Recent AI

ECE30007 Intro to AI Project

outline

- Early Stories
- New Applications
- Issues

beginning of neural networks

- **Deep learning is all about deep neural networks.**

- **1949: Hebbian learning**
  - Donald Hebb: the father of neural networks
- **1958: (single layer) Perceptron**
  - Frank Rosenblatt
  - ← Marvin Minsky, 1969
- **1986: Multilayer Perceptron (Backpropagation)**
  - David Rumelhart, Geoffrey Hinton, and Ronald Williams
  - ← Vladimir Vapnik and Corinna Cortes, 1995 SVM
- **2006: Deep neural networks**
  - Geoffrey Hinton and Ruslan Salakhutdinov

cybernetics

↓

connectionism (PDP)

↓

deep learning


D. Hebb (1904-1985)


F. Rosenblatt (1928-1971)


M. Minsky (1927-)


D. Rumelhart (1942-2011)


V. Vapnik

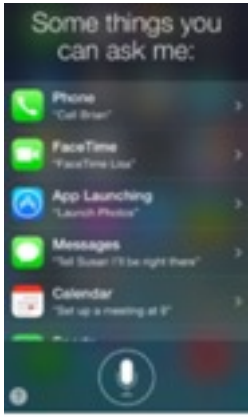
- In the 1960s, some believed that a workable AI system was just 10 years away.
- In the 1980s, a wave of commercial start-ups collapsed ("AI-winter") (Nov. 24, 2012, NYT)

AI on the way

- computer is getting smarter.
  - IBM Watson's victory on Jeopardy! (2011)
  - Apple Siri's speech understanding (2011)
  - Google Now's intention recognition (2012)

Deep Blue (IBM, May 11, 1997)

IBM Watson



Apple Siri

Google Now

on media

- TV shows and newspapers on deep learning!

The New York Times

How Many Computers to Identify a Cat? 16,000

Jun. 25, 2012

Scientists See Promise in Deep-Learning Programs

Nov. 23, 2012

WIRED

- Researcher dreams up machines that learn without humans (Jun. 27, 2013)

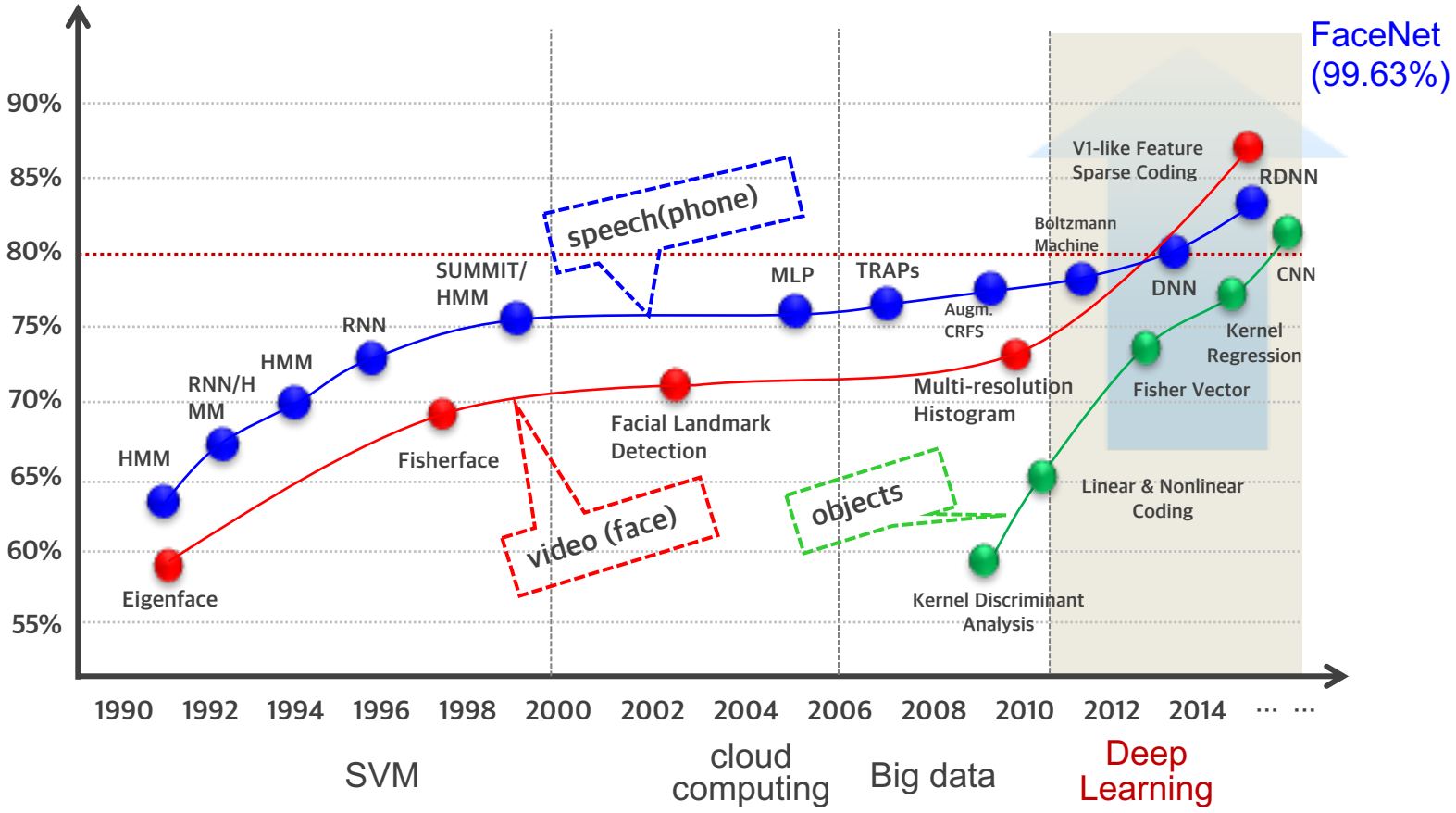
THE GLOBE AND MAIL

CANADA'S NATIONAL NEWSPAPER • FOUNDED 1844

- Google taps U of Toronto professor to teach context to computers (Mar. 11, 2013)

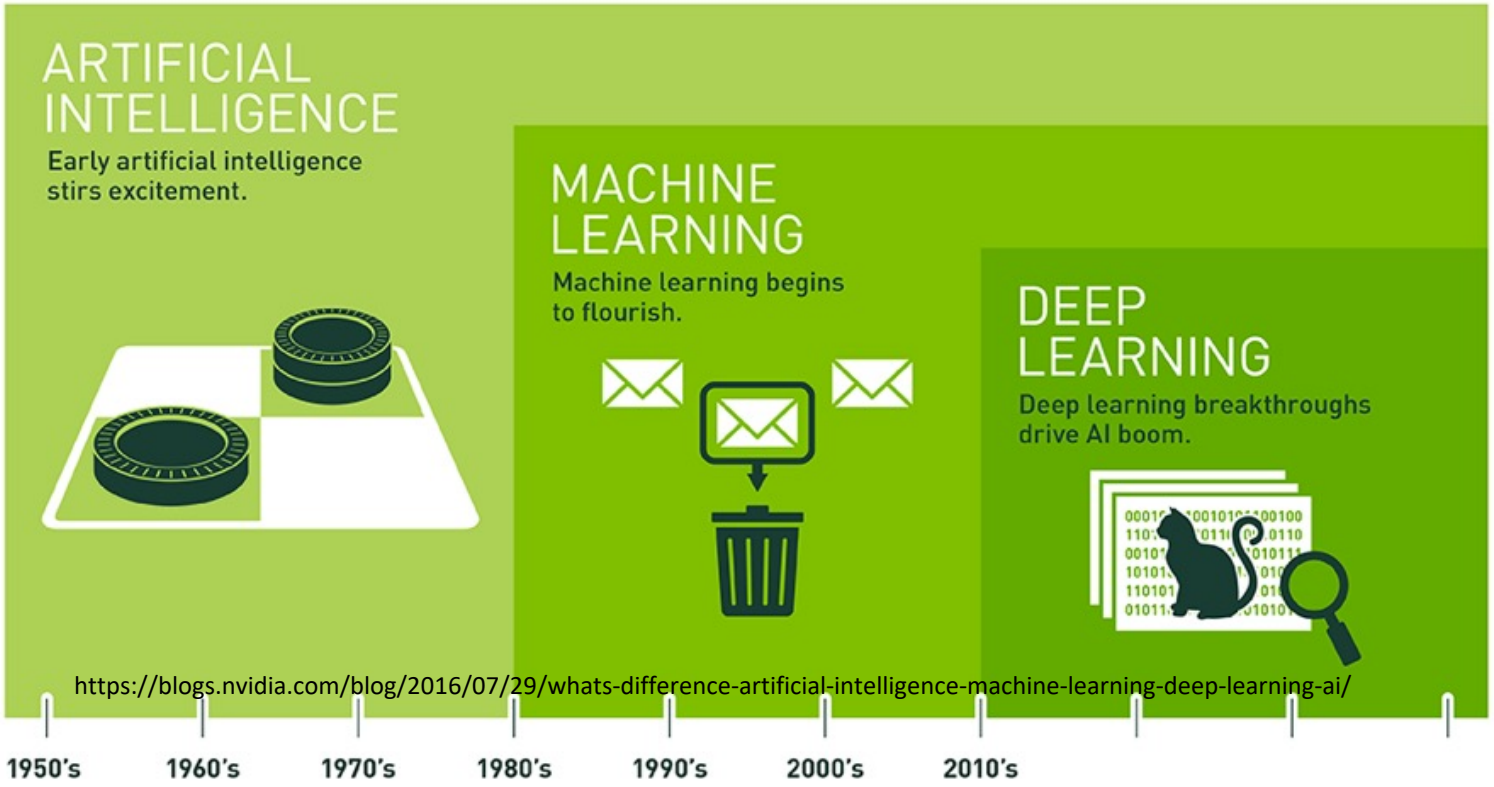
trends in pattern recognition

- upheaval in pattern recognition due to deep learning



recent advances in AI are attributed to deep learning

- why deep learning (DL)?
  - to understand complex problems, our model should be powerful enough.
  - DL is expressive and generalizing well (distributed representation)



competitions

- German traffic sign recognition competition:
  - deep learning outperformed human visual system (IDISA `11).
- Merck Drug Discovery
  - A deep learning group won the contest (U of Toronto, `12).

