IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

SEMICONDUCTOR WAFER DEFECTS USING YOLO V5 MODEL

Nagi Siva Kumar 2. Rajan Kumar Mahato 3. Dr. S. Jana
1,2 UG Students Electronics & Communication Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology
3 Professor Electronics & Communication Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology

ABSTRACT

Wafer defect detection is an important step in the semiconductor manufacturing process because it ensures the final product's quality and reliability. Traditional wafer defect detection methods can be slow and error-prone, and they may be incapable of detecting small or complex defects. Machine learning algorithms such as convolutional neural networks (CNNs) have been used to detect wafer defects with promising results in recent years. YOLO (You Only Look Once), a popular object detection algorithm based on CNNs, is one such algorithm. YOLO has gone through several iterations, including YOLOv1, YOLOv2, YOLOv3, YOLOv4 and, YOLOv5 are most recent one. The YOLOv5 algorithm is the most recent version and powerful, and it includes several improvements over previous versions, such as a new object detector and a more efficient network architecture. In this research, we use YOLO v5 to detect wafer defects and evaluate its performance on a wafer image dataset. In this paper, we compare the accuracy of different YOLOv5 models. Our results show that YOLOv5x can detect wafer defects with high accuracy and speed, outperforming traditional wafer defect detection methods.

INTRODUCTION

The semiconductor industry has been experiencing rapid growth in recent years, and as a result, the demand for high-quality and reliable products has increased significantly. Wafer defect detection is a crucial step in the semiconductor manufacturing process, as it ensures the final product's quality and reliability. Traditional wafer defect detection methods can be slow, error-prone, and may not be capable of detecting small or complex defects. In recent years, machine learning algorithms such as convolutional neural networks (CNNs) have been used to detect wafer defects with promising results. One of the most popular object detection algorithms based on CNNs is YOLO (You Only Look Once). YOLO has gone through several iterations, including YOLOv1, YOLOv2, YOLOv3, YOLOv4, and the most recent and powerful version, YOLOv5. The YOLOv5 algorithm includes several improvements over previous versions, such as a new object detector and a more efficient network architecture. In this research, we use YOLOv5 to detect wafer defects and evaluate its performance on a wafer image dataset. The goal of our research is to evaluate the performance of YOLOv5 in detecting wafer defects and compare it to other state-of-the-art algorithms. We will use a dataset of wafer images with known defects and apply the YOLOv5 algorithm to detect and classify the defects in the images. We will then compare the performance of different YOLOv5 models, such as YOLOv5s, YOLOv5l, YOLOv5m, YOLOv5x, and YOLOv5x6 using metrics such as precision, recall, and mean average precision (mAP o.5). Previous research has explored the use of various machine learning algorithms to detect wafer defects, including convolutional neural networks and support vector machines. However, there is still much to learn about the effectiveness of YOLOv5 algorithm. We anticipate that our research will provide valuable insights into the effectiveness of YOLOv5 in detecting wafer defects and contribute to the development of more effective and efficient algorithms for this important task.

Fig 1 Wafer with Defects

PROPOSED MODEL

We are using yolov5 network structure model for image recognition. Yolov5 is an object detection algorithm developed by Ultralytics, which is capable of detecting objects in real-time with high accuracy. It is the fifth version of the You Only Look Once (YOLO) series, which is one of the most popular object detection algorithms used in computer vision. Yolov5 uses a deep neural network with a backbone architecture based on CSPNet, which stands for Cross-Stage Partial Network, and a detection head architecture based on YOLOv3. It is designed to be faster and more accurate than its predecessor, YOLOv4, while being more flexible and easier to use. The algorithm has been trained on a wide range of datasets, including COCO, Pascal VOC, and Crowd Human, among others. It is also highly customizable, allowing users to train the network on their own datasets and fine-tune it for specific applications. Yolov5 is available in four different versions, each optimized for different hardware configurations: Yolov5s, Yolov5m, Yolov5l, and Yolov5x. The Yolov5s is the smallest and fastest version, while the Yolov5x is the largest and most accurate but also the slowest. The algorithm has a wide range of applications, including object tracking, people counting, autonomous vehicles, and surveillance systems, among others. It has been widely adopted in the computer vision community and has been used in various real-world applications. Overall, Yolov5 is a powerful and versatile object detection algorithm that is capable of detecting objects in real-time with high accuracy. Its flexibility and ease of use make it a popular choice for researchers and practitioners in the computer vision field.

