



IRIMAS UNIVERSITÉ
HAUTE-ALSACE

Institut de Recherche en Informatique, Mathématiques, Automatique et Signal

Intelligence Artificielle : La révolution Deep Learning

LYCÉE SCHWEITZER

Jonathan Weber, Maître de Conférences en Informatique

Jeudi 7 avril 2022

🐦 @jjmweber / 🌐 <https://www.jonathan-weber.eu> / 🗣️ JjmWeber

1. La révolution du Deep Learning
2. Le Deep Learning, qu'est-ce donc ?
3. Quelques exemples de travaux Mulhousiens
4. Takeway message

La révolution du Deep Learning



Figure 1 – MIT Technology Review : Brenden Lake

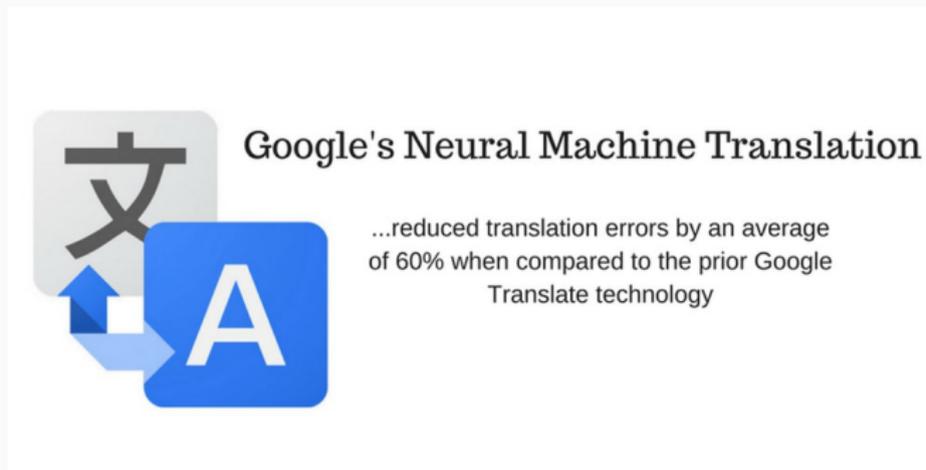


Figure 2 – Dilmanj : Poorya Zeynalzadeh

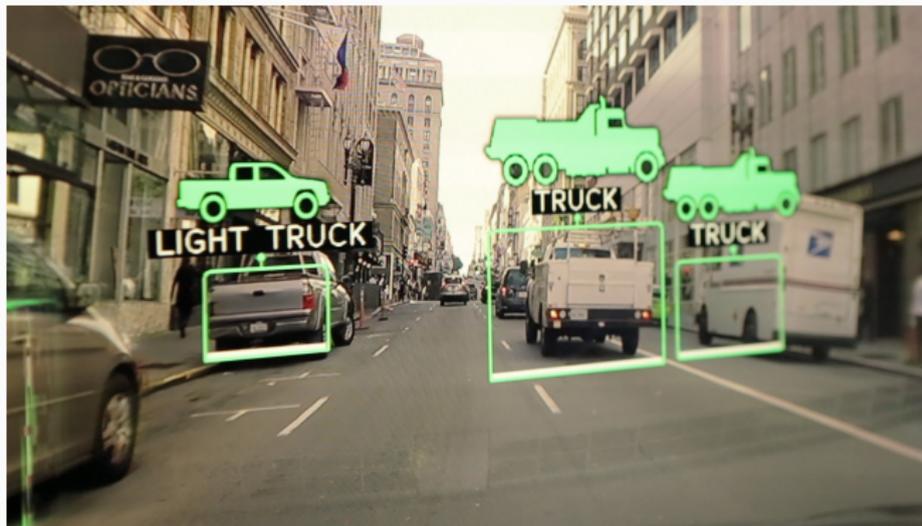


Figure 3 – Verne Global : Nick Dale

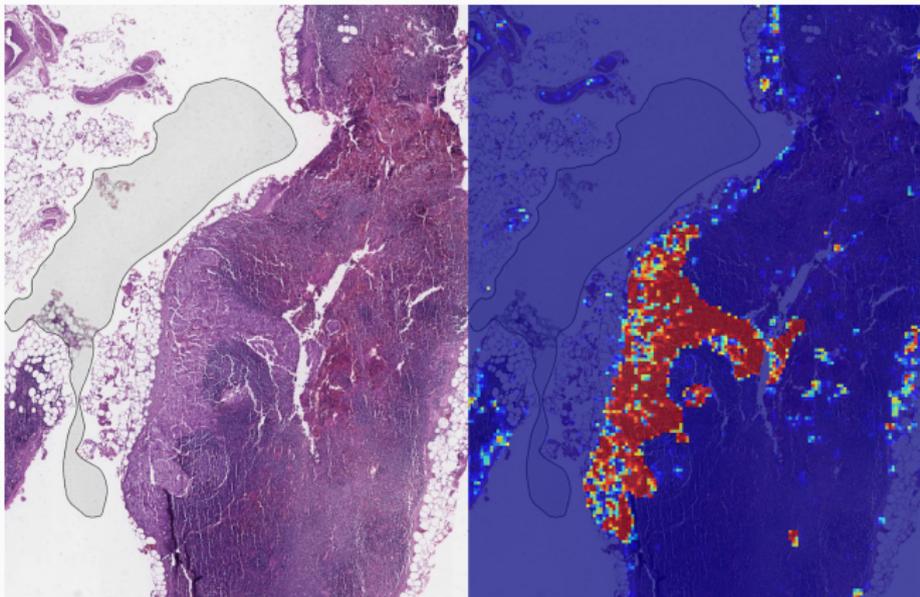
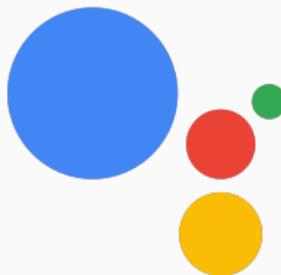


Figure 4 – Venture Beat : Kyle Wiggers



(a) Apple Siri



(b) Google Assistant



(c) Amazon Alexa



(d) Microsoft Cortana



Figure 5 – BBC : Technology

Détection de Malware (Nvidia MalConv)



Figure 6 – Cyber.tn/?p=3943



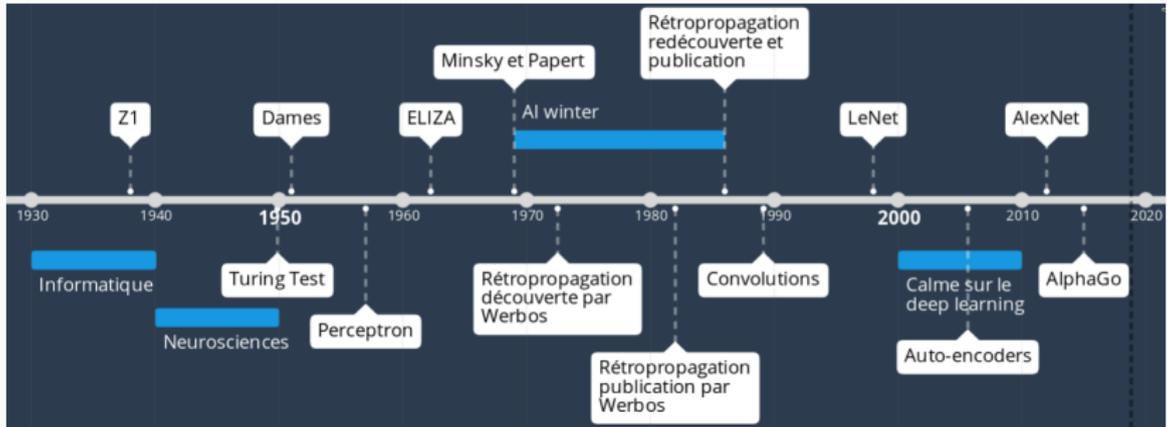
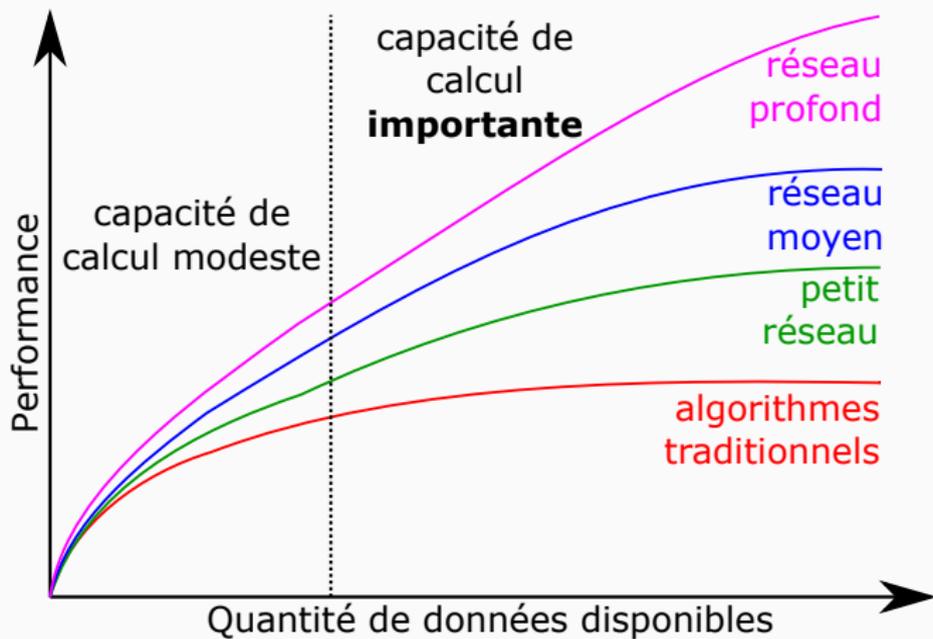


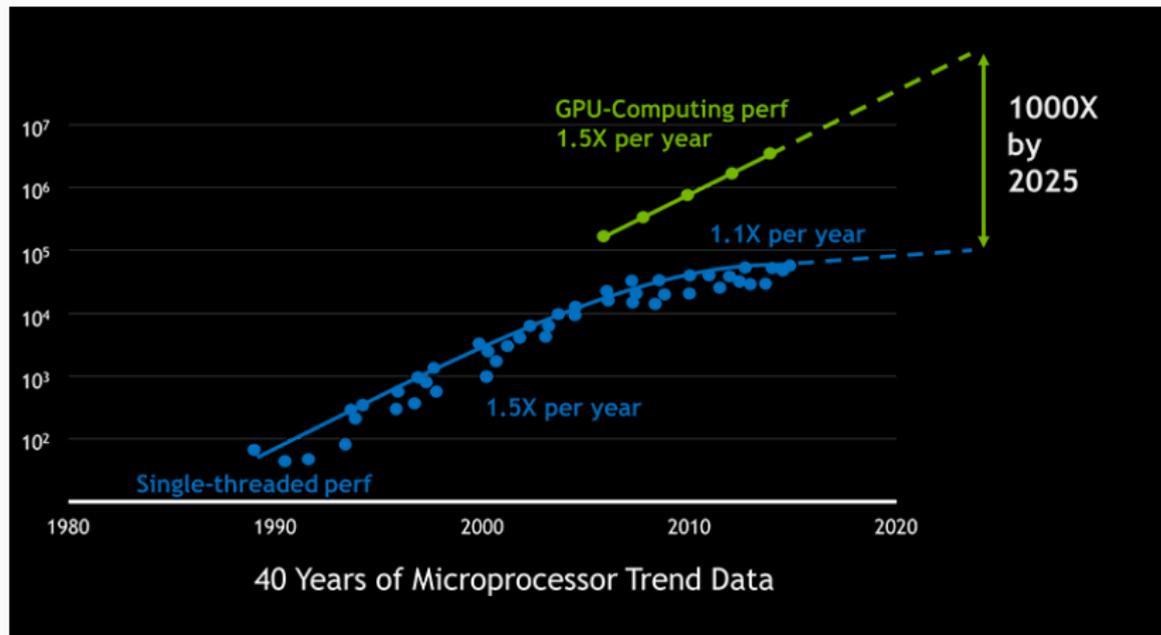
Figure 7 – Chronologie de l'Intelligence Artificielle et du Deep Learning

. <https://time.graphics/line/169596>

Pourquoi maintenant ?



