Supplementary Material for TPW YOLO

TPW-YOLO for microaneurysm detection with tile-phase consistency and workload-aware ranking

Reference numbers correspond to the main manuscript. Cohorts, seeds, matching rules, and operating thresholds follow the main methods.

Supplementary Methods

Annotation correspondence and paired tile views are shown in Figure S1. The qualitative comparison and candidate selection before and after ranking are shown in Figures S2 and S3, respectively.

Comparator implementation and training exposure

FPN–FSAF retained green-channel enhancement, gamma 0.8, a 5 × 5 median filter, pretrained ResNet34 with CBAM, a P2–P7 pyramid, online level selection, and an RoIAlign classifier [9, 24]. It used native-resolution 200 × 200 patches with 112-pixel overlap. The local adaptation used standard ResNet34 widths, required inside-box points with P2 fallback for tiny targets, and used softplus regression distances with generalized IoU loss. Unified used its MA candidate-generation and pretrained VGG16 RoI stream, with a 700 × 700 working image and four overlapping 500 × 500 patches [20].

SSPNET used pretrained ResNet50, attention-based scale selection, Trident RPN, and the released two-output foreground/background RoI classifier [8]. It sampled one 640 × 512 patch per image per epoch and used full-image inference; random sampling replaced the unavailable ICNegSampler. PMNPDR combined matched-filter candidates with a 32 × 32 patch CNN [19]. This partial adaptation approximated two-means thresholding and omitted curvelet enhancement and mutual-information/differential-evolution optimization.

Comparisons used a shared partition and nominal optimization schedule while retaining each pipeline's sampling procedure. The epoch count therefore did not equalize sample exposure or training cost: FPN–FSAF used 100,589 patches per epoch, whereas TPCL used 11,650 pair-sampling slots. PMNPDR, SSPNET, and Unified recorded 8,523, 119, and 478 EMA updates, respectively. TPW-YOLO additionally used two OOF detector fits for IRBR training. Reported comparisons concern these local implementations; published scores from other partitions were not combined with them.

The four paper-derived pipelines used batch size 4, nominal batch size 64 with gradient accumulation, gradient clipping at 10, and the common optimizer and checkpoint settings in the main manuscript. Unavailable components were adapted as stated above; these experiments evaluate local implementations rather than published results on different partitions.

Boundary and phase diagnostics

Boundary recall was evaluated at the pooled 4-FPI threshold. For each lesion, the largest center-to-border distance among its covering tiles was divided by 320 and assigned to [0, 0.25), [0.25, 0.5), [0.5, 0.75), or [0.75, 1]. The first bin identifies lesions close to a boundary even in their most central covering tile. Separately evaluated phase grids supplied both-phase detection coverage and, for lesions detected in both grids at the candidate floor, half the absolute score difference, half the absolute within-image rank-percentile difference, and Euclidean center displacement. These conditional diagnostics characterize changes between the two grid placements.

Candidate coverage was measured at detector confidence floor 10^{-4} using the same 12-pixel matching rule (Table S8). Annotation-based reconstruction of epoch 29 quantified lesion pairs meeting detection-retention and inset-visibility criteria before model-dependent assignment (Table S9). The n-scale seed-0 A4 run additionally recorded geometric pairs, positive assignments in both views, and batches without assigned

pairs during active-consistency epochs 16–30 (Table S12). Boundary contrasts used paired image resampling for the same 161 nearest-boundary lesions.

Image partition and visit grouping

ADCIS identifies each e-ophtha source folder as a patient visit (https://www.adcis.net/en/third-party/e-ophtha/). In the evaluated image-level partition, 21 source folders contain both training and test images (36 training and 21 test images). Twenty-three source folders span the two OOF folds. Image-level OOF fitting therefore does not exclude other images from the same visit, and evaluation estimates performance under this image-level partition.

Supplementary Tables

Table S1 Individual Sn@4 FPI and 95% image-bootstrap intervals by scale and seed.

Scale and configuration	Seed 0	Seed 1	Seed 2
s A3 Random shift	41.94 [36.86, 46.05]	-	-
s A4 Full TPCL	42.77 [38.16, 46.58]	-	-
s A6 TPW-YOLO	43.39 [38.47, 47.41]	-	-
s A7 Reduced features	42.70 [37.78, 46.72]	-	-
n A3 Random shift	41.45 [36.49, 45.87]	37.09 [33.09, 40.68]	40.62 [36.20, 44.53]
n A4 Full TPCL	43.94 [38.63, 48.49]	43.94 [39.27, 48.12]	41.38 [36.61, 45.47]
n A6 TPW-YOLO	45.19 [39.68, 50.21]	45.05 [40.70, 48.96]	43.18 [38.15, 47.42]
n A7 Reduced features	42.77 [37.29, 47.32]	44.22 [39.54, 48.30]	42.15 [37.27, 46.18]
n TPCL without gate	41.59 [36.44, 46.14]	42.91 [38.21, 47.28]	40.83 [35.99, 44.90]
n Score consistency only	41.04 [35.86, 45.47]	36.26 [31.59, 40.38]	41.38 [36.43, 45.59]
n Location consistency only	44.01 [38.63, 48.74]	43.94 [39.30, 48.12]	42.28 [37.57, 46.50]

Intervals use 2,000 image resamples at each fitted model's fixed full-sample operating threshold. A dash denotes an unperformed s-scale seed. Results from different scales are reported separately.

Table S2 Source-specific Sn@4 and intervals for each scale and seed.