German traffic sign recognition competition

- single-image, multi-class
- more than 40 classes
- more than 50K images in total
- large, lifelike database

rank	team	method	accuracy (%)
1	IDSIA	Committee of CNNs	99.46
2	INI	Human performance	98.84
3	Sermanet	Multi-scale CNNs	98.31
4	CAOR	Random Forests	96.14

2012 New drug discovery contest

- Detect a new potential molecule structures for new drugs

- enter the contest at the last minute
- relatively small set of data

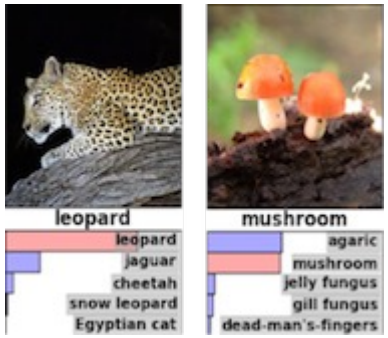
object recognition

- object recognition on images
- Deep learning made a big progress compared to hand-crafted feature extraction.

ImageNet

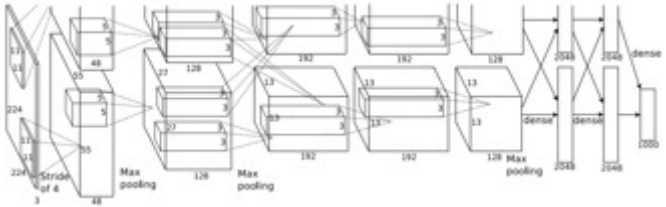
- ImageNet data set
  - 22K categories
  - 15M labeled images

Recognition task
(Top-5)



network structure

253440–186624–64896–64896–43264–4096–4096–1000



- 8 layers: 650K neurons, 60M parameters

- Trained on 2 GPUs
- 5-6 days of training (90 iterations)

Trained filters

- 1000/10,184 categories
- 1.2M/8.9M training images
- 50K validation
- 150K test

(Krizhevsky et al 2012)

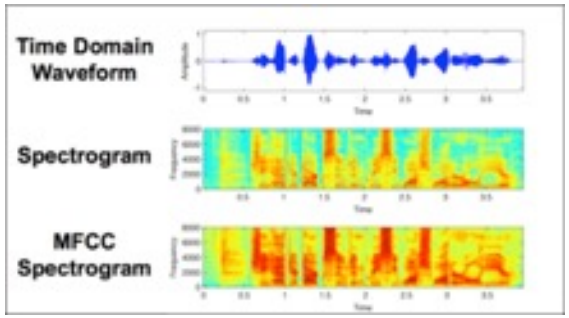
results on ILSVRC-2010/ImageNet2009 (error %)

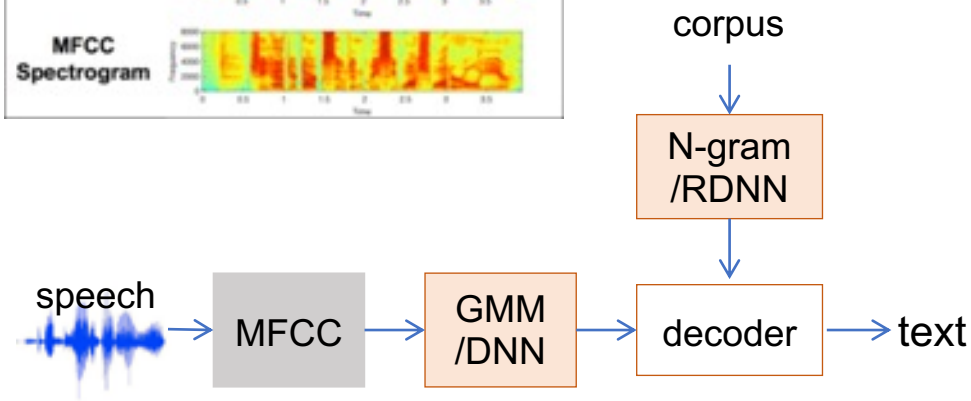
	previous SOTA	deep learning (CNNs)
top-1	45.7 / 78.1	37.5 / **67.4**
top-5	25.7 / 60.9	17.0 / **40.9**

3.57% (as of 2015)

speech recognition

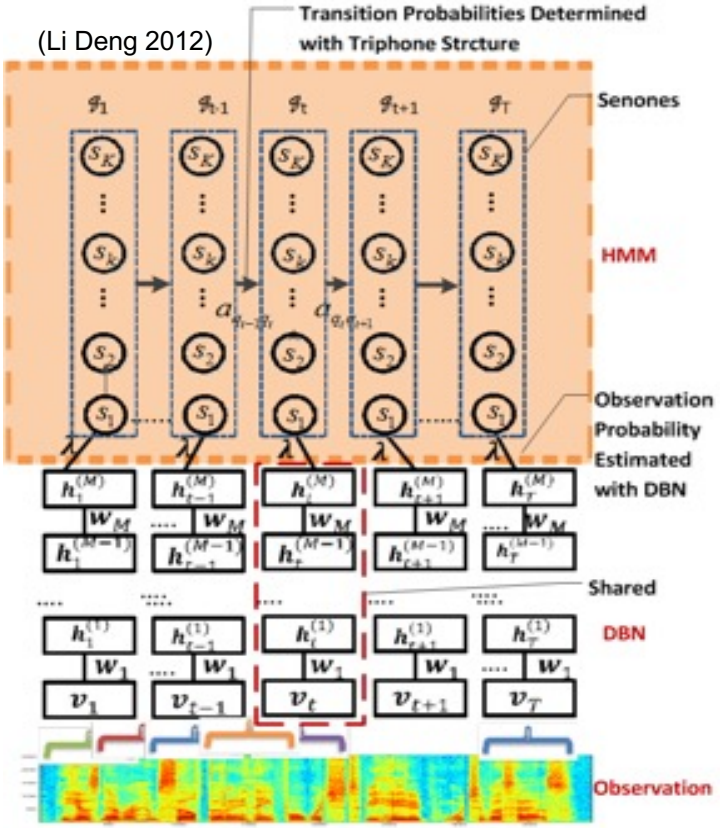
- In speech recognition systems, the phoneme recognition module was implemented with deep learning. (Google, `12)
- Recurrent deep neural network has been proposed. (`13)





GMM: P(o|s)
DNN: P(s|o)

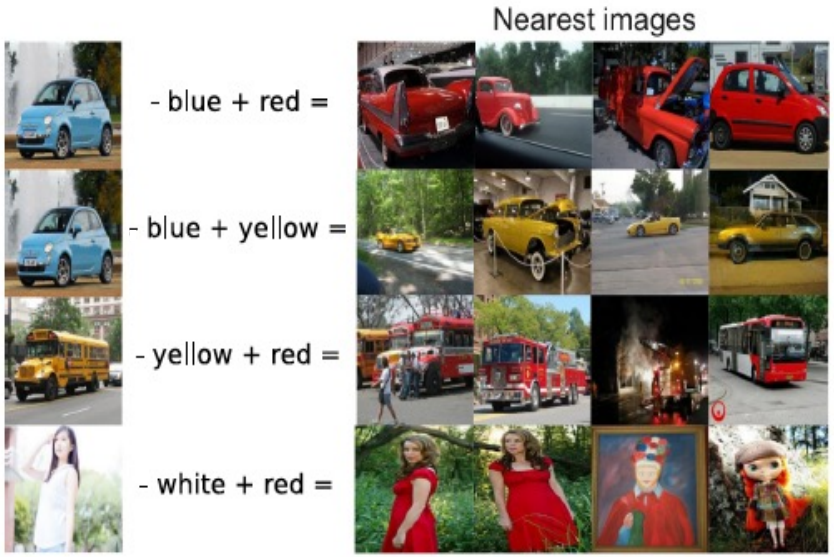
5-gram: $P(w_t|w_{t-1:t-4})$
RDNN: $P(w_t|w_{t-1}, \ldots w_1)$



New Applications

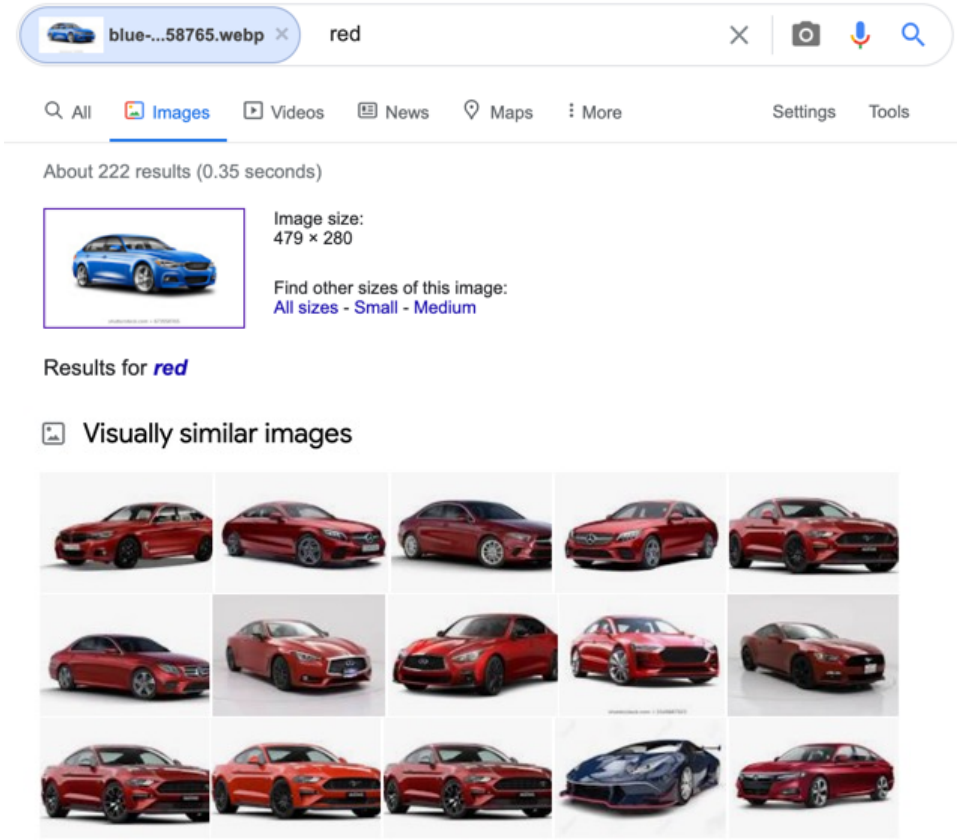
algebra with image and text

- data regularities are captured in multimodal vector space. (cf. grounding)
  - algebra is possible in the space.

