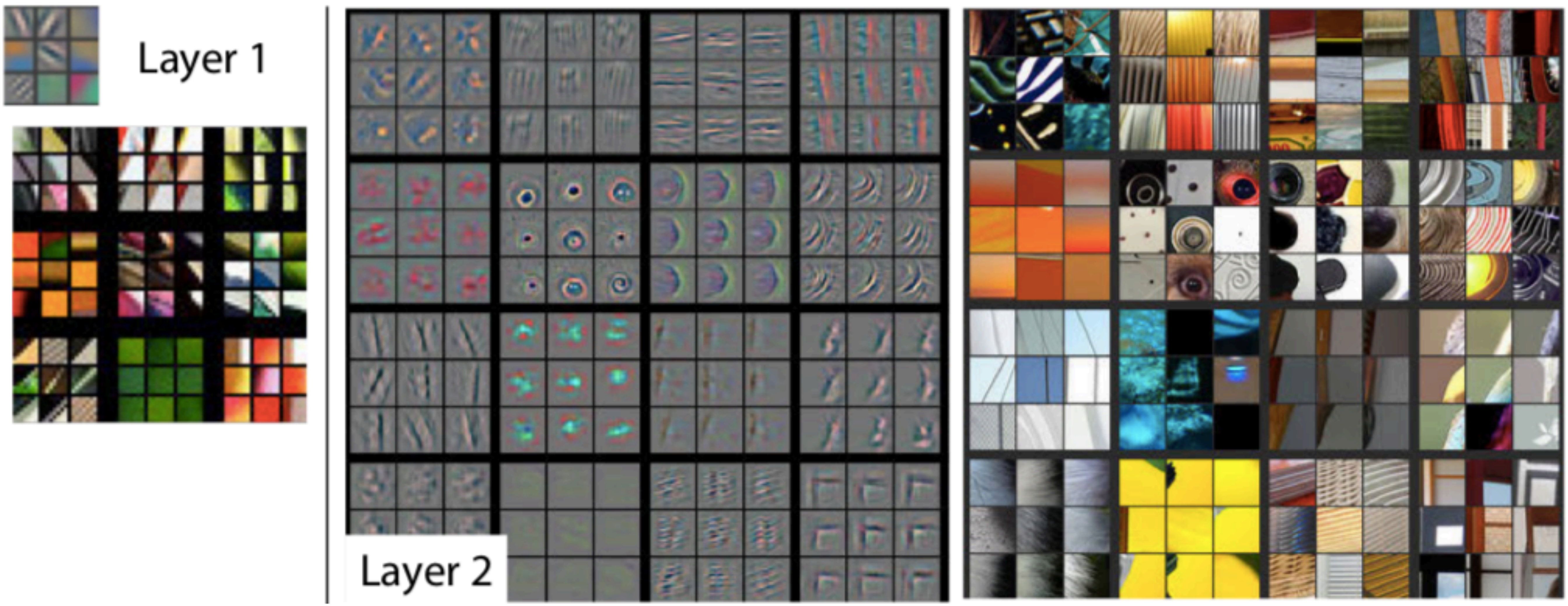
Now: Learned features: End-to-end Learning



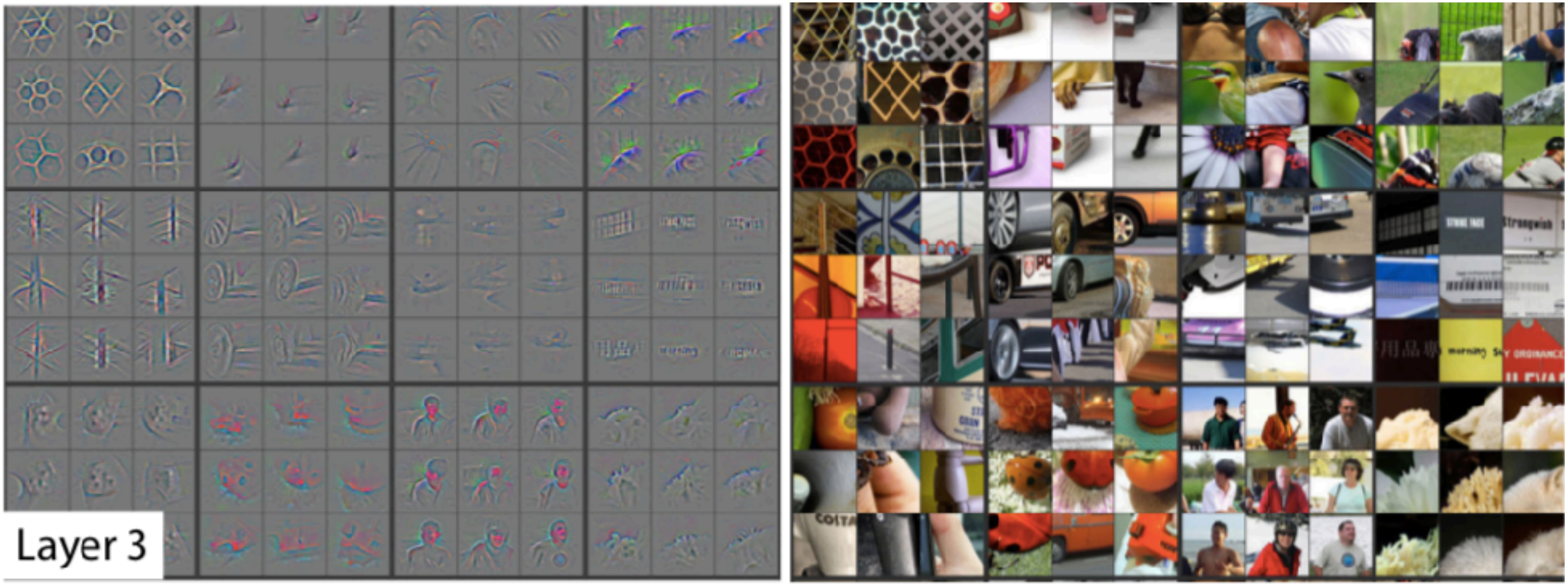
[Yann Le Cun]



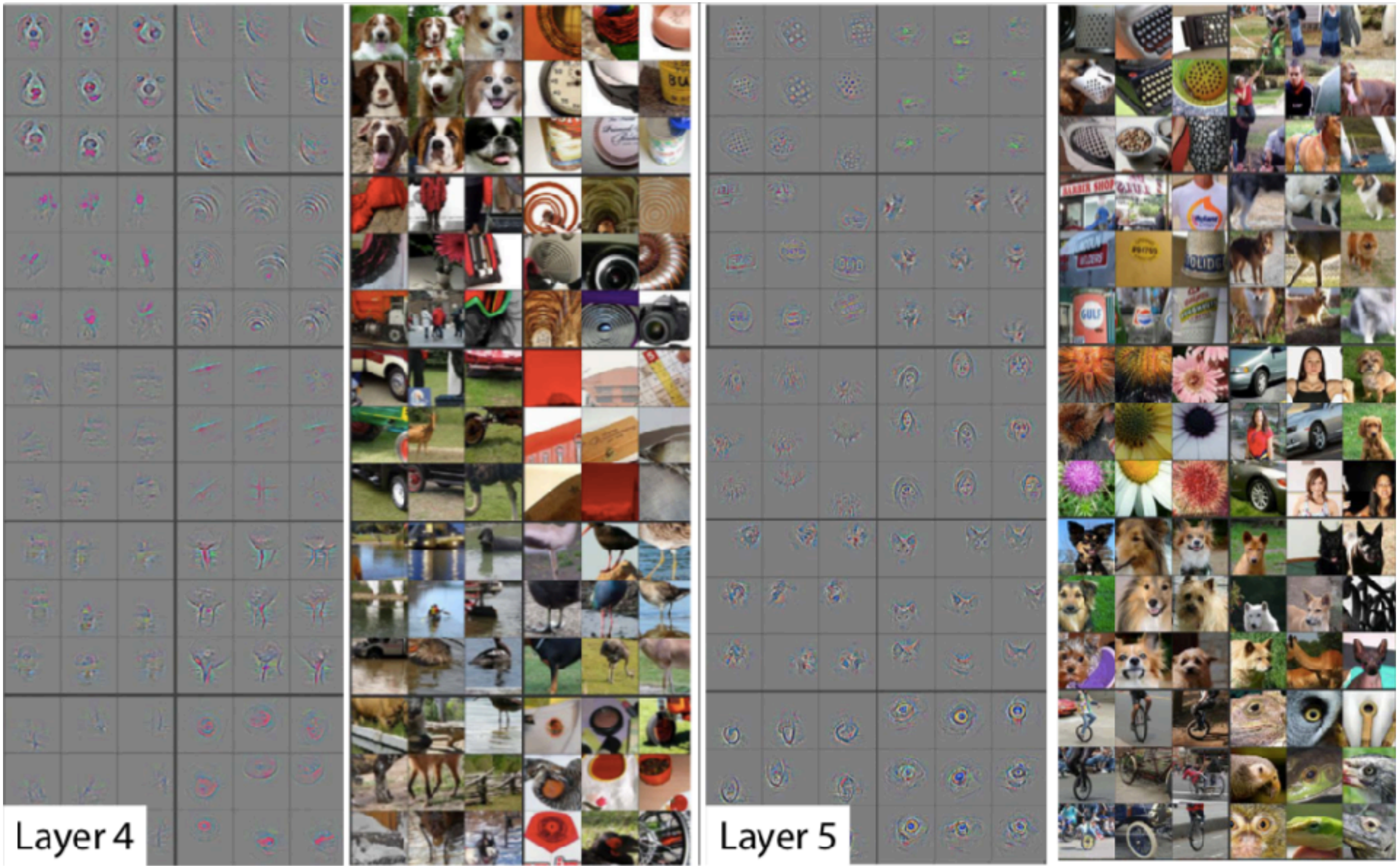
[Zeiler & Fergus 2013]



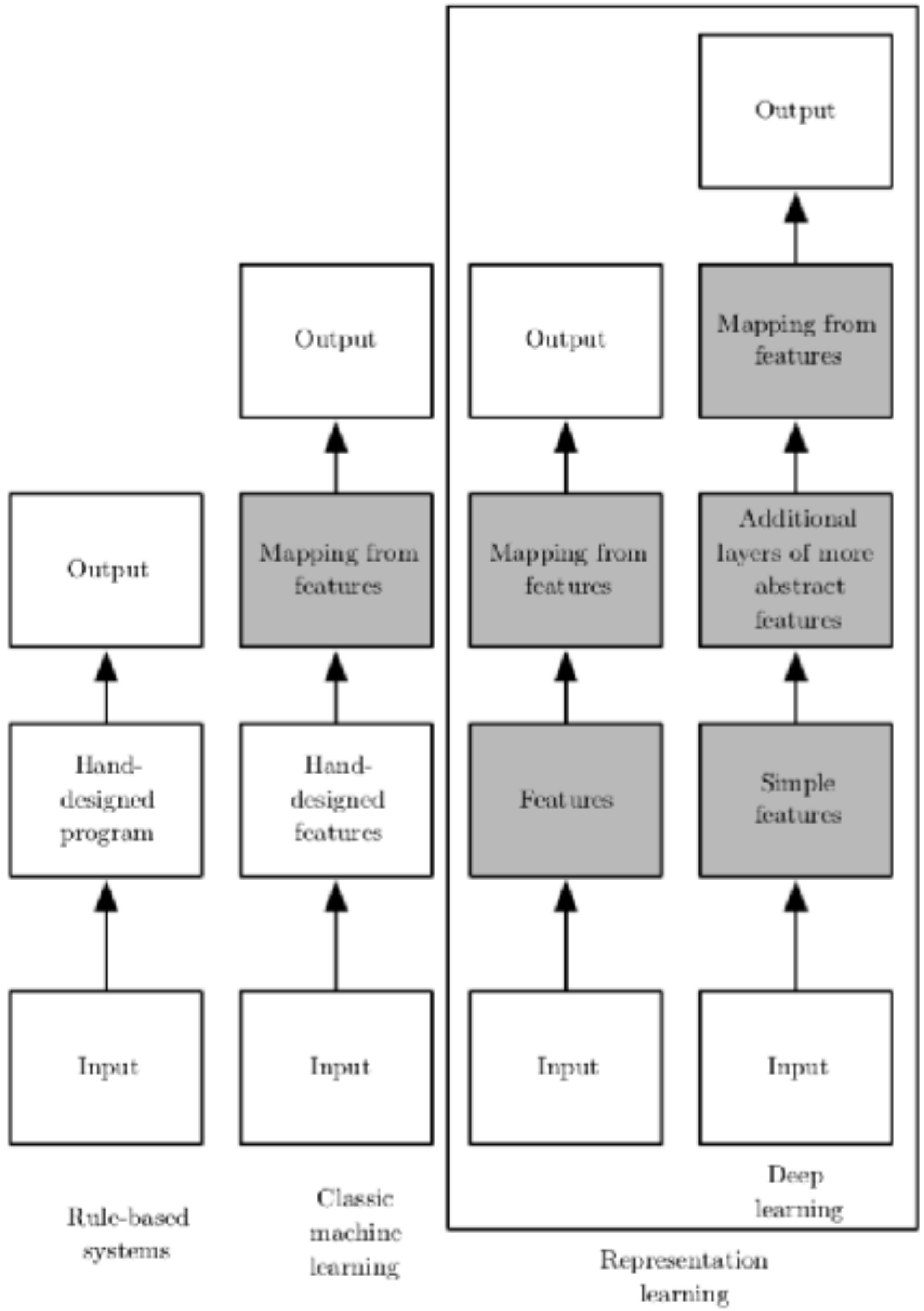
[Zeiler & Fergus 2013]



[Zeiler & Fergus 2013]



[Zeiler & Fergus 2013]



Goodfellow et al. - Deep Learning - 2016

Why now? Why not sooner?

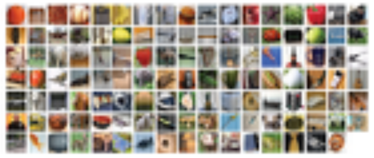
1. A lot more (labeled) data

2. A lot more computing power

3. Knowledge on how to train deep networks

1. A lot more (labeled) data

Data

IMAGENET  14 millions images
21000 labels (WordNet)

Wikipedia 5 millions articles in the English Wikipedia

Common Crawl 2 billions web pages

Labels

amazon mechanicalturk **Low-cost** global, 24x7 workforce

reCAPTCHA **Free** global, 24x7 workforce

2. A lot more computing power

CPUs: 20 cores

GPUs: 4000 cores

Big Sur

aws

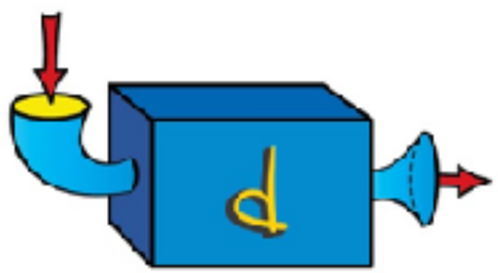
Distributed Computing

Platforms

TensorFlow PYTORCH Keras

3. Knowledge on how to train deep networks

Learning uses Gradient - Derivative



[Robert Ghrist - Calculus - U. of Penn]

$$\frac{\texttt{change in output}}{\texttt{change in input}} = \frac{\mathbf{dy}}{\mathbf{dx}} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{(x+h) - x} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

$$f(x+h) = f(x) + \mathbf{d}h$$

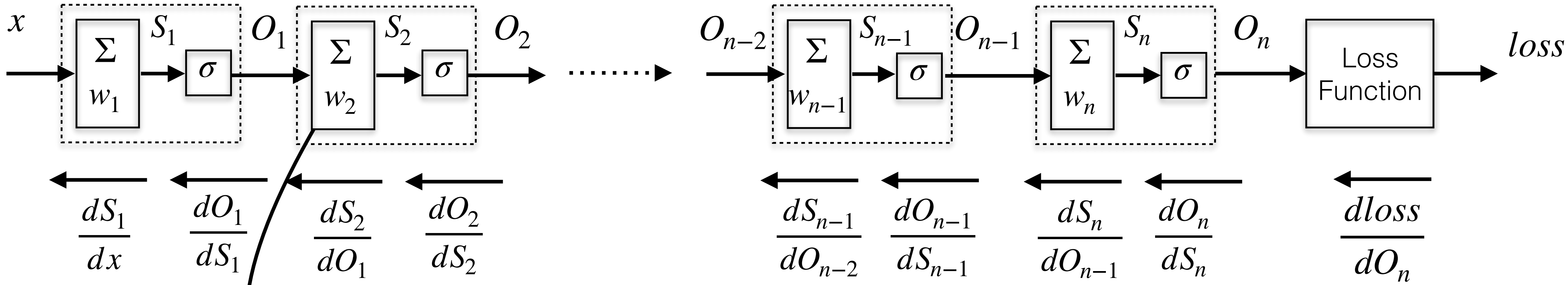
$$\xrightarrow{x} \boxed{\times w} \xrightarrow{y} \quad y = xw$$

$$(\; x + \underline{\mathbf{h}} \;)\, w = xw + \underline{\mathbf{h}}w \longrightarrow \frac{dy}{dx} = w$$

$$x\, (\; w + \underline{\mathbf{h}} \;) = xw + x\underline{\mathbf{h}} \longrightarrow \frac{dy}{dw} = x$$

$$w_{new} = w_{old} - learning\_rate * \frac{dy}{dw}$$

Chain rule of gradients (multiplicative)



$$w_{2new} = w_{2old} - lr * \frac{dloss}{dw2}$$

$$w_{2new} = w_{2old} - lr * \frac{dloss}{dO_n} * \frac{dO_n}{dS_n} * \frac{dS_n}{dO_{n-1}} * \frac{dO_{n-1}}{dS_{n-1}} * \frac{dS_{n-1}}{dO_{n-2}} * \ldots * \frac{dO_2}{dS_2} * O_1$$

$$w_{2new} = w_{2old} - lr * 10 * 10 * 10 * 10 * 10 * \ldots * 10 * 10 \quad \approx w_{2old} - lr * \infty$$ **Exploding** gradient

$$w_{2new} = w_{2old} - lr * 0.10 * 0.10 * 0.10 * \ldots * 0.10 * 0.10 \quad \approx w_{2old} - lr * \mathbf{0.0}$$ **Vanishing** gradient

