

DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **27 mars 2025**

Nom de famille et prénom de l'auteur. e : **Monsieur Alexandre DEVILLERS**

Titre de la thèse : Structuration de représentations visuelles pour améliorer la généralisation en apprentissage auto-supervisé

Résumé



L'apprentissage de représentations est devenu un pilier central de l'intelligence artificielle moderne, jouant un rôle clé dans les avancées récentes de domaines tels que la vision par ordinateur et le traitement du langage naturel. Avec l'intérêt croissant pour l'apprentissage auto-supervisé, où les modèles apprennent de manière autonome à partir de données brutes sans supervision humaine, l'apprentissage des représentations est devenu encore plus important. Il permet de donner un sens à ces données brutes en extrayant des caractéristiques pertinentes. De plus, ce cadre autonome favorise l'apprentissage de représentations plus générales grâce à l'absence d'une labellisation spécifique — ce qui les rend agnostiques aux tâches aval — tout en tirant parti des grandes quantités de données brutes disponibles. Néanmoins, la difficulté réside dans la recherche d'un signal de supervision, accessible uniquement à partir des données d'entrée, mais suffisamment pertinent pour structurer des représentations générales offrant de bonnes performances sur les tâches aval. Les méthodes récentes d'apprentissage auto-supervisé de représentations visuelles utilisent comme supervision des tâches prétextes de discrimination d'instances, qui ont démontré un fort potentiel pour générer des représentations riches, réutilisables et transférables à un large éventail de tâches aval, surpassant parfois même les approches supervisées. Le principe de discrimination d'instances repose sur l'idée que des entrées similaires doivent être projetées vers des points similaires dans l'espace des représentations. En pratique, cela est généralement réalisé grâce à une architecture siamoise, qui traite deux vues augmentées

d'une même entrée à travers des réseaux identiques. Ces vues sont générées de manière auto-supervisée en appliquant des transformations — aussi appelées augmentations — sur une même image, produisant des paires sémantiquement similaires mais visuellement distinctes. La tâche prétexte vise ensuite à aligner les sorties des deux vues, encourageant le réseau à construire des représentations invariantes aux augmentations, mettant ainsi l'accent sur l'apprentissage des motifs visuels partagés entre les vues. Cet apprentissage, fondé sur une tâche prétexte visant à capturer une invariance, se distingue des méthodes historiques, telles que celles basées sur la reconstruction, qui cherchent à reconstruire une image à partir de sa représentation. En effet, la discrimination d'instances se focalise sur un objectif orienté structure, et le succès de ces approches met en évidence l'importance d'explorer les propriétés structurelles des représentations apprises, non pas uniquement comme un outil pratique pour concevoir des tâches prétextes, mais comme une façon directe pour améliorer leur qualité. Cette thèse s'inscrit dans cette perspective en explorant comment la structure des représentations — notamment l'invariance, la sensibilité et l'équivariance — peut être exploitée pour améliorer la généralisation dans l'apprentissage des représentations visuelles. Cette problématique est abordée à travers des sous-questions spécifiques, chacune liée à une contribution de la thèse. Ces sous-questions examinent la structure via divers moyens, tels que la modification de la distribution des données, l'ajout d'aspects variationnels, l'utilisation de l'équivariance, ou encore la corrélations entre performances et sous-propriétés structurelles. Ces travaux ont ainsi permis de mettre en lumière que la structure des représentations joue un rôle important dans la généralisation et montrent donc qu'elle constitue donc un levier efficace pour améliorer les performances.

Mots-clés :

Apprentissage de représentation, Apprentissage auto-supervisé, Apprentissage profond, Discrimination d'instances, Structuration d'espace latent, Invariance et équivariance,