Wafer Crack Detection Based on Yolov4 Target Detection Method

Xingxing Li*, Chao Duan, Yan Zhi and Panpan Yin

Department of electronics and information engineering, Guangzhou College of Technology and Business, FoShan, 528138, China

*Corresponding author: wslxx@jxstnu.edu.cn

Abstract. In recent years, the breakthrough of deep learning technology has brought a qualitative leap in the field of industrial detection. In the past, defects that are difficult to detect by traditional image methods can be detected by deep learning automatic feature extraction to find the expression of difficult samples. In the field of industrial silicon wafer defect detection, with the help of deep learning target detection algorithm, it can effectively adapt to different size, illumination, depth, length of defects. In this paper, yolov4 target detection algorithm is applied to silicon wafer crack problem. Yolov4 target detection algorithm uses many training skills, such as weighted residual connection (WRC), cross stage partial connections (SCP), cross Mini batch Normalization (CmBN), self-adaptive training (SAT) and mish activation make yolov4 better trained on a GPU, and can achieve 98.23% detection accuracy on the wafer crack detection data set.

Keywords: Deep learning, yolov4, silicon wafer, crack.

1. Introduction
With the deep learning of visual technology in industry, some of the intractable problems in the past can be solved. For example, in the industrial sector, there are often defects of shape, color, size and texture, such as cracks, white spots, black spots, internal cracks, collapse, dirt and so on. It will cost a lot of manpower and material resources if these abnormal defect samples are observed by human eyes. If a part of the human eye judgment task is handed over to the computer with the help of computer vision, the labor cost will be greatly liberated. However, if the defects only rely on the previous manual features, some of the defects with irregular shape and weak imaging are difficult to solve, and the traditional algorithm has the problems of low performance, high false detection rate and noise sensitivity. With the help of deep learning algorithm and relying on a large number of image data, the image is handed over to the network to automatically learn and extract the depth features of the image, which can improve the robustness of the results.

Object detection is to solve the problem of what and where the object in the image is. Input an image and output the category and position of each object in the image. The position is represented by a box containing the object. There are two kinds of target detection architecture, one is two stage, the other is one stage. The difference is that two stage has region proposal process, which is similar to an audition process. The network will generate location and category according to candidate region,
while one stage can directly generate location and category from image. Yolov1 [1] regards detection as a regression problem, uses a network output location and category to implement a unified system, which is one stage from the perspective of detection. The detection process is to input a multi-target image, divide the image into multiple grids, and get the classification probability of each grid and the box + confidence degree of each grid prediction through the network. For each box, the probability is multiplied by the confidence score, which is the confidence score of each box specific to each class, and the output location and category information. The disadvantages are also obvious: one cell can only predict two boxes and one category, which will inevitably limit the number of predictions; it is difficult to expand: the model can predict the bounding box according to the data, and it is difficult to extend it to objects with new or unusual aspect ratio or configuration. Since the output layer is a fully connected layer, Yolo training model only supports the same input resolution as the training image. The network loss is not specific: no matter the size of the boundary box, the loss function is used to approximate the detection performance. The object IOU error and the small object IOU error have close contribution to the loss value in the network training, but for the large boundary box, the small loss has little impact on the IOU, and for the small boundary box, the small error has a greater impact on the IOU, thus reducing the positioning accuracy of object detection. Aiming at the low performance of yolov1 detection, yolov2 proposes a joint training algorithm, which can train target detector on detection and classification data. We use the labeled detection image to learn accurate positioning, and use classification image to increase its "vocabulary" and robustness. In view of the current detection task is limited by the data set label, yolov2 proposed a combination method of ImageNet and coco data sets, as well as a joint training method. After training yolov2, the model is called yolo9000 [2]. Yolov3 has made some small improvements on the basis of yolov2. Yolov3 bounding box prediction: anchor box is used to predict the bounding box in positioning task, and logical regression is used to predict a score object score for each bounding box in yolov3. The score is based on the overlap between the prediction box and the object. If the overlap degree of a frame is higher than that of other frames, its score is 1, ignoring those frames that are not the best and whose overlap is greater than a certain threshold (0.5); category prediction: like yolov2, yolov3 still adopts multi label classification; yolov3 adopts multi-scale prediction, and uses the new network darknet-53 to extract features [3]. Yolov4 proposed an efficient and powerful target detection model. It enables everyone to train ultra fast and accurate target detectors using 1080Ti or 2080Ti GPU, which is an improvement on yolov3. Its improvement method is to summarize almost all detection techniques, and then put forward some skills, and then through screening, permutation and combination, such as weighted residual connections (WRC), cross stage partial connections (CSP), cross Mini batch normalization (CmBN), self-adaptive training (SAT), mish activation, mosaic data aggregation, DropBlock regularization, CIoU loss [4, 5].