$\Rightarrow$ We want all terms to be centered around 1.0

Tricks / knowledge to be able to train deep networks

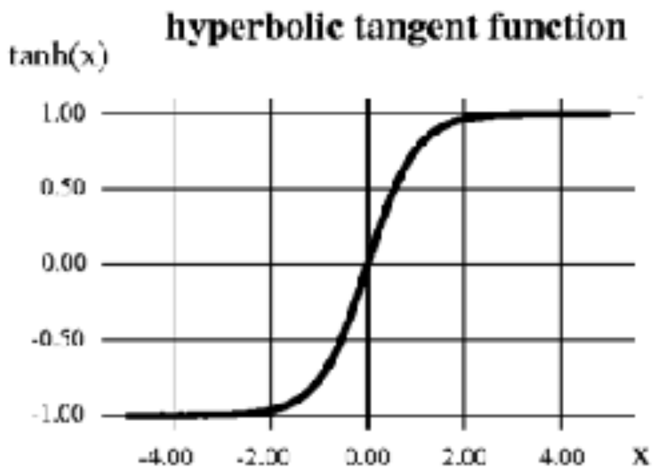
⇒ We want all terms to be centered around 1.0

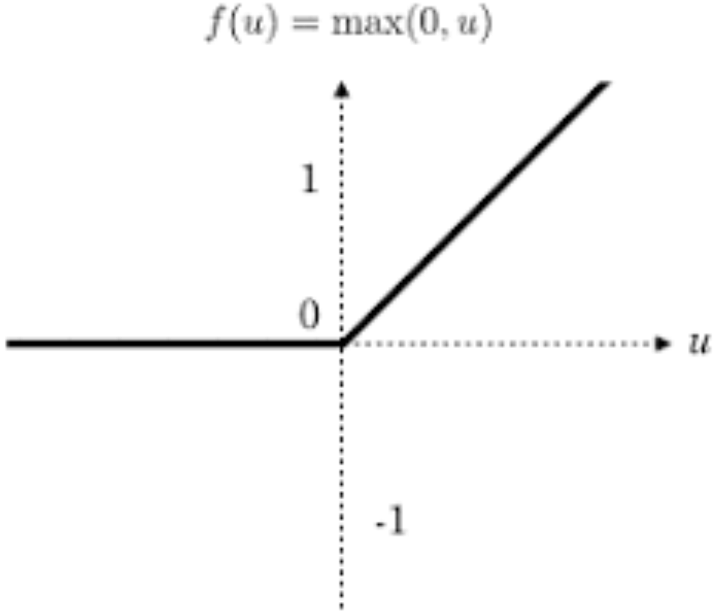
Weight initialization: inversely proportional to number of inputs (e.g. Xavier's initialization scheme)

$$S = w_1 * x_1 + w_2 * x_2 + w_3 * x_3 + \ldots + w_n * x_n$$

BatchNorm: scales inputs at each layer to be centered around 0.0 with variance 1.0 (unit Gaussian)

Non-linearity function: sigmoid have zero gradients —> use ReLU



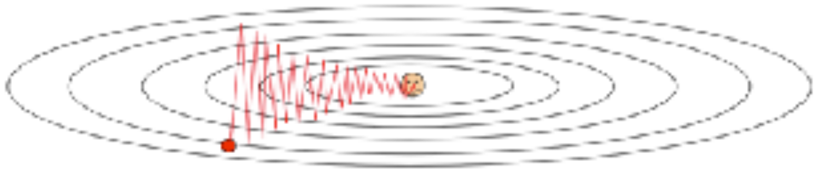


Tricks / knowledge to be able to train deep networks

DropOut: randomly ignore some neurons during training (better generalization & avoids overfitting)

Analogy with diffused versus focused mode of learning (Barbara Oakley - Learning how to learn)
(do not rely on just a few neurons)

Update rule:



CS231n - Stanford

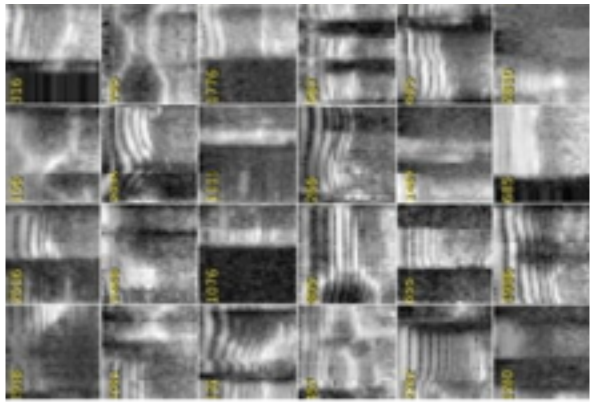
Vanilla update: $w = w - learning\_rate * dw$

Adam update: $w = w - learning\_rate * mb/(sqrt(vb) + eps)$

(Per parameter adaptive learning rate methods with momentum)

Current Application Domains

Speech processing, Image processing, Natural Language Processing



Speech Processing

State of the art voice recognition accuracy (2012)

Text 2 Speech (e.g. DeepMind WaveNet)

Voice modeling; speaker encoding, voice cloning (e.g. Deep Voice - Baidu)

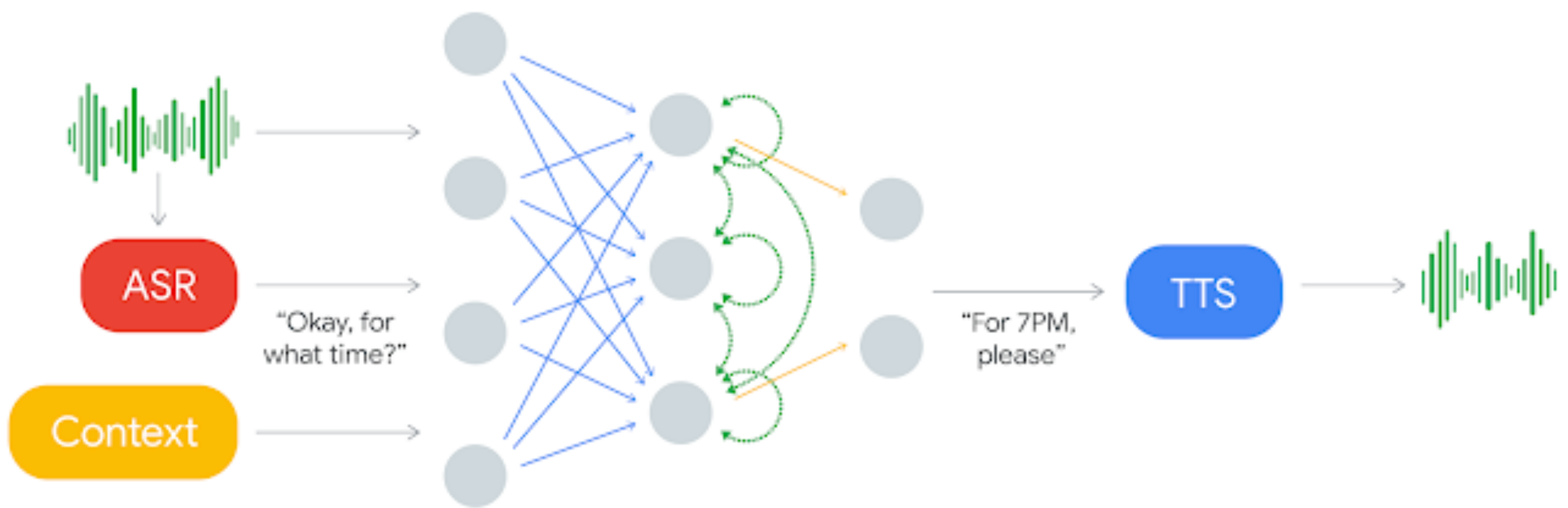
Speech Processing - Quiz

Scheduling a hair salon appointment

Calling a restaurant

Speech Processing - Google Duplex

https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html



- Conducting natural conversations to carry out real world tasks over the phone.
- Completing specific tasks
- Speak normally as you would to another person
- Duplex is constrained to closed domains
- Domains that are sufficiently narrow that they can be explored extensively
- It CANNOT carry out general conversations
- Built to sound natural
- The system calls real businesses to make some appointments

IBM Project Debatter

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

First AI system that can debate humans on complex topics

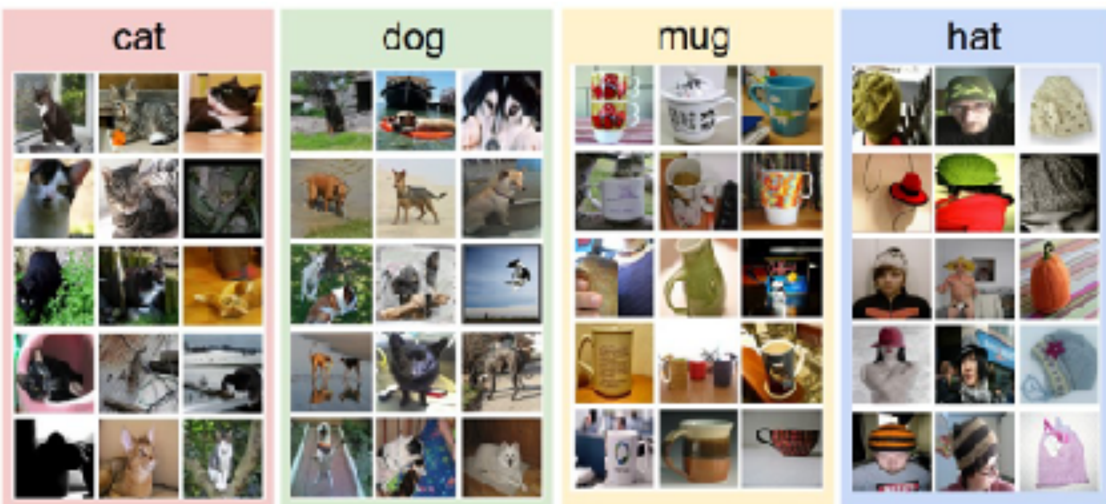
2019-02-11: Project debater versus Harish Natarajan (2012 European champion)

Will help people reason by providing compelling, evidence-based arguments

2019-01-11: Speech By Crowd : E.g. "Gambling should be banned"

Image - Video Processing

ImageNet: Yearly challenge in large scale visual recognition (image-net.org)



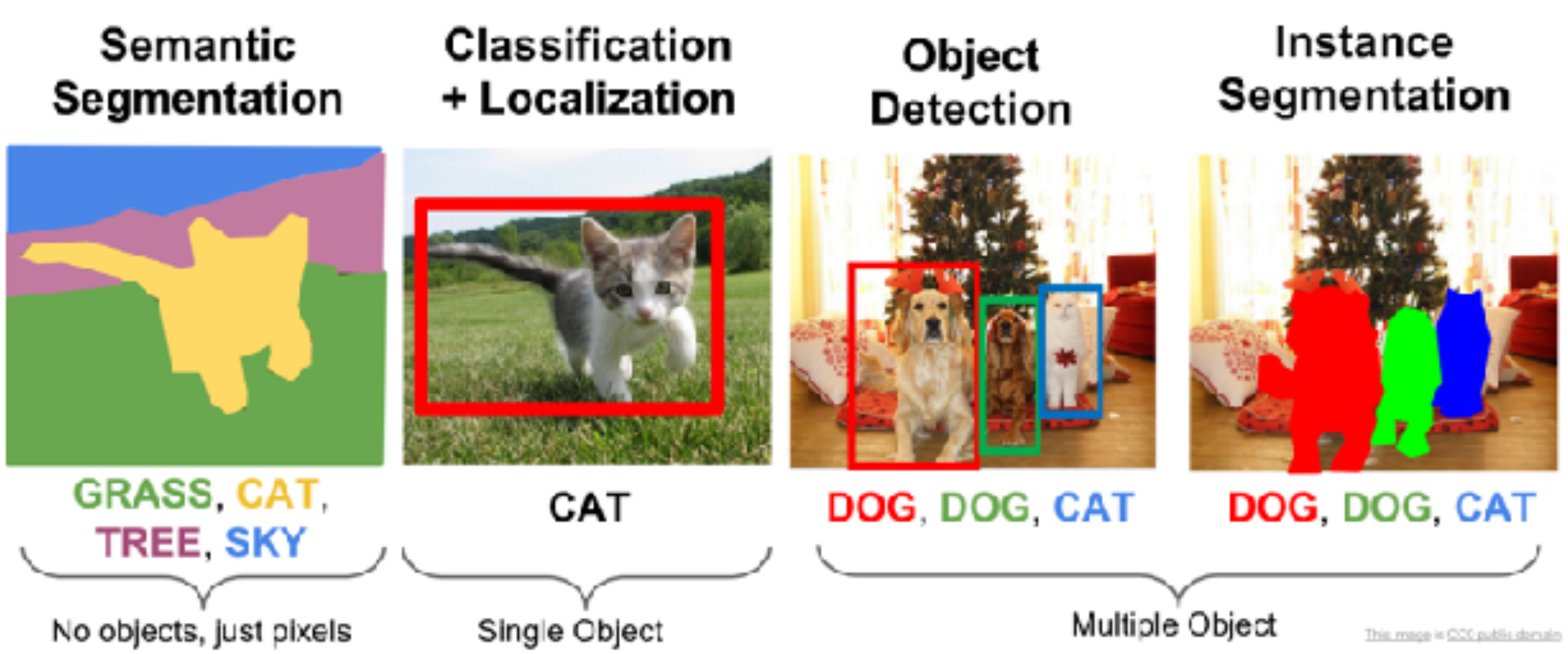
14 million images, 21,000 labels (WordNet)

Model running locally on phones

Face recognition

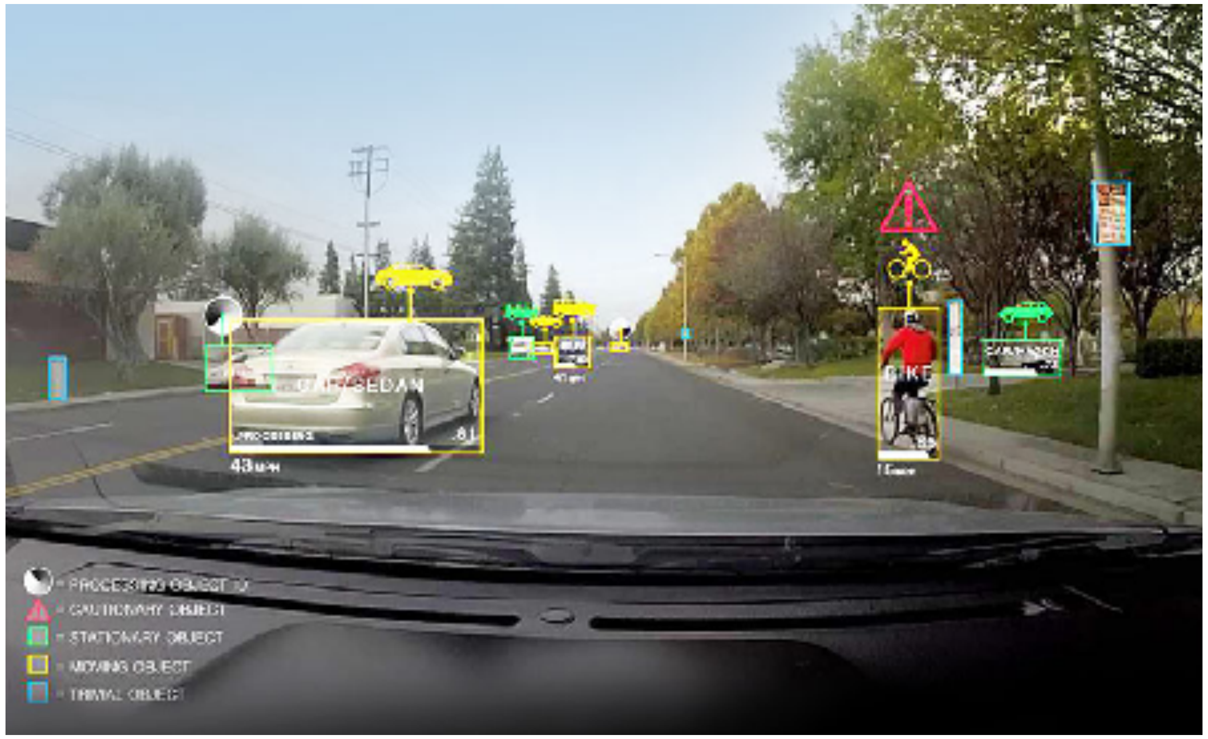
Classify each image in 1,000s of categories

