

Modèles Multiéchelles et Réseaux de Neurones Convolutifs



COLLÈGE
DE FRANCE
—1530—

Cours 1

- *Introduction aux réseaux de Neurones Convolutifs*

Stéphane Mallat

- High-dimensional $x = (x(1), \dots, x(d)) \in \mathbb{R}^d$:
- **Classification:** estimate a class label $f(x)$
given n sample values $\{x_i, y_i = f(x_i)\}_{i \leq n}$

Image Classification $d = 10^6$

Huge variability
inside classes

Anchor



Joshua Tree



Beaver



Lotus



Water Lily



Prediction: Supervised Learning

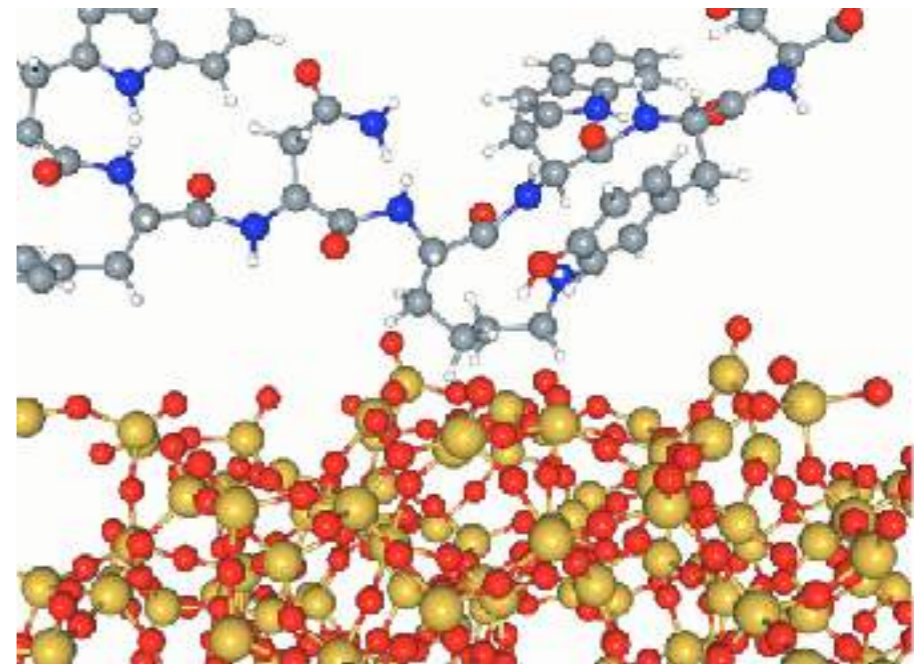
- High-dimensional $x = (x(1), \dots, x(d)) \in \mathbb{R}^d$:
- **Regression:** approximate a *functional* $f(x)$
given n sample values $\{x_i, y_i = f(x_i) \in \mathbb{R}\}_{i \leq n}$

Physics: energy $f(x)$ of a state vector x

Astronomy

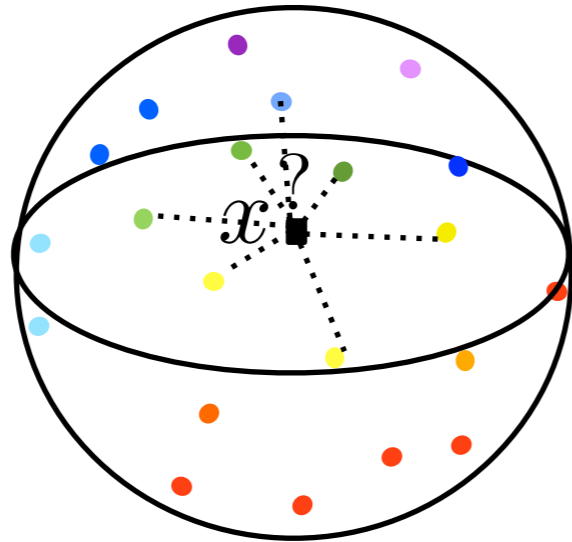


Quantum Chemistry



Curse of Dimensionality

- $f(x)$ can be approximated from examples $\{x_i, f(x_i)\}_i$ by local interpolation if f is regular and there are close examples:



- Need $n \geq \epsilon^{-d}$ points to cover $[0, 1]^d$ at a Euclidean distance ϵ
Problem: $\|x - x_i\|$ is always large
- To estimate $f(x)$ when x is in a high-dimensional Ω requires *strong regularity* of f in Ω : what regularity ?

