Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Lightweight Dual-Path Network with Dynamic Distillation for Industrial LED Defect Recognition

Jun Deng
Guangdong University of Science and Technology
Jiaming Zhong
Guangdong University of Science and Technology
Bingchun Jiang
Guangdong University of Science and Technology
Yanping Li
Guangdong University of Science and Technology
Bangquan xie
Great Bay University
Chaoen Tan
Neusoft Institute Guangdong
Qinchuan Lei
Guangdong University of Science and Technology
Nannan Li
nnli@must.edu.mo
Macau University of Science and Technology
Fuqin Deng
Wuyi University
Min Zhao
Guangdong University of Science and Technology

Article

Keywords: Led defects recognition, Lightweight network, Dynamic classifier, Dual-Level Distillation

Posted Date: August 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10557038/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Lightweight Dual-Path Network with Dynamic Distillation for Industrial LED Defect Recognition

Jun Deng[1], Jiaming Zhong[1], Bingchun Jiang[1], Yanping Li[1], Bangquan Xie[4], Chaoen Tan[5], Qinchuan Lei[1], Nannan Li[2†], Fuqin Deng[3,4†], Min Zhao[1]

***Abstract*—Defect recognition is critical in LED manufacturing, but it faces challenges including imbalanced data, limited edge computing resources, and fine-grained defect similarity. This work proposes a lightweight dual-path network with dynamic distillation for efficient LED defect recognition. The method includes a dynamic dual-path classifier with adaptive hierarchical fusion and dynamic logits fusion to enhance fine-grained discrimination. A dual-level distillation scheme transfers feature and logits knowledge from teacher to student networks. Experiments on a self-built LED-Defect-6K dataset and IFLYTEK 2022 LED dataset show 97.17% and 96.31% accuracy, demonstrating strong performance and generalization for industrial imbalanced scenarios. The source code and dataset will be released at https://github.com/JiaMingZhong2621/LED.**

***Index Terms*—Led defects recognition, Lightweight network, Dynamic classifier, Dual-Level Distillation.**

I. Introduction

LED technology has been extensively adopted in modern manufacturing and display industries. Nevertheless, surface defects such as bubble defects, loose pins, and glue overflow can significantly compromise product reliability and result in substantial economic losses [1], [2]. These issues highlight the necessity of developing accurate and efficient automated LED defect recognition systems.

Early LED defect recognition approaches primarily relied on template matching or handcrafted-feature-based machine learning methods. Template matching algorithms measure similarity between a reference template and an input image [3]–[5]. Although computationally efficient, their performance deteriorates under noise, illumination changes, positional deviations, and other variations inherent to real manufacturing environments. Handcrafted-feature-based models extract statistical or morphological descriptors and classify them using traditional classifiers such as SVMs [6]. However, their generalization capability is heavily constrained by the quality of manually designed features, making them insufficient for modeling diverse and fine-grained defect patterns.

The emergence of deep learning has substantially advanced automated defect recognition. Chen et al. [7] introduced SGL-FSPNet, a lightweight transformer-based framework that improves fabric defect detection through global–local feature interaction and efficient multi-scale feature aggregation. Zhang et al. [8] proposed MCAM-Net, a multi-scale convolutional attention network that employs channel-prioritized multi-scale subspace attention and multi-scale convolutional blocks to enhance low-contrast defect detection. Li et al. [9] proposed RetiNeXt, a lightweight segmentation network that combines frequency-domain enhancement and global–local feature modeling to improve the representation of fine structures. Chen et al. [10] proposed LBG-YOLO, integrating lightweight multiscale convolutions and a bidirectional feature pyramid network. Zhong et al. [11] developed a YOLO-based framework for Micro LED recognition and incorporated a class-balanced loss to address long-tailed distributions. Despite these advancements, significant challenges remain in practical industrial settings. Recognition systems deployed on production lines must process high-throughput image streams under strict latency and hardware constraints, while edge devices often operate with limited computational capacity. Existing LED defect recognition approaches [10]–[15] typically rely on CNN, YOLO, or ViT-based architectures, many of which are computationally intensive and therefore unsuitable for embedded deployment. Although lightweight models offer reduced computational overhead, they often suffer from degraded feature representation and limited discriminative capacity, especially when dealing with visually ambiguous or imbalanced defect categories. Consequently, designing a lightweight yet highly discriminative model remains a critical challenge.

Knowledge distillation (KD) provides an effective mechanism for enhancing lightweight models by transferring knowledge from high-capacity teacher networks to compact student models [16]–[19]. For example, Yi et al. [20] developed a specialized insulator-defect segmentation dataset and proposed a double knowledge-distillation framework to achieve superior detection performance using a lightweight model. Liu et al. [21] introduced a self-distillation strategy to transfer knowledge within the model itself and designed an improved KL-divergence loss to enhance the accuracy of the student network. Existing KD approaches can generally be categorized into feature-based and logit-based methods. Feature-

[1]College of Mechanical and Electrical Engineering, Guangdong University of Science and Technology, Dongguan 523000, China. [2]School of Computer Science and Engineering, Faculty of Innovation Engineering, Macau University of Science and Technology, Macau, China. [3]School of Electronics and Information Engineering, the Wuyi University, Jiangmen, China. [4]College of Advanced Engineering, Great Bay University, Dongguan 523000, Guangdong, China. [5]College of Artificial Intelligence, Neusoft Institute Guangdong, Foshan, Guangdong, China. †Corresponding authors: Nannan Li (email: nnli@must.edu.mo); Fuqin Deng (email: 519726155@qq.com).

This work was partially supported by the National Natural Science Foundation of China (Grant No. 62073274), the Design and Development of Non-contact Wafer Handling Manipulator (No. GKY-2025BSQDK-17), the Wuyi University-Hong Kong-Macau Joint Funding Scheme (Grant Nos. 2022WGALH17, 2021WGALH18), the Guangdong Province Key Construction Discipline Research Ability Enhancement Project (Grant No. 2024ZDJS071), and the Key Fields Special Project for Ordinary Universities in Guangdong Province (Grant No. 2024ZDZX2083).

based methods focus on representation alignment, whereas logit-based methods emphasize the transfer of class-relation information encoded in teacher predictions. Although KD has shown promising progress in various defect detection and segmentation tasks, its application in LED defect recognition remains largely unexplored. Moreover, existing methods typically exploit only a single form of supervision, which may limit their ability to simultaneously enhance representation quality and decision consistency in challenging defect recognition scenarios.

Beyond model compression, learning discriminative feature representations is equally important for defect recognition. Wang et al. [22] proposed CoDeR, a prototype-guided regularization framework that improves fine-grained visual classification by enforcing inter-class separation and guiding hierarchical representation learning. However, such representation-learning strategies have rarely been investigated for industrial LED defect recognition.

In addition, LED defect recognition is complicated by substantial intra-class variability and pronounced inter-class similarity. Variations in material properties, manufacturing conditions, and environmental factors introduce large appearance differences within the same defect category. At the same time, several defect types share highly similar visual patterns, which increases the likelihood of confusion during classification. The limited availability of large-scale annotated datasets further restricts the model's ability to learn robust representations, particularly for rare defect categories.

