Table S1 Ablation on DCT block size k in DHF-FPN (PVELAD EL2021).

k	mAP@50	mAP@50–95	Precision	Recall
4	0.642	0.427	0.699	0.644
8	0.735	0.443	0.792	0.762
16	**0.757**	**0.482**	**0.890**	0.637
32	0.668	0.456	0.807	0.566

Table S2 Stability of the fixed binary high-pass mask (3 seeds: 42, 123, 456).

Seed	Precision	Recall	mAP@50	mAP@50–95
42	0.784	0.544	0.601	0.398
123	0.801	0.580	0.630	0.418
456	0.851	0.727	0.763	0.489
Mean±Std	0.812±0.035	0.617±0.096	0.665±0.084	0.435±0.047

Table S3 Sensitivity analysis of GA-Head hidden channels (*hidc*) on PVELAD EL2021.

hidc	mAP@50	mAP@50–95	Precision	Recall
256	0.634	0.413	0.821	0.481
384	0.604	0.404	0.767	0.551
512 (default)	**0.725**	**0.473**	**0.870**	**0.553**

Supplementary Material

R1-2. DCT Mask and Frequency Split Study (DHF-FPN)

We observe that $k = 16$ yields the best mAP@50–95 (0.482 vs. 0.443 at $k = 8$, +0.039 / +8.8%), while $k = 4$ degrades performance, consistent with stronger boundary effects and over-partitioning. Accordingly, we adopt $k = 16$ as the default in Sec. 3.2.1.

The fixed mask shows stable performance across seeds (mAP@50–95 = 0.435±0.047) without introducing extra learnable frequency parameters.

R1-3. GA-Head Hyperparameters and Architecture Details

Statistical significance.

we conduct a two-sided paired t-test on mAP@50–95 across the three seeds. The gains of our GA-Head over the standard head and the w/o-DCN variant are statistically significant ($p < 0.001$).

Table S4 Replacing GA-Head with alternative detection heads (mean±std over 3 seeds).

Head	mAP@50	mAP@50–95	Δ vs. GA-Head
GA-Head (Ours)	**0.708±0.032**	**0.441±0.010**	–
Standard Detect	0.568±0.021	0.376±0.016	-14.7%
Decoupled (w/o DCN)	0.535±0.018	0.355±0.003	-19.5%
DyHead	0.567±0.004	0.383±0.004	-13.2%

Table S5 Ablation on insertion positions of DCA and DHF-FPN on PVELAD (100 epochs, seed=42).

Setting	DCA pos.	DHF-FPN pos.	Params(M)	mAP@50	mAP@50–95	Precision	Recall
P3-only	P3	P3	2.33	0.722	0.446	0.903	0.619
P5-only	P5	–	2.55	0.682	0.434	0.894	0.598
All-level (default)	P3–P5	P3–P4	3.01	0.667	0.427	0.842	0.660

Table S6 Comparison of high/low-frequency fusion strategies in DHF-FPN (mean±std over 3 seeds).

Fusion strategy	Precision	Recall	mAP@50	mAP@50–95
multiply_add (Ours)	0.827±0.095	0.672±0.070	0.697±0.023	**0.444±0.011**
add_only	0.839±0.039	0.598±0.054	0.650±0.049	0.423±0.024
concat	0.780±0.016	0.567±0.007	0.615±0.050	0.385±0.010

Insertion position ablation of DCA/DHF-FPN

Enhancing high-resolution features (P3-only) yields the best mAP, consistent with many PV defects being small and high-frequency dominated. The default multi-level setting provides higher recall, which is preferable when missing defects is costly.

R1-4. Fusion Strategy Ablation and γ Visualization

Eq. (4) in the main paper is a lightweight gated residual fusion, where γ controls the injection strength of the high-frequency branch. A channel-attention variant can be obtained by extending γ from a scalar to a channel-wise vector, which increases flexibility but also introduces extra parameters and potential overfitting risk. We leave a thorough evaluation of channel-wise gating to future work.