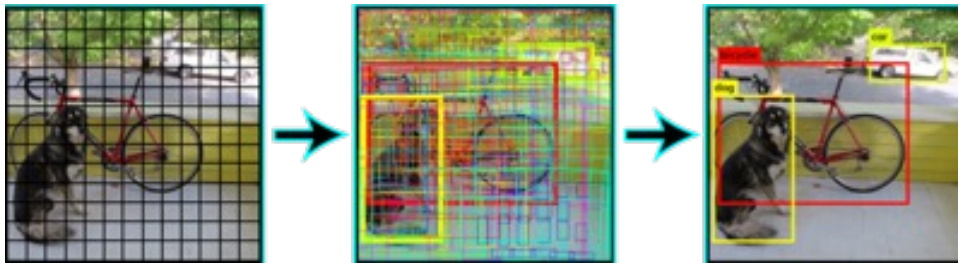


from (Kiros 2014)

only possible in a multimodal representation
(in a Euclidean space)



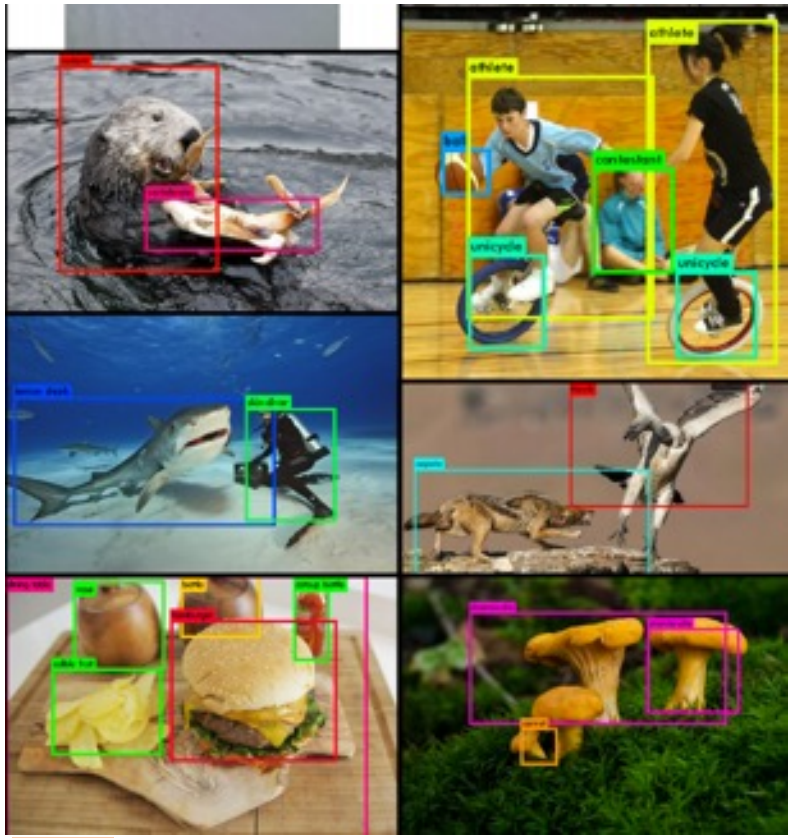
real-time object detection



Detection Frameworks	Train	mAP	FPS
Fast R-CNN [5]	2007+2012	70.0	0.5
Faster R-CNN VGG-16[15]	2007+2012	73.2	7
Faster R-CNN ResNet[6]	2007+2012	76.4	5
YOLO [14]	2007+2012	63.4	45
SSD300 [11]	2007+2012	74.3	46
SSD500 [11]	2007+2012	76.8	19
YOLOv2 288 × 288	2007+2012	69.0	91
YOLOv2 352 × 352	2007+2012	73.7	81
YOLOv2 416 × 416	2007+2012	76.8	67
YOLOv2 480 × 480	2007+2012	77.8	59
YOLOv2 544 × 544	2007+2012	**78.6**	40

You Only Look Once (YOLO)

https://pjreddie.com/darknet/yolo/



click

(Redmon and Farhadi 2016)

game

- game with reinforcement learning

Atari Breakout

Supermario

AlphaStar

DeepMind's StarCraft-playing AI beats 99.8 per cent of human gamers

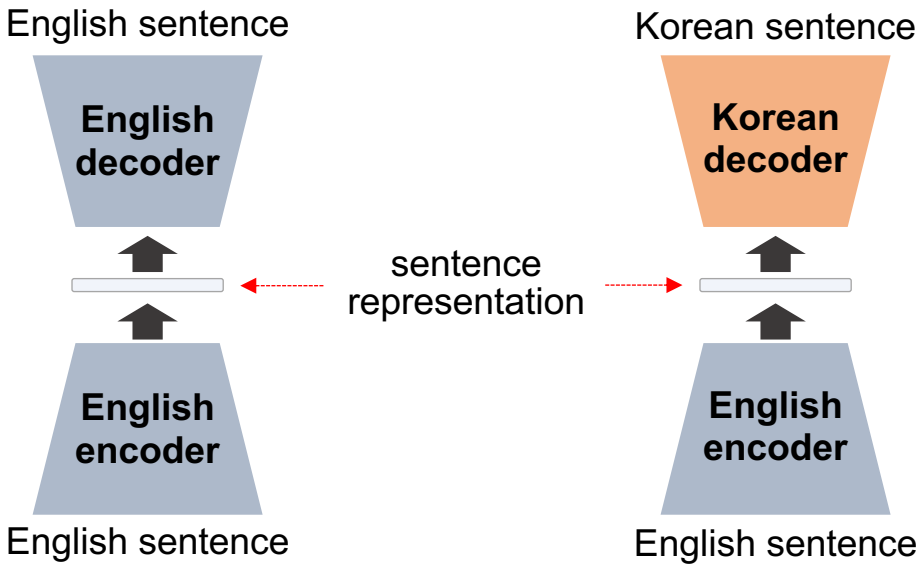
TECHNOLOGY 30 October 2019

By Donna Lu

neural machine translation (NMT)

- machine translation (MT) as encoding/decoding.
  - there is a sentence representation (bottleneck) layer.

- end-to-end neural networks
  - encodes the variable length input sentence into a fixed dimensional vector
  - decodes the representation to the output sentence.



(Bengio 2014)

data for NMT

e.g., English-German

unique words: 709K vs. 1,553K
total words: 117M vs. 110M
lines: 4,507,701 (4.5M)

Eng

Moreover , the Santa Lucia railway station is just 5 minutes away while other major sights such as the Rialto Bridge and St. Mark 's Square can quickly and easily be reached with a 15 to 20 minutes ' walk .

Have you already thought over how to present this holiday to your sweetheart this time ?

I come from the Prievidza region , which has a strong mining tradition .

