PBSD-Net: Prismatic Battery Surface Defect Detection via Sliding Slice Amplification and Shunted Dynamic Snake Convolution

Ying Xu*, Bo Li*, Yikui Zhai*, *Senior Member, IEEE,* Feng Ke*, *Member, IEEE,* Jianhong Zhou*, Pasquale Coscia, *Senior Member, IEEE*, Angelo Genovese, *Senior Member, IEEE,* Vincenzo Piuri, *Fellow, IEEE,* and Fabio Scotti, *Senior Member, IEEE*

***Abstract*—Automatically detecting surface defects in prismatic battery is crucial for ensuring quality meets established standards. Traditional methods face challenges in accurately identifying these defects due to their minute and varied shapes and high density of distribution. To address these issues, we propose an innovative network for prismatic battery surface defect (PBSD-Net), which employs shunted dynamic snake convolution and focal modulation to detect surface defects in prismatic battery. This network is integrated into the 2D-AOI system. Firstly, we introduce sliding slice amplification (SSA) as a training strategy to enhance the network's ability to recognize densely clustered tiny defects. Secondly, we develop a novel method using the shunted dynamic snake convolution (SDSC) module and focal modulation (FM) to improve the extraction of deformation features, thereby addressing complex and sporadically scattered surface defects. By integrating the SDSC module and FM mechanism, the receptive field of the defect feature extraction network is expanded, enabling the acquisition of comprehensive defect edge features. Additionally, we introduce the quality focal loss (QFL) function to effectively tackle the issue of imbalanced sample types. Experimental results on the PBSD-RGB dataset demonstrate that our method achieves a mAP@50 of 85.8%, representing an improvement of approximately 7.7% over the baseline network. We have applied the PBSD-Net to an automatic defect detection system in a well-known battery production company. This enhancement significantly boosts the accuracy of surface defect detection in prismatic battery. The relevant code is at the https://github.com/yikuizhai/PBSD-Net.**

***Note to Practitioners*—Despite the existing limitations in the application of deep learning for automatic surface defect detection in the manufacturing industry, particularly within the context of battery production, we have introduced PBSD-Net. This innovative system demonstrates superior performance in identifying minute defects across various scales on battery surfaces. By integrating PBSD-Net with an automated detection framework and a high-performance server, we have developed a detection solution that is both efficient and robust. Empirical results obtained from real-world testing at a battery manufacturing facility underscore the significant potential of this approach for practical industrial applications. This advancement not only enhances the accuracy and efficiency of defect detection but also provides a solid foundation for the intelligent automation of battery production processes, thereby contributing to the ongoing technological progress in the manufacturing sector.**

***Index Terms*—Shunted dynamic snake convolution, object detection, prismatic battery surface defect detection, automatic defect detection system.**

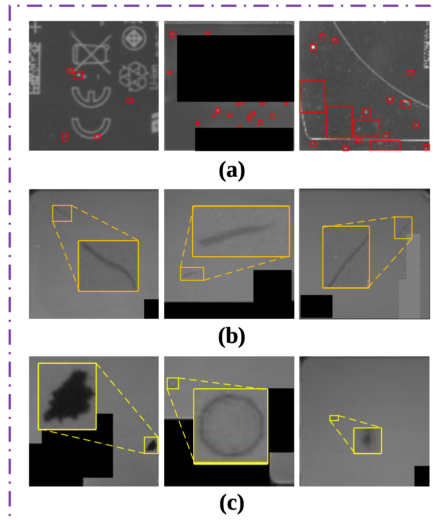

Fig. 1. Common types of surface defects on prismatic battery. (a) Dust. (b) Scratch. (c) Stain.

I. Introduction

With the advent of the digital age, intelligent electronic products, such as smartphones and laptops, have become integral components of our daily lives and professional pursuits. The prismatic battery, serving as the intelligent electronic products' energy storage and supply system, is crucial in determining the device's operational lifespan. In other words, the standard of the prismatic battery is an essential

This study was funded by Guangdong Basic and Applied Basic Research Foundation (No. 2021A1515011576), Guangdong Science and Technology Planning Project (No. 2021A0505030080, No. 2021A0505060011), Guangdong Higher Education Innovation and Strengthening School Project (No. 2020ZDZX3031, No. 2022ZDZX1032, No. 2023ZDZX1029), Wuyi University Hong Kong and Macao Joint Research and Development Fund (No. 2022WGALH19), Guangdong Jiangmen Science and Technology Research Project (No. 2220002000246, No. 2023760300070008390).

Ying Xu, Bo Li, Yikui Zhai and Jianghong Zhou are with the School of Electronics and Information Engineering, Wuyi University, Jiangmen 529020, China (e-mail: xuying117@163.com; libo_winner@163.com; yikuizhai@163.com; jackychou_lab@126.com)). Corresponding author: Yikui Zhai.

Feng Ke is with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: fengke@scut.edu.cn).

Pasquale Coscia, Angelo Genovese, Vincenzo Piuri, and Fabio Scotti are with Depatimento di Information, Università , Degli Studi di Milano, via Celoria 18, 20133 Milano (MI), Italy (e-mail: pasquale.coscia@unimi.it; angelo.genovese@unimi.it; vincenzo.piuri@unimi.it; fabio.scotti@unimi.it).

*Contributed to this work equally.

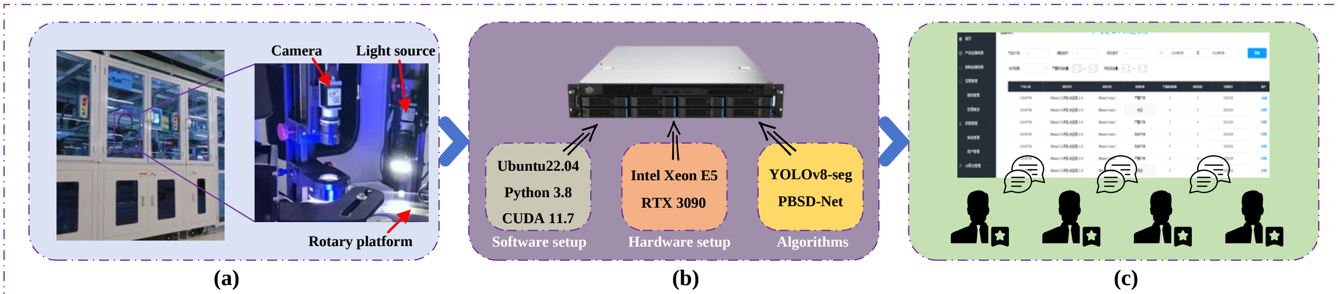

Fig. 2. Overall framework of the automatic system testing. (a) 2D-AOI system. (b) Detection system. (c) Expert evaluation.

element in assessing the normal functioning of an intelligent electronic product [1]. Presently, leading smartphone manufacturers, including Apple, Huawei, and Xiaomi, rely on third-party companies for the production of their battery. These third-party companies are under immense pressure to ensure high-quality control, particularly in the area of surface defect detection in prismatic battery [2]. Due to factors such as raw materials, manufacturing processes, operational procedures, and workshop environments, battery are bound to exhibit scratch, stain, and other imperfections, which can have a severe impact on product quality. The presence of defective battery in the market can harm the reputation of the company and pose a significant safety risk to users. Consequently, automatically identifying surface imperfections in industrial product is a crucial element of the industrial manufacturing process [3]. Undoubtedly, surface defect detection in battery is an indispensable aspect for ensuring battery quality. Moreover, effective control over such defects can avert economic losses due to scrapping and enhance product quality, thus promoting the market competitiveness of energy companies and fostering the development of related industries.

Currently, the assessment of the surface quality of prismatic battery relies on experienced workers, who are skill-intensive, inefficient, and imprecise. Owing to advancements in computational power, machine vision-based defect detection technology has become one of the primary methods for addressing the aforementioned issues. Consequently, the development of automatic defect detection algorithms based on prismatic battery surface images is of significant practical importance for enhancing the quality of battery. This method can significantly enhance the automation and production efficiency in prismatic battery manufacturing, greatly lower the labor intensity and costs tied to quality inspection, and offers broad application potential [4]. In the domain of defect detection using machine vision, object detection stands as the most commonly utilized task. Recently, techniques for object detection grounded in CNNs have made swift progress [5], [6], attracting substantial interest from researchers focused on industrial applications [7], [8]. Additionally, CNN-based object detection techniques have been successfully employed for identifying defects in steel strips [9], [10], [11], printed circuit board (PCB) [12], [13], and surface defects in solar cells [14], [15], [16], [17]. For instance, Chen et al. [10] introduced a block named enhanced deformation-feature extraction (EDE) for detecting complex, irregular defects in steel strips. Using deformable convolutions, they expanded the receptive field to capture complete defect textures. They also replaced the spatial pyramid pooling (SPP) with a coordination attention (CA) module, enhancing the ability of network to locate defects effectively. In [13], An enhanced YOLOv5 [18] framework, named YOLO-HMC, has been introduced for detecting defects on PCBs. This framework incorporates the HorNet structure, which greatly improves the capability to extract features of small defects within intricate PCB backgrounds. Chen [17] proposed a detector with bidirectional-path group-wise attention (BPGA), and a groundbreaking framework designed for defect detection in solar cells. Consequently, it significantly enhances the feature representation abilities of CNN models.

However, academic research on surface defect detection in prismatic battery has not received much attention. Currently, the academic community lacks datasets specifically tailored for detecting surface defects on prismatic battery. Most existing studies on industrial defect detection rely on publicly available datasets from other industrial domains, such as steel and solar cells [19], [20]. Constructing dedicated datasets requires considerable time and manpower, and the screening and labeling of data require trained personnel, which often becomes a significant deterrent for researchers. As a result, many studies focus on publicly available datasets to reduce research costs. While detection models developed from these datasets may not be suitable for addressing the unique challenges of prismatic battery surface defect detection. Thus, there is an urgent need for an effective, dedicated dataset specifically for prismatic battery surface defects. For prismatic battery defect detection, there are three main challenges: 1) In certain prismatic battery manufacturing facilities lacking dust-free environments, dust inevitably accumulates during production, making dust contamination a key factor affecting battery appearance. The challenge arises because this dust is densely packed and microscopic, making it difficult to detect. 2) Common surface defects on prismatic battery, such as scratch and stain, exhibit intricate and variable multiscale features, as shown in Fig.1. The complexity of these defect features poses a significant challenge for detection models. 3) In actual production, the quantity of dust is much higher than that of other defects, which can easily cause the model to overlook defects with fewer samples, leading to poor detection performance for those defects. These factors collectively contribute to the significant challenges faced in detecting surface defects in prismatic battery.

To address these issues, as illustrated in Fig.2, we have developed a training strategy tailored for dense and tiny targets, designed a convolution module for edge feature extraction, and introduced a focal modulation mechanism to enhance YOLOv5's capability in detecting surface defects of prismatic battery, in combination with the 2D-AOI system. The algorithm deployment process is as follows: Firstly, the workshop sends the manufactured batteries to the 2D-AOI system for imaging. Photos of both the front and opposite sites of the batteries are saved based on their serial numbers and uploaded to the server for storage. Taking into account customer data security, our algorithm is deployed on a server within the customer's intranet. Once the 2D-AOI system sends the inspection command, the positioning algorithm, YOLOv8-seg, is executed to detect the batteries. Post-processing is then applied to crop out the batteries. Finally, the detection algorithm for prismatic battery surface defect (PBSD-Net), performs inference on the cropped battery images, saving the detection results of both sides of each battery for further evaluation by expert. This automated approach ensures more efficient and accurate detection processes. The key contributions of this study are as follows:

1) The prismatic battery surface defect dataset in RGB (PBSD-RGB), is constructed, comprising three defect categories: dust, scratch, and stain. A total of 1370 images of battery sized 1145 × 800 pixels were annotated by experts to ensure accuracy and reliability in this study.

