Underfitting Heuristic Segmentation Models for Superior Neural Results

Jack D. Carson[1], Gabriel LeBlanc[2]

[1]Department of Electrical and Computer Engineering, University of Tulsa
[2]Department of Chemistry and Biochemistry, University of Tulsa

Abstract

As datasets for deep machine learning architectures have grown exponentially, the technique of machine-assisted dataset generation has become increasingly necessary. Typically, this technique takes the form of network generating validation datasets or labels for another network, which are fine-tuned by human feedback. This paper presents a model, trained entirely on auto-generated validation masks, to identify defects in images of electrochemical depositions. A U-Net segmentation architecture was trained by underfitting the model on a low quality validation dataset which was created with an entirely deterministic kernel convolution of a 512×512 image. The model has been shown to exceed the quality of its training data in terms of lack of noise, generalizability, and correctness when tested on a calibration sheet of laser-engraved film defects. Furthermore, the dataset was created with very minimal human fine-tuning or adjustment, which was necessary for localizing hundreds of small cracks and spots per chemical film photo sample. Due to the small network, training was complete in only 7 epochs, taking a total of 6 minutes on average to completely train. The proposed technique is believed to be applicable to similar, basic segmentation tasks; however, the effectiveness of it for pretraining largely depends on the nature of the data to segmented productively without the use of machine learning.

Keywords: *U-Net, Semantic Segmentation, Biomedical Imaging, Image Processing*

1. Introduction

The advent of artificial intelligence (AI) has revolutionized numerous fields, including materials science and manufacturing. In particular, AI has shown immense potential in the analysis and quality control of electrochemical films used in the production of perovskite solar cells (PSCs). PSCs have emerged as a promising technology in the field of photovoltaics due to their high power conversion efficiency and cost-effective fabrication processes [3]. Our laboratory focuses on the roll-to-roll production of PSCs through the use of electrodeposition. In this process, conductive indium tin oxide (ITO) plastic substrates are plated with a thin layer of a desired chemical; when complete, the process creates a stack sufficient to synthesize an efficient electrode. However, the presence of defects such as scratches, spots, and holes can significantly impair the functionality of PSCs. These defects can lead to nonradiative recombination, which lowers the photoluminescence quantum yields (PLQYs), device efficiency, and stability. As the demand for renewable energy increases, efficient and accurate methods to detect defects in the manufacturing process of perovskite solar cells will become essential.

Obtaining a computation model that is capable of quickly and precisely identifying hundreds of small-scale defects in chemical films would be of valuable assistance to scientists working on

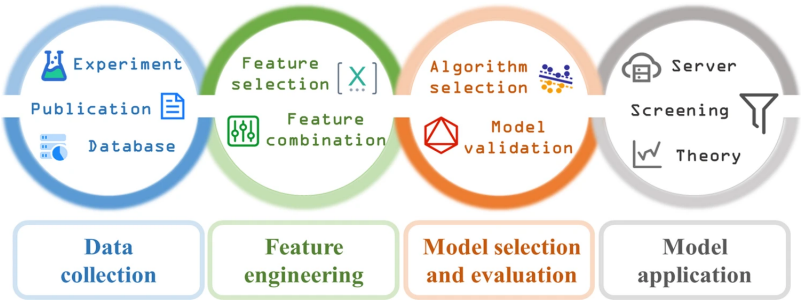

Figure 1. Workflow of Scientific Machine Learning (Xu et al., 2023) [4]

optimizing the electrodeposition procedures and in manufacturing solar cells at increasingly large scales. Moreover, a quantitative model gives a more precise, reliable, and standardized approach to treatment and evaluation of films. Synthesized films are non-homogeneous and contain patches of varying thickness and concentration. Areas of defect within a film can be identified by deformations in color that are representative of experimental error from air bubbles, crystalization, and other phenomenon that affect the thickness and homogeneity of a reaction. Although color changes are fairly obvious and well-controlled in this context, determining the exact area or intensity of film defects when attempting to optimize procedures is challenging.

