Course 4: Deep Learning



OZ Course 4: Deep Learning

2025

Course 4: Deep Learning



Summary

Last session

- Unsupervised learning discover structure from unlabeled data
- 2 Clustering
- 3 Decomposition
- Preprocessing and feature selection

Today's session

- Multi-Layer Perceptron
- Convolutional Neural Networks
- Transformers

Course 4: Deep Learning

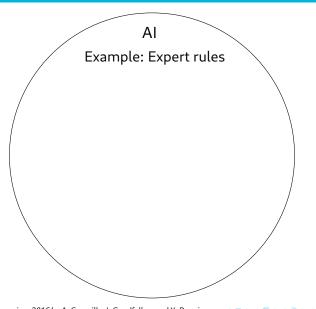
-Summary

2025-

Last session Clustering Decomposition

Unsupervised learning discover structure from unlabeled data

Multi-Layer Perceptron Convolutional Neural Transformers Preprocessing and feature



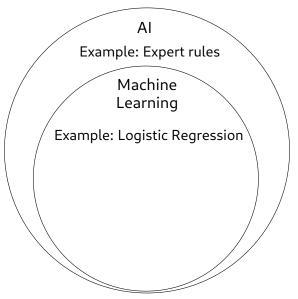
Course 4: Deep Learning

-Global overview...



Deep Learning is a particular case of Machine Learning.

Adapted from Deep Learning, 2016 by A. Courville, I. Goodfellow and Y. Bengio IMT-Atlantique



Course 4: Deep Learning

—Global overview…



Deep Learning is a particular case of Machine Learning.

Adapted from Deep Learning, 2016 by A. Courville, I. Goodfellow and Y. Bengio

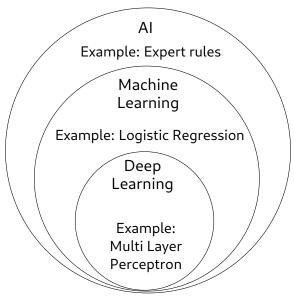
IMT-Atlantique

Course 4: Deep Learning

4 □ > 4 ⑤ > 4 ⓒ > 4 ⓒ >

Course 4: Deep Learning

3/27



Course 4: Deep Learning

2025

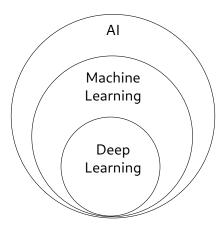
└─Global overview...



Deep Learning is a particular case of Machine Learning.

Adapted from Deep Learning, 2016 by A. Courville, I. Goodfellow and Y. Bengio

IMT-Atlantique Course 4: Deep Learning 3/2



Hence, Deep Learning methods can be supervised, unsupervised, ...

S Course 4: Deep Learning

–Global overview…



Deep Learning in a nutshell (1/3)

Course 4: Deep Learning

Deep Learning in a nutshell (1/3)

Installation of a finishing of a finishing start of the s

Tentative definition

Using *deep* Artificial Neural Networks. Optimized on *large* datasets.

We generally talk about "Neural Networks" instead of "Artificial Neural Networks", but the latter is the most accurate terminology!

The strength of Deep Learning lies in using a lot (a lot) of data.

2025

Deep Learning in a nutshell (1/3)

Course 4: Deep Learning

Deep Learning i

Deep Learning in a nutshell (1/3)

icentatus efficience distinguished per Artificial Neural Networks. Optimized on large datasets. We generally talk about "Neural Networks" instead of "Artificial Neural Networks", for the latter is the most accurate terminology of the strength of Deep Learning lies in using a lot fa feld of data.

Tentative definition

Using *deep* Artificial Neural Networks. Optimized on *large* datasets.

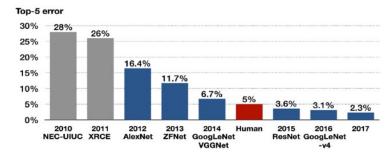
We generally talk about "Neural Networks" instead of "Artificial Neural Networks", but the latter is the most accurate terminology!

The strength of Deep Learning lies in using a lot (a **lot**) of data.

Deep Learning in a nutshell (2/3)

A major breakthrough in image classification:

IMT-Atlantique



Source: Kang, D. Y., Duong, H. P., & Park, J. C. (2020). Application of deep learning in dentistry and implantology. Journal of implantology and applied sciences, 24(3), 148-181.

