AI-Based Rapid Defect Detection Method for Display Screen Appearance

Shujuan Yin, Yuyu Liu, Xingqun Jiang
BOE Technology Group Co., Ltd., China

Abstract

The display manufacturing industry faces growing challenges as pixel sizes shrink and product complexity increases. Traditional defect detection algorithms, while widely adopted, suffer from three critical limitations: (1) Scenario-specific customization requiring labor-intensive parameter adjustments during product transitions; (2) Compromised efficiency due to manual balancing between over-detection (false positives) and under-detection (missed defects); (3) Inability to handle emerging defect types in advanced manufacturing processes.

This study proposes a universal AI-driven inspection framework integrating a dual-stage detection architecture: A high-speed screening module rapidly filters defect-free regions (＞99% elimination rate), followed by a deep learning-based classification module for precise defect identification. Implemented with an optimized YOLO architecture incorporating global attention mechanisms and dimensionality reduction, our solution achieves a 3% to 5% higher accuracy than baseline models while maintaining real-time processing capabilities. The system demonstrates strong adaptability across diverse defect types (scratches, ITO defects, particle contamination, etc.) and industrial scenarios, with deployment timelines reduced by 68% compared to conventional methods.

Author Keywords

Display manufacturing; Automated optical inspection; Deep learning; Attention mechanism; Multi-scale defect detection

1、Introduction

The traditional method involves using conventional image processing techniques to extract defects, and then classifying them based on their characteristics. This helps distinguish between true defects (scratches, marks, overflow of adhesive, etc.) and false defects (dust, fibers, dirt, grease, etc.). This approach requires high imaging quality, where the defects are clearly visible and can be easily extracted. Additionally, it relies on distinct features of the defects to determine their authenticity. If these necessary conditions are not met, it can lead to significant interference, resulting in a high number of over-killed and under-killed.

Based on the above analysis, traditional detection methods require customized lighting and can only detect a limited number of defect types that meet certain conditions, which significantly restrict the application scenarios and types of defects in appearance inspection.

Taking the appearance inspection of display screens as an example, Figure 1-1 shows the typical categories of defects in display screen appearance.

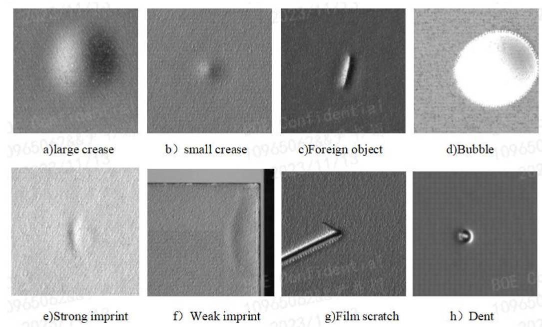

Figure 1-1 Typical defect categories of the appearance of display screen

In this study, to reduce costs, minimize debugging parameters, enhance the ease of software debugging, and improve system stability, a new detection technology in needed,which has more relaxed requirements for the imaging environment and is more inclusive of defect imaging characteristics. This technology should not only meet existing detection performance and functionality but also possess higher scalability and a broader range of applicability.

2、Proposed Algorithm

In today's rapidly advancing AI landscape, all industries are being redefined through AI, and quality inspection is no exception. A new quality inspection algorithm integrates AI technologies, offering greater flexibility in handling photo requirements and defect characteristics, while also achieving higher accuracy in defect identification.

The main detection algorithm is divided into two stages. The first stage is preliminary screening, which performs an initial selection of detection objects to filter out images with suspected defects. The primary goal of this step is to leverage the high efficiency of the screening algorithm to exclude images that are definitely free of defects, leaving only those that may contain defects. This significantly improves detection efficiency. The second stage is identification, where the remaining images that may contain defects undergo defect recognition to distinguish between true defects and false defects.

The core technical algorithm of the AI component utilizes deep learning techniques to train a model on a large number of categorized images, including true defect classes, false defect classes, and good product classes. This model can identify images with suspected defects and make accurate classifications, ultimately pinpointing true defects. The key advantage of using deep learning lies in its ability to bypass traditional complex feature extraction and recognition techniques, instead learning directly from big data to produce effective and precise results. While the efficiency of deep learning is slightly lower, the preliminary screening process ensures that the overall detection efficiency meets the practical requirements of industrial applications.

