IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 November 2024, accepted 14 December 2024, date of publication 24 December 2024, date of current version 3 January 2025.

Digital Object Identifier 10.1109/ACCESS.2024.3521937

RESEARCH ARTICLE

Automated Morphology Detection of Nail-Fold Capillaries Through Enhanced Object Detection Network

HANG THI PHUONG NGUYEN[1], (Member, IEEE), AND HIEYONG JEONG[1,2], (Member, IEEE)
[1]Department of Artificial Intelligence Convergence, Chonnam National University, Buk-gu, Gwangju 61186, Republic of Korea
[2]Graduate School of Medicine, Osaka University, Suita, Osaka 565-0871, Japan

Corresponding author: Hieyong Jeong (h.jeong@jnu.ac.kr)

This work was supported in part by the Basic Science Research Program through the National Research Foundation (NRF) of Korea Grant funded by the Ministry of Education under Grant NRF-2021R1I1A3055210; in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) through the Artificial Intelligence Convergence Innovation Human Resources Development Grant funded by Korean Government through the Ministry of Science and ICT (MSIT), South Korea, under Grant IITP-2023-RS-2023-00256629; and in part by MSIT through the Information Technology Research Center (ITRC) Support Program by the Institute for Information Communications Technology Planning Evaluation (IITP) under Grant IITP-2024-RS-2024-00437718.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Osaka University Hospital under Application No. 18546-6, December 07, 2021, and performed in line with the Declaration of Helsinki.

ABSTRACT The analysis of nail-fold anatomy can effectively evaluate microcirculation and diagnose vascular-related diseases. Early detection of these conditions is crucial due to the risk of severe complications if intervention is delayed. Extensive research supports the notion that nail-fold capillary morphology serves as a critical biomarker for various disease processes, with the degree of capillary structural damage potentially reflecting the involvement of internal organs. This study proposes a non-invasive methodology for detecting nail-fold capillary morphology by integrating an object detection model for improvement within a deep learning framework. We conducted an ablation study to enhance YOLOv8's performance in detecting nail-fold capillaries and classifying their morphology. Our enhancements included adding a detection layer to improve the detection of various-sized objects, implementing Efficient Channel Attention (ECA) mechanisms, and incorporating data augmentation techniques and hyper-parameter tuning. These modifications yielded a notable improvement in mean Average Precision at IoU 0.50 (mAP@50), with increases of 3.7% in mAP, 3.6% in precision, and 2.5% in recall compared to the baseline YOLOv8 model. This culminated in a mAP@50 score of 79.9%. We also utilized Slicing-Aided Hyperinference (SAHI) to enhance inference performance on untrained multi-scale images and smaller capillaries, demonstrating significant effectiveness in real-time testing scenarios. The results from this research are promising for advancing early-stage diabetes detection using nail-fold image analysis and could potentially enable real-time applications in clinical environments.

INDEX TERMS Nail-fold capillaries, YOLOv8, ECA attention, diabetes, capillaroscopy, SAHI, object detection, deep learning.

I. INTRODUCTION

Nail-fold capillaries are located beneath the proximal nail-fold region of the fingers, oriented in a parallel arrangement along the dermis, as illustrated by the red rectangle in Figure 1(c). These nail-fold microvascular abnormalities are distinguishing features and play a significant role in the pathogenesis of various systemic conditions, including rheumatoid arthritis (RA) [1], systemic sclerosis (SSc) [2], dermatomyositis (DM) [3], systemic lupus erythematosus

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy.

(SLE) [4], [5], and diabetes mellitus [6]. The diagnostic evaluation of nail-fold capillary alterations can be instrumental for early disease detection and monitoring, often preceding the onset of clinical symptoms. Manifestations of these abnormalities may include [1], [2], [3], [6], and [7]:

- Abnormal morphologies: Deviations from normative capillary shape and structure, indicating potential underlying pathology.
- Sparse capillary density: A reduced number of discernible capillaries in the nail-fold area may correlate with systemic vascular compromise.
- Microaneurysms or hemorrhages: Localized capillary dilation or bleeding episodes indicate vascular integrity disruption within the microcirculation.

These capillary assessments are vital for clinicians to establish early therapeutic interventions and longitudinal evaluations in affected patient populations.

Diabetes mellitus (DM) is an increasingly prevalent chronic condition that presents significant public health challenges, particularly in developing and newly industrialized countries. Notably, the condition is largely preventable, underscoring the need for timely interventions. In the Americas, diabetes is responsible for 5.9% of mortality in adults aged 20 and older, with type 2 diabetes (T2D) disproportionately contributing to this burden [8], [9]. The global incidence of T2D has risen sharply among adolescents and young adults from 1990 to 2019, and projections indicate that by 2030, around 470 million people will be living with prediabetes, reinforcing the critical need for effective preventive strategies [10]. DM is defined by persistent hyperglycemia, characterized by consistently elevated blood glucose levels [11]. This chronic state of hyperglycemia leads to extensive damage to multiple organ systems, particularly the vascular system, resulting in both microvascular and macrovascular complications [6], [8]. Extensive research has elucidated the strong correlation between chronic hyperglycemia and alterations in capillary architecture. In diabetic patients, vascular neogenesis occurs as a compensatory response to tissue ischemia, leading to significant morphological changes within the microcirculation, such as increased capillary diameter, enhanced tortuosity, and reduced vessel density [12]. These alterations indicate disease progression, declining glycemic control, and the onset of systemic complications [13]. Thus, the importance of early diabetes detection cannot be overstated. It is crucial for optimizing therapeutic interventions and reducing the risk of complications.

Nail-fold capillaroscopy (NFC) is a non-invasive, user-friendly, and cost-effective technique providing critical insights into viable tissues' microvascular abnormalities. Historically, the NFC has been employed to examine the microvascular alterations in nail-fold capillaries associated with various rheumatologic conditions [14]. The increasing literature on NFC in rheumatic diseases highlights the diversity of research applications and the technique's efficacy in clinical settings. Furthermore, The NFC has proven beneficial in investigating other systemic disorders characterized by alterations in capillary architecture, such as diabetes mellitus [15] and schizophrenia [16], among others [17].

The NFC technique employs a capillaroscope to capture high-resolution images of nail-fold capillaries. After acquiring these images, clinicians, with their invaluable expertise, conduct visual assessments and manual measurements of various parameters, including the presence of dilated capillaries, capillary morphology, and capillary density quantified as the number of loops per millimeter. This clinical analysis, driven by their expertise, is crucial, but ensuring consistent and standardized evaluations across healthcare providers remains a significant challenge. Efforts have been made to automate the analysis of capillary images, but variability in capillary size has yet to improve outcomes. This inconsistency in size limits the training of machine learning models and ultimately results in suboptimal predictive performance, which could lead to misdiagnosis or delayed treatment for patients. To address this limitation, we proposed an enhanced-deep-learning-based method for automated nail-fold capillaroscopy image analysis. This approach leveraged object detection algorithms to detect and classify individual nail-fold capillaries' morphology automatically, crucial parameters for medical diagnostics in general and early diabetes mellitus diagnosis in particular because it reflects the involvement of internal organs [6], [18]. Furthermore, the enhanced deep learning framework employed in our method demonstrated the ability to detect capillaries within multi-scale, untrained microscopic images. The main contributions include:

- We performed ablation studies employing various feature fusion techniques, including the Feature Pyramid Network (FPN), the FPN-path-aggregation network (FPN-PAN), and the bidirectional Feature Pyramid Network (BiFPN), all integrated with You Only Look Once version 8 (YOLOv8). The primary objective was to evaluate the effectiveness of FPN-PAN within the context of our specific dataset.
- We introduced enhancements to the YOLOv8, focusing on integrating detection layers designed explicitly for various-sized object detection. Additionally, we implemented the Efficient Channel Attention (ECA) module to refine feature extraction. Our approach included advanced data augmentation techniques and rigorous hyper-parameter optimization. These modifications markedly improved model performance relative to the YOLOv8 baseline, demonstrating their efficacy in various detection tasks.
- To demonstrate the effectiveness of our proposed model, we conducted a comparative analysis of multiple object detection models. Comparison results included several versions of the YOLO frameworks, Faster R-CNN, the Real-Time Detection Transformer (RT-DETR), and the scalable EfficientDet model.
- We implemented Slicing-Aided-Hyper-Inference (SAHI) methodologies to improve inference performance

on multi-magnification, untrained images, thereby augmenting the detection capabilities of small objects.

The paper is organized as follows: Section II comprehensively reviews existing literature on nail-fold capillary detection techniques. Section III details the experimental framework and the dataset utilized in the study. Section IV delineates the proposed models in depth. Section V discusses the results from the ablation experiments conducted and the comparison results. Section VI provides a thorough and rigorous analysis, instilling confidence in the validity of our findings. Finally, Section VII summarizes the conclusions and recommends future research directions.

II. RELATED WORKS

Previous studies have examined a range of traditional object detection methodologies rooted in computer vision, such as template matching, Object-based Image Analysis (OBIA), and knowledge-based detection techniques [19]. Demir et al. employed the Contrast-Limited-Adaptive-Histogram-Equalization (CLAHE) method to enhance the visibility of capillaries in imaging. By improving image contrast, this technique facilitates visualization and subsequent analysis of these small vascular structures. Additionally, the authors employed an adjustable thresholding mechanism to optimize detection accuracy. Nonetheless, this approach may require manual parameter tuning, which can be influenced by external variables such as ambient lighting conditions and dermal thickness [20].

Over the past decade, the field of object detection has witnessed a profound transformation, largely propelled by the emergence of deep neural networks [21]. These networks have played a pivotal role in the development of robust object detectors, which can be broadly categorized into two main types: CNN-based models and Vision Transformer models.

CNN-based approaches can be classified into one-stage and two-stage methods. One-stage detectors, such as YOLO (2015-2024) and EfficientDet (2019) [22], are particularly efficient for real-time applications that demand swift processing. They achieve this by using a unified architecture to process the entire input image in a single pass, enabling direct predictions of both object classes and bounding box coordinates. In contrast, two-stage methods like R-CNN (2014), Fast R-CNN (2015), and Faster R-CNN (2016) [23] first generate region proposals before classifying these proposals, making them relatively more complex and slower. On the other hand, Vision Transformers have emerged as a promising avenue for object detection. Models like the Detection Transformer (DETR) [24] introduced in 2020 and the Real-Time Detection Transformer (RT-DETR) [25] released in 2023 apply transformer architectures to object detection tasks, showcasing the potential for high performance, especially in scenarios where integrating context and global reasoning is beneficial.

Vision transformer-based object detection models excel in processing extensive datasets and managing intricate detection challenges, albeit at the cost of significant computational resources [26]. The original DEtection TRansformer (DETR) model, introduced in 2020 [24], employs transformers to simultaneously predict bounding boxes and class labels from an input image. Its architecture is built upon a Convolutional Neural Network (CNN) backbone, a transformer encoder-decoder structure, and dedicated prediction heads. However, DETR's prolonged inference times render it less suitable for real-time applications. RT-DETR, released in 2023 [25], enhances the model's adaptability by providing flexible speed tuning options through adjustable decoder layers to mitigate this limitation. This allows for modifications tailored to specific application requirements without necessitating retraining of the entire model. Despite these advancements, RT-DETR prefers larger datasets over smaller ones and demands significant computational resources.