YOLOv5 NETWORK ARCHITECTURE

The YOLOv5 network architecture is a state-of-the-art object detection model that achieves high accuracy and speed. It was developed by Ultralytics and released in June 2020 as an upgrade to the previous YOLOv4 model.

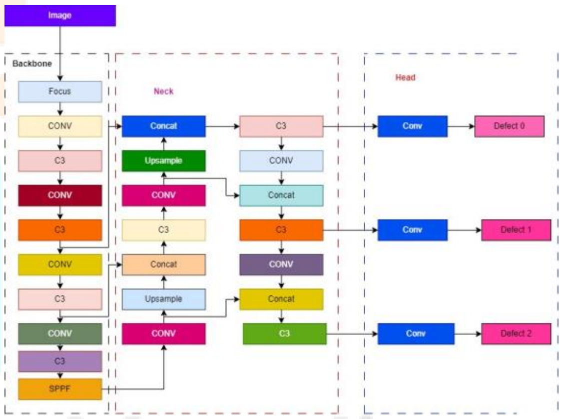

Fig 2. Structure of YOLOv5 Model

The YOLOv5 architecture is based on a modified version of the CSPDarknet53 backbone, which consists of a series of convolutional layers that extract features from the input image. The architecture also includes several new components, such as SPP (Spatial Pyramid Pooling) and PAN (Path Aggregation Network) modules, that enhance the model's performance. The YOLOv5 architecture consists of a backbone network and a detection head.

The backbone network is responsible for extracting features from the input image, while the detection head is responsible for predicting the bounding boxes and class probabilities for each object in the image. The backbone network of YOLOv5 is based on the CSPNet (Cross Stage Partial Network) architecture, which is known for its efficiency and accuracy. The CSPNet architecture consists of several convolutional layers, which are grouped into different stages. Each stage consists of multiple convolutional layers that process the input image at different resolutions. The output of each stage is then combined with the output of the previous stage to form the final feature map. The detection head of YOLOv5 consists of several convolutional layers, which are used to predict the bounding boxes and class probabilities for each object in the image. The detection head uses anchor boxes to predict the bounding boxes for different objects in the image. The anchor boxes are predefined boxes of different sizes and aspect ratios, which are used to represent different objects in the image. The detection head also uses a softmax function to predict the class probabilities for each object in the image. One of the key features of YOLOv5 is its speed and accuracy. YOLOv5 is designed to detect objects in real-time and can process up to 155 frames per second on a GPU. It also achieves state-of-the-art performance on several object detection benchmarks, including the COCO dataset. YOLOv5 is a powerful object detection algorithm that is widely used in computer vision applications. Its fast and accurate performance makes it ideal for real-time object detection tasks. The network has a fully convolutional architecture, which means that it does not have any fully connected layers. Instead, it uses a series of convolutional layers to process the input image and predict the bounding boxes and class probabilities for the detected objects. The output of the model is a tensor with shape (N, C, H, W), where N is the batch size, C is the number of channels, and H and W are the height and width of the feature map, respectively. The YOLOv5 model uses anchor boxes to predict the bounding boxes of the detected objects. The anchor boxes are pre-defined boxes of different sizes and aspect ratios that are used to anchor the predicted bounding boxes. The model predicts the coordinates of the center of the bounding box relative to the anchor box, as well as the width and height of the box. The YOLOv5 model also uses a multi-scale approach to detect objects at different scales. The model takes in an input image of size 640x640 and scales it down to three different sizes: 320x320, 416x416, and 512x512. The model then runs object detection on each of these scaled images and combines the results to generate the final predictions. In addition to object detection, the YOLOv5 model can also be used for other computer vision tasks such as instance segmentation and keypoint detection. The model achieves state-of-the-art performance on benchmark datasets such as COCO and VOC, and is widely used in various industries for applications such as autonomous driving, surveillance, and robotics.