The main detection algorithm also involves pre-processing. The pre-processing stage handles the raw images captured by the equipment, with the goal of segmenting the detection areas

ISSN 0097-996X/25/5603-1591-$1.00 © 2025

based on the product's morphology. Different areas have distinct imaging characteristics, requiring tailored detection methods. For display screen products, the detection areas are mainly divided into three parts: the contour area, the terminal circuit area, and the display area (Figure 2-1).

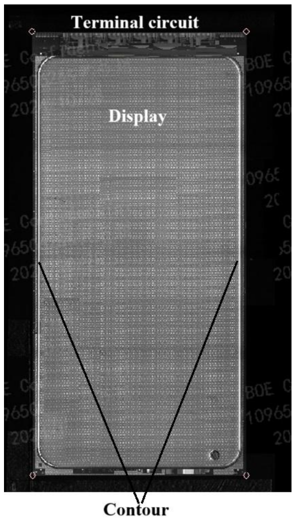

Figure 2-1 Display screen parts

The three parts have significant differences in imaging characteristics and produce different types of defects. For example, the contour area represents the peripheral region of the product and exhibits edge-related features. Its main defects include undercutting, edge collapse, burrs, etc. The terminal circuit area refers to the gold finger area at the end of the product, which has a complex background circuit. Its main defects include scratches, missing corners, burns, and edge-related defects. The display area represents the main visible region of the product, characterized by uniform imaging background and regular textures. Its main defects include dimples, creases, scratches, burns, foreign objects within the film, etc. Among these three parts, the display area accounts for the largest proportion in product imaging and has the strictest detection standards. It is the most challenging aspect in terms of both detection accuracy and efficiency. This article will focus on providing detailed explanations of the detection process for the display area.

The display area detection uses AI-based classification. The first stage is screening, where the system performs an initial selection of the objects to be inspected. This step uses a screening algorithm to quickly filter out images that might contain defects. Its main goal is to exclude images that are calculated to be defect-free. Most of the images here are good products, making up over 99% of the total, while only the portion that might have defects is kept. These calculations are done on the CPU, using multi-core parallel processing to boost detection efficiency.

The second stage is recognition. At this stage, the system performs a more detailed check on the portion retained after the initial screening. Using deep learning, it identifies defects and employs AI-based recognition to classify them precisely. During this process, any irrelevant items are filtered out, while actual defects are accurately detected.

The overall detection process is illustrated in Figure 2-2.

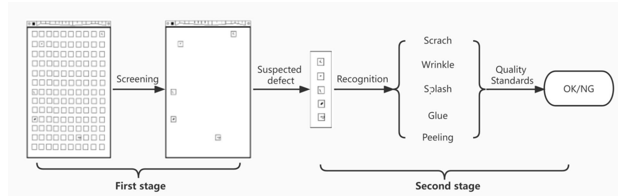

Figure 2-2 Defect detection overall flowchart

3、Technology Realization

In the previous section, we summarized the overall detection process for the display screen, with a particular focus on the display area. In this section, we will delve into the detailed implementation steps for detecting the display area.

3.1 Small Window Segmentation Design

The image exhibits visible non-uniformity due to inconsistencies in the light source and deformations caused by the platform's suction, as illustrated in Figure 3-1. While flat-field correction can partially address the light source's unevenness, further improvements in processing algorithms are needed to enhance image quality in the short term. This non-uniformity negatively affects the detection of central defects, particularly issues like dents, folds, scratches, stains, and ITO defects.

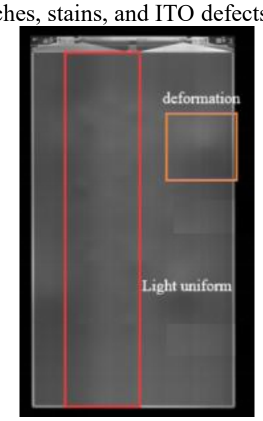

Figure 3-1: Non-uniformity image

In such cases, using global detection would result in a significant number of false positives. However, local regions typically exhibit relatively good uniformity. If we treat these local regions as the smallest detection units, it can effectively address optical issues. To solve this problem, we divide the image into multiple smaller local regions for detection.

The specific steps are as follows, as illustrated in Figure 3-2:

1、Divide the detection area into regions according to the specified number of threads, and define the starting points and parameters for each thread's detection area.

