Deep Learning - Part 2

Aymeric DIEULEVEUT

DS4M - April 2025

Outline

CNN

We consider an image classification task: we want to recognize which object is on an image.

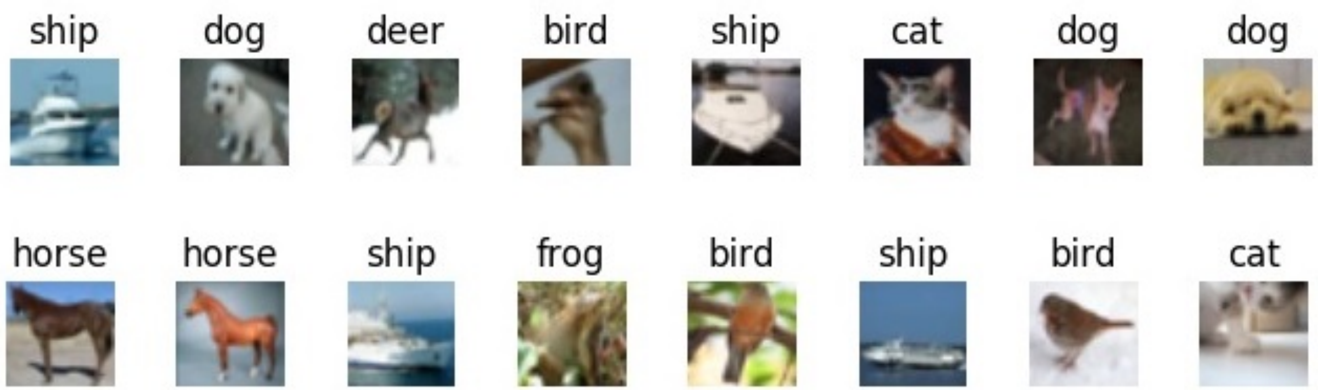


Figure: CIFAR dataset

The input is an image: about 10^3 to 10^7 pixels: these are our inputs.
Problems: The multi-layer perceptron:

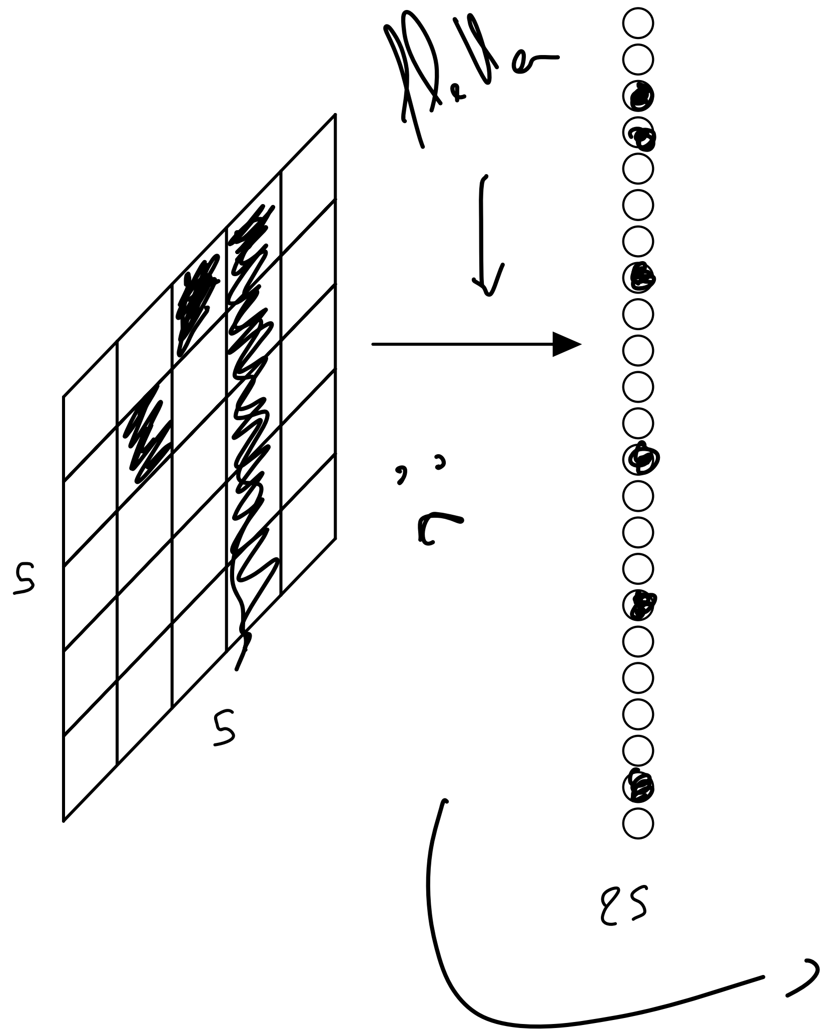
- Does not take into account local information. It would work similarly on any permutation of the pixels!
- Has, just on the first layer, as many parameters as there are inputs!

Convolutional Layer

Instead of making a linear combination of all the pixels, we consider a weighted average of a moving window of pixels, of size 3x3, or 5x5...

Convolutional Layer

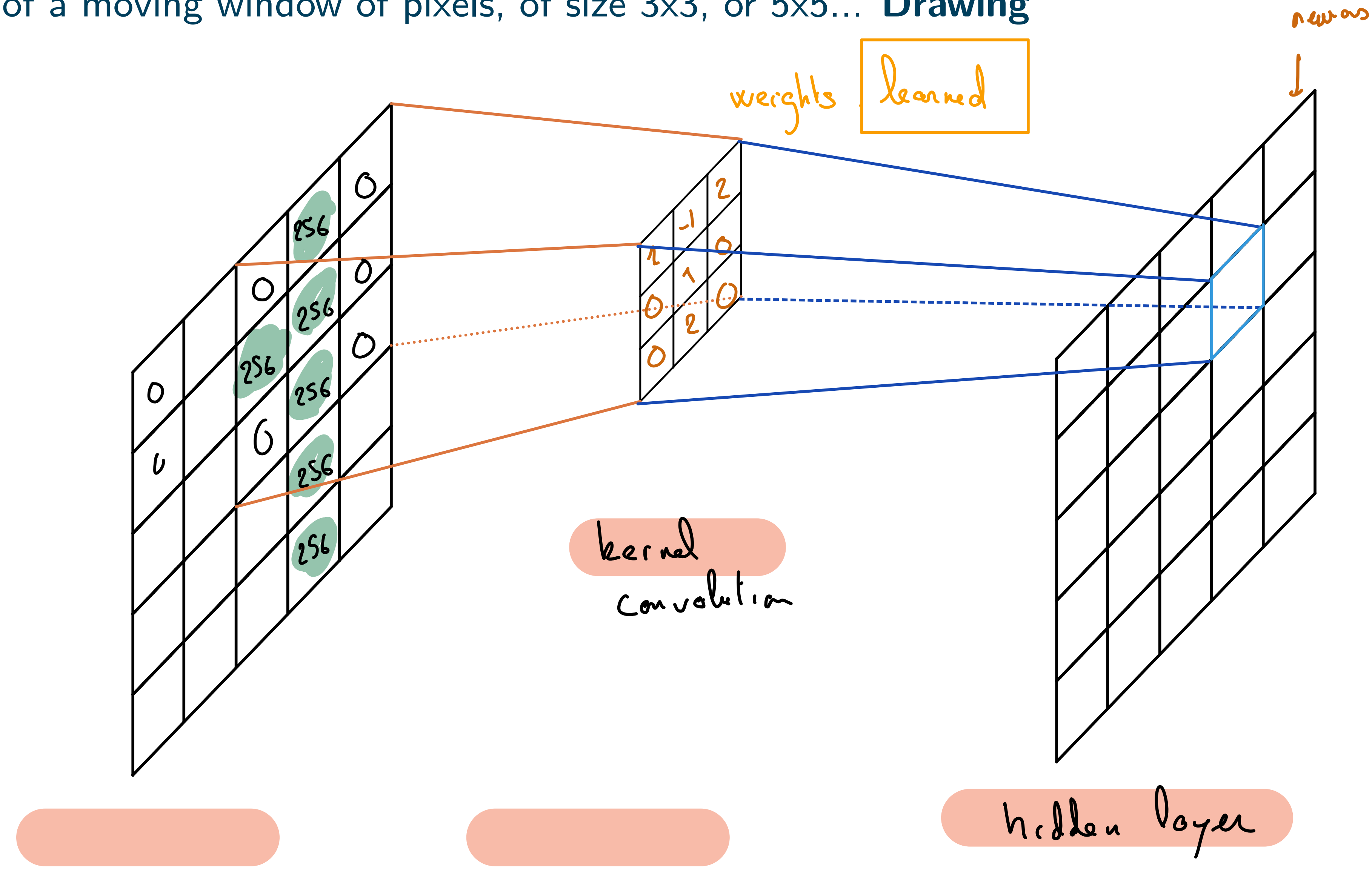
Instead of making a linear combination of all the pixels, we consider a weighted average of a moving window of pixels, of size 3x3, or 5x5...



loses proximity informatio

Convolutional Layer

Instead of making a linear combination of all the pixels, we consider a weighted average of a moving window of pixels, of size 3x3, or 5x5... **Drawing**



Convolutional Layer

0	0	0	0	0	0	...
0	156	155	156	158	158	...
0	153	154	157	159	159	...
0	149	151	155	158	159	...
0	146	146	149	153	158	...
0	145	143	143	148	158	...
...	...	...	...	...	...	...

Input Channel #1 (Red)

0	0	0	0	0	0	...
0	167	166	167	169	169	...
0	164	165	168	170	170	...
0	160	162	166	169	170	...
0	156	156	159	163	168	...
0	155	153	153	158	168	...
...	...	...	...	...	...	...

Input Channel #2 (Green)

0	0	0	0	0	0	...
0	163	162	163	165	165	...
0	160	161	164	166	166	...
0	156	158	162	165	166	...
0	155	155	158	162	167	...
0	154	152	152	157	167	...
...	...	...	...	...	...	...

Input Channel #3 (Blue)