Pixelwise classification



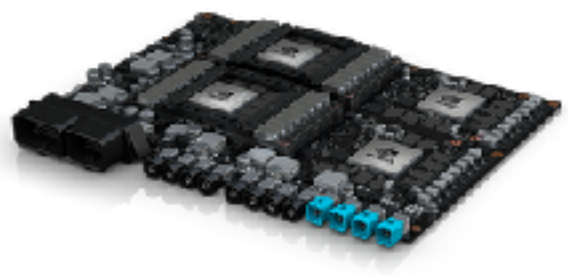
cs231n.stanford.edu

Self-driving car



[NVIDIA Drive]

Self-driving platform: Nvidia Drive

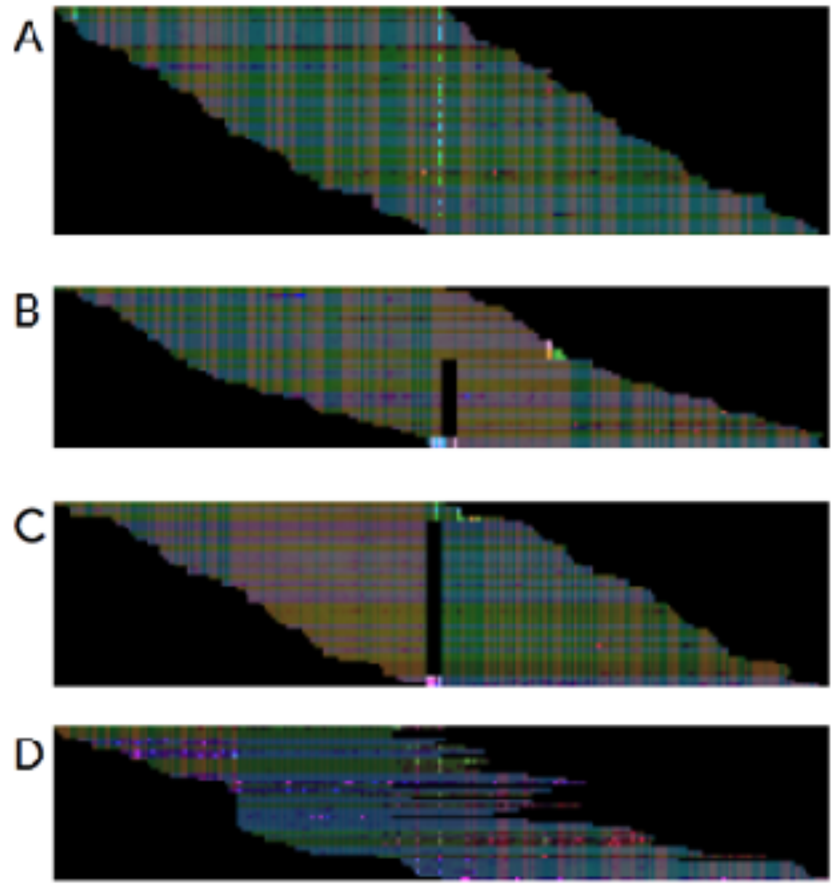


Course: selfdrivingcars.mit.edu

Genome Sequencing

Google DeepVariant: research.google.com/teams/brain/genomics/

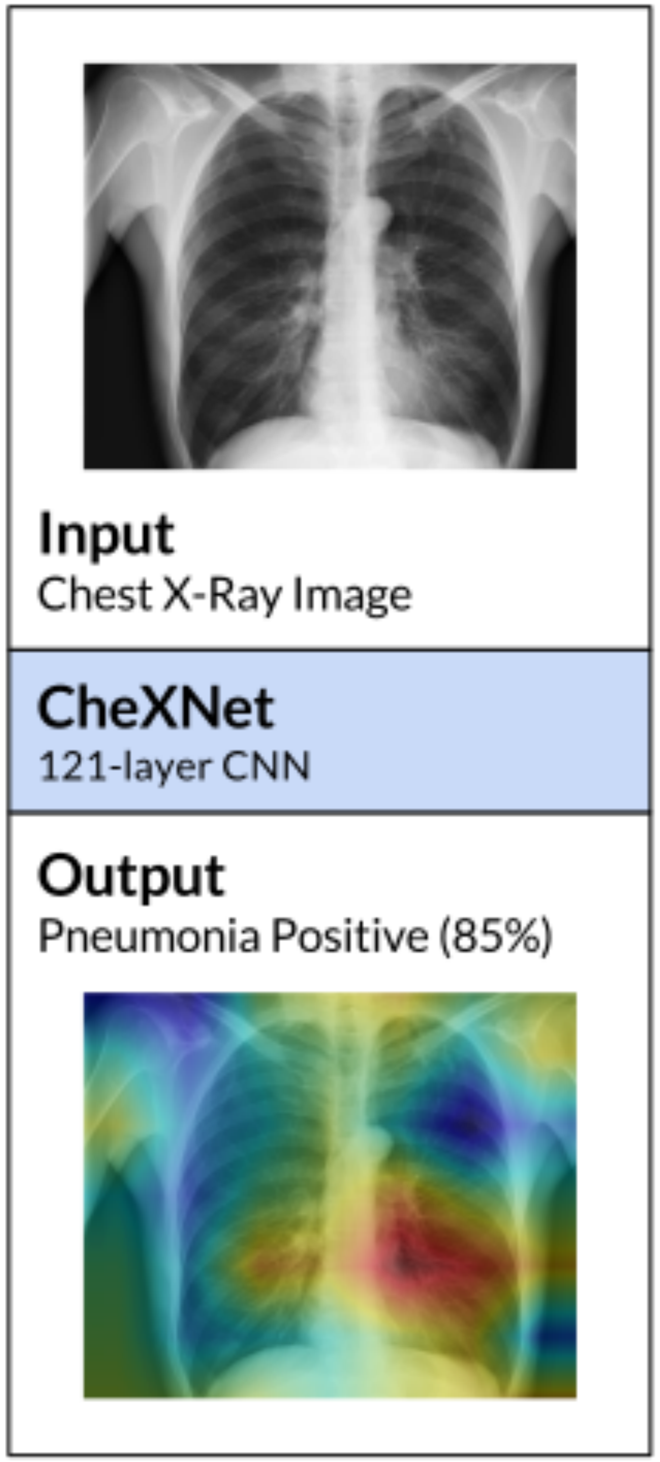
Turn problem of calling variants into a vision problem



[DeepVariant]

Medical Application: Radiology

Look for papers on arxiv.org



[CheXNet - Stanford]

CheXNet: Detecting pneumonia on Chest X-Rays - Stanford

121- layer convolutional neural network

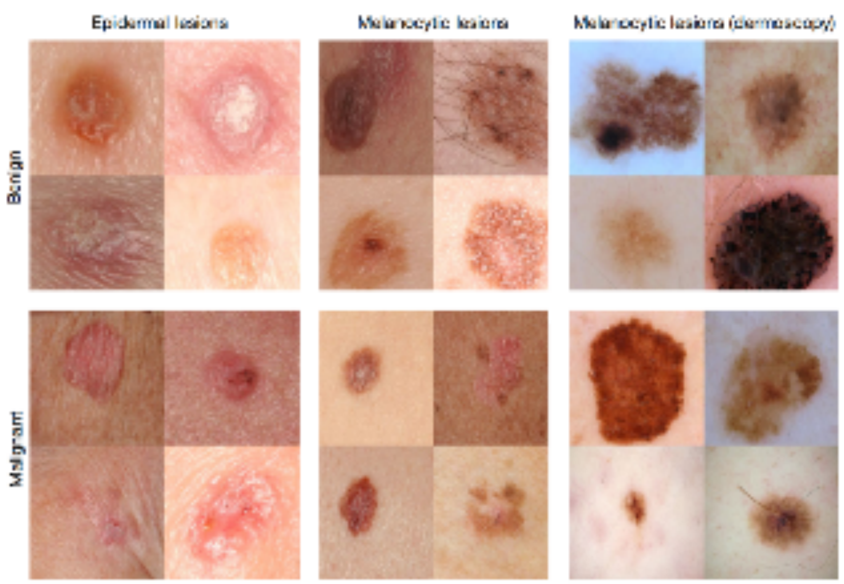
ChestX-ray 14 public dataset: 100,000 images, 14 diseases

Heatmap localizing areas most indicative of pneumonia

Exceeds average radiologist performance

Medical Application: Dermatology

[Dermatologist-level classification of skin cancer with deep neural networks - Stanford - Nature Feb. 2017]



cs.stanford.edu/people/esteva/nature

Google Inception V3 model pre-trained on ImageNet

130,000 clinical images - 2,000 diseases

Tree structured taxonomy

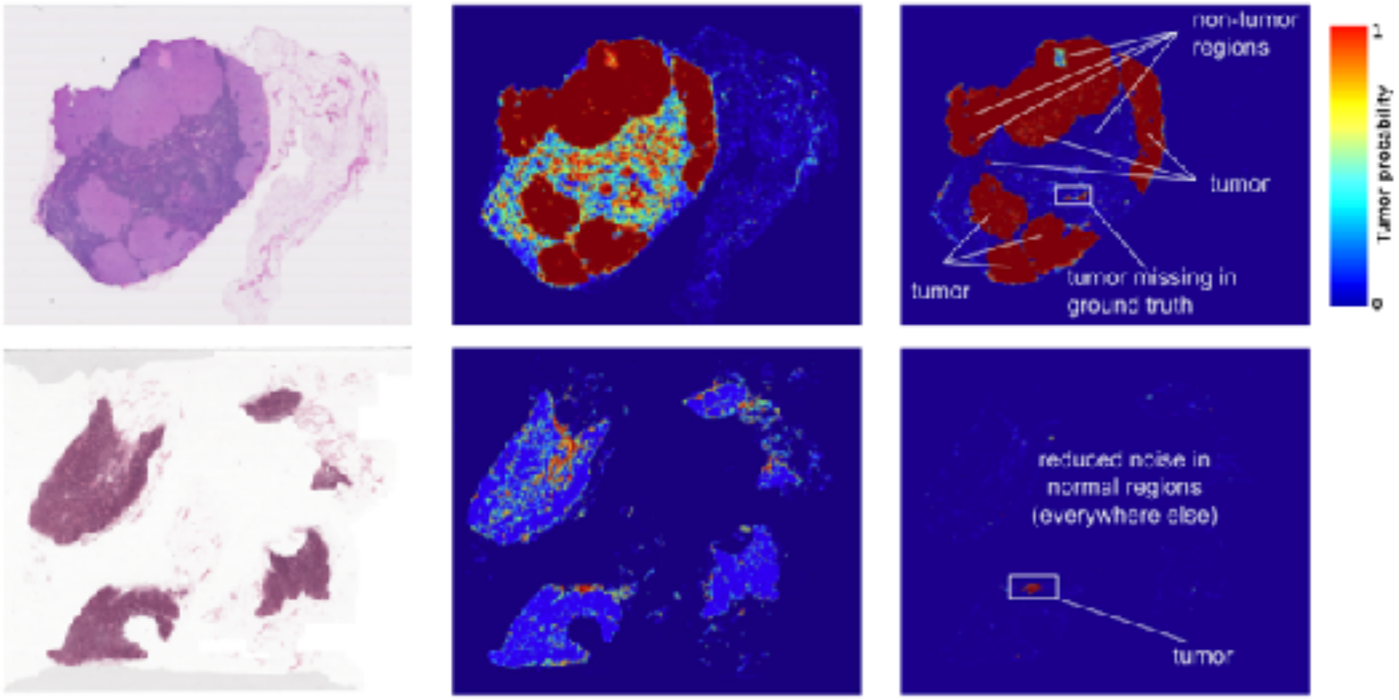
Medical Application: Pathology

Detection of breast cancer

System complementing a pathologist's workflow

camelyon17.grand-challenge.org

challenge is to evaluate new and existing algorithms for automated detection and classification of breast cancer metastases in whole-slide images of histological lymph node sections



[research.googleblog.com/2017/03/assisting-pathologists-in-detecting.html]

Inception - GoogLeNet model

Medical Application: Diabetic Retinopathy

[Wikipedia]

Normal vision

With diabetic retinopathy

Leading cause of blindness

Affects up to 80% of people who have had diabetes for 20 years or more

Can be treated if done before retina is severely damaged

[research.googleblog.com Nov. 2016]

IDx-DR: eyediagnosis.net

2018-04-11: US FDA approves Deep Learning system for Diabetic Retinopathy

https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357.htm

Quiz

adversarial image

Toaster

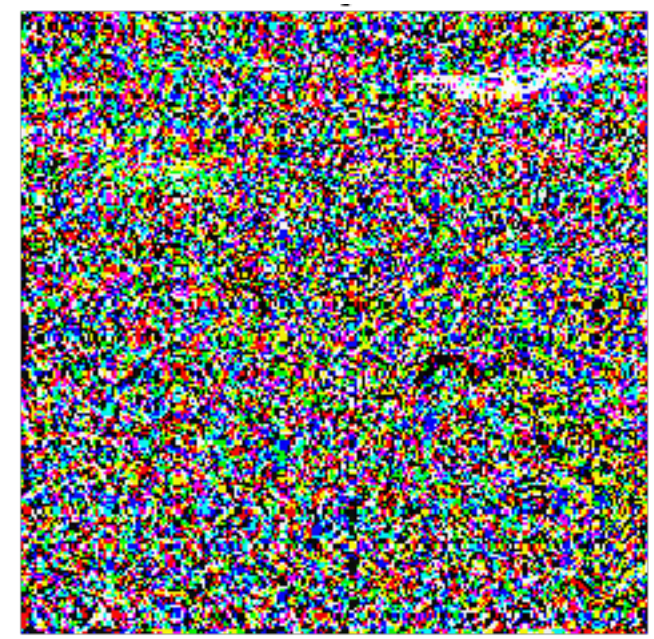
=

original image

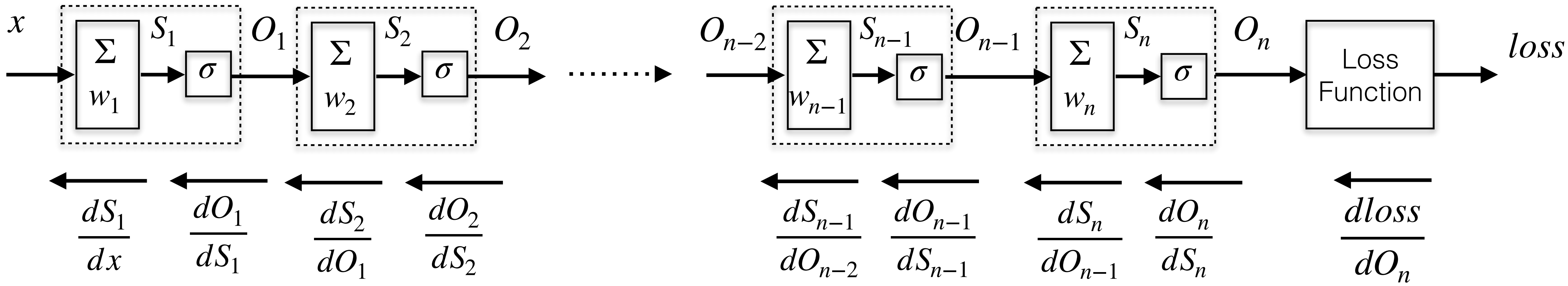
Sports car

+

adversarial noise

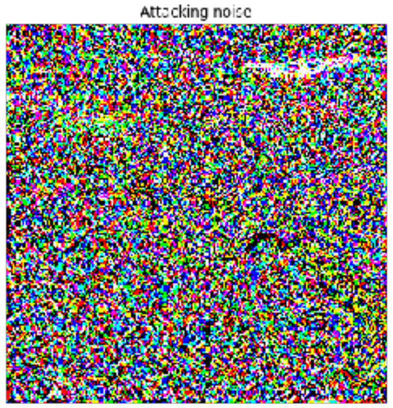


Adversarial images

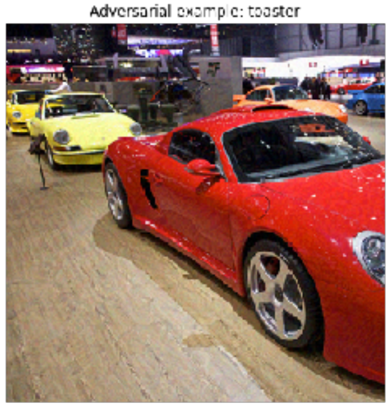


$$x_{new} = x_{old} - lr * \frac{dS_1}{dx} * \frac{dO_1}{dS_1} * \frac{dS_2}{dO_1} * \frac{dO_2}{dS_2} * \ldots * \frac{dS_{n-1}}{dO_{n-2}} * \frac{dO_{n-1}}{dS_{n-1}} * \frac{dS_n}{dO_{n-1}} * \frac{dO_n}{dS_n} * \frac{dloss}{dO_n}$$

