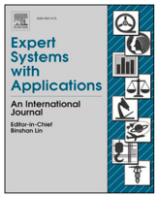
Expert Systems With Applications 289 (2025) 128305

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



Progressive contrastive representation learning for defect diagnosis in aluminum disk substrates with a bio-inspired vision sensor

Peng Chen [a,b], Ruijin Zhang [a], Changbo He [c,*], Yaqiang Jin [d], Shuai Fan [e], Junyu Qi [f], Chengning Zhou [g], Chun Zhang [a]

[a] College of Engineering, Shantou University, Shantou, 515063, Guangdong, China
[b] Key Laboratory of Intelligent Manufacturing Technology, Ministry of Education, Shantou, 515063, Guangdong, China
[c] College of Electrical Engineering and Automation, Anhui University, Hefei, 230601, China
[d] School of Qilu Transportation, Shandong University, Jinan, 250061, Shandong, China
[e] School of Mechanical and Electrical Engineering, Chengdu University of Technology, Chengdu, 610059, Sichuan, China
[f] Electronics & Drives, Reutlingen University, Reutlingen, 72762, Germany
[g] Nuclear Power Institute of China, Chengdu, 610213, Sichuan, China

ARTICLE INFO

Keywords:
Aluminum disk substrates
Fault diagnosis
Bio-inspired vision sensor
Contrastive learning
Semi-supervised learning
Event stream imaging
Multi-stage loss function

ABSTRACT

Traditional industrial surface defect detection using CCD/CMOS cameras faces limitations in detecting minute defects on aluminum substrates in dynamic industrial scenes. While event cameras can capture small defects through event information flow (EIF), they struggle with noise-related challenges. Additionally, current contrastive learning frameworks employ simplified loss computation mechanisms that limit their capability to discover and characterize novel categories. This study addresses these challenges through two main innovations: (1) a novel event stream imaging technique that combines equal time window segmentation, overlapping windows, and Gaussian filtering to enhance data quality, and (2) Progressive Contrastive Representation Learning (PCRL), a sophisticated framework for handling both known and unknown fault classes. The PCRL framework overcomes traditional limitations through a four-stage loss function strategy and structured analysis of intra-cluster and inter-cluster relationships, enabling better feature extraction and model convergence for previously unseen patterns. Through systematic validation, our approach demonstrates superior performance in noise reduction, unknown fault detection, and precise surface defect classification, offering a robust solution for industrial applications. Through systematic validation, our methodology exhibits exceptional efficacy in mitigating signal noise, identifying novel fault patterns, and executing high-precision surface anomaly categorization, yielding a comprehensive accuracy rate of 93.51 % across all test cases and demonstrating remarkable performance with 95.36 % accuracy in previously unencountered fault categories, thereby presenting a sophisticated solution for industrial implementations while establishing a foundation for subsequent scholarly investigations in this field.

1. Introduction

Surface defect detection of industrial products has emerged as a fundamental and indispensable component of intelligent manufacturing systems, particularly as quality control demands become increasingly stringent (Chen et al., 2025a, 2023; Mezher & Marble, 2024; Ozdemir & Koc, 2024). Although the current fault diagnosis and detection technology predominantly relies on CCD/CMOS sensors, which have undergone extensive development and refinement over decades and consequently achieved remarkable results across diverse applications such as semiconductor wafer inspection (Cheng et al., 2023), automotive paint surface quality control (He et al., 2024), and textile fabric defect detection (Kahraman & Durmuşoğlu, 2023), these conventional sensors nevertheless face inherent limitations. Specifically, due to the discrete sensing methodology inherent to CCD/CMOS sensors, they encounter significant challenges in detecting minute and subtle defects, especially in highly dynamic industrial environments where rapid movement and varying lighting conditions prevail.

In response to these technological constraints, and in pursuit of more effective visual perception capabilities in high-dynamic industrial

* Corresponding author.
E-mail addresses: pengchen@alu.uestc.edu.cn, dr.pengchen@foxmail.com (P. Chen), 842962781@qq.com (R. Zhang), changbh@ahu.edu.cn (C. He), yaqiang.jin@outlook.com (Y. Jin), fanshuai@cdut.edu.cn (S. Fan), junyu.qi@reutlingen-university.de (J. Qi), chengningzhou@foxmail.com (C. Zhou), chunzhang.st@foxmail.com (C. Zhang).

https://doi.org/10.1016/j.eswa.2025.128305
Received 22 November 2024; Received in revised form 7 May 2025; Accepted 22 May 2025
Available online 25 May 2025

scenarios, researchers (Snyder et al., 2023; Zhong et al., 2024) have begun exploring alternative approaches by introducing biologically-inspired event cameras into the domain of industrial quality inspection. The operational principle of these event cameras represents a paradigm shift in imaging technology, whereby individual pixels autonomously generate events whenever they detect intensity changes exceeding a predetermined threshold, operating independently of adjacent pixels. This fundamental difference distinguishes them from traditional CCD/CMOS cameras, which capture sequential static frames at fixed intervals. Instead, event cameras generate an asynchronous Event Information Flow (EIF), which encompasses comprehensive environmental data with unprecedented temporal precision. Consequently, these innovative sensors achieve superior temporal resolution and significantly enhanced dynamic range for environmental perception, making them particularly well-suited for visual applications in challenging, high-dynamic scenarios. Given that EIF represents a novel data format with unique characteristics and processing requirements, numerous researchers have initiated comprehensive investigations into Event Information Flow Processing (EIFP) methodologies, seeking to fully harness and optimize the exceptional capabilities of event cameras for industrial applications.

In the rapidly evolving field of surface defect detection, event cameras have emerged as a transformative and prominent tool that is increasingly being employed by researchers across diverse applications, particularly due to their unique capabilities in handling high-speed dynamics and challenging lighting conditions (Chen et al., 2025b,c; Gamage et al., 2023; Schaefer et al., 2022; Wang et al., 2023; Xin et al., 2025). Contemporary research has addressed several critical challenges in this domain. Notably, to overcome the limitations of existing event camera feature tracking methods, which have traditionally relied on either handcrafted approaches or methods requiring extensive parameter tuning while lacking generalization capabilities, Messikommer et al. (2023) have introduced a groundbreaking data-driven feature tracker. This innovative system not only combines low-latency event processing with a novel frame attention module but also uniquely leverages events to track features detected in grayscale frames. Furthermore, through the implementation of a novel self-supervision strategy, their approach has demonstrated remarkable performance improvements, achieving up to 130 % enhancement in relative feature age while maintaining minimal latency, even when transferring zero-shot from synthetic to real data. In parallel developments, when addressing the multifaceted challenges inherent in MAV-based civil infrastructure inspection under variable lighting conditions, Gamage et al. (2023) have developed an innovative approach that synergistically combines event cameras with Spiking Neural Networks (SNNs). This integrated system has not only demonstrated substantially superior classification accuracy under dynamic lighting conditions compared to traditional image-based methods but has also achieved remarkable energy efficiency improvements of 65–135 times when deployed on neuromorphic hardware versus conventional Artificial Neural Networks (ANNs). Additionally, in the domain of stereo vision processing, Garg (2023) has introduced a sophisticated cooperative stereo method utilizing dual fixed Dynamic Vision Sensors (DVS) cameras, wherein their approach has achieved more than 50 % reduction in observed average error compared to existing state-of-the-art methods, with comprehensive validation conducted across multiple public stereo event datasets. To address the inherent limitations of existing event camera object detectors in managing diverse object velocities while maintaining high temporal resolution, Liu et al. (2023) have developed a comprehensive motion-robust and high-speed detection pipeline. This sophisticated system incorporates both a Temporal Active Focus (TAF) representation and a Bifurcated Folding Module (BFM). In addition, to optimize detection accuracy while preserving computational efficiency, they have introduced a lightweight Agile Event Detector (AED) alongside an innovative data augmentation method, consequently demonstrating competitive performance across multiple metrics including accuracy, efficiency, and parameter count in real-scene datasets.

Of particular significance is the observation that the asynchronous event information stream captured by event cameras bears substantial similarity to multivariate time series in terms of storage architecture. Building upon this insight, researchers such as Schaefer et al. (2022) and Xie et al. (2022) have successfully implemented methods based on Graph Convolutional Neural Networks (GCN) to process the asynchronous Event Information Flow (EIF). Moreover, considering the biomimetic nature of event cameras, numerous scholars have leveraged SNNs to enhance the biological realism of artificial neural networks and facilitate direct processing of asynchronous event information streams. Building upon these fundamental concepts, Wang et al. (2023) have systematically adapted and extended the object detection algorithm framework originally designed for CCD/CMOS cameras. In addressing the performance limitations of SNNs in object detection tasks, Kim et al. (2020) have introduced two significant methodological advances: channel-wise normalization and signed neuron with imbalanced threshold. These revolutionary developments have culminated in the creation of Spiking-YOLO, which represents the first spike-based object detection model of its kind. Most notably, this pioneering system achieves comparable performance (98 %) to Tiny YOLO while demonstrating exceptional improvements in both energy efficiency (280 times less energy consumption) and convergence speed (2.3 to 4 times faster) compared to previous SNN conversion methods. The sampling principle of event cameras, which is fundamentally based on temporal brightness thresholding, inherently introduces significant noise artifacts into the captured data stream. Furthermore, environmental and operational factors, such as subtle mechanical vibrations, momentary interruptions in camera functionality, and fluctuations in ambient illumination conditions, can substantially degrade the quality and reliability of the event stream signal. These technical constraints pose fundamental challenges in achieving consistent and accurate event-based sensing.

Nevertheless, despite considerable technological advances in event-based vision systems, several critical challenges continue to impede the widespread practical deployment of event-based detection frameworks. While the implementation of semi-supervised models has emerged as a promising approach to address the challenge of novel class discovery in surface defect detection scenarios, these methodologies are constrained by significant limitations. Most notably, current approaches exhibit strong dependencies on data quality, wherein model performance is inextricably linked to the fidelity and representativeness of the training dataset. Moreover, existing contrastive learning frameworks typically employ relatively simplistic loss computation mechanisms, which consequently restrict their capacity to effectively discover and characterize novel categories within the data distribution. The inherent complexity of analyzing multidimensional relationships both within and between data clusters introduces additional computational and interpretative challenges, particularly in the context of clustering outcome analysis and validation. Therefore, these methodological limitations can be systematically categorized and summarized as detailed below.

1. Event Stream Quality Dependencies: The inherent characteristics of event cameras, particularly their temporal brightness thresholding mechanism, introduce significant noise artifacts into the data stream. Environmental factors such as mechanical vibrations, camera interruptions, and ambient lighting variations further compromise data quality. This fundamental limitation in data acquisition directly impacts the downstream processing and analysis capabilities of event-based detection systems.
2. Representation Learning Constraints: Current contrastive learning frameworks employ simplified loss computation mechanisms that limit their capability to discover and characterize novel categories. The absence of sophisticated multi-step analytical processes impairs the model's ability to build robust representations of previously unseen patterns.
3. Structural Analysis Challenges: The complexity of analyzing relationships within and between data clusters poses significant

difficulties. The high sensitivity of clustering outcomes to methodological assumptions often leads to interpretation uncertainties, affecting both standard clustering and over-clustering scenarios.
4. Emergent Pattern Recognition Limitations: Despite their proficiency in identifying known fault patterns, existing systems demonstrate significant limitations in detecting novel fault types. This shortcoming becomes particularly critical in dynamic industrial environments where new fault patterns continuously emerge and evolve.

To address the aforementioned challenges in surface defect detection systems, this study presents a comprehensive two-fold approach. Firstly, it introduces an innovative event stream imaging technique that effectively denoises and optimizes image data transformed from event stream data. Specifically, the proposed method integrates three key components: equal time window segmentation, overlapping multiple time windows, and Gaussian filtering, which work synergistically to enhance data quality and reduce noise artifacts. Subsequently, the study introduces Progressive Contrastive Representation Learning (PCRL), a sophisticated methodological framework that simultaneously addresses two critical challenges: the pervasive issue of unknown fault classes and the efficient utilization of limited known fault class data.

Within this proposed PCRL framework, we conduct a rigorous and systematic investigation into the complex dynamics of both intra-cluster and inter-cluster relationships, with particular emphasis on their manifestation within over-clustered and standard-clustered fault classes. Through this meticulous exploration, the feature extraction capabilities at each successive stage of the model undergo iterative refinement, thereby not only accelerating the model's convergence trajectory but also substantially enhancing its capacity to generalize effectively to previously unencountered data patterns. Furthermore, this methodical process ensures that the model's learning trajectory maintains optimal alignment with its predetermined objectives, while simultaneously adapting to emerging patterns in the data stream. Consequently, the principle contribution can be summarized as follows.

1. Novel Event Stream Imaging Technique: This study introduces an innovative approach to event stream data processing that effectively denoises and optimizes image data transformation. The proposed method seamlessly integrates three key components: equal time window segmentation, overlapping multiple time windows, and Gaussian filtering, working synergistically to enhance data quality and reduce noise artifacts, thereby establishing a robust foundation for subsequent fault detection processes.
2. Development of Progressive Contrastive Representation Learning (PCRL): This investigation presents a sophisticated approach that specifically addresses the fundamental challenge of unidentified fault classes. By strategically utilizing a limited yet carefully curated selection of known fault classes, this methodology effectively bridges the gap in unknown fault detection, thereby advancing the field's capabilities in fault identification and classification.
3. Comprehensive Framework for Analyzing Cluster Dynamics: The study introduces a structured analytical framework that meticulously examines the intricate dynamics of intra-cluster and inter-cluster relationships, while simultaneously considering both over-clustered and standard-clustered fault classifications. This detailed exploration not only enhances our understanding of complex fault relationships but also serves as a foundation for optimizing model performance through improved feature representation.
4. Implementation of Multi-Step Loss Function Strategies in Contrastive Learning: A sophisticated multi-step framework for loss functions, extending from pre-training Stage 0 through Stage 3, has been carefully designed and implemented. This hierarchical strategy effectively addresses the inherent limitations of conventional direct methods, which often struggle to capture the nuanced complexities of novel or previously unrecognized category instances. Through the implementation of this tiered loss function approach, the model achieves enhanced adaptability and demonstrates superior performance in recognizing diverse data distributions.

