Fundamental Research on Electronic Image Recognition of Cylindrical Zno Nanorods Based on Deep Learning

Yuzhou Liang*
School of Eletronic Information and Eletrical Engineering, Shanghai Jiaotong University, Shanghai, China

*Corresponding author e-mail: 373810863@qq.com

Abstract. ZnO is recognized as one of the most important photonic materials in the blue-violet region due to its straight-width band gap and large excitation bonding energy. Since ZnO nanorod array performs superior optical and field emission properties, a lot of efforts have been made in the fabrication of a vertically ordered ZnO nanorod array. The shape and size of ZnO nanorods have a significant effect on PEC property. In order to efficiently recognize and measure the shape and size of ZnO nanorods, a new method based on deep learning model mask r-cnn is proposed to detect cylindrical ZnO nanorods. The SEM images of ZnO nanorods were used as a data set for training. Adjust the size of the bounding boxes that model generated to make it more suitable for the data set. At the same time, improve the NMS (non-maximum suppression) algorithm to reduce the missing detection rate, and achieve a good detection effect on the SEM images of ZnO nanorods.

1. Introduction
ZnO is a semiconductor material with piezoelectric and photoelectric properties. The band gap is 3.37 eV at room temperature and the exciton binding energy is 60 meV [1]. Compared with other common semiconductor photoelectrochemical decomposing agents such as TiO2, WO3, Fe2O3, ZnO is one of the most important photoelectrochemical decomposing water materials due to its high electron mobility, low manufacturing cost and good light trapping performance. The one-dimensional ZnO nano structure has the characteristics of large specific surface area, fast electron transport rate, and quantum confinement effect. In recent years, various devices based on ZnO nanorod array have shown board application in many fields such as quantum dot-sensitized solar cells, photoelectrochemical decomposition of water, gas sensitive devices [2, 3, and 4].

The shape and size of the ZnO nanorods have effects on the PEC property. The ZnO nanorods are usually observed by using scanning electron microscope (SEM), however this will consume a lot of time, manpower and material resources. Using image recognition [5] algorithm to identify ZnO nanorods is an efficient new measurement method. Since the convolutional neural network has achieved great success in image classification [6] on the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [7, 8], object detection algorithms based on deep learning have been widely used in various object detection tasks. In this paper, the SEM images of ZnO nanorods are used as a data set, and the convolutional neural network model mask r-cnn is trained to detect cylindrical ZnO nanorods.
2. Improved Mask r-cnn model

2.1. Mask r-cnn model

Mask r-cnn [9] is an instance segmentation algorithm based on Faster r-cnn [10] proposed by The Kaiming. Mask r-cnn is a convolutional neural network model based on ROIs (regions of interest). The model structure is shown below.

The steps of mask r-cnn are:

1. The input pictures are put into the backbone consists of ResNet101 [11] and FPN [12] network for feature extraction to obtain the feature maps after it is cropped.

2. Each pixel of the feature maps generates a predetermined number of bounding boxes as candidate ROIs.

3. These candidate ROIs are sent to the RPN(regions proposal network) for binary classification (foreground or background) and bounding boxes regression, and some candidate ROIs are filtered out by the NMS [13] algorithm.

4. Do ROIAlign operation on the remaining ROIs to associate the original images with the feature maps.

5. Classification, accurate bounding boxes regression for these ROIs, and generate the masks of the object in the FCN (fully convolutional network) [14].

2.2. Evaluation of the model

The prediction results of the model can be divided into four categories: TP (true positive samples), FP (false positive samples), TN (true negative samples), and FN (false negative samples). Define the accuracy and recall rate of the model as:

\[
p = \frac{TP}{TP + FP} \quad (1)
\]

\[
r = \frac{TP}{TP + FN} \quad (2)
\]

The performance of the model is evaluated by the average accuracy. Plot the curve with accuracy and recall and the average accuracy is the area under the curve, as shown in formula (9):
$$AP = \int_0^1 p(r)dr$$ (3)

2.3. Optimization of model

Feature maps with size of 16×16 are obtained after convolutional layers, and each pixel of the feature map generates bounding boxes with a determined quantity and size. Each pixel in the original model generates nine bounding boxes of size (128, 256, and 512) and an aspect ratio of (0.5, 1, and 2). Considering that many small-sized nanostructures exist in the images, increase the bounding boxes with size of 32 and 64. Considering that most of the ZnO nanorods are relatively slender, adjust the aspect ratio of the bounding boxes.

After the bounding boxes is generated, each object corresponds to multiple bounding boxes, but each object in the final detection only corresponds to one bounding box, so the NMS algorithm is used to filter out some bounding boxes. After the bounding boxes are sent to the RPN, each box obtains a confidence score. The NMS algorithm filters out some bounding boxes according to the confidence scores and the IOU between different bounding boxes. The IOU between the two candidate boxes is calculated by the following formula:

$$IOU = \frac{\text{Intersection}}{\text{Union}}$$ (4)

The steps of NMS are:

1. Select a box with the highest confidence score from the bounding boxes set and put it in the final ROIs set, and remove the box from the original set.
2. Calculate the IOU of the box with the highest score and the remaining boxes in turn. If the IOU is greater or equal to a threshold, then the box is considered to correspond to the same object as the box with the highest score, so the box is removed.
3. Repeat the above steps until there are no remaining boxes in the original bounding boxes set.

