# Data Science coding challenge

Andrey Krutilin 08.02.2023

supper&supper

# Contents

## Coding Challenge

1) Crop Yield Prediction

2) 3D Ground Segmentation

3) Weld Defects Detection

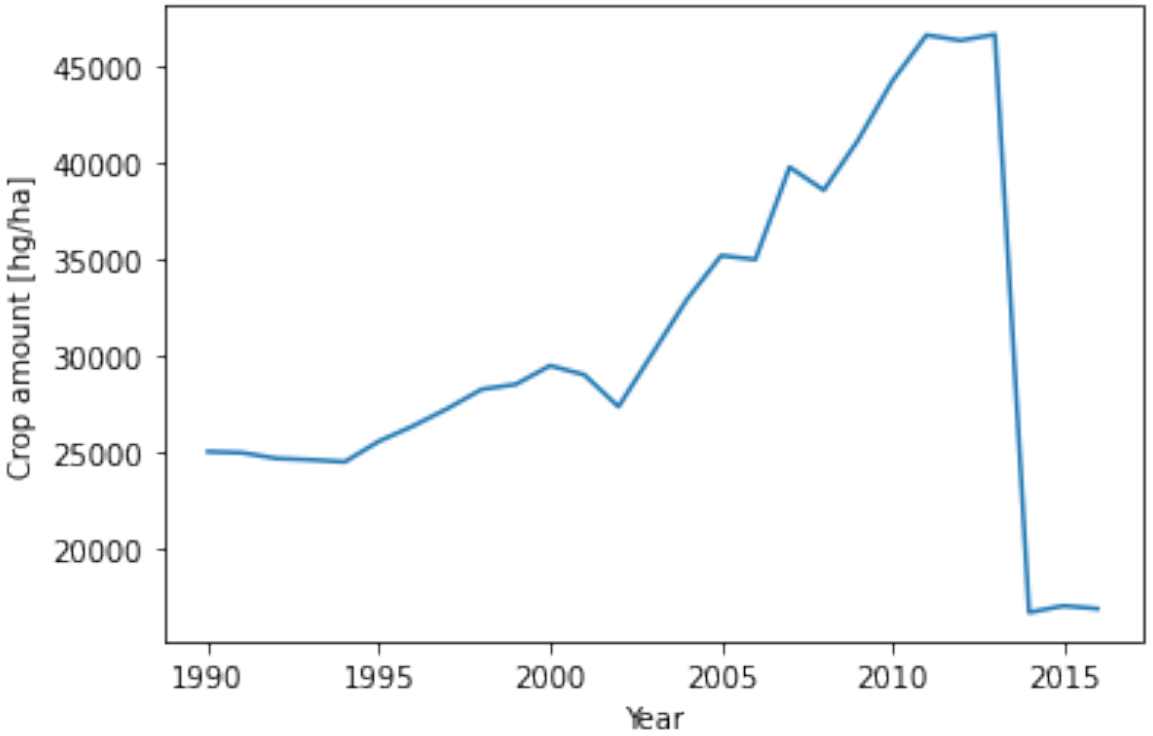
# Crop Yield Prediction

supper&supper

# Crop Yield – Introduction

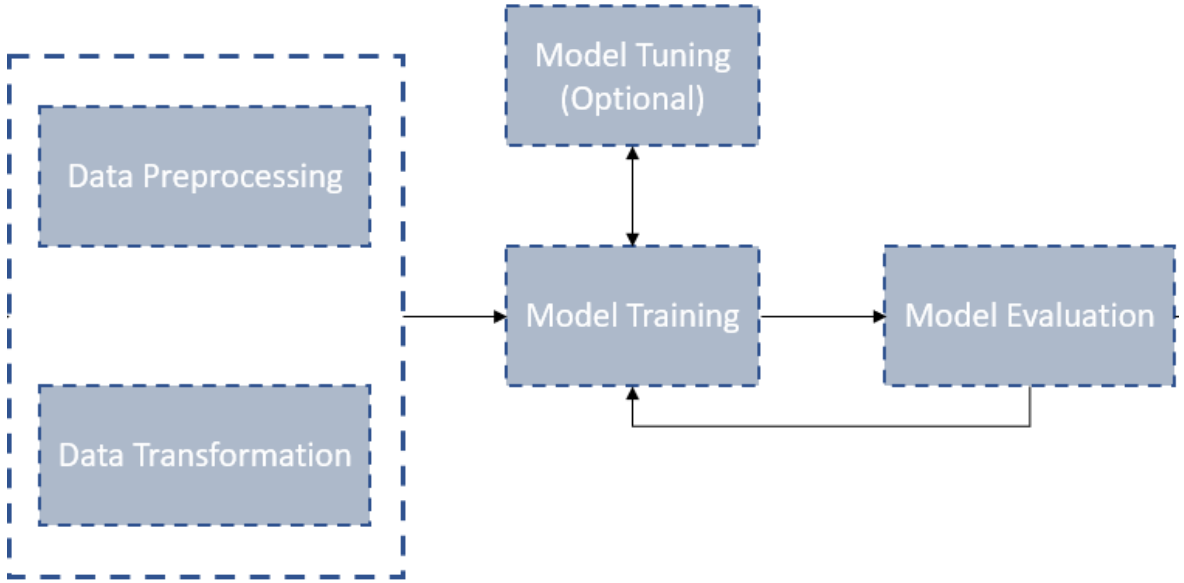
## Datasets

- Crop yield is essential for a growing population
  - Understanding the impact of various factors will improve food production
- Datasets contain over 200 countries
- The dataset timeline overlaps for all between 1990 – 2016
- Hypothesis: Crop yield is a function
  - Rainfall
  - Temperature
  - Pesticides
  - Geolocation



# Exploratory Data Analysis and Machine Learning

- Data preprocessing:
  - Exploratory data analysis
  - Merging data in a single table
- Model training:
  - The target is a continuous variable
    - Random forest regressor (best performance)
    - Linear regression
    - Bayes ridge
- Model evaluation:
  - R-squared: 0.97
  - RMSE: 13 900
  - An important feature for crop yield: pesticides, while rainfall is the least important



# UK as a New Item for Model Evaluation

- Question:
  - Hypothetical use case of rolling out the model to unseen country
- UK as unseen datapoint:
  - Linear regression, fandom forest perform
  - Results:
    - R-squared: 0.95
    - RMSE: 38 000
- Outlook
  - Cohorts are neglected and a model is applied
  - Rainfall dataset average multiple decades
  - Apply ensemble ML algorithms