This paper is systematically organized into four interconnected sections that collectively provide a comprehensive understanding of the research. Initially, Section 2 examines the relevant theoretical frameworks, thereby establishing both the conceptual foundations and contextual backdrop for this investigation. Subsequently, Section 3 introduces and elaborates on the proposed Progressive Contrastive Representation Learning (PCRL), providing an in-depth examination of its architecture and mechanisms. Building upon this, Section 4 presents and analyzes the experimental results obtained through model evaluation through case study, thus enabling a critical assessment of its performance. Finally, Section 5 synthesizes the principal findings while discussing their implications and suggesting future research directions.

2. Related works

In this section, we establish the theoretical foundations that underpin our research methodology and contextualize it within the existing literature. The discourse begins by elucidating the fundamental principles of data measurement from event cameras, which are essential for understanding the temporal dynamics of visual information processing. Subsequently, we thoroughly examine the theoretical basis of semi-supervised learning, whereby we explore both its conventional paradigms and emerging frameworks. This examination naturally extends to encompass novel class discovery methods and their associated open-world assumption, which represents a significant departure from traditional closed-world classification scenarios. Furthermore, we conduct a comprehensive analysis of the latest developments in semi-supervised learning, with particular emphasis on its applications in dynamic visual processing and its integration with event-based sensing technologies.

2.1. Fundamental principles of event-based visual data acquisition

Event cameras, which represent a paradigm shift from conventional frame-based sensors, operate based on asynchronous detection of luminance differences (Yang et al., 2023). These sophisticated devices capture events through two fundamental and interconnected processes: (1) Light sensor conversion: The photosensitive sensor array at each pixel location continuously converts incoming photons into electrical signals, thereby providing instantaneous feedback regarding the current light intensity. This process occurs independently at each pixel, enabling parallel processing of visual information. (2) Difference circuit: Subsequently, a dedicated differential circuit computes the temporal derivative of light intensity between the current and previous measurements. When this calculated difference surpasses a predetermined threshold, an event is triggered and logged with precise temporal resolution. This process can be formally expressed through the following mathematical relationship.

$$\Delta L\left(\mathbf{x}_k, t_k\right) = L\left(\mathbf{x}_k, t_k\right) - L\left(\mathbf{x}_k, t_k - \Delta t_k\right) \tag{1}$$

where $\mathbf{x}_k$ denotes the spatial coordinates of the pixel, t_k represents the current temporal instance, and Δt_k is the temporal sampling interval between consecutive measurements. When the computed luminance change, ΔL, exceeds the empirically determined threshold, an event is registered in the camera's output stream. Although various event camera architectures may incorporate different supplementary information in their output, they invariably record the spatial coordinates of detected events, thus maintaining the fundamental capability of spatial localization.

As shown in Fig. 1, the working process of the event camera is illustrated by an example of a black dot on a disk rotating at a constant speed. Due to the large difference between the gray value of the black dot and the surrounding area, when the black dot moves in a circle, the gray value on both sides of the contour will change significantly during

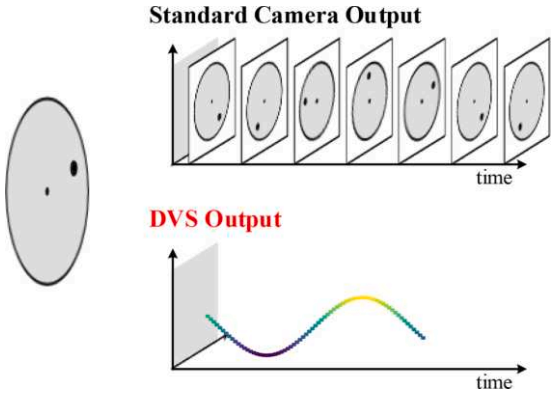


Fig. 1. Schematic representation of the working principle of an event camera.

the rotation process. The change can be captured by the event camera to record the time, coordinates and other information of the event. From the perspective of time, the event stream collected by the event camera appears as a spiral.

Event stream imaging, a fundamental cornerstone within the broader domain of neuromorphic vision, emerged from the pioneering development of Dynamic Vision Sensors (DVS), whereby initial research endeavors predominantly focused on the intricate synthesis of event streams with conventional intensity images. This innovative approach was notably exemplified by the Development of Asynchronous Vision Sensors (DAVIS), which revolutionized the capture of rapid brightness variations in dynamic scenes through asynchronous event streams, while simultaneously achieving both High Dynamic Range (HDR) imaging and low-latency performance through the seamless integration of sparse event data with traditional grayscale information. Despite these remarkable achievements, existing methodologies continue to encounter substantial challenges in effectively suppressing noise artifacts and adequately resolving motion blur during high-speed imaging scenarios, which has consequently catalyzed exploration into advanced multimodal data fusion techniques for enhanced system robustness. Recent breakthrough studies have made significant contributions, including Zou et al. (2025) who developed sophisticated keyframe-guided recurrent convolutional networks for HDR video reconstruction and validated their approach on the EventHDR dataset, while Wan et al. (2022) implemented event-driven optical flow estimation techniques that explicitly model temporal correlations between event clusters and scene motion. These developments underscore an increasingly pronounced trend toward the simultaneous exploitation of both spatial and temporal event features, which promises to overcome traditional technological bottlenecks while advancing our theoretical understanding of neuromorphic vision systems, thereby potentially influencing future research directions in related fields such as computer vision and artificial intelligence.

2.2. Novel class discovery: theoretical framework and implementation paradigm

Novel Class Discovery (NCD) (Chen et al., 2024c; Li et al., 2023a; Liu et al., 2023) represents a sophisticated machine learning paradigm that transcends traditional supervised learning boundaries by enabling the identification of previously unknown classes within unlabeled datasets while leveraging limited labeled data as an auxiliary learning mechanism. Unlike conventional supervised learning frameworks, NCD operates under the fundamental premise that the class distribution in the training data is inherently distinct from that of the test data. This paradigm shift necessitates a dual capability: maintaining proficiency in classifying known categories while simultaneously developing the ability to recognize and categorize emergent classes from unlabeled data.

The NCD methodology typically encompasses a two-phase implementation strategy: (1) Representation Learning Phase: Initially, the model utilizes labeled data to develop robust feature representations, establishing a foundational understanding of discriminative characteristics that can generalize beyond known classes. (2) Discovery Phase: Subsequently, these learned representations facilitate the identification and categorization of novel classes within unlabeled data through sophisticated clustering or classification mechanisms. In NCD, we will usually have a total loss function L that combines the supervised loss of the known class with the unsupervised loss of the new class:

$$\mathcal{L} = \lambda \mathcal{L}_{\text{sup}} + (1-\lambda)\mathcal{L}_{\text{unsup}} \tag{2}$$

where $\mathcal{L}_{\text{sup}}$ is the supervised loss on the known class data. $\mathcal{L}_{\text{unsup}}$ is an unsupervised loss on the new class of data. λ is a hyperparameter that controls the balance between the two. In addition, contrastive learning is widely used in NCD to improve the discrimination ability of the model on new classes of data, so the unsupervised loss based on contrastive learning is usually introduced:

$$\mathcal{L}_{contrastive} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp\left(\text{sim}\left(z_i, z_j\right)/\tau\right)}{\sum_{k=1}^{2N} \mathbb{I}_{[k\neq i]} \exp\left(\text{sim}\left(z_i, z_k\right)/\tau\right)} \tag{3}$$

where z_i and z_j denote the embedding of positive sample pairs. $\text{sim}(z_i, z_j)$ is a similarity measure for embeddings (e.g., cosine similarity). τ is the temperature hyperparameter used to adjust the sensitivity of the similarity. The function $\mathbb{I}_{[k\neq i]}$ is an indicator function whose range is $[0, 1]$, taking the value 1 if and only if $k \neq i$.

This methodological approach proves particularly valuable in addressing open-world scenarios, where class distributions are dynamic and data sources are heterogeneous. The framework's ability to adapt to emerging categories makes it especially relevant for contemporary machine learning applications that must operate in evolving, real-world environments.

2.3. Contrastive and semi-supervised learning: advanced frameworks for limited-label scenarios

Contrastive Learning (Chen et al., 2024a, 2025b; Li et al., 2023b; Sampath et al., 2023) has emerged as a fundamental and increasingly prominent unsupervised learning paradigm in contemporary machine learning research, wherein the primary objective is to systematically maximize the similarity between positive pairs of samples while concurrently minimizing it for negative samples. Throughout the iterative training process, contrastive learning algorithms methodically develop robust feature representations by optimizing this dual objective function. A particularly significant advantage of this sophisticated approach lies in its inherent capacity to train models without necessitating labeled data, thereby rendering it exceptionally valuable in domains where label acquisition presents substantial technical, temporal, or financial challenges. Consequently, contrastive learning has garnered widespread adoption across diverse computational tasks, encompassing, but not limited to, image classification, object detection, image generation, and text representation learning. Classical contrastive learning approaches, exemplified by SimCLR (Chen et al., 2020), implement a sophisticated and systematically structured process to obtain positive and negative pairs. Specifically, these methods employ carefully designed random data augmentation techniques on individual samples twice, thereby generating two distinct yet semantically consistent augmented views of the same data point. These comprehensive augmentation operations encompass a diverse array of transformations, including, but not limited to, random cropping, color adjustments, Gaussian blurring, and geometric flipping. Furthermore, since both augmented views inherently originate from the same source sample, SimCLR naturally considers them as a positive pair, whereas augmented views derived from different samples are systematically designated as negative pairs, irrespective of their underlying class membership.

In parallel developments, Semi-supervised Learning (Manivannan, 2023; Park & Kim, 2025) represents a sophisticated hybrid approach that

strategically combines a limited quantity of labeled data with an extensive collection of unlabeled data during the training phase. When compared to traditional fully supervised learning paradigms, this methodology substantially reduces the critical dependence on large-scale labeled datasets while effectively leveraging the rich contextual information embedded within unlabeled data. Consequently, this innovative approach has demonstrated particular efficacy in domains where labeled data acquisition presents significant challenges, such as medical image analysis, natural language processing, and advanced computer vision applications. The systematic enhancement of semi-supervised learning models occurs through two fundamental and complementary mechanisms: primarily, through the sophisticated exploitation of unlabeled data structures, whereby models leverage the inherent statistical distribution of unlabeled data to discern underlying patterns and relationships. For instance, advanced clustering algorithms and contrastive sample generation techniques facilitate a more comprehensive understanding of data distribution, which subsequently enhances generalization capabilities to labeled data. Secondarily, through pseudo-labeling mechanisms, where models generate provisional labels for unlabeled data and systematically incorporate these into the supervised learning process, thereby iteratively improving performance through self-refinement, despite the inherent limitations in labeled data availability.

In recent years, several notable and innovative semi-supervised learning architectures have been developed, including DTC (Han et al., 2019), IIC (Li et al., 2023a), ORCA (Cao, 2024; Xiao et al., 2024), ASS-Bert (Sun et al., 2023), and OpenNCD (Liu et al., 2023), each contributing unique methodological advances to the field. Specifically, DTC employs sophisticated transfer learning principles by initially training a robust base model on labeled data before systematically adapting it to recognize novel class samples. In contrast, IIC utilizes symmetric KL divergence to quantify both inter-class and intra-class similarities, methodically incorporating these measurements into the loss function optimization process. ORCA introduces several innovative architectural components, including adaptive supervised learning mechanisms, pairwise objectives, and sophisticated model regularization techniques. ASS-Bert innovatively combines active learning with semi-supervised BERT, selecting uncertain code samples for manual labeling while pseudo-labeling high-confidence unlabeled data to address data scarcity in smart contract vulnerability detection. Furthermore, OpenNCD extends traditional contrastive learning principles by utilizing an expanded set of prototypes beyond the conventional number of classes, implementing progressive clustering during training, and establishing both prototype-level and prototype group-level similarity metrics.

3. The proposed methodology of progressive contrastive representation learning (PCRL)

In this section, we present a comprehensive examination of the foundational theoretical framework that underpins our proposed Progressive Contrastive Representation Learning (PCRL) methodology. The fundamental principles of PCRL, which will be meticulously analyzed in Section 3.1, constitute a tripartite approach comprised of three distinct yet inherently interconnected stages: Initially, we implement a neighbor-based positive pair selection mechanism, whereby each sample is systematically paired with its nearest neighbor based on feature similarity metrics. Subsequently, this process evolves into an overcluster-based selection paradigm, during which positive pairs are stochastically selected from within the same overclustered partition, thereby enabling broader feature space exploration. Finally, the methodology culminates in a standard cluster-based selection phase, wherein samples residing within identical standard clusters are randomly paired to reinforce learned representations. Moreover, the technical intricacies of the pre-training phase, the systematic sample pairing procedures, and their associated methodological components will be extensively elaborated upon in subsequent sections, specifically spanning from Section 3.2 to Section 3.4, thus providing a thorough understanding of the entire framework's operational mechanics.

