Computers in Biology and Medicine 170 (2024) 108055

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Enhancing medical image analysis with unsupervised domain adaptation approach across microscopes and magnifications

Talha Ilyas [a,b], Khubaib Ahmad [a,b], Dewa Made Sri Arsa [a,b,c], Yong Chae Jeong [b,d], Hyongsuk Kim [a,b,*]

[a] Division of Electronics and Information Engineering, Jeonbuk National University, Jeonju, 54896, Republic of Korea
[b] Core Research Institute of Intelligent Robots, Jeonbuk National University, Jeonju, 54896, Republic of Korea
[c] Department of Information Technology, Universitas Udayana, Bali, 80361, Indonesia
[d] Division of Electronics Engineering, Jeonbuk National University, Jeonju, 54896, Republic of Korea

ARTICLE INFO

Keywords:
Medical imaging
Segmentation
Convolutional neural networks
Malaria
Microscopy
Varying magnifications

ABSTRACT

In the domain of medical image analysis, deep learning models are heralding a revolution, especially in detecting complex and nuanced features characteristic of diseases like tumors and cancers. However, the robustness and adaptability of these models across varied imaging conditions and magnifications remain a formidable challenge. This paper introduces the Fourier Adaptive Recognition System (FARS), a pioneering model primarily engineered to address adaptability in malarial parasite recognition. Yet, the foundational principles guiding FARS lend themselves seamlessly to broader applications, including tumor and cancer diagnostics. FARS capitalizes on the untapped potential of transitioning from bounding box labels to richer semantic segmentation labels, enabling a more refined examination of microscopy slides. With the integration of adversarial training and the Color Domain Aware Fourier Domain Adaptation (F2DA), the model ensures consistent feature extraction across diverse microscopy configurations. The further inclusion of category-dependent context attention amplifies FARS's cross-domain versatility. Evidenced by a substantial elevation in cross-magnification performance from 31.3% mAP to 55.19% mAP and a 15.68% boost in cross-domain adaptability, FARS positions itself as a significant advancement in malarial parasite recognition. Furthermore, the core methodologies of FARS can serve as a blueprint for enhancing precision in other realms of medical image analysis, especially in the complex terrains of tumor and cancer imaging. The code is available at; https://github.com/Mr-TalhaIlyas/FARS.

1. Introduction

Malaria, a life-threatening disease triggered by the Plasmodium parasite, continues to be a global health burden, predominantly affecting the resource-limited regions of Africa, Asia, and Latin America. Anopheles mosquitoes serve as vectors, facilitating the transmission of the parasite to humans through a single bite. Symptoms of malaria span from fever, nausea, and abdominal pain to more severe manifestations such as organ failure and seizures, potentially leading to death. According to the World Malaria Report 2021, there was a notable increase in malaria cases in 2020, with 247 million clinical cases and 619,000 deaths, indicating an exacerbation of the challenge faced by health systems globally [1]. The majority of these deaths occur in Africa, with children under five being the most vulnerable demographic. Despite making up only 1.7% of malaria cases worldwide, India also experiences a significant toll, especially among children [2].

The Anopheles mosquito carries four species of Plasmodium: P.vivax, P.malariae, P.falciparum, and P.ovlae, with P.vivax being most prevalent in Asia and South America [2,3]. The parasite undergoes four stages of infection in its vertebrate hosts, which can be detected via microscopic examination of peripherial blood smears (PBS) [4,5]. These stages include the ring, schizont, trophozoite, and gametocyte stages, each presenting unique characteristics within red blood cells (RBCs). [2]. Blood smear examinations utilize thick and thin films, offering varying degrees of sensitivity and specificity in parasite detection. While this method is the diagnostic gold standard, its practical application in low-resource settings is hindered by the need for sophisticated microscopes and skilled technicians [2,6] . This highlights the necessity for computerized systems, where artificial intelligence (AI) has shown great potential in medical diagnostics.

* Corresponding author.
E-mail addresses: talha@jbnu.ac.kr (T. Ilyas), hskim@jbnu.ac.kr (H. Kim).

https://doi.org/10.1016/j.compbiomed.2024.108055
Received 6 October 2023; Received in revised form 5 January 2024; Accepted 26 January 2024
Available online 29 January 2024

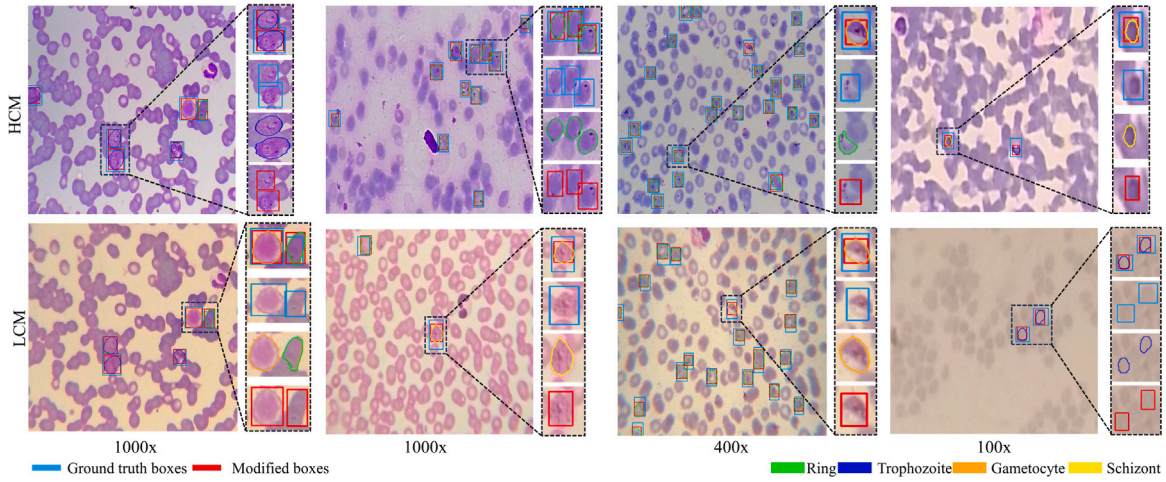

Fig. 1. Sample Images from the M5 Dataset: The images showcase slides captured under both HCM and LCM microscopes at different magnifications. Detailed views on the right side of each image compare various annotation methods. It becomes evident that traditional bounding boxes often fail to tightly surround the target object, resulting in frequent overlaps with adjacent nuclei. However, pixel-level semantic segmentation labels offer a more precise location and distinct separation of each nucleus, including those instances where they overlap or touch. The figure also introduces modified bounding boxes, which are derived from the auto-generated segmentation labels. Notably, these modified boxes exhibit a tighter fit around the objects, thereby reducing unnecessary background inclusions.

AI-driven computer-aided diagnostics have made significant strides in the detection and characterization of malaria infections, assisting in clinical decision-making processes [7–11]. However, these models often falter when faced with a domain shift between the training and testing datasets. For example, a model trained on high-magnification images from high-cost microscopes (HCMs) may struggle when applied to images from low-cost microscopes (LCMs), even if the magnification level is similar. This difficulty arises because LCM images often have a limited field of view and differ markedly in appearance, clarity, and color distributions, as illustrated in Fig. 1. To address these disparities, domain adaptation algorithms have become essential. These algorithms facilitate the transfer of knowledge from one domain (such as HCM images) to a related but distinct target domain (like LCM images). This approach, particularly unsupervised domain adaptation (UDA), is gaining traction as it does not require annotated target data (LCM images) for model training, thereby enhancing model adaptability in diverse medical imaging scenarios.

Our research addresses two critical domain adaptation challenges in malaria diagnosis using microscopic image analysis: (1) transitioning from high-cost microscope (HCM) images to low-cost microscope (LCM) images and (2) adapting from higher to lower magnification slides. These adaptations are crucial due to the predominant use of LCMs and lower magnification in resource-limited settings where malaria is most prevalent.

To tackle the first challenge, our approach incorporate the color domain aware Fourier domain adaptation (CDAFDA→F2DA) algorithm which significantly enhances the compatibility between HCM and LCM images. This algorithm adeptly transfers the textural details from HCM images to LCM images, effectively bridging the gap in image quality and detail resolution. By doing so, it ensures that diagnostic features prominent in HCM images are retained and recognizable in LCM images, facilitating accurate malaria diagnosis even with less sophisticated equipment.

Addressing the second challenge, our proposed unsupervised domain adaptation (UDA) training strategy effectively transfers the intricate features learned from higher magnification images to enable precise parasite detection in lower magnification slides. Fig. 2 shows HCM captured slides at different magnifications. This is particularly significant, as lower magnification slides are more commonly used in field settings. By enabling this feature transfer, our model overcomes the traditional limitations of lower magnification in identifying detailed parasitic features, thereby enhancing the diagnostic accuracy in diverse field conditions.

Furthermore, in the realm of medical image annotation, labeling the data for machine learning training is a significant challenge, particularly because it requires the involvement of domain experts [12,13]. Most medical image datasets utilize bounding box labels [14–16], which are quicker to generate but offer less precise information about the region of interest [17,18]. Unlike natural images, medical images often feature overlapping entities (nuclei), making bounding box annotations less robust for applications like malaria diagnosis where exact boundary detection of overlapping cells is vital, see Fig. 1. Thus, to advance the capabilities of deep learning algorithms in detecting malaria, it is crucial to shift towards more refined annotation techniques like semantic segmentation, which provide precise pixel-level details. Although this process is more time-consuming and resource-intensive, it potentially facilitates robust feature extraction and diagnosis, significantly improving detection accuracy. Building upon these insights, to expand the available annotated data, we present a simple and practical method for transforming existing datasets with bounding box annotations into datasets with semantic segmentation annotations. This conversion improves precision and robustness, enabling more reliable feature extraction.

In response to these needs, our research presents a comprehensive solution to two of the most pressing challenges in AI-driven malaria diagnostics: adapting from high-end, high-magnification microscopy to more commonly available, lower magnification microscopy, and ensuring effective diagnosis using less sophisticated LCMs. Through these innovations, we aim to enhance the global fight against malaria, particularly in regions most affected by this devastating disease.

2. Related works

The rising interest in medical imaging analysis and advanced computer vision and deep learning models has paved the way for significant strides in malaria parasite diagnosis. In this section, we delve into the various methodologies used in analyzing thin blood smears, highlighting their complexities and contributions to overcoming current challenges. We then explore cutting-edge object detection and segmentation techniques in unsupervised domain adaptation (UDA) for medical imaging. Lastly, we discuss the evolution of segmentation learning from bounding boxes.

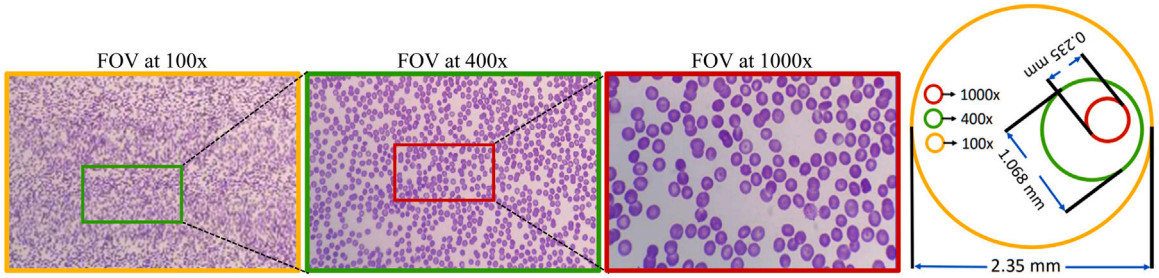

Fig. 2. Comparative visualization of field-of-view in blood smear slides at varying magnifications, illustrating the increasing level of detail and decreased viewing area as magnification increases.

2.1. Malarial parasite recognition

Recent advancements in deep learning, especially Convolutional Neural Networks (CNNs), have significantly outperformed traditional classifiers in malaria diagnosis. Research by Liang et al. [19] and Dong et al. [20] achieved over 95% accuracy in differentiating infected from uninfected cells. Gopakumar et al. applied CNNs to focus stacks, enhancing both sensitivity (97.06%) and specificity (98.50%) in detecting malaria parasites. Angel et al. developed a CNN model that successfully identifies malaria-infected red blood cells (RBCs), boasting an impressive 99.24% accuracy. This model not only distinguishes normal RBCs but also identifies those with other inclusions. Further, Maqsood et al. [10] introduced a computational model leveraging natural language processing for classifying malaria parasite proteins, achieving notable success with a genetic algorithm-based ensemble approach.