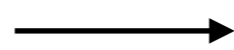


Linearize with a Representation

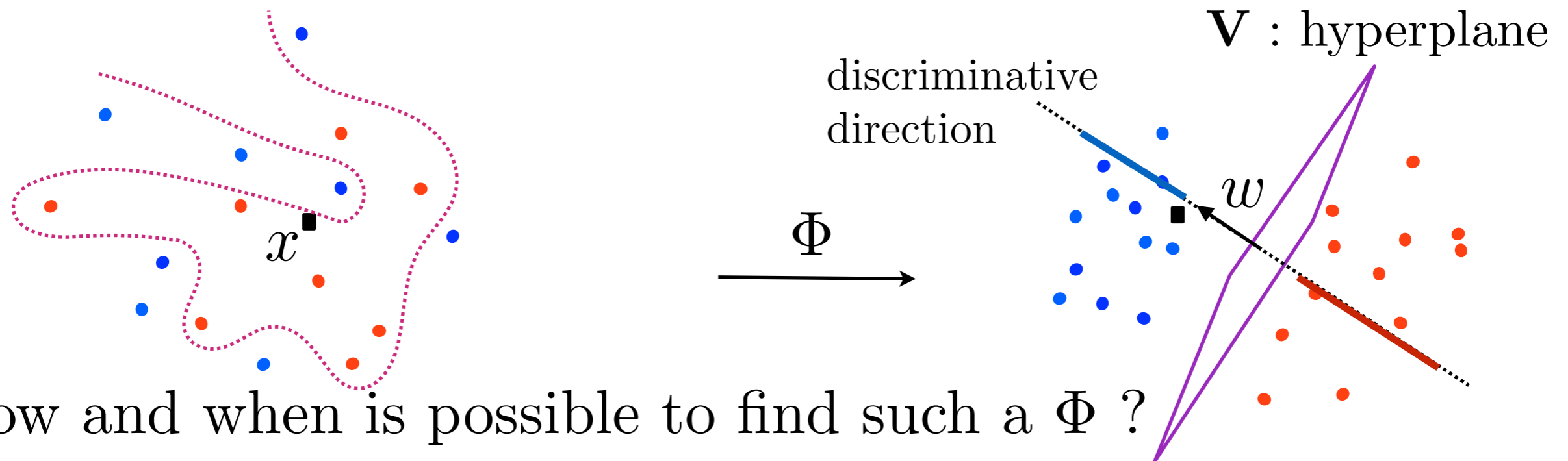
Change of variable $\Phi(x) = \{\phi_k(x)\}_{k \leq d'}$ to nearly linearize class boundaries, and approximate $f(x)$ by:

$$\tilde{f}(x) = \text{sign}(\langle w, \Phi(x) \rangle + b) = \text{sign}\left(\sum_k w_k v'_k + b\right)$$

$$x = (v_1, \dots, v_d)$$



$$\Phi(x) = (v'_1, \dots, v'_d)$$



- How and when is possible to find such a Φ ?
 - Only from a priori information on $f(x)$
 - Learned from data, with some a priori : neural networks



Challenges de Données

challengedata.ens.fr

- Données: images, séries temporelles, textes
- Calculer une prédiction \tilde{y} de y à partir de données x
- Exemples supervisés $\{x_i, y_i\}_{i \leq n}$
- Données de test $\{x_i^t\}_{i \leq n_t}$
- On doit calculer une prédiction \tilde{y}_i^t de chaque y_i^t inconnu
- Erreur évaluée par une fonction de risque $r(\tilde{y}, y)$

$$\text{Score} = n_t^{-1} \sum_{i=1}^{n_t} r(\tilde{y}_i^t, y_i^t)$$

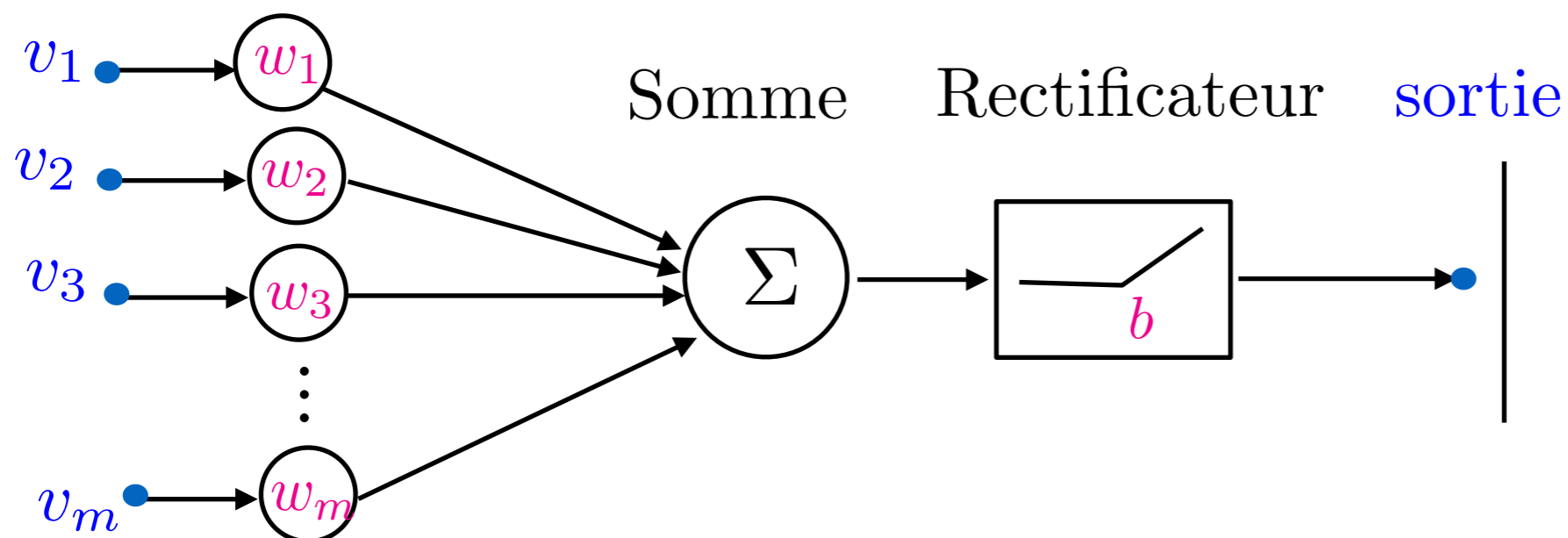
- Maximum 2 soumissions par jour
- Résultat sur données cachées: intermédiaire 19/3/20, final 15/12/20.
- Oral entre le 23 et 27 Mars pour étudiants de M2 du MVA et PSL.