Scale configuration source	Seed 0	Seed 1	Seed 2
s A3 Random shift idrid	32.26 [27.66, 36.06]	-	-
s A3 Random shift roc	29.07 [16.16, 58.83]	-	-
s A3 Random shift eophtha	59.49 [51.49, 66.18]	-	-
s A4 Full TPCL idrid	32.90 [27.75, 36.97]	-	-
s A4 Full TPCL roc	27.91 [16.46, 56.42]	-	-
s A4 Full TPCL eophtha	56.93 [49.11, 63.71]	-	-
s A6 TPW-YOLO idrid	33.64 [28.81, 37.54]	-	-
s A6 TPW-YOLO roc	23.26 [13.33, 46.56]	-	-
s A6 TPW-YOLO eophtha	55.47 [46.99, 62.92]	-	-

Scale configuration source	Seed 0	Seed 1	Seed 2
s A7 Reduced features idrid	34.01 [29.42, 37.69]	-	-
s A7 Reduced features roc	19.77 [10.00, 44.68]	-	-
s A7 Reduced features eophtha	53.28 [45.00, 60.24]	-	-
n A3 Random shift idrid	35.30 [30.68, 39.64]	12.26 [9.83, 14.70]	35.39 [30.53, 39.56]
n A3 Random shift roc	24.42 [14.55, 46.43]	32.56 [22.06, 57.58]	18.60 [11.21, 34.15]
n A3 Random shift eophtha	55.84 [47.62, 63.83]	55.47 [47.09, 63.09]	55.11 [46.29, 63.38]
n A4 Full TPCL idrid	35.58 [30.95, 39.73]	36.59 [32.06, 40.81]	34.38 [29.61, 38.37]
n A4 Full TPCL roc	23.26 [13.45, 51.07]	27.91 [16.98, 50.00]	22.09 [13.21, 43.59]
n A4 Full TPCL eophtha	57.66 [48.90, 66.14]	56.57 [48.60, 64.49]	55.11 [46.19, 63.28]
n A6 TPW-YOLO idrid	38.71 [33.54, 43.55]	36.77 [32.91, 40.74]	35.21 [30.37, 39.17]
n A6 TPW-YOLO roc	23.26 [14.28, 48.96]	24.42 [14.61, 41.94]	19.77 [13.04, 36.37]
n A6 TPW-YOLO eophtha	50.73 [40.87, 60.77]	48.54 [39.15, 56.92]	53.28 [44.83, 61.61]
n A7 Reduced features idrid	35.39 [30.40, 39.52]	35.67 [31.60, 39.65]	34.19 [29.62, 38.24]
n A7 Reduced features roc	19.77 [11.56, 40.00]	18.60 [8.42, 44.44]	19.77 [12.12, 40.74]
n A7 Reduced features eophtha	53.65 [45.95, 60.69]	50.36 [42.65, 57.89]	56.57 [47.69, 65.09]
n TPCL without gate idrid	36.68 [32.00, 41.05]	35.58 [30.94, 39.77]	33.09 [29.53, 36.38]
n TPCL without gate roc	25.58 [13.39, 53.85]	26.74 [16.95, 48.28]	20.93 [12.09, 41.03]
n TPCL without gate eophtha	52.19 [43.52, 61.09]	56.93 [49.36, 64.19]	53.65 [45.56, 61.02]
n Score consistency only idrid	35.21 [30.18, 39.79]	6.36 [4.76, 7.78]	33.92 [29.59, 38.07]
n Score consistency only roc	23.26 [12.19, 48.40]	27.91 [16.82, 51.72]	18.60 [11.11, 37.94]
n Score consistency only eophtha	55.47 [47.02, 63.67]	54.74 [46.32, 62.83]	53.28 [43.99, 61.95]
n Location consistency only idrid	35.12 [30.51, 39.34]	36.87 [32.27, 41.22]	34.01 [29.91, 37.55]
n Location consistency only roc	25.58 [15.33, 51.86]	30.23 [18.46, 55.18]	20.93 [12.33, 42.60]
n Location consistency only eophtha	56.20 [48.06, 64.32]	56.20 [48.16, 63.51]	55.47 [45.49, 64.96]

Intervals use source-specific full-sample 4-FPI thresholds. ROC includes 10 test images and 86 lesions.

Table S3 Archived computational measurements for individual runs.

Scale / seed	Configuration	Mean ms/image	P90 ms/image	GPU MiB
s/0	A3 Random shift	1381.7	3793.2	137.1
s/0	A4 Full TPCL	1361.6	3307.0	137.1
s/0	A6 TPW-YOLO	1362.1	3307.7	137.1
s/0	A7 Reduced features	1361.9	3307.5	137.1
n/0	A3 Random shift	2141.8	5500.1	173.6
n/0	A4 Full TPCL	2275.7	6148.1	173.8
n/0	A6 TPW-YOLO	2276.1	6148.6	173.8
n/0	A7 Reduced features	2276.1	6148.8	173.8
n/0	TPCL without gate	2069.4	5186.1	174.0
n/0	Score consistency only	2354.1	6068.7	173.6
n/0	Location consistency only	2346.9	6356.8	173.6
n/1	A3 Random shift	2201.2	5880.8	173.5
n/1	A4 Full TPCL	2357.6	6326.1	173.6
n/1	A6 TPW-YOLO	2357.9	6326.6	173.6
n/1	A7 Reduced features	2357.9	6326.6	173.6
n/1	TPCL without gate	2574.4	7117.2	173.6
n/1	Score consistency only	2420.6	6785.1	173.6
n/1	Location consistency only	4514.7	12354.2	173.6
n/2	A3 Random shift	4249.2	12707.8	173.6
n/2	A4 Full TPCL	2144.8	6603.3	173.6
n/2	A6 TPW-YOLO	2145.1	6603.8	173.6
n/2	A7 Reduced features	2145.1	6603.8	173.6
n/2	TPCL without gate	2239.0	6953.6	173.6
n/2	Score consistency only	2215.3	6684.2	173.6
n/2	Location consistency only	2011.4	6226.5	173.5

Latency is per original image, excluding file reading. A6/A7 reuse A4 extraction times and add CPU scoring time. Records were acquired across sessions. MiB denotes 2^{20} bytes.

