This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Harmful or Helpful? When Precision-Recall Balance Governs Safe QGFL Transfer in Malaria Detection

Thabang Fenge Isaka[1][0009−0000−5216−112X], Claire Wynne[2][0000−0002−0481−8366], and Jane Courtney[1][0000−0002−9175−7855]

[1] School of Electrical and Electronic Engineering, Technological University Dublin, Dublin, Ireland
d23125116@mytudublin.ie
[2] School of Biological, Health and Sports Sciences, Technological University Dublin, Dublin, Ireland

Abstract. Malaria's burden falls hardest where reliable microscopy is scarcest, so a detector tuned in one context must stay safe when redeployed in another. This tempts a shortcut: carrying a recall-boosting loss, Quality-Guided Focal Loss (QGFL), unchanged from one detector to the next. We test it by transferring QGFL across three detectors, YOLOv8s, YOLOv11s and RT-DETR, on three malaria datasets, across 135 training runs. The shift towards recall is universal, but its effect on F1 is not. Aggressive re-weighting helps a minority and quietly harms the rest. What decides the outcome is one measurable property of the untouched baseline, its precision–recall balance, not its architecture or accuracy. Matched baselines with opposite balance moved in opposite directions, and the verdict survives a recall-weighted re-scoring. This yields a pre-deployment check, applying the aggressive loss only when the baseline is precision-dominant and flagging harmful transfer before retraining. The lesson reaches past this loss: reliability comes from matching the intervention to the baseline's error profile, not from the strongest one. Code: https://github.com/ThabangIsaac1/qgfl-transfer-malaria.

Keywords: Medical Image Analysis · Object Detection · Loss Function Transfer · Precision-Recall Trade-off · Malaria Detection · Computational Pathology

1 Introduction

Malaria's burden is heaviest in sub-Saharan Africa, where reliable microscopy is scarcest at the point of care [16]. Automated parasite detection therefore matters most when a detector is deployed to a new or under-resourced site, not the laboratory that tuned it. The task couples a sparse target under extreme class imbalance with an asymmetric error cost: a missed parasite delays treatment while excess false positives burden the reviewing expert [16,12]. Focal weighting [11] and hard-example mining [14] address this imbalance for one detector

at a fixed operating point. Such a loss is routinely carried to a new detector without testing whether the precision–recall trade-off that justified it still holds. This blind loss-function transfer is the risk.

Quality-Guided Focal Loss (QGFL) couples localisation-quality weighting with difficulty-based focusing [7,10,18], trading precision for recall to recover sparse positives. Its published validation used a single RetinaNet baseline, whose native loss is the Focal Loss QGFL extends, that was recall-deficient ($R < P$) on all three datasets [7]. There, a loss shifting towards recall tends to raise F1, so the reported gains were structurally favoured by that baseline rather than shown to be a property of the loss. Whether they survive a baseline that is not recall-deficient was acknowledged as open but never tested, because the architecture was held fixed.

Malaria detection has since moved to anchor-free and attention-based YOLO families [8,9,17] and set-prediction transformers such as RT-DETR [19,1]. These need not begin from RetinaNet's recall-deficient balance, so QGFL's recall trade may help some and harm others. Transfer has precedent, though not for a fixed loss under a changing imbalance regime. A learned augmentation policy ports across detectors and datasets, but alters only training, not the deployed model [20]. A searched loss surpasses handcrafted losses across RetinaNet, Faster R-CNN, and Deformable DETR [15], but only on balanced benchmarks, absent sparse detection's precision–recall stakes. How a hand-designed loss behaves when both architecture and imbalance regime change is uncharacterised.