2) Sliding slice amplification (SSA), a new training strategy, is designed to dynamically slice and amplify critical regions of an image. This enhances the capability of network to detect densely clustered tiny defects.

3) PBSD-Net is proposed based on YOLOv5 for prismatic battery surface defect detection, in which a novel shunted dynamic snake convolution (SDSC) module is integrated to enhance global and multi-scale feature extraction, effectively improving detection precision and reliability.

4) The quality focal (QFL) loss function was introduced to alleviate the sample imbalance problem, achieving an mAP@50 of 85.8% on the PBSD-RGB dataset, which represents a 7.7% improvement over the baseline network.

The following sections of this paper are organized as follows. In Section II, related works is delineated. Following that, Section III delves into the presented training strategy and the approach to modular design, along with an examination of the loss function. Section IV is dedicated to the presentation of experimental findings. Section V provides a detailed overview of the algorithms' deployment and testing processes. Lastly, Section VI encapsulates conclusion of the study.

II. Related Work

A. Dense small object detection

Dense small object detection is a specialized area within object detection that focuses on identifying small-sized objects densely packed within an image, which occupy fewer pixels and often exhibit limited distinguishable features [21], [22], [23]. Detecting small objects remains a persistent challenge in object detection, demanding meticulous approaches and advanced methodologies. Currently, mainstream methods fall into two main categories: altering the receptive field of convolutional kernels and adding attention mechanisms. Building on YOLOv3, Xu et al. [24] proposed an enhancement to feature extraction performance. Their approach introduces the notion of multi-receptive fields, with the objective of amalgamating diverse receptive fields to elevate the accuracy in detecting smaller objects. In [25], the utilization of kroenecker convolution (K-conv) within the receptive field block (RFB) section dedicated to feature extraction was implemented to expand the receptive field while mitigating the loss of local contextual details. This strategic application enhances the spatial coverage of individual pixels across the feature map, thereby extending its scope of overall meaning and specific contextual details. Consequently, this augmentation significantly fortifies the fidelity of object extraction, particularly for diminutive targets. Jiang et al. [26] devised an augmented effective focusing feature through the integration of light module alongside attention mechanism modules. This strategy capitalizes on the distinct significance of various features within disparate channels, all while minimizing the network's computational overhead. Notably, this approach demonstrates remarkable efficacy in the detection of minute and moderate sized defects during the PCB manufacturing process. Zhou et al. [27] utilized a dense attention mechanism to steer profound attention cues towards the decoding phase. This guided integration of multiscale deep features by the decoder culminates in the production of a comprehensive saliency map. Notably, the dense attention scheme orchestrates the emphasis of deep features towards regions of defect, fostering heightened focus on areas of interest. Building on YOLOX, Xuan et al. [28] leveraged the coordinate attention mechanism (CA) to augment the detection efficacy of minor imperfections on PCB defects. Although attention mechanisms have improved model performance, they also introduce complex computational overhead. In order to reduce this complexity, Yang et al. [29] proposed the focal modulation mechanism to replace attention module. This approach offers enhanced efficacy and efficiency in capturing token interactions, making it a more potent mechanism for modeling. Additionally, Akyon et al. [30] adopted a strategy called slicing aided hyper inference (SAHI) only during the model inference stage. They conducted extensive experiments within the YOLO series and demonstrated that this inference strategy yields excellent results for dense small object detection.

The problem of dense small targets in battery surface defects highlights the need for exploring efficient small object detection strategies, which remains a worthy area of research.

B. Multi-scale object detection

In practical scenarios, target defects often exhibit uncertain sizes and positions in images, presenting a multiscale characteristic challenge. In computer vision, the focus on multiscale object detection using deep learning techniques is crucial for accurately identifying targets of interest across varying scales in images. YOLOv3 [31] incorporates multiscale feature fusion and prediction, introducing the feature pyramid network (FPN) to extract multiscale features. FPN integrates low-level coarse features with high-level fine-grained features through additional lateral connections and upsampling operations within the network, thus obtaining feature maps with

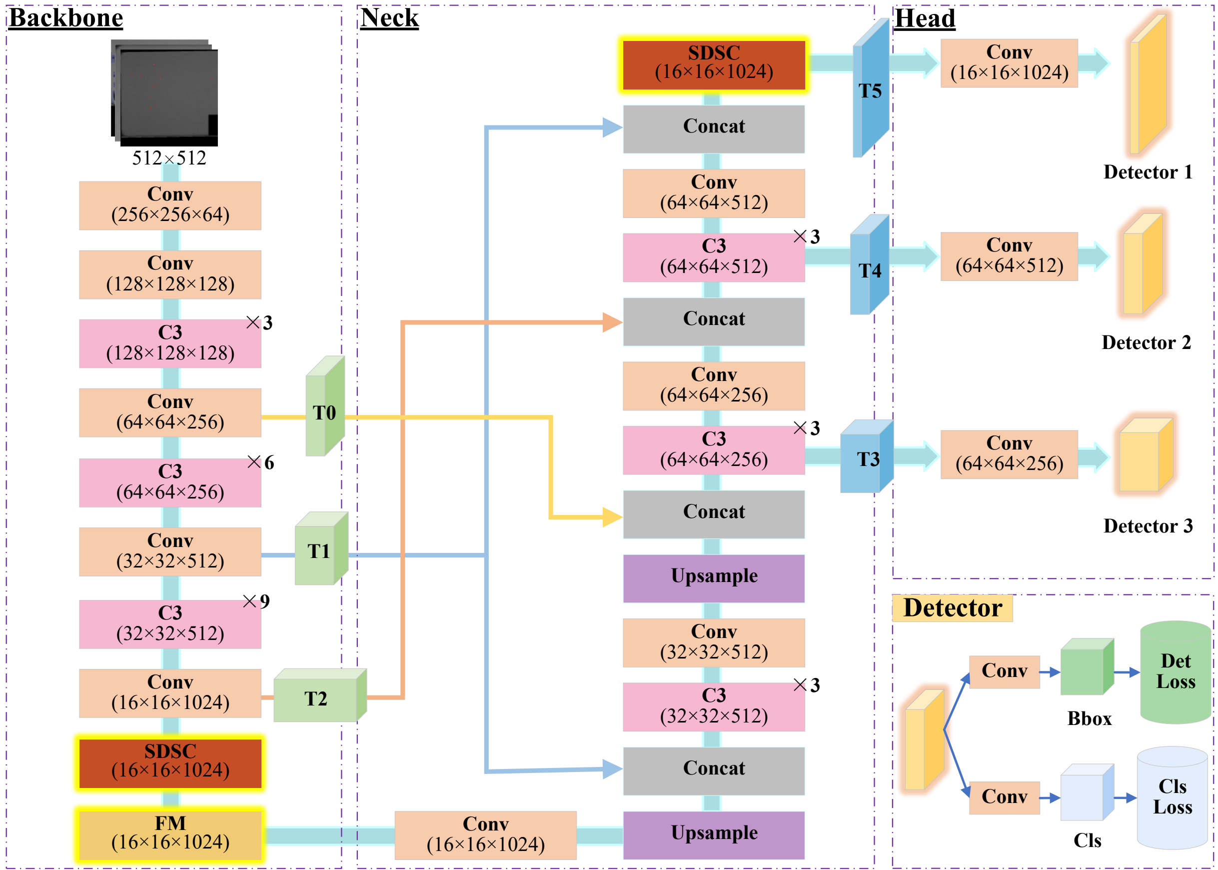

Fig. 3. Overall framework of the proposed PBSD-Net.

rich semantic information and high resolution. This enhances the detection capability for objects of different scales. The introduction of FPN has significantly influenced subsequent research on multiscale objects [32], [33], [34], [35]. Cui et al. [11] proposed a global affinity module (GAM) to discern semantic information from high-level features, guiding the localization and integration of lower-level details. They subsequently implemented a scale interaction module (SIM) to facilitate the amalgamation and interaction of feature data across different layers, yielding remarkable results in detecting objects across multiple scales. Luo et al. [36] designed the adaptive scale equilibrium pyramid convolution (ASEPC), which utilizes intra-layer feature correlations to fuse features across different scales. ASEPC demonstrated effectiveness in integrating shallow and deep layer information on public datasets DAGM, CAMO, and NEU-DET. To enhance the network's expressive power without increasing its depth or width, Chen et al. [37] proposed dynamic convolution. However, dynamic convolution requires substantial technical resources when dealing with extremely complex target scales, limiting its feature-expression capability. Addressing this issue, Qi et al. [38] introduced dynamic snake convolution, allowing convolution kernels to adaptively fit structural learning features without deviating too far from the target structure under constraint conditions. This approach achieved favorable results in cardiac and cerebral vascular segmentation.

In battery surface defect detection, the issue of multi-scale detection is a pressing concern. Researching efficient multi-scale defect target feature extraction modules is highly necessary.

C. Sample imbalance in object dection

In object detection tasks, the issue of sample imbalance is one of the key factors affecting model performance. Common forms of sample imbalance include class imbalance and the imbalance between positive and negative samples. To address this issue, researchers have proposed various loss function-based solutions. BCEWithLogitsLoss is a commonly used loss function in object detection. However, when dealing with sample imbalance, BCEWithLogitsLoss faces challenges related to the disproportion between positive and negative samples, often leading the model to favor majority classes, which results in poor detection performance for minority classes. Slide Loss [39] adjusts the decay curve of the loss value, further enhancing the model's ability to learn from hard-to-classify examples. Slide Loss is designed to gradually reduce the contribution of easy samples, allowing the model to focus on difficult examples, particularly in highly imbalanced datasets. This loss function demonstrates robustness in scenarios with extreme sample imbalance. Focal Loss [40] was introduced to address the severe imbalance between positive and negative samples. Its core idea is to adjust the sample's loss weight, reducing the impact of easy-to-classify examples (such as background or majority classes), and thus increasing focus on hard-to-classify examples. In detection tasks, Focal Loss introduces a modulating factor based on standard cross-entropy loss to dynamically control the contribution of each sample, effectively mitigating the problem of extreme positive-negative sample imbalance. Quality Focal Loss (QFL) [41] is an improved version of Focal Loss that addresses prediction quality issues. Unlike traditional Focal Loss, QFL not only considers the classification of positive and negative samples but also measures the quality of the bounding boxes. By combining the classification score with the Intersection over Union (IoU), QFL aims to reduce penalties for low-quality boxes, guiding the model to generate higher-quality detection results. By simultaneously optimizing both classification and quality, QFL improves both prediction accuracy and reliability while addressing sample imbalance.

Although sample imbalance has been widely studied on public datasets, its application in battery surface detection is lacking. Therefore, exploring a loss function that is tailored to battery surface detection tasks remains an open topic.

III. METHODOLOGY

From the above review, it is evident that in recent years, the rapid evolution of deep learning-enabled object detection methodologies has significantly advanced the application of surface defect detection in industrial products. This advancement has led to enhanced precision and efficiency in models to varying degrees. However, challenges persist in identifying surface irregularities in prismatic battery due to shared characteristics among defect classes, varied and intricate anomaly sizes, and their clustered arrangements. These factors collectively present hurdles to achieving precise defect detection. Therefore, to improve the precision and applicability of object detection algorithms for detecting surface defects in prismatic battery, inspired by methods described in the literature [10], [13], [29], [30], [33], [38], [41], [42], [43], this study utilizes the established YOLOv5 as the foundation for the object detection model and develops a prismatic battery surface defect detection network based on SDSC and FM, termed PBSD-Net, as depicted in Fig.3.