Table S4 Boundary and conditional phase diagnostics for individual runs.

Scale seed configuration	Bin 1 (%)	Bin 2 (%)	Bin 3 (%)	Bin 4 (%)	Both phases (%)	Score fluctuation
s/0 A3	44.72	42.08	41.85	39.23	85.26	0.0558
s/0 A4	44.72	42.74	43.06	40.33	84.64	0.0575
s/0 A6	48.45	43.40	43.26	39.23	84.64	0.1001
s/0 A7	46.58	42.74	42.86	38.67	84.64	0.0691
n/0 A3	47.20	41.25	39.84	41.44	87.06	0.0596
n/0 A4	50.31	43.89	41.45	45.30	87.89	0.0602
n/0 A6	49.69	42.08	45.27	51.38	87.89	0.1000
n/0 A7	40.37	40.59	42.05	54.14	87.89	0.0448
n/0 no_gate	40.99	43.07	41.25	38.12	87.61	0.0587

Scale seed configuration	Bin 1 (%)	Bin 2 (%)	Bin 3 (%)	Bin 4 (%)	Both phases (%)	Score fluctuation
n/0 score_only	48.45	40.76	39.64	39.23	88.24	0.0586
n/0 loc_only	50.93	44.22	41.45	44.20	87.75	0.0620
n/1 A3	34.16	36.30	37.63	40.88	88.10	0.0541
n/1 A4	44.10	43.40	44.67	43.65	88.44	0.0528
n/1 A6	47.20	44.88	44.06	46.41	88.44	0.1086
n/1 A7	45.34	44.55	44.47	41.44	88.44	0.0580
n/1 no_gate	40.99	42.57	43.46	44.20	88.10	0.0559
n/1 score_only	36.02	36.14	37.22	34.25	88.37	0.0534
n/1 loc_only	43.48	43.73	44.67	43.09	88.24	0.0565
n/2 A3	45.96	41.09	39.84	36.46	87.34	0.0604
n/2 A4	47.83	41.58	40.85	36.46	87.68	0.0607
n/2 A6	49.69	41.91	42.86	42.54	87.68	0.1005
n/2 A7	45.96	41.09	42.25	41.99	87.68	0.1092
n/2 no_gate	47.83	42.08	39.24	34.81	88.10	0.0562
n/2 score_only	47.20	41.91	40.64	36.46	88.30	0.0560
n/2 loc_only	47.20	42.90	40.64	40.33	87.68	0.0574

Each row is one fitted-model result. Boundary bins are [0, 0.25), [0.25, 0.5), [0.5, 0.75), and [0.75, 1]. Phase metrics use candidates detected in both grids at the candidate floor. Raw score fluctuation depends on the scoring scale.

Table S5 Additional local comparator adaptations.

Local adaptation	Sn@4 FPI (%) [95% CI]	FROC score	AP50 (%)
DRN + ROI adaptation [10]	0.00 [0.00, 0.00]	0.000099	0.000161
PRISM-DR adaptation [12]	0.69 [0.15, 1.64]	0.004548	0.108637

Both adaptations use the common 67-image test partition and lesion matcher. Their implementation details below define the scope of these supplementary results.

DRN + ROI used a dilated ResNet50 with RoI processing and approximate morphology, including line lengths 3–18 and up to 120 candidates per patch [10]. Candidate centers covered 47/1,445 reference lesions within 12 pixels before regression. Its 4-FPI cutoff retained no candidates, giving zero sensitivity for this local adaptation.

PRISM-DR used the released retinal cropper, green-channel processing, and three-of-five voting among locally retrained MA detectors [12]. Five seeds on the common split replaced the original cross-validation ensemble, and inference used 640-pixel tiles with 20% overlap rather than 1,280-pixel tiles with 25% overlap. Released detector weights were excluded to avoid potential overlap with the local IDRiD test partition. Source-specific FROC values were 0.0001, 0.1346, and 0.1913 on IDRiD, ROC, and e-ophtha, respectively.

Table S6 Source-specific comparative sensitivity and image-bootstrap intervals.

Method	IDRiD Sn@4 (%) [95% CI]	ROC Sn@4 (%) [95% CI]	e-ophtha Sn@4 (%) [95% CI]
TPW-YOLO (A6)	33.64 [28.81, 37.54]	23.26 [13.33, 46.56]	55.47 [46.99, 62.92]
FPN + FSAF adaptation [9, 24]	29.03 [25.70, 31.99]	25.58 [17.57, 44.45]	42.34 [32.12, 51.73]
Unified MA stream [20]	21.38 [17.42, 25.34]	12.79 [3.11, 33.87]	54.74 [46.97, 63.12]
SSPNET + Trident adaptation [8]	0.92 [0.22, 1.64]	22.09 [13.54, 36.67]	26.64 [18.86, 34.34]
PMNPDR partial adaptation [19]	0.65 [0.11, 1.35]	13.95 [7.29, 25.53]	18.25 [12.23, 24.29]

Intervals use 2,000 resamples of source images at each fitted model's source-specific full-sample 4-FPI threshold.

Table S7 Within-cohort paired sensitivity differences at pooled 4-FPI thresholds.

