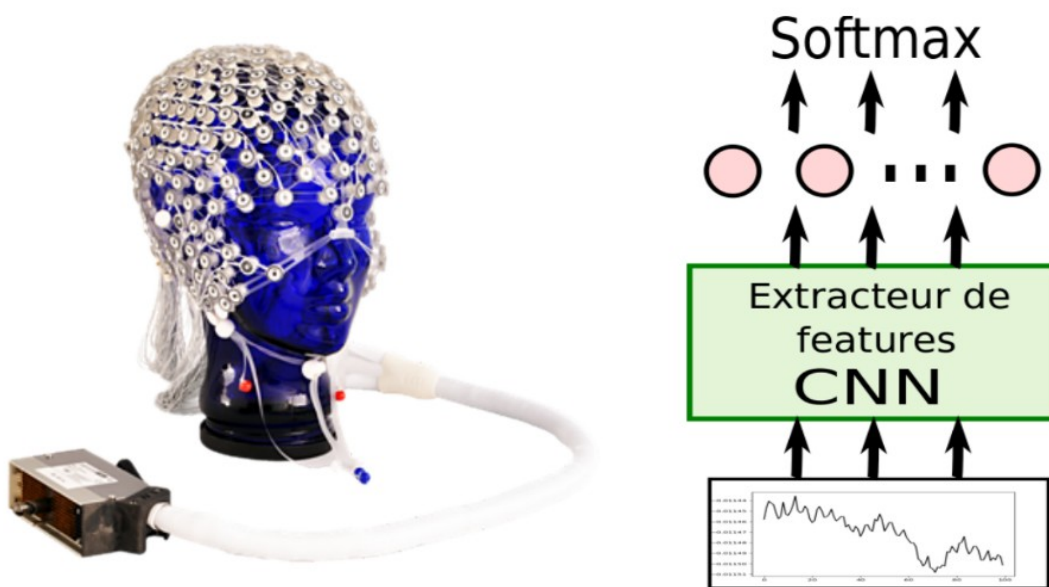


## Apprentissage profond pour l'analyse d'électroencéphalogrammes (EEG) long terme

L'objectif de la thèse est de proposer des méthodes reposant sur l'apprentissage profond pour l'analyse d'électroencéphalogrammes (EEG) long terme, c'est à dire correspondant à des acquisitions de l'ordre de plusieurs dizaines de minutes. La thèse est adossée à une base de données collectée à l'institut du cerveau et de la moelle épinière (ICM), qui contient des EEG uniques acquis à une fréquence de 250 Hz pendant des durées de 20 à 30 minutes, pour plus d'une centaines de patients. Deux tâches applicatives ayant de forts enjeux médicaux sont ciblées : la classification d'états de conscience (coma, état végétatif, état de conscience minimale, et état conscient) et la prédiction d'identité.



*Illustration 1: Dispositif d'acquisition d'EEG et apprentissage profond*

Le premier axe de recherche envisagé pour la thèse consiste à mener une réflexion sur les architectures de réseaux de neurones adaptées à l'analyse d'EEG long terme. Si les réseaux de neurones profonds connaissent un regain d'intérêt fort depuis le succès historique remporté au challenge ImageNet 2012 [KSH12], les architectures pertinentes pour les données EEG nécessitent une étude particulière. Nous proposons d'explorer des architectures convolutives, très efficaces pour extraire des représentations hiérarchiques locales des données de types signaux pauvres en sémantique. Une première réflexion consistera à définir des filtres de convolution dont la réponse fréquentielle et le champ réceptif soient adaptés au type de pattern recherché pour caractériser le contenu des EEG ; nous nous appuierons sur l'expertise de l'ICM sur ce point, et axerons notre analyse sur un intervalle fréquentiel de  $[0,5 ; 45]$  Hz. Un premier réseau convolutif profond de taille «moyenne» sera défini ( $\sim 10^4$  paramètres), et évalué pour la tâche de reconnaissance d'identité et de prédiction d'états de conscience. Cette étude initiale sera basée sur l'analyse de fragments temporels d'EEG de 800 ms. Les performances des réseaux convolutifs seront comparées aux méthodes de références disponibles à l'ICM, basées sur l'extraction de marqueurs manuellement extraits des données, et l'utilisation de méthodes d'apprentissage

«plates» (i.e. non profondes) comme des forêts aléatoires.