Motivated by the limitations of existing lightweight defect recognition methods and the challenges posed by highly imbalanced and visually ambiguous LED defect data, we propose a lightweight dual-path network with dynamic distillation. The framework comprises two complementary components. A Dynamic Dual-Path Classifier performs multi-level feature aggregation and adaptively balances the contributions of the two paths to learn more discriminative representations. Meanwhile, a Dual-Level Distillation scheme transfers both feature-level and logit-level knowledge from a high-capacity teacher encoder to a compact student encoder, enabling the student to achieve strong recognition performance with substantially lower computational cost. To facilitate rigorous evaluation, we further construct a real-world LED defect dataset containing eight highly imbalanced defect categories.

The key contributions of this work are summarized as follows:

- **Lightweight dual-path architecture.** The proposed method integrates an Adaptive Hierarchical Fusion module for multi-level feature aggregation and a Dynamic Dual-Path Classifier with adaptive logits fusion. This design enhances fine-grained discriminability, effectively reduces inter-class confusion, and maintains high computational efficiency.
- **Dual-level dynamic distillation strategy.** A two-stage dynamic distillation mechanism is introduced to transfer both feature-level and logit-level knowledge from a high-capacity teacher encoder to a lightweight student encoder. By selectively emphasizing informative representations, the proposed method improves robustness to visually similar, rare, and hard-to-recognize defect categories.
- **LED-Defect-6K dataset.** A real-world dataset consisting of 6,024 LED images across eight defect categories is constructed. The dataset exhibits a highly imbalanced distribution and enables comprehensive evaluation of the proposed method under realistic industrial conditions.

The remainder of this paper is organized as follows. Section II presents the overall framework and describes the key architectural components of the proposed approach. Section III introduces the datasets, evaluation metrics, and implementation details, and reports both comparative and ablation results. Section IV summarizes the main findings of this work.

II. Methods

The overall framework of the proposed method is shown in Fig. 1, which consists of two main components: Dual-Level Distillation and the Dynamic Dual-Path Classifier. The initially collected LED images x_i are used as inputs to the model. Each image x_i is fed into both the student and teacher encoders, which independently extract feature representations denoted as z_i^s and z_i^t, respectively. Next, the AHF module operates on the student feature z_i^s to extract multi-level representations and enhance the discriminative capability of the learned features. Finally, in the DLF module, the MLP logits h_i^s and the AHF logits h_i^{AHF} are fused in a learnable manner to model inter-class confusion and produce the final logits h_i^{final}, where the fusion weight β is generated by a dedicated β Generator. To transfer prior knowledge from the teacher and jointly optimize the learnable parameters in both DDPC and DLD, multiple losses are employed, including the focal cross-entropy loss $\mathcal{L}_{FCE}$ between h_i^{final} and the label y_i, the contrastive loss $\mathcal{L}_{NCE}$ between z_i^e and z_i^t, and the KL divergence loss $\mathcal{L}_{KLD}$.

A. Dynamic Dual-Path Classifier

To satisfy the computational requirements of industrial deployment, we adopt EfficientNet-B0 [23] as the student encoder. Owing to its compound-scaling design, EfficientNet-B0 achieves an excellent balance between accuracy and efficiency, enabling the proposed framework to learn discriminative LED defect representations while maintaining low computational overhead.

Adaptive Hierarchical Fusion. To fully exploit the diversity of multi-level features extracted by the image encoder $\mathcal{V}$, the proposed AHF module transforms and fuses features from different depths. This allows AHF to preserve fine-grained information from low-level features while effectively incorporating the semantic abstraction of high-level features. As illustrated in Fig. 1 and Fig. 2, AHF contains multiple Feature Alignment Modules (FAM), each consisting of a 1×1 projection layer, a normalization block, a nonlinear activation, and an adaptive pooling operator. Three levels of features, f_i^1, f_i^2, and f_i^3, are first extracted from $\mathcal{V}$. These features are then transformed by identical FAM to obtain z_i^1, z_i^2, and z_i^3. The resulting features are fused through a weighted fusion