We characterise this transfer across YOLOv8s, YOLOv11s, and RT-DETR on three datasets spanning *P. falciparum*, *P. vivax*, and multi-species smears (135 runs, five seeds each), an exploratory screen of nine configurations. Aggressive QGFL is not safe by default: the shift towards recall is universal but its effect on net F1 is not, improving three and degrading six. Our contributions are threefold: (i) a systematic study of the transfer safety of a hand-designed detection loss across CNN and transformer detectors, resolving where it helps and where it harms rather than a pooled average; (ii) evidence that the outcome is governed by the baseline precision–recall balance, not architecture or accuracy, with two configurations at matched baseline F1 but opposite balance moving in opposite directions; and (iii) a pre-deployment screening rule that adopts aggressive QGFL only when the baseline is precision-dominant ($P/R > 1.5$), robust to a recall-weighted (F2) re-scoring.

2 Methods

We compare each detector against itself under three classification losses to isolate loss-function transfer from architectural confounds, characterise reliability across imbalance regimes, and probe whether strong and lower-intensity perturbations behave alike. The nine architecture–dataset configurations are not statistically independent, sharing datasets and architectures; we therefore limit confirmatory testing to a sign test on the directional shift and treat the configuration-level regression and matched-baseline comparison as exploratory at $n = 9$.

2.1 Quality-Guided Focal Loss

QGFL [7] extends Focal Loss by reweighting each example's classification loss according to its difficulty and the detector's IoU-aware target quality. Let z be the predicted logit, $p{=}\sigma(z)$, and $y \in [0,1]$ the detector's quality (soft) target, which already embeds the IoU of the matched box; QGFL modulates the binary cross-entropy $\ell_{\mathrm{BCE}}(z,y)$ as

$$\mathrm{QGFL} = \alpha_c\,(1-p_t)^{\gamma_{\mathrm{eff}}}\,(1+w_q)\,\ell_{\mathrm{BCE}}(z,y), \qquad p_t = py + (1-p)(1-y), \quad (1)$$

with class weight α_c, a difficulty-thresholded focusing exponent

$$\gamma_{\mathrm{eff}} = \gamma_0 + (\gamma^c_{\max} - \gamma_0)\,\frac{\max\big((1-p_t)-\tau,\,0\big)}{1-\tau}, \quad (2)$$

and a quality weight that amplifies predictions far from the target,

$$w_q = \min\big(\max(|p-y|-m,\,0)^{\beta},\,10\big). \quad (3)$$

The IoU thus enters through y, not a separate term; $\gamma_0{=}2$ and $\beta{=}2$ throughout, and the remaining parameters are given in Table 1. A per-epoch factor scaling the QGFL term decays linearly from 2.0 to 0.5 over training.

Table 1 lists the two parameterisations we evaluate. The aggressive setting uses the RetinaNet parameters from [7], concentrating loss on infected examples with steep focusing; the calibrated setting halves the focusing intensities and rebalances α towards the class frequencies, a softer perturbation for stronger baselines.

Table 1. QGFL settings. Aggressive follows the original [7] parameterisation; calibrated reduces all focusing intensities.

Parameter	Aggressive (Original)	Calibrated
α_{inf} / α_{uninf}	0.9 / 0.1	0.35 / 0.65
$\gamma_{\max,\mathrm{inf}}$ / $\gamma_{\max,\mathrm{uninf}}$	8.0 / 4.0	4.0 / 2.0
Difficulty threshold	0.925	0.6
Quality margin	0.5	0.3

2.2 Architecture Integration and Training

We integrate the same QGFL, a BCE-based extension of Focal Loss, into three modern detectors: for the anchor-free YOLOv8s and YOLOv11s [8,9] it replaces the BCE classification term in the detection head, and for the transformer-based RT-DETR [19] it replaces the VarifocalLoss term inside the Hungarian matching objective. Box regression and distribution focal loss are left untouched. Because QGFL is applied identically while each baseline keeps its native classification loss, transferring it to RT-DETR also swaps VarifocalLoss for a BCE base. We

evaluate transfer within each detector's native setup rather than artificially standardising the baselines, and treat the resulting BCE/VarifocalLoss confound as a scope condition.