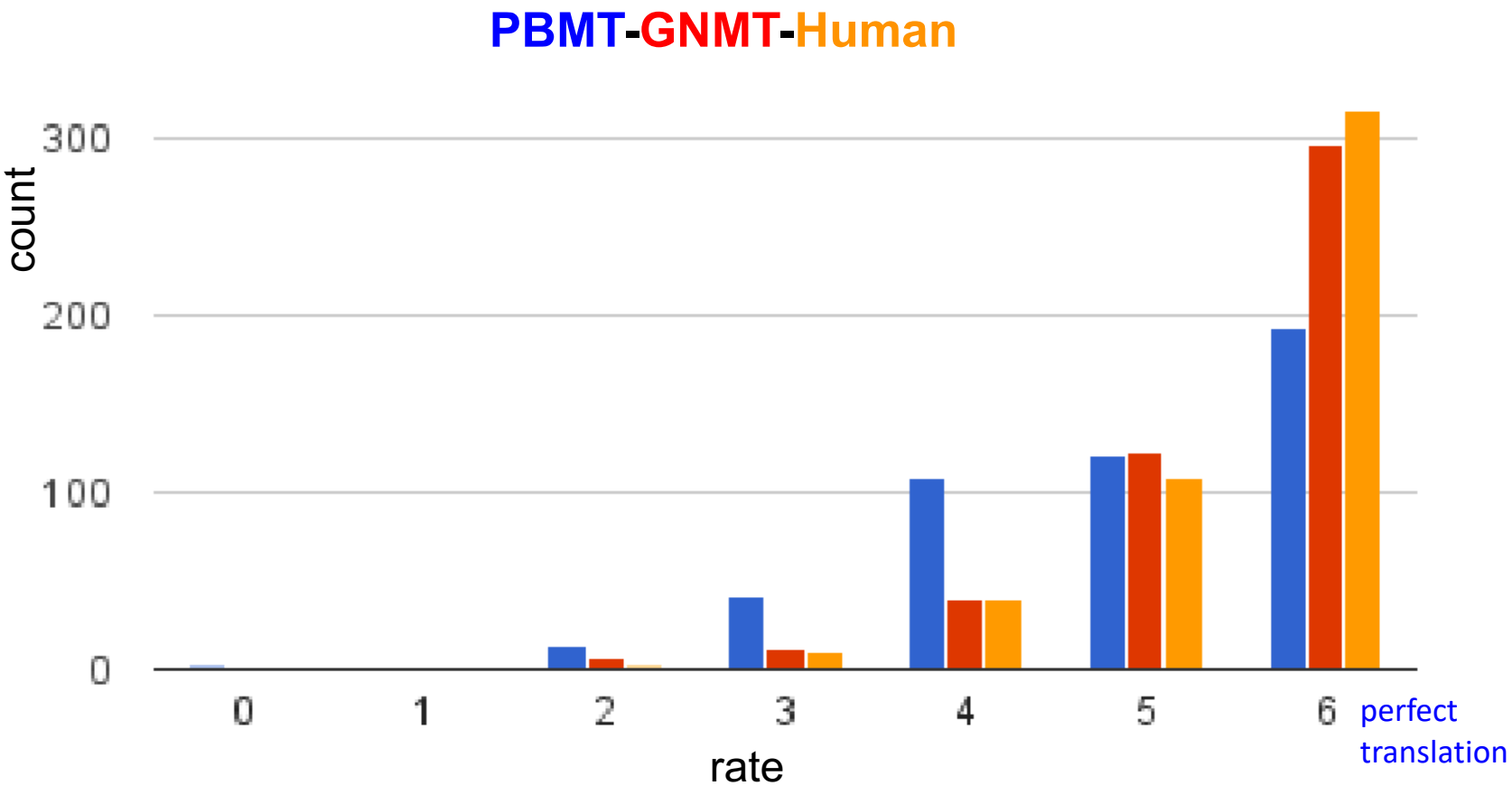
German

Ausserdem befindet sich der Santa Lucia Bahnhof in nur 5 Minuten Entfernung , während andere wichtige Sehenswürdigkeiten , wie die Rialto Brücke und der San Marco Platz schnell und einfach in einem 15- bis 20-minütigen Spaziergang erreicht werden können .

Eine langeZeit war die Stadt unter dem Einfluss von vielen Imperien und Völker .

Ich komme aus der Region Prievidza , in der es eine starke Tradition des Bergbaus gibt .

Google NMT performance



(Wu et al, 2016)

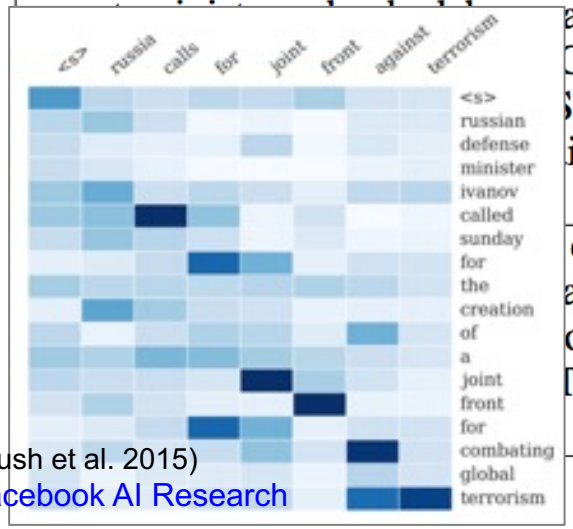
summarization is kind of translation!

From text to news headline using the same model as the NMT

(Lopyrev 2015) Stanford

Text 1st paragraph	Actual Headline	Predicted Headline
1. At least 72 people died and scores more were hurt when a truck crowded with pilgrims plunged into a gorge in the desert state of Rajasthan on Friday, police told the press trust of India.	Urgent: truck crashes killing 72 pilgrims in India	At least 72 dead in Indian road accident
2. Sudanese president Omer Al-Bashir has announced his refusal of discharging a govern- accused by the CC) of commit- Sudanese region ily reported on	Sudanese president refuses to discharge state minister indicted by ICC	Sudanese president refuses to of alleged war crimes
ousted Taliban ama Bin Laden of condolences, Tuesday on Al-	Taliban leader says Bin Laden still alive	Urgent: Bin Laden alive, says Taliban chief

Gigaword corpus

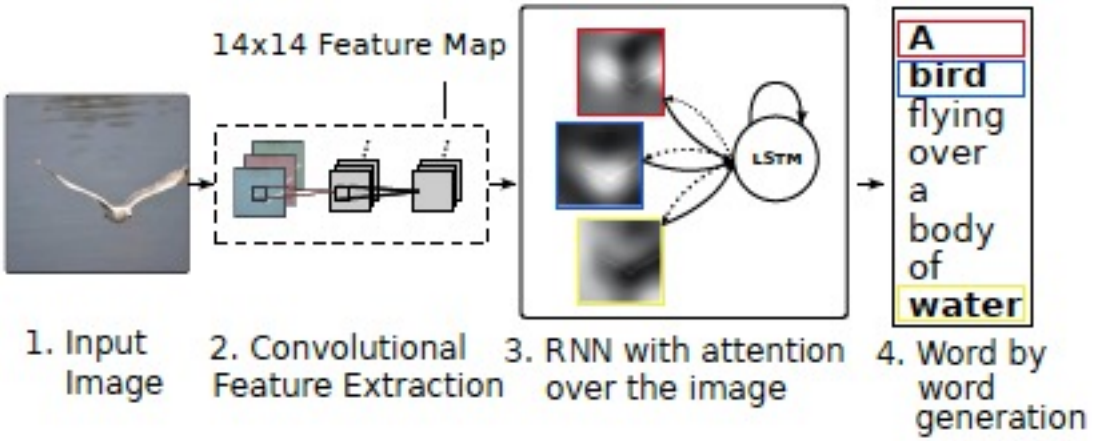


(Rush et al. 2015)
Facebook AI Research

auto caption generation

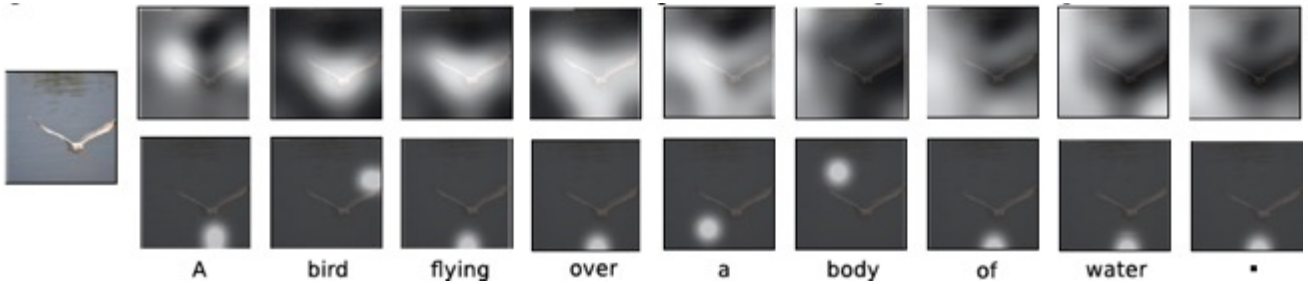
- Attention lets you know where to look at.

show attend and tell



A stop sign is on a road with a mountain in the background.

A giraffe standing in a forest with trees in the background.



(Xu et al 2015)

qustion and answering system

- based on text

Task 1: Single Supporting Fact
Mary went to the bathroom.
John moved to the hallway.
Mary travelled to the office.
Where is Mary? A:office

Task 2: Two Supporting Facts
John is in the playground.
John picked up the football.
Bob went to the kitchen.
Where is the football? A:playground

bAbi dataset

- based on image or figure

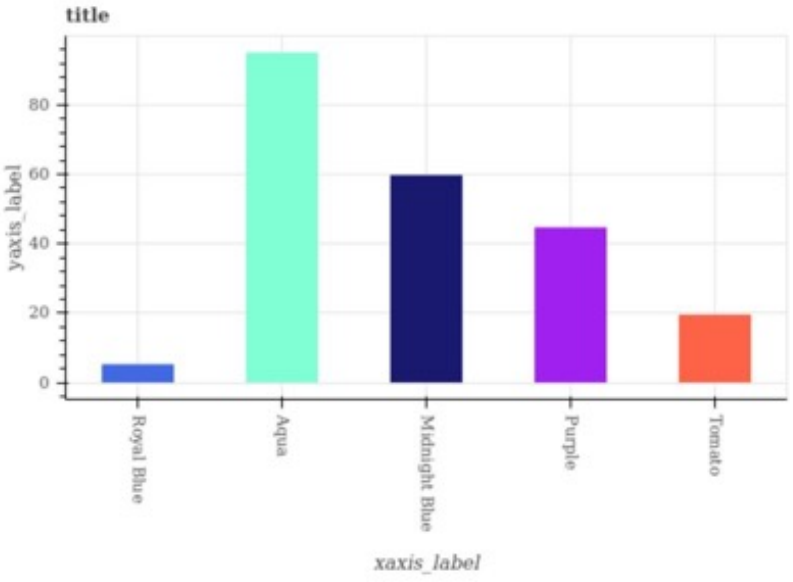
Where is the child sitting?
fridge

arms


How many children are in the bed?
2

1


from www.visualqa.org



Q: Is Aqua the maximum?
A: Yes

Q: Is Midnight Blue greater than Aqua?
A: No

Q: Is Midnight Blue less than Aqua?
A: Yes

Q: Is Purple the high median?
A: Yes

Q: Is Tomato the low median?
A: No

FigureQA from Maluuba

deep writing

A JAPANESE AI PROGRAM JUST WROTE A SHORT NOVEL, AND IT ALMOST WON A LITERARY PRIZE

By **Chloe Olewitz** — Updated March 24, 2016 11:30 am

https://www.digitaltrends.com/cool-tech/japanese-ai-writes-novel-passes-first-round-nationanl-literary-prize/

Harry Potter: Written by Artificial Intelligence

I trained an LSTM Recurrent Neural Network (a deep learning algorithm) on the first four Harry Potter books. I then asked it to produce a chapter based on what it learned. Here's the chapter. (I added a bit of formatting to aid readability)

https://medium.com/deep-writing/harry-potter-written-by-artificial-intelligence-8a9431803da6

