

Scruter pour mieux comprendre : Deep Learning et mécanismes d'attention

Christian Wolf
INRIA-Chroma, CITI, LIRIS, INSA-Lyon





= ?

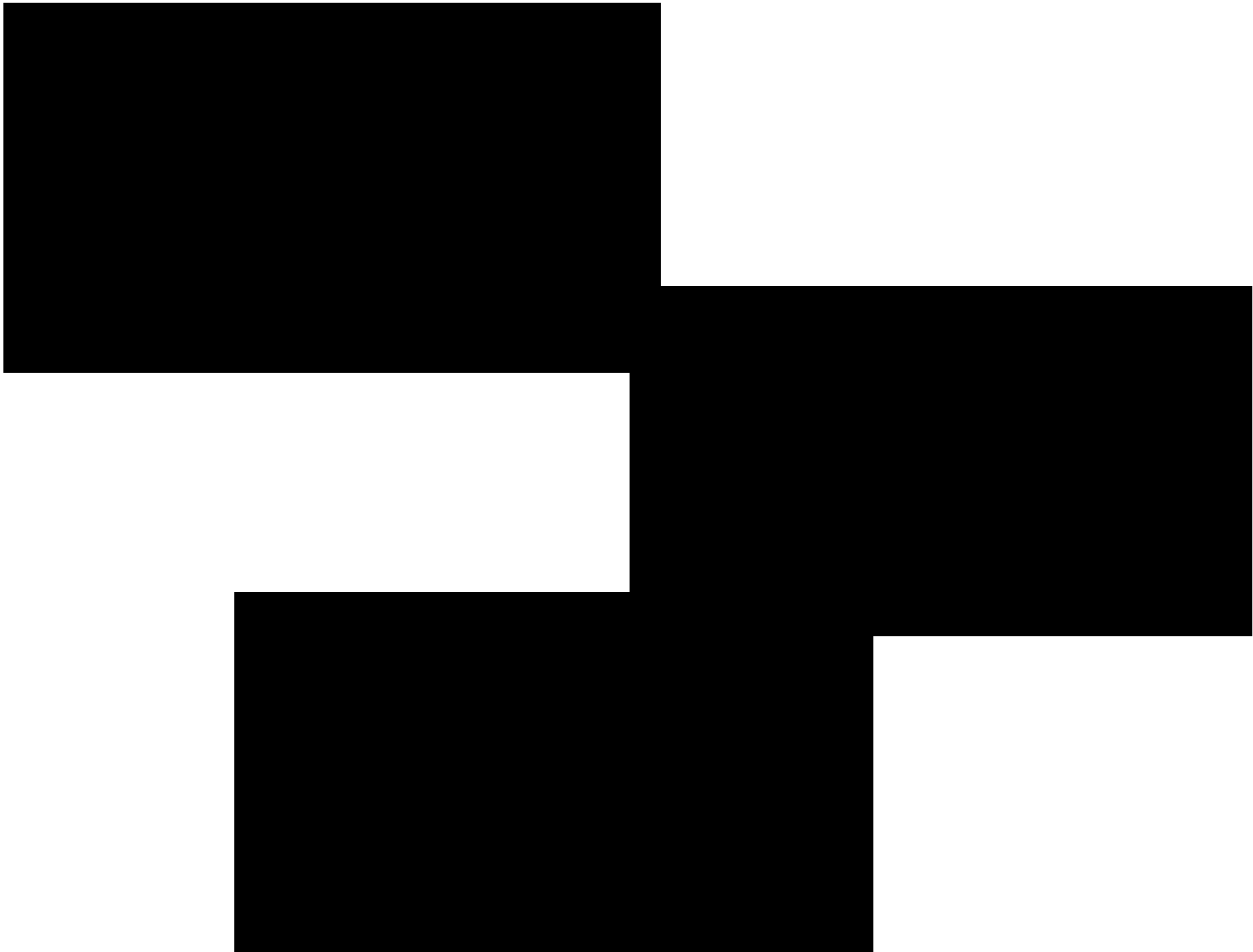
[illegible]



Ami ou pas?
Sourire ou Courir?



Ou suis-je?



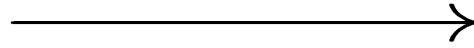
Intelligence et pensée critique

Pensée critique et créativité

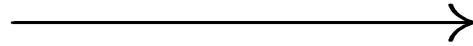
Alors, avant de s'inquiéter de nos compétences face à ce monde présumé d'intelligences artificielles – disons, IA pour désigner cette croyance, nous nommerons autrement les éléments scientifiques établis -, nous devons aiguïser nos compétences humaines clés : la pensée critique et la créativité. Tout d'abord, si nous faisons preuve de pensée critique, il faudrait commencer par nous questionner sur l'expression IA. Est-ce que le terme d'intelligence est pertinent pour désigner des applications informatiques basées notamment sur l'apprentissage machine (*machine learning*) ? Ces algorithmes ont pour objectif de développer des systèmes capables de capter, de traiter et de réagir face à des informations (massives) selon des mécanismes qui s'adaptent au contexte ou aux données pour maximiser les chances d'atteindre les objectifs définis pour le système. Ce comportement qui peut paraître "intelligent" a été créé par des humains et présente des limites liées d'une part à la capacité humaine actuelle de

Blog « Binaire », Le monde, 14 Mai 2018

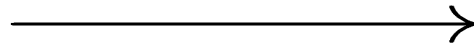
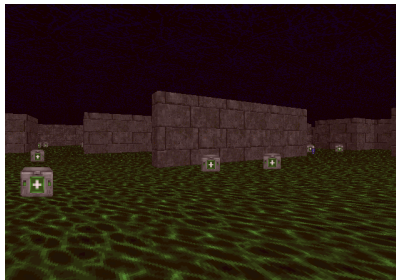
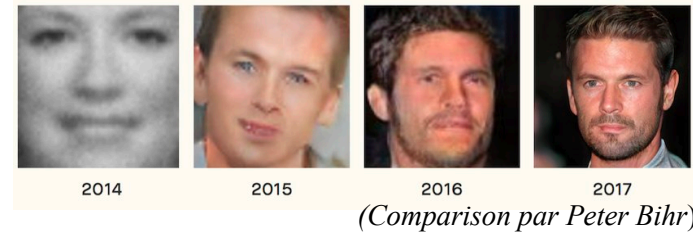
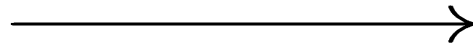
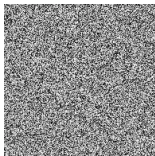
Apprendre à prendre de décisions



{chien, chat, pokemon, helicoptre, ...}



{0, 1, ... 24, 25, 26, ..., 98, 99, ...}



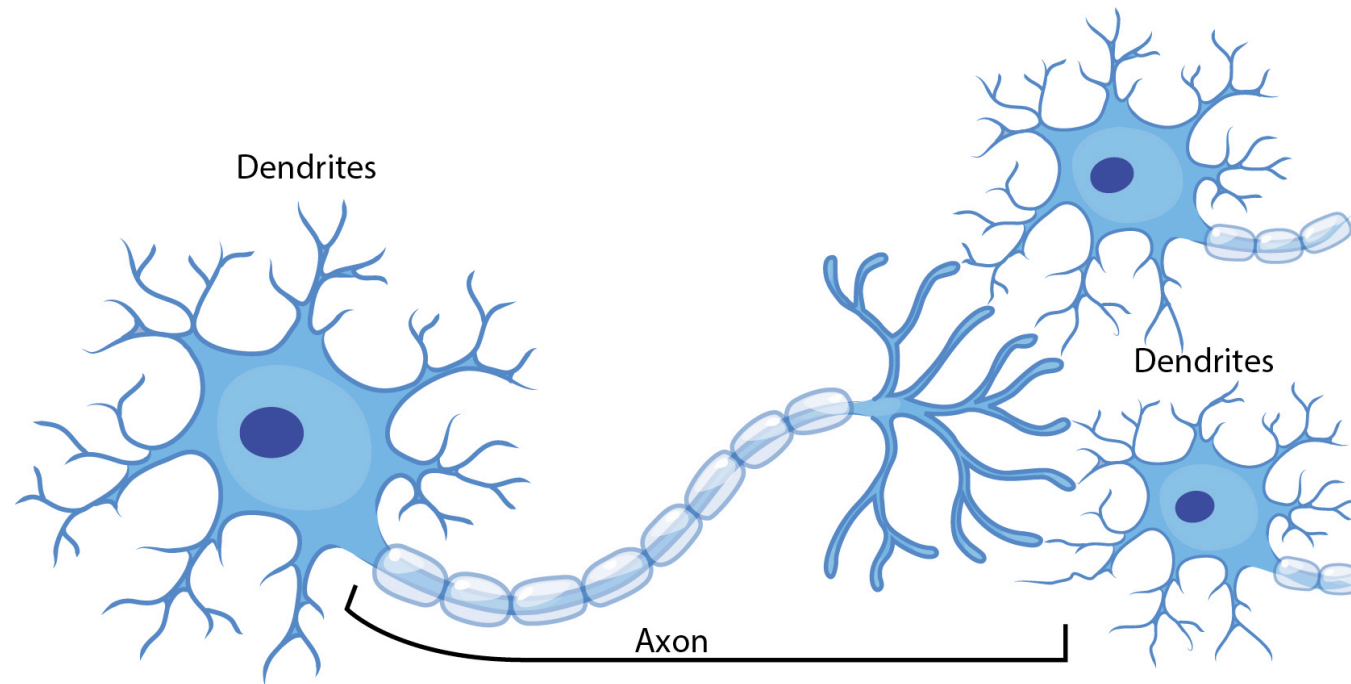
{Gauche, droite, avancer, reculer, ...}

L'objectif est de prédire une valeur y à partir d'une entrée observée x

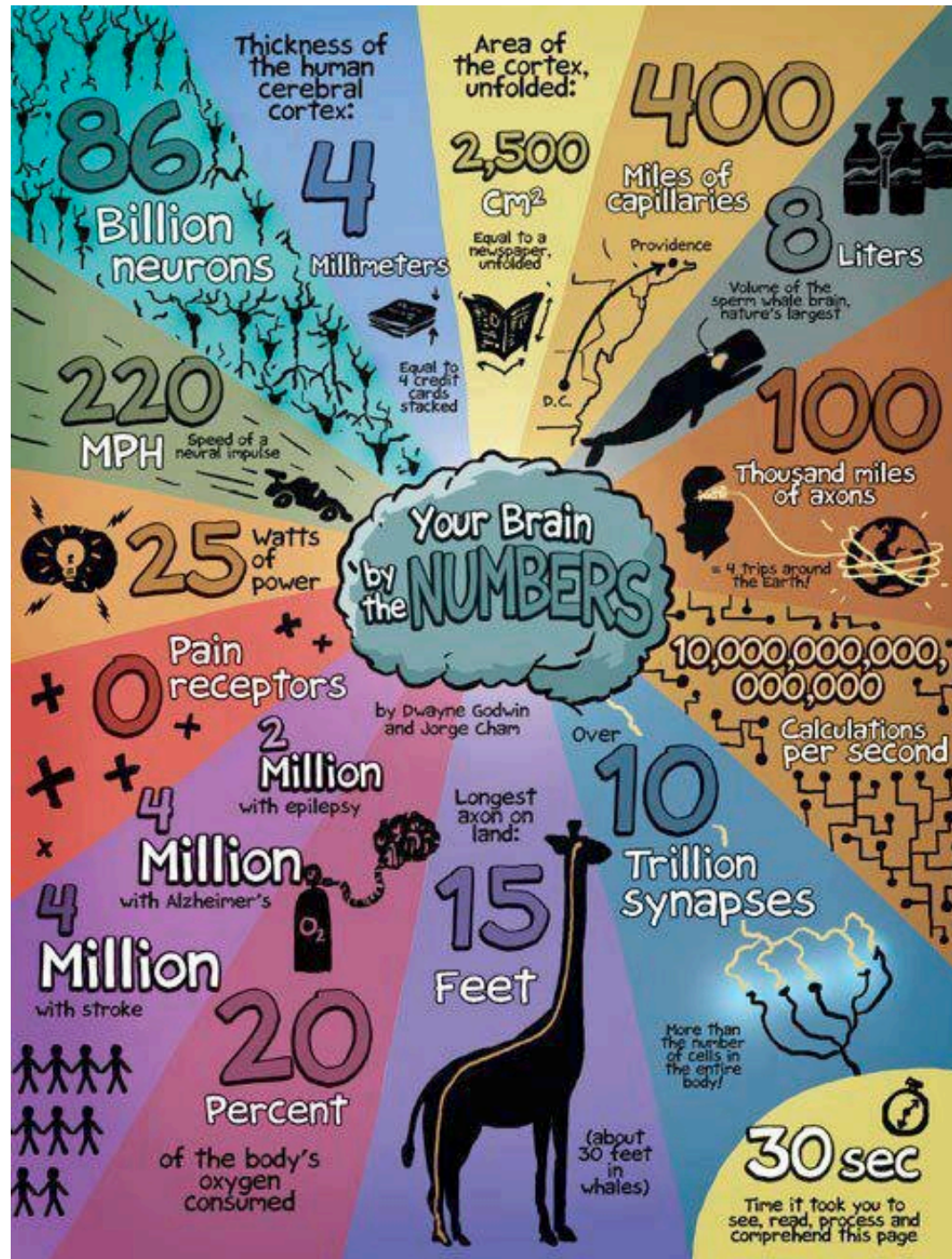
$$y = h(x, \theta)$$

Les paramètres θ sont appris à partir de données.