The NMS algorithm has disadvantages. The removal of the bounding boxes depends on the threshold, but it is difficult to set a suitable threshold. There are many overlapping nanorods in the images. If the threshold is set inappropriately, it is easy to filter out some bounding boxes that should not be filtered out.

Considering this situation, the SOFT-NMS algorithm [12] is used instead of the NMS algorithm. Different from the NMS algorithm, when the IOU of the box \(b_i\) and the box \(b_m\) with the highest score is greater than the threshold \(T\), the SOFT-NMS algorithm does not directly filter out the box \(b_i\), but uses a penalty coefficient \(S_i\) to reduce the confidence of the \(b_i\). \(S_i\) is calculated as follows:

$$S_i = \begin{cases} S_i, & IOU(b_m, b_i) < T \\ \frac{IOU(b_m, b_i)^2}{\sigma}, & IOU(b_m, b_i) \geq T \end{cases}$$ (5)

Where \(\sigma\) is usually set as 0.5. After multiple iterations, the bounding boxes with low scores will be filtered out.

3. Experiment

This paper selects tensorflow as the framework for to implement mask r-cnn and use the method of transfer learning to pre-train the model on the COCO dataset. Although there are no ZnO nanorods on the COCO dataset, the model can learn a lot of other feature information after pre-training, which can help the model to converge faster. Finally, train the model with the ZnO nanorods data set.
3.1. Dataset
The SEM images of ZnO nanorods were used as the data set, which included a training set of 177 images and a validation set of 45 images. The size of the images were $512 \times 384$ after being cropped. The ZnO nanorods contained in the data set are mainly cylindrical and prismatic. Use VIA as the labeling tool to label the cylindrical ZnO nanorods in the images to generate the .json dataset file.

3.2. Size of bounding boxes
Using different aspect ratio to obtain the performance of the model:

| aspect ratio | (0.5, 1, 2) | (0.3, 1, 3) | (0.2, 1, 5) |
|--------------|-------------|-------------|-------------|
| AP           | 0.739       | 0.762       | 0.761       |

When the aspect ratio is (0.3, 1, 3), the model has better performance, so the aspect ratio used in the training is (0.3, 1, 3).

3.3. Detection result
The output of Mask r-cnn are the type, bounding boxes and mask of the detected object, the position and size of the object are described by the bounding boxes, and the outline of the object is described by the mask.
3.4. Comparison of model performance before and after optimization

Use the data set to train the model before and after optimization. Obtain the performance comparison.

![Figure 4. Model performance. (a). Before optimization. (b). After optimization.](image)

It can be seen that the improved model achieves better performance, higher AP, and the model has better robustness. For the same image, the detection results before and after optimization is as follows:

![Figure 5. Detection results. (a). Before optimization. (b). after optimization.](image)

It can be seen that the improved model achieves a better detection effect and effectively reduces the missed detection rate for the same image.

4. Conclusion

The deep learning method can effectively identify ZnO nanorods and achieve a good detection effect in the SEM images. By adjusting the size of the bounding boxes and optimizing the NMS algorithm, better performance is achieved on the same data set, and the model is more suitable for the recognition task of ZnO nanorods. And using deep learning to identify nanostructures can be used for many other applications. Knowing the SEM magnification, the size of the nanostructure can be calculated according to the size of the bounding boxes and the pixels of the images. In future research work, more applications of deep learning in nanostructure identification and measurement will be carried out.
References

[1] Fung S, Xu X L, Zhao Y W, et al. Gallium/aluminum interdiffusion between n-GaN and sapphire [J]. Journal of Applied Physics, 1998, 84 (4) :2355-2357.

[2] Bu Y Y, Chen Z Y. Effect of hydrogen treatment on the photoelectrochemical properties of quantum dots sensitized ZnO nanorod array [J]. Journal of Power Sources, 2014, 272 (3) :647-653.

[3] Xu S, Wang Z L. One-dimensional ZnO nanostructures: solution growth and functional properties [J]. Nano Research, 2011, 4 (11): 1013-1098.

[4] Udom I, Ram M K, Stefanakos E K, et al. One dimensional-ZnO nanostructures: synthesis, properties and environmental applications [J]. Materials Science in Semiconductor Processing, 2013, 16 (6): 2070-2083.

[5] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]// Proc of International Conference on Neural Information Processing Systems. Curran Associates Inc., 2012: 1097-1105.

[6] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C]// Advances in neural information processing systems. 2012: 1097-1105.

[7] Deng J, Dong W, Socher R, et al. Imagenet: A larger-scale hierarchical image database [C]// 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.

[8] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge [J]. International journal of computer vision, 2015, 115(3): 211-252.

[9] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.

[10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [C]// Advances in Neural Information Processing Systems. 2015:91-99.

[11] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning [C]// Thirty-First AAAI Conference on Artificial Intelligence. 2017.

[12] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection [Z/OL]. arXiv preprint,arXiv:1612.03144v2 (2017-04-19).

[13] NEUBECK A, VAN GOOL L. Efficient non-maximum suppression [C]// Proceedings of the 18th International Conference on Pattern Recognition. 2006, 3:850-855.

[14] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [A]. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. [C]. 2015:3431-3440.