MaLDAM: Masked Localized Domain Adaptation for Malaria Detection in Low-Cost Microscopic Images

Muditha Fernando[1], Tishan Rathnasekara[1], Saeedha Nazar[1]*, Avishka Perera[1], and Tharindu Kaluarachchi[1]

University of Moratuwa, Sri Lanka
fernandoiamd.19@uom.lk, rathnasekarats.20@uom.lk, nazarfs.20@uom.lk, pereramat2000@gmail.com, tikzoxs@gmail.com

Abstract. Malaria diagnosis in resource-deficient regions increasingly relies on low-cost microscopes (LCMs), which offer an affordable alternative to high-cost microscopes (HCMs). Creating labeled LCM datasets is laborious because of poor image quality and limited field of view, and models trained on annotated HCM images often underperform on LCMs due to domain shift. Existing domain adaptation methods primarily focus on global feature alignment, despite malaria detection being a dense prediction task where local cellular characteristics are fundamental to microscopy-based examination. In this work, we propose MaLDAM, a cell-aware domain adaptation framework for LCM malaria detection that jointly performs global and masked local feature alignment between paired HCM and LCM images. By leveraging precomputed cell-containing patch masks, our method focuses adaptation on diagnostically relevant cellular regions while reducing the influence of background variations. We further apply microscopy-specific augmentations to improve robustness to low-cost imaging artefacts. On the M5-dataset, our method improves mean average precision (mAP) by **8.5** at 1000×, **7.3** at 400×, and **3** at 100× magnification without additional inference latency over the previous state-of-the-art. MaLDAM demonstrates more robust domain-invariant representation learning for practical malaria diagnosis using affordable microscopy devices. Our code is publicly available: https://github.com/TishanSathruwan/MaLDAM.

Keywords: Domain Adaptation · Low-Cost Microscopy · Parasitemia Detection

1 Introduction

Malaria remains a serious global health burden. Over the past decade, between 2015 and 2024, the estimated number of malaria cases increased from 230 million to 282 million (22.6%), while deaths increased from 578,000 to 610,000 (5.5%) [24]. This corresponds to more than one malaria-related death every

* Corresponding author.

minute. Still, malaria is both preventable and curable with early diagnosis. Microscopy is an established method for parasite-based malaria diagnosis [23]. However, malaria continues to disproportionately affect tropical and sub-tropical regions with poor healthcare infrastructure, making early detection a bottleneck. To improve diagnostic accessibility in such resource-deficient areas, low-cost microscopes (LCMs), which are over 70% cheaper than high-cost microscopes (HCMs), have emerged as a feasible alternative [6,20].

Malaria microscopy requires trained personnel to scan stained blood films through a microscope's limited field of view (FoV), a process that may take over 10 minutes depending on the examiner's experience [1]. To streamline this process, early computer-aided methods relied on classical machine learning and computer vision techniques [12,21], while subsequent approaches adopted deep learning methods [1,7,9,17]. However, these methods generally assume similar imaging conditions during training and testing. In reality, the inferior lenses and smaller FoV of LCMs produce lower-quality images, making annotation more time-consuming and limiting the creation of labeled datasets [6,20]. A practical solution is therefore to train models on annotated HCM images and adapt them to unlabeled LCM images, a concept known as domain adaptation (DA) [6,20]. To address this, Sultani *et al.* [20] introduced paired HCM–LCM images and partially supervised DA for infected-cell detection and parasite life-stage classification. CodaMal [6] extended this setting to end-to-end contrastive learning. Going beyond malaria detection, FARS [10] proposed unsupervised DA with adversarial learning and Fourier-domain style transfer for segmentation.

Although existing HCM→LCM domain adaptation methods have shown promising improvements [6,20], their cross-domain alignment is primarily global. However, malaria detection prediction task that depends on subtle cell-level morphology. Further, recent self-supervised learning methods [3,22] shows that global feature alignment alone does not guarantee more accurate object detection. Although FARS incorporates local spatial modeling, it performs unsupervised DA and does not exploit instance-level HCM–LCM correspondences, limiting its ability to learn cross-microscope cell correspondences unlike partially supervised approaches [6,20]. Past methods [6,10,20] do not explicitly decouple foreground cellular regions from background smear, allowing background-dominated features to influence adaptation. Additionally, the contrastive global alignment objective in CodaMal relies on explicit negative pairs, which increase memory and computational demands [2,4]. In contrast, recent approaches learn invariant features more naturally by avoiding pairwise negatives through information maximization [2,3,26]. Furthermore, the generic augmentations used by previous methods [6] (*e.g.*, random cropping, scaling, mix-up), do not capture LCM-specific acquisition artefacts. Prior DA research [13,15] show that realistic, domain-specific augmentations can improve performance by promoting domain-invariant representations.