Sunspring | A Sci-Fi Short Film Starring Thomas Middleditch - YouTube

https://www.youtube.com/watch?v=LY7x2Ihqjmc

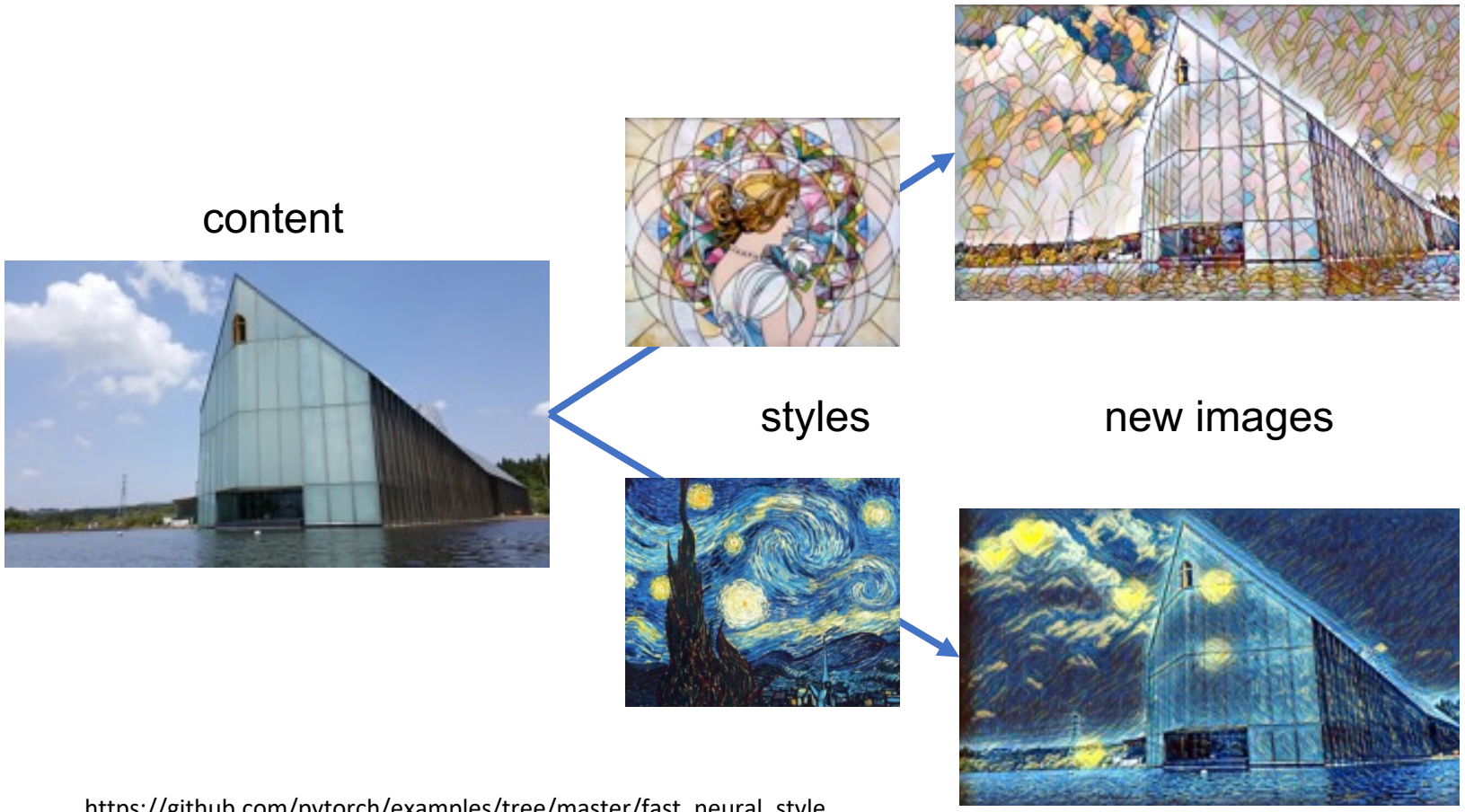
Jun 9, 2016 - Uploaded by Ars Technica Videos

In the wake of Google's AI Go victory, filmmaker Oscar Sharp turned to his technologist collaborator Ross ...

click

deepart

- **style and content representations are separable**



https://github.com/pytorch/examples/tree/master/fast_neural_style

(Gatys et al 2015)

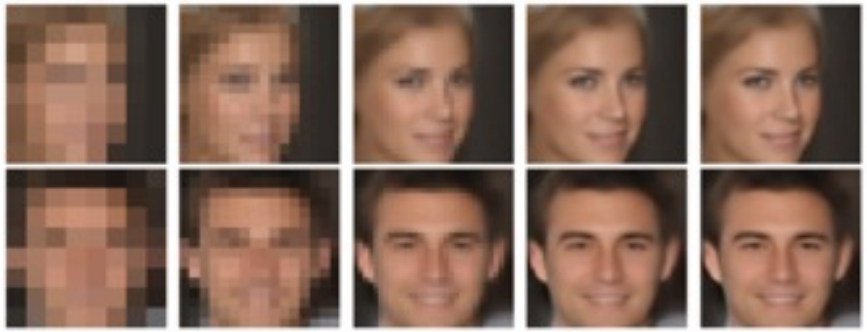
changing attributes of face

Interpolation between different attributes

(Lample et al., 2017)

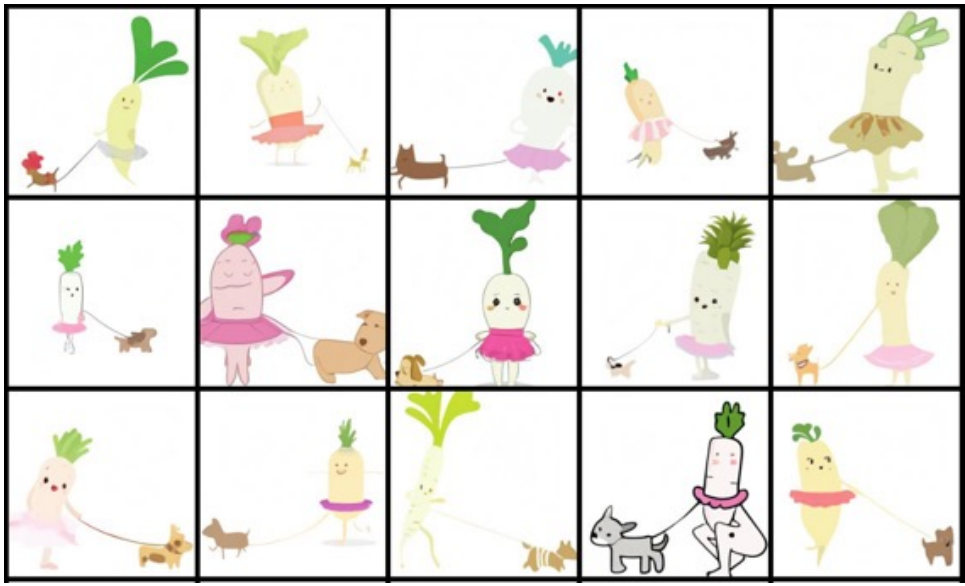
image generation

- super resolution or adding color to gray image
- translation from text to image



(Kolesnikov and Lampert 2017)

“A baby daikon radish in a tutu walking a dog”



deep learning and self-driving

How AI is Making Self-Driving Cars Smarter

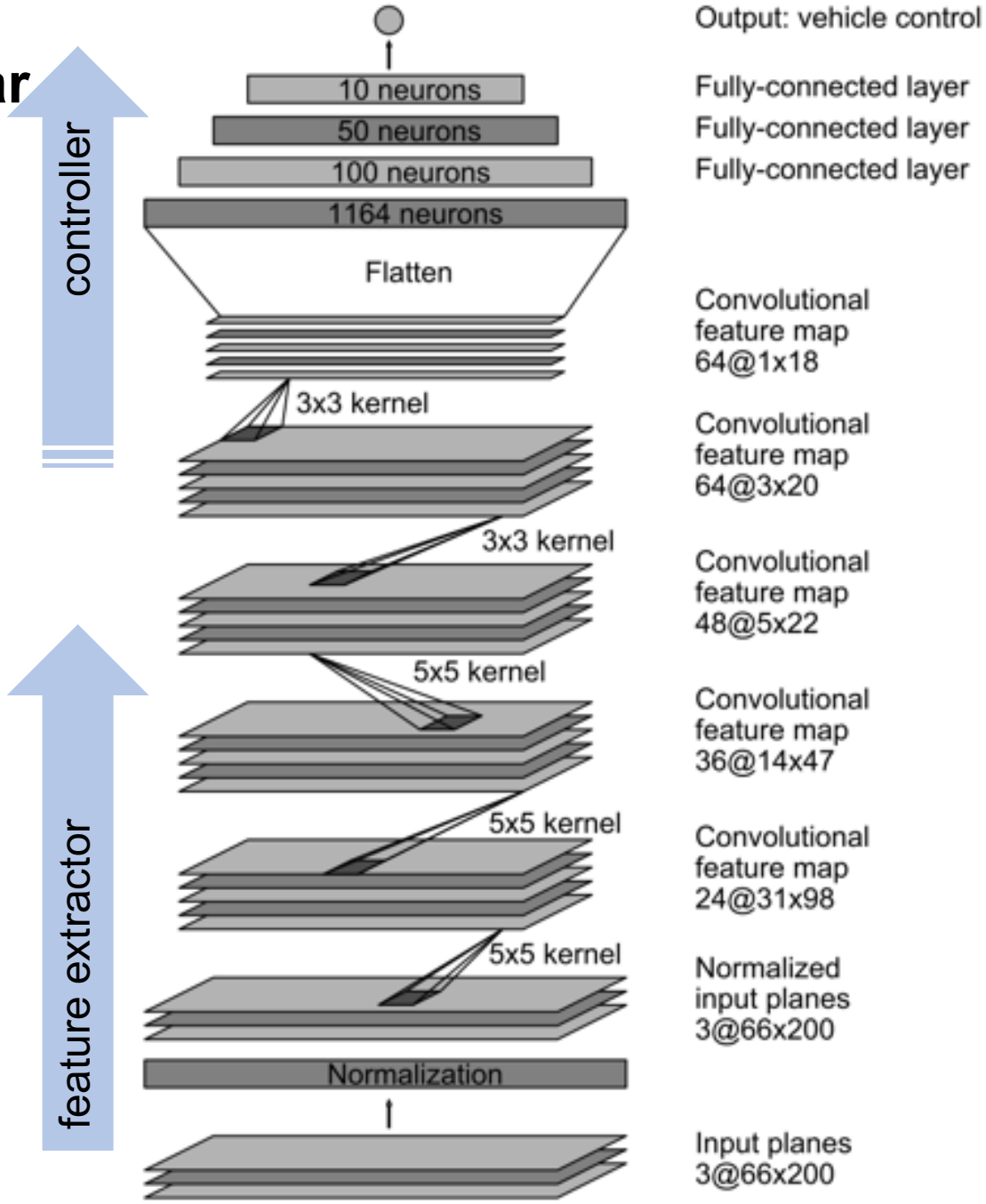
The stage is set for artificial intelligence to dominate our roads. Here's how artificial intelligence is improving self-driving cars.

https://blogs.nvidia.com/blog/2016/05/06/self-driving-cars-3/

CNNs in self-driving car

CNN architecture.