- Confronter la « théorie » à l'algorithmique et aux résultats
- Le projet est l'occasion de se poser des questions
- Développer une représentation $\Phi(x)$ basée sur des a priori
- Eventuellement comparer avec des réseaux de neurones

Neurone Informatique

- Début dans les années 1950 : inspiration biologique
McCulloch & Pitts : modèle de neurone de grenouilles

entrées



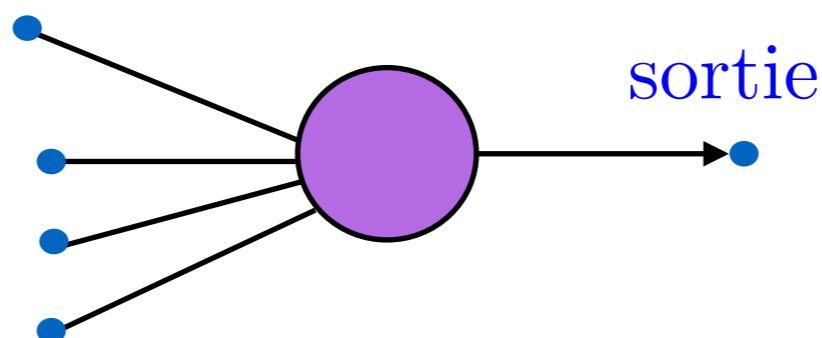
Parcimonie

$$0 \text{ si } \sum_{k=1}^m v_k w_k < b$$

$$\sum_{k=1}^m v_k w_k - b \text{ sinon}$$

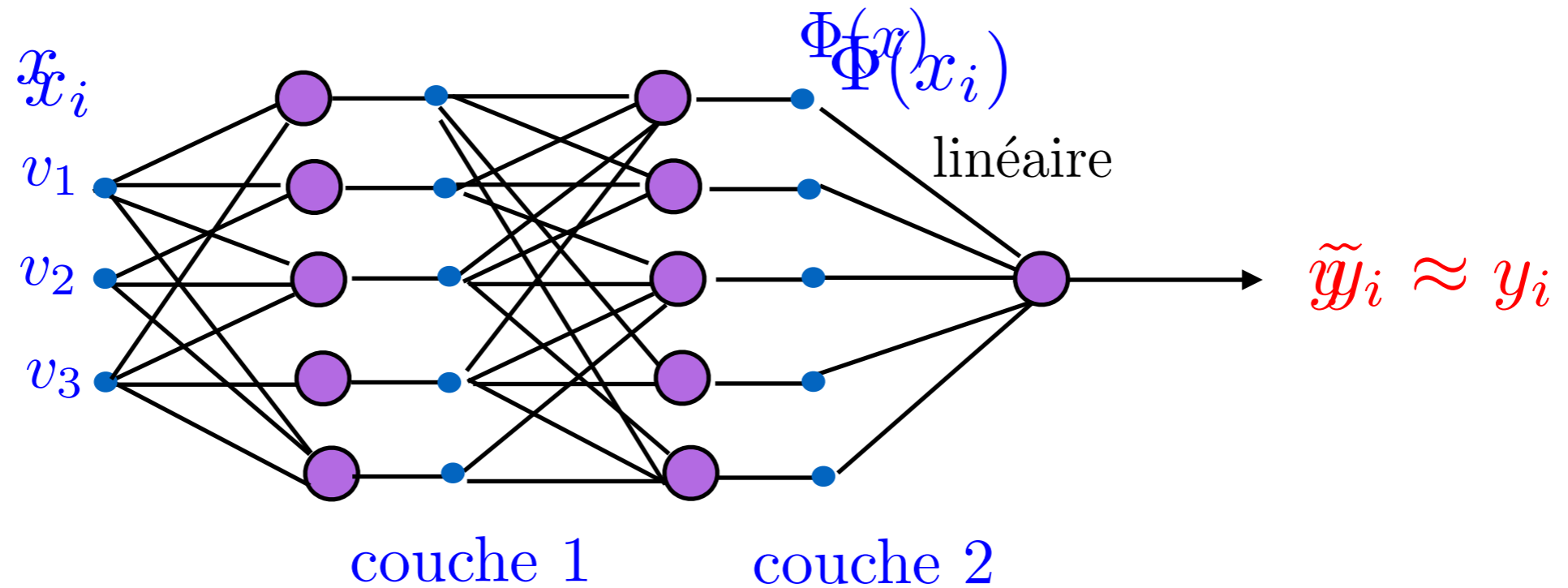
Un neurone est paramétré par les w_k et le seuil b

entrées





Réseaux de Neurones

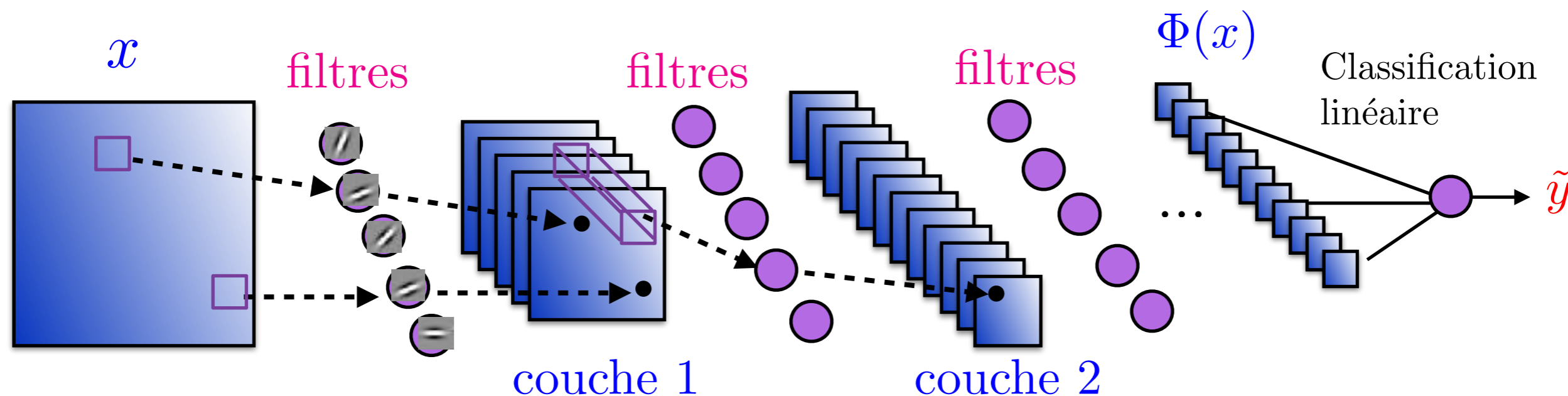


Apprentissage: optimise les **paramètres** pour minimiser les erreurs sur les exemples. Problèmes difficiles d'optimisation.