To address the HCM→LCM domain shift, we introduce **MaLDAM**, a **Ma**sked **L**ocalized **D**omain **A**daptation framework for LCM **M**alaria detection. MaLDAM is a partially supervised DA framework that leverages HCM–LCM image

correspondence. This enables the model to learn not only *where cells are* located but also *which cells correspond across microscope domains*. Our framework jointly optimizes object detection and DA. It consists of an offline segmentation step to isolate foreground, cell-containing image patches in both HCM and LCM images. The resulting masks are used only during training, incurring no additional inference-time latency. For DA, we introduce a novel DA loss ($\mathcal{L}_{\text{DA}}$) that jointly aligns global image context and local cellular representations. The global feature alignment step aligns image-level semantics across domains. Inspired by how malaria diagnosis fundamentally relies on examining cell-level morphological features, we introduce a masked local feature alignment step using the precomputed foreground cellular masks to enforce cell-aware learning and suppress irrelevant background variations. Moving beyond contrastive learning, we adopt VICReg-based (Sec. 2.4) feature alignment to learn smooth and natural invariances without requiring negative sampling. To further promote domain-invariant learning, we apply physically motivated photometric degradations as augmentations to HCM images (*e.g.*, aggressive blurring, staining shifts, and Gaussian noise), simulating LCM-specific imaging artefacts. For object detection, we use the standard YOLO [18] detection loss ($\mathcal{L}_{\text{DET}}$). Following established testing protocols [20], we evaluate our method on the large-scale M5-dataset and achieve a new state-of-the-art (SOTA) performance in infected-cell localization and parasite life-stage classification on unseen LCM images.

The contributions of our paper are as follows:

1. We propose MaLDAM, a partially supervised end-to-end DA learning framework that introduces a novel masked local feature alignment strategy. To the best of our knowledge, it is the first dedicated effort for cell-aware domain adaptation in malaria microscopy.
2. As part of the framework, we incorporate domain-altering photometric augmentations (blur, noise, and staining) to shift HCM images toward the LCM domain and expose the model to LCM-specific appearance variations.
3. Our approach achieves significant improvements of **8.5** on 1000× magnification FOV, **7.3** on 400× FOV and **3** on 100× FOV in terms of mean average precision (mAP), without incurring any additional inference-time latency compared to the previous SOTA [6].

2 Method

2.1 Problem Formulation and Data Preparation

We introduce MaLDAM to address the microscope induced domain shift in blood smear analysis. Given a paired training dataset $\mathcal{D}_{\text{train}} = \{(X_h^i, X_l^i, Y_h^i)\}_{i=1}^N$, where $X_h^i, X_l^i \in \mathbb{R}^{H\times W\times 3}$ are HCM and LCM images capturing the same FOV of a blood smear. Minor spatial misalignments among corresponding image pairs exist due to hardware limitations (details in [20]). Ground-truth annotations Y_h^i provide parasite bounding boxes and life-stage labels (Plasmodium vivax: ring, trophozoite, schizont, gametocyte) for HCM images. Leveraging these labels, our framework learns a detection model invariant to HCM–LCM distributional