Pourquoi maintenant ?



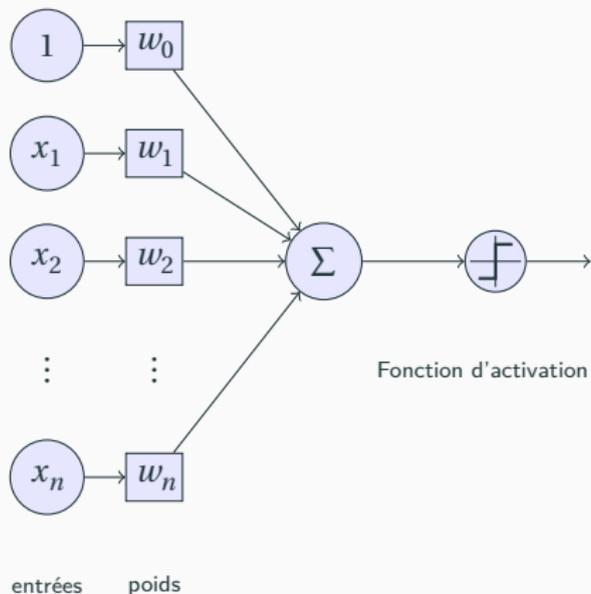
source : nvidia.com

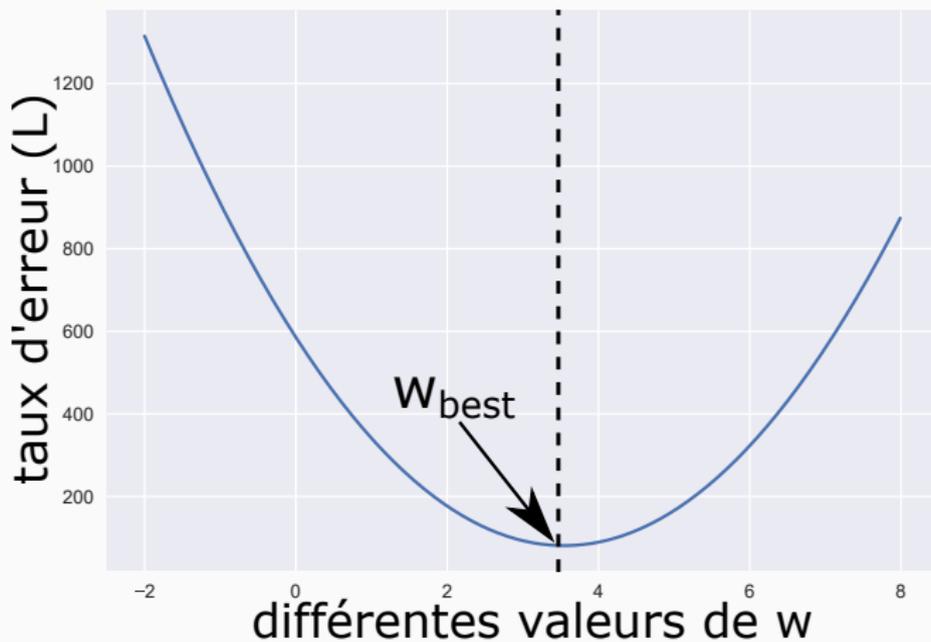
Pourquoi le deep learning se développe maintenant ?

- ▷ Quantité des données disponible
- ▷ Capacité de calcul disponible
- ▷ Algorithmes (plus profonds)

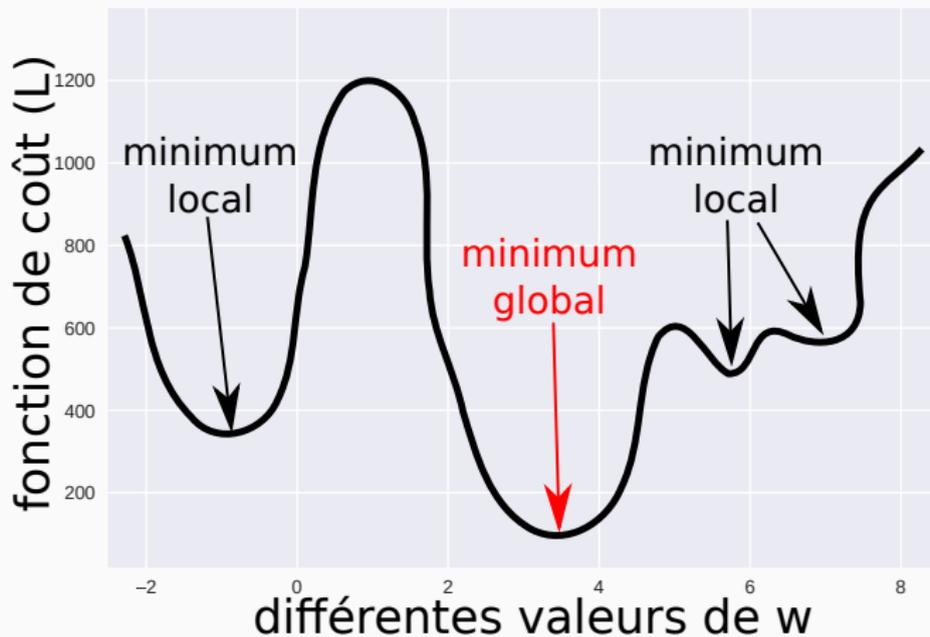
Le Deep Learning, qu'est-ce donc ?

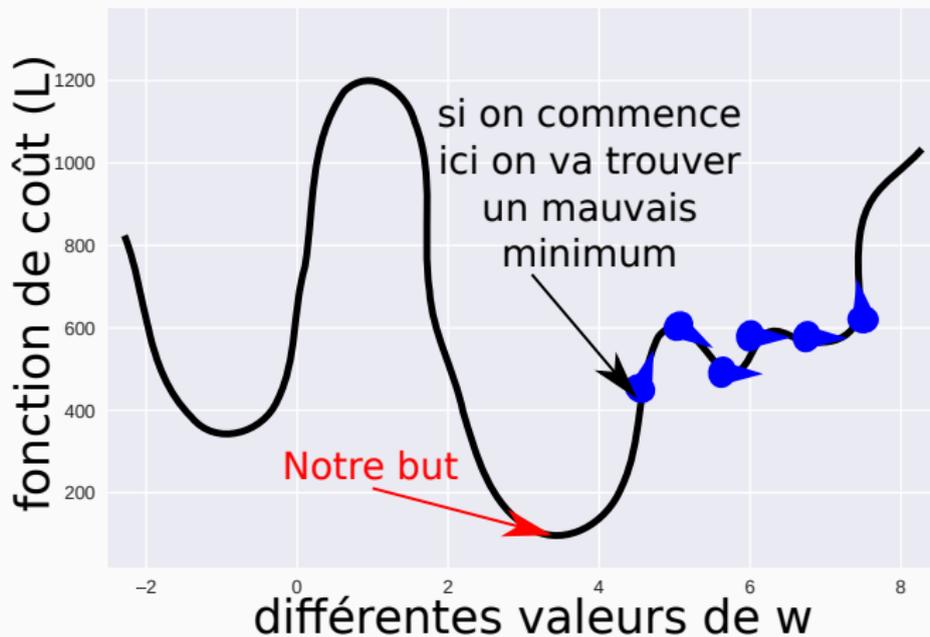
Perceptron :