YOLO (You Only Look Once) represents a significant advancement in object detection, enabling the processing of images in a single forward pass through a convolutional neural network. The model employs a grid-based approach, allowing each cell to make simultaneous predictions regarding bounding boxes, objectness scores, and class probabilities. This architecture confers notable advantages in terms of speed compared to other one-stage methods, rendering YOLO particularly well-suited for real-time applications [27]. In 2021, research conducted by Tan et al. identified YOLOv3 as presenting an optimal compromise between detection speed and real-time performance when contrasted with RetinaNet and SSD [28]. Subsequent studies have established that YOLOv5 outperforms traditional two-stage models such as Faster R-CNN in critical metrics, including precision, recall, processing speed, and training efficiency, particularly evident in specialized tasks like screw head detection [29]. Most recently, Huang et al. [30] introduced enhancements to YOLOv8, featuring a new detection head designed to improve detection performance across varying object sizes. Their findings indicate that YOLOv8 significantly exceeds YOLOv3 and YOLOv5 in terms of accuracy for small object detection tasks.

KV et al. [12] conducted a study using nail-fold capillaries, where they implemented YOLOv2 and YOLOv3 to analyze a dataset of 600 images from healthy and diabetic individuals. Their objective was to accurately localize capillaries within the nail-fold region and categorize them into five distinct classes: normal, wide, elongated, tortuous, and hemorrhagic. They devised a mathematical model to refine further patient classification, classifying individuals as either ''Normal'' or ''Diabetic''. In 2023, Nguyen et al. advanced this research by leveraging YOLOv5 and BiFPN to enhance capillary detection precision. They developed a nail-fold capillaroscopy dataset through microscopy, thoroughly annotating it into four classes: ''hairpin'', ''crossing'', ''bushy'', and ''tortuous''. Their model yielded a notable mean Average Precision at IoU 0.5 (mAP@50) score of 74.2% [31]. While these studies underscore the effectiveness of deep learning models, particularly YOLO variants, in detecting microvascular alterations via nail-fold imagery, they fail to

address the inherent challenges of small object detection in this domain. Moreover, the scalability of these findings to untrained multi-scale images, which better reflect real-world clinical scenarios, remains an open question.

In recent decades, there has been a growing emphasis on improving the robustness of deep learning models, resulting in various proposed methodologies to achieve this objective. Critical strategies for augmenting model robustness encompass [32]:

- Data: Data augmentation is an effective strategy for enhancing model robustness by expanding the diversity of the training dataset. This approach facilitates the model's invariance to specific features or concepts, ultimately improving generalization performance [33].
- Architecture: Architectural robustness strategies encompass a range of approaches, from optimizing individual layers to refining entire network architectures. Enhancing network performance can be achieved by increasing the depth or width of the architecture, which involves adding layers or expanding the number of neurons in each layer. Furthermore, refining the learning algorithm and incorporating attention mechanisms allow models to compute weighted feature maps, enabling them to better adapt to various imaging corruptions [34], [35].
- Optimization: Optimization-based strategies improve robustness by adjusting learning objectives and conducting hyper-parameter tuning. These methods enhance resilience against corrupted inputs while promoting the acquisition of informative and discriminative features. Modifications to the loss function or the training protocol are implemented to mitigate the risk of the model learning shortcuts that could hinder generalization, ultimately reducing vulnerability to brittleness [32].

We opted for YOLOv8 architecture in this study due to its superior accuracy and efficient inference speed. This choice is particularly beneficial for scenarios involving limited datasets, constrained computational resources, and applications demanding real-time performance. Our research concentrated on several strategies to bolster YOLO's robustness, specifically focusing on data augmentation techniques, architectural improvements, and optimization methods.

III. EXPERIMENTAL SYSTEM

A. HUMAN PARTICIPANTS

The study involved 126 individuals diagnosed with type 2 diabetes mellitus and 76 healthy controls for acquiring nail-fold capillary images. The diabetic cohort consisted of participants with verified diabetes histories sourced from Osaka University Hospital, while the control group was drawn from Chonnam National University. Both sets of images were captured under identical experimental conditions. Notably, the two groups had no significant body mass index (BMI) discrepancies. All participants, out of their own volition, received a comprehensive briefing regarding the study's objectives and methodologies, after which they provided informed consent to participate voluntarily, a testament to their respect for research.

The study protocol adhered to the ethical principles outlined in the Declaration of Helsinki and received approval from the relevant ethics committee (No. 18546-6, December 07, 2021).

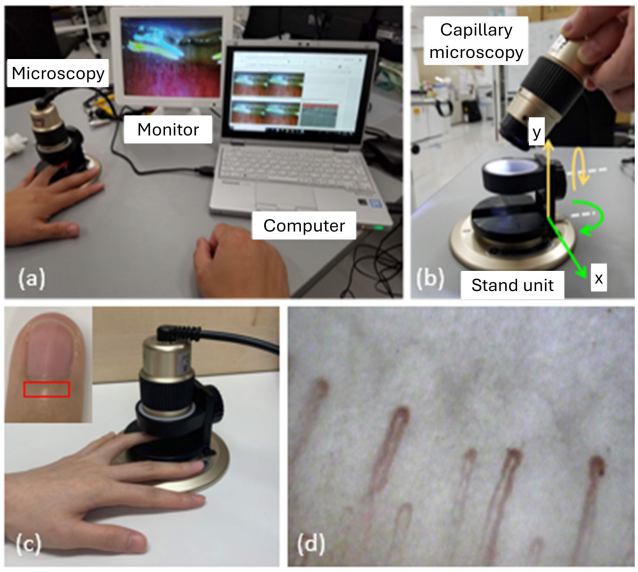

FIGURE 1. A description of an experimental system and environment for recording nail-fold capillaries: (a) represents the total system configuration, (b) represents the detail stand for regulating the microscope, (c) represents how to record capillaries, and (d) represents the recorded results of a healthy participant.

TABLE 1. The microscope hardware specifications.

Feature	Description
Camera Unit	
Magnification	100x to 410x on 14-inch monitor
Dimension	ϕ45 (external diameter) × 96 (length) mm without attachment ϕ45 (external diameter) × 106 (length) mm with attachment
Weight	140g with attachment Attachment (Focus cap): 10g
Stand Unit	
Dimension	Base unit: ϕ120 mm Camera holder unit: ϕ58 mm, height: 72.5 mm
Overall height (with Camera Unit)	132 mm, except protrusion
Weight	250g
Range of the movement	10 mm in a crosswise direction; 10 mm vertically

B. EXPERIMENTAL EQUIPMENT

Figure 1(a) illustrates the schematic of the experimental setup, comprising a microscope, a monitor, and a computer for real-time image visualization and analysis. Capillary images were obtained using the GOKO BSCAN-Z (GOKO Imaging Devices, Kanagawa, Japan), a compact vertical cylindrical microscopy system noted for its enhanced imaging capabilities (see Table 1 for specifications). The

GOKO BSCAN-Z features a robust zoom function that allows for high-resolution imaging without the need for lens interchange. This function, coupled with a distinct black central aperture, ensures clear and detailed images and provides a smooth live video capture devoid of time lags. As shown in Figure 1(b), the microscopy holder mounted on the stand allows for 10 mm vertical movement along the y-axis, facilitating precise adjustments between upper and lower positions. The finger stand, a crucial component, is designed for 10 mm of lateral movement along the x-axis. This lateral movement allows for optimal positioning of the finger from left to right, which is essential for enhancing image capture. For z-axis adjustments, participants can modulate their finger positioning to achieve clear views of the capillaries. The system is complemented by a 14-inch monitor, delivering real-time visualization of capillary images across a spectrum of magnifications, from low to high levels [36].

To mitigate the impact of temperature on capillary diameter measurements, participants were acclimatized for 15∼20 minutes in a temperature-controlled environment maintained at 22∼25 °C before assessment. The evaluation encompassed eight fingers, excluding the thumbs, from each hand. Each finger underwent a cleansing procedure, followed by applying immersion oil in the nail fold region to improve image resolution and clarity. As illustrated in Figure 1(c), participants were seated comfortably, with their fingers on a designated stand beneath the capillary microscope. The red rectangle in the figure delineates the nail-fold region of interest. The researcher ensured optimal alignment for real-time observation on computer display, capturing images for subsequent analysis. Figure 1(d) presents a representative nail-fold capillary image acquired at a magnification of 390×.

C. DATASET CONSTRUCTION

Maldonado et al. [37] have identified a higher prevalence of specific capillary morphologies in patients with diabetes mellitus, including tortuosity, bushy capillaries, capillary ectasia, neoformation, microhemorrhage, bizarre capillaries, and aneurysms. Furthermore, the work of Abd El-Khalik et al. [38] suggests that this population's most prevalent capillary types are tortuous and branched capillaries and precapillary oedema. Additionally, Shikama et al. [39] established a significant positive correlation between the frequency of crossing capillaries in the nail fold and the risk of type 2 diabetes mellitus. Building on these findings, our study aims to identify and classify four distinct morphologies of nail-fold capillaries into four classes: ''hairpin'', ''tortuous'', ''crossing'', and ''bushy''.

- Hairpin capillaries are characterized by capillaries that resemble hairpins with a U-shape [40].
- Tortuous capillaries have multiple curves and curled but do not cross over themselves [40].
- Crossing capillaries: the limbs cross once, twice, or many [39].
- Bushy capillaries are characterized by small and multiple buds along the limb branches [40].

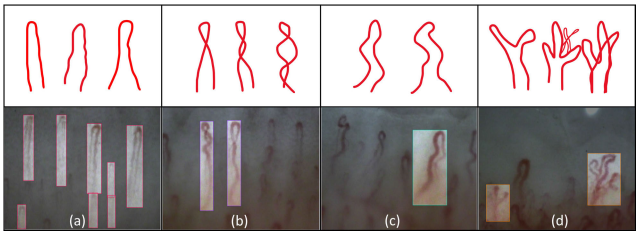

FIGURE 2. Schematic illustrations and corresponding images of four distinct classes of nail-fold capillary morphologies: (a) hairpin capillaries, (b) crossing capillaries, (c) tortuous capillaries, (d) bushy and branching capillaries.

Figure 2 depicts four distinct classes of nail-fold capillaries through schematic illustrations and corresponding images: (a) hairpin capillaries, (b) crossing capillaries, (c) tortuous capillaries, and (d) bushy and branching capillaries. The dataset comprises 3,283 images from diabetic patients (an average of 26 images per human participant) and 3,412 images from healthy controls (an average of 44.89 images per participant), all captured at 390× magnification with a resolution of 640 × 480. After filtering out duplicates and low-quality images, a refined subset of 1,279 images was selected for subsequent analysis. This dataset was partitioned into an 80% training set and a 20% validation set to facilitate model development and evaluation. The capillary images were manually labeled using the modified Openlabelling tool [41], categorizing the images into four classes: ''hairpin'', ''crossing'', ''torturous'', and ''bushy''. Random brightness and contrast adjustments were employed to augment the training set, doubling the training dataset to 2,046 images. A separate test set was created, consisting of 19 multi-scale images acquired at various magnifications ranging from 100× to 400×.

