*applied sciences*

MDPI

*Article*

# SwinIrisNet: A Hybrid Deep Learning Framework for Robust Iris Segmentation

**Tresor Lisungu Oteko * and Kingsley A. Ogudo**

Department of Electrical and Electronics Engineering Technology, University of Johannesburg, Johannesburg 0524, South Africa; kingsleyo@uj.ac.za

* Correspondence: lisungu.tresor1@gmail.com

**Abstract**

Accurate iris segmentation remains a fundamental challenge in iris biometric recognition and medical image analysis, particularly in challenging scenarios such as non-cooperative acquisition conditions involving variable illumination, partial occlusions, degraded image quality, and diverse unconstrained environments. Prevailing segmentation algorithms exhibit limited robustness when confronted with such challenges, and the disparity between near-infrared (NIR) and visible-light imaging modalities further compounds the complexity of achieving a robust segmentation outcome. To address these challenges, this paper introduces SwinIrisNet, a hybrid deep learning architecture that integrates Swin Transformer and convolutional neural network (CNN) branches within a U-Net framework for robust iris segmentation. The Swin Transformer branch leverages hierarchical window-based self-attention to capture global contextual dependencies, whereas the CNN branch extracts fine-grained local features essential for precise boundary delineation. A memory-efficient cross-attention fusion module combines these complementary feature representations, further enhanced by a Convolutional Block Attention Module (CBAM), Atrous Spatial Pyramid Pooling (ASPP), and attention-gated skip connections for multi-scale context aggregation. An extensive evaluation is conducted across four publicly available benchmark datasets, including UBIRIS.v2, IITD, CASIA-Thousand, and MMU.v1, encompassing both visible-light and NIR imaging environments. The proposed architecture yields F1 values of 0.9612–0.9672, Dice coefficients of 0.9489–0.9519, mIoU values of 0.9266–0.9450, precision values of 0.9565–0.9633, recall values of 0.9600–0.9672, and classification accuracies of 99.51–99.53%, with NICE1 error rates of 0.57–0.60% and NICE2 values of 1.82–2.24%, confirming pixel-level segmentation quality. Cross-database generalization experiments further demonstrate that SwinIrisNet learns transferable iris representations and generalizes effectively across heterogeneous imaging sources, with the strongest transfer occurring in the NIR-to-visible direction. A comparative analysis against existing algorithms demonstrates that the proposed architecture attains substantial performance improvements over several existing segmentation networks when evaluated on identical benchmark databases, surpassing them across the majority of qualitative and quantitative metrics while maintaining a marginally lower memory footprint.

**Keywords:** biometric recognition; hybrid deep learning; iris segmentation; multi-scale feature extraction; Swin Transformer; U-Net architecture

Check for updates

Academic Editors: Martina Iammarino, Chiara Verdone and Stefano Galantucci

Received: 21 July 2026
Revised: 27 August 2026
Accepted: 3 September 2026
Published: 7 September 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

## 1. Introduction

Biometric recognition systems have been extensively deployed across a diverse range of domains, covering security-critical applications, financial services, law enforcement, border control, healthcare, among numerous other applications. Among available biometric modalities, iris recognition is regarded as one of the most reliable due to its exceptional discriminative capacity, stability over time, and inherent resistance to presentation attacks. The iris, with its unique and complex texture patterns, provides a rich source of discriminative features for identity authentication and verification. However, the efficacy of iris recognition systems is critically dependent on accurate segmentation of the iris. Iris segmentation is the process of precisely localizing and extracting the iris region from eye images while excluding adjacent anatomical structures such as eyelids, eyelashes, sclera, and pupils [1]. Beyond its main application in biometric authentication systems, iris segmentation finds practical application in medical image analysis, where it serves as an important preprocessing stage for non-invasive screening and computer-aided diagnosis of ocular diseases, including glaucoma (e.g., angle-closure assessment), cataract, uveitis, and iris tumors or anterior-segment abnormalities. It also enables pupillometry for neurological evaluation, gaze and fatigue monitoring in telemedicine, and secure patient identification in electronic health records (EHRs), supporting reliable and effective smart-healthcare applications [2,3].

Despite significant advances in deep learning-based segmentation methods, iris segmentation remains a challenging task, particularly when dealing with heterogeneous datasets that exhibit variations in image quality, resolution, illumination conditions, occlusions, and acquisition environments. Traditional approaches based on circular Hough transforms or integro-differential operators often fail to offer acceptable results under non-ideal conditions [4–6], while early deep learning methods using standard U-Net architectures with VGG/ResNet backbones [7,8] or Fully Convolutional Networks (FCNs) [9] are inherently limited in their ability to simultaneously capture both global features and fine-grained local details necessary for accurate boundary delineation.

Recent Vision Transformer architectures, particularly Swin Transformer, have demonstrated strong performance in extracting long-range dependencies and global context through hierarchical window-based self-attention mechanisms [10–13]. However, pure Transformer-based approaches often lack the inductive biases and local feature extraction strengths that CNNs inherently possess, properties that are crucial for detecting subtle iris boundaries and handling fine spatial details. This insight motivates the design of hybrid architectures capable of harnessing the complementary strengths of both paradigms [1].

This paper introduces a lightweight hybrid deep learning architecture that unifies Swin Transformer and CNN branches within a U-Net framework, specifically designed for robust iris segmentation across heterogeneous datasets. The main contributions of this paper are summarized as follows:

i. We propose SwinIrisNet, a hybrid U-Net architecture that integrates Swin Transformer and CNN branches for iris segmentation. The architecture leverages Swin Transformers for global context and long-range dependencies via hierarchical window-based self-attention, while the CNN branch captures local features and fine spatial details critical for precise boundary delineation.
ii. We propose a memory-efficient cross-attention fusion mechanism that combines Transformer and CNN feature streams via channel attention, enhanced by the Convolutional Block Attention Module (CBAM). Learnable attention weights adaptively emphasize salient feature representations from both branches, while a memory-optimized configuration enables training on standard hardware without compromising segmentation performance.

iii. Cross-dataset generalization ability: The proposed architecture generalizes reliably across datasets with comparable image acquisition characteristics, including under challenging conditions such as occlusions, illumination variations, and real-world biometric variability, with minimal train–validation disparity and the strongest transfer occurring in the NIR-to-visible direction.
iv. We conduct a comprehensive benchmark evaluation on four different benchmark databases, including IITD, UBIRIS.v2, CASIA-Thousand, and MMU.v1, to establish competitive performance benchmarks for iris segmentation.

The rest of this paper is structured as follows. Section 2 discusses related works on iris segmentation, encompassing traditional methods, deep learning approaches, and hybrid architectures. Section 3 presents the proposed Swin–CNN hybrid architecture, covering the encoder design, fusion mechanism, and decoder structure. Section 4 outlines the experimental configuration, benchmark datasets, and performance evaluation criteria employed. Section 5 presents the experimental results for the proposed model and its performance across four datasets. Section 6 presents a discussion, including comparisons with state-of-the-art methods and an analysis of challenges and limitations. Section 7 concludes this paper and summarizes its key contributions.

## 2. Related Works

This section reviews the related literature on iris segmentation to contextualize the current state of research and identify existing gaps. Iris segmentation is fundamental to the effectiveness of an iris recognition system. It establishes the foundation for all subsequent processes, including feature extraction, normalization, matching, and decision-making. Despite some efforts made in iris segmentation, there are still challenges to address in the area of iris segmentation, especially in complex scenarios (e.g., non-cooperative environments, illumination variations, deepfake scenarios, etc.). Traditional methods [4,14–17] have proven inadequate for such complex scenarios, necessitating the adoption of deep learning approaches for iris segmentation. More research work is required to address the challenges related to noisy iris images, off-angle iris images, occlusions, reflections, iris images captured at a distance, iris images captured under different lighting conditions, etc.

Several studies [4,16–21] have employed traditional techniques for iris segmentation, including Daugman's rubber sheet model, integro-differential operator, Hough transform, and active contour models. However, these techniques face significant limitations that hinder their efficacy in practical applications. These methods are often susceptible to noise and occlusions, including specular reflections, poor lighting conditions, eyelashes, and eyelids, which can affect segmentation performance across variable image quality conditions. Additionally, traditional methods lack robustness and generalization, often failing to adapt to different environments or diverse populations. Moreover, the reliance on manual parameter tuning further limits their practicality, rendering these approaches time-intensive and poorly scalable for deployment in production settings.

With the emergence of deep learning-based techniques, research focus has shifted to deep learning-based algorithms for iris segmentation tasks. Compared with traditional approaches, deep learning algorithms [9,10,22–29] provide significant performance gains by overcoming the limitations outlined above. These methods enhance robustness to noise and occlusions through automatic feature learning and end-to-end training, adapting to various artifacts and partial occlusions. They handle variable image quality challenges better, using adaptive models and data augmentation techniques. Deep learning models also exhibit improved generalization and adaptability, benefiting from large-scale training and transfer learning. Automated hyperparameter optimization simplifies parameter tuning, while

higher accuracy and real-time processing capabilities make deep learning approaches more effective for iris segmentation in practical applications, such as medical image analysis.

### 2.1. Traditional Methods

The ring-shaped integro-differential algorithm proposed by Daugman [30] was one of the first geometry-based iris segmentation techniques, in which the iris region is assumed to be circular, bounded by the pupil on one side and the sclera on the other. This is categorized as a boundary-based category of iris segmentation. The algorithm has been widely used by many researchers for iris segmentation [30,31]. There are two main boundary-based techniques widely used for iris segmentation: the integro-differential operator method and the Hough transform approach [6,16].

The second category of iris segmentation approaches is pixel-based iris segmentation, which focuses on the pixels of the iris region and distinguishes them from non-iris regions [24,32]. These methods first extract distinctive appearance characteristics from the iris's neighborhood of pixels and then establish classifiers for iris pixel categorization. However, pixel classification approaches have proven to be inefficient in challenging scenarios; they are easily affected by the presence of eyelashes and eyelids, which impacts the accuracy of iris segmentation [27].

### 2.2. Deep Learning Based Algorithms

The third category of iris segmentation techniques encompasses machine learning-based iris segmentation techniques, in which deep learning algorithms are used to address challenges that the preceding two methods cannot resolve. Deep learning-based segmentation methods yield markedly improved performance over conventional techniques.

In recent years, CNN-based algorithms have attracted significant research interest, demonstrating consistent superiority over traditional methods in semantic segmentation tasks. Several studies [7,25,27,33] have employed transfer learning by leveraging pre-trained models and modifying only fully connected layers to achieve the desired segmentation outcome. Deep learning-based methods not only offer improved segmentation accuracy but also provide the computational efficiency needed for real-world applications.

The authors of [22] proposed a deep learning-based architecture based on an improved U-Net that produced improved accuracy compared to geometry-based methods [16,17,22]. The authors employed a combination of fully dilated convolution and U-Net, referred to as the FD-UNet architecture. This combination helped improve U-Net performance and produce improved accuracy when trained and tested on three datasets, including the CASIA Iris Interval, UBIRIS.v2, and ND-IRIS-0405 datasets.

Chen et al. 2019 [27] developed an iris segmentation algorithm based on a CNN architecture incorporating dense blocks, termed DFCN. In this architecture, the encoder of the proposed system is composed of dense blocks, while the decoder outputs the predicted masks using a transpose technique. They evaluated the proposed method using three distinctive datasets, namely CASIA Iris Interval-V4, UBIRIS.v2, and IITD. The experimental results showed that their method outperformed most existing methods.

Sardar et al. (2020) [26] proposed a new variant of the U-Net architecture for iris segmentation to handle unconstrained environments, including but not limited to reflection, poor lighting, and occlusions. They also employed squeeze-and-expand modules that helped achieve faster training while enhancing storage efficiency through a reduction in the number of trainable parameters. To evaluate the proposed method, the authors used three datasets: the IITD, NICE I, and CASIA Iris Interval-V4 datasets. The proposed algorithm demonstrated a certain level of robustness and is capable of accurate segmentation on datasets with insufficient annotations.

Khaki et al. [7] presented a deep learning-based iris segmentation approach combining the U-Net and ResNet architectures, called ISUR. It involves two main stages, namely attention mask detection and iris region isolation, both based on distinctive CNNs. To evaluate the proposed method, two datasets, namely UBIRIS and CASIA Iris at Distance, were used. The results showed that ISUR demonstrates improved performance compared to the U-Net-only architecture.

Mathur et al. [28] introduced a joint synthesis framework capable of simultaneously generating eye images and their associated segmentation masks for the sclera, iris, and pupil (SIP) regions. The methodology employs a progressive generative adversarial network to produce synthetic SIP masks, while a Pix2Pix architecture is utilized to reconstruct realistic eye images conditioned on these masks. Experimental results indicate that the synthesized data contributes to measurable gains in segmentation performance.

Malgheet et al. [34] developed an iris segmentation algorithm based on a Dense U-Net architecture, integrating dense connectivity with the U-Net framework to exploit its inherent advantages of reduced trainable parameters and proven efficacy in semantic segmentation. The proposed method yielded a segmentation accuracy of 98.36% and an F1 of 97.07% and demonstrated robustness in complex scenarios, such as the diagnosis of diseases affecting the iris.

The authors of [27] proposed an attention-guided U-Net architecture for robust iris segmentation, called ATT-UNet. The research work involved the introduction of an attention process into the standard U-Net to accurately discriminate the iris region from other eye regions. The attention mechanism incorporated into the system detects, to some extent, inefficiencies in the attention mask generation. The proposed algorithm was trained and evaluated using the CASIA V4 Distance and UBIRIS V4 datasets.

### *2.3. Swin Transformer-Based Algorithms*

This section presents a concise literature review of selected state-of-the-art algorithms proposed in the literature that build upon the Swin Transformer architecture. Researchers widely adopt Swin Transformer to address segmentation tasks, including iris segmentation, which plays a crucial role in robust iris authentication systems. Several authors who leveraged this architecture as the backbone of their proposed approaches have reported promising outcomes, underscoring its capacity to capture both local and global contextual information, which is essential for accurately delineating fine-grained anatomical structures such as the iris boundary [1,34].

Gao and Bourlai [11] proposed SwinIris, an iris recognition approach employing a fine-tuned Swin Transformer coupled with linear layers, reporting classification accuracy ranging from 95.14% to 99.56% across four distinct CASIA iris datasets. Although the algorithm demonstrated strong performance across all four datasets, it underperformed other existing state-of-the-art methods by over 4.7% [1,35].

Qurru et al. 2023 [35] proposed an improved CNN-based network architecture for image segmentation. The proposed system combined a CNN and Transformer for global extraction of semantic information and local characteristics. The method was evaluated and tested using medical datasets related to CT scans and ultrasound images. The developed method can also be leveraged in iris segmentation to provide a more reliable tool for iris-based applications.

Meng and Bao [1] developed a hybrid deep learning framework that integrates a bilateral segmentation backbone combining Swin Transformer with convolutional neural networks. Their architecture incorporates a multi-scale feature extraction module to capture spatial information at varying resolutions, alongside a channel attention mechanism to

enhance iris region discriminability. When assessed on the IITD and UBIRIS.v2 iris datasets, the model obtained F1 values of 0.9844 and 0.9774, respectively.

Sundaram et al. [13] proposed a comparative study of Vision Transformer, Swin Transformer, and Transformer in Transformer in image classification tasks, and the results showed that Swin Transformer achieved an excellent balance between accuracy and computational efficiency compared to Transformer in Transformer (TNT).

Huynh et al. [36] presented RadialSwin-Unet, a lightweight segmentation architecture that leverages Swin-based attention within a polar coordinate representation aligned to the iris's radial–angular structure. Evaluated on the CASIA-IrisV4 and IITD benchmarks, the model attained a Dice score of 0.9885 and contained 13× fewer parameters than U-Net, offering a strong accuracy–efficiency balance for resource-constrained biometric systems.

## 3. Methodology

### 3.1. Overview of U-Net Architecture

In our proposed SwinIrisNet method, the U-Net architecture is one of the core components; therefore, we briefly introduce the U-Net architecture before discussing the proposed methodology. The U-Net framework is widely employed for segmentation tasks, especially in the field of medical image analysis [7,8,33,35,37].

U-Net follows an encoder–decoder topology, with a contracting path for hierarchical feature extraction and an expansive path for progressive spatial recovery, as depicted in Figure 1. The encoding pathway captures multi-resolution feature hierarchies through repeated application of $3 \times 3$ convolutional kernels with ReLU activations, followed by $2 \times 2$ max-pooling layers that systematically downsample the spatial dimensions while expanding the feature channel capacity. The bottleneck retains maximally compressed feature representations that encode the highest-level semantic information. In the decoder, each stage performs bilinear upsampling followed by a $2 \times 2$ transposed convolution to halve the channel dimensionality. Skip connections concatenate cropped encoder feature maps with the upsampled representations, recovering fine-grained spatial detail. Two successive $3 \times 3$ convolutional layers with ReLU activations then refine the reconstructed features. The output layer applies a $1 \times 1$ convolution to reduce the 64-channel feature volume to the target number of classes, yielding the dense segmentation prediction [7,8].