Un neurone en « VO »



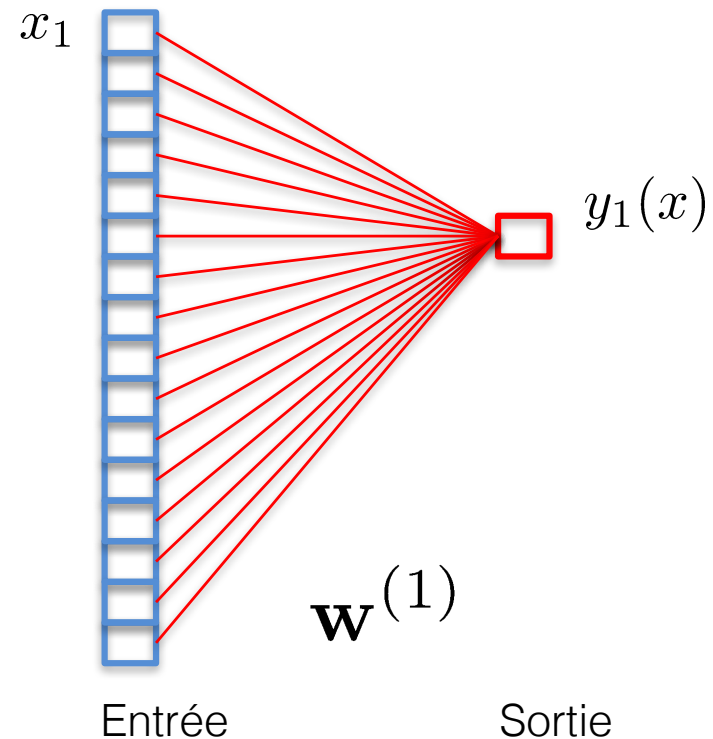
Devin K. Phillips



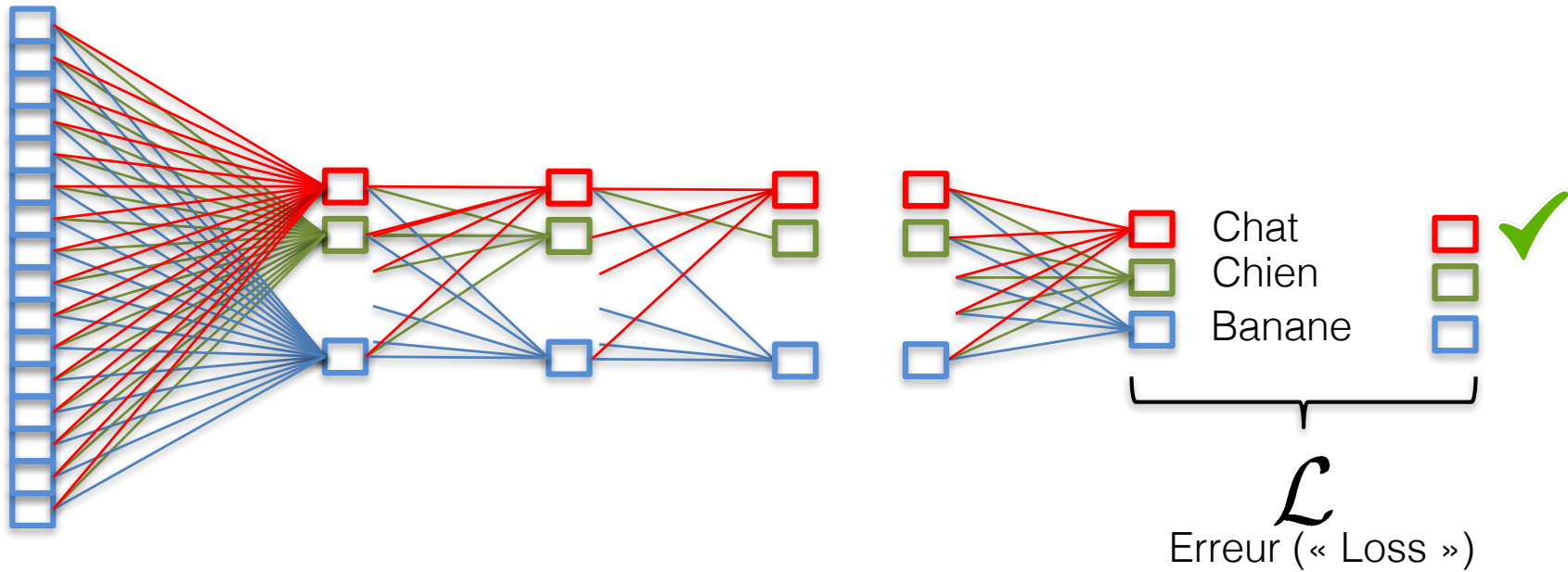
Réseaux de neurones

« Perceptron »

$$y(\mathbf{x}, \mathbf{w}) = \sum_{i=0}^D \mathbf{w}_i \mathbf{x}_i$$

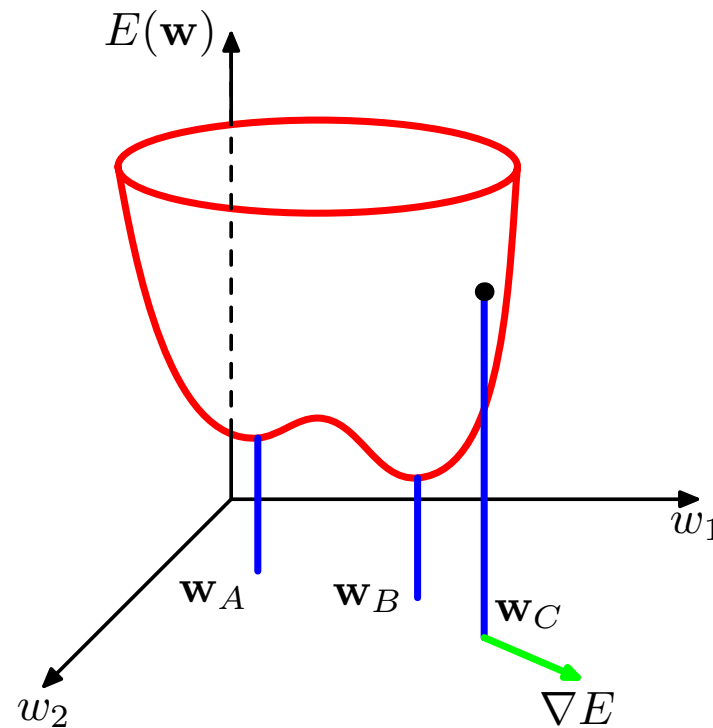


Réseaux « profonds »



Descente de gradients

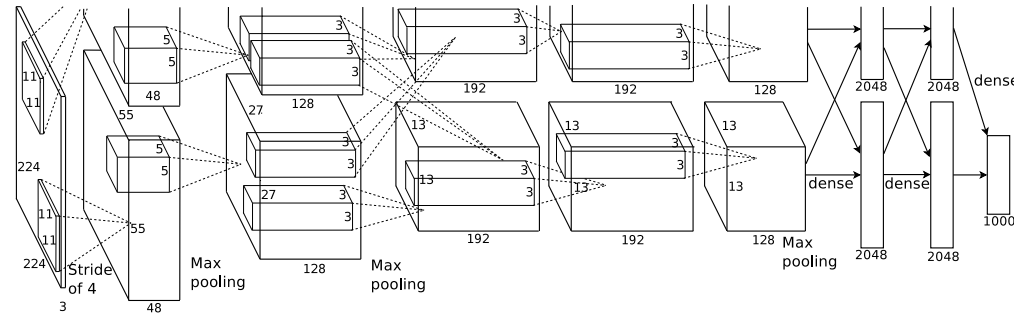
Minimisation de l'erreur commise sur des données connues (étiquetées).



Blocage possible dans un minimum local (et on s'en fiche)

Le deep du deep

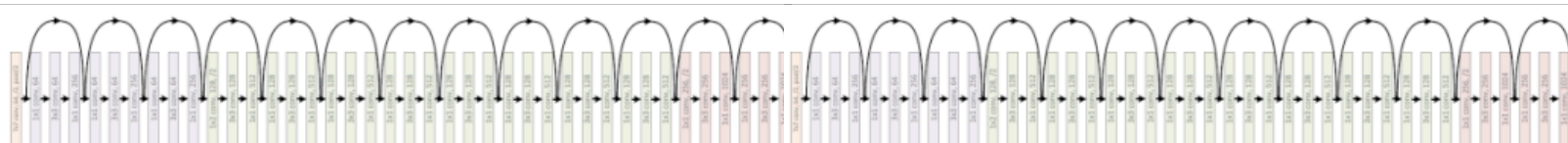
2012 : AlexNet, 8 couches entraînées sur 2 GPUs. Nouvelle technique : dropout



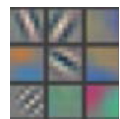
2014 : GoogLeNet, 20 couches. Nouvelle technique : intermediate supervision



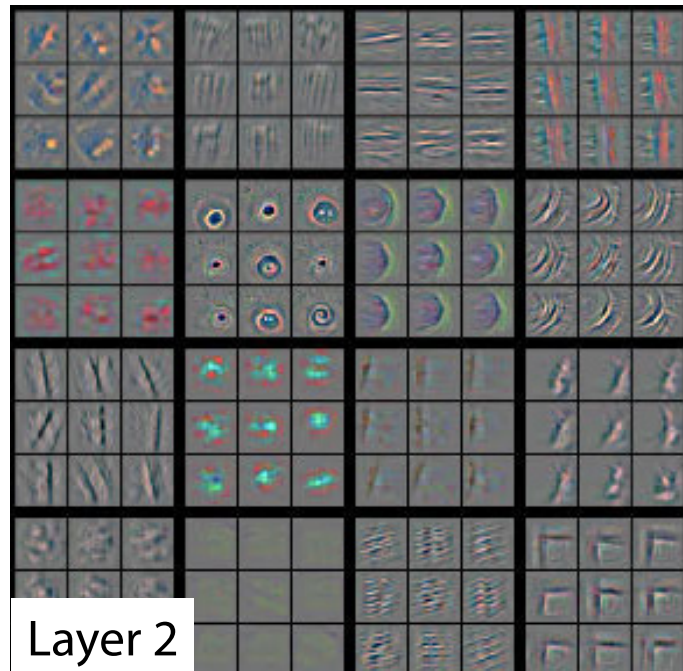
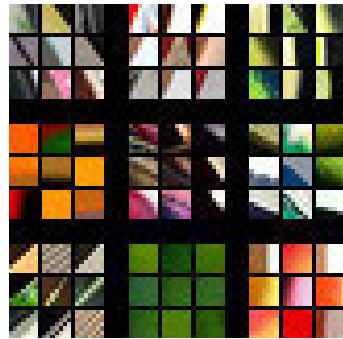
2015 : Microsoft research, 150 couches (!!). Nouvelle technique : *residual learning*



