International Journal of Innovative Computing, Information and Control
Volume **21**, Number **2**, April **2025**

 ISSN 1349-4198
pp. **457–467**

A SELF-SUPERVISED METHOD FOR TRILATERAL SENSING: LOCALIZATION, ANGULAR MEASUREMENT, AND DEFECT DETECTION

Xiang Li, Dong Zhou* and Xiangyu Shao

Department of Control Science and Engineering
Harbin Institute of Technology
No. 92, West Dazhi Street, Nangang District, Harbin 150001, P. R. China
{ lixiang_soa; xiangyushao }@hit.edu.cn
*Corresponding author: dongzhou@hit.edu.cn

Received August 2024; revised November 2024

Abstract. *In manufacturing, defect detection during production is crucial. However, the scarcity of defect samples increases costs for visual systems. To address this, we propose a self-supervised pipeline for transistor chip defect detection. It requires only qualified samples for training and achieves localization, angle measurement, and defect detection simultaneously. The method includes Integral Feature-Based Localization (IFBL) and Rotation-Driven Anomaly Detection (RDAD) modules. IFBL effectively locates chips using rotational invariant integral features, while RDAD combines angle measurement with defect detection. While training the network to understand image angles, it also acquired defect detection capabilities. Experimental results show our method outperforms baselines, achieving positioning errors of 0.7511 mm and 2.476°, with a 96.1% AUC in defect detection. These findings demonstrate the effectiveness of our approach, offering a cost-effective solution for the manufacturing industry.*
Keywords: Visual inspection, Rotated object detection, Angular measurement, Anomaly detection

1. **Introduction.** In the manufacturing industry, detecting defects in electronic components, such as transistor chips, is crucial to ensuring system stability and preventing severe consequences. Automated Optical Inspection (AOI) inspects the pin and packaging integrity of components. Typically, after electrical testing, chips are randomly placed on a tray, photographed at close range, and any defective products are promptly removed by a mechanical device.

The visual inspection process involves three tasks: chip localization followed by rotation angle measurement and defect detection. Initially, as chip orientations are unknown, existing data-driven object detection methods require significant amounts of annotated data at various angles, imposing cost pressures. In the subsequent defect detection phase, even greater challenges arise in data acquisition. Moreover, even with the utilization of unsupervised learning models, the small scale and subtle features of defects often make such approaches ineffective.

To address these challenges, we propose a self-supervised pipeline for product inspection using unlabeled image data. This pipeline integrates localization, angular measurement, and defect detection tasks through two modules: Integral Feature-Based Localization (IFBL) and Rotation-Driven Anomaly Detection (RDAD). The IFBL module extracts chip

DOI: 10.24507/ijicic.21.02.457

center coordinates using rotational-invariant integral features. The RDAD module employs a network to regress chip orientations, thereby effectively performing defect detection tasks by leveraging the network's incidental structural understanding.

To validate the effectiveness of the proposed method, we collected and utilized two datasets: the Transistor Localization Dataset and the Transistor Defect Dataset. For defect detection, we introduced the MVTec AD dataset [1] to validate the practicality of our algorithm and compared it with state-of-the-art methods. The experimental results indicate impressive performance achieved by our proposed method, confirming the reliability and efficacy of our approach, supporting its application in manufacturing.

2. **Related Works.** Regarding the three tasks discussed in this paper – localization, angular measurement, and defect detection – most existing methods commonly treat localization and angular measurement as a combined problem within the scope of rotated object detection, while defect detection is handled as an independent task. This section reviews the current research in these two areas. In contrast, our approach innovatively integrates the latter two tasks – angular measurement and defect detection – into a unified framework. Although the approaches differ, the underlying techniques and challenges are similar, making the review of these areas valuable for our method.

2.1. **Rotated object detection.** Rotated object detection is a significant issue in computer vision, and there have been many effective methods proposed recently. For example, Lu et al. [2] developed OSKDet, an orientation-sensitive heatmap based rotated detector. It utilizes keypoints to predict a heatmap for rotated target representation, enhancing modeling capabilities for improved localization and detection quality. Yu et al. [3] proposed a rotating object detector for remote sensing images, integrating convolutional dynamic adaptive matching. Their approach includes an angle estimator in the backbone network to align the convolutional kernel with the target orientation, ensuring precise feature extraction aligned with spatial targets.