### 3.2. Memory-Efficient Architecture Design in SwinIrisNet

Figure 2 illustrates the overall framework of the proposed iris segmentation network, hereafter referred to as SwinIrisNet. It is designed with memory efficiency as a core architectural principle, achieved through a series of careful design decisions that collectively reduce memory requirements by several orders of magnitude. It is built on a U-Net encoder–decoder framework with a dual-branch hybrid encoder that runs a Swin Transformer branch for global context and long-range spatial relationships in parallel with a CNN branch that uses residual blocks to capture fine local details and enable precise iris boundary delineation. Their outputs are combined at each scale through a memory-efficient cross-attention fusion module augmented with the Convolutional Block Attention Module (CBAM). This fusion strategy employs global average pooling on both feature streams to generate combined context vectors, followed by learnable attention weights that adaptively emphasize relevant features.

Let $H$ and $W$ denote the spatial height and width of a given feature map (in pixels/tokens), $C$ the channel dimension (e.g., $48, 96, 192, 384$ across successive Swin stages), $B$ the batch size (set to 4 in SwinIrisNet), and $N = H \cdot W$ the total token count. The standard Swin Transformer configuration employs an embedding dimension of 96 with layer depths

of $[2, 2, 6, 2]$. In contrast, SwinIrisNet adopts a reduced embedding dimension of 48 and modified stage depths of $[2, 2, 4, 2]$ to achieve a more lightweight model. Since Transformer memory scales quadratically with channel dimension in query–key–value ($QKV$) projections ($3 \times C \times C$ per head), halving the embedding dimension reduces attention memory by a factor of 4. The shallower stage 3, comprising four blocks rather than six, further eliminates two full blocks of activations from memory usage. Window-based self-attention restricts token interactions to local $7 \times 7$ windows, reducing attention complexity from $\mathcal{O}(N^2)$ to $\mathcal{O}(H \cdot W)$ and making it linear in image size. Complementing this, reduced CNN filters $[32, 64, 128, 256]$ and decoder filters $[128, 64, 32, 16, 8]$ directly halve activation memory relative to standard configurations, since activation memory scales linearly with channel count.

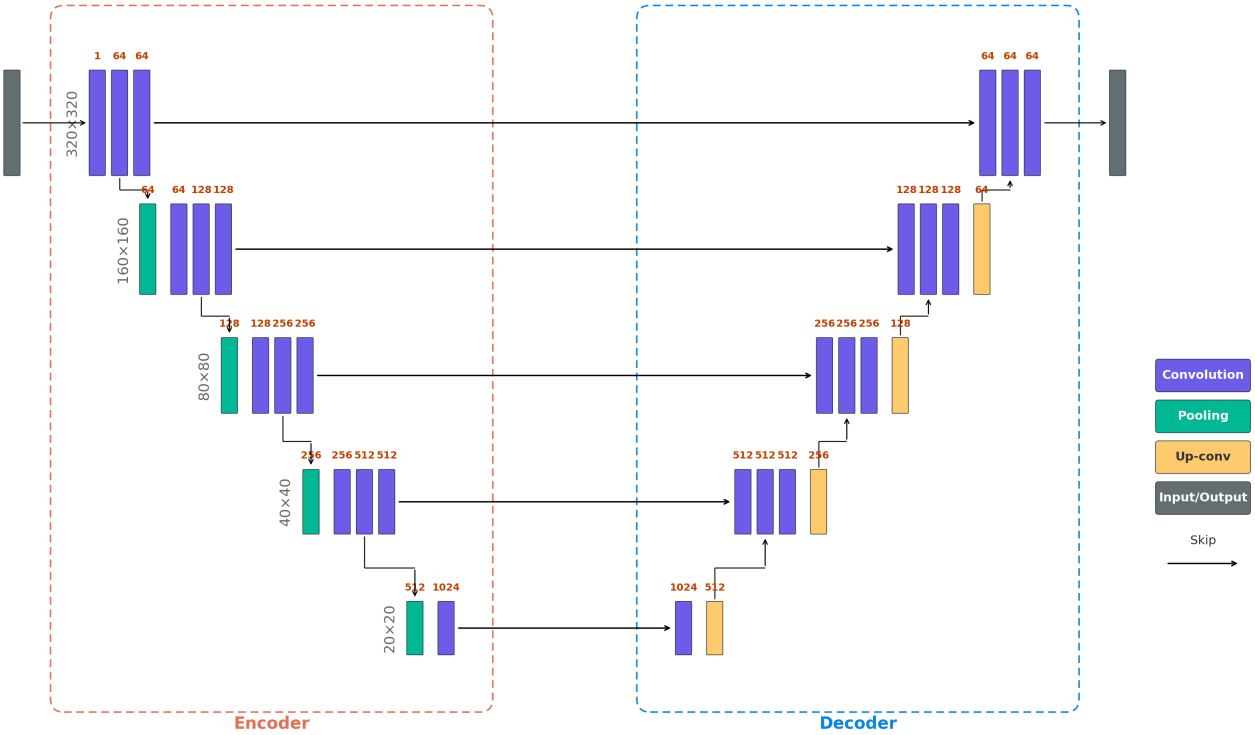



**Figure 1.** Basic structure of U-Net.

The CBAM module applies a reduction ratio of 8 to its fully connected (FC) channel attention layers, while patch merging avoids storing intermediate feature maps during spatial downsampling. Most importantly, the cross-attention fusion module replaces spatial cross-attention with global average pooling, followed by channel-wise fully connected layers, reducing per-stage fusion memory complexity from $\mathcal{O}(B \cdot H^2 \cdot W^2)$ to $\mathcal{O}(B \cdot C)$. For a $64 \times 64$ feature map with $C = 48$ and $B = 4$, this reduction translates from about $\sim$67 million units to merely 192 units, representing a decrease of roughly five orders of magnitude. Each Swin Transformer block (ST-Block) integrates window-based and shifted-window multi-head self-attention mechanisms for cross-window feature interaction. Each residual block (ResBlock) comprises two sequential $3 \times 3$ convolutional layers with batch normalization and ReLU activation, followed by a residual connection.

The SwinIrisNet decoder consists of five sequential decoder blocks with progressively decreasing filter counts $[128, 64, 32, 16, 8]$, each performing bilinear $2\times$ upsampling followed by an attention-gated skip connection from the corresponding encoder fusion stage, concatenation, and two $3 \times 3$ convolutional layers with batch normalization and ReLU activation. The attention gates selectively focus on iris-relevant features in the skip connections while suppressing irrelevant background features such as eyelashes and sclera. Spatial

dropout is applied between blocks for regularization, and a final $1 \times 1$ convolution with sigmoid activation produces the binary segmentation mask at the original input resolution of $256 \times 256 \times 1$.

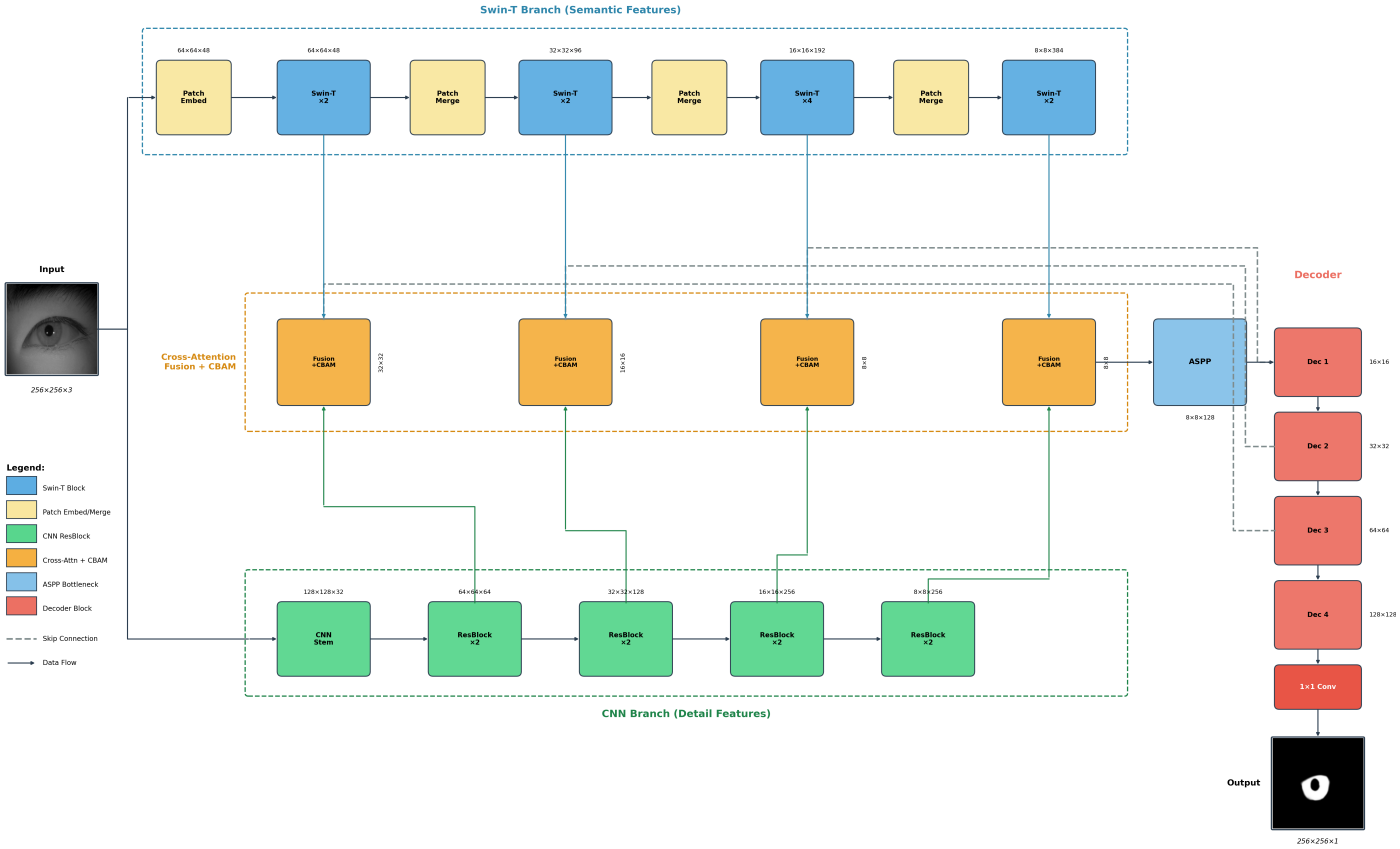



**Figure 2.** Architecture of the proposed SwinIrisNet method for iris segmentation.

The choice of architecture for SwinIrisNet is motivated by the complementary strengths of its two branches: the Swin Transformer branch captures the long-range global context needed to localize the iris region under challenging conditions (occlusions, illumination variations, gaze variations, etc.), while the parallel CNN branch preserves the local, high-frequency detail required for precise pupil and limbus boundaries. Rather than concatenating these cues, we fuse them with a cross-attention module followed by a CBAM, so that channel and spatial attention adaptively emphasize the more reliable branch at each location, and an ASPP bottleneck aggregates multi-scale context before decoding. Unlike recent hybrid Transformer–CNN segmentation networks (e.g., TransUNet and Swin-Unet) that adopt a serial configuration by pairing a Transformer encoder with a convolutional decoder, SwinIrisNet employs parallel Swin Transformer and CNN branches whose multi-scale features are integrated via a cross-attention fusion module with a CBAM at each resolution level, explicitly balancing global semantic representations with local boundary details.

Together, these design choices enable SwinIrisNet to train on a single 24 GB NVIDIA A10G, making high-performance iris segmentation accessible without enterprise-tier resources. Figure 3 depicts the structure of the feature fusion module and the decoder block of the proposed SwinIrisNet model, respectively. Table 1 summarizes the details of the proposed SwinIrisNet structure.

In our implementation, SwinIrisNet is trained with a combined loss comprising binary cross-entropy ($\mathcal{L}_{BCE}$), Dice ($\mathcal{L}_{Dice}$), and dilated boundary Dice ($\mathcal{L}_{Boundary}$) to balance pixel-wise accuracy, region overlap, and boundary localization. Boundary localization is

critical for downstream iris recognition, where precise boundary delineation affects feature extraction quality [7]. For reproducibility, we specify the complete experimental protocol, including the dataset split (70/15/15), input resolution (256 × 256), preprocessing steps, optimizer, learning-rate schedule, and loss weighting, with all hyperparameter values listed in the Table A1.

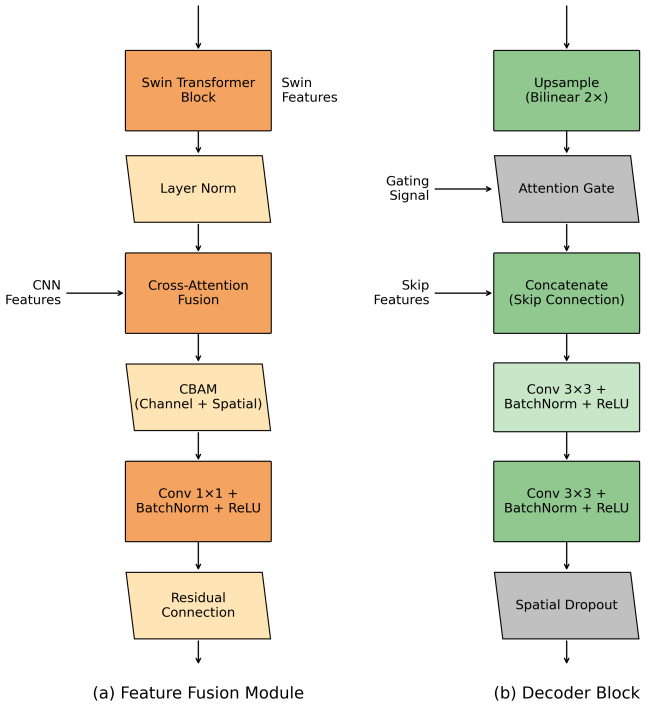



**Figure 3.** SwinIrisNet: component-level view of the feature fusion and decoding stages.

**Table 1.** Layer-wise specifications of the SwinIrisNet architecture.

| Component | Input Dimensions (H × W × C) | Layer Operation | Stride | Output Dimensions (H × W × C) |
|---|---|---|---|---|
| Swin Transformer Branch | 256 × 256 × 3 | Patch Embedding (4 × 4 Conv) | 4 | 64 × 64 × 48 |
| | 64 × 64 × 48 | ST-Block × 2 | 1 | 64 × 64 × 48 |
| | 64 × 64 × 48 | Patch Merging | 2 | 32 × 32 × 96 |
| | 32 × 32 × 96 | ST-Block × 2 | 1 | 32 × 32 × 96 |
| | 32 × 32 × 96 | Patch Merging | 2 | 16 × 16 × 192 |
| | 16 × 16 × 192 | ST-Block × 4 | 1 | 16 × 16 × 192 |
| | 16 × 16 × 192 | Patch Merging | 2 | 8 × 8 × 384 |
| | 8 × 8 × 384 | ST-Block × 2 | 1 | 8 × 8 × 384 |
| CNN Branch | 256 × 256 × 3 | 3 × 3 Conv→BN→ReLU | 2 | 128 × 128 × 32 |
| | 128 × 128 × 32 | 3 × 3 Conv→BN→ReLU × 2 | 1 | 128 × 128 × 32 |
| | 128 × 128 × 32 | ResBlock × 2 | 1 | 128 × 128 × 32 |
| | 128 × 128 × 32 | ResBlock × 2 | 2 | 64 × 64 × 64 |
| | 64 × 64 × 64 | ResBlock × 2 | 2 | 32 × 32 × 128 |
| | 32 × 32 × 128 | ResBlock × 2 | 2 | 16 × 16 × 256 |
| Feature Fusion | Stage 1 | Cross-Attention Fusion + CBAM | – | 64 × 64 × 48 |
| | Stage 2 | Cross-Attention Fusion + CBAM | – | 32 × 32 × 96 |
| | Stage 3 | Cross-Attention Fusion + CBAM | – | 16 × 16 × 192 |
| | Stage 4 | Cross-Attention Fusion + CBAM | – | 8 × 8 × 384 |
| Bottleneck | 8 × 8 × 384 | ASPP (rates: 6, 12, 18) | 1 | 8 × 8 × 128 |
| Decoder | 8 × 8 × 128 | Upsample + Attention Gate + Conv | 2 | 16 × 16 × 128 |
| | 16 × 16 × 128 | Upsample + Attention Gate + Conv | 2 | 32 × 32 × 64 |
| | 32 × 32 × 64 | Upsample + Attention Gate + Conv | 2 | 64 × 64 × 32 |
| | 64 × 64 × 32 | Upsample + Conv | 2 | 128 × 128 × 16 |
| | 128 × 128 × 16 | Upsample + Conv | 2 | 256 × 256 × 8 |
| | 256 × 256 × 8 | 1 × 1 Conv + Sigmoid | 1 | 256 × 256 × 1 |

## 4. Experimental Setup

### 4.1. Overview of Iris Image Datasets

To assess the robustness and generalizability of the proposed method, we conducted extensive experiments on four publicly available iris datasets captured under diverse and challenging conditions: UBIRIS.v2 [38], IITD, CASIA-Thousand, and MMU.v1 [39]. Collectively, these datasets encompass a wide spectrum of real-world acquisition scenarios, including variations in illumination, off-angle acquisition, and at-distance imaging, which are characteristics of unconstrained environments that remain open challenges in biometric research. The ground-truth annotations utilized in this work for all four datasets were made available by the WaveLab laboratory at the University of Salzburg, referred to as IrisSeg-EP [40]. UBIRIS.v2 [38] contains iris images from 261 subjects acquired under unconstrained visible-wavelength conditions at distances of 4–8 m, encompassing on-the-move capture, off-axis gaze, occlusions, specular reflections, and eyeglasses. For this study, the dataset was partitioned into 1572 training, 330 validation (17 subjects), and 330 test images (17 subjects). IITD [39] contains 2240 NIR iris images from 224 subjects (10 samples each: 5 left, 5 right) captured using CMOS, JPC1000, and JIRIS sensors at a resolution of 320 × 240 pixels [11]. CASIA-Thousand [39], the first publicly available thousand-subject iris dataset, was curated by the Chinese Academy of Sciences. Intra-class variation arises primarily from eyeglasses and specular reflections. A subject-disjoint evaluation was conducted on 1000 images. MMU.v1 [39] comprises iris images from 45 subjects captured using the LG IrisAccess 2200 sensor (sourced from LG Electronics, Seoul, Republic of Korea) under controlled indoor conditions in BMP format at 320 × 240 pixels. We used 310 images for training, 70 for validation, and 70 for testing. The key attributes of the four datasets are summarized in Table 2.