Details for the human evaluation: Russakovsk, Dieg et al.. ImageNet Large Scale Visual Recognition Challenge, https://arxiv.org/pdf/1409.0575.pdf



Course 4: Deep Learning 5/27

Course 4: Deep Learning

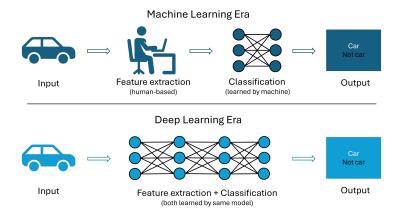
2025

Deep Learning in a nutshell (2/3)



The landscape of Machine Learning changed in 2012: Deep Neural Networks, a technique used in a minority of cases until then, suddenly won the Image Classification contest on ImageNet, a standard image classification dataset. From this point, Deep Neural Networks became mainstream, and the performance of Deep Neural Network models skyrocketed.

Deep Learning in a nutshell (3/3)



 $Inspired from \ https://www.softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/softwaretestinghelp.com/data-mining-vs-machine-learning-vs-machine-l$



Course 4: Deep Learning

—Deep Learning in a nutshell (3/3)



IMT-Atlantique

2025-10

-Outline

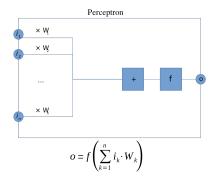
Multi-layer Perceptron



Perceptron

1943, first known implementation in 1957

Perceptron is a nonlinear operation in which weights W are trainable.



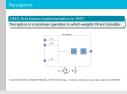
 $Source: By \ Mat \ the \ w \ at \ English \ Wikipedia, \ CC \ BY-SA \ 3.0, \ https://commons.wikimedia.org/w/index.php?curid=23766733$



Course 4: Deep Learning

Multi-layer Perceptron

Perceptron



The Perceptron is a matrix multiplication, followed by a nonlinear function f(). It is important to say that the Perceptron is old!! Hence, AI is not a brand new thing, but an old research domain.

Perceptron

1943, first known implementation in 1957

Perceptron is a nonlinear operation in which weights W are trainable.

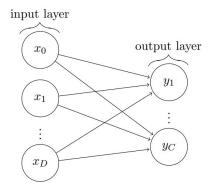


Figure: The arrows represent the weights W.

2025-10-20

Course 4: Deep Learning
—Multi-layer Perceptron

-Perceptron



The Perceptron is a matrix multiplication, followed by a nonlinear function f(). It is important to say that the Perceptron is old!! Hence, AI is not a brand new thing, but an old research domain.

Source: https://davidstutz.de/illustrating-convolutional-neural-networks-in-latex-with-tikz/ 1/2 / 2/2

Optimizing the weights

Loss

■ Prediction:
$$y = f\left(\sum_{d=0}^{D} x_d W_d\right)$$

- Ground truth: ŷ
- Loss (one example:) $\mathcal{L}(x, W, \hat{y}) = d(y, \hat{y})$ (ex: $d(y, \hat{y}) = ||y - \hat{y}||_2^2$)
- Loss (*i* examples): $J(W) = \sum_{i} \mathcal{L}(x^{(i)}, W, \hat{y}^{(i)})$

Gradient descent

- Compute the gradient: $\frac{\partial J(W)}{\partial W}$ (high dimensional derivative)
- Update weights: $W \leftarrow W \eta \frac{\partial J(W)}{\partial W}$

Course 4: Deep Learning

Multi-layer Perceptron

Optimizing the weights

potentialing the weights one production $y = f\left(\frac{C}{2}x_0W_0\right)$ a Production $y = f\left(\frac{C}{2}x_0W_0\right)$ a Concent borth y is a loss (one around $\xi(x_0,W,y) = dy,y$) (or $c_0(x,y,y) = (y,y)$) (or $c_0(x,y,y) = (y,y)$) in a loss (or annothed $\xi(x_0,W,y) = (y,y)$) and $\xi(x_0,y) = (y,y) = (y,y)$ (or annothed $\xi(x_0,y) = (y,y) = (y,y)$) of the product $\frac{C_0(x_0,y)}{2}$ (by the small product $\frac{C_0(x_0,y)}{2}$ (by the small product $\frac{C_0(x_0,y)}{2}$) $\frac{C_0(x_0,y)}{2}$

The Perceptron is a matrix multiplication, followed by a nonlinear function f(). It is important to say that the Perceptron is old!! Hence, AI is not a brand new thing, but an old research domain.