Current research primarily focuses on complex scenarios such as remote sensing, relying on supervised Convolutional Neural Networks (CNNs) that require extensive labeled data. In contrast, our study targets the manufacturing environment, which has simpler backgrounds but frequent product appearance changes, making high labeling costs prohibitive and these methods impractical. Additionally, most existing approaches perform angle recognition and object detection simultaneously, whereas we defer angle recognition to the subsequent defect detection stage.

2.2. **Defect detection.** Anomaly detection methods can be fully supervised, unsupervised, or weakly supervised. Our research falls under the unsupervised category, requiring no labeled anomalous samples during training. Akcay et al. [4] introduced GANomaly, based on Generative Adversarial Networks (GANs), utilizing encoder-decoder-encoder sub-networks within the generator. This enables the model to map the input image to a lower-dimensional vector, which is then used to reconstruct the output image. Schlegl et al. introduced f-AnoGAN [5] for clinical imaging anomaly detection, noted for its speed and accuracy. CutPaste [6] by Li et al. employed self-supervised learning and innovative data augmentation for effective defect detection and localization. Xie et al. [7] proposed the DC-EMV2YOLOV4 model to address the sample quality imbalance problem, achieving promising results in metal surface defect detection applications. Kim et al. [8] introduced a classification model for quality control of shape in rubber manufacturing defect detection. This method integrates image data and shape size, offering a novel approach for early applications in fields with limited labeled data.

While these methods have shown impressive results in detecting texture defects within their respective application domains, existing models, particularly those based on GANs, exhibit significant performance degradation when normal samples exhibit a wide distribution. The starting point of this study lies precisely in addressing this scenario.

3. **Proposed Methods.** Our proposed method aims to address the challenge of simultaneously performing two regression tasks and one classification task: localization, angular measurement, and defect detection, using only images of qualified products without labeled data.

The proposed pipeline is shown in Figure 1. In the localization stage, we propose an Integral Feature-Based Localization (IFBL) method as a rotation-invariant criterion. By using a single image as a template, this method rapidly identifies positions in the image that match the template's texture. It is computationally efficient and robust, making it well-suited for product localization in manufacturing. When product appearances change, the localization algorithm can be updated simply by replacing the template image, eliminating the need for additional human-labeled training data.

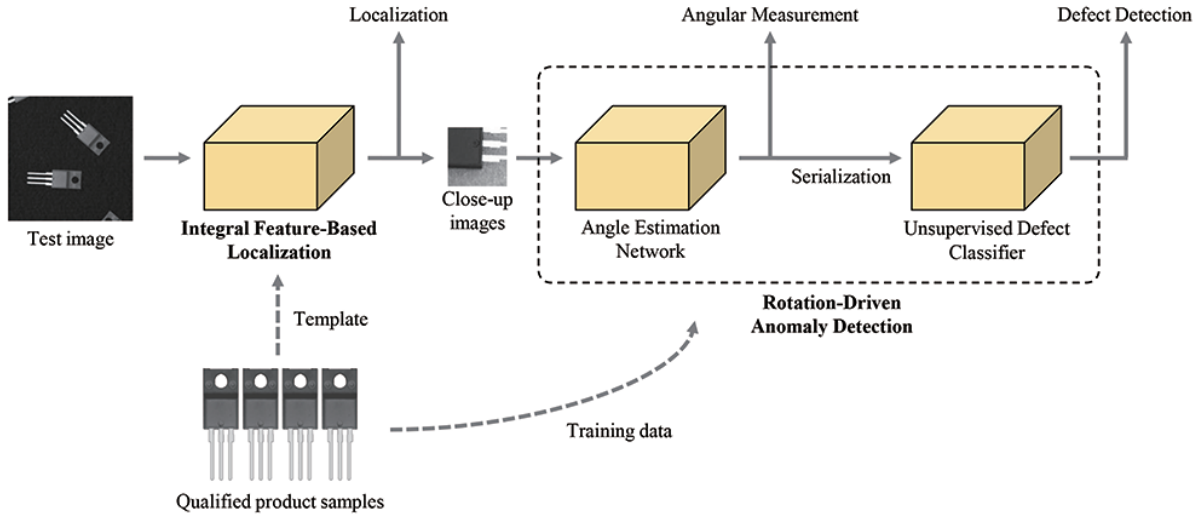
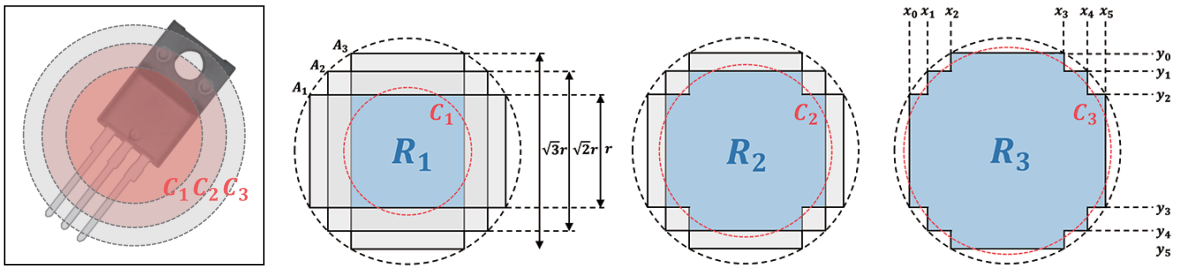