Nous explorerons ensuite des extensions de ce premier réseau de référence autour des aspects suivants. Tout d'abord, nous proposerons un réseau sensiblement plus profond, le nombre de paramètres étant un aspect crucial pour augmenter les performances des modèles de deep learning. Afin d'améliorer la robustesse au sur-apprentissage, nous explorerons deux pistes principales. Tout d'abord, nous aurons recours à des techniques d'augmentation de données, pour lesquelles une réflexion particulière sera menée pour adapter le type de transformation préservant l'information de classe au contenu des EEG. D'autre part, nous explorerons des méthodes de transfert de connaissances entre tâches. Les bases de données publiques étant rares sur les EEG long terme, nous utiliserons des méthodes d'apprentissage multi-tâches [K17], dont le principe consiste à utiliser des objectifs d'apprentissage différents mais corrélés pour partager des représentations intermédiaires entre tâches. Nous nous appuyerons sur les compétences dans l'équipe MSDMA pour mettre en place des schémas d'apprentissage multi-tâches reposant sur la définition d'une tâche principale [MTH+18]. Enfin, nous étudierons des réseaux de neurones récurrents, par leur capacité explicite de modélisation de l'information temporelle. Nous étudierons des architectures récurrentes classiques comme les LSTM [HS97] ou les GRU [CMG+14], et proposerons également des architectures hybride convolutives-récurrentes.

Le second axe de recherche concerne la prise en compte de l'information spatiale entre les différentes électrodes pour l'analyse du contenu de signaux EEG. Les architectures de réseaux de neurones profonds étudiées jusqu'ici reposent sur l'analyse du signal moyenné sur les différentes électrodes. Bien que ce type d'approche soit utilisée dans la littérature [SBM+18], une information précieuse et fortement caractéristique du signal est ignorée. Nous étudierons la possibilité de conserver l'information spatiale entre les différentes électrodes. Pour cela, nous étudierons des méthodes basées sur la convolution entre les signaux d'électrodes voisines. A nouveau, la nature spécifique des données sera tirée à profit pour adapter la notion de topologie et de voisinage aux différentes électrodes des EEG. Nous explorerons les convolutions sur graphe, revisitées récemment pour l'apprentissage avec des modèles de réseaux de neurones profonds, i.e. «Graph Convolutional Networks» (GCN) [BZS+13]. Les GCN permettent de généraliser les outils de convolutions à une cardinalité de voisinage indéterminée. Bien que ce formalisme repose sur la définition de la convolution dans le domaine spectral, on se retrouve rapidement confronté à des problèmes de temps de calcul. Nous étudierons des méthodes permettant le passage à l'échelle de ces méthodes, e.g. [KW16, MBB17]. Nous pourrions mettre à profit les travaux de l'équipe Vertigo sur les recherches de motifs spatio-temporels pertinents pour améliorer ces modèles [VF19]. Afin de valider la pertinence des modèles de GCN proposés, nous comparerons les performances obtenues avec la méthode de référence de l'axe 1, i.e. sans information spatiale, ainsi qu'à une approche encodant de manière naïve l'information spatiale des différentes électrodes en concaténant les signaux.

Le troisième axe de la thèse consiste à localiser temporellement les fragments d'EEG responsables de l'information de supervision annotée dans les données. Ceci ouvre la voie à la possibilité d'interprétation du modèle de prédiction du réseau de neurones. Deux directions seront explorées pour cela.

