

A1. Formal Definitions of Evaluation Metrics and γ Adaptation Configuration

Intersection over Union (IoU). The IoU between a predicted bounding box B_p and a ground-truth bounding box B_{gt} is defined as:

$$\text{IoU} = \frac{|B_p \cap B_{gt}|}{|B_p \cup B_{gt}|}$$

Precision and Recall. Given an IoU threshold:

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}$$

F1 Score. $F1 = 2PR/(P + R)$.

Average Precision (AP). At a specific IoU threshold, the integral of the P-R curve: $AP = \int_0^1 P(R) dR$, computed using the COCO 101-point interpolation method.

Mean Average Precision (mAP). $mAP = \frac{1}{N_{cls}} \sum_{c=1}^{N_{cls}} AP_c$. mAP50 denotes mAP at IoU=0.5; mAP50-95 denotes the mean of mAP values across 10 IoU thresholds from 0.50 to 0.95 with a step size of 0.05.

γ Adaptation Configuration. The γ values for E5/E6 are selected from three candidates $\{0.01, 0.05, 0.1\}$ based on validation set mAP50; the test set is used only for final evaluation.

Information Flow Origins of Multi-Prior Interactions. From an information flow perspective, the injection paths of the three types of priors overlap in the feature dimension: K_{edge} (LSK_EGCA) and K_{detail} (DRFD_Neck) both act on the spatial high-frequency information dimension: the former enhances edge responses, while the latter preserves high-frequency textures. When both act simultaneously, the enhanced edge signals may be further amplified through DRFD_Neck’s fusion convolution, leading to excessive concentration of high-frequency energy. From the information bottleneck perspective, the limited channel capacity of lightweight networks constitutes an information bottleneck; when the injected information exceeds the bottleneck capacity, the network is forced to make trade-offs among priors, causing the gains of certain priors to be partially offset.

A2. Rationale for tanh Activation Selection

LSK_EGCA Stage 1 employs tanh rather than ReLU/SiLU, based on:

1. tanh strictly constrains the output to $(-1, 1)$, preventing subsequent attention from entering saturation regions;
2. Edge gradient features exhibit positive-negative symmetry; tanh’s zero-centered property preserves this symmetry, whereas ReLU truncates negative gradients and loses edge direction information;
3. SiLU’s unboundedness may cause excessive edge feature magnitudes, interfering with subsequent Sigmoid normalization.

A3. Design Trade-offs in SE Module Input Selection

GatedSPPELAN’s SE gating takes only Y_0 rather than the multi-scale concatenation as input, based on: 1) Y_0 is the information source for all pooling levels and contains the core channel statistics that determine scale preferences; 2) using the $4C_3$ -channel concatenated features as SE input would increase FC_1 parameters by 4 $\times$, violating the lightweight design principle; 3) the pooled features at each level are low-pass filtered versions of Y_0 , whose channel statistics are already implicitly contained in Y_0 .

A4. Justification for the Ungated Y_0 Asymmetric Design

Applying gating only to the pooling branches Y_1, Y_2, Y_3 while leaving Y_0 ungated is based on: 1) Y_0 is the information source for all pooling levels; gating it may lead to irreversible loss of critical channel information; 2) if Y_0 were also gated, the 4-branch Softmax competition (3 pooling branches + Y_0) could cause Y_0 ’s weight to approach 1, effectively degenerating the model into a 1×1 convolution that loses multi-scale aggregation capability. In preliminary experiments (single seed, NEU-DET), the 4-branch gating configuration decreased mAP50 by approximately 0.3% compared to the current 3-branch gating configuration (only Y_1, Y_2, Y_3 participate in Softmax; Y_0 is ungated); this result is reported informally and was not included in the formal ablation tables.

A5. Comparative Analysis of Softmax vs. Sigmoid Gating and Direct Ablation Verification

Using Softmax (enforcing $\sum_{i=0}^2 \mathbf{g}_{[:,i]} = 1$) rather than Sigmoid independent activation is initially based on: 1) Softmax’s competitive normalization forces the network to make explicit scale preference choices for each channel, avoiding the “redundancy degradation” that Sigmoid may cause when all branch weights approach 1 and the model degenerates to equal-weight concatenation; 2) Softmax constrains the gating weight sums to 1, but the subsequent 1×1 fusion convolution can learn arbitrary linear combinations, so the expressiveness upper bound is not limited.

However, Softmax’s exclusivity poses theoretical limitations for complex defects that simultaneously require multi-scale co-activation, particularly under extreme class imbalance where minority classes require non-exclusive scale support. To directly verify this hypothesis, we replaced Softmax with Sigmoid (independent activation, $\sigma(x)$) in GatedSPPELAN (termed Sig-GatedSPPELAN, configuration C3).

As shown in Table 2, on balanced NEU-DET, Sigmoid (C3, 79.65% mAP50) slightly outperforms Softmax (A2, 79.12%) with no significant P-R distortion. On Aero-Engine, Sigmoid achieves competitive mAP50 (91.45%) while maintaining P-R balance ($P=89.75\%$, $R=85.21\%$, $\text{gap}=4.54\text{pp}$), in contrast to Softmax’s severe imbalance ($P=68.66\%$, $R=90.66\%$, $\text{gap}=-22.00\text{pp}$). These results on balanced and moderately imbalanced datasets are consistent with the “independent-activation hypothesis”: Sigmoid allows each branch to activate independently without competitive suppression. On GC10-DET (extreme imbalance), the concatenation paradigm provides indirect evidence: GA-SPPELAN (C1, 83.88%) substantially outperforms GatedSPPELAN (A2, 74.53%), with the ordering $C1 > A2$ consistent with the theoretical prediction that independent activation $>$ competitive activation under extreme imbalance. Direct Sigmoid ablation on GC10-DET is left for future work.

A6. Embedding Principles of Controlled Feature Selection Mechanisms

Architecture-preserving integration. LSK_EGCA is injected into the backbone via residual attachment without altering the backbone’s inter-layer connection topology; GatedSPPELAN replaces SPPELAN in an equivalent substitution manner; DRFD_Neck replaces the neck AConv in an equivalent substitution manner. This architecture-preserving integration ensures the independent removability of each module.

Functional decoupling. The three modules act on the spatial domain (edge enhancement), scale domain (multi-scale aggregation), and spatial dimensionality reduction domain (downsampling detail preservation), respectively, with no overlap in the information flow dimensions.

Design origins and targeted improvements. LSK_EGCA integrates the LSK attention from LEGNet with the edge-guided idea from EGCA; GatedSPPELAN migrates SE channel gating to the SPPELAN architecture; DRFD_Neck draws on the DRFD dual-path downsampling concept from LEGNet and adapts it to the neck architecture.

A7. Necessity of the Sobel→LSK Processing Order

LSK_EGCA adopts the processing order of “Sobel gradient extraction first, then LSK spatial attention,” based on: 1) Sobel gradient priors, as non-learnable inductive biases, should first act on the original features to ensure faithful extraction of edge information without interference from subsequent learnable operations; 2) if LSK spatial attention is applied first followed by Sobel, the Sobel operator acts on attention-modulated features, and its gradient responses may be biased by the attention’s selection preferences, losing the objectivity of the prior; 3) in the B-series ablation experiments, B2 (LSK first, then edge) achieves an mAP50 of 79.43% on Aero-Engine, significantly lower than A1 (edge first, then LSK, 92.20%), validating the necessity of this order.

