



utbm
université de technologie
Belfort-Montbéliard

UBFC
UNIVERSITÉ
BOURGOGNE FRANCHE-COMTÉ



SPIM

école doctorale **sciences pour l'ingénieur et microtechniques**

Titre de la thèse : Génération de vérités terrain incomplètes et fiables en utilisant des modèles profonds pour l'apprentissage en ligne

Laboratoire d'accueil : Connaissance et Intelligence Artificielle Distribuées (CIAD - UR7533)

Spécialité du doctorat préparé : Informatique

Mots-clefs : apprentissage en ligne, apprentissage profond (CNN, GAN, Transformers), régression, prise de décision adaptative, jeu de données de fine-tuning, modèle de vérification, données évolutives, fonctions de perte, apprentissage basé sur les espaces de projection, connaissance contextuelle.

Descriptif détaillé de la thèse

L'apprentissage en ligne, également connu sous le nom d'apprentissage itératif, continu ou adaptatif, est une approche d'apprentissage automatique dans laquelle un modèle est entraîné à partir de données en constante évolution, tout en s'adaptant aux nouvelles informations au fur et à mesure de leur arrivée. Contrairement à l'apprentissage hors ligne traditionnel, où un modèle est entraîné sur un ensemble de données fixe, l'apprentissage en ligne permet d'améliorer les performances des modèles dans des environnements changeants et de prendre des décisions adaptatives en temps réel, offrant ainsi des solutions plus agiles et réactives aux problèmes du monde réel. Cette approche trouve des applications dans de nombreux domaines couvrant la robotique, les véhicules autonomes, la vidéosurveillance (estimation de la profondeur, estimation du flux optique, segmentation sémantique, génération d'images, classification, régression, etc.).

L'apprentissage en ligne permet de résoudre la problématique liée aux données d'apprentissage, généralement insuffisantes pour couvrir ou prédire tous les cas possibles dans la réalisation d'une tâche. Cependant, la difficulté de cette méthode est l'indisponibilité des annotations pour les données d'entrée. Certaines approches proposent de continuer l'apprentissage en ligne en se basant sur la prédiction du modèle lui-même (modèle pré-entraîné). Le problème est que les prédictions, utilisées comme vérité terrain en fonction d'un degré de confiance, impactent le processus de fine-tuning du modèle d'apprentissage en ligne pour aboutir finalement à des prédictions moins précises sur de nouvelles données.

Dans ce projet de thèse, de nouvelles techniques d'apprentissage efficaces seront développées pour relever ces défis. L'approche que nous proposons consiste tout d'abord à construire de nouveaux modèles profonds (CNN, GAN, Transformers) qui seront entraînés pour vérifier les sorties du modèle d'apprentissage en ligne. Cette première étape permet de collecter une base de fine-tuning composée d'entrées et de vérités terrain ne comportant que les prédictions les plus fiables (les vérités terrain seront certainement incomplètes). Cette base servira dans une seconde étape à fine-tuner le modèle d'apprentissage en ligne. Ce processus est répété autant que nécessaire (plusieurs stratégies à prévoir) pour pouvoir améliorer continuellement les performances du modèle en ligne. La méthodologie de développement des modèles envisagés tiendra compte de plusieurs niveaux d'analyse : modèles avec différentes stratégies d'apprentissage (cascade, collaboration, compétition, apprentissage à base d'espaces de projection, etc.), architecture du modèle, fonction de perte, etc.

Pour la tâche de segmentation sémantique, par exemple, l'idée consiste à intégrer un modèle de vérification pour prendre en entrée la sortie prédite du modèle de segmentation en ligne. Le modèle de vérification génère une carte d'erreur sémantique pour estimer les prédictions erronées. Ensuite, en comparant cette carte d'erreur sémantique (vue comme un masque) avec la carte sémantique prédite, nous pouvons localiser les prédictions erronées et ainsi corriger la sortie du modèle de segmentation en ligne. En suivant cette approche, nous construisons une base de données constituée d'images et de cartes sémantiques corrigées, ne contenant que les prédictions les plus fiables, qui serviront de vérités de terrain pour affiner le modèle de segmentation en ligne, contribuant ainsi à améliorer continuellement sa performance.