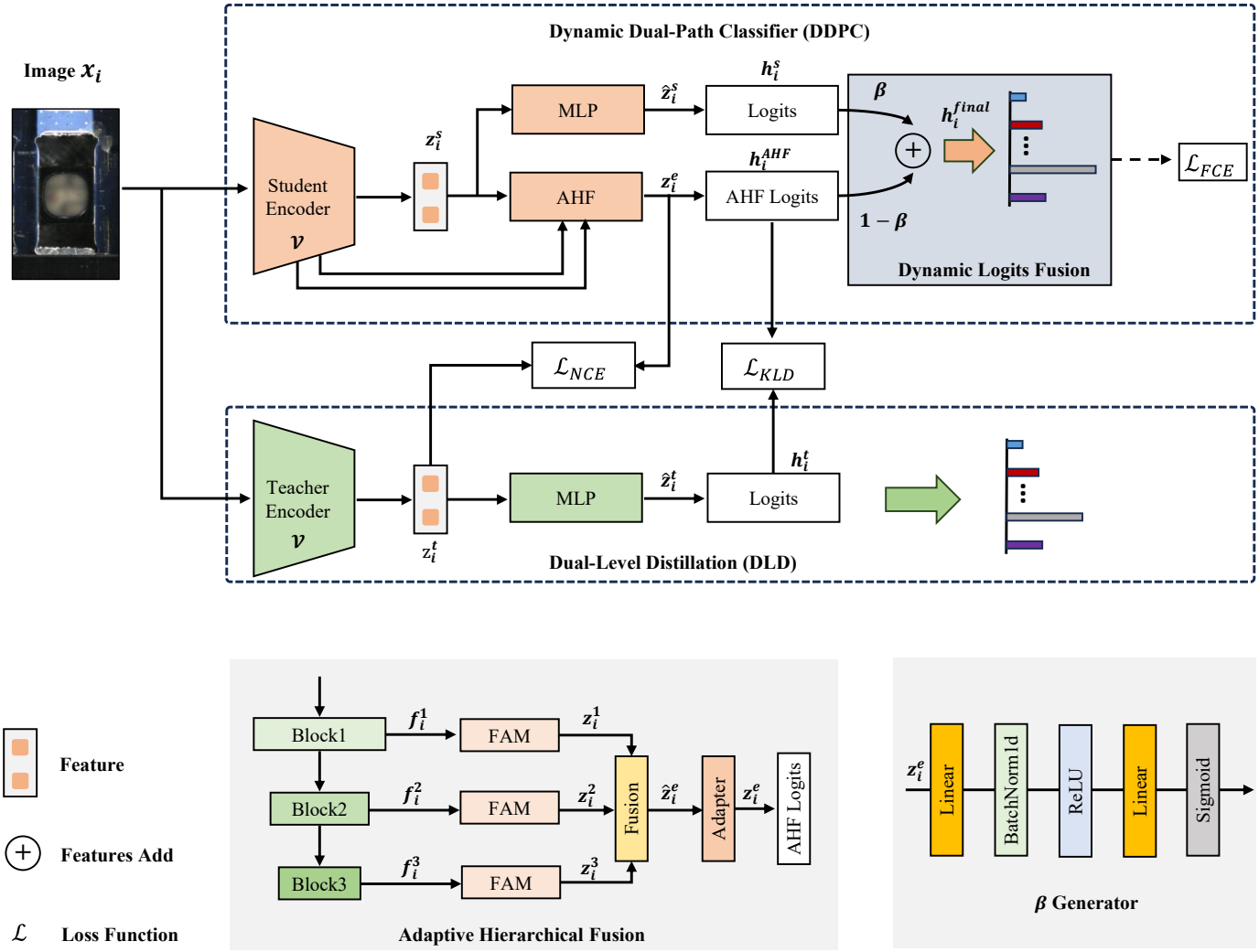


Fig. 1. Overview of the proposed framework, which comprises two key components: Dual-Level Distillation and the Dynamic Dual-Path Classifier. The model is optimized using the focal cross-entropy loss $\mathcal{L} * FCE$, contrastive loss $\mathcal{L} * NCE$, and KL-divergence loss $\mathcal{L}_{KLD}$. In Dynamic Logits Fusion, the β Generator predicts an adaptive weight β to fuse the logits h_i^s and h_i^{AHF}. Details of the Adaptive Hierarchical Fusion module are provided in Section II-A.

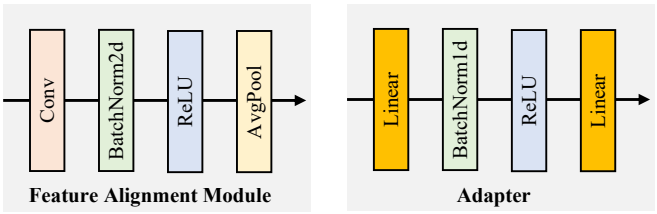


Fig. 2. Details of the FAM and Adapter.

mechanism to produce a fused representation $\hat{z}_i^e$. Finally, an enhanced feature z_i^e is obtained through the Adapter module, whose structure is shown in Fig. 2. The image encoder follows the EfficientNet-B0 architecture.

Dynamic Logits Fusion. To mitigate inter-class confusion, a Dynamic Logits Fusion (DLF) module is introduced to adaptively integrate the MLP logits h_i^s with the AHF logits h_i^{AHF}. A learnable β Generator produces a dynamic fusion weight β to balance the contributions of the two branches. The MLP branch consists of two linear layers, while the AHF branch provides multi-level aggregated features. The final fused logits are computed as:

$$h_i^{final} = \beta h_i^s + (1-\beta) h_i^{AHF}, \quad \beta = G_\beta(z_i^e), \tag{1}$$

where G_β denotes the β Generator, composed of two linear layers, a batch normalization block, a ReLU activation, and a sigmoid gating function.

By introducing the Dynamic Dual-Path Classifier, the method adaptively adjusts the relative contributions of the two branches for each input sample. This dynamic weighting mechanism improves the model's ability to resolve inter-class ambiguity and produces more robust and discriminative logits.

B. Dual-Level Distillation

To enable accurate LED defect recognition under resource-constrained industrial deployments, a Dual-Level Distillation scheme is introduced to transfer the discriminative capability of a high-capacity teacher model to a lightweight student network. This scheme jointly preserves high-level semantic decision information and low-level feature representations, allowing the student to achieve competitive accuracy while maintaining computational efficiency.

Logit-Level Distillation. To align the student's predictive distribution with that of the teacher, a temperature-scaled KL divergence is employed [16]:

$$\mathcal{L}_{KLD} = \mathrm{KL}\left(\mathrm{softmax}\left(\frac{h_i^t}{T}\right) \parallel \mathrm{softmax}\left(\frac{h_i^{AHF}}{T}\right)\right) T^2, \tag{2}$$

where T denotes the temperature coefficient. This objective encourages the student to capture the teacher's softened class relationships, enabling richer modeling of inter-class similarity than is possible with hard labels alone.

Feature-Level Distillation. To further ensure that the student inherits the teacher's representation ability, an InfoNCE-based contrastive loss is adopted [24]:

$$\mathcal{L}_{\mathrm{NCE}} = -\log \frac{\exp\big(\mathrm{sim}(z_i^e, z_i^t)/\tau\big)}{\sum_{j=1}^{B} \exp\big(\mathrm{sim}(z_i^e, z_j^t)/\tau\big)}, \tag{3}$$

where $\mathrm{sim}(\cdot)$ denotes cosine similarity and τ is a learnable temperature factor. This formulation aligns each student feature z_i^e with its corresponding teacher feature z_i^t, while simultaneously enforcing separation from non-matching teacher features, thereby improving feature discriminability.

C. Training Objective

During optimization, three complementary loss functions are employed to jointly supervise the training process: the focal loss $\mathcal{L}_{FCE}$, the contrastive loss $\mathcal{L}_{NCE}$, and the logit-level distillation loss $\mathcal{L}_{KLD}$. Together, these objectives guide the student network to learn both accurate decision boundaries and discriminative feature representations. The overall training objective is formulated as

$$\mathcal{L} = \mathcal{L}_{FCE}\Big(h_i^{final}, y_i\Big) + \mathcal{L}_{NCE} + \mathcal{L}_{KLD}. \tag{4}$$

III. Experiments

This section presents comprehensive experimental evaluations designed to assess the effectiveness of the proposed method and to quantify the contribution of each component to the overall performance.

A. Datasets

In this paper, two LED defect datasets from different industrial scenarios are used to evaluate the effectiveness and generalizability of the proposed method.

LED-Defect-6K dataset: The LED-Defect-6K dataset is constructed to evaluate the effectiveness of the proposed method. It contains eight defect categories: foreign object defects (FOB), excess glue (EG), excessive bright defect area (EBDA), bubble defects (BB), missing glue or material (MGM), glue overflow (GO), loose pin (LP), and pin sticking with glue (PSG). The dataset consists of 6,024 images with a resolution of 224×411 pixels and is divided into 4,214 training images, 1,203 validation images, and 607 testing images.

IFLYTEK 2022 LED dataset [25]: This open-source dataset was released by IFLYTEK in 2022 for LED chip production and packaging defect recognition. The images are captured under multi-angle illumination, including two low-angle lights, one ultraviolet light, and one ring light. The dataset contains four categories: normal (NO), glue-related defects (EG), bubble defects (BB), and foreign object defects (FOB). A total of 4,349 images are provided, split into 3,043 training images, 868 validation images, and 438 testing images.

Fig. 5 presents the category distributions of the two datasets. As shown, the datasets exhibit markedly different imbalance characteristics. The IFLYTEK 2022 LED dataset follows a pronounced long-tailed distribution: the NO class overwhelmingly dominates, whereas most defect categories contain only a limited number of samples. In contrast, the LED-Defect-6K dataset displays a moderately imbalanced multi-class distribution—categories such as PSG, LP, GO, BB, and EG appear with comparable frequencies, while tail classes (e.g., MGM) contain substantially fewer samples.