3.1. Progressive contrastive representation learning (PCRL)

This subsection presents a comprehensive elucidation of the proposed Progressive Contrastive Representation Learning (PCRL) methodology, as illustrated in Fig. 2, which has been specifically developed to identify and categorize various forms of defects. The methodology incorporates an innovative two-layer contrastive learning architecture that effectively leverages the limited quantity of known defect classes in samples. Through this approach, the model demonstrates remarkable capability in identifying both previously documented and novel defect classes in real-world. Consequently, this methodology addresses a significant challenge in defect diagnosis: the identification of unknown class defects in open-world of aluminum disk substrates data. The contrastive learning technique facilitates the construction of a sophisticated

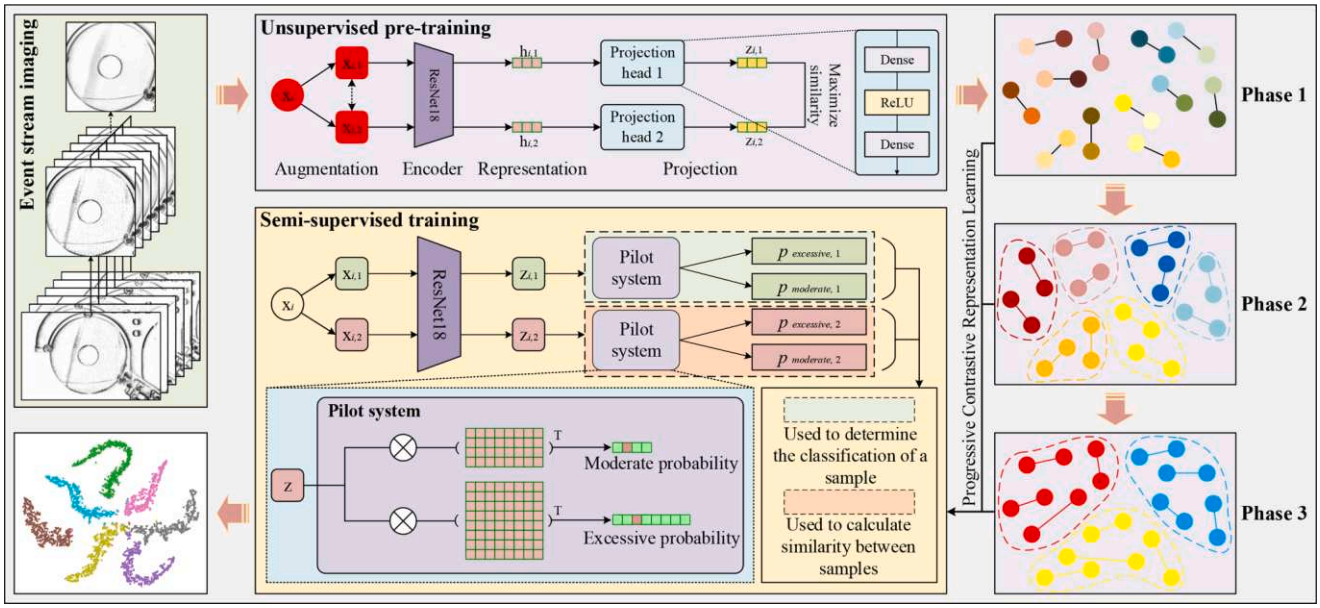


Fig. 2. The proposed architecture of progressive contrastive representation learning.

feature space wherein similar defect classes are systematically clustered while maintaining appropriate separation between dissimilar categories, thereby enabling the model to extrapolate its learned representations to previously unencountered defect classes.

The model implements a sophisticated progressive learning paradigm wherein the intricacy of feature extraction systematically evolves in tandem with the model's developing capabilities. Although conventional approaches often rely on intensifying random augmentation to amplify the distinction between positive sample pairs, this methodology encounters fundamental limitations, as excessive augmentation can potentially obscure or deteriorate essential sample characteristics, thereby compromising the learning process. Consequently, as demonstrated in Fig. 2, this section introduces a meticulously designed contrastive learning framework that not only promotes sample diversity but also ensures smooth progression through four interconnected stages: (1) Pre-training-based feature initialization, through which the SimCLR model establishes the fundamental feature extraction framework via unsupervised learning; (2) Initial neighbor-based positive pair selection, whereby each sample systematically identifies its nearest neighbor as a positive pair; (3) Overcluster-based selection, wherein samples residing within the same overclustered partition are stochastically designated as positive pairs; and (4) Standard cluster-based selection, through which samples within identical standard clusters undergo random pairing. It should be noted that the projection head consists of a two-layer MLP during pre-training: a first fully connected layer followed by ReLU activation, and a second layer mapping to a lower-dimensional embedding space. It transforms encoder outputs into normalized contrastive representations during pretraining but is discarded in downstream tasks, retaining only the encoder.

This methodologically rigorous progression ensures that the discriminative disparity between positive pairs incrementally increases throughout the learning trajectory, thereby enabling the model to synthesize and interpret sample characteristics from progressively broader perspectives. The implementation of this graduated approach is fundamentally crucial, particularly since attempting to directly employ Phase 3-level positive sample selection at the initialization stage would prove ineffective, primarily due to the inherent random initialization characteristics of deep learning architectures and the initial stochastic distribution of samples in high-dimensional feature space. Instead, the progressive strategy facilitates a structured learning pathway where each stage systematically builds upon and reinforces the learning outcomes of its predecessor, thus ensuring that samples exhibiting small cosine distances in high-dimensional space demonstrate genuine and meaningful similarity.

The architectural framework comprises an encoder E, specifically implementing ResNet-18 architecture, coupled with a sophisticated pilot system that incorporates two trainable tensor structures: $C_{excessive}$ and $C_{moderate}$. The $C_{excessive}$ tensor, characterized by dimensions $(n * k, dim)$, facilitates over-classification operations, where n represents the initial class count, k denotes the over-clustering factor, and dim corresponds to the encoder's output feature dimensionality. Within this structure, each row of $C_{excessive}$ functions as a prototype, essentially encoding a characteristic feature vector for a specific class designation. The classification process encompasses feature extraction followed by probability assignment through cosine similarity computations, mathematically expressed as:

$$p_{excessive} = \frac{\exp\left(\frac{1}{\tau}\mathbf{z} \cdot {C_{excessive}}^{\top}\right)}{\sum \exp\left(\frac{1}{\tau}\mathbf{z} \cdot {C_{excessive}}^{\top}\right)} \tag{4}$$

In parallel, $C_{moderate}$, configured with dimensions (n, dim), executes standard classification operations adhering to similar principles while generating an n-dimensional probability distribution vector:

$$p_{moderate} = \frac{\exp\left(\frac{1}{\tau}\mathbf{z} \cdot {C_{moderate}}^{\top}\right)}{\sum \exp\left(\frac{1}{\tau}\mathbf{z} \cdot {C_{moderate}}^{\top}\right)} \tag{5}$$

The comprehensive learning process methodically progresses through these three distinct stages, with each phase representing increasingly sophisticated levels of contrastive learning, while simultaneously incorporating progressively complex optimization strategies designed to enhance the model's discriminative capabilities across diverse fault classes. This structured approach ensures robust feature learning and classification performance across varying levels of data complexity.

3.2. Encoder pre-training: enhancement through stochastic data augmentation methodology

During the initial pre-training phase, the model undergoes fundamental unsupervised learning, wherein the loss function is deliberately structured to align with the SimCLR framework. This phase exclusively focuses on training the encoder E, while the subsequent pilot system remains inactive. The methodology employs stochastic augmentation to generate paired positive samples, whereby two augmented variants of each input image are processed through encoder E to extract corresponding feature vectors. In the implementation process, each original sample undergoes dual augmentation transformations, effectively expanding the initial dataset of $N + M$ samples to $2(N + M)$ augmented instances. Within this augmented dataset, pairs derived from the same source constitute positive samples, while the remaining $2(N + M) - 2$ instances serve as negative samples. Subsequently, the SimCLR contrastive learning framework is applied for model pre-training, governed by the following loss function:

$$\mathcal{L}_{pre} = \frac{1}{2(N+M)} \sum_{k=1}^{N+M} \left(l_{2k,2k-1} + l_{2k-1,2k}\right) \tag{6}$$

$$l_{i,j} = -\log \frac{\exp\left(\mathrm{sim}\left(\mathbf{z}_i, \mathbf{z}_j\right)/\tau\right)}{\sum_{k=1}^{2(N+M)} \mathbb{I}_{[k\neq i]} \exp\left(\mathrm{sim}\left(\mathbf{z}_i, \mathbf{z}_k\right)/\tau\right)} \tag{7}$$

$$\mathrm{sim}\left(\mathbf{z}_i, \mathbf{z}_j\right) = \frac{\mathbf{z}_i^{\top}\mathbf{z}_j}{\tau\|\mathbf{z}_i\|\|\mathbf{z}_j\|} \tag{8}$$

where N denotes the number of labeled data and M denotes the number of unlabeled data. $\mathbf{z}_i$ and $\mathbf{z}_j$ represent the feature vectors obtained from two augmented images generated from the same sample after feature extraction by encoder E. The parameter τ denotes the temperature coefficient, which controls the concentration level of the distribution. The function $\mathbb{I}_{[k\neq i]}$ is an indicator function whose range is $[0, 1]$, taking the value 1 if and only if $k \neq i$.

3.3. Phase 1: nearest neighbor-based positive sample selection and multi-objective optimization

This initial phase encompasses the concurrent training of both $C_{excessive}$ and $C_{moderate}$ classifiers through a comprehensive multi-objective optimization framework. The total loss function integrates multiple components: the supervised cross-entropy loss for labeled instances, the similarity-based loss for positive sample pairs derived from nearest neighbors, and two distribution regularization terms that address both standard-classified and over-classified class distributions.

Primarily, for the supervised component, we employ the conventional cross-entropy (CE) loss function on the labeled data subset. Given that the available ground truth labels correspond to broader class categories rather than fine-grained subdivisions, the cross-entropy loss is computed using the probability distribution $p_{moderate}$ obtained through $C_{moderate}$ and the corresponding true labels. Consequently, this supervisory signal exclusively contributes to the optimization of $C_{moderate}$. The cross-entropy loss function $\mathcal{L}_{CE}$ is formally defined as:

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{k=1}^{N} \sum_{i\in D_l} y_i \log\left(p_{moderate,i}\right) \tag{9}$$

where D_l represents the labeled dataset, y_i denotes the ground truth label of sample i, $p_{moderate,i}$ indicates the moderate assignment probability for sample i, and N represents the total number of labeled samples.

In parallel, we introduce a nearest neighbor similarity loss specifically designed for training $C_{excessive}$. This approach diverges from conventional contrastive learning methods by selecting nearest neighbor samples, rather than augmented variants of the same instance, as positive pairs. This selection strategy is justified by the encoder's acquired representational capabilities during pre-training, which ensures that proximate samples exhibit substantial similarity. Furthermore, the inherent variability between nearest neighbor samples, which exceeds that of augmented variants, facilitates more robust feature learning. The nearest neighbor similarity loss function $\mathcal{L}_{sim}^{p1}$ is expressed as:

$$\mathcal{L}_{sim}^{p1} = -\frac{1}{N+M}\sum_{i\in D}\log\left(sim\left(p_{excessive,i}, p'_{excessive,i}\right)\right) \tag{10}$$

where D denotes the entire dataset, sim denotes the cosine similarity metric, and $p_{excessive_i}$ and $p'_{excessive_i}$ represent the over-classified probability assignments for sample i and its corresponding neighbor, respectively.

While these loss components provide essential learning signals, they do not explicitly constrain the class distribution of unlabeled samples, potentially leading to model collapse where samples cluster into a single class, leaving numerous prototypes underutilized. To mitigate this issue, we incorporate distribution regularization through KL divergence:

$$\mathcal{L}_{reg} = KL\left(p_{prior} \| mean\left(p_{moderate}\right)\right) \tag{11}$$

where p_{prior} represents the theoretical prior distribution of samples.

Given the practical challenges in determining the true prior distribution, we approximate it with a uniform distribution. Although this assumption may seem restrictive, empirical evidence suggests that it does not significantly impact model performance, except in cases with highly skewed class distributions. The modified distribution regularization terms $\mathcal{L}_{reg1}$ and $\mathcal{L}_{reg2}$ are thus formulated as:

$$\begin{aligned}\mathcal{L}_{reg1} &= KL\left(Q \| mean\left(p_{moderate}\right)\right)\\ \mathcal{L}_{reg2} &= KL\left(Q \| mean\left(p_{excessive}\right)\right)\end{aligned} \tag{12}$$

where $\mathcal{L}_{reg1}$ and $\mathcal{L}_{reg2}$ correspond to the standard-classified and over-classified distribution regularization terms respectively, with Q representing the uniform distribution.

Ultimately, the comprehensive loss function for Phase 1 integrates all components through:

$$\mathcal{L}_1 = \mathcal{L}_{CE} + \mathcal{L}_{sim}^{p1} + \alpha\mathcal{L}_{reg1} + \beta\mathcal{L}_{reg2} \tag{13}$$

where α and β serve as balancing coefficients for the respective distribution regularization terms, enabling adaptive control over their relative contributions to the overall optimization objective.

3.4. Phase 2: refined intra-class cohesion and inter-class discrimination through fine-grained sample pair selection

Building upon the foundation established in Phase 1, this phase implements a more sophisticated data mining strategy that focuses on two critical objectives: (1) enhancing cohesion within over-classified categories and (2) establishing clear boundaries between different over-classified categories within the same standard-classified class. The model's representational capacity is significantly enhanced through the strategic selection of positive sample pairs within over-classified categories, while simultaneously implementing mechanisms to distinguish between distinct over-classified categories. This refined approach effectively identifies and excludes misclassified samples from their respective categories. While maintaining the cross-entropy loss and distribution regularization terms from Phase 1, this phase introduces two novel loss functions specifically designed to achieve these objectives.

In this refined framework, the classification criterion is derived from the assignment probability of the first enhanced sample, where $p_{excessive}$ is utilized to compute the probability distribution across over-classified categories. The final classification is determined by selecting the over-classified category with the highest probability value. Consequently, all samples are systematically partitioned into $n * k$ distinct subsets, denoted as $\{g_1, g_2, g_3, \ldots g_{n*k}\}$, where samples within each g_i are considered to belong to the same over-classified category. The corresponding similarity loss function for over-classified categories, $\mathcal{L}_{sim}^{p2}$, is formulated as:

$$\mathcal{L}_{sim}^{p2} = -\frac{1}{nk}\sum_{i=1}^{nk}\sum_{j\in g_i}\log\left(\mathrm{sim}\left(p_{\mathrm{excessive},j}, p_{\mathrm{excessive},j'}\right)\right) \tag{14}$$

where j' represents a randomly selected sample from the same fine-grained category as j, ensuring intra-class comparative learning.