Sports car



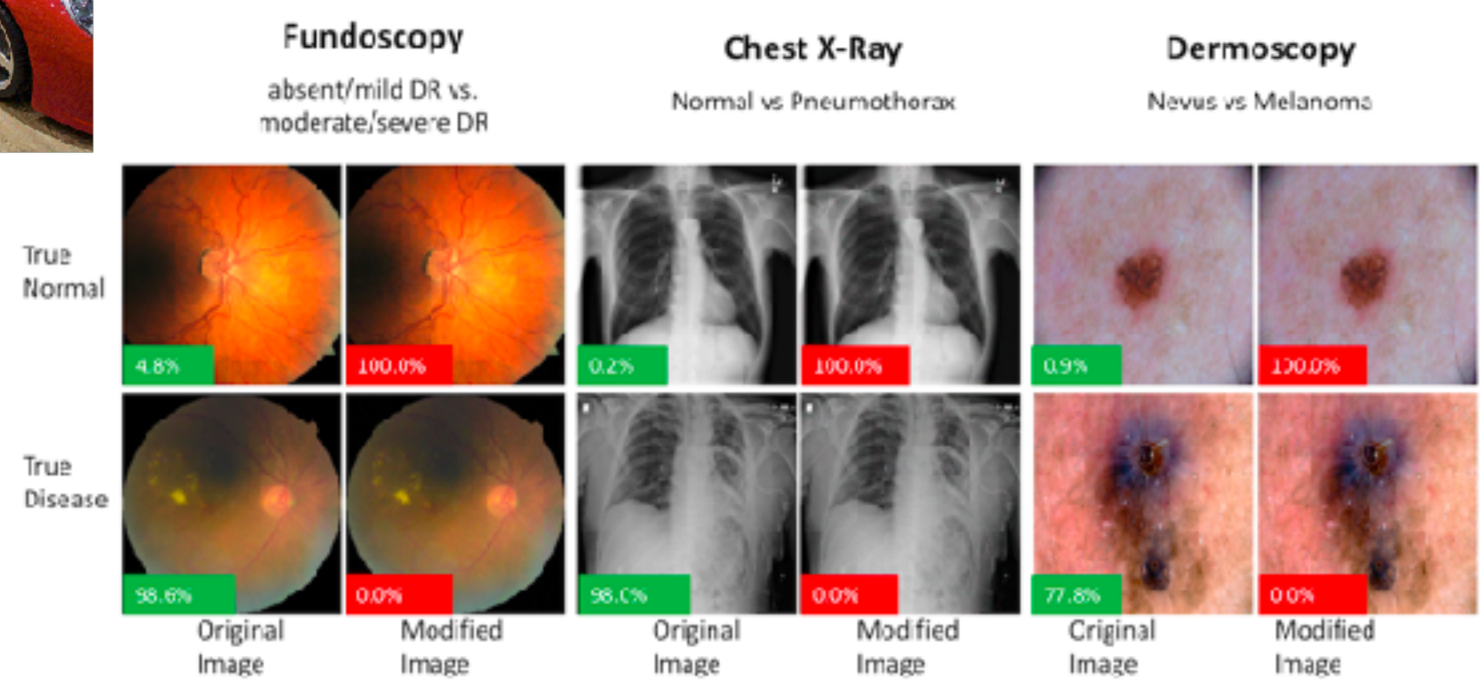
[blog.xix.ai]



Toaster

Potential for insurance frauds

Need for defense strategies



[Adversarial Attacks Against Medical Deep Learning Systems - Finlayson et al. - April 2018]

Adversarial attacks

Speed Limit 45

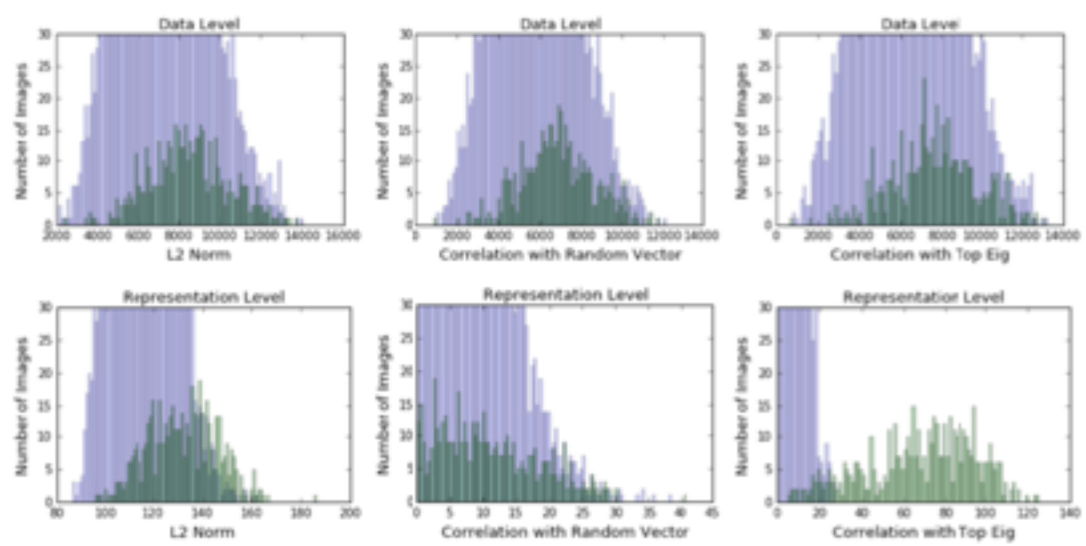
Microwave

Phone

Kevin Eykholt et al. - CVPR 2018

Adversarial attacks: counter measure (NeurIPS 2018)

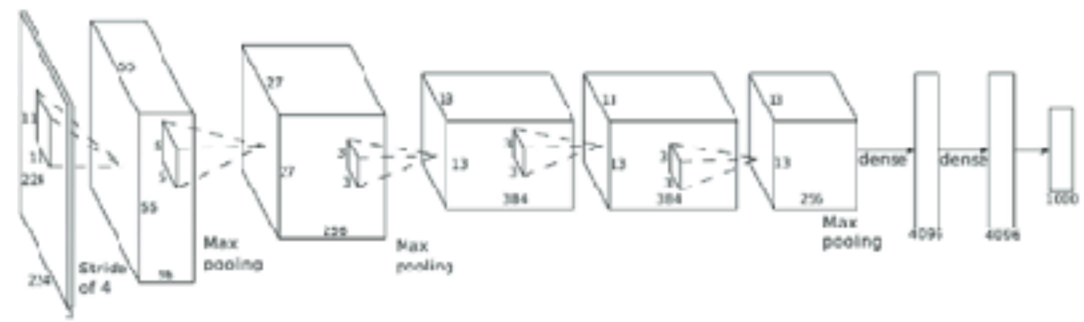
Spectral signatures on backdoor attacks - Tran et al.: Remove corrupted images based on SVD



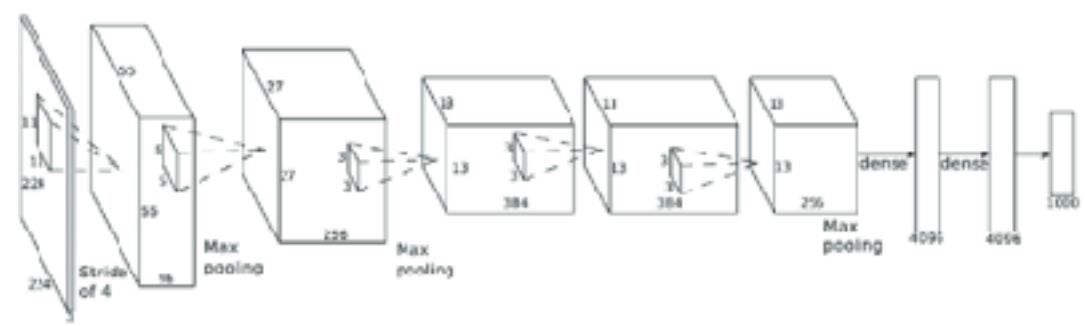
Neural Artistic Style

[A Neural Algorithm of Artistic Style by Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge, 2015]

Content image



Style image

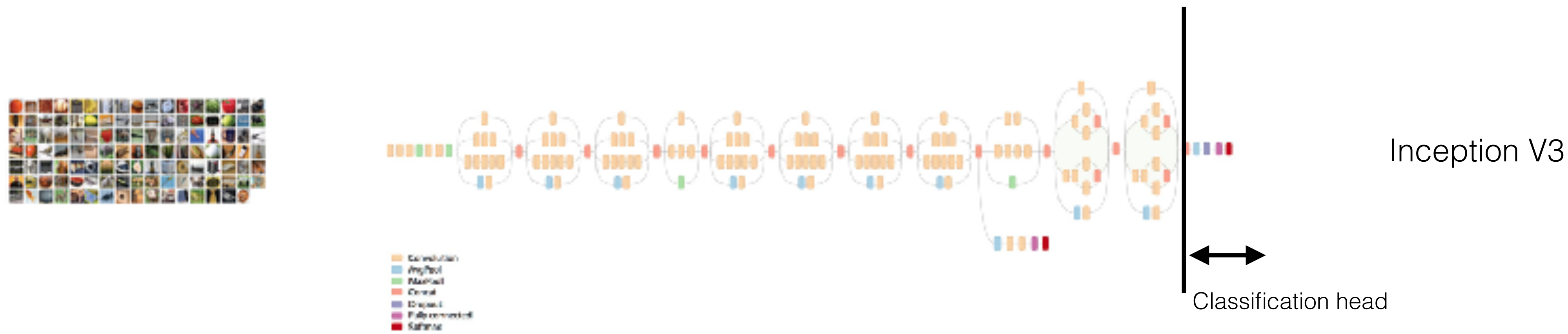


Extract covariance matrices of activations of all layers from the given image style
Optimize over content image: get content of content image and covariance matrices of style image

Do your own @ DeepArt.io

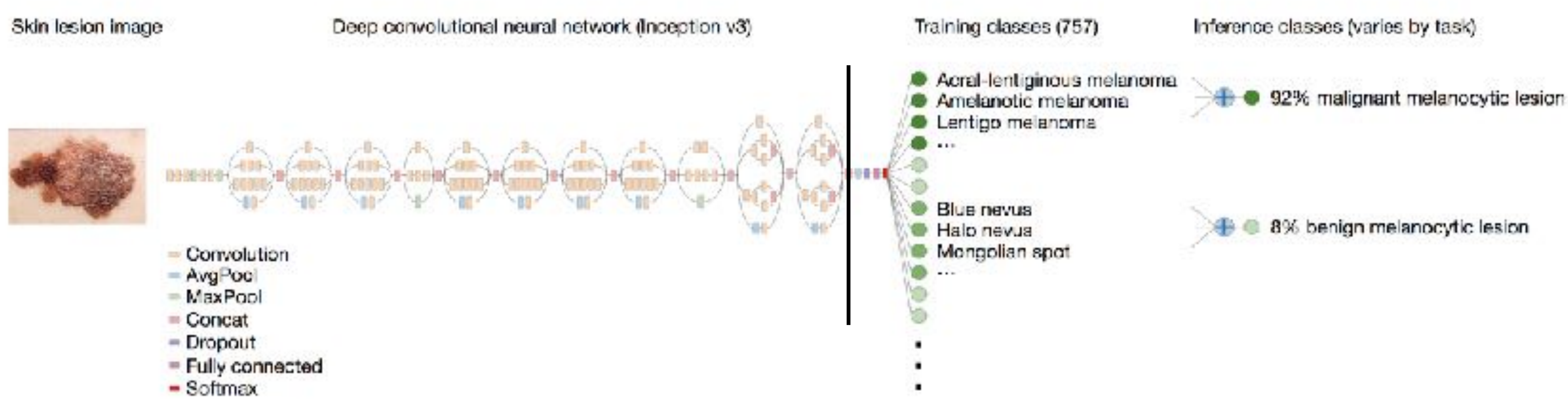
Meta Learning - Transfer Learning

Take a model pre-trained on ImageNet (e.g. available from keras.io)



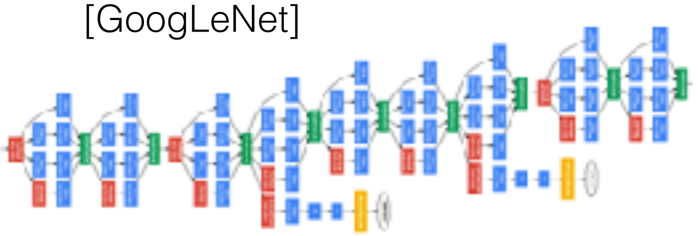
Replace classification head and re-train on new much smaller data set

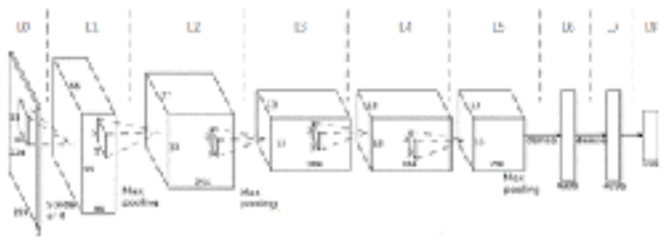
[Andre Esteva - Stanford - Skin Cancer Classification with Deep Learning - 2017]

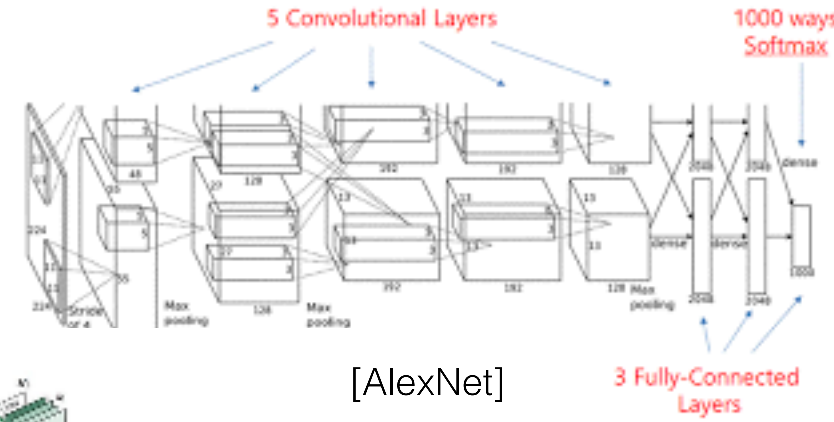


Searching for Network Architecture

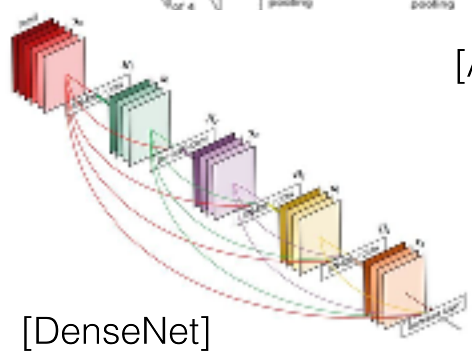
[GoogLeNet]