Beyond class imbalance, both datasets also exhibit significant visual ambiguity, as illustrated in Fig. 3 and Fig. 4. Several defect types show strong inter-class similarity (e.g., BB vs. GO, LP vs. PSG), which increases the likelihood of misclassification. Meanwhile, certain categories suffer from large intra-class variance, further complicating the development of a robust defect recognition model.

B. Evaluation Metrics

Given that LED defect recognition is a multi-class classification task, multiple quantitative metrics are employed to comprehensively assess the performance of the proposed method. These include overall classification accuracy (Acc), precision (P), recall (R), F1-score (F1), and the model parameter count (Params).

Among these metrics, Params is used to evaluate the computational complexity of the model, where lower values indicate better suitability for deployment on devices with limited computational resources. In contrast, higher values of Acc, P, R, and F1 reflect stronger recognition performance and more reliable classification capability [26], [27]. The metrics are defined as follows:

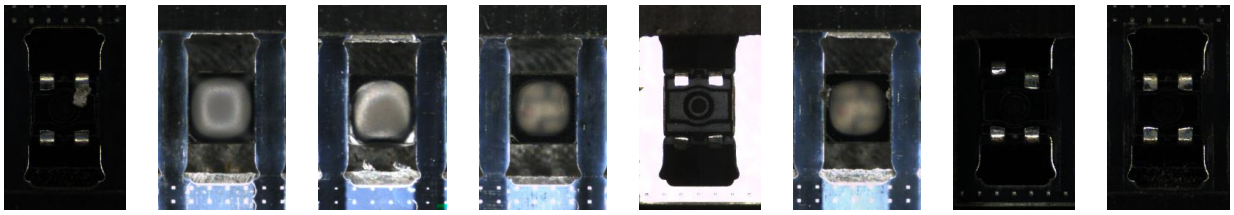


Fig. 3. Examples from the LED-Defect-6K dataset.

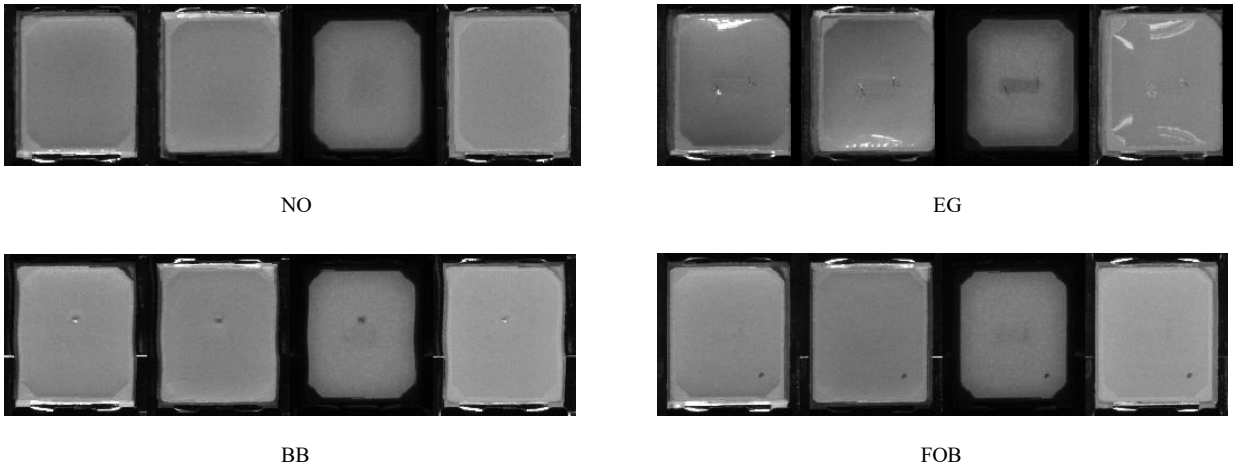


Fig. 4. Examples from the IFLYTEK 2022 LED defect dataset.

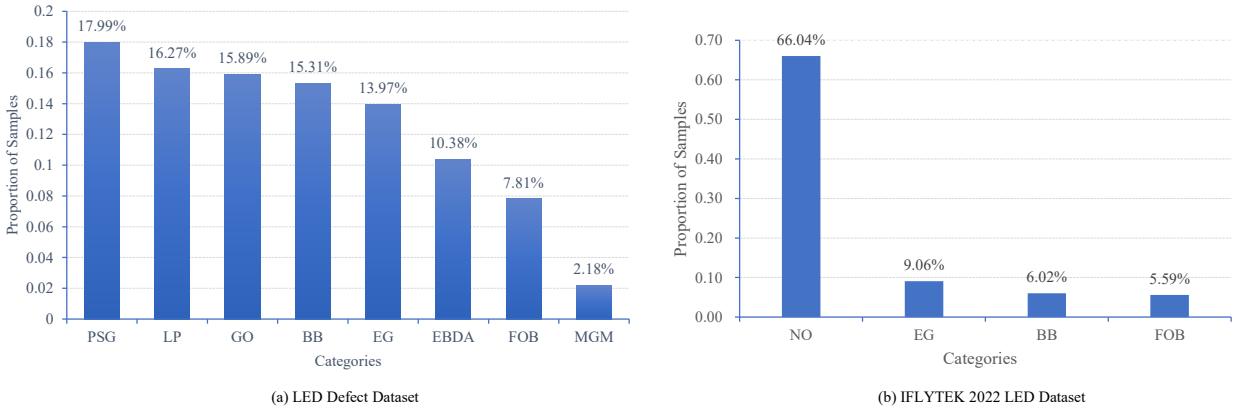


Fig. 5. Category distribution of the LED-Defect-6K dataset and the IFLYTEK 2022 LED defect dataset, illustrating the class imbalance issue.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, mAcc = \frac{1}{C}\sum_{c=1}^{C} Acc_c,$$
$$P = \frac{TP}{TP + FP}, mP = \frac{1}{C}\sum_{c=1}^{C} P_c,$$
$$R = \frac{TP}{TP + FN}, mR = \frac{1}{C}\sum_{c=1}^{C} R_c,$$
$$F1 = \frac{2 \times P \times R}{P + R}, mF1 = \frac{2 \times mP \times mR}{mP + mR}. \quad (5)$$

where TP and TN denote the numbers of true positive and true negative samples, respectively; FP represents the number of false positive samples, and FN denotes the number of false negative samples. C is the total number of LED defect categories.

C. Implementation Details

All experiments are conducted using the PyTorch framework [35]. The Adam optimizer [36] is adopted with a learning rate of 1×10^{-4} and a weight decay of 1×10^{-4}. The models are trained for 50 epochs with a batch size of 16,

TABLE I
RESULTS OF THE MACC(%), MP(%), MR(%), MF1(%) SCORES OF DIFFERENT MODELS ON THE LED-DEFECT-6K DATASET.