All models were trained for up to 200 epochs (early stopping, patience 20; the best-fitness checkpoint is evaluated) at batch size 16 and 640×640 input resolution using the Ultralytics framework, each detector under its native optimiser. The YOLO detectors used SGD ($\mathrm{lr} = 0.01$, $\mathrm{momentum} = 0.95$, weight $\mathrm{decay} = 0.0005$) with a linear decay to $0.01 \times$ the initial rate; RT-DETR used the framework's default AdamW. Evaluation uses a confidence threshold of 0.25 and an IoU threshold of 0.45 with class-agnostic NMS, following Guemas et al. [4]. The low confidence threshold reflects the clinical priority of not missing sparse infected cells; class-agnostic NMS prevents a single cell being retained twice when the infected- and uninfected-class heads both fire on it, otherwise inflating detected counts and apparent false-positive rates. All runs used a single NVIDIA RTX 8000 GPU (48 GB).

2.3 Experimental Design

Table 2 summarises the three source datasets[3]: a small laboratory *P. falciparum* set (D1 [3]), a mid-scale *P. vivax* patient set (D2 [6]), and a large multi-species multi-site set (D3 [4]), all standardised to binary infected/uninfected detection; splits follow each dataset's original train/val/test partition. Infected cells are the minority throughout, roughly 13 to 70 uninfected per infected cell across splits. The full factorial design crosses 3 datasets, 3 architectures, and 3 loss settings (baseline, aggressive QGFL, calibrated QGFL), each combination trained over 5 random seeds (42, 123, 456, 7, 2024) for 135 training runs in total. The primary evaluation metric is infected-class F1, chosen for its sensitivity to minority-class performance; precision and recall are reported alongside it. The abundant uninfected class stays consistently high throughout. We summarise each architecture–dataset configuration's operating regime by its baseline precision-to-recall ratio (P/R), the per-seed mean of infected-class precision divided by the per-seed mean of infected-class recall, used as the screening variable. Effect sizes are reported as Hedges' g [5] with magnitude thresholds following Cohen [2].

Table 2. Dataset characteristics. Splits follow each dataset's original train/val/test partition.

Dataset	Images	Train	Val	Test	Inf. %
D1 [3]	398	225	57	116	6.6
D2 [6]	1,328	966	242	120	2.9
D3 [4]	29,228	20,830	3,890	4,508	2.6

[3] Public dataset sources: D1 https://data.mendeley.com/datasets/j55fyhtxn4/1; D2 https://bbbc.broadinstitute.org/BBBC041; D3 https://zenodo.org/records/8358829.

3 Results

3.1 A Universal Shift, Conditional in Effect

Across all nine architecture–dataset configurations, aggressive QGFL reduced infected-class precision and increased recall. Because these configurations are not statistically independent, we read the sign test as a conservative directional check: 9/9 precision decreases and 9/9 recall increases ($p = 0.004$). This direction is mechanistically expected, the aggressive setting concentrating loss on infected-class examples (Table 1) at the expense of uninfected ones, raising recall while eroding false-positive suppression; the contribution is the conditional magnitude across baseline regimes, unknown beyond RetinaNet.

Whether the shift helps or harms is conditional (Table 3). On D2, whose baselines were the most recall-deficient, aggressive QGFL improved infected-class F1 across all three architectures (up to ΔF1 of +0.258); on the recall-sufficient D1 and D3 baselines it degraded F1 (worst −0.478, on D3 RT-DETR). The ‡ markers on the D3 YOLO configurations reflect very low within-group variance (baseline SD below 0.01); on D3 RT-DETR the same marker reflects a genuinely large effect at a normal SD of 0.092. Calibrated QGFL was weaker and mixed, improving all D1 configurations but degrading two of three on D2.

Table 3. Infected-class F1 (mean ± std, $n = 5$ seeds). ΔF1 is the change from the per-config baseline. Hedges' g: $^{\dagger}|g| > 1.0$, $^{\ddagger}|g| > 2.0$. **Bold** marks a large-effect improvement. The ‡ markers can reflect very low seed variance rather than a large absolute change.