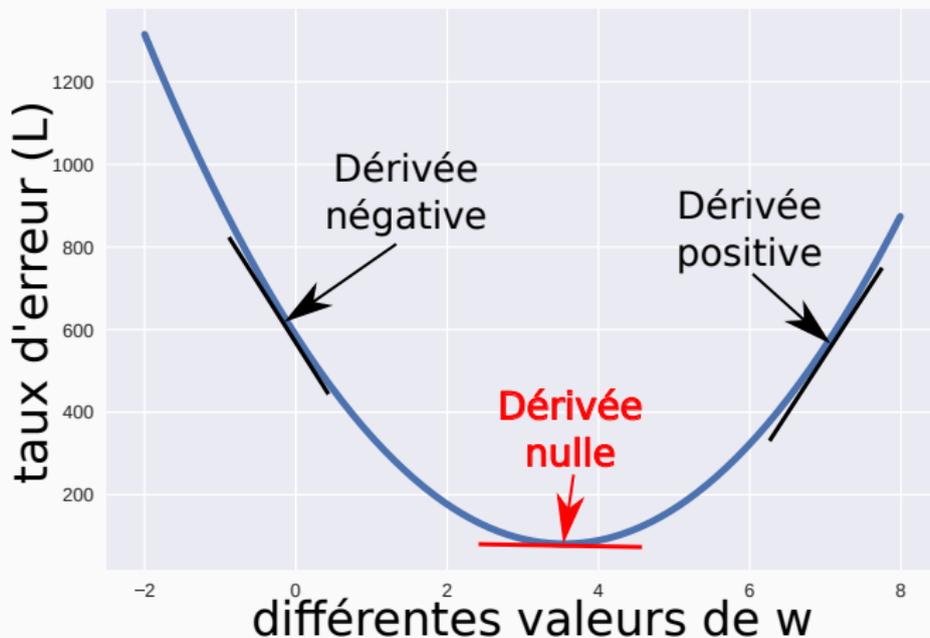
Minimiser une fonction de coût

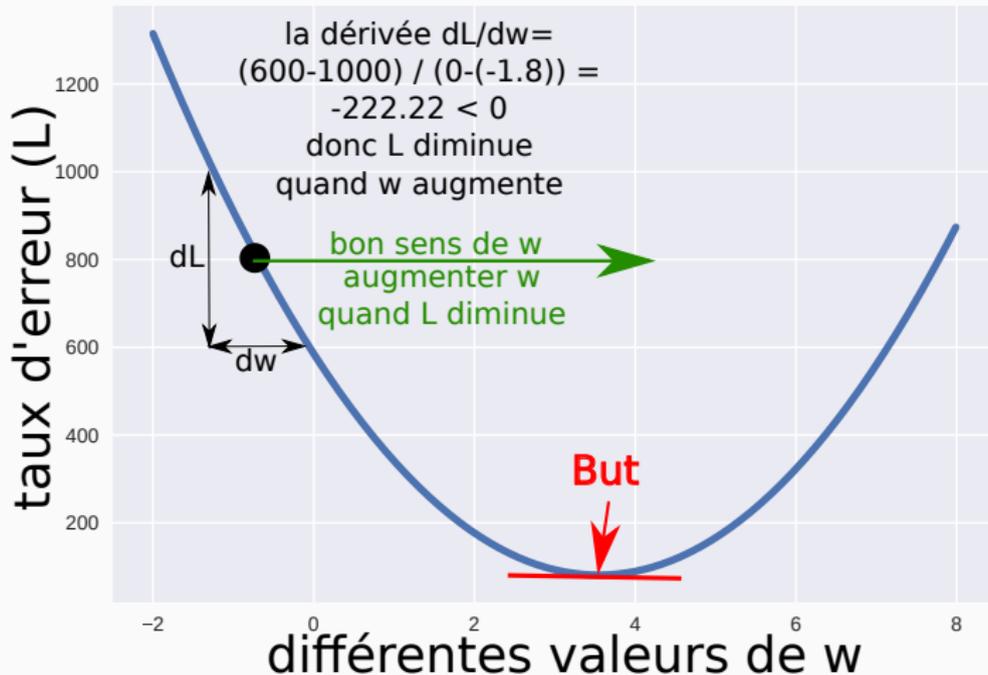




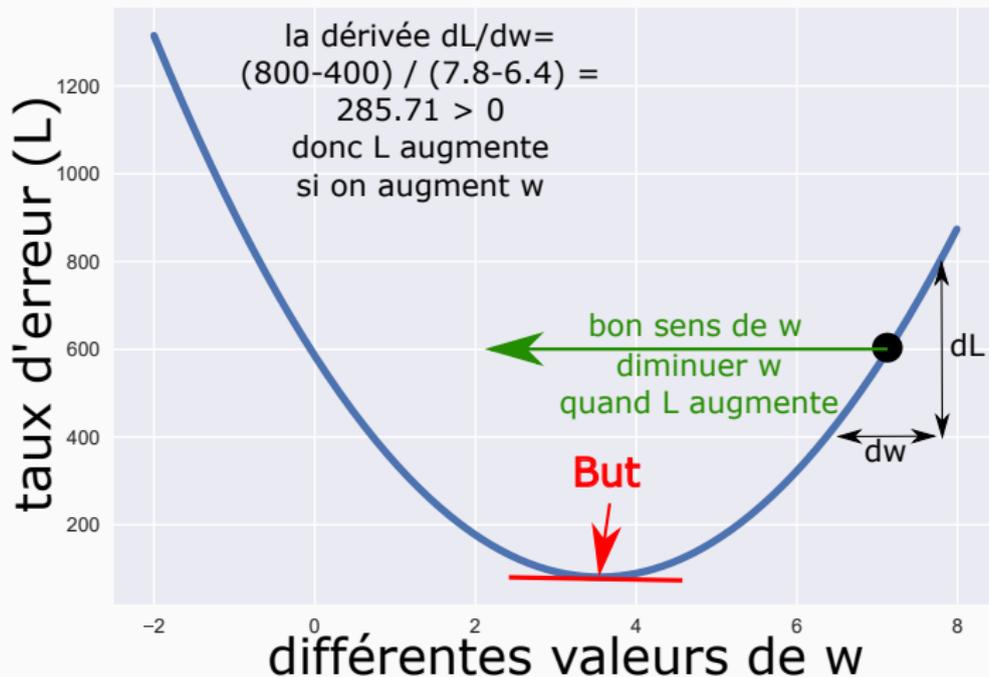
Dérivée (ou gradient)

Dérivée \Rightarrow augmentation partielle (très faible) de w





Le coût augmente



Comment varier w ?

Si w augmente, le taux d'erreur L augmente :

- ▷ C'est-à-dire qu'on est dans une montée
- ▷ Il faut diminuer w

Comment varier w ?

Si w augmente, le taux d'erreur L augmente :

- ▷ C'est-à-dire qu'on est dans une montée
- ▷ Il faut diminuer w

Si w diminue, le taux d'erreur L diminue :

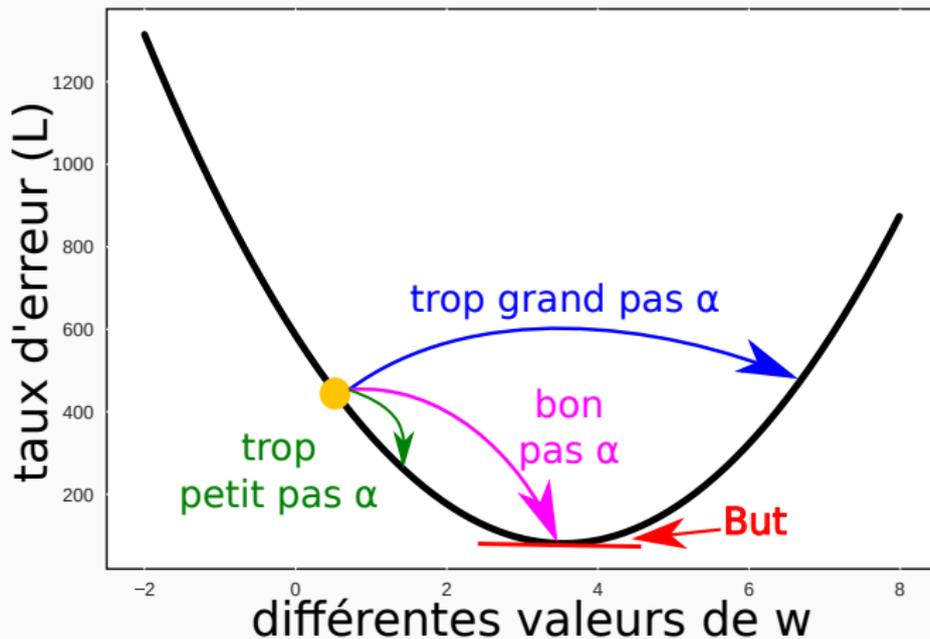
- ▷ C'est-à-dire on est dans une descente
- ▷ Il faut augmenter w

Comment varier w ?

$$w = w - \alpha \frac{\partial L}{\partial w} \quad (1)$$

- ▷ w : l'ensemble des paramètres à apprendre
- ▷ L : la fonction de coût qu'on veut minimiser
- ▷ α : le taux d'apprentissage qui contrôle la variation de w
- ▷ “-” : (le signe moins) signifie qu'on varie w dans le sens inverse de la variation de $L \implies$ on minimise l'erreur

Pourquoi a-t-on besoin de α le taux d'apprentissage ?



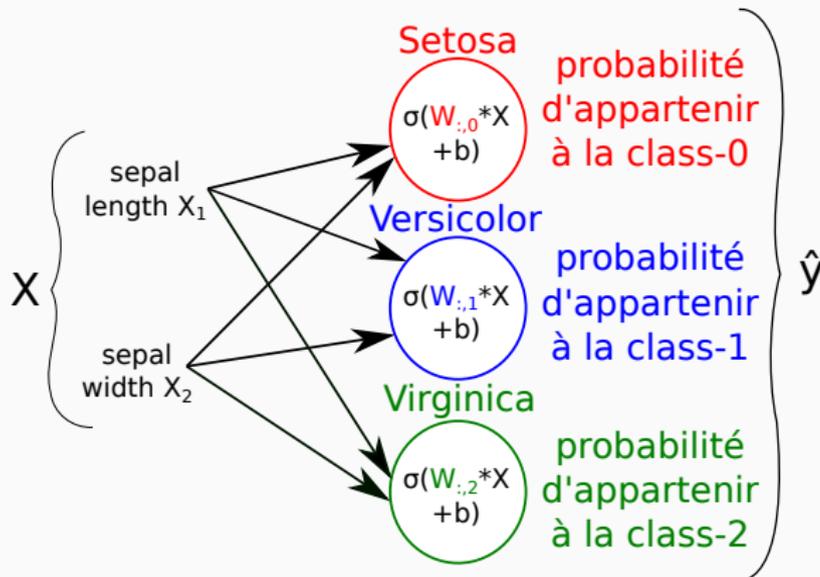
Comment choisir α :

- ▷ Un très grand α peut amener à rater l'objectif
- ▷ Un très petit α peut amener à une convergence trop lente nécessitant trop d'itérations

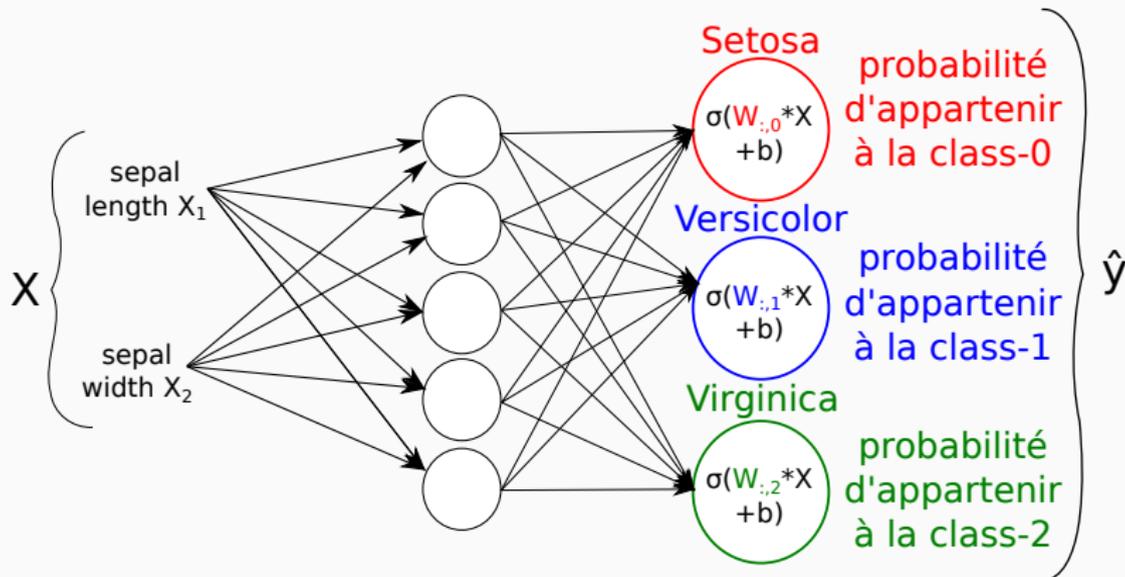
Comment choisir α :