En ce qui concerne l'approche d'entraînement du modèle de vérification, nous collecterons des données publiques, comprenant des images ainsi que leurs cartes sémantiques de vérité terrain correspondantes, représentant des segments sémantiques sans erreur. Ensuite, nous utilisons des modèles de segmentation sémantique issus de la littérature, qui prennent en entrée les images pour

générer leurs segments sémantiques respectifs, qui contiennent forcément des erreurs de segmentation. L'étape suivante consiste à créer des masques indiquant les erreurs de prédiction en appliquant une correspondance basée sur les segments entre les cartes sémantiques de vérité terrain et celles prédites par les modèles de segmentation de la littérature. De cette manière, nous constituons un ensemble de données comprenant des paires d'images : les cartes sémantiques prédites et leurs masques associés d'erreurs de prédiction, qui serviront de vérité terrain pour entraîner notre modèle de vérification. Ce dernier fonctionnera comme un discriminateur-régresseur, estimant l'erreur de prédiction de la segmentation sémantique pixel par pixel. Une fois entraîné, il sera intégré dans le processus d'apprentissage en ligne afin d'estimer les erreurs de prédiction du modèle de segmentation sémantique en ligne.

Il est à noter que la fonction de perte privilégiée est celle de l'entropie croisée binaire (Binary Cross Entropy Loss), l'objectif étant de minimiser l'écart pixel par pixel entre les prédictions et les vérités terrain (masques d'erreurs de prédiction). Cependant, nous envisageons de proposer de nouvelles fonctions de perte plus adaptées. Nous envisageons d'enrichir le modèle de vérification avec des connaissances contextuelles (météo, environnement, ...) lors de l'entraînement. Ces connaissances peuvent être directes ou générées à partir de représentation et raisonnement ontologiques.

Cette méthodologie sera développée et adaptée pour d'autres tâches de vision (estimation de la profondeur, estimation du flux optique, détection d'objets, segmentation panoptique, amélioration d'images...) avec des données de type image (RGB, thermique, événementielle) ou provenant d'autres modalités (LIDAR, RADAR).

Références bibliographiques

- [1] Volpi, Riccardo, et al. "On the road to online adaptation for semantic image segmentation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
- [2] Mai, Zheda, et al. "Online continual learning in image classification: An empirical survey." Neurocomputing 469 (2022): 28-51.
- [3] Yu, Xiaofan, et al. "SCALE: Online Self-Supervised Lifelong Learning without Prior Knowledge." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
- [4] Ghunaim, Yasir, et al. "Real-time evaluation in online continual learning: A new hope." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
- [5] Saha, Gobinda, and Kaushik Roy. "Saliency guided experience packing for replay in continual learning." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2023.
- [6] Du, Fei, et al. "Efficient Perturbation Inference and Expandable Network for continual learning." Neural Networks 159 (2023): 97-106.
- [7] Maltoni, Davide, and Vincenzo Lomonaco. "Continuous learning in single-incremental-task scenarios." Neural Networks 116 (2019): 56-73.
- [8] Lesort, Timothée, et al. "Continual learning for robotics: Definition, framework, learning strategies, opportunities and challenges." Information fusion 58 (2020): 52-68.