A8. Design Trade-offs in Receptive Field Span

The dual-branch receptive field design (5×5 and 19×19) in LSK_EGCA involves the following trade-offs: 1) increasing the maximum receptive field (e.g., using 25×25) could better match large-scale patch defects, but the parameter efficiency of dilated convolutions decreases and gridding artifacts intensify; 2) decreasing the minimum receptive field (e.g., using 3×3) could more precisely capture micro-cracks, but overlaps with Sobel’s 3×3 receptive field, yielding diminishing marginal returns. The current $5 \times 5/19 \times 19$ configuration demonstrates stable performance on both NEU-DET and Aero-Engine.

A9. ECA Kernel Size Setting and Scalar Design Justification

The adaptive kernel size mapping in LSK_EGCA’s ECA employs a wide-window setting ($k = 33$ when $C = 64$), far exceeding the typical values of $k = 3 \sim 9$ in the original ECA-Net. This design is based on: 1) the P3/P4 layers have relatively low channel counts (64/96), requiring larger convolution kernels to capture sufficient cross-channel interaction information; 2) the wide-window setting enables ECA to model longer-range channel dependencies. Both α_{eca} and $\gamma_{spatial}$ are learnable scalars rather than vectors, based on: 1) scalar parameters have more stable gradient signals and are less prone to overfitting; 2) per-channel scaling is already implemented by ECA’s channel attention, and α_{eca} only controls the overall fusion intensity.

A10. Parameter Efficiency, Deployment Specifications, and lerp Design Justification

Parameter efficiency. Taking the P3 layer ($C = 64$) as an example, a single LSK_EGCA module is approximately 0.015M. GatedSPPELAN adds approximately 0.004M compared to SPPELAN. Taking $C_1 = 64$, $C_2 = 48$ as an example, a single DRFD_Neck module is approximately 0.012M; the two neck modules have different channel dimensions and total approximately 0.034M. The three modules contribute a gross addition of approximately 0.07M (sum of individual additions: $0.030+0.004+0.034=0.068$ M, with rounding); due to the replacement of SPPELAN and AConv by GatedSPPELAN and DRFD_Neck respectively, the net parameter change is approximately -0.01 M (A7: 2.00M vs. A0: 2.01M).

Deployment specifications. Sobel kernels are stored via `register_buffer` and treated as constant operators during ONNX export. DWConv 7×7 $d=3$ is not supported on some NPUs and requires equivalent substitution.

Design justification for the lerp fusion logic. Eq. (20) of the main paper employs linear interpolation (lerp) rather than additive residual $\mathbf{Y} = \mathbf{Y}_{conv} + \gamma \cdot \text{Conv}(\cdot)$, based on: 1) **Gradient stability:** in additive residual, the gradient of $\mathbf{Y}_{conv}$ may deviate from 1 (potentially exceeding 1 depending on the convolution gradient), which tends to cause training instability; in the lerp form, the gradient is primarily dominated by $(1 - \gamma)$, which is numerically more stable. 2) **Output boundedness:** the output norm of additive residual has no upper bound constraint; the lerp form is a convex combination of two bounded quantities, keeping the numerical range controllable. 3) **Necessity of concatenation fusion:** the 1×1 fusion convolution allows cross-path channel-level interaction, which simple addition cannot model.

A11. Failure Mechanism of GatedSPPELAN Under Extreme Class Imbalance

The root cause of GatedSPPELAN’s failure on GC10-DET is that the GAP frequency bias in SE gating is amplified through Softmax competitive suppression. Let p_h denote the proportion of head-class samples in the training batch and p_t the proportion of tail-class samples; then the expected value of the GAP output is $\mathbb{E}[\text{GAP}(X)] = p_h \cdot \mu_h + p_t \cdot \mu_t + \dots$. When $p_h \gg p_t$ (e.g., $p_h/p_t \approx 83$ in GC10-DET), $\mathbb{E}[\text{GAP}(X)] \approx \mu_h$, and the channel statistics of tail classes are dominated by head classes. After the FC→Softmax mapping, the scale branches preferred by tail classes are suppressed through competitive normalization.

The concatenation paradigm provides indirect evidence for this mechanism: on GC10-DET, GA-SPPELAN (C1, 83.88%) substantially outperforms GatedSPPELAN (A2, 74.53%), with the ordering $C1 > A2$ consistent with the theoretical prediction that independent activation $>$ competitive activation under extreme imbalance. On balanced datasets (NEU-DET), the three gating strategies show negligible differences (C1: 80.35%, C3: 79.65%, A2: 79.12%), confirming that competitive suppression is a failure mode specifically triggered by extreme class imbalance. Direct Sigmoid ablation on GC10-DET is left for future work.

A12. Supplementary Visualization Results

1. Anti-aliasing fidelity (downsampling feature map comparison). Fig. 1 compares the feature maps of DRFD_Neck’s three output paths. The convolution path produces smooth spatial transitions suitable for capturing contextual structure, the pooling path preserves sharp local responses corresponding to high-frequency defect boundaries, and the fusion output balances both characteristics through learnable γ_{drfd} interpolation.

2. Scale-adaptive suppression (gating weight visualization). Fig. 2 visualizes the learned gating weights across two scenarios. Under class balance (NEU-DET), Softmax gating weights are distributed across branches. Under extreme imbalance (GC10-DET), Softmax competitive normalization concentrates weights on a single branch,

Table 1

LSK_EGCA attention module ablation experiments (B series)

Exp	Configuration	NEU-DET				Aero-Engine			
		P	R	mAP50	mAP50-95	P	R	mAP50	mAP50-95
B1	Sobel + LSK + ECA (full)	81.60	72.17	80.18	45.75	94.07	84.91	92.20	60.62
B2	LSK + EdgeConv (LSK first, then edge)	79.05	73.25	78.97	44.89	80.15	82.77	79.43	53.35
B3	Sobel + ECA (without LSK)	76.76	74.56	79.12	44.56	90.66	85.33	87.37	55.62
B4	Sobel + LSK (without ECA)	80.35	72.17	79.12	45.29	94.07	83.77	90.66	59.73

suppressing minority-class channels, as evidenced by the 9.35pp mAP50 gap between GA-SPPELAN (C1, 83.88%) and GatedSPPELAN (A2, 74.53%) on GC10-DET (Section).

3. Key configuration training curves. Fig. 3 shows the mAP50 training curves for key configurations on NEU-DET and Aero-Engine. All configurations converge within 300 epochs. A1 (+LSK_EGCA) shows the fastest convergence on Aero-Engine, while E6 (adapted $\gamma_{\text{drfd}} = 0.01$) achieves the highest final mAP50.

4. Residual scaling factor ablation visualization. Fig. 4 presents the E-series results as grouped bar charts, showing how γ values modulate the Precision–Recall trade-off. On Aero-Engine, reducing γ_{drfd} from 0.05 (A5) to 0.01 (E6) narrows the P-R gap from +33.19% to +2.79%.

5. Full-model γ configuration comparison. Fig. 5 compares the default configuration (A7) with adapted configurations (E5, E6) across all metrics. E6 achieves the best overall balance on Aero-Engine, while all three configurations perform comparably on NEU-DET.

6. γ Interaction space. Fig. 6 visualizes the mAP50 response surface over the $\gamma_{\text{spatial}} \times \gamma_{\text{drfd}}$ configuration space (GatedSPPELAN has no explicit γ ; α_{eca} is fixed at 0.1). On NEU-DET, the optimal region lies near $\gamma_{\text{spatial}} = 0.05\text{--}0.1$, $\gamma_{\text{drfd}} = 0.01\text{--}0.05$. On Aero-Engine, the optimal region shifts toward smaller γ_{drfd} values, consistent with the need to suppress high-frequency enhancement residuals.