- ▷ Un très grand α peut amener à rater l'objectif
 - ▷ Un très petit α peut amener à une convergence trop lente nécessitant trop d'itérations
- ⇒ Il faut choisir α pour avoir un équilibre entre le temps de calcul disponible et la précision du minimum qu'on veut atteindre

Un réseau à une seule couche (régression logistique)



Un réseau à plusieurs couches (une seule couche cachée)



Nécessaire pour apprendre des fonctions non-linéaires

$$\hat{y} = h(X) \quad | \quad h \text{ étant une fonction non-linéaire} \quad (2)$$

Nécessaire pour apprendre des fonctions non-linéaires

$$\hat{y} = h(X) \quad | \quad h \text{ étant une fonction non-linéaire} \quad (2)$$

La régression logistique donne une décision linéaire car il n'y a pas d'interaction non-linéaire entre les termes X_1 et X_2 par exemple

$$\hat{y} = \sigma(W_1 * X_1 + W_2 * X_2 + b) \quad (3)$$

Nécessaire pour apprendre des fonctions non-linéaires

$$\hat{y} = h(X) \quad | \quad h \text{ étant une fonction non-linéaire} \quad (2)$$

La régression logistique donne une décision linéaire car il n'y a pas d'interaction non-linéaire entre les termes X_1 et X_2 par exemple

$$\hat{y} = \sigma(W_1 * X_1 + W_2 * X_2 + b) \quad (3)$$

Pour la rendre non-linéaire il faut ajouter des termes d'interaction

$$\hat{y} = \sigma(W_1 * X_1 + W_2 * X_2 + W_3 * X_1 * X_2 + W_4 * X_1^2 + W_5 * X_2^2 + \dots + b) \quad (4)$$

Nécessaire pour apprendre des fonctions non-linéaires

$$\hat{y} = h(X) \quad | \quad h \text{ étant une fonction non-linéaire} \quad (2)$$

La régression logistique donne une décision linéaire car il n'y a pas d'interaction non-linéaire entre les termes X_1 et X_2 par exemple

$$\hat{y} = \sigma(W_1 * X_1 + W_2 * X_2 + b) \quad (3)$$

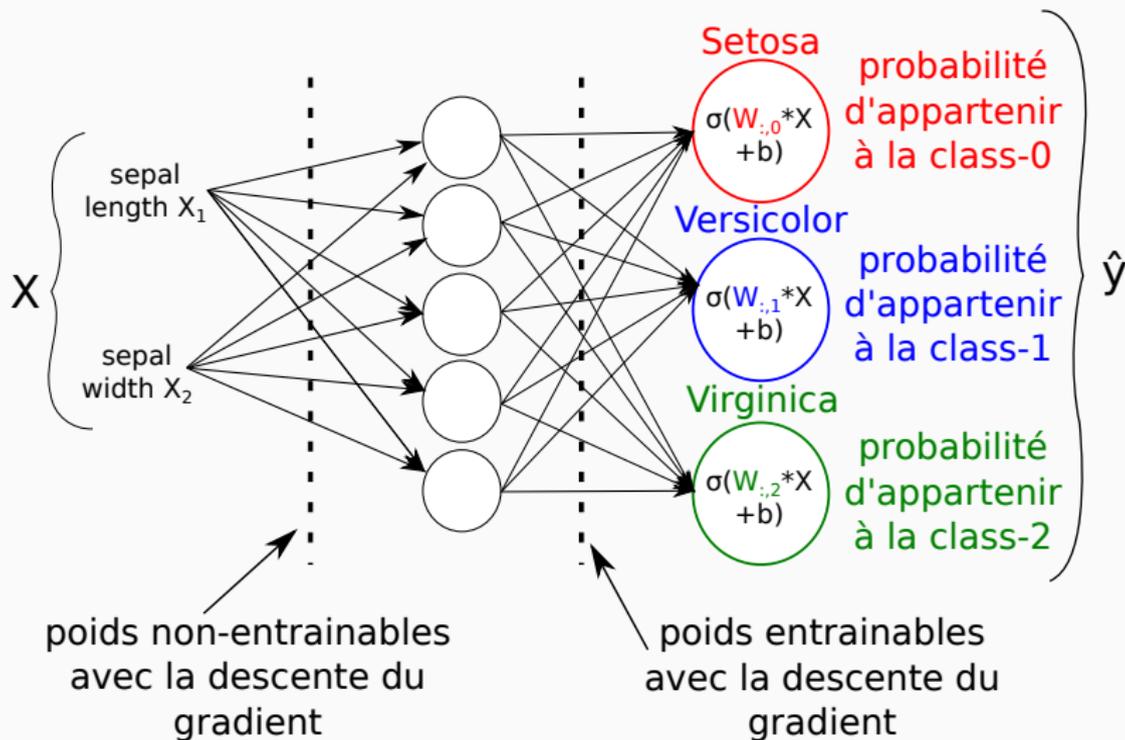
Pour la rendre non-linéaire il faut ajouter des termes d'interaction

$$\hat{y} = \sigma(W_1 * X_1 + W_2 * X_2 + W_3 * X_1 * X_2 + W_4 * X_1^2 + W_5 * X_2^2 + \dots + b) \quad (4)$$

Ceci est difficile à faire pour plusieurs raisons :

- ▷ Grand nombre d'attributs X : une image a 128×128 pixels
- ▷ On ne connaît pas quelle sorte d'interaction : $X_1^2 * \sqrt{X_2}$ ou $X_1^2 * X_2^3$?

Pourquoi est-ce difficile d'entraîner des réseaux profonds ?



On connaît uniquement ce que doit-être la sortie d'un neurone qui appartient à la dernière couche (non-cachée)

On connaît uniquement ce que doit-être la sortie d'un neurone qui appartient à la dernière couche (non-cachée)

⇒ On peut calculer l'erreur pour ces neurones

On connaît uniquement ce que doit-être la sortie d'un neurone qui appartient à la dernière couche (non-cachée)

⇒ On peut calculer l'erreur pour ces neurones

⇒ On peut appliquer la descente du gradient

On connaît uniquement ce que doit-être la sortie d'un neurone qui appartient à la dernière couche (non-cachée)

⇒ **On peut calculer l'erreur pour ces neurones**

⇒ **On peut appliquer la descente du gradient**

On ne connaît pas ce que doit-être la sortie d'un neurone appartenant à une couche cachée

On connaît uniquement ce que doit-être la sortie d'un neurone qui appartient à la dernière couche (non-cachée)

⇒ **On peut calculer l'erreur pour ces neurones**

⇒ **On peut appliquer la descente du gradient**

On ne connaît pas ce que doit-être la sortie d'un neurone appartenant à une couche cachée

⇒ **On ne peut pas calculer l'erreur pour ces neurones**

On connaît uniquement ce que doit-être la sortie d'un neurone qui appartient à la dernière couche (non-cachée)

⇒ **On peut calculer l'erreur pour ces neurones**

⇒ **On peut appliquer la descente du gradient**

On ne connaît pas ce que doit-être la sortie d'un neurone appartenant à une couche cachée

⇒ **On ne peut pas calculer l'erreur pour ces neurones**

⇒ **On ne peut appliquer directement la descente du gradient**

Pourquoi est-ce difficile d'entraîner des réseaux profonds ?

On connaît uniquement ce que doit-être la sortie d'un neurone qui appartient à la dernière couche (non-cachée)

⇒ **On peut calculer l'erreur pour ces neurones**

⇒ **On peut appliquer la descente du gradient**

On ne connaît pas ce que doit-être la sortie d'un neurone appartenant à une couche cachée

⇒ **On ne peut pas calculer l'erreur pour ces neurones**

⇒ **On ne peut appliquer directement la descente du gradient**

Il faut trouver un moyen pour estimer l'erreur pour un tel neurone appartenant à une couche cachée : rétropropagation

Deep learning : a fancy word for deep neural networks

Deep learning : a fancy word for deep neural networks

Deep learning \equiv Réseau de neurones profonds

Deep learning : a fancy word for deep neural networks

Deep learning \equiv Réseau de neurones profonds

Réseau de neurones profonds \equiv Réseau à plusieurs couches

Deep learning : a fancy word for deep neural networks

Deep learning \equiv Réseau de neurones profonds

Réseau de neurones profonds \equiv Réseau à plusieurs couches

Le succès du deep learning est dû à l'entraînement des réseaux de neurones à plusieurs couches pour apprendre des décisions non-linéaires

Quelques exemples de travaux Mulhousiens



- ▷ Université de Haute-Alsace
- ▷ Institut IRIMAS
- ▷ Département Informatique
- ▷ MSD : Modélisation et Science des Données



Prof. G. Forestier



Prof. P.A. Muller



Prof. M. Hassenforder



Dr. Jonathan Weber



Dr. Frédéric Fondement



Dr. Maxime Devanne

Non permanents

▷ Doctorants



M. Bamond



M. Čučulović



M. Hamri



R. Heckenauer



M. Negash



G. Pialla



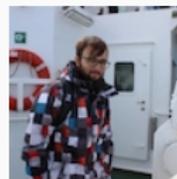
A. Roux



O. Schirm

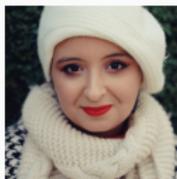


T. Terefe



R. Wenger

▷ Post-doctorants



Dr. A. Ben Hamida



Dr. B. Cuan



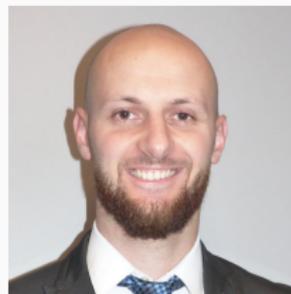
Dr. O. Jrad



Dr. H. Ismail Fawaz



Dr. B. Lafabregue



Dr. B. Latard



Dr. Bendali-Braham

IRIMAS



Pr. L. Idoumghar



Pr. J.P. Lauffenburger



Pr. C. Cudel



Dr. T. Josso-Laurain

Université de Strasbourg



Pr. C. Wemmert



Pr. A. Puissant

Institut Saint-Louis



Dr. S. Changey



Dr. N. Hueber

U. Paris



Dr. C. Kurtz

MHH (All.)