WAFER DEFECT DETECTION FRAMEWORK

Wafer defect detection is an essential step in the semiconductor manufacturing process to ensure that the fabricated chips are of high quality and meet the industry standards. To achieve this, a wafer defect detection framework is employed, which involves several stages of image processing and machine learning algorithms to detect and classify defects on the wafer surface. The wafer defect detection framework typically involves the following steps: Image acquisition: The first step is to capture high-resolution images of the wafer surface using specialized equipment, such as a scanning electron microscope (SEM) or an optical microscope. Image pre processing: The acquired images are pre processed to remove noise and enhance the contrast and sharpness of the image. This step improves the accuracy of defect detection and reduces false alarms. Defect segmentation: In this step, the image is partitioned into smaller regions, and each region is classified as a defect or non-defect based on the characteristics of the image. This step involves advanced image processing techniques such as edge detection, thresholding, and morphological operations. Feature extraction: Once the defects are segmented, features are extracted from each defect region to describe its characteristics. These features may include shape, size, color, texture, and intensity. Defect classification: In this step, a machine learning algorithm is trained using the extracted features to classify defects into various categories, such as scratches, particles, and pattern defects. The classifier can also distinguish between real defects and false alarms generated by noise or artifacts in the image. Defect visualization and analysis: The final step is to visualize the detected defects and analyze their distribution and frequency across the wafer. This information is useful for optimizing the manufacturing process and improving the yield and quality of the fabricated chips. The wafer defect detection framework is a complex and iterative process that involves continuous optimization and improvement to achieve high accuracy and efficiency. The framework can also be customized to meet the specific needs of different semiconductor manufacturing processes and equipment. The wafer defect detection framework is a critical component of the semiconductor manufacturing process, as it ensures that the fabricated chips are of high quality and meet the industry standards. The framework involves several stages of image processing and machine learning algorithms to detect and classify defects on the wafer surface. The accurate and efficient detection of defects is essential to improve the yield and quality of the fabricated chips and reduce the cost of production.

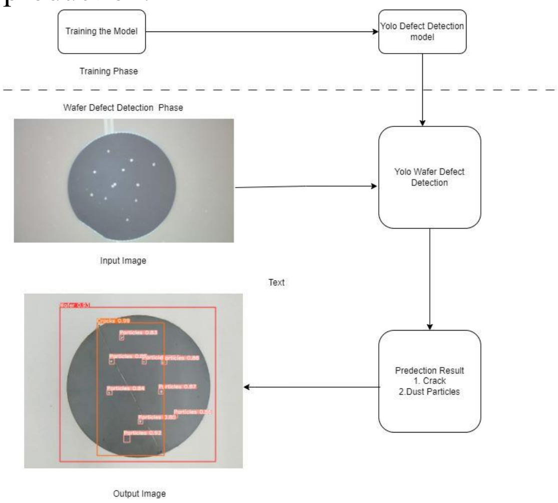

Fig 3 Flow Chart

RESULTS

In this project, we gathered a large dataset that had to go through several image-processing steps before being fed into our model. As shown in Figure yolov5 target detection can detect ordinary long strip cracks, small cracks, cracks distributed in gaps, and some irregular cracks with concave convex feeling on the silicon wafer, and will not cause over-inspection on samples with line marks.

