# Semi-supervised segmentation with AutoML

## Project description

In this project I demonstrate an app that leverages semi-supervised learning to aid labelling and segmentation of data that could be used for further analysis or training of a deep learning model.

For demonstration purposes we will use an example of a Molybdenum Oxide Crystal (<https://deben.co.uk/detectors/stem-detector/>) recorded using annular dark field (ADF) STEM imaging.

A schema of the JEOL ARM 200CF electron microscope at the University of Oxford Materials Department is shown below.

<figure>

Often images taken from such instruments are processed manually. Here we want to segment the various phases of the crystal.

<figure>

Figure credit:

## Segmentation method

A few months ago, the Dash – Plotly team (link) created the dash-canvas tool (link). This inspired me to create a segmentation tool like WEKA or Ilastik in python.

The workflow is as follows:

<figure>

For the interactive segmentation the dash canvas tool was used to paint over the relevant pixels within the image

<GIF>

For the image pre-processing, a set of filters similar to the WEKA segmentation approach was used

<figure>

For creating the machine learning classifier, an autoML tool such as TPOT was used. TPOT allows the automation of ML pipeline building by combining genetic programming and data pre-processing workflows from scikit-image.

<figure>

For demonstration purposes, the optimiser was run for 20mins to create the pipeline shown below that was used for segmentation.

## Demonstration of the workflow as a web application

The complete workflow describe here can be packaged into a web application shown in the video. In this demonstration a MoS2 crystal’s phases are identified by painting over them and then used to train a classifier model for segmentation purposes.

<video>

## Closing remarks and things to improve on

Here we have seen an example of a semi-supervised machine learning pipeline to segment images from various imaging experiments. The strength of this approach is that it leverages the scientist’s expertise over a generic mechanical Turk service to provide improved labelling for more advanced machine learning pipelines.

This is a relatively simple setup however there are many things to improve on depending on the sample. One example is to use relevant filters depending on the material analysed and the other is to optimise the classifier even more by letter the genetic algorithm fine-tune for longer.