Pr. F. Feuerhake

CYU



Pr. J. Longhi

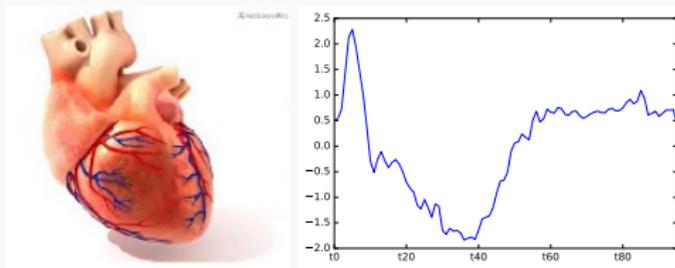
Monash (Aus.)



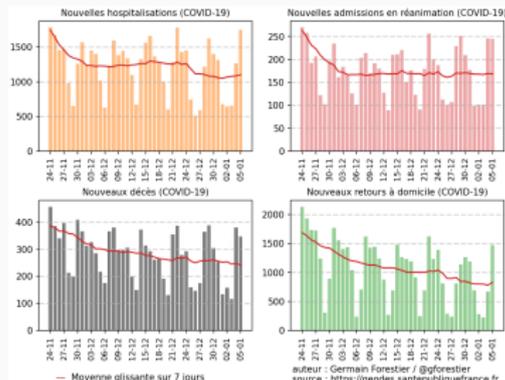
Pr. G. Webb

Les séries temporelles sont omni-présentes

- ▷ Suivi de l'évolution d'une valeur numérique ou symbolique dans le temps
- ▷ Si plusieurs valeurs, on parlera de séries multidimensionnelles
- ▷ Type de données présent dans de nombreux domaines d'application

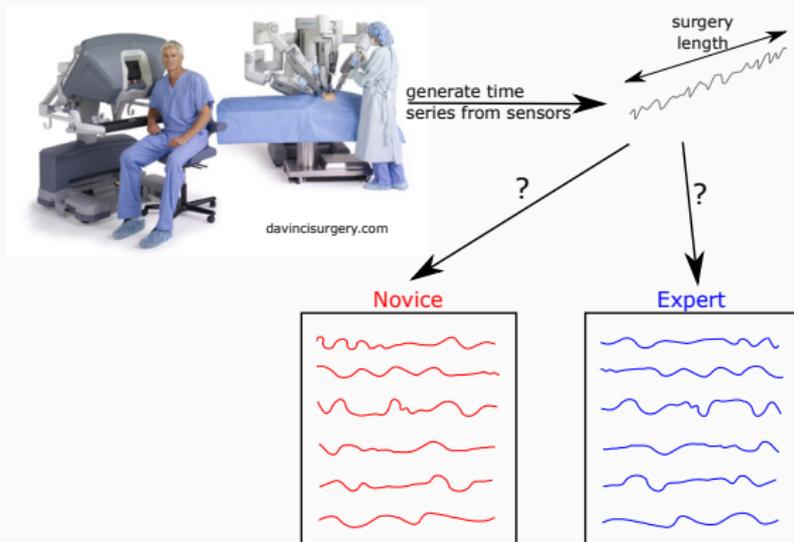


↑
A
→
X
←
B
B



La classification de séries temporelles

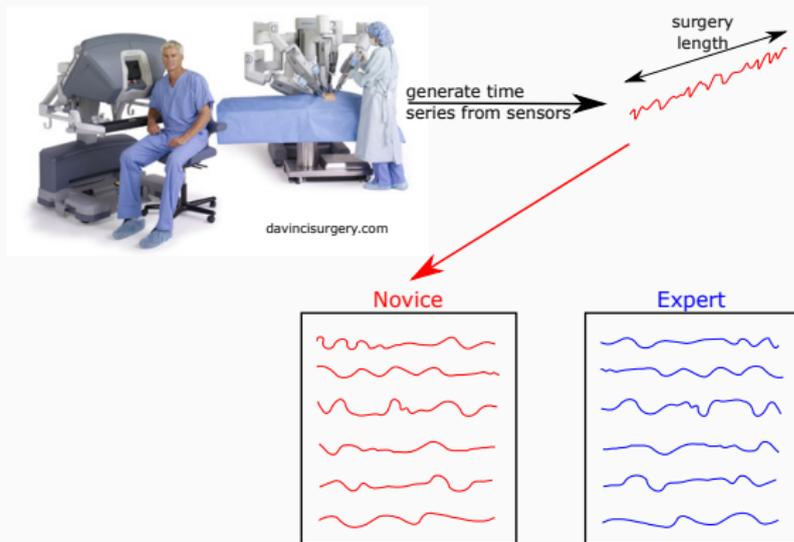
- ▷ Les séries temporelles sont regroupées par classe (par exemple Novice / Expert)
- ▷ L'objectif est d'affecter une classe à de nouvelles séries temporelles



 Petitjean, F., et al. (2014). Dynamic time warping averaging of time series allows faster and more accurate classification. IEEE International Conference on Data Mining

La classification de séries temporelles

- ▷ Les séries temporelles sont regroupées par classe (par exemple Novice / Expert)
- ▷ L'objectif est d'affecter une classe à de nouvelles séries temporelles

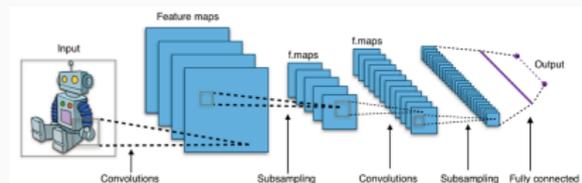


 Petitjean, F., et al. (2014). Dynamic time warping averaging of time series allows faster and more accurate classification. IEEE International Conference on Data Mining

L'apprentissage profond (deep learning)

Idée :

- ▷ Réseau de neurones avec plus de couches
- ▷ « Neurones » plus complexes
- ⇒ Possible grâce aux cartes graphiques



source : commons.wikimedia.org

Résultats :

- ▷ Révolution dans le domaine de la vision par ordinateur [1]
- ▷ Performances équivalentes aux humains en reconnaissance d'images [2]
- ▷ Adoptée en Traitement Automatique des Langues (TAL) :
 - ▷ Traduction automatique
 - ▷ Apprentissage de plongement lexical
- ▷ État de l'art pour la reconnaissance de la parole [3]

- [1] Krizhevsky, A., et al. (2012). Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems
- [2] Szegedy, C., et al. (2015). Going deeper with convolutions. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition
- [3] Hinton, G., et al. (2012). Deep neural networks for acoustic modeling in speech recognition : The shared views of four research groups. IEEE Signal Processing Magazine

Pourquoi l'appliquer aux séries temporelles ?

- ▷ Performant pour l'analyse de données *séquentielles* (texte, audio, etc.) [1]
- ▷ Capable de détecter des caractéristiques invariantes temporellement
 - ▷ Similairement aux filtres invariants spatialement dans les images
 - ▷ Nécessite moins de données pour les données 1D (séries temporelles) ?
- ▷ Capable de gérer des séries multidimensionnelles

 [1] Bahdanau, D., et al. (2015). Neural machine translation by jointly learning to align and translate, In Proceedings of International Conference on Learning Representations

Pourquoi l'appliquer aux séries temporelles ?

- ▷ Performant pour l'analyse de données *séquentielles* (texte, audio, etc.) [1]
- ▷ Capable de détecter des caractéristiques invariantes temporellement
 - ▷ Similairement aux filtres invariants spatialement dans les images
 - ▷ Nécessite moins de données pour les données 1D (séries temporelles) ?
- ▷ Capable de gérer des séries multidimensionnelles



source : miro.medium.com



source : i.morioh.com

■ [1] Bahdanau, D., et al. (2015). Neural machine translation by jointly learning to align and translate, In Proceedings of International Conference on Learning Representations

Pourquoi l'appliquer aux séries temporelles ?

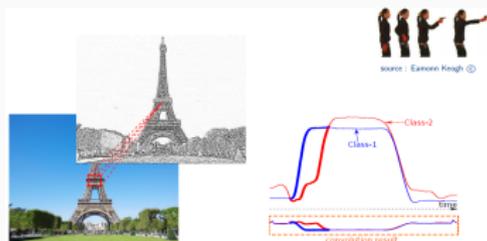
- ▷ Performant pour l'analyse de données *séquentielles* (texte, audio, etc.) [1]
- ▷ Capable de détecter des caractéristiques invariantes temporellement
 - ▷ Similairement aux filtres invariants spatialement dans les images
 - ▷ Nécessite moins de données pour les données 1D (séries temporelles) ?
- ▷ Capable de gérer des séries multidimensionnelles



source : miro.medium.com



source : i.morioh.com



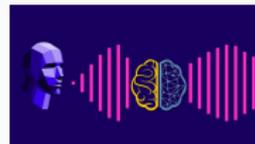
[1] Bahdanau, D., et al. (2015). Neural machine translation by jointly learning to align and translate, In Proceedings of International Conference on Learning Representations

Pourquoi l'appliquer aux séries temporelles ?

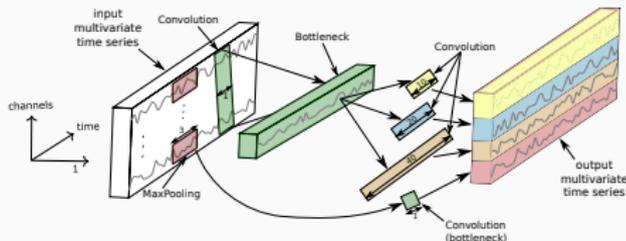
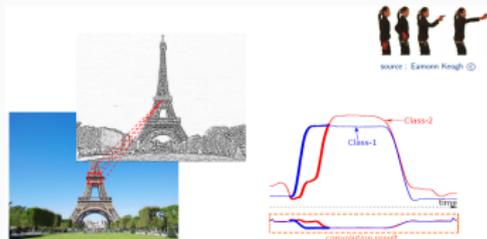
- ▷ Performant pour l'analyse de données *séquentielles* (texte, audio, etc.) [1]
- ▷ Capable de détecter des caractéristiques invariantes temporellement
 - ▷ Similairement aux filtres invariants spatialement dans les images
 - ▷ Nécessite moins de données pour les données 1D (séries temporelles) ?
- ▷ Capable de gérer des séries multidimensionnelles



source : miro.medium.com



source : i.morioh.com

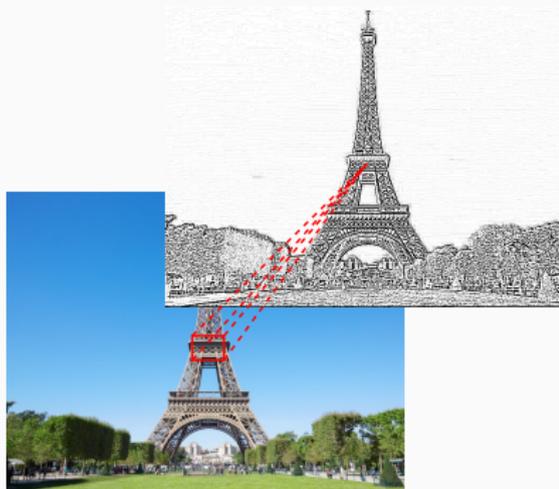


[1] Bahdanau, D., et al. (2015). Neural machine translation by jointly learning to align and translate, In Proceedings of International Conference on Learning Representations

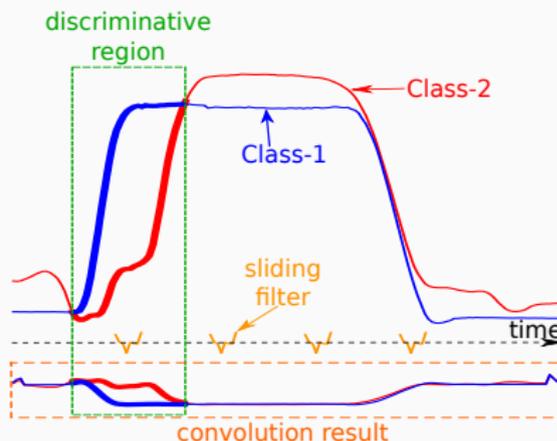
Convolution : image vs. séries temporelles



source : Eamonn Keogh ©



Résultat de l'application d'une convolution de détection de contours.