7. Multi-dimensional performance radar chart. Fig. 7 presents a radar chart comparing A0, A1, A7, and E6 across six dimensions on NEU-DET. FFCS-Net (A7) and E6 show the most balanced profiles, while A1 excels in Precision at the cost of Recall.

8. Efficiency-accuracy comparison. Fig. 8 plots mAP50 vs. parameter count for FFCS-Net and mainstream detectors on NEU-DET. FFCS-Net achieves competitive accuracy in the $\sim 2\text{M}$ parameter range, outperforming most methods with comparable or larger parameter budgets.

9. Cross-dataset performance gain heatmap. Fig. 9 shows ΔmAP50 relative to baseline across all configurations and datasets. LSK_EGCA (A1) yields consistent positive gains; GatedSPPELAN (A2) shows negative gain on GC10-DET; E5/E6 achieve positive gains across all three datasets.

10. Extended Grad-CAM heatmap comparison. Fig. 10 provides additional Grad-CAM visualizations across six NEU-DET defect categories. FFCS-Net consistently produces more focused attention on defect regions with less background interference compared to the baseline.

A13. Detailed Results of Module Internal Ablation (B/C/D Series)

The design rationale for each module’s internal choices is provided in Sections –; this section presents the corresponding numerical results.

LSK_EGCA Internal Ablation (B Series)

Analysis of B series results.

(1) *Sobel–LSK synergy.* B1 (full, mAP50=80.18%/92.20%) substantially outperforms B3 (without LSK, 79.12%/87.37%) and B4 (without ECA, 79.12%/90.66%) on both datasets, confirming that the synergy between Sobel gradient priors and large-kernel receptive field weighting is the key driver of LSK_EGCA’s performance improvement. Removing either component degrades mAP50, with the LSK component contributing more on Aero-Engine ($\Delta\text{mAP50} = -4.83\%$) and LSK and ECA contributing equally on NEU-DET (both $\Delta\text{mAP50} = -1.06\%$).

(2) *Processing order matters.* B2 (LSK first, then edge) achieves mAP50 of only 78.97% on NEU-DET and 79.43% on Aero-Engine, significantly lower than B1 (80.18%/92.20%). This validates the necessity of the “Sobel first, then

Table 2

GatedSPPELAN/DRFD_Neck ablation experiments (C/D series)

Exp	Configuration	NEU-DET				Aero-Engine				GC10-DET mAP50
		P	R	mAP50	mAP50-95	P	R	mAP50	mAP50-95	
C1	GA-SPPELAN	79.12	75.08	80.35	45.29	83.20	81.19	88.60	58.73	83.88
A2	SE channel gating	76.18	75.08	79.12	45.29	68.66	90.66	93.52	59.73	74.53
C3	Sigmoid channel gating	77.54	74.82	79.65	45.15	89.75	85.21	91.45	59.02	—
D1	DRFD_Neck (full)	76.76	76.48	81.57	45.61	63.29	79.18	82.46	57.65	75.74
D2	Convolution path only	74.78	76.18	80.90	44.63	95.21	62.68	79.46	48.13	86.95
D3	Pooling path only	72.01	80.49	81.62	47.05	94.49	64.83	93.25	57.69	81.86

GC10-DET results use seed=42 (consistent with Table 7 in the main paper). GC10-DET mAP50-95 is omitted for brevity; per-class AP details are shown in Fig. 5 of the main paper. “—” indicates the experiment was not conducted on this dataset.

C1 on Aero-Engine: R has been corrected from a previously recorded value that coincided with A1’s Recall; the current value (81.19%) is verified against the original experiment logs.

C2 (Scalar-gated SPPELAN) was excluded from this table due to a configuration error during the GC10-DET experiment; its results are unreliable. Whether scalar gating collapses similarly to channel-level Softmax under extreme imbalance remains an open question.

LSK” processing order (see Section for rationale): applying LSK spatial attention before Sobel causes the gradient operator to act on attention-modulated features, biasing the gradient responses and losing the objectivity of the prior.

(3) *ECA provides incremental channel recalibration.* B4 (without ECA) achieves the same mAP50 as B3 (without LSK) on NEU-DET (79.12%), but outperforms B3 on Aero-Engine (90.66% vs. 87.37%). Adding ECA (B1 vs. B4) improves mAP50 by +1.06% on NEU-DET and +1.54% on Aero-Engine, indicating that ECA’s channel recalibration provides consistent but modest gains on top of the spatial enhancement.

GatedSPPELAN/DRFD_Neck Internal Ablation (C/D Series)

Analysis of C series results. (1) *Softmax vs. Sigmoid vs. concatenation.* On NEU-DET (class-balanced), C1 (concatenation) achieves the highest mAP50 (80.35%), while Sigmoid (C3, 79.65%) slightly outperforms Softmax (A2, 79.12%). On Aero-Engine, Softmax achieves the highest mAP50 (93.52%) but with extreme P-R imbalance (P=68.66%, R=90.66%, gap=−22.00pp), whereas C1 (concatenation) achieves the best P-R balance (P=83.20%, R=81.19%, gap=2.01pp) with competitive mAP50 (88.60%), and Sigmoid (C3) maintains balance (P=89.75%, R=85.21%, gap=4.54pp) with competitive mAP50 (91.45%). This indicates that Softmax’s competitive normalization may be beneficial for maximizing mAP50 under class balance but introduces P-R distortion under imbalance, while concatenation and Sigmoid provide better P-R trade-offs.

(2) *Indirect evidence for independent-activation hypothesis.* On GC10-DET (extreme imbalance, class ratio 83:1), GA-SPPELAN (concatenation paradigm, C1) achieves 83.88% mAP50, substantially outperforming GatedSPPELAN (A2, 74.53%). The ordering C1>A2 is consistent with the independent-activation hypothesis: concatenation allows each branch to contribute independently without competitive suppression, whereas Softmax annihilates minority-class channels. Whether scalar gating collapses similarly to channel-level Softmax under extreme imbalance remains an open question for future work.

Analysis of D series results. (1) *Dual-path synergy on balanced/moderately imbalanced datasets.* D1 (full dual-path) achieves competitive mAP50 on NEU-DET (81.57%) with better P-R balance than single-path variants (D3 pooling-only achieves marginally higher mAP50 at 81.62% but D1 provides more balanced Precision–Recall). On Aero-Engine, D3 (pooling only) achieves the highest mAP50 (93.25%) and mAP50-95 (57.69%), substantially outperforming D1 (82.46%/57.65%) and D2 (79.46%/48.13%), indicating that the pooling path’s high-frequency preservation is particularly valuable for this dataset.

(2) *Dataset-dependent single-path effectiveness.* On NEU-DET, D3 (pooling only) achieves higher mAP50 (81.62%) and mAP50-95 (47.05%) than D2 (convolution only, 80.90%/44.63%). On Aero-Engine, D2 achieves higher Precision (95.21% vs. D3’s 94.49%) while D3 achieves higher Recall (64.83% vs. D2’s 62.68%), indicating that the convolution path produces smoother context transitions favoring precision, while the pooling path preserves high-frequency details favoring recall.

