Journal of Electronic Research and Application, 2026, Volume 10, Issue 6
http://ojs.bbwpublisher.com/index.php/JERA
Online ISSN: 2208-3510
Print ISSN: 2208-3502

Architectural Design of RT-DETR-L for PCB Surface Defect Detection: A Systematic Comparison of Attention Mechanisms, Backbone Replacement, and Cross-Dataset Generalization

Tao Huang*

School of Electronics and Information Engineering, Liaoning Technical University, Huludap 125105, Liaoning, China

***Corresponding author:** Tao Huang, phamvanlinh65116@gmail.com

Abstract: Based on RT-DETR-L, this paper systematically compares five attention mechanisms (SE, CBAM, CA, ECA, and EMA) at the P3/P4/P5 outputs of the feature-pyramid neck under identical training conditions, and evaluates FasterNet backbone replacement and a P2 small-object detection head as complementary improvements. Experiments reveal a performance gap of up to 4.29 percentage points (CA: 93.04% to EMA: 97.33% in mAP_{50}), indicating that the choice of attention mechanism has a substantial impact on RT-DETR-type PCB detectors. EMA achieves the best mAP_{50} (97.33%) and the highest $mAP_{50:95}$ (56.45%); ECA offers a competitive trade-off without increasing GFLOPs (96.69%); CA performs worst (93.04%), a 3.34 pp drop below the baseline, tentatively attributed to an architectural conflict with the AIFI encoder. FasterNet backbone replacement trades accuracy for efficiency (31% fewer parameters, 40% lower GFLOPs); and, when trained from scratch on the second dataset DeepPCB, the EMA variant again yields the largest gain (mAP_{50}89.33%, +4.83 pp over the baseline), showing that the improvement is not specific to a single dataset.

Keywords: PCB defect detection; RT-DETR; Attention mechanism comparison; EMA; FasterNet; Cross-dataset generalization

Online publication: Jul 7, 2026

1. Introduction

Surface defects in printed circuit board (PCB) manufacturing—open circuit, short, spur, spurious copper, missing hole, and mouse bite— lead to product failure and increased recall costs if left undetected [1]. Deep learning has greatly advanced PCB defect detection: the YOLO series attains strong accuracy at high throughput, and various PCB-oriented YOLO variants further adapt it to manufacturing constraints. The Transformer-based RT-DETR overcomes the NMS bottleneck via bipartite matching and has drawn wide

attention in PCB detection: HSA-RTDETR integrates a FasterNet-EMA backbone and reaches 96.9% mAP_{50}, while other works improve small-object localization by introducing shallow P2 features or achieve light-weighting by drastically reducing parameters [2,3].

However, two research gaps remain: (1) a systematic comparison of different attention mechanisms at the neck (after the backbone, before AIFI) of the RT-DETR-L feature pyramid is still absent in PCB detection; (2) systematic evaluation on more than one PCB dataset for RT-DETR-type detectors is largely missing.

The main contributions of this paper are:

the first controlled comparison of five attention mechanisms (SE, CBAM, CA, ECA, EMA) at the P3/P4/P5 neck outputs of RT-DETR-L, revealing a 4.29 pp performance gap;

(1) Quantifying the effect of neck-level EMA insertion, achieving the best mAP_{50} (97.33%) and $mAP_{50:95}$ (56.45%);

(2) Faithfully reporting the accuracy–efficiency trade-off of FasterNet backbone replacement (with the FasterNet+P2 configuration evaluated on DeepPCB);

(3) Providing a multi-variant comparison on a second dataset (DeepPCB), where the EMA variant again yields the largest gain (89.33%), showing the improvement is not specific to one dataset.

2. Proposed method

2.1. RT-DETR-L baseline architecture

RT-DETR-L adopts an HGNetv2 backbone, an AIFI Transformer encoder, a RepC3 FPN neck, and an RTDETRDecoder for end-to-end detection [2]. With 640×640 input, the backbone outputs three-level features P3 (80^2), P4 (40^2), and P5 (20^2); the AIFI encoder applies 8-head self-attention (embedding dimension 256, feed-forward dimension 1024) to the highest-level feature; the decoder requires no NMS thanks to bipartite matching. Because backbone replacement makes pretrained weights incompatible, all modified models are trained from scratch (pretrained=False)..

