Cu_2O Nanocrystal Facet Segmentation for Photoelectrochemical Water Splitting

Rouven Spiess
SCPD student – Artificial Intelligence Graduate Certificate
Department of Computer Science
Stanford University
rouvens@stanford.edu | https://github.com/pan0rama

***Abstract*—Photoelectrochemical (PEC) water splitting offers a promising path for sustainable production of hydrogen (H_2) fuel from solar energy [1]. The potential of PEC devices made from metal-oxide crystals is demonstrated. However, their effectiveness still remains a matter of study. The photocatalytic efficiency depends strongly on the exposed crystal facets. While it is possible to drive crystal synthesis such as to expose a given facet, in many cases, the exact distribution of exposed facets remains unknown. We propose an algorithm that determines the distribution of facets from Scanning Electron Microscope (SEM) images. Cuprous oxide (Cu_2O) nanocrystals are interesting study examples, as they are non-toxic, relatively abundant and efficient solar-to-hydrogen energy converters. We train and test several algorithms for semantic segmentation and obtain a mean Intersection over Union (mIoU) score of 0.52 with a U-NET 5-stage network.**

I. Introduction

The surface properties of metal-oxydes depend strongly on the exposure of certain facets. In the case of Cu_2O nanocrystals, which has applications in photocatalysis, organocatalysis, and sensing, any of these properties are more or less attainable, depending on the exposed facets [2]. To our knowledge, a quantitative method for facet exposure determination is not available today. The goal of this project is to build an appropriate CNN architecture that is able to perform semantic segmentation of nanocrystal facets. In this manner, physical and chemical properties of sythesized nanocrystals can be linked to their respective crystal shapes, or facet composition. We use semantic segmentation algorithms usually applied to medical imagery and scene understanding [6]. The input to the algorithm is a SEM image of Cu_2O nanocrystals. The output is a color image with predicted facets. In this type of problem, the amount of data is small, as SEM images are relatively expensive, and it would be difficult to train a common deep neural network. However, appropriate architectures paired with data augmentation can take advantage of spatial information and converge even on small datasets.

II. Related work

Image classification of the presence of facets was achieved on a $BiVO_4$ nanocrystal dataset with a shallow CNN [3]. The idea of facet classification expands to pixel-wise labeling (semantic segmentation), to determine the exposed surface of several facets, with crystallographic orientations (100), (110) and (111) in the case of Cu_2O nanocrystals.

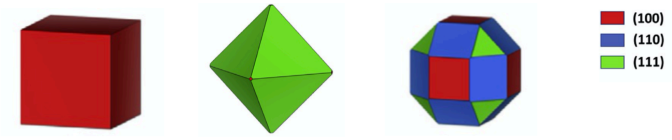

Fig. 1. Cubic, octahedral and 26-facet polyhedral Cu2o nanocrystals [4]

Semantic segmentation encompasses localisation, which provides not only the classes but also additional information regarding the spatial location of those classes. Semantic segmentation makes dense predictions inferring labels for every pixel.

III. Dataset and Features

A. Dataset

The dataset consists of a total of 26 SEM images with size ranging from 139x106 to 370x256 pixels. It depicts cubic, octahedral and 26-facet polyhedra crystals. The ground truth masks are manually labeled according to the indications in Figure 1. As the dataset is small, an appropriate data augmentation technique is key to have a neural network that converges while reducing overfitting of the model.