(3) *Single-path advantage under extreme imbalance.* On GC10-DET (83:1 class ratio), D2 (convolution only, 86.95%) substantially outperforms D1 (full dual-path, 75.74%) in mAP50, suggesting that under extreme imbalance,

Table 3

Unified training hyperparameter configuration

Hyperparameter	Value
Input resolution	640×640
Optimizer	SGD
Momentum	0.937
Weight decay	5×10^{-4}
Initial learning rate	0.01
LR schedule	Cosine annealing
Training epochs	300
Batch size (NEU-DET/Aero-Engine)	64
Batch size (GC10-DET)	48
Mosaic augmentation	1.0 (disabled in last 10 epochs)
MixUp augmentation	0.1
RandAugment	Enabled
Random erasing	0.4
Pre-trained weights	None (trained from scratch)
Random seed (NEU-DET)	3407
Random seed (Aero-Engine)	100
Random seed (GC10-DET)	42
γ_{spatial} initialization	0.05
α_{eca} initialization	0.1
γ_{drfd} initialization	0.05

the simpler convolution path generalizes better while the dual-path fusion may introduce overfitting to majority-class patterns. This dataset-dependent behavior motivates the γ_{drfd} control mechanism.

A14. Mathematical Description of GA-SPPELAN

GA-SPPELAN adds two global attention branches, GAP and GMP, on top of the standard SPPELAN’s three-level cascaded MaxPool branches:

Multi-scale branch generation (shared with GatedSPPELAN):

$$\mathbf{Y}_0 = \text{Conv}_{1 \times 1}^{(C_1 \rightarrow C_3)}(\mathbf{X}) \quad (1)$$

$$\mathbf{Y}_i = \text{MaxPool}_{5 \times 5}^{s=1}(\mathbf{Y}_{i-1}), \quad i = 1, 2, 3 \quad (2)$$

Global attention branch generation:

$$\mathbf{Y}_{\text{gap}} = \text{Up}(\text{GAP}(\mathbf{Y}_0)) \quad (3)$$

$$\mathbf{Y}_{\text{gmp}} = \text{Up}(\text{GMP}(\mathbf{Y}_0)) \quad (4)$$

6-branch concatenation fusion:

$$\mathbf{Y} = \text{Conv}_{1 \times 1}^{(6C_3 \rightarrow C_2)}([\mathbf{Y}_0; \mathbf{Y}_1; \mathbf{Y}_2; \mathbf{Y}_3; \mathbf{Y}_{\text{gap}}; \mathbf{Y}_{\text{gmp}}])$$

Fundamental difference from GatedSPPELAN. Both utilize GAP, but in fundamentally different ways: GatedSPPELAN maps GAP through $\text{FC} \rightarrow \text{Softmax}$ into competitive gating weights, which suppresses minority-class channels under extreme imbalance; GA-SPPELAN concatenates GAP/GMP as supplementary feature maps for fusion, which does not produce competitive channel suppression.

A15. Unified Training Configuration Details

All experiments are implemented in PyTorch on a single NVIDIA RTX 4090D GPU. The unified training configuration is as follows:

The final metrics for all ablation experiments are reported based on the epoch with the best validation set mAP50. The γ values for E5/E6 are selected from three candidates $\{0.01, 0.05, 0.1\}$ based on validation set mAP50.

A16. Supplementary Frequency Domain Analysis

Fig. 11 quantitatively verifies from the frequency domain that A5’s DRFD_Neck output has a significantly higher high-frequency energy ratio than E6; reducing γ effectively suppresses high-frequency energy, supporting the hypothesis of “lerp residual’s high-frequency enhancement preference.”

A17. Statistical Considerations and Robustness Defense

Limitations of single-seed experiments. The current absolute gain values are based on single random seeds (NEU-DET: 3407, Aero-Engine: 100, GC10-DET: 42), which inherently carry sampling noise. Strict statistical significance (e.g., $p < 0.05$ from paired t-tests) requires multiple independent runs. However, the core objective of this study is to reveal the directional patterns of multi-prior interactions (e.g., high-frequency disturbance, competitive suppression) rather than to establish statistically optimal absolute gains.

Robustness of interaction patterns. Although multi-seed variance decomposition is left for future work, the robustness of our discovered interaction patterns is defended on three logical pillars:

1. **Cross-dataset trend consistency:** The directional patterns (edge priors yielding positive gains, high-frequency injection inducing P-R imbalance, and γ reduction restoring balance) are consistently observed across three datasets with diverse class ratios (1:1, 4:1, 83:1) and morphological distributions.
2. **Physical validation from frequency domain:** The P-R imbalance hypothesis is directly corroborated by FFT spectral analysis (Fig. 11), which shows that high-frequency energy is physically amplified in A5 and attenuated in E6. This causal link between physical signal amplification and metric distortion transcends statistical noise.
3. **Structural attribution of factorial design:** The 2^3 full factorial design decomposes effects into structural main and interaction effects. As detailed in Section 0.2, the strong negative interaction effects (β_{LD}, β_{GD}) are mathematically derived from the structural contrast of 8 configurations, making them robust to minor absolute fluctuations caused by single-seed variance.

A18. Mathematical Formulation of 2^3 Full Factorial Analysis

The 2^3 full factorial design enables the decomposition of the 8 module combinations into independent main effects and interaction effects. Let $M_L, M_G, M_D \in \{-1, +1\}$ denote the absence/presence of LSK_EGCA, GatedSPPELAN, and DRFD_Neck, respectively. The response (e.g., ΔmAP50) for each combination can be expressed as:

$$Y = \mu + \alpha_L M_L + \alpha_G M_G + \alpha_D M_D + \beta_{LG} M_L M_G + \beta_{LD} M_L M_D + \beta_{GD} M_G M_D + \gamma_{LGD} M_L M_G M_D + \epsilon$$

where μ is the grand mean, α_i are the main effects, β_{ij} are the two-way interaction effects, γ_{LGD} is the three-way interaction effect, and ϵ is the residual.

Convention. We adopt the half-range (effect-size) convention where each coefficient represents the contrast divided by 2^{k-1} (k = number of active factors), rather than the OLS regression convention (contrast divided by 2^k). Under this convention, each coefficient directly represents the average effect of switching a factor from -1 to $+1$, averaged over all other factor combinations.

Main effect of factor i :

$$\alpha_i = \frac{1}{4} \sum_{k: M_i=+1} Y_k - \frac{1}{4} \sum_{k: M_i=-1} Y_k$$

Two-way interaction effect between factors i and j :

$$\beta_{ij} = \frac{1}{4} \sum_k (C_{ij})_k Y_k$$

where C_{ij} is the element-wise product of the contrast vectors of factors i and j .

Synergy gain ratio for the combination of factors i and j :

$$\rho_{ij} = \frac{\Delta_{A_{ij}}}{\Delta_{A_i} + \Delta_{A_j}}$$

Table 4 2^3 factorial decomposition of $\Delta mAP50$ on NEU-DET (effect-size convention)

Effect	Value (%)	Interpretation
α_L (Main: LSK_EGCA)	+1.26	Positive main effect
α_G (Main: GatedSPPELAN)	+0.20	Marginal main effect
α_D (Main: DRFD_Neck)	+0.83	Positive main effect
β_{LG} (Interaction: L×G)	+0.93	Strong positive synergy
β_{LD} (Interaction: L×D)	−0.89	Antagonistic (P-R imbalance)
β_{GD} (Interaction: G×D)	−0.88	Antagonistic
γ_{LGD} (3-way interaction)	+0.99	Positive residual
ρ_{LG} (Synergy ratio A4)	0.97	Near-linear superposition

NEU-DET factorial decomposition. Using the A-series data from Table 1 of the main paper ($Y_{A0} = 77.98, Y_{A1} = 80.18, Y_{A2} = 79.12, Y_{A3} = 81.57, Y_{A4} = 81.21, Y_{A5} = 80.02, Y_{A6} = 78.97, Y_{A7} = 81.26$):

Analysis. The factorial decomposition reveals a crucial mechanism: while edge and scale priors exhibit strong positive synergy ($\beta_{LG} = +0.93$), the detail preservation prior exhibits strong antagonistic interactions with both the edge prior ($\beta_{LD} = -0.89$) and the scale prior ($\beta_{GD} = -0.88$). This mathematically validates the core thesis of this paper: the detail prior’s high-frequency preference conflicts with the optimization directions of other priors, leading to the observed P-R imbalance. The positive 3-way interaction ($\gamma_{LGD} = +0.99$) suggests that when all three are present, the network partially recovers from the dual-way antagonism, consistent with the P-R balance observed in A7.