A. Overall Network Architecture

Based on YOLOv5 (anchor-free) [18], we have developed an architecture named PBSD-Net, depicted in Fig.3. The core architectural framework consists of three main components: backbone, neck, and head.

1) **Backbone**: The backbone facilitates the extraction of key features essential for target identification, constituting a pivotal aspect of the framework. Due to the dense distribution and diverse sizes of surface defects on prismatic battery, higher demands are placed on the backbone. Therefore, we have redesigned a feature extraction module called SDSC specifically for dense and multi-scale targets. Firstly, drawing inspiration from the concept of dynamic snake convolution [38], we utilize its powerful feature extraction capability to fully capture the global edge sizes of targets. Then, to obtain richer gradient information, we refer to the idea of the gradient flow main branch in YOLOv7 [43] and position the dynamic snake convolution accordingly, thereby achieving higher precision and spatial perception capability. Since the last convolutional layer output carries rich full-scale feature information, we embed SDSC behind it, hoping SDSC can obtain as much useful target feature information as possible. In addition, we have replaced the original SPPF, a technology used for image recognition and object detection that functions to extract and encode features from images at different scales, with focal modulation [29]. Focal modulation is a feature enhancement technique that shows potential for improving performance in image recognition and object detection tasks through its unique mechanism.

2) **Neck**: Processes and integrates features from the shallow, middle, and deep layers of the Backbone. The original model's Feature Pyramid Network (FPN) combined with the Path Aggregation Network (PAN) faces challenges in accurately locating defect objects in complex backgrounds. This difficulty arises because defects often occupy only a small portion of the battery's surface area, making them hard to detect with precision. Moreover, the integration of PAN demands additional computational resources, thereby impacting detection speed. To address these issues, we have opted to retain only the FPN while integrating SDSC module within the final feature layer. This strategic combination aims to achieve several objectives: enhanced detection accuracy by focusing on the FPN and SDSC, improving the model's ability to detect small and subtle defects in complex backgrounds; reduced computational overhead by eliminating PAN, thus improving detection speed without compromising accuracy; improved feature integration by incorporating SDSC in the last layer, enhancing the processing of detailed features and ensuring minute global targets are more accurately captured; minimized information loss by mitigating the loss of information between input and output feature maps, preserving essential details throughout the feature extraction process; and optimized network efficiency by fine-tuning the network's feature processing efficacy, leading to more robust and reliable defect detection, particularly where defects are small and dispersed across the battery surface. This redesigned architecture, focusing on a streamlined yet powerful feature integration strategy, positions the model to perform more effectively in detecting defects under challenging conditions, while maintaining a balance between accuracy and computational efficiency.

3) **Head**: The primary function of the head in our model is to predict target categories and their positions. To optimize efficiency for edge deployment, a 1×1 convolution is used for prediction, enhancing inference speed by reducing computational complexity. Bounding box regression employs a loss formulation combining QFL and CIoU Loss. QFL addresses class imbalance by giving higher weights to difficult examples, while CIoU Loss ensures accurate and robust bounding box predictions by considering overlap, center distance, and aspect ratio. For box filtering, non-maximum suppression (NMS) is used to eliminate redundant and overlapping boxes, ensuring precision in detection results. This design enables efficient, accurate predictions suitable for real-time applications on resource-constrained edge devices.

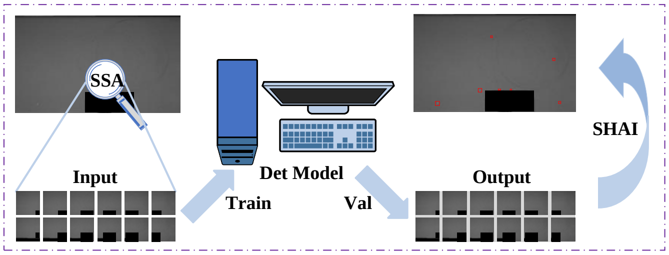

Fig. 4. Training strategy and inference process.

B. SSA Training Strategy

Due to the production process, batteries inevitably accumulate minute dust and stain, which adversely affect their aesthetic appeal. Therefore, detecting dust and stain is crucial in battery appearance inspection. However, traditional manual quality inspection finds it challenging to accurately quantify the degree of contamination caused by these small defects. Meanwhile, in the field of detecting small objects, prevalent deep learning approaches often incorporate attention mechanisms into their network architectures. This addition enhances the network's ability to accurately identify and locate tiny objects against complex backgrounds. While attention mechanisms improve the network's ability to detect small objects, they also increase its complexity, resulting in higher computational demands. SAHI [30] is an inference strategy that enhances the model's ability to detect small objects during the inference stage. The principle of SAHI is as follows.

1) **Image slicing**. SAHI initially divides the large-sized image into multiple smaller-sized image blocks. To comply with the input size requirements of object detection models, these image blocks are usually reduced in size compared to the initial image. During the process of slicing, overlapping regions can be set to avoid inaccurate detection caused by target being cut off at the edges.

2) **Individual detection**. For each image slice, SAHI employs a pre-selected object detection model (such as YOLO, etc.) for independent object detection. Each slice is treated as a separate image for analysis, and the model will output the categories, positions, and confidences of all detected objects in that slice.

3) **Result aggregation**. After the detection process is completed, SAHI aggregates the detection results obtained from all image slices back into the original image. This step takes into account the overlapping regions between slices and processes redundant detection results in these areas using specific algorithms, such as NMS, to merge or select the best detection boxes.

Drawing inspiration from SAHI, we devised a training strategy tailored for detecting small target defects, as depicted in Fig.4. In essence, we mimicked the human observation of small objects using a magnifying glass. In other words, the model learns target features as if it were examining them through a magnifying glass. Of course, setting appropriate window sizes and sliding speeds requires experimentation to adjust them to meet task requirements. Here, the window size indicates the actual size of the image input to the network, while the sliding rate denotes how quickly or slowly the network learns from the entire image. We represent the overlap between the forward and backward (up and down) sliding windows as shown in Equation (1).

$$SSA\_sliding\_ratio = 1 - coverage. \tag{1}$$

After training the model, SAHI is employed for final inference, mapping the results of small image detections back to the original image.

C. SDSC Module

SDSC, as depicted in Fig.5(a), builds upon the C3 module by introducing gradient flow diversion to enable the network to acquire richer gradient flow information. The core feature extraction module is replaced with the DysnakeConv module, drawing inspiration from Deformable Convolution [44]. By altering the shape of the convolution kernel, DysnakeConv focuses on the core structural features of tubular structures. The convolution kernel of DysnakeConv can conform freely to learn features of structures while remaining constrained to the target structure. Upon utilizing SSA, we were surprised to find that hidden features of defects were visualized, with some hidden features exhibiting tubular-like structures. As shown in Fig.1, the defects processed with SSA amplify the complex edge information for object detection. Therefore, we adopted DysnakeConv, as depicted in Fig.5(b). For a 2D convolutional kernel (3 × 3) K_s, it is represented as:

$$K_s = \{(k_x \pm i, k_y \pm i) \mid i \in \{0, 1\}\}. \tag{2}$$

To grant more flexibility to the convolutional kernel and enable it to concentrate on intricate geometric features of the target, inspired by deformable convolutions, the introduction of a deformation offset Δ was adopted. Nevertheless, when the model is given unrestricted freedom to learn deformation offsets, the receptive field frequently strays from the intended target, particularly in the case of elongated tubular structures. Therefore, Qi et al. [38] utilized an iterative approach, depicted in Fig.5(c), wherein they sequentially determined the subsequent position of each target for processing. This method ensured consistent focus without the perceptual range extending excessively due to significant deformation offsets.

In the dynamic snake convolution approach, the convolutional kernel is linearized along the x and y axes. For instance, with a kernel size 9 × 9, each grid position $K_{i\pm s}$, where s ranges from 0 to 4, represents the horizontal distance from the central grid $(x_i \pm s, y_i \pm s)$. The selection of these grid positions within the kernel K follows an accumulative process. Starting from the central position K_i, each subsequent position K_{i+1} relative to K_i is adjusted by an offset Δ chosen from the set $[-1, 1]$. Accumulating these offsets Σ ensures the linearity of the convolutional kernel's structure. The variation along the x and y axes, as depicted in Fig.5(c), manifests in both positive and negative directions as follows:

$$K_{i,j+d} = \begin{cases} (x_{i+d}, y_{i+\Delta})_x = \left(x_i + d, y_i + \sum_i^{i+d} \Delta\right) \\ \\ (x_{j+\Delta}, y_{j+d})_y = \left(x_j + \sum_j^{j+d} \Delta, y_j + d\right) \end{cases}, \tag{3}$$

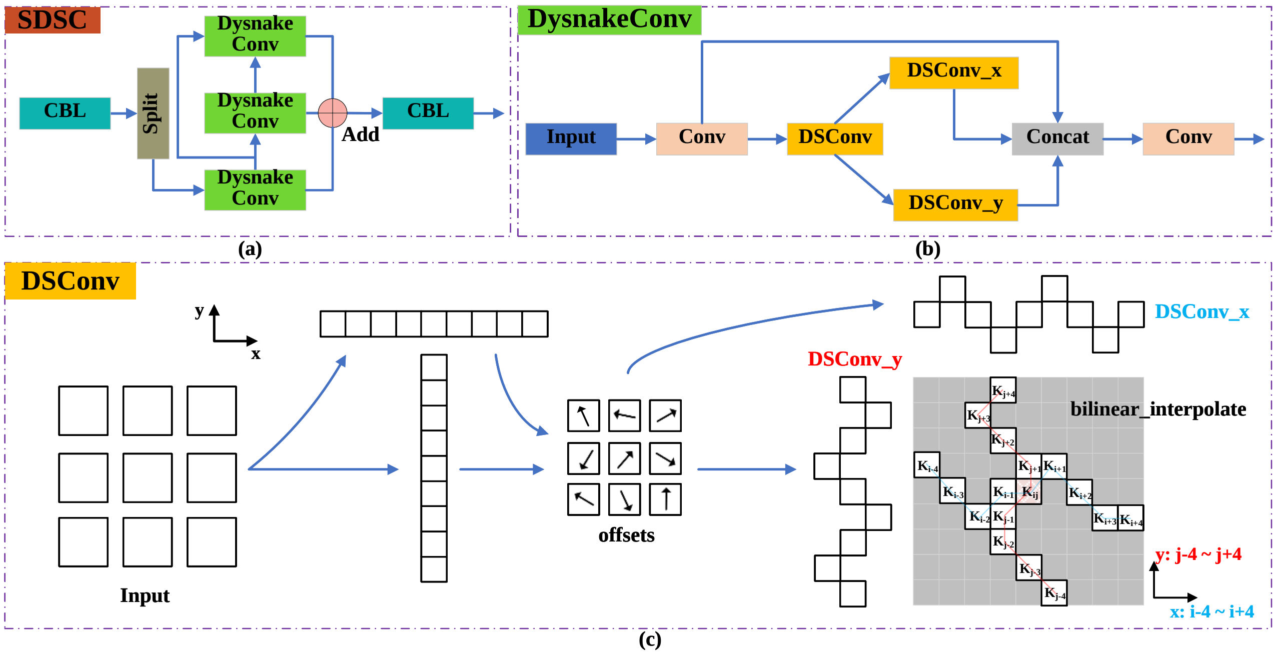

Fig. 5. SDSC Module. (a) SDSC Structure. (b) DysnakeConv Structure. (c) DSConv Strategy.