-1	-1	1
0	1	-1
0	1	1

Kernel Channel #1

1	0	0
1	-1	-1
1	0	-1

Kernel Channel #2

0	1	1
0	1	0
1	-1	1

Kernel Channel #3

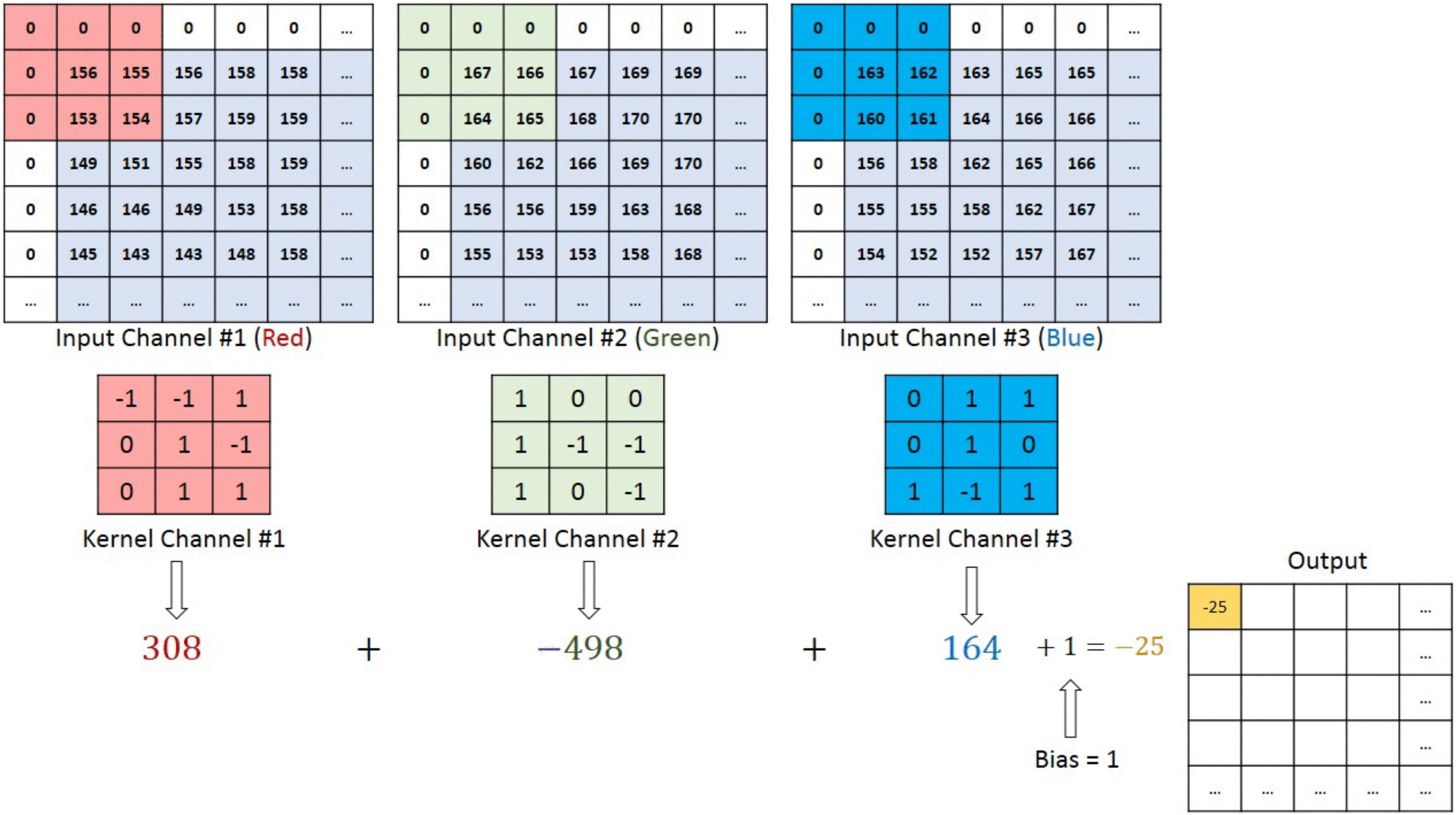
$$308 + -498 + 164 + 1 = -25$$

Bias = 1

Output

-25				...
				...
				...
				...
...	...	...	...	...

Convolutional Layer



Convolutional Layer

Input Channel #1 (Red)

0	0	0	0	0	0	...
0	156	155	156	158	158	...
0	153	154	157	159	159	...
0	149	151	155	158	159	...
0	146	146	149	153	158	...
0	145	143	143	148	158	...
...	...	...	...	...	...	...

Input Channel #2 (Green)

0	0	0	0	0	0	...
0	167	166	167	169	169	...
0	164	165	168	170	170	...
0	160	162	166	169	170	...
0	156	156	159	163	168	...
0	155	153	153	158	168	...
...	...	...	...	...	...	...

Input Channel #3 (Blue)

0	0	0	0	0	0	...
0	163	162	163	165	165	...
0	160	161	164	166	166	...
0	156	158	162	165	166	...
0	155	155	158	162	167	...
0	154	152	152	157	167	...
...	...	...	...	...	...	...

Kernel Channel #1

-1	-1	1
0	1	-1
0	1	1

Kernel Channel #2

1	0	0
1	-1	-1
1	0	-1

Kernel Channel #3

0	1	1
0	1	0
1	-1	1

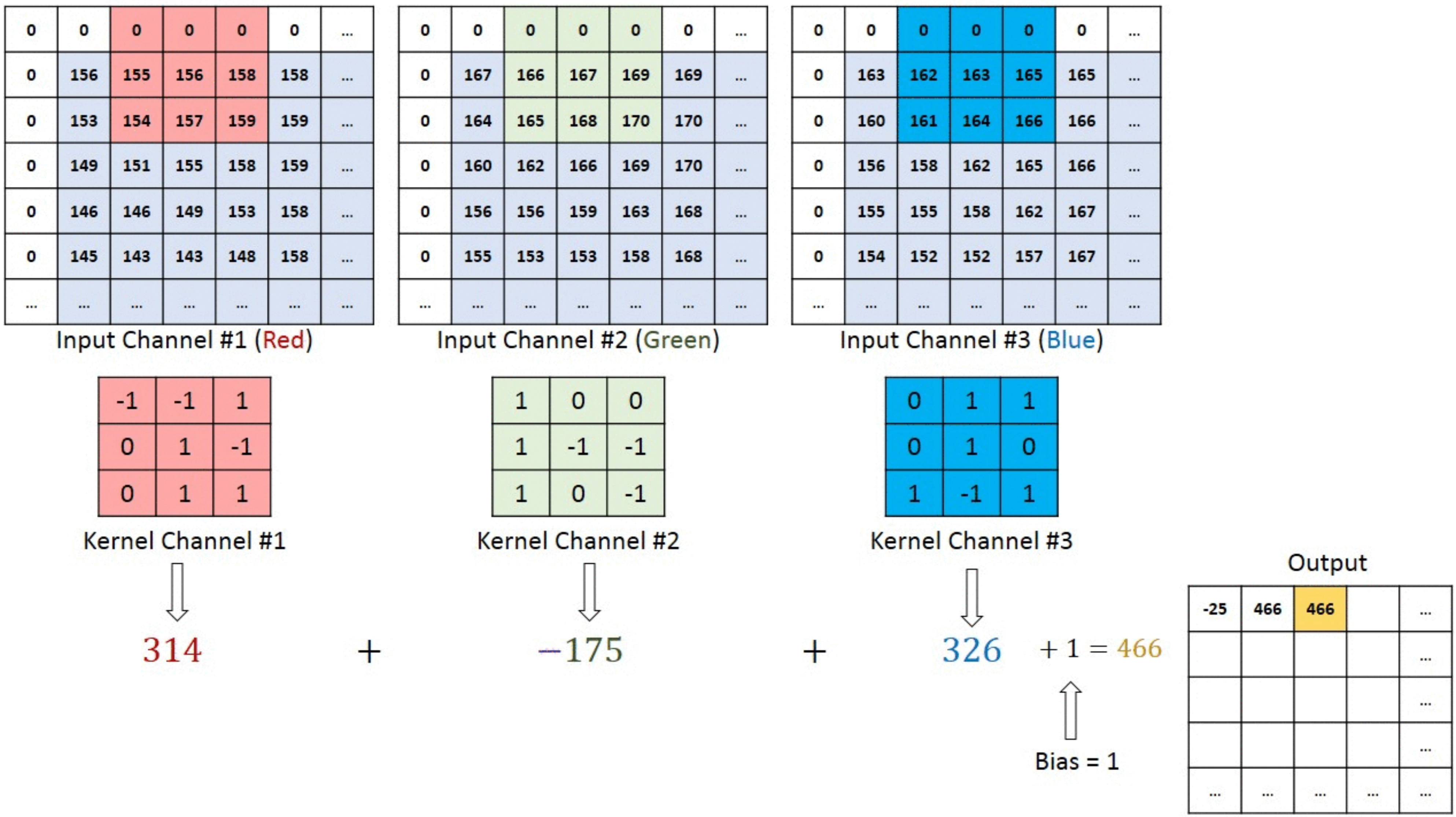
$$310 \quad + \quad -170 \quad + \quad 325 \quad +1 = 466$$

Bias = 1

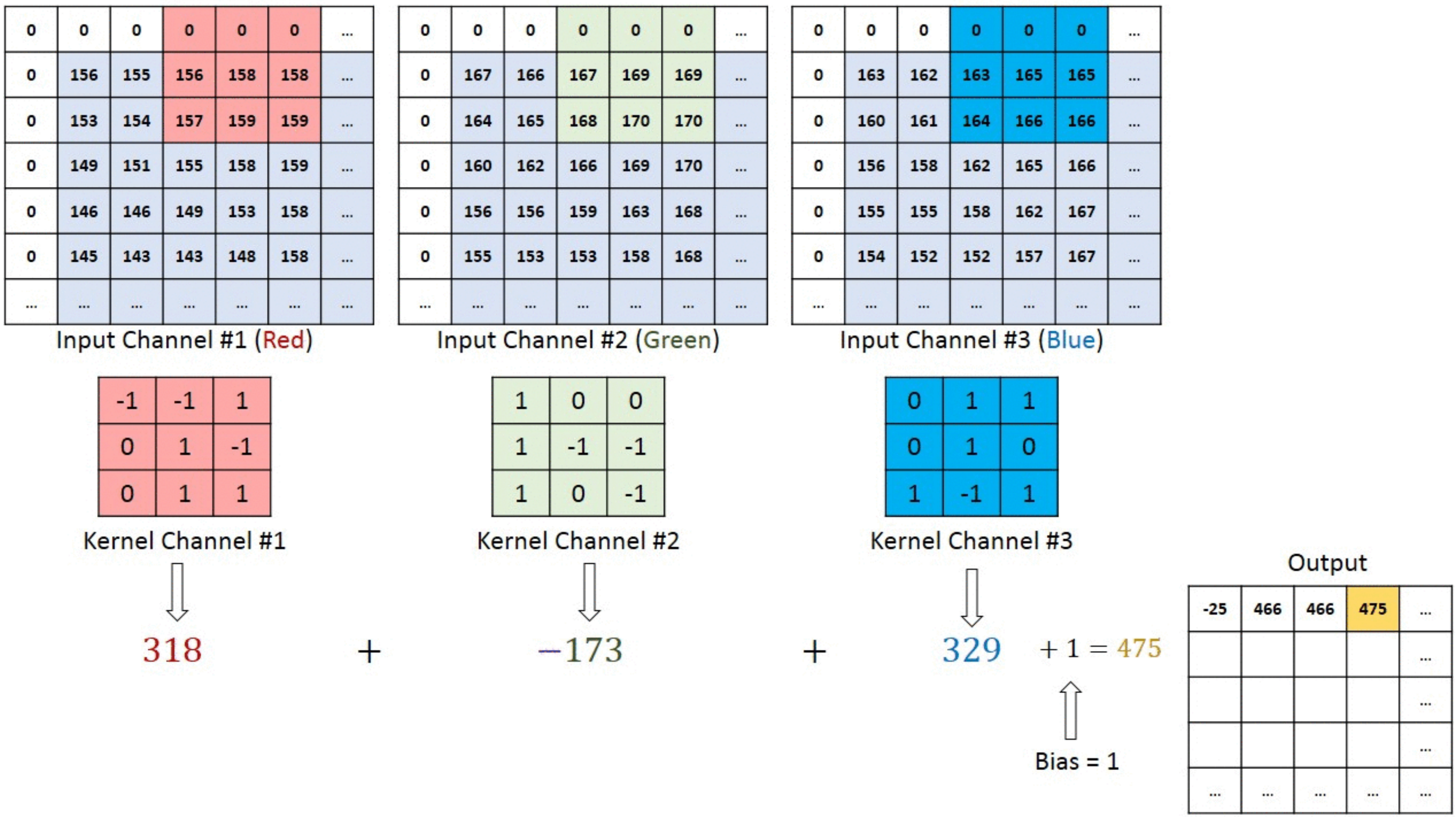
Output

-25	466			...
				...
				...
				...
...	...	...	...	...

Convolutional Layer



Convolutional Layer



Convolutional Layer

0	0	0	0	0	0	...
0	156	155	156	158	158	...
0	153	154	157	159	159	...
0	149	151	155	158	159	...
0	146	146	149	153	158	...
0	145	143	143	148	158	...
...	...	...	...	...	...	...

Input Channel #1 (Red)

0	0	0	0	0	0	...
0	167	166	167	169	169	...
0	164	165	168	170	170	...
0	160	162	166	169	170	...
0	156	156	159	163	168	...
0	155	153	153	158	168	...
...	...	...	...	...	...	...

Input Channel #2 (Green)

0	0	0	0	0	0	...
0	163	162	163	165	165	...
0	160	161	164	166	166	...
0	156	158	162	165	166	...
0	155	155	158	162	167	...
0	154	152	152	157	167	...
...	...	...	...	...	...	...