**Table 2.** Key attributes of the four iris datasets used for experimental validation.

| Characteristic | UBIRIS.v2 | CASIA-Thousand | IITD | MMU.v1 |
|---|---|---|---|---|
| Size of dataset | 2232 | 1000 | 2240 | 450 |
| Image Format | TIFF | JPEG | BMP | BMP |
| Image Size | 400 × 300 pixels | 640 × 480 pixels | 320 × 240 pixels | 320 × 240 pixels |
| Wavelength | Visible Light | Near-Infrared | Near-Infrared | Near-Infrared |
| Color | RGB | gray-level | gray-level | gray-level |
| Splitting approach | subject-disjoint | subject-disjoint | subject-disjoint | subject-disjoint |

### 4.2. Data Augmentation

To improve the generalization of SwinIrisNet and mitigate possible overfitting, on-the-fly data augmentation was applied to the training set only, whereas validation and test images were resized to 256 × 256 without any augmentation. All iris images were first resized to 256 × 256 and normalized to the range $[0, 1]$. Each augmentation was applied stochastically per iris image according to the probabilities shown in Table 3. Based on the probabilities used, at least 90% of the iris images in the training set were augmented, roughly doubling the effective amount of training data per epoch.

### 4.3. Implementation Details

We implemented the proposed architecture using TensorFlow [41]. All experiments were conducted on a computing instance equipped with a single NVIDIA A10G (24 GB GDDR6 memory), 4 vCPUs, 16 GB of system RAM, and 250 GB of NVMe SSD storage, optimized for deep learning workloads. For each of the four iris datasets, data were partitioned into training, validation, and test splits using a 70:15:15 ratio, with the split at the subject level ensuring that the same subject or eye does not appear across the training, validation, and test sets. The input image size was set to 256 × 256, with the identical

augmentation technique applied uniformly across all datasets. The network was trained for 50 epochs with an initial learning rate of $1 \times 10^{-4}$ and a mini-batch configuration of four samples. The optimizer used for training all SwinIrisNet variants was AdamW, with an initial learning rate of 0.0001. Table A1 summarizes the hyperparameters and training details necessary for reproducibility. The detailed specifications and characteristics of each dataset employed in our experiments are presented in Table 2.

**Table 3.** Data augmentation operations, parameters, and their probabilities used during training.

| Augmentation | Parameters | Probability |
|---|---|---|
| Horizontal flip | – | 0.50 |
| Rotation | $\pm 15^{\circ}$ | 0.30 |
| Brightness/contrast | $\pm 0.2$ each | 0.30 |
| Hue/saturation/value | $\pm 10/\pm 20/\pm 10$ | 0.30 |
| Elastic transform | $\alpha = 50, \sigma = 2.5$ | 0.15 |
| Grid distortion | 5 steps, distort limit 0.1 | 0.15 |
| Gaussian noise | variance $[5, 30]$ | 0.20 |
| Blur (Gaussian or motion) | kernel limit 3 | 0.20 |
| CLAHE | clip limit 2.0 | 0.09 |
| Sharpen | $\alpha$ $[0.1, 0.3]$, lightness $[0.5, 1.0]$ | 0.09 |

### *4.4. Performance Evaluation Metrics*

To evaluate SwinIrisNet and benchmark it against existing state-of-the-art iris segmentation algorithms, we used a comprehensive set of evaluation metrics: the Dice coefficient, mean intersection over union (mIoU), F1, accuracy, precision, specificity, training loss, and the NICE1 and NICE2 error scores. The mathematical formulations and brief descriptions of these metrics are as follows [1,22]:

$$\text{Dice} = \frac{2 \times |Y \cap \hat{Y}|}{|Y| + |\hat{Y}|} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{IoU} = \frac{|Y \cap \hat{Y}|}{|Y \cup \hat{Y}|} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{2}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{7}$$

$$\text{Error Rate} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{8}$$

$$\text{NICE}_1 = \frac{|Y \oplus \hat{Y}|}{|Y|} = \frac{\text{Disagreeing Pixels}}{\text{Ground Truth Iris Pixels}} \tag{9}$$

$$\mathrm{NICE}_2 = 1 - \frac{2 \times |\partial Y \cap \partial \hat{Y}|}{|\partial Y| + |\partial \hat{Y}|} \tag{10}$$

where $Y$ represents the ground-truth mask; $\hat{Y}$ represents the predicted mask; FP, FN, TP, and TN represent the numbers of false positives, false negatives, true positives, and true negatives, respectively; $\oplus$ denotes the XOR operation; and $\partial$ denotes the boundary operator.

For F1, mIoU, precision, specificity, accuracy, recall, and Dice, values closer to 1 indicate superior segmentation performance, whereas values approaching 0 reflect poor performance. Conversely, for loss, NICE1, and NICE2, values closer to 0 indicate superior performance, while values approaching 1 reflect degraded performance.

To balance segmentation accuracy and iris boundary precision, SwinIrisNet employs a composite loss function (Equation (14)) that integrates three complementary components: Dice loss (Equation (11)) for spatial overlap optimization, binary cross-entropy loss (Equation (12)) for pixel-wise classification, and boundary loss (Equation (13)) for iris contour fidelity.

$$\mathcal{L}_{\mathrm{Dice}} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i + \epsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + \epsilon} \tag{11}$$

$$\mathcal{L}_{\mathrm{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \tag{12}$$

$$\mathcal{L}_{\mathrm{Boundary}} = \frac{1}{N}\sum_{i=1}^{N}(\nabla y_i - \nabla \hat{y}_i)^2 \tag{13}$$

$$\mathcal{L}_{\mathrm{total}} = \alpha \cdot \mathcal{L}_{\mathrm{BCE}} + \beta \cdot \mathcal{L}_{\mathrm{Dice}} + \gamma \cdot \mathcal{L}_{\mathrm{Boundary}}. \tag{14}$$

$N$ represents the total number of pixels in the image, $y_i$ indicates the ground-truth label assigned to pixel $i$, $\hat{y}_i$ represents the predicted probability output by the model for pixel $i$, and $\epsilon$ is a small smoothing constant introduced to avoid division by zero. $\nabla Y_i$ represents the spatial gradient of the ground-truth mask at pixel $i$; $\nabla \hat{Y}_i$ represents the spatial gradient of the predicted mask at pixel $i$; and $\alpha$, $\beta$, and $\gamma$ represent the weighting coefficients for the binary cross-entropy (BCE), Dice, and boundary loss components, respectively.

## 5. Experimental Results

This section details the empirical assessment of the proposed SwinIrisNet architecture for iris segmentation. The model was systematically trained and evaluated across four benchmark datasets encompassing diverse acquisition conditions, including variations in illumination, occlusion, capture distance, and sensor noise.

### *5.1. Ablation Experiments on Two Datasets*

To isolate the contribution of each component, four variants were trained under identical hyperparameters and evaluated on the CASIA-Thousand and UBIRIS.v2 datasets: (i) a Swin Transformer-only encoder, (ii) a CNN-only residual encoder, (iii) SwinIrisNet *w/o* boundary loss (BL), and (iv) the final SwinIrisNet with the boundary-aware loss. The experimental results for each segmentation network variant tested on these two datasets are reported in Table 4.

The results in Table 4 show that SwinIrisNet achieved better segmentation results than the other single-branch variants. On the CASIA-Thousand dataset, compared to the single-branch network variants, SwinIrisNet's F1 improved by 3.95% and 1.34%, whereas NICE2 improved by 10.43% and 14.27%. On UBIRIS.v2, SwinIrisNet displayed mIoU improvements of 2.82% and 2.92% and NICE2 improvements of 20.80% and 15.74%.

**Table 4.** Ablation experiments on two iris datasets: CASIA-Thousand and UBIRIS.v2.

| Category | Method | Precision | Accuracy | mIoU | Recall | F1 | Dice | NICE1 (%) | NICE2 (%) |
|---|---|---|---|---|---|---|---|---|---|
| CASIA-Thousand | Swin Transformer | 0.9289 | 99.29 | 0.8850 | 0.9493 | 0.9388 | 0.9062 | 0.620 | 2.032 |
| | CNN | 0.9483 | 99.35 | 0.9078 | 0.9550 | 0.9514 | 0.9409 | 0.488 | 2.123 |
| | SwinIrisNet *w/o* BL | 0.9493 | 99.44 | 0.9114 | 0.9591 | 0.9534 | 0.9459 | 0.43 | 2.092 |
| | **SwinIrisNet (Final)** | **0.9552** | **99.46** | **0.920** | **0.9612** | **0.9605** | **0.9477** | **0.60** | **1.82** |
| UBIRIS.v2 | Swin Transformer | 0.9459 | 99.33 | 0.9045 | 0.9538 | 0.9497 | 0.9380 | 0.67 | 2.50 |
| | CNN | 0.9423 | 99.41 | 0.9036 | 0.9566 | 0.9492 | 0.9143 | 0.59 | 2.35 |
| | SwinIrisNet *w/o* BL | 0.9313 | 99.38 | 0.9005 | 0.9647 | 0.9475 | 0.9167 | 0.62 | 2.07 |
| | **SwinIrisNet (Final)** | **0.9563** | **0.99.54** | **0.9300** | **0.9672** | **0.9624** | **0.9519** | **0.59** | **1.98** |

Note: Bold represents the best iris segmentation network (our SwinIrisNet) in comparison with the ablation variants.

Adding the boundary loss to the SwinIrisNet *w/o* BL variant reduced NICE2 by 13.0% on CASIA-Thousand and improved F1 by 0.21%, Dice by 0.18%, and mIoU by 0.86%. On UBIRIS.v2, NICE2 decreased by 4.3%, with gains of 2.50% in precision, 2.95% in mIoU, 1.49% in F1, and 3.52% in Dice. This differential effect is a direct consequence of the formulation of the boundary term, which computes the mean-squared error between the Sobel gradient-magnitude maps of the predicted and ground-truth masks. As a high-pass operator, it provides an informative signal, predominantly in regions where boundaries are well-defined and consistently annotated. On the NIR CASIA-Thousand dataset, in which the pupil and limbus boundaries exhibit high contrast and precise labeling, this resulted in a significant NICE2 reduction of 13.0%. Conversely, on the visible-light UBIRIS.v2 dataset, in which boundaries are inherently blurred, exhibit greater variability, and residual error is dominated by occlusion artifacts rather than boundary localization imprecision, the same metric improved by only 4.3% despite considerably larger region-level gains in Dice and mIoU. Gradient-based boundary losses therefore demonstrate strong efficacy on sharp near-infrared edges but remain susceptible to high-frequency noise in diffuse visible-light imagery. Noise-robust alternatives, such as distance-transform, Hausdorff, or smoothed soft-boundary losses, are better suited to such conditions and constitute a promising direction for future investigation.

The final SwinIrisNet network demonstrated competitive segmentation performance on both the CASIA-Thousand and UBIRIS.v2 datasets, supporting two claims: (1) the hybrid encoder yields consistent gains over single-modal baselines and (2) the boundary-aware loss is necessary for translating the added capacity into measurable improvements in both datasets, serving as an edge-supervision signal on clean iris images and as a regularizer that stabilizes training on noisier iris images. The qualitative segmentation results of the various network configurations on the CASIA-Thousand and UBIRIS.v2 datasets are presented in Figure 4 and Figure 5, respectively.

### *5.2. Experimental Results on Four Different Datasets*

Table 5 presents a comprehensive summary of SwinIrisNet's segmentation performance across all four benchmark iris datasets: IITD, UBIRIS.v2, CASIA-Thousand, and MMU.v1. The method achieved pixel-level accuracy greater than 99.51%, specificity greater than 0.9967, and precision greater than 0.9565 on every dataset, confirming reliable discrimination of iris and non-iris regions with minimal false positives. Recall values consistently exceeded 0.96 (peaking at 0.9672 on CASIA-Thousand), while mIoU values ranged from 0.9266 to 0.9450, indicating close agreement with ground-truth masks. F1 values ranged from 0.9612 to 0.9672 and Dice coefficients from 0.9489 to 0.9519, reflecting a balanced precision–recall trade-off. NICE1 errors remained below 0.60% across all

datasets (lowest, 0.57% on MMU.v1), and NICE2 errors ranged from 1.82 to 2.24%, with the highest value observed on CASIA-Thousand. These results confirm that SwinIris-Net generalizes effectively across datasets with varying acquisition protocols, resolutions, and occlusion levels.

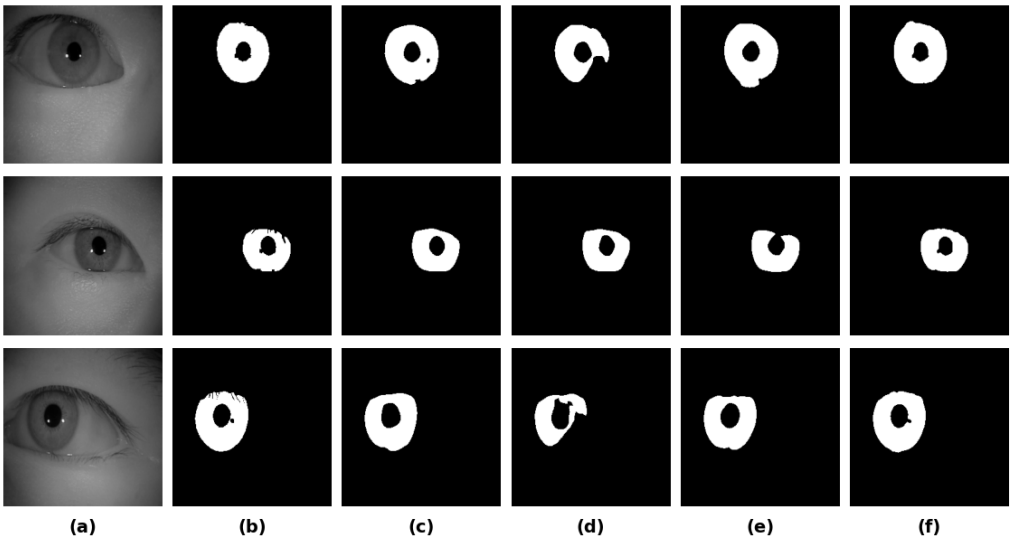



**Figure 4.** Segmentation results of different methods on CASIA-Thousand: (**a**) original iris image, (**b**) ground truth, (**c**) Swin Transformer only, (**d**) CNN only, (**e**) SwinIrisNet *w/o* BL, (**f**) final SwinIrisNet.