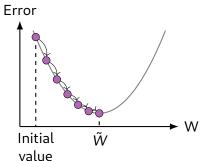
2025

Gradient descent

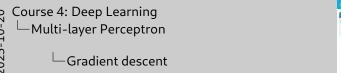
Intuition behind the gradient descent

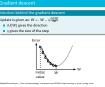
Update is given as: $W \leftarrow W - \eta \frac{\partial J(W)}{\partial W}$

- $\partial J(W)$ gives the direction
- \blacksquare η gives the size of the step



Adapted from https://tex.stackexchange.com/questions/561921/replicating-a-plot-using-tikz





The gradient follows the increase of the error function, hence the inverse of the gradient follows its decrease. Parameter η , the learning rate, gives the size of the step to take at each iteration. If too small, the model will slowly converge. If too large, the model can be unstable and never reach the optimal solution. In practice, setting the learning rate is not easy.

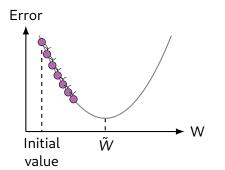
Gradient descent

Intuition behind the gradient descent

Update is given as: $W \leftarrow W - \eta \frac{\partial J(W)}{\partial W}$

- lacksquare $\partial J(W)$ gives the direction
- lacksquare η gives the size of the step

Small step:



Course 4: Deep Learning

Multi-layer Perceptron

Gradient descent



Small step:

The gradient follows the increase of the error function, hence the inverse of the gradient follows its decrease. Parameter η , the learning rate, gives the size of the step to take at each iteration. If too small, the model will slowly converge. If too large, the model can be unstable and never reach the optimal solution. In practice, setting the learning rate is not easy.

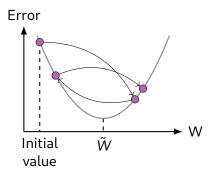
Gradient descent

Intuition behind the gradient descent

Update is given as: $W \leftarrow W - \eta \frac{\partial J(W)}{\partial W}$

- $\partial J(W)$ gives the direction
- \blacksquare η gives the size of the step

Large step:





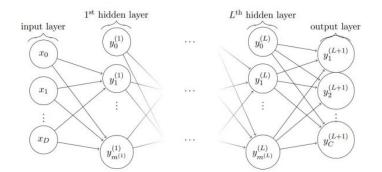


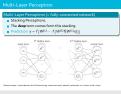
The gradient follows the increase of the error function, hence the inverse of the gradient follows its decrease. Parameter η , the learning rate, gives the size of the step to take at each iteration. If too small, the model will slowly converge. If too large, the model can be unstable and never reach the optimal solution. In practice, setting the learning rate is not easy.

Multi-Layer Perceptron

Multi-Layer Perceptron (= fully-connected network)

- Stacking Perceptrons.
- The *deep* term comes form this stacking
- Prediction: $y = f(W^{(L)} \cdots f(W^{(2)} f(W^{(1)} x)))$





The bias are removed from the equations for simplicity, but say that they exist orally.

Source: https://davidstutz.de/illustrating-convolutional-neural-networks-in-latex_with-tikz/

- **Each** building block $\mathbf{y}^{(l+1)}$ is called a **layer**.

Course 4: Deep Learning Multi-layer Perceptron

-Neural Networks

2025

 $\mathbf{v}^{(l+1)} = f(\mathbf{W}^{(l)}\mathbf{v}^{(l)} + \mathbf{b}^{(l)}) = \phi^{(l)}(\mathbf{v}^{(l)})$ ■ Each building block v(1+1) is called a laver

Definitions

$$\mathbf{v}^{(l+1)} = f(\mathbf{W}^{(l)}\mathbf{v}^{(l)} + \mathbf{b}^{(l)}) = \phi^{(l)}(\mathbf{v}^{(l)})$$

- **Each** building block $\mathbf{y}^{(l+1)}$ is called a **layer**.
- One element i of a layer $(y_i^{(l)})$ is called a **neuron** (both as input and output).
- The nonlinear function *f* is called the **activation function**.
- W^(l) are called weights.
- **b**(l) is called the **bias**.

Note: while each layer can have a different activation function f, it is standard that each layer uses the same.

Course 4: Deep Learning

Multi-layer Perceptron

Neural Networks

Neural Networks

Definitions $y^{(i+1)} = (\chi W^{(i)} y^{(i)} + b^{(i)}) = \phi^{(i)} V^{(i)}$ If Each building block $y^{(i+1)}$ is called a flavor

If the object $y^{(i+1)}$ is called a neuron (both as input output).

If the neurons of a layer $(y^{(i)})$ is called a neuron (both as input output).