IV. METHODS

A. BASELINE MODEL

YOLOv8 [42] represents a state-of-the-art single-shot object detection framework acclaimed for its optimal balance of speed and accuracy, which is critical in inference tasks. At its core, YOLOv8 employs CSPDarknet53 as the backbone architecture, a sophisticated deep neural network that adeptly extracts features at multiple resolutions through a progressive downsampling approach. A notable advancement in YOLOv8 is its C2f module represented in figure 5, which improves upon the C3 block of YOLOv5 by seamlessly fusing high-level semantic features with low-level spatial context [43]. This sophisticated integration significantly boosts the model's detection performance, particularly for small objects, making it especially relevant for applications such as nail-fold capillaroscopy analysis.

In the neck component, YOLOv8 employs an advanced fusion mechanism called FPN-PAN. This mechanism is significant as it integrates the architecture of the Feature

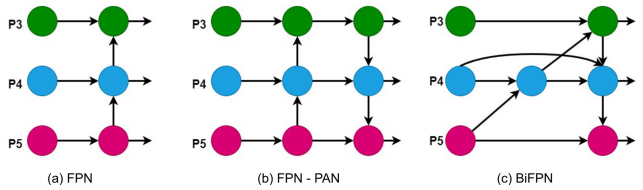

FIGURE 3. The typical three kinds of feature fusion structures: (a) Feature Pyramid Network (FPN), (b) Feature Pyramid Network and Path Aggregation Networks (FPN - PAN), (c) Bi-directional Feature Pyramid Network (BiFPN).

Pyramid Network (FPN) and Path Aggregation Network (PAN), creating multi-scale feature pyramids. These pyramids are crucial as they allow for a more effective fusion of features derived from various depths of the backbone network, resulting in enhanced feature representations that are crucial for object detection tasks.

YOLOv8 architecture features a multi-head detection system, where each head produces predictions for bounding boxes, confidence scores, and class probabilities across various scales in the input image. The model employs binary cross-entropy loss (BCE Loss) for the classification component. For the regression aspect, it utilizes a hybrid of distribution focal loss (DFL) and Complete Intersection over Union (CIoU) loss, significantly boosting performance, especially for detecting smaller objects.

B. FEATURE NETWORKS

The Feature Pyramid Network (FPN) effectively integrates multi-scale image features to improve object detection accuracy. As depicted in Figure 3(a), FPN [44] architecture utilizes input features from layers 3 to 5, represented as $\vec{P_{in}} = (P_3^{in}, P_4^{in}, P_5^{in})$. By leveraging these varying levels of feature maps, the network enhances the representation of objects at different scales, facilitating improved detection performance across a range of object sizes.

For an input size of 640 × 640, the feature levels $(P_3^{in}, P_4^{in}, P_5^{in})$ correspond to resolutions of 80 × 80, 40 × 40, and 20 × 20, respectively. In the FPN context, multi-scale features are synthesized via a top-down pathway, allowing information integration across various resolutions.

$$\begin{aligned} P_5^{out} &= Conv(P_5^{in}) \\ P_4^{out} &= Conv(P_4^{in} + Resize(P_5^{out})) \\ P_3^{out} &= Conv(P_3^{in} + Resize(P_4^{out})) \end{aligned} \tag{1}$$

FPN presents a practical method for feature fusion, yet its unidirectional information flow constrains it from higher to lower levels. To mitigate this limitation, FPN-PAN [45] implements an additional bottom-up pathway, as illustrated in Figure 3(b). This "PAN" mechanism employs downsampling, similar to the upsampling operation (Resize) depicted in Equation 1 to seamlessly integrate lower-level features into the higher levels, thereby enhancing the comprehensive exchange of information.

The BiFPN [22] further advances this concept by incorporating a direct connection from the original input feature map to the output nodes at each level, facilitating the fusion of an expanded range of features. This adjustment enriches feature representation by incorporating low-level details directly into the final predictions, as shown in Figure 3(c). Additionally, the BiFPN optimizes the network architecture by removing nodes with only a single input edge, streamlining the overall design.

C. CHANNEL ATTENTION NETWORKS

Attention networks have significantly enhanced the efficacy of deep convolutional neural networks (CNNs) in various computer vision applications. They enable CNNs to prioritize critical information in inputs while effectively reducing the incorporation of irrelevant background data. Over the years, many innovative attention mechanisms have emerged in the field. In our study, we employ several attention modules, namely the Convolutional Block Attention Module (CBAM) [46] and the Global Attention Mechanism (GAM) [47], to benchmark against our chosen mechanism, the Efficient Channel Attention (ECA) [48]. We choose ECA due to its utilization of efficient 1D convolution, which strikes a balance between computational efficiency and performance enhancement, all while adding minimal parameters and incurring negligible overhead. This characteristic makes ECA particularly advantageous for real-time applications, thereby significantly boosting the overall performance of our model [48].

The ECA mechanism [48] presents an innovative strategy for facilitating channel-wise feature interactions without necessitating dimensionality reduction while maintaining a low model complexity. As illustrated in Figure 4, ECA employs global average pooling to create a comprehensive channel descriptor that effectively encapsulates the overall feature distribution across channels. The module further captures local cross-channel interactions by assessing each channel's k proximal channels. The channel weights are dynamically computed via a 1D convolution operation, where the kernel size k is adaptively determined based on the input channel dimension (C). This kernel size k plays a critical role in defining the scope of local cross-channel interaction within ECA. The representation of ECA can be formulated as follows:

$$\omega = \sigma(Conv_k^{1d}(y)) \tag{2}$$

where $Conv_k^{1d}$ indicated the 1D convolution utilized a kernel size k, y is the aggregated feature, and a sigmoid activation function (σ).

To effectively determine the coverage area, the authors propose making the kernel size k for the 1D convolution adaptively proportional to the number of channels C. Consequently, the local region of coverage can be dynamically calculated as follows:

$$k = \psi(C) = |\frac{log_2(C)}{\gamma} + \frac{b}{\gamma}|_{odd} \tag{3}$$

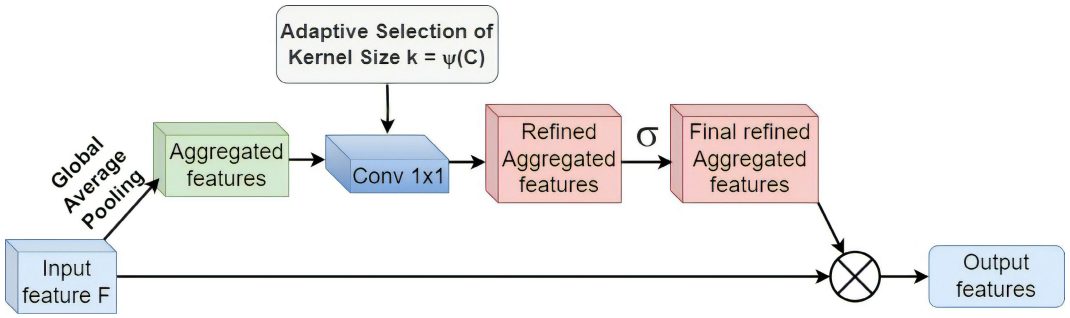

FIGURE 4. The architecture of Efficient Channel Attention (ECA).

where C is the channel dimension, $C = \phi(k) = 2^{(\gamma * k - b)}$, $|k|_{odd}$ represents the nearest odd number of k, and the hyper-parameters b and γ are configured as 2 and 1, respectively.

D. ENHANCED BASELINE MODEL FOR OBJECT DETECTION

Our proposed model significantly enhances YOLOv8 architecture, as Figure 5 depicts. The modifications we've made to the original framework, enclosed within the red dashed boxes, and the additional connection represented by the bold black line, are crucial advancements in object detection. Detecting small objects accurately is a formidable challenge in YOLOv8n architecture. This difficulty stems from two key factors: the limited feature representation in deep feature maps due to the small size of the objects and the significant downsampling employed.

Our model's original architecture, utilizing an input size of 640 × 640 and defining a minimum detection scale of 80 × 80, presented a challenge for objects with bounding boxes smaller than this threshold. To address this issue and improve various-sized object detection capabilities, we incorporated an additional input from the second level of the backbone network (C2f layer 2) into the neck component, establishing a detection layer with a 160 × 160 scale, referred to as the **p2** detection layer, as depicted in Figure 5. This modification significantly enhances the model's ability to detect objects across a broader spectrum, with resolutions of 160 × 160, 80 × 80, 40 × 40, and 20 × 20.

We incorporated the Efficient Channel Attention (ECA) modules after C2f layers to selectively enhance the most relevant channels in an image, thereby augmenting feature representation. ECA stands out for its computational efficiency and ease of implementation relative to other attention mechanisms. To assess the efficacy of ECA, we performed an ablation study, contrasting its performance with that of various alternative attention methods.

E. HYPER-PARAMETER TUNING

Table 2 outlines the hyper-parameter configurations utilized during our model training process. We made noteworthy adjustments to the default values of the mosaic and mixup data augmentation hyper-parameters introducing more variability and complexity into the training data. Figure 6 illustrates the mosaic data augmentation and mixup data augmentation in the nail-fold capillary dataset. The mosaic augmentation technique integrates four images into a single training instance by stitching them together and applying a random cutout, resulting in the final Mosaic image [49], [50]. Conversely, the mixup generates a novel image by averaging two images, guided by a weighting parameter, while linearly blending their respective labels to form a corresponding pair [50], [51]. These augmentation strategies allow the model to engage with a more diverse set of nail-fold capillary scales and classes, spatial positions, and arrangements, which enhances its detection capabilities for smaller objects, boosts generalization to previously unseen data, and counteract overfitting risks. This, in turn, contributes to improved model robustness, a key aspect of our model's performance. In our implementation, the values of mosaic and mixup hyper-parameters were optimized by Grid Search. The mosaic data augmentation factor was reduced to 0.5 (default: 1.0). At the same time, the mixup was set to 0.5 (default: 0.0). These values are the probability of applying these data augmentations in the training process ranging from 0.0 to 1.0. These modifications explore the potential advantages of fine-tuning these specific hyper-parameters relative to the default settings applied in YOLOv8n.

TABLE 2. Used settings for hyper-parameter tuning.

