Rethinking SMT Defect Inspection from Human Visual Cognition: Modules that Boost Defect Performance

Hefeng Zhou[1], **Liyang Dong**[2,3], **Shuo Zhang**[2,3], **Yuanbin Wang**[1], **Yan Wu**[1], **Jiong Lou**[1], **Jie Li**[1,*]
[1]Shanghai Jiao Tong University
[2]COSMOPlat Institute of Industrial Intelligence (Qingdao) Co., Ltd
[3]State Key Laboratory of Massive Personalized Customization System and Technology

Abstract

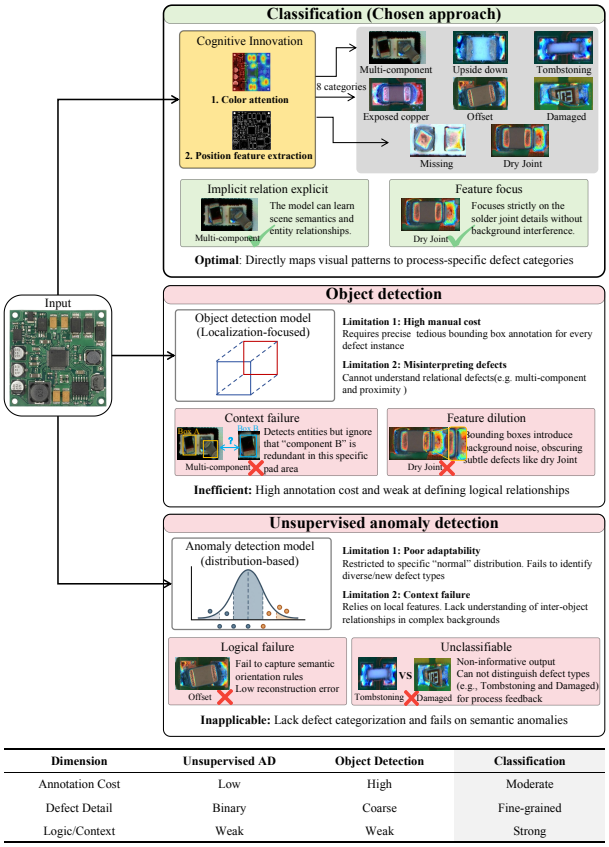
Human visual recognition is shaped by constraints such as sensitivity to peripheral evidence, integration of complementary visual cues, and tolerance for ambiguous category boundaries. We ask whether these principles can improve automated inspection in Surface Mount Technology (SMT), where defects are often subtle, off-center, and difficult to distinguish categorically. We implement these ideas in a standard classifier using lightweight modules for position-aware normalization, Fourier and color/position cue integration, and ambiguity-aware learning over confusable defect classes. Experiments on industrial and public PCB datasets show that these cognitively inspired modifications improve defect recall under low false-alarm conditions and encourage more consistent focus on defect-relevant regions. The results suggest that human-inspired perceptual constraints offer a useful framework for designing robust machine vision systems in applied settings.

Keywords: computer vision, industrial applications.

Introduction

Computer vision has made substantial progress in generic image recognition, and many benchmark settings now appear technically mature. A substantial gap often persists between benchmark performance and dependable deployment in domains where errors carry high costs and operating conditions impose strict constraints. Industrial inspection and medical imaging exemplify this gap: models must be accurate, stable under distribution shifts, and efficient enough to run on production hardware with strict latency and memory budgets (Jeevan & Sethi, 2024). In such settings, success depends not only on improving average accuracy, but also on reducing false alarms and missed defects under realistic process variability (Khanam et al., 2024).

A useful way to rethink industrial recognition problems is to return to the cognitive principles that originally inspired computer vision. Human vision is often described as a hierarchy spanning early, intermediate, and high-level processing, with selective attention allocating limited resources to task-relevant regions and integrating spatial priors with learned feature representations. These principles are especially relevant when target signals are subtle, sparse, or located near the visual periphery (Islam et al., 2024). Surface Mount Technology (SMT) defect inspection provides a concrete testbed for this perspective. In Automatic Optical Inspection (AOI), systems must verify thousands of printed circuit boards daily, each containing diverse components and inspection points (Ebayyeh & Mousavi, 2020). Defects include missing components, displacement, misorientation, solder-related failures, and surface damage (Sankar et al., 2022). This setting introduces three challenges that are also cognitively meaningful: defect evidence may appear in variable positions across crops, category boundaries are often ambiguous, and deployment constraints require lightweight and stable solutions (A. Wang et al., 2024).