If the neurons in extra first a called the activation functions $y^{(i+1)} = y^{(i)} = y^{(i)}$

Definitions

$$\mathbf{v}^{(l+1)} = f(\mathbf{W}^{(l)}\mathbf{v}^{(l)} + \mathbf{b}^{(l)}) = \phi^{(l)}(\mathbf{v}^{(l)})$$

- **Each** building block $\mathbf{y}^{(l+1)}$ is called a **layer**.
- One element i of a layer $(y_i^{(l)})$ is called a **neuron** (both as input and output).
- The nonlinear function f is called the activation function.



Course 4: Deep Learning Multi-layer Perceptron

-Neural Networks

 $\mathbf{v}^{(i+1)} = f(\mathbf{W}^{(i)}\mathbf{v}^{(i)} + \mathbf{b}^{(i)}) = \phi^{(i)}(\mathbf{v}^{(i)})$

 Each building block v⁽¹⁺¹⁾ is called a laver One element i of a layer (y_i⁽ⁱ⁾) is called a neuron (both as input an The nonlinear function f is called the activation function

Definitions

$$\mathbf{v}^{(l+1)} = f(\mathbf{W}^{(l)}\mathbf{v}^{(l)} + \mathbf{b}^{(l)}) = \phi^{(l)}(\mathbf{v}^{(l)})$$

- **Each** building block $\mathbf{y}^{(l+1)}$ is called a **layer**.
- One element i of a layer $(y_i^{(l)})$ is called a **neuron** (both as input and output).
- The nonlinear function *f* is called the **activation function**.
- **W**^(l) are called **weights**.
- **b**(l) is called the **bias**.

Note: while each layer can have a different activation function f, it is standard that each layer uses the same.

Course 4: Deep Learning

Multi-layer Perceptron

Neural Networks

Neural Networks

Definitions $\mathbf{y}^{(t+2)} = \mathbf{q}(\mathbf{y}^{(t)}\mathbf{y}^{(t)} + \mathbf{b}^{(t)}) = \phi^{(t)}\mathbf{y}^{(t)}$ at Each building block $\mathbf{y}^{(t+2)}$ is called a layer.

at One element of a layer $(\mathbf{y}_i^{(t)})$ is called a neuron (both as input output) in the production of the produ

Definitions

$$\mathbf{v}^{(l+1)} = f(\mathbf{W}^{(l)}\mathbf{v}^{(l)} + \mathbf{b}^{(l)}) = \phi^{(l)}(\mathbf{v}^{(l)})$$

- **Each** building block $\mathbf{y}^{(l+1)}$ is called a **layer**.
- One element i of a layer $(y_i^{(l)})$ is called a **neuron** (both as input and output).
- The nonlinear function *f* is called the **activation function**.
- **W**^(l) are called **weights**.
- **b**(l) is called the **bias**.

Note: while each layer can have a different activation function f, it is standard that each layer uses the same.

Course 4: Deep Learning

Multi-layer Perceptron

Neural Networks

Neural Networks $\begin{aligned} & y^{n+1} - (y u^n v^n + v^n) - \phi^n (y^n) \\ & = \text{Sub-building block } y^{n+1} \in \text{cutled is layer.} \end{aligned}$ $& = \text{Due determent of a layer } (y^n)^n \text{ such as warve (light has input.}$ & = The motivate princh for called the activation function. $& = w^n \text{ in a cutled weights.}$

Definitions

$$\mathbf{v}^{(l+1)} = f(\mathbf{W}^{(l)}\mathbf{v}^{(l)} + \mathbf{b}^{(l)}) = \phi^{(l)}(\mathbf{v}^{(l)})$$

- **Each** building block $\mathbf{y}^{(l+1)}$ is called a **layer**.
- One element i of a layer $(y_i^{(l)})$ is called a **neuron** (both as input and output).
- The nonlinear function *f* is called the **activation function**.
- **W**^(l) are called **weights**.
- **b**(l) is called the **bias**.

Note: while each layer can have a different activation function f, it is standard that each layer uses the same.

Course 4: Deep Learning

Multi-layer Perceptron

Neural Networks

Neural Networks Definitions $y^{(i+1)} = (y_i w^{(i)} y^{(i)} + b^{(i)}) = \phi^{(i)}(y^{(i)})$ If Each building block $y^{(i+1)}$ is called a layer. If Does element in of a layer $(y^{(i)})$ is called a newern (both as in The nonlinear function if is called the activation function $y^{(i)}$ is called the activation function.