Hyper-parametric configuration	Description
batch	8
epochs	300
mosaic	**0.5** (default: 1.0)
mix-up	**0.5** (default: 0.0)
pretrained	yolov8n

F. SLICING AIDED HYPER INFERENCE

Slicing Aided Hyper Inference (SAHI) [52] presents an innovative methodology to enhance small object detection within computer vision. This technique utilizes a robust framework that employs image slicing, segmenting the input into overlapping patches to improve small objects' visibility and detection accuracy. The model ensures that small objects

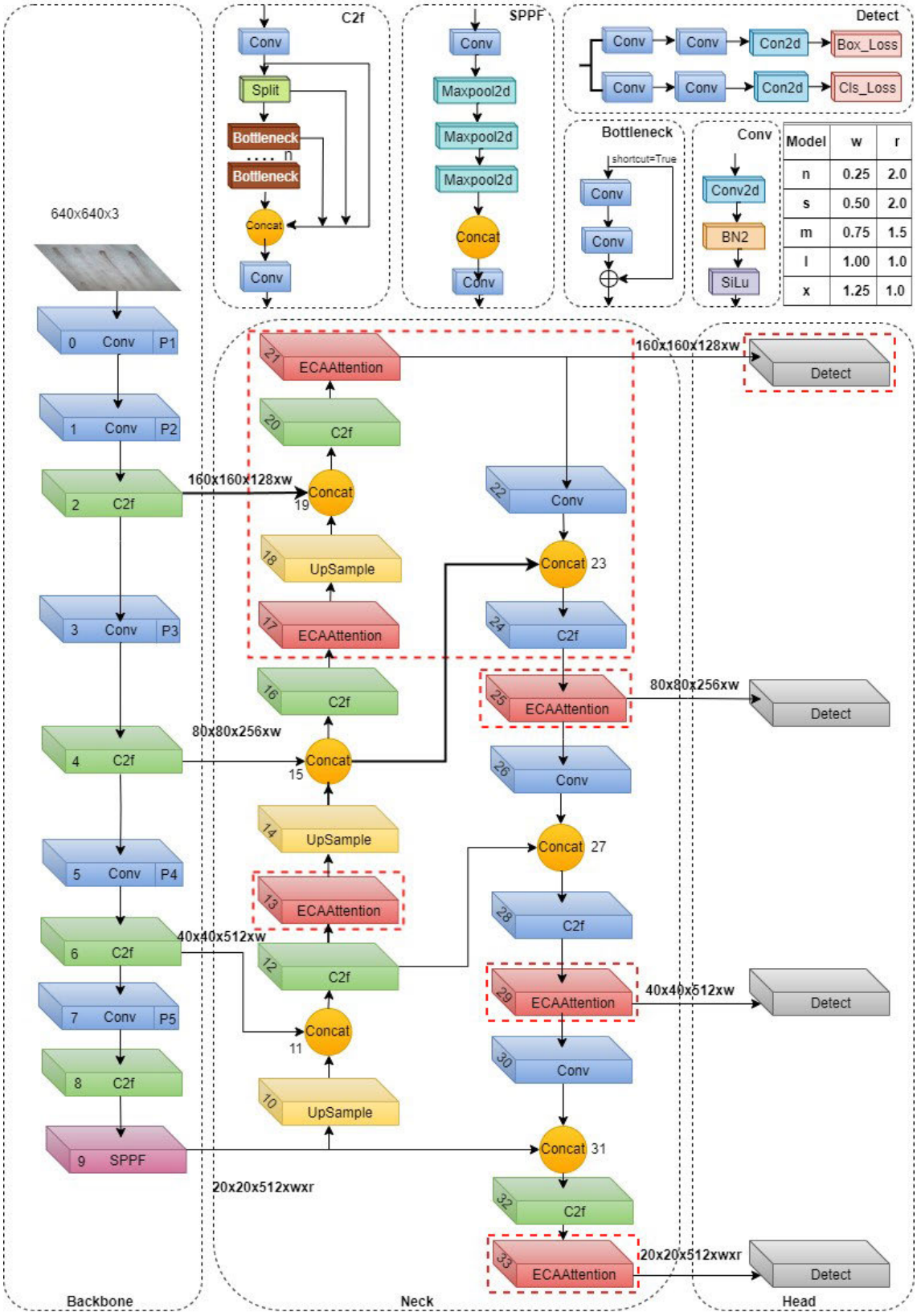

FIGURE 5. Our enhanced YOLOv8 architecture for nail-fold capillary detection.

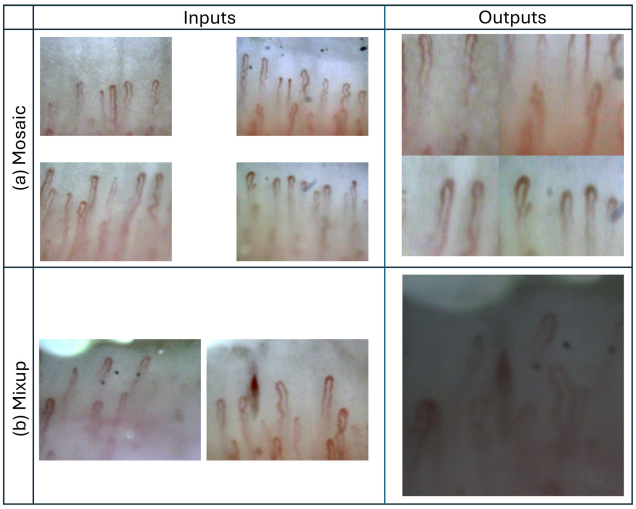

FIGURE 6. The illustration of the mosaic data augmentation and mixup data augmentation in the nail-fold capillary dataset.

are adequately represented within the neural network's processing pipeline.

To specifically address the challenge of detecting various-sized capillaries at lower magnifications, such as 100× or 200×, we introduce a solution that synergizes our enhanced YOLOv8n model optimized through training on a custom dataset with the SAHI framework. By harnessing the high-performance attributes of the refined YOLOv8n architecture in conjunction with SAHI's targeted approach to tiny object detection, we achieve significant advancements in the detection capabilities for minute capillaries.

G. PERFORMANCE METRICS

In this study, several metrics were used to evaluate model performance, including Precision, Recall, and mean Average Precision (mAP), which were calculated using True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN) [53]. GFLOPS, or giga floating-point operations per second, is a measure of a model's computational complexity, with a higher FLOPS count indicating increased model complexity, more processing steps, and longer computation time. Frames per second (FPS) refers to the frame rate, representing the number of individual images a model can process each second. Low FPS can negatively impact the model's ability to perform real-time tracking effectively.

$$FPS = \frac{1}{inference\_time} \tag{4}$$

where *inference_time* is the total time taken by a model to process one image.

Precision is the proportion of true positives among all positive predictions, evaluating the model's ability to minimize false positives:

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

Recall is defined as the proportion of true positives among all actual positives, measuring the model's ability to identify all instances of a class:

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

Intersection over Union (IoU) quantifies the overlap between a predicted bounding box and the ground truth bounding box.

Average Precision (AP) calculates the area under the precision-recall curve, providing a single value that summarizes the model's precision and recall performance.

Mean Average Precision (mAP) is mean average precision.

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{7}$$

where AP_i is the AP in the ith class and N is the number of classes. mAP@50 calculated at an IoU threshold of 0.50 across all classes. An analogous metric is the mean Average Precision between thresholds 0.5 and 0.95 (mAP@50-95). It calculates the average precision across a range of IoU thresholds from 0.5 to 0.95, offering a more balanced and comprehensive assessment of the model's performance across varying levels of detection difficulty.

Mean absolute percentage error (MAPE) is a performance metric for evaluating the precision of prediction quantities relative to actual quantities.

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{8}$$

where y_i is the actual value, $\hat{y}_i$ is the prediction value and n is the number of fitted points.

V. RESULTS

A. RESULTS OF ABLATION STUDY

All experiments were meticulously conducted on an NVIDIA GeForce RTX 4090 GPU with a driver version 550.54.14 and 24GB of memory. The deep learning framework utilized for the experiments is PyTorch 2.1.1, with CUDA 11.8 running on Python 3.8.18, ensuring the highest standards of accuracy and reliability in our results.

Table 3 presents the outcomes of ablation studies that assessed the effects of various feature fusion techniques—specifically FPN-PAN, BiFPN, and FPN—on the performance of the YOLOv8 model in the context of nail-fold capillaroscopy image analysis. The FPN-PAN configuration demonstrated the most substantial performance gains, achieving a mean Average Precision at an Intersection over the Union (IoU) threshold of 50% (mAP@50) of 76.2%, alongside the highest precision rate at 70.2%. These outcomes underscore the significant potential of FPN-PAN in image analysis. Moreover, it maintained the lowest parameter count compared to the other methodologies explored. While BiFPN excelled in recall, recording an actual identification rate of 74.4%, FPN-PAN distinguished itself by delivering

TABLE 3. The result of ablation experiments with feature fusion methods.

Model	Precision	Recall	mAP@50	mAP@50 - 95	Parameters	GFLOPS
YOLOv8n + FPN-PAN	**0.702**	0.7	**0.762**	**0.437**	**3006428**	**8.1**
YOLOv8n + BiFPN	0.674	**0.744**	0.757	0.434	3022812	**8.1**
YOLOv8n + FPN	0.649	0.743	0.757	0.436	4663516	19.0

TABLE 4. The result of ablation experiments with attention modules.

Model	Precision	Recall	mAP@50	mAP@50 - 95	Parameters	GFLOPS
YOLOv8n	0.709	**0.714**	0.762	0.435	3006428	**8.1**
YOLOv8n + CBAM	0.704	0.713	0.76	0.435	3109796	8.2
YOLOv8n + GAM	**0.719**	0.702	0.762	0.432	3687708	9.5
YOLOv8n + ECA	0.682	0.712	**0.764**	**0.436**	3006440	**8.1**

TABLE 5. The result of ablation experiments with our proposed robustness methods. "Aug" means data augmentation.

Model	p2	ECA	Aug	Hyper-parameter	Precision	Recall	mAP@50	mAP@50-95	Parameter	GFLOPS
YOLOv8n					0.709	0.714	0.762	0.435	3006428	**8.1**
YOLOv8n	✓				0.716	0.726	0.776	0.434	2921568	12.2
YOLOv8n		✓			0.682	0.712	0.764	0.436	3006440	**8.1**
YOLOv8n			✓		0.693	0.695	0.741	0.433	3006428	**8.1**
YOLOv8n				✓	0.703	0.743	0.77	0.449	3006428	**8.1**
YOLOv8n		✓	✓		0.693	0.735	0.774	0.447	3011624	8.2
YOLOv8n		✓	✓	✓	0.728	**0.744**	0.786	0.45	3006440	**8.1**
YOLOv8n	✓	✓	✓	✓	**0.745**	0.739	**0.799**	**0.463**	**2668266**	9.3