2、Starting from the defined starting point, calculate the number of vertical and horizontal divisions possible based on the small window size. Then, sequentially split the region into small windows and apply the detection algorithm to each area.

3、Shift the starting position of the small window by 1/3 of its width and repeat step 2 to create additional divisions.

4、Move the starting position of the small window down by 1/3 of its height and repeat step 2 to generate further divisions.

5、Split the remaining area into small windows based on the small window size and apply the liquid crystal area detection algorithm to each specific area.

6、Gather and combine the results, then conclude the detection process.

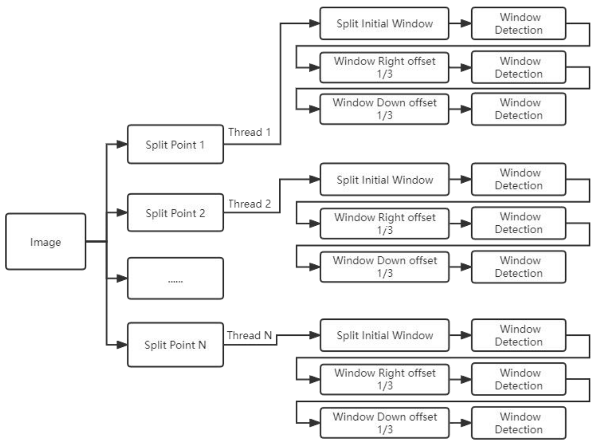

Figure 3-2: Small window detection process

During actual detection, images often exhibit periodic stripe or grid interference caused by the structure of the liquid crystal grid. To address this, when performing small window detection, if the parameters include texture removal, the small image is converted to the frequency domain and subjected to specific bandpass filtering to reduce or eliminate the interference. As a result, the design of the small window size must take this into account, and a size of 256*256 pixels is typically used. This size has been determined through testing and practical experience, balancing efficiency and uniformity.

3.2 AI Recognition Design

AI recognition serves as the second step in defect detection, where images potentially containing defects are analyzed to distinguish between true and false defects. The core approach involves using deep learning techniques to train a model on a large dataset of categorized images, including true defects, false defects, and defect-free products. This model is then applied to identify and classify suspected defects in images, ultimately isolating the true defects. The key advantage of deep learning is that it replaces traditional, complex feature extraction and recognition methods with data-driven learning, delivering highly accurate and effective results. While the process is slightly less efficient, the overall efficiency meets practical application requirements thanks to the support of pre-screening steps.

The core of the entire solution lies in the generation of defect models and product parameter models. In this case, both the reasoning algorithm for defect information and the algorithm for product parameter information utilize the same algorithm model, which is optimized based on the open-source YOLO [1,2]segmentation model. The core structure of the YOLO open-source model is illustrated in Figure 3-3.

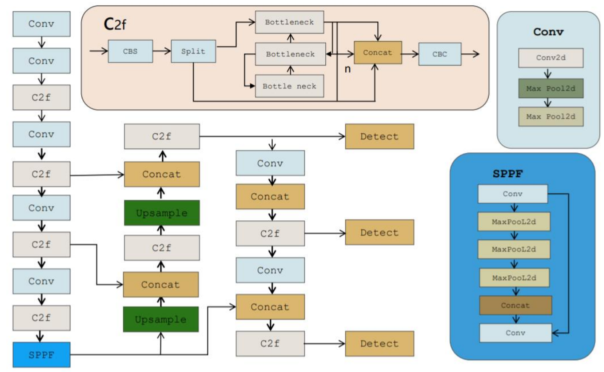

Figure 3-3 YOLO open-source model

The disadvantages of the open-source YOLO model:

1、It processes high-resolution images relatively slowly.

2、Its accuracy is low in highly complex and crowded scenes.

3、It struggles to detect rare or irregular physical forms, as well as those with weak imaging, due to limited sample availability.

For industrial scenarios, several key characteristics stand out:

1、High Efficiency Demands: Industrial environments require extremely high processing efficiency to match production capacity.

2、Complex Microscopic Backgrounds: Most industrial scenes involve microscopic views with highly complex and variable backgrounds.

3、Random and Subtle Defects: Defects in industrial settings often appear randomly and may have very weak imaging, sometimes nearing the limits of human vision.

Limited Sample Availability: Industrial samples are typically scarce, posing challenges for training and validation.