Dimension	Unsupervised AD	Object Detection	Classification
Annotation Cost	Low	High	Moderate
Defect Detail	Binary	Coarse	Fine-grained
Logic/Context	Weak	Weak	Strong

Figure 1: Visual illustration of why existing methods struggle with SMT defect tasks.

* corresponding author. This work was supported in part by the National Key R&D Program of China under Grant 2024YFB2705300, the National Natural Science Foundation of China (NSFC) under Grants 62232011 and 62402315, the Shanghai Science and Technology Innovation Action Plan under Grant 24BC3201200, and the Key Research and Development Program of Shandong Province under Grant 2024CXPT068.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

Existing approaches address parts of this problem but often fail under the combined demands of component diversity, multiple defect types, and strict deployment budgets. Rule-based AOI pipelines are sensitive to illumination and viewpoint variation and require substantial manual engineering (Taha et al., 2014). Object detection and segmentation methods struggle when defects are defined less by distinct objects than by subtle relational or semantic cues (Fontana et al., 2024; J. Wang et al., 2022). Unsupervised anomaly detection is limited by component diversity and by the need to distinguish rare but acceptable variations from true defects (Batzner et al., 2024). Pure image classification remains an efficient end-to-end solution, but standard classifiers may perform poorly when recognition depends on peripheral evidence, spatial reasoning, or fine-grained cue integration (Ge et al., 2022). These limitations motivate a cognitively grounded reformulation of SMT defect inspection that explicitly incorporates position sensitivity, feature integration, and graded ambiguity into the learning process.

Motivated by these observations, we revisit image classification as a core building block and strengthen it using cognitively grounded design principles. We target three failure modes that frequently appear in production. First, position sensitivity and center biased predictions can cause misses when defect evidence lies near image boundaries. Second, task specific cues in SMT images often depend on frequency structure, local regularity, and spatial configuration, which may not be adequately captured by a standard backbone alone (S. Wang et al., 2019). Third, label ambiguity and confusable defect categories can distort training objectives when a single label is forced to represent multiple plausible interpretations. Our method addresses these issues through a spatially aware preprocessing strategy, task tailored feature extraction and fusion, and an objective that accounts for structured confusability among defect categories, all while remaining compatible with industrial resource limits (Wu et al., 2022).

In summary, this work makes the following contributions:

- We analyze SMT AOI crop classification and identify three practical challenges that systematically affect recognition: uncertain evidence location (often near crop boundaries), overlapping cues across defect categories, and asymmetric inspection costs.
- We propose lightweight, deployment-compatible modules inspired by these constraints, including position-aware random translation during training, compact Fourier and color/position feature augmentation, and an ambiguity-aware objective for predefined confusable defect pairs.
- We evaluate under held-out testing and a production-oriented coverage check (reported separately), using inspection-level metrics (OK FPR and NG recall) together with stress tests and behavioral analyses that probe robustness to color precision and evidence location.

Method

Materials

Industrial data and protocols. Our primary materials are component-level image crops collected from a real Surface Mount Technology (SMT) production line using an in-house Automatic Optical Inspection (AOI) system. Board-level images are acquired by fixed FOV industrial cameras under multi-angle illumination, and inspection points are localized and cropped by a rule-assisted alignment procedure relative to a golden reference board. Crop sizes vary widely with component geometry, approximately 50×30 to 1500×1000 pixels, yielding substantial variability in scale and in the spatial location of defect evidence within the crop (Zervakis et al., 2004). Unless otherwise noted, datasets are split at the image level into 70% training and 30% testing by random sampling. In addition, to match common industrial validation practice, we also report an industry-style backtesting setting in which all available historical data are used for training and performance is evaluated retrospectively on the same historical set. The industrial dataset contains over 170k images in total. We report both protocols to contrast standard held-out evaluation with production-oriented coverage requirements.

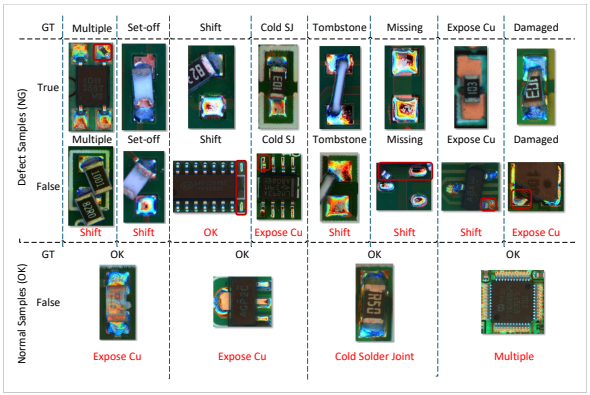


Figure 2: **Samples collected in industrial testing.** Defective (NG) samples are shown in the upper section, normal (OK) samples in the lower section. For NG defects (8 types), **black** denotes correct classification; **red** indicates misclassified samples (containing multiple defects or prone to human/model error). Red in the OK column marks false positives. All displayed data are real and not synthetic.