Method	mAcc(%)	mP(%)	mR(%)	mF1(%)	params(M)
ResNet18 [28]	96.51	96.69	95.79	96.14	11.69
CLIP-RN50 [29]	96.41	96.77	95.89	96.25	38.32
RegNetX-400MF [30]	95.61	95.34	94.89	95.03	5.10
ShuffleNet-V2 [31]	96.17	96.22	95.17	95.60	1.26
EfficientNet-V2 [32]	96.84	97.19	96.10	96.63	20.19
EfficientNet-B0 [23]	96.15	96.09	94.75	95.37	4.02
CARE-S0 [33]	86.82	87.47	86.82	86.66	7.3
MicroViTv2 [34]	91.83	92.01	91.83	91.68	6.4
Proposed Model	**97.17**	**97.59**	**96.46**	**96.93**	5.41

TABLE II
RESULTS OF THE MACC(%), MP(%), MR(%), MF1(%) SCORES OF DIFFERENT MODELS ON THE IFLYTEK 2022 LED DEFECT DATASET.

Method	mAcc(%)	mP(%)	mR(%)	mF1(%)	params(M)
RegNetX-400MF [30]	92.74	87.41	76.16	78.60	5.10
ShuffleNet-V2 [31]	95.16	93.71	85.05	88.32	1.26
EfficientNet-B0 [23]	95.02	93.05	85.80	88.61	4.02
CARE-S0 [33]	86.75	82.12	86.76	83.38	7.3
MicroViTv2 [34]	91.32	85.80	91.32	88.33	6.4
Proposed Model	**96.31**	**95.53**	**89.11**	**91.94**	5.41

TABLE III
RESULTS OF THE P(%), R(%), F1(%) SCORES OF DIFFERENT MODELS ON THE LED-DEFECT-6K DATASET.

Class	RegNetX_400MF			ShuffleNet_V2			EfficientNet_B0			Proposed method		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
FOB	87.80	85.71	86.75	90.70	**92.86**	91.76	94.59	83.33	88.61	**100.00**	90.48	**95.00**
EG	**95.89**	94.59	95.24	94.74	97.30	96.00	93.42	95.95	94.67	93.59	**98.65**	**96.05**
EBDA	94.83	**100.00**	97.35	98.18	98.18	98.18	94.83	**100.00**	97.35	**100.00**	**100.00**	**100.00**
BB	93.98	**97.50**	95.71	96.30	**97.50**	96.89	96.25	96.25	96.25	**98.73**	**97.50**	**98.11**
MGM	92.86	**100.00**	**96.30**	92.31	92.31	92.31	**100.00**	92.31	96.00	**100.00**	92.31	96.00
GO	98.70	91.57	95.00	98.72	92.77	95.65	**100.00**	92.77	96.25	**100.00**	**95.18**	**97.53**
LP	96.43	95.29	95.86	**96.47**	96.47	**96.47**	93.33	**98.82**	96.00	94.32	97.65	95.95
PSG	96.84	97.87	97.35	**100.00**	**98.94**	**99.47**	96.84	97.87	97.35	96.88	**98.94**	97.89
Avg	94.67	95.32	94.95	95.93	95.79	95.84	96.16	94.66	95.31	**97.94**	**96.34**	**97.07**

and five-fold cross-validation is employed to ensure reliable performance evaluation. To improve generalization, standard data augmentation techniques are applied during training, including random cropping, horizontal flipping, and center cropping. In particular, random 224×224 crops are generated for each input image. All experiments are performed on a single NVIDIA RTX 3060 GPU.

D. Performance Comparison

In this section, the proposed method is evaluated on two previously unseen datasets and compared with several state-of-the-art lightweight models to assess its effectiveness and generalizability. Table I and Table II summarize the results on the LED-Defect-6K dataset and the IFLYTEK 2022 LED dataset, respectively, in terms of four performance metrics: mAcc, mP, mR, and mF1.

For the LED-Defect-6K dataset, the proposed method achieves mAcc, mP, mR, and mF1 of 97.17%, 97.59 96.46%, and 96.93%, outperforming the strongest SOTA competitor by 0.33%, 0.40%, 0.36%, and 0.30%, respectively. Despite its performance advantages, the proposed model is highly compact and contains only 5.41M parameters, which is approximately 26.8% of the parameter size of the strongest competing method. Compared with the baseline EfficientNet-B0, it further achieves performance gains of 1.02%, 1.50%, 1.71%, and 1.56% across the four evaluation metrics.

On the IFLYTEK 2022 LED dataset, the proposed method attains 96.50% mAcc, 95.53% mP, 89.11% mR, and 91.94% mF1, surpassing the best-performing state-of-the-art model by 1.15%, 1.82%, 4.06%, and 3.62%, respectively. The substantial improvement in mR (+4.06%) demonstrates the method's strong capability in addressing severe category imbalance. This gain is particularly noteworthy because many SOTA models were originally designed for large-scale datasets with long-tailed distributions, when applied to small-scale, highly imbalanced industrial LED defect datasets, these models often fail to learn stable and discriminative feature representations. In contrast, the proposed method enhances both feature robustness and decision reliability, enabling consistent improvements across diverse and imbalanced scenarios.

TABLE IV

COMPARISON BETWEEN EFFICIENTNET-B0 (BASELINE) AND THE PROPOSED METHOD ON THE LED-DEFECT-6K DATASET UNDER 5-FOLD CROSS-VALIDATION. BOLD VALUES DENOTE THE BEST PERFORMANCE, AND THE NUMBERS IN PARENTHESES INDICATE THE IMPROVEMENTS OVER THE BASELINE (MARKED IN GREEN).

Fold	Acc(%)	P(%)	R(%)	F1(%)
Fold 1	96.05 → **97.20** (+1.15)	96.47 → **97.42** (+0.95)	95.62 → **96.89** (+1.27)	95.94 → **97.08** (+1.14)
Fold 2	96.87 → **97.20** (+0.33)	96.50 → **97.54** (+1.04)	95.76 → **96.98** (+1.22)	96.04 → **97.19** (+1.15)
Fold 3	96.05 → **96.87** (+0.82)	96.34 → **97.42** (+1.08)	93.70 → **95.79** (+2.09)	94.75 → **96.47** (+1.72)
Fold 4	95.55 → **97.03** (+1.48)	95.68 → **97.55** (+1.87)	93.41 → **95.90** (+2.49)	94.37 → **96.60** (+2.23)
Fold 5	96.21 → **97.53** (+1.32)	96.45 → **98.03** (+1.58)	95.26 → **96.74** (+1.48)	95.76 → **97.33** (+1.57)
Mean	96.15 → **97.17** (+1.02)	96.09 → **97.59** (+1.50)	94.75 → **96.46** (+1.71)	95.37 → **96.93** (+1.56)

TABLE V

ABLATION EXPERIMENT RESULTS OF EACH MODULE ON THE LED-DEFECT-6K DATASET AND THE IFLYTEK 2022 LED DATASET.