$$K_{i,j-d} = \begin{cases} (x_{i-d}, y_{i+\Delta})_x = \left(x_i - d, y_i + \sum_{i-d}^{i} \Delta\right) \\ (x_{j+\Delta}, y_{j-d})_y = \left(x_j + \sum_{j-d}^{j} \Delta, y_j - d\right) \end{cases} . \quad (4)$$

As the offset Δ is commonly expressed in decimal format, while coordinates are typically integers, bilinear interpolation is employed, denoted by:

$$\boldsymbol{K} = \sum_{K'} b\left(K_x, K_x'\right) \cdot b\left(K_y, K_y'\right) \cdot \boldsymbol{K'}. \quad (5)$$

In equations (3) and (4), $\boldsymbol{K}$ represents the fractional indices, $\boldsymbol{K'}$ denotes all integer spatial positions, and b denotes the bilinear interpolation filter, which can be decomposed into two separate one-dimensional kernels. This setup ensures avoidance of redundancy, while adopting a scholarly style and reorganizing sentence structure.

$$\boldsymbol{K_F} = \boldsymbol{K_{D_1}} \oplus \boldsymbol{K_{D_2}} \oplus \boldsymbol{K_{D_3}}, \quad (6)$$

where $\boldsymbol{K_{D_1}}$, $\boldsymbol{K_{D_2}}$, and $\boldsymbol{K_{D_3}}$ represent the processing through three DysnakeConv operations, which capture different levels of geometric information and edge details from the input feature map. $\boldsymbol{K_F}$ represents the fusion of feature information from different levels. This helps extract local edge information and also introduces multi-scale feature fusion between different convolutional layers.

D. Focal Modulation Mechanism

The success of Transformer depends on the self-attention mechanism, which supports global input dependent interactions, but despite these advantages, due to the self-attention quadratic computational complexity is less efficient, especially for high resolution inputs. In contrast, FM greatly reduces computational complexity and can effectively model interactions between tokens without using self-attention [35]. As shown in Fig.6, the core ideas of FM are as follows.

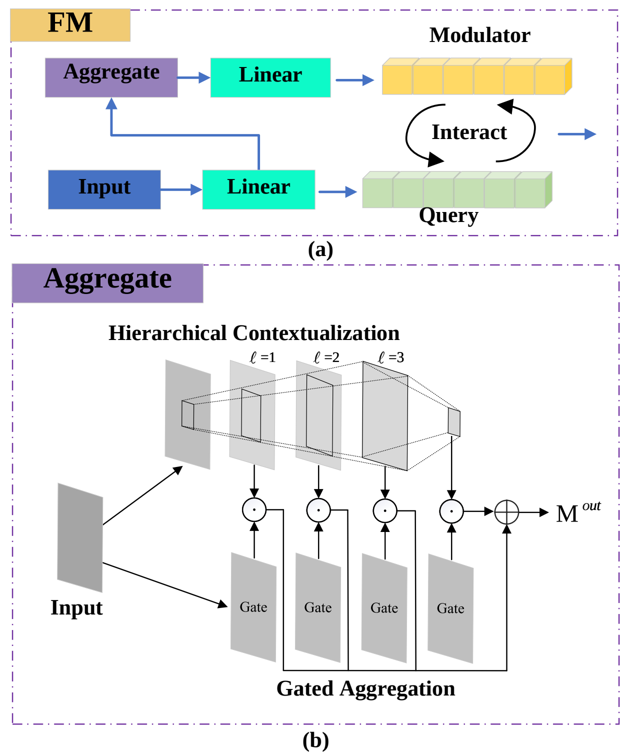

Fig. 6. FM Mechanism. (a) FM Structure. (b) Aggregate Strategy.

1) **Pre-linear Projection**: The input feature x is transformed into three components: query vector $\boldsymbol{q}$, context feature $\boldsymbol{m_i^\ell}$, and gating weights $\boldsymbol{g_i^\ell}$ through a convolution operation:

$$[\boldsymbol{q},\ \boldsymbol{m}_i^\ell,\ \boldsymbol{g}_i^\ell] = \mathrm{Split}(\mathrm{Conv}_{dw}(x),\ (C,\ L+1)), \quad (7)$$

Fig. 7. Empirical Size Distributions of Surface Defects in the PBSD-RGB Dataset: Dust, Scratch, and Stain Categories with Individually Scaled Axes.

where C is the number of input channels, and L is the number of focal levels.

2) **Context Aggregation**: For each focal level ℓ, the context feature $\boldsymbol{m_i^\ell}$ is processed through a series of convolutions, and the outputs are aggregated by weighted summation using the gating weights to form the total context feature $\boldsymbol{ctx_{all}}$:

$$\boldsymbol{ctx_{all}} = \sum_{\ell=1}^{L} \mathrm{Conv}(\boldsymbol{m}_i^\ell) \cdot \boldsymbol{g_i^\ell}[:, \ell]. \tag{8}$$

Then, global context information is extracted via global average pooling and weighted using the gating weight from the last focal level:

$$\boldsymbol{x_i} = \boldsymbol{ctx_{all}} + \mathrm{GlobalPooling}(\boldsymbol{m_i^\ell}) \cdot \boldsymbol{g_i^\ell}[:, \ell]. \tag{9}$$

3) **Modulation**: Using a linear transformation $\mathrm{h}(\cdot)$ to derive the modulator $\mathbf{M} = \mathrm{h}(\mathbf{M}^{\mathrm{out}})$. This modulator is then fused with the query $\boldsymbol{x_i}$ by element-level multiplication $\odot$ to give the final output $\boldsymbol{y_i}$:

$$\boldsymbol{y}_i = \mathrm{q}\left(\boldsymbol{x}_i\right) \odot \mathrm{h}\left(\sum_{\ell=1}^{\ell+1} \boldsymbol{m}_i^\ell \cdot \boldsymbol{g}_i^\ell\right), \tag{10}$$

where $\mathrm{q}(\cdot)$ is a query mapping function.

E. Quality Focal Loss

Due to the extreme imbalance in defect category samples, the model tends to bias toward categories with more samples, resulting in a significant discount in training effectiveness. Meanwhile, YOLOv5 trains the classification score and IoU score separately during training, but considers them together during inference. This leads to significant discrepancies between training and inference performance. To address these issues, we replaced the original BCEWithLogitsLoss of YOLOv5 with Quality Focal Loss, while retaining CIoU Loss for computing bounding box losses.

$$QFL(\sigma) = y\log(\sigma) - |y-\sigma|^2((1-y)\log(1-\sigma). \tag{11}$$

Quality Focal Loss integrates classification and IoU scores by weighting classification labels with IoU values, reflecting the overlap between predicted and ground truth bounding boxes, and selecting the lowest-cost match for optimal alignment in cases of multiple matches. It aligns predicted bounding boxes with actual ones by calculating IoU and selecting those exceeding a threshold as final predictions. Each box is assigned a class score, which, after sigmoid transformation, corresponds to σ in Equation (9), where y represents the label multiplied by the corresponding IoU score, signifying the joint product of classification and IoU scores.

IV. EXPERIMENTS AND RESULTS

A. Implementation Particulars

The experiments used PyTorch 1.13.1, with a learning rate starting at 0.0001 and increasing to 0.01 over 200 epochs, without pre-trained weights. AdamW was the optimizer, and all trials ran on a Nvidia RTX 3060 GPU.

*1) **PBSD-RGB Dataset**:* As shown in Fig. 2, we utilized AOI camera to capture images of the front and back sides of battery. Due to the presence of considerable irrelevant background information in the battery images, we annotated the battery positions using Labelme and trained a lightweight segmentation model based on YOLOv8 [45] to preprocess the images, segment the battery, and perform subsequent annotation processing. Currently, our PBSD-RGB Dataset comprises three defect categories: dust, scratch, and stain. A total of 1370 battery images (including both front and back sides) with dimensions of 1145 × 800 pixels were annotated, containing 27899 instances of dust, 1018 instances of scratch, and 1920 instances of stain.The dataset was partitioned into training, validation, and test subsets in an 8:1:1 ratio. The evaluation metrics for subsequent comparative experiments are conducted on the test set.

To illustrate the size distribution of defects, Fig.7 the width and height distributions for dust, scratch, and stain. Dust is small, with widths ranging from 5 to 20 pixels and heights from 5 to 15 pixels. Scratch has a wider size range, with widths from 10 to 200 pixels and heights from 10 to 175 pixels, reflecting their elongated nature. Stain is slightly larger than dust, with widths between 10 to 30 pixels and heights between 5 to 25 pixels. These distributions align with the typical characteristics of each defect: dust appears as a small white dot or fine particle; scratch manifests as a long, dark linear mark from mechanical contact; stain shows as a blurry patch or ring caused by residual contamination. These distributions are derived from manual annotations and validated against real-world inspection data to ensure representativeness.

*2) **Object Detection Evaluation**:* In the task of detecting surface defects on prismatic battery, similar to the YOLO series, the evaluation metrics include Precision (P), Recall (R), and mAP. These metrics can be defined by the following formulas:

$$P = \frac{TP}{TP+FP}, \tag{12}$$

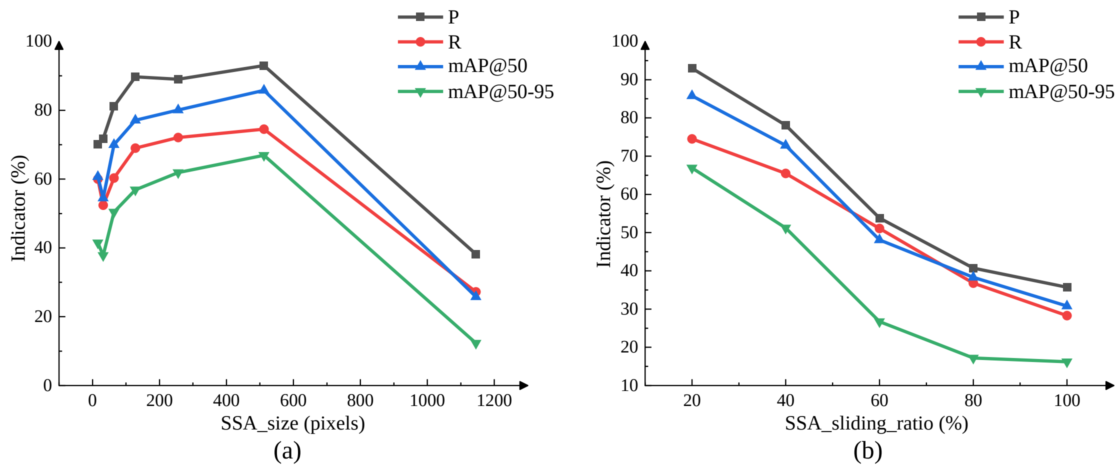

Fig. 8. Comparison of Different SSA Configurations. (a) Comparison of Different Sliding Window Sizes. (b) Comparison of Different Sliding Window Rates.

$$R = \frac{TP}{TP + FN}, \tag{13}$$

$$AP = \int_0^1 P(r)dr, \tag{14}$$

$$mAP = \frac{\sum_{i=1}^{n} AP_i}{n}, \tag{15}$$

where TP stands for true positives, FN for false negatives, FP for false positives, and n for the number of classes. When mAP@50 denotes the mAP value at IoU $= 0.5$. The metric mAP@50-95 refers to the average precision values computed over IoU thresholds ranging from 0.5 to 0.95 at 5% intervals. This results in ten distinct mAP values, which are then averaged to yield the final score.