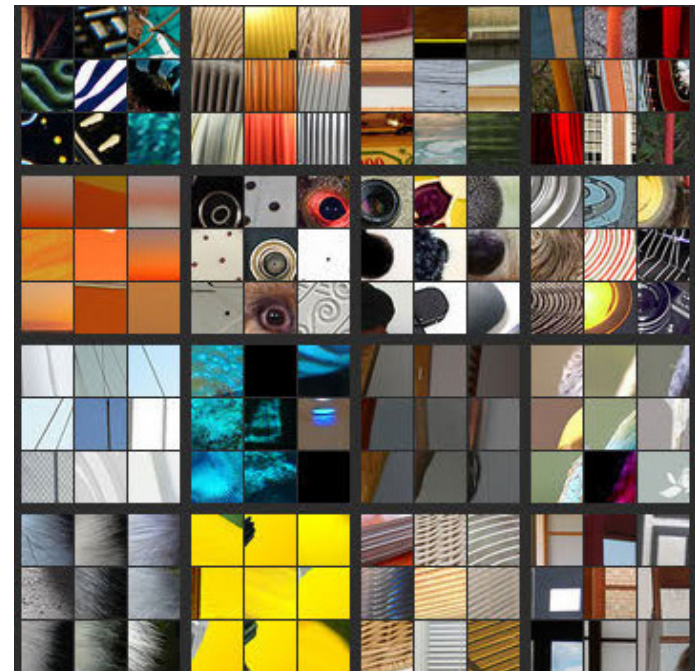
Visualisation



Layer 1

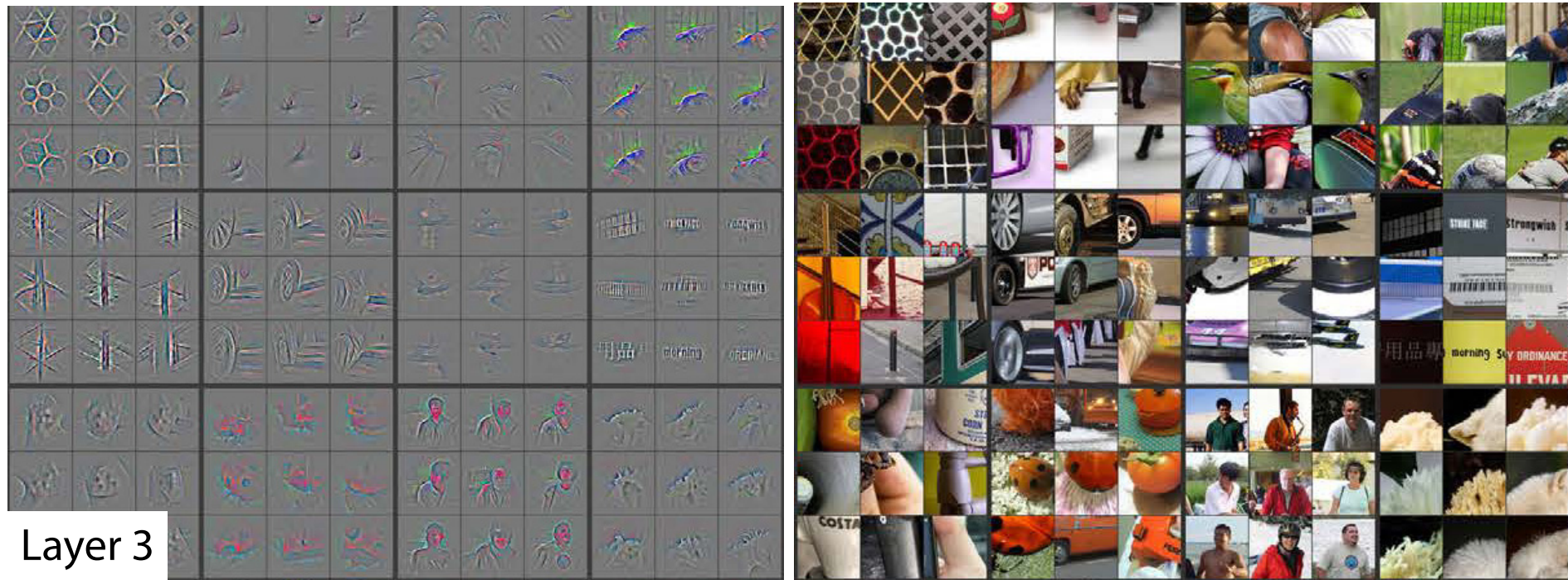


Layer 2



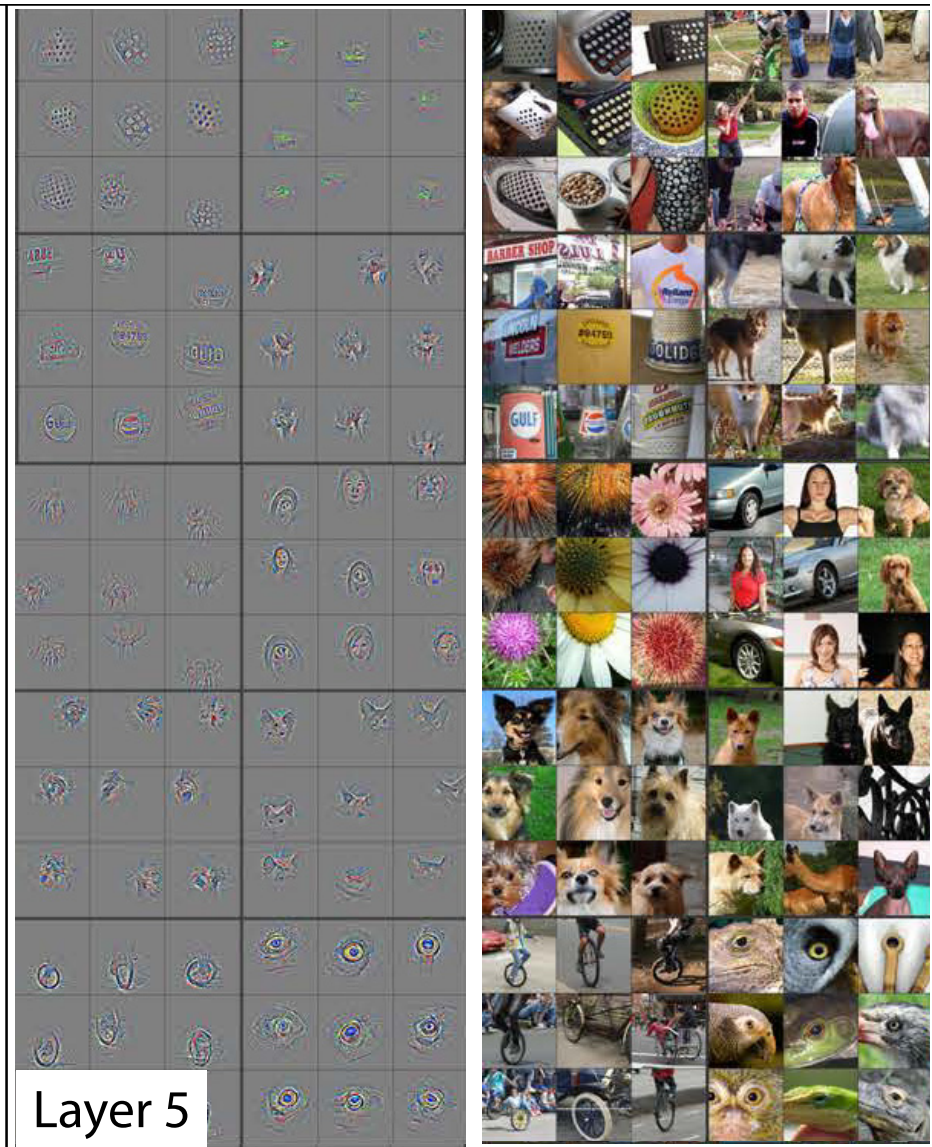
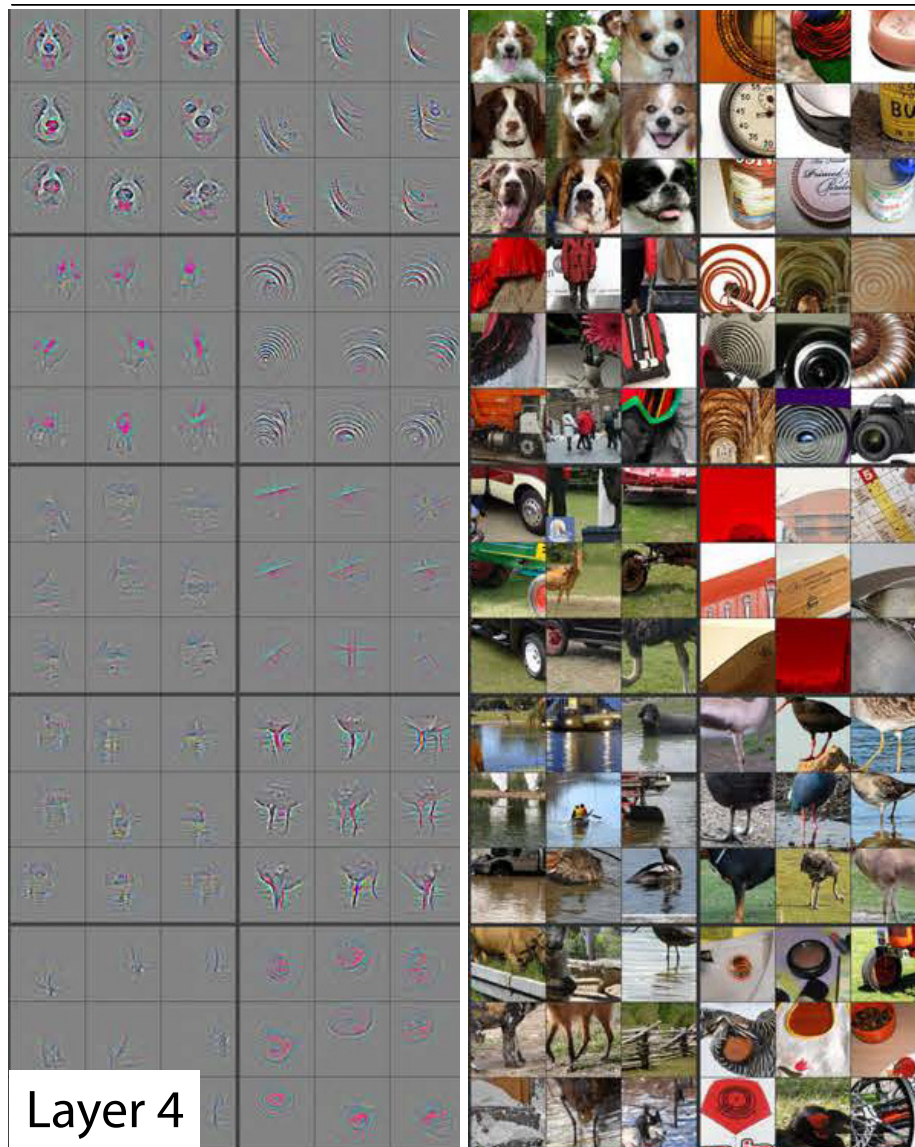
[Zeiler and Fergus,
ECCV 2014]

Visualisation



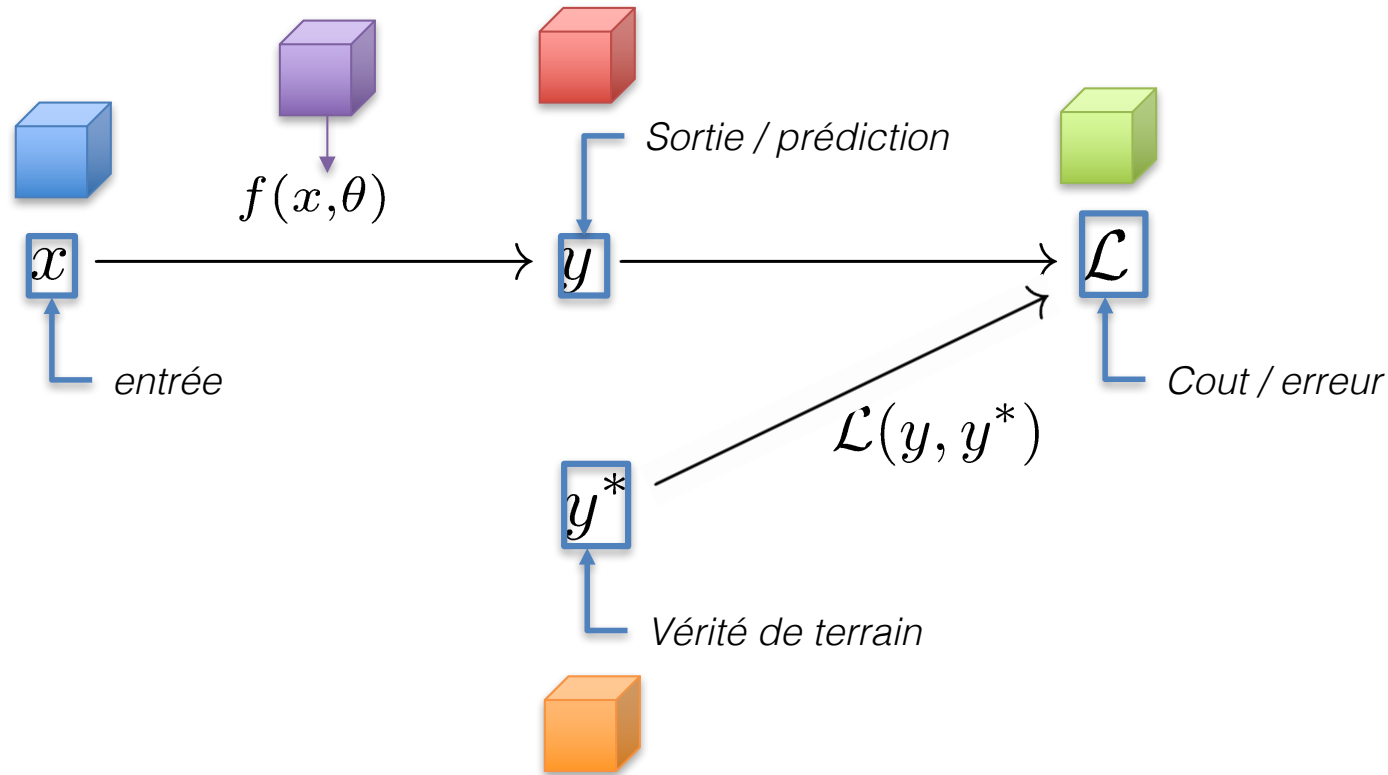
[Zeiler and Fergus,
ECCV 2014]

Visualisation



Comment coder tout ça?

Entrées, sorties et paramètres sont des tenseurs de plusieurs dimensions.



PyTorch (Facebook) : exemple

Calcul impératif et non pas symbolique (!). Débogage (relativement) facile.

```
1 class Net(nn.Module):
2
3     def __init__(self):
4         super(Net, self).__init__()
5
6         # Set up my layers
7         self.fc1 = nn.Linear(7*7*64, 1024)
8         self.fc2 = nn.Linear(1024, 10)
9
10        # Set up loss function and optimizer
11        crossentropy = nn.CrossEntropyLoss()
12        optimizer = optim.SGD(net.parameters(), lr=0.001, mo
13
14    def train(self, xinput, ylabel):
15
16        x = self.fc1(xinput)
17        x = F.relu(x)
18        youtput = self.fc2(x)
19
20        loss = crossentropy(youtput, ylabel)
21
22        print("Debug:", xinput, youtput, x)
23
24        loss.backward()
25
26        optimizer.step()
```

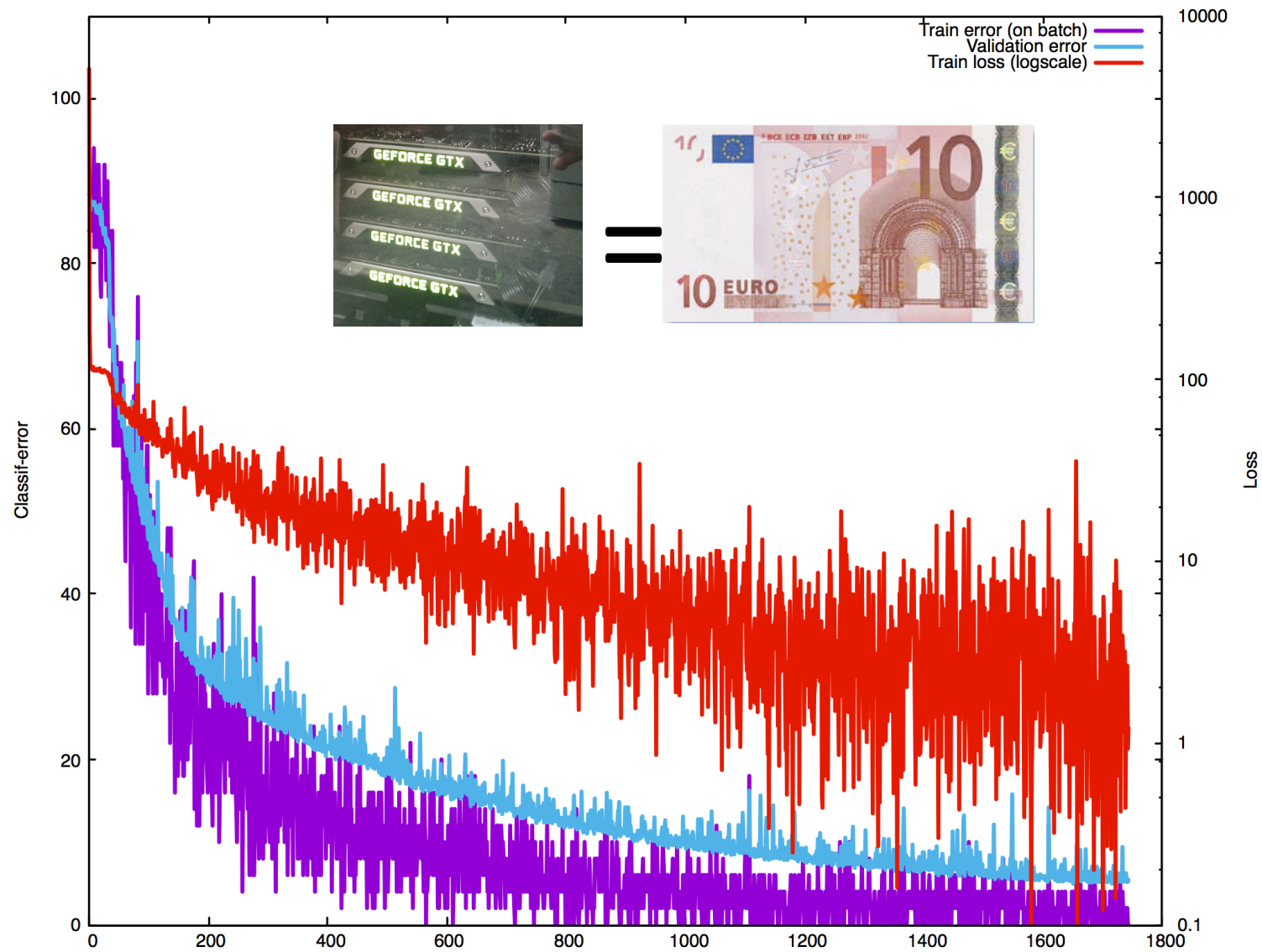
Code du pass « forward »
traversant le réseau : ici, 2
couches nommées fc1, fc2