Input Channel #3 (Blue)

-1	-1	1
0	1	-1
0	1	1

Kernel Channel #1

1	0	0
1	-1	-1
1	0	-1

Kernel Channel #2

0	1	1
0	1	0
1	-1	1

Kernel Channel #3

$$298 + -491 + 487 + 1 = 295$$

Bias = 1

Output

-25	466	466	475	...
295				...
				...
				...
...	...	...	...	...

Convolutional Layer

0	0	0	0	0	0	...
0	156	155	156	158	158	...
0	153	154	157	159	159	...
0	149	151	155	158	159	...
0	146	146	149	153	158	...
0	145	143	143	148	158	...
...	...	...	...	...	...	...

Input Channel #1 (Red)

0	0	0	0	0	0	...
0	167	166	167	169	169	...
0	164	165	168	170	170	...
0	160	162	166	169	170	...
0	156	156	159	163	168	...
0	155	153	153	158	168	...
...	...	...	...	...	...	...

Input Channel #2 (Green)

0	0	0	0	0	0	...
0	163	162	163	165	165	...
0	160	161	164	166	166	...
0	156	158	162	165	166	...
0	155	155	158	162	167	...
0	154	152	152	157	167	...
...	...	...	...	...	...	...

Input Channel #3 (Blue)

-1	-1	1
0	1	-1
0	1	1

Kernel Channel #1

1	0	0
1	-1	-1
1	0	-1

Kernel Channel #2

0	1	1
0	1	0
1	-1	1

Kernel Channel #3

$$148 \quad + \quad -8 \quad + \quad 646 \quad + 1 = 787$$

Bias = 1

Output

-25	466	466	475	...
295	787			...
				...
				...
...	...	...	...	...

Summary

CNN:

- allow to take spacial information into account
- reduce the number of parameters per layer. We can have deeper networks!

Summary

CNN:

- allow to take spacial information into account
- reduce the number of parameters per layer. We can have deeper networks!

LeNet 1998:

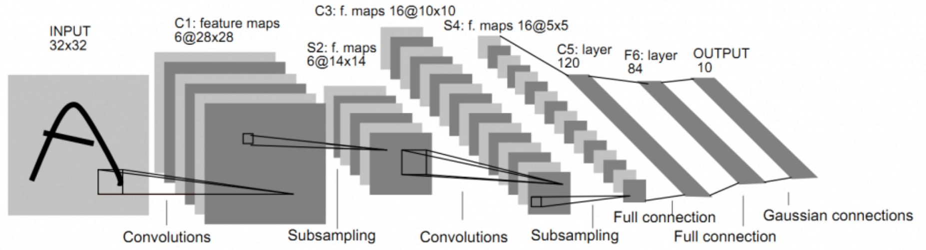
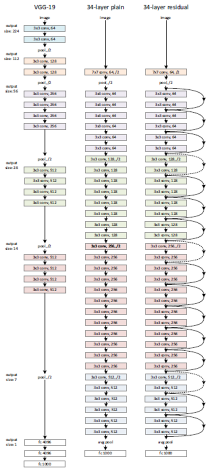


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

8 layers, 60k parameters.

Resnet 2016



$\sim$40 Layers, 140M parameters

Outline

Outline

Image classification - CNN Tree

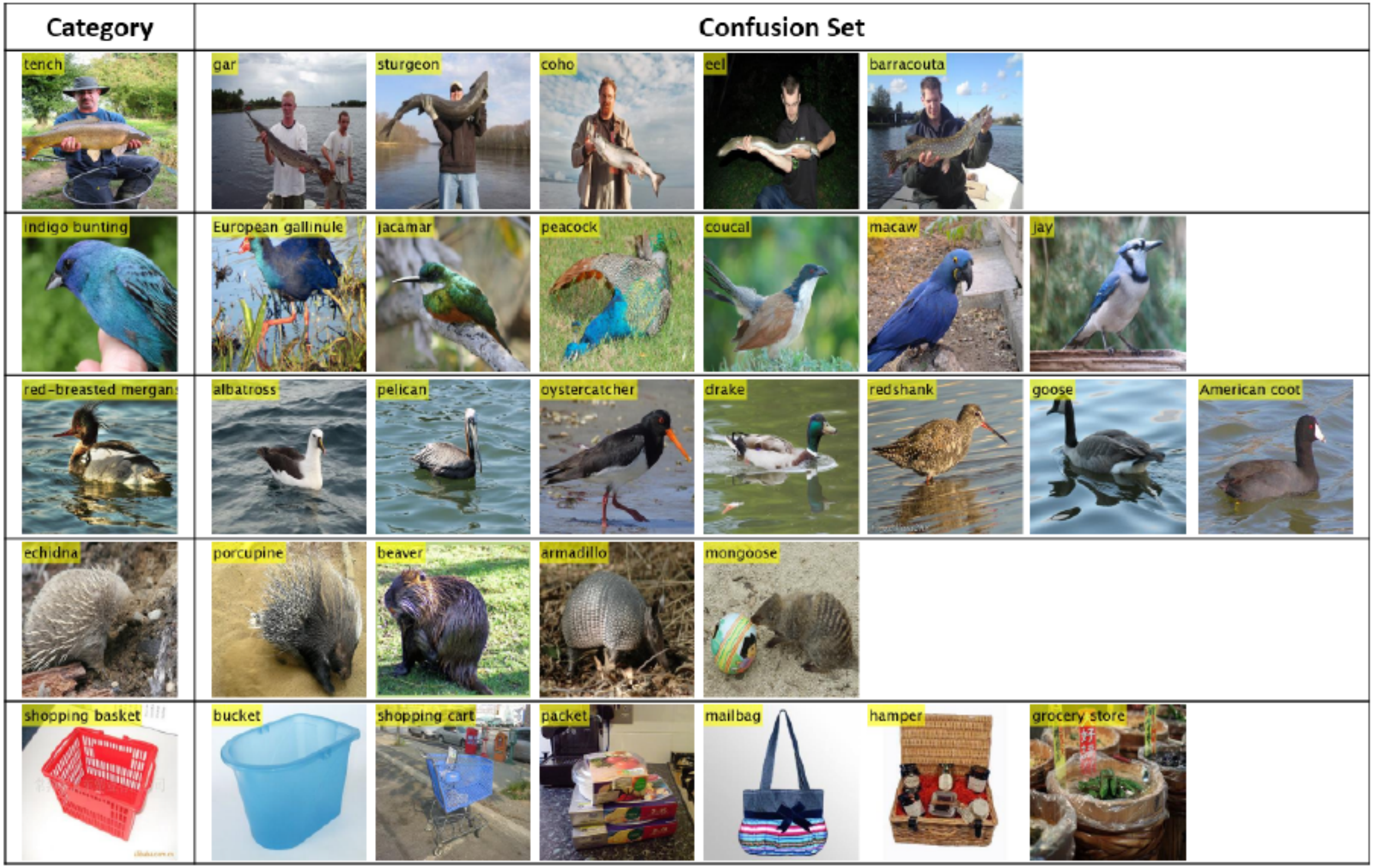


Figure: Confusion set outputs by AlexNet softmax prediction on validation set of ILSVRC 2015.

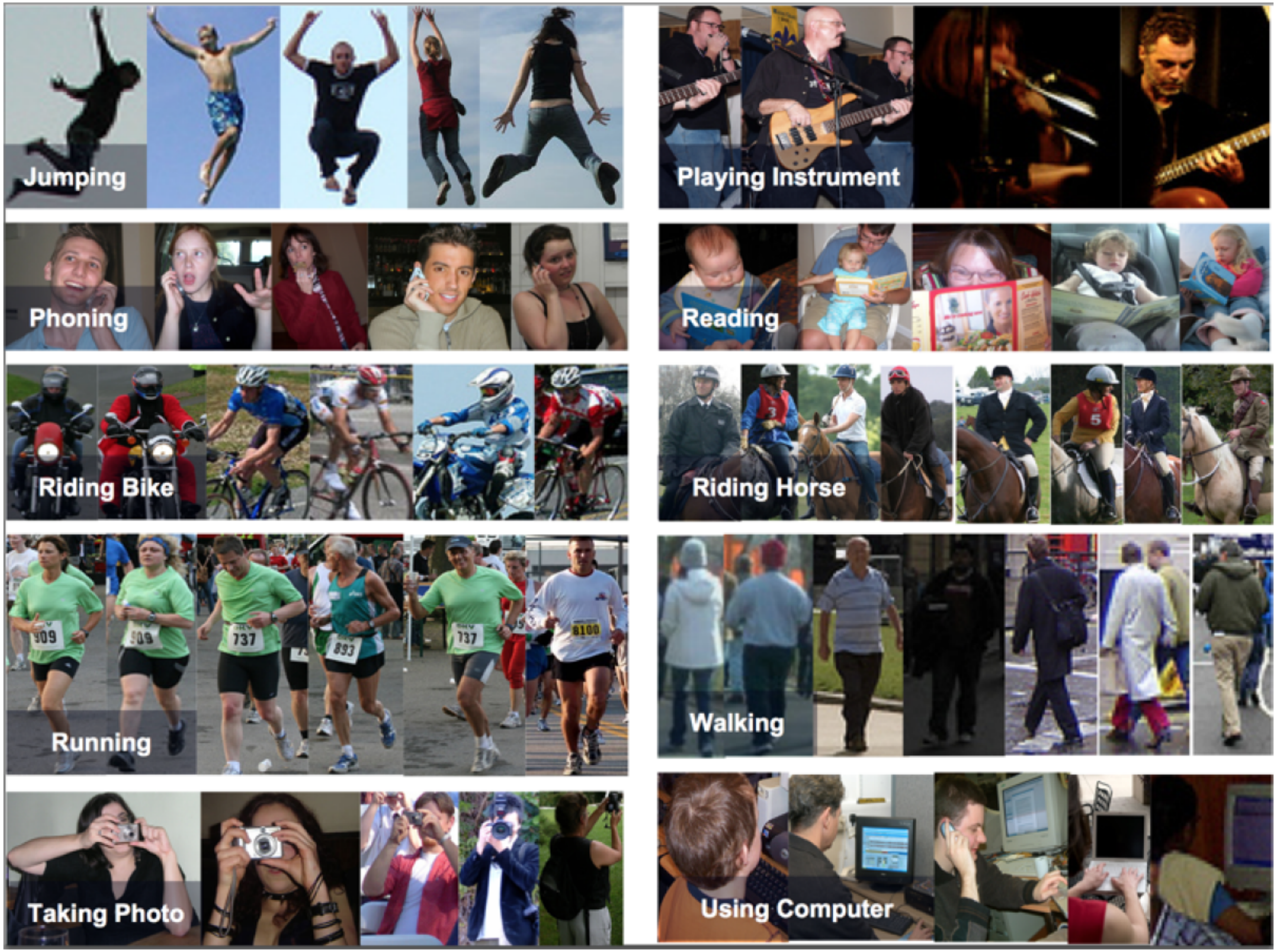
Image classification - CNN Tree

Figure: Top label is given by basic AlexNet CNN while bottom one is given by CNNTree (green color corresponds to a correct prediction)

