

# SOUTENANCE DE THÈSE

Unité de recherche Roberval  
Unité de recherche en mécanique, énergie et électricité

**Tuan Le**

sur le sujet :

Titre de thèse :

**Semi-Supervised Deep Learning for Robust Anomaly Detection in Smart Manufacturing Systems**

Le mardi 10 décembre 2024 à 10h  
à l'université de technologie de Compiègne  
Centre Pierre Guillaumat - amphi L.202

Devant le jury composé de :

- M<sup>me</sup> Anne Barros, professeure des universités, CentraleSupélec, université-Saclay, Gif-sur-Yvette, rapporteur
- M<sup>me</sup> Zohra Cherfi-Boulanger, professeure des universités, université de technologie de Compiègne, laboratoire Roberval, directrice de thèse
- M. Alexandre Durupt, maître de conférences, université de technologie de Compiègne, laboratoire Roberval, examinateur
- M<sup>me</sup> Mitra Fouladirad, professeure des universités, École Centrale Marseille, Aix Marseille Université, Marseille, rapporteur
- M. Antoine Grall, professeur des universités, université de technologie de Troyes, examinateur
- M. Abdessamad Kobi, professeur des universités, Polytech Angers, examinateur
- M<sup>me</sup> Amélie Ponchet-Durupt, enseignante chercheuse, université de technologie de Compiègne, laboratoire Roberval, examinatrice
- M. Hai-Canh Vu, enseignant chercheur, université de technologie de Compiègne, laboratoire Roberval, directeur de thèse

Invités :

- M. Nassim Boudaoud, maître de conférences, université de technologie de Compiègne
- M. Julien Moreau, maître de conférences, université de technologie de Compiègne

## **Preface**

Anomaly detection (AD) in monitoring manufacturing systems involves identifying data points or events that deviate from the system's normal state. Modern AD methods for monitoring smart systems must handle a large amount of data, often called Big Data, from various domains such as engine temperature, bearing vibration, and product images. To effectively mine data, AD methods based on deep learning are gaining attention due to their ability to extract important features from large datasets without manual effort. However, the performance of these models remains limited in practical applications due to factors such as data noise or new types of anomalies. Therefore, in this thesis, we propose different solutions to improve the performance of these deep learning-based AD methods and explore their applications across different data domains, including images and time series.

First, we demonstrate that if trained through a semi-supervised process (i.e., only trained on normal data), anomaly detection (AD) methods experience reduced performance if the training data contains noise or outliers. We propose a method to separate noise/outliers from the training data and train the model simultaneously. We experimented with this method on time-dependent data domains (time series) that contain information about the temperature and vibration of machines in industrial systems. The results show that this method outperforms current state-of-the-art models when the training data contains noise or outliers.

In the case of spatially dependent data, such as images, we have found that explicit noise or outliers, like blurs or discolored pixels, are not common in systems that use modern devices designed to capture images under optimal conditions. Therefore, we focus on detecting a different type of anomaly in this context, a logical anomaly involving incorrect combinations of normal objects. We propose a method for creating anomalies by altering the vectors representing the image's general information. We then train advanced models with these synthesized anomalies. The results indicate that incorporating these synthesized anomalies improves the model's performance by 1-2% without affecting computation time or model complexity.

In summary, the thesis outlines different methods to improve the performance of deep learning-based AD methods used in smart manufacturing systems. These methods aim to make AD models more resilient and reliable in real-world applications.

## **Préface**

La détection d'anomalies (AD) dans la surveillance des systèmes de fabrication consiste à identifier les points de données ou événements qui dévient de l'état normal du système. Les méthodes modernes de détection d'anomalies pour les systèmes intelligents doivent traiter un grand volume de données, souvent appelé Big Data, provenant de divers domaines tels que la température des moteurs, les vibrations des roulements et les images de produits. Pour exploiter efficacement ces données, les méthodes de détection d'anomalies basées sur l'apprentissage profond suscitent de plus en plus d'intérêt grâce à leur capacité à extraire des caractéristiques importantes de grands ensembles de données sans intervention manuelle. Cependant, les performances de ces modèles restent limitées dans les applications pratiques en raison de facteurs tels que le bruit des données ou l'apparition de nouveaux types d'anomalies. Ainsi, dans cette thèse, nous proposons différentes solutions pour améliorer les performances de ces méthodes basées sur l'apprentissage profond et explorons leurs applications dans divers domaines de données, notamment les images et les séries temporelles.

Tout d'abord, nous démontrons que, lorsqu'elles sont entraînées via un processus semi-supervisé (c'est-à-dire uniquement sur des données normales), les méthodes de détection d'anomalies (AD) voient leurs performances diminuer si les données d'entraînement contiennent du bruit ou des données aberrantes. Nous proposons une méthode permettant de séparer le bruit et les aberrations des données d'entraînement tout en formant le modèle simultanément. Nous avons expérimenté cette méthode sur des données temporelles (séries chronologiques) contenant des informations sur la température et les vibrations des machines dans les systèmes industriels. Les résultats montrent que cette méthode surpasse les modèles les plus avancés lorsque les données d'entraînement contiennent du bruit ou des aberrations.

Dans le cas des données à dépendance spatiale, comme les images, nous avons constaté que les bruits explicites ou les aberrations, tels que le flou ou les pixels décolorés, sont peu fréquents dans les systèmes utilisant des dispositifs modernes conçus pour capturer des images dans des conditions optimales. Par conséquent, nous nous concentrons sur la détection d'un autre type d'anomalie dans ce contexte : une anomalie logique impliquant des combinaisons incorrectes d'objets normaux. Nous proposons une méthode de création d'anomalies en modifiant les vecteurs représentant les informations générales de l'image. Nous entraînons ensuite des modèles avancés avec ces anomalies synthétisées. Les résultats montrent que l'intégration de ces anomalies synthétisées améliore les performances des modèles de 1 à 2 % sans affecter le temps de calcul ni la complexité du modèle.

En résumé, cette thèse présente différentes méthodes visant à améliorer les performances des techniques de détection d'anomalies basées sur l'apprentissage profond dans les systèmes de fabrication intelligents. Ces méthodes visent à rendre les modèles de détection d'anomalies plus robustes et fiables dans des applications réelles.