Comparaison avec les
étiquettes, calcul de l'erreur

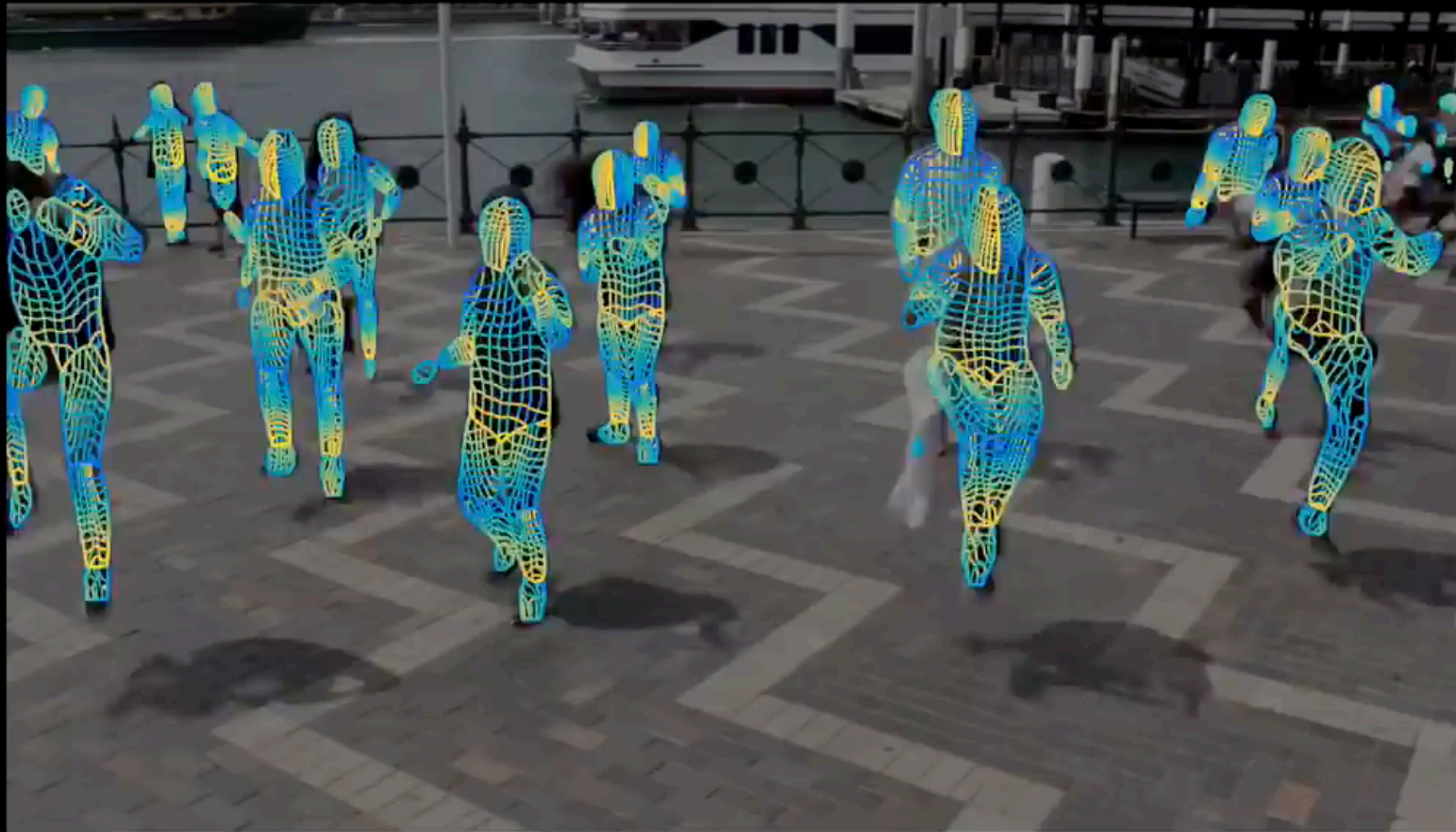
Calcul **automatique** du
gradient de l'erreur

Mise à jour **automatique**
du modèle
(=entraînement).

May the loss go down



Quelques applications



Video Source: <https://www.youtube.com/watch?v=2DiQUX11YaY>

[Güler, Neverova, Kokkinos, CVPR 2018]

Apprendre à expliquer les images



[Karpathy et al, 2015]

Générer des images

Sentence

A sheep by another sheep standing on the grass with sky above and a boat in the ocean by a tree behind the sheep



[Johnson, Gupta, Fei-Fei, CVPR 2018]



[Siarohin, Sangineto, Lathulière, Sebe, CVPR 2018]

Apprendre à contrôler



Edward
Beeching
M2R @ INRIA,
CITI,
INSA-Lyon



Christian Wolf
INRIA, LIRIS, CITI
INSA-Lyon

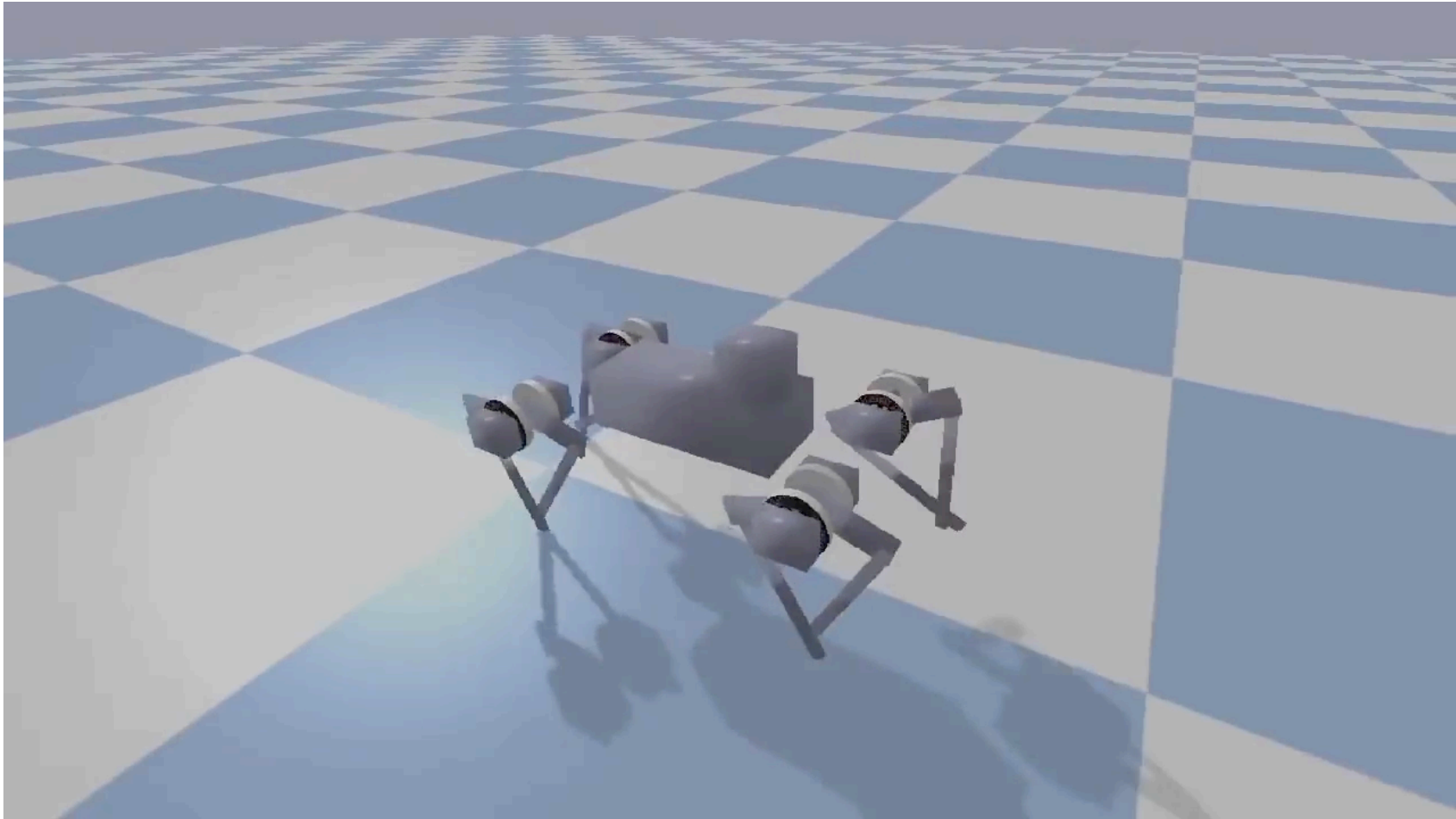


Jilles Dibangoye
INRIA, CITI,
INSA-Lyon



Olivier Simonin
INRIA, CITI,
INSA-Lyon

Apprendre à contrôler



[Tan, Zhang, Coumans, Iscen, Bai, Hafner, Bohez, Vanhouke, RSS 2018]

Scruter pour mieux
comprendre

Apprendre à traduire

Nous avons une quantité massive de textes (livres, articles) traduits par des traducteurs professionnels.

Peut-on apprendre à traduire uniquement à partir de données, sans aucune notion de grammaire et de langue?

Все счастливые семьи похожи друг на друга, каждая несчастливая семья несчастлива по-своему..

Toutes les familles heureuses se ressemblent. Chaque famille malheureuse, au contraire, l'est à sa façon.

Happy families are all alike. Every unhappy family is unhappy in its own way.

Alle glücklichen Familien gleichen einander, jede unglückliche Familie ist auf ihre eigene Weise unglücklich

[L. Tolstoy, 1873]

Traduction par un humain



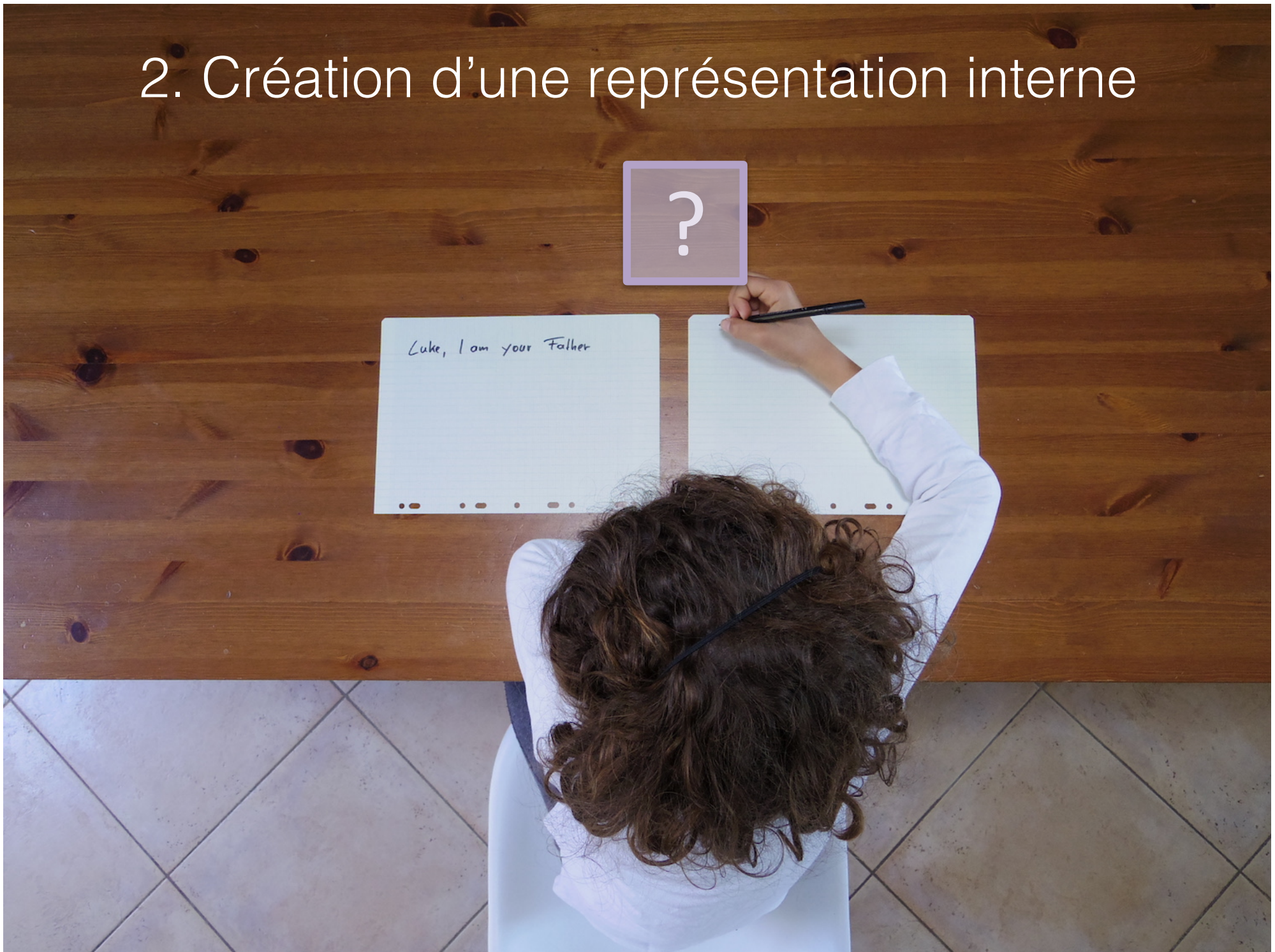
1. Lecture



2. Création d'une représentation interne

?