Outline

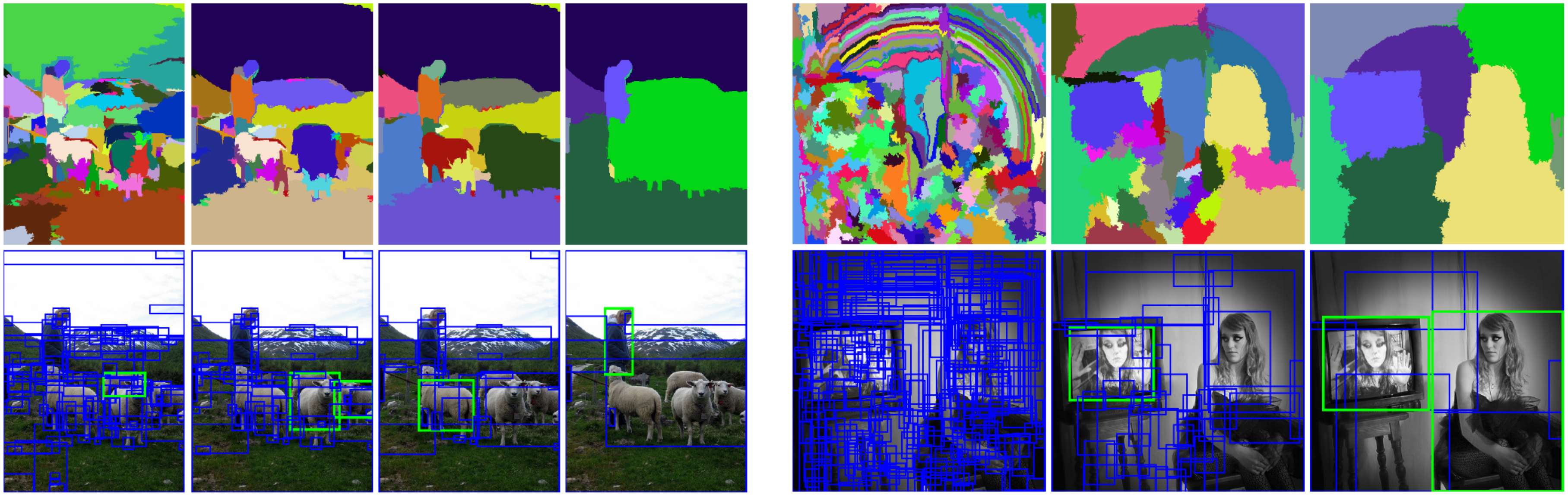
Pose estimation - Deeppose

Action recognition



Outline

Object detection and segmentation



(a) (b)

Object detection - YOLO, SDD

More recently, YOLO (You Only Look Once) and SSD (Single Shot Detector) allow single pipeline detection that directly predicts class labels.

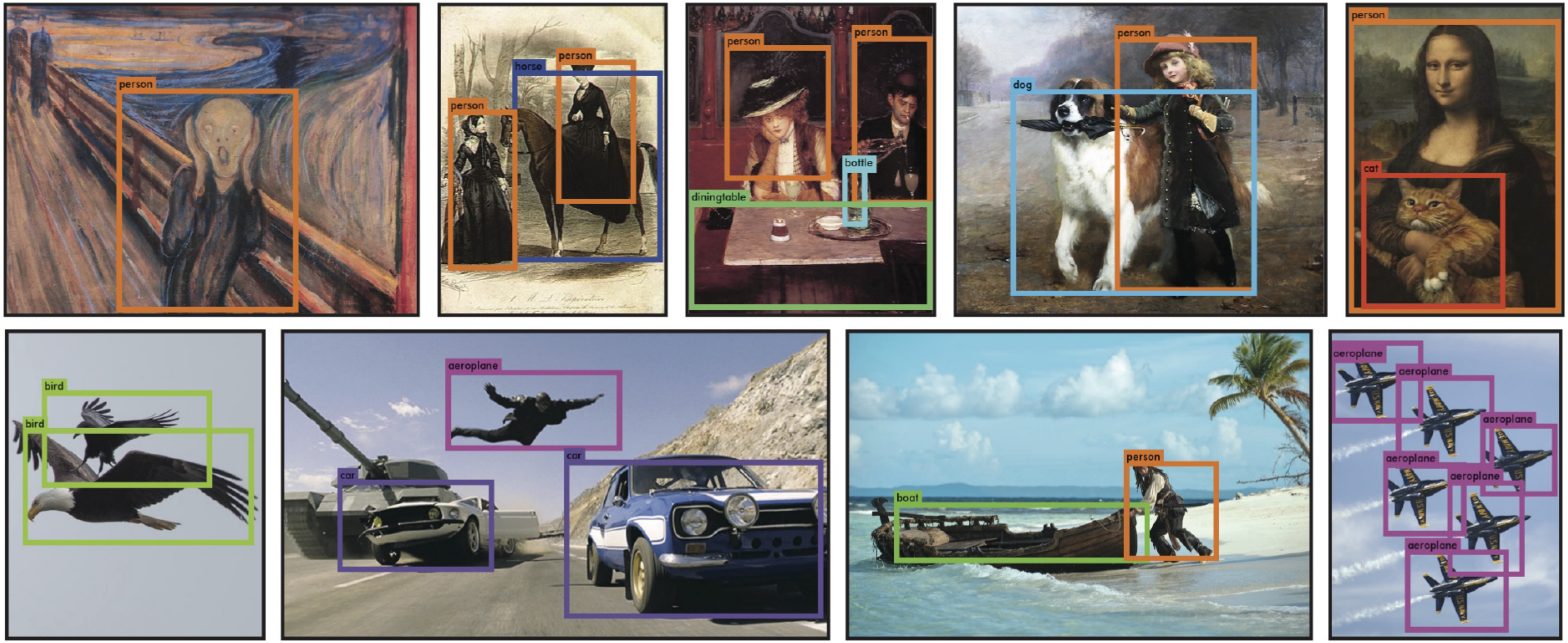
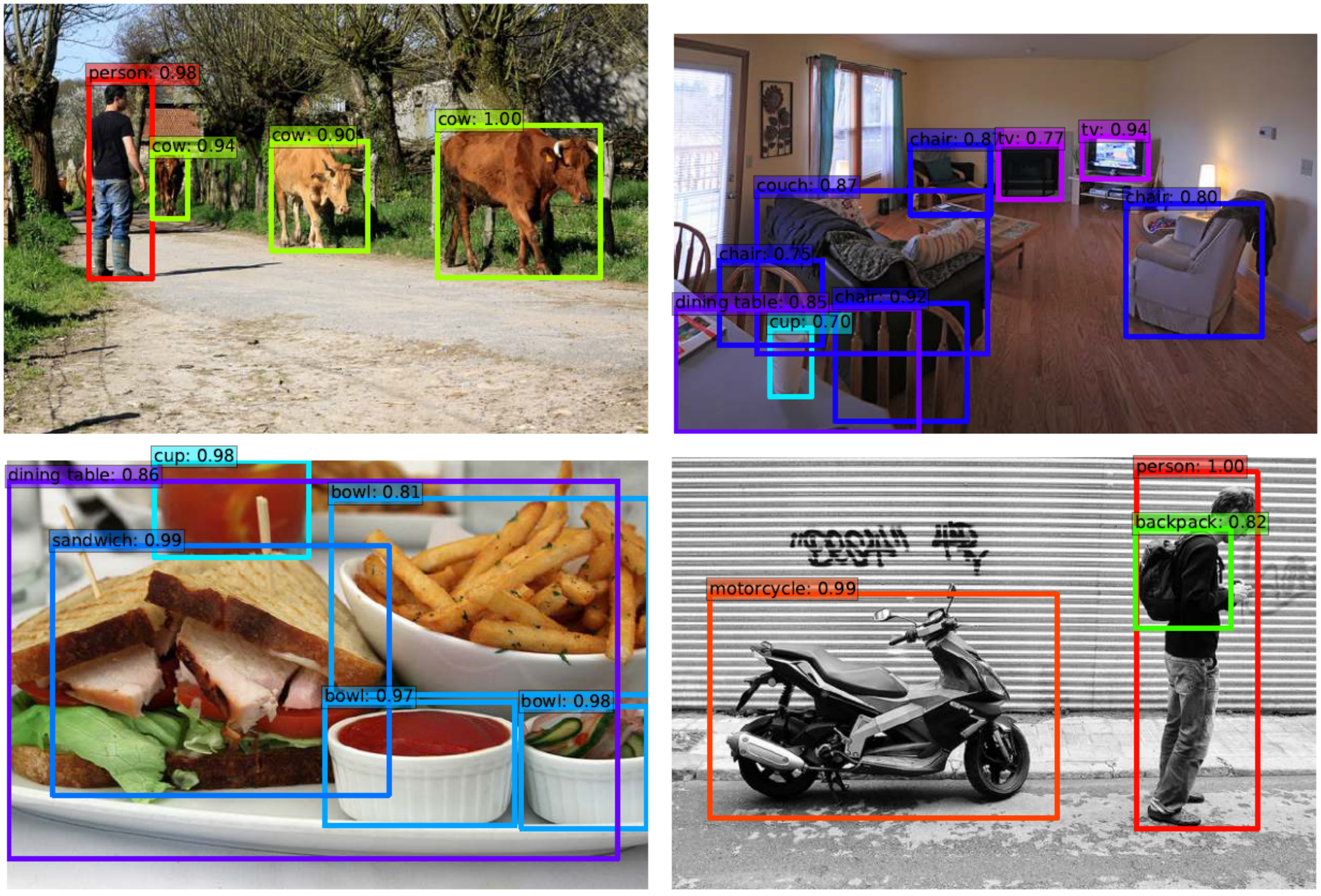


Figure: YOLO results

SSD

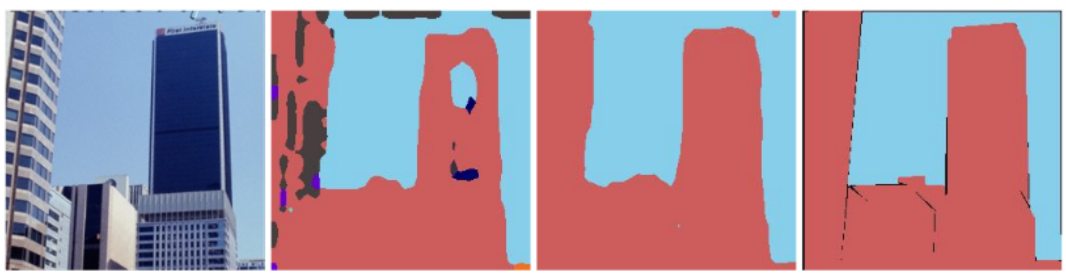


person: 0.98
cow: 0.94
cow: 0.90
cow: 1.00
chair: 0.81
tv: 0.77
tv: 0.94
couch: 0.87
chair: 0.80
chair: 0.75
dining table: 0.85
chair: 0.92
cup: 0.70
cup: 0.98
dining table: 0.86
bowl: 0.81
sandwich: 0.99
bowl: 0.97
bowl: 0.98
person: 1.00
backpack: 0.82
motorcycle: 0.99

Outline

Scene labeling - DAG-RNN

Input Image CNN DAG-RNN Ground Truth



Input Image CNN DAG-RNN Ground Truth

Outline

Object tracking

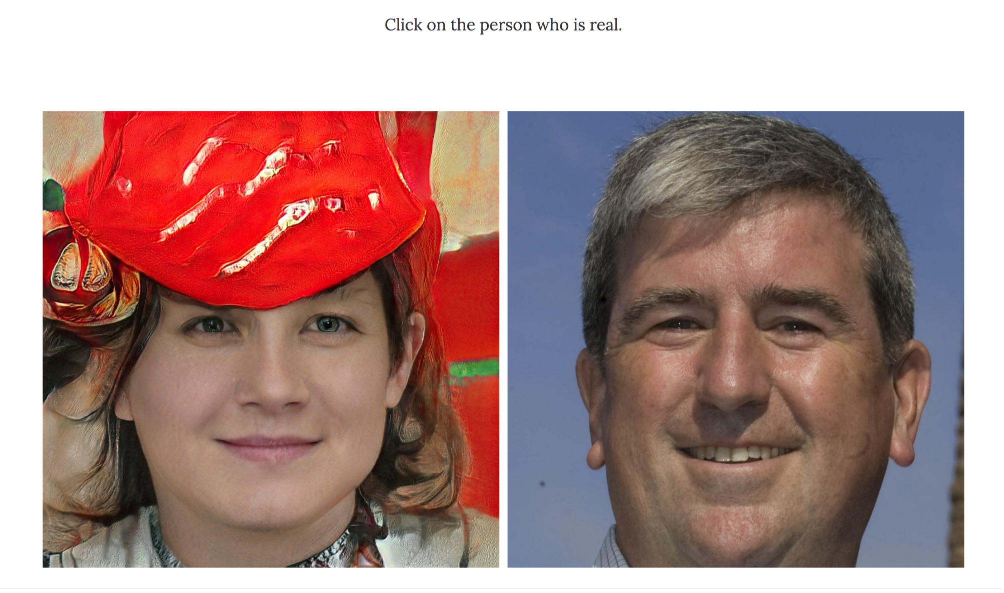
Object Detection in a video:

- YOLO `https://pjreddie.com/darknet/yolo/`
- YOLO - James Bond `https://www.youtube.com/watch?v=VOC3huqHrss`

Deep Learning - Many many other applications!