Information à priori: elle est dans l'architecture du réseau.

- Les paramètres sont invariants par translations: **filtres**

Y. LeCun



Centaines de millions de paramètres

Prédictions exceptionnelles et *génériques*: pour l'image, parole, diagnostiques, traduction, en physique...

Le $\Phi(x)$ a aplati les frontières. Comment ?

Importance de l'architecture: symétrie, multiéchelle

ImageNet Data Basis

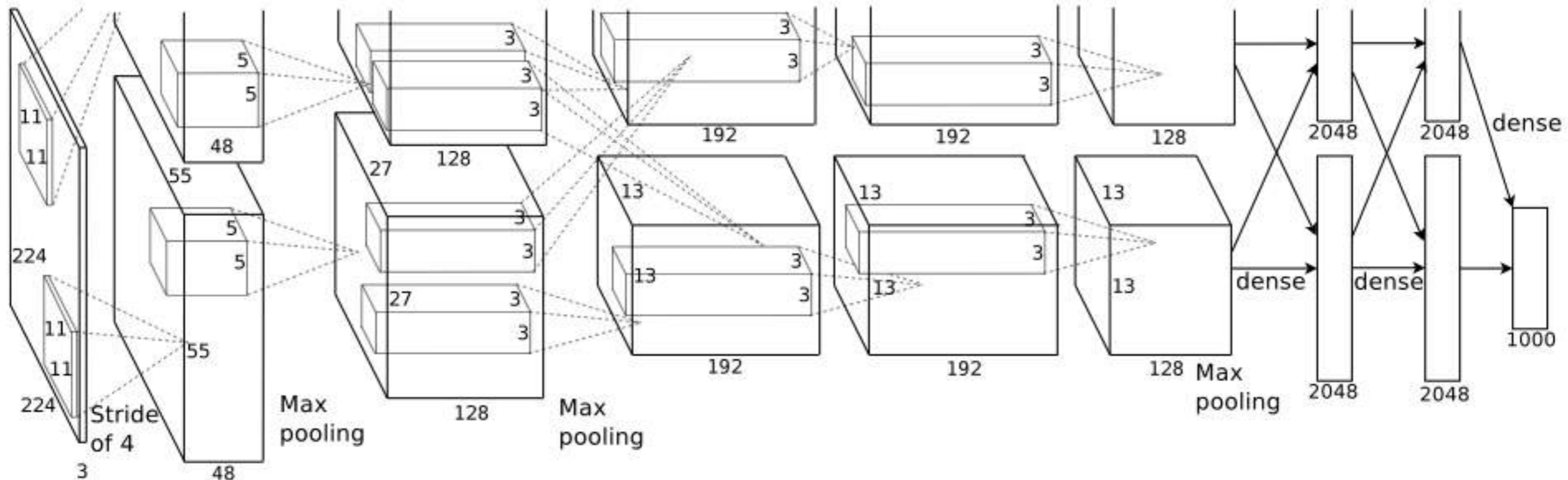
- Data basis with 1 million images and 2000 classes



Alex Deep Convolution Network

A. Krizhevsky, Sutskever, Hinton

- Imagenet supervised training: $1.2 \cdot 10^6$ examples, 10^3 classes
15.3% testing error in 2012



New networks with 3% errors.

Up to 150 layers!

Pourquoi cela marche ?



Image Classification



mite

container ship

motor scooter

leopard

	mite black widow cockroach tick starfish		container ship lifeboat amphibian fireboat drilling platform		motor scooter go-kart moped bumper car golfcart		leopard jaguar cheetah snow leopard Egyptian cat
--	---	--	---	--	--	--	---