- 9 layers
- 27M connections
- 250K parameters



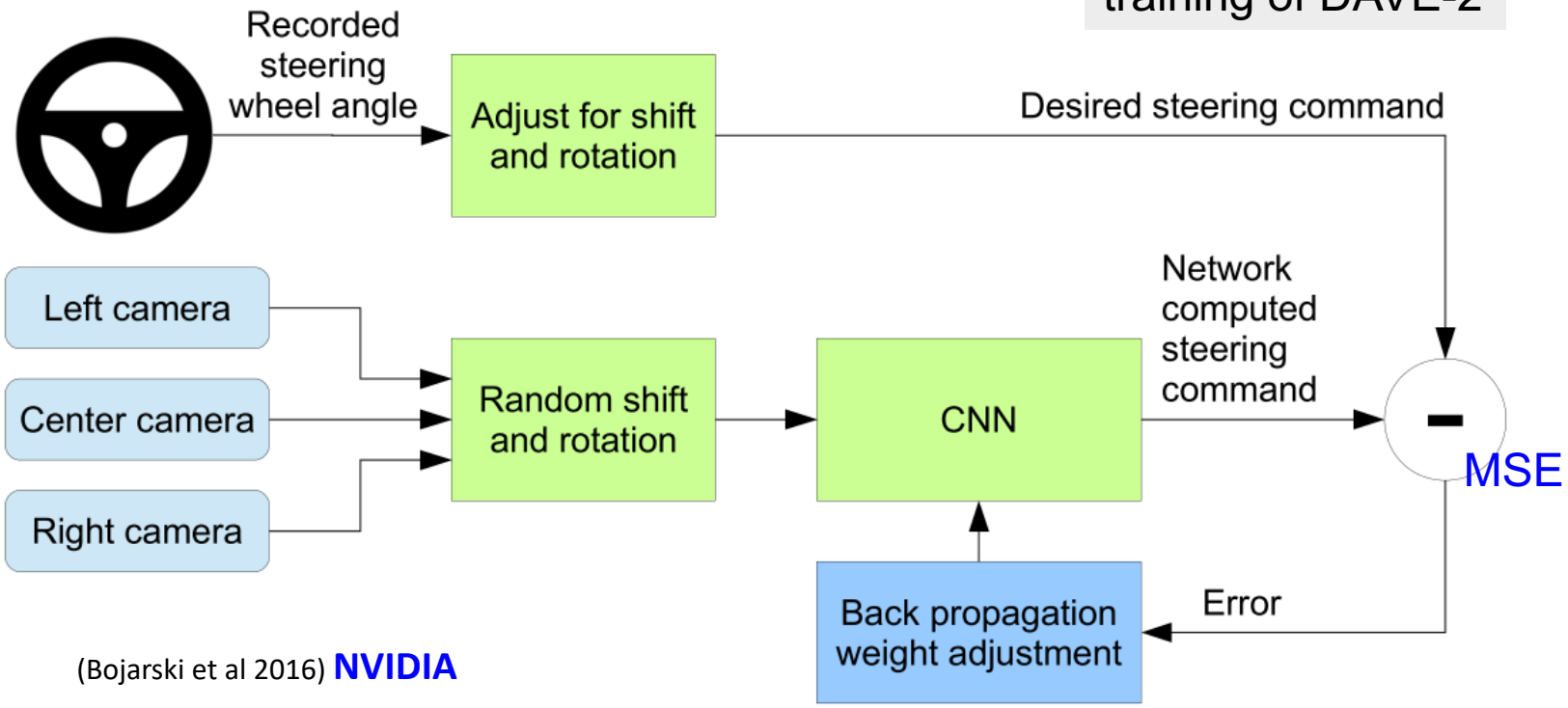
(Bojarski et al 2016)

deep learning in self-driving car

- supervisor: human driver
- input: camera images
- output: wheel steering (and braking)

avoid “if, then, else” rules

training of DAVE-2



(Bojarski et al 2016) **NVIDIA**

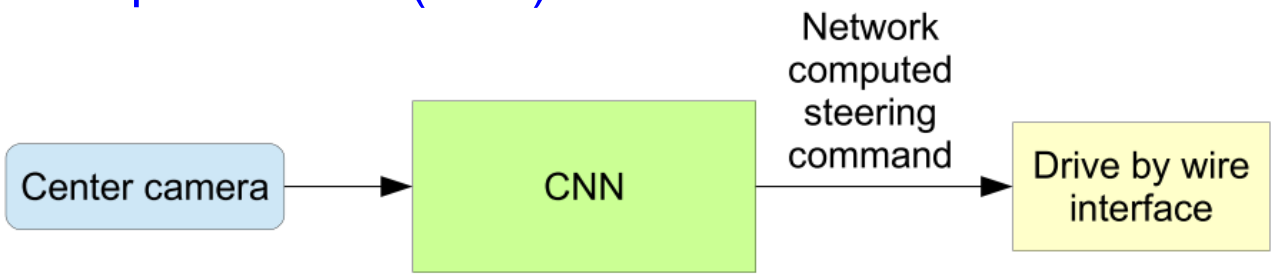
CNNs in self-driving car

- supervisor: human driver
- input: camera images
- output: wheel steering (and braking)

avoid “if, then, else” rules

testing of DAVE-2

30 frames per second (FPS)



(Bojarski et al 2016)

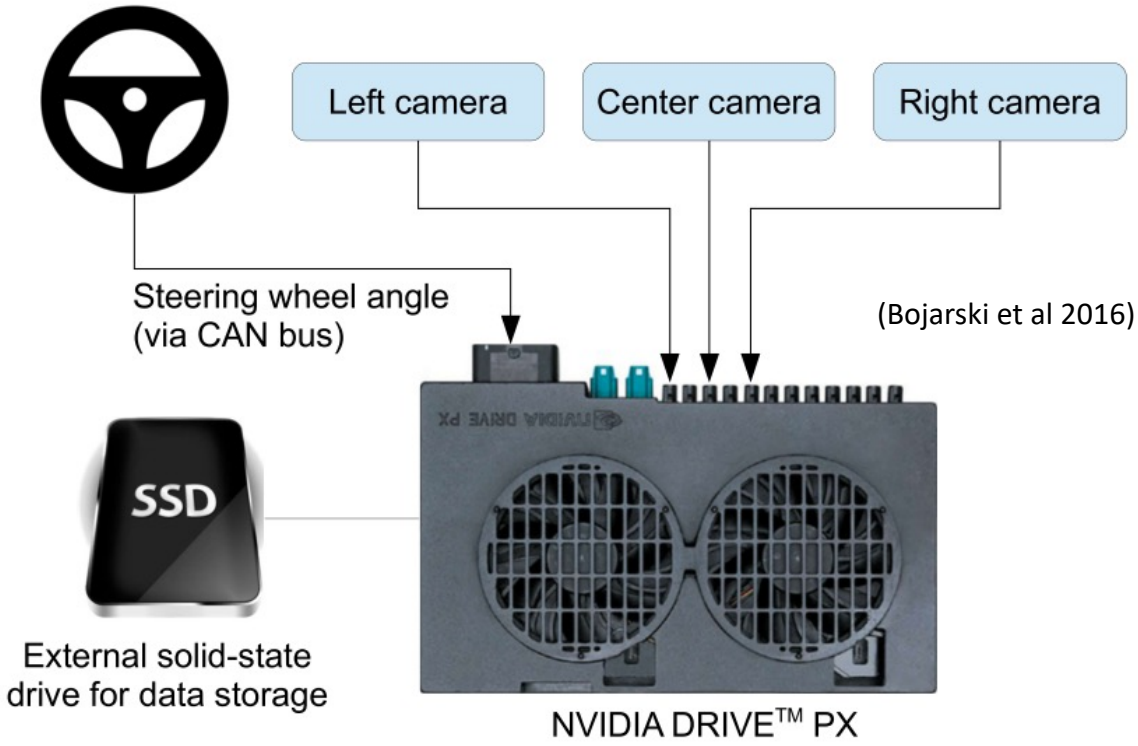
data collection

- a wide variety of roads a diverse set of lighting and weather conditions

- two-lane roads with and without lane markings
- residential roads with parked cars, tunnels, unpaved roads
- clear, cloudy, foggy, snowy, and rainy weather
- both day and night.