Multi-stage classification models have also shown promise. Loh et al. [3] employed Mask R-CNN for rapid and accurate detection of malaria parasites in RBCs. Kristofer et al. [21] combined connected component analysis with a pre-trained Inception V3 model for effective parasite detection and species identification. Courosh et al. [22] created an automated malaria diagnostic system using object detection and CNNs, demonstrating strong field-level detection capabilities. Charles et al. [23] developed a three-stage classifier for cell categorization in thin blood smears, achieving high specificity.

A significant hurdle in malaria diagnosis is the complex life cycle of the parasites, including stages like gametocytes, rings, trophozoites, and schizonts. Variations in imaging equipment and procedures, as well as inconsistencies across different laboratories, add to the complexity of accurate detection. LCMs, while affordable, have limitations like restricted field of view (FOV) and lower image clarity, making the detection and stage confirmation of malaria cells challenging and time-consuming. Consequently, the development of domain-adaptive algorithms that can transition from high-cost to low-cost microscopy (HCM→LCM), and vice versa, is vital for overcoming these imaging discrepancies.

2.2. Unsupervised domain adaptation

In ***object detection based domain adaptation***, pioneering work by Chen et al. [24] with AdaptRCNN introduced two gradient reversal layers (GRL) for aligning features at both image and instance levels. This spurred the development of several two-stage detectors. Saito et al.'s [25] DA-Detection framework highlighted the effectiveness of strong local and weak global feature alignment, while Zhu et al. [26] emphasized instance-level alignment using Region Proposal Network (RPN) proposals. Despite these advances, precise instance-level representations and multi-modal instance information were often overlooked.

To address these gaps, Xu et al. [27] introduced the use of a relational graph for more accurate instance-level feature representation, leading to a graph-induced prototype alignment (GPA). GPA also incorporated contrastive Loss to mitigate the effects of class imbalance in domain adaptation, refining the training process.

However, these methods have limitations. Primarily, they focus on aligning features at a single scale, specifically at the RPN stage of Faster-RCNN. This approach is less effective for complex domain adaptation tasks, such as adapting microscopy slides from different domains (HCM and LCM) and across various magnifications (100x, 400x, and 1000x).

In the realm of adaptive malarial parasite detection, Sultani et al. combined adversarial training with CycleGAN [28] for image-to-image translation, aligning source and target domain images, to which we refer to as M5RCNN for ease of discussion. Their approach further employed ranking and triplet losses to align RPN features of Faster-RCNN across domains, achieving a cross-domain performance of 37.5% mean Average Precision (mAP). However, their method did not account for the alignment of features at varying magnifications, leading to suboptimal performance.

Segmentation-based domain adaptation has also seen significant developments. Hoffman et al.'s [29] CyCADA integrated adversarial adaptation at both pixel and feature levels, using cycle-consistency and semantic losses for structural and semantic consistency. Vu et al.'s [30] ADVENT leveraged entropy loss to penalize low-confidence predictions in the target domain, which was later on proved to be ineffective in regions with low entropy by [31]. Tsai et al. [32] introduced AdaptSegNet, assuming pixel-level predictions as structured outputs for efficient domain adaptation through adversarial learning.

Despite the advancements, GAN-based UDA frameworks often face challenges, including added complexity, increased computational cost, and performance degradation with limited training data [31]. Additionally, adversarial learning can be counterproductive for semantic segmentation due to complex representations and difficulties in stabilizing training [33].

Improving upon CyCADA, Zhou et al. [34] proposed APA2SegNet, adding anatomy content consistency regularization to ensure preservation of object-specific content during unpaired domain adaptation. Xing et al. [35] introduced a differentiable, stochastic data augmentation module to reduce discriminator overfitting, addressing challenges in limited target domain data settings.

In a different approach, some researchers have focused on domain alignment by blending input and output samples from both source and target domains at the pixel level. For example, Tranheden et al. [36] proposed the innovative idea of mixing images and corresponding labels or pseudo-labels from both domains. This creates a set of highly perturbed samples used for training. Expanding on this concept, Mattolin et al.'s [37] ConfMix strategically combines samples based on regional confidence in target pseudo detections, facilitating a smoother learning transition and thereby enhancing target domain detection accuracy.

However, these methods of high-level domain mixing, primarily effective in standard domain transfer tasks like transitioning from HCM to LCM, encounters limitations in more complex scenarios. Specifically,

in our research context, where the adaptation needs to account for varying magnifications across microscopes, these methods show a marked decrease in effectiveness. Our experiments indicate that while domain mixing improves results under the same magnification, performance drops significantly when magnifications change, such as from 1000x to 400x.

To address the unique challenges in UDA for malaria detection, several methods have been proposed. Sen et al. [38] utilized graph convolutional networks for domain-adaptive classification, while Rehamn et al. [39] introduced a modified distribution matching loss for CycleGAN to counter feature hallucination in medical image synthesis. Ramarolahy et al. [40] used GANs to generate synthetic images, augmenting malaria datasets and enhancing cross-domain robustness.

Our work diverges from these methods by resolving two domain adaptation problems within a unified network. We employ a color domain aware Fourier domain adaptation (F2DA) algorithm to address variations in staining procedures between HCM and LCM. By transferring the stain and texture style from HCM to LCM images, and aligning their low-frequency domain spectrum components, we effectively bridge distribution gaps caused by differing staining techniques, eliminating the need of CycleGAN. Additionally, the variations in magnifications are addressed through focused adversarial training on feature alignment. This method not only aligns image styles but also ensures the extraction of critical nuclei-specific features for robust recognition.

2.3. Segmentation learning with bounding boxes

The intricacy of malaria parasite detection, particularly due to the overlapping nature of nuclei in blood smears, calls for a shift from traditional bounding box annotations to more precise pixel-level segmentation labels [17,41]. This transition is crucial for enhancing detection accuracy. Pioneering this shift, Rajchl et al. [42] introduced DeepCut, a method for deriving semantic segmentations from images with bounding box annotations. By employing an iterative energy minimization process within a conditional random field (CRF) and concurrently refining a CNN model's parameters, they achieved substantial improvements in brain and lung segmentation accuracy. Similarly, Ou et al. [43] developed the BBox-Guided Segmentor, a weakly-supervised segmentation pipeline that leverages bounding box input within an adversarial framework, resulting in notable improvements in stroke lesion segmentation.

Our proposed model diverges from these methods by eliminating the need for bounding box annotations during training. Instead, it utilizes auto-generated segmentation labels, streamlining the training process and reducing the reliance on manual annotations. Additionally, during inference, our model operates directly on the input image, forgoing the need for bounding boxes to identify the region of interest. This approach simplifies both training and inference processes, enhancing the model's efficiency and practicality.

The main contributions of our framework are:

- We introduce a color domain aware Fourier domain adaptation (F2DA) algorithm to bridge the gap between high-cost microscopy (HCM) and low-cost microscopy (LCM). This algorithm effectively handles variations in imaging equipment and procedures, enhancing the detection process's robustness.
- By transitioning from bounding box to pixel-level segmentation labels, we enable more accurate boundary detection, crucial for dealing with overlapping nuclei and improving overall robustness.
- Our model uniquely requires only auto-generated segmentation labels during training and relies solely on input images during inference. This approach significantly simplifies the training and inference framework, reducing the annotation burden and increasing efficiency.
- We address domain, life-cycle, and magnification variations simultaneously with a novel encoder–decoder architecture. This integrated approach employs category-dependent context attention and adversarial optimization, offering a robust and reliable solution for cross-domain, multi-stage, and multi-magnification malaria parasite recognition.

In summary, our proposed framework provides an innovative, all-encompassing solution to the complex challenges of malaria parasite detection. It integrates domain adaptation, pixel-level segmentation, and a novel architecture to handle multiple variations effectively.

3. Materials and methods

3.1. Dataset

For our study, we selected the M5 dataset [6], a large-scale multi-microscope multi-magnification malaria dataset. Sample images from M5 dataset are shown in Fig. 1. There are 7543 images in the M5 collection with a total of 20,331 labeled nuclei. Several factors influenced our decision to use this dataset as the foundation for our research.

Firstly, the M5 dataset addresses the specific challenges associated with low-cost microscopes (LCM) commonly found in resource-constrained areas [4,44]. These microscopes have a limited field of view and lower image clarity due to the utilization of low-quality lenses. By including images captured using both LCM and high-cost microscopes (HCM), the M5 dataset allows us to examine the impact of microscope variations (domain shift) on malaria detection accuracy. This is particularly important as more than 70% of the population relies on low-cost microscopes for malaria diagnosis.

Secondly, the M5 dataset provides images captured at multiple magnifications, namely 1000x, 400x, and 100x. Fig. 2 visually compares the FOV of slides captured by a microscope at these magnifications. This feature enables us to investigate the influence of different magnifications on malaria parasite detection. Each magnification scale offers a unique perspective and level of detail, and understanding their impact is essential for developing robust and reliable malaria detection algorithms.

Moreover, the M5 dataset focuses on Plasmodium vivax (P. vivax) malaria, one of the most prevalent and dangerous types of malaria. By analyzing data specifically from P. vivax-infected patients, we ensure that our research addresses the practical and clinical relevance of malaria diagnosis.

Additionally, the M5 dataset includes manual annotations done by expert medical professionals. The annotations are provided in the form of bounding boxes. But unlike objects in natural images, nuclei tend to overlap with each other, shown in Fig. 1. As a result, the bounding box for one instance often covers other nuclei, which negatively impacts the robustness of all bounding box-based detection algorithms. So we convert these bounding box labels to pixel precise segmentation labels using a simple yet real time practical approach that can be applied to any existing dataset, see next section for details. Unlike recent works, we only use these auto-generated segmentation labels for training our network and during inference only original image is provided as an input to generate the final predictions.

By utilizing this dataset, we develop and evaluate our malaria detection framework in a realistic and clinically relevant context.

3.2. Limitations of bounding box annotations and the importance of pixel-precise segmentation

When labeling data with bounding boxes instead of pixel-precise segmentation, there is a potential for human error, leading to bounding boxes that are not tightly around the target object and may include neighboring objects. This can occur due to various reasons, such as the subjective judgment of the annotator, difficulties in accurately

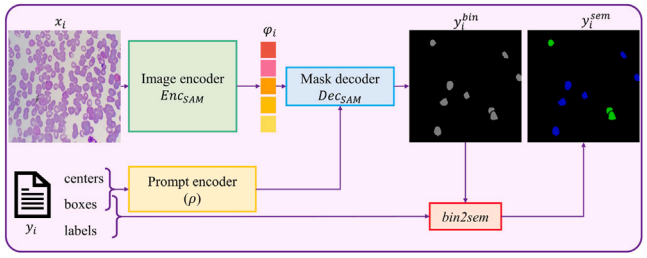

Fig. 3. Overall process of automatic semantic segmentation labels generation.

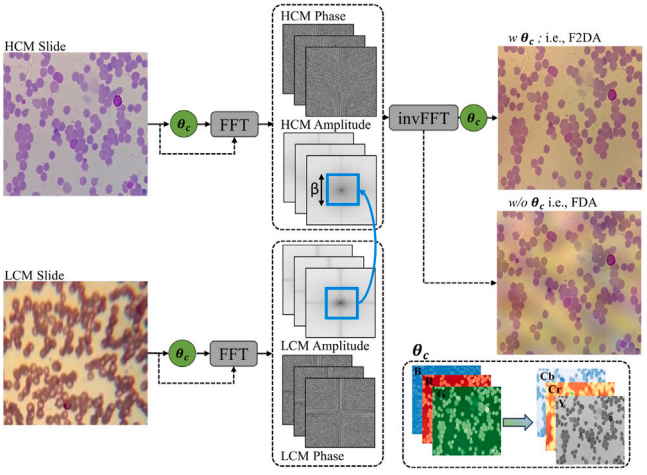

Fig. 4. Overview of F2DA algorithm: Transformation of a HCM image to mimic the LCM 'style' while preserving its semantic content. The algorithm infuses the style from a randomly sampled LCM image/slide by swapping its low-frequency spectral component with that of the HCM image, resulting in an 'LCM-stylized' HCM image. Beta controls the amount of FFT amplitude to be swapped between domains.

identifying the object boundaries, or limitations in the annotation tools. The imprecise nature of bounding box annotations can result in a few consequences. Firstly, it can lead to misrepresentation of the true extent and location of the target nuclei. This can affect downstream tasks that rely on precise nuclei localization or boundary detection, such as instance segmentation or fine-grained analysis. Secondly, the inclusion of neighboring nuclei or smear background within the bounding box introduces ambiguity and may confuse the learning algorithms during training, potentially affecting their ability to generalize accurately.