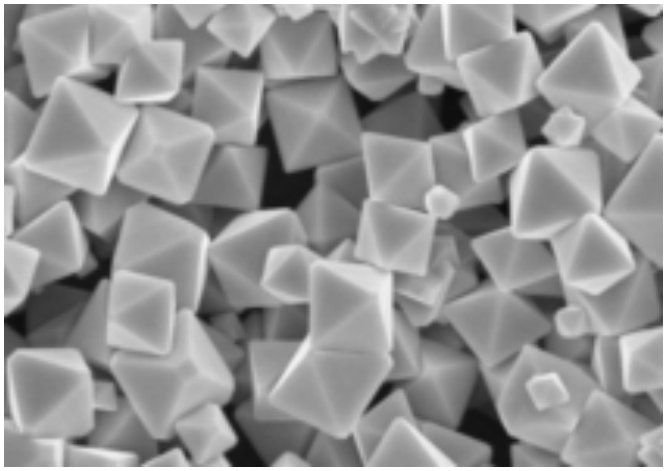

Fig. 2. SEM image of octahedral Cu_2O nanocrystals

B. Data Augmentation

Typically, SEM images show large variations of contrast, no preferential orientation, noise, shear and distortion created by external perturbations during the scanning process. To preserve feature representation, and as the magnification varies largely among the images, they are rescaled so that the pixel size of the nanocrystals is analogous. Furthermore, the following random transformations are applied sequentially: Cropping with size 96x96 pixels, horizontal and vertical mirroring, rotation with angle +/-10, shear with angle +/-3, variations in brightness, contrast, zoom and elastic distortion on a 5x5 grid [5].

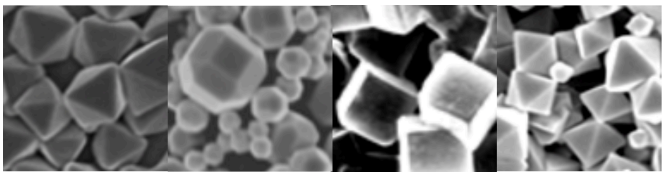

Fig. 3. Examples after augmentation, showing variations in contrast.

There are 22 pictures that are augmented for the training set, and 4 for the test set, in order to obtain 4000/250 pictures after augmentation.

IV. Methods

U-NET is a type of architecture that was proposed for biomedical semantic segmentation [6] . It consists of a contracting path (encoder) and an expansive path (decoder). During the contraction, the spatial information is reduced while contextual information is increased. The expansive path combines context and spatial information through a sequence of up-convolutions and concatenations with high-resolution features from the contracting path. Each stage of the encoder consists of two 3x3 convolutions, followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling. At each stage, the number of feature maps is doubled, while halving the resolution.

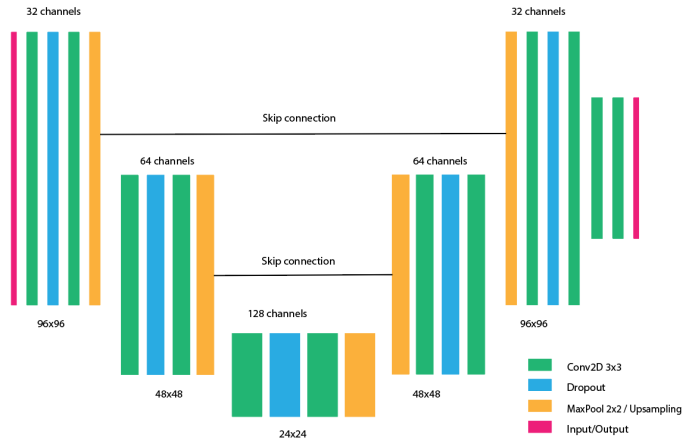

Fig. 4. U-NET mini (5-stage) architecture

In what follows, we consider the further U-NET architecture variants [7]:

- U-NET mini (5 stage U-NET)
- U-NET Vanilla CNN encoder
- U-NET VGG-16 pretrained encoder

The first is a small version of the original U-NET with 5 stages instead of 9. The second has a default CNN encoder. Lastly, we use a U-NET with VGG-16 encoder with weights pre-trained on the ImageNet dataset.

TABLE I
Architecture Training Parameters

Architecture	# Parameters
U-NET mini	473,203
U-NET Vanilla CNN encoder	4,504,192
U-NET VGG-16 trained on ImageNet	12,351,219

The loss function for all models is the cross entropy error function, defined as:

$$\mathcal{L}(y, \hat{y}) = \sum_{i=1}^{P} \sum_{j=1}^{C} y \cdot log(\hat{y_{ic}}), \tag{1}$$

were we are summing over all pixels P and classes C.

V. Experiments/Results/Discussion

A. Training

The networks are trained using minibatch gradient-descent with batch size 8. A batch size large enough will ensure that each batch has a balanced set of classes, as many images only represent one class. While training accuracy still increases for larger batch sizes, it also increases training duration remarkably. We choose ADAM optimizer with learning rate 0.001.