To quantify the dissimilarity between over-classified categories within the same standard-classified class, we employ the Kullback–Leibler (KL) divergence. However, given the inherent asymmetry of traditional KL divergence in measuring information loss between probability distributions, we adopt the symmetric KL divergence, which provides a more comprehensive measure by considering bidirectional information flow. This choice is particularly appropriate as our objective is not to compare predicted distributions against true distributions, but rather to assess the distributional differences between samples within the same cluster. The symmetric KL divergence D_{KL} is defined as:

$$D_{KL}(p,q) = \frac{KL(p\|q) + KL(q\|p)}{2} \tag{15}$$

To guide the model in establishing clear boundaries between over-classified categories within the same standard-classified class, we utilize the average negative D_{KL} as a loss function. The process begins by partitioning samples into n standard classes, $\{G_1, G_2, G_3, \ldots G_n\}$, based on $p_{moderate}$. Subsequently, each standard class G_i is further subdivided into multiple over-classified categories $\{g_{i,1}, g_{i,2}, g_{i,3}, \ldots\}$. The average assignment probability for over-classified categories within each standard-classified class is computed as:

$$p_{i,j} = \frac{1}{N_{i,j}}\sum_{j\in g_{i,j}} p_{excessive_j} \tag{16}$$

where $g_{i,j}$ denotes the jth over-classified class of the ith standard-classified class, and $N_{i,j}$ represents the number of samples belonging to the jth over-classified class of the ith standard-classified class.

Building upon this foundation, we calculate the average loss of the ith largest class $\mathcal{L}_{D_{KL}}^{i}$ using the average assignment probabilities. The specific calculation formula is defined as:

$$\mathcal{L}_{D_{KL}}^{i} = -\frac{1}{\binom{\mathrm{class}}{2}}\sum_{x,y} D_{KL}\left(p_{i,x}, p_{i,y}\right) \tag{17}$$

where $p_{i,x}$ and $p_{i,y}$ represent the average assignment probabilities of any two small classes within the same large class, $class$ denotes the number of over-classified classes in the standard-classified class, and $\binom{\mathrm{class}}{2}$ signifies the number of possible combinations. To obtain a comprehensive measure across all standard-classified classes, the average symmetric KL divergence loss $\mathcal{L}_{D_{KL}}^{p2}$ is calculated as:

$$\mathcal{L}_{D_{KL}}^{p2} = \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_{D_{KL}}^{i} \tag{18}$$

The comprehensive loss function for this phase integrates all components through:

$$\mathcal{L}_2 = \mathcal{L}_{CE} + \mathcal{L}_{sim}^{p2} + \alpha\mathcal{L}_{reg1} + \beta\mathcal{L}_{reg2} + \gamma\mathcal{L}_{D_{KL}}^{p2} \tag{19}$$

The weighting parameters α, β, and γ play a crucial role in balancing the contributions of different loss components, enabling precise control over the model's learning dynamics and ensuring optimal performance in both intra-class cohesion and inter-class discrimination tasks. This hierarchical approach to loss calculation ensures that the model effectively learns both local (within over-classified categories) and global (across standard-classified categories) discriminative features, while maintaining the structural integrity of the classification hierarchy.

3.5. Phase 3: hierarchical consolidation through standard-class-based positive pair selection and inter-class discrimination

This phase represents the culmination of the progressive contrastive representation learning framework, focusing on two critical objectives: (1) reinforcing intra-class cohesion within standard-classified categories and (2) establishing clear discriminative boundaries between different standard-classified categories. While maintaining structural parallels with Phase 2, this phase implements these objectives through modified versions of the cross-entropy loss and distribution regularization terms, alongside redefined formulations of $\mathcal{L}_{sim}$ and $\mathcal{L}_{D_{KL}}$.

The classification mechanism in this phase utilizes the assignment probability derived from the initial augmented sample as its foundational criterion. Specifically, $p_{moderate}$ is employed to compute the probability distribution across standard classes, with class assignment determined by the maximum probability value. This process results in the partitioning of all samples into n distinct sets, denoted as $\{G_1, G_2, G_3, \dots G_n\}$, where samples within each G_i are considered to belong to the same standard-classified category. The corresponding similarity loss function for standard-classified categories, $\mathcal{L}_{sim}^{p3}$, is formulated as:

$$\mathcal{L}_{sim}^{p3} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j\in G_i}\log\left(\mathrm{sim}\left(p_{moderate,j}, p_{moderate,j'}\right)\right) \tag{20}$$

where j' represents a randomly selected sample from the same standard-classified category as j, ensuring intra-class comparative learning.

To quantify the dissimilarity between standard-classified categories, we employ the symmetric Kullback–Leibler (KL) divergence D_{KL}, which provides a robust measure of distributional differences between all possible pairs of standard-classified categories. The computation proceeds through two steps: first, calculating the average assignment probability for each standard-classified category, and second, determining the overall dissimilarity loss $\mathcal{L}_{D_{KL}}^{p3}$ across all category pairs:

$$p_i = \frac{1}{N_i}\sum_{j\in C_i} p_{moderate,j} \tag{21}$$

$$\mathcal{L}_{D_{KL}}^{p3} = -\frac{1}{\binom{n}{2}}\sum D_{KL}(p_i, p_j) \tag{22}$$

where C_i encompasses all samples belonging to the ith standard-classified category, N_i represents the cardinality of this category, and p_i and p_j denote the average assignment probabilities for any two distinct standard-classified categories. The denominator $\binom{n}{2}$ accounts for all possible pairwise combinations of standard-classified categories.

The comprehensive loss function for this final phase integrates all components through:

$$\mathcal{L}_3 = \mathcal{L}_{CE} + \mathcal{L}_{sim}^{p3} + \alpha\mathcal{L}_{reg1} + \beta\mathcal{L}_{reg2} + \gamma\mathcal{L}_{D_{KL}}^{p3} \tag{23}$$

where the weighting parameters α, β, and γ serve to balance the contributions of different loss components, enabling precise control over the learning dynamics of the model.

Based on the theoretical framework outlined above, we present a detailed implementation of this Progressive Contrastive Representation Learning (PCRL) methodology in Algorithm 1, which systematically orchestrates the progression through all three phases to achieve optimal representational learning.

Algorithm 1 Progressive contrastive representation learning framework.

Unsupervised pre-training

Input: Dataset: $D = D_{label} + D_{unlabel} = \{X_i\}_{i=1}^{N+M}$

Create augments for P_t : $D_p = \{X_{i,1}, X_{i,2}\}_{i=1}^{N+M} \xleftarrow{\text{Random augmentation}} D$

Pairs derived from the same source within the augmented dataset constitute positive samples

while not converge **do** Train on P_t

Calculate the output: $l_{i,j} = -\log \frac{\exp\left(\mathrm{sim}(\mathbf{z}_i,\mathbf{z}_j)/\tau\right)}{\sum_{k=1}^{2(N+M)} \mathbb{I}_{[k\neq i]} \exp\left(\mathrm{sim}(\mathbf{z}_i,\mathbf{z}_k)/\tau\right)}$

Calculate the loss $\mathcal{L}_{pre} = \frac{1}{2(N+M)}\sum_{k=1}^{N+M}\left(l_{2k,2k-1} + l_{2k-1,2k}\right)$

Update parameters by back propagation

end while

Output: optimized parameter for P_t

Phase 1: Nearest Neighbor-based Positive Pair Selection

Input: Dataset: $D = \{X_i\}_{i=1}^{N+M}$

while not converge **do** Train

Calculate excessive classification probability: $p_{excessive} = \frac{\exp\left(\frac{1}{\tau}\mathbf{z}\cdot C_{excessive}{}^{\top}\right)}{\sum\exp\left(\frac{1}{\tau}\mathbf{z}\cdot C_{excessive}{}^{\top}\right)}$

Calculate moderate classification probability: $p_{moderate} = \frac{\exp\left(\frac{1}{\tau}\mathbf{z}\cdot C_{moderate}{}^{\top}\right)}{\sum\exp\left(\frac{1}{\tau}\mathbf{z}\cdot C_{moderate}{}^{\top}\right)}$

Calculate $\mathcal{L}_{CE}$, $\mathcal{L}_{sim}^{p1}$, $\mathcal{L}_{reg1}$, $\mathcal{L}_{reg2}$

Calculate the total loss of this phase: $\mathcal{L}_1 = \mathcal{L}_{CE} + \mathcal{L}_{sim}^{p1} + \alpha\mathcal{L}_{reg1} + \beta\mathcal{L}_{reg2}$

Update parameters by back propagation

end while

Output: optimized parameter

Phase 2–3: Over-Class-Based Positive Pair Selection & Standard-Class-Based Positive Pair Selection

Input: Dataset: $D = \{X_i\}_{i=1}^{N+M}$

for $t = 2$ to 3 **do**

while not converge **do** Train

Calculate excessive classification probability $p_{excessive}$ and moderate classification probability $p_{moderate}$

Calculate the average symmetric KL divergence loss $\mathcal{L}_{D_{KL}}^{i}$: $\mathcal{L}_{D_{KL}}^{i} = -\frac{1}{\binom{\text{class}}{2}}\sum_{x,y} D_{KL}(p_{i,x}, p_{i,y})$

Calculate $\mathcal{L}_{CE}$, $\mathcal{L}_{sim}^{pt}$, $\mathcal{L}_{reg1}$, $\mathcal{L}_{reg2}$ and $\mathcal{L}_{D_{KL}}^{pt}$

Calculate the total loss of this phase: $\mathcal{L}_t = \mathcal{L}_{CE} + \mathcal{L}_{sim}^{pt} + \alpha\mathcal{L}_{reg1} + \beta\mathcal{L}_{reg2} + \gamma\mathcal{L}_{D_{KL}}^{pt}$

Update parameters by back propagation

end while

end for

Output: Make predictions for D

4. Experimental validation and analysis

In this section, we begin by offering a detailed overview of the dataset, alongside the classification of defects, which is elaborated upon in Section 4.1. Following this, Section 4.2 delves into the methodologies and techniques employed for event-based signal processing and analysis. To conclude, we present the experimental validation and comparative analysis in Section 4.3, where the effectiveness and performance of the proposed methods are rigorously evaluated against existing approaches.

4.1. Dataset description and defect categorization

This research utilizes the comprehensive surface inspection dataset introduced by Guo et al. (2023), which consists of event-based recordings systematically organized into 200 distinct segments and stored in the h5 file format. The dataset encompasses both defective and non-defective surface classifications, wherein the defective category is

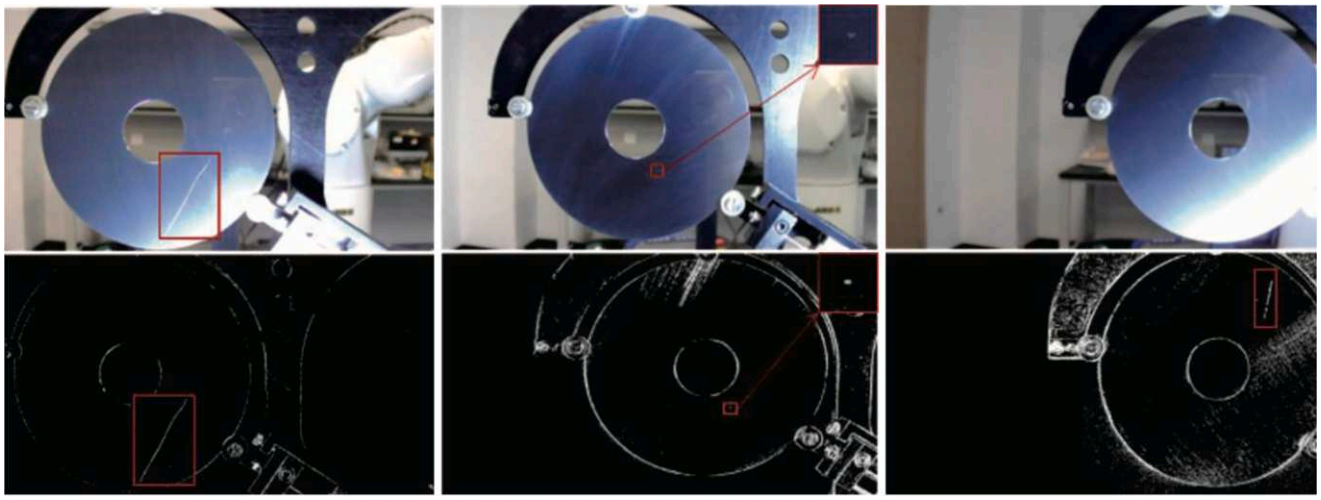


Fig. 3. Defective dataset description: (a) Stain, (b) Spot, (c) Scratch.

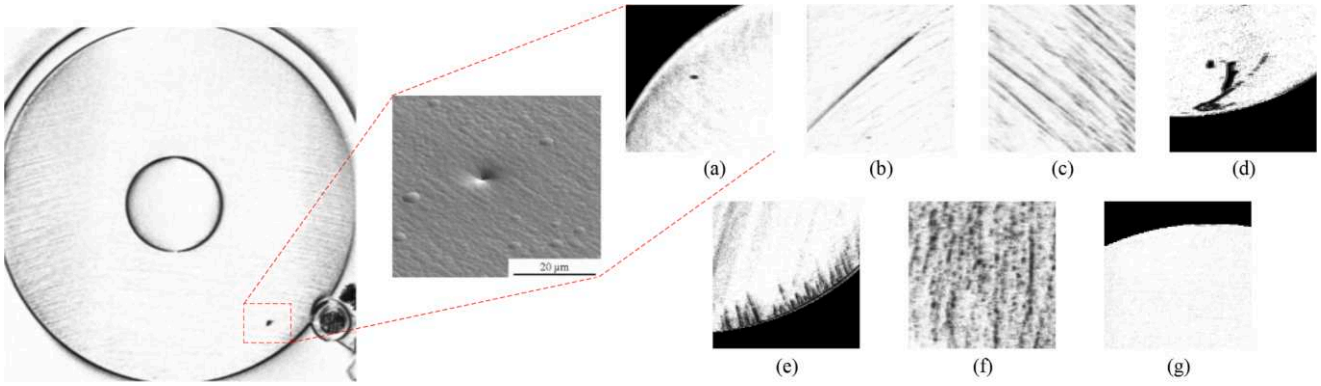


Fig. 4. Visual defects comparisons: (a) Punctuate , (b) Scratch , (c) Multi-scratch, (d) Blot, (e) Edge, (f) Mess mottle, and (g) Perfect.

further subdivided into three primary defect types: spot defects (previously categorized as Punctuate and Mess mottle), scratch defects (formerly classified as Scratch and Multi-scratch), and stain defects (previously designated as Blot and Edge).