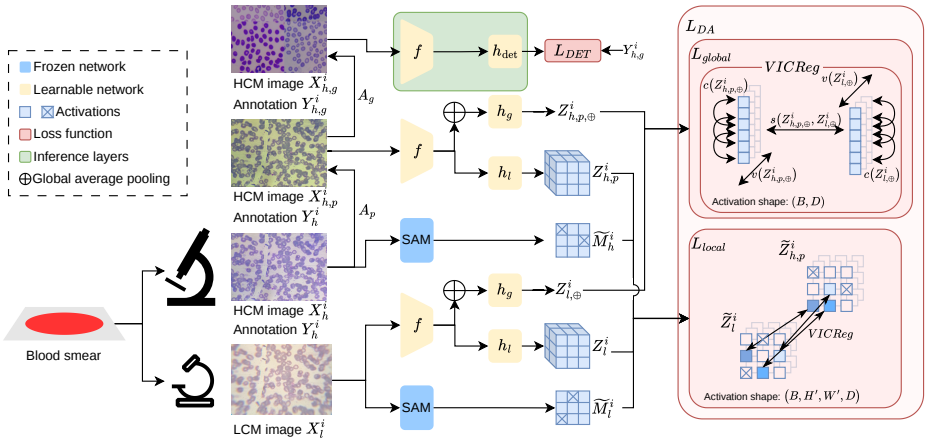


Fig. 1: **Architecture diagram of MaLDAM.** A shared backbone (f) extracts features from paired HCM and LCM images. Geometrically augmented HCM views ($X^i_{h,g}$) supervise the object detection head (h_{det}). Photometrically degraded HCM ($X^i_{h,p}$) and LCM (X^i_l) representations are aligned via the non-contrastive VICReg [2] objective ($\mathcal{L}_{\text{DA}}$), where invariance $s(\cdot)$, variance $v(\cdot)$, and covariance $c(\cdot)$ terms ensure domain alignment without requiring negative samples. This is applied to both global alignment ($\mathcal{L}_{\text{global}}$) and masked local alignment ($\mathcal{L}_{\text{local}}$). The latter explicitly matches cell-centric dense features filtered by masks ($\tilde{M}^i_h, \tilde{M}^i_l$). At inference, only f and h_{det} are used.

shifts, enabling reliable parasite detection on unseen LCM images $\mathcal{D}_{\text{test}}$ during inference.

To isolate cell-containing patches from background, we precompute binary segmentation masks $M^i_h, M^i_l \in 0, 1^{H \times W}$ using Cellpose-SAM [16] as an offline preprocessing step for both modalities, where 1 indicates the presence of a cell. Cellpose-SAM is preferred over classical thresholding, which is sensitive to imaging artefacts and non-uniform backgrounds in LCM images.

2.2 Feature Extraction

To balance DA and object detection, two complementary views are constructed from each HCM image X^i_h: a photometrically transformed view $X^i_{h,\text{p}}$ for DA feature alignment, and a geometrically augmented view $X^i_{h,\text{g}}$ for object detection. $X^i_{h,\text{p}}$ is generated via asymmetric photometric degradations A_p (*e.g.*, blurring, Gaussian noise, targeted stain shifts) to induce non-geometric perturbations that simulate LCM characteristics. $X^i_{h,\text{p}}$ is then subjected to object detection-specific geometric augmentations A_g (*e.g.*, random cropping, scaling, mix-up, mosaic) to obtain $X^i_{h,\text{g}}$ and is applied alongside corresponding ground-truth boxes to obtain $Y^i_{h,\text{g}}$. All three images are processed by a shared YOLOv5 backbone [8] denoted as $f(\cdot)$:

$$F^i_{h,p} = f(X^i_{h,p}), \quad F^i_l = f(X^i_l), \quad F^i_{h,g} = f(X^i_{h,g}). \tag{1}$$

2.3 Detection Objectives

The geometrically augmented HCM image features, $F^i_{h,g}$, are routed to a task-specific detection head, denoted as $h_{\text{det}}(\cdot)$. This head predicts bounding box coordinates, class labels, and objectness scores $\hat{Y}^i_h = h_{\text{det}}\left(F^i_{h,g}\right)$. The predictions are supervised using the corresponding ground-truth annotations $Y^i_{h,\text{g}}$. Following [6], we employ the standard YOLO [18] objective:

$$\mathcal{L}_{\text{DET}} = \mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{loc}} + \mathcal{L}_{\text{obj}}. \tag{2}$$

Here, $\mathcal{L}_{\text{cls}}$, $\mathcal{L}_{\text{loc}}$, and $\mathcal{L}_{\text{obj}}$ denote the classification loss, localization loss, and objectness loss, respectively, computed between predictions $\hat{Y}^i_h$ and ground truth $Y^i_{h,\text{g}}$. By retaining the original detection architecture and optimization pipeline established in [6], we isolate the contribution of our proposed cross-domain modules, ensuring a fair comparison with the baseline.