1. Image and video (super-resolution, 3D, Image captioning), medical applications...
2. Self Driving cars (scene segmentation, etc.)
3. In NLP (language, text), automatic translation, voice recognition, text generation, next word completion ... (Virtual assistants)
4. Recommendation systems...
5. For time series, complex datasets too...

But also beyond the supervised settings: using adversarial networks (GANS) for generation of images, faces, voice, etc.



http://www.whichfaceisreal.com/

Outline

Limits and problems

Limits of Deep Learning: $\rightarrow$

Limits and problems

Limits of Deep Learning: →

- Interpretability
- Vulnerability, failures, biases

Limits and problems

Limits of Deep Learning: →

- Interpretability
- Vulnerability, failures, biases
- Size of the models: Applicability, Ethical, Energy & Ecological concerns (Ai has both a positive and a huge negative carbon footprint)!

Limits and problems

Limits of Deep Learning: →

- Interpretability
- Vulnerability, failures, biases
- Size of the models: Applicability, Ethical, Energy & Ecological concerns (Ai has both a positive and a huge negative carbon footprint)!
- Theory!

Limits and problems

Limits of Deep Learning: →

- Interpretability
- Vulnerability, failures, biases
- Size of the models: Applicability, Ethical, Energy & Ecological concerns (Ai has both a positive and a huge negative carbon footprint)!
- Theory!
- Ethical and practical concerns

Limits and problems

Limits of Deep Learning: →

- Interpretability
- Vulnerability, failures, biases
- Size of the models: Applicability, Ethical, Energy & Ecological concerns (Ai has both a positive and a huge negative carbon footprint)!
- Theory!
- Ethical and practical concerns
- Privacy

Limits and problems

Limits of Deep Learning: →

- Interpretability
- Vulnerability, failures, biases
- Size of the models: Applicability, Ethical, Energy & Ecological concerns (Ai has both a positive and a huge negative carbon footprint)!
- Theory!
- Ethical and practical concerns
- Privacy
- Source tracking, Ownership, IP

Limits and problems

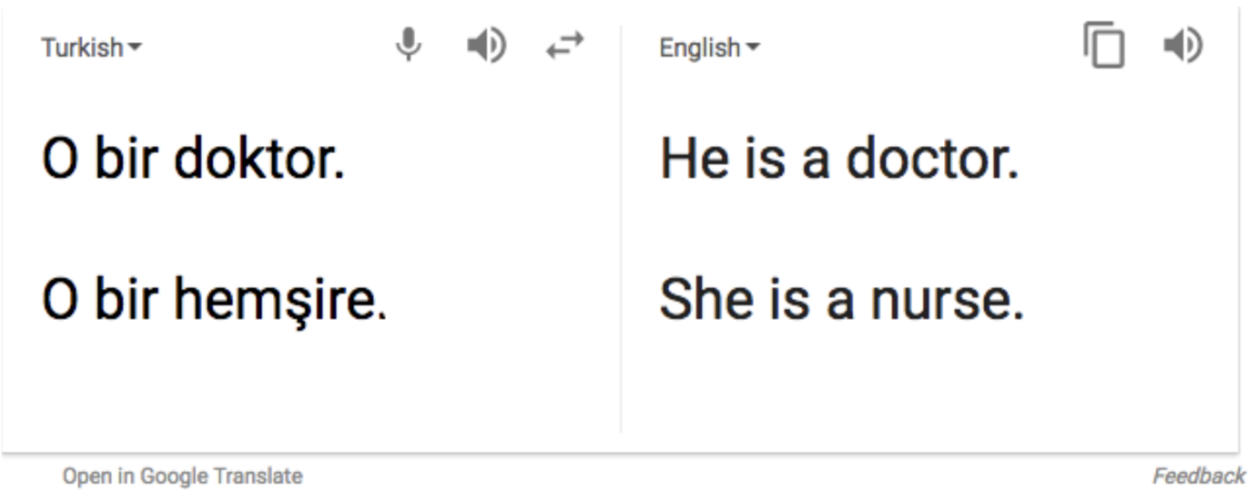
Limits of Deep Learning: →

- Interpretability
- Vulnerability, failures, biases
- Size of the models: Applicability, Ethical, Energy & Ecological concerns (Ai has both a positive and a huge negative carbon footprint)!
- Theory!
- Ethical and practical concerns
- Privacy
- Source tracking, Ownership, IP
- Alignment

Ethical issues

1. **Bias**
2. **Interpretability**
3. **Ecological Impact of AI and Energy Consumption**
4. **Ethics and Responsibility in AI**
5. **Data Privacy Protection**
6. **Alignment**
7. **Copyright issues**
8. **Societal challenges**

Biases


Turkish
English
O bir doktor.
He is a doctor.
O bir hemşire.
She is a nurse.
Open in Google Translate
Feedback

Car accidents

Female drivers and right front passengers are approximately

17 percent more likely

to be killed

in a car crash than a male occupant of the same age.

Any seatbelt-wearing female vehicle occupant has

73 percent greater odds of being

seriously injured

in a frontal car crash than the odds of a seatbelt-wearing male occupant being injured in the same kind and severity of crash.

Sources: NHTSA and the journal Traffic Injury Prevention

Analysis of crash and injury data compiled from the National Automotive Sampling System Crashworthiness Data System for the years 1998 to 2015.

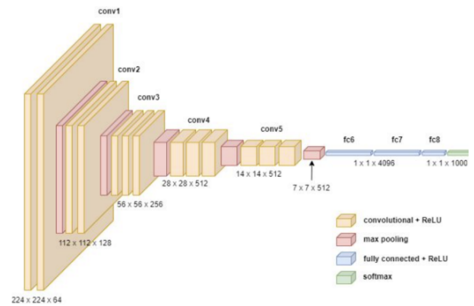
Bias in crash test

https://www.consumerreports.org/car-safety/crash-test-bias-how-male-focused-testing-puts-female-drivers-at-risk/

Interpretability

Very hard questions:

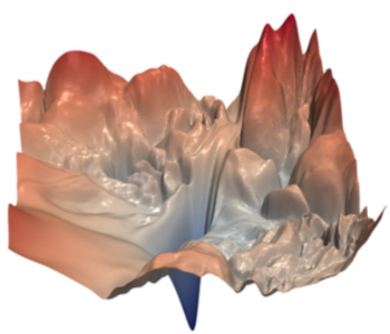
1. How can I interpret the impact of each feature in my prediction?



VGGNet, 16 layers, up to 95M parameters, trained on 1 billion images

2. What is the impact of each training point in my model?

   $\rightarrow$ (data valuation, sourcing)



⚠ Hard to avoid! Model size makes interpretability and learning dynamics always more complex.

Home News Tech Finance Leadership Well Recommends Fortune 500

TECH· A.I.

OpenAI reportedly wants to build 5-gigawatt data centers, and nobody knows who could supply that much power

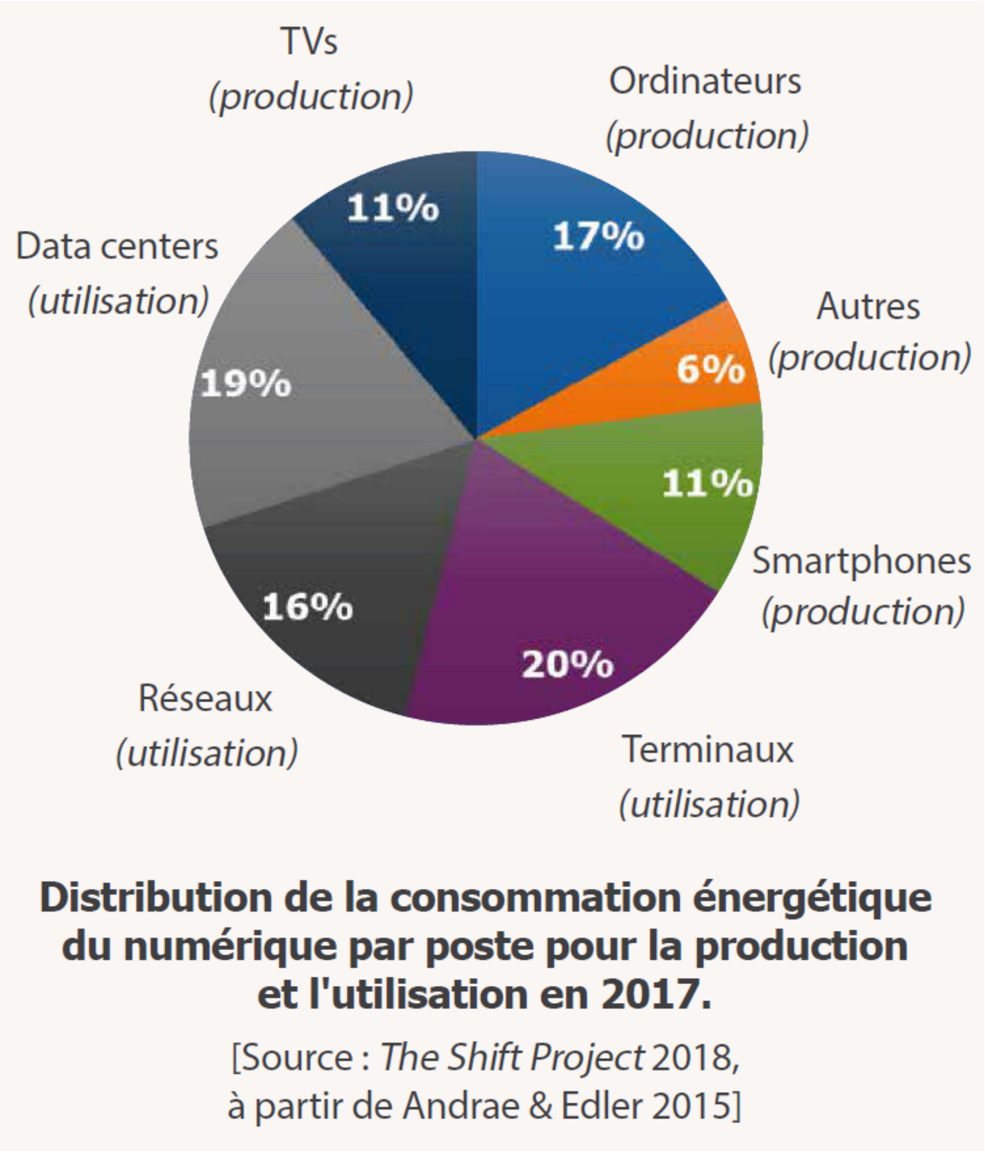
BY DAVID MEYER
September 27, 2024 at 1:59 PM GMT+2

Sam Altman, chief executive officer of OpenAI.
DAVID PAUL MORRIS—BLOOMBERG/GETTY IMAGES

Figure: Yahoo finance, 27 September 2024 source

⚠ Hard to avoid! Model size and training keeps increasing.