Defect taxonomy and labeling convention. Each crop is assigned one of nine labels: OK and eight defect types (Multi-item, Set-off, Tombstone, Expose Cu, Shift, Damaged, Missing, Cold Solder Joint). In production data, multiple defect cues may co-occur within a single crop, which introduces inherent label ambiguity. To maintain a lightweight single-label classification setup compatible with deployment constraints, annotators assign a single label using a predefined priority order when multiple defect cues are present. We also maintain a predefined set of confusable defect relations that are treated as

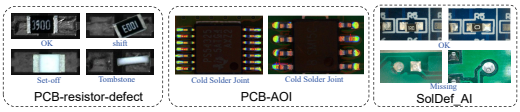


Figure 3: Public SMT defect dataset.

partially overlapping in analysis and in the training objective (Song et al., 2019).

Annotation and quality control. All industrial images are manually annotated following written labeling guidelines. Labels are reviewed by multiple annotators through a multi-pass verification procedure to reduce noise in ambiguous cases and to ensure consistency with AOI practice.

Additional datasets for generalization. To test whether observed effects generalize beyond our in-house data, we also evaluate on public PCB-related datasets (PCB-AOI, SolDef-AI, and a resistor defect dataset), mapping their defect categories to the closest SMT categories when applicable. As a natural-image control, we additionally report results on Tiny-ImageNet to assess whether changes targeting industrial cues degrade generic recognition behavior (Le & Yang, 2015). Details are shown in Figure 3.

Proposed Method

Our goal is to improve defect recognition in AOI component crops by explicitly modeling two cognitive mismatches with standard object-centric classification. First, evidence location is uncertain: cues may occur near crop boundaries due to upstream localization/alignment and boundary defects. Second, evidence is not strictly exclusive across categories: several defect types share cues and can co-occur, while the labeling protocol enforces a single label via a priority rule. We introduce minimal modules that (i) reduce position-dependent learning, (ii) integrate complementary low-level cues, and (iii) encode graded penalties for confusable categories, keeping the backbone and overall training recipe otherwise standard (Bhattacharya & Cloutier, 2022).

Position-aware input normalization Convolutional models trained on natural-image benchmarks often internalize a center prior because discriminative evidence is typically object-centered. In inspection crops, however, diagnostic cues may occur anywhere within the crop and are frequently near boundaries due to imperfect localization and edge-adjacent defects. To reduce this mismatch, we use a position-aware training preprocessor: each crop is embedded into a larger canvas via symmetric padding, then randomly moved within the canvas while preserving pixel content. This increases the positional variability of identical evidence across epochs and encourages location-robust decisions driven by local appearance rather than absolute coordinates. Importantly, we treat position not only as a nuisance factor to randomize, but also as a controllable cue: downstream, our color+position-aware feature augmentation explicitly injects lightweight spatial/appearance signals (Ge et al., 2022) into the representation so that the model can integrate "where" and "what" information when it is informative. At test time, we deterministically center the padded crop to avoid adding evaluation noise.

Learning with structured label ambiguity Although each training image is assigned a single ground-truth label for operational reasons, some defect categories share overlapping visual cues and often co-occur in practice. Under ordinary cross-entropy training, these relations are treated as equally incompatible, which can lead the model to allocate excessive evidence against plausible alternatives and to become overconfident in ambiguous cases. We incorporate prior knowledge about confusable defect relations by augmenting the learning objective with a soft overlap constraint. Concretely, we specify a small set of category relations identified by inspection guidelines and error patterns, and we relax the penalty for confusions within these relations. This preserves the simplicity of single-label training while allowing the model to express graded uncertainty for visually similar defects, which better matches the ambiguity structure of industrial AOI data (Yu et al., 2022).

Classifier and training objective We use a standard convolutional image classifier that maps an input crop to a categorical distribution over the nine defect labels. In addition to the ordinary classification loss, we explicitly encode two task-relevant pressures that are not captured by accuracy alone: the priority of detecting defects (NG recall) and the graded severity of confusing visually similar defect types (Deng et al., 2018).