To facilitate detailed defect analysis and comparison, these primary categories are further refined into specific subcategories based on their morphological characteristics, as illustrated in Fig. 4. The dimensional specifications of these defects have been precisely quantified through high-resolution measurements: spot defects are characterized by circular or irregular shapes with minimum diameters not exceeding 200 μm, while linear defects exhibit longitudinal patterns with minimum depth measurements below 10 μm, as documented in Fig. 3.

The experimental data acquisition protocol was implemented under controlled conditions, whereby the event camera remained in a fixed position while the aluminum substrate moved through the field of view at a constant velocity. To ensure robust model development and validation, the dataset underwent systematic partitioning: 180 event stream sequences were allocated to the training set, including 60 unlabelled sequences for unsupervised learning applications, while the remaining 20 sequences were reserved for testing purposes. Through meticulous preprocessing procedures, the training set yielded 15,768 key frames, comprising 5256 unlabelled event frames and 10,512 labelled event frames, whereas the test set generated 1780 frames. To establish statistical significance and ensure algorithmic reproducibility, the dataset partitioning process was repeated across 10 independent iterations, thereby minimizing potential sampling bias and enabling robust cross-validation procedures.

4.2. Event-based signal processing and analysis

This subsection presents a comprehensive methodology for converting event stream data into conventional image representations through temporal processing techniques. The process primarily involves the segmentation of event streams into uniform temporal intervals (time windows), where the careful selection of these intervals substantially influences the resultant image quality. To address the dual challenges of minor camera vibrations and brief camera interruptions, this research implements the Hough circle transformation algorithm to precisely locate the storage medium's center and subsequently perform targeted cropping operations. This approach effectively isolates the storage medium from its surrounding background, ensuring robust detection regardless of the medium's relative displacement to the camera position.

Given the inherent noise characteristics of event cameras and subtle ambient lighting variations, this study employs a dual-stage noise reduction strategy combining overlapping temporal windows with Gaussian filtering. The raw event stream data is structured in an $n \times 4$ matrix format, where each row corresponds to a single event characterized by four parameters: grayscale value, timestamp, and spatial coordinates (x, y) of the triggered pixel.

The fundamental visualization approach treats events as discrete scatter points plotted on a two-dimensional coordinate system. However, due to the horizontal motion of the storage medium during data acquisition, direct visualization of the complete event stream results in significant motion artifacts, analogous to motion blur in conventional

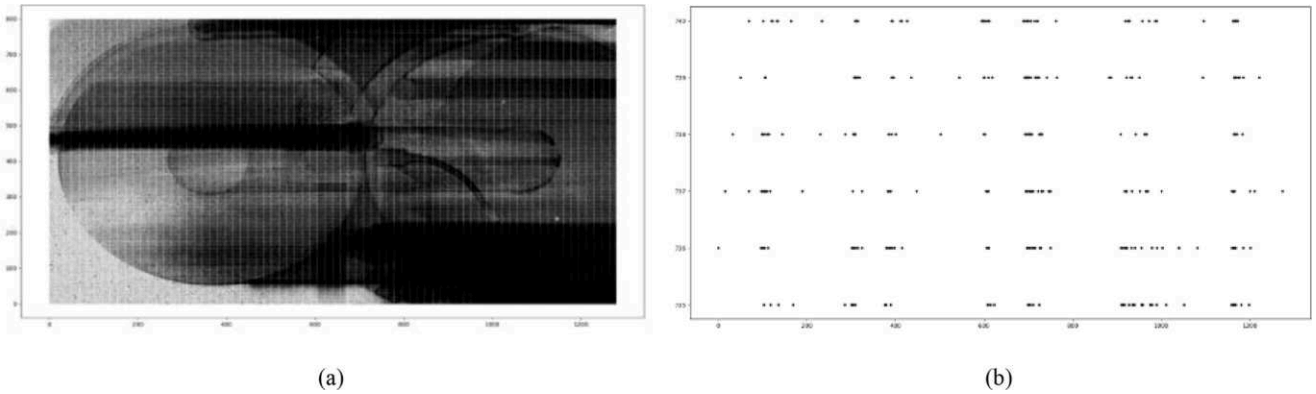


Fig. 5. Comparative visualization of event stream data: (a) Cumulative representation of all events demonstrating motion artifacts, (b) Instantaneous event distribution at a specific temporal instance showing data sparsity.

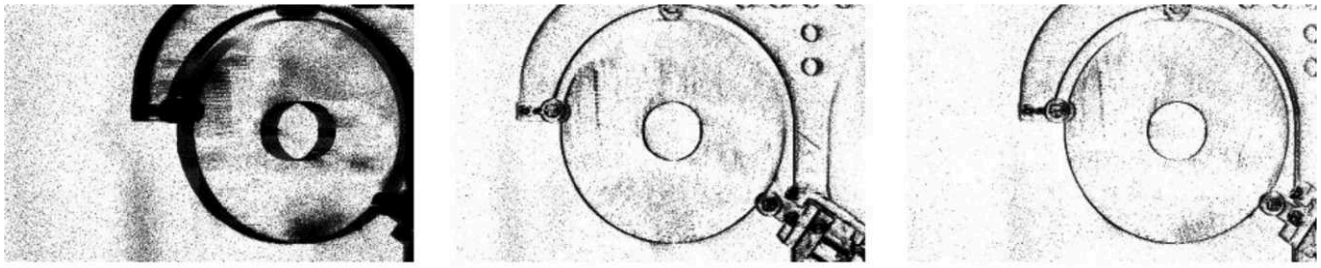


Fig. 6. Temporal window analysis demonstrating the effect of varying integration intervals: (a) Coarse temporal resolution ($\Delta = 0.1$ s), (b) Intermediate temporal resolution ($\Delta = 0.01$ s), (c) Fine temporal resolution ($\Delta = 0.001$ s).

frame-based cameras. Furthermore, instantaneous event visualization proves inadequate due to the sparse temporal nature of event data.

To overcome these limitations, we implement a temporal windowing approach, segmenting the event stream into equal-duration intervals and aggregating events within each window, as illustrated in Fig. 5.

Although the initial imaging results clearly delineate the storage medium's contour, the inherent characteristics of event cameras introduce considerable noise that could significantly impact subsequent model training. To address this challenge, we implement an advanced multiple time window overlapping methodology. Specifically, the event stream is initially partitioned into segments with a precise temporal resolution of 0.001 s, followed by sequential scatter plot generation, as illustrated in Fig. 6. Subsequently, we employ the Hough circle transformation to determine the storage medium's central coordinates, enabling precise image cropping. The process culminates in the superposition of all temporal windows corresponding to the sample, yielding a comprehensive final image. To further enhance image quality, we apply Gaussian filtering as the terminal step in our noise reduction pipeline.

The experimental data acquisition was conducted using a stationary event camera while the storage medium underwent horizontal translation within the camera's field of view. The event stream data is archived in .h5 file format, with each file encapsulating the complete event sequence for a single storage medium. Upon parsing, the data yields an $n4$ matrix, where each event is represented as $event = (gs, t, x, y)$. The grayscale values (gs) range from 0 to 255, while the spatial coordinates span x ∈ [0,1279] and y ∈ [0,799], corresponding to the camera's resolution. Individual .h5 files contain 7–11 s of event data, comprising tens of millions of events. The storage medium's physical dimensions are characterized by outer and inner ring diameters of 700 and 100 pixels, respectively, as determined through Hough circle transformation.

Following careful examination of aluminum disk substrates degradation patterns, our comprehensive dataset incorporates seven distinct categories of physical conditions that commonly manifest in optical storage media: punctuate, scratch, multi-scratch, blot, edge, mess mottle, and perfect specimens serving as control samples. In order to facilitate detailed analysis, we implement an event stream imaging protocol, through which we systematically extract regions of interest measuring 64×64 pixels that effectively capture the characteristic features of various damage patterns. Subsequently, to enhance the robustness and generalization capacity of our model while maintaining data authenticity, these extracted samples undergo strategic rotational augmentation at four cardinal angles (0°, 90°, 180°, and 270°), thereby effectively quadrupling the size of our training dataset without introducing artificial variations. As illustrated comprehensively in Fig. 7, the sophisticated data preprocessing pipeline consists of several critical stages for transforming raw event data. Initially, a Hough circle detection algorithm identifies and marks the image center with a red circle. This detected center then serves as a reference point for overlapping window processing, where green boxes define key shear boundaries around the circle. The pipeline continues with Gaussian filtering operations to reduce noise and enhance signal quality. Through these sequential transformations, the raw data is refined into optimized image representations that are well-suited for downstream analysis and classification tasks.

4.3. Experimental validation and comparative analysis

To rigorously evaluate the effectiveness and advantages of the proposed Progressive Contrastive Representation Learning (PCRL) model, we conducted a comprehensive validation utilizing the aforementioned datasets. In order to ensure a thorough comparative analysis, we systematically employed several well-established state-of-the-art semi-supervised learning techniques, including DTC (Han et al., 2019), ORCA (Cao, 2024; Xiao et al., 2024), IIC (Li et al., 2023a), and OpenNCD (Liu et al., 2023) as benchmarks. These particular models were deliberately selected due to their proven track record of robust performance and significant relevance within the domain, thereby providing a substantive foundation for meaningful comparative assessments.

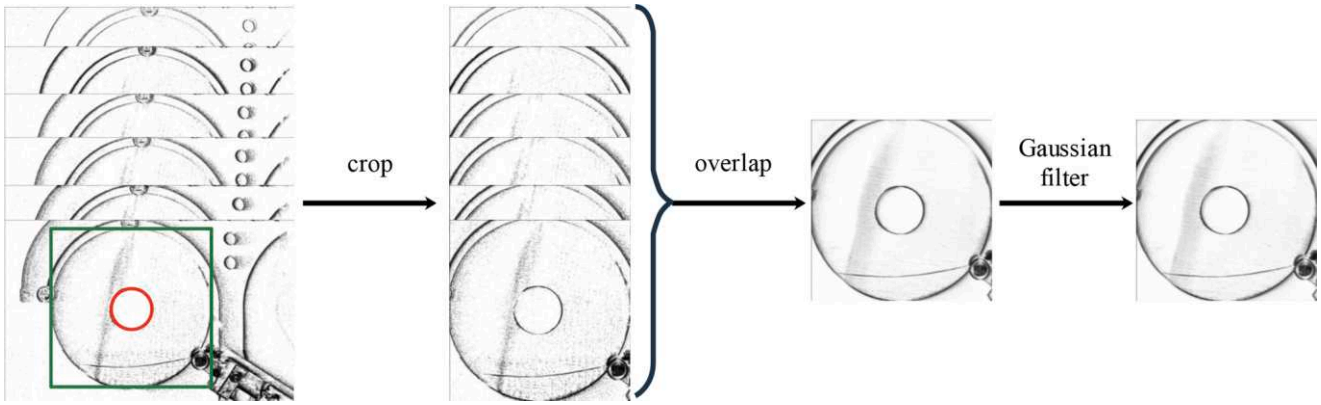


Fig. 7. Sequential visualization of the multi-stage event stream processing pipeline.

A significant methodological advancement in this research lies in the comprehensive refinement of the random augmentation protocol, which systematically incorporates three fundamental components: chromatic adjustment, geometric flipping, and selective cropping. Among these, the cropping mechanism represents a particularly noteworthy departure from the conventional SimCLR approach, as it has been specifically tailored to address the unique challenges presented by storage medium datasets. This modification was primarily driven by the distinctive characteristics of storage medium defect patterns, wherein defective regions typically constitute a minimal fraction of the total image area. Traditional random cropping methodologies would frequently eliminate these crucial defective regions, thereby potentially compromising the model's learning efficacy and overall performance. To overcome this substantial limitation, we introduce a novel critical area-based random cropping strategy, which has been meticulously developed through comprehensive dataset analysis. Our approach capitalizes on the observation that defective regions consistently exhibit substantially higher pixel intensity gradients relative to their surrounding areas. Leveraging this intrinsic characteristic, the methodology incorporates a sophisticated initial peak detection phase that precisely identifies critical regions while systematically excluding peripheral areas beyond the storage medium boundaries. These strategically identified locations, formally designated as key points, subsequently serve as essential reference points for the cropping operation, which preserves between 25 % and 100 % of the original image area.

As illustrated in Fig. 8, where red markers distinctly indicate algorithmically detected key points, the margin parameters are systematically computed as the distances from these key points to the image boundaries across all four cardinal directions. The expansion algorithm, exemplified by the dx_l parameter, is governed by the following mathematical formulation: parameter, follows the formula:

$$len = dx_l/2 + random(dx_l/2) \tag{24}$$

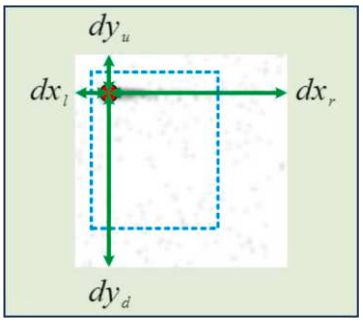


Fig. 8. Diagram illustrating the clipping procedure for key regions.

This rigorous algorithmic approach ensures the consistent retention of at least half the margin in each direction while maintaining a minimum preserved area threshold of 25 % of the original image. Moreover, this enhanced augmentation protocol, which seamlessly integrates all three augmentation operations, is systematically implemented throughout the subsequent three phases of the methodology, thereby ensuring consistency and robustness in the overall approach.

For the neural network architecture, we employ ResNet-18 as the feature extraction backbone, configured with a 32-dimensional output. The over-clustering factor $k = 5$ results in 35 distinct clusters (5×7). We utilize the Adam optimizer with a learning rate of 0.05 and batch size of 512. Additional hyperparameters include: temperature coefficient $\tau = 5$; regularization weights $(\alpha, \beta) = (10, 10)$; D_{KL} loss weight $\gamma = 0.1$. The training protocol comprises 400 epochs, distributed across four stages: Pre-training (300 epochs), Phase 1 (20 epochs), Phase 2 (30 epochs), and Phase 3 (50 epochs).