Environmental issues

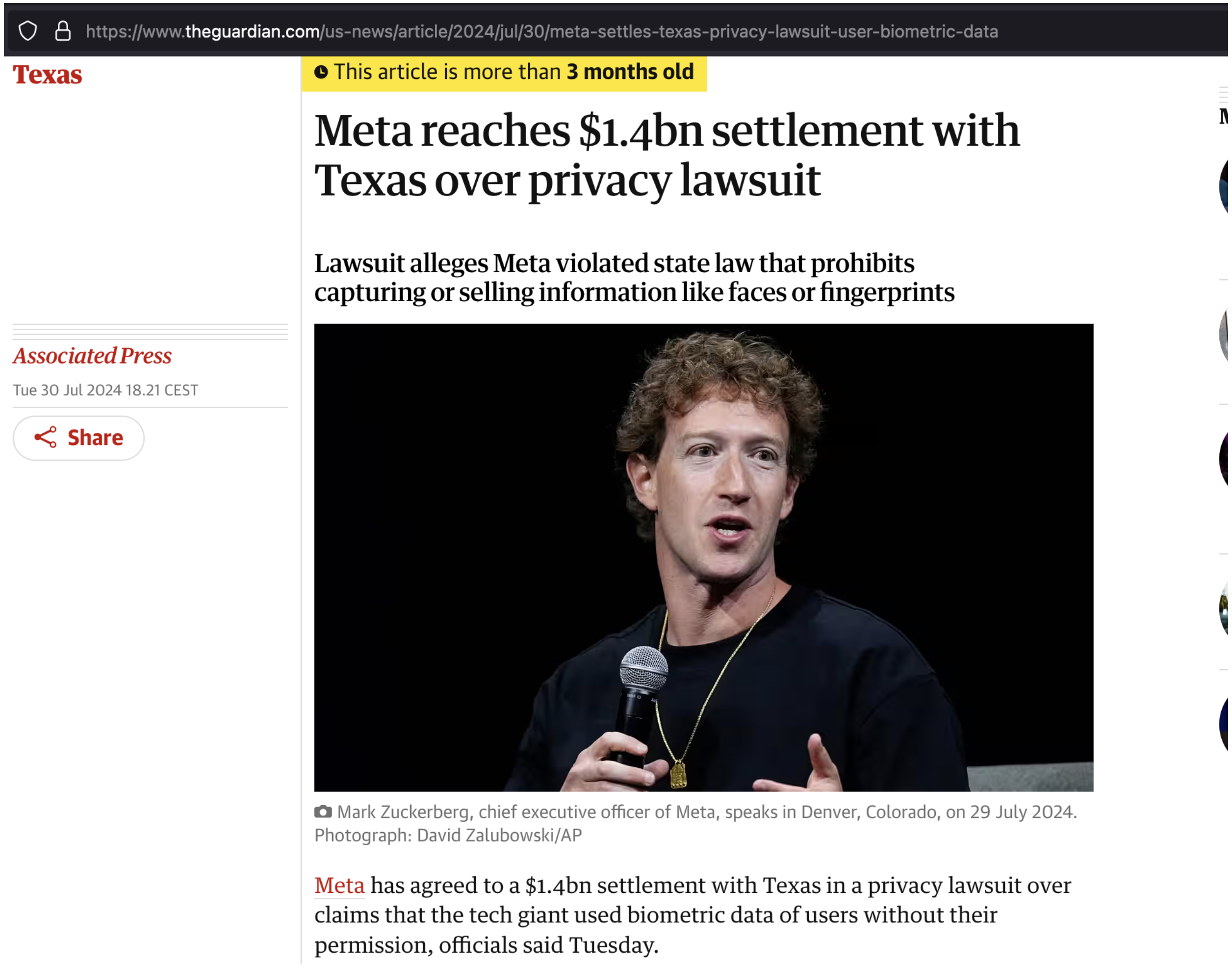


The Shift Project `https://theshiftproject.org/lean-ict/`

Ethics and responsibility

1. Military applications
2. Face recognition

Data privacy



https://www.theguardian.com/us-news/article/2024/jul/30/meta-settles-texas-privacy-lawsuit-user-biometric-data

Texas

This article is more than **3 months old**

Meta reaches $1.4bn settlement with Texas over privacy lawsuit

Lawsuit alleges Meta violated state law that prohibits capturing or selling information like faces or fingerprints

Associated Press

Tue 30 Jul 2024 18.21 CEST

Share

Mark Zuckerberg, chief executive officer of Meta, speaks in Denver, Colorado, on 29 July 2024. Photograph: David Zalubowski/AP

Meta has agreed to a $1.4bn settlement with Texas in a privacy lawsuit over claims that the tech giant used biometric data of users without their permission, officials said Tuesday.

Figure: The guardian, 2024

⚠ Ensuring privacy during training is extremely hard.

Alignment


FAST COMPANY

CO.DESIGN TECH WORK LIFE NEWS IMPACT PODCASTS VIDEO RECOMMENDER INNOVATION FESTI

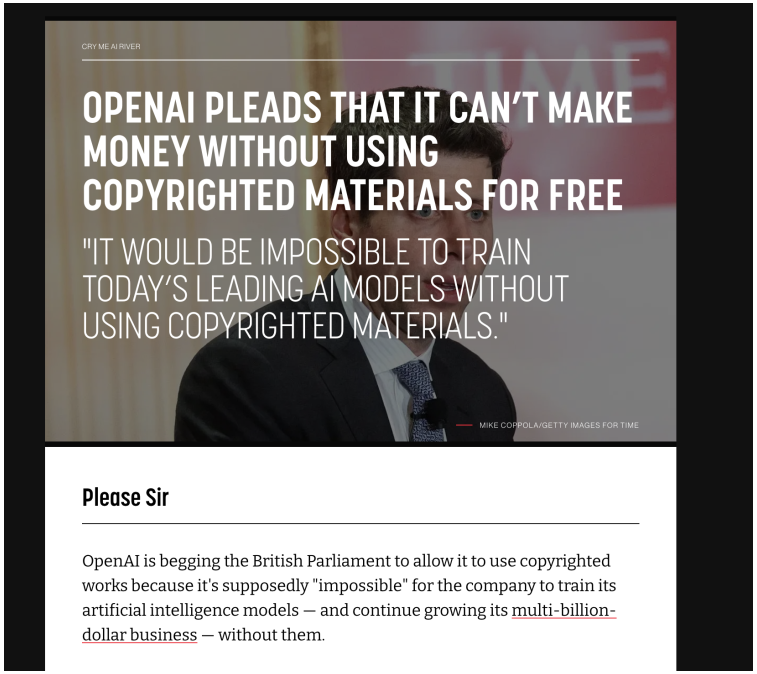
11-06-17 | PLATFORM WARS

Netflix CEO Reed Hastings: Sleep Is Our Competition

For Netflix, the battle for domination goes far beyond which TV remote to pick up.

CEO Reed Hastings [Photo: Mondileinchen/Wikimedia Commons]

Copyrighted Data



CRY ME AI RIVER

OPENAI PLEADS THAT IT CAN'T MAKE MONEY WITHOUT USING COPYRIGHTED MATERIALS FOR FREE

"IT WOULD BE IMPOSSIBLE TO TRAIN TODAY'S LEADING AI MODELS WITHOUT USING COPYRIGHTED MATERIALS."

MIKE COPPOLA/GETTY IMAGES FOR TIME

Please Sir

OpenAI is begging the British Parliament to allow it to use copyrighted works because it's supposedly "impossible" for the company to train its artificial intelligence models — and continue growing its multi-billion-dollar business — without them.

The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work

Millions of articles from The New York Times were used to train chatbots that now compete with it, the lawsuit said.

Share full article

1.3K

⚠ **Foundation Models and Intellectual Property are not easily compatible**

Societal challenges

Society and jobs

- **Transformation of Current Jobs and Emergence of New Roles**: New positions focused on the development, management, and supervision of AI systems.
- **Societal Risks and Benefits of AI**: Risks such as job loss, algorithmic discrimination, and privacy protection.

Society and democracy, crime, etc.

- **Democratic and Political Challenges** Mass manipulation, echo chambers, hacking.
- Criminal usage (e.g., deepfakes)

Society and Sovereignty

- **Concentration of Major AI Players** $\rightarrow$ competition, innovation, and benefit distribution.
- The Sovereign AI Challenge: The concept of sovereign AI implies a country's capacity to develop, deploy, and control its own AI technologies, raising concerns about technological dependence.

Cost of storing data

- Money: 300.000 US dollars in Google Cloud to store 1 petabyte during one year.

Cost of storing data

- Money: 300.000 US dollars in Google Cloud to store 1 petabyte during one year.
- Data centers and environment
  - 2% of the total electricity consumption in the US.
  - 626 billion liters of water.
  - 2% of total global greenhouse emissions.

⚠ Hard to avoid - model performance is directly related to data.

THE VERGE TECH REVIEWS SCIENCE CREATORS ENTERTAINMENT VIDEO MORE

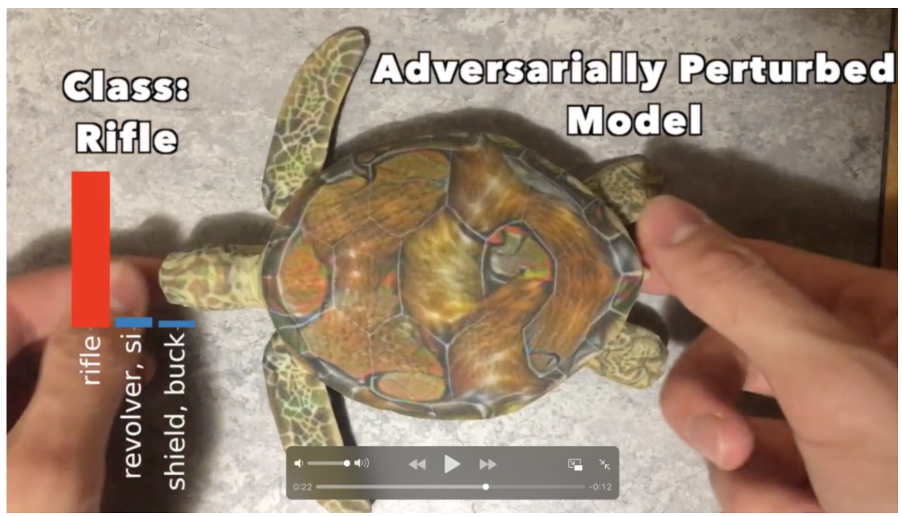
MICROSOFT WEB TL;DR

Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day

By James Vincent | Mar 24, 2016, 6:43am EDT

Via The Guardian | Source TayandYou (Twitter)

SHARE



CNN are not always “robust” to adversarial attacks !