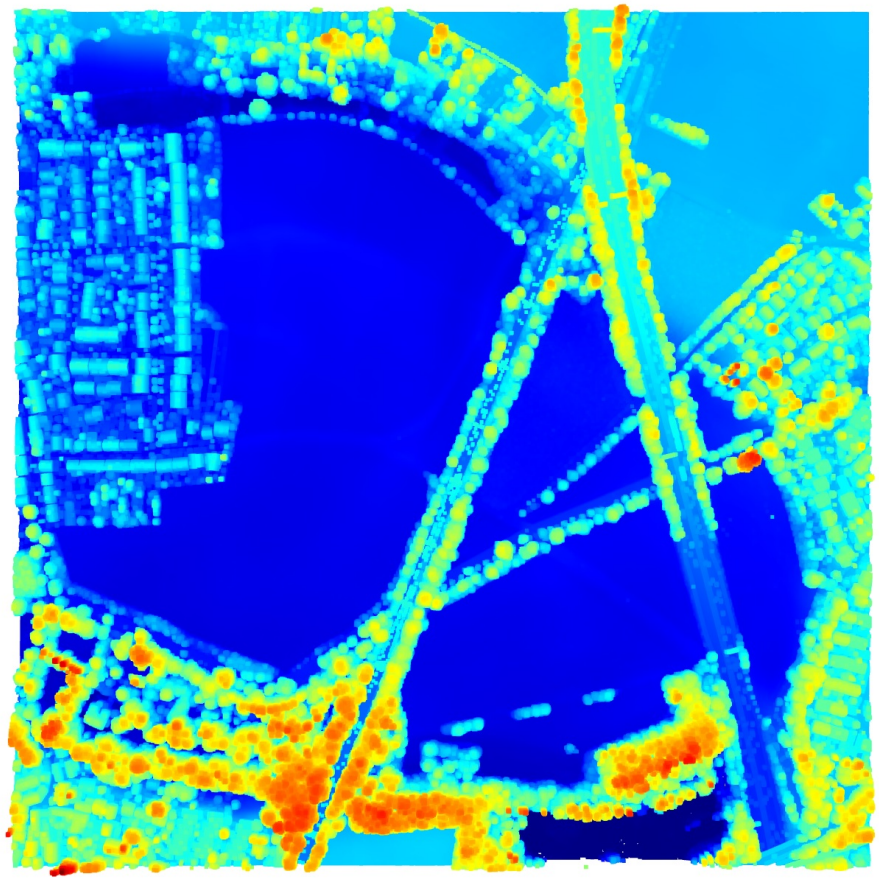
# 3D Ground Classification

supper&supper

# Point Cloud of a Landscape in Bird View

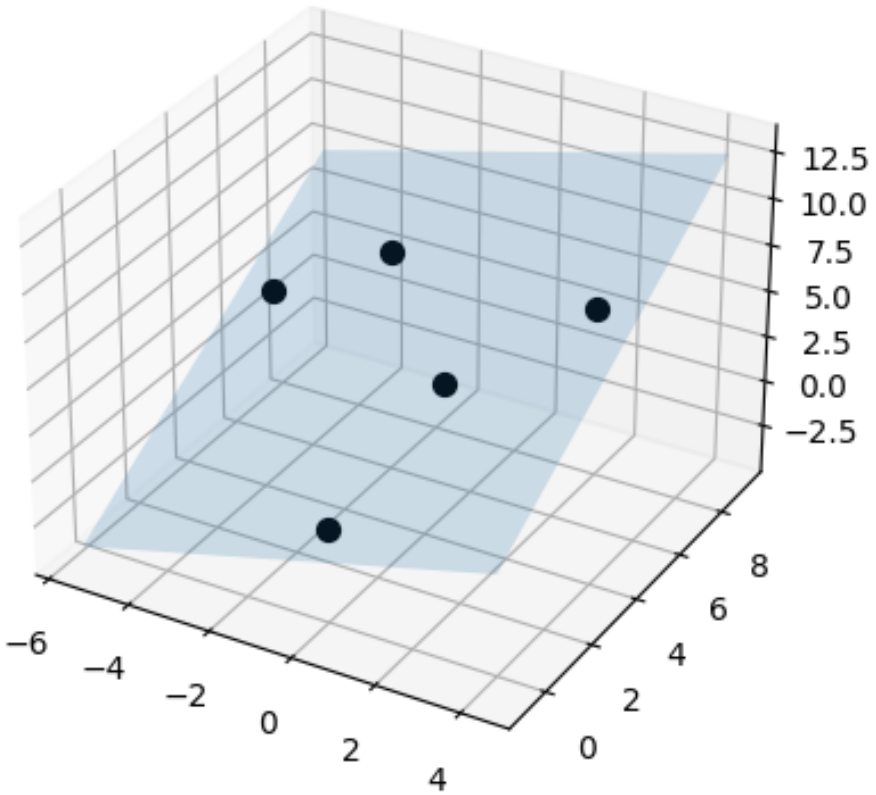
## Lidar Created Map Containing a Landscape Profile

- Point clouds are used for many purposes, including:
  - Construction progress tracking
  - 3D model creation
- Map consists of 10 mio. points each with an unique X, Y, and Z coordinates
- Blue regions are closer to the ground, while red areas are elevated
- First approach:
  - Matplotlib, but data processing is too slow



# Ground Segmentation with a Plane

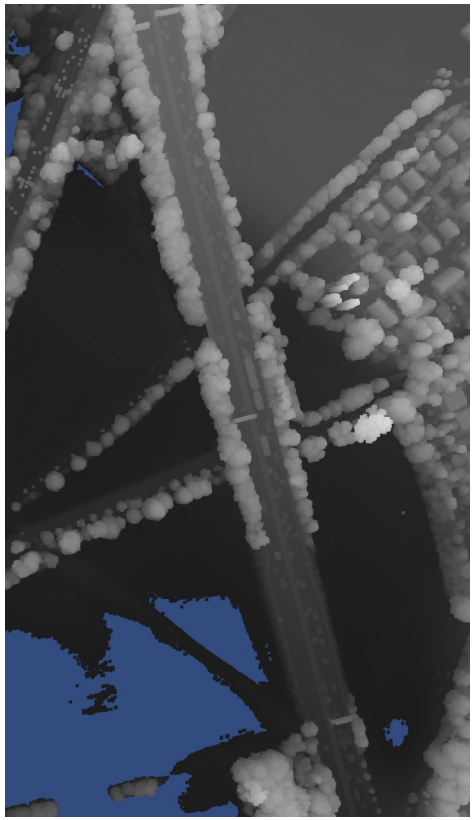
- The plane is perpendicularly placed in Z-direction
- Adjusting the height of the plane selects more objects