grille

mushroom

cherry

Madagascar cat

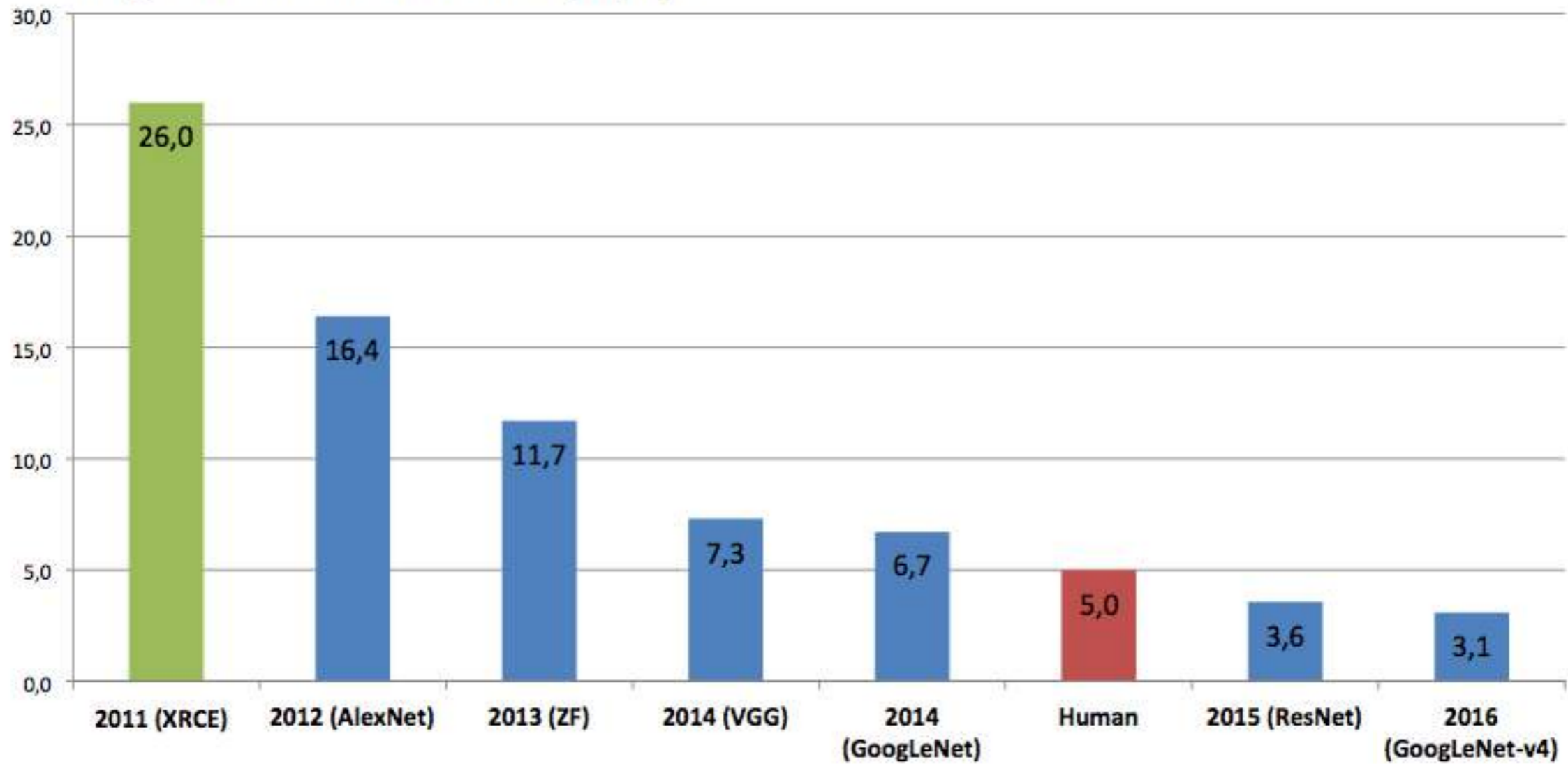
	grille convertible pickup beach wagon fire engine		mushroom agaric mushroom jelly fungus gill fungus dead-man's-fingers		cherry dalmatian grape elderberry ffordshire bullterrier currant		Madagascar cat squirrel monkey spider monkey titi indri howler monkey
--	--	--	--	--	--	--	---

Quels types de "connaissance" ont été apprises ?



Erreurs sur ImageNet

ImageNet Classification Error (Top 5)



Why Understanding ?

Szegedy, Zaremba, Sutskever, Bruna, Erhan, Goodfellow, Fergus

$$x + \epsilon = \tilde{x} \quad \text{with} \quad \|\epsilon\| < 10^{-2} \|x\|$$



correctly
classified

classified as
ostrich

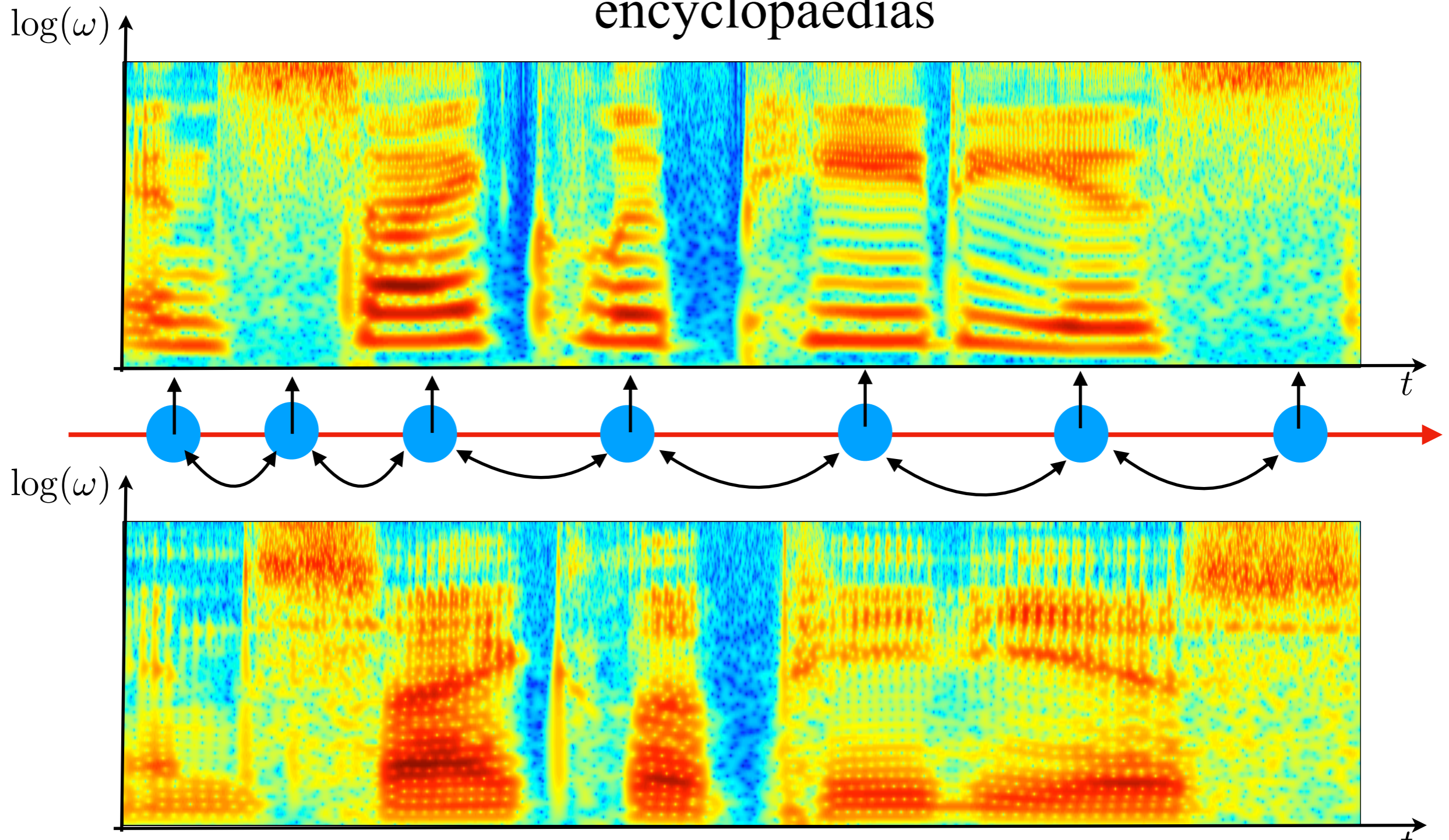
- Trial and error testing can not guarantee reliability.