2.2. Neck-level EMA Multi-scale Attention (Main Contribution)

The EMA module is inserted at the P3/P4/P5 outputs, immediately before the input projection layer of the AIFI encoder. In simplified form, EMA can be summarized as two parallel branches operating on $\mathbf{X}\in R^{C\times H\times W}$ (the full module additionally applies grouped cross-spatial learning, see [4]):

$$\mathbf{f}_c\left(\mathbf{X}\right)=\sigma\left(\mathbf{W}_1\cdot GAP_{1D}\left(\mathbf{X}\right)\right)\otimes\mathbf{X},\mathbf{f}_s\left(\mathbf{X}\right)=\sigma\left(\mathbf{W}_3*\mathbf{X}\right)\otimes\mathbf{X},\mathbf{Y}=\mathbf{f}_c+\mathbf{f}_s$$

Unlike SE, EMA does not compress the channel dimension; compared with CBAM and CA, its parallel design avoids the sequential bottleneck. Our neck-level insertion differs from the backbone-integrated scheme of HSA-RTDETR, acting on post-backbone features before multi-scale fusion [3, 5–7].

2.3. FasterNet Backbone Replacement and P2 Detection Head

HGNetv2 is replaced with FasterNet, whose core partial convolution (PConv) apples standard convoutin only to $C/4$ channels: $PConv(\mathbf{X})=\left[Conv(\mathbf{X}_{1:C/4});\mathbf{X}_{C/4+1:C}\right]$, greatly reducing computation [8]. By leaving the remaining $3C/4$ channels untouched, PConv also avoids much of the memory-access overhead that bottlenecks depthwise-separable designs, which is the main reason FasterNet attains higher realized

throughput than its FLOPs alone would suggest. On top of the FasterNet backbone, an additional P2 feature map (160×160, stride 4) is introduced as a fourth-scale detection head to improve the localization of small targets such as spurs and missing holes, following the established idea of introducing shallow features in RT-DETR to improve small-object localization. The P2 variant is evaluated only on DeepPCB and is not independently ablated on pcb-cropped. **Figure 1** contrasts the baseline (a) with the proposed EMA-enhanced architecture (b).

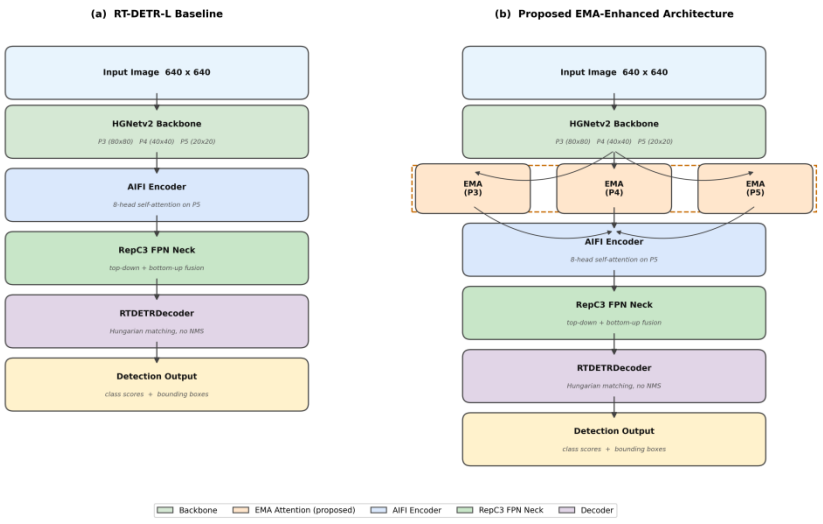

Figure 1. Architecture comparison. **(a)** RT-DETR-L baseline (HGNetv2 backbone); **(b)** the proposed EMA-enhanced architecture: EMA modules (orange) inserted at the P3/P4/P5 outputs, followed by the AIFI encoder → RepC3 FPN → RTDETRDecoder to produce detections.

3. Experiments

3.1. Datasets and implementation details

3.1.1. Pcb-cropped (main benchmark)

From the Peking University PCB defect image library, 693 original images and six defect classes (open circuit, short, mouse bite, spur, missing hole, spurious copper). The split is first performed at the original-image level in an 8:1:1 ratio. 640×640 Sliding-window cropping is applied independently to avoid data leakage fully, yielding 4 379 cropped images (3 507 train / 439 val / 433 test) and 5 266 annotated instances.

3.1.2. DeepPCB

1 000 training and 500 test images at (binarized CCD image pairs, with a class definition different from pcb-cropped), used only as a second dataset to assess whether the improvement generalizes beyond pcb-cropped [9] 640×640 .

3.1.3. Training configuration

AdamW optimizer; lr = 1e-4; weight decay 1e-4; cosine schedule; 2 warmup epochs; batch 16; 100 epochs (early stopping patience = 30); AMP = True; seed 42. FPS is measured on an NVIDIA RTX 4090 D with batch = 1 (50 warmup + 300 timed runs).

3.1.4. Evaluation metrics

We report mAP_{50} and $mAP_{50:95}$ (COCO-style, averaged over IoU thresholds 0.50:0.05:0.95), together with precision and recall at the default confidence threshold; parameter count, GFLOPs, and FPS serve as efficiency indicators. All reported accuracy numbers are obtained on the held-out test split, never on the validation split used for early stopping, so that no model selection leaks into the test measurement.

