Masterarbeit zu vergeben

Data Augmentation of Track Irregularities using Advanced Schemes of Generative Adversarial Networks and Frequency-based Loss Function

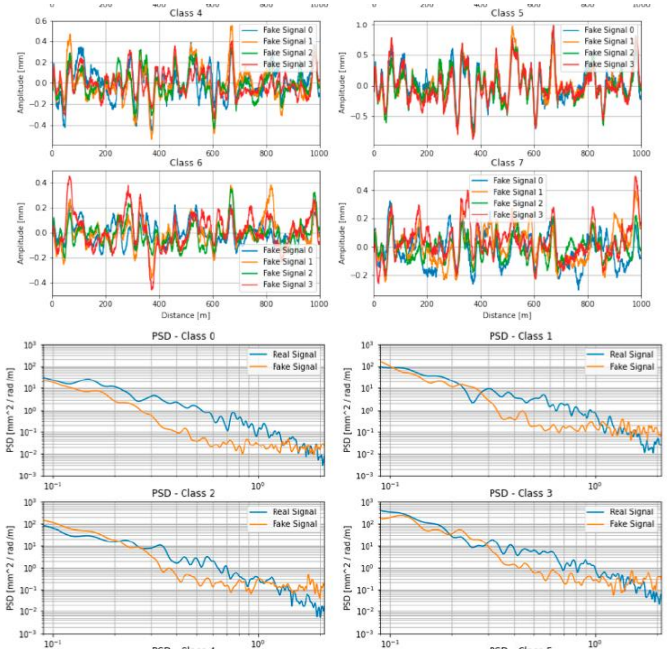
Data Augmentation von Gleisunregelmäßigkeiten mithilfe fortgeschrittener Schemata von Generative Adversarial Networks und frequenzbasierter Verlustfunktion

To perform classification tasks, Machine Learning (ML) algorithms require large amounts of information for training and learning. However, there are still applications where it is difficult, not affordable or even impossible to gather data. Railway systems is a typical example of this situation. In such cases, the available training information could be so sparse, that the use of ML techniques is no longer suitable. A solution to the issues of data sparsity is the data augmentation approach, where synthetic data is generated from just a few instances of information, increasing the amount of available, useful training examples.

Generative Adversarial Networks (GAN) have arisen as a promising approach to perform data augmentation. In this, two neural networks, a Generator and a Discriminator, compete against each other: the generator tries to produce synthetic data, similar to the real one, while the discriminator intents to recognise whether the data is real or it is the outcome of the generator. This interaction leads, theoretically, to the generation of synthetic data virtually indistinguishable from the real one.

The objective of this work is the further development of an existing Generator of Track Irregularities (time-series-like data) implemented using a Conditional GAN (CGAN). The purpose is the synthesis of realistic Track Irregularities signals which can be used in further research to train ML algorithms for Fault Diagnosis on the railway track.

In the framework of this thesis, several network architectures for the generator and discriminator will be created and compared, including fully connected (FCN), convolutional (CNN) and recurrent neural network (RNN) models. To improve the performance of the generative process and the quality of the synthetic signals, frequency features will be included in the loss function (Frequency-based Mean Square Error loss) and advanced GAN schemes will be employed, including Wasserstein GAN (WGAN), WGAN with Gradient Penalization (WGAN-GP) and Time-Series GAN (TSGAN). The synthesis process will be evaluated, among others, by using frequency approaches, dimensionality reduction techniques (PCA, t-SNE) and via the performance of the augmented dataset in classification tasks.



Bei Interesse wenden Sie sich bitte an:
Héctor Alberto Fernández Bobadilla, M.Eng.
Institut für Eisenbahn- und Verkehrswesen, Pfaffenwaldring 7, 70569 Stuttgart
Telefon: 0711 685-69255 , hector.fernadez@ievvwi.uni-stuttgart.de