Method	LED-Defect-6K dataset				IFLYTEK 2022 LED dataset			
	mAcc(%)	mP(%)	mR(%)	mF1(%)	mAcc(%)	mP(%)	mR(%)	mF1(%)
EfficientNet-B0	96.15	96.09	94.75	95.37	95.02	93.05	85.80	88.61
EfficientNet-B0 + DDPC	96.64	96.99	95.28	96.01	95.62	94.98	86.91	90.16
EfficientNet-B0 + DDPC + DLD	**97.17**	**97.59**	**96.46**	**96.93**	**96.31**	**95.53**	**89.11**	**91.94**

TABLE VI

ABLATION STUDY EVALUATING THE EFFECTS OF DLD AND DDPC UNDER DIFFERENT TEACHER–STUDENT CONFIGURATIONS. THE BASELINE USES EFFICIENTNET-B0 AS THE STUDENT ENCODER WITHOUT DLD OR DDPC.

Method	**Acc (%)**	**P (%)**	**R (%)**	**F1 (%)**
EfficientNet-B0 (Baseline)	96.15	96.09	94.75	95.37
Proposed Model (Teacher Encoder: RN50)	96.97	97.34	96.30	96.70
Proposed Model (Teacher Encoder: EfficientNet-V2)	97.17	97.59	96.46	96.93

TABLE VII

RESULTS OF COMPLEXITY ANALYSIS OF DIFFERENT MODELS.

Model	RegNetX_400MF	ShuffleNet_V2	EfficientNet_B0	CARE-S0	Proposed Model
Params (M)	5.10	1.26	4.01	5.30	5.40
FLOPs (G)	0.43	0.15	0.41	0.72	0.80
FPS	366.95	401.67	334.00	276.81	300.33

Tables III report the precision, recall, and F1-score of different models for each defect category in the LED-Defect-6K dataset. Compared with lightweight state-of-the-art models, the proposed method achieves the highest precision in multiple categories, including FOB, EBDA, MGM, and BB. In particular, for the FOB defect, the method attains the highest F1-score of 95%, representing a 6.39% improvement over EfficientNet-B0.

As illustrated in Fig. 5, MGM and FOB belong to rare categories, while BB and GO represent categories prone to inter-class confusion. The results indicate that the proposed DLD and DDPC modules substantially strengthen the capability to recognize both rare categories and visually similar categories, effectively addressing the issues of imbalance in data distribution and inter-class similarity. Moreover, the proposed method achieves the highest average precision, recall, and F1-score across all compared models on the LED-Defect-6K dataset, demonstrating superior overall performance and enhanced generalization capability for LED defect recognition.

Furthermore, Table IV reports the results of the 5-fold cross-validation experiments on the LED-Defect-6K dataset, comparing the proposed method with the EfficientNet-B0 baseline. Across all folds, the proposed model consistently attains higher accuracy, precision, recall, and F1-score. The performance gains remain stable, typically ranging from 0.3% to 2.5% depending on the fold and metric. Notably, several folds exhibit pronounced improvements in recall, highlighting the ability of the proposed method to correctly identify minority and visually ambiguous defect categories. The averaged results further verify its consistent advantage, showing improvements of 1.02%, 1.50%, 1.71%, and 1.56% in accuracy, precision, recall, and F1-score, respectively, compared with the baseline.

E. Ablation Experiments

To quantify the contribution of each component in the proposed framework, we perform a series of ablation studies on the LED-Defect-6K and IFLYTEK 2022 LED datasets. Unless otherwise stated, EfficientNet-B0 serves as the student encoder, while EfficientNet-V2 is adopted as the teacher encoder.

Effect Analysis of Each Component. This work introduces two key components, DLD and DDPC. To quantify their indi-

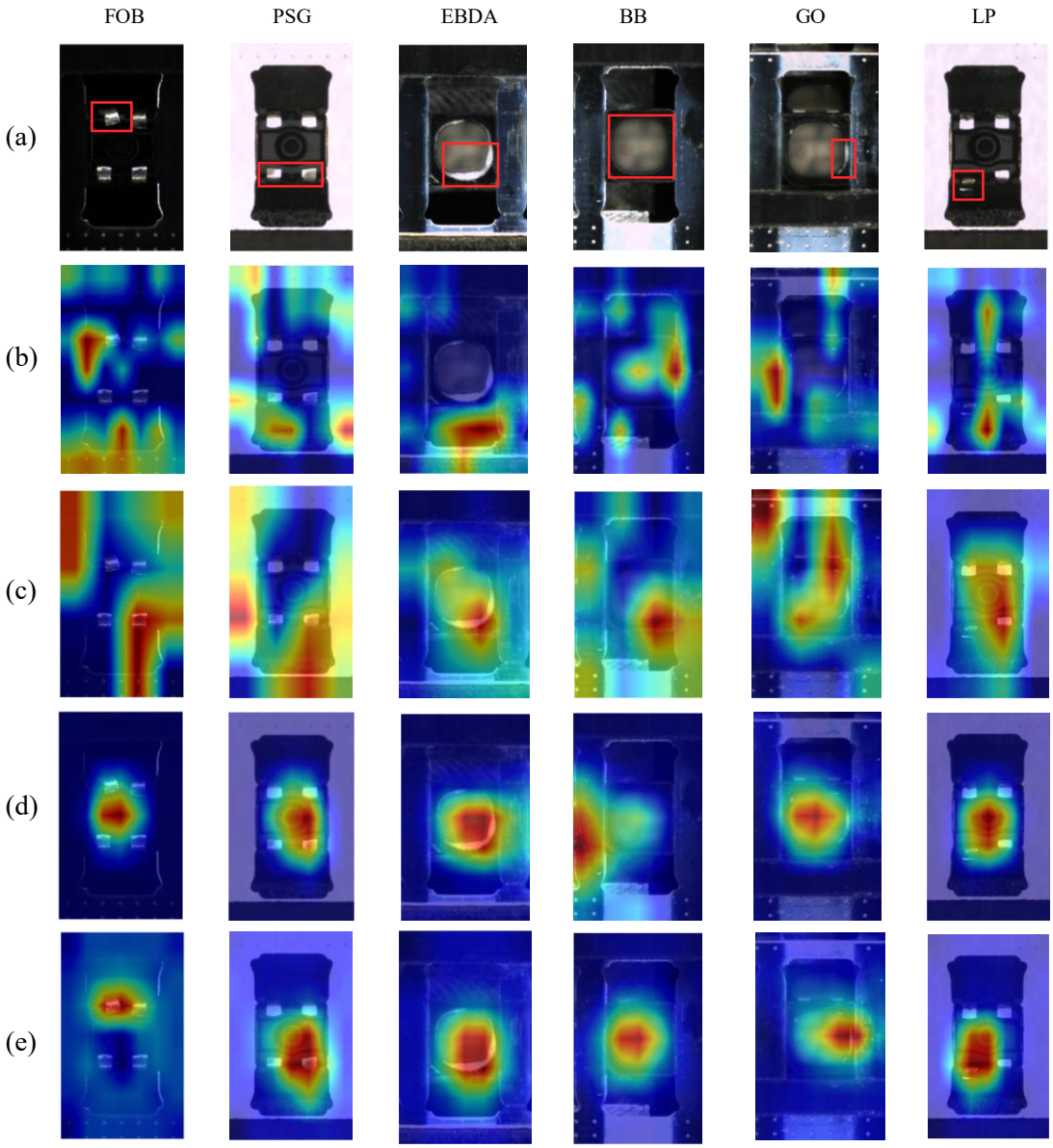


Fig. 6. Qualitative comparisons of Grad-CAM visualizations across six LED defect categories from different recognition models. (a) Original images, (b) CARE-S0, (c) MicroViTv2, (d) EfficientNet-B0, (e) Proposed model.

vidual contributions, a series of ablation studies are conducted, with results summarized in Table V. On the LED-Defect-6K dataset, incorporating DDPC alone improves mAcc, mP, mR, and mF1 by 0.49%, 0.90%, 0.53%, and 0.64%, respectively. When DLD is further integrated, the gains increase to 1.02%, 1.50%, 1.71%, and 1.56%. On the IFLYTEK 2022 LED dataset, the improvements become more pronounced. The combination of DLD and DDPC leads to increases of 1.29%, 2.48%, 3.31%, and 3.33% across the four metrics. These results demonstrate that the proposed components effectively capture inter-class confusion and alleviate class imbalance, thereby making substantial contributions to the overall performance improvements of the framework.

