Evaluating Crystal Facets for Photoelectrochemical Water Splitting
Semantic segmentation of Cu$_2$O nanocrystals

Rouven Spiess
rouvens@stanford.edu

Motivation
Photoelectrochemical Water Splitting could be a promising way for producing Hydrogen from Solar Energy. Cuprous oxide Cu$_2$O Microcrystals are interesting candidates, as of their abundance on earth and non-toxicity. The electrocatalitic activity depends largely on the facet type, thus, a method for determining the exposed facets is paramount to increase devices performance.

Dataset
26 Scanning Electronic Microscopy (SEM) images representing cubic, octahedral and 26-facet polyhedra crystals. The total amount of data is 1MB. The images are labelled manually to deliver the ground truth images.

Data Augmentation
As the dataset is small, a proper data augmentation strategy is key. We can take advantage that SEM images are rotation invariant by applying random rotations and augmenting the data massively. Other transformations include shear, grid distortion and zoom.

Network and Training
U-NET architecture is suitable for semantic segmentation and small datasets. The following U-NET variants are tested:

- U-NET Mini
- Vanilla CNN U-NET
- VGG-16 U-NET pret-trained on ImageNet

Overall, the models present a high variance problem, which can only moderately be mitigated by regularization techniques.

Results

Discussion and Future Work

Semantic segmentation problems can be solved with low amount of data. SEM images provide enough spatial information to be augmented. U-NET architectures, with either low number of parameters or pretrained weights are powerful methods to train on limited datasets. As project continuation, Synthetic data generation, where the crystal geometries are build from scratch, could be an interesting alternative to manual labelling, which is time consuming. This should be coupled with a domain transfer technique, e.g. by training a GAN to match the original distribution.