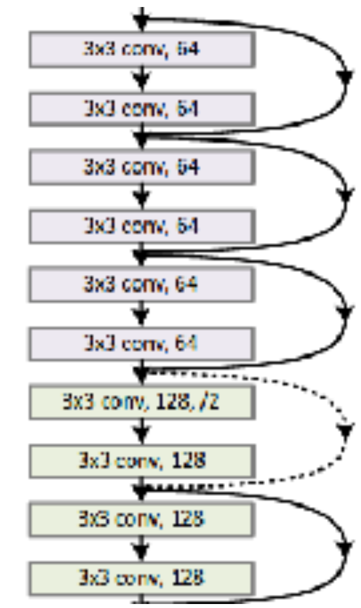




[AlexNet]



[DenseNet]

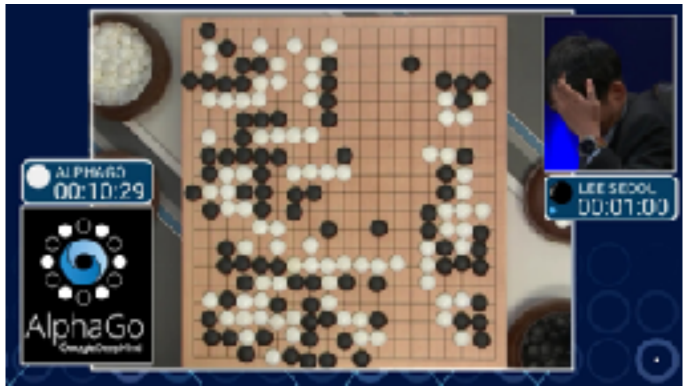


[ResNet]

- Manual task: search for optimal model architecture for given dataset
  - Time consuming - performed by skilled professionals
- Efficient Neural Architecture Search (NAS - ENAS): 2018-02
  - Google AutoML: cloud.google.com/automl
  - AutoKeras: autokeras.com
    - Transfer learning + Neural Architecture Search
    - Network built layer-wise in search for optimal performance

Tabula Rasa: Learning from scratch

Towards **unsupervised** learning (no labeled data)



AlphaGo (2015)

- Data: Use thousands of games
- Handcrafted features
- Networks: 2 (policy + evaluation)
- Rollouts: fast random games to predict moves
- Big Data, Big Processing Power

AlphaGo Zero (2017)

- Data: **None**
- Handcrafted features: **None**
- Networks: 1
- Rollouts: **None** (computed by network)
- No Data, Less Processing Power
  - Better algorithms better than power and data
- Not constrained by limits of human knowledge

deepmind.com/blog/alphago-zero-learning-scratch/

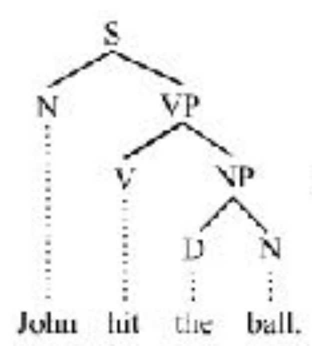
Natural Language Processing

Surfaces: talk, talked, talking, TALK, talks

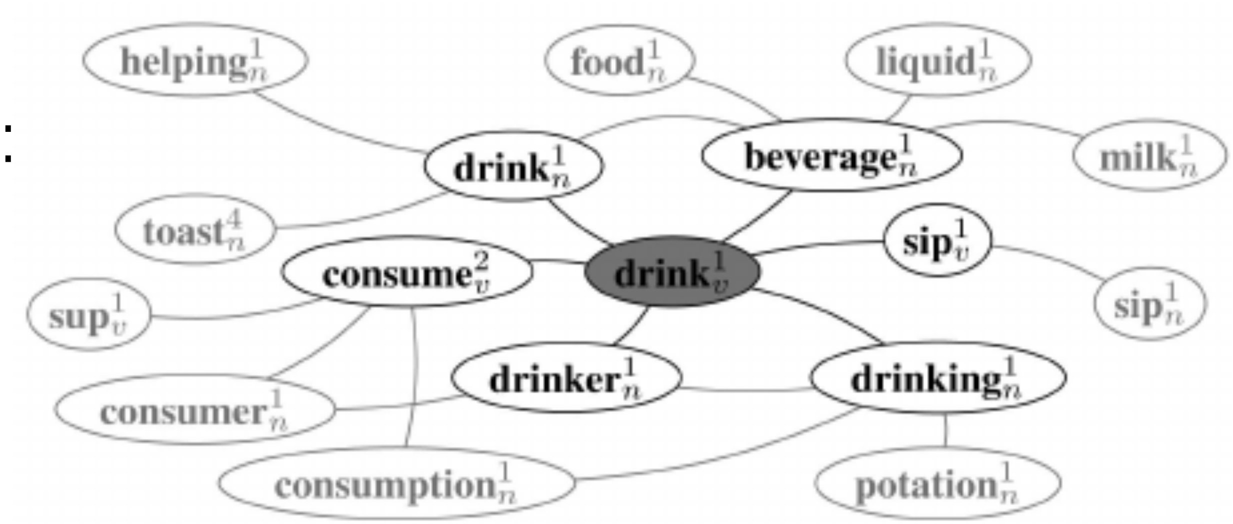
Lemma: talk

Features: word ending in “ly”, word starting with a capital letter, …

Parse trees:



Ontology:



Related words: hotel, motel, hostel, auberge

Unique ID: __UNK__ = 0, the =1, he = 2, to = 3, …, boy = 1,357, …

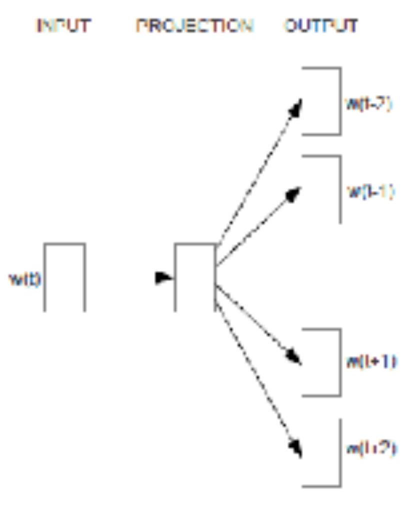
One hot encoding: boy = [0000….0000**1**0000…..00000]

Natural Language Processing (NLP)

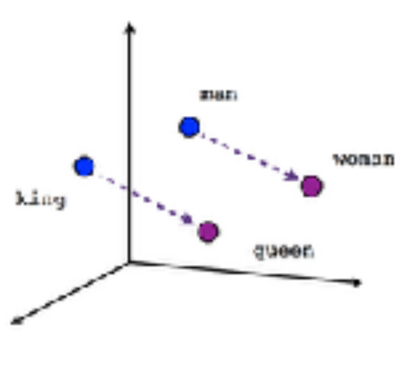
Word embeddings: words are represented as 100-300 dimensional vectors

- Vectors learned on lots of unlabeled text: e.g. Wikipedia, Common Crawl
- Most popular tools: Word2Vec (Mikolov), GloVe (Stanford)
- Train your own vectors: gensim package

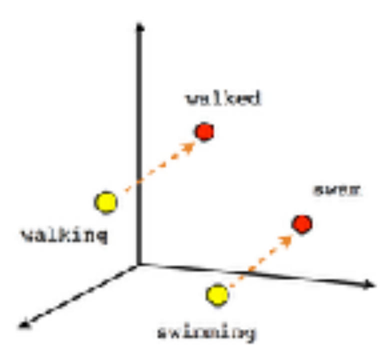
[Mikolov - Word2Vec]



Skip-gram



Male-Female



Verb tense

Words appearing in similar context

Nearest words to frog:

1. frogs
2. toad
3. litoria
4. leptodactylidae
5. rana
6. lizard
7. eleutherodactylus

nlp.stanford.edu/projects/glove/

cs224n.stanford.edu

NLP - Better Word Embeddings

- One word —> one vector   unadvised ⟶ [0.23, -0.12, …, 1.28, 0.26]

- One word —> sum of several vectors [Enriching word vectors with subword information - Facebook AI Research 2017]

  unadvised = {<unadvised> + <un + una + nad + adv + dvi + vis + ise + sed + ed>}

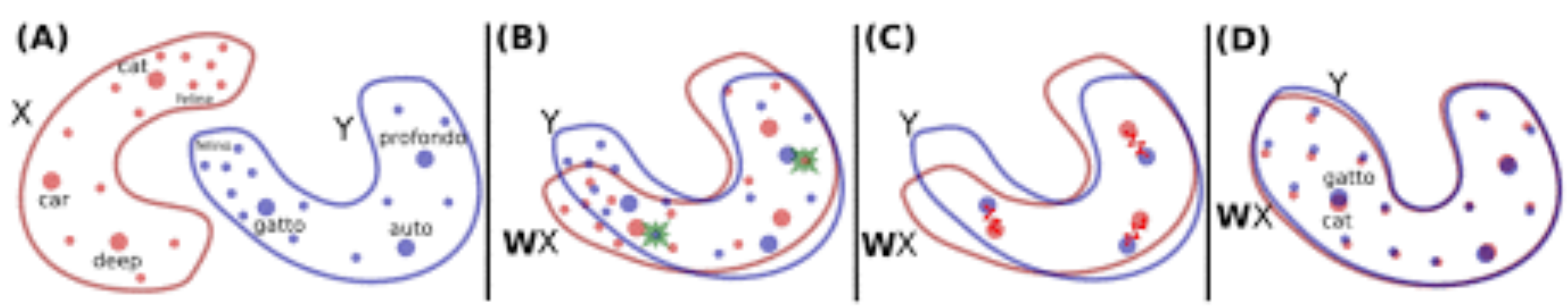
- One word —> Gaussian probability distributions [Multi Model Word Distributions - Cornell 2017]

  captures uncertainty and polysemy

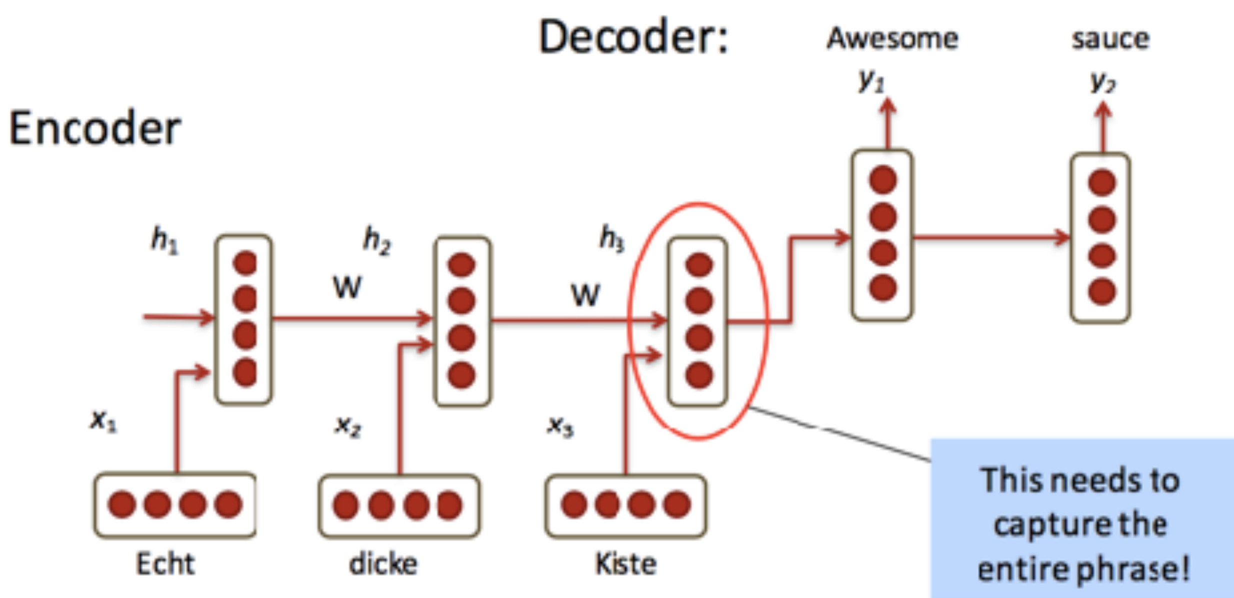
- Word Translation without Parallel Data - Facebook - ICLR 2018

  Structure learnt in embedding spaces similar among different languages (Mikolov 2013)

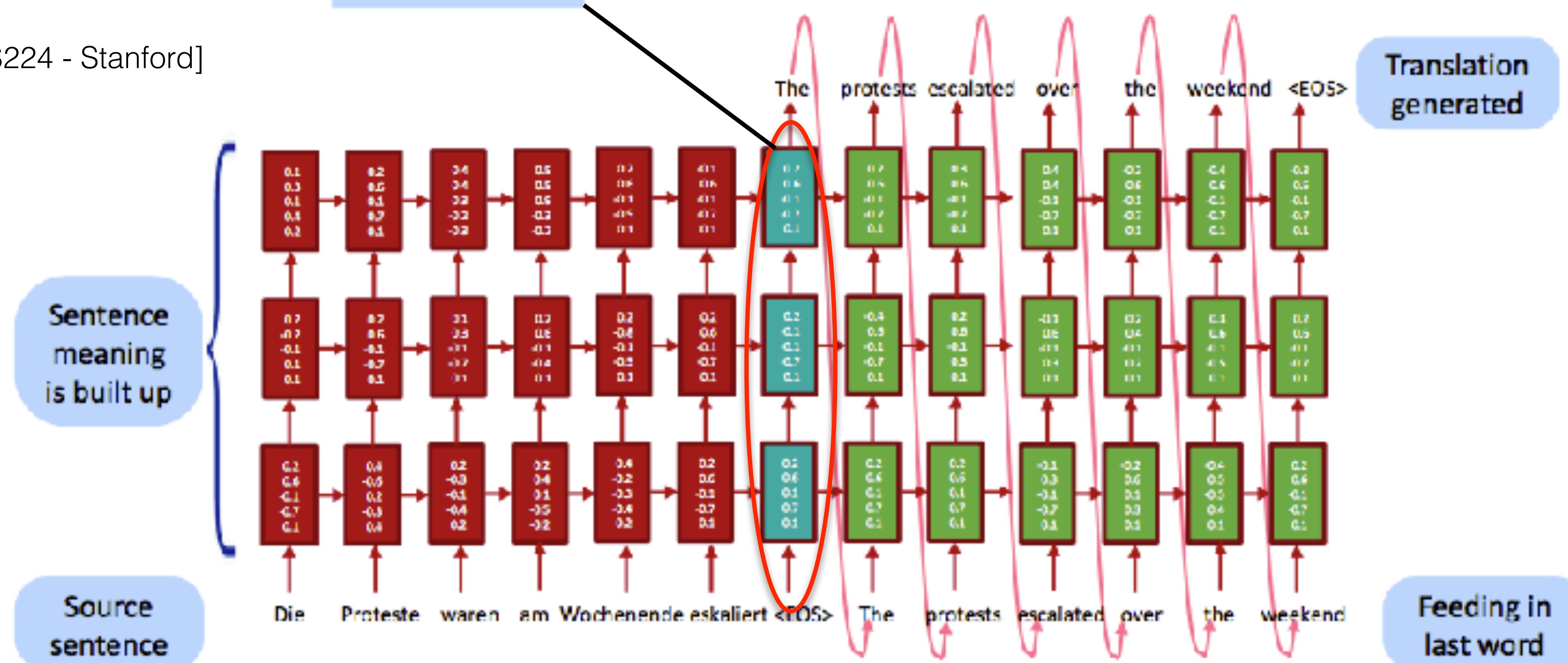
  The Procrustes problem



NLP - Neural Machine Translation



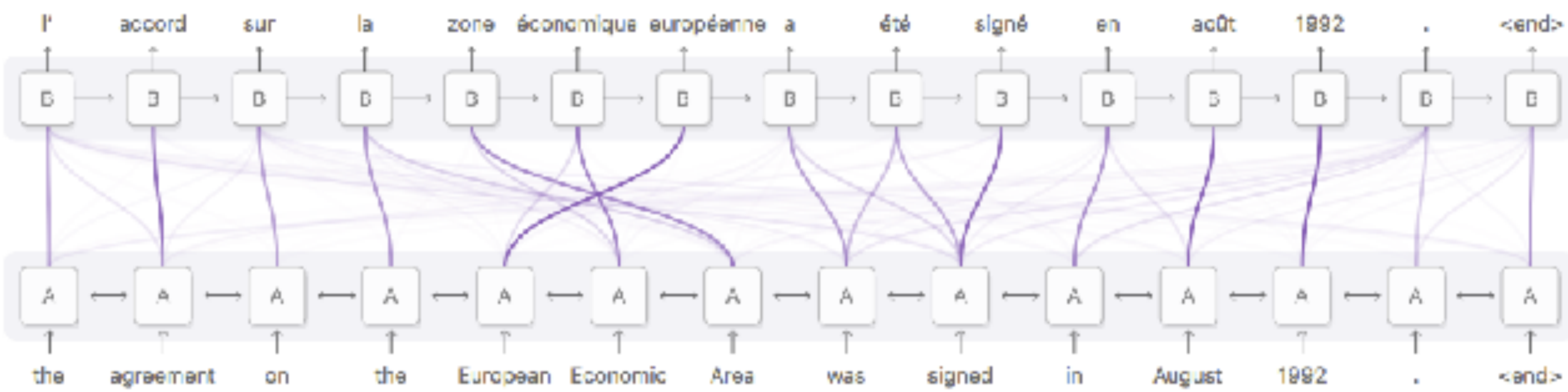
[CS224 - Stanford]