Scale / seed	Lesion set	Contrast	Difference (pp)	Paired 95% interval (pp)
s/0	All lesions	A4-A3	+0.83	[-0.38, +2.26]
s/0	All lesions	A6-A4	+0.62	[-0.34, +1.51]
s/0	All lesions	A6-A3	+1.45	[+0.14, +2.82]
s/0	All lesions	A6-A7	+0.69	[-0.24, +1.49]
s/0	Boundary	A4-A3	+0.00	[-4.29, +4.19]
s/0	Boundary	A6-A4	+3.73	[+1.23, +6.29]
s/0	Boundary	A6-A3	+3.73	[-0.91, +8.21]
n/0	All lesions	A4-A3	+2.49	[+1.26, +3.65]
n/0	All lesions	A6-A4	+1.25	[-0.54, +3.28]
n/0	All lesions	A6-A3	+3.74	[+1.91, +5.39]
n/0	All lesions	A6-A7	+2.42	[+0.77, +4.05]
n/0	All lesions	A4-loc_only	-0.07	[-1.46, +1.28]
n/0	All lesions	A4-no_gate	+2.35	[+0.44, +4.11]
n/0	All lesions	A3-score_only	+0.42	[-0.79, +1.64]
n/1	All lesions	A4-A3	+6.85	[+5.34, +8.35]
n/1	All lesions	A6-A4	+1.11	[-0.16, +2.40]
n/1	All lesions	A6-A3	+7.96	[+6.51, +9.45]
n/1	All lesions	A6-A7	+0.83	[-0.36, +2.19]
n/1	All lesions	A4-loc_only	+0.00	[-1.34, +1.23]
n/1	All lesions	A4-no_gate	+1.04	[-0.22, +2.35]
n/1	All lesions	A3-score_only	+0.83	[-0.23, +1.95]
n/2	All lesions	A4-A3	+0.76	[-0.40, +1.86]
n/2	All lesions	A6-A4	+1.80	[+0.58, +3.02]
n/2	All lesions	A6-A3	+2.56	[+0.88, +4.11]
n/2	All lesions	A6-A7	+1.04	[+0.23, +1.89]
n/2	All lesions	A4-loc_only	-0.90	[-2.03, +0.25]
n/2	All lesions	A4-no_gate	+0.55	[-0.72, +1.75]
n/2	All lesions	A3-score_only	-0.76	[-1.98, +0.50]

Scale / seed	Lesion set	Contrast	Difference (pp)	Paired 95% interval (pp)
n/0	Boundary	A4-A3	+3.11	[-0.65, +6.25]
n/0	Boundary	A6-A4	-0.62	[-4.42, +3.59]
n/0	Boundary	A6-A3	+2.48	[-2.91, +7.14]
n/1	Boundary	A4-A3	+9.94	[+5.22, +15.04]
n/1	Boundary	A6-A4	+3.11	[-1.01, +7.62]
n/1	Boundary	A6-A3	+13.04	[+8.62, +17.76]
n/2	Boundary	A4-A3	+1.86	[-2.06, +5.59]
n/2	Boundary	A6-A4	+1.86	[-2.86, +6.54]
n/2	Boundary	A6-A3	+3.73	[-1.41, +7.83]

Positive values favor the first named configuration. All-lesion and nearest-boundary denominators are 1,445 and 161. Paired image-bootstrap intervals use 2,000 resamples, RNG seed 710, and fixed full-sample thresholds. All n-scale contrasts use matched architecture and initialization.

Table S8 Coverage of the stored candidate pool.

Scale / seed	Configuration	Candidates	Matched / 1445	Pool recall (%)
s/0	A3	29520	1268	87.75
s/0	A4	29291	1269	87.82
s/0	A6	29291	1269	87.82
n/0	A3	38609	1295	89.62
n/0	A4	39741	1302	90.10
n/0	A6	39741	1302	90.10
n/1	A3	39026	1300	89.97
n/1	A4	40038	1302	90.10
n/1	A6	40038	1302	90.10
n/2	A3	35451	1295	89.62
n/2	A4	36266	1296	89.69
n/2	A6	36266	1296	89.69

Coverage uses all stored candidates at detector floor 10−4 and GT-first matching within 12 pixels. A6 shares A4 boxes. This metric describes pool availability before imposing a finite FPI constraint.

Table S9 Annotation-only reconstruction of visibility-gated pairing.

Seed	Eligible observations	Eligible / slot	Eligible / common (%)	Unique lesions covered (%)	Zero-eligible slots (%)
0	10577	0.908	90.22	87.19	66.64
1	10297	0.884	89.49	83.29	66.75
2	10389	0.892	88.92	85.83	66.84

Each epoch-29 reconstruction uses 212 images, 3,614 lesions, and 11,650 cyclic pair slots. Common observations meet detection retention in both views; eligible observations also meet inset-region visibility. Counts include repeated lesion observations before model-dependent positive assignment. Zero-eligible slots describe individual sampling slots.

Table S10 Descriptive cross-method computational records.

Method	Mean ms/image	P90 ms/image	Peak allocated GPU memory (MiB)
TPW-YOLO (A6)	1362.1	3307.7	137.1
FPN + FSAF adaptation [9, 24]	10050.4	21070.0	122.4
Unified MA stream [20]	4459.5	5765.8	711.1
SSPNET + Trident adaptation [8]	714.8	1139.1	8139.2
PMNPDR partial adaptation [19]	1807.7	3733.6	539.6

Records include method-specific preprocessing and candidate extraction, excluding file reading. Measurements were acquired in different sessions. TPW-YOLO denotes the primary s-scale seed-0 system.

Table S11 Complete per-run detection and operating-point measurements.