Let y and $\hat{y}$ denote the ground-truth label and predicted label, respectively. Our overall loss is

$$\mathcal{L} = \mathcal{L}_{Bce}(y, \hat{y}) + \mathcal{L}_{NGRecall}(y, \hat{y}) + \lambda \cdot \ell_{SOD}(y, \hat{y}), \quad (1)$$

where $\mathcal{L}_{Bce}$ is the base classification term, $\mathcal{L}_{NGRecall}$ is a defect-focused term that prioritizes separating defective (NG) from non-defective (OK) parts, and ℓ_{SOD} is the Similar Overlap Defect term that down-weights penalties for a small set of confusable defect relations.

The term $\mathcal{L}_{NGRecall}$ inherits from the binary cross-entropy form by collapsing the nine-way labels into a binary decision, OK versus NG, where NG contains the eight defect classes. This aligns training with inspection practice, in which the primary concern is whether a part is defective at all, and it mitigates cases where high overall accuracy can coexist with low recall on rare defects.

To model structured ambiguity among visually similar defects, we define three symmetric confusable relations: (a) Shift versus Missing, (b) Tombstone versus Shift, (c) Multi-item versus Shift. Here, $\hat{y}$ denotes the model prediction used in training (i.e., the softmax probability output rather than the argmax label), so the SOD term is computed in a differentiable manner by applying the graded costs in expectation over the

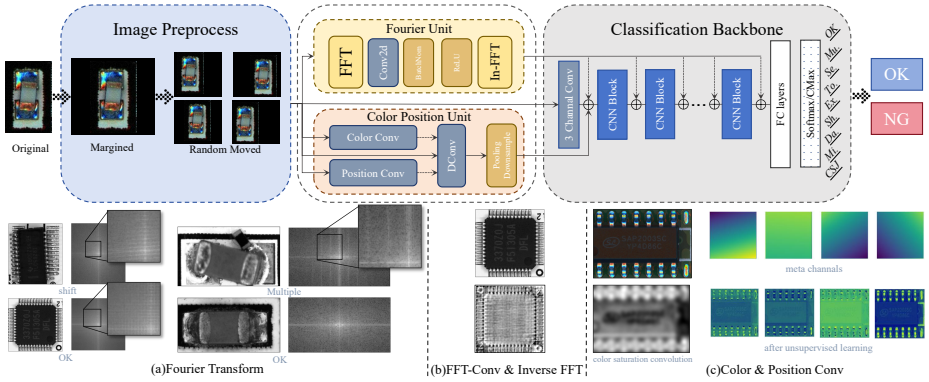


Figure 4: **Proposed framework and feature extraction units.** The image is input directly into the backbone and fused via the feature extraction unit. The dotted line in the feature extraction unit represents the mask that only provides guidance. The dotted line in the backbone represents that a part may be selected for fusion, which is determined during the construction of the backbone. **(a)** Fourier unit extracts defect features as low-level input. **(b)** Feature map after Fourier transformation. **(c)** Color saturation convolution and trained channel mask highlight critical regions. See Table for the abbreviations of defect.

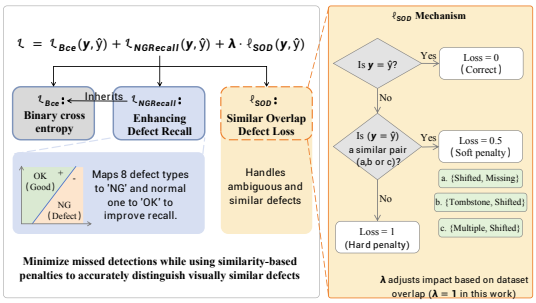


Figure 5: Loss design description.

predicted distribution. The Similar Overlap Defect cost is

$$\ell_{SOD}(y,\hat{y}) = \begin{cases} 0, y = \hat{y} \\ 0.5, \textit{if situation a. or b. or c.} \\ 1, \textit{Except for the above situations} \end{cases} \quad (2)$$

which implements a graded notion of error severity: confusions within the confusable set are treated as less informative mistakes than confusions between unrelated categories. This choice reflects the fact that these categories partially overlap in their visual evidence and can co-occur in production, while the labeling protocol requires a single label.

The hyperparameter λ controls how strongly the model is encouraged to treat these confusable relations as partially overlapping. Larger values are appropriate when overlapping cues and near-boundary cases are more common. We set $\lambda = 1$ in our design. The loss follows the procedure in Figure 5.

Evaluation logic We evaluate the method under two validation protocols that correspond to different inferential goals. The held-out split assesses generalization to unseen images drawn from the same sampling process using a 70% training and 30% testing split at the image level. The industry-style backtesting assesses coverage and stability under the full historical variability encountered in production. Across both protocols, we report class-wise and inspection-relevant metrics, with particular emphasis on defect recall and false positive rate on OK parts, which reflect asymmetric costs in inspection workflows (Volkau et al., 2019).