Résultat de l'application d'une convolution discriminante apprise sur le jeu de données GunPoint.

Contexte : Science des données chirurgicales

- ▷ Problème : Évaluation subjective et inexplicable [1]
- ▷ Besoin : Amélioration du retour lors des entraînements chirurgicaux [2]
 1. Objectif
 2. Précis
 3. Explicable

Da Vinci Surgical System

- ▷ Multiple capteurs \Rightarrow séries multidimensionnelles
- ▷ Échantillonnage haute fréquence \Rightarrow big data
- ▷ Données hétérogènes \Rightarrow modèles hybrides



source : Mark Clifford ©

- [1] Kassahun, Y., et al. (2016). Surgical robotics beyond enhanced dexterity instrumentation : a survey of machine learning techniques and their role in intelligent and autonomous surgical actions. *International journal of computer assisted radiology and surgery*, 11(4), 553-568.
- [2] Islam, G., et al. (2016). Affordable, web-based surgical skill training and evaluation tool. *Journal of biomedical informatics*, 59, 102-114.

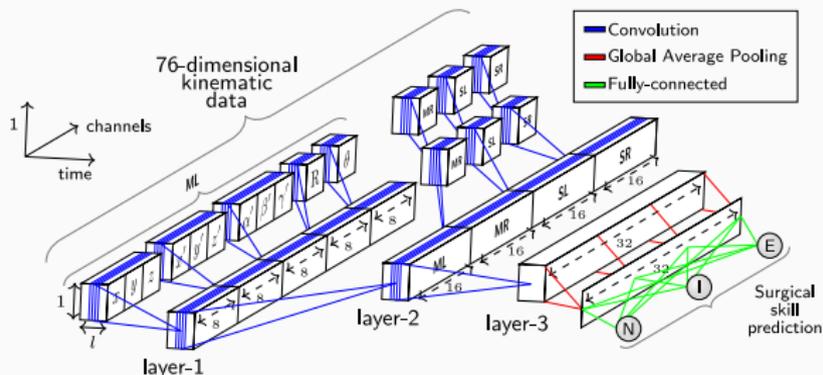
Notre approche pour la classification de chirurgies

Jeu de données JIGSAWS [1]



Les 3 tâches chirurgicales : suture, nouage, passage d'aiguille

Notre approche FCN pour l'évaluation de compétences chirurgicales [2]



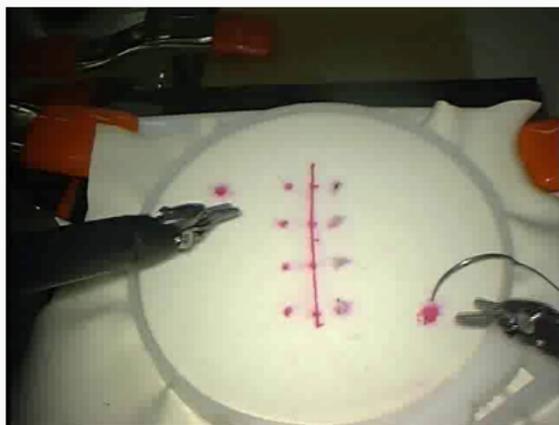
- [1] Gao, Y., et al. (2014). JHU-ISI gesture and skill assessment working set (JIGSAWS) : A surgical activity dataset for human motion modeling. In MICCAI Workshop : M2CAI (Vol. 3, p. 3).
- [2] Ismail Fawaz, H., et al. (2018). Evaluating surgical skills from kinematic data using convolutional neural networks. Int. Conf. On Medical Image Computing and Computer Assisted Intervention (MICCAI)

- ▷ Pas d'optimisation d'hyper-paramètres \implies architecture identique pour toutes les tâches

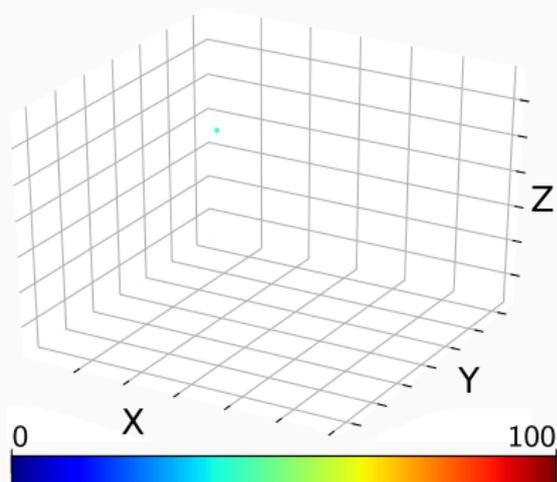
Méthode	Suture		Passage d'aiguille		Nouage	
	Micro	Macro	Micro	Macro	Micro	Macro
S-HMM [1]	97.4	n/a	96.2	n/a	94.4	n/a
ApEn [2]	100	n/a	100	n/a	99.9	n/a
Sax-Vsm [3]	89.7	86.7	96.3	95.8	61.1	53.3
Notre approche	100	100	100	100	92.1	93.2

🔗 Code et résultats disponibles sur <https://github.com/hfawaz/miccai18>

- 📄 [1] Tao, L., et al. (2012). Sparse hidden markov models for surgical gesture classification and skill evaluation. In International conference on information processing in computer-assisted interventions (pp. 167-177).
- 📄 [2] Zia, A., & Essa, I. (2018). Automated surgical skill assessment in RMIS training. International journal of computer assisted radiology and surgery, 13(5), 731-739.
- 📄 [3] Forestier, G., et al. (2017). Discovering discriminative and interpretable patterns for surgical motion analysis. In Proceedings of Artificial Intelligence in Medicine in Europe (pp. 136-145).

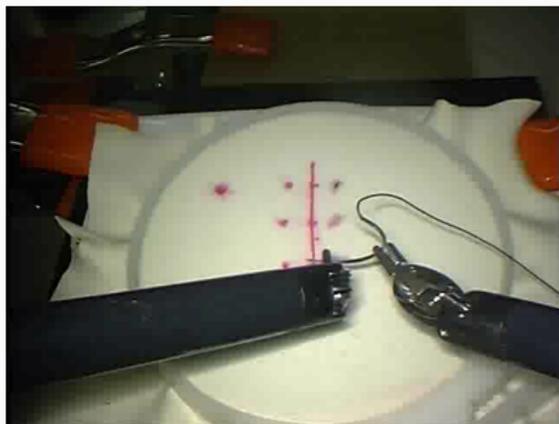


Dernière trame du 4^{ème} essai du sujet H (Novice) pour la suture

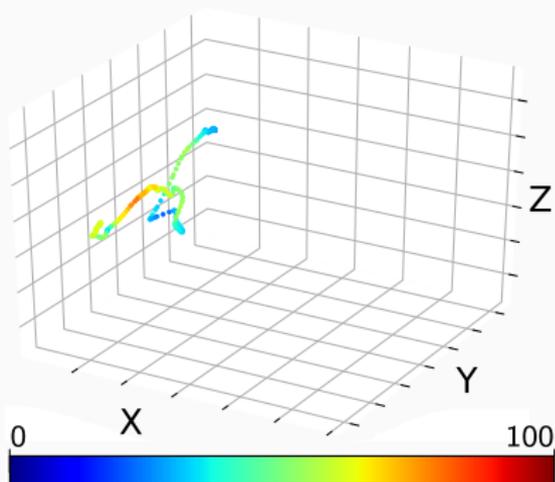


% of contribution in a classification

Trajectoire du manipulateur gauche

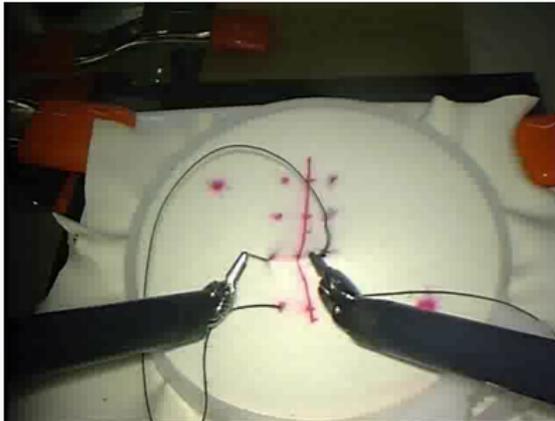


Dernière trame du 4^{ème} essai du sujet H (Novice) pour la suture

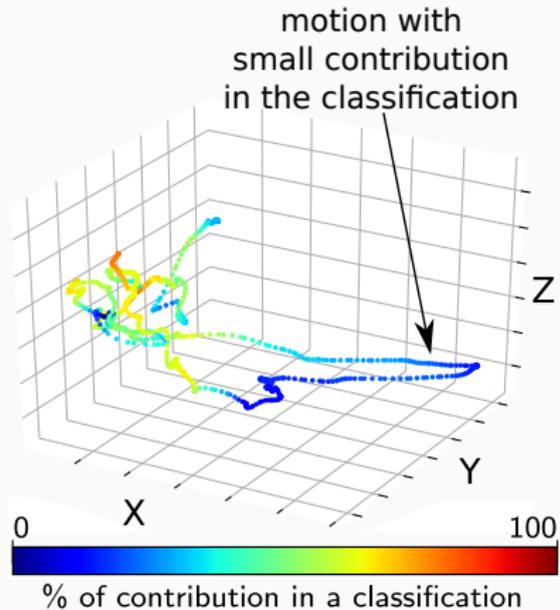


% of contribution in a classification

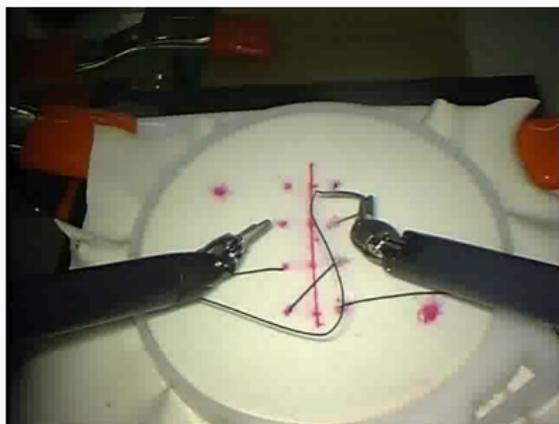
Trajectoire du manipulateur gauche



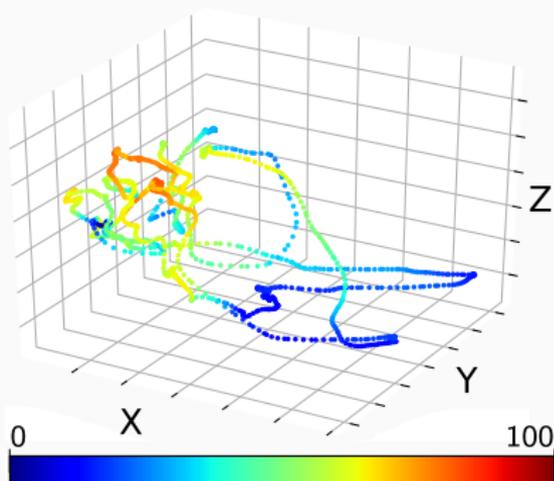
Dernière trame du 4^{ème} essai du sujet H (Novice) pour la suture