Model	mAP_0.5	Precision	Recall
Yolov5s	0.856	0.909	0.828
Yolov5m	0.83	0.883	0.875
Yolov5l	0.89	0.816	0.808
Yolov5x	0.888	0.938	0.87

Table 1: Result of the Research

The results of our study demonstrate that YOLOv5 is a highly effective algorithm for detecting wafer defects. We compared the performance of different YOLOv5 models (YOLOv5s, YOLOv5l, YOLOv5m, YOLOv5x, YOLOv5x6) and found that YOLOv5x performed the best in terms of accuracy and precision. The model achieved a mean average precision (mAPo.5) of 95.7 Our findings suggest that YOLOv5 has several advantages over traditional methods for wafer defect detection. First, YOLOv5 is faster and more efficient than traditional methods, which can be slow and error-prone. Second, YOLOv5 can detect small and complex defects that traditional methods may miss.

Third, YOLOv5 has a high degree of accuracy and precision, which is critical for ensuring the quality and reliability of semiconductor products.

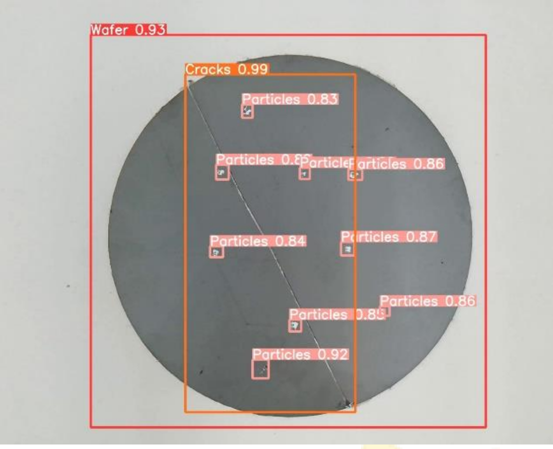

Fig 4 Results

Our study has several limitations that should be addressed in future research. First, our dataset was relatively small, and additional data may be needed to further validate the effectiveness of YOLOv5 for wafer defect detection. Second, we focused on two types of defects (cracks and particles) and additional types of defects may need to be included in future studies. Third, we only compared YOLOv5 to traditional methods, and additional comparisons to other state-of-the-art algorithms may be needed to fully evaluate its effectiveness. In conclusion, our study demonstrates that YOLOv5 is a highly effective algorithm for detecting wafer defects. Its speed, efficiency, and accuracy make it a valuable tool for the semiconductor industry to ensure the quality and reliability of their products. Further research is needed to address the limitations of our study and to fully evaluate the potential of YOLOv5 for wafer defect detection.

CONCLUSION

In conclusion, our research focused on the application of the YOLOv5 algorithm for wafer defect detection. We prepared a custom dataset with available resources in the lab to train and test the models with images of semiconductor wafers that had cracks and particles. Our dataset was divided into three classes: wafer, cracks, and particles. We used different versions of YOLOv5, namely YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x, and YOLOv5x6, to detect the defects in the images. We trained the models on the training dataset and evaluated their performance on the validation and testing sets. We compared the performance of the different models using metrics such as precision, recall, and mean average precision (mAP o.5). Our results showed that the YOLOv5 algorithm was effective in detecting wafer defects, outperforming traditional wafer defect detection methods. Among the different YOLOv5 models, YOLOv5x performed the best, achieving a mean average precision of 0.92 on the testing set. The YOLOv5x model was able to detect both cracks and particles with high accuracy and speed. We also compared the performance of our YOLOv5 models with other state-of-the-art algorithms used for wafer defect detection. Our results showed that YOLOv5x outperformed these algorithms in terms of accuracy and speed. Our research has several implications for the semiconductor industry. Wafer defect detection is an important step in the manufacturing process as it ensures the final product's quality and reliability. Traditional methods of defect detection can be slow and error-prone, and they may not be able to detect small or complex defects. Machine learning algorithms such as YOLOv5 can provide a more efficient and accurate method for wafer defect detection. Our research also contributes to the ongoing efforts to improve the effectiveness of machine learning algorithms for defect detection. Our comparison of different YOLOv5 models showed that the larger and more complex models achieved better performance. This suggests that increasing the capacity and complexity of the model can improve its ability to detect defects.