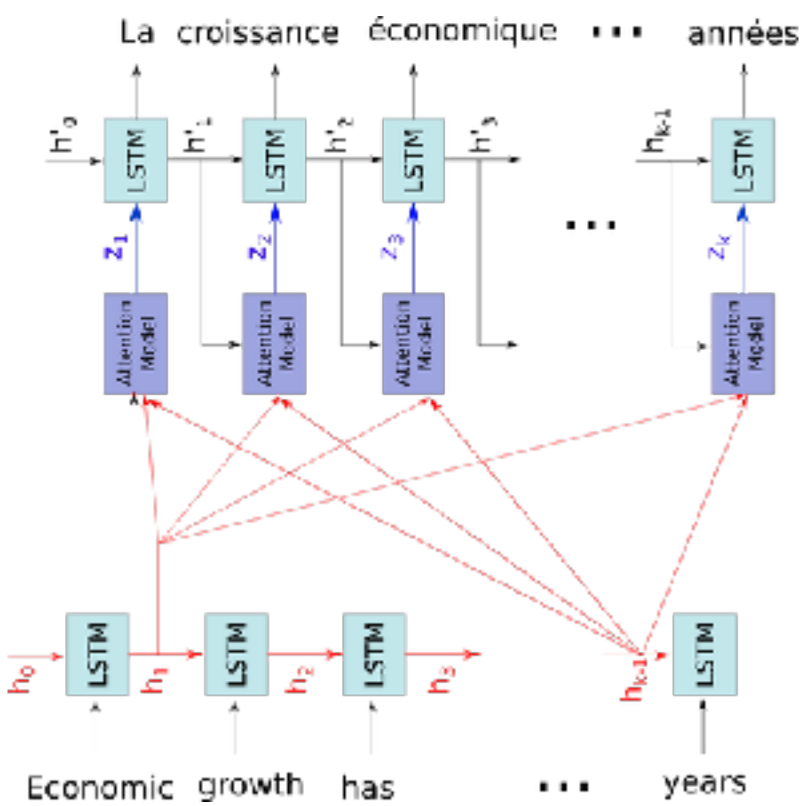
[CS224 - Stanford]

NLP - Attention Mechanism

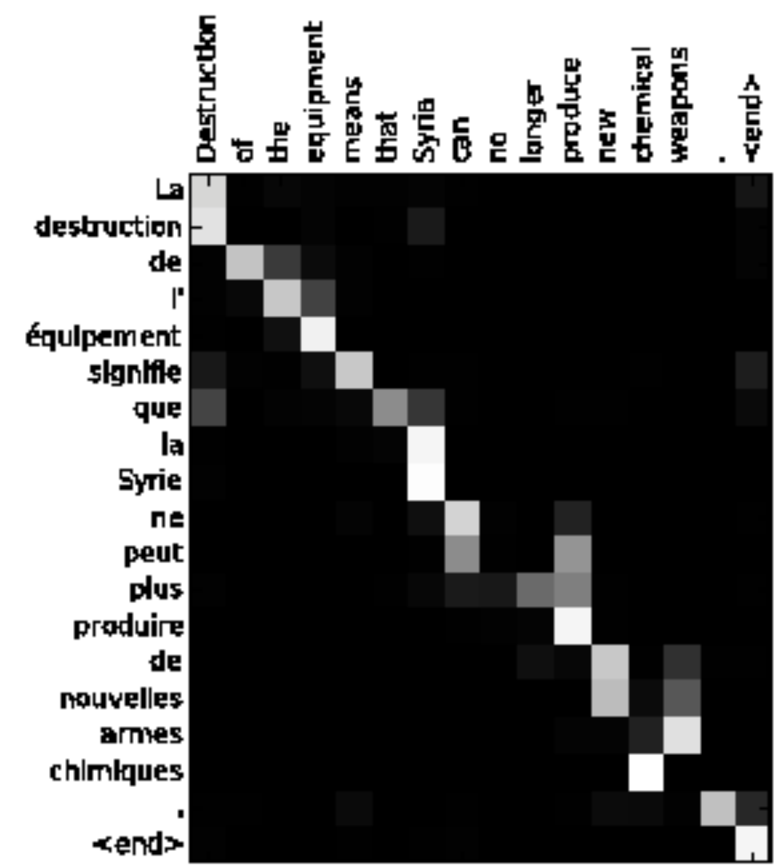
[Neural Machine Translation by Jointly Learning to Align and Translate - Bahdanau, Cho, Bengio (2015)]



[Olah & Carter - distill.pub]



[blog.heuritech.com]



[Bahdanau, Cho, Bengio - 2015]

Google Translate: nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html

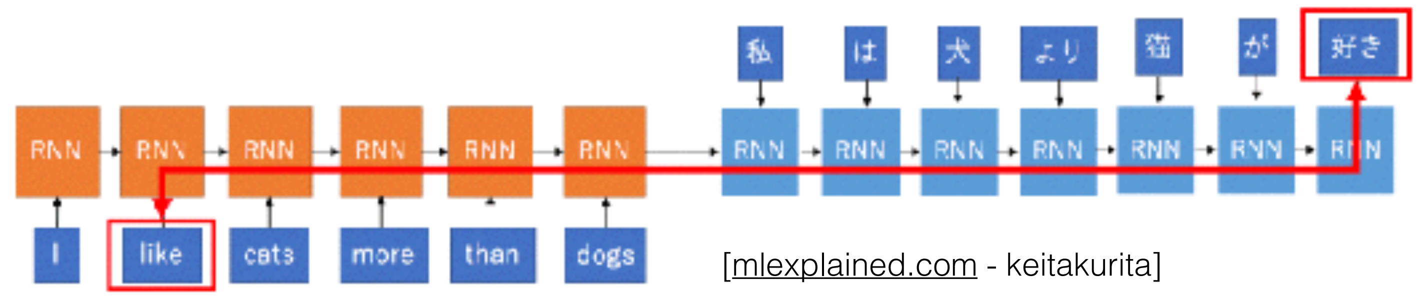
NLP - Attention Is All You Need

[A novel approach to neural machine translation - Facebook (May 2017)]

[Fully parallel Text Generation - SalesForce (Nov 2017)]

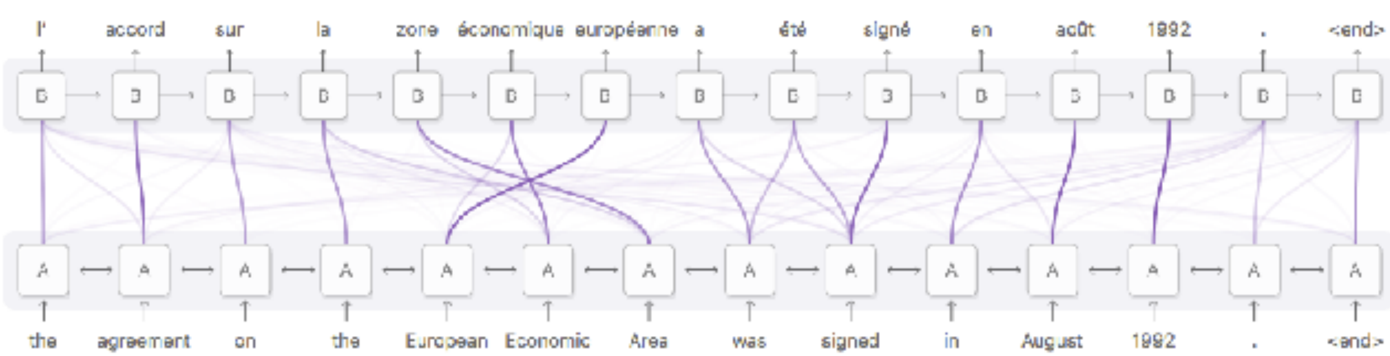
[Attention Is All You Need - Google (Dec 2017)]

1. Recurrent Neural Networks (RNN)



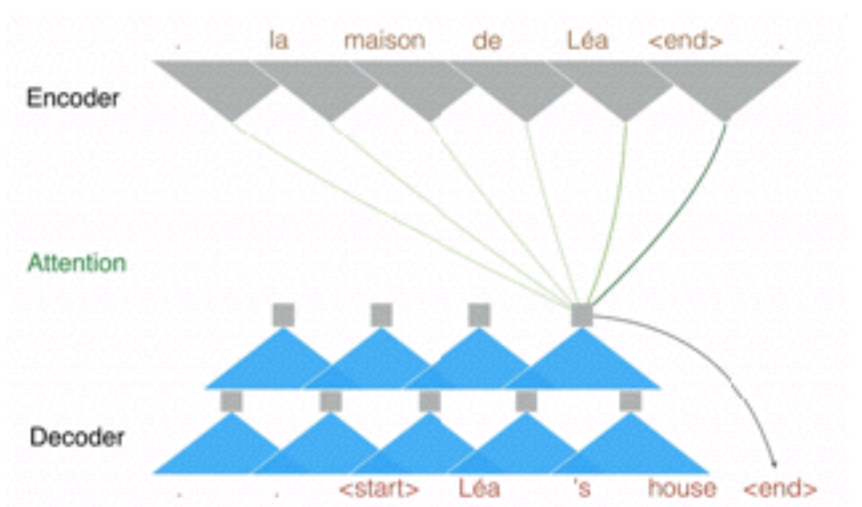
[mlexplained.com - keitakurita]

2. Recurrent Neural Networks (RNN) with Attention

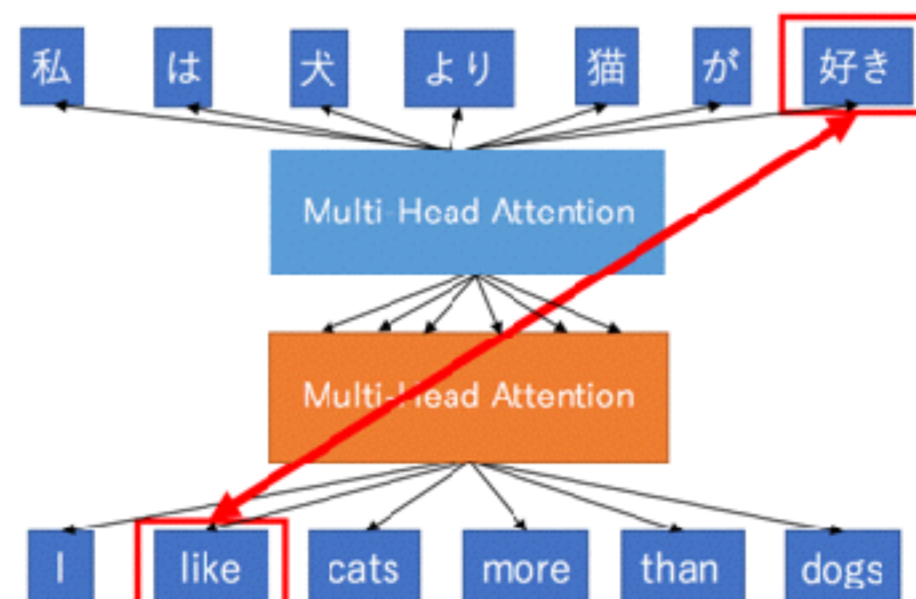


[Olah & Carter - distill.pub]

3. Attention Only - Nor more RNN (Faster)



[Facebook - Gehring

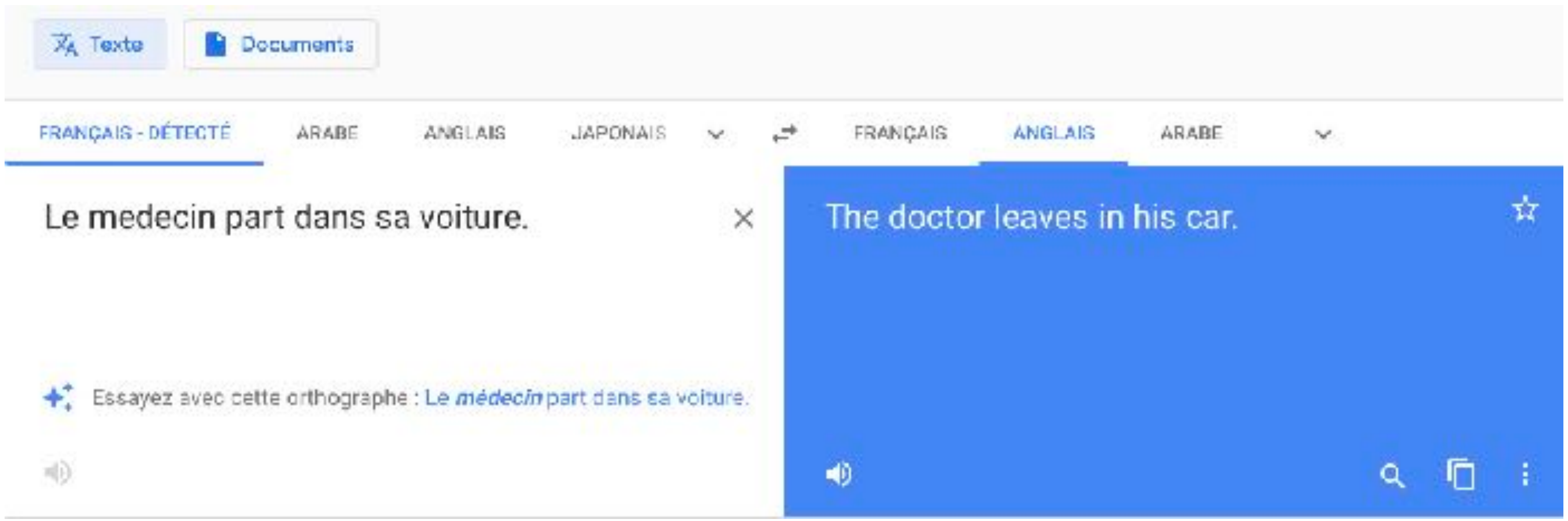


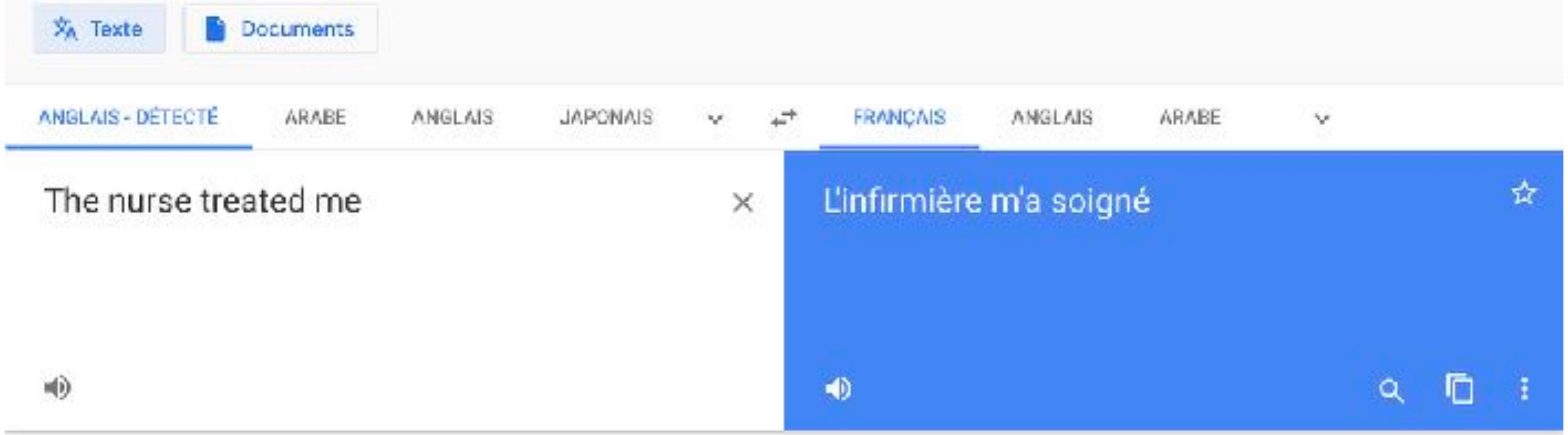
[mlexplained.com - keitakurita]

Neural Machine Translation - Reliability issues

(reliability issues highlighted in a post by Sharon Zhou - sharonzhou.me)