Progress in the development of machine learning (ML) algorithms shows potential for application in computer-aided automatic data analysis of chemical films through photoanalysis. The multi-level processing of convolutional neural networks (CNNs) have been surpassing humans for several years at many vision tasks, especially those that involve the simultaneous processing of numerous characteristics [1]. CNNs rely on identifying small artifacts such as corners, dots, or grain patterns that comprise a larger structure within an image. As such, it is sensible to assume that the abstract shapes of film defects could be recognized based on a characteristic of hard edges outlining the color of a hole or crack in the deposition. Individual CNNs are limited in the dimensionality of their output data. More robust encoder-decoder models that contain nested convolutions such as the U-Net architecture become necessary for identifying defects in an image, as would be desired for capturing size, intensity, and location. Although deeper and more modern encoder-decoder architectures such as Vision Transformers (ViT) and U2-Nets have become popular recently, for the sake of simplicity and attention to the topic of data fitting, a U-Net was designed after Ronneberger et al. [2]

2. Dataset

U-Net architectures are usually trained on a paired image and a validation mask of equal dimensionality, which identify the ground truth segmentation region for the image. This type of dataset is especially labor-intensive to create. The segmentation area must be manually selected for each image, and, due to the relative complexity of the task, large datasets are necessary to fully train a suitable model. This presents additional issues when analysis of the data requires technical expertise (such as in the case of tumor segmentation) or data is especially abstract (as in the case of our example). For biomedical and scientific semantic segmentation, the creation of dataset is usually the preventative factor to the use of a ML model. Modern computer vision data processing techniques such as image augmentation, as delineated thoroughly in the Ronneberger et al. paper, have lightened the burden of dataset generation substantially, allowing the creation of synthetic training data based on original data. However, even labelling enough samples for image augmentation can be a laborious task. Finally in our specific usage, the density of defects within a film makes manual segmentation especially challenging and subject to human error. False negatives arising from human error are especially damaging, due to introducing contradictory training data that would be further exacerbated by intense data augmentation.

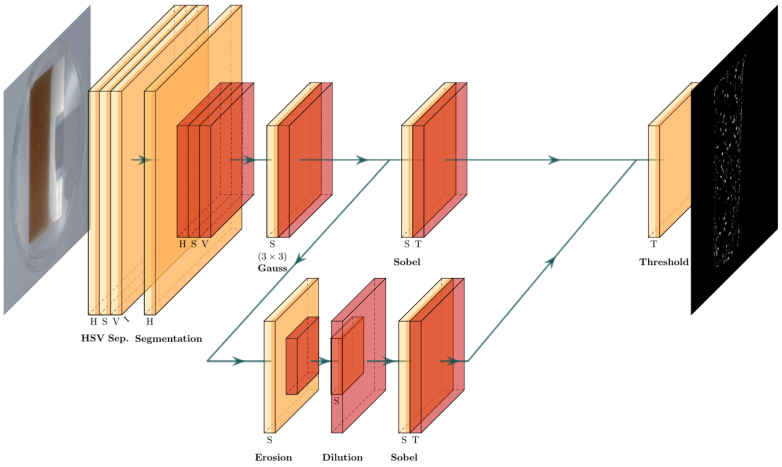

Figure 2. Heuristic Architecture. Layers are 1-dimensional, separated from the HSV channels. No shrinking or pooling convolutions are applied.

3. Heuristic

To address the issue of dataset generation for these sorts of simple segmentation tasks, we created a provisional heuristic relying on basic image transforms to identify defects by color. Images of films were first converted from RGB to HSV, providing deeper insights into the nature of the color without blending. Experimentally, it was shown that the color change exhibited by undesirable regions occurred primarily over the saturation channel, which was elected to be processed with a Sobel 7×7 edge-finding kernel in the vertical and horizontal directions; these were blended to form a complete error mask. Another pass was attempted using the same Sobel procedure, but it was first eroded and dilated using the relevant OpenCV image processing functionality. This served to artificially lower the quality of the image sample such that a less noisy result could be ascertained by controlling the first pass through a bitwise AND operation.

This heuristic model suffers from several problems, making its usage alone unsuitable as a computational model for the desired segmentation task.