B. Ablation Experiment

*1) **Comparison of Different SSA Settings**:* Based on the proposed PBSD-Net, we investigated the influence of different SSA settings on the experimental outcomes. As shown in Fig.8(a), we conducted comparative experiments with SSA_sliding_ratio set to 0.2 on image sizes of 16 × 16, 32 × 32, 64 × 64, 128 × 128, 512 × 512, and the original size (1145 × 800). Among them, the size of 512 × 512 achieved the best performance, with P, R, mAP@50, and mAP@50-95 reaching 93%, 74.5%, 85.8%, and 66.9%, respectively. As illustrated in Fig.8(b), using the SSA_size of 512 × 512, we further explored the effect of different SSA_sliding_ratio values: 0.2, 0.4, 0.6, 0.8, and 1. The results show that detection performance gradually deteriorates as the SSA_sliding_ratio increases. This phenomenon can be attributed to three factors: Larger sliding ratios cause each window to cover broader areas, thereby diluting critical local features and weakening the model's sensitivity to fine-grained defects; Overlapping regions introduced by large sliding ratios generate redundant candidate boxes, which increase the burden on classification and localization; As the number of candidate boxes grows, the merging process—especially for large-scale defects like scratches—becomes more error-prone, negatively impacting final detection accuracy. To address redundancy, our method incorporates non-maximum suppression (NMS) to suppress overlapping bounding boxes. Nevertheless, careful selection of the SSA_sliding_ratio remains crucial. Although a higher sliding ratio increases coverage, it introduces trade-offs in feature clarity and box merging. Therefore, to balance detection quality and computational efficiency, we selected an SSA_size of 512 × 512 and an SSA_sliding_ratio of 0.2 in the subsequent experiments.

*2) **Comparative Analysis of Various Parts**:* To assess the proposed method's effectiveness, we utilized YOLOv5 (anchor-free) as the baseline network and progressively tested the performance of each enhancement (SSA, SDSC, and FM). Table I shows the results of ablation experiments for each added improvement. From the experimental results in Table I, primarily, regarding training strategy, SSA enhances the representation of surface defects on phone battery, thereby emphasizing their location and characteristics. This method effectively boosts the detection network model's ability to learn surface defect features on phone battery. Compared with the baseline network YOLOv5's 23.2%, the mAP increased to 78.1%. To verify the effectiveness of SDSC at different feature layers, we conducted stepwise ablation experiments on the Backbone and Neck networks. As shown in Fig.9, T0, T1, and T2 represent the shallow, intermediate, and deep feature maps of the Backbone, respectively. Deeper feature maps provide more global information. The experimental results in Table I demonstrate that extracting features after T2 with SDSC achieves the best performance. Furthermore, to validate the effectiveness of FM, we replaced the original SPPF with FM after SDSC-T2, as shown in Fig.9(d). Compared to SPPF, FM not only handles input images of different sizes, providing more flexible feature extraction capabilities, but also more accurately identifies and localizes objects within images, especially useful for detecting small or hard-to-spot objects in complex backgrounds. The use of FM significantly enhanced the model's ability to extract defect features and locate targets, raising the mAP to 81.6%.

Finally, to validate the impact of SDSC in the Neck network, ablation experiments were performed on the T3, T4, and T5 feature maps, representing small, medium, and large target sizes, respectively, as shown in Fig.9(e)-(g). Large target sizes often contain more detailed feature information, facilitating the detection of small objects. By integrating

TABLE I
ABLATION EXPERIMENTAL RESULTS

Baseline	SSA	Backbone				Neck			mAP@50 (%)	Params	GFLOPs
		SDSC-T0	SDSC-T1	SDSC-T2	FM	SDSC-T3	SDSC-T4	SDSC-T5			
✓	×	×	×	×	×	×	×	×	23.2	25.1M	59.7
✓	✓	×	×	×	×	×	×	×	78.1	25.1M	59.7
✓	✓	✓	×	×	×	×	×	×	77.1	25M	62.8
✓	✓	×	✓	×	×	×	×	×	78.3	25.4M	63.9
✓	✓	×	×	✓	×	×	×	×	79.3	21.7M	62.5
✓	✓	×	×	✓	✓	×	×	×	81.6	22.1M	62.9
✓	✓	×	×	✓	✓	✓	×	×	81.9	22.3M	63.7
✓	✓	×	×	✓	✓	×	✓	×	82.5	22.4M	64.1
✓	✓	×	×	✓	✓	×	×	✓	**83.3**	23.5M	65.2

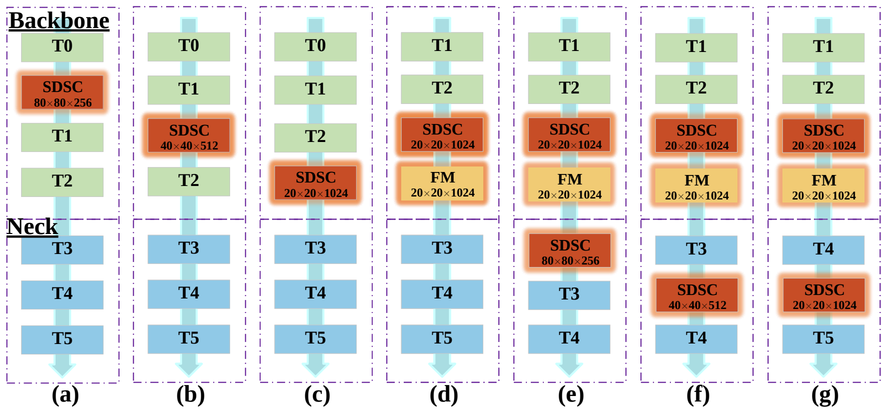

Fig. 9. Comparison of the Ablation Structures in Different Locations. (a)-(c) The structure in which SDSC is embedded following the T0-T2 feature layers. (d) Based on (c), the structure where FM replaces SPPF. (e)-(g) Based on (d), the structure in which SDSC is embedded before the T3-T4 feature layers.

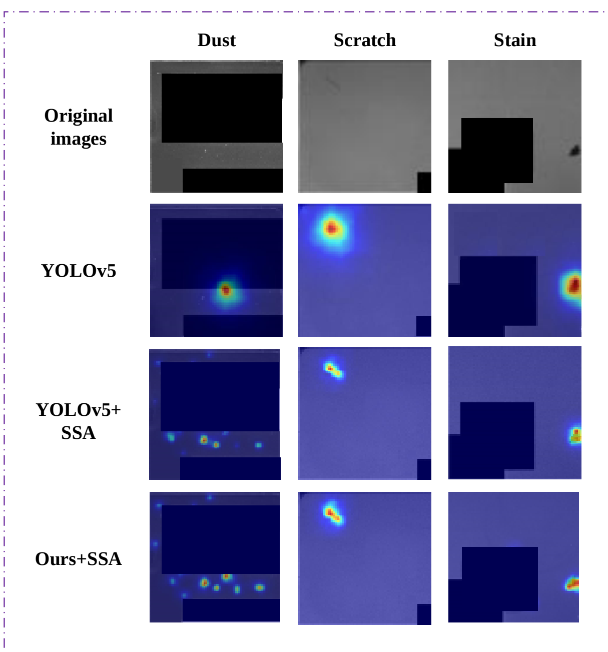

Fig. 10. Contrasted with Baseline Network Heatmap.

TABLE II
DIFFERENT CLASSIFY LOSS EXPERIMENTAL RESULTS

Methods	mAP@50(%)	Dust	Scratch	Strain
OurWork+BCEWithLogitsLoss	83.3	70.1	87.6	92.2
OurWork+SlideLoss	84.3	72.1	87.6	93.3
OurWork+FocalLoss	84.2	70.6	88.4	93.6
OurWork+QFL	**85.8**	**73.2**	**88.5**	**95.6**

SDSC into the T3 layer of the Backbone and the T5 layer of the Neck, and designing multiple sets of dynamic snake convolutions, we extracted the final feature layer. The SDSC module, inspired by gradient shunting, retains rich gradient information, enabling the network to achieve higher-quality spatial information aggregation, capturing edge information more accurately. Notably, due to the SDSC module's strong edge feature extraction capabilities, the detection network captures a broader range of defect scale features, improving the mAP@50 by over 4% compared to the baseline. Additionally, to further explore the effectiveness of the proposed method, feature heatmap analyses were conducted for both the baseline and the proposed method, as shown in Fig.10. Compared to the baseline network, PBSD-Net demonstrates a greater focus on small targets and better edge information extraction for complex defects.

C. Comparative Experiment

*1) **Comparison of Various Classification Loss Functions**:* To assess the efficacy of QFL, this study juxtaposes its algorithm against BCEWithLogitsLoss, SlideLoss, and FocalLoss. Each employs CIoU Loss for calculating bounding box losses. The comparative findings are delineated within Table II. The comparative analysis indicates that, compared to the original BCEWithLogitsLoss, QFL enhances the mAP value of Dust by about 2% by combining classification scores and IoU scores. Additionally, QFL directs attention to the less common examples of Scratch and Strain, resulting in improved mAP values for each. SlideLoss is a commonly used classification

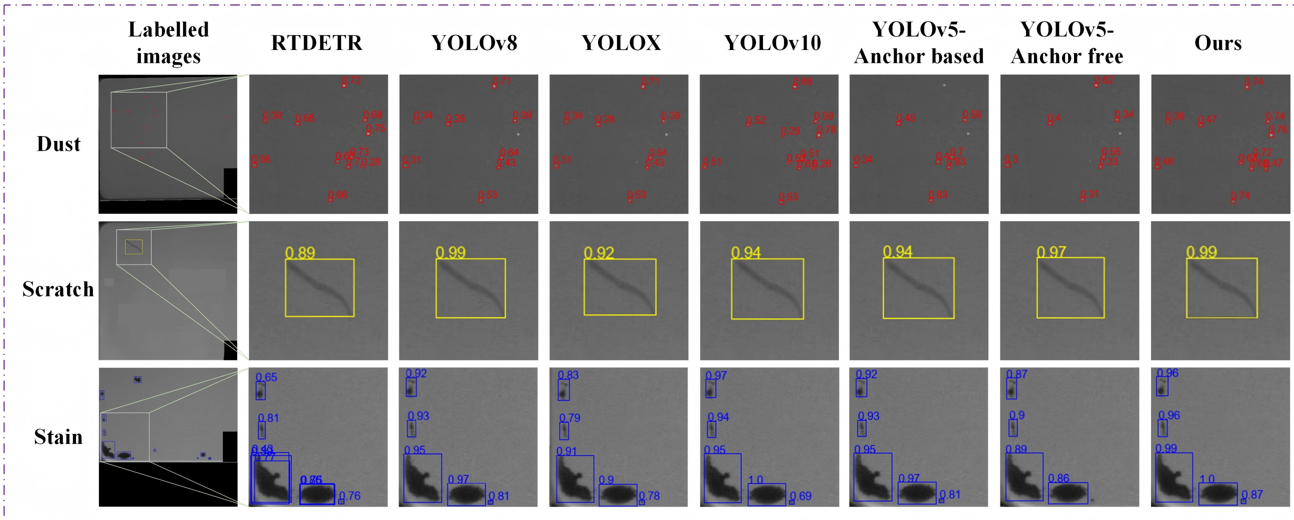

Fig. 11. Comparison of different methods on PBSD-RGB.