FIGURE 1. The proposed pipeline

Subsequently, the proposed Rotation-Driven Anomaly Detection (RDAD) framework simultaneously performs angular measurement and defect detection. The angle estimation network is a critical component of the process. It demonstrates high angular measurement accuracy for qualified samples while producing significant errors for defective samples. By evaluating the accuracy of angle predictions, defect detection can be indirectly achieved. This approach is based on the insight that if a network can accurately predict the rotation angle of an object, it must have an understanding of the object's structure, thereby inherently possessing the potential for defect detection.

3.1. **Integral Feature-Based Localization.** The Integral Feature-Based Localization (IFBL) module takes images containing multiple transistor chip samples with uncertain orientations as input, deriving the central position of each transistor as output. Due to the inability to train with extensive annotated data, we cannot rely on data-driven methods. Instead, we employ a sliding window strategy. At each position, we use specific criteria to compare with the template image and utilize the matching score for position suggestion. The criteria must meet the following requirements:

1) Rotational invariance, ensuring consistent values for targets at any angle;



(a) (b)

FIGURE 2. (a) Concentric circular regions on the template image; (b) construction of the Integral Feature

2) Computational efficiency, ensuring minimal computation in each iteration;
3) Strong discrimination and robustness against interference.

To meet these requirements, we propose Integral Features, as shown in Figure 2(a). This method defines three concentric circular regions on the template image and calculates the normalized pixel mean within these regions. With RGB channels, this results in a 9-dimensional feature vector. However, directly calculating the pixel mean within circular regions is computationally intensive. To improve efficiency, we approximate this calculation using several rectangular regions during image traversal.

As shown in Figure 2(b), for a circular region with a radius r, three internal rectangles A_i can be inscribed with heights of r, $\sqrt{2}r$, $\sqrt{3}r$, respectively. From these rectangles, three composite rectangular regions R_i can be derived, each of which approximates the shape of the circle C_i. Table 1 lists detailed information on these regions, including their shape definition and area, the radius of the corresponding C_i, and the Intersection over Union (IoU) between them.

TABLE 1. Approximation details

	Shape definition	Area	Radius of C_i	IoU
R_1	$A_1 \cap A_2 \cap A_3$	r^2	0.56419r	83.36%
R_2	$(A_1 \cap A_2) \cup (A_1 \cap A_3)$	$\left(2\sqrt{2}-1\right) r^2$	0.76289r	89.83%
R_3	$A_1 \cup A_2 \cup A_3$	$\left(2\sqrt{3}-2\sqrt{2}+2\right) r^2$	0.91594r	92.69%

Therefore, we can approximate the circular regions C_i with the rectangular regions R_i. Since they are composed of rectangles, the pixel intensity sum S_{Ri} in region R_i can be quickly computed using integral images, specifically:

$$\begin{cases} S_{R1} = I(x_3, y_3) + I(x_2, y_2) - I(x_2, y_3) - I(x_3, y_2) \\ S_{R2} = I(x_4, y_3) + I(x_1, y_2) + I(x_3, y_4) + I(x_2, y_1) - I(x_4, y_2) \\ \qquad - I(x_1, y_3) - I(x_3, y_1) - I(x_2, y_4) - S_{R1} \\ S_{R3} = I(x_5, y_3) + I(x_0, y_2) + I(x_4, y_4) + I(x_3, y_5) + I(x_2, y_0) \\ \qquad + I(x_1, y_1) - I(x_4, y_1) - I(x_1, y_4) - I(x_5, y_2) - I(x_0, y_3) \\ \qquad - I(x_3, y_0) - I(x_2, y_5) - S_{R1} - S_{R2} \end{cases} \tag{1}$$