Experiment

Our evaluation is organized around two task-level claims that are easy to miss when PCBA inspection is treated as ordinary object-centric classification. The first claim is about positional uncertainty: defect evidence is not reliably centered within AOI crops, so a model with center preference should fail disproportionately on boundary-localized defects. The second claim is about *structured ambiguity*: several defect categories share partially overlapping visual cues, so forcing strict mutual exclusivity during training should produce systematic confusions and overconfident errors. We test whether the proposed preprocessing and loss design shift model behavior in the intended directions under both conventional generalization testing and production-oriented verification.

Environment and training settings

For deployment, we use TensorRT inference on an industrial computer equipped with an RTX 3060 (12GB). Unless otherwise stated, models are trained for up to 200 epochs with input size 256×256 and 9 output classes. The batch size is 128 for training and 64 for inference. We use SGD with initial learning rate 0.001 and step decay factor $\gamma = 0.9$ every 10 epochs. The optimization target is the proposed objective in Eq. 1, including the structured ambiguity term in Eq. 2.

Data sources

The main dataset contains over 170k manually annotated component crops collected from a real SMT production line. It includes the majority of normal (OK) samples and eight defect types: Missing (Mi.), Cold Solder Joint (CSJ), Shift (Sh.),

Table 1: Performance comparison on dataset (%): overall performance, single-item recall and cost

Method	Overall Performance (%)				Single Item Recall (%)								Cost(time for per batch)	
	Recall	NG Recall	Fpr	Acc.	OK	Multiple	Set-off	Tombstone	Expose Cu	Shift	Damaged	Cold Solder Joint	TimeCost(s)	MemCost(Mb)
RepViT-1.1 (A. Wang et al., 2024)	47.12	71.61	4.81	88.23	94.57	24.59	35.59	34.41	77.96	46.61	10.71	80.15	0.0397s	1302
RepViT-2.3	57.25	75.57	3.12	91.48	97.53	31.15	50.08	24.73	87.98	54.68	53.57	81.22	0.0611s	1568
MobileViT-xxs (Mehta & Rastegari, 2021)	83.25	94.79	0.66	97.42	98.49	73.30	91.53	77.41	92.98	86.22	46.43	96.09	0.0411s	995
MobileViT-xs	84.17	95.41	0.53	97.89	98.83	72.60	93.22	80.65	90.98	90.42	53.57	95.06	0.0561s	1368
Efficient-B0	88.33	95.79	0.41	98.91	99.64	90.63	90.68	82.85	87.90	98.71	46.43	97.99	0.0124s	1558
Efficient-B3	89.64	96.26	0.36	98.09	99.80	90.40	95.34	85.97	89.94	96.45	64.29	99.00	0.0176s	2254
Efficient-B5	92.52	96.75	0.31	98.66	99.46	90.87	93.22	87.77	90.04	97.52	78.57	99.06	0.0211s	2792
RegNet-Y-400MF	85.43	93.65	0.68	98.14	99.27	80.58	89.83	87.95	88.98	85.25	53.57	98.01	0.0153s	1220
RegNet-Y-800MF	85.39	95.36	0.49	98.17	99.40	78.74	92.37	84.93	91.03	88.37	50.00	95.04	0.0241s	1255
RegNet-Y-1600MF	86.01	95.72	0.53	98.11	98.85	83.61	90.68	77.41	89.89	91.39	46.43	97.03	0.0186s	1358
ResNet-18	74.63	92.16	1.33	94.07	98.72	60.42	82.20	57.61	84.94	88.16	39.29	88.15	0.0023s	946
ResNet-34	77.10	93.11	1.14	97.28	96.03	70.02	74.58	72.45	89.07	74.06	35.71	88.04	0.0029s	988
ResNet-50	86.67	94.29	0.74	97.12	98.20	78.69	85.59	80.63	89.99	89.34	50.00	94.08	0.0045s	1756
ResNeXt (Xie et al., 2017)	88.75	91.55	0.72	98.12	98.45	79.52	86.95	81.97	90.76	89.73	53.57	97.24	0.0049s	1816
ResNeSt (Zhang et al., 2022)	90.21	94.85	0.59	98.78	98.82	80.46	89.48	84.62	91.83	90.21	53.57	95.87	0.0052s	1875
ACmix (Pan et al., 2022)	91.97	95.28	0.66	99.23	99.06	81.05	90.98	85.91	92.42	90.57	78.57	96.72	0.0053s	1987
Ours(ResNet50)	94.07	97.32	0.69	98.99	97.09	81.27	93.75	87.18	93.77	89.80	71.43	98.05	0.0051s	1812
Ours(EfficientNet-B0)	95.23	97.96	0.37	99.16	98.69	93.22	95.83	89.74	93.02	98.92	71.43	99.14	0.0126s	1715
ResNet50*	97.12	98.17	0.49	99.47	98.79	96.50	99.07	96.32	99.19	96.05	97.12	99.15	—	—
Ours*(ResNet50)	99.26	99.88	0.04	99.72	99.84	99.27	99.89	99.87	99.96	98.60	97.06	99.63	—	—