1. Biased data - Biased translation

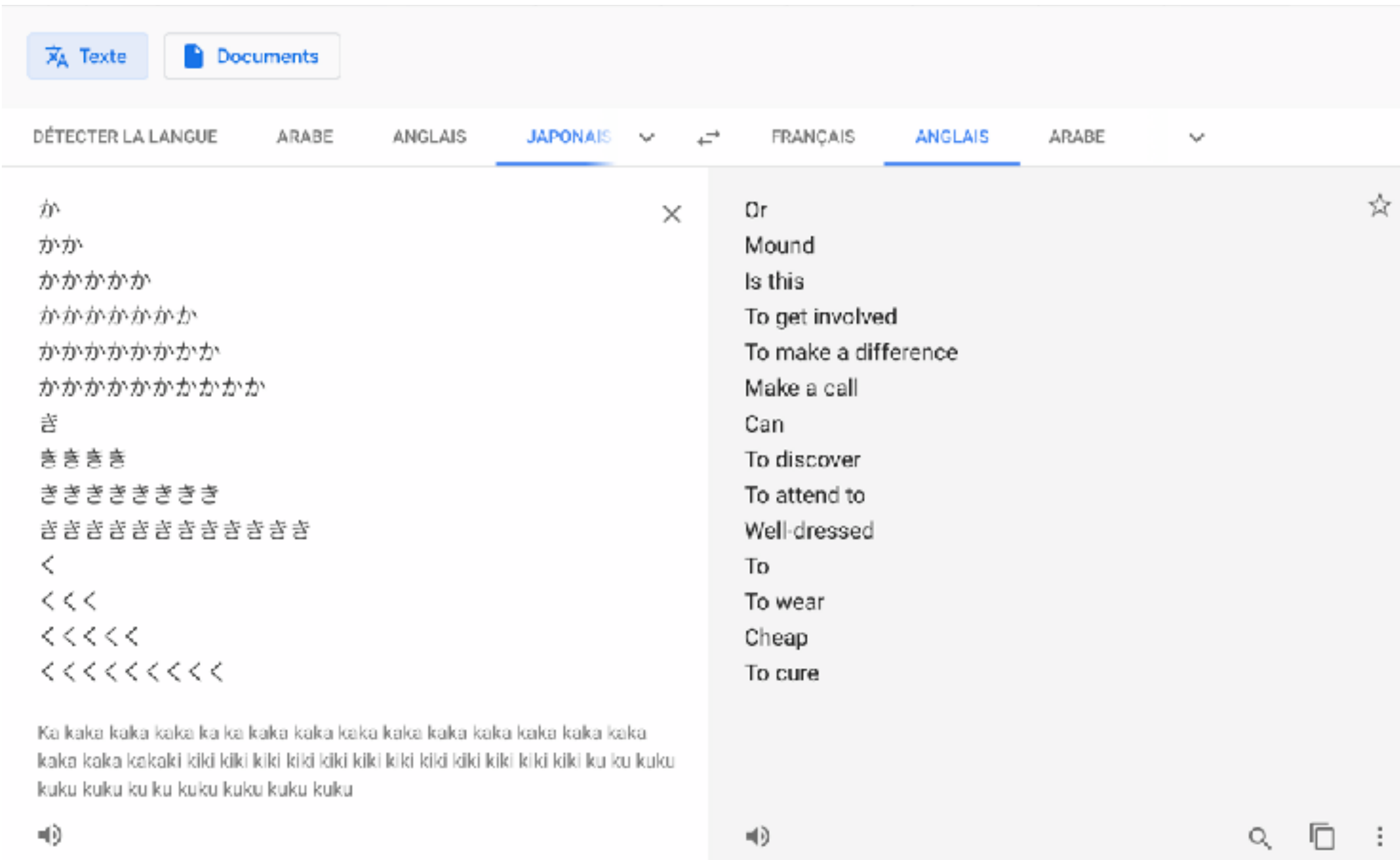




Neural Machine Translation - Reliability issues

(reliability issues highlighted in a post by Sharon Zhou - sharonzhou.me)

2. Unseen data during training



Neural Machine Translation - Reliability issues

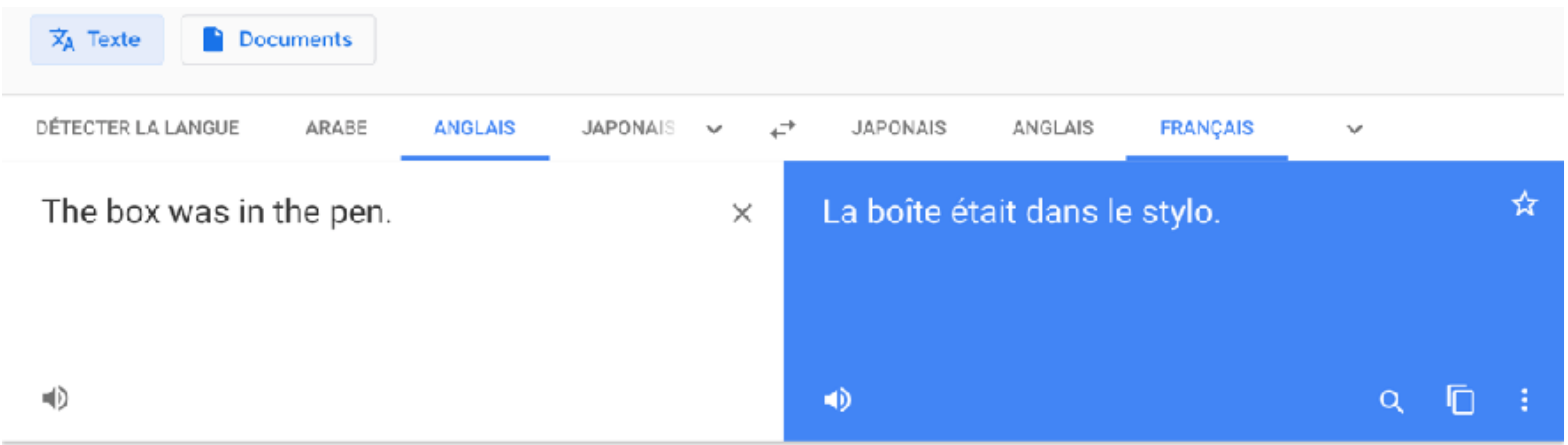
(reliability issues highlighted in a post by Sharon Zhou - sharonzhou.me)

3. Memory : translating single sentences

Neural Machine Translation - Reliability issues

(reliability issues highlighted in a post by Sharon Zhou - sharonzhou.me)

4. Lacking world knowledge



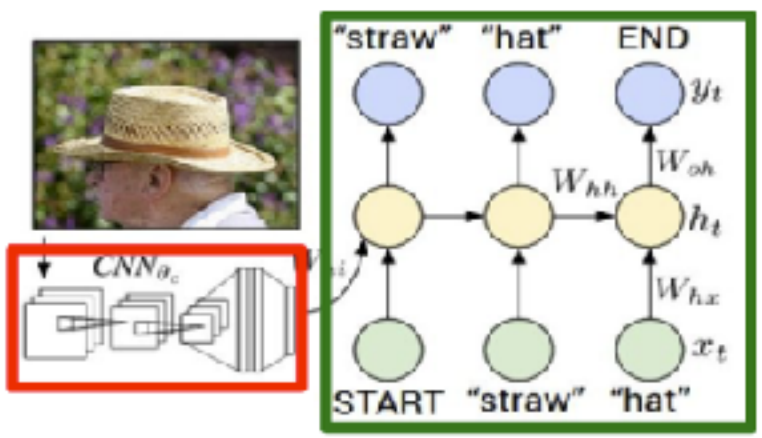
The Future

- Training more data in the amount of time
- OpenNMT.net: Open Source Neural Machine Translation System

Vision - NLP: Image annotation

[CS231n.stanford.edu]

Recurrent Neural Network



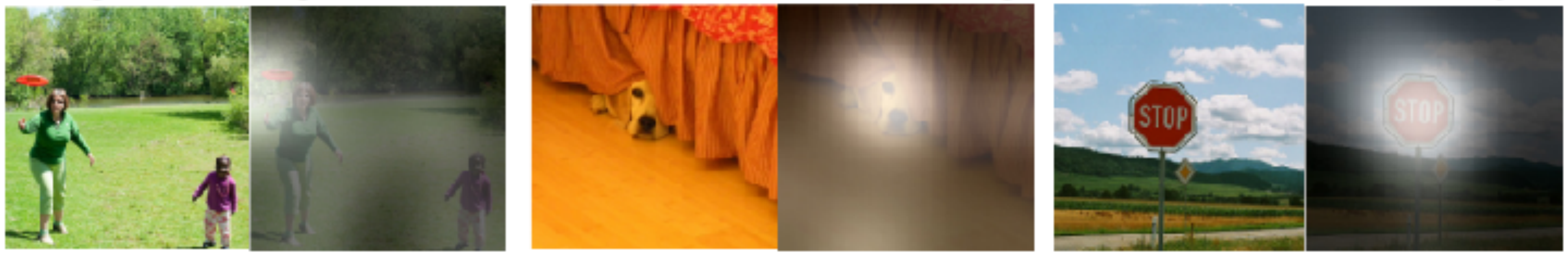
Convolutional Neural Network

[CS231n.stanford.edu]

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

Figure 3. Examples of attending to the correct object (*white* indicates the attended regions, *underlines* indicated the corresponding word)



A woman is throwing a frisbee in a park. A dog is standing on a hardwood floor. A stop sign is on a road with a mountain in the background.

[Show, Attend and Tell: Neural Image Caption Generation with Visual Attention - 2016 (UofM, UofT)]

getidentifi.com (cloudsight.ai)

Generative Adversarial Networks (GANs)

Ian Goodfellow - 2014

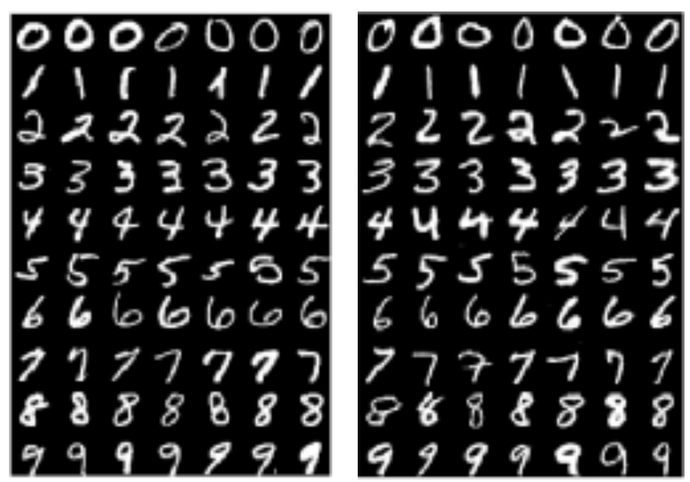
“What I cannot create, I do not understand.” — Richard P. Feynman

Karras et al. (NVIDIA) - ICLR 2018

Yanghua Jin et al. - 2017

Only two real images (Spolier: Row 1, C and Row 2, D) Vue.ai

Elgammal et al. ICCC 2017



Real (MNIST) Generated

Alec Radford et al. - ICLR 2016

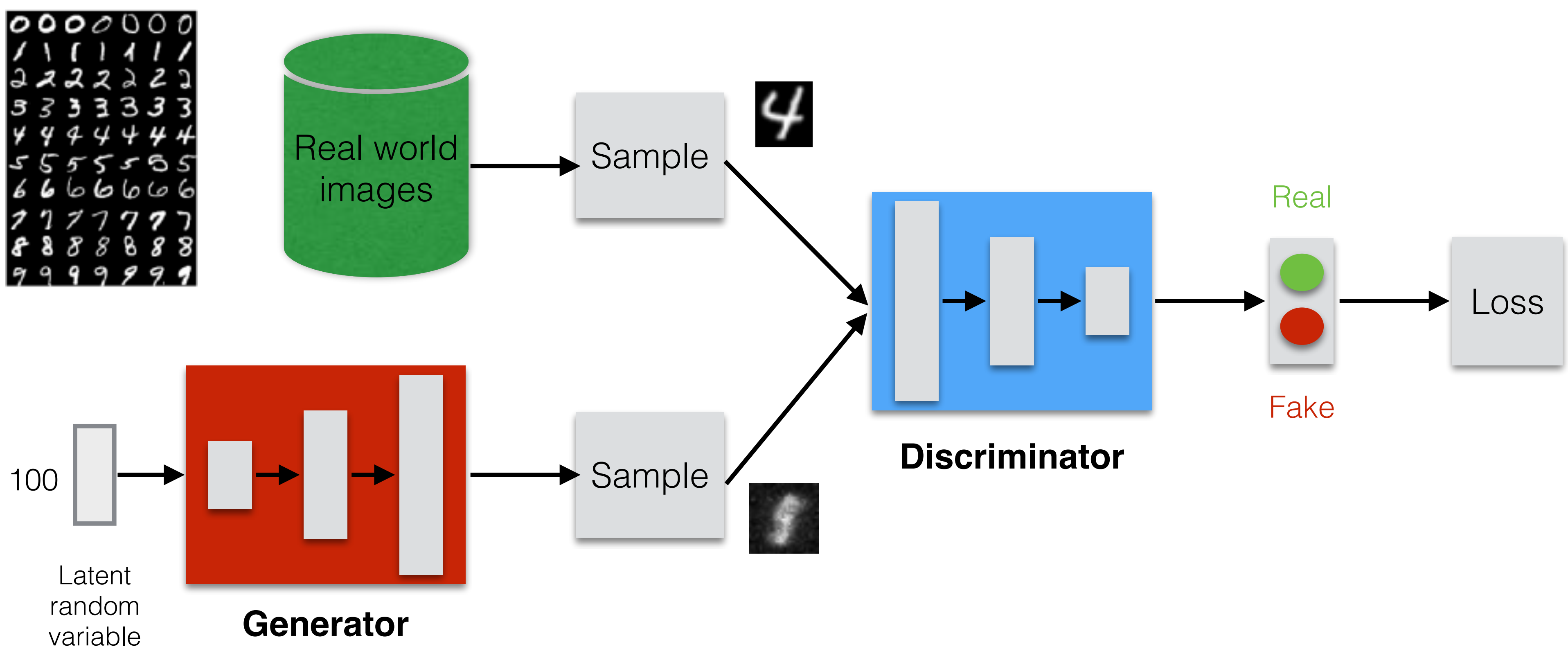
(b) Handbag images (input) & **Generated** shoe images (output)

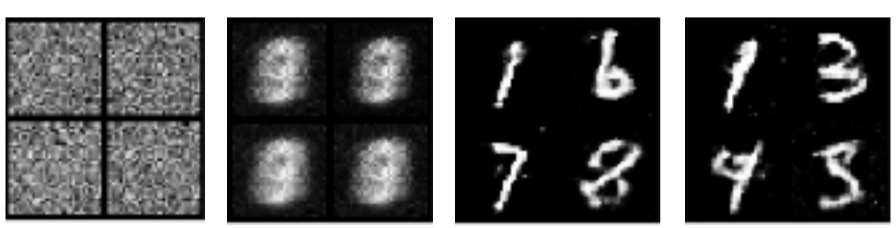
DiscoGAN

- Fashion industry: Alibaba, Amazon, …
- Generation of music…

Generative Adversarial Networks (GANs)