Here, $I(x_i, y_i)$ represents the pixel value of the integral image at position (x_i, y_i), as indicated in Figure 2(b), which can be determined by the center coordinates (x_c, y_c) and the radius r of the circle:

$$x_i = \begin{cases} x_c - \dfrac{\sqrt{3-i}}{2}r, & i \le 2 \\ x_c + \dfrac{\sqrt{3-i}}{2}r, & i > 2 \end{cases} \qquad y_i = \begin{cases} y_c - \dfrac{\sqrt{3-i}}{2}r, & i \le 2 \\ y_c + \dfrac{\sqrt{3-i}}{2}r, & i > 2 \end{cases} \tag{2}$$

It can be observed that regardless value of the radius r, utilizing the above approximation method only requires accessing the values of 24 points in the integral image and performing 24 addition operations to obtain approximate pixel sums for the three circular regions. After introducing the area of R_i, we can calculate its normalized pixel mean.

For three-channel image data, each computation yields a 9-dimensional vector as the Integral Feature. At each position in the sliding window, the Euclidean distance between the Integral Feature of the image at that position and that of the template image is calculated. A smaller distance indicates a better matching, resulting in a Gaussian blob suggesting the position. The center of each Gaussian blob corresponds to the detected target's center. Since the Integral Feature of the template image needs to be computed only once, we can use the exact C_i concentric circular regions for feature calculation rather than rectangular approximations, ensuring the accuracy of the template image's feature values.

3.2. **Rotation-Driven Anomaly Detection.** The Rotation-Driven Anomaly Detection (RDAD) module takes close-up images of transistor chips as input, with some positional variation due to localization errors, and outputs the angle of the chip and defect detection results. The angle is derived by the Angle Estimation Network, and defect detection is performed by an unsupervised defect classifier.

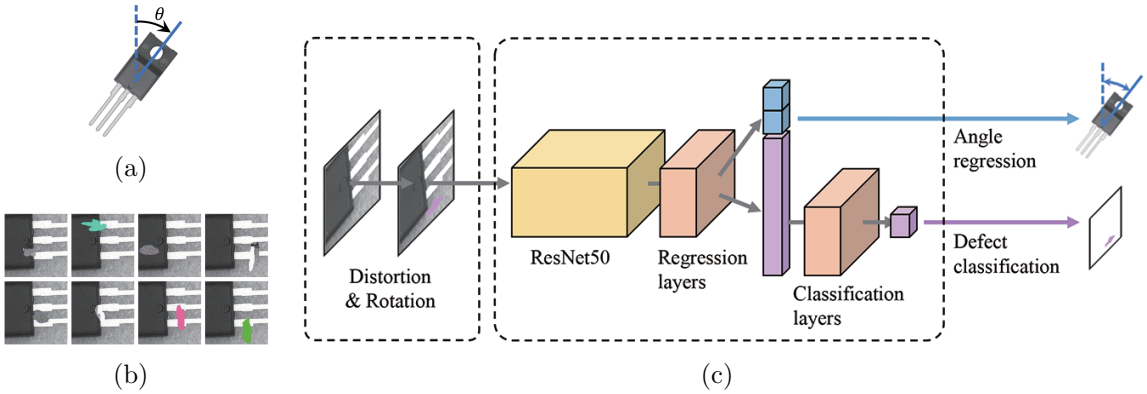


FIGURE 3. (a) Angle definition; (b) sample cases of random distortion; (c) architecture of the Angle Estimation Network

3.2.1. *Angle Estimation Network.* The rotational angle of transistor chips is defined as illustrated in Figure 3(a), where it is represented as a scalar $\theta \in [0, 360)$. In our dataset, all samples are oriented similarly to avoid additional annotation costs, and we assume they share the same angle value. To mitigate numerical ambiguity caused by angle wrapping, we use a 2-dimensional vector $(-\sin\theta, \cos\theta)$ to represent θ. During training, we manually rotate the images by a random angle. To avoid edge distortions caused by rotating rectangular images, we only keep the inscribed circular region of the images, covering the remaining parts with black. This rotated angle is then used as training label. We use the Mean Squared Error (MSE) loss function to calculate the regression error:

$$\text{Loss}_R = \frac{1}{N}\sum_{i=1}^{N} \|\hat{y}_i - y_i\|_2^2 \tag{3}$$

where N is the number of samples, $\hat{y}_i$ is the predicted vector output by the model, and $y_i = (-\sin\theta_i, \cos\theta_i)$ is the ground truth vector determined by the manually rotated angle θ_i. The notation $\|\cdot\|_2^2$ represents the squared L2 norm.