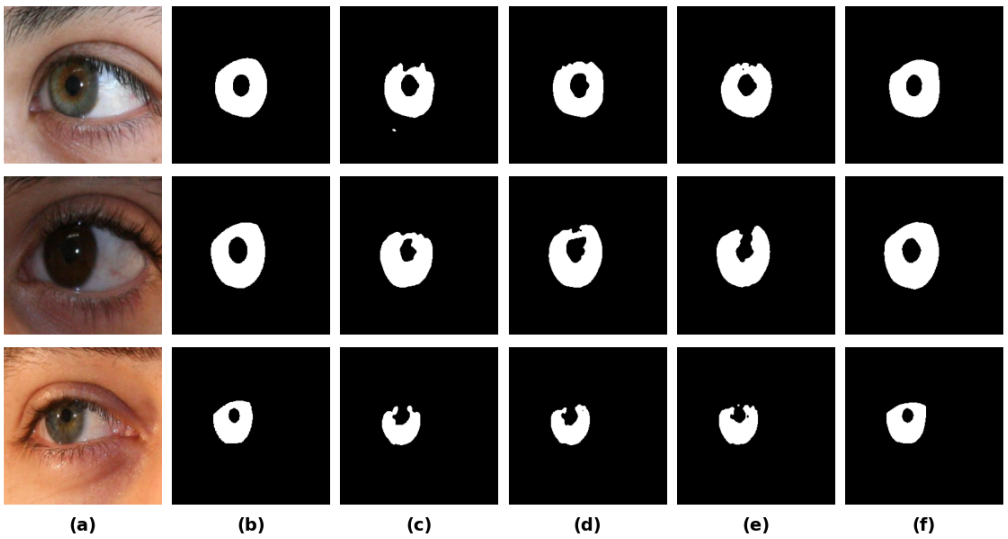



**Figure 5.** Segmentation results of different methods on UBIRIS.v2: (**a**) original iris image, (**b**) ground truth, (**c**) Swin Transformer variant, (**d**) CNN variant, (**e**) SwinIrisNet *w/o* BL, (**f**) final SwinIrisNet.

**Table 5.** SwinIrisNet segmentation performance on four different iris datasets.

| Dataset | Loss | Precision | Accuracy | mIoU | Recall | Specificity | F1 | Dice | NICE1 (%) | NICE2 (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| UBIRIS.v2 | 0.0479 | 0.9565 | 99.51 | 0.9300 | 0.9626 | 0.9967 | 0.9672 | 0.9519 | 0.59 | 2.07 |
| IITD | 0.0594 | 0.9633 | 99.53 | 0.9450 | 0.9635 | 0.9969 | 0.9662 | 0.9489 | 0.60 | 2.24 |
| CASIA-Thousand | 0.0478 | 0.9565 | 99.51 | 0.9266 | 0.9672 | 0.9967 | 0.9612 | 0.9519 | 0.60 | 1.82 |
| MMU.v1 | 0.0475 | 0.9608 | 99.52 | 0.9277 | 0.9640 | 0.9969 | 0.9622 | 0.9497 | 0.57 | 2.210 |

Figure 6 shows that loss dropped sharply in the early epochs and that the key metrics (precision, recall, mIoU, and F1) stabilized around epoch 18 on all four datasets, indicating rapid and consistent convergence.

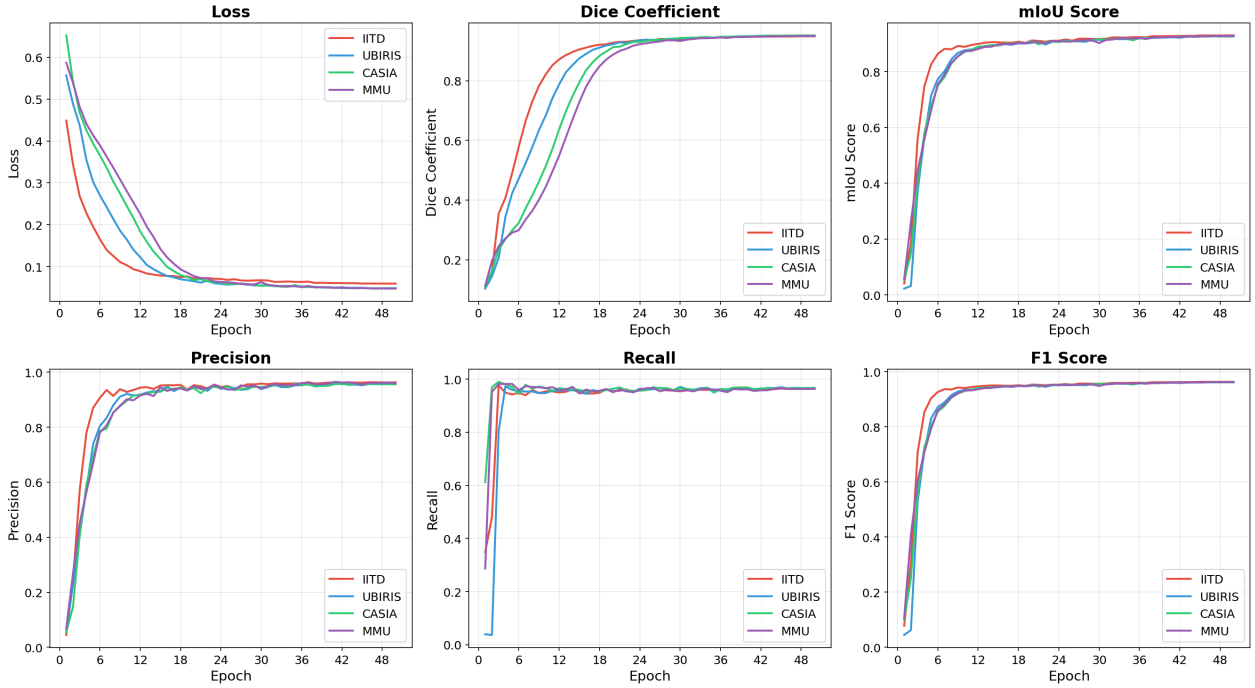



**Figure 6.** Segmentation results for SwinIrisNet (loss, Dice, precision, recall, F1, and mIoU) across the UBIRIS.v2, CASIA-Thousand, IITD, and MMU.v1 datasets.

Figure 7 confirms training accuracy convergence around the 12th epoch, while other metrics such as Dice, loss, NICE1, and NICE2 stabilized around the 30th epoch. All models were trained for a maximum of 50 epochs. To prevent overfitting, we applied early stopping to monitor validation loss with a patience of 15 epochs and a minimum-improvement threshold of $1 \times 10^{-4}$; upon triggering, the weights were restored to those of the best epoch, and the checkpoint maximizing the validation Dice coefficient was retained as the final model. The 50-epoch value thus served as an upper training budget rather than a fixed stopping point. This budget was chosen because hybrid Swin Transformer–CNN models, whose Transformer components lack the local inductive biases of convolutions, converge more slowly than purely convolutional networks, an effect amplified by our extensive data augmentation [1,11,13]. The early-stopping criterion was not triggered during training, indicating that the validation loss was still improving at epoch 50 without a prolonged stagnation phase. The same 50-epoch budget and early-stopping protocol were applied uniformly across all ablation variants for fair comparison with existing methods on the same or similar datasets [1,27], ensuring that performance differences reflected architectural design rather than training duration.

Figure 8 illustrates the segmentation results for SwinIrisNet on the UBIRIS.v2 dataset. Each row represents a different subject, illustrating performance across diverse iris appearances, including variations in pigmentation, pupil size, eyelash occlusion, and capture conditions. The predicted masks closely aligned with ground-truth annotations at both the iris–sclera and iris–pupil boundaries, confirming that SwinIrisNet accurately captured the ring-shaped structure of the iris under the challenging visible-wavelength conditions of UBIRIS.v2.

Figure 9 depicts the SwinIrisNet segmentation outcomes on the UBIRIS.v2 dataset, where green indicates true positives, red indicates false positives, and blue marks false negatives, validating precise iris boundary delineation under varying gaze angles and occlusion conditions.

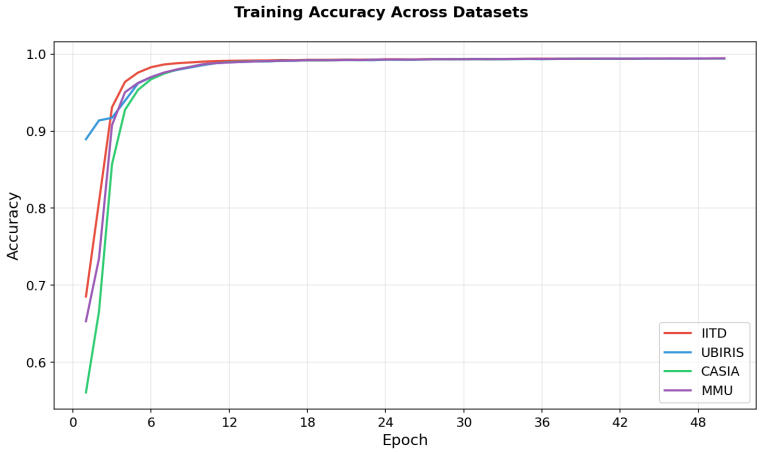



**Figure 7.** Training accuracy curves across the four iris datasets over 50 epochs.

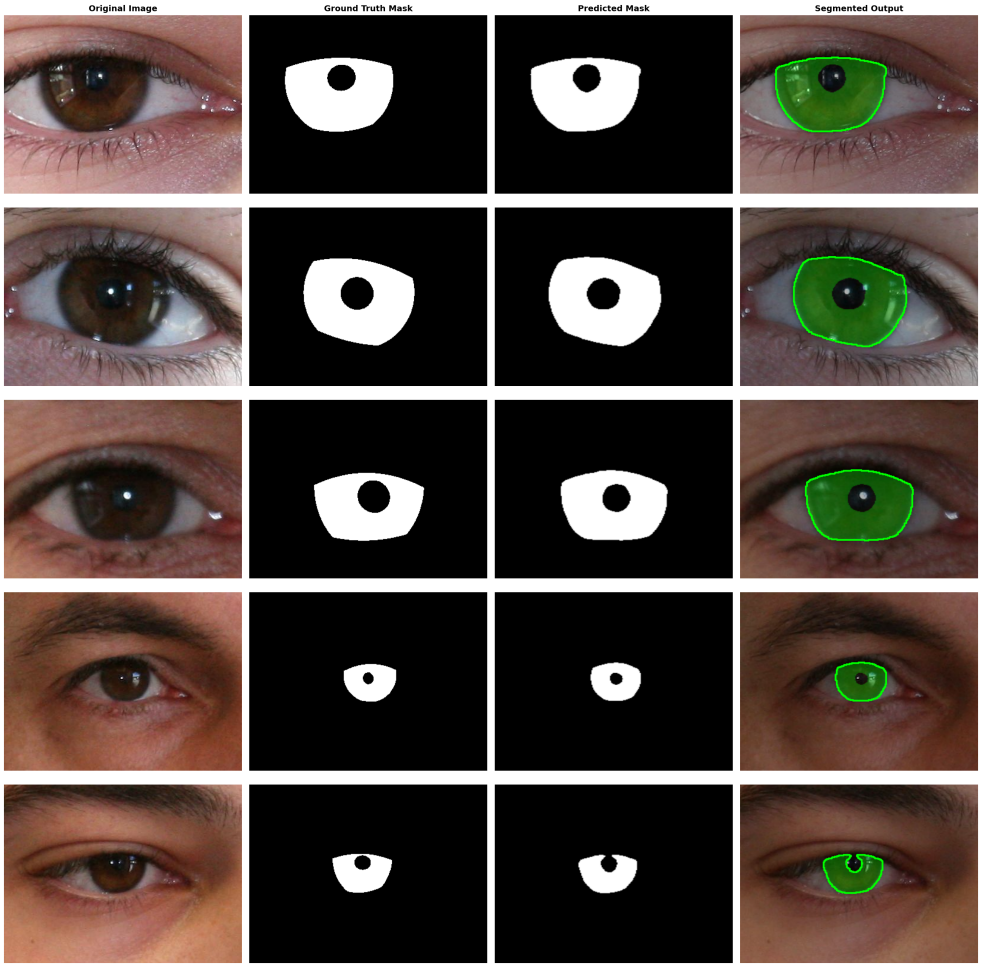



**Figure 8.** Segmentation outputs of SwinIrisNet on the UBIRIS.v2 dataset. From **left** to **right**: the original iris image, the corresponding ground-truth mask, the predicted segmentation mask, and the overlay on the original image.

Figure 10 presents the segmentation results of SwinIrisNet on the IITD iris dataset. The predicted masks exhibit strong spatial similarity to the reference annotations, accurately delineating both the iris–sclera boundary and the pupillary margin. Since SwinIrisNet operates at a fixed $256 \times 256$ input resolution, the predicted masks were rescaled to the native IITD input resolution $(320 \times 240)$ via post-processing to maintain spatial correspondence. This rescaling introduced slightly quantized edges along the iris boundaries, visible as

minor step-like discontinuities in some rows, although overall segmentation quality and topological accuracy remained intact.

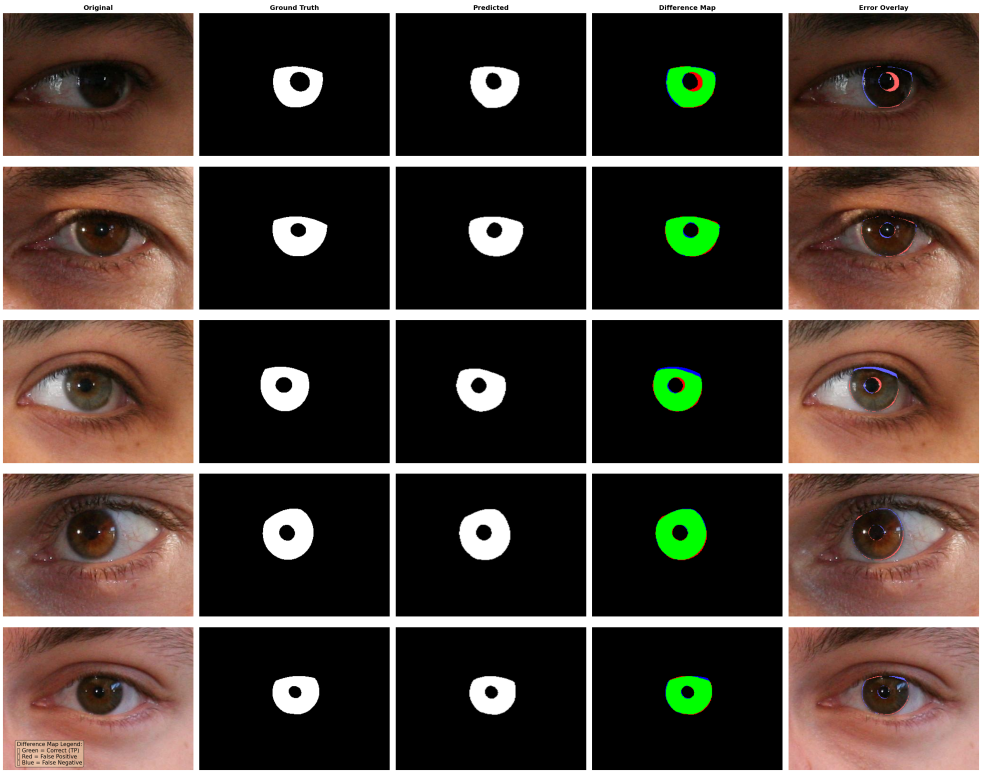



**Figure 9.** SwinIrisNet segmentation performance on the UBIRIS.v2 dataset under off-axis gaze conditions.

Figure 11 shows the qualitative segmentation results of SwinIrisNet on the MMU.v1 dataset. MMU.v1 images exhibit substantial variability, including eyelid and eyelash occlusion, off-axis capture, and non-circular iris boundaries. SwinIrisNet reliably reproduced these irregular iris shapes and preserved the pupil region across all tested samples, demonstrating its ability to generalize beyond the canonical circular iris assumption.

To assess SwinIrisNet's behavior under challenging conditions, Figure 12 shows complex samples from both the UBIRIS.v2 and CASIA-Thousand datasets, including the input, predicted boundary overlay, ground-truth mask, and predicted mask, with regions of deviation highlighted. On the visible-light UBIRIS.v2 dataset, accuracy degraded under heavy eyelid/eyelash occlusion (with boundary leakage into the sclera). On the NIR CASIA-Thousand dataset, the method was more robust, with residual errors near the pupillary boundary and around specular reflections. Much of the fragmentation in the raw output was removed by the morphological post-processing used in our quantitative evaluation; the raw output is shown here to reflect the model's intrinsic behavior. The results show that SwinIrisNet exhibits some weaknesses on noisy images, particularly complex visible-light images. Observed failure modes include occlusion-induced errors, fragmentation under low contrast, and spurious holes near specular reflections. As our datasets were acquired under relatively controlled protocols and do not systematically span severe blur, motion, extreme illumination, or sensor noise, we refrain from making strong robustness claims under such conditions.