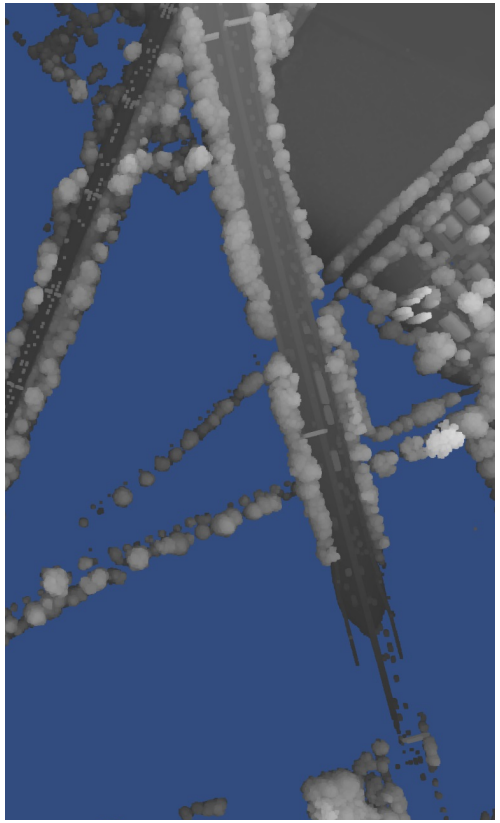
# Higher Plane Classifies more Points as Ground

- A plane intersects with the landscape
- The intersection plane is blue
- Problem: Plane is not selective