Regarding network architecture, we utilized ResNet50 [9] as the backbone for feature extraction, followed by the addition of a regression network consisting of four linear layers. The output of the regression network is a 16-dimensional vector, where the first 2 dimensions represent the predicted angle $\hat{y}_i$. As a matter of fact, this network architecture is capable of effectively measuring angles, as evidenced by achieving an average error as low as 1.301° on our dataset. However, this network structure alone is insufficient for subsequent defect detection tasks, necessitating the introduction of additional training strategies.

Firstly, to ensure the network's responsiveness to potential defects, we introduced random distortion to the images. Specifically, as depicted in Figure 3(b), we applied disturbances to random locations in the images, replacing textures with random colors, Gaussian noise, or copies of other regions, all resulting in semantic discontinuity in localized areas of the images to aid the network in acquiring defect detection capabilities. Additionally, we fed the last 14 dimensions of the regression network's output into a classification network to obtain an additional classification prediction corresponding to whether the input image was distorted. During training, both distorted images and their original counterparts are paired and fed into the angle prediction network, trained using the cross-entropy loss function denoted as Loss_C.

Ultimately, the structure of the Angle Estimation Network is depicted in Figure 3(c), where the loss function for training is defined as

$$\text{Loss} = \begin{cases} \text{Loss}_C + \text{Loss}_R, & \text{for original image input} \\ \text{Loss}_C, & \text{for distorted image input} \end{cases} \tag{4}$$

It is noteworthy that for distorted image input, the angle regression error is not computed. This deliberate omission aims to guide the network to ignore angular measurements of anomalous images, facilitating the design of subsequent unsupervised classifiers. Additionally, although it helps improve regression accuracy, the two dimensions of the angular output are not L2 normalized for the same reason.

We apply the improved network to all possible angles of all samples in our Transistor Defect Dataset. The Kernel Density Estimation (KDE) of the angle prediction errors is shown in Figure 4(a). As can be seen, for normal samples, the network achieves an average error of 2.476°, with the majority of errors remaining below 10°. Although the angle measurement capability slightly decreased compared to the network without Loss_C, it still meets practical requirements. For abnormal samples, the network shows an average error of 7.481°, which is significantly higher than for normal samples, reflecting the effect of the network improvement. However, this is not sufficient for direct defect detection, and we still rely on the unsupervised defect classifier.

3.2.2. *Unsupervised defect classifier.* Constructed and trained as described, the angle estimation network exhibits better performance in predicting angles for normal images. Thanks to its unique training approach, the network tends to perform relatively worse in angle predictions for defective images. By leveraging the unsupervised Support Vector Data Description (SVDD) [10] classifier, we can detect defects in original images based on angle prediction values.

To achieve this, we manually rotate the test images by increments of 1° throughout a full rotation and feed all 360 resulting images into the angle estimation network to obtain a sequence of angle predictions. We adopt this serialization approach for several reasons:

1) During the prediction phase, the true angles of the input images are unknown, making it impossible to calculate measurement errors. Rotating the images manually covers all possible angle scenarios, thus circumventing this issue;
2) The fluctuation range of measurement errors for individual images is too large to support defect detection tasks.

Figure 4(c) shows the serialized angle prediction values for several normal and abnormal samples. It can be observed that there are discrepancies in these values between normal and abnormal samples. Given the unknown initial angles of input images, it is impractical to assume the starting position of the sequence. Therefore, during the training of the unsupervised classifier, we randomly reordered the angle prediction sequence. Figure 4(b) illustrates the training and predicting process of the unsupervised SVDD classifier.