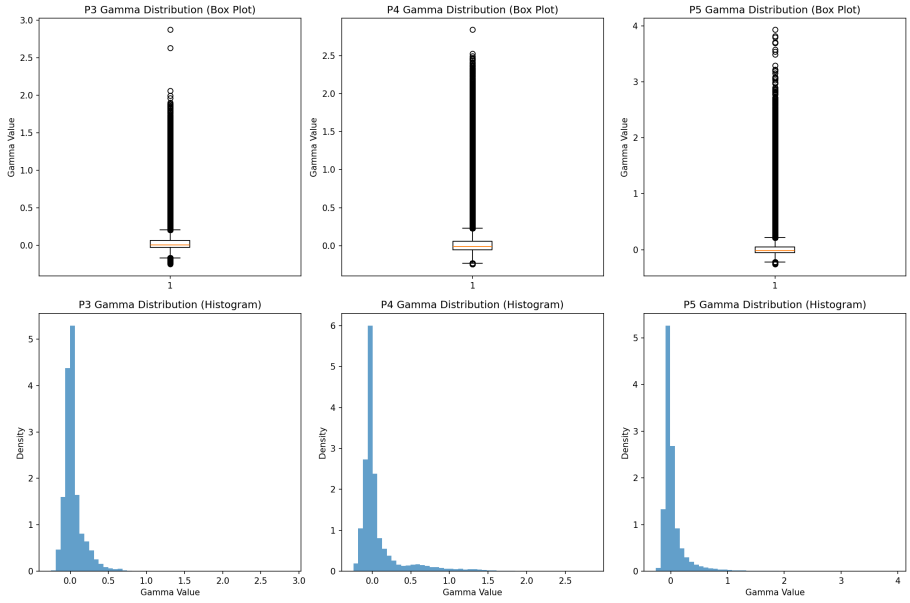

Fig. S1 Distribution of learned γ across FPN levels (P3/P4/P5) over three seeds. Shallow levels tend to prefer larger γ (stronger high-frequency injection), while deeper levels prefer smaller γ (more semantic-preserving), indicating hierarchical frequency preference.

Table S7 Per-seed results and differences (YOLO-PVELAD minus YOLO11n). Seeds are 42/123/456. Metrics are reported on PVELAD EL2021.

Seed	YOLO11n				YOLO-PVELAD				Diff. (Ours - YOLO11n)			
	mAP@50	mAP@50–95	P	R	mAP@50	mAP@50–95	P	R	ΔmAP@50	ΔmAP@50–95	ΔP	ΔR
42	59.0	36.4	80.2	47.9	58.7	38.6	86.9	55.1	-0.3	+2.2	+6.7	+7.2
123	65.7	38.1	72.1	54.5	75.5	46.4	89.7	73.0	+9.8	+8.3	+17.6	+18.5
456	57.3	36.2	89.8	54.3	75.3	43.6	88.4	78.5	+18.0	+7.4	-1.4	+24.2
Mean±Std	60.7±3.6	36.9±0.8	80.7±7.3	52.2±3.0	69.8±7.9	42.9±3.3	88.3±1.2	68.9±10.0	+9.2	+6.0	+7.6	+16.6

R1-5. Per-seed results and significance (YOLO-PVELAD vs. YOLO11n)

Significance testing (paired).

We conduct a paired t-test on the per-seed mAP@50–95 differences between YOLO-PVELAD and YOLO11n ($n = 3$). We therefore emphasize transparent mean±std reporting and per-seed results, and will increase the number of runs when stricter statistical validation is required.

R1-7. Standard Alternative Controls for Each Module

Deformable sampling brings a clear improvement over standard convolutional heads, validating that geometry-adaptive receptive fields are crucial for elongated and irregular PV defects.

Table S8 Effect of deformable sampling in GA-Head (mean±std over 3 seeds). All variants keep the same backbone/neck and only change the detection head.

Head variant	Params(M)	mAP_{50}	$mAP_{50:95}$	Recall
GA-Head (with DyDCNv2)	∼3.0	0.687±0.018	0.437±0.009	0.640±0.028
Decoupled head (no DCN)	∼2.6	0.519±0.008	0.354±0.003	0.484±0.014
Standard YOLO head	∼2.5	0.533±0.022	0.368±0.008	0.524±0.074

Table S9 Ablation on shared vs. separate feature towers in task decomposition (100 epochs, seed=42).

Setting	Tower	Params(M)	mAP_{50}	$mAP_{50:95}$	Precision	Recall
Shared-tower (default)	shared	3.01	0.667	0.427	0.842	0.660
Separate-tower	separate	3.13	0.601	0.410	0.829	0.576

The shared-tower design achieves better accuracy with fewer parameters and higher recall, suggesting that moderate feature sharing can stabilize optimization and promote beneficial interaction between semantic classification and geometric regression.

R1-8. Data-driven Challenging Subset Definition and Visualization

We compute subset thresholds on the training split: $a_{20} = 0.0019$ for relative area and $r_{80} = 8.41$ for aspect ratio. Fig. S2 shows the distributions, and Fig. S3 shows the area–aspect-ratio scatter plot.