2.4 Domain Adaptive Feature Alignment

VICReg Preliminaries To align representations across domains without relying on large batch sizes or negative sample mining, we adopt the Variance-Invariance-Covariance Regularization (VICReg) framework [2]. VICReg is a non-contrastive, information-maximization approach. Given two sets of embeddings Z^P and Z^Q , the VICReg criterion is defined as,

$$\mathcal{L}_{\text{VIC}}(Z^P, Z^Q) = \lambda s(Z^P, Z^Q) + \mu\big[v(Z^P) + v(Z^Q)\big] + \nu\big[c(Z^P) + c(Z^Q)\big] \tag{3}$$

where $s(\cdot)$, $v(\cdot)$, and $c(\cdot)$ denote the invariance (mean squared error), variance (hinge loss), and covariance (feature decorrelation) regularization terms, scaled by weights λ, μ, and ν following the standard VICReg formulation [2].

To achieve comprehensive domain invariance, our alignment strategy is decomposed into a global context criterion and a novel masked local criterion. The photometrically transformed HCM image features $F^i_{h,p}$ and their paired LCM counterparts F^i_l are used for domain-adaptive feature alignment.

Global Feature Alignment To align high-level semantic context, we apply global average (GAP) to the feature maps $F^i_{h,p}$ and F^i_l, yielding vectors $F^i_{h,p,\oplus}, F^i_{l,\oplus} \in \mathbb{R}^C$. These are then mapped through a global non-linear projection head $h_g(\cdot)$ into a latent space, producing $Z^i_{h,p,\oplus}, Z^i_{l,\oplus} \in \mathbb{R}^D$. The global alignment loss $\mathcal{L}_{\text{global}}$, is defined using the standard VICReg objective as $\mathcal{L}_{\text{VIC}}(Z^i_{h,p,\oplus}, Z^i_{l,\oplus})$ applied to these global embeddings.

Masked Local Feature Alignment Global alignment cannot capture subtle parasite morphology, especially given the mechanical misalignments between LCM and HCM. Moreover, directly aligning dense maps risks matching irrelevant background; to address this, we introduce a Masked Local Matching strategy. Unpooled feature maps $F^i_{h,p}$ and F^i_l are projected via a local projection network $(h_l(\cdot))$ to dense embeddings $Z^i_{h,p}, Z^i_l \in \mathbb{R}^{D\times H'\times W'}$. Concurrently, segmentation masks are mapped to a $H' \times W'$ spatial grid, computing the foreground proportion in each patch. Patches exceeding a threshold $T = 0.2$ are marked as valid

cell-centric regions, forming binary masks $\tilde{M}_h^i, \tilde{M}_l^i \in \{0,1\}^{H' \times W'}$. Where, 0 and 1 indicate foreground and background, respectively. Here, $\tilde{M}_h^i$ is refined using HCM bounding box labels Y_h^i to guarantee no infected cell is discarded. Applying these masks to dense embeddings isolates valid cell-centric features $\tilde{Z}_{h,p}^i$ and $\tilde{Z}_l^i$, filtering out background noise. To account for spatial misalignments between modalities, we perform feature-based matching. For each vector $z_h^p \in \tilde{Z}_{h,p}^i$, we find its nearest neighbor in $\tilde{Z}_l^i$ using ℓ_2 distance, retain the top-k most confident patches ($k = 20$), and compute the local loss $\mathcal{L}_{\text{local}}$ via the VICReg criterion.

The DA objective balances global and local losses with factor α, while the overall loss combines detection and DA terms weighted by λ_{DA}:

$$\mathcal{L}_{\text{DA}} = \alpha \mathcal{L}_{\text{global}} + (1 - \alpha)\mathcal{L}_{\text{local}} \tag{4}$$

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{DET}} + \lambda_{\text{DA}}\, \mathcal{L}_{\text{DA}} \tag{5}$$

3 Experiments

3.1 Dataset and evaluation metrics

We conduct experiments on the multi-microscope multi-magnification malaria (M5) dataset [20], which comprises paired HCM and LCM blood smear images acquired at magnifications of 1000×, 400×, and 100×. The dataset contains 7,542 images with 20,331 annotated *Plasmodium vivax* parasites spanning four developmental stages: ring, trophozoite, schizont, and gametocyte. We use the HCM training labels and deliberately discard the LCM training labels to simulate the DA problem. The cross-microscope performance is then evaluated on LCM. Following most related work [6,20], detection performance is primarily measured using mean Average Precision at an IoU threshold of 0.5 (mAP@0.5). We additionally report precision and recall to provide a more comprehensive evaluation of detection quality (Sec. 3.3).