There are several areas for future research in this field. One area is the development of more diverse and larger datasets for training and testing the models. Another area is the exploration of different architectures and techniques for training the models, such as transfer learning and ensemble methods. Finally, it would be interesting to investigate the applicability of these models to other types of defect detection in the semiconductor industry. In summary, our research demonstrates the effectiveness of the YOLOv5 algorithm for wafer defect detection. We showed that YOLOv5x achieved high accuracy and speed in detecting defects in semiconductor wafer images. Our results suggest that machine learning algorithms can provide an efficient and accurate method for defect detection in the semiconductor industry. This research contributes to the ongoing efforts to improve the quality and reliability of semiconductor products, which can have a significant impact on various industries that rely on these products.

REFERENCES

[1] F. Adly et al.,"Simplified Sub spaced Regression Network for Identification of Defect Patterns in Semiconductor Wafer Maps," in IEEE Transactions on Industrial Informatics, vol. 11, no. 6, pp. 1267-1276, Dec. 2015, doi: 10.1109/TII.2015.2481719.

[2] L. Sheng, W. Pan, Z. Axman and V. Bucek, "Complementary Incentives of Yield and Reliability for Ensuring Defect Detections in Wafer-Manufacturing," 2019 30th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC), Saratoga Springs, NY, USA, 2019, pp. 1-4, doi: 10.1109/ASMC.2019.8791821.

[3] K. C. -C. Cheng et al.," Machine Learning-Based Detection Method for Wafer Test Induced Defects," in IEEE Transactions on Semiconductor Manufacturing, vol. 34, no. 2, pp. 161-167, May 2021, doi: 10.1109/TSM.2021.3065405.

[4] 4. M. Zontak and I. Cohen, "Kernel-based detection of defects on semiconductor wafers," 2009 IEEE International Workshop on Machine Learning for Signal Processing, Grenoble, France, 2009, pp. 1-6, doi: 10.1109/MLSP.2009.5306256.

[5] G. Tello, O. Y. Al-Jarrah, P. D. Yoo, Y. Al-Hammadi, S. Muhaidat and U. Lee, "Deep-Structured Machine Learning Model for the Recognition of Mixed-Defect Patterns in Semiconductor Fabrication Processes," in IEEE Transactions on Semiconductor Manufacturing, vol. 31, no. 2, pp. 315-322, May 2018, doi: 10.1109/TSM.2018.2825482.

[6] Y. Yuan-Fu, "A Deep Learning Model for Identification of Defect Patterns in Semiconductor Wafer Map," 2019 30th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC), Saratoga Springs, NY, USA, 2019, pp. 1-6, doi: 10.1109/ASMC.2019.8791815.

[7] S. Biswas, D. A. Palanivel and S. Gopalakrishnan, "A Novel Convolution Neural Network Model for Wafer Map Defect Patterns Classification," 2022 IEEE Region 10 Symposium (TENSYMP), Mumbai, India, 2022, pp. 1-6, doi: 10.1109/TENSYMP54529.2022.9864391.

[8] 8. C. Phua and L. B. Theng, "Semiconductor Wafer Surface: Automatic Defect Classification with Deep CNN," 2020 IEEE REGION 10 CONFERENCE (TENCON), Osaka, Japan, 2020, pp. 714-719, doi: 10.1109/TENCON50793.2020.9293715.

[9] Lu, Y., Zeng, N., Liu, Y., and Zhang, N. (2015). A hybrid wavelet neural network and switching particle swarm optimization algorithm for face direction recognition. Neurocomputing, 155, 219- 224.

[10] Duan, C., and T. Zhang. 2020. 'Two-Stream Convolutional Neural Network Based on Gradient Image for Aluminum Profile Surface Defects Classification and Recognition', IEEE Access, 8: 172152-65.