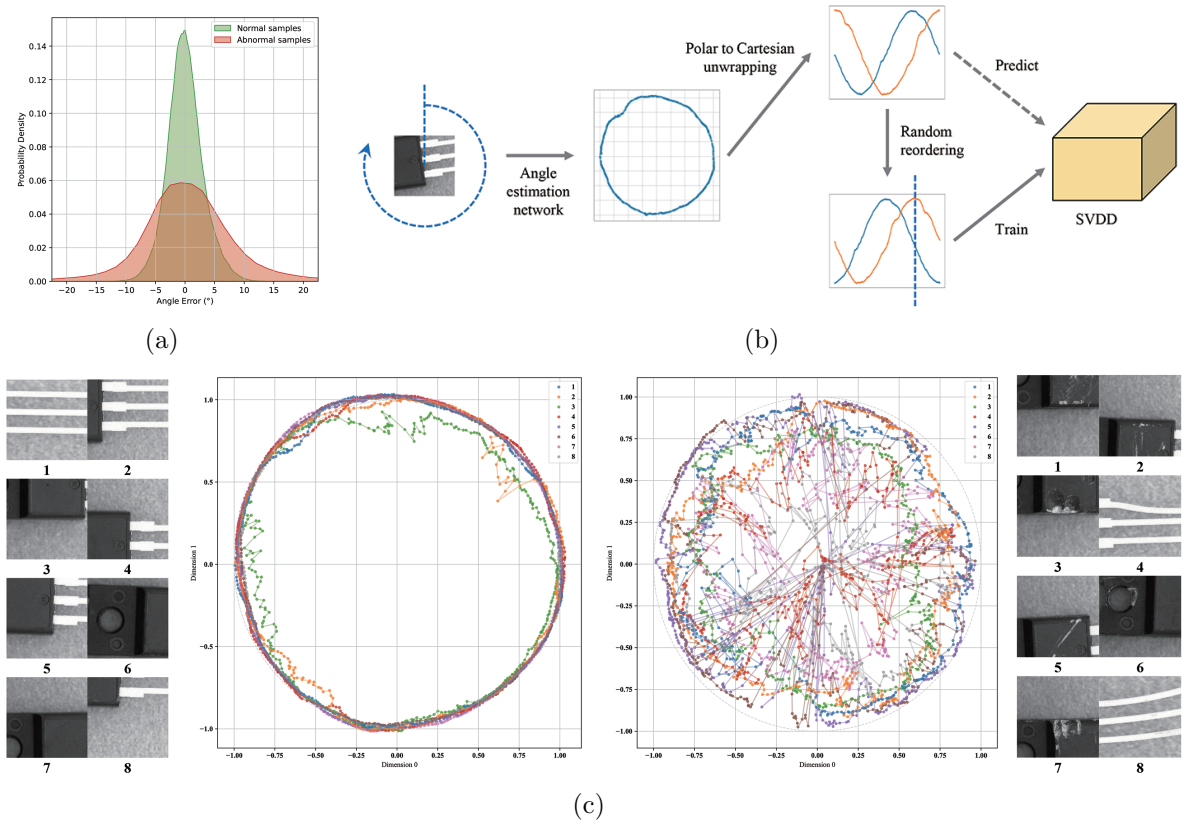


FIGURE 4. (color online) (a) Distribution of angle prediction errors for normal and abnormal samples; (b) training process of the unsupervised defect classifier; (c) sample cases of serialized data

It is worth noting that while the angle estimation network itself does possess classification outputs, these outputs are not applicable. They only demonstrate detection capabilities on artificially distorted images and serve merely as auxiliary during its training process.

4. **Experiment.**

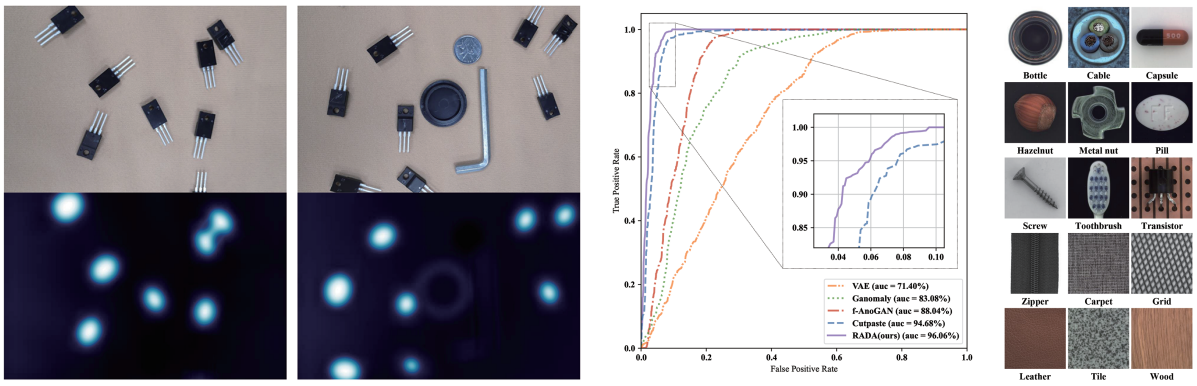
4.1. **Evaluation of localization performance.** To evaluate the effectiveness of the proposed Integral Feature-Based Localization method, we constructed a Transistor Localization Dataset containing 260 images. Each image in the dataset contains multiple transistor chip targets with annotated center positions. These annotations were only used

to assess the experimental results and did not participate in the algorithm execution process. The dataset was crafted to include images with uneven textured backgrounds, non-uniform lighting, irrelevant objects as interference, and target with defects to comprehensively assess the stability and recognition capabilities of the algorithm.

