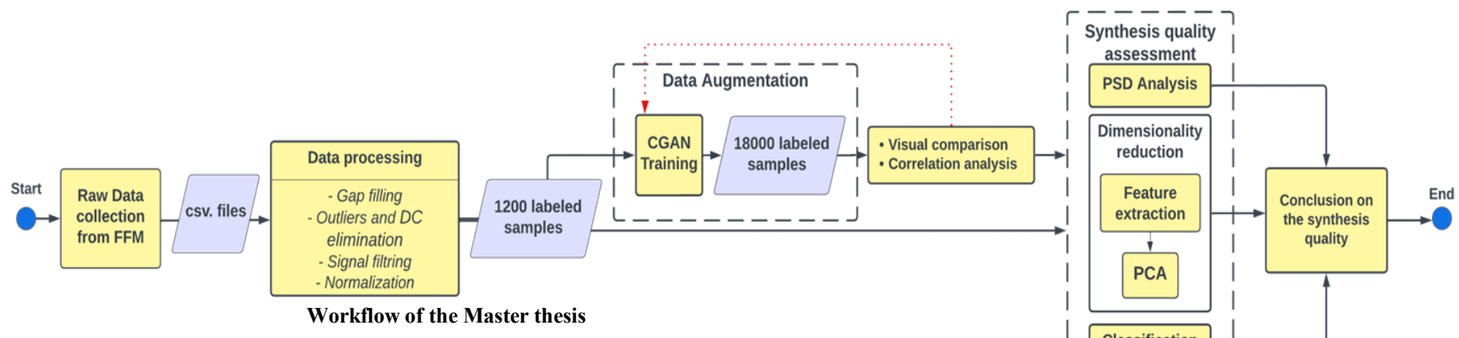
Masterarbeit

Data Augmentation in Vehicle-Track Scale Model for Fault Classification Performance Improvement of Machine Learning Algorithms

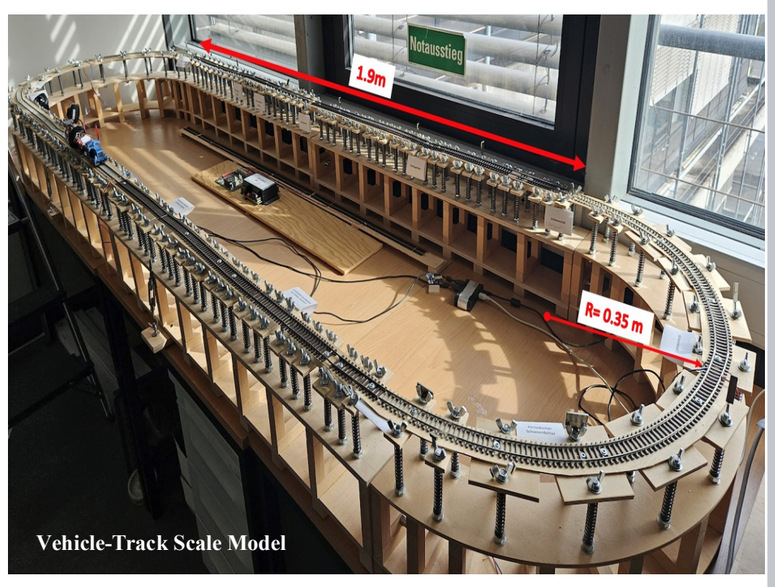
Railway provides a reliable mode of transportation that offers attractive solutions to energy and environmental concerns. However, maintaining railway infrastructure is a significant challenge. Condition Monitoring (CM) and Fault Detection and Isolation (FDI) methods can provide effective solutions, relying on Machine Learning (ML) to process large volumes of data. The training of ML algorithms is however problematic given their data-driven nature and the scarcity and imbalance frequently observed in the real-world data available for training these algorithms. Data Augmentation (DA) has emerged as a fundamental technique to address this problem through several techniques, of which Generative Adversarial Networks (GANs) are proving to be one of the most effective approaches.

This study investigates the application of DA using GAN in FDI for railway systems. The research focuses on data synthesis for a Vehicle-Track-Scale model (FFM) to create a balanced dataset for ML applications on fault classification. The process involves collecting and processing data from the physical model to obtain real samples of the track faults in the FFM.