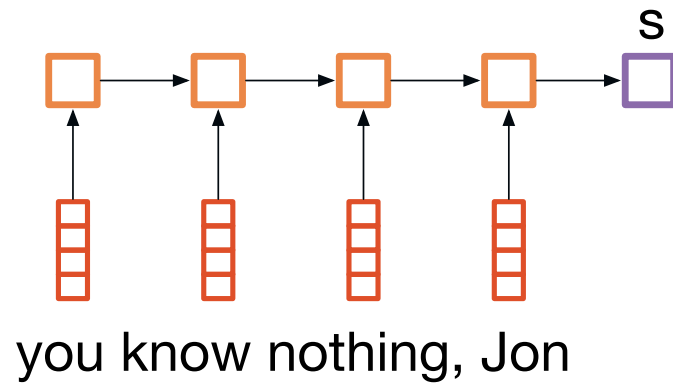
Luke, I am your Father



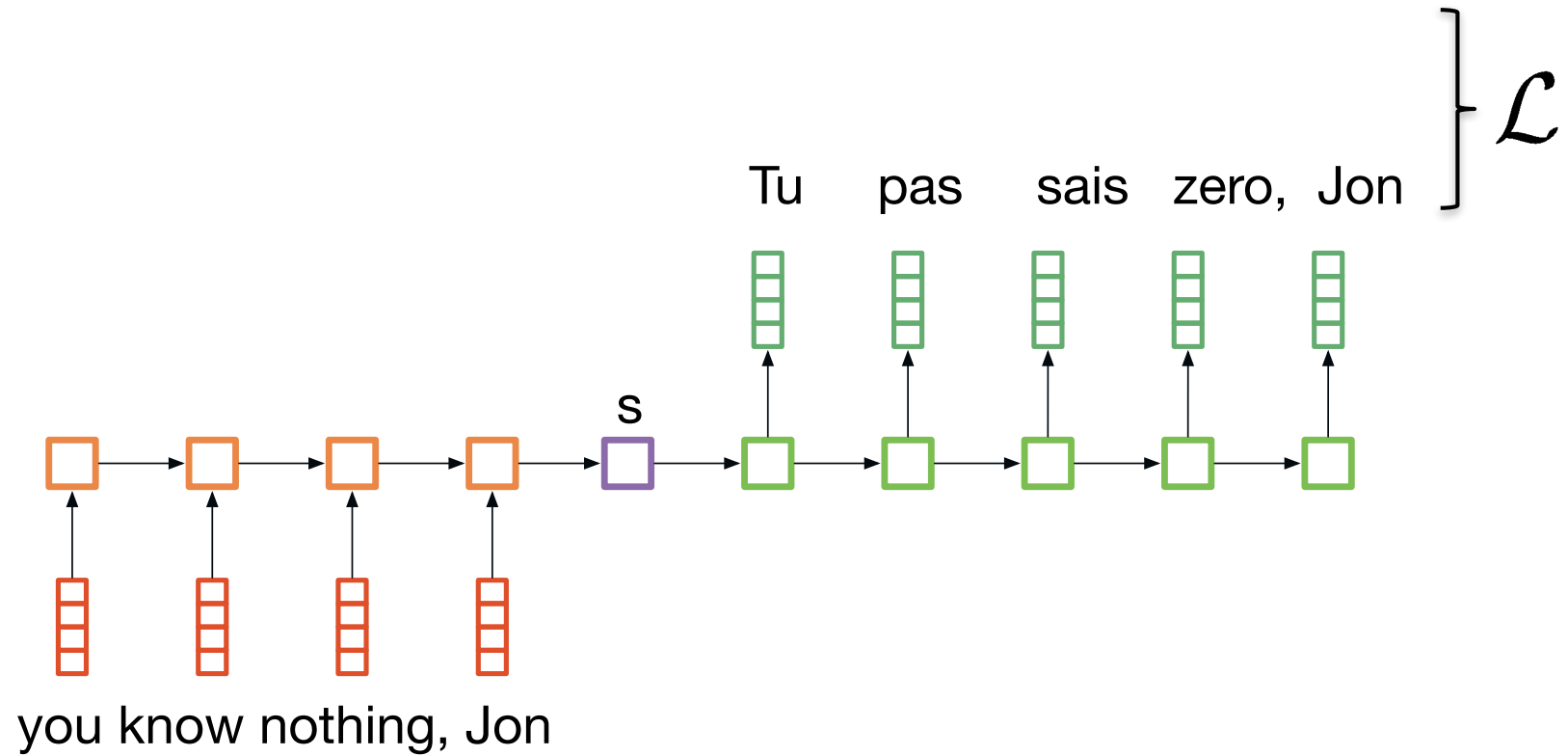
3. Ecriture de la traduction



Apprendre à traduire

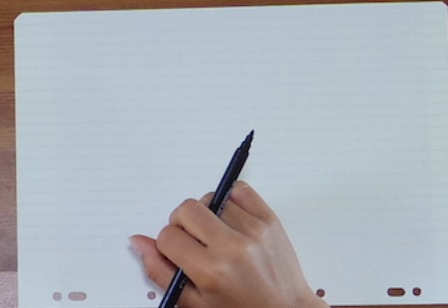


Apprendre à traduire



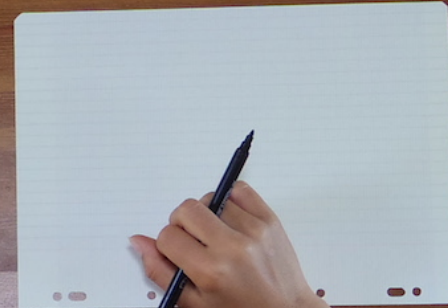
Traduction d'un long texte par un humain ?

If you really want to hear about it,
the first thing you probably want to
know is where I was born, and what
my lousy childhood was like, and how
my parents were occupied and all before
they had me, and all that David Copperfield



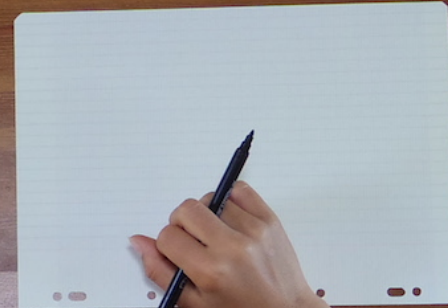
Lecture du texte entier

If you really want to hear about it,
the first thing you probably want to
know is where I was born, and what
my lousy childhood was like, and how
my parents were occupied and all before
they had me, and all that David Copperfield



Lecture attentive de la première partie

If you really want to hear about it,
the first thing you probably want to
know is when I was born, and what
my lousy childhood was like, and how
my parents were occupied and all before
they had me, and all that David Copperfield



Traduction et écriture de cette partie

If you really want to hear about it,
the first thing you probably want to
know is where I was born, and what
my lousy childhood was like, and how
my parents were occupied and all before
they had me, and all that David Copperfield

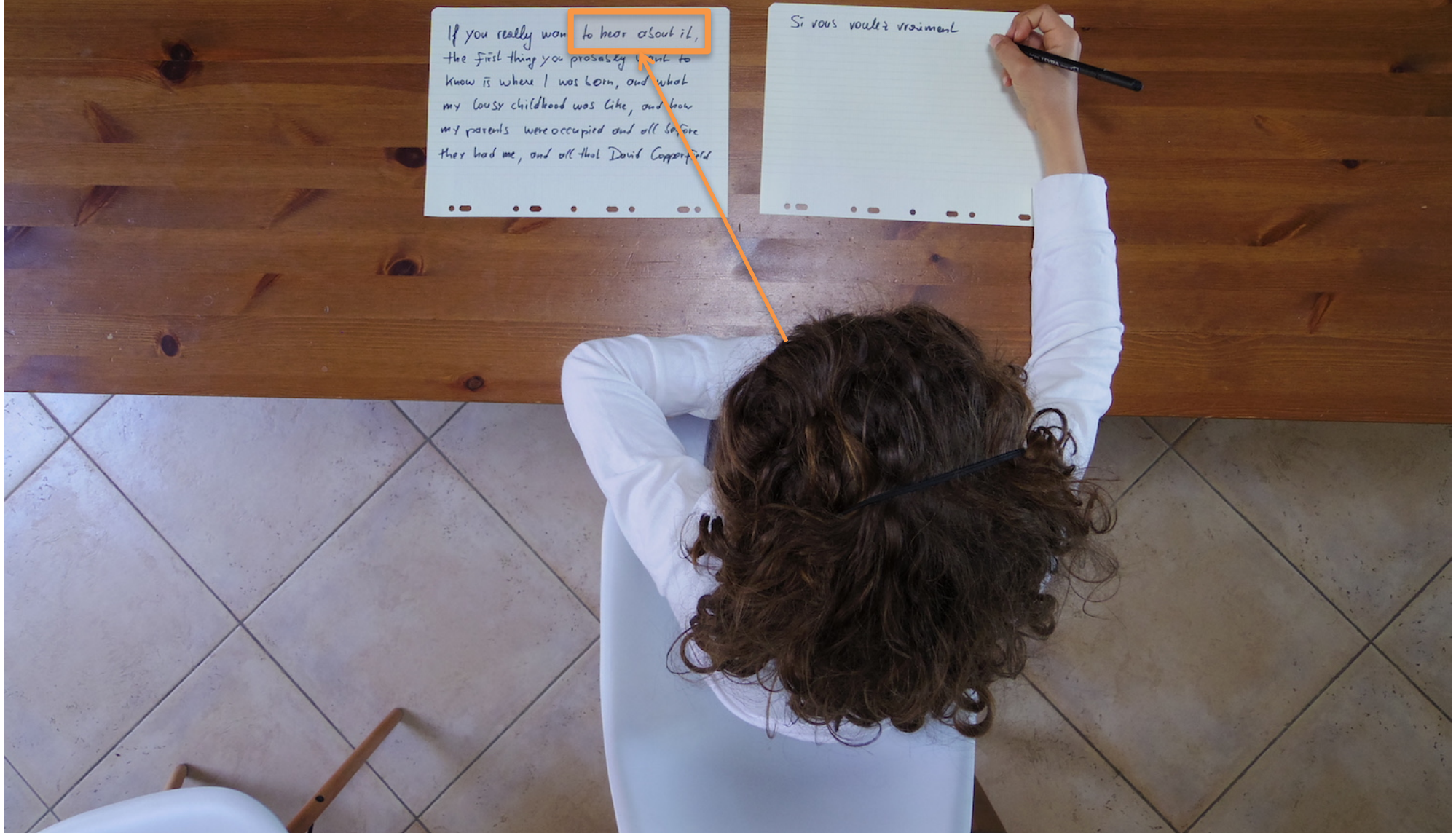
Si vous voulez vraiment



Lecture attentive de la prochaine partie

If you really want to hear about it,
the first thing you probably want to
know is where I was born, and what
my lousy childhood was like, and how
my parents were occupied and all before
they had me, and all that David Copperfield

Si vous voulez vraiment



Traduction et écriture de cette partie

If you really want to hear about it,
the first thing you probably want to
know is where I was born, and what
my lousy childhood was like, and how
my parents were occupied and all before
they had me, and all that David Copperfield

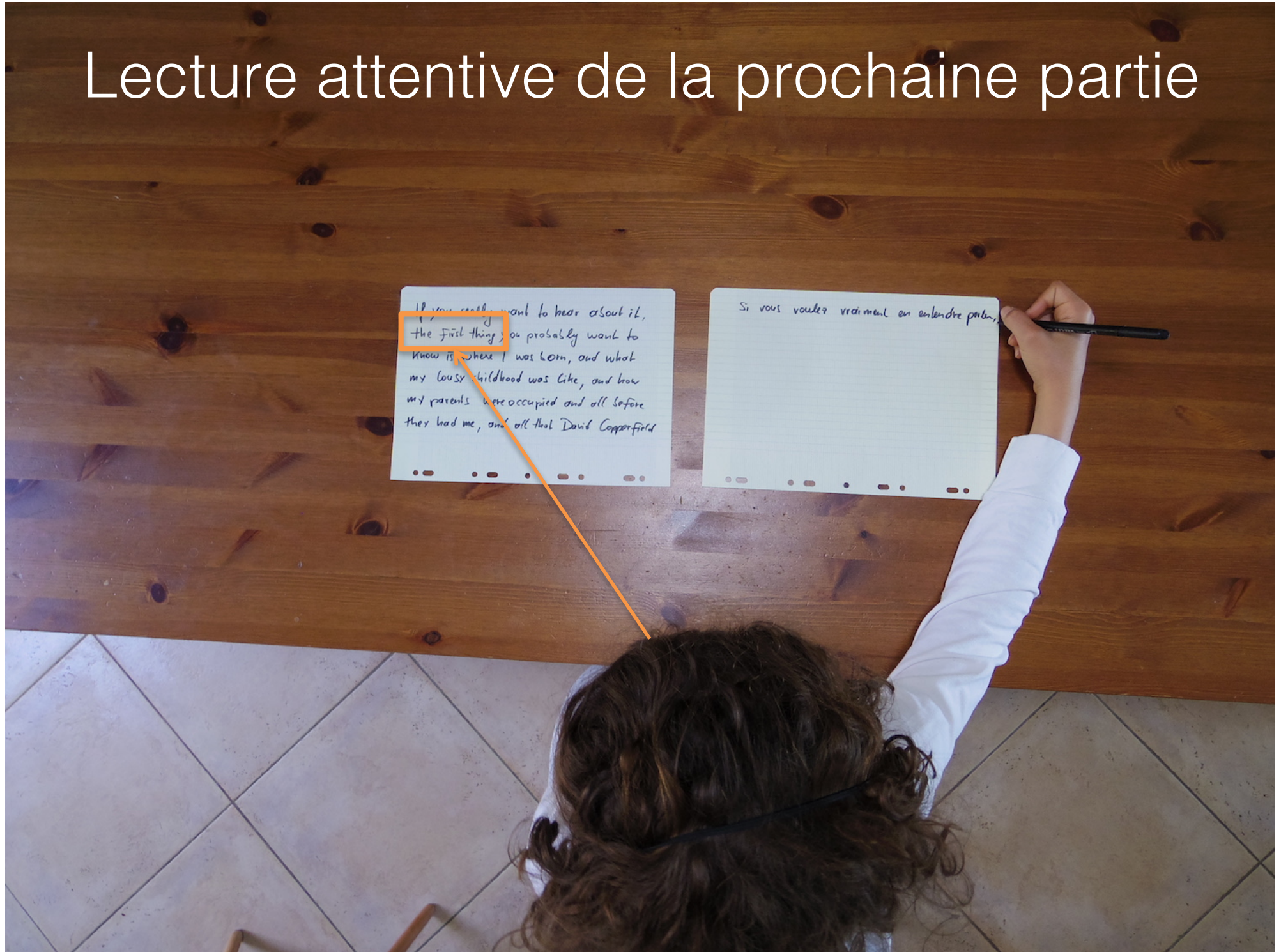
Si vous voulez vraiment en entendre plus



Lecture attentive de la prochaine partie

If you really want to hear about it,
the first thing you probably want to
know is where I was born, and what
my lousy childhood was like, and how
my parents were occupied and all before
they had me, and all that David Copperfield