Considering these requirements, the following optimizations will be applied to the YOLO model:

Building on the YOLO model, a global attention mechanism[3,4,5,6] has been introduced to enhance performance by maintaining the relationship between spatial and channel information. This improvement enables the model to prioritize critical information from large datasets while filtering out less relevant data. Additionally, input dimensionality reduction is implemented to boost algorithm efficiency.

In industrial applications, detecting extremely small defects is crucial, and product parameter information is often intricate and complex, with limited sample availability[7]. Furthermore, production capacity demands impose exceptionally high efficiency requirements. To address these challenges, this solution integrates a global attention mechanism to increase detection efficiency, optimize scene relationships, and enhance accuracy.

The schematic diagram of the global attention mechanism module is shown in Figure 3-4.

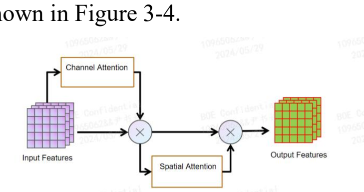

Figure 3-4 Global attention mechanism module

The global attention mechanism module includes a total of three processes, which are as follows:

1、Added after the backbone SPPF.

2、Combined with the three C2f layers in the neck and Detect.
3、Added after the C2f in the neck.
The integration of these processes is illustrated as Figure 3-5.

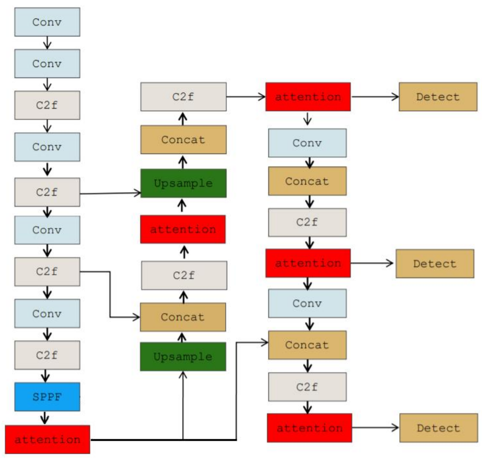

Figure 3-5 Integration of processes

4、Experimental Result

We trained both the original YOLO model and the improved model using a dataset of 300,000 images from the industrial field. The training results are presented in Figure 4-1.

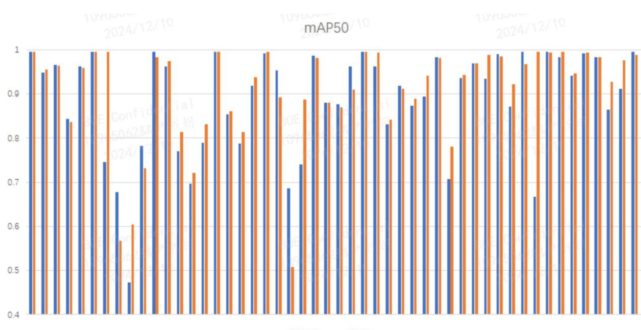

Figure 4-1 Training comparison results

As illustrated in the chart, our solution exhibits notably higher accuracy than the original model across the majority of categories. Our solution achieves a 3% to 5% improvement in accuracy over baseline models, while reducing deployment timelines by 68% compared to traditional methods.

5、Conclusion

In conclusion, the key to display defect detection lies in three aspects: first, the design of the batch detection architecture; second, the design of high-speed detection algorithms; and third, the design of high-precision algorithms. By incorporating appropriate attention mechanisms into the model, detection accuracy can be significantly improved. Additionally, dimensional reduction techniques can effectively enhance efficiency. This approach is not only applicable to the display field but also highly versatile, capable of being transferred to other industrial application scenarios.

6、References

1. Ultralytics: YOLOv8 Documentation. https://docs.ultralytics.com/ (2023). Accessed: 2023-10-01
2. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection. In: CVPR (2016)
3. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: CBAM: Convolutional Block Attention Module. In: ECCV (2018)
4. Hu, J., Shen, L., Sun, G.: Squeeze-and-Excitation Networks. In: CVPR (2018)
5. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is All You Need. In: NeurIPS (2017)
6. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In: CVPR (2020)
7. Liu, Y., Zhang, L., Wang, F., Zhang, D.: Industrial Defect Detection with Deep Learning: A Survey. IEEE Transactions on Industrial Informatics (2021)