Tombstone (To.), Set-off (Se.), Expose Cu (Ex.), Multi-item (Mu.), and Damaged (Da.). Industrial defects are rare, so we sample the dataset to reflect the real-world class imbalance. We also apply data augmentation and synthetic generation, but these samples are used for training only.

To probe transfer beyond the proprietary distribution, we additionally incorporate three public PCB datasets, PCB-AoI, SolDef-AI, and PCB-resistor-defect, which overlap with our defect taxonomy and provide secondary benchmarks for fine-tuning and evaluation. Our in-house task further includes three more challenging defect categories that do not appear in these public datasets, and we mark them as not applicable in the public-data setting. Details are shown in Table 2.

Table 2: Statistics of datasets used in this work.

Source	Use	From	OK	Missing	Cold Solder Joint	Shift	Tombstone	Set-off	Expose Cu	Multiple	Damaged
Ours	Train	Real Data	81355	4000	834	1140	489	900	6000	1140	42
		Augmentation	80000	2000	834	1140	500	300	1200	1140	420
	Test	Real Data	77165	4196	623	929	93	118	4025	427	28
Public	Train	Real Data	200	50	30	50	50	50	-	-	-
		Augmentation	400	50	30	50	50	50	-	-	-
	Test	Real Data	50	10	10	10	10	10	-	-	-

Protocols, preprocessing, and evaluation criteria

We report two validation protocols that serve different inferential goals. The held-out protocol uses a 7:3 train/test split at the image level to estimate generalization under the same acquisition process. The production backtesting protocol (marked with an asterisk) trains on the full historical dataset and evaluates on the same pool as a coverage/regression check for industrial deployment: it verifies that the model handles historically observed (including rare) defect modes and remains stable across historical variability, but it is not an unbiased estimate of out-of-sample error. All experiments use the unified position-aware preprocessing described in Sec. 2: during training, crops are resized and padded into a 256×256 canvas and randomly moved within a small margin so the same defect evidence appears at multiple spatial positions across epochs, mitigating center bias when cues are edge-localized.

Because inspection cost is asymmetric, we evaluate with both inspection-level (grouped) and class-level (strict) criteria. Let OK denote the normal class and NG denote the union of the eight defect classes. Collapsing the nine-way predictions into OK vs. NG yields grouped confusion counts: TP' (true positives; NG correctly predicted as NG), FN' (false negatives; NG misclassified as OK), FP' (false positives; OK misclassified as NG), and TN' (true negatives; OK correctly predicted as OK). Under this grouping, we define the defect-detection recall (NG Recall) as the grouped true positive rate,

$$\text{NG Recall} = \text{TPR}' = \frac{TP'}{TP' + FN'}, \tag{3}$$

and we define the false positive rate on OK parts as

$$\text{FPR} = \frac{FP'}{FP' + TN'}. \tag{4}$$

In addition to these inspection-level metrics, we report per-class recall across the nine labels (strict multi-class recall) to identify which defect types benefit from the proposed changes and to characterize residual confusions among visually similar categories. We also include accuracy (Acc.), a commonly used metric in standard classification.

Overall results on the production-line dataset

We compare our approach with standard CNN backbones (He et al., 2016; Tan & Le, 2021) and mobile-friendly transformers (Mehta & Rastegari, 2021; A. Wang et al., 2024) under matched training settings. Our method improves grouped Recall and reduces FPR across backbones, giving a better operating point for production inspection with asymmetric costs. Improvements are strongest for boundary or off-center defects and for visually overlapping categories, consistent with

position-aware preprocessing and structured ambiguity learning. Backtesting results marked with * are for deployment verification rather than unbiased generalization. The blue and gray highlights indicate the plug-and-play module comparisons that deserve primary attention.

Transfer to public datasets

To evaluate transfer, we augment the training set with public training samples and fine-tune the best production model for 20 epochs. We evaluate on the public test split. Gains on overlapping categories suggest the method captures task-relevant cues that transfer across components and imaging conditions, rather than line-specific artifacts (Table 3).