Fixed kernel size Due to a fixed, small kernel size, defects larger than 14px in diameter cannot be accurately segmented. This issue became more apparant when image quality increased, and large, intentional defects were created.

Oversensitive In order to accurately segment the full area of true positives, the threshold of sensitivity was lowered such that noise was easily introduced into the image through single pixel color aberrations and compression marks.

High False Positive Rate The use of such a small kernel window led to a highly inaccurate output. The segmentation lacked spacial context (as would be provided through a repeated pooling) and would often report minute lighting differences and reflections as flagrant rows of defects.

Despite significant issues with the model, its high sensitivity showed a very low false-negative rate. As discussed earlier, type II statistical errors are much more confounding to the model. We hypothesized that if the model was deliberately undertrained to a dataset generated by the heuristic, it would continue to be trained on the correctly identified defects, but would struggle to train as quickly on the random noise and false positives. If you cut off training early, you would be left with a model that is not perfect, but essentially pre-trained under complete self-supervision.

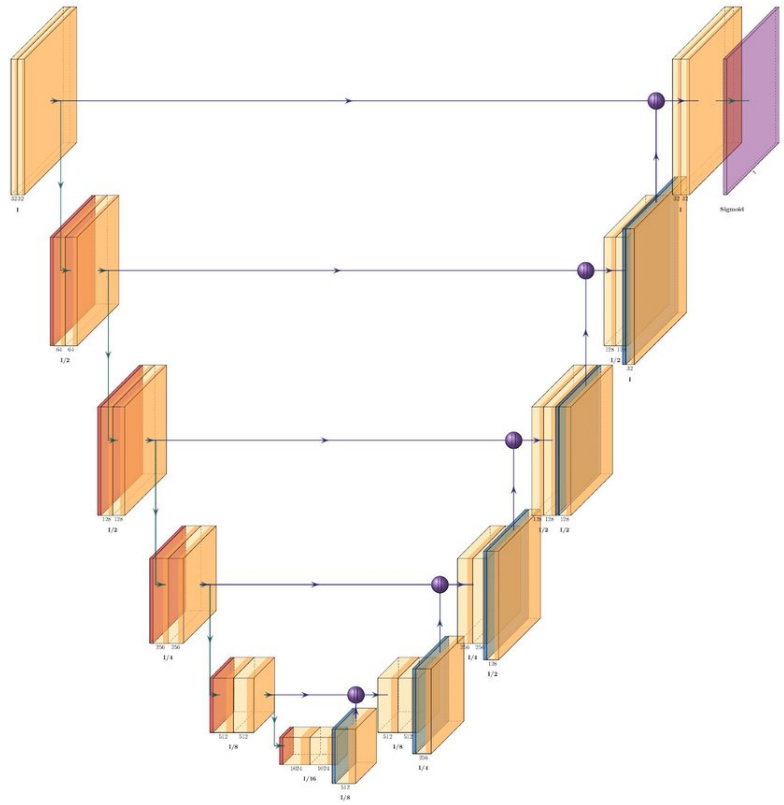

Figure 3. U-Net Architecture. Visualization of multi-level convolution and pooling (Tokime et al., 2019).

4. Neural Model

To train on the heuristic, the U-Net model begins with an initial convolutional block that processes the input data through multiple kernels. Following this, a series of 'Down' modules are sequentially linked, progressively increasing the number of filters with each layer while reducing the spatial dimensions. Each 'Down' module comprises a Max Pooling operation and a double convolution operation, successively abstracting the information and increasing the feature space dimensions. In the subsequent decoder part of the architecture, a set of 'Up' modules symmetrically mirrors the encoder's structure, gradually decreasing the feature space dimensions while expanding the spatial dimensions. These 'Up' modules perform an upsampling operation followed by a double convolution, combining the high-level features learned from the encoder with the detailed spatial information from the decoder. The final module of the architecture consolidates the processed features into the desired output classes of rasterized floats.

This U-Net model's beauty lies in its symmetry and in its capacity to process an input through a series of convolutions, pooling, and upsampling operations, while maintaining a balance between the spatial and feature space dimensions. This ensures high-resolution output, making it a suitable choice for precise biomedical image segmentation tasks.