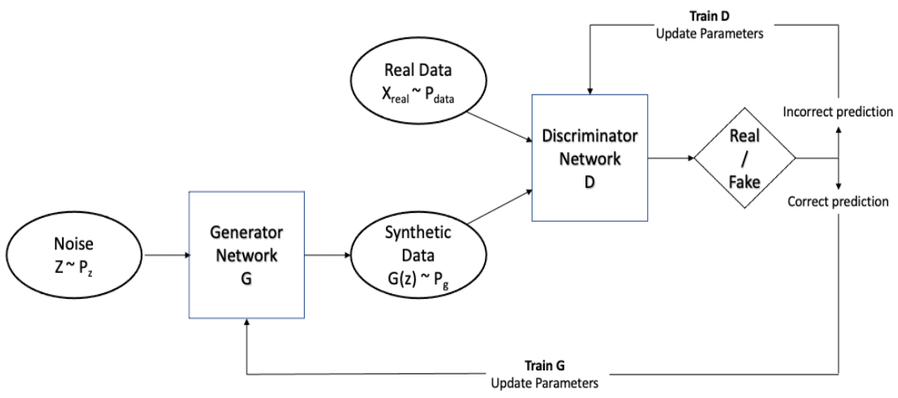
Yahya Bouchikhi



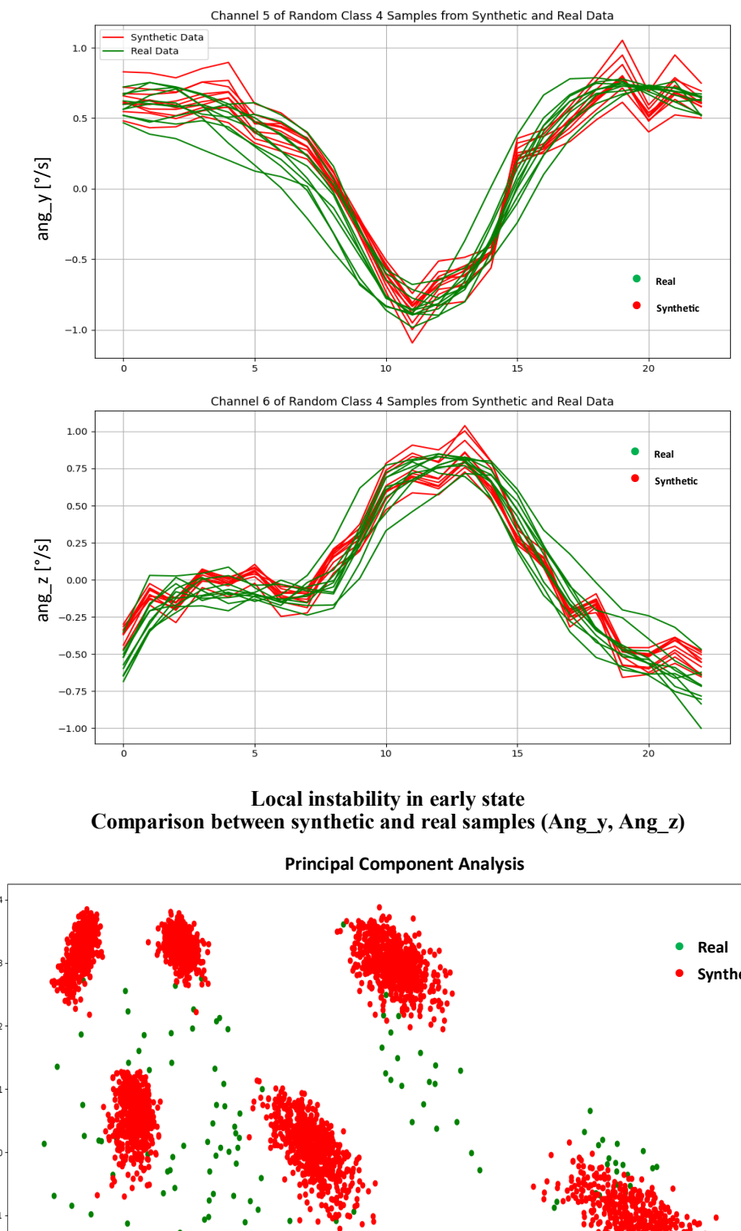
Workflow of the Master thesis



Vehicle-Track Scale Model



GANs basic Structure



Local instability in early state
Comparison between synthetic and real samples (Ang_y, Ang_z)