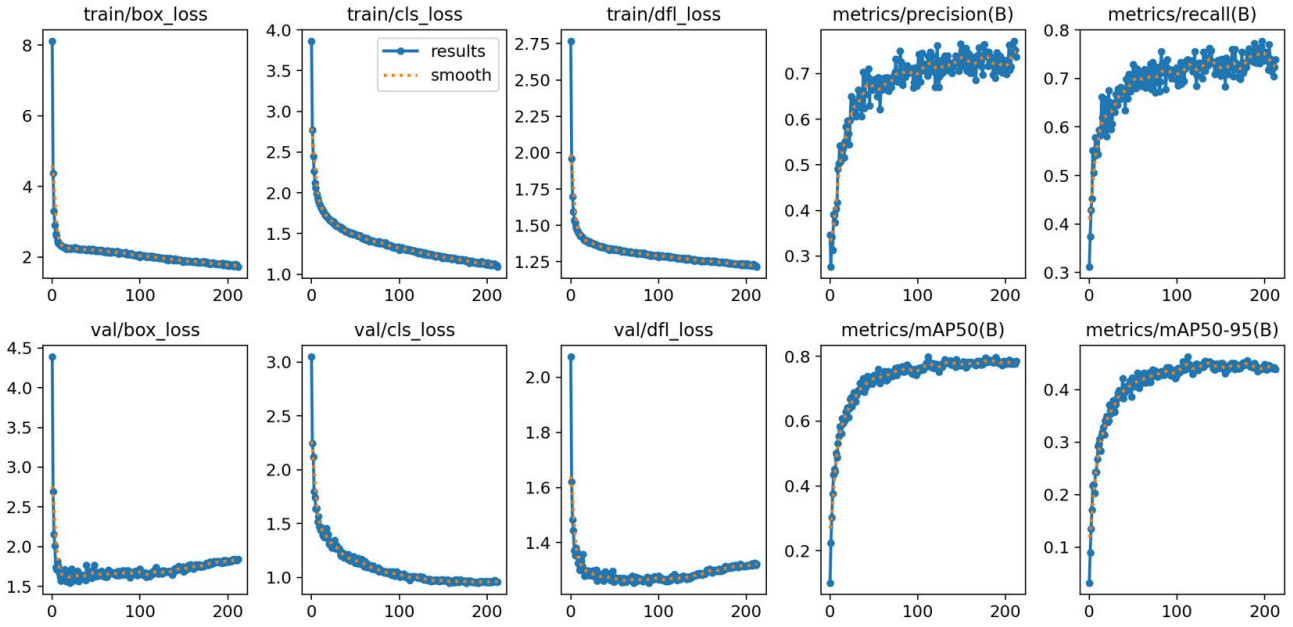

FIGURE 7. The training and validation curves for various performance metrics: box loss, distribution focal loss, classification loss, precision, recall, and mean average precision (mAP).

a more comprehensive balance of mAP@50, precision, and parameter efficiency.

Table 4 summarizes the results of ablation studies assessing the effectiveness of different attention modules—CBAM, GAM, and ECA—when integrated into the YOLOv8 framework. The ECA module demonstrated superior performance, enhancing 0.2% without significantly increasing the number of parameters, adding only 12 to the model. This minimal parameter increase underscores the scalability of the ECA module. Additionally, it maintains the GFLOPS efficiency comparable to YOLOv8n. These findings highlight the efficacy of ECA attention in enhancing model performance while keeping the complexity low.

Table 5 displays the ablation studies' outcomes on our proposed robust models, which integrate an additional p2 detection layer, ECA attention, data augmentation, and

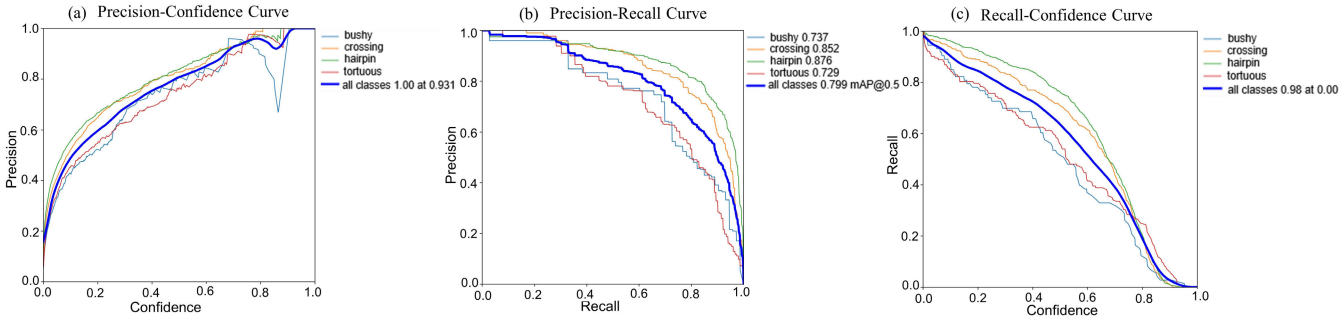

FIGURE 8. (a) The Precision-Confidence Curve, (b) Precision-Recall Curve, and (c) Recall-Confidence Curve of our proposed model in four classes: bushy, crossing, hairpin, tortuous, and all classes.

hyperparameter optimization. Incorporating the p2 detection layer, ECA attention mechanism, and hyperparameter tuning yielded performance enhancements in nail fold capillary detection, resulting in increases of 1.4%, 0.2%, and 0.8% in mAP@50, respectively. In contrast, data augmentation decreased YOLOv8n's detection accuracy by 2.1% in mAP@50 for this specific application. This decrease can be attributed to the introduction of noisy or irrelevant data during the augmentation process, which the model struggles to effectively utilize. Notably, combining data augmentation with either ECA attention or the p2 detection layer significantly improved detection robustness. For instance, the "YOLOv8n + ECA + Aug" configuration demonstrated a 1% improvement in mAP@50 over "YOLOv8n + ECA" without data augmentation. This indicates that the baseline YOLOv8n model struggles to leverage augmented data effectively, whereas our refined YOLOv8n model adapts proficiently, resulting in enhanced generalization capabilities with the integration of data augmentation.

The results from the "YOLOv8n + p2" configuration underscore the effectiveness of the p2 detection layer, with substantial enhancements of 0.7%, 1.2%, and 1.4% in Precision, Recall, and mAP@50, respectively. The integration of the p2 detection layer has also led to a 2.8% decrease in the total parameter count, a testament to the balance achieved between model complexity and parameter reduction. This balance is facilitated by the p2 head functioning at a higher resolution with reduced channel usage, effectively managing the trade-off between granularity and model complexity. The reduction in total parameters is a clear indication of this architectural modification's success in maintaining model efficiency while enhancing performance.

Our enhanced YOLOv8 model incorporates several key advancements: the introduction of an additional detection layer, implementation of Efficient Channel Attention (ECA), various data augmentation techniques, and finely-tuned hyperparameters. These enhancements lead to considerable performance gains over the baseline YOLOv8n model, with precision, recall, mAP@50, and mAP@50-95 improving by 3.6%, 2.5%, 3.7%, and 2.8%, respectively. Notably, these improvements are achieved with a reduced parameter count—11.3% lower than YOLOv8n—thanks to the efficient utilization of 1D convolutions in the ECA block and model optimization. This optimization was carefully balanced, ensuring efficacy for small object detection by maintaining a low computational complexity of 9.3 GFLOPS, demonstrating superior efficiency compared to the 1.2 GFLOPS of YOLOv8n.

The training and validation loss curves for box, classification, and distribution focal loss are illustrated in Figure 7, along with precision, recall, and mean Average Precision (mAP) metrics across the training epochs. The box loss quantifies the error in bounding box predictions, while the classification loss assesses the accuracy of object category predictions. It's important to note that the distribution focal loss plays a critical role in addressing class imbalance by emphasizing the training of difficult-to-classify instances. Notably, after 200 epochs, the model exhibits a significant improvement in precision, recall, and mAP, achieving stabilization by epoch 217, at which point training was terminated.

In Figure 8, we present the precision-recall metrics for our proposed model. Figure 8(a) illustrates the precision-confidence curve, where maximum precision reaches one at a confidence threshold of 0.931. Figure 8(b) displays the precision-recall curve, highlighting a mean average precision (mAP) of 0.799 across all classes at an IoU threshold set to 0.5. The individual class mAPs are as follows: bushy (0.737), crossing (0.852), hairpin (0.876), and tortuous (0.729). Lastly, Figure 8(c) shows the recall curve, providing additional insights into the model's performance across different confidence levels.

B. RESULTS OF COMPARATIVE ANALYSIS

Figure 9 illustrates a qualitative comparison of detection outcomes between the baseline YOLOv8n and the enhanced YOLOv8n model under varying conditions: (a) multiple capillaries, (b) image blur, (c) backlight scenarios, and (d) 100× magnification. The displayed results include bounding boxes, class labels, and confidence scores for capillary detection, where the confidence score reflects the model's certainty in its classifications.

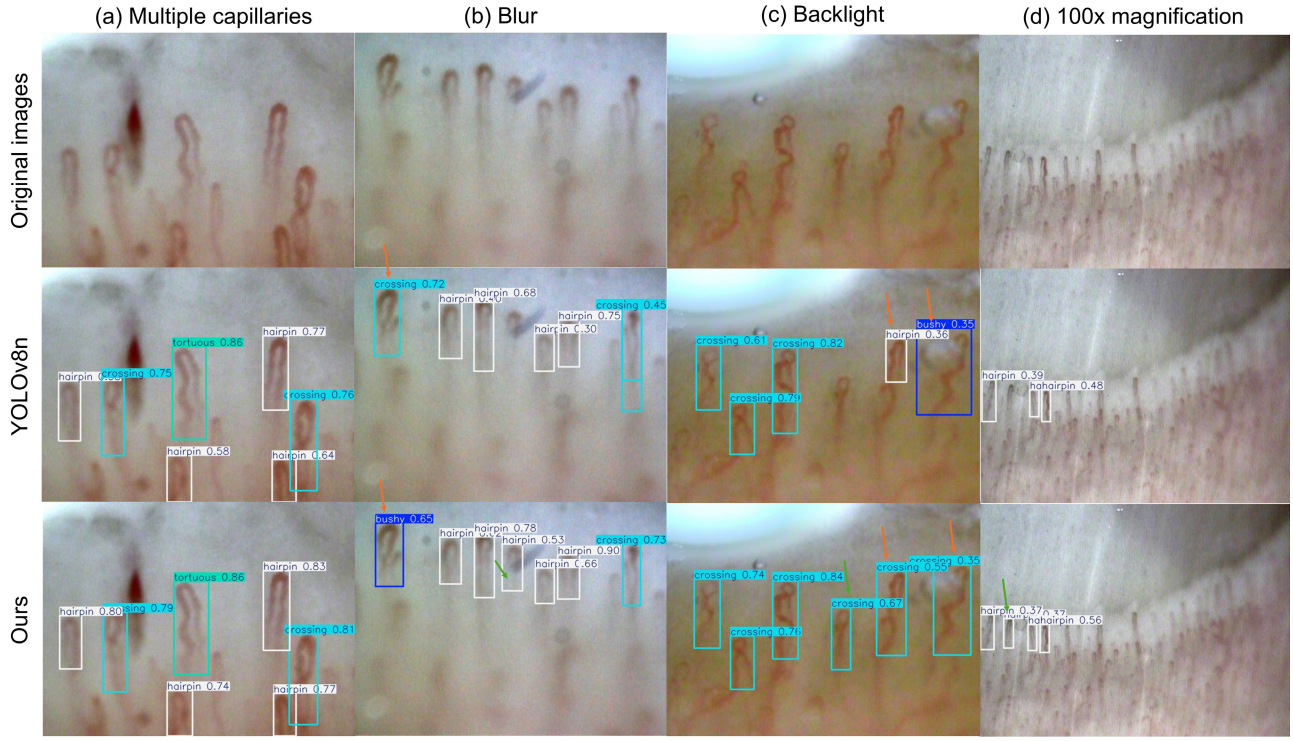

FIGURE 9. The comparison of the qualitative results between the original YOLOv8n and our improved YOLOv8n in various conditions: (a) Multiple capillaries, (b) Blur image, (c) Backlight, and (d) 100x magnification.