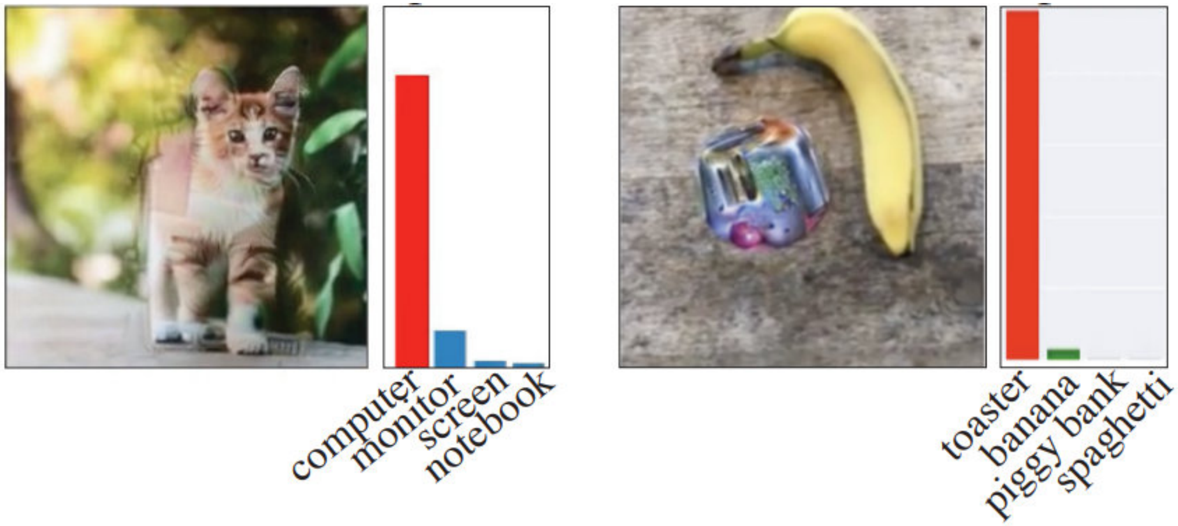
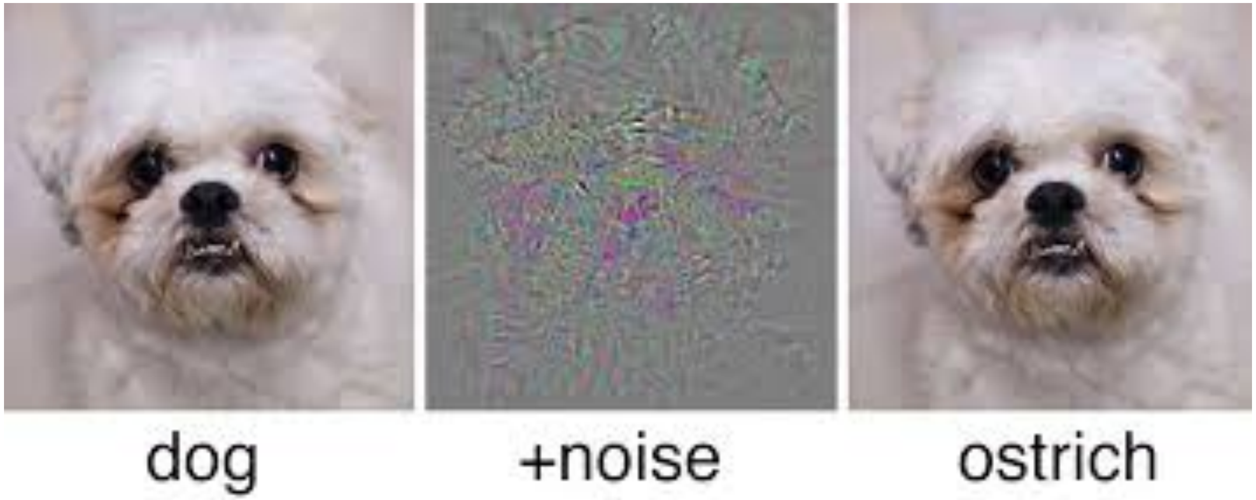


Figure: What happened here ??

Deep Learning: failures

CNN are not always “robust” to adversarial attacks !



Adding a small well chosen noise can completely fool a CNN !
↬ Can we trust deep networks on cars, medical applications, planes...?

Deep Learning: questionable applications

Are some applications just bad (they all already exist :( ) ?

- Deep Fakes: insert someone in a video.
- Underground exploration: finding new fossil resources.
- Voice imitation: what if I cannot check who is calling me.
- Military applications + automatic weapons.
- Mass surveillance.

Deep Learning: questionable applications

Are some applications just bad (they all already exist :( ) ?

- Deep Fakes: insert someone in a video.
- Underground exploration: finding new fossil resources.
- Voice imitation: what if I cannot check who is calling me.
- Military applications + automatic weapons.
- Mass surveillance.

Deep Learning: questionable applications

💡 A challenging problem: chain of responsibility. The same tools are used for positive and negative applications, e.g.:

- mass surveillance
- tumor recognition

are both based on image recognition...
Good news ! the research community is very aware of the situation.
Some references:

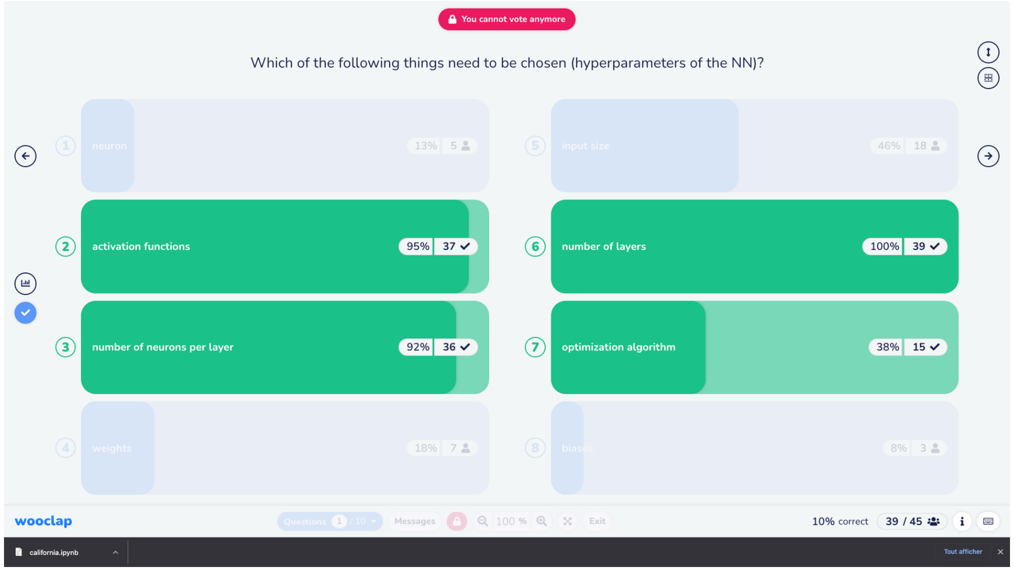
- Asilomar AI Principles `https://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-oheigeartaigh.pdf` `https://futureoflife.org/ai-principles/`
- European guidelines `https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai`
- AI for Good `https://ai4good.org/`

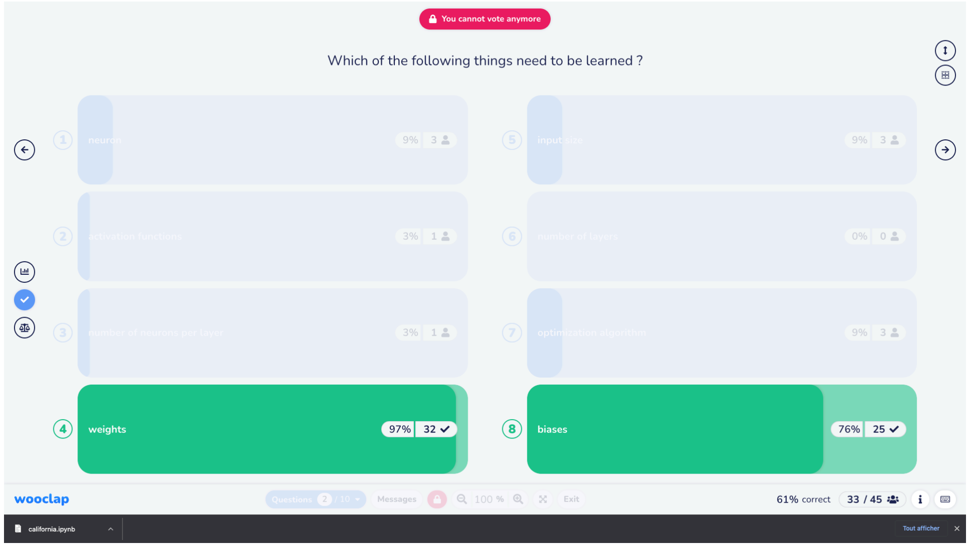
Conclusion: Deep Learning in one slide

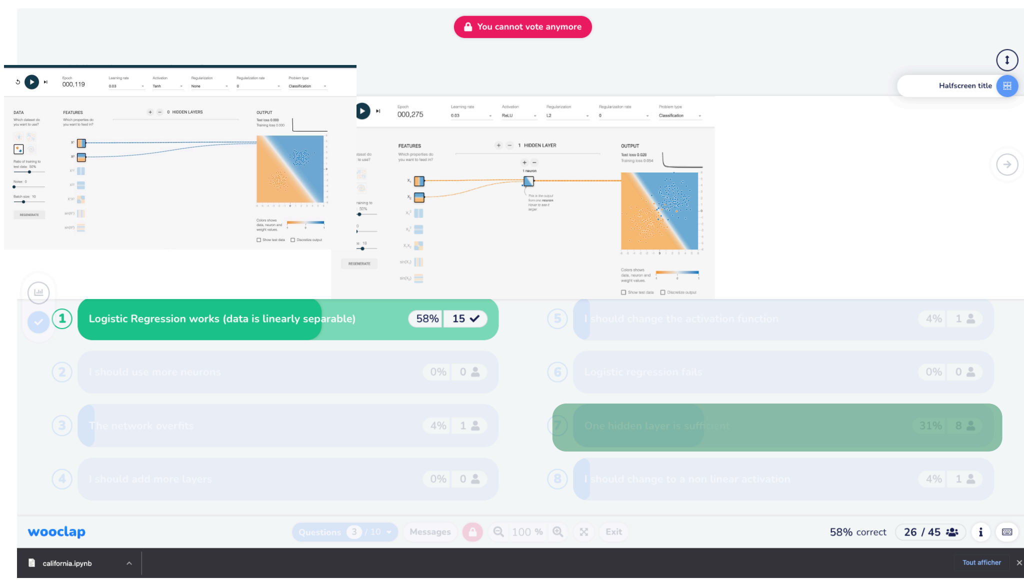
- **How does it work:**
  - Automatically learn representations of observations
  - Learn highly non-linear models.
- **What does it require:**
  - Large datasets with structure
  - Computational power
- **Why now:**
  - Combination of the 2 points above
  - investment!
- **Some Applications**
  - Image classification; object / face recognition
  - Self driving cars
  - Automatic Translation, Information extraction
  - Caption Generation
  - Ads, recommendation systems, etc.

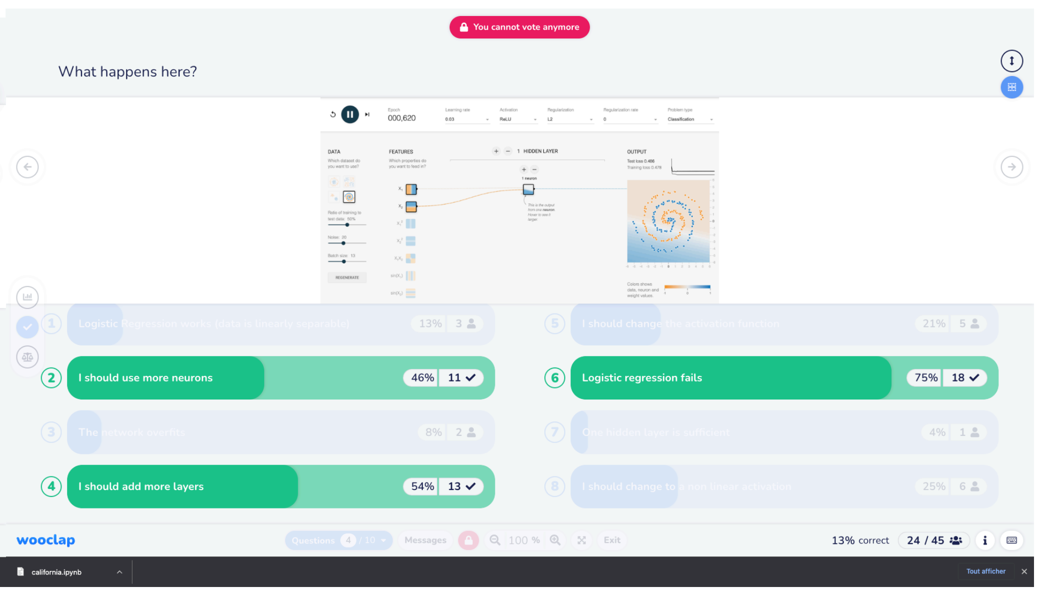
Goals: Understanding core concepts of Deep Learning.