		Baseline	Aggressive QGFL		Calibrated QGFL	
Dataset	Model	F1	F1	ΔF1	F1	ΔF1
D1	RT-DETR	0.627 ± 0.200	0.439 ± 0.049	$-0.188^{\dagger}$	0.795 ± 0.026	$+\mathbf{0.169}^{\dagger}$
	YOLOv11s	0.835 ± 0.025	0.490 ± 0.083	$-0.344^{\ddagger}$	0.841 ± 0.055	$+0.006$
	YOLOv8s	0.765 ± 0.025	0.441 ± 0.064	$-0.324^{\ddagger}$	0.821 ± 0.070	$+0.056$
D2	RT-DETR	0.627 ± 0.065	0.702 ± 0.063	$+\mathbf{0.075}^{\dagger}$	0.556 ± 0.047	$-0.071^{\dagger}$
	YOLOv11s	0.456 ± 0.116	0.598 ± 0.070	$+\mathbf{0.143}^{\dagger}$	0.380 ± 0.054	-0.075
	YOLOv8s	0.356 ± 0.093	0.614 ± 0.026	$+\mathbf{0.258}^{\ddagger}$	0.458 ± 0.053	$+\mathbf{0.102}^{\dagger}$
D3	RT-DETR	0.809 ± 0.092	0.331 ± 0.093	$-0.478^{\ddagger}$	0.880 ± 0.018	$+0.072$
	YOLOv11s	0.921 ± 0.004	0.848 ± 0.019	$-0.074^{\ddagger}$	0.903 ± 0.004	$-0.018^{\ddagger}$
	YOLOv8s	0.911 ± 0.006	0.769 ± 0.010	$-0.141^{\ddagger}$	0.893 ± 0.006	$-0.018^{\ddagger}$

3.2 Predicting Harm vs. Benefit with Baseline P/R Balance

To anticipate the outcome before retraining, we tested baseline metrics as predictors of the F1 change (ΔF1) under aggressive QGFL. The baseline P/R ratio was the most mechanistically interpretable, with an exploratory association robust to leave-one-out cross-validation ($R^2 = 0.71$, $p = 0.0045$; per-fold R^2 0.599–0.794; Figure 1). Baseline recall, an inverse view of the same mechanism, fit

marginally better ($R^2 = 0.76$, $p = 0.0022$) and baseline F1 worse ($R^2 = 0.52$, $p = 0.0284$). These results are consistent with the P/R balance, rather than overall performance level, shaping QGFL's net effect.

These results translate into a pre-deployment screening rule. Compute the baseline infected-class P/R ratio on a representative validation set: if the baseline is precision-dominant (P/R > 1.5 in our experiments), aggressive QGFL is a plausible intervention; if it is recall-sufficient or near-balanced (P/R $\lesssim 1.0$), it is not recommended. The empirical transition lies near P/R ≈ 1.5, between the highest recall-sufficient configuration (harmed) and the lowest precision-dominant configuration (benefited). Re-scoring with the recall-weighted F2 (β=2), computed from the same per-seed precision and recall, leaves this unchanged: six of nine configurations still degrade and baseline P/R predicts ΔF2 at least as well ($R^2 = 0.75$) with the same cutoff, so the screen is robust to the clinically motivated recall weighting.