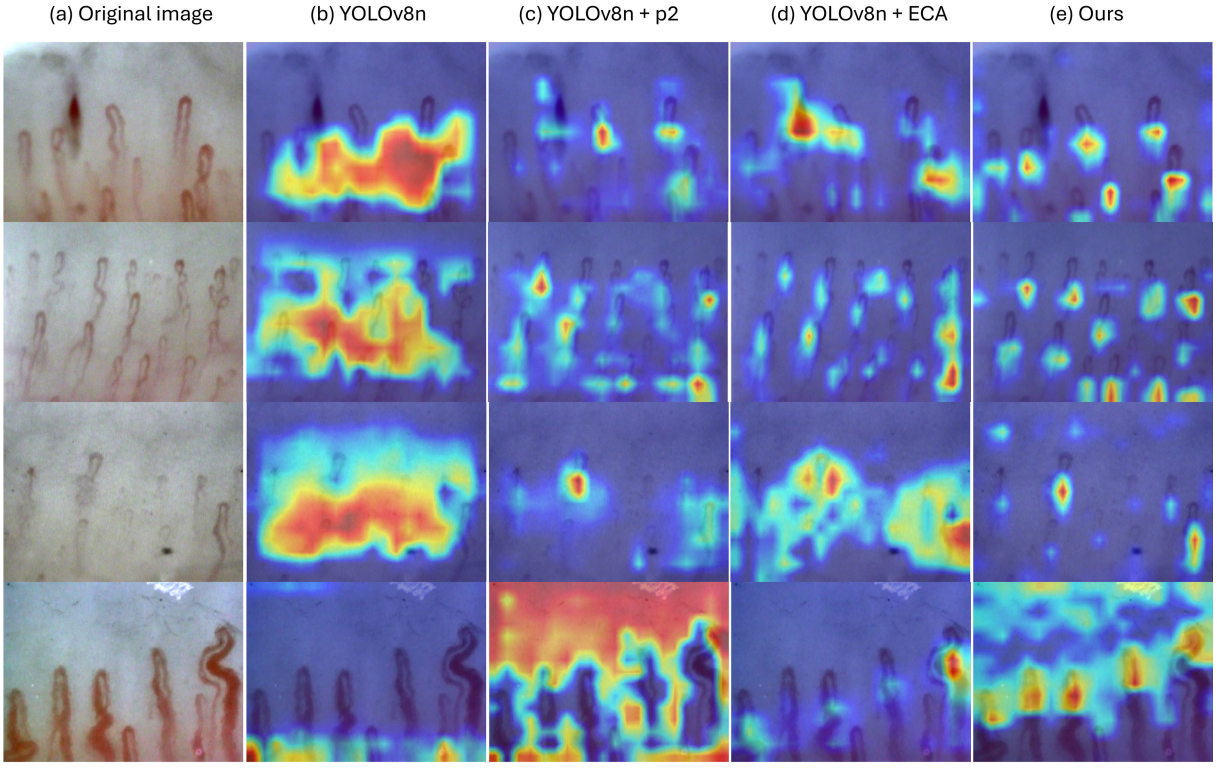

FIGURE 10. The comparison of heatmaps generated with Eigen-CAM: (a) Original image, (b) YOLOv8n, (c) YOLOv8n + p2, (d) YOLOv8n + ECA and (e) Our improved YOLOv8n.

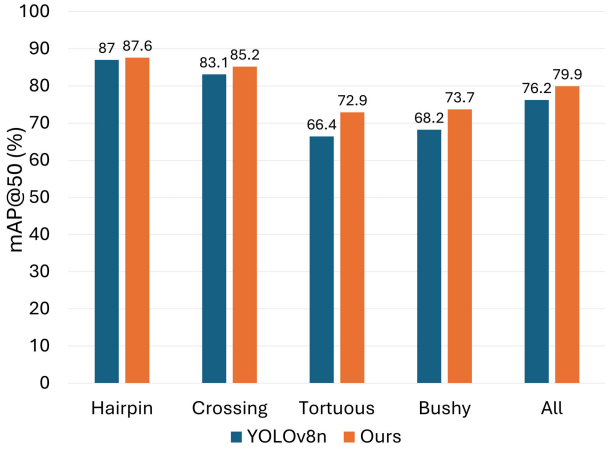

FIGURE 11. Category mAP@50 (%) comparison between YOLOv8n and our improved YOLOv8.

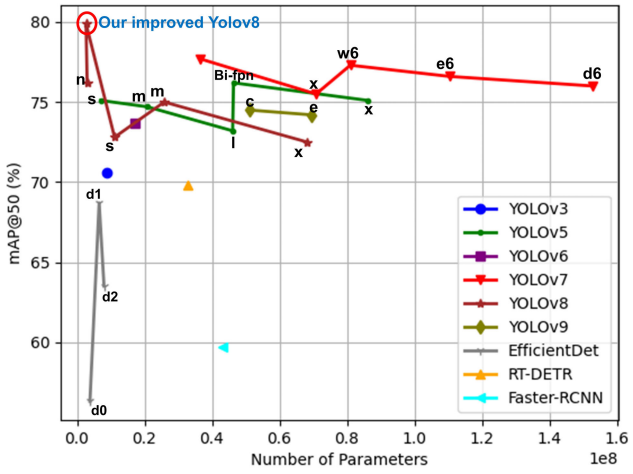

FIGURE 12. The performance comparison of object detection models trained on Nailfold capillary dataset.

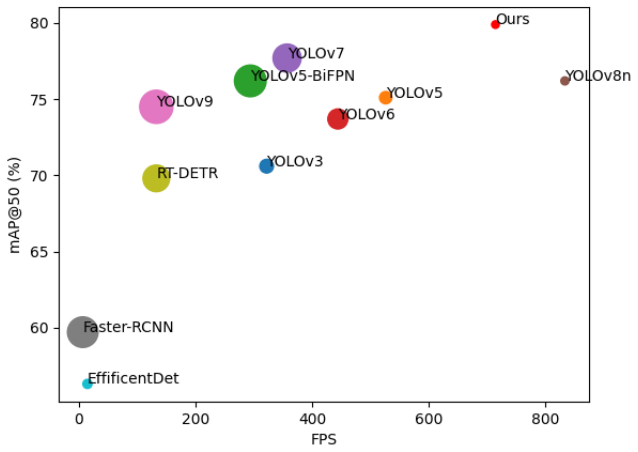

FIGURE 13. Nailfold capillary detection model performance comparison, FPS: frame rate, mAP: mean Average Precision. The radius of the circle represents the number of model parameters. The larger the radius, the greater the number of model parameters.

The improved model demonstrates a superior performance across all tested conditions, typically yielding higher confidence scores for capillary identification than the original YOLOv8n. Notably, the orange arrows in the blur and backlight conditions indicate instances of misclassification by the original YOLOv8n, which our improved model accurately identified, providing reassurance about its accuracy. Moreover, the green arrows highlight capillary detection achieved by our model that the original YOLOv8n missed in the blur, backlight, and 100× magnification contexts.

Figure 10 presents an Eigen-CAM visualization that elucidates model predictions across five sample images, specifically contrasting (b) the original YOLOv8n, (c) YOLOv8n with an added p2 detection layer, (d) YOLOv8n equipped with ECA, and (e) our enhanced version of YOLOv8n. The heatmaps, derived from Eigen-CAM utilizing class activation maps (CAM), delineate the focal regions within the visual data the model prioritizes for its predictions, thus providing critical insights into the model's attention distribution and confidence levels.

For this visualization, the second-to-last layer was selected to avoid interference from the detection output of the final layer in YOLOv8. The heatmaps highlight that the brighter regions signify areas of intense attention during inference. Notably, the models incorporating the p2 detection layer, the ECA, and our improved YOLOv8n exhibit a more refined focus on nail fold capillaries. They also effectively downplay attention to irrelevant background elements, reassuring the audience about their efficiency. In stark contrast, the original YOLOv8n tends to distribute its attention more evenly between capillaries and adjacent backgrounds, which could lead to unnecessary computational complexity and diminished accuracy. Our improved YOLOv8n, however, achieves superior precision in targeting capillary regions, including smaller capillaries, demonstrating its efficacy in this specific context.

Figure 11 illustrates the enhanced accuracy across various object categories relative to the baseline YOLOv8 model. Notably, there has been a noticeable and consistent increase in mAP@50 for all classes since implementing our refined model. The most significant improvements are observed in the "tortuous" and "bushy" categories, which experienced gains of 6.5% and 5.5%, respectively. These categories, which represent a more specialized subset of capillaries within the dataset, are the specific focus of our research. Substantial advancements in detection performance for these features are a testament to our efforts.

Figure 12 presents a comparative analysis of multiple object detection models applied to our nail-fold capillaroscopy dataset, a crucial resource in studying microvascular abnormalities in various diseases. This includes YOLOv3 [54], as referenced in paper [12], and various iterations of YOLOv5 [43] (specifically, the small, medium, large, and

extra-large versions), along with Bi-FPN from paper [31]. We also evaluate YOLOv6 [55], YOLOv7 [56] (covering versions YOLOv7, x, w6, e6, d6), YOLOv8 [42] (including versions nano, small, medium, and extra-large), and a modified iteration introduced in this work. Additionally, we assess YOLOv9 [57] (versions c and e), EfficientDet [22] (d0, d1, d2, and d3-d7, noting that latter versions encountered memory issues), along with the Transformer-based Real-Time Detection Transformer (RT-DETR) [25] and Faster R-CNN [23].

In practical applications, models are typically optimized to enhance both accuracy and inference speed, as rapid inference is essential for effective deployment. Figure 13 illustrates the frames per second (FPS) performance across several object detection architectures, including the YOLO series, Faster R-CNN, EfficientDet, the transformer-based RT-DETR, and our enhanced YOLOv8 model. In this visualization, the radius of each circle indicates the model's parameter count, with larger circles representing greater parameter complexity. Our model, with the lowest parameter count, delivers the best mAP@50, with 714 FPS ranking just behind YOLOv8n (833 FPS). This optimal interplay between efficiency and accuracy underscores its superiority, making it a compelling choice for real-time object detection applications.

In figures 12 and 13, the nail-fold capillary detection performance of RT-DETR, Faster R-CNN, and EfficientDet is observed to be inferior to that of the YOLO family, with these models also demonstrating longer inference times. Consequently, they are not well-suited for this real-time application. The suboptimal performance of Faster R-CNN can be attributed to its limited ability to detect objects in low-resolution images, primarily due to its weaker capacity to capture local textures. In contrast, EfficientDet demands substantial memory for training and exhibits prolonged inference durations. While RT-DETR offers speed improvements over DETR, its suitability for small datasets remains limited. Whereas, YOLO-based models (from YOLOv3 through YOLOv9) demonstrate exceptional accuracy, with our optimized YOLOv8n model outperforming all alternatives even with a reduced parameter count. These findings emphasize the model's efficacy for real-time analysis in nail-fold capillaroscopy applications. In summary, our research indicates that the YOLOv8n model is the most suitable for our specific application, offering a balance of accuracy and efficiency.

