Segmentation of Cracked Silicon Wafer Image Based on Deep Learning

Xingxing Li*, Panpan Yin, Chao Duan, Ningxing Wang

School of Engineering, Guangzhou College of Technology and Business, Foshan, 528138, China

*Corresponding author: wslxx@jxstnu.edu.cn

Abstract. With the development of the new energy industry, a large number of silicon wafers need to be tested for production quality through the automation industry. The development of deep learning technology has brought huge technological improvements to the industrial quality inspection industry. Through the image segmentation technology based on deep learning, it can accurately divide the defects existing in the silicon wafer. In this paper, the UNet deep learning network is used to segment the hidden cracks in the silicon wafer. The network can extract the shallow semantic features in the silicon wafer well. It uses 5,000 samples collected on the industrial site as the training set, 1,000 pieces the sample is used as the test set, and the segmentation accuracy IOU can reach 58.7%.

Keywords: Deep learning technology, image segmentation technology, silicon wafer, UNet

1. Introduction

As the deep learning vision technology shines in the industry, some difficult problems that have appeared in the industry in the past can be solved. For example, in the industry, there are often defects of indeterminate shape, color, size, and texture, such as cracks, white spots, black spots, internal cracks, internal collapse, dirt and so on. These abnormal blemish samples rely on human eyes to observe it will consume a lot of manpower and material resources. If computer vision is used, part of the human eye judgment task is handed over to the computer, which will greatly liberate labor costs. However, if the defects that appear only rely on the previous manual features, some of the irregular shapes and weak imaging defects are difficult to solve, and traditional algorithms have problems such as low performance, high error detection rate, and noise sensitivity. With the help of deep learning algorithms and relying on a large amount of image data, the image is handed over to the network to automatically learn and extract the deep features of the image, which can improve the robustness of the results [1].

Image segmentation technology is to solve the specific location information of the specific target in the image. It is an accurate pixel-level classification problem. It is necessary to classify each pixel in the image. The purpose of this article is to segment specific defects such as cracks in silicon wafers at the pixel level, and apply specific colors to masks to visualize the cracks appearing in specific locations in the image.
This article uses the deep learning UNet [2] segmentation model as the experimental baseline. After the UNet [2] network was proposed, it has been widely used for the segmentation of medical images. Nowadays, there are more and more natural image understandings. Semantic segmentation and target detection SOTA models began to pay attention to and use U-shaped structures, such as semantic segmentation Discriminative Feature Network (DFN) (CVPR2018) [3], target detection Feature Pyramid Networks for Object Detection (FPN) (CVPR 2017) [4], etc.

In this paper, the UNet network is selected as the baseline. The main reason is that industrial images are similar to medical images, both of which are two-channel grayscale images, and the hidden flaws are similar to the shape of tumors and bone cracks in medical images, and such as flaws. The shallow semantic features such as the edge and contour of the image play an important role in the correctness of pixel classification. On the contrary, some deep network models cannot better capture the shallow semantic features. Therefore, this article chooses the UNet network structure as the deep learning network baseline. Because the purpose of this experiment is to distinguish between cracked and non-cracked areas, a pixel-level binary classification problem can be seen. This article is followed by the sigmoid function after the output of the network. Using the Binary Cross Entropy loss function, the final segmentation accuracy of IOU is 58.7% on the test set.

2. Principle

2.1. UNet network model structure

The structure of UNet mainly has two biggest characteristics, U-shaped structure and skip-connection. The network structure is shown in Figure 1.

![Figure 1. Schematic diagram of UNet network model structure.](image)

The network architecture is as shown in the figure above, which can be described as consisting of a contraction path (left) and an expansion path. The shrinking path is the same as the traditional convolutional network. It consists of an unfilled convolution with a convolution kernel size of 3×3. After each convolution, it passes through the ReLU function and the largest pool with a size of 2×2 and a step.
distance of 2. Composition. This maximum pooling is the process of down-sampling. After down-sampling, the channels are doubled. The expansion path consists of a 2×2 upper convolution, the output channels of the upper convolution are half of the original, and then are connected in series with the corresponding feature map (after cropping) (to obtain channels of the same size as the original), and then pass through two sizes of 3×3 convolution and ReLU function. The corresponding cropping feature map is necessary, because there will be loss of boundary pixels in the process of our convolution. In the last layer, the desired target type is obtained through convolution with a convolution kernel size of 1×1. In this network, there are 23 convolutional layers.