Trajectoire du manipulateur gauche



Dernière trame du 4^{ème} essai du sujet H (Novice) pour la suture

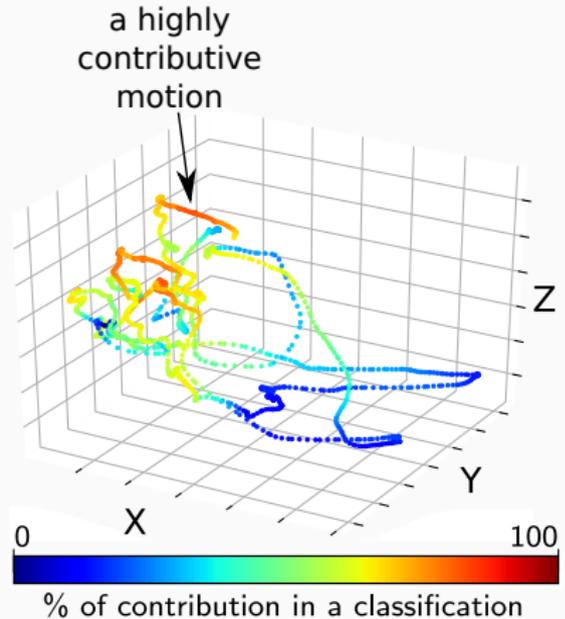


% of contribution in a classification

Trajectoire du manipulateur gauche



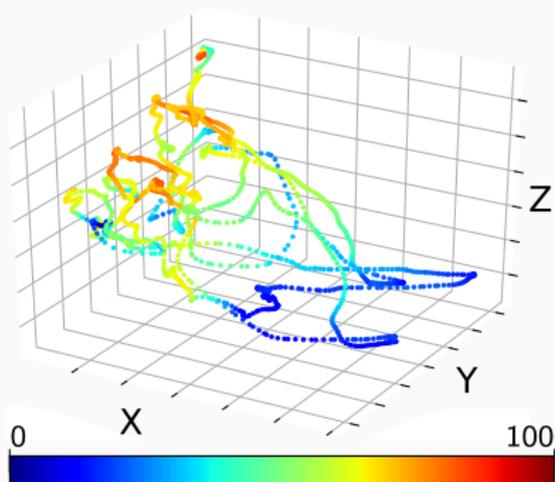
Dernière trame du 4^{ème} essai du sujet H (Novice) pour la suture



Trajectoire du manipulateur gauche

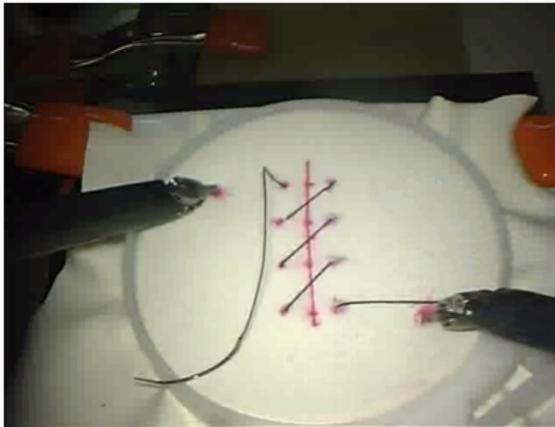


Dernière trame du 4^{ème} essai du sujet H (Novice) pour la suture

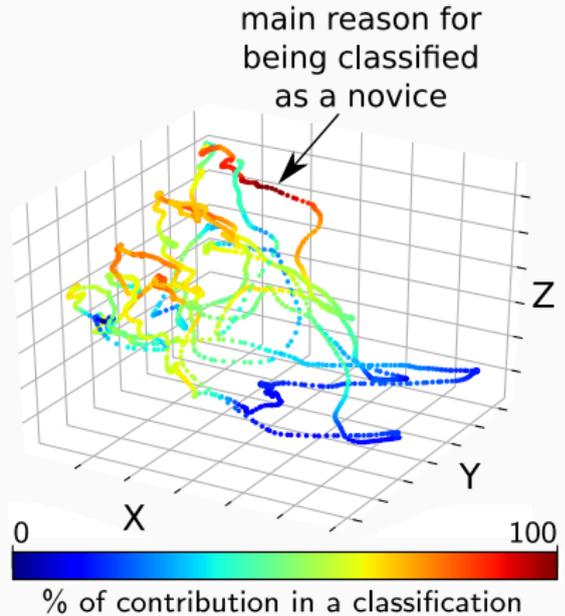


% of contribution in a classification

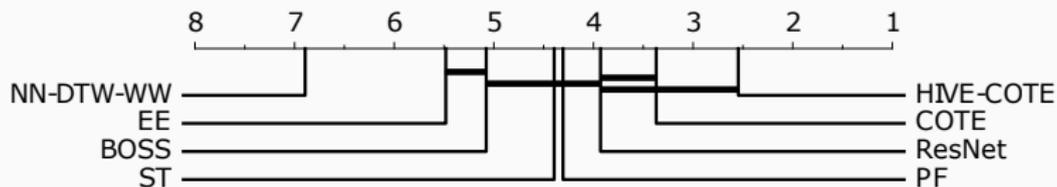
Trajectoire du manipulateur gauche



Dernière trame du 4^{ème} essai du sujet H (Novice) pour la suture

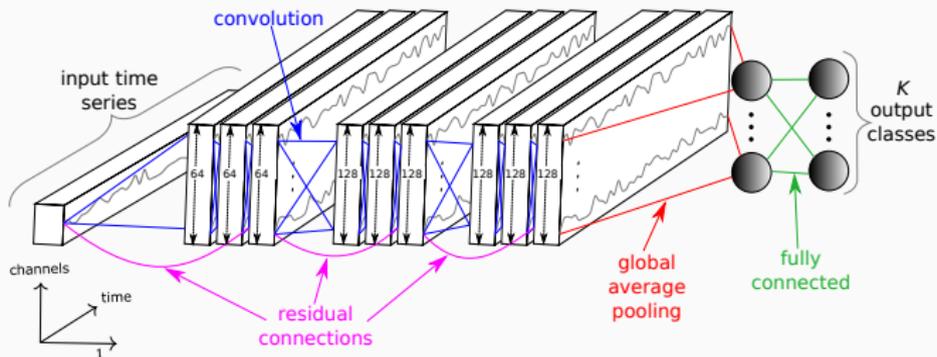


Classification de séries temporelles (CST)

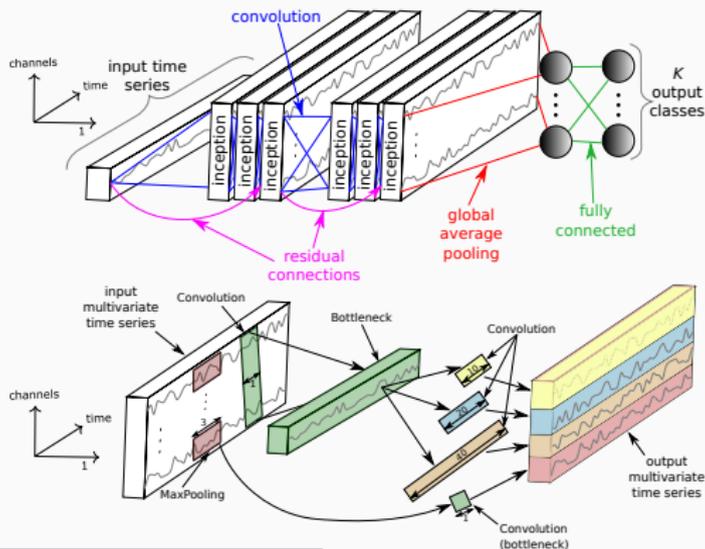
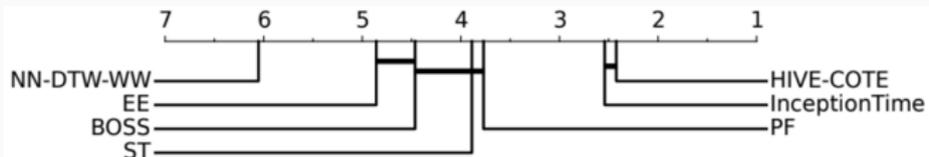


▷ ResNet obtient des performances similaires à HIVE-COTE

- ▷ Mais entraînement et prédiction plus rapides
- ▷ Besoin de plus de jeux de données
- ▷ Besoin de plus gros jeux de données