Note: while each laver can have a different activation function f. it is

N/I is called the blas

standard that each layer uses the same

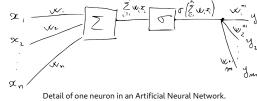
Why is it called Neural Network?

"Neurons" may be seen as **loosely** inspired from the human brain.



https://en.wikipedia.org/wiki/Neural_ network_(machine_learning)#/media/

File:Neuron3.png



* Late note: I forgot the bias in the graphic.

This is an **analogy**, artificial neural networks are **not** following the human brain (in general).



2025

10-

Course 4: Deep Learning

Multi-layer Perceptron

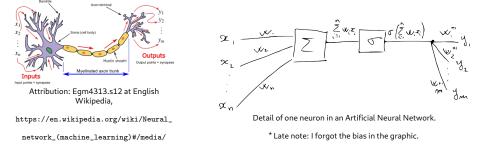
└─Why is it called Neural Network?



Why is it called Neural Network?

File:Neuron3.png

"Neurons" may be seen as **loosely** inspired from the human brain.



This is an **analogy**, artificial neural networks are **not** following the human brain (in general).



Course 4: Deep Learning Multi-layer Perceptron

10-

-Why is it called Neural Network?

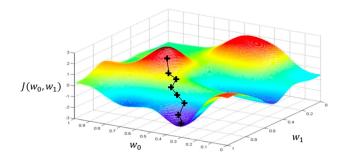


IMT-Atlantique Course 4: Deep Learning

Gradient descent for deep neural networks

Backpropagation

- Gradient descent for all layers (chain rule).
- Simplified equation: $\frac{\partial J(W)}{\partial W} = \frac{\partial J(W)}{\partial W^{(L)}} \frac{\partial W^{(L)}}{\partial W^{(L-1)}} \frac{\partial W^{(L-1)}}{\partial W^{(L-2)}} \cdots \frac{\partial W^{(2)}}{\partial W^{(1)}}$
- The error **backpropagates** through the network (reverse path)
- Computationally efficient, but finds a local minimum (at best)



Course 4: Deep Learning

Source: http://introtodeeplearning.com/

IMT-Atlantique

(ロト 4個 h 4 種 h 4 種 h) 種 り 9 (で)

Course 4: Deep Learning
—Multi-layer Perceptron

The error backpropagates

Computationally efficient,

—Gradient descent for deep neural networks

Careful: the given equation is not correct, just conveniently simplified. For better details, refer to https://en.wikipedia.org/wiki/Backpropagation.

Batch

Course 4: Deep Learning -Multi-layer Perceptron -Batch

10-

2025

■ The i examples are divided in batches (small excerpt) Allows one to train without loading the whole dataset in memory Accelerates the learning phase

Batch

- The *i* examples are divided in *batches* (small excerpt)
- Allows one to train without loading the whole dataset in memory
- Accelerates the learning phase

Limits of MLP

Course 4: Deep Learning -Multi-layer Perceptron Limits of MLP

10-

2025

Limits of Multi-Layer Perceptrons

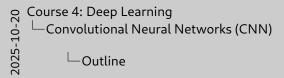
- Computationally heavy for large inputs
- Large number of parameters: prone to overfitting
- No notion of structure in the input: everything is a vector

 Large number of parameters: prone to overfitting No notion of structure in the input: everything is a vector

Outline

1 Multi-laver Perceptron

- 2 Convolutional Neural Networks (CNN)
- **3** Transformers
- 4 State of Deep Learning today



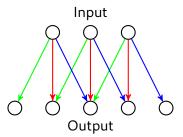


Convolutional Neural Network (1/5)

Principle

- Applying a kernel to the input, on small parts of the image at a time.
- Weights of the kernel are **learned** and **shared**!
- 2D convolution was a game-changer for image processing
- Translation invariance

Convolutional layer



IMT-Atlantique



Course 4: Deep Learning 16/

Course 4: Deep Learning

Convolutional Neural Networks (CNN)

202

—Convolutional Neural Network (1/5)

Convolution neural networks are the most common architecture for neural networks nowadays. They are particularly successful for image processing, were not challenged until very recently, with Vision Transformers (see Transformers in the last slides). The translation invariance is really important for image processing, as objects can be anywhere on the image. It contributed to the success of Convolutional Neural Networks.