The network adopts the common Encoder-Decoder structure, and adds to the original structure the operation of directly intercepting information from the encoder and placing it in the decoder. This operation can effectively retain the edge detail information in the original image and prevent excessive edges. Loss of information. In order to ensure seamless splicing of the output segment mapping, the size of the input image needs to be carefully selected to ensure that all Max Pooling operations are applied to layers with even x-size and even y-size.

The network does not have any fully connected layers, and only uses the effective part of each convolution, that is, the segmentation map contains only pixels that can obtain complete context in the input image. This strategy allows seamless segmentation of arbitrarily large areas through the overlapping block strategy image. In order to predict the pixels in the border area of the image, the missing context is extrapolated by mirroring the input image. This tiling strategy is important for applying the network to large images, otherwise the resolution will be limited by GPU memory.

2.2. The dataset has a small amount of available data
There is very little training data available, and too much data enhancement is used by applying elastic deformation to the available training images. This allows the network to learn the invariance of such deformations without needing to see these transformations in the annotated image corpus. This is especially important in biomedical segmentation, because deformation was once the most common change in tissue and can effectively simulate real deformation. Dosovitskiy et al. have demonstrated the added value of learning invariant data in the context of unsupervised feature learning.

A smooth deformation is generated by using a random displacement vector in a 3×3 coarse grid. The displacement is sampled from the Gaussian distribution. The Gaussian distribution has a standard deviation of ten pixels. The offset of each pixel is obtained by bicubic interpolation.

2.3. Binary Cross Entropy
The task of this paper is to segment the hidden and non-hidden cracked areas by pixels, so it is a two-class classification problem. The Binary Cross Entropy loss function is used for training. The function expression of Binary Cross Entropy is shown in formula 1.

\[
Loss(y', y) = \sum_{i=0}^{k=1} y_i \cdot \log y_i - (1 - y_i) \cdot \log(1 - y_i)
\]

(1)

When using, you need to add the Sigmoid function in front of the network layer, because there are only positive and negative examples, and the sum of the probabilities of the two is 1, then only one probability needs to be predicted.

2.4. Experimental data
The experimental data in this paper uses the hidden crack image taken by optical imaging in an industrial scene as shown in Figure 2 below.
In Fig. 2, the upper left corner is a normal crack image, the upper right corner is a crack image with a small round hole, and the lower left corner is a long strip of crack, which spans the entire silicon wafer. Compared with the cracked image in the lower left corner, the ordinary line mark in the lower right corner is caused by industrial cutting. It also spans the lateral area of the entire silicon wafer. The shape is relatively regular, and there are no hidden cracks and depressions. Black mark shape, this kind of line mark sample is easy to cause great interference to the detection, industrial-grade cracked images show strange shapes, and often the industry collects real samples with a long period of time and samples are in short supply.

3. Experiments
This article sorted out 5000 images of hidden training samples, 1000 test samples, and used the UNet deep learning segmentation model for training. The test segmentation results are shown in Figure 3 below.

Figure 2. Example image of wafer cracking.

Figure 3. Example image of wafer segmentation result.
Figure 3 shows the result of silicon segmentation. The hidden crack appears in the lower left corner of the image. The result of image segmentation is wrapped in red lines. Using the deep learning segmentation algorithm, not only can this long strip of hidden crack be segmented, but also divide small cracks, dot-shaped cracks, and cracks appearing at the edges. This article uses 5000 training images and 1000 test images, and finally the IOU accuracy on the test set can reach 58.7%.

4. Conclusion
In this paper, the UNet segmentation algorithm is applied to the problem of silicon chip cracking. This segmentation method can automatically speed up the shallow semantic information of the image, and by fusing the features of down-sampling and up-sampling across the handover, it can solve the defect segmentation problem of industrial images. The experiment achieved 58.7% of the IOU segmentation results on the test set.

Acknowledgments
This work was financially supported by fund project, that is, Young Talents in Higher Education of Guangdong, China, (No. 2021KQNCX125 and No. 2019KQNCX232).

References
[1] Li X, Duan C, Zhi Y, et al. Wafer Crack Detection Based on Yolov4 Target Detection Method [J]. Journal of Physics Conference Series, 2021, 1802(2): 022101.
[2] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation [J]. Springer, Cham, 2015.
[3] Yu C, Wang J, Peng C, et al. Learning a Discriminative Feature Network for Semantic Segmentation [C]/ 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.
[4] Lin T Y, Dollar P, Girshick R, et al. Feature Pyramid Networks for Object Detection [C]/ 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2017.