Laboratoire de Mathématiques et Informatique pour la Complexité et les Systèmes

MICS

Présente

L'AVIS DE SOUTENANCE

de **Monsieur Léo Fillioux**

Laboratoire MICS, CentraleSupélec, Université Paris Saclay, qui soutiendra publiquement ses travaux de thèse de doctorat intitulés :

«Foundation models for computer vision: adaptation, robustness and application to medical imaging.»

Sous la Direction de Madame Maria Vakalopoulou, Monsieur Stergios Christodoulidis et Monsieur Paul-Henry Cournède.

Le Lundi 29 Juin 2026 à 14h

À l'école CentraleSupélec, dans le **Théâtre Rousseau** – Salle E.070, Bâtiment Eiffel.

Membres du jury :

Ipek Oguz, Rapportrice & Examinatrice, Vanderbilt University

Dimitris Samaras, Rapporteur & Examinateur, Stony Brook University

Spyros Gidaris, Examinateur, Valeo.ai

Caroline Petitjean, Examinatrice, Université de Rouen

Nicolas Thome, Examinateur, Sorbonne Université

Gül Varol, Examinatrice, École des Ponts ParisTech

Résumé :

Au cours des dernières années, les modèles de fondation ont fondamentalement transformé le paysage de la recherche dans le deep learning, en particulier dans le domaine de la vision par ordinateur, en repoussant les limites de l'état de l'art et en améliorant considérablement la capacité de généralisation des modèles. Cela a donné lieu à de nouveaux défis, dont certains sont inhérents aux modèles de fondation, tandis que d'autres sont amplifiés par l'ampleur de l'adoption généralisée de ces modèles.

L'adaptation étudie comment modifier efficacement les poids de modèles préentrainés pour s'adapter à des changements de distribution, tout en maintenant la généralisabilité du modèle de base. Ceci est crucial pour étendre les domaines dans lesquels ces puissants modèles peuvent être utilisés, même lorsque très peu de données sont disponibles. Dans la première partie de cette thèse, nous introduisons deux méthodes pour l'adaptation sous différents scénarios. La première porte sur l'adaptation de Transformers dans un contexte à faible nombre d'images et permet de modéliser naturellement l'incertitude dans la prédiction. De plus, nous présentons également une méthode qui met l'accent sur la calibration pour l'adaptation de modèles de vision-langage. La robustesse est définie comme la capacité d'un modèle à maintenir une performance stable sous différents scénarios, tels que les changements de distribution. Bien qu'elle soit essentielle pour les systèmes de deep learning, les performances élevées des modèles de fondation ont accéléré leur déploiement dans des applications sensibles, rendant la robustesse d'autant plus cruciale. Nous analysons comment les modèles de fondation de vision performent dans différents scénarios réalistes dans le cadre de la prédiction conformale, permettant aux modèles de prédire un ensemble de prédictions, plutôt qu'une prédiction unique. Nous proposons également un cadre plus réaliste pour la généralisation inter-domaines, avec une distribution de données déséquilibrée. Ce contexte compliqué a un impact négatif sur les méthodes existantes, auquel nous proposons de remédier par une solution dédiée. Enfin, nous présentons, à notre connaissance, le premier modèle de fondation et le premier benchmark pour les organoïdes, un modèle tumoral tridimensionnel, tous deux contribuant ainsi à accélérer la recherche en matière de traitements oncologiques.

Abstract:

Foundation models have fundamentally reshaped the landscape of deep learning in recent years, particularly within computer vision, by pushing the state of the art and significantly enhancing model generalizability. This has given rise to new challenges and questions, some of which are inherent to foundation models, while others are augmented by the scale and the widespread adoption of such models. Adaptation studies how to efficiently update the weights of a pretrained model to adjust to a distribution shift, while still maintaining the generalizability of the original model. This is crucial for extending the domains in which such powerful models can be used, even when very limited data is available. In the first part of this thesis, we introduce two methods for adaptation under different scenarios. The former addresses the adaptation of vision Transformers under a few-shot regime, and enables natural modeling of the prediction uncertainty. Furthermore, we also propose a method that puts emphasis on calibration for test-time prompt tuning of vision-language models. Robustness defines a model's capacity to maintain stable performance under varying scenarios, such as distribution shifts. While critical for any deep learning system, the high performance of foundation models has accelerated their deployment in sensitive applications, making robustness all the more crucial. We analyze how vision foundation models perform across different real-world settings under the conformal prediction framework, allowing models to predict a set of predictions rather than point predictions. We also introduce a more realistic framework for domain generalization, featuring an imbalanced data distribution. This challenging setting negatively impacts existing methods. We propose a novel framework for addressing this important issue. Finally, we introduce, to the best of our knowledge, the first foundation model and benchmark for organoids, a groundbreaking three-dimensional tumor model, thereby participating in the acceleration of oncology treatment research.