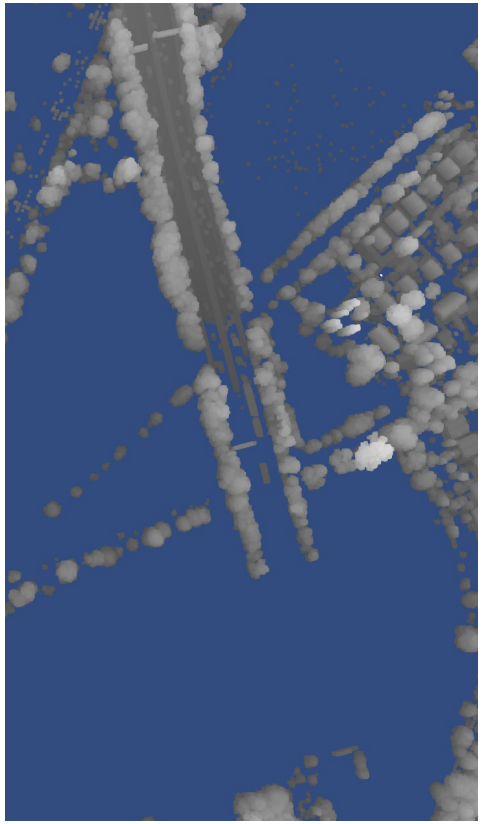
- Threshold Z = 5000



- Threshold Z = 5400



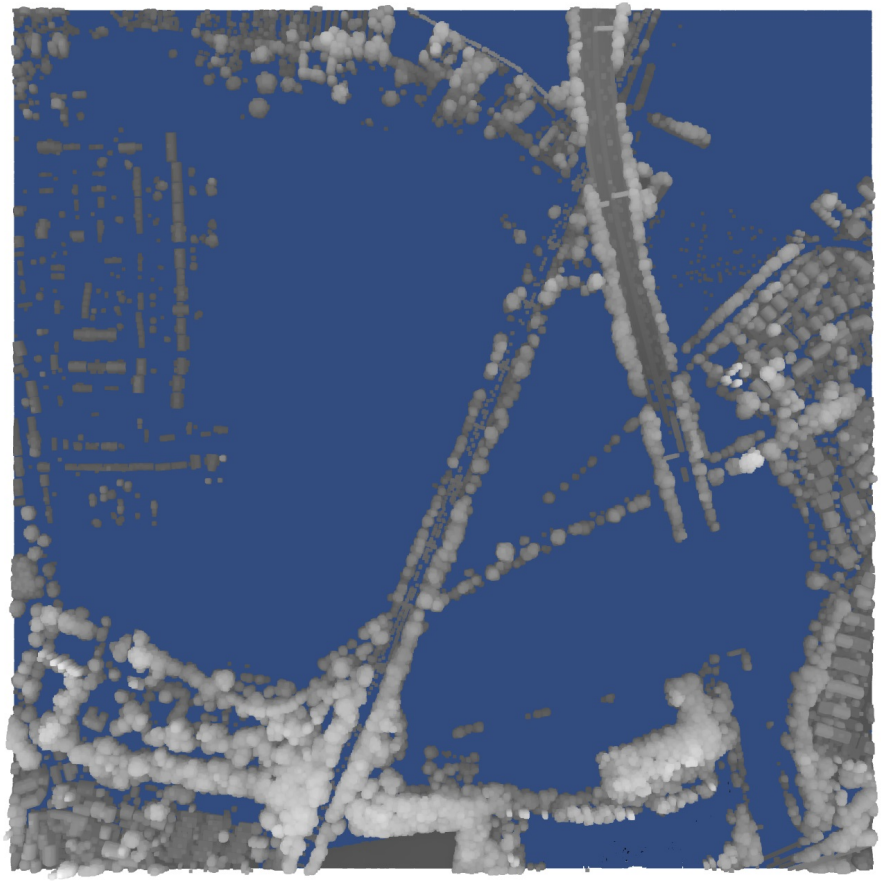
- Threshold Z = 5900



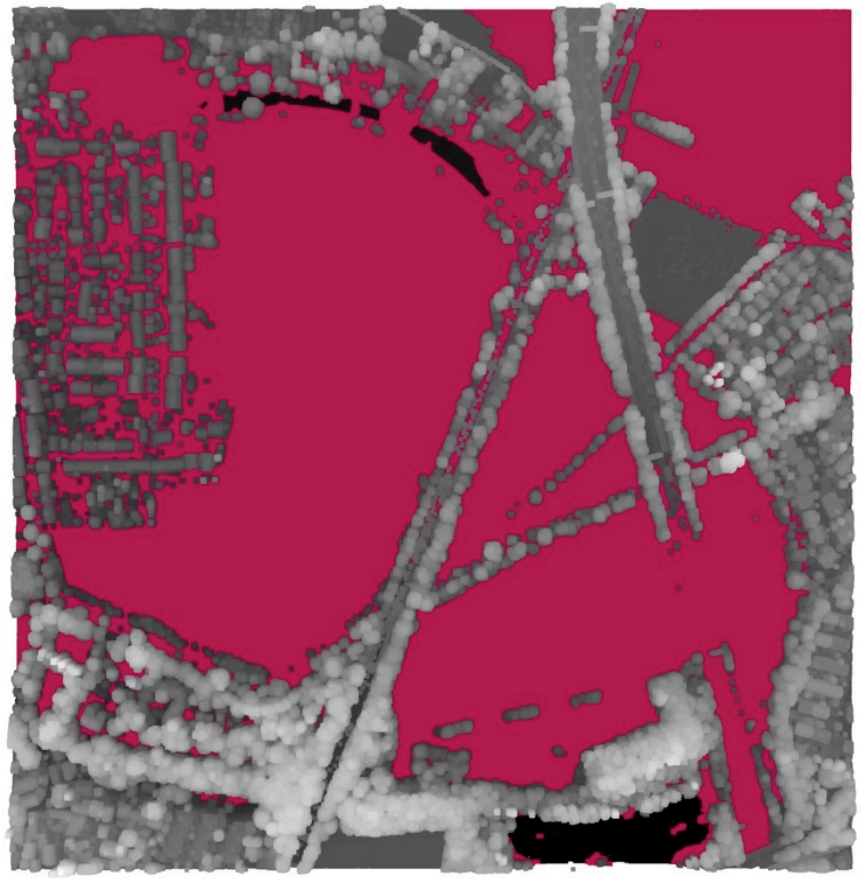
# Build in Method in Open3d Library

