Zno SEM Image Segmentation Based on Deep Learning

Linmao Li

School of Electronic, Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China
lilinmao@sjtu.edu.cn

Abstract. The radius of nanomaterials, which will affect the specific surface area of the nanowires and other functions, is important for the optoelectronic application of nanomaterials. The Scanning Electron Microscopy (SEM) is an effective method to observe the spatial morphology of nanowires. However, the current measurement of topographical features mainly uses manual methods, which will bring about instability errors, especially when measuring a large number of them. Deep learning provides an efficient, fast way to identify and segment nanowire SEM images. Through deep learning methods, the spatial characteristics of nanomaterials can be measured quickly, which aim to explain the relationship between features and radio and television applications from a statistical point of view. In this paper, we design a deep learning image recognition system to measure the radius.

1. Introduction

In nature, green plants convert solar energy into chemical energy through photosynthesis [1]. Inspired by this idea, the so-called artificial photosynthesis uses water splitting reactions store the energy in hydrogen [2]. Hydrogen can be produced by photoelectrochemical (PEC) cells under the sun. This technology, which is a very environmentally friendly, does not produce carbon dioxide compare with consumes energy. The spatial structure of ZnO, which has been widely use in artificial photosynthesis, nanomaterials is very important for the PEC water splitting. ZnO nanowire arrays (NWAs) have direct pathways for charge carriers and a high light absorption [3] [4]. For the PEC cell, the surface area of ZnO NWAS photoanode is one of the key factors for improving the PEC water splitting performance [5]. However, the surface area of the ZnO NWAS is only roughly and quantitatively measured. So it is important to measure the surface area more precisely, which is conducive to quantitative correlation between surface area and PEC performance. If we could measure the radius of the nanowire, the measurement will be more precisely.

Scanning electron microscopy (SEM) is one of the most important tools for observing the spatial geometry of one-dimensional nanomaterials. The SEM image of the nanomaterials is a complex collection of many similar nanostructures, which are interlaced and randomly located, leading to the difficult measurement work. In the traditional measurement method, it mainly relies on manual measurement. There are two fatal flaws. First, manual measurement cannot be automated. Measuring images requires huge resources and time costs, and can only be applied to a small number of samples. Second, manual measurement is highly susceptible to the subjective factors of the surveying personnel, leading to the inconsistency of labeling results among different surveyors. Therefore, a more scientific method is needed for measurement.
In recent years, image segmentation technology has developed rapidly, and can effectively extract desired image features. The hundreds of traditional image segmentation method mainly distinguishes the target on the pixel set, the process is not only complicated, but also has low efficiency and precision. Each segmentation method is limited to a specific segmentation object, and there is no universal one. There is still a lot of works to be done on methods, versatility and anti-jamming capabilities. Since the Alexnet was introduced in 2013, the Convolutional Neural Networks (CNN) has gained wide attention with its excellent implementation effect and strong generalization ability [6], [7], and has formed an image segmentation technology based on CNN, which provides a new image segmentation. CNN-based image segmentation model has strong versatility and anti-interference. Up to now, a variety of CNN-based image segmentation models have been derived, such as FCN [8], MaskRCNN [9] and other classic models. In this paper, the improved MaskRCNN network is used to train and extract the structural features of nanomaterials, and the required nanomaterial structural features are successfully segmented.

Compared with other detection methods, the accuracy and robustness of CNN have been significantly improved. Girshick first proposed an object detection network RCNN [10] based on CNN. For an image, the RCNN first extracts 2000 candidate frames by the SS method, and then inputs the images intercepted by each candidate frame into the CNN to extract features and classify them. Although this method has greatly improved the accuracy rate compared with the previous method, it makes the RCNN have serious repetitive work in extracting features, and the calculation amount is very large. Then, the author proposed Fast-RCNN [11], which performs feature extraction on the entire image only, and realizes feature sharing through a ROI layer, which is greatly accelerated under the premise of maintaining high performance. However, in the Fast-RCNN candidate box extraction is still through the SS method, the speed is still very limited. So, the author also proposed Faster-RCNN to increase the speed. By replacing the SS with a CNN to obtain a candidate box, this improves the overall performance of the network, and the speed is increased by about 10 times. Although FasterRCNN [12] achieves object detection, its ROI layer adopts a very rough mapping method, so there is a large error in positioning. At the time, He KaiMing improved the object positioning capability through the ROI align method, and added the FCN network to the branch layer to achieve high-precision image segmentation.