In comparison, pixel-precise segmentation labels provide a more detailed and accurate representation of nuclei boundaries. They offer pixel-level delineation, enabling precise localization and separation of individual nuclei, even in cases of overlapping or touching instances. This level of annotation facilitates more robust and fine-grained analysis, as the segmentation masks provide explicit information about the exact shape and boundaries of each object. Therefore, transitioning from bounding box labels to pixel-precise segmentation labels is crucial for improving detection accuracy and addressing the limitations associated with bounding box annotations.

3.3. Automatic generation of segmentation labels

Consider two domains of labeled datasets: one from High-Cost Microscopy, denoted $D_S = \left\{(x_{i,m}^S, y_{i,m}^S)\right\}_{i=1,m}^{N,M}$, and the other from LCM, given by $D_T = \left\{(x_{i,m}^T, y_{i,m}^T)\right\}_{i=1,m}^{N,M}$. Here, N represents the total number of available images in each dataset, and M represents the available magnifications ($M \subset \{100x, 400x, 1000x\}$). Both domains have an equal number of labeled images and available magnifications. For brevity, we omit the microscopy and magnification superscripts in subsequent expressions. In this notation, $x_i \in \mathbb{R}^{H\times W\times 3}$ is RGB-image and $y_i = \left\{[(t,l),(b,r)]_j, k_j\right\}_{j=1}^{B}$ corresponds to an XML file containing the top-left and bottom-right corners of jth-nuclei instance, with k_j representing the corresponding class label, here B is the total number of bounding box annotations provided for ith image.

To automate the process of generating segmentation labels, we utilize the Segment Anything Model (SAM) [45]. SAM is a robust binary segmentation model trained on an extensive dataset of 11 million images with 1 billion masks and is gaining popularity in medical image analysis [46–49]. However, since SAM is designed for general purposes, we need to generate appropriate prompts to ensure reliable segmentation of nuclei in our specific dataset, which consists of multiple classes. To overcome this limitation, we perform several post-processing steps on each generated mask, leveraging the corresponding bounding box and class labels provided in the original dataset. The entire process is illustrated in Fig. 3.

First, we extract the bounding box coordinates ($[(t,l),(b,r)]_j$) and class labels (k_j) from the available annotation files. Each bounding box is defined by its top left corner (t,l) and bottom right corner (b,r). From these coordinates, we calculate the center point (c,d) of the object enclosed within the bounding box using Eq. (1).

$$(c_j - d_j) = \left(\frac{t_j - b_j}{2}, \frac{l_j - r_j}{2}\right) \tag{1}$$

Next, we generate a prompt query using the bounding boxes and center points. This prompt query, along with the feature embeddings of the input image x_i, generated by the SAM model's encoder (Enc_{SAM}), is fed into the decoder of SAM (Dec_{SAM}).

$$\varphi_i = Enc_{SAM}\left(x_i\right) \tag{2}$$

$$y_i^{bin} = Dec_{SAM}\left(\varphi_i,\ \rho\left[[(t,l),(b,r)]_j, (c,d)_j\right]_{j=1}^{B}\right) \tag{3}$$

Dec_{SAM} outputs a binary mask (y_i^{bin}) that highlights all the detected nuclei as foreground based on the input prompt, while considering the remaining regions as background. The prompt encoder, denoted as ρ, generates positional encodings [50] that are then input to Dec_{SAM} to guide the generation of the binary mask. Finally, we pass this binary mask and a newly generated prompt derived from the bounding boxes and labels provided in the original annotation files through the *bin2sem* block to generate the final semantic segmentation label $y_i^{seg} \in \mathbb{R}^{H\times W\times K}$.

$$y_i^{seg} = bin2sem\left(y_i^{bin},\ \left\{[(t,l),(b,r)]_j, k_j\right\}_{j=1}^{B}\right) \tag{4}$$

The pseudo code for *bin2sem* block is shown in *Algorithm*1. To ensure the validity and reliability of the automatically generated semantic labels compared to the original bounding box labels, we convert the semantic labels back into bounding box annotations. We then calculate the mean Average Precision (mAP) between the original annotations and the regenerated bounding box annotations. Remarkably, we found the mAP between these two annotations sets to be 99.57%, indicating the high reliability and similarity of the auto-generated segmentation labels to the original annotations provided with the dataset. Both the original annotations and the auto-generated segmentation labels are used for comprehensive evaluation, as discussed in the Results and Discussion section.

3.4. Color domain aware fourier domain adaptation (F2DA)

In addressing the challenges of malaria detection, our research adopts the innovative Color Domain Aware Fourier Domain Adaptation (F2DA) strategy. This approach targets the distributional discrepancies between images obtained from High-Cost Microscopy (HCM) and Low-Cost Microscopy (LCM), as illustrated in Fig. 4. F2DA is an advanced adaptation of Fourier Domain Adaptation (FDA), designed to

Algorithm 1 Pseudo Code for converting binary masks to semantic masks

```
function BIN_2_SEM(binary_mask, coords, det_classes, class_dict)
    s ← INIT_ZEROS(binary_mask.shape)
    labeled, num_features ← PERFORM_CONNECTED_COMPONENT_LABELING(binary_mask)
    for blob in range(1, num_features + 1) do
        xmin, ymin, xmax, ymax ← CALCULATE_BLOB_BOUNDING_BOX(labeled, blob)
        max_iou ← 0
        max_class ← 0
        for box in range(len(coords)) do
            iou ← CALCULATE_IOU(coords[box], (xmin, ymin, xmax, ymax))
            if iou > max_iou then
                max_iou ← iou
                max_class ← GET_CLASS(det_classes[box], class_dict)
            end if
        end for
        s ← ASSIGN_CLASS_TO_BLOB(s, labeled, blob, max_class)
    end for
    return s
end function
```

transfer textural features across domains effectively while minimizing variances.

At the core of F2DA is the Fast Fourier Transform (FFT), applied to both HCM and LCM images. FFT transforms these images into frequency representations, which encompass crucial texture and visual features. For each image, the FDA algorithm calculates both amplitude and phase components. The pivotal step in FDA involves substituting the amplitude spectrum of the source (HCM) image with that of the target (LCM) image while maintaining the source image's phase spectrum.

To enhance the FDA process specifically for malaria detection, we utilize the YCrCb color space instead of the traditional RGB format. Each channel in the YCrCb space is individually subjected to FFT, significantly affecting the extracted frequency components and texture features. The Y channel, emphasizing the green channel, offers enhanced contrast between red blood cells (RBCs) and the background, thereby enriching the frequency components for texture feature analysis. Concurrently, the Cr and Cb chrominance channels provide critical color information, aiding in distinguishing RBCs from malaria parasites.

Mathematically, F2DA operates as follows:

Given an HCM image x^s_{RGB} in the RGB domain, we convert it to the YCrCb domain, x^s_{YCrCb}, using the transformation function θ_c:

$$x^s_{YCrCb} = \theta_c(x^s_{RGB}) \tag{5}$$

This conversion is fundamental for manipulating the frequency components effectively. Applying the Fourier transform $\mathcal{F}$ to x^s_{YCrCb}, we decompose it into its amplitude $\mathcal{F}_A$ and phase $\mathcal{F}_\phi$ components:

$$\mathcal{F}(x^s_{YCrCb})(u,v) = \sum_{h,w} x^s_{YCrCb}(h,w)e^{-j2\pi(hu/W+wv/H)} \tag{6}$$

A mask Z is then defined, with a parameter β governing the extent of amplitude transfer, as shown in Fig. 4:

$$Z(h,w) = 1 \text{ for } (h,w) \in [-\beta W : \beta W, -\beta H : \beta H] \tag{7}$$

FDA is applied between transformed images from $\mathcal{D}_S$ and $\mathcal{D}_T$:

$$\begin{aligned} &x^s_{YCrCb} \rightarrow x^t_{YCrCb} = \\ &\mathcal{F}^{-1}([\mathcal{F}_A(x^s_{YCrCb}) \odot Z_\beta + \mathcal{F}_A(x^t_{YCrCb}) \\ &\qquad \odot(1 - Z_\beta), \mathcal{F}_\phi(x^s_{YCrCb})]) \end{aligned} \tag{8}$$

The adapted image is then reverted back to the RGB domain:

$$x^t_{RGB} = \theta_c^{-1}(x^t_{YCrCb}) \tag{9}$$

F2DA directs the algorithm's attention towards specific nuclei features, regardless of the inconsistencies in image quality or magnification. This adaptation aims to maximize confusion between HCM and LCM captured images, thus minimizing the model's dependency on the texture of the smear background or stain intensity. The method incorporates adversarial learning, where a discriminator is trained to maximize this confusion between the source (HCM) and the target (LCM) representations. As a result, our model focuses on the most pertinent features for malaria detection, providing a robust solution to the challenges of varying image conditions and allowing for more reliable malaria detection across microscopes and magnifications.

In summary, our adoption and implementation of FDA, particularly with the use of the YCrCb color space, i.e., F2DA, represents a strategic approach to overcome domain-specific variabilities in malaria detection. This method ensures the robustness and efficacy of our model by concentrating on specific nuclei features crucial for malaria detection, rather than being sidetracked by variable quality of microscopy images or differing stain intensities. The clear differences in stain transfers between simple FDA and F2DA can be seen in Fig. 4, further illustrating the benefits of our approach.

3.5. Proposed network architecture

In this study, we introduce an encoder–decoder model, denoted by (ϕ), with learnable parameters w specifically designed for malaria parasite detection, as illustrated in Figs. 5 and 6. We will first describe the fundamental components of the encoder and decoder. Subsequently, we outline their integration within the FARS framework and conclude with details on our adversarial training methodology.

One of the primary challenges in automated image analysis for malaria detection is dealing with inhomogeneous intensities and noise in the images. Traditional approaches often resort to various denoising algorithms [51] and normalization techniques [52,53] to enhance performance. However, these methods can sometimes lead to the loss of vital information and an increase in computational demands. Our proposed approach circumvents these pitfalls by facilitating effective malaria recognition without the need for additional denoising and normalization techniques.