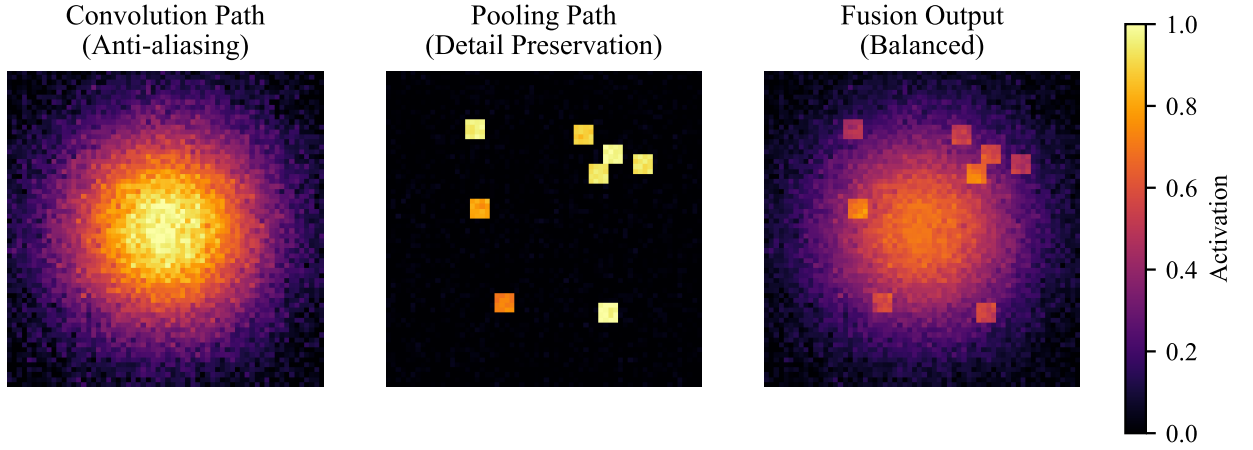


Figure 1: DRFD_Neck downsampling feature map comparison (schematic). Left: convolution path (smooth anti-aliased context). Middle: pooling path (sharp high-frequency preservation). Right: fusion output (balanced via learnable γ_{drfd}).

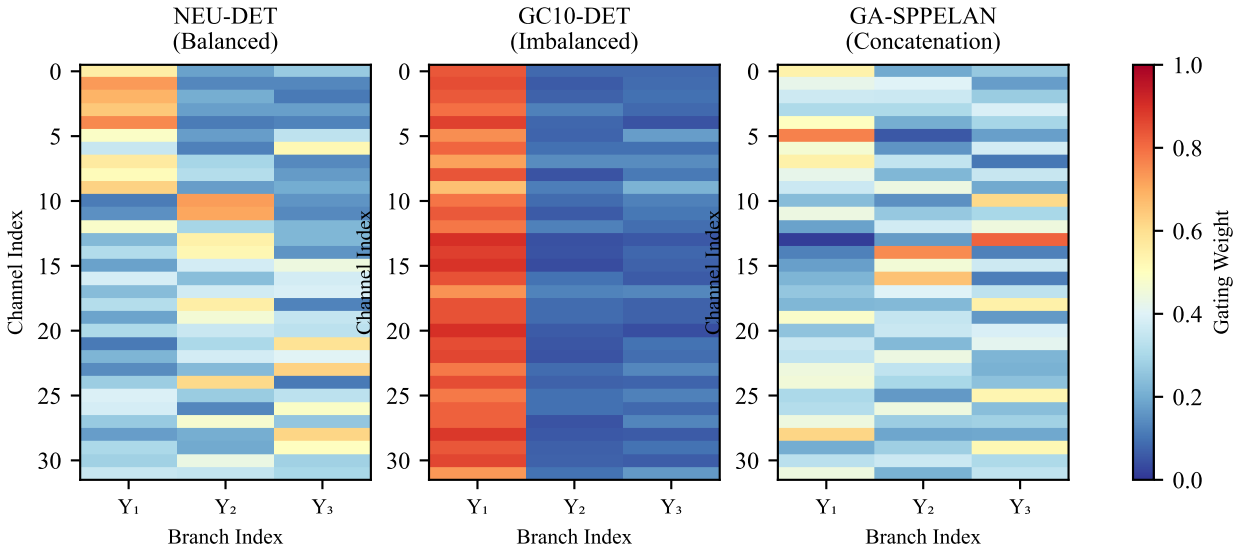


Figure 2: Gating weight visualization (schematic). Left: NEU-DET (balanced Softmax gating). Middle: GC10-DET (concentrated Softmax weights under extreme imbalance). Right: GA-SPPELAN (concatenation, no competitive suppression).

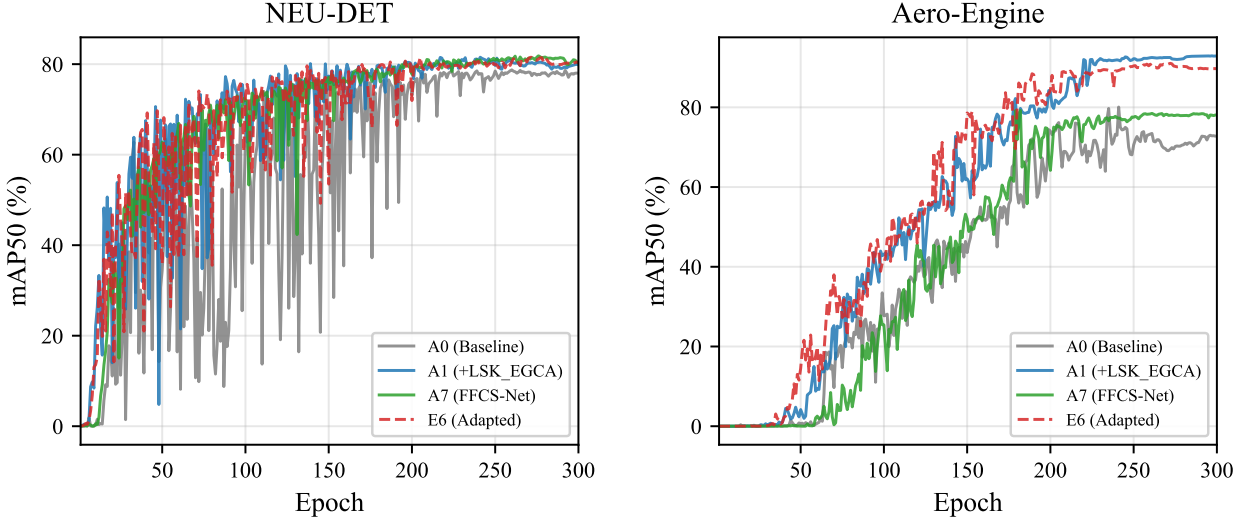


Figure 3: Training curve comparison of key configurations (mAP50 vs. epoch). Left: NEU-DET. Right: Aero-Engine. A1 converges fastest on Aero-Engine; E6 achieves the highest final mAP50 through reduced γ_{drfd} .