VI. DISCUSSION

We enhance the detection accuracy of tiny capillaries by integrating the model with the SAHI (Slicing Aided Hyper Inference) framework. The comparative analysis of the improved YOLOv8n model's performance, and without SAHI, is presented in Table 6. After training our improved model, optimal weights are saved in the best.pt file, and the SAHI technique is integrated for subsequent inference tasks.

Figure 14 illustrates the model's performance across multi-scale images at 100×, 200×, and 390× magnifications, both with and without SAHI integration. Notably, at the highest magnification (390×), the detection outcomes of the standalone and SAHI-integrated models exhibit minimal differences. However, SAHI significantly enhances the detection of tiny capillaries at lower magnifications (100× and 200×), resulting in a 67.32% increase in detection volume and a 35.6% reduction in Mean Absolute Percentage Error (MAPE). Importantly, integrating SAHI reduces inference time, achieving up to 50 frames per second (FPS), making it not just suitable but highly practical for real-time applications.

These findings underscore that SAHI is particularly advantageous for low-magnification images, improving small capillary detection. In contrast, the enhanced YOLOv8n model alone is adequate for high-magnification (390×) scenarios. Notably, the figures represent performance on untrained multi-scale images at magnifications of 100×, 200×, and 390×, which deviated from the training dataset's magnification of 390×.

In this study, we employed several techniques to enhance the robustness of the YOLOv8n model, achieving significant improvements in detecting nail-fold capillaries. The year 2024 marked the introduction of YOLOv11, the latest iteration in the YOLO series of object detection models, which redefined the state-of-the-art with its superior accuracy, speed, and efficiency [58]. Table 7 provides a comparative analysis of the results obtained by applying our robustness techniques to YOLOv11. Compared to YOLOv8, YOLOv11 demonstrates advantages in model simplicity and computational speed, requiring fewer parameters. Specifically, YOLOv11n reduces the number of parameters and GFLOPS by 14% and 22%, respectively, compared to YOLOv8n. However, the mAP@50 of YOLOv11n is lower than that of YOLOv8n on the nail-fold capillaries dataset. When integrating our proposed techniques into YOLOv11n, we observed improvements of 1.1%, 1.6%, 2.9%, and 3% in Precision, Recall, mAP@50, and mAP@50-95, respectively, while achieving a 6.3% reduction in the number of parameters. Despite these advancements, the combination of our proposed methods with YOLOv8 outperforms YOLOv11, with gains of 3.6%, 0.3%, 1.6%, and 0.4% in Precision, Recall, mAP@50, and mAP@50-95, respectively. However, YOLOv11 demonstrates its strengths in speed and model complexity, making it a valuable addition to the YOLO family.

In real-world applications, privacy and security concerns often limit the number of samples that can be collected and labeled for each class, sometimes to just a few. Techniques such as small-sample learning, zero-shot learning, and the development of domain-specific large models offer promising solutions to address these practical challenges [59], [60]. Manually annotating the boundaries of blood vessels presents significant challenges in practice, which requires

TABLE 6. The result of comparison between our improved YOLOv8n with and without SAHI on test samples.

Model	Inference time (ms)	FPS	Model detection	Actual capillaries	MAPE(%)	Classification accuracy (%)
Ours	**1.4**	714	101	191	47.1	95
Ours + SAHI	20	50	**169**	191	**11.5**	**97**

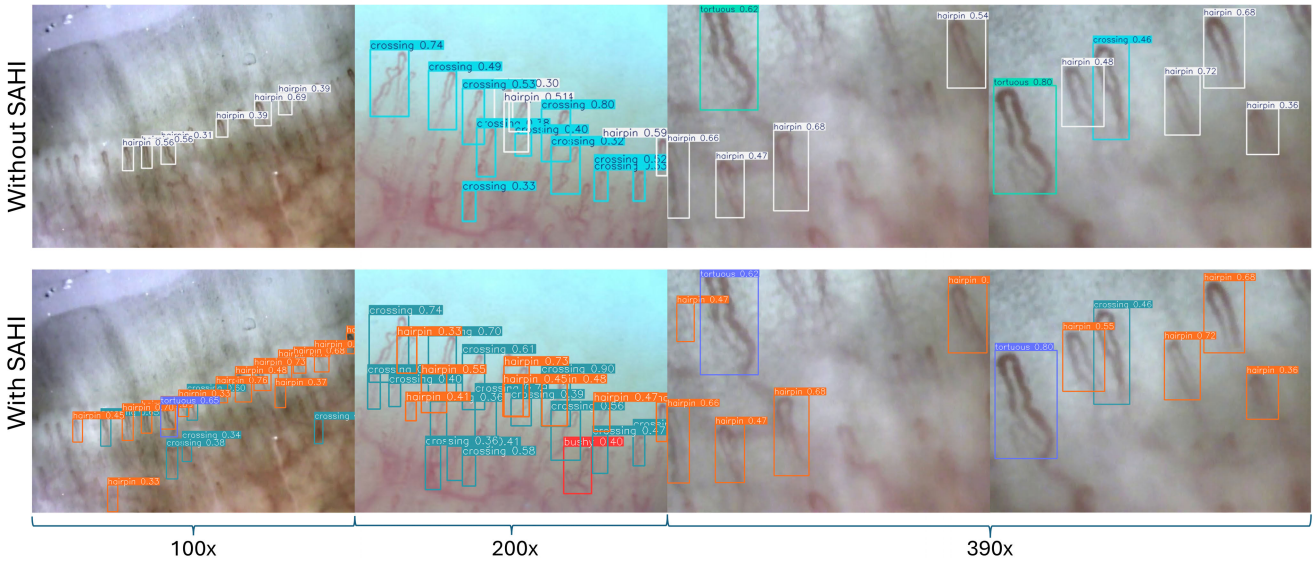

FIGURE 14. The results of multi-scale untrained images of our model with SAHI and without SAHI.

TABLE 7. The result of our techniques implementation for model robustness in YOLOv11.

Model	Precision	Recall	mAP@50	mAP@50-95	Parameter	GFLOPS
YOLOv8n	0.709	0.714	0.762	0.435	3006428	8.1
YOLOv11n	0.698	0.72	0.754	0.429	2582932	**6.3**
Our proposed techniques + YOLOv11n	0.709	0.736	0.783	0.459	**2419290**	7.5
Ours	**0.745**	**0.739**	**0.799**	**0.463**	2668266	9.3

collaboration between clinical experts and AI engineers. One potential solution to address this difficulty is using high-resolution images to improve the accuracy of manual annotations. Additionally, post-processing techniques can be applied to filter out unnecessary detections. To ensure robust evaluation, metrics such as Average Precision (AP) at varying object sizes (e.g., medium, small) can be employed to capture performance across different sizes of objects better.

VII. CONCLUSION

This study presents an enhanced YOLOv8 model designed to analyze non-invasive nail-fold capillaroscopy images. We made essential modifications to optimize small capillary detection, including increasing detection layers in the model's head, integrating Efficient Channel Attention (ECA) modules, fine-tuning hyper-parameters, and applying advanced data augmentation techniques. These enhancements result in significant performance gains over the baseline YOLOv8n model, with improvements of 3.6% in precision, 2.5% in recall, 3.7% in mAP@50, and 2.8% in mAP@50-95. Notably, the model maintains a low parameter count and computational efficiency, making it suitable for real-time applications.

Additionally, we investigate the combined utilization of Slicing-Aided-Hyper-Inference (SAHI) techniques to address the challenges associated with small object detection in nail-fold capillary images captured at low magnification. These techniques, which consider the scale of objects in the image, lead to significant performance enhancements, particularly in untrained and multi-magnification scenarios involving minute capillaries. The results suggest the potential development of a clinical tool that aids healthcare professionals in detecting and assessing nail-fold capillary morphology, which is crucial for counting capillary numbers and serves as an essential indicator in the early diagnosis of various diseases. Future research will aim to broaden the model's applicability to diabetic patient detection, further illustrating its potential clinical significance.

ACKNOWLEDGMENT

The authors would like to express a huge thank you to the participants who joined their experiments.

REFERENCES

[1] G. A. Scardina and P. Messina, "Microvascular abnormalities in patients with rheumatoid arthritis," *Ann. Anatomy—Anatomischer Anzeiger*, vol. 188, no. 5, pp. 425–429, Sep. 2006, doi: 10.1016/j.aanat.2006.04.004.

[2] D. Saygin, K. B. Highland, and A. R. Tonelli, "Microvascular involvement in systemic sclerosis and systemic lupus erythematosus," *Microcirculation*, vol. 26, no. 3, Apr. 2019, Art. no. e12440, doi: 10.1111/micc.12440.

[3] Z. Barth, B. N. Witczak, B. Flatø, A. Koller, I. Sjaastad, and H. Sanner, "Assessment of microvascular abnormalities by nailfold capillaroscopy in juvenile dermatomyositis after medium-to long-term followup," *Arthritis Care Res.*, vol. 70, no. 5, pp. 768–776, May 2018, doi: 10.1002/acr.23338.

[4] M. Cojocaru, I. M. Cojocaru, I. Silosi, and C. D. Vrabie, "Modified open labelling," *Maedica*, vol. 6, no. 4, p. 330, Oct. 2011.

[5] P. Anyfanti, E. Angeloudi, A. Dara, A. Arvanitaki, E. Bekiari, G. D. Kitas, and T. Dimitroulas, "Nailfold videocapillaroscopy for the evaluation of peripheral microangiopathy in rheumatoid arthritis," *Life*, vol. 12, no. 8, p. 1167, Jul. 2022, doi: 10.3390/life12081167.

[6] P.-C. Hsu, P.-Y. Liao, H.-H. Chang, J. Y. Chiang, Y.-C. Huang, and L.-C. Lo, "Nailfold capillary abnormalities are associated with type 2 diabetes progression and correlated with peripheral neuropathy," *Medicine*, vol. 95, no. 52, p. e5714, Dec. 2016, doi: 10.1097/md.0000000000005714.

[7] D. Rennie, "Nailfold dermatoscopy in general practice," *Austral. Family Physician*, vol. 44, no. 11, pp. 809–812, Nov. 2015.

[8] World Health Org. (WHO). (2024). *Diabetes*. [Online]. Available: https://www.who.int/europe/news-room/fact-sheets/item/diabetes

[9] World Health Org. (WHO). (2023). *Global Report on Diabetes*. [Online]. Available: https://iris.who.int/bitstream/handle/10665/204871/9789241565257_eng.pdf

[10] A. G. Tabák, C. Herder, W. Rathmann, E. J. Brunner, and M. Kivimäki, "Prediabetes: A high-risk state for diabetes development," *Lancet*, vol. 379, no. 9833, pp. 2279–2290, Jun. 2012.

[11] American Diabetes Association, "Diagnosis and classification of diabetes mellitus," *Diabetes Care*, vol. 33, pp. S62–S69, Jan. 2010, doi: 10.2337/dc10-S062.