Convolutional Neural Network (2/5)

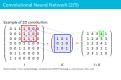
Example of 2D convolution:

Source: https://tex.stackexchange.com/questions/437007/drawing-a-convolution-with-tikz

Course 4: Deep Learning

Convolutional Neural Networks (CNN)

-Convolutional Neural Network (2/5)

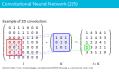


Convolutional Neural Network (2/5)

Example of 2D convolution:

Source: https://tex.stackexchange.com/questions/437007/drawing-a-convolution-with-tikz

Course 4: Deep Learning
Convolutional Neural Networks (CNN)
Convolutional Neural Network (2/5)



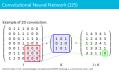
Convolutional Neural Network (2/5)

Example of 2D convolution:

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0$$

Source: https://tex.stackexchange.com/questions/437007/drawing-a-convolution-with-tikz

Course 4: Deep Learning
Convolutional Neural Networks (CNN)
Convolutional Neural Network (2/5)



Convolutional Neural Network (3/5)

Example of 2D pooling:

1	2	3	1
1	1	1	1
2	3	1	6
8	1	4	5





maxpool, kernel 2, stride 2

Course 4: Deep Learning
Convolutional Neural Networks (CNN)
Convolutional Neural Network (3/5)



2025-

Convolutional Neural Network (3/5)

Example of 2D pooling:

1	2	3	1
1	1	1	1
2	3	1	6
8	1	4	5





maxpool, kernel 2, stride 2

Course 4: Deep Learning
Convolutional Neural Networks (CNN)
Convolutional Neural Network (3/5)



2025-

Convolutional Neural Network (3/5)

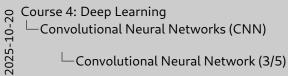
Example of 2D pooling:

1	2	3	1
1	1	1	1
2	3	1	6
8	1	4	5



238-

maxpool, kernel 2, stride 2





Convolutional Neural Network (3/5)

Example of 2D pooling:

1	2	3	1
1	1	1	1
2	3	1	6
8	1	4	5





maxpool, kernel 2, stride 2

Course 4: Deep Learning
Convolutional Neural Networks (CNN)
Convolutional Neural Network (3/5)

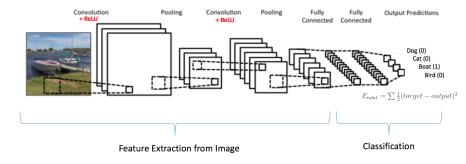


2025-

Convolutional Neural Network (4/5)

And repeat...

- Convolutional neural network: mainly Convolution + Pooling.
- ...But many other components may be added! (batch norm, dropout, skip connections, concatenation, ...)



Source: https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

IMT-Atlantique Course 4: Deep Learning 19/27

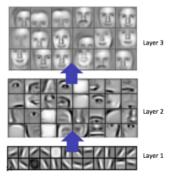
Course 4: Deep Learning
Convolutional Neural Networks (CNN)
Convolutional Neural Network (4/5)



Convolutional Neural Network (5/5)

Why convolutions?

- Kernels capture important information in images
- The kernels become more and more complex with the depth of the network



Course 4: Deep Learning
Convolutional Neural Networks (CNN)
Convolutional Neural Network (5/5)

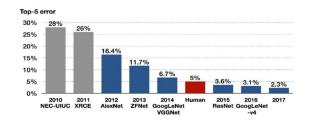


Convolution are able to catch simple shapes (lines, edges) which turn into complex shapes in the subsequent layers.

Source: https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

IMT-Atlantique Course 4: Deep Learning 20/27

What happened in 2012?



A combination of...

IMT-Atlantique

- Convolutional neural networks
- A very large dataset (ImageNet)
- Clever tricks (ex: data augmentation, i.e. altering image during training, very standard in Deep Learning)

Course 4: Deep Learning

■ The use of GPUs for computation

Course 4: Deep Learning

Convolutional Neural Networks (CNN)

-What happened in 2012?



- A very large dataset (ImageNet)
 Clever tricks (ex: data augmentation, i.e. altering image during
- training, very standard in Deep Learning)

Outline

Multi-laver Percentror

- 2 Convolutional Neural Networks (CNN
- **3** Transformers
- 4 State of Deep Learning today



Outline

B Made days Pursuption

Commissional National Mathematic (CMM)

Transformers

State of Deep Learning today

Transformers

Course 4: Deep Learning

Transformers

Transformers

2025

clanded antificture newadays If No convolution If Section statistics what should be important for context? If Used for fast, image, audio, ...

Standard architecture nowadays

- No convolution
- Based on attention: what should be important for context?
- Used for text, image, audio, ...