3.2. Ablation study

Table [tab:ablation] progressively adds improvement modules starting from the RT-DETR-L baseline.

The EMA-only variant achieves the highest mAP_{50} (97.33%, + 0.95 pp) and $mAP_{50:95}$ (56.45%, + 1.75 pp). The FasterNet variant drops 1.07 pp in accuracy but reduces parameters by 31% (32.82 → 22.78 M), GFLOPs by 40% (108.0 → 65.3), and raises FPS to 11.3. The combined FasterNet+EMA variant (95.64%) falls below both the EMA-only configuration and the baseline, indicating that the restriction of PConv on channel diversity weakens the discriminative feature variance that EMA can exploit.

Figure 2 shows the training process of the proposed EMA variant: all losses converge smoothly, mAP_{50} and $mAP_{50:95}$ stabilize after about 40 epochs with no sign of overfitting, confirming convergence stability under from-scratch training (pretrained = False).

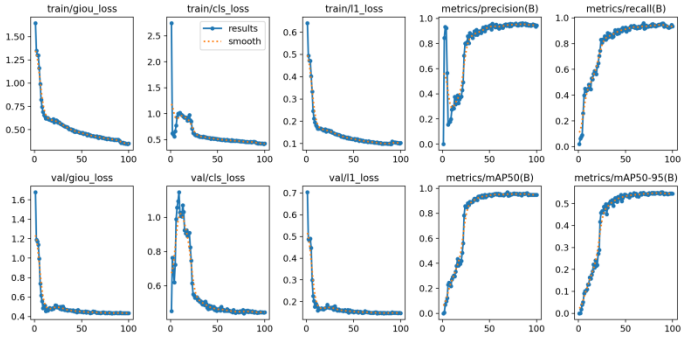

Figure 2. Training curves of the proposed EMA variant on pcb-cropped (100 epochs). Top row: training losses, precision, and recall; bottom row: validation losses, mAP_{50}, and $mAP_{50:95}$.

3.3. Attention mechanism comparison

All five attention mechanisms are integrated at the P3/P4/P5 neck outputs of the original HGNetv2 backbone of RT-DETR-L under identical training conditions.

The mAP_{50} ranking is EMA (97.33%) > ECA (96.69%) > None (96.38%) > CBAM (95.55%) > SE (95.43%) > CA (93.04%), with a maximum gap of **4.29 pp**. CA shows a notable drop (−3.34 pp); we tentatively attribute this to a conflict between its horizontal–vertical coordinate-decomposed encoding and the holistic spatial query mechanism of the AIFI encoder, with the exact mechanism awaiting further verification via feature-map visualization and similar means.

Two broader patterns emerge. First, only EMA and ECA surpass the no-attention baseline, whereas SE, CBAM, and CA all degrade it. This suggests that, at the neck of a DETR-style detector, attention that compresses or sequentially reshapes the channel description—SE squeezes channels, CBAM applies channel-

then-spatial attention in sequence, and CA factorizes space into two 1-D directions—can disturb the feature distribution that the downstream AIFI encoder expects. Second, the two mechanisms share a common trait: they preserve the full channel description. ECA does so through a parameter-free 1-D local cross-channel interaction, leaving GFLOPs identical to the baseline, while EMA does so through parallel channel and multi-scale spatial branches without channel reduction. EMA's extra multi-scale spatial modeling explains why it edges out ECA (+0.64 pp mAP_{50}) at the cost of 11.5 more GFLOPs, making ECA the preferable choice when compute is constrained and EMA the preferable choice when accuracy is paramount [10] .

3.4. Comparison with state-of-the-art methods

† Cited from the original paper, using PKU-PCB original images (no sliding-window cropping); FPS measured on different hardware, not directly comparable, provided for method reference only.

The proposed EMA variant achieves the highest mAP_{50} (97.33%) and $mAP_{50:95}$ (56.45%) under the unified protocol, slightly above YOLOv8l (97.32%/55.69%). The end-to-end FPS of the YOLO series (76.7–112.7) is markedly higher than that of the RT-DETR series (8.7–11.3), mainly because NMS post-processing is highly engineering-optimized; the FasterNet variant (11.3 FPS) has the highest throughput among the RT-DETR family.