Table 3: Performance in public datasets: recall rate

Method	OK	Mi.	CSJ	Sh.	To.	Se.
EfficientNet-B0	0.8	1.0	0.7	0.9	0.7	0.8
Ours(EfficientNet-B0)	1.0	0.9	0.8	1.0	0.8	0.9
ResNet50*	1.0	1.0	0.9	0.9	0.8	1.0
Ours*(ResNet50)	1.0	1.0	1.0	1.0	0.9	1.0

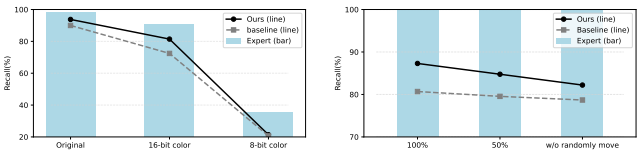


Figure 6: Comparison of color downsample (left) and randomly move effect (right).

Ablation and behavioral interpretation

We ablate key components under matched settings with a ResNet50 backbone to understand which design choices drive inspection behavior. Removing the loss design yields the largest degradation (**Recall↓5.54%, NG Recall↓2.29%, ACC↓1.07%**), suggesting that the proposed objective meaningfully reshapes the model's decision policy for industrial ambiguity: it encourages less overconfident separation among visually confusable defect classes and better prioritizes defect-oriented recognition. Removing the Fourier+Color Position feature extraction and fusion unit also reduces performance (**Recall↓3.21%, NG Recall↓2.35%, ACC↓0.47%**), while OK FPR changes only slightly. This indicates that the additional channels mainly improve feature separability for subtle/peripheral defect evidence rather than simply shifting the operating threshold. Using the Fourier unit or the Color Position unit alone provides limited benefit, implying that the two streams act complementarily.

Case study and interpretability

We test whether the proposed modules change model behavior in ways consistent with perceptual factors in SMT inspection using controlled input perturbations (Fig. 6). We reduce color precision by color downsampling to probe the contribution of color-aware cues, and we vary the random-translation magnitude in our position-aware preprocessing to probe robustness to evidence location. We also include human expert results on the same inputs to relate these perturbations to human perception. Our method is more robust to color downsampling and benefits more consistently from larger translation magnitude than the baseline (Fig. 6). These stress tests support our perceptually motivated design, and the attribution maps suggest a more consistent focus on defect-relevant regions (Fig. 7).

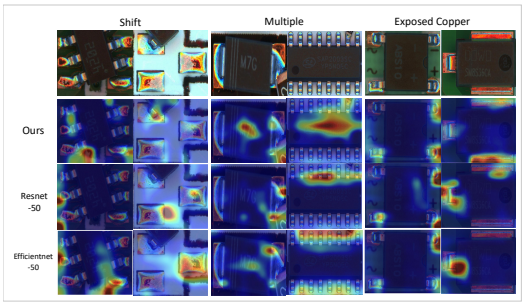


Figure 7: Interpretability of Representative Defect Cases.

Discussion

Our results suggest that modeling inspection crop recognition as classification under perceptual constraints yields more reliable decisions. Position aware normalization mitigates center bias when diagnostic cues lie near crop boundaries, and Fourier plus color and position augmentation provides complementary low level signals that improve recognition of subtle defects. The ambiguity aware objective further shifts learning toward defect detection by reducing penalties among known confusable classes, consistent with the ablations.

Several limitations remain. Single label supervision with a priority rule cannot represent defect co occurrence and may conflate true visual ambiguity with labeling conventions. The confusable pair structure is hand specified and may not transfer across lines, component mixes, or imaging conditions; backtesting offers operational coverage but does not replace out of distribution evaluation. Future work should study multi label and uncertainty aware training, learn confusability from data, and run controlled stress tests including systematic translations, peripheral cue perturbations, and illumination shifts to determine when these biases help and when additional mechanisms are needed.

Conclusion

We show that SMT AOI defect recognition improves when framed as classification under cognitive constraints rather than object-centric recognition. Modules for position robustness, low-level cue integration, and graded treatment of confusable classes increase defect recall while maintaining low false-alarm rates under deployment limits. More broadly, aligning inductive biases and training objectives with the task yields more predictable evidence use and error patterns in inspection.

References

Batzner, K., Heckler, L., & König, R. (2024). Efficientad: Accurate visual anomaly detection at millisecond-level latencies. *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 128–138.