- 72 hours
  (as of March 28, 2016)

in training, use data
only staying in a lane



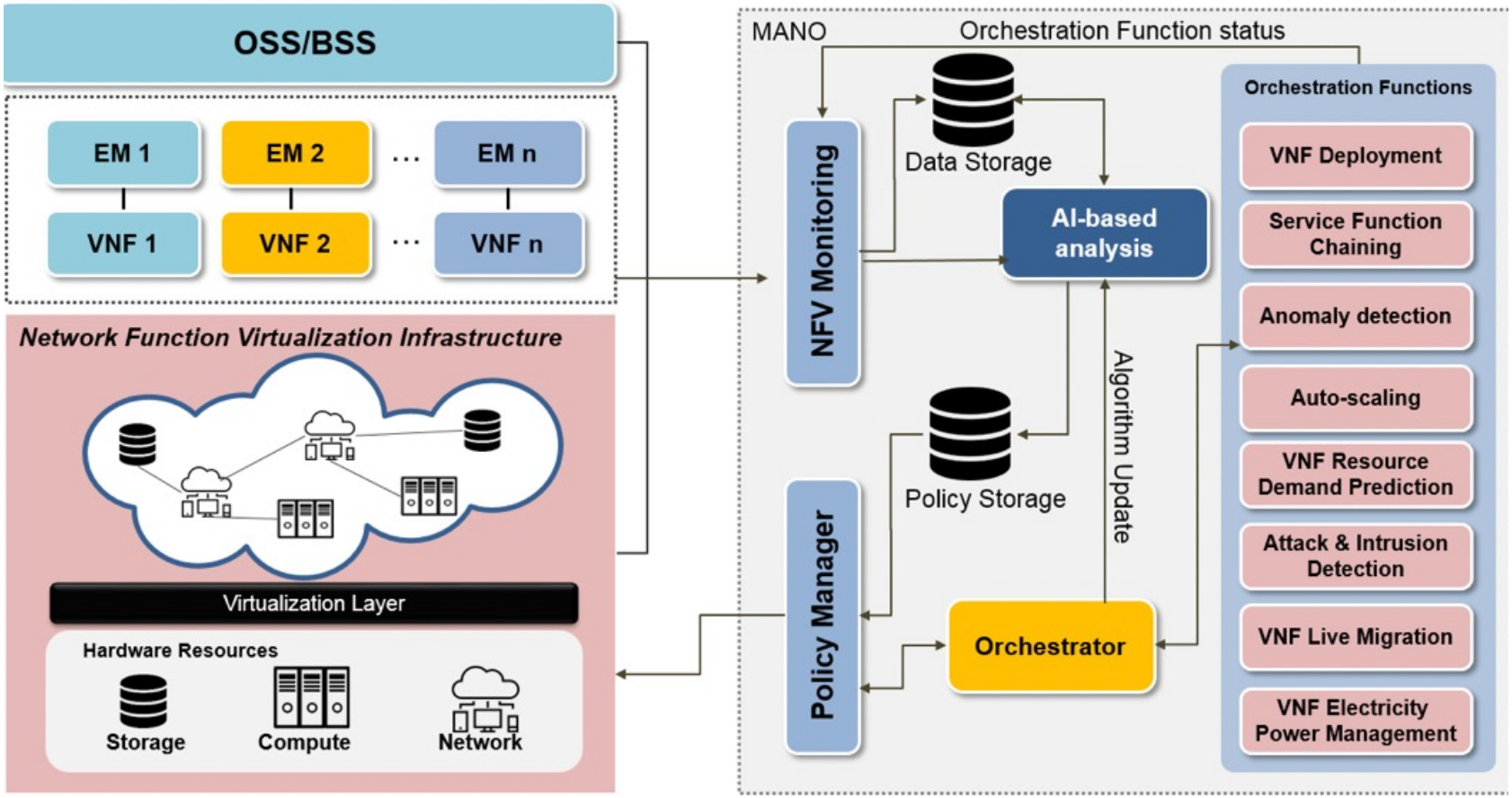
(Bojarski et al 2016)

AI for smart factory

images from the web

AI for the internet network management

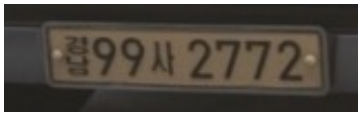
Artificial intelligence to manage virtual network functions



video analysis for container

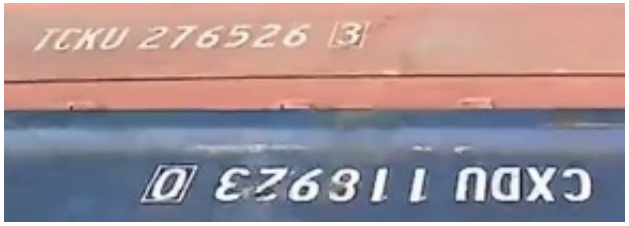
License Plate

LP Type 1



LP Type 2

Container ID



Tag Recognition

issues

model compression

- light model for deep learning on embedded system or smartphones

speech recognition on mobile phone

image recognition on mobile phone

images adapted from the web

issues in self-driving car

Researchers Find a Malicious Way to Meddle with Autonomous Cars

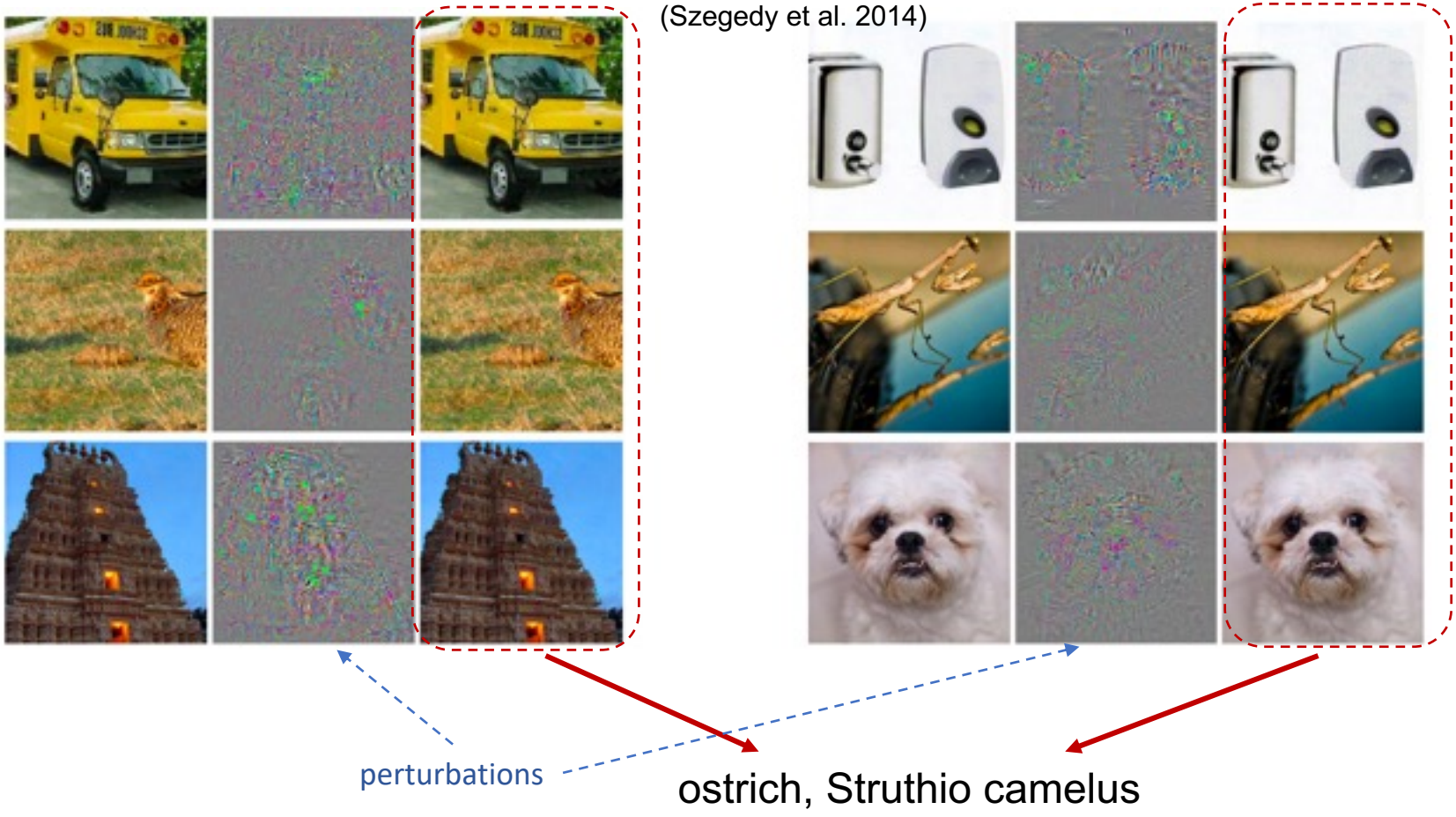
AUGUST 4, 2017 AT 11:06 AM BY MARK HARRIS | PHOTOGRAPHY BY UNIVERSITY OF WASHINGTON

SHARE TWEET

with some noise, they are all recognized as speed limit 45!

adversarial examples

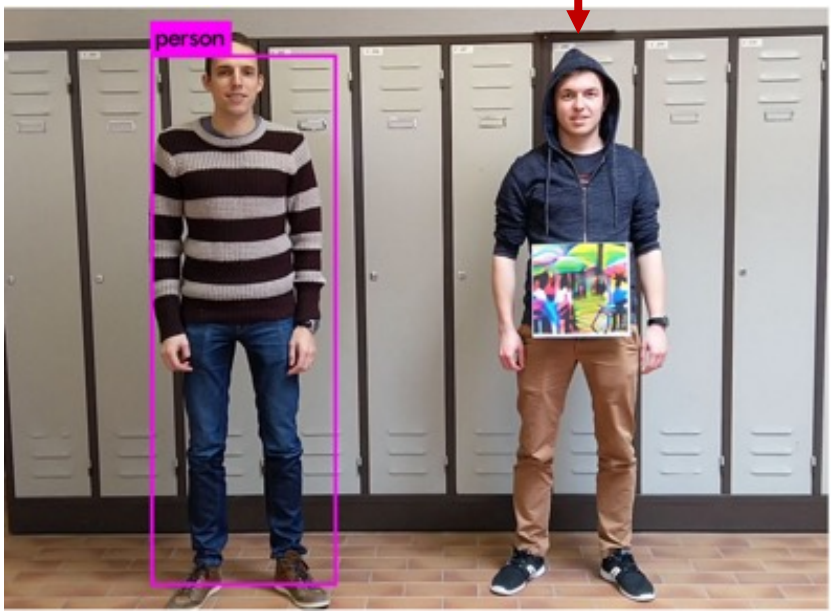
after adding some perturbation, they become all ostrich!