The experimental process involved applying the Integral Feature-Based Localization method to our constructed dataset. The algorithm's center coordinates were compared with the annotated ground truth values. Our experiments were conducted on a high-performance computing platform equipped with an Intel Xeon Gold 6132 processor and an NVIDIA Tesla P100 GPU, providing the necessary computational power to achieve reliable and precise results.

Figure 5(a) shows some experimental results. By applying our method, each expected target position forms a Gaussian Blob based on the matching degree. The centroid of each Gaussian blob is used as the predicted position. It is evident that the proposed method is stable in locating transistor targets and effectively avoids interference. Quantitatively, across all transistor targets in the dataset, our method achieved an average localization error of 0.7511 mm, a miss rate of 0.429%, and a false positive rate of 0%. These experimental findings demonstrate the robustness and accuracy of our proposed Integral Feature-Based Localization method for transistor chip localization.

4.2. **Assessment of defect detection performance.** To comprehensively evaluate the performance of the proposed Rotation-Driven Anomaly Detection (RDAD) method, we collected a Transistor Defect Dataset comprising 556 positive and 533 abnormal images,



(a) (b) (c)

FIGURE 5. (a) Localization results of the proposed IFBL; (b) ROC curves for the proposed RADA and other models; (c) examples of MVTec AD dataset

TABLE 2. Sample cases

	1	2	3	4	5	6	7	8	9	10
VAE	×							×		
GANomaly [4]	×	×	×	×						
f-AnoGAN [5]	×	×	×		×					
CutPaste [6]	×	×	×	×		×	×		×	
Our method	×	×	×	×	×	×	×	×		×

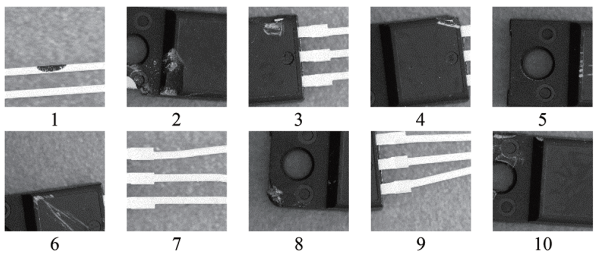


Table 3. Performance of various models

		VAE	GANomaly [4]	f-AnoGAN [5]	CutPaste [6]	Our method
Our transistor dataset		0.714	0.831	0.880	0.947	**0.961**
MVTec AD (oriented)	Cable	0.582	0.790	0.809	0.788	**0.811**
	Capsule	0.626	0.651	0.724	0.883	**0.913**
	Pill	0.553	0.591	0.760	0.905	**0.950**
	Screw	0.588	0.729	0.781	0.873	**0.880**
	Toothbrush	0.578	0.548	0.590	0.857	**0.895**
	Transistor	0.817	0.646	0.687	0.914	**0.922**
	Zipper	0.716	0.718	0.546	**0.996**	0.947
MVTec AD (non-oriented)	Bottle	0.735	0.711	0.800	**0.990**	0.513
	Hazelnut	0.422	0.829	0.865	**0.955**	0.769
	Metal nut	0.740	**0.765**	0.624	0.762	0.706
	Carpet	0.621	**0.749**	0.670	0.732	0.664
	Grid	0.679	0.695	0.747	**0.998**	0.781
	Leather	0.421	0.869	0.952	**0.999**	0.675
	Tile	0.574	0.928	0.748	**0.984**	0.558
	Wood	0.611	0.943	0.959	**0.991**	0.578

covering various common defect types encountered in transistor chip production. Combining this dataset with the MVTec AD dataset, we randomly cropped images to augment the diversity of normal data. The dataset samples are listed in Figure 5(c).

We compared the RDAD method with VAE and state-of-the-art techniques in anomaly detection, including GANomaly [4], f-AnoGAN [5], and CutPaste [6]. The Receiver Operating Characteristic (ROC) curves for defect detection of these models on our transistor dataset are shown in Figure 5(b), and Table 2 presents the detection results for some defect samples, an "×" mark indicates that the model alarmed for the corresponding defect sample. Additionally, Table 3 presents the Area Under the Curve (AUC%) results for all datasets. Overall, the CutPaste exhibited the best performance among the baseline methods. The proposed RDAD demonstrated impressive performance as well, especially in our transistor dataset and the oriented categories of the MVTec AD dataset.