Since the candidate frame-based algorithms belong to the deep learning model of two-stage detection, the detection speed is difficult to increase again. Later, Redmon et al. proposed a new object detection algorithm YOLO [13], which only needs to detect the feature map once instead of the consume 2000 times, which belongs to the first level detection. As a completely new architecture, he achieved high accuracy and a higher detection speed than the RCNN series.

2. Algorithms and theories

2.1. Image Sharpening

In the process of acquiring the SEM image of the nanomaterial, for various reasons, some images may have edge blurring problems. The edge of an image refers to the area in the image where the gradation changes abruptly. The change in image gradation can be reflected by the gradient of the gray scale distribution. Given a continuous image \( f(x, y) \), the directional derivative takes a local maximum in the normal direction of the edge. Edge blurring can make the boundaries of nanomaterials difficult to determine, which can cause difficulties in labeling work and cause unnecessary labeling errors. In order to reduce the effects of such adverse effects, it is necessary to use image sharpening technology, which aims to make the edges clear. We use several common image processing methods for edge sharpening. Among them, the Laplace operator is sensitive to noise and has good effect in the case of less noise; the Robert operator has better effect on steep low-noise images, especially the image with more edges and negative 45 degrees, but the positioning accuracy is poor. The Prewitt operator extracts the image edge of the gray gradient gradually, but does not consider the influence of the
distance of the adjacent points on the current pixel; the Sobel operator considers the comprehensive factor and has better image processing effect on more noise. The operators are defined as follow:

\[
\nabla f(x, y) = [f(x-1, y+1) + 2f(x, y+1) + f(x+1, y+1) - f(x-1, y-1) - 2f(x, y-1) - f(x+1, y-1)]
\]

(1)

\[
\nabla^2 f(x, y) = 4f(x, y) - f(x+1, y) - f(x-1, y) - f(x, y+1) - f(x, y-1)
\]

(5)

The formula (1) and (2) is about Sobel, (3) and (4) is about Prewitt, (5) is about Laplace.

2.2. Segmentation

The SEM image is first processed by CONV Layers to generate the feature map. Since the Resnet network is deep enough to extract the image features effectively, we use Resnet as the CONV Layers. After obtaining the feature map, the feature map is input into the RPN network. RPN is also a convolution network, which can generate a series of coordinates of X, Y, W, and H. Each coordinate also corresponds to a category, and the classification result. These coordinates form a series of rectangular candidate regions, and each candidate region is subjected to a frame regression operation and filtered by the NMS algorithm. Perform a ROI Align operation on each candidate region obtained by the RPN, and map it more accurately onto the feature map after ROI processing. These feature map are then used for two branches: one is the classification and border regression branches, which are used to generate the regression box and the prediction category; the other is the Mask branch, which uses the FCN algorithm for pixel-level image segmentation. The network structure of MaskRCNN is shown as follow:

![Figure 1. The MaskRCNN framework for segmentation](image-url)
3. Experiment

3.1. Our approach
First, we use image sharpening to pre-process the image to make the edges of the image clear. Then use data augmentation to expand the data set, and then input the processed images into two networks, and the obtained results are filtered by an AND gate to get the final segmented results. The flow is shown in Figure 2.

![Figure 2. The structure of our segmentation](image)

3.2. Data process
The SEM image of nanomaterials has edge blurring problem, which will bring inconvenience to label work, so we use Laplace algorithm for edge sharpening. We compare the effects of various other edge algorithms. It can be seen from the figure that the Prewitt algorithm has little effect on the edge sharpening effect of nanomaterials, and the edges are still blurred, because it is more suitable for gray from the composition of the Prewitt operator.