TABLE III
DIFFERENT OBJECTION DETECTION METHODS EXPERIMENTAL RESULTS ON PBSD-RGB

Methods	Params	mAP@50(%)	Dust	Scratch	Strain
YOLOv7	21M	75	76.7	59.9	48
YOLOv8	25.8M	78.9	64.1	83.2	89.3
YOLOX	25.3M	80.5	76.8	78	86.6
RTDETR	41.9M	80	**79.9**	72.8	87.2
YOLOv9	20.2M	80.4	69.9	80.5	90.9
YOLOv10	16.5M	82	69.7	83.8	92.5
YOLOv5-Gold	27.8M	79.1	69.1	77.1	91.2
YOLOv5-SCConv	43.4M	75.7	60.3	75.8	90.9
YOLOv5-Anchor based	20.9M	74.2	62.7	75.5	84.4
YOLOv5-Anchor free	25.1	78.1	63.3	84.4	86.5
Ours	23.5M	**85.8**	73.2	**88.5**	**95.6**

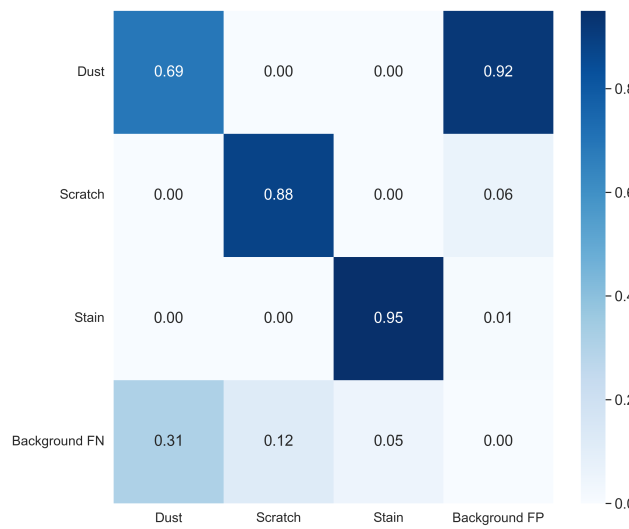

Fig. 12. Confusion matrix of the PBSD-Net on PBSD-RGB.

loss function to address sample imbalance, focusing on challenging training examples. Dust, due to its small shape and dense location, represents difficult training examples within the dataset. Compared to BCEWithLogitsLoss, Dust's mAP improves by 2%, demonstrating the model's focus on challenging examples. Similarly, FocalLoss is another commonly used classification loss function aimed at addressing class sample imbalance issues. Compared to BCEWithLogitsLoss, the mAP values of the less common Scratch and Strain are improved. However, SlideLoss and FocalLoss fail to account for the interplay between classification scores and IoU scores in training, consequently leading to reduced mAP during inference. The QFL used in this paper effectively handles classification scores and IoU scores during training, while also mitigating class sample imbalance issues, thus achieving optimal detection performance.

2) ***Comparative Analysis of Various Defect Detection Algorithms on PBSD-RGB:*** To assess the holistic efficacy of PBSD-Net, this paper compares the detection performance of various object detection network models, including YOLOv5-anchor based [18], YOLOv5-anchor free [18], YOLOv7 [43], YOLOv8-v10 [45], RTDETR [46], YOLOv5-Gold [47], YOLOv5-SCConv [48] and YOLOX [49]. As shown in Fig.11, red, yellow, and blue respectively represent dust, scratch, and strain. From the detection image results, it can be seen that YOLOv8, RTDETR, and YOLOX perform well in detecting dust defects. However, none of them can fully detect dust defects, and even false positives occur. The similarity between the background of the surface defect image of the prismatic battery and dust, coupled with the dense distribution and small size of dust particles, makes it challenging for the detection network to fully capture its positional information, thus resulting in detection errors. The method proposed in this paper effectively enhances the detection of such defects through QFL, which helps the detection network focus on hard-to-train samples. Compared with other YOLOv5 and its improved methods, the method proposed shows better improvement. When detecting scratch and strain defects, the SDSC module proposed can accurately capture the scale-complex defect features through dynamic snake convolution, and has the highest confidence compared to other detection methods.

Moreover, the experimental findings presented in Table III delineate the performance disparity between PBSD-Net and conventional algorithms. From the comparison experiment

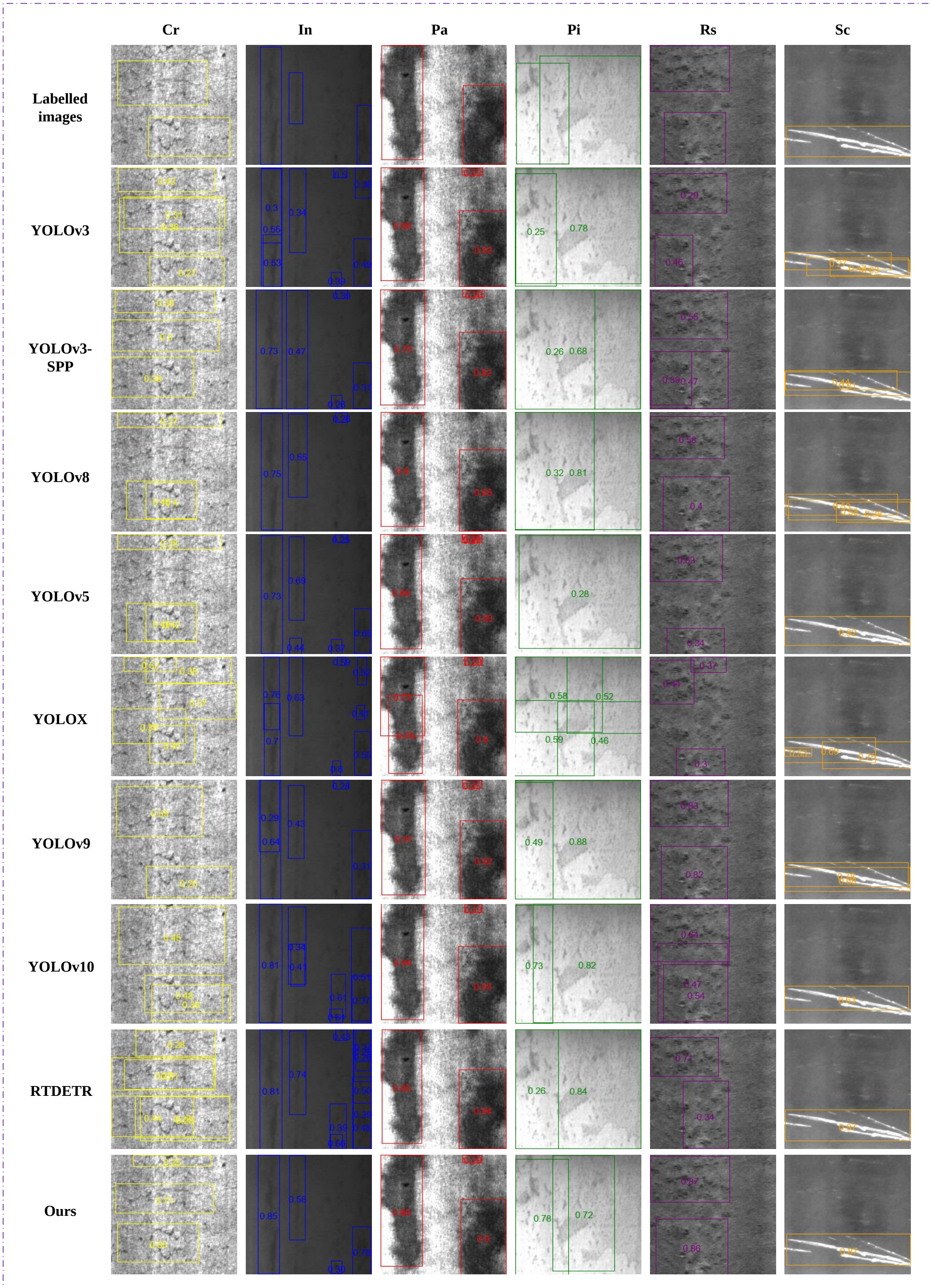

Fig. 13. Comparison of different methods on NEU-DET.

results, PBSD-Net shows significant improvement compared to other YOLO series networks in detecting scratch and strain defects. This stems from the limited feature extraction capacity of YOLO series networks when confronted with extensive irrelevant noise and intricate, dynamic features. This study presents the application of SSA for image enlargement, proposes the SDSC module and FM mechanism to integrate dynamic snake convolution into the feature extraction process of the detection network, and introduces QFL to aid the network in prioritizing difficult-to-train samples, thereby enhancing the network's feature focus. Although PBSD-Net's performance in detecting dust defects, as measured by AP value, is relatively average compared to other detection networks such as YOLOv7, RT-DETR, and YOLOX, which is characteristic of typical YOLO models, this can be attributed to the excessive background noise interference associated with dust defects. Moreover, there remains a challenge concerning the insufficient denoising capability of the detection network during the learning phase. However, when it comes to detecting other defects, PBSD-Net demonstrates significant prowess, achieving an mAP of 85.8%, surpassing alternative detection algorithms. This underscores PBSD-Net's effectiveness in enhancing surface defect detection accuracy, particularly in prismatic battery applications. Compared to the number of parameters in YOLOv10, our method still has room for further improvement.

To further analyze the model's limitations, we examine the confusion matrix shown in Fig.12. It can be observed that the detection accuracy for the Dust category is relatively lower (0.69) compared with scratch (0.88) and stain (0.95). Meanwhile, the false negative rate for Dust reaches 0.31, and the background false positive rate is also high (0.92), indicating that a considerable number of Dust samples were missed, and many background regions were misclassified as Dust. This phenomenon can be attributed to the visual characteristics of Dust defects, which are often small in size, low in contrast, and lack clear structural features, making them more likely to be confused with background textures.

*3) **Comparative Analysis of Various Defect Detection Algorithms on NEU-DET**:* To evaluate the generalization performance of PBSD-Net, we conducted comparative experiments on the NEU-DET dataset [19]. The NEU-DET dataset is a steel surface defect detection dataset released by Northeastern University (NEU). This dataset collects six typical surface defects of hot-rolled steel strips (Cr, In, Pa, Pi, Rs, and Sc). As shown in Table IV , we compared our method with both classical YOLO-based detection models (e.g., YOLOv3 to YOLOv10, YOLOX) and the Transformer-based RTDETR. PBSD-Net (Ours) outperformed all other methods with the highest mAP of 79.2%, demonstrating strong overall detection capability. In particular, our method achieved the best performance in detecting Pi and Rs defects, which are characterized by complex and varied scale features. This validates the effectiveness of our multi-granularity modeling strategy. However, due to interference from background textures, the performance on Cr defects was relatively lower. Despite this, the comprehensive advantage of PBSD-Net over other models highlights its robustness and applicability in industrial defect detection tasks.