Si vous voulez vraiment en entendre parler,



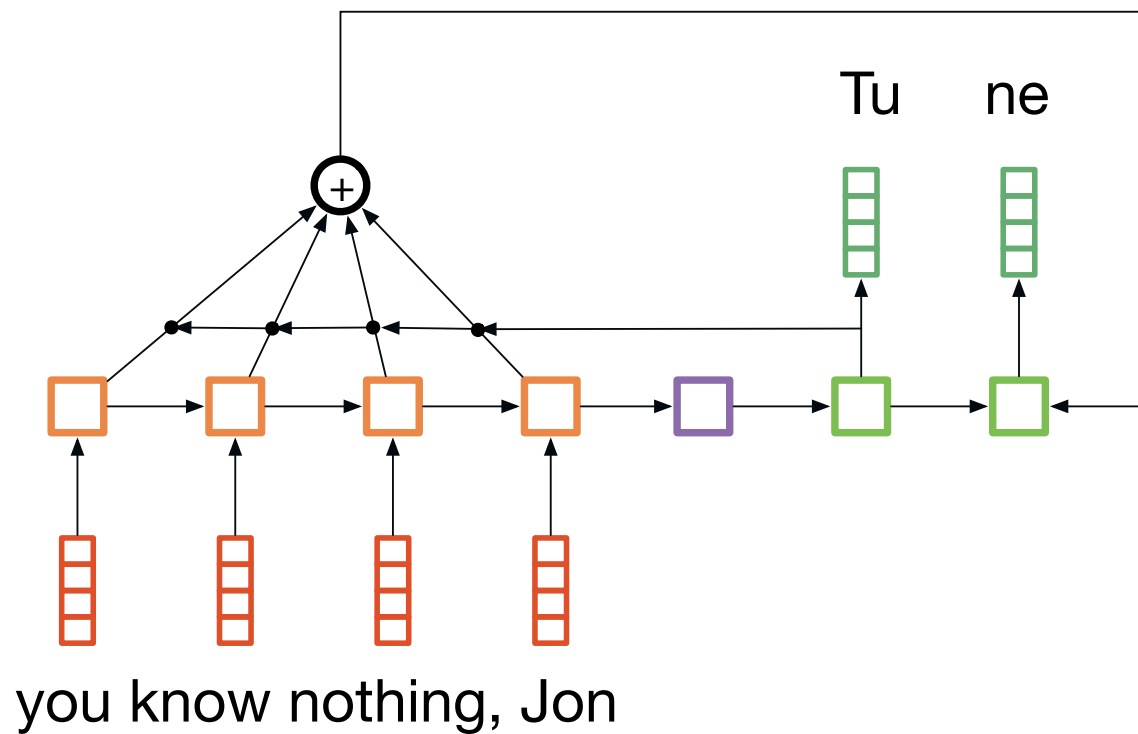
Traduction et écriture de cette partie

If you really want to hear about it,
the first thing you probably want to
know is where I was born, and what
my lousy childhood was like, and how
my parents were occupied and all before
they had me, and all that David Copperfield

Si vous voulez vraiment
la première chose



Apprendre à traduire



[Bahdanau, Cho, Bengio, 2016]

Traduction : exemples

VO

If you really want to hear about it, the first thing you'll probably want to know where I was born, and what my lousy childhood was like, and how my parents were occupied and all before they had me, and all that David Copperfield kind of crap, but I don't feel like going into it, if you want to know the truth. In the first place, that stuff bores me, and in the second place, my parents would have two hemorrhages apiece if I told anything pretty personal about them.

VF Google

Si vous voulez vraiment en entendre parler, la première chose que vous voudrez probablement savoir est où je suis né, et ce que mon enfance moche était, et comment mes parents étaient occupés et tout ce qu'ils avaient avant moi, et tout ce que David Copperfield, mais je n'ai pas envie d'y aller, si tu veux savoir la vérité. En premier lieu, cela m'ennuie, et en second lieu, mes parents auraient deux hémorragies si je leur racontais quelque chose de très personnel.

VF officielle

Si vous voulez vraiment que je vous dise, alors sûrement la première chose que vous aller demander c'est où je suis né, et à quoi ça ressemblait, ma saloperie d'enfance et c'est que faisaient mes parents avant de m'avoir, et toutes ces conneries à la David Copperfield, mais j'ai pas envie de raconter ça et tout. Primo, ce genre de trucs ça me rase, et secundo, mes parents ils auraient chacun une attaque ou même deux chacun, si je me mettais à baratiner sur leur compte quelque chose d'un peu personnel.

J.D. Salinger, « the catcher in the Rye »

Traduction: exemples

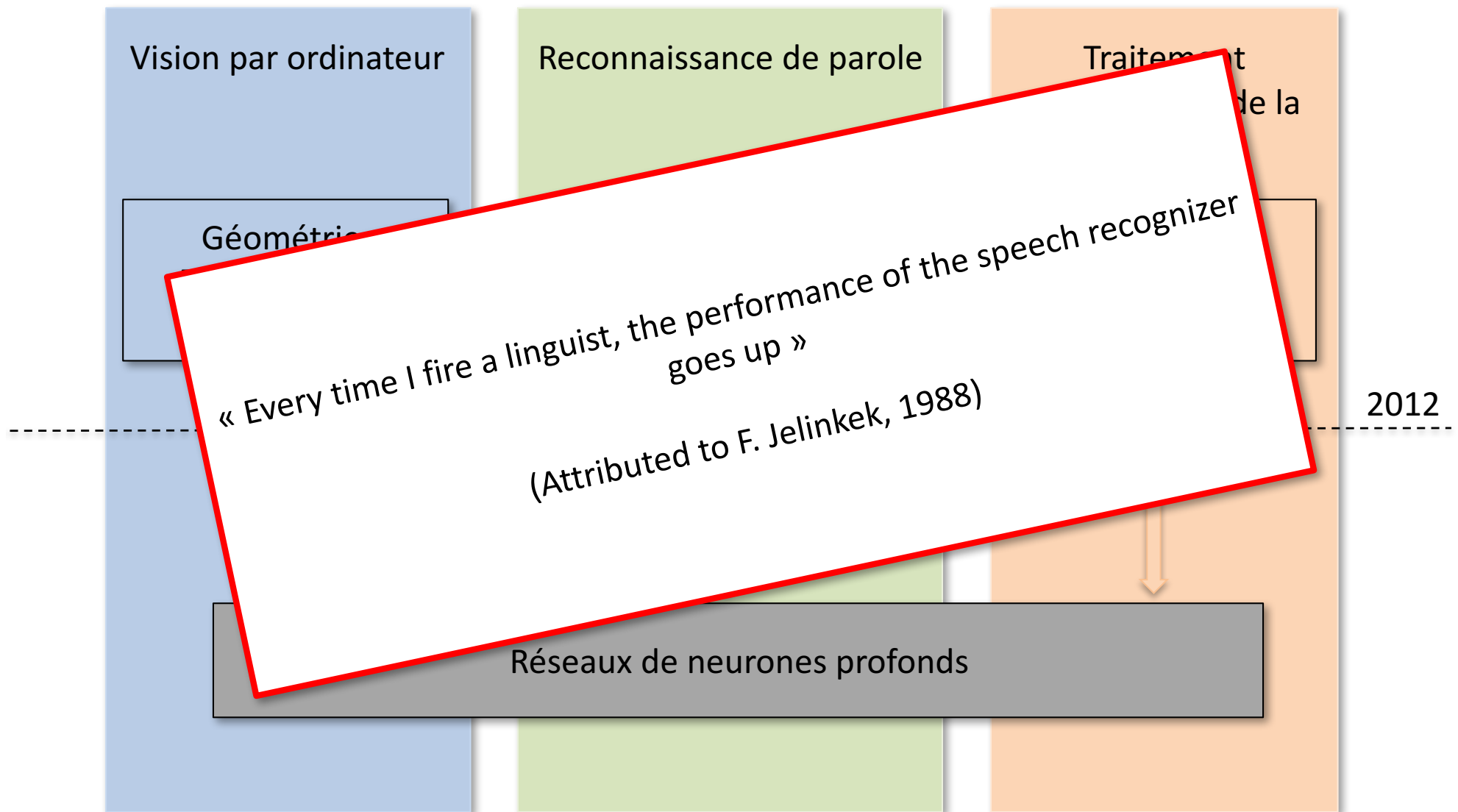
VO

Future robots will be trained, rather than programmed, for almost all of their tasks and sub-problems: perception, planning and navigation, behavior and control. However, obtaining gains through learning from massive amounts of data will not be as easy as in vision, speech, language processing, where machine learning was successful in the recent past. In contrast to these problems, which are often solved through supervised learning, teaching robots and agents to act autonomously requires learning from interactions with an environment

VF Google

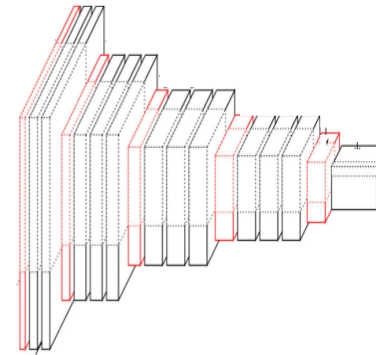
Les futurs robots seront **formés**, plutôt que programmés, pour presque toutes leurs tâches et sous-problèmes: perception, planification et navigation, comportement et contrôle. Cependant, obtenir des gains en apprenant des quantités massives de données ne sera pas aussi facile que dans la vision, la parole, le traitement du langage, où l'apprentissage automatique a été un succès récent. Contrairement à ces problèmes, qui sont souvent résolus par l'apprentissage supervisé, **enseigner** aux robots et aux agents à agir de manière autonome nécessite d'apprendre à partir d'interactions avec un environnement.

Convergence des modèles



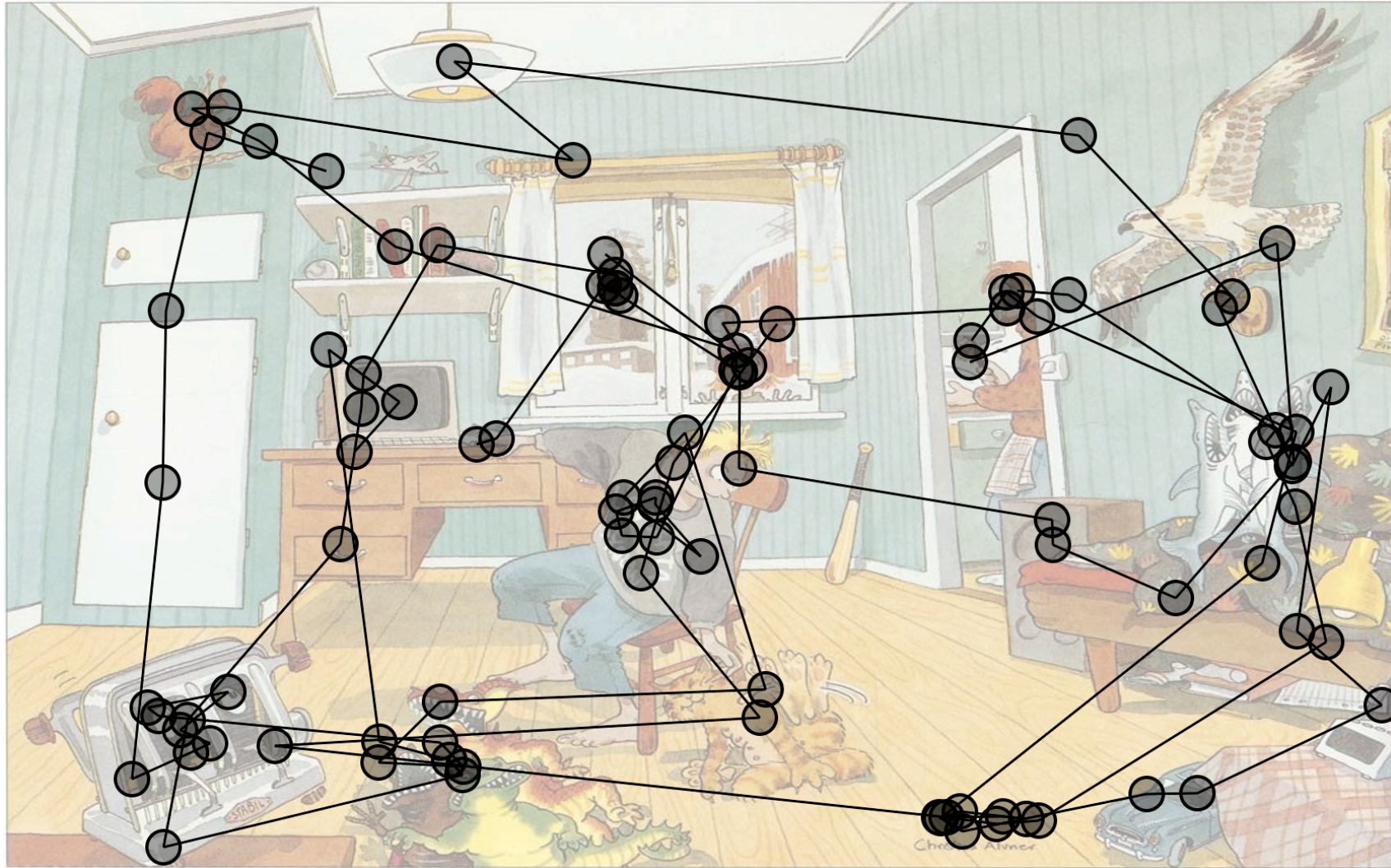
Modèles d'attention visuelle

Est-ce que Paul est dans la pièce?



Oui/No

Human attention: gaze patterns



[Johansson, Holsanova, Dewhurst, Holmqvist, 2012]

Modèles d'attention visuelle



Figure : Matthieu Cord, UPMC/LIP6

[Durand, Mordan, Thome, Cord, CVPR 2017]

Modèles d'attention visuelle

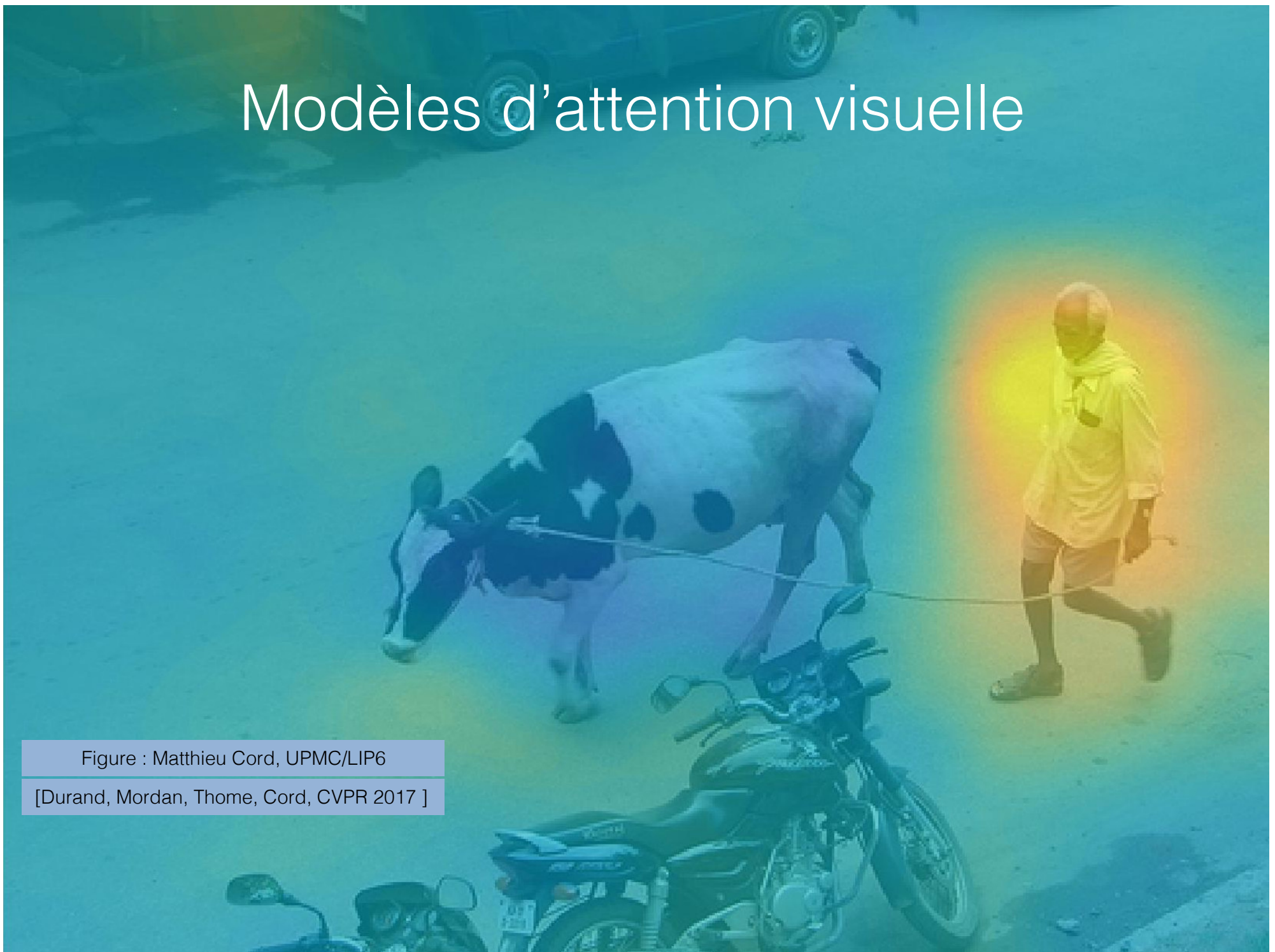


Figure : Matthieu Cord, UPMC/LIP6

[Durand, Mordan, Thome, Cord, CVPR 2017]

Modèles d'attention visuelle

Figure : Matthieu Cord, UPMC/LIP6

[Durand, Mordan, Thome, Cord, CVPR 2017]



Modèles d'attention visuelle

Figure : Matthieu Cord, UPMC/LIP6

[Durand, Mordan, Thome, Cord, CVPR 2017]



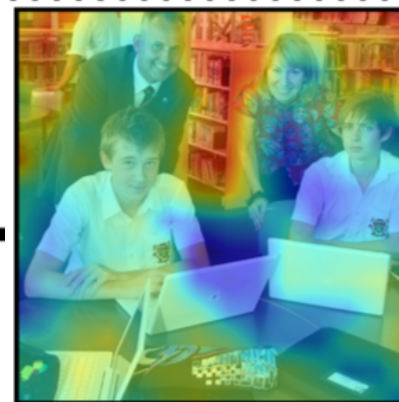
Scruter ... pour répondre à une question

What is sitting on the desk
in front of the boys?



Laptops

What are on the shelves
in the background?



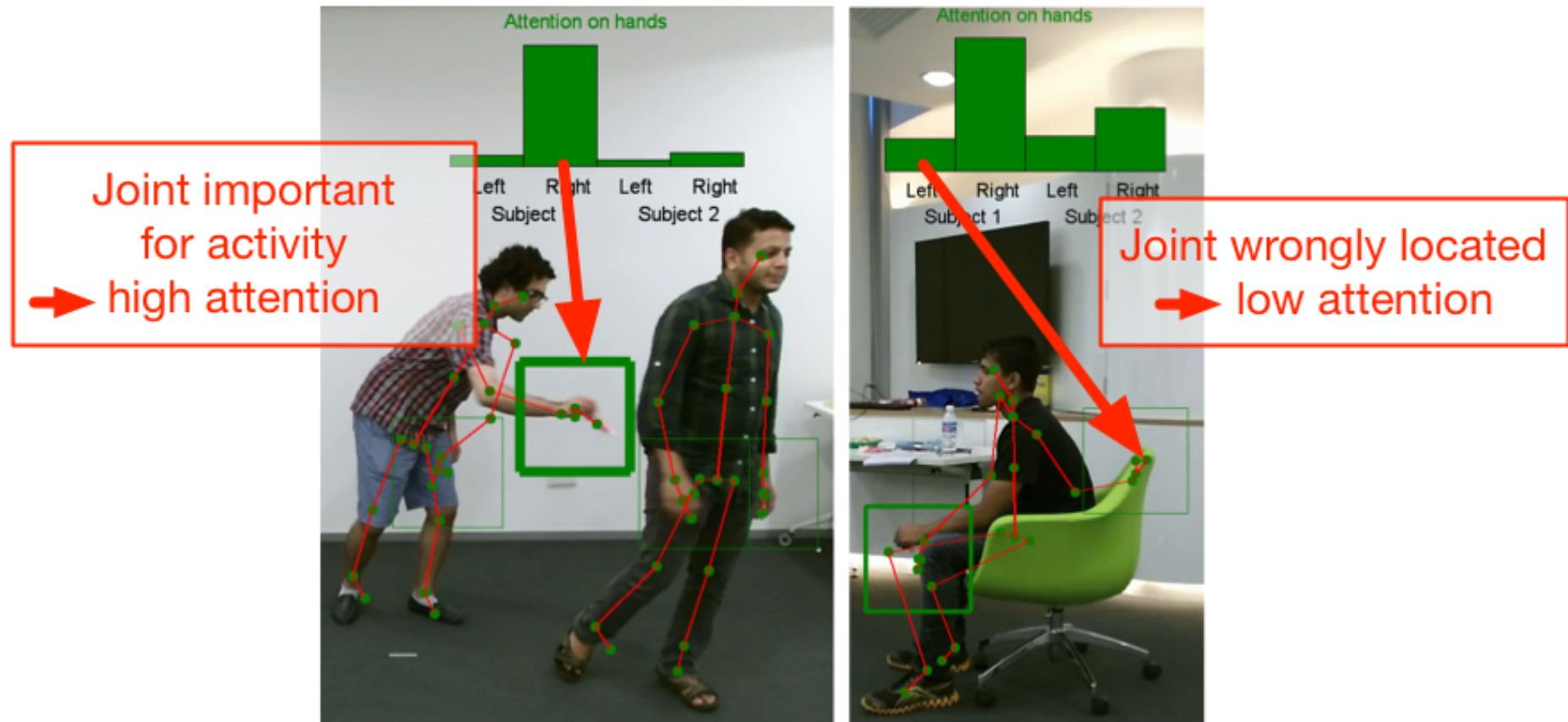
Books

[Ben-Younes, R. Cadene, N. Thome, M. Cord, ICCV 2017]



[Baradel, Wolf, Mille, ICCV-W-
Hands in Action, 2017]

Attention sur les parties du corps



Fabien Baradel,
Phd @ LIRIS,
INSA-Lyon

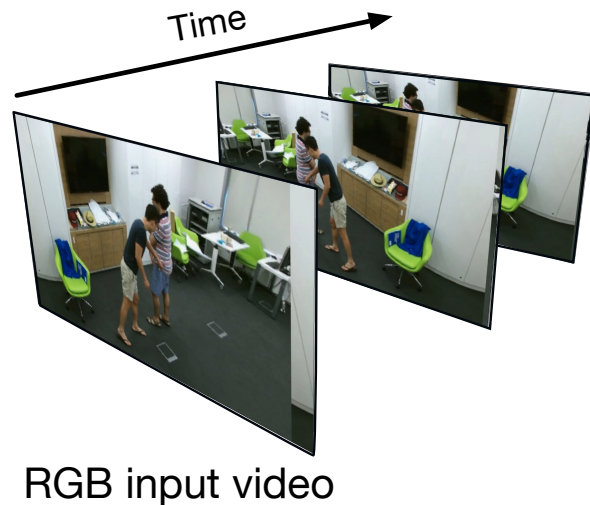


Christian Wolf
INRIA, LIRIS, CITI
INSA-Lyon



Julien Mille
LI, INSA Val de
Loire

Attention libre et dynamique



1. Learn where to attend
2. Learn how to track attended glimpse points (assign glimpses to semantic entities)
3. Learn how to recognize activities from a collection of tracked semantic entities

[Baradel, Wolf, Mille, Taylor, CVPR 2018]



Fabien Baradel,
Phd @ LIRIS,
INSA-Lyon



Christian Wolf
INRIA, LIRIS, CITI
INSA-Lyon

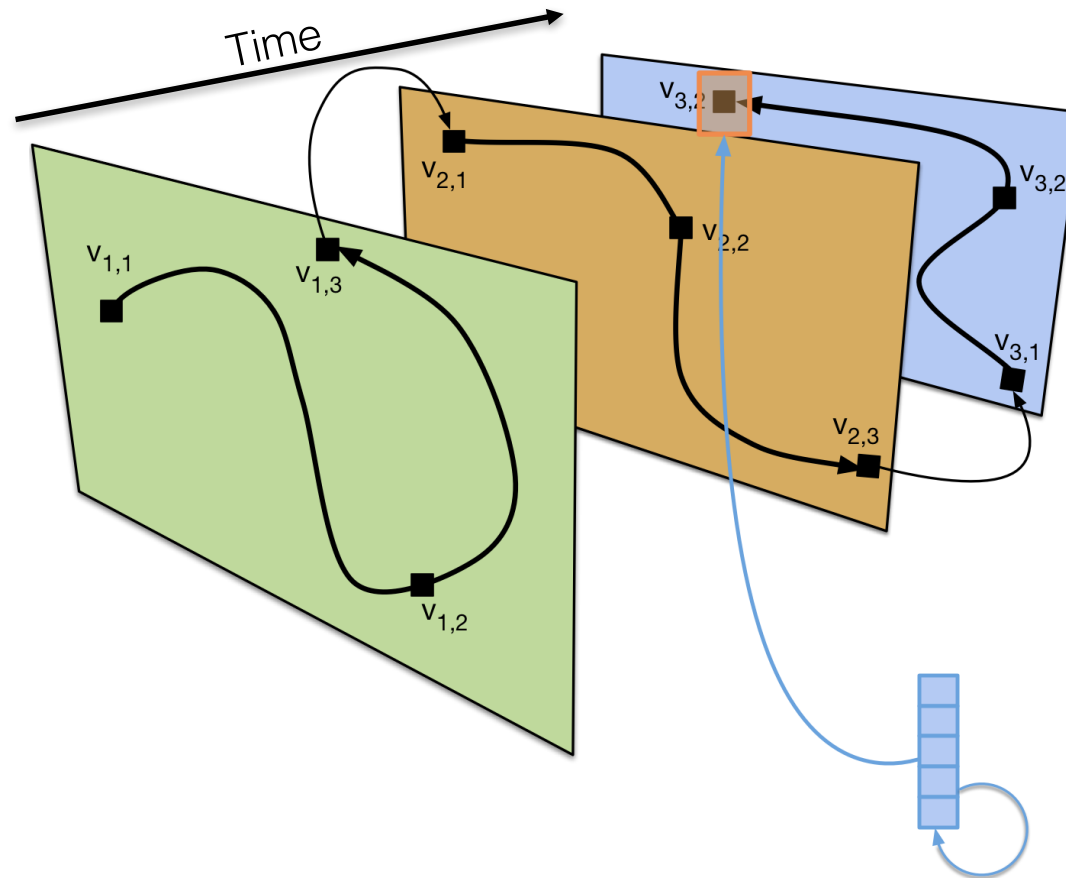


Julien Mille
LI, INSA Val de
Loire



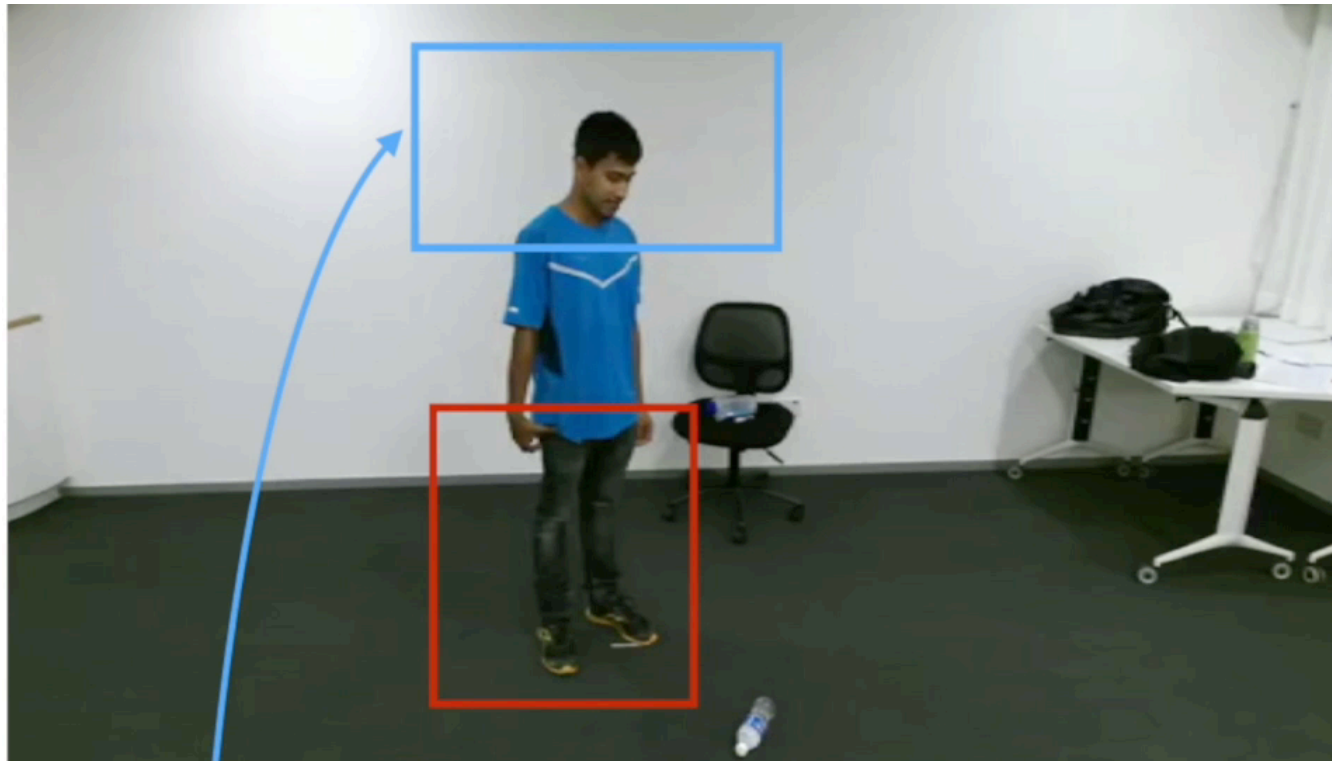
Graham W. Taylor
Univ. of Guelph,
Vector Institut