3.2 Implementation details

Our framework is implemented in PyTorch and trained on two NVIDIA RTX 4090 GPUs, each with 24 GB memory. We initialize the backbone $f(\cdot)$ and $h_{det}(\cdot)$ using MS-COCO [11] pretrained YOLOv5 [8] weights. The projection networks $h_g(\cdot)$ and $h_l(\cdot)$, are both 2-layer MLPs which are randomly initialized. $T = 0.2$ is used for foreground patch selection and $k = 20$ for the number of patches used for local feature matching. λ, μ, and ν is set to 25, 25, 1 following [2]. The DA objective combines global and local losses with $\alpha = 0.8$, and the overall loss combines detection and DA terms with $\lambda_{\text{DA}} = 0.05$. Other hyper-parameters including batch size $B = 32$ and learning rate of 0.01 are kept unchanged from CodaMal [6] for fair comparison.

3.3 Results

Following the experimental protocol of [20], we evaluate inter-microscopic generalization across multiple field-of-view (FOV) settings (100×, 400×, and 1000×). As summarized in Table 1, our method achieves consistent improvements over

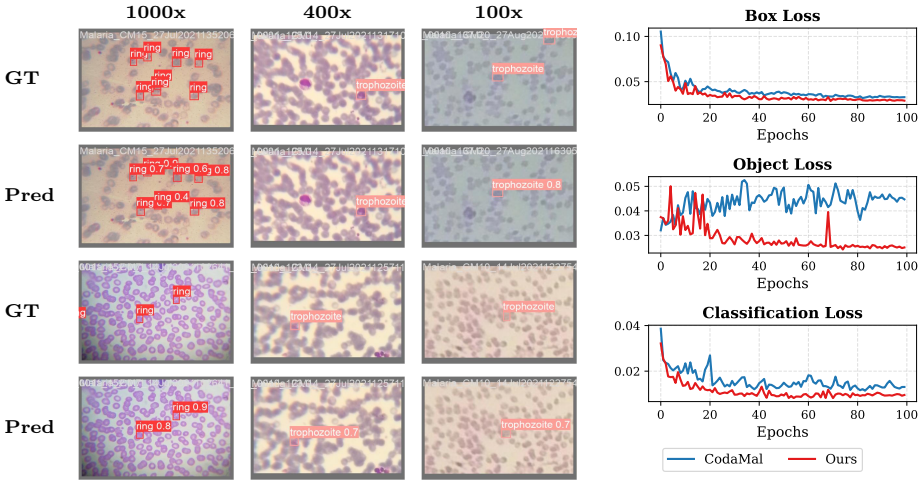
Table 1: **Comparison with previous methods on the M5-dataset [20].** 'Params' denotes trainable parameters (millions), and 'Inf.' denotes inference time (ms) on the same hardware as CodaMal. HCM→LCM denotes models trained on HCM labels and tested on LCM images at magnifications 1000×, 400×, and 100× (mAP@0.5 IoU). * denotes deviations from reported values after local reproduction. Best values are bold; second-best are underlined. † values indicate fully supervised LCM→LCM performance, *i.e.*, training with LCM labels and testing on LCM, representing an upper-bound for YOLOv5 backbone performance. Both MaLDAM and CodaMal use the same YOLOv5 backbone in a partially supervised DA setting. MaLDAM achieves higher mAP, approaching or even surpassing the supervised upper-bound, while maintaining the same inference speed.

Method	**Params**	**Inf.**	**HCM→LCM**		
	(M)	**(ms)**	**1000x**	**400x**	**100x**
YOLOv5 [8] (supervised LCM→LCM)	20.9	8.9	62.6†	61.9†	30.1†
Chen *et al.* [5] CVPR'18	43.7	184	17.6	21.5	-
Saito *et al.* [19] CVPR'19	43.7	184	24.8	21.4	-
Xu *et al.* [25] CVPR'20	43.7	184	15.5	21.6	-
Faster R-CNN [20]	43.7	184	17.1	26.7	-
Sultani *et al.* [20] CVPR'22	43.7	184	37.5	33.8	-
FARS+ [10] CBM'24	49.5	17.8	43.71	45.0	-
CodaMal [6] ICIP'24	**21.2**	**8.9**	53.6	48.7	-
CodaMal [6] ICIP'24 (reproduced)	**21.2**	**8.9**	51.8*	48.3*	28.3*
MaLDAM (Ours)	21.4	**8.9**	**62.1**	**56.0**	**31.3**