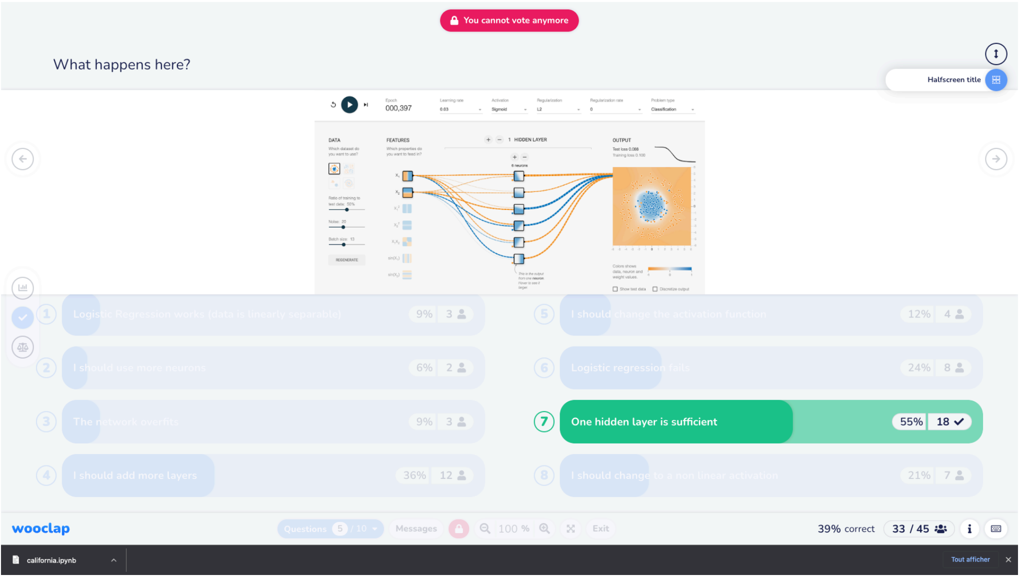
- When and how it works on paper (data, models, architecture)
- How to implement a simple neural network with Python.
- Overview of some of the main applications and challenges.

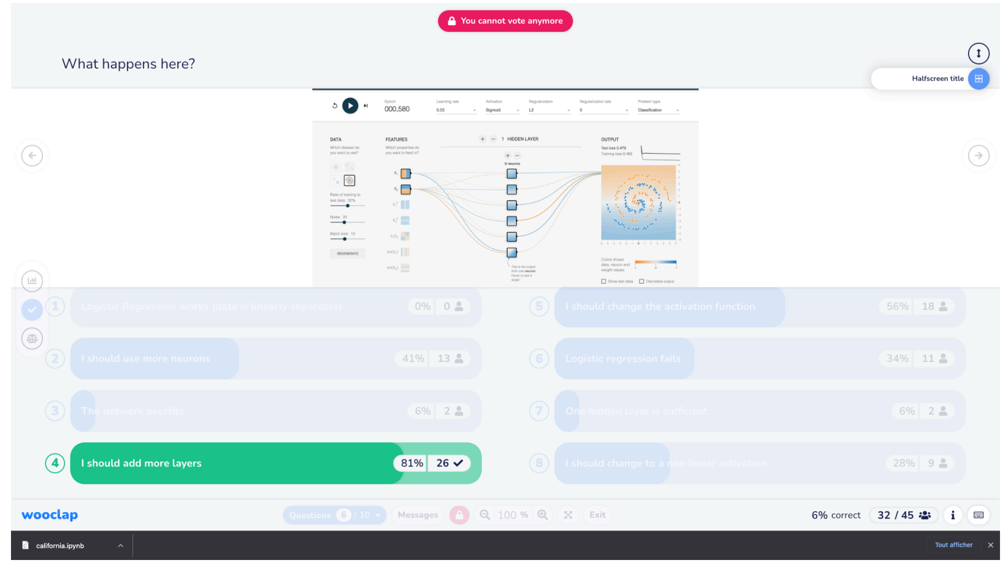

You cannot vote anymore
Which of the following things need to be chosen (hyperparameters of the NN)?
neuron 13% 5
activation functions 95% 37
number of neurons per layer 92% 36
weights 18% 7
input size 46% 18
number of layers 100% 39
optimization algorithm 38% 15
biases 8% 3
wooclap
10% correct 39 / 45

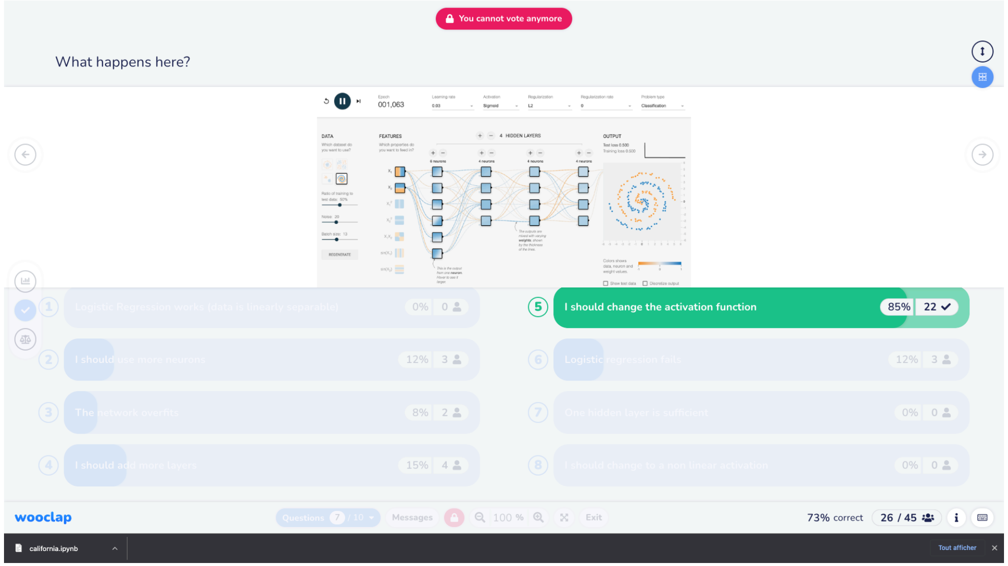

You cannot vote anymore
Which of the following things need to be learned ?
neuron 9% 3
activation functions 3% 1
number of neurons per layer 3% 1
weights 97% 32
input size 9% 3
number of layers 0% 0
optimization algorithm 9% 3
biases 76% 25
wooclap
61% correct 33 / 45

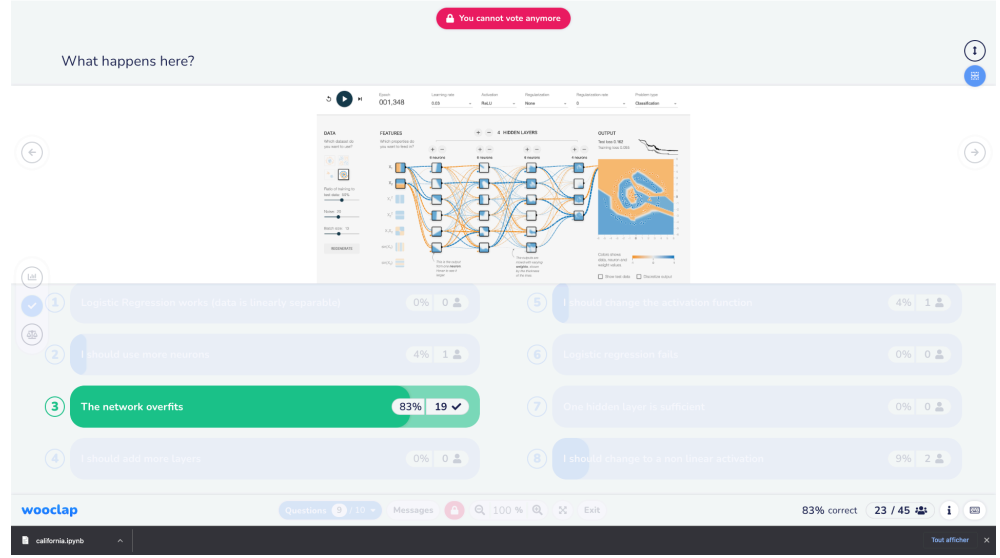

You cannot vote anymore
Halfscreen title
Logistic Regression works (data is linearly separable) 58% 15
I should use more neurons 0% 0
The network overfits 4% 1
I should add more layers 0% 0
I should change the activation function 4% 1
Logistic regression fails 0% 0
One hidden layer is sufficient 31% 8
I should change to a non linear activation 4% 1
wooclap
58% correct 26 / 45

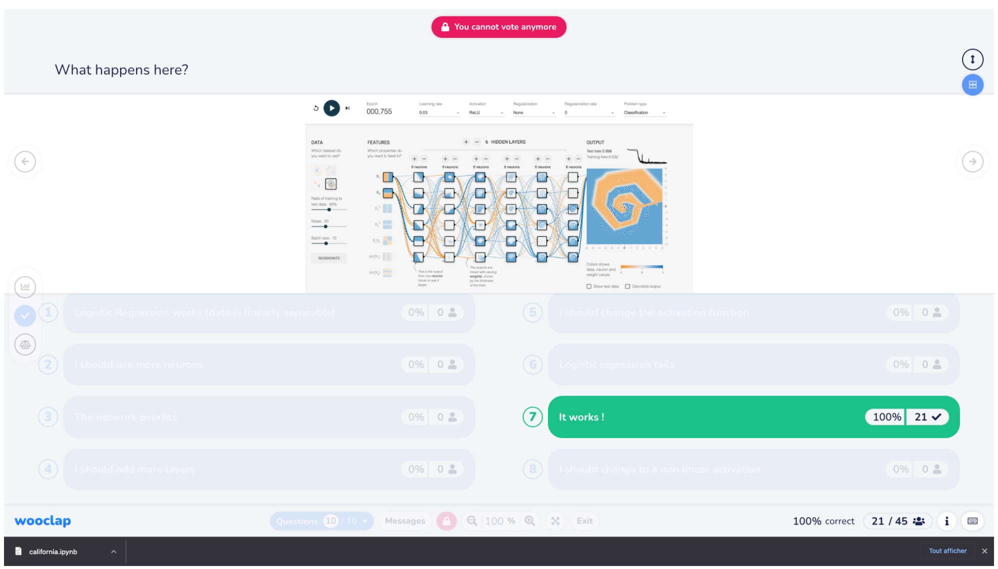

You cannot vote anymore
What happens here?
Logistic Regression works (data is linearly separable) 13% 3
I should use more neurons 46% 11
The network overfits 8% 2
I should add more layers 54% 13
I should change the activation function 21% 5
Logistic regression fails 75% 18
One hidden layer is sufficient 4% 1
I should change to a non linear activation 25% 6
wooclap
13% correct 24 / 45


You cannot vote anymore
What happens here?
Halfscreen title
Logistic Regression works (data is linearly separable) 9% 3
I should use more neurons 6% 2
The network overfits 9% 3
I should add more layers 36% 12
I should change the activation function 12% 4
Logistic regression fails 24% 8
One hidden layer is sufficient 55% 18
I should change to a non linear activation 21% 7
wooclap
39% correct 33 / 45


You cannot vote anymore
What happens here?
Halfscreen title
Logistic Regression works (data is linearly separable) 0% 0
I should use more neurons 41% 13
The network overfits 6% 2
I should add more layers 81% 26
I should change the activation function 56% 18
Logistic regression fails 34% 11
One hidden layer is sufficient 6% 2
I should change to a non linear activation 28% 9
wooclap
6% correct 32 / 45


You cannot vote anymore
What happens here?
5 I should change the activation function 85% 22
wooclap
73% correct 26 / 45
california.ipynb
Tout afficher


You cannot vote anymore
What happens here?
3 The network overfits 83% 19
wooclap
83% correct 23 / 45
california.ipynb
Tout afficher


You cannot vote anymore
What happens here?
7 It works ! 100% 21
wooclap
100% correct 21 / 45
california.ipynb
Tout afficher