Profil demandé

- Un diplôme de master ou équivalent en informatique, vision par ordinateur, apprentissage machine, robotique ou autres spécialités connexes.
- Un très bon niveau de maîtrise en programmation orientée-objet (C++, Python) et dans les outils d'apprentissage machine/profond (PyTorch, TensorFlow, Matlab) est nécessaire..
- Une bonne maîtrise de l'anglais (oral et écrit) est exigée.
- Des connaissances de l'environnement ROS seront appréciées

Critères de sélection préférés

- Adéquation de la formation avec le sujet de thèse et niveau des résultats académiques
- Fortes compétences en programmation et en apprentissage automatique/deep learning, avec une expérience pratique dans l'utilisation de frameworks tels que PyTorch et TensorFlow.
- Vision et projection sur le sujet de thèse
- Fortes compétences en communication et en rédaction scientifique.

Caractéristiques personnelles

- Rigueur scientifique et curiosité intellectuelle.
- Capacité à travailler en équipe de manière autonome.
- Aptitude à communiquer efficacement ses idées et ses résultats, tant à l'écrit qu'à l'oral.
- Motivation et engagement envers la réalisation des objectifs de la thèse.

Financement : MESRI Etablissement

Dossier à envoyer pour le **7 juin 2024**

Début du contrat : septembre/octobre 2024

Salaire mensuel brut : 1975€

Direction de la thèse :

Prof. Yassine RUICHEK (yassine.ruichek@utbm.fr)

Encadrement de la thèse : co-directeur(s) et co-encadrant(s)

Abderrazak CHAHI (abderrazak.chahi@utbm.fr) / co-encadrant

Les candidats sont invités à soumettre leur candidature aux encadrants de la thèse. Celle-ci doit comprendre les documents suivants :

- CV
- Lettre de motivation
- Relevés de notes
- Lettres de référence

Thesis title: Generating incomplete and reliable ground truths for online learning using deep models

Host Laboratory : Connaissance et Intelligence Artificielle Distribuées (CIAD - UR7533)

Speciality : Computer Science

Keywords: online learning, deep learning (CNN, GAN, Transformers), regression, adaptive decision-making, fine-tuning dataset, verification model, evolving data, loss functions, learning based on projection spaces, contextual knowledge

Job description

Online learning, also known as iterative, continuous, or adaptive learning, is a machine learning approach in which a model is trained on continuously evolving data, adapting to new information as it becomes available. Unlike traditional offline learning, which relies on a fixed dataset, online learning allows for improved model performance in dynamic environments and facilitates real-time adaptive decision-making. Consequently, it provides more agile and responsive solutions to real-world problems. This approach finds applications in various fields, including robotics, autonomous vehicles, and video surveillance (such as depth estimation, optical flow estimation, semantic segmentation, image generation, classification, regression, etc.).

Online learning helps address the issue of inadequate training data, which is usually insufficient to cover or predict all possible cases in performing a task. However, a significant challenge with this method is the unavailability of annotations for input data. Some approaches suggest continuing online learning based on the model's own predictions (pre-trained model). However, relying on these predictions as ground truth, based on a confidence level, can impact the fine-tuning process of the online learning model, resulting in less accurate predictions on new data.

In this thesis project, new effective learning techniques will be developed to address these challenges. Our proposed approach involves first building new deep models (such as CNNs, GANs, and Transformers) that will be trained to verify the outputs of the online learning model. This initial step enables the collection of a fine-tuning dataset composed of inputs and ground truths containing only the most reliable predictions (although the ground truths will likely be incomplete). In a second step, this dataset will be used to fine-tune the online learning model. This process will be repeated as necessary (with several strategies to be considered) to continuously improve the model's online performance. The methodology for developing the envisaged models will consider several levels of analysis, including models with different learning strategies (such as cascade, collaboration, competition, learning based on projection spaces, etc.), model architecture, loss function, and more.

For the semantic segmentation task, for example, the idea is to integrate a verification model to take as input the predicted output of the online segmentation model. The verification model generates a semantic error map to estimate erroneous predictions. Then, by comparing this semantic error map (seen as a mask) with the predicted semantic map, we can locate the erroneous predictions and thus correct the output of the online segmentation model. Following this approach, we build a database consisting of images and corrected semantic maps, containing only the most reliable predictions, which will serve as ground truths to refine the online segmentation model, thereby continuously improving its performance.