5. Results

5.1. Implementation

For the experiment our heuristic was programmed using OpenCV image transforms, and the neural model was designed using PyTorch. The implementation is publicly available on the GitHub repository along with the weights. The model has been implemented into the TU analytical chemistry software suite, T.U.D.E.L., where experimental evaluation was performed. Model training was performed on a single Google Colab NVIDIA V100 GPU under the 32GB configuration.

5.2. Training

Training was evaluated for the model using the Adam optimizer with a learning rate of 0.001 for the first five epochs and was reduced to 0.0003 for the remaining epochs. For model evaluation during training, the 10% training set split and the calibration sheet were evaluated based on Binary Cross-Entropy at the completion of each epoch. The weights were saved for the following experiments at the local minimum for calibration sheet loss (epoch 6), although the model continued to train to overfitting. Following epoch 10, the training and testing average loss continued to decrease in lockstep, while the validation loss showed an continuous positive trend from the minimum. Both the validation and test evaluations were necessary to distinguish between the fit on the low quality validation dataset against the high-quality manual segmentation on the calibration sheet. The minimum loss for the validation split was achieved at epoch 18, demonstrating that the model was underfit to achieve the minimum loss over the verified data. Please note more standardized units for segmentation quality were not used for evaluation due to the small scale of defect masks (50% of correctly identified defects were less than 100px in size). Dice Score and Jaccard Score were shown to provide negatively biased and noisy metrics for evaluation.

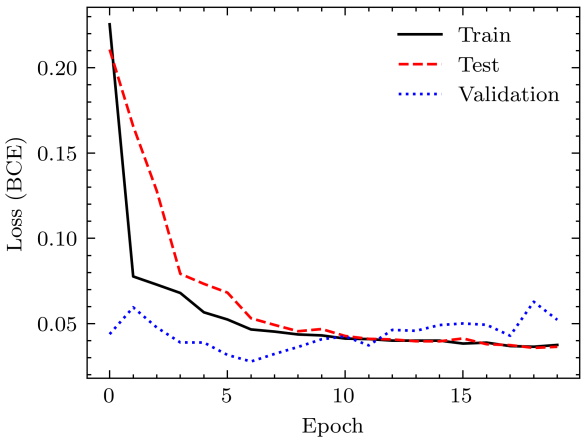

Figure 4. Performance of model by epoch. (Averaged from three training sessions)

5.3. Experiments

After completion of the training, the model was tested and qualitatively evaluated on a separate film dataset. Compared to the original model, the neural model showed significantly less noise and sensitivity than the training heuristic. Furthermore, it displayed good choice and consistency in its segmentation of errors (which are thresholded at 0.2 for the visualization). The model, in contrast to its training data, rarely selects false positives, but instead shows a greater preference for false negatives. This is separately alarming, but an easier problem to combat in the fine-tuning stage of a deeper model. This problem is hypothesized to also be less apparent in more modern U-Net-based architectures, the implementation of which were outside the scope of this learning experiment.

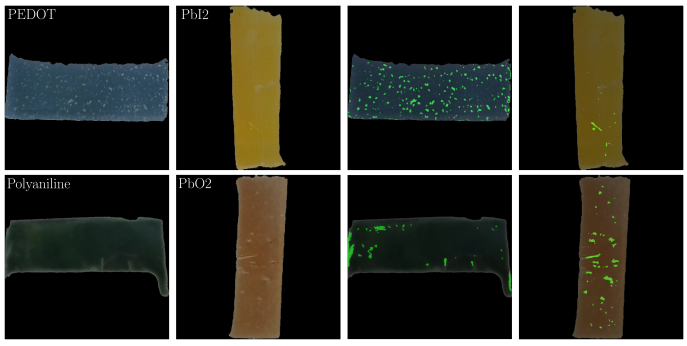

Figure 5. Predictions on the film dataset. Original film samples labeled by film material. Remaining films shown with highlighted segmentation mask.