3.5. Cross-dataset generalization: DeepPCB

Table 1. DeepPCB cross-dataset generalization results (500 test images; all three models trained from scratch on the DeepPCB training set)

Method	Params (M)	GFLOPs	FPS	mAP_{50} (%)	$mAP_{50:95}$ (%)
RT-DETR-L (baseline)	32.82	108.0	8.5	84.50	57.55
RT-DETR-L+EMA (ours)	33.68	119.5	8.8	**89.33**	**61.08**
FasterNet+P2	26.03	152.0	9.4	85.66	58.89
EMA vs. baseline	+0.86	+11.5	+0.3	+4.83	+3.53

When all three models are trained from scratch on DeepPCB, the EMA variant reaches 89.33% mAP_{50} (+4.83 pp) and improves $mAP_{50:95}$ by 3.53 pp. Because DeepPCB differs from pcb-cropped in imaging modality and class definition, this indicates that the gain from neck-level EMA attention is not specific to a single dataset but holds on a second, distinct dataset. Unlike most existing RT-DETR-type PCB studies that report results on only one dataset, our multi-variant comparison across two datasets shows that the choice of attention mechanism contributes robustly to detection quality. Note that FasterNet+P2 has the highest GFLOPs (152.0) in this table because the high-resolution P2 head (160×160) adds substantially more computation than the FasterNet backbone saves.

3.6. Qualitative results and error analysis

Figure 3 shows the normalized confusion matrix of the proposed EMA variant on the pcb-cropped test set: the diagonal recall of all six classes is no lower than 0.96 (short reaches 1.00), indicating strong inter-class discrimination; the main errors come from the background column (spurious_copper missed at 0.32, open_circuit at 0.18), reflecting that confusing tiny, low-contrast defects with copper-foil texture remains a core

challenge in PCB detection.

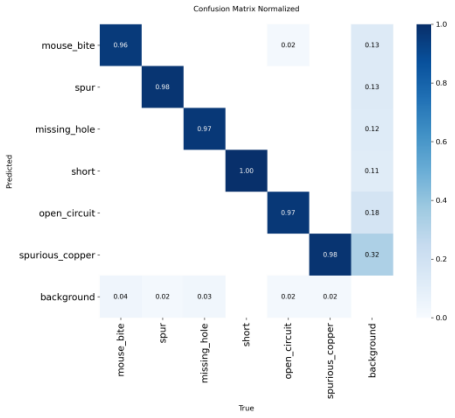

Figure 3. Normalized confusion matrix of the proposed EMA variant on the PCB-cropped test set. The diagonal gives per-class recall; the rightmost column gives the proportion missed as background.

4. Conclusion

Based on RT-DETR-L, this paper systematically compares five neck attention mechanisms and reveals a 4.29 pp performance gap: EMA achieves the best mAP_{50} (97.33%) and $mAP_{50:95}$ (56.45%), both surpassing the baseline; CA drops 3.34 pp, tentatively attributed to an architectural conflict with the AIFI encoder; FasterNet trades 1.07 pp of accuracy for 31% fewer parameters and 40% lower GFLOPs; and the EMA gain also holds on the second dataset DeepPCB (89.33%, 4.83 pp over the baseline). Future work will study the FasterNet–EMA negative interaction and knowledge distillation for edge deployment.

Data availability

pcb-cropped is available from the author upon reasonable request; DeepPCB is publicly available (https://github.com/tangsanli5201/DeepPCB).

Author contributions

T.H. conceived the study, designed and performed the experiments, analyzed the data, and wrote the manuscript.

Disclosure statement

The author declares no conflict of interest.

References

[1] Ling Q, Isa N, 2023, Printed Circuit Board Defect Detection Methods Based on Image Processing, Machine Learning and Deep Learning: A Survey. IEEE Access, 11: 15921–15944.

[2] Zhao Y, Lv W, Xu S, et al., 2024, DETRs Beat YOLOs on Real-Time Object Detection. CVPR.

[3] Wang Y, Wu B, Zhang L, et al., 2025, Enhanced PCB Defect Detection Via HSA-RTDETR on RT-DETR. Scientific Reports, 15: 31783.

[4] Ouyang D, He S, Zhang G, et al., 2023, Efficient Multi-Scale Attention Module with Cross-Spatial Learning. ICASSP.

[5] Hu J, Shen L, Sun G, 2018, Squeeze-and-Excitation Networks. CVPR: 7132–7141.

[6] Woo S, Park J, Lee J, et al., 2018, CBAM: Convolutional Block Attention Module. ECCV: 3–19.

[7] Hou Q, Zhou D, Feng J, 2021, Coordinate Attention for Efficient Mobile Network Design. CVPR: 13713–13722.

[8] Chen J, Kao S, He H, et al., 2023, Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. CVPR: 12021–12031.

[9] Tang S, He F, Huang X, et al., 2019, On-Line PCB Defect Detector on a New PCB Defect Dataset. arXiv: 1902.06197.

[10] Wang Q, Wu B, Zhu P, et al., 2020, ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. CVPR: 11534–11542.