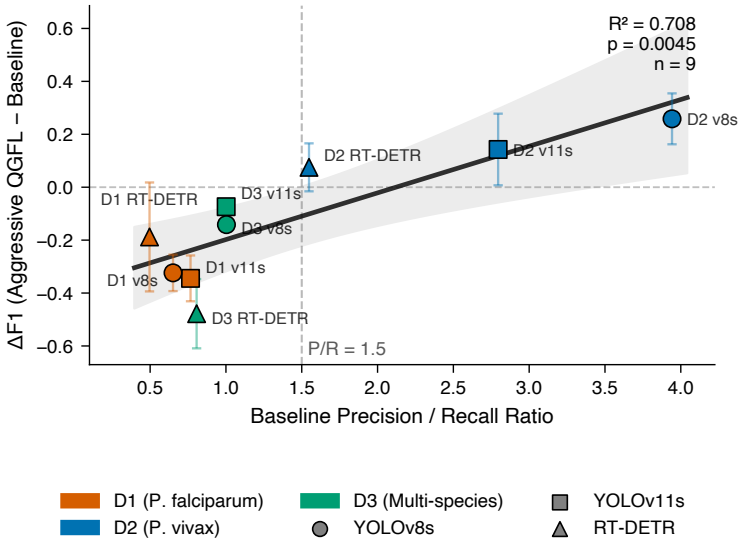

Fig. 1. Baseline P/R ratio vs. ΔF1 under aggressive QGFL ($n = 9$, exploratory). Horizontal dashed line: ΔF1=0 (points below it lost F1); vertical dashed line: P/R=1.5 screening cutoff. Vertical bars: ± 1 SD of ΔF1.

3.3 Isolating P/R from Performance Level

The regression alone cannot separate P/R balance from overall performance level; a matched-baseline pair can. Figure 2 contrasts one: D1 RT-DETR and D2 RT-DETR reached identical baseline F1 (0.627) through opposite P/R profiles, D1 recall-sufficient and D2 precision-dominant. Aggressive QGFL then moved them in opposite directions, ΔF1 $= -0.188$ versus $+0.075$. The D1 baseline shows high seed-to-seed variance (SD $= 0.200$, Table 3), but both shifts are large-effect ($|g| > 1.0$, Table 3), so the matched comparison holds.

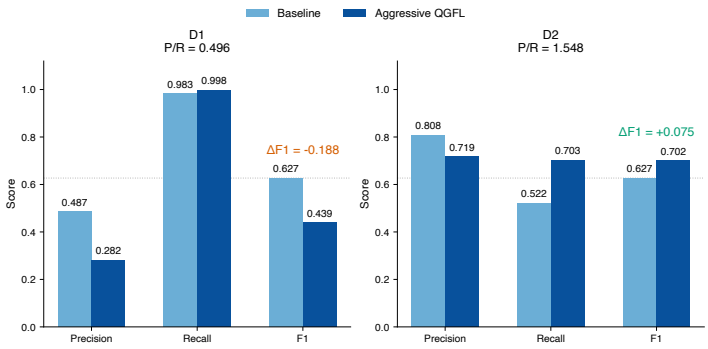

Fig. 2. Matched-baseline comparison: D1 and D2 RT-DETR start from identical baseline F1 (0.627) but opposite P/R. Each panel: precision, recall, and F1 for baseline vs. aggressive QGFL.

3.4 Missed Infections versus False Positives

The precision–recall shift translates into a detection-error trade-off, which Figure 3 makes visible at the image level. On a recall-deficient D2 case the baseline YOLOv8s misses the only infected cell, and QGFL recovers it. On a recall-sufficient D3 image, aggressive QGFL adds false positives where the baseline was already correct. Across the YOLO configurations this trade-off is consistent even where F1 declined: the recall gain recovers missed infections, the precision loss adds background false positives. RT-DETR does not follow this YOLO pattern on D1, consistent with its VarifocalLoss baseline and query-based head rather than YOLO's BCE head (Sec. 4, item 2).

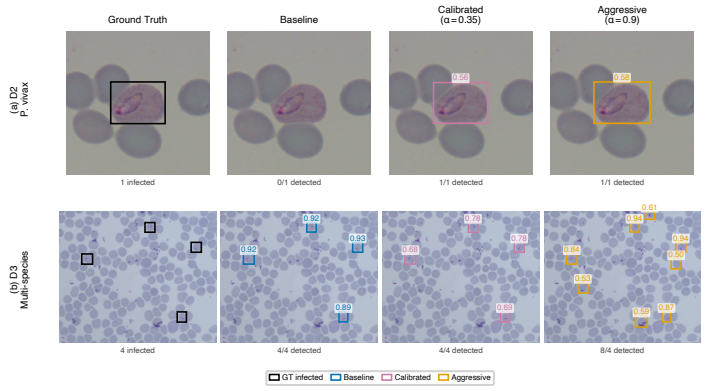

Fig. 3. Qualitative YOLOv8s detections (columns: ground truth, baseline, calibrated QGFL, aggressive QGFL). Box labels are infected-class confidence; the display threshold is 0.5. (a) D2 (*P. vivax*): the baseline misses the infected cell and QGFL recovers it. (b) D3 (multi-species): aggressive QGFL adds low-confidence background false positives, while baseline and calibrated stay accurate.