3.5.1. Encoder building blocks

To enhance the extraction and acquisition of information-rich features, we draw inspiration from contemporary hybrid neural network architectures that amalgamate the spatial comprehension of Convolutional Neural Networks (CNNs) with the contextual reasoning prowess of Transformers [54]. This fusion empowers these architectures to adeptly apprehend intricate nuances in medical images like microscopic slides, all the while embracing a more expansive perspective. This

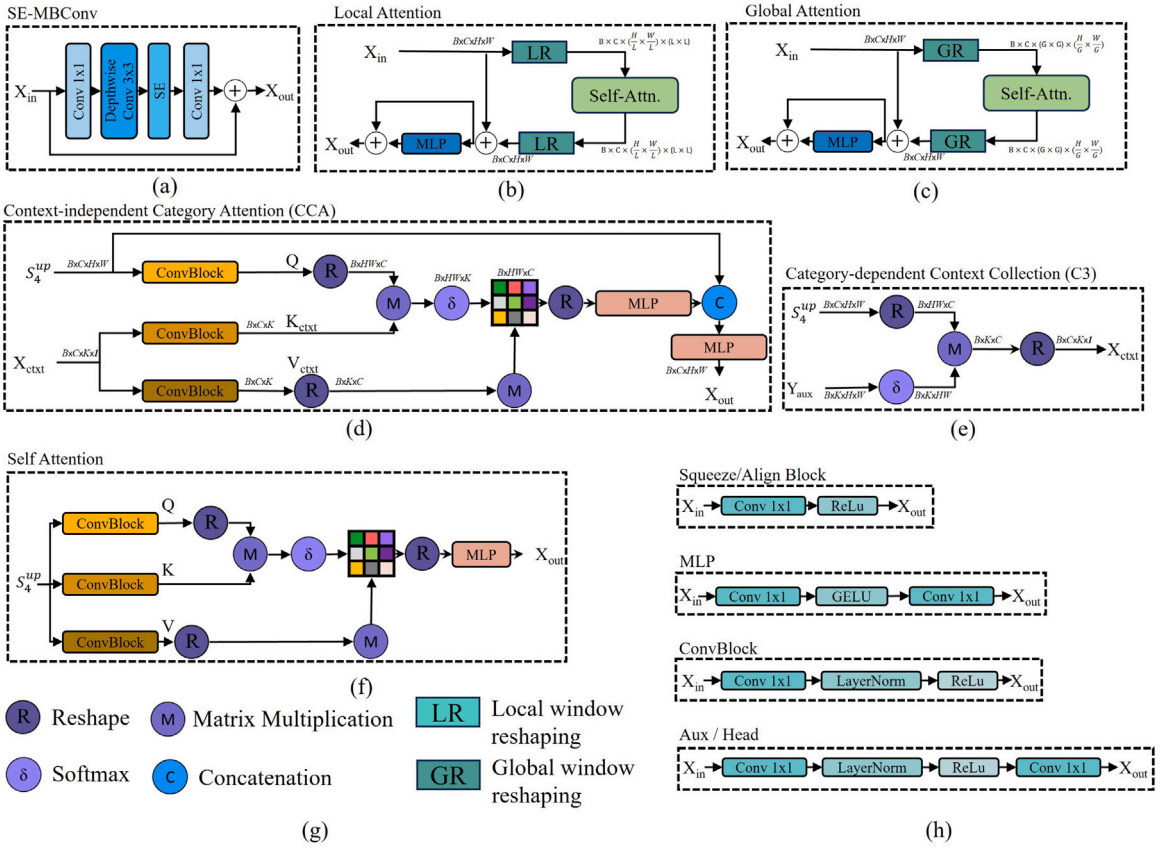

Fig. 5. Schematic representations of building blocks in the proposed framework:. (a) SE-MBConv Block, (b) local attention block, (c) global attention block, (d) context independent category attention block (CCA) (e) category dependent context collection ($C3$) (f) conventional self attention, (g) legends, (f) other sequential blocks. Full network architecture is show in Fig. 6.

holistic approach proves indispensable in precisely delineating structures and identifying abnormalities in microscopic slides, significantly enhancing diagnostic accuracy.

SE-MBConv Blocks: At its core, this block employs inverted residual blocks [55] but substitutes the bottleneck with depth-wise convolutions to cut computational complexity. We further refine each MobileNet-v2 block by applying the squeeze-excitation algorithm [56]. This module forms the foundational stem block of our proposed framework, depicted in Fig. 6. Integrating this with the transformer-styled attention block enhances the network's training stability and generalization capability. The encoder's primary stem block includes two SE-MBConv blocks, with the initial block having a stride of 2. Subsequent stages involve Max-GL blocks (discussed subsequently), with channels in each block organized as $N_{ch} \in \{96, 192, 384, 768\}$. We repeat each block N_s times at each stage S_s, where $N_s \in \{2, 6, 14, 2\}$.

Multi-Axis Global Local Attention (Max-GL): Building on spatially enriched features, we adapt and reconfigure the multi-axis attention blocks as suggested by [57]. These blocks, originally conceptualized for broader classification tasks, are modified here for meticulous segmentation. Our redesigned block incorporates both local and global attention, striking a balance between focused and broad perspectives, thus efficiently processing inputs regardless of their resolution. This architecture's details are furnished in Fig. 5.

Local Attention: For an input feature map $X \in \mathbb{R}^{H \times w \times C}$, we avoid applying attention to the flattened spatial dimension, as seen in previous literature [54,58]. Instead, we first reshape the incoming features into a tensor of shape $(\frac{H}{L} \times \frac{W}{L}, L \times L, C)$, effectively partitioning them into non-overlapping local windows, each of size $L \times L$. Applying self-attention within these local windows is equivalent to attending within a small region. Subsequently, the features are reshaped back into feature maps of shape $(H \times W \times C)$ and processed through an MLP layer before being passed to a global attention block, as illustrated in Fig. 5 (b).

Global Attention: To achieve sparse global attention, we reshape the input feature map $X \in \mathbb{R}^{H \times W \times C}$ into a feature map of shape $(G \times G, \frac{H}{G} \times \frac{W}{G}, C)$ using a fixed $G \times G$ uniform grid. This results in windows with adaptive sizes of $\frac{H}{G} \times \frac{W}{G}$. Applying self-attention to these decomposed uniform grids $(G \times G)$ is equivalent to dilated global attention. This approach allows for sparse global interactions at linear time, enabling our model to capture global information efficiently at all stages.

A sequential combination of these three blocks (SE-MBConv, Global, and Local Attention) results in the multi-axis global-local attention blocks (Max-GL). For fine-grained feature extraction from input microscopic slides, we initially process incoming features using the SE-MBConv block. We then apply global attention to obtain an overall context and structure of the slides, followed by finer detail enhancement at the local level using the local attention block. This sequence of feature extraction blocks is referred to as multi-axis global-local self-attention (Max-GL). In subsequent stages, we stack these Max-GL blocks to extract features at multiple resolutions.

We denote S as the number of stacked Max-GL blocks in the encoder, and S_s as the feature map at the s_th stage. C, H, and W represent the channel number, height, and width of the Max-GL input. In the last three stages of the Max-GL backbone, we extract feature maps and pass them on to the decoder for generating the final predictions.

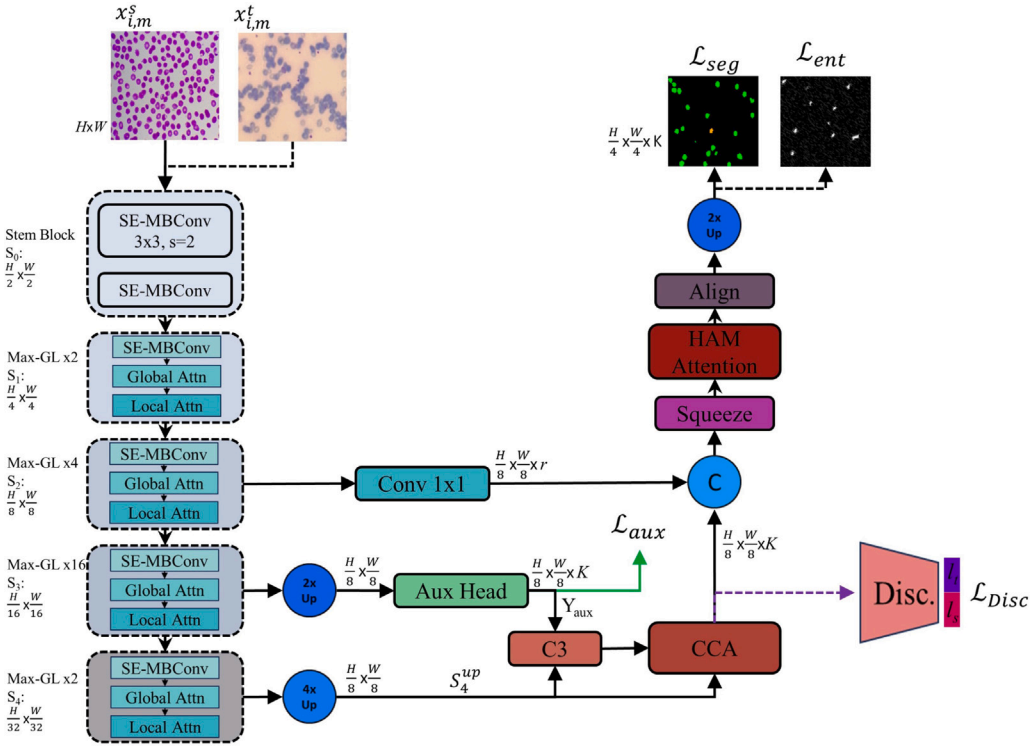

Fig. 6. Architecture of proposed framework. Here, $x^s_{i,m}$ and $x^t_{i,m}$ denote the i_{th} HCM and LCM slide, respectively, each with a specific magnification factor m. During adversarial training, the encoder–decoder and discriminator networks alternate in training, with the discriminators kept frozen while the encoder–decoder is updated, and vice versa. The green arrow indicates the auxiliary loss employed for deep supervision, while the purple arrows illustrate the flow of the forward pass during discriminator training. During the inference stage, the discriminator branch is omitted, and only the encoder–decoder is utilized. For a detailed breakdown of each architectural block, refer to Fig. 5.

3.5.2. Decoder building blocks

Our decoder design is characterized by its simplicity and efficiency, featuring deep supervision through an auxiliary branch for aligning features from two distinct domains, namely HCM and LCM. This deep feature alignment plays a crucial role in stabilizing adversarial training, as explained in the following section.

To enhance and regulate the flow of information from the encoder to the decoder, we propose a modified self-attention module. This module comprises two sub-modules, which we refer to as Category-Dependent Context Collection (C3) and Context-Independent Category Attention (CCA). These sub-modules work together to generate robust features for cross-domain recognition. A visual comparison between the self-attention and modified context attention is provided in Fig. 5 (d) and (f).

Category-Dependent Context Collection (C3): The $C3$ module enriches nuclei instance feature embeddings by considering their interactions and relationships with the background smear and surrounding Red Blood Cells (RBCs). To achieve this, the output features from the last two stages of the backbone network, S_4 and S_3, are upscaled by a factor of 4x and 2x, respectively, to match the spatial dimensions of the features at stage S_2. The upscaling is accomplished using bilinear upsampling ($\hat{B}$), as shown by the following equations:

$$S_4^{up} = \hat{B}\left(S_4,\ scale_factor = 4\right) \tag{10}$$

$$S_3^{up} = \hat{B}\left(S_3,\ scale_factor = 2\right) \tag{11}$$

Soft nuclei predictions (Y_{aux}) are generated by passing the upsampled S_3^{up} features through an auxiliary head.

$$y_{aux} = AuxHead\left(S_3^{up}\right) \tag{12}$$

Here $y_{aux} \in \mathbb{R}^{H/8\times W/8\times K}$, and K is number of unique nuclei categories present in the M5 dataset. The soft object predictions and the S_{up}^4 features are subsequently processed through a class context collection ($C3$) module. This module captures pairwise interactions between nuclei and other RBCs, computing their spatial relationships and appearance similarities to refine the nuclei feature embeddings. The refined feature is denoted as x_{ctxt}. Mathematically the procedure of C3 module can be expressed as in equation below:

$$x_{ctxt} = \mathcal{R}\left(\mathcal{R}\left(S_3^{up}\right) \otimes \delta(y_{aux})\right) \tag{13}$$

Here $\mathcal{R}$ represents the reshape operation, $\otimes$ is the matrix multiplication and δ is SoftMax activation.

Context-Independent Category Attention (CCA): Next the CCA module utilizes the S_4^{up} feature maps to generate queries, while the x_{ctxt} feature maps are used to generate key (K_{ctxt}) and value (V_{ctxt}) feature maps. These feature maps serve as the Q (query), K(key), and V(value) for calculating global contextual information and long-range dependencies in our proposed Context-Independent Category Attention (CCA) module, as illustrated in Fig. 6 (d). The equation for the context attention module can be expressed as follows (Eq. (9)). Now the equation for context attention module can be written as:

$$\widehat{CA}(Q, K_{ctxt}, V_{ctxt}) = \delta\left(\frac{QK_{ctxt}^{trans.}}{\sqrt{d_{k_{ctxt}}}}\right)V_{ctxt} \tag{14}$$

$$\text{CA} = \text{MLP}\left[\text{MLP}\left\{\mathcal{R}\left(\widehat{CA}\right)\right\}\right]' S_4^{up} \tag{15}$$

Here, MLP and © represents multi-layer perceptron and concatenation operation. The final predictions, y_{out}, are generated by combining the output features of the CA module with the feature maps from the third-last stage of the backbone. Before the combination, we reduce the channel dimensions of the S_2 feature maps via 1×1 convolution following [62], generating S_2^r feature maps which are then processed through a HAM attention module [63]. This module models global

Table 1
Comparison with state-of-the-art object detectors. Training and testing are done on HCM slides, evaluation metric used is *mAP*.