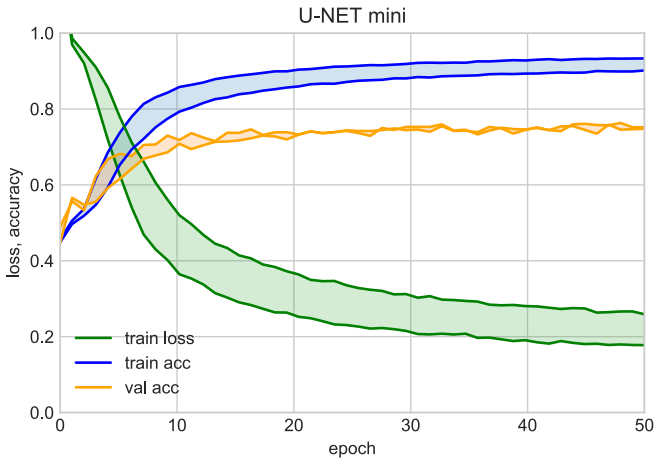

Fig. 5. Training on U-NET mini network

The best performance is achieved by the U-NET architecture with 5-stages (U-NET mini). Considering that human-level performance estimation is at 0.95, then there is a 2% training error and 19% test error (for the best model), hence, the models

are suffering from a high variance problem (overfitting). In order to reduce variance, several options are considered, including increasing the dropout rate for all layers. While this effectively decreased the training accuracy, no change was observed for the validation accuracy. This is attributed to a hard limit in optimisation arising from small dataset. Figure 5 depicts this behavior for dropout rates of 0.1 and 0.3.

B. Performance and Metrics

As evaluation metric, the Intersection over Union (IoU) is appropriate for semantic segmentation [8]:

$$IoU = \frac{target \cup prediction}{target \cap prediction} \tag{2}$$

As for a typical semantic segmentation application, a target $IoU > 0.5$ is desirable. The resulting model performance is shown on table II after hyperparameter tuning.

TABLE II
ARCHITECTURE PERFORMANCE

Architecture	train acc	val acc	val IoU
U-NET mini	0.93	0.76	**0.60**
U-NET Vanilla CNN	0.92	0.71	0.51
U-NET VGG-16	0.96	0.76	0.58

C. Predictions

Prediction images are computed from the U-NET 5-stage architecture after training during 50 epochs, achieving a mean IoU score of 0.60 on the validation set, and 0.52 on the test set, figures 6 to 11:

Fig. 6. (100)

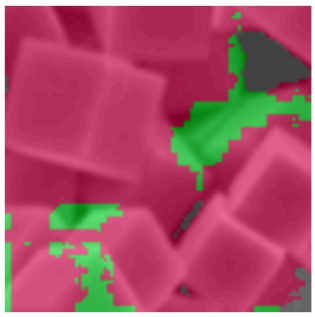

Fig. 7. (100)

An error analysis of 250 predicted images has exposed the following failure modes:

- Cubic crystals are inclined and look like octahedra crystals (figure 7)
- Facets are locally close to each other, and merge or disappear (figure 11)

As the application concerns facet presence and composition, the class wise IoU and pixel accuracy PA are further compared for 250 predicted images, table III.

An IoU score of 0.5 means that a class is represented to 50% correctly on the prediction image. However, the pixel accuracy PA is also of interest as precise localisation is not an issue for this application. Either way, class (110) is best predicted by the algorithm. The images containing 26-facet polyedra lower the overall score for classes (100) and (111). Indeed, segmentation is more challenging task here, as distinct facets are close to each other and have low densities, e.g. only 3% for class (111). One possible explanation for the detailed prediction error at the edges of the facets is related to information loss from the contracting path. Due to downsampling from the 2x2 max pooling operations, finer details than 4 pixels (U-NET 5-stage architecture has 2 downsampling stages) cannot be represented on the predicted image.

Fig. 8. (111)

Fig. 9. (111)

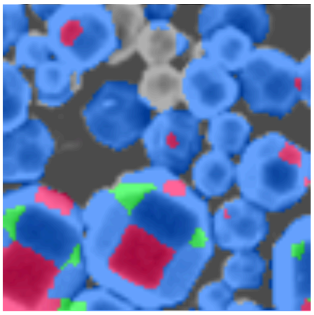

Fig. 10. (100), (111) and (110)

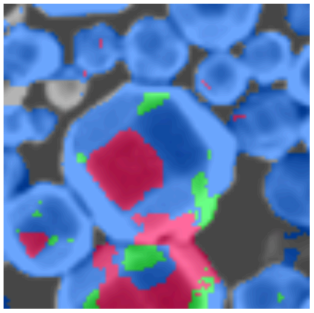

Fig. 11. (100), (111) and (110)

