CHAPTER I
INTRODUCTION

Chapter I provides the foundational context, direction, and scope of this research. The discussion encompasses the background, problem formulation, research objectives, significance of the study, and scope limitations. Each subsection is organized systematically in alignment with the topic under investigation, a comparative study of the CAS-UNet architecture incorporating Cross Fusion Channel Attention and Coordinate Attention mechanisms for retinal blood vessel segmentation.

1.1 Background

The retina is a thin membranous layer located at the posterior segment of the eye, housing a complex vascular network. This layer functions as the site of phototransduction, converting incident light into visual signals processed by the brain, such that even subtle structural changes in the retinal vasculature may serve as early indicators of visual impairment [1][2]. Morphological alterations in retinal blood vessels are well established as consequences of various ocular and systemic pathologies [3]. Among these, diabetic retinopathy, a microvascular complication of diabetes mellitus, involves progressive damage to retinal capillaries due to uncontrolled blood glucose levels, which may ultimately lead to irreversible blindness if not detected and managed in its early stages [4]. According to the World Health Organization (WHO), diabetic retinopathy accounts for approximately 4.8% of blindness cases among the estimated 39 million blind individuals worldwide [5].

Fundus imaging has consequently become indispensable for the early detection of vascular structural changes. It enables non-invasive visualization of the complete retinal architecture, including the vasculature, macula, and optic disc. Manual analysis of fundus images by ophthalmologists remains the clinical standard for retinal disease diagnosis; however, this approach is inherently limited by diagnostic subjectivity dependent on individual expertise, substantial time requirements, particularly in high-volume clinical settings, and inequitable geographic access to specialist care [2][6]. These limitations have been a primary driver for the development of automated fundus image analysis based on machine learning.

Substantial evidence indicates that vascular structural alterations including arteriolar widening, narrowing, or increased vessel density, function as early indicators of disease progression. Chakraborty et al. (2023) demonstrated that a unified deep learning framework for the simultaneous segmentation of blood vessels, exudates, and microaneurysms achieves disease classification accuracy of 97.4% [6]. Furthermore, quantitative vascular features such as vessel density and tortuosity exhibit significant correlation with diabetic retinopathy severity ($r = 0.82$), establishing vascular morphology as a valuable biomarker for early-stage diagnosis [7]. These findings confirm that image segmentation is a fundamental step in separating vascular structures from the retinal background, enabling precise analysis of vascular abnormalities. The primary challenges in retinal vessel segmentation arise from the thin, finely branching vascular architecture, its low contrast against background tissue, and susceptibility to noise and uneven illumination [8].

Driven by advances in deep learning, U-Net, a CNN-based encoder-decoder architecture equipped with skip connections has become one of the most widely adopted frameworks for medical image segmentation, as it simultaneously preserves fine spatial detail and extracts deep hierarchical feature representations [6]. Its symmetric U-shaped structure, formed by successive downsampling and upsampling pathways, enables effective multi-scale feature learning [9]. Although U-Net has proven effective in image segmentation tasks [4][11], its relatively shallow encoder constrains its capacity to model complex image patterns [11]. As a result, models frequently struggle to detect small and intricately branching vessels, since fine-grained features are susceptible to loss through repeated downsampling and upsampling operations [3].

One established strategy for improving segmentation performance is the integration of attention mechanisms into the U-Net framework. Such mechanisms enable the model to selectively focus on informative features while suppressing irrelevant background activations, thereby enhancing sensitivity to thin and branching vascular structures. You et al. (2023) proposed CAS-UNet, an improved

U-Net with an attention mechanism, incorporating Cross-Fusion Channel Attention (CFCA), an Additive Attention Gate (AG+), and SoftPool. This architecture achieved accuracy of 96.68% and sensitivity of 83.21% on the CHASE_DB1 dataset, and accuracy of 95.86% with sensitivity of 83.75% on DRIVE, with each dataset evaluated independently [12].

The CFCA module in CAS-UNet enriches feature representations through a cross-channel fusion mechanism. However, the reliance of CFCA, and other conventional channel attention variants, on Global Average Pooling (GAP) to aggregate spatial information into a single scalar per channel poses a risk of discarding critical positional cues. Such cues are essential in retinal vessel segmentation, given the tubular, branching, and geometrically oriented nature of the vascular structures [13].

In contrast, Hou et al. (2021) introduced Coordinate Attention (CA), a mechanism designed to jointly integrate channel and spatial information by factorizing feature aggregation along two independent one-dimensional directions, horizontal and vertical. This approach allows the model to preserve spatial coordinate information while efficiently capturing long-range inter-channel dependencies, making it particularly well-suited for maintaining the geometric continuity and directional orientation of retinal blood vessels.

Despite the demonstrated efficacy of CAS-UNet with CFCA, no systematic study has yet compared the effectiveness of CFCA and CA within the same CAS-UNet architectural framework. This research therefore focuses on a comparative evaluation of CAS-UNet with Cross-Fusion Channel Attention against CAS-UNet with Coordinate Attention for the task of retinal blood vessel segmentation. Experiments are conducted on a combined dataset comprising DRIVE (adult population) and CHASE_DB1 (paediatric population) to enable an objective analysis of performance differences, contribute empirical evidence toward the selection of an optimal attention mechanism, and establish a methodological foundation for developing more efficient and clinically applicable deep learning models.

1.2 Research Questions

Based on the foregoing background, the research questions addressed in this study are formulated as follows:

1. How can the CFCA and CA attention mechanisms be effectively implemented within the CAS-UNet architecture to address the loss of fine structural detail in retinal blood vessel segmentation?
2. How does the performance of the CFCA module which prioritizes inter-channel correlation compare with that of the CA module which integrates spatial positional information in generating an optimal retinal vessel segmentation map on the combined DRIVE and CHASE_DB1 dataset?

1.3 Research Objectives

Research objectives represent the targets and anticipated outcomes of a study. Accordingly, this research aims to:

1. Develop and analyze retinal blood vessel segmentation models based on the CAS-UNet architecture by integrating CFCA and CA attention mechanisms to preserve fine vascular structural detail in retinal images.
2. Identify the most effective attention mechanism for retaining retinal vascular structure by comparatively evaluating the segmentation performance of CAS-UNet with CFCA and CAS-UNet with CA on the combined DRIVE and CHASE_DB1 dataset.

1.4 Research Significance

This research aims to deliver contributions across the following dimensions:

1. To advance knowledge in medical image processing and deep learning by providing a comparative analysis of the influence of CFCA and CA attention mechanisms within the CAS-UNet architecture on retinal vessel segmentation performance, evaluated on a fundus dataset encompassing diverse vascular morphological characteristics.
2. To provide a methodological reference for future research on the selection and evaluation of attention mechanisms in retinal vessel segmentation

models, particularly through the use of a combined dataset representing adult (DRIVE) and paediatric (CHASE_DB1) populations, enabling a more comprehensive assessment of model performance and architectural efficiency.

3. To demonstrate the potential of a more accurate and efficient retinal vessel segmentation system as a component of computer-aided diagnostic workflows. Improved vessel segmentation can assist clinicians in identifying retinal vascular abnormalities earlier and more objectively, thereby supporting the diagnosis of retinal diseases such as diabetic retinopathy.

1.5 Scope and Limitations

To ensure a focused scope of investigation, the following limitations are defined:

1. This study employs two publicly available retinal fundus image datasets: DRIVE and CHASE_DB1.
2. The scope of this research is limited to retinal blood vessel segmentation.
3. Model performance evaluation is restricted to standard segmentation metrics: accuracy, sensitivity, specificity, F1-score, and Intersection over Union (IoU).