Training Magnification	FCOS [62]			RetinaNet [63]			YOLO [64]			Faster R-CNN [65]			FARS		
	Test Magnifications (HCM→ HCM); Metric = mAP														
	1000x	400x	100x	1000x	400x	100x	1000x	400x	100x	1000x	400x	100x	1000x	400x	100x
1000x	36.8	13.5	0	43.1	29.7	0	62.8	36.7	0	66.8	31.3	0	67.13	55.19	5.6
400x	31.4	29.1	1.9	32.9	34.0	1.8	55.2	56.6	4.5	56.9	61.1	1.4	66.1	63.62	13.02
100x	9.4	14.8	8.9	10.2	15.4	16.3	10.5	3.9	20.1	25.4	31.9	31.5	24.67	30.82	33.5

context learning as a low-rank recovery problem with non-negative matrix factorization (NMF2D) as its solution (see [59] for details). To match channel dimensions and improve alignment before and after HAM attention, we apply a 1×1 convolution followed by a *ReLu* activation. The final output feature map is upsampled by a factor of 2 and forwarded to the prediction head to generate the final predictions, $y_{out} \in \mathbb{R}^{H/4 \times W/4 \times K}$.

3.6. Adversarial training

The overall objective of our methodology is to train the model using High Cost Microscopy (HCM) slides such that it produces dependable predictions when applied to Low Cost Microscopy (LCM) slides, even though it was never trained on them. We aim for the network to focus on the textural features of malarial parasites, making it agnostic to the type of microscope used or the stain intensity of the slides. To achieve this, we apply adversarial training to minimize the difference in features produced when HCM or LCM slides are inputted, which ensures feature alignment.

In our adversarial training approach, we leverage the PatchGAN [60] framework, a fully convolutional discriminator (θ), to classify slide features based on whether they originate from HCM or LCM. To align the deep features extracted from both types of microscopies, we utilize two discriminators: one for the decoder features (θ_v) and another for the encoder features (θ_v^{aux}), each with its respective learnable parameters (v).

For HCM slides, our segmentation network (φ_w) predicts a K-dimensional soft segmentation map, $p_i = \varphi_w(x_i^S)$. This prediction is facilitated by minimizing the combined loss, composed of the cross-entropy loss ($\mathcal{L}_{CE}$) and the Lovasz-Softmax loss ($\mathcal{L}_{LS}$). While cross-entropy loss combined with mutual information has been explored in other image classification contexts [61], its integration alone might not be optimal for dense (pixel-level) recognition tasks. Therefore, we incorporated the Lovasz-Softmax loss to enhance robust feature extraction from images. The cross-entropy loss measures the discrepancy between the ground truth (y_i^{sem}) and the predicted probability map (p_i). This combined loss approach encourages the extraction of robust features that, in turn, allow the model to generate reliable predictions, irrespective of the type of microscope used. The segmentation loss involving both cross-entropy and Lovasz-Softmax loss can be expressed as follows:

$$\mathcal{L}_{seg}(x_i^S, y_i^{sem}) = \mathcal{L}_{CE}\left(y_i^{sem},\ \phi^w\left(x_i^S\right)\right) + \lambda_{LS}\mathcal{L}_{LS}\left(y_i^{sem},\ \phi^w\left(x_i^S\right)\right) \tag{16}$$

Here, λ_{LS} is the balancing coefficient that manages the contribution of the Lovasz-Softmax loss in the combined loss function, we set it to 0.5 in our experiments. We refer to this variant of out model as FARS+ and when we set the value of λ_{LS} to 0 then we refer to it as FARS.

In the case of LCM slides, we turn to entropy minimization to enhance prediction certainty. This strategy is necessary because we do not utilize the labels provided with LCM slides for training. An entropy map (e_i) is generated for each LCM input (x_i^T) to represent pixel-wise entropies of the network's predictions (p_i) for the LCM domain. Entropy is calculated by following equation:

$$e_i = -\frac{1}{\log(K)} \sum_{k=0}^{K} \varphi^w\left(x_i^T\right) . \log(\varphi^w\left(x_i^T\right)) \tag{17}$$

However, direct entropy minimization proved ineffective in low-entropy regions. Unlike previous methods like Advent [30], which directly minimize entropy, we adopted a robust entropy minimization approach. This involves using a carbonnier penalty function to more heavily penalize high-entropy predictions when $\eta > 0.5$, thereby improving feature alignment. The resulting entropy loss ($\mathcal{L}_{ent}$) is then expressed as:

$$\mathcal{L}_{ent}(x_i^T) = \left(\frac{1}{N}\sum_{i=0}^{N} {e_i}^2 + 0.0001^2\right)^{\eta} \tag{18}$$

The probability distributions of classes, as determined from the encoder and decoder's features (p_i^{aux} and p_i respectively), are then forwarded to the corresponding discriminators. The role of these discriminators is to classify whether a given input belongs to the HCM or LCM domain, based on which they assign a value of $\mathbb{1}$ for HCM slides and $\mathbb{0}$ for LCM slides. The discriminators are trained using cross-entropy loss follows:

$$\mathcal{L}_D = \mathcal{L}_{ce}\left(\theta^v(x_i^S),\ \mathbb{1}\right) + \mathcal{L}_{ce}\left(\theta^v(x_i^T), \mathbb{0}\right) + \lambda_{aux}\left(\mathcal{L}_{ce}\left(\theta^v_{aux}(x_i^S),\ \mathbb{1}\right) + \mathcal{L}_{ce}\left(\theta^v_{aux}(x_i^T),\ \mathbb{0}\right)\right) \tag{19}$$

While this may seem relatively straightforward, the distinction in textural features between slides captured with low-cost and high-cost microscopes poses a huge challenge. To enhance this complexity and compel the discriminators to backpropagate more constructive gradients for training the segmentation network, we employ the F2DA algorithm. This approach transfers the texture of LCM slides onto HCM slides, thereby maximizing the confusion and shifting the network's focus towards the unique and distinguishable features of malarial parasites, rather than the variations in stain and smear background.

By minimizing the cross-entropy loss between the discriminator's predictions for LCM slides and the labels of the HCM slides, we ensure that the model can effectively navigate the variations in slide preparation and capture conditions.

$$\mathcal{L}_{adv} = \mathcal{L}_{ce}\left(\theta^v(x_i^T),\ \mathbb{1}\right) \tag{20}$$

Now the total loss for segmentation network can be written as:

$$\mathcal{L}_{total} = \mathcal{L}_{seg} + \lambda_{ent}\mathcal{L}_{ent} + \lambda_{adv}\mathcal{L}_{adv} \tag{21}$$

To summarize, our adversarial training approach effectively aligns features derived from both HCM and LCM slides. This strategy enhances the model's capacity for generalization, providing more reliable predictions across various microscopy settings and stain intensities, and can be easily adapted for various magnifications. The end result is a model that is more robust and reliable for detecting malaria parasites under different conditions

Table 2
Comparison with state-of-the-art segmentation networks. Training and testing are done on HCM slides, evaluation metric used is *mPQ*.

Training Magnification	U-Net [66]			DeepLabv3+ [67]			nnUNet [68]			SegNext [69]			FARS		
	Test Magnifications (HCM→HCM); Metric = mPQ														
	1000x	400x	100x	1000x	400x	100x	1000x	400x	100x	1000x	400x	100x	1000x	400x	100x
1000x	60.1	42.1	2.1	68.2	54.5	7.1	71.23	51.5	4.7	72.57	58.6	9.5	78.04	64.07	5.9
400x	45.6	46.2	3.2	66.3	62.7	8.3	69.2	59.9	7.6	73.4	69.8	11.2	77.11	78.78	13.57
100x	25.7	33.1	25.9	30.1	27.3	30.2	31.03	28.57	29.8	32.4	29.5	30.6	35.72	33.58	39.14

3.7. Implementation details

Our framework was implemented using the PyTorch toolbox, on a single NVIDIA RTX-3090 GPU with a memory capacity of 24 GB. We utilized the High Cost Microscopy (HCM) and Low Cost Microscopy (LCM) divisions provided in the original M5 dataset. For HCM, all image-label pairs were employed for training, validation, and testing splits. However, for LCM, we did not use the labels of the training split, we only utilized the test split labels for LCM images to evaluate our model's performance.

We trained the segmentation network using the Stochastic Gradient Descent (SGD) optimizer with a weight decay set at 5×10^{-4}. For the discriminators, the Adam optimizer was employed and a cosine decay policy was applied to the learning rate of the segmentation network, which was initially set at 0.001 and warmed up for the first 2000 iterations. Conversely, for the discriminators, we adopted a polynomial decay policy with an initial learning rate of 10^{-4}.

4. Results

Given the low-quality nature of slides captured by Low-Cost Microscopes (LCM), identifying malaria parasites in these slides can be significantly challenging, even for field experts, compared to slides captured by High-Cost Microscopes (HCM). In the original M5 dataset, field experts exclusively labeled the HCM slides at 1000x magnification. Consequently, in our experiments, we similarly utilize HCM captured slide splits for training the network across different magnifications. The LCM split is employed for cross-domain (i.e., microscope) and cross-magnification evaluation experiments and comparisons. The central aim of our proposed framework, Fourier Adaptive Recognition System (FARS), is to demonstrate that a model trained on data which is more readily labeled by humans can perform equally well on a dataset that is difficult for field experts to label. This not only validates the efficacy of FARS but also underscores its significant practical implications in the field of malaria detection.

4.1. Evaluation metrics

In assessing the performance of our framework for malaria parasite detection, we aim to provide a comprehensive and rigorous evaluation. To achieve this, we compare our framework with both widely used semantic segmentation and object detection benchmarks.

For comparison with semantic segmentation networks, we employ the Mean Panoptic Quality (mPQ) as the evaluation metric. The mPQ is chosen over Dice Coefficient Score (DCS) and Mean Intersection over Union (mIOU) as it more effectively gauges both segmentation and detection quality in a unified and reliable manner. DCS and mIOU, while useful, have a susceptibility to over-penalization in regions of overlap, potentially skewing the evaluation [14]. This can be particularly problematic in samples that contain hard-to-identify nuclei, such as those that are poorly stained or blurry. In contrast, the mPQ offers an in-depth, robust analysis of the framework's performance, effectively demonstrating the system's ability to adapt from HCM to LCM while maintaining precision and accuracy. The PQ is defined as follows;

$$PQ = \frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|} \times \frac{\sum_{(x,y)\in TP} IoU\,(x,y)}{|TP|} \tag{22}$$

Here, TP, FP, FN, and IoU represent true positives, false positives, false negatives, and Intersection over Union, respectively.

In object detection, the PASCAL-style mean Average Precision (mAP) is employed. This metric is pivotal for comparing our model's performance with existing object detection models. The mAP is computed by averaging the precision at a series of recall levels. Following the Pascal VOC Challenge standards, we average precision over recall levels [0, 0.1, …, 1], and the mAP is calculated across all classes. The Average Precision (AP) and mAP are defined as follows:

$$AP = \frac{1}{11} \sum_{r\in\{0,0.1,\ldots,1\}} p_{\text{interp}}(r) \tag{23}$$

$$p_{\text{interp}}(r) = \max_{\tilde{r}:\tilde{r}\geq r} p(\tilde{r}) \tag{24}$$

Here, $p_{\text{interp}}(r)$ represents the interpolated precision at recall r, and $p(\tilde{r})$ is the measured precision at recall $\tilde{r}$. The mAP is then the mean of AP computed for all classes. For our evaluation, we compute the mAP at an Intersection over Union (IoU) threshold of 0.5 to account for the complexity of the scenarios in our task.

In sum, our evaluation approach provides a thorough and accurate analysis of the framework's performance, emphasizing its proficiency in adapting from HCM to LCM while maintaining a high degree of accuracy and precision in malaria parasite detection. Next we first compare the performance of our approach with bounding boxes based object detection models and next we compare with state-of-the-art segmentation models.

4.2. Object detectors vs. Segmentors

Tables 1 and 2 highlight the comparative analysis of our proposed fourier adaptive recognition system (FARS) with other state-of-the-art object detectors and segmentation networks. Each model was independently trained at all three magnification levels using high-cost microscope (HCM) slides.