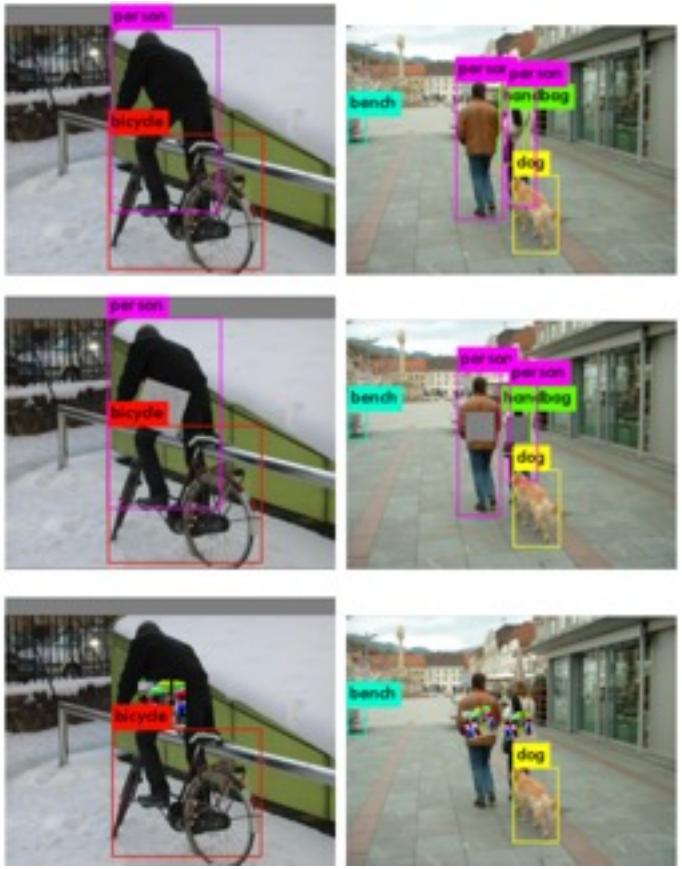


adversarial patches

"Fooling automated surveillance cameras: adversarial patches to attack person detection" (Thys et al, 2019)

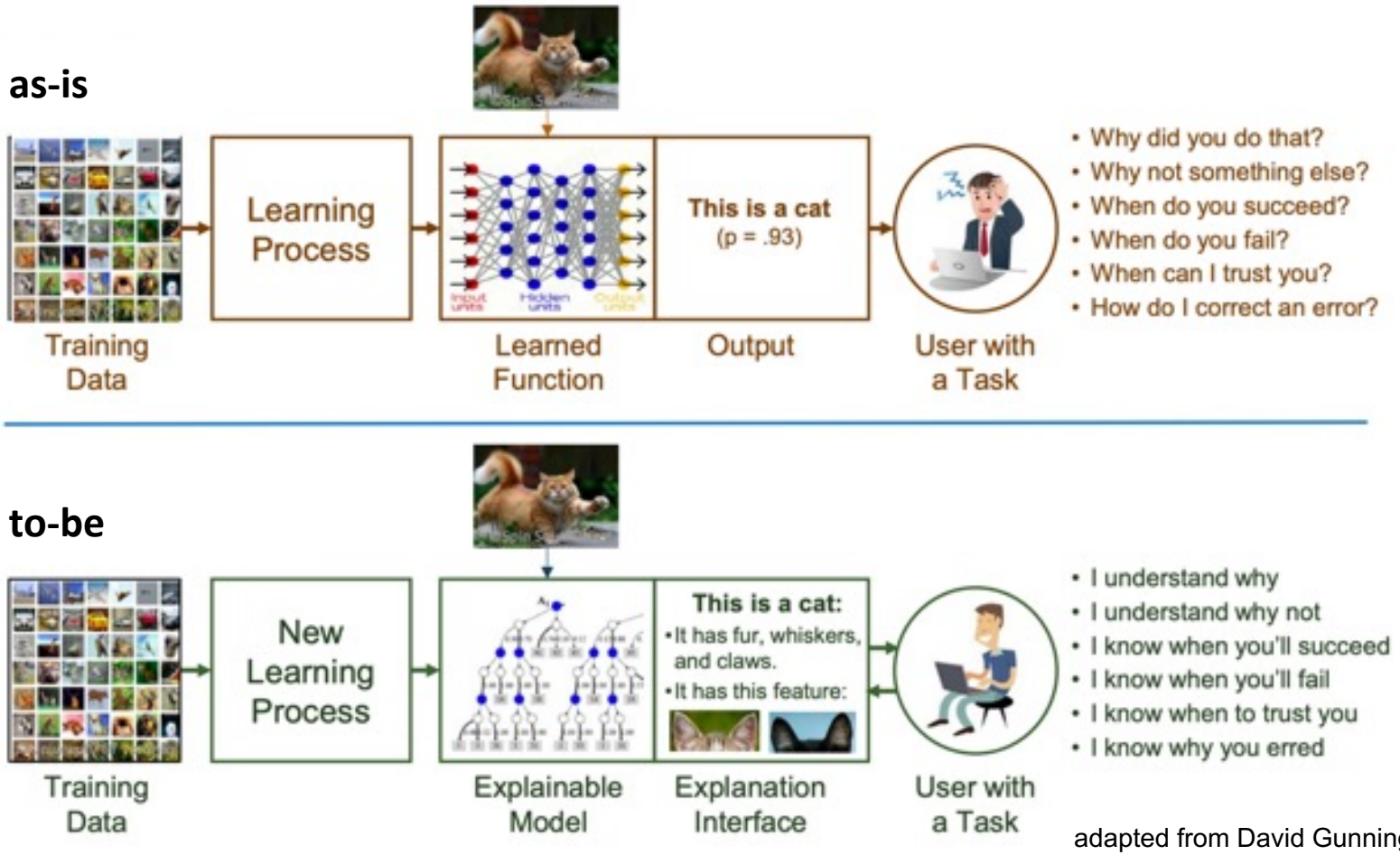
model cannot detect this guy!





explainable AI (XAI)

- explosion of AI applications with great success
  - XAI is essential for users to trust the autonomous AI



adapted from David Gunning's slide

meta learning

- meta learning is learning to learn.
  - if you have learned lots of tasks, then you can learn for a new task easily.
- one of essential features of AGI

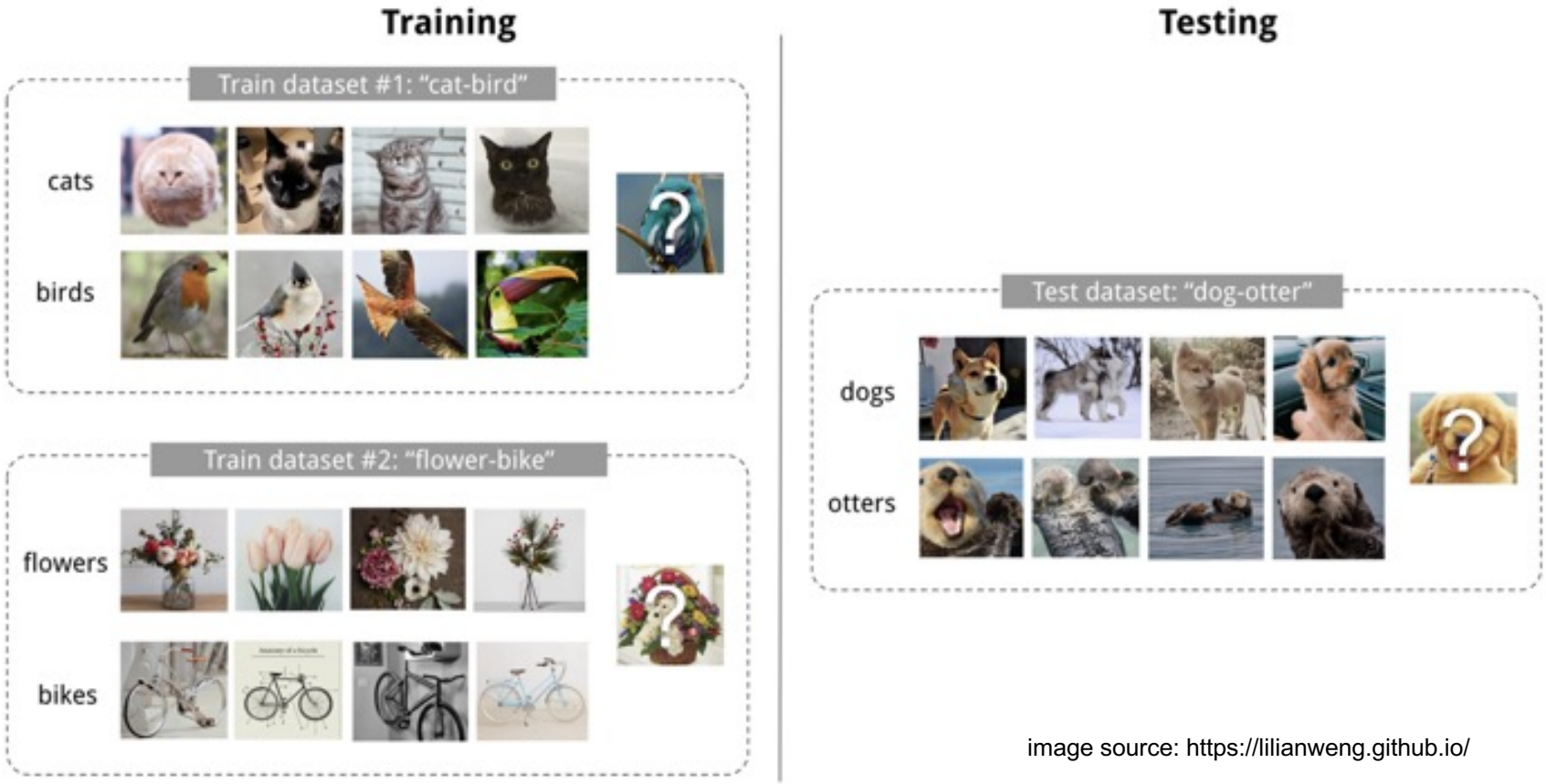


image source: https://lilianweng.github.io/