R1-9. Inference speed testing protocol.

We report inference latency/FPS and peak memory under a unified setting: batch size = 1, FP32 precision, 10 warmup runs and 100 timed runs. To ensure accurate GPU timing and avoid asynchronous bias, latency is measured using CUDA events (or equivalently enforcing `torch.cuda.synchronize()` before and after timing). Peak memory is recorded as the maximum allocated GPU memory during inference.

We agree that edge-device benchmarks (e.g., Jetson with TensorRT FP16/INT8) are important. Due to limited access to such hardware during revision, we do not report measured edge-device benchmarks. We will add edge-device measurements in future work when the hardware becomes available.

R1-10. Multi-resolution complexity analysis.

To characterize how computation and memory scale with input resolution, we report Params and FLOPs at 416/640/1280. For reference, MACs are computed as MACs ≈ FLOPs/2 under the standard 2-FLOPs-per-MAC convention. We also report peak

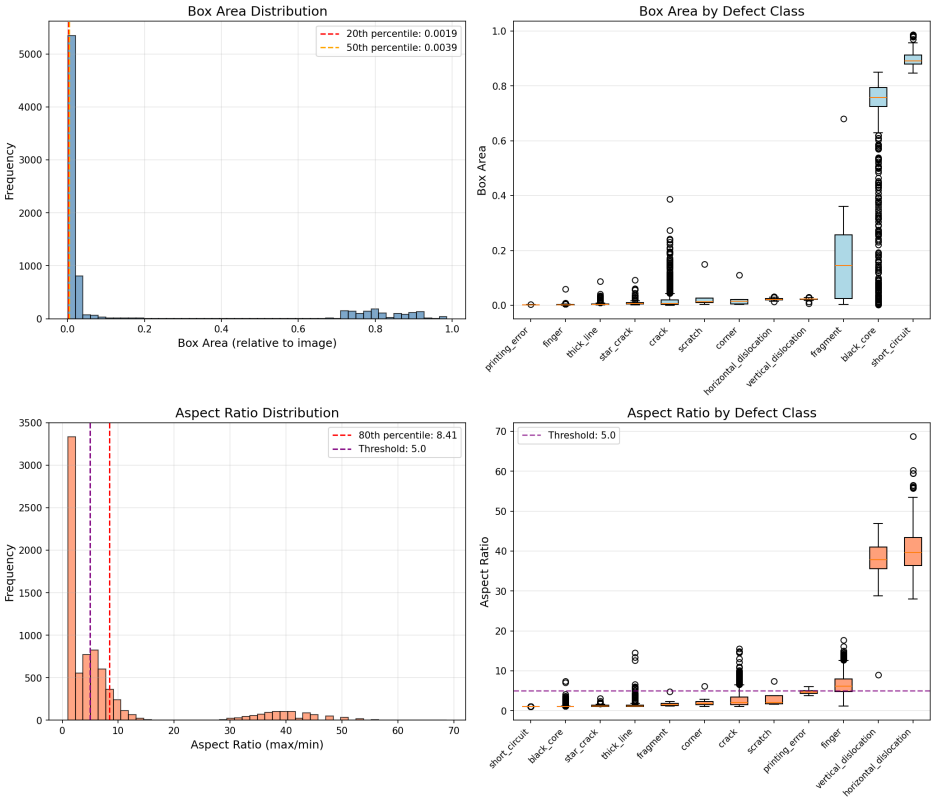

Fig. S2 Distributions for challenging subset thresholds. Relative box area and aspect ratio distributions. Thresholds are computed on the training split ($a_{20} = 0.0019$, $r_{80} = 8.41$).

Table S10 Inference speed and memory under a unified protocol (FP32, batch=1). All results are measured at input resolution 640×640 using the protocol described above.

Model	Params(M)	FLOPs(G)	Latency(ms)	FPS	Memory(MB)
YOLO11n (baseline)	2.62	3.31	4.96	201.7	72.1
YOLO11n + DCA (ELA)	1.86	2.92	4.78	209.1	79.6
YOLO11n + DyHead	3.14	3.88	8.31	120.4	111.4
YOLO-PVELAD (ours)	**3.01**	**4.90**	**7.43**	**134.7**	**109.1**

memory (maximum allocated GPU memory during inference) as a practical proxy for activation/memory cost (at 640×640 under the same protocol as R1-9).

R1-11. Preprocessing pipeline.