Scale / seed	Configuration	Sn@4 (%)	FROC	AP50 (%)	FPI at Sn50
s/0	A3 Random shift	41.94	0.2521	39.57	7.49
s/0	A4 Full TPCL	42.77	0.2570	41.13	6.64
s/0	A6 TPW-YOLO	43.39	0.2635	39.67	6.60
s/0	A7 Reduced features	42.70	0.2482	38.12	9.91
n/0	A3 Random shift	41.45	0.2093	36.77	7.37
n/0	A4 Full TPCL	43.94	0.2791	40.43	6.12
n/0	A6 TPW-YOLO	45.19	0.2839	38.53	6.25
n/0	A7 Reduced features	42.77	0.2644	35.48	9.33
n/0	TPCL without gate	41.59	0.2607	40.29	7.60
n/0	Score consistency only	41.04	0.2147	37.03	6.73
n/0	Location consistency only	44.01	0.2665	39.02	6.04
n/1	A3 Random shift	37.09	0.1496	33.72	8.00
n/1	A4 Full TPCL	43.94	0.2668	40.57	6.19
n/1	A6 TPW-YOLO	45.05	0.2739	36.36	8.12
n/1	A7 Reduced features	44.22	0.2577	36.16	10.67
n/1	TPCL without gate	42.91	0.2706	40.80	6.34
n/1	Score consistency only	36.26	0.1456	34.02	7.51
n/1	Location consistency only	43.94	0.2724	41.00	5.99
n/2	A3 Random shift	40.62	0.2631	39.59	6.75
n/2	A4 Full TPCL	41.38	0.2501	39.32	6.84
n/2	A6 TPW-YOLO	43.18	0.2531	37.73	6.97
n/2	A7 Reduced features	42.15	0.2566	39.24	6.64
n/2	TPCL without gate	40.83	0.2508	38.97	7.30
n/2	Score consistency only	41.38	0.2455	38.23	7.00
n/2	Location consistency only	42.28	0.2602	39.61	7.06

Values are individual fitted-model observations, with separate operating thresholds. FPI at Sn50 refers to at least 50% lesion sensitivity in the retained pool. Scale-specific runs are kept separate.

Table S12 Observed pair and batch counts in the n-scale seed-0 A4 fit.

Epoch	Batches	Pair slots	Geometric pairs	Assigned pairs	Zero-pair batches
16	2913	11650	10290	10290	556
17	2913	11650	10647	10647	554
18	2913	11650	10301	10301	600
19	2913	11650	10269	10269	595
20	2913	11650	10714	10714	570
21	2913	11650	10910	10910	585
22	2913	11650	10723	10723	531
23	2913	11650	10346	10346	555
24	2913	11650	10566	10566	542
25	2913	11650	10450	10450	585
26	2913	11650	10654	10654	529
27	2913	11650	10327	10327	598
28	2913	11650	10629	10629	581
29	2913	11650	10414	10414	619
30	2913	11650	10577	10577	552
Total	43695	174750	157817	157817	8552

Counts include repeated lesion observations during active-consistency epochs 16–30. Assigned pairs have positive detection assignments in both views. All recorded geometric pairs passed assignment; 19.57% of batches had no assigned pair. Table S9 gives the separate annotation-only reconstruction.

Table S13 Source allocation at the pooled operating thresholds in the n cohort.

Seed / system	Source	TP / GT	FP	FP/image	Sensitivity (%)
0 / A3	eophtha	102/274	50	1.67	37.23
0 / A3	idrid	486/1085	198	7.33	44.79
0 / A3	roc	11/86	16	1.60	12.79
0 / A4	eophtha	99/274	20	0.67	36.13
0 / A4	idrid	524/1085	237	8.78	48.29
0 / A4	roc	12/86	11	1.10	13.95
0 / A6	eophtha	94/274	17	0.57	34.31
0 / A6	idrid	547/1085	243	9.00	50.41
0 / A6	roc	12/86	8	0.80	13.95
1 / A3	eophtha	90/274	15	0.50	32.85
1 / A3	idrid	434/1085	246	9.11	40.00
1 / A3	roc	12/86	7	0.70	13.95
1 / A4	eophtha	103/274	24	0.80	37.59
1 / A4	idrid	520/1085	233	8.63	47.93
1 / A4	roc	12/86	9	0.90	13.95
1 / A6	eophtha	108/274	27	0.90	39.42
1 / A6	idrid	529/1085	231	8.56	48.76
1 / A6	roc	14/86	7	0.70	16.28
2 / A3	eophtha	96/274	19	0.63	35.04
2 / A3	idrid	481/1085	241	8.93	44.33
2 / A3	roc	10/86	6	0.60	11.63

Seed / system	Source	TP / GT	FP	FP/image	Sensitivity (%)
2 / A4	eophtha	95/274	16	0.53	34.67
2 / A4	idrid	494/1085	244	9.04	45.53
2 / A4	roc	9/86	6	0.60	10.47
2 / A6	eophtha	99/274	15	0.50	36.13
2 / A6	idrid	515/1085	247	9.15	47.47
2 / A6	roc	10/86	5	0.50	11.63

A single pooled operating threshold is applied to all sources within each seed and system. The FP budget is shared across 67 images. Source-specific sensitivities in Table S2 instead use each source's own threshold curve.

Table S14 Lesion sensitivity (%) at the seven FPI budgets in the primary s-scale seed-0 experiment.