Bhattacharya, A., & Cloutier, S. G. (2022). End-to-end deep learning framework for printed circuit board manufacturing defect classification. *Scientific Reports*. https://doi.org/10.1038/s41598-022-16302-3

Deng, Y.-S., Luo, A.-C., & Dai, M.-J. (2018). Building an automatic defect verification system using deep neural network for pcb defect classification. *2018 4th International Conference on Frontiers of Signal Processing (ICFSP)*. https://doi.org/10.1109/icfsp.2018.8552045

Ebayyeh, A. A. R. M. A., & Mousavi, A. (2020). A review and analysis of automatic optical inspection and quality monitoring methods in electronics industry. *IEEE Access*, 183192–183271. https://doi.org/10.1109/access.2020.3029127

Fontana, G., Calabrese, M., Agnusdei, L., Papadia, G., & Del Prete, A. (2024). Soldef_ai: An open source pcb dataset for mask r-cnn defect detection in soldering processes of electronic components. *Journal of Manufacturing and Materials Processing*, *8*(3), 117.

Ge, Y., Xiao, Y., Xu, Z., Wang, X., & Itti, L. (2022). Contributions of shape, texture, and color in visual recognition. *European Conference on Computer Vision*, 369–386.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Islam, M. A., Kowal, M., Jia, S., Derpanis, K. G., & Bruce, N. D. (2024). Position, padding and predictions: A deeper look at position information in cnns. *International Journal of Computer Vision*, 1–22.

Jeevan, P., & Sethi, A. (2024). Which backbone to use: A resource-efficient domain specific comparison for computer vision. *arXiv preprint arXiv:2406.05612*.

Khanam, R., Hussain, M., Hill, R., & Allen, P. (2024). A comprehensive review of convolutional neural networks for defect detection in industrial applications. *IEEE Access*, *12*, 94250–94295. https://doi.org/10.1109/ACCESS.2024.3425166

Le, Y., & Yang, X. (2015). Tiny imagenet visual recognition challenge. *CS 231N*, *7*(7), 3.

Mehta, S., & Rastegari, M. (2021). Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv:2110.02178*.

Pan, X., Ge, C., Lu, R., Song, S., Chen, G., Huang, Z., & Huang, G. (2022). On the integration of self-attention and convolution. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 815–825.

Sankar, V. U., Lakshmi, G., & Sankar, Y. S. (2022). A review of various defects in pcb. *Journal of Electronic Testing*, 481–491. https://doi.org/10.1007/s10836-022-06026-7

Song, J.-D., Kim, Y.-G., & Park, T.-H. (2019). Smt defect classification by feature extraction region optimization and machine learning. *The International Journal of Advanced Manufacturing Technology*, 1303–1313. https://doi.org/10.1007/s00170-018-3022-6

Taha, E. M., Emary, E., & Moustafa, K. (2014). Automatic optical inspection for pcb manufacturing: A survey. *Int. J. Sci. Eng. Res*, *5*(7), 1095–1102.

Tan, M., & Le, Q. (2021). Efficientnetv2: Smaller models and faster training. *International conference on machine learning*, 10096–10106.

Volkau, I., Mujeeb, A., Wenting, D., Marius, E., & Alexei, S. (2019). Detection defect in printed circuit boards using unsupervised feature extraction upon transfer learning. *2019 International Conference on Cyberworlds (CW)*, 101–108.

Wang, A., Chen, H., Lin, Z., Han, J., & Ding, G. (2024). Repvit: Revisiting mobile cnn from vit perspective. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15909–15920.

Wang, J., Zhou, H., & Yu, X. (2022). Pgtrnet: Two-phase weakly supervised object detection with pseudo ground truth refinement. *ICASSP 2022*, 2245–2249.

Wang, S., Han, K., & Jin, J. (2019). Review of image low-level feature extraction methods for content-based image retrieval. *Sensor Review*, 783–809. https://doi.org/10.1108/sr-04-2019-0092

Wu, H., Lei, R., & Peng, Y. (2022). Pcbnet: A lightweight convolutional neural network for defect inspection in surface mount technology. *IEEE Transactions on Instrumentation and Measurement*, 1–14. https://doi.org/10.1109/tim.2022.3193183

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1492–1500.

Yu, C., Zhao, Y., & Xu, Z. (2022, January). Smt component defection reassessment based on siamese network. In *Communications in computer and information science,methods and applications for modeling and simulation of complex systems* (pp. 75–85). https://doi.org/10.1007/978-981-19-9195-0_7

Zervakis, M., Goumas, S., & Rovithakis, G. (2004). A bayesian framework for multilead smd post-placement quality inspection. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 440–453. https://doi.org/10.1109/tsmcb.2003.817037

Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., Sun, Y., He, T., Mueller, J., Manmatha, R., et al. (2022). Resnest: Split-attention networks. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2736–2746.