4 Limitations and Future Work

We study one representative recall-trading loss, not the best remedy for class imbalance. Examining a single intervention isolates the governing factor, the baseline precision–recall balance, so we make no claim that QGFL outperforms re-sampling, cost-sensitive learning, or threshold calibration. Several scope conditions bound where this evidence applies. (1) *Statistical*: the regressions are exploratory at $n = 9$, with non-independent observations from a crossed 3×3 design, and the two QGFL settings vary multiple parameters jointly, leaving unresolved which parameter drives the effect, though the governing-factor claim rests on the matched-baseline comparison, not the regression. (2) *Architecture confound*: QGFL uses a BCE base while the baselines keep their native losses (BCE for YOLO, VarifocalLoss for RT-DETR), and each architecture uses its native optimiser (SGD for YOLO, AdamW for RT-DETR); cross-architecture transfer therefore mixes the reweighting with base-loss and optimiser differences, all held fixed within the RT-DETR matched pair. (3) *Screening-rule coverage*: no configuration falls between P/R $\approx$ 1.0 and 1.5, so this transition is interpolated, a screening cutoff rather than a calibrated threshold; and because the underlying regression covered only aggressive QGFL, the rule does not extend to the calibrated setting. (4) *Generalisation*: the screen was derived from small-scale detectors (YOLOv8s, YOLOv11s, one RT-DETR) and three public datasets, none from African clinical cohorts or sites, so its transfer to larger families, further datasets, and African deployment conditions is untested.

The most direct extension tests whether the same P/R screen governs other recall-trading interventions, cost-sensitive reweighting and confidence-threshold calibration, and whether it holds within QGFL's own quality-aware lineage, GFL and VarifocalNet [10,18]. This would establish whether the predictor generalises beyond QGFL, and populating the unobserved 1.0–1.5 P/R interval would sharpen the boundary. The binary task also sets aside finer species and lifecycle-stage labels; whether the P/R screen transfers to these tasks, where per-class imbalance is more severe, remains open.

5 Conclusion

Loss-function transfer across detector architectures is not safe by default. Aggressive QGFL harmed more configurations than it helped, and the outcome was governed not by the detector family or its overall accuracy but by the baseline precision–recall balance, which a matched-baseline comparison isolated directly. Because that balance is measurable on the untouched baseline, harm can be screened before retraining. Apply aggressive re-weighting only when the detector is precision-dominant, an operating-point check of the kind now emphasised for clinical AI [13]. The broader lesson is that reliability comes not from adopting the strongest loss but from matching an intervention to the error profile of the baseline it modifies.