Reconnaissance de la parole

- Depuis les années 60.
- Jusqu'en 2010, basé sur le calcul de spectrogrammes et modèles de chaînes de Markov.

encyclopaedias

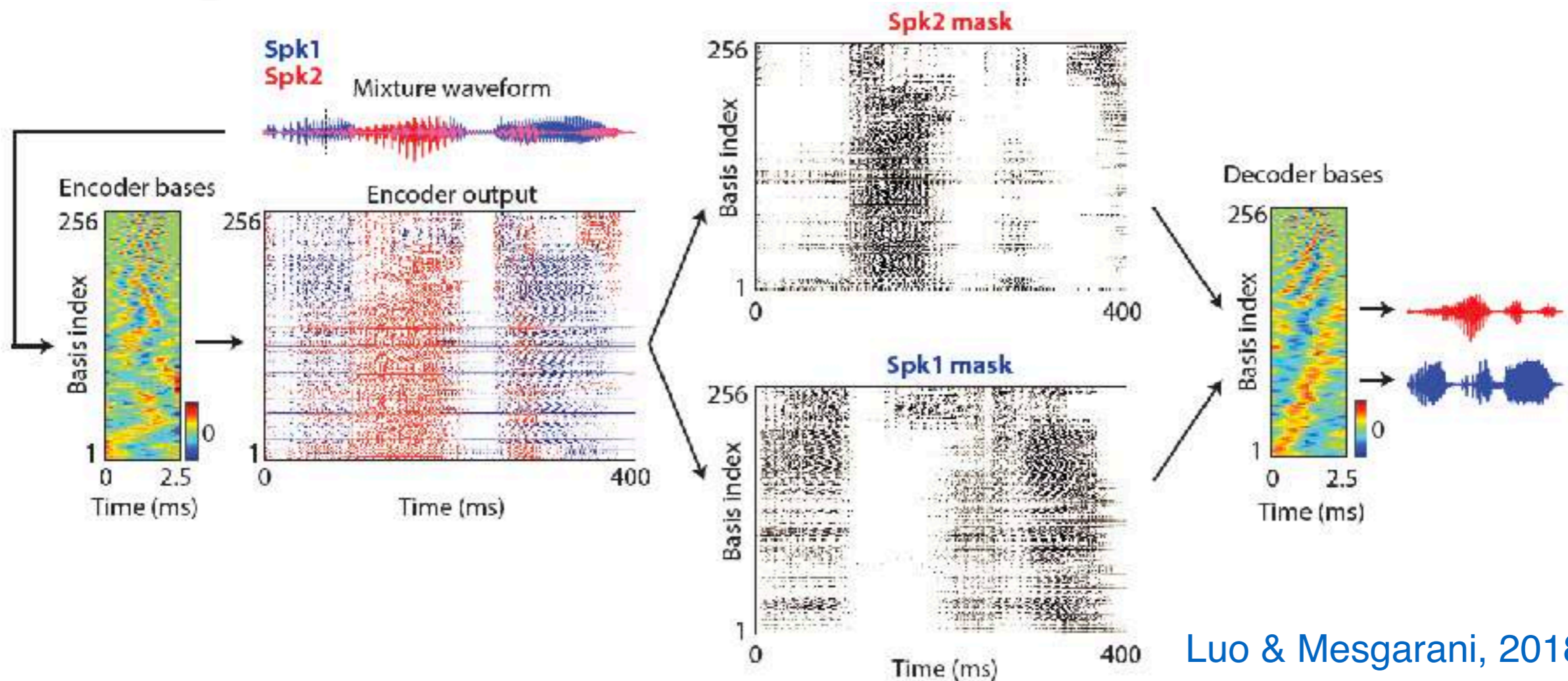


Reconnaissance Audio

- Reconnaissance de la parole, de la musique:



- Séparation de sources:





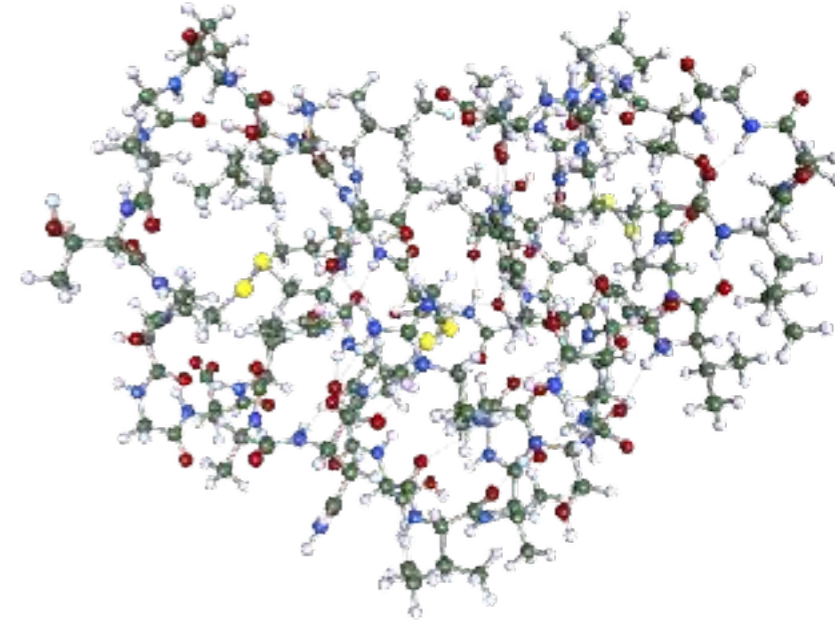
Astrophysique



Dynamique des fluides



Chimie Quantique

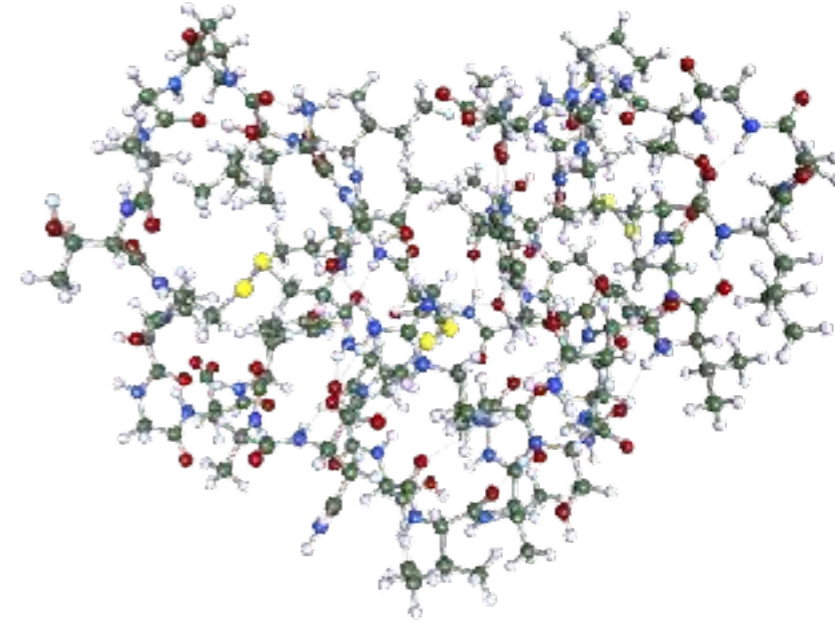


- Lois de la physique connues: équations de Newton, Boltzmann, Maxwell, Navier-Stokes, Schrodinger... mais calcul difficile des solutions résultant d'interactions d'un grand nombre de « corps élémentaires ».
- Simulations numériques: à partir des équations fondamentales.
- Peut on prédire une solution par régression à partir d'une base de données de solutions et quelques informations à priori ?
De mieux en mieux avec des réseaux de neurones profonds...

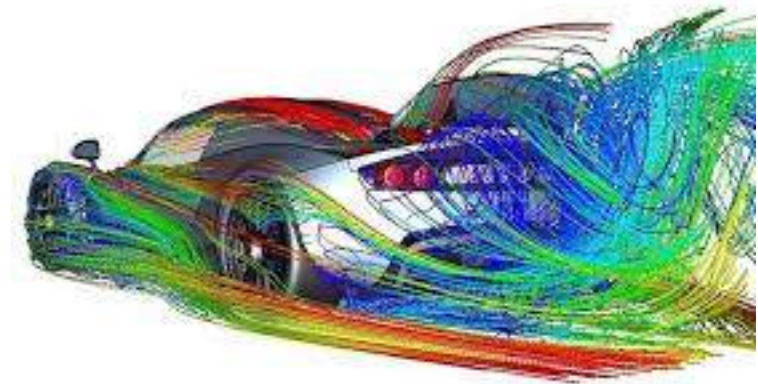
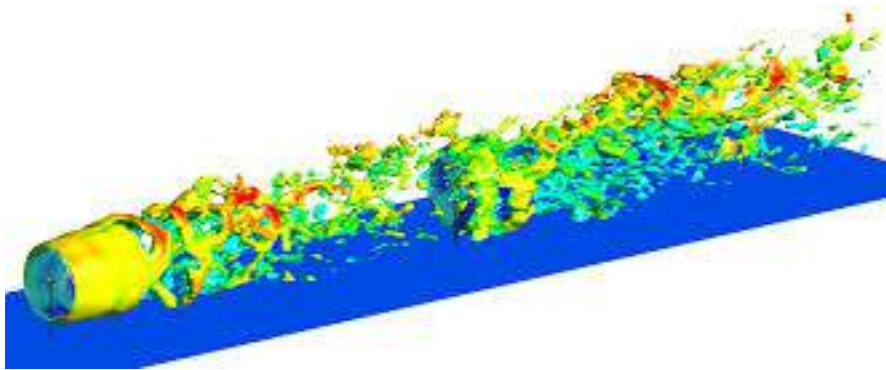


Applications Industrielles

Nouveaux matériaux, molécules pharmaceutiques:



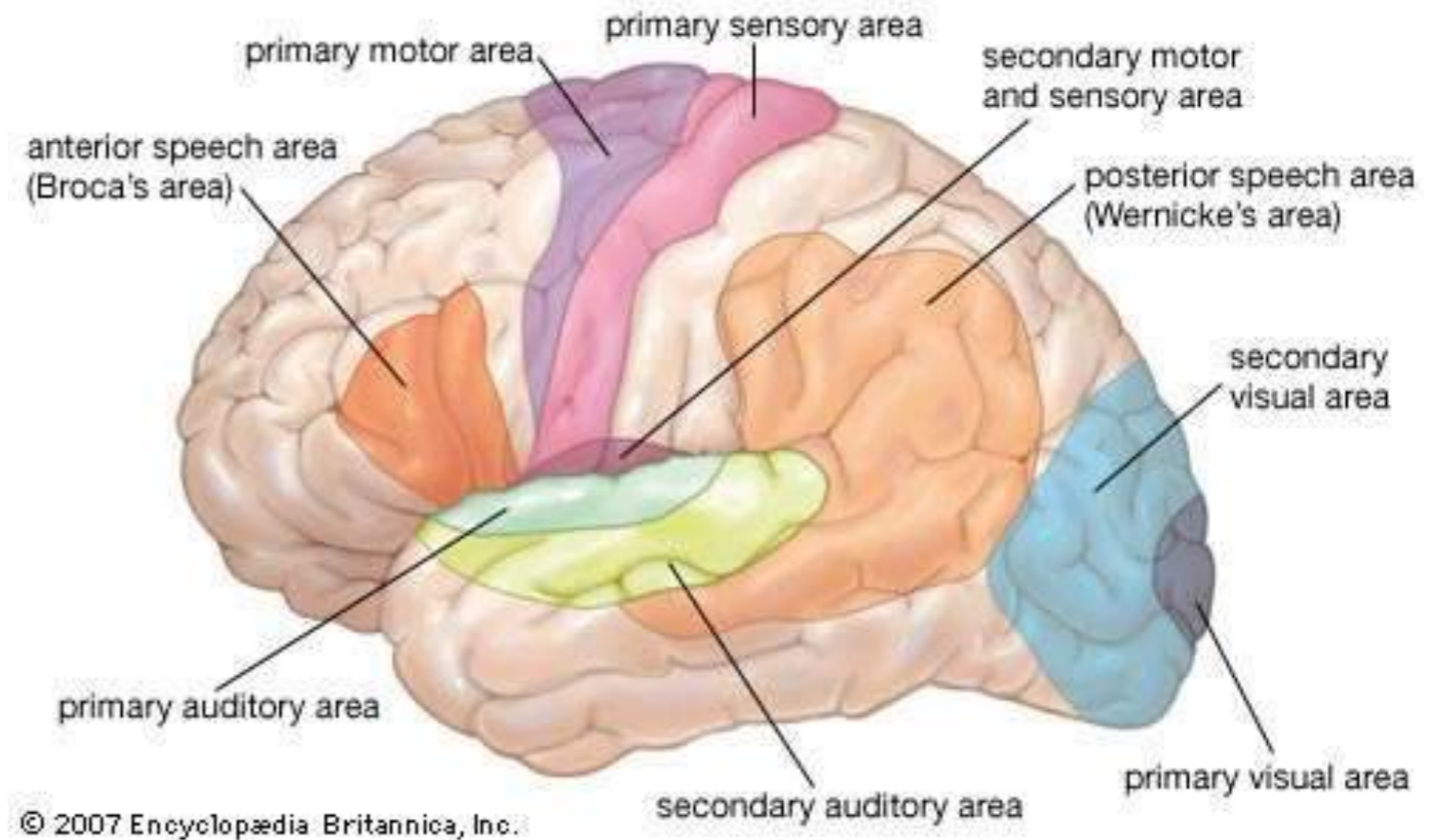
Simulations numériques rapide ?



Neurone



Cortex



- Quelle similarité entre réseaux de neurones artificiels et biologie ? traitements, niveaux fonctionnels, modèles...

Perception Visuelle

- Reconnaissance d'animaux: 150ms



Questions

- Pourquoi ces architectures sont elles « génériques » ? Appliquées à des image, sons, physique, médecine, langage naturel...
- 3 types de problèmes:
 - Estimation: analyser l'erreur de généralisation
 - Optimisation: minimiser l'erreur empirique
 - **Approximation: lien avec l'architecture**
- Quelle est l'information a priori ?
- Pourquoi des convolutions ?
- Quels types de filtres ?
- Role des non-linéarités ?
- Liens avec la neurophysiologie de la perception: image et audio ?

Contenu du Cours

- Estimation, optimisation et **approximation** d'un réseau
- Régularité: séparation, symétries, parcimonie.
- Symétrie: convolution et analyse de Fourier
- Séparation et parcimonie: principe d'incertitude et représentations temps-fréquence
- Neurophysiologie de l'audition et de la vision
- Transformées multiéchelles et ondelettes
- Invariance par translations, rotations et déformations
- Classification d'images et de sons: SIFT et MFCC
- Invariants multiéchelles: réseaux d'ondelettes et scattering
- Régression en chimie quantique
- Ce qui manque: apprentissage de symétries ou de formes ?



- 22 Janvier: Présentation de 8 challenges de données
- 29 Janvier: Présentation de 8 challenges de données
- 5 Février: Présentation des gagnants des challenges 2019
- 12 Février: *Geoffroy Peeters*, Applications à l'audio et musique
- 19 Février: Pas de cours et de séminaire.
- 26 Février : *Simon Thorpe*, Perception visuelle et CNN
- 4 Mars: *Gilles Wainrib*, Diagnostique médical interprétable
- 11 Mars : *Shihab Shamma*, Modèles physiologique de l'audition
- 18 Mars: *Edouard Oyallon*, Au-delà de l'information a priori