

**Titre de la thèse :** Apprentissage profond de l'espace de données pour des tâches en vision par ordinateur

**Laboratoire d'accueil :** Connaissance et Intelligence Artificielle Distribuées (CIAD) – <http://www.ciad-lab.fr>

Equipe Perception de l'Environnement et Navigation Autonome (PENA) – <https://epan-utbm.github.io/>

**Spécialité du doctorat préparé :** Informatique

**Mots-clefs :** génération d'images, classification, régression, apprentissage profond (CNN, GAN, Transformers), décomposition d'espace de données, apprentissage supervisé et non supervisé

**Descriptif détaillé de la thèse :**

La génération d'images est une technique qui consiste à générer une image à partir d'une ou plusieurs images appartenant ou non au même domaine. Elle a été introduite récemment, avec le développement de modèles d'apprentissage profond. Elle est utilisée dans de nombreuses applications couvrant la robotique, les véhicules autonomes, la vidéosurveillance (estimation de la profondeur, estimation du flux optique, segmentation sémantique, etc.).

Classiquement, les approches de génération d'images basées sur l'apprentissage profond sont conçues pour générer des images aussi proches que possible des images cibles prédéfinies. L'apprentissage de telles approches est principalement guidé par différentes fonctions de perte qui jugent la similarité entre l'image générée et sa vérité terrain au niveau de l'intensité du pixel. Cela conduit généralement à des images générées floues et à des détails haute fréquence corrompus. De plus, la dépendance au niveau des pixels est l'une des principales raisons du problème dit de dépendance de scène où la plupart des approches d'apprentissage profond échouent considérablement lorsqu'elles sont testées sur des images qui ne faisaient pas partie de leurs données d'entraînement. Dans ce projet de doctorat, de nouvelles approches d'apprentissage efficaces seront développées pour améliorer ce type de génération d'images et relever le défi des images non vues (unseen).

L'objectif est de développer de nouvelles approches d'apprentissage permettant de pallier la problématique liée aux données, généralement insuffisantes pour couvrir ou prédire tous les cas possibles dans la résolution d'un problème ou la réalisation d'une tâche. L'approche qui nous intéresse consiste à construire de nouveaux modèles profonds (CNN, GAN, Transformers) pour apprendre des espaces de données (au lieu des données elles-mêmes) dans un cadre supervisé et non supervisé, en considérant plusieurs niveaux d'analyse : décomposition (en composants singuliers par exemple), modèles avec différentes stratégies

(cascade, collaboration, compétition, etc.), architecture du modèle, fonction de perte, etc. Les recherches considéreront également d'autres tâches telles que la classification et la régression.

Cette thèse s'inscrit dans le cadre du projet DSDNet financé par le Conseil de la région Bourgogne-Franche-Comté (France). Ce projet lui-même s'inscrit dans le cadre du programme MOBILITECH regroupant des activités de recherche sur les véhicules autonomes et la robotique.

**Références bibliographiques :**

Mandal, M., & Vipparthi, S. K. (2020). Scene independency matters: An empirical study of scene dependent and scene independent evaluation for CNN-based change detection. IEEE Transactions on Intelligent Transportation Systems.

Giraldo, J. H., Javed, S., Werghi, N., & Bouwmans, T. (2021). Graph CNN for Moving Object Detection in Complex Environments from Unseen Videos. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 225-233).

Tezcan, O., Ishwar, P., & Konrad, J. (2020). BSUV-Net: A fully-convolutional neural network for background subtraction of unseen videos. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 2774-2783).

Mandal, M., Dhar, V., Mishra, A., Vipparthi, S. K., & Abdel-Mottaleb, M. (2020). 3DCD: Scene independent end-to-end spatiotemporal feature learning framework for change detection in unseen videos. IEEE Transactions on Image Processing, 30, 546-558.

Giraldo, J. H., & Bouwmans, T. (2020, October). Semi-supervised background subtraction of unseen videos: Minimization of the total variation of graph signals. In 2020 IEEE International Conference on Image Processing (ICIP) (pp. 3224-3228). IEEE.

Kas, M., Kajo, I., & Ruichek, Y. (2022). Coarse-to-fine SVD-GAN based framework for enhanced frame synthesis. Engineering Applications of Artificial Intelligence, 110, 104699.