Acknowledgments. Funded by Taighde Éireann – Research Ireland through the Research Ireland Centre for Research Training in Machine Learning (18/CRT/6183).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: Computer Vision – ECCV 2020. LNCS, vol. 12346, pp. 213–229. Springer (2020). https://doi.org/10.1007/978-3-030-58452-8_13
2. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates, 2nd edn. (1988)
3. Davidson, M.S., Andradi-Brown, C., Yahiya, S., Chmielewski, J., O'Donnell, A.J., Gurung, P., Jeninga, M.D., Prommana, P., Andrew, D.W., Petter, M., Uthaipibull, C., Boyle, M.J., Ashdown, G.W., Dvorin, J.D., Reece, S.E., Wilson, D.W., Cunningham, K.A., Ando, D.M., Dimon, M., Baum, J.: Automated detection and staging of malaria parasites from cytological smears using convolutional neural networks. Biological Imaging **1**, e2 (2021). https://doi.org/10.1017/S2633903X21000015
4. Guemas, E., Routier, B., Ghelfenstein-Ferreira, T., Cordier, C., Hartuis, S., Marion, B., Bertout, S., Varlet-Marie, E., Costa, D., Pasquier, G.: Automatic patient-level recognition of four *Plasmodium* species on thin blood smear by a real-time detection transformer (RT-DETR) object detection algorithm: a proof-of-concept and evaluation. Microbiology Spectrum **12**(2), e01440–23 (2024). https://doi.org/10.1128/spectrum.01440-23
5. Hedges, L.V., Olkin, I.: Statistical Methods for Meta-Analysis. Academic Press (1985)
6. Hung, J., Carpenter, A.: Applying faster R-CNN for object detection on malaria images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 808–813 (2017). https://doi.org/10.1109/CVPRW.2017.112
7. Isaka, T.F., Wynne, C., Courtney, J.: Quality-guided focal loss: Enhancing minority class detection in haematological imaging. In: Tommasino, C., Russo, C., Bernardini, M. (eds.) Artificial Intelligence for Biomedical Data. pp. 197–211. Springer Nature Switzerland, Cham (2026). https://doi.org/10.1007/978-3-032-17216-7_16
8. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics YOLOv8. https://github.com/ultralytics/ultralytics (2023), accessed: 2024
9. Khanam, R., Hussain, M.: YOLOv11: An overview of the key architectural enhancements. arXiv preprint arXiv:2410.17725 (2024)
10. Li, X., Wang, W., Wu, L., Chen, S., Hu, X., Li, J., Tang, J., Yang, J.: Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. In: Advances in Neural Information Processing Systems 33 (NeurIPS 2020) (2020)
11. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 2999–3007 (2017). https://doi.org/10.1109/ICCV.2017.324

12. Mosquera, C., Ferrer, L., Milone, D.H., Luna, D., Ferrante, E.: Class imbalance on medical image classification: towards better evaluation practices for discrimination and calibration performance. European Radiology **34**(12), 7895–7903 (2024). https://doi.org/10.1007/s00330-024-10834-0
13. Rudolph, J., Huemmer, C., Preuhs, A., Buizza, G., Dinkel, J., Koliogiannis, V., Fink, N., Goller, S.S., Schwarze, V., Heimer, M., Hoppe, B.F., Liebig, T., Ricke, J., Sabel, B.O., Rueckel, J.: Threshold optimization in AI chest radiography analysis: integrating real-world data and clinical subgroups. European Radiology Experimental **9**, 95 (2025). https://doi.org/10.1186/s41747-025-00632-8
14. Shrivastava, A., Gupta, A., Girshick, R.: Training region-based object detectors with online hard example mining. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 761–769 (2016). https://doi.org/10.1109/CVPR.2016.89
15. Tao, C., Li, Z., Zhu, X., Huang, G., Liu, Y., Dai, J.: Searching parameterized ap loss for object detection. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. NIPS '21, Curran Associates Inc., Red Hook, NY, USA (2021)
16. World Health Organization: World malaria report 2024 (2024)
17. Zedda, L., Loddo, A., Di Ruberto, C.: YOLO-PAM: Parasite-attention-based model for efficient malaria detection. Journal of Imaging **9**(12), 266 (2023). https://doi.org/10.3390/jimaging9120266
18. Zhang, H., Wang, Y., Dayoub, F., Sünderhauf, N.: VarifocalNet: An IoU-aware dense object detector. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8514–8523 (2021)
19. Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., Chen, J.: DETRs beat YOLOs on real-time object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 16965–16974 (2024)
20. Zoph, B., Cubuk, E.D., Ghiasi, G., Lin, T.Y., Shlens, J., Le, Q.V.: Learning data augmentation strategies for object detection. In: Computer Vision – ECCV 2020. LNCS, vol. 12372, pp. 562–578. Springer (2020). https://doi.org/10.1007/978-3-030-58583-9_34