existing approaches, surpassing the previous SOTA [6] by 8.5 and 7.3 mAP@0.5 on the 1000×, 400× settings respectively, and by 3 over our reproduced CodaMal baseline on the 100× setting. In addition to the mAP gains, our method achieves precision/recall of (0.80/0.49), (0.80/0.49), and (0.35/0.38) across the 1000×, 400×, and 100× settings, respectively, compared to CodaMal's (0.58/0.47), (0.83/0.41), and (0.58/0.24). Our method consistently improves recall across all FOV settings while maintaining competitive precision. Qualitative examples are presented in Fig. 2a, with additional high-resolution visualizations provided in the supplementary material. Fig. 2b shows that our domain alignment objectives yield lower and more stable detection losses. At 100× MaLDAM surpasses supervised LCM→LCM, where the poor quality of LCM images makes HCM→LCM DA more effective than direct LCM training.

3.4 Ablation study

Table 2 shows that replacing the contrastive global alignment in [6] with global VICReg alignment yields a 2.5 mAP@0.5 improvement. The domain-specific augmentation further improves contrastive and VICReg alignments by 6.1 and 5.6 respectively, highlighting the importance of domain-invariant feature learning under microscope-induced appearance variations. Building upon this configura-



(a) Qualitative results at different LCM magnifications.

(b) Detection loss curves at LCM 1000× magnification.

Fig. 2: **Qualitative results and detection loss curves.** The qualitative results align with MaLDAM's reduced losses. The stable convergence in MaLDAM's objectness loss can be attribtued to VICReg-based DA, which encourages informative and decorrelated features without relying on negatives.

Table 2: **Ablation study on the M5-dataset [20] at 1000× magnification.** HCM→LCM denotes training on HCM labels with geometric augmentations and testing on LCM; if only this is ticked, it uses the YOLOv5 [8] backbone without DA modules. 'Con. Align.' indicates contrastive global alignment, 'VICReg Align.' denotes VICReg based global alignment, 'Domain specific Augmentations' refers to the proposed LCM tailored photometric degradations, 'Local Align.' is VICReg-based local feature alignment, and 'Mask Guidance' restricts local alignment to foreground cell regions.

HCM→LCM	Con. Align.	VICReg Align.	Domain-specific Augmentations	Local Align.	Masked Guidance	LCM mAP50
✓	✗	✗	✗	✗	✗	49.7
✓	✓	✗	✗	✗	✗	51.8
✓	✗	✓	✗	✗	✗	54.3
✓	✓	✗	✓	✗	✗	57.9
✓	✗	✓	✓	✗	✗	59.9
✓	✗	✓	✓	✓	✗	60.1
✓	✗	✓	✓	✓	✓	62.1

tion, incorporating local alignment provides a modest 0.2 gain, suggesting that dense matching remains susceptible to background noise. Applying mask guidance increases performance by 2, showing that the true benefit of local alignment arises when focused on foreground cellular regions. The ablation therefore supports **masked local alignment as a combined mechanism**.

For local matching, (k=10,20,40,100) yields (59.3,62.1,60.4,60.3) mAP@0.5, respectively, supporting (k=20).

4 Conclusion

We introduced MaLDAM, which combines global domain alignment with masked cell-aware local alignment for HCM→LCM malaria detection. Experiments on M5 [20] demonstrate improved detection across three magnifications, with ablations showing that foreground masking is critical for making local alignment effective. The present evaluation is limited to M5, *P. vivax*, and a YOLOv5 detector, and LCM segmentation errors may affect the local alignment objective during training. Future work can explore robustness to segmentation quality and generalizability across datasets, parasite species, and detector architectures.

Acknowledgments. The authors gratefully acknowledge Pekoe Pte. Ltd. for generously providing the computational resources that supported this work.