FARS displays superior performance over previous models. For instance, it enhances mAP by about 0.33% (66.8% vs 67.13%) at a 1000x test-magnification when trained at the same 1000x magnification. However, when the test magnification is reduced to 400x, the performance differential widens to 20% (31.3% vs. 55.19%) favoring FARS. We attribute this significant improvement to our approach of converting bounding box labels into pixel-precise segmentation labels. This strategy enables precise localization and robust feature extraction of malarial parasites—even in the case of overlapping or touching instances.

What is particularly noteworthy is the reliability of FARS when training magnification drops to 400x. Contrarily, traditional object detection models suffer noticeable performance degradation across different test magnifications. For instance, when trained at 400x and tested from 400x to 1000x, the mAP for these models falls from 61.1% to 56.9%. Conversely, FARS's accuracy improves from 63.62% to 66.1%, demonstrating its robustness and adaptability across various magnifications.

Segmentation networks compared to object detectors display more stability and reliability across different magnifications. For instance, at 100x training magnification, there is a significant performance gap at 1000x and 400x test magnifications between object detectors and

Table 3
Comparison with object detection based domain adaptation algorithms.

Method	HCM→LCM (Metric = mAP)			
	1000x→1000x	400x→400x	1000x→400x	400x→1000x
GPA [27]	15.5	21.6	0.5	19.91
DA-Det. [25]	24.8	21.4	4.3	23.32
AdaptRCNN [24]	17.6	21.5	3.67	27.65
M5RCNN [6]	37.5	33.8	5.83	31.09
ConfMix [37]	36.8	34.47	4.79	21.71
FARS	36.04	35.31	15.68	41.16
FARS+	43.71	45.0	24.48	43.17

Table 4
Comparison with segmentation based domain adaptation algorithms.

Method	HCM→LCM (Metric = mPQ)			
	1000x→1000x	400x→400x	1000x→400x	400x→1000x
Advent [30]	10.15	15.63	9.24	18.76
AdaptSegNet [32]	13.45	19.87	10.35	21.47
CyCADA [28]	27.67	32.07	25.81	28.93
APA2Seg-Net [34]	28.42	27.64	21.54	22.39
FARS	38.97	38.31	36.67	42.16
FARS+	57.48	56.97	39.75	59.86

segmentation-based models. Notably, most detectors display no detection (0% mAP) at all, whereas segmentation models exhibit at least 2.1% mPQ. In FARS's case, this jumps to 5.9% mPQ and 5.6% mAP. It is important to mention that a true positive prediction in both mPQ and mAP calculations is considered only when there is more than 50% IOU with the ground truth instance of the same class.

Another key observation is that object detection models tend to struggle with accuracy at 100x, even when trained at this magnification, whereas segmentation models fare better. This is likely because, even with HCM, the internal structure of cells, critical for malaria cell detection and stage classification, is less discernible at 100x. Given the low accuracy at this level, we prioritize the 1000x and 400x magnifications in subsequent experiments. Interestingly, models trained on 400x magnification yield reasonable results at 1000x, even rivaling models trained at 1000x. The reverse scenario, however, does not hold true, potentially due to deep learning models' tendency for overfitting.

Given that the field of view (FOV) at 400x covers approximately 20 FOVs of 1000x, we can expedite malaria cell detection by roughly 20 times at 400x, compared to scanning at 1000x. This aspect further enhances the practicality and efficiency of our FARS framework.

4.3. Experiments on domain adaptation

We performed experiments to analyze the adaptability of different models from one microscope to another, i.e., High-Cost Microscope (HCM) to Low-Cost Microscope (LCM), as depicted in Tables 3 and 4. We also examined how a network trained on a particular microscope (e.g., HCM) and at a specific magnification (say 1000x) would perform on a slide from a different microscope and at a different magnification. This can be seen in the final two columns of Tables 3 and 4, marked as "HCM→LCM" and "1000x→400x".

Our proposed model, FARS, was compared against both object detection (Table 3) and segmentation-based (Table 4) domain adaptation algorithms. We also introduced a variant of FARS, termed FARS+, where we combined the Lovasz-Softmax loss with the cross entropy loss, see Eq. (11), during the adversarial training stage to align decoder features.

The Lovasz-Softmax loss is particularly effective in this context as it directly optimizes the mean Intersection-over-Union (IoU) loss, which is critical for semantic segmentation tasks, especially those with multiple classes. It prioritizes predictions based on the degree of error, then sequentially calculates how each error impacts the IoU score. The predictions that decrease the IoU score the most are penalized more heavily. This leads to more effective handling of misclassified instances and ultimately to improved performance across varying domains and magnifications. However, we found this loss combination decreased performance when testing in the same domain (HCM→HCM) as detailed in Table 5.

As observed in Tables 3 and 4, FARS, even with simple cross entropy loss, outperforms previous segmentation and object detection based approaches. For instance, under same magnification domain adaptation (400x→400x; HCM→LCM), FARS improved the mAP from 34.47% to 35.31%. The performance leap is even more significant, jumping from 4.79% to 15.68% and from 21.71% to 41.16% mAP under 1000x→400x and 400x→1000x domain adaptation settings, respectively.

Interestingly, recent methodologies like ConfMix, which utilize confidence-based merging strategies, exhibit comparable performance to FARS in certain scenarios. However, their effectiveness diminishes significantly under varying magnification scenarios. This limitation is clearly exhibited in our experimental outcomes, particularly in the 1000x→400x setting shown in Table 3. Here, ConfMix's performance witnesses a notable decline, dropping from 36.8% to 4.79% mAP, mirroring the trends observed in other contemporary methods such as M5RCNN and AdaptRCNN. This decline in performance can be attributed to the differences in FOVs associated with different magnifications, as visually represented in Fig. 2. In contrast, under the same evaluation conditions, both FARS and FARS+ achieve mAPs of 15.68% and 24.48%, respectively. This demonstrates the effectiveness of FARS in extracting robust features that generalize well across different domains and magnifications.

The effectiveness of FARS in context of segmentation-based domain adaptation models is similarly impressive, as shown in Table 4. Utilizing a straightforward focal loss, FARS outperforms traditional GAN-based methods, i.e., CyCADA and APA2SegNet. The diminished performance of these GAN-based approaches is often due to the increased training complexity introduced by tools like CycleGAN, which are essential for image-level domain alignment. This complexity can lead to less efficient training and hindered adaptability in diverse domain scenarios.

In contrast, the success of FARS is largely attributable to two key factors. Firstly, its category-dependent context attention mechanism plays a crucial role. This feature allows FARS to focus specifically on relevant textural and contextual details within the microscopic slides, enhancing its accuracy and reliability in segmenting and identifying malaria parasites across different microscopy types.

Secondly, and perhaps more significantly, FARS leverages the training-free F2DA algorithm for pixel-level domain alignment. This approach, in contrast to the more complex GAN-based methods, does not require additional training for texture transfer. The F2DA algorithm efficiently bridges the domain gap at a pixel level, aligning textural features without the added burden of extensive training. This not only simplifies the model's training process but also ensures more consistent performance across various domains and magnifications.

In essence, FARS's combination of category-dependent context attention and the efficient use of the F2DA algorithm for pixel-level domain alignment culminates in a robust, efficient, and highly adaptable framework.

Moreover, when Lovasz loss is used, FARS's performance is further enhanced. For instance in Table 3 the mPQ under the 1000x→1000x setting rising from 36.04% to 43.71%, and in Table 4, mAP under the 400x→1000x setting jumping from 42.16% to 59.86%. Despite the slight drop in performance in the same domain evaluation with Lovasz loss (as shown in Tables 5 and 6), it is evident that using Lovasz loss significantly bolsters FARS's performance across various domains and magnifications (see Fig. 7).

Table 5
Ablation studies on effect of each component of FARS+ framework on *cross-magnification* performance. *bPQ* represents binary panoptic quality where all the malaria infected cells are considered one class and rest is background.

Experiments	mAP			mPQ			bPQ		
Testing Magnifications →	1000x	400x	100x	1000x	400x	100x	1000x	400x	100x
HCM→HCM; Training Magnification = 400x									
RS	40.09	41.36	4.21	51.43	56.81	3.4	53.27	57.8	4.03
RS+Aug	42.6	45.8	7.63	53.18	58.29	5.49	54.63	60.3	5.91
RS+Aug+Adv	64.1	63.62	13.02	77.11	78.78	13.57	80.89	81.16	14.3
RS+Aug+Adv+F2DA$^{L\rightarrow H}$	65.45	60.77	6.31	72.67	72.34	7.51	74.9	73.88	8.01
RS+Aug+Adv+F2DA$^{(L\Leftrightarrow H)}$	56.35	54.86	4.03	68.66	70.3	9.9	72.84	72.24	12.45
RS+Aug+Adv+ F2DA$^{(L\Leftrightarrow H)}$ +LovaSz	57.96	55.82	9.23	74.22	75.96	15.28	78.9	78.99	18.58
HCM→HCM; Training Magnification = 1000x									
RS	50.87	40.15	0.2	63.42	50.74	0.16	65.79	52.63	0.19
RS+Aug	55.61	43.27	0.65	66.58	55.07	0.59	68.97	57.55	0.62
RS+Aug+Adv	67.13	55.19	5.6	78.04	64.07	0.59	81.34	66.85	0.69
RS+Aug+Adv+F2DA$^{L\rightarrow H}$	65.91	50.27	3.08	68.62	57.81	1.09	70.32	58.9	1.13
RS+Aug+Adv+F2DA$^{(L\Leftrightarrow H)}$	66.79	51.21	2.72	77.63	65.32	1.44	79.82	68.19	2.3
RS+Aug+Adv+ F2DA$^{(L\Leftrightarrow H)}$ +LovaSz	66.61	53.82	0.9	82.02	69.05	4.52	85.23	72.33	2.65

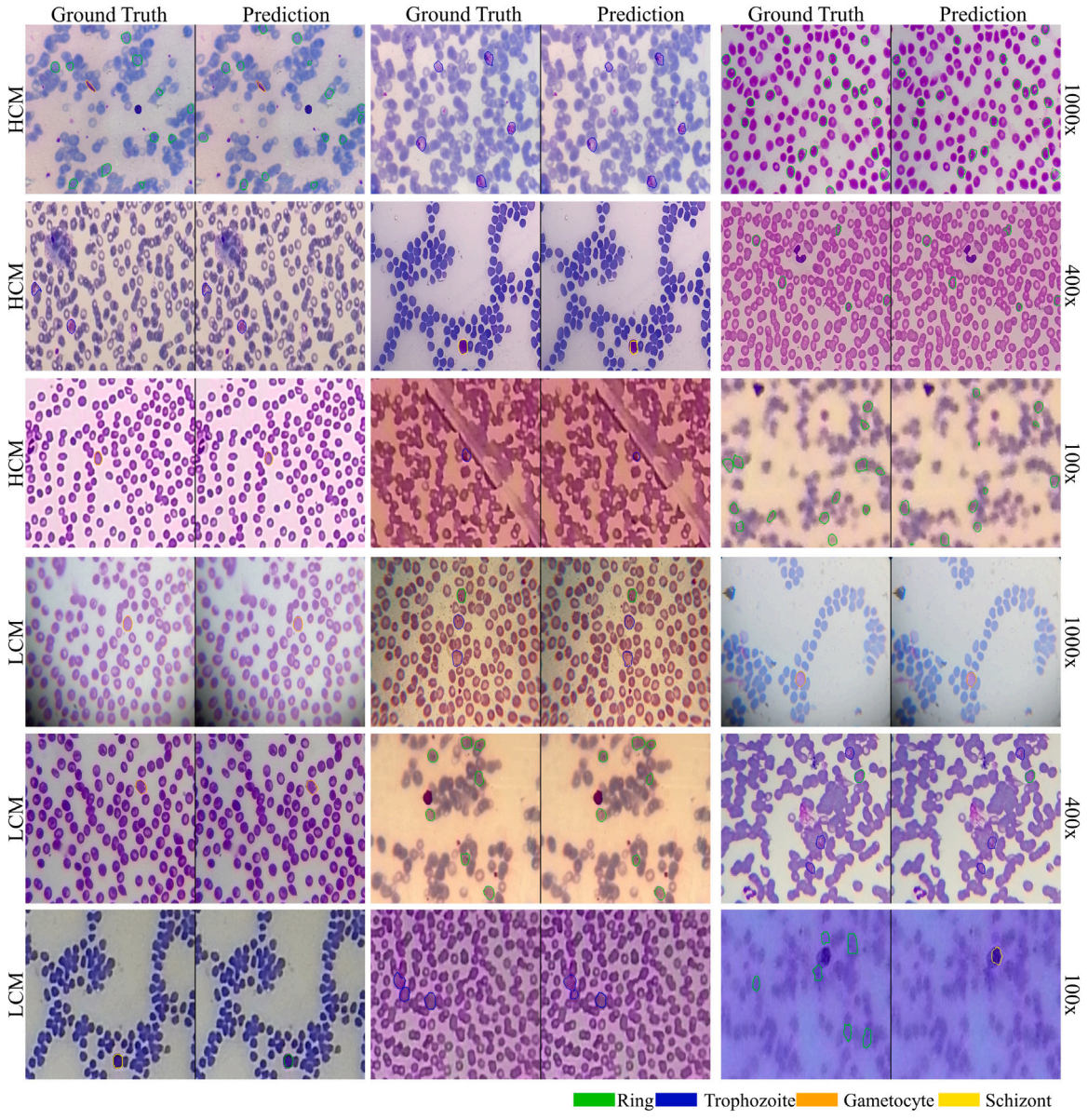

Fig. 7. Visual results of FARS on M5 dataset across different domains and magnifications. It is worth noting that here the network was only trained on HCM slides at 400x magnification.