Method	0.125	0.25	0.5	1	2	4	8
TPW-YOLO (A6)	7.61	10.10	15.92	22.84	32.94	43.39	51.63
A3 Random shift	4.36	9.83	13.98	23.04	32.18	41.94	51.14
A4 Full TPCL	5.61	9.00	15.36	22.91	30.93	42.77	53.36
FPN + FSAF adaptation [9, 24]	3.67	7.40	11.83	17.92	26.92	35.43	41.52
Unified MA stream [20]	3.88	5.47	8.58	8.58	22.08	30.38	37.16
SSPNET + Trident adaptation [8]	1.52	2.01	3.04	4.22	5.54	7.06	9.07
PMNPDR partial adaptation [19]	0.07	0.14	0.14	0.76	1.31	2.70	6.44
A7 Reduced features	5.74	7.75	14.88	22.21	32.73	42.70	47.75

Each operating point uses its own attainable threshold. A3 and A4 are the matched random-shift and full-TPCL controls; A7 is the reduced-feature ranker.

Table S15 Source-specific comparative performance.

Method	IDRiD Sn@4 (%) / FROC	ROC Sn@4 (%) / FROC	e-ophtha MA Sn@4 (%) / FROC
TPW-YOLO (A6)	33.64 / 0.1938	23.26 / 0.1694	55.47 / 0.4077
FPN + FSAF adaptation [9, 24]	29.03 / 0.1695	25.58 / 0.1445	42.34 / 0.2732
Unified MA stream [20]	21.38 / 0.1194	12.79 / 0.0930	54.74 / 0.3306
SSPNET + Trident adaptation [8]	0.92 / 0.0063	22.09 / 0.1346	26.64 / 0.1658
PMNPDR partial adaptation [19]	0.65 / 0.0037	13.95 / 0.0814	18.25 / 0.0938

Each source uses its own operating curve for Sn@4 and FROC. IDRiD, ROC, and e-ophtha contain 27/1,085, 10/86, and 30/274 test images/lesions, respectively. Source-specific intervals are provided in Table S6.

Table S16 Incremental IRBR scoring time on reused candidate records.

Cohort / seed	A4 mean ms/image	A6 mean ms/image	Added CPU ms/image
s / 0	1361.60	1362.12	0.52
n / 0	2275.72	2276.05	0.34
n / 1	2357.58	2357.94	0.36
n / 2	2144.76	2145.10	0.34

A6 reuses A4 candidate extraction and adds measured CPU ranking time. Within-run increments isolate scoring overhead; archived detector times from different sessions are reported in Table S3.

Table S17 Source allocation at the primary s-scale pooled operating thresholds.

System	Source	TP / GT	FP	FP/image	Sensitivity (%)
A4	eophtha	114/274	33	1.10	41.61
A4	idrid	491/1085	226	8.37	45.25
A4	roc	13/86	8	0.80	15.12
A6	eophtha	118/274	31	1.03	43.07
A6	idrid	498/1085	231	8.56	45.90
A6	roc	11/86	6	0.60	12.79

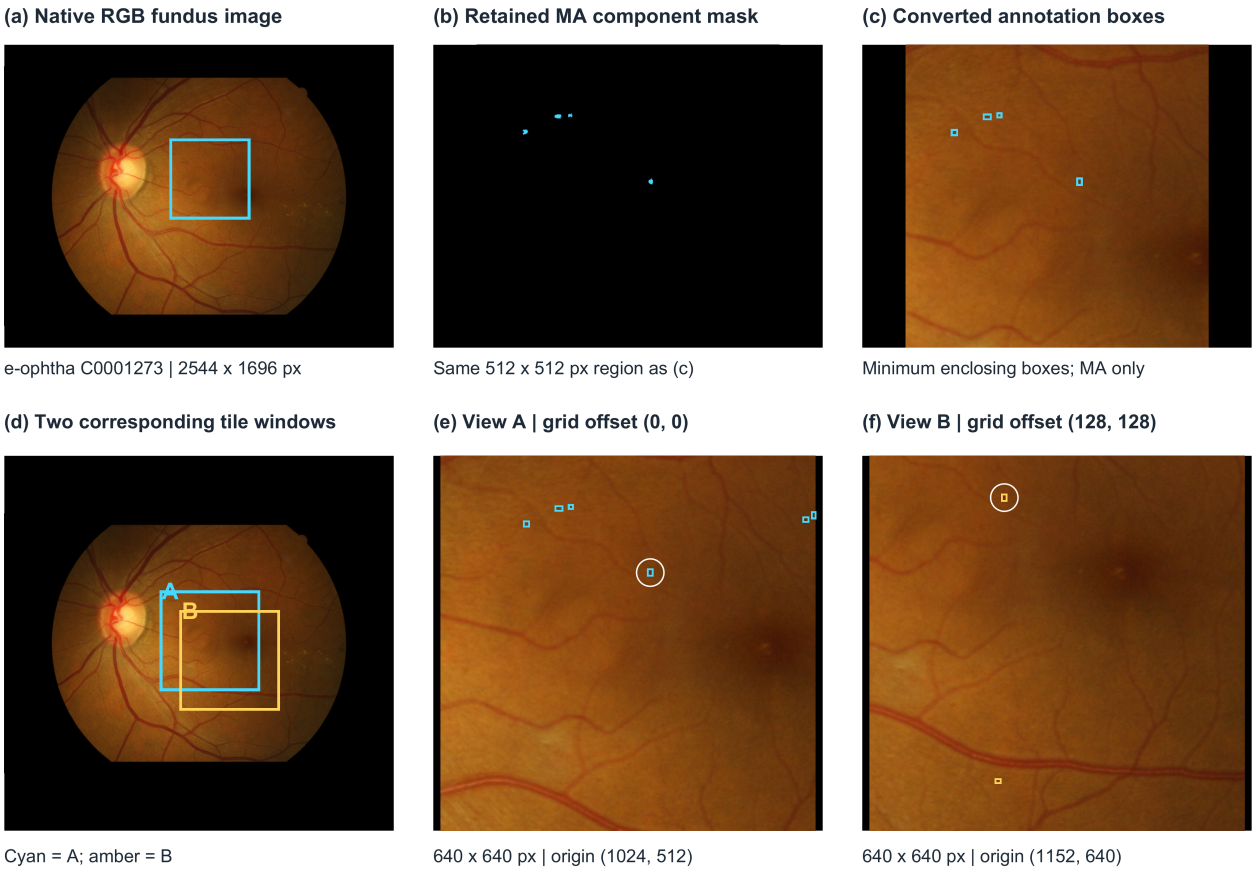
One global threshold is applied across the three sources within each system. A4 totals 618 TP and 267 FP; A6 totals 627 TP and 268 FP. Table S2 instead uses source-specific thresholds. Table S13 provides complete n-scale allocations.

