

HABILITATION A DIRIGER DES RECHERCHES

Date de la soutenance : **12 décembre 2024**

Nom de famille et prénom de l'auteur : **Monsieur LEFORT Mathieu**

Titre des travaux : « *Apprentissage de représentations structurées et perception multimodale, vers des systèmes sensorimoteurs plus autonomes* »



Résumé

Mes axes de recherche s'inscrivent entre les sciences cognitives/neuroscience et l'intelligence artificielle, avec une emphase claire sur ce dernier domaine. Alors que mes travaux initiaux durant ma thèse se situaient en neuroscience computationnelle, en apportant une vision informatique à certains mécanismes de calcul et d'apprentissage de structures corticales, j'ai ensuite progressivement orienté mes recherches vers l'apprentissage automatique, en lien également avec la vision par ordinateur. Je conserve toujours certains liens avec les neurosciences et les sciences cognitives, et un plus important avec la psychophysique, mais je les utilise principalement comme une source d'inspiration pour obtenir des informations sur les processus cérébraux pour les transférer à des systèmes artificiels. Mes recherches sont ainsi interdisciplinaires, ce qui est facilité et supporté par mes collaborations récurrentes avec des collègues de ces autres domaines.

Je présente dans le chapitre 2 certains travaux sur des méthodes d'apprentissage auto-supervisé (SSL) de représentations visuelles. Premièrement je montre des travaux préliminaires sur une approche utilisant un mécanisme simple à base de saillance pour construire automatiquement des paires de proto-objets reliés sémantiquement pour l'apprentissage de représentations visuelles, et sur une étude dédiée au rôle de l'action dans la structure et l'apprentissage de représentations visuelles pertinentes lors de la réception de parties d'images. Deuxièmement, en nous appuyant sur ces idées, nous avons proposé un nouveau module basé sur l'équivariance, i.e. une transformation dans l'espace latent peut être prédite par l'augmentation appliquée à l'image, qui peut être ajouté aux architectures SSL de l'état de l'art. En augmentation la quantité d'information de l'entrée qui doit être conservé dans les représentations, puisqu'elles doivent pouvoir être projetées sur deux espaces, l'un invariant et l'autre équivariant aux augmentations, il améliore les performances de classification et ouvre la voie à des représentations plus générales et robustes. Enfin, nous avons systématiquement étudié les structures des représentations qui émergent des méthodes de discrimination d'instance. Nous avons empiriquement observé qu'elles étaient organisées d'une manière à favoriser la discrimination entre classes, avec une haute similarité intra-classe et une faible inter-classe, alors qu'elles sont apprises sur une tâche de discrimination d'instance. Bien que ce mécanisme ne soit pas complètement compris, ces observations peuvent aider à améliorer les méthodes SSL et soulever des questions de recherche sur la transférabilité de ces structures à d'autres tâches avales.

Dans le chapitre 3, je présente un modèle pour la fusion multisensorielle avec un mécanisme de compétition/fusion décentralisé basé sur les champs neuronaux dynamiques (DNF), qui s'appuie sur une représentation topologiquement organisée des espaces sensoriels. D'abord nous proposons un cadre intégré pour souligner les similarités entre les DNF et les autres algorithmes de prise de décision de la psychophysique à la robotique, et

illustrer la flexibilité de ses propriétés spatio-temporelles sur un ensemble de divers scénarii. Ensuite, nous utilisons les DNF pour modéliser la fusion audio-visuelle de participants humains à une expérience psychophysique de localisation auditive. Ils obtiennent des résultats similaires à l'état de l'art classique en modélisation, basé sur l'inférence bayésienne, tout en étant plus ancrés dans les substrats neuronaux. Cela ouvre la voie à une modélisation comportementale au niveau du participant, plutôt qu'à celui de la population, et pourrait être rendu compatible avec des mécanismes de perception active. De plus, couplé à l'apprentissage topologique de l'espace sensoriel, nous avons validé que les propriétés du modèle étaient qualitativement préservées, jusqu'à ce que la topologie soit trop distordue. De manière intéressante, nous avons montré que l'importance d'un signal dans la fusion pouvait s'appuyer sur la densité relative de la topologie apprise, qui est reliée à la densité de senseurs. Ce mécanisme doit encore être validé sur nos données psychophysique. Cependant, cela pourrait être utilisé pour une évaluation autonome et en ligne de la pertinence d'un senseur dans la perception multimodale d'agents artificiels.

Le chapitre 4 est dédié à la dimension temporelle de l'apprentissage. Dans une première contribution nous avons proposé un mécanisme d'apprentissage actif, qui augmente la fréquence des données/tâches mal classifiées dans le jeu de données d'apprentissage pour les prochaines époques. Bien que relativement simple, ce mécanisme améliore les performances de classification lorsque les données sont fortement biaisées en aidant le modèle à éviter d'apprendre les corrélations fallacieuses. Un mécanisme similaire, couplé à une forme d'apprentissage multi-tâches, améliore les performances d'apprentissage bout en bout de multiplication. De manière intéressante, il induit d'une certaine manière un transfert de l'apprentissage des sous-opérations à celle de la multiplication globale. La seconde contribution s'inscrit dans le contexte de l'apprentissage incrémental de classes non supervisé. Nous avons proposé d'augmenter un auto-encodeur variationnel, utilisant une mixture de gaussiennes pour modéliser les différents composants de la distribution, avec des tests statistiques pour détecter de nouveaux composants ou reconnaître des précédents dans le flux de données. Cela permet de découvrir automatiquement le nombre de catégories présentes tout en conservant des performances de classification raisonnables.