GANs from Scratch 1: A deep introduction - Diego Gomez Mosquera

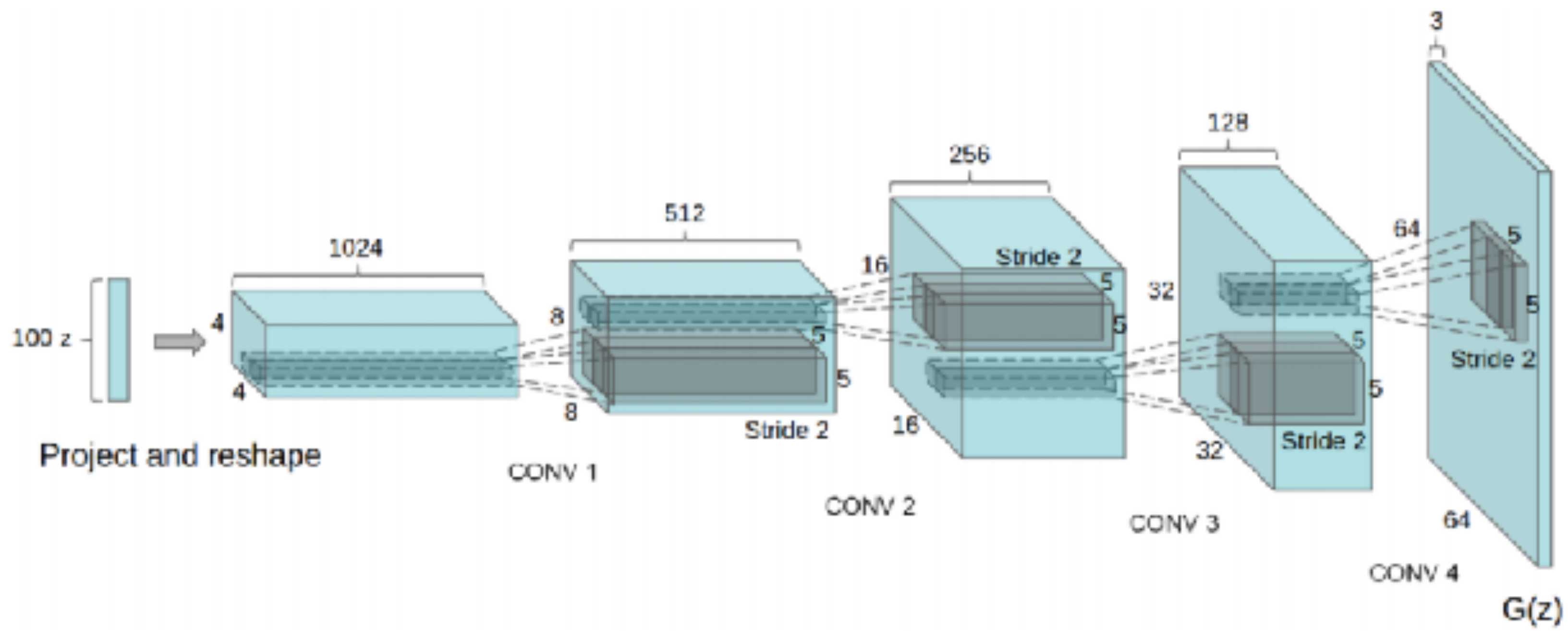




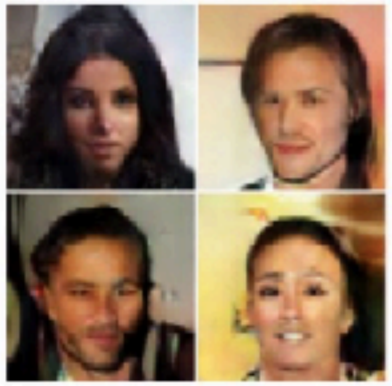
iter 0 iter 3 iter 100 iter 200

Alec Radford et al. - ICLR 2016

DCGAN



DCGAN
11/2015



EBGAN-PT
9/2016

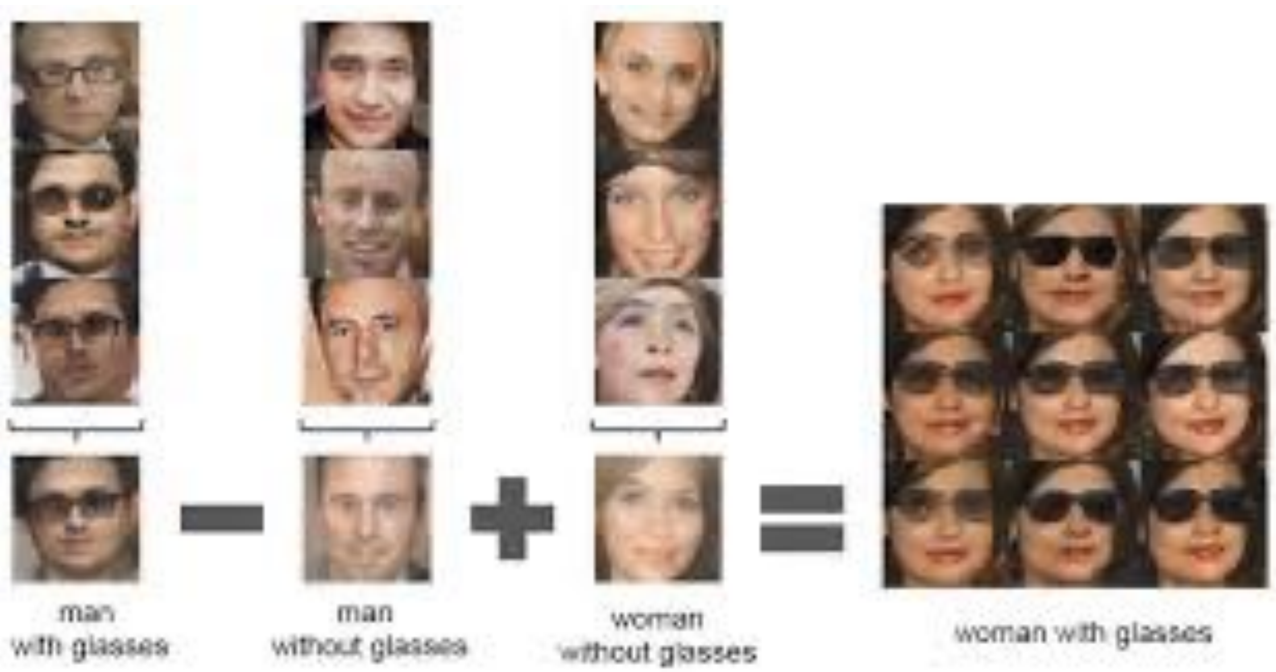
BEGAN
3/2017
128 × 128

Progressive GAN
10/2017
1024 x 1024

https://medium.com/@jonathan_hui

thispersondoesnotexist.com

whichfaceisreal.com



UNSUPERVISED REPRESENTATION LEARNING WITH **DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS**
Alec Radford & Luke Metz - ICLR 2016

Robbie Barrat

github.com/robbiebarrat/art-DCGAN

Code: Open source from Robbie Barrat

Code and Deep Learning Expertise:

Robbie Barrat

github.com/robbiebarrat/art-DCGAN

Who: **Caselles**, Fautrel, Vernier

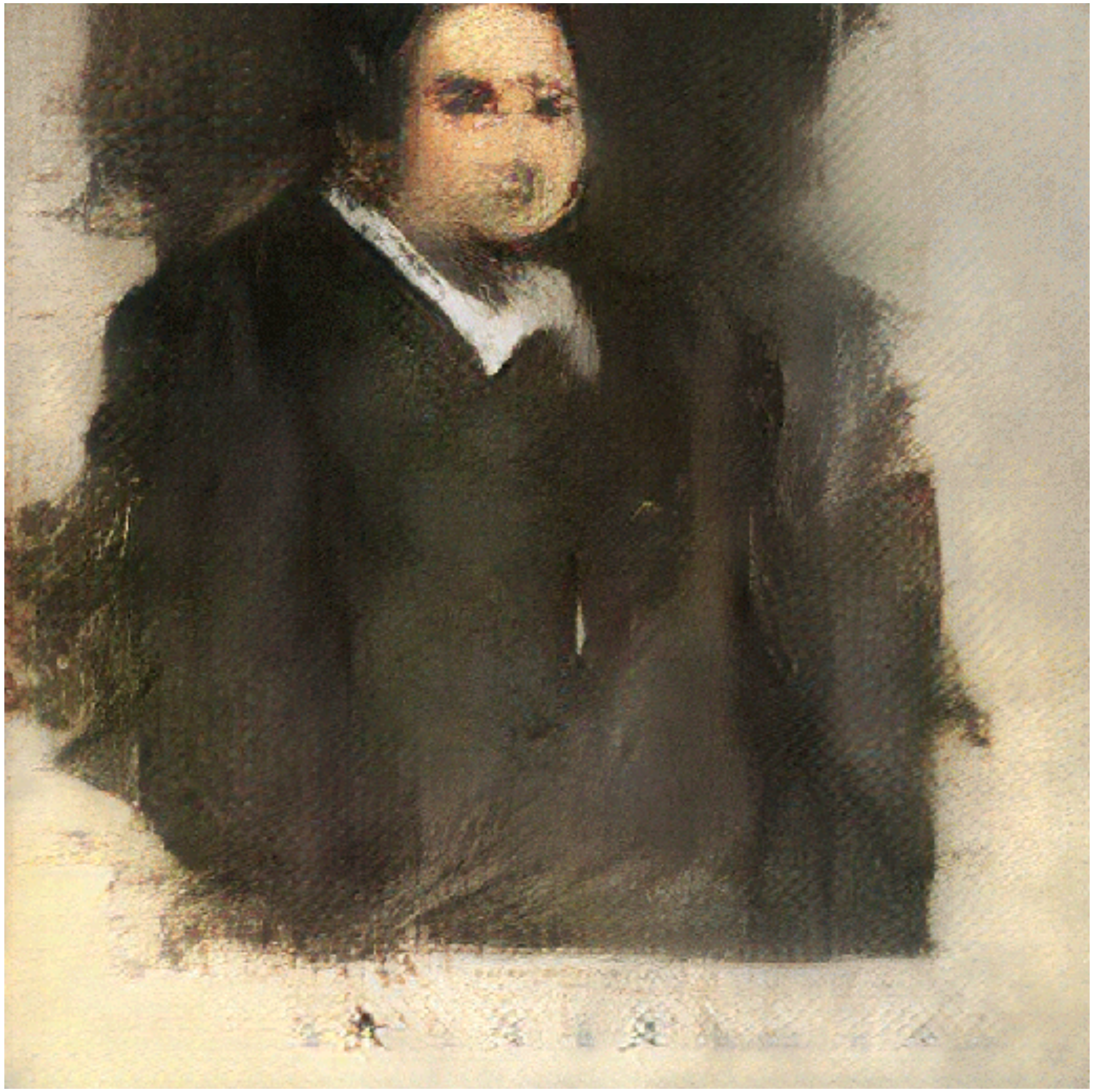
Expertise: Marketing

Site: http://obvious-art.com

Auction house: Christie's New York

When: October 25th, 2018

Price sold: 432 500 USD



https://www.theverge.com/2018/10/23/18013190/ai-art-portrait-auction-christies-belamy-obvious-robbie-barrat-gans

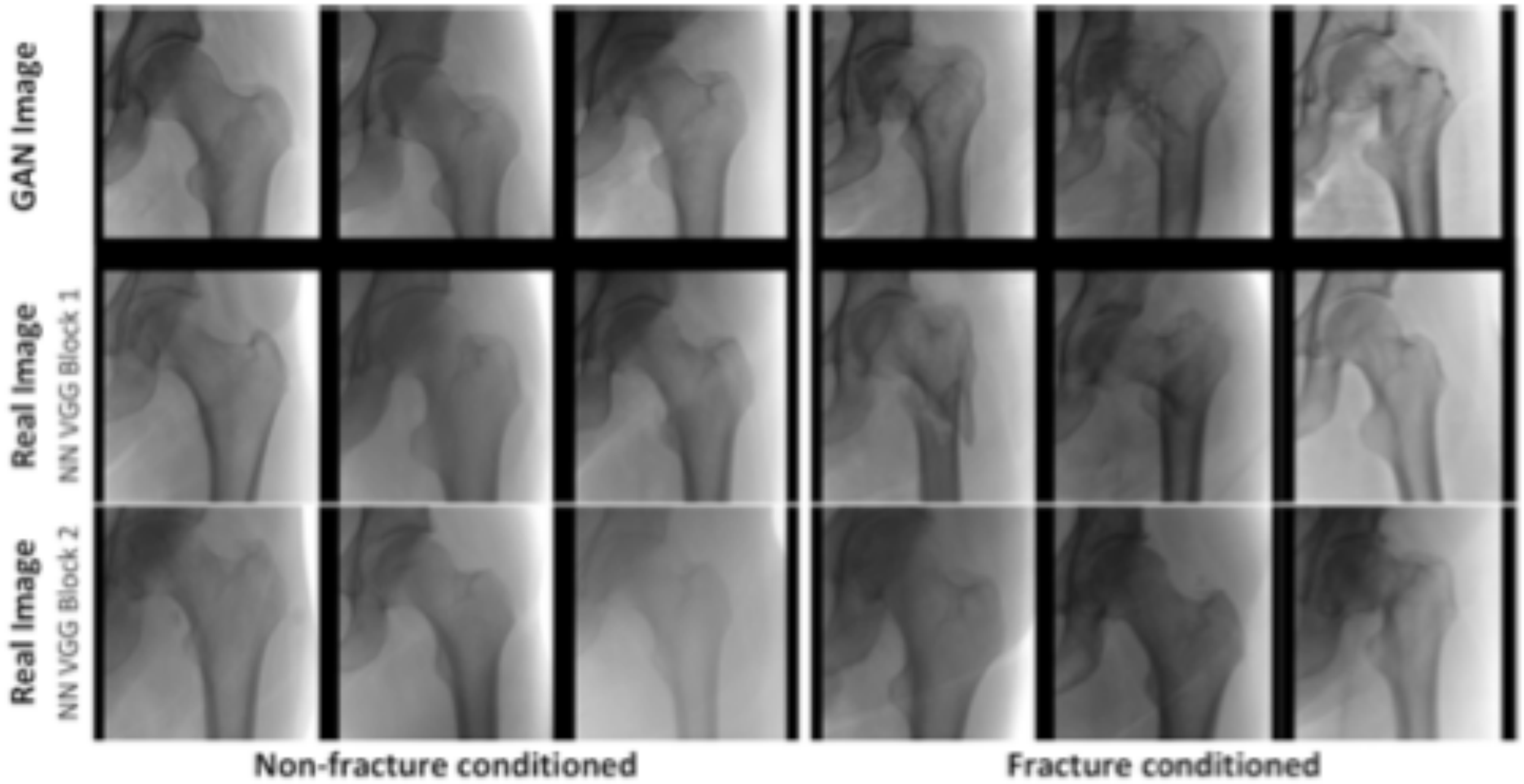
https://github.com/robbiebarrat/art-DCGAN/issues/3

StackGAN++

StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks

2017

NeurIPS 2018 - Application of GANs



https://arxiv.org/pdf/1812.01547.pdf

The GAN Zoo

https://github.com/hindupuravinash/the-gan-zoo

- 3D-ED-GAN - Shape Inpainting using 3D Generative Adversarial Network and Recurrent Convolutional Networks
- 3D-GAN - Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling (github)
- 3D-IWGAN - Improved Adversarial Systems for 3D Object Generation and Reconstruction (github)
- 3D-PhysNet - 3D-PhysNet: Learning the Intuitive Physics of Non-Rigid Object Deformations
- 3D-RecGAN - 3D Object Reconstruction from a Single Depth View with Adversarial Learning (github)
- ABC-GAN - ABC-GAN: Adaptive Blur and Control for improved training stability of Generative Adversarial Networks (github)
- ABC-GAN - GANs for LIFE: Generative Adversarial Networks for Likelihood Free Inference
- AC-GAN - Conditional Image Synthesis With Auxiliary Classifier GANs
- acGAN - Face Aging With Conditional Generative Adversarial Networks
- ACGAN - Coverless Information Hiding Based on Generative adversarial networks
- acGAN - On-line Adaptative Curriculum Learning for GANs
- ACtuAL - ACtuAL: Actor-Critic Under Adversarial Learning
- AdaGAN - AdaGAN: Boosting Generative Models
- Adaptive GAN - Customizing an Adversarial Example Generator with Class-Conditional GANs
- AdvEntuRe - AdvEntuRe: Adversarial Training for Textual Entailment with Knowledge-Guided Examples
- AdvGAN - Generating adversarial examples with adversarial networks
- AE-GAN - AE-GAN: adversarial eliminating with GAN
- AE-OT - Latent Space Optimal Transport for Generative Models
- .....
- Text2Shape - Text2Shape: Generating Shapes from Natural Language by Learning Joint Embeddings
- textGAN - Generating Text via Adversarial Training
- TextureGAN - TextureGAN: Controlling Deep Image Synthesis with Texture Patches
- TGAN - Temporal Generative Adversarial Nets
- TGAN - Tensorizing Generative Adversarial Nets
- TGAN - Tensor-Generative Adversarial Network with Two-dimensional Sparse Coding: Application to Real-time Indoor Localization
- TGANs-C - To Create What You Tell: Generating Videos from Captions
- tiny-GAN - Analysis of Nonautonomous Adversarial Systems
- TP-GAN - Beyond Face Rotation: Global and Local Perception GAN for Photorealistic and Identity Preserving Frontal View Synthesis
- TreeGAN - TreeGAN: Syntax-Aware Sequence Generation with Generative Adversarial Networks
- Triple-GAN - Triple Generative Adversarial Nets
- ....

Resources

Courses: cs231n.stanford.edu, cs224n.stanford.edu, Udacity PyTorch

News: deeplearningweekly.com deeplearning.ai/thebatch

Blogs: colah.github.io, karpathy.github.io

Podcast: MIT AI, Talking Machines

Tools: pytorch.org, tensorflow.org, keras.io, mxnet, spaCy, gensim, fastText.cc

Free GPU: Google Colab

Papers: arxiv.org