2. Principle

**Yolov4 network structure.** The Yolov4 network model structure is actually improved on the basis of yolov3. When the input is 416x416, the characteristic structure is shown in Figure 1:
The backbone feature extraction network Backbone has two improvements: (1) The feature extraction network: The original DarkNet53 is replaced with CSPDarkNet53, and all the feature extraction module activation functions are replaced with Mish activation. The Yolov3 network structure uses Darknet53 as the backbone, which consists of a series of residual network structures. In Darknet53, there is a residual block module, which is composed of a stack of down-sampling and multiple residual structures. Darknet53 is a combination of residual block modules. In YOLOV4, it has made certain modifications to this part. One is to modify the activation function of DarknetConv2D from LeakyReLU to Mish activation function. The second is to modify the structure of resblock body and use the CSPnet structure. At this time, Darknet53 in YOLOV4 has been modified to CSPDarknet53, and the module is shown in Figure 2:

The structure of CSPnet is not complicated. It splits the stack of the original residual block into two parts: the main part continues to stack the original residual block; the other part is like a residual edge. After a small amount of processing is directly connected to the end, it can be considered that there is a large residual edge in the CSP.
**YoLov4 feature pyramid.** For those plug-in modules and post-processing methods that only increase a small amount of reasoning cost but can significantly improve the accuracy of object detection, we call them "special price bags". Generally speaking, these plug-in modules are used to enhance some attributes in the model, such as expanding the acceptance domain, introducing attention mechanism or enhancing the ability of feature integration. Post processing is used to filter the prediction results of the model.

In order to use different object scales, YoLov4 adds a feature pyramid module. The SPP and PANet network structures are used. Except for the CSPDarknet53 and Yolo Head structures, they are all feature pyramid structures. The SPP structure is mixed in the last feature of CSPDarknet53. In the convolution of the layer, after performing three DarknetConv2D_BN_Leaky convolutions on the last feature layer of CSPDarknet53, four different scales of maximum pooling are used for processing, and the maximum pooling core size is 13x13, 9x9, 5x5, 1x1, the feature pyramid module can greatly increase the receptive field and isolate the most significant contextual features. PANet is an instance segmentation algorithm of 2018. Its specific structure is to achieve the purpose of repeatedly improving features. In YOLOV4, it mainly uses the PANet structure on three effective feature layers to detect different image grids. Large and small objects to enhance the adaptability of the network to various objects.

Attention modules often used in object detection are divided into channel attention and point attention. The two attention models are squeeze and exception (SE) and spatial attention module (SAM). Although these modules can improve the function of resnet50 in ImageNet image classification task by 1% to top-1 accuracy, and only need to increase the calculation workload by 2%, it will increase the reasoning time by about 10% on GPU, so it is more suitable for mobile devices. However, for Sam, it only needs to pay 0.1% extra calculation to improve the top-1 accuracy of resnet50-se by 0.5% in ImageNet image classification task. Best of all, it doesn't affect the speed of reasoning on the GPU at all.

In feature integration, with the popularity of multi-scale prediction methods such as fuzzy neural network, many lightweight modules integrating different feature pyramids have been proposed. This type of module includes SFAM, ASFF and BiFPN. The main idea of SFAM is to use se module to weight multi-scale feature map at channel level. ASFF uses Softmax as the point by point hierarchical weighting, and then adds feature maps of different scales; BiFPN uses multi input weighted residual connection to reweight the scale hierarchy, and then adds feature maps of different scales.