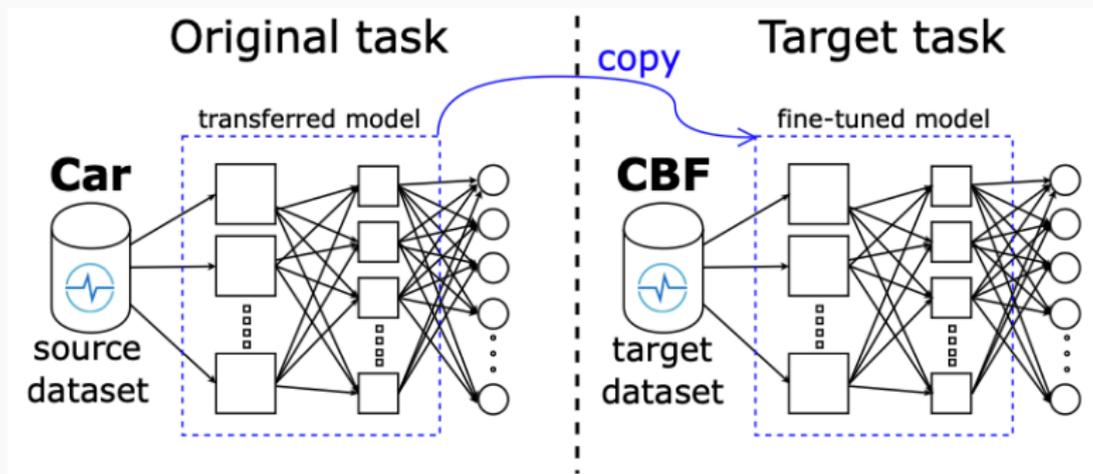


CST : InceptionTime



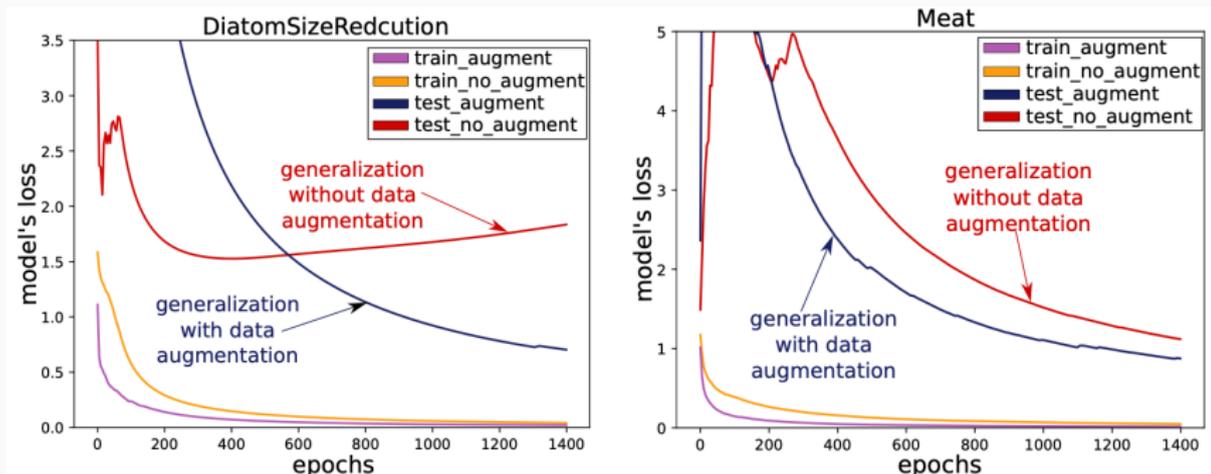
. Fawaz, H. I. et.al (2020). InceptionTime : Finding AlexNet for Time Series Classification. Data Mining and Knowledge Discovery, 34(6), 1936-1962.

- ▷ Est-il possible de partager de la connaissance entre jeux de données ?



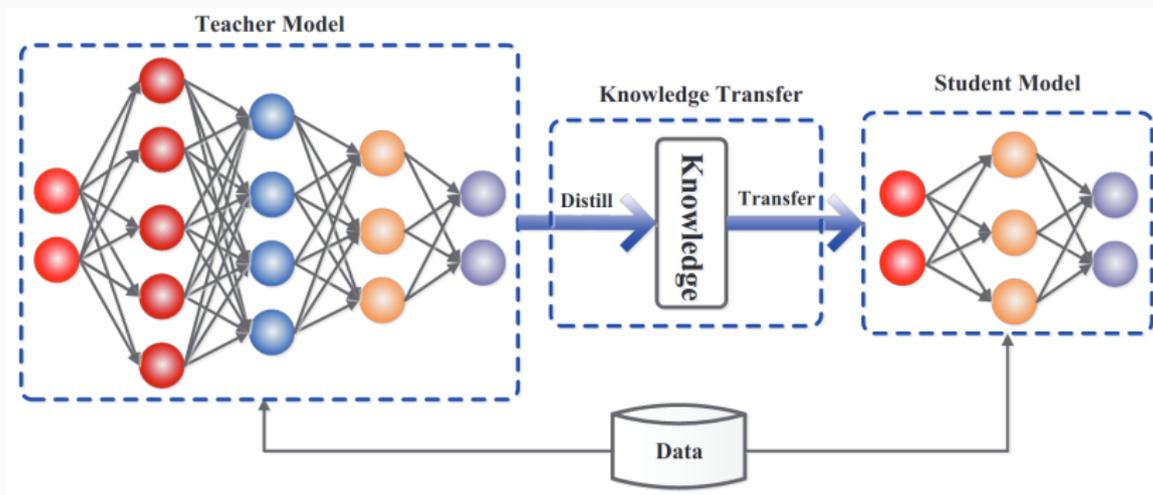
. Ismail Fawaz, H. et.al (2018). Transfer learning for time series classification. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 1367-1376). IEEE

▷ Est-ce que cela permet une meilleur généralisation ?



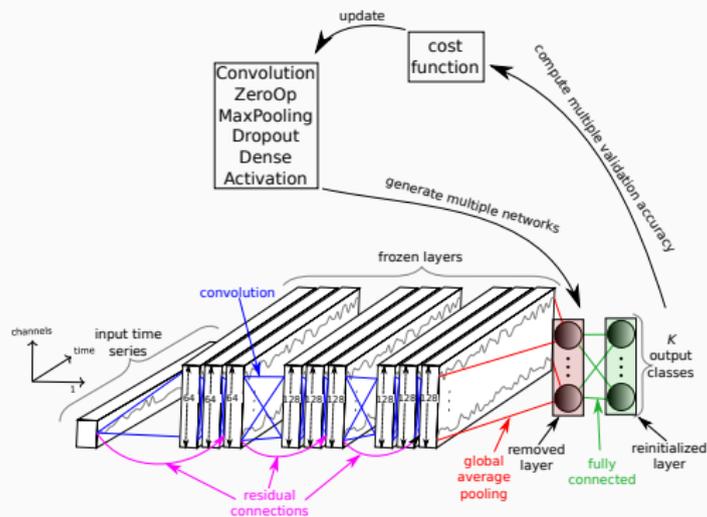
. Ismail Fawaz, H. et.al (2018). Data augmentation using synthetic data for time series classification with deep residual networks. In ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data. Dublin, Ireland

▷ Réseaux plus petits donc plus rapides ?



Neural architecture search

- ▷ Collaboration avec l'équipe OMeGa d'IRIMAS
- ▷ Travaux initiés lors des thèses de Hojjat Rakhshani et Hassan Ismail Fawaz [1]
- ▷ NAS-T : meilleur que HIVE-COTE mais pas significativement

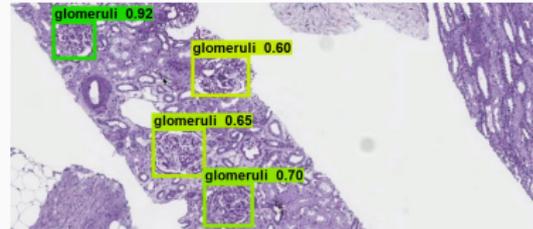


- ▷ Futures perspectives :
 - ▷ Appliquer NAS-T à d'autres architectures profondes et d'autres problèmes
 - ▷ Metaheuristique pour l'ensemble du réseau

[1] Rakhshani, H., et al. (2020). Neural Architecture Search for Time Series Classification. IEEE International Joint Conference on Neural Networks (IJCNN).

Analyse d'images histopathologiques

- ▷ Méthodes d'apprentissage profond
- ▷ Précédents travaux sur la détection temps réel de glomérules (rein) [1]
- ▷ Projet AiCOLO : Cancer colorectal
 - ▷ Post-doctorat de Amina Ben Hamida (2020-2023)
 - ▷ Thèse de Robin Heckenauer (2019-2022)
 - ▷ Données : Images histopathologiques du colon
 - ⇒ Classification de tissus
 - ⇒ Prédiction de marqueurs génétiques
- ▷ Projet MI*EDGE : Glioblastome (cerveau)
 - ▷ Post-doctorat de Bonan Cuan (2020-2023)
 - ▷ Données : Images histopathologiques et IRM du cerveau
 - ⇒ Prédiction de rechute



[1] Heckenauer, R., et al. (2020). Real-time detection of glomeruli in renal pathology. IEEE International Symposium on Computer Based Medical Systems (CBMS).

Apprentissage profond pour la navigation

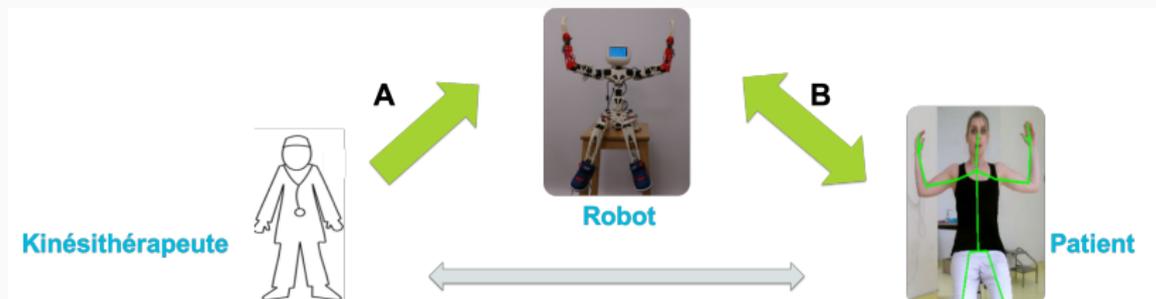
- ▷ Collaboration avec l'Institut Saint-Louis et le département ASI d'IRIMAS
- ▷ Thèse d'Alicia Roux (2020-2023)
- ▷ Estimation état courant (position, vitesse, direction) d'une munition en se basant sur les capteurs embarqués
- ▷ Deux approches :
 1. End-to-end
 2. Hybride : réseau pour l'estimation dynamique de matrices de Kalman



- ▷ DELEGATION, ANR JCJC
 - ▷ Deep learning for generating human motion
 - ▷ Application à la rééducation assistée par un robot kiné
- ▷ REVEIL, Fondation Paul Bennetot
 - ▷ Étude de la rééducation en réalité virtuelle

Collaboration :

- ▷ Centre de Réadaptation de Mulhouse
- ▷ CHRU, Brest



Takeway message

Deep learning

- ▷ Il n'y a pas d'intelligence artificielle
- ▷ Ce ne sont que des systèmes statistiques
- ▷ Possible grâce à l'augmentation de la puissance de calcul
- ▷ Mais avec des performances incroyables dans de nombreux domaines !

Merci à

- ▷ Hassan Ismail Fawaz, pour certains de ses slides -
 <https://hfawaz.github.io/>
- ▷ Maxime Devanne, pour certains de ses slides -
 <https://www.maxime-devanne.com/>
- ▷ Germain Forestier, pour certains de ses slides -
 <https://germain-forestier.info/>

Si plus tard, vous voulez travailler dans le deep learning, contactez-nous !