Attention libre et dynamique



[Baradel, **Wolf**, Mille, Taylor,
CVPR 2018]

Attention libre et dynamique



[Baradel, Wolf, Mille, Taylor, CVPR 2018]

Intelligence naturelle vs. artificielle

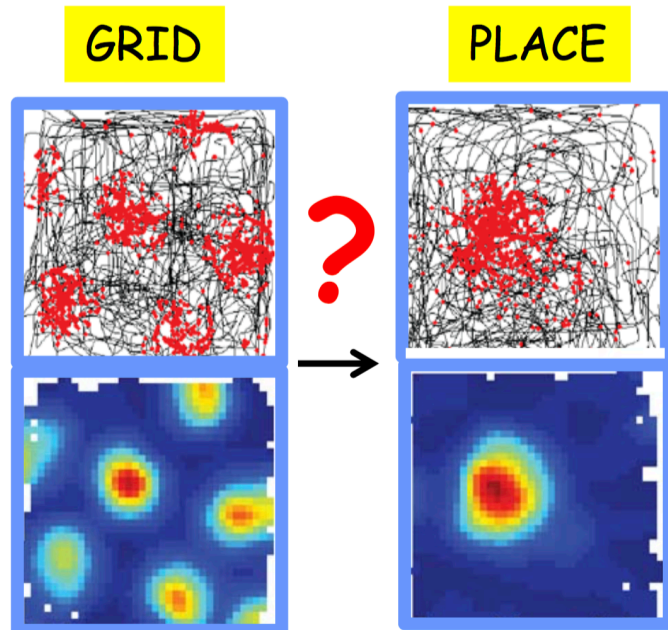


Photo: A. Mahmoud
May-Britt Moser
Prize share: 1/4

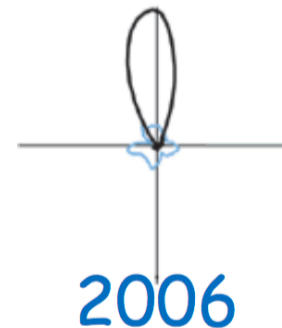


Photo: A. Mahmoud
Edvard I. Moser
Prize share: 1/4

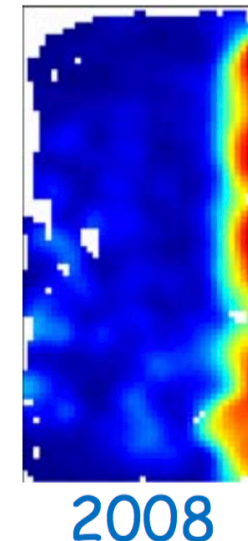
Prix Nobel 2014 de Médecine attribué pour des travaux découvrant des cellules de positionnement dans le cerveau.



Direction de la tête



Bordures



Intelligence naturelle vs. artificielle

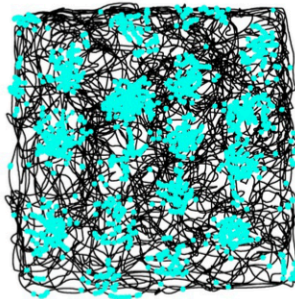


Photo: A. Mahmoud
May-Britt Moser
Prize share: 1/4

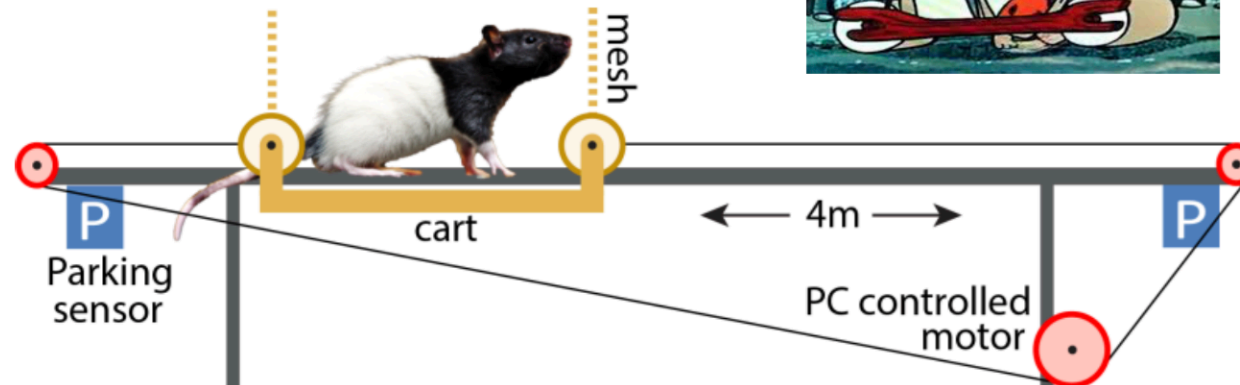


Photo: A. Mahmoud
Edvard I. Moser
Prize share: 1/4

Prix Nobel 2014 de Médecine attribué pour des travaux découvrant des cellules de positionnement dans le cerveau.

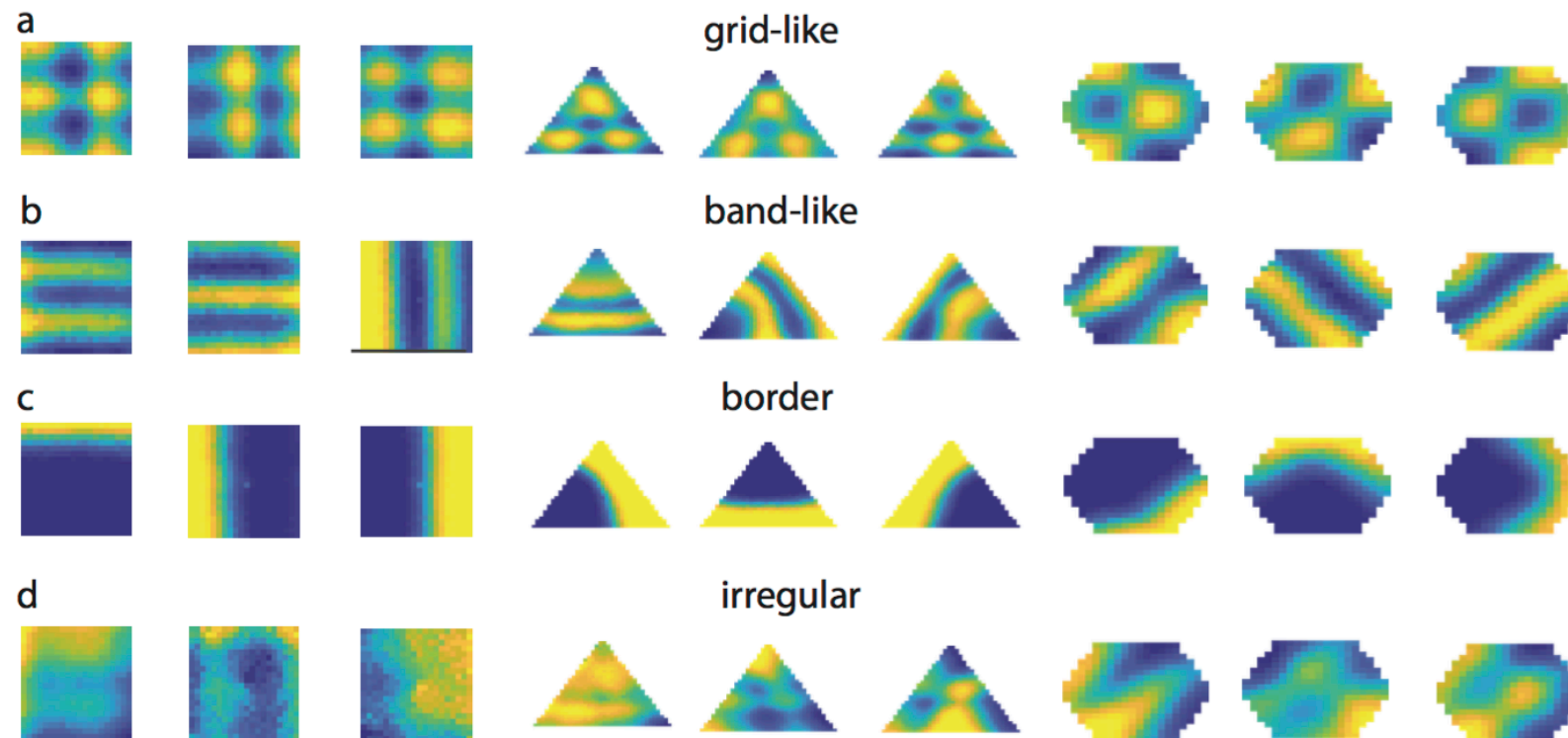


Speed cells are necessary for updating the grid pattern in accordance with the animal's movement
(distance=speed x time)



Intelligence naturelle vs. artificielle

2018 : découverte des mêmes cellules ... dans un réseau de neurones **artificiel** entraîné à se localiser dans un labyrinthe.



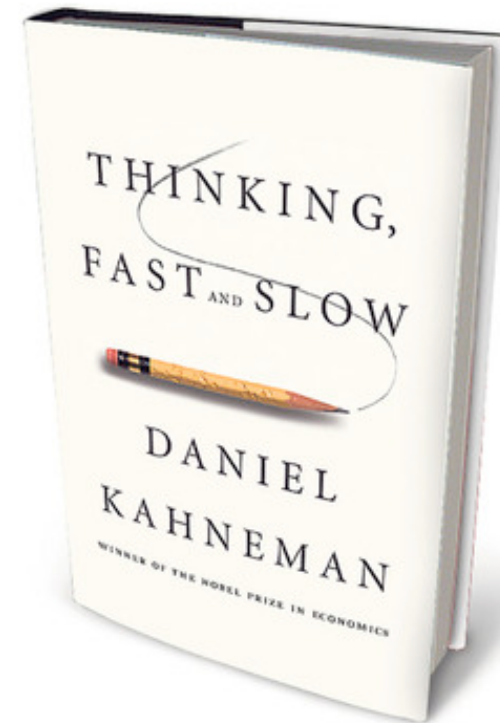
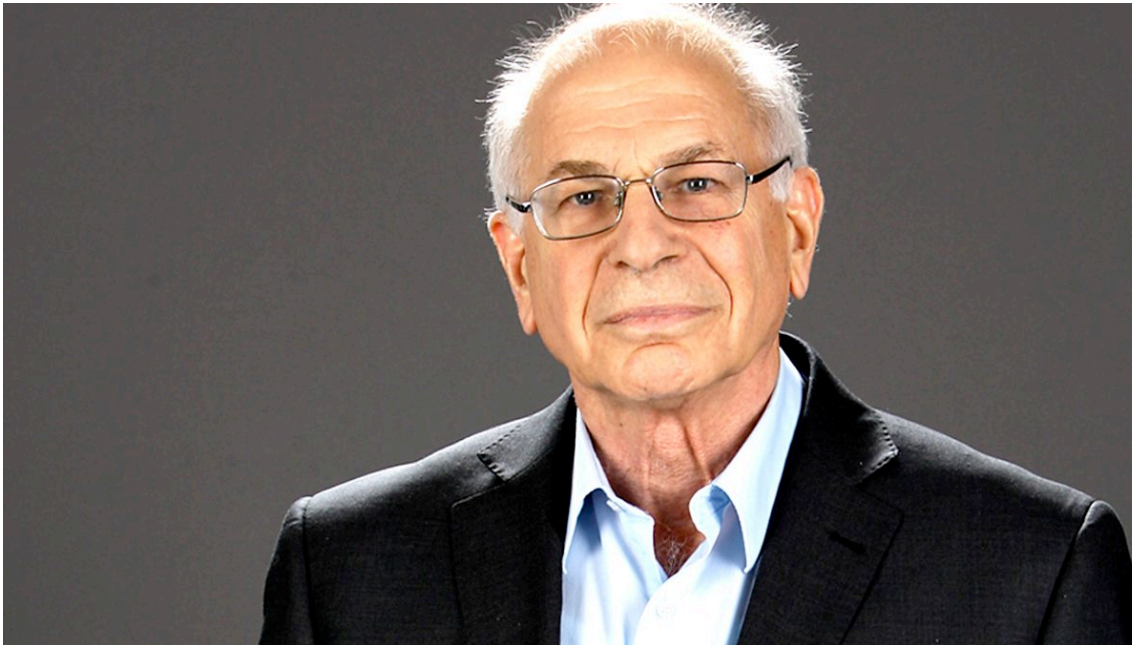
[Cueva, Wei, ICLR 2018]

Raisonnement: qu'est-ce qui s'est passé?



La psychologie humaine

- Daniel Kahnemann (Prix Nobel in 2002)
- Livre: « *Thinking Fast and Slow* »



Tâches cognitives



$$24 * 17 = ?$$

Deux systèmes

System 1

- Surveille en permanence l'environnement et l'esprit
- Pas d'attention spécifique
- Génère continuellement des évaluations / jugements sans efforts, même en présence de données insuffisantes. Saute aux conclusions
- Commet des erreurs

System 2

- Reçoit des questions ou les génère
- Dirige l'attention et recherche la mémoire pour trouver les réponses
- Nécessite (éventuellement beaucoup) d'effort
- Plus fiable

Où est l'apprentissage automatique aujourd'hui?

Mon évaluation personnelle :

- Nous sommes très focalisés sur les tâches liées au Système 1 :
 - Nos systèmes sautent toujours aux conclusions, quelle que soit la quantité d'information disponible
 - Mécanismes de type « feed-forward »
 - Le contexte joue un rôle important dans la recherche
- Nos « Système 1 » sont fortement spécialisés pour des tâches spécifiques (souvent la classification)
- Nous allons dans le sens du système 2 :
 - Mécanismes d'attention
 - Mémoires neuronales
 - Raisonnements de haut niveau (???)

Conclusion



- Le futur du codage est l'entraînement
- Le *Deep Learning* nécessite beaucoup de données (annotées) : **big data**!
- Gourmand de ressources (GPU, clusters, GPU, clusters de GPUs, FPGA, ASIC).
- Nécessite une expertise pour l'entraînement ("*babysitting*").
- Une forme de raisonnement va sans paraître. Laquelle? Symbolique? Dérivable?
- Convergence (en modèles et en applications) entre vision, robotique, TAL, parole ... automatique (?)
- Classification, régression, génération de données ... pour quand IA "forte"?