TABLE IV
DIFFERENT METHODS EXPERIMENTAL RESULTS ON NEU-DET

Methods	mAP@50(%)	Cr	In	Pa	Pi	Rs	Sc
YOLOv3	74.4	45.3	83.1	92.5	76.7	66.8	81.9
YOLOv3-SPP	75.2	46.4	85.7	92.6	82.8	70.7	72.7
YOLOv5	73.0	46.4	85.0	93.0	75.5	54.9	83.3
YOLOv7	75.4	49.0	83.4	**96.5**	79.4	65.5	78.4
YOLOv8	75.8	49.8	86.0	93.0	79.3	64.6	81.9
YOLOX	76.1	50.1	86.3	91.3	86.9	69.9	72.1
YOLOv9	72.0	50.0	81.9	93.5	74.0	62.0	70.9
YOLOv10	73.6	43.1	84.1	96.1	78.2	61.9	78.4
RTDETR	72.2	39.4	73.6	90.0	82.4	63.2	**83.6**
Ours	**79.2**	**52.1**	**86.1**	91.8	**84.2**	**77.4**	83.4

As shown in Fig.13, the performance of various classical object detection methods on the NEU-DET dataset is compared, including YOLOv3 [31], YOLOv3-SPP [50], YOLOv8-v10, YOLOv5-Anchor free, YOLOX, YOLOv9, YOLOv10, RT-DETR, and the proposed method. In the Cr category, although all methods can detect some targets, our method achieves results closer to the labeled images in terms of both the number and the positioning of the detection boxes, demonstrating superior accuracy. For the In category, while the detection results of most methods are generally similar, our method exhibits higher precision in both detection accuracy and box positioning, achieving better alignment with the labeled images. In the Pa category, the detection results of YOLOv3, YOLOv3-SPP, and YOLOv9 are similar but suffer from missed and false detections. Methods such as YOLOv8, YOLOv5-Anchor free, YOLOX, YOLOv10, and RTDETR show slight improvements. However, our method demonstrates the highest accuracy in terms of both the number and the positioning of detection boxes, achieving nearly perfect consistency with the labeled images. For the Pi category, differences in detection performance among the methods are more pronounced. While YOLOv3, YOLOv3-SPP, and YOLOv9 provide accurate detection box positions, their results are insufficient in terms of the number of detections. YOLOv8, YOLOv5-Anchor free, and YOLOv10 achieve better results but exhibit some false detections. YOLOX and RTDETR generate more complete detection results, whereas our method achieves the highest consistency with the labeled images in both the quantity and positioning of detection boxes. In the Rs category, various methods exhibit varying degrees of missed and false detections. YOLOv8, YOLOv5-Anchor free, YOLOX, YOLOv9, YOLOv10, and RTDETR perform relatively well but still have notable errors. In contrast, our method achieves results that are nearly identical to the labeled images in terms of both box quantity and positioning, demonstrating superior accuracy. Finally, in the Sc category, although the detection results across all methods are relatively close, our method exhibits higher accuracy in both the number and positioning of detection boxes, showing better consistency with the labeled images.

V. PRACTICAL APPLICATION VERIFICATION

This innovative study aims to address the practical issue of battery surface defect detection. In collaboration with a renowned battery subcontracting company, we developed a

TABLE V
THE TEST RESULTS ON DIFFERENT PLATFORMS IN PBSD-RGB

GPU	Methods	Params	FPS	P(%)	R(%)	mAP@50(%)
RTX 3060 (12G)	YOLOv8-seg	21.9M	125	99.7	99.5	99.5
	PBSD-Net	23.5M	45	38.2	27.2	23.2
	PBSD-Net + SAHI	-	11($\downarrow$34)	87.1($\uparrow$48.9)	80.3($\uparrow$53.1)	85.7($\uparrow$62.5)
	YOLOv8-seg + PBSD-Net + SAHI	-	10($\downarrow$35)	87.0($\uparrow$48.8)	80.4($\uparrow$53.2)	85.7($\uparrow$62.5)
RTX 3090 (24G)	YOLOv8-seg	21.9M	234	99.8	99.5	99.5
	PBSD-Net	23.5M	77	38.3	27.4	23.9
	PBSD-Net + SAHI	-	21($\downarrow$56)	87.1($\uparrow$48.8)	80.1($\uparrow$52.7)	85.5($\uparrow$61.3)
	YOLOv8-seg + PBSD-Net + SAHI	-	19($\downarrow$58)	80.0($\uparrow$48.9)	80.0($\uparrow$52.6)	85.5($\uparrow$61.3)

testing system, as shown in Fig.2, which consists of three components: 2D-AOI system, detection system, and expert evaluation. The 2D-AOI system includes two cameras, a rotatable platform, and a light source. The platform rotates to deliver the batteries into the inspection area, where two symmetrically placed cameras capture images. This system is capable of simultaneously imaging both sides of the battery, and the images are subsequently fed into the detection system. The detection system comprises hardware, software setups, and algorithms. The hardware setup includes a high-performance image processing server equipped with an Intel Xeon E5 CPU, Nvidia RTX 3090 GPU, and 64GB of memory. The software operates on Ubuntu 22.04, featuring Python 3.8, CUDA 11.7, and PyTorch 1.13.1. The algorithms include a positioning algorithm, YOLOv8-seg, and a defect detection algorithm, PBSD-Net. Finally, the detection results were verified through expert engineer evaluations.

Algorithm 1 Battery Defect Detection Algorithm

Input: Locator model $f_L(\cdot;\theta_L)$, Detector model $f_D(\cdot;\theta_D)$, raw battery image set D_B, input image path `input_path`, cropped image save path `cropped_path`, detection result save path `detected_path`
Output: Save cropped images and detection results
for each image I_B in D_B **do**
  $I_L \leftarrow$ LoadImage(`input_path`)
  $C \leftarrow f_L(I_L;\theta_L)$
  $I_C \leftarrow$ Crop(I_L, C)
  SaveImage(I_C, `cropped_path`)
  $I_D \leftarrow$ LoadImage(`path1`)
  $R \leftarrow f_D(I_D;\theta_D)$
  MarkDefects(I_D, R)
  SaveImage(I_D, `detected_path`)
end for

The battery positioning and defect detection algorithms were evaluated on both an experimental platform equipped with an RTX 3060 (12 GB) and a deployment platform with an RTX 3090 (24 GB). As shown in Table V, to facilitate deployment, the trained model weights were converted to ONNX format. Performance was assessed on the annotated test set using standard evaluation metrics, including frames per second (FPS), mean average precision at IoU threshold 0.5 (mAP@50), precision (P), and recall (R). First, YOLOv8-seg and PBSD-Net were tested independently. Due to the relatively simple nature of the battery localization task, YOLOv8-seg achieved high performance with P = 99.7–99.8%, R = 99.5%, and mAP@50 = 99.5%. Subsequently, PBSD-Net was deployed with the slicing aided hyper inference (SAHI) framework. Although the FPS decreased, the detection accuracy improved significantly, with mAP@50 increasing to 85.5–85.7%, P rising from 38.2% to 87.1%, and R from 27.2% to 80.3%. Finally, the localization and detection algorithms were integrated to simulate the complete deployment pipeline, as illustrated in Algorithm 1. The end-to-end system achieved an mAP@50 of approximately 85%, an FPS of around 19, and balanced precision-recall performance (P = 80.0–87.1%, R = 80.0–80.4%). Although the current system does not meet the strict requirements of real-time detection, it has been acknowledged by users for improving the efficiency of battery surface inspection.

VI. CONCLUSION

A series of automatic detection problems for surface defects on prismatic battery were addressed by proposing the SSA training strategy and SDSC module. Firstly, a dataset for surface defects on prismatic battery was proposed. Secondly, to tackle the issue of dense and tiny defects in the images, SSA was introduced as a training strategy for data. Lastly, the SDSC module combined with FM was proposed to extract the morphological features of the defects. Experimental results on the self-made PBSD-RGB dataset demonstrated that the algorithm achieved a mAP@50 of 85.8%, indicating robust feature extraction capabilities in the presence of irregular defect features and ensuring efficient detection. In practical applications, our work has effectively improved the detection accuracy of automated inspection systems. While the proposed methods yielded promising results, their applicability is restricted to non-real-time processing scenarios due to the primary concern of prolonged model training and inference time associated with SSA. Additionally, the model is susceptible to interference from background noise during detection, leading to degraded performance. Therefore, the following three aspects will be the focus of future research: 1) Exploring a novel lightweight network to improve detection efficiency and address the issue of slow model inference speed; 2) Improving the training strategy to allow the model to adaptively adjust the zooming size and sliding rate; 3) Enhancing the model's anti-interference ability against background noise information.

REFERENCES

[1] S. M. Lambert *et al.*, "Rapid nondestructive-testing technique for in-line quality control of Li-ion batteries," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 4017–4026, Dec. 2016.

[2] H. Zhang, X. Di, and Y. Zhang, "Real-time CU-net-based welding quality inspection algorithm in battery production," *IEEE Trans. Ind. Electron.*, vol. 67, no. 12, pp. 10 942–10 950, Dec. 2020.

[3] H. Yang *et al.*, "Multi-Category Decomposition Editing Network for the Accurate Visual Inspection of Texture Defects," *IEEE Trans. Autom. Sci. Eng.*, pp. 1–11, 2023.

[4] H. R. Chi *et al.*, "Guest editorial: Next-generation network automation for industrial internet-of-things in Industry 5.0," *IEEE Trans. Ind. Inf.*, vol. 19, no. 2, pp. 2062–2064, Oct. 2022.

[5] W. Zhou, Q. Guo, J. Lei, L. Yu, and J. N. Hwang, "ECFFNet: Effective and Consistent Feature Fusion Network for RGB-T Salient Object Detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1224–1235, March 2022.

[6] W. Zhou, Y. Zhu, J. Lei, J. Wan, and L. Yu, "APNet: Adversarial Learning Assistance and Perceived Importance Fusion Network for All-Day RGB-T Salient Object Detection," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 6, no. 4, pp. 957–968, Aug. 2022.

[7] J. Cui *et al.*, "Reslt: Residual learning for long-tailed recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3695–3706, May 2022.

[8] A. J. Deo, S. K. Behera, and D. P. Das, "Online monitoring of iron ore pellet size distribution using lightweight convolutional neural network," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1974–1985, 2024.

[9] H. Yao, W. Yu, and X. Wang, "A feature memory rearrangement network for visual inspection of textured surface defects toward edge intelligent manufacturing," *IEEE Trans. Autom. Sci. Eng.*, vol. 20, no. 4, pp. 2616–2635, Oct. 2023.

[10] X. Dong, C. J. Taylor, and T. F. Cootes, "Defect classification and detection using a multitask deep one-class CNN," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 3, pp. 1719–1730, Jul. 2022.

[11] W. Cui *et al.*, "Autocorrelation Aware Aggregation Network for Salient Object Detection of Strip Steel Surface Defects," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, Jun. 2023, art no. 5019412.

[12] S. Carlos and T. Du-Ming, "Accurate Vision-Based PCB Positioning Using Cosine-Convolutional Neural Networks," *IEEE Trans. Autom. Sci. Eng.*, pp. 1–10, 2023.

[13] M. Yuan *et al.*, "YOLO-HMC: An Improved Method for PCB Surface Defect Detection," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–11, 2024, art no. 2001611.

[14] X. Tao, L. Peng, Y. Tao, and C. Ye, "Inspection of defects in weld using differential array ECT probe and deep learning algorithm," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021, art no. 2512209.

[15] B. Su, H. Chen, K. Liu, and W. Liu, "RCAG-Net: Residual channelwise attention gate network for hot spot defect detection of photovoltaic farms," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–14, 2021, art no. 3510514.

[16] B. Du, Y. He, Y. He, J. Duan, and Y. Zhang, "Intelligent classification of silicon photovoltaic cell defects based on eddy current thermography and convolution neural network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6242–6251, Oct. 2020.

[17] H. Chen *et al.*, "Detection of surface defects in solar cells by bidirectional-path feature pyramid group-wise attention detector," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–9, 2022, art no. 5025609.