To ensure reproducibility for low-contrast EL imagery, we explicitly report the input bit depth and preprocessing used in all experiments. The EL images are stored as 8-bit

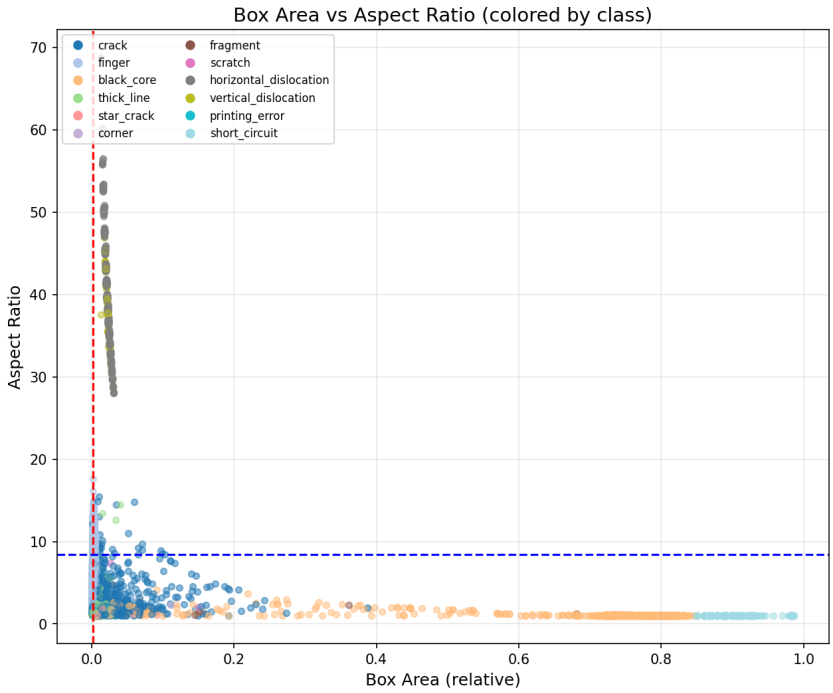

Fig. S3 Area vs. aspect ratio scatter plot. Each point corresponds to a ground-truth box. The quantile thresholds ($a_{20} = 0.0019$, $r_{80} = 8.41$) define the small and elongated subsets.

Table S11 Multi-resolution complexity profile (batch=1). FLOPs are reported under a consistent counting rule; MACs are shown for reference (MACs≈FLOPs/2).

Model	Params(M)	FLOPs (G)			MACs (G)			Peak Mem (MB)
		416	640	1280	416	640	1280	640
YOLO11n (baseline)	2.62	1.40	3.31	13.23	0.70	1.66	6.62	72.1
YOLO11-HSFPN	1.86	1.23	2.92	11.67	0.62	1.46	5.84	79.6
YOLO11-DyHead	3.14	1.64	3.88	15.51	0.82	1.94	7.76	111.4
Ours (ELA-HSFPN-TADDH)	**3.01**	**2.09**	**4.90**	**19.45**	**1.05**	**2.45**	**9.73**	**109.1**

after conversion from the original 16-bit data. We apply linear normalization by dividing pixel values by 255.0 and do not perform mean subtraction. We intentionally do not use histogram equalization (e.g., CLAHE) or denoising, so that the benefit comes from the proposed DHF-FPN frequency enhancement rather than external contrast enhancement.

R1-11. Augmentation settings.

We enable Mosaic augmentation during training with probability $p = 1.0$ and adopt HSV augmentation with $(h, s, v) = (0.015, 0.7, 0.4)$.

Table S12 Data preprocessing and augmentation settings used in all experiments.

Setting	Value
Bit depth	8-bit (converted from original 16-bit)
Normalization	/255.0 (no mean subtraction)
Histogram equalization (CLAHE)	Not used
Denoising	Not used
Mosaic augmentation	Enabled, $p = 1.0$
HSV augmentation	$h = 0.015,\ s = 0.7,\ v = 0.4$

Table S13 Robustness under common image degradations (mAP@50–95). Results are evaluated on the PVELAD validation split.

Degradation	Setting	YOLO11n	Ours	Relative gain
Gaussian noise	σ=10	0.343	**0.418**	+21.9%
Gaussian noise	σ=15	0.319	**0.383**	+20.1%
Gaussian blur	k=7	0.374	**0.414**	+10.7%
Gaussian blur	k=11	0.357	**0.381**	+6.7%
Contrast	f=0.5	0.256	**0.292**	+14.1%
Contrast	f=1.5	0.286	**0.384**	+34.3%

Fig. S4 Robustness curves under noise/blur/contrast. We plot mAP@50–95 versus degradation strength for YOLO11n and our method.