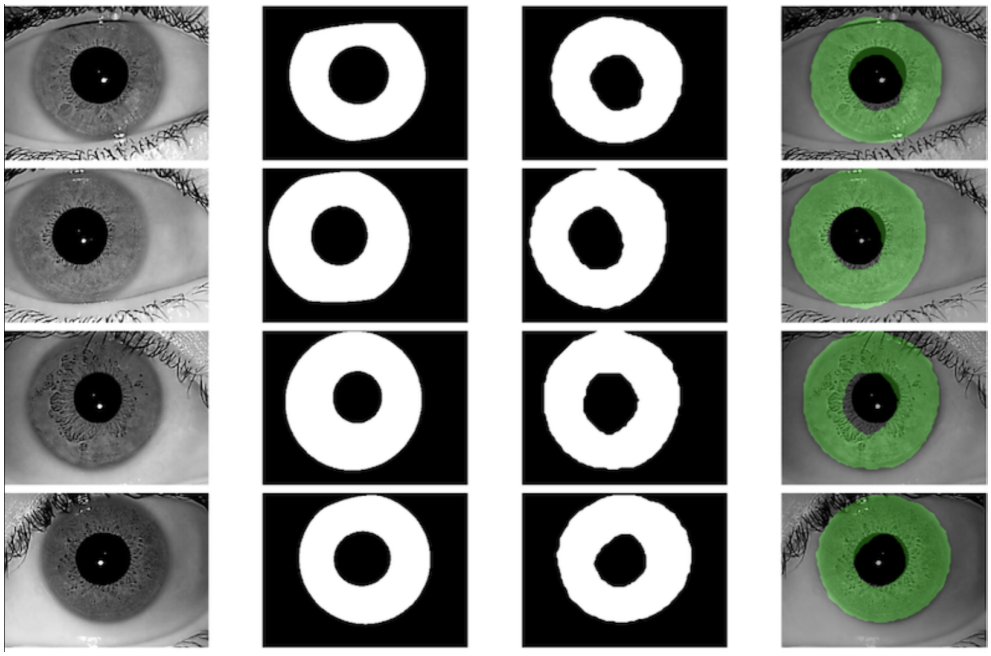


**Figure 10.** Segmentation results of SwinIrisNet on the IITD iris dataset. From **left** to **right**: the original iris image, the corresponding ground-truth mask, the predicted segmentation mask, and the overlay on the original image.

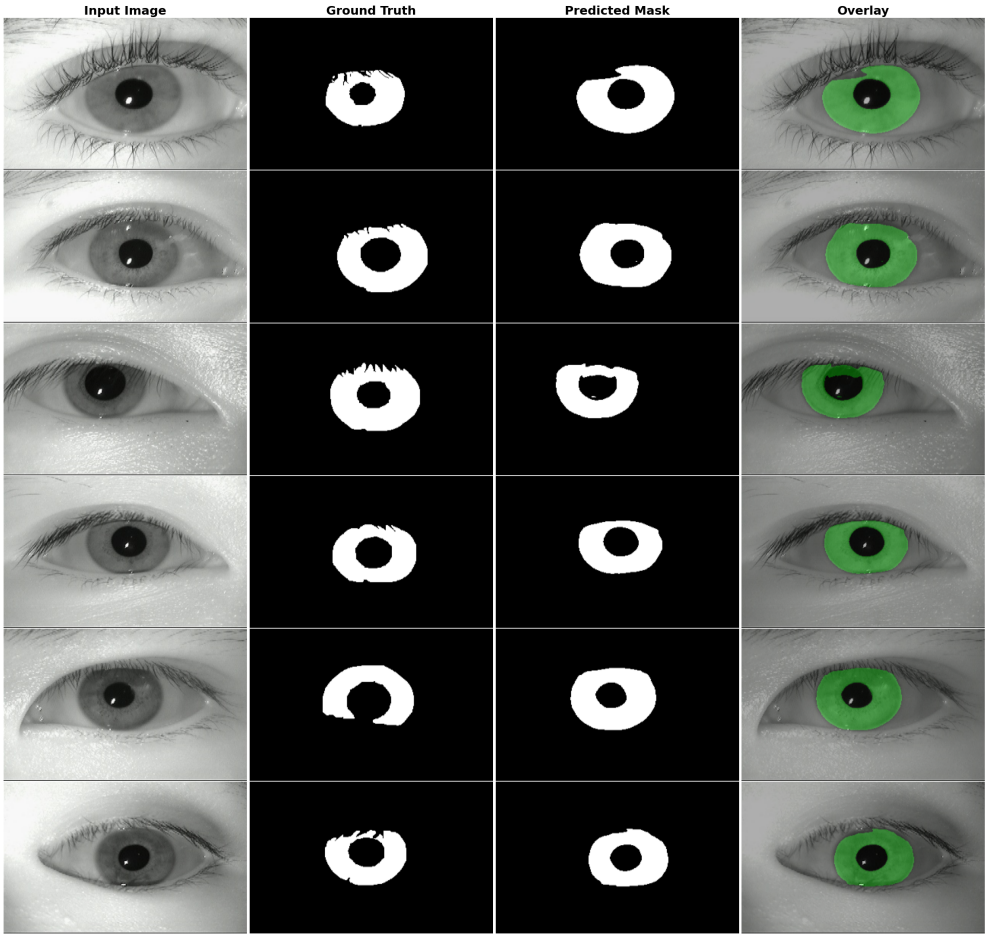



**Figure 11.** Iris segmentation results of SwinIrisNet on the MMU.v1 dataset. From **left** to **right**: the original iris image, the corresponding ground-truth mask, the predicted segmentation mask, and the overlay on the original image.

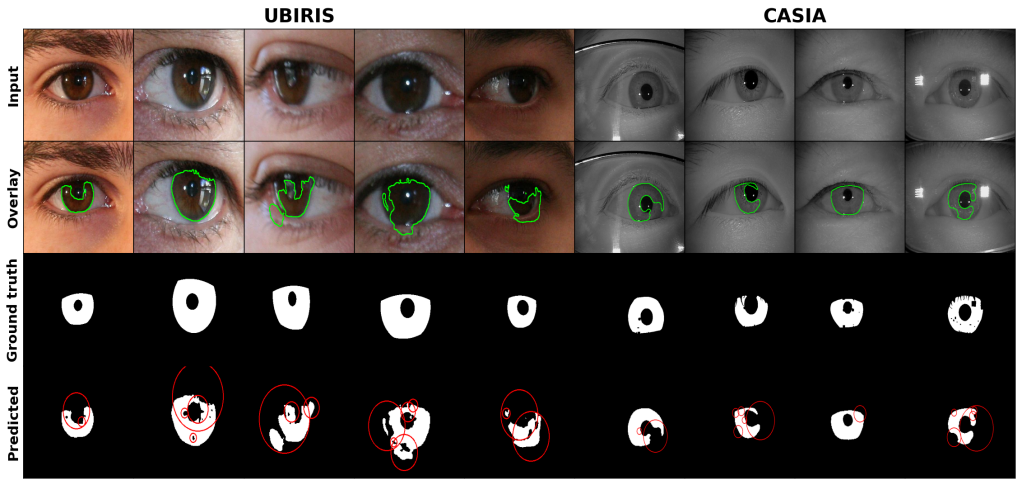



**Figure 12.** Segmentation results on challenging samples (UBIRIS.v2 **left**, CASIA-Thousand **right**). Rows: input, predicted boundary overlay (green), ground-truth mask, and predicted mask, with deviations from the ground truth circled in red.

## 6. Discussion

This section provides a critical analysis of the segmentation performance of SwinIrisNet in relation to both conventional model-based approaches [14,19,34,42–51] and contemporary deep learning-based methods [1,9,11,23,27,29,35,42,43,52–59], evaluated under identical experimental conditions using the same benchmark datasets. The majority of methods selected from the literature employ computational resources comparable to or greater than those utilized for the proposed SwinIrisNet architecture. The quantitative results of these comparisons are consolidated in Tables 6–9. Overall, the empirical results show that SwinIrisNet achieves performance comparable to that of the highest-performing methods [1,57] reported in the current literature. The subsequent discussion contextualizes these findings by examining both the merits and inherent constraints of the proposed method. To ensure a comprehensive evaluation, four iris datasets are employed, namely UBIRIS.v2, IITD, CASIA-Thousand, and MMU.v1, each presenting distinct acquisition challenges that collectively span a broad spectrum of real-world imaging variability.

### *6.1. Comparative Evaluation Against Existing Iris Segmentation Methods*

On the UBIRIS.v2 dataset, our proposed algorithm outperforms all traditional algorithms, including GST [42], OSIRIS [43], Wahet [44], Ifpp [34], and Caht [45], across all metrics. Compared to the GST method, which is the best-performing traditional algorithm, SwinIrisNet improves recall, F1, and precision by 53.41%, 56.79%, and 57.41%, respectively. Compared to Ifpp, SwinIrisNet improves NICE1, NICE2, and F1 by 23.27%, 35.92%, and 67.73%, respectively. Compared with the method of Chen et al. [27], the proposed method improves accuracy and F1 by 0.04% and 0.66%, respectively, while NICE1 is equal at 0.52%, although our method scores 1.74% lower in NICE2. Compared with FD-UNET [22], SwinIrisNet improves F1, NICE1, and NICE2 by 1.91%, 2.06%, and 91.03%, respectively.

On the IITD dataset, among the selected conventional iris segmentation methods, OSIRIS exhibits the best performance. Compared to OSIRIS, SwinIrisNet demonstrates improvements of 10.47%, 2.29%, 4.39%, 4.81%, and 4.28% in precision, recall, F1, NICE1, and NICE2, respectively. Among the evaluated CNN-based approaches, the method proposed by Y. Chen et al. [27] demonstrates the strongest overall performance. Relative to this baseline, SwinIrisNet improves accuracy by 0.69% and NICE1 by 0.41%, while exhibiting marginal reductions of 1.5% in F1 and 1.918% in NICE2.

**Table 6.** Comparison with existing iris segmentation algorithms on the UBIRIS.v2 iris dataset.

| Category | Method | Precision | Accuracy | mIoU | Recall | F1 | NICE1 (%) | NICE2 (%) |
|---|---|---|---|---|---|---|---|---|
| Traditional Methods | GST [42] | 0.3931 | – | – | 0.4259 | 0.3993 | – | – |
| | OSIRIS [43] | 0.2276 | – | – | 0.2740 | 0.2368 | – | – |
| | Wahet [44] | – | – | – | – | 0.1977 | 27.43 | 44.98 |
| | Caht [45] | – | – | – | – | 0.1048 | 12.26 | 48.09 |
| | Ifpp [34] | – | – | – | – | 0.2899 | 23.79 | 39.70 |
| CNN-Based Methods | FD-UNET [22] | – | – | – | – | 0.9481 | 0.40 | 2.58 |
| | FISNET [23] | 0.937 | 98.4 | 0.8188 | 0.848 | 0.8877 | – | – |
| | IrisParseNet [52] | 0.935 | 97.9 | 0.7265 | 0.751 | 0.8299 | – | – |
| | M. R. Sumi et al. [35] | 0.9089 | – | 0.9178 | – | – | – | – |
| | N. A. Mashudi et al. [25] | – | 93.66 | – | – | 0.9456 | – | – |
| | Y. Chen et al. [27] | 0.9592 | 99.47 | – | 0.9620 | 0.9606 | 0.52 | 2.04 |
| | MS-Net [28] | 0.9618 | 96.13 | – | 0.9618 | 0.9616 | – | – |
| | Y. Meng and T. Bao [1] | – | – | 0.9566 | – | 0.9774 | – | 1.96 |
| | OSIRIS [43] | 0.1729 | – | – | 0.2646 | 0.1865 | – | – |
| | FCEDNs-original [53] | – | – | – | – | 0.7691 | – | 12.49 |
| | FCEDNs-basic [53] | – | – | – | – | 0.7700 | – | 11.17 |
| | FCEDNs-bayesian-b [53] | – | – | – | – | 0.8407 | – | 11.16 |
| | Linknet [54] | – | – | 0.9195 | – | 0.9567 | – | 3.16 |
| | PixlSegNet [55] | 0.965 | 98.2 | 0.7872 | 0.796 | 0.8635 | – | – |
| | DeepLabV3 [56] | – | – | 0.7024 | – | 0.8755 | – | – |
| | RTV-L [56] | – | – | 0.7401 | – | 0.8597 | – | – |
| | DMS-UNet [57] | – | – | 0.9474 | – | 0.9725 | – | 2.48 |
| | UNet [58] | – | – | 0.9362 | – | 0.9553 | – | – |
| Proposed | SwinIrisNet | 0.9565 | 99.51 | 0.9300 | 0.9600 | 0.9672 | 0.590 | 2.070 |

On the CASIA-Thousand dataset, the proposed SwinIrisNet achieves an accuracy of 99.51%, outperforming most traditional approaches [14,19,44,49,50] and achieving performance comparable to that of Masek [48]. Among the CNN-based segmentation methods, the approach of Y. Chen et al. [27] represents the current state of the art. Relative to this baseline, SwinIrisNet yields an accuracy improvement of 0.67% and an NICE1 improvement of 0.58%, while scoring 2% and 2.35% lower in F1 and NICE2, respectively. Compared with the three FCEDN variants proposed by Jalilian and Uhl [53], SwinIrisNet outperforms them across all evaluated metrics.

On the MMU.v1 dataset, SwinIrisNet achieves 99.52% accuracy, exceeding the traditional methods RCHT [6], IT + RCHT [6], and IDO [30], which yield accuracies of 98.22%, 99.11%, and 97.78%, respectively. Among the deep learning baselines, the proposed model likewise outperforms U-Net [37], LinkNet [37], and FD-DenseNet [37], whose accuracies range from 99.01% to 99.10%.

Figure 13 presents a performance comparison between the proposed SwinIrisNet and several established iris segmentation approaches on the UBIRIS.v2 and IITD benchmarks. On UBIRIS.v2, SwinIrisNet achieves the highest F1 of 0.9662 and a competitive mIoU of 0.93, substantially outperforming earlier architectures. On the IITD dataset, SwinIrisNet produces an mIoU of 0.9450 alongside a strong F1 of 0.9650 that ranks among the highest reported values for this benchmark.

### *6.2. Comparison with the Hybrid Swin–CNN Approach of Meng and Bao [1]*

SwinIrisNet shares the two-branch Swin–CNN approach of Meng and Bao [1] but differs in architecture, training, data handling, and evaluation scope; these differences summarized in Table 10 must be considered when interpreting the reported results.

**Table 7.** Comparison with existing iris segmentation algorithms on the IITD iris dataset.

| Category | Method | Prec. | Acc. | mIoU | Recall | F1 | Dice | NICE1 (%) | NICE2 (%) |
|---|---|---|---|---|---|---|---|---|---|
| Traditional Methods | GST [42] | 0.8586 | – | – | 0.9006 | 0.8660 | – | – | – |
| | OSIRIS [43] | 0.9101 | – | – | 0.9406 | 0.9223 | – | 5.55 | 7.57 |
| | Wahet [44] | – | – | – | – | – | – | 13.77 | 17.62 |
| | Gangwar et al. [46] | 0.9370 | – | – | 0.9533 | 0.9437 | – | – | – |
| | Ifpp [34] | – | 97.60 | – | – | – | – | – | – |
| | S. Ahmad et al. [47] | – | – | – | – | 0.9520 | – | – | – |
| | Caht [45] | 0.7887 | – | – | 0.9660 | 0.8628 | – | 11.38 | 15.16 |
| | Masek [48] | 0.9045 | – | – | 0.8283 | 0.8530 | – | – | – |
| CNN-Based Methods | FCN [9] | – | – | – | – | 0.9744 | – | – | – |
| | GAN [9] | – | – | – | – | 0.9584 | – | – | – |
| | S. K. Balasubramanian et al. [60] | 0.9703 | 99.42 | – | – | 0.9805 | – | – | – |
| | Y. Chen et al. [27] | 0.9818 | 98.84 | – | 0.9806 | 0.9812 | – | 1.15 | 1.37 |
| | Y. Meng and T. Bao [1] | – | – | 0.9609 | – | 0.9800 | – | – | 2.12 |
| | OSIRIS [43] | – | – | – | – | 0.9220 | – | – | – |
| | FCEDNs-original [53] | – | – | – | – | 0.8661 | – | 5.91 | 6.59 |
| | FCEDNs-basic [53] | – | – | – | – | 0.8892 | – | 5.39 | 5.94 |
| | FCEDNs-bayesian-b [53] | – | – | – | – | 0.8489 | – | 6.82 | 7.01 |
| | Linknet [54] | – | – | 0.9595 | – | 0.9793 | – | – | 1.18 |
| | PixlSegNet [55] | 0.9832 | 98.41 | – | 0.9852 | 0.9841 | – | – | – |
| | DMS-UNet [57] | – | – | 0.9603 | – | 0.9797 | – | – | 1.76 |
| Proposed | SwinIrisNet | 0.9633 | 99.53 | 0.9450 | 0.9635 | 0.9662 | 0.9489 | 0.60 | 2.24 |

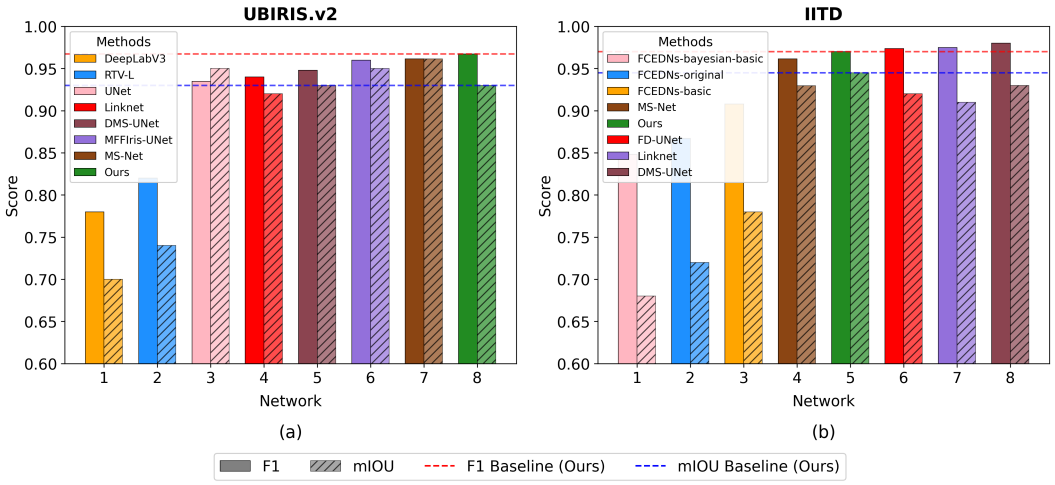



**Figure 13.** Comparison with existing CNN-based iris segmentation algorithms using F1 and mIoU metrics on two datasets: (**a**) UBIRIS.v2, (**b**) IITD.