## Comparison between

- Threshold = 5900



- Plane segmentation method

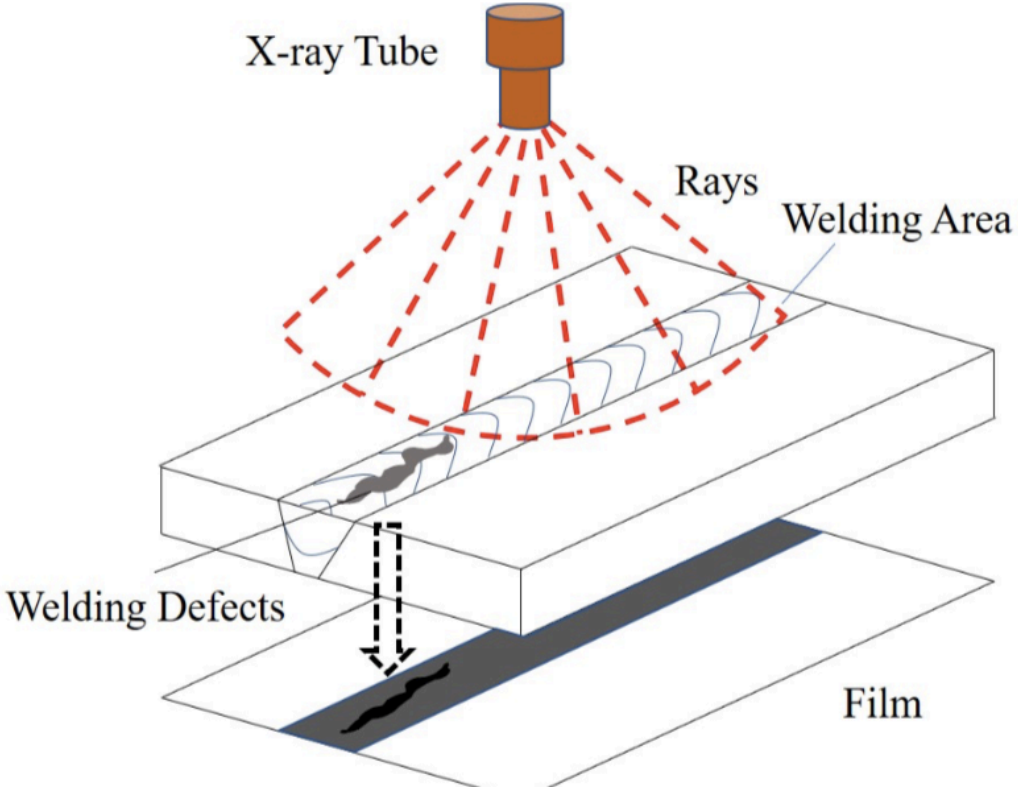


# Weld Defects Localisation

supper&supper

# Weld Defects - Introduction

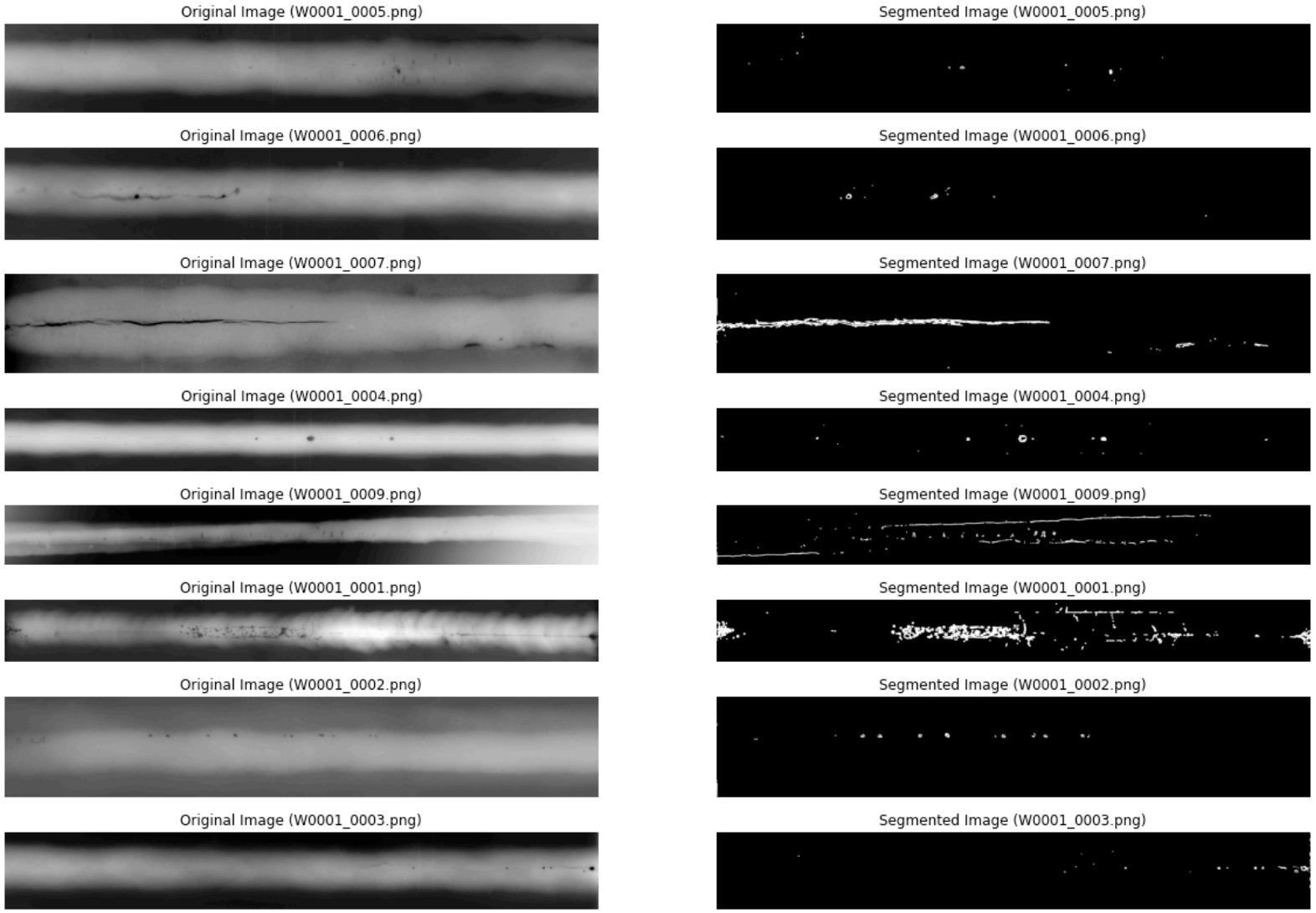
- Missed or defect welds can have a impact on safety issues and have negative consequences for businesses
- Measures: Quality control with X-ray images possible
- First approach:
  - Increasing the contrast
  - Sobel operator
  - Laplace operator