4.3.1. *Baseline model selection and configuration*

In this experimental framework, DTC, ORCA, IIC, and OpenNCD serve as baseline models, whereby ORCA and OpenNCD maintain identical partition structures of the dataset as PCRL. Nevertheless, it should be noted that the dataset partition methodology for DTC and IIC differs fundamentally, as it treats the known and unknown classes as completely disjoint sets, focusing solely on unknown class identification. To facilitate the extension of these two models for known class recognition, we incorporated known class samples as components of the unknown class, subsequently detecting and reporting performance metrics such as accuracy in a manner consistent with standard unknown class evaluation protocols. Across all experimental configurations, we maintained a consistent allocation of 10 % labeled data within known classes.

4.3.2. *Multi-dimensional performance metrics framework*

Our evaluation framework encompasses five distinct metrics: (1) Known class accuracy, which is derived by extracting and analyzing samples with true labels from known classes (specifically, the top 4 classes); (2) Unknown class accuracy, determined through the extraction and analysis of samples whose real labels belong to new classes (specifically, the last 3 classes); (3) All class accuracy, calculated through the implementation of the Hungarian algorithm to determine optimal matching schemes based on comprehensive clustering results; (4) Normalized Mutual Information (NMI), which quantifies the correlation between clustering results and true labels, with values ranging from $[0, 1]$, wherein higher values indicate stronger correlations; and (5) Adjusted Rand Index (ARI), which measures clustering agreement with true class labels across a $[-1, 1]$ range, where positive values indicate stronger agreement, zero suggests random matching, and negative values represent divergence from true labels.

Table 1
Comparative analysis of classification performance metrics and the inference efficiency across state-of-the-art models.

Methods	Overall	Known	Unknown	NMI	ARI	Time(s)
DTC	65.23 %	82.49 %	71.20 %	0.6660	0.5691	891.94
IIC	83.22 %	75.18 %	93.32 %	0.7113	0.6410	950.23
ORCA	91.82 %	91.32 %	92.45 %	0.8294	0.8259	640.81
ASSBert	79.49 %	78.67 %	80.51 %	0.6533	0.4992	923.85
OpenNCD	91.90 %	91.12 %	92.89 %	0.8178	0.8249	682.54
PCRL (Ours)	**93.51 %**	**92.02 %**	**95.36 %**	**0.8497**	**0.8572**	549.43

Table 2
Component-wise ablation study results.

Methods	Overall	Known	Unknown
$w/o\ \mathcal{L}_{sim}^{p2}$	81.2728 %	90.8184 %	67.9927 %
$w/o\ \mathcal{L}_{SKL}^{p2}$	78.5103 %	90.9496 %	61.2043 %
$w/o\ \mathcal{L}_{sim}^{p3}$	76.3431 %	90.2938 %	56.9343 %
$w/o\ \mathcal{L}_{SKL}^{p3}$	85.3937 %	90.8709 %	77.7737 %
PCRL (Ours)	**93.5060 %**	**92.0232 %**	**95.3649 %**

Thus, the results of multi-dimensional performance metrics are shown in Table 1. It is obvious that our proposed PCRL is far superior to other models used for comparison in terms of accuracy, NMI and ARI. Specifically, the total accuracy of PCRL is as high as 93.51 %, the accuracy of known class and the accuracy of unknown class reach a high level of 92.02 % and 95.36 %, respectively. NMI and ARI are better than other models, reaching 0.8497 and 0.8572. Our proposed model demonstrates exceptional computational efficiency, completing the training process in merely 549.43 s while maintaining competitive classification accuracy. This performance represents a significant advancement when compared to established state-of-the-art frameworks, particularly considering that prominent models such as IIC and ASSBert require substantially longer processing times of 950.23 and 923.85 s, respectively. Additionally, when evaluating inference efficiency against contemporary architectures, our model outperforms several notable benchmarks, including DTC (891.94 s), ORCA (640.81 s), and OpenNCD (682.54 s). The marked reduction in computational overhead, coupled with the preservation of classification accuracy, underscores the model's optimization capabilities and its potential for real-world applications where processing time constraints are critical considerations.

4.3.3. Component-wise ablation analysis

To systematically investigate the individual contributions of various loss function components within our proposed methodology, we conducted a series of meticulously designed ablation studies. These analyses specifically focused on examining the impacts of both the similarity loss $\mathcal{L}_{sim}$ and the Kullback–Leibler divergence loss $\mathcal{L}_{D_{KL}}$ during the second and third phases of model operation. For enhanced clarity and precision in our analysis, we designated $\mathcal{L}_{sim}^{p2}$ and $\mathcal{L}_{D_{KL}}^{p2}$ to represent the similarity and inter-class dissimilarity losses in Phase 2, respectively, while $\mathcal{L}_{sim}^{p3}$ and $\mathcal{L}_{D_{KL}}^{p3}$ denote their counterparts in Phase 3. Through systematic component removal and subsequent performance monitoring, we established a comprehensive understanding of each element's contribution to overall model effectiveness.

The results presented in Table 2 demonstrate that each module makes substantial positive contributions to the model's overall functionality and performance metrics. Notably, the removal of $\mathcal{L}_{sim}^{p3}$ resulted in a particularly significant degradation of performance metrics, underscoring its critical importance. This finding is especially relevant given that Phase 3 represents the culmination of the contrastive learning process, playing an instrumental role in facilitating large-category aggregation and establishing the fundamental framework for final sample classification. Consequently, the inclusion of $\mathcal{L}_{sim}^{p3}$ proves essential for maintaining optimal model functionality and achieving superior performance outcomes.

4.3.4. Feature distribution visualization through T-SNE analysis

The T-SNE visualization results, as illustrated in Fig. 9, demonstrate that the proposed PCRL method effectively characterizes features into distinctly separated clusters corresponding to various fault categories. Furthermore, this sophisticated classification framework not only encompasses known fault classes, including punctate, scratch, multi-scratch, and blot, but also successfully extends to accommodate previously unknown classes such as edge, mess mottle, and perfect. Additionally, these comprehensive visualizations empirically validate that the PCRL model excels at capturing and representing the intricate patterns inherent in the data associated with known classes, while simultaneously exhibiting remarkable adaptability in fault diagnosis when confronted with novel, unknown classes.

Notably, the pronounced proximity of data points within each known fault category strongly indicates robust intra-class similarity, thereby substantiating the effectiveness of the PCRL model. Concurrently, the distinct separation between different clusters underscores the model's

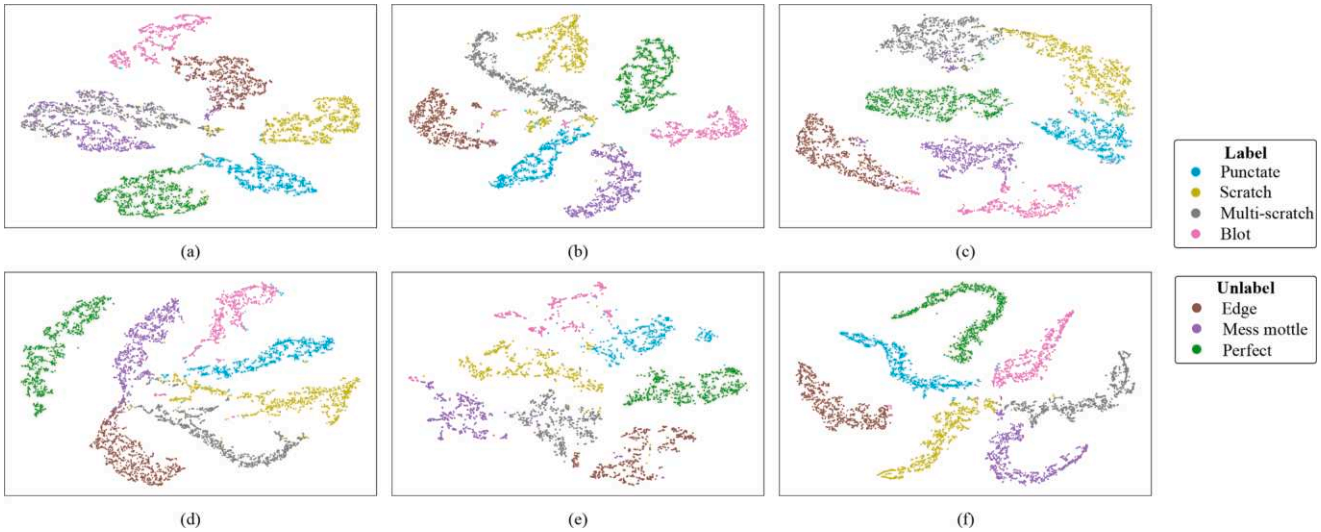


Fig. 9. The T-SNE visualization of learned feature representations: Comparative analysis across six models with 4 known classes (10 % labeled) and 3 unknown classes on (a) DTC, (b) IIC, (c) ORCA, (d) ASSBert, (e) OpenNCD and (f) PCRL (Ours).

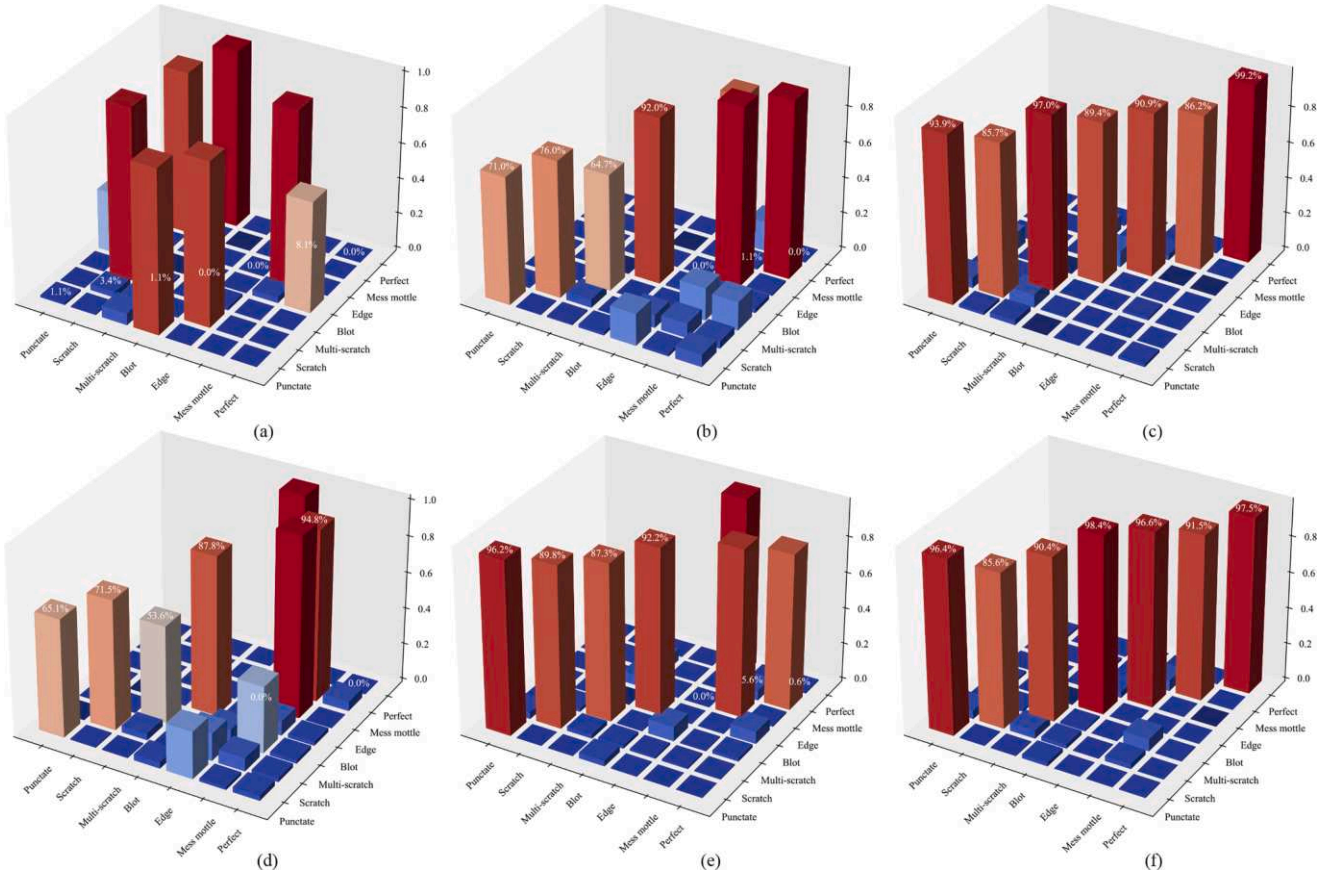


Fig. 10. Multi-class performance analysis through Confusion Matrices: Comparative evaluation of six models with 4 known classes (10 % labeled) and 3 unknown classes on (a) DTC, (b) IIC, (c) ORCA, (d) ASSBert, (e) OpenNCD and (f) PCRL (Ours).

exceptional proficiency in discriminating various fault types, consequently enhancing diagnostic accuracy. Conversely, T-SNE visualizations derived from alternative semi-supervised learning approaches, specifically DTC and IIC, frequently exhibit overlapping clusters, particularly when processing unknown fault classes. This overlapping phenomenon fundamentally indicates a substantial reduction in feature extraction effectiveness and classification performance compared to the PCRL methodology.

In the specific context of the ORCA model, while certain known classes such as multi-scratch and blot can be identified through well-distributed clustering, significant challenges emerge in distinguishing classes like mess mottle, which frequently exhibit overlap with multi-scratch and edge classifications. Moreover, faults from known classes, particularly blot, are often erroneously classified as unknown classes, thereby highlighting the model's inherent limitations in effectively generalizing across diverse data instances. The T-SNE visualization reveals a significant limitation in ASSBert's performance, where despite its effectiveness in distinguishing unknown classes, the model exhibits substantial confusion and overlap when classifying known categories, resulting in considerable misclassification among established class boundaries. In contrast, the T-SNE visualizations associated with the PCRL approach consistently demonstrate superior separability between known and unknown faults, thus not only reinforcing the model's exceptional classification capabilities but also underscoring its robust diagnostic performance, particularly in complex fault identification scenarios.