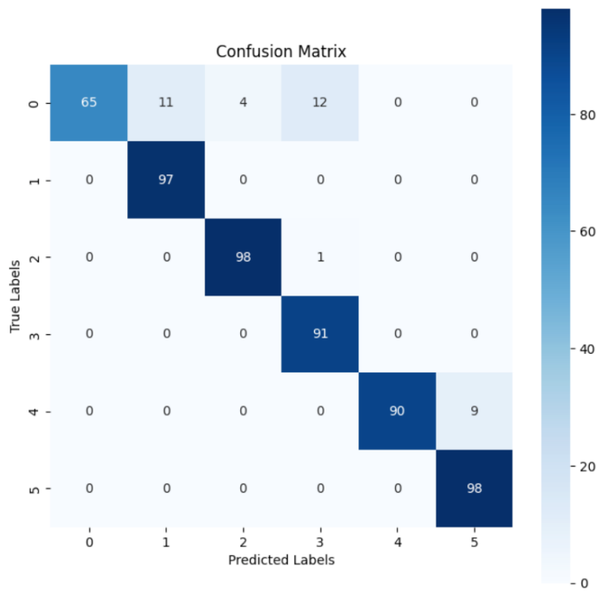
Results of PCA applied on real and synthetic data

6 clusters correspond to the six classes/defects. Most of the real points overlap with the synthetic points, indicating that GAN was able to reproduce most of the time and frequency features of the real data.

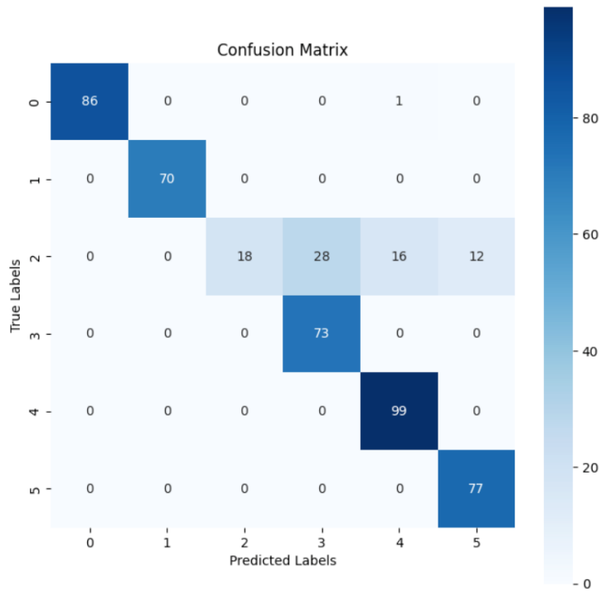
The training process provides promising results. The trained GAN is then used to generate data, which is then compared to the real data to assess their similarities in terms of time and frequency features. The synthesis quality assessment includes techniques such as dimensionality reduction via PCA and PSD comparison, as well as the performance of the classification scenarios. The findings demonstrate that synthetic data significantly enhances the training data's quality, leading to notable improvements in the performance of machine learning classifiers. This proposed solution effectively addresses the challenge of data scarcity. Data augmentation (DA) proves to be a valuable approach in generating high-quality training data for track fault classifiers, particularly when it is supplemented with high-quality real data during the GAN training process.

Scenario	TRTR	TRTS	TSTR	TSTS	Mixed
Accuracy	89.06%	86.67%	96.18%	100.00%	96.97%

FCN performance on classification scenarios



FCN : TSTR Confusion matrix



FCN : Mixed Confusion matrix

Master Thesis by Yahya Bouchikhi
Examiner: Prof. Dr.-Ing Ullrich Martin
Supervising Tutor: M.Eng. Héctor Alberto Fernández Bobadilla
Processing Period May - November 2023

Vorgelegt an der Universität Stuttgart

Institut für Eisenbahn- und Verkehrswesen

E-2023/16