

Titre : Apprentissage profond diversifié pour une meilleure généralisation à des domaines inconnus

Résumé de la thèse :

De plus en plus d'applications embarquées, confrontées à des environnements diversifiés, changeants et non-contrôlés, nécessitent un très haut degré d'adaptabilité et de fortes capacités d'analyse pour mener leur tâche à bien. La pré-programmation d'actions n'est plus suffisante pour effectuer ces nouveaux types de tâches, et est donc en train d'être remplacée par un paradigme prometteur : l'apprentissage profond.

Cependant, les réseaux de neurones sont vulnérables à des changements de distribution (ou domaine) de données entre l'apprentissage et l'utilisation. Ce défaut apparent empêche un déploiement fréquent dans des produits embarqués. De plus, il est impossible de récolter et d'ajouter au jeu de données d'entraînement suffisamment de données pour prévenir tous les changements de distribution possibles, à cause de leur importante diversité.

L'origine de cette vulnérabilité se trouve en partie dans le comportement d'apprentissage de raccourcis des réseaux de neurones profonds: seulement les motifs prédictifs les plus efficaces, aussi fallacieux soient-ils, sont appris, les autres sont fortement ignorés. Confronté à une nouvelle distribution, dans laquelle les motifs prédictifs sont partiellement différents, un réseau qui aura appris un ensemble limité de caractéristiques sera statistiquement moins capable de prendre une décision correcte.

En collaboration avec Thales Land and Air Systems, l'objectif de ce travail est donc de développer des solutions pour atténuer la chute de performance des réseaux de neurones lors d'un changement de domaine. Ce travail comprend deux contributions principales.

Premièrement, nous proposons une nouvelle architecture générative de réseaux de neurones qui permet de limiter l'ampleur de l'apprentissage de raccourcis dans un contexte sous-exploré. Les précédents travaux de l'état de l'art reposaient explicitement sur des éléments particuliers, dont les raccourcis ne sont pas alignés avec la majorité des autres données, et augmentaient leur importance dans la procédure d'apprentissage. Dans cette contribution, nous avons montré que ces éléments n'étaient pas nécessaires pour éviter les raccourcis, grâce à des jeux de données synthétiques construits dans cet objectif, et avons de plus validé l'efficacité de notre approche sur un jeu de données réaliste.

Dans un deuxième temps, le travail présenté ici se concentre sur des situations de changement de domaines plus générales, dans lesquelles un unique domaine est disponible pendant l'apprentissage. L'adaptation au moment du test aux données a émergé comme un ensemble de stratégies prometteur pour efficacement améliorer les performances quand de nouveaux domaines sont rencontrés lors de l'utilisation. Cependant, ces méthodes s'appuient sur des modèles entraînés avec la procédure standard, qui ne permet pas l'apprentissage de la totalité des motifs prédictifs, comme nous l'avons vu. Nous proposons donc une modification de la procédure d'entraînement complémentaire avec l'adaptation au moment du test. Notre méthode permet l'apprentissage des motifs naturellement appris ainsi que des motifs habituellement ignorés, et par conséquent, permet une adaptation plus approfondie. Sur la base de plusieurs expériences, nous montrons que notre approche améliore la qualité des prédictions sur des domaines jamais rencontrés au moment de l'apprentissage.

Mots-clés : apprentissage profond, généralisation de domaine, adaptation au moment du test, apprentissage des raccourcis, apprentissage débiaisé.

Title : Deep diversity learning for better generalization to unseen domains

Abstract of the thesis :

A growing number of embedded applications, confronted with diversified, shifting, and uncontrolled environments, require an increased degree of adaptability and analysis capabilities to fulfill their task. Pre-programmed actions are no longer able to deal with these new sets of tasks and are therefore being replaced by a promising paradigm: deep learning.

However, deep neural networks are susceptible to data distribution shifts occurring between training and use. This apparent flaw prevents the widespread deployment of deep networks in embedded products. Furthermore, it is impossible to gather and add enough data to the training set to cover all possible shifts due to their tremendous diversity.

The origin of this vulnerability lies partly in the shortcut-learning behaviour of deep networks: they learn only the most efficient patterns, no matter how spurious, and completely disregard the others. Confronted with a new distribution in which the predictive patterns are partially different, a network that learned a limited subset of features would be less likely to be able to make a proper decision.\\

In collaboration with Thales Land and Air Systems, this work therefore aims to develop solutions to mitigate the domain shifts performance drop in deep networks. This work has two main contributions.

Firstly, we propose a new deep generative architecture that mitigates the shortcut-learning behavior in an under-explored setting. Previous state-of-the-art works relied explicitly on shortcut-contrary samples and increased their importance in the training procedure. In this work, we demonstrated on several different synthetic benchmarks that such particular samples were not needed for shortcut avoidance and further confirmed the effectiveness of our approach on a realistic benchmark.\\

Secondly, the work presented here focuses on more general and realistic domain shift situations in which only a single domain is available during training. Test-time adaptation to the encountered data has emerged as a promising set of strategies to efficiently increase performance when facing new domains at use time. They, however, rely on a model trained with the standard procedure, which, as previously stated, will ignore some predictive patterns. We propose a training-time approach complementary to test-time adaptation. Our method seeks to learn both the patterns learned through standard training and the normally "hidden" ones, enabling a more thorough test-time adaptation. Based on extensive experiments, we show that our approach improves the quality of predictions on domains unseen at training-time.

Keywords : deep learning, domain generalization, test-time adaptation, shortcut-learning, bias avoidance, covariance shifts.