4.3.5. Quantitative performance assessment via confusion matrix analysis

To facilitate a more rigorous and comprehensive quantitative analysis of the diagnostic results, we employed confusion matrix (CM) analysis, with the resultant findings illustrated in Fig. 10. The empirical evidence demonstrates that when the proposed PCRL method undergoes training with 3000 samples per category, it consistently achieves remarkable accuracy across all fault classifications. Specifically, the method demonstrates exceptional performance metrics, achieving accuracy rates of 96.4 %, 85.6 %, 90.4 %, and 98.4 % for Punctate, Scratch, Multi-scratch, and Blot classifications, respectively. Furthermore, the PCRL methodology exhibits robust performance in detecting previously unknown categories, achieving accuracy rates of 96.6 % for Edge detection, 91.5 % for Mess mottle identification, and 97.5 % for Perfect classification.

It is pertinent to note that both IIC and DTC methodologies are fundamentally constrained by their problem formulation, wherein the models are exclusively utilized for novel category identification. To ensure equitable comparative analysis, during the evaluation of IIC and DTC performance metrics, known categories were necessarily incorporated as constituent elements of the novel category set. Consequently, while the DTC semi-supervised learning model exhibits minor deviations in confusion matrix distribution—attributable to the absence of true labels for new classes and reliance on clustering mechanisms—it nevertheless demonstrates suboptimal fault identification capabilities.

Moreover, the IIC model exhibits merely marginal accuracy in distinguishing known fault classes associated with storage medium impairments, with surprisingly low recognition rates of 71.0 %, 76.0 %, and 64.7 % for punctate, scratch, and multi-scratch classifications, respectively. The classification accuracy of ASSBert on known classes is relatively low, only the accuracy of Blot class is high, reaching 87.8 %, but the classification accuracy of unknown classes is high, especially the accuracy of Mess mottle reaches 94.8 %. Although alternative

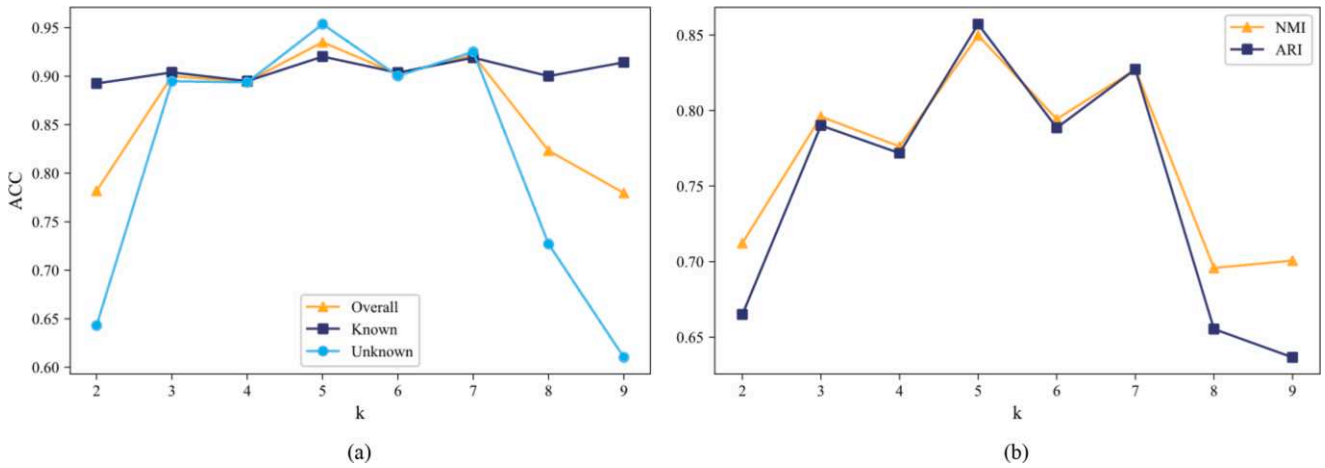


Fig. 11. (a) The impact of over-clustering factor k on accuracy (b) The impact of over-clustering factor k on NMI and ARI.

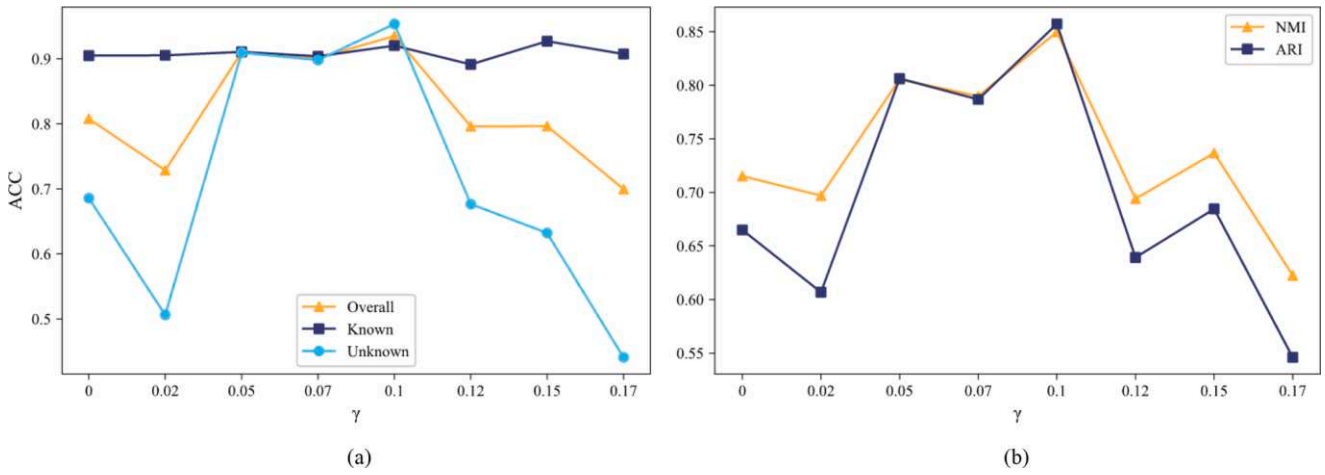


Fig. 12. Performance sensitivity analysis: (a) The impact of symmetric KL divergence weight γ on accuracy (b) The impact of symmetric KL divergence weight γ on NMI and ARI.

semi-supervised models such as ORCA and OpenNCD demonstrate high accuracy in identifying known fault classes, they manifest significant limitations in unknown fault classification, indicating substantial generalization deficiencies. For instance, while the ORCA model accurately identifies certain fault types (Punctate: 93.9 %, Multi-scratch: 97.0 %, Perfect: 99.2 %), it exhibits diminished recognition accuracy for other classifications such as Scratch (85.7 %) and Mess mottle (86.2 %).

4.3.6. *Hyperparameter sensitivity analysis and optimization*

This section presents a comprehensive investigation of model performance across varying values of the hyperparameter k, specifically examining the range from $k = 2$ to $k = 9$. As evidenced in Fig. 11, model accuracy demonstrates remarkable stability within the over-clustering multiple range of 3 to 7, achieving optimal performance at $k = 5$. However, a notable performance degradation is observed when k approaches 9.

This phenomenon can be attributed to the fundamental role of k in controlling over-clustering multiplication: while theoretically, larger k values should not inherently compromise model performance through the creation of additional small clusters, practical hardware limitations necessitate batch-wise data processing. Given that the internal similarity calculations proposed in this methodology are batch-specific (with a batch-size constraint of 512), excessive cluster subdivision results in insufficient samples per small class. This data sparsity consequently leads to inadequate information provision for the model, culminating in decreased performance metrics. Therefore, based on these empirical findings, we strongly recommend maintaining the over-clustering factor k at approximately 5 for optimal performance.

Furthermore, we conducted a comprehensive sensitivity analysis of the model's performance across various gamma values, specifically examining $\gamma = 0$, $\gamma = 0.02$, $\gamma = 0.05$, $\gamma = 0.07$, $\gamma = 0.1$, $\gamma = 0.12$, $\gamma = 0.15$, and $\gamma = 0.17$. The resultant performance metrics, as illustrated in Fig. 12, reveal several significant patterns. Notably, while the classification accuracy for known classes demonstrates remarkable resilience to variations in γ, the parameter exhibits substantial influence over the classification accuracy of novel classes, as well as the Normalized Mutual Information (NMI) and Adjusted Rand Index (ARI) metrics.

This phenomenon can be attributed to the fundamental role of symmetric KL divergence in class separation. When γ assumes extremely small values, the Phase 2 training process fails to effectively discriminate between distinct classes within the same larger classification group-a critical objective of this phase. The failure to accomplish this task effectively nullifies the Phase 2 process, essentially creating a direct transition from Phase 1 to Phase 3, inevitably resulting in degraded model performance. Conversely, when γ assumes excessively large values, it induces substantial repulsive forces between different subclasses within the same major classification during Phase 2 training. This leads to frequent perturbations in the model's sample classification structure, thereby significantly increasing the probability of non-convergence. Consequently, the judicious selection of γ proves instrumental in optimizing the model's ultimate performance, with empirical evidence suggesting that a γ value of 0.1 achieves optimal results.

Table 3
Comparison of key hyperparameter tuning performance: Over-clustering factor k and symmetric KL divergence weight γ.

Over-clustering factor						Symmetric KL divergence					
k	Overall	Known	Unknown	NMI	ARI	γ	Overall	Known	Unknown	NMI	ARI
7 2	78.18 %	89.25 %	64.30 %	0.7122	0.6650	0.00	80.77 %	90.48 %	68.57 %	0.7152	0.6649
3	89.99 %	90.39 %	89.48 %	0.7960	0.7901	0.02	72.85 %	90.53 %	50.62 %	0.6969	0.6068
4	89.44 %	89.49 %	89.37 %	0.7762	0.7718	0.05	91.01 %	91.06 %	90.94 %	0.8056	0.8062
5	**93.50 %**	**92.02 %**	**95.36 %**	**0.8497**	**0.8572**	0.07	90.13 %	90.39 %	89.81 %	0.7898	0.7867
6	90.21 %	90.36 %	90.03 %	0.7944	0.7886	**0.10**	**93.50 %**	92.02 %	**95.36 %**	**0.8497**	**0.8572**
7	92.17 %	91.90 %	92.51 %	0.8272	0.8273	0.12	79.61 %	89.14 %	67.66 %	0.6940	0.6392
8	82.33 %	90.01 %	72.70 %	0.6958	0.6554	0.15	79.64 %	**92.71 %**	63.21 %	0.7367	0.6845
9	77.96 %	91.43 %	61.02 %	0.7007	0.6366	0.17	69.95 %	90.73 %	44.08 %	0.6225	0.5460

Table 4
Quantitative comparison of Area Under ROC Curve (AUC) metrics across models.

Classes	DTC	IIC	ORCA	ASSBert	OpenNCD	PCRL (Ours)
Punctate	0.52	0.90	0.99	0.90	0.50	**1.00**
Scratch	0.36	0.93	0.99	0.92	0.60	**0.99**
Multi-scratch	0.18	0.89	0.99	0.84	0.42	**0.99**
Blot	0.76	0.98	0.99	0.96	0.59	**1.00**
Edge	0.45	0.72	1.00	0.75	0.51	**1.00**
Mess mottle	0.50	0.58	0.99	0.88	0.52	**0.99**
Perfect	0.18	0.79	1.00	0.77	0.38	**1.00**

In addition, the overall results of the influence of taking different values of k and γ on the model performance are presented in Table 3.

4.3.7. Model performance characterization through ROC analysis

The Receiver Operating Characteristic (ROC) curves, presented in Fig. 13, provide a comprehensive visualization of model performance by plotting the true positive rate against the false positive rate across varying threshold configurations. This sophisticated evaluation metric offers particularly valuable insights into the model's discriminative capabilities, wherein curves approaching unity indicate superior fault diagnosis performance.

As evidenced in Table 4, the ROC curves associated with our proposed PCRL model demonstrate near-optimal performance (approaching unity) in identifying seven distinct aluminum disk substrates damage categories, encompassing broken, chipping, crack, and normal conditions. In contrast, the DTC model exhibits significant performance limitations across damage type identification tasks, with particularly notable deficiencies in Multi-scratch and Perfect detection, achieving remarkably low AUC values of 18 % for both categories. While both IIC and ORCA models demonstrate reasonably competitive performance overall, the IIC model specifically shows substantial limitations in identifying unknown classes, achieving suboptimal AUC values of 72 %, 58 %, and 79 % for Edge, Mess mottle, and Perfect classifications, respectively. The experimental results further indicate that ASSBert's classification performance exhibits notable limitations, particularly in discriminating between Edge and Perfect classes, where the model demonstrates suboptimal accuracy. Specifically, the Area Under the Curve (AUC) metrics for these categories reach only 75 % and 77 % respectively, suggesting that while the model maintains moderate discriminative capability, it falls short of the robust performance levels typically desired for reliable automated classification systems. Moreover, these relatively modest AUC values underscore the inherent challenges in distinguishing between subtle variations in these particular classes, thereby highlighting the need for potential architectural improvements or enhanced feature engineering approaches to boost the model's classification efficacy. Furthermore, the OpenNCD framework demonstrates consistently subpar performance, with AUC valuesD hovering around 50 % across all seven damage categories. These comprehensive comparative analyses provide compelling evidence for the superior capability of the proposed PCRL methodology in accurately identifying both known and unknown damage classes while maintaining consistently high detection rates across all categories.