Table 6
Ablation studies on effect of each component of FARS+ framework on $cross-domain$ performance. bPQ represents binary panoptic quality where all the malaria infected cells are considered one class and rest is background.

Experiments HCM→LCM	Training Meg → Test Meg					
	1000x→1000x			400x→400x		
	mAP	mPQ	bPQ	mAP	mPQ	bPQ
RS+Aug+Adv	10.15	23.87	25.93	42.71	49.02	50.13
RS+Aug+ F2DA$^{L\rightarrow H}$	12.56	26.37	27.58	31.09	33.78	35.6
RS+Aug+Adv+F2DA$^{L\rightarrow H}$	25.15	40.42	41.81	35.31	38.31	40.28
RS+Aug+Adv+F2DA$^{(L\Leftrightarrow H)}$	36.04	38.97	41.94	42.26	50.48	51.74
RS+Aug+Adv+F2DA$^{(L\Leftrightarrow H)}$ +LovaSz	43.71	57.84	62.25	45.0	56.97	60.87

Table 7
Ablation studies on effect of each component of FARS+ framework on $cross-magnification$ and $cross-domain$ performance. bPQ represents binary panoptic quality where all the malaria infected cells are considered one class and rest is background.

Experiments HCM→LCM	Training Meg → Test Meg					
	1000x→400x			400x→1000x		
	mAP	mPQ	bPQ	mAP	mPQ	bPQ
RS+Aug+Adv	12.68	23.58	25.22	44.96	58.35	60.66
RS+Aug+ F2DA$^{L\rightarrow H}$	12.03	20.09	20.74	41.61	42.16	44.38
RS+Aug+Adv+F2DA$^{L\rightarrow H}$	13.37	21.71	23.09	38.74	40.62	43.93
RS+Aug+Adv+F2DA$^{(L\Leftrightarrow H)}$	15.68	36.67	27.7	40.85	54.78	58.25
RS+Aug+Adv+F2DA$^{(L\Leftrightarrow H)}$ +LovaSz	24.48	39.75	44.33	43.17	59.86	64.84

4.4. Qualitative analysis

A visual comparative analysis is presented in the Fig. 7, where the ground truth and prediction results of our FARS+ model, trained on HCM slides at 400x magnification, are shown. We emphasize that the network was never trained using labels from any other magnification or microscope slides, demonstrating the effectiveness of unsupervised domain adaptation. Fig. 7 reveals the robust performance of FARS+ across different microscopes and magnifications.

However, it is worth noting that at 100x magnification, the images from both microscopes (HCM and LCM) lack visible texture or distinct parasitic structure, which results in numerous false detections at this magnification. Classifying cells at 400x and 1000x magnifications is complex due to transitions between parasite stages and multiple infections. Traditional diagnostic methods like light microscopy and Rapid Diagnostic Tests (RDTs) are time-consuming and costly, while other techniques, e.g., Polymerase Chain Reaction (PCR), are impractical for field deployment. Our model, FARS+, offers an efficient solution for malaria diagnosis by using unsupervised domain adaptation to provide accurate predictions across different domains and magnifications, making it suitable for rapid, large-scale use in resource-limited settings.

5. Discussion

To determine the optimal configurations for our FARS framework, we carried out a series of ablation studies, examining every possible combination of the proposed frameworks. This includes various combinations of our novel encoder–decoder architecture referred to as (RS) here, adversarial training objective (Adv), color domain-aware Fourier domain adaptation (F2DA), and loss functions. Tables 5 represent the performance contributions of each component of our framework across different magnifications, while Tables 6 and 7 demonstrate the impact across both domains and magnifications, respectively. For a comprehensive understanding, it is advised to examine all three tables concurrently.

In the subsequent discussion, we will streamline the focus on the mAP metric for 1000x magnification under same domain (HCM→HCM), cross domain (HCM→LCM), and one cross magnification setting (1000x→400x). The trend for the other metric, mPQ, is similar to the performance observed for mAP across all settings and is detailed in Tables 5 through 7.

Beginning with our baseline recognition system (RS) and incorporating standard data augmentations (Aug) to mitigate overfitting, we observed a notable improvement in mAP performance by 5%. This improvement underscores the importance of data augmentation in deep network applications, as they are generally data-hungry and require diverse training examples to enhance reliability and robustness. The positive impact of these augmentations, supported by existing literature [70–72], suggests the potential for further enhancements using an expanded dataset in future iterations of our research.

Integrating the adversarial objective (Adv) into our framework marked a significant leap in performance. In the same domain setting, the mAP surged from 55.61% to 67.13%, highlighting the efficacy of this approach. When we extended our evaluation to cross-domain scenarios (as reported in Tables 6 and 7), starting with the MD and Adv combination, the initial mAP for cross-domain accuracy at 1000x magnification was 10.15%. Meanwhile, in cross-magnification scenarios (1000x→400x), the starting point was 12.68% mAP.

In our next exploration, we applied texture transfer via F2DA from LCM to HCM slides ($F2DA^{L\rightarrow H}$). While there was a slight dip in same domain evaluation performance, the cross-domain performance more than doubled, going from 10.15% to 25.15% mAP. Cross-domain adaptation and cross magnification experiments saw an increase from 12.68% to 13.37% mAP. We attributed these results to adversarial training's ability to blur the distinction between source and target domain features, which aligns the features across both domains, thereby making them microscope-agnostic. To further enhance this 'confusion', we transferred the texture of slides between both microscopes (LCM to HCM and vice versa, denoted as HCM⇆LCM). This had a minor improvement for the same domain setting (from 65.91% to 66.79% mAP), but significantly improved cross-domain performance from 25.15% to 36.04% mAP, and increased mAP in cross-magnification settings (1000x→400x) from 13.37% to 15.68%. Incorporating the Lovasz-Softmax loss into our model further elevated the cross-domain and cross magnification mAP to 43.71% and 24.48% respectively. Looking forward, we propose exploring capsule network-based methods for future research, considering their potential in enhancing image classification tasks due to their unique feature representation capabilities [73–76].

While initial observation of the evaluation performance under the same domain setting (HCM→HCM) at 400x magnification might question the use of the F2DA algorithm and Lovasz loss function, due to a drop in performance from 63.62% mAP to 55.82% mAP, the contrast becomes clear when cross-domain and cross-magnification experiments for 400x magnification are taken into account. The additions do indeed enhance the performance of FARS across domains and magnifications.

In conclusion, our work has been successful in delivering on its main objective: the development of an adaptive recognition system that improves performance across different domains and magnifications. The methodological enhancements we have incorporated in the FARS framework, including the incorporation of adversarial training, Fourier Domain Adaptation, and Lovasz-Softmax loss function, have collectively contributed to making our model robust and versatile. While performance within the same domain might sometimes seem compromised, it is crucial to recognize that the real strength of our framework lies in its adaptability across varying domains and magnifications.

6. Conclusion

This research presents an inventive approach in the field of medical imaging and disease diagnosis. The proposed Fourier Adaptive Recognition System (FARS) is uniquely capable of adaptability across different domains (microscpes) and magnifications, a critical feature for detecting diseases like malaria from thin smear microscopic slides. By first incorporating a effective and reliable conversion of bounding box labels to semantic segmentation labels, FARS provides a more granular

analysis of the slides. Additionally, the use of adversarial training and color domain aware fourier domain adaptation (F2DA) enables the extraction of robust and microscope-agnostic features. The model's architecture also benefits from a category-dependent context attention module, which further enhances its adaptability and precision. Through the synergistic employment of multiple loss functions, FARS demonstrates significant performance improvements across diverse domains and magnifications.

Empirical data from our experiments underline FARS's superior performance. For example, in standard domain adaptation scenarios from HCM to LCM, FARS improved the mAP remarkably, from 34.47% to 45.0%. In more challenging cross-domain and cross-magnification settings (HCM→LCM+1000x→400x), the mAP enhancement was even more pronounced, soaring from 4.79% to 15.68%. These results validate FARS's effectiveness in diverse operational conditions.

While there is a marginal dip in performance in same-domain (HCM→HCM) and same-magnification (1000x→1000x) scenarios – from 67.13% to 66.61% in mAP – it is essential to emphasize that FARS's main strength is its adaptability across different microscopy settings. Its versatile nature provides a strong foundation for future research and development in other diseases and imaging techniques, thereby potentially amplifying its impact on computer aided diagnostic systems.

CRediT authorship contribution statement

Talha Ilyas: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. **Khubaib Ahmad:** Formal analysis, Validation, Writing – review & editing. **Dewa Made Sri Arsa:** Formal analysis, Validation, Writing – review & editing. **Yong Chae Jeong:** Funding acquisition, Resources, Writing – review & editing. **Hyongsuk Kim:** Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the Crop and Weed Project administered through the Agricultural Science and Technology Development Cooperation Research Program, South Korea (PJ015720) and by the National Research Foundation of Korea (NRF), South Korea grant funded by the Korea government (NRF-2019R1A2C1011297 and NRF-2019R1A6A1A09031717).