Regarding the training approach of the verification model, we will collect public data, including images and their corresponding ground truth semantic maps, representing error-free semantic segments. Then, we use semantic segmentation models from the literature, which take images as input to generate their respective semantic segments, which inevitably contain segmentation errors. The next step is to create masks indicating prediction errors by applying a segment-based correspondence between the ground truth semantic maps and those predicted by the models from the literature. In this way, we create a dataset comprising pairs of images: the predicted semantic maps and their associated prediction error masks, which will serve as ground truth to train our verification model. The latter will function as a discriminator-regressor, estimating the prediction error of semantic segmentation pixel by pixel. Once trained, it will be integrated into the online learning process to estimate prediction errors of the online semantic segmentation model.

It is important to note that the preferred loss function is binary cross-entropy loss, aimed at minimizing the pixel-by-pixel difference between predictions and ground truths (prediction error masks). However, we plan to propose new loss functions that are more suitable. We intend to enrich the

verification model with contextual knowledge (such as weather and environment) during training. This knowledge can be obtained directly or generated from ontological representation and reasoning.

This methodology will be further developed and adapted for other computer vision tasks, including depth estimation, optical flow estimation, object detection, panoptic segmentation, image enhancement, etc., using image data (RGB, thermal, event-based) or data from other modalities (LIDAR, RADAR).

Bibliography

[1] Volpi, Riccardo, et al. "On the road to online adaptation for semantic image segmentation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

[2] Mai, Zheda, et al. "Online continual learning in image classification: An empirical survey." Neurocomputing 469 (2022): 28-51.

[3] Yu, Xiaofan, et al. "SCALE: Online Self-Supervised Lifelong Learning without Prior Knowledge." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

[4] Ghunaim, Yasir, et al. "Real-time evaluation in online continual learning: A new hope." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

[5] Saha, Gobinda, and Kaushik Roy. "Saliency guided experience packing for replay in continual learning." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2023.

[6] Du, Fei, et al. "Efficient Perturbation Inference and Expandable Network for continual learning." Neural Networks 159 (2023): 97-106.

[7] Maltoni, Davide, and Vincenzo Lomonaco. "Continuous learning in single-incremental-task scenarios." Neural Networks 116 (2019): 56-73.

[8] Lesort, Timothée, et al. "Continual learning for robotics: Definition, framework, learning strategies, opportunities and challenges." Information fusion 58 (2020): 52-68.

Applicant profile

- Holder of Master's degree or equivalent in computer science, computer vision, machine/deep learning, or related fields.
- Advanced knowledge and practice in object-oriented programming (C++, Python) and machine/deep learning tools (PyTorch, TensorFlow, Matlab platforms) are required.
- Advanced level in English writing and speaking is required.
- Knowledge in ROS framework will be appreciated.

Preferred selection criteria:

- Adequacy background/thesis subject and level of academic results
- Advanced programming skills in machine learning and deep learning, with practical experience using frameworks such as PyTorch and TensorFlow.
- Thesis subject understanding and projection.
- Excellent communication and scientific writing skills.

Personal characteristics:

- Scientific rigor and curiosity.
- Ability to work independently within a team.
- Proficiency in effectively communicating ideas and results, both in writing and orally.
- Motivation and commitment to achieving the objectives of the thesis.

Funding Institution: MESRI

Documents to be sent by **June 7, 2024**

Start of contract: September/October 2024

Gross monthly salary: 2100€

Thesis director

Prof. Yassine RUICHEK (yassine.ruichek@utbm.fr)

Thesis supervision: co-director(s) and co-supervisor(s)

Abderrazak CHAHI (abderrazak.chahi@utbm.fr) / co-supervisor

Applicants are invited to submit their application to the thesis supervisors. It must include the following documents:

- CV
- Cover letter
- Transcripts
- Letters of reference