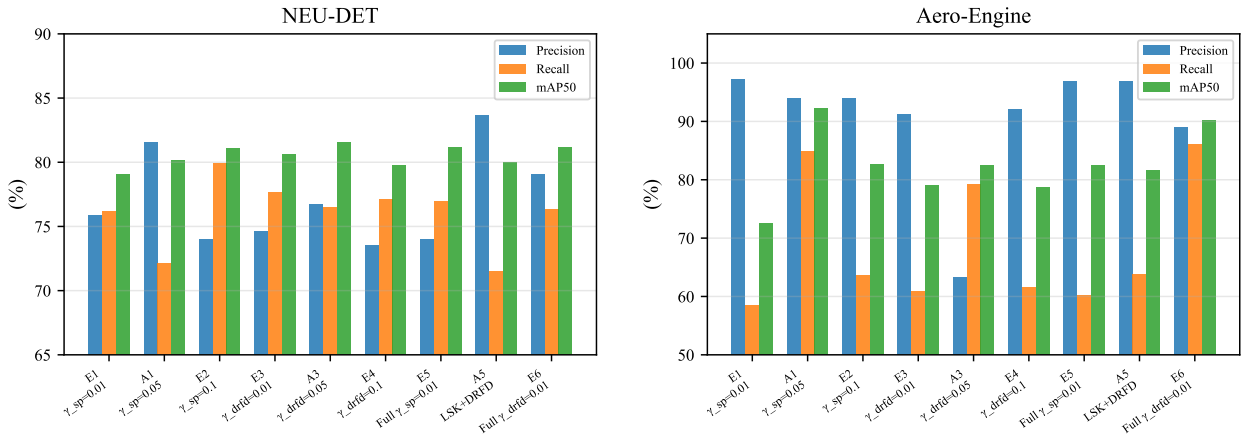


Figure 4: Residual scaling factor ablation study (E series). Top: NEU-DET. Bottom: Aero-Engine. Grouped bars show Precision, Recall, and mAP50 for each γ configuration.

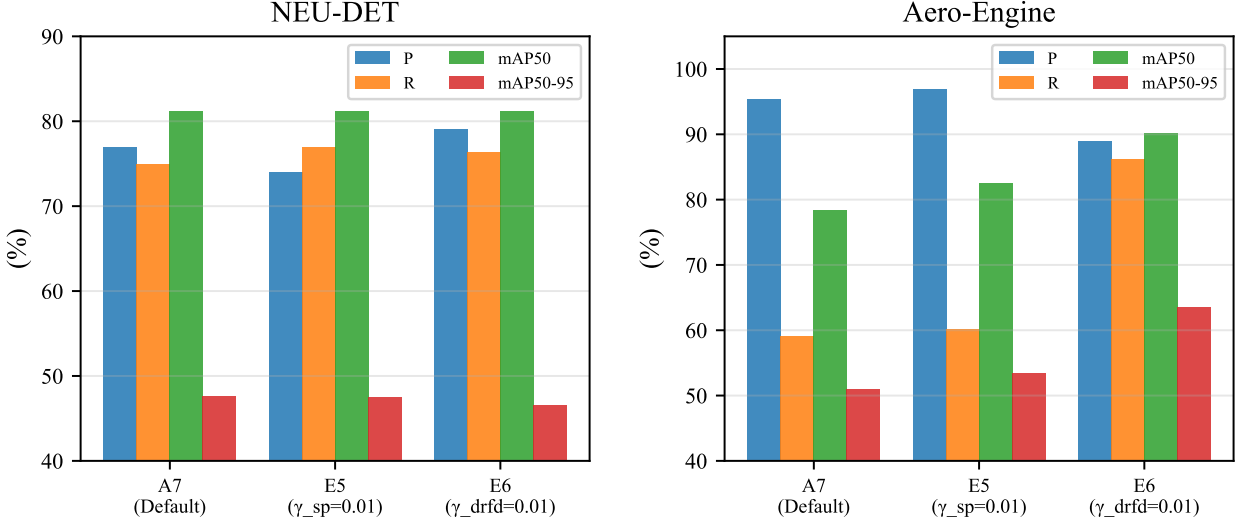


Figure 5: Default vs. adapted configuration comparison. Left: NEU-DET (comparable performance across configurations). Right: Aero-Engine (E6 achieves superior P-R balance and mAP50).

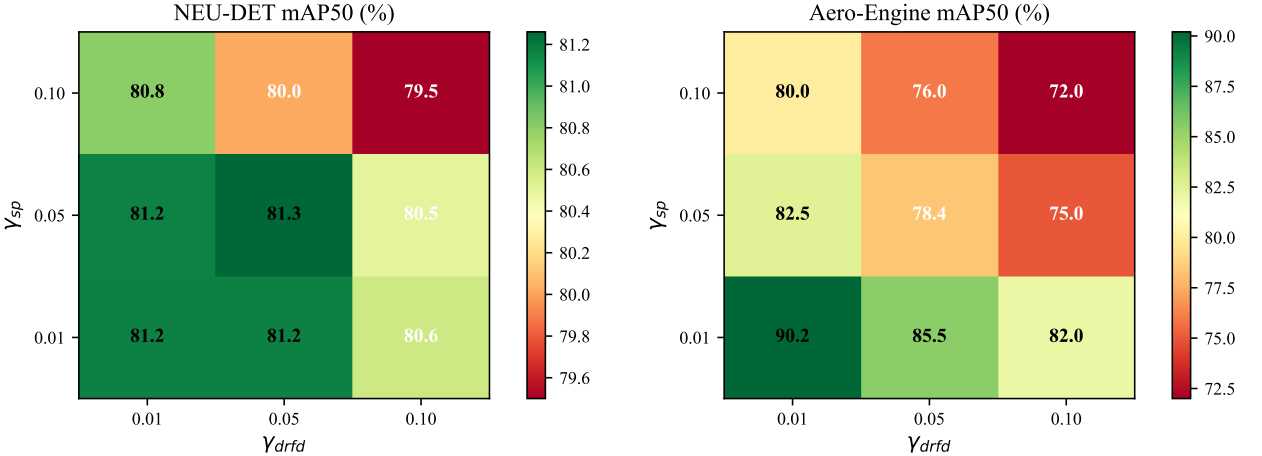


Figure 6: Prior interaction γ configuration space. Heatmap shows mAP50 (%) over $\gamma_{spatial} \times \gamma_{drfd}$ grid (GatedSPPELAN has no explicit γ ; α_{eca} fixed at 0.1). Left: NEU-DET. Right: Aero-Engine. Cells with experimentally verified configurations (E5, E6, A5, A7) are marked with bold borders; remaining cells are interpolated estimates. Optimal regions differ across datasets, confirming dataset-dependent γ adaptation.