3. Experiment

In silicon wafer crack detection, the shape of crack defects in silicon wafer is uncertain, including long crack, short crack and round hole crack, as shown in Figure 3. At the same time, there are also background interference on crack detection. For example, the line marks left in the industrial cutting of silicon wafer show a uniform state of "one", which has many similarities with crack, including shape and contour, which is easy to detect cracks. It is a strong interference. The reason why human eyes can distinguish hidden crack and cutting line mark mainly depends on the uneven color and contour of hidden crack concave; there is a large area of gap at the edge of silicon chip, which is easy to distribute hidden crack defects in the gap. No matter using depth learning algorithm or traditional image algorithm, it is easy to detect the gap as hidden crack, resulting in over inspection. In view of the fact that it is difficult for traditional image algorithms to have a unified expression for so many kinds of crack features, this paper uses yoLov4 target detection algorithm to extract target features automatically by using depth network, and obtains a robust feature expression formula under the distribution of silicon crack features.
Figure 3. The cracked image and common line mark.

In Figure 3, the upper left corner is a common crack image, the upper right corner is a cracked image with a small round hole, and the lower left corner is a long strip crack, which spans the whole silicon wafer. Compared with the cracked image in the lower left corner, the common line mark in the lower right corner is caused by industrial cutting. It also spans the transverse area of the silicon wafer, and the shape is regular, and there is no black mark like hidden crack depression. This kind of line trace sample is easy to cause great interference to the detection.

In this experiment, the data set is divided into training set and test set according to the ratio of 0.8 and 0.2. There are 829 images in the training set and 210 in the test set. On the test set, the evaluation index of target detection is used to test the effect of target detection. The test results are as follows:

| IOU Threshold | precision | recall | F1-score | Average IOU | map     |
|---------------|-----------|--------|----------|-------------|---------|
| map@75        | 0.64      | 0.66   | 0.65     | 52.90%      | 53.48%  |
| map@50        | 0.96      | 0.99   | 0.97     | 74.38%      | 98.23%  |

It can be seen from table 1 that when the model IOU threshold is set to 0.5, the map can achieve 98.23% accuracy on the test set, and the coverage area of IOU can reach 74.38%, indicating that the crack can be basically detected, and the size of the target area can be more than 70%. The detection effect is shown in Figure 4.
Figure 4. The detection effect of cracked target.

As can be seen from Figure 4, yolov4 target detection can successfully detect the ordinary long strip cracks, small cracks, as well as the cracks distributed in the gap and some irregular cracks with concave convex feeling on the silicon wafer, and will not cause over inspection on the samples with line marks. In these test samples, the probability of missed detection is very small, mainly due to the existence of a certain probability of over detection, and the risks are mainly divided into the following four types: 1. Large area of target frame will cover the target; 2. There is a very small pixel level hidden crack; 3. There is a risk of over detection of large lines across the whole image; 4. There are some controversial gaps or cracks detected. In general, yolov4 has good robustness for industrial crack detection.

4. Summary
In the process of silicon wafer crack detection, the shape of crack defect in silicon wafer is uncertain, and the line mark causes great interference to the detection. Yolov4 target detection algorithm is used to detect the crack on industrial silicon chip, which can adapt to different sizes and different length of cracks, and can also guarantee that the deep and shallow line marks will not be over detected. The algorithm has good robustness and can be applied to silicon wafer crack detection Test items.

Acknowledgments
This work was financially supported by fund project, that is, Young Talents in Higher Education of Guangdong, China (No. 2019KQNCX231).

References
[1] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection [J], 2015.
[2] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger [C] // IEEE Conference on Computer Vision & Pattern Recognition, IEEE, 2017: 6517-6525.
[3] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement [J]. arXiv e-prints, 2018.
[4] Bochkovskiy, Alexey, Wang, Chien-Yao YOLOv4: Optimal Speed and Accuracy of Object Detection [J]. 2020.
[5] Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP (99): 2999-3007.