Tout d'abord, nous étudierons des méthodes d'apprentissage faiblement supervisé, adaptées notamment pour effectuer des prédictions locales (ici de détection temporelles) à

partir de modèles entraînés avec une supervision globale, e.g. un label précisant la classe d'un EEG long terme. Un aspect central dans les approches d'apprentissage faiblement supervisé concerne la définition de la fonction de pooling utilisée pour convertir les prédictions locales en labels globaux sur les données. Nous nous appuyerons sur les travaux récents menés dans l'équipe MSDMA sur ce point, notamment par l'utilisation de fonction de pooling de type «negative evidence» [DTC19]. Une adaptation de ces approches sera menée à deux niveaux. D'une part, nous étudierons la façon dont l'information issue des données multi-électrodes des EEG long terme peut être prise en compte. D'autre part, nous explorerons l'apprentissage global des réseaux profonds («end-to-end training») basé sur l'optimisation de fonctions objectif faiblement supervisées.

Un second type d'approche pour localiser temporellement les signatures EEG consistera à explorer les modèles attentionnels. Des approches issues de la communauté du traitement du langage naturel [BCG+14] sont aujourd'hui intensivement utilisées pour introduire une information de localisation sur des données comme le texte ou plus pauvres en sémantique comme les images, et sont utilisés dans les modèles état de l'art en légndage d'image («Image Captioning») [XBK+15] ou en réponse à des questions visuelles. Nous nous baserons sur les compétence de l'équipe MSDMA sur ces aspects, notamment sur la proposition récente de modèles attentionnelles plus riches par rapport aux cartes d'attention visuelles utilisées dans la littérature, permettant un raisonnement itératif pour la localisation de zones d'intérêt [CBT+19]. Une réflexion sera menée pour prendre en compte la nature des données EEG multicanal dans le processus de sélection temporelle.

[BCG+14] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473, 2014

[BJL+17] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. "Geometric deep learning: going beyond euclidean data". In: IEEE Signal Processing Magazine 34.4 (2017), pp. 18–42.

[BZS+13] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. "Spectral networks and locally connected networks on graphs". In: arXiv preprint arXiv:1312.6203 (2013).

[CBT+19] R. Cadene, H. Ben-younes, N. Thome, M. Cord. MUREL: Multimodal Relational Reasoning for Visual Question Answering. CVPR 2019.

[DTC19] Thibaut Durand, Nicolas Thome, Matthieu Cord. Exploiting Negative Evidence for Deep Latent Structured Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41(2), pp. 337-351, 2019.

[CMG+14] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. Learning phrase representations using rnn encoder-decoder for statistical machine translation, EMNLP 2014.

[HS97] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. Neural Comput., 9(8):1735–1780.

[K17] Iasonas Kokkinos. UberNet: Training a universal convolutional neural network for low-,

mid-, and high-level vision using diverse datasets and limited memory. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (NIPS), 2012.

[KW16] Thomas N Kipf and Max Welling. “Semi-supervised classification with graph convolutional networks”. In: arXiv preprint arXiv:1609.02907 (2016).

[MBB17] Federico Monti, Michael Bronstein, and Xavier Bresson. “Geometric matrix completion with recurrent multi-graph neural networks”. In: Advances in Neural Information Processing Systems. 2017, pp. 3697–3707.

[MTH+18] Taylor Mordan, Nicolas Thome, Gilles Henaff, and Matthieu Cord. Revisiting multi-task learning with ROCK: a deep residual auxiliary block for visual detection. In Advances in Neural Information Processing Systems (NIPS), December 2018.

[SBM+18] Arnaud Sors, Stéphane Bonnet, Sébastien Mirek, Laurent Vercueil, Jean-François Payen: A convolutional neural network for sleep stage scoring from raw single-channel EEG. Biomed. Signal Proc. and Control 42: 107-114 (2018)

[VF19] Tiphaine Viard, Raphaël Fournier-S'niehotta: Augmenting content-based rating prediction with link stream features. *Computer Networks*, 150, 127–133 (2019).

[XBK+15] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In Proceedings of the 32nd International Conference on Machine Learning, volume 37 of Proceedings of Machine Learning Research, pages 2048–2057, Lille, France, 07–09 Jul 2015.