# Master Thesis: Closing the Gap – Analyzing and improving the behavior of synthetic training data for a Neural Network image segmentation task.

## Background

The development of Deep Neural Networks (DNNs) demands for large amounts of training data, especially for computer vision tasks. One method to circumvent labour intensive labeling tasks is by generating synthetic training data and using the model – trained on synthetic data – for the real segmentation task (transfer learning). This data must be as realistic as possible to ensure that the trained model is generalizing well on real-world images.

Generative Adversarial Networks (GANs) have become one of the most promising technologies in the field of deep generative models over the last years. They have been successfully applied in image generation and image super-resolution.

## Description

At sewts, one of our core research areas is identifying different areas in a given image of a textile, specifically identifying the seam areas. To create a larger training dataset, we have developed a pipeline of finite element method (FEM) simulations and renderings to create highly realistic synthetical representations of textiles. We use those images to augment our datasets and speed up the development of our computer vision algorithms, but reliability and generalizability of this approach has yet to be investigated. The proposed thesis aims to develop and implement a scientifically accurate benchmarking procedure to assess the influence of synthetic training data on deep learning model development for a given image segmentation model. This benchmark shall be followed by an exploration of methods found in literature to improve the synthetic data, for example employing GANs. The results of this optimization are then to be assessed with the developed benchmarking procedure. The following work packages are particularly relevant:

* State-of-the-art research regarding the use of synthetical training data for deep learning and corresponding benchmarking procedures
* State-of-the-art research regarding style transfer methods and specifically the use of General Adversarial Networks (GANs) for image generation/image enhancement
* Generation of a representative benchmark dataset of synthetic and real textile images
* Definition of a deep learning training pipeline (hyperparameters, pre-processing etc.) to be used for a reproducible benchmarking process
* Design / implementation / execution of a benchmarking procedure to assess the influence of synthetic train data on the model quality
* Employment of Neural Network diagnostic methods (comparing output and activation functions, etc.) to find meaningful differences between models that have been trained on the synthetic and real data in hopes of gaining insight into the “black box” and what features are important for the given seam segmentation task.
* Design/implementation of one or several image enhancement procedures towards improving quality of the synthetic training data
* Benchmark the influence of the enhanced training data on the model quality