



**Laboratoire de Mathématiques et Informatique pour la Complexité et les
Systèmes
MICS**

Présente

L'AVIS DE SOUTENANCE

De **Monsieur Marin Scalbert**

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

*“Mind the domain gap: Improving Deep Learning models generaliza7on across
diverse visual domains”*

Sous la Direction de Monsieur Marc AIGUIER, Madame Maria VAKALOPOULOU et le co-
encadrement de Florent COUZINIÉ-DEVY

Le jeudi 28 novembre 2024 à 13h00

À CentraleSupélec, Bâtiment Bouygues, **Amphithéâtre sc.071**.

Membres du jury :

Rémi FLAMARY, Rapporteur & Examineur, École Polytechnique
Amaury HABRARD, Rapporteur & Examineur, Université Jean Monnet
Céline HUDELOT, Examinatrice, CentraleSupélec
Dimitris SAMARAS, Examineur, Stony Brook university
Christian DESROSIERS, Examineur, École de Technologie Supérieure

Résumé:

Au cours des deux dernières décennies, la disponibilité des données et l'amélioration de la puissance de calcul ont permis au Deep Learning de se développer dans de nombreux domaines tels que la conduite autonome, la télédétection, l'imagerie médicale, etc. Par ailleurs, il présente également un potentiel pour faire progresser la pathologie digitale. Cependant, son utilisation clinique reste difficile en raison de la variabilité des protocoles utilisés pour l'acquisition des données de pathologie. Cette variabilité entraîne un "décalage de domaine" qui se produit lorsque les distributions de données d'entraînement et de test diffèrent, ce qui peut s'accompagner d'une baisse de performance des modèles sur les données de test. Ce phénomène ne se limite pas à la pathologie digitale. Cette thèse vise donc à améliorer la généralisation des modèles de Deep Learning face au décalage de domaine, en abordant des problèmes de complexité croissante et en proposant de nouvelles solutions appliquées à la fois à des données de pathologie digitale et/ou de vision par ordinateur plus classiques. Le Chapitre 2 aborde le problème d'adaptation de domaine non supervisée multi-sources en introduisant une méthode appelée CMSDA. Cette méthode s'appuie sur des avancées récentes en apprentissage contrastif et semi-supervisé, pour mieux aligner les distributions conditionnelles conditionnées à la classe pour les domaines sources et adapter progressivement le modèle au domaine cible. Ce travail montre que l'apprentissage contrastif supervisé peut aligner efficacement les distributions sources et conduit à des gains de performance significatifs pour des domaines cibles avec un décalage de domaine important. Le Chapitre 3 explore la généralisation de domaine multi-sources, un scénario plus complexe où les données du domaine cible sont inaccessibles pendant l'entraînement. Une technique d'augmentation de données, applicable au moment de l'inférence et basée sur un modèle de traduction d'image en image avec des stratégies d'ensembling utilisant des discriminateurs de domaine, est proposée. Par ailleurs, nous démontrons que ce modèle, entraîné sur des données différentes de la tâche finale (e.g., provenant de différents organes, cancers, etc.), peut être utilisé pour augmenter des données et entraîner des modèles invariants aux protocoles. Le Chapitre 4 aborde le problème de généralisation de domaine non supervisée, visant à développer des modèles auto-supervisés invariants au domaine qui, une fois affinés, généralisent à des domaines cibles jusqu'alors non rencontrés. Inspirés par les méthodes précédentes basées sur l'apprentissage contrastif, nous émettons l'hypothèse que la suppression de la variabilité de style au sein d'un lot d'exemples peut permettre d'obtenir des représentations auto-supervisées invariantes au domaine sans nécessiter l'utilisation de labels de domaine. Pour valider cette hypothèse, nous proposons une méthode dénommée BSS, pour standardiser les styles d'image au sein d'un lot. BSS s'intègre facilement aux méthodes d'apprentissage auto-supervisé existantes et offre des avantages significatifs par rapport aux approches de généralisation de domaine non supervisée précédentes. BSS favorise l'invariance par domaine sans l'utilisation de labels de domaine ni d'architecture spécifique, conduisant à une meilleure simplicité et flexibilité. Enfin, le Chapitre 5 se focalise sur le problème d'apprentissage par instances multiples (MIL), particulièrement adapté à la résolution de tâches globales au niveau des lames de pathologie digitale. Compte tenu des défis liés au chargement des tuiles en mémoire, à leur augmentation et aux calculs de leurs représentations lors de l'entraînement de modèles de MIL, un nouveau modèle génératif est proposé pour permettre l'augmentation des représentations de tuiles. Cette méthode s'avère rapide et offre surtout un meilleur contrôle pour l'augmentation des représentations comparée aux méthodes précédentes.

Abstract:

In the last two decades, advancements in data availability and computing power have propelled Deep Learning into many fields such as autonomous driving, remote sensing, medical imaging, etc. It has a real potential for advancing Digital Pathology. However, using these models in clinical practice is challenging due to protocol variability arising when acquiring pathology data. Such variability results in a phenomenon known as domain shift which occurs when training and testing data distributions differ. Domain shift often results in a drop in model generalization. This phenomenon is not restricted to digital pathology but affects nearly all fields where Deep Learning is applied. Thus, this thesis aims to improve the generalization of Deep Learning models in the context of domain shift, addressing problems of increasing complexity, proposing new solutions, and applying them to both digital pathology and conventional computer vision data. Chapter 2 addresses the problem of Multi-Source Unsupervised Domain Adaptation (MS-UDA) by introducing a new method called Contrastive Multi-Source Domain Adaptation (CMSDA). This method leverages recent advances in Contrastive learning and Semi-Supervised Learning, to better align class-conditional distributions across source domains and progressively adapt the model to the target domain. This work demonstrates that supervised contrastive learning can effectively align source class-conditional distributions, resulting in significant performance gains for target domains experiencing substantial domain shift. Chapter 3 explores Multi-Source Domain Generalization (MS-DG), a more challenging scenario where target domain data is inaccessible during training. A novel Test-Time Augmentation (TTA) technique, based on image-to-image translation with tailored ensembling strategies using domain discriminators, is proposed. Additionally, it is shown that the same image-to-image translation model, trained on unrelated histopathology data (e.g., from different organs, cancers, etc.), can serve as a strong data augmentation technique for developing protocol-invariant models. Chapter 4 addresses the Unsupervised Domain Generalization (UDG) problem, aiming to develop domain-invariant SSL models that, when fine-tuned, generalize well to unseen target domains. Inspired by contrastive-based UDG methods, we hypothesize that removing style variability within a batch of examples can achieve domain-invariant SSL representations without requiring domain labels. To test this hypothesis, we propose Batch Styles Standardization (BSS) to standardize image styles within a batch. BSS integrates seamlessly into existing SSL methods, offering significant advantages over previous UDG approaches by enforcing domain invariance without domain labels or specific architecture components, and by providing simplicity and flexibility. Finally, Chapter 5 focuses on the Multiple Instance Learning (MIL) problem, particularly suited to solving slide-level tasks in digital pathology. Given the computational challenges of loading patches, performing patch-level data augmentation, and extracting features when training deep MIL models, a novel generative model is introduced to provide faster and more controllable data augmentation at the features level than previous methods.