![Figure 3. (a) Is processed by Laplace operator, (b) is processed by Prewitt operator. (c) Is processed by Sobel operator](image)

SEM image do not belong to this type, Laplace and Sobel algorithms play a better edge sharpening effect. Although the Sobel operator can achieve certain effects, its effect is slightly worse than the Laplace algorithm, and there are still some fuzzy problems for some edges. MaskRCNN is trained on the VOC dataset first, and the trained model is retained for migration learning. The purpose of this is to make the model have a very strong feature extraction ability and improve generalization ability. In the network training, we only retain the feature extraction part of the model, and the rest of the network structure like RPN, classification we will updated parameter.
3.3. The segmentation
A total of 1770 samples were used to create a data set, 1450 samples were used as training sets. The remaining 23 sheets were used as test sets. The result is shown in table a. The results of the identification are measured by the three indicators of RC, PR and MAP. Here PR is precision rate, RC is recall rate, mAP is mean Average Precision. The definition of PR and RC as follows:

\[ PR = \frac{TP}{TP + FN} \]  
\[ RC = \frac{TP}{TP + FP} \]

TP means that positive samples are correctly identified, and FN means that positive samples are identified as negative samples. FN means that a negative sample is identified as a positive sample. A higher PR means that the correct ratio in the sample being detected is higher, but there may be a large number of undetected correct samples. A higher RC means that the higher the proportion of all correct samples detected, but there may be a large number of negative samples that are determined to be positive samples. Map combines these two indicators to more comprehensively determine network performance, but lacks locality.

![Figure 4. the segment result with MaskRCNN and YOLO](image)

There are a large number of nanomaterials in the SEM diagram. These nanomaterials are interlaced and have different shapes. There are many interference features. In image segmentation, these features are inevitably misidentified as we want. After the detection of MaskRCNN, we found that there are a lot of misidentifications in the detection results. The main type of misidentification is FN, and the negative samples are recognized as positive samples. We need to filter the results of the recognition so that the FN is reduced to improve the accuracy of the recognition. So we use another detection technology YOLO to perform object detection on nanomaterials, and use the detection results to combine with MaskRCNN to screen the results. Only when the results are detected as positive samples on both sides, the test is passed, otherwise it is judged as Negative sample. After feature screening, our segmentation achieves the PR improvement of about 10% with a decrease of RC about 4%. The final result is shown in table 1.

| METHOD          | PR  | RC  | MAP |
|-----------------|-----|-----|-----|
| MASKRCNN        | 0.72| 0.91| 0.86|
| MASKRCNN+YOLO   | 0.82| 0.87| 0.84|
4. Conclusion
The segmentation of nanowires are crucial in this paper, especially in the analysis of relationship between the radius and photoelectric performance. We propose an efficient method for SEM image of ZnO detection which include segmentation method and yolo Feature screening. The result show that maskRCNN is ha good robustness and accuracy in SEM image of ZnO. The combination of segment method and feature screening can achieve higher PR which we more care about.

References
[1] Osterloh F. Inorganic nanostructures for photoelectrochemical and photocatalytic water splitting [J]. CHEMICAL SOCIETY REVIEWS, 2013, 42(6):2294-2320.
[2] Zhang, Zhong J. Metal oxide nanomaterials for solar hydrogen generation from photoelectrochemical water splitting [J]. MRS Bulletin, 2011, 36(01):48-55.
[3] Bai Z, Yan X, Chen X, et al. ZnO nanowire array ultraviolet photodetectors with self-powered properties [J]. Current Applied Physics, 2013, 13(1):165–169.
[4] Zhang Y, Yan X, Yang Y, et al. Scanning Probe Study on the Piezotronic Effect in ZnO Nanomaterials and Nanodevices [J]. Advanced Materials, 2012, 24(34):0-0.
[5] He Jun, Ge Hong, Wang Yufeng. Overview of image segmentation algorithms [J]. Computer engineering and Science,2009,31(12):58-61.
[6] Alex Krizhevsky, I Sutskever, G Hinton. ImageNet Classification with Deep Convolutional Neural Networks [J]. Advances in neural information processing systems, 2012, 25(2).
[7] Lecun, Yann, Bottou, Leon, Bengio, Y, Haffner, Patrick. (1998). Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE. 86. 2278 - 2324. 10.1109/5.726791. O
[8] A. Raj, D. Maturana, and S. Scherer, “Multi-scale convolutional architecture for semantic segmentation,” Carnegie Mellon University, Pittsburgh
[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 580–587.
[10] R. Girshick, “Fast R-CNN,” in Proc. Int. Conf. Comput. Vis. (ICCV) Dec. 2015, pp. 1440–1448.
[11] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” arXiv:1506.01497v3 [cs.CV] 6 Jan 2016
[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once:Unified, real-time object detection,” in Proc. IEEE Conf. Comput. Vis.Pattern Recognit. (CVPR), Jun. 2015, pp. 779–788.