[18] G. Jocher, "YOLOv5-Master," Accessed: Nov. 22, 2022, [Online]. Available: https://github.com/ultralytics/yolov5.

[19] Y. He, K. Song, Q. Meng *et al.*, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1493–1504, Apr. 2020.

[20] B. Su, Z. Zhou, and H. Chen, "PVEL-AD: A Large-Scale Open-World Dataset for Photovoltaic Cell Anomaly Detection," *IEEE Trans. Ind. Inf.*, vol. 19, no. 1, pp. 404–413, Jan. 2023.

[21] Q. Xu *et al.*, "Heterogeneous Graph Transformer for Multiple Tiny Object Tracking in RGB-T Videos," *IEEE Trans. Multimedia.*, vol. 26, pp. 9383–9397, 2024.

[22] H. I. Liu *et al.*, "A DeNoising FPN With Transformer R-CNN for Tiny Object Detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, no. 4704415, pp. 1–15, 2024.

[23] F. Han and X. Lang, "A Fast Magnetic Flux Leakage Small Defect Detection Network," *IEEE Trans. Ind. Inf.*, vol. 19, no. 12, pp. 11 941–11 948, Dec. 2023.

[24] D. Xu and Y. Wu, "MRFF-YOLO: a multi-receptive fields fusion network for remote sensing target detection," *Remote Sens.*, vol. 12, no. 19, p. 3118, Sept. 2020.

[25] J. Wang *et al.*, "Adaptive Receptive Field Enhancement Network Based on Attention Mechanism for Detecting the Small Target in the Aerial Image," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, no. 5600118, pp. 1–18, 2024.

[26] W. Jiang *et al.*, "PCB defects target detection combining multi-scale and attention mechanism," *Eng. Appl. Artif. Intell.*, vol. 123, p. 106359, 2023.

[27] X. Zhou *et al.*, "Dense Attention-Guided Cascaded Network for Salient Object Detection of Strip Steel Surface Defects," *IEEE Trans. Instrum. Meas.*, vol. 71, no. 5004914, pp. 1–14, 2022.

[28] W. Xuan *et al.*, "A lightweight modified YOLOX network using coordinate attention mechanism for PCB surface defect detection," *IEEE Sensors J.*, vol. 22, no. 21, pp. 20 910–20 920, Nov. 2022.

[29] J. Yang *et al.*, "Focal modulation networks," in *Adv. Neu. In. Proc. Sys. (NIPS)*, vol. 35, 2022, pp. 4203–4217.

[30] F. C. Akyon, S. O. Altinuc, and A. Temizel, "Slicing aided hyper inference and fine-tuning for small object detection," in *2022 IEEE Int. Conf. Imag. Proc. (ICIP)*, Bordeaux, France, 2022, pp. 966–970.

[31] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, arXiv:1804.02767.

[32] Y. Zeng *et al.*, "Learning pyramid-context encoder network for high-quality image inpainting," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 1486–1494.

[33] W. Zhou, Y. Zhu, J. Lei, R. Yang, and L. Yu, "LSNet: Lightweight Spatial Boosting Network for Detecting Salient Objects in RGB-Thermal Images," *IEEE Trans. Image Process.*, vol. 32, pp. 1329–1340, 2023.

[34] W. Zhou, Y. Zhu, J. Lei, J. Wan, and L. Yu, "CCAFNet: Crossflow and Cross-Scale Adaptive Fusion Network for Detecting Salient Objects in RGB-D Images," *IEEE Trans. Multimedia.*, vol. 24, pp. 2192–2204, 2022.

[35] J. Yang *et al.*, "Focal modulation networks," in *Proc. 36th Int. Conf. NeurIPS*, ser. NIPS '22, 2024.

[36] Q. Luo *et al.*, "CDDNet: Camouflaged Defect Detection Network for Steel Surface," *IEEE Trans. Instrum. Meas.*, vol. 73, no. 5000313, pp. 1–13, 2024.

[37] Y. Chen *et al.*, "Dynamic convolution: Attention over convolution kernels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 11 030–11 039.

[38] Y. Qi *et al.*, "Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation," in *Proc. IEEE/CVF Int. Conf. Compute. Vis. (ICCV)*, 2023, pp. 6070–6079.

[39] Y. Z, H. H *et al.*, "Yolo-facev2: A scale and occlusion aware face detector," *Pattern Recognit.*, vol. 155, p. 110714, 2024.

[40] R. T. Y and D. G, "Focal loss for dense object detection," in *Proc. IEEE/CVF Int. Conf. Compute. Vis. (ICCV)*, 2017, pp. 2980–2988.

[41] L. X *et al.*, "Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 21 002–21 012, 2020.

[42] G. Yang *et al.*, "Face mask recognition system with YOLOv5 based on image recognition," in *Proc. IEEE 6th Int. Conf. Comput. Commun. (ICCC)*, Dec. 2020, pp. 1398–1404, chengdu, China.

[43] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.

[44] J. Dai *et al.*, "Deformable convolutional networks," in *Proc. IEEE/CVF Int. Conf. Compute. Vis. (ICCV)*, 2017, pp. 764–773.

[45] Ultralytics, "YOLOv10," Github, Oct. 2024, [Online]. Available: https://github.com/ultralytics/ultralytics.

[46] W. Lv *et al.*, "DETRs beat YOLOs on real-time object detection," 2023.

[47] C. Wang, W. He, Y. Nie *et al.*, "Gold-YOLO: Efficient object detector via gather-and-distribute mechanism," in *Adv. Neu. In. Proc. Sys. (NIPS)*, 2024, p. 36.

[48] J. Li, Y. Wen, and L. He, "Scconv: spatial and channel reconstruction convolution for feature redundancy," in *Proc. IEEE/CVF Int. Conf. Compute. Vis. (ICCV)*, 2023, pp. 6153–6162.

[49] Z. Ge *et al.*, "YOLOX: Exceeding YOLO series in 2021," 2021, arXiv:2107.08430.

[50] K. He, X. Zhang, S. Ren *et al.*, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sept. 2015.

Ying Xu received the BS. and MS. degrees in automation, and control theory and control engineering from the Wuhan University of Science and Technology, Wuhan, China, in 2004 and 2008, respectively, and the Ph.D. degree in control theory and control engineering from the South China University of Science and Engineering, Guangzhou, China, in 2013.

She is currently an Associate Professor with the School of information and communication engineering, Wuyi University, Jiangmen, Guangdong, China. Her current research interests include image processing, deep learning, and biometric extraction.

Bo Li received the B.S. degree in communication engineering from Wuyi University, Jiangmen, China, in 2020, where he is currently working toward the master's in on industrial visual detection degree with the School of information and communication engineering.

His research interests include computer vision, representational contrastive learning, image processing, and object detection.

Yikui Zhai (Senior Member, IEEE) received the bachelor's degree in optical electronics information and communication engineering and the master's degree in signal and information processing from Shantou University, Shantou, China, in 2004 and 2007, respectively, and the Ph.D. degree in signal and information processing from Beihang University, Beijing, China, in 2013.

Since 2007, he has been with the School of Electronics and Information Engineering, Wuyi University, Jiangmen, China, where he is currently a Full Professor and Associate Dean. He has been a Visiting Scholar with the Department of Computer Science, University of Milan, Milan, Italy, from 2016 to 2017, and from 2023 to 2024. His research interests include image processing, deep learning, and pattern recognition.

Feng Ke (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronic engineering from the South China University of Technology (SCUT), Guangzhou, China, in 1998, 2001, and 2004, respectively.

From 2013 to 2014, he was a Visiting Scholar with the Department of Electronic Engineering, Missouri University of Science and Technology, Rolla, MO, USA. He is currently with the School of Electronic and Information Engineering, SCUT, and also with the Pazhou Laboratory, Guangzhou. He is the Vice Director with the Intelligent Communication Network and Computer Engineering Technology Engineering Center, SCUT, and the Chief Scientist with Guangdong Chinese Network Electrical Engineering Research Institute, Guangzhou, and also an Expert Advisor with Guangzhou Science and Technology Bureau, Guangzhou, and Guangdong Science and Technology Department, Guangzhou. His research interests include deep learning, resource allocation, signal processing.

Jianhong Zhou received the B.S. degree in communication engineering in 2020 from Wuyi University, Jiangmen, China, where he is currently working toward the master's in on industrial visual detection and drone vehicle detection degree with the School of Electronics and Information Engineering.

His research interests include computer vision, representational contrastive learning, image processing, and object detection.

Pasquale Coscia (Senior Member, IEEE) received the Ph.D. degree in Industrial and Information Engineering from the Universitá degli Studi della Campania "Luigi Vanvitelli", Italy, in 2019.

From 2019 to 2022, he was a post-doctoral researcher at the Universitá degli Studi di Padova, Italy. He is currently an Assistant Professor at the Department of Computer Science of the Universitá degli Studi di Milano, Italy. He is Co-chair of the Intelligent Measurement Systems Technical Committee (TC-22) of the IEEE Instrumentation and Measurement Society since 2023. His research activities are focused on theoretical, methodological, and applied aspects of computational intelligence for signal and image processing.

Angelo Genovese (Senior Member, IEEE) received the Ph.D. degree in computer science from the Università degli Studi di Milano, Milan, Italy, in 2014. He has been a Post-Doctoral Research Fellow incomputer science with the Università degli Studidi Milano, since 2014. He has been a Visiting Researcher with the University of Toronto, Toronto, ON, Canada. Original results have been publishedin over 30 articles in international journals, proceedings of international conferences, books, and book chapters. His current research interests include signal and image processing, three-dimensional reconstruction, computational intelligence technologies for biometric systems, industrial and environmental monitoring systems, and design methodologies and algorithms for self-adapting systems.

Dr.Genovese is an Associate Editor of the *Journal of Ambient Intelligence-and Humanized Computing* (Springer).

Vincenzo Piuri (Fellow, IEEE) received the M.S. and Ph.D. degrees in computer engineering from Politecnico di Milan, Milan, Italy, in 1984 and 1988, respectively.

He was the Department Chair with the University of Milan, Milan, from 2007 to 2012, where he has been a Full Professor since 2000. He was an Associate Professor with Politecnico di Milan from 1992 to 2000, a Visiting Professor with The University of Texas at Austin, Austin, TX, USA from 1996 to 1999, and a Visiting Researcher with George Mason University, Fairfax, VA, USA, from 2012 to 2016. He founded a startup company, Sensure srl, Bergamo, Italy, in the area of intelligent systems for industrial applications (leading it from 2007 to 2010) and was active in industrial research projects with several companies. His main research and industrial application interests are intelligent systems, computational intelligence, pattern analysis and recognition, machine learning, signal and image processing, biometrics, intelligent measurement systems, industrial applications, distributed processing systems, Internet-of-Things, cloud computing, fault tolerance, application-specific digital processing architectures, and arithmetic architectures.

Dr. Piuri is an ACM Fellow.

Fabio Scotti (Senior Member, IEEE) received the Ph.D. degree in computer engineering from the Politecnico di Milan, Milan, Italy, in 2003.

He has been an Associate Professor of computer science with the Universitá degli Studi di Milan, Crema, Italy, since 2015. His original results have been published in more than 100 articles in international journals, proceedings of international conferences, books, book chapters, and patents. His research interests include biometric systems, machine learning and computational intelligence, signal and image processing, theory and applications of neural networks, 3-D reconstruction, industrial applications, intelligent measurement systems, and high-level system design.

Dr. Scotti is an Associate Editor with the IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS and *Soft Computing* (Springer). He has been an Associate Editor with the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, and a Guest Coeditor for the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT.