

# Bachelor-/Masterthesis

## in the field of data science and artificial intelligence

### When do we have „enough“ labeled data in the context of AI? Finding a saturation curve for annotation amounts

#### ■ Background

Thermography can aid in the detection of heat anomalies - such as those pertaining to leakages in district heating network pipelines. However, such infrared images are recorded in an urban context and therefore contain various hotspots stemming from common yet irrelevant features (like people, cars, manholes, or street lamps). Identifying these different classes is a task that deep learning models have been shown to excel at in the realm of image analysis. Such models are influenced by their data basis and the amount of available, labeled information significantly impacts a model's performance. However, the annotation process is cumbersome and time-consuming and the question of when "enough" data has been generated to achieve an acceptable model performance inevitably arises.

#### ■ Your contribution

The key aim of this study lies with finding a saturation curve for annotation amounts in the context of multi-class semantic image segmentation through systematic experimentation with an existing convolutional neural network and real-world, multi-spectral (RGBT) UAV-based data. It will therefore include:

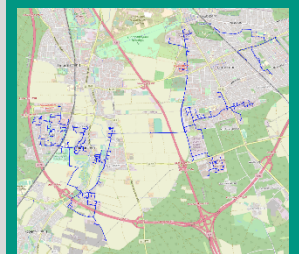
- research into existing literature to create an overview of method options
- applying a select method to the existing model and dataset in Python to analyze the different types of classes and influencing factors
- finding saturation curves for all the classes or a globally applicable one

#### ■ Requirements

- independent, structured way of working with an enthusiasm for scientific research, coding, and working with real-world data
- programming skills (particularly in Python) and knowledge of deep learning / artificial intelligence helpful, though not explicitly required
- proficiency in English

#### ■ Interested?

Please contact Elena Vollmer ([elena.vollmer@kit.edu](mailto:elena.vollmer@kit.edu)) with your application.  
Starting date: as soon as possible



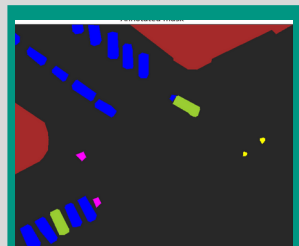
District heating network, displayed in QGIS



Exemplary thermal images acquired via Unmanned Aircraft System (UAS)



Example of an annotated image



Example of polygon annotations converted into an annotation mask