R2-2. Domain-specific Methods and Comparability Notes

Why we summarize instead of directly mixing numbers.

Recent PV/industrial defect works may differ from our setting in task definition (classification vs. detection), imaging modality (RGB/thermal/EL), dataset availability, and evaluation metrics. To avoid misleading cross-paper comparisons, we provide the following summary table and keep the main-paper Table 2 strictly reproducible on the same dataset and protocol.

R2-5. Robustness to Image Degradation and Failure Case Analysis

Robustness protocol.

We evaluate robustness on the PVELAD validation split (900 images) using the best checkpoint trained with seed 42. We apply three degradations: Gaussian noise ($\sigma \in \{5, 10, 15, 20, 25\}$), Gaussian blur (kernel $\in \{3, 5, 7, 9, 11\}$), and contrast scaling ($f \in \{0.5, 0.75, 1.0, 1.25, 1.5\}$).

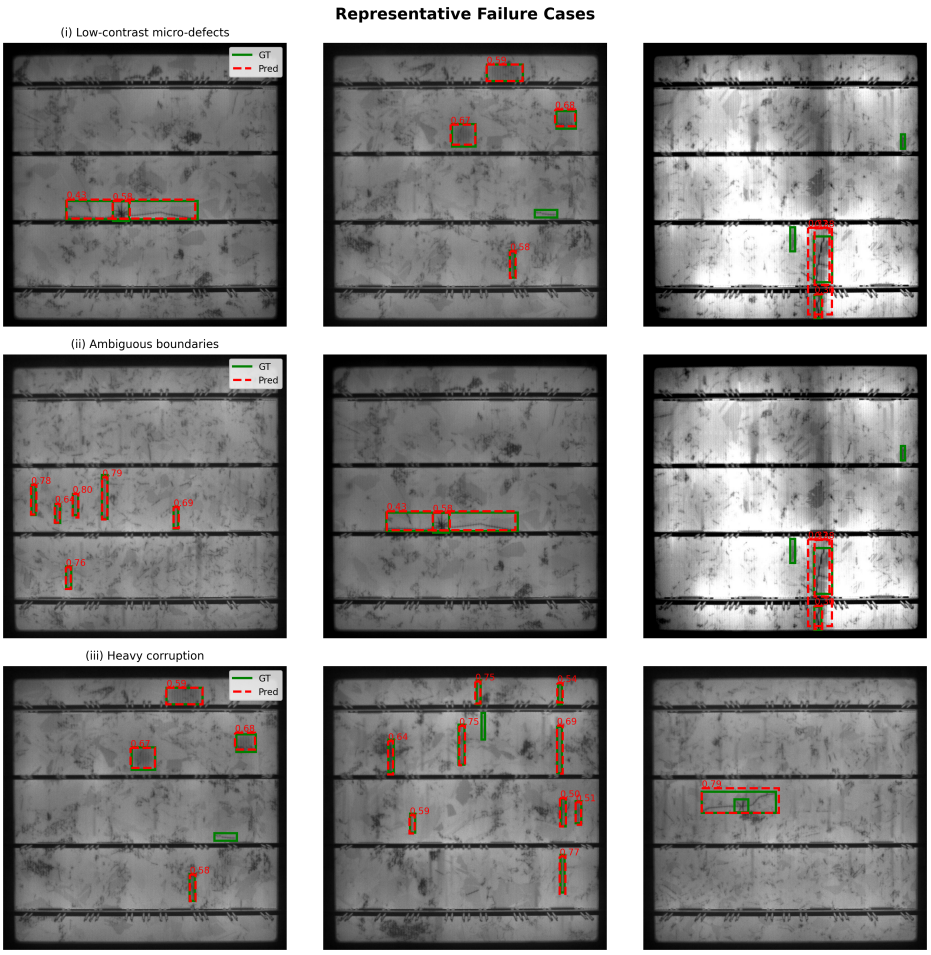

Fig. S5 Representative failure cases and failure-mode taxonomy. We show samples where YOLO11n and our method fail, together with brief failure-mode annotations.

Failure cases and limitations.

To provide a balanced assessment, we include representative failure cases and summarize typical failure modes: (i) extremely low-contrast micro-defects near the noise floor, (ii) ambiguous boundaries where multiple defect types overlap, (iii) heavy corruption (e.g., very strong noise) where all methods degrade. See Fig. S5 for qualitative examples.