References

[1] W. Team, Malaria Report, World Health Organization, WHO, 2023.
[2] N. Sengar, R. Burget, M.K. Dutta, A vision transformer based approach for analysis of plasmodium vivax life cycle for malaria prediction using thin blood smear microscopic images, Comput. Methods Programs Biomed. 224 (2022) 106996.
[3] D.R. Loh, W.X. Yong, J. Yapeter, K. Subburaj, R. Chandramohanadas, A deep learning approach to the screening of malaria infection: Automated and rapid cell counting, object detection and instance segmentation using Mask R-CNN, Comput. Med. Imaging Graph. 88 (2021) 101845.
[4] C.K. Murray, R.A. Gasser Jr., A.J. Magill, R.S. Miller, Update on rapid diagnostic testing for malaria, Clin. Microbiol. Rev. 21 (1) (2008) 97–110.
[5] R. Shouval, J.A. Fein, B. Savani, M. Mohty, A. Nagler, Machine learning and artificial intelligence in haematology, Br. J. Haematol. 192 (2) (2021) 239–250.
[6] W. Sultani, W. Nawaz, S. Javed, M.S. Danish, A. Saadia, M. Ali, Towards low-cost and efficient malaria detection, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2022, pp. 20655–20664.
[7] Y.W. Lee, J.W. Choi, E.-H. Shin, Machine learning model for predicting malaria using clinical information, Comput. Biol. Med. 129 (2021) 104151.
[8] S. Rahman, B. Azam, S.U. Khan, M. Awais, I. Ali, et al., Automatic identification of abnormal blood smear images using color and morphology variation of RBCS and central pallor, Comput. Med. Imaging Graph. 87 (2021) 101813.
[9] A. Molina, J. Rodellar, L. Boldú, A. Acevedo, S. Alférez, A. Merino, Automatic identification of malaria and other red blood cell inclusions using convolutional neural networks, Comput. Biol. Med. 136 (2021) 104680.
[10] M. Hayat, M. Tahir, F.K. Alarfaj, R. Alturki, F. Gazzawe, NLP-BCH-Ens: NLP-based intelligent computational model for discrimination of malaria parasite, Comput. Biol. Med. 149 (2022) 105962.
[11] A.F. Al-Battal, I.R. Lerman, T.Q. Nguyen, Multi-path decoder U-Net: A weakly trained real-time segmentation network for object detection and localization in ultrasound scans, Comput. Med. Imaging Graph. 107 (2023) 102205.
[12] Z. Salahuddin, H.C. Woodruff, A. Chatterjee, P. Lambin, Transparency of deep neural networks for medical image analysis: A review of interpretability methods, Comput. Biol. Med. 140 (2022) 105111.
[13] N. Tajbakhsh, J.Y. Shin, S.R. Gurudu, R.T. Hurst, C.B. Kendall, M.B. Gotway, J. Liang, Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Trans. Med. Imaging 35 (5) (2016) 1299–1312.
[14] S. Graham, Q.D. Vu, S.E.A. Raza, A. Azam, Y.W. Tsang, J.T. Kwak, N. Rajpoot, Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images, Med. Image Anal. 58 (2019) 101563.
[15] J. Gamper, N. Alemi Koohbanani, K. Benet, A. Khuram, N. Rajpoot, Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification, in: Digital Pathology: 15th European Congress, ECDP 2019, Warwick, UK, April 10–13, 2019, Proceedings, vol. 15, Springer, 2019, pp. 11–19.
[16] S. Graham, M. Jahanifar, A. Azam, M. Nimir, Y.-W. Tsang, K. Dodd, E. Hero, H. Sahota, A. Tank, K. Benes, et al., Lizard: a large-scale dataset for colonic nuclear instance segmentation and classification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 684–693.
[17] S. Chen, C. Ding, M. Liu, J. Cheng, D. Tao, CPP-net: Context-aware polygon proposal network for nucleus segmentation, IEEE Trans. Image Process. 32 (2023) 980–994.
[18] D. Maini, A.K. Aggarwal, Camera position estimation using 2D image dataset, Int. J. Innov. Eng. Technol. 10 (2018) 199–203.
[19] S. Moon, S. Lee, H. Kim, L.H. Freitas-Junior, M. Kang, L. Ayong, M.A. Hansen, An image analysis algorithm for malaria parasite stage classification and viability quantification, PLoS One 8 (4) (2013) e61812.
[20] K. Prasad, J. Winter, U.M. Bhat, R.V. Acharya, G.K. Prabhu, Image analysis approach for development of a decision support system for detection of malaria parasites in thin blood smear images, J. Digit. Imag. 25 (2012) 542–549.
[21] K.E.D. Peñas, P.T. Rivera, P.C. Naval, Malaria parasite detection and species identification on thin blood smears using a convolutional neural network, in: 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE, IEEE, 2017, pp. 1–6.
[22] C. Mehanian, M. Jaiswal, C. Delahunt, C. Thompson, M. Horning, L. Hu, T. Ostbye, S. McGuire, M. Mehanian, C. Champlin, et al., Computer-automated malaria diagnosis and quantitation using convolutional neural networks, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 116–125.
[23] C.B. Delahunt, M.S. Jaiswal, M.P. Horning, S. Janko, C.M. Thompson, S. Kulhare, L. Hu, T. Ostbye, G. Yun, R. Gebrehiwot, et al., Fully-automated patient-level malaria assessment on field-prepared thin blood film microscopy images, in: 2019 IEEE Global Humanitarian Technology Conference, GHTC, IEEE, 2019, pp. 1–8.
[24] Y. Chen, W. Li, C. Sakaridis, D. Dai, L. Van Gool, Domain adaptive faster r-cnn for object detection in the wild, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3339–3348.
[25] K. Saito, Y. Ushiku, T. Harada, K. Saenko, Strong-weak distribution alignment for adaptive object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 6956–6965.
[26] X. Zhu, J. Pang, C. Yang, J. Shi, D. Lin, Adapting object detectors via selective cross-domain alignment, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 687–696.
[27] M. Xu, H. Wang, B. Ni, Q. Tian, W. Zhang, Cross-domain detection via graph-induced prototype alignment, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 12355–12364.
[28] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.
[29] J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko, A. Efros, T. Darrell, Cycada: Cycle-consistent adversarial domain adaptation, in: International Conference on Machine Learning, PMLR, 2018, pp. 1989–1998.
[30] T.-H. Vu, H. Jain, M. Bucher, M. Cord, P. Pérez, Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 2517–2526.
[31] Y. Yang, S. Soatto, Fda: Fourier domain adaptation for semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 4085–4095.

[32] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang, M. Chandraker, Learning to adapt structured output space for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7472–7481.
[33] M. Long, Y. Cao, J. Wang, M. Jordan, Learning transferable features with deep adaptation networks, in: International Conference on Machine Learning, PMLR, 2015, pp. 97–105.
[34] B. Zhou, Z. Augenfeld, J. Chapiro, S.K. Zhou, C. Liu, J.S. Duncan, Anatomy-guided multimodal registration by learning segmentation without ground truth: Application to intraprocedural CBCT/MR liver segmentation and registration, Med. Image Anal. 71 (2021) 102041.
[35] F. Xing, X. Yang, T.C. Cornish, D. Ghosh, Learning with limited target data to detect cells in cross-modality images, Med. Image Anal. 90 (2023) 102969.
[36] W. Tranheden, V. Olsson, J. Pinto, L. Svensson, Dacs: Domain adaptation via cross-domain mixed sampling, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 1379–1389.
[37] G. Mattolin, L. Zanella, E. Ricci, Y. Wang, ConfMix: Unsupervised domain adaptation for object detection via confidence-based mixing, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 423–433.
[38] S. Li, Z. Du, X. Meng, Y. Zhang, Multi-stage malaria parasite recognition by deep learning, GigaScience 10 (6) (2021) giab040.
[39] A. Rahman, M.S. Rahman, M. Mahdy, 3C-GAN: class-consistent CycleGAN for malaria domain adaptation model, Biomed. Phys. Eng. Express 7 (5) (2021) 055002.
[40] R.T.C. Ramarolahy, E.O. Gyasi, A. Crimi, Classification and generation of microscopy images with plasmodium falciparum via artificial neural networks using low cost settings, in: Domain Adaptation and Representation Transfer, and Affordable Healthcare and AI for Resource Diverse Global Health: Third MICCAI Workshop, DART 2021, and First MICCAI Workshop, FAIR 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27 and October 1, 2021, Proceedings, vol. 3, Springer, 2021, pp. 147–157.
[41] A. Srivastava, V. Singhal, A.K. Aggarawal, Comparative analysis of multimodal medical image fusion using PCA and wavelet transforms, Int. J. Latest Technol. Eng. Manag. Appl. Sci. (IJLTEMAS) VI (2017).
[42] M. Rajchl, M.C. Lee, O. Oktay, K. Kamnitsas, J. Passerat-Palmbach, W. Bai, M. Damodaram, M.A. Rutherford, J.V. Hajnal, B. Kainz, et al., Deepcut: Object segmentation from bounding box annotations using convolutional neural networks, IEEE Trans. Med. Imaging 36 (2) (2016) 674–683.
[43] Y. Ou, S.X. Huang, K.K. Wong, J. Cummock, J. Volpi, J.Z. Wang, S.T. Wong, BBox-Guided Segmentor: Leveraging expert knowledge for accurate stroke lesion segmentation using weakly supervised bounding box prior, Comput. Med. Imaging Graph. 107 (2023) 102236.
[44] P.M. Tedder, J.R. Bradford, C.J. Needham, G.A. McConkey, A.J. Bulpitt, D.R. Westhead, Gene function prediction using semantic similarity clustering and enrichment analysis in the malaria parasite Plasmodium falciparum, Bioinformatics 26 (19) (2010) 2431–2437.
[45] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A.C. Berg, W.-Y. Lo, et al., Segment anything, 2023, arXiv preprint arXiv:2304.02643.
[46] S. He, R. Bao, J. Li, P.E. Grant, Y. Ou, Accuracy of segment-anything model (sam) in medical image segmentation tasks, 2023, arXiv preprint arXiv:2304.09324.
[47] J. Wu, R. Fu, H. Fang, Y. Liu, Z. Wang, Y. Xu, Y. Jin, T. Arbel, Medical sam adapter: Adapting segment anything model for medical image segmentation, 2023, arXiv preprint arXiv:2304.12620.
[48] Y. Huang, X. Yang, L. Liu, H. Zhou, A. Chang, X. Zhou, R. Chen, J. Yu, J. Chen, C. Chen, et al., Segment anything model for medical images? 2023, arXiv preprint arXiv:2304.14660.
[49] A.K. Aggarwal, Biological Tomato Leaf disease classification using deep learning framework, Int. J. Biol. Biomed. Eng. 16 (1) (2022) 241–244.
[50] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, R. Ng, Fourier features let networks learn high frequency functions in low dimensional domains, Adv. Neural Inf. Process. Syst. 33 (2020) 7537–7547.
[51] E. Goceri, Evaluation of denoising techniques to remove speckle and Gaussian noise from dermoscopy images, Comput. Biol. Med. 152 (2023) 106474.
[52] E. Goceri, Intensity normalization in brain MR images using spatially varying distribution matching, in: 11th Int. Conf. on Computer Graphics, Visualization, Computer Vision and Image Processing, CGVCVIP 2017, 2017, pp. 300–4.
[53] E. Goceri, Fully automated and adaptive intensity normalization using statistical features for brain MR images, Celal Bayar Univ. J. Sci. 14 (1) (2018) 125–134.
[54] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, 2020, arXiv preprint arXiv:2010.11929.
[55] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.
[56] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[57] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, Y. Li, Maxvit: Multi-axis vision transformer, in: European Conference on Computer Vision, Springer, 2022, pp. 459–479.
[58] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012–10022.
[59] Z. Geng, M.-H. Guo, H. Chen, X. Li, K. Wei, Z. Lin, Is attention better than matrix decomposition? 2021, arXiv preprint arXiv:2109.04553.
[60] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.
[61] E. Göçeri, An application for automated diagnosis of facial dermatological diseases, İzmir Katip Çelebi Üniversitesi Sağlık Bilimleri Fakültesi Dergisi 6 (3) (2021) 91–99.
[62] Z. Tian, C. Shen, H. Chen, T. He, Fcos: Fully convolutional one-stage object detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 9627–9636.
[63] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.
[64] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779–788.
[65] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, in: Advances in Neural Information Processing Systems, vol. 28, 2015.
[66] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III, vol. 18, Springer, 2015, pp. 234–241.
[67] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 801–818.
[68] F. Isensee, P.F. Jaeger, S.A. Kohl, J. Petersen, K.H. Maier-Hein, nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, Nat. Methods 18 (2) (2021) 203–211.
[69] M.-H. Guo, C.-Z. Lu, Q. Hou, Z. Liu, M.-M. Cheng, S.-M. Hu, Segnext: Rethinking convolutional attention design for semantic segmentation, Adv. Neural Inf. Process. Syst. 35 (2022) 1140–1156.
[70] E. Goceri, Medical image data augmentation: techniques, comparisons and interpretations, Artif. Intell. Rev. (2023) 1–45.
[71] E. Goceri, Comparison of the impacts of dermoscopy image augmentation methods on skin cancer classification and a new augmentation method with wavelet packets, Int. J. Imaging Syst. Technol. (2023).
[72] E. Goceri, Image augmentation for deep learning based lesion classification from skin images, in: 2020 IEEE 4th International Conference on Image Processing, Applications and Systems, IPAS, IEEE, 2020, pp. 144–148.
[73] E. Goceri, Capsule neural networks in classification of skin lesions, in: International Conference on Computer Graphics, Visualization, Computer Vision and Image Processing, 2021, pp. 29–36.
[74] E. Goceri, Classification of skin cancer using adjustable and fully convolutional capsule layers, Biomed. Signal Process. Control 85 (2023) 104949.
[75] A. Kumar, A. Banno, S. Ono, T. Oishi, K. Ikeuchi, Global coordinate adjustment of the 3D survey models under unstable GPS condition, Seisan Kenkyu 65 (2) (2013) 91–95.
[76] A. Kumar, Y. Sato, T. Oishi, S. Ono, K. Ikeuchi, Improving gps position accuracy by identification of reflected gps signals using range data for modeling of urban structures, Seisan Kenkyu 66 (2) (2014) 101–107.