Architecture and efficiency: SwinIrisNet uses a Swin embedding dimension of 48 (versus 96 in [1]) and channel-wise cross-attention fusion in place of their MFIEM/CAMM modules. Since Transformer cost scales quadratically with embedding width, SwinIrisNet's Transformer branch requires approximately one-third of the computation per image, while its $4 \times 256 \times 256$ configuration consumes $\sim 12\times$ less activation memory per step than the $32 \times 224 \times 224$ setting of [1]. The smaller batch entails roughly eight times more optimization steps per epoch. SwinIrisNet comprises $\sim$13.58 M parameters and 20 GFLOPs as reported in Table 11 at $256 \times 256$; Meng and Bao do not report complexity metrics.

**Table 8.** Quantitative benchmarking of SwinIrisNet against established segmentation approaches on the CASIA-Thousand dataset.

| Category | Method | Prec. | Acc. | mIoU | Recall | F1 | Dice | NICE1 (%) | NICE2 (%) |
|---|---|---|---|---|---|---|---|---|---|
| Traditional Methods | R. Bharathi et al. [49] | – | 98.83 | – | – | – | – | – | – |
| | MCHT [50] | – | 97.02 | – | – | – | – | – | – |
| | CHT [50] | – | 73.38 | – | – | – | – | – | – |
| | Integro-Differential [14] | – | 98.81 | – | – | – | – | – | – |
| | Hough Transform [14] | – | 95.77 | – | – | – | – | – | – |
| | K. Nguyen et al. [14] | – | 99.10 | – | – | – | – | – | – |
| | Masek [48] | – | 96.51 | – | – | – | – | – | – |
| | S. Ziauddin et al. [19] | – | 99.50 | – | – | – | – | – | – |
| CNN-Based Methods | FCN+MCNN [59] | – | 95.63 | – | – | – | – | – | – |
| | YH. Li et al. [29] | – | 97.70 | 0.989 | – | – | – | – | – |
| | R. Gao & T. Bourlai [11] | – | 99.56 | – | – | – | – | – | – |
| | FISNETC [23] | 0.971 | 99.40 | 0.945 | 0.970 | – | – | – | – |
| | IrisParseNet [52] | 0.978 | 96.80 | 0.610 | 0.618 | 0.748 | – | – | – |
| | U-Net [35] | – | – | 0.9508 | – | 0.9507 | – | – | – |
| | U-Net-W [35] | – | – | 0.9465 | – | 0.9457 | – | – | – |
| | U-Net++ [35] | – | – | 0.9526 | – | 0.9522 | – | – | – |
| | FCN [9] | – | – | – | – | 0.9442 | – | – | – |
| | GAN [9] | – | – | – | – | 0.9538 | – | – | – |
| | Y. Chen et al. [27] | 0.9818 | 98.84 | – | 0.9806 | 0.9812 | – | 1.15 | 1.37 |
| | Y. Meng & T. Bao [1] | – | – | 0.9609 | – | 0.9800 | – | – | 2.12 |
| | FCEDNs-original [53] | – | – | – | – | 0.8661 | – | 5.91 | 6.59 |
| | FCEDNs-basic [53] | – | – | – | – | 0.8892 | – | 5.39 | 5.94 |
| | FCEDNs-bayesian-b [53] | – | – | – | – | 0.8489 | – | 6.82 | 7.01 |
| | Linknet [54] | – | – | 0.9595 | – | 0.9793 | – | – | 1.18 |
| | PixlSegNet [55] | 0.990 | 98.90 | 0.874 | 0.880 | 0.929 | – | – | – |
| | OSIRIS [43] | – | – | 0.8851 | – | 0.8778 | – | – | – |
| Proposed | SwinIrisNet | 0.9565 | 99.51 | 0.9266 | 0.9672 | 0.9612 | 0.9519 | 0.60 | 1.82 |

**Table 9.** Comparison with existing iris segmentation algorithms on the MMU.v1 dataset.

| Category | Methods | Accuracy |
|---|---|---|
| Traditional Methods | RCHT [6] | 98.22 |
| | IT + RCHT [6] | 99.11 |
| | IDO [30] | 97.78 |
| CNN-Based Methods | U-Net [37] | 99.10 |
| | LinkNet [37] | 99.03 |
| | FD-DenseNet [37] | 99.01 |
| **Proposed** | **SwinIrisNet** | **99.52** |

Note: Bold marks the best accuracy (our SwinIrisNet).

Loss and augmentation: Meng and Bao optimize a pure Dice loss without reporting Dice values, whereas SwinIrisNet employs a composite Dice, binary cross-entropy, and boundary loss for precise iris–sclera and iris–pupil delineation. SwinIrisNet further applies extensive on-the-fly augmentation (see Table 3), approximately doubling the effective training data per epoch. Meng and Bao train without augmentation, which reduces training complexity but increases overfitting risk given the limited dataset sizes; our augmentation strategy directly targets robustness to appearance and geometric variation.

**Table 10.** Configuration comparison with Meng and Bao [1].

| Aspect | SwinIrisNet (Ours) | Meng and Bao |
|---|---|---|
| Input/Swin embed dim. | $256 \times 256/48$ | $224 \times 224/96$ |
| Branch fusion | Channel-wise cross-attention | MFIEM + CAMM |
| Optimizer/LR | AdamW/$10^{-4}$ warmup-cosine | Adam/$10^{-3}$ fixed |
| Epochs/batches | 50/4 | 50/32 |
| Loss | Dice + BCE + boundary | Dice only |
| Augmentation | Yes (Table 3) | None |
| Datasets/metrics | 4 datasets, full metric set | 2 datasets, 3–4 metrics |
| GPU (24 GB) | NVIDIA A10G | RTX 3090 |

On the two shared datasets, Meng and Bao's method outperforms SwinIrisNet in three of the four reported metrics, while SwinIrisNet outperforms [1] in NICE2; on average, all differences remain within 1.5% (Tables 6 and 8). These marginal differences are consistent with our use of a boundary-aware composite loss and a lighter model, accompanied by broader metric and dataset coverage.

Cross-dataset generalization: Meng and Bao assess generalization only qualitatively through visual examples without providing quantitative cross-dataset results, whereas SwinIrisNet reports quantitative transfer performance (Table 12). Overall, SwinIrisNet trades a small margin in in-dataset overlap for markedly lower memory cost, a boundary-aware loss, augmentation-driven robustness, broader metric reporting, and quantified cross-dataset generalization.

### *6.3. Memory Efficiency and Comparison with Existing Methods*

As shown in Table 11, SwinIrisNet (13.58M parameters, 20.0 GFLOPs) is 2.5× lighter than U-Net [22], 3.2× smaller than ResNet-50 U-Net [61], 2.3× more compact than IrisParseNet [52], and uses 24% fewer parameters than DeepLabV3+ [56]. The method proposed by C. Wang et al. [52] achieves comparable FLOPs but requires 2.3× more parameters. With a CPU inference time of 183 ms, SwinIrisNet is 3.6× faster than U-Net and 2.4× faster than DeepLabV3+. IrisParseNet and C. Wang et al.'s method achieve slightly lower latency (150 ms) owing to their single-branch designs, but both require more than double the parameters. This trade-off positions SwinIrisNet as a memory-efficient architecture well-suited for resource-constrained deployment scenarios where limited computational resources are available.

**Table 11.** Computational resource consumption of existing baseline architectures and SwinIrisNet ablation variants in terms of parameters (M), FLOPs (G), and CPU inference time (ms).

| Method | Params (M) | FLOPs (G) | Infer (ms, CPU) |
|---|---|---|---|
| U-Net [22] | 33.53 | 66.51 | 650 |
| ResNet-50 U-Net [61] | 43.9 | ∼80 | — |
| DeepLabV3+ [56] | 17.78 | 22.87 | 440 |
| IrisParseNet [52] | 31.28 | 263.56 | 150 |
| C. Wang et al. [52] | 31.2 | 19.6 | 150 |
| Ablation variants (this work) | | | |
| Swin Transformer | 7.08 | 6.3 | 208 |
| CNN | 3.71 | 10.7 | 235 |
| SwinIrisNet $w/o\ L_b$ | 12.34 | 20.0 | 201 |
| SwinIrisNet | 13.58 | 20.0 | 183 |

### *6.4. Cross-Dataset Generalization*

To thoroughly assess the cross-dataset generalization ability of the proposed SwinIrisNet, two cross-domain inference experiments were conducted, in which the trained model was applied to previously unseen iris images originating from a dataset entirely excluded from the training process.

#### 6.4.1. Cross-Dataset Qualitative Generalization Analysis

In the first experiment, SwinIrisNet, trained exclusively on the UBIRIS.v2 dataset (visible wavelength, RGB), was evaluated on unseen samples from the CASIA-Thousand dataset (near-infrared), as depicted in Figure 14. In the second experiment, the model trained solely on CASIA-Thousand was applied to unseen UBIRIS.v2 images (visible wavelength, RGB), as illustrated in Figure 15.

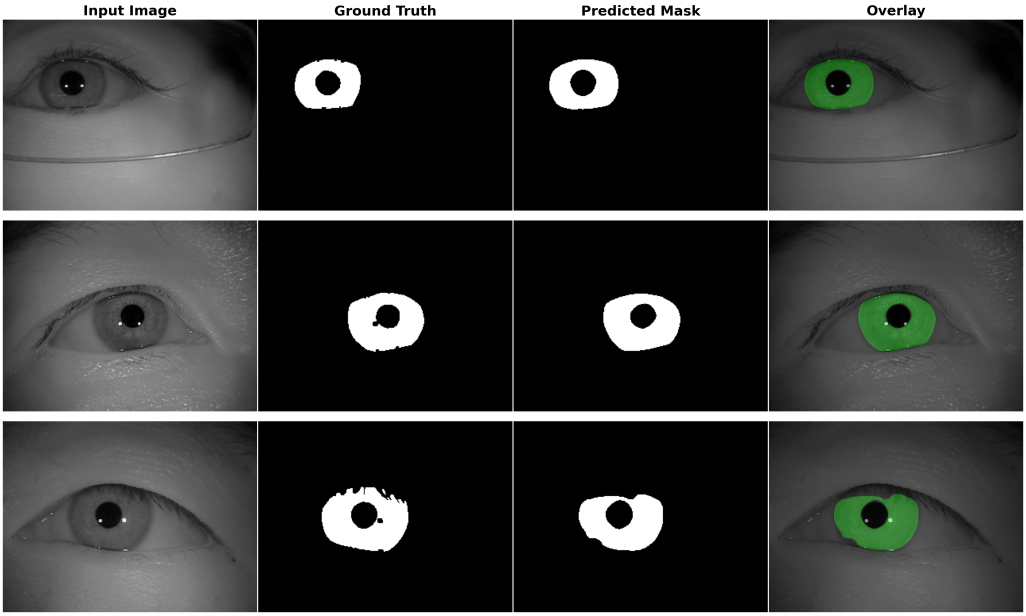



**Figure 14.** Segmentation of unseen CASIA iris images using SwinIrisNet trained on the UBIRIS.v2 dataset.

Both experimental scenarios yield high-fidelity segmentation results, as evidenced by the close similarity between the predicted masks and ground-truth annotations. These findings substantiate that the proposed SwinIrisNet generalizes effectively across disparate imaging modalities (near-infrared and visible-light), heterogeneous sensor configurations, and divergent acquisition conditions without necessitating retraining or domain-specific robust fine-tuning. This cross-dataset robustness is attributable to the hybrid encoder architecture of SwinIrisNet, wherein the Swin Transformer branch captures domain-invariant global iris structural representations while the convolutional branch preserves fine-grained local boundary features that remain consistent across imaging wavelengths.

#### 6.4.2. Cross-Dataset Quantitative Generalization Analysis

Table 12 summarizes the cross-dataset generalization performance of the proposed SwinIrisNet across eight train–test combinations involving four datasets: UBIRIS.v2, CASIA-Thousand, IITD, and MMU.v1. Three key observations emerge from the obtained results.

First, models trained on CASIA-Thousand, IITD, and MMU.v1 all transfer effectively to UBIRIS.v2, consistently achieving Dice scores above 0.96, mIoU values exceeding 0.92, and NICE1 values below 0.51%. This strong generalization suggests that features learned

from NIR imagery capture iris representations that remain discriminative under the visible-wavelength acquisition conditions of UBIRIS.v2. The low standard deviations ($\pm 0.02$) further confirm stable performance across the entire UBIRIS.v2 test set.

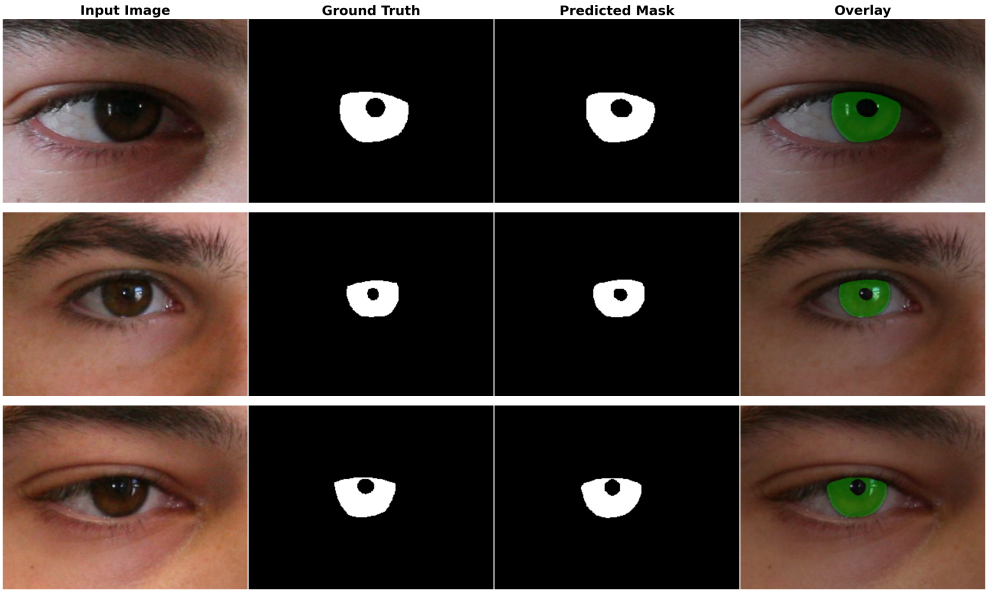



**Figure 15.** Segmentation of unseen UBIRIS.v2 iris images using SwinIrisNet trained on the CASIA-Thousand dataset.

Second, transfer in the opposite direction is notably asymmetric. UBIRIS.v2 to CASIA-Thousand achieves a Dice of $0.57 \pm 0.37$, while IITD to CASIA-Thousand and MMU.v1 to CASIA-Thousand yield Dice scores of $0.47 \pm 0.35$ and $0.55 \pm 0.36$, respectively. The elevated standard deviations indicate bimodal segmentation behavior: the model segments certain CASIA subjects successfully while failing on others whose NIR-specific characteristics fall outside the scope of the training data.

Third, all models trained on the remaining three datasets (IITD, UBIRIS.v2, and CASIA-Thousand) yield limited segmentation performance when evaluated on MMU.v1 test samples (Dice, 0.40–0.41; mIoU, 0.27–0.28). The relatively low variance ($\pm 0.17$) indicates consistent under-segmentation rather than sporadic failure, likely due to MMU.v1's smaller iris-to-frame ratio and its distinct sensor acquisition characteristics.

Regarding boundary precision, NICE2 remains elevated across all cross-dataset pairs (32–91%), even for the successful transfers to UBIRIS.v2. For example, CASIA to UBIRIS achieves a Dice of 0.96 but an NICE2 of 32%, compared to a within-dataset NICE2 of 1.82%. This large difference indicates that while the model accurately localizes the iris region, fine boundary alignment degrades when sensor-specific edge characteristics differ between training and test distributions. The loss weights $(\alpha_{\text{BCE}}, \beta_{\text{Dice}}, \gamma_{\text{Boundary}}) = (0.35, 0.40, 0.25)$ were tuned on the validation split and fixed across datasets. No sensitivity analysis perturbing $(\alpha, \beta, \gamma)$ against NICE1/NICE2 or automatic weight balancing was employed, and this will be considered for future work.