Influence of Different Teacher Encoder. Table VI evaluates the impact of different teacher encoders within the proposed Teacher–Student framework. The baseline configuration employs EfficientNet-B0 as the student encoder without DLD or DDPC. When RN50 is used as the teacher, consistent improvements are observed across all evaluation metrics. Replacing RN50 with EfficientNet-V2 leads to additional gains and produces the best overall performance. These results demonstrate that the proposed method is compatible with a wide range of backbone choices and can effectively benefit from stronger teacher models, underscoring its versatility and generalizability.

TABLE VIII
COMPLEXITY ANALYSIS OF THE PROPOSED METHOD.

Model	Params(M)	FLOPs(G)
EfficientNet-B0	4.01	0.41
+DLD	4.01	0.41
+DDPC	5.40	0.80
Full Model	5.40	0.80

F. Visualization Analysis

To further quantitatively evaluate the effectiveness of our proposed method, Fig. 6 (a)-(e) provides Grad-CAM visualizations derived from the final feature-extraction layer across six typical industrial component defect types. Qualitative results reveal that CARE-S0, MicroViTv2, and EfficientNet-B0 frequently struggle to pinpoint critical defective regions accurately. These baseline methods tend to erroneously activate irrelevant background pixels for FOB, PSG, LP and GO defects, leading to misleading feature attribution. By contrast, as illustrated in Fig. 6 (e), our proposed model reliably activates only discriminative fault-relevant regions. Specifically for FOB, PSG and GO samples, the network tightly concentrates its activation on genuine defective zones while effectively suppressing responses from useless background content.

Overall, the Grad-CAM visualizations show that the proposed model captures more discriminative and compact features from LED defect images, leading to more reliable recognition performance.

G. Model Complexity Analysis

To evaluate the deployment efficiency of the proposed method, we compare its computational complexity with several representative lightweight architectures, including RegNetX_400MF, ShuffleNet_V2, EfficientNet-B0, and CARE-S0. The results are reported in Table VII in terms of parameters, FLOPs, and FPS.

Compared with EfficientNet-B0, the proposed model increases the parameter count from 4.01M to 5.40M and the FLOPs from 0.41G to 0.80G. Despite this increase, the model maintains a high inference speed of 300.33 FPS, which remains suitable for real-time industrial inspection scenarios. Compared with CARE-S0, a recent lightweight industrial recognition model, the proposed method introduces only a marginal increase in computational cost while achieving significantly superior recognition performance, demonstrating a favorable accuracy-efficiency trade-off.

To further analyze the source of the additional complexity, Table VIII presents the complexity contribution of each component. Introducing the DLD module does not increase either parameters or FLOPs, indicating that the distillation strategy only affects the training stage and incurs no additional inference cost. In contrast, the DDPC module increases the parameter count from 4.01M to 5.40M and the FLOPs from 0.41G to 0.80G, showing that most of the additional computation originates from the dual-path classification architecture. Nevertheless, the overall complexity remains low compared with many modern recognition networks, while the proposed framework consistently achieves superior classification accuracy across both LED-Defect-6K and IFLYTEK LED datasets. These results demonstrate that the proposed DLD and DDPC modules effectively enhance feature discrimination and defect recognition performance with acceptable computational overhead, making the framework practical for real-world industrial deployment.

H. Discussion and Analysis

Analysis of Dual-Level Distillation. The proposed DLD transfers knowledge at both the feature and logit levels. Specifically, $\mathcal{L}_{NCE}$ aligns student features with teacher representations, encouraging semantically discriminative embeddings and reducing intra-class variation. Meanwhile, $\mathcal{L}_{KLD}$ distills the class relationships encoded in teacher predictions, enhancing the discrimination of visually similar defect categories. By jointly constraining the representation and decision spaces, DLD improves intra-class compactness and inter-class separability, which is particularly beneficial for LED defect recognition with high intra-class variation and strong inter-class similarity.

Analysis of Dynamic Logits Fusion. The proposed DLF adaptively combines the logits from the student and AHF branches using a learnable coefficient β. Unlike fixed-weight fusion, DLF assigns sample-specific fusion weights, enabling dynamic exploitation of complementary classification cues from different feature hierarchies. For confident samples, the model emphasizes the more reliable branch, while for ambiguous samples, it leverages information from both branches. This adaptive strategy improves classification robustness and mitigates inter-class confusion with negligible computational overhead.

IV. CONCLUSION

In this paper, a lightweight network is proposed to address the challenges of imbalanced data distributions, limited computational resources, substantial intraclass variability, and strong interclass similarity in LED defect classification. The proposed approach integrates two key components, the Dynamic Dual-Path Classifier and Dual-Level Distillation, enabling the model to effectively reduce intraclass variation, alleviate interclass confusion, and operate under strict computational constraints. The DDPC serves as a discriminative classification component that enhances the adaptability of feature representations by extracting and fusing multi-level information. Through Adaptive Hierarchical Fusion and a

learnable β-generator, it explicitly models interclass confusion and produces clearer and more discriminative classification logits. The DLD component further transfers both semantic-level and feature-level knowledge from a high-capacity teacher model, enabling the model to achieve strong recognition performance with significantly fewer parameters. Extensive experiments demonstrate that the proposed method yields substantial accuracy gains on two LED defect datasets, validating its effectiveness in handling class imbalance and distinguishing visually ambiguous defect categories.

Data Availability

The code used in this study is available at https://github.com/JiaMingZhong2621/LED. The datasets used in this study are available from the corresponding author upon reasonable request.

Funding

This work was partially supported by the National Natural Science Foundation of China (Grant No. 62073274), the Design and Development of Non-contact Wafer Handling Manipulator (No. GKY-2025BSQDK-17), the Wuyi University-Hong Kong-Macau Joint Funding Scheme (Grant Nos. 2022WGALH17 and 2021WGALH18), the Guangdong Province Key Construction Discipline Research Ability Enhancement Project (Grant No. 2024ZDJS071), and the Key Fields Special Project for Ordinary Universities in Guangdong Province (Grant No. 2024ZDZX2083).