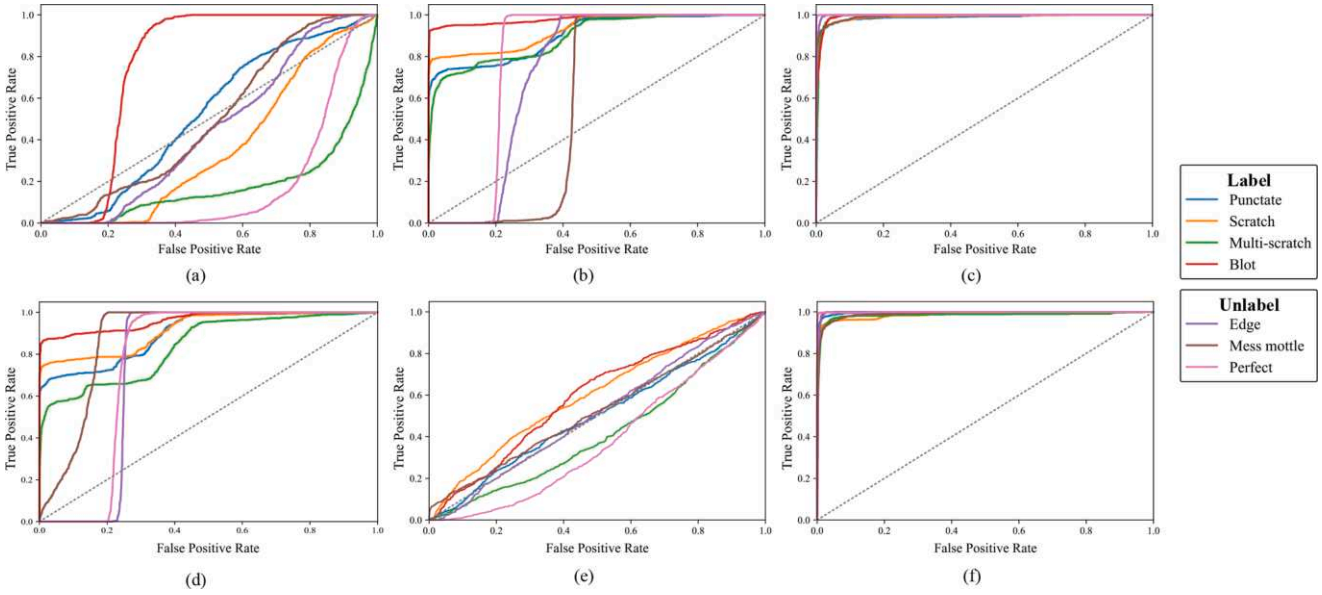


Fig. 13. The Receiver Operating Characteristic (ROC) analysis with 4 known classes (10 % labeled) and 3 unknown classes on (a) DTC, (b) IIC, (c) ORCA, (d) ASSBert, (e) OpenNCD and (f) PCRL (Ours).

4.3.8. Analytical assessment of methodological benefits

The entire progressive contrastive learning process begins by utilizing augmented samples as positive pairs during the pre-training phase. This process enables the model to acquire basic representation capabilities without labeled data. However, the knowledge gained remains limited due to the extreme similarity between samples and their augmented counterparts. To address this limitation, we subsequently employ each sample's nearest neighbors as positive pairs. Through the guidance of the loss function, the model learns to bring these more substantially different but highly similar samples closer in the feature space, thereby enhancing its representation capacity. Building upon this foundation, we introduce over-clustering techniques to achieve finer sample clustering. By treating samples within the same cluster as positive pairs, we intentionally increase the model's learning difficulty while enabling it to capture more abstract and beneficial patterns. At this stage, we implement symmetric KL divergence as the loss function to guide the model in distinguishing subtle subclass variations within broader categories. This approach resembles assigning students advanced problems beyond standard curriculum requirements to strengthen their inductive reasoning capabilities. In the final phase, we establish same-class samples as positive pairs while continuing to employ symmetric KL divergence. The loss function now serves dual purposes: compressing intra-class feature distributions while expanding inter-class differences, thereby optimizing the model for its ultimate task objectives. Notably, our loss function incorporates two critical components: (1) Cross-entropy loss for labeled samples ensures proper clustering direction in our semi-supervised framework, and (2) KL divergence between the model's cluster distribution and uniform distribution prevents feature space collapse. This is a critical safeguard given contrastive learning's inherent tendency to map all samples to proximate locations without explicit constraints. The core methodology follows a progressive challenge mechanism: as the model's feature extraction capability improves, we systematically introduce more demanding learning tasks. This approach mirrors human cognitive development patterns, where gradual exposure to increasingly complex problems facilitates deeper understanding. The structured escalation of learning difficulty ultimately constitutes the fundamental basis for our method's success.

5. Conclusion

This paper presents an enhanced event stream imaging technique and the PCRL framework, addressing the challenges in industrial surface defect detection, particularly for aluminum substrates. The experimental results demonstrate that our proposed approach successfully overcomes the limitations of traditional CCD/CMOS cameras and effectively manages the noise issues inherent in event camera-based detection systems. The multi-step loss function strategy, combined with sophisticated analysis of intra-cluster and inter-cluster relationships, has proven particularly effective in improving feature extraction capabilities and model convergence. Our framework shows superior performance in both known and unknown fault class detection, achieving robust and reliable results in real industrial settings. The systematic validation through case studies confirms the practical applicability and effectiveness of PCRL in industrial defect detection scenarios.

Future work should focus on extending the PCRL framework to handle multi-modal sensor fusion and investigating its applicability to real-time adaptive learning scenarios in dynamic industrial environments, with particular emphasis on optimizing the computational efficiency of the contrastive learning process.

Code availability

To promote reproducibility and facilitate future research, the code implementations developed in this study will be made publicly available through the Smart Sensing and Sustainable Diagnosis Prognostics Lab (S3DP-LAB)'s digital archival platform upon publication. The platform provides comprehensive version control and long-term preservation capabilities, ensuring sustained accessibility and reusability of our research outputs. Researchers can access the complete codebase, including implementation details, documentation, and example usage through our institutional repository at https://drpengchen.vip.cpolar.cn/research-work/.

CRediT authorship contribution statement

Peng Chen: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition; **Ruijin Zhang:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization; **Changbo He:** Software, Investigation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition; **Yaqiang Jin:** Software, Investigation, Resources, Project administration; **Shuai Fan:** Methodology, Investigation, Resources, Project administration; **Junyu Qi:** Conceptualization, Methodology, Investigation, Resources, Project administration; **Chengning Zhou:** Conceptualization, Methodology, Investigation, Resources, Project administration, Funding acquisition; **Chun Zhang:** Conceptualization, Methodology, Software, Data curation, Writing – review & editing, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Partial funding for this research has been provided by several sources, including the National Natural Science Foundation of China under Grant 52105111, and 52305085, the Guangdong Basic and Applied Basic Research Foundation under Grant 2025A1515012256, the Shantou University (STU) Scientific Research Initiation Grant under Grant NTF21029, the Industry-Academia Cooperation Project from the Guangdong Institute of Special Equipment Inspection and Research Shunde Branch under Grant XTJ-KY01-202503-030, the Enterprise Collaboration Project from the National Excellent Engineer Innovation Research Institute for Advanced Manufacturing Industry in Foshan of Guangdong-Hong Kong-Macao Greater Bay Area under Grant NSJH2025008, the China Postdoctoral Science Foundation under Grant 2023M740021, the Natural Science Foundation of Anhui Province under Grant 2108085QE229, the Nuclear Power Institute of China Original Foundation under Grant KJCX2022YC111, and the Natural Science Foundation of Sichuan Province under Grant 2023NSFSC0861.

References

Cao, K. (2024). Enhancing Machine Learning With Data-Efficient Methods. Ph.D. thesis. Stanford University.

Chen, P., Ma, J., He, C., Jin, Y., & Fan, S. (2025a). Semi-supervised consistency models for automated defect detection in carbon fiber composite structures with limited data. *Measurement Science and Technology*, *36*(4), 046109.

Chen, P., Ma, Z., Xu, C., Jin, Y., & Zhou, C. (2024a). Self-supervised transfer learning for remote wear evaluation in machine tool elements with imaging transmission attenuation. *IEEE Internet of Things Journal*, *11*, 23045–23054.

Chen, P., Ma, Z., Xu, C., Zhang, M., Li, H., Zheng, K., & Jin, Y. (2024b). Scale-aware domain adaptation for surface defects detection on machine tool components in contaminant measurements. *IEEE Transactions on Instrumentation and Measurement*, *74*, (pp. 1–9).

Chen, P., Wu, Y., Fan, S., He, C., Jin, Y., Qi, J., & Zhou, C. (2025b). Adaptive signal regime for identifying transient shifts: A novel approach toward fault diagnosis in wind turbine systems. *Ocean Engineering*, *325*, 120798.

Chen, P., Wu, Y., Xu, C., Huang, C.-G., Zhang, M., & Yuan, J. (2025c). Interference suppression of nonstationary signals for bearing diagnosis under transient noise measurements. *IEEE Transactions on Reliability*, (pp. 1–15).

Chen, P., Xu, C., Ma, Z., & Jin, Y. (2023). A mixed samples-driven methodology based on denoising diffusion probabilistic model for identifying damage in carbon fiber composite structures. *IEEE Transactions on Instrumentation and Measurement*, *72*(3513411), 1–11.

Chen, P., Zhang, R., Fan, S., Guo, J., & Yang, X. (2024c). Step-wise contrastive representation learning for diagnosing unknown defective categories in planetary gearboxes. *Knowledge-Based Systems*, *309*, 112863.

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597–1607). PMLR.

Cheng, J., Wen, G., He, X., Liu, X., Hu, Y., & Mei, S. (2023). Achieving the defect transfer detection of semiconductor wafer by a novel prototype learning based semantic segmentation network. *IEEE Transactions on Instrumentation and Measurement*, *73*, (pp. 1–12).

Gamage, U. K., Zanatta, L., Fumagalli, M., Cadena, C., & Tolu, S. (2023). Event-based classification of defects in civil infrastructures with artificial and spiking neural networks. In *International work-conference on artificial neural networks* (pp. 629–640). Springer.

Garg, A. (2023). The dynamic vision sensor stereo vision calibration inspired by biology. In *2023 2nd international conference on futuristic technologies (INCOFT)* (pp. 1–7).

Guo, A., Ma, J., Dian, R., Ma, F., Wu, J., & Li, S. (2023). Surface defect detection competition with a bio-inspired vision sensor. *National Science Review*, *10*(6), nwad130.

Han, K., Vedaldi, A., & Zisserman, A. (2019). Learning to discover novel visual categories via deep transfer clustering. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 8401–8409).

He, Y., Wu, B., Mao, J., Jiang, W., Fu, J., & Hu, S. (2024). An effective MID-based visual defect detection method for specular car body surface. *Journal of Manufacturing Systems*, *72*, 154–162.

Kahraman, Y., & Durmuşoğlu, A. (2023). Deep learning-based fabric defect detection: A review. *Textile Research Journal*, *93*(5–6), 1485–1503.

Kim, S., Park, S., Na, B., & Yoon, S. (2020). Spiking-yolo: Spiking neural network for energy-efficient object detection. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 11270–11277). (vol. *34*).

Li, W., Fan, Z., Huo, J., & Gao, Y. (2023a). Modeling inter-class and intra-class constraints in novel class discovery. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3449–3458).

Li, W., Zhang, H., Wang, G., Xiong, G., Zhao, M., Li, G., & Li, R. (2023b). Deep learning based online metallic surface defect detection method for wire and arc additive manufacturing. *Robotics and Computer-Integrated Manufacturing*, *80*, 102470.

Liu, B., Xu, C., Yang, W., Yu, H., & Yu, L. (2023). Motion robust high-speed light-weighted object detection with event camera. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1–13.

Liu, J., Wang, Y., Zhang, T., Fan, Y., Yang, Q., & Shao, J. (2023). Open-world Semi-supervised Novel Class Discovery. arXiv e-prints, arXiv:2305.13095.

Manivannan, S. (2023). Collaborative deep semi-supervised learning with knowledge distillation for surface defect classification. *Computers & Industrial Engineering*, *186*, 109766.

Messikommer, N., Fang, C., Gehrig, M., & Scaramuzza, D. (2023). Data-driven feature tracking for event cameras. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5642–5651).

Mezher, A. M., & Marble, A. E. (2024). Computer vision defect detection on unseen backgrounds for manufacturing inspection. *Expert Systems with Applications*, *243*, 122749.

Ozdemir, R., & Koc, M. (2024). On the enhancement of semi-supervised deep learning-based railway defect detection using pseudo-labels. *Expert Systems with Applications*, *251*, 124105.

Park, J.-E., & Kim, Y.-K. (2025). Semi-supervised learning for steel surface inspection using magnetic flux leakage signal. *Journal of Intelligent Manufacturing*, *36*(2), (pp. 1021–1031).

Sampath, V., Maurtua, I., Martín, J. J. A., Rivera, A., Molina, J., & Gutierrez, A. (2023). Attention-guided multitask learning for surface defect identification. *IEEE Transactions on Industrial Informatics*, *19*(9), 9713–9721.

Schaefer, S., Gehrig, D., & Scaramuzza, D. (2022). Aegnn: Asynchronous event-based graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12371–12381).

Snyder, S., Thompson, H., Kaiser, M. A.-A., Schwartz, G., Jaiswal, A., & Parsa, M. (2023). Object motion sensitivity: A bio-inspired solution to the ego-motion problem for event-based cameras. arXiv preprint arXiv:2303.14114

Sun, X., Tu, L., Zhang, J., Cai, J., Li, B., & Wang, Y. (2023). Assbert: Active and semi-supervised bert for smart contract vulnerability detection. *Journal of Information Security and Applications*, *73*, 103423.

Wan, Z., Dai, Y., & Mao, Y. (2022). Learning dense and continuous optical flow from an event camera. *IEEE Transactions on Image Processing*, *31*, 7237–7251.

Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2023). Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7464–7475).

Xiao, R., Feng, L., Tang, K., Zhao, J., Li, Y., Chen, G., & Wang, H. (2024). Targeted representation alignment for open-world semi-supervised learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 23072–23082).

Xie, B., Deng, Y., Shao, Z., Liu, H., & Li, Y. (2022). Vmv-gcn: Volumetric multi-view based graph cnn for event stream classification. *IEEE Robotics and Automation Letters*, *7*(2), 1976–1983.

Xin, G., Chen, Y., Li, L., Chen, C., Liu, Z., & Antoni, J. (2025). Complex symplectic geometry mode decomposition and a novel time-frequency fault feature extraction method. *IEEE Transactions on Instrumentation and Measurement*, *74*, (pp. 1–10).

Yang, Y., Pan, L., & Liu, L. (2023). Event camera data pre-training. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10699–10709).

Zhong, X., Yu, Z., & Gu, X. (2024). Recent advances in bio-inspired vision sensor: A review. *Journal of Circuits, Systems and Computers*, *33*(16), (p. 2430008).

Zou, Y., Fu, Y., Takatani, T., & Zheng, Y. (2025). EventHDR: From event to high-speed HDR videos and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *47*(1), 32–50.