The cross-dataset generalization results demonstrate that SwinIrisNet transfers effectively across datasets sharing similar acquisition characteristics. When evaluated across NIR iris datasets and on the visible-wavelength target (UBIRIS dataset), the model achieves Dice, mIoU, and F1 values of approximately 0.96 with correspondingly low NICE1 errors (0.5%), confirming that the learned representations generalize well when source and target domains share comparable imaging conditions. The reduced performance observed for certain source-target pairs reflects the pronounced domain gap between visible-wavelength and near-infrared imagery, compounded by inter-dataset differences in spatial resolution,

standoff distance, illumination, occlusion patterns, and annotation conventions. This degradation is consistent with well-documented behavior of deep segmentation networks under distribution shift and is not architecture-specific. The elevated NICE1 and NICE2 scores in these cases indicate systematic under- or over-segmentation when the target modality is underrepresented during training, rather than stochastic pixel-level errors [62–64]. Nevertheless, even under these challenging cross-modality conditions, the model partially segments a substantial proportion of images. We have included this as a limitation of our method in Section 6.5.

**Table 12.** Cross-dataset generalization results of SwinIrisNet.

| Train → Test | Dice | mIoU | F1 | NICE1 (%) | NICE2 (%) | N-Sample |
|---|---|---|---|---|---|---|
| UBIRIS → CASIA | 0.5704 ± 0.3696 | 0.4878 ± 0.3472 | 0.5704 ± 0.3696 | 3.543 ± 2.764 | 68.469 ± 11.690 | 100 |
| UBIRIS → MMU | 0.4097 ± 0.1732 | 0.2738 ± 0.1505 | 0.4097 ± 0.1732 | 9.189 ± 2.854 | 89.981 ± 4.408 | 49 |
| CASIA → UBIRIS | 0.9609 ± 0.0205 | 0.9255 ± 0.0365 | 0.9609 ± 0.0205 | 0.505 ± 0.351 | 32.031 ± 10.789 | 150 |
| CASIA → MMU | 0.4113 ± 0.1690 | 0.2740 ± 0.1429 | 0.4113 ± 0.1690 | 9.354 ± 2.936 | 90.870 ± 4.099 | 49 |
| IITD → UBIRIS | 0.9607 ± 0.0211 | 0.9251 ± 0.0375 | 0.9607 ± 0.0211 | 0.503 ± 0.343 | 31.814 ± 10.587 | 150 |
| IITD → CASIA | 0.4709 ± 0.3496 | 0.3776 ± 0.3082 | 0.4709 ± 0.3496 | 4.091 ± 2.365 | 77.161 ± 9.025 | 100 |
| MMU → UBIRIS | 0.9610 ± 0.0204 | 0.9257 ± 0.0366 | 0.9610 ± 0.0204 | 0.499 ± 0.351 | 31.566 ± 10.593 | 150 |
| MMU → CASIA | 0.5523 ± 0.3620 | 0.4654 ± 0.3395 | 0.5523 ± 0.3620 | 3.537 ± 2.568 | 72.556 ± 11.323 | 100 |

### *6.5. Challenges and Limitations*

Despite its competitive performance, SwinIrisNet has some limitations. Its memory-optimized design with an embedding dimension of 48 and reduced CNN and decoder filters limits representational capacity compared to full-scale backbones such as Swin-Tiny or ResNet-50, which may affect performance on larger and more diverse datasets. The fixed 256 × 256 input resolution also introduces aspect ratio distortion on datasets such as IITD and MMU.v1, which can slightly reduce boundary smoothness. Resizing the predicted masks back to the native resolution through post-processing interpolation can introduce quantization artifacts along the iris contours, potentially degrading boundary precision. Additionally, we acknowledge the gap in cross-dataset generalization as a limitation of single-dataset training and plan to address it through domain adaptation and multi-modal training strategies in future work. Nonetheless, SwinIrisNet remains well-suited for biometric iris recognition, access control, forensic identification, and medical eye imaging applications where accurate iris boundary delineation is essential.

### *6.6. Downstream Recognition Validation*

To validate that segmentation quality extends beyond pixel-level metrics (Dice, mIoU) and translates into practical recognition performance, we constructed a complete iris recognition pipeline on top of the predicted masks and evaluated it on the near-infrared (CASIA-Thousand) and visible-light (UBIRIS.v2) test splits.

The pipeline follows the classical recognition stages. Each segmented iris is normalized into a fixed polar representation using Daugman's rubber-sheet model, unwrapping the annular region between the inner (pupil) and outer (limbus) boundaries into a 64 × 512 strip. The strip is encoded with a bank of multi-scale one-dimensional Log-Gabor filters whose phase response is quantized into a binary IrisCode. Matching uses the fractional Hamming distance with an angular bit-shift search to compensate for in-plane eye rotation, comparing only mutually valid (non-occluded) bits. Eyelashes and specular reflections, together with the eyelid sectors, are excluded through an occlusion mask.

Recognition is evaluated on samples of 100 images from the test splits. Genuine comparisons are formed between images sharing the same subject and eye, and impostor comparisons are formed between all remaining pairs; the left and right irises of a subject

are treated as distinct identities, following standard iris-recognition practice. We report the equal error rate (EER), the decidability index $d'$, and the rank-1 correct recognition rate (CRR). Importantly, the encoder and matcher are held fixed, so that the only variable is the source of the iris mask: we compare recognition driven by the ground-truth masks against recognition driven by the masks predicted by the proposed segmentation network. This isolates the effect of segmentation quality on downstream recognition.

On both test splits, the predicted masks closely match the ground truth, with a mean Dice of 0.955 (minimum above 0.90; no sample below 0.80). This high agreement is the prerequisite for a meaningful recognition comparison: if the predicted iris regions align with the ground truth, downstream recognition should be preserved.

Table 13 reports the recognition performance for the ground-truth and predicted masks on both the CASIA-Thousand and UBIRIS.v2 datasets. Recognition driven by the predicted masks is similar to recognition driven by the ground-truth masks: on CASIA-Thousand, the equal error rates are 6.19% (GT) and 6.71% (predicted), and on UBIRIS.v2, the predicted-mask pipeline attains an EER and a CRR on par with, or better than, the ground-truth pipeline. In all cases, the impostor Hamming-distance distribution is centered near 0.485, close to the value of 0.5 expected for uncorrelated iris codes, while the genuine distribution is well separated. These results confirm that the segmentation output preserves the identity information required for recognition.

**Table 13.** Downstream iris recognition: comparison of ground-truth (GT) and predicted masks under an identical encoder and matcher.

| Dataset | Mask Source | EER (%) | $d'$ | Rank-1 CRR (%) |
|---|---|---|---|---|
| CASIA-Thousand (NIR) | Ground truth | 6.19 | 2.40 | 83.7 |
| CASIA-Thousand (NIR) | Predicted | 6.71 | 2.19 | 88.5 |
| UBIRIS.v2 (visible) | Ground truth | 10.43 | 2.23 | 67.4 |
| UBIRIS.v2 (visible) | Predicted | 4.64 | 2.61 | 90.2 |

Figure 16 illustrates the recognition pipeline on representative CASIA-Thousand samples: the input eye, the segmented iris with fitted inner and outer boundaries, the normalized rubber-sheet strip, and the binary IrisCode. Figure 17 shows example verifications, where a probe is matched against the gallery and the nearest identity is retrieved.

These results confirm that the proposed segmentation supports accurate downstream recognition: driven by the predicted masks, the pipeline attains single-digit to low double-digit equal error rates and rank-1 recognition rates that match those obtained with ground-truth masks, with impostor scores concentrated near the uncorrelated baseline of 0.5. The high mask fidelity (mean Dice, 0.955) explains this preservation of performance.

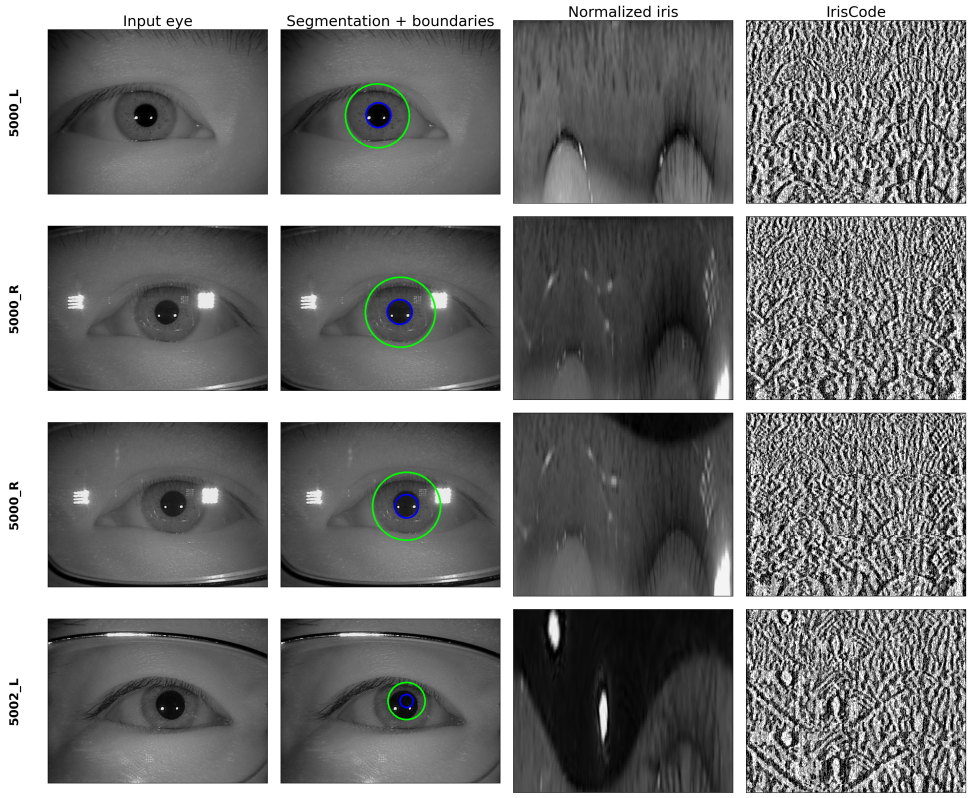



**Figure 16.** Recognition pipeline stages: input eye, segmented iris with fitted pupil/limbus boundaries (blue = inner pupil boundary, green = outer limbus/iris boundary), normalized iris strip, and IrisCode template.

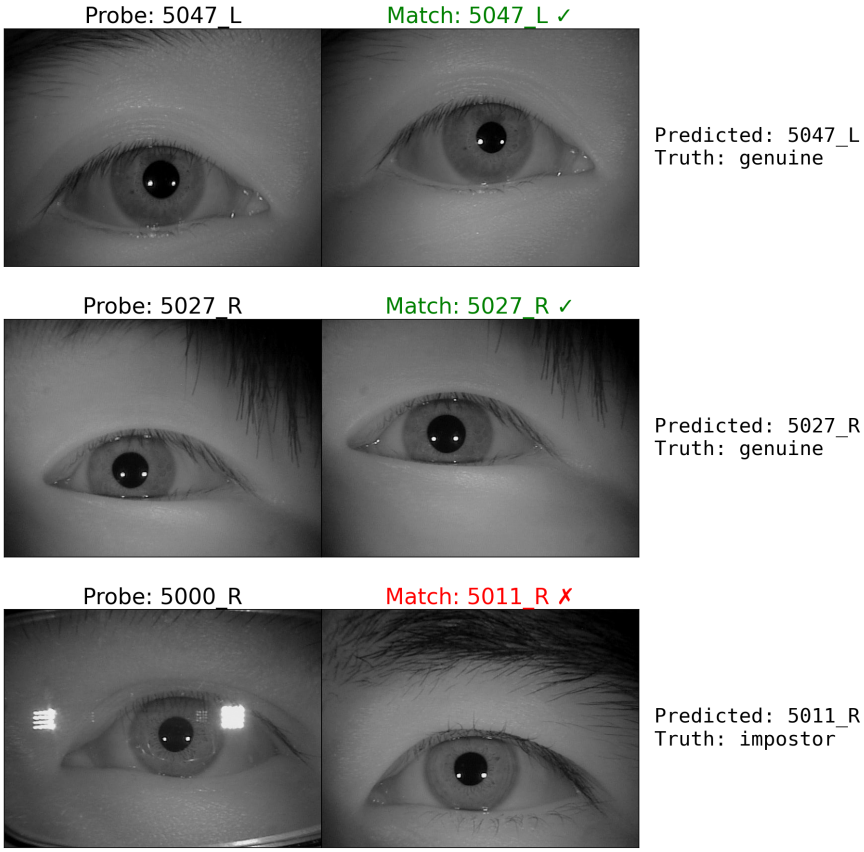



**Figure 17.** Example verifications on CASIA. Each probe is matched to the nearest gallery identity. "✓" indicates a correct verification and "×" an incorrect one, relative to the genuine/impostor ground truth.

## 7. Conclusions and Future Studies

The accuracy of iris segmentation directly governs the reliability of downstream recognition systems. In this study, we proposed SwinIrisNet, a memory-optimized hybrid deep learning architecture that combines Swin Transformer and convolutional neural networks within a U-Net framework. The Swin Transformer branch captures long-range contextual dependencies through hierarchical window-based self-attention, while the CNN branch extracts fine-grained local features essential for precise boundary delineation. A memory-efficient cross-attention fusion module enhanced by CBAM enables refined feature selection across both branches, and the decoder integrates ASPP for multi-scale context aggregation with attention gates for improved skip connections.

SwinIrisNet was evaluated on four publicly available iris datasets (UBIRIS.v2, IITD, CASIA-Thousand, and MMU.v1), encompassing challenging imaging conditions such as illumination variations, occlusions, and off-axis gaze angles. Performance was assessed using multiple quantitative and qualitative metrics, including F1, precision, accuracy, Dice coefficient, recall, specificity, loss, mIoU, and NICE1/NICE2. Experimental results demonstrate that SwinIrisNet outperforms several existing methods across multiple evaluation metrics while achieving competitive performance against the best-performing approaches in the field. Cross-dataset evaluation further indicates that the proposed architecture exhibits promising transferability across heterogeneous datasets. Furthermore, the segmentation achieved by SwinIrisNet translates into strong downstream recognition performance on the test sets, confirming that the predicted masks preserve sufficient identity-discriminative information for reliable iris recognition. The memory-efficient design makes SwinIrisNet particularly suitable for biometric and medical imaging applications operating under constrained computational resources.

We acknowledge that cross-dataset generalization remains constrained by single-dataset training. In future work, we intend to explore network pruning, noise-robust boundary formulations (e.g., distance-transform or Hausdorff-based losses), and domain adaptation strategies to further reduce the computational footprint and improve cross-dataset generalization, particularly in the visible-to-NIR transfer direction.

**Author Contributions:** Conceptualization, T.L.O.; methodology, T.L.O.; software, T.L.O.; validation, T.L.O. and K.A.O.; formal analysis, T.L.O.; investigation, T.L.O. and K.A.O.; resources, T.L.O. and K.A.O.; data curation, T.L.O.; writing—original draft preparation, T.L.O.; writing—review and editing, T.L.O. and K.A.O.; visualization, T.L.O.; supervision, K.A.O.; project administration, K.A.O.; funding acquisition, K.A.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the University of Johannesburg: "University of Johannesburg/Research Fund; KA_Ogudo/UJ/DEET/2024_Research Cost Center, UJ Library_APC-2026_South Africa". The APC was funded by a grant from the University of Johannesburg Library Research Funds (UJ).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets used in this study are publicly available. The UBIRIS.v2 dataset is available upon request from University of Beira Interior, Covilhã, Portugal at https://iris.di.ubi.pt/ubiris2AppForm.pdf, accessed on 31 July 2023. The IITD iris dataset is available from the Indian Institute of Technology Delhi. The CASIA-Thousand dataset is available from the Chinese Academy of Sciences Institute of Automation. The MMU.v1 dataset is available from the Multimedia University, Malaysia.