We observed that certain types in the MVTec AD dataset, such as those marked as "non-oriented" in Table 3, lack inherent orientation features in their shapes. In such cases, the angle is solely determined by the cropping position rather than the object's orientation itself. This circumstance may introduce misleading information during the training of angle estimation networks, thereby limiting the applicability of our method. This aspect should be recognized as a limitation of our approach.

5. **Conclusions.** This study proposes a self-supervised pipeline for defect detection in transistor chips, capable of localization, angle measurement, and defect detection tasks without the need for annotated datasets. The experimental results demonstrate that our approach, leveraging integral features, achieves robust accuracy and interference resistance when detecting multiple similar targets in varying angles. Additionally, we introduce a Rotation-Driven Anomaly Detection framework, that detects defects and measures angles simultaneously. Experimental results indicate the superiority of our method in various scenarios, providing a cost-effective and efficient solution for AOI tasks. However, we acknowledge limitations in our approach, particularly its inapplicability to targets lacking

inherent orientation features and the need for improved real-time defect detection. These limitations are expected to be addressed in future research endeavors.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (No. U23A20346, No. 62173107). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] P. Bergmann, M. Fauser, D. Sattlegger and C. Steger, MVTec AD – A comprehensive real-world dataset for unsupervised anomaly detection, *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.9592-9600, 2019.

[2] D. Lu, D. Li, Y. Li and S. Wang, OSKDet: Orientation-sensitive keypoint localization for rotated object detection, *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[3] L. Yu, Y. Zhou, X. Li, S. Hu and D. Jing, A rotating object detector with convolutional dynamic adaptive matching, *Applied Sciences*, vol.14, no.2, 633, 2024.

[4] S. Akcay, A. Atapour-Abarghouei and T. P. Breckon, GANomaly: Semisupervised anomaly detection via adversarial training, in *Computer Vision – ACCV 2018. Lecture Notes in Computer Science*, C. Jawahar, H. Li, G. Mori and K. Schindler (eds.), Cham, Springer, 2019.

[5] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs and U. Schmidt-Erfurth, F-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks, *Med. Image Anal.*, vol.54, pp.30-44, 2019.

[6] C.-L. Li, K. Sohn, J. Yoon and T. Pfister, CutPaste: Selfsupervised learning for anomaly detection and localization, *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.9659-9669, 2021.

[7] X. Xie, C. Li, Y. Liu, J. Song, J. Ahn and Z. Zhang, Application of YOLOV4 algorithm with integrated attention mechanism in metal surface defect detection, *International Journal of Innovative Computing, Information and Control*, vol.19, no.2, pp.447-463, 2023.

[8] H. Kim, H. Bae, J. Park and M. Seo, Ensemble method for shape defect classification of rubber products, *ICIC Express Letters*, vol.17, no.8, pp.873-880, 2023.

[9] K. He et al., Deep residual learning for image recognition, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[10] D. M. Tax anf R. P. Duin, Support vector data description, *Machine Learning*, vol.54, pp.45-66, 2004.

Author Biography

Xiang Li received the B.S. degree in Automation and M.E. degree in Control Science and Engineering from Harbin Institute of Technology, Harbin, China, in 2015 and 2018, respectively, where he is currently working toward the Ph.D.

His current research interests include pattern recognition, unsupervised learning, and automated optical inspection.

Dong Zhou received the B.S. degree in Automation from the Harbin Engineering University, Harbin, China, in 2018 and the Ph.D. degree in Control Science and Engineering from the Harbin Institute of Technology, Harbin, China, in 2023.

He is currently a postdoc in the Department of Control Science and Engineering, Harbin Institute of Technology, His research interests include visual object tracking, deep reinforcement learning, vision-language robotics and embodied intelligence.

Xiangyu Shao received a B.S. degree in Automation from the Harbin Engineering University, Harbin, China, in 2016, and M.S. and Ph.D. degrees in Control Science and Engineering from the Harbin Institute of Technology, Harbin, China, in 2018 and 2022, respectively.

He is currently an Assistant Professor with the Department of Control Science and Engineering, Harbin Institute of Technology. His research interests include space robots, soft robots, sliding mode control and fractional order control.