Author Contributions

J.D. and J.Z. conceived the study and designed the methodology. J.D. performed the experiments and analyzed the data. J.Z. supervised the project and contributed to the manuscript preparation. B.J., Y.L., B.X., C.T., and Q.L. assisted with data collection and experimental analysis. N.L. and F.D. provided guidance on the research design and manuscript revision. M.Z. supervised the research and revised the manuscript. All authors reviewed and approved the final manuscript.

Competing Interests

The authors declare no competing interests.

References

[1] L. Wei, J. Cai, K. Wen, and C. Zhang, "Local–global lightweight vit model for mini/micro-led-chip defect recognition," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106247, 2023.

[2] M. Chen, J. Chen, C. Li, Q. Wang, and K. Takamasu, "Defect detection of microled with low distinction based on deep learning," *Optics and Lasers in Engineering*, vol. 173, p. 107924, 2024.

[3] D.-M. Tsai, I.-Y. Chiang, and Y.-H. Tsai, "A shift-tolerant dissimilarity measure for surface defect detection," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 128–137, 2011.

[4] H. Wang, J. Zhang, Y. Tian, H. Chen, H. Sun, and K. Liu, "A simple guidance template-based defect detection method for strip steel surfaces," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2798–2809, 2018.

[5] W.-H. Weng, C.-Y. Tsai, C.-Y. Hung, and C.-F. J. Kuo, "Development of an adaptive template for fast detection of lithographic patterns of light-emitting diode chips," *The International Journal of Advanced Manufacturing Technology*, vol. 117, no. 11, pp. 3297–3321, 2021.

[6] C.-F. J. Kuo, C.-P. Tung, and W.-H. Weng, "Applying the support vector machine with optimal parameter design into an automatic inspection system for classifying micro-defects on surfaces of light-emitting diode chips," *Journal of Intelligent Manufacturing*, vol. 30, no. 2, pp. 727–741, 2019.

[7] Y. Chen, H. Liu, J. Zhuang, and J. Liang, "Enhanced fabric defect detection via snake-shaped global–local feature shrinking network," *The Visual Computer*, vol. 42, no. 6, p. 242, 2026.

[8] Q. Zhang, X. Dong, J. Xiong, H. Zhu, Q. Wang, J. Wang, and W. Dai, "Mcam-net: multi-scale convolutional attention for enhanced industrial surface defect detection," *The Visual Computer*, vol. 42, no. 4, p. 182, 2026.

[9] Y. Wen, S. Shen, W. Shi, W. Cao, L. Bi, X. Yang, and B. Sheng, "A lightweight depthwise separable convnet with frequency-domain enhancement for retinal vessel segmentation," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 21, no. 12, pp. 1–23, 2025.

[10] M. Chen, S. Han, and C. Li, "Efficient micro-led defect detection based on microscopic vision and deep learning," *Optics and Lasers in Engineering*, vol. 177, p. 108116, 2024.

[11] Z. Zhong, C. Li, M. Chen, H. Wu, and T. Kiyoshi, "Micro led defect detection with self-attention mechanism-based neural network," *Digital Signal Processing*, vol. 149, p. 104474, 2024.

[12] H. Zhang, T. Liu, K. Tang, J. Liu, W. Tang, and W. Wang, "Superpixel-based regional anomaly detection for led chips using local statistical features," *Optics and Lasers in Engineering*, vol. 195, p. 109281, 2025.

[13] Y. Lin, S. Pan, J. Yu, Y. Hong, F. Wang, L. Zheng, J. Tang, and S. Chen, "Mdca-detr: Detr with multi-channel deformable convolution and coordinate attention for mini-led wafer surface defects detection," *Optics and Lasers in Engineering*, vol. 193, p. 109082, 2025.

[14] Y. Lin, S. Pan, J. Yu, Y. Hong, F. Wang, J. Tang, L. Zheng, and S. Chen, "Mbgb-detector: A multi-branch gradient backhaul lightweight model for mini-led surface defect detection," *Computers in Industry*, vol. 164, p. 104204, 2025.

[15] J. Zhang, N. Chen, M. Li, Y. Zhang, X. Suo, R. Li, and J. Liu, "Lddp-net: A lightweight neural network with dual decoding paths for defect segmentation of led chips," *Sensors*, vol. 25, no. 2, p. 425, 2025.

[16] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[17] A. Andonian, S. Chen, and R. Hamid, "Robust cross-modal representation learning with progressive self-distillation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16430–16441, 2022.

[18] J.-X. Zhang, G. Wan, J. Gao, and Z.-H. Ling, "Audio-visual representation learning via knowledge distillation from speech foundation models," *Pattern Recognition*, vol. 162, p. 111432, 2025.

[19] Y. Tian, S. Pei, X. Zhang, C. Zhang, and N. V. Chawla, "Knowledge distillation on graphs: A survey," *ACM Computing Surveys*, vol. 57, no. 8, pp. 1–16, 2025.

[20] L. Yi, J. Song, Y. Huang, L. Zhou, J. Ding, C. Li, and Z. Ning, "A double knowledge distillation framework for insulator defect detection," *IEEE Transactions on Industrial Informatics*, vol. 21, no. 11, pp. 9126–9136, 2025.

[21] J. Liu, H. Li, F. Zuo, Z. Zhao, and S. Lu, "Kd-lightnet: A lightweight network based on knowledge distillation for industrial defect detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–13, 2023.

[22] Z. Wang, T. Li, Y. Qin, and B. Sheng, "Contrastive decoupling: Dynamic regularization for enhanced fine-grained image classification," *IEEE Transactions on Circuits and Systems for Video Technology*, 2026.

[23] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, pp. 6105–6114, PMLR, 2019.

[24] A. v. d. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.

[25] "LED Defect Dataset on AI Studio." https://aistudio.baidu.com/datasetdetail/152832. Accessed: 2022-07-14.

[26] X. Liu, J. Liu, Z. Wang, L. Wang, and H. Zhang, "Basic-class and cross-class hybrid feature learning for class-imbalanced weld defect recognition," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 9, pp. 9436–9446, 2022.

[27] Y. Zhang, Q. Zhan, and Z. Ma, "Efficientnet-eca: A lightweight network based on efficient channel attention for class-imbalanced welding defects

classification," *Advanced Engineering Informatics*, vol. 62, p. 102737, 2024.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[29] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*, pp. 8748–8763, PmLR, 2021.

[30] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollár, "Designing network design spaces," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10428–10436, 2020.

[31] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 116–131, 2018.

[32] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *International conference on machine learning*, pp. 10096–10106, PMLR, 2021.

[33] Y. Zhou, Q. Xu, J. Cui, J. Zhou, J. Zhang, R. Hong, and H. Zhang, "Care transformer: Mobile-friendly linear visual transformer via decoupled dual interaction," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 20135–20145, 2025.

[34] N. Setyawan, C.-C. Sun, M.-H. Hsu, W.-K. Kuo, and J.-W. Hsieh, "Microvitv2: Beyond the flops for edge energy-friendly vision transformers," in *2026 IEEE International Conference on Image Processing (ICIP)*, (Tampere, Finland), 2026.

[35] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[36] D. P. Kingma, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.