Transformers

Course 4: Deep Learning

Transformers

Transformers

2025

Standard architecture nowadays

- No convolution
- Based on attention: what should be important for context?
- Used for text, image, audio, ...

Will be presented in the next class (foundation models)!

Transformers

Sandard architecture newadays

in No convolution

in Based on artention what should be important for context?

Usude for test, impay, audio, ...

Will be presented in the next class (foundation models)

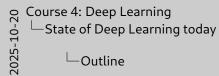
Outline

Multi-laver Percentror

2 Convolutional Neural Networks (CNN

3 Transformers

4 State of Deep Learning today



State-of-the-Art nowadays

Image classification

- Image classification for a single dataset is (almost) solved
- Challenges of adapting models to unseen datasets
- Challenges when data is scarce
- Specific domains with few variability or complex classification are still challenging (ex: medical imaging)

Large Language Models

- Large Language Models caught everyone's attention (ChatGPT)
- Challenges of reducing their resources (data/power)
- May hallucinate: lack of robustness

Many other domains

Multimodal models (DALL-E, ...), Audio, Games, Video, ...

Course 4: Deep Learning

State of Deep Learning today

-State-of-the-Art nowadays

State- of the-Art novadays
Impug classification
Impug classification
Impug classification for a single dissert is binnest solved
Impug classification for a single dissert of binnest solved
Impug classification for a single dissert of binnest solved
Impug classification dissert of binnest solved
Impug classification are stated to binnest solved in the single solved

4)Q(**▷**

State-of-the-Art nowadays

Image classification

- Image classification for a single dataset is (almost) solved
- Challenges of adapting models to unseen datasets
- Challenges when data is scarce
- Specific domains with few variability or complex classification are still challenging (ex: medical imaging)

Large Language Models

- Large Language Models caught everyone's attention (ChatGPT)
- Challenges of reducing their resources (data/power)
- May hallucinate: lack of robustness

Many other domains

Multimodal models (DALL-E, ...), Audio, Games, Video, ...

Course 4: Deep Learning

State of Deep Learning today

State-of-the-Art nowadays

mage classification

Il mage classification for a single dataset is (almost) solved

il challenges of adapting models to unseen datasets

il challenges when data is scarce

il Spacific domains with flow variability or compiles classification are
still challenging (ex: medical imaging)

Large Language Models caught everyone's attention (ChatGPT
 Challenges of reducing their resources (data/power)
 May hallucinate: lack of robustness

IMT-Atlantique Course 4: Deep Learning 23/27

State-of-the-Art nowadays

Image classification

- Image classification for a single dataset is (almost) solved
- Challenges of adapting models to unseen datasets
- Challenges when data is scarce
- Specific domains with few variability or complex classification are still challenging (ex: medical imaging)

Large Language Models

- Large Language Models caught everyone's attention (ChatGPT)
- Challenges of reducing their resources (data/power)
- May hallucinate: lack of robustness

Many other domains

Multimodal models (DALL-E, ...), Audio, Games, Video, ...

Course 4: Deep Learning

State of Deep Learning today

State-of-the-Art nowadays

Image classification for a single dataset is (almost) solved Challenges of adapting models to unseen datasets Challenges when data is carcer Specific domains with few variability or complex classification are still challenging (ex: medical kinaging)

Large Language Models caught everyone's attention (ChatGPT
 Challenges of reducing their resources (data/power)
 May hallucinate: lack of robustness

rother domains modal models (DALL-E, ...), Audio, Games, Video

IMT-Atlantique Course 4: Deep Learning 23/27

Focus on Large Language Models

Many models

- GPT (Open-AI)
- LLaMA (Meta)
- Gemini (Google)
- Mistral 8x7B (MistralAI)
- Many others... And more to come!

- The network learns to reconstruct masked words
- No supervision!
- Allows to leverage immense datasets (ex: GPT-3 was learned on an **Internet scale** dataset)

Course 4: Deep Learning State of Deep Learning today

-Focus on Large Language Models

- LLaMA (Meta) Gemini (Google)
- Mistral 8x7B (MistralAI)

Focus on Large Language Models

Many models

- GPT (Open-AI)
- LLaMA (Meta)
- Gemini (Google)
- Mistral 8x7B (MistralAI)
- Many others... And more to come!

Masked Language Modeling

How are **you** doing today? \rightarrow How are ... doing today?

Course 4: Deep Learning State of Deep Learning today

-Focus on Large Language Models

- Mistral Rv7R (Mistral A1)

How are you doing today? → How are ... doing today?