**Acknowledgments:** This study utilizes several publicly available iris image datasets for experimental evaluation. The datasets employed include the CASIA collection, made accessible by the Chinese

Academy of Sciences' Institute of Automation; the IITD iris dataset, developed at the Indian Institute of Technology Delhi; the UBIRIS.v2 collection, maintained by the Computer Science Department at the University of Beira Interior; and the MMU.v1 dataset, created by the Multimedia University (MMU) in Malaysia. The authors acknowledge the use of Amazon Quick (version: 0.1000.20017) and Grammarly v.1.174.1.0 as AI-assisted tools for language refinement and grammatical correction to enhance the clarity of this manuscript. All scientific content, experimental design, analysis, interpretation, and conclusions remain solely the work of the authors.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A

**Table A1.** Complete hyperparameter settings for SwinIrisNet (identical across datasets unless noted).

| Group | Hyperparameter | Value |
|---|---|---|
| Swin Transformer branch | Embedding dimension | 48 |
| | Stage depths | (2, 2, 4, 2) |
| | Attention heads | (3, 6, 12, 24) |
| | Window size/shift size | 7/3 |
| | Attn. MLP dropout | 0.0/0.0 |
| CNN branch | Residual filters | (32, 64, 128, 256) |
| | Patch/stem size | 4 |
| Decoder/head | Decoder filters | (128, 64, 32, 16, 8) |
| | BatchNorm/dropout | Yes/0.2 |
| | Spatial dropout | 0.1 |
| | Output activation | Sigmoid |
| Loss (Equation (14)) | $\alpha_{\mathrm{BCE}}$ | 0.35 |
| | $\beta_{\mathrm{Dice}}$ | 0.40 |
| | $\gamma_{\mathrm{Boundary}}$ | 0.25 |
| | Tversky $\alpha/\beta$, focal $\gamma$ | 0.5/0.5/2.0 |
| Optimization | Optimizer | AdamW |
| | Weight decay | 0.05 |
| | Initial learning rate | $1 \times 10^{-4}$ |
| | Min learning rate | $1 \times 10^{-8}$ |
| | LR schedule/warmup | Warmup-cosine/5 epochs |
| | Gradient clip (norm) | 1.0 |
| | $L_2$ regularization | $1 \times 10^{-4}$ |
| | Early-stopping patience | 15 |
| Data/training | Input size | $256 \times 256 \times 3$ |
| | Batch size | 4 |
| | Epochs | 50 |
| | Train/val/test split | 70/15/15 |

## References

1. Meng, Y.; Bao, T. Towards More Accurate and Complete Heterogeneous Iris Segmentation Using a Hybrid Deep Learning Approach. *J. Imaging* **2022**, *8*, 246. [CrossRef] [PubMed]
2. Fu, H.; Xu, Y.; Lin, S.; Wong, D.W.K.; Baskaran, M.; Mahesh, M. Angle-Closure Detection in Anterior Segment OCT Based on Multilevel Deep Network. *IEEE Trans. Cybern.* **2020**, *50*, 3358–3366. [CrossRef] [PubMed]
3. Sumi, M.R.; Das, P.; Hossain, A.; Dey, S.; Schuckers, S. A Comprehensive Evaluation of Iris Segmentation on Benchmarking Datasets. *Sensors* **2024**, *24*, 7079. [CrossRef] [PubMed]
4. Roy, D.A.; Soni, U.S. IRIS segmentation using Daugman's method. In Proceedings of the International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai, India, 3–5 March 2016; pp. 2668–2676.
5. Tan, T.; He, Z.; Sun, Z. Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. *Image Vis. Comput.* **2010**, *28*, 223–230. [CrossRef]
6. Umer, S.; Dhara, B.C. A fast iris localization using inversion transform and restricted circular Hough transform. In Proceedings of the 8th International Conference on Advances in Pattern Recognition (ICAPR), Kolkata, India, 4–7 January 2015; pp. 1–6.

7. Khaki, A.; Aghagolzadeh, A.; Cami, B.R. ISUR: Iris Segmentation based on UNet and ResNet. In Proceedings of the 11th International Conference on Computer Engineering and Knowledge (ICCKE), Mashhad, Iran, 18–19 November 2021; pp. 1–7.
8. Zeng, Z.; Xie, W.; Zhang, Y.; Lu, Y. RIC-Unet: An Improved Neural Network Based on Unet for Nuclei Segmentation in Histology Images. *IEEE Access* **2019**, *7*, 21420–21428. [CrossRef]
9. Bezerra, C.S.; Laroca, R.; Lucio, D.R.; Severo, E.; Oliveira, L.F.; Britto, A.S. Robust Iris Segmentation Based on Fully Convolutional Networks and Generative Adversarial Networks. In Proceedings of the 31st SIBGRAPI Conference on Graphics, Patterns and Images, Parana, Brazil, 29 October–1 November 2018; pp. 281–288.
10. Ma, B.; Qiu, D.; Xiong, Z.; Hu, Y.; Deng, L.; Yuan, H. SS-SwinUnet: A Distillation Method of Swin Transformer for Superior Ocular Image Segmentation. In Proceedings of the IEEE BIBM, Lisbon, Portugal, 3–6 December 2024; pp. 2291–2296.
11. Gao, R.; Bourlai, T. On Designing a SwinIris Transformer Based Iris Recognition System. *IEEE Access* **2024**, *12*, 30723–30737. [CrossRef]
12. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 9992–10002.
13. Sundaram, D.M.; Kasukurthi, Y.; Jayachandran, S. Comparative Study of Vision Transformer (ViT), Swin Transformer, and Transformer in Transformer (TNT) on Image Classification. In Proceedings of the 2025 5th International Conference on Soft Computing for Security Applications (ICSCSA), Salem, India, 4–6 August 2025; pp. 1469–1473.
14. Nguyen, K.; Fookes, C.; Sridharan, S. Fusing shrinking and expanding active contour models for robust iris segmentation. In Proceedings of the 10th ISSPA, Kuala Lumpur, Malaysia, 13–16 May 2010; pp. 185–188.
15. Daugman, J. How iris recognition works. *IEEE Trans. Circuits Syst. Video Technol.* **2004**, *14*, 21–30. [CrossRef]
16. Koh, J.; Govindaraju, V.; Chaudhary, V. A robust iris localization method using an active contour model and Hough transform. In Proceedings of the ICPR, Istanbul, Turkey, 23–26 August 2010; pp. 2852–2856.
17. Banerjee, S.; Mery, D. Iris segmentation using geodesic active contours and GrabCut. *Lect. Notes Comput. Sci.* **2016**, *9555*, 48–60. [CrossRef]
18. Jalil, N.A.; Sahak, R.; Saparon, A. A Comparison of Iris Localization Techniques for Pattern Recognition Analysis. In Proceedings of the Sixth Asia Modelling Symposium, Bali, Indonesia, 29–31 May 2012; pp. 75–80.
19. Ziauddin, S.; Dailey, M.N. A robust hybrid iris localization technique. In Proceedings of the 6th ECTI-CON, Chonburi, Thailand, 6–9 May 2009; pp. 1058–1061.
20. Kumar, V.; Asati, A.; Gupta, A. Iris localization based on integro-differential operator for unconstrained infrared iris images. In Proceedings of the ISPCC, Waknaghat, India, 24–26 September 2015; pp. 277–281.
21. Nsaef, A.K.; Jaafar, A.; Jassim, K.N. Enhancement segmentation technique for iris recognition system based on Daugman's Integro-differential operator. In Proceedings of the 2012 International Symposium on Instrumentation & Measurement, Sensor Network and Automation (IMSNA), Sanya, China, 25–28 August 2012; pp. 71–75.
22. Zhang, W.; Lu, X.; Gu, Y.; Liu, Y.; Meng, X.; Li, J. A Robust Iris Segmentation Scheme Based on Improved U-Net. *IEEE Access* **2019**, *7*, 85082–85089. [CrossRef]
23. Sharma, G.; Jaswal, G.; Nigam, A.; Ramachandra, R. FISNET: A Learnable Fusion-Based Iris Segmentation Network Improving Robustness Across NIR and VIS Modalities. *IEEE Access* **2025**, *13*, 101472–101490. [CrossRef]
24. Tan, C.-W.; Kumar, A. Automated segmentation of iris images using visible wavelength face images. In Proceedings of the CVPR Workshops, Colorado Springs, CO, USA, 20–25 June 2011; pp. 9–14.
25. Mashudi, N.A.; Ahmad, N.; Noor, N.M. Dynamic U-Net Using Residual Network for Iris Segmentation. In Proceedings of the ICSIPA, Kuala Terengganu, Malaysia, 12–14 October 2021; pp. 117–121.
26. Sardar, M.; Banerjee, S.; Mitra, S. Iris Segmentation Using Interactive Deep Learning. *IEEE Access* **2020**, *8*, 219322–219330. [CrossRef]
27. Chen, Y.; Wang, W.; Zeng, Z.; Wang, Y. An Adaptive CNNs Technology for Robust Iris Segmentation. *IEEE Access* **2019**, *7*, 64517–64532. [CrossRef]
28. Mathur, A.; Patel, M.; Vyas, R. A Co-Generation Framework Using Progressive GAN and Pix2Pix for Enhanced Sclera–Iris–Pupil (SIP) Segmentation. *IEEE Access* **2026**, *14*, 24288–24297. [CrossRef]
29. Li, Y.-H.; Putri, W.R.; Aslam, M.S.; Chang, C.-C. Robust Iris Segmentation Algorithm in Non-Cooperative Environments Using Interleaved Residual U-Net. *Sensors* **2021**, *21*, 1434. [CrossRef] [PubMed]
30. Daugman, J.G. High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans. Pattern Anal. Mach. Intell.* **1993**, *15*, 1148–1161. [CrossRef]
31. He, Z.; Tan, T.; Sun, Z.; Qiu, X. Toward Accurate and Fast Iris Segmentation for Iris Biometrics. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 1670–1684. [CrossRef] [PubMed]

32. Sokolova, N.; Taschwer, M.; Sarny, S.; Putzgruber-Adamitsch, D.; Schoeffmann, K. Pixel-Based Iris and Pupil Segmentation in Cataract Surgery Videos Using Mask R-CNN. In Proceedings of the IEEE 17th ISBI Workshops, Iowa City, IA, USA, 3–7 April 2020; pp. 1–4.
33. Mashudi, N.A.; Ahmad, N.; Noor, N.M. Hybrid U-Net With ResNet Iris Segmentation Method for Occlusion by Eyelids and Eyelash in Iris Recognition. *IEEE J. Sel. Top. Signal Process.* **2025**, *19*, 1406–1419. [CrossRef]
34. Malgheet, J.R.; Manshor, N.B.; Affendey, L.S.; Halin, A.B.A. MS-Net: Multi-Segmentation Network for the Iris Region Using Deep Learning in an Unconstrained Environment. *IEEE Access* **2023**, *11*, 59368–59385. [CrossRef]
35. AL Qurri, A.; Almekkawy,M. Improved UNet with Attention for Medical Image Segmentation. *Sensors* **2023**, *23*, 8589. [CrossRef] [PubMed]
36. Huynh, T.-T.; Huynh, D.-T.; Duong, C.-S.; Nguyen, H.-S.; Nguyen, Q.H. Polar Topology Transformers with Anatomical Skip Connections for Efficient Iris Segmentation. *IEEE Open J. Comput. Soc.* **2026**, *7*, 1003–1014.. [CrossRef]
37. Ganeeva, Y.; Myasnikov, E. Using Convolutional Neural Networks for Segmentation of Iris Images. In Proceedings of the FarEastCon, Vladivostok, Russia, 6–9 October 2020; pp. 1–4.
38. Proenca, H.; Filipe, S.; Santos, R.; Oliveira, J.; Alexandre, L.A. The UBIRIS.v2: A Database of Visible Wavelength Iris Images Captured On-the-Move and At-a-Distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1529–1535. [CrossRef] [PubMed]
39. Kumar, A.; Passi, A. Comparison and combination of iris matchers for reliable personal authentication. *Pattern Recognit.* **2010**, *43*, 1016–1026. [CrossRef]
40. Hofbauer, H.; Alonso-Fernandez, F.; Wild, P.; Bigun, J.; Uhl, A. A Ground Truth for Iris Segmentation. In Proceedings of the 22nd ICPR, Stockholm, Sweden, 24–28 August 2014; pp. 527–532.
41. TensorFlow. Available online: https://www.tensorflow.org/ (accessed on 1 January 2024).
42. Alonso-Fernandez, F.; Bigun, J. Iris boundaries segmentation using the generalized structure tensor. A study on the effects of image degradation. In Proceedings of the IEEE 5th International Conference on Biometrics: Theory, Applications and Systems, Arlington, VA, USA, 23–25 September 2012; pp. 426–431.
43. Sutra, G.; Dorizzi, B.; Garcia-Salicetti, S.; Othman, N. *A Biometric Reference System for Iris—Osiris Version 4.1*; Télécom SudParis, France, Tech. Rep.: Évry, France, 2012.
44. Uhl, A.; Wild, P. Weighted adaptive Hough and ellipsopolar transforms for real-time iris segmentation. In Proceedings of the 5th IAPR ICB, New Delhi, India, 29 March–1 April 2012; pp. 283–290.
45. Rathgeb, C.; Uhl, A.; Wild, P. *Iris Biometrics: From Segmentation to Template Security*; Springer: New York, NY, USA, 2012; Volume 59.
46. Gangwar, A.; Joshi, A.; Singh, A.; Alonso-Fernandez, F.; Bigun, J. IrisSeg: A fast and robust iris segmentation framework for non-ideal iris images. In Proceedings of the International Conference on Biometrics (ICB), Halmstad, Sweden, 13–16 June 2016; pp. 1–8.
47. Ahmad, S.; Fuller, B. Unconstrained Iris Segmentation Using Convolutional Neural Networks. *arXiv* **2019**, arXiv:1812.08245.
48. Masek, L. Recognition of Human Iris Patterns for Biometric Identification. Available online: http://www.peterkovesi.com/studentprojects/libor/sourcecode.html (accessed on 1 January 2024).
49. Bharathi, R.; Anandaraju, M.B. Iris Recognition System Using Circular Hough Transform and HOG. In Proceedings of the ICRASET, Mandya, India, 21–22 November 2024; pp. 1–4.
50. Ismail, S.; Ali, F.H.M. Noisy Iris Segmentation Using Modified Circular Hough Transform (MCHT). In Proceedings of the 3rd ICCR, Dubai, United Arab Emirates, 3–4 July 2025; pp. 1–6.
51. Wild, P.; Hofbauer, H.; Ferryman, J.; Uhl, A. Segmentation-Level Fusion for Iris Recognition. In Proceedings of the BIOSIG, Darmstadt, Germany, 9–11 September 2015; pp. 1–6.
52. Wang, C.; Muhammad, J.; Wang, Y.; He, Z.; Sun, Z. Towards Complete and Accurate Iris Segmentation Using Deep Multi-Task Attention Network for Non-Cooperative Iris Recognition. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 2944–2959. [CrossRef]
53. Jalilian, E.; Uhl, A. Iris Segmentation Using Fully Convolutional Encoder–Decoder Networks. In *Advances in Computer Vision and Pattern Recognition*; Springer: New York, NY, USA, 2017; pp. 133–155.
54. Chaurasia, A.; Culurciello, E. LinkNet: Exploiting encoder representations for efficient semantic segmentation. In Proceedings of the IEEE VCIP, St. Petersburg, FL, USA, 10–13 December 2017.
55. Jha, R.R.; Jaswal, G.; Gupta, D.; Saini, S.; Nigam, A. PixISegNet: Pixel-level iris segmentation network using convolutional encoder–decoder with stacked hourglass bottleneck. *IET Biom.* **2020**, *9*, 11–24. [CrossRef]
56. You, X.; Zhao, P.; Mu, X.; Bai, K.; Lian, S. Heterogeneous Noise Iris Segmentation Based on Attention Mechanism and Dense Multiscale Features. *Laser Optoelectron. Prog.* **2022**, *59*, 0410006.
57. Huo, G.; Lin, D.; Yuan, M.; Yang, Z.; Niu, Y. Heterogeneous iris segmentation method based on modified U-Net. *J. Electron. Imaging* **2021**, *30*, 063015. [CrossRef]
58. Lozej, J.; Meden, B.; Struc, V.; Peer, P. End-to-End Iris Segmentation Using U-Net. In Proceedings of the IEEE IWOBI, San Carlos, Costa Rica, 18–20 July 2018; pp. 1–6.

59. Tobji, R.; Di, W.; Ayoub, N. FMnet: Iris Segmentation and Recognition by Using Fully and Multi-Scale CNN for Biometric Security. *Appl. Sci.* **2019**, *9*, 2042. [CrossRef]
60. Balasubramanian, S.K.; Jeganathan, V.; Subramani, T. Deep Learning-Based Iris Segmentation Algorithm for Effective Iris Recognition System. *Proc. Eng. Technol. Innov.* **2023**, *23*, 60–70. [CrossRef]
61. Wei, Y.; Zeng, A.; Zhang, X. RAG-Net: ResNet-50 attention gate network for accurate iris segmentation. *ET Image Process* **2022**, *16*, 3057–3066. [CrossRef]
62. Zhou, K.; Liu, Z.; Qiao, Y.; Xiang, T.; Loy, C.C. Domain Generalization: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 4396–4415. [CrossRef] [PubMed]
63. Arsalan, M.; Naqvi, R.A.; Kim, D.S.; Nguyen, P.H.; Owais, M.; Park, K.R. IrisDenseNet: Robust Iris Segmentation Using Densely Connected Fully Convolutional Networks in the Images by Visible Light and Near-Infrared Light Camera Sensors. *Sensors* **2018**, *18*, 1501. [CrossRef] [PubMed]
64. Tsai, Y.-H.; Hung, W.-C.; Schulter, S.; Sohn, K.; Yang, M.-H.; Chandraker, M. Learning to Adapt Structured Output Space for Semantic Segmentation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7472–7481.


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.