**Profil demandé :**

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

**Financement :** Conseil de la Région Bourgogne Franche-Comté

**Dossier (CV, Notes, lettres de références, ...) à envoyer pour : 20 mai 2022**

**Début du contrat :** septembre/octobre 2022

**Direction / Codirection de la thèse :** Prof. Yassine RUICHEK (yassine.ruichek@utbm.fr)

**PhD title:** Data space deep learning for visual tasks

**Host laboratory:** Connaissance et Intelligence Artificielle Distribuées (CIAD) – <http://www.ciad-lab.fr>

**Host team:** Environment Perception and Autonomous Navigation (EPAN) – <https://epan-utbm.github.io/>

**Speciality of PhD:** Computer Science

**Keywords:** image generation, classification, regression, deep learning (CNN, GAN, Transformers), Data space decomposition, supervised and unsupervised learning

**Job description:**

Image generation is a technique that consists of generating an image from one or more images belonging to the same domain or not. It was introduced recently, with the development of deep learning models. It is used in many applications covering robotics, autonomous vehicles, video surveillance (depth estimation, optical flow estimation, semantic segmentation, etc.).

Classically, deep learning-based image generation approaches are designed to generate images as close as possible to predefined target images. The learning of such approaches is mainly guided by different loss functions that judge the similarity between the generated image and its ground truth at the pixel intensity level. This usually leads to blurry generated images and corrupted high-frequency details. Additionally, pixel-level dependency is one of the main reasons for the so-called scene dependency problem where most deep learning approaches fail drastically when tested on images that were not part of their training data. In this PhD project, effective learning approaches will be developed to improve this type of image generation and deal with the unseen images challenge.

The objective is to develop new learning approaches that overcome the problem related to data, which is generally insufficient to cover or predict all possible cases in solving a problem or performing a task. The approach we are interested in consists of building new deep models (CNN, GAN, Transformers) for learning data spaces (instead of data itself) in a supervised and unsupervised framework, considering several levels of analysis: decomposition (into singular components for example), models with different strategies (cascade, collaboration, competition, etc.), model architecture, loss function, etc.

The research study will also consider other tasks such as classification and regression.

This thesis is part of the DSDNet project funded by the Bourgogne-Franche-Comté regional Council (France). This part is itself part of the MOBILITECH program gathering research activities on autonomous vehicles and robotics.

**References:**

Mandal, M., & Vipparthi, S. K. (2020). Scene independency matters: An empirical study of scene dependent and scene independent evaluation for CNN-based change detection. IEEE Transactions on Intelligent Transportation Systems.

Giraldo, J. H., Javed, S., Werghi, N., & Bouwmans, T. (2021). Graph CNN for Moving Object Detection in Complex Environments from Unseen Videos. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 225-233).

Tezcan, O., Ishwar, P., & Konrad, J. (2020). BSUV-Net: A fully-convolutional neural network for background subtraction of unseen videos. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 2774-2783).

Mandal, M., Dhar, V., Mishra, A., Vipparthi, S. K., & Abdel-Mottaleb, M. (2020). 3DCD: Scene independent end-to-end spatiotemporal feature learning framework for change detection in unseen videos. IEEE Transactions on Image Processing, 30, 546-558.

Giraldo, J. H., & Bouwmans, T. (2020, October). Semi-supervised background subtraction of unseen videos: Minimization of the total variation of graph signals. In 2020 IEEE International Conference on Image Processing (ICIP) (pp. 3224-3228). IEEE.

Kas, M., Kajo, I., & Ruichek, Y. (2022). Coarse-to-fine SVD-GAN based framework for enhanced frame synthesis. Engineering Applications of Artificial Intelligence, 110, 104699.

**Candidate Profile:**

- Master degree in computer science, computer vision, machine learning, robotics or related field.
- Advanced knowledge and practice in object-oriented programming (C++, Python) and machine learning tools (deep learning platforms) are required.
- Advanced level in English writing and speaking is required.
- Knowledge in ROS framework will be appreciated.

**Financing Institution:** Bourgogne Franche-Comté Regional Council (France)

**Application (CV, scores, reference letters, ...) deadline:** 20 May 2022

**Start of contract:** September/October 2022

**Supervisor(s):** Prof. Yassine RUICHEK (yassine.ruichek@utbm.fr)