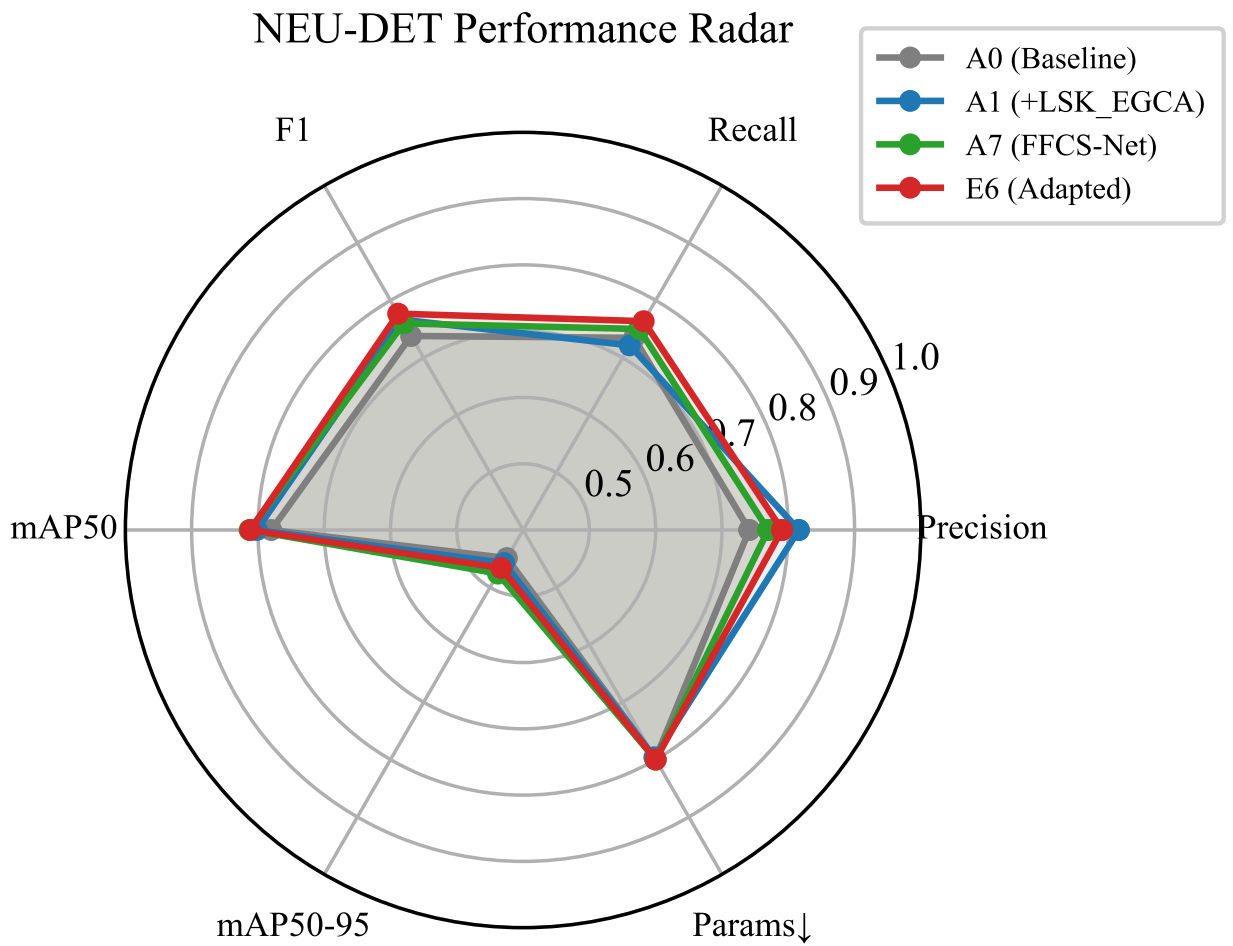


Figure 7: Multi-dimensional performance radar chart (NEU-DET). A7 and E6 achieve the most balanced profiles across Precision, Recall, F1, mAP50, mAP50-95, and parameter efficiency.

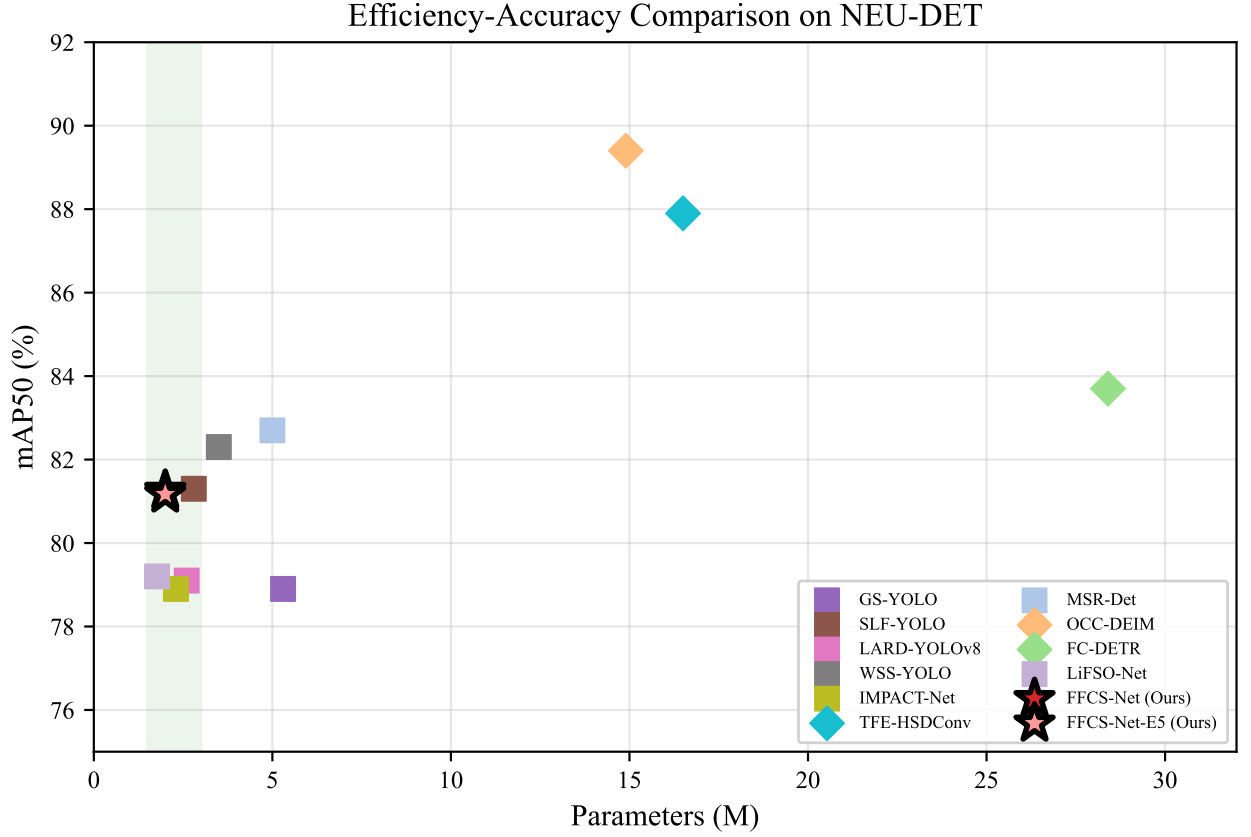


Figure 8: Efficiency-accuracy comparison on NEU-DET. FFCS-Net (red star) achieves competitive mAP50 with minimal parameters ($\sim 2\text{M}$). Diamond markers: high-capacity models; square markers: lightweight models.