Furthermore, we ablated a Polyaniline film with 25 micron laser defects in a raster grid. BRAD: THIS SENTENCE IS NOT COMPLETE: This "calibration sheet" was designed to be unambiguously segmented by a human, and a more accurate test for the model's performance on controlled data (as opposed to natural macroscopic films, where there is a great deal of ambiguity as to what

is considered a "defect" which would confound any rigorous testing). The heuristic and neural model were evaluated on this calibration sheet and compared to the ground truth. The heuristic model had extreme trouble with this sample. The heuristic was oversensitive to the minimal color changes that can be most easily recorded under a microscope, leading to essentially complete noise. However, the neural model was able to quite easily segment the regions with zero false positives (other than systemic noise around the edge that can be easily controlled for). However, the neural model also had a number of false negatives, especially near the edge of the colored region. It also exhibited strange behavior in its inaccurate localization, shifting the rightmost columns of defects around 4 pixels to the right, undermining the overlap correlation between the ground truth and prediction. This phenomenon is interesting for further testing, but is also a problem that can be addressed in the fine-tuning stage.

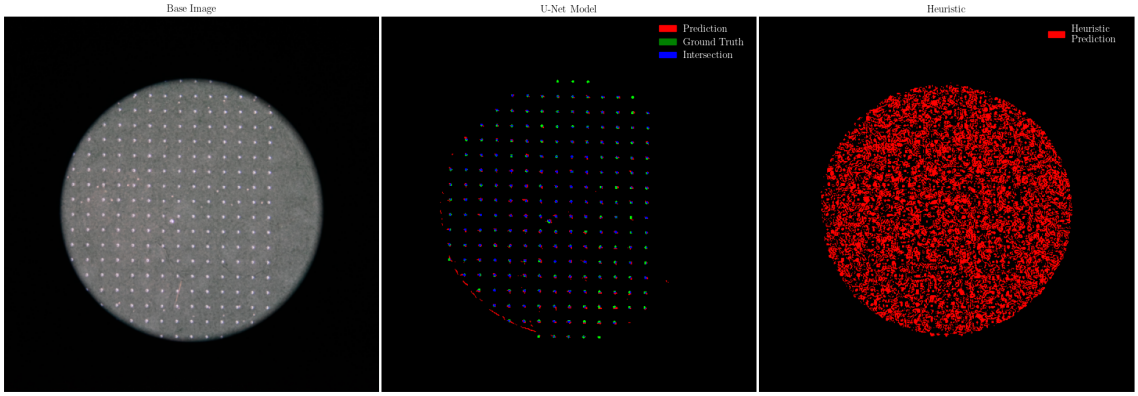

Figure 6. Left: Original image; Middle: Neural prediction against Ground Truth; Right: Heuristic prediction.

6. Conclusions

Our model was successfully trained using entirely a self-supervision mechanism on low-quality data. Furthermore, it was shown to exhibit loss minima on manually generated masks before the loss minima on testing masks, implying that the model achieves best results when undertrained. The use of self-supervision and machine-assisted dataset generation tremendously reduced the amount of manual effort needed to create a practicable model. We hypothesize that this technique is applicable to other segmentation tasks on more difficult datasets. However, its efficacy depends largely on the ability for a dataset to be segmented productively without the use of machine learning, restricting the technique to more controlled segmentation environments such as biomedical and microscopic segmentation, especially in environments where color data can be preserved.

References

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. Imagenet classification with deep convolutional neural networks. *Neural Information Processing Systems*, 25, 01 2012.

[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[3] Tianhao Wu, Zhenzhen Qin, Yanbo Wang, Yongzhen Wu, Wei Chen, Shufang Zhang, Molang Cai, Songyuan Dai, Jing Zhang, Jian Liu, Zhongmin Zhou, Xiao Liu, Hiroshi Segawa, Hairen Tan, Qunwei Tang, Junfeng Fang, Yaowen Li, Liming Ding, Zhijun Ning, Yabing Qi, Yiqiang Zhang, and Liyuan Han. The main progress of perovskite solar cells in 2020–2021. *Nano-Micro Letters*, 13(1):152, Jul 2021.

[4] Pengcheng Xu, Xiaobo Ji, Minjie Li, and Wencong Lu. Small data machine learning in materials science. *npj Computational Materials*, 9(1):42, Mar 2023.