[12] K. V. Suma, S. Selvi, P. Nanda, M. Shetty, M. Vikas, and K. Awasthi, "Deep learning approach to nailfold capillaroscopy based diabetes mellitus detection," *Int. J. Online Biomed. Eng. (iJOE)*, vol. 18, no. 6, pp. 95–109, May 2022.

[13] G. Maldonado, A. Chacko, R. Lichtenberg, M. Ionescu, and C. Rios, "Nailfold capillaroscopy in diabetes mellitus: A case of neo-angiogenesis after achieving normoglycemia," *Oxford Med. Case Rep.*, vol. 2022, no. 9, Sep. 2022, Art. no. omac088.

[14] A. Rajaei, P. Dehghan, and A. Amiri, "Nailfold capillaroscopy in 430 patients with rheumatoid arthritis," *Caspian J. Internal Med.*, vol. 8, no. 4, pp. 269–274, Jan. 2017.

[15] B. I. Chazan, M. C. Balodimos, R. L. Lavine, and L. Koncz, "Capillaries of the nailfold of the toe in diabetes mellitus," *Microvascular Res.*, vol. 2, no. 4, pp. 504–507, Oct. 1970.

[16] A. D. Pokorny, R. B. Mefferd Jr., and G. M. Stevens, "Nailfold capillary plexus and familial schizophrenia," *J. Nervous Mental Disease*, vol. 150, no. 2, pp. 146–154, Feb. 1970.

[17] J. Tian, Y. Xie, M. Li, J. Oatts, Y. Han, Y. Yang, Y. Shi, Y. Sun, J. Sang, K. Cao, C. Xin, L. Siloka, H. Wang, and N. Wang, "The relationship between nailfold microcirculation and retinal microcirculation in healthy subjects," *Frontiers Physiol.*, vol. 11, p. 880, Jul. 2020.

[18] K. Bogusz-Górna, A. Polańska, A. Dańczak-Pazdrowska, R. Żaba, M. Sumińska, P. Fichna, and A. Kędzia, "Non-invasive detection of early microvascular changes in juveniles with type 1 diabetes," *Cardiovascular Diabetol.*, vol. 22, no. 1, p. 285, Oct. 2023.

[19] M. Abdalla Helmy Mohamed Abdou, T. Tuyen Truong, E. Jul, and P. Ferreira, "Deep learning and computer vision techniques for microcirculation analysis: A review," 2022, *arXiv:2205.05493*.

[20] S. U. Demir, R. Hakimzadeh, R. H. Hargraves, K. R. Ward, E. V. Myer, and K. Najarian, "An automated method for analysis of microcirculation videos for accurate assessment of tissue perfusion," *BMC Med. Imag.*, vol. 12, no. 1, pp. 1–3, Dec. 2012.

[21] A. Banitalebi-Dehkordi, A. Amirkhani, and A. Mohammadinasab, "EBCDet: Energy-based curriculum for robust domain adaptive object detection," *IEEE Access*, vol. 11, pp. 77810–77825, 2023, doi: 10.1109/ACCESS.2023.3298369.

[22] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 10778–10787.

[23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[24] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.* Glasgow, U.K.: Springer, Jan. 2020, pp. 213–229.

[25] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "DETRs beat YOLOs on real-time object detection," 2023, *arXiv:2304.08069*.

[26] H. Zhao, X. Peng, S. Wang, J.-B. Li, J.-S. Pan, X. Su, and X. Liu, "Improved object detection method for unmanned driving based on transformers," *Frontiers Neurorobotics*, vol. 18, May 2024, Art. no. 1342126.

[27] M. Hussain, "YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection," *Machines*, vol. 11, no. 7, p. 677, Jun. 2023.

[28] L. Tan, T. Huangfu, L. Wu, and W. Chen, "Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification," *BMC Med. Informat. Decis. Making*, vol. 21, no. 1, pp. 1–11, Dec. 2021.

[29] B. Karbouj, G. A. Topalian-Rivas, and J. Krüger, "Comparative performance evaluation of one-stage and two-stage object detectors for screw head detection and classification in disassembly processes," *Proc. CIRP*, vol. 122, pp. 527–532, Jan. 2024.

[30] H. Huang, B. Wang, J. Xiao, and T. Zhu, "Improved small-object detection using YOLOv8: A comparative study," *Appl. Comput. Eng.*, vol. 41, no. 1, pp. 80–88, Feb. 2024.

[31] H. T. P. Nguyen, S. Ko, and H. Jeong, "Deep-learning-based capillary detection," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2023, pp. 4932–4934.

[32] Y. Zhao, "Improvement of the robustness of deep learning models against adversarial attacks," *Appl. Comput. Eng.*, vol. 75, no. 1, pp. 285–289, Jul. 2024.

[33] A. Amirkhani, A. Khosravian, M. Masih-Tehrani, and H. Kashiani, "Robust semantic segmentation with multi-teacher knowledge distillation," *IEEE Access*, vol. 9, pp. 119049–119066, 2021, doi: 10.1109/ACCESS.2021.3107841.

[34] H. Huang, Y. Wang, S. M. Erfani, Q. Gu, J. Bailey, and X. Ma, "Exploring architectural ingredients of adversarially robust deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2021, pp. 5545–5559.

[35] H. Ichiwara, H. Ito, K. Yamamoto, H. Mori, and T. Ogata, "Spatial attention point network for deep-learning-based robust autonomous robot motion generation," 2021, *arXiv:2103.01598*.

[36] Goko Imag. Devices. *Imaging Products*. Accessed: Oct. 11, 2023. [Online]. Available: https://www.gokocamera.com/english/ev/

[37] G. Maldonado, R. Guerrero, C. Paredes, and C. Ríos, "Nailfold capillaroscopy in diabetes mellitus," *Microvascular Res.*, vol. 112, pp. 41–46, Jul. 2017.

[38] D. M. Abd El-Khalik, E. A. Hafez, H. E. Hassan, A. E. Mahmoud, D. M. Ashour, and N. A. Morshedy, "Nail folds capillaries abnormalities associated with type 2 diabetes mellitus progression and correlation with diabetic retinopathy," *Clin. Med. Insights: Endocrinol. Diabetes*, vol. 15, Sep. 2022, Art. no. 11795514221122828.

[39] M. Shikama, N. Sonoda, A. Morimoto, S. Suga, T. Tajima, J. Kozawa, N. Maeda, M. Otsuki, T. Matsuoka, I. Shimomura, and Y. Ohno, "Association of crossing capillaries in the finger nailfold with diabetic retinopathy in type 2 diabetes mellitus," *J. Diabetes Invest.*, vol. 12, no. 6, pp. 1007–1014, Jun. 2021.

[40] C. Grover, D. Jakhar, A. Mishra, and A. Singal, "Nail-fold capillaroscopy for the dermatologists," *Indian J. Dermatol., Venereol. Leprol.*, vol. 88, no. 3, pp. 300–312, Oct. 2021.

[41] I. Goncharov. Accessed: Jun. 26, 2023. [Online]. Available: https://github.com/ivangrov/ModifiedOpenLabelling

[42] G. Jocher, A. Chaurasia, and J. Qiu. (2023). *Ultralytics YOLO (Version 8.0.0)*. [Online]. Available: https://github.com/ultralytics/ultralytics

[43] G. Jocher et al., "Ultralytics/YOLOv5: V7. 0-YOLO5 sota real-time instance segmentation," Zenodo, Nov. 2022, doi: 10.5281/zenodo.7347926.

[44] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 936–944.

[45] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 8759–8768.
[46] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Munich, Germany, Sep. 2018, pp. 3–19.
[47] Y. Liu, Z. Shao, and N. Hoffmann, "Global attention mechanism: Retain information to enhance channel-spatial interactions," 2021, *arXiv:2112.05561*.
[48] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 11531–11539.
[49] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[50] N. Crasto, "Class imbalance in object detection: An experimental diagnosis and study of mitigation strategies," 2024, *arXiv:2403.07113*.
[51] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," 2017, *arXiv:1710.09412*.
[52] F. C. Akyon, S. Onur Altinuc, and A. Temizel, "Slicing aided hyper inference and fine-tuning for small object detection," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Bordeaux, France, Oct. 2022, pp. 966–970.
[53] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Niteri, Brazil, Jul. 2020, pp. 237–242.
[54] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[55] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, Y. Li, B. Zhang, Y. Liang, L. Zhou, X. Xu, X. Chu, X. Wei, and X. Wei, "YOLOv6: A single-stage object detection framework for industrial applications," 2022, *arXiv:2209.02976*.
[56] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Vancouver, BC, Canada, Jun. 2023, pp. 7464–7475.
[57] C.-Y. Wang, I.-H. Yeh, and H.-Y. Mark Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," 2024, *arXiv:2402.13616*.
[58] Ultralytics. (2024). *Ultralytics Yolov11*. Accessed: Nov. 20, 2024. [Online]. Available: https://docs.ultralytics.com/models/yolo11/s
[59] X. Hao, S. Yang, R. Liu, Z. Feng, T. Peng, and B. Huang, "VSLM: Virtual signal large model for few-shot wideband signal detection and recognition," *IEEE Trans. Wireless Commun.*, early access, Nov. 19, 2024, doi: 10.1109/TWC.2024.3496813.
[60] C. Li, K. Luo, L. Yang, S. Li, H. Wang, X. Zhang, and Z. Liao, "A zero-shot fault detection method for UAV sensors based on a novel CVAE-GAN model," *IEEE Sensors J.*, vol. 24, no. 14, pp. 23239–23254, Jul. 15, 2024, doi: 10.1109/JSEN.2024.3405630.

HANG THI PHUONG NGUYEN (Member, IEEE) received the degree from the Department of Information Systems, Hanoi University of Science and Technology, Vietnam, in 2018, and the M.S. degree in artificial intelligence convergence from Chonnam National University, in 2022, where she is currently pursuing the Ph.D. degree with the Department of Artificial Intelligence Convergence. After that, she was with Viettel Business Solutions Corporation as a Software Engineer (full-time). In September 2020, she received the master's-course scholarship at Chonnam National University. Her research interests include image processing, human posture, and healthcare. She received awards from international and domestic conferences, such as the Excellent Paper in ICAII 2021 and the Best Student Paper Award in IW-FCV 2021.

HIEYONG JEONG (Member, IEEE) received the Ph.D. degree in mechanical engineering from Osaka University, Japan, in 2009. From April 2009 to November 2013, he was a Senior Research Engineer (full-time) with Samsung Heavy Industries Company Ltd., Daejeon, Republic of Korea. From November 2014 to August 2019, he was with the Graduate School of Medicine, Osaka University, as an Associate Professor (full-time). Since September 2019, he has been operating as a Professor (full-time) with the Department of Artificial Intelligence Convergence, Chonnam National University, Republic of Korea. His research interests include human–robot interaction, human posture, and healthcare. He received awards from international and domestic conferences, such as the Excellent Paper in ICAIIC 2021, the Best Student Paper Award in IW-FCV 2021, and the Best Poster Presentation Award in IEEE LifeTech 2019.