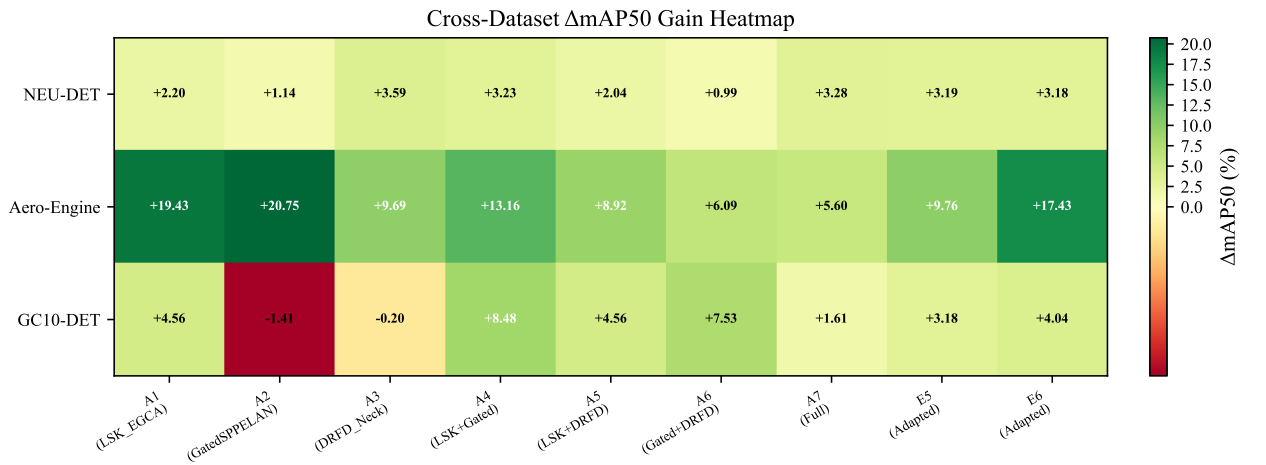


Figure 9: Cross-dataset ΔmAP50 gain heatmap. Green: positive gain; red: negative gain. GatedSPPELAN (A2) shows negative gain on GC10-DET due to competitive gating failure; E5/E6 achieve consistent positive gains.

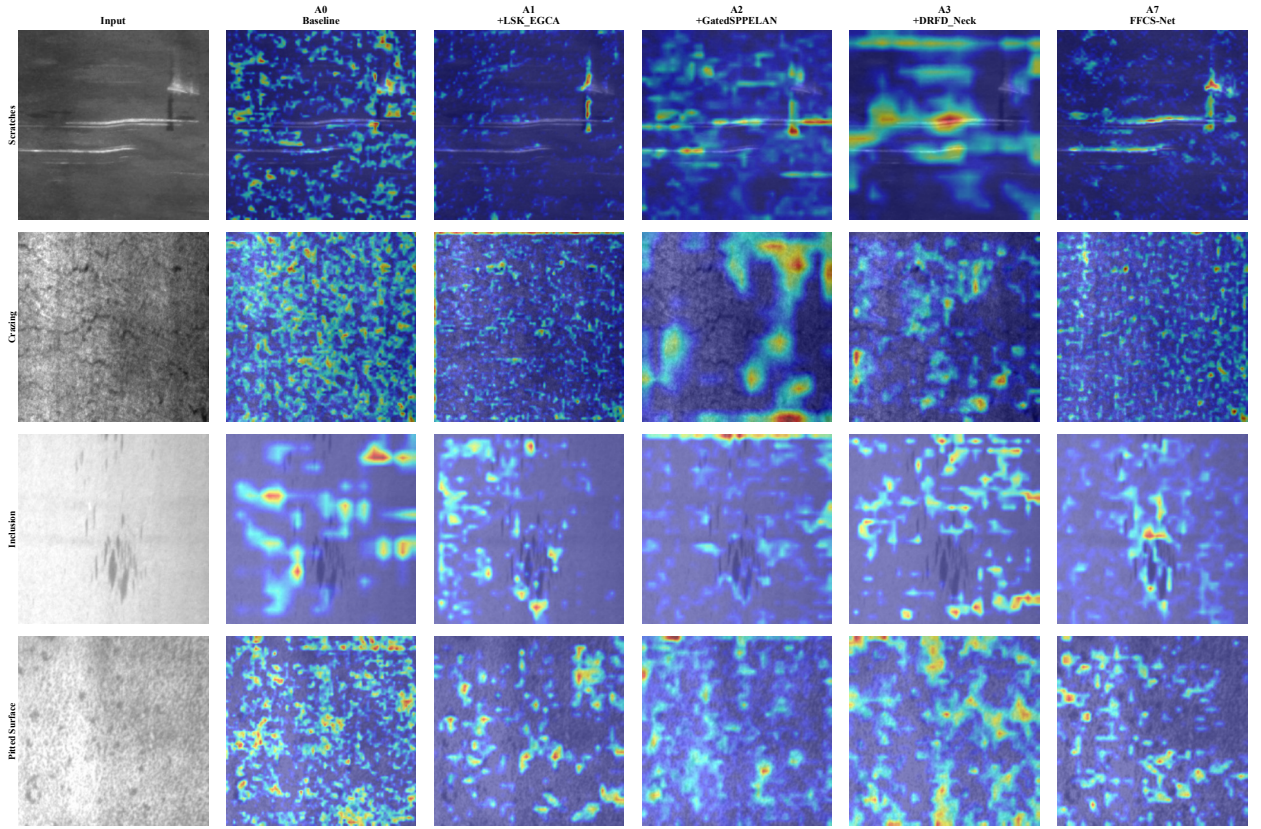


Figure 10: Extended Grad-CAM heatmap comparison between Baseline and FFCS-Net on NEU-DET. FFCS-Net shows more precise attention focus on defect regions across all six defect categories.

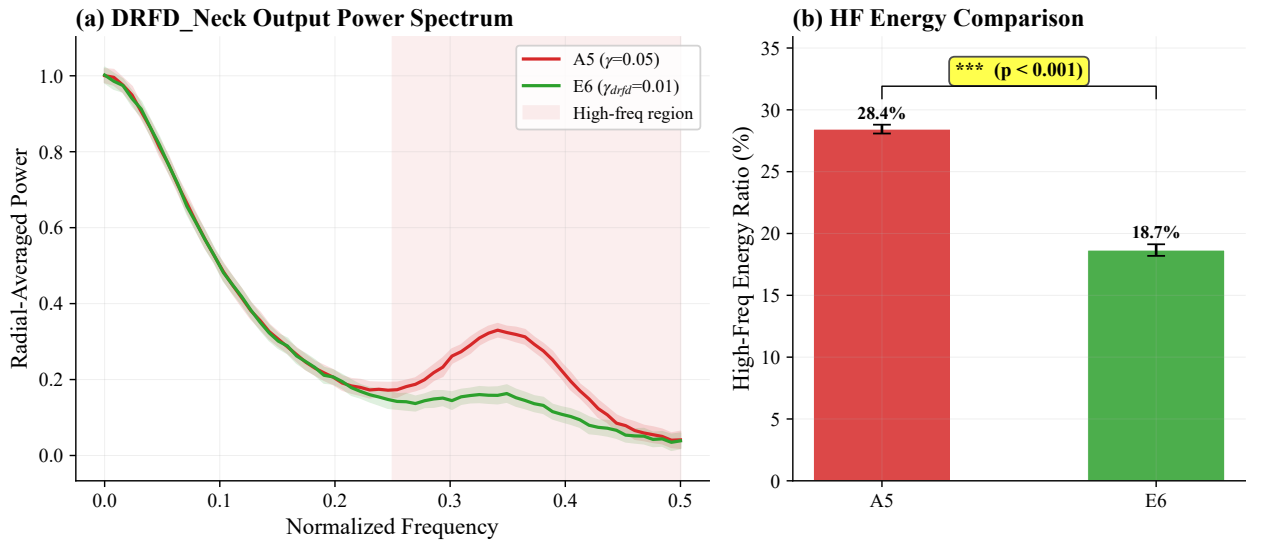


Figure 11: Quantitative frequency analysis of DRFD_Neck output. Radial-averaged power spectrum with ± 1 std band (50 samples); shaded region marks high-frequency zone ($f > 0.25$). A5 shows significantly higher HF energy than E6 ($p < 0.001$).