Focus on Large Language Models

Many models

- GPT (Open-AI)
- LLaMA (Meta)
- Gemini (Google)
- Mistral 8x7B (MistralAI)
- Many others... And more to come!

Masked Language Modeling

How are **you** doing today? \rightarrow How are ... doing today?

- The network learns to reconstruct masked words.
- No supervision!
- Allows to leverage immense datasets (ex: GPT-3 was learned on an Internet scale dataset)

Course 4: Deep Learning State of Deep Learning today

-Focus on Large Language Models

LLaMA (Meta)

Gemini (Google) Mistral 8x7B (MistralAI) Many others... And more to come!

How are you doing today? → How are ... doing today?

No supervision!

 Allows to leverage immense datasets (ex: GPT-3 was learned or an Internet scale dataset)

Large Language Models are greedy

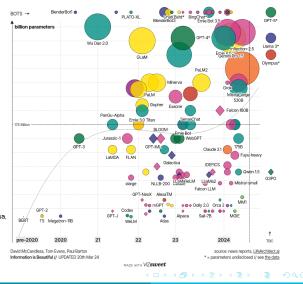
Model sizes

- AlexNet (2012):62 Millionparameters
- GPT-3 (2020): 175 Billion parameters

Image source:
https://informationisbeautiful.

net/visualizations/

the-rise-of-generative-ai-large-langua



■ Amazon-owned Anthropic Apple Chinese Google Meta / Facebook Microsoft OpenAl Other

Course 4: Deep Learning

—State of Deep Learning today

Large Language Models are greedy



2025-

Deep Learning

Conclusion

- Deep Learning algorithms: **powerful** without feature extraction
- They require **a lot** of data to be trained
- The architecture plays an important role

Common criticisms

- Hard to interpret
- Reproduce biases from data
- May require massive amounts of energetic consumption

Going further

- Details and maths behind IA: https://youtu.be/aircAruvnKk
- Ethics and reflexions (french): Science4all & M.Phi

https://woutu.be/HbFadtOxs4k|https://woutu.be/ XJsAOsTOBo

Course 4: Deep Learning

State of Deep Learning today

Deep Learning

eep Learning
onclusion
II Deep Learning algorithms: powerful without
III They require a lot of data to be trained
III The architecture plays an important nole
III The architecture plays an important nole
III The architecture plays an important nole
III They require the second of the plays and important nole
III They require the second of the plays and important nole
III They require the second of the plays and in the plays are the second of the plays are the second of the plays are the second of the plays are the p

Deep Learning

Conclusion

- Deep Learning algorithms: **powerful** without feature extraction
- They require **a lot** of data to be trained
- The architecture plays an important role

Common criticisms

- Hard to interpret
- Reproduce biases from data
- May require massive amounts of energetic consumption

Course 4: Deep Learning

Going further

IMT-Atlantique

- Details and maths behind IA: https://youtu.be/aircAruvnKk
- Ethics and reflexions (french): Science4all & M.Phi

https://voutu.be/HbFadtOys4k|https://voutu.be/ XJs4OsTOBo

Course 4: Deep Learning

State of Deep Learning today

Deep Learning

202

eep Learning
onclusion

Deep Learning algorithms: powerful without fea

They require a let of data to be trained

The architecture plays an important role
one mon criticisms

Renrodure biases from data

rther ils and maths behind IA: https://journ.https://doi.org/ is and reflexions (french): Science4all & M.Phi

Deep Learning

Conclusion

- Deep Learning algorithms: **powerful** without feature extraction
- They require **a lot** of data to be trained
- The architecture plays an important role

Common criticisms

- Hard to interpret
- Reproduce biases from data
- May require massive amounts of energetic consumption

Going further

- Details and maths behind IA: https://youtu.be/aircAruvnKk
- Ethics and reflexions (french): Science4all & M.Phi

https://youtu.be/HbFadtOxs4k|https://youtu.be/_XJsAQsTOBo

Course 4: Deep Learning

State of Deep Learning today

Deep Learning

sep Lemmg

Dept. Service of the serv

■ Ethics and reflexions (french): Science4all & M.Phi

Practical session

Course 4: Deep Learning

State of Deep Learning today

Practical session

2025

Lab

B Lab Pytorch: manipulating the basics of PyTorch

B Lab Modality: try a first baseline on your modality

Lab

- Lab Pytorch: manipulating the basics of PyTorch
- Lab Modality: try a first baseline on your modality