Table S18 Boundary and phase diagnostics for the matched within-scale sequences.

Scale / seed	Configuration	Boundary TP/161	Both phases (%)	Rank fluctuation	Shift (px)
s / 0	A3	72/161	85.26	0.0486	1.532
s / 0	A4	72/161	84.64	0.0499	1.485
s / 0	A6	78/161	84.64	0.0735	1.485
n / 0	A3	76/161	87.06	0.0552	1.584
n / 0	A4	81/161	87.89	0.0552	1.541
n / 0	A6	80/161	87.89	0.0939	1.541
n / 1	A3	55/161	88.10	0.0525	1.573
n / 1	A4	71/161	88.44	0.0508	1.550
n / 1	A6	76/161	88.44	0.1107	1.550
n / 2	A3	74/161	87.34	0.0548	1.615
n / 2	A4	77/161	87.68	0.0603	1.665
n / 2	A6	80/161	87.68	0.0776	1.665

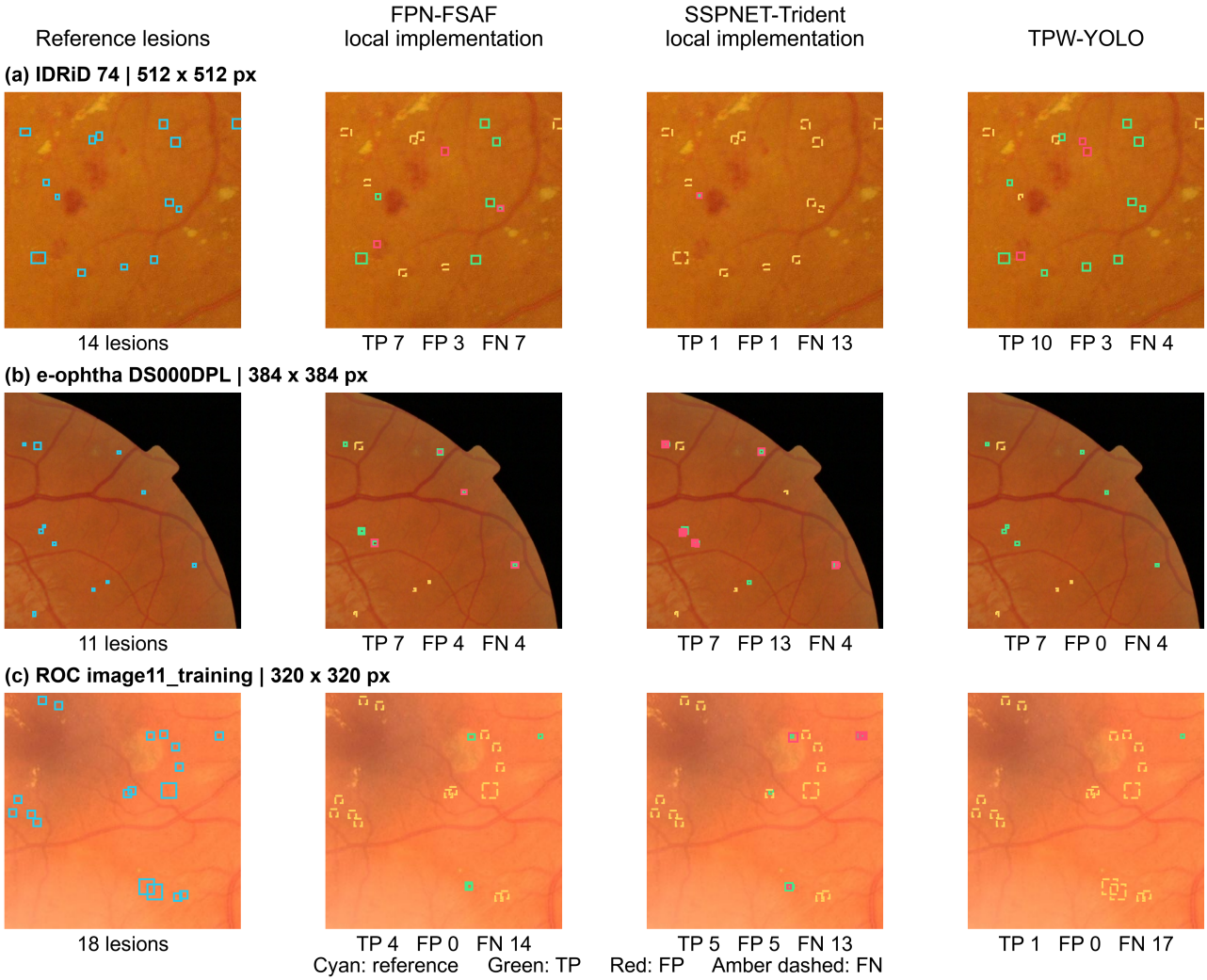
The boundary denominator is 161 lesions defined using the primary grid. Phase diagnostics use lesions detected in both independently evaluated grids at the candidate floor. Complete per-run measurements appear in Table S4.