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

References

1. Arshad, Q.A., Ali, M., Hassan, S., Chen, C., Imran, A., Rasul, G., Sultani, W.: A dataset and benchmark for malaria life-cycle classification in thin blood smear images. Neural Computing and Applications 34, 4473–4485 (2022).
2. Bardes, A., Ponce, J., LeCun, Y.: VICReg: Variance-invariance-covariance regularization for self-supervised learning. In: International Conference on Learning Representations (2022).
3. Bardes, A., Ponce, J., LeCun, Y.: VICRegL: Self-supervised learning of local visual features. In: Advances in Neural Information Processing Systems, vol. 35, pp. 8799–8810 (2022).
4. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: Proceedings of the 37th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 119, pp. 1597–1607. PMLR (2020).
5. Chen, Y., Li, W., Sakaridis, C., Dai, D., Van Gool, L.: Domain adaptive Faster R-CNN for object detection in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3339–3348 (2018)
6. Dave, I.R., de Blegiers, T., Chen, C., Shah, M.: CodaMal: Contrastive domain adaptation for malaria detection in low-cost microscopes. In: 2024 IEEE International Conference on Image Processing (ICIP), pp. 3848–3853. IEEE (2024).
7. Davidson, M.S., Andradi-Brown, C., Yahiya, S., et al.: Automated detection and staging of malaria parasites from cytological smears using convolutional neural networks. Biological Imaging 1, e2 (2021).
8. Jocher, G., Stoken, A., Borovec, J., NanoCode012, Zhang, Y., Chaurasia, A., Kwon, Y., Verdenius, S.: YOLOv5: State-of-the-art object detection. GitHub repository, https://github.com/ultralytics/yolov5 (2020)
9. 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, pp. 56–61. IEEE (2017).

10. Ilyas, T., Ahmad, K., Arsa, D.M.S., Jeong, Y.C., Kim, H.: Enhancing medical image analysis with unsupervised domain adaptation approach across microscopes and magnifications. Computers in Biology and Medicine **170**, 108055 (2024).
11. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common objects in context. In: European Conference on Computer Vision, pp. 740–755. Springer (2014)
12. Linder, N., Turkki, R., Walliander, M., Mårtensson, A., Diwan, V., Rahtu, E., Pietikäinen, M., Lundin, M., Lundin, J.: A malaria diagnostic tool based on computer vision screening and visualization of *Plasmodium falciparum* candidate areas in digitized blood smears. PLoS ONE 9(8), e104855 (2014).
13. Liu, X., Yoo, C., Xing, F., Oh, H., El Fakhri, G., Kang, J.-W., Woo, J.: Deep unsupervised domain adaptation: A review of recent advances and perspectives. APSIPA Transactions on Signal and Information Processing **12**(1), e25 (2022).
14. Ljosa, V., Sokolnicki, K.L., Carpenter, A.E.: Annotated high-throughput microscopy image sets for validation. Nature Methods. 9, 637–637 (2012).
15. Orbes-Arteaga, M., Varsavsky, T., Sørensen, L., Nielsen, M., Pai, A., Ourselin, S., Modat, M., Cardoso, M.J.: Augmentation based unsupervised domain adaptation. arXiv preprint arXiv:2202.11486 (2022).
16. Pachitariu, M., Rariden, M., Stringer, C.: Cellpose-SAM: Superhuman generalization for cellular segmentation. bioRxiv (2025).
17. Rajaraman, S., Antani, S.K., Poostchi, M., Silamut, K., Hossain, M.A., Maude, R.J., Jaeger, S., Thoma, G.: Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. PeerJ 6, e4568 (2018).
18. Redmon, J., Divvala, S., Girshick, R., Farhadi A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 779-–788 (2016).
19. Saito, K., Ushiku, Y., Harada, T., Saenko, K.: Strong-weak distribution alignment for adaptive object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6956–6965 (2019)
20. Sultani, W., Nawaz, W., Javed, S., Danish, M.S., Saadia, A., Ali, M.: Towards low-cost and efficient malaria detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20655–20664 (2022).
21. Tek, F.B., Dempster, A.G., Kale, I.: Malaria parasite detection in peripheral blood images. In: Proceedings of the British Machine Vision Conference, pp. 347–356 (2006).
22. Wang, X., Zhang, R., Shen, C., Kong, T., Li, L.: Dense contrastive learning for self-supervised visual pre-training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3024–3033 (2021).
23. World Health Organization: Malaria diagnosis: Microscopy. https://www.who.int/teams/global-malaria-programme/case-management/diagnosis/microscopy, last accessed 2026/06/24
24. World Health Organization: World malaria report 2025: addressing the threat of antimalarial drug resistance. World Health Organization, Geneva (2025). https://www.who.int/publications/i/item/9789240117822, last accessed 2026/06/24
25. Xu, M., Wang, H., Ni, B., Tian, Q., Zhang, W.: Cross-domain detection via graph-induced prototype alignment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12355–12364 (2020)
26. Zbontar, J., Jing, L., Misra, I., LeCun, Y., Deny, S.: Barlow Twins: Self-supervised learning via redundancy reduction. In: International Conference on Machine Learning, pp. 12310–12320. PMLR (2021)