# Canny Edge Detection

## Detecting Areas with a High Contrast

- Areas with sharp contrast are detected
- Afterward area of interest is increased
- Problem: High contrast areas but no defect
- Outlook: ML models on labeled datasets



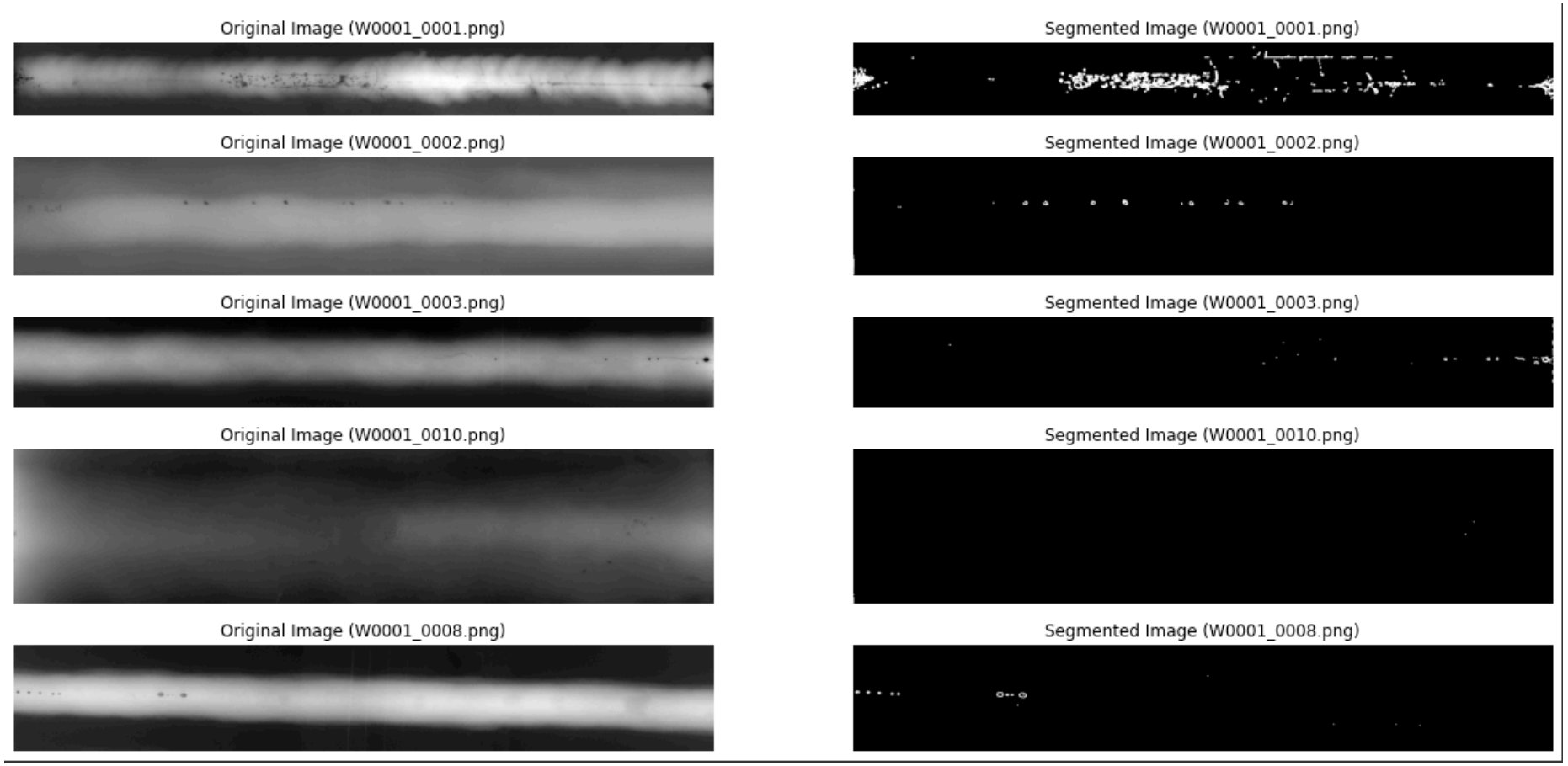
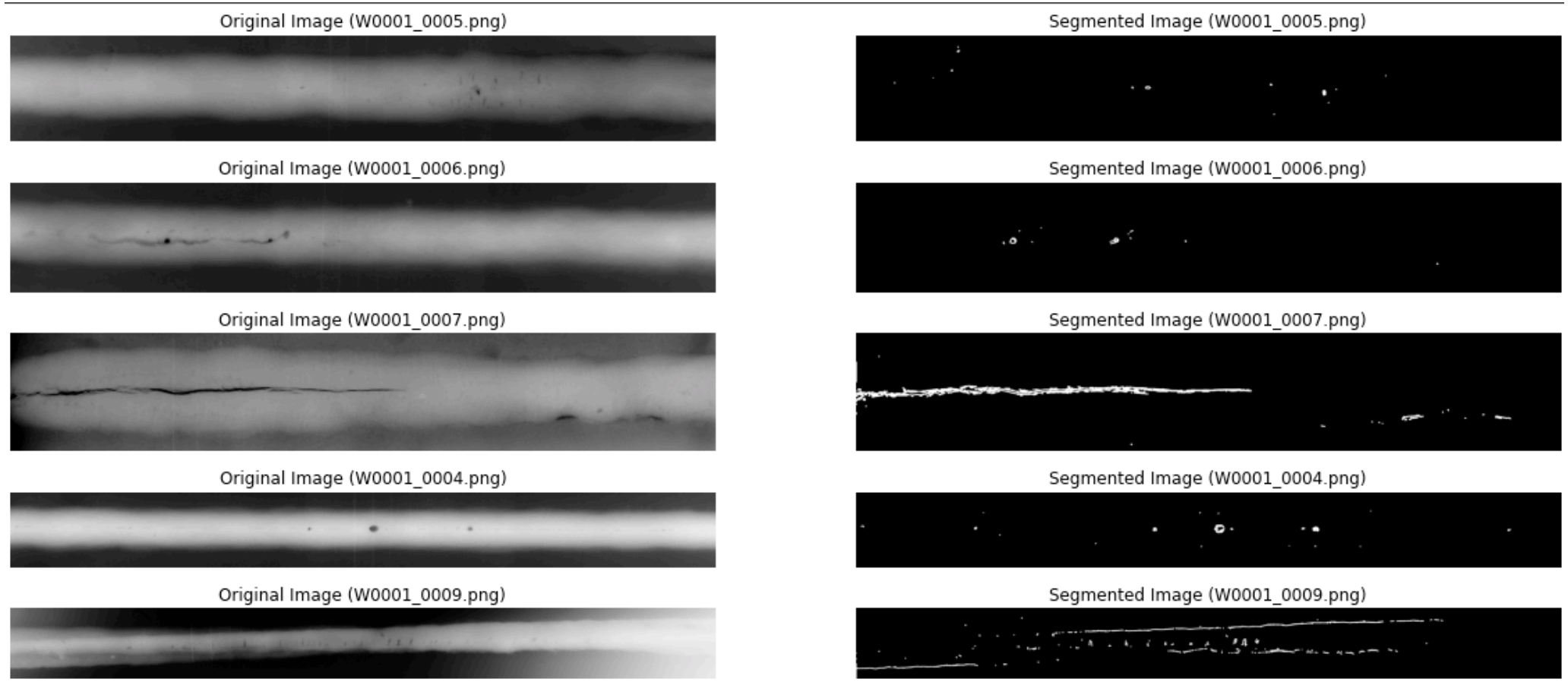
# References


- Crop Yield
  - Pandas, Sklearn documentation
- Point Cloud Ground Segmentation
  - Open3d documentation website
  - Arcgis, https//:desktop.arcgis.com/de/arcmap/10.3/manage-data/las-dataset/lidar-point-classification.htm
- Welding Defects Detection
  - A Lighter and Faster YOLO for Weld Defect Detection of X-ray Image, Moyun Liu, Youping Chen, Lei He, Yang Zhang, Jingming Xie, J. OF Latex Class Files, VOL. 14, NO. 8, August 2015
  - Automatic Welding Defect Detection (https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8758813)
  - OpenCV documentation website

# BACKUP

supper&supper

# Weld Defects Detection – Backup 1



# Weld Defects Detection – Backup 2