Figure S1



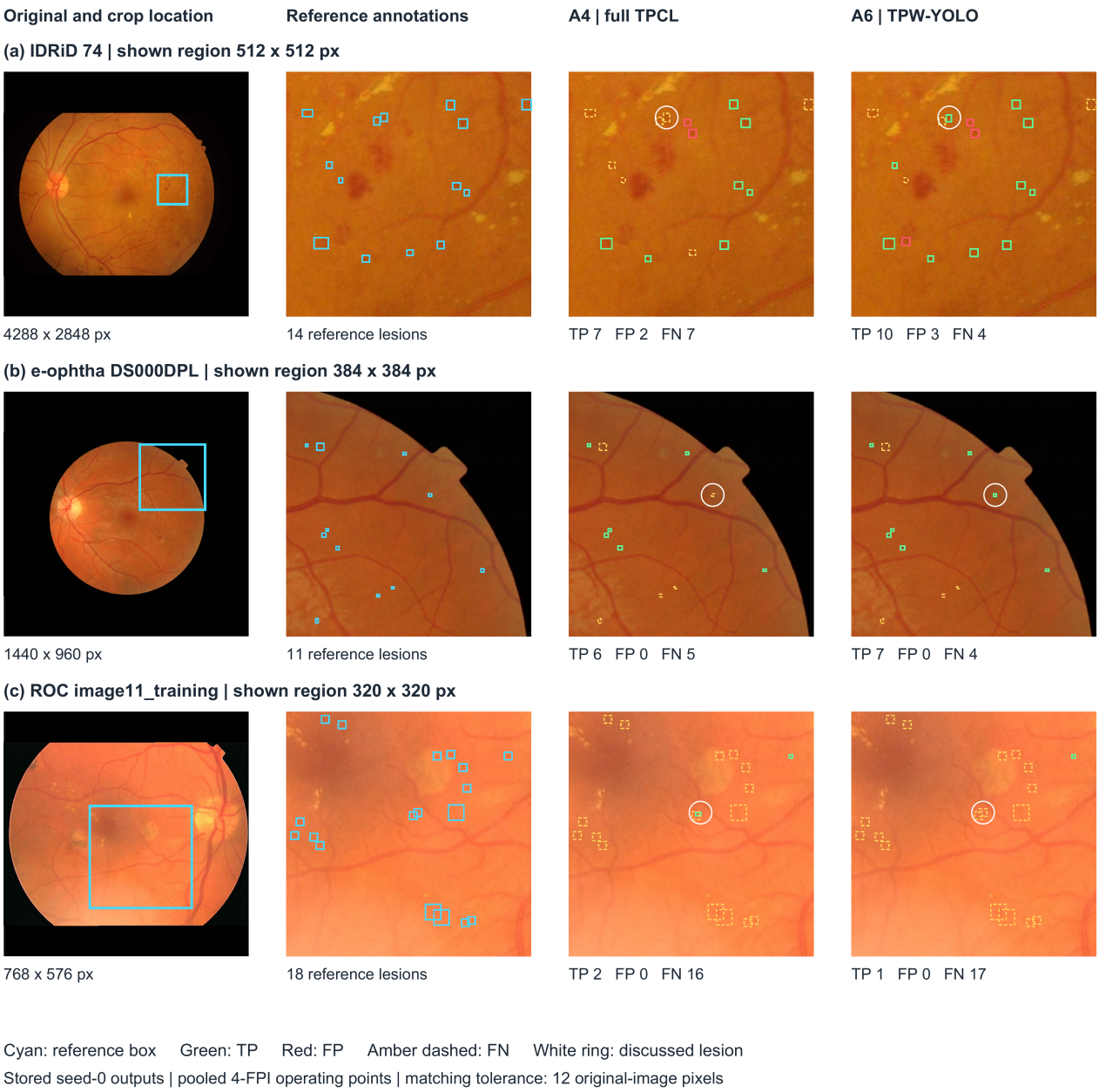
Annotation correspondence in paired tile views. (a) Native-resolution e-ophtha training image C0001273; the cyan square marks the enlarged region. (b,c) Component mask and enclosing boxes in the same 512 x 512-pixel region. (d) Corresponding tile windows. (e,f) Native 640 x 640-pixel crops at origins (1024, 512) and (1152, 640), respectively. Cyan and amber denote views A and B; white rings identify the same annotated lesion. The grids have offsets (0, 0) and (128, 128).

Figure S2



Qualitative comparison at the primary s-scale pooled 4-FPI operating points. Rows show identical regions from (a) IDRiD 74, (b) e-ophtha DS000DPL, and (c) ROC image11_training. Columns show reference annotations, the local FPN-FSAF and SSPNET-Trident implementations, and TPW-YOLO. Cyan denotes reference boxes, green true positives, red false positives, and dashed amber missed lesions. Each method uses its own pooled score threshold. One-to-one matching within 12 original-image pixels was performed on full images before cropping. Counts describe the displayed regions; the examples were selected to illustrate differing outcomes.

Figure S3



Candidate selection before and after IRBR at the primary s-scale pooled 4-FPI operating points. Rows show (a) IDRiD 74, (b) e-ophtha DS000DPL, and (c) ROC image11_training. Columns show the original image and crop location, reference annotations, A4 predictions, and A6 predictions. Cyan denotes reference lesions, green true positives, red false positives, and dashed amber missed lesions. White rings mark a lesion selected only by A6 in (a,b) and only by A4 in (c). Counts refer to displayed crops after full-image matching. A4 and A6 share candidate boxes and use their respective pooled thresholds.