TABLE III
CLASS WISE SCORES FOR U-NET MINI ON TEST SET

Class	IoU	PA
(100)	0.48	0.54
(111)	0.44	0.52
(110)	0.54	0.81
background	0.62	0.79
mean	**0.52**	**0.66**

VI. CONCLUSION/FUTURE WORK

The U-NET 5-stage architecture achieves a minimally acceptable performance on the Cu_2O dataset, with an IoU score of 0.52. The benefits of transfer learning are reduced by the large number of parameters in which the model with VGG-16 encoder has to train on, respectively this could also be the reason why the U-NET mini performs well. However, all models have limited performance on adjacent and small facets. For an industrial setting, the pixel accuracy PA is important, and should be increased for the (100) and (111) classes on

can be successfully applied to SEM images, given their spatial information content.

Feasibility of crystal facet segmentation with variants of U-NET architectures have been demonstrated. In future scope, the models prediction capabilities could be further enhanced with larger datasets, including other type of crystal shapes. As manual labeling is time consuming and a limiting factor, synthetic data generation could offer an interesting path to a unified algorithm for many nanocrystal shapes.

VII. Contributions

SEM images were provided by Thomas A. Gill, Department of Mechanical Engineering at Stanford University.

References

[1] Sun Shaodong, Song Xiaoping, Sun Yuexia, Deng Dongchu, Yang Zhimao. (2012). *The crystal-facet-dependent effect of polyhedral Cu2O nanocrystals on photocatalytic activity.* Catal. Sci. Technol.. 2. 925-930. 10.1039 https://arxiv.org/abs/1505.04597

[2] Shang, Y. Guo, L. (2015). FacetControlled Synthetic Strategy of Cu2OBased Crystals for Catalysis and Sensing. Adv. Sci., 2: 1500140.

[3] Zixi Liu, Wanling Liu, Jiyao Yuan *Bismuth Vanadate (111) Facet Detection*. CS229 Final Report, December, 13, 2018. http://cs229.stanford.edu/proj2018/report/215.pdf

[4] Ferrer, Mateus Fabris, Guilherme Faria, Bruno Martins, Joo Moreira, Mrio Sambrano, J.. (2019). Quantitative evaluation of the surface stability and morphological changes of Cu2O particles. Heliyon. 5(10):e02500 October 2019. https://www.sciencedirect.com/science/article/pii/S2405844019361602

[5] https://github.com/mdbloice/Augmentor

[6] Ronneberger, O., Fischer, P., Brox, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. (2015). https://arxiv.org/abs/1505.04597

[7] https://github.com/divamgupta/image-segmentation-keras

[8] Ulku I., Akagunduz E., *A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D Images.* (2020). https://arxiv.org/abs/1912.10230v1

REFERENCES

[1] Sun Shaodong, Song Xiaoping, Sun Yuexia, Deng Dongchu, Yang Zhimao. (2012). *The crystal-facet-dependent effect of polyhedral Cu2O nanocrystals on photocatalytic activity.* Catal. Sci. Technol.. 2. 925-930. 10.1039 https://arxiv.org/abs/1505.04597

[2] Shang, Y. Guo, L. (2015). FacetControlled Synthetic Strategy of Cu2OBased Crystals for Catalysis and Sensing. Adv. Sci., 2: 1500140.

[3] Zixi Liu, Wanling Liu, Jiyao Yuan *Bismuth Vanadate (111) Facet Detection*. CS229 Final Report, December, 13, 2018. http://cs229.stanford.edu/proj2018/report/215.pdf

[4] Ferrer, Mateus Fabris, Guilherme Faria, Bruno Martins, Joo Moreira, Mrio Sambrano, J.. (2019). Quantitative evaluation of the surface stability and morphological changes of Cu2O particles. Heliyon. 5(10):e02500 October 2019.

[5] Ronneberger, O., Fischer, P., Brox, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. (2015). https://arxiv.org/abs/1505.04597

[6] Ulku I., Akagunduz E., *A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D Images.* (2020). https://arxiv.org/abs/1912.10230v1

[7] https://github.com